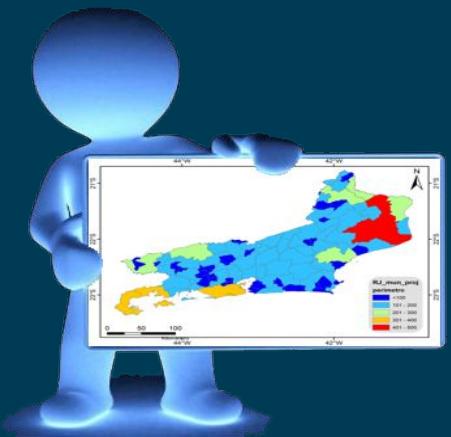


# Tópicos Especiais em RGV II

## Modelagem de adequabilidade ambiental



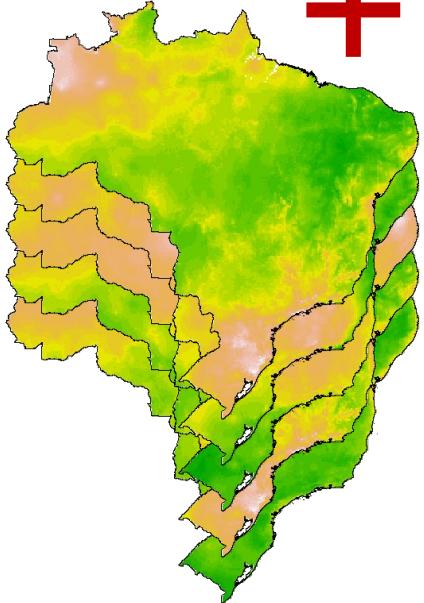
Diogo S. B. Rocha

## Registros de ocorrência



Dados de entrada

Variáveis preditoras



Precipitação

Temperatura

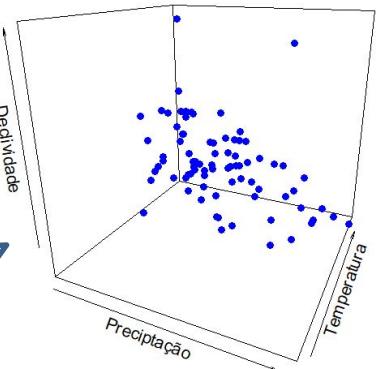
Topografia

Variável X

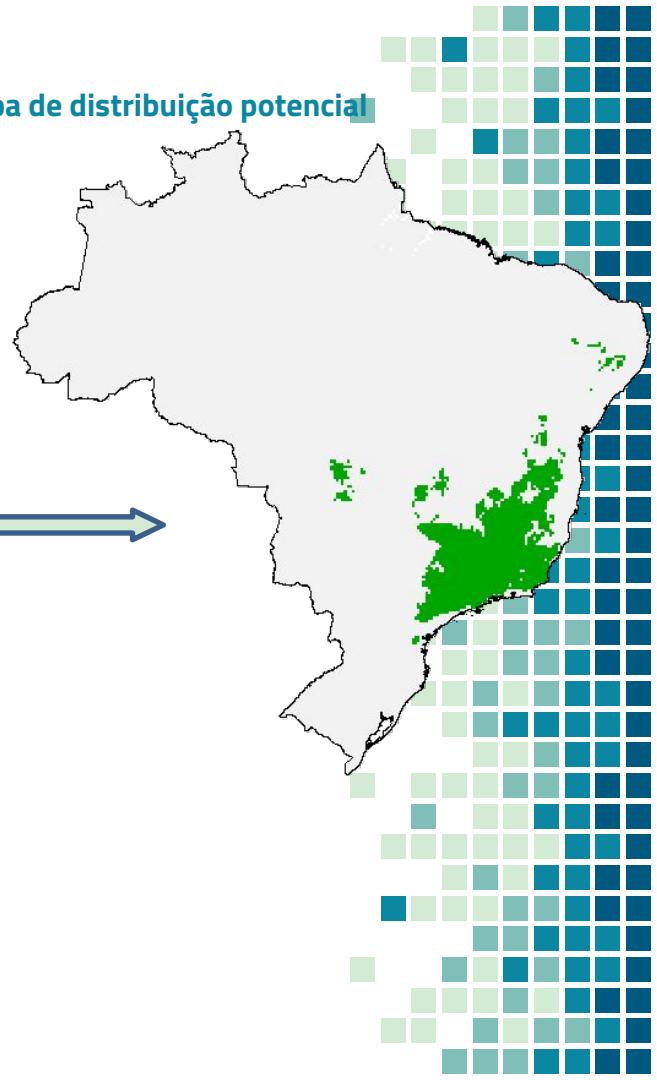
Variável Y

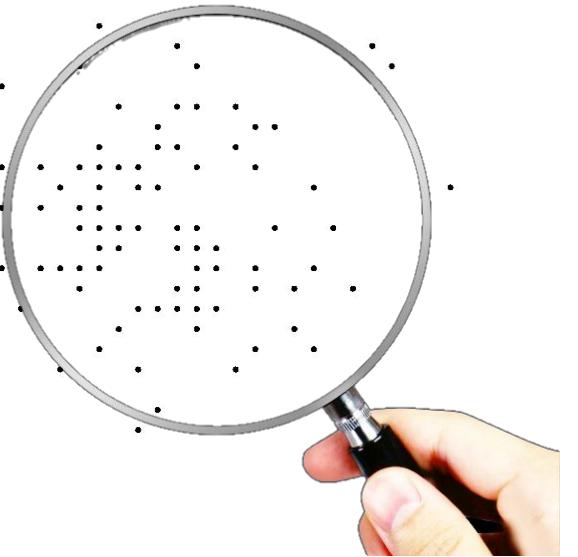
## Algoritmos de modelagem

(Bioclim, GLM, GAM, ANN, MAxEnt, etc.)



## Mapa de distribuição potencial





# Registros de ocorrência

Aquisição de dados



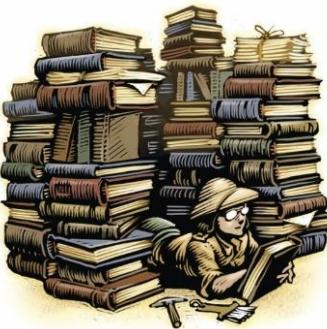
# Fonte de dados



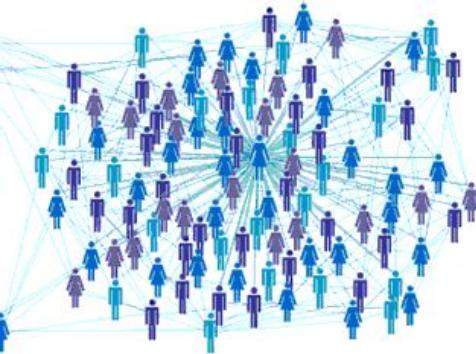
Global Biodiversity  
Information Facility

*species* link

Literatura  
acadêmica  
especializada



Redes de  
colaboração





# GBIF

Global Biodiversity  
Information Facility

Occurrence records

1.707.961.747

Datasets

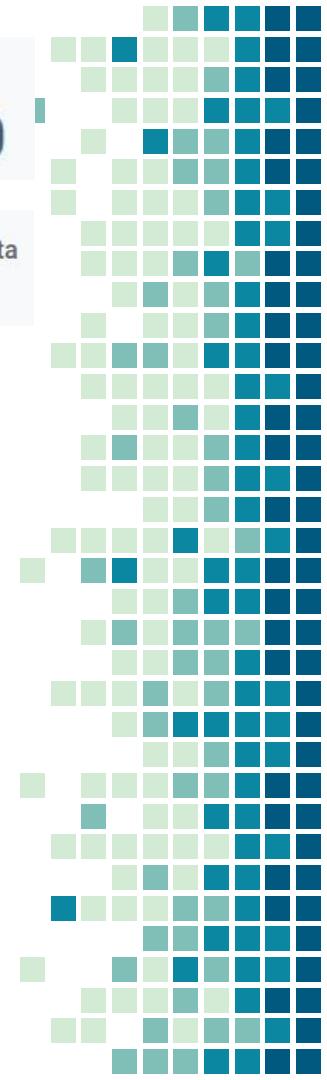
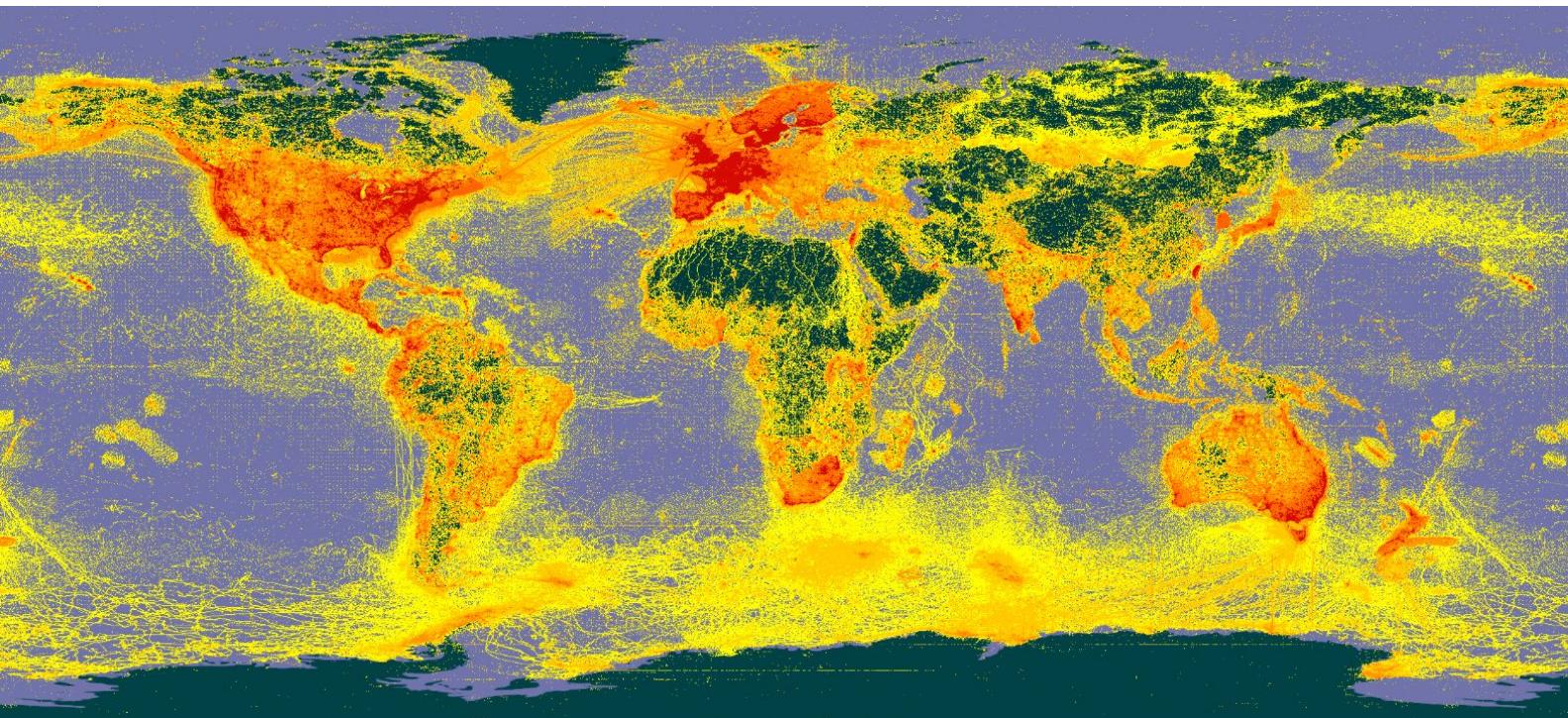
60.210

