

# Explorando a visualização de dados no R com *tidyverse*

Ministrante: Jeanne Franco

Meetup R-Ladies Natal

22/03/2024

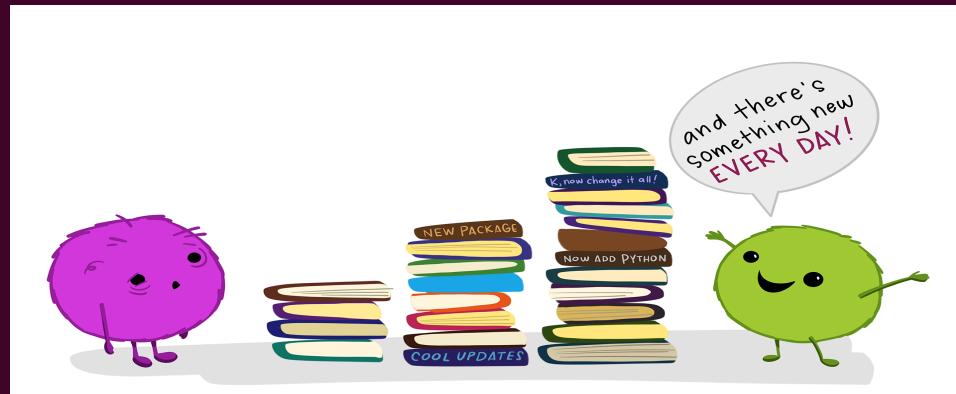
Um pouco sobre mim

# Formação e experiência com o R

- Mestrado em Ciências Florestais, graduação e doutorado em Ecologia (UFRN);
- Coorganizadora da comunidade R-Ladies desde 2020;
- Colaboradora da empresa Beta Analítica;
- Pacotes: *tidyverse* e *rmarkdown*;
- Contato: LinkedIn, Github e Instagram.

# Conteúdo

- Introdução ao pacote *tidyverse*;
- Introdução aos pacotes auxiliares *readr*, *tibble*, *forcats* e *dplyr*;
- Introdução ao pacote de visualização de dados *ggplot2*;
- Teoria da Gramática de Gráficos;
- Principais tipos de gráficos;
- Prática no RStudio.



