

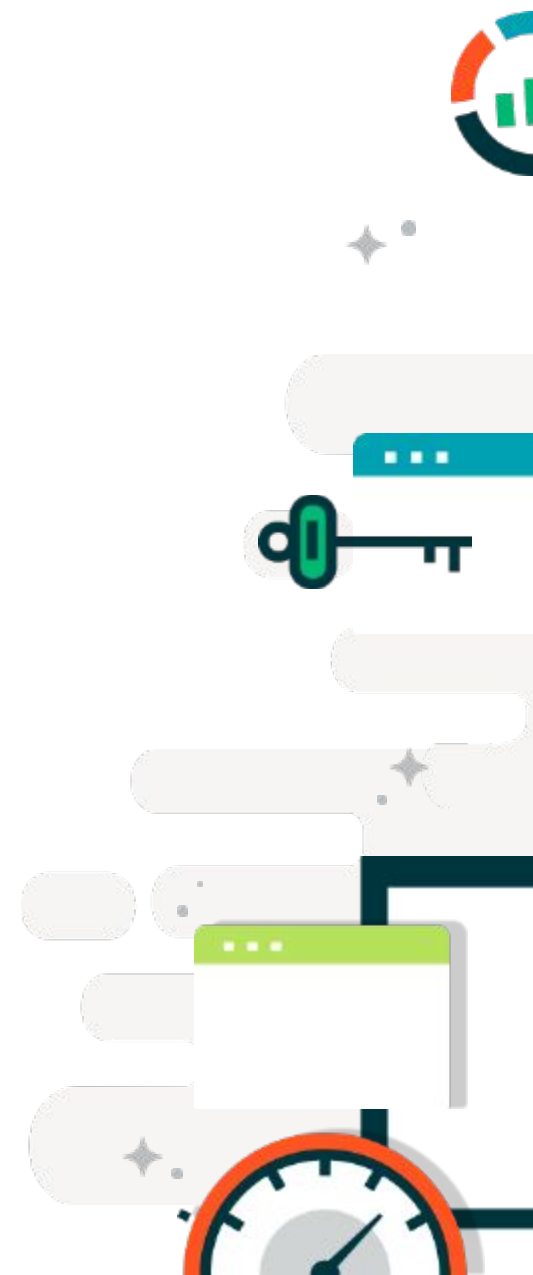
# Introdução à Análise de Dados com Pandas

Prof Raphael Campos



# Ementa da Unidade

- Pandas Series
- Pandas DataFrames
- Indexação
- Transformações com Apply
- Groupby
- Mesclando tabelas
- Visualização de Dados



# O que é Pandas?

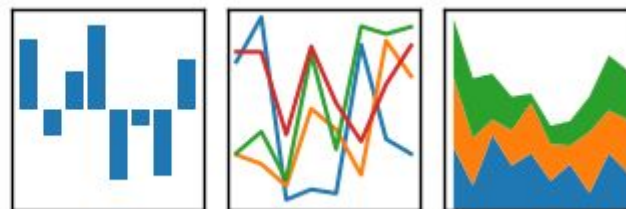


# Pandas

É uma biblioteca open source do eco-sistema Scipy para análise de dados.

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



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



# Series



# Series

É um objeto uni-dimensional similar ao ndarray do numpy ou a lista do python.

```
0          7
1    Heisenberg
2          3.14
3    Happy Eating!
dtype: object
```

```
A          7
B    Heisenberg
C          3.14
D    Happy Eating!
dtype: object
```

```
A          7
B          8
C          3.14
D          10
dtype: float
```



# DataFrames



# DataFrame

É uma estrutura de dados tabular composta de linhas e colunas, similar a uma planilha excel.

	losses	team	wins	year
0	5	Bears	11	2010
1	8	Bears	8	2011
2	6	Bears	10	2012
3	1	Packers	15	2011





