

# Introdução à modelagem de distribuição de espécies usando a linguagem R

Mastozóologos Organizados em uma Conferência Online (MOCÓ)

---

Maurício H. Vancine

UNESP - Rio Claro

Laboratório de Ecologia Espacial e Conservação (LEEC)

25-26/09/2021

Primeiramente, gostaria de agradecê-los pela presença na oficina, sejam muito bem-vindas e bem-vindos!

Gostaria de agradecer também ao MOCÓ pela oportunidade de oferecer essa oficina!

# Agradecer especialmente ao João Giovanelli



# E por fim, agradecer aos meus professores e colegas:

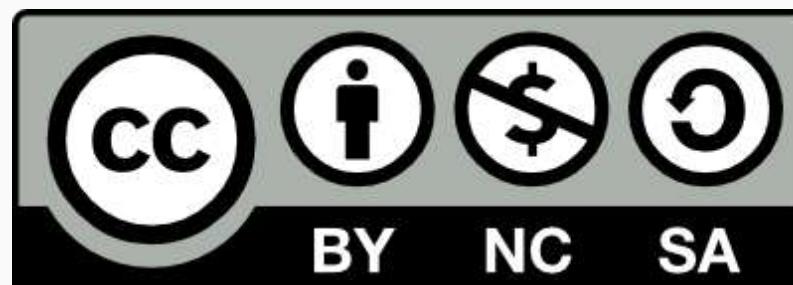
- Prof. Milton Cezar Ribeiro
- Prof. Tadeu de Siqueira Barros
- Flavia Pinto
- Vinicius Tonetti
- Prof. Thadeu Sobral-Souza
- Natalia Stefanini da Silveira
- João Giovanelli
- Prof. Thiago Sanna
- Renata Muylert
- Julia Oshima
- Carol Bello
- Prof. Matheus Lima-Ribeiro
- Profa. Carina Terrible
- Priscila Lemes
- Pablo Hendrigo Alves de Melo
- Bruna de França Gomes

# Licença

## Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)

- **Atribuição** — Você deve dar o crédito apropriado, prover um link para a licença e indicar se mudanças foram feitas. Você deve fazê-lo em qualquer circunstância razoável, mas de nenhuma maneira que sugira que o licenciante apoia você ou o seu uso
- **NãoComercial** — Você não pode usar o material para fins comerciais
- **Compartilhagual** — Se você remixar, transformar, ou criar a partir do material, tem de distribuir as suas contribuições sob a mesma licença que o original.

Saiba mais em: [Creative Commons](#)



# 1 Apresentações

# 1.1 Apresentações

## Maurício Vancine

Ecólogo (2014) | Mestre em Zoologia (2018) | Doutorando em Ecologia (2020)

### Pesquisa

- Ecologia Espacial
- Modelagem Ecológica
- Ecologia e Conservação de Anfíbios

### Habilidades

- Modelos de Distribuição de Espécies (SDMs)
- Análise de Dados Ecológicos e Geoespaciais
- *Open source* [R, QGIS, GRASS GIS, GNU/Linux, ...]

### Informações:

[mauriciovancine.github.io](https://mauriciovancine.github.io)

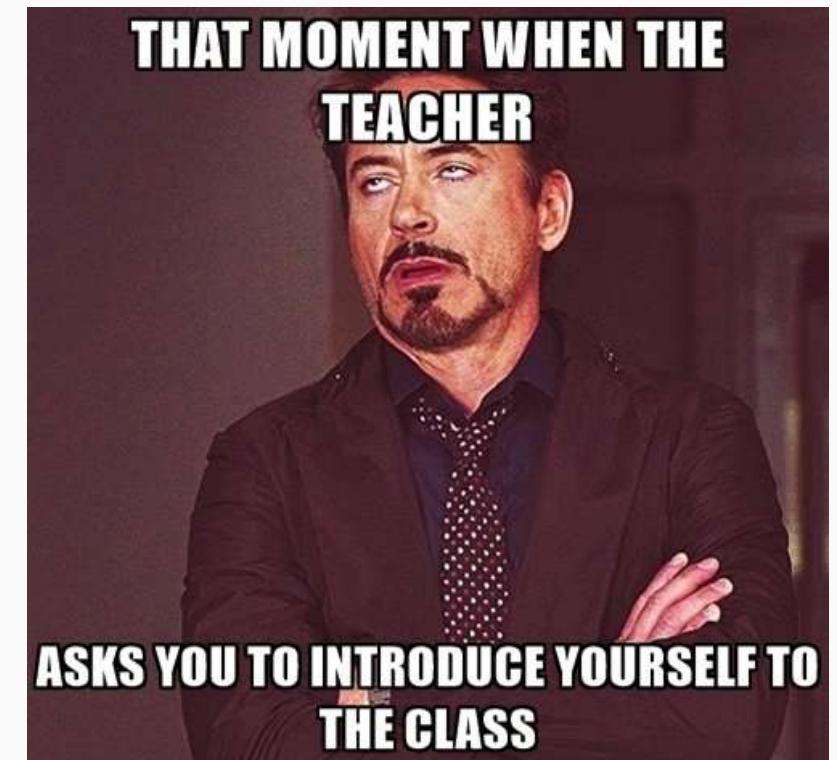


# 1.1 Apresentações

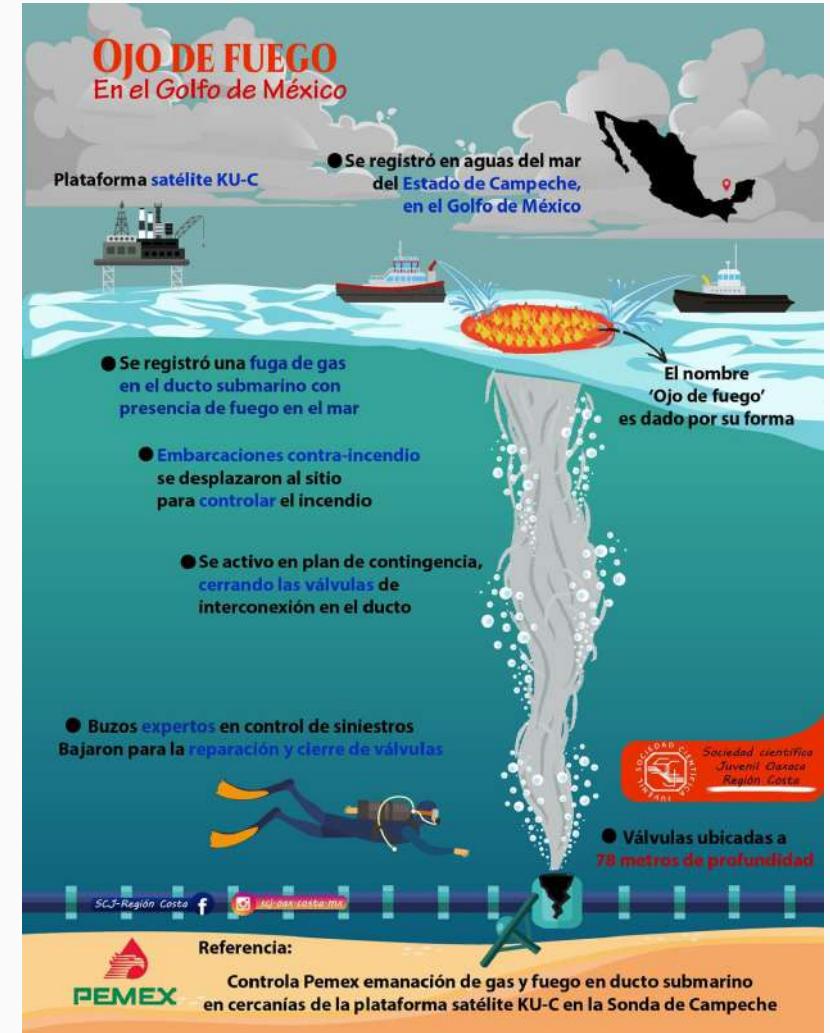
Participantes (~10 segundos)

Pode abrir a câmera e microfone, se puder e quiser =]

1. Nome
2. Formação
3. O que faz ou pensa em fazer da vida?
4. Conhece Modelos de Distribuição de Espécies?
5. O que espera desta oficina?
6. O que descobriram nessa pandemia?



# O oceano pode pegar fogo...



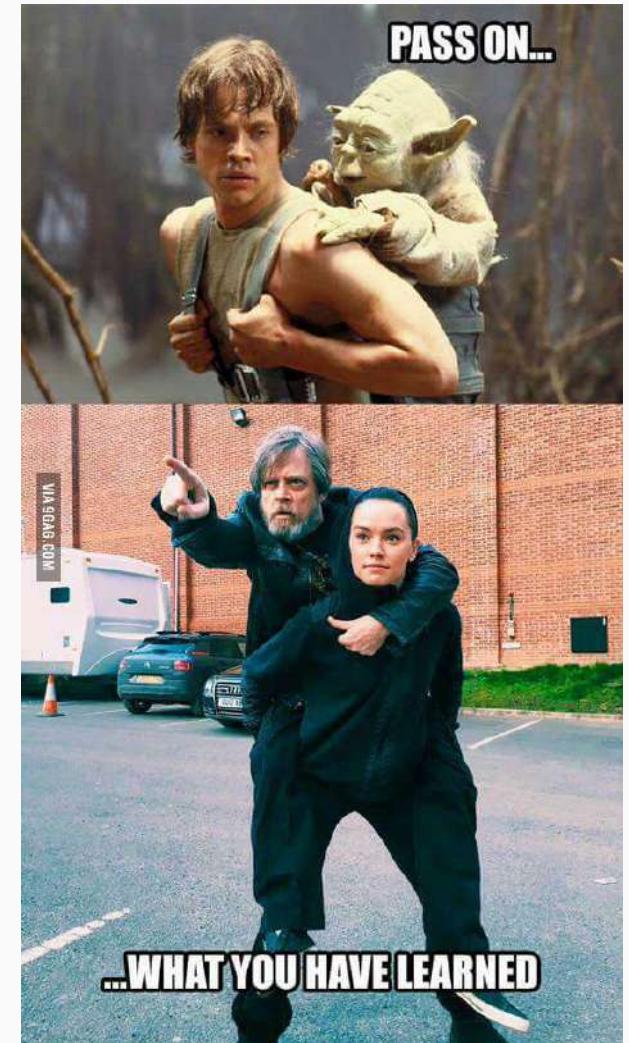
# 1.2 Oficina

## Carga horária

- 06 horas

## Dinâmica

- Tudo vai ser muito rápido
- O principal intuito é fazer uma introdução ao tema
- Vou indicar material (mas muito mesmo...)



## 1.2 Oficina

Espaço seguro e amigável

Sintam-se à vontade para me interromper e tirar dúvidas

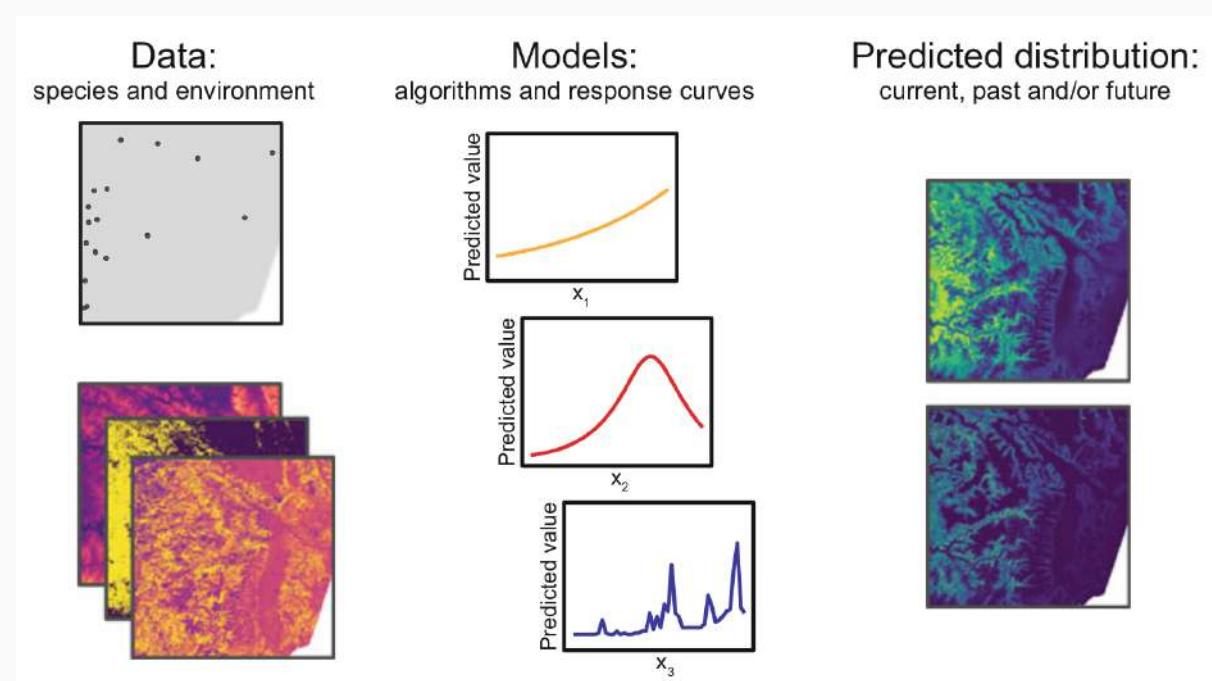


# 1.2 Oficina

## Conteúdo

### 2 Teoria de nicho ecológico e distribuição de espécies (90 min)

1. Visão geral dos Modelos de Distribuição de Espécies
2. Nicho ecológico e distribuição de espécies
3. Tipos de modelos
4. Padronização para criação dos modelos
  - 4.1. Conceituação
  - 4.2. Preparação dos dados
  - 4.3. Ajuste dos modelos
  - 4.4. Avaliação dos modelos
  - 4.5. Predições no tempo e no espaço



[Fletcher & Fortin \(2018\)](#), [Zurell et al. \(2020\)](#)

# 1.2 Oficina

## Conteúdo

### 3 Introdução à linguagem R (20 min)

1. Linguagem R
2. RStudio
3. Console
4. Scripts
5. Operadores
6. Objetos
7. Funções
8. Pacotes
9. Ajuda
10. Diretório de trabalho
11. tidyverse (pipe - %>%)

[R](#), [RStudio](#)

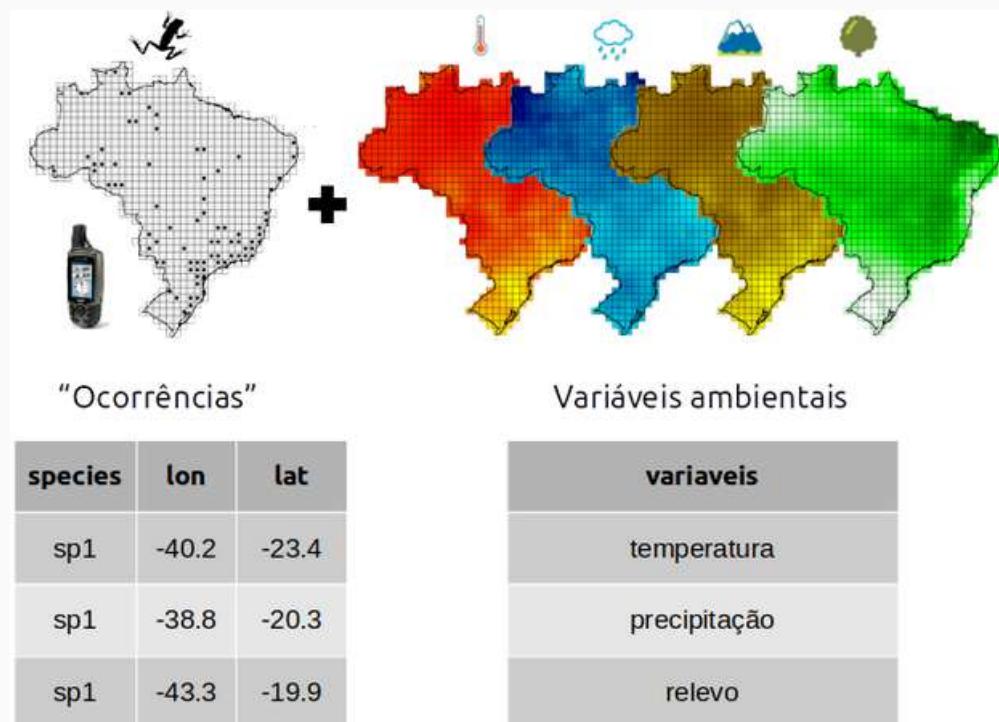


# 1.2 Oficina

## Conteúdo

### 4 Preparação dos dados (ocorrências e variáveis) (60 min)

1. Dados geoespaciais
2. Dados de ocorrências
3. Bases de dados de ocorrências
4. Download de dados de ocorrência
5. Limpeza de dados de ocorrência
6. Dados ambientais
7. Base de dados ambientais
8. Download de dados ambientais
9. Colinearidade de dados ambientais

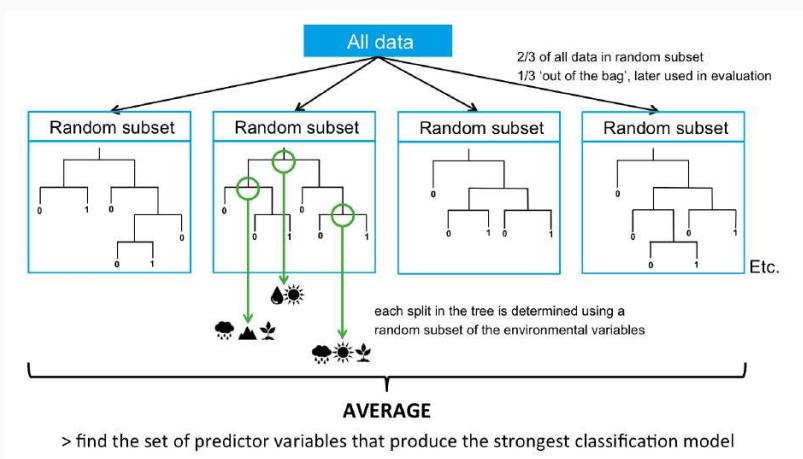
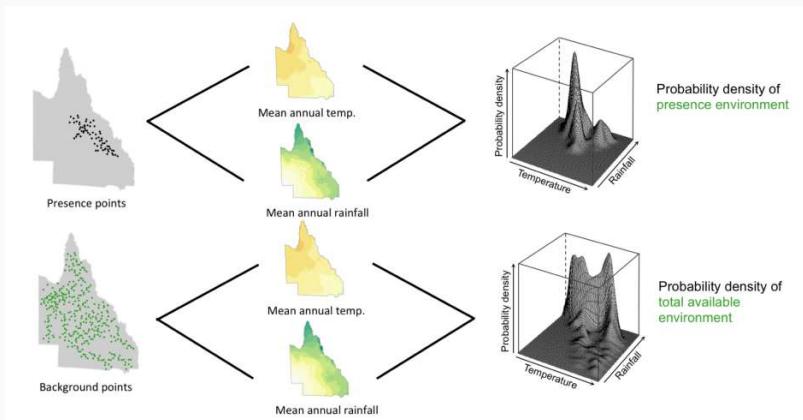
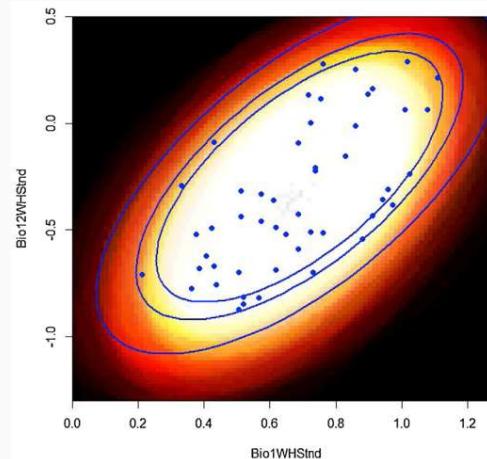
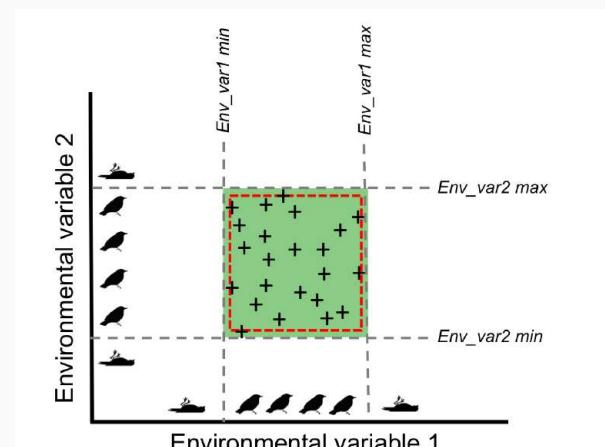


# 1.2 Oficina

## Conteúdo

### 5 Ajuste dos modelos (30 min)

1. Tipos de métodos ou algoritmos
2. Somente-presença (BIOCLIM, DOMAIN e Mahalanobis)
3. Presença-(pseudo)ausência (GLM, GAM, Random Forest e SVM)
4. Presença-background (MaxEnt)
5. Outros métodos ou algoritmos



# 1.2 Oficina

## Conteúdo

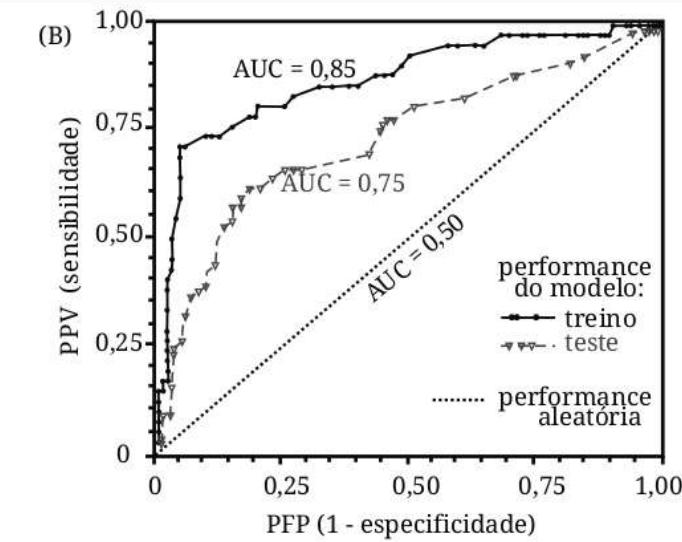
### 6 Avaliação dos modelos (30 min)

1. Partição dos dados (treino e teste)
2. Limiar de corte dos modelos
3. Matriz de confusão
4. Métricas de avaliação

(A)

		OBSERVADO		total	
PREDITO	+	positivo a	falso positivo b		
	-	falso negativo c	negativo d		
		T+	T-	N	

PPV (sensibilidade) =  $a/(a+c)$   
PNV (especificidade) =  $d/(b+d)$   
PFP =  $b/(b+d) = 1 - \text{especificidade}$



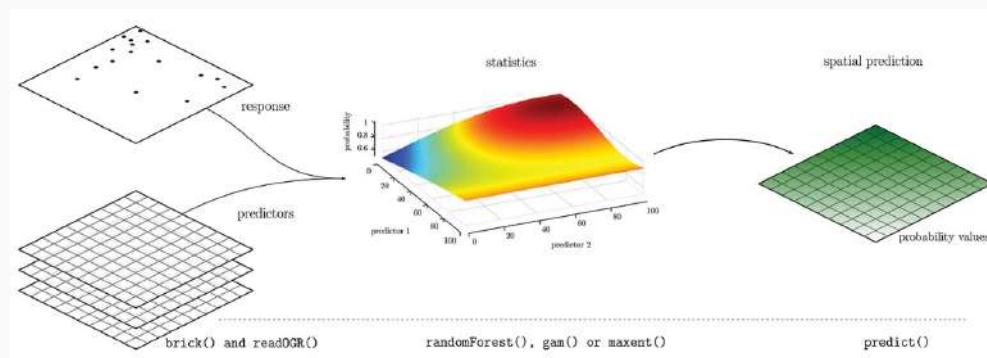
[Lima-Ribeiro & Diniz-Filho \(2013\)](#)

# 1.2 Oficina

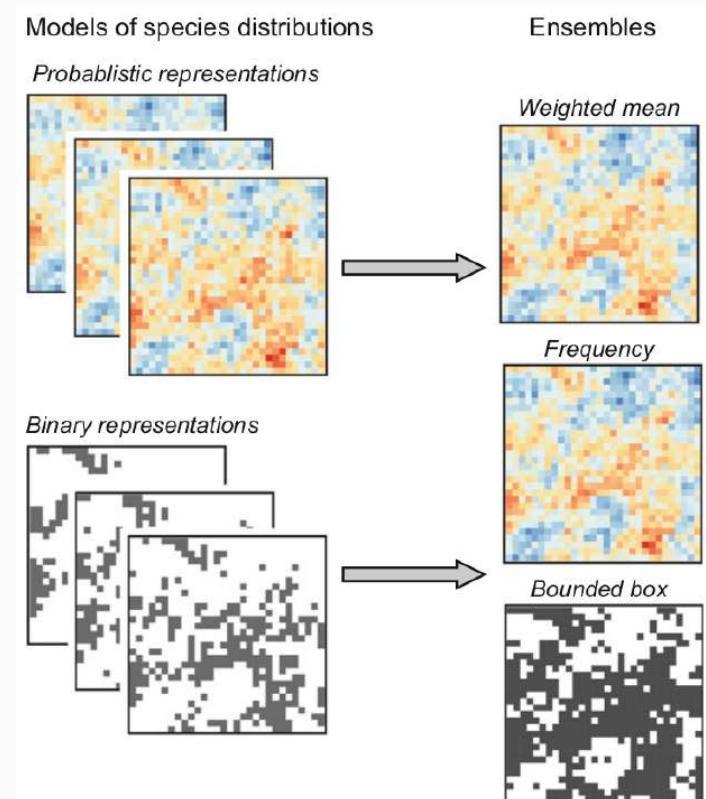
## Conteúdo

### 7 Predição e consenso dos modelos (30 min)

1. Predição dos modelos
2. Consenso dos modelos
3. Consenso por frequênciā
4. Consenso por média
5. Consenso por média ponderada
6. Incertezas



[Wegmann et al. \(2016\)](#), [Fletcher & Fortin \(2018\)](#).

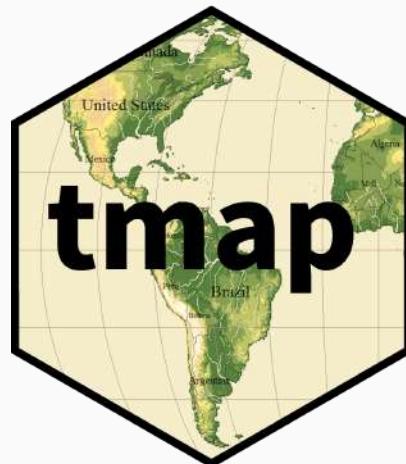
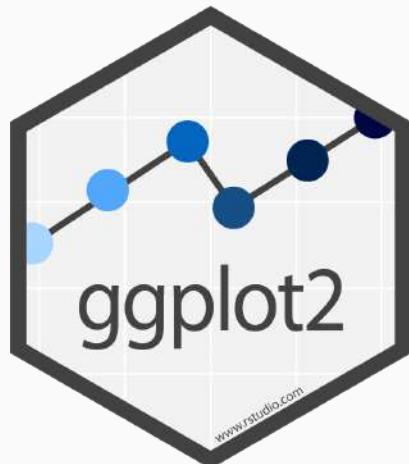


# 1.2 Oficina

## Conteúdo

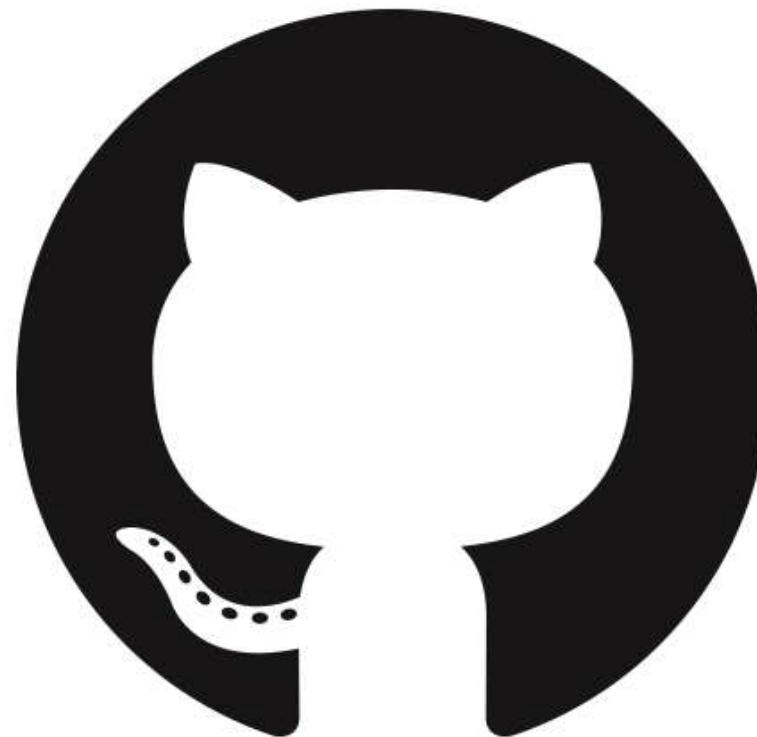
### 8 Composição de mapas (20 min)

1. Composição de mapas
2. Mapas estáticos
3. Mapas interativos



## 1.3 Material

O material da oficina pode ser acessado nesse repositório



[Repositório no GitHub](#)

# 1.4 Slides

Todos os slides estão disponibilizados em .html



[slides](#)

[xaringan](#)

# 1.5 Conferir os computadores



**Software** é aquilo que você xinga.  
**Hardware** é aquilo que você chuta.

Dúvidas?

## 2 Teoria de nicho ecológico e distribuição de espécies

# 2.1 Visão geral dos SDMs

## Terminologia

### 1. Modelos de Distribuição de Espécies (*Species Distribution Models* - SDMs)

- Caracteriza a distribuição da espécie **sem transferir** para outras condições no espaço ou no tempo
- Prediz a distribuição com maior precisão incorporando elementos de dispersão e colonização

### 2. Modelos de Nicho Ecológico (*Ecological Niche Models* - ENMs)

- Estima as preferências ambientais da espécie para **poder transferir** para outras condições no espaço ou no tempo

### 3. Modelos de Adequabilidade de Habitat (*Habitat Suitability Models* - HSMs)

- Reflete melhor a base dos modelos: quantificar a **relação espécie-ambiente** (habitat)
- Modelos nem sempre precisam prever a **distribuição geográfica** da entidade modelada
- Aplicável a entidades que **não sejam espécies** (e.g. genes ou comunidades)
- Aplicável a casos em que um **nicho é parcialmente capturado** em um modelo (área menor que a distribuição total)

Em comum: todos são **modelos**. Mas o que são  
modelos?

## 2.1 Visão geral dos SDMs

Modelos [científicos]

Idealizações da **realidade** e não a própria **realidade**



[Silva & Catelli \(2019\)](#)

# 2.1 Visão geral dos SDMs

## Modelos [científicos]

**Idealizações da realidade e não a própria realidade**

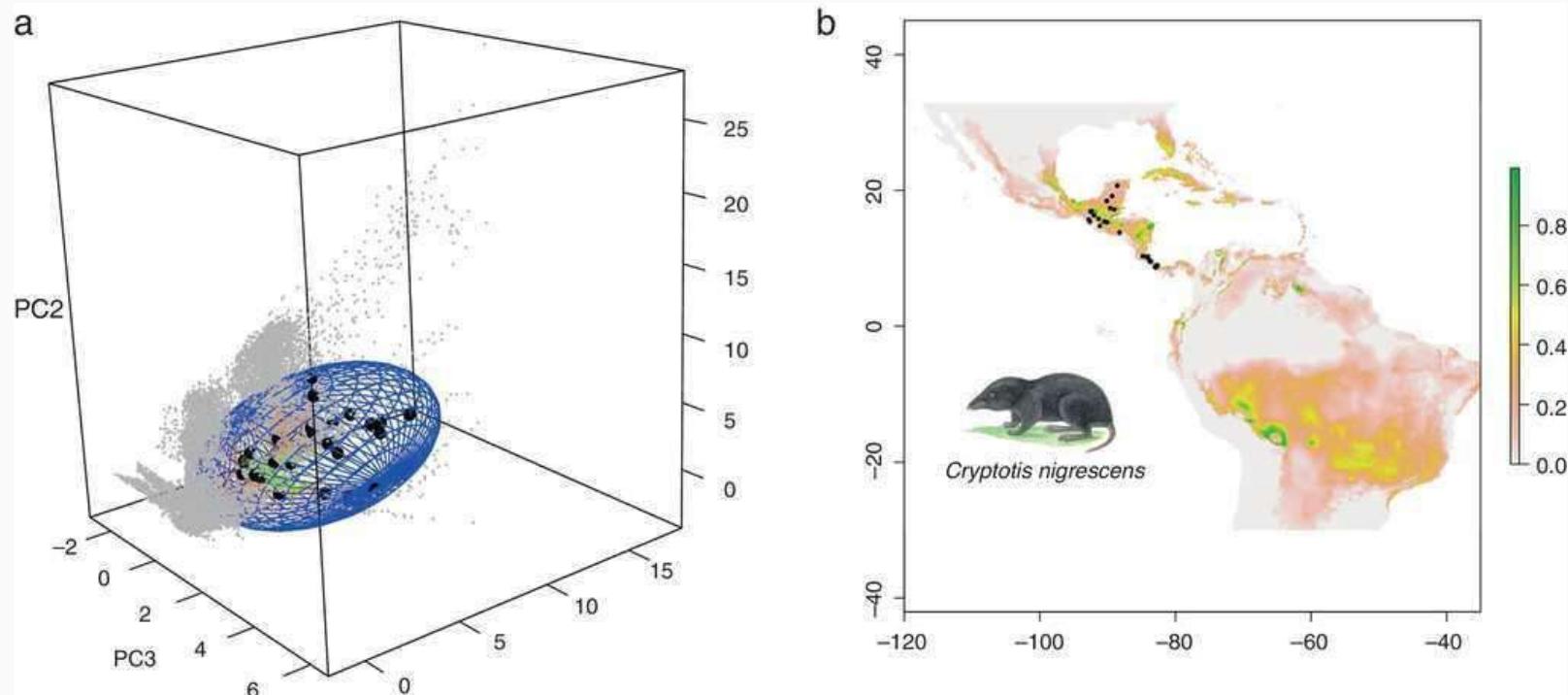


(...) essentially, all models are wrong, but some are useful. George E. P. Box (1919-2013)

# 2.1 Visão geral dos SDMs

## Objetivo

Predizer a **distribuição de espécies** (ou entidades) a partir do **nicho ecológico**



[Soberón et al. \(2017\)](#)

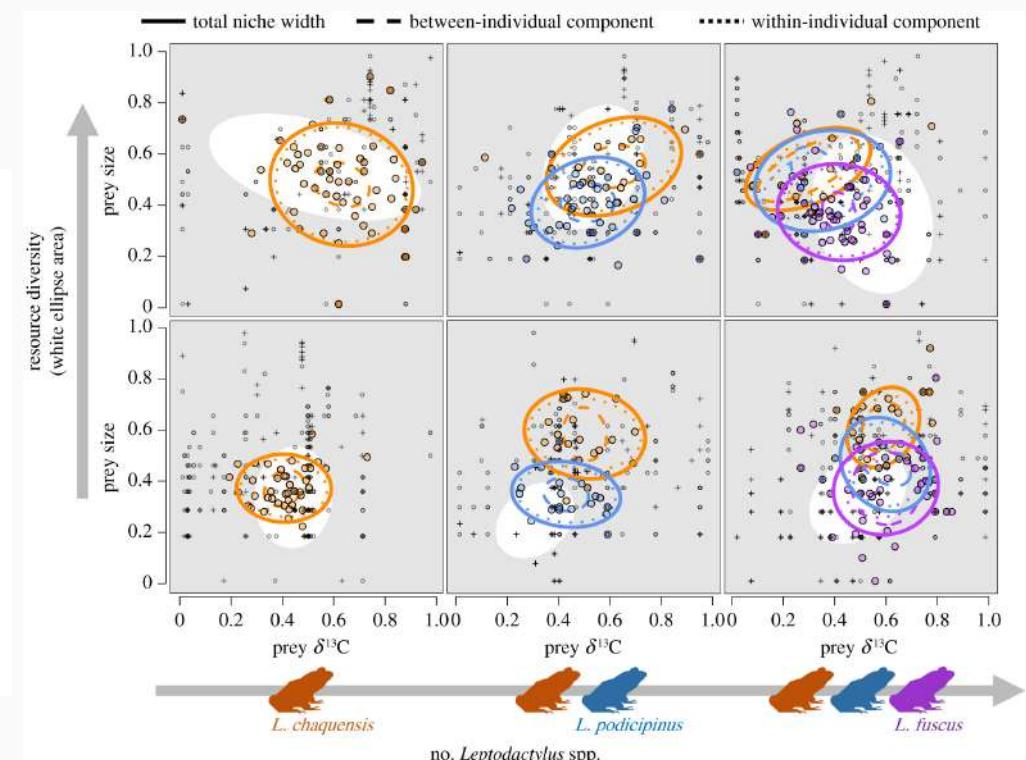
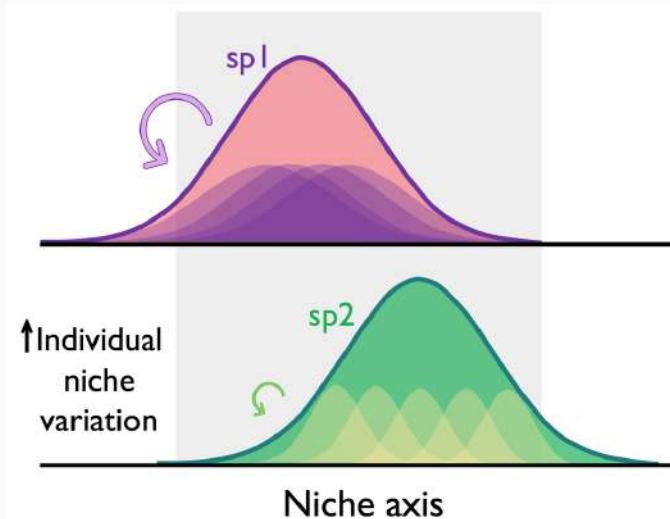
Revista Mexicana de Biodiversidad. 2017;88:437–41

# 2.1 Visão geral dos SDMs

Entretanto...

Estimar/analisar o **nicho das espécies** sem inferir sua distribuição geográfica

- Sobreposição de nicho
- Amplitude do nicho
- Montagem de comunidades
- Respostas ambientais
- Diversidade funcional
- ...



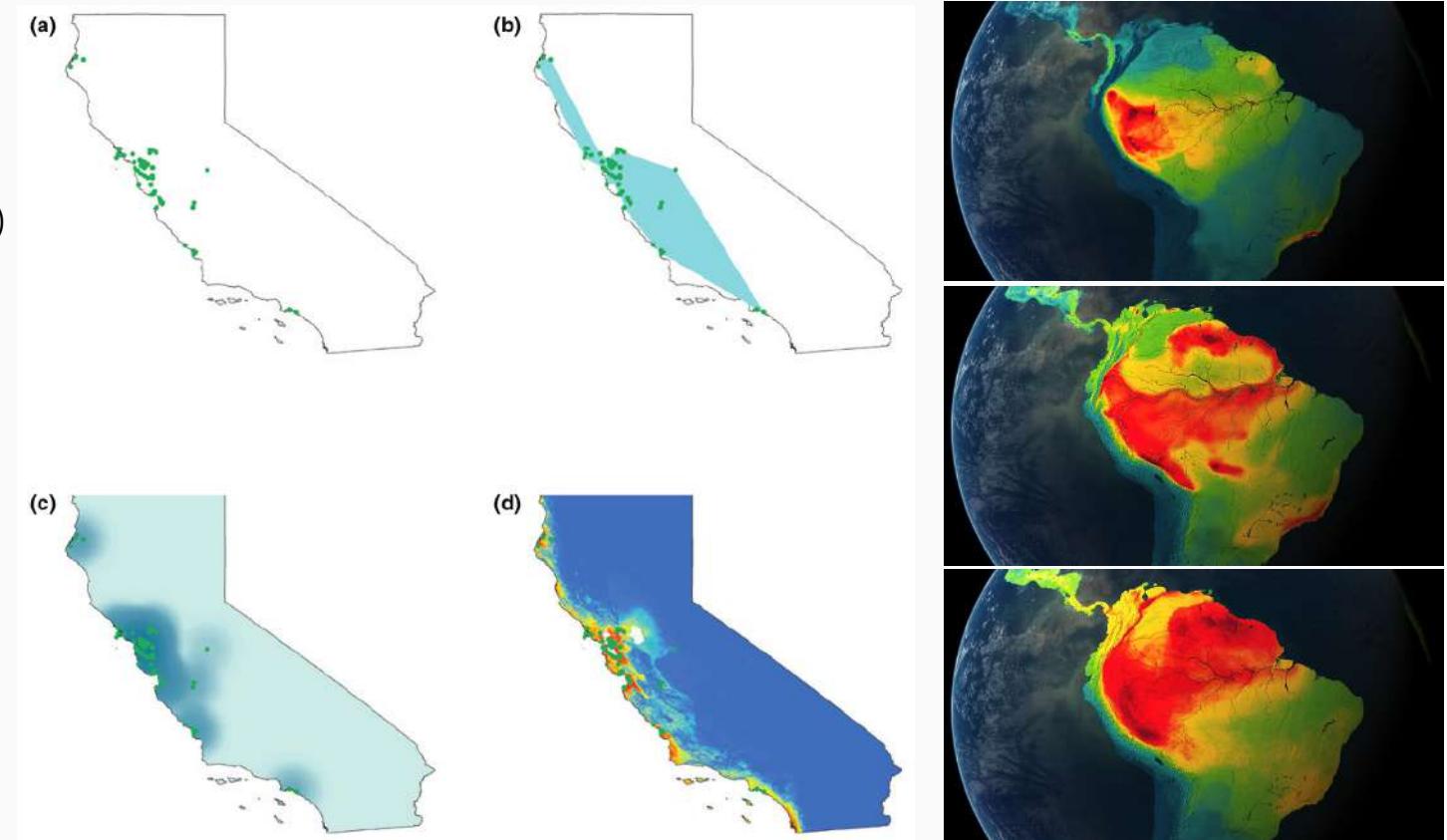
[Site pessoal do Raul Costa-Pereira](#), [Costa-Pereira et al. \(2019\)](#)

# 2.1 Visão geral dos SDMs

Entretanto...

Inferir a **distribuição das espécies** sem estimar o nicho ecológico

- Buffers
- Densidade de Kernel
- Mínimo Polígono Convexo (MPC)
- Interpolações (IDW, Spline, Krigagem)
- Opinião de especialistas (IUCN)
- ...



[Bahn & McGill \(2007\)](#),

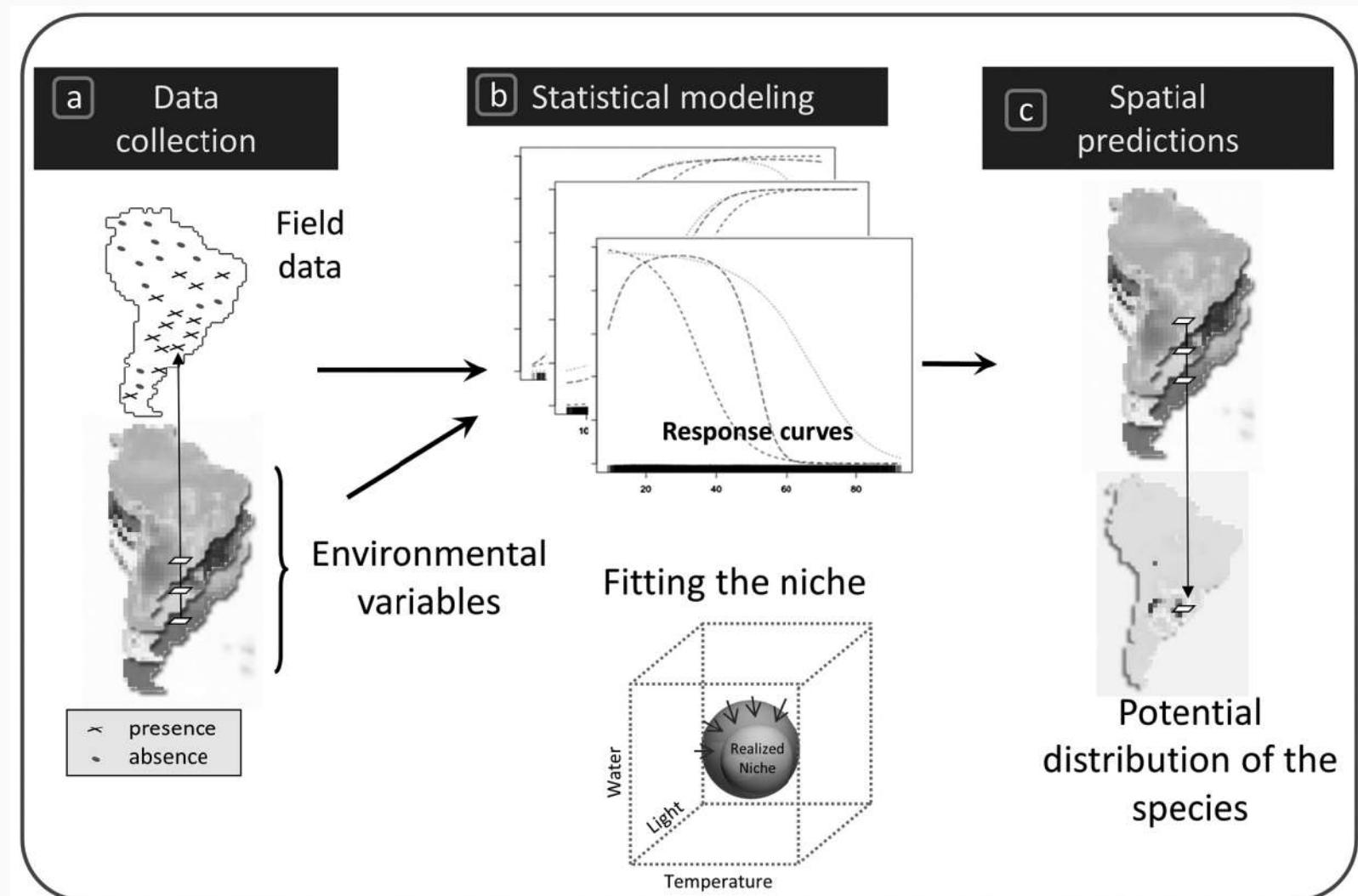
[Jenkins et al. \(2013\)](#),

[BiodiversityMapping.org](#),

[IUCN \(2021\)](#),

[Warren \(2012\)](#).

## 2.1 Visão geral dos SDMs



[Guisan et al. \(2017\)](#)

# O que é nicho ecológico?

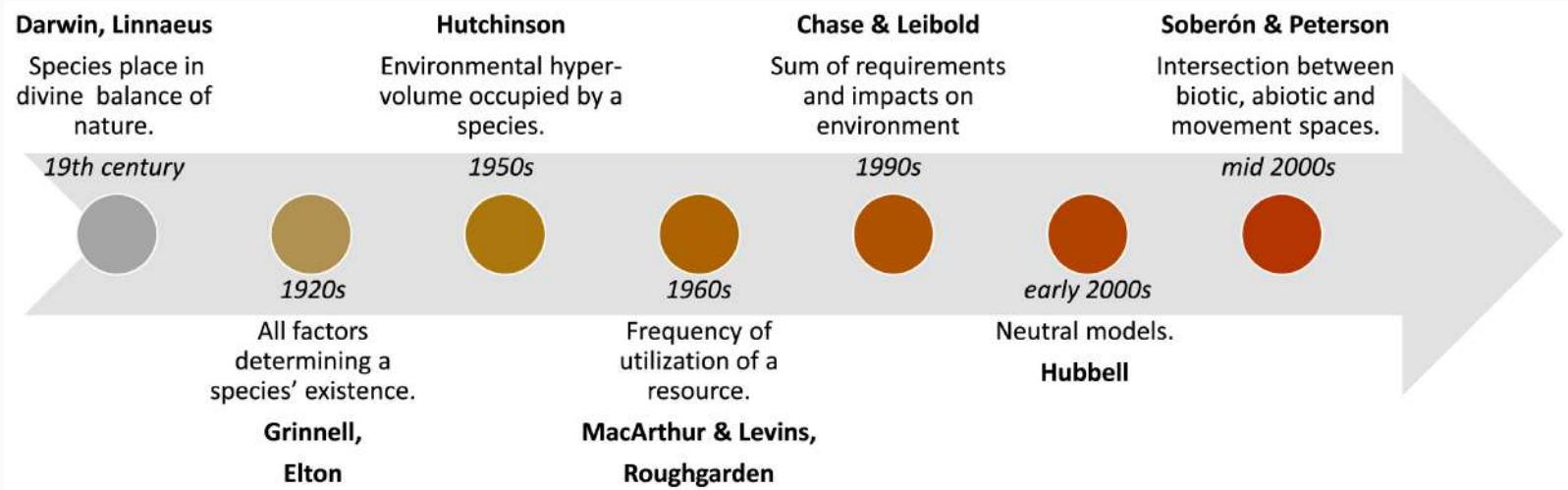
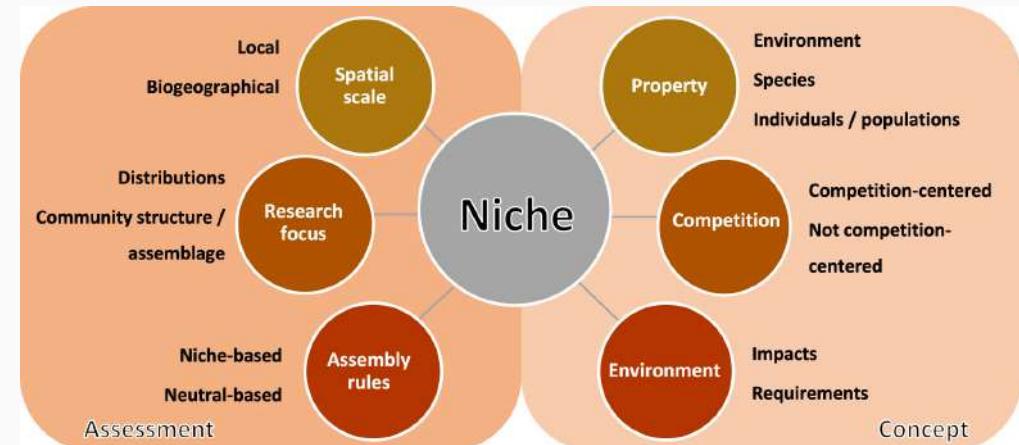
# 2.2 Nicho ecológico e distribuição de espécies

## Nicho ecológico

### 1. Modelagem de nicho ecológico

### 2. Coexistência entre espécies

### 3. Metacomunidades

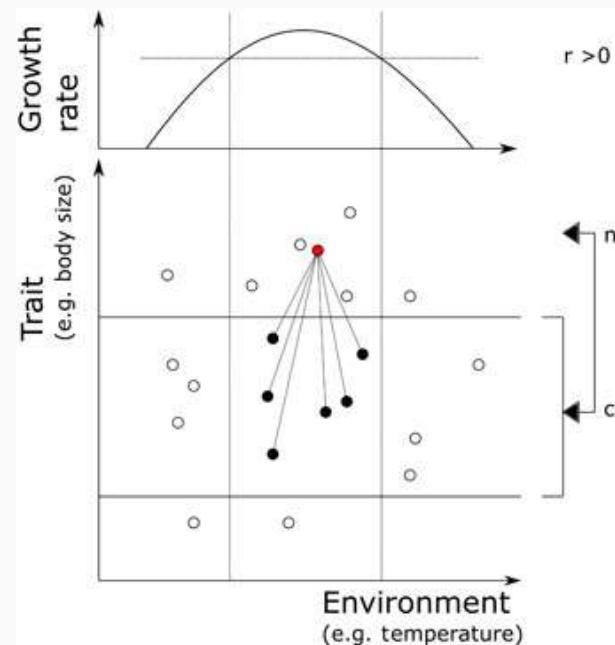


# 2.2 Nicho ecológico e distribuição de espécies

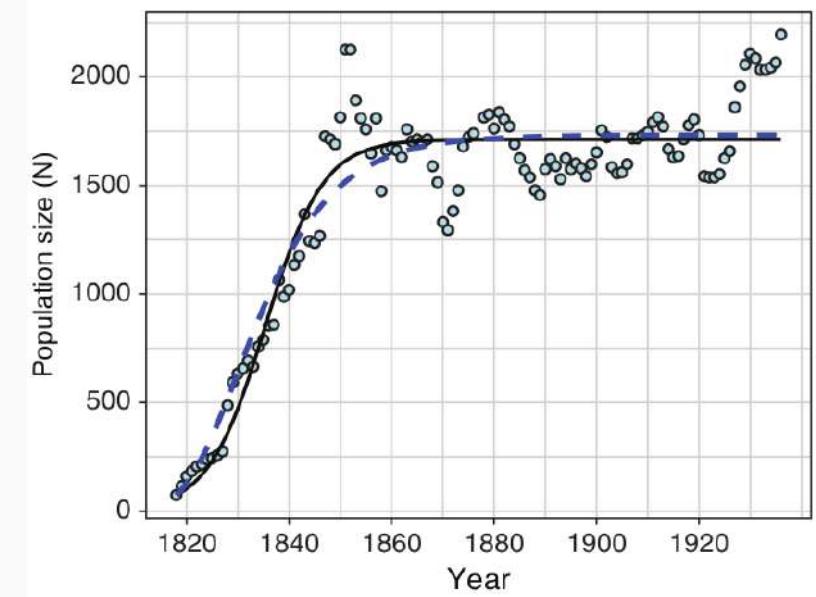
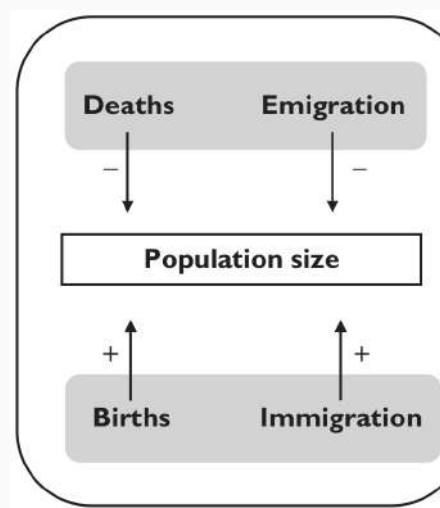
## Nicho ecológico

### 1. Modelagem de nicho ecológico

- **Combinações de requisitos** que permitem a uma população **sobreviver e crescer ( $r > 0$ )** em um **determinado lugar**



$$\frac{dN}{dt} = rN \left[ 1 - \frac{N}{K} \right]$$



[Chase & Leibold \(2003\)](#), [Holt \(2009\)](#), [Soberón \(2010\)](#), [Ovaskainen et al. \(2016\)](#), [Gravel et al. \(2019\)](#), [Murray & Sandercock \(2019\)](#)

# 2.2 Nicho ecológico e distribuição de espécies

## Nicho ecológico

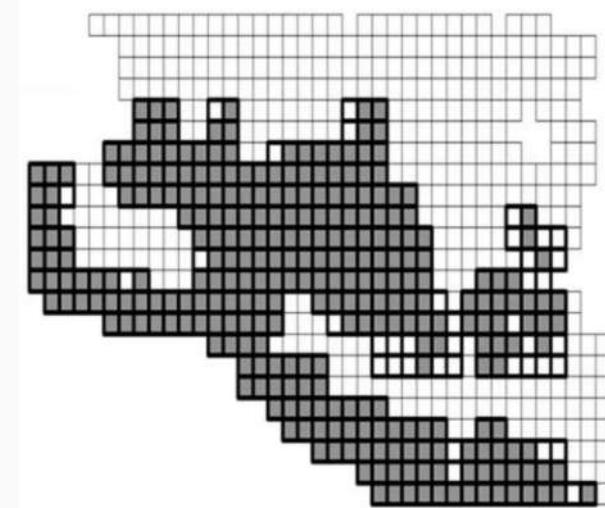
### 1. Modelagem de nicho ecológico

- **Requisitos:** componente abiótico, biótico e movimento

$$\frac{1}{x_{i,g}} \frac{dx_{i,g}}{dt} = r_{i,g}(e_g) - \varphi_{i,g}(e_g, R_{i,g}; x_g) + \Psi_{i,g}(x_i; T_i)$$

- $\frac{1}{x_{i,g}} \frac{dx_{i,g}}{dt}$ : **taxa de crescimento per capita de uma população**, onde  $x_{i,g}$  é a densidade da espécie  $i$  na célula  $g$
- $r_{i,g}(e_g)$ : termo de **crescimento intríseco** da população, onde  $e_g$  são as condições abióticas da espécie  $i$  na célula  $g$
- $\varphi_{i,g}(e_g, R_{i,g}; x_g)$ : termo de **regulação**, onde  $R_{i,g}$  são as interações com outras espécies (incluindo recursos), e  $x_g$  densidades populacionais das espécies ocorrendo na célula  $g$
- $\Psi_{i,j}(x_i; T_i)$ : termo de **movimento**, onde  $x_i$  representa a densidade populacional da espécie  $i$  em cada célula de  $\mathbf{G}$ , e  $T_i$  expressa os movimentos intercelulares da espécie  $i$

[Soberón \(2010\)](#), [Peterson et al. \(2011\)](#).

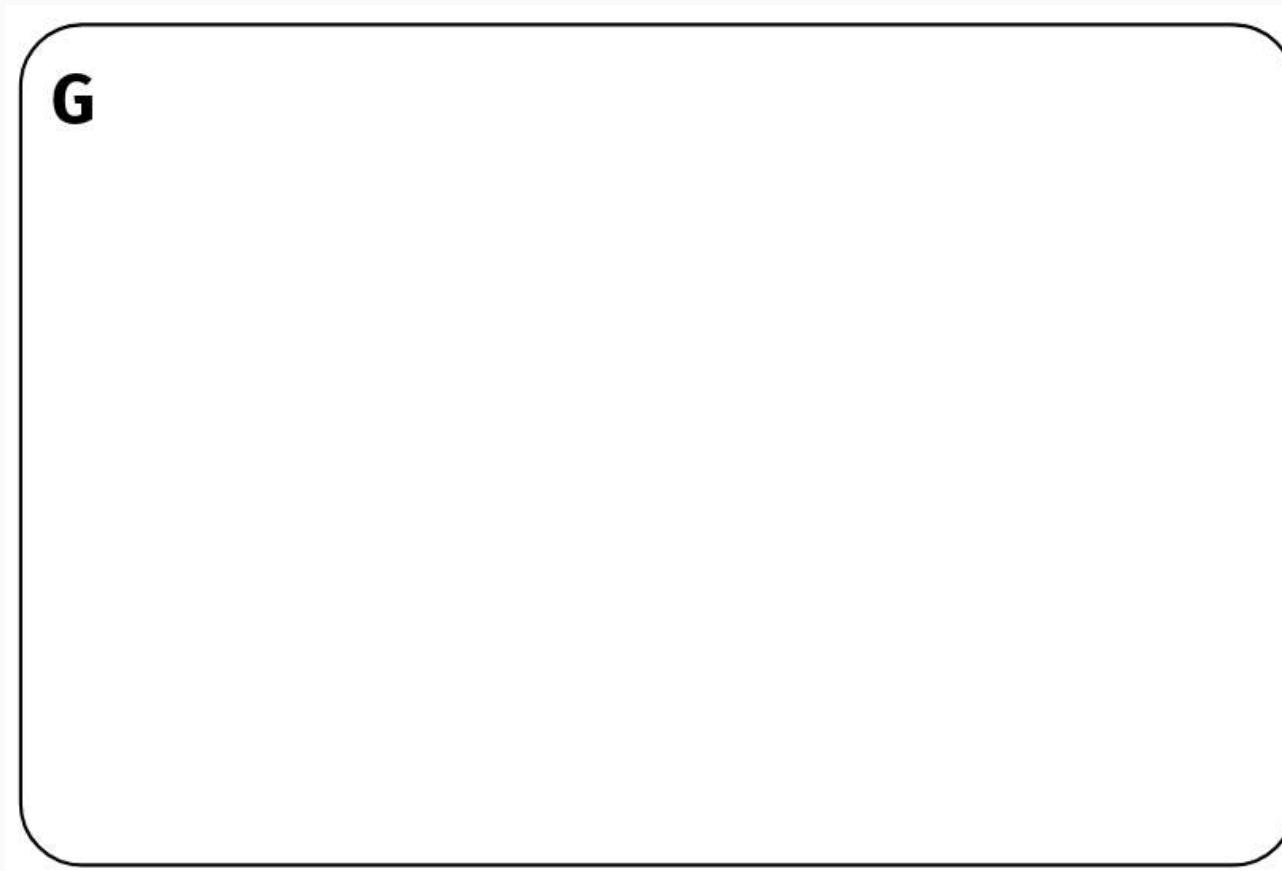


Meio complicado, não?

Vamos olhar para os mesmos conceitos de  
**requerimentos** (abiótico, biótico e movimento), mas  
agora desenhando e com sapos...

