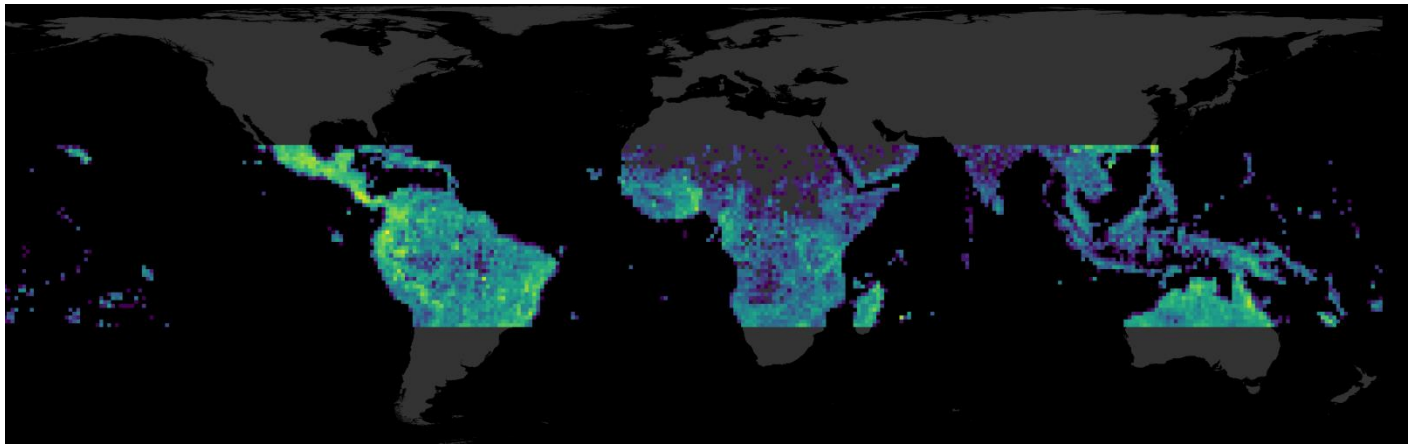
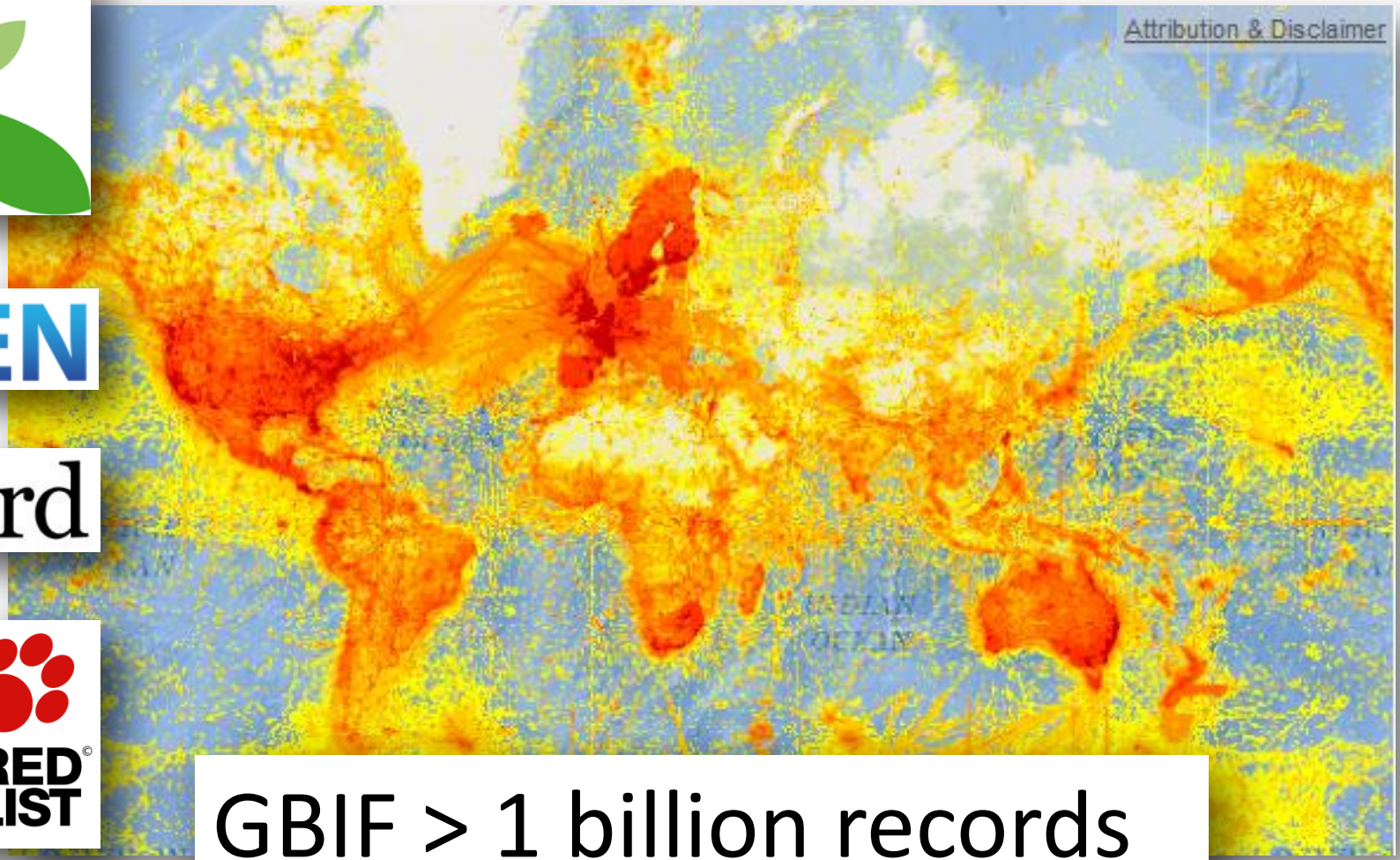
Biodiversity Data in Ecology and Biogeography



Alexandre Antonelli & Alexander Zizka

Species distribution data – a revolution



GBIF

BIEN

eBird

IUCN RED LIST

GBIF > 1 billion records

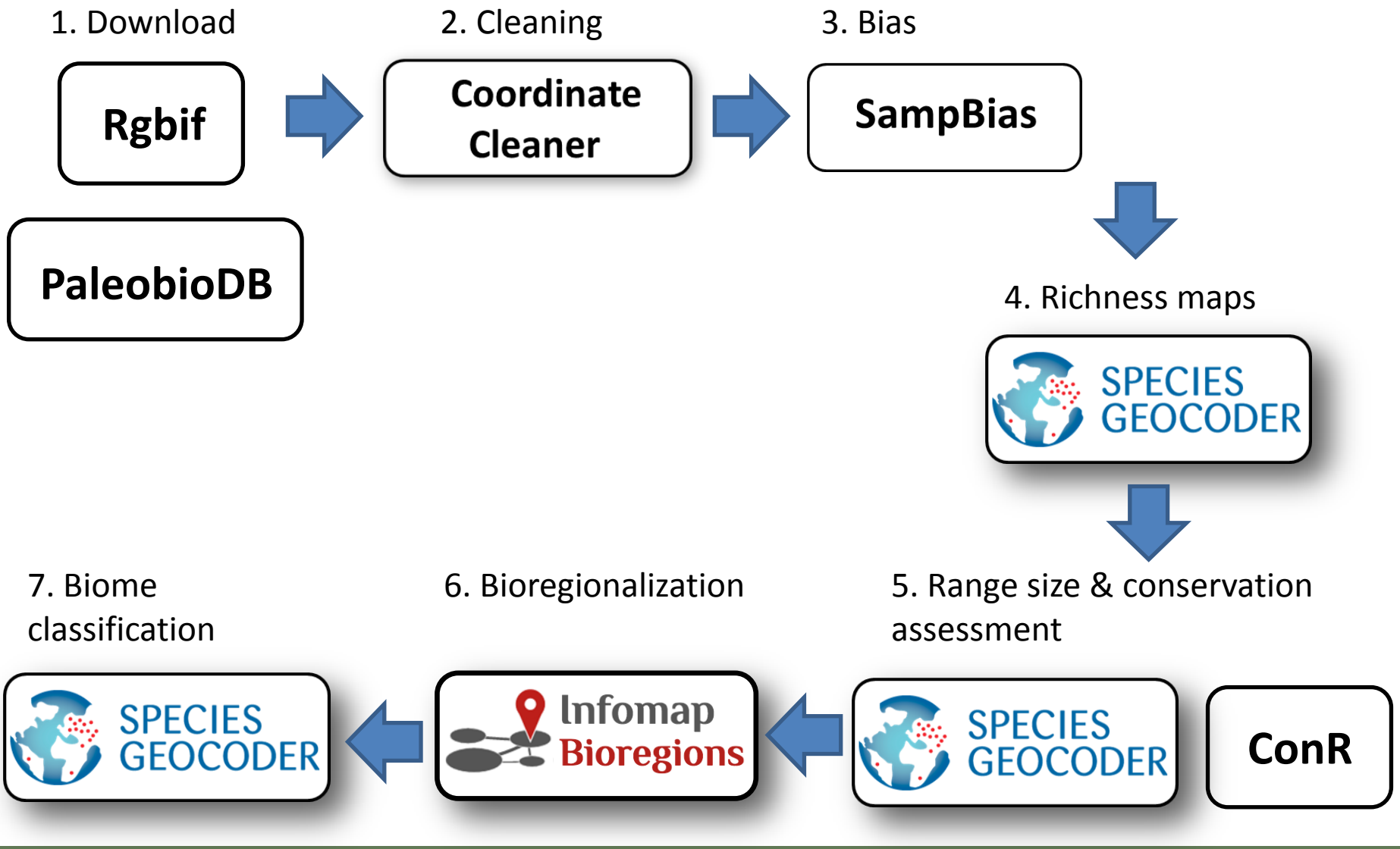
Fantastic!

- How can we use these data in biogeography and ecology?

How can we use these data in biogeography and ecology?

