

# R para Ciência de Dados 2

dplyr++

Rafael Vetromille

10/09/2020

## O pacote dplyr

```
# Carregando o pacote  
shhh(library(tidyverse))
```

### A função bind\_rows()

Vamos usar a função `bind_rows()` para juntarmos duas bases com as mesmas colunas. Neste caso, a função `bind_rows()` é equivalente à função `rbind()`.

```
# Juntando duas bases  
  
imdb_2015 <- readr::read_rds("./data/imdb_por_ano/imdb_2015.rds")  
imdb_2016 <- readr::read_rds("./data/imdb_por_ano/imdb_2016.rds")  
  
bind_rows(imdb_2015, imdb_2016) %>% head()  
  
## # A tibble: 6 x 15  
##   titulo    ano diretor duracao cor generos pais classificacao orcamento  
##   <chr> <int> <chr>    <int> <chr> <chr> <chr> <chr>          <int>  
## 1 Aveng~ 2015 Joss W~    141 Color Action~ USA A partir de ~ 250000000  
## 2 Juras~ 2015 Colin ~    124 Color Action~ USA A partir de ~ 150000000  
## 3 Furio~ 2015 James ~    140 Color Action~ USA A partir de ~ 190000000  
## 4 The G~ 2015 Peter ~     93 Color Advent~ USA Livre          NA  
## 5 Jupit~ 2015 Lana W~    127 Color Action~ USA A partir de ~ 176000000  
## 6 Insid~ 2015 Pete D~     95 Color Advent~ USA Livre          175000000  
## # ... with 6 more variables: receita <int>, nota_imdb <dbl>,  
## # likes_facebook <int>, ator_1 <chr>, ator_2 <chr>, ator_3 <chr>
```

**Observação:** tanto para a `bind_rows()` quanto para a `rbind()`, a ordem das colunas em ambas as bases pode ser diferente. As colunas são empilhadas pelo nome.

Podemos também usar a função `bind_rows()` para juntar várias tabelas. Aqui, todas as tabelas continuam tendo as mesmas colunas.

```
arquivos <- list.files("./data/imdb_por_ano/", full.names = TRUE)

teste <- arquivos %>%
  purrr::map(readr::read_rds) %>%
  bind_rows()
```

ou

```
teste <- arquivos %>%
  purrr::map_dfr(readr::read_rds)
```