

Estatística aplicada à Biologia

Aula 2 – Organização e apresentação de dados

Regras básicas para planilhar dados

- Observações nas linhas
 - Uma unidade amostral em cada linha
- Variáveis nas colunas
 - Tratamentos experimentais, variáveis preditoras

Anatomia de uma tabela

Exemplo 1.5 Apresentação de dados em tabela

Tabela 1.1

População residente no Brasil, segundo o sexo, de acordo com o Censo Demográfico 2010

Sexo	População residente
Homens	93.406.990
Mulheres	97.348.809
Total	190.755.799

Fonte: Censo Demográfico 2010. IBGE (2011).²

²Disponível em: <<http://www.ibge.gov.br>>.

Exemplo 1.5 Apresentação de dados em tabela

Número

Tabela 1.1

População residente no Brasil, segundo o sexo, de acordo com o Censo Demográfico 2010

Sexo	População residente
Homens	93.406.990
Mulheres	97.348.809
Total	190.755.799

Fonte: Censo Demográfico 2010. IBGE (2011).²

²Disponível em: <<http://www.ibge.gov.br>>.

Exemplo 1.5 Apresentação de dados em tabela

Tabela 1.1

População residente no Brasil, segundo o sexo, de acordo com o Censo Demográfico 2010

Sexo	População residente
Homens	93.406.990
Mulheres	97.348.809
Total	190.755.799

Fonte: Censo Demográfico 2010. IBGE (2011).²

²Disponível em: <<http://www.ibge.gov.br>>.

 **Título**

Exemplo 1.5 Apresentação de dados em tabela

Tabela 1.1

População residente no Brasil, segundo o sexo, de acordo com o Censo Demográfico 2010

Cabeçalho →

Sexo	População residente
Homens	93.406.990
Mulheres	97.348.809
Total	190.755.799

Fonte: Censo Demográfico 2010. IBGE (2011).²

²Disponível em: <<http://www.ibge.gov.br>>.

Exemplo 1.5 Apresentação de dados em tabela

Tabela 1.1

População residente no Brasil, segundo o sexo, de acordo com o Censo Demográfico 2010

Sexo	População residente
Homens	93.406.990
Mulheres	97.348.809
Total	190.755.799

Linha => Observação



Fonte: Censo Demográfico 2010. IBGE (2011).²

²Disponível em: <<http://www.ibge.gov.br>>.

Exemplo 1.5 Apresentação de dados em tabela

Tabela 1.1

População residente no Brasil, segundo o sexo, de acordo com o Censo Demográfico 2010

Sexo	População residente
Homens	113.166.990
Mulheres	97.348.809
Total	190.755.799

Colunas => variáveis

Fonte: Censo Demográfico 2010. IBGE (2011).²

²Disponível em: <<http://www.ibge.gov.br>>.

Exemplo 1.5 Apresentação de dados em tabela

Tabela 1.1

População residente no Brasil, segundo o sexo, de acordo com o Censo Demográfico 2010

Sexo	População residente
Homens	93.406.990
Mulheres	97.348.809
Total	190.755.799

Célula => um único dado

Fonte: Censo Demográfico 2010. IBGE (2011).²

²Disponível em: <<http://www.ibge.gov.br>>.

Exemplo 1.5 Apresentação de dados em tabela

Tabela 1.1

População residente no Brasil, segundo o sexo, de acordo com o Censo Demográfico 2010

Sexo	População residente
Homens	93.406.990
Mulheres	97.348.809
Total	190.755.799

Fonte: Censo Demográfico 2010. IBGE (2011).²

Fonte (apenas caso não tenha sido produzida pelo autor)

²Disponível em: <<http://www.ibge.gov.br>>.

Tabela para dados qualitativos (Frequência)

Tabela 2: Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB segundo o grau de instrução

Grau de instrução	Frequência (n_i)	Proporção (fr_i)	Porcentagem ($100fr_i$)
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67

Fonte: Tabela 1 - Bussab e Morettin(2002)

Tabela 2: Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB segundo o grau de instrução

Grau de instrução	Frequência (n_i)	Proporção (fr_i)	Porcentagem ($100fr_i$)
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67

Fonte: Tabela 1 - Bussab e Morettin(2002)

Uma única variável

Tabela 2: Frequências e porcentagens dos 36 empregados da seção de orçamentos da Companhia MB segundo o grau de instrução

Grau de instrução	Frequência (n_i)	Proporção (fr_i)	Porcentagem ($100fr_i$)
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67

Tabela de contingência

Tabela 1.5

Raça ou etnia das voluntárias segundo o grupo

Raça ou grupo étnico	Grupo	
	Intervenção	Rotina
Branca	356	396
Asiática	92	72
Outro	42	42

Fonte: Crowther, CA *et alii*. Effect of Treatment of Gestational Diabetes Mellitus on Pregnancy Outcomes. N Engl J Med 2005; 352:2.477-2.486, June 16, 2005.

Tabela de contingência

Tabela 1.6

Raça ou etnia das voluntárias segundo o grupo

Raça ou grupo étnico	Grupo		Total
	Intervenção	Rotina	
Branca	356	396	752
Asiática	92	72	164
Outro	42	42	84
Total	490	510	1.000

Duas variáveis

Fonte: Crowther, CA *et alii*. Effect of Treatment of Gestational Diabetes Mellitus on Pregnancy Outcomes. N Engl J Med 2005; 352: 2.477-2.486, June 16, 2005.

Dica: Usadas em testes de
Qui-quadrado e teste G

Tabela de frequência para dados ordinais

Tabela 1.3

Opinião dos brasileiros sobre o técnico de futebol

Resposta	Frequência
Ótimo ou bom	1.075
Regular	1.506
Péssimo	2.635
Não sabe/ não respondeu	161
Total	5.377

Fonte: dimassantos.com.br/pesquisa-aponta-tite-para-futuro-tecnico-da-selecao. Acesso em: Setembro de 2014.