- Data quality
- Sampling effort

Analysis pipeline



Exercises

- 1. Downloading geographic occurrence data from GBIF **(R)**
- 2. Combine with data from the field & cleaning **(R)**
- 3. Quantifying sampling bias **(R + GUI)**
- 4. Species richness maps **(R)**
- 5. Range size and conservation assessment **(R)**
- 6. Bioregionalization **(GUI)**
- 7. Biome classification **(R)**

1. Download - Rgbif

Rgbif

- A programmatic interface to the Web Service methods provided by the Global Biodiversity Information Facility
- includes functions for searching for taxonomic names, retrieving information on data providers, getting species occurrence records, getting counts of occurrence records

2. Data cleaning - CoordinateCleaner

An R-package to identify potentially problematic records based on gazetteers

https://github.com/azizka/CoordinateCleaner

2. CoordinateCleaner

Coordinate Cleaner

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

Invalid or missing

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

Invalid or missing

Zeros

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

Invalid or missing
Zeros
Seas

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

Invalid or missing
Zeros
Seas
Country centroids

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

Invalid or missing
Zeros
Seas
Country centroids
Capitals

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

Invalid or missing
Zeros
Seas
Country centroids
Capitals
Institutions

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

Invalid or missing
Zeros
Seas
Country centroids
Capitals
Institutions
Wrong country

2. Data cleaning - Common coordinate errors

Coordinate Cleaner

Invalid or missing
Zeros
Seas
Country centroids
Capitals
Institutions
Wrong country
Outliers

2. Data cleaning - Workflow

Coordinate Cleaner

```r
flags <- CleanCoordinates(x = dat, lon = "decimalLongitude", lat = "decimalLatitude",
                          countries = "countryCode",
                          species = "species",
                          countrycheck = T,
                          outliers = T) # most test are on by default
```

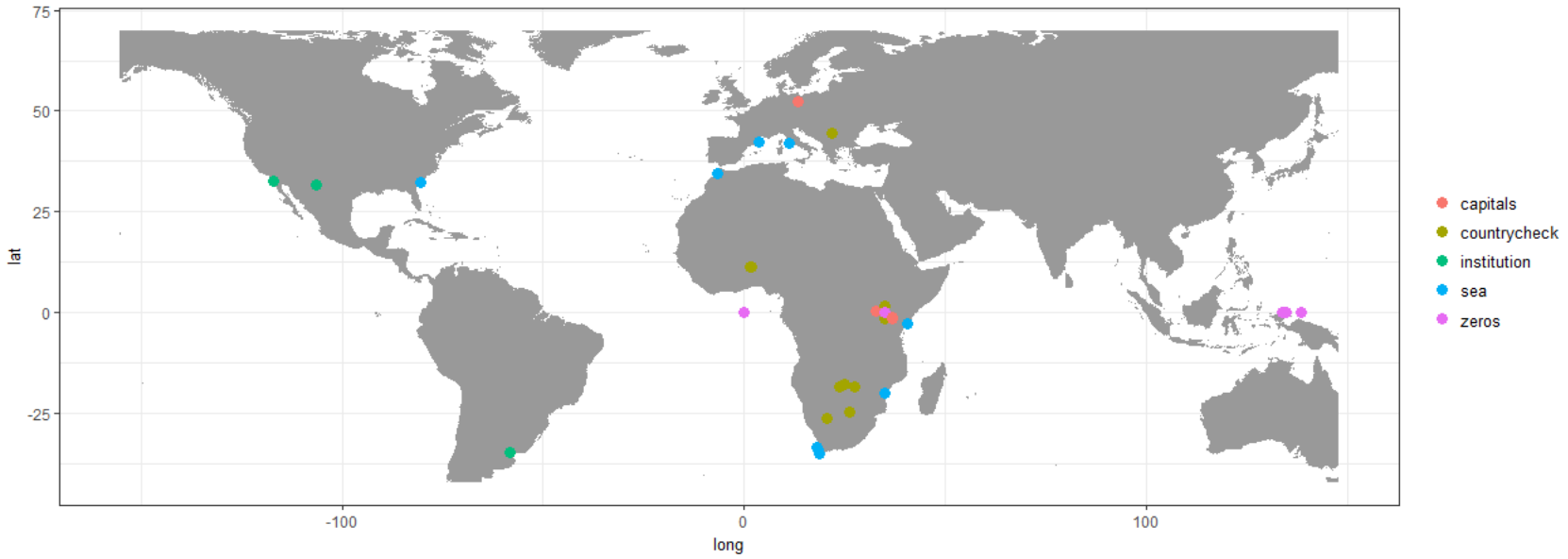
```
## running validity test
## running zero coordinate test
## flagged 27 records
## running capitals test
## flagged 5 records
## running centroids test
## flagged 22 records
## running seas test
## flagged 14 records
## running countrycheck test
## flagged 16 records
## running GBIF test
## flagged 0 records
## running institutions test
## flagged 9 records
## flagged 67 of 1332 records, EQ = 0.05
```

```
##   decimallongitude decimallatitude validity zeros capitals centroids  sea
## 1          1.46994        11.45179     TRUE  TRUE     TRUE      TRUE TRUE
## 2          1.47035        11.46455     TRUE  TRUE     TRUE      TRUE TRUE
## 3          1.49126        11.44658     TRUE  TRUE     TRUE      TRUE TRUE
## 4          1.58874        11.40803     TRUE  TRUE     TRUE      TRUE TRUE
## 5          1.47939        11.44816     TRUE  TRUE     TRUE      TRUE TRUE
## 6          1.35556        11.24228     TRUE  TRUE     TRUE      TRUE TRUE
##   countrycheck gbif institution summary
## 1         TRUE TRUE        TRUE    TRUE
## 2        FALSE TRUE        TRUE   FALSE
## 3         TRUE TRUE        TRUE    TRUE
## 4        FALSE TRUE        TRUE   FALSE
## 5         TRUE TRUE        TRUE    TRUE
## 6         TRUE TRUE        TRUE    TRUE
```

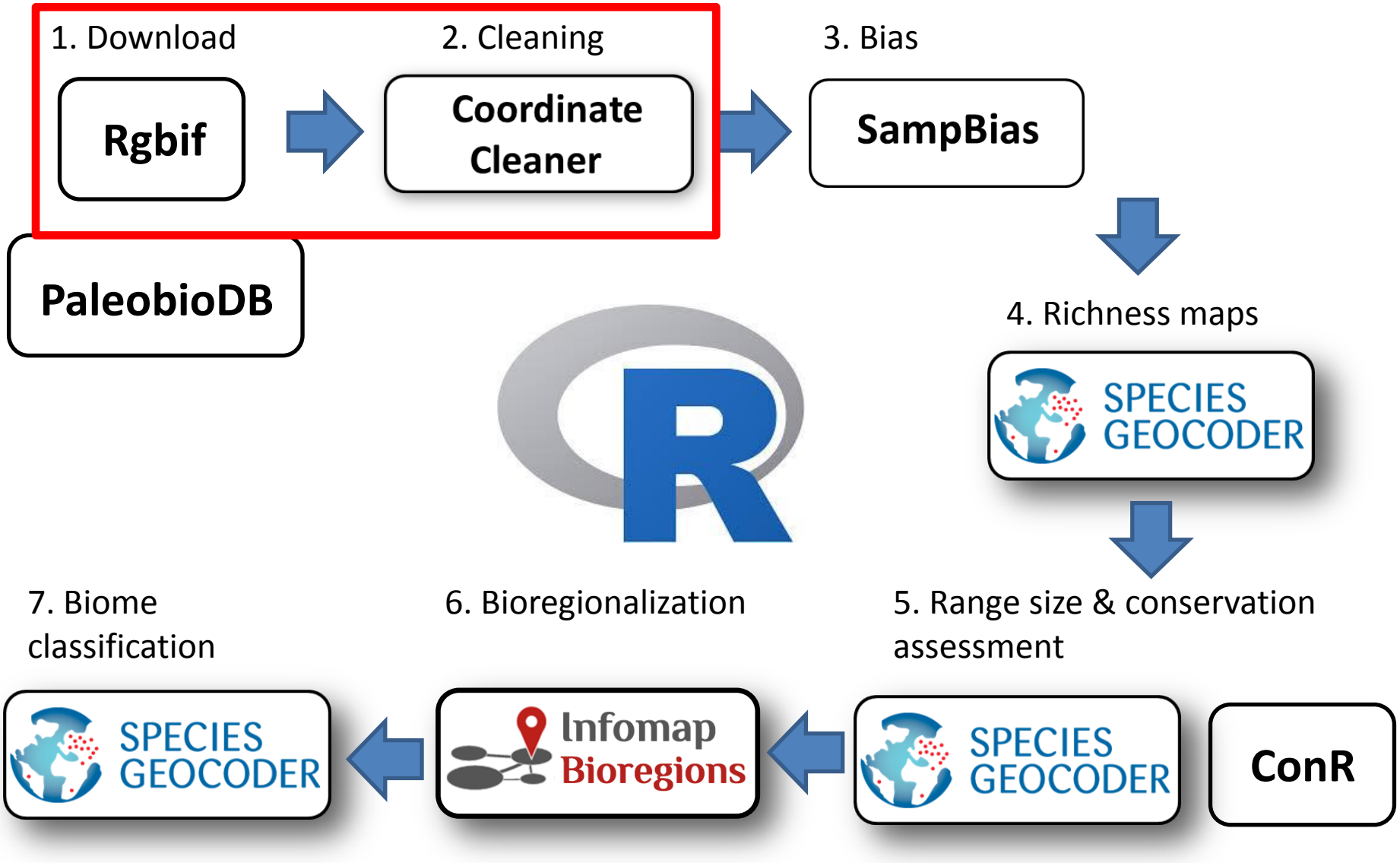
https://github.com/azizka/speciesgeocodeR

