

# Les arbres isolés dans Open Street Map

Olivier Leroy

24 octobre 2018

## Jeux de données

Les données sont issues d'un import d'OSM via <https://www.geofabrik.de/>. L'import c'est fait dans une BD Postgresql (10) avec les extensions postgis, postgis\_topology et hstore.

## Création de la BD

La base à d'abord été créée : en psql, et connecté en postgres (sudo -i -u postgres)

```
CREATE DATABASE [sonpetitnom];
CREATE EXTENSION postgis;
CREATE EXTENSION postgis_topology;
CREATE EXTENSION hstore;
```

## import/peuplement de la BD

Puis importée via osm2pgsql

```
sudo -u postgres osm2pgsql -d [sonpetitnom] --slim -C 8000--hstore --multi-geometry
/home/lo82302h/Téléchargements/france-latest.osm.bz2 -W
```

-d indique la DB osmdbfrance

- --slim permet de sauver les données temporaires dans la DB plutot que dans la RAM

- --hstore indique qu'il y un type de données hstore (spécifique à PostgreSQL)

-C indique la RAM à utilisée

- --multi-geometry va créer des geometries multiples plutôt que de séparer en simple

-W demande le code ici de postgres

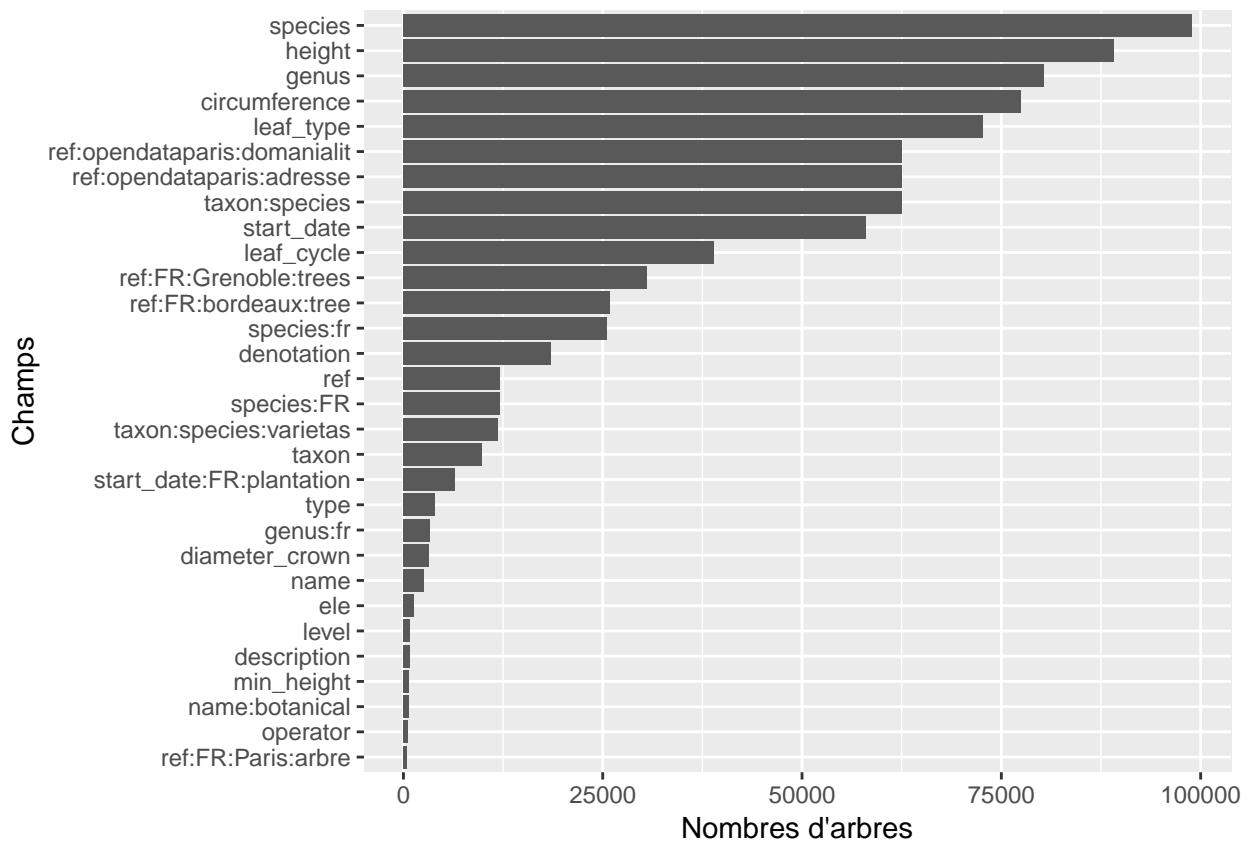