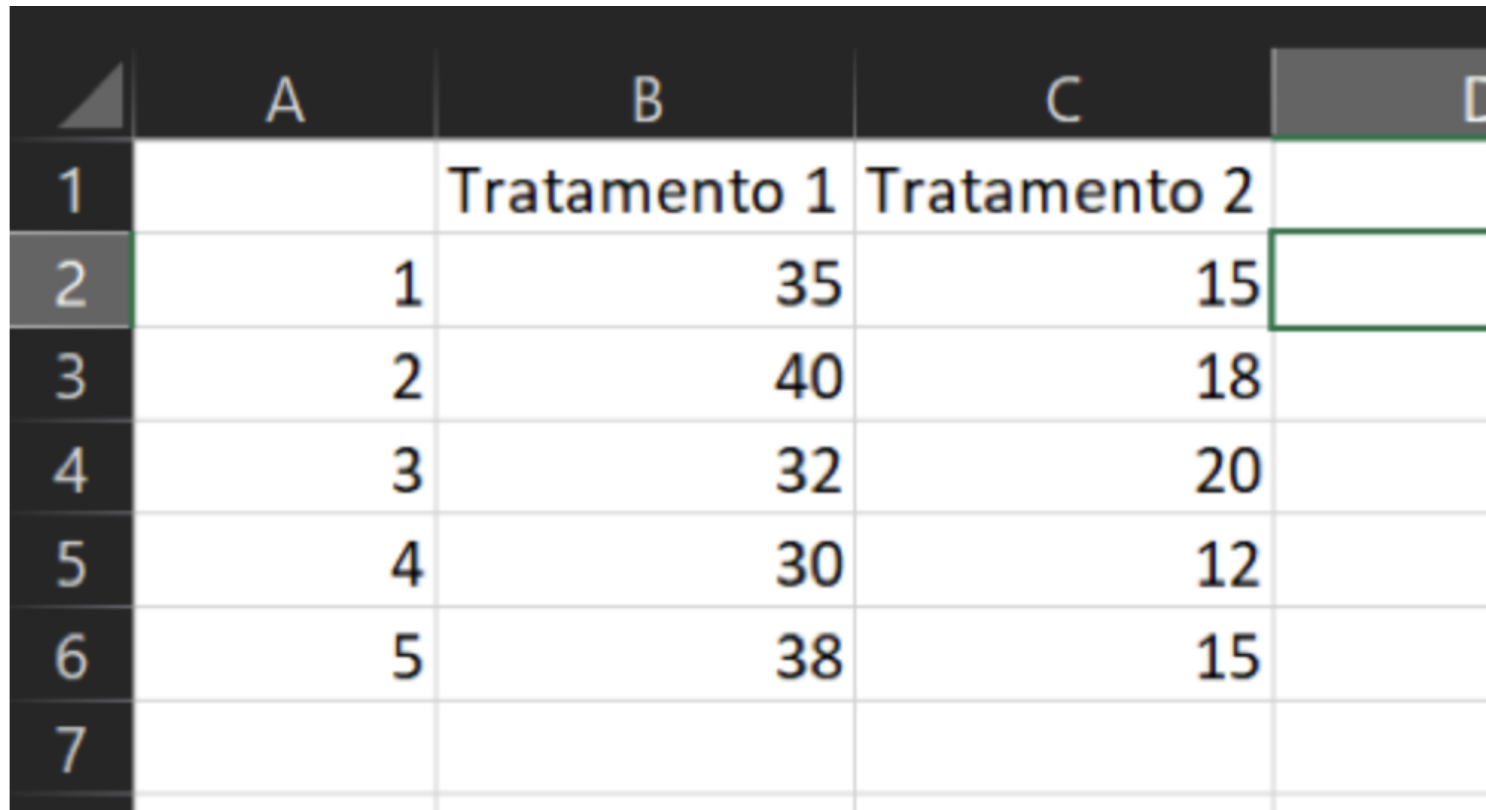
Tabela para dados quantitativos

Tabela 1: Informações sobre estado civil, grau de instrução, número de filhos, salário (expresso como fração do salário mínimo), idade (medida em anos e meses) e procedência de 36 empregados da seção de orçamentos da Companhia MB.

N ^o	Estado civil	Grau de instrução	N ^o de filhos	Salário (× sal. mín.)	Idade		Região de procedência
					anos	meses	
1	Solteiro	ensino fundamental	—	4	26	3	Interior
2	Casado	ensino fundamental	1	4,56	32	10	Capital
3	Casado	ensino fundamental	2	5,25	36	5	Capital
4	Solteiro	ensino médio	—	5,73	21	10	Outra
5	Solteiro	ensino fundamental	—	6,26	41	7	Outra
.	
.	
.	
36	Casado	superior	3	23,3	42	2	Interior

Fonte: Bussab e Morettin (2002)

Como não planilhar dados



	A	B	C	D
1		Tratamento 1	Tratamento 2	
2	1	35	15	
3	2	40	18	
4	3	32	20	
5	4	30	12	
6	5	38	15	
7				

Fig 1. O terror de toda pessoa que precisa analisar dados alheios: uma planilha mal feita em formato Excel.

	A	B	
1	grupo_experimental	comprimento_corpo	
2	com_predadores	35	
3	com_predadores	40	
4	com_predadores	32	
5	com_predadores	30	
6	com_predadores	38	
7	sem_predadores	15	
8	sem_predadores	18	
9	sem_predadores	20	
10	sem_predadores	12	
11	sem_predadores	15	
12			

Fig 2. Uma tabela corretamente formatada traz paz ao coração de qualquer cientista. Salve o formato “variável-coluna, medida-linha”!

	A	B	
1	grupo_experimental	comprimento_corpo	
2	com_predadores	35	
3	com_predadores	40	
4	com_predadores	32	
5	com_predadores	30	
6	com_predadores	38	
7	sem_predadores	15	
8	sem_predadores	18	
9	sem_predadores	20	
10	sem_predadores	12	
11	sem_predadores	15	
12			

- Uma observação por linha
- Uma variável por coluna
- Todas as colunas nomeadas/identificadas

Fig 2. Uma tabela corretamente formatada traz paz ao coração de qualquer cientista. Salve o formato “variável-coluna, medida-linha”!

	A	B
1	grupo_experimental	comprimento_corpo
2	com_predadores	35
3	com_predadores	40
4	com_predadores	32
5	com_predadores	30
6	com_predadores	38
7	sem_predadores	15
8	sem_predadores	18
9	sem_predadores	20
10	sem_predadores	12
11	sem_predadores	15
12		