## 2.2 Nicho ecológico e distribuição de espécies

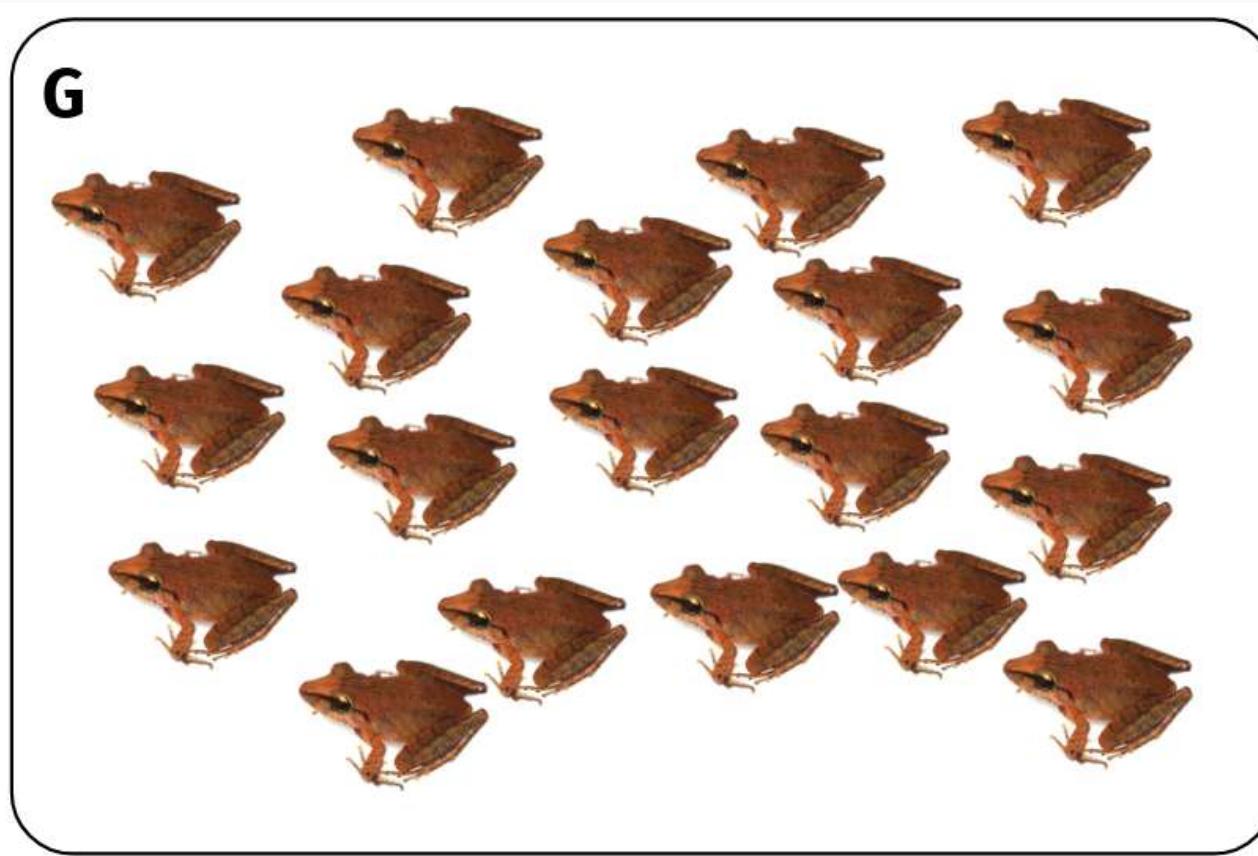
### Espaço geográfico (G)



[Soberón & Peterson \(2005\)](#), [Peterson et al. \(2011\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

### Distribuição de espécies

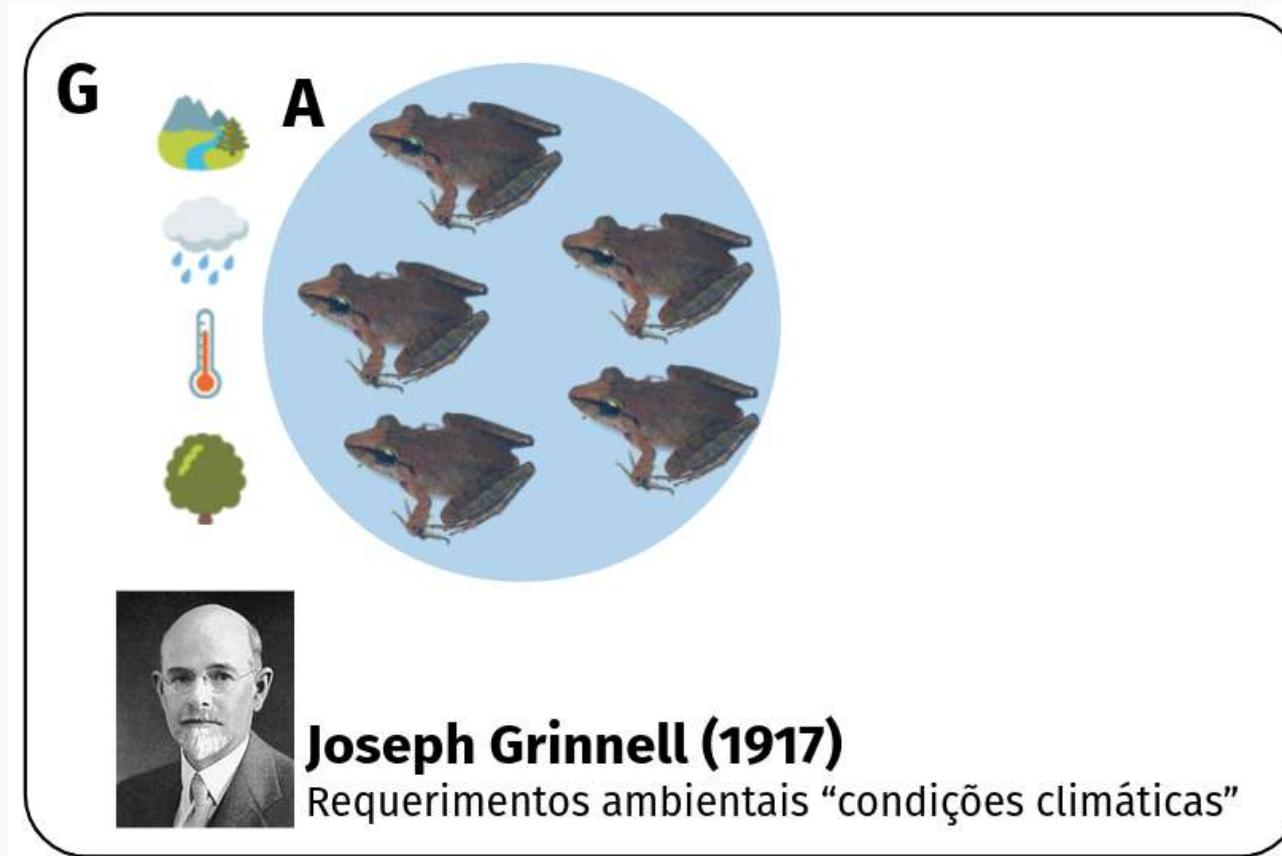


[Soberón & Peterson \(2005\)](#), [Peterson et al. \(2011\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

### Condições abióticas (A)

$r_{i,g}(e \rightarrow g)$

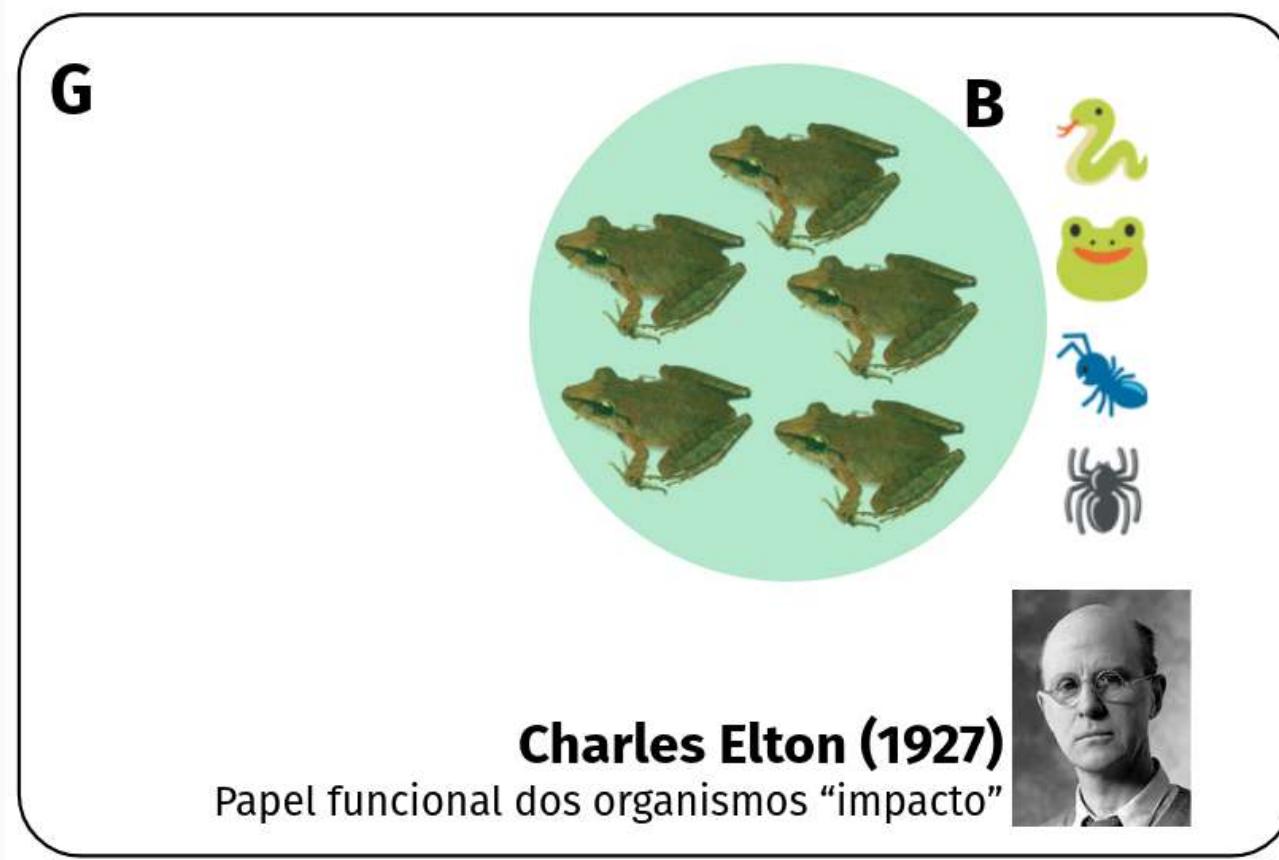
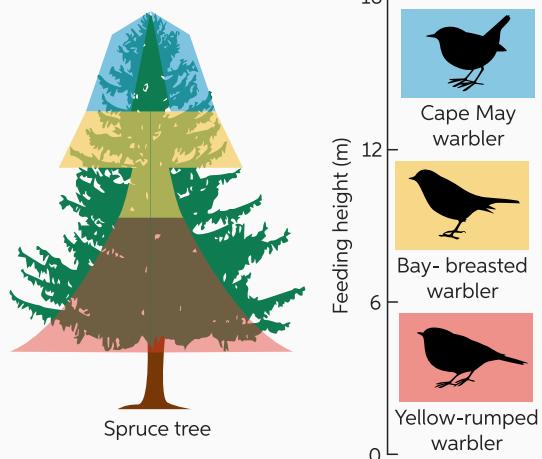


[Grinnell \(1917\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#)

## 2.2 Nicho ecológico e distribuição de espécies

### Condições bióticas (B)

$$\varphi_{i,g}(e_g, R_{i,g}; x_g)$$



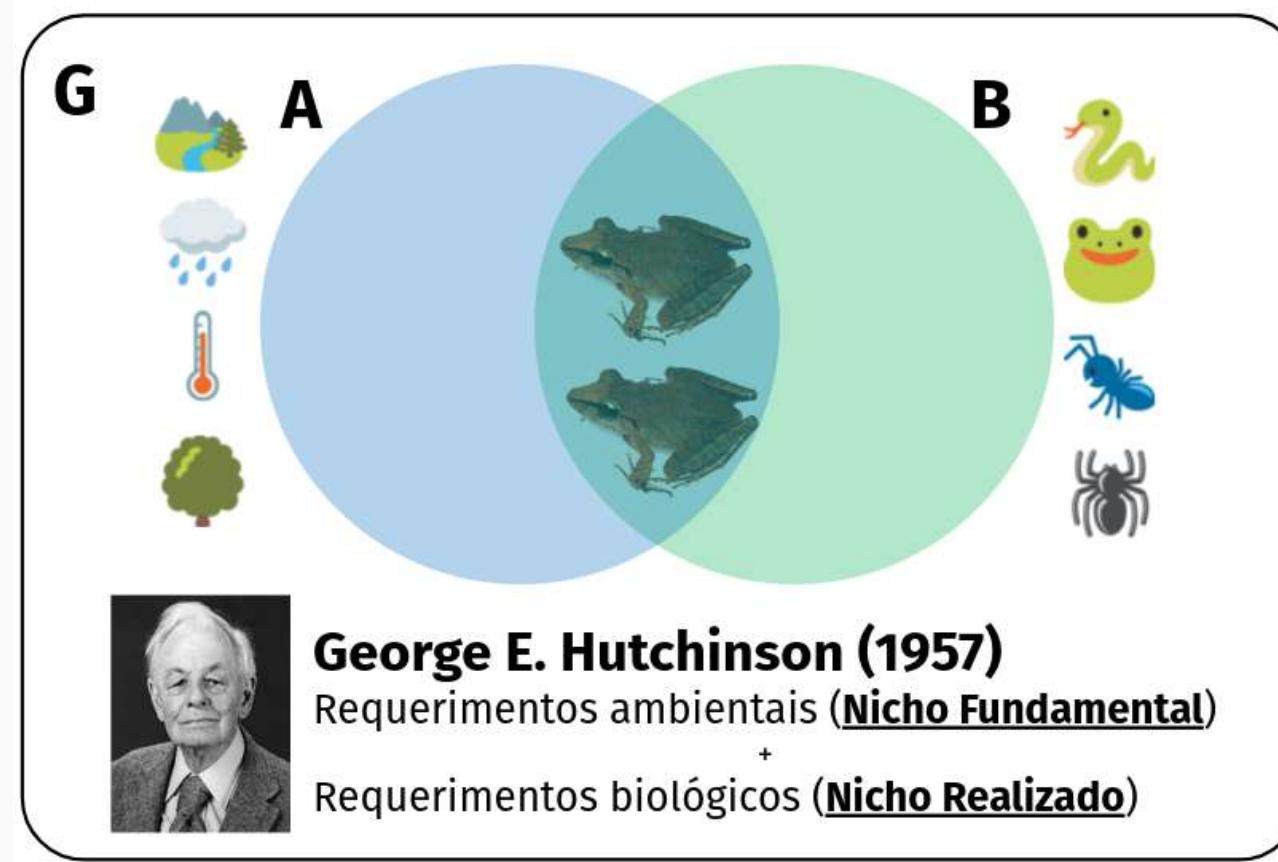
[Elton \(1927\)](#), [MacArthur & Levins \(1967\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

Condições abióticas (A) e bióticas (B)

$$+r_{i,g}(e \rightarrow g)$$

$$-\varphi_{i,g}(e \rightarrow g, R^{\rightarrow}{}_{i,g}; x^{\rightarrow}{}_{g})$$

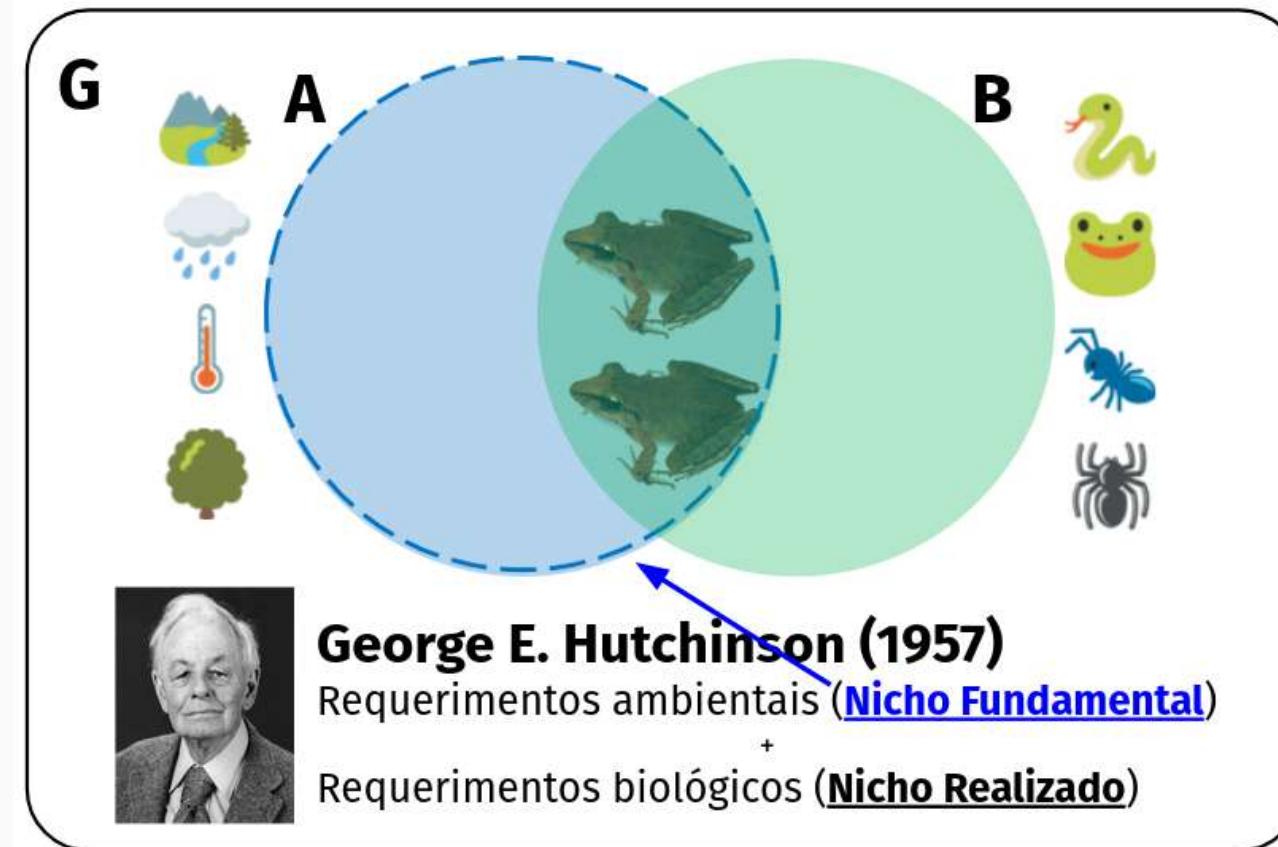


[Hutchinson \(1957\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

**Nicho fundamental:** todas as condições ambientais favoráveis

$+r_{i,g}(e \rightarrow g)$



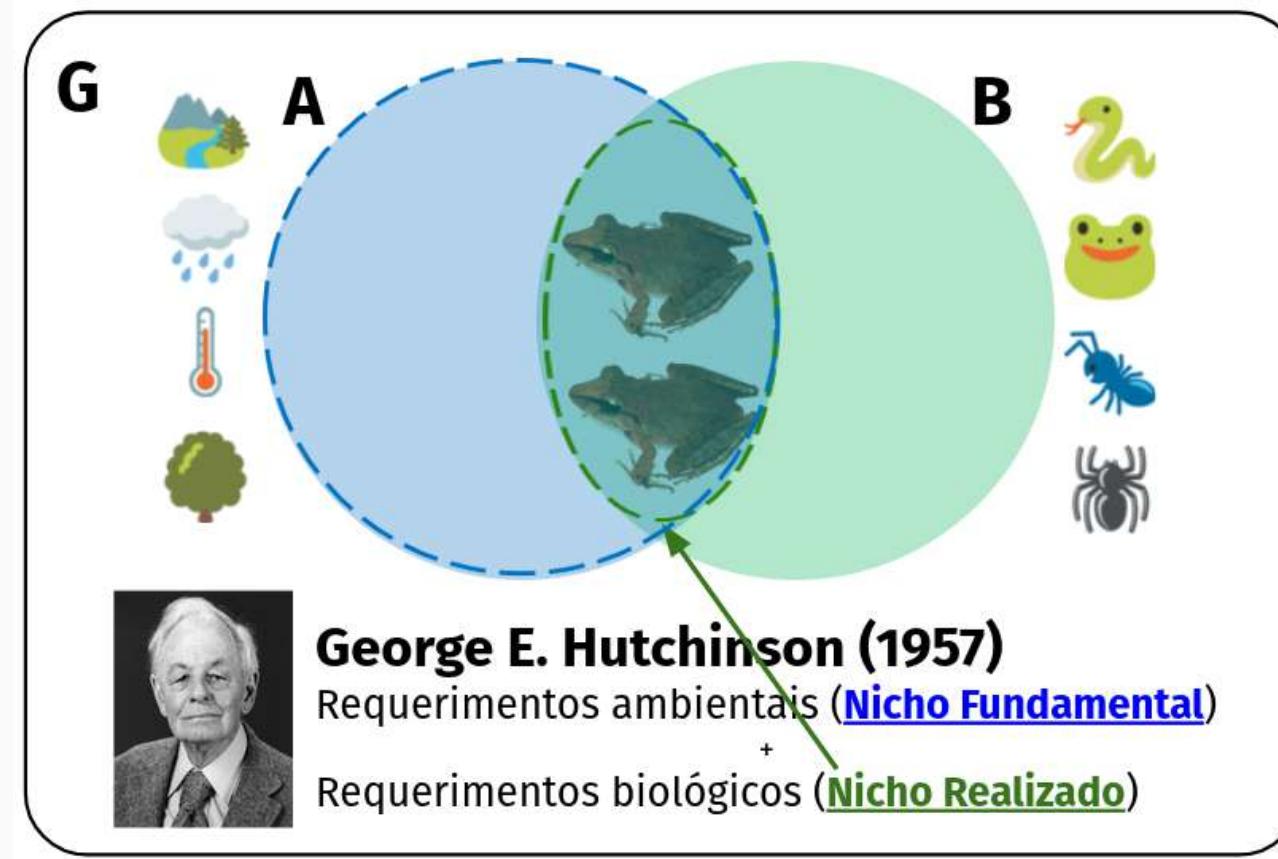
[Hutchinson \(1957\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

**Nicho realizado:** todas as condições ambientais favoráveis limitadas pela interação

$$+r_{i,g}(e \rightarrow g)$$

$$-\varphi_{i,g}(e \rightarrow g, R'_{i,g}; x'_{i,g})$$



[Hutchinson \(1957\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#).

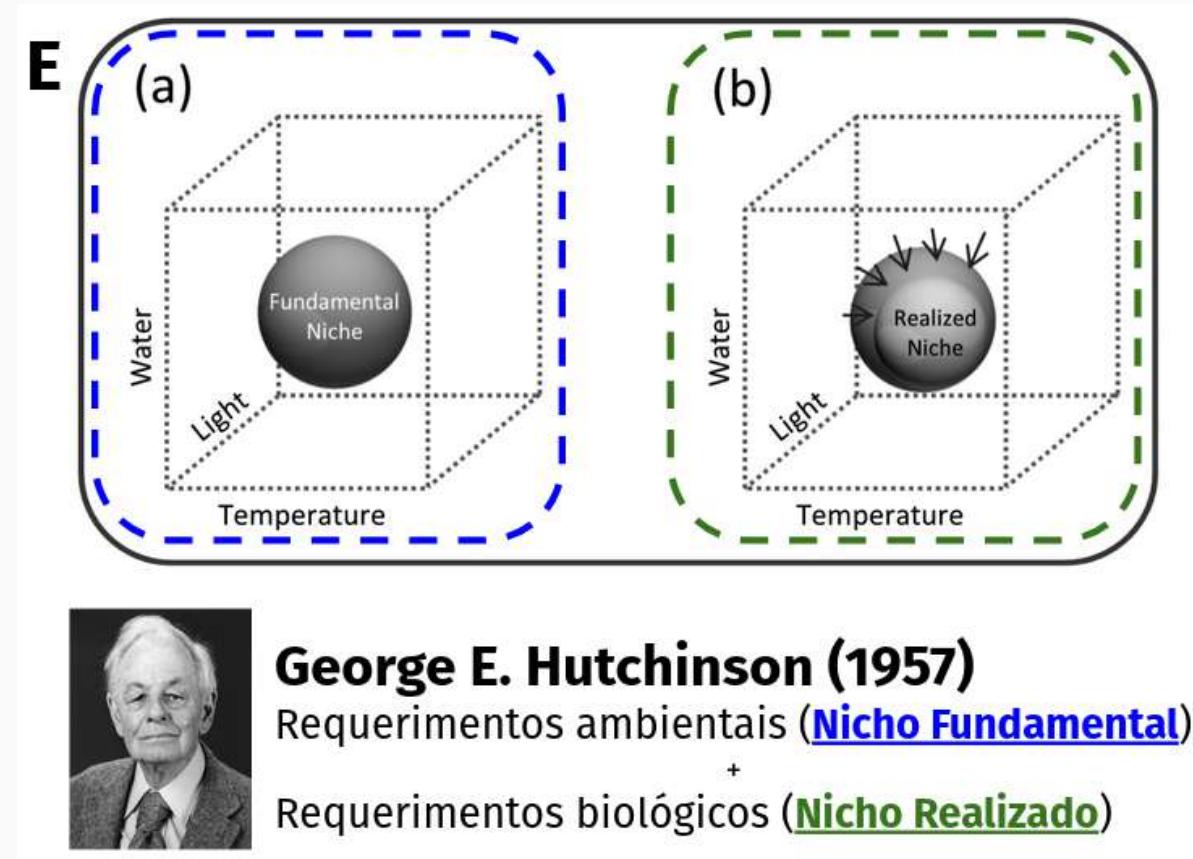
Agora vamos olhar para as **condições ambientais e bióticas** no **Espaço Ambiental (E)**, ou seja, apenas para os **valores** das condições ambientais e as limitações bióticas

## 2.2 Nicho ecológico e distribuição de espécies

### Nicho: Hipervolume n-dimensional

$$+r_{i,g}(e \rightarrow g)$$

$$-\varphi_{i,g}(e \rightarrow g, R^i, g; x^i_g)$$



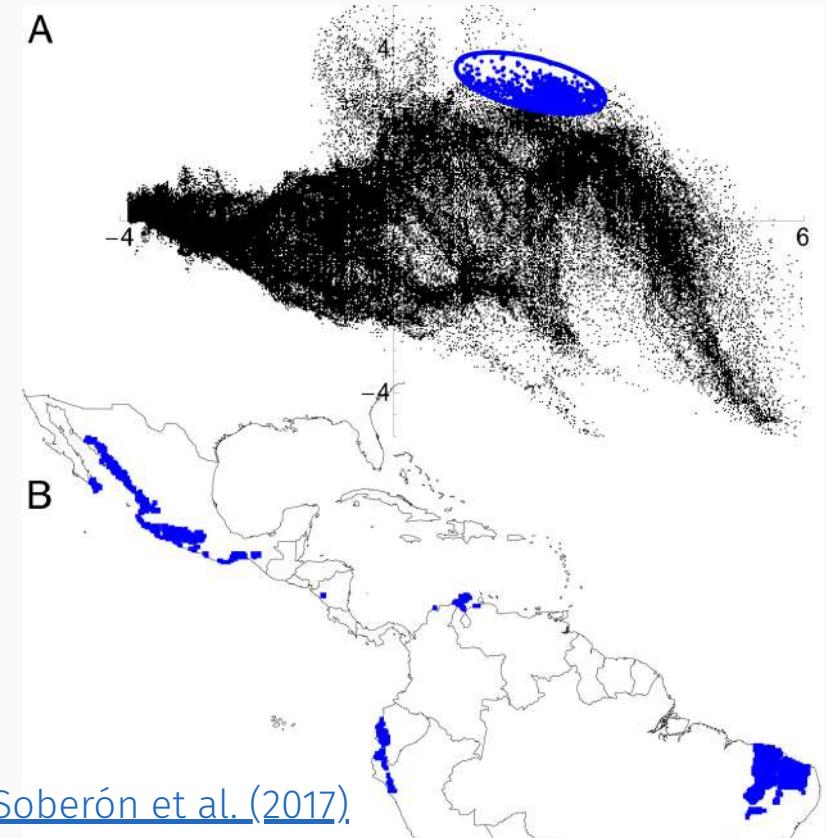
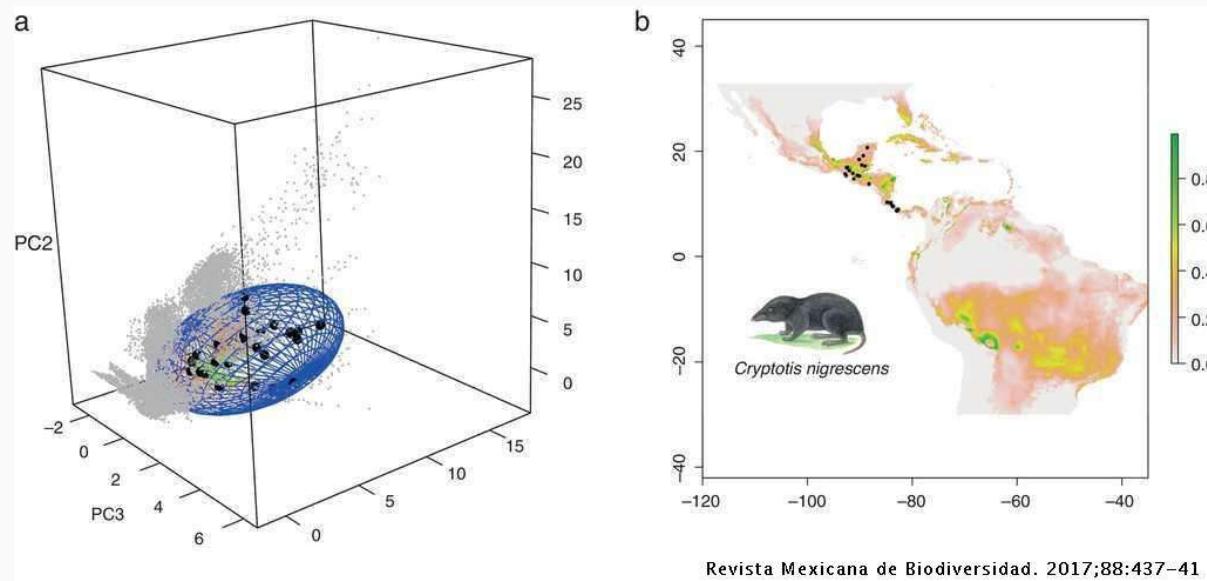
[Hutchinson \(1957\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#), [Guisan et al. \(2017\)](#).

**Oxe 1:** notaram que falei do **Espaço Geográfico (G)** e do  
**Espaço Ambiental (E)?**

## 2.2 Nicho ecológico e distribuição de espécies

Dualidade Hutchinsoniana

Inferir a **distribuição geográfica** de uma espécie a partir do **nicho ecológico**

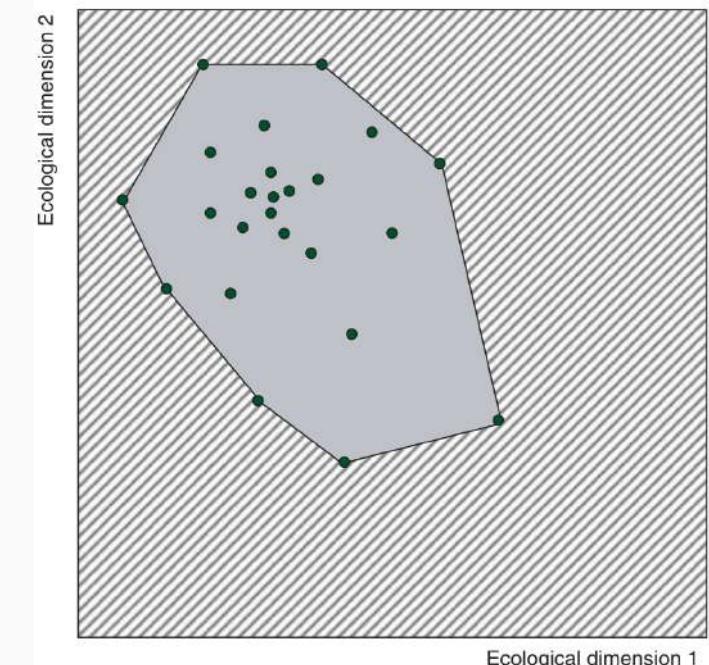
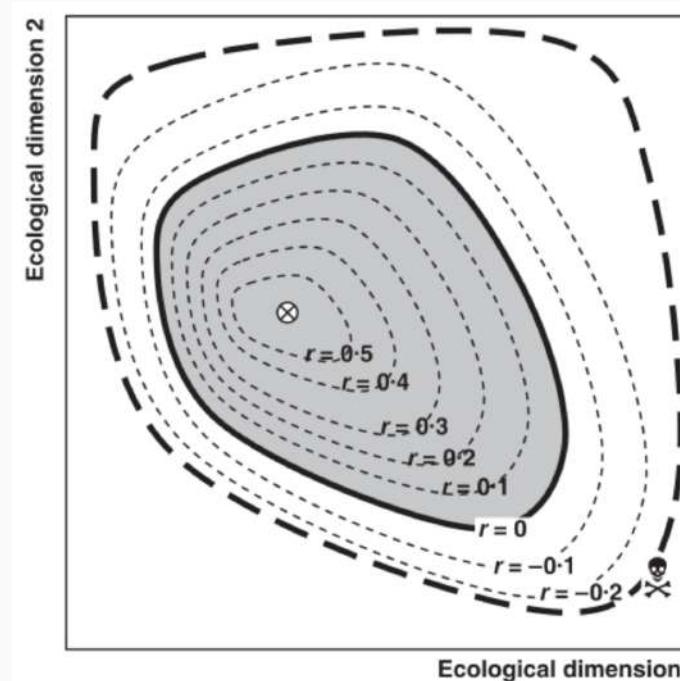
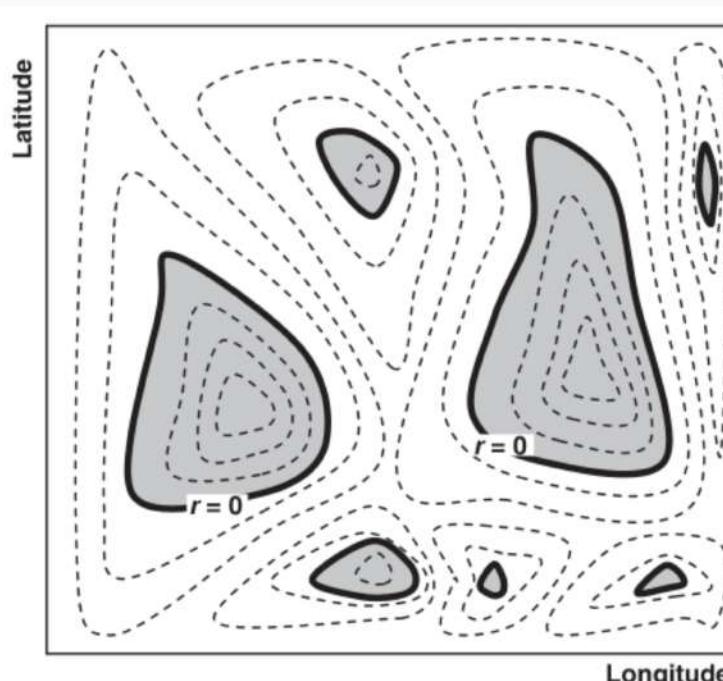


[Colwell & Rangel \(2009\)](#), [Soberón & Nakamura \(2009\)](#), [Peterson et al. \(2011\)](#), [Soberón et al. \(2017\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

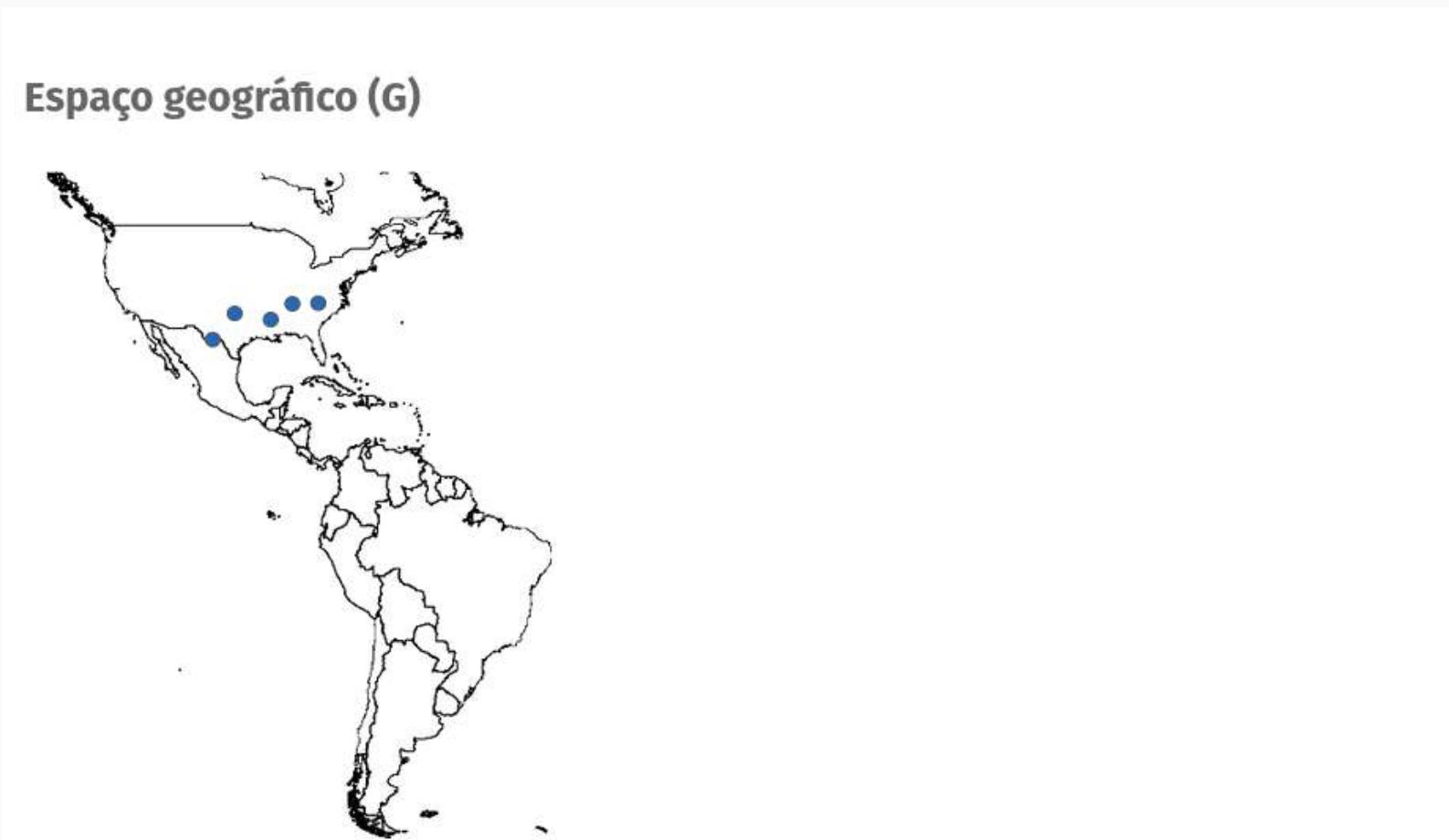
### Dualidade Hutchinsoniana e Estimativa do Nicho Ecológico

- Podemos associar a **taxa de crescimento intrínseca ( $r$ )** com as condições ambientais
- Essas condições favoráveis no Espaço Ambiental (E) possuem **diferentes localizações** no Espaço Geográfico (G)
- O modelo vai **estimar o nicho ecológico** no Espaço Ambiental (E) e **predizer** a distribuição no Espaço Geográfico (G)



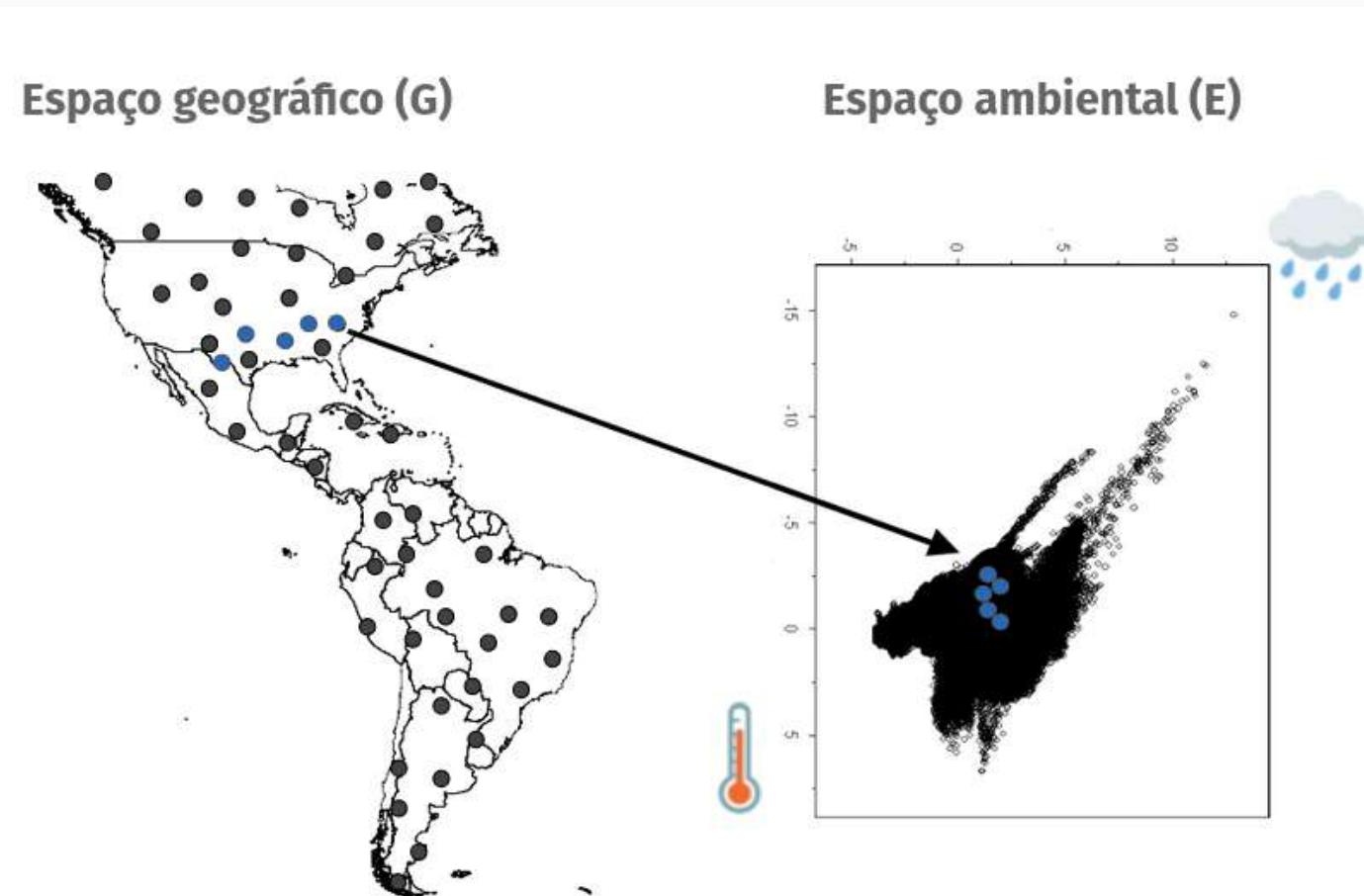
## 2.2 Nicho ecológico e distribuição de espécies

### Ocorrências no espaço geográfico



## 2.2 Nicho ecológico e distribuição de espécies

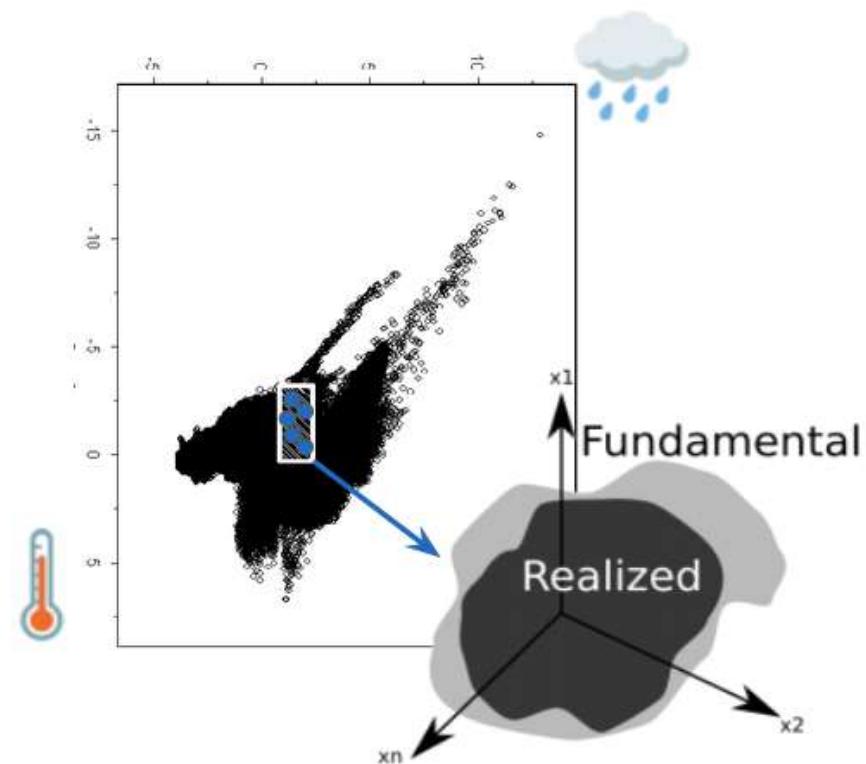
Condições ambientais para as ocorrências e para a área modelada



## 2.2 Nicho ecológico e distribuição de espécies

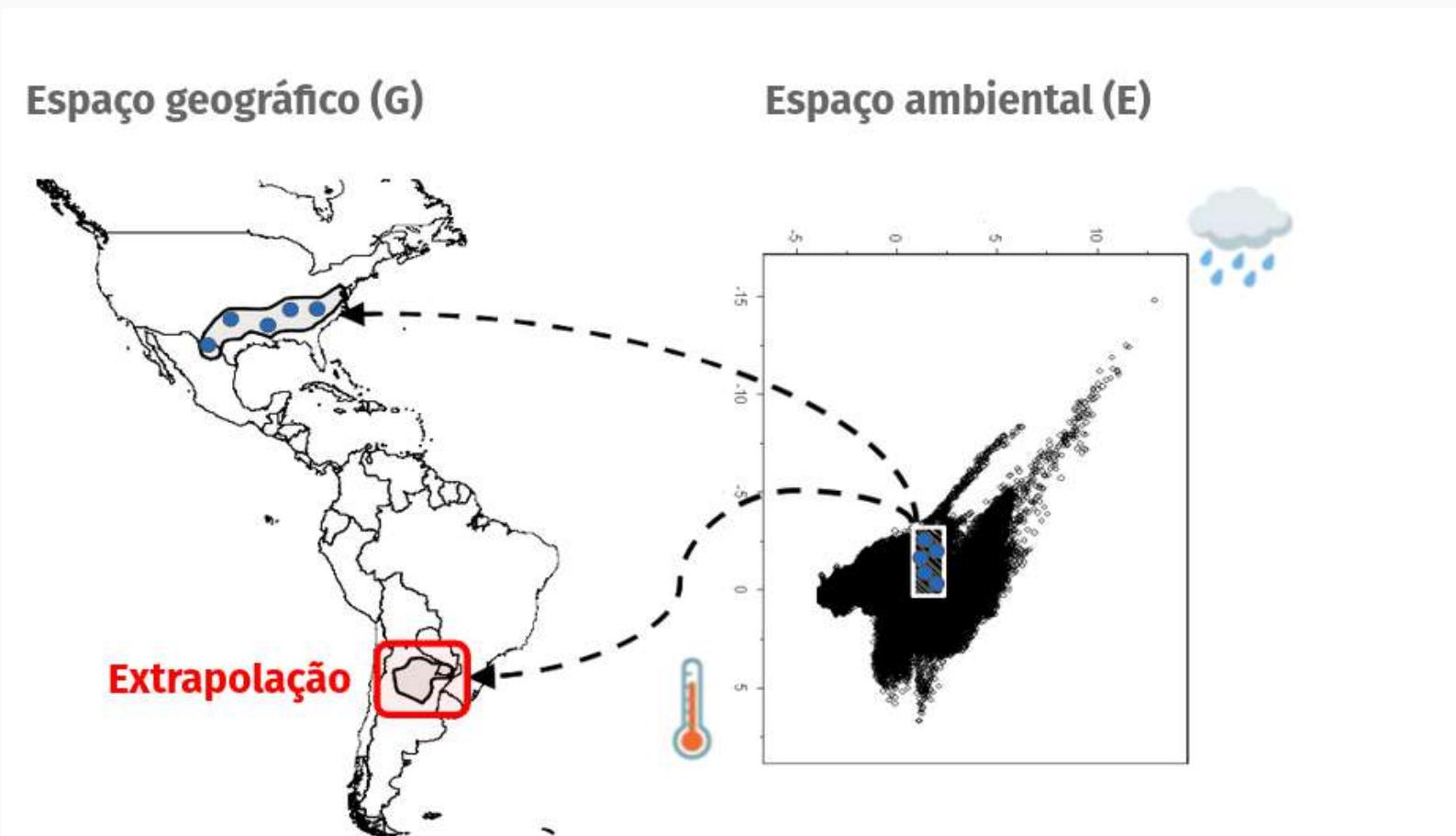
### Estimativa do nicho realizado

Espaço ambiental (E)



## 2.2 Nicho ecológico e distribuição de espécies

Predição geográfica do nicho realizado estimado



Oxe2! Pode ser que uma mesma condição ambiental no Espaço Ambiental (E) represente mais de um local no Espaço Geográfico (G)!

Como contornar essa extração?

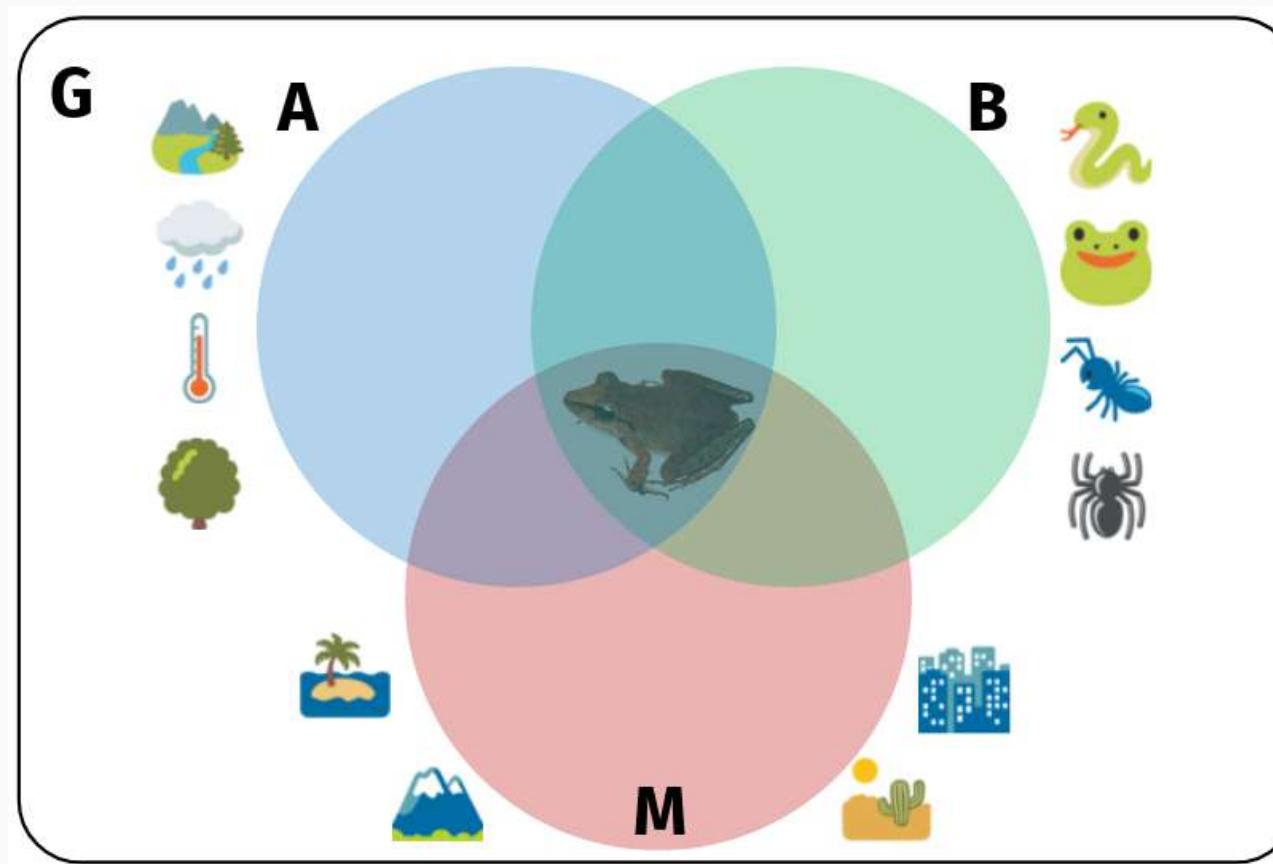
## 2.2 Nicho ecológico e distribuição de espécies

Limitações pela área acessível (dispersão ou movimento) (M)

$$+r_{i,g}(e \rightarrow g)$$

$$-\varphi_{i,g}(e \rightarrow g, R_{i,g}; x \rightarrow g)$$

$$+\Psi_{i,j}(x_i; T_i)$$



[Hutchinson \(1957\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#), [Barve et al. \(2011\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

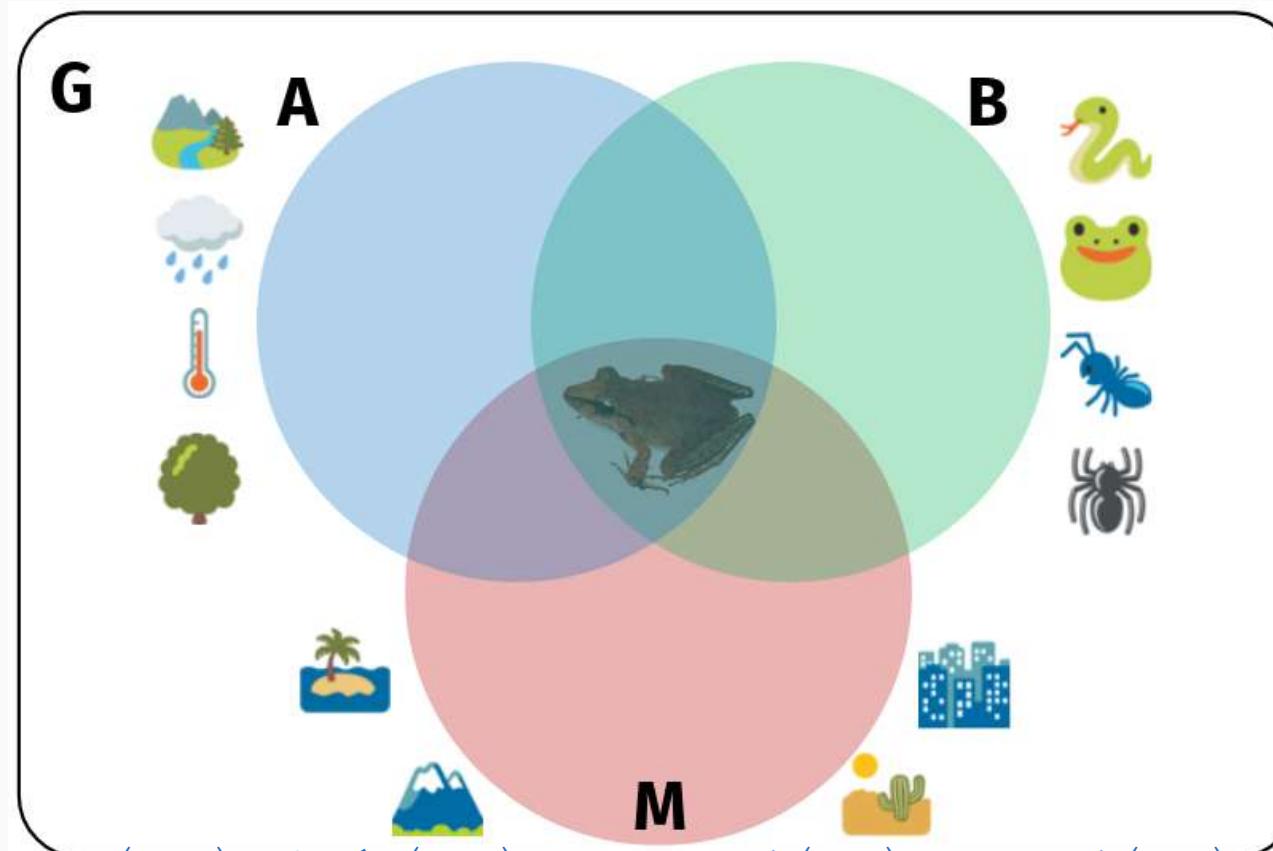
Limitações pela área acessível (dispersão ou movimento) (M)

$$\frac{1}{x_{i,g}} \frac{dx_{i,g}}{dt} =$$

$$+r_{i,g}(e \rightarrow g)$$

$$-\varphi_{i,g}(e \rightarrow g, R \rightarrow i,g; x \rightarrow g)$$

$$+\Psi_{i,j}(x \rightarrow i; T_i)$$



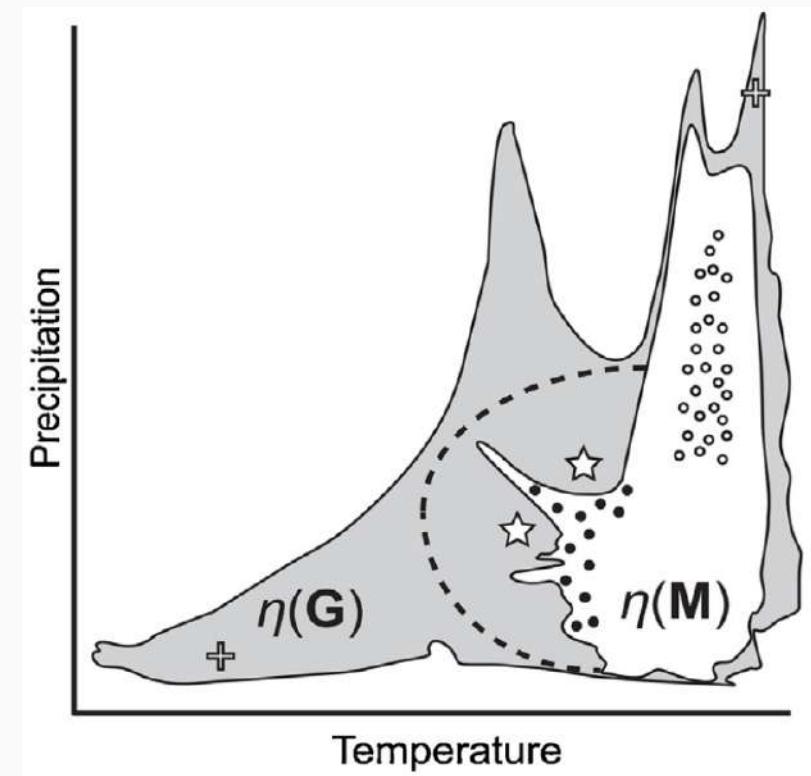
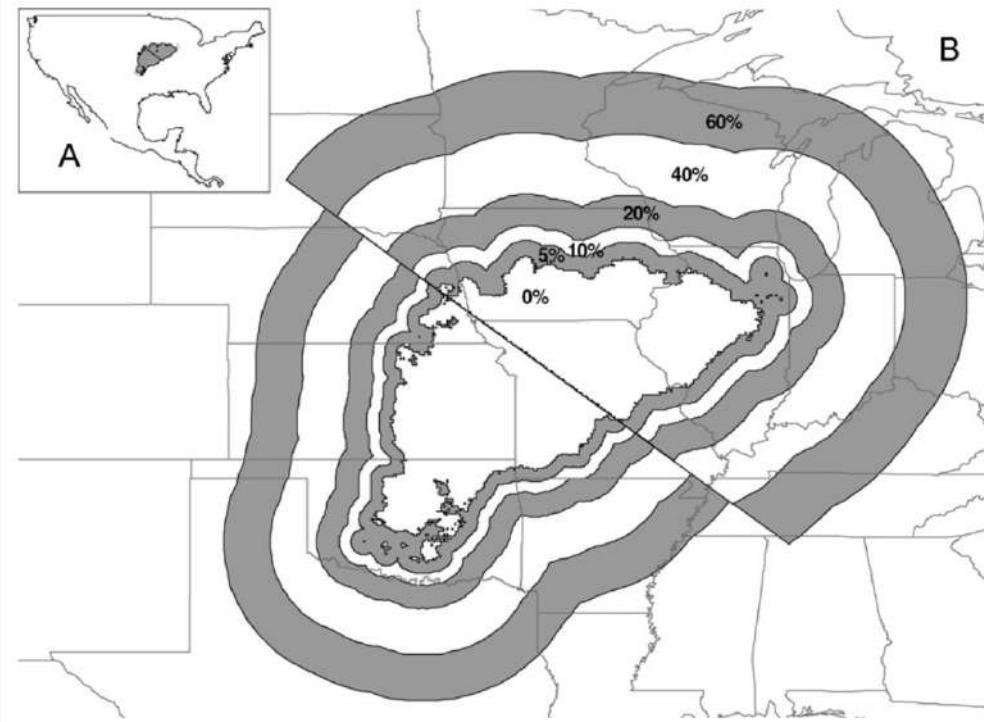
[Hutchinson \(1957\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#), [Barve et al. \(2011\)](#).

E que limite usar para criar os modelos?

## 2.2 Nicho ecológico e distribuição de espécies

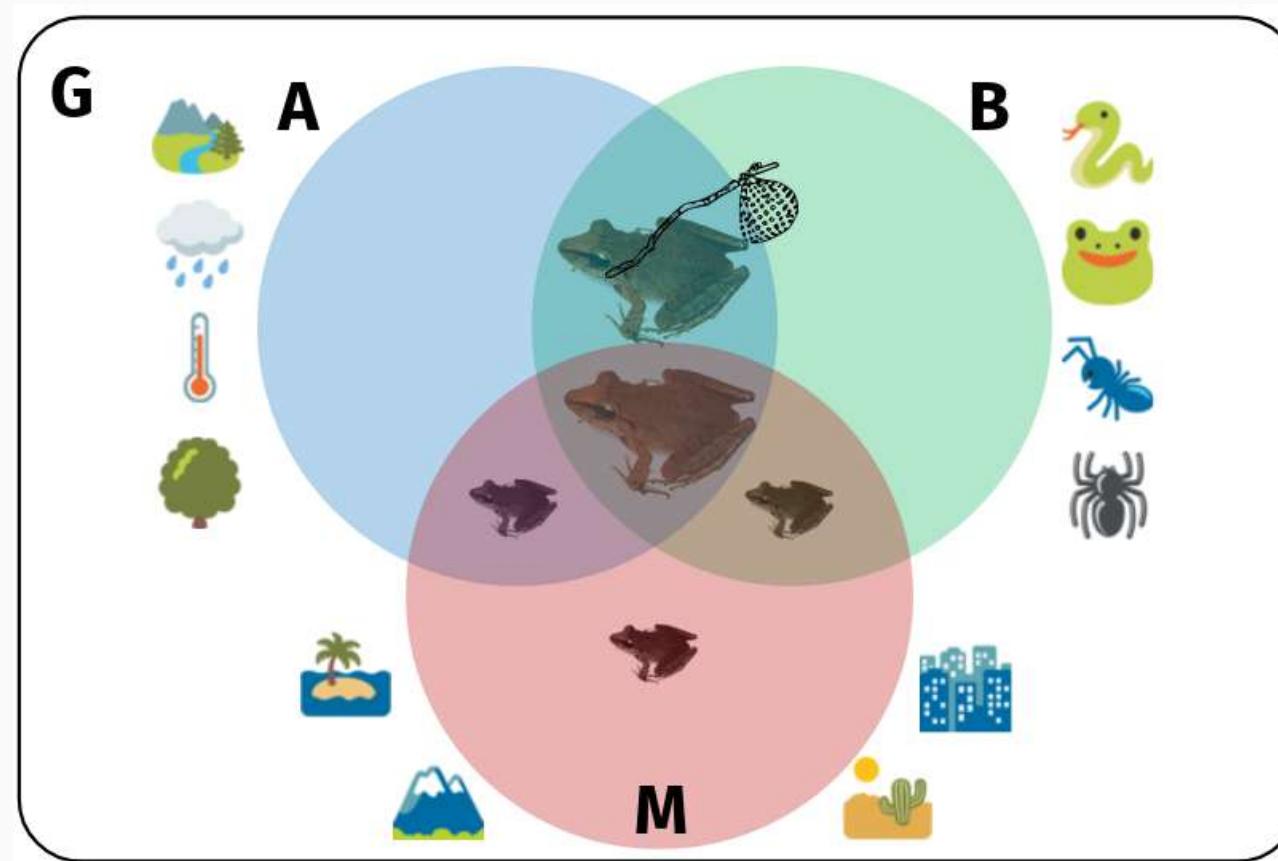
### Limite da área acessível (dispersão ou movimento) (M)

- O limite deve considerar as **condições de dispersão históricas** das espécies



## 2.2 Nicho ecológico e distribuição de espécies

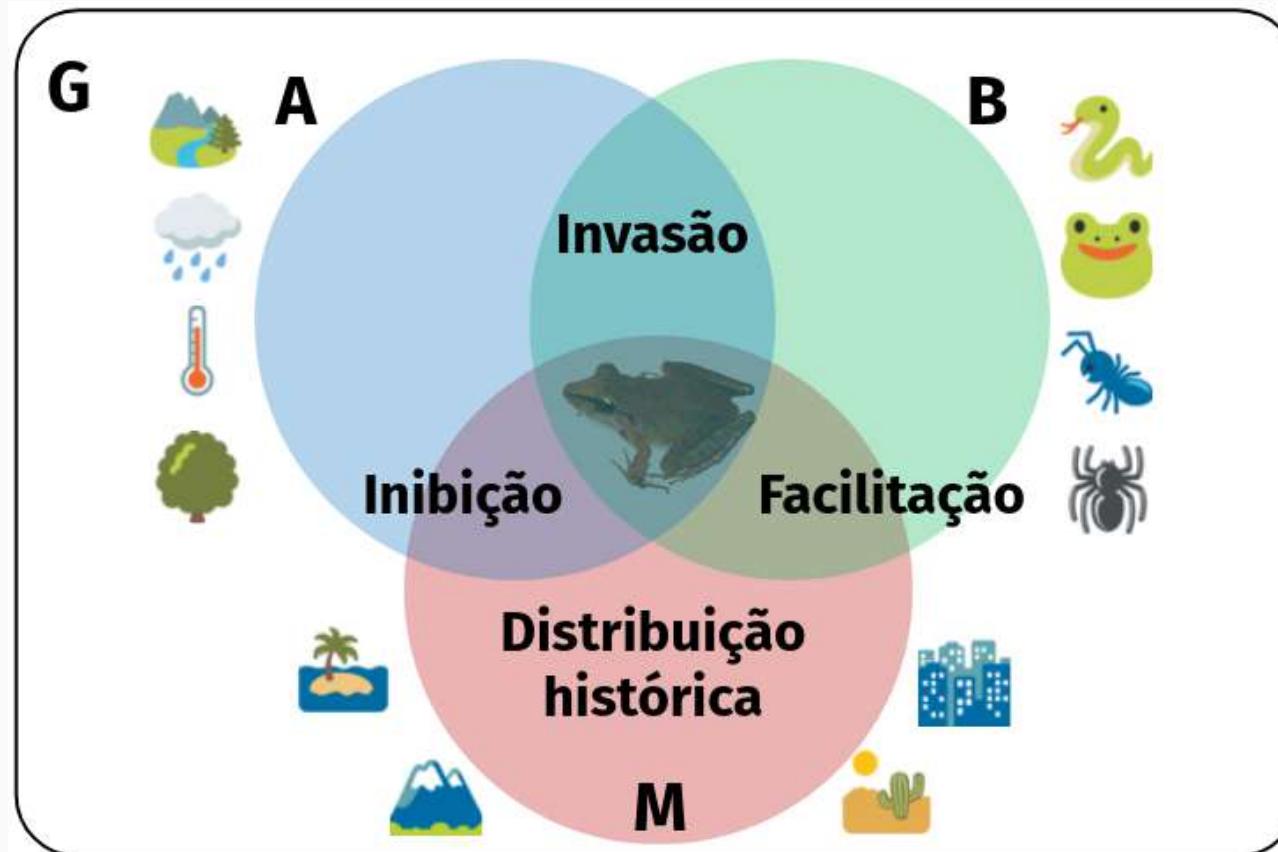
Populações fonte e ralo (source-sink)



[Hutchinson \(1957\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#), [Barve et al. \(2011\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

Populações fonte e ralo (source-sink)



[Hutchinson \(1957\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#), [Barve et al. \(2011\)](#)

## 2.2 Nicho ecológico e distribuição de espécies

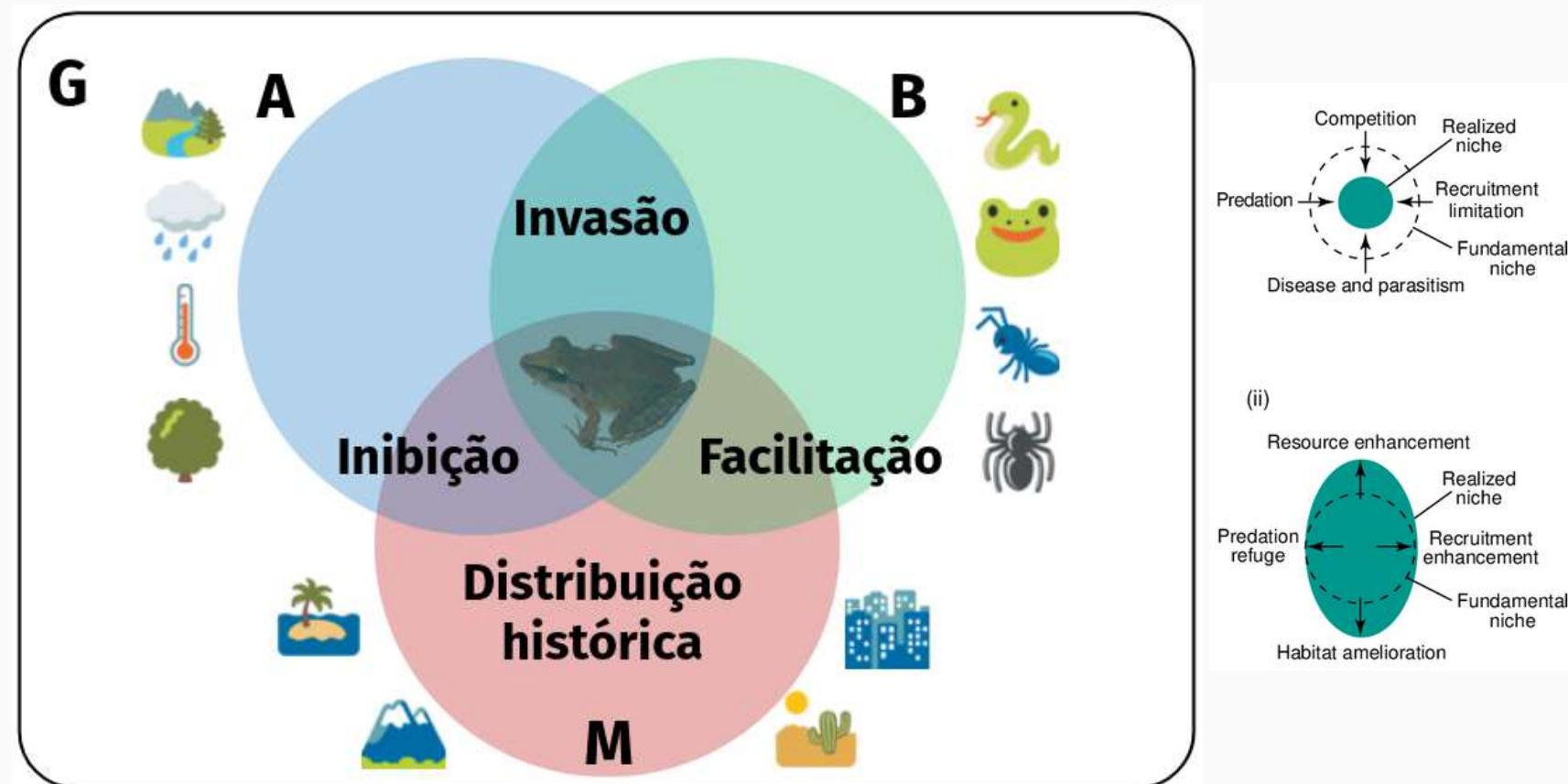
### Populações fonte e ralo (source-sink)

#### Inibição

- Competição
- Predação
- Parasitismo
- Amensalismo

#### Facilitação

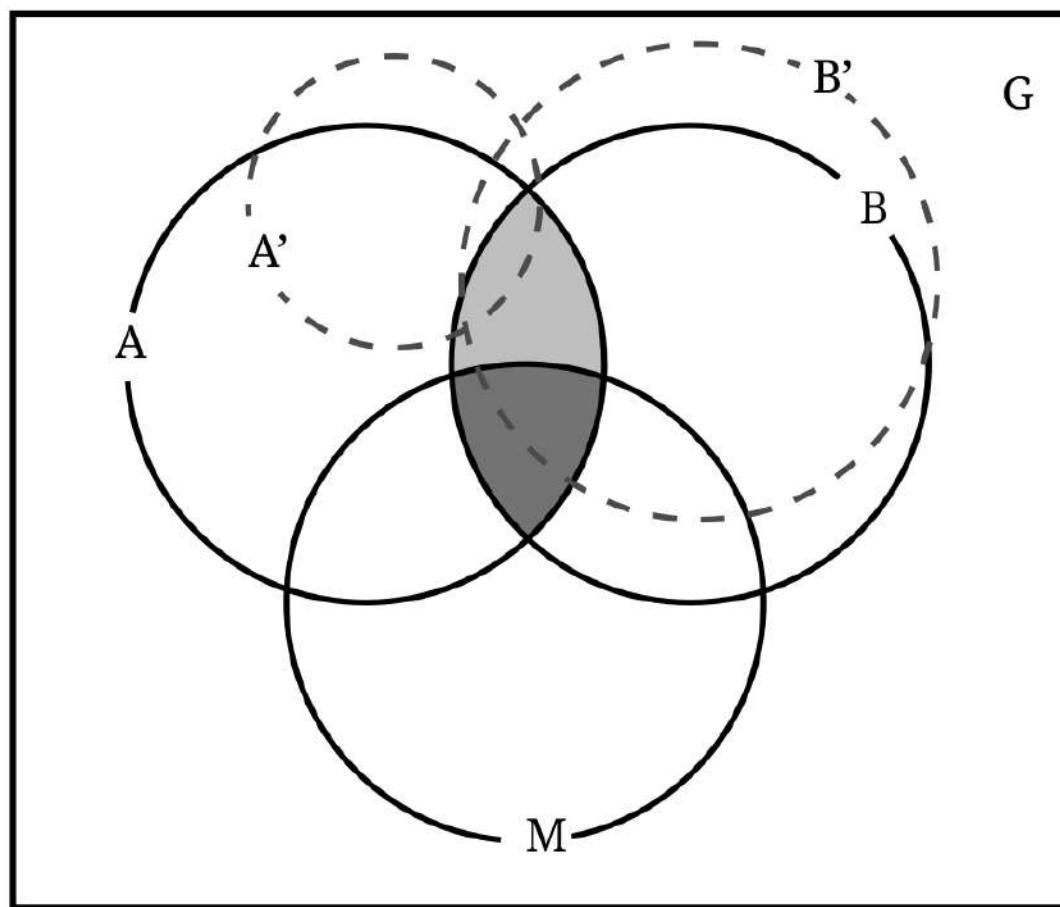
- Mutualismo
- Cooperação
- Comensalismo
- Inquilinismo
- Epifitismo



[Hutchinson \(1957\)](#), [Bruno et al. \(2003\)](#), [Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#).

