

Mulheres da Ciência

Análise de dados em R

MARIA LETÍCIA MACHADO
AGOSTO, 2018

A oficina

- O que é e para que serve a **análise exploratória de dados**
- Como a **programação** pode facilitar o trabalho com dados
- Introdução ao ambiente **R e Rstudio**
 - Objetos
 - Operadores lógicos
 - Matrizes
 - Dataframes
 - Funções base
 - Gráficos



Tipos de variáveis

Clássico:

Variáveis qualitativas

Variáveis quantitativas

Ou ainda:

Nominais (usadas, novas, rasgadas...)

Ordinais (excelente, bom, satisfatório, ruim, péssimo)

Discretas (9, 8, 12...)

Contínuas (10,5 mm, 8,34 mm...)

**Transformar
informação em
conhecimento**

Dados não estruturados

“Temos 5 bolas verdes com diâmetro de 43 mm cada”

Dados estruturados

```
>> # Computadores fazem certo o que você pediu errado
>>
>> quantidade <- 5
>> cor <- “verde”
>> diâmetro <- 43
>>
```

O fluxo de trabalho com dados



A **análise exploratória** é o primeiro contato com a informação disponível.

Ela é parte da fase de **perguntas**, mas também pode ser a **resposta por si só**.

Figura: Escola de dados



E a programação nessa história?

Programando em R

mpg-plot.R *

```
1 library(ggplot2)
2
3 ggplot(mpg, aes(x = displ, y = hwy)) +
4   geom_point(aes(colour = class))
5
```

Editor

1:1 (Top Level)

R Script

Console ~/Documents/r4ds/data-analysis/

```
> library(ggplot2)
> ggplot(mpg, aes(x = displ, y = hwy)) +
+   geom_point(aes(colour = class))
> |
```

Console

Environment

History

Import Dataset

Global Environment

Environment is empty

Files

Plots

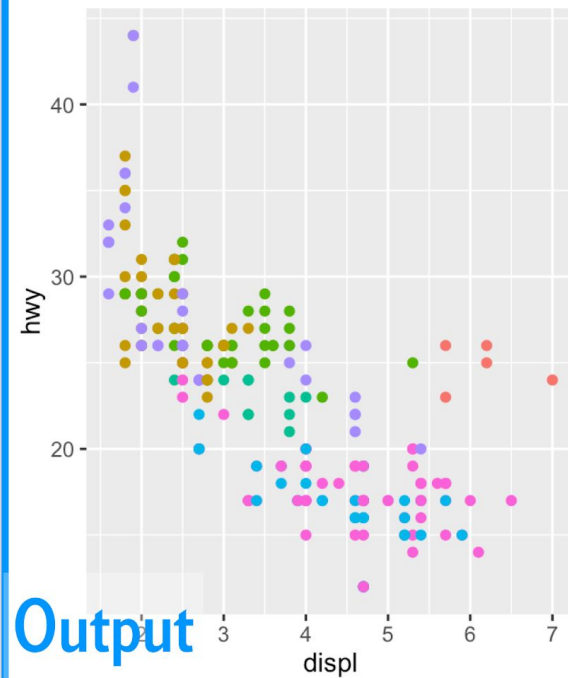
Packages

Help

Viewer

Zoom

Export



Output

Conceitos chave

Função

É como chamamos os comandos em R

```
Ex.:    sqrt(9)
      [1] 3
```

Objeto

Todo objeto possui estado (classes) e comportamentos (funções)

```
Ex.:    x <- 20
      x
      [1] 20
```

5 comandos básicos

#1 Operações matemáticas

$1 + 5$

$360 - 35$

$38 * 2$

$40 / 9$

2^3

#2 Atribuir valores a objetos

`salario <- 4500`

`gastos <- 3100`

`poupanca <- 4500 - 3100`

5 comandos básicos

#3 Requisitar informações

```
print(poupanca)
```

#4 Guardar mais de um dado

```
transacoes_bancarias<- c (salario, gastos, poupanca)
```