2. Data cleaning - Workflow

Coordinate Cleaner



Analysis pipeline



- Webpage

Analysis pipeline

1. Download

Rgbif

PaleobioDB

2. Cleaning

Coordinate Cleaner

3. Bias

SampBias

4. Richness maps

SPECIES GEOCODER

5. Range size & conservation assessment

SPECIES GEOCODER

ConR

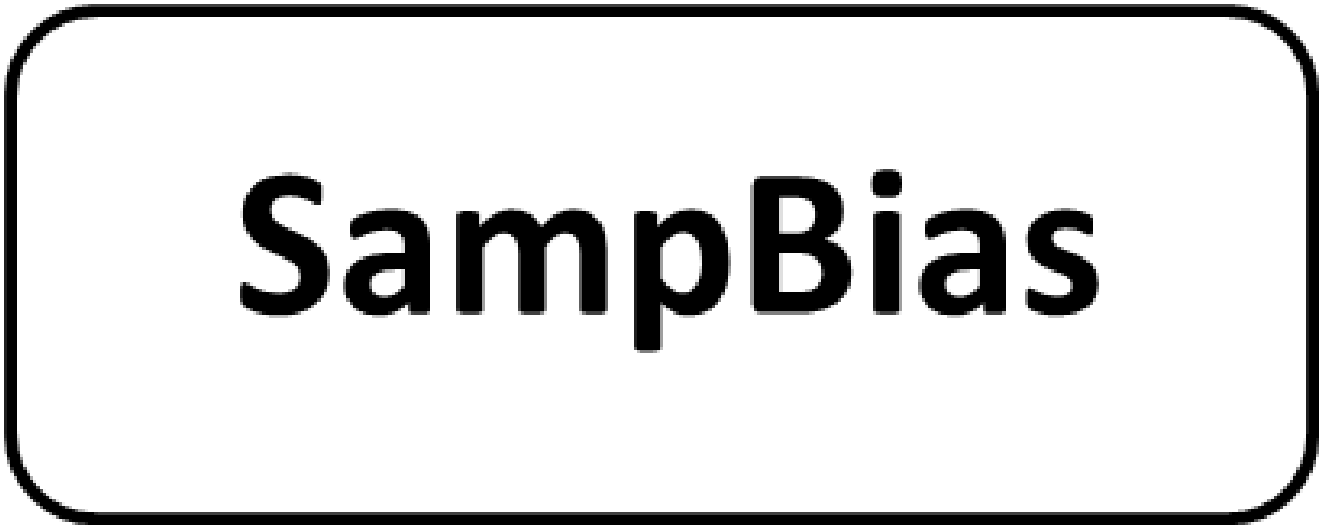
6. Bioregionalization

Infomap Bioregions

7. Biome classification

SPECIES GEOCODER

3. Bias



3. Bias - Background

SampBias

- Relation between anthropogenic geographic features and occurrence records

- Based on gazetteers and frequency distributions of distances

3. Bias - Background

SampBias

3. Bias - Background

SampBias

3. Bias - Background

SampBias

- Roads

3. Bias - background

SampBias

- Roads
- Cities

3. Bias - Background

SampBias

- Roads
- Cities
- Airports

3. Bias - Background

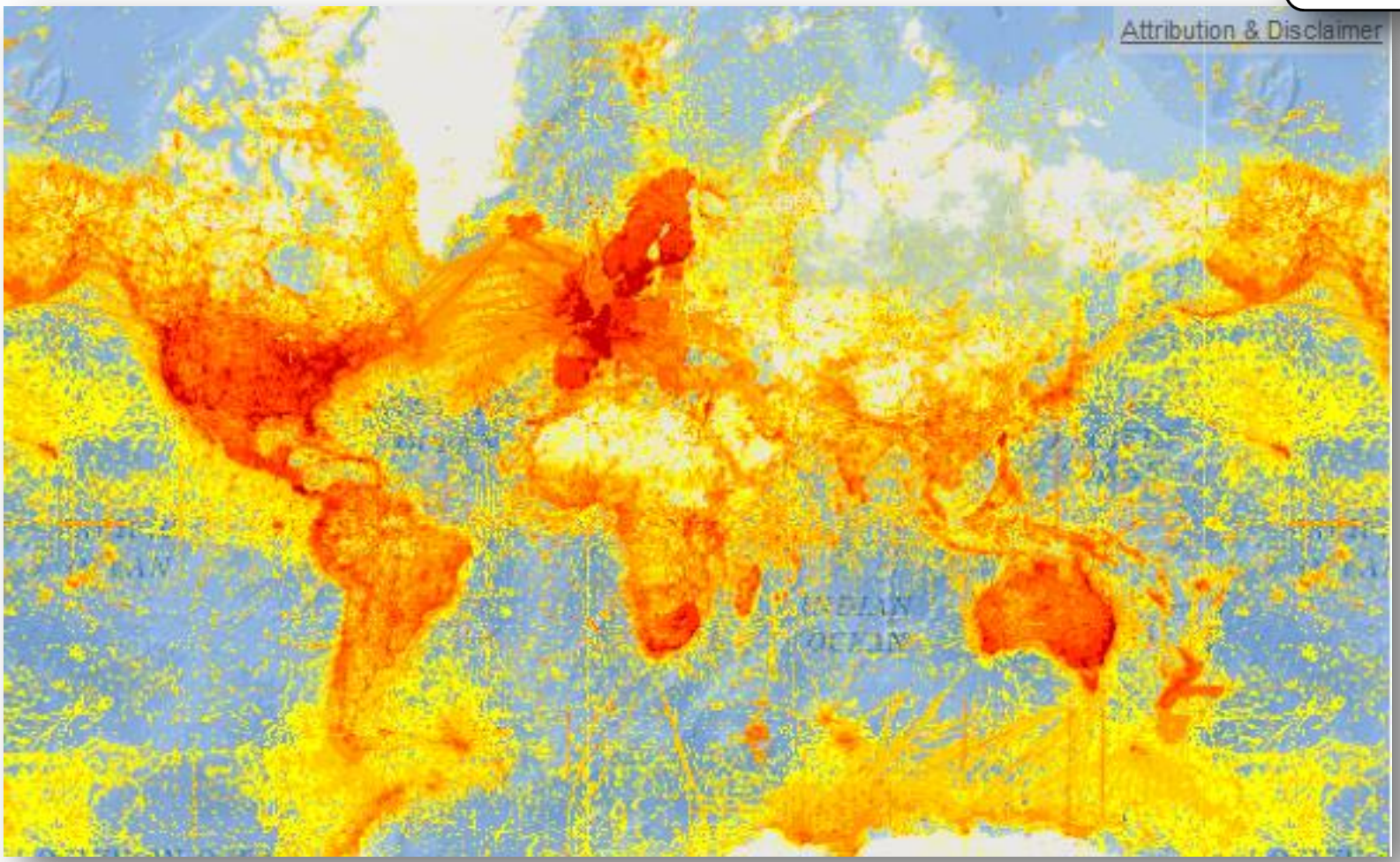
SampBias

- Roads
- Cities
- Airports
- Rivers

Sampling increases with accessibility

