

R para Data Science

Solução dos exercícios

Jeidsan A. da C. Pereira

2023-10-25

Conteúdo

Introdução	7
I Explorar	9
1 Visualização de dados com ggplot2	11
1.1 Introdução	11
1.2 Primeiros passos	11
1.3 Mapeamentos estéticos	13
1.4 Problemas comuns	13
1.5 Facetas	14
1.6 Objetos geométricos	14
1.7 Transformações estatísticas	14
1.8 Ajustes de posição	14
1.9 Sistemas de coordenadas	14
1.10 A gramática em camadas de gráficos	14
2 Fluxo de trabalho: o básico	15
3 Transformação de dados com dplyr	17
4 Fluxo de trabalho: scripts	19
5 Análise exploratória de dados	21
6 Fluxo de trabalho: projetos	23

II	Wrangle	25
7	Tibbles com <code>tibble</code>	27
8	Importando dados com <code>readr</code>	29
9	Arrumando dados com <code>tidyr</code>	31
10	Dados relacionais com <code>dplyr</code>	33
11	Strings com <code>stringr</code>	35
12	Fatores com <code>forcats</code>	37
13	Datas e horas com <code>lubridate</code>	39
III	Programar	41
14	Pipes com <code>magrittr</code>	43
15	Funções	45
16	Vetores	47
17	Iteração com <code>purrr</code>	49
18	(PART) Modelar	51
19	O básico de modelos com <code>modelr</code>	53
20	Construção de modelos	55
21	Muitos modelos com <code>purrr</code> e <code>broom</code>	57
IV	Comunicar	59
22	R Markdown	61

<i>CONTEÚDO</i>	5
23 Gráficos para comunicação com ggplot2	63
24 Formatos R Markdown	65
25 Fluxo de trabalho de R Markdown	67

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```


Introdução

Esta página serviu para estudo e prática com o pacote R Bookdown e contém a solução encontrada por mim para os exercícios propostos no livro R para Data Science, de Hadley Wickham e Garret Golemund, publicado no Brasil em 2019 pela Alta Books Editora [Wickham and Golemund, 2019].

Por se tratar de um produto construído durante o processo de aprendizagem, o conteúdo pode conter erros, tanto no texto em si, como na lógica utilizada para solução dos exercícios.

Dúvidas ou sugestões de melhoria podem ser encaminhadas para o e-mail *jeidsan.pereira@gmail.com*.

Parte I

Explorar

Capítulo 1

Visualização de dados com ggplot2

Para a correta execução dos códigos desse capítulo, utilizaremos algumas configurações específicas.

Inicialmente, precisaremos carregar o pacote `nycflights13`, que contém os dados de todos os voos da cidade de Nova York em 2013.

```
library(nycflights13)
```

1.1 Introdução

Não temos exercícios nesta seção.

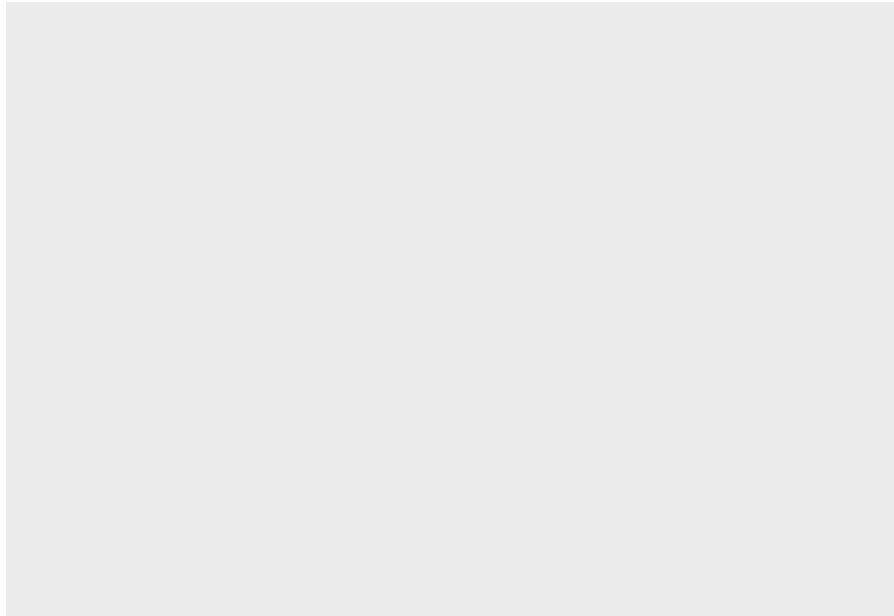
1.2 Primeiros passos

Exercício 1.2.1

Execute `ggplot(data=mpg); .` O que você vê?

