Manipulação e Apresentação de Dados

Guilherme Bovi Ambrosano

Dia 2 - 08/08/2024

Sumário

tidyverse

tidyverse

O `tidyverse` é um conjunto de pacotes do R, incluindo:

- `forcats`
- `stringr`
- `lubridate`
- `tibble`
- `readr`
- `readxl`
- `dplyr`
- `tidyr`
- `purrr`
- `ggplot2`

tibble

tibble

```
library(tibble)
load("../dados/Dados.RData")

dados
```

```
##                    val
## Trat1_Rep1 15.78291
## Trat1_Rep2 16.14339
## Trat1_Rep3 15.74257
## Trat1_Rep4 14.94049
## Trat1_Rep5 15.02677
## Trat2_Rep1 16.72542
## Trat2_Rep2 17.44605
## Trat2_Rep3 16.42591
## Trat2_Rep4 17.72766
## Trat2_Rep5 16.87040
## Trat3_Rep1 18.33888
## Trat3_Rep2 18.47856
## Trat3_Rep3 18.67104
## Trat3_Rep4 17.86359
```

tibble

```
as_tibble(dados)
```

```
## # A tibble: 20 x 1
##       val
##     <dbl>
##  1  15.8
##  2  16.1
##  3  15.7
##  4  14.9
##  5  15.0
##  6  16.7
##  7  17.4
##  8  16.4
##  9  17.7
## 10  16.9
## 11  18.3
## 12  18.5
## 13  18.7
## 14  17.9
## 15  17.8
```

tibble

```
# rownames_to_column, column_to_rownames
rownames_to_column(dados, "id")
```

```
##                id      val
## 1   Trat1_Rep1 15.78291
## 2   Trat1_Rep2 16.14339
## 3   Trat1_Rep3 15.74257
## 4   Trat1_Rep4 14.94049
## 5   Trat1_Rep5 15.02677
## 6   Trat2_Rep1 16.72542
## 7   Trat2_Rep2 17.44605
## 8   Trat2_Rep3 16.42591
## 9   Trat2_Rep4 17.72766
## 10 Trat2_Rep5 16.87040
## 11 Trat3_Rep1 18.33888
## 12 Trat3_Rep2 18.47856
## 13 Trat3_Rep3 18.67104
## 14 Trat3_Rep4 17.86359
## 15 Trat3_Rep5 17.75907
## 16 Trat4_Rep1 21.25240
```

tibble

```
tibble(numerico=c(1,2,3),
       caractere=c("a","b","c"),
       lógico=c(T,T,F),
       fator=factor(c(1,1,2)))
```

```
## # A tibble: 3 x 4
##   numerico caractere lógico fator
##      <dbl> <chr>     <lgl>  <fct>
## 1        1 a         TRUE   1
## 2        2 b         TRUE   1
## 3        3 c         FALSE  2
```

tibble

```r
# enframe, deframe
(palavras <- sapply(c("p", "g", "m", "sap"), paste0, "ato"))
```

```
##         p         g         m       sap
##   "pato"    "gato"    "mato" "sapato"
```

```r
enframe(palavras)
```

```
## # A tibble: 4 x 2
##   name  value
##   <chr> <chr>
## 1 p     pato
## 2 g     gato
## 3 m     mato
## 4 sap   sapato
```

Pipes (%>%)

Pipes

```r
funcao <- function(arg1, arg2, arg3) {
  return(c(arg1, arg2, arg3))
}

funcao(1,2,3)
```

```
## [1] 1 2 3
```

Pipes

```
funcao(1,2,3)
```

```
## [1] 1 2 3
```

```
1 %>%
  funcao(2,3)
```

```
## [1] 1 2 3
```

```
1 %>%
  funcao(2,3) %>%
  funcao(4,5)
```

```
## [1] 1 2 3 4 5
```

```
c(c(1, 2, 3), 4, 5)
```

```
## [1] 1 2 3 4 5
```

Pipes

```
funcao(1,2,3)
```

```
## [1] 1 2 3
```

```
1 %>%
  funcao(2,3)
```

```
## [1] 1 2 3
```

```
1 %>%
  funcao(2,3) %>%
  funcao(4,5)
```

```
## [1] 1 2 3 4 5
```

```
c(c(1, 2, 3), 4, 5)
```

```
## [1] 1 2 3 4 5
```

Pipes

```
funcao(1,2,3)
```

```
## [1] 1 2 3
```

```
1 %>%
  funcao(2,3)
```

```
## [1] 1 2 3
```

```
1 %>%
  funcao(2,3) %>%
  funcao(4,5)
```

```
## [1] 1 2 3 4 5
```

```
c(c(1, 2, 3), 4, 5)
```

```
## [1] 1 2 3 4 5
```

Pipes

```
funcao(1,2,3)
```

```
## [1] 1 2 3
```

```
1 %>%
  funcao(2,3)
```

```
## [1] 1 2 3
```

```
1 %>%
  funcao(2,3) %>%
  funcao(4,5)
```

```
## [1] 1 2 3 4 5
```

```
c(c(1, 2, 3), 4, 5)
```

```
## [1] 1 2 3 4 5
```

Pipes

```r
adicionar_coluna <- function(dados, coluna) {
  dados$coluna2 <- coluna
  return(dados)
}

tibble(coluna1=c(1,2,3)) %>%
  adicionar_coluna(c(4,5,6))
```

```
## # A tibble: 3 x 2
##   coluna1 coluna2
##     <dbl>   <dbl>
## 1       1       4
## 2       2       5
## 3       3       6
```

Pipes

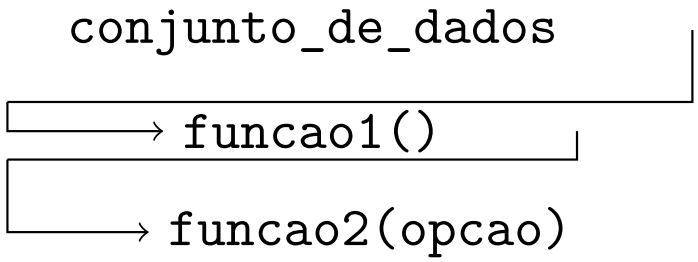
```
adicionar_coluna <- function(dados, ...) {
  colunas <- list(...)
  sapply(seq(1, length(colunas)),
         function(i) {
           dados[,names(colunas)[i]] <<- colunas[[i]]
         })
  return(dados)
}

tibble(coluna1=c(1,2,3)) %>%
  adicionar_coluna(coluna2=c(4,5,6),
                   coluna3=c(7,8,9))
```

```
## # A tibble: 3 x 3
##   coluna1 coluna2 coluna3
##     <dbl>   <dbl>   <dbl>
## 1       1       4       7
## 2       2       5       8
## 3       3       6       9
```

Pipes

```
conjunto_de_dados %>%
    funcao1() %>%
   funcao2(opcao)
```

Pipes



Pipes

`conjunto_de_dados` $\longrightarrow$ `funcao1()` $\longrightarrow$ `funcao2(opcao)`

Pipes

`funcao1(conjunto_de_dados)` $\longrightarrow$ `funcao2(opcao)`

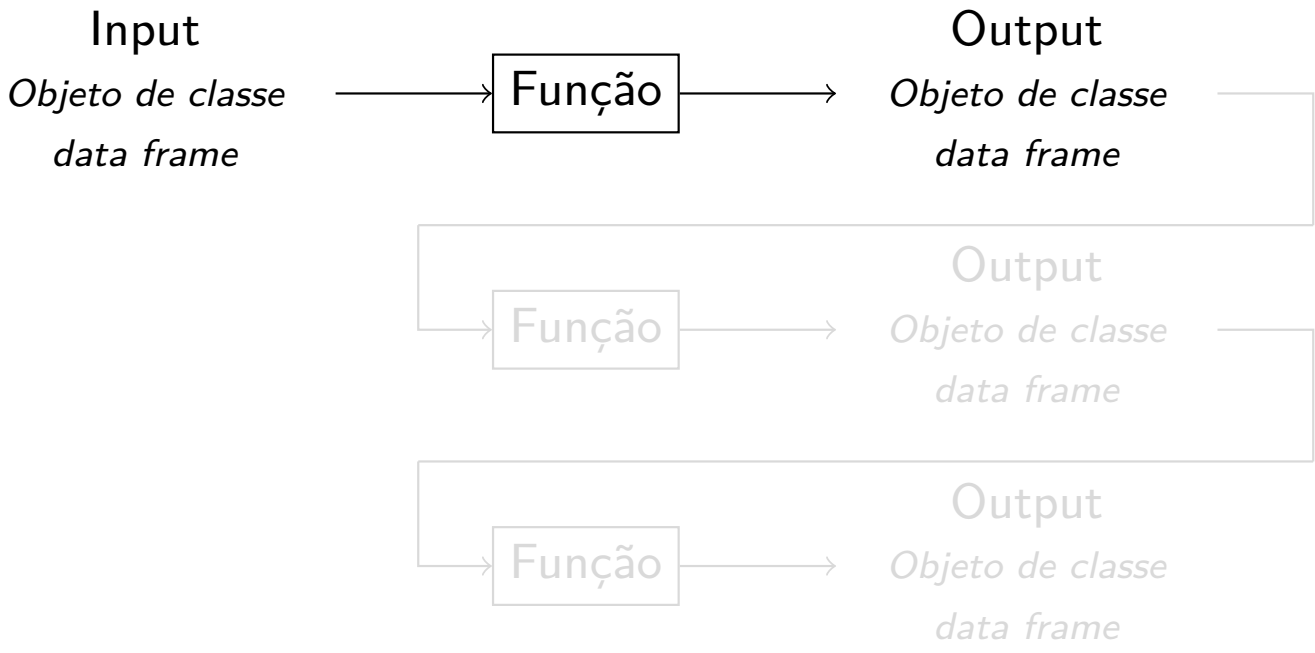
Pipes

```
funcao2(funcao1(conjunto_de_dados), opcao)
```