## Analyse exploratoire

### stats de base

Au moment de l'import (mi-octobre) il y avait 762167 lignes correspondant à des arbres isolés (natural=tree). Ils étaient renseignés par 27 champs (en plus de 'tags' et 'natural') complétés par 146 champs dans tags. Cela ne fait donc rien que pour les arbres isolés 173 champs.

Species est le champs le plus renseigné : 98874 valeurs non nulles, soit près de 13%.

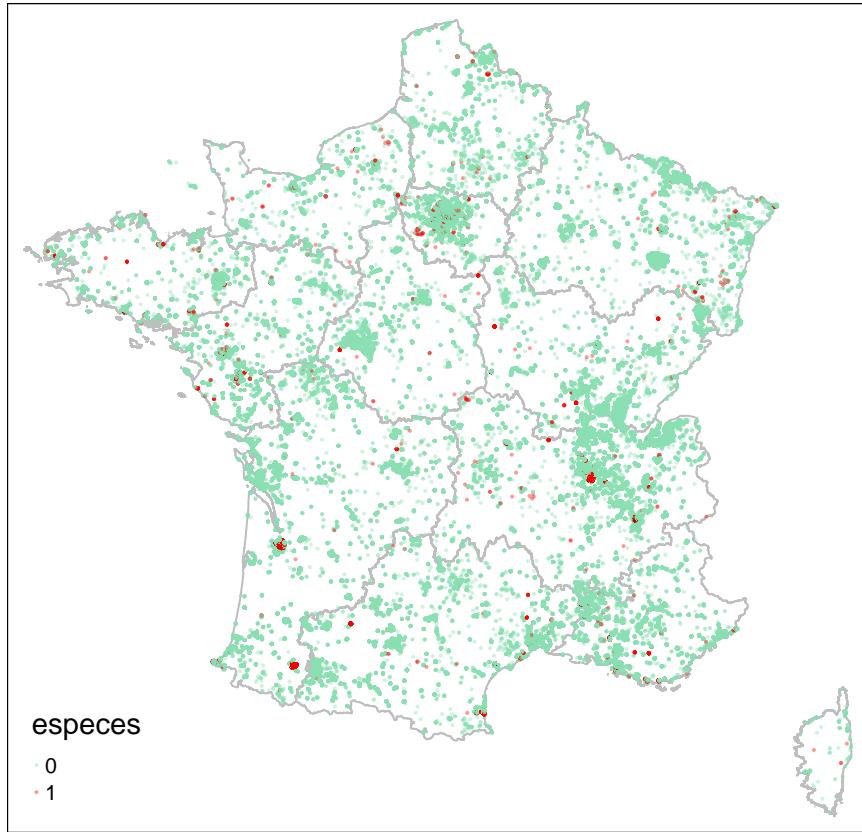
```
names_champs[1:30,] %>%
  ggplot( aes(x = reorder(champs, V1), y = V1)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  ylab("Nombres d'arbres") +
  xlab("Champs")
```



### Répartition du champ “species”

On a une répartition très urbaine des arbres isolés référencés dans OSM. Difficile à dire si il y a une logique sur ceux possédant une information de type “espèce”, il semble cependant qu'il y ait des “clusters”

```
tm_shape(st_simplify(st_geometry(france.shp)), dTolerance = 100) + # attention il a un simplify pour a
  tm_borders("grey") +
  tm_shape(species.shp) +
  tm_dots(alpha = 0.4, col = "especes", palette = c("#8be0b3", "red"))
```



Il y a 1813 valeurs différentes. Le wiki indique que “species” doit correspondre au nom latin (binomial).

