

Visualização de dados com ‘ggplot2’ e ‘dplyr’

Ministrante: Jeanne Franco

Meetup R-Ladies Ribeirão Preto

Um pouco sobre mim

Formação e experiência com o R



- Doutorado em Ecologia (UFRN);
- Colaboradora da empresa Beta Analítica;
- Coorganizadora da comunidade R-Ladies Natal;
- Pacotes: rmarkdown e tidyverse (ggplot2, dplyr, tidyr, stringr, readr, purrr, tibble eforcats);
- Contato: [LinkedIn](#), [Github](#) e [Instagram](#).

Conteúdo

- Introdução aos pacotes dplyr e ggplot2;
- Teoria da Gramática de Gráficos;
- Principais tipos de gráficos;
- Etapas iniciais para criação de gráficos;
- Prática no RStudio.



Introdução aos pacotes dplyr e ggplot2

Introdução: o que é o pacote dplyr?

- Dplyr é um pacote que fornece um conjunto de funções que ajudam a resolver os desafios mais comuns da manipulação de dados.



| Arte: Allison Horst

Conjunto de funções do pacote dplyr

- `mutate()`: adiciona novas variáveis em função das variáveis existentes;
- `select()`: seleção de variáveis;
- `filter()`: filtra valores de cada variável;
- `summarise()`: reduz múltiplos valores a um único resumo;
- `arrange()`: alterna a ordem dos valores de cada variável.



Introdução: o que é o pacote ggplot2?

- O ggplot2 é um pacote dedicado à visualização de dados que produz gráficos elegantes.



| Arte: Allison Horst

Introdução: o que é o pacote ggplot2?

- Criado por Hadley Wickham enquanto ele era estudante em Iowa State University.



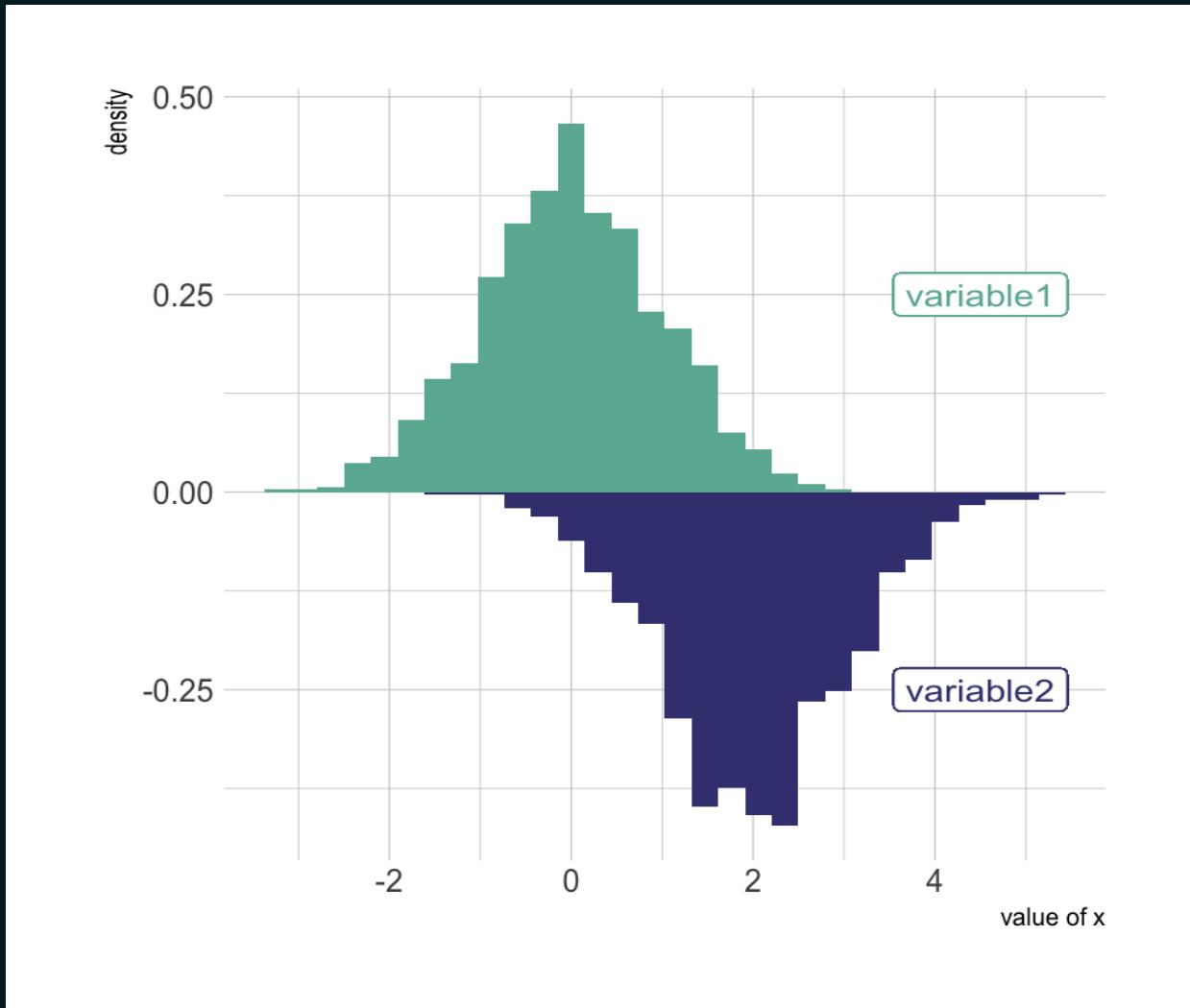
Introdução: o que é o pacote ggplot2?

