



Data Location Quality at GBIF

John Waller

ZOOLOGICAL MUSEUM, COPENHAGEN | 30 SEPTEMBER 2019



Known issues

A part of the GBIF processing is to flag occurrences that have suspicious fields

73,187,457	Country derived from coordinates	30,703,810	Identified date invalid	25,092,657	Individual count invalid		
24,830,645	Taxon match higherrank	23,154,770	Depth non numeric	21,324,799	Elevation non numeric		
19,669,424	References uri invalid	19,207,110	Continent invalid	16,692,087	Coordinate precision invalid		
13,945,585	Geodetic datum invalid	7,975,303	Recorded date invalid	7,431,202	Taxon match none	7,194,473	Taxon match fuzzy
7,060,470	Recorded date mismatch	6,930,309	Type status invalid	5,851,867	Country invalid	4,121,996	Identified date unlikely
4,054,825	Modified date invalid	3,338,055	Elevation min/max swapped	2,397,674	Country coordinate mismatch		
1,902,925	Coordinate uncertainty meters invalid	1,462,159	Recorded date unlikely	1,199,976	Basis of record invalid		
1,190,450	Zero coordinate	1,050,891	Multimedia uri invalid	964,428	Coordinate invalid	361,550	Depth min/max swapped
347,519	Presumed swapped coordinate	315,923	Presumed negated longitude	185,996	Country mismatch		
157,902	Modified date unlikely	143,347	Coordinate out of range	85,880	Presumed negated latitude	85,699	Depth unlikely
42,924	Depth not metric	42,861	Coordinate reprojection suspicious	27,831	Elevation unlikely	5,796	Elevation not metric
2,330	Multimedia date invalid	495	Coordinate reprojection failed				

Fossils

There are fossils among your results. That can mean species occurrences at unexpected locations

Living specimens

Your search includes living specimens such as occurrences in botanical and zoological gardens.