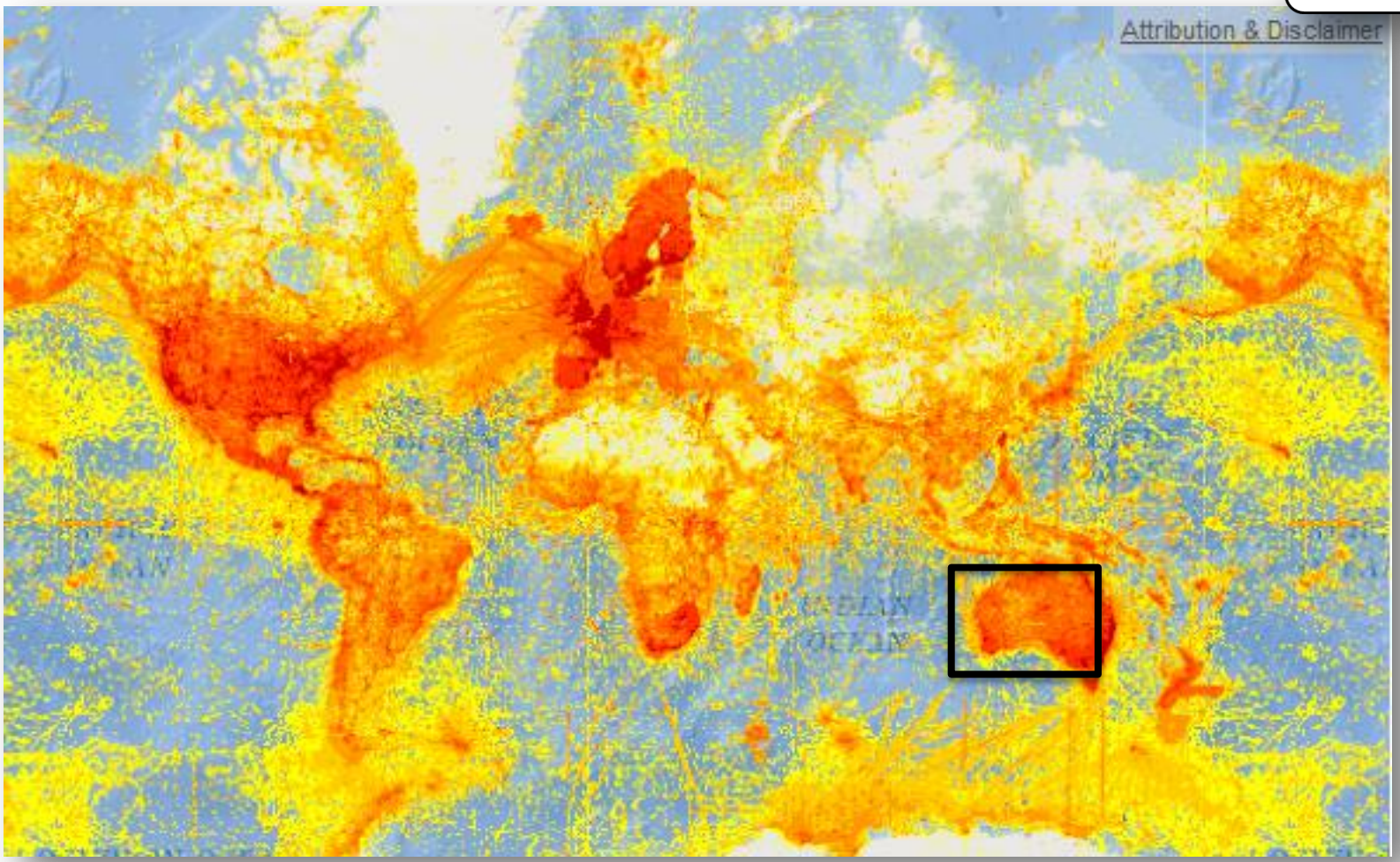
3. Bias - Background

SampBias



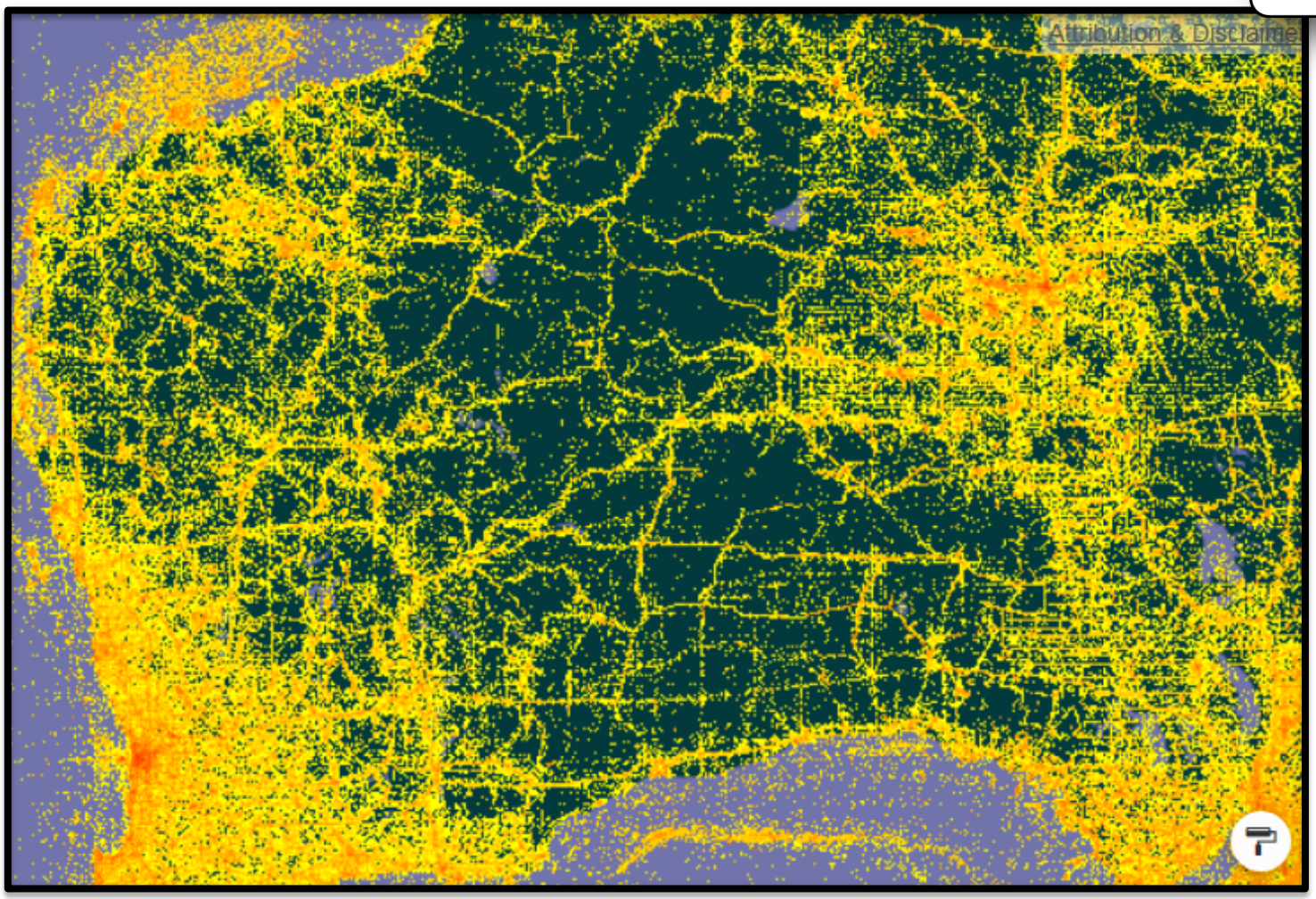
3. Bias - Background

SampBias



3. Bias - Background

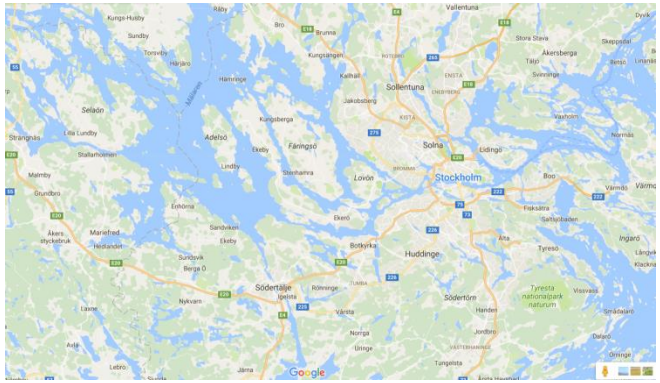
SampBias



3. Bias - How does it work?

SampBias

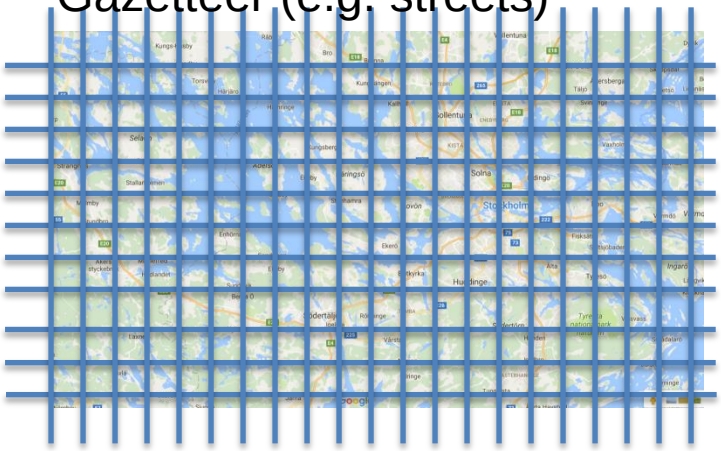
Gazetteer (e.g. streets)



3. Bias - How does it work?

SampBias

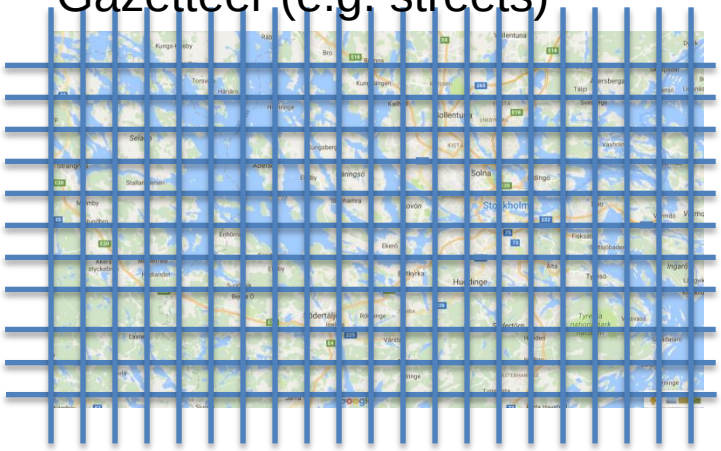
Gazetteer (e.g. streets)



3. Bias - How does it work?

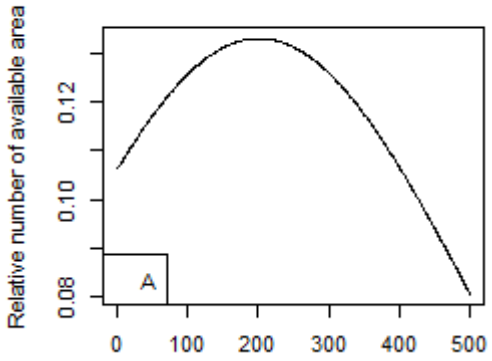
SampBias

Gazetteer (e.g. streets)



Minimum distance

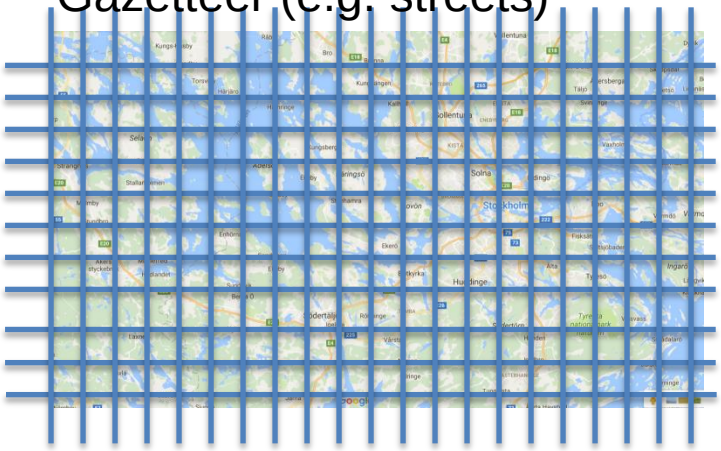
Expected distribution



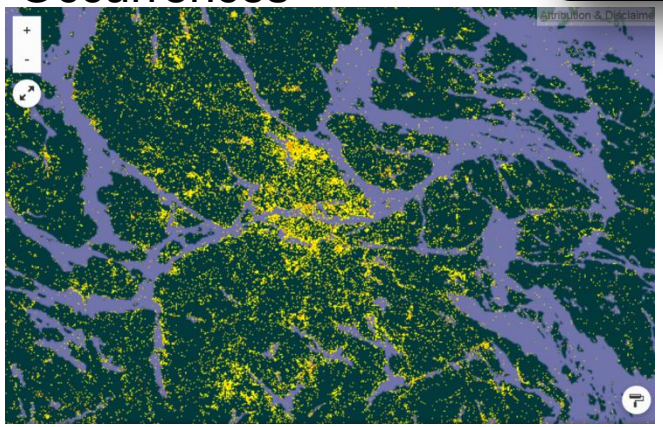
3. Bias - How does it work?

SampBias

Gazetteer (e.g. streets)

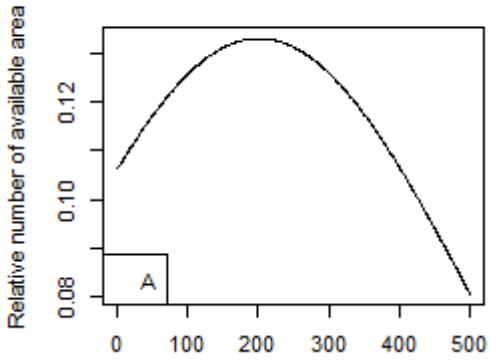


Occurrences



Minimum distance

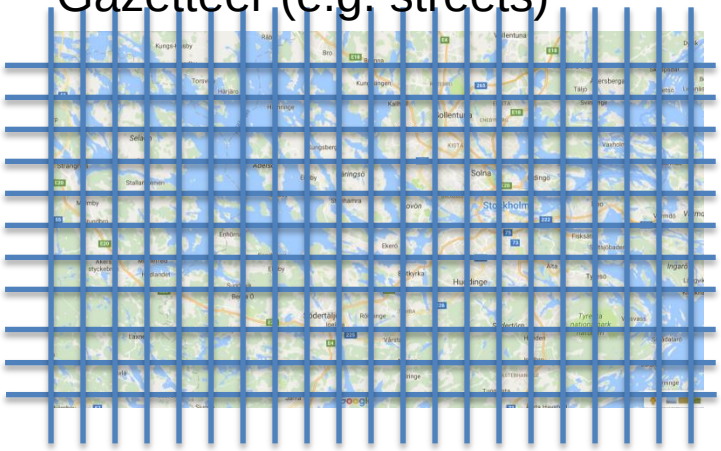
Expected distribution



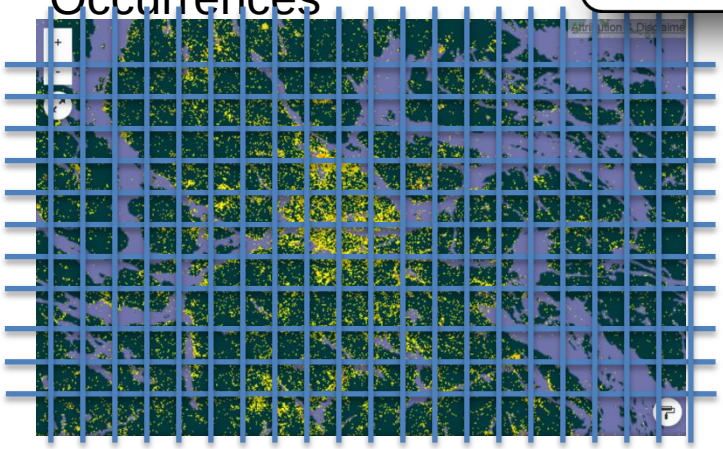
3. Bias - How does it work?

SampBias

Gazetteer (e.g. streets)