Publishing institutions

1.692

Peer-reviewed papers using data

5.831

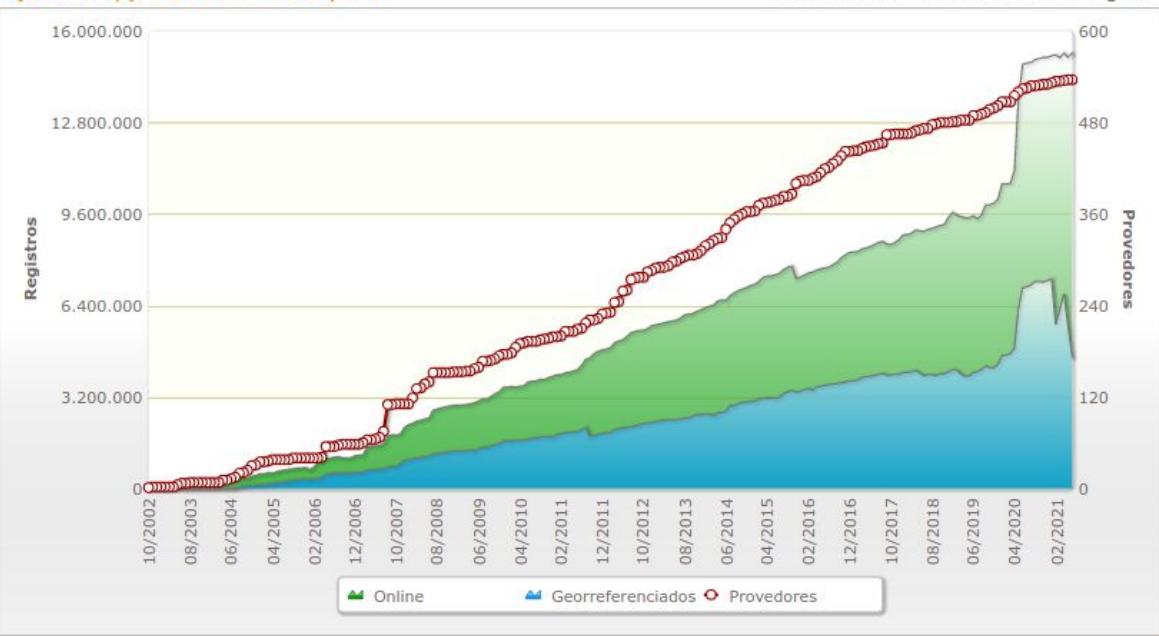


## Indicadores

Os indicadores apresentados são baseados nos dados disponibilizados pelas coleções participantes da rede speciesLink. São relatórios dinâmicos ou diários apresentados em formato de mapas e gráficos. O indicador reflete apenas a análise dos dados que estão na Internet, podendo, portanto, não refletir a realidade da coleção cuja maior parte do acervo não está digitalizada. [Clique aqui para visualizar a tabela Contribuição x Dependência de registros por estado.](#)

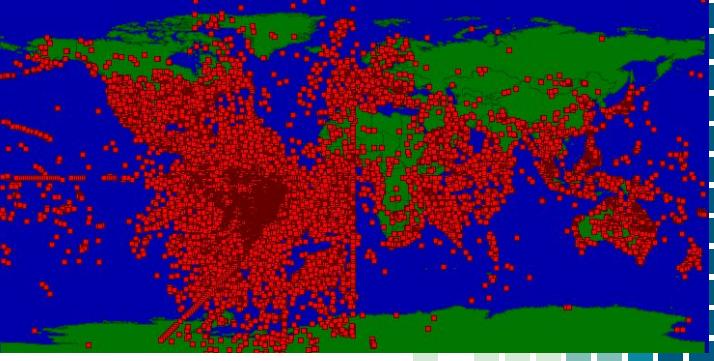
Veja outras opções de indicadores aqui ↓

todas as redes - todos os acervos - registros



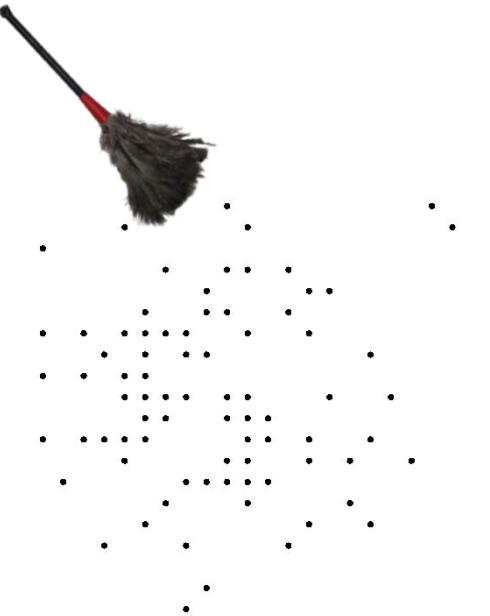
Histórico do fluxo de registros disponível na rede. São apresentadas as médias mensais, tanto do número total de registros online, como também do número de registros georreferenciados. A linha vermelha representa o número mensal de provedores de dados (coleções biológicas).

Atualizado em 11/06/21 06:01



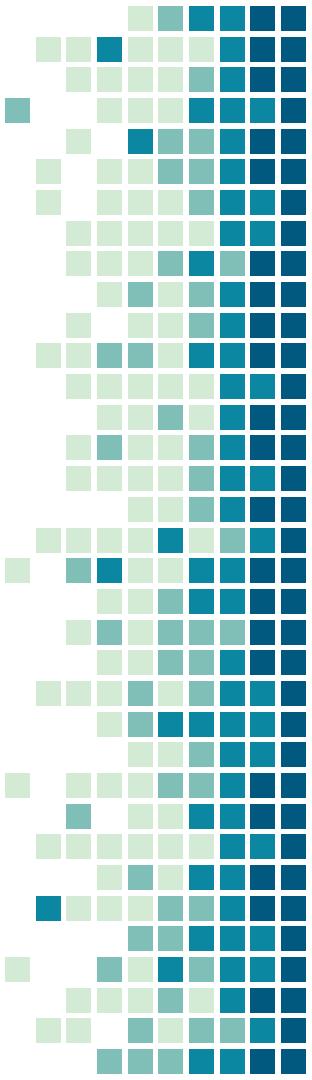
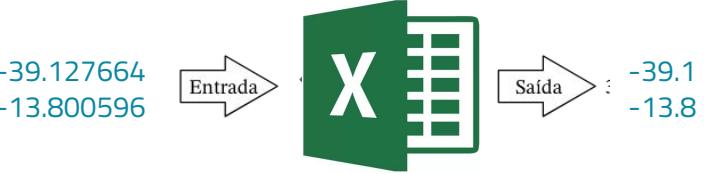
# Registros de ocorrência

Limpeza de dados



# Qualidade dos dados

- Armazenamento e difusão;
  - Arredondamentos automáticos em planilhas;
- Manuseio dos dados;
  - As colunas estão corretamente identificadas?
  - Cuidado com a junção de dados de diferentes projetos, diferentes objetivos, diferentes precisões...



# Qualidade taxonômica

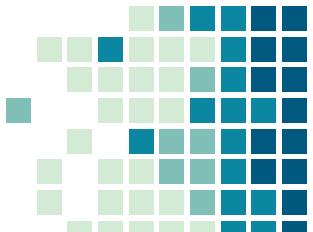
## Catalogue of Life



FLORA DO BRASIL 2020 - ALGAS, FUNGOS E PLANTAS



# Qualidade taxonômica



FLORA 2020

Plantminer

 FLORA DO BRASIL 2020 - ALGAS, FUNGOS E

**LIMPAR FILTRO** **CONSULTAR**

**Nome**

Grupo: Angiospermas | Família: todos  
Gênero: Espécie:   
Autor: Nome Popular:   
Nome Completo:  Cactaceae *Cereus macrogonus* K.Schum.

**Descrição** **Imagens**

Apenas Taxons com Descrição Sómente com imagens

**Forma de Vida e Substrato**

Forma de Vida: todos | Substrato: todos

**Abrangência Geográfica**

Ocorre no Brasil: qualquer | Ocorrência: ocorre em:

Endemismo: todos | Origem: todos

**Plantminer**

Select a database:

- Brazilian Flora 2020
- The Plant List
- API & About the data
- Source code on Github

Gustavo Carvalho  
Profile on Github

**Brazilian Flora 2020 - Flora R package frontend**

**1. Name options**

Replace synonyms  
 Correct misspelled names  
 Remove authors (slow)

How conservative the name guessing should be? Lower values are less conservative and may result in incorrect suggestions.