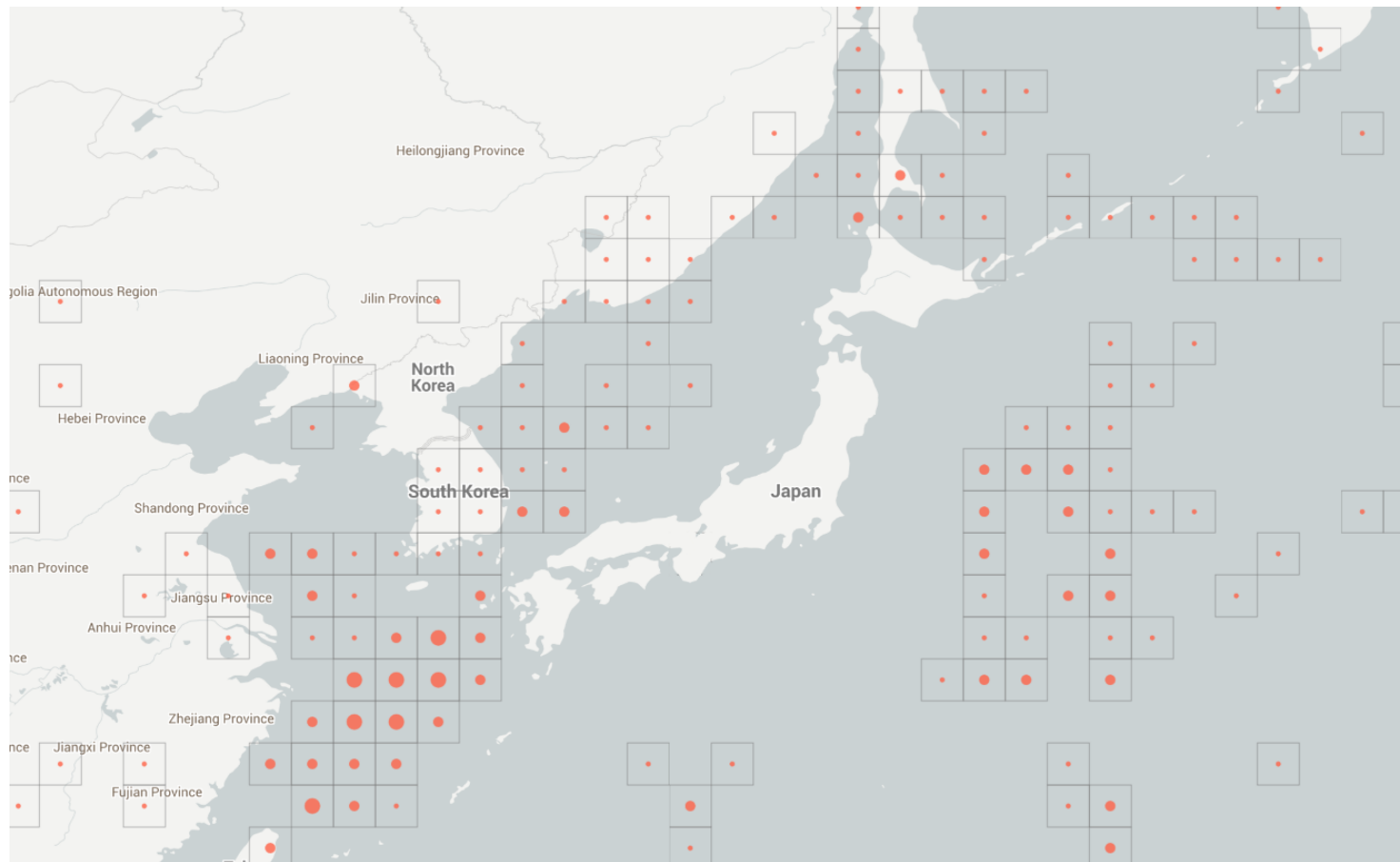


GBIF

Global Biodiversity
Information Facility

2,397,674

Country coordinate mismatch

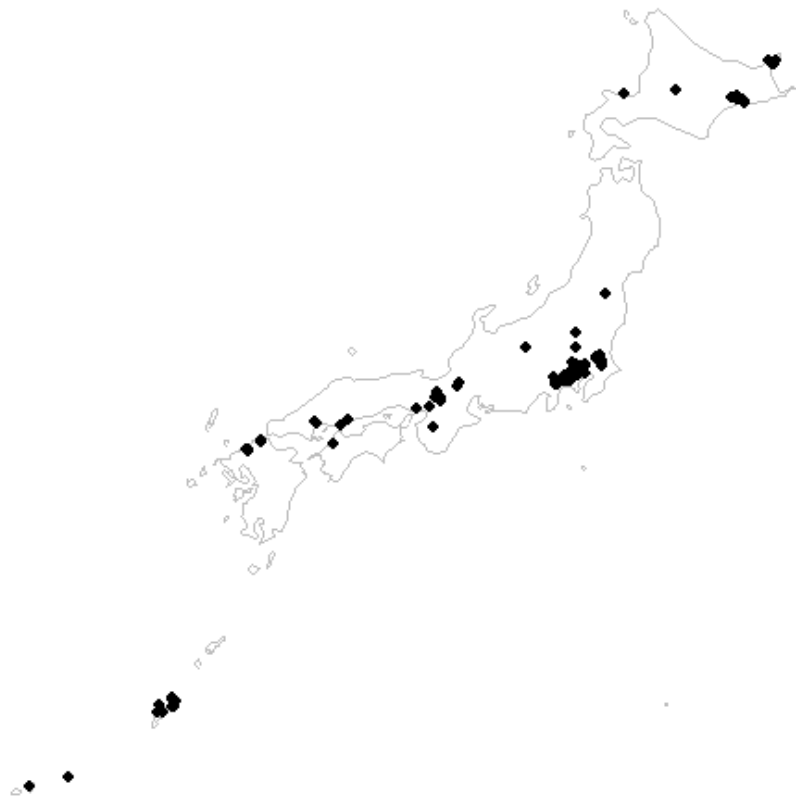


1,190,450

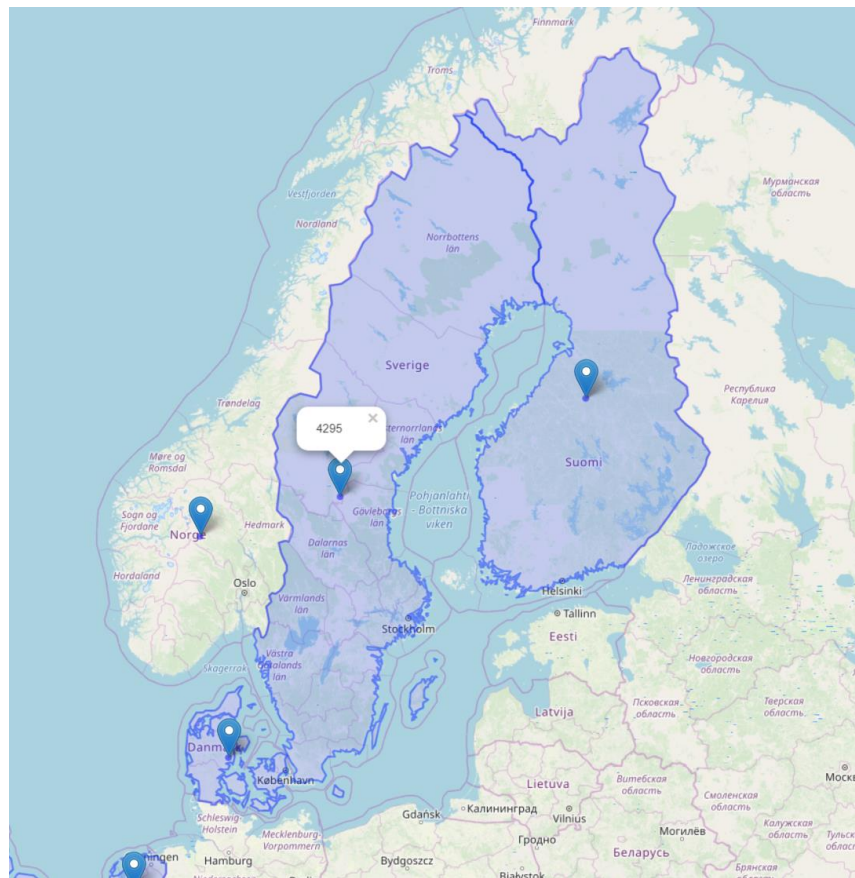
Zero coordinate

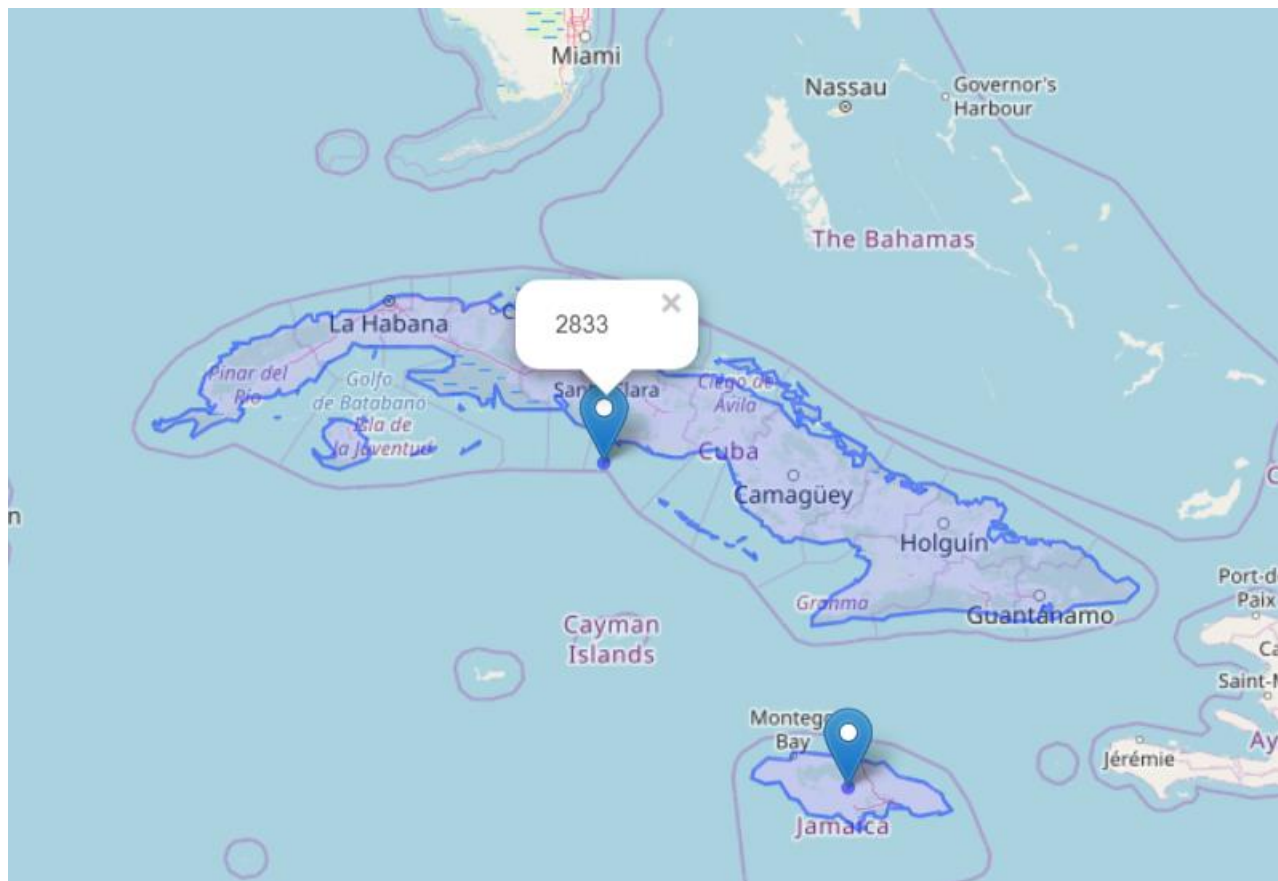


New Flags in Development...

