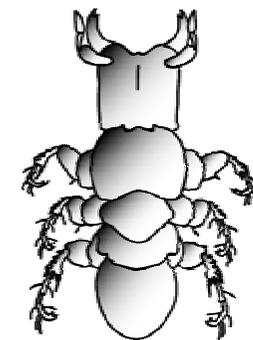
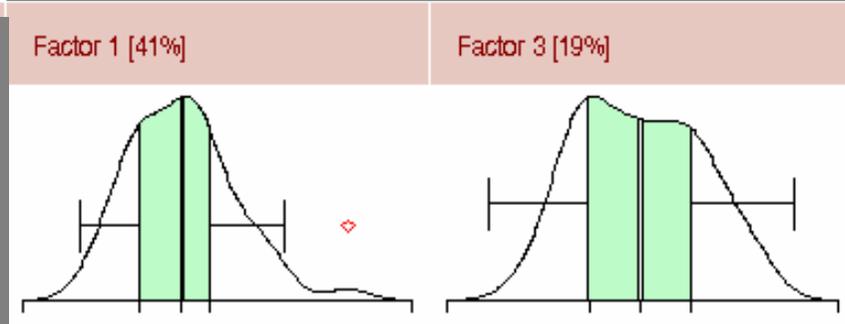
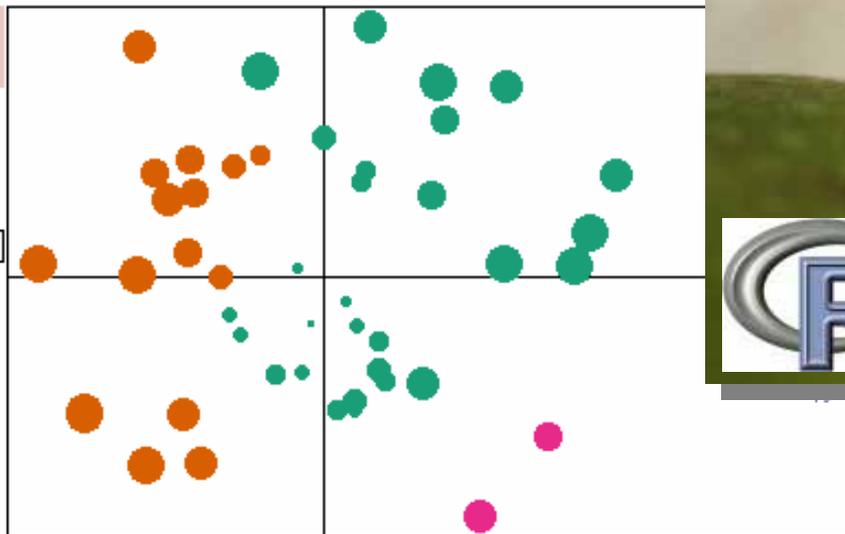
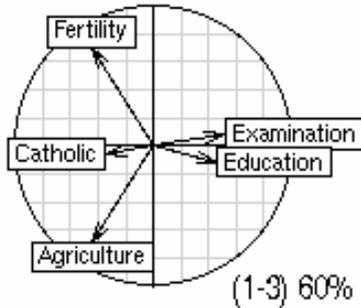


Uso da linguagem R para análise de dados em ecologia

PCA 5 vars
princomp(x = data, cor = cor)



Objetivo da aula

**Apresentar os princípios básicos de
Análise Exploratória de Dados.**

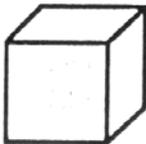
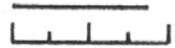
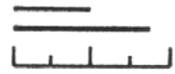
Análise Exploratória de Dados (AED)

- ✓ O que é AED?
- ✓ Quem utiliza AED?
- ✓ Objetivos da AED:
 - conhecer os dados
 - detectar padrões
 - orientar os testes de hipóteses
 - ajudar a interpretar os testes de hipóteses
- ✓ Apresentação gráfica é muito útil para AED

Análise Exploratória de Dados (AED)

Princípios básicos da apresentação gráfica:

- ✓ Ressaltar os padrões de interesse;
- ✓ Manter a estrutura dos dados, de forma que o leitor possa reconstruir os dados a partir da figura;
- ✓ A figura deve ter uma razão **dado:tinta** alta;
- ✓ As figuras **não** devem distorcer, exagerar ou apagar os dados.