- Gráficos bonitos e detalhados;
- Versatilidade que permite transmitir as principais mensagens dos gráficos.

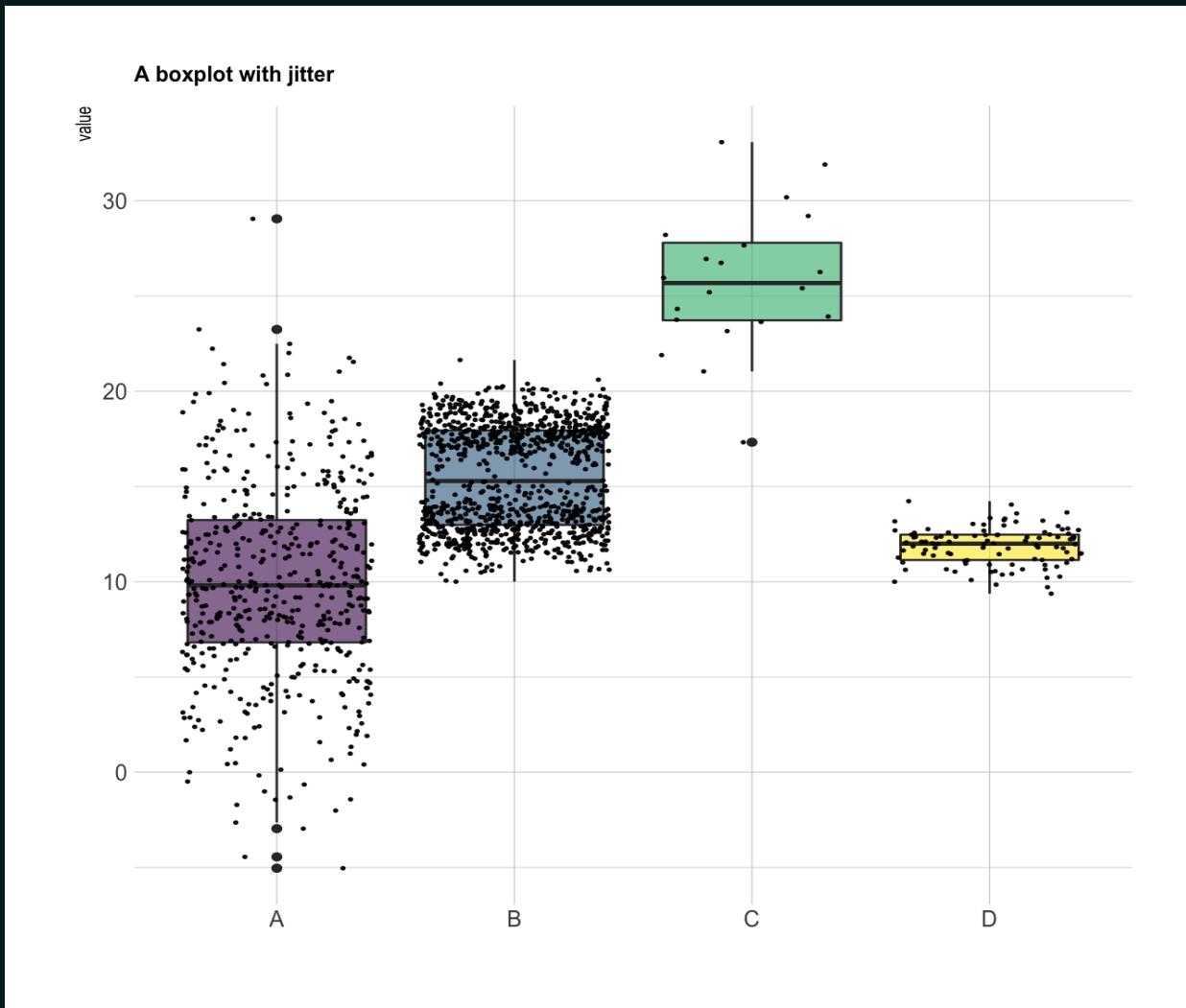


| Arte: Allison Horst

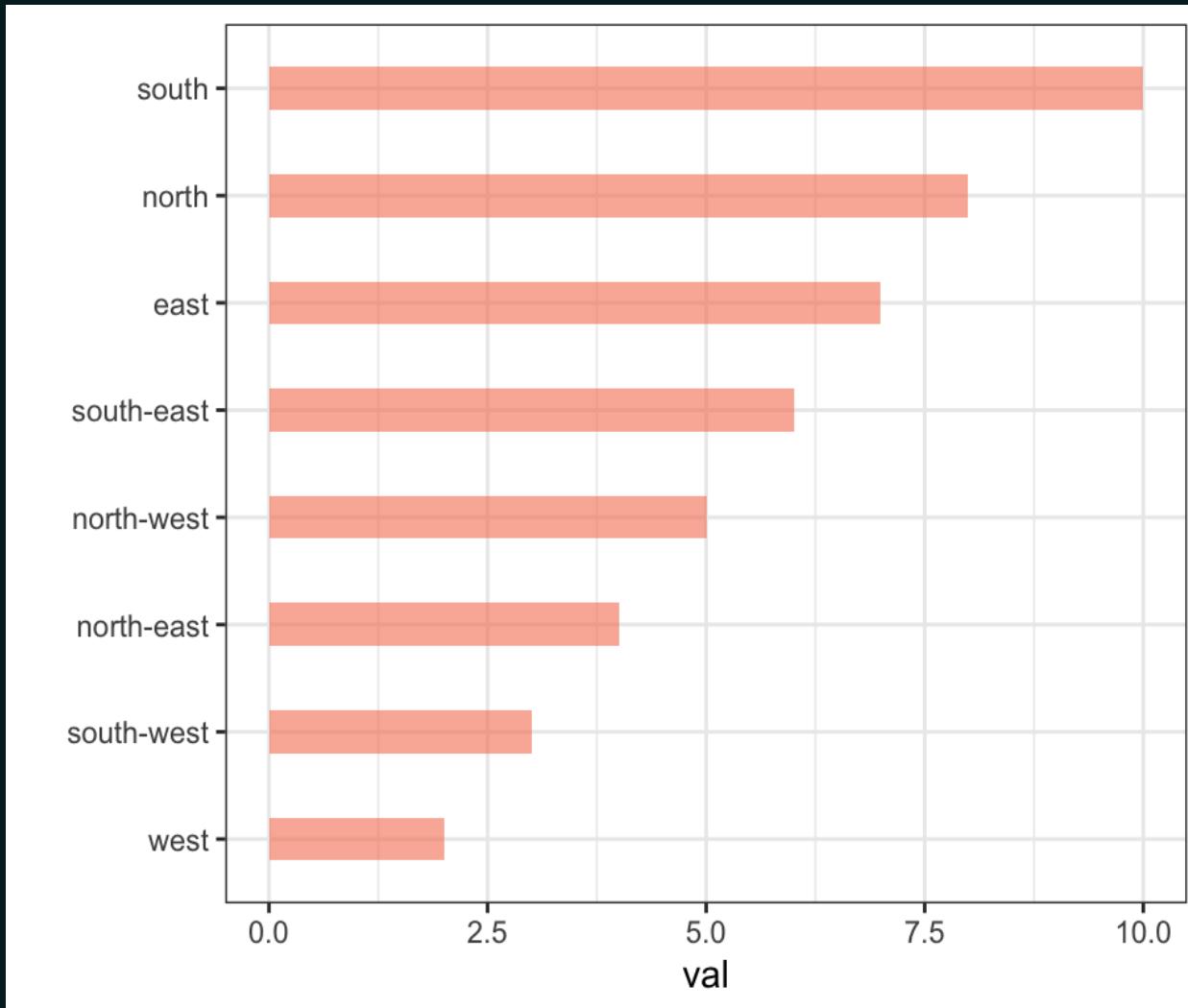
