Análises multidimensionais em estudos de campo

Prof. Dr. Diogo Borges Provete
Setor de Ecologia – INBIO – UFMS
diogo.provete@ufms.br
diogoprovete.weebly.com

Cronograma

Segunda-feira: manhã

- Introdução às análises multidimensionais, objetivos e principais usos
- Recapitulação: Correlação e Covariância
- Plano Euclidiano
- Rudimentos de Álgebra de Matrizes: Autovetores, Autovalores, autoanálise
- Combinação linear de variáveis aleatórias

Segunda-feira: tarde

- Coeficientes de distância métricos, semi-métricos e não-métricos
- Transformações e padronização de dados
- Prática: Introdução ao R, importação de dados, e cálculo de matrizes de distância

Cronograma

Terça-feira	• Agrupamento hierárquico, não-hierárquico (k-means), espécies indicadoras (IndVal) • Prática à tarde
Quarta-feira	• Ordenações irrestritas: PCA, PCoA, nMDS, CA • Prática à tarde
Quinta-feita	• Ordenação restrita (canônica): CCA, RDA, pRDA, db-RDA, LDA • Prática à tarde
Sexta-feira	• Métodos para relacionar dois conjuntos de dados: Mantel, Mantel parcial, RMR, Procrustes/Protest, PERMANOVA, betadisper, permdist

Ao final do curso

1

Avaliação

- Duas: prova teórica e trabalho prático para entregar uma semana depois com análise de dados multivariados e interpretação

2

Resumos e mapas mentais são fornecidos ao final da disciplina no Moodle

3

Feedback disciplina

Avisos

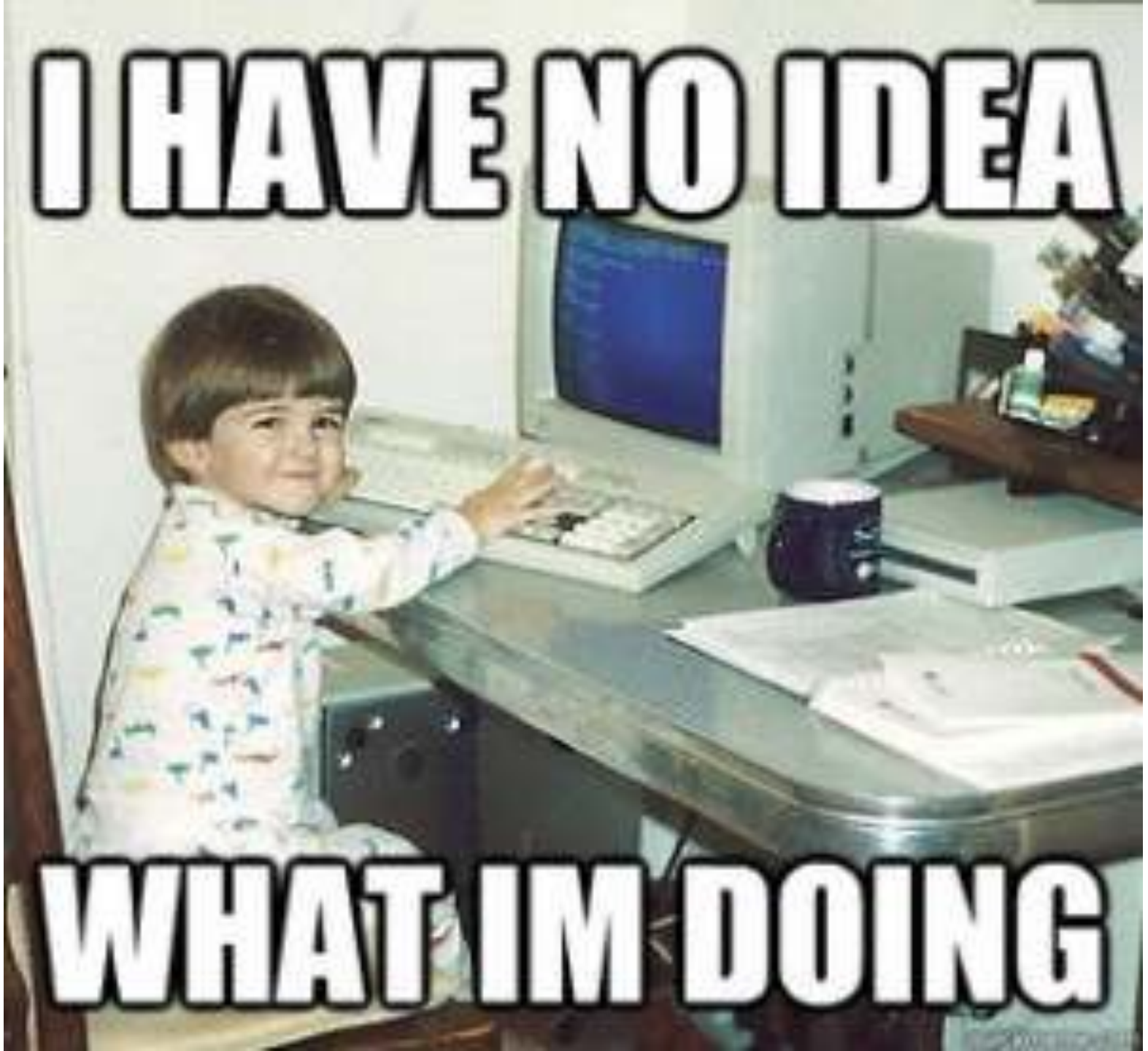
Curso inteiramente no Moodle

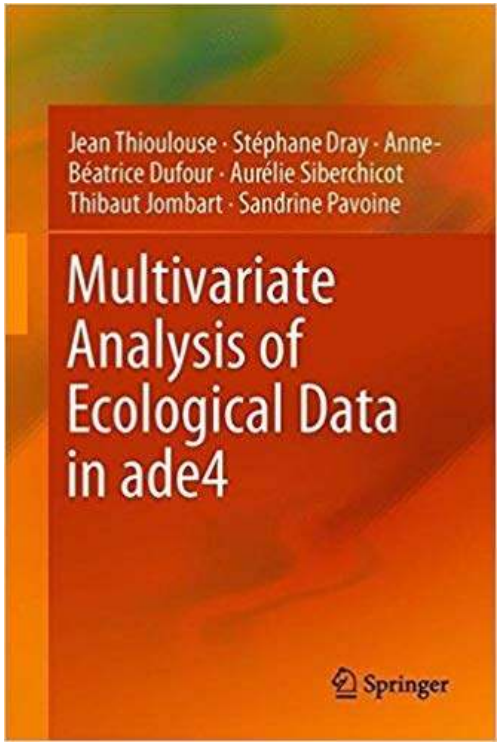
https://ava.ufms.br/course/view.php?id=228

Slides, vídeos, bibliografia, envio do trabalho e prova teórica

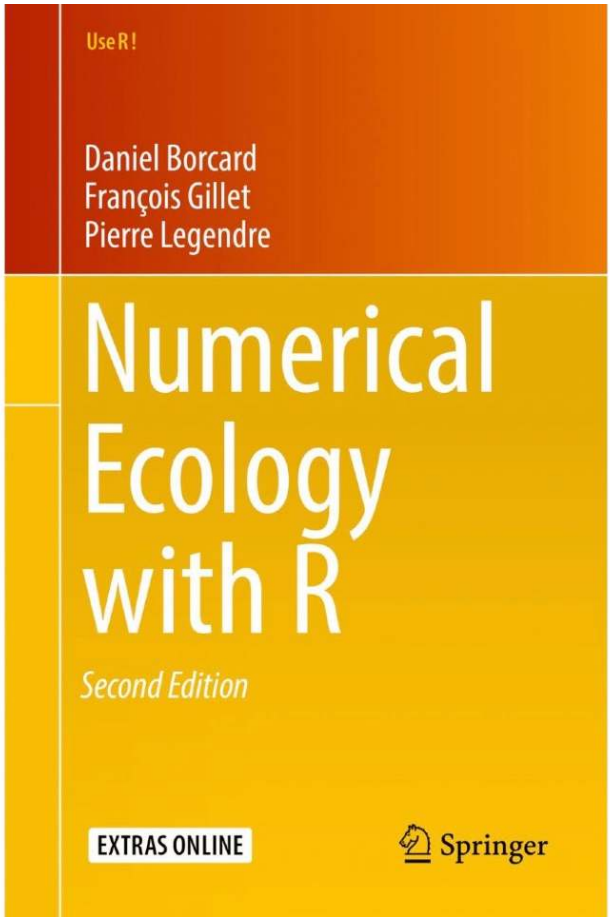
Logo, não precisa copiar o slide, melhor acompanhar a aula e PERGUNTAR!

Scripts do R já serão fornecidos, alunos devem acompanhar junto

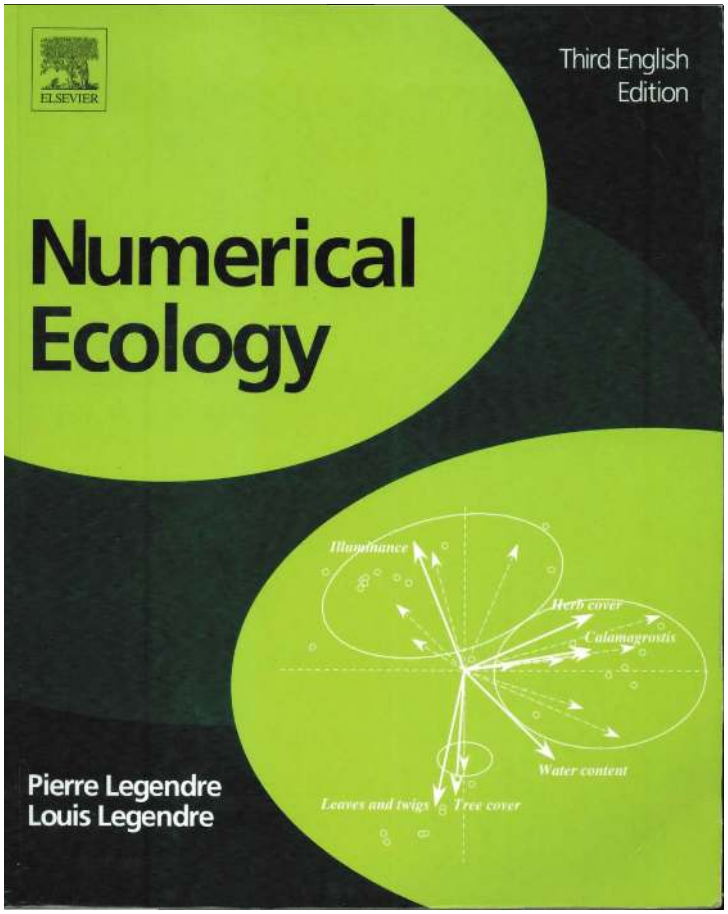

I HAVE NO IDEA
WHAT IM DOING



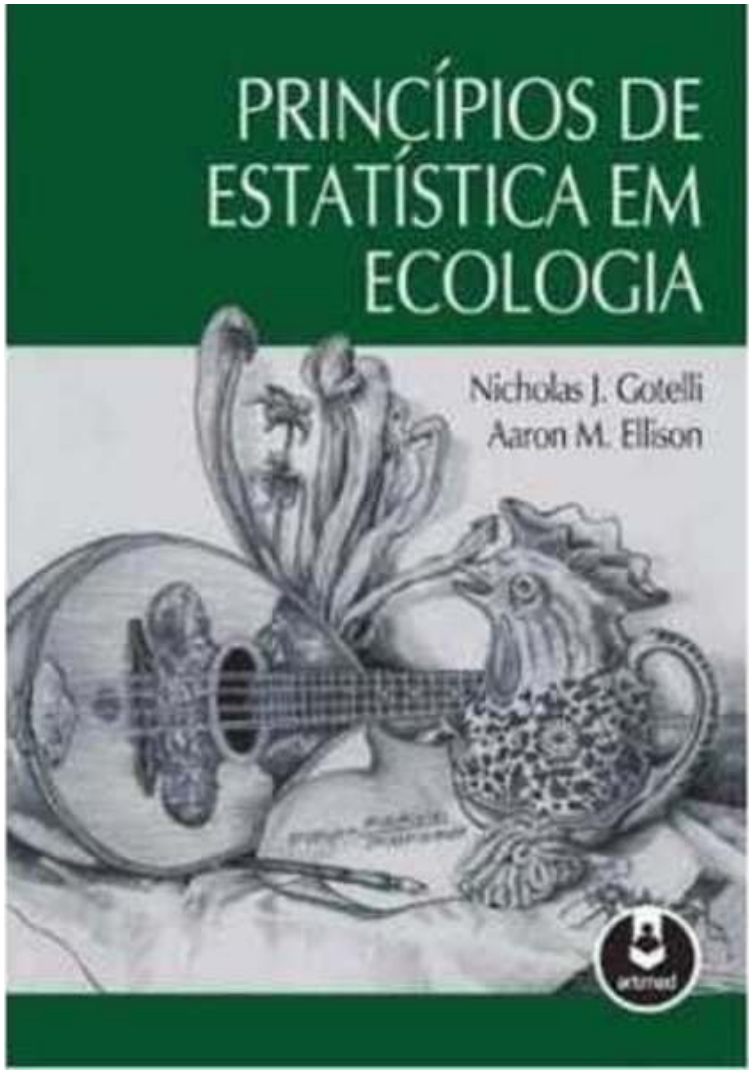
2019

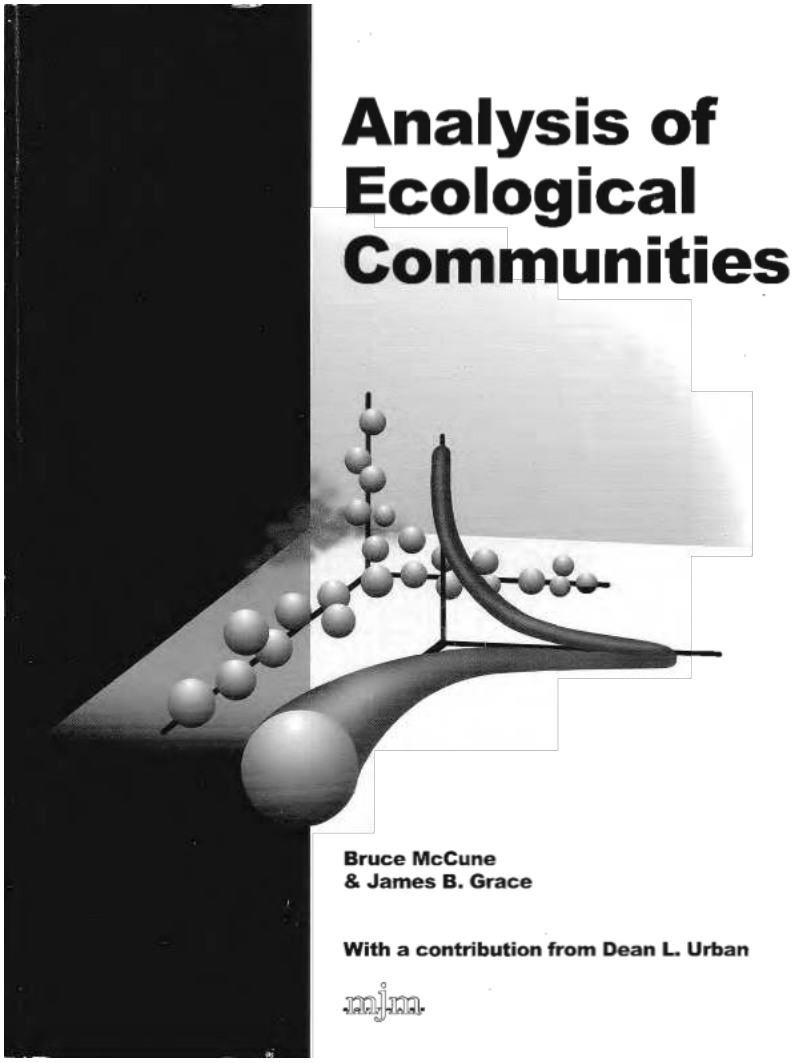


2018



2012





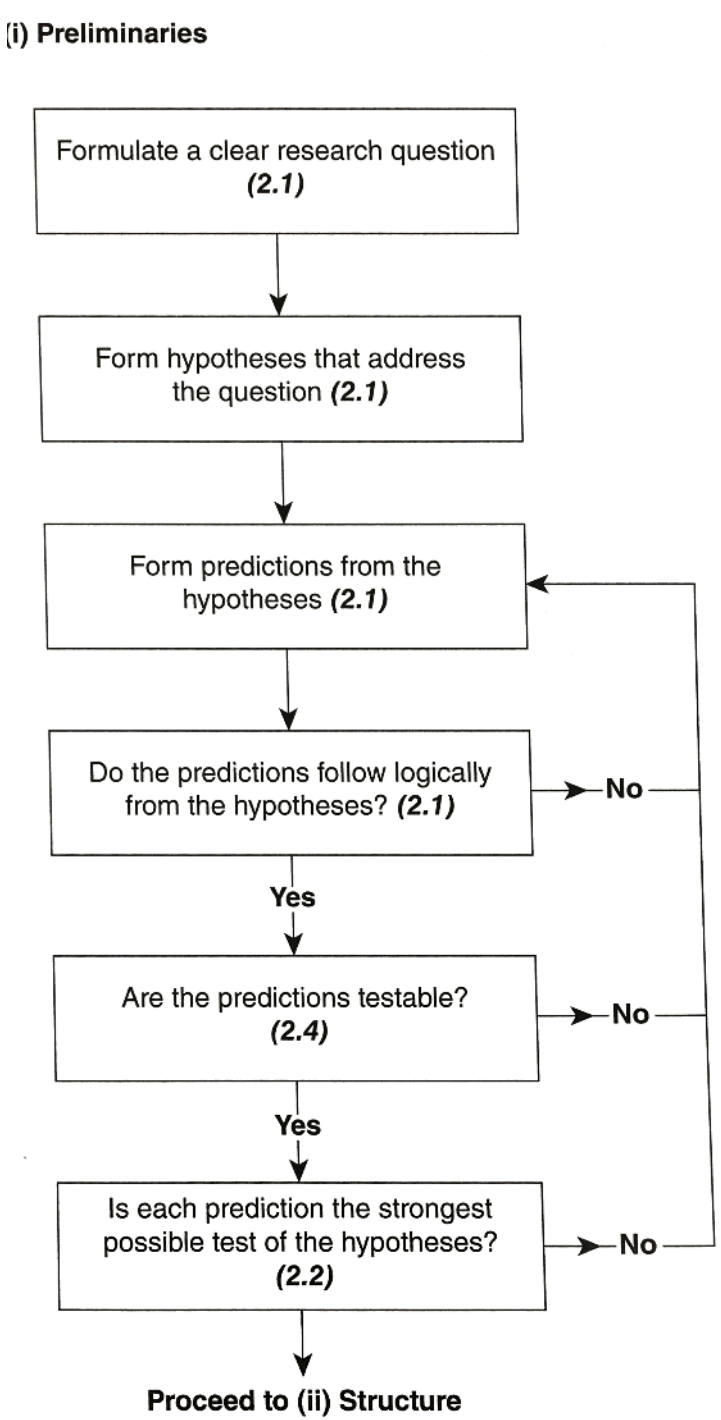
2002

Cap 12 sobre multivariada
Cap. 7 sobre desenhos amostrais

Todos prontos?