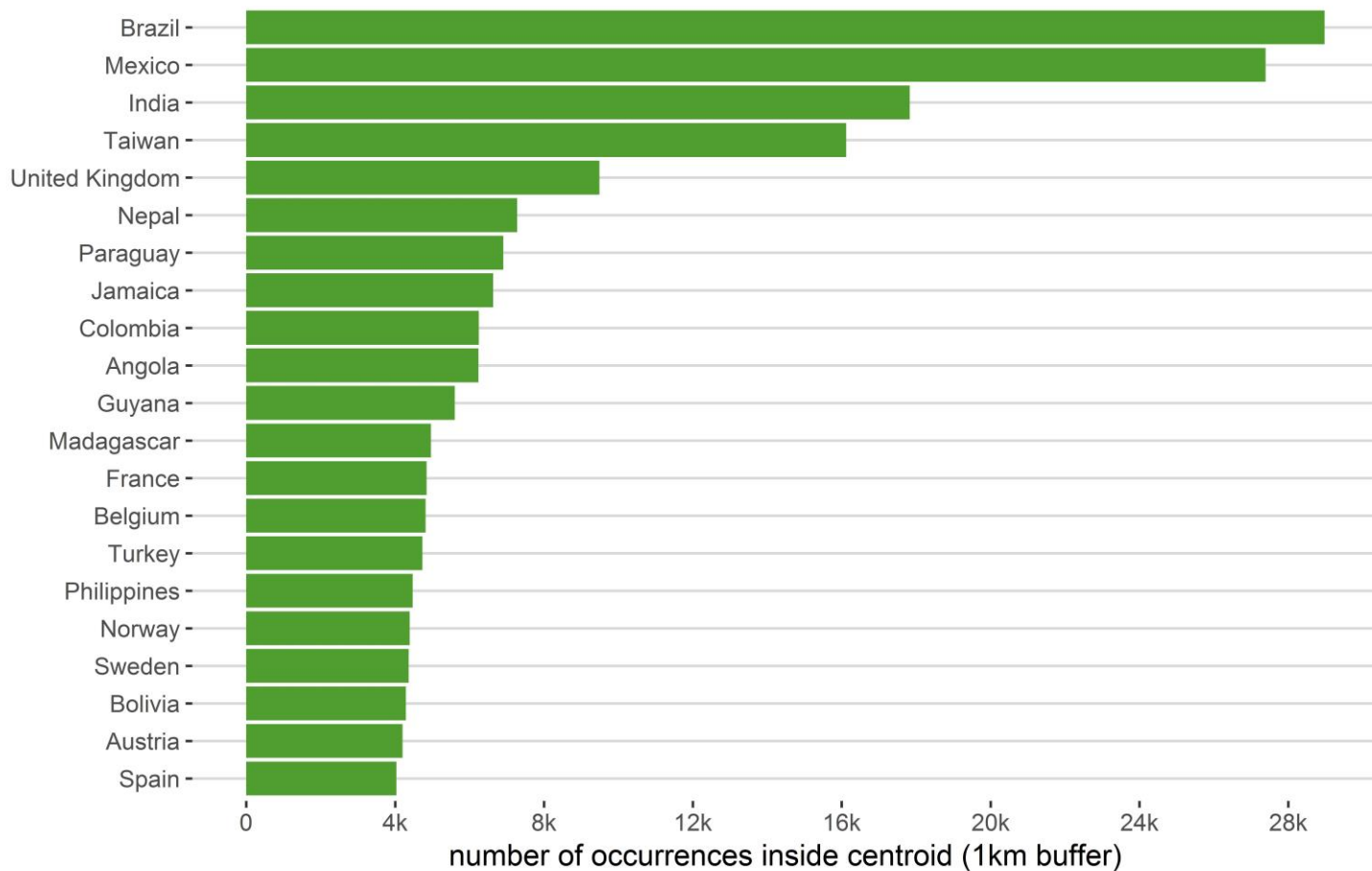








Top Country Centroids





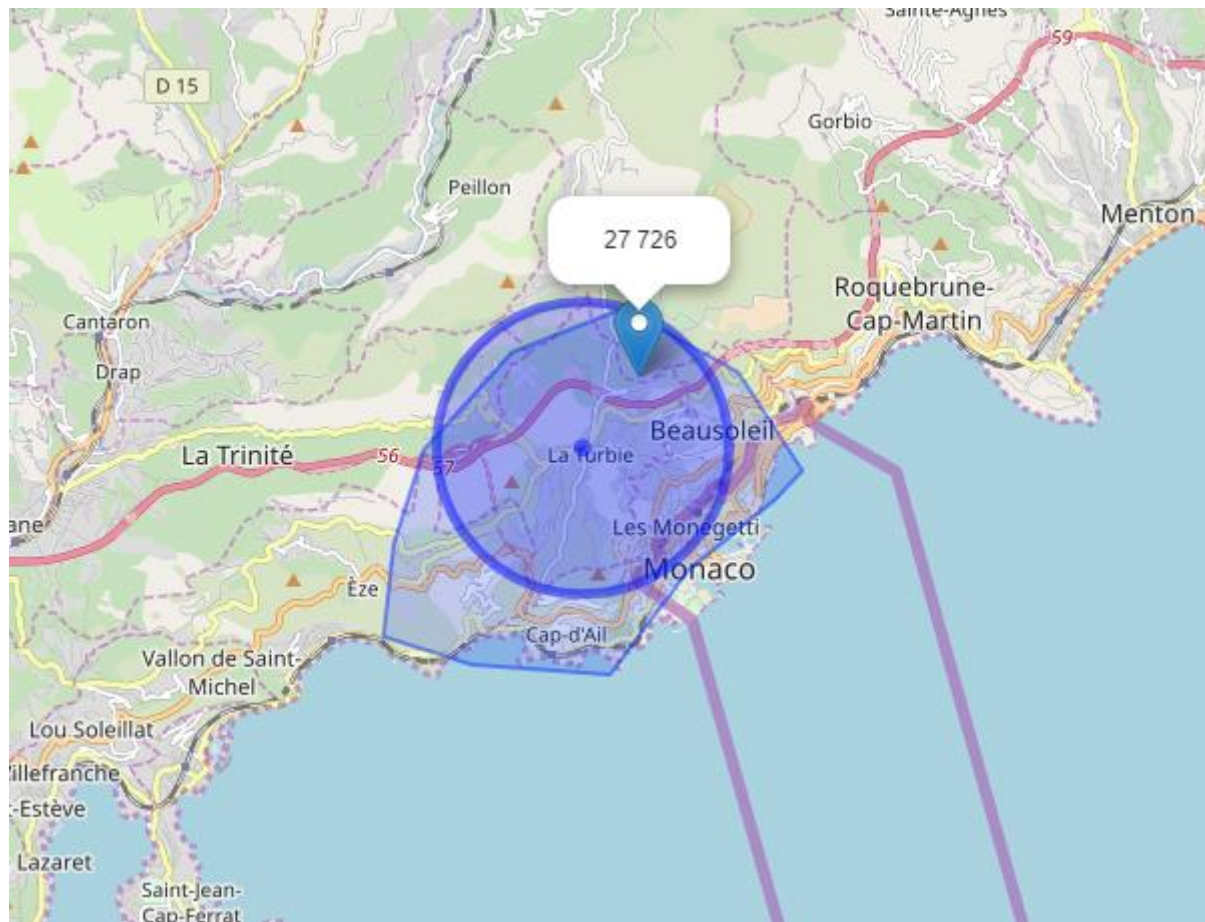
GBIF

Global Biodiversity
Information Facility

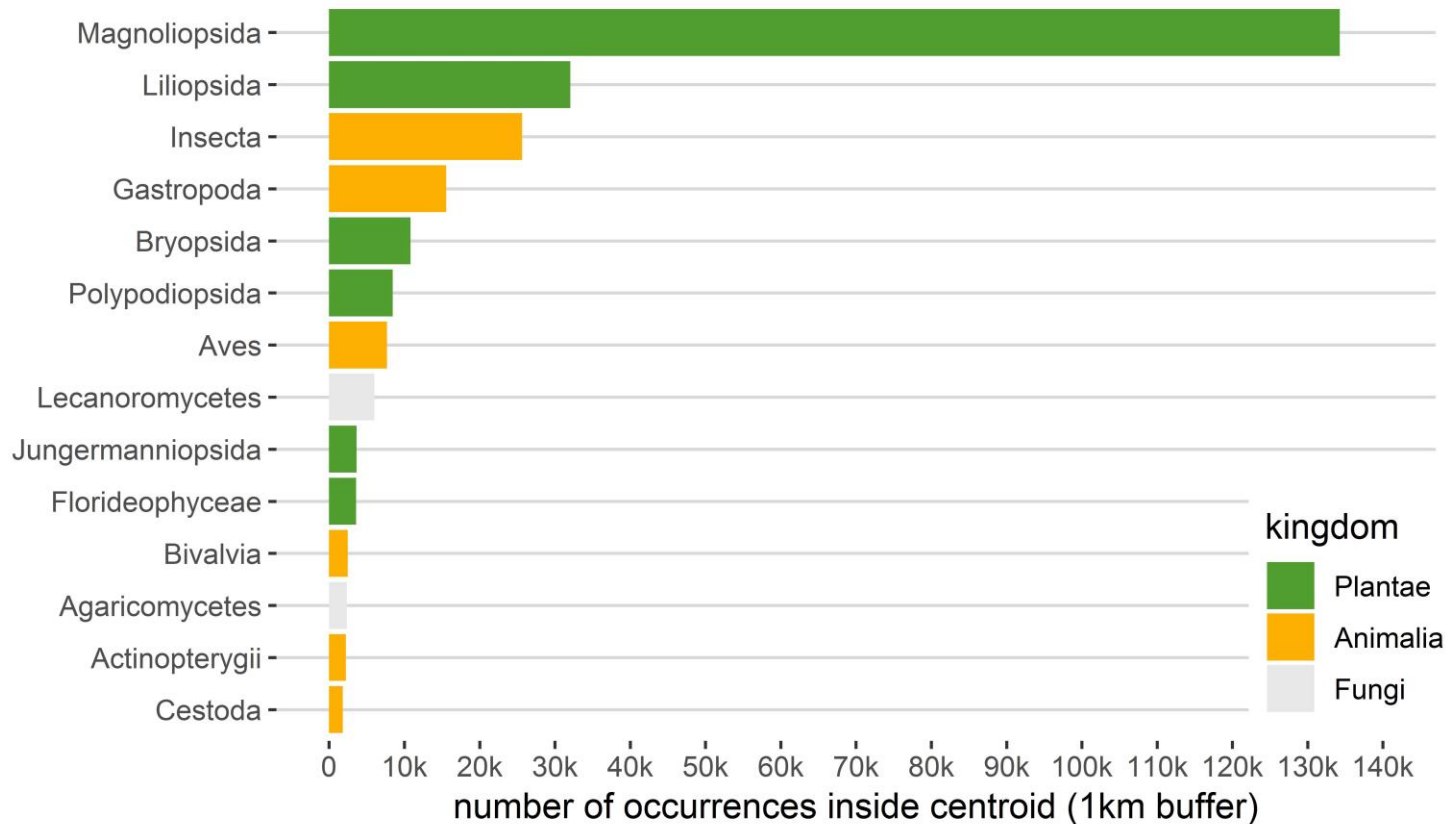