BETTER

1. Position along a common scale

2. Position along identical scales

3. Length

4. Angle/Slope

5. Area

6. Volume

7. Shading: color, saturation, density

WORSE



William S. Cleveland

Análise Exploratória de Dados (AED)

Mais dicas sobre apresentação gráfica:

1. linhas não podem obscurecer os dados
2. dados não devem cair sobre os eixos
3. evitar linhas de grade
4. dados sobrepostos devem ser claramente distinguíveis (diminuir o tamanho dos pontos, separar os dados em gráficos diferentes, agitar os pontos etc)
5. o gráfico deve ser legível após redução na publicação ou em projeções em seminários

Análise Exploratória de Dados (AED)

Tabela vs. gráfico

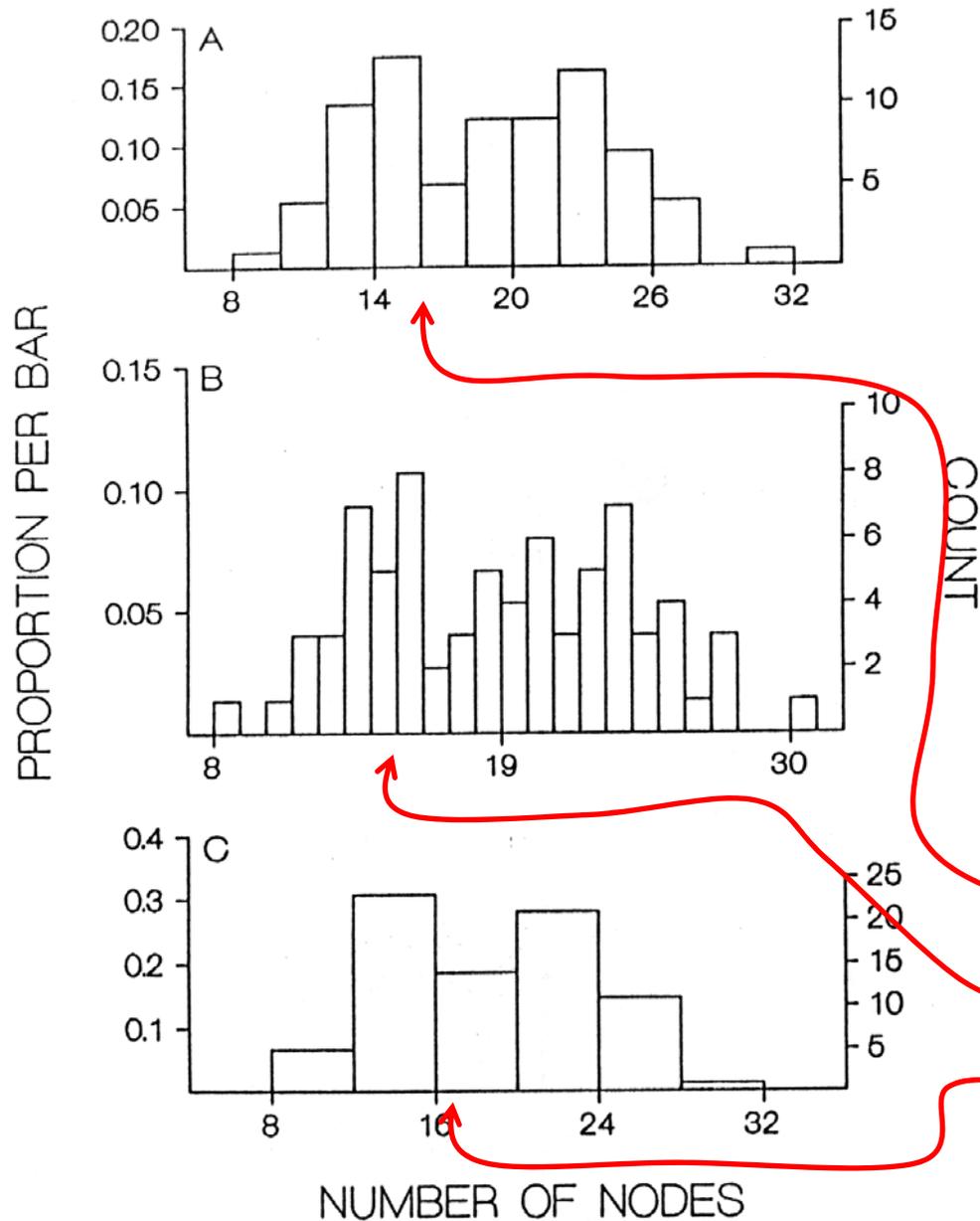
- ✓ Gráfico é melhor para **mostrar rapidamente** relações entre variáveis e dados multidimensionais de forma compreensível.
- ✓ Tabela mostra melhor valores exatos