Diferenças estudos observacionais e experimentais

- Tudo começa com uma boa pergunta
  - Mas o que diabos é uma “boa pergunta”?
  - Hipótese científica vs. estatística
  - Predições
- Conceitos imporantes
  - Réplica, unidade amostral, independência amostral
- Diferença entre estudos observacionais e experimentais
  - Experimentos: controle, aleatorização, produzem relação causal
  - Observacionais: confundimento, 3ª variável



Ruxton & Colegrave 2016

Dicas para fazer uma boa pergunta

- Uma pergunta científica deve ser **testável** por meio de observações e /ou experimentos
- Para algo ser **testável,** ele tem de ser **mensurável**
- Uma *boa* pergunta deve conter as variáveis resposta (ou dependente) e preditora(s) (ou independente)
- Atente para o verbo usado, ele vai guiar o tipo de teste

Exemplos de pergunta

- Qual a **relação** entre X e Y?
- Qual o **efeito** do fator X no Y?
- Como o fator x **influencia** o fator Y?
- Como a mudança em X **afeta** Y?

- **Evite** questões do tipo sim/não ou pras quais a resposta é sim/não

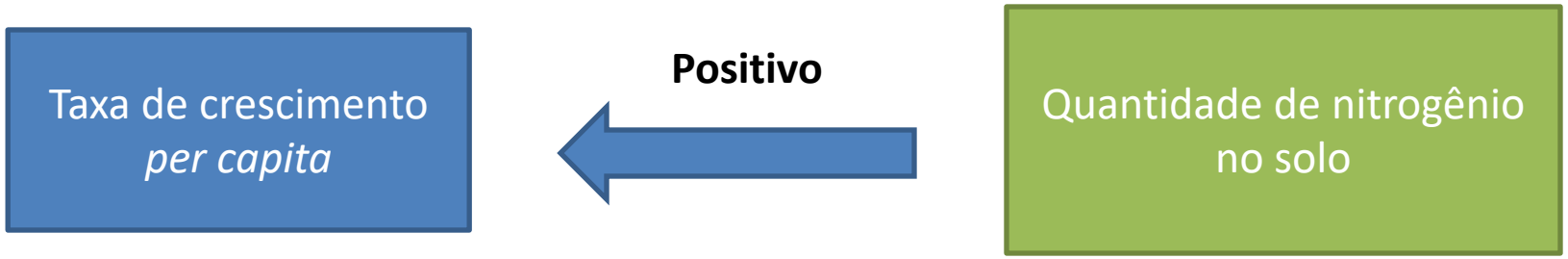
Exemplos

- Existe diferença entre as áreas?
- Qual a influência do tipo de solo no crescimento de *Pinus*?
- Qual o efeito da quantidade de nitrogênio no solo na taxa de crescimento *per capita* de árvores adultas do gênero *Pinus*?
- A quantidade de nitrogênio no solo aumenta a taxa de crescimento per capita de *Pinus*?

Como podemos melhorar nossas perguntas?

- Existe **diferença** entre as áreas?
- Qual a **influência** do tipo de solo no crescimento de *Pinus*?
- Qual o **efeito** da quantidade de nitrogênio do solo na taxa de crescimento *per capita* de árvores adultas do gênero *Pinus*?
- A quantidade de nitrogênio no solo **aumenta** a taxa de crescimento per capita de *Pinus*?

Fluxograma com perguntas e predições



Hipótese: O nitrogênio está presente em várias moléculas orgânicas envolvidas no processo de crescimento de plantas em geral. **Logo**, espero que quanto maior a quantidade de nitrogênio disponível no solo, maior seja a taxa de crescimento *per capita* de uma população de plantas do gênero *Pinus*

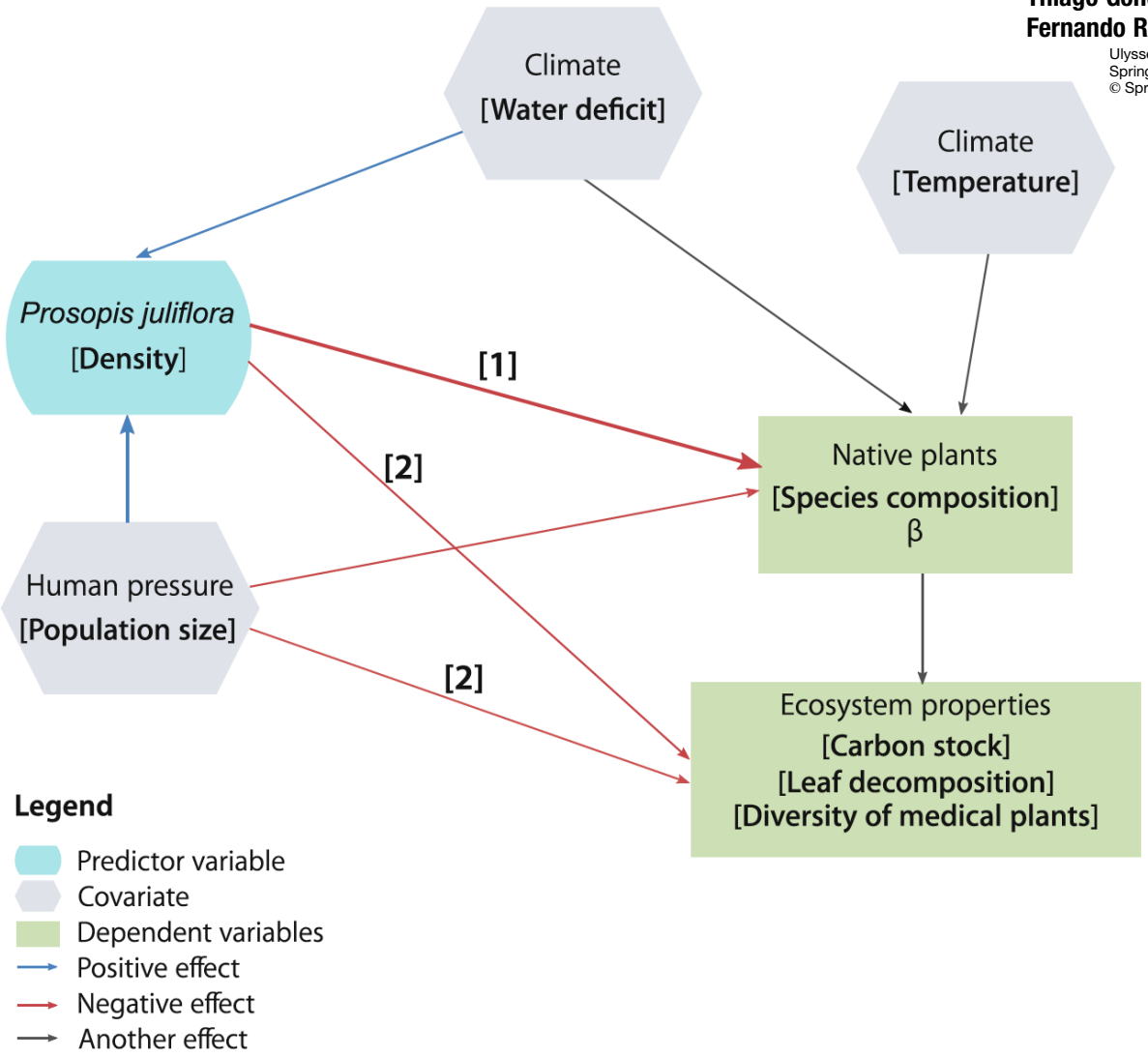
Cap. 2 Ruxton & Colegrave

Going Back to Basics: How to Master the Art of Making Scientifically Sound Questions

Thiago Gonçalves-Souza, Diogo B. Provete, Michel V. Garey, Fernando R. da Silva, and Ulysses Paulino Albuquerque

Ulysses Paulino Albuquerque et al. (eds.), *Methods and Techniques in Ethnobiology and Ethnoecology*, Springer Protocols Handbooks, https://doi.org/10.1007/978-1-4939-8919-5_7,

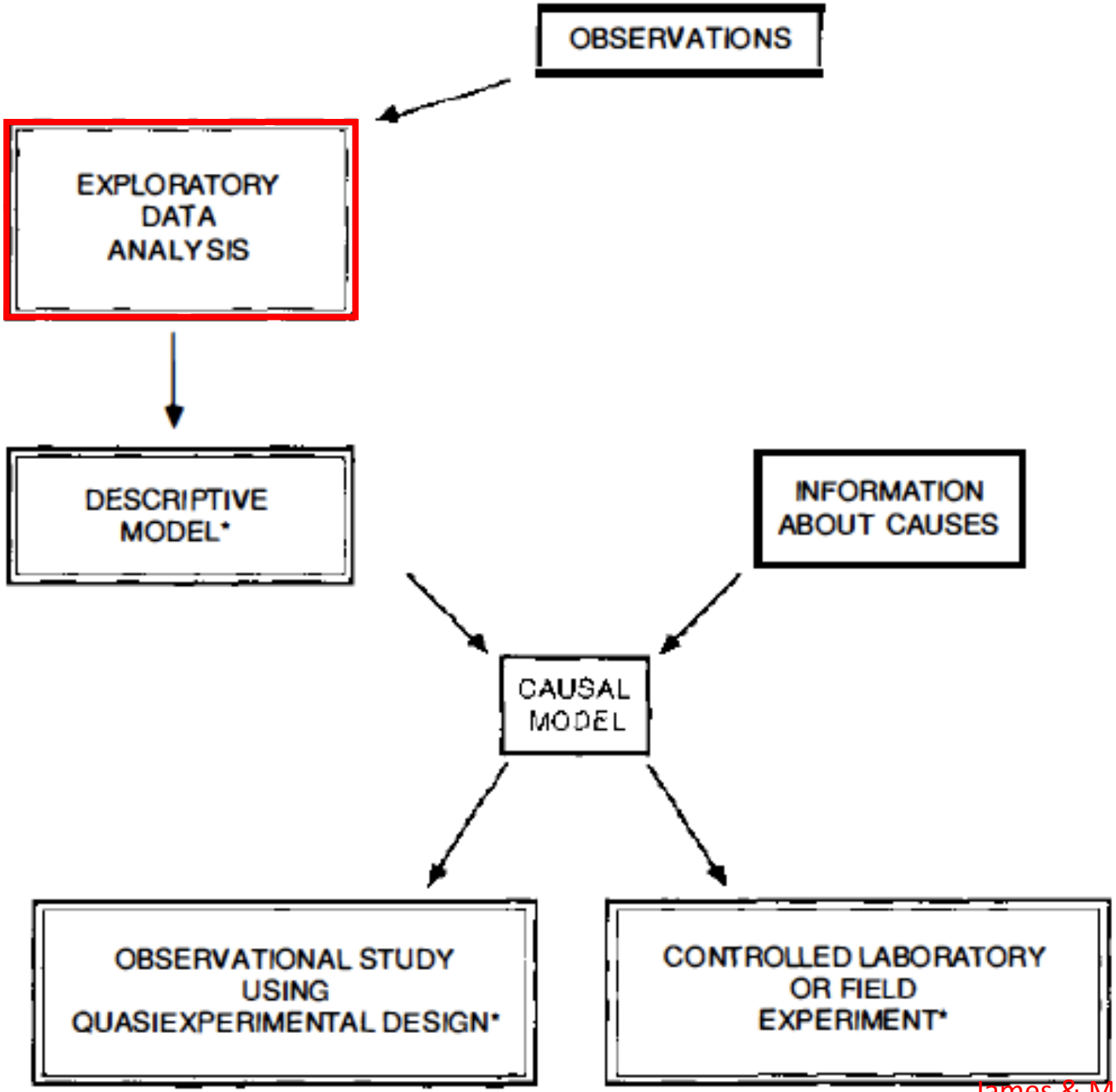
a) Flowchart



Conceitos importantes

- **Unidade amostral**: Plot de floresta (população de árvores)
- **Réplica**: vários plots distribuídos na floresta
- Melhor replicar plots ou florestas? (11.2.1 Ruxton & Colegrave)
  - Depende da pergunta e onde está a maior variabilidade dos dados
- Tamanho do efeito: diferença na média entre grupos ou tratamentos
  - Poder do teste
  - Quantos plots? Quantas florestas? “regra” dos 10 (Gotteli & Ellison, 2004 cap 6) => 10 u.a. pra cada variável preditora

- Variabilidade natural nos dados => estudos piloto (2.4 Ruxton & Colegrave)

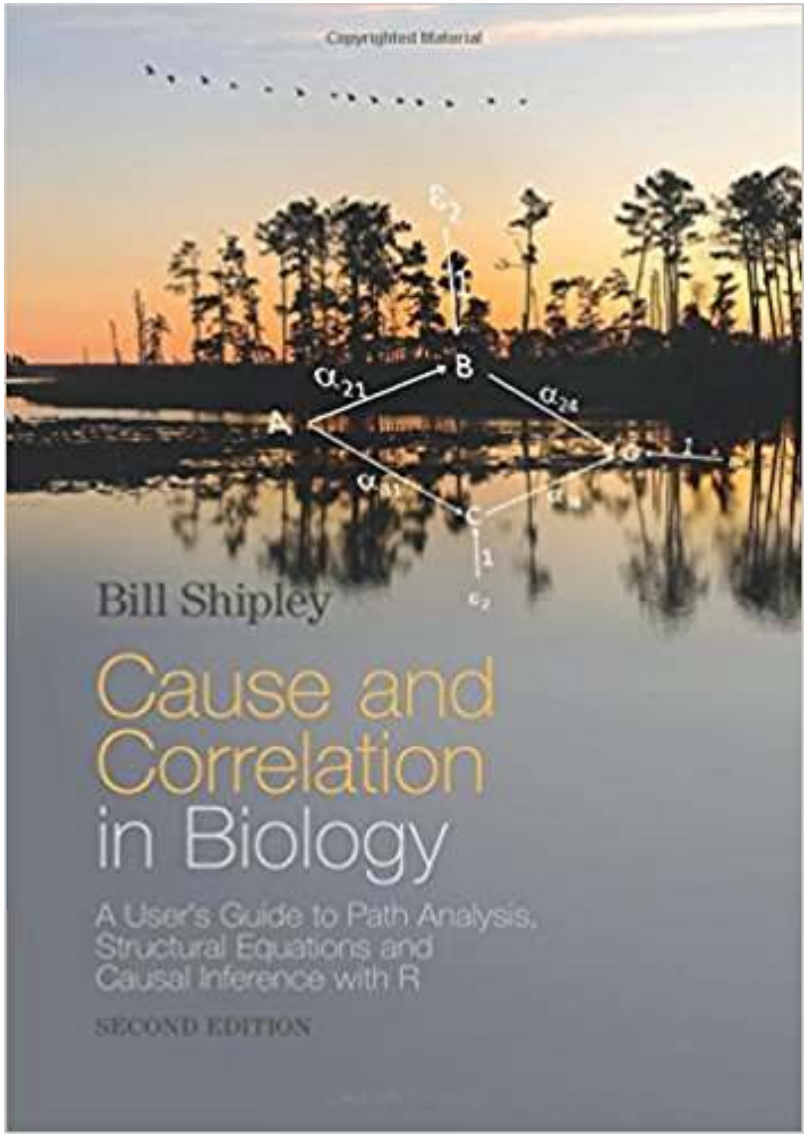


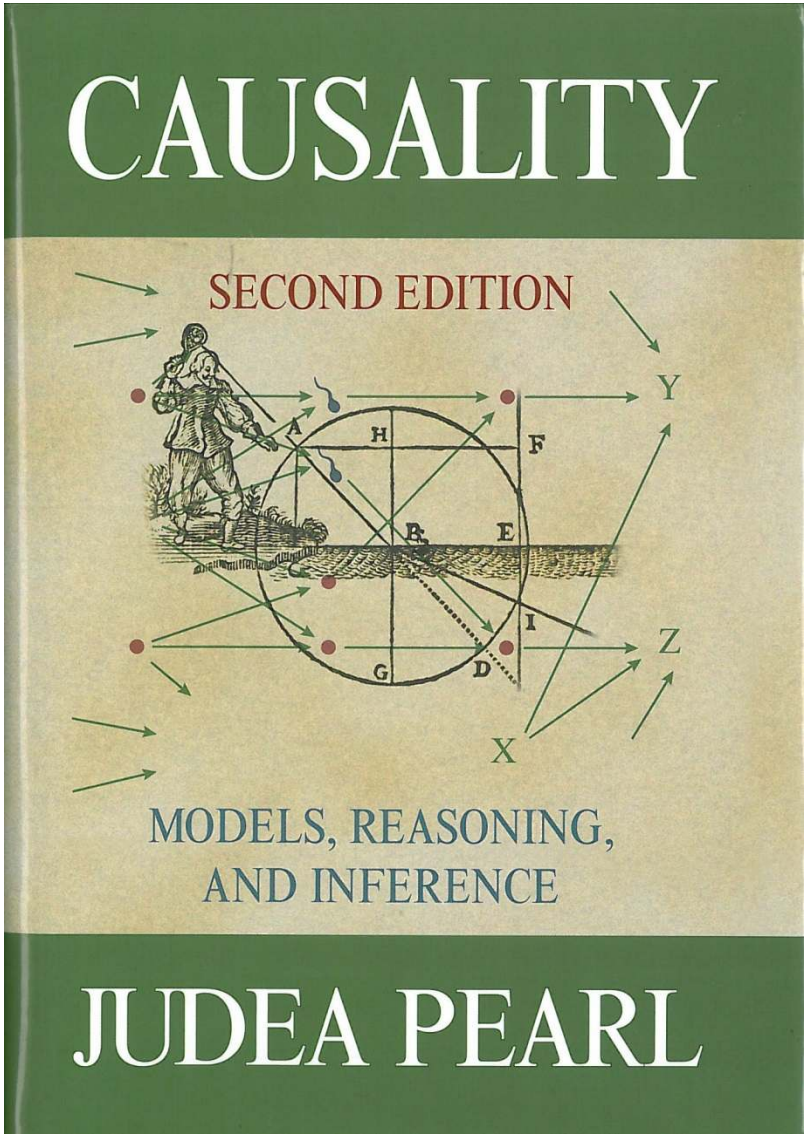
James & McCulloch (1990)

Diferença entre estudos observacionais e experimentais

- Estudos observacionais não permitem inferir **causalidade**, relação causal entre variáveis, porque não isolam variáveis confundidoras e as réplicas podem não ser independentes
  - Análise de rota (Path analysis – Structural Equation Modelling)
  - Incluir covariável no modelo? Análise exploratória de dados? Partição dos dados?

