

**Técnicas estatísticas:
teoria e prática
(*R Programming*)**



PÓS-GRADUAÇÃO

Elaborando gráficos estatísticos com o R


Bloco 1

Marcelo Tavares de Lima





► Objetivos

- Apresentar os principais gráficos estatísticos.
 - Apresentar os principais comandos da linguagem R para elaboração de gráficos estatísticos.
 - Desenvolver exemplos de aplicação para elaboração de gráficos na linguagem R.
- 



► Introdução

- O que são gráficos estatísticos?
- Quais suas vantagens?
- A linguagem R para elaboração de gráficos estatísticos.



► Introdução

- Bussab e Morettin (2017, p. 6) afirmam que “os métodos gráficos têm encontrado um uso cada vez maior devido ao seu forte apelo visual”.
- O uso de gráficos, em geral, é mais fácil de ser compreendido quando comparado a informações contidas em tabelas ou resumos numéricos.



► Introdução

- Chambers et. al. (1983 apud Bussab e Morettin, 2017, p. 6) afirmam que gráficos são utilizados para diversos fins, tais como:
 - Buscar padrões e relações.
 - Confirmar (ou não) certas expectativas que se tinha sobre os dados.
 - Descobrir novos fenômenos; confirmar (ou não) suposições feitas sobre os procedimentos estatísticos usados.
 - Apresentar resultados de modo mais rápido e fácil.



► Gráficos estatísticos com o R

- Vamos utilizar a IDE (*Integrated Development Environment* ou Ambiente de Desenvolvimento Integrado) Rstudio.
- Existem gráficos apropriados para cada tipo de variável.

► Gráficos estatísticos com o R

- Para a execução dos comandos é necessário baixar os dados no seu computador e, em seguida, fazer a importação para o R através da IDE RStudio via menu de opções:

File → Import Dataset →
From Text (base)...

Selecionar a pasta onde os
dados estão armazenados →
Import.



► Gráficos estatísticos com o R

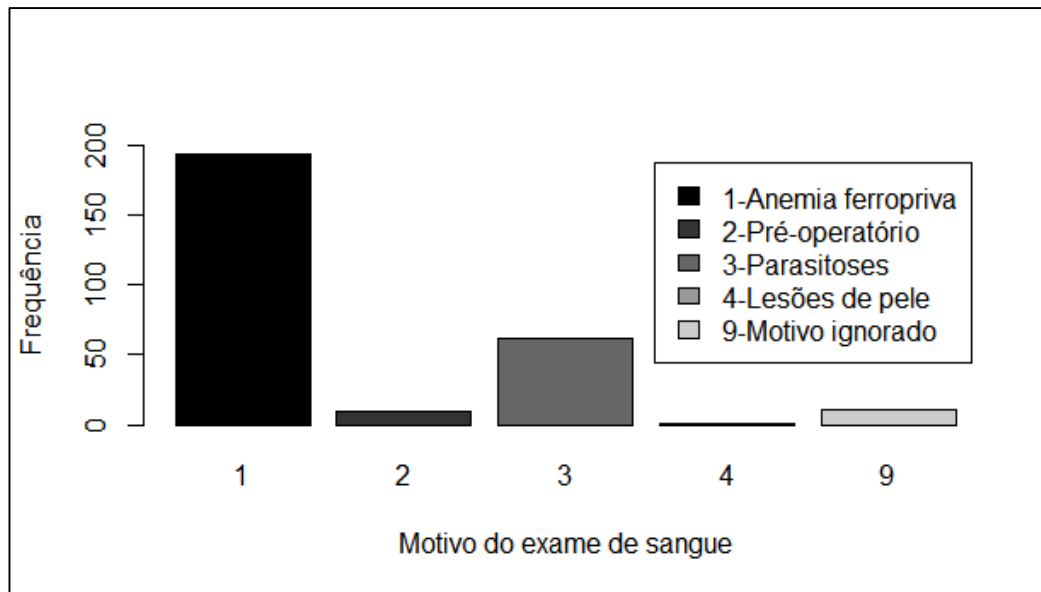
- O banco de dados “Toxo” (Siqueira; Tibúrcio, 2011) tem inúmeras variáveis, no entanto, iremos utilizar a variável referente ao motivo da realização de exame de sangue para produzir um gráfico de colunas.

► Gráficos estatísticos com o R

```
barplot(table(Exame),  
xlab="Motivo do exame de sangue",  
ylab="Frequência",  
ylim=c(0,200),  
legend.text=c("1-Anemia ferropriva",  
"2-Pré-operatório",  
"3-Parasitoses",  
"4-Lesões de pele",  
"9-Motivo ignorado"),  
col=gray(0:5/5))
```

► Gráficos estatísticos com o R

Figura 1 – Gráfico de barras para o motivo de exame



Fonte: adaptada de Siqueira e Tibúrcio (2011).

► Gráficos estatísticos com o R

- Vamos elaborar um gráfico de setores (pizza).

Tabela 1 – Escolaridade de pessoas de uma amostra.

