

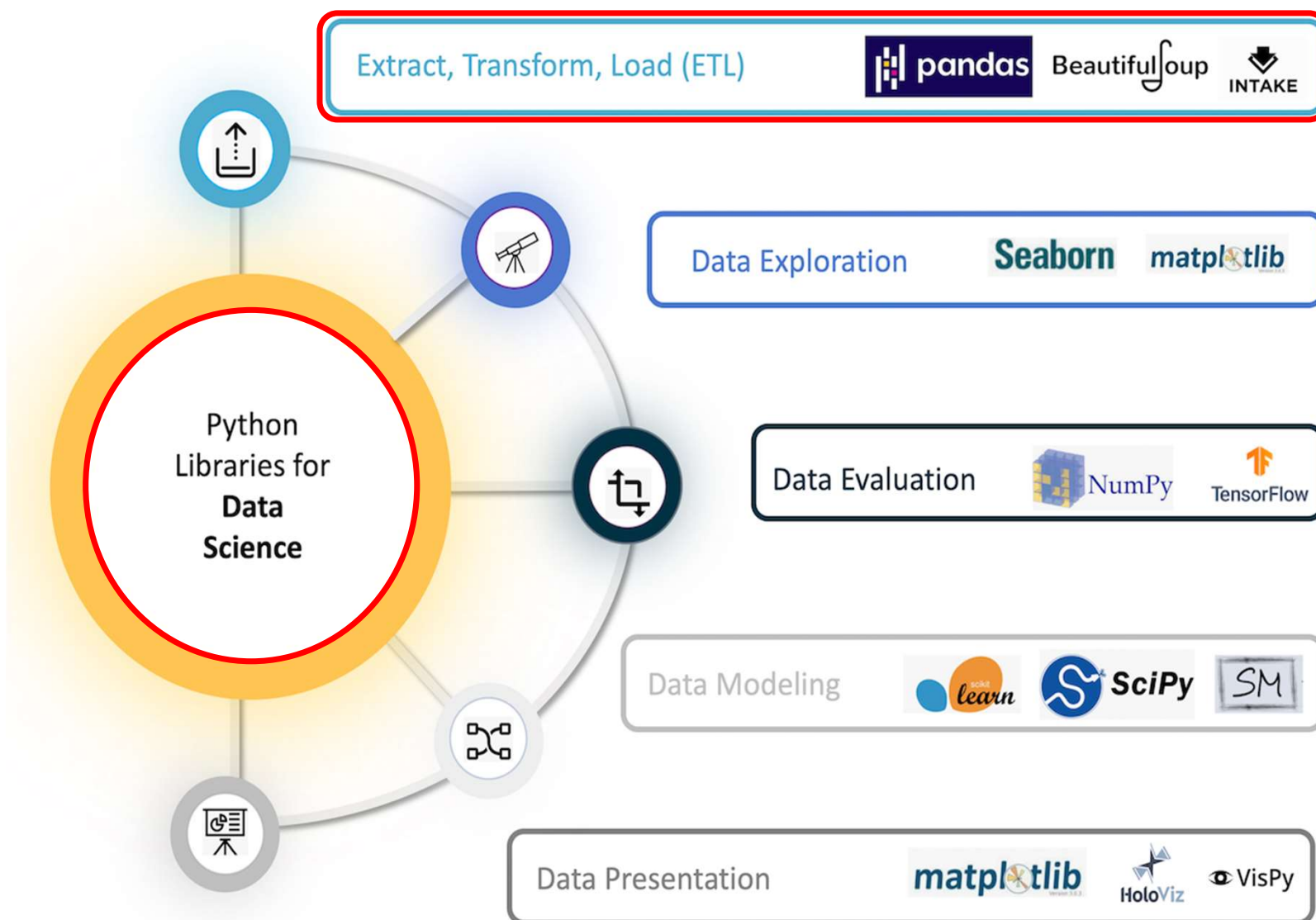


Prof. Dr. Jorge Zavaleta

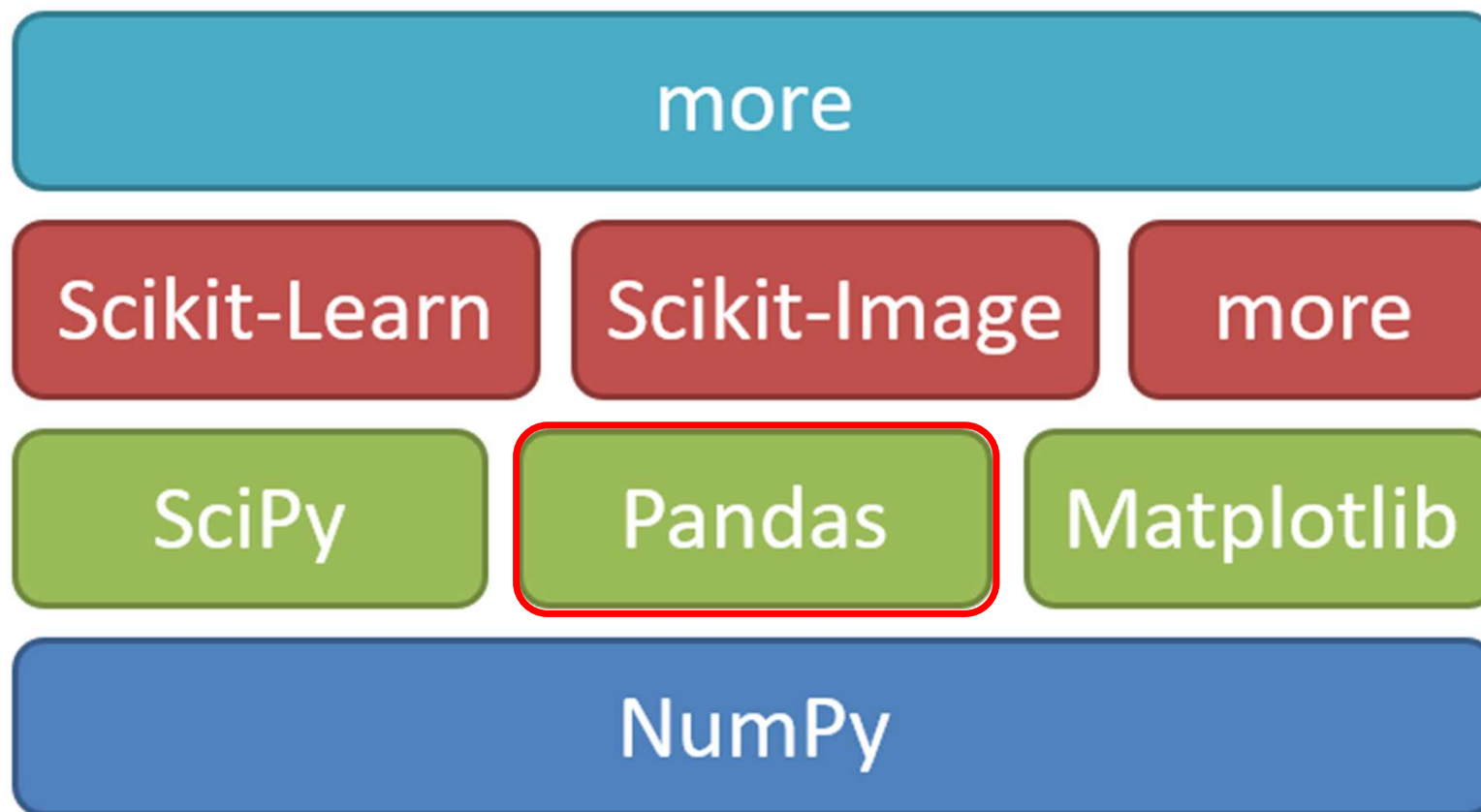
Departamento de Ciências Ambientais (DCA)

Universidade Federal Rural do Rio de Janeiro (UFRRJ)

Pesquisador de Pós-doutorado (PDJ/CNPq)



A pilha do Python



O que é Pandas?

