

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. Pour avoir les noms des contributeurs et la date il faut s'identifier sur le site.

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 --cache 10000 --number-processes 2 --hstore -x --multi
```

-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)

- -x ajoute les infos extras comme le contributeurs

-C -cache indique la RAM à utilisée

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

-W demande le code ici de postgres

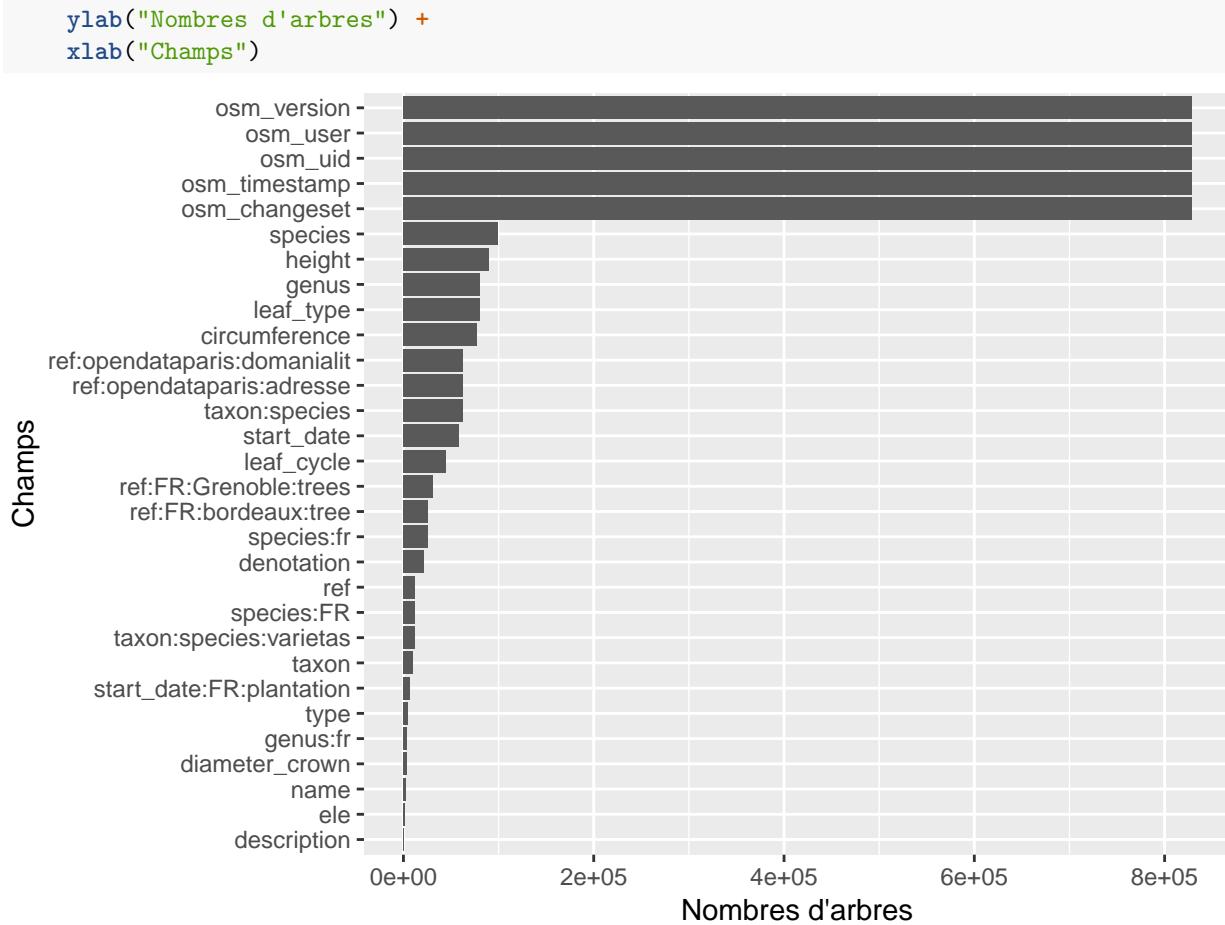
Analyse exploratoire

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 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()
```



