

DATA ANALYTICS
VISUALIZAÇÃO DE DADOS

AULA 02

SUMÁRIO

O QUE VEM POR AÍ?	3
CONHEÇA SOBRE O ASSUNTO	4
HANDS ON	9
O QUE VOCÊ VIU NESTA AULA?	10
REFERÊNCIAS	11

EMASP

O QUE VEM POR AÍ?

Até aqui você teve uma noção de como carregar fontes de dados de maneiras distintas e aprendeu algumas manipulações base para sua jornada.

Como meio de complementar suas análises, essa aula irá te mostrar como enxergar a proporcionalidade dos dados, utilizar uma biblioteca gráfica bem bacana, chamada Seaborn, e a como visualizar suas análises com ela.

A base de dados você encontra aqui <https://github.com/alura-tech/pos-datascience-introducao-a-visualizacao/archive/refs/heads/dados.zip>.

Agora vamos para cima deste maravilhoso mundo dos dados!

IMPORTANTE: Não esqueça que temos um desafio para essa disciplina, que faz parte da sua jornada de aprendizado. Você pode acessá-lo na área de atividades da plataforma FIAP ON! Caso precise de ajuda, basta entrar em contato conosco pelo Discord.

CONHEÇA SOBRE O ASSUNTO

É muito comum nós, que somos da área de dados e que utilizamos Python para manipular, chegemos em um ponto que nos leva à seguinte pergunta: será que existe algo além do Matplotlib?

E a resposta é: claro que existe!

A biblioteca Seaborn é a prova viva disso. Famosa por ter gráficos lindos e com um leque de variedades interessantes, ela vem para abalar nossos corações.

Dê uma olhada na página inicial:

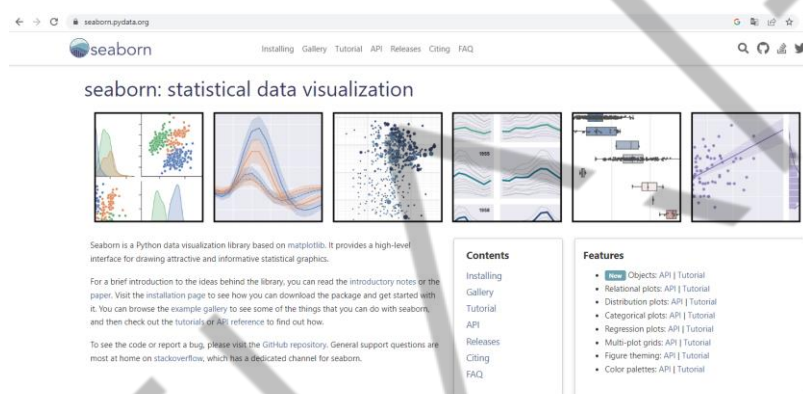


Figura 1 – Seaborn
Fonte: seaborn.pydata.org (2023)

O Seaborn é uma biblioteca de visualização de dados baseada em Matplotlib para o Python. É projetada para fornecer uma interface de alto nível para criar gráficos estatísticos atrativos e informativos. A biblioteca Seaborn suporta vários tipos de gráficos, incluindo gráficos de distribuição, gráficos de regressão, gráficos de violino, gráficos de barra, gráficos de área, entre outros. Além disso, a biblioteca inclui uma série de temas e paletas de cores para personalizar a aparência dos gráficos.

Com essa ferramenta, torna-se mais fácil trabalhar com dados categóricos, além de fornecer recursos avançados de visualização de dados, como plotagem de séries temporais, mapas de calor e gráficos de matriz de correlação. Ele também permite ajustar facilmente modelos estatísticos aos seus dados, como regressão linear e regressão logística.

Para você compreender um pouquinho do que estou dizendo, dê uma olhada na primeira parte da videoaula, onde mostramos a você como importar a biblioteca e fazer o primeiro gráfico.

