Aula prática de relatórios

R para Ciência de Dados I

Beatriz, Tereza e turma

2022-09-26

## R Markdown

[](https://quarto.org/)

Segundo Leão and Santiago (2022),

é dito que (Leão and Santiago 2022) ….

* Cuidado com os caminhos!

library(tidyverse)  
library(dados)

imdb <- read\_rds("../dados/imdb.rds")

glimpse(imdb)

## Rows: 28,490  
## Columns: 20  
## $ id\_filme <chr> "tt0023352", "tt0037946", "tt0216204", "tt0171889…  
## $ titulo <chr> "Prestige", "Nob Hill", "The Shade", "Viewer Disc…  
## $ ano <dbl> 1931, 1945, 1999, 1998, 1987, 1945, 1949, 1956, 2…  
## $ data\_lancamento <chr> "1932-01-22", "1945-11-15", "2000-03-01", "2012-0…  
## $ generos <chr> "Adventure, Drama", "Drama, Musical", "Drama", "C…  
## $ duracao <dbl> 71, 95, 83, 105, 133, 91, 81, 98, 50, 85, 267, 11…  
## $ pais <chr> "USA", "USA", "USA", "USA", "USA", "USA", "USA", …  
## $ idioma <chr> "English", "English", "English", "English", "Engl…  
## $ orcamento <dbl> NA, NA, 400000, NA, 20000000, NA, NA, 1505000, 50…  
## $ receita <dbl> NA, NA, NA, NA, 67331309, NA, NA, NA, NA, NA, NA,…  
## $ receita\_eua <dbl> NA, NA, NA, NA, 51249404, NA, NA, NA, NA, NA, NA,…  
## $ nota\_imdb <dbl> 5.7, 6.3, 7.1, 3.4, 7.2, 7.1, 6.3, 7.0, 5.7, 3.1,…  
## $ num\_avaliacoes <dbl> 240, 246, 102, 111, 26257, 1639, 310, 162, 115, 1…  
## $ direcao <chr> "Tay Garnett", "Henry Hathaway", "Raphaël Nadjari…  
## $ roteiro <chr> "Harry Hervey, Tay Garnett", "Wanda Tuchock, Norm…  
## $ producao <chr> "RKO Pathé Pictures", "Twentieth Century Fox", "F…  
## $ elenco <chr> "Ann Harding, Adolphe Menjou, Melvyn Douglas, Ian…  
## $ descricao <chr> "A woman travels to a French penal colony in Indo…  
## $ num\_criticas\_publico <dbl> 12, 11, 1, 5, 142, 35, 8, 9, 4, 12, 4, 502, 6, 35…  
## $ num\_criticas\_critica <dbl> 2, 2, 1, 3, 62, 10, 5, NA, 5, 7, 7, 161, 1, 18, 8…

filmes\_nota\_boa <- imdb %>%   
 filter(nota\_imdb > 8, num\_avaliacoes > 10000)  
  
top\_filmes <- filmes\_nota\_boa %>%   
 arrange(desc(nota\_imdb)) %>%   
 slice(1:10) %>%   
 pull(titulo) %>%   
 knitr::combine\_words(and = " e ")

### Inline code!

A base IMDB possui 28490 filmes, sendo 130 filmes com nota maior do que 8 tendo mais do que 10 mil avaliações. Os filmes com melhor nota são The Shawshank Redemption, The Godfather, The Godfather: Part II, 12 Angry Men, Pulp Fiction, Schindler’s List, Wheels, Forrest Gump, Star Wars: Episode V - The Empire Strikes Back, e Goodfellas.

### Tabelas

| Col1 | Col2 | Col3 |
| --- | --- | --- |
| 28490 |  |  |
|  |  |  |
|  |  |  |

filmes\_nota\_boa %>%   
 mutate(lucro = receita - orcamento) %>%   
 slice\_max(order\_by = lucro, n = 10) %>%   
 select(titulo, ano, nota\_imdb, lucro) %>%   
 knitr::kable()