**Variável qualitativa
(fator com 2 níveis)**

**Variável quantitativa
(contínua)**

Dica: tabela similar à usada para teste t

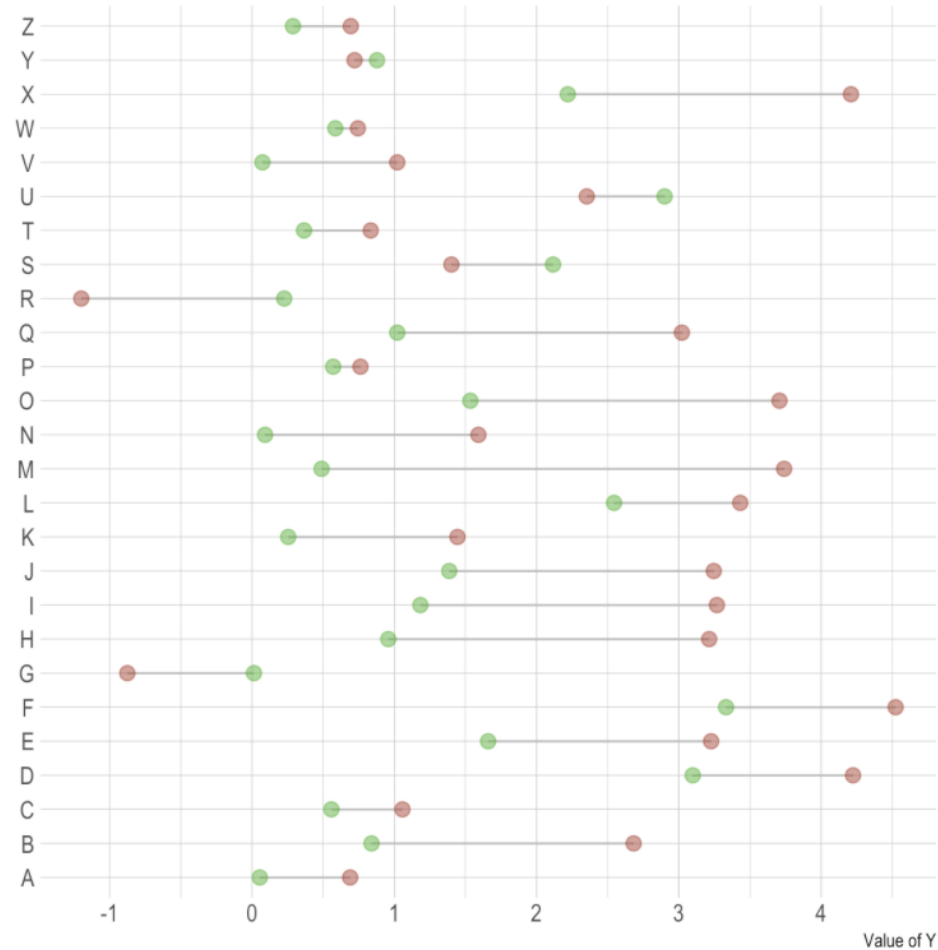
Fig 2. Uma tabela corretamente formatada traz paz ao coração de qualquer cientista. Salve o formato “variável-coluna, medida-linha”!

Tipos de gráficos

Tipos de gráficos

- Gráficos para variáveis qualitativas
 - Pizza (setores)
 - Pirulito
- Gráficos para variáveis quantitativas
 - Dispersão (pontos)
 - Histogramas
 - Barra
 - Boxplot
 - Linha

Pirulito (lollipop plots)



Frequência de 2 grupos

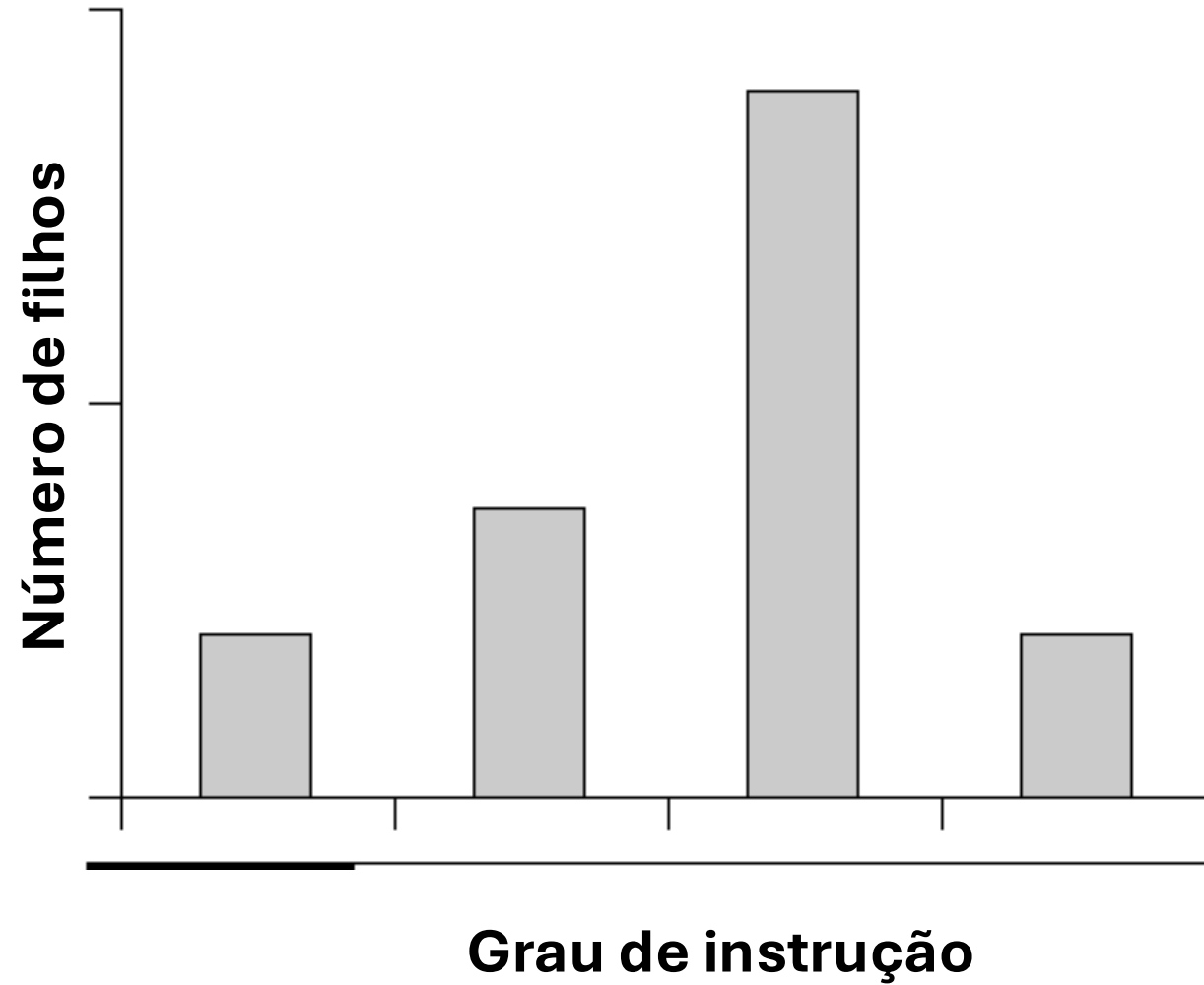
Tabela 2.1

Você acha que a cirurgia melhorou sua aparência?

