

ANÁLISE EXPLORATÓRIA DE DADOS COM R

PARTE II VISUALIZAÇÃO

Professor Walter Subiza

WALTER HUMBERTO SUBIZA PINA

Lotação

Gerência de Imageamento
Coordenação de Cartografia
Diretoria de Geociências

Contato

2142 4908
walter.pina@ibge.gov.br

REPOSITÓRIO

O material está disponível no link:

<https://github.com/HumbertoSubiza/AED-II>

Os dados e o software usados foram criados especificamente para a oficina pelo instrutor ou são de fonte aberta (*open source**).

- * “Dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura.”, *Ministério do Planejamento, Desenvolvimento e Gestão*

O PORQUÊ DO CURSO

“The simple graph has brought more information to the data analyst’s mind than any other device.”

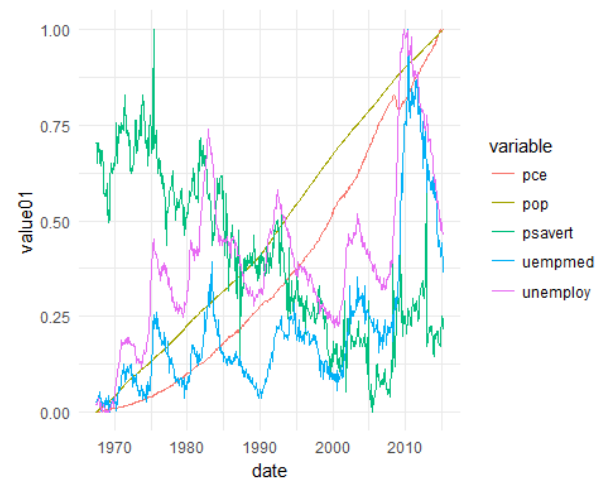
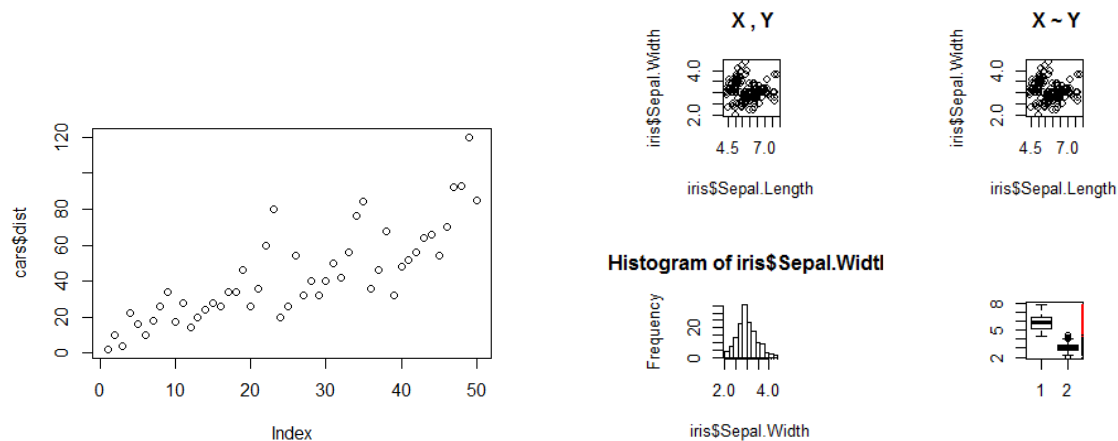
“The greatest value of a picture is when it forces us to notice what we never expected to see.”

—

John Tukey

RESUMO VISUAL

Curso Elaboração de Gráficos | Pedro Silva



O objetivo de um gráfico ou figura é apresentar um **resumo visual** que seja mais informativo que números ou dados.

