

Exercício 7: Manipulação de dataframes com R e tidyverse

Introdução

Este exercício tem como objetivo familiarizar pesquisadores(as) com as funções básicas do R e com a biblioteca tidyverse, especialmente o pacote dplyr. Utilizaremos o conjunto de dados penguins para realizar diversas manipulações em dataframes.

Carregando as bibliotecas e os dados

Code snippet

```
library(tidyverse)
library(palmerpenguins)
```

Exercícios

1. Exploração inicial:

- Visualize as primeiras 10 linhas do dataset.
- Obtenha um resumo estatístico das variáveis numéricas.
- Verifique se há valores missing (na) no dataset.

2. Filtrando dados:

- Selecione apenas os pinguins da espécie "Adelie".
- Filtre os dados para incluir apenas os pinguins com massa corporal acima de 5000 g.
- Crie um novo dataframe com os pinguins da ilha "Torgersen" e com comprimento do bico maior que 40 mm.

3. Selecionando colunas:

- Selecione apenas as colunas "species", "island" e "body_mass_g".
- Crie um novo dataframe com as colunas "species", "bill_length_mm" e "flipper_length_mm".

4. Criando novas variáveis:

- Crie uma nova coluna chamada "body_mass_kg" convertendo a massa corporal para quilogramas.

- Crie uma coluna booleana indicando se o comprimento do bico é maior que a média.

5. Agrupando e resumindo dados:

- Calcule a média da massa corporal para cada espécie.
- Conte o número de pinguins em cada ilha.

6. Ordenando dados:

- Ordene os dados por espécie em ordem alfabética e, dentro de cada espécie, por massa corporal em ordem decrescente.

Expandindo o Exercício

- **Visualização:** Use ggplot2 para criar gráficos exploratórios dos dados.
- **Join:** Combine o dataset penguins com outro dataset (se disponível) usando left_join(), right_join(), etc.
- **Manipulação de datas:** Se o dataset tiver datas, utilize funções do pacote lubridate para manipulá-las.
- **Funções personalizadas:** Crie suas próprias funções para realizar tarefas específicas.
- **Observações:**
 - A biblioteca tidyverse oferece uma sintaxe intuitiva e concisa para manipulação de dados.
 - O pipe (%>%) permite encadear operações de forma clara e eficiente.
 - Explore a documentação do dplyr para descobrir mais funções e opções.