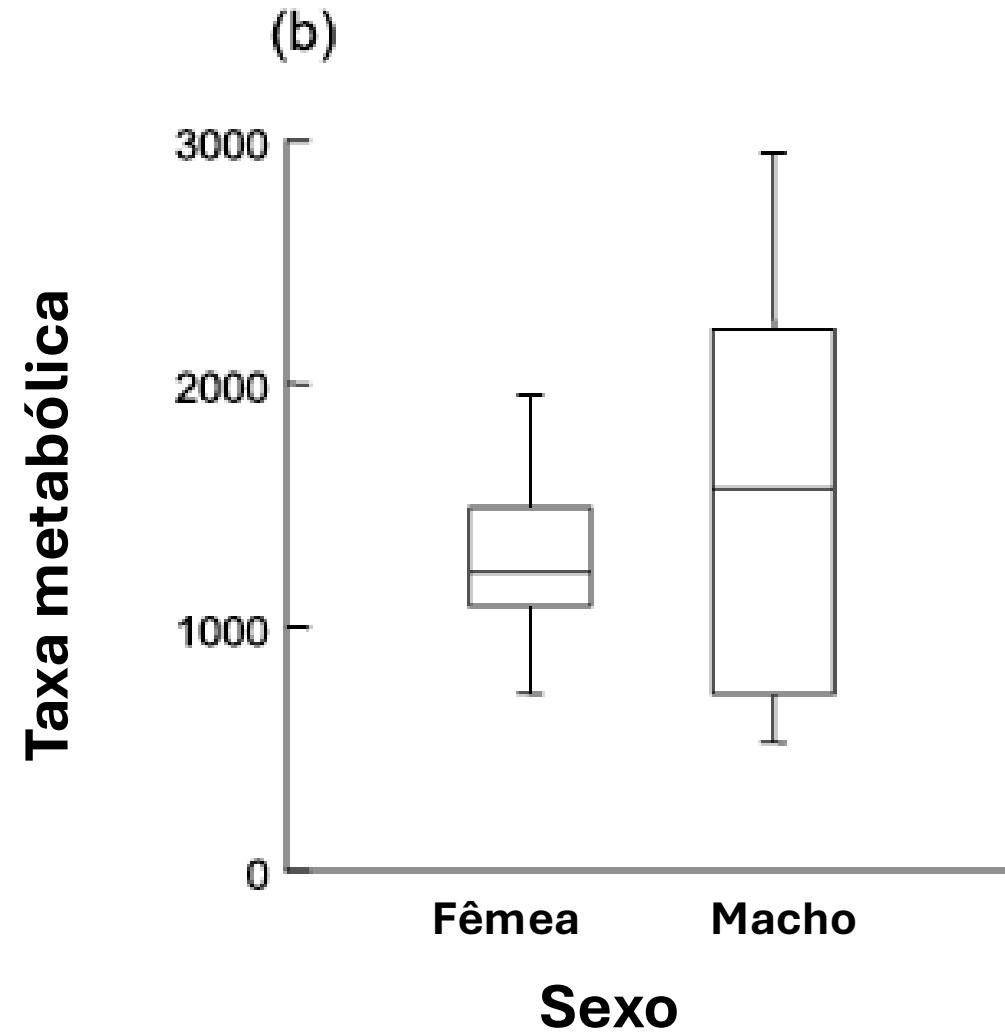
Resposta	Frequência	Porcentagem
Sim	66	66
Em parte	20	20
Não	8	8
Sem resposta	6	6
Total	100	100



FIGURA 2.1 Você acha que a cirurgia melhorou sua aparência?



Box-plot



Anatomia de um Box-plot

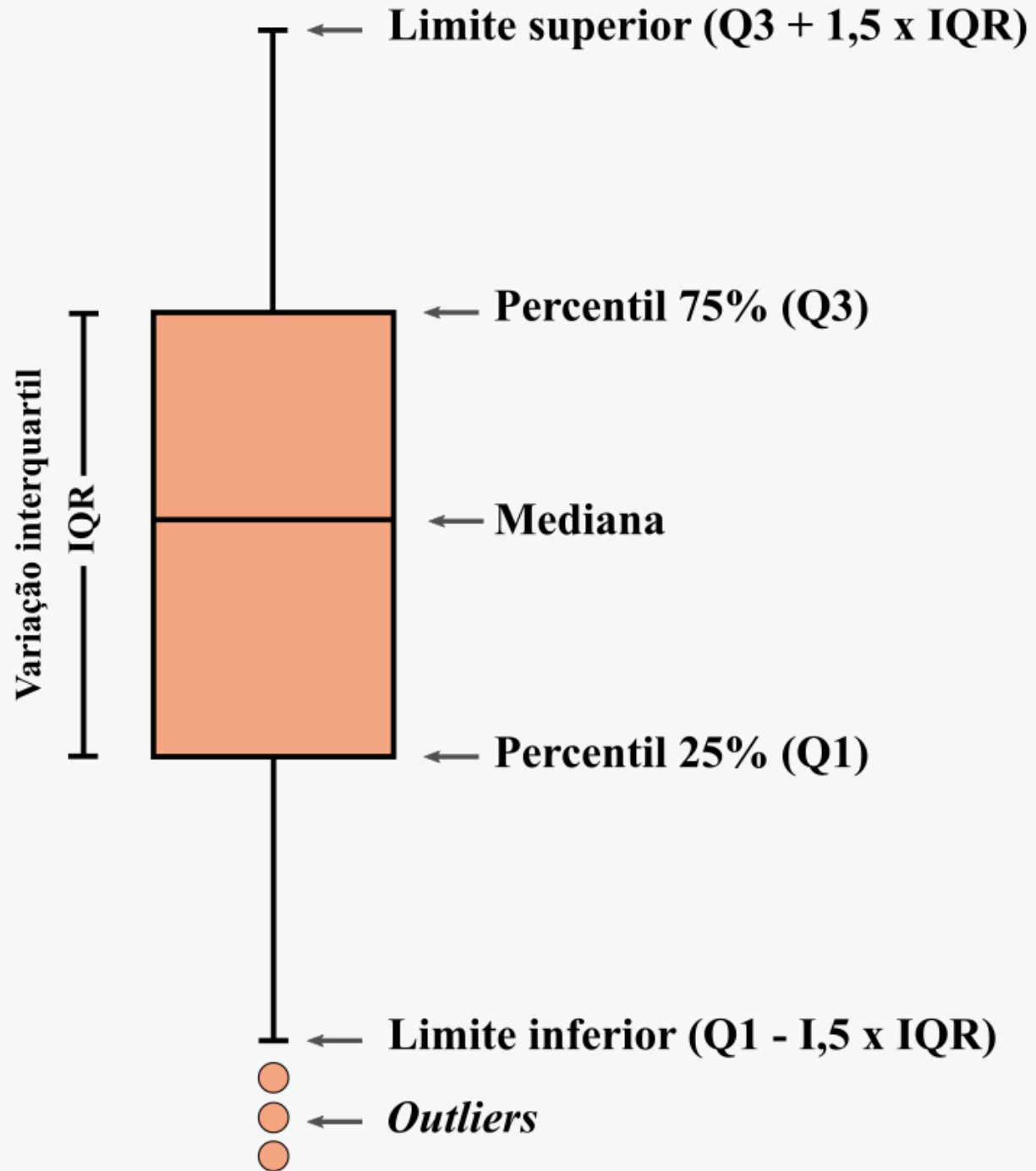
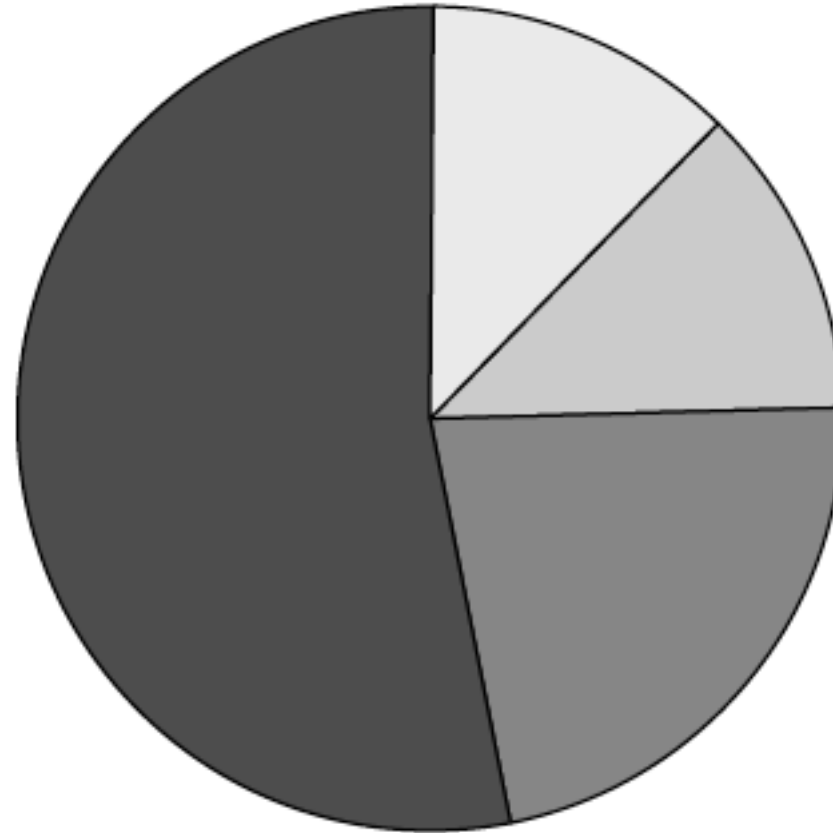


