

BIG DATA



LÁBDATA



FUNDAÇÃO
INSTITUTO DE
ADMINISTRAÇÃO

Introdução ao Big Data

Tema da Aula: **Visualização dos dados com Python**

Coordenação:

Prof. Dr. Adolpho Walter
Pimazzi Canton

Profa. Dra. Alessandra de
Ávila Montini

Prof.: **Dino Magri**

Data: **05 de Dezembro de 2018**

- Contatos:

- E-mail: professor.dinomagri@gmail.com
- Twitter: https://twitter.com/prof_dinomagri
- LinkedIn: <http://www.linkedin.com/in/dinomagri>
- Site: <http://www.dinomagri.com>

Coordenação:

Prof. Dr. Adolpho Walter
Pimazzi Canton

Profa. Dra. Alessandra de
Ávila Montini

Currículo

- **(2014-Presente)** – Professor no curso de Extensão, Pós e MBA na Fundação Instituto de Administração (FIA) – www.fia.com.br
- **(2013-Presente)** – Pesquisa e Desenvolvimento no Laboratório de Arquitetura e Redes de Computadores (LARC) na Universidade de São Paulo – www.larc.usp.br
- **(2013)** – Professor no MBA em Desenvolvimento de Inovações Tecnológicas para WEB na IMED Passo Fundo – RS – www.imed.edu.br
- **(2012)** – Bacharel em Ciência da Computação pela Universidade do Estado de Santa Catarina (UDESC) – www.cct.udesc.br
- **(2009/2010)** – Pesquisador e Desenvolvedor no Centro de Computação Gráfica – Guimarães – Portugal – www.ccg.pt
- **Lattes:** <http://lattes.cnpq.br/5673884504184733>

Material das aulas

- Material das aulas estão disponíveis em:
 - <https://urls.dinomagri.com/posmba-turma9>
 - **Senha:** turma9
 - **Data Expiração:** 31-12-2018

Material das aulas

- Caso esteja utilizando seu próprio computador, realize o download de todos os arquivos e salve na **Área de Trabalho** para facilitar o acesso.
 - Lembre-se de instalar os softwares necessários conforme descrito no documento de Instalação (**InstalaçãoPython3v1.1.pdf**).
- Nos computadores da FIA os arquivos já estão disponíveis, bem como a instalação dos softwares necessários.

Conteúdo da Aula

- Objetivo
- Visualização dos dados com Python
- Exercícios
- Material Extra
- Referências

Conteúdo da Aula

- **Objetivo**
- Visualização dos dados com Python
- Exercícios
- Material Extra
- Referências

Objetivo

- Objetivo dessa aula é introduzir as **bibliotecas de visualização de dados** disponíveis no **Python**.

Conteúdo da Aula

- Objetivo
- **Visualização dos dados com Python**
- Exercícios
- Material Extra
- Referências

Visualização dos dados com Python

- Python tem diversas bibliotecas de visualização de dados.
- Existem bibliotecas para propósitos específicos, como rastrear o movimento dos olhos (**GazeParser**), até visualização em tempo real de redes neurais (**pastalog**).
- Uma lista das bibliotecas de visualização existente no PyPI - <https://pypi.python.org/pypi?:action=browse&c=399>

Visualização dos dados com Python

- Porém existem bibliotecas genéricas que podem ser utilizadas para diversos fins, como:
 - Matplotlib
 - Seaborn
 - Bokeh

