



# Python para Data Science

---

Prof. Dr. Jorge Zavaleta

[zavaleta.jorge@gmail.com](mailto:zavaleta.jorge@gmail.com)



# Sobre – Jorge Zavaleta



Curriculo  
**Lattes**

- Doutor em Engenharia de Sistemas e Computação - UFRJ
- Mestre em Ciência da Computação – UFRGS
- Licenciado em Matemática



DIRETÓRIO DOS GRUPOS  
DE PESQUISA NO BRASIL  
**Lattes**



- Áreas de Interesse e Pesquisa:
  - Inteligência Artificial
  - *Machine Learning*
  - *Deep Learning*
  - Ciência de Dados Aplicados a: Educação, Saúde e Meteorologia.
  - Linguagem Java, R e Python.

# Agenda

- O que Data Science?
- Configuração do Ambiente de Trabalho
- Jupyter Notebook
- Python Básico
- Análise de Dados usando Numpy
- Análise de Dados usando pandas
- Análise de Dados – Exercícios pandas
- Visualização de Dados – Matplotlib
- Visualização de Dados – Seaborn
- Visualização de Dados – Visualização de dados Pandas

# O que é Data Science?

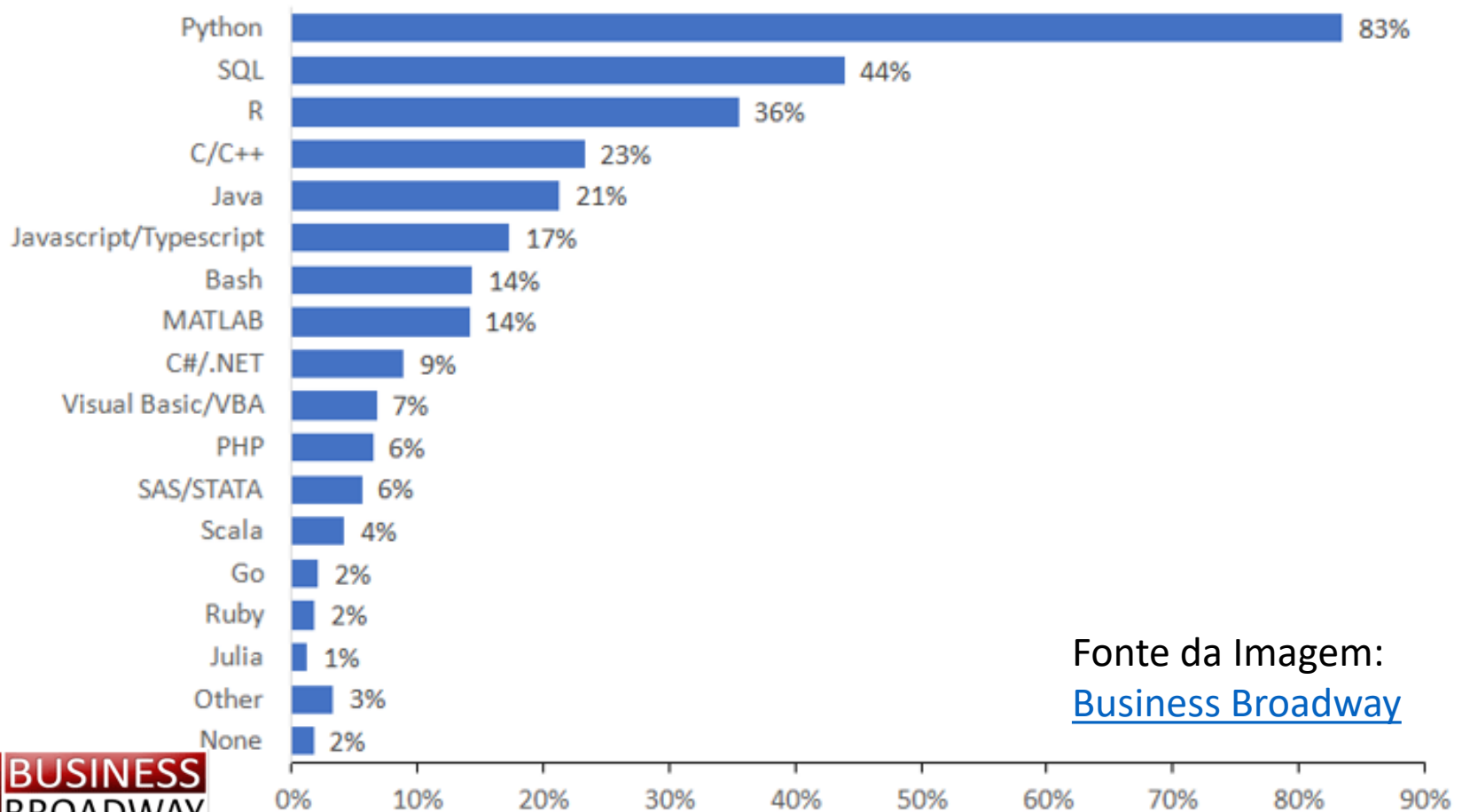
- **Ciência de dados (Data Science - DS)** é o estudo dos dados.
- Ciência de Dados, é o atual termo para a ciência que **analisa dados**, combinando a **estatística** com **aprendizado de máquina/mineração de dados** e **tecnologias de base de dados**, para responder ao desafio que o **Big Data** apresenta.