0.7 0.9 1

**4. Processed taxa**

Columns might be automatically removed from display to fit the width of your screen. IDs are links to taxa on the [Brazilian Flora 2020 website](#), which is the source of all data used here. Please cite them accordingly. Threat statuses are determined by [CNC Flora](#) and follow the IUCN convention. **Dataset last updated on 29th Jan 2019**

Show 10 entries Search:

ID	scientific name	original search	search str	taxon status	notes	taxon rank	family	threat status
1	9668 <i>Miconia albicans</i> (Sw.) Triana	<i>Miconia</i> <i>albicans</i>	<i>Miconia</i> <i>albicans</i>	accepted		species	Melastomataceae	
2	10699 <i>Myrcia guianensis</i> (Aubl.) DC.	<i>Myrcya</i> <i>lingua</i>	<i>Myrcia</i> <i>guianensis</i>	accepted	was misspelled replaced synonym	species	Myrtaceae	LC
3	115 <i>Fabaceae</i> Lindl.	<i>Fabaceae</i> sp.1	<i>Fabaceae</i> sp.1	accepted		family	Fabaceae	

Showing 1 to 3 of 3 entries Previous 1 Next

**2. Life form & more**

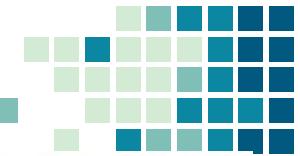
Life form  
 Habitat  
 Vegetation type  
 Occurrence  
 Show common names  
 Establishment

**5. Download**

Comma-delimited  Semicolon-delimited  Tab-delimited  Phylomatic taxa

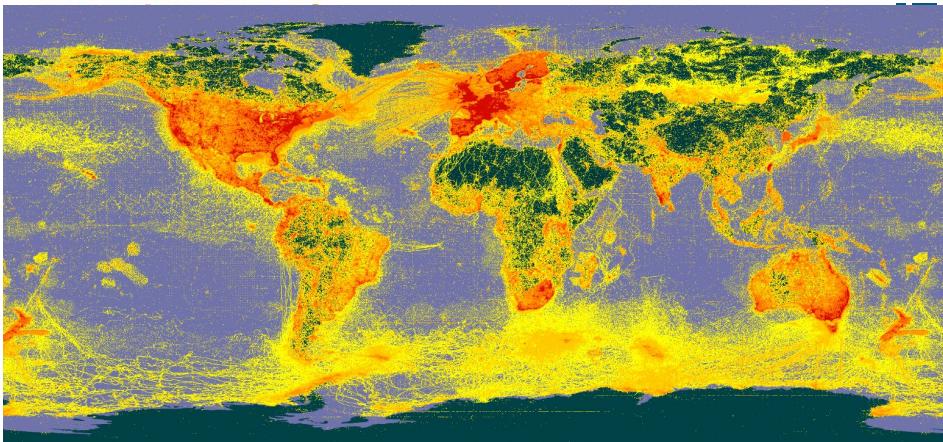
Gustavo Carvalho - Laboratório de Fenologia, Unesp Rio Claro, Brasil  
gustavo.bio@gmail.com

# Qualidade de georeferenciamento

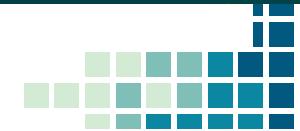
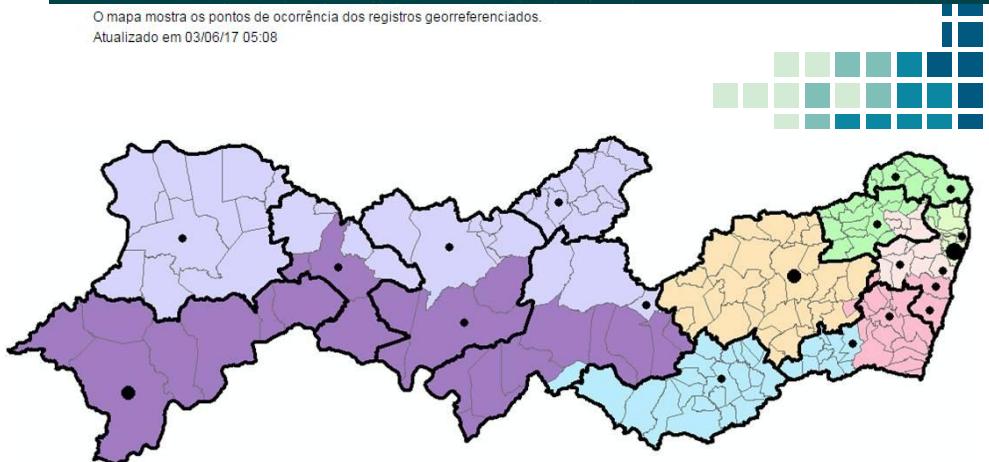


## Precisão

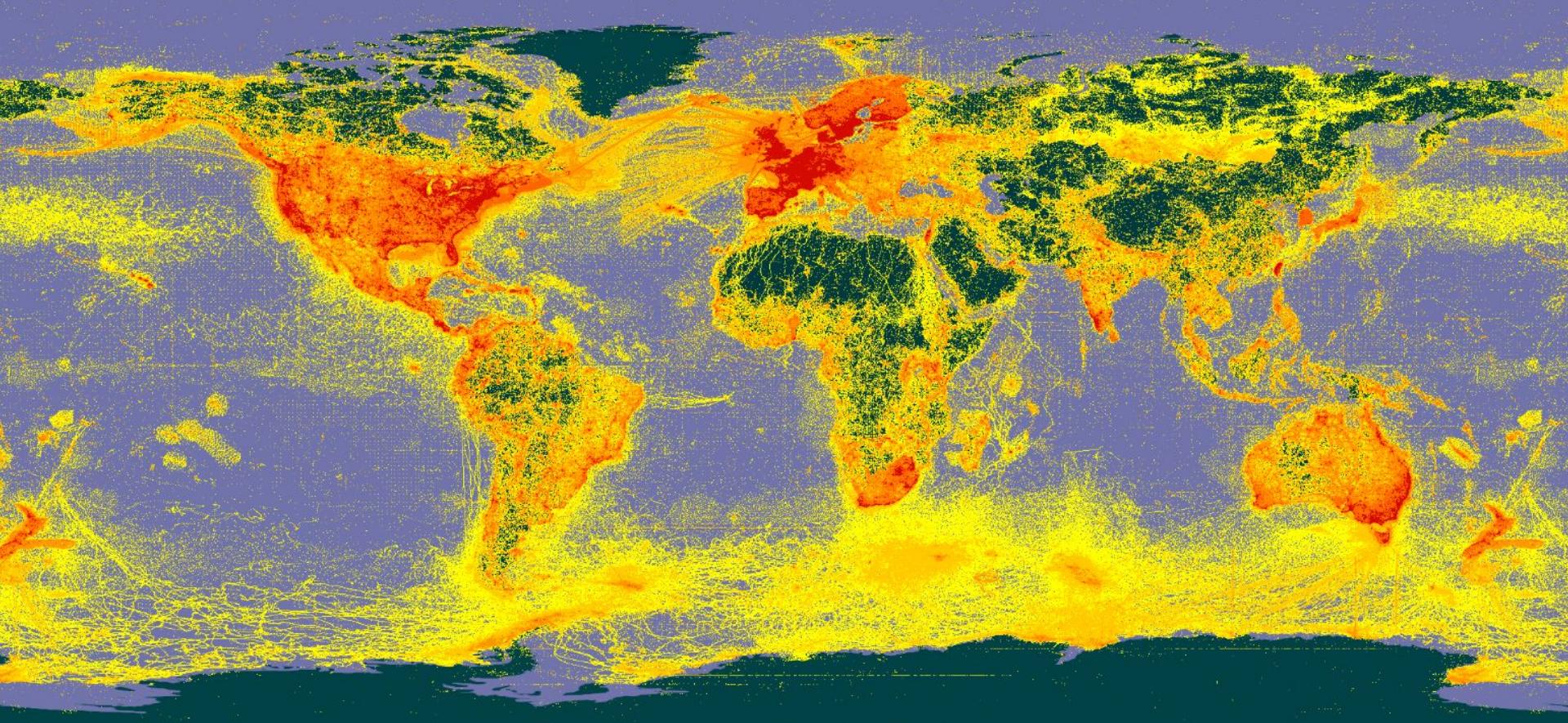
- Dados próprios (coleta de campo, GPS)
  - Georreferenciados *a posteriori* (literatura, rede)
- 
- Ponto (indivíduo coletado)
  - Área (fragmento de vegetação)
  - Localidade (fazenda, UCs, bairro, estrada)
  - Município



O mapa mostra os pontos de ocorrência dos registros georreferenciados.  
Atualizado em 03/06/17 05:08



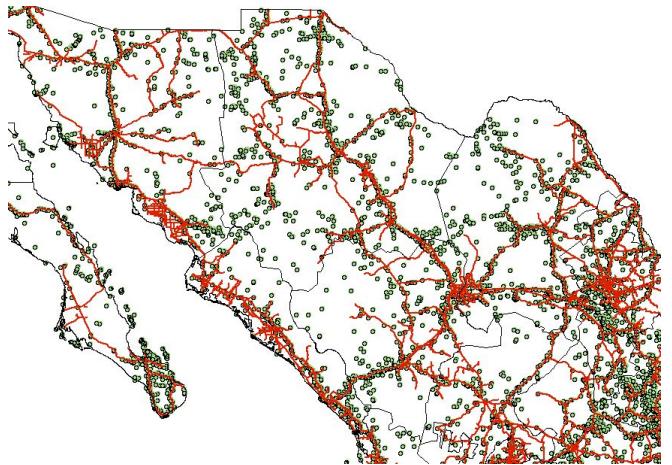
# Qualidade de georeferenciamento.