matplotlib

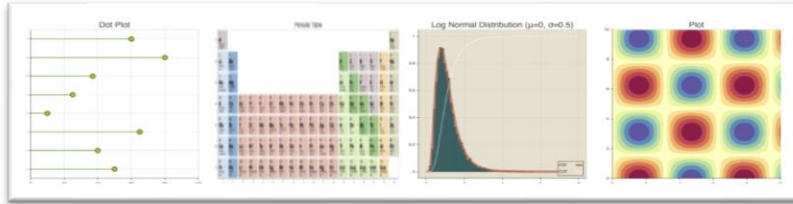
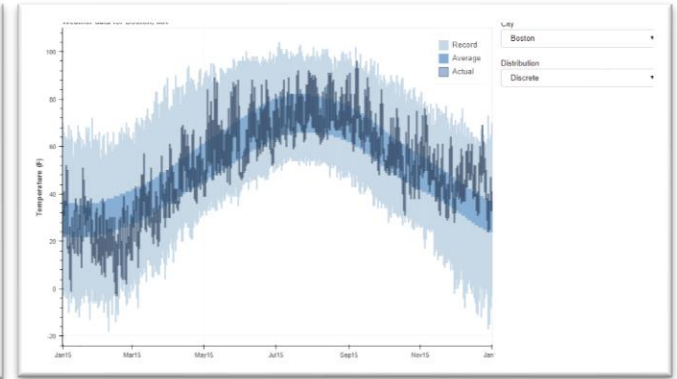
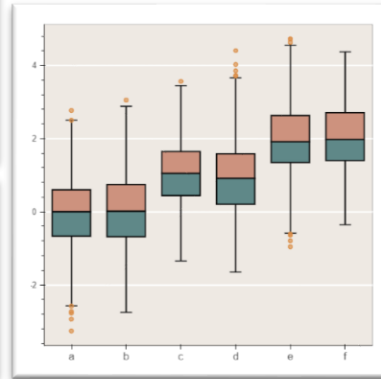
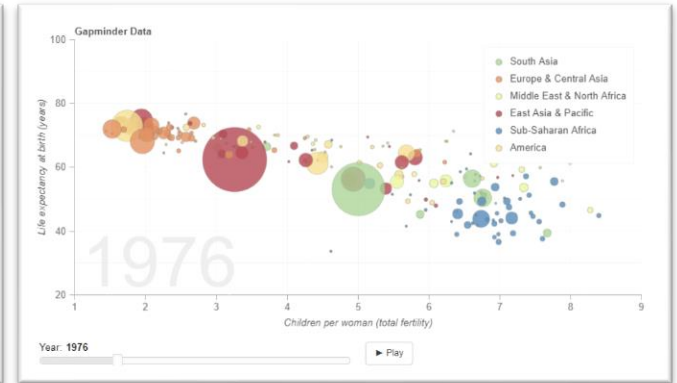
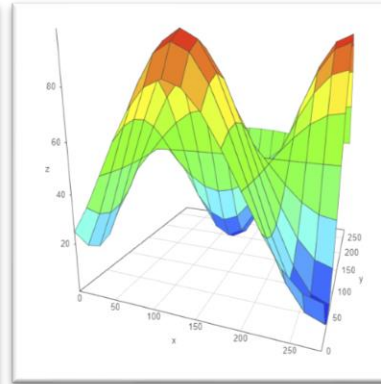
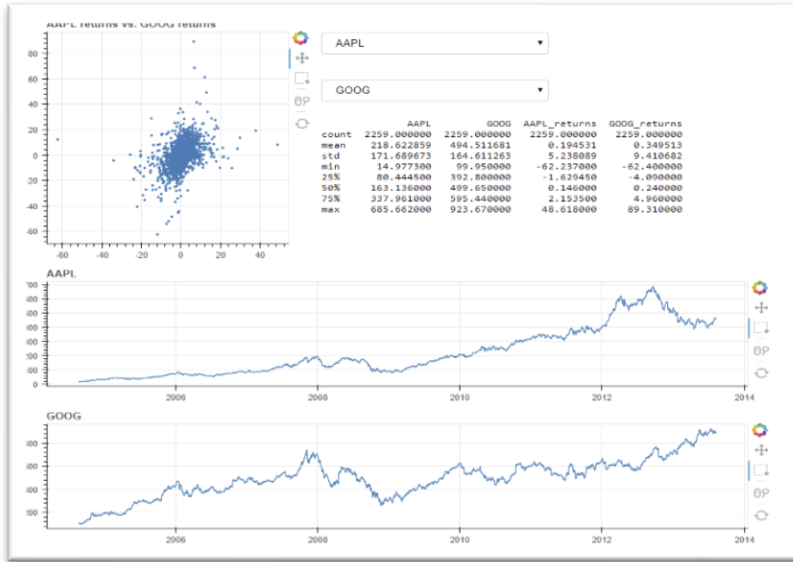
- É uma das bibliotecas de visualização mais antiga do Python (2002), porém muito utilizada ainda.
- Funciona muito bem para realizarmos análises iniciais no dados, ter uma noção do que temos. Porém ela não é muito útil para a criação de gráficos com qualidade de publicação rápida e fácil.
- Ela é muito poderosa, porém complexa!
- Um dos pontos mais criticados é o estilo padrão, que passa uma sensação dos anos 90. Porém a versão 2.0 terá algumas melhorias nesse sentido:
http://matplotlib.org/style_changes.html
- Galeria de exemplos: <http://matplotlib.org/examples/index.html>