## 2.2 Nicho ecológico e distribuição de espécies

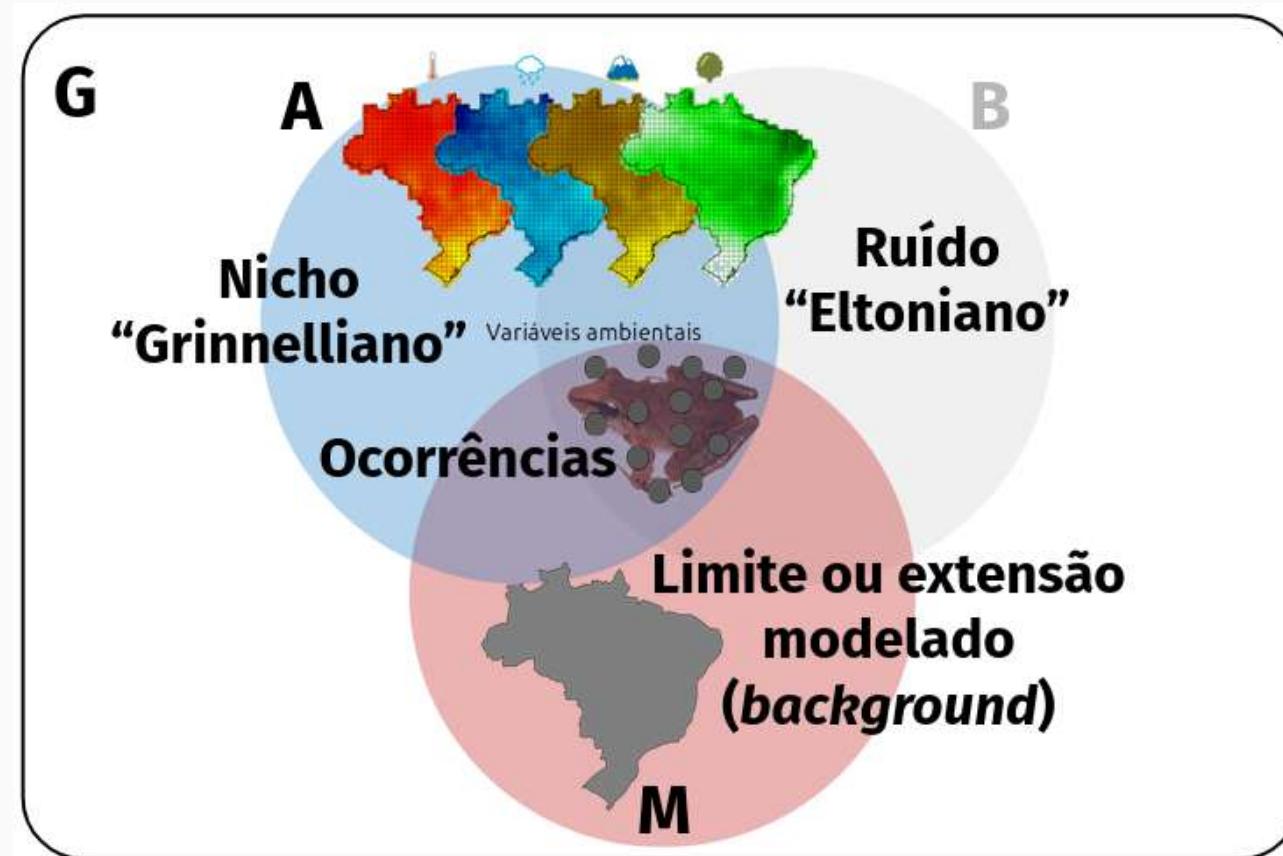
Variações das condições abióticas e bióticas em **função do tempo**



[Lima-Ribeiro & Diniz-Filho \(2013\)](#)

## 2.2 Nicho ecológico e distribuição de espécies

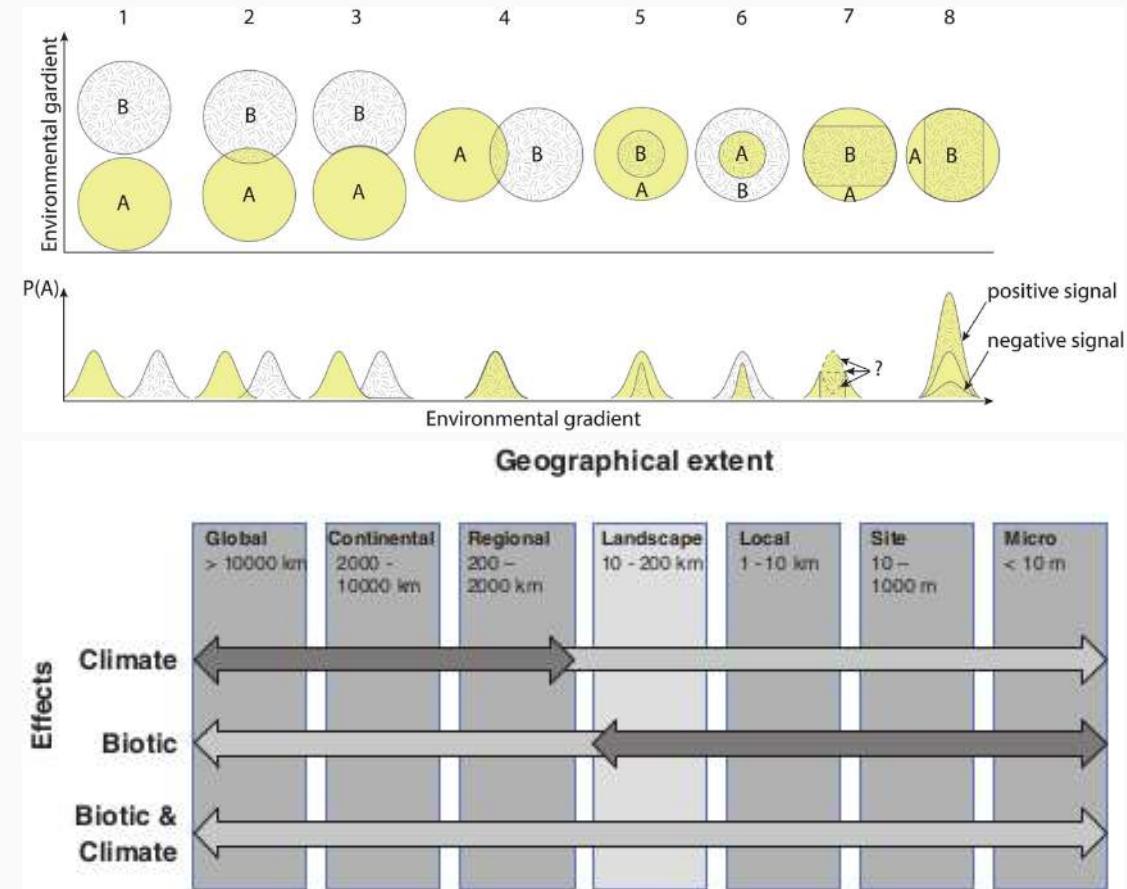
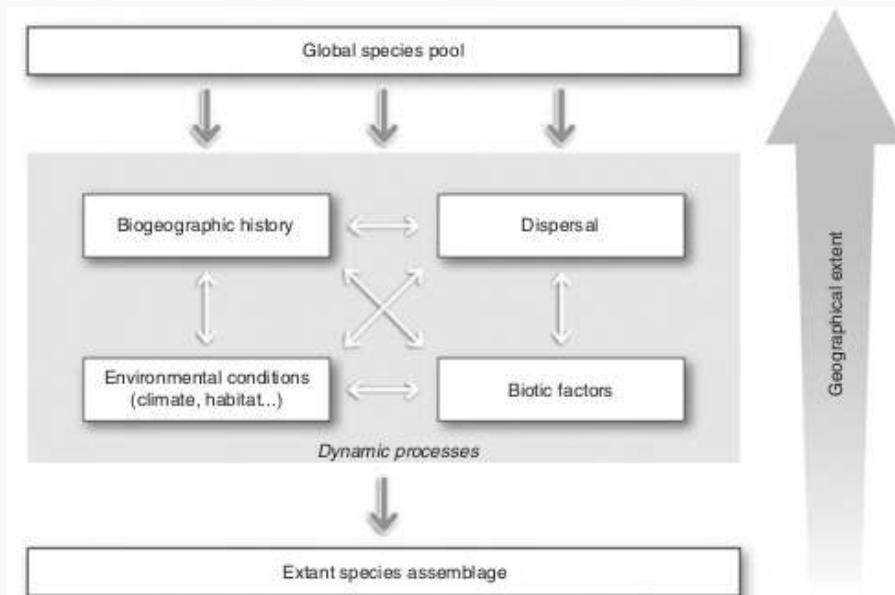
O que de fato modelamos?



[Soberón & Peterson \(2005\)](#), [Soberón \(2007\)](#), [Peterson et al. \(2011\)](#)

## 2.2 Nicho ecológico e distribuição de espécies

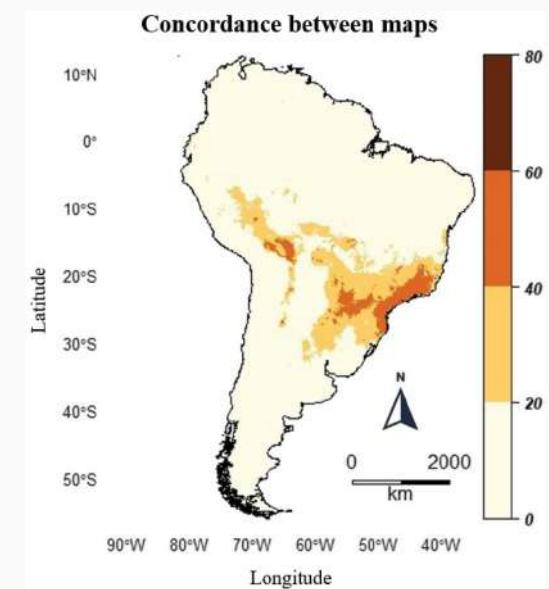
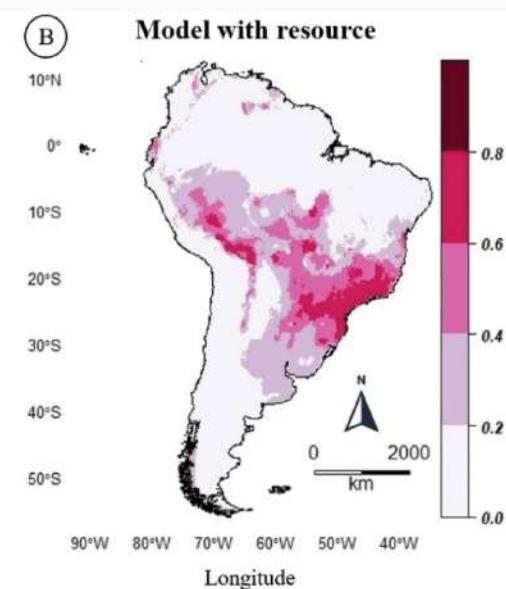
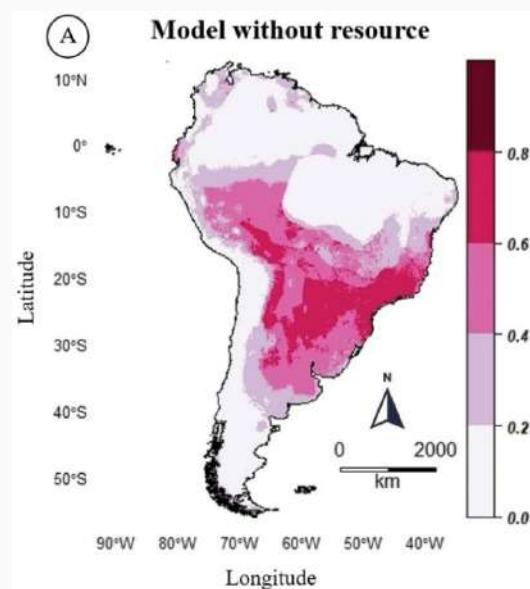
"Ruído Eltoniano": desconsideramos as condições bióticas



[Colwell & Rangel \(2009\)](#), [Soberón & Nakamura \(2009\)](#), [Wisz et al. \(2013\)](#), [Dormann et al. \(2018\)](#)

## 2.2 Nicho ecológico e distribuição de espécies

Entretanto...

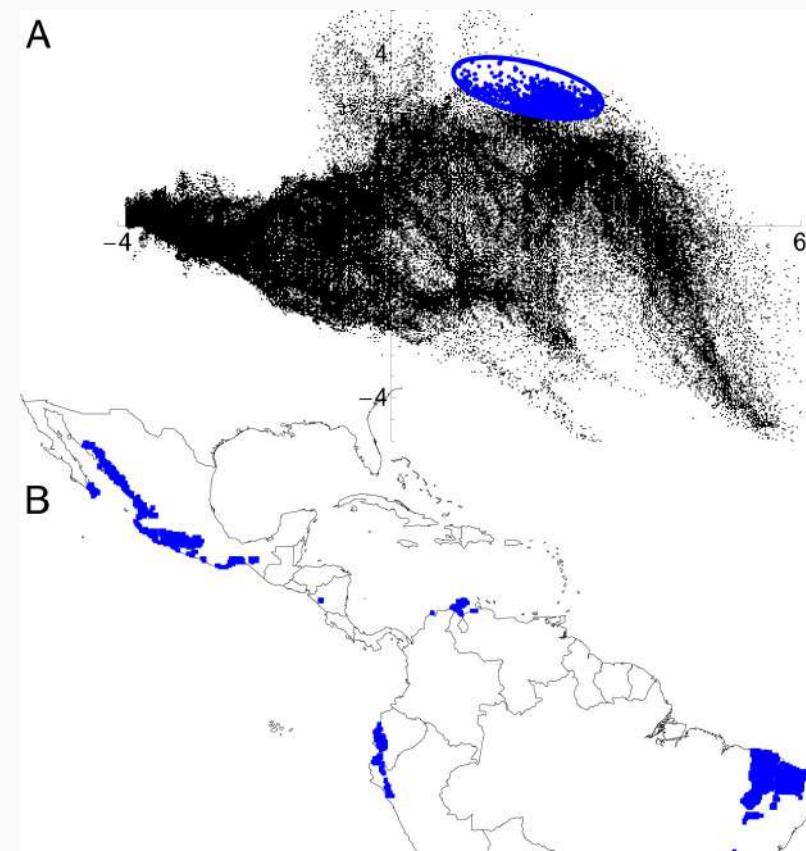


- Esquilo: *Sciurus aestuans* (catinguelê); Palmeira: *Syagrus romanzoffiana* (jerivá)  
Imagens: [Wikipedia](#), [Aves e Árvores](#), [Preuss & Padial \(2021\)](#).

# 2.2 Nicho ecológico e distribuição de espécies

## Pressupostos

- **Estabilidade do nicho ecológico:** as espécies mantêm o mesmo nicho ecológico ao longo do tempo, ou, pelo menos, entre os períodos utilizados na modelagem
- **Equilíbio espécie-clima:** distribuição de uma espécie é considerada em equilíbrio com o clima quando ela está presente em todas as áreas climaticamente adequadas à sua ocorrência e não ocorre nas regiões inadequadas
- **Controle taxonômico:** certeza em relação à taxonomia da espécie para as ocorrências (discutível)
- **"Ruído Eltoniano":** as interações bióticas não afetam a distribuição da espécie na macroescala



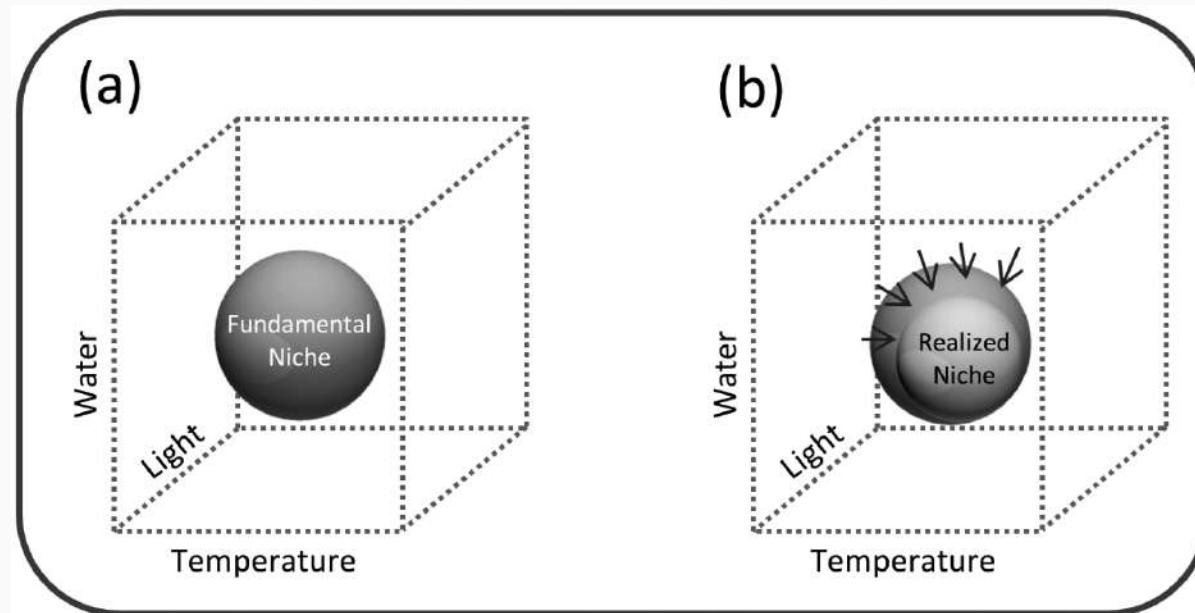
Bem, mas esses modelos estimam o **nicho realizado** sendo denominados **Modelos Correlativos**, pois correlacionam as ocorrências com as condições ambientais

Há ainda os **Modelos Mecanísticos** que estimam o  
**Nicho Fundamental**

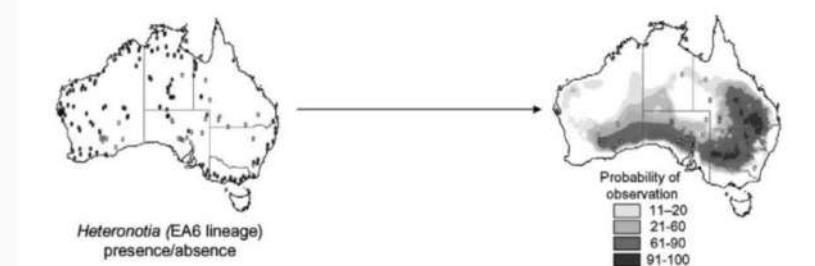
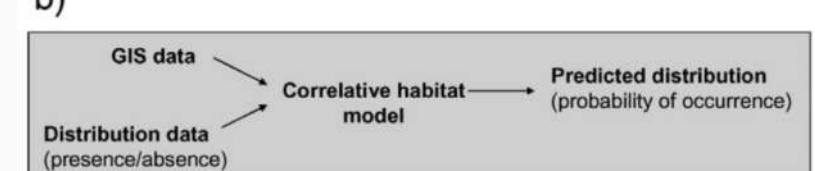
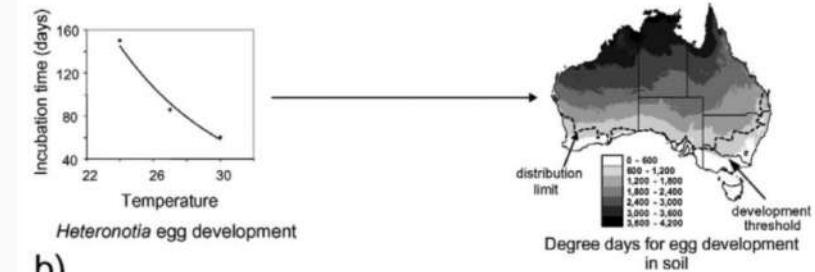
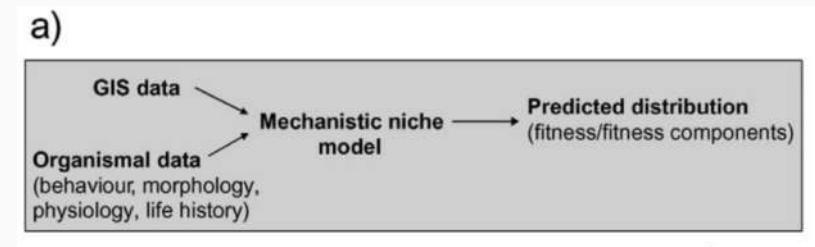
## 2.3 Tipos de modelos

### Modelos Mecanísticos

- Experimento laboratoriais para definir o nicho fundamental
- Variáveis ambientais altamente acuradas



[Kearney \(2006\)](#), [Guisan et al. \(2017\)](#)



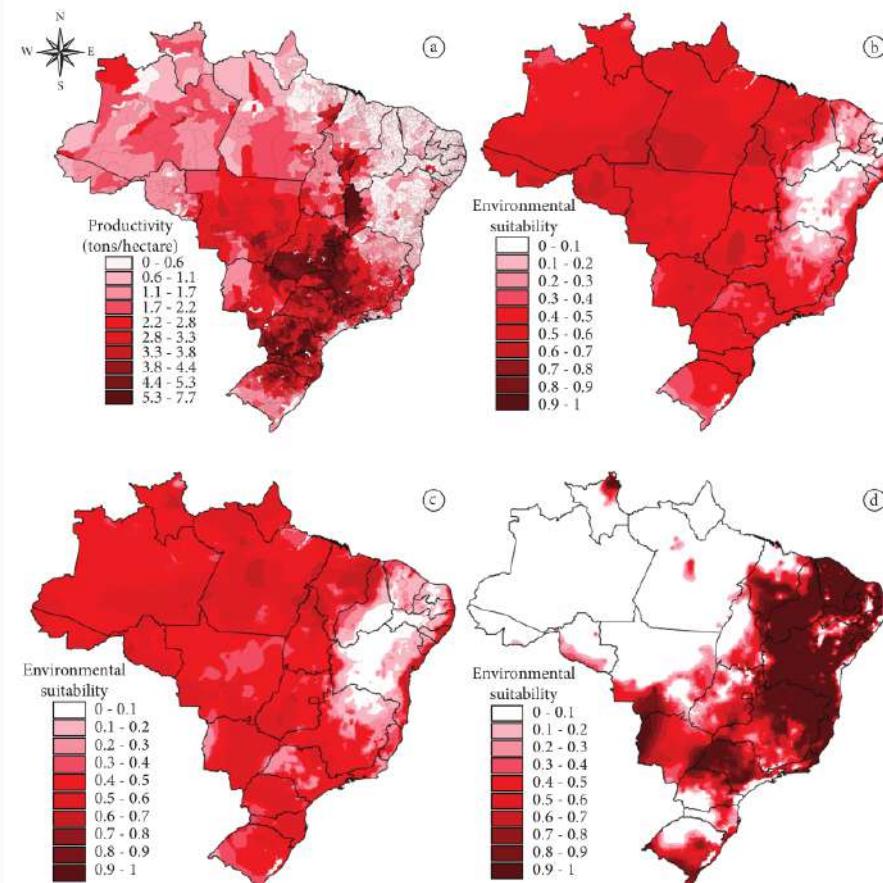
# 2.3 Tipos de modelos

## Modelos Mecanísticos

- (a) Produtividade de *Zea mays* nos municípios brasileiros
- (b) Modelo correlativo (MaxEnt)
- (c) Híbrido (MaxEnt + Saída do Plantgro)
- (d) Modelo mecanístico (Plantgro - DIVA-GIS)



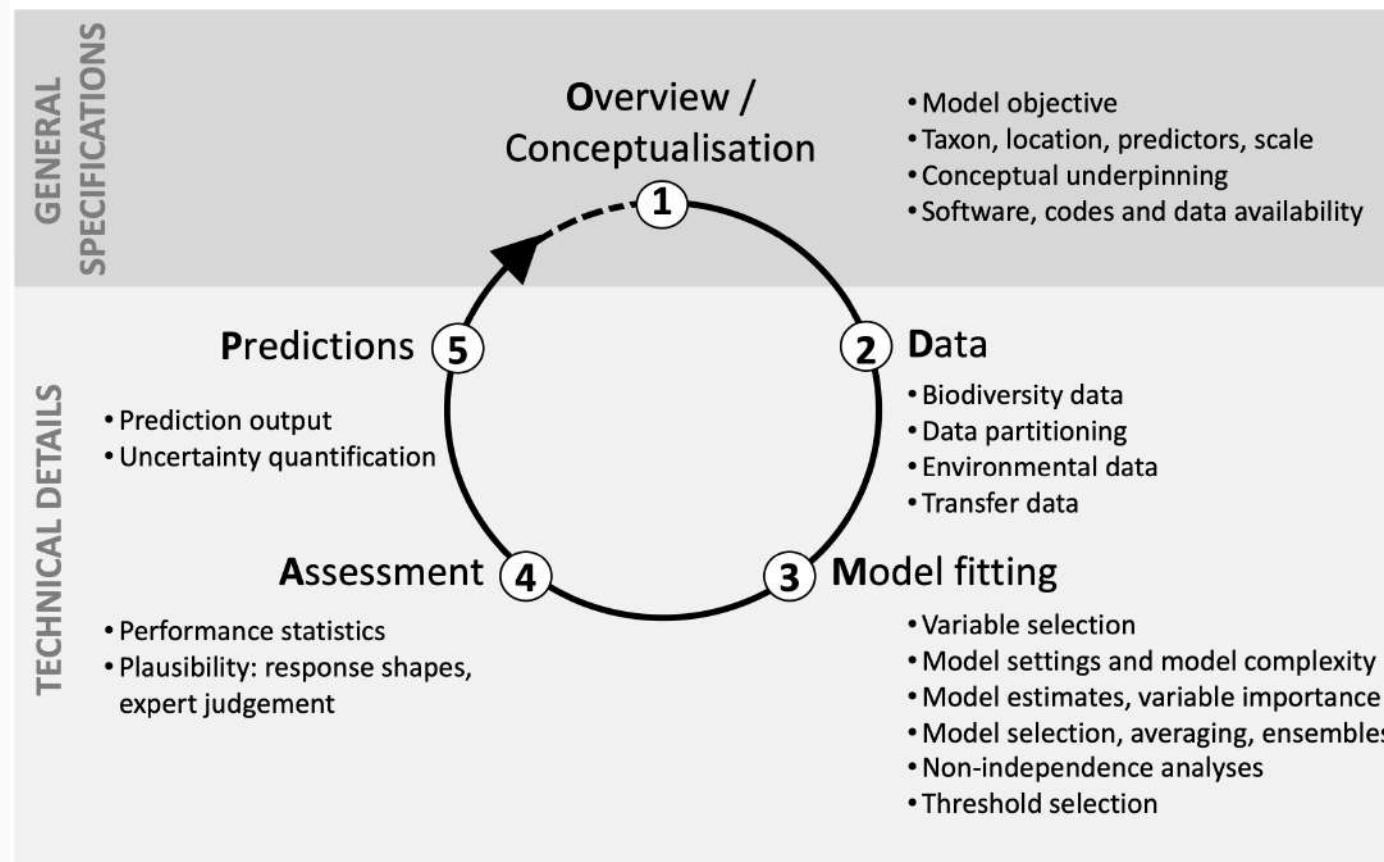
[Nabout et al. \(2012\)](#)



Dúvidas?

## 2.4 Padronização para criação dos modelos

### Protocolo padrão para modelos de distribuição de espécies



# 2.4 Padronização para criação dos modelos

## Protocolo padrão para modelos de distribuição de espécies

(A) ODMAP v1.0 | What is ODMAP? | How to use this app | Create a protocol | Protocol viewer | Upload / Import

### What is ODMAP?

Species distribution models (SDMs) constitute the most common class of biodiversity models. The advent of ready-to-use software packages and increasing availability of digital geo-information have considerably assisted the application of SDMs in recent years enabling their use in informing conservation and management, and quantifying impacts from global change.

(B) ODMAP v1.0 | What is ODMAP? | How to use this app | Create a protocol | Protocol viewer | Upload / Import

Progress  
Overview  
Data

1. Overview 2. Data 3. Model 4. Assessment 5. Prediction

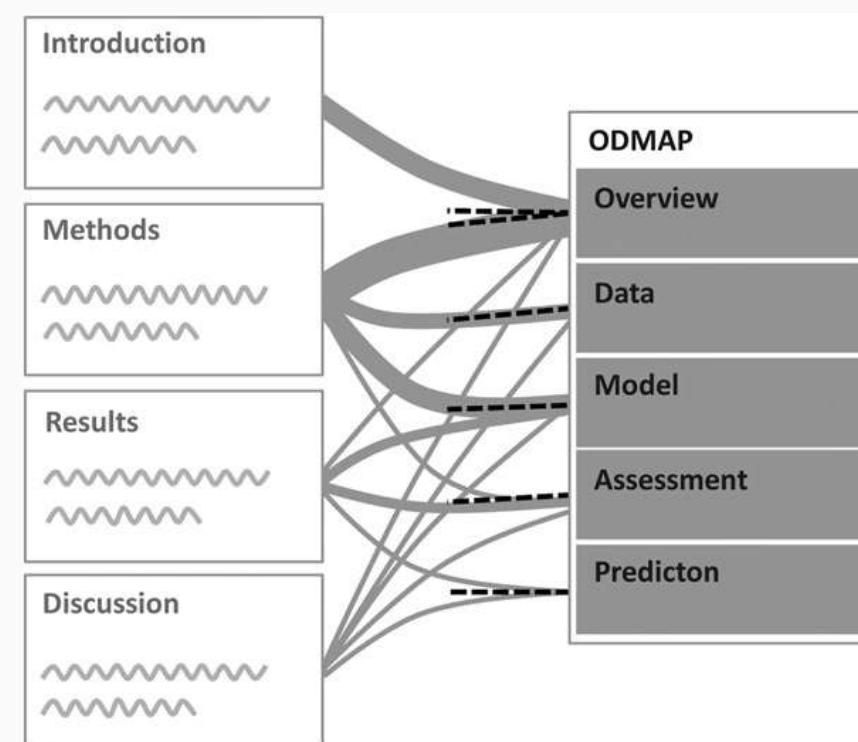
Give a brief overview of all important parts of your study.  
Authorship  
Predicting the distribution of shrub species in southern California from climate and terrain-derived variables

(C) ODMAP v1.0 | What is ODMAP? | How to use this app | Create a protocol | Protocol viewer | Upload / Import

- ODMAP protocol -  
Predicting the distribution of shrub species in southern California from climate and terrain-derived variables  
Authors: Janet Franklin  
Contact:  
Date: 2020-03-02

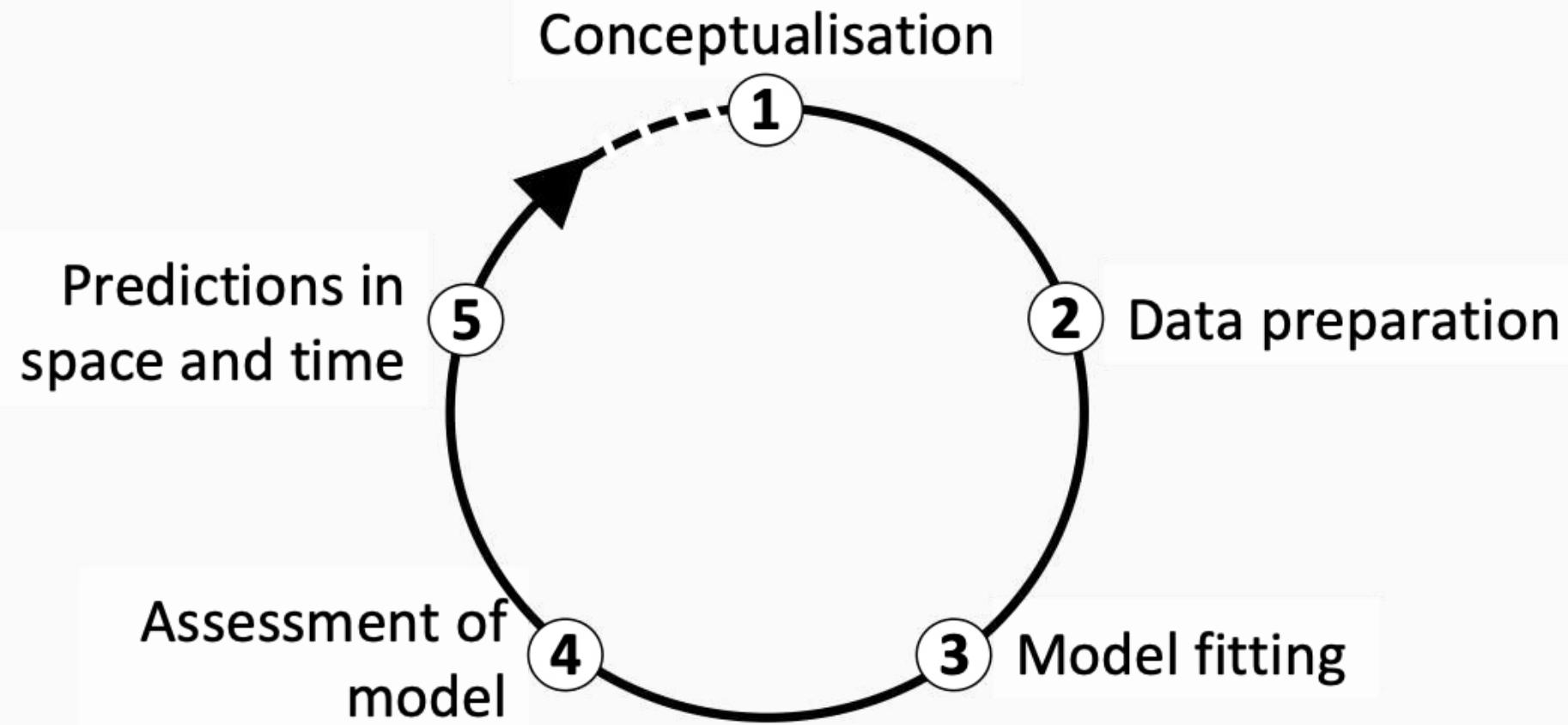
(D) ODMAP v1.0 | What is ODMAP? | How to use this app | Create a protocol | Protocol viewer | Upload / Import

There are two options for importing data into your ODMAP protocol  
(1) Upload an ODMAP protocol (.csv)  
This option is convenient if you want to edit or resume working on a previously saved ODMAP protocol.  
(2) Upload an RMM file (.RDS or .csv)



## 2.4 Padronização para criação dos modelos

Protocolo padrão para modelos de distribuição de espécies



# 2.4 Padronização para criação dos modelos

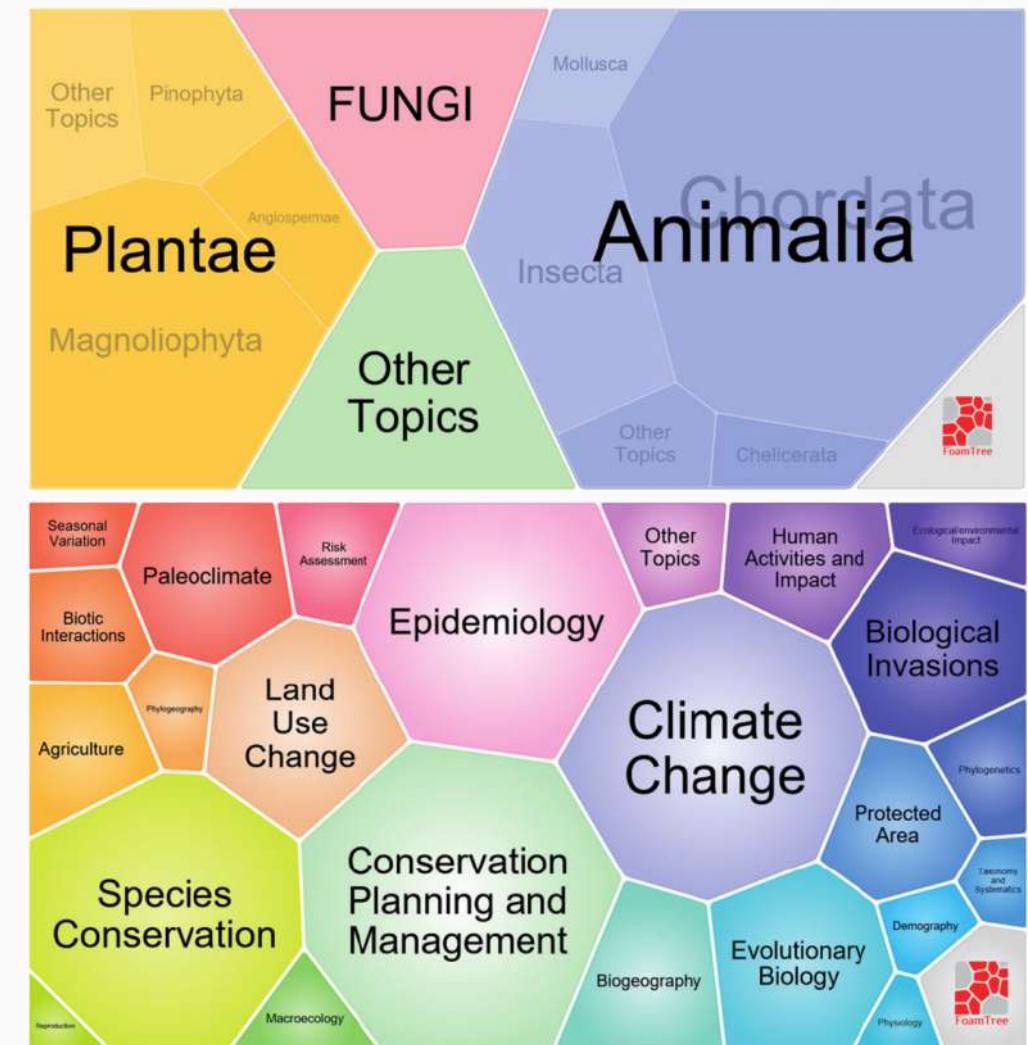
## 1. Conceitualização

### Temas

1. Distribuição de espécies
2. Padrões de diversidade
3. Mudanças climáticas (passado e futuro)
4. Invasão biológicas
5. Transmissão de doenças
6. Interações entre espécies
7. Processos de extinção
8. Conservação-evolução de nicho
9. Estabelecer refúgios climáticos
10. Estabelecimento e eficiência de áreas protegidas

**"Species Distribution Modeling in Latin America:  
A 25-Year Retrospective Review"**

[Urbina-Cardona et al. \(2019\)](#)

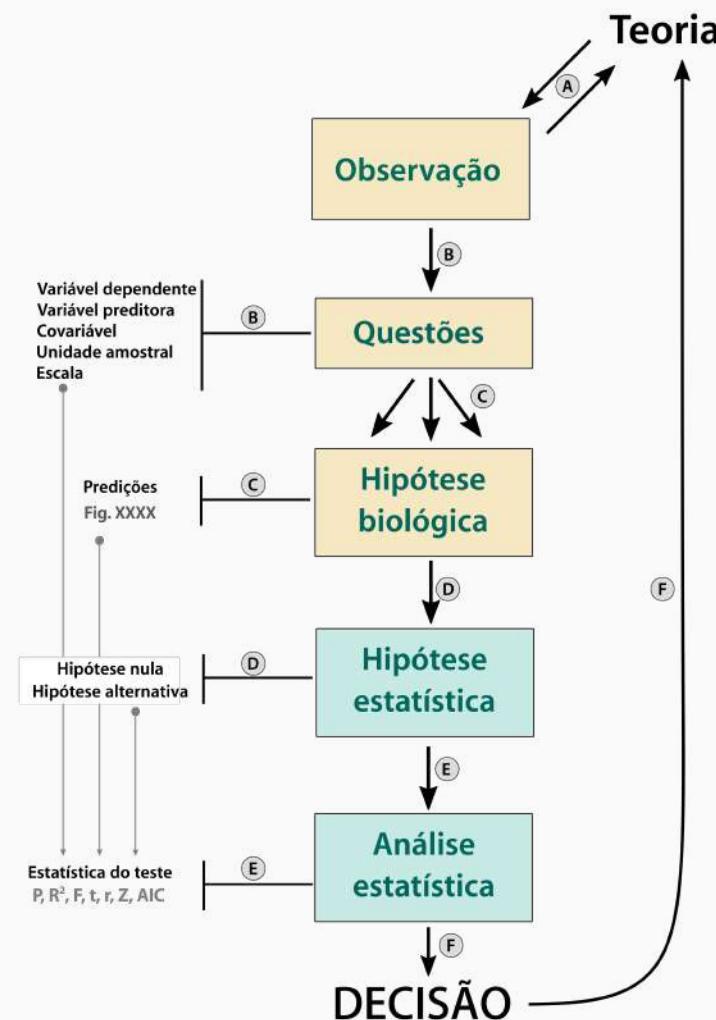


# 2.4 Padronização para criação dos modelos

## 1. Conceitualização

### Temas

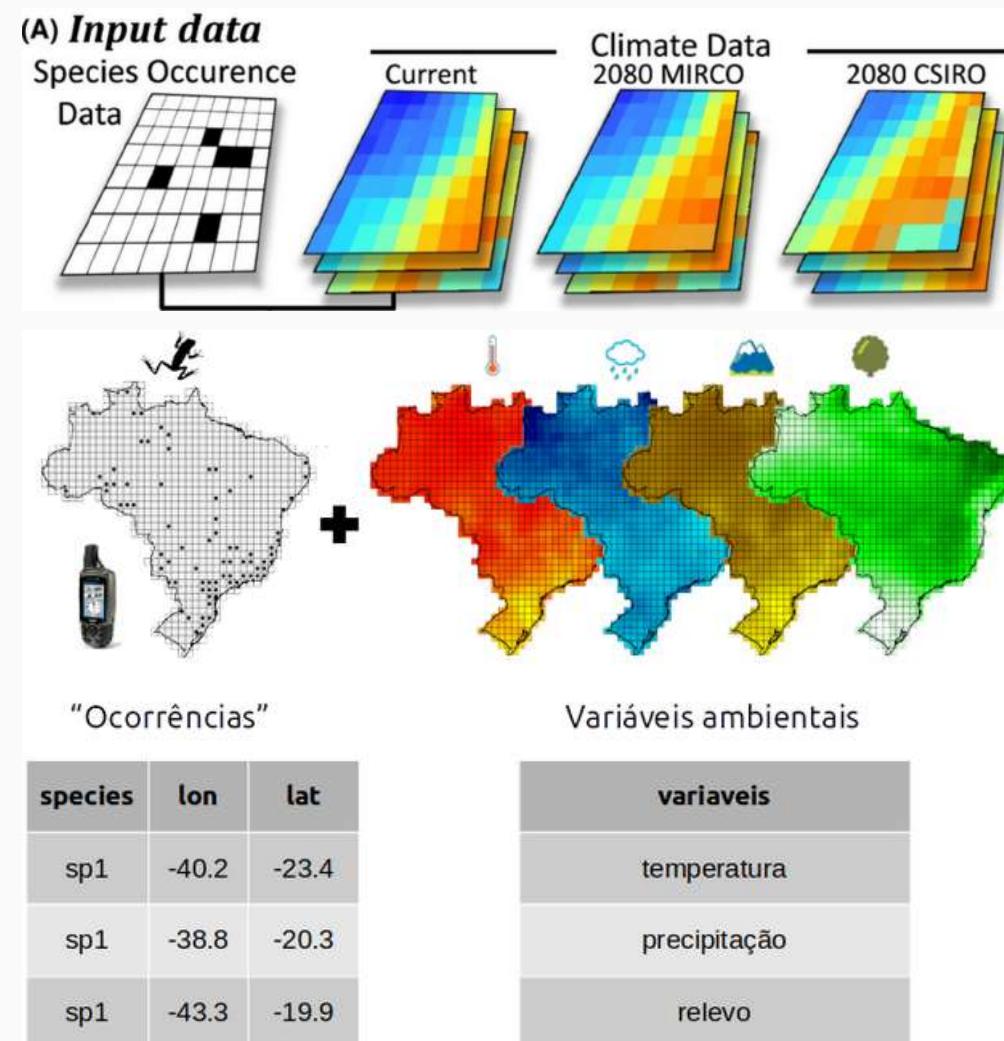
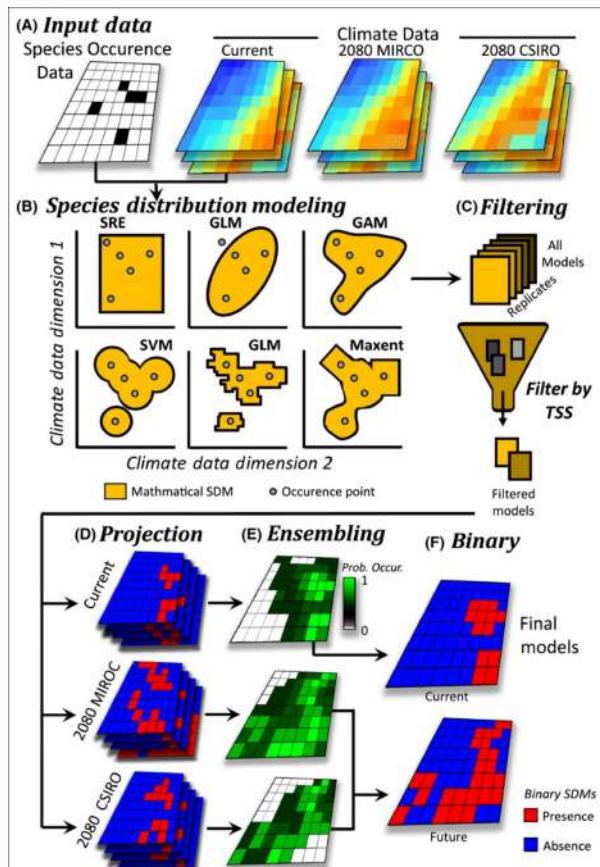
1. Distribuição de espécies
2. Padrões de diversidade
3. Mudanças climáticas (passado e futuro)
4. Invasão biológicas
5. Transmissão de doenças
6. Interações entre espécies
7. Processos de extinção
8. Conservação-evolução de nicho
9. Estabelecer refúgios climáticos
10. Estabelecimento e eficiência de áreas protegidas



Adaptado de Gonçalves-Souza et al. (2019).

# 2.4 Padronização para criação dos modelos

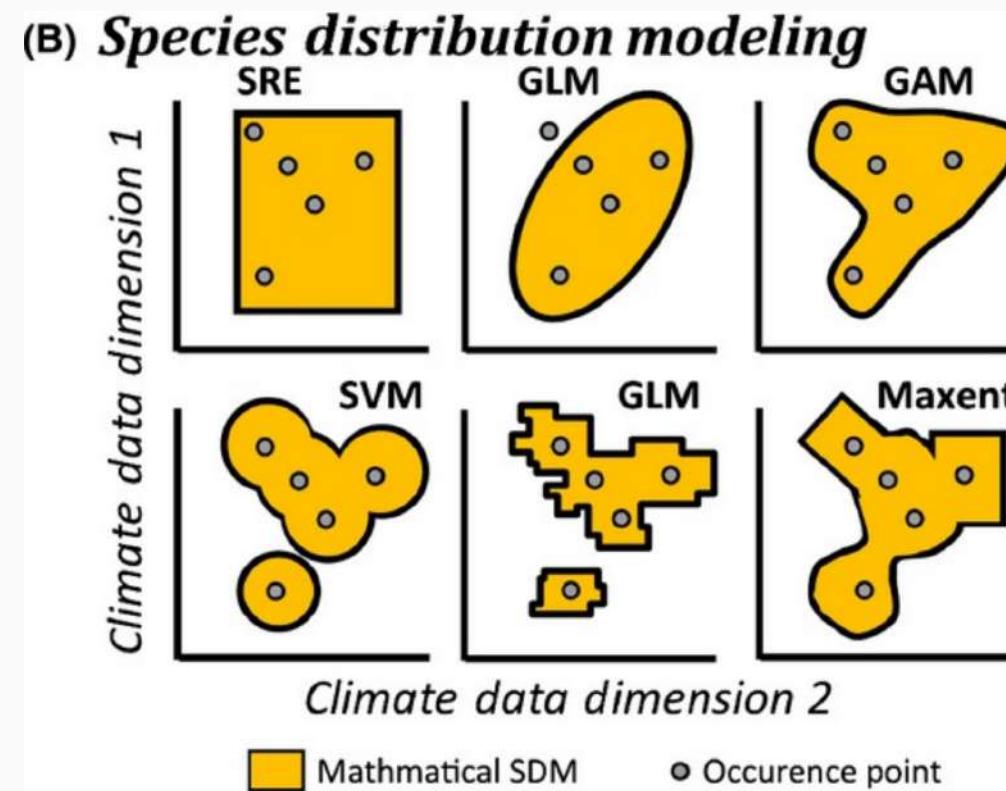
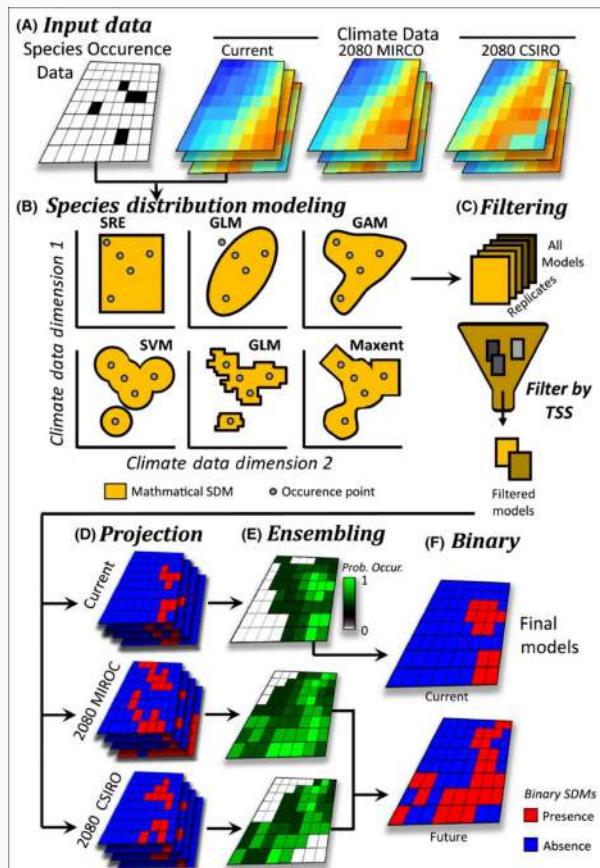
## 2. Preparação dos dados



[Brown & Yoder \(2015\)](#)

# 2.4 Padronização para criação dos modelos

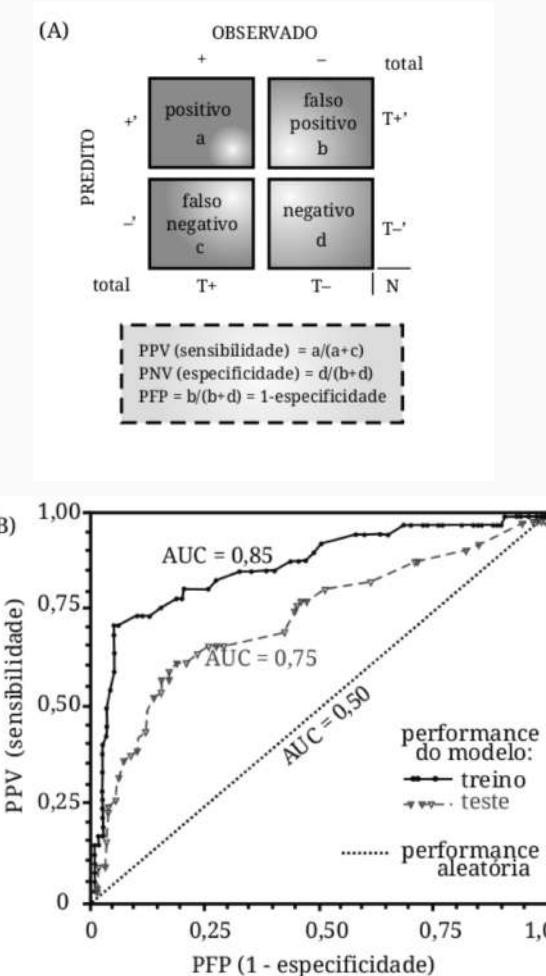
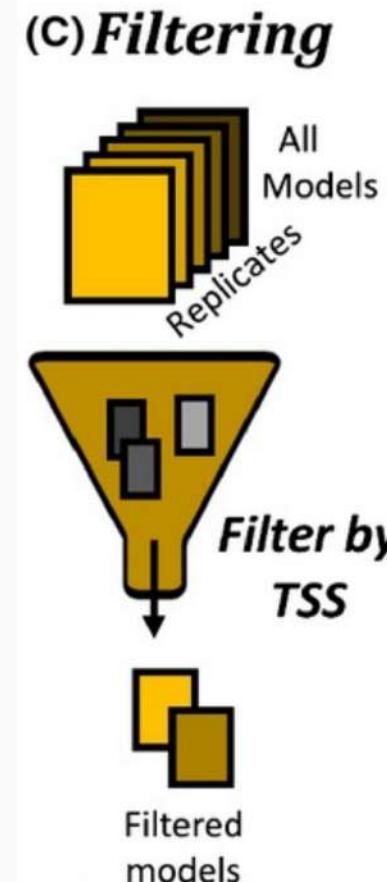
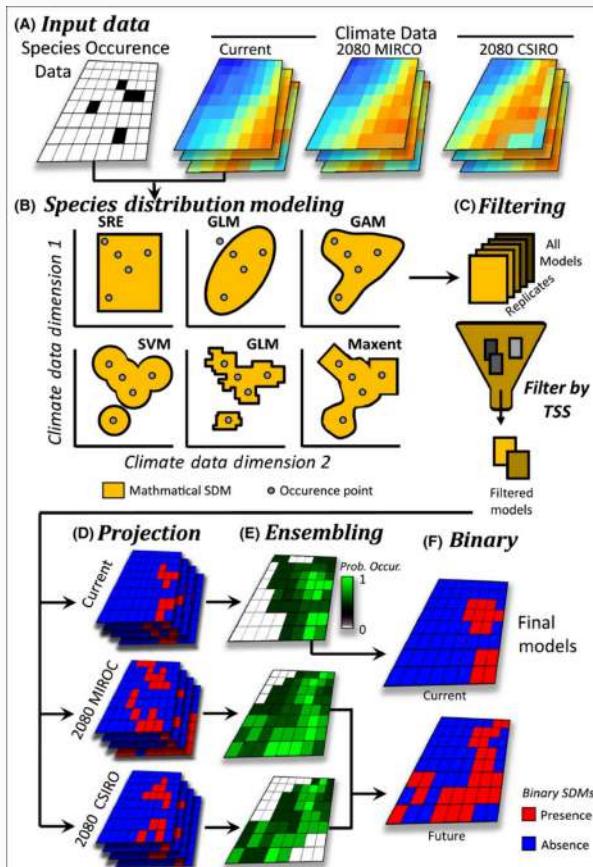
## 3. Ajuste dos modelos



Brown & Yoder (2015)

# 2.4 Padronização para criação dos modelos

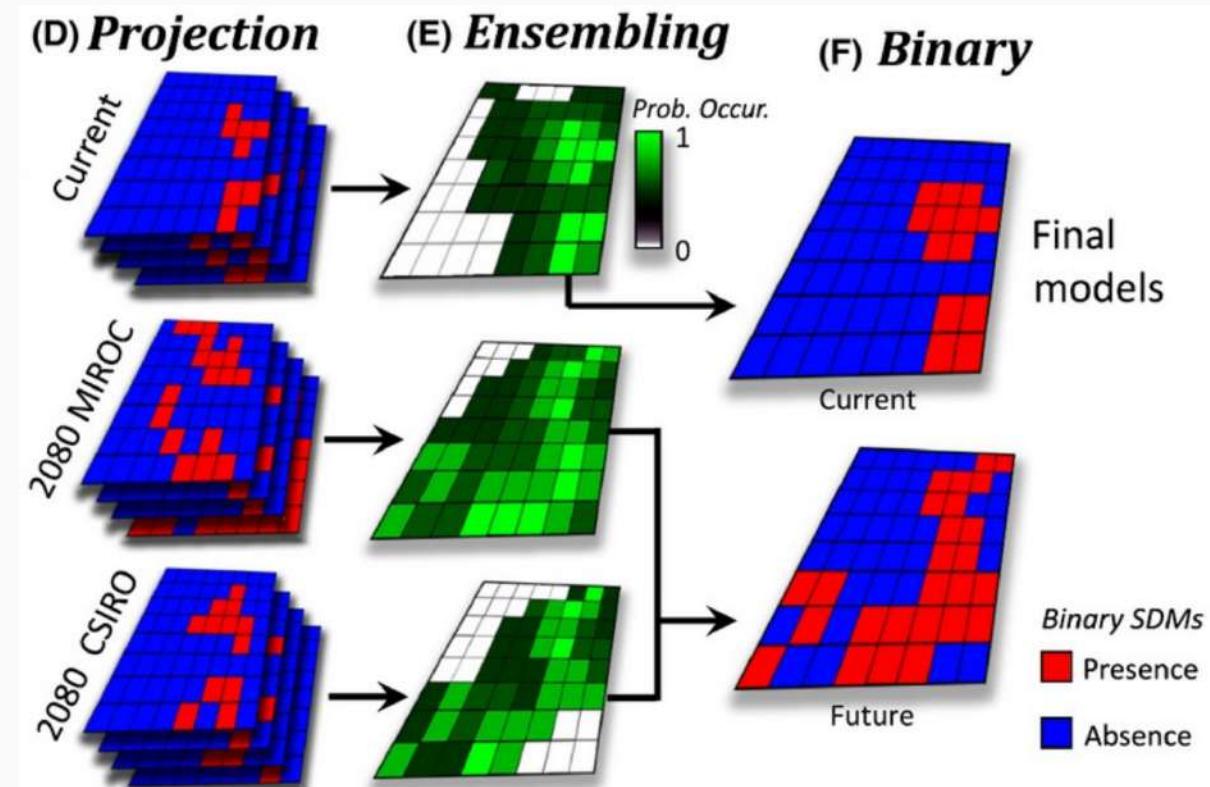
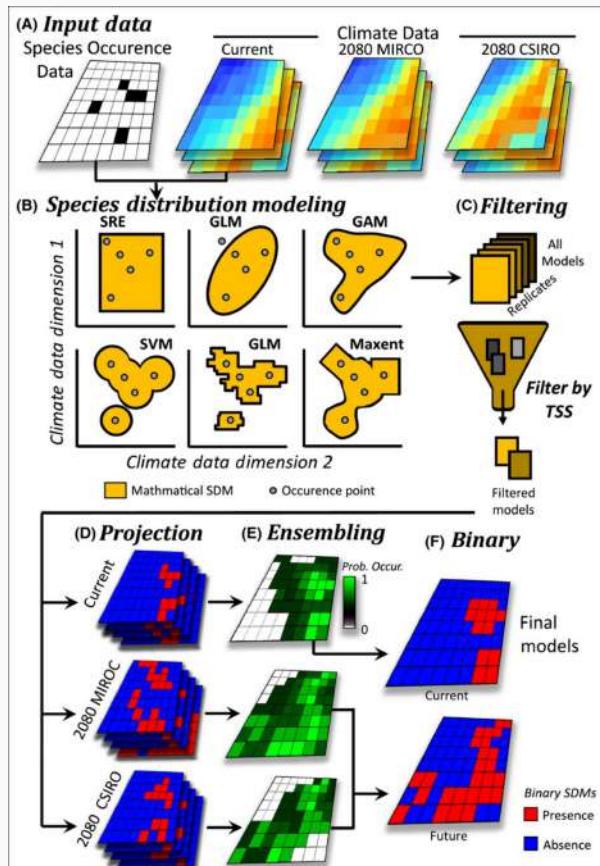
## 4. Avaliação (do ajuste) dos modelos



[Brown & Yoder \(2015\)](#), [Lima-Ribeiro & Diniz-Filho \(2013\)](#)

# 2.4 Padronização para criação dos modelos

## 5. Predições no tempo e no espaço



Dúvidas?