# Principais vieses de amostragem

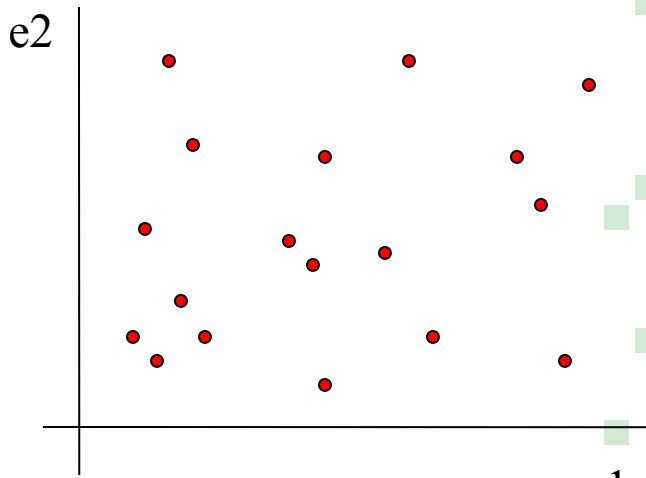
## Efeito museu / Estradas / Rios

Visualização das coletas no espaço geográfica/espacial

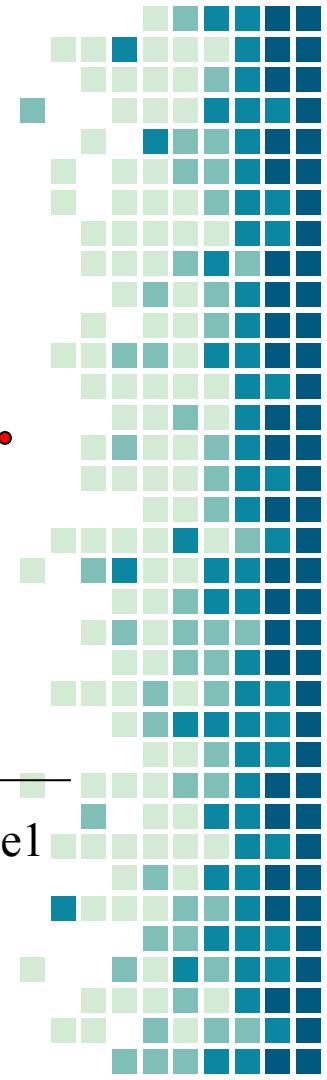


Espaço geográfico

Visualização das coletas no espaço ambiental



Realizar uma análise prévia dos dados para verificar se as informações dos pontos de ocorrência estão bem distribuídas no espaço geográfico e ambiental gerando tabelas cruzando os pontos de ocorrência (lat, lon) e as variáveis ambientais utilizadas (temp, prec, elev etc).





# Viés de amostragem

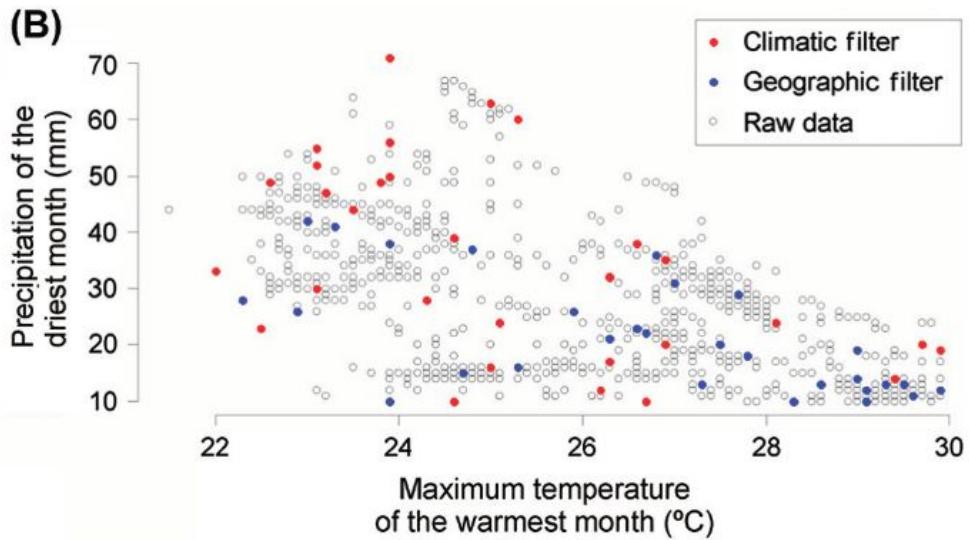
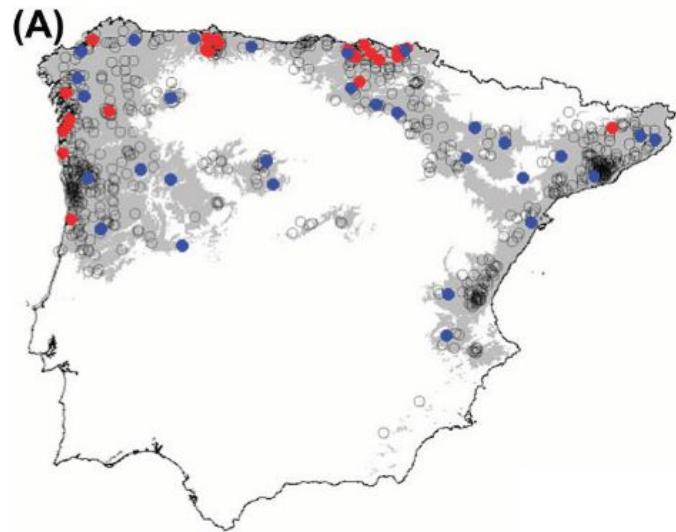


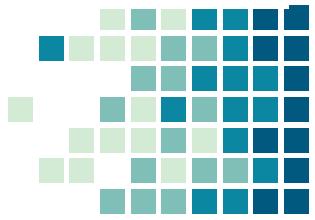
Figure 4. Differences between geographic (blue points) and climatic (red points) filters applied to the data sets in geographic space (A) and in environmental space (B), the latter shown for the two climatic variables used in the environmental filtering. Here, we provide examples for one experiment (bias of the raw data set: distance to populated areas; sample size for the filtered data set: 30 points).

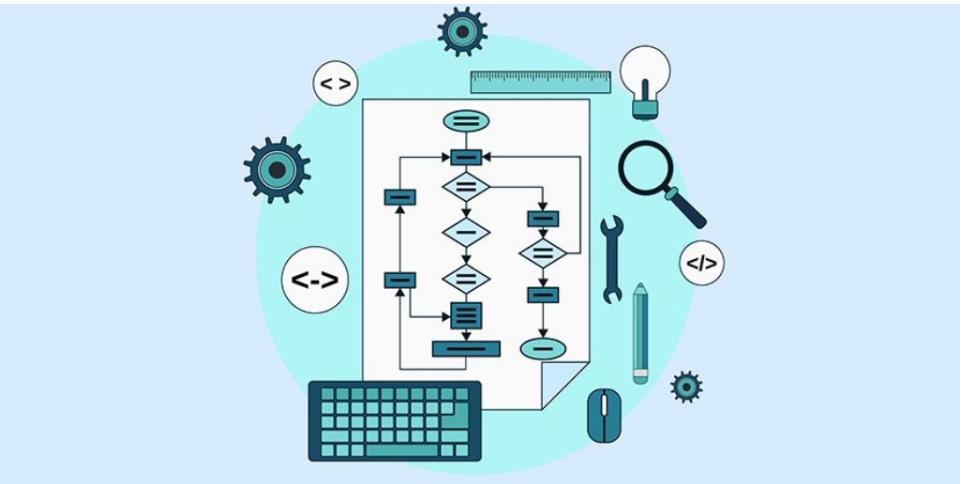
Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models

Sara Varela, Robert P. Anderson, Raúl García-Valdés and Federico Fernández-González

*Ecography* 37: 1084–1091, 2014

doi: 10.1111/j.1600-0587.2013.00441.x



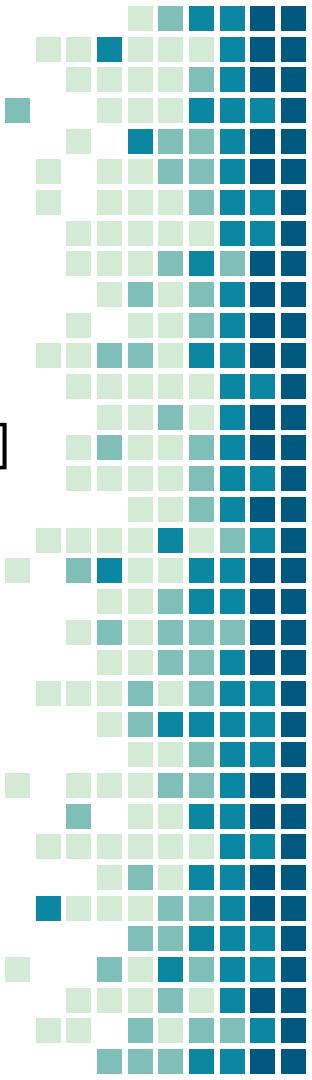


# Registros de ocorrência

Limpeza de dados automatizada



# Data Cleaning



## Pacotes:

### Species data cleaning



Documentation for package 'spfilt' version 1.1

- [DESCRIPTION file](#).

### Help Pages

[Eugenia\\_aurata](#) Eugenia aurata occurrence data.

[example](#) Rasetr South America.

[filt](#) Mark occurrences with the name informed different from the coordinate in a user supplied vector file

[filt\\_generic](#) Marks occurrences based on user supplied vector file

[Leucaena\\_leucocephala](#) Leucaena leucocephala occurrence data.

[remove\\_dups](#) Remove duplicated and NAs

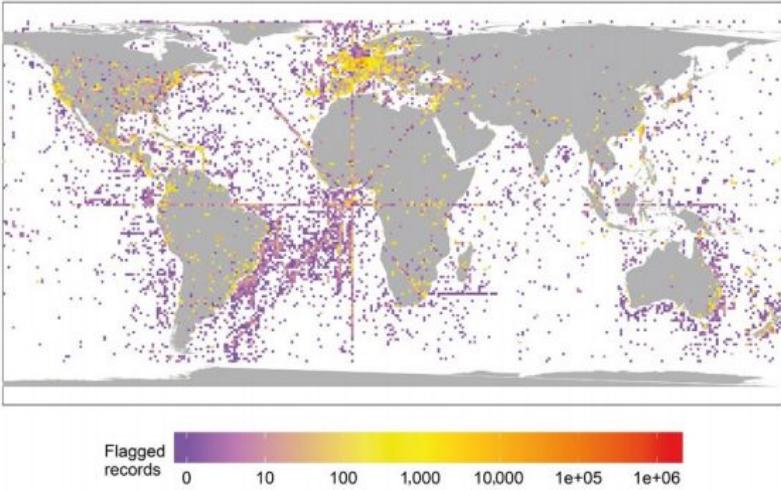
[remove\\_pts](#) Removing species with few records.

