

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 plutôt que dans la RAM
- -hstore indique qu'il y a un type de données hstore (spécifique à PostgreSQL)
- C indique la RAM à utiliser
- -multi-geometry va créer des géométries 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 champ 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")
```

