

PROGRAMA AVANÇADO EM DATA SCIENCE E DECISÃO

Lista de exercícios

Nome: _____

- 1) Crie um vetor com os números 15, 10, 21, 30, 52 e 60. A partir desse vetor, calcule:
 - a) a média;
 - b) o desvio padrão;
 - c) indique o valor armazenado na terceira posição desse vetor utilizando [] e
 - d) apresente o comprimento do vetor utilizando uma função para isso.
- 2) Faça uma função para determinar a sequência de Fibonacci até o n-ésimo termo.
- 3) Utilizando o R crie um *data frame* (data frame ou tibble) com as seguintes informações e nomes de variáveis:

id	idade	categoria	satisfacao
1	35	tecnologia	70
2	24	romance	80
3	31	ficção	93
4	29	tecnologia	79
5	20	tecnologia	95
6	19	tecnologia	88
7	42	ficção	85
8	54	romance	91
9	49	tecnologia	100
10	60	ficção	79

A partir desse *data frame*, realize as seguintes tarefas:

- a) Defina dois objetos contendo, respectivamente, o número de linhas e de colunas desse *data frame*.
- b) Obtenha um vetor com os dados da coluna idade.
- c) Crie um novo objeto considerando apenas as linhas 1 e 2 desse *data frame*.
- d) Cálculo da idade média.
- e) Satisfação mediana.
- f) Contagem de categorias.
- g) Idade média de acordo com as categorias.
- h) Gráfico de dispersão de idade por satisfação.
- i) Refaça o item anterior indicando a cor dos pontos de acordo com a categoria.
- j) Refaça o item i) considerando uma linha de tendência.

PROGRAMA AVANÇADO EM DATA SCIENCE E DECISÃO

4) O banco de dados *diamonds* contido no pacote *ggplot2* apresenta informações de preço e característica de 53.940 diamantes (para mais detalhe, digite “?*diamonds*” no console).

Considerando esse banco de dados, execute as seguintes tarefas:

- a) Utilizando a função *count*, calcule a frequência absoluta das categorias da variável *color* ordenada de forma decrescente.
 - b) Faça a contagem das combinações de *cut* e *color*.
 - c) Crie a variável *ratio* dada pela razão entre *price* e *table*.
 - d) Calcule a média e desvio padrão do preço para cada categoria de *cut*.
 - e) Calcule as medidas do item anterior considerando apenas dados com *table* acima de 50.
 - f) Calcule o preço médio de acordo com a variável *cut* e faça um gráfico de colunas (*geom_col*) com as cores de acordo com *cut*.
 - g) Faça um gráfico de dispersão da variável *price* por *depth*.
 - h) Facete o gráfico do item anterior de acordo com *cut*.
- 5) Instale a biblioteca *gapminder* e carregue os dados *gapminder* utilizando o código abaixo e responda (assista o vídeo em <https://bit.ly/3Cvq2Vq>).

```
install.packages("gapminder")
```

```
library(gapminder)
```

```
?gapminder
```

- a) Quantas observações há no total?
- b) Para quais anos (*year*) há registros na base de dados?
- c) Quantas observações há para cada continente?
- d) Quantas observações há para cada continente em cada um dos anos considerados?
- e) Quantas observações há para cada continente em cada um dos anos considerados levando em conta apenas os dados a partir de 2000?
- f) Qual a média e desvio-padrão do PIB per capita (*gdpPercap*) dos continentes de acordo com o ano levando em conta apenas dados a partir de 2000?
- g) Faça o gráfico de linhas de expectativa de vida (*lifeExp*) de acordo com o ano (*year*) para os países (*country*).
- h) Obtenha a expectativa de vida média por continente ao longo dos anos.
- i) Faça um gráfico de linhas de expectativa de vida média ao longo dos anos para os continentes.