

# Exercícios

Data Science with R in Healthcare

André Peralta-Santos, Pedro Casaca

Setembro 2023

## Introdução

Esta série de exercícios é destinada aos praticantes do curso de Curta Duração *Data Science with R in Healthcare* que desejam melhorar as suas capacidades de programação em R e análise de dados. Os exercícios abrangem desde a organização de projetos até a análise estatística e visualização de dados. Para elaborar estes exercícios são necessários dois ficheiros adicionais no formato ".xlsx".

## Exercício 1

1. Crie um projeto em R com o nome *data-science-ensp*.
2. Organize o projeto seguindo as boas práticas, criando as pastas necessárias.
3. Crie um script em R Markdown com saída em formato HTML, incluindo ao menos uma imagem.
4. Crie outro script em R Markdown com saída em formato .docx, e inclua referências bibliográficas.

## Exercício 2

1. Copie o arquivo "hipoxia.xlsx" para a pasta *data*.
2. Crie um novo script em R Markdown com saída em formato HTML.
3. Importe o arquivo "hipoxia.xlsx" para a memória do RStudio.
4. Verifique o conteúdo do arquivo importado.
5. Crie um novo conjunto de dados com as seguintes variáveis:
  - `iv_morphin_eq`
  - `age`

- female
  - race
  - bmi
  - smoking
  - diabetes
  - propofol\_induction
  - crystalloids
  - colloids
6. Filtre o conjunto para incluir apenas indivíduos com mais de 25 anos de idade.
  7. Crie uma variável que seja a soma das variáveis *crystalloids* e *colloids*.
  8. Calcule as médias e desvios-padrão do *bmi* agrupados por sexo.

### Exercício 3

1. Copie o arquivo "municípios.xlsx" para a pasta *data*.
2. Importe o arquivo "municípios.xlsx" para a memória do RStudio.
3. Calcule a frequência de indivíduos por sexo e por município.

### Exercício 4

Escolha uma das seguintes opções:

- Crie um gráfico que relacione as variáveis *bmi* e *min\_sao2*.

OU

- Crie um gráfico que mostre a mediana e os intervalos interquartis das variáveis *bmi* e *diabetes*.

### Exercício 5

1. Construa um modelo estatístico para determinar os fatores de risco associados à utilização de CPAP, considerando as variáveis *age*, *female*, *race*, *bmi* e *sleeptime*.