

# Introdução à Programação e à R - Parte 2

Curso de Programação - 2025

João Pedro de Freitas Gomes

14 de Maio, 2025



1. Tidyverse

2. Prática

# Tidyverse

---

- Em 2007, um estatístico chamado Hadley Wickham publicou sua tese de Doutorado, intitulada *Practical tools for exploring data and models*.
- Nela, ele propôs dois pacotes para o R que mudariam a forma que se usa R. Um deles, o reshape, se tornaria a base para o que hoje conhecemos como Tidyverse.
- O outro pacote, intitulado ggplot2, é utilizado até hoje dentro deste ecossistema, e é a principal forma que utilizamos para visualizar dados.

# Datasets Tidy

- Cada variável é uma coluna
- Cada observação é uma linha
- Cada célula é um único valor

country	year	cases	population
Afghanistan	1999	1845	19987071
Afghanistan	2000	1866	20095360
Brazil	1999	31737	17206362
Brazil	2000	81488	17404898
China	1999	21258	127215272
China	2000	21366	12802583

variables

country	year	cases	population
Afghanistan	1999	1845	19987071
Afghanistan	2000	1866	20095360
Brazil	1999	31737	17206362
Brazil	2000	81488	17404898
China	1999	21258	127215272
China	2000	21366	12802583

observations

country	year	cases	population
Afghanistan	1999	1845	19987071
Afghanistan	2000	1866	20095360
Brazil	1999	31737	17206362
Brazil	2000	81488	17404898
China	1999	21258	127215272
China	2000	21366	12802583

values

```
flights %>%  
  filter(dest == "IAH") %>%  
    group_by(year, month, day) %>%  
      summarize(  
        arr_delay = mean(arr_delay, na.rm = TRUE)  
      )
```

- *filter()*
- *distinct()*
- *group\_by()*
- *summarize()*
- *count()*
- *select()*
- *mutate()*

# Exemplo com Visualização de Dados

```
flights %>%  
  filter(dest == "IAH") %>%  
  ggplot(aes(x=arr_delay, y=dep_delay)) + geom_point()
```



# Prática

---

- Utilizando as ferramentas que aprendemos na aula, analise, na base voês, qual companhia aérea(*carrier*) possui os piores atrasos médios.