# 3 Introdução ao R

# 3.1 Linguagem R

## Definição

O R é uma **linguagem de programação livre** (*open source*), direcionada à **manipulação, análise e visualização de dados**, com diversas **expansões** (*pacotes*) para o uso de **dados com formatos específicos**



## 3.1 Linguagem R

E por que usar o R?

WHAT GIVES PEOPLE  
FEELINGS OF POWER



# 3.1 Linguagem R

## Cinco motivos para usar R

1. R é completamente **gratuito**
2. Grandes pessoas da comunidade **disponibilizam** seu trabalho em R
3. R possui um ecossistema que incentiva a **reprodutibilidade**
4. R tem uma **comunidade** vibrante e crescente
5. Os focos da linguagem são **modelagem, visualização e análise de dados**

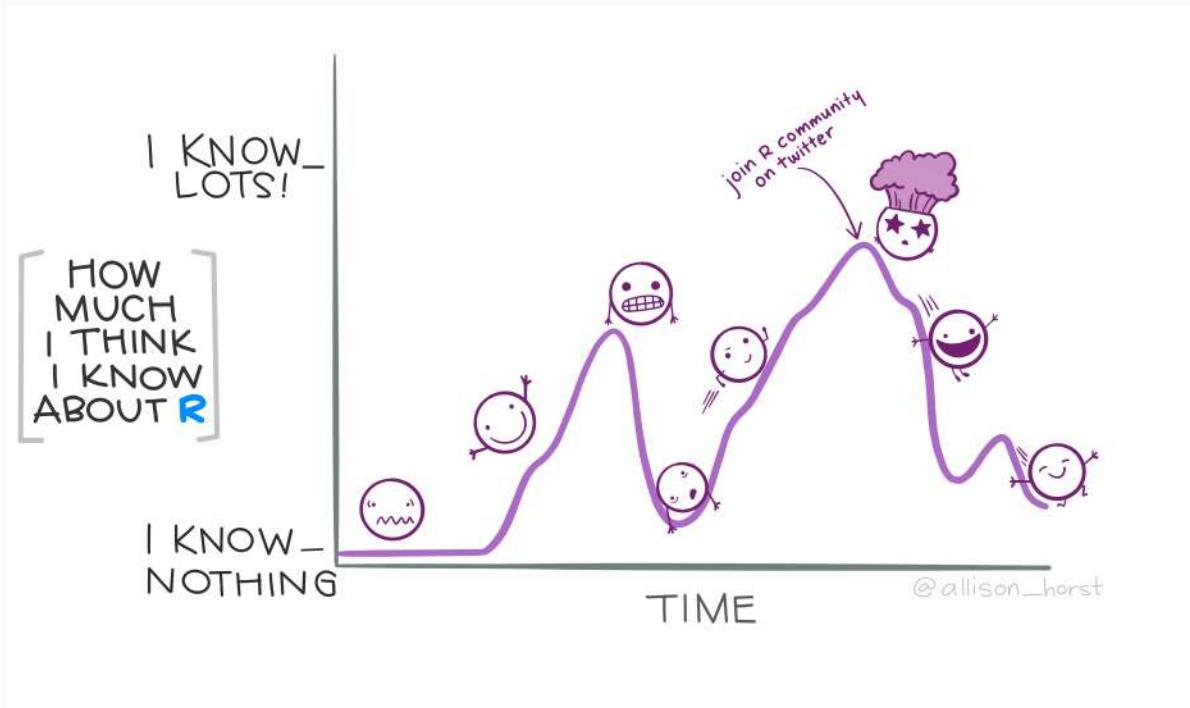


[Curso-R](#)

**5** MOTIVOS  
PARA  
USAR R

# 3.1 Linguagem R

É legal, mas nem sempre é fácil...



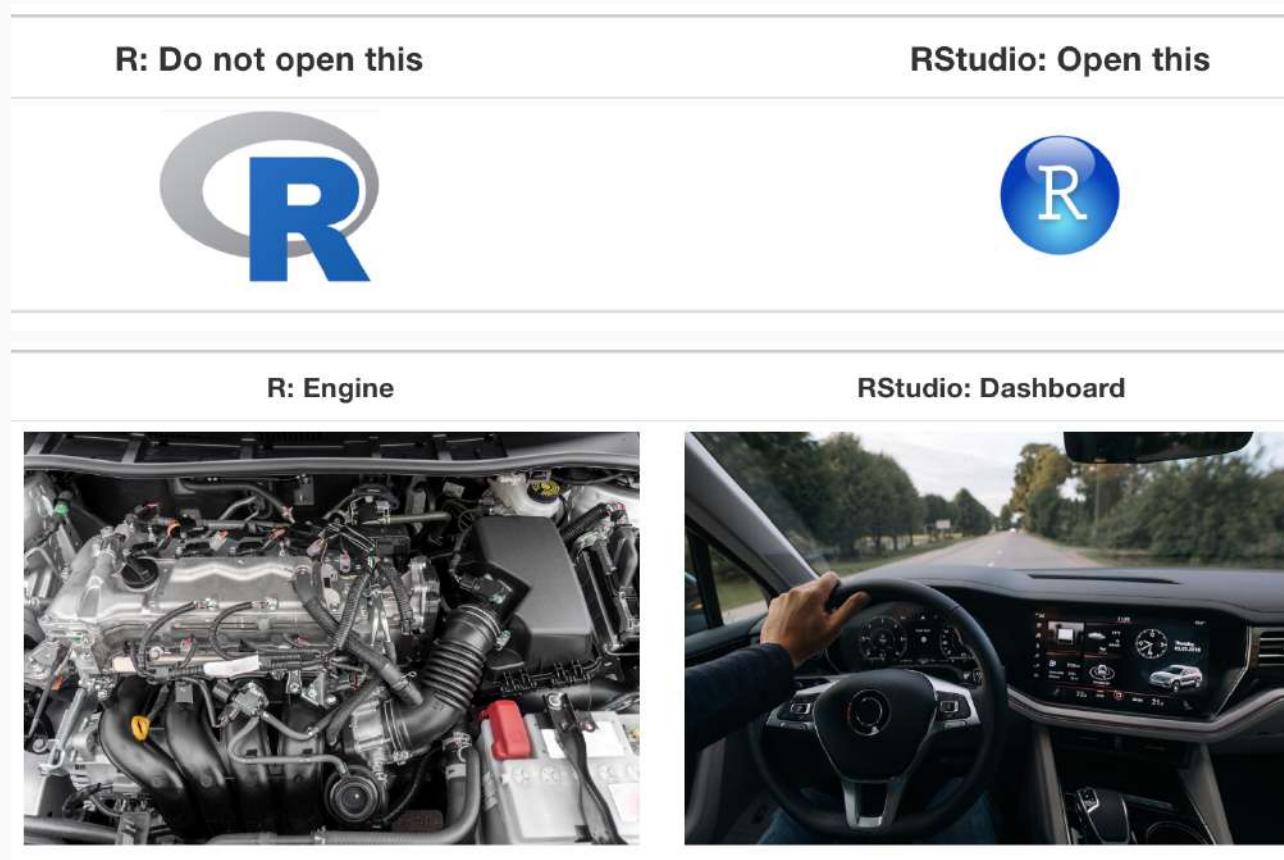
[@allison\\_horst](https://allisonhorst.com)



# 3.2 RStudio

## IDE

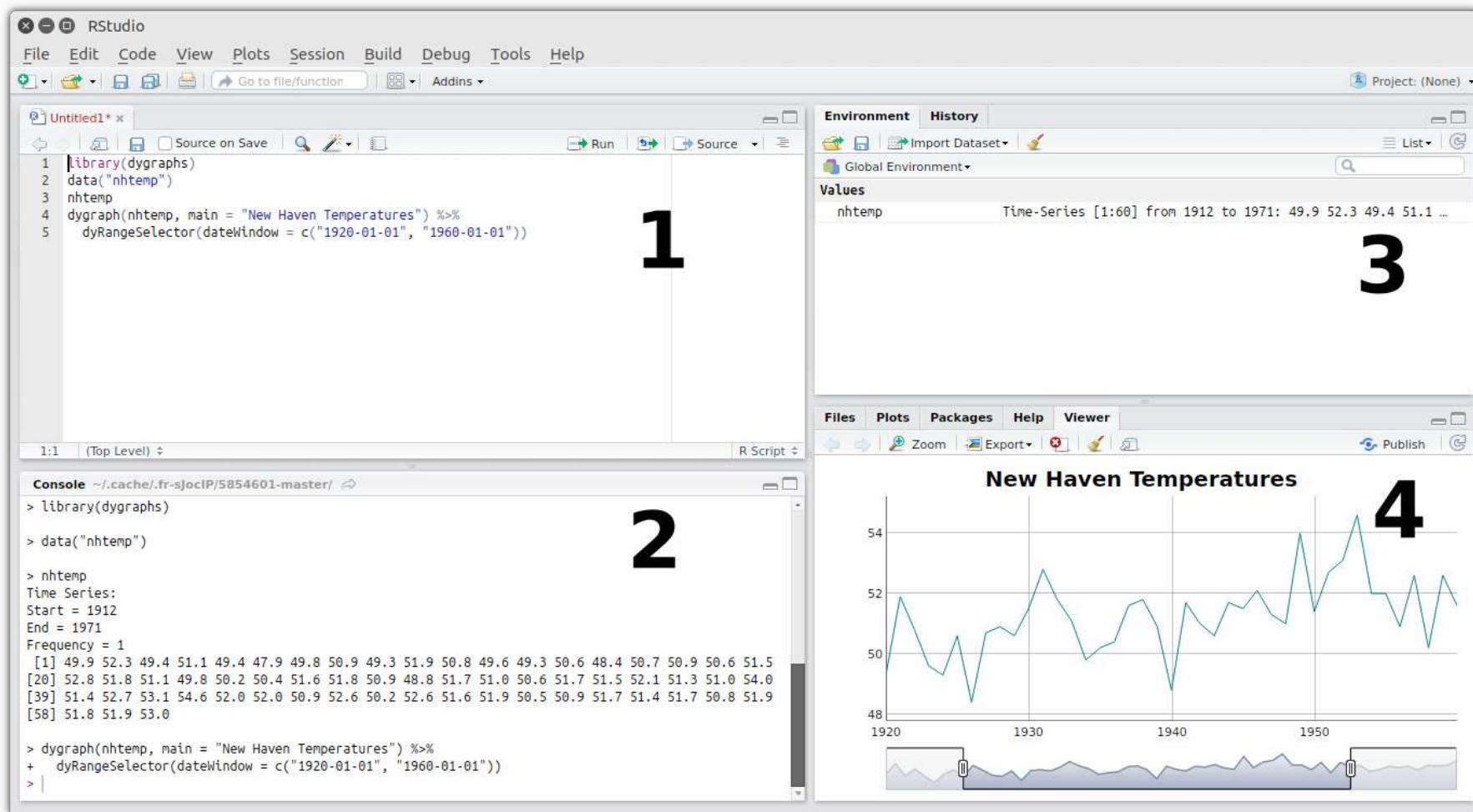
*Integrated Development Environment ou Ambiente de Desenvolvimento Integrado*



[Ismay & Kim \(2020\)](#)

# 3.2 RStudio

## Interface



# 3.2 RStudio

## Janelas e abas

**1. Editor/Script:** é onde escrevemos nossos códigos em R ou R Markdown

**2. Console:** é onde os códigos são rodados e vemos as saídas

**3. Environment:** painel com todos os objetos criados na sessão

**3. History:** painel com o histórico dos códigos rodados

**3. Connection:** painel para conectar banco de dados

**3. Git:** painel do controle de versão

**3. Tutorial:** painel de tutoriais

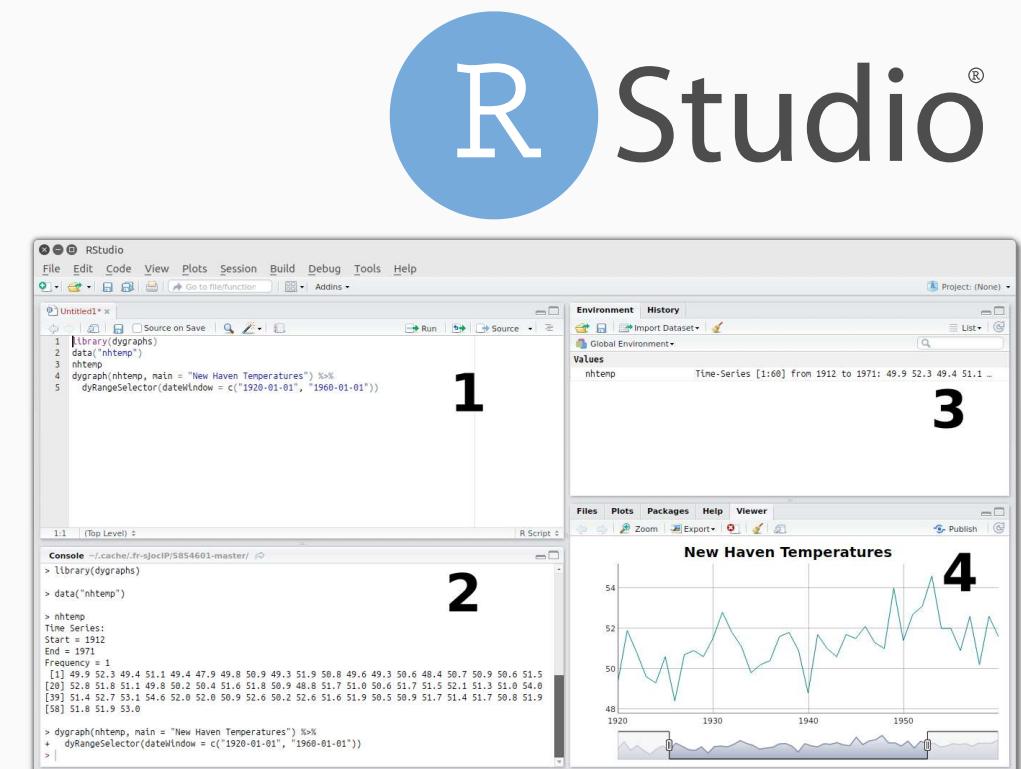
**4. Files:** painel que mostra os arquivos no diretório de trabalho

**4. Plots:** painel onde os gráficos são apresentados

**4. Packages:** painel que lista os pacotes

**4. Help:** painel onde a documentação das funções é exibida

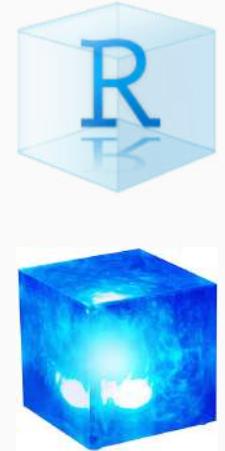
**4. Viewer:** painel de visualização



# 3.2 RStudio

## Projeto R (.Rproj) (Tesseract)

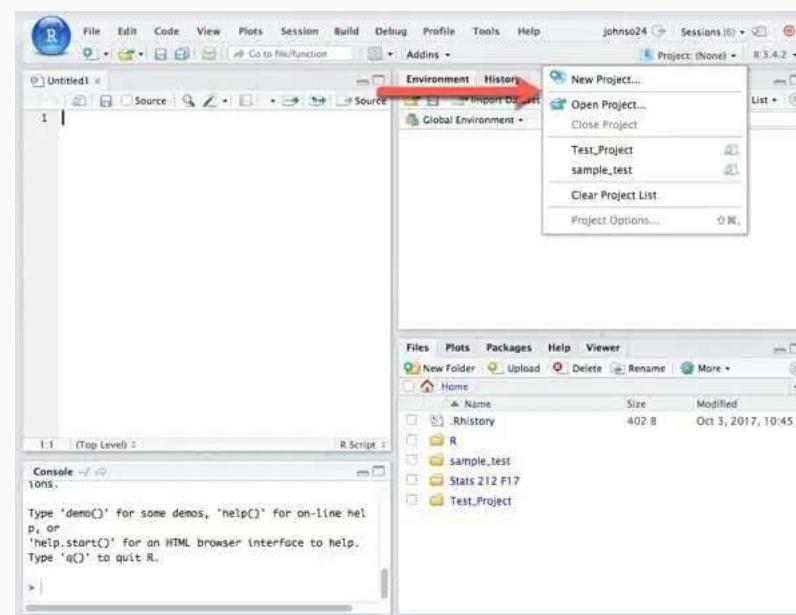
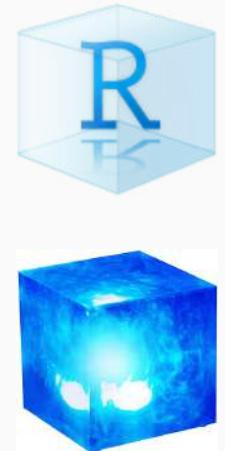
- Facilita o trabalho em múltiplos ambientes
- Cada projeto possui seu diretório, documentos e workspace
- Permite versionamento



## 3.2 RStudio

### Projeto R (.Rproj)

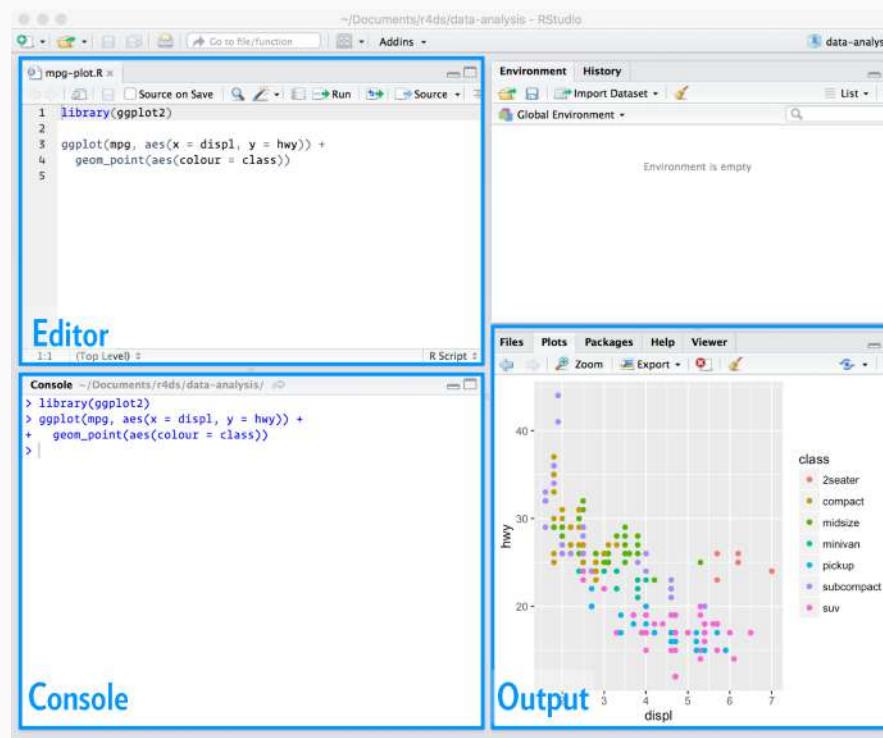
- Sempre **abram o RStudio** pelo arquivo **.Rproj**
- Ou **definam o projeto** depois de abrir o RStudio



### 3.3 Console

#### Console

O console é onde a versão da linguagem R instalada é carregada para executar os códigos da linguagem R



### 3.3 Console

#### Console

O console é onde a versão da linguagem R instalada é carregada para executar os códigos da linguagem R

```
10 + 2
```

```
## [1] 12
```

```
1:42
```

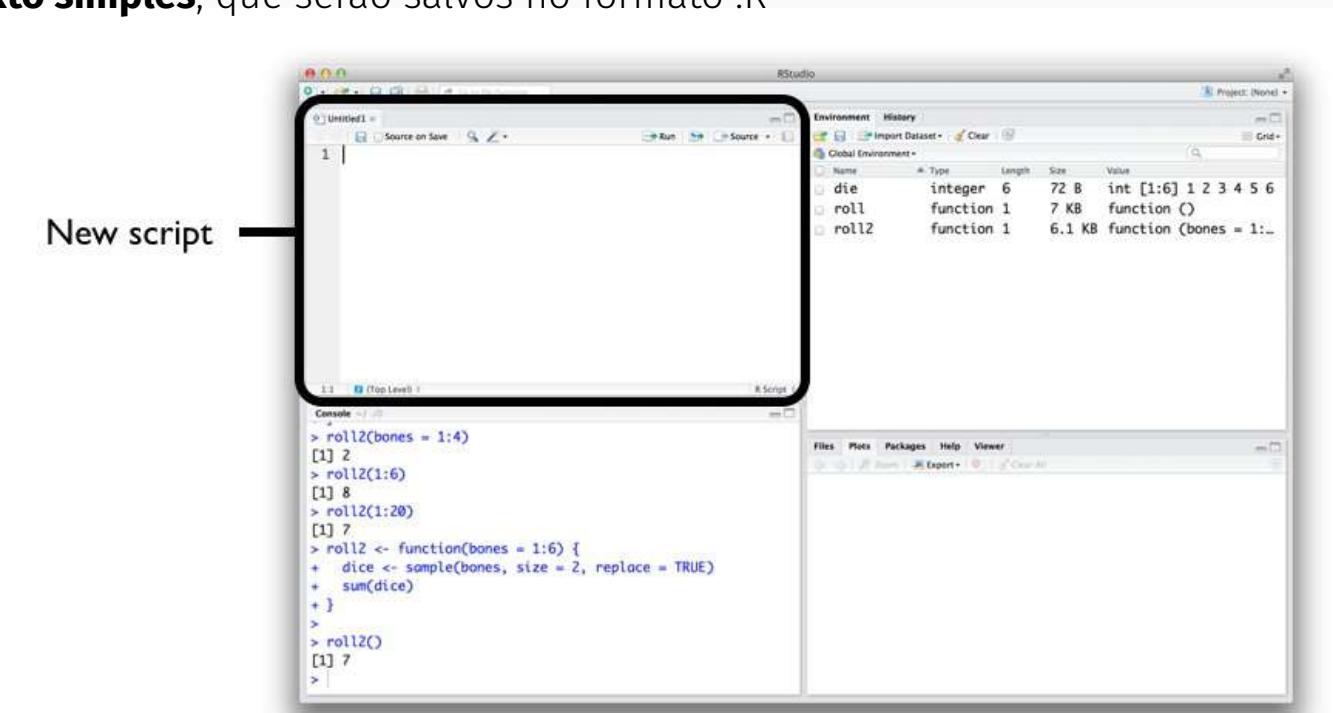
```
## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
```

# 3.4 Scripts

## Script (.R)

Atalho: Ctrl + Shift + N

- São **rascunhos** dos códigos
- Será neles que os **códigos serão escritos** e depois **enviados ao console do R**
- São **arquivos de texto simples**, que serão salvos no formato .R



A screenshot of the RStudio interface. A new script file titled "Untitled1.R" is open in the top-left pane, indicated by a black arrow pointing to it with the text "New script". The code in the script file is:

```
1 | 
|> roll2(bones = 1:4)
| [1] 2
|> roll2(1:6)
| [1] 8
|> roll2(1:20)
| [1] 7
|> roll2 <- function(bones = 1:6) {
+   dice <- sample(bones, size = 2, replace = TRUE)
+   sum(dice)
+ }
|>
|> roll2()
| [1] 7
|>
```

The bottom-left pane shows the R Console with the same code history. The right side of the interface includes the Environment, History, Global Environment, and Files panes.

## 3.4 Scripts

Todos os **códigos** devem ser digitados preferencialmente no **script**

Deixem o **cursor** em **qualquer local da linha** e executem essa linha utilizando essa **combinação**:

**Atalho:** Ctrl + Enter

Vamos testar:

```
1
```

```
## [1] 1
```

```
1 + 2
```

```
## [1] 3
```

E é isso que faremos pelo resto de nossas vidas...

## 3.4 Scripts

### Comentários (#)

Comentários **não são lidos** pelo R e **descrevem informações** em nosso script

São representados pelo **#** (hash) ou **#'** (hash-linha)

### Informações sobre os códigos

```
## comentarios  
# o r nao le o codigo depois do # (hash)  
42 # essas palavras nao sao executadas, apenas o 42
```

```
## [1] 42
```

## 3.4 Scripts

### Comentários (#)

Comentários **não são lidos** pelo R e **descrevem informações** em nosso script

São representados pelo **#** (hash) ou **#'** (hash-linha)

### Cabeçalho

```
#' ---
#' title: Introdução à modelagem de distribuição de espécies usando a linguagem R
#' author: seu nome
#' date: 25-09-2021
#' ---
```

# 3.5 Operadores

## Operadores

### Operadores aritméticos (Números)

Operador	Descrição	Uso
+	Adição	$a + b$
-	Subtração	$a - b$
*	Multiplicação	$a * b$
/	Divisão	$a / b$
%%	Resto da divisão	$a \% b$
%/%	Quociente da divisão	$a \%/% b$
^	Potenciação	$a^b$

### Operadores relacionais (TRUE|FALSE)

Operador	Descrição	Uso
<	Menor	$a < b$
>	Maior	$a > b$
==	Igual	$a == b$
<=	Menor ou igual	$a <= b$
>=	Maior ou igual	$a >= b$
!=	Não igual (diferente)	$a != b$

# 3.5 Operadores

## Ordem das operações aritméticas

^ >> \* ou / >> + ou -

```
# sem especificar - segue a ordem  
1 * 2 + 2 / 2 ^ 2
```

```
## [1] 2.5
```

```
# especificando - segue a ordem dos parênteses  
((1 * 2) + (2 / 2)) ^ 2
```

```
## [1] 9
```

# 3.6 Objetos

## Atribuição (<-)

Objetos são palavras que **atribuímos** dados

A atribuição possibilita a **manipulação** de dados ou resultados de análises

Utilizaremos o símbolo "<" (**menor**) seguido de "-" (**menos**), sem espaço!!!

**palavra <- dados**

**Atalho:** Alt + -



## 3.6 Objetos

Vamos atribuir o **valor 10** à palavra **obj10**

```
## atribuicao - simbolo (←)  
obj10 ← 10
```

Agora a palavra **obj10** vale **10**

Mas não aconteceu nada....



## 3.6 Objetos

Sempre **confira** a atribuição!!!

Chame o objeto **novamente**!!!

```
## atribuicao - simbolo (←)
obj10 ← 10
obj10
```

```
## [1] 10
```

Outro exemplo

```
## atribuicao - simbolo (←)
obj2 ← 2
obj2
```

```
## [1] 2
```

## 3.6 Objetos

CUIDADO!

O R **sobrescreve** os valores dos objetos com o **mesmo nome**!

```
# sobrescreve o valor dos objetos
obj ← 100
obj
```

```
## [1] 100
```

```
# obj agora vale 2
obj ← 2
obj
```

```
## [1] 2
```

# 3.7 Funções

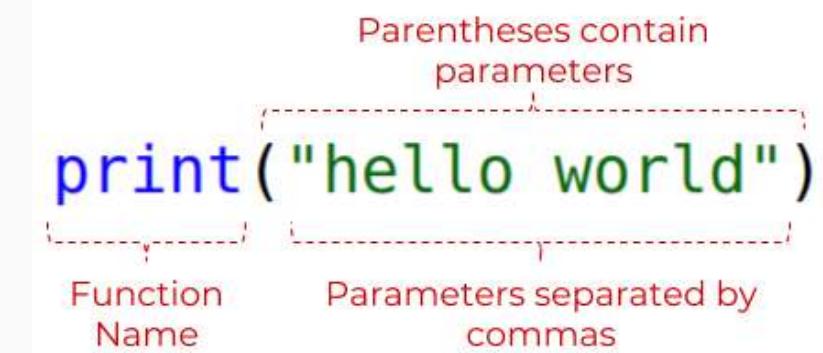
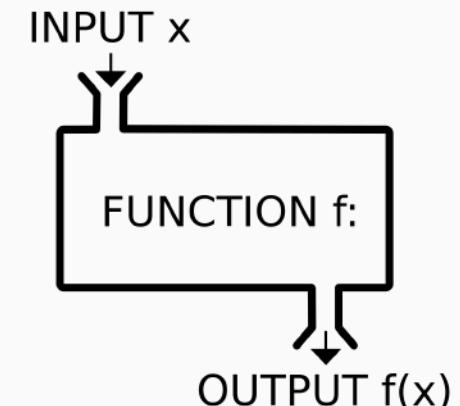
## Funções

**Códigos** que realizam **operações** em **argumentos**

Estrutura de uma função:

```
nome_da_funcao(argumento1, argumento2)
```

1. **Nome da função**: remete ao que ela faz
2. **Parênteses**: limitam a função
3. **Argumentos**: onde a função atuará
4. **Vírgulas**: separam os argumentos



# 3.7 Funções

## Exemplos

```
# soma  
sum(10, 2)
```

```
## [1] 12
```

```
# soma de objetos  
sum(obj10, obj2)
```

```
## [1] 12
```

```
# soma de objetos atribuidos a objetos  
obj_sum ← sum(obj10, obj2)
```

Nesse momento vocês devem estar se perguntando:  
como raios vou saber o nome das funções?!



# Uma maracutaia para ajudar!

A screenshot of a Google search results page. The search query "r non-numeric argument to mathematical function" is entered into the search bar, with a red arrow pointing to the search term. Below the search bar, the "All" tab is selected, along with other options like Videos, Images, Shopping, News, More, Settings, and Tools. The search results indicate approximately 465,000 results found in 0.42 seconds. The top result is a link to a Stack Overflow question titled "r - Non-numeric argument to mathematical function - Stack Overflow". The link URL is stackoverflow.com/questions/.../non-numeric-argument-to-mathematical-function. The snippet from the answer describes a common R programming error where a non-numeric data frame is passed to a function expecting a numeric vector. It includes a code example: > is.numeric(iris[,-5]) [1] FALSE > str(iris[,-5]) 'data.frame': 150 obs.

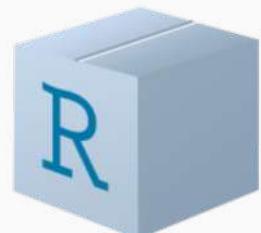
E de onde vêm as funções?!

## 3.8 Pacotes

### Coleção de funções para executar **tarefas específicas**

Duas fontes:

- **CRAN** (*finalizados*)
- **GitHub** (*em desenvolvimento*)



### Verificar pacotes carregados

```
# verificar pacotes carregados  
search()
```

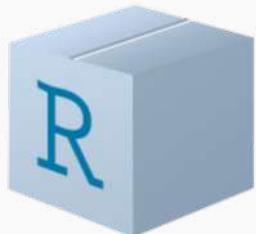
```
## [1] ".GlobalEnv"          "package:vegan"        "package:lattice"      "package:permute"     "package:countdown"   "package:forca"  
## [8] "package:dplyr"         "package:purrr"        "package:readr"        "package:tidyverse"    "package:tibble"      "package:ggplot2"  
## [15] "package:xaringan"      "package:pagedown"     "tools:rstudio"       "package:stats"       "package:graphics"    "package:grDevices"  
## [22] "package:datasets"      "package:methods"      "Autoloads"           "package:base"
```

## 3.8 Pacotes

**Coleção de funções para executar tarefas específicas**

Duas fontes:

- **CRAN** (*finalizados*)
- **GitHub** (*em desenvolvimento*)

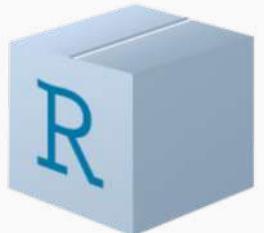


**Verificar pacotes instalados**

```
# verificar pacotes instalados  
library()
```

## 3.8 Pacotes

Ex.: pacote vegan



Fontes:

Pacotes do CRAN

<https://cran.r-project.org/web/packages/vegan/index.html>

Pacotes do GitHub

<https://github.com/vegandevs/vegan>

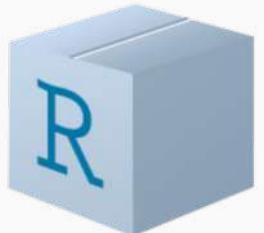
## 3.8 Pacotes

### Instalar pacotes

1. Instala-se apenas **uma vez**
2. **Precisa** estar conectado à **internet**
3. O **nome do pacote precisa** estar entre **aspas**
4. Função (CRAN):

```
install.packages()
```

```
# instalar pacotes  
install.packages("vegan")
```



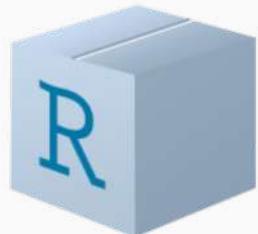
## 3.8 Pacotes

### Carregar pacotes

1. Carrega-se **toda vez** que se abre **uma nova sessão do R**
2. **Não precisa** estar conectado à **internet**
3. O **nome do pacote não precisa** estar entre **aspas**
4. Funções:

`library()` ou `require()`

```
# carregar pacotes
library(vegan)
```



## 3.8 Pacotes

Exemplos:

**vegan** – análises de comunidades

**sf** – manipulação de vetores

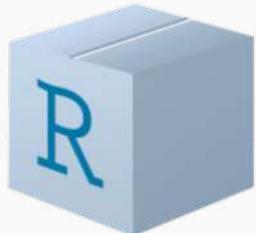
**raster** – manipulação de rasters

**ggplot2** – gráficos

**tmaps** – mapas temáticos

**bblme** – seleção de modelos (AIC)

**tidyverse** – data science



## 3.9 Ajuda (Help)

### Descreve as informações de uma função

```
## ajuda  
# descreve as informacoes de uma funcao  
help("mean") # arquivo .html  
  
?mean
```

- **Description:** resumo da função
- **Usage:** como utilizar a função e quais os seus argumentos
- **Arguments:** detalha os argumentos e como os mesmos devem ser especificados
- **Details:** detalhes importantes para se usar a função
- **Value:** mostra como interpretar a saída (*output*) da função (os resultados)
- **Note:** notas gerais sobre a função
- **Authors:** autores da função
- **References:** referências bibliográficas para os métodos usados para construção da função
- **See also:** funções relacionadas
- **Examples:** exemplos do uso da função. Às vezes pode ser útil copiar esse trecho e colar no R para ver como funciona e como usar a função

# 3.9 Ajuda (Help)

mean {base} R Documentation

## Arithmetic Mean

### Description

Generic function for the (trimmed) arithmetic mean.

### Usage

```
mean(x, ...)

## Default S3 method:
mean(x, trim = 0, na.rm = FALSE, ...)
```

### Arguments

- x An R object. Currently there are methods for numeric/logical vectors and [date](#), [date-time](#) and [time interval](#) objects. Complex vectors are allowed for trim = 0, only.
- trim the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.
- na.rm a logical value indicating whether NA values should be stripped before the computation proceeds.
- ... further arguments passed to or from other methods.

### Value

If trim is zero (the default), the arithmetic mean of the values in x is computed, as a numeric or complex vector of length one. If x is not logical (coerced to numeric), numeric (including integer) or complex, `NA_real_` is returned, with a warning.

If trim is non-zero, a symmetrically trimmed mean is computed with a fraction of trim observations deleted from each end before the mean is computed.

### References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

### See Also

[weighted.mean](#), [mean.POSIXct](#), [colMeans](#) for row and column means.

### Examples

```
x <- c(0:10, 50)
xm <- mean(x)
c(xm, mean(x, trim = 0.10))
```

### 3.10 Diretório de trabalho

O diretório de trabalho é o endereço da pasta (diretório) de onde o R importará ou para onde exportará nossos dados

Podemos utilizar o próprio RStudio para tal tarefa, indo em:

| Session > Set Work Directory > Choose Directory...

**Atalho:** Ctrl + Shift + H



## 3.10 Diretório de trabalho

O diretório de trabalho é o endereço da pasta (diretório) de onde o R importará ou para onde exportará nossos dados

Podemos utilizar funções no script

```
# definir o diretorio de trabalho
setwd("/home/mude/data/github/workshop-r-sdm")

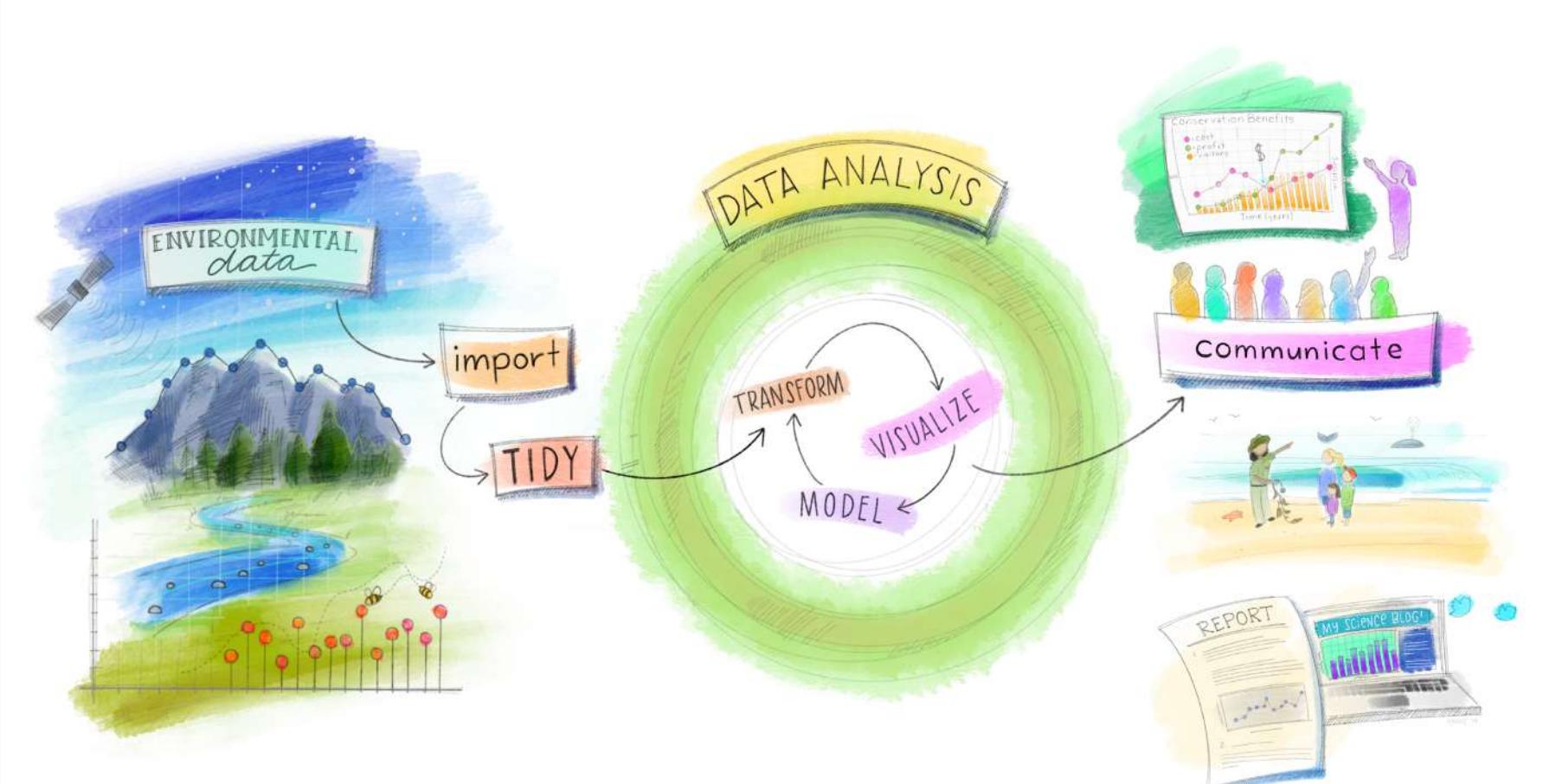
# verificar o diretorio
getwd()

# listar os arquivos no diretorio
dir()
```

*tidyverse*

## 3.11 tidyverse

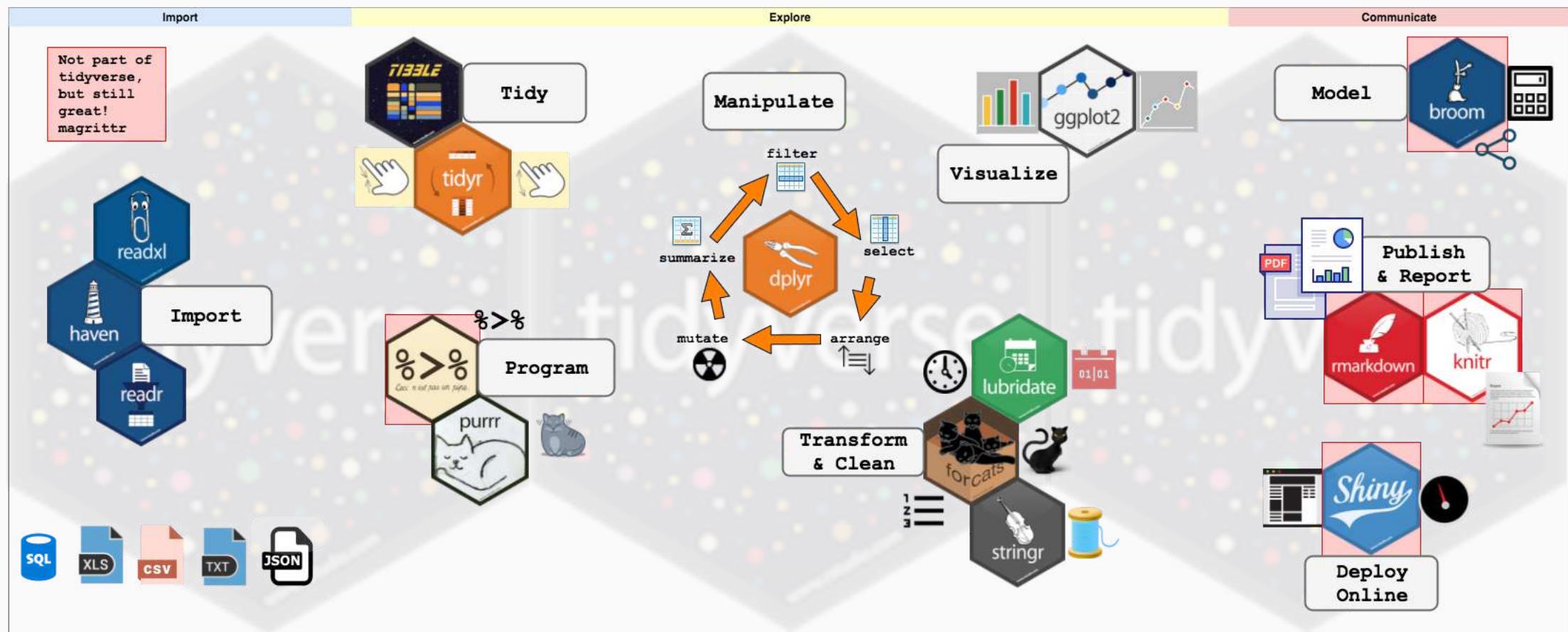
O tidyverse é um **fluxo de trabalho** designado para **Data Science**



Fonte: [@allison\\_horst](#)

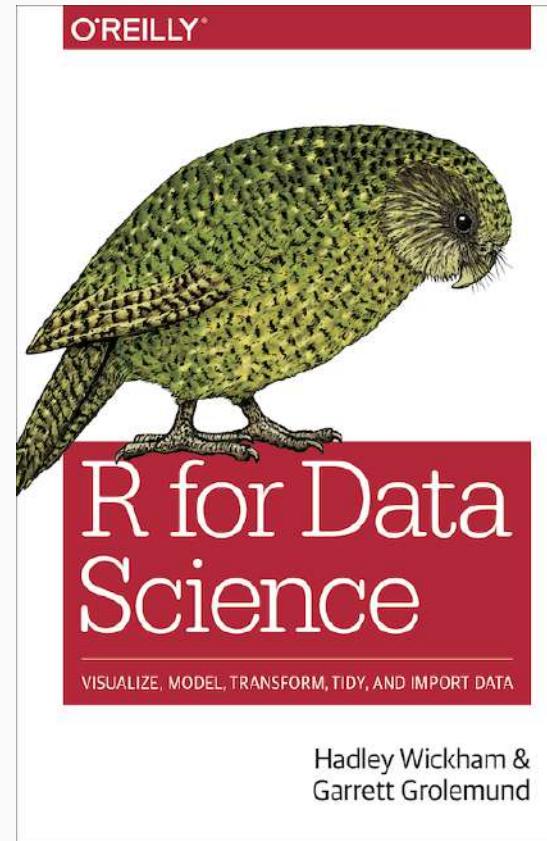
# 3.11 tidyverse

## Pacotes e funcionalidades



# 3.11 tidyverse

R for Data Science (2017)



[Wickham & Grolemund \(2016\).](#)

## 3.11 tidyverse

magrittr (pipe - %>%)

René Magritte (1898-1967)



## 3.11 tidyverse

magrittr (pipe - %>%)

**Atalho:** Ctrl + Shift + M

```
# sem pipe  
sqrt(sum(1:100))
```

```
## [1] 71.06335
```

### Composite Functions

$$(f \circ g)(x) = ?$$

$$(g \circ f)(x) = ?$$

# 3.11 tidyverse

## magrittr (pipe - %>%)

**Atalho:** Ctrl + Shift + M

```
# sem pipe  
sqrt(sum(1:100))
```

```
## [1] 71.06335
```



# 3.11 tidyverse

## magrittr (pipe - %>%)

**Atalho:** Ctrl + Shift + M

```
# com pipe  
1:100 %>%  
  sum() %>%  
  sqrt()
```

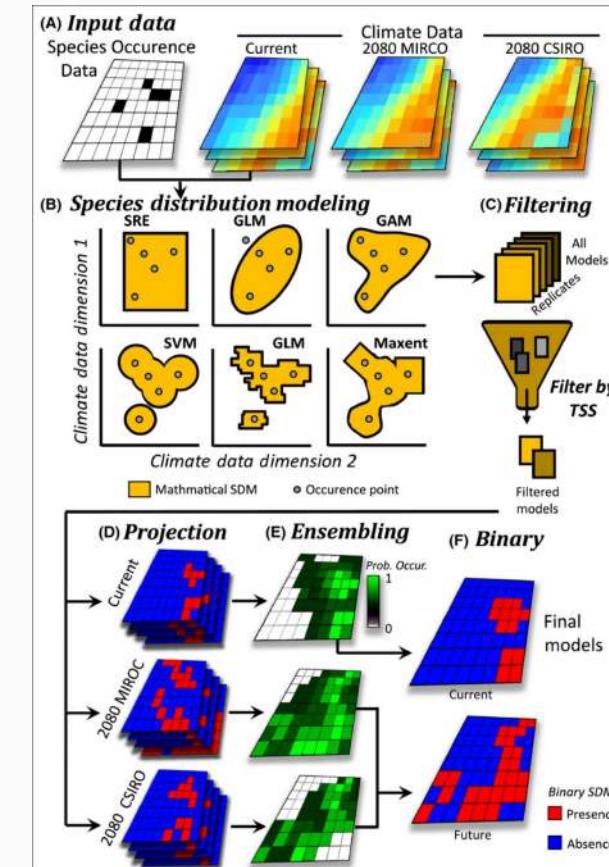
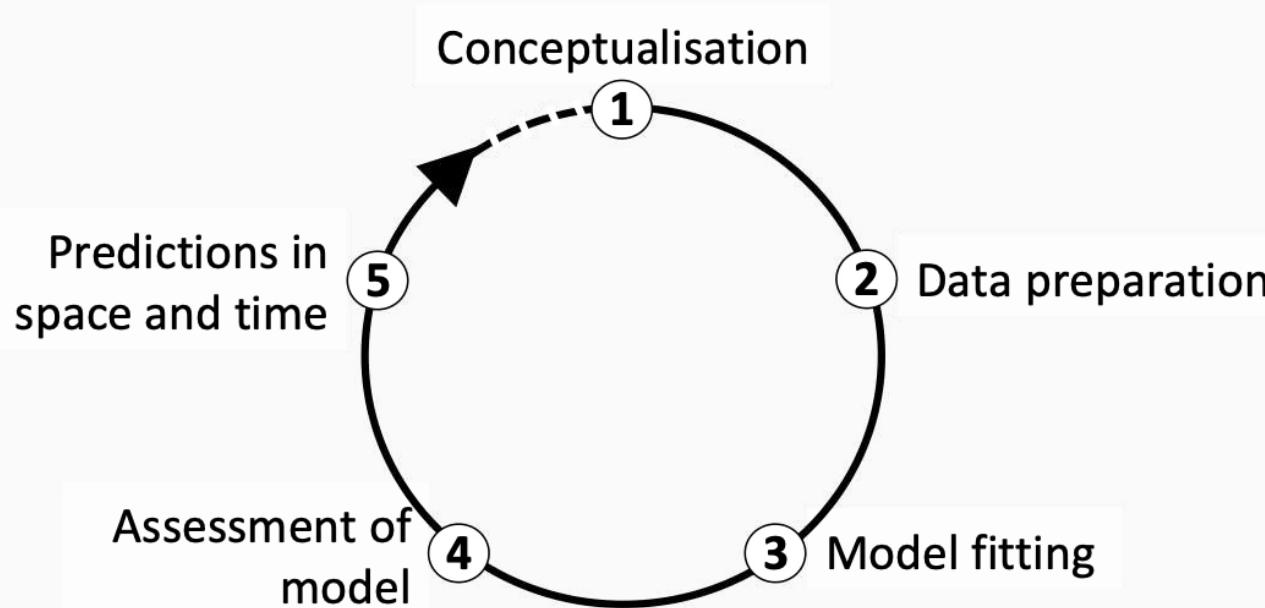
```
## [1] 71.06335
```



# Dúvidas?

## 2.4 Padronização para criação dos modelos

Protocolo padrão para modelos de distribuição de espécies

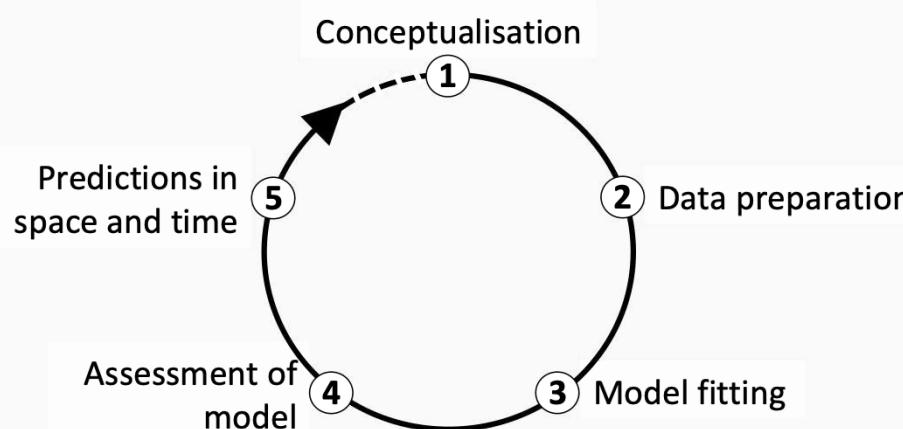


[Brown & Yoder \(2015\)](#), [Zurell et al. \(2020\)](#)

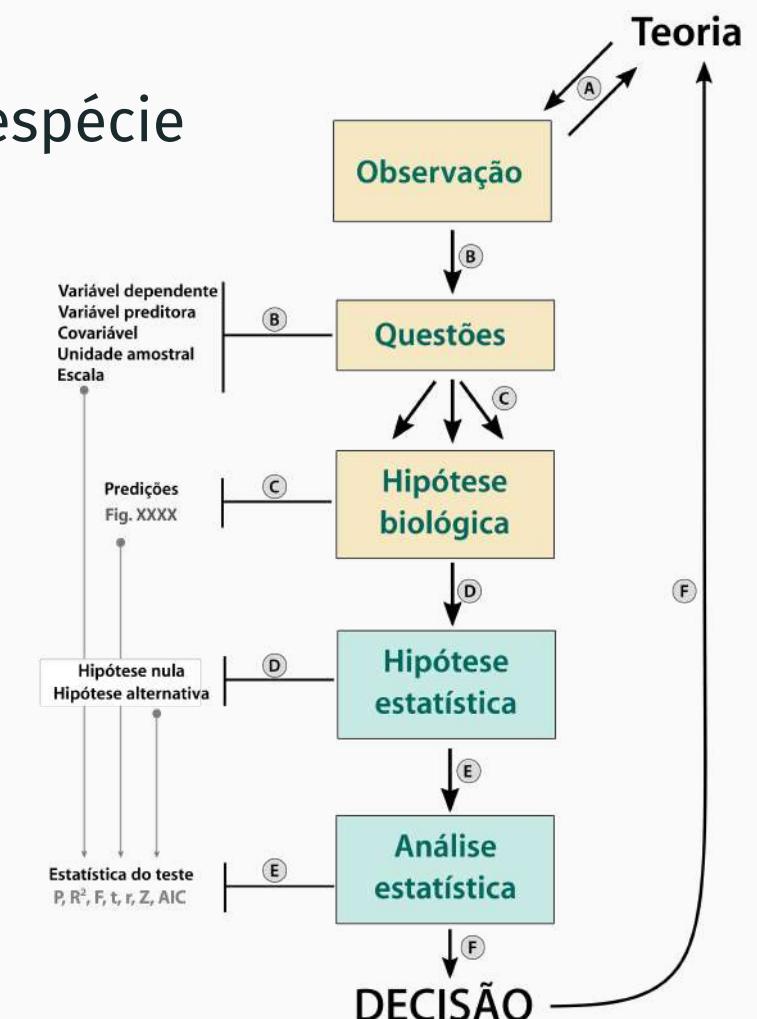
# 2.4 Padronização para criação dos modelos

## 1. Conceitualização

**Objetivo:** Modelar a distribuição potencial de uma espécie



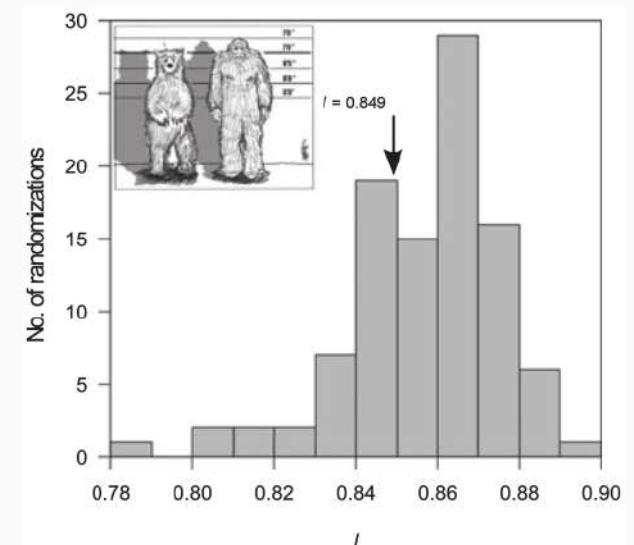
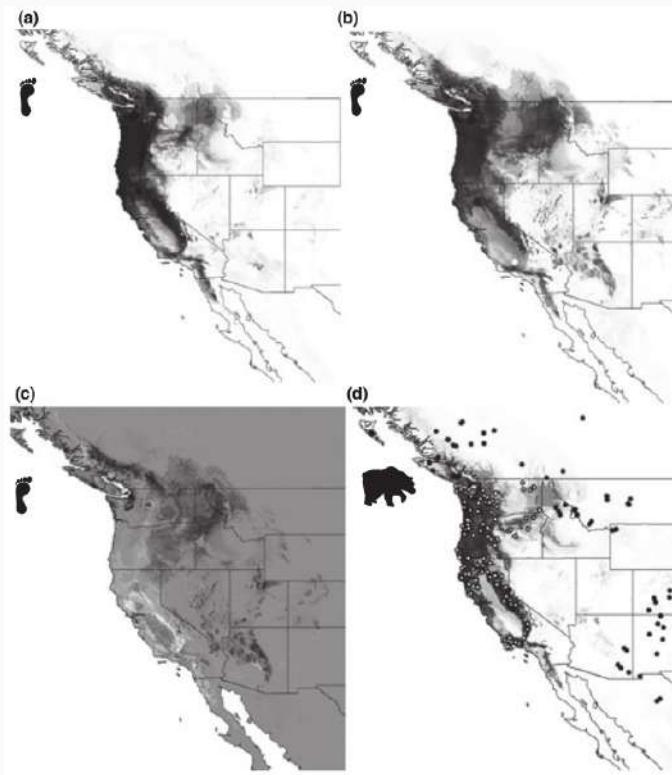
Adaptado de Gonçalves-Souza et al. (2019), Zurell et al. (2020)



# 2.4 Padronização para criação dos modelos

## 1. Conceitualização

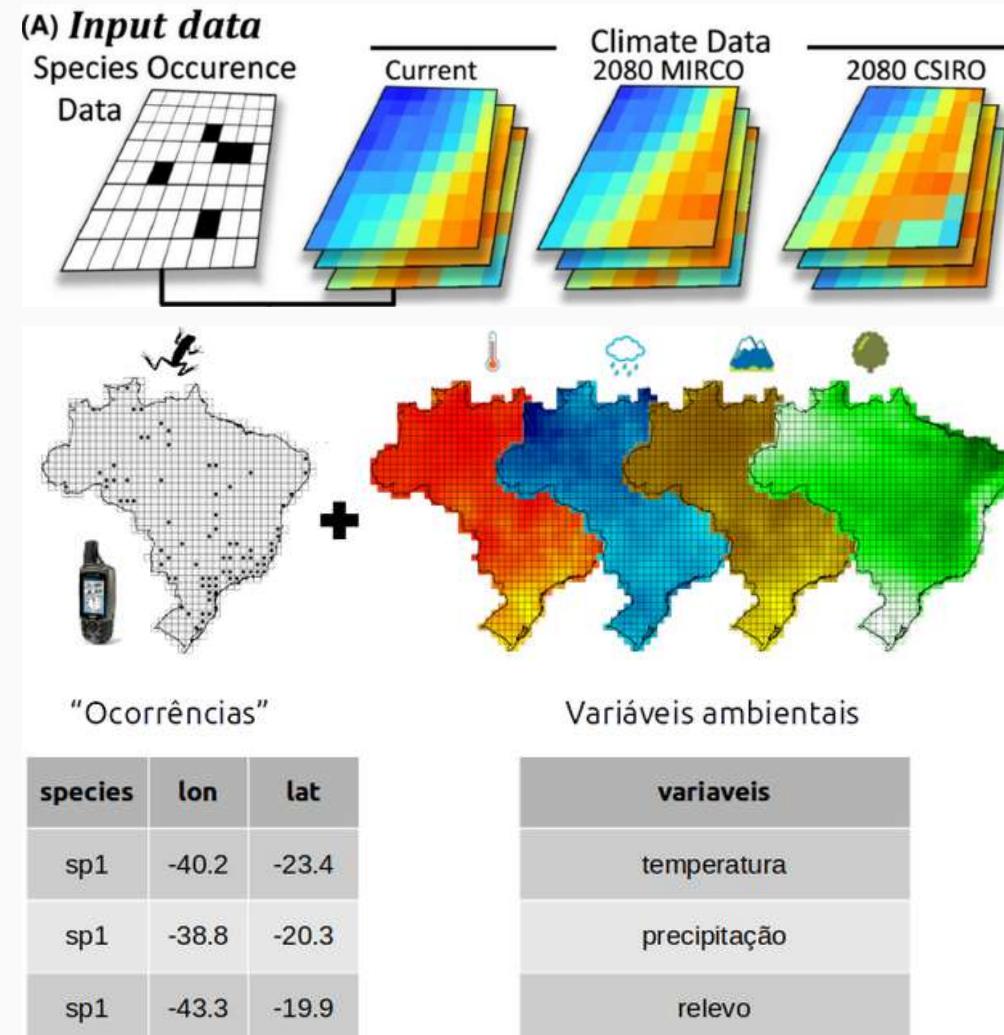
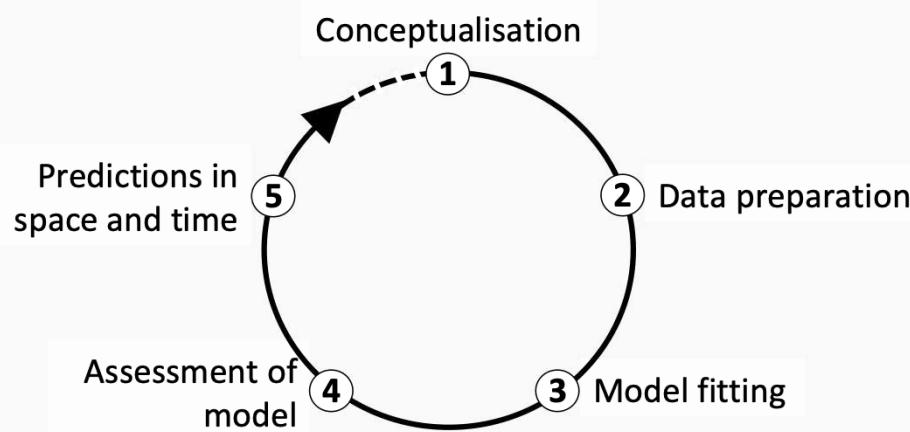
**Objetivo:** Modelar a distribuição potencial de uma espécie



- Sasquatch (pé-grande), *Ursus americanus* (urso-negro), [Lozier et al. \(2009\)](#), [Sykes et al. \(2014\)](#).