Tipos de gráficos para AED

Dados univariados

- ✓ Histograma: gráfico de densidade. Não confundir com gráfico de barras
 - Desvantagens: **esconde** os dados (**mostra apenas intervalos**), as classes são **arbitrárias** e ocupa posição baixa na escala de Cleveland.

Histograma



Número de nós por planta em 75 indivíduos de *Ailanthus altissima*.

A) Intervalo = 2

B) Intervalo = 1

C) Intervalo = 4

Tipos de gráficos para AED

Haste e folha

- ✓ Dados brutos são apresentados
- ✓ Mesmo problema visual que o histograma: classes são arbitrárias

0 8
1 0111
1 2223333333
1H 4444455555555
1 66777
1M 888889999
2 000000111
2H 2222233333333
2 4445555
2 6777
2
3 0

Cada linha é uma haste

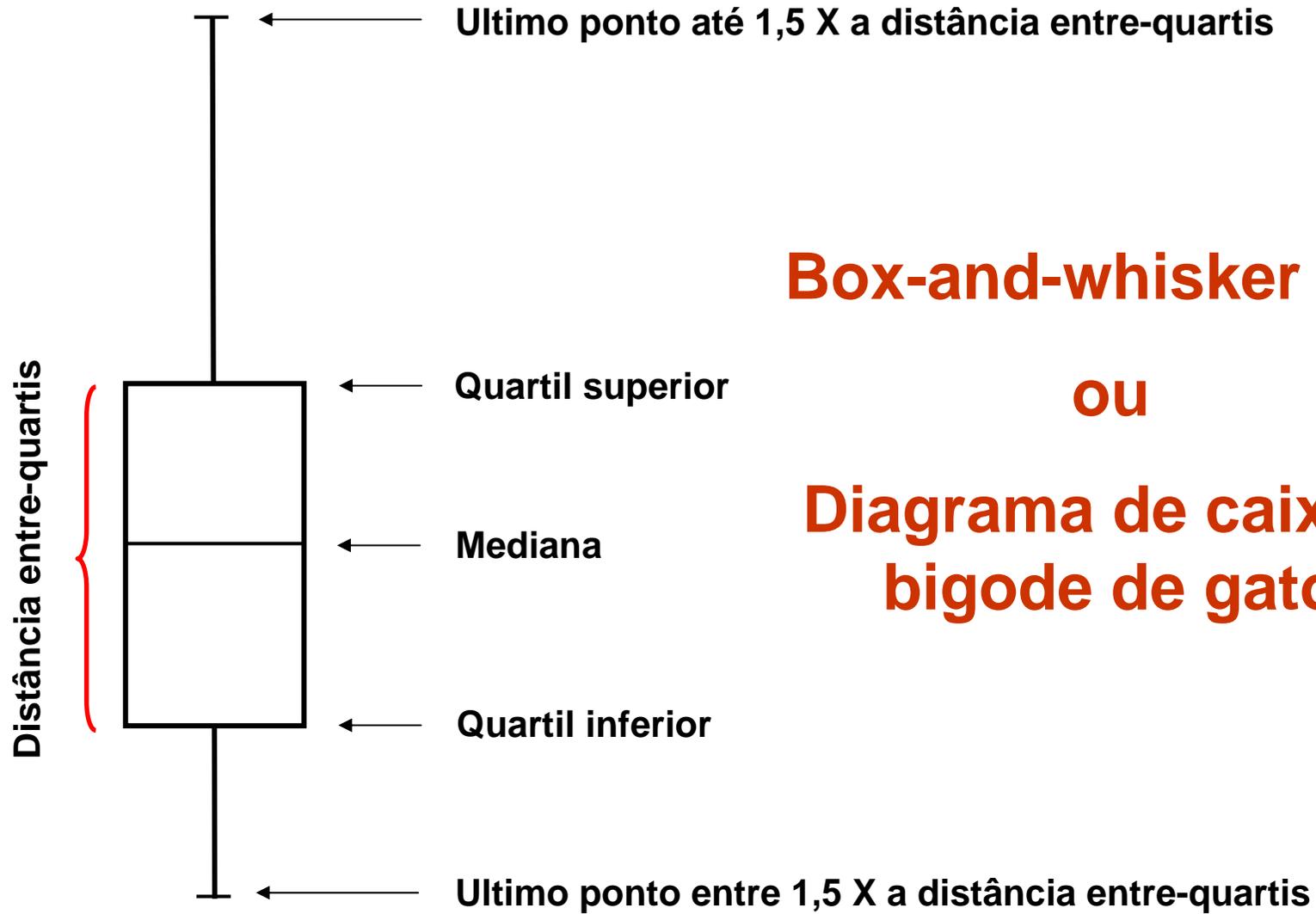
Cada dado é uma folha

Exemplo:

08
10, 11, 11, 11

M = mediana
H = quartil

* ← Valor extremo: + que 1,5 X a distância entre-quartis



Box-and-whisker plot

ou

Diagrama de caixa e bigode de gato

Tipos de gráficos para AED

Histogramas e **haste e folha** podem representar possível bimodalidade nos dados