Gráfico de "pizza"



Região de procedência

Figure 19.12 A basic two-dimensional pie chart of the same data as in Figure 19.1.

Gráfico de dispersão (scatterplot)

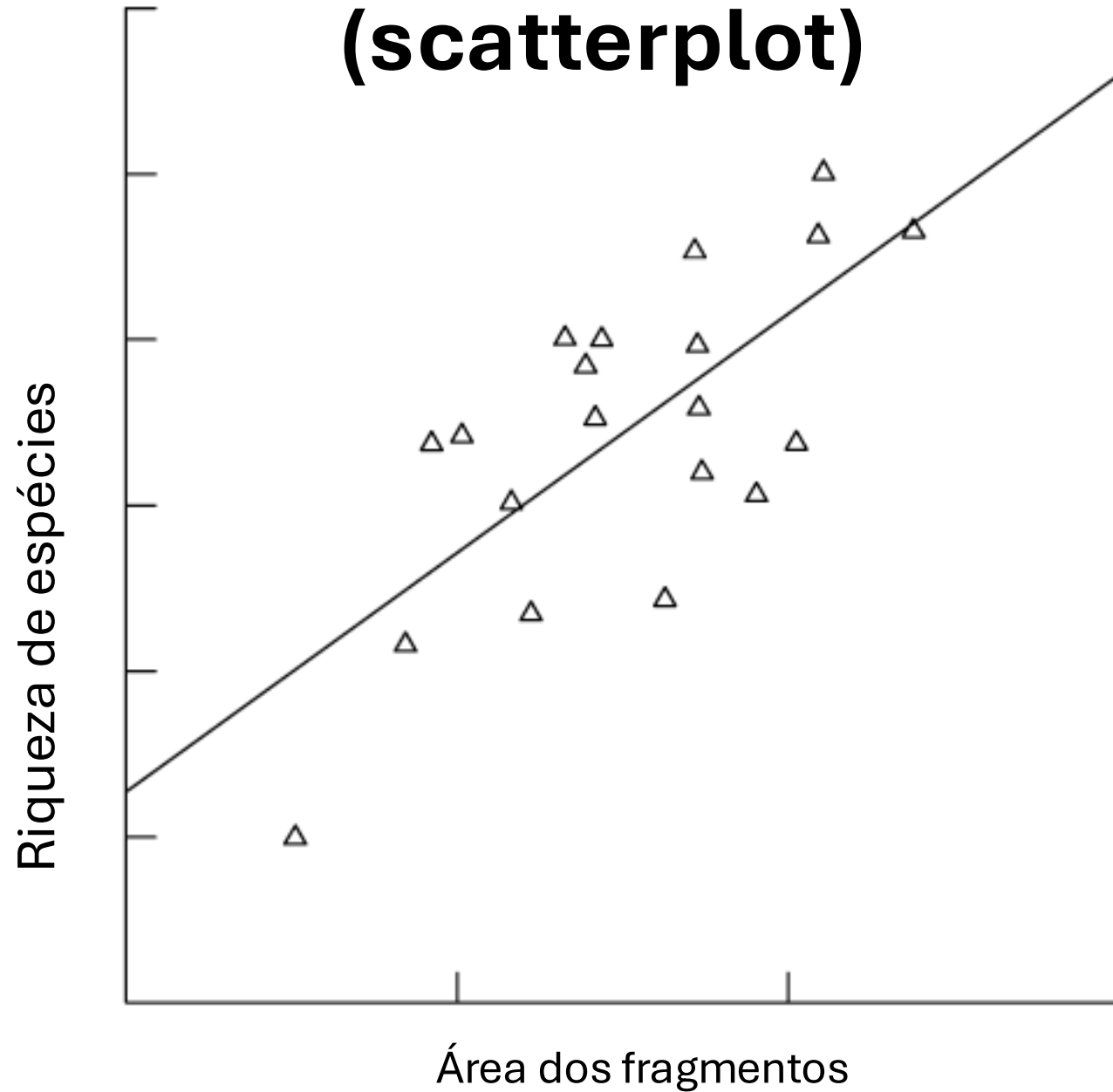
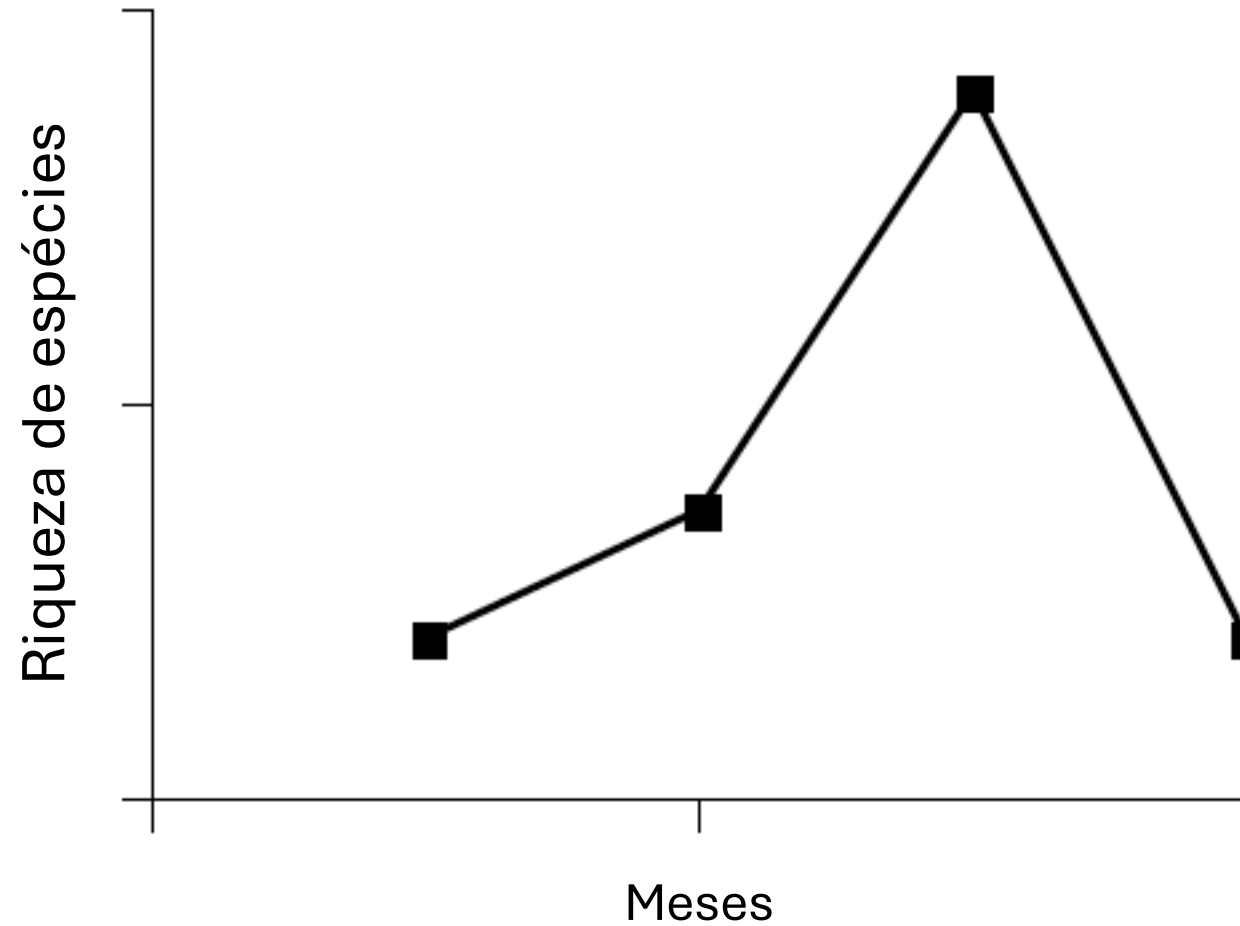
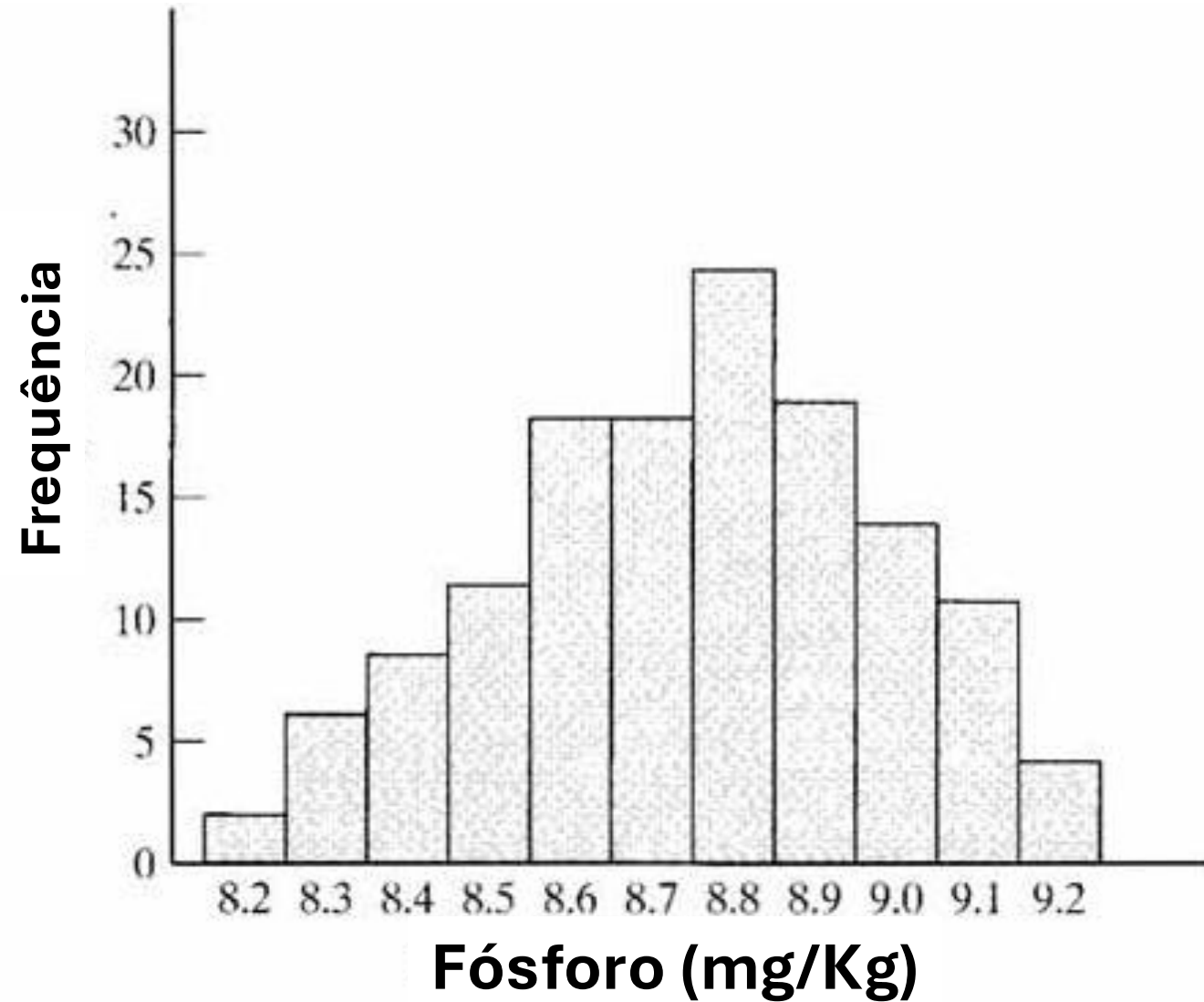