# 2.4 Padronização para criação dos modelos

## 2. Preparação dos dados

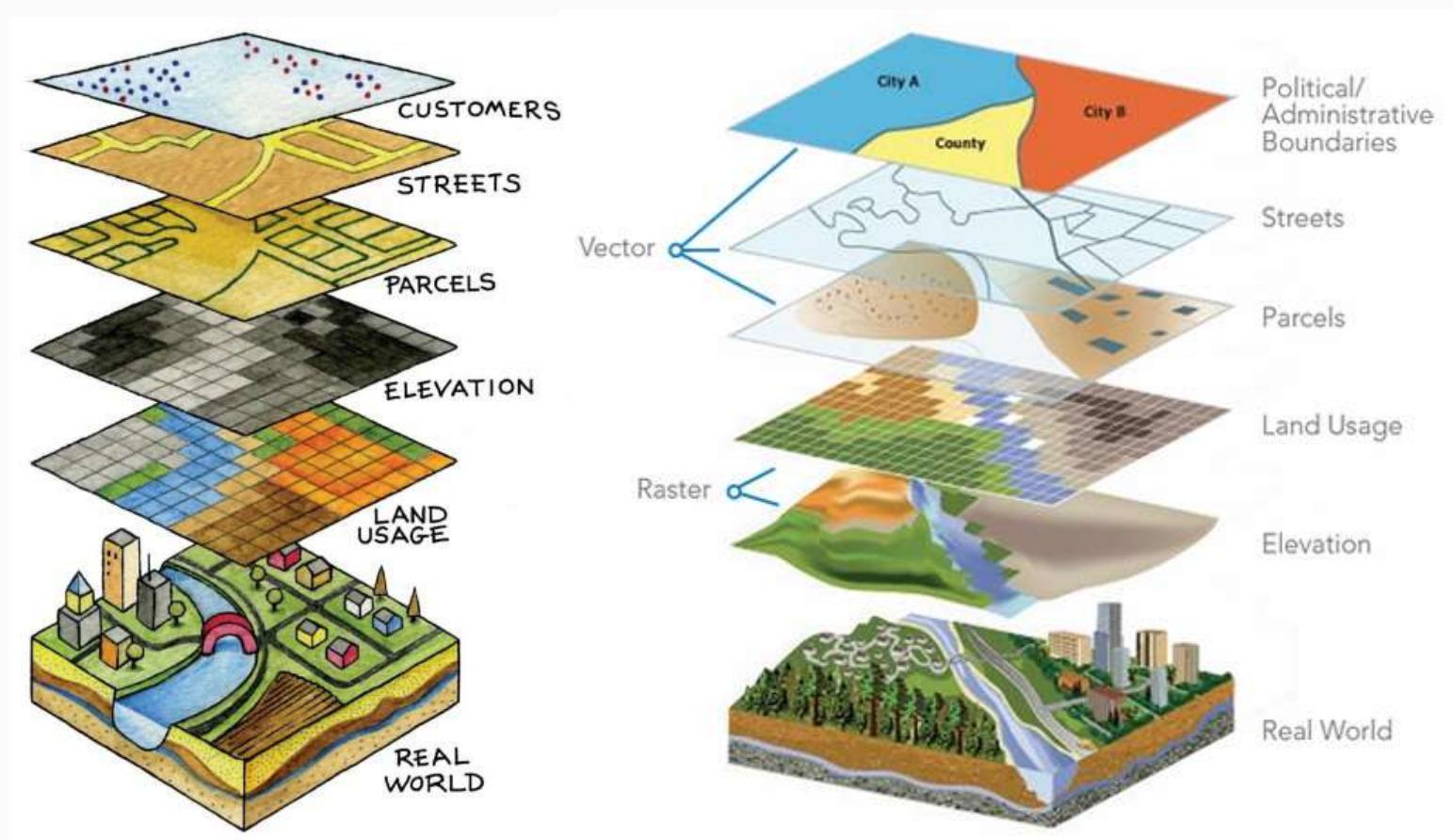


[Brown & Yoder \(2015\)](#)

# 4 Preparação dos dados (ocorrências e variáveis)

# 4.1 Dados geoespaciais

## Vetor e raster



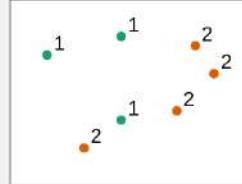
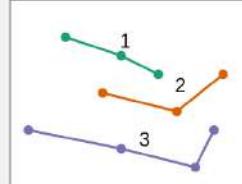
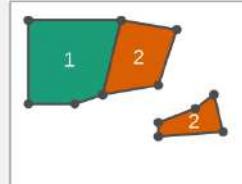
[Campbell & Shin \(2012\)](#),  
[ESRI \(2019\)](#)

# 4.1 Dados geoespaciais

## Vetores

### Pontos, linhas e polígonos + Tabela de atributos

Geometrias	Entidade espacial	Representação	Atributos																																																												
Pontos	 	 	<table border="1"><thead><tr><th>FID</th><th>Município</th><th>Hidrografia</th><th>Vazão</th></tr></thead><tbody><tr><td>1</td><td>Rio Claro</td><td>Nascente</td><td>0,2</td></tr><tr><td>2</td><td>Rio Claro</td><td>Nascente</td><td>0,8</td></tr><tr><td>3</td><td>Rio Claro</td><td>Nascente</td><td>1,1</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td></tr></tbody></table> <table border="1"><thead><tr><th>FID</th><th>Município</th><th>Hidrografia</th><th>Vazão</th></tr></thead><tbody><tr><td>1</td><td>Rio Claro</td><td>Rios</td><td>2,4</td></tr><tr><td>2</td><td>Rio Claro</td><td>Rios</td><td>3,1</td></tr><tr><td>3</td><td>Rio Claro</td><td>Rios</td><td>7,7</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td></tr></tbody></table> <table border="1"><thead><tr><th>FID</th><th>Município</th><th>Uso</th><th>Área</th></tr></thead><tbody><tr><td>1</td><td>Rio Claro</td><td>Floresta</td><td>10,1</td></tr><tr><td>2</td><td>Rio Claro</td><td>Floresta</td><td>19,8</td></tr><tr><td>3</td><td>Rio Claro</td><td>Floresta</td><td>50,2</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td></tr></tbody></table>	FID	Município	Hidrografia	Vazão	1	Rio Claro	Nascente	0,2	2	Rio Claro	Nascente	0,8	3	Rio Claro	Nascente	1,1	...	...	...	...	FID	Município	Hidrografia	Vazão	1	Rio Claro	Rios	2,4	2	Rio Claro	Rios	3,1	3	Rio Claro	Rios	7,7	...	...	...	...	FID	Município	Uso	Área	1	Rio Claro	Floresta	10,1	2	Rio Claro	Floresta	19,8	3	Rio Claro	Floresta	50,2	...	...	...	...
FID	Município	Hidrografia	Vazão																																																												
1	Rio Claro	Nascente	0,2																																																												
2	Rio Claro	Nascente	0,8																																																												
3	Rio Claro	Nascente	1,1																																																												
...	...	...	...																																																												
FID	Município	Hidrografia	Vazão																																																												
1	Rio Claro	Rios	2,4																																																												
2	Rio Claro	Rios	3,1																																																												
3	Rio Claro	Rios	7,7																																																												
...	...	...	...																																																												
FID	Município	Uso	Área																																																												
1	Rio Claro	Floresta	10,1																																																												
2	Rio Claro	Floresta	19,8																																																												
3	Rio Claro	Floresta	50,2																																																												
...	...	...	...																																																												
Linhas																																																															
Polígonos	 	 																																																													

	<p>Example attributes for point data</p> <table border="1"><thead><tr><th>ID</th><th>name</th><th>has</th><th>evergreen</th></tr></thead><tbody><tr><td>1</td><td>Broadleaf</td><td>Leaves</td><td>FALSE</td></tr><tr><td>2</td><td>Conifer</td><td>Needles</td><td>TRUE</td></tr></tbody></table>	ID	name	has	evergreen	1	Broadleaf	Leaves	FALSE	2	Conifer	Needles	TRUE				
ID	name	has	evergreen														
1	Broadleaf	Leaves	FALSE														
2	Conifer	Needles	TRUE														
	<p>Example attributes for line data</p> <table border="1"><thead><tr><th>ID</th><th>name</th><th>lanes</th><th>cycling</th></tr></thead><tbody><tr><td>1</td><td>Road A</td><td>4</td><td>FALSE</td></tr><tr><td>2</td><td>Road B</td><td>3</td><td>TRUE</td></tr><tr><td>3</td><td>Road C</td><td>2</td><td>TRUE</td></tr></tbody></table>	ID	name	lanes	cycling	1	Road A	4	FALSE	2	Road B	3	TRUE	3	Road C	2	TRUE
ID	name	lanes	cycling														
1	Road A	4	FALSE														
2	Road B	3	TRUE														
3	Road C	2	TRUE														
	<p>Example attributes for polygon data</p> <table border="1"><thead><tr><th>ID</th><th>name</th><th>population</th><th>touristic</th></tr></thead><tbody><tr><td>1</td><td>Country A</td><td>1000</td><td>FALSE</td></tr><tr><td>2</td><td>Country B</td><td>500</td><td>TRUE</td></tr></tbody></table>	ID	name	population	touristic	1	Country A	1000	FALSE	2	Country B	500	TRUE				
ID	name	population	touristic														
1	Country A	1000	FALSE														
2	Country B	500	TRUE														

# 4.1 Dados geoespaciais

## Formato de arquivos vetoriais

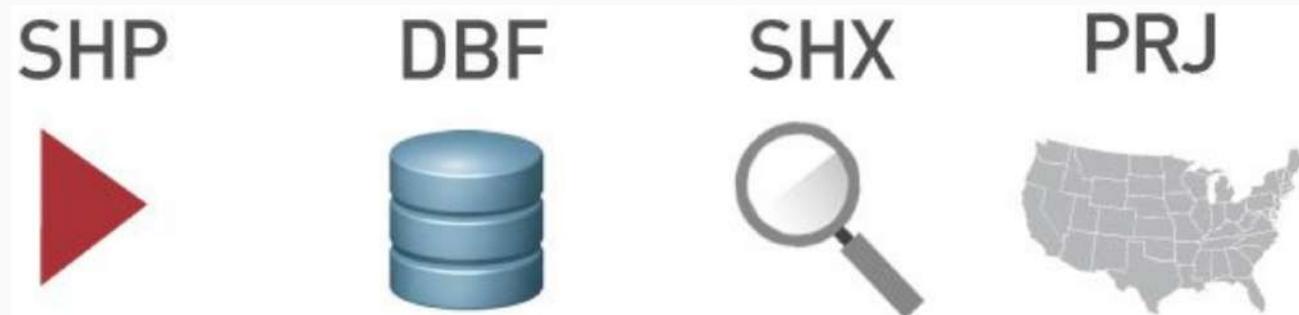
### Shapefile

**.shp:** contém as informações da feição (desenho)

**.dbf:** tabela que contém as informações (colunas) sobre cada feição (linhas)

**.shx:** arquivo de índice que une o .shp ao .dbf

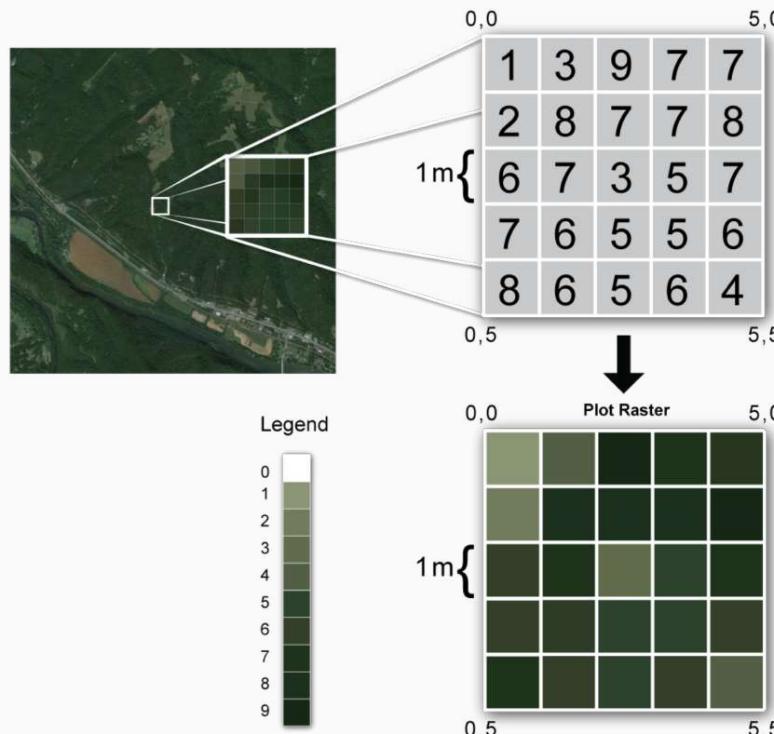
**.prj:** contém as informações do sistema de referências de coordenadas



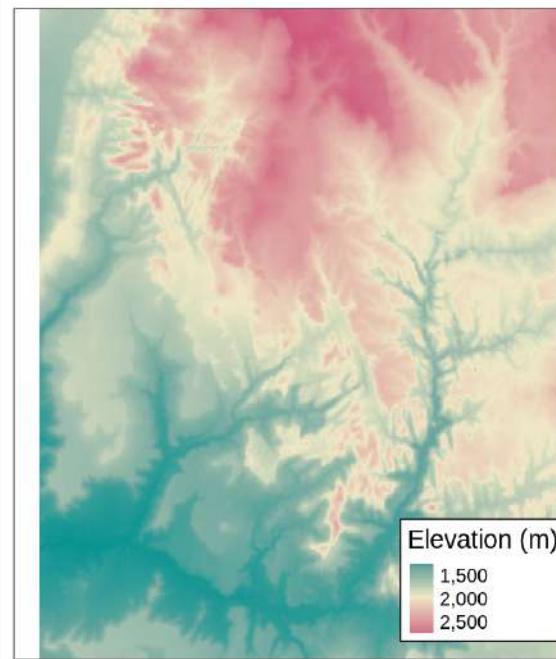
# 4.1 Dados geoespaciais

## Gride ou raster

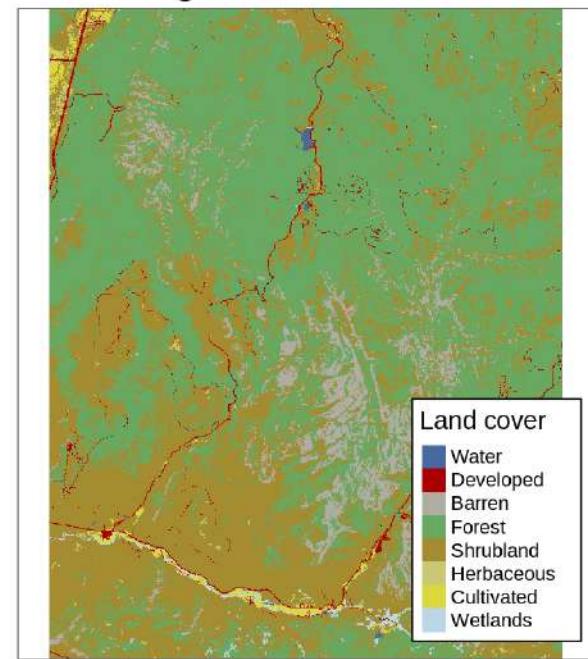
Matriz de valores - contínuos ou categóricos



A. Continuous data



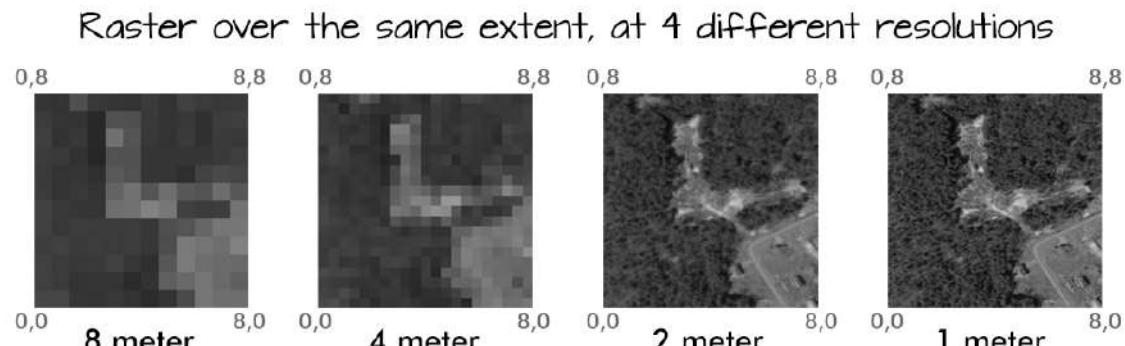
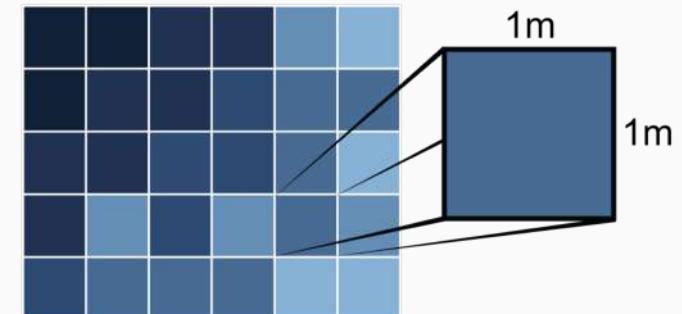
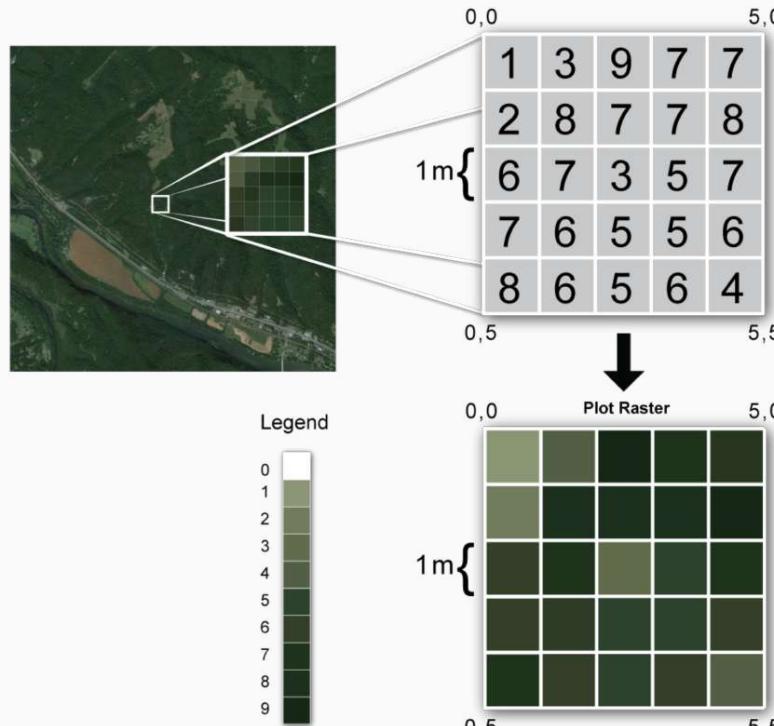
B. Categorical data



# 4.1 Dados geoespaciais

## Gride ou raster

### Propriedades - Extensão e resolução

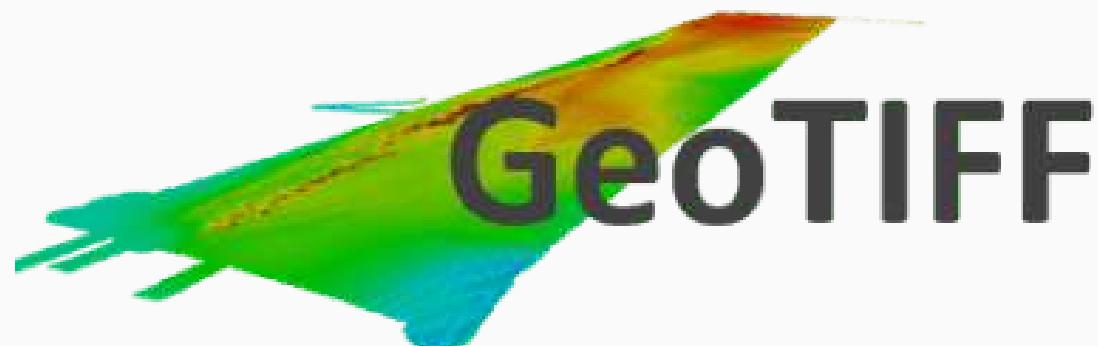


[National Ecological Observatory Network \(NEON\)](#)

# 4.1 Dados geoespaciais

Gride ou raster - Formato - GeoTiff

.tif: arquivo matricial georreferenciamento com informações de georreferenciamento



[LuminFire Developers](#), [GeoTIFF](#)

# 4.1 Dados geoespaciais

## Sistema de coordenadas

### Geográfico (graus)

#### 1. Graus, minutos e segundos

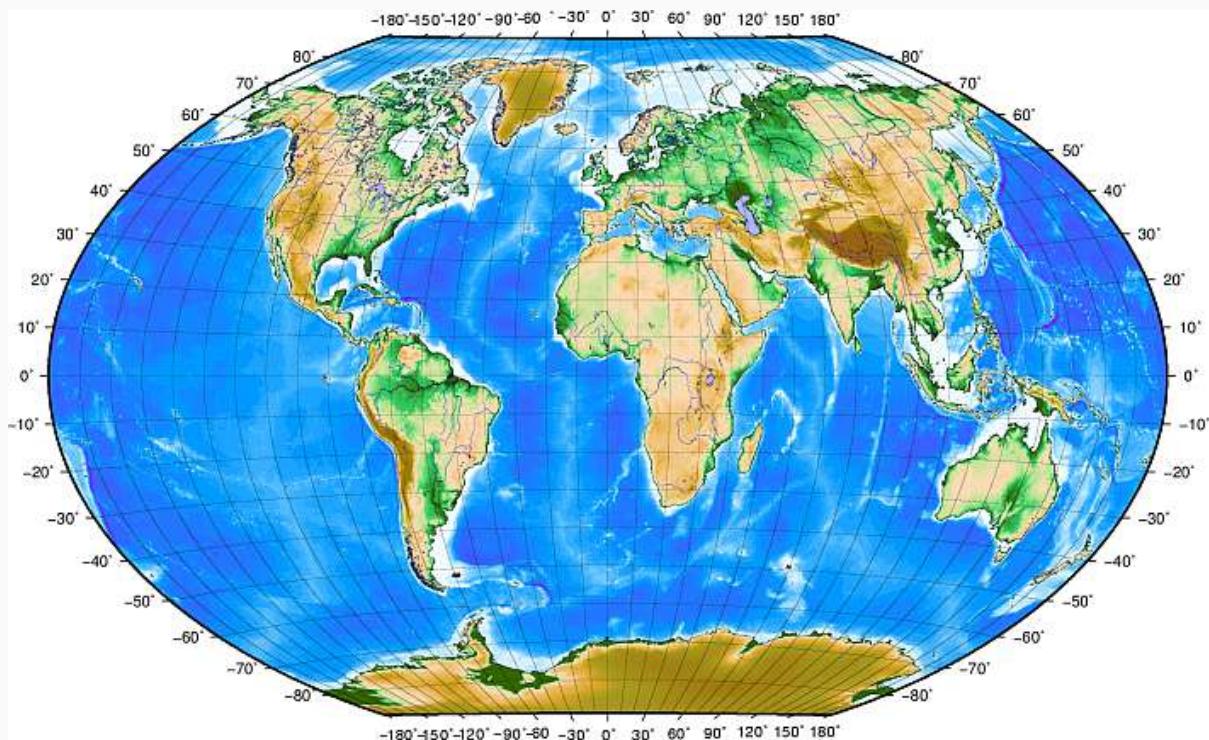
- Longitude:  $42^{\circ} 42'42''$  O
- Latitude:  $23^{\circ} 23'23''$  S

#### 2. Graus decimais

- Longitude: -42.71167
- Latitude: -23.38972

**Conversão:**  $42 + (42/60) + (42/3600) = 42.71167$

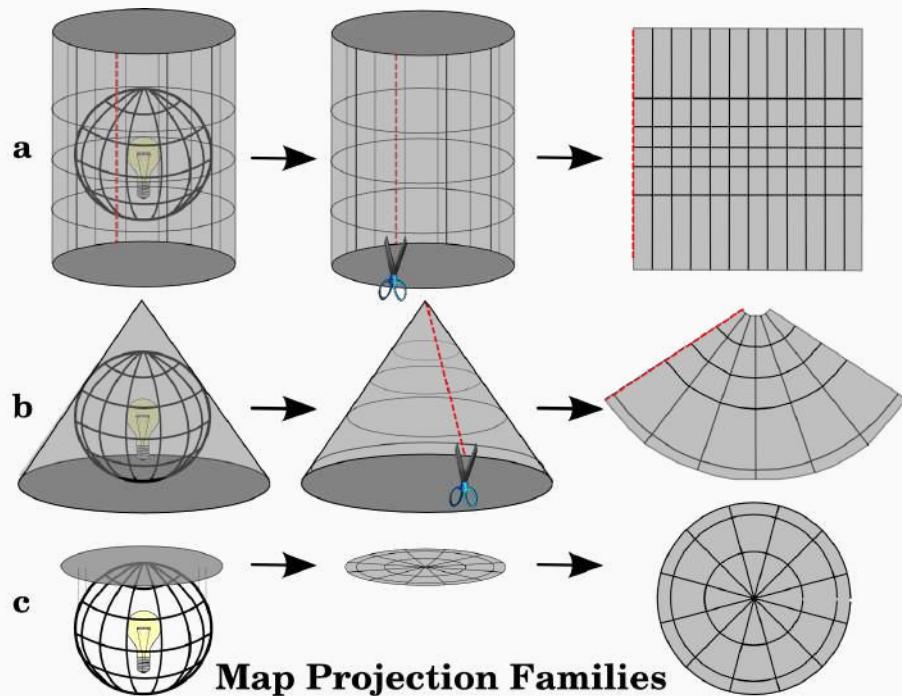
**Conversão:**  $23 + (23/60) + (23/3600) = 23.38972$



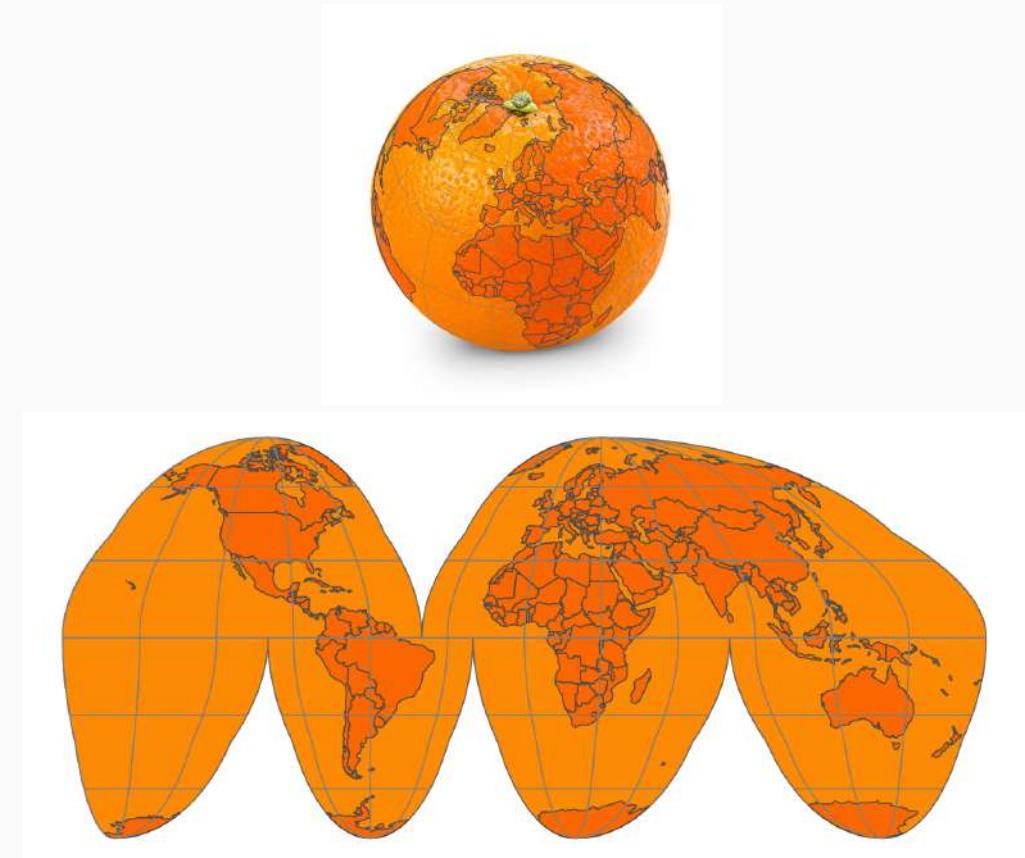
# 4.1 Dados geoespaciais

## Sistema de coordenadas

Projetado (metros)



[Guia do Usuário QGIS 3.16](#), Tennekes & Nowosad (in prep.).



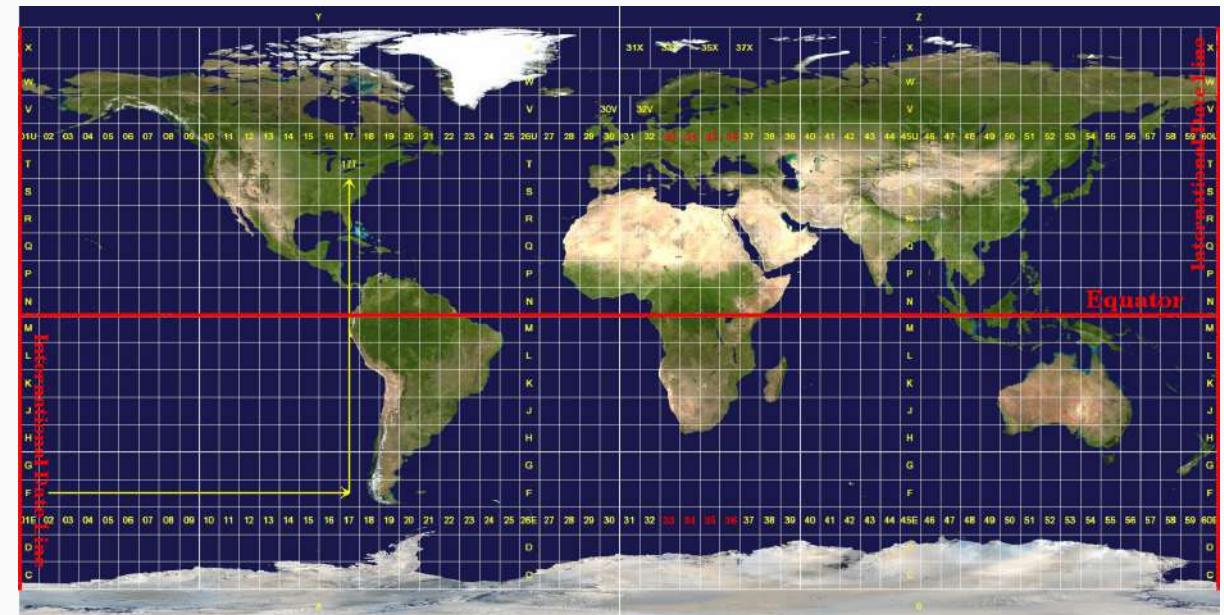
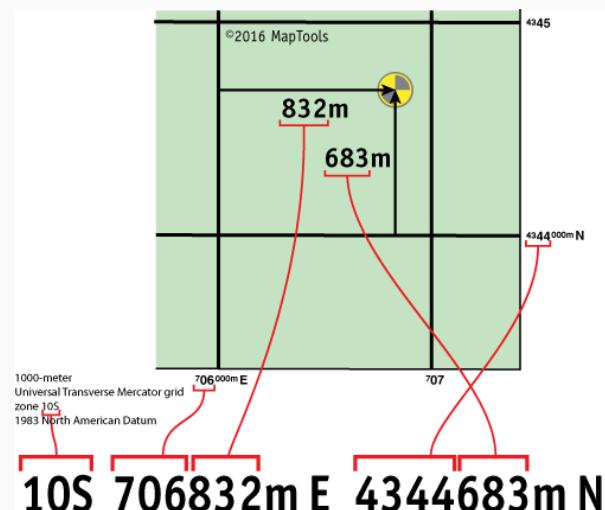
# 4.1 Dados geoespaciais

## Sistema de coordenadas

Projetado (metros) - Universal Transversa de Mercator (UTM)

### 1. Zona e metros

- X UTM: 706832 m E
- Y UTM: 4344683 m N
- Zona: 10S



# 4.1 Dados geoespaciais

## Datum

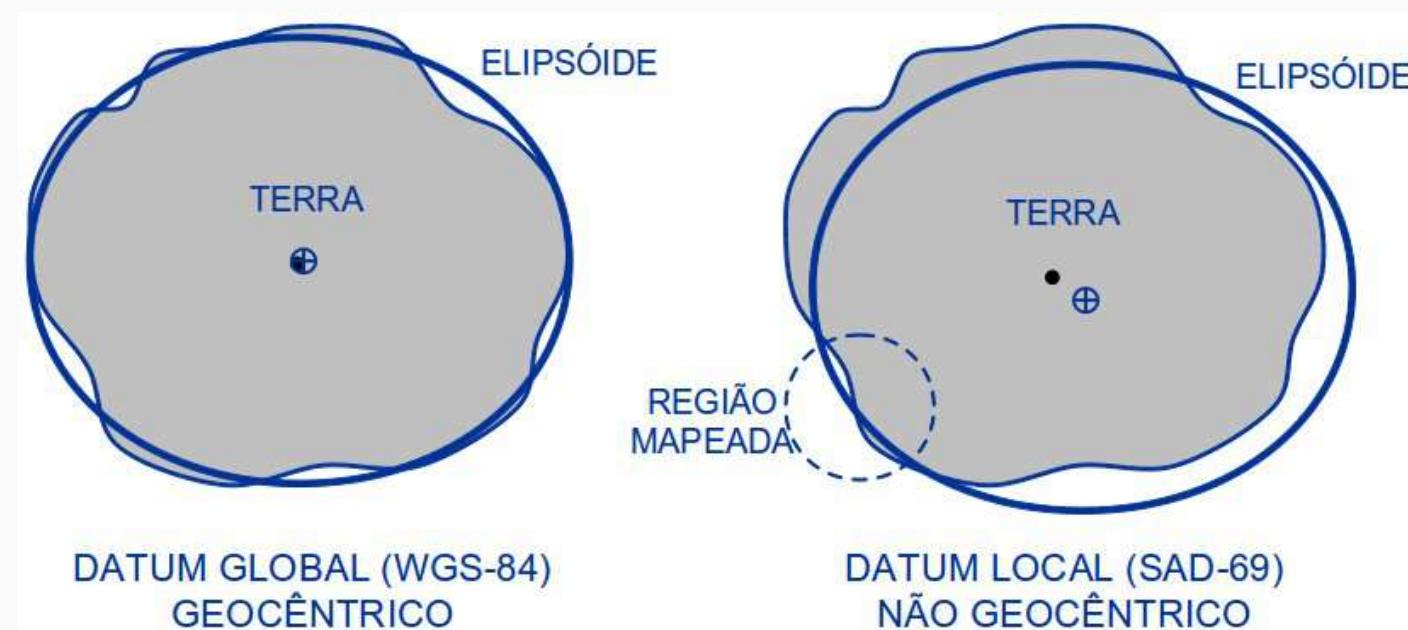
Relação do sistema de coordenadas com a superfície da Terra

### 1. Geocêntricos

- WGS 84
- SIRGAS 2000

### 2. Topocêntricos

- SAD 69
- Córrego Alegre



**Conversão:** não é trivial e requer muita trigonometria

# 4.1 Dados geoespaciais

## Datum

Relação do sistema de coordenadas com a superfície da Terra

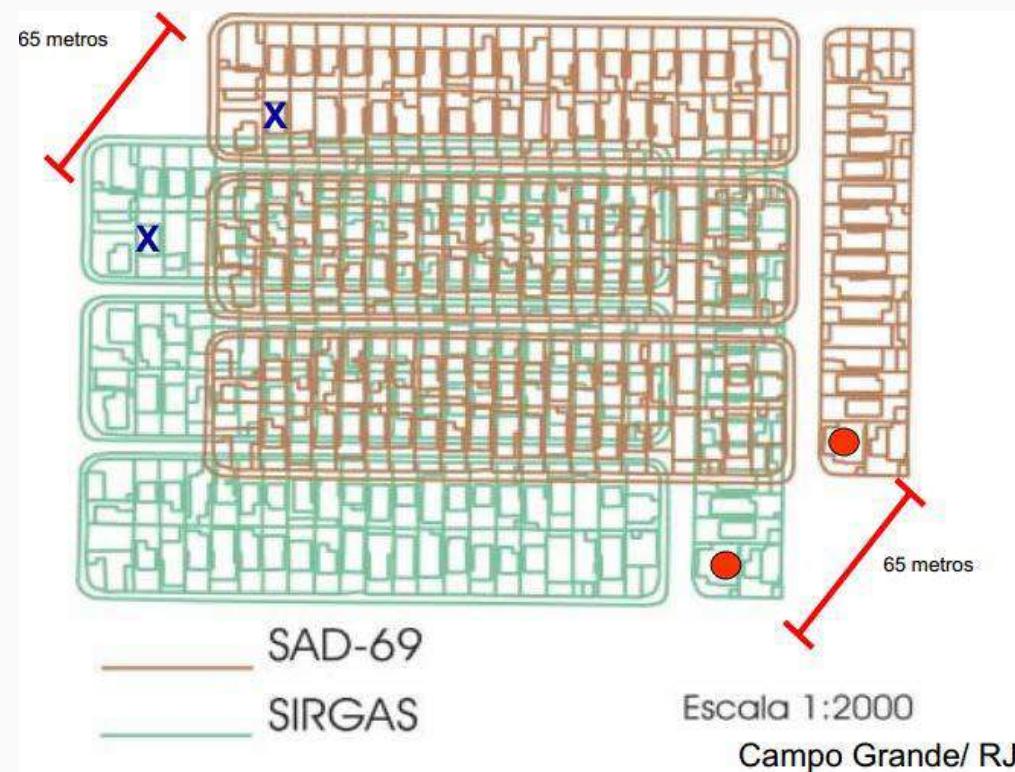
### 1. Geocêntricos

- SIRGAS 2000

### 2. Topocêntricos

- SAD 69

**Diferença:** cerca de 65 metros



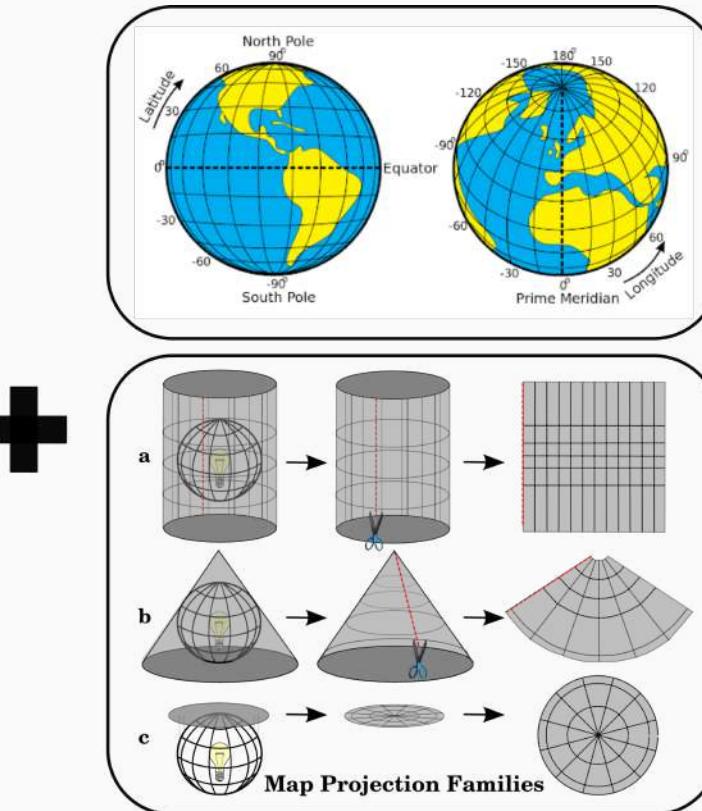
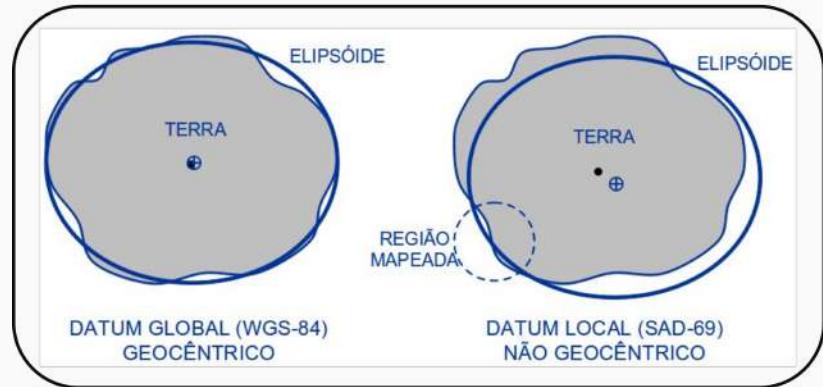
Recapitulando: o Sistema de Referência de Coordenadas (CRS) é composto por:

1. Sistema de coordenadas (geográfico ou projetado)
2. Datum (relação do sistema de coordenadas com a superfície da Terra)

# 4.1 Dados geoespaciais

Sistema de Referência de Coordenadas (*Coordinate Reference System*)

Datum + Sistema de Coordenadas (geográfico ou projetado)



## 4.1 Dados geoespaciais

EPSG (*European Petroleum Survey Group*)

Possui códigos numéricos para quase todos os CRSs



# 4.1 Dados geoespaciais

EPSG (*European Petroleum Survey Group*)

EPSG: 4326 [WGS84 Geográfico](#)

EPSG: 4674 [SIRGAS2000 Geográfico](#)

EPSG: 32723 [WGS84 UTM23S](#)

EPSG: 31983 [SIRGAS2000 UTM23S](#)



Coordinate Systems Worldwide

# 4.1 Dados geoespaciais

## Principais fontes de dados vetoriais

- [\*\*IBGE\*\*](#): limites territoriais e censitários do Brasil
- [\*\*FBDS\*\*](#): uso da terra, APP e hidrografia - Mata Atlântica e Cerrado
- [\*\*GeoBank\*\*](#): dados geológicos do Brasil
- [\*\*Pastagem.org\*\*](#): dados de pastagens e gado para o Brasil
- [\*\*CanaSat\*\*](#): dados de cana-de-açúcar para o Brasil
- [\*\*CSR Maps\*\*](#): diversos dados vetoriais para o Brasil
- [\*\*Ecoregions\*\*](#): dados de biorregiões e biomas do mundo
- [\*\*GADM\*\*](#): limites de áreas administrativas do mundo
- [\*\*Natural Earth\*\*](#): diversos limites para o mundo
- [\*\*Protected Planet\*\*](#): áreas protegidas para o mundo
- [\*\*UN Biodiversity Lab\*\*](#): Diversas bases de dados para o mundo |
- [\*\*HydroSHEDS\*\*](#): informações hidrológicas do mundo
- [\*\*Global Roads Inventory Project \(GRIP\)\*\*](#): dados de estradas do mundo todo

# 4.1 Dados geoespaciais

## Principais fontes de dados matriciais (raster)

- [\*\*MapBiomas\*\*](#): uso e cobertura da terra para o Brasil, diversos anos
- [\*\*Bahlu\*\*](#): Distribuições históricas de terras agrícolas e pastagens para todo o Brasil de 1940 a 2012
- [\*\*USGS\*\*](#): dados de diversos satélites livres para o mundo
- [\*\*SRTM\*\*](#): dados de elevação para o mundo
- [\*\*Global Forest Watch\*\*](#): dados de uso e cobertura para todo o planeta
- [\*\*Geoservice Maps\*\*](#): dados de elevação e florestas para o mundo
- [\*\*GlobCover\*\*](#): dados de uso e cobertura da terra para todo o planeta
- [\*\*Global Human Footprint\*\*](#): dados de pegada ecológica para o mundo
- [\*\*Land-Use Harmonization \(LUH2\)\*\*](#): dados atuais e previsões de uso da terra
- [\*\*SoilGrids\*\*](#): dados de solo para o mundo
- [\*\*WorldClim\*\*](#): dados climáticos para o mundo
- [\*\*CHELSA\*\*](#): dados climáticos para o mundo
- [\*\*EarthEnv\*\*](#): dados de cobertura da terra, nuvens, relevo e hidrografia
- [\*\*MARSPEC\*\*](#): dados de condições do oceano para o mundo
- [\*\*Bio-ORACLE\*\*](#): dados de condições do oceano para o mundo

# 4.1 Dados geoespaciais

## Principais fontes de dados em pacotes no R

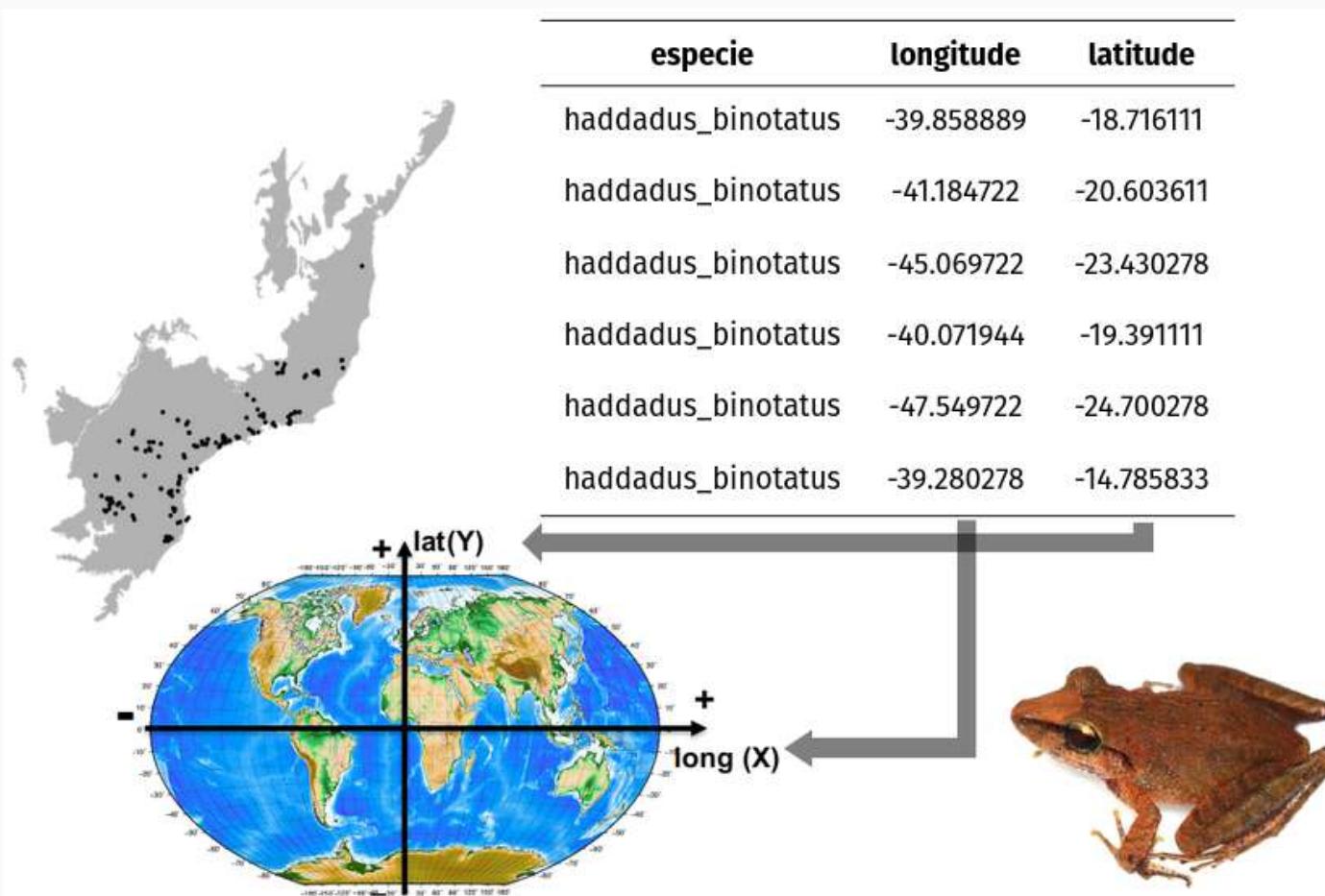
- [\*\*geobr\*\*](#): carrega Shapefiles de Conjuntos de Dados Espaciais Oficiais do Brasil
- [\*\*rnatulearth\*\*](#): dados do mapa mundial da Natural Earth
- [\*\*rworldmap\*\*](#): mapeando Dados Globais
- [\*\*spData\*\*](#): conjuntos de dados para análise espacial
- [\*\*OpenStreetMap\*\*](#): acesso para abrir imagens raster de mapas de ruas
- [\*\*osmdata\*\*](#): baixe e importe dados do OpenStreetMap
- [\*\*geonames\*\*](#): interface para o serviço da Web de consulta espacial "Geonames"
- [\*\*rgbif\*\*](#): interface para o Global 'Biodiversity' Information Facility API
- [\*\*maptools\*\*](#): ferramentas para lidar com objetos espaciais
- [\*\*marmap\*\*](#): importar, traçar e analisar dados batimétricos e topográficos
- [\*\*envirem\*\*](#): geração de Variáveis ENVIREM
- [\*\*sdpredictors\*\*](#): conjuntos de dados preditor de modelagem de distribuição de espécies

Dúvidas?

## 4.2 Dados de ocorrências

## 4.2 Dados de ocorrências

### Formato



## 4.2 Dados de ocorrências

### Fontes

- Coletas em campo (amostragens em campo)
- Literatura (artigos, data papers, ...)
- Naturalistas e ciência cidadã (e-Bird, iNaturalist, ...)
- Coleções científicas e museus (Museu Nacional, MZUSP, CFHB, ...)
- Banco de dados (GBIF, SpeciesLink, ...)



**Incêndio no Museu Nacional (RJ) 02/09/2018.** Foto: [Ricardo Moraes/Reuters](#)

## 4.2 Dados de ocorrências

### Banco de dados

- [Global Biodiversity Information Facility \(GBIF\)](#)
- [iNaturalist](#)
- [VertNet](#)
- [eBird](#)
- [iDigBio](#)
- [Ocean Biogeographic Information System \(OBIS\)](#)
- [Botanical Information and Ecology Network \(BIEN\)](#)
- [speciesLink](#)



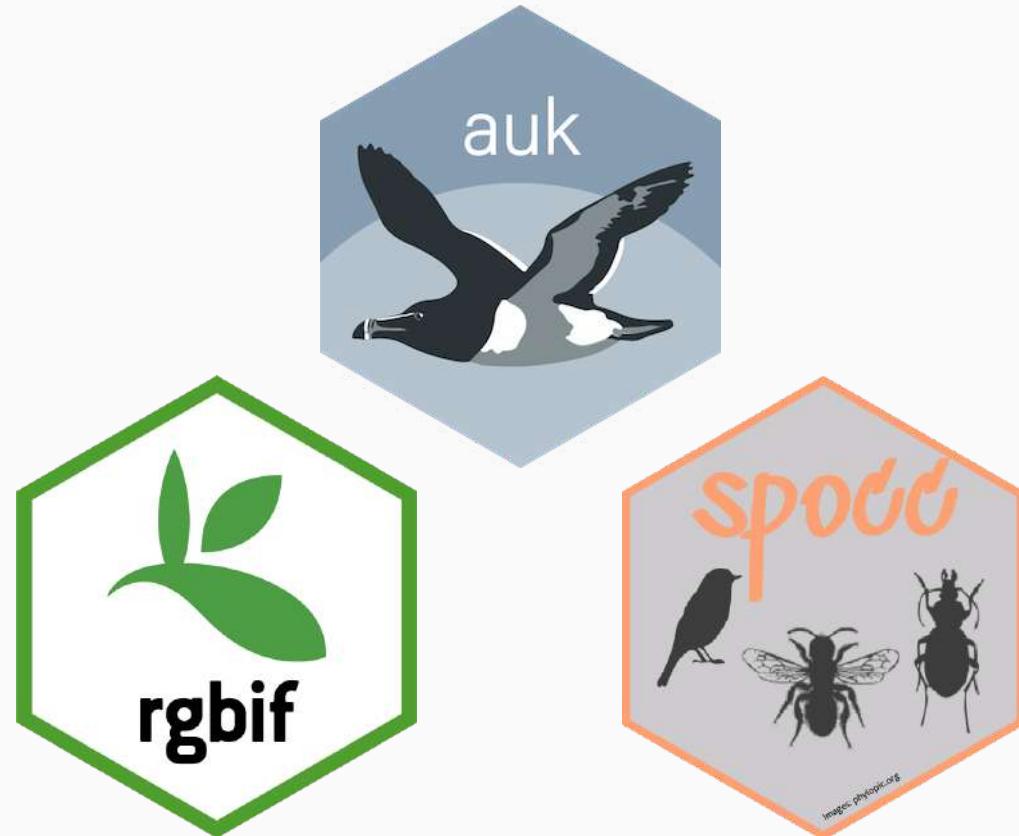
[Maitner et al. \(2017\)](#)



# 4.2 Dados de ocorrências

## Pacotes no R

- [spocc](#)
- [rgbif](#)
- [rvertnet](#)
- [rebird](#)
- [auk](#)
- [ridigbio](#)
- [BIEN](#)
- [taxize](#)
- [CoordinateCleaner](#)
- [sampbias](#)
- [spThin](#)
- [occCite](#)



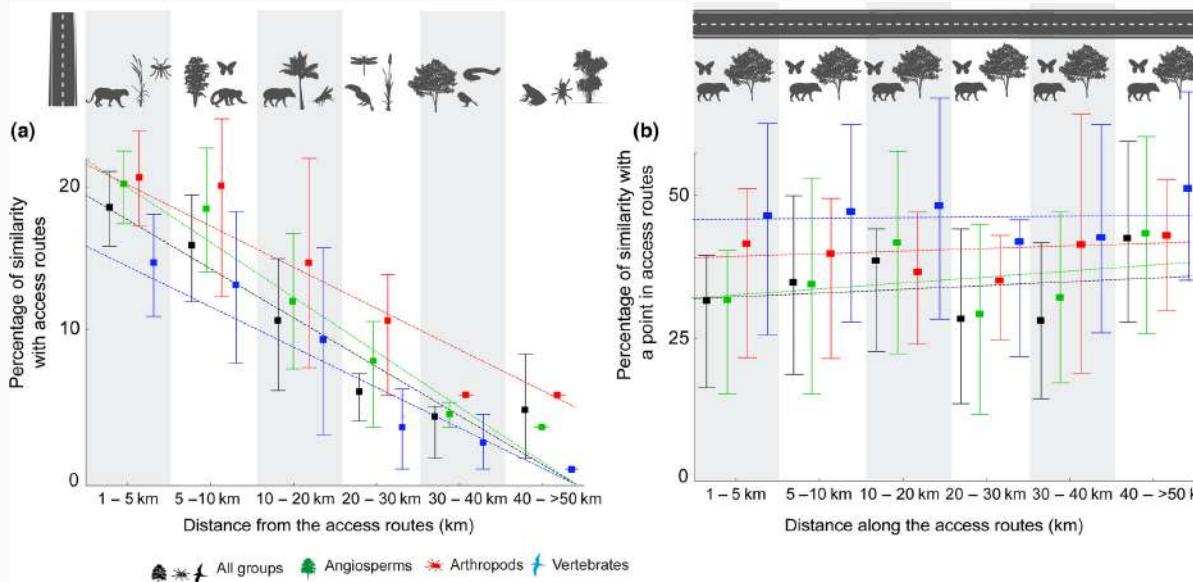
[Aiello-Lammens et al. \(2015\)](#), [Maitner et al. \(2017\)](#), [Zizka et al. \(2019\)](#), [Zizka et al. \(2020\)](#), [Owens et al. \(2021\)](#).

Mas o mundo não é tão simples e coletar dados em campo é complicado...

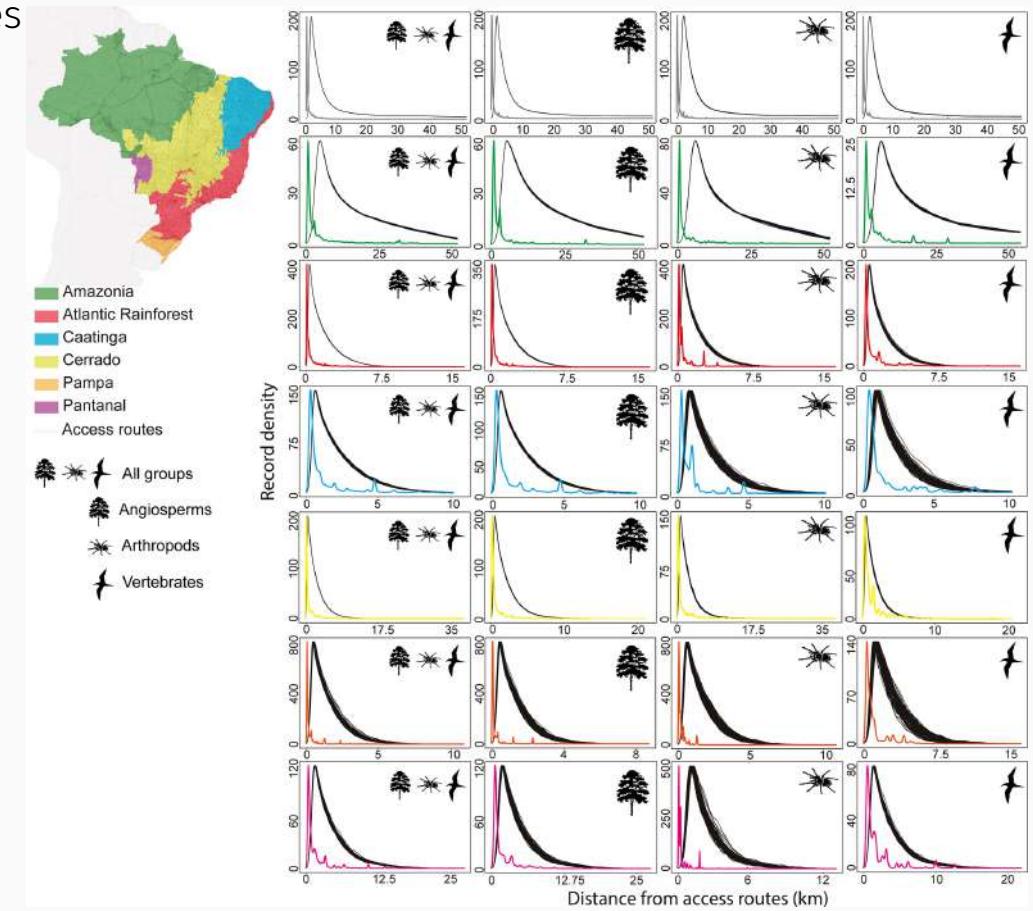
## 4.2 Dados de ocorrências

### Viés de amostragem - Brasil

- 1.144.629 (total) e 882.468 (válidos) ocorrências para 4345 espécies
- Grupos: vertebrados , artrópodes e angiospermas
- Todos as ocorrências < 1km de rotas de acesso (estradas e rios)



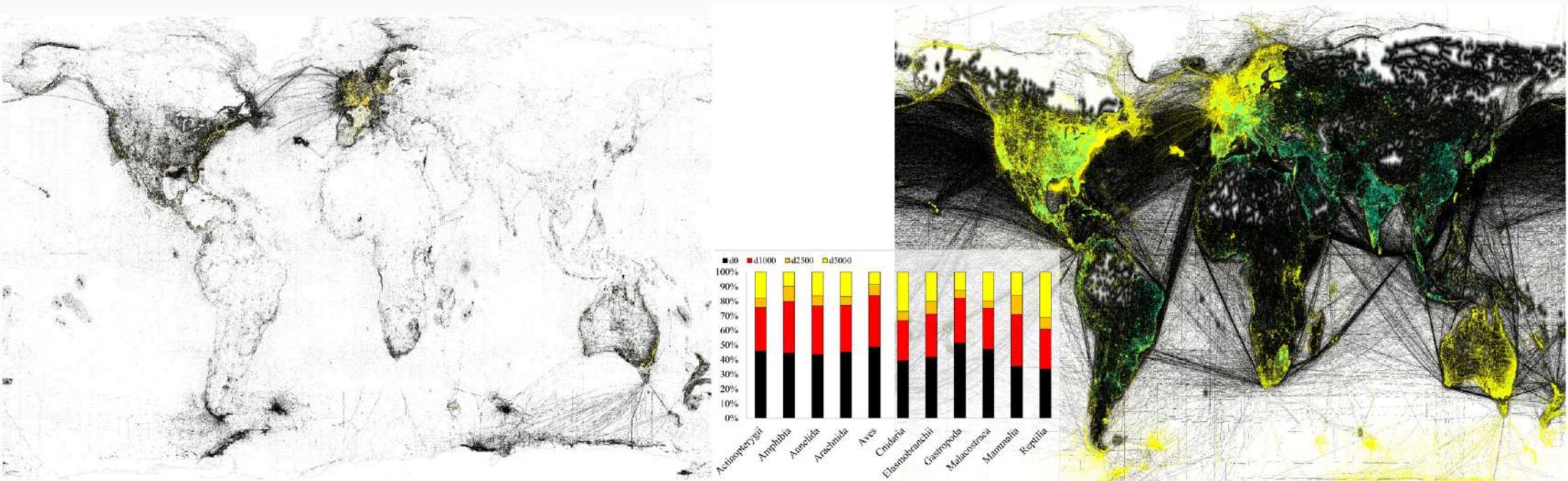
Oliveira et al. (2016)



## 4.2 Dados de ocorrências

### Viés de amostragem - Mundo

- 742 milhões de ocorrências de 374.900 espécies
- Representando 6,74% do globo amostrado



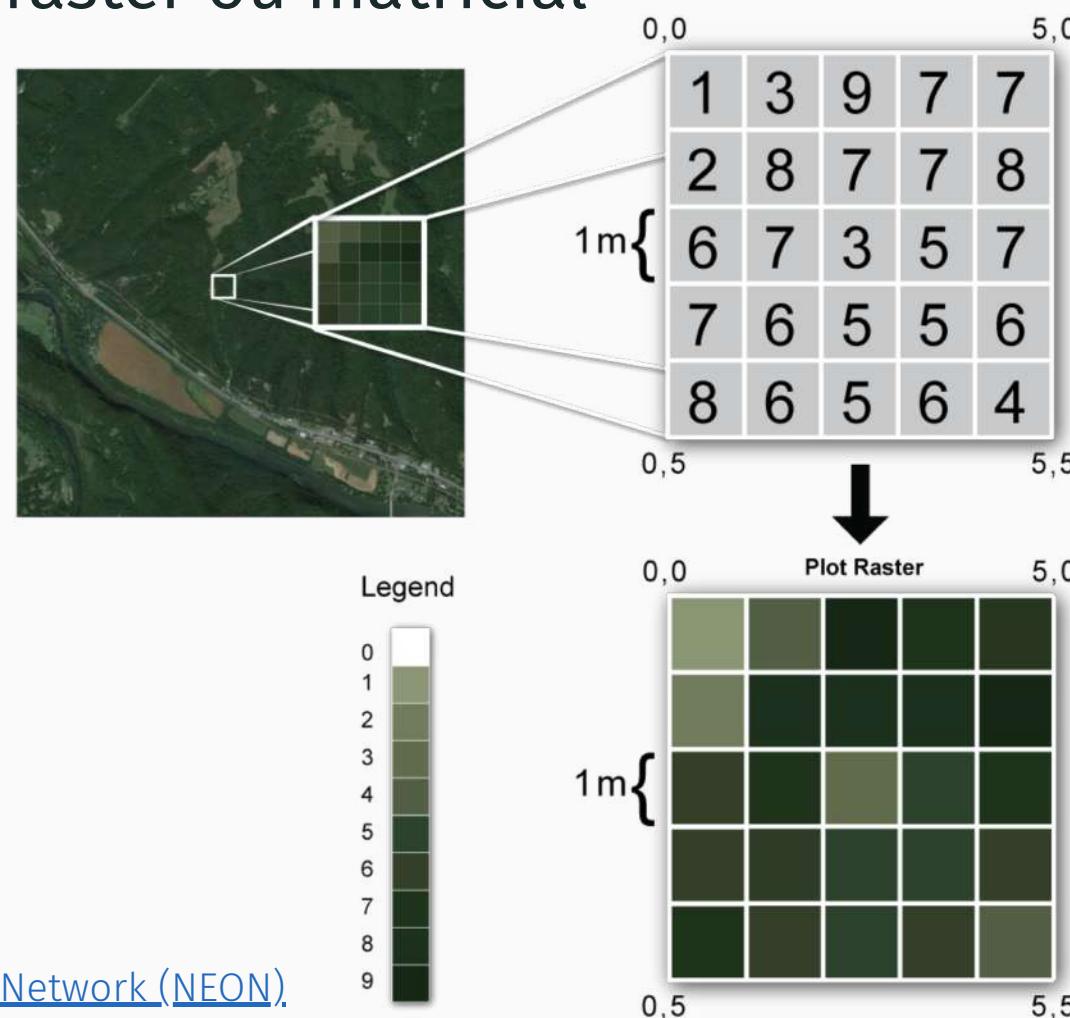
[Hughes et al. \(2021\)](#)

Bora voltar para o R para baixar as ocorrências!

## 4.3 Dados ambientais

## 4.3 Dados ambientais

Formato: gride ou raster ou matricial



# 4.3 Dados ambientais

## Bases de dados

- [WorldClim](#)
- [CHELSA](#)
- [MERRAclim](#)
- [ENVIREM](#)
- [PaleoClim](#)
- [ecoClimate](#)
- [EarthEnv](#)
- [SRTM](#)
- [Land-Use Harmonization \(LUH2\)](#)
- [GlobCover](#)
- [Global Human Footprint](#)
- [SoilGrids](#)
- [MARSPEC](#)
- [Bio-ORACLE](#)



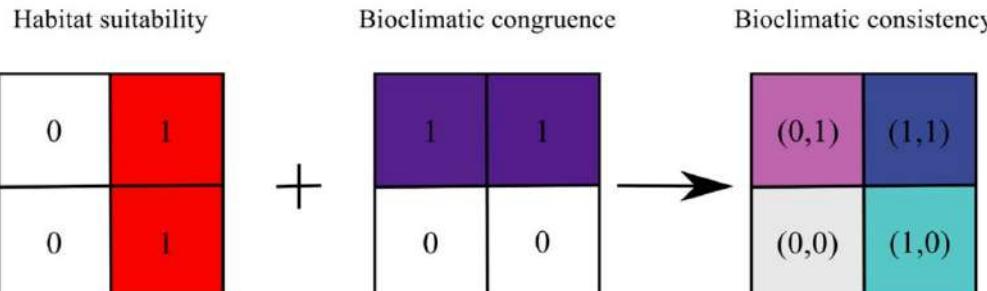
E tem muita diferença entre as bases de dados?