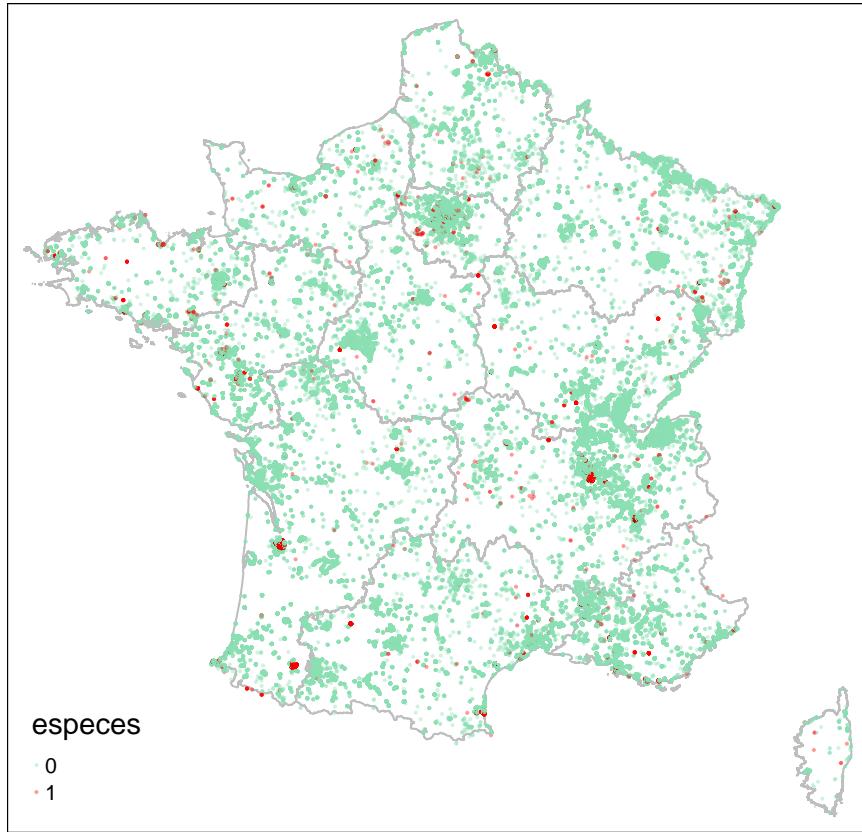
Répartition du champ “species”