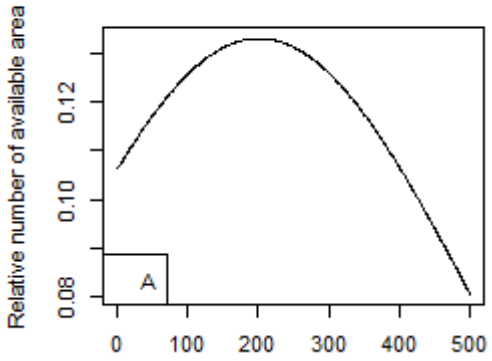


Occurrences

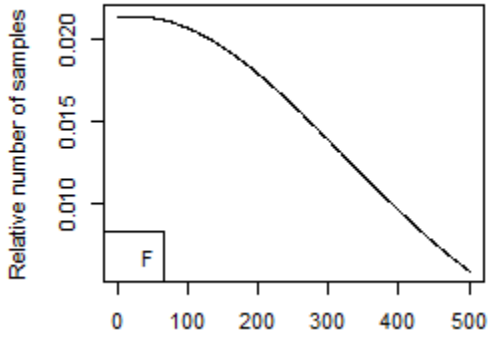


Minimum distance

Expected distribution

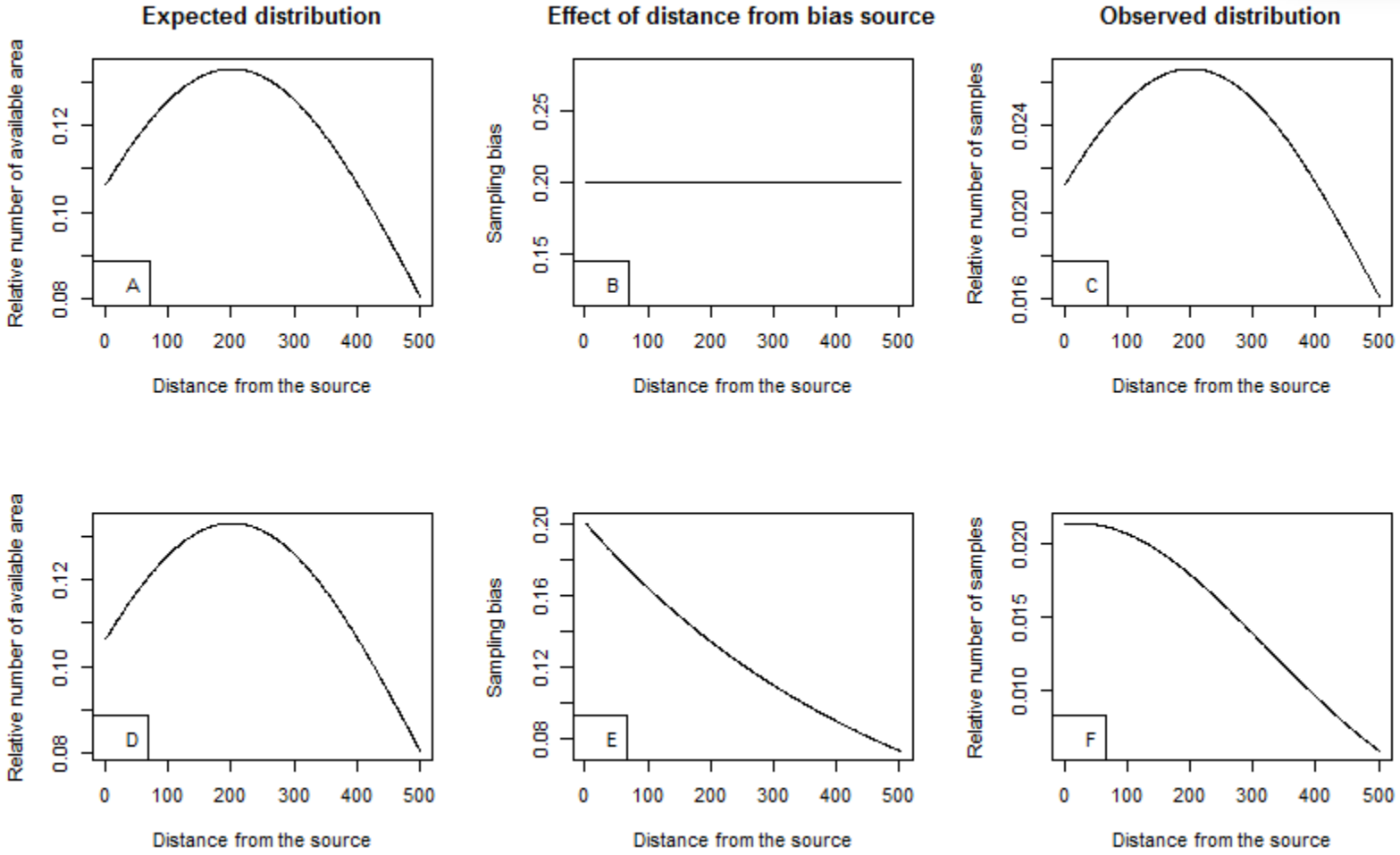


Observed distribution



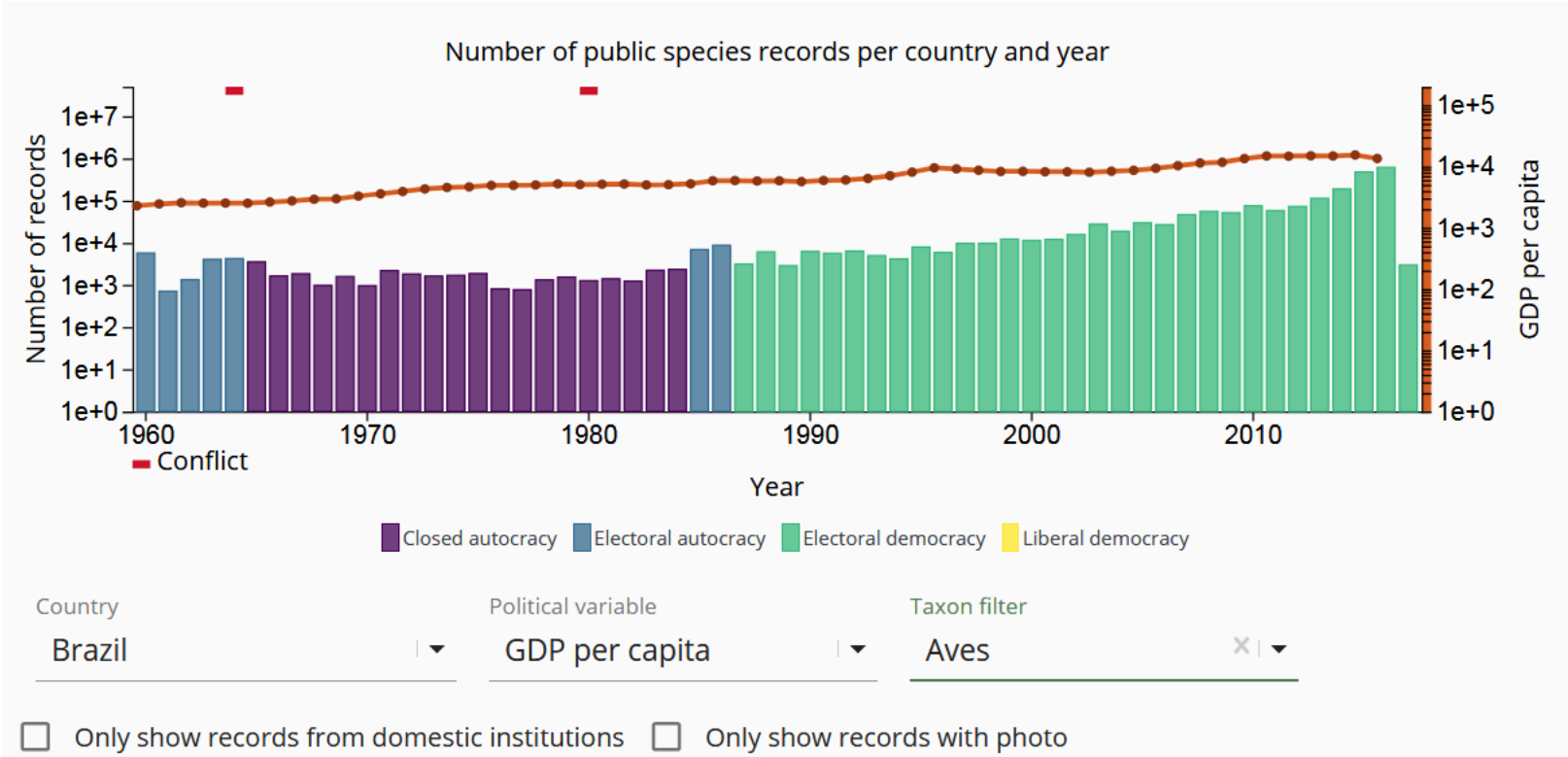
3. Bias - How does it work?

SampBias



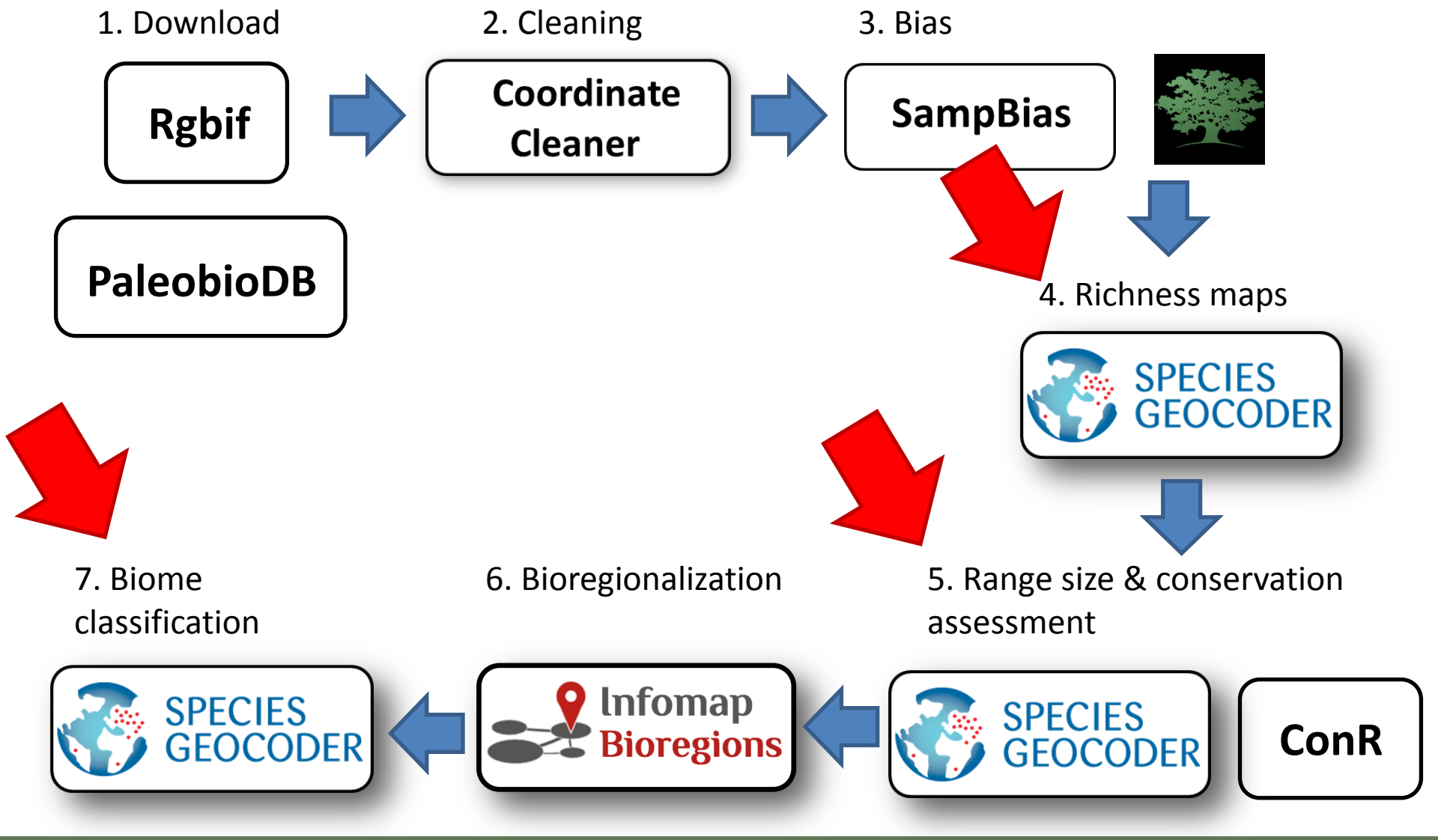
3. Bias - BioDem

- Explore the connection between political systems and biodiversity record collection