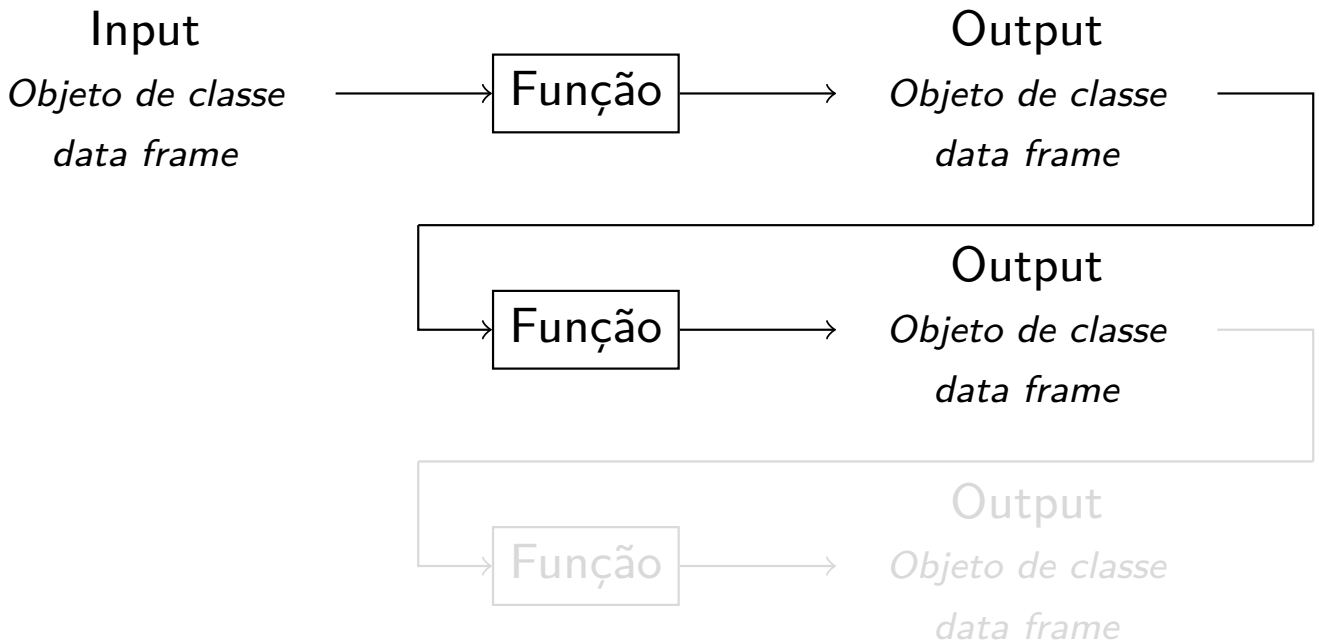
Verbos do dplyr

select, filter, mutate, group_by e summarise

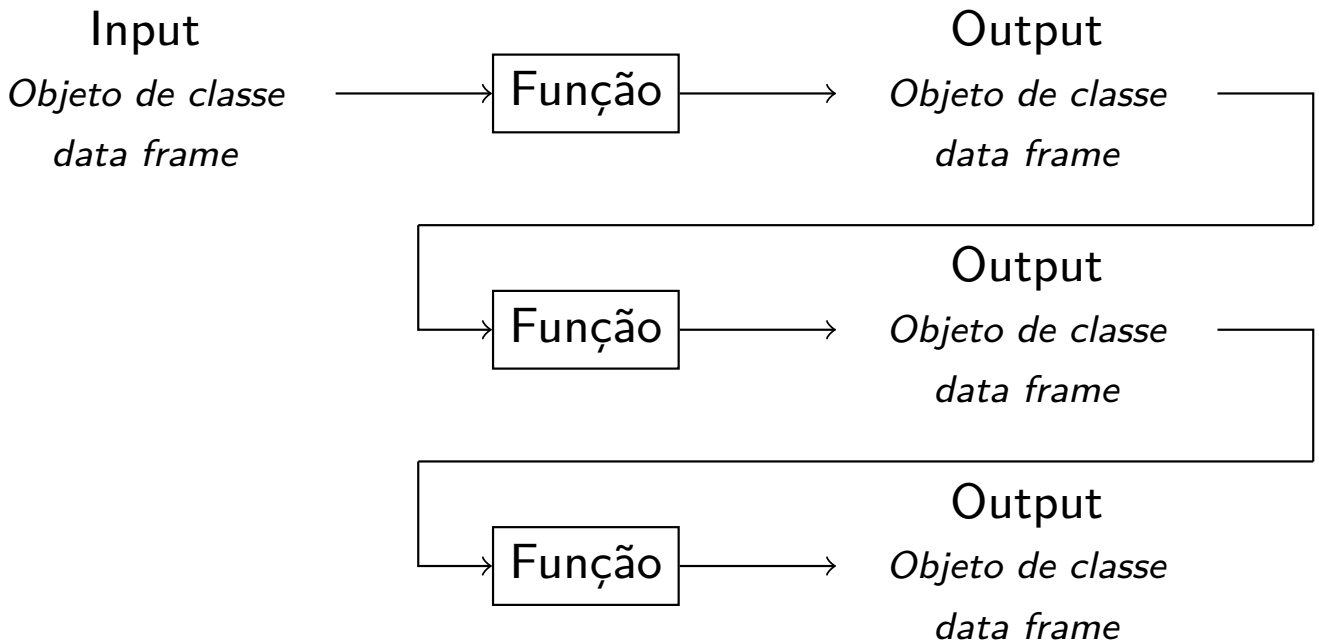
Verbos do dplyr



Verbos do dplyr



Verbos do dplyr



dplyr

- Verbo `select`: selecionar colunas
- Verbo `filter`: filtrar linhas

Verbo select

Verbo `select`: selecionar colunas

```
mtcars
```

```
##                    mpg cyl  disp  hp drat    wt  qsec vs am
## Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1
## Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1
## Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1
## Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0
## Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0
## Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0
## Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0
## Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0
## Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0
## Merc 280C         17.8   6 167.6 123 3.92 3.440 18.90  1  0
## Merc 450SE        16.4   8 275.8 180 3.07 4.070 17.40  0  0
## Merc 450SL        17.3   8 275.8 180 3.07 3.730 17.60  0  0
## Merc 450SLC       15.2   8 275.8 180 3.07 3.780 18.00  0  0
## Cadillac Fleetwood 10.4  8 472.0 205 2.93 5.250 17.98  0  0
```

Verbo select

Verbo `select`: selecionar colunas

```
library(dplyr)
mtcars %>%
  select(mpg, cyl, disp)
```

```
##                    mpg cyl  disp
## Mazda RX4         21.0   6 160.0
## Mazda RX4 Wag     21.0   6 160.0
## Datsun 710        22.8   4 108.0
## Hornet 4 Drive    21.4   6 258.0
## Hornet Sportabout 18.7   8 360.0
## Valiant           18.1   6 225.0
## Duster 360        14.3   8 360.0
## Merc 240D         24.4   4 146.7
## Merc 230          22.8   4 140.8
## Merc 280          19.2   6 167.6
## Merc 280C         17.8   6 167.6
## Merc 450SE        16.4   8 275.8
## Merc 450SL        17.3   8 275.8
```

Verbo select

Verbo `select`: selecionar colunas

```
mtcars %>%
  select(-mpg, -cyl, -disp, -gear,
         -am, -vs, -wt, -carb)
```

```
##                    hp drat  qsec
## Mazda RX4         110 3.90 16.46
## Mazda RX4 Wag     110 3.90 17.02
## Datsun 710         93 3.85 18.61
## Hornet 4 Drive    110 3.08 19.44
## Hornet Sportabout 175 3.15 17.02
## Valiant           105 2.76 20.22
## Duster 360        245 3.21 15.84
## Merc 240D          62 3.69 20.00
## Merc 230           95 3.92 22.90
## Merc 280          123 3.92 18.30
## Merc 280C         123 3.92 18.90
## Merc 450SE        180 3.07 17.40
## Merc 450SL        180 3.07 17.60
```