CONTANDO UMA HISTÓRIA

Curso Elaboração de Gráficos | Pedro Silva

Gráfico de barras - cores acorde tipo de transmissão

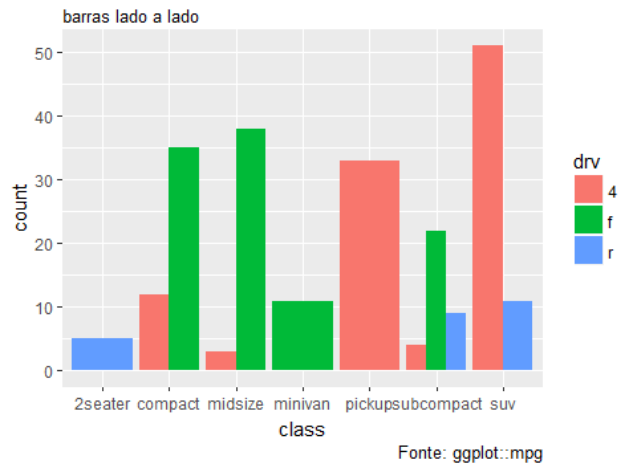
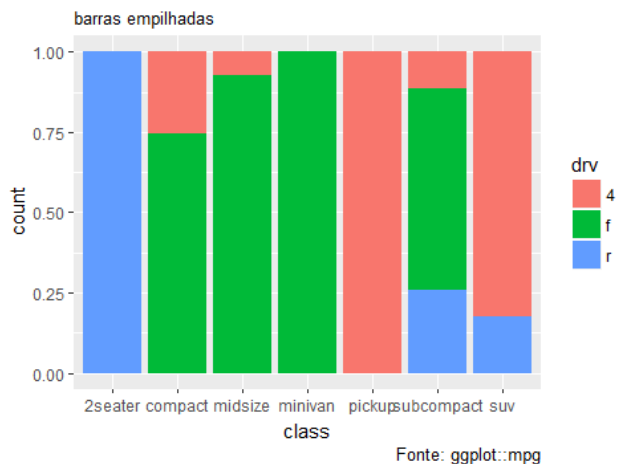


Gráfico de barras - cores acorde tipo de transmissão



O gráfico deve contar uma **história** ou **transmitir uma mensagem**.

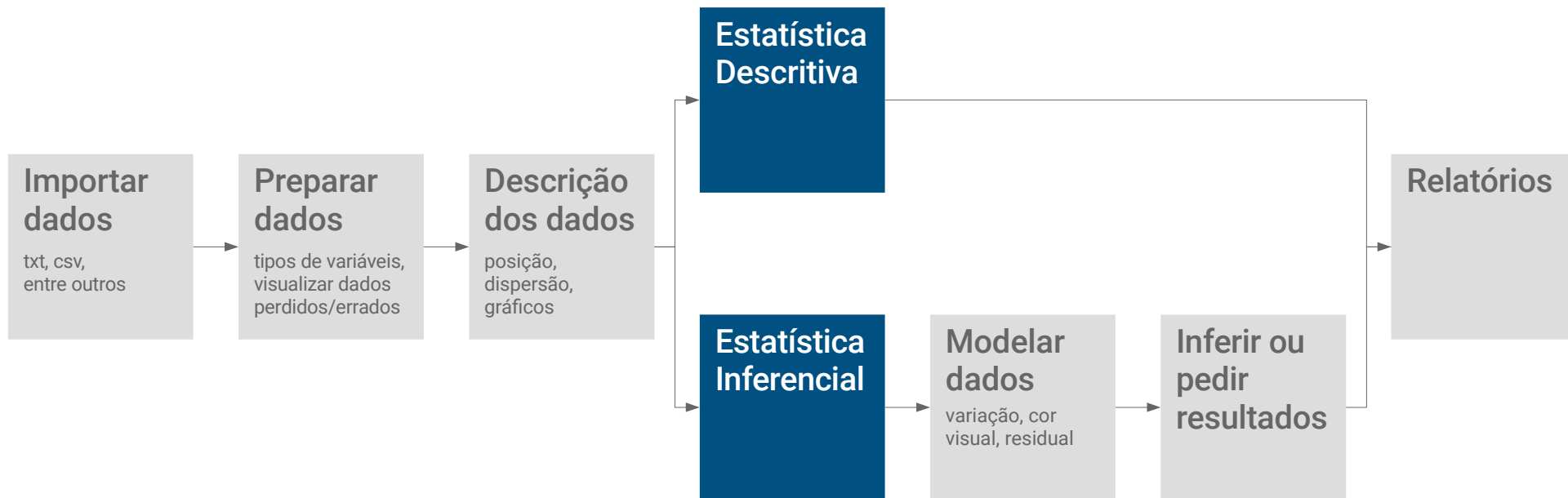
Um gráfico é bom quando há o correto entendimento da história pelo leitor.