- Pandas é uma biblioteca para Ciência de Dados de código aberto (open source), construída sobre a linguagem Python, e que providencia uma abordagem rápida e flexível, com estruturas robustas para se trabalhar com dados relacionais (ou rotulados) de maneira simples e intuitiva.
- É uma ferramenta de análise e manipulação de dados, rápida, poderosa, flexível e fácil de usar.

Pandas - Instalação

- A maneira mais fácil e simples de instalar, segundo a própria documentação do Pandas, é instalando a distribuição do Anaconda.
 - *conda install conda-forge::pandas*
- <https://www.python.org/>
 - `pip install pandas`

Pandas - estruturas de dados

- As principais estruturas de dados que o pandas fornece são:
 - Series
 - DataFrames
- Geração de estatísticas descritivas sobre dados
- Limpeza de dados usando funções pandas integradas
- Operações de dados frequentes para subconjuntos, filtragem, inserção, exclusão e agregação de dados
- Mesclagem de vários conjuntos de dados usando dataframes

Pandas – principais estruturas

Series

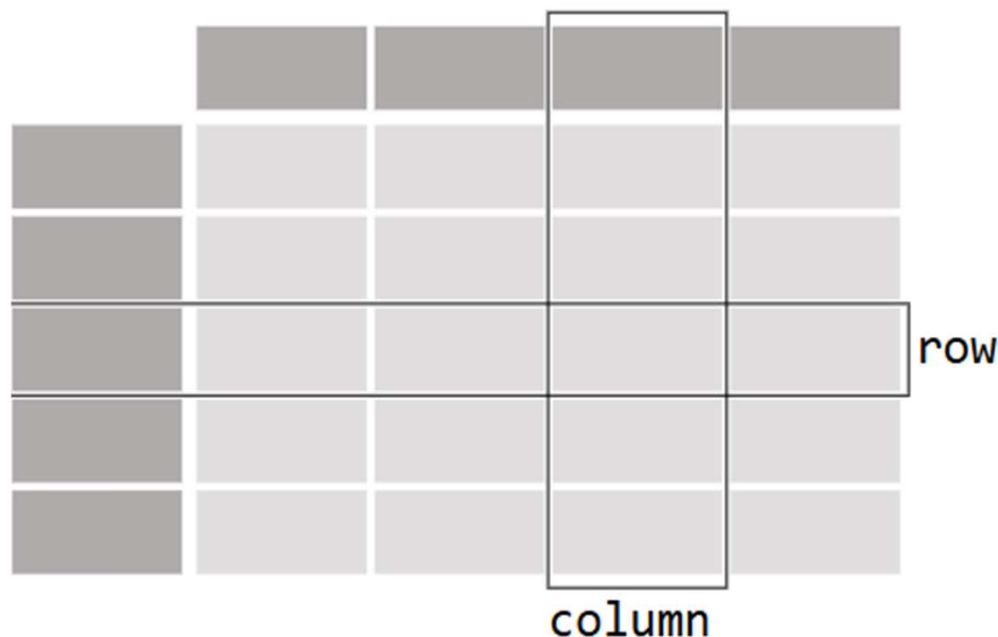
Series


```
# serie
import pandas as pd
idade = pd.Series([12,40,50,34,21],name='idade')
idade
```

```
0    12
1    40
2    50
3    34
4    21
Name: idade, dtype: int64
```