Fonte da Imagem:

<http://www.kiwidatascience.com/>

# Linguagens de Programação



Fonte da Imagem:  
[Business Broadway](#)

# Ferramentas para DS

**R**

usado para  
análise  
estatística

**BigML**

ML para DS de fácil  
uso

**Tableau**

simplifica dados  
brutos em formato  
compreensível

**SAS**

facilita o análise,  
reportes, e  
mineração de dados

**MatLab**

usado para  
processamento de  
dados

**Ggplot2**

usado para  
Visualização gráfica

**Matplotlib**

geração de gráficos  
com análise de  
dados

**Jupyter**

ajuda a  
transformar e  
visualização de  
dados

# Configuração do Ambiente

- O curso será feito em **JUPYTER**
- Pode ser usado qualquer ambiente de programação
- As aulas estão disponíveis no [GitHub](#)
- [https://github.com/zavaleta/Python\\_Data\\_Science\\_UFF\\_2020](https://github.com/zavaleta/Python_Data_Science_UFF_2020)
- Será usado o Python 3 através do ambiente ANACONDA

# Configuração do Ambiente

- Instalação para Usuários **EXPERIENTES**:
- O Python poderá ser instalado usando “*Python package manager*”, usando o comando, **pip**, ao invés do ambiente Anaconda.
- Abrir o prompt de comandos e usar:

**pip3 install jupyter**  
ou  
**pip install jupyter**



# Configuração do Ambiente

- Instalação para Usuários **NOVOS**:
- Para usuários **novos** recomenda-se instalar o ambiente [Anaconda](#).
  - ✓ <https://www.anaconda.com/distribution/>
- Acessar o site do **JUPYTER** e seguir o roteiro para a instalação
  - <https://jupyter.org/>



# Python Básico

- **Variáveis e tipos de dados**

- Atribuição de variáveis
- Cálculos com variáveis
- Tipos e Conversão de tipos

- **Strings**

- Operações e métodos das strings

- **Listas**

- Selecionar elementos de uma lista
- Operações com listas
- Métodos das listas

## Libraries

### Import libraries

```
>>> import numpy
>>> import numpy as np
```

### Selective import

```
>>> from math import pi
```

pandas   
Data analysis

 learn  
Machine learning

 NumPy  
Scientific computing

 matplotlib  
2D plotting

## Install Python

  
**ANACONDA**  
Leading open data science platform  
powered by Python

  
**spyder**  
Free IDE that is included  
with Anaconda

  
**jupyter**  
Create and share  
documents with live code,  
visualizations, text, ...

# Jupyter