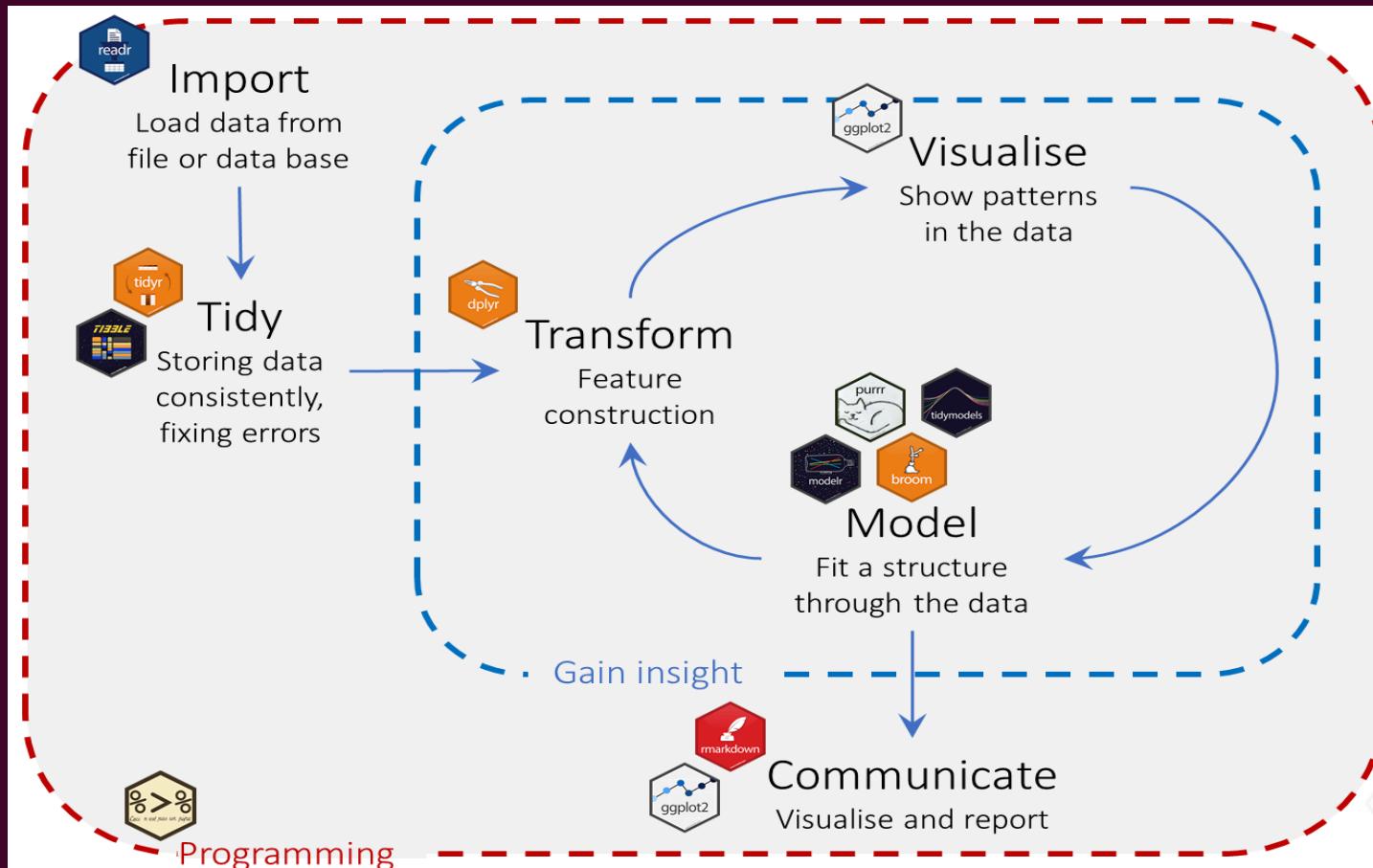
# Introdução ao pacote *tidyverse*

# Pacote *tidyverse*

'Meta-pacote' que apresenta um conjunto de pacotes do R projetados para ciência de dados.



# Pacote *tidyverse*



# Pacotes do *tidyverse*



- *dplyr*: manipulação de dados.
- *tidyr*: organização de dados.
- *readr*: importação de dados.
- *purrr*: ferramentas para trabalhar com funções e vetores.
- *tibble*: data.frames modernos e simplificados.
- *stringr*: ferramentas para trabalhar com caracteres.
- *forcats*: ferramentas para lidar com variáveis categóricas (fatores).
- *ggplot2*: visualização de dados.

# Introdução aos pacotes auxiliares na visualização de dados *readr*, *tibble*, *forcats* e *dplyr*

# Pacote *readr*

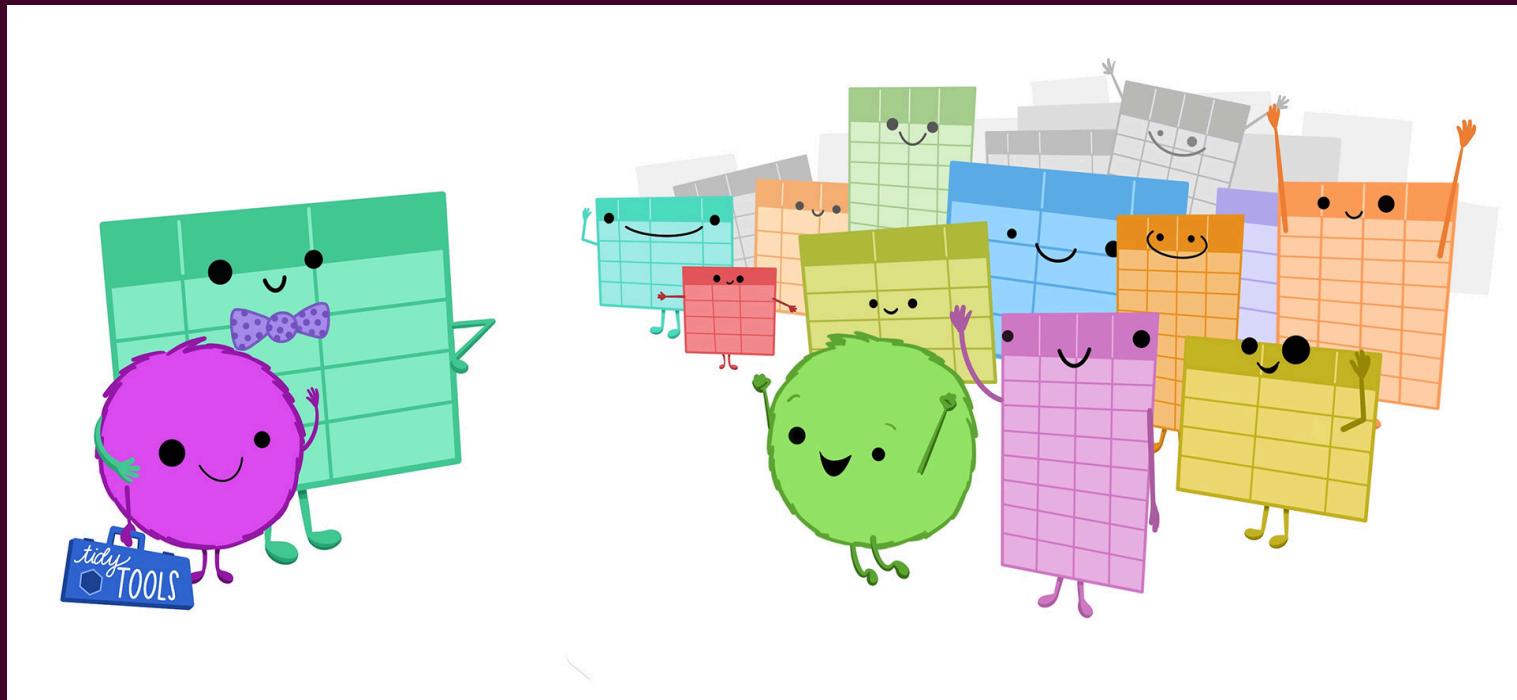
Promove uma amigável forma de ler arquivos.