On trouve plusieurs écart à cette règle :

- Les cas d'une information plus précise : le cultivar ou la variété de mentionné (avec plusieurs syntaxes possibles)
- Les cas de nomenclatures différentes : première lettre du descripteur
- Les erreurs :
  - Des plantes non ligneuses "iris versicolor" par exemple.
  - Uniquement le nom du genre : "platanus" / "Platanus"
  - Des noms vernaculaires, dans des langues variées : "Olivette blanche"
  - Des choses étranges : "OEIL"
  - etc...

### Prenons le cas des platanes

Pour savoir où en sont les botanistes j'ai utilisé TRNS

```
#on enlève la géométrie
species.dat <- st_set_geometry(species.shp, value = NULL)
# un premier grep pour se rapprocher
temp <- unique(species.dat$species)[grep(pattern = "atan", unique(species.dat$species))]
# on enlève les érables, tillieul et robinier
paste("on obtient", length(temp[grep(pattern = "^A|T|R", temp, invert = T)]), "descripteurs pour platanus")
## [1] "on obtient 32 descripteurs pour platanes"
species.dat %>%
  filter(species %in% temp[grep(pattern = "^A|T|R", temp, invert = T)]) %>%
```

```

group_by(species) %>%
summarize(comptage = n()) %>%
arrange(desc(comptage)) %>%
print(n = Inf)

## # A tibble: 32 x 2
##   species           comptage
##   <chr>              <int>
## 1 Platanus acerifolia      8737
## 2 Platanus x hispanica     5456
## 3 Platanus orientalis       243
## 4 Platanus x acerifolia     231
## 5 platanus                  193
## 6 Platanus platanor          79
## 7 Platane                      75
## 8 Platanus hispanica          71
## 9 Platanus ×hispanica          57
## 10 Platañes                     42
## 11 Platanus                      40
## 12 platane                      32
## 13 Platanus ×acerifolia          25
## 14 Platanus xacerifolia          13
## 15 Platanus orientalis Digitata      7
## 16 Platanus x acerfoilia          7
## 17 Platanus X acerifolia          7
## 18 Platanus X hispanica          7
## 19 Murier Platane                  6
## 20 Platanus occidentalis          6
## 21 Platanus orientalis L.          3
## 22 Platane à Feuille d'Érable        2
## 23 Platanus n. sp.                  2
## 24 platane commun                  1
## 25 platanoides                   1
## 26 Platanus × acerifolia          1
## 27 Platanus acerifolia            1
## 28 Platanus Orientalis            1
## 29 Platanus orientalis 'Digitata'    1
## 30 Platanus sp.                    1
## 31 Platanus x Hispánicus           1
## 32 platanus xacerifolia           1

```

Ces 32 descripteurs couvrent 15350 arbres. Avant de regarder la botanique en détail il y a plusieurs problèmes.

### Les noms de genre :

Le nom de genre devrait être mis dans l'attribut Key:genus. Selon le wiki osm pour ce dernier il doit être en latin et capitalisé. On retrouve dans cette catégorie : platanus, Platanus, platane, platanes, Platanus sp ou encore Platanus n. sp.

### les fautes de frappes

Elles sont de plusieurs types: \* Il y a l'absence de capitalisation du genre, la capitalisation du nom d'espèce et celle du croisement (moins tranché) \* La coquille : Platanus acerifolia

## les arbres de reveries et d'OSM

```
esp <- read.csv("nomsp_nomverma.csv", sep = "\t") # lecture du fichier

species.dat %>%
  filter(species %in% esp$Species) %>%
  group_by(species) %>%
  summarize(comptage = n()) %>%
  arrange(desc(comptage))

## # A tibble: 105 x 2
##   species           comptage
##   <chr>              <int>
## 1 Quercus robur      2902
## 2 Liquidambar styraciflua 2482
## 3 Fraxinus excelsior 2061
## 4 Celtis australis   1999
## 5 Tilia cordata      1990
## 6 Aesculus hippocastanum 1947
## 7 Carpinus betulus    1937
## 8 Acer platanoides    1858
## 9 Acer pseudoplatanus 1839
## 10 Betula pendula     1328
## # ... with 95 more rows
```