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 33 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: 33 x 2
##   species           comptage
##   <chr>              <int>
## 1 Platanus acerifolia      8738
## 2 Platanus x hispanica     5482
## 3 Platanus orientalis       243
## 4 Platanus x acerifolia     231
## 5 platanus                  193
## 6 Platanus platanor          79
## 7 Platane                      75
## 8 Platanus hispanica          72
## 9 platane                     59
## 10 Platanus ×hispanica        57
## 11 Platanes                   42
## 12 Platanus                   40
## 13 Platanus ×acerifolia        25
## 14 Platanus xacerifolia        13
## 15 Platanus × hispanica        10
## 16 Platanus orientalis Digitata    7
## 17 Platanus x acerfoilia        7
## 18 Platanus X acerifolia        7
## 19 Platanus X hispanica        7
## 20 Murier Platane                 6
## 21 Platanus occidentalis         6
## 22 Platanus orientalis L.          3
## 23 Platane à Feuille d'Érable      2
## 24 Platanus n. sp.                  2
## 25 platane commun                  1
## 26 platanoides                   1
## 27 Platanus × acerifolia         1
## 28 Platanus acerifolia            1
## 29 Platanus Orientalis             1
## 30 Platanus orientalis 'Digitata'  1
## 31 Platanus sp.                      1
## 32 Platanus x Hispnicus            1
## 33 platanus xacerifolia            1

```

Ces 32 descripteurs couvrent 15350 arbres. Avant de regarder la botanique en detail 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 categorie : platanus, Platanus, platane, platanes, Platanus sp ou encore Platanus n. sp.

les fautes de frappes

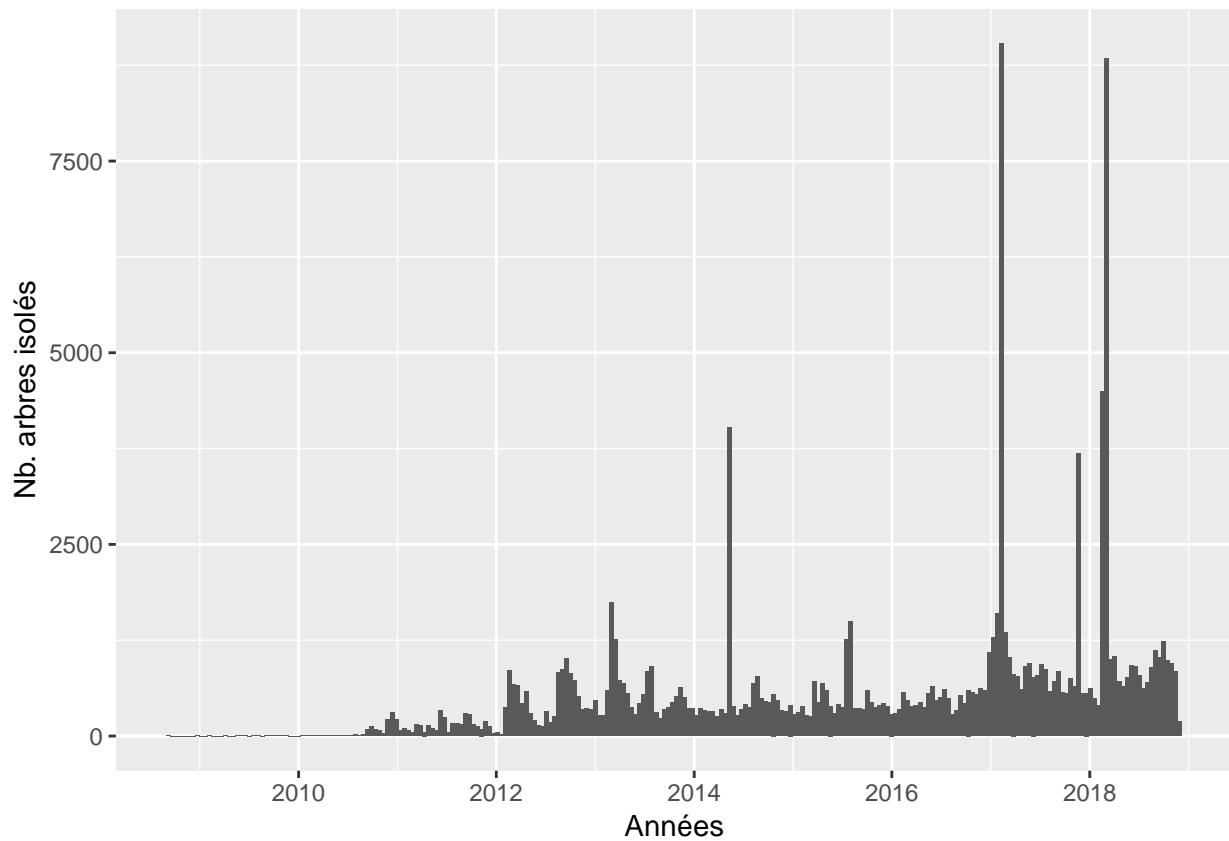
Elles sont des plusieurs types: * Il y a l'absence de capitalisation du genre, la capitalisation du nom d'espèce et celle du croisement (moins tranché) * La coquille : Platanus acerifolia

Les débats botaniques

Il y aussi les noms latins non tranchés chez les botanistes, les problèmes avec les croisements, les cultivars et variétés mis au niveau de l'espèce, etc.

Utilisateurs et source de données