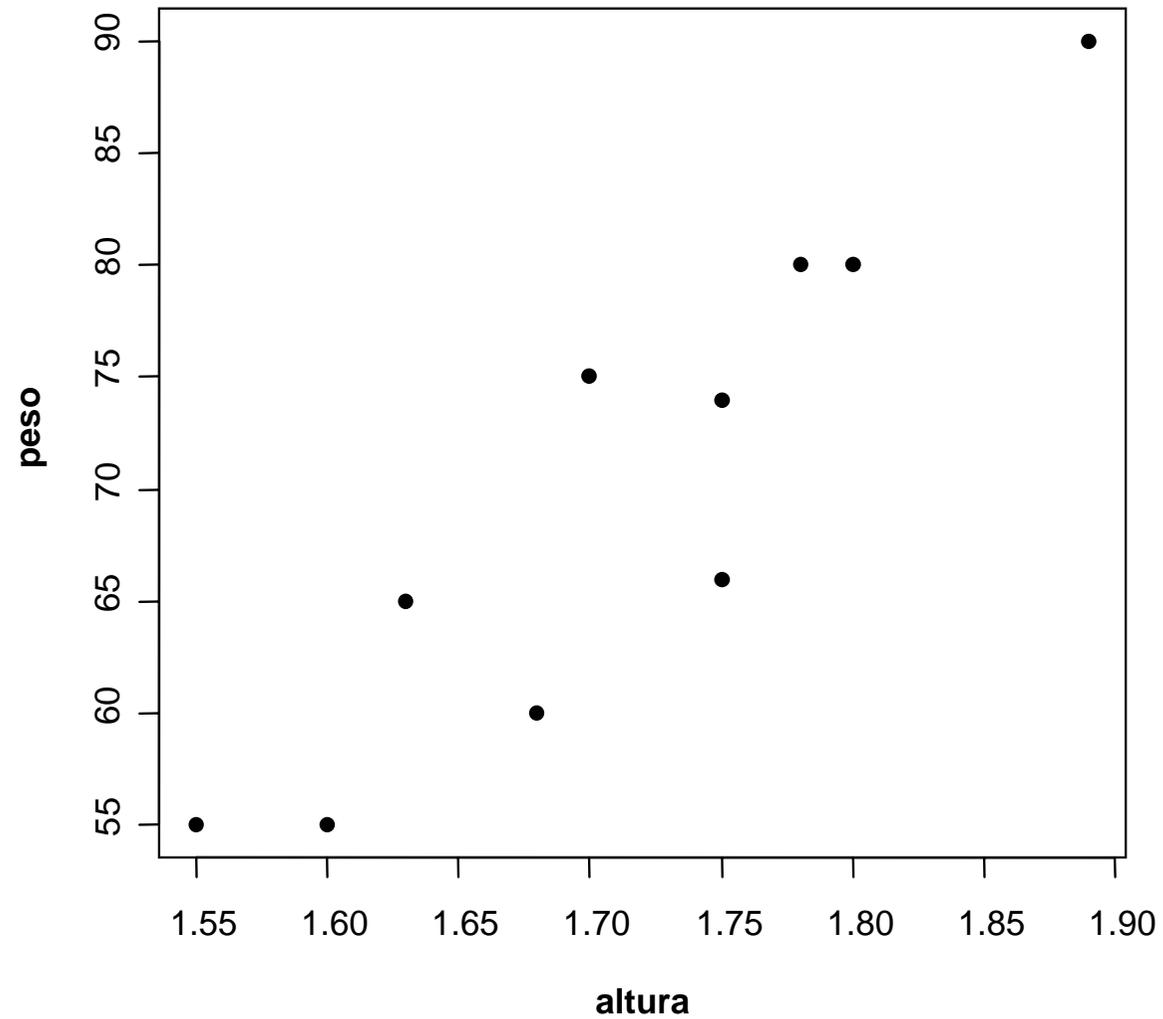
Box-plot escondem esse padrão

Exemplos no R

Dados bivariados

Dispersão

- ✓ Permite analisar a relação entre duas variáveis



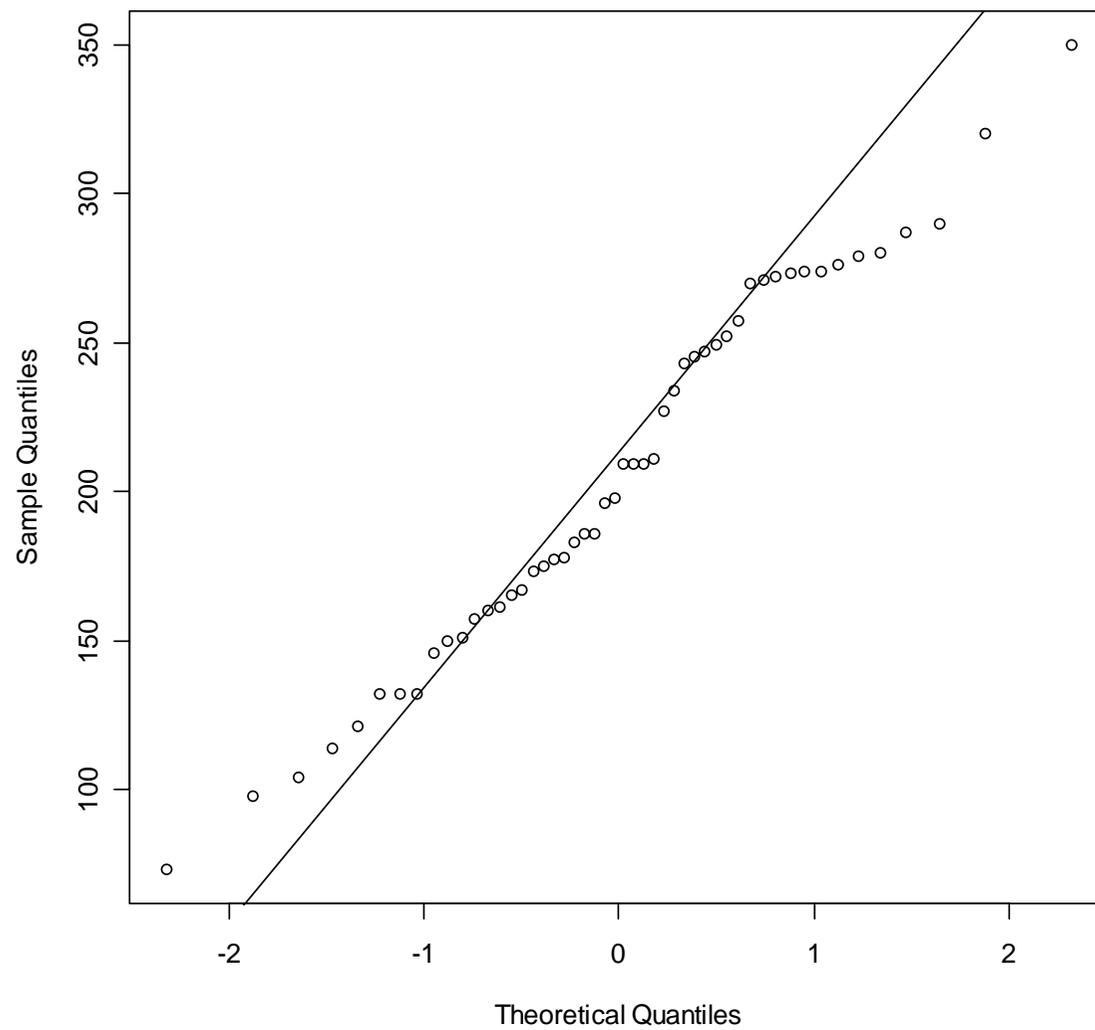
Tipos de gráficos para AED

Quantil-Quantil

- ✓ Permite comparar a forma de duas distribuições
- ✓ Muito usado para **avaliar** se os dados têm **distribuição parecida** com alguma **distribuição teórica** (Normal, Poisson etc).