- **read\_csv()**: valores separados por vírgula (CSV) (ponto como marcador decimal) - (,).
- **read\_csv2()**: valores separados por ponto e vírgula (e vírgula para decimais) - (;).
- **read\_tsv()**: valores separados por tabulação (TSV) - (\t).
- **read\_delim()**: permite o usuário especificar o tipo de separador.
- **read\_table()**: arquivos separados por espaços em branco.
- **read\_excel()**, **read\_xls()** e **read\_xlsx**: leitura de arquivos (.xls ou .xlsx) em planilhas de Excel - Arquivos sem separador explícito.

# Pacote *tibble*

Versão moderna de data.frames. Tibbles possuem um método de visualização aprimorado e mais fácil de usar com grandes conjuntos de dados.



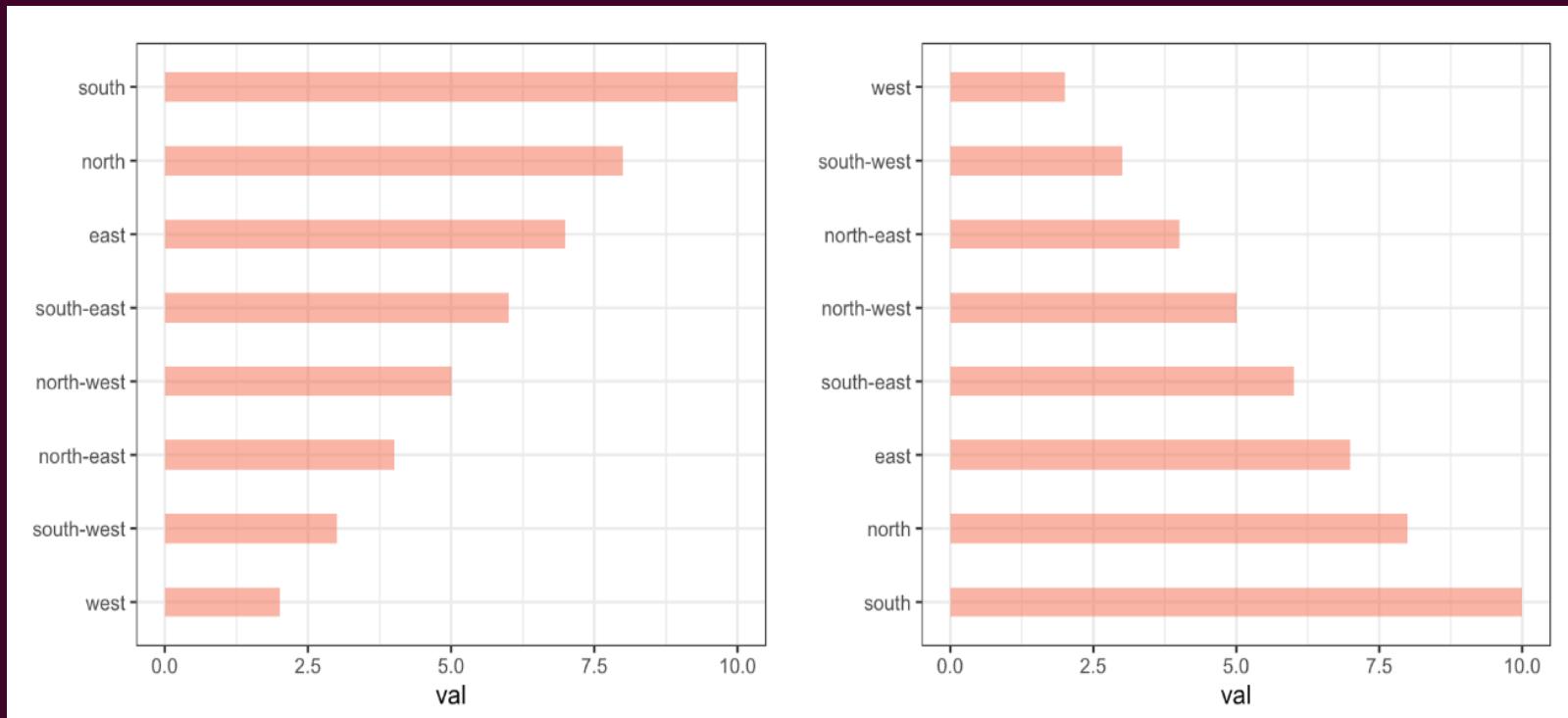
# Características do pacote *tibble*



- **tibble** = "tbl"; **data.frame** = "data.frame".
- Estrutura de dados com **variáveis nas colunas e observações nas linhas**;
- Quando você imprime um tibble, ele **mostra apenas as primeiras dez linhas e todas as colunas que cabem em uma tela**.
- Ele **nunca altera o tipo de uma entrada** (ou seja, não há mais `stringsAsFactors = FALSE!`).
- Ele também **imprime uma descrição abreviada do tipo de coluna** e usa estilos de fonte e cores para realçar.