 Abra o arquivo "**aula7-parte4-extra-matplotlib.ipynb**"

bokeh

- **Bokeh** é uma biblioteca de visualização interativa que roda nos principais navegadores.
- Foi criado pela Continuum Analytics.
- Possibilita construir de forma simples e rápida gráficos elegantes, concisos e de alto desempenho em grandes volumes de dados.

bokeh



bokeh

- As funcionalidades da biblioteca estão disponíveis em dois níveis de interfaces para o usuário:
 - Nível alto – **bokeh.plotting** permite compor os efeitos visuais.
 - Nível baixo – **bokeh.models** fornece maior flexibilidade para os desenvolvedores de aplicações.

bokeh

```
<html>  
  <script>...</script>  
  <div>  
    <canvas>...</canvas>  
  </div>  
</html>
```



BokehJS

json

Bokeh
Python

rbokeh
R

bokeh-lua
lua

bokeh.jl
julia

Fonte: <https://goo.gl/esSgAo>

bokeh

- Para instalar:
 - `pip3 install bokeh`
 - `six` – utilitários de compatibilidade entre Python 2 e 3
 - `Requests` – Requisições HTTP
 - `Numpy` – ☺
 - `PyYAML` – Analisador de YAML
 - `python-dateutil` – Extensões para o módulo padrão `datetime` do Python
 - `Jinja2` – Mecanismo rápido e simples para modelos (template)
 - `Tornado` – Framework web e biblioteca de rede assíncrona
 - `MarkupSafe` – Implementa marcações XML/HTML/XHTML

bokeh

- Vamos utilizar a interface `bokeh.plotting`, para criar alguns gráficos. Para isso, existem alguns componentes principais:
 - A função **figure** para criar um novo objeto de plotagem.
 - Utilizar as funções **output_file** ou **output_notebook** para determinar a saída.
 - Executar as funções **show** e **save** para mostrar ou salvar os gráficos gerados.

bokeh

- `figure` – Contém diversos métodos para serem utilizados na criação de gráficos.
- O formato de saída pode ser:
 - `output_file` – gera um HTML simples com o código da visualização
 - `output_notebook` – mostra a visualização no próprio notebook
- `show` – método utilizado para mostrar os gráficos criados.
- `save` – salva um arquivo HTML com os respectivos gráficos do documento atual.



Abra o arquivo "**aula7-parte1-10minutos.ipynb**"

Conteúdo da Aula

- Objetivo
- Visualização dos dados com Python
- **Exercícios**
- Material Extra
- Referências

Exercício 1

- Utilizando o arquivo `tweets_10min.csv` iremos criar um gráfico de barras com os 10 usuários mais populares.



