

# Ciência de Dados com R

## Exercícios para Praticar



1. Crie os objetos:

- `nome` com seu nome (não esqueça das aspas);
- `idade` com a sua idade;
- `ano` com seu ano de nascimento;
- `r` sendo `TRUE` se você já conhecia o R, e `FALSE` caso contrário.

Execute as seguintes operações e comente os resultados:

- `idade + ano`
- `ano + r`
- `idade - r`
- `r*nome`
- `rep(nome, idade)`

2. A partir da função `seq`, crie uma sequência de valores `x` variando de -4 à 4 com 10 valores (utilize o parâmetro `length = 10`). Daí, crie uma variável `y` contendo os valores de `x` elevados ao quadrado e faça `plot(x, y, type="l")`.

3. Repita o gráfico anterior utilizando `length = 1000` para gerar `x`. Qual a diferença entre os gráficos?

4. Crie o vetor `s <- seq(0, 1000, by = 30)` e responda:

- Qual é a posição do número 510?
- Quantos elementos existem nesse vetor?
- Qual é a diferença entre o último e o vigésimo número?
- Sem utilizar a função `mean(s)`, qual é a média de `s`? (Dica: a média é a soma dos valores de `s`, dividido pela quantidade de elementos desse vetor)

5. Crie o vetor `w <- c(5, 10, 50, 30, 1, 0, 2, 300, 100, 60)`. A partir de `w`, utilize funções do R para chegar no vetor (100, 60, 50, 30, 10, 5, 2, 1).

6. Crie a matriz:

```
frutas <- matrix(c("maçã", "banana", "cereja", "laranja", "uva", "abacaxi",  
                  "pera", "melão", "figo"), nrow = 3, ncol = 3)
```

e responda:

- Como você acessaria apenas o elemento "uva"?
- Como você acessaria apenas os elementos "laranja", "uva" e "abacaxi"?
- Como você acessaria apenas os elementos "maçã", "pera", "cereja" e "figo"?

- (d) Adicione uma coluna com as frutas "manga", "melancia" e "acerola" e depois uma linha com as frutas "carambola", "caqui", "morango" e "pitaya".
7. A partir de uma idade, crie um código que printe: "menor de idade", se a idade for menor que 18 anos; "adulto", se a idade for maior ou igual a 18 e menor que 60 anos; e "idoso", se a idade for maior ou igual a 60 anos.
8. Crie uma função que recebe como argumento uma idade e que retorne: "menor de idade", se a idade for menor que 18 anos; "adulto", se a idade for maior ou igual a 18 e menor que 60 anos; e "idoso", se a idade for maior ou igual a 60 anos.
9. Crie os dataframes a seguir:
- ```
dados1 <- data.frame(Treino = c("Força", "Resistência", "Outro"),
                     Pulso = c(100, 150, 120), Duração = c(60, 30, 45))

dados2 <- data.frame(Passos = c(3000, 6000, 2000), Calorias = c(300, 400, 300))
```
- e faça o que se pede:
- (a) Junte os dois dataframes em um único dataframe chamado `dados`.
- (b) Encontre a média das `Calorias` e do `Pulso`.
- (c) Remova a primeira linha e a terceira coluna do dataframe `dados`.
10. Com o auxílio das funções do pacote `dplyr`, faça o que se pede:
- (a) Selecione 5 variáveis da base de dados `flights` e salve no objeto `dados`
- (b) Renomeie as variáveis do dataframe `dados` e salve no objeto `dados1`
- (c) Ordene o dataframe `dados1` de acordo com uma variável de sua escolha salve no objeto `dados2`
- (d) Filtre o dataframe `dados2` para remover linhas de seu interesse e salve no objeto `dados3`
- (e) Modifique uma coluna do seu interesse do dataframe `dados3` e salve no objeto `dados_final`
- (f) Repita esse processo utilizando o operador `%>%`
11. A partir da base de dados `mpg` e com o auxílio das funções do pacote `ggplot2`, faça o que se pede:
- (a) Crie um dataframe `dados_grafico` com as variáveis quantitativas `motor`, `mpgc` e `mpgr` e mais 2 variáveis qualitativas de sua escolha
- (b) Para cada uma das variáveis qualitativas, faça um gráfico de barras
- (c) Para cada uma das variáveis quantitativas, faça um histograma
- (d) Explore a associação entre cada uma das variáveis do dataframe `dados_grafico` com a variável `mpgr`
12. A partir da base de dados `mpg` e com o auxílio das funções do pacote `ggplot2`, faça o que se pede:
- (a) Crie um dataframe `dados_ajuste` com a variável `mpgr` e mais 2 variáveis quaisquer da base `mpg`
- (b) Faça uma análise gráfica das variáveis individualmente
- (c) Explore a associação entre cada uma das variáveis do dataframe `dados_ajuste` com a variável `mpgr`
- (d) Ajuste modelos de regressão
- (e) Escolha o melhor modelo através do AIC