Grau de instrução	Frequência	Porcentagem
Fundamental	12	33,33
Médio	18	50,00
Superior	6	16,67
Total	36	100,00

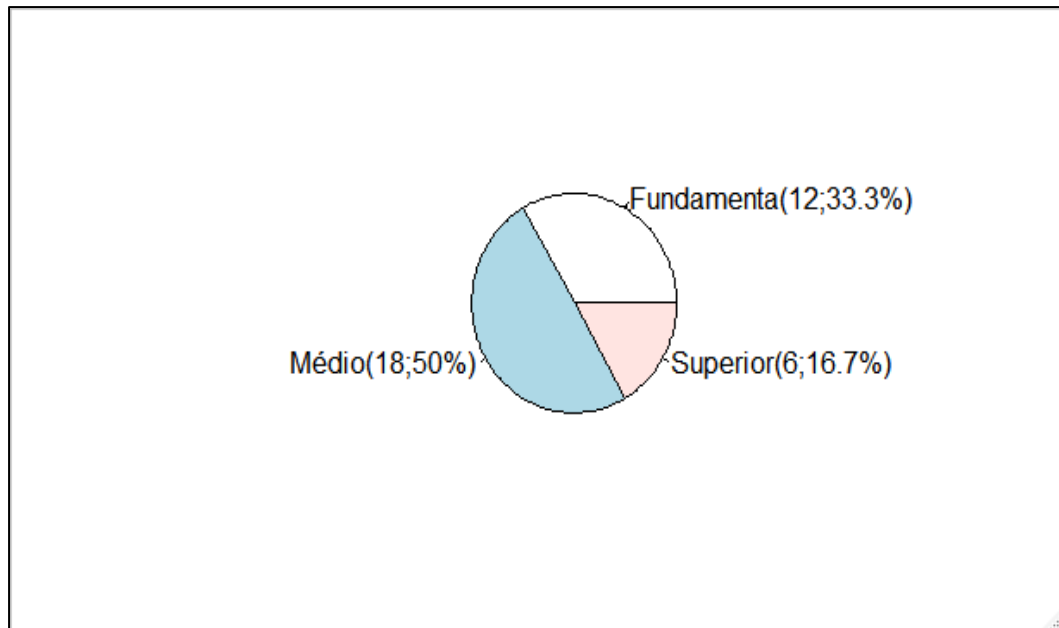
Fonte: adaptada de Bussab e Morettin (2017).

► Gráficos estatísticos com o R

```
# é separada a variável escolaridade do banco de dados tab2_1 para
# facilitar a execução dos comandos
escolaridade<-tab2_1$grau_instrucao
# é criada a tabela para a variável escolaridade
x<-table(escolaridade)
# são criados rótulos para as categorias
# da variável escolaridade
rotulos<-paste( c("Fundamental","Médio",
"Superior"),"(",c(12,18,6),";",
round(c(33.33,50,16.67),1),"%)",sep="")
# é criado o gráfico de pizzas com os
argumentos de rotulação e legenda
pie(x,labels=rotulos)
```

► Gráficos estatísticos com o R

Figura 2 – Gráfico de setores.



Fonte: adaptada de Bussab e Morettin (2017).

► Gráficos estatísticos com o R

Vamos elaborar um gráfico de dispersão unidimensional

tabela de frequência de valores da variável para verificar as frequências

```
table(tab2_1$n_filhos)
```

#resultado

```
0 1 2 3 5 # valores da variável
```

```
4 5 7 3 1 # frequência dos valores da variável
```

```
stripchart(tab2_1$n_filhos, # base de dados
```

```
  method = "stack", # elementos empilhados
```

```
  offset = 1, # espaçamento entre os elementos
```

```
  pch = 19, # formato do elemento (19 = circular)
```

```
  col="darkblue", # cor do elemento
```

```
  ylim=c(0,7), # eixo vertical
```

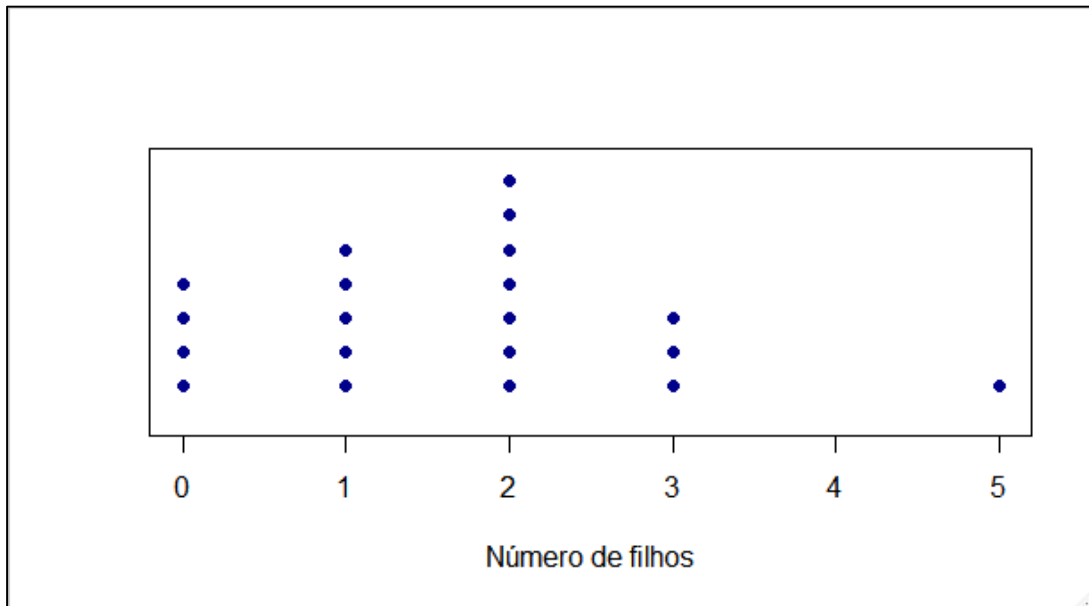
```
  xlab="Número de filhos", # rótulo horizontal
```

```
  cex=1 # tamanho dos elementos.
```

```
)
```

► Gráficos estatísticos com o R

Figura 3 – Gráfico de dispersão unidimensional.