Para entender o poder da ferramenta e como é simples, na segunda parte da videoaula, mostramos uma maneira de gerar três informações com o Scatterplot do Seaborn, utilizando três linhas:

```
plt.figure(figsize=(7,7))
sns.scatterplot(data = gastos_e_populacao, x = "populacao", y
="gastos_2021/Ago")
sns.scatterplot(data = gastos_e_populacao, x = "populacao", y
="gastos_2021/Jul")
sns.scatterplot(data = gastos_e_populacao, x = "populacao", y
="gastos_2021/Jun")
```

Cujo resultado é:

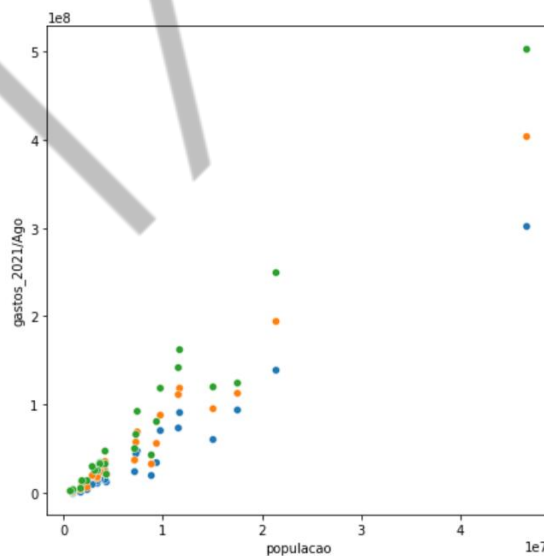


Figura 2 – Scatter múltiplo
Fonte: Elaborado pelo autor (2023)

Veja que os dados foram diferenciados por cor sem declararmos, ou seja, a biblioteca gerou automaticamente, mas podemos mudar os parâmetros e colocar outras cores.

Em comparação com o Matplotlib, o Seaborn pode ser uma opção melhor em algumas situações:

Estilo visual: O Seaborn vem com um estilo padrão que é mais atrativo e profissional do que o Matplotlib. Isso pode ser útil quando você precisa criar visualizações para apresentações ou relatórios.

Plotagem avançada: O Seaborn fornece suporte a vários tipos de gráficos comuns, como gráficos de densidade, histogramas, gráficos de dispersão e muito mais. Além disso, é possível criar visualizações mais avançadas, como gráficos de correlação, gráficos de matriz e mapas de calor.

Análise de dados: O Seaborn inclui recursos para análise exploratória de dados, como gráficos de distribuição e testes estatísticos. Isso pode ser útil para entender rapidamente a estrutura de seus dados e identificar tendências ou padrões.

A biblioteca Seaborn pode ser uma opção melhor do que o Matplotlib quando você precisa de visualizações mais atrativas e com recursos avançados para análise exploratória de dados. No entanto, é importante notar que o Matplotlib ainda é uma biblioteca poderosa e versátil, que pode ser usada em combinação com o Seaborn.

Vamos ver algumas ilustrações de código!

Gráfico de dispersão:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Carregar o dataset de exemplo iris
iris = sns.load_dataset("iris")