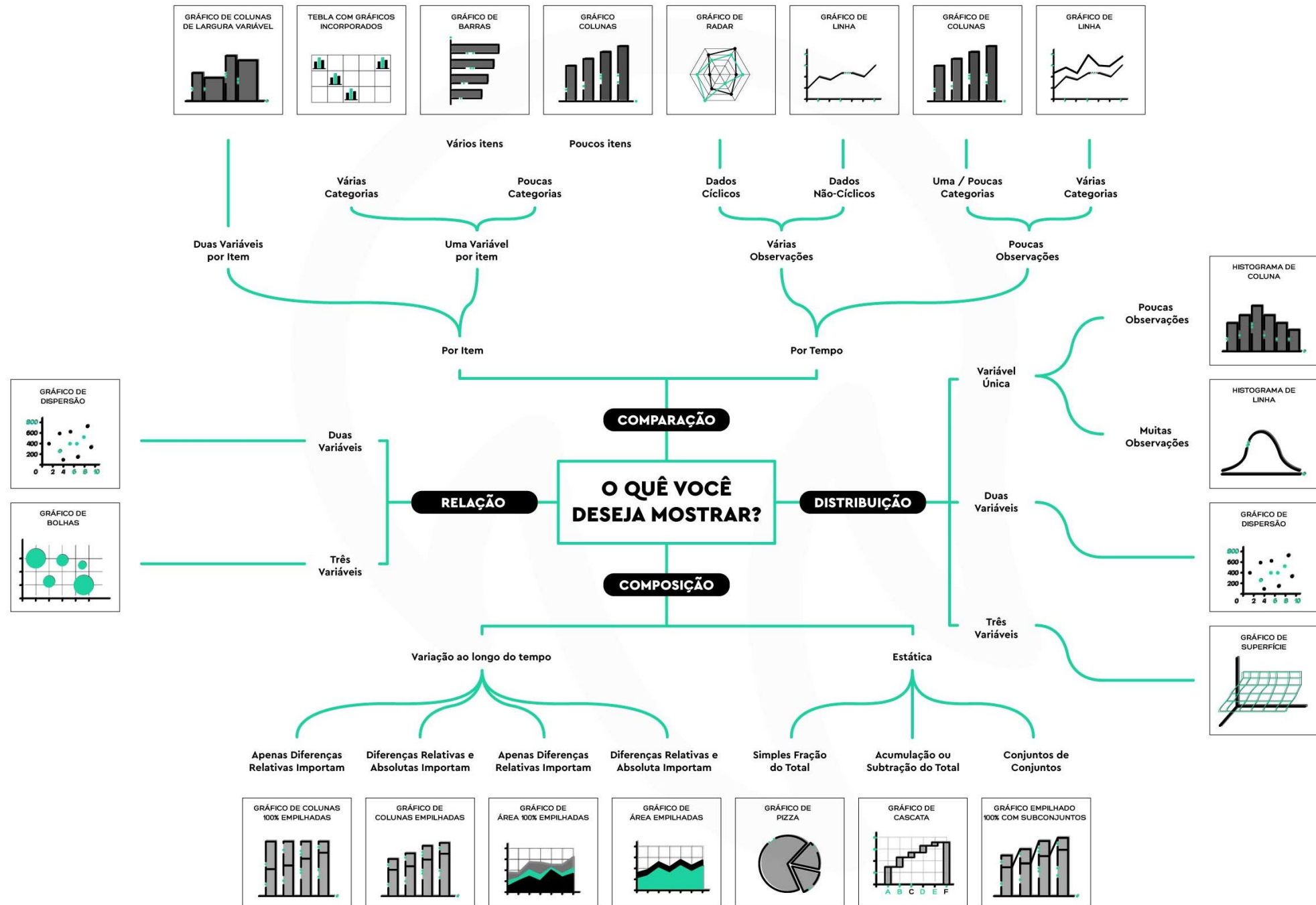
Gráfico de Linha



Histograma



Como escolher qual gráfico usar?



Quando usar gráficos vs. tabelas?

Vantagens

- **dos Gráficos:**

- **Sintetizam os dados e facilitam o entendimento de padrões**
- **Informam rápida e concisamente sobre a variabilidade** dos dados
- Permitem visualizar a distribuição das variáveis
- Podem ser **impactantes** e levar a insights que as palavras por si só não conseguem transmitir, comunicando padrões, redes e sistemas complexos

- **das Tabelas:**

- São ideais para quando a precisão de cada valor é importante para o leitor
- Apresentam **grandes volumes de dados precisos** de forma organizada

Limitações

- **dos Gráficos:**

- Apresentar muitos pontos de dados em gráficos como scatter plots, box plots ou histogramas pode **sacrificar a legibilidade**
- A interpretação baseada apenas em gráficos pode ser não confiável, pois é possível construir gráficos muito diferentes com base nos mesmos dados

- **das Tabelas:**

- Não são ideais para **exibir tendências ou relacionamentos** entre variáveis.
- Podem **sobrecarregar o leitor** se forem muito grandes ou contiverem excesso de informações.