Fonte: adaptada de Bussab e Morettin (2017).

► Gráficos estatísticos com o R

Vamos elaborar um histograma.

Para verificar um resumo da variável para descobrir o valor mínimo e máximo da idade.

```
summary(tab2_1$idade_anos)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
20.00	30.00	34.50	34.58	40.00	48.00

Para construir um histograma com a variável idade (em

anos) do banco de dados tab2_1, com alguns parâmetros

determinados como, título do eixo horizontal e vertical,

valores de idade divididos em classe de valores de

amplitude igual a cinco, sem título principal no gráfico

com valores de frequência absoluta no eixo vertical e

variando de 0 a 15 e rótulos em cada coluna e,

colunas de cor azul claro.

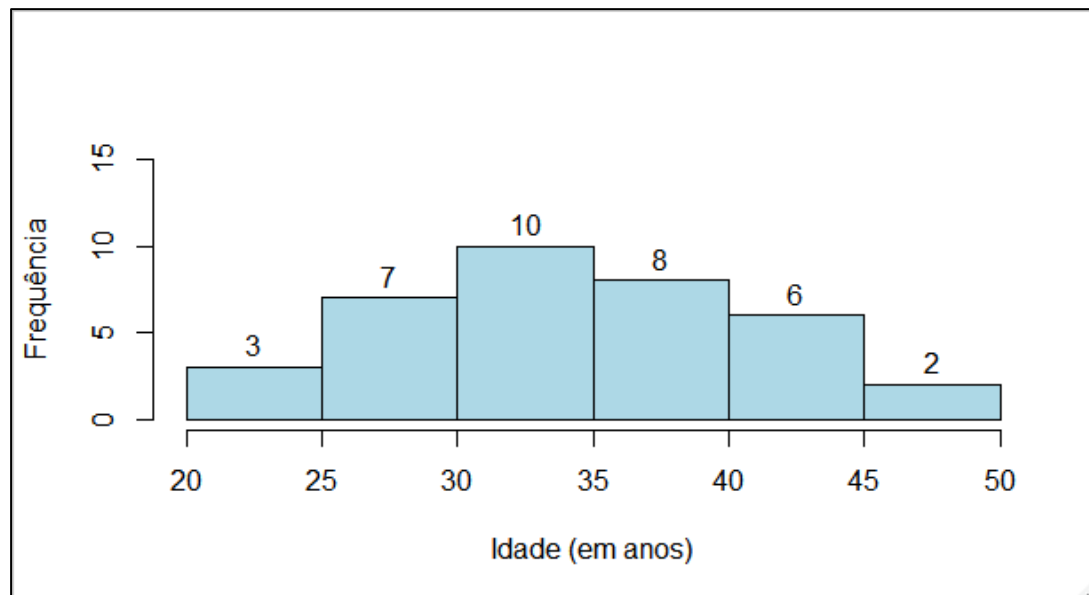
```
hist(tab2_1$idade_anos,xlab="Idade (em anos)",
```

```
ylab="Frequência", main='', col="lightblue",
```

```
breaks=seq(20,50, by=5), labels=TRUE,ylim=c(0,15))
```

► Gráficos estatísticos com o R

Figura 4 – Histograma



Fonte: adaptada de Bussab e Morettin (2017).

PÓS-GRADUAÇÃO

Elaborando gráficos estatísticos com o R

Bloco 2

Marcelo Tavares de Lima



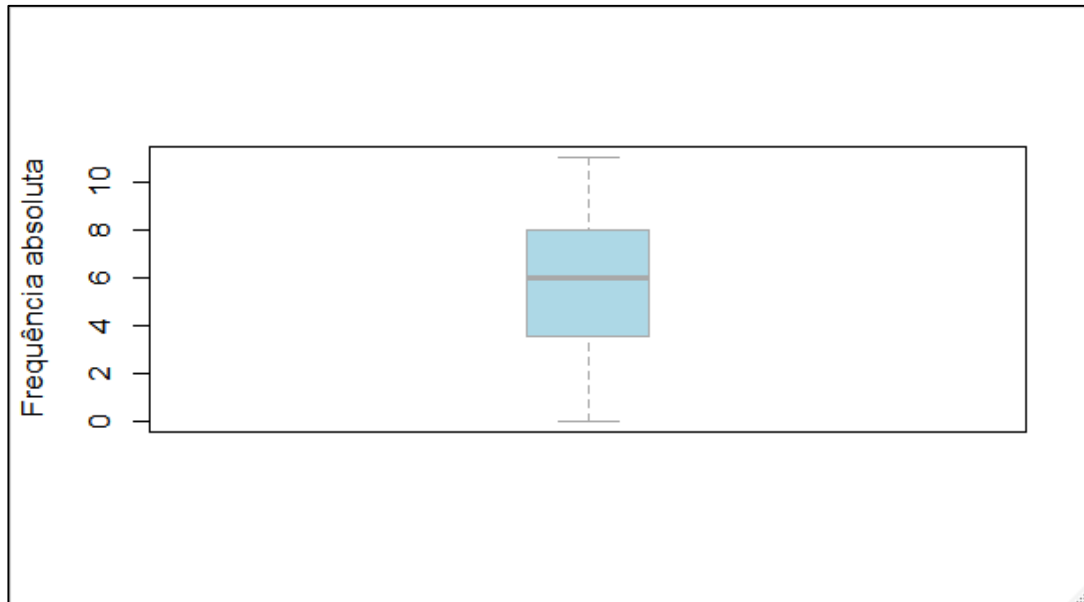
► Gráficos estatísticos com R (continuação)

Vamos elaborar um boxplot.