Verbo select

Verbo `select`: selecionar colunas

```
names(iris)
```

```
## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
```

```
iris %>%
  select(starts_with("Petal")) %>% head()
```

```
##   Petal.Length Petal.Width
## 1          1.4         0.2
## 2          1.4         0.2
## 3          1.3         0.2
## 4          1.5         0.2
## 5          1.4         0.2
## 6          1.7         0.4
```

```
iris %>%
  select(ends_with("Width")) %>% head()
```

```
##   Sepal.Width Petal.Width
## 1         3.5         0.2
## 2         3.0         0.2
## 3         3.2         0.2
## 4         3.1         0.2
## 5         3.6         0.2
```

Verbo select

Verbo `select`: selecionar colunas

```
names(iris)
```

```
## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
```

```
iris %>%
  select(starts_with("Petal")) %>% head()
```

```
##   Petal.Length Petal.Width
## 1          1.4         0.2
## 2          1.4         0.2
## 3          1.3         0.2
## 4          1.5         0.2
## 5          1.4         0.2
## 6          1.7         0.4
```

```
iris %>%
  select(ends_with("Width")) %>% head()
```

```
##   Sepal.Width Petal.Width
## 1         3.5         0.2
## 2         3.0         0.2
## 3         3.2         0.2
## 4         3.1         0.2
## 5         3.6         0.2
```

Verbo select

Verbo select: selecionar colunas

```
names(iris)
```

```
## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
```

```
iris %>%
  select(starts_with("Petal")) %>% head()
```

```
##   Petal.Length Petal.Width
## 1          1.4         0.2
## 2          1.4         0.2
## 3          1.3         0.2
## 4          1.5         0.2
## 5          1.4         0.2
## 6          1.7         0.4
```

```
iris %>%
  select(ends_with("Width")) %>% head()
```

```
##   Sepal.Width Petal.Width
## 1         3.5         0.2
## 2         3.0         0.2
## 3         3.2         0.2
## 4         3.1         0.2
## 5         3.6         0.2
```

Verbo select

Verbo `select`: selecionar colunas

```
iris %>%
  select(matches("[SP]e[tp]al\\.[WL][ie].g?th"))
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1            5.1         3.5          1.4         0.2
## 2            4.9         3.0          1.4         0.2
## 3            4.7         3.2          1.3         0.2
## 4            4.6         3.1          1.5         0.2
## 5            5.0         3.6          1.4         0.2
## 6            5.4         3.9          1.7         0.4
## 7            4.6         3.4          1.4         0.3
## 8            5.0         3.4          1.5         0.2
## 9            4.4         2.9          1.4         0.2
## 10           4.9         3.1          1.5         0.1
## 11           5.4         3.7          1.5         0.2
## 12           4.8         3.4          1.6         0.2
## 13           4.8         3.0          1.4         0.1
## 14           4.3         3.0          1.1         0.1
## 15           5.8         4.0          1.2         0.2
## 16           5.7         4.4          1.5         0.4
```

Verbo filter

Verbo `filter`: filtrar linhas

```
mtcars %>%
  filter(hp > 250)
```

```
##                  mpg cyl disp  hp drat   wt qsec vs am gear ca
## Ford Pantera L 15.8   8  351 264 4.22 3.17 14.5  0  1    5
## Maserati Bora  15.0   8  301 335 3.54 3.57 14.6  0  1    5
```

Verbo filter

```
mtcars %>%
  filter(hp > 150, hp < 200)
```

```
##                    mpg cyl  disp  hp drat    wt  qsec vs am g
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0
## Merc 450SE        16.4   8 275.8 180 3.07 4.070 17.40  0  0
## Merc 450SL        17.3   8 275.8 180 3.07 3.730 17.60  0  0
## Merc 450SLC       15.2   8 275.8 180 3.07 3.780 18.00  0  0
## Pontiac Firebird  19.2   8 400.0 175 3.08 3.845 17.05  0  0
## Ferrari Dino      19.7   6 145.0 175 3.62 2.770 15.50  0  1
```

Verbo filter

```
mtcars %>%
  filter(hp > 150 & hp < 200)
```

```
##                    mpg cyl  disp  hp drat    wt  qsec vs am g
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0
## Merc 450SE        16.4   8 275.8 180 3.07 4.070 17.40  0  0
## Merc 450SL        17.3   8 275.8 180 3.07 3.730 17.60  0  0
## Merc 450SLC       15.2   8 275.8 180 3.07 3.780 18.00  0  0
## Pontiac Firebird  19.2   8 400.0 175 3.08 3.845 17.05  0  0
## Ferrari Dino      19.7   6 145.0 175 3.62 2.770 15.50  0  1
```

Verbo filter

```
mtcars %>%
  filter(hp > 300 | hp < 60)
```

```
##                  mpg cyl  disp  hp drat    wt  qsec vs am gear
## Honda Civic     30.4   4  75.7  52 4.93 1.615 18.52  1  1    4
## Maserati Bora 15.0   8 301.0 335 3.54 3.570 14.60  0  1    5
```

Verbo filter

```
library(readxl)
nomes <- names(read_xls("../dados/diario2023.xls",
                        skip=7, n_max = 0))
diario2023 <- read_xls("../dados/diario2023.xls", skip=10,
                       col_names=nomes) %>%
  filter_all(any_vars(!is.na(.)))

head(diario2023)
```

```
## # A tibble: 6 x 24
##   TIMESTAMP           RECORD BattV_Avg Tar_AVG UR_inst Vvento_ms_AVG
##   <dttm>               <dbl>     <dbl>   <dbl>   <dbl>          <dbl>
## 1 2022-12-31 00:00:00   8102      12.5    20.3    98             0.1
## 2 2022-12-31 00:15:00   8103      12.5    20.2    98.3           0
## 3 2022-12-31 00:30:00   8104      12.5    20.1    98.5           0
## 4 2022-12-31 00:45:00   8105      12.5    20.1    98.3           0
## 5 2022-12-31 01:00:00   8106      12.5    20.1    98.4           0
## 6 2022-12-31 01:15:00   8107      12.5    20.1    98.3           0
## # i 17 more variables: Qg_AVG <dbl>, PAR_AVG <dbl>, Rn_Avg <dbl>,
## #   Chuva_mm <dbl>, Patm_kPa_AVG <dbl>, rQg_AVG <dbl>, Qatm_AVG <dbl
## #   Qsup_AVG <dbl>, Boc_AVG <dbl>, Bol_AVG <dbl>, Albedo_Avg <dbl>,
```

dplyr

- Verbo `select`: selecionar colunas
- Verbo `filter`: filtrar linhas
- Verbo `mutate`: criar novas colunas

Verbo mutate

Verbo `mutate`: criar novas colunas

```
tibble(aleatorio1=rnorm(10))
```

```
## # A tibble: 10 x 1
##    aleatorio1
##         <dbl>
##  1      1.64
##  2      0.547
##  3      0.837
##  4      0.929
##  5     -0.382
##  6      0.193
##  7      1.02
##  8     -1.03
##  9     -1.69
## 10      0.172
```

Verbo mutate

Verbo `mutate`: criar novas colunas

```
tibble(aleatorio1=rnorm(10)) %>%
  mutate(aleatorio2 = rnorm(10))
```

```
## # A tibble: 10 x 2
##    aleatorio1 aleatorio2
##         <dbl>      <dbl>
##  1   -0.481      -0.430
##  2    0.0294      0.579
##  3   -0.677       0.264
##  4    0.614       2.01
##  5    0.473      -0.605
##  6    2.04        0.666
##  7   -0.491      -1.08
##  8    0.237       0.345
##  9    0.155       0.0111
## 10    0.00711     1.34
```

Verbo mutate

Verbo `mutate`: criar novas colunas

```
tibble(aleatorio1=rnorm(10)) %>%
  mutate(aleatorio2 = rnorm(10)) %>%
  mutate(soma = aleatorio1 + aleatorio2)
```

```
## # A tibble: 10 x 3
##    aleatorio1 aleatorio2    soma
##         <dbl>      <dbl>   <dbl>
##  1   -1.35         0.510  -0.841
##  2    0.612        0.195   0.808
##  3   -0.329        0.0176 -0.311
##  4    0.0106      -0.0668 -0.0562
##  5   -0.513       -0.213  -0.726
##  6    0.00633      1.64    1.65
##  7   -0.115        0.195   0.0801
##  8    0.753        0.381   1.13
##  9    0.425        0.290   0.715
## 10   -0.433       -1.54   -1.97
```

Verbo mutate

```r
tamanhos <- tibble(tam = c(1.5, 1.9, 1.6, 1.8, 2.0, 1.7))
```

```r
#tamanhos %>%
#  mutate(classe =
#           ifelse(tam < median(tam),
#                  "Pequeno", "Grande"))

tamanhos %>%
  mutate(classe = case_when(
    tam < median(tam) ~ "Pequeno",
    T ~ "Grande"))
```

```
## # A tibble: 6 x 2
##     tam classe
##   <dbl> <chr>
## 1   1.5 Pequeno
## 2   1.9 Grande
## 3   1.6 Pequeno
## 4   1.8 Grande
## 5   2   Grande
## 6   1.7 Pequeno
```