States/Provinces also have centroids



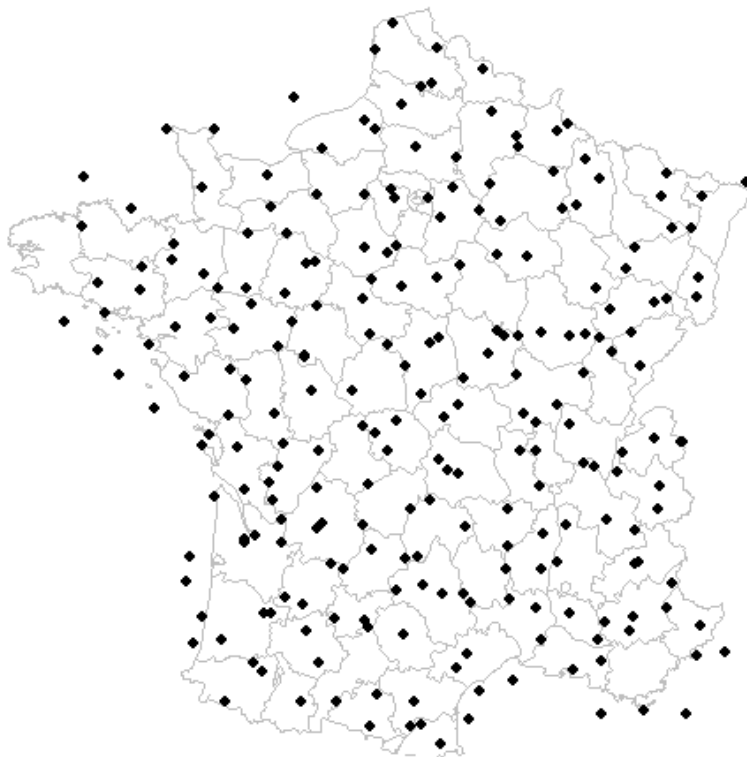
Some areas are too small to flag



Mostly plants in centroids

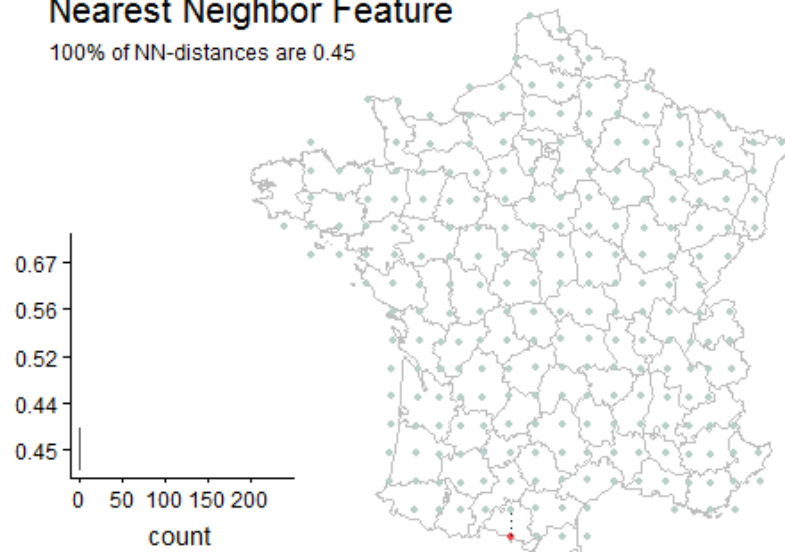


Gridded Datasets