# 4.3 Dados ambientais

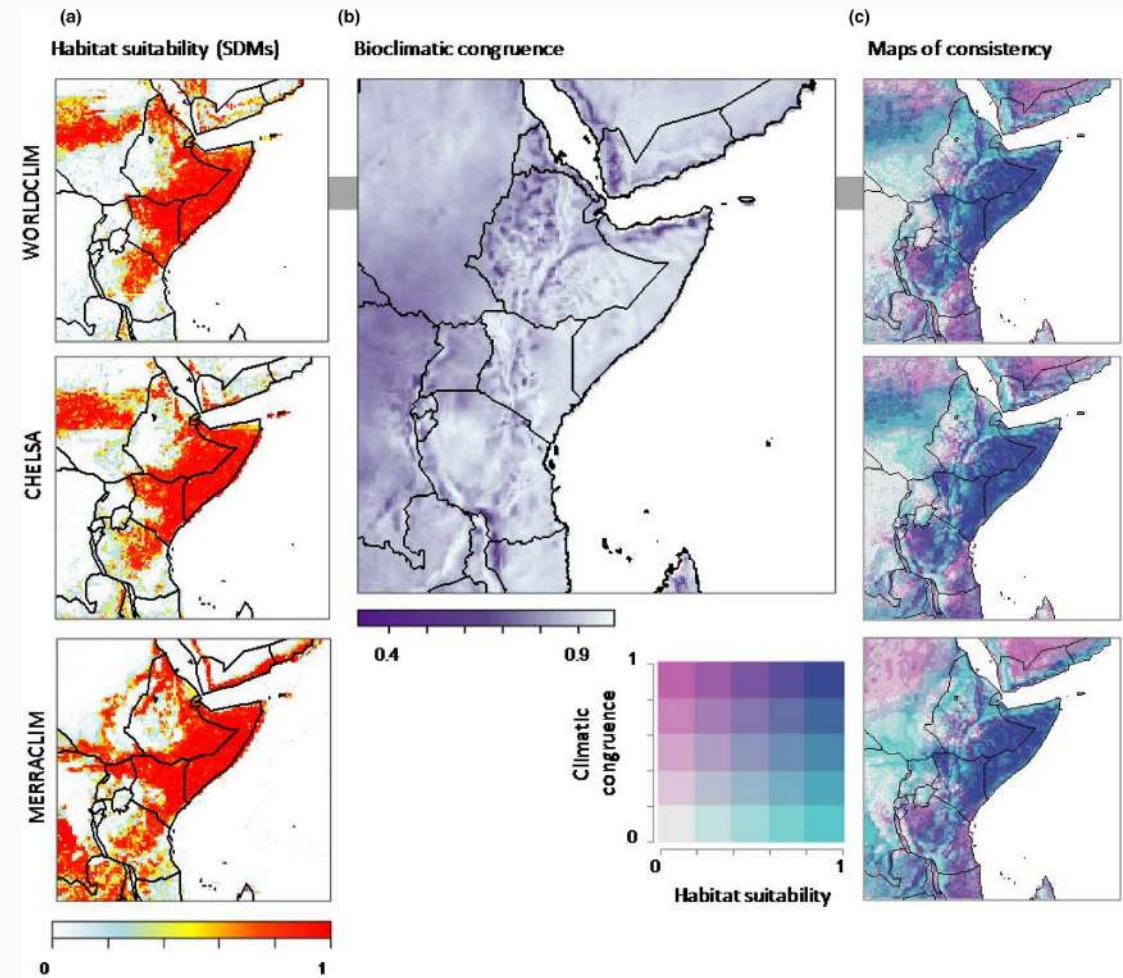
## Diferenças entre as bases de dados bioclimáticas

### Correlação

	MARSPEC	Bio-ORACLE	MERRAclim
Mean annual temperature of sea surface			
MARSPEC	1.000		
Bio-ORACLE	0.997	1.000	
MERRAclim	0.992	0.994	1.000
	WorldClim	CHELSA	MERRAclim
Mean annual temperature (mean annual precipitation)			
WorldClim	1.000		
CHELSA	0.997 (0.953)	1.000	
MERRAclim	0.991	0.994	1.000

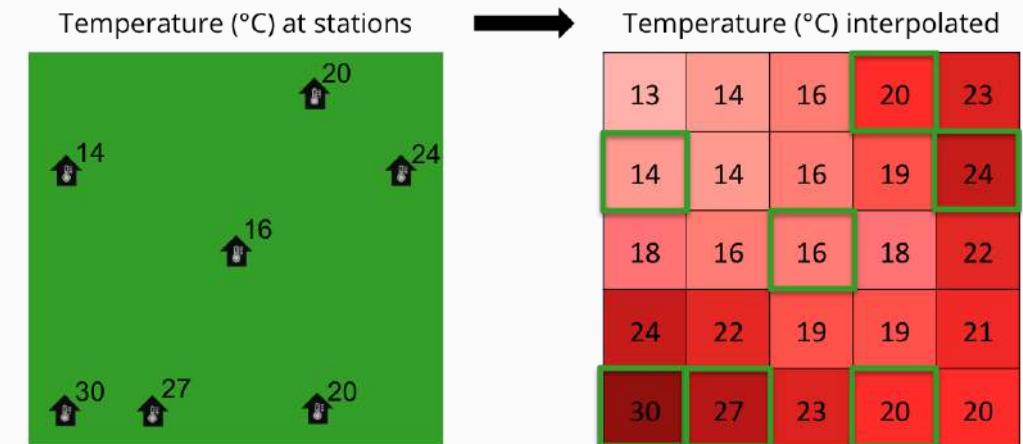
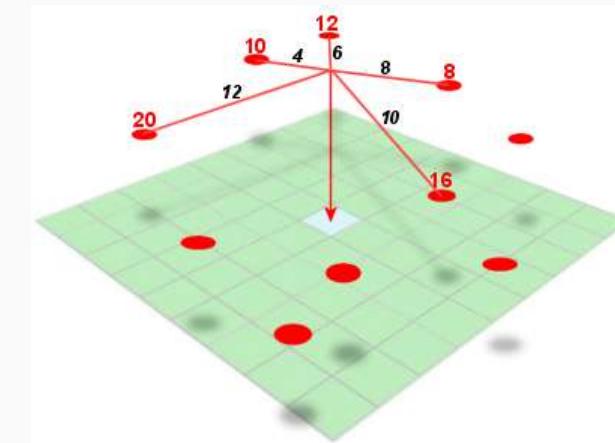
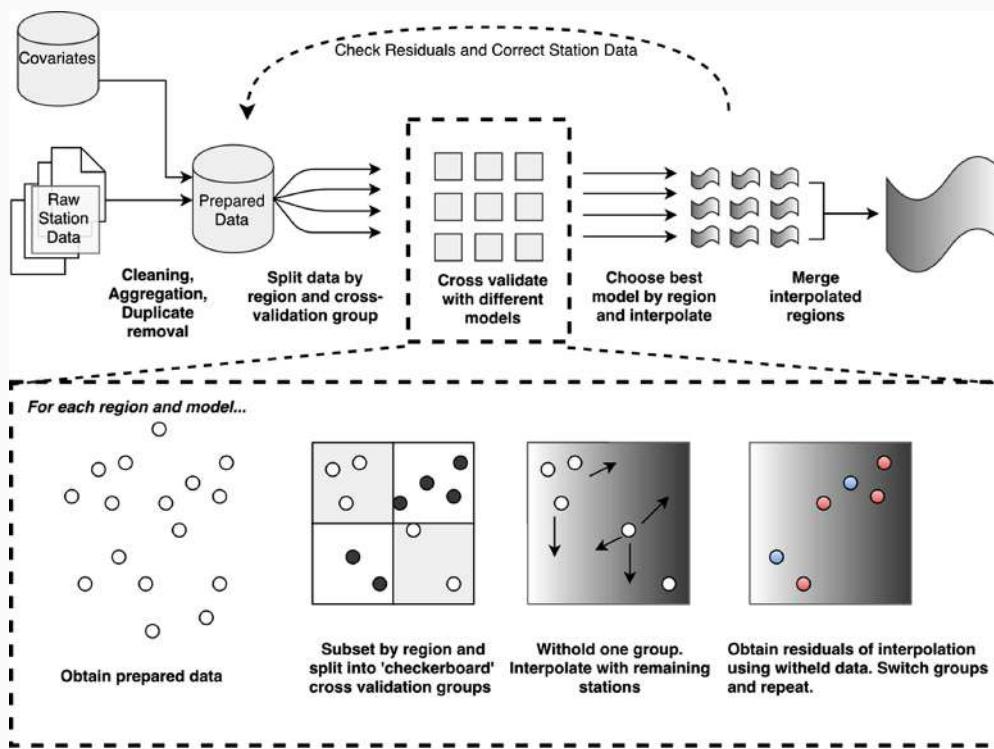


[Morales-Barbero & Vega-Álvarez \(2018\)](#)



# 4.3 Dados ambientais

## Interpolação



[Biodiversity and Climate Change Virtual Laboratory \(BCCVL\)](#),  
[National Ecological Observatory Network \(NEON\)](#),  
Fick & Hijmans (2017)

Adapted from [http://planet.botany.uwc.ac.za/nisl/GIS/spatial/chap\\_1\\_11.h](http://planet.botany.uwc.ac.za/nisl/GIS/spatial/chap_1_11.h)

## 4.3 Dados ambientais

### Variáveis bioclimáticas

Variável	Tipo	Descrição
BIO01	Temperatura	Temp. média anual
BIO02	Temperatura	Variação diurna média da temp.
BIO03	Temperatura	Isotermalidade
BIO04	Temperatura	Sazonalidade da temp.
BIO05	Temperatura	Temp. máx. do mês mais quente
BIO06	Temperatura	Temp. mín. do mês mais frio
BIO07	Temperatura	Amplitude térmica anual
BIO08	Temperatura	Temp. média do trim. mais úmido
BIO09	Temperatura	Temp. média do trim. mais seco
BIO10	Temperatura	Temp. média do trim. mais quente
BIO11	Temperatura	Temp. média do trim. mais frio

Variável	Tipo	Descrição
BIO12	Precipitação	Prec. anual
BIO13	Precipitação	Prec. do mês mais chuvoso
BIO14	Precipitação	Prec. do mês mais seco
BIO15	Precipitação	Sazonalidade da prec.
BIO16	Precipitação	Prec. do trim. mais chuvoso
BIO17	Precipitação	Prec. do trim. mais seco
BIO18	Precipitação	Prec. do trim. mais quente
BIO19	Precipitação	Prec. do trim. mais frio

\*Temp. = Temperatura; Prec. = Precipitação;  
trim. = Trimestre;

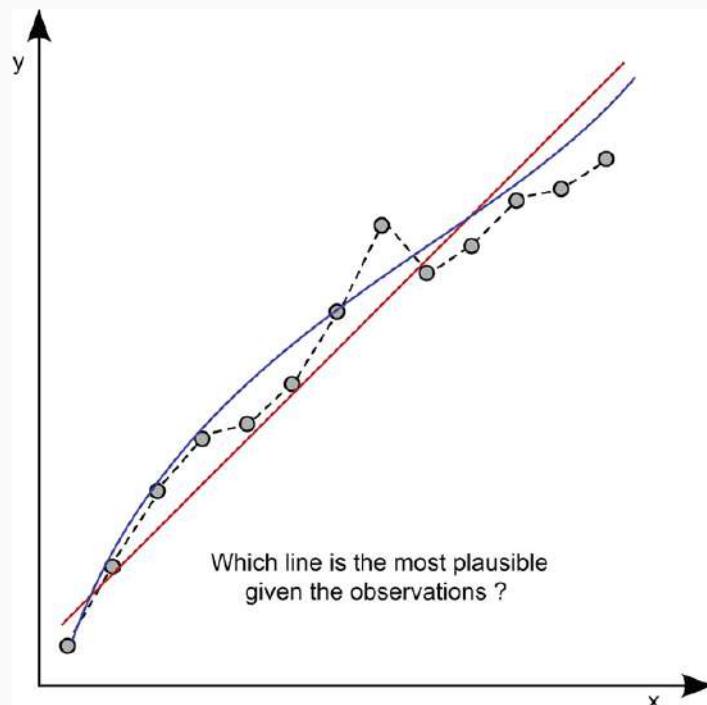
[O'Donnell & Ignizio \(2012\)](#)

Mas informação demais quase sempre atrapalha...

## 4.3 Dados ambientais

### Princípio da Parcimônia (Por que complicar esse modelo aí?)

- William of Ockham's (1285–1348)
- "É fútil fazer com mais o que pode ser feito com menos"



[Coelho et al. \(2018\)](#), [What does Occam's razor mean? \(2019\)](#).

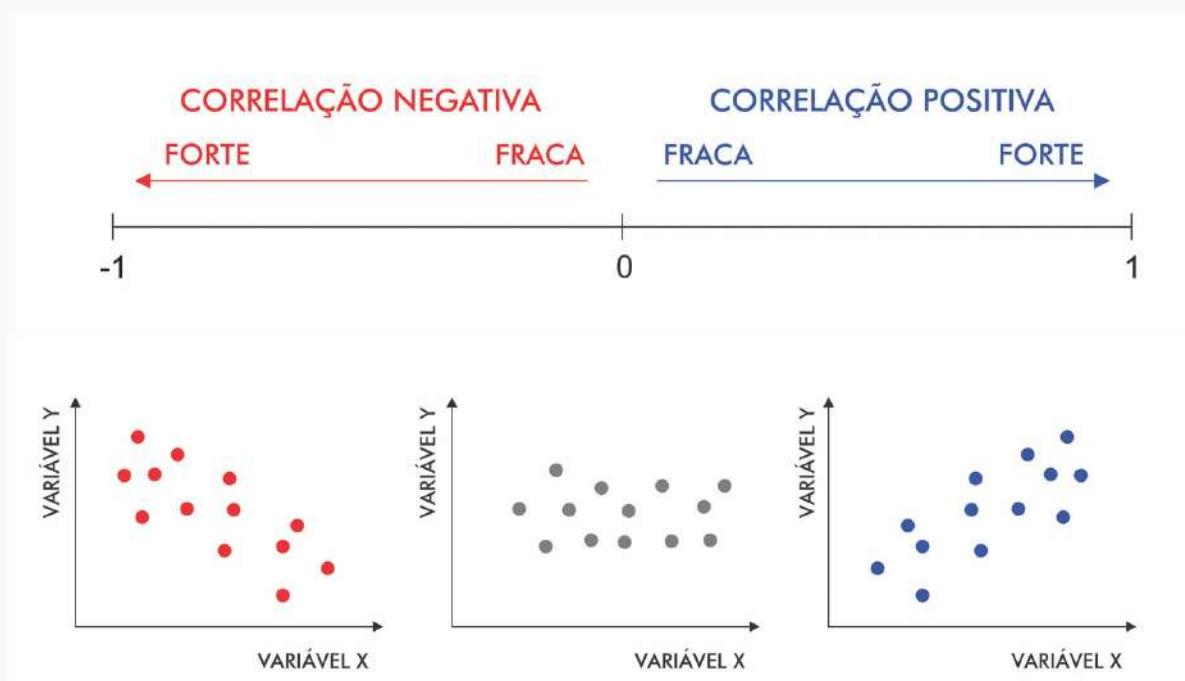


## 4.3 Dados ambientais

### Colinearidade - Correlação (Pearson ou Spearman)

- Mede o grau de relação (positiva ou negativa) entre duas variáveis
- $r > |0.7|$ : indica alta correlação

$$r = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$



[Zuur et al. \(2010\)](#), [Dormann et al. \(2012\)](#), [Naimi et al. \(2014\)](#), [Feng et al. 2019](#), [Coeficientes de correlação](#)

## 4.3 Dados ambientais

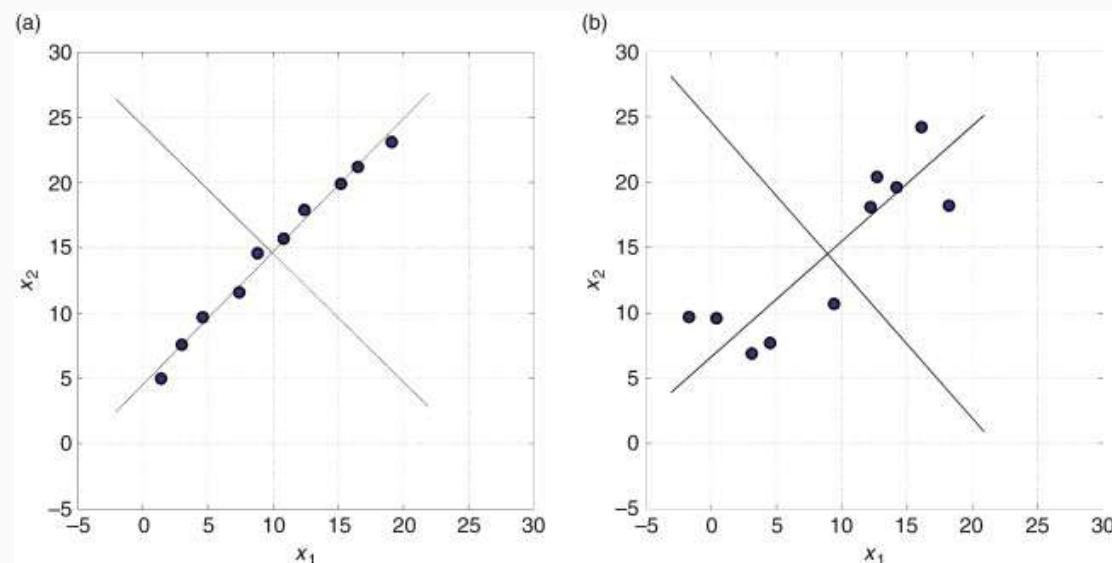
### Colinearidade - Fator de Inflação de Variância (VIF)

- Medida da quantidade de multicolinearidade em um conjunto de múltiplas variáveis
- VIF = 1: não há multicolinearidade entre as variáveis
- VIF > 2 ou VIF > 5 ou VIF > 10: indica alta multicolinearidade (depende do autor)

- $VIF_i = \frac{1}{1 - R_i^2}$

- (a)  $r = 0.8946$  e  $VIF = 5.007$

- (b)  $r = 0.7697$  e  $VIF = 2.453$



[Zuur et al. \(2010\)](#), [Dormann et al. \(2012\)](#), [Naimi et al. \(2014\)](#), [Feng et al. 2019](#), [VIF](#)

Dúvidas?

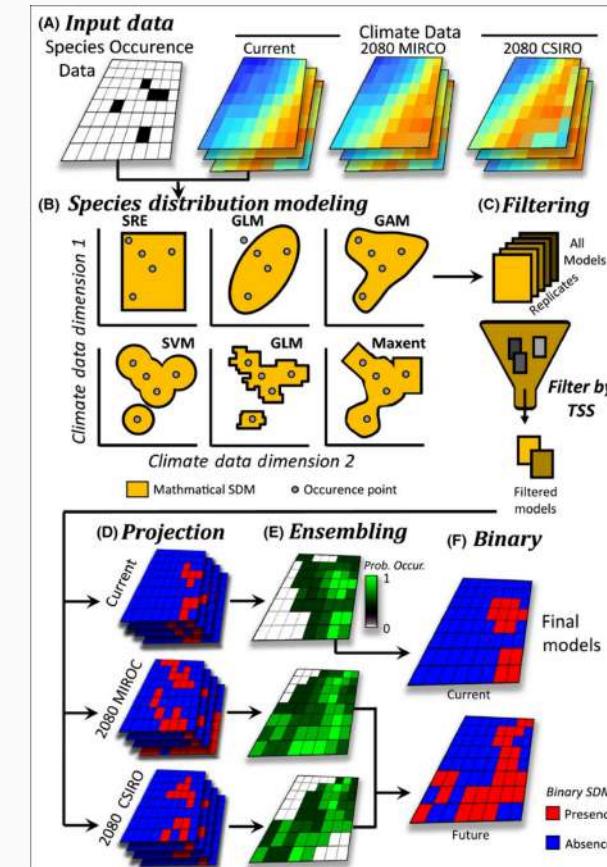
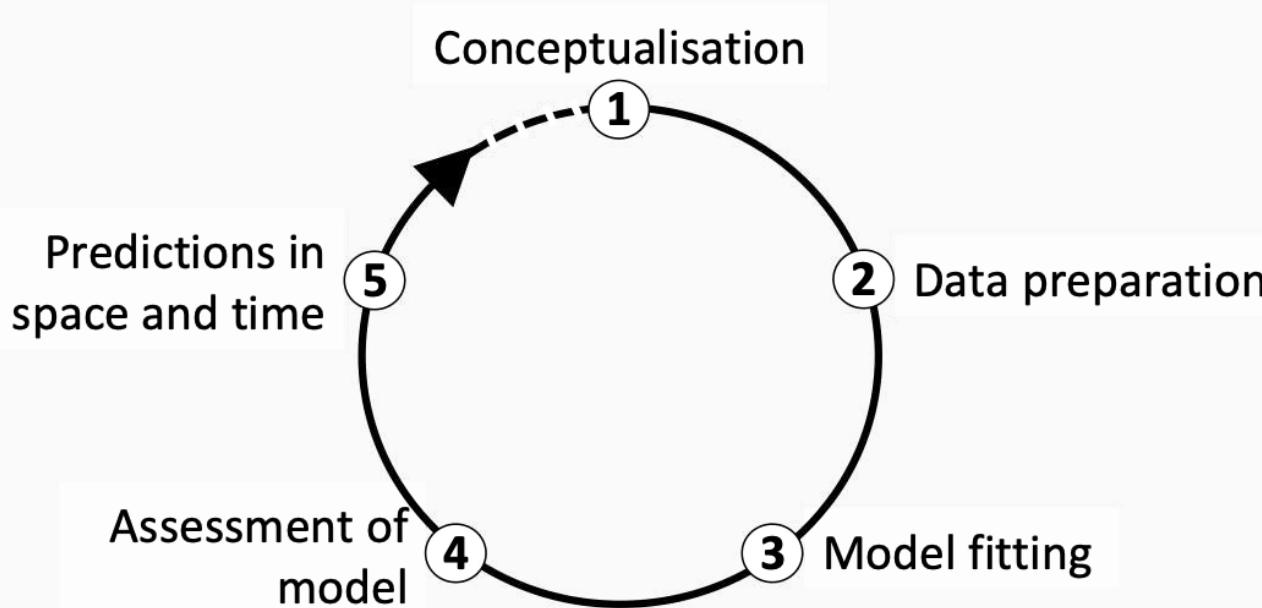
Bora voltar para o R para baixar as variáveis!

Encerramos por hoje =]

Bom dia povo! Vamos recapitular!

# 2.4 Padronização para criação dos modelos

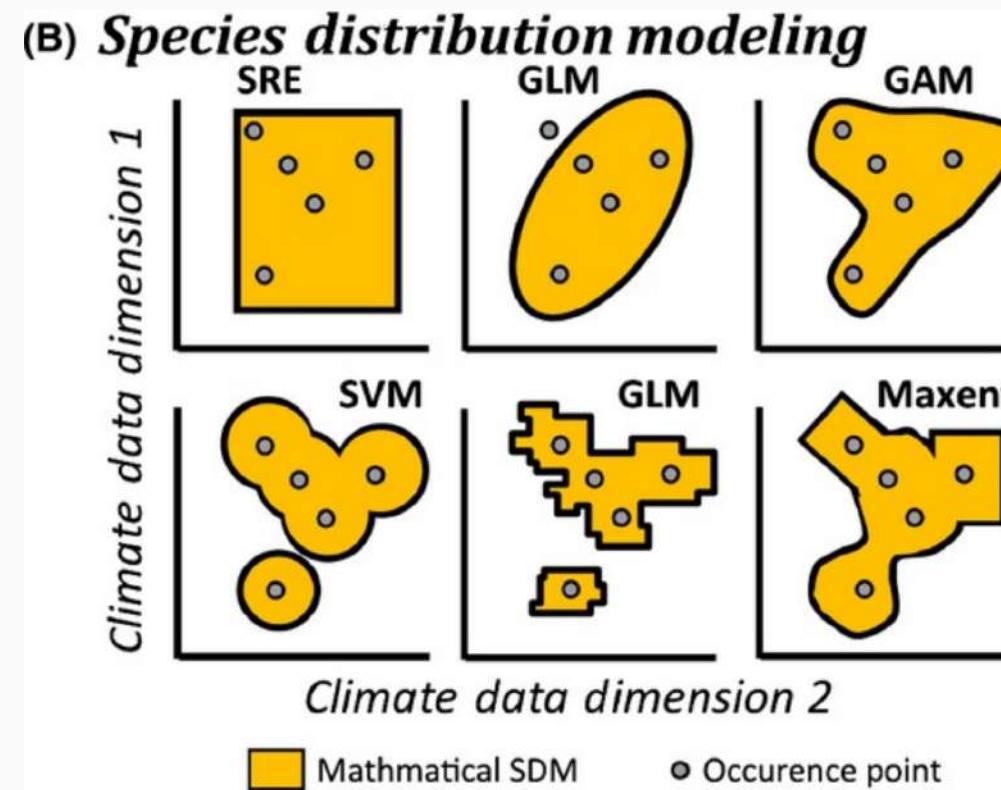
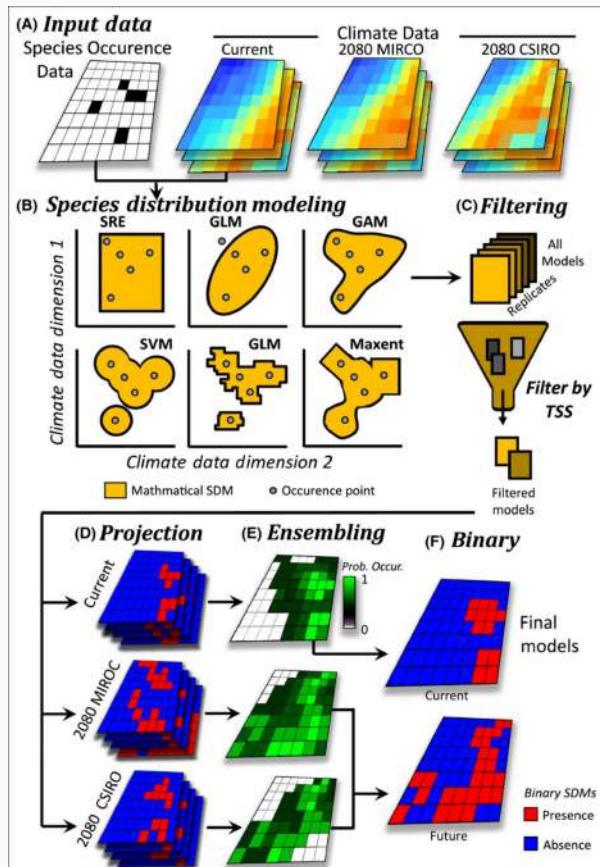
Protocolo padrão para modelos de distribuição de espécies



[Brown & Yoder \(2015\)](#), [Zurell et al. \(2020\)](#)

# 2.4 Padronização para criação dos modelos

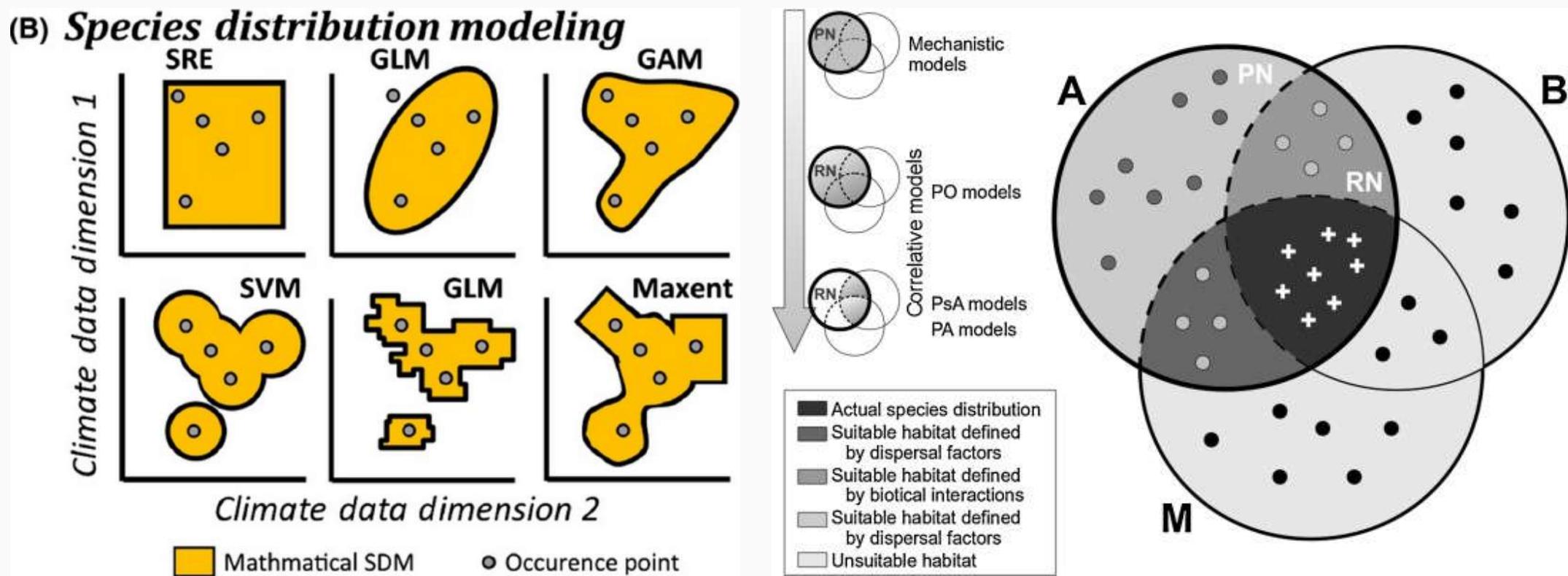
## 3. Ajuste dos modelos



# 5 Ajuste dos modelos

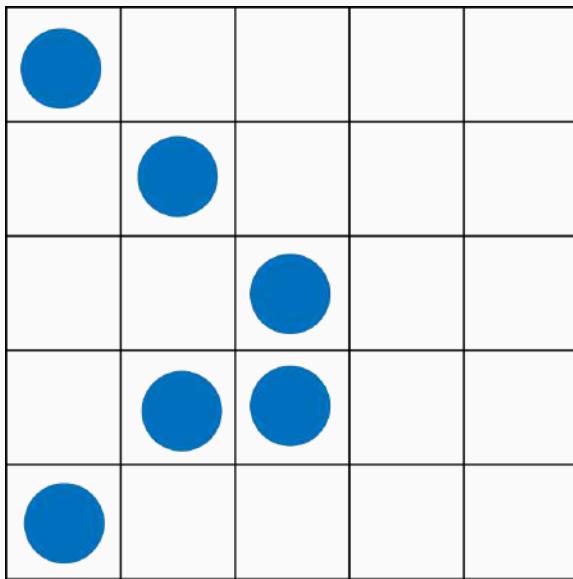
# 5.1 Tipos de métodos ou algoritmos

Modela diferentes áreas do Diagrama BAM

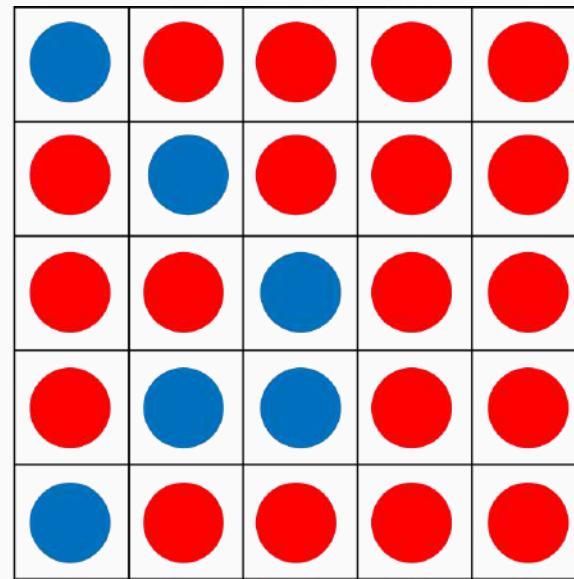


# 5.1 Tipos de métodos ou algoritmos

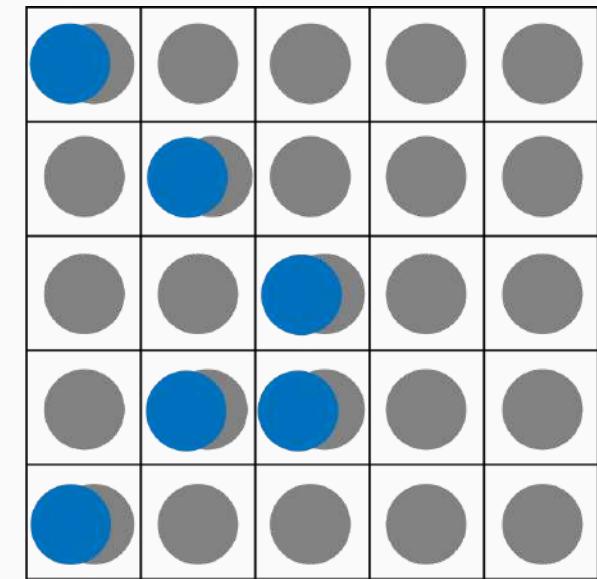
Depende dos dados de presença e ausência da espécie



● Presença



● Presença  
● Ausência

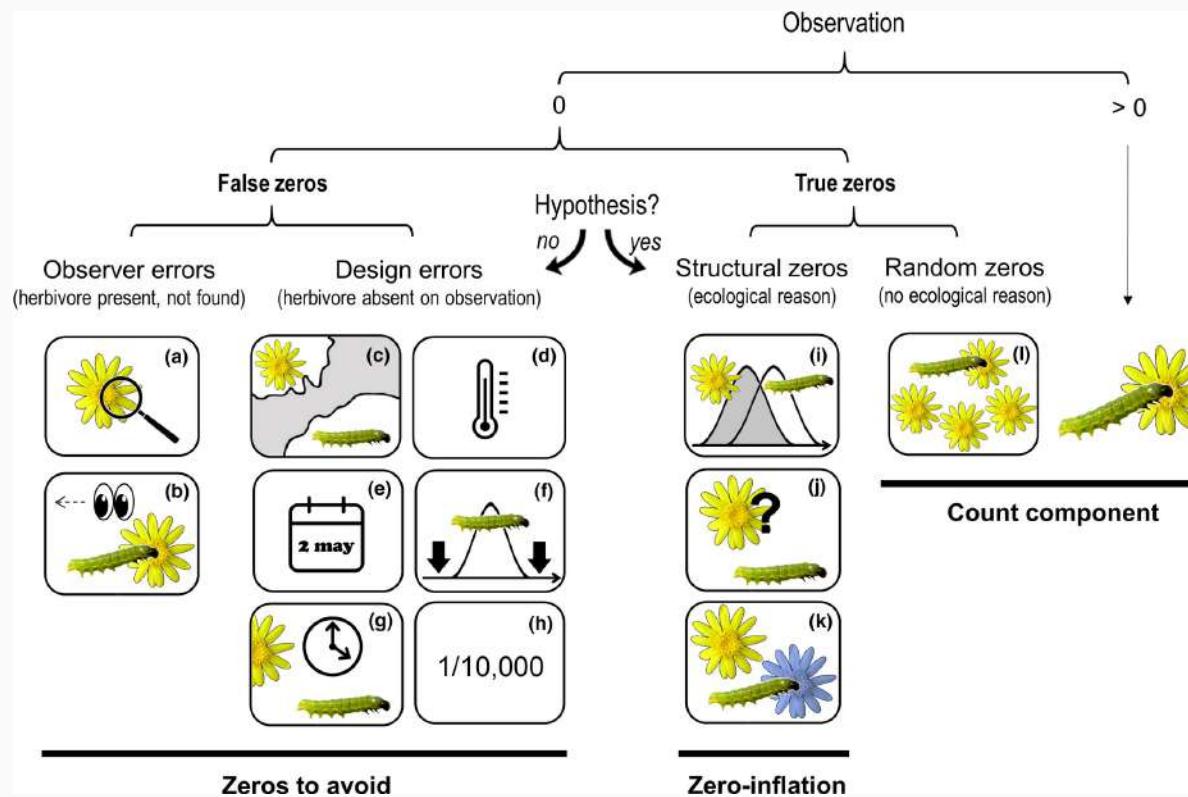


● Presença  
● Ausência  
● Background

# 5.1 Tipos de métodos ou algoritmos

## Ausências em Ecologia...

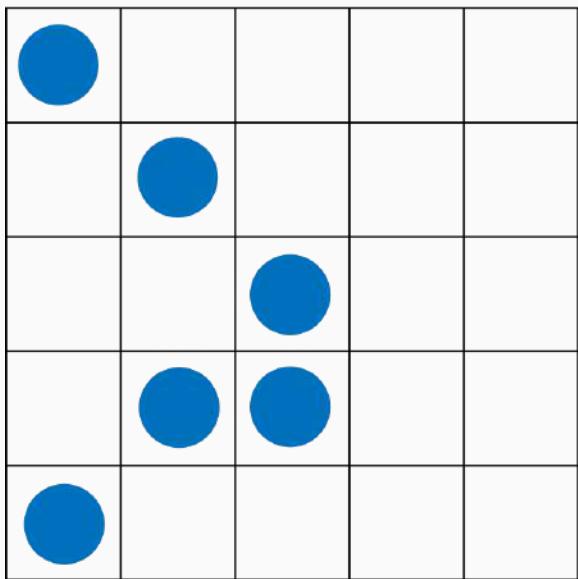
- **Zeros falsos:** falta de experiência do observador (a – b) ou projeto experimental pobre (c – h)
- **Zeros verdadeiros:** relacionados ao sistema ecológico em estudo - zeros estruturais (i – k) ou emergentes da variabilidade de amostragem - zeros aleatórios (l)



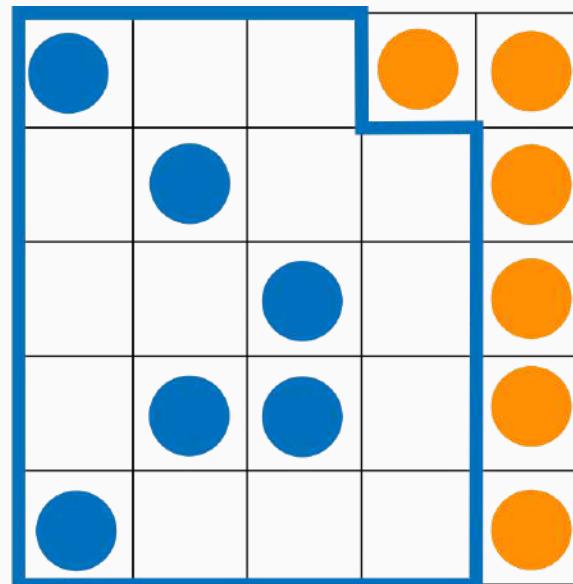
[Blasco-Moreno et al. \(2019\)](#)

# 5.1 Tipos de métodos ou algoritmos

Depende dos dados de presença e ausência da espécie

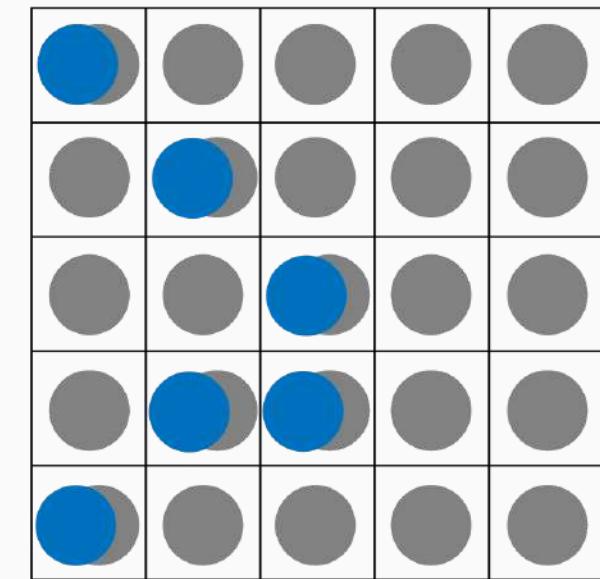


● Presença



● Presença

● Pseudo-ausência



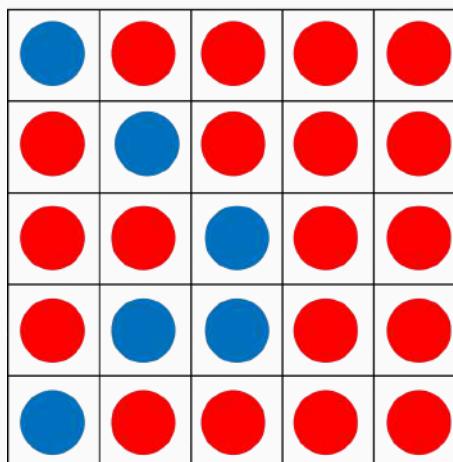
● Presença

● Background

# 5.1 Tipos de métodos ou algoritmos

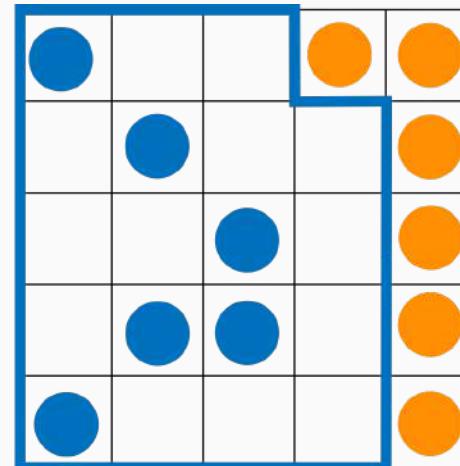
## Ausência vs pseudo-ausência vs background

- **Ausência:** dados de "ausências verdadeiras" (quando disponíveis)
- **Pseudo-ausência:** dados de ausência gerados por amostragem aleatória (seguindo ou não regras geográficas ou ambientais), geralmente em número igual ao número de presenças
- **Background:** dados gerados para amostrar as condições da área de estudo (quanto mais, melhor)



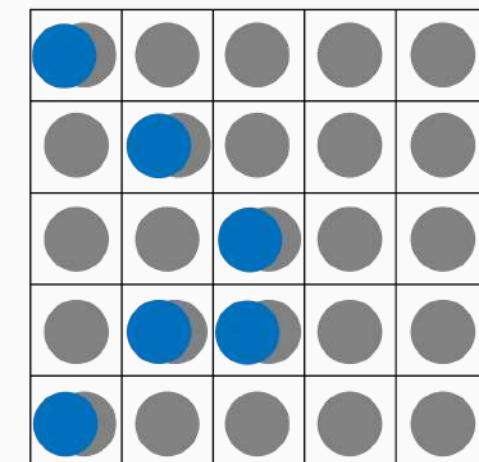
● Presença

● Ausência



● Presença

● Pseudo-ausência



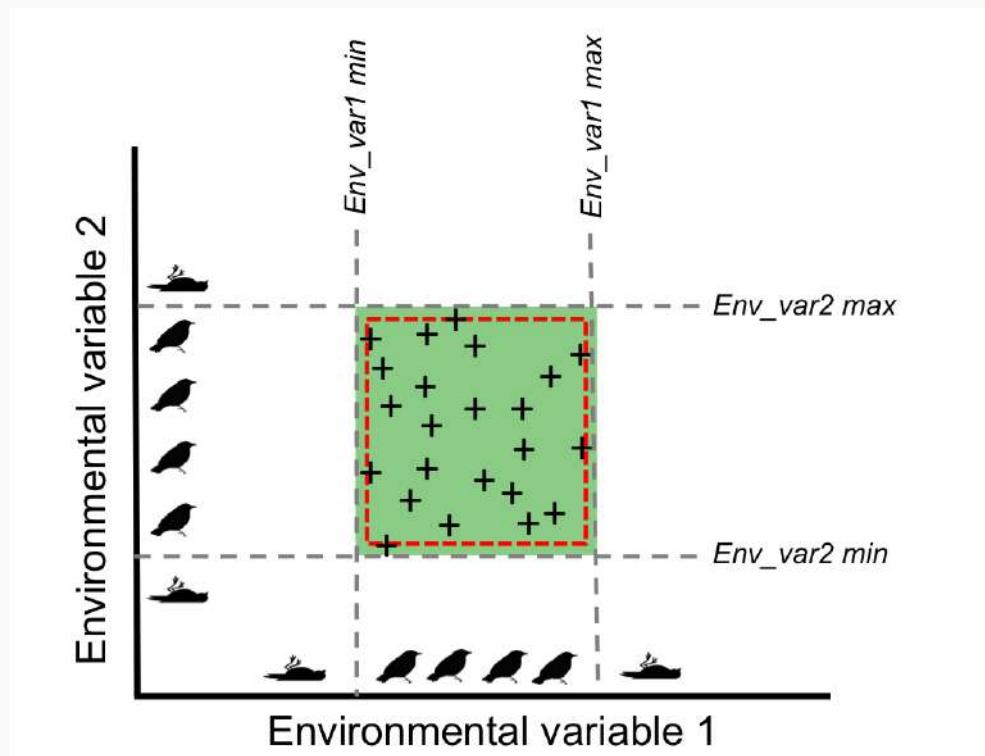
● Presença

● Background

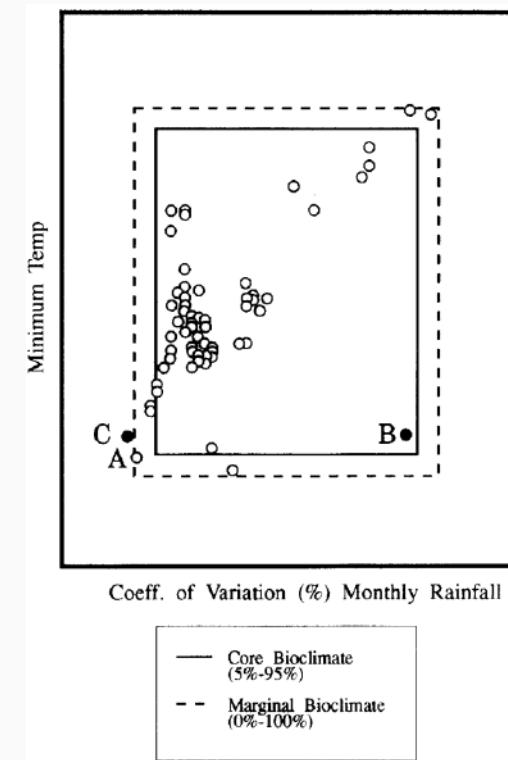
## 5.2 Somente-presença

### Bioclim (*Envelope score*)

- Envelope climático com escores



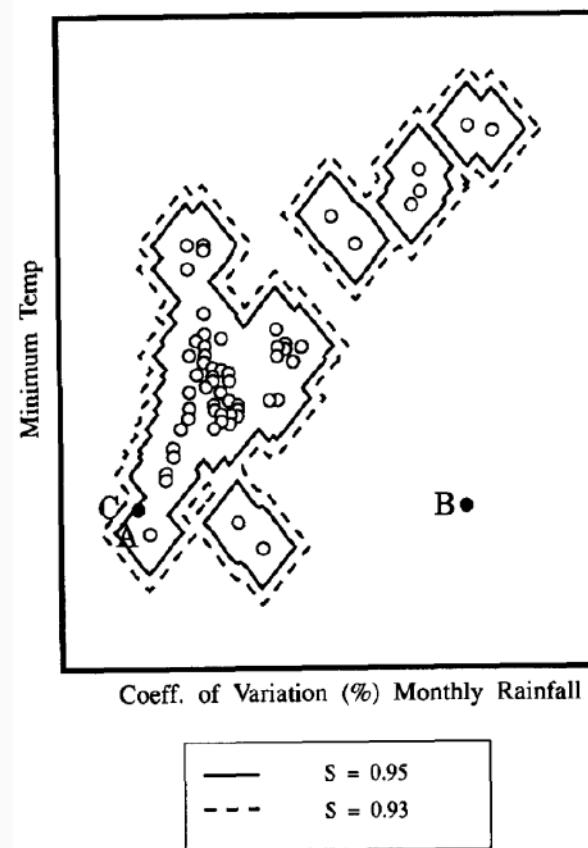
[Nix \(1986\)](#), [Booth et al. \(2014\)](#)



## 5.2 Somente-presença

### Domain (Distântica de Gower)

- Distância ponto a ponto

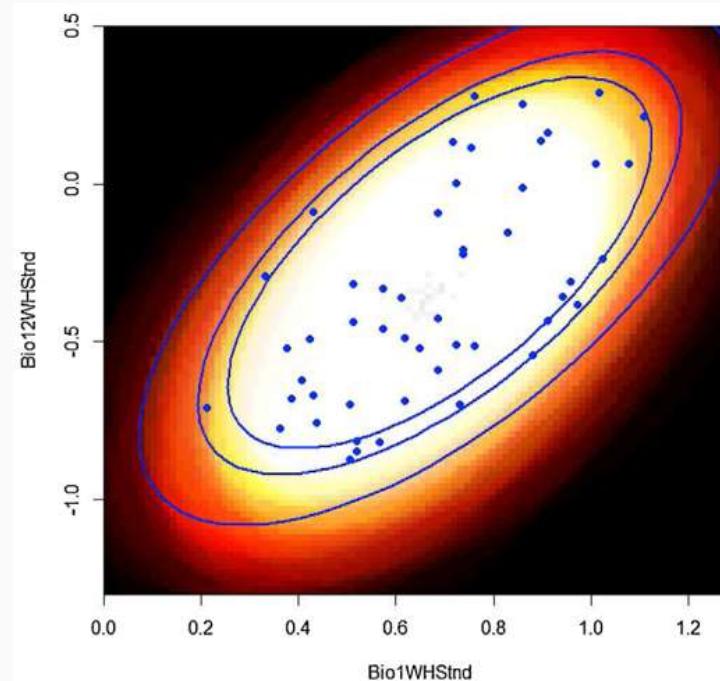
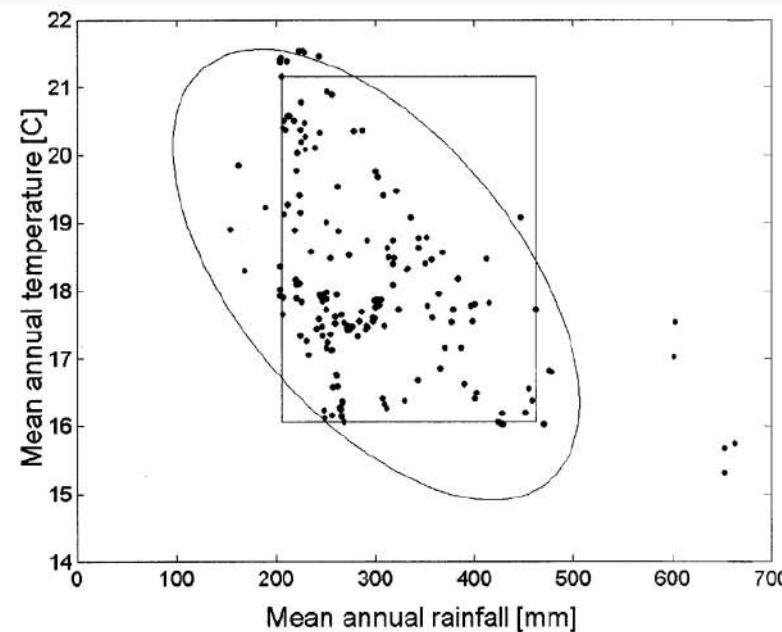


[Carpenter et al. \(1993\)](#)

## 5.2 Somente-presença

### Mahalanobis (Distância de Mahalanobis)

- Envelope de Mahalanobis é **oblíquo** (inclinado) lidando com correlações entre as variáveis climáticas
- Envelope de Mahalanobis é **elíptico**, refletindo melhor a tendência central da teoria do nicho
- Previsões são baseadas em **todos os dados**, sendo menos sensíveis a valores discrepantes

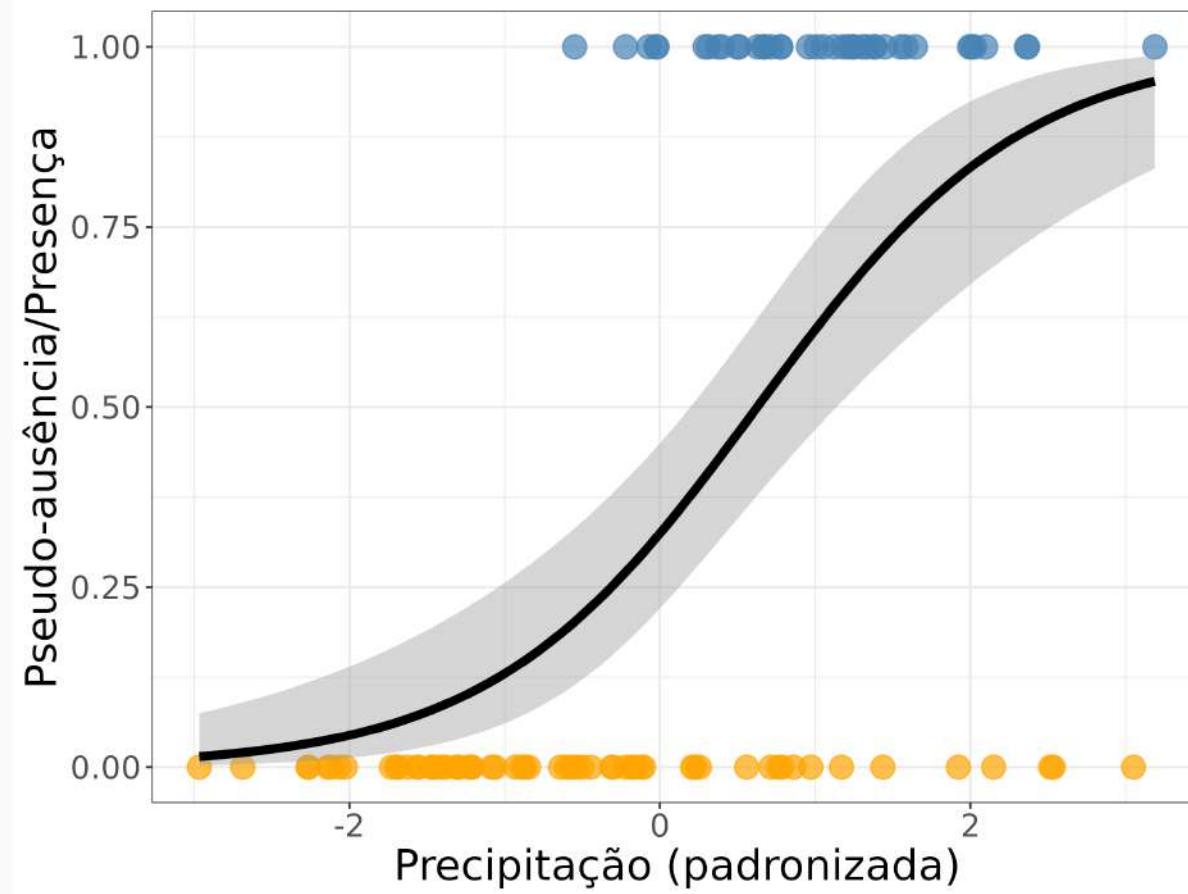


[Farber & Kadmon \(2003\)](#), [Jiménez et al. \(2019\)](#)

## 5.4 Presença-(pseudo)ausência

### GLM (Modelos Lineares Generalizados - *Generalized Linear Models*)

- Distribuição de erros binomial

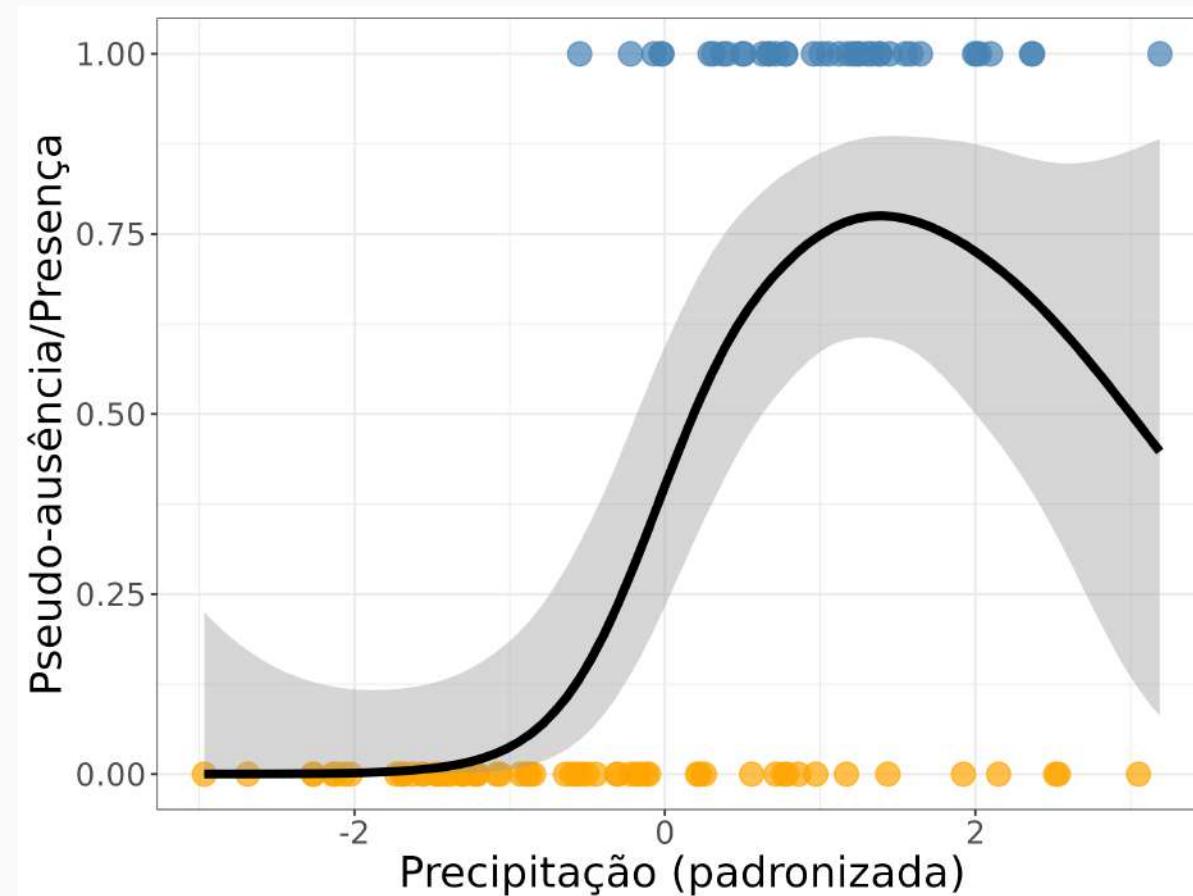


[Guisan et al. \(2002\)](#)

## 5.4 Presença-(pseudo)ausência

### GAM (Modelos Generalizados Aditivos - *Generalized Additive Model*)

- Distribuição de erros binomial
- Funções de suavização

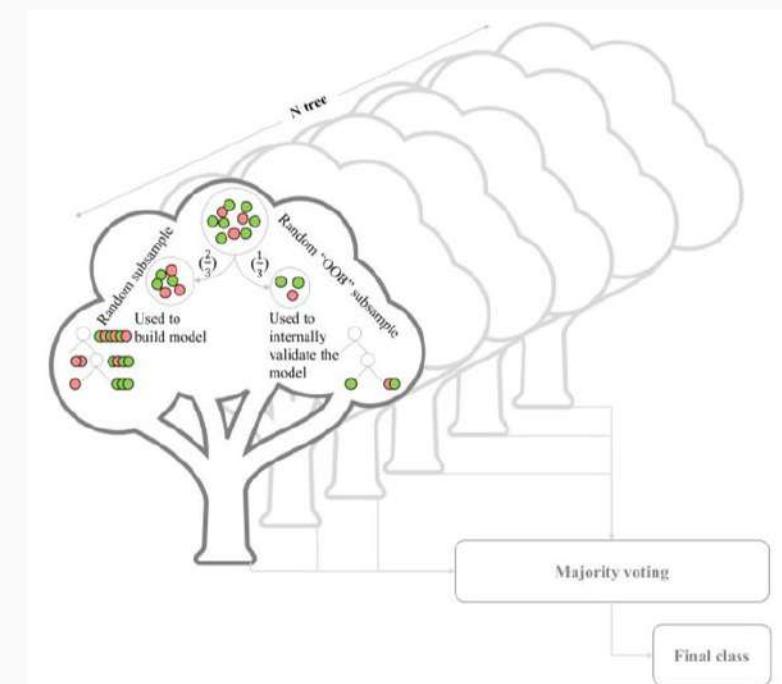
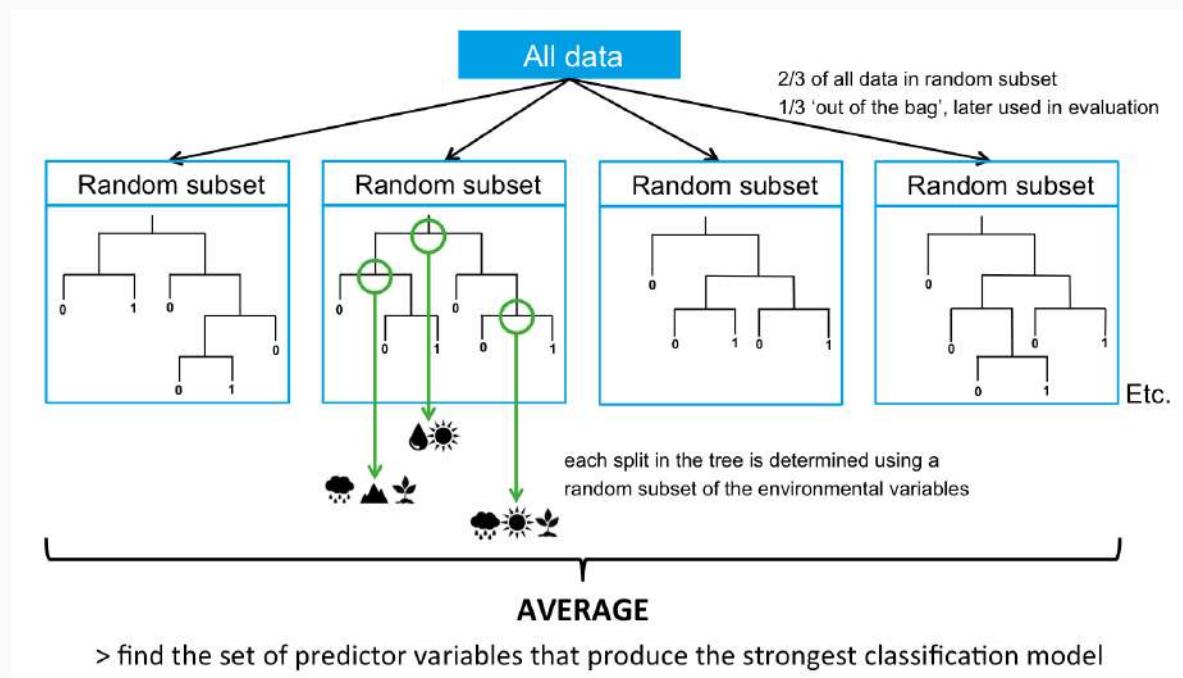


[Guisan et al. \(2002\)](#)

# 5.4 Presença-(pseudo)ausência

## Random Forest

- Cria diversas **árvores de decisão** ('floresta') por **bootstrap** (reamostragem)
- Combina os **resultados** de cada árvore em uma última classificação (categórico) ou regressão (contínuo)

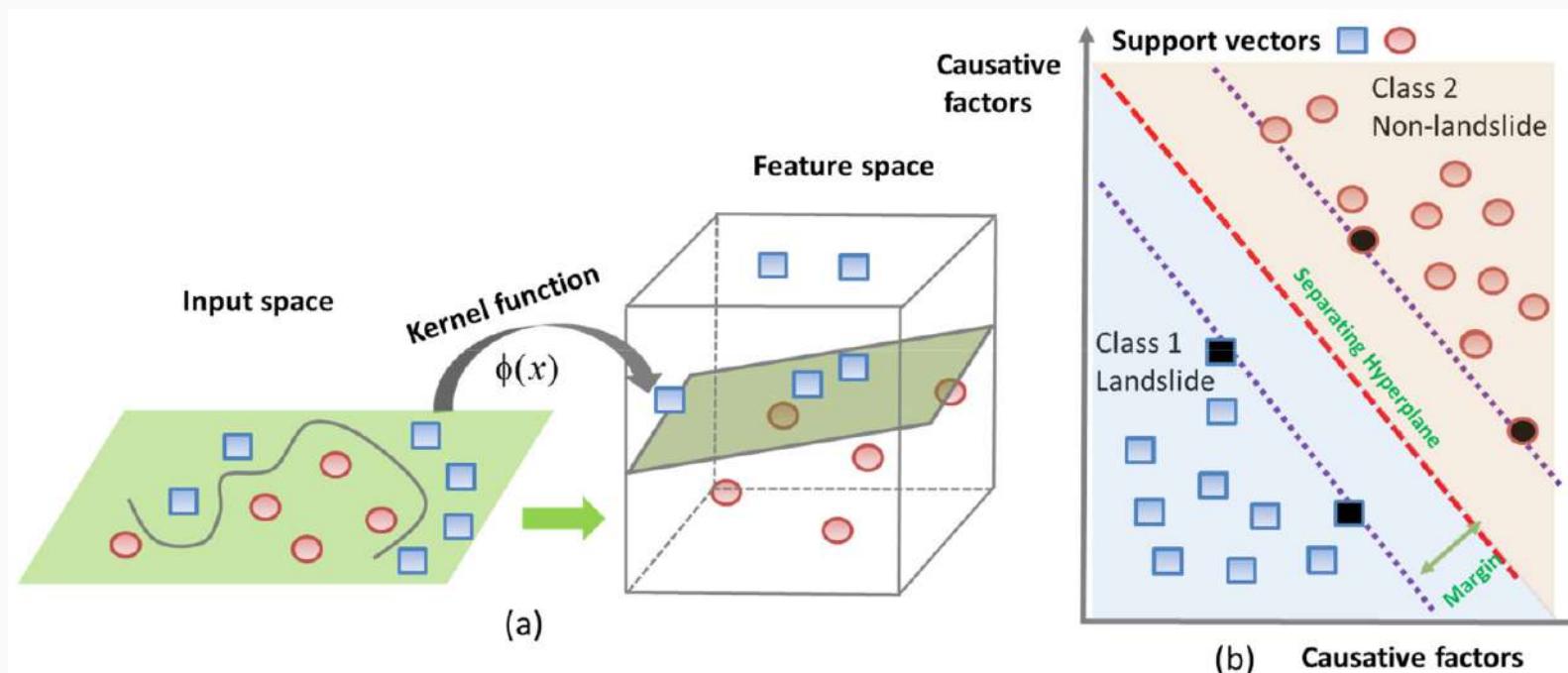


[Breiman \(2001\)](#), [Liaw & Wiener \(2002\)](#), [Prasad et al. \(2006\)](#), [Cutler et al. \(2007\)](#), [Sheykhmousa et al. \(2020\)](#), [BCCVL](#)

# 5.4 Presença-(pseudo)ausência

## SVM (*Support Vector Machine*)

- Cria **hiperplanos** através de **vetores de suporte** considerando features
- Faz a classificação (categórico) ou regressão (contínuo) para **separar** os dados em duas ou mais partes

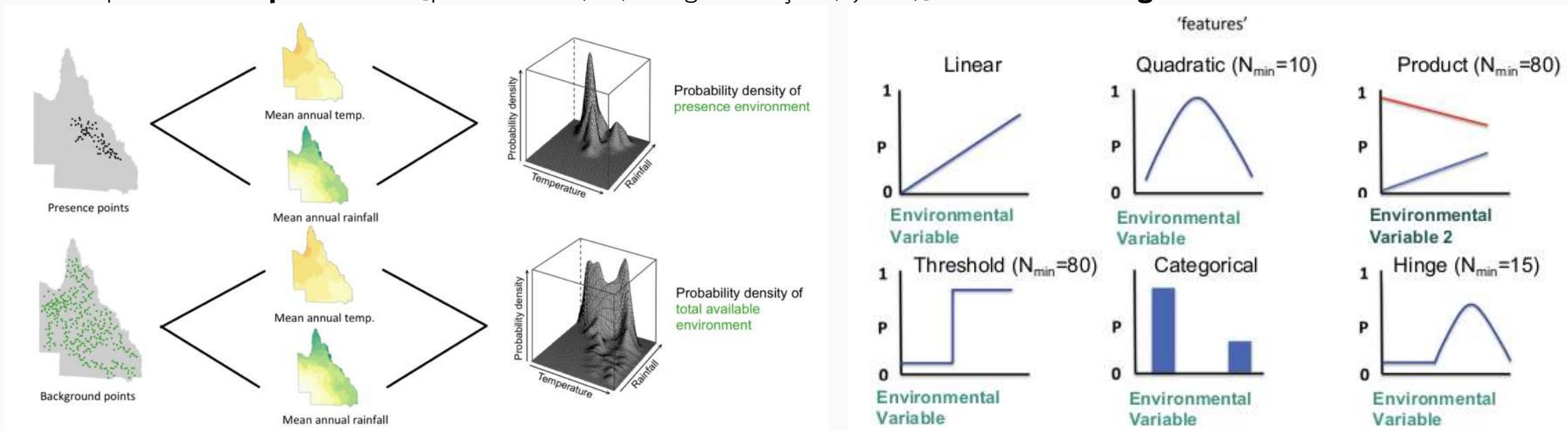


[Tax & Duin \(2004\)](#), [Guo et al. \(2005\)](#), [Drake et al. \(2006\)](#), [Dou et al. \(2019\)](#).