#5 Plotar informações num gráfico

```
plot(transacoes_bancarias)
```

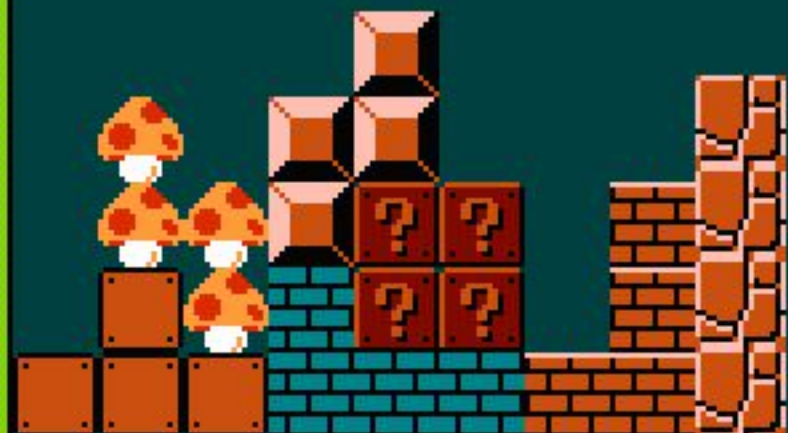
```
plot(transacoes_bancarias, type="l")
```

```
?plot
```

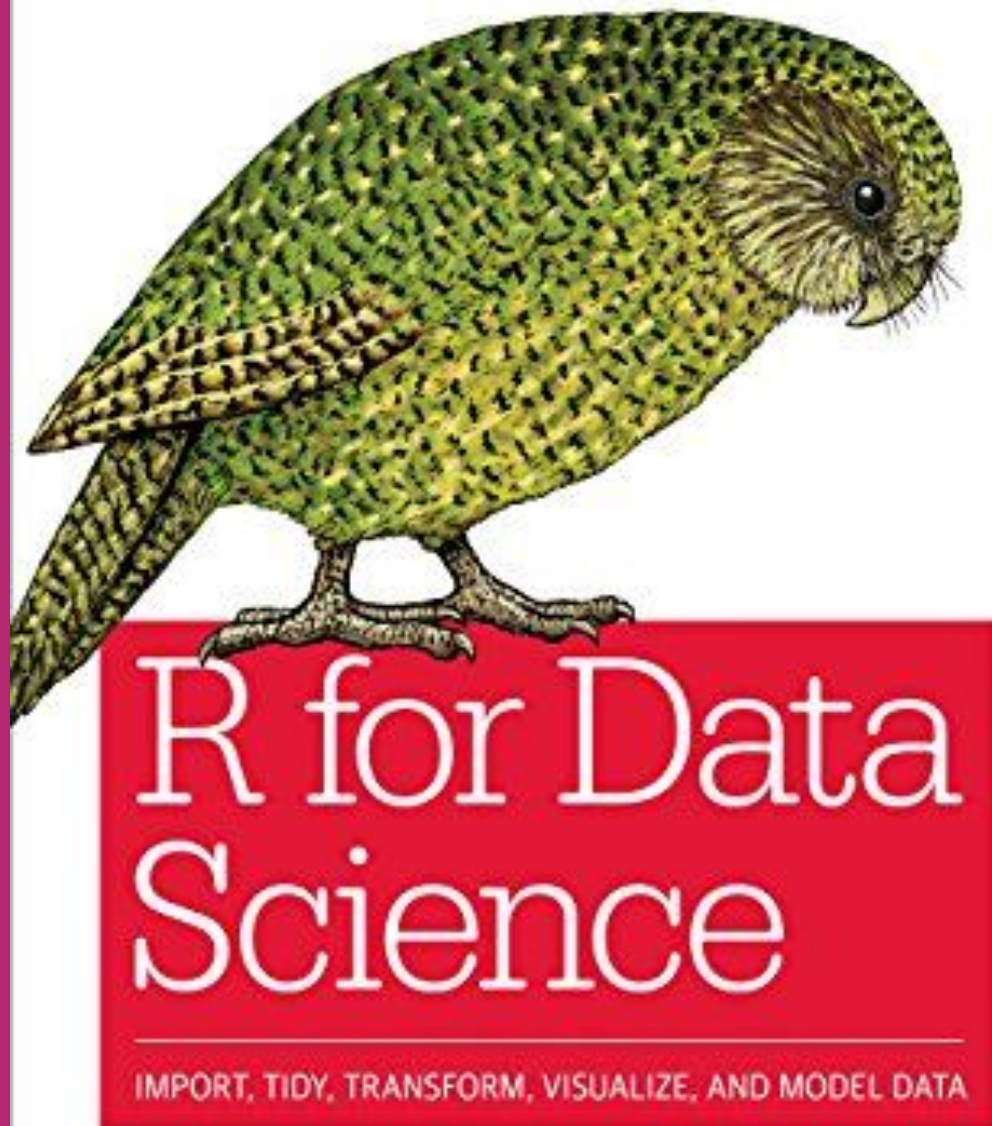
HIGH SCORE!



YOU WIN!



Materiais de estudo



Hadley Wickham &
Garrett Grolemund

Maria Letícia Machado

Linkedin: letiiciaam

Github: leticiamachado