Exemplo de gráfico - ggplot2



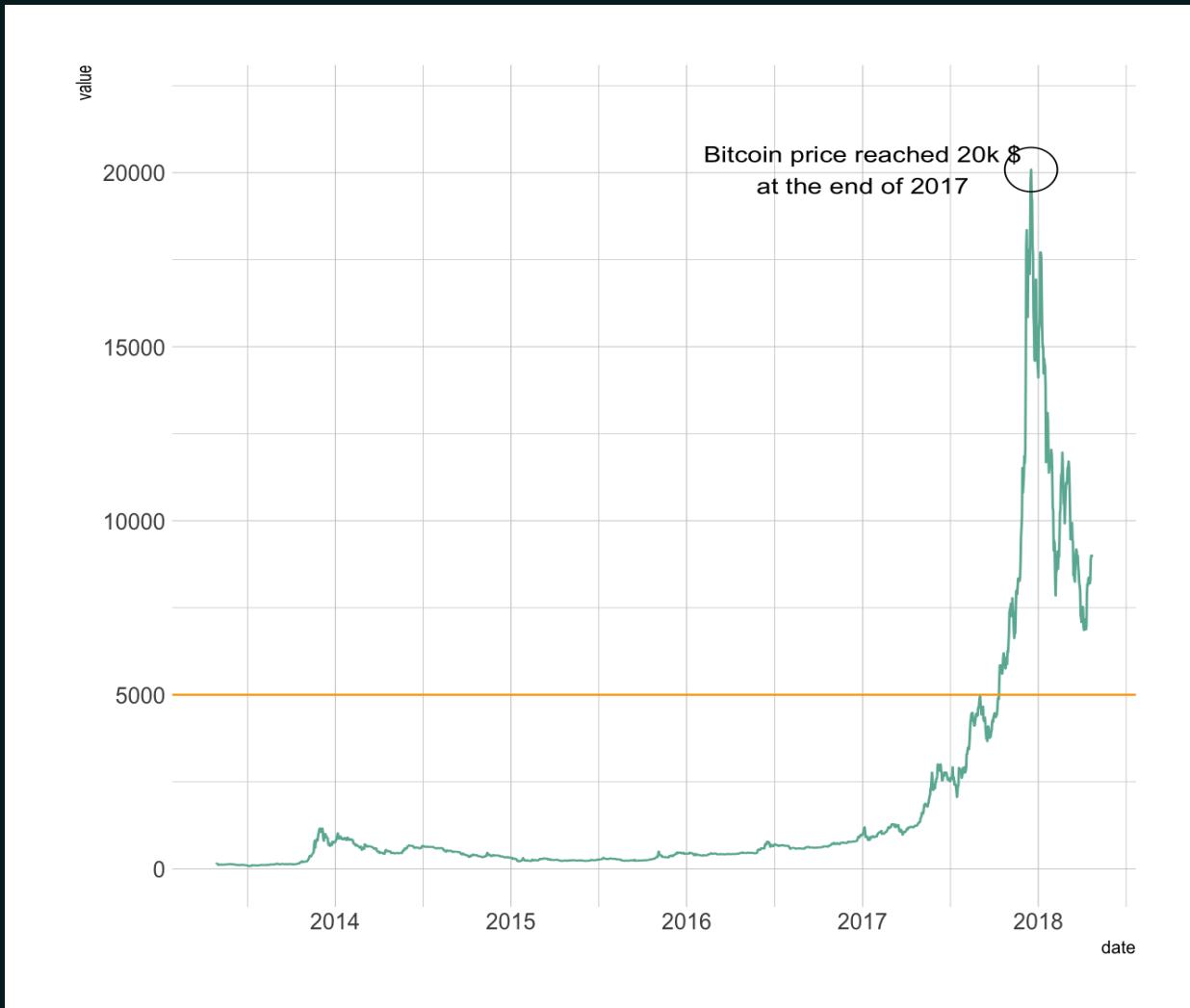
Exemplo de gráfico - ggplot2



Exemplo de gráfico - ggplot2



Exemplo de gráfico - ggplot2



Definição geral do pacote ggplot2

Mapeia as suas variáveis através da camada **aesthetics**, e informa a **geometria** (pontos, linhas, barras, histogramas) que será desenhada sobre um **sistema específico de coordenadas**.