```
boxplot(tab2_1$idade_meses,  
        pch="*", # tipo de marcador dos outliers (só aparece se tiver um valor  
        identificado como outlier  
        col="lightblue", # cor do preenchimento do boxplot  
        border="darkgrey", # cor da linha do box plot  
        boxwex=0.3, # Tamanho da caixa  
        ylab="Frequência absoluta" # rótulo eixo vertical  
)
```

► Gráficos estatísticos com R (continuação)

Figura 5 – Boxplot.



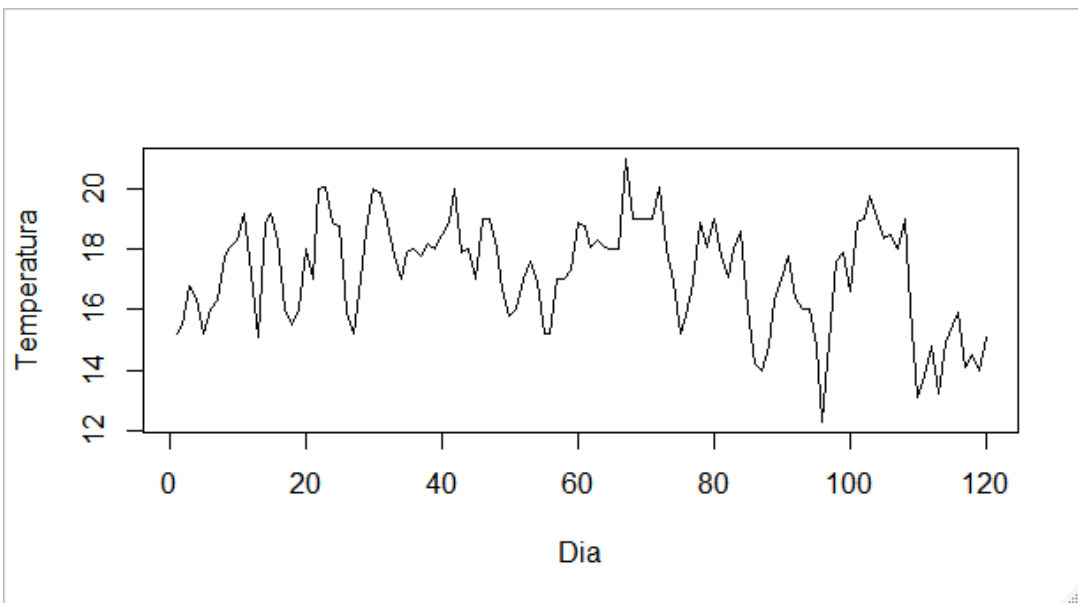
Fonte: Adaptado de Bussab e Morettin (2017).

► Gráficos estatísticos com R (continuação)

Vamos elaborar um gráfico de séries temporais.

```
plot.ts(cd_poluicao$temp, xlab="Dia", ylab="Temperatura")
```

Figura 6 – Série temporal da Temperatura (°C).



Fonte: adaptada de Bussab e Morettin (2017).

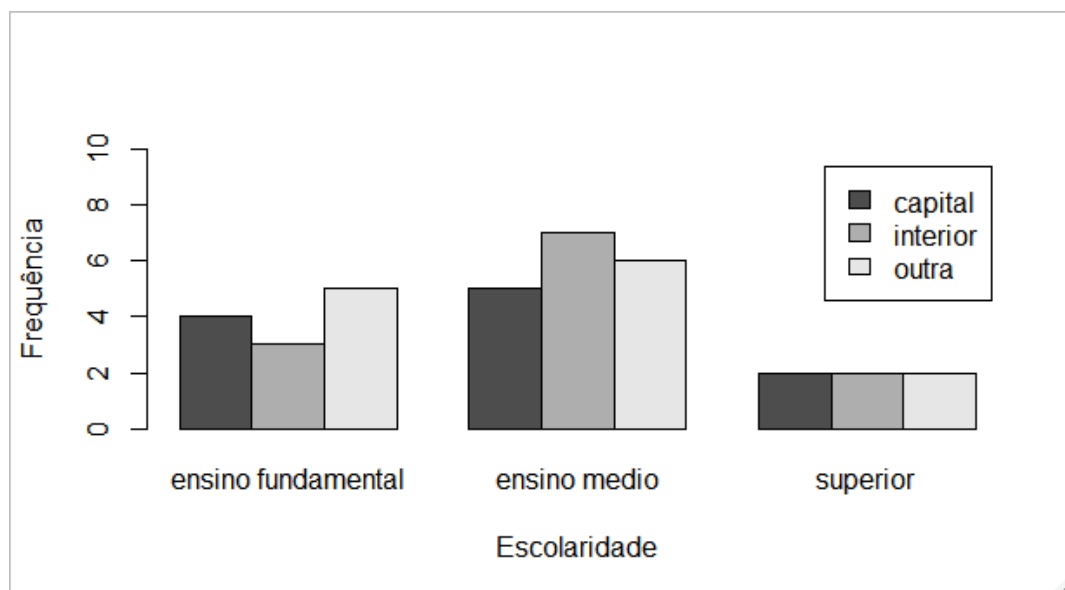
► Gráficos estatísticos com R (continuação)

Vamos elaborar um gráfico de colunas para duas variáveis.