# Criar um gráfico de distribuição usando o atributo "sepal_width"
sns.distplot(iris["sepal_width"])

# Mostrar o gráfico
plt.show()
```

Resultado:

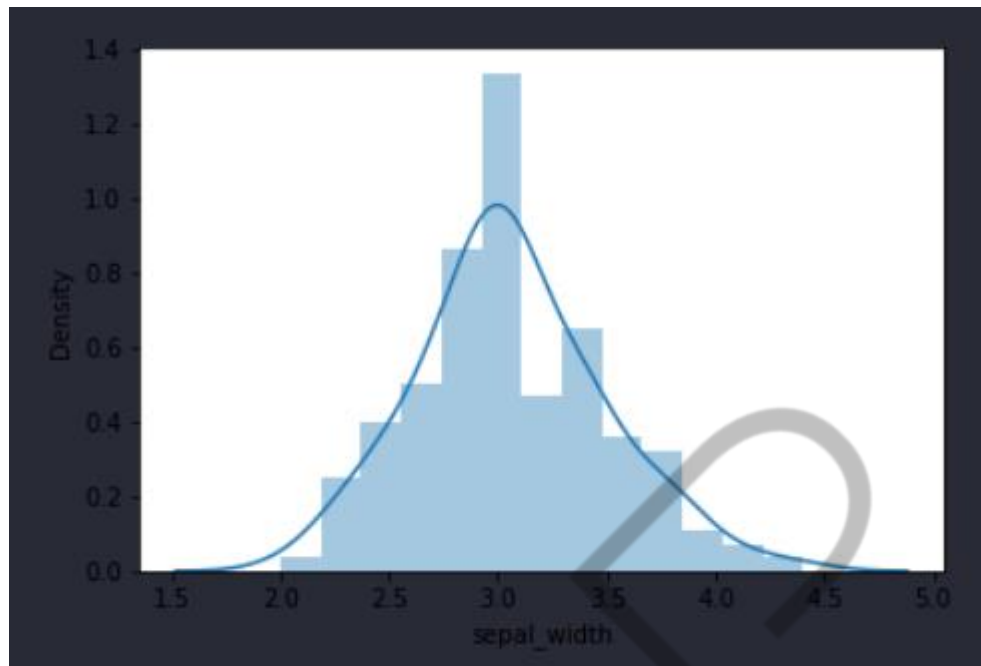


Figura 3 – Gráfico de dispersão
Fonte: Elaborado pelo autor (2023)

Gráfico de Regressão Linear:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Carregar o dataset de exemplo tips
tips = sns.load_dataset("tips")

# Criar um gráfico de regressão linear usando "total_bill" como variável
independente e "tip" como variável dependente
sns.regplot(x="total_bill", y="tip", data=tips)

# Mostrar o gráfico
plt.show()
```

Resultado:

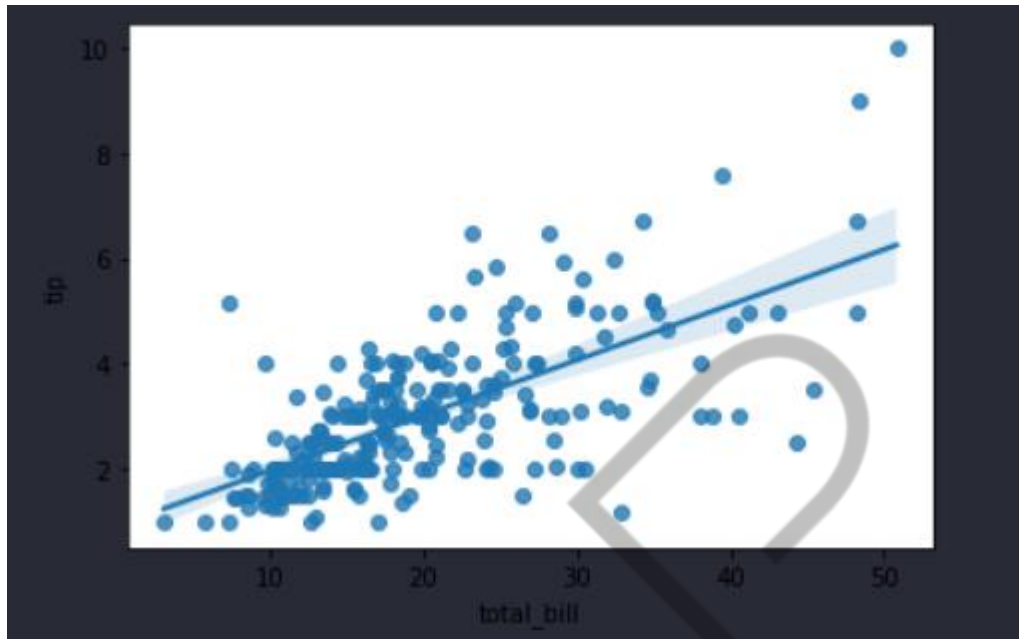


Figura 4 – Regressão Linear
Fonte: Elaborado pelo autor (2023)

Gráfico de Barra:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Carregar o dataset de exemplo tips
tips = sns.load_dataset("tips")