http://bio-dem.surge.sh/

Analysis pipeline

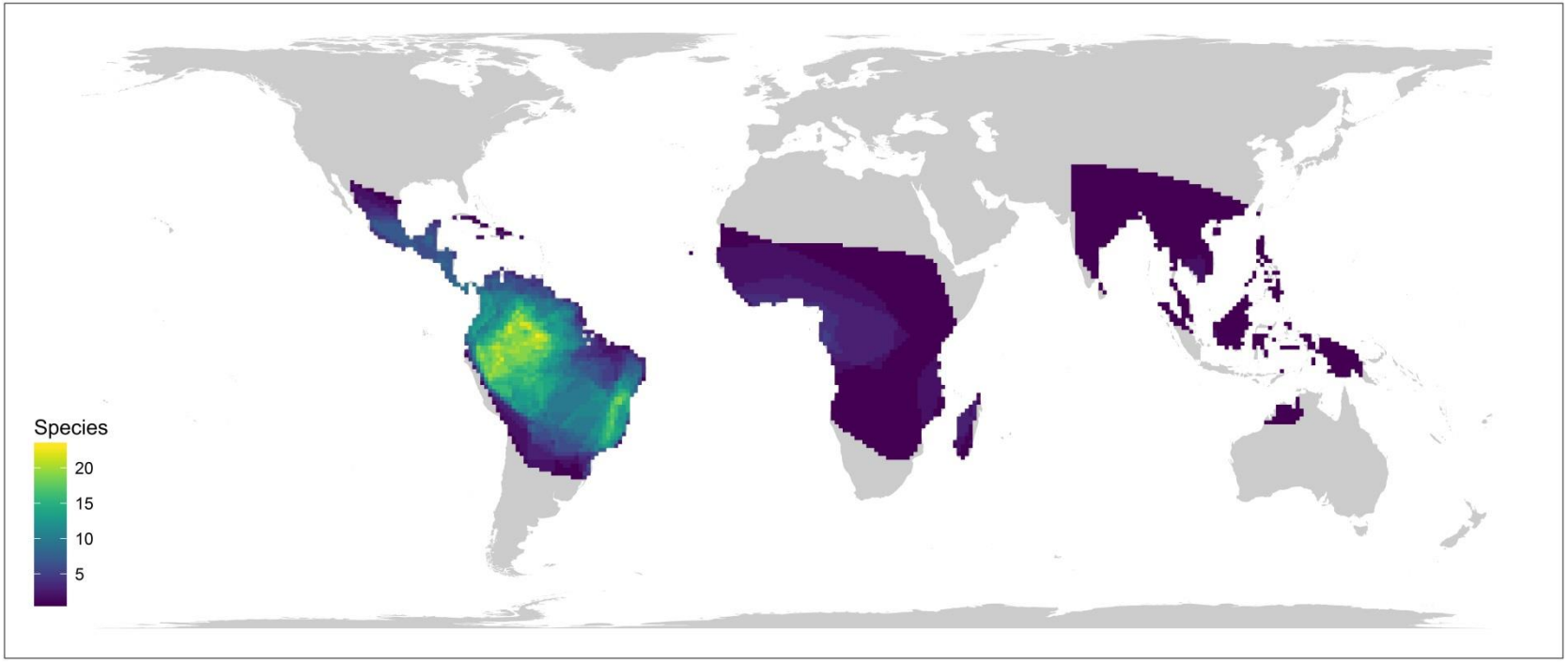


4., 5. & 7. Richness maps, range size and biome classification

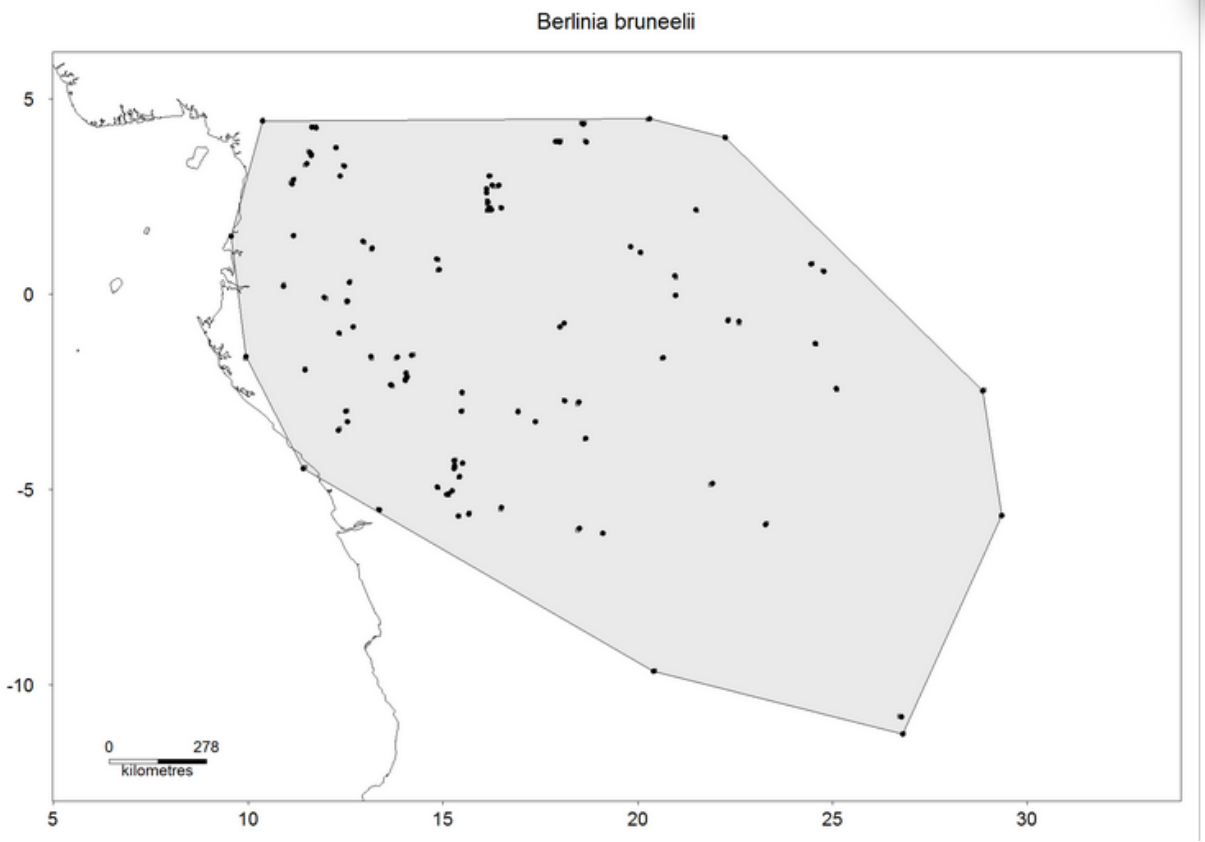
Töpel, Zizka *et al.* 2017 *Syst Biol*
Zizka *et al.* 2015 *bioRxiv*

4. Species richness maps

- Approximate species ranges
- Visualize species richness patterns



5. Range size - EOO



5. Conservation assessment - ConR

ConR

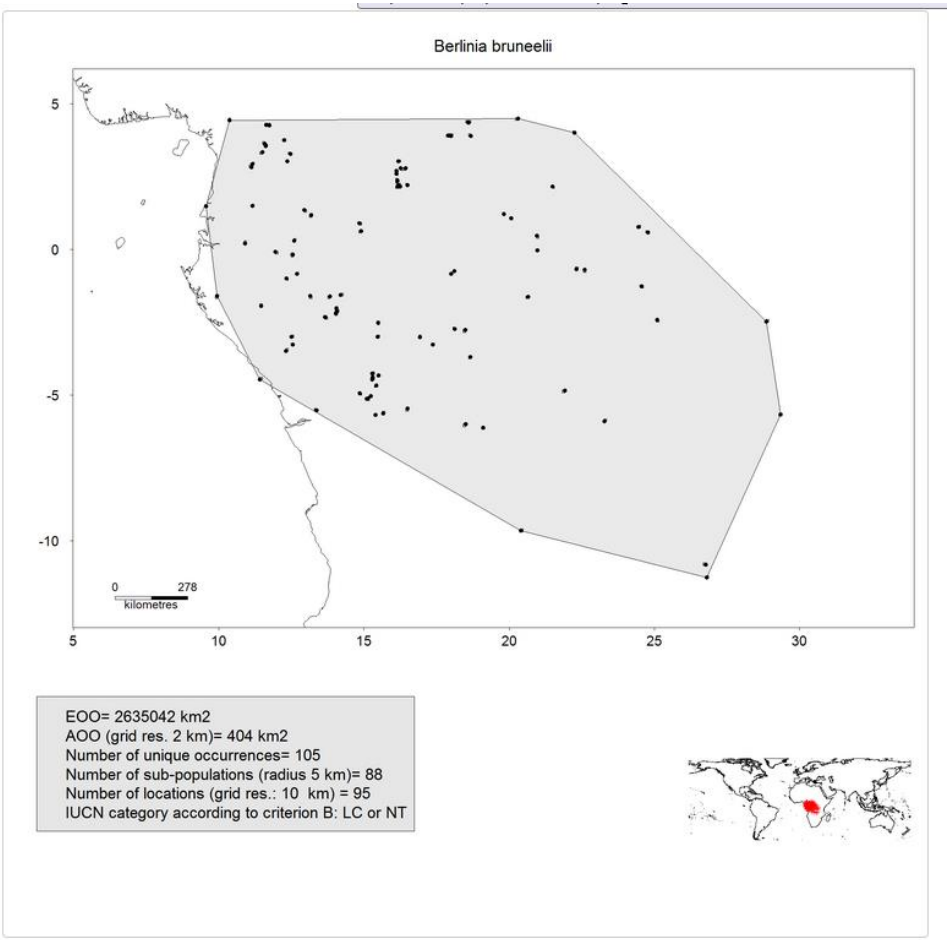
Computation of Parameters Used in Preliminary Assessment of Conservation Status

ddlat	ddlon	tax	higher.tax.rank	coly
-4.46667	11.4167	Berlinia bruneelii	Fabaceae	1827
-5.66667	29.3500	Berlinia bruneelii	Fabaceae	1989
3.88333	18.6833	Berlinia bruneelii	Fabaceae	1980
4.48333	20.3000	Berlinia bruneelii	Fabaceae	1805
-2.76667	18.4833	Berlinia bruneelii	Fabaceae	1788
-1.26667	24.5500	Berlinia bruneelii	Fabaceae	1993