Característica / Aspecto	Gráficos	Tabelas
Finalidade Principal	<p>Apresentar tendências e relacionamentos entre variáveis. Sintetizar dados e facilitar o entendimento de padrões. Visualizar a distribuição das variáveis.</p>	<p>Apresentar dados precisos para referência. Exibir grandes quantidades de dados que poderiam sobrecarregar um gráfico. Fornecer visões gerais sistemáticas e informações de fundo.</p>
Vantagens	<p>Síntese e clareza de padrões: Rápida compreensão de variabilidade e distribuição. Impacto visual: Enriquecem a argumentação e são cruciais para comunicar sistemas complexos. Revelam padrões: Permitem ver a forma da distribuição dos dados e a variabilidade dos pontos. Preservam dados individuais (e.g., ramo-e-folhas).</p>	<p>Precisão e detalhe: Exibem valores exatos e grandes volumes de dados. Visão sistemática: Úteis para dados que exigem atenção individual e para informações de fundo.</p>
Quando Preferir	<p>Ao focar em: tendências, comparações de distribuições (contínuas entre categorias), relações entre variáveis contínuas, e a apresentação de valores médios com medidas de variação. Para um número pequeno de categorias em gráficos de setores. Para contar uma história mais abrangente com múltiplos formatos. Para dados contínuos em grande número (histograma) ou dados discretos (diagrama de linhas). Para dados contínuos em pequeno número (gráfico de pontos).</p>	<p>Ao focar em: dados para referência, exibição de grandes volumes de dados precisos, e fornecimento de visões gerais sistemáticas ou informações de fundo. Quando todos os dados precisam de atenção igual e não há uma tendência clara a ser destacada.</p>
Limitações	<p>Podem ofuscar dados precisos se a finalidade for referência. Muitos tipos mostram apenas valores médios, ocultando dados individuais. Legibilidade comprometida com excesso de pontos. Gráficos 3D podem dificultar compreensão de proporções. A interpretação pode ser influenciada pela forma de construção, levando a não confiabilidade se analisado isoladamente.</p>	<p>Não são eficazes para exibir tendências ou relacionamentos entre variáveis. Podem sobrecarregar o leitor se muito grandes ou com excesso de informação. Geralmente não devem ser o elemento principal de uma "figura" por poderem ofuscar outros visuais. Não destacam características específicas dos dados visualmente tão bem quanto um gráfico.</p>

COMMENTARY

Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting

Katharina Gerstner^{*,1,2}, David Moreno-Mateos^{3,4}, Jessica Gurevitch⁵, Michael Beckmann², Stephan Kambach^{6,7}, Holly P. Jones⁸ and Ralf Seppelt^{2,9}

Boas práticas para reportar resultados na forma de texto, tabela e gráficos

Permite extrair e compilar informações tanto de tabelas quanto de gráficos

Examples of poor reporting

Methods

We tested the effects of intensified grazing on plant growth and plant diversity in several fields **1**

We used 9–12 replicates of each species and treatment combination **2**

The study was conducted in the Eifel mountains in Southern Germany at 50-210817, 6-692549 **3**

Results

[...] on average **4** regularly grazed fields yielded 12.66 g m⁻² (± 4.3) **5** plant biomass [...] Intensified grazing for species A did not show significant effects **6** while all other species showed a significant increase ($P < 0.05$) **7** response to the grazing treatment [...]

Tables

Treatment	Number of samples	Plant biomass (mean) 5
Control	42	12.66 \pm 4.3
Intensified	42	9.21 \pm 2.2

Figures

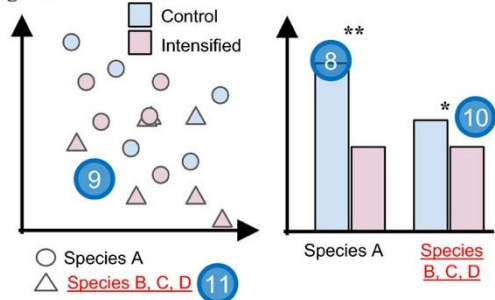


Figure 1 [...] Bars express mean plant biomass by grazing treatment **5**

Reporting requirements

- 1 Provide detailed information on sampling area of study (especially for observational or experimental work conducted in the field).
- 2 Report the exact number of replicates for each mean reported (e.g., for each species) for each treatment in a table or figure.
- 3 Always report exact location as coordinates (preferably decimal degrees) and give Northing/Easting and units (°, ', "). If exact locations are not revealed on purpose (i.e. to protect endangered species) provide the reason in the paper.
- 4 Explain what the means represent (e.g. means of individual plants, of plots, etc.), and report how replicates were assigned and number of independent replicates (i.e., sample size in which the means were calculated).
- 5 Identify units and the measure of variation when reporting it (SE, SD, 95% CI, etc.), especially in tables and figure captions.
- 6 Report all outcomes, significant as well as those that are not statistically significant; avoid "n.s.". Alternatively, some results can be shown in the supplementary material.
- 7 Report actual *P*-values, test statistics and sample sizes, rather than " > 0.05 " or asterisks. Report test-statistics alongside *P*-values.
- 8 Show means with variation (and identify nature of measure of variation) in figures.
- 9 Avoid overlapping points in scatterplots, use transparency and indicate full overlap.
- 10 If significance is labeled with asterisks in figures, provide the exact *P*-value in text or table as well
- 11 Report all data separately for all treatments and treatment combinations and observation units.

Examples of good reporting

Methods

[...] We tested the effects of increased grazing on plant growth in seven field plots sized 4 ha each. [...] We used nine replicate plots for species A and D and 12 replicates for species B and C in each treatment. Plots were seeded with one of the four species each in April 2017. Treatments were assigned to plots in a randomized block design, with.... [...]

[...] The study was conducted in the Eifel mountains in Southern Germany, 1km South of Gees at 50-210817° N, 6-692549° E [...]

Results

[...] on average plant biomass across all species in control plots was 12.66 g m⁻² (± 4.3 SD) [...]. Intensified grazing for species A did not show significant effects ($F_{1,8} = 1.07$, $P = 0.32$). Species B responded to grazing with increased growth ($F_{1,11} = 6.01$, $P = 0.03$) ... [text continues with reporting of results for all other species]

Tables

Treatment	Number of samples	Plant biomass (g m ⁻²) (mean \pm SD)
Control	42	12.66 \pm 4.3
Intensified	42	9.21 \pm 2.2

Figures

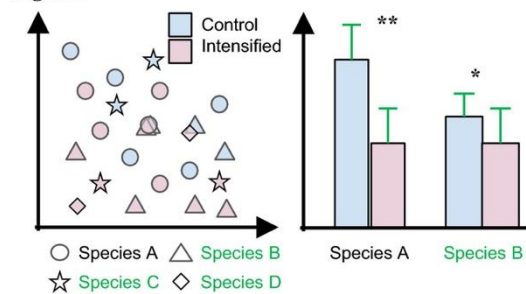


Figure 1 [...] Bars express mean and standard error for plant biomass [g m⁻²] by grazing treatment. Sample sizes were $N = 14$ and $N = 16$ for A in control and intensified treatments, respectively, and $N = \dots$ [...]