# Criar um gráfico de barra agrupado por "smoker" e "day"
sns.barplot(x="day", y="total_bill", hue="smoker", data=tips)

# Mostrar o gráfico
plt.show()
```


Resultado:

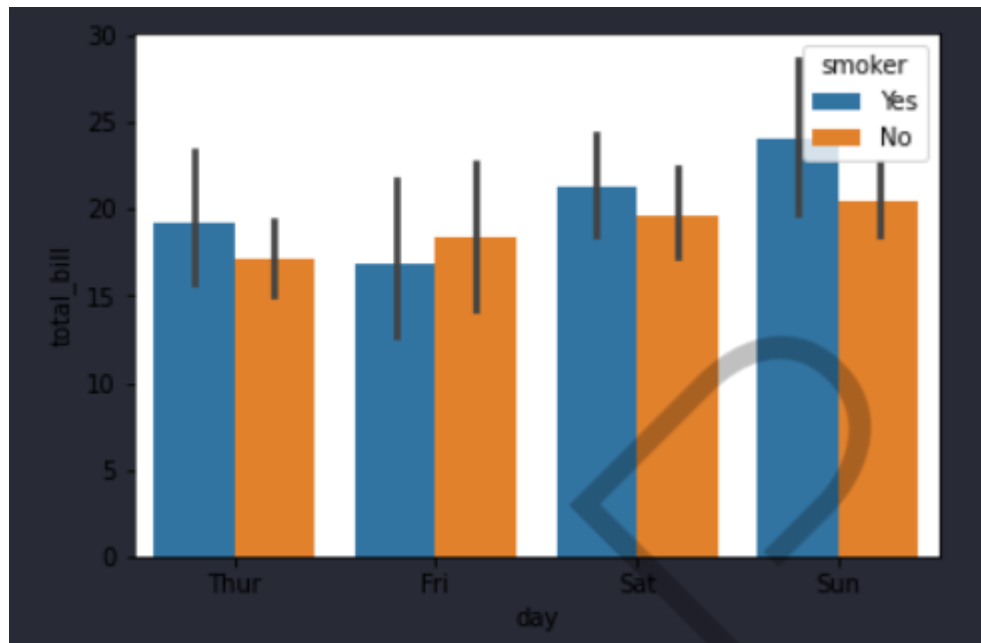


Figura 5 – Gráfico de Barra
Fonte: Elaborado pelo autor (2023)

HANDS ON

Agora chegou o momento de ver, na prática, como começar a importação dos dados e trabalhar com eles via programação. Não se limite apenas ao código disponibilizado no hands on, é sempre bom procurar a documentação das bibliotecas e explorar novas funcionalidades, combinado?!

O link para os notebooks utilizados estão em nosso git <<https://github.com/alura-tech/pos-datascience-introducao-a-visualizacao/archive/refs/heads/aula2.zip>> para que você possa consultar e seguir com os seus estudos.

Examine atentamente toda a possibilidade na documentação oficial do Seaborn e verifique os tipos de gráficos ideais para os seus problemas.

O QUE VOCÊ VIU NESTA AULA?

Como construir os primeiros gráficos com Seaborn; como fazer comparações dos resultados de forma proporcional à população; como analisar gráficos como de gastos_por_habitantes X população e gastos X população, comparando suas diferenças e gerando hipóteses e o que são e como usar Type Hints.

IMPORTANTE: não esqueça de acessar o documento que contém o desafio da disciplina, para que assim você possa aprimorar os seus conhecimentos de maneira sólida, além, é claro, de interagir no Discord e participar das lives!

REFERÊNCIAS

DOCUMENTAÇÃO PANDAS. <<https://pandas.pydata.org/>>. Acesso em: 08 fev. 2023.

DOCUMENTAÇÃO SEABORN. <<https://seaborn.pydata.org/>>. Acesso em 08 fev. 2023.

GOOGLE COLAB. <<https://colab.research.google.com/>>. Acesso em: 08 fev. 2023.

EMENDAS

PALAVRAS-CHAVE

Python. Seaborn. Dataframe.

EMAP

The background is a dark blue field filled with numerous small, light blue dots, resembling a starry sky. Overlaid on this are several large, flowing, wavy lines in shades of teal, blue, and yellow. These lines create a sense of movement and depth. Scattered throughout the composition are various geometric shapes: a large circle containing the number '7' in the upper center, a smaller circle on the left, a cross-like shape in the lower left, a small circle below it, and a hexagon in the bottom right corner. The overall aesthetic is futuristic and technological.

POSTECH