- Estudos experimentais permitem inferir causalidade, pois a **atribuição** das unidades amostrais aos tratamentos é **aleatória**


Copyrighted Material
Bill Shipley
Cause and Correlation in Biology
A User's Guide to Path Analysis, Structural Equations and Causal Inference with R
SECOND EDITION


CAUSALITY
SECOND EDITION
MODELS, REASONING, AND INFERENCE
JUDEA PEARL

Estudos experimentais

- Aleatorização é a chave!
  - Efeitos de confundimento vão estar em todos os tratamentos, mas de forma não enviesada
- Desvantagens
  - Escala espacial restrita (Colaboração pode ajudar)
  - Impossível fazer com organismos grandes
  - Custos para estabelecer
  - Limites quantidade de tratamentos

Fazer um estudo experimental ou correlativo?

Fazer um estudo experimental ou correlativo?

Não existe resposta única e nem sempre é possível realizar um experimento

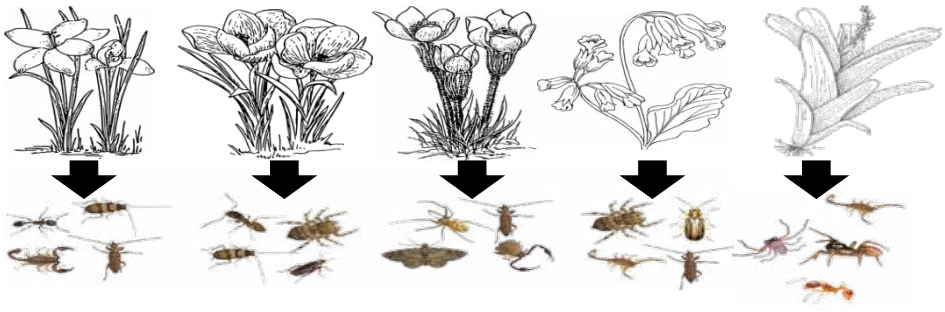
Mais sobre isso no Cap 3 Ruxton & Colegrave

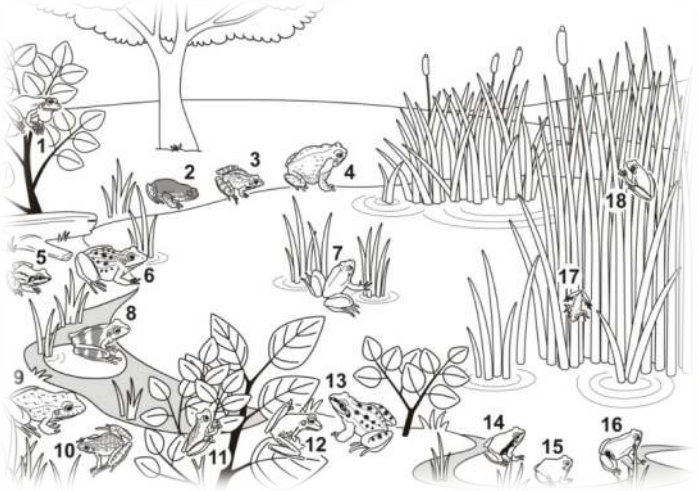
Dois exemplos de estudos observacionais bastante comuns em biologia animal

Resposta de comunidades biológicas

Principal questão

As variáveis ambientais influenciam a composição de espécies?





