Introdução à Programação e à R - Parte 2

Curso de Programação - 2023

João Pedro de Freitas Gomes

15 de Agosto, 2023

FEA - USP

Made
centro de
pesquisa em
macroeconomia
das desigualdades

Resumo

1. Resolução do Exercício Proposto na Aula Anterior
2. Tidyverse
3. Prática

Resolução do Exercício Proposto na Aula Anterior

Resolução

```
numeros <- 1:20
numeros_2 <- c()
for(i in numeros){
    if(numeros[i] %% 2 ==0){
        numeros_2[i] <- numeros[i]**2
    }else{
        numeros_2[i] <- numeros[i]*2
    }
}
```

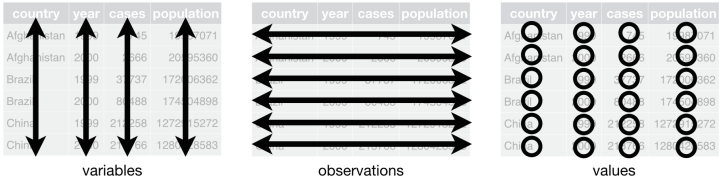
Tidyverse

Contexto

- 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



Dplyr

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

Dplyr - Transformando Dados

- *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

Exercício

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