Verbo mutate

```
tamanhos <- tibble(tam = c(1.5, 1.9, 1.6, 1.8, 2.0, 1.7))
```

```
#tamanhos %>%
#  mutate(classe =
#           ifelse(tam < median(tam),
#                  "Pequeno", "Grande"))

tamanhos %>%
  mutate(classe = case_when(
    tam < median(tam) ~ "Pequeno",
    T ~ "Grande"))
```

```
## # A tibble: 6 x 2
##     tam classe
##   <dbl> <chr>
## 1   1.5 Pequeno
## 2   1.9 Grande
## 3   1.6 Pequeno
## 4   1.8 Grande
## 5   2   Grande
## 6   1.7 Pequeno
```

Verbo mutate

```
#tamanhos %>%
#  mutate(classe = ifelse(
#    tam < quantile(tam, .25), "Muito pequeno",
#    ifelse(
#      tam < median(tam), "Pequeno", "Grande")))

tamanhos %>%
  mutate(classe = case_when(
    tam < quantile(tam,.25) ~ "Muito pequeno",
    tam < median(tam) ~ "Pequeno",
    T ~ "Grande"))
```

```
## # A tibble: 6 x 2
##     tam classe
##   <dbl> <chr>
## 1   1.5 Muito pequeno
## 2   1.9 Grande
## 3   1.6 Muito pequeno
## 4   1.8 Grande
## 5   2   Grande
## 6   1.7 Pequeno
```

dplyr

- Verbo `select`: selecionar colunas
- Verbo `filter`: filtrar linhas
- Verbo `mutate`: criar novas colunas
- Verbos `group_by` e `summarise`: agrupar e resumir

Verbos group_by e summarise

```
mtcars %>%
  group_by(cyl) %>%
  summarise(media_hp = mean(hp))
```

```
## # A tibble: 3 x 2
##     cyl media_hp
##   <dbl>    <dbl>
## 1     4     82.6
## 2     6    122.
## 3     8    209.
```

Verbos group_by e summarise

```
read_xlsx("../dados/Planilha.xlsx")
```

```
## # A tibble: 30 x 4
##    Tamanho Cor    Repetição Medida
##    <chr>   <chr>      <dbl>  <dbl>
##  1 Grande  Claro          1    1.8
##  2 Grande  Claro          2   15.4
##  3 Grande  Claro          3    5.2
##  4 Grande  Claro          4    2.2
##  5 Grande  Claro          5   12.4
##  6 Grande  Escuro         1   20
##  7 Grande  Escuro         2    5.4
##  8 Grande  Escuro         3   NA
##  9 Grande  Escuro         4    5.8
## 10 Grande  Escuro         5    5.4
## # i 20 more rows
```

Verbos group_by e summarise

```
(medidas <- read_xlsx("../dados/Planilha.xlsx") %>%
  mutate(Tamanho = factor(Tamanho,
                          levels=c("Pequeno", "Médio",
                                   "Grande")),
         Cor = factor(Cor, levels=c("Claro", "Escuro")),
         Repetição = factor(Repetição)))
```

```
## # A tibble: 30 x 4
##    Tamanho Cor    Repetição Medida
##    <fct>   <fct>  <fct>       <dbl>
##  1 Grande  Claro  1             1.8
##  2 Grande  Claro  2            15.4
##  3 Grande  Claro  3             5.2
##  4 Grande  Claro  4             2.2
##  5 Grande  Claro  5            12.4
##  6 Grande  Escuro 1            20
##  7 Grande  Escuro 2             5.4
##  8 Grande  Escuro 3            NA
##  9 Grande  Escuro 4             5.8
## 10 Grande  Escuro 5             5.4
## # i 20 more rows
```

Verbos group_by e summarise

```
medidas %>%
  group_by(Cor) %>%
  summarise(média=mean(Medida))
```

```
## # A tibble: 2 x 2
##   Cor    média
##   <fct>  <dbl>
## 1 Claro     NA
## 2 Escuro    NA
```

Verbos group_by e summarise

```
medidas %>%
  group_by(Cor) %>%
  summarise(média=mean(Medida, na.rm=T))
```

```
## # A tibble: 2 x 2
##   Cor    média
##   <fct>  <dbl>
## 1 Claro   13.4
## 2 Escuro  19.9
```

Verbos group_by e summarise

```
medidas %>%
  group_by(Cor) %>%
  summarise_at(vars(Medida), list(média=mean), na.rm=T)
```

```
## # A tibble: 2 x 2
##   Cor    média
##   <fct>  <dbl>
## 1 Claro   13.4
## 2 Escuro  19.9
```

Verbos group_by e summarise

```
medidas %>%
  group_by(Cor) %>%
  summarise_at(vars(Medida), list(média=mean,
                                  desvio=sd), na.rm=T)
```

```
## # A tibble: 2 x 3
##   Cor    média desvio
##   <fct>  <dbl>  <dbl>
## 1 Claro   13.4   8.06
## 2 Escuro  19.9  16.7
```

Verbos group_by e summarise

```
medidas %>%
  group_by(Cor, Tamanho) %>%
  summarise_at(vars(Medida), list(média=mean,
                                  desvio=sd), na.rm=T)
```

```
## # A tibble: 6 x 4
## # Groups:   Cor [2]
##   Cor    Tamanho média desvio
##   <fct>  <fct>   <dbl>  <dbl>
## 1 Claro  Pequeno 15.1    2.26
## 2 Claro  Médio   19.3    9.64
## 3 Claro  Grande   7.4    6.17
## 4 Escuro Pequeno 40.1   10.6
## 5 Escuro Médio   10.4    6.03
## 6 Escuro Grande   9.15   7.24
```

Verbos group_by e summarise

```
medidas %>%
  group_by(Cor, Tamanho) %>%
  summarise_at(vars(Medida), list(média=mean,
                                  desvio=sd), na.rm=T) %>%
  summarise_at(vars(média), max)
```

```
## # A tibble: 2 x 2
##   Cor    média
##   <fct>  <dbl>
## 1 Claro   19.3
## 2 Escuro  40.1
```

Verbos group_by e summarise

```
medidas %>%
  group_by(Cor, Tamanho) %>%
  summarise_at(vars(Medida), list(média=mean,
                                  desvio=sd), na.rm=T) %>%
  summarise_at(vars(média), max) %>%
  summarise_at(vars(média), sum)
```

```
## # A tibble: 1 x 1
##   média
##   <dbl>
## 1  59.5
```

Verbos group_by e summarise

```
medidas %>%
  group_by(Cor, Tamanho) %>%
  summarise_at(vars(Medida), list(média=mean,
                                  desvio=sd), na.rm=T) %>%
  ungroup()
```

```
## # A tibble: 6 x 4
##   Cor    Tamanho média desvio
##   <fct>  <fct>   <dbl>  <dbl>
## 1 Claro  Pequeno 15.1    2.26
## 2 Claro  Médio   19.3    9.64
## 3 Claro  Grande   7.4    6.17
## 4 Escuro Pequeno 40.1   10.6
## 5 Escuro Médio   10.4    6.03
## 6 Escuro Grande   9.15   7.24
```

dplyr

- Verbo `select`: selecionar colunas
- Verbo `filter`: filtrar linhas
- Verbo `mutate`: criar novas colunas
- Verbos `group_by` e `summarise`: agrupar e resumir
- Verbo `arrange`: ordenar uma variável de maneira crescente ou decrescent (`desc`)
- Verbo `slice`: selecionar linhas pelo número

Verbos arrange e slice

```
mtcars %>%
  arrange(hp)
```

```
##                    mpg cyl  disp  hp drat    wt  qsec vs am
## Honda Civic       30.4   4  75.7  52 4.93 1.615 18.52  1  1
## Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0
## Toyota Corolla    33.9   4  71.1  65 4.22 1.835 19.90  1  1
## Fiat 128          32.4   4  78.7  66 4.08 2.200 19.47  1  1
## Fiat X1-9         27.3   4  79.0  66 4.08 1.935 18.90  1  1
## Porsche 914-2     26.0   4 120.3  91 4.43 2.140 16.70  0  1
## Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1
## Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0
## Toyota Corona     21.5   4 120.1  97 3.70 2.465 20.01  1  0
## Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0
## Volvo 142E        21.4   4 121.0 109 4.11 2.780 18.60  1  1
## Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1
## Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1
## Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0
## Lotus Europa      30.4   4  95.1 113 3.77 1.513 16.90  1  1
## Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0
```