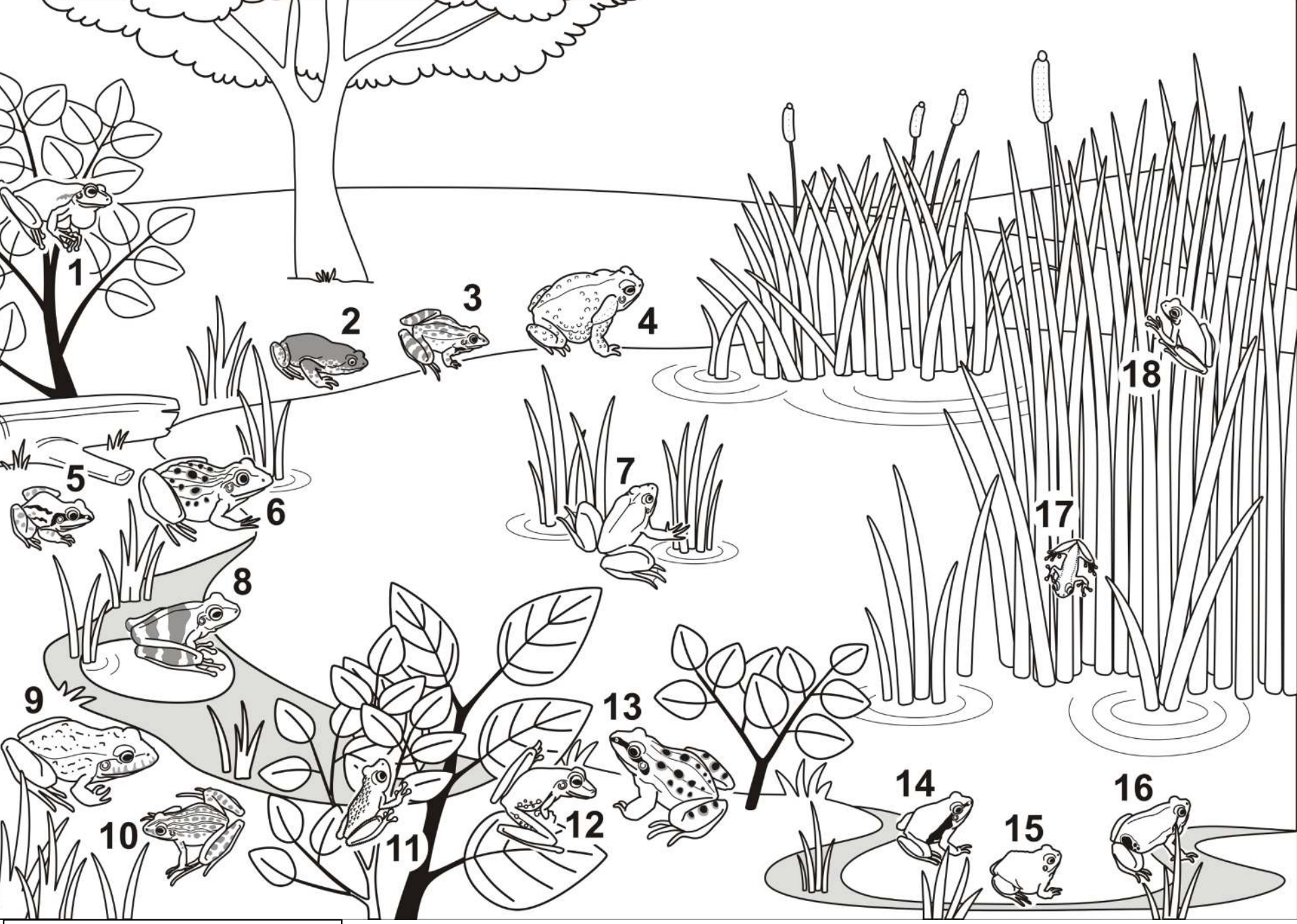
Principal teoria

Nicho ecológico

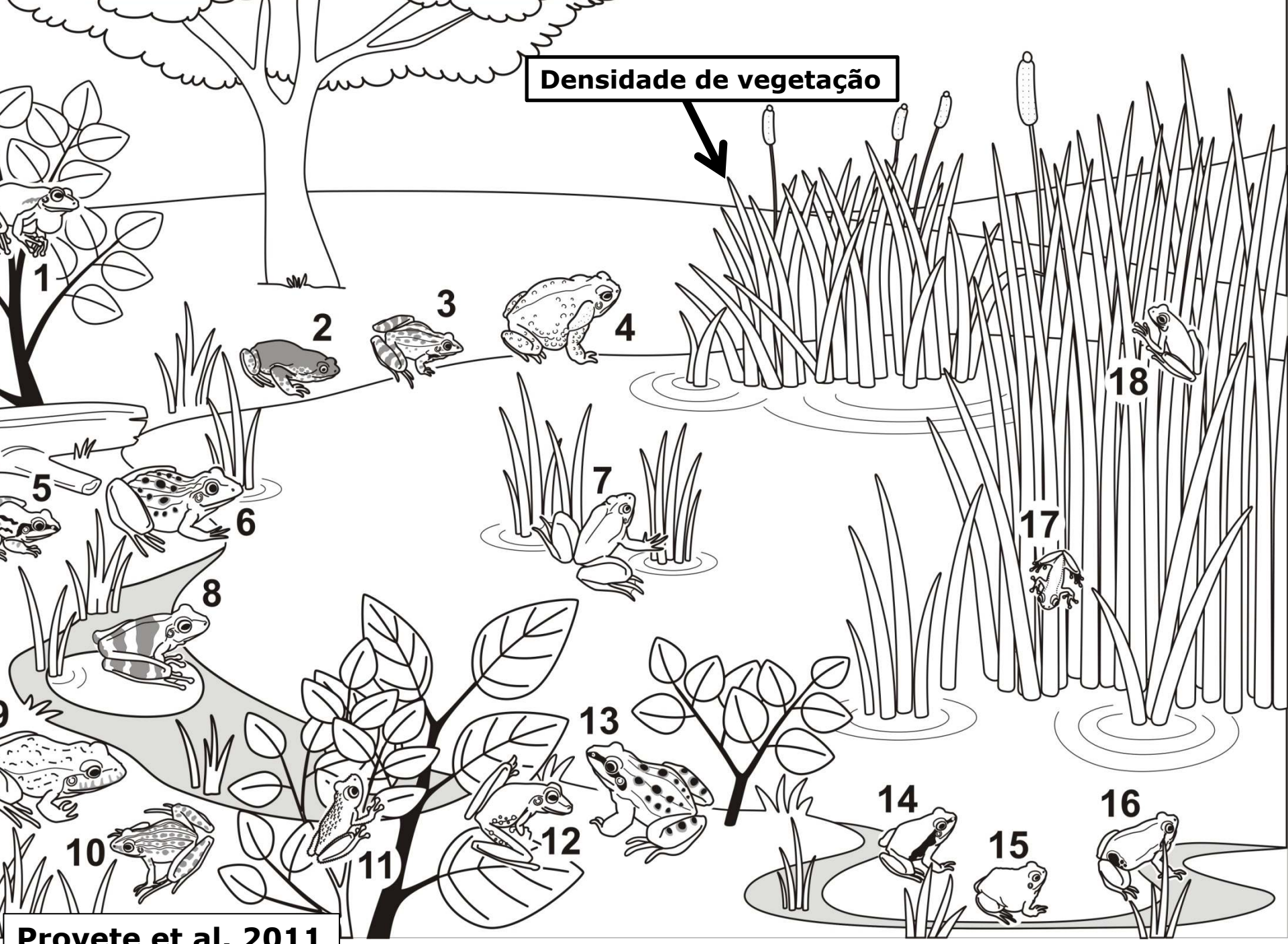
Species sorting/metacomunidades

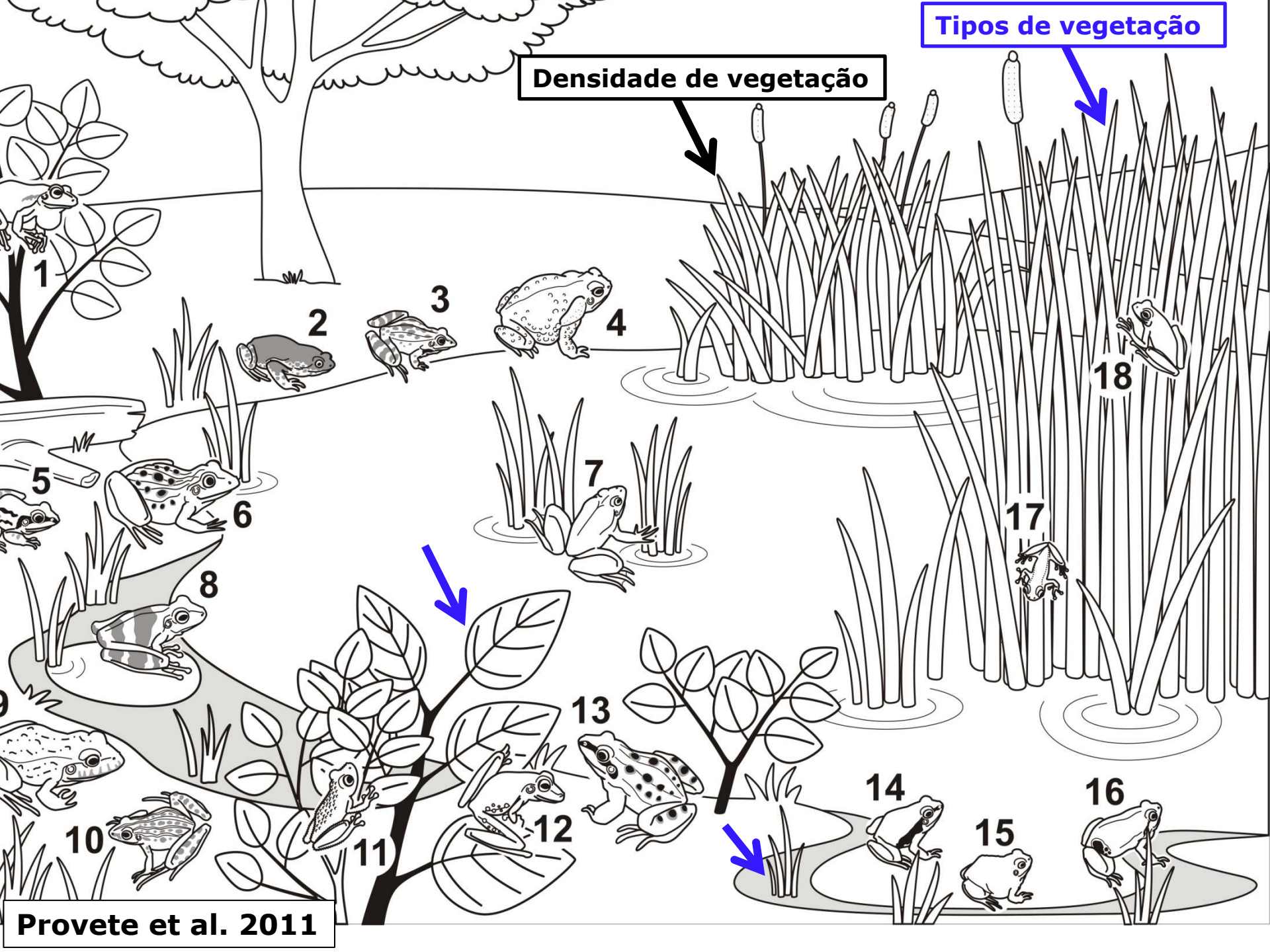
Unidade amostral

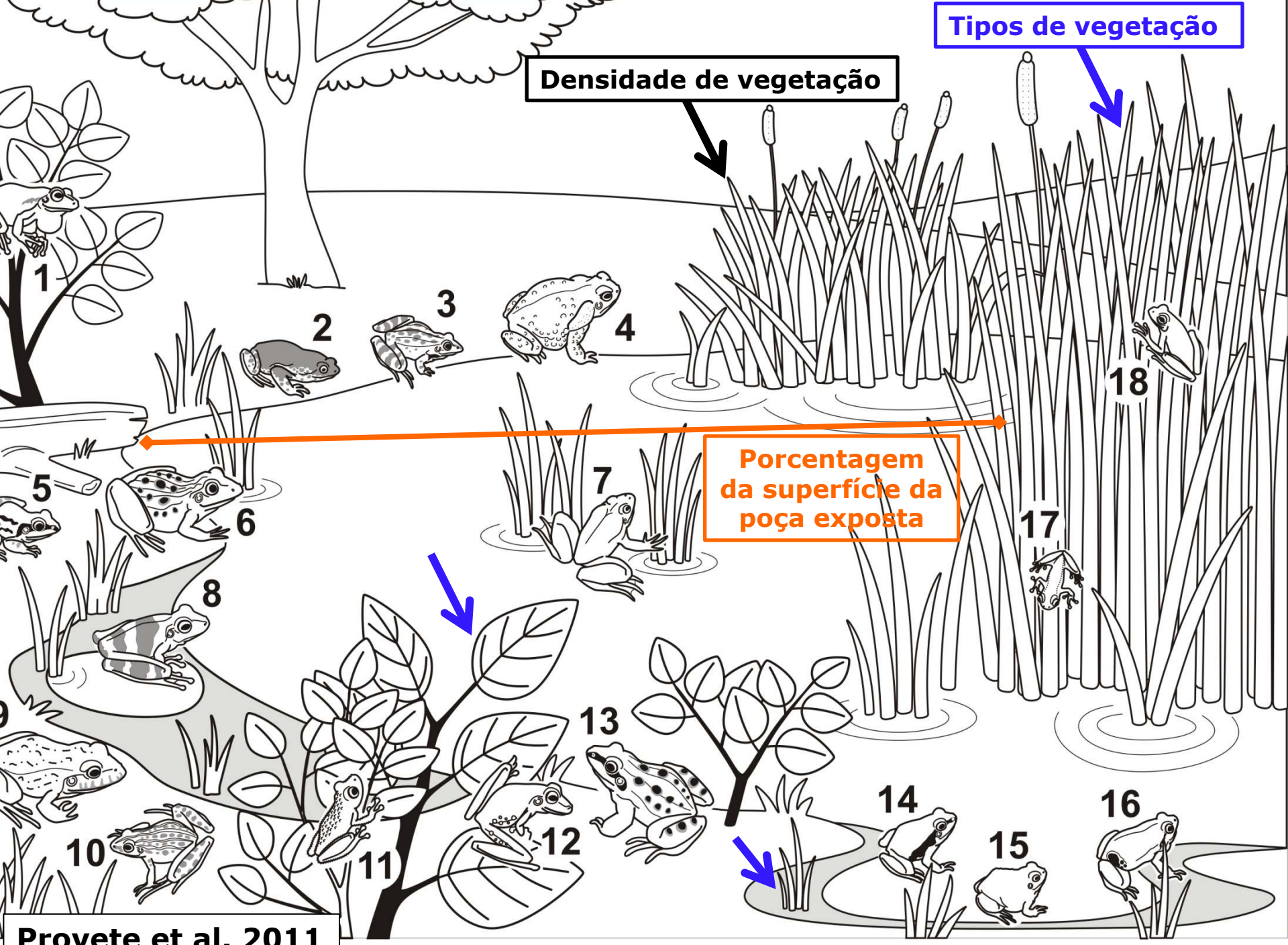
Poças

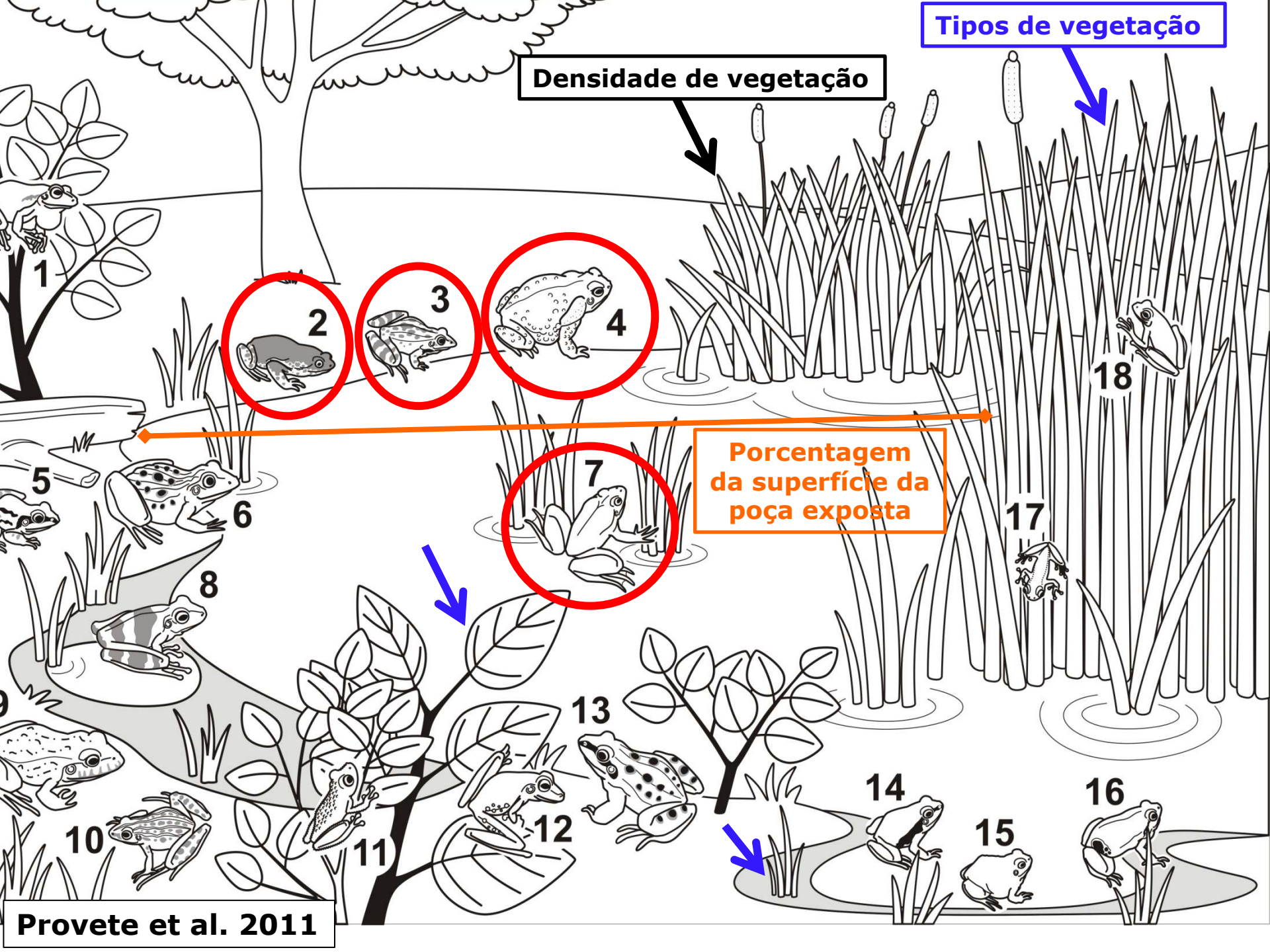


Provete et al. 2011


Densidade de vegetação
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Provete et al. 2011


Tipos de vegetação
Densidade de vegetação
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Provete et al. 2011


Tipos de vegetação
Densidade de vegetação
Porcentagem da superfície da poça exposta
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Provete et al. 2011

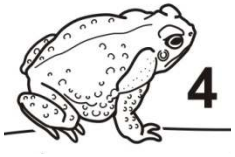

Tipos de vegetação
Densidade de vegetação
Porcentagem da superfície da poça exposta
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Provete et al. 2011

Tipos de vegetação

Densidade de vegetação

Porcentagem da superfície da poça exposta

X



Y

Pergunta

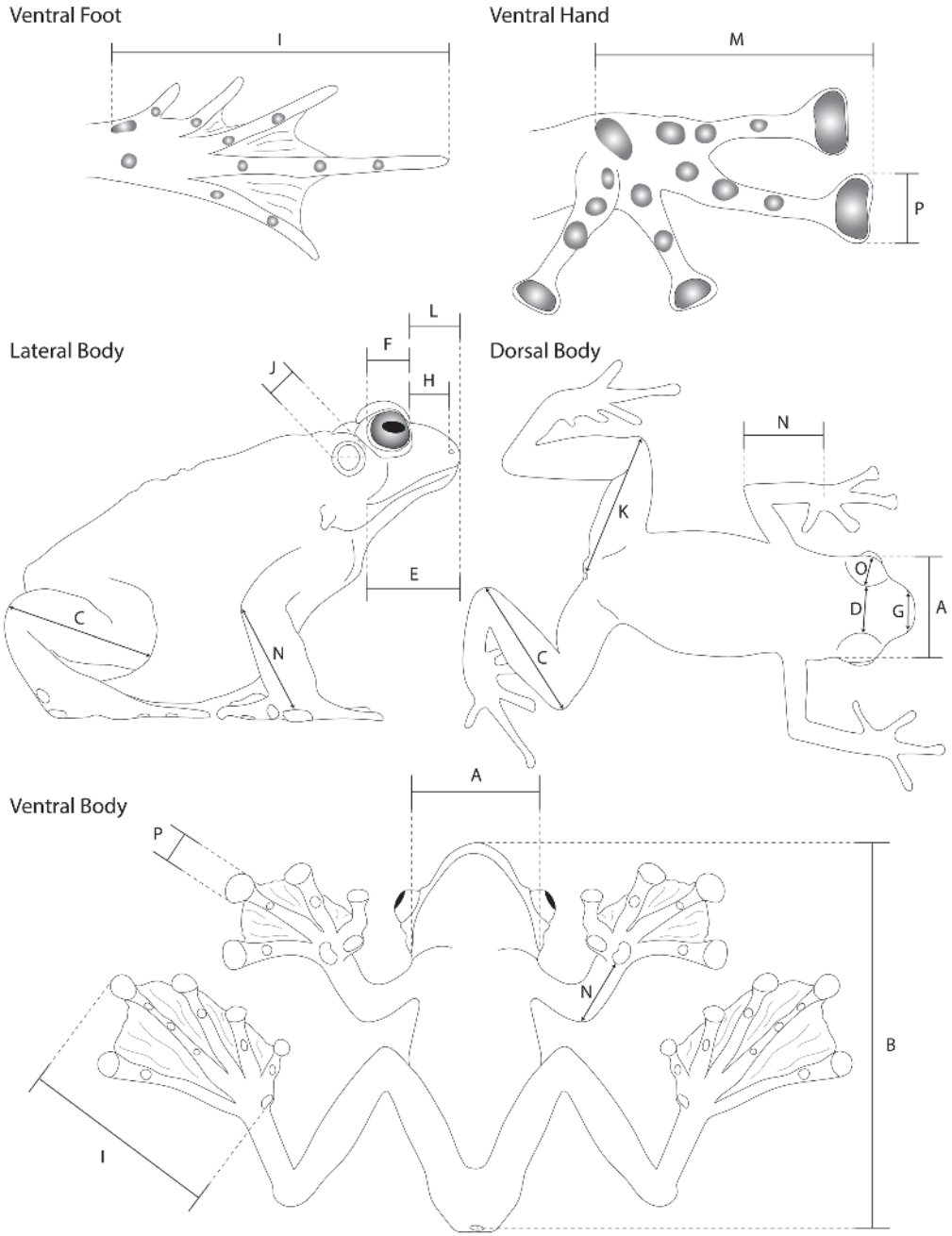
Cladogênese promoveu divergência morfológica em grupos de espécies?

Principal teoria

Teoria da evolução por seleção natural
Modos de especiação

Unidade amostral

Indivíduos de uma dada espécie ou grupo de espécies

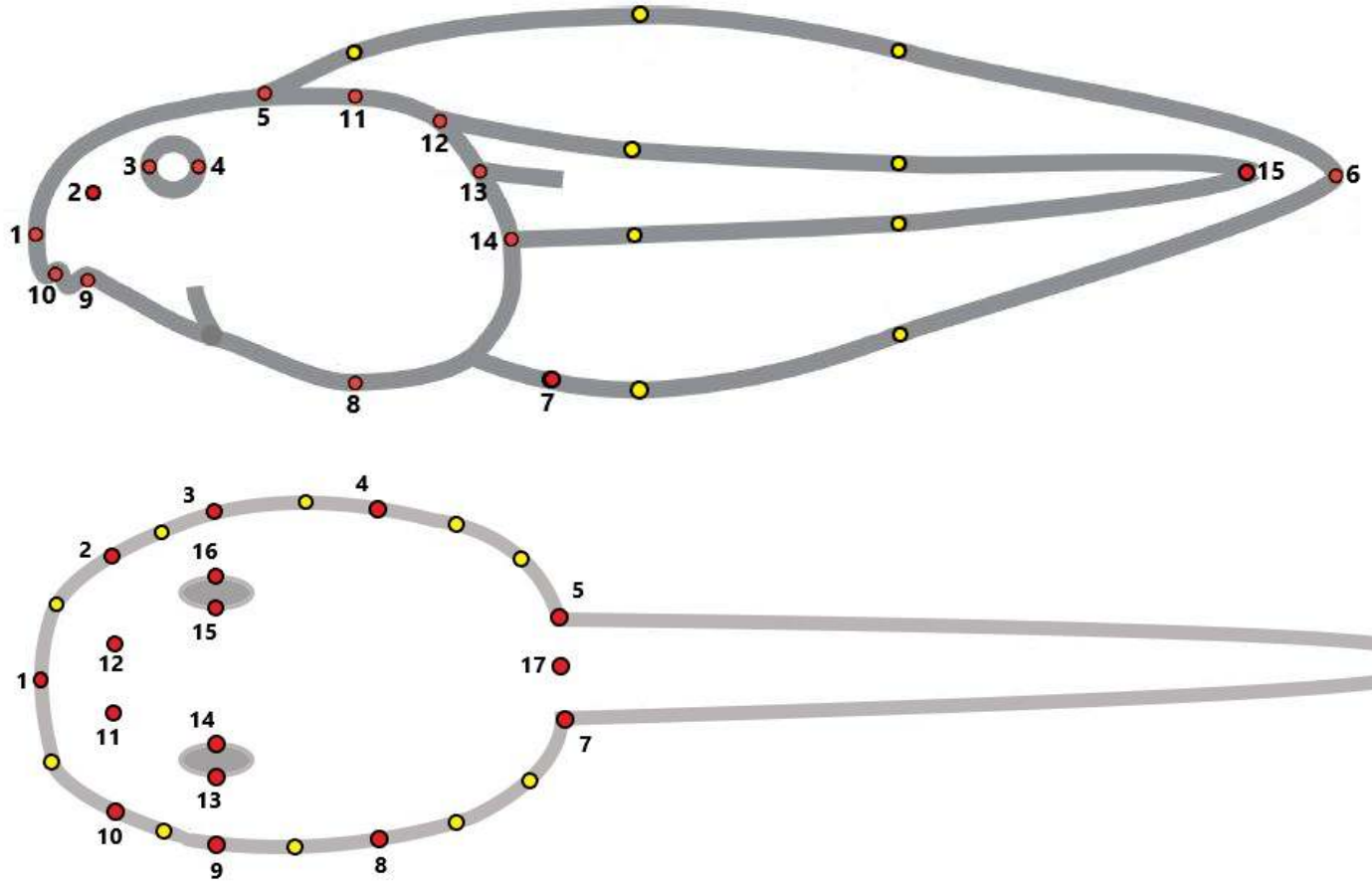


Watters et al. 2016 Zootaxa

Morfometria tradicional (distâncias lineares)

Conjunto de dados multivariado

Morfometria geométrica



Slide Marcos Severgnini

Tabela 1. Planilha modelo para análises estatística, com unidades amostrais nas linhas, e variáveis dependentes e independentes nas colunas

	v. dependente1	v. dependente2	...	v. dependente *n*	v. independente1	v. independente2	...	v. independente m
unid.amostral1	2.593	3.789		n_1	2.177	3.318		m_1
unid.amostral2	2.326	1.000		n_2	2.910	2.575		m_2
unid.amostral3	2.190	1.828		n_3	5.007	3.128		m_3
unid.amostral4	2.883	3.207		n_4	5.479	4.250		m_4
unid.amostral5	1.828	1.810		n_5	1.404	3.298		m_5
unid.amostral6	3.657	2.760		n_6	2.614	3.491		m_6
...	...	...		...	...	...		...
unid.amostral n	n_1	n_2		n_n	m_1	m_2		m_n

TABLE 4: Descriptive statistics of measurements (in mm) of males (N = 321) and females (N = 191) of *Rhinella major* (SD = standard deviation).

Characters	Average		Minimum		Maximum		SD	
	Male	Female	Male	Female	Male	Female	Male	Female
SVL	53.78	54.33	35.8	33.9	72.8	81.1	7.50	8.77
HW	18.05	17.92	12.0	12.1	25.2	34.4	2.21	2.68
HL	13.66	13.40	10.1	8.9	17.6	17.8	1.38	1.67
IND	2.06	2.04	1.4	1.3	3.0	3.4	0.30	0.32
SW	5.33	5.24	3.8	3.5	7.5	7.4	0.67	0.71
END	3.68	3.73	2.7	2.8	4.5	5.4	0.35	0.41
ESD	5.67	5.67	4.2	4.1	7.3	7.4	0.55	0.64
IOD1	6.34	6.37	4.7	4.7	8.4	8.9	0.79	0.86
IOD2	5.47	5.49	3.5	3.5	7.8	9.1	0.74	0.92
ED	4.73	4.56	3.5	3.2	6.4	6.7	0.53	0.61
TD	2.49	2.28	1.4	1.2	3.8	3.5	0.40	0.45
TH	3.18	3.00	1.9	2.1	4.5	4.6	0.45	0.47
EW	4.07	4.00	2.6	2.7	6.5	5.4	0.44	0.49
PGW	9.48	9.63	4.8	6.1	13.8	15.2	1.64	1.74
PGH	7.41	7.35	3.8	4.8	10.7	10.8	1.19	1.30
STCL	2.85	2.80	1.9	1.7	4.1	4.1	0.39	0.50
POS	4.81	4.65	3.3	3.0	6.8	6.9	0.65	0.73
THL	11.88	11.78	8.3	8.0	16.1	17.1	1.56	1.72
TIL	19.21	18.32	12.2	12.0	28.2	29.1	3.32	3.15
TAL	18.18	17.10	11.9	9.5	27.8	26.0	3.05	2.88
HAL	12.17	11.61	8.1	7.2	17.7	17.9	1.91	2.02
FOL	18.68	17.37	12.2	10.8	27.5	25.1	2.75	2.69

Narvaes & Rodrigues 2009

Logo, precisamos de métodos multidimensionais para lidar com uma grande quantidade de dados

- **Objetivos Principais**
  - Reduzir a dimensionalidade dos dados para encontrar as principais estruturas
  - Medir (dis)similaridade entre objetos e descritores
  - Encontrar grupos
- **Análises Exploratórias vs. Testes de Hipótese**
  - Modelagem de uma matriz em função de variáveis explanatórias

Table 2.1. Examples of objects and attributes in ecological matrices.

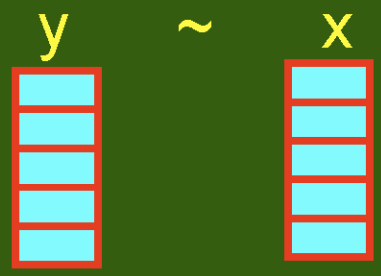
Type of Study	Objects	Attributes
Community analysis	Sample plots Stands Community types	Species Molecular markers Structures or functions Environmental factors Time of sample
Niche-space analysis	Individuals Populations Species Guilds	Resources used or provided Environmental optima, limits, or responses Physicochemical characteristics of resources Habitats
Behavioral analysis	Individuals Populations Species	Activities Response to stimuli Test scores
Taxonomic analysis	Individuals (specimens) Populations Species	Morphological characters Nucleotide positions Isozyme presence Secondary chemicals
Functional or guild analysis	Individuals Populations Species Higher taxa	Life history characteristics Morphology Ecological functions Ecological preferences

McCune & Grace 2002

Análises Multivariadas

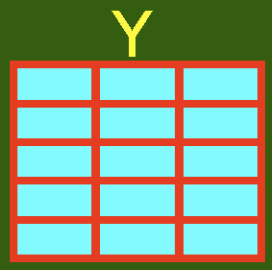
Univariadas --> y ~ x

Regressão, Anova, Reg. Logística



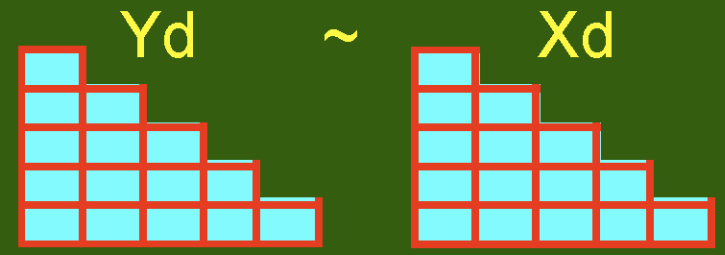
Multivariadas exploratórias --> Y

Ordenação e Classificação



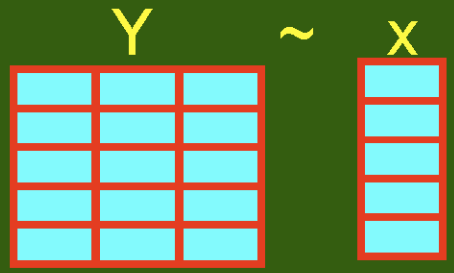
Multivariadas (testes) --> Yd ~ Xd

Teste de Mantel



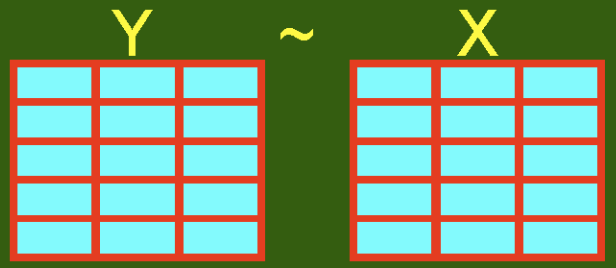
Multivariadas (testes) --> Y ~ x

Manova, db-Manova



Multivariada (ordenações restritas) --> Y ~X

CCA, RDA, CapScale



Slide Adriano S. Melo

Table 5.1 Solutions to multivariate analytical challenges posed by the basic properties of ecological community data (Box 5.1). The various classes of problems and their solutions are explained in the remainder of this book.

