

Trilha 03

Indicações de Soluções das Atividades de Aprofundamento

Explorando a Base de Dados Diamonds do pacote ggplot2

Diamonds é uma base de dados que contém os preços e outros atributos de quase 54.000 diamantes. De fato, são 53940 linhas e 10 variáveis, mostradas a seguir:

Variável	Descrição
price	preço em dólares americanos (\$326–\$18,823)
carat	peso do diamante (0.2–5.01)
cut	qualidade do corte (Fair, Good, Very Good, Premium, Ideal)
color	cor do diamante, indo de J (pior) a D (melhor)
clarity	medida de quão claro é o diamante (I1 (pior), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (melhor))
x	comprimento em mm (0–10.74)
y	largura em mm (0–58.9)
z	profundidade em mm (0–31.8)
depth	percentual de profundidade total = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43–79)
table	largura do topo do diamante relativo ao ponto mais largo (43–95)

A leitura dos dados pode ser feita com o comando abaixo:

```
# caso não tenha, instale a biblioteca ggplot2
# install.packages("ggplot2")
library(ggplot2)
data(diamonds)
diamantes <- diamonds
```

Após carregar a base de dados no R, vamos a algumas explorações:

- Qual é a estrutura do conjunto de dados “diamantes”?
- Explore a parte inicial e a final do conjunto de dados.
- Faça alguns sumários estatísticos para entender melhor a base de dados.
- A saída da função `summary()` está de acordo com a descrição mostrada anteriormente?

- Explore a variável price, seguindo o modelo de exploração.
- Veja a distribuição da variável (histograma); observe a faixa de valores da variável e também.
- Explore também as variáveis carat, cut, color, clarity, x, y, z, depth e table, seguindo o modelo de exploração.
- Crie boxplots para as variáveis numéricas; veja se existem dados anormais (outliers).
- Utilize as variáveis categóricas para fazer o facetamento dos dados, mostrando alguns gráficos com 2 ou mais variáveis contínuas lado a lado.

Para cada resultado/gráfico obtido, explique e discuta-os, de modo a construir um relatório de exploração dos dados, que deverá ser submetido.