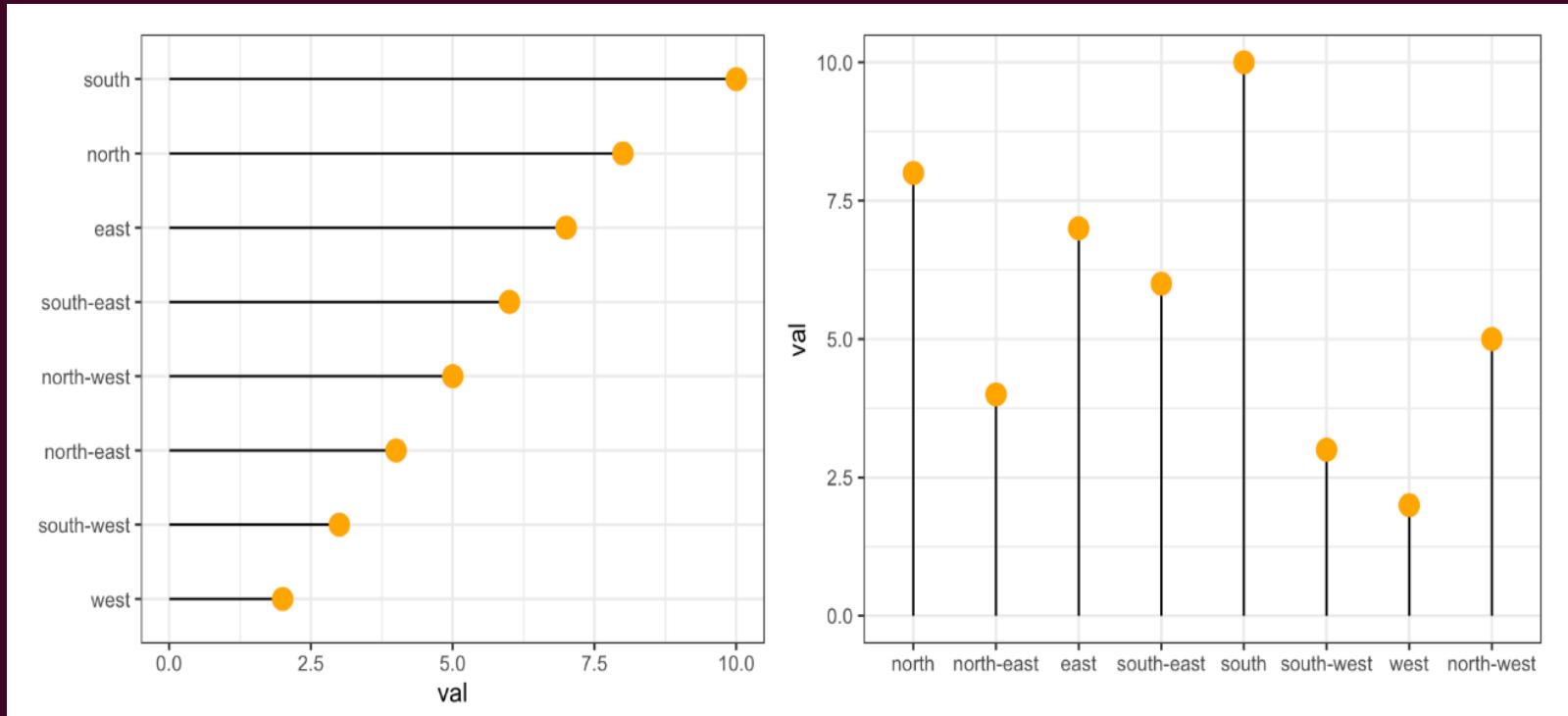
# Pacote *forcats*

Lida com variáveis categóricas. Reordena vetores de caracteres e níveis para melhorar a exibição do gráfico.



# Pacote *forcats*

Lida com variáveis categóricas. Reordena vetores de caracteres e níveis para melhorar a exibição do gráfico.



# Funções do pacote *forcats*



- `fct_reorder()`: reordena um fator por uma outra variável.
- `fct_infreq()`: reordena um único fator pela frequência dos valores.
- `fct_relevel()`: altera a ordem de um fator manualmente.
- `fct_lump()`: une diversos fatores similares em novas variáveis categóricas.