Verbos arrange e slice

```
mtcars %>%
  arrange(desc(hp))
```

```
##                    mpg cyl  disp  hp drat    wt  qsec vs am
## Maserati Bora     15.0   8 301.0 335 3.54 3.570 14.60  0  1
## Ford Pantera L    15.8   8 351.0 264 4.22 3.170 14.50  0  1
## Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0
## Camaro Z28        13.3   8 350.0 245 3.73 3.840 15.41  0  0
## Chrysler Imperial 14.7   8 440.0 230 3.23 5.345 17.42  0  0
## Lincoln Continental 10.4 8 460.0 215 3.00 5.424 17.82  0  0
## Cadillac Fleetwood 10.4  8 472.0 205 2.93 5.250 17.98  0  0
## Merc 450SE        16.4   8 275.8 180 3.07 4.070 17.40  0  0
## Merc 450SL        17.3   8 275.8 180 3.07 3.730 17.60  0  0
## Merc 450SLC       15.2   8 275.8 180 3.07 3.780 18.00  0  0
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0
## Pontiac Firebird  19.2   8 400.0 175 3.08 3.845 17.05  0  0
## Ferrari Dino      19.7   6 145.0 175 3.62 2.770 15.50  0  1
## Dodge Challenger  15.5   8 318.0 150 2.76 3.520 16.87  0  0
## AMC Javelin       15.2   8 304.0 150 3.15 3.435 17.30  0  0
```

Verbos arrange e slice

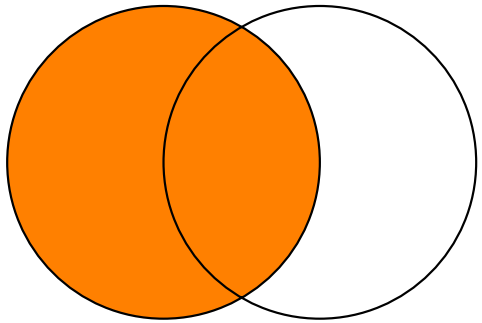
```
mtcars %>%
  arrange(desc(hp)) %>%
  slice(1:5)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am ge
## Maserati Bora     15.0   8  301 335 3.54 3.570 14.60  0  1
## Ford Pantera L    15.8   8  351 264 4.22 3.170 14.50  0  1
## Duster 360        14.3   8  360 245 3.21 3.570 15.84  0  0
## Camaro Z28        13.3   8  350 245 3.73 3.840 15.41  0  0
## Chrysler Imperial 14.7   8  440 230 3.23 5.345 17.42  0  0
```

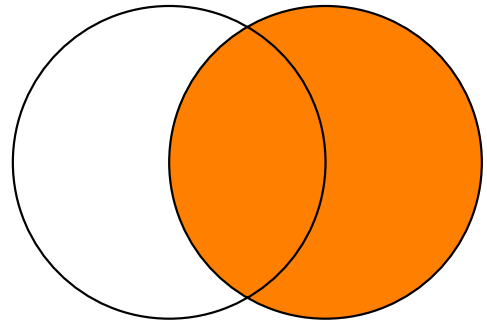
Verbos arrange e slice

```
mtcars %>%
  rownames_to_column("carro") %>%
  arrange(desc(hp)) %>%
  slice(1:5) %>%
  select(carro, hp) %>%
  deframe()
```

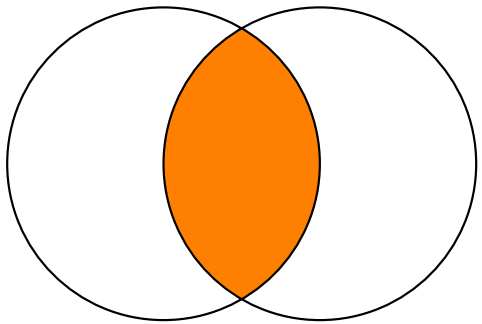
```
##      Maserati Bora    Ford Pantera L         Duster 360
##                335                264                245
## Chrysler Imperial
##                230
```



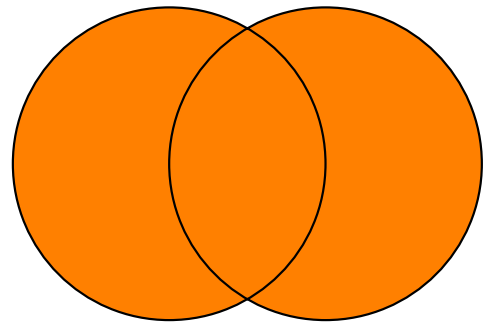
Left join



Right join



Inner join



Full join

União de conjuntos de dados

```r
usuarios <- read_xlsx("../dados/Transacoes.xlsx", "usuarios")
transacoes <- read_xlsx("../dados/Transacoes.xlsx", "transacoes")

usuarios
```

```
## # A tibble: 5 x 4
##   usuario nome     sobrenome dataNasc
##   <chr>   <chr>    <chr>     <chr>
## 1 U0001   Arnaldo  Moreno    08/02/2007
## 2 U0002   Isa      Soares    05/04/1971
## 3 U0003   Teodósio Gomes     10/07/1995
## 4 U0004   Roberto  Jorge     30/08/1987
## 5 U0005   Corina   Cruz      08/03/1990
```

```r
transacoes
```

```
## # A tibble: 9 x 3
##   usuario quantidade data
##   <chr>        <dbl> <dttm>
## 1 U0002         3372  2022-07-07 00:00:00
## 2 U0004        25761. 2022-04-06 00:00:00
## 3 U0005      1011250. 2021-01-10 00:00:00
## 4 U0006        27697. 2021-01-04 00:00:00
## 5 U0002       613021. 2022-06-27 00:00:00
## 6 U0001       170038. 2022-09-25 00:00:00
## 7 U0004       290199. 2022-05-03 00:00:00
## 8 U0002      2031465. 2022-06-10 00:00:00
## 9 U0002       989311. 2022-10-14 00:00:00
```

União de conjuntos de dados

```
transacoes %>%
  left_join(usuarios,by="usuario")
```

```
## # A tibble: 9 x 6
##   usuario quantidade data                nome    sobrenome dataNasc
##   <chr>        <dbl> <dttm>              <chr>   <chr>     <chr>
## 1 U0002         3372 2022-07-07 00:00:00 Isa     Soares    05/04/197
## 2 U0004        25761. 2022-04-06 00:00:00 Roberto Jorge     30/08/198
## 3 U0005      1011250. 2021-01-10 00:00:00 Corina  Cruz      08/03/199
## 4 U0006        27697. 2021-01-04 00:00:00 <NA>    <NA>      <NA>
## 5 U0002       613021. 2022-06-27 00:00:00 Isa     Soares    05/04/197
## 6 U0001       170038. 2022-09-25 00:00:00 Arnaldo Moreno    08/02/200
## 7 U0004       290199. 2022-05-03 00:00:00 Roberto Jorge     30/08/198
## 8 U0002      2031465. 2022-06-10 00:00:00 Isa     Soares    05/04/197
## 9 U0002       989311. 2022-10-14 00:00:00 Isa     Soares    05/04/197
```

União de conjuntos de dados

```
transacoes %>%
  right_join(usuarios,by="usuario")
```

```
## # A tibble: 9 x 6
##   usuario quantidade data                nome     sobrenome dataNasc
##   <chr>        <dbl> <dttm>              <chr>    <chr>     <chr>
## 1 U0002         3372  2022-07-07 00:00:00 Isa      Soares    05/04/19
## 2 U0004        25761. 2022-04-06 00:00:00 Roberto  Jorge     30/08/19
## 3 U0005      1011250. 2021-01-10 00:00:00 Corina   Cruz      08/03/19
## 4 U0002       613021. 2022-06-27 00:00:00 Isa      Soares    05/04/19
## 5 U0001       170038. 2022-09-25 00:00:00 Arnaldo  Moreno    08/02/20
## 6 U0004       290199. 2022-05-03 00:00:00 Roberto  Jorge     30/08/19
## 7 U0002      2031465. 2022-06-10 00:00:00 Isa      Soares    05/04/19
## 8 U0002       989311. 2022-10-14 00:00:00 Isa      Soares    05/04/19
## 9 U0003           NA  NA                  Teodósio Gomes     10/07/19
```

União de conjuntos de dados

```
transacoes %>%
  inner_join(usuarios,by="usuario")
```

```
## # A tibble: 8 x 6
##   usuario quantidade data                nome    sobrenome dataNasc
##   <chr>        <dbl> <dttm>              <chr>   <chr>     <chr>
## 1 U0002         3372  2022-07-07 00:00:00 Isa     Soares    05/04/197
## 2 U0004        25761. 2022-04-06 00:00:00 Roberto Jorge     30/08/198
## 3 U0005      1011250. 2021-01-10 00:00:00 Corina  Cruz      08/03/199
## 4 U0002       613021. 2022-06-27 00:00:00 Isa     Soares    05/04/197
## 5 U0001       170038. 2022-09-25 00:00:00 Arnaldo Moreno    08/02/200
## 6 U0004       290199. 2022-05-03 00:00:00 Roberto Jorge     30/08/198
## 7 U0002      2031465. 2022-06-10 00:00:00 Isa     Soares    05/04/197
## 8 U0002       989311. 2022-10-14 00:00:00 Isa     Soares    05/04/197
```