Pandas – principais estruturas

DataFrame

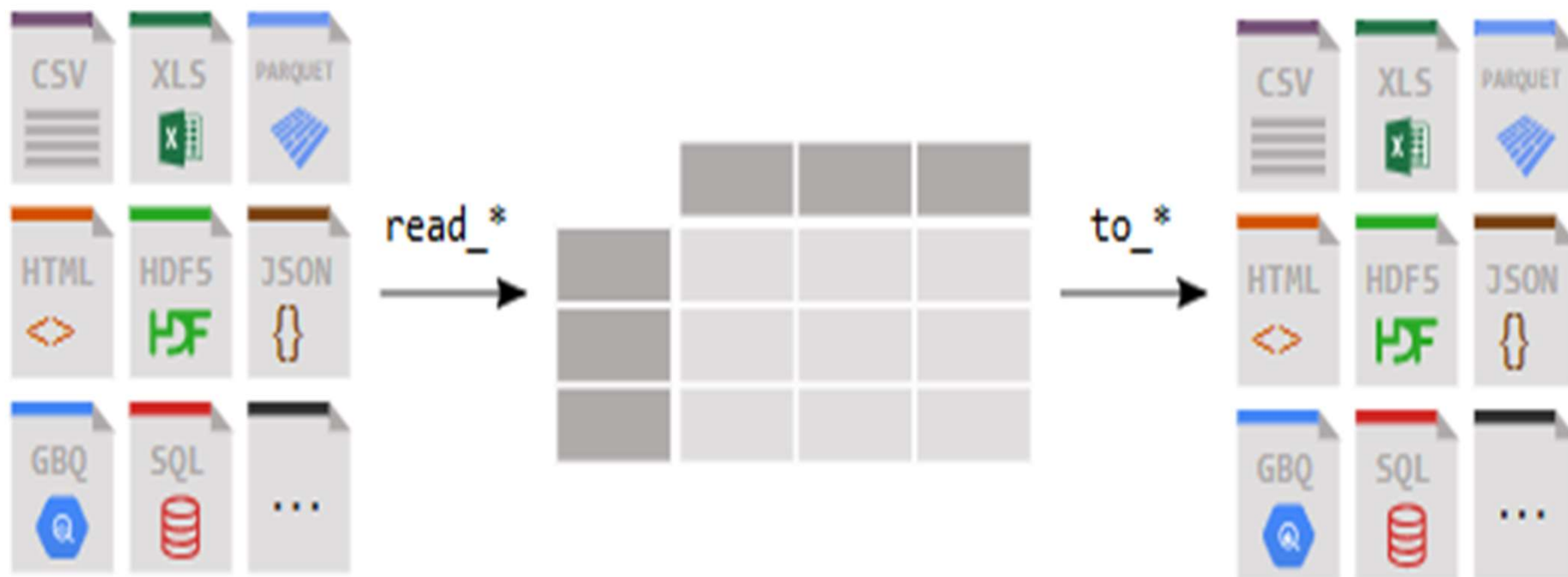


DataFrame

```
#dataframe
df = pd.DataFrame(
    {
        "Nome": [
            "Braund, Mr. Owen quipuzco",
            "Allen, Mr. William Guillermo",
            "Bonnell, Miss. Eliza",
        ],
        "Idade": [21, 35, 58],
        "Sexo": ["male", "male", "female"],
    }
)
df
```

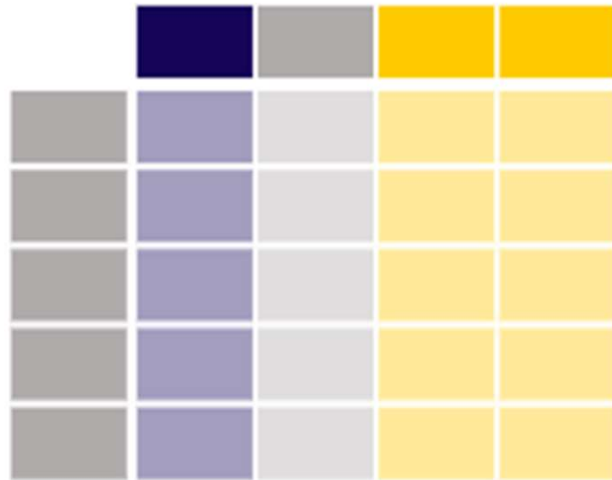
	Nome	Idade	Sexo
0	Braund, Mr. Owen quipuzco	21	male
1	Allen, Mr. William Guillermo	35	male
2	Bonnell, Miss. Eliza	58	female