# Pacote *dplyr*

Pacote que fornece um conjunto de funções que ajudam a resolver os desafios mais comuns da manipulação de dados.

**dplyr::filter()**

KEEP ROWS THAT  
satisfy  
*your CONDITIONS*

keep rows from... this data... ONLY IF... type is "otter" AND site is "bay"

```
filter(df, `type == "otter" & `site == "bay")
```

type	food	site
otter	urchin	bay
Shark	seal	channel
otter	abalone	bay
otter	crab	channel

@allison\_horst

# Pacote *dplyr*

Pacote que fornece um conjunto de funções que ajudam a resolver os desafios mais comuns da manipulação de dados.



# Funções do pacote *dplyr*



- **mutate()**: cria 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;
- **rename()**: renomeia nomes de variáveis.

# Introdução ao pacote de visualização de dados *ggplot2*

# Pacote *ggplot2*

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

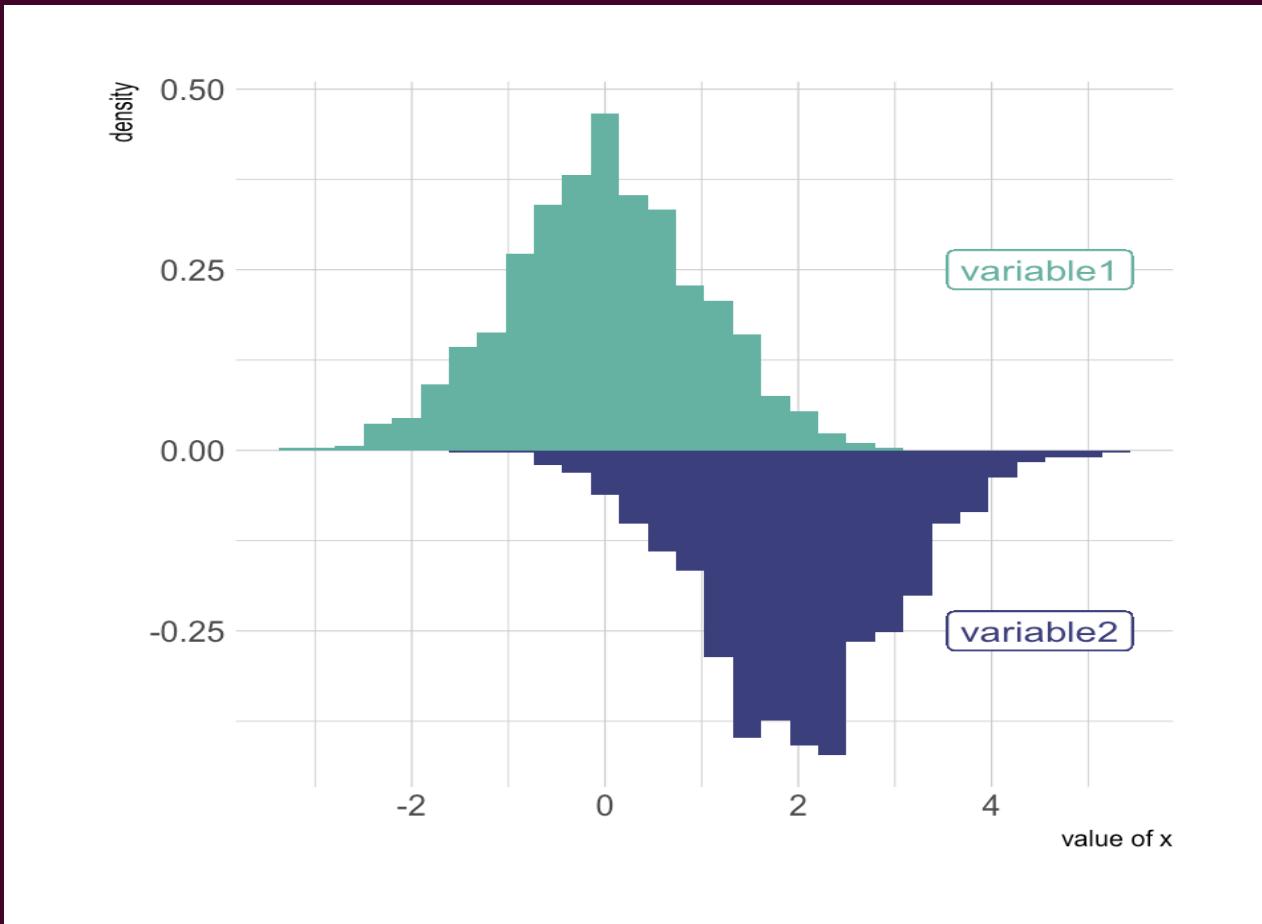


# Pacote *ggplot2*

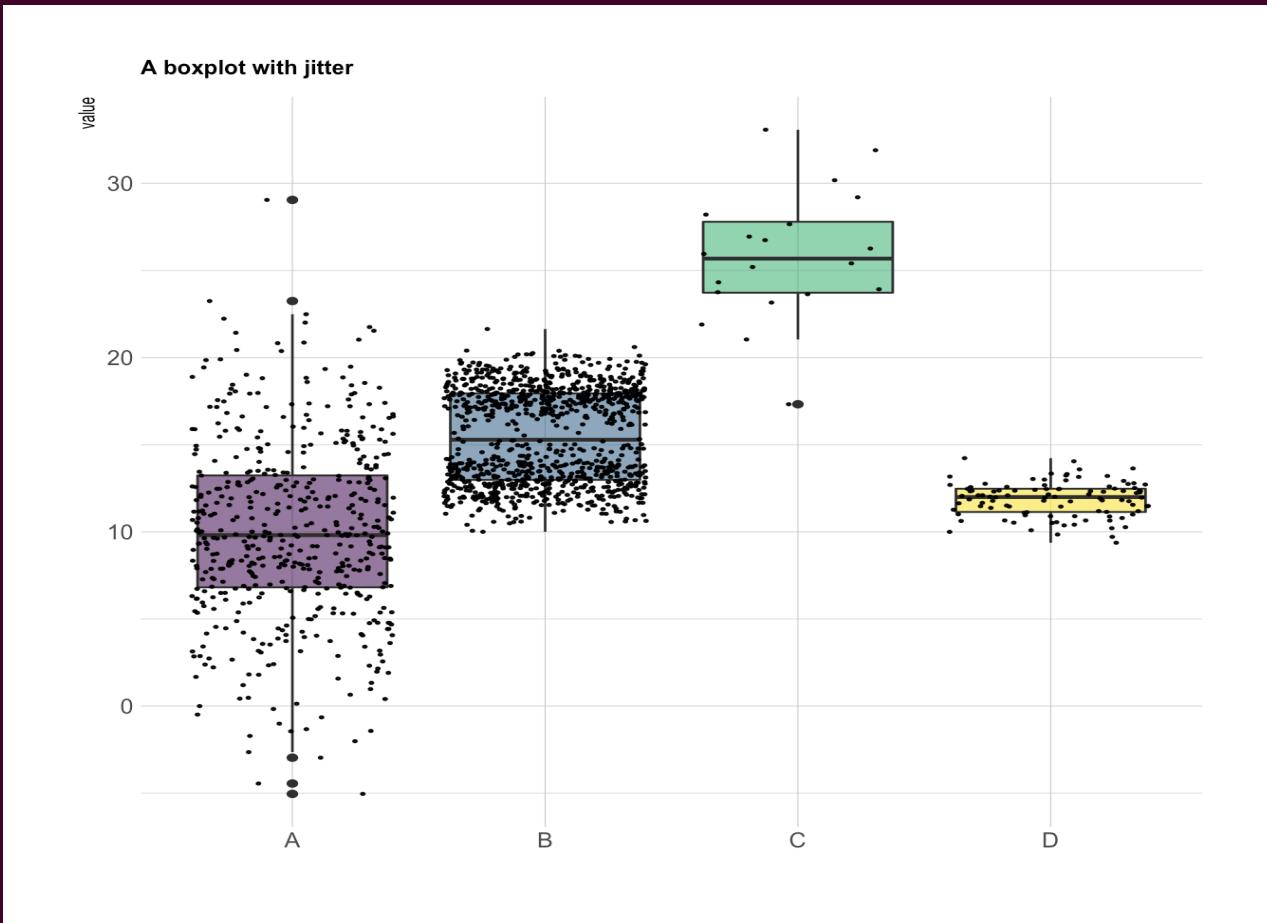
Gráficos elegantes e versatéis que permitem adição de detalhes e transmissão das principais mensagens.