# Visualização

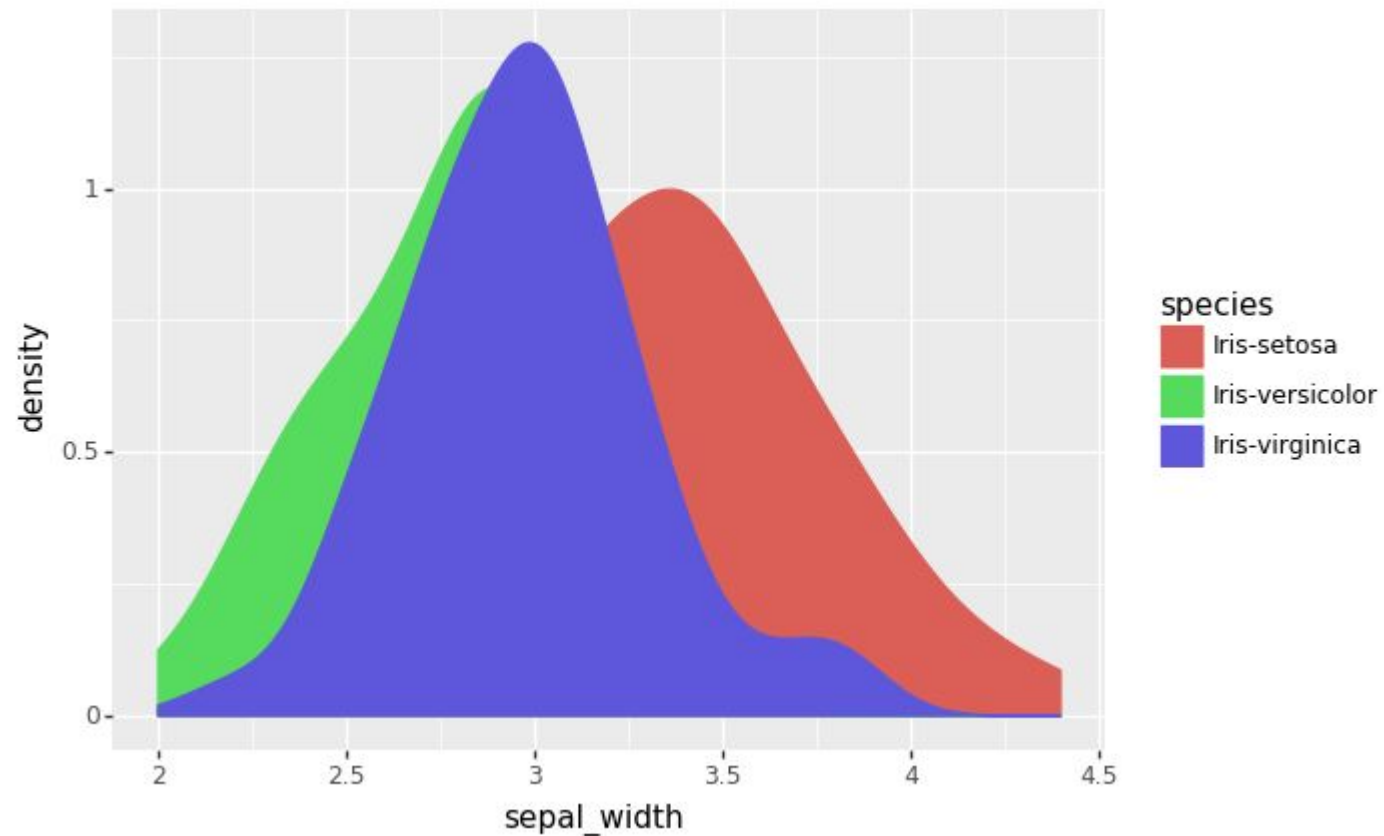


# Matplotlib

Matplotlib é uma biblioteca Python para traçar gráficos 2D que produz figuras de qualidade em vários formatos distintos



# Gráficos Estadísticos



# Seaborn

É uma biblioteca de visualização do Python baseada no Matplotlib.

Ela provê uma interface de alto-nível para traçar gráficos estatísticos.



# Plotnine

É uma implementação de uma gramática de gráficos em Python, baseada no ggplot2.





# Referências



# Referências

<https://github.com/raphaelcampos/introducao-linguagens-estatisticas/>

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

<https://github.com/has2k1/plotnine>

<https://matplotlib.org/>

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