```
attach(tab2_1) # anexar o banco de dados
tabela<-table(reg_procedencia, grau_instrucao)
tabela # verificar as frequências da tabela
barplot(tabela, xlab='Escolaridade',
        ylab='Frequência', ylim=c(0,10),
        legend=TRUE, beside=TRUE)
```

► Gráficos estatísticos com R (continuação)

Figura 7 – Gráfico de colunas.



Fonte: adaptada de Bussab e Morettin (2017).

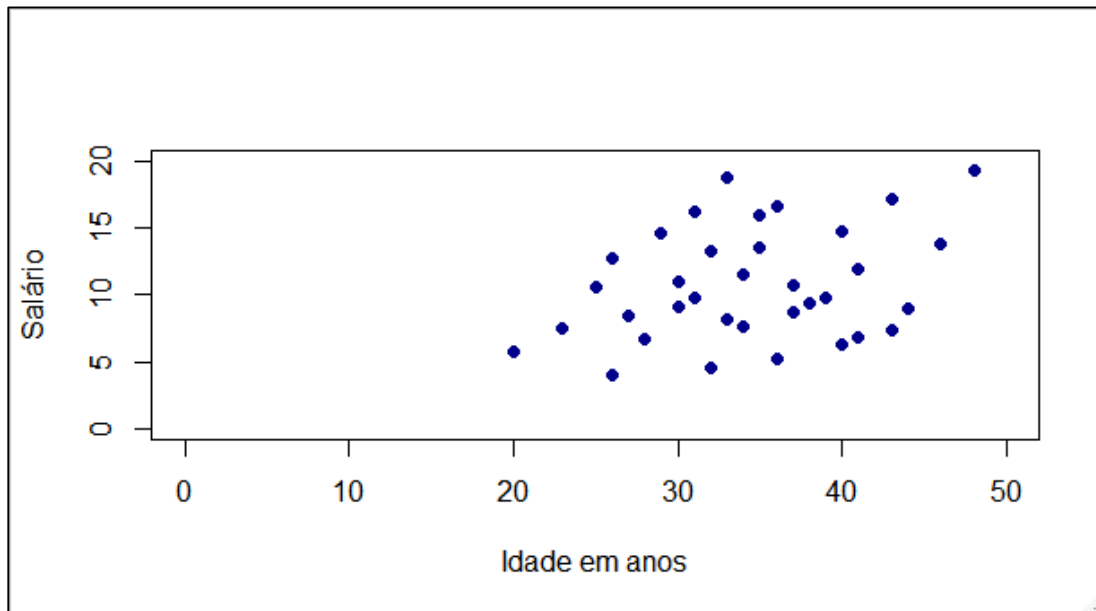
► Gráficos estatísticos com R (continuação)

Vamos fazer um gráfico de dispersão.

```
plot(tab2_1$IdadeAnos, # variável do eixo horizontal
      tab2_1$salario, # variável do eixo vertical
      ylim = c(0,20), # faixa de valores do eixo y
      xlim=c(0,50), # faixa de valores do eixo x
      pch=16, # tipo de elemento gráfico para (x,y)
      col="darkblue", # cor dos elementos gráficos
      xlab = "Idade em anos", # rótulo do eixo x
      ylab = "salário" # rótulo do eixo y
      )
```

► Gráficos estatísticos com R (continuação)

Figura 8 – Gráfico de dispersão.



Fonte: elaborada pelo autor.

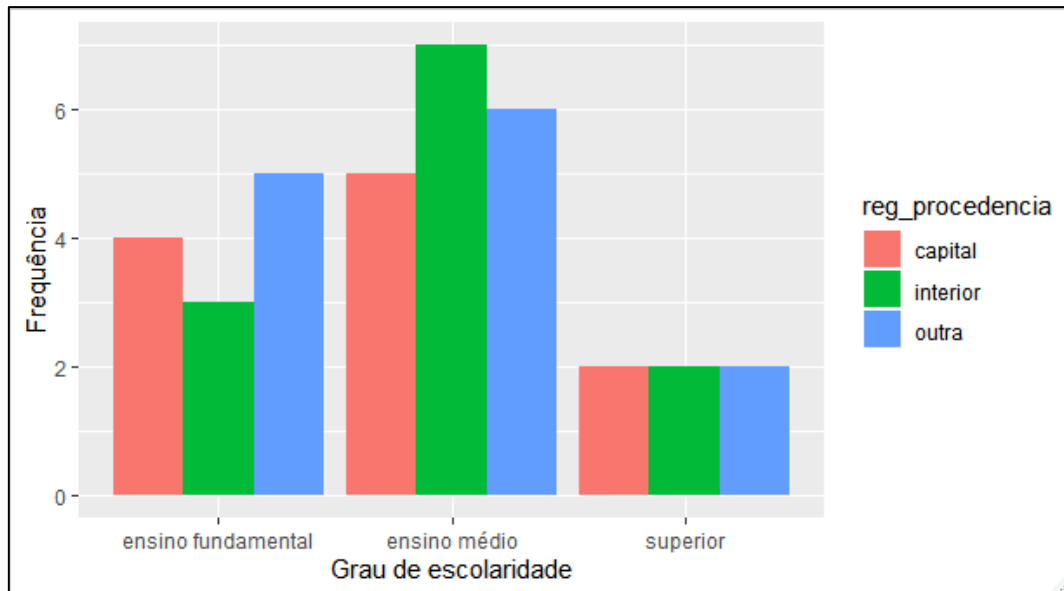
► Gráficos estatísticos com R (continuação)

Vamos elaborar um gráfico de colunas com o pacote ggplot2.