Quantil-quantil

Normal Q-Q Plot

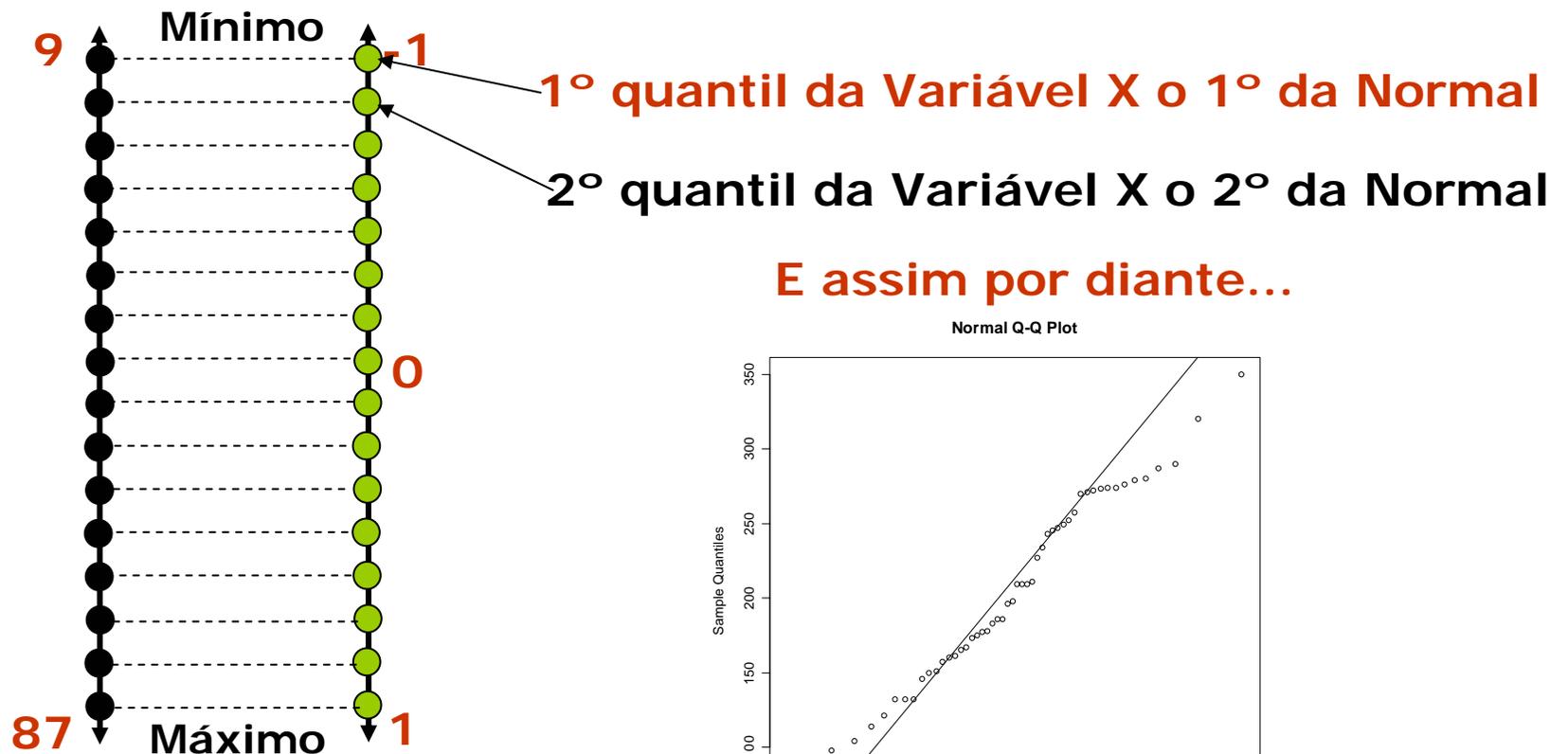


Quantil-quantil

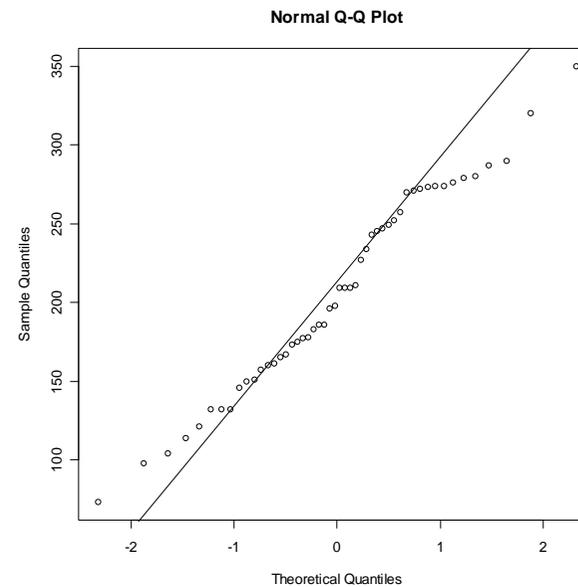
1º) Ordenar as variáveis

2º) Plotar os quantis uns contra os outros