[world](#) World map.

Município; Estado; País ...

- Coordenadas 0,0 [[Null Island](#)]
- Longitude e Latitude invertidas
- Sinal trocado
- ( -, - ; +, - ; -, + )

(a)



Received: 26 March 2018 | Accepted: 21 December 2018

DOI: 10.1111/2041-210X.13152

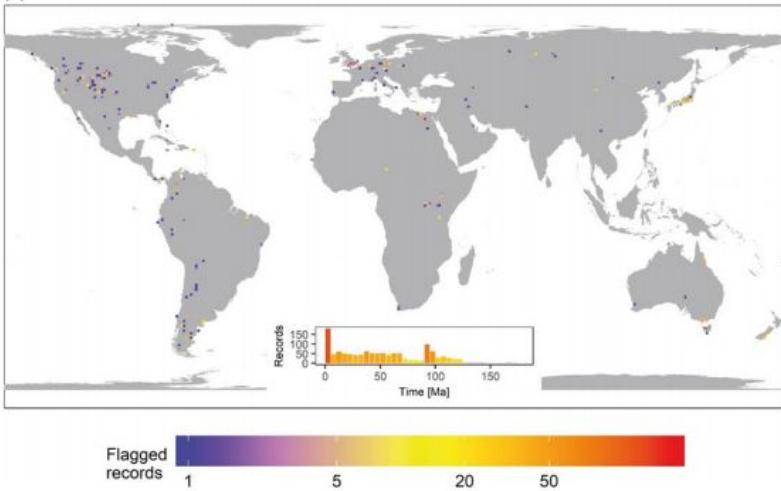
**APPLICATION****COORDINATECLEANER: Standardized cleaning of occurrence records from biological collection databases**

Alexander Zizka<sup>1,2,3</sup> | Daniele Silvestro<sup>1,2,4</sup> | Tobias Andermann<sup>1,2</sup> | Josué Azevedo<sup>1,2</sup> | Camila Duarte Ritter<sup>1,2,5</sup> | Daniel Edler<sup>1,2,6</sup> | Harith Farooq<sup>1,2,7,8</sup> | Andrei Herdean<sup>1</sup> | María Ariza<sup>9</sup> | Ruud Scharn<sup>2,10</sup> | Sten Svantesson<sup>1</sup> | Niklas Wengström<sup>1</sup> | Vera Zizka<sup>11</sup> | Alexandre Antonelli<sup>1,2,12</sup>

<sup>1</sup>Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden; <sup>2</sup>Göteborg Global Biodiversity Centre, Göteborg, Sweden; <sup>3</sup>German Center for Integrative Biodiversity Research (iDiv), Leipzig, Germany; <sup>4</sup>Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; <sup>5</sup>Department of Eukaryotic Microbiology, University of Duisburg-Essen, Essen, Germany; <sup>6</sup>Integrated Science Lab, Department of Physics, Umeå University, Umeå, Sweden; <sup>7</sup>Departamento de Biología & CESAM, Universidade de Aveiro, Aveiro, Umeå, Portugal; <sup>8</sup>Faculty of Natural Sciences at Lúrio University, Universidade de Aveiro, Pemba, Mozambique; <sup>9</sup>Natural History Museum, University of Oslo, Oslo, Norway; <sup>10</sup>Department of Earth Sciences, University of Gothenburg, Göteborg, Sweden; <sup>11</sup>Faculty of Biology, University Duisburg-Essen, Essen, Germany and <sup>12</sup>Göteborg Botanical Garden, Göteborg, Sweden

**FIGURE 1** The number of species occurrence records flagged by COORDINATECLEANER in empirical datasets, per  $100 \times 100$  km grid cell. Warmer colours indicate more flagged records. (a) Flowering plants from the Global Biodiversity Information Facility (c. 91M; Global Biodiversity Information Facility, 2017) (b) Angiosperm fossils from PBDB (c. 19,000; PBDB, 2018). Note the logarithmic scale

(b)



**How to cite this article:** Zizka A, Silvestro D, Andermann T, et al. COORDINATECLEANER: Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol Evol*. 2019;00:1–8. <https://doi.org/10.1111/2041-210X.13152>

**TABLE 1** Geographical and temporal tests implemented in the COORDINATECLEANER package

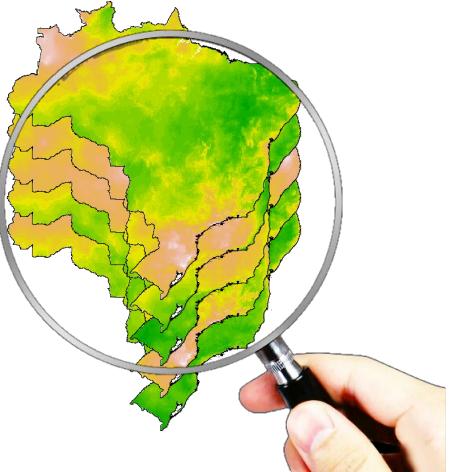
Test function	Level	Flags	Main error source	GBIF (%)	PBDB (%)
cc_cap	REC	Radius around country capitals	Imprecise geo-referencing based on vague locality description	1.1	-
cc_cen	REC	Radius around country and province centroids	Imprecise geo-referencing based on vague locality description	1.8	1
cc_coun	REC	Records outside indicated country borders	Various, e.g. swapped latitude and longitude	-	-
cc_dupl	REC	Records from one species with identical coordinates	Various, e.g. duplicates from various institutions, records from genetic sequencing data	-	-
cc_equ	REC	Records with identical lon/lat	Data entry errors	1.6	1
cc_gbif	REC	Radius around the GBIF headquarters in Copenhagen	Data entry errors, erroneous geo-referencing	0	0
cc_inst	REC	Radius around biodiversity institutions	Cultivated/captured individuals, data entry errors	0.8	0
cc_iucn	REC	Records outside external range polygon	Naturalized individuals, data entry errors	-	-
cc_outl	REC	Geographically isolated records of a species	Various, e.g. swapped latitude and longitude	-	-
cc_sea	REC	Records located within oceans	Various, e.g. swapped latitude and longitude	0.1	-
cc_urb	REC	Records from within urban areas	Cultivated individuals, old records	-	-
cc_val	REC	Records outside lat/lon coordinate system	Data entry errors, e.g. wrong decimal delimiter	0	0
cc_zero	REC	Plain zeros in the coordinates and a radius around (0/0)	Data entry errors, failed geo-referencing	1.6	0.01
cd_ddmm	DS	Over proportional drop of records at 0.6	Erroneous conversion from dd.mm to dd.dd	4.1% datasets	-
cd_round	DS	Decimal periodicity or over proportional number of zero decimals	Rasterized or rounded data	18.5% datasets	-
cf_age	FOS/REC	Temporal outliers in fossil age or collection year	Various	-	-
cf_equal	FOS	General time validity	Data entry errors	-	0
cf_range	FOS	Overly imprecise age ranges	Lack of data	-	3.3
cf_outl	FOS	Outliers in space-time	Data entry error	-	2.1

REC, record-level; DS, dataset-level; FOS, fossil-level; dd.mm, degree minute annotation; dd.dd, decimal degree annotation; GBIF, Global Biodiversity

**How to cite this article:** Zizka A, Silvestro D, Andermann T, et al. COORDINATECLEANER: Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol Evol.* 2019;00:1–8. <https://doi.org/10.1111/2041-210X.13152>