União de conjuntos de dados

```
transacoes %>%
  full_join(usuarios,by="usuario")
```

```
## # A tibble: 10 x 6
##    usuario quantidade data                nome     sobrenome dataNas
##    <chr>        <dbl> <dttm>              <chr>    <chr>     <chr>
##  1 U0002         3372  2022-07-07 00:00:00 Isa      Soares    05/04/1
##  2 U0004        25761. 2022-04-06 00:00:00 Roberto  Jorge     30/08/1
##  3 U0005      1011250. 2021-01-10 00:00:00 Corina   Cruz      08/03/1
##  4 U0006        27697. 2021-01-04 00:00:00 <NA>     <NA>      <NA>
##  5 U0002       613021. 2022-06-27 00:00:00 Isa      Soares    05/04/1
##  6 U0001       170038. 2022-09-25 00:00:00 Arnaldo  Moreno    08/02/2
##  7 U0004       290199. 2022-05-03 00:00:00 Roberto  Jorge     30/08/1
##  8 U0002      2031465. 2022-06-10 00:00:00 Isa      Soares    05/04/1
##  9 U0002       989311. 2022-10-14 00:00:00 Isa      Soares    05/04/1
## 10 U0003            NA  NA                 Teodósio Gomes     10/07/1
```

tidyr

pivot, unnest, separate, unite

Funções pivot_longer e pivot_wider

```
(avaliacoes <-
   tibble("Avaliação 1" = c(1.50, 1.55, 1.54),
          "Avaliação 2" = c(1.51, 1.56, 1.54)))
```

```
## # A tibble: 3 x 2
##   `Avaliação 1` `Avaliação 2`
##           <dbl>         <dbl>
## 1           1.5          1.51
## 2           1.55         1.56
## 3           1.54         1.54
```

```
(estados <-
   tibble(Região = c(rep("Sul", 3), rep("Centro-Oeste", 3)),
          ID = c(1,2,3, 1,2,3),
          Estado = c("RS", "PR", "SC", "MS", "MT", "GO")))
```

```
## # A tibble: 6 x 3
##   Região          ID Estado
##   <chr>        <dbl> <chr>
## 1 Sul              1 RS
## 2 Sul              2 PR
## 3 Sul              3 SC
## 4 Centro-Oeste     1 MS
## 5 Centro-Oeste     2 MT
## 6 Centro-Oeste     3 GO
```

Funções pivot_longer e pivot_wider

```
(avaliacoes <-
   tibble("Avaliação 1" = c(1.50, 1.55, 1.54),
          "Avaliação 2" = c(1.51, 1.56, 1.54)))
```

```
## # A tibble: 3 x 2
##   `Avaliação 1` `Avaliação 2`
##           <dbl>         <dbl>
## 1           1.5           1.51
## 2           1.55          1.56
## 3           1.54          1.54
```

```
(estados <-
   tibble(Região = c(rep("Sul", 3), rep("Centro-Oeste", 3)),
          ID = c(1,2,3, 1,2,3),
          Estado = c("RS", "PR", "SC", "MS", "MT", "GO")))
```

```
## # A tibble: 6 x 3
##   Região          ID Estado
##   <chr>        <dbl> <chr>
## 1 Sul              1 RS
## 2 Sul              2 PR
## 3 Sul              3 SC
## 4 Centro-Oeste     1 MS
## 5 Centro-Oeste     2 MT
## 6 Centro-Oeste     3 GO
```

Funções pivot_longer e pivot_wider

```
library(tidyr)

avaliacoes %>%
  pivot_longer(cols=c("Avaliação 1", "Avaliação 2"),
               names_to="Avaliação", values_to="Altura (cm)")
```

```
## # A tibble: 6 x 2
##   Avaliação   `Altura (cm)`
##   <chr>               <dbl>
## 1 Avaliação 1          1.5
## 2 Avaliação 2          1.51
## 3 Avaliação 1          1.55
## 4 Avaliação 2          1.56
## 5 Avaliação 1          1.54
## 6 Avaliação 2          1.54
```

Funções pivot_longer e pivot_wider

```
estados %>%
  pivot_wider(id_cols=ID, names_from=Região,
              values_from=Estado)
```

```
## # A tibble: 3 x 3
##      ID Sul   `Centro-Oeste`
##   <dbl> <chr> <chr>
## 1     1 RS    MS
## 2     2 PR    MT
## 3     3 SC    GO
```

```
iris
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width     Spec
## 1            5.1         3.5          1.4         0.2      set
## 2            4.9         3.0          1.4         0.2      set
## 3            4.7         3.2          1.3         0.2      set
## 4            4.6         3.1          1.5         0.2      set
## 5            5.0         3.6          1.4         0.2      set
## 6            5.4         3.9          1.7         0.4      set
## 7            4.6         3.4          1.4         0.3      set
## 8            5.0         3.4          1.5         0.2      set
## 9            4.4         2.9          1.4         0.2      set
## 10           4.9         3.1          1.5         0.1      set
## 11           5.4         3.7          1.5         0.2      set
## 12           4.8         3.4          1.6         0.2      set
## 13           4.8         3.0          1.4         0.1      set
## 14           4.3         3.0          1.1         0.1      set
## 15           5.8         4.0          1.2         0.2      set
## 16           5.7         4.4          1.5         0.4      set
## 17           5.4         3.9          1.3         0.4      set
```

tidyr

```
iris %>%
  pivot_longer(c(Petal.Length, Sepal.Length,
                 Petal.Width, Sepal.Width))
```

```
## # A tibble: 600 x 3
##    Species name         value
##    <fct>   <chr>        <dbl>
##  1 setosa  Petal.Length   1.4
##  2 setosa  Sepal.Length   5.1
##  3 setosa  Petal.Width    0.2
##  4 setosa  Sepal.Width    3.5
##  5 setosa  Petal.Length   1.4
##  6 setosa  Sepal.Length   4.9
##  7 setosa  Petal.Width    0.2
##  8 setosa  Sepal.Width    3
##  9 setosa  Petal.Length   1.3
## 10 setosa  Sepal.Length   4.7
## # i 590 more rows
```

tidyr

```
iris %>%
  pivot_longer(-Species)
```

```
## # A tibble: 600 x 3
##    Species name         value
##    <fct>   <chr>        <dbl>
##  1 setosa  Sepal.Length   5.1
##  2 setosa  Sepal.Width    3.5
##  3 setosa  Petal.Length   1.4
##  4 setosa  Petal.Width    0.2
##  5 setosa  Sepal.Length   4.9
##  6 setosa  Sepal.Width    3
##  7 setosa  Petal.Length   1.4
##  8 setosa  Petal.Width    0.2
##  9 setosa  Sepal.Length   4.7
## 10 setosa  Sepal.Width    3.2
## # i 590 more rows
```

tidyr

```
iris %>%
  pivot_longer(c(starts_with("Petal"),
                 starts_with("Sepal")))
```

```
## # A tibble: 600 x 3
##    Species name         value
##    <fct>   <chr>        <dbl>
##  1 setosa  Petal.Length   1.4
##  2 setosa  Petal.Width    0.2
##  3 setosa  Sepal.Length   5.1
##  4 setosa  Sepal.Width    3.5
##  5 setosa  Petal.Length   1.4
##  6 setosa  Petal.Width    0.2
##  7 setosa  Sepal.Length   4.9
##  8 setosa  Sepal.Width    3
##  9 setosa  Petal.Length   1.3
## 10 setosa  Petal.Width    0.2
## # i 590 more rows
```

tidyr

```
iris %>%
  pivot_longer(c(starts_with("Petal"),
                 starts_with("Sepal"))) %>%
  separate(name, c("Parte da flor", "Medida"))
```

```
## # A tibble: 600 x 4
##    Species `Parte da flor` Medida value
##    <fct>   <chr>           <chr>  <dbl>
##  1 setosa  Petal           Length   1.4
##  2 setosa  Petal           Width    0.2
##  3 setosa  Sepal           Length   5.1
##  4 setosa  Sepal           Width    3.5
##  5 setosa  Petal           Length   1.4
##  6 setosa  Petal           Width    0.2
##  7 setosa  Sepal           Length   4.9
##  8 setosa  Sepal           Width    3
##  9 setosa  Petal           Length   1.3
## 10 setosa  Petal           Width    0.2
## # i 590 more rows
```

tidyr

```
dados
```

```
##                    val
## Trat1_Rep1 15.78291
## Trat1_Rep2 16.14339
## Trat1_Rep3 15.74257
## Trat1_Rep4 14.94049
## Trat1_Rep5 15.02677
## Trat2_Rep1 16.72542
## Trat2_Rep2 17.44605
## Trat2_Rep3 16.42591
## Trat2_Rep4 17.72766
## Trat2_Rep5 16.87040
## Trat3_Rep1 18.33888
## Trat3_Rep2 18.47856
## Trat3_Rep3 18.67104
## Trat3_Rep4 17.86359
## Trat3_Rep5 17.75907
## Trat4_Rep1 21.25240
## Trat4_Rep2 22.04975
```

tidyr

```
dados %>%
  rownames_to_column("id")
```

```
##              id      val
## 1   Trat1_Rep1 15.78291
## 2   Trat1_Rep2 16.14339
## 3   Trat1_Rep3 15.74257
## 4   Trat1_Rep4 14.94049
## 5   Trat1_Rep5 15.02677
## 6   Trat2_Rep1 16.72542
## 7   Trat2_Rep2 17.44605
## 8   Trat2_Rep3 16.42591
## 9   Trat2_Rep4 17.72766
## 10  Trat2_Rep5 16.87040
## 11  Trat3_Rep1 18.33888
## 12  Trat3_Rep2 18.47856
## 13  Trat3_Rep3 18.67104
## 14  Trat3_Rep4 17.86359
## 15  Trat3_Rep5 17.75907
## 16  Trat4_Rep1 21.25240
```

