

# Python para Ciência de Dados

## Matplotlib Básico



Luiz Alberto

Ciência da Computação

May 13, 2019

# Matplotlib

## Definição:

Biblioteca python para plotagem de gráficos 2D (incluindo 3D) ([www.matplotlib.org](http://www.matplotlib.org)).

- Simplicidade de utilização
- Desenvolvimento gradual e interativo
- Grande controle sobre os elementos gráficos
- Exportação em formatos PNG, PDF, SVG e EPS

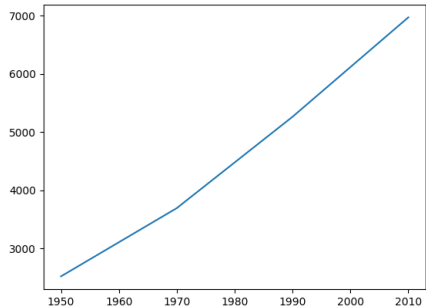
# Vizualização de Dados

- Muito importante na visualização de dados
  - Explorar os dados
  - Apresentar "insights"



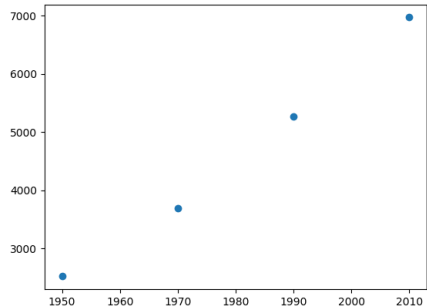
# Line plot

```
1 import matplotlib.pyplot as plt
2 anos = [1950, 1970, 1990, 2010]
3 pops = [2519, 3692, 5263
4         , 6972]
5
6 plt.plot(anos, pops)
7 #           X           Y
8
9 plt.show()
```



# Scatter plot

```
1 import matplotlib.pyplot as plt
2 anos = [1950, 1970, 1990, 2010]
3 pops = [2519, 3692, 5263
4         , 6972]
5
6 plt.scatter(anos, pops)
7 #           X           Y
8
9 plt.show()
```

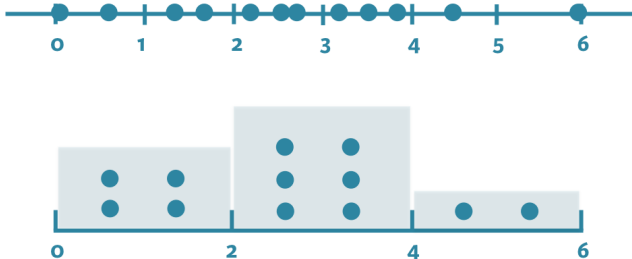


# Hora de colocar a mão na massa (1)

Salvar cada um dos exercícios a seguir em um arquivo separado.

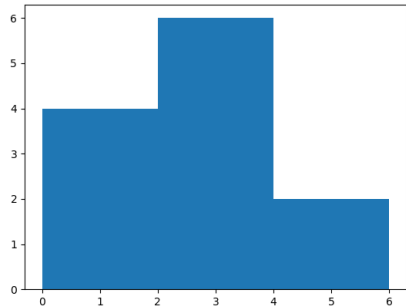
# Histogram

- Utilizado para explorar dados
- Fornece uma ideia da distribuição dos dados



# Histogram

```
1 import matplotlib.pyplot as plt
2 valores = [
3     0,    0.6, 1.4, 1.6
4     , 2.2, 2.5, 2.6, 3.2
5     , 3.5, 3.9, 4.2, 6
6 ]
7
8 plt.hist(valores, bins=3)
9 plt.show()
```



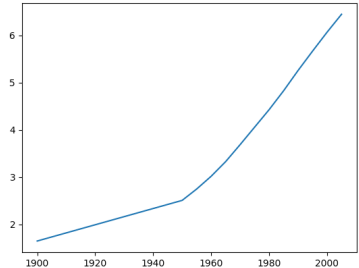


# Customização

- Existem muitas opções
  - Diferentes tipos de gráficos
  - Diversas customizações
- A escolha depende
  - Dados
  - Estória a ser contada

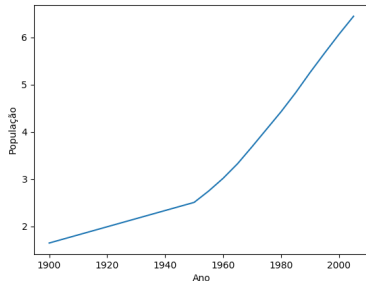
# Customização

```
1 import matplotlib.pyplot as plt
2 anos = [
3     1900, 1950, 1955, 1960, 1965,
4     1970, 1975, 1980, 1985, 1990,
5     1995, 2000, 2005
6 ]
7 pops = [
8     1.65, 2.51, 2.75, 3.02, 3.33,
9     3.69, 4.06, 4.43, 4.83, 5.26,
10    5.67, 6.07, 6.45
11 ]
12
13 plt.plot(anos, pops)
14 #           X           Y
15
16 plt.show()
```



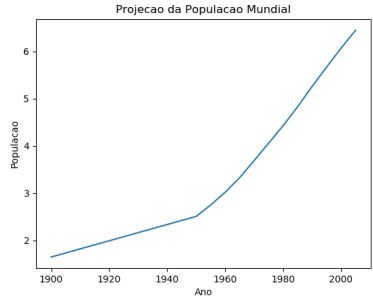
# Títulos dos eixos X e Y

```
1 import matplotlib.pyplot as plt
2 anos = [
3     1900, 1950, 1955, 1960, 1965,
4     1970, 1975, 1980, 1985, 1990,
5     1995, 2000, 2005
6 ]
7 pops = [
8     1.65, 2.51, 2.75, 3.02, 3.33,
9     3.69, 4.06, 4.43, 4.83, 5.26,
10    5.67, 6.07, 6.45
11 ]
12
13 plt.plot(anos, pops)
14 plt.xlabel('Ano')
15 plt.ylabel('Populacao')
16 plt.show()
```



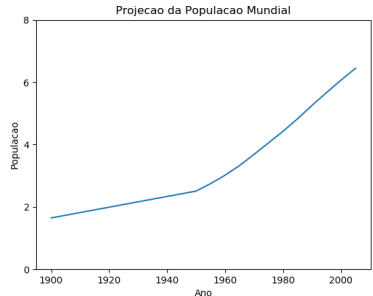
# Título Principal

```
1 plt.plot(anos, pops)
2 plt.xlabel('Ano')
3 plt.ylabel('Populacao')
4 plt.title('Projecao da Populacao
5           Mundial')
6 plt.show()
```



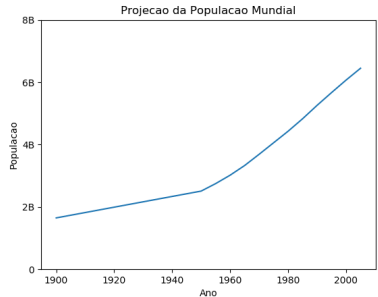
# Ticks (1)

```
1 plt.plot(anos, pops)
2 plt.xlabel('Ano')
3 plt.ylabel('Populacao')
4 plt.title('Projecao da Populacao
5           Mundial')
6 plt.yticks([0, 2, 4, 6, 8])
7 plt.show()
```