Abra o arquivo "**aula7-parte2-10mais.ipynb**"

Exercício 2

- Utilizando o arquivo `tweets_10min.csv` iremos utilizar a latitude e longitude para plotar no Google Maps a localização dos usuários.
- Também iremos adicionar informações complementares (localização e usuário) em cada ponto gerado no gráfico.
- Para isso, temos que realizar 5 passos:
 1. Carregar o conjunto de dados
 2. Criar um mapa
 3. Gerar a API para acessar o Google Maps
 4. Adicionar os tweets no mapa (lat, long)
 5. Adicionar as informações complementares



Abra o arquivo "[aula7-parte3-google-maps.ipynb](#)"

Conteúdo da Aula

- Objetivo
- Visualização dos dados com Python
- Exercícios
- **Material Extra**
- Referências

Parte 5 – Extra - Visualização de dados Geográficos

- Esse exercício extra apresenta uma forma de adicionar dados em um gráfico de outra fonte de mapa.
- Pode ser:
 - OpenStreetMap - <https://www.openstreetmap.org/>
 - Stamen - <http://maps.stamen.com>
 - MapQuest - <http://maps.stamen.com>
 - Google Maps - <https://www.google.com.br/maps>
 - **Algun servidor próprio**
- Foi utilizado o mesmo conjunto de dados (`tweets_10min.csv`).
- O processo é relativamente o mesmo que o do Google Maps, com a necessidade de converter a latitude e longitude para o formato Web Mercator para que os pontos sejam adicionados no mapa corretamente.

 Abra o arquivo "**aula7-parte5-extra-qualquer-fonte-mapa.ipynb**"

Parte 6 – Extra – Criando o próprio mapa

- Nesse exercício extra iremos utilizar uma função para criar o Mapa utilizando as fronteiras dos países do mundo.
- Serão utilizados os dados dos tweets recuperados anteriormente.
- Para isso, temos **4** passos a serem realizados:
 1. Criar um mapa
 2. Separar as colunas que queremos (date, lat, long, location, user, text)
 3. Adicionar os tweets no mapa (lat, long)
 4. Adicionar as informações complementares.

Parte 6 – Extra – Criando o próprio mapa

- Passo 1 – Criar um mapa
 - Iremos utilizar uma função pronta que está disponível no livro **Spark for Python Developers, Amit Nandi, PACKT, 2015**.
 - Essa função pega as coordenadas de fronteiras de todos os países disponíveis em <https://github.com/hargup/WIGI-website> e cria a estrutura necessária (**ColumnDataSource**) para que possamos adicionar os pontos de latitude e longitude.

Parte 6 – Extra – Criando o próprio mapa

- Passo 2 – Separar os dados do Twitter.
 - Iremos utilizar a biblioteca pandas para separar os dados que queremos bem como realizar pequenos ajustes.
- Passo 3 – Adicionar os Tweets no Mapa
 - Utilizando o DataFrame criado, iremos gerar os pontos (scatter) no mapa criado anteriormente.

Parte 6 – Extra – Criando o próprio mapa

- Passo 4 - Adicionar as informações complementares
 - Usuário
 - Local



Abra o arquivo "**aula7-parte5-extra-criando-o-proprio-mapa.ipynb**"

Conteúdo da Aula

- Objetivo
- Visualização dos dados com Python
- Exercícios
- Material Extra
- **Referências**

Referências Bibliográficas

- **Learning Python Data Visualization** – Shai Vaingast – Apress, 2014.
- **Beginning Python Visualization** – Shai Vaingast – Apress, 2014.
- **Mastering pandas** – Femi Anthony – Packt Publishing, 2015.
- Referência Bokeh -
<http://bokeh.pydata.org/en/latest/docs/reference.html>

Referências Bibliográficas

- **Data Science from Scratch** – Joel Grus – O'Reilly, 2015.
- **Python for Data Analysis** – Wes McKinney – USA: O'Reilly, 2013.
- As referências de links utilizados podem ser visualizados em <http://urls.dinomagri.com/refs>