```
g<-ggplot(data = tab2_1) +  
  geom_bar(  
    mapping = aes(x = grau_instrucao, # variável  
                  fill = reg_procedencia), # preenchimento das colunas  
    position = "dodge" # colunas lado a lado  
  )  
# Rótulos(títulos)  
g <- g +  
  labs(  
    y = 'Frequência', # rótulo vertical  
    x = 'Grau de escolaridade' # rótulo horizontal  
  )  
g # impressão do gráfico
```

► Gráficos estatísticos com R (continuação)

Figura 9 – Gráfico de colunas com o ggplot2.



Fonte: elaborada pelo autor.

PÓS-GRADUAÇÃO

Teoria em prática

Bloco 3

Marcelo Tavares de Lima



► Elaborando gráfico estatístico no RStudio

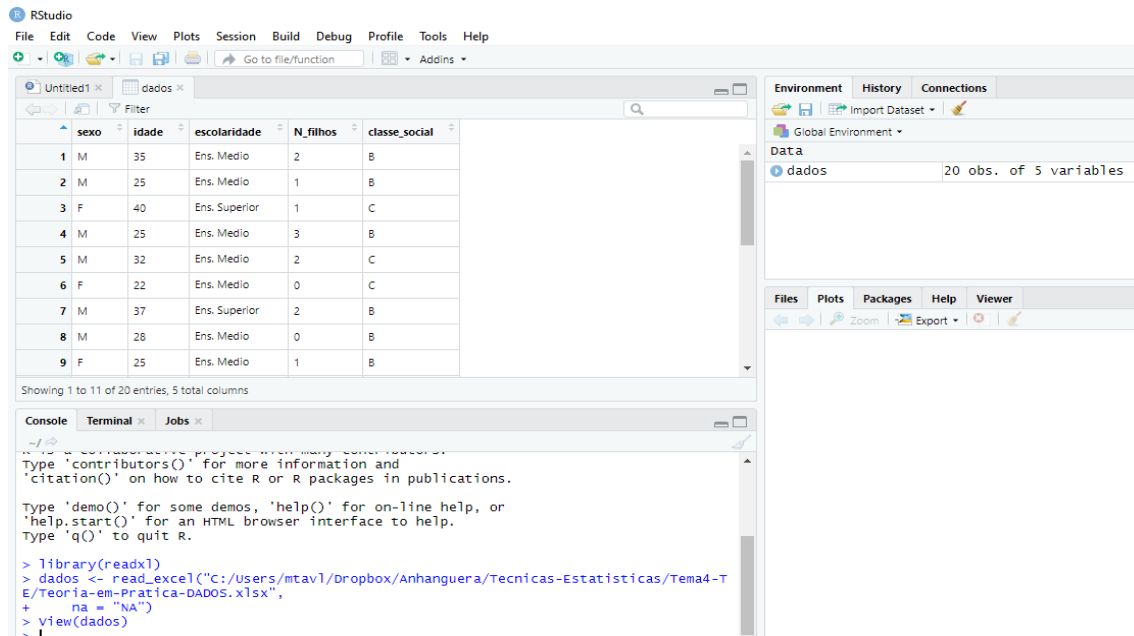
Figura 10 – Conjunto de dados

	A	B	C	D	E
1	sexo	idade	escolaridade	N_filhos	classe_social
4	F	40	Ens. Superior	1	C
5	M	25	Ens. Medio	3	B
6	M	32	Ens. Medio	2	C
7	F	22	Ens. Medio	0	C
8	M	37	Ens. Superior	2	B
9	M	28	Ens. Medio	0	B
10	F	25	Ens. Medio	1	B
11	F	39	Ens. Superior	2	C
12	M	35	Ens. Fundamental	1	B
13	F	21	Ens. Fundamental	0	A
14	F	27	NA	0	A
15	F	45	Ens. Medio	2	C
16	M	57	Pos-Graduacao	4	C
17	F	33	Ens. Medio	2	A
18	M	36	Ens. Fundamental	0	B
19	M	35	Ens. Medio	2	C
20	M	33	Ens. Medio	2	B
21	F	22	Ens. Superior	0	C

Fonte: Costa (2012)

► Manipulação de dados no RStudio

Figura 11 – Conjunto de dados importados para o RStudio



The screenshot displays the RStudio environment with a data frame named 'dados' loaded. The data is shown in a table with 20 rows and 5 columns. The columns are 'sexo', 'idade', 'escolaridade', 'N_filhos', and 'classe_social'. The first 9 rows are visible, showing a mix of male and female individuals with various ages, education levels, and number of children. The console shows the R code used to load the data from an Excel file.

	sexo	idade	escolaridade	N_filhos	classe_social
1	M	35	Ens. Medio	2	B
2	M	25	Ens. Medio	1	B
3	F	40	Ens. Superior	1	C
4	M	25	Ens. Medio	3	B
5	M	32	Ens. Medio	2	C
6	F	22	Ens. Medio	0	C
7	M	37	Ens. Superior	2	B
8	M	28	Ens. Medio	0	B
9	F	25	Ens. Medio	1	B