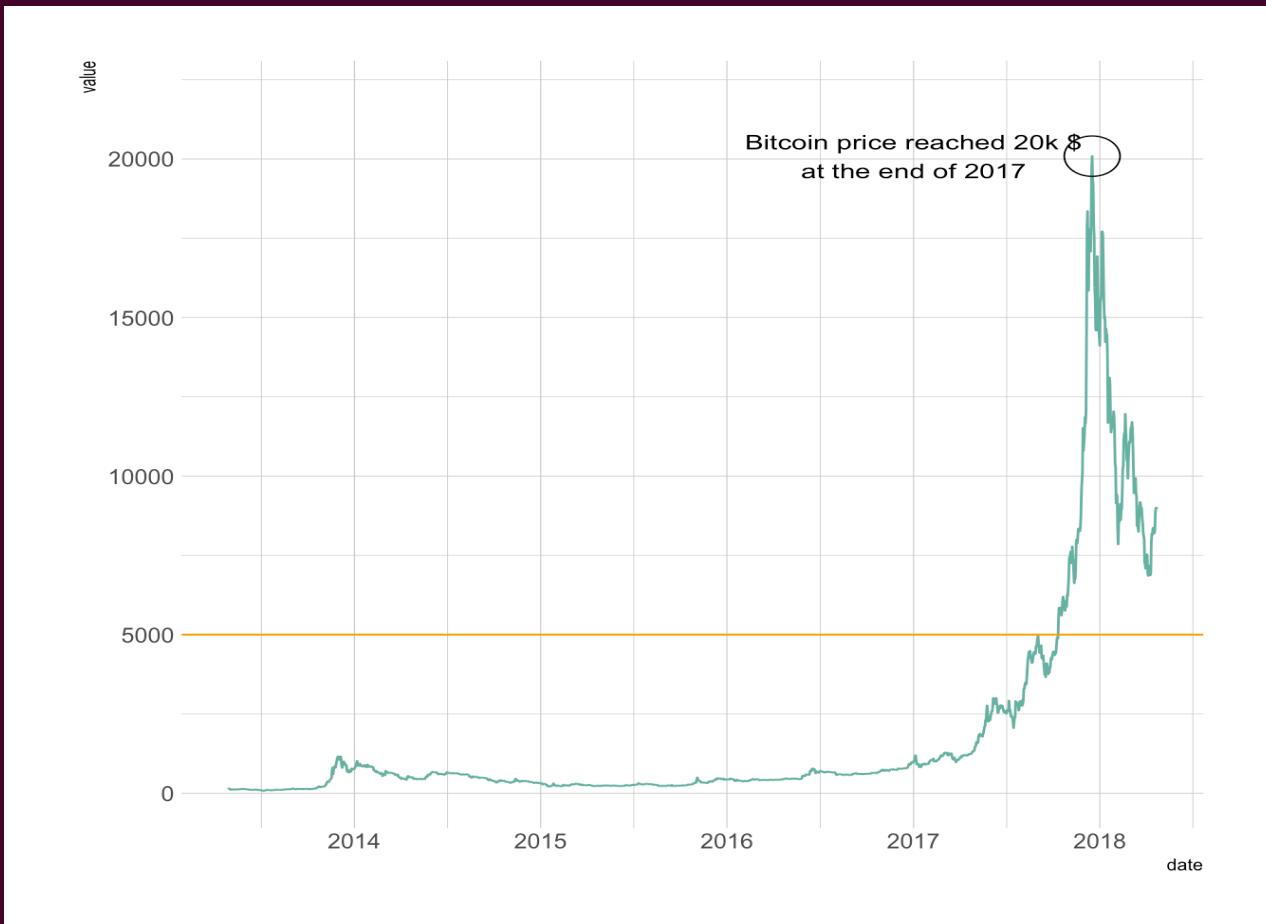
# Exemplo de gráfico



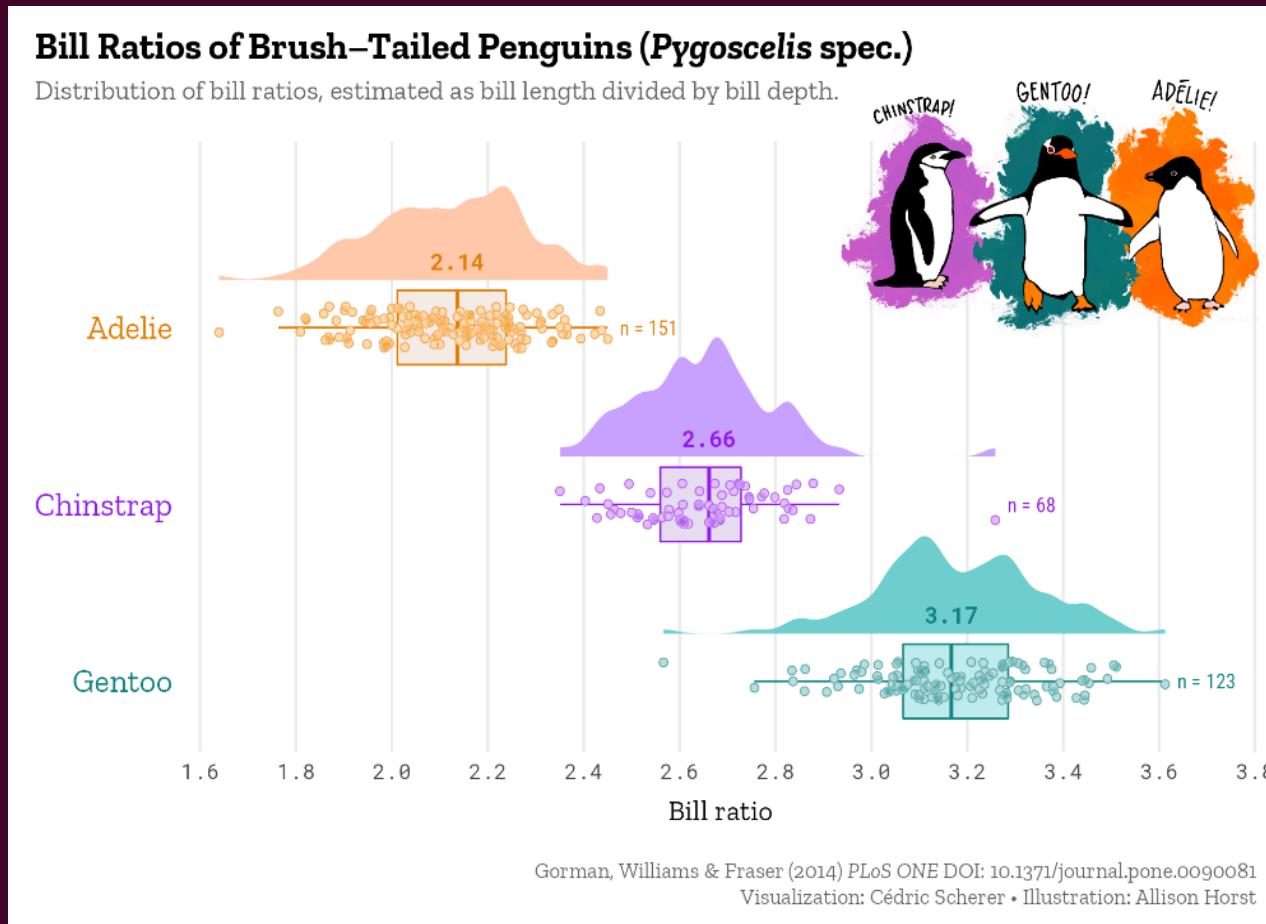
# Exemplo de gráfico



# Exemplo de gráfico



# Exemplo de gráfico



# Ideia de camadas

Implementação da Gramática de Gráficos;

Conjunto independente de componentes compostos em diferentes formas.



# Teoria da Gramática de Gráficos

## Camadas do ggplot2



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

# Tipos de gráficos com ggplot2

# Tipos de gráficos ggplot2



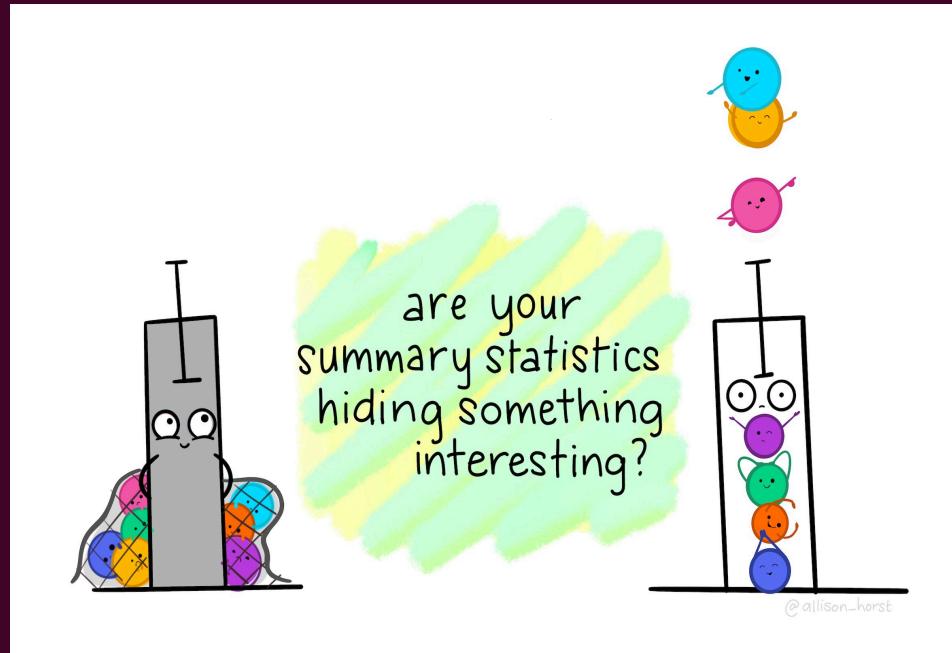
- Histogramas;
- Densidade;
- Boxplot;
- Violino;
- Barras;
- Dispersão;
- Linhas;
- Circular barplot;
- Gráficos de área;
- Gráficos dotplot;
- Multivariadas;
- Gráficos animados e interativos;
- Gráficos de textura, wordcloud e mapas.

→ Galeria de gráficos com ggplot2

→ Opções de geometrias do ggplot2

# Ressalvas na visualização de dados

"Suas estatísticas de resumo escondem algo interessante?"



→ Advertências para produção de gráficos

# Recomendação de referências

- 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*
- Pacote *tidyverse*
- Advertências para produção de gráficos
- Artes by Allison Horst

# Recomendação de livro

