



Visualização de dados

Higor Alexandre Duarte Mascarenhas



Visualização

- Um dos primeiros artifícios de comunicação da humanidade;
- “Uma imagem vale mais que mil palavras”;
- A ideia principal de uma visualização de dados por meio de imagens, gráficos ou cartografia é simplificar conteúdos de forma a permitir a compreensão de uma ideia geral do todo;
- Visualização envolve a arte de facilitar a percepção do todo.



Visualização

- A visualização é um recurso que facilita o processo de compreensão e de tomada de decisão em praticamente todas as áreas do conhecimento.
- É possível processar uma imagem de 250 megapixels e tomar uma decisão a partir de uma imagem nunca vista anteriormente em poucos segundos



Visualização

- Corresponde a um dos pontos mais importantes do Data Science. Se todos os procedimentos foram bem feitos, mas a entrega não atende as necessidades, tudo que foi feito antes perde o sentido.
- É de extrema importância entregar bem uma análise de dados para quem vai consumi-la.
- Importante saber explorar todos os recursos possíveis no processo de visualização, de forma a demonstrar com clareza o resultado da análise, impedindo que sejam tomadas decisões equivocadas por má interpretação da visualização.



Visualização

- Dominar o processo de visualização é uma arte, há muitas estratégias de se fazer isso.
- Entender para que serve cada tipo de gráfico, e a melhor forma de visualizar os dados, é parte de um diferencial muito grande dos profissionais de Data Science.
- Com o site abaixo, é possível explorar para que cada tipo de gráfico pode ser utilizado.
 - <https://datavizproject.com/>



Bibliotecas mais usadas - Matplotlib

- Biblioteca mais comum para visualização de dados em Python.
- Possui diversas variações de visualizações e é usada como base para diversas outras bibliotecas disponíveis.
- Através do site <https://matplotlib.org/> você consegue ter a variedade dos gráficos presentes na biblioteca, bem como ver o código para efetuar cada um dos gráficos.
- Como carregar a biblioteca:
 - `import matplotlib.pyplot as plt`



Matplotlib

- Estrutura do Matplotlib

- 1. Geração de uma tela vazia (gráfico em branco)
- 2. Formatação da tela (eixos, título, etc)
- 3. Preenchimento da tela com dados - Aqui escolhemos qual o tipo de gráfico iremos usar
- 4. Formatação dos dados (legenda, etc)




Bibliotecas mais usadas - Seaborn

- Biblioteca construída baseada em Matplotlib.
- Foi projetada para funcionar em harmonia com os objetos do tipo Dataframe Pandas.
- Possui variações mais elaboradas das visualizações e detalhes estéticos mais elaborados;
- Através do site <https://seaborn.pydata.org> você consegue ter a variedade dos gráficos presentes na biblioteca, bem como ver o código para efetuar cada um dos gráficos.
- Como carregar a biblioteca:
 - `Import seaborn as sns`



Matplotlib x Seaborn

- A biblioteca Matplotlib por ser a base para grande parte das outras bibliotecas, é uma das mais completas, entretanto, muitas das suas configurações devem ser efetuadas manualmente, com linhas de programação.
- A biblioteca Seaborn já se apresenta com um conjunto de parâmetros que se autoconfiguram, para se adaptar ao seu conjunto de dados, e muitas das informações não precisam ser explicitadas, elas já estão configuradas.



Apache Superset (em alguns casos, pode utilizar Python)

- <https://superset.apache.org/>
- Ex:
- <http://educa-ibict.fcc.org.br:8088/superset/dashboard/12/>
- Outro exemplo:
- <https://ocid.ibict.br/dados-pt/?omeka-set=15462&omeka-group=Dados>



Sem Python



Flourish



➤ <https://app.flourish.studio/projects>





Charts

➤ <https://charts.livegap.com/?lan=pt#TypesofCharts>



■ Demonstração