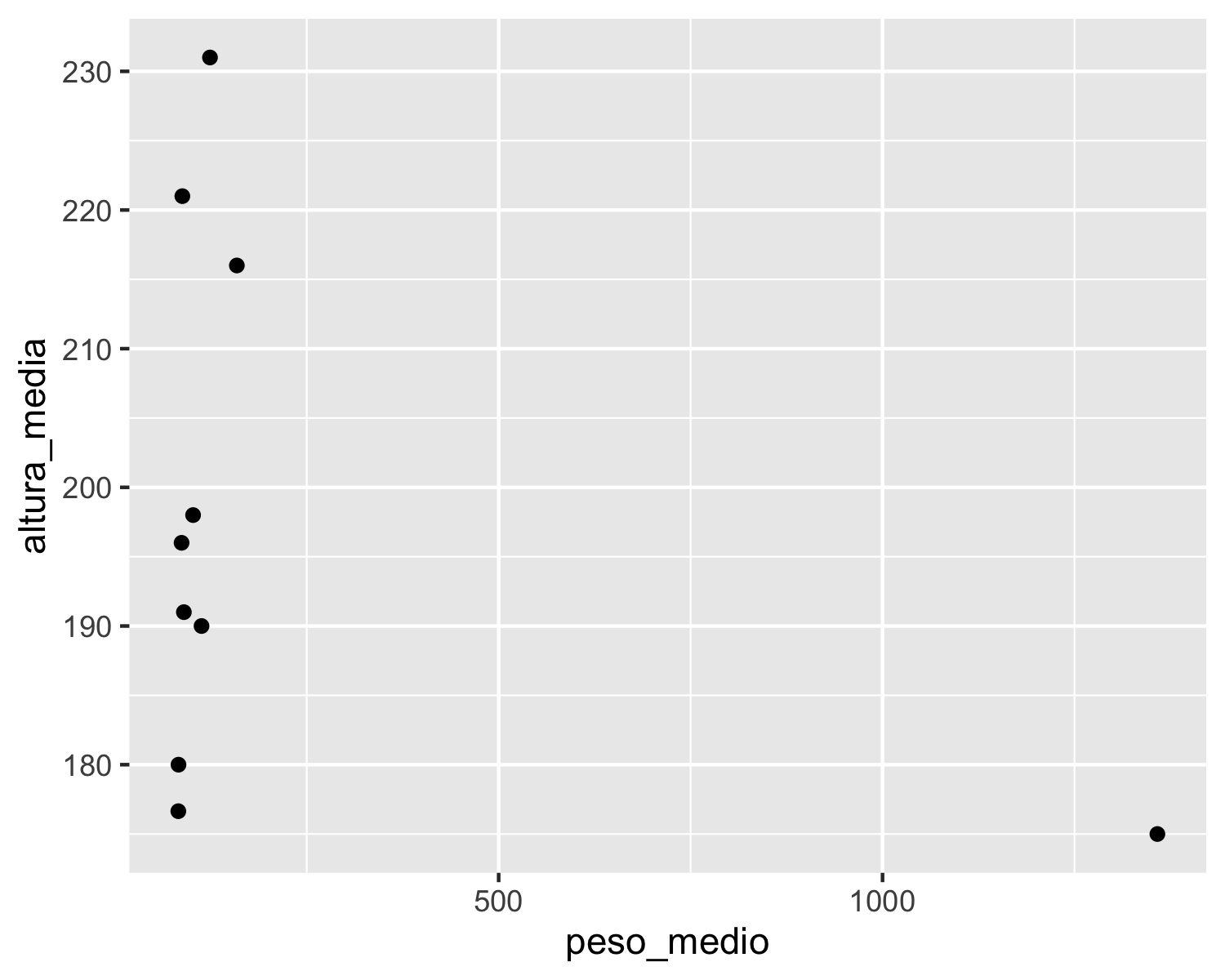
| titulo | ano | nota\_imdb | lucro |
| --- | --- | --- | --- |
| Avengers: Endgame | 2019 | 8.4 | 2441800564 |
| Avengers: Infinity War | 2018 | 8.4 | 1727359754 |
| Jurassic Park | 1993 | 8.1 | 970758854 |
| The Lion King | 1994 | 8.5 | 923511805 |
| Toy Story 3 | 2010 | 8.3 | 866969703 |
| Star Wars | 1977 | 8.6 | 764768912 |
| Inside Out | 2015 | 8.1 | 683071350 |
| The Sixth Sense | 1999 | 8.1 | 632806292 |
| Coco | 2017 | 8.4 | 632083670 |
| Forrest Gump | 1994 | 8.8 | 623229452 |

base\_tabela <- filmes\_nota\_boa %>%   
 mutate(lucro = receita - orcamento) %>%   
 slice\_max(order\_by = lucro, n = 10) %>%   
 mutate(lucro\_formatado = lucro / 1000000,  
 # imaginando se fosse em reais  
 lucro\_formatado = scales::dollar(round(lucro\_formatado), prefix = "R$ ", big.mark = ".", decimal.mark = ",")  
 ) %>%   
 select(Nome = titulo, Ano = ano, `Nota IMDB` = nota\_imdb, `Lucro (em milhões de reais)` = lucro\_formatado)   
  
  
 knitr::kable(base\_tabela)

| Nome | Ano | Nota IMDB | Lucro (em milhões de reais) |
| --- | --- | --- | --- |
| Avengers: Endgame | 2019 | 8.4 | R$ 2.442 |
| Avengers: Infinity War | 2018 | 8.4 | R$ 1.727 |
| Jurassic Park | 1993 | 8.1 | R$ 971 |
| The Lion King | 1994 | 8.5 | R$ 924 |
| Toy Story 3 | 2010 | 8.3 | R$ 867 |
| Star Wars | 1977 | 8.6 | R$ 765 |
| Inside Out | 2015 | 8.1 | R$ 683 |
| The Sixth Sense | 1999 | 8.1 | R$ 633 |
| Coco | 2017 | 8.4 | R$ 632 |
| Forrest Gump | 1994 | 8.8 | R$ 623 |

# base\_tabela %>%   
# DT::datatable()

peso\_altura <- dados\_starwars |>   
 drop\_na(especie) |>   
 group\_by(especie) |>   
 summarise(peso\_medio = mean(massa, na.rm = TRUE),  
 altura\_media = mean(altura, na.rm = TRUE)) |>   
 slice\_max(order\_by = peso\_medio, n = 10) |>   
 ungroup()  
  
peso\_altura |>   
 ggplot() +  
 geom\_point(aes(x = peso\_medio, y = altura\_media))





## Referências

Leão, Suiane Rodrigues, and Alvany Maria dos Santos Santiago. 2022. “Cenário Das Barragens de Rejeito: Conhecer Para Evitar Novas Catástrofes.” *Ambiente & Sociedade* 25. <https://doi.org/10.1590/1809-4422asoc20210066r1vu2022l2ao>.