Pandas - io

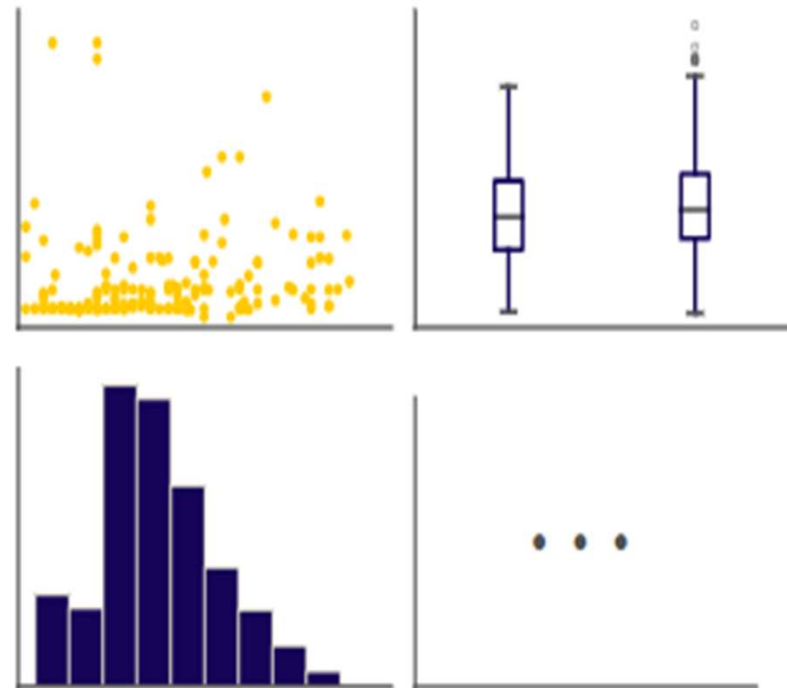


Fonte: https://pandas.pydata.org/docs/getting_started/intro_tutorials/02_read_write.html

Pandas - plots



`.plot.*`



Fonte: https://pandas.pydata.org/docs/getting_started/intro_tutorials/02_read_write.html

Pandas - métodos

Data Importing

```
pd.read_csv()  
pd.read_table()  
pd.read_excel()  
pd.read_sql()  
pd.read_json()  
pd.read_html()  
pd.DataFrame()  
pd.concat()  
pd.series()  
pd.date_range()
```

Data Cleaning

```
pd.fillna()  
pd.dropna()  
pd.sort_values()  
pd.apply()  
pd.groupby()  
pd.append()  
pd.join()  
pd.rename()  
pd.to_csv()  
pd.set_index()
```

Data Statistic

```
pd.head()  
pd.tail()  
pd.describe()  
pd.info()  
pd.mean()  
pd.median()  
pd.count()  
pd.std()  
pd.max()  
pd.min()
```

Pandas – alguns métodos

- `unique()` - retorna uma lista dos elementos únicos no dataset.
- `nunique()` – retorna o total dos valores distintos das colunas.
- `describe()` – retorna a estatísticas descritivas de tendência central.
- `sort_values()` – retorna valores das colunas ordenados – (de)crescente.
- `value_counts()` - retorna o valor da contagem para cada coluna.
- `isnull()` - descobre se há algum valor ausente no array ou dataset
- `fillna()` - preencher os valores faltantes com base no item especificado.
- `groupby()` – usada para agrupar usando alguma função específica.
- `map()` – mapear series e substituir valores por outro valor associado.
- `apply()` – ajuda a modificar os dados de acordo com uma condição.
- `pivot()` - remodela o dataset por valores de índice ou coluna

Bibliografia

<https://pandas.pydata.org/>