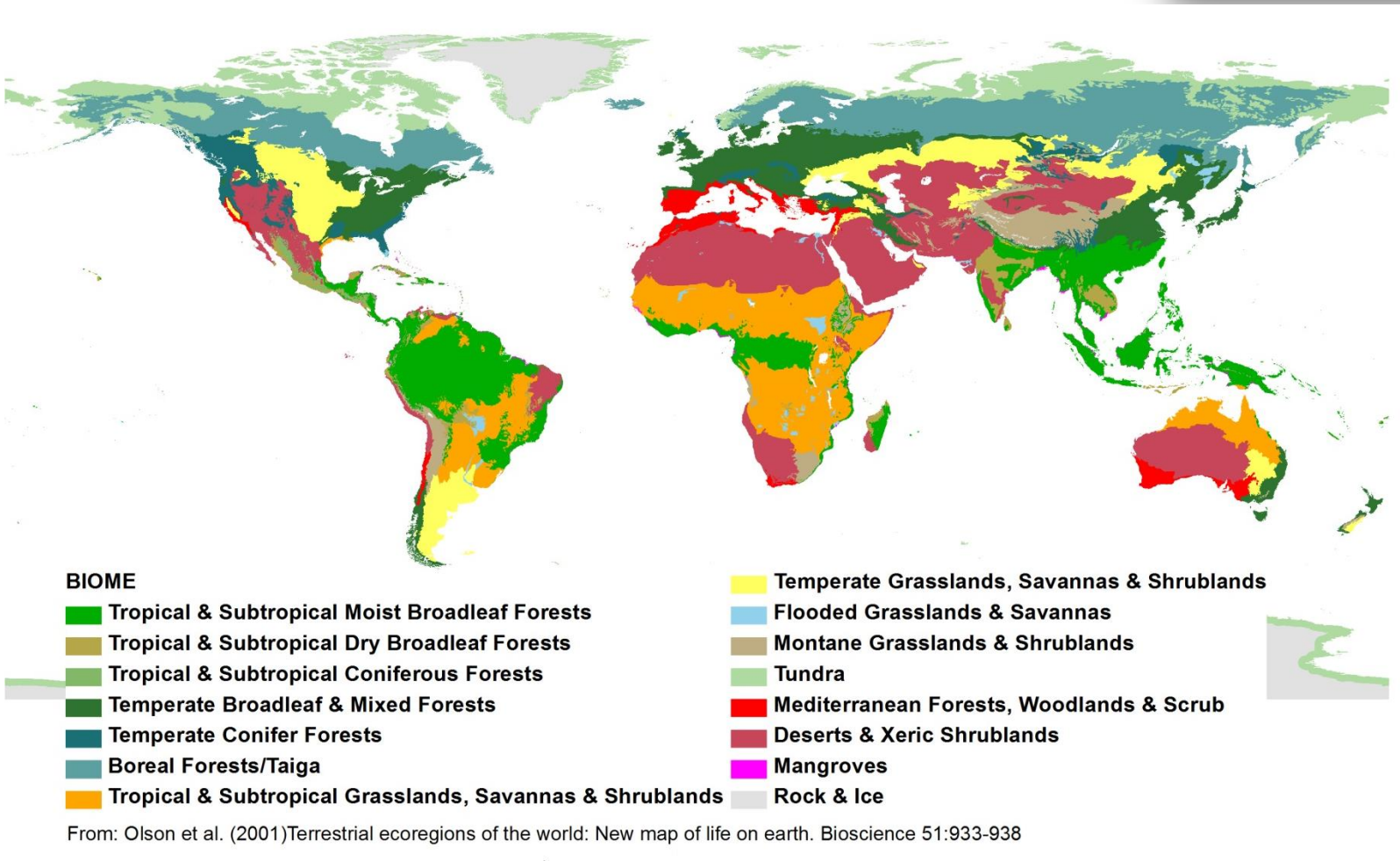
```
IUCN.eval(MyData)
```

5. Conservation assessment - ConR

ConR

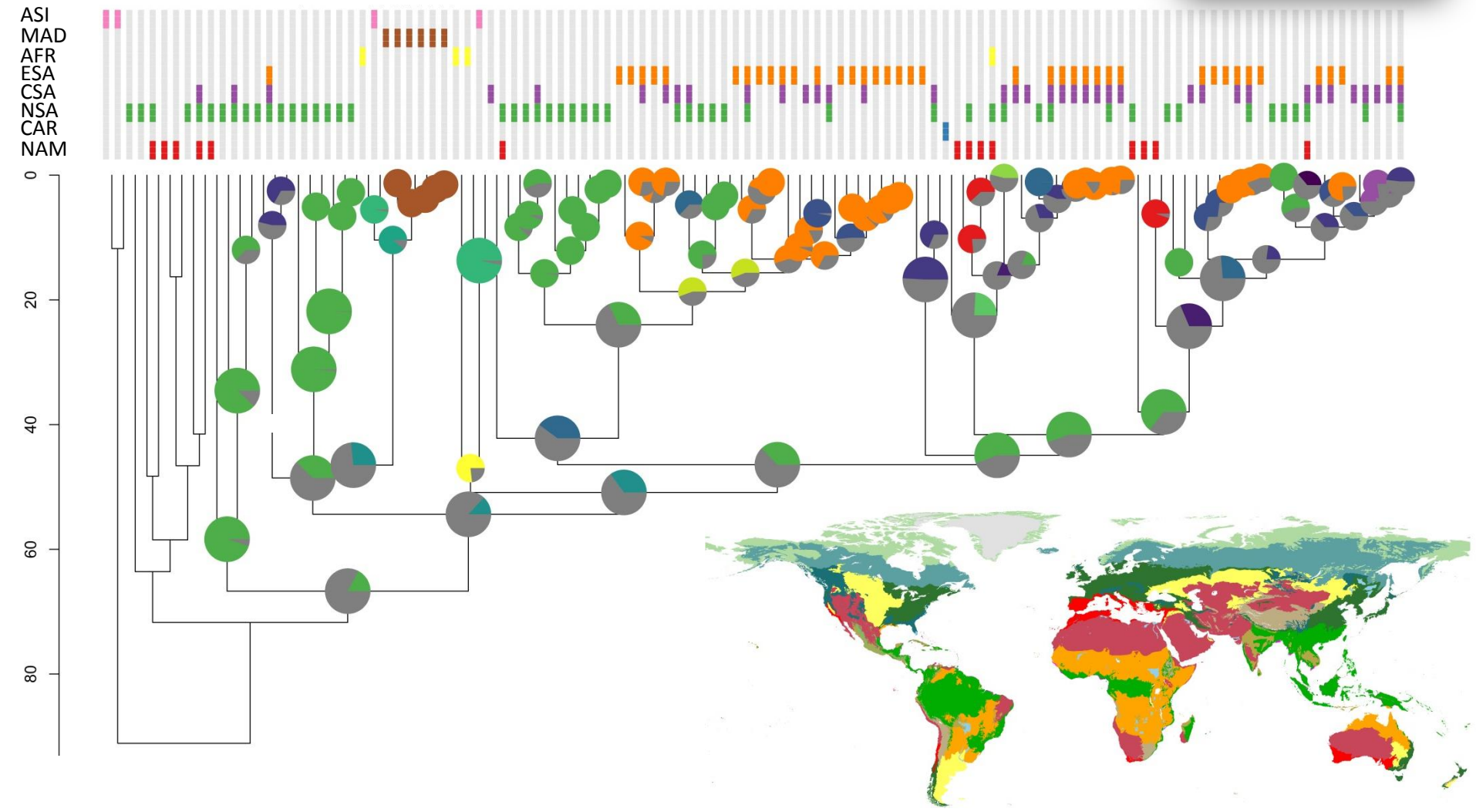


7. Biome classification



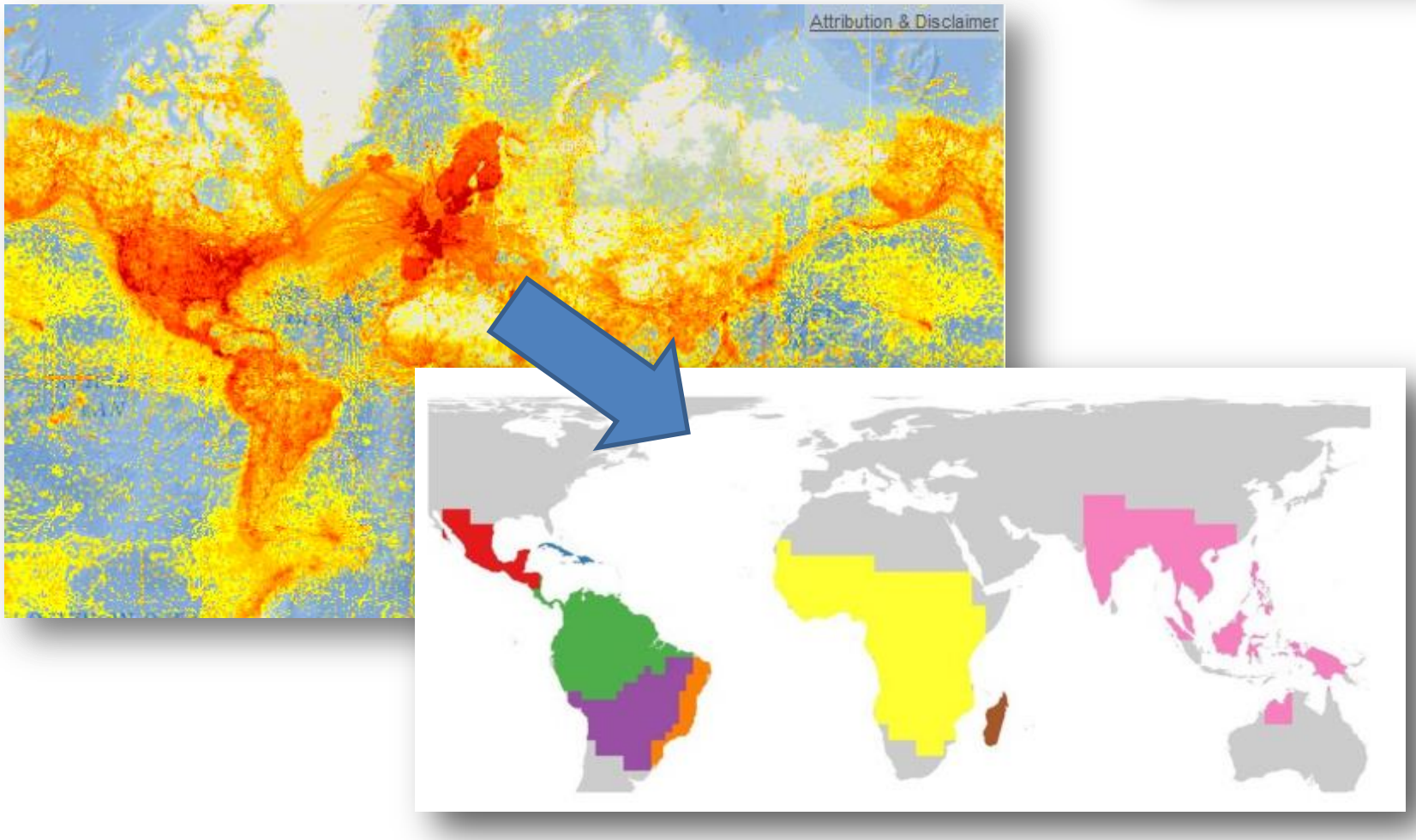
From: Olson et al. (2001)Terrestrial ecoregions of the world: New map of life on earth. Bioscience 51:933-938