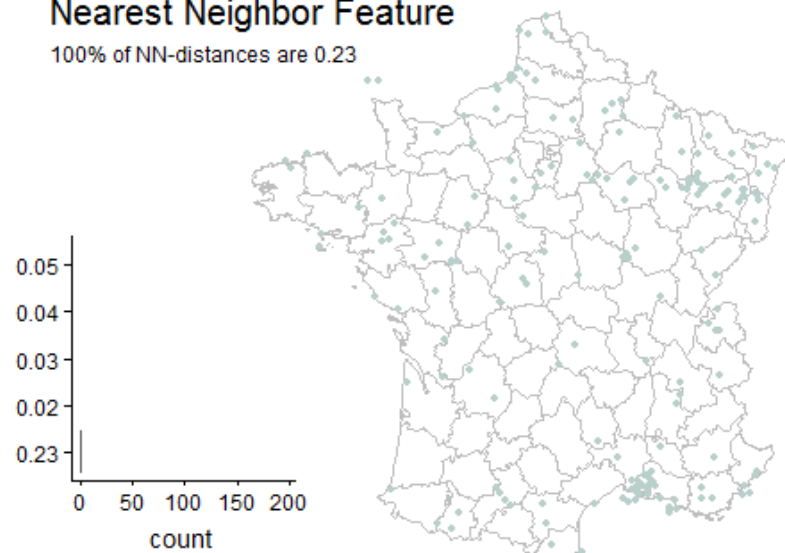
Nearest Neighbor Feature

100% of NN-distances are 0.45



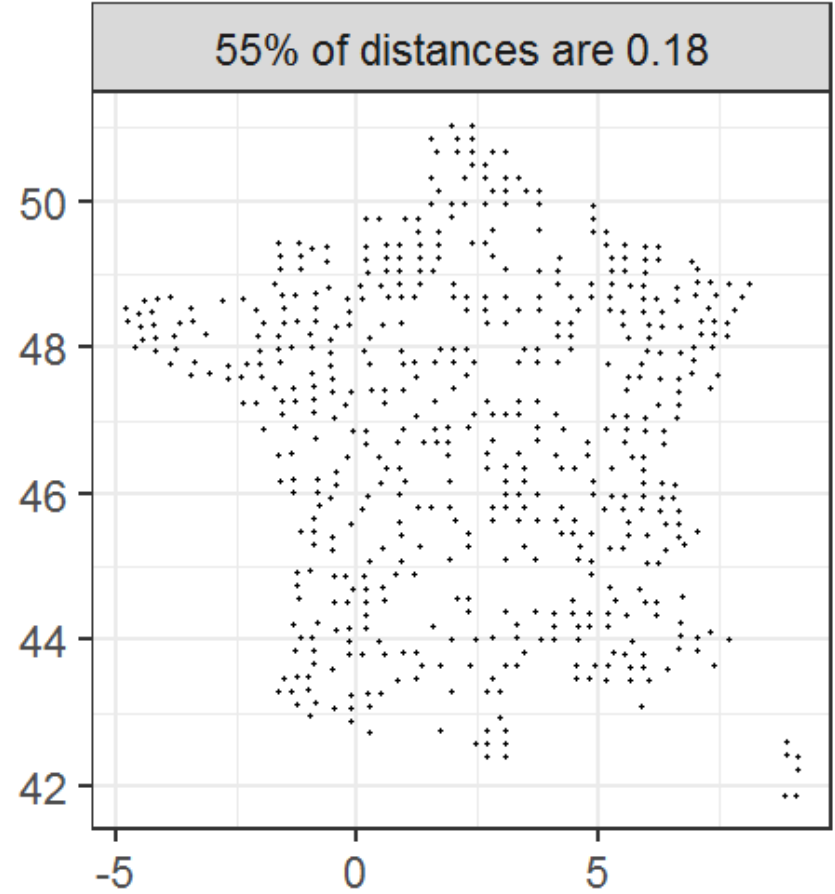
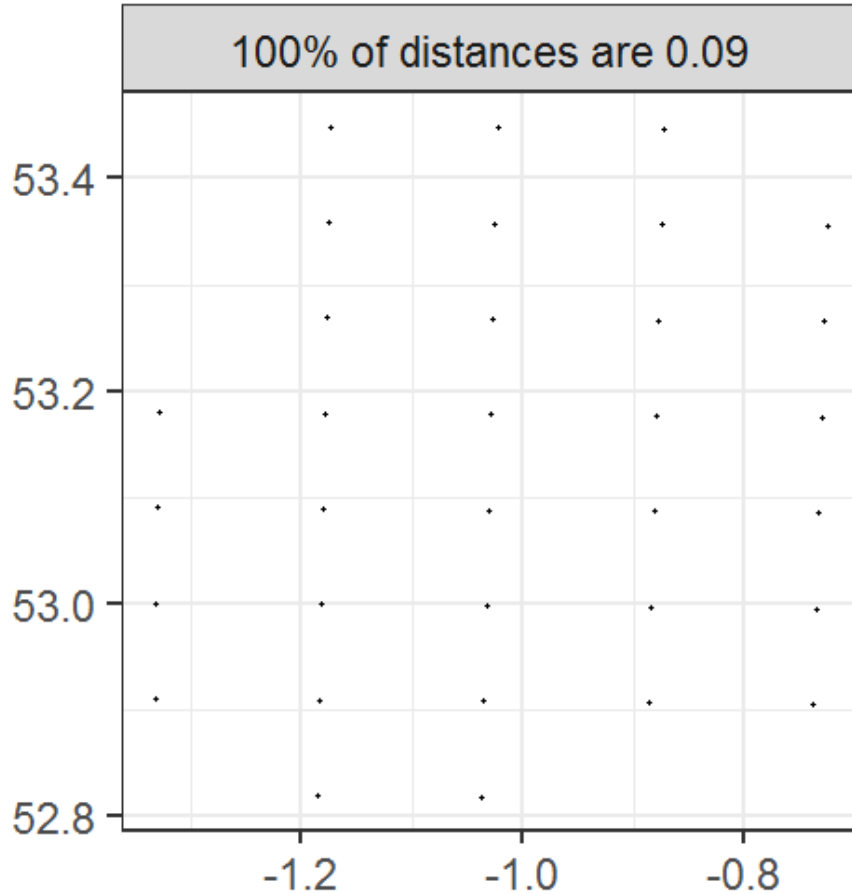
Nearest Neighbor Feature