PRINCÍPIOS GERAIS

Curso Elaboração de Gráficos | Pedro Silva

- Mostre os dados;
- Diga a verdade;
- Siga um propósito definido:
exploração, interpretação e comunicação;
- Use o tipo de gráfico adequado para cada conjunto de dados;
- Destaque os dados com o mínimo de lixo gráfico;
- Inclua rótulos e descrições indicando: título, conteúdo ou objetivo, o que representa cada eixo e a fonte dos dados.

AED E MODELADO DE DADOS



VISUALIZAÇÃO DE DADOS

Finalidade Exploratória

análise gráfica de dados,
para conhecer, preparar
e interpretar os dados (AED).

Finalidade Explanatória

informar, explicar, demonstrar para um
público mais amplo ou especializado
(apresentações e relatórios finais).

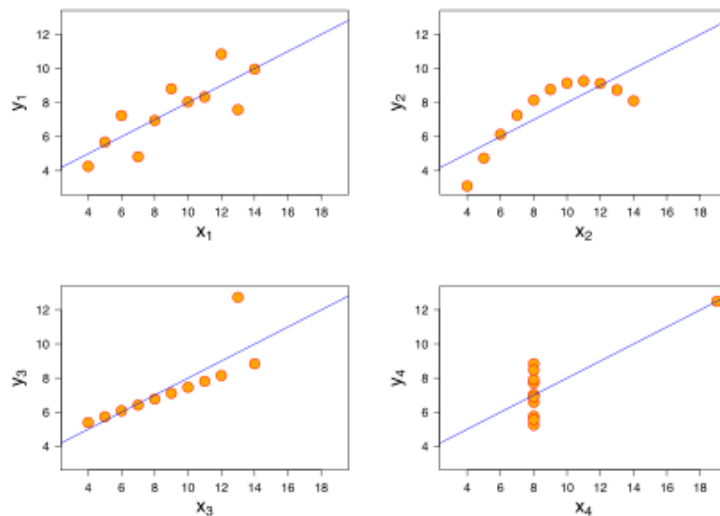
QUARTETO DE ANSCOMBE

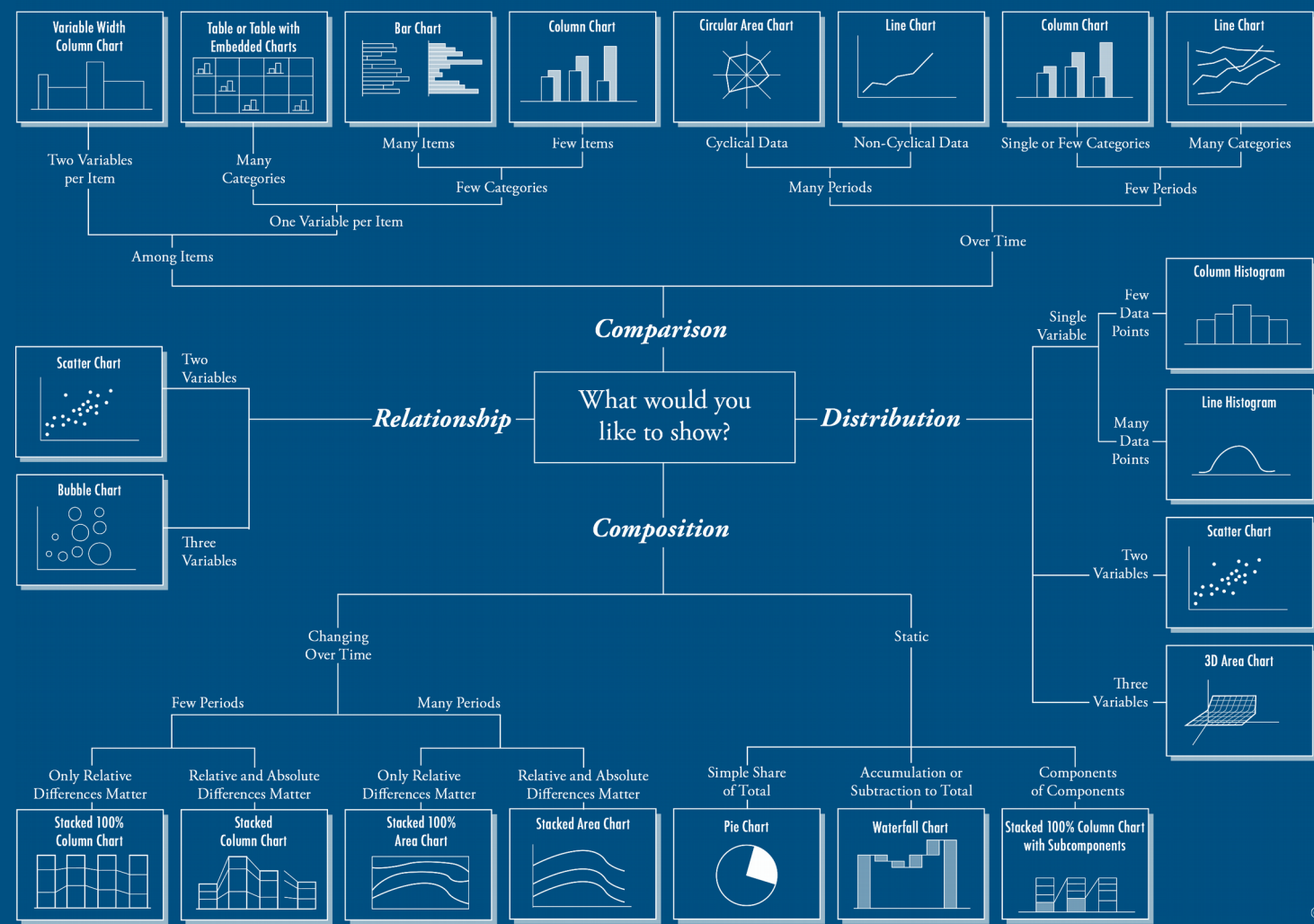
“Quatro conjuntos de dados que aparentam ser idênticos quando descritos por certas técnicas de estatística descritiva, mas que são distintos quando exibidos graficamente.”

