

1

Lab01 (análise de dados)

- Objetivo do código: O código está apenas fazendo uma análise de dados disponíveis no conjunto de dados "penguins" usando a biblioteca Seaborn.

Importação de bibliotecas:

- `import seaborn as sns` : Seaborn é uma biblioteca de visualização de dados em Python baseada no Matplotlib. Ele fornece uma interface de alto nível para criar gráficos estatísticos atraentes e informativos.
- `import matplotlib as plt` : Matplotlib é uma biblioteca de plotagem de baixo nível em Python, usada para criar gráficos estáticos, gráficos interativos e painéis de visualização.

```
import seaborn as sns
import matplotlib as plt
```

Carregamento dos dados:

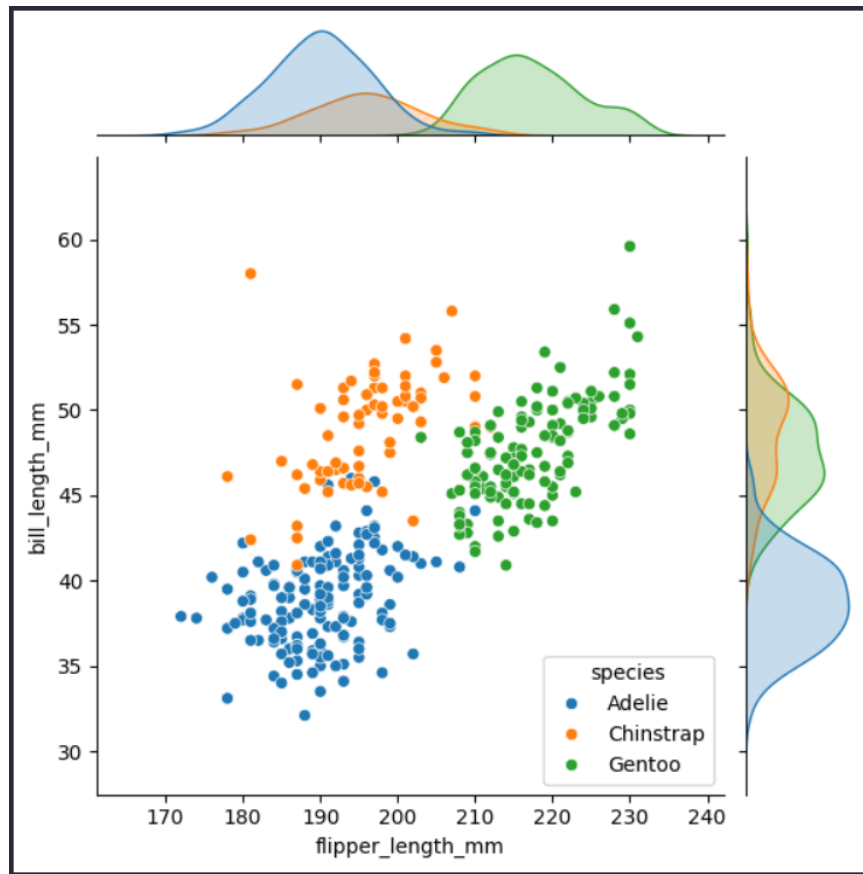
- `df = sns.load_dataset("penguins")` : Isso carrega um conjunto de dados de exemplo chamado "penguins" do pacote Seaborn e o armazena em um DataFrame do Pandas chamado `df`. Este conjunto de dados contém informações sobre pinguins, como comprimento do bico, comprimento da nadadeira, espécie, etc.

```
df = sns.load_dataset("penguins")
sns.jointplot(data=df, x="flipper_length_mm", y="bill_length_
```

Visualização:

- `sns.jointplot()` : Esta função cria um gráfico de dispersão com duas variáveis, além de histogramas de cada variável nas margens.

- `data=df` : Especifica o DataFrame a ser usado como fonte de dados.
 - `x="flipper_length_mm", y="bill_length_mm"` : Define as variáveis do DataFrame a serem plotadas nos eixos x e y, respectivamente. Aqui, estamos plotando o comprimento da nadadeira no eixo x e o comprimento do bico no eixo y.
 - `hue="species"` : Essa opção separa os pontos no gráfico com base em uma terceira variável (nesse caso, a espécie de pinguim), colorindo-os de forma diferente. Isso ajuda a visualizar a distribuição das duas variáveis em relação à espécie de pinguim.
- Um DataFrame é uma maneira conveniente de organizar e manipular dados em linhas e colunas. Cada coluna em um DataFrame representa uma variável e cada linha representa uma observação ou entrada de dados.
 - DataFrame `df` contém dados sobre pinguins, onde cada linha representa um pinguim individual e cada coluna representa uma característica desse pinguim, como comprimento da nadadeira, comprimento do bico, espécie, sexo, etc. O DataFrame permite acessar, visualizar e manipular esses dados de forma eficiente, facilitando a análise e a criação de visualizações.



- O gráfico foi gerado através da biblioteca Seaborn, que é construída sobre o Matplotlib