100% of NN-distances are 0.23

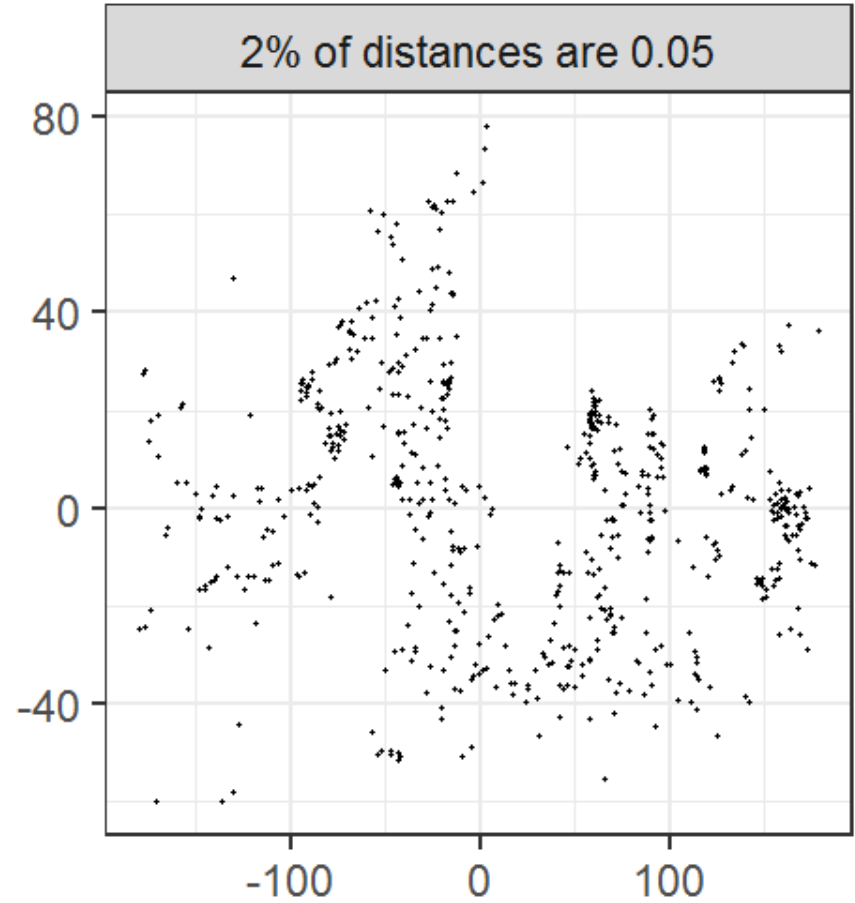
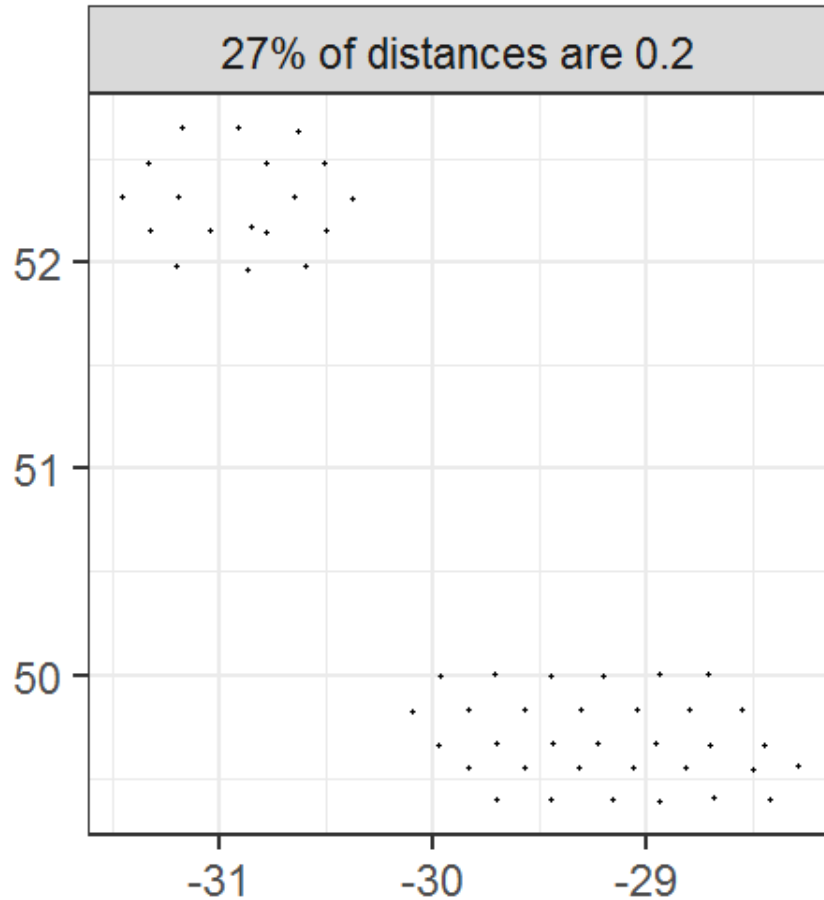


**GBIF**Global Biodiversity
Information Facility

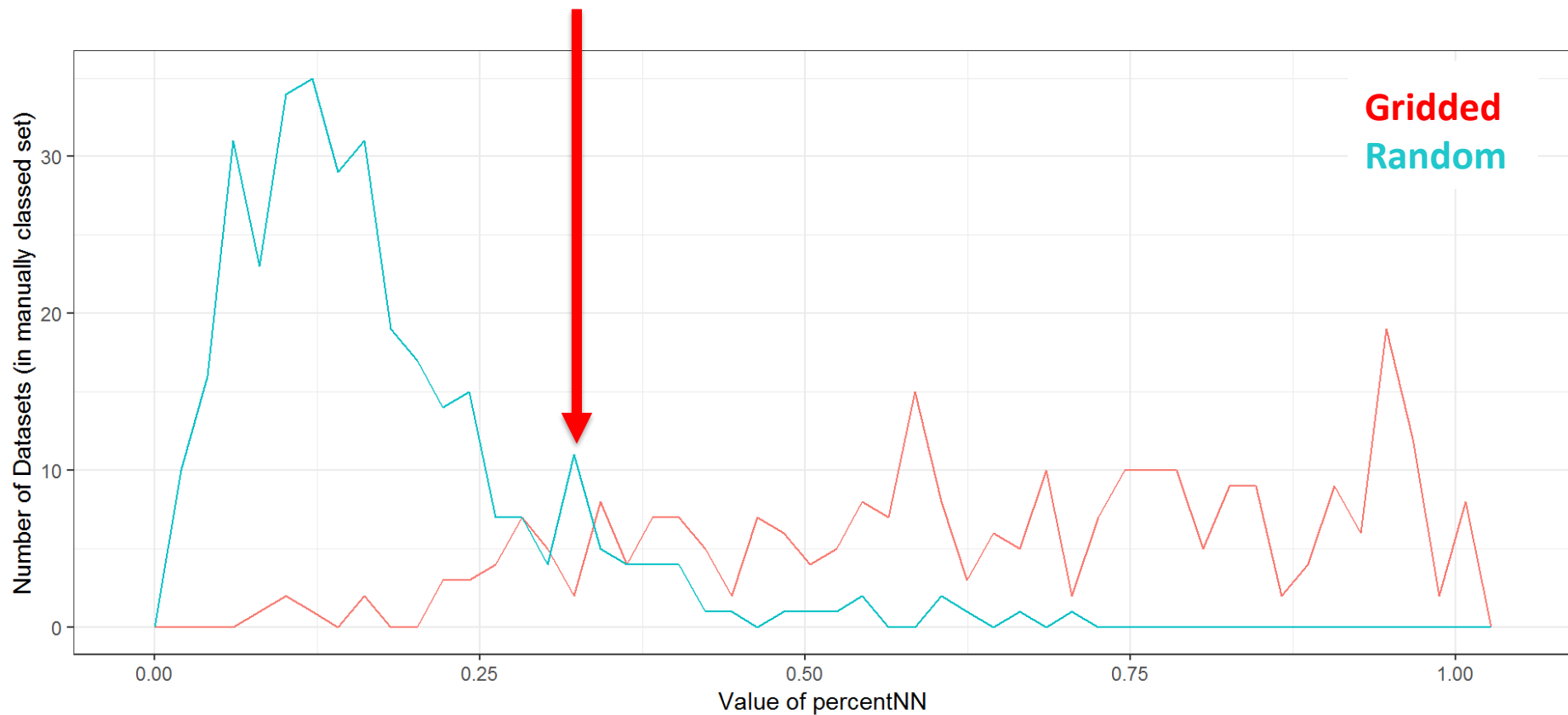
Spectrum of “griddyness”