# 5.3 Presença-background

## MaxEnt (Máxima Entropia)

- Encontra parâmetros que **aproximam a forma da distribuição** (maximiza a entropia) dos valores ambientais do background e em função dos valores ambientais das presenças
- Usa diferentes curvas ('features') para **ajustar a máxima entropia** e achar os valores de adequabilidade ambiental
- Dependente de **parâmetros** [prevalência (0.5) e regularização (ajuste)] e **limite do background**



# 5.4 Outros métodos ou algoritmos

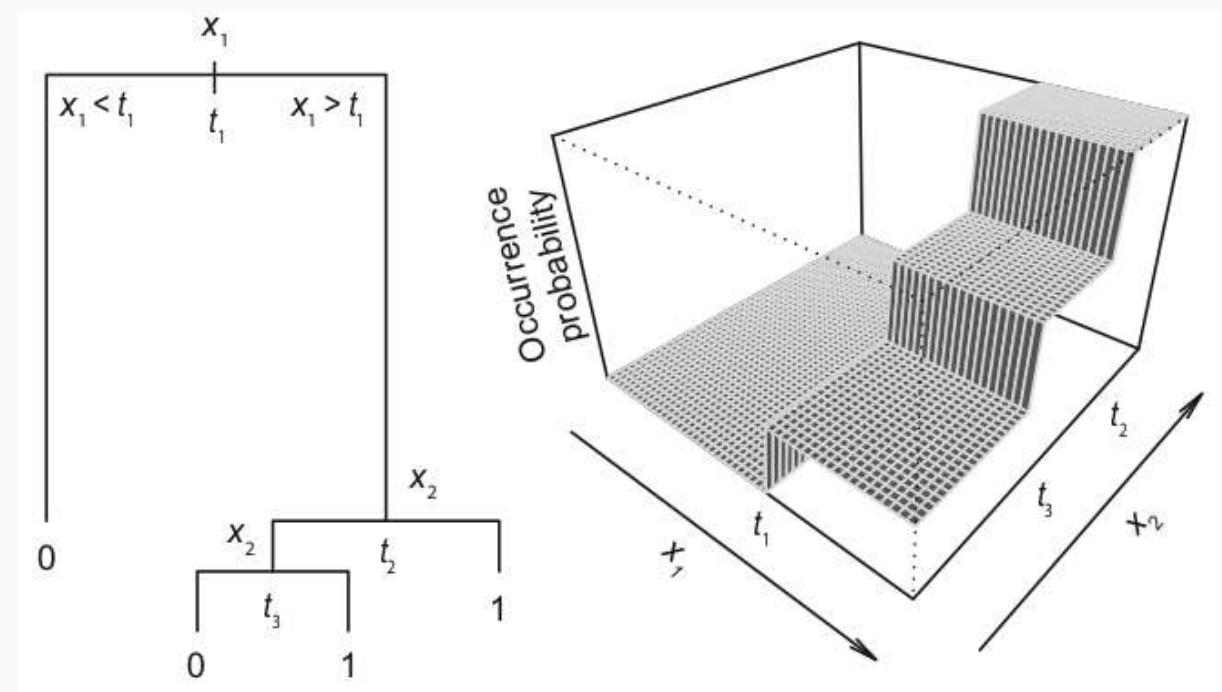
Há ainda diversos outros métodos ou algoritmos

## Presença-(pseudo)ausência

- Classification and regression trees (CART)
- Multivariate adaptive regression spline (MARS)
- Mixture discriminant analysis (MDA)
- Artificial neural networks (ANN)
- Boosted Regression Trees (BRT)
- Bayesian Gaussian Process

## Presença-background

- Ecological Niche Factor Analysis (ENFA)
- Genetic Algorithm for Rule-set Production (GARP)
- Maxlike

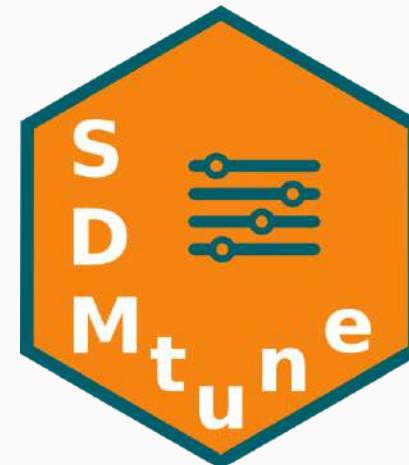


[Fletcher & Fortin \(2018\)](#)

# 5.5 Principais pacotes no R

Diversos pacotes implementados para ajustes de SDMs no R

- [dismo](#)
- [biomod2](#)
- [sdm](#)
- [ENMTML](#)
- [modelR](#)
- [SDMtune](#)
- [ENMeval](#)
- [wallace](#)
- [ssdm](#)
- [ShinyBIOMOD](#)



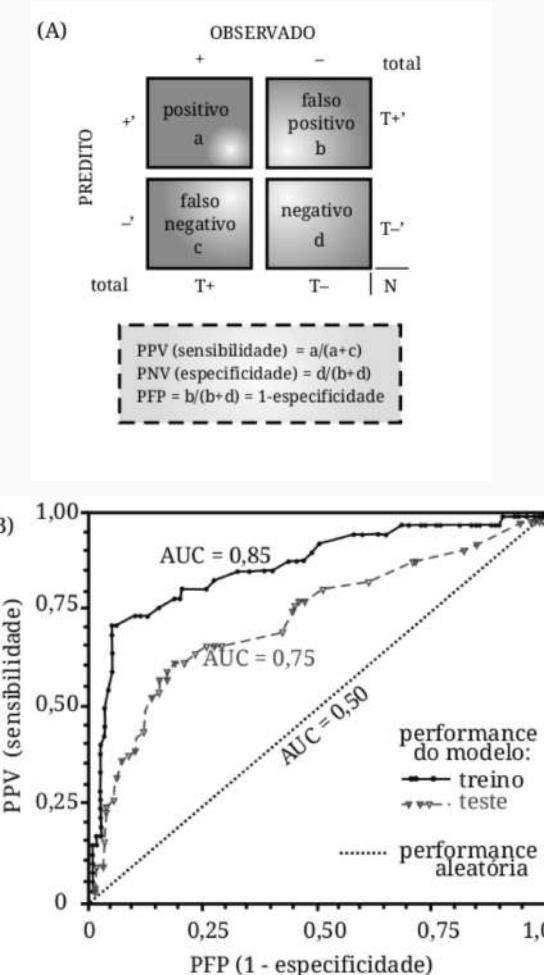
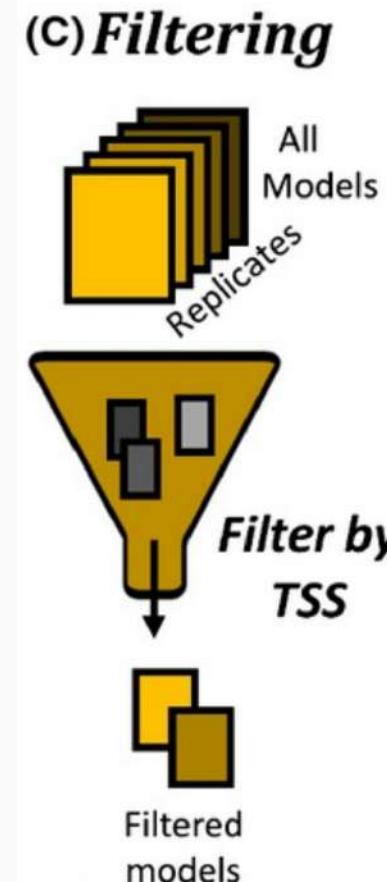
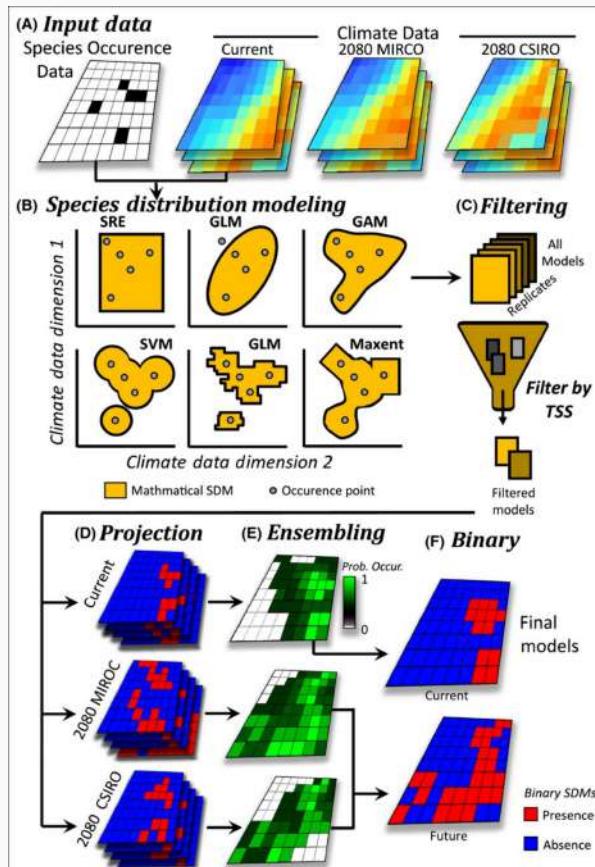
[Lista de pacotes no R](#)

Uma vez ajustados nossos modelos, devemos avaliar seu **desempenho** antes de fazer as previsões para todo o Espaço Geográfico (G)

**Desempenho:** esse modelo aí é melhor que jogar uma moeda (0 ou 1) para cada pixel?

# 2.4 Padronização para criação dos modelos

## 4. Avaliação (do ajuste) dos modelos



[Brown & Yoder \(2015\)](#), [Lima-Ribeiro & Diniz-Filho \(2013\)](#)

# 6 Avaliação dos modelos

# 6.1 Partição dos dados (treino e teste)

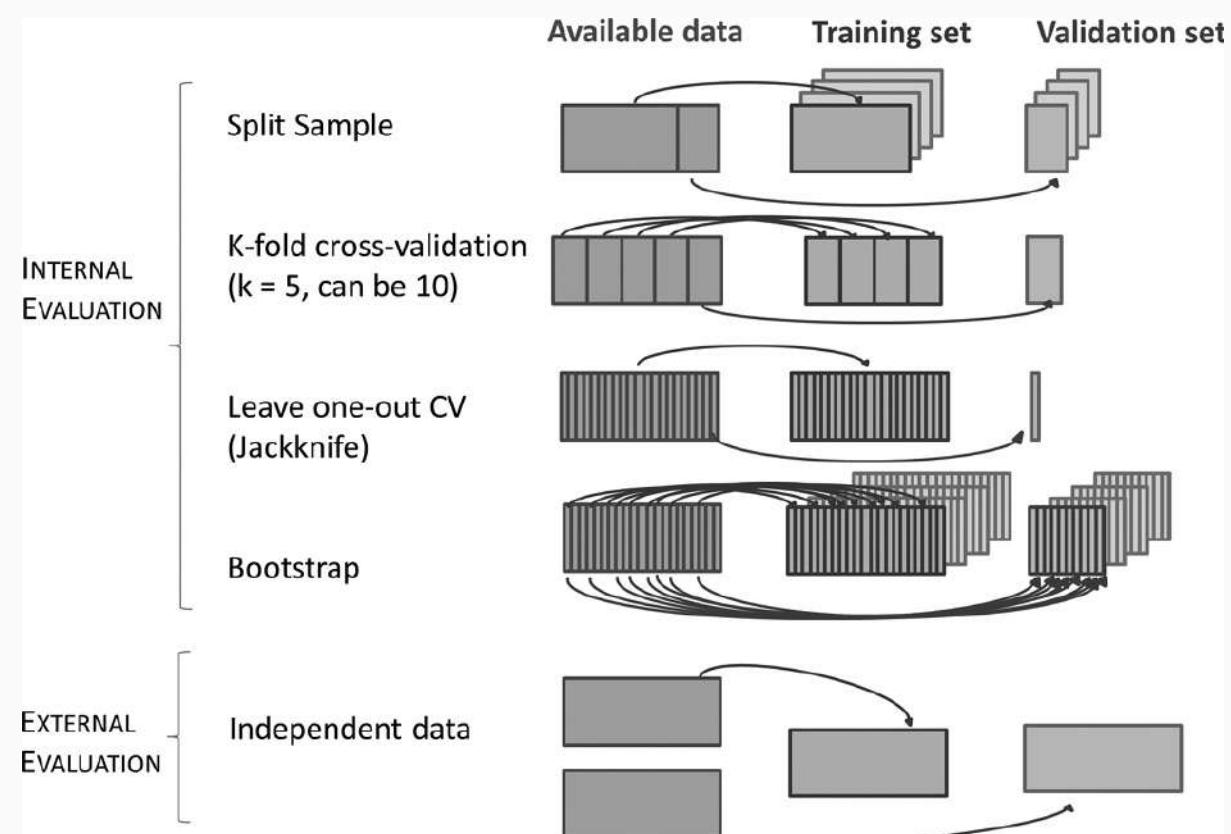
## Partição dos dados de presença (ou pseudo-ausência)

### Treino (flecha superior)

- Dados usados para **ajustar** o modelo

### Teste (flecha inferior)

- Dados usados para **avaliar** o modelo



[Guisan et al. \(2017\)](#)

# 6.1 Partição dos dados (treino e teste)

## Partição dos dados de presença (ou pseudo-ausência)

### Treino (flecha superior)

- Dados usados para **ajustar** o modelo

### Teste (flecha inferior)

- Dados usados para **avaliar** o modelo

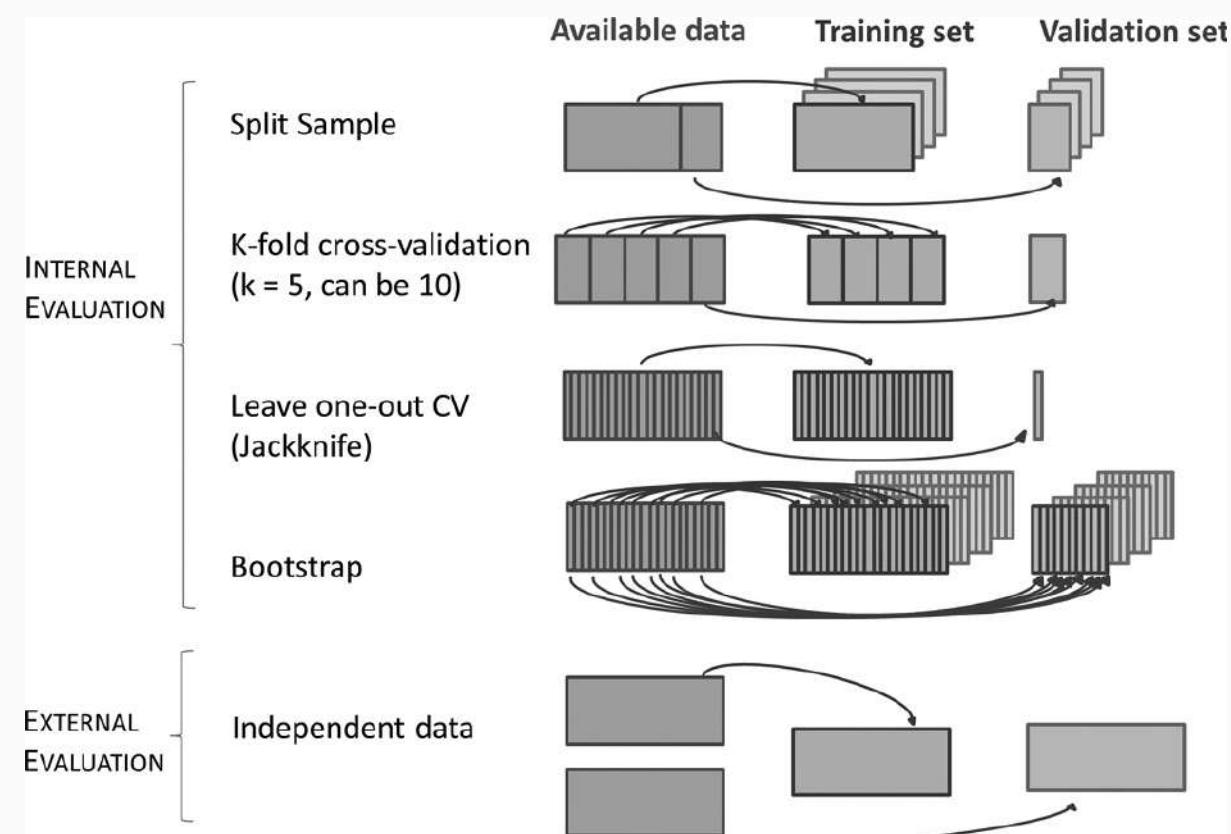
### Tipos de avaliação

- Interna**

Partição dos mesmos dados em treino e teste

- Externa**

Dados de teste não entram no treino



# 6.1 Partição dos dados (treino e teste)

## Tipos de partições para avaliação interna

- **Split sample**

Divide aleatoriamente as ocorrências em dois conjuntos, geralmente 70% para treino e 30% para teste e faz repetições aleatórias desses conjuntos

- **K-fold cross-validation**

Divide as ocorrências em conjuntos iguais ( $k$ ), um dos conjuntos vai para teste e os outros para treino, alternando até todas as partes irem para teste

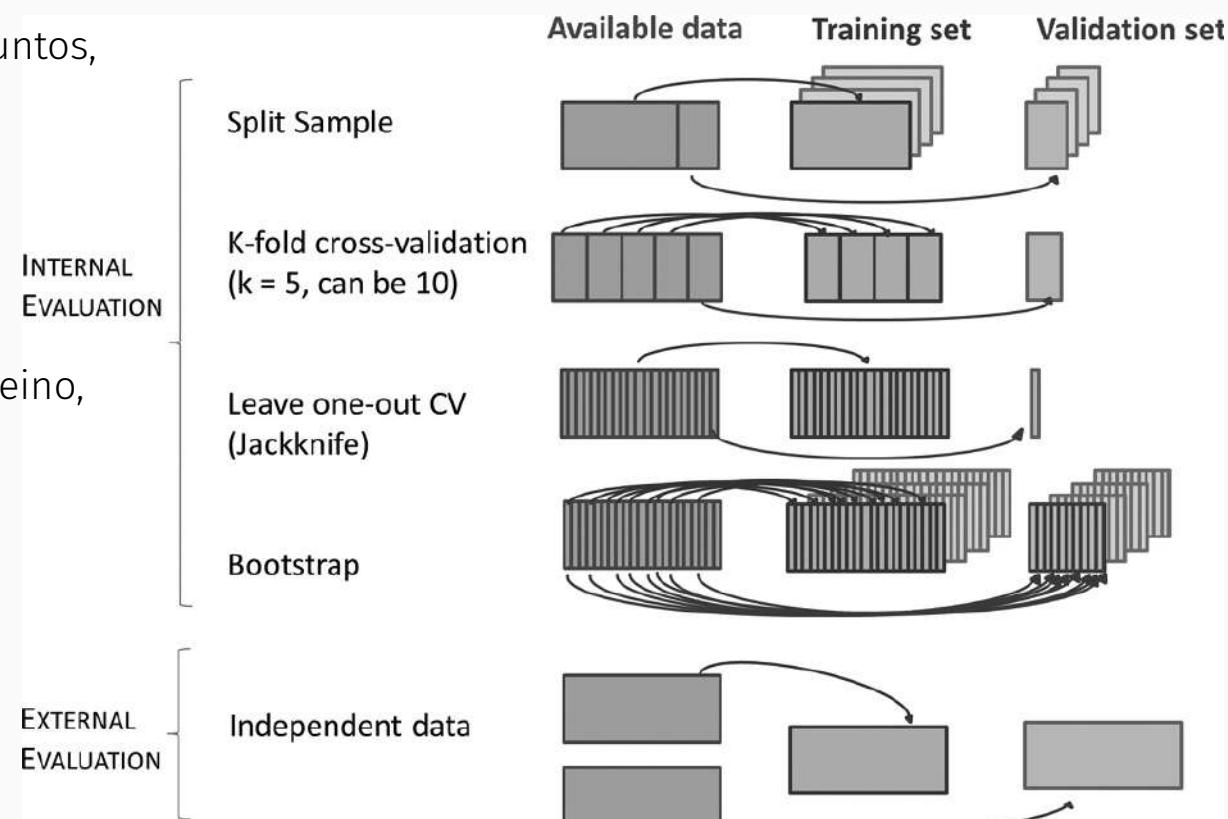
- **Leave one-out CV (Jackknife)**

Uma ocorrência para teste e o restante para treino, alternando até todas as ocorrências irem para teste

- **Bootstrap**

Reamostragem (mesmas ocorrências podem ir para treino e teste), usado mais para medir incertezas dos ajustes dos modelos

[Guisan et al. \(2017\)](#)

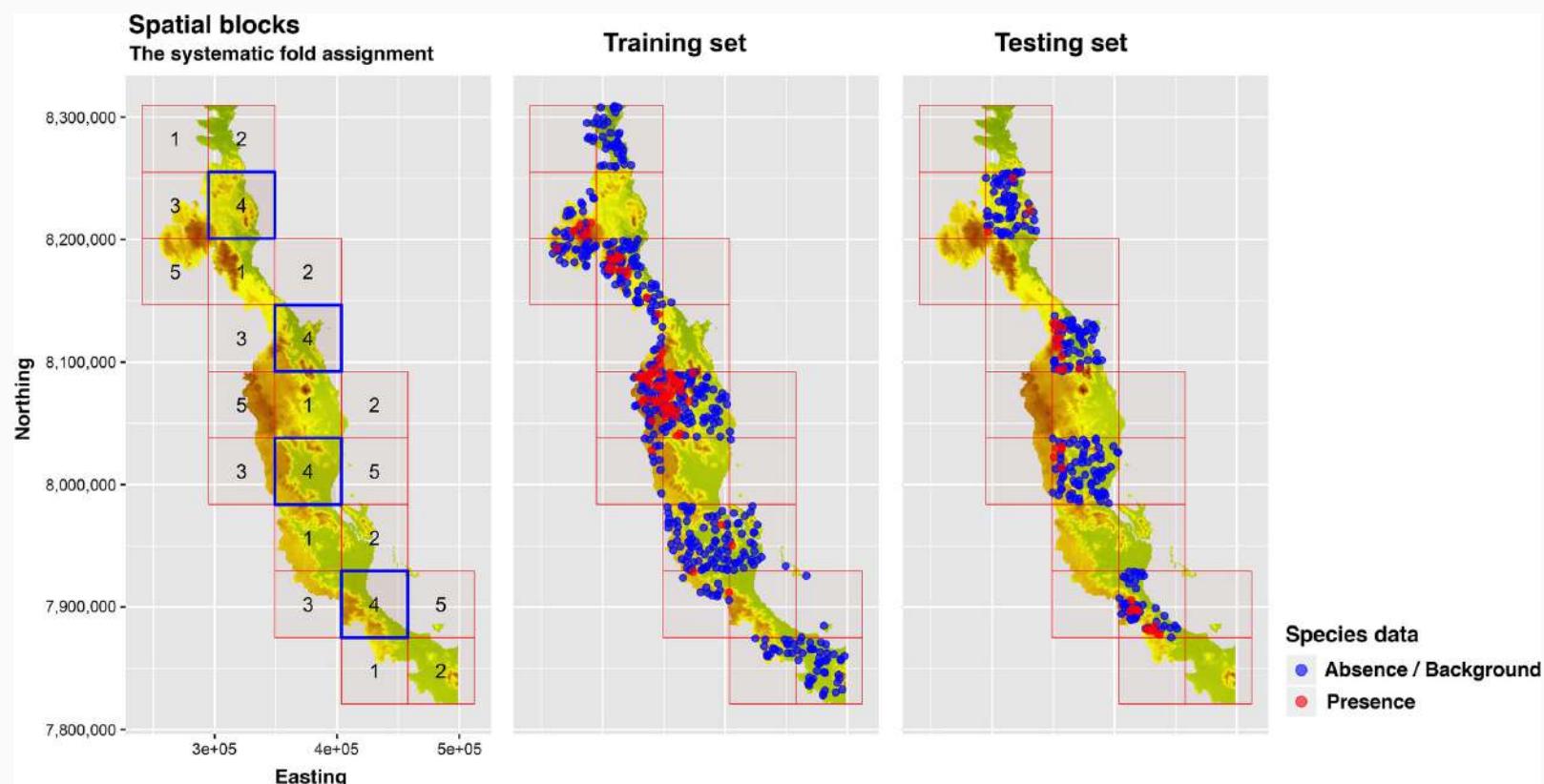


E como particionar esses dados?

# 6.1 Partição dos dados (treino e teste)

Tipos de partições para avaliação interna

Geralmente aleatória, mas pode ser por regras geográficas (G) ou ambientais (E)



# 6.1 Partição dos dados (treino e teste)

## Tipos de partições para avaliação interna

- **Split sample**

Divide aleatoriamente as ocorrências em dois conjuntos, geralmente 70% para treino e 30% para teste e faz repetições aleatórias desses conjuntos

- **K-fold cross-validation**

Divide as ocorrências em conjuntos iguais ( $k$ ), um dos conjuntos vai para teste e as outras para treino, alternando até todas as partes irem para teste

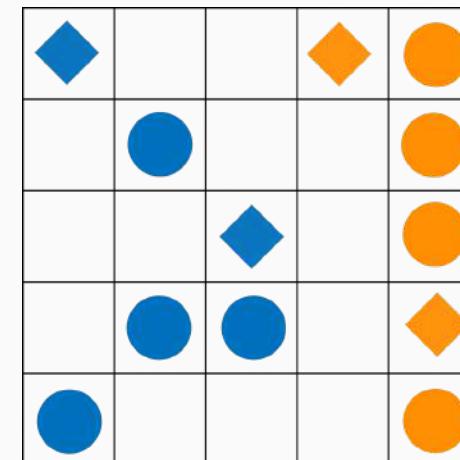
- **Leave one-out CV (Jackknife)**

Uma ocorrência para teste e o restante para treino, alterando até todas as ocorrências irem para teste

- **Bootstrap**

Reamostragem (mesmas ocorrências podem ir para treino e teste), usado mais para medir incertezas dos ajustes dos modelos

[Guisan et al. \(2017\)](#)



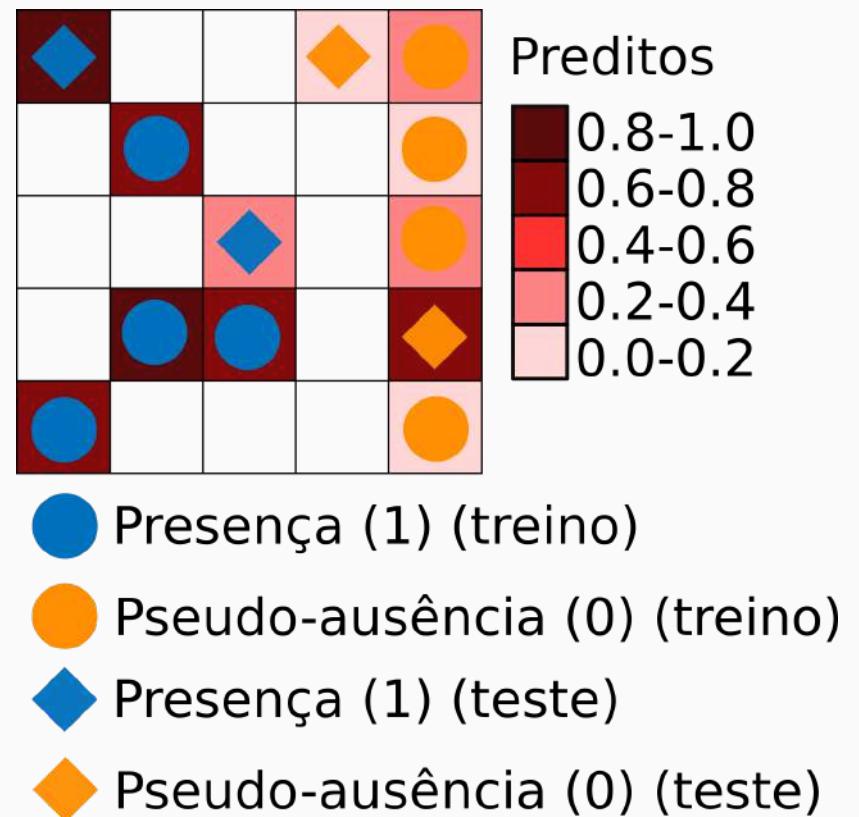
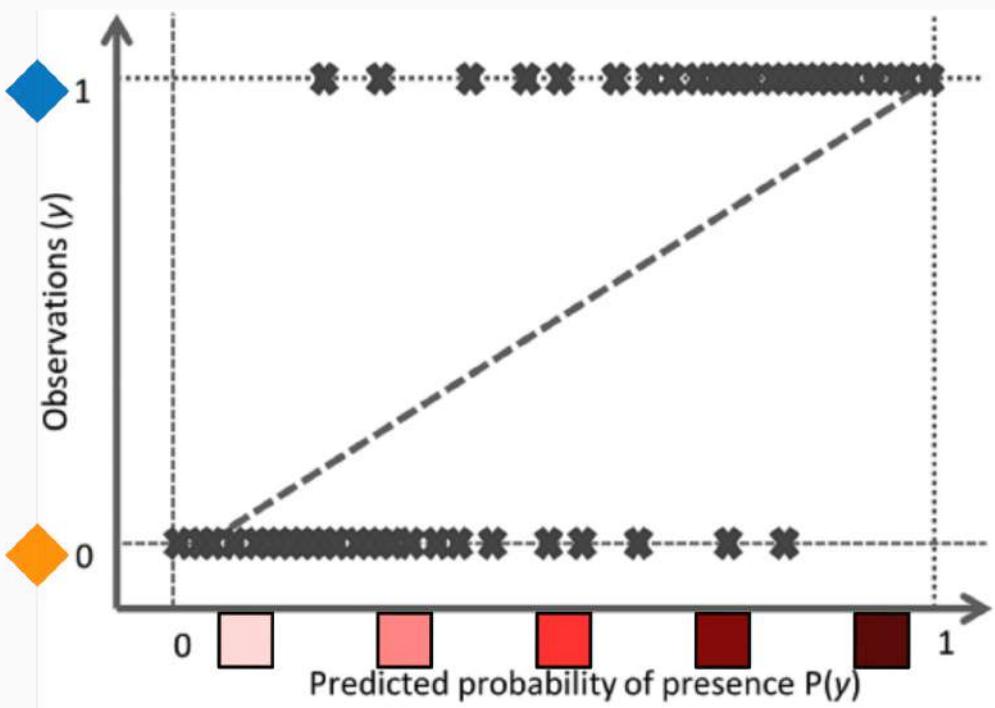
- Presença (treino)
- Pseudo-ausência (treino)
- ◆ Presença (teste)
- ◆ Pseudo-ausência (teste)

Vamos para o R preparar e particionar os nossos dados, e ajustar nossos modelos

Avaliar a **calibração dos modelos**: meu modelo acerta  
as ocorrências e (pseudo)ausências do conjunto de  
**teste?**

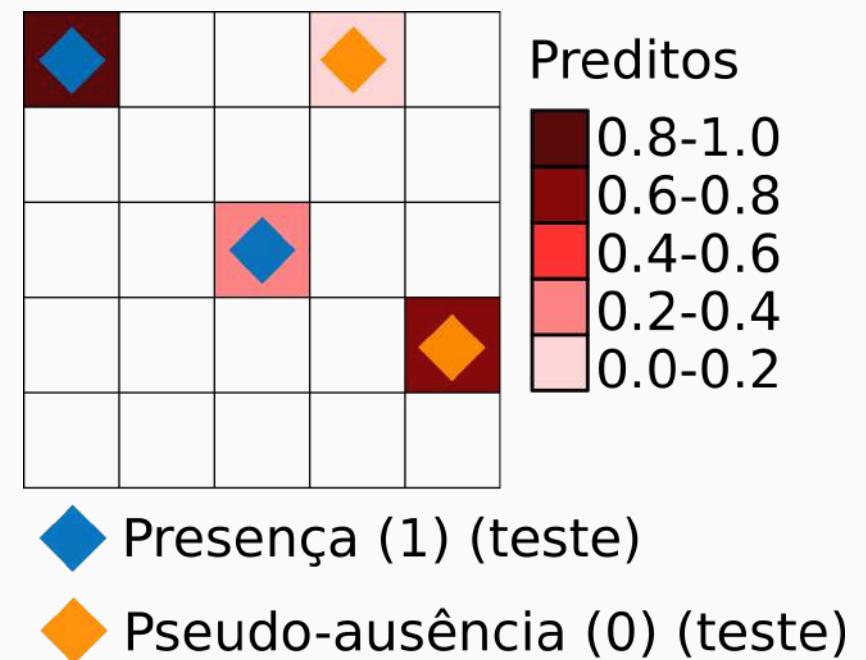
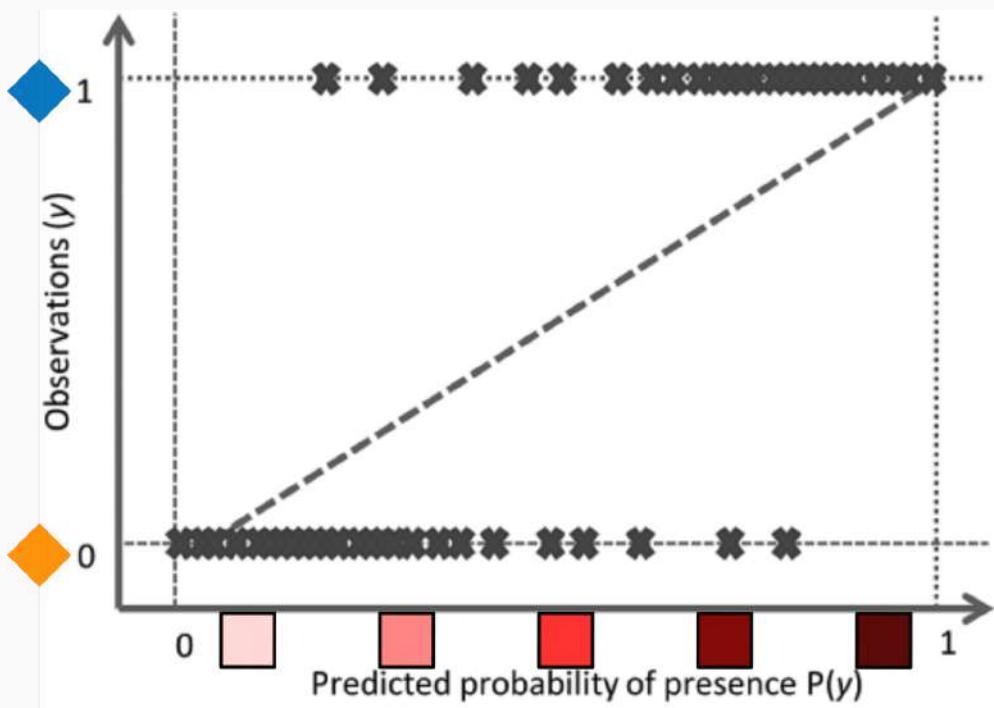
## 6.2 Calibração dos modelos

Valores observados vs valores preditos [treino e teste]



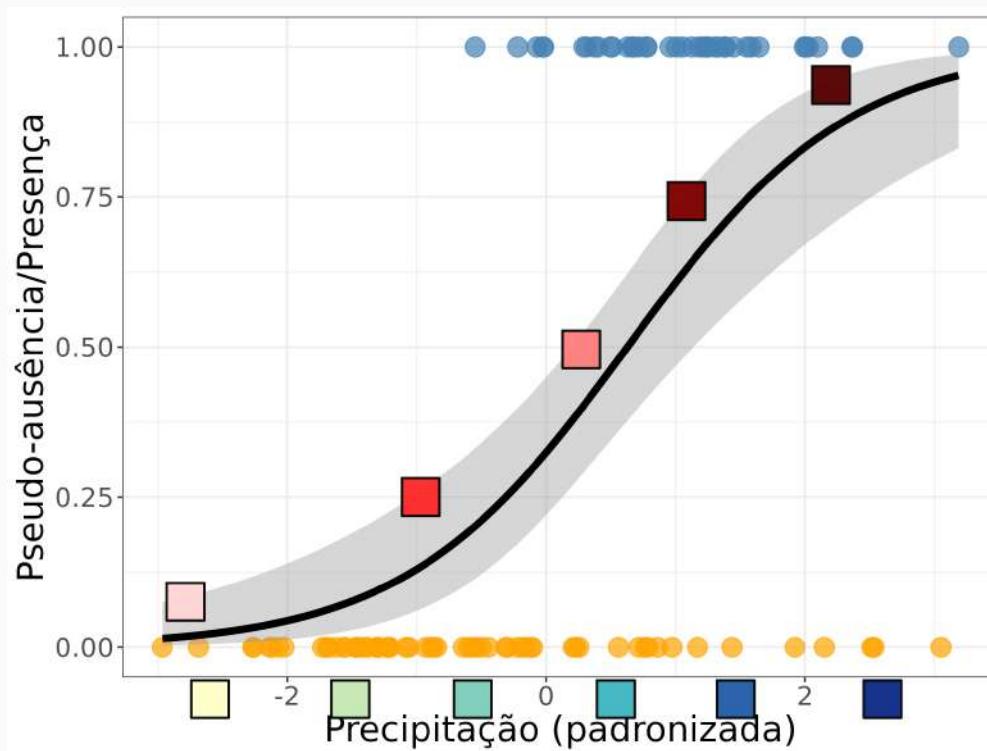
## 6.2 Calibração dos modelos

Valores observados vs valores preditos [teste]



## 6.2 Calibração dos modelos

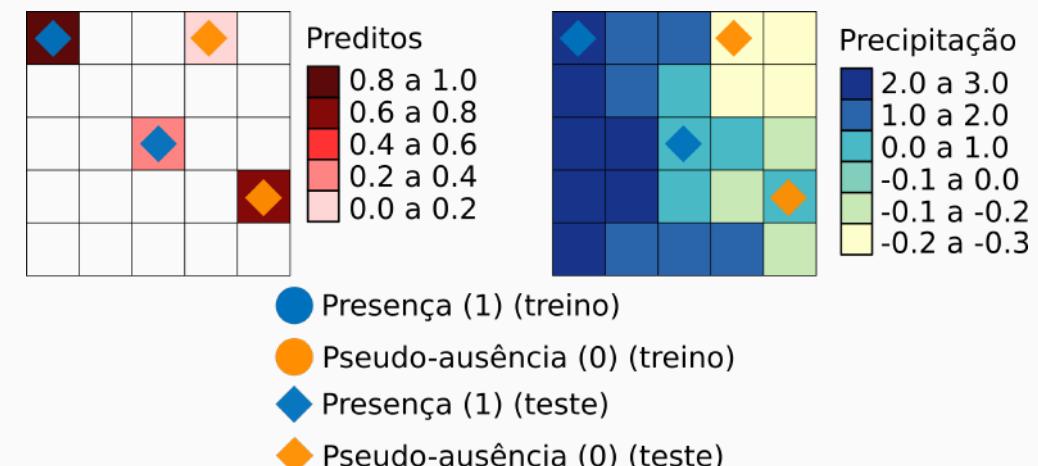
Valores preditos ( $P(x)$ ) a partir dos valores da variável preditora ( $x$ )



$$\text{Modelo : } P(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

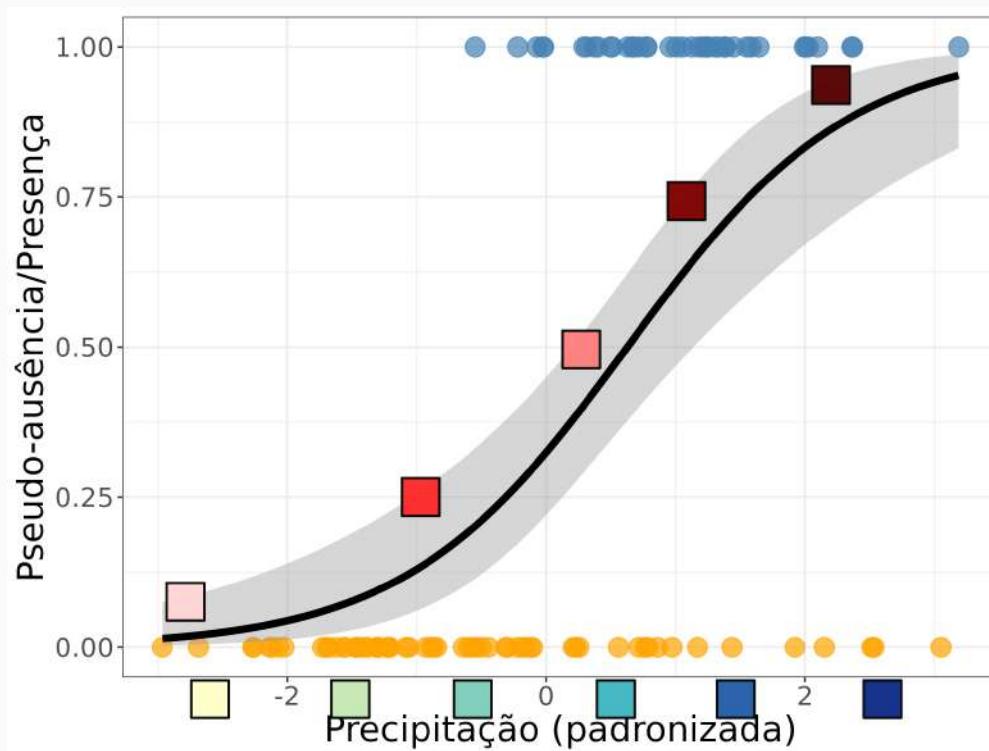
$$\text{Parametros : } \beta_1 = 1.169; \beta_0 = -0.733$$

$$\text{Modelo : } P(x) = \frac{1}{1 + e^{(-0.733 + 1.169x)}}$$



## 6.2 Calibração dos modelos

Valores preditos ( $P(x)$ ) a partir dos valores da variável preditora ( $x$ )



$$\text{Modelo : } P(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$$\text{Parametros : } \beta_1 = 1.169; \beta_0 = -0.733$$

$$\text{Modelo : } P(x) = \frac{1}{1 + e^{(-0.733 + 1.169x)}}$$

Seja  $x$  os valores de precipitação padronizados:

Para :  $x = -2; P(x) = 0.04431945; y = 0$

Para :  $x = -1; P(x) = 0.1298823; y = 0$

Para :  $x = 0; P(x) = 0.3245367; y = 0$

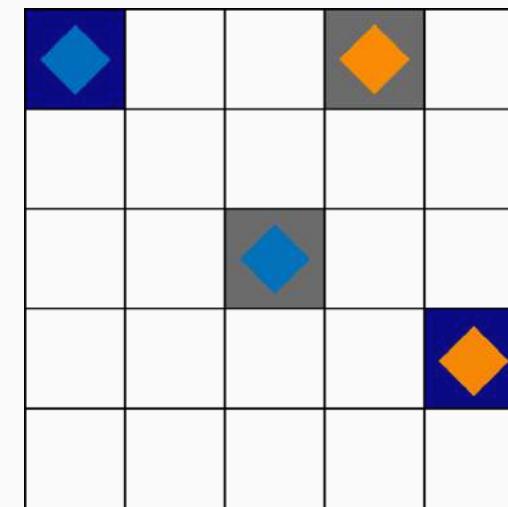
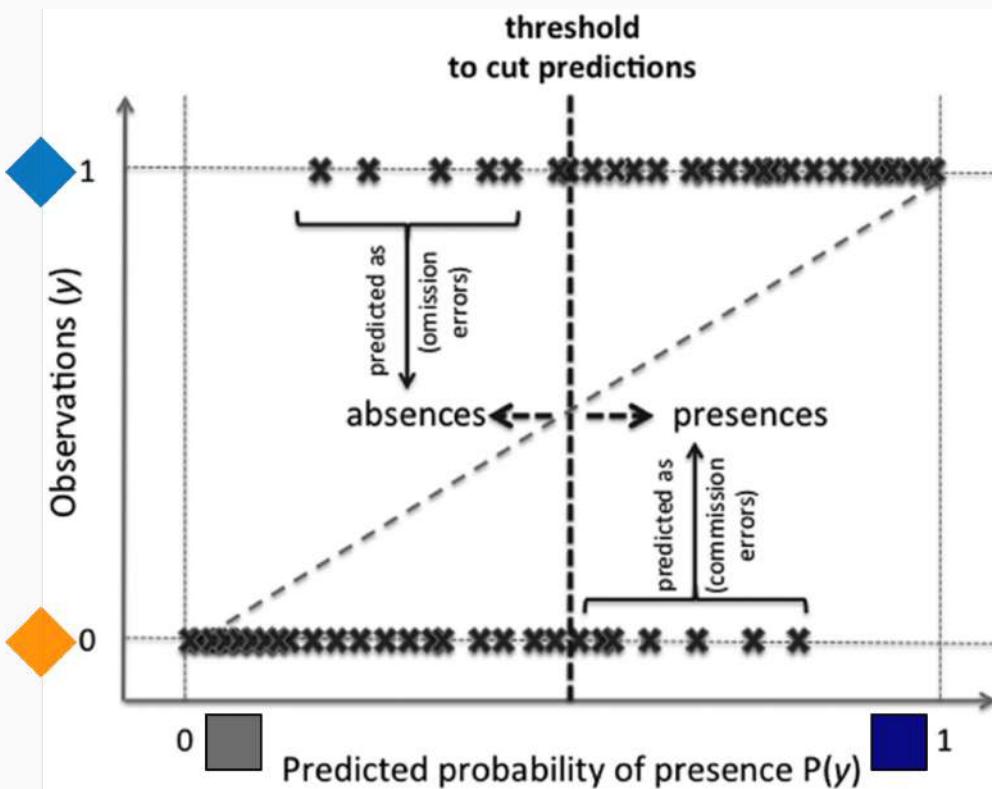
Para :  $x = 1; P(x) = 0.6073055; y = 1$

Para :  $x = 2; P(x) = 0.832716; y = 1$

Dado que a predição é contínua [0 a 1], vamos considerar algum **valor predito como limiar** (e.g. 0.5) para considerar se o **modelo acertou** as presenças e pseudo-ausências

## 6.3 Limiar de corte dos modelos

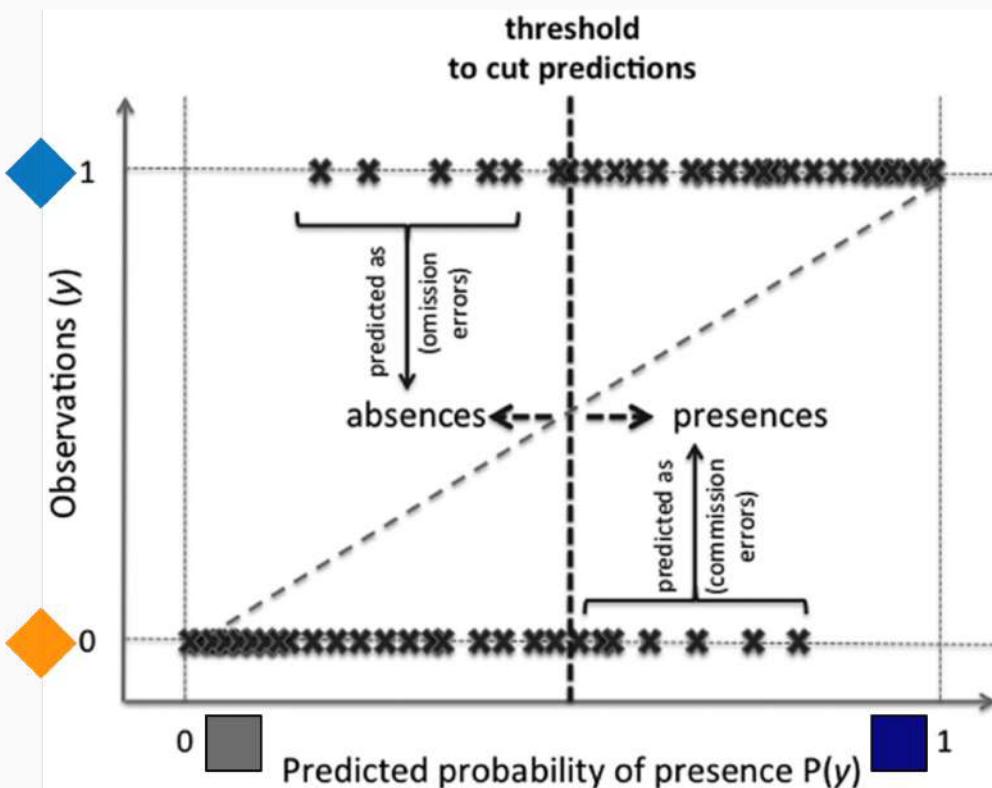
**Limiaria de corte (thresholds)** para considerar os acertos e erros do modelo



- ◆ Presença (1) (teste)
- ◆ Pseudo-ausência (0) (teste)

# 6.3 Limiar de corte dos modelos

## Tipos de erros



### Erro de omissão

A observação é de presença (1), mas a predição é de ausência (0) - **ausência falsa**

### Erro de comissão

A observação é de ausência (0), mas a predição é de presença (1) - **presença falsa**

# 6.4 Matriz de confusão

## Contabilizar os acertos e erros em presenças e pseudo-ausências

- Matriz de confusão

Matriz de confusão (2x2)		Observado		
Preditivo	Presença (1)	Presença (1)	Ausência (0)	Soma
	Ausência (0)	Presença verdadeira (TP)	Presença falsa (FP)	TP+FP
		Ausência falsa (FA)	Ausência verdadeira (TA)	FA+TA
	Soma	TP+FA	FP+TA	TP+FP+FA+TA

Matriz de confusão (2x2)		Observado		
Preditivo	Presença (1)	Presença (1)	Ausência (0)	Soma
	Ausência (0)			2
				2
	Soma	2	2	4


Preditos  
(thr = 0.5)

1  
 0

Presença (1) (teste)

Pseudo-ausência (0) (teste)

# 6.5 Métricas de avaliação

## Métricas que contabilizam os acertos e erros

- **Sensibilidade** (proporção de presenças verdadeiras - TPR)[0:1]
- **Especificidade** (proporção de ausências verdadeiras - TAR)[0:1]
- **AUC** (probabilidade de classificação correta)[0:1]
- **Kappa** (medida de concordância da classificação)[-1:1]
- **TSS** (compara o número de classificações corretas, menos aquelas atribuídas aleatoriamente) [-1:1]

		Observado		
		Presença (1)	Ausência (0)	Soma
Preditivo	Presença (1)	a	b	a+b
	Ausência (0)	c	d	c+d
	Soma	a+c	b+d	N

[Fielding & Bell \(1997\)](#)

Measure	Calculation
Prevalence	(a + c)/N
Overall diagnostic power	(b + d)/N
Correct classification rate	(a + d)/N
Sensitivity	a/(a + c)
Specificity	d/(b + d)
False positive rate	b/(b + d)
False negative rate	c/(a + c)
Positive predictive power (PPP)	a/(a + b)
Negative predictive power (NPP)	d/(c + d)
Misclassification rate	(b + c)/N
Odds-ratio	(ad)/(cb)
Kappa	$[(a + d) - (((a + c)(a + b) + (b + d)(c + d))/N)]/[N - (((a + c)(a + b) + (b + d)(c + d))/N)]$ $[-a.\ln(a)-b.\ln(b)-c.\ln(c)-d.\ln(d)+(a+b).\ln(a+b)+(c+d).\ln(c+d)]/[N.\ln N - ((a+c).\ln(a+c) + (b+d).\ln(b+d))]$
NMI n(s)	

# 6.5 Métricas de avaliação

## Métricas independentes de limiares

- **Sensibilidade** (proporção de presenças verdadeiras - TPR)[0:1]

$$\text{sensibilidade} = a/(a + c)$$

- **Especificidade** (proporção de ausências verdadeiras - TAR)[0:1]

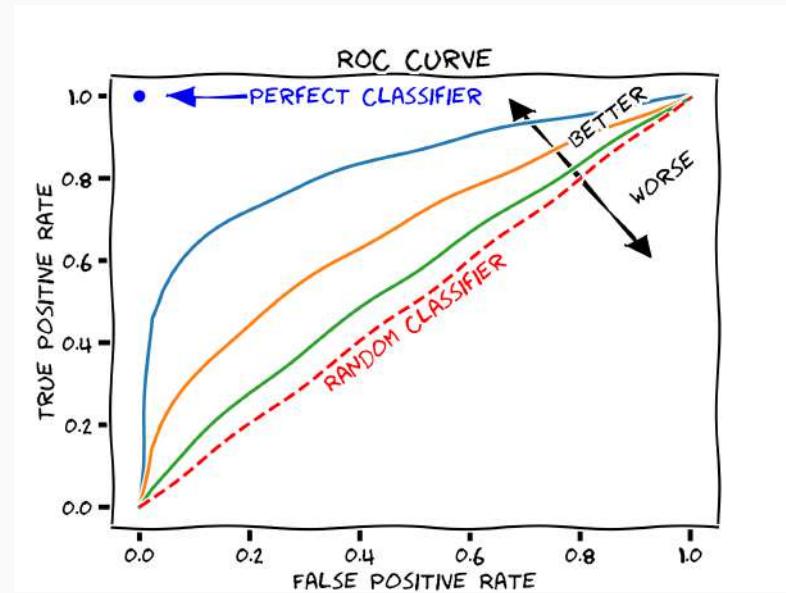
$$\text{especificidade} = d/(b + d)$$

- **1 - Especificidade** (proporção de presenças falsas - FPR)[0:1]

$$1 - \text{especificidade} = b/(b + d)$$

- **ROC** (sensibilidade ~ (1 - especificidade))

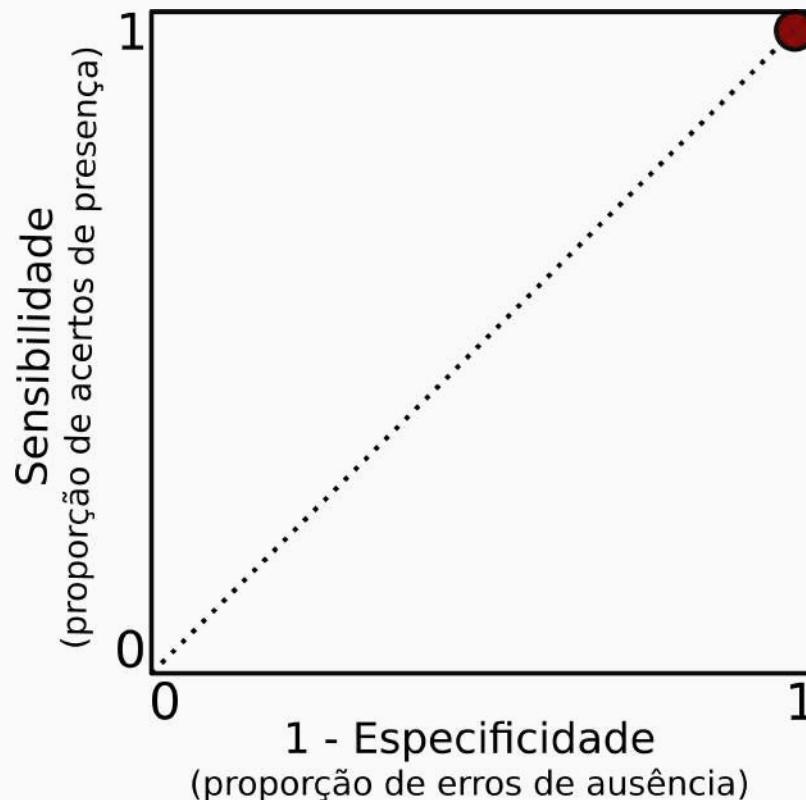
- **AUC** (probabilidade de classificação correta)[0:1]



		<b>Observado</b>		
		Presença (1)	Ausência (0)	Soma
<b>Preditó</b>	Presença (1)	a	b	a+b
	Ausência (0)	c	d	c+d
	Soma	a+c	b+d	N

## 6.5 Métricas de avaliação

ROC (Característica de Operação do Receptor - *Receiver Operating Characteristic*)



Limiar	Sensibilidade	1-Especificidade	Ponto
0.00	1	1	●

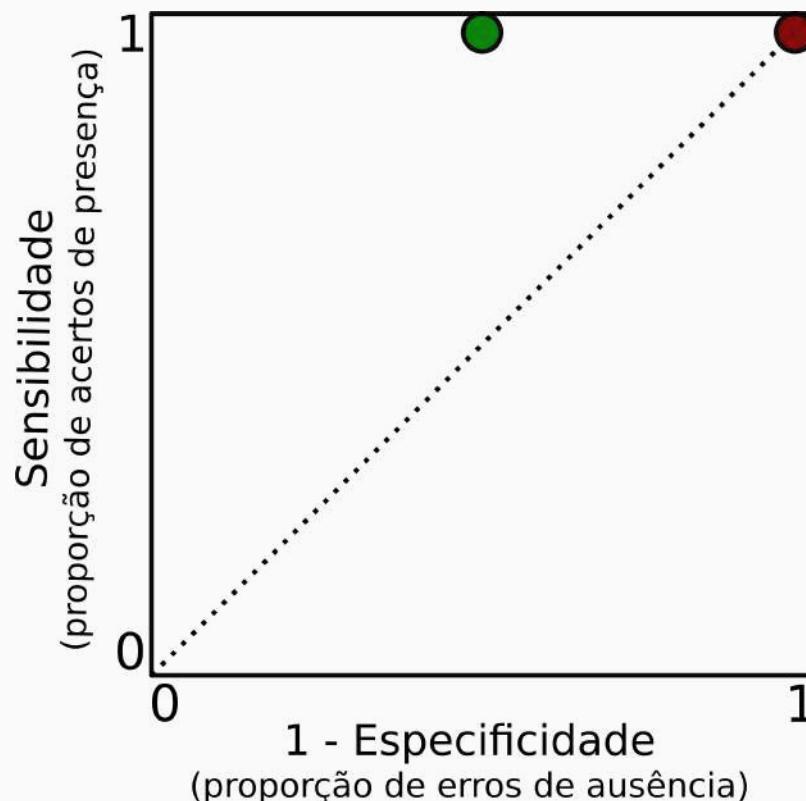
Limiar: 0.0  
Sensibilidade: 1  
1-Especificidade: 1

		Observado		
		Presença (1)	Ausência (0)	Soma
Predito	Presença (1)	2	2	4
	Ausência (0)	0	0	0
	Soma	2	2	4

Preditos (thr = 0.00)  
■ 1  
■ 0  
◆ Presença (1)  
◆ Pseudo-ausência (0)

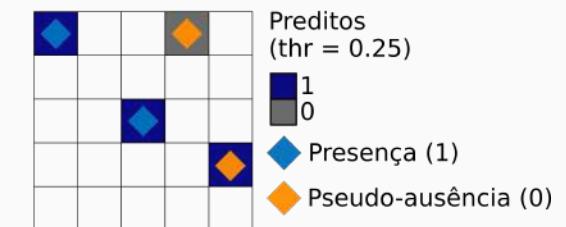
## 6.5 Métricas de avaliação

ROC (Característica de Operação do Receptor - *Receiver Operating Characteristic*)



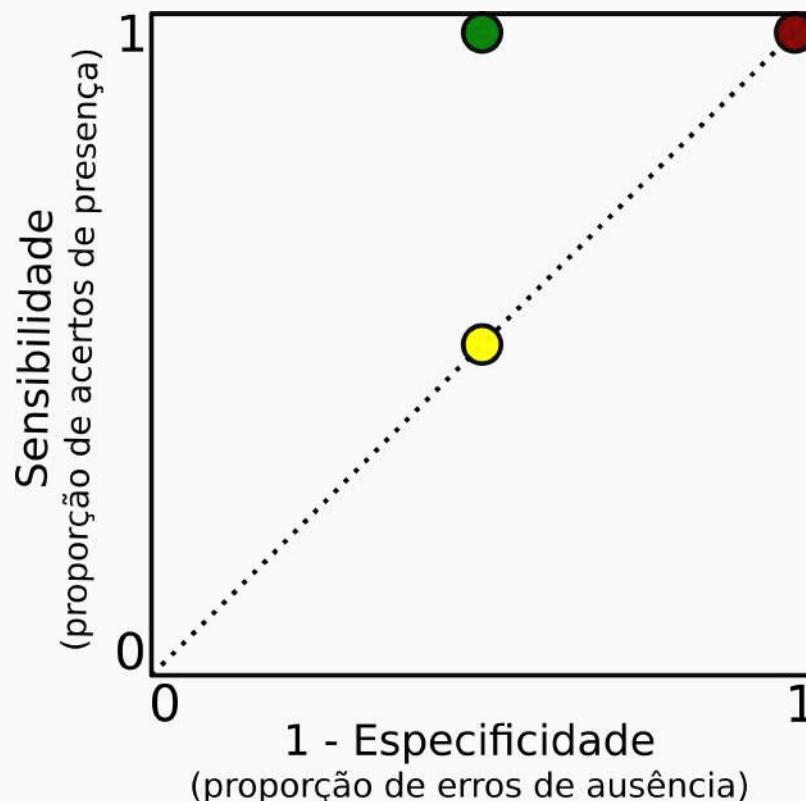
Limiar	Sensibilidade	1-Especificidade	Ponto
0.25	1	0.5	● (green)
0.00	1	1	● (red)

		Observado		
		Presença (1)	Ausência (0)	Soma
Predito	Presença (1)	2	1	3
	Ausência (0)	0	1	1
	Soma	2	2	4



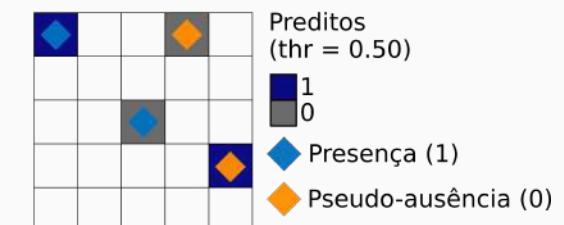
## 6.5 Métricas de avaliação

ROC (Característica de Operação do Receptor - *Receiver Operating Characteristic*)