Variável Normal



E assim por diante...



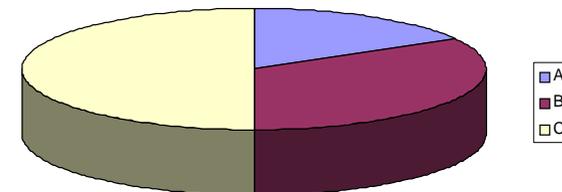
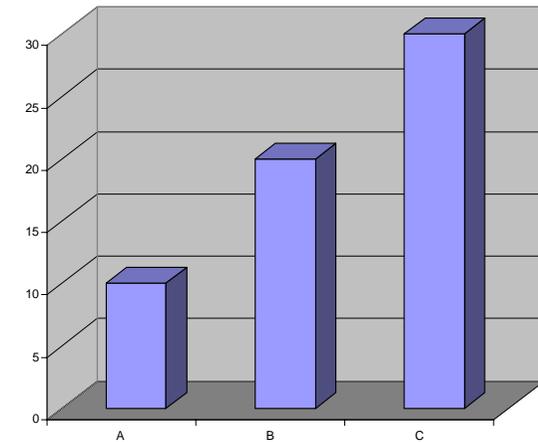
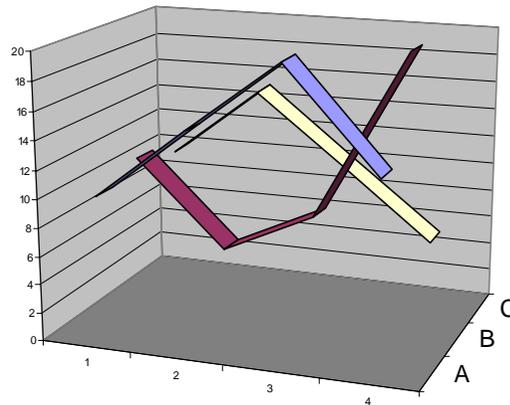
Exemplos no R