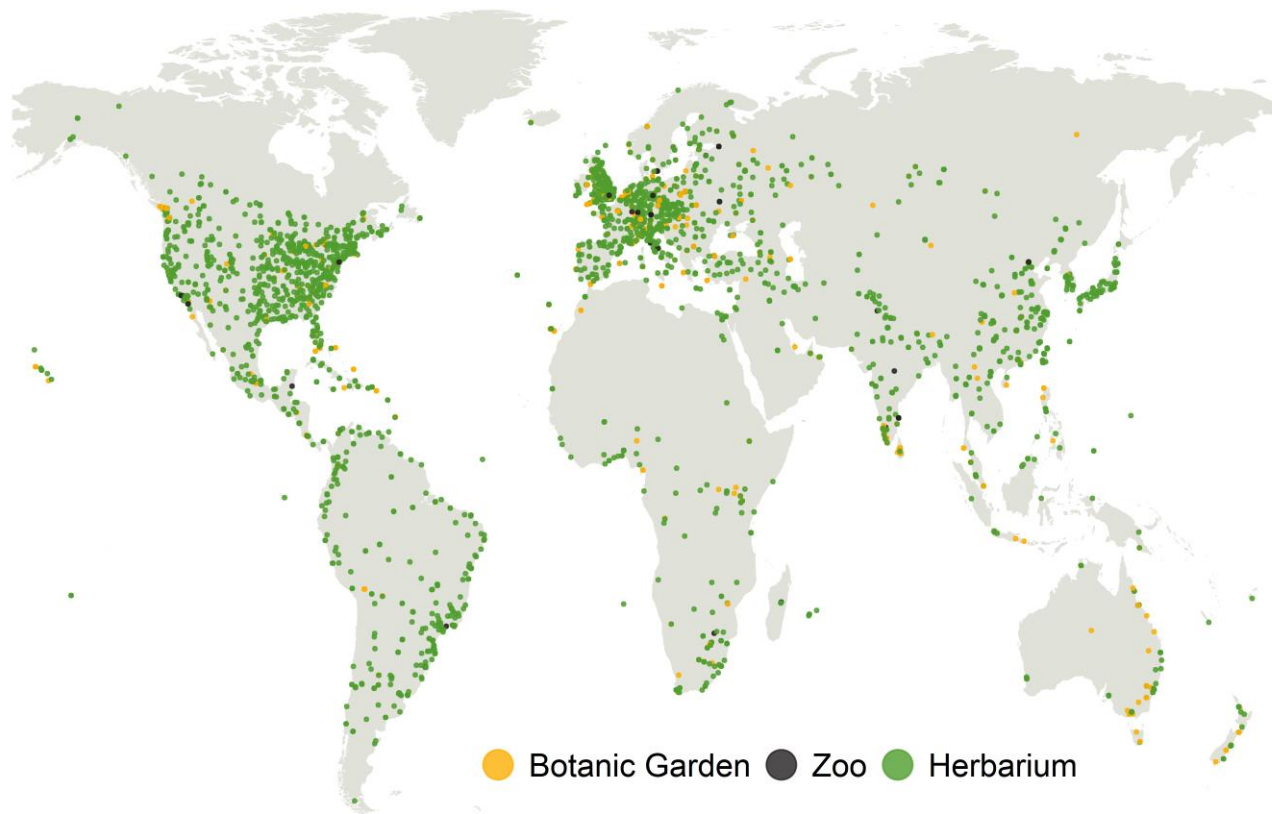
Spectrum of “griddyness”



30% good cut off point



- There are around **400** gridded datasets within GBIF.
- **France publishes the most gridded datasets** with 228 gridded datasets.
- Not all rasterized or gridded datasets are a perfect square and there is a spectrum of griddyness.
- At least **75M** occurrence records are gridded above 0.01 degrees.
- Most datasets (370) are **gridded at 0.1°** and only **12 datasets are gridded at greater than 0.40°** and 33 datasets are gridded **below 0.1°**.



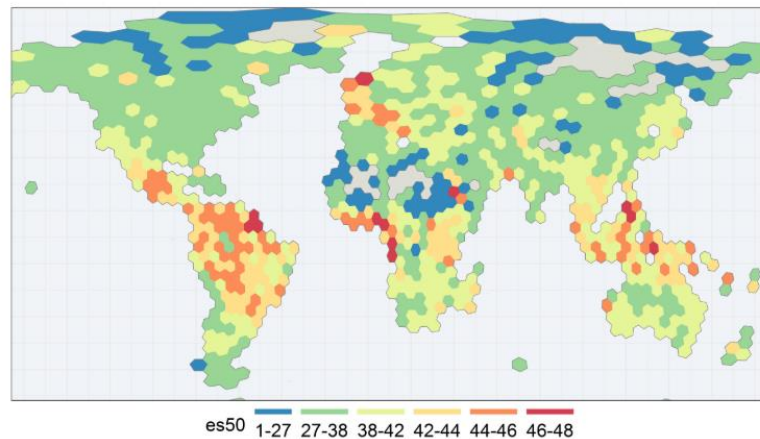


<https://data-blog.gbif.org>

Exploring es50 for GBIF

John Waller

2019-07-10 · GBIF



[jpg](#) | [pdf](#) | [svg](#) | [code](#)

It has been suggested that GBIF could make **es50** maps similar to what organizations like **OBIS** are already doing. I decided to make one for **land animals** (graph above). [link to code](#)

es50 (Hulbert index) is the statistically expected number of unique species in a random sample of 50 occurrence records, and is an indicator of biodiversity richness. The score can be computed without random sampling, but the mean of infinite random sampling will produce the same result.