tidyr

```
dados %>%
  rownames_to_column("id") %>%
  separate(id, c("Trat", "Rep"))
```

```
##      Trat  Rep      val
## 1   Trat1 Rep1 15.78291
## 2   Trat1 Rep2 16.14339
## 3   Trat1 Rep3 15.74257
## 4   Trat1 Rep4 14.94049
## 5   Trat1 Rep5 15.02677
## 6   Trat2 Rep1 16.72542
## 7   Trat2 Rep2 17.44605
## 8   Trat2 Rep3 16.42591
## 9   Trat2 Rep4 17.72766
## 10 Trat2 Rep5 16.87040
## 11 Trat3 Rep1 18.33888
## 12 Trat3 Rep2 18.47856
## 13 Trat3 Rep3 18.67104
## 14 Trat3 Rep4 17.86359
```

```
medidas
```

```
## # A tibble: 30 x 4
##    Tamanho Cor    Repetição Medida
##    <fct>   <fct>  <fct>       <dbl>
##  1 Grande  Claro  1             1.8
##  2 Grande  Claro  2            15.4
##  3 Grande  Claro  3             5.2
##  4 Grande  Claro  4             2.2
##  5 Grande  Claro  5            12.4
##  6 Grande  Escuro 1            20
##  7 Grande  Escuro 2             5.4
##  8 Grande  Escuro 3            NA
##  9 Grande  Escuro 4             5.8
## 10 Grande  Escuro 5             5.4
## # i 20 more rows
```

```
medidas %>%
  unite(ID, c(Tamanho, Cor, Repetição))
```

```
## # A tibble: 30 x 2
##    ID                Medida
##    <chr>              <dbl>
##  1 Grande_Claro_1       1.8
##  2 Grande_Claro_2      15.4
##  3 Grande_Claro_3       5.2
##  4 Grande_Claro_4       2.2
##  5 Grande_Claro_5      12.4
##  6 Grande_Escuro_1     20
##  7 Grande_Escuro_2      5.4
##  8 Grande_Escuro_3     NA
##  9 Grande_Escuro_4      5.8
## 10 Grande_Escuro_5      5.4
## # i 20 more rows
```

tidyr

```
medidas %>%
  unite(ID, c(Tamanho, Cor, Repetição)) %>%
  column_to_rownames("ID")
```

```
##                      Medida
## Grande_Claro_1      1.80000
## Grande_Claro_2     15.40000
## Grande_Claro_3      5.20000
## Grande_Claro_4      2.20000
## Grande_Claro_5     12.40000
## Grande_Escuro_1    20.00000
## Grande_Escuro_2     5.40000
## Grande_Escuro_3          NA
## Grande_Escuro_4     5.80000
## Grande_Escuro_5     5.40000
## Médio_Claro_1            NA
## Médio_Claro_2       7.00000
## Médio_Claro_3      16.33333
## Médio_Claro_4      27.33333
## Médio_Claro_5      26.66667
```

```
excel_sheets("../dados/CaribbeanMaize.xlsx")
```

```
## [1] "Antigua"    "StVincent"
```

tidyr

```
lapply(c("Antigua", "StVincent"),
       read_xlsx, path="../dados/CaribbeanMaize.xlsx") %>%
  lapply(pivot_longer, c(-block, -plot), names_to="site", values
  tibble(isle=c("Antigua", "StVincent"),
         area=c(280, 345),
         dados=.) %>%
  unnest(dados)
```

```
## # A tibble: 612 x 6
##    isle     area block  plot site  yield
##    <chr>   <dbl> <chr> <dbl> <chr> <dbl>
##  1 Antigua   280 B1        1 DBAN   4.96
##  2 Antigua   280 B1        1 LFAN   2.92
##  3 Antigua   280 B1        1 TEAN   1.27
##  4 Antigua   280 B1        1 WEAN   4.02
##  5 Antigua   280 B1        1 WLAN   2
##  6 Antigua   280 B1        1 NSAN   2.43
##  7 Antigua   280 B1        1 OVAN   2.04
##  8 Antigua   280 B1        1 ORAN   5.26
##  9 Antigua   280 B1        2 DBAN   3.94
```

```r
medidas %>%
  mutate(Medida = replace_na(Medida, 0))
```

```
## # A tibble: 30 x 4
##    Tamanho Cor    Repetição Medida
##    <fct>   <fct>  <fct>       <dbl>
##  1 Grande  Claro  1             1.8
##  2 Grande  Claro  2            15.4
##  3 Grande  Claro  3             5.2
##  4 Grande  Claro  4             2.2
##  5 Grande  Claro  5            12.4
##  6 Grande  Escuro 1            20
##  7 Grande  Escuro 2             5.4
##  8 Grande  Escuro 3             0
##  9 Grande  Escuro 4             5.8
## 10 Grande  Escuro 5             5.4
## # i 20 more rows
```

Funções do pacote `forcats` para valores ausentes

```
library(forcats)

# fct_na_value_to_level
medidas %>%
  mutate(Cor = fct_na_level_to_value(Cor, "Claro"))
```

```
## # A tibble: 30 x 4
##    Tamanho Cor    Repetição Medida
##    <fct>   <fct>  <fct>       <dbl>
##  1 Grande  <NA>   1             1.8
##  2 Grande  <NA>   2            15.4
##  3 Grande  <NA>   3             5.2
##  4 Grande  <NA>   4             2.2
##  5 Grande  <NA>   5            12.4
##  6 Grande  Escuro 1            20
##  7 Grande  Escuro 2             5.4
##  8 Grande  Escuro 3            NA
##  9 Grande  Escuro 4             5.8
```

Exemplo

Exemplo

```
library(readr)
(linhas <- read_lines("http://www.leb.esalq.usp.br/leb/exceldados/DCE20
```

```
##    [1] "=================================================================
##    [2] "No  ANO    DIA    MES    R.GLOBA  INSO-PRECIPIUMIDADE  VENTO
##    [3] "                              2  LACAO TACAO RELATIV MAXIMO
##    [4] "                         cal/cm.   h/d    mm      %      m/s
##    [5] "=================================================================
##    [6] "1    2023     1      JAN      492    6,1   10,9      83   10,2
##    [7] "2    2023     2      JAN      514    6,7    4,3      80   12,0
##    [8] "3    2023     3      JAN      434    4,6    1,5      84   11,1
##    [9] "4    2023     4      JAN      331    1,9   16,8      89   10,4
##   [10] "5    2023     5      JAN      167    0,0    9,1      88    8,9
##   [11] "6    2023     6      JAN      308    1,3    0,0      78   12,3
##   [12] "7    2023     7      JAN      339    2,1    0,0      74   11,3
##   [13] "8    2023     8      JAN      435    4,6    0,3      82    8,6
##   [14] "9    2023     9      JAN      260    0,0    4,6      89    8,5
##   [15] "10   2023    10      JAN      329    1,8   48,8      90    8,1
##   [16] "11   2023    11      JAN      360    2,7    8,0      89    6,4
##   [17] "12   2023    12      JAN      408    3,9    3,6      85    6,7
##   [18] "13   2023    13      JAN      458    5,2   83,2      88    2,6
```

Exemplo

```
library(stringr)
which(str_starts(linhas, "="))
```

```
##   [1]   1   5  37  46  50  82  91  95 127 136 140 172 181 185
## [20] 275 307 316 320 352 361 365 397 406 410 442 451 455 487
```

```
length(which(str_starts(linhas, "=")))
```

```
## [1] 37
```

```
which(str_starts(linhas, "="))[c(1, seq(3, 37, 3))] # inicios
```

```
##  [1]   1  37  82 127 172 217 262 307 352 397 442 487 532
```

```
which(str_starts(linhas, "="))[c(2, seq(5,37,3), 37)] # finais
```

```
##  [1]   5  50  95 140 185 230 275 320 365 410 455 500 541
```

Exemplo

```
library(stringr)
which(str_starts(linhas, "="))
```

```
##   [1]   1   5  37  46  50  82  91  95 127 136 140 172 181 185
## [20] 275 307 316 320 352 361 365 397 406 410 442 451 455 487
```

```
length(which(str_starts(linhas, "=")))
```

```
## [1] 37
```

```
which(str_starts(linhas, "="))[c(1, seq(3, 37, 3))] # inicios
```

```
##   [1]   1  37  82 127 172 217 262 307 352 397 442 487 532
```

```
which(str_starts(linhas, "="))[c(2, seq(5,37,3), 37)] # finais
```

```
##   [1]   5  50  95 140 185 230 275 320 365 410 455 500 541
```