Diagramas 3D

✓ Usar gráficos tridimensionais **APENAS** se tiver 3 variáveis

✓ Nunca usar em trabalhos científicos:

1. Barras com volume
2. Pizza com volume
3. Tiras para série temporais
4. Etc



Diagramas 3D

Dispesão 3D

Pacote lattice

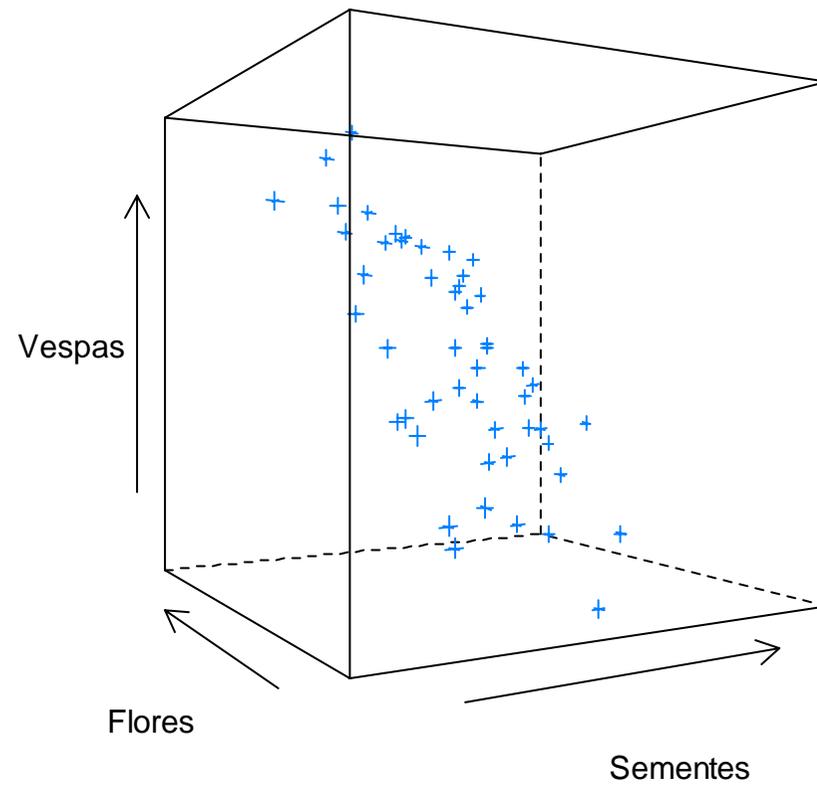
cloud()

Fórmula: $y \sim x * z$

EX:

```
cloud(y ~ x * z, data = nome.df,  
screen=list(x = -90, y = 30), distance = .4, zoom = .6)
```

Dispersão 3D



Sugestões de leitura

Ellison, A. M. 1993. Exploratory data analysis and graphic display. In: Scheiner, S. M. (ed.), *Design and analysis of ecological experiments*. Chapman & Hall, pp. 14-45.

McGill, R., Tukey, J. W. and Larsen, W. A. 1978. Variations of Box Plots. *Am. Statist.* 32: 12-16.