Class of problem	Example solution, appropriate for community data	Solutions based on a linear model, usually inappropriate for community data
Measure distances in multidimensional space	Sørensen distance (proportionate city-block distance)	Euclidean distance or correlation-based distance
Test hypothesis of no multivariate difference between two or more groups (one-way classification)	MRPP or Mantel test, using Sørensen distance	one-factor MANOVA
Single factor repeated measures, randomized blocks, or paired sample	blocked MRPP	randomized complete block MANOVA, repeated measures MANOVA
Partition variation among levels in nested sampling	nonparametric MANOVA (=NPMANOVA)	univariate nested ANOVA
Two-factor or multi-factor design with interactions	NPMANOVA	MANOVA
Evaluate species discrimination in one-way classification	Indicator Species Analysis	Discriminant analysis
Extract synthetic gradient (ordination)	Nonmetric multidimensional scaling (NMS) using Sørensen (Bray-Curtis) distance	Principal components analysis (PCA)
Assign scores on environmental gradients to new sample units, on basis of species composition	NMS scores	linear equations from PCA

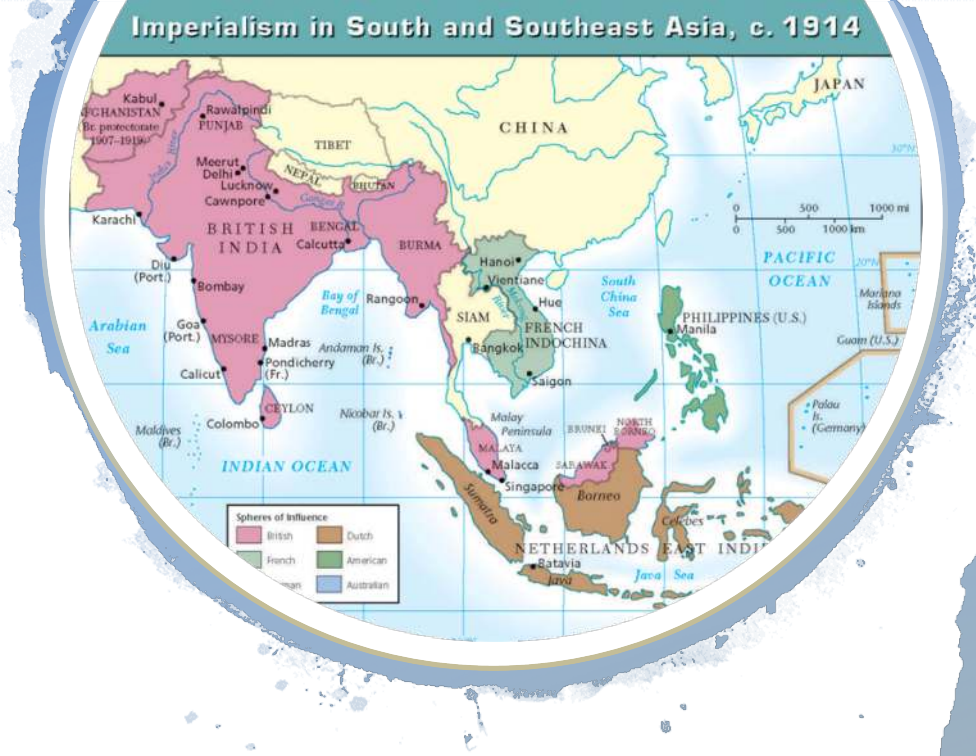
McCune & Grace 2002

Introdução

- Concebidas para análises psicométricas e sociológicas na virada do século XIX para XX.

- Introduzidas na ecologia inicialmente na classificação de comunidades vegetais.


Senta que
lá vem a
história!!



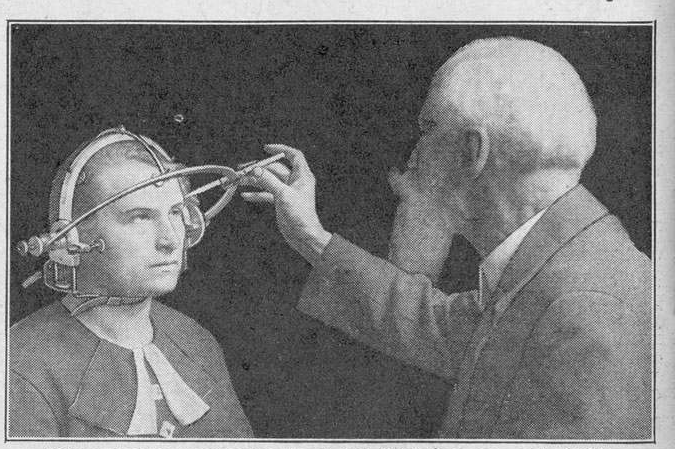
Partilha da África e Ásia

ps://www.nuttyhistory.com/scramble-for-se-asia.html

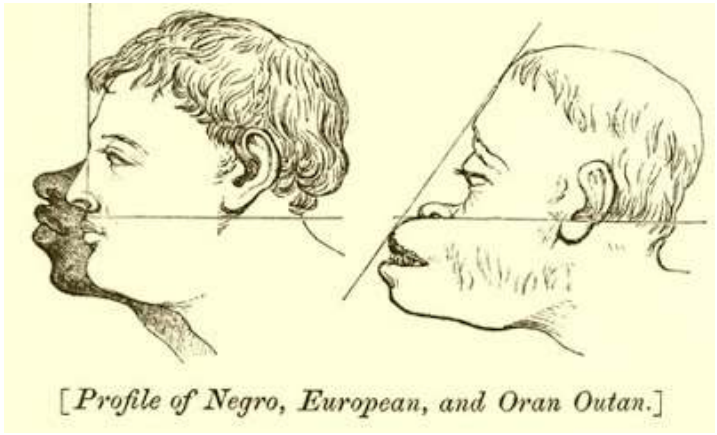
Movimento eugenista e a "justificativa científica" para a dominação da África e Ásia por Europeus no Séc XIX

Measuring the Head to Determine Ability

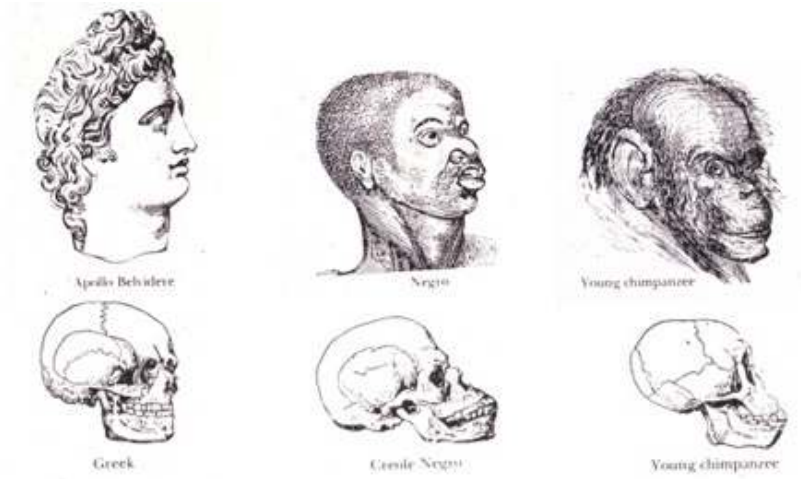
PROF. BURGER-VILLINGER, of the Humboldt College in Berlin, invented the Plastometer shown in the photograph, which is used for determining race symptoms and professional abilities by taking cranial measurements of the subject. The apparatus is fastened to the subject's head, and a three-dimensional scale is provided for obtaining records.



This apparatus measures the head and determines race and ability.

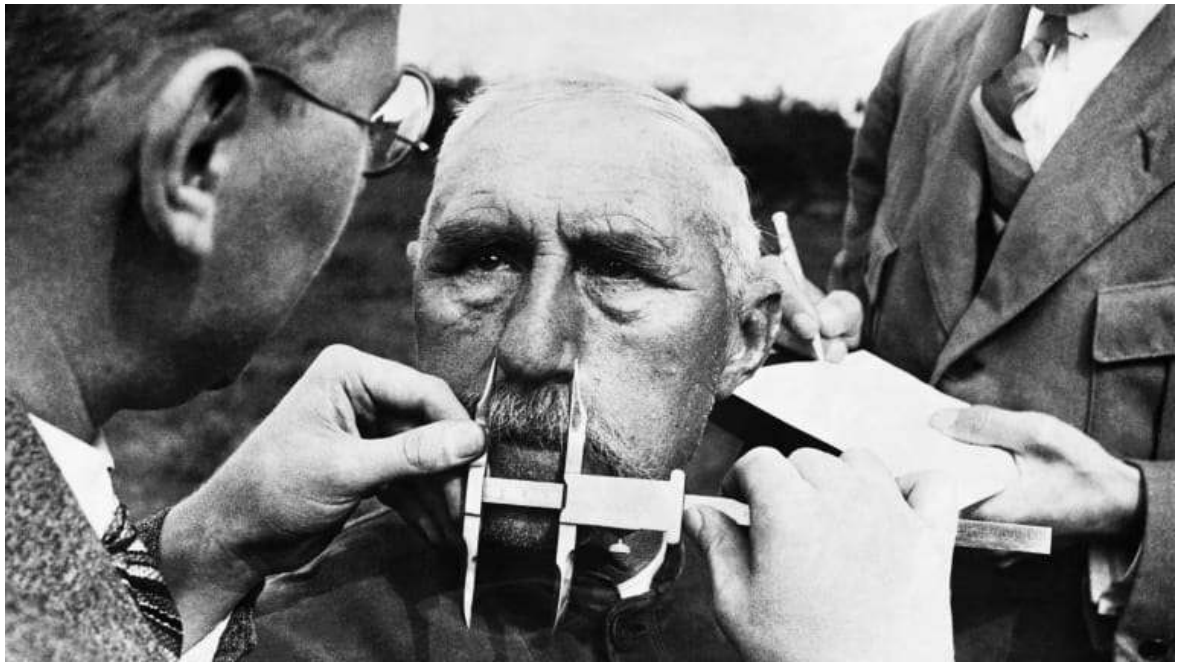


[*Profile of Negro, European, and Oran Outan.*]



http://www.ibamendes.com/2009/12/medindo-cabecas.html
http://www.understandingrace.com/history/science/measuring_race.html

Culminando no genocídio da 2ª Guerra Mundial e o mito da “raça ariana”



https://www.scientificamerican.com/article/neurologists-role-in-nazi-racial-hygiene-only-now-comes-to-light/

Genocídio dos Herero e Namaquas na Namíbia 1904-8

BBC Sign in | News | Sport | Weather | Shop | Reel | Travel | More | Search

NEWS

Home | Video | World | UK | Business | Tech | Science | Stories | Entertainment & Arts | Health | World News TV | More

World | Africa | Asia | Australia | Europe | Latin America | Middle East | US & Canada

Germany returns skulls of Namibian genocide victims

29 August 2018

Share

Leading Herero activist Esther Muinjangue is still demanding an apology from Germany

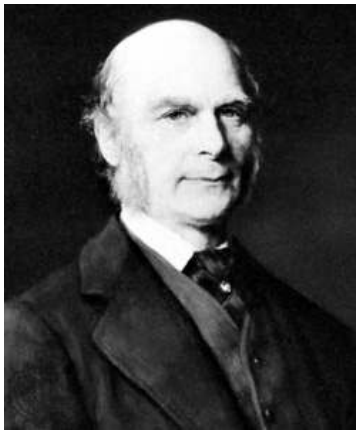
Top Stories

EU reveals no-deal Brexit plans
50 minutes ago

SA issues arrest warrant for Grace Mugabe
16 minutes ago

Elon Musk unveils LA transport tunnel
4 hours ago

Mais um pouco de história



- Francis Galton (1822-1911): Antropólogo inglês, primo de Charles Darwin, defendia idéias de *eugenia*
- Cunhou o termo ”regressão” no contexto de estudos de herdabilidade de caracteres quantitativos em 1886
- Mediu **altura** de filhos e seus pais para investigar o padrão de herdabilidade desta característica

> “the average regression of the offspring is a constant fraction of their respective mid-parental deviations”.

ANTHROPOLOGICAL MISCELLANEA.

Regression *towards* Mediocrity *in* Hereditary Stature.

By Francis Galton, F.R.S., &c.

1886

Assessment in Education: Principles, Policy & Practice
Vol. 19, No. 2, May 2012, 147–158

Routledge
Taylor & Francis Group

Francis Galton, measurement, psychometrics and social progress

Harvey Goldstein*

VII. *Mathematical Contributions to the Theory of Evolution.*—III. *Regression, Heredity, and Panmixia.*

By KARL PEARSON, *University College, London.*

Communicated by Professor HENRICI, *F.R.S.*

Received September 28,—Read November 28, 1895.
Revised November 29, 1895.

Define coeficiente de regressão

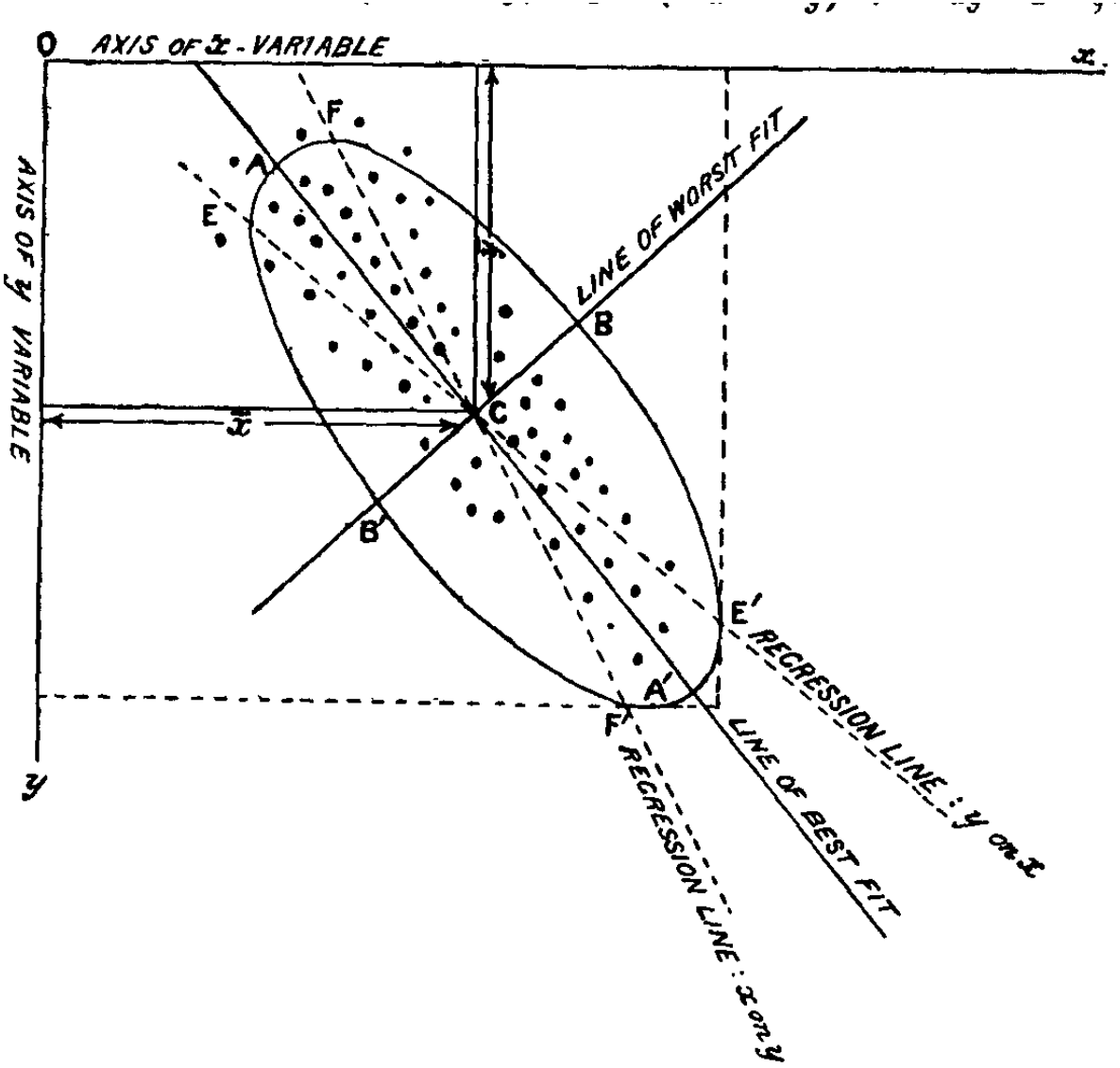
Karl Pearson (1857-1936)

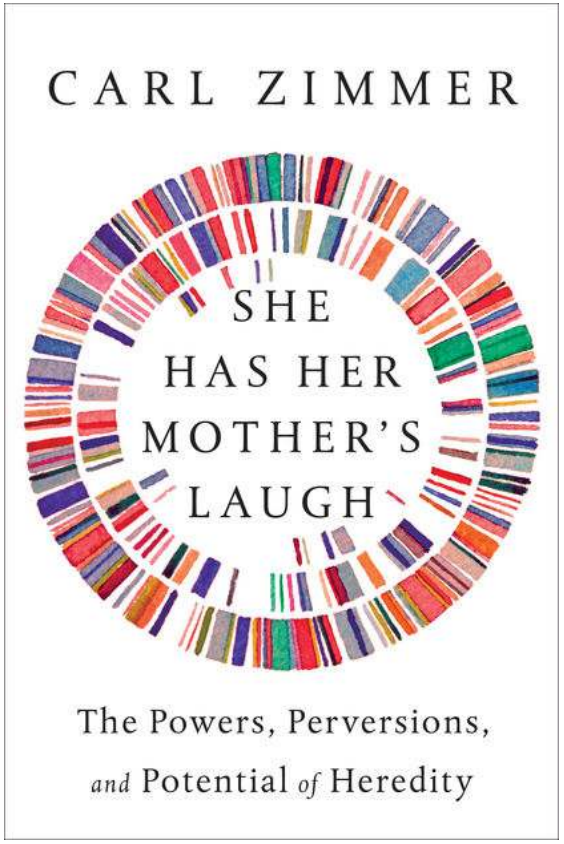
LIII. *On Lines and Planes of Closest Fit to Systems of Points in Space.*

By KARL PEARSON, *F.R.S., University College, London* *.