# Dados Abióticos

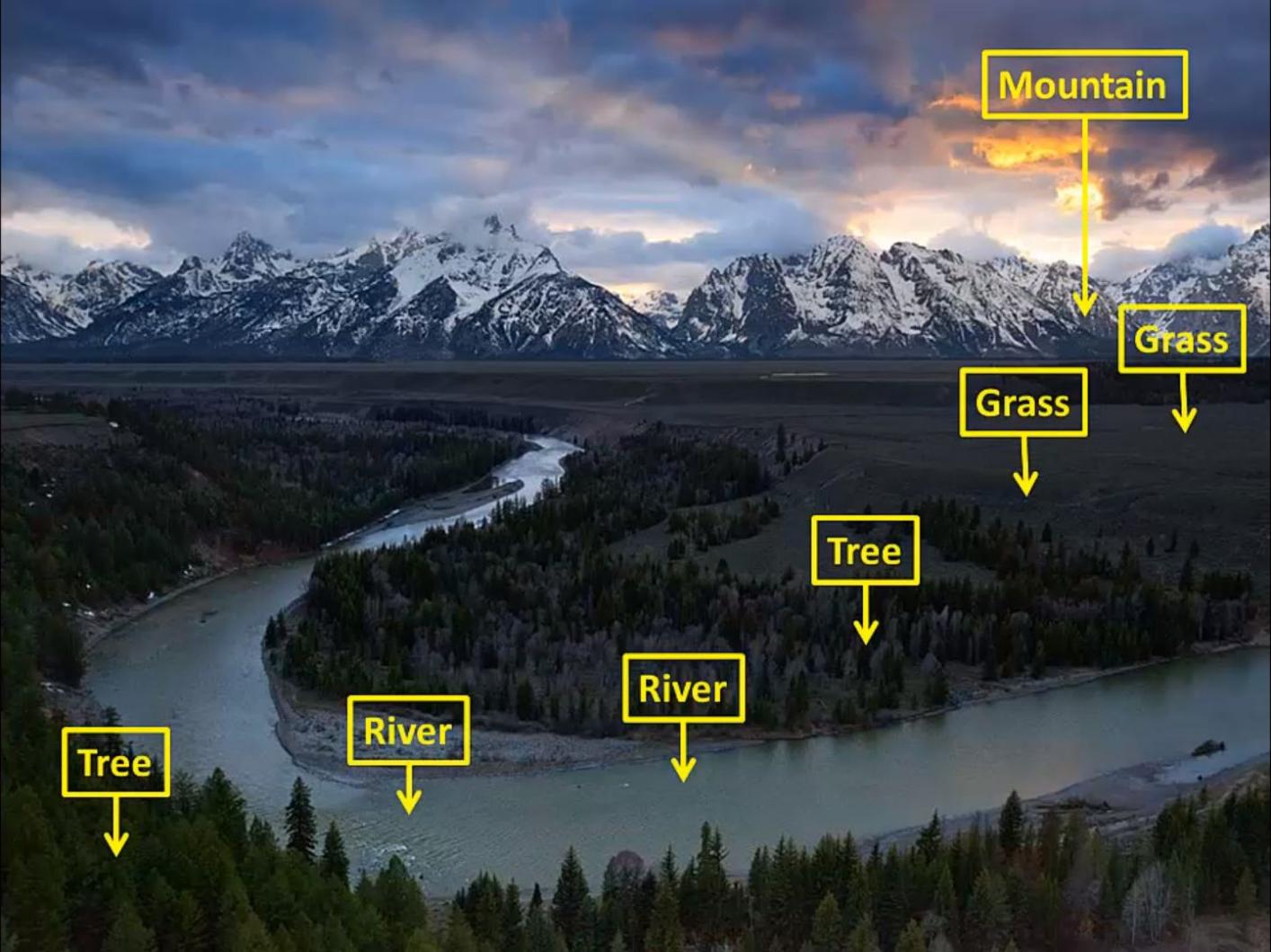
Aquisição



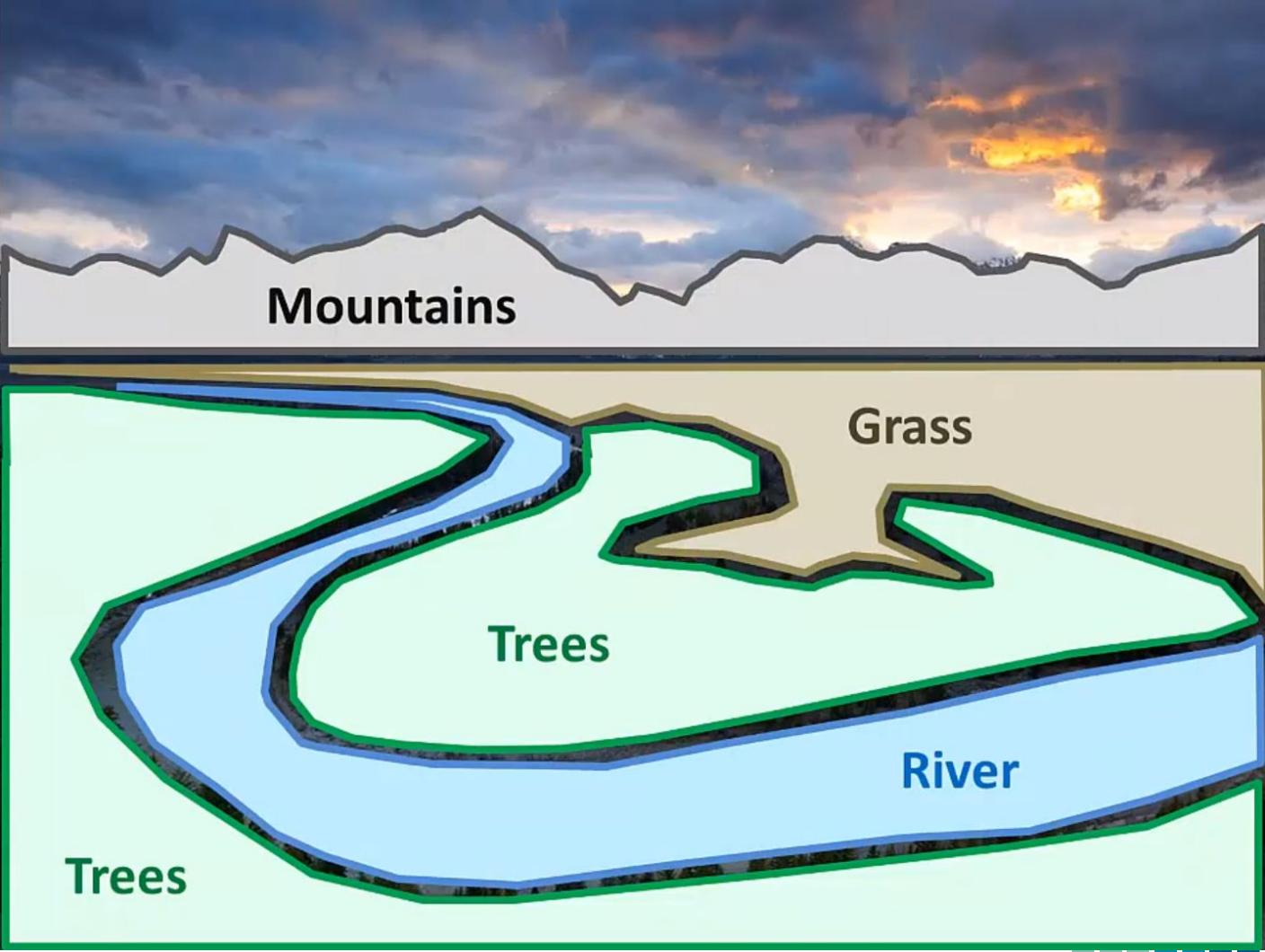
# Tipos de dados espaciais



# Mundo Real

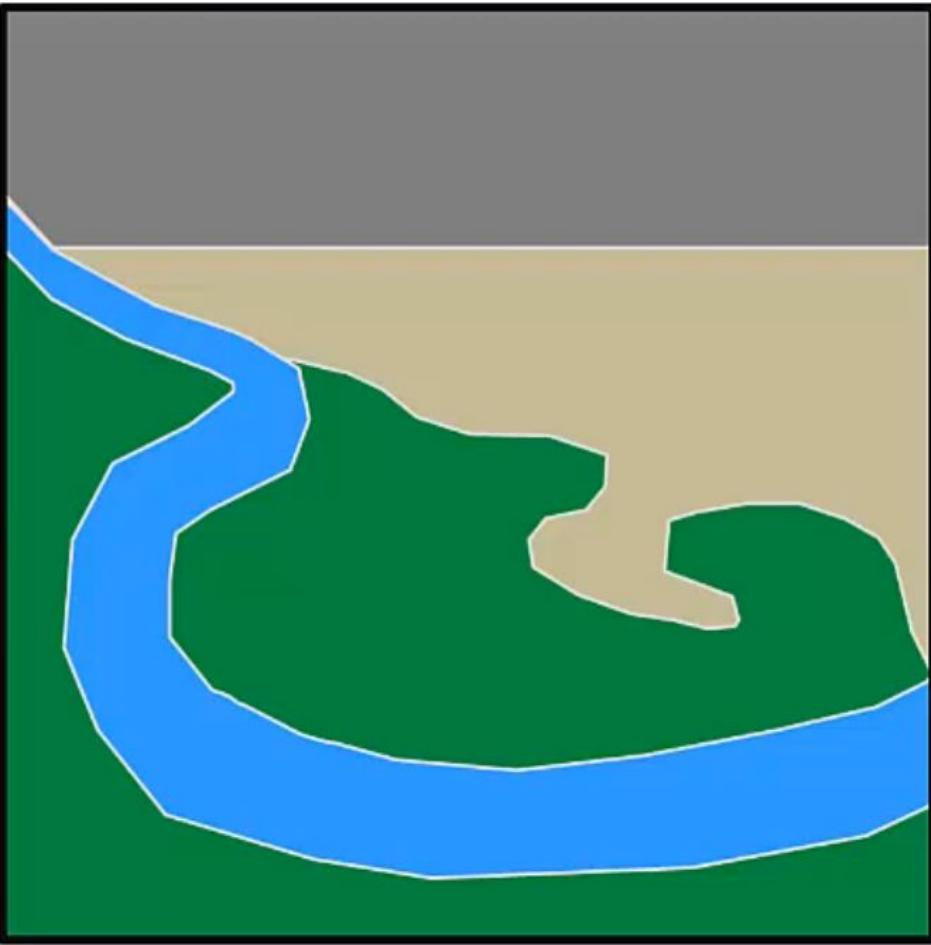


# Representação

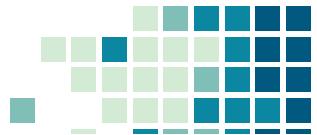
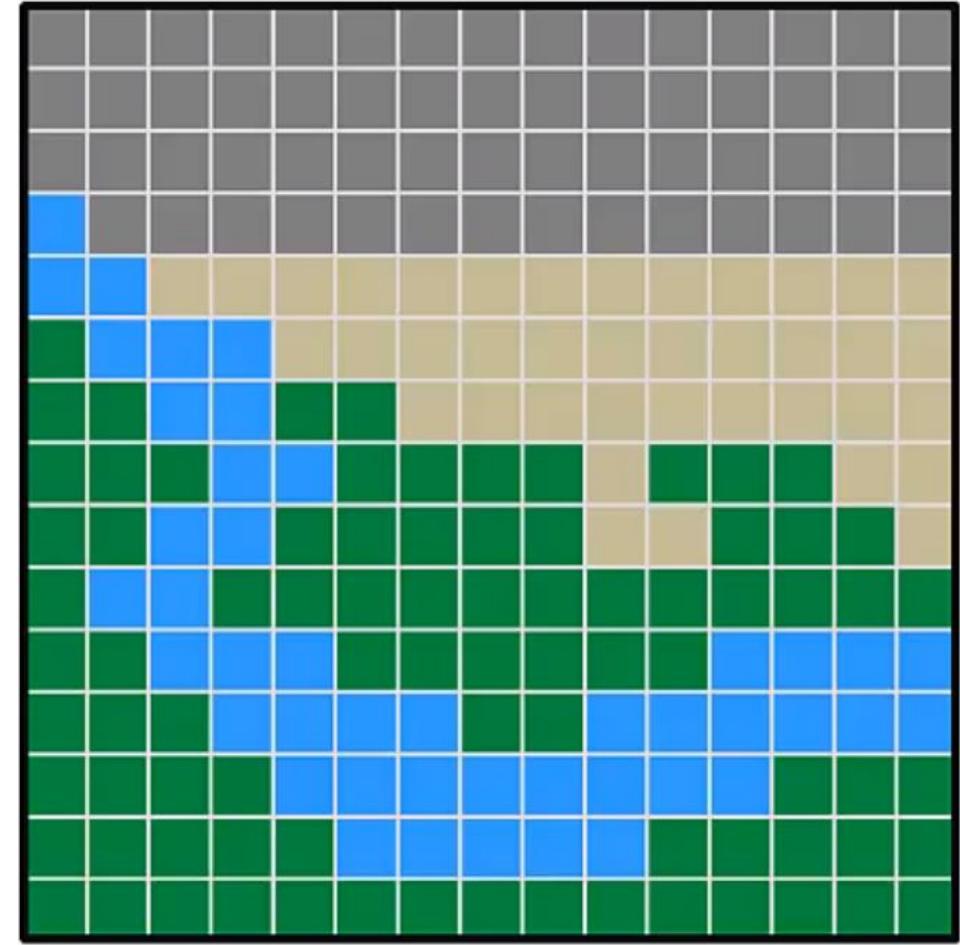