Demonstra-se a importância de visualizar os dados antes de analisá-los e o efeito dos outliers nas propriedades estatísticas.”

—

Wikipedia





FUNÇÕES

Duas funções serão apresentadas para a visualização:

- **plot()**: base para AED e para gráficos explanatórios;
- **ggplot2()**: tidyverse para gráficos explanatórios.

FUNÇÕES | PLOT()

Função genérica que depende dos dados de entrada. Apresenta dispersão, histogramas, boxplot, barras (barplot), girassol (sunflowerplot), pairs, contour etc.

```
# data frame  
plot(iris)  
# simple plot de uma variável fator  
plot(iris$Species)
```

Consulte a biblioteca *graphics*.

FUNÇÕES | GGLOT2()

“Gráfico é um mapeamento de dados a partir de atributos estéticos de objetos geométricos.”

—

Leland Wilkinson

FUNÇÕES | GGLOT2()

Baseado em Leland Wilkinson, *Grammar of Graphics*, 1999, há dois princípios:

- o gráfico é composto por diferentes camadas de elementos;
- os gráficos com significado se obtém através de elementos estéticos adicionados.

Baseado nesses conceitos, Hadley Wickman estabelece que um gráfico pode ser construído a partir de camadas que são adicionadas sucessivamente. Assim, surgiu o pacote ggplot2.

ELEMENTOS DOS GRÁFICOS

Elementos básicos	Descrições
Dados (data)	O conjunto que contém os dados a apresentar, por exemplo, um dataframe.
Estética (aesthetics)	Atributos estéticos como: mapeamento da posição das variáveis, forma, cor, tamanho etc.
Geometrias (geometries)	Os elementos visuais com os quais representamos os dados (pontos, linhas, barras etc.)

ELEMENTOS DOS GRÁFICOS

Elementos complementares	Descrições
Fatores (facets)	Gráficos múltiplos por categoria.
Estatísticas (statistics)	Representação dos dados para ajudar na compreensão.
Coordenadas (coordinates)	O espaço no qual os dados serão representados.
Temas (themes)	Aparência geral.

RESUMO DAS VARIÁVEIS

Dados	Variáveis de interesse
Estética	x,y col, fill size, labels alpha, shape lwd, ltype
Geometrias	point line histogram bar boxplot etc.
Fatores	columns rows
Estatísticas	binning smoothing descriptive inferential
Coordenadas	cartesian fixed polar limits
Temas	classic dark light bw ?themes

ANÁLISE EXPLORATÓRIA DE DADOS COM R

PARTE II VISUALIZAÇÃO

Professor Walter Subiza