Exemplo

```
library(stringr)
which(str_starts(linhas, "="))
```

```
##  [1]   1   5  37  46  50  82  91  95 127 136 140 172 181 185
## [20] 275 307 316 320 352 361 365 397 406 410 442 451 455 487
```

```
length(which(str_starts(linhas, "=")))
```

```
## [1] 37
```

```
which(str_starts(linhas, "="))[c(1, seq(3, 37, 3))] # inicios
```

```
##  [1]   1  37  82 127 172 217 262 307 352 397 442 487 532
```

```
which(str_starts(linhas, "="))[c(2, seq(5,37,3), 37)] # finais
```

```
##  [1]   5  50  95 140 185 230 275 320 365 410 455 500 541
```

Exemplo

```
inicio <- which(str_starts(linhas, "="))[c(1, seq(3, 37, 3))]
final <- which(str_starts(linhas, "="))[c(2, seq(5,37,3), 37)]

pular <- sapply(seq(1,13), function(i){
  seq(inicio[i], final[i])
})

pular
```

```
## [[1]]
## [1] 1 2 3 4 5
##
## [[2]]
##  [1] 37 38 39 40 41 42 43 44 45 46 47 48 49 50
##
## [[3]]
##  [1] 82 83 84 85 86 87 88 89 90 91 92 93 94 95
##
## [[4]]
##  [1] 127 128 129 130 131 132 133 134 135 136 137 138 139 140
##
```

Exemplo

```
pular2 <- which(linhas=="")

linhas2 <- linhas[-c(unlist(pular),pular2)]
```

Exemplo

```
head(linhas2)
```

```
## [1] "1    2023    1    JAN    492    6,1   10,9    83   10,2
## [2] "2    2023    2    JAN    514    6,7    4,3    80   12,0
## [3] "3    2023    3    JAN    434    4,6    1,5    84   11,1
## [4] "4    2023    4    JAN    331    1,9   16,8    89   10,4
## [5] "5    2023    5    JAN    167    0,0    9,1    88    8,9
## [6] "6    2023    6    JAN    308    1,3    0,0    78   12,3
```

```
as_tibble(linhas2)
```

```
## # A tibble: 366 x 1
##    value
##    <chr>
##  1 1    2023    1    JAN    492    6,1   10,9    83   10,2    5
##  2 2    2023    2    JAN    514    6,7    4,3    80   12,0    6
##  3 3    2023    3    JAN    434    4,6    1,5    84   11,1    6
##  4 4    2023    4    JAN    331    1,9   16,8    89   10,4    9
##  5 5    2023    5    JAN    167    0,0    9,1    88    8,9   12
##  6 6    2023    6    JAN    308    1,3    0,0    78   12,3   15
##  7 7    2023    7    JAN    339    2,1    0,0    74   11,3   15
##  8 8    2023    8    JAN    435    4,6    0,3    82    8,6   10
```

Exemplo

```
head(linhas2)
```

```
## [1] "1    2023    1    JAN    492    6,1   10,9    83   10,2
## [2] "2    2023    2    JAN    514    6,7    4,3    80   12,0
## [3] "3    2023    3    JAN    434    4,6    1,5    84   11,1
## [4] "4    2023    4    JAN    331    1,9   16,8    89   10,4
## [5] "5    2023    5    JAN    167    0,0    9,1    88    8,9
## [6] "6    2023    6    JAN    308    1,3    0,0    78   12,3
```

```
as_tibble(linhas2)
```

```
## # A tibble: 366 x 1
##    value
##    <chr>
##  1 1    2023    1    JAN    492    6,1   10,9    83   10,2    5
##  2 2    2023    2    JAN    514    6,7    4,3    80   12,0    6
##  3 3    2023    3    JAN    434    4,6    1,5    84   11,1    6
##  4 4    2023    4    JAN    331    1,9   16,8    89   10,4    9
##  5 5    2023    5    JAN    167    0,0    9,1    88    8,9   12
##  6 6    2023    6    JAN    308    1,3    0,0    78   12,3   15
##  7 7    2023    7    JAN    339    2,1    0,0    74   11,3   15
##  8 8    2023    8    JAN    435    4,6    0,3    82    8,6   10
```

Exemplo

```
linhas[c(2, 3, 4, 6)]
```

```
## [1] "No  ANO    DIA    MES    R.GLOBA  INSO-PRECIPIUMIDADE  VENTO VENTO  TEMPER TEMPER TEMPER EVAPO-"
## [2] "                           2  LACAO TACAO RELATIV MAXIMO MEDIO  MAXIMA MINIMA  MEDIA RACAO"
## [3] "                       cal/cm.   h/d    mm     %      m/s   km/h  grau C grau C grau Cmm"
## [4] "1    2023     1     JAN      492    6,1   10,9     83   10,2    5,1   29,7   19,5   24,6  5,95
```

```
str_replace_all(linhas[6], " +", " ")
```

```
## [1] "1 2023 1 JAN 492 6,1 10,9 83 10,2 5,1 29,7 19,5 24,6 5,95 29,7 19,5 24,6 10,9"
```

```
linhas2[60]
```

```
## [1] "60   2023    29     FEV"
```

Exemplo

```
as_tibble(linhas2) %>%
  separate(value,
           c("No", "ANO", "DIA", "MES", "R.GLOBA",
             "INSOLACAO", "PRECIPITACAO", "UMIDADE RELATIV",
             "VENTO MAXIMO", "VENTO MEDIO", "TEMPER MAXIMA",
             "TEMPER MINIMA", "TEMPER MEDIA", "EVAPORACAO"),
           sep=" +")
```

```
## # A tibble: 366 x 14
##    No    ANO   DIA   MES   R.GLOBA INSOLACAO PRECIPITACAO `UMIDADE R
##    <chr> <chr> <chr> <chr> <chr>   <chr>     <chr>         <chr>
##  1 1     2023  1     JAN   492     6,1       10,9          83
##  2 2     2023  2     JAN   514     6,7       4,3           80
##  3 3     2023  3     JAN   434     4,6       1,5           84
##  4 4     2023  4     JAN   331     1,9       16,8          89
##  5 5     2023  5     JAN   167     0,0       9,1           88
##  6 6     2023  6     JAN   308     1,3       0,0           78
##  7 7     2023  7     JAN   339     2,1       0,0           74
##  8 8     2023  8     JAN   435     4,6       0,3           82
##  9 9     2023  9     JAN   260     0,0       4,6           89
## 10 10    2023  10    JAN   329     1,8       48,8          90
```

Exemplo

```
as_tibble(linhas2) %>%
  separate(value,
           c("No", "ANO", "DIA", "MES", "R.GLOBA",
             "INSOLACAO", "PRECIPITACAO", "UMIDADE RELATIV",
             "VENTO MAXIMO", "VENTO MEDIO", "TEMPER MAXIMA",
             "TEMPER MINIMA", "TEMPER MEDIA", "EVAPORACAO"),
           sep=" +") %>%
  unite(Data, DIA, MES, ANO)
```

```
## # A tibble: 366 x 12
##    No    Data    R.GLOBA INSOLACAO PRECIPITACAO `UMIDADE RELATIV` `V
##    <chr> <chr>   <chr>   <chr>     <chr>         <chr>              <c
##  1 1     1_JAN_~ 492     6,1       10,9          83                 10
##  2 2     2_JAN_~ 514     6,7       4,3           80                 12
##  3 3     3_JAN_~ 434     4,6       1,5           84                 11
##  4 4     4_JAN_~ 331     1,9       16,8          89                 10
##  5 5     5_JAN_~ 167     0,0       9,1           88                 8,
##  6 6     6_JAN_~ 308     1,3       0,0           78                 12
##  7 7     7_JAN_~ 339     2,1       0,0           74                 11
##  8 8     8_JAN_~ 435     4,6       0,3           82                 8,
##  9 9     9_JAN_~ 260     0,0       4,6           89                 8,
```

Exemplo

```
as_tibble(linhas2) %>%
  separate(value,
           c("No", "ANO", "DIA", "MES", "R.GLOBA",
             "INSOLACAO", "PRECIPITACAO", "UMIDADE RELATIV",
             "VENTO MAXIMO", "VENTO MEDIO", "TEMPER MAXIMA",
             "TEMPER MINIMA", "TEMPER MEDIA", "EVAPORACAO"),
           sep=" +") %>%
  unite(Data, DIA, MES, ANO) %>%
  mutate(Data=lubridate::dmy(Data)) %>%
  mutate_at(vars(-Data), str_replace, ",", ".") %>%
  mutate_at(vars(-Data), parse_number)
```

```
## # A tibble: 366 x 12
##       No Data       R.GLOBA INSOLACAO PRECIPITACAO `UMIDADE RELATIV`
##    <dbl> <date>       <dbl>     <dbl>         <dbl>              <dbl>
##  1     1 2023-01-01     492       6.1          10.9                 83
##  2     2 2023-01-02     514       6.7           4.3                 80
##  3     3 2023-01-03     434       4.6           1.5                 84
##  4     4 2023-01-04     331       1.9          16.8                 89
##  5     5 2023-01-05     167       0             9.1                 88
##  6     6 2023-01-06     308       1.3           0                   78
```