```
# import des contributeurs et du moment de la contribution
user.dat <- dbGetQuery(con, "SELECT DISTINCT tags -> 'osm_timestamp' AS ts, tags -> 'osm_user' AS user
FROM planet_osm_point
WHERE planet_osm_point.natural = 'tree';")
user.dat$ts <- as_date(user.dat$ts) # on passe en POSIX juste date
# j'ai pris la 15aine mais on est presque de l'ordre du jour
ggplot(user.dat, aes(x = ts)) +
  geom_histogram(binwidth = 15) +
  xlab("Années") +
  ylab("Nb. arbres isolés")
```



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)) %>%
print(n = Inf)

## # A tibble: 107 x 2
##   species           comptage
##   <chr>              <int>
## 1 Quercus robur      2889
## 2 Liquidambar styraciflua 2482
## 3 Fraxinus excelsior    2061
## 4 Celtis australis     2000
## 5 Tilia cordata       1991
## 6 Aesculus hippocastanum 1961
## 7 Carpinus betulus     1939
## 8 Acer platanoides     1872
## 9 Acer pseudoplatanus    1838
## 10 Prunus serrulata     1328
## 11 Betula pendula       1313
## 12 Tilia platyphyllos    1282
## 13 Acer saccharinum     1236
## 14 Liriodendron tulipifera 1152
## 15 Quercus cerris       1107
## 16 Magnolia grandiflora    1101
## 17 Fraxinus angustifolia 1024
## 18 Sophora japonica      979
## 19 Acer campestre        897
## 20 Populus nigra        759
## 21 Quercus palustris      747
## 22 Robinia pseudoacacia  734
## 23 Cercis siliquastrum    633
## 24 Koelreuteria paniculata 583
## 25 Gleditsia triacanthos    566
## 26 Olea europaea        517
## 27 Quercus rubra        513
## 28 Acer negundo         490
## 29 Quercus ilex          467
## 30 Alnus cordata        466
## 31 Betula utilis          451
## 32 Alnus glutinosa        410
## 33 Fagus sylvatica       379
## 34 Ginkgo biloba         376
## 35 Morus alba          350
## 36 Fraxinus ornus        311
## 37 Prunus avium          238
## 38 Populus alba          195
## 39 Buxus sempervirens     182
## 40 Juglans nigra        175
## 41 Corylus avellana       173
## 42 Ulmus minor          162
## 43 Quercus petraea        139
## 44 Prunus padus          137
## 45 Prunus laurocerasus     133
## 46 Acer monspessulanum     124
## 47 Ailanthus altissima      123
## 48 Castanea sativa        106

```

## 49	<i>Laurus nobilis</i>	101
## 50	<i>Albizia julibrissin</i>	98
## 51	<i>Salix alba</i>	94
## 52	<i>Ilex aquifolium</i>	86
## 53	<i>Platanus hispanica</i>	72
## 54	<i>Quercus pubescens</i>	59
## 55	<i>Broussonetia papyrifera</i>	57
## 56	<i>Ostrya carpinifolia</i>	54
## 57	<i>Punica granatum</i>	52
## 58	<i>Melia azedarach</i>	50
## 59	<i>Betula pubescens</i>	49
## 60	<i>Ficus carica</i>	48
## 61	<i>Juglans regia</i>	48
## 62	<i>Sorbus aucuparia</i>	46
## 63	<i>Morus nigra</i>	41
## 64	<i>Nerium oleander</i>	41
## 65	<i>Sorbus aria</i>	39
## 66	<i>Eriobotrya japonica</i>	33
## 67	<i>Sorbus domestica</i>	30
## 68	<i>Arbutus unedo</i>	28
## 69	<i>Crataegus monogyna</i>	27
## 70	<i>Ligustrum vulgare</i>	26
## 71	<i>Populus tremula</i>	26
## 72	<i>Sorbus torminalis</i>	23
## 73	<i>Mespilus germanica</i>	18
## 74	<i>Syringa vulgaris</i>	17
## 75	<i>Acer opalus</i>	16
## 76	<i>Cotinus coggygria</i>	14
## 77	<i>Prunus dulcis</i>	13
## 78	<i>Cornus mas</i>	12
## 79	<i>Prunus spinosa</i>	12
## 80	<i>Crataegus laevigata</i>	11
## 81	<i>Pittosporum tobira</i>	11
## 82	<i>Prunus serotina</i>	11
## 83	<i>Pyrus nivalis</i>	11
## 84	<i>Diospyros kaki</i>	10
## 85	<i>Salix caprea</i>	9
## 86	<i>Tilia europaea</i>	9
## 87	<i>Photinia serrulata</i>	8
## 88	<i>Rhus typhina</i>	8
## 89	<i>Pistacia lentiscus</i>	6
## 90	<i>Sambucus nigra</i>	6
## 91	<i>Laburnum anagyroides</i>	4
## 92	<i>Populus trichocarpa</i>	4
## 93	<i>Cornus sanguinea</i>	3
## 94	<i>Euonymus europaeus</i>	3
## 95	<i>Phillyrea angustifolia</i>	3
## 96	<i>Sorbus latifolia</i>	3
## 97	<i>Elaeagnus angustifolia</i>	2
## 98	<i>Juniperus oxycedrus</i>	2
## 99	<i>Crataegus azarolus</i>	1
## 100	<i>Malus sylvestris</i>	1
## 101	<i>Quercus coccifera</i>	1
## 102	<i>Rhamnus alaternus</i>	1

## 103 Ruscus aculeatus	1
## 104 Salix fragilis	1
## 105 Viburnum opulus	1
## 106 Viburnum tinus	1
## 107 Vitex agnus-castus	1