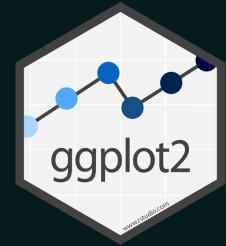
Outras informações como **cor**, **tamanho**, **legenda**, **rótulos** de variáveis e **temas** são definidos para detalhar os gráficos.

[Site do pacote ggplot2](#)

Tipos de gráficos com ggplot2

Tipos de gráficos ggplot2

- Histogramas;
- Densidade;
- Boxplot;
- Violino;
- Barras;
- Dispersão;
- Linhas;
- Circular barplot;



Tipos de gráficos ggplot2



- Gráficos de área;
- Gráficos dotplot;
- Multivariadas;
- Gráficos animados e interativos;
- Gráficos de textura;
- Wordcloud;
- Heat maps.

Galeria de gráficos com ggplot2

Opções de geometrias do ggplot2

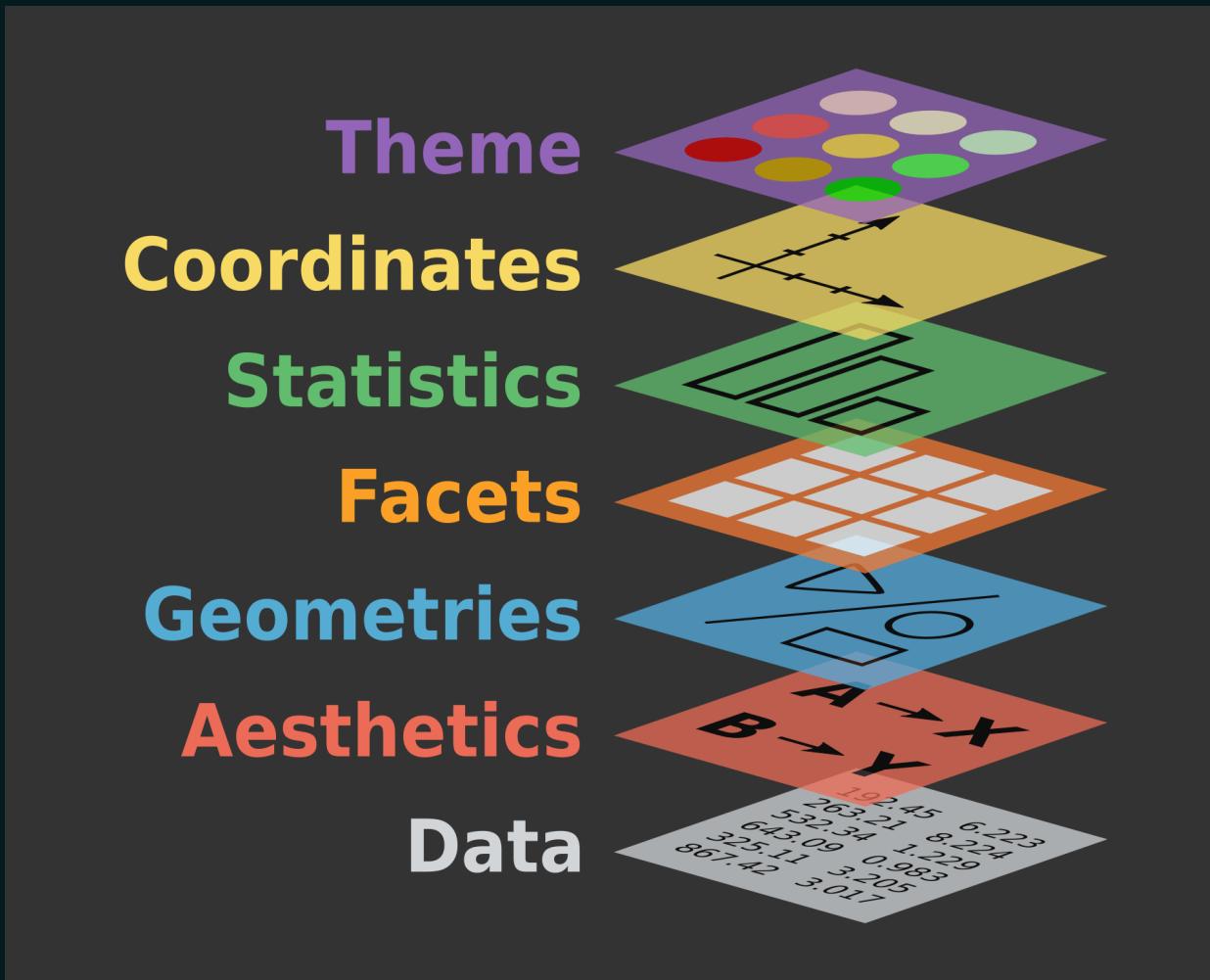
Teoria da Gramática de Gráficos

Ideia de camadas

- Implementação da Gramática de Gráficos;
- Conjunto independente de componentes compostos em diferentes formas.



Teoria da Gramática de Gráficos



Teoria da Gramática de Gráficos

Camadas do ggplot2

- Dados
- Variáveis
- Geometrias
- Estatísticas
- Escalas
- Sistemas de coordenadas
- Facetas
- Rótulos e legendas
- Temas

Etapas para criação de gráficos usando o ggplot2

Etapas iniciais necessárias para criação dos gráficos

1º Definir o diretório de trabalho:

```
# Para saber qual seu diretório atual  
getwd()  
  
# Para criar um novo diretório de trabalho  
dir.create("evento_rladies_ribeirao_preto")  
  
# Para estabelecer o novo diretório criado  
setwd("evento_rladies_ribeirao_preto")  
  
# Estabelecer diretório manualmente  
Session > Set Working Directory > Choose Directory
```

Etapas iniciais necessárias para criação dos gráficos

2º Carregar pacotes:

```
# Pacote para produção dos gráficos
install.packages("ggplot2")
library(ggplot2)

# Pacote usado para análises descritivas
install.packages("dplyr")
library(dplyr)

# Conjunto de pacotes para manipulação e visualização de dados
install.packages("devtools")
devtools::install_github("tidyverse/tidyverse")
library(tidyverse)

# Pacote auxiliar para mudar cores do gráfico
install.packages("devtools")