Solução.

```
ggplot(data=mpg) +  
  tema
```



É exibido um quadro em branco. Este quadro contém o sistema de coordenadas sobre o qual serão desenhados os gráficos que pretendemos exibir.

Exercício 1.2.2

Quantas linhas existem em `mtcars`? Quantas colunas?

Solução.

```
dim(mtcars)
```

```
## [1] 32 11
```

R.: Existem 32 linhas e 11 colunas.

Exercício 1.2.3

O que a variável `drv` descreve?

Solução. Executamos o comando `?mpg` no console no R e a página de ajuda foi aberta. Nela encontramos o significado de cada variável do conjunto de dados.

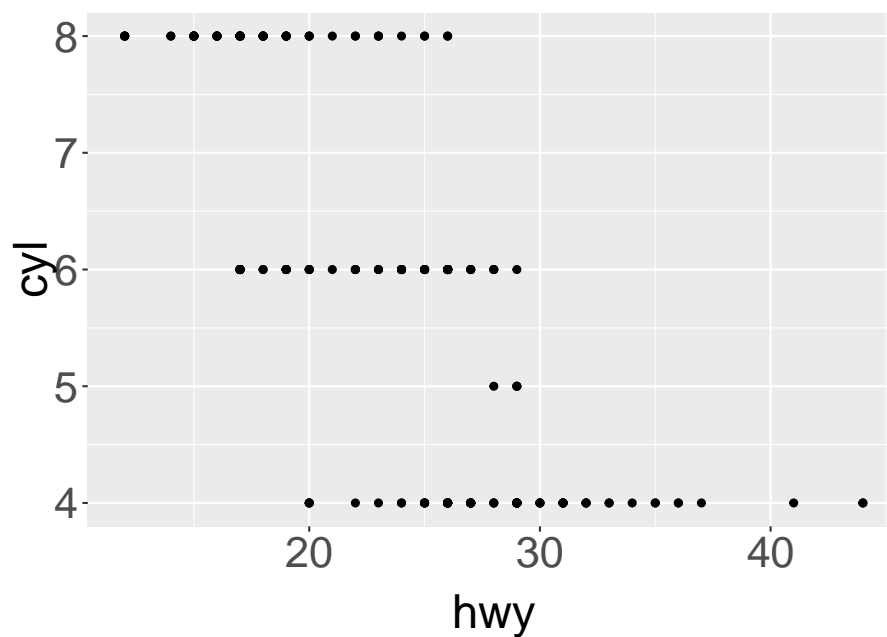
A variável descreve o tipo de tração dos carros analisados, onde `f` significa tração dianteira, `r` significa tração traseira e `4` significa tração nas quatro rodas.

Exercício 1.2.4

Faça um gráfico de dispersão de `hwy` *versus* `cyl`.

Solução.

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = hwy, y = cyl)) +  
  tema
```

**1.3 Mapeamentos estéticos****Exercício 1.3.1**

Exercício 1.1. x

Solução. x

1.4 Problemas comuns

Exercício 1.2. x

1.5 Facetas

Exercício 1.3. [x](#)

1.6 Objetos geométricos

Exercício 1.4. [x](#)

1.7 Transformações estatísticas

Exercício 1.5. [x](#)

1.8 Ajustes de posição

Exercício 1.6. [x](#)

1.9 Sistemas de coordenadas

Exercício 1.7. [x](#)

1.10 A gramática em camadas de gráficos

Exercício 1.8. [x](#)

Capítulo 2

Fluxo de trabalho: o básico

Capítulo 3

Transformação de dados com dplyr

Capítulo 4

Fluxo de trabalho: scripts

Capítulo 5

Análise exploratória de dados

Capítulo 6

Fluxo de trabalho: projetos

Parte II

Wrangle

Capítulo 7

Tibbles com tibble

Capítulo 8

Importando dados com readr

Capítulo 9

Arrumando dados com tidyr

Capítulo 10

Dados relacionais com dplyr

Capítulo 11

Strings com stringr

Capítulo 12

Fatores com forcats

Capítulo 13

Datas e horas com lubridate

Parte III

Programar

Capítulo 14

Pipes com `magrittr`

Capítulo 15

Funções

Capítulo 16

Vetores

Capítulo 17

Iteração com purrr

Capítulo 18

(PART) Modelar

Capítulo 19

O básico de modelos com modelr

Capítulo 20

Construção de modelos

Capítulo 21

Muitos modelos com purrr e broom

Parte IV

Comunicar

Capítulo 22

R Markdown

Capítulo 23

Gráficos para comunicação com ggplot2

Capítulo 24

Formatos R Markdown

Capítulo 25

Fluxo de trabalho de R Markdown

Bibliografia

Hadley Wickham and Garrett Golemund. *R para Data Science*. Alta Books, Rio de Janeiro, 2019.