# Tipos de dados espaciais

Vetorial

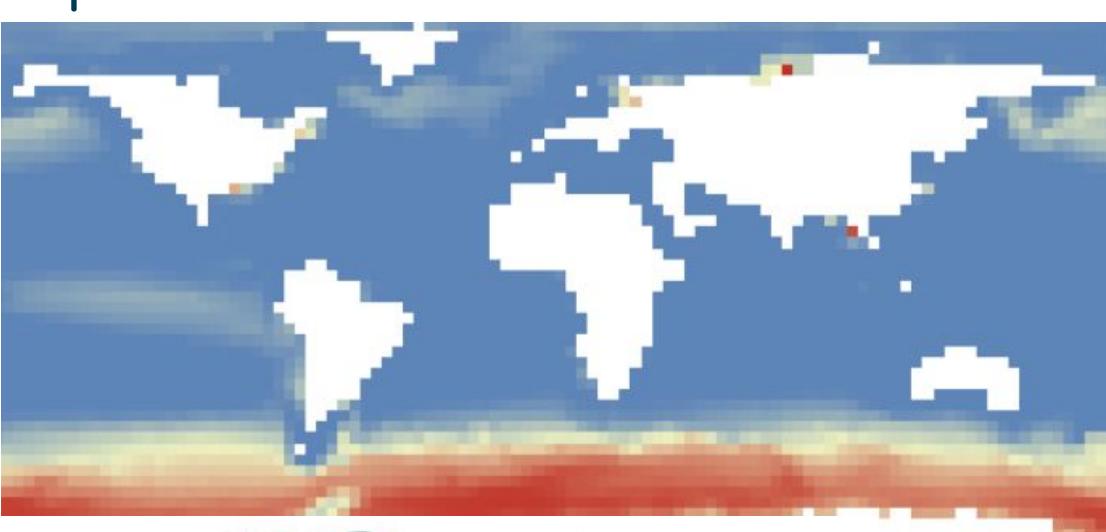


Matricial

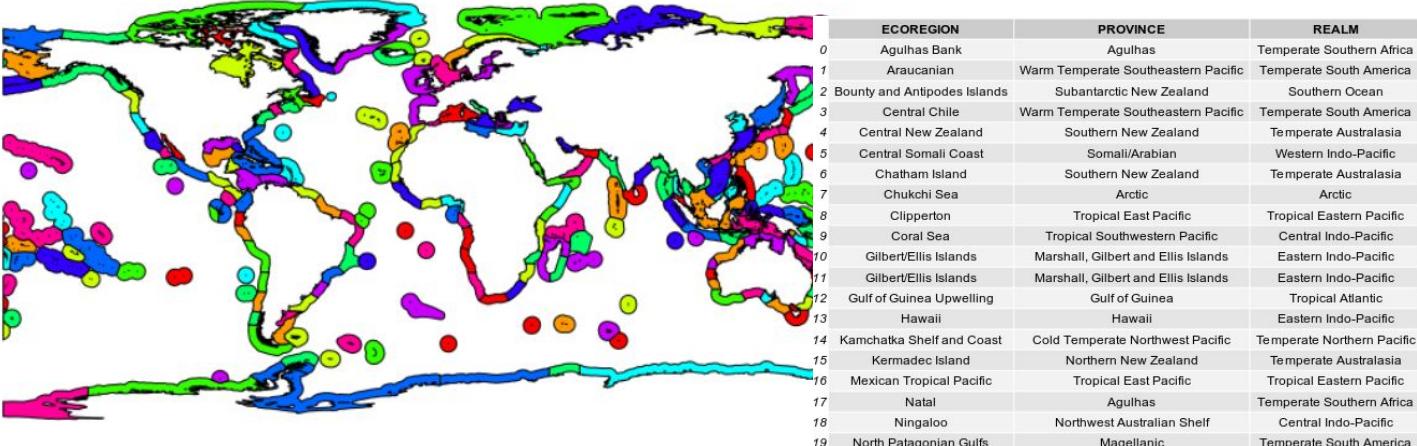


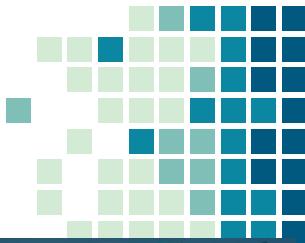
# Tipos de dados espaciais

- Matricial
  - Raster
  - netCDF-4
- Vetorial
  - Shapefile



	4	4	3	3	3
4	4	4	4	1	1
4	3	3	1	1	1
4	3	3	1	1	1
4	3	3	2	2	1
4	4	2	2	2	





# Fonte de dados abióticos

## WorldClim - Global Climate Data

*Free climate data for ecological modeling and GIS*



# Dados Abióticos

Seleção



# Seleção de variáveis

BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality ( $P_2/P_7$ ) (\* 100)

BIO4 = Temperature Seasonality (standard deviation \*100)

BIO5 = Max Temperature of Warmest Month

**BIO6 = Min Temperature of Coldest Month**

BIO7 = Temperature Annual Range ( $P_5-P_6$ )

BIO8 = Mean Temperature of Wettest Quarter

BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

**BIO11 = Mean Temperature of Coldest Quarter**

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month

BIO15 = Precipitation Seasonality (Coefficient of Variation)

BIO16 = Precipitation of Wettest Quarter

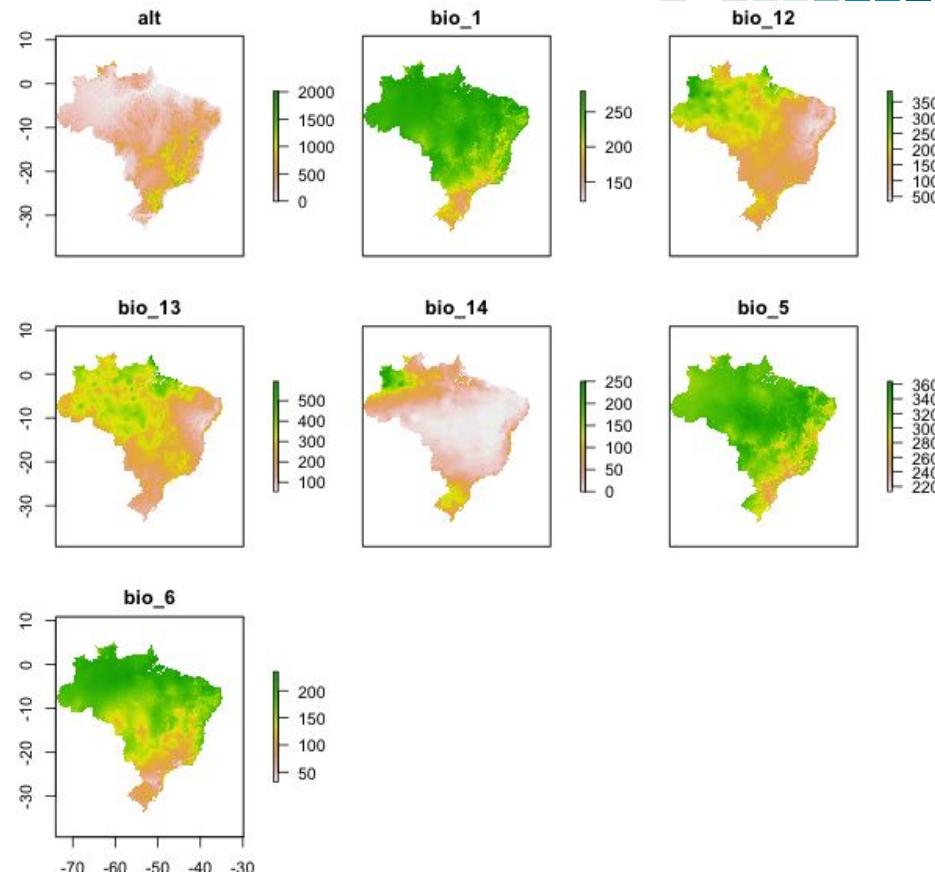
BIO17 = Precipitation of Driest Quarter

BIO18 = Precipitation of Warmest Quarter

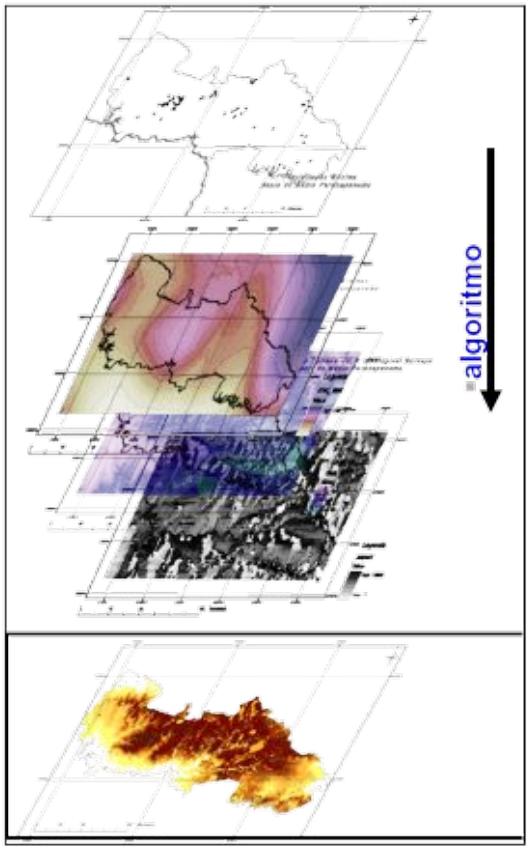
BIO19 = Precipitation of Coldest Quarter

Modelo Digital de Elevação (Altitude)