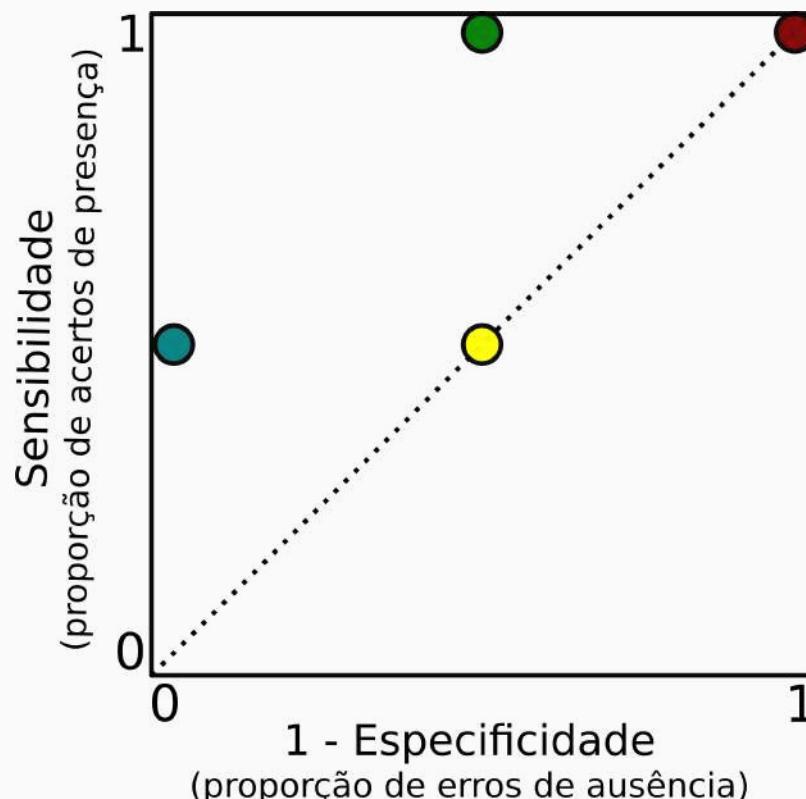
Limiar	Sensibilidade	1-Especificidade	Ponto
0.50	0.5	0.5	Yellow point
0.25	1	0.5	Green point
0.00	1	1	Red point

		Observado		
		Presença (1)	Ausência (0)	Soma
Predito	Presença (1)	1	1	2
	Ausência (0)	1	1	2
	Soma	2	2	4



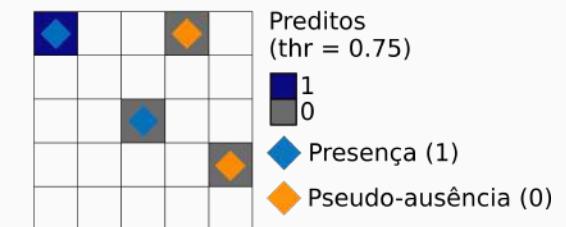
## 6.5 Métricas de avaliação

ROC (Característica de Operação do Receptor - *Receiver Operating Characteristic*)



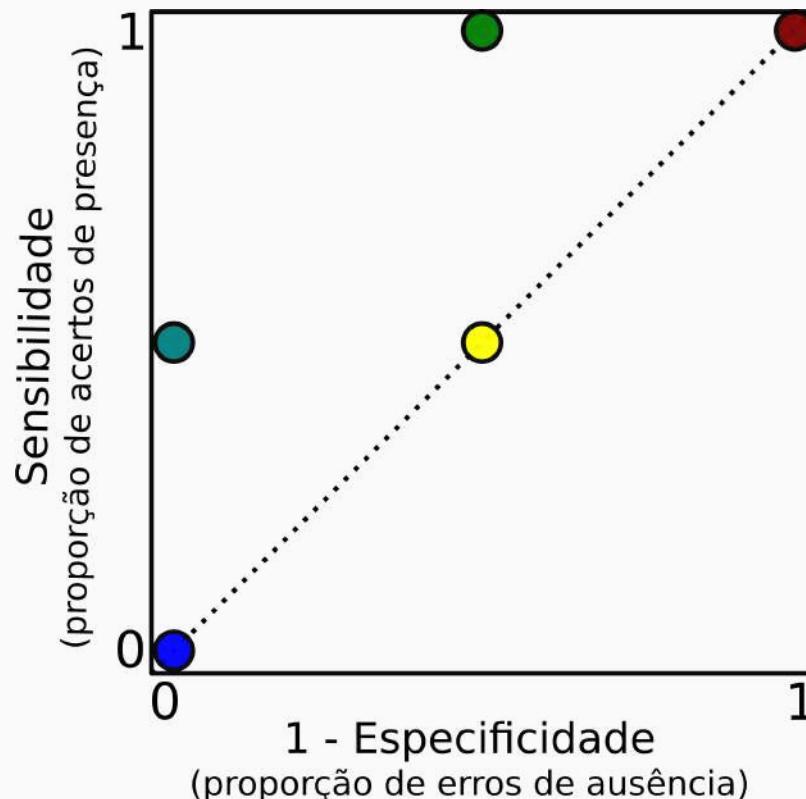
Limiar	Sensibilidade	1-Especificidade	Ponto
0.75	0.5	0	● (teal)
0.50	0.5	0.5	● (yellow)
0.25	1	0.5	● (green)
0.00	1	1	● (red)

		Observado		
		Presença (1)	Ausência (0)	Soma
Preditos	Presença (1)	1	0	1
	Ausência (0)	1	2	3
	Soma	2	2	4



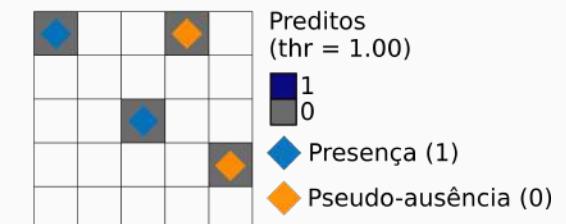
## 6.5 Métricas de avaliação

ROC (Característica de Operação do Receptor - *Receiver Operating Characteristic*)



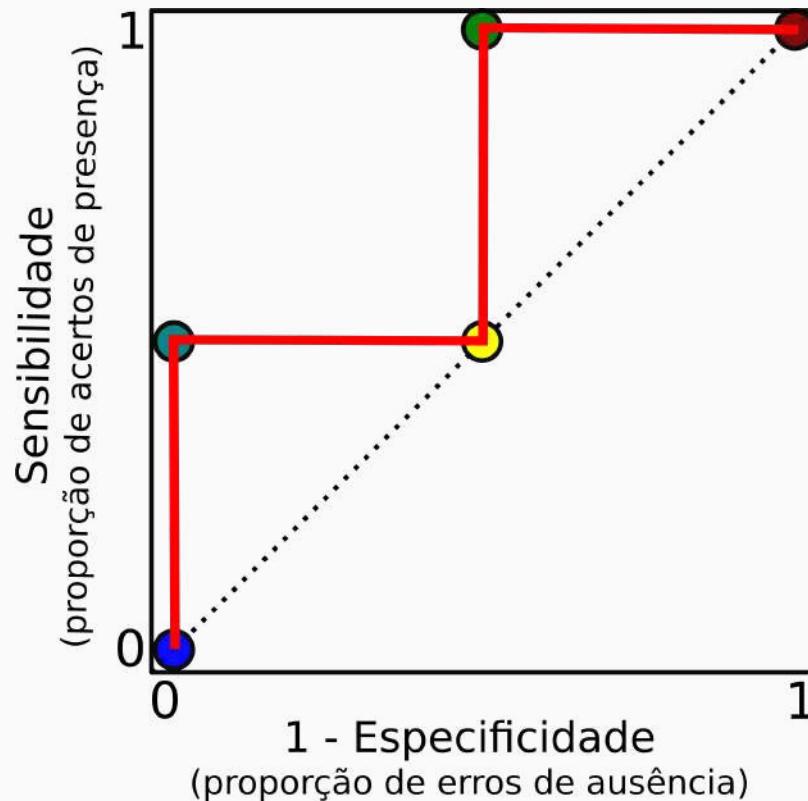
Limiar	Sensibilidade	1-Especificidade	Ponto
1.00	0	0	● (blue)
0.75	0.5	0	● (teal)
0.50	0.5	0.5	● (yellow)
0.25	1	0.5	● (green)
0.00	1	1	● (red)

		Observado		
		Presença (1)	Ausência (0)	Soma
Preditos	Presença (1)	0	0	0
	Ausência (0)	2	2	4
	Soma	2	2	4

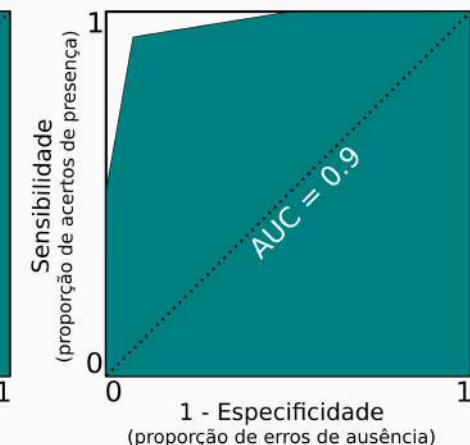
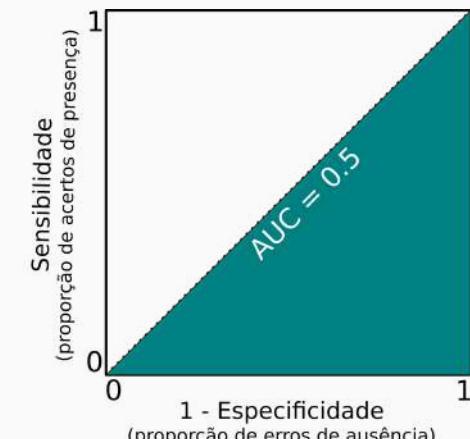
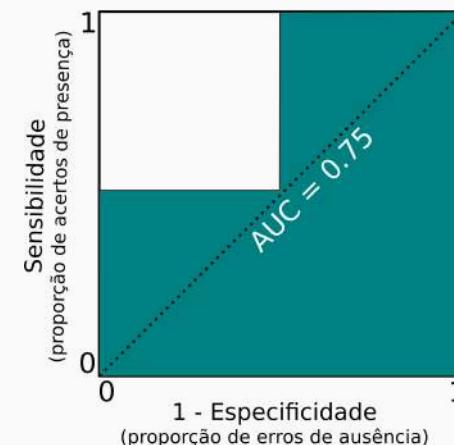


# 6.5 Métricas de avaliação

## AUC (Área Abaixo da Curva - Area Under the Curve)

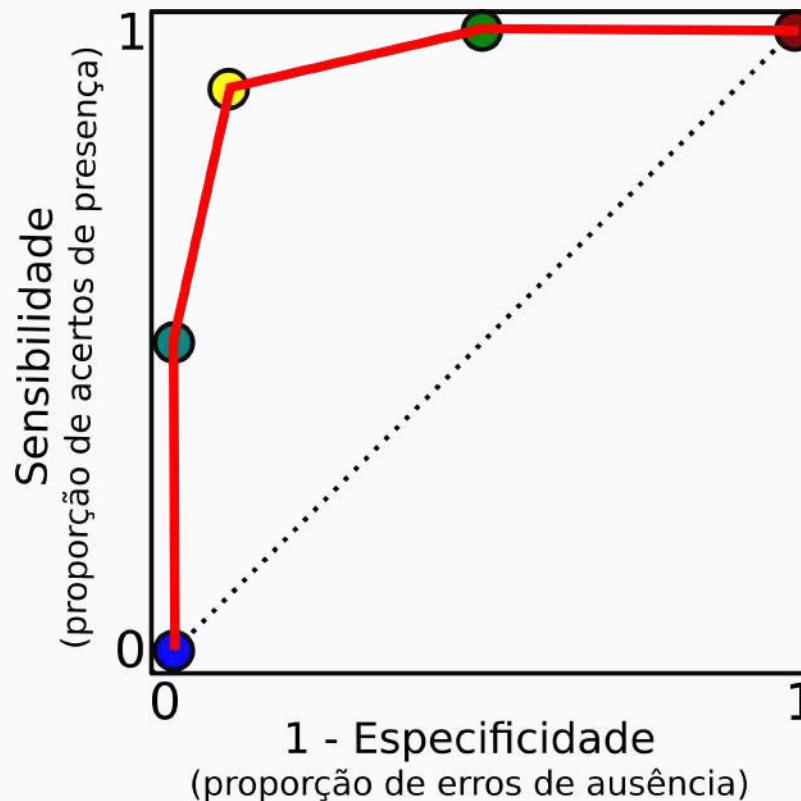


Valores de AUC	Interpretação
$AUC > 0.90$	Excelente
$0.80 < AUC < 0.90$	Bom
$0.70 < AUC < 0.80$	Razoável
$0.60 < AUC < 0.70$	Pobre
$0.50 < AUC < 0.60$	Falhou
$AUC < 0.50$	Contra-predições

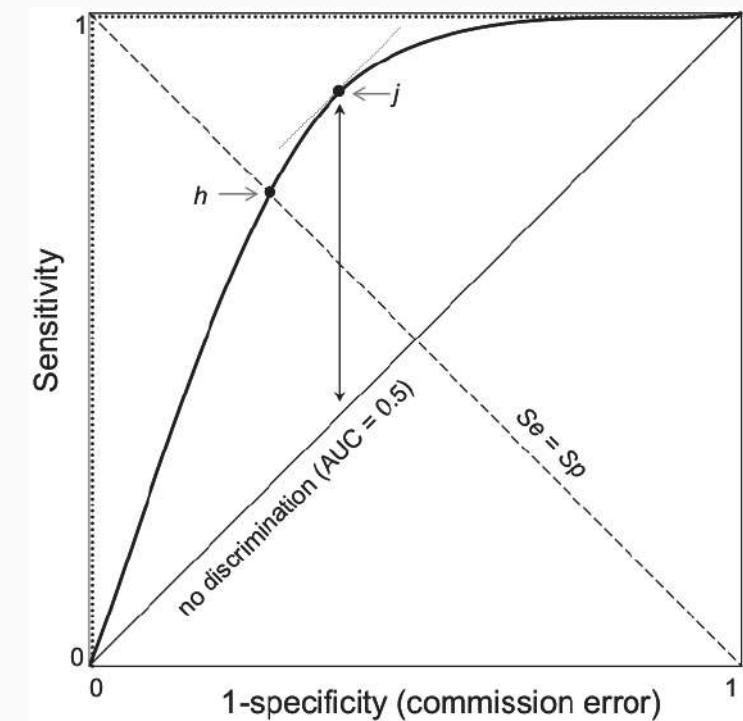


# 6.5 Métricas de avaliação

## Métricas dependentes de limiares



- Max-SSS  
Limiar que maximiza a soma da sensibilidade e especificidade
- LPT  
Limiar para o menor valor predito para as ocorrências



# 6.5 Métricas de avaliação

## Métricas dependentes de limiares

### Kappa ( $\kappa$ ) [-1 a +1]

$$\kappa = \frac{\frac{a+d}{N} - \frac{(a+b)(a+c)+(c+d)(b+d)}{N^2}}{1 - \frac{(a+b)(a+c)+(c+d)(b+d)}{N^2}} = \frac{P_o - P_e}{1 - P_e}, \text{ onde } P = \text{acurácia}$$

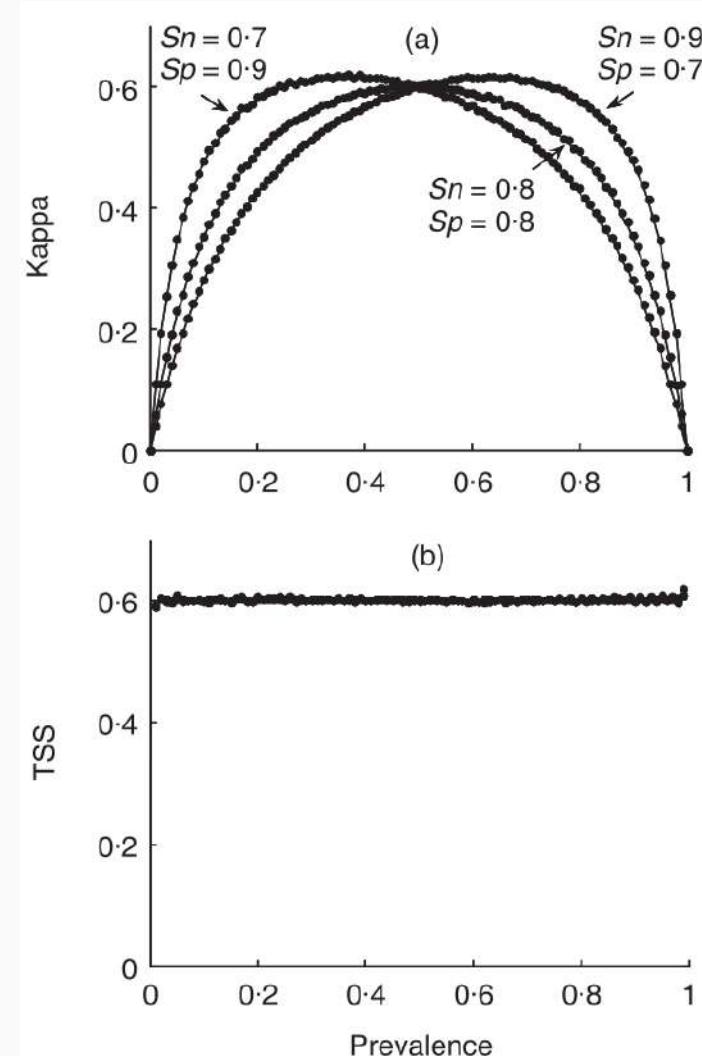
- $\kappa > 0.75$ : Excelente
- $0.40 > \kappa > 0.75$ : Bom
- $\kappa < 0.40$ : pobre

### TSS (*True Skill Statistic*) [-1 a +1]

$$TSS = \frac{ad-bc}{(a+c)(b+d)} = \text{Sensitividade} + \text{Especificidade} - 1$$

- $TSS \cong +1$ : concordância perfeita
- $TSS \leq 0$ : desempenho não melhor do que aleatório
- Indicativos:  $TSS > 0.5$ ,  $TSS > 0.7$ ,  $TSS > 0.8$  e  $TSS > 0.9$

[Landis & Koch \(1977\)](#), [Fielding & Bell \(1997\)](#), [Allouche et al. \(2006\)](#).

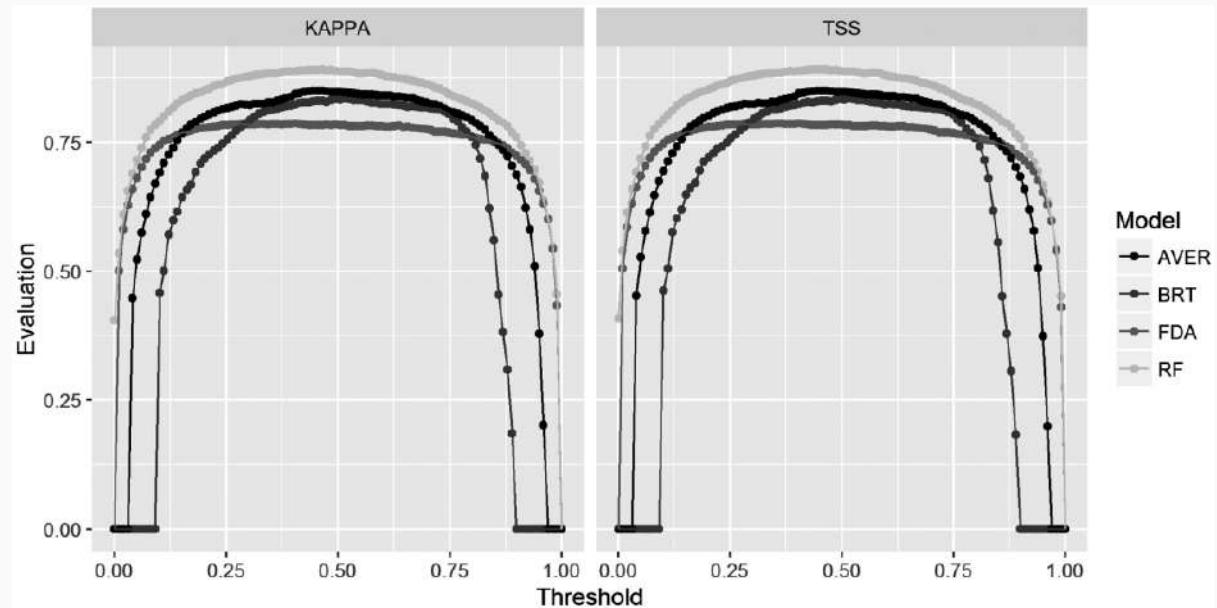


# 6.5 Métricas de avaliação

## Métricas dependentes de limiares

### Valores máximos de Kappa e TSS

- Calcular **vários valores das estatísticas** para diferentes liminares
- Liu et al. (2013) mostraram que o limiar que maximiza as estatísticas (max-TSS ou max-Kappa) é equivalente ao limiar de **máxima soma da sensibilidade e especificidade** (max-SSS)
- Essa estratégia gera o **mesmo limiar** para métodos de somente-presença, presença-pseudo-ausência e presença-background



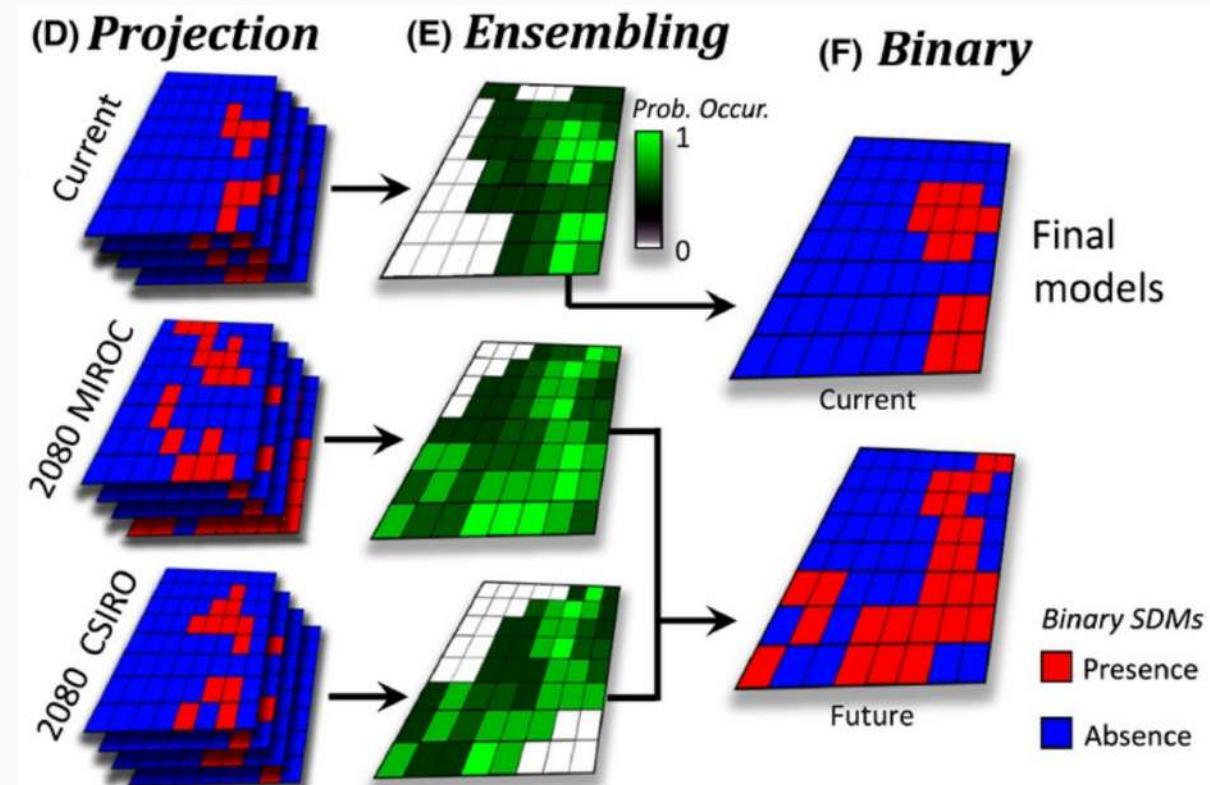
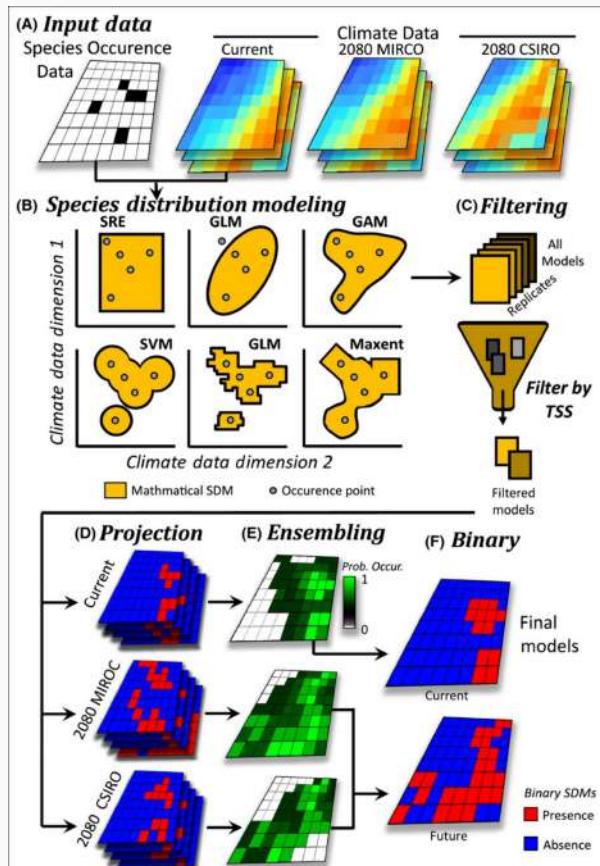
[Allouche et al. \(2006\)](#), [Liu et al. \(2013\)](#), [Guisan et al. \(2017\)](#),

Vamos para o R avaliar nossos modelos

Uma vez que **avaliamos a calibração dos modelos**  
(acertam as presenças e pseudo-ausências) podemos  
fazer a **predição** para áreas desconhecidas

# 2.4 Padronização para criação dos modelos

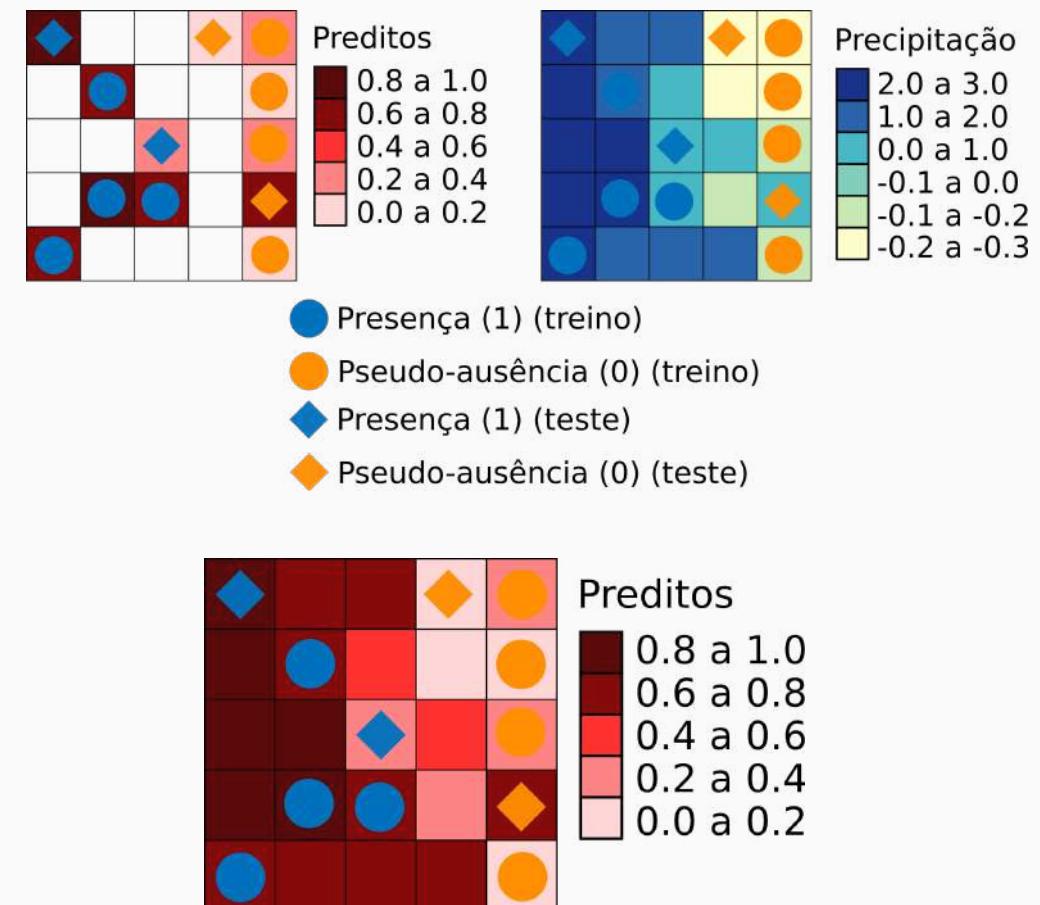
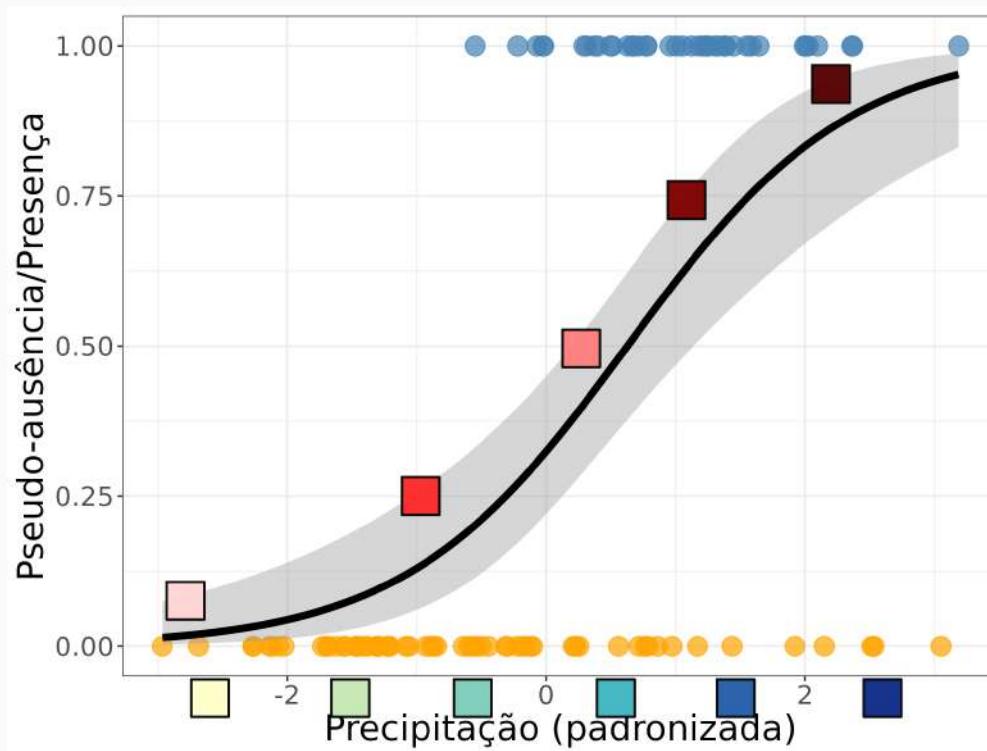
## 5. Predições no tempo e no espaço



# 7 Predição e consenso dos modelos

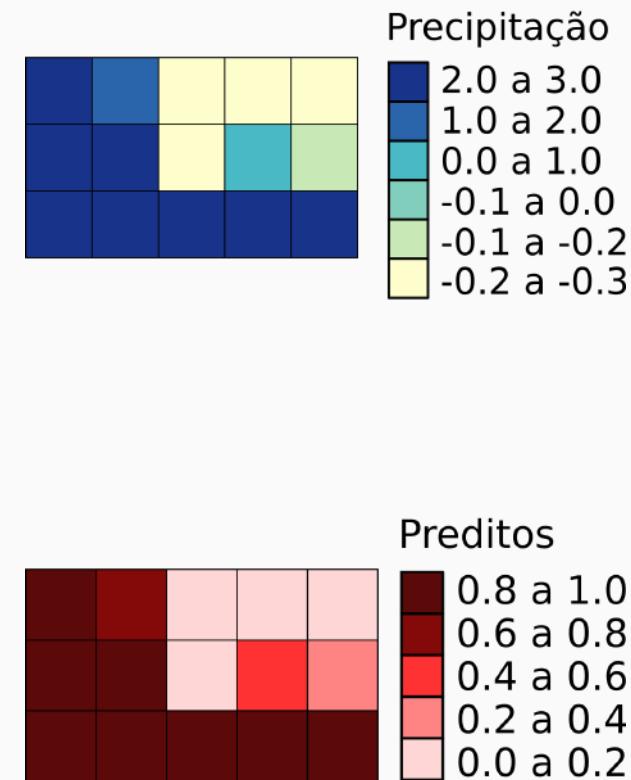
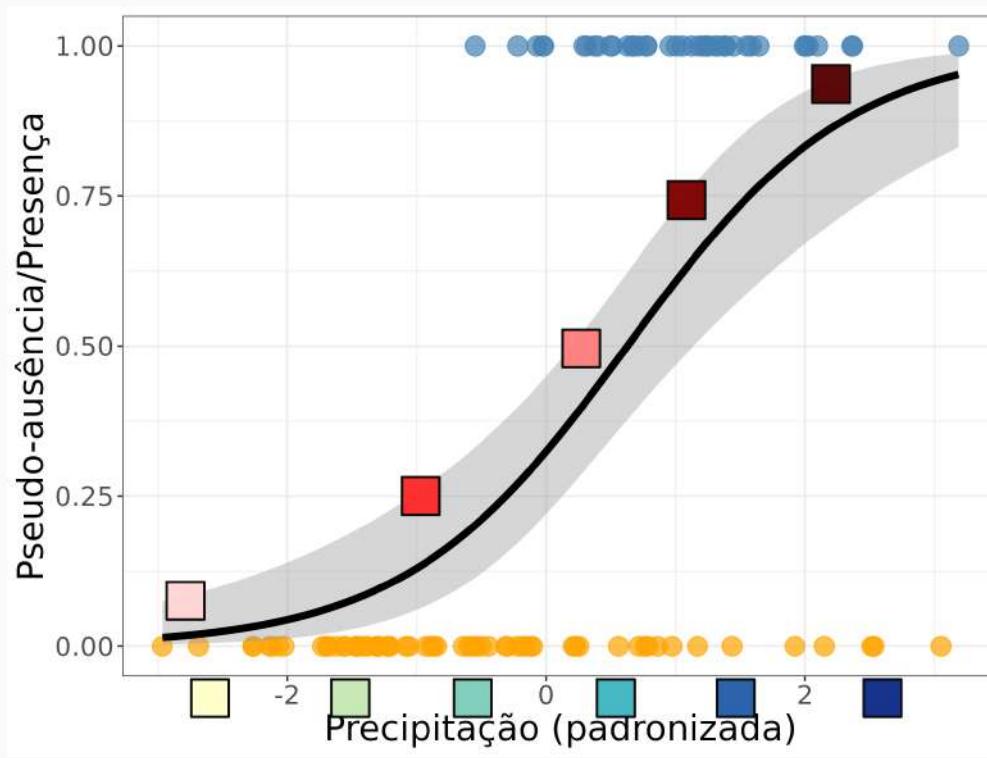
# 7.1 Predição dos modelos

## Predição para a área modelada



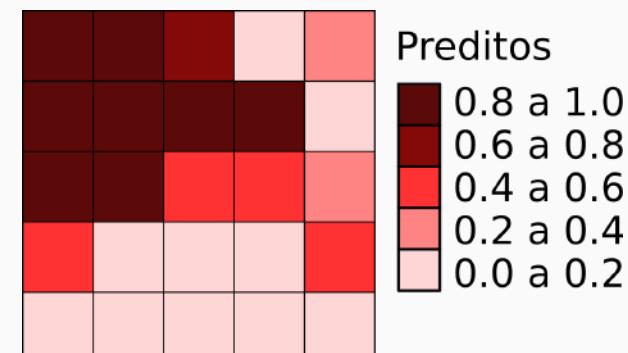
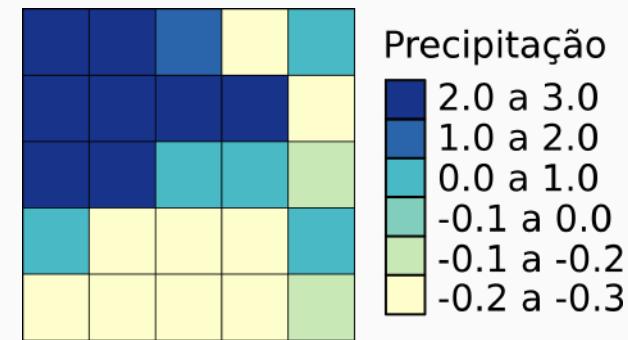
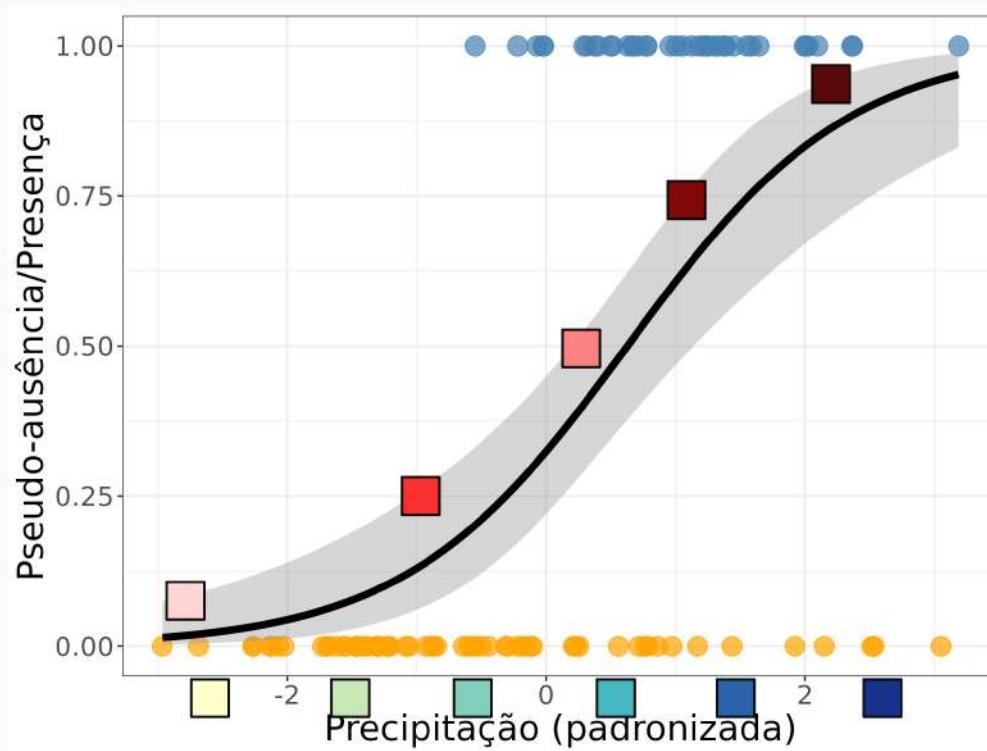
# 7.1 Predição dos modelos

## Predição para outras áreas (espaço)



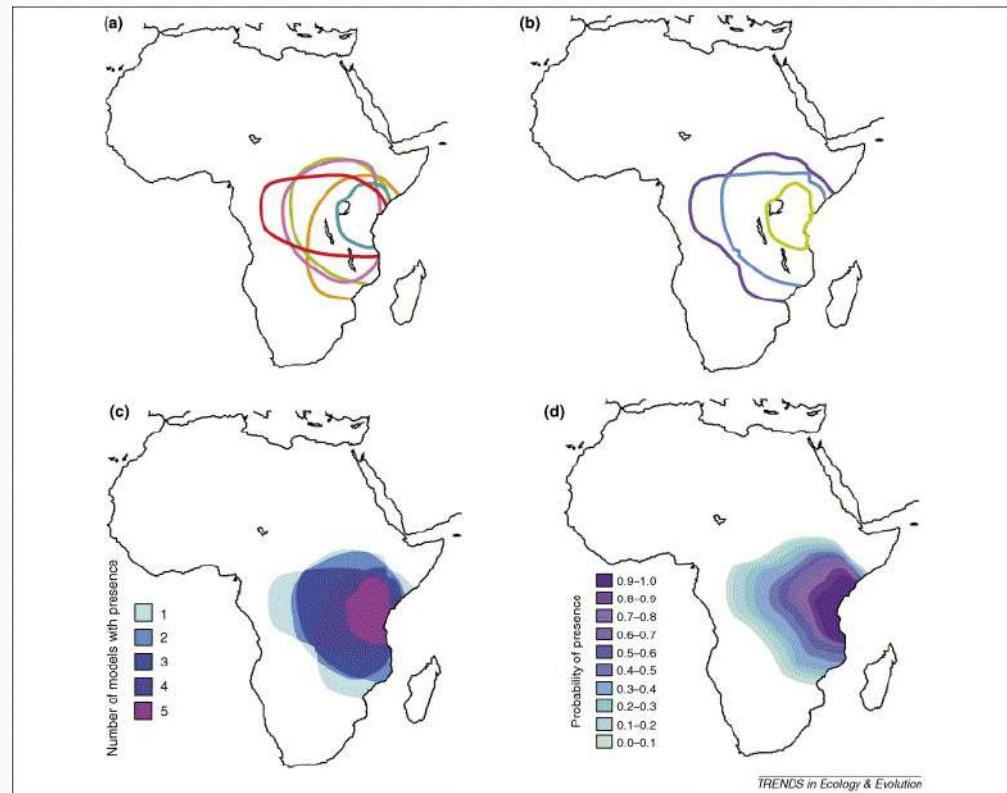
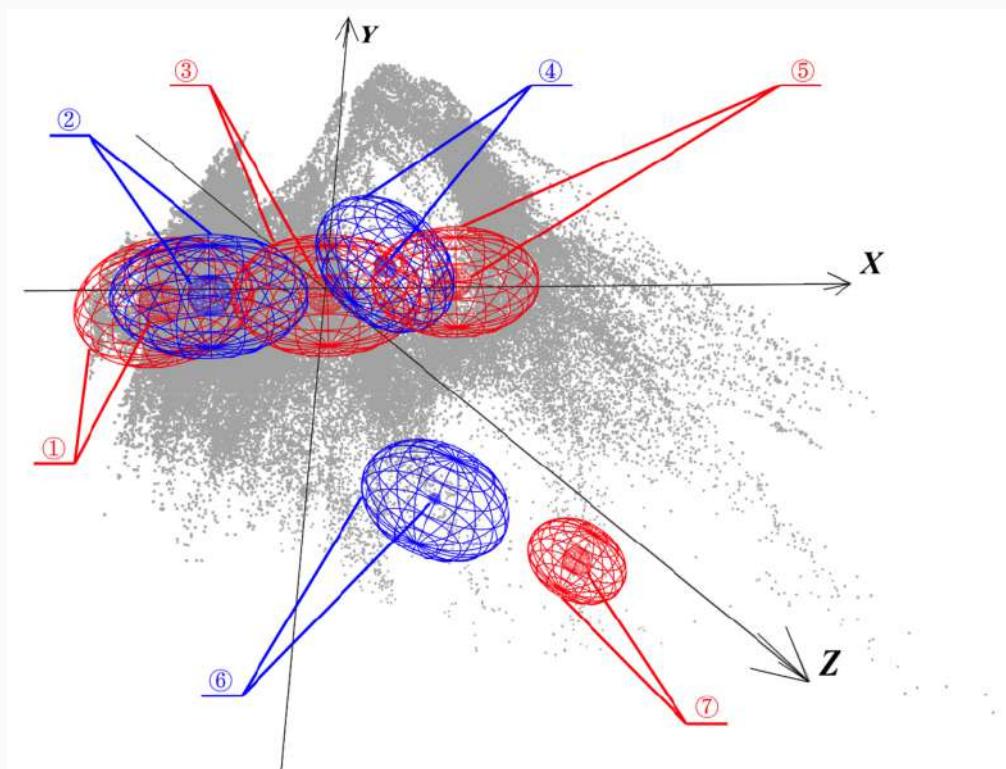
# 7.1 Predição dos modelos

Predição para outros períodos no passado ou no futuro (tempo)



## 7.2 Consenso dos modelos

Diferentes estimativas do nicho - diferentes previsões geográficas

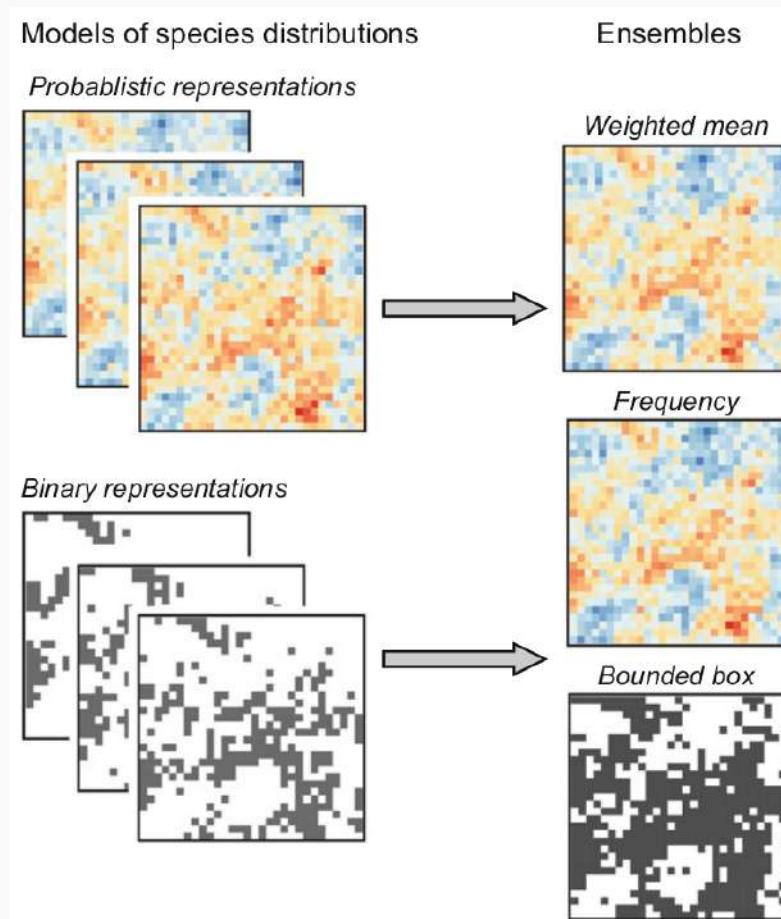


[Araújo & New \(2007\)](#) [Qiao et al. \(2015\)](#)

## 7.2 Consenso dos modelos

### Tipos de consensos

- Frequência
- Média
- Média ponderada (AUC ou TSS)



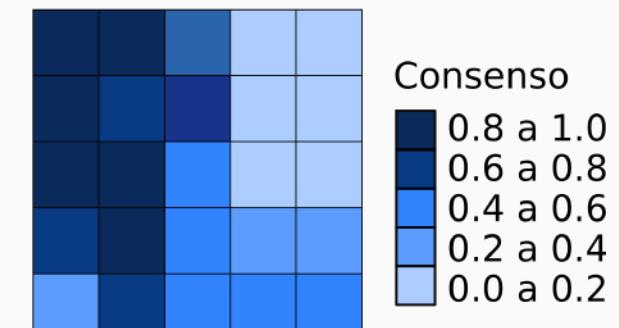
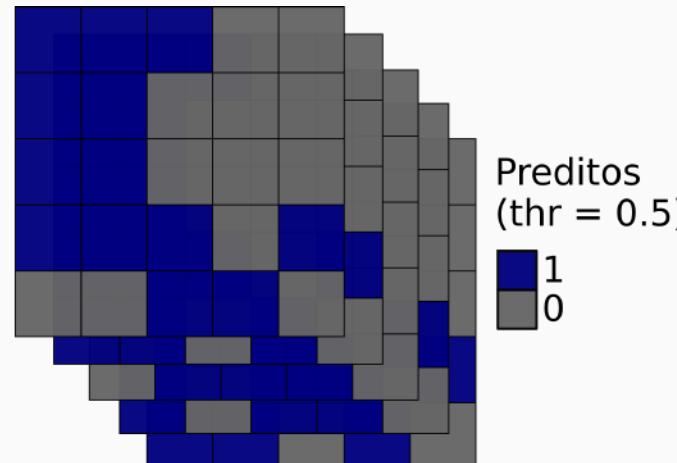
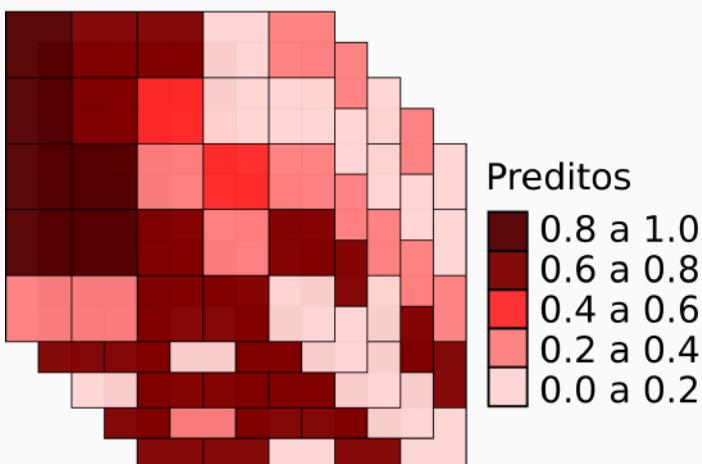
## 7.3 Consenso por frequênci

Contabilizar quantas vezes um pixel foi previsto como presença

1. Uso de um **limiar de corte para binarizar** os modelos [0 e 1]

2. **Soma** da quantidade de pixels e divisão pelo total de modelos [0 a 1]

$$\text{consenso} = \sum_{i=1}^N (\text{pixel}_i \geq \text{limiar}_j)$$



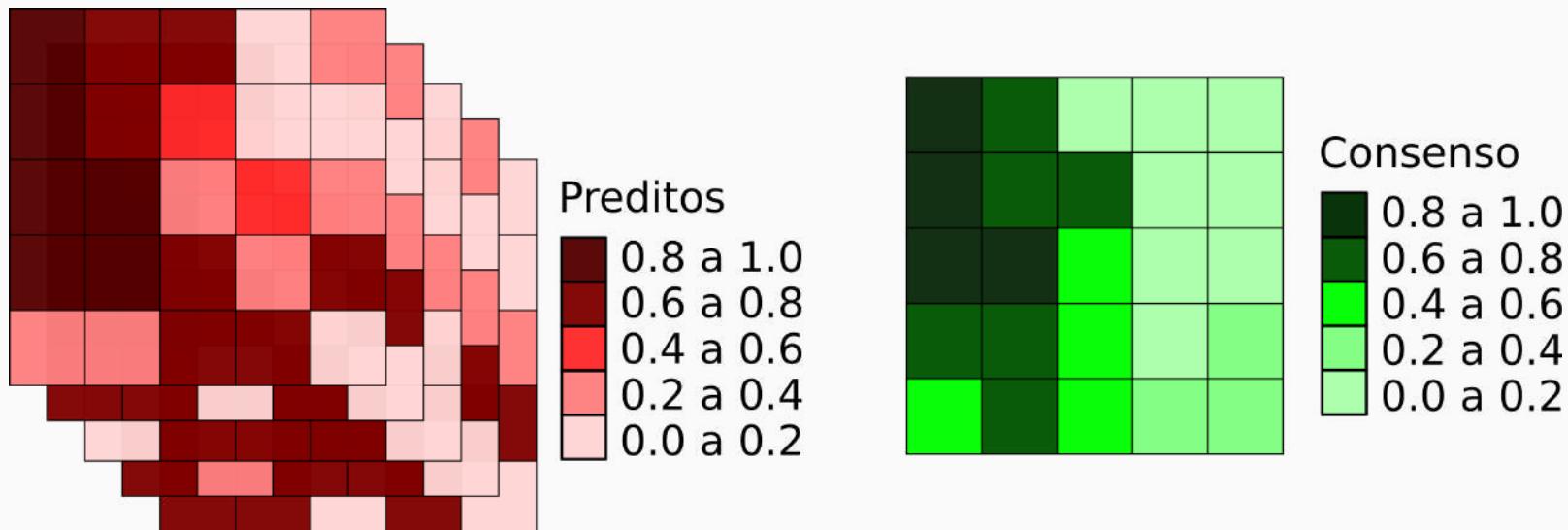
## 7.4 Consenso por média

Calcular a média dos valores de adequabilidade para cada pixel

1. Padronização dos valores de adequabilidade [0 a 1]

2. Cálculo da **média aritmética** para cada pixel

$$\text{consenso} = \frac{1}{N} \sum_{i=1}^N (\text{pixel}_i)$$



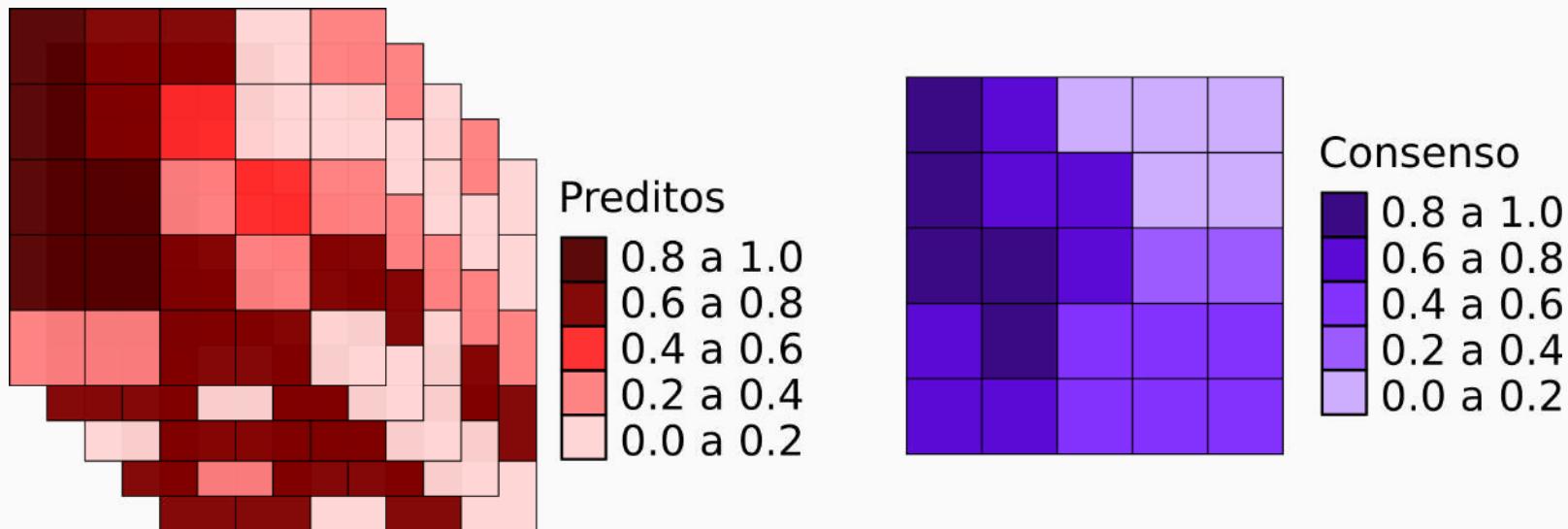
## 7.5 Consenso por média ponderada

Calcular a média ponderada dos valores de adeq. para cada pixel

1. Padronização dos valores de adequabilidade [0 a 1]

2. Cálculo da **média ponderada** pelos valores de AUC ou TSS para cada pixel

$$\text{consenso} = \frac{\sum_{i=1}^N (w_i \cdot \text{pixel}_i)}{\sum_{i=1}^N (w_i)}$$



# 7.6 Incertezas

Calcular um valor de incerteza para cada pixel

1. Padronização dos valores de adequabilidade [0 a 1]

2. Cálculo do desvio padrão ou partição da variância para cada pixel

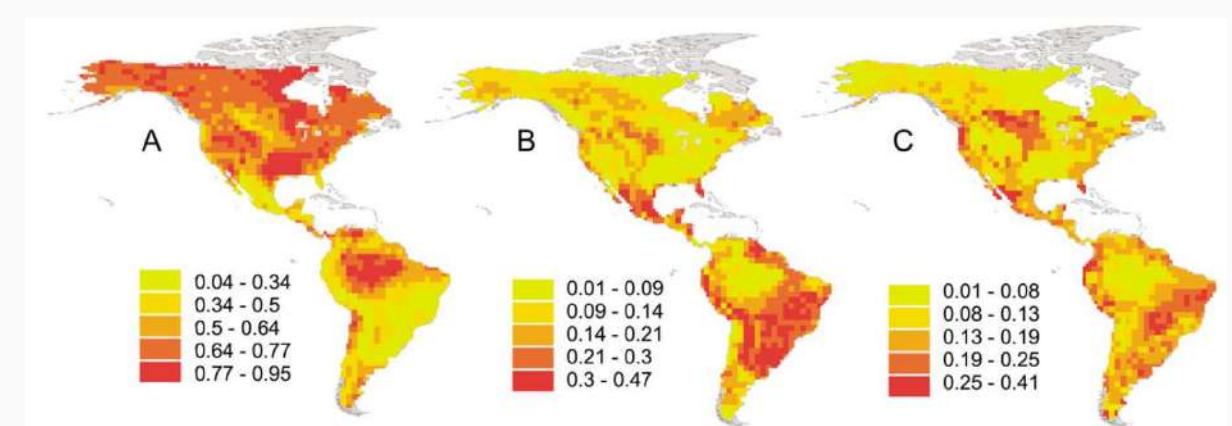
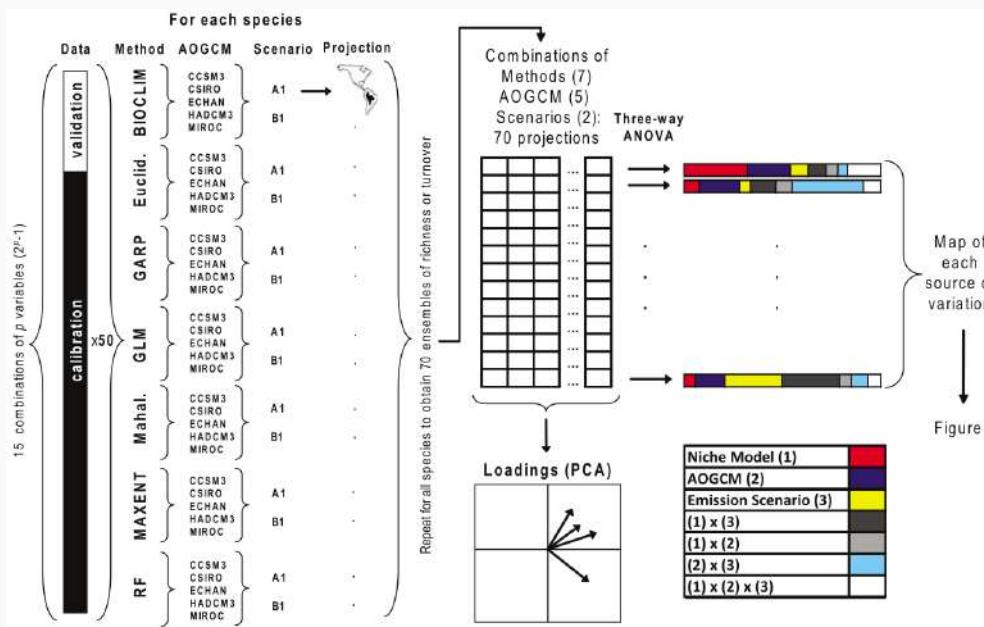


Figure 3. Proportion of the total sum of squares accounted for by SDM (A), AOGCM (B) and the interaction between these factors (C).

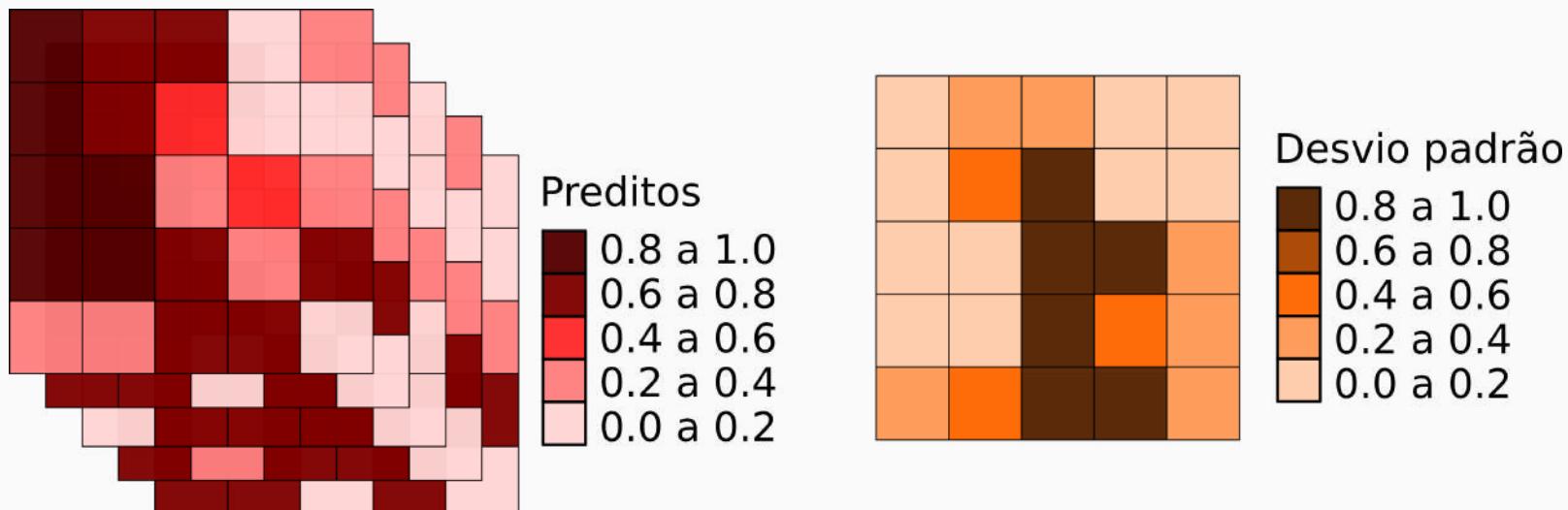
# 7.6 Incertezas

Calcular um valor de incerteza para cada pixel

1. Padronização dos valores de adequabilidade [0 a 1]

2. Cálculo do **desvio padrão** ou **partição da variância** para cada pixel

$$\text{desvio} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{pixel}_i - \bar{\text{pixel}})^2}$$



Vamos para o R predizer e fazer o consenso dos nossos  
modelos

# Principais artigos

## Português

- [De Marco, P., & Siquerá, M. F. \(2009\). Como determinar a distribuição potencial de espécies sob uma abordagem conservacionista? 5, 12.](#)
- [Giannini, T. C. et al. \(2012\). Desafios atuais da modelagem preditiva de distribuição de espécies. Rodriguésia, 63\(3\), 733–749.](#)

## Inglês

- [Araújo, M. B. et al. \(2019\). Standards for distribution models in biodiversity assessments. Science Advances, 5\(1\), eaat4858.](#)
- [Feng, X. et al. \(2019\). A checklist for maximizing reproducibility of ecological niche models. Nature Ecology & Evolution, 3\(10\), 1382–1395.](#)
- [Zurell, D. et al. \(2020\). A standard protocol for reporting species distribution models. Ecography, 43\(9\), 1261–1277.](#)
- [Sillero, N., & Barbosa, A. M. \(2021\). Common mistakes in ecological niche models. International Journal of Geographical Information Science, 35\(2\), 213–226.](#)
- [Sillero, N. et al. \(2021\). Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. Ecological Modelling, 456, 109671.](#)

# Livros

## Português

- [Lima-Ribeiro, M., & Diniz-Filho, J. \(2013\). Modelos ecológicos ea extinção da megafauna: Clima e homem na América do Sul.](#)

## Inglês

- [Franklin, J., & Miller, J. A. \(2009\). Mapping species distributions: Spatial inference and prediction. Cambridge University Press.](#)
- [Peterson, A. T. \(Org.\). \(2011\). Ecological niches and geographic distributions. Princeton University Press.](#)
- [Peterson, A. T. \(2014\). Mapping disease transmission risk: Enriching models using biogeography and ecology. Johns Hopkins University Press.](#)
- [Wegmann, M., Leutner, B., & Dech, S. \(Orgs.\). \(2016\). Remote sensing and GIS for ecologists: Using open source software. Pelagic Publishing.](#)
- [Guisan, A., Thuiller, W., & Zimmermann, N. E. \(2017\). Habitat Suitability and Distribution Models: With Applications in R. Cambridge University Press.](#)

# Cursos on-line

## Português

- [Prof. Ubirajara de Oliveira](#)
- [Profa. Luara Tourinho](#)

## Inglês

- [Prof. Townsend Peterson](#)
- [Profa. Damaris Zurell](#)
- [Prof. Babak Naimi](#)
- [Species Distribution Modeling Remote with Sensing \(NASA\)](#)
- [BCCVL Training](#)

Muito obrigado!

# Maurício Vancine

Contatos:

✉ [mauricio.vancine@gmail.com](mailto:mauricio.vancine@gmail.com)

🐦 [@mauriciovancine](https://twitter.com/mauriciovancine)

🐙 [mauriciovancine](https://github.com/mauriciovancine)

🔗 [mauriciovancine.github.io](https://mauriciovancine.github.io)



Slides criados via pacote [xaringan](#) e tema [Metropolis](#). Animação dos sapos por [@probzz](#).