remotes::install_github("mtennekes/cols4all")
library(cols4all)
```

Etapas iniciais necessárias para criação dos gráficos

3º Criar ou carregar conjunto de dados:

```
# Criar novo conjunto de dados
nomes = c(rep("A", 20), rep("B", 8), rep("C", 30), rep("D", 80))
valores = c(sample(2:5, 20, replace = T), sample(4:10, 8, replace = T),
            sample(1:7, 30, replace = T), sample(3:8, 80, replace = T))

# Gerar tabela
dados = data.frame(nomes, valores)

# Carregar conjunto de dados do R
datasets::iris

# Importar tabela csv, excel ou txt do computador
read.csv()
read.table()
readxl::read.xlsx()
```

Construção de gráficos no RStudio

Construção de gráficos - Exercícios

Boxplot

- Carregar e fazer a filtragem dos dados da tabela;
- Definir as variáveis;
- Definir o tipo de gráfico;
- Definir uma cor para cada categoria do boxplot;
- Definir um tema.

Construção de gráficos - Exercícios

Dispersão

- Definir as variáveis e o tipo de gráfico;
- Adicionar a camada da linha de tendência;
- Adicionar a escala de cores;
- Definir os títulos dos eixos;
- Acrescentar um tema ao gráfico e os detalhes dos eixos;
- Modificar o tamanho da fonte dos títulos dos eixos;
- Adicionar legenda.

Construção de gráficos - Exercícios

Barras

- Calcular a média e o erro padrão com o pacote dplyr;
- Definir as variáveis e o tipo de gráfico;
- Definir os títulos dos eixos;
- Acrescentar um tema ao gráfico;
- Reduzir a largura das barras e retirar a legenda;
- Mudar os nomes dos eixos;
- Ordenar as barras da maior para menor.

Referências

- Wickham, H; Grolemund, G. R for Data Science: Import, Tidy, Transform, Visualize and Model Data. 1. ed. Canada: O'Reilly, 2016. 520 p.
- Wickham, H. ggplot2: Elegant Graphics for Data Analysis. 2. ed. Houston: Springer, 2016. 268 p.
- Rougier, N. P. et al. Ten Simple Rules for Better Figures. PLoS Computational Biology, v. 10, p. 1-7, sep. 2014.
- Galeria de gráficos com ggplot2
- Camadas do ggplot2
- Pacote dplyr
- Advertências para produção de gráficos