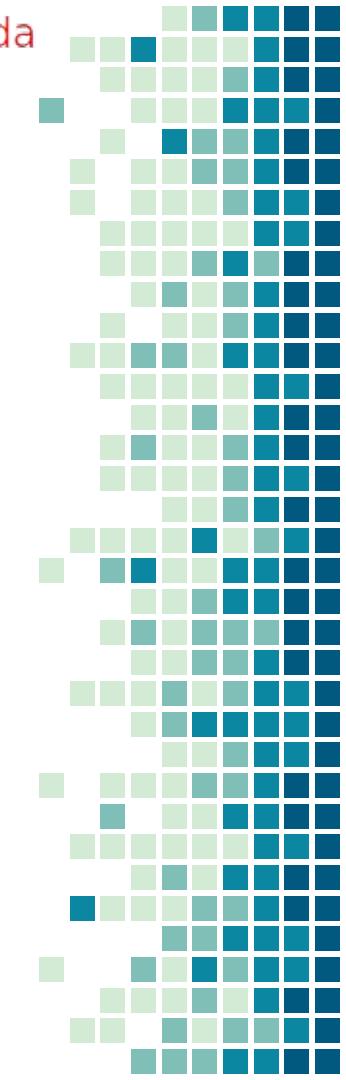
20 camadas



# Exemplo de tabela de dados, com os valores ambientais adicionados, utilizada para se avaliar o nicho



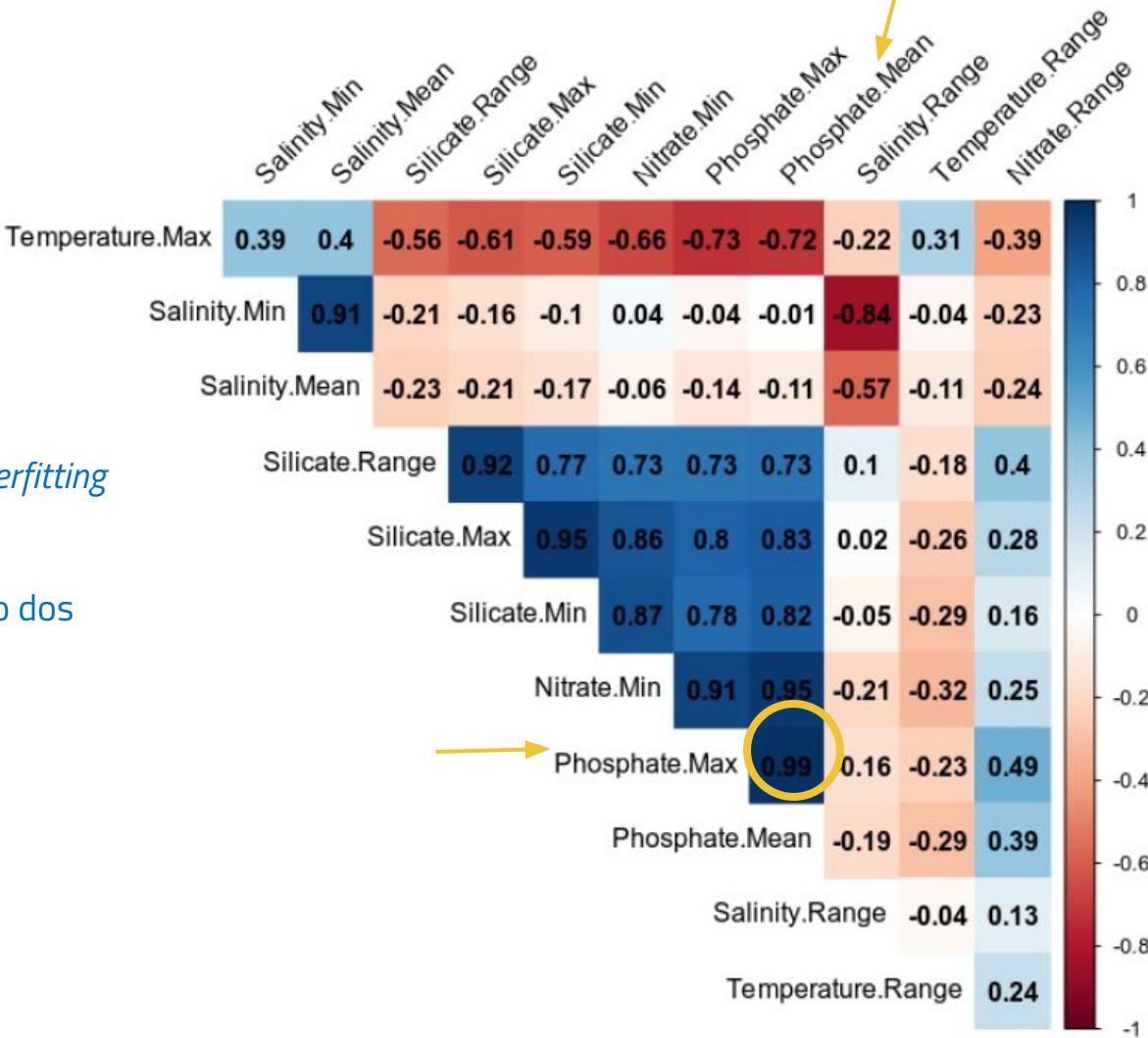
1	Especie	Long	Lat	temp medi	temp max	temp min	prec	elevacao
2	Tibouchina arborea	-44.9933	-23.2458	17.9255	25.9	8.08	1925	966
3	Tibouchina arborea	-44.96	-23.07	16.4255	24.4	6	1521	920
4	Tibouchina arborea	-45.7438	-22.8652	17.6185	25.2	7.37	1770	931
5	Tibouchina arborea	-44.3631	-22.8086	17.0965	25.3903	6.51	1598	1052
6	Tibouchina arborea	-43.04	-22.65	23.1536	30.9	15.44	1284	13
7	Tibouchina arborea	-42.4	-22.5864	23.0058	30.3	15.54	1157	85
8	Tibouchina arborea	-42.385	-22.5864	23.0058	30.3	15.54	1157	85
9	Tibouchina arborea	43.4464	22.5833	20.9221	29.1	12.48	1755	846
10	Tibouchina arborea	43.4464	22.5833	20.9221	29.1	12.48	1755	846
11	Tibouchina arborea	-42.3461	-22.5322	23.0058	30.3	15.54	1157	85
12	Tibouchina arborea	-43.18	-22.51	17.7855	25.6	9.38	1782	925
13	Tibouchina arborea	42.25	22.5	20.8284	28.4806	12.91	1319	217
14	Tibouchina arborea	-44.56	-22.5	17.3771	25.5903	6.5	1706	752
15	Tibouchina arborea	-42.1639	-22.4756	20.8284	28.4806	12.91	1319	182
16	Tibouchina arborea	-42.1639	-22.4756	20.8284	28.4806	12.91	1319	182
17	Tibouchina arborea	-42.65	-22.46	19.6895	27.7903	11.13	1515	421
18	Tibouchina arborea	-43.005	-22.4586	18.8675	27.1	9.95	1653	981
19	Tibouchina arborea	-42.8333	-22.4333	19.6895	27.7903	11.13	1515	110
20	Tibouchina arborea	-42.57	-22.36	18.212	26.2806	9.81	1501	1892
21	Tibouchina arborea	-41.9917	-21.9483	19.5472	27.5	10.46	1374	712
22	Tibouchina arborea	-41.9917	-21.9483	19.5472	27.5	10.46	1374	712
23	Tibouchina arborea	-41.1017	-20.5269	20.3125	28.671	11.6194	1265	670
24	Tibouchina arborea	-41.1017	-20.5269	20.3125	28.671	11.6194	1265	670
25	Tibouchina arborea	-41.075	-20.5175	20.3125	28.671	11.6194	1265	670
26	Tibouchina arborea	40.615	-20.4123	22.9451	30.3806	15.8226	1183	283
27	Tibouchina arborea	-40.6592	-20.3633	20.3335	28.171	12.3194	1278	652
28	Tibouchina arborea	-40.66	-20.36	20.3335	28.171	12.3194	1278	652
29	Tibouchina arborea	-40.4819	-20.3025	24.3821	31.4806	17.92	1139	461
30	Tibouchina arborea	40.5508	-20.1342	21.3829	28.9806	13.8226	1253	728
31	Tibouchina arborea	-40.8603	-20.0717	19.109	27.171	10.4194	1343	915
32	Tibouchina arborea	-40.8603	-20.0717	19.109	27.171	10.4194	1343	915
33	Tibouchina arborea	-40.5964	-20.0542	21.6286	29.2806	13.9258	1253	478
34	Tibouchina arborea	-40.5422	-20.0125	20.8205	28.4806	12.9226	1291	748



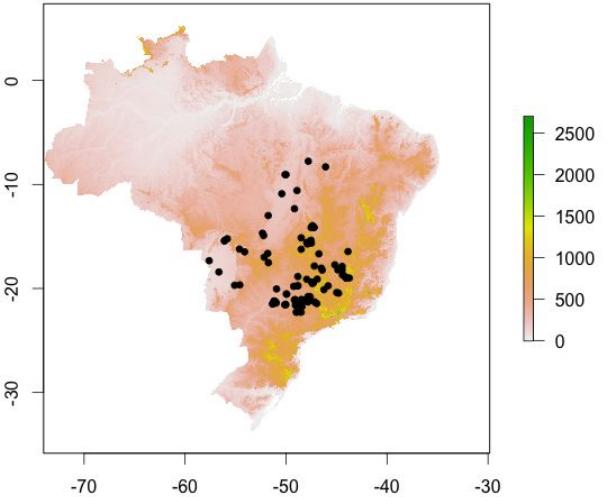
# Correlação entre os dados espaciais

## Consequência:

- Aumentamos a chance de ocorrer um *overfitting* (sobre-ajuste) do modelo. Modelos cujos resultados refletem a própria distribuição dos pontos e nada mais.



# Extração de variáveis ambientais nas ocorrências



	alt	bio_1	bio_12	bio_13	bio_14	bio_5	bio_6
[1.]	495	232	1282	244	13	309	121
[2.]	1286	212	1826	355	4	284	137
[3.]	921	198	1523	282	19	273	88
[4.]	409	259	1962	344	5	342	174
[5.]	320	259	1602	287	2	341	161
[6.]	563	207	1210	216	26	293	100