[Obis Definition here](#)

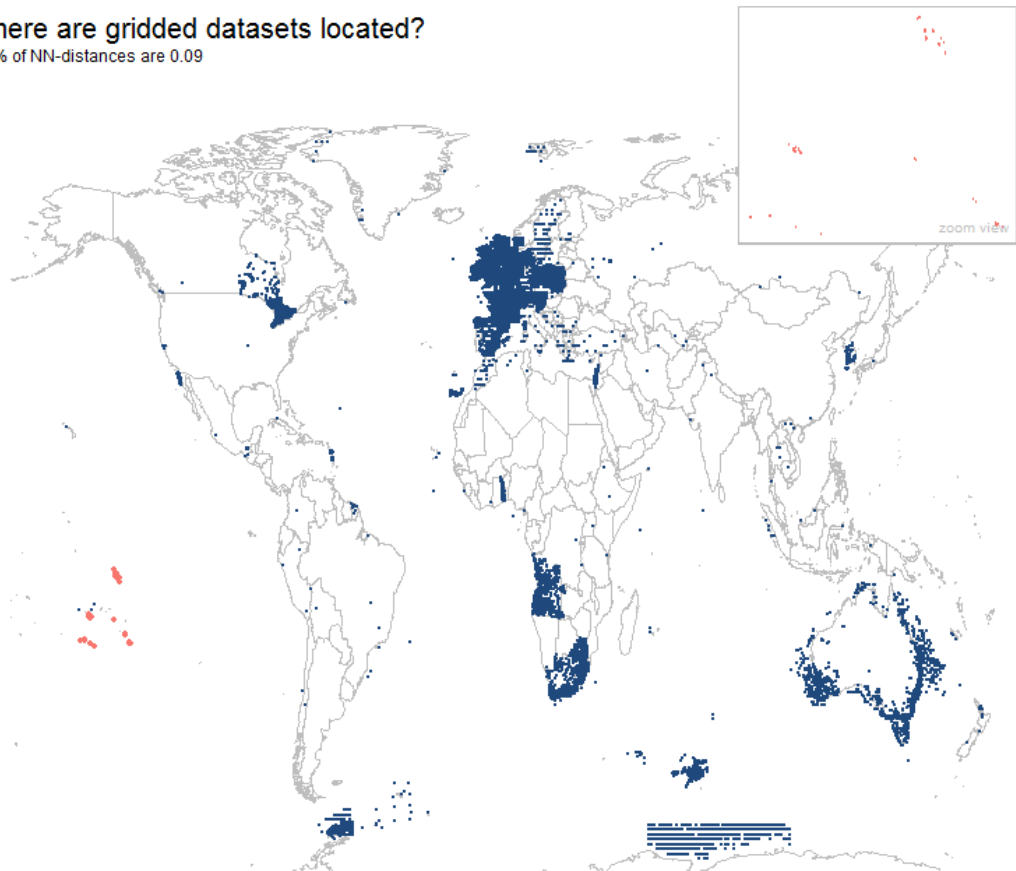
Dataset

Centroids Records

1. Natural History Museum Collection Specimens	77 K
2. The Vascular Plant Collection at the Botanische Staatssammlung München	34 K
3. Geneva Herbarium – General Collection (G)	21 K
4. Geographically tagged INSDC sequences	15 K
5. MAL	14 K
6. Geneva Herbarium – De Candolle's Prodromus	12 K
7. A global database for the distributions of crop wild relatives	7 K
8. Field Museum of Natural History Botany Seed Plant Collection	5 K
9. Botanical Museum, Copenhagen, the Phycology Herbarium	4 K
10. Field Museum of Natural History (Botany) Lichen Collection	4 K
11. National Museum of Natural History Luxembourg	4 K
12. Norwegian Biodiversity Information Centre - Other datasets	3 K
13. United States National Plant Germplasm System Collection	2 K
14. The Fungal Collection at the Senckenberg Museum für Naturkunde Görlitz	2 K
15. Field Museum of Natural History (Botany) Bryophyte Collection	2 K

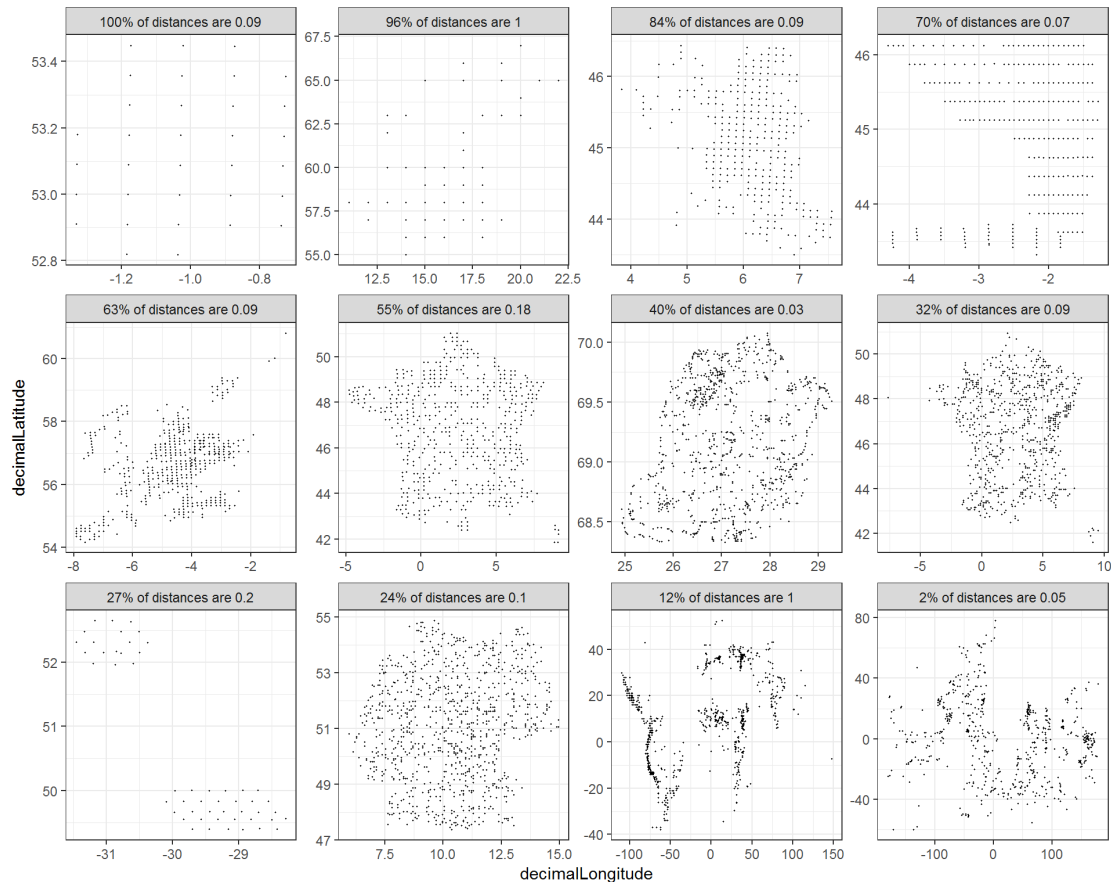
Where are gridded datasets located?

61% of NN-distances are 0.09





Mostly plants in centroids





Coordinate Rounded

Location

Term	Interpreted	Original	Remarks
Coordinate uncertainty in metres	5	5	
Country or area	New Zealand	28.523457182°N, 80.683094159°W	"HEY, CHECK OUT THIS SPECIFIC SAND GRAIN!"
Country code	NZ		
Decimal latitude	-42.426067	-42.4260667758	Coordinate rounded
Decimal longitude	173.680679	173.6806788483	Coordinate rounded
Geodetic datum	WGS84		Geodetic datum assumed WGS84
State province	Canterbury	Canterbury	
Verbatim locality	Pacific Ocean, , , NZ	Pacific Ocean, , , NZ	