7. Biome classification

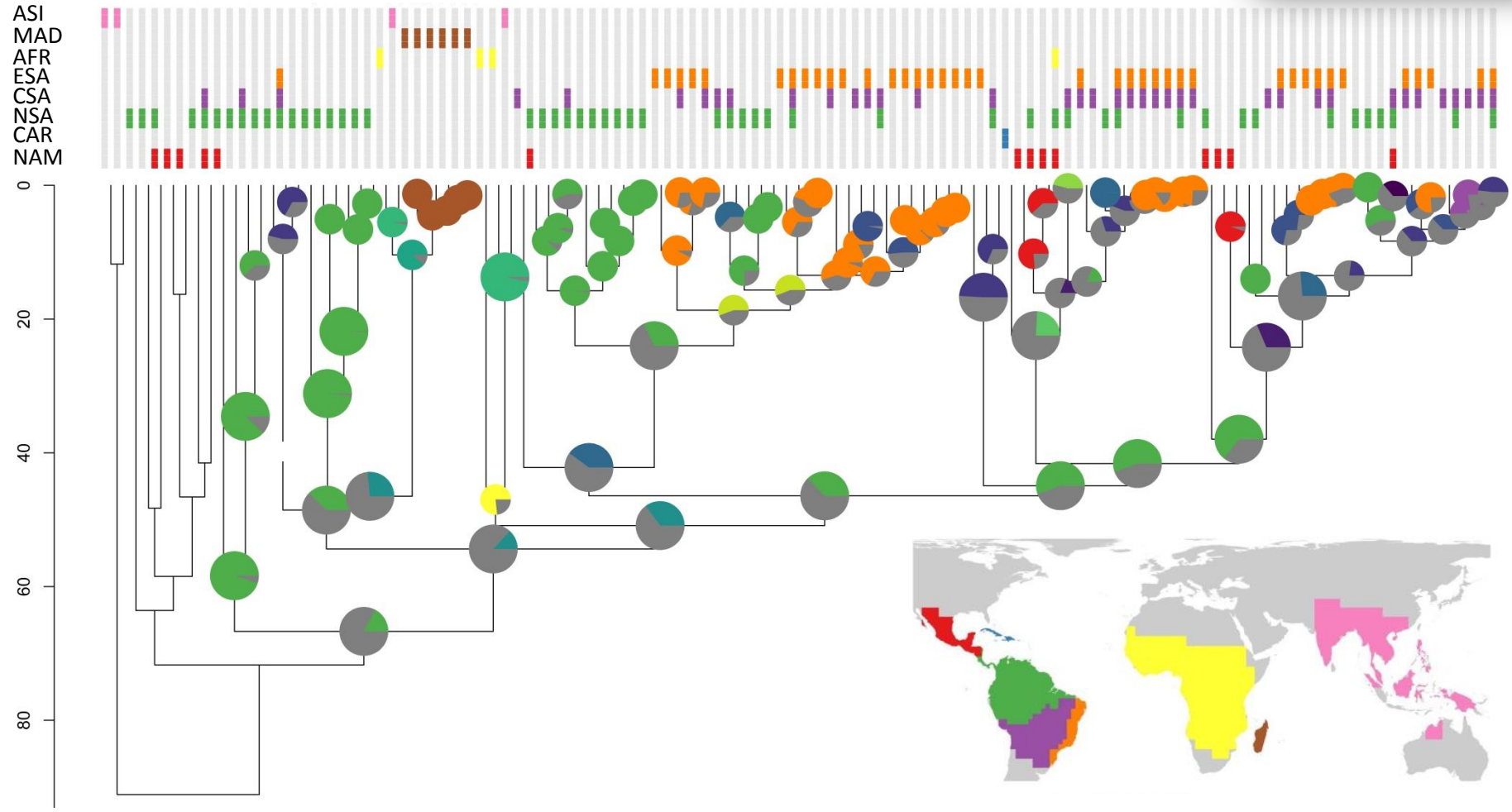


6. Bioregionalization

Infomap Bioregions



6. Bioregionalization



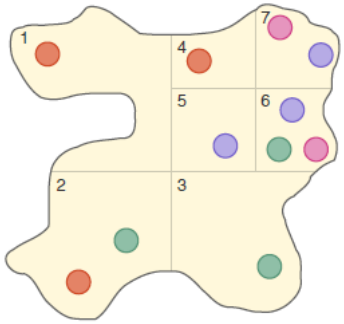
6. Bioregionalization - How does it work?

Infomap Bioregions

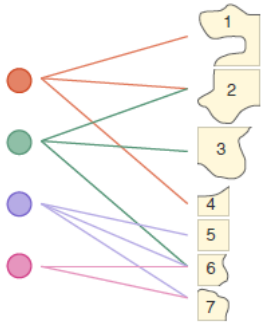
a) Input data

```
Name, Lat, Long
Species A, 6.9, 1.0
Species A, 6.7, 4.7
Species A, 1.3, 2.8
Species B, 2.3, 2.9
Species B, 1.8, 6.4
Species B, 4.6, 6.6
Species C, 4.7, 5.3
Species C, 5.5, 6.8
Species C, 6.8, 7.6
Species D, 7.4, 6.5
Species D, 4.6, 7.4
```

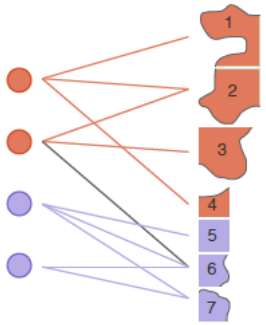
Quadtree mapping



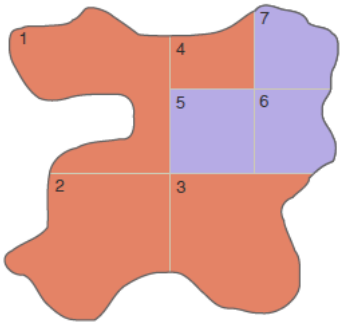
Network mapping



Network clustering



Bioregions

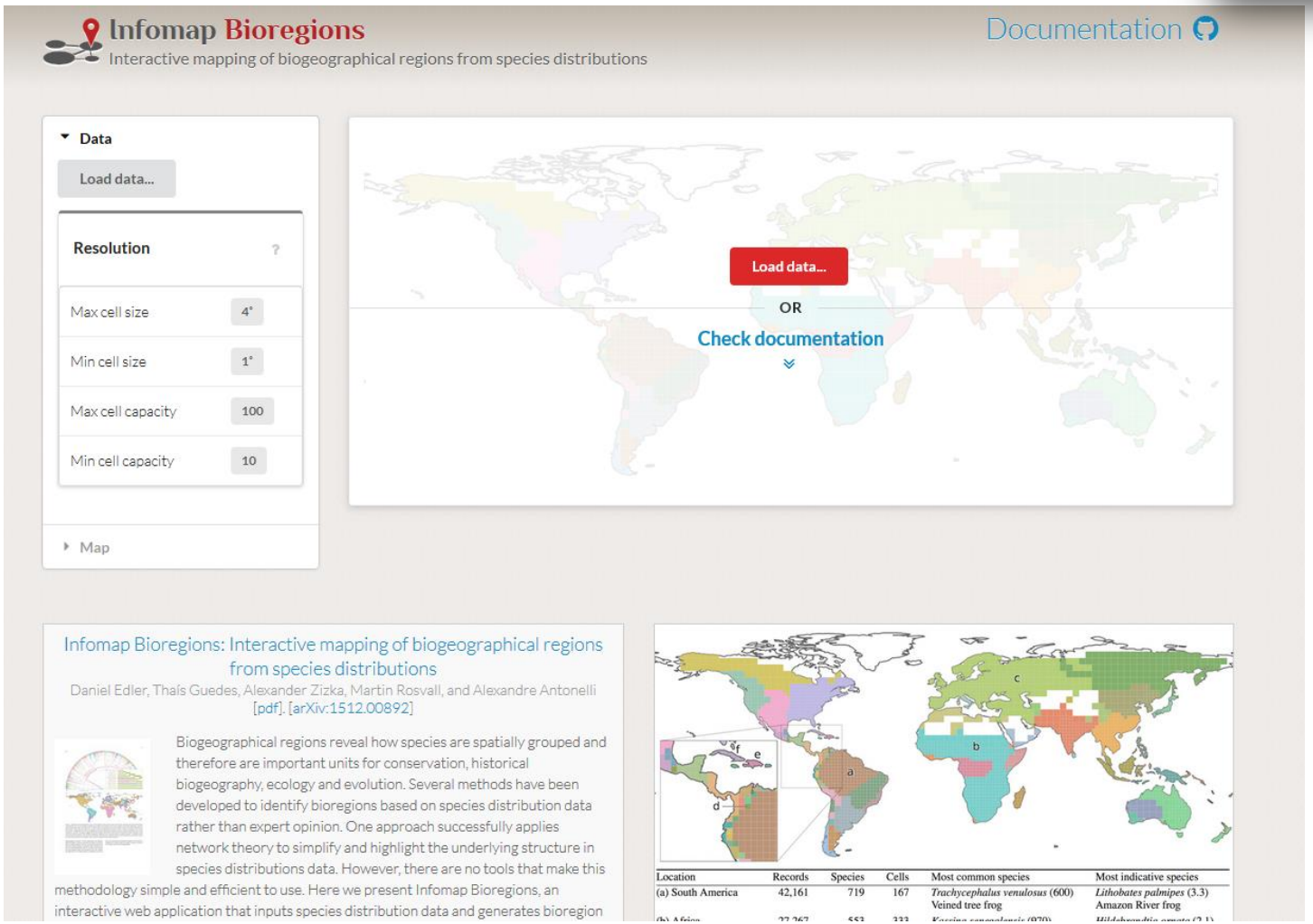


Exporting

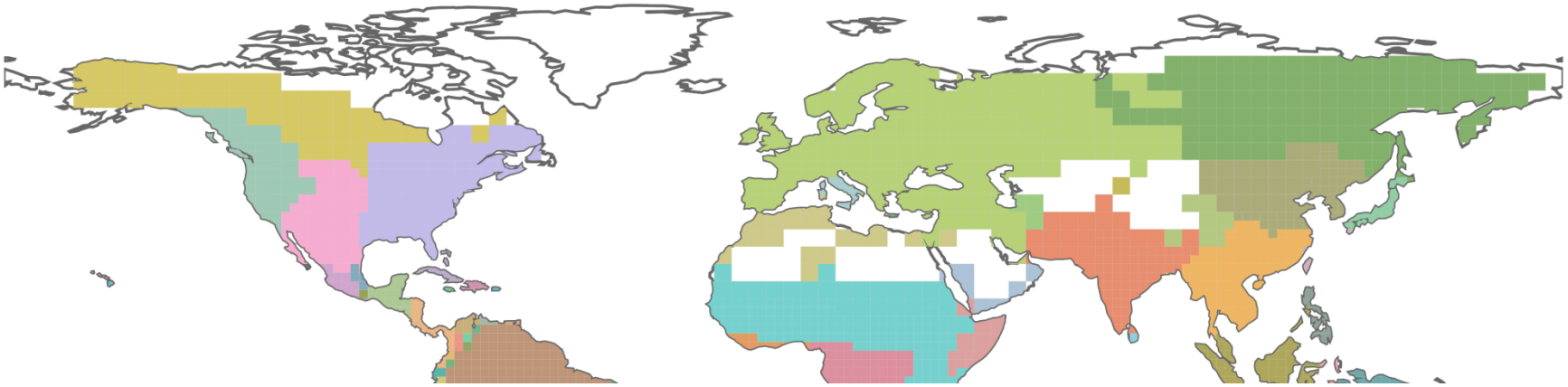
f) Output data

- svg/png
- csv
- NEXUS
- shapefile
- GeoJSON

Bioregionalization - Workflow



6. Bioregionalization - Output



Location	Records	Species	Cells	Most common species (records)	Most indicative species (score)
(a) South America	42,161	719	167	*Trachycephalus venulosus* (600) Veined tree frog	*Lithobates palmipes* (3.3) Amazon River frog
(b) Africa	27,267	553	333	*Kassina senegalensis* (970) Senegal running frog	*Hildebrandtia ornata* (2.1) African ornate frog
(c) Eurasia	13,083	103	313	*Rana arvalis* (1,547) Moor frog	*Triturus cristatus* (1.3) Northern crested newt
(d) Andes	121	75	1	*Pristimantis nervicus* (13) –	*Pristimantis nervicus* (157) –
(e) Hispaniola	181	65	4	*Hypsiboas heilprini* (28) Los Bracitos tree frog	*Osteopilus vastus* (73) Hispaniola tree frog
(f) Cuba	214	61	4	*Osteopilus septentrionalis* (28) Cuban tree frog	*Eleutherodactylus varleyi* (73) –

Edler *et al.* 2017

Today's schedule

- 8.00-9.00: Finish exercises 1-2
- 9.00-9.30: Zizka's presentation of further analyses and discussion
- 9.30-15.00: Exercises 3-6 on your clade of interest. *(If time allows, additional analyses.)*
- 15.00-17.00: Preparation of slides for tomorrow
- https://github.com/azizka/Biodiversity_Data_from_Field_to_Yield/tree/master/field_data

Project presentation

- 5 minutes, 4-5 slides
  - S1: General information on your clade (#spp/gen, global distribution, a few images)
  - S2: How many records in GBIF & PaleoDB (E1)? How many potentially wrong records in GBIF (E2)?
  - S3: Richness map (E3), biome classification (E7)
  - S4: Bioregionalization (for the larger taxon, e.g. order or family) (E6)
  - S5: Additional analysis if time allows (e.g., conservation status, mapping on a phylogeny, major spatial biases, … see Tutorials for examples)

Obrigado!

SampBias

Coordinate Cleaner

ConR

Rgbif

http://antonelli-lab.net/resources.php
https://github.com/azizka/CoordinateCleaner
https://github.com/azizka/sampbias