

Jeidsan A. da C. Pereira

R para Data Science

Solução dos exercícios

To Shao Yong (邵雍),
for sharing a secret joy with simple words;

月到天心处，风来水面时。
一般清意味，料得少人知。

and

To Hongzhi Zhengjue (宏智禅师),
for sharing the peace of an ending life with simple words.

梦幻空华，六十七年；
白鸟淹没，秋水连天。

Conteúdo

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

II Wrangle	19
7 Tibbles com <code>tibble</code>	21
8 Importando dados com <code>readr</code>	23
9 Arrumando dados com <code>tidyr</code>	25
10 Dados relacionais com <code>dplyr</code>	27
11 Strings com <code>stringr</code>	29
12 Fatores com <code>forcats</code>	31
13 Datas e horas com <code>lubridate</code>	33
III Programar	35
14 Pipes com <code>magrittr</code>	37
15 Funções	39
16 Vetores	41
17 Iteração com <code>purrr</code>	43
18 (PART) Modelar	45
19 O básico de modelos com <code>modelr</code>	47
20 Construção de modelos	49
21 Muitos modelos com <code>purrr</code> e <code>broom</code>	51
IV Comunicar	53
22 R Markdown	55
23 Gráficos para comunicação com <code>ggplot2</code>	57

<i>Contents</i>	v
24 Formatos R Markdown	59
25 Fluxo de trabalho de R Markdown	61



Prefácio

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*¹.

¹<mailto:jeidsan.pereira@gmail.com>



Parte I

Explorar



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.

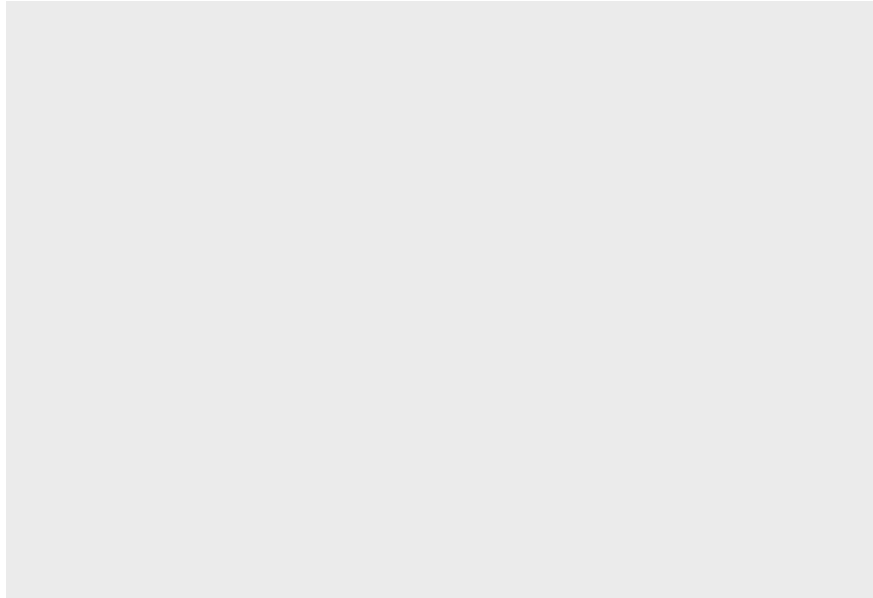
1.2 Primeiros passos

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.

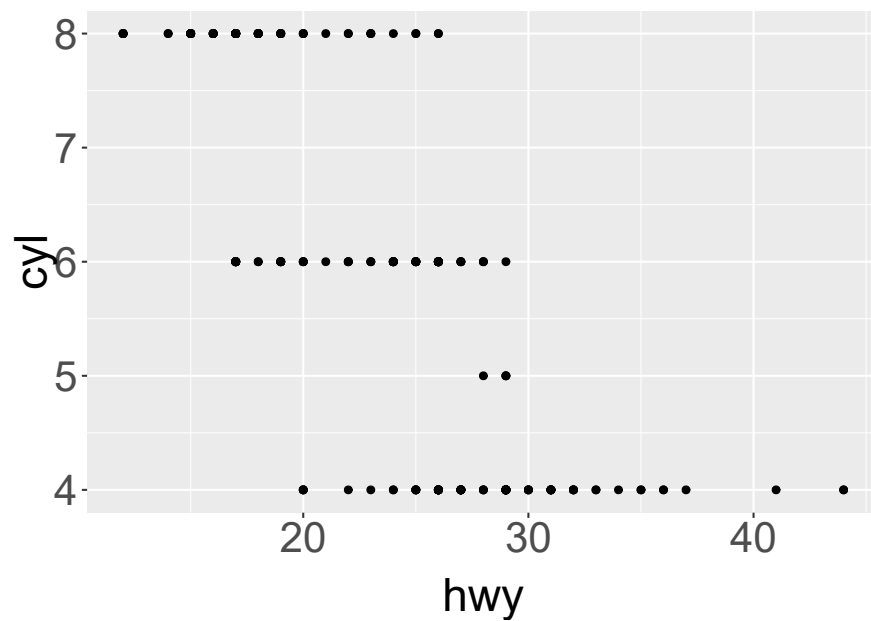
A variável descreve o tipo de tração dos carros analisados, onde `f` significa tração dianteira, `r` significa tração traseira e `4` significa tração nas quatro rodas.

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



2

Fluxo de trabalho: o básico



3

Transformação de dados com `dplyr`



4

Fluxo de trabalho: scripts



5

Análise exploratória de dados



6

Fluxo de trabalho: projetos



Parte II

Wrangle



7

Tibbles com tibble



8

Importando dados com readr



9

Arrumando dados com tidyr



10

Dados relacionais com dplyr



11

Strings com stringr



12

Fatores com forcats



13

Datas e horas com lubridate



Parte III

Programar



14

Pipes com magrittr



15

Funções



16

Vetores



17

Iteração com *purrr*



18

(PART) Modelar



19

O básico de modelos com `modelr`



20

Construção de modelos



21

Muitos modelos com purrr e broom



Parte IV

Comunicar



22

R Markdown



23

Gráficos para comunicação com `ggplot2`



24

Formatos R Markdown



25

Fluxo de trabalho de R Markdown



Bibliografia

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