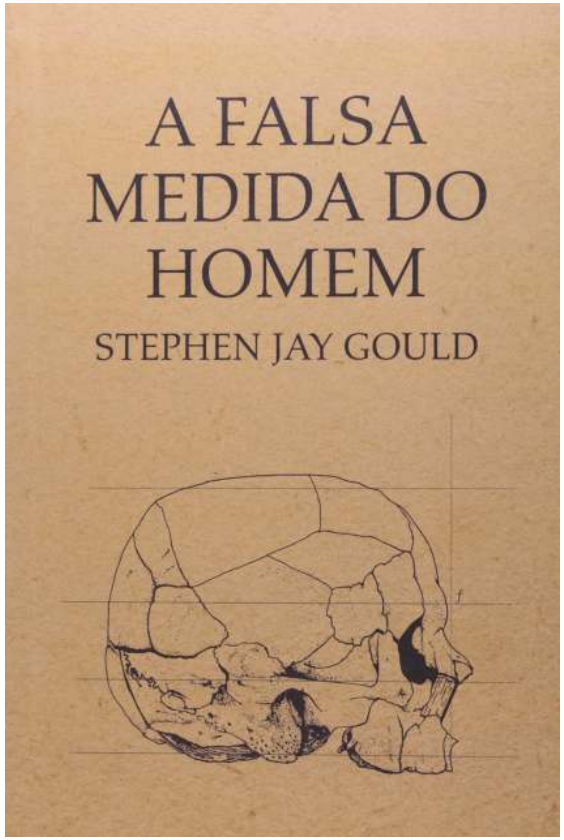
(1) IN many physical, statistical, and biological investigations it is desirable to represent a system of points in plane, three, or higher dimensioned space by the "best-fitting" straight line or plane. Analytically this consists in taking

$$y = a_0 + a_1 x, \quad \text{or} \quad z = a_0 + a_1 x + b_1 y,$$
$$\text{or} \quad z = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots + a_n x_n,$$


O
AXIS OF x - VARIABLE
x
AXIS OF y VARIABLE
y
F
A
E
LINE OF WORST FIT
B
C
B'
E'
REGRESSION LINE : y on x
A'
LINE OF BEST FIT
REGRESSION LINE : x on y


CARL ZIMMER
SHE HAS HER MOTHER'S LAUGH
The Powers, Perversions, and Potential of Heredity


david salsburg
UMA SENHORA TOMA CHÁ...


A FALSA MEDIDA DO HOMEM
STEPHEN JAY GOULD

Introdução

- Concebidas para análises psicométricas e sociológicas.

- Introduzidas na ecologia inicialmente na classificação de comunidades vegetais.

OBJECTIVE METHODS FOR THE CLASSIFICATION OF VEGETATION

III. AN ESSAY IN THE USE OF FACTOR ANALYSIS*

By D. W. GOODALL†

(*Manuscript received April* 5, 1954)

1914-2018 (Eutanásia)

AN ORDINATION OF THE UPLAND FOREST COMMUNITIES OF SOUTHERN WISCONSIN*

J. Roger Bray† and J. T. Curtis
Department of Botany, University of Minnesota, Minneapolis, Minnesota
Department of Botany, University of Wisconsin, Madison, Wisconsin

Ecological Monographs 1957

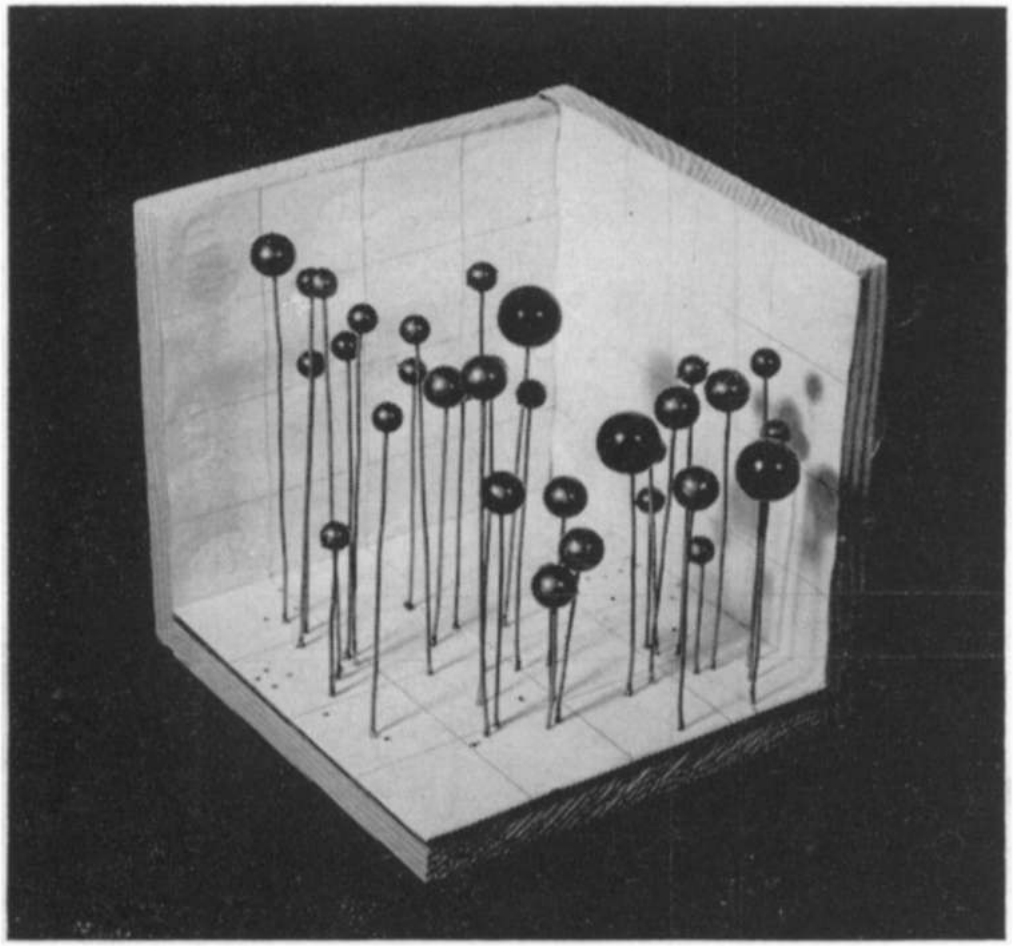


FIG. 7. Three-dimensional model of the dominance behavior of *Quercus borealis* within the ordination. The 3 sizes of spheres indicate the top 3 quartiles of dominance per acre. Stands of the lowest quartile and without the species are represented with holes which appear as dots in the figure. The x axis is on base of model at front from left to right; y axis on base from front to rear; z axis in vertical plane from below to above.

Ordenação Polar

1º método de ordenação para dados de contagem

http://ecovirtual.ib.usp.br/doku.php?id=ecovirt:roteiro:comuni:comuni_order

Biol. Rev. (1967), **49**, *pp.* 207–264

GRADIENT ANALYSIS OF VEGETATION*

By

R. H. WHITTAKER

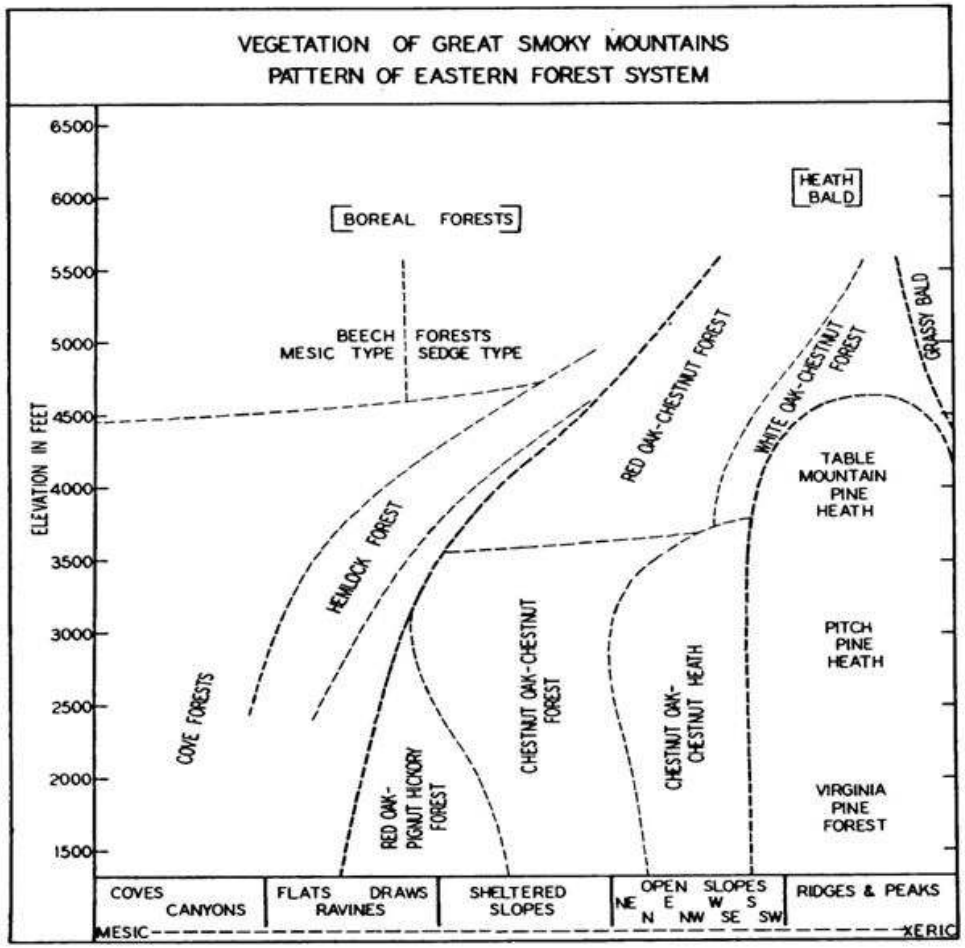
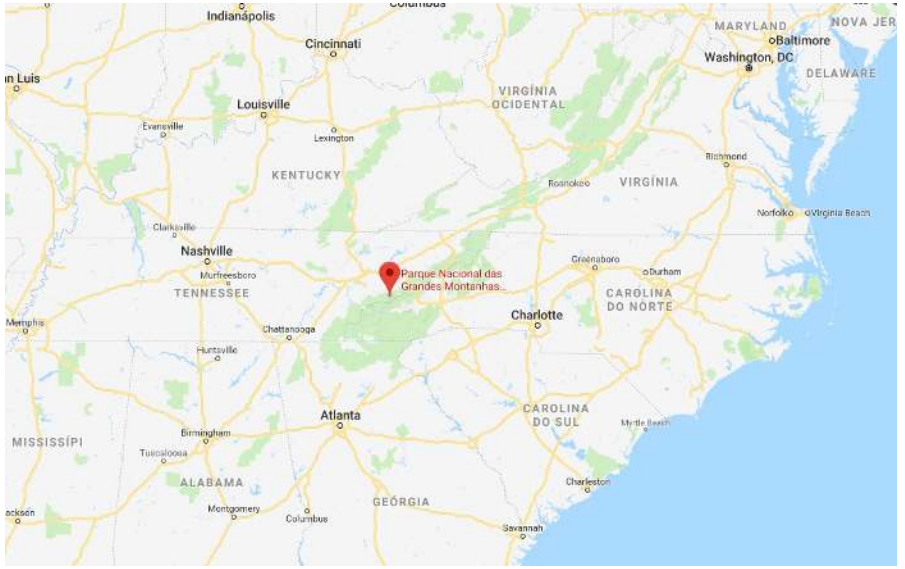


FIG. 19. (Vegetation of Great Smoky Mountains, pattern of Eastern Forest System.)



From Whittaker 1956

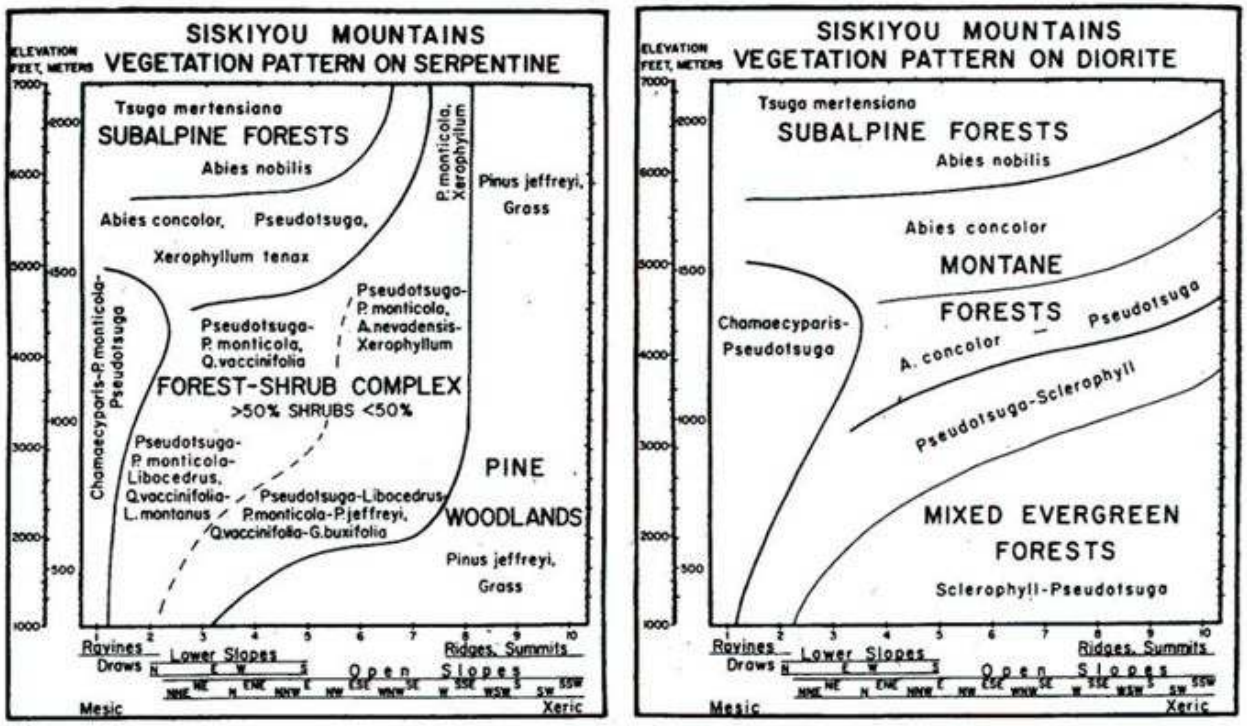
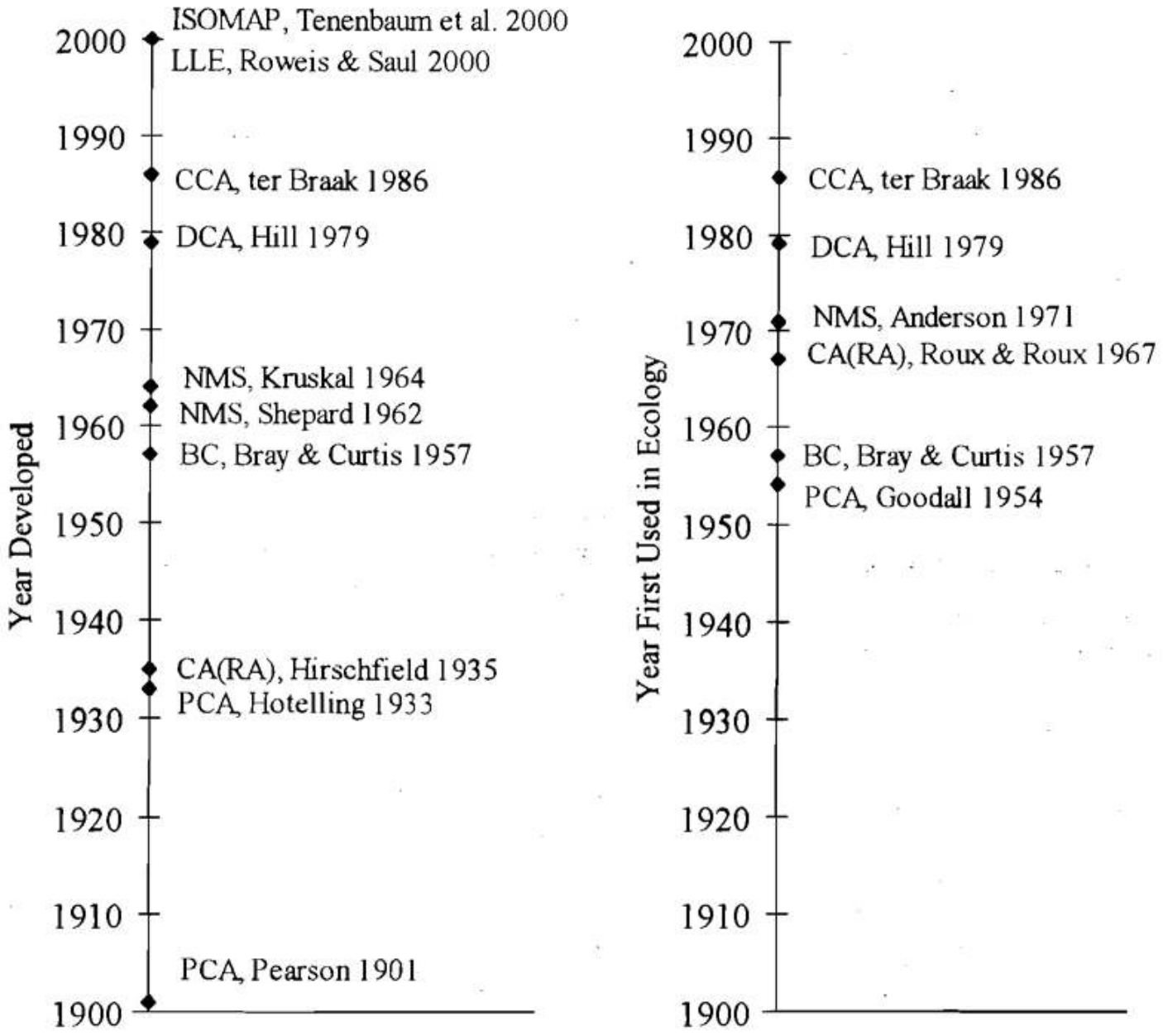