```
> library(readxl)
> dados <- read_excel("C:/users/mtav1/Dropbox/Anhanguera/Tecnicas-Estatisticas/Tema4-T
E/Teoria-em-Pratica-DADOS.xlsx",
+   na = "NA")
> view(dados)
> |
```

Fonte: captura de tela de RStudio.

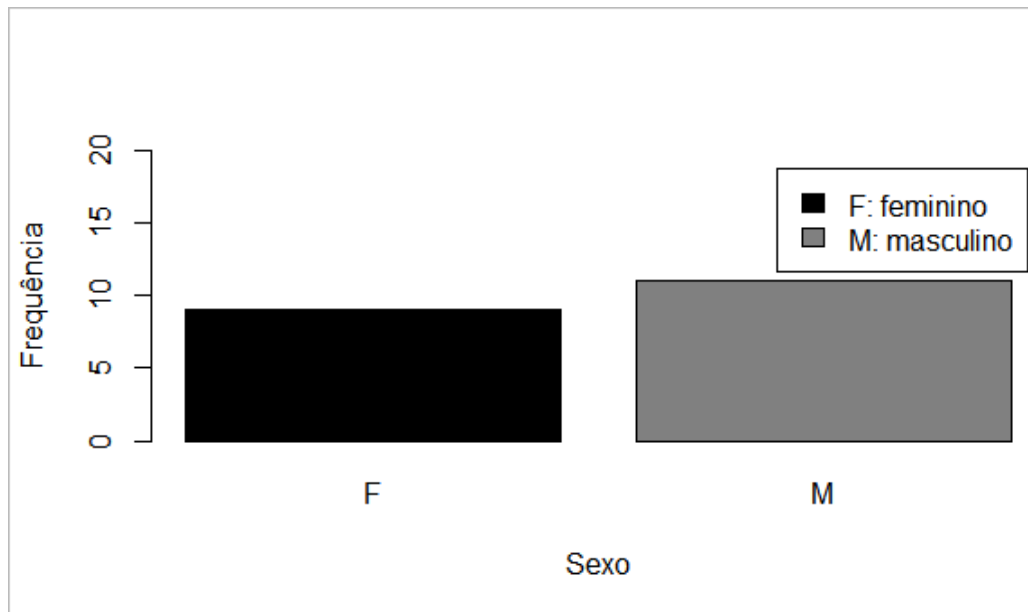
► Manipulação de dados no RStudio

Comando:

```
barplot(table(dados$sexo),  
        xlab="Sexo",  
        ylab="Frequência",  
        ylim = c(0,20),  
        legend.text = c("F: feminino", "M: masculino"),  
        col = gray(0:2/2)  
)
```


► Manipulação de dados no RStudio

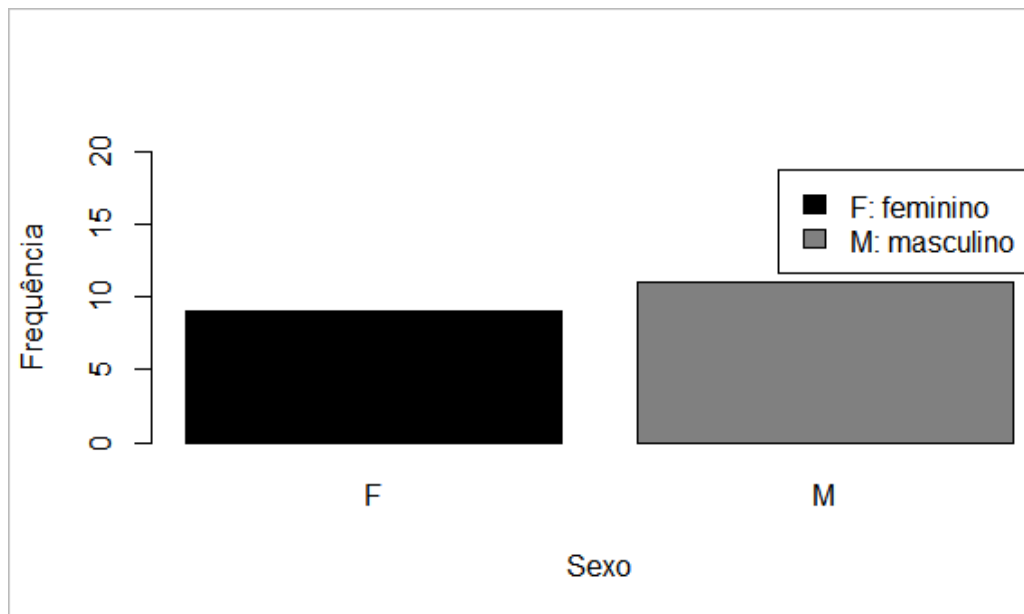
Figura 10 – Gráfico de colunas



Fonte: elaborada pelo autor.

► Manipulação de dados no RStudio

Figura 10 – Gráfico de colunas



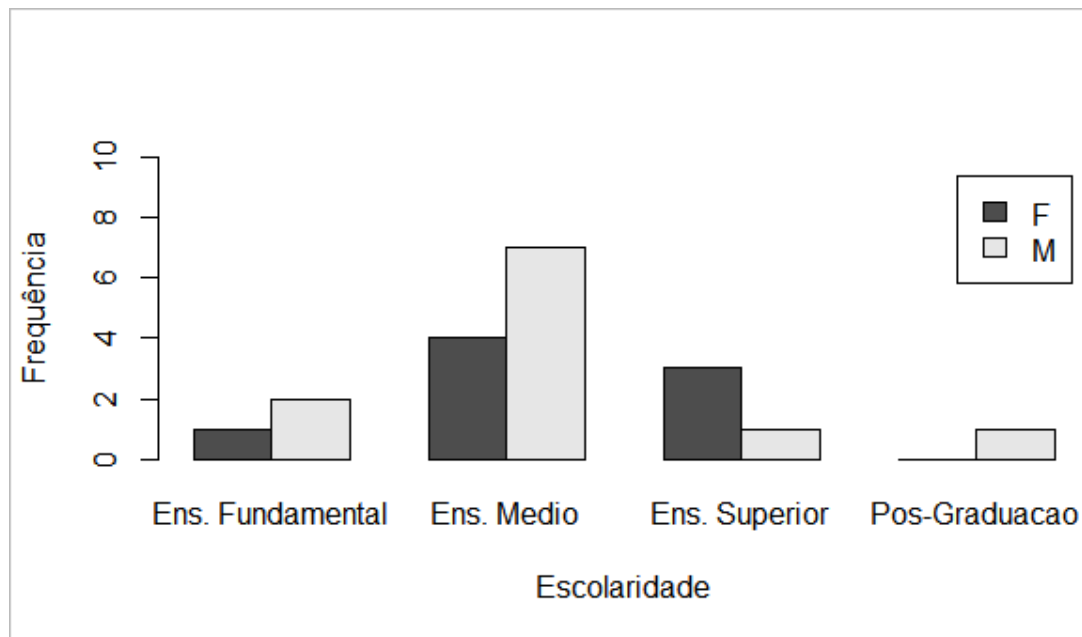
Fonte: elaboração do autor.

► Manipulação de dados no RStudio

