CEFET/RJ

Manipulação de dados

Eduardo Ogasawara
eduardo.ogasawara@cefet-rj.br
https://eic.cefet-rj.br/~eogasawara

Criando uma tabela básica

```r
weight <- c(60, 72, 57, 90, 95, 72)
height <- c(1.75, 1.80, 1.65, 1.90, 1.74, 1.91)
subject <- c("A", "B", "C", "D", "E", "F")
d <- data.frame(weight=weight, height=height, subject=subject)
head(d)
```

```
##   weight height subject
## 1     60   1.75       A
## 2     72   1.80       B
## 3     57   1.65       C
## 4     90   1.90       D
## 5     95   1.74       E
## 6     72   1.91       F
```

Criando uma tabela auxiliar

```r
subject <- c("A", "B", "C", "D", "E", "F")
state <- c("RJ", "SP", "MG", "RJ", "SP", "MG")
ds <- data.frame(subject=subject, state=state)
head(d)
```

```
##   weight height subject
## 1     60   1.75       A
## 2     72   1.80       B
## 3     57   1.65       C
## 4     90   1.90       D
## 5     95   1.74       E
## 6     72   1.91       F
```

Integração de dados por junção

```r
dsm <- merge(d, ds, by.x="subject", by.y="subject")
head(dsm)
```

```
##   subject weight height state
## 1       A     60   1.75    RJ
## 2       B     72   1.80    SP
## 3       C     57   1.65    MG
## 4       D     90   1.90    RJ
## 5       E     95   1.74    SP
## 6       F     72   1.91    MG
```

Transformações

```
library(dplyr)
```

Conjunto de dados do pipeline são passados pelo operador |> para função seguinte no primeiro parâmetro da função de modo implícito

Pipeline: filtragem, projeção e ordenação

```
result <- dsm |>
    filter(height>1.7) |>
    select(subject, weight, height) |>
    arrange(height)
head(result)
```

```
##   subject weight height
## 1       E     95   1.74
## 2       A     60   1.75
## 3       B     72   1.80
## 4       D     90   1.90
## 5       F     72   1.91
```

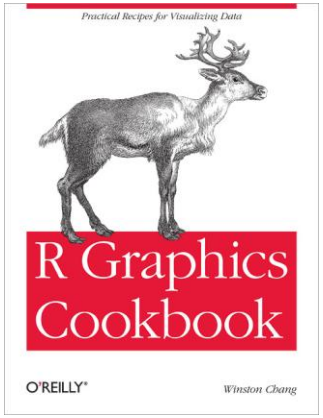
Pipeline: agregação de dados

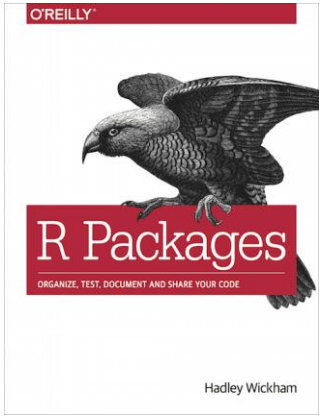
```
result <- dsm |>
    group_by(state) |>
    summarize(count = n(), height = mean(height))
head(result)
```

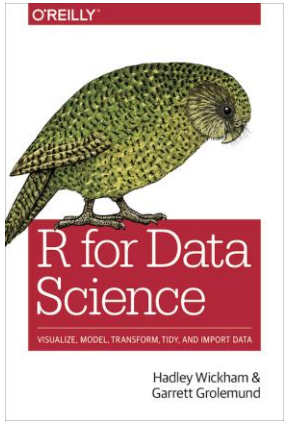
```
## # A tibble: 3 × 3
##   state count height
##   <chr> <int>  <dbl>
## 1 MG        2   1.78
## 2 RJ        2   1.82
## 3 SP        2   1.77
```

Referências

Material: https://eic.cefet-rj.br/~eogasawara/tutorial-r







Hands-on Programming with R: https://rstudio-education.github.io/hopr/index.html
R Graphics Cookbook: https://r-graphics.org
R Packages: https://r-pkgs.org/index.html
R for Data Science: https://r4ds.had.co.nz

https://rstudio-education.github.io/hopr/basics.html

[1] G. Grolemund, 2014, Hands-On Programming with R: Write Your Own Functions and Simulations. O'Reilly Media, Inc.