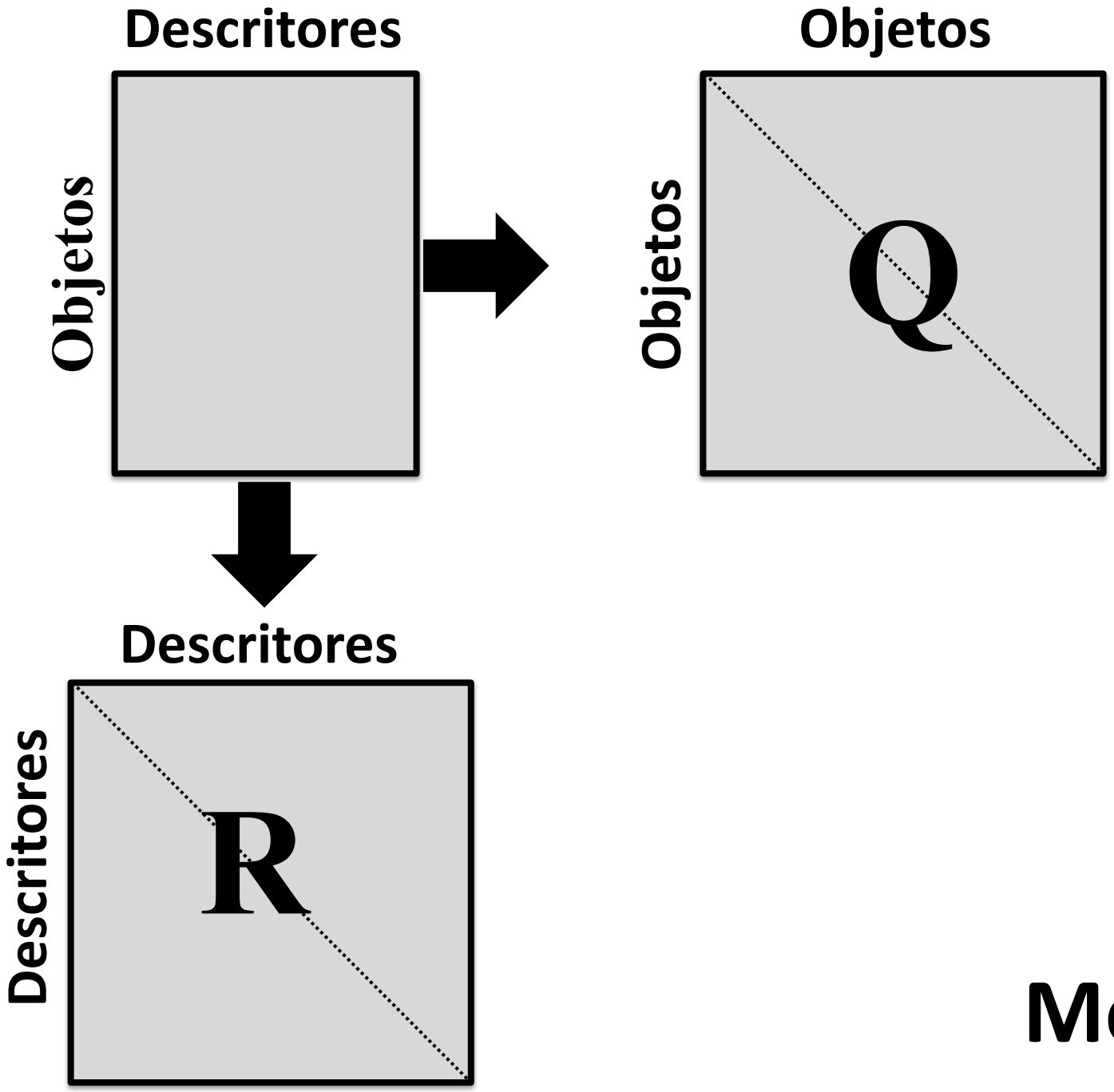
Fig. 5. Mosaic diagrams for vegetation on quartz diorite (right) and serpentines (left) in the Siskiyou Mountains, southwestern Oregon (WHITTAKER 1960, cf. WHITTAKER 1956, WHITTAKER & NIERING 1965, 1968b). Vegetation samples were classified (into dominance-types) and plotted by elevation and topographic position on the chart. Boundaries for community-types were drawn at average positions of transitions between types. Vegetation on each parent material is considered a complex pattern of continuously intergrading communities (WHITTAKER 1960).



McCune & Grace 2002


DEU POR HOJE!!!
CHEGA DE HISTÓRIA!!
GERADORMEMES.COM

Como os métodos funcionam de forma geral



Modo R e Q

As duas matrizes típicas

Espécies

	Sp1	Sp2	Sp3	Sp4	Spn
Área 1	2	12	14	0	2
Área 2	10	0	11	0	15
Área 3	18	19	10	1	19
Área 4	6	21	9	0	8
Área 5	0	0	1	18	0
Área 6	0	0	2	9	0
Área 7	0	0	0	19	0
Área 8	1	0	1	21	0

Variáveis ambientais

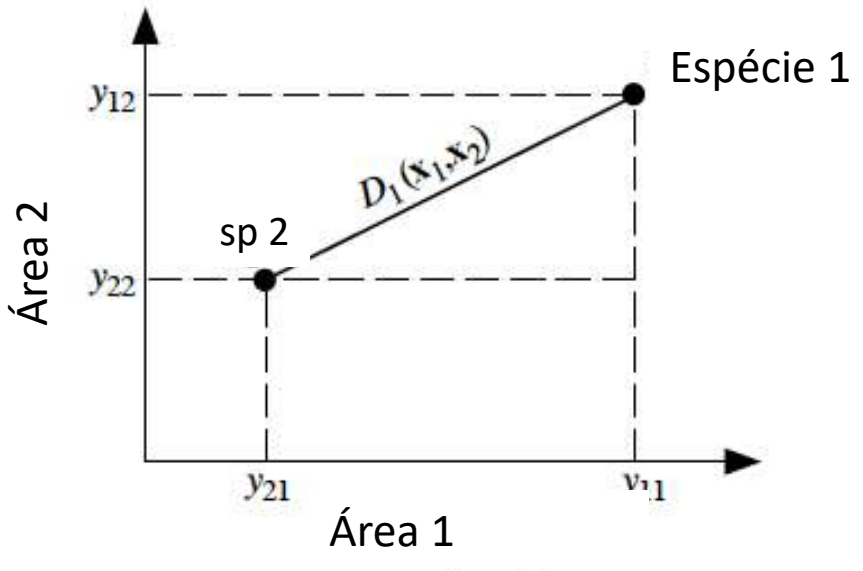
Temp	pH	O_2	CO_2
2	12	14	0
10	0	11	0
18	19	10	1
6	21	9	0
0	0	1	18
0	0	2	9
0	0	0	19
1	0	1	21

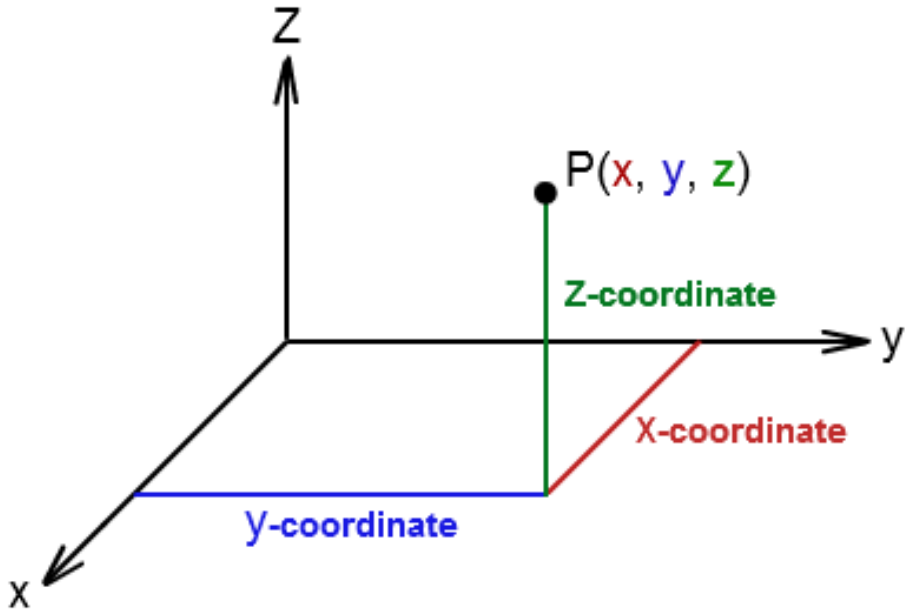
Comecemos por uma matriz

Espécies

	Sp1	Sp2	Sp3	Sp4	Spn
Área 1	2	12	14	0	2
Área 2	10	0	11	0	15
Área 3	18	19	10	1	19
Área 4	6	21	9	0	8
Área 5	0	0	1	18	0
Área 6	0	0	2	9	0
Área 7	0	0	0	19	0
Área 8	1	0	1	21	0

Representação no Espaço Euclidiano

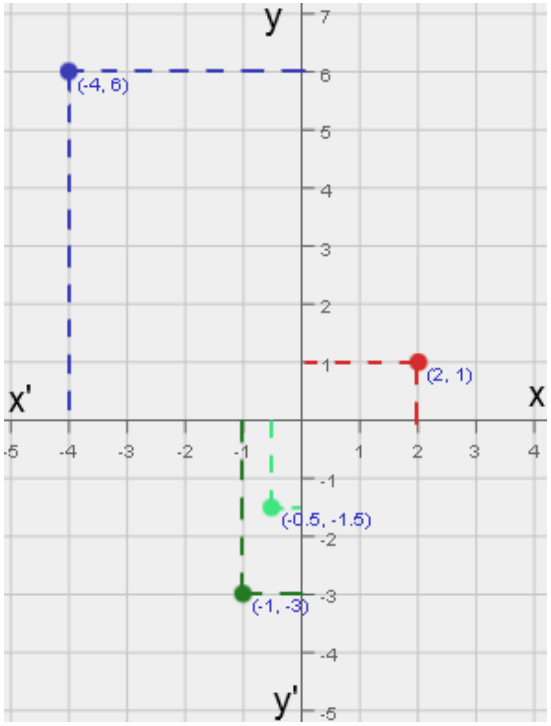




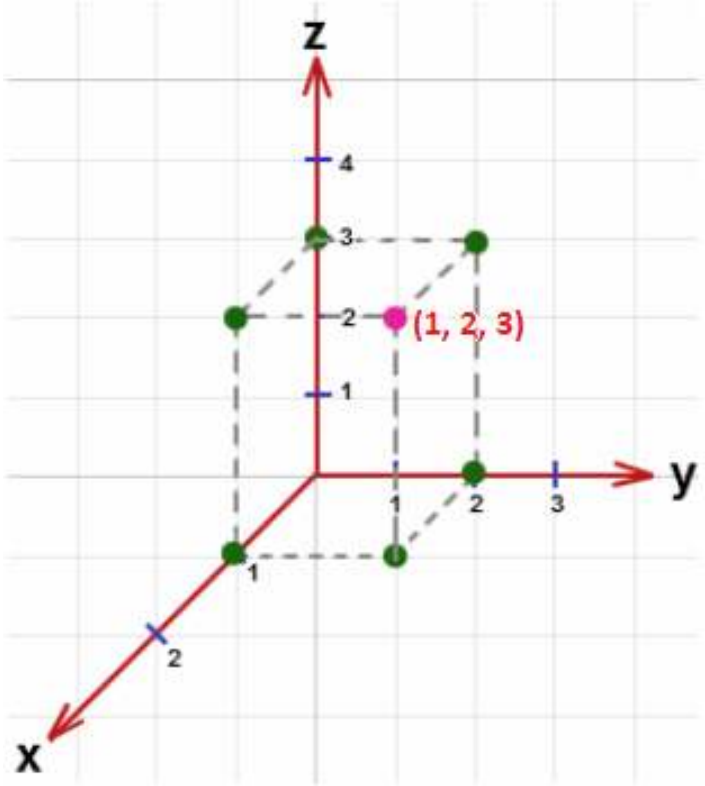
3D

Legendre & Legendre 2012

Medidas de distância no espaço Euclidiano



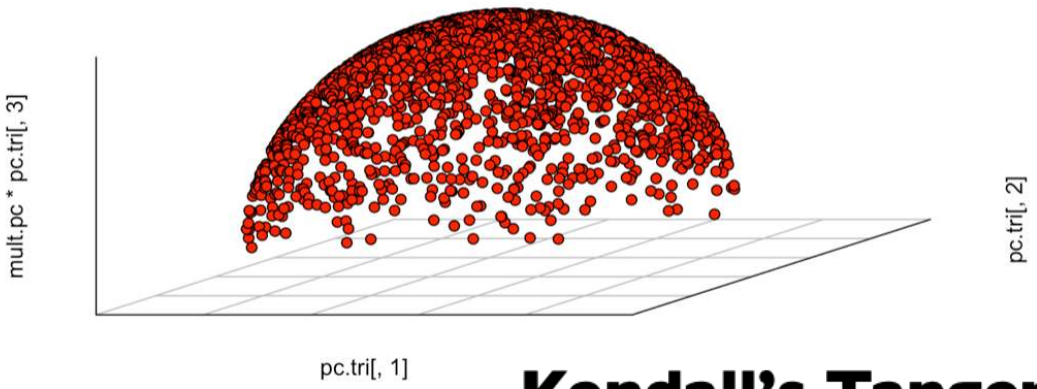
2D



3D

https://math.tutorvista.com/algebra/Euclidean-space.html

- **Shape space is curved!**

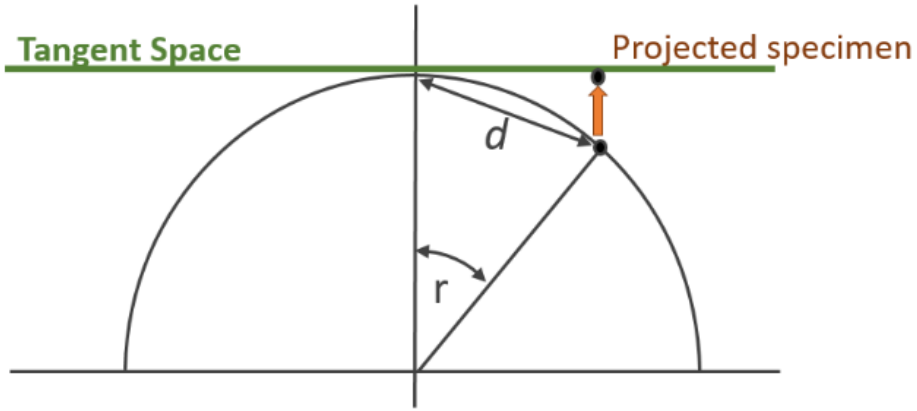


Um pequeno parênteses

Kendall's Tangent Space Coordinates

- Orthogonal projection of shapes to linear tangent space

$$\mathbf{X}' = \mathbf{X}\left(\mathbf{I_{kp}} - \mathbf{X_c^T}(\mathbf{X_c}\mathbf{X_c^T})^{-1}\mathbf{X_c}\right)$$



Slides Antigoni Kaliantzoupolous

Rudimentos de álgebra de matrizes

Combinações lineares

- Combinação de várias variáveis (vetores) que são multiplicadas por constantes e adicionadas a outras variáveis

- Por exemplo: combinação linear de x, y, z pode ser escrita como

$$ax+by+cz$$

Por que isso é importante?

- Métodos de ordenação basicamente encontram combinações lineares, maximizando as constantes
- Ou no caso de ordenações restritas (duas ou mais matrizes), restringem uma combinação linear por outra combinação linear
- Diferentes métodos maximizam diferentes coeficientes/constantes. Por exemplo, PCA encontra correlação linear entre as variáveis

Correlação de Pearson

Coeficiente de Correlação

$$r_{Y_1 Y_2} = \frac{\sum_{i=1}^{n} [(y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)]}{\sqrt{\sum_{i=1}^{n} (y_{i1} - \bar{y}_1)^2 \sum_{i=1}^{n} (y_{i2} - \bar{y}_2)^2}}$$

Relação linear entre duas variáveis, mas sem assumir uma dependência funcional entre as duas

$-1 < r > 1$

Coeficiente de Correlação

$$r_{Y_1 Y_2} = \frac{\sum_{i=1}^{n} [(y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)]}{\sqrt{\sum_{i=1}^{n} (y_{i1} - \bar{y}_1)^2 \sum_{i=1}^{n} (y_{i2} - \bar{y}_2)^2}}$$

Soma dos produtos cruzados (Covariância)

Desvio padrão das duas variáveis

Coeficiente de Correlação

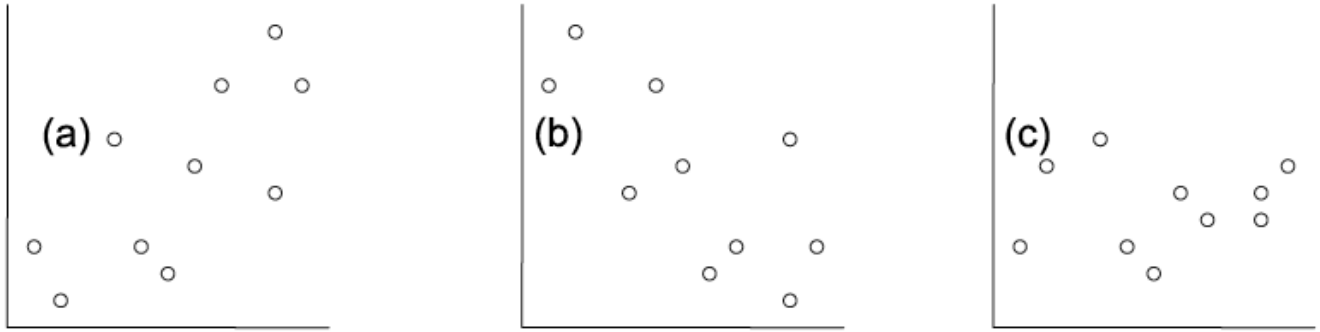
$$r_{Y_1 Y_2} = \frac{\sum_{i=1}^{n}[(y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)]}{\sqrt{\sum_{i=1}^{n}(y_{i1} - \bar{y}_1)^2 \sum_{i=1}^{n}(y_{i2} - \bar{y}_2)^2}}$$

Pode ser positivo ou negativo

Sempre positivo

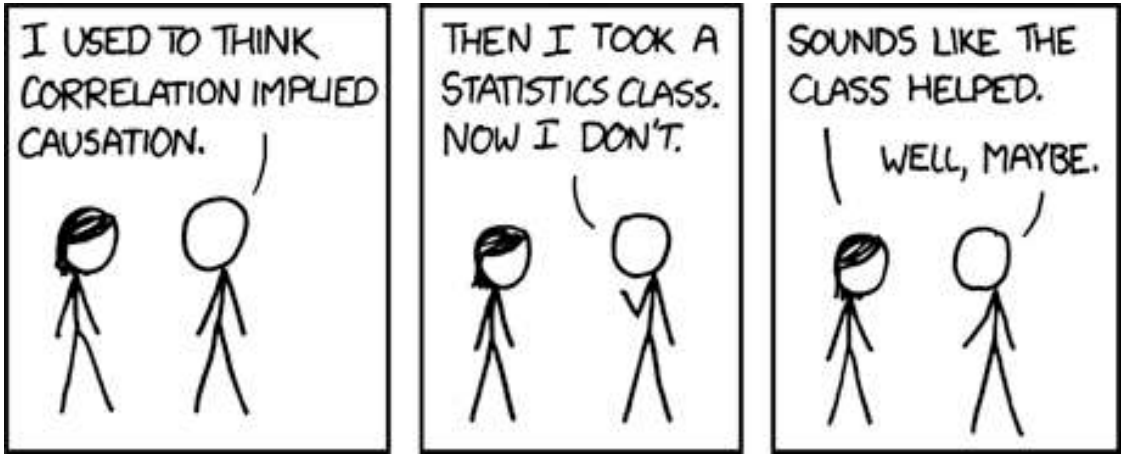
Se o numerador for negativo, o r será negativo e vice-versa

Figure 5.2 Scatterplots illustrating (a) a positive linear relationship ($r = 0.72$), (b) a negative linear relationship ($r = -0.72$), (c) and (d) no relationship ($r = 0.10$ and -0.17), respectively, and (e) a nonlinear relationship ($r = 0.08$).



Quinn & Keough 2002

CORRELATION
DOES NOT IMPLY
CAUSATION



Teste de hipótese do r de Pearson

- Utiliza um teste t:

$$t = \frac{r}{s_r}$$

Erro padrão do *r*

$$s_r = \sqrt{\frac{(1 - r^2)}{(n - 2)}}$$

- Testa a hipótese nula de que $\rho = 0$

Matriz de correlação Matriz de covariância

$$\mathbf{P} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) & \mathrm{Cov}(X,Z) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) & \mathrm{Cov}(Y,Z) \\ \mathrm{Cov}(X,Z) & \mathrm{Cov}(Y,Z) & \mathrm{Var}(Z) \end{bmatrix}$$

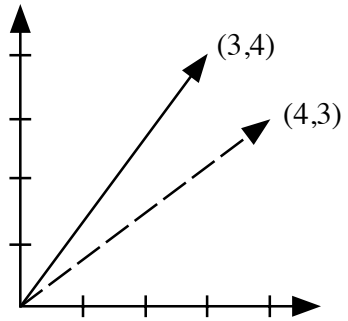
Mais alguns conceitos importantes

- Escalar = número inteiro com o qual geralmente se faz operações com matrizes (e.g., multiplicação ou adição)
  - Pode também multiplicar vetores, fazendo com que aumentem ou diminuam de tamanho
- Vetor = coluna de uma matriz
- Autovetor = vetor não nulo que muda somente quando é multiplicado por um escalar
- Autovalor = número inteiro (escalar) que multiplica um vetor, sendo portanto múltiplo deste

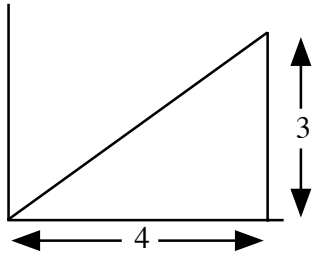
Vetores

A (column) vector is noted as follows:

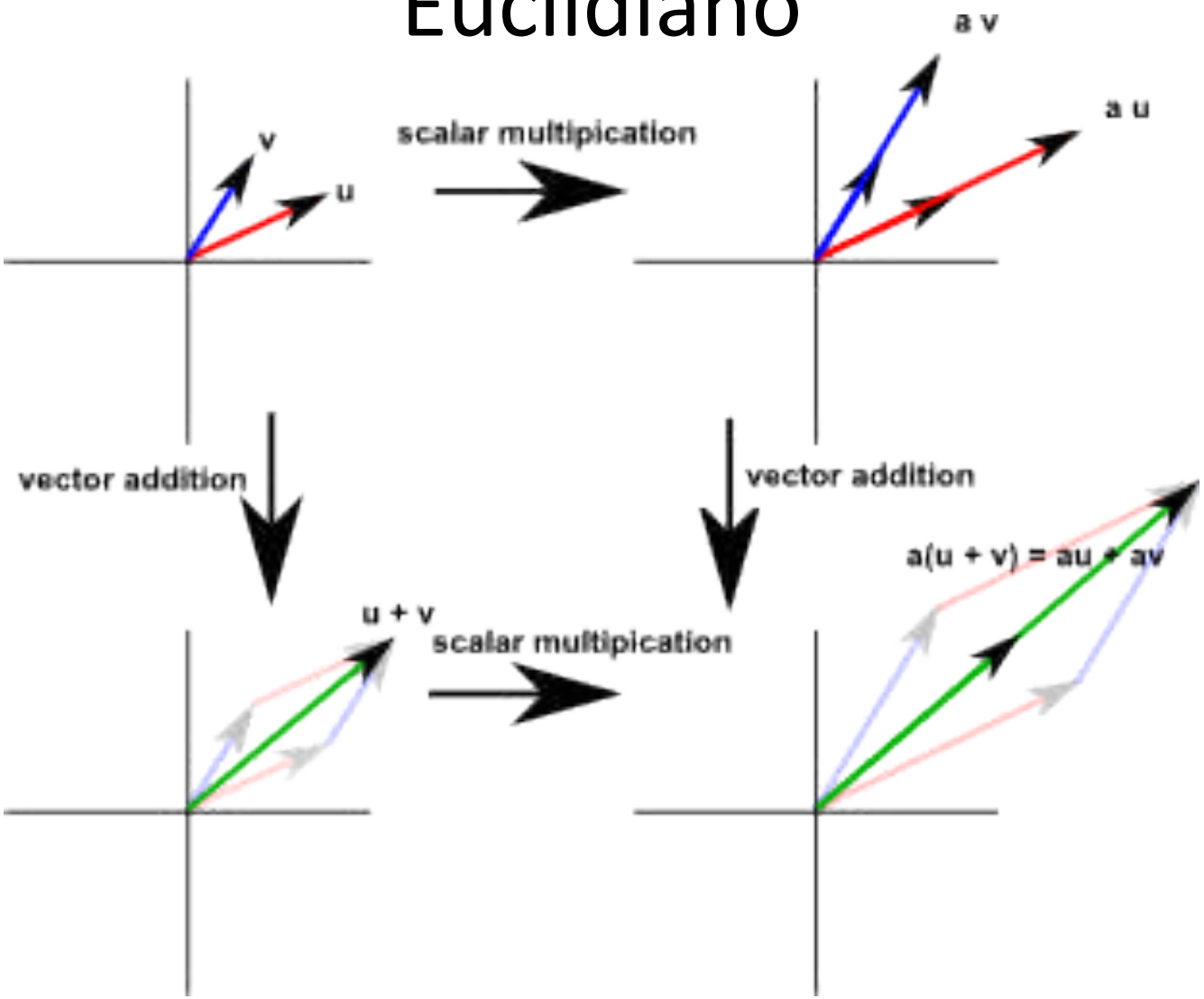
$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$$



Using the Pythagorean theorem, it is easy to calculate the length of any vector. For example, the length of vector [4 3]' is that of the hypotenuse of a right triangle with base 4 and height 3:



Operações com vetores no espaço Euclidiano



http://www.euclideanspace.com/maths/geometry/space/euclidean/index.htm

Autovetores e Autovalores

O autovetor de uma matriz é encontrado pelo resultado da multiplicação de um vetor pela matriz, que é igual a ƛ vezes o vetor. Esse vetor do resultado passa a ser o autovetor, e o ƛ o seu respectivo autovalor.

Portanto, os autovetores e autovalores da matriz **A** podem ser encontrados com a seguinte equação:

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$


poderia ser legal,e interessante...
mais não foi.
vk.com/yoda_advice

Autovetores e Autovalores

- A derivação desta equação garante que os autovetores produzidos são ortogonais entre si
- Ortogonais (90º) = variam em direções independentes

- Para mais detalhes veja p. 92-4 L&L


num intendi nada que
você disse ai fera

Multiplicação de matrizes

■ **Exemplo 1.17** Determine AB, dados

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$$

Inicialmente note que A tem 3 colunas e B tem 3 linhas, deste modo, existe o produto AB; para calculá-lo, multiplicamos cada *linha* de A por cada *coluna* de B e somamos os resultados destas multiplicações:

$$\begin{aligned} AB &= \begin{bmatrix} 1^{\text{a}} \text{ linha de } A \text{ vezes } 1^{\text{a}} \text{ coluna de } B \\ 2^{\text{a}} \text{ linha de } A \text{ vezes } 1^{\text{a}} \text{ coluna de } B \end{bmatrix} \\ &= \begin{bmatrix} 1 \cdot 7 + 2 \cdot 8 + 3 \cdot 9 \\ 4 \cdot 7 + 5 \cdot 8 + 6 \cdot 9 \end{bmatrix} \\ &= \begin{bmatrix} 7 + 16 + 27 \\ 28 + 40 + 63 \end{bmatrix} \\ &= \begin{bmatrix} 50 \\ 131 \end{bmatrix} \end{aligned}$$

■

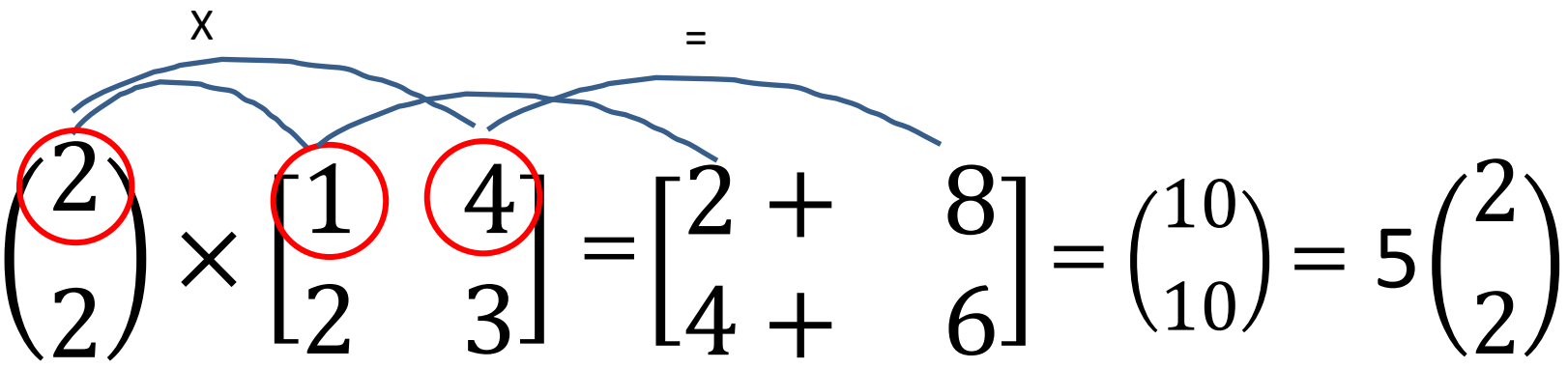
https://books.google.com.br/books?id=Spi2BAAAQBAJ&printsec=frontcover&dq=matemática+ensino+médio&hl=pt-BR&sa=X&ved=0ahUKEwi0oKLVuunhAhUXH7kGHZFHDiUQ6AEIKTAA#v=onepage&q=matemática%20ensino%20médio&f=false

Um exemplo ajuda?

Aqui temos um vetor (2 2) que
será multiplicado pela seguinte matriz:

$$\begin{pmatrix} 2 \\ 2 \end{pmatrix} \times \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$$

Um exemplo ajuda?



Um exemplo ajuda?

O vetor inicial é o mesmo final. O número 5 é então um escalar (número inteiro). Dizemos então que o número 5 é um autovalor e o vetor (2 2) é um autovetor

$$\begin{pmatrix}2\\2\end{pmatrix}\times\begin{bmatrix}1 & 4\\2 & 3\end{bmatrix}=\begin{bmatrix}2+ & 8\\4+ & 6\end{bmatrix}=\begin{pmatrix}10\\10\end{pmatrix}=5\begin{pmatrix}2\\2\end{pmatrix}$$

http://setosa.io/ev/eigenvectors-and-eigenvalues/

Calcular os autovetores e autovalores de uma matriz é equivalente a fazer auto-análise da matriz

https://en.wikipedia.org/wiki/Eigendecomposition_of_a_matrix#Eigendecomposition_of_a_matrix

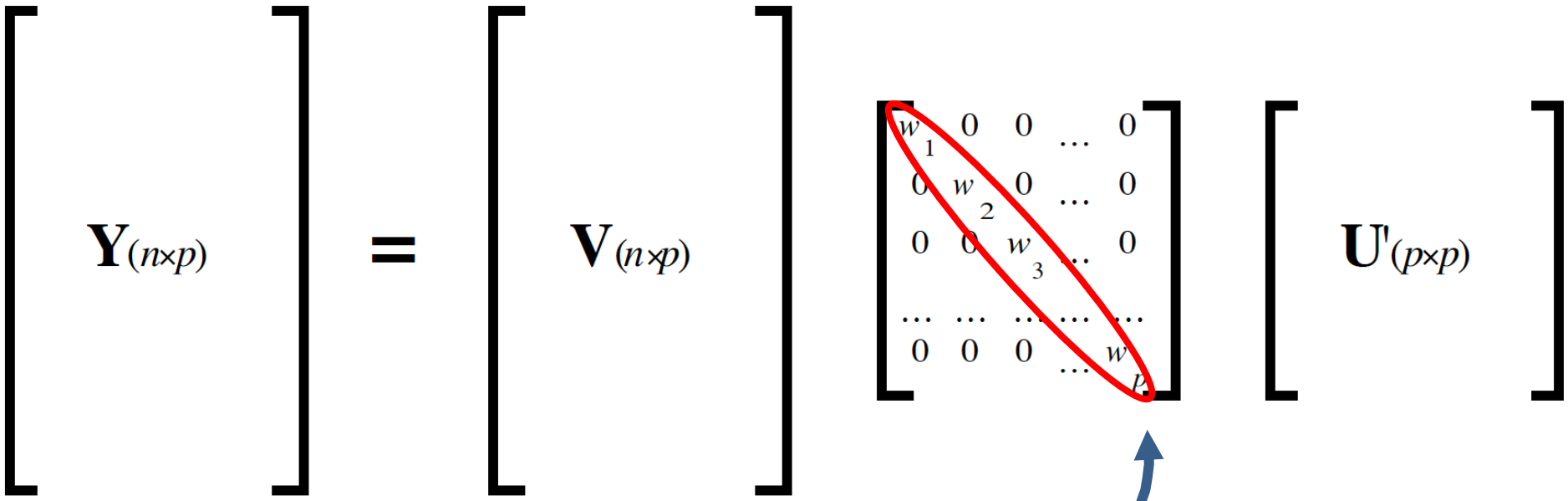
https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors#Eigenvalues_and_eigenvectors_of_matrices

Outra forma de decompor uma matriz é usar Singular Value Decomposition

$$\mathbf{Y}(n\times p) = \mathbf{V}(n\times k)\ \mathbf{W}(\text{diagonal}, k\times k)\ \mathbf{U}'(k\times p)$$

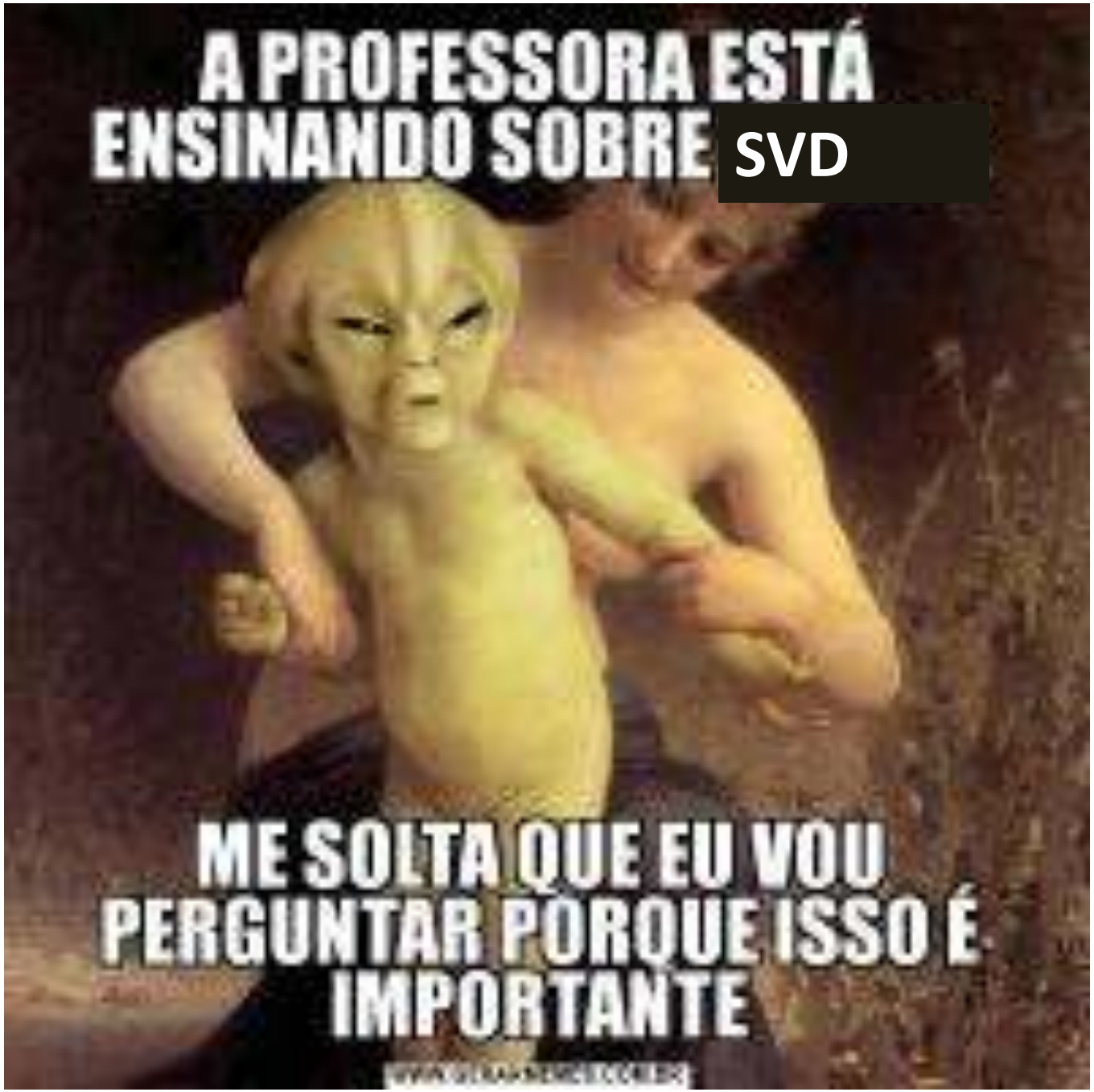
$$\begin{bmatrix} \\ \mathbf{Y}_{(n\times p)} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{V}_{(n\times p)} \\ \\ \end{bmatrix} \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & w_p \end{bmatrix} \begin{bmatrix} \\ \mathbf{U}'_{(p\times p)} \\ \\ \end{bmatrix}$$

$$\mathbf{Y}(n\times p) = \mathbf{V}(n\times k)\ \mathbf{W}(\text{diagonal}, k\times k)\ \mathbf{U}'(k\times p)$$



Valores singulares da matriz **Y**

(rank da matriz = no. Valores singulares > 0)


A PROFESSORA ESTÁ ENSINANDO SOBRE
SVD
ME SOLTA QUE EU VOU PERGUNTAR PORQUE ISSO É IMPORTANTE

Vários métodos de ordenação utilizam SVD, por exemplo, Correspondência Canônica, e também pode ser usado em PCA

Be Happy 'Cause Its Lunch Time

KeepCalmAndPosters.com