```
barplot(table(dados$sexo,dados$escolaridade),  
        xlab='Escolaridade',  
        ylab='Frequência', ylim=c(0,10),  
        legend=TRUE,  
        beside=TRUE)
```

► Manipulação de dados no RStudio

Figura 11 – Gráfico de colunas



Fonte: elaboração do autor.

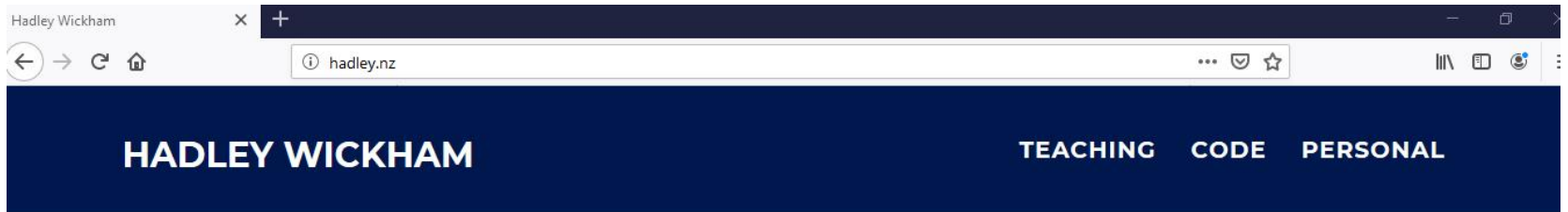
Dica do professor

Bloco 4

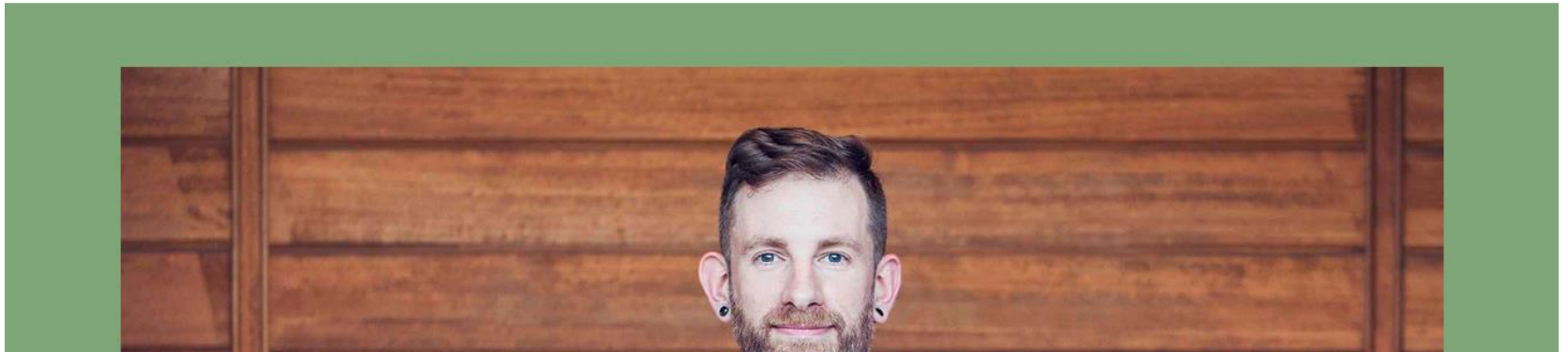
Marcelo Tavares de Lima



► Criador do ggplot2



Hi! I'm Hadley Wickham, Chief Scientist at [RStudio](#), and an Adjunct Professor of Statistics at the [University of Auckland](#), [Stanford University](#), and [Rice University](#). I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand but I currently live in Houston, TX with my partner and dog.





► Referências

BUSSAB, Wilton.; MORETTIN, Pedro A. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017. 554p.

COSTA, G. G. de O. **Curso de estatística inferencial e probabilidades**: teoria e prática. São Paulo: Atlas, 2012.

SIQUEIRA, A. L., TIBÚRCIO, J. D. **Estatística na área da saúde**: conceitos, metodologia, aplicações e prática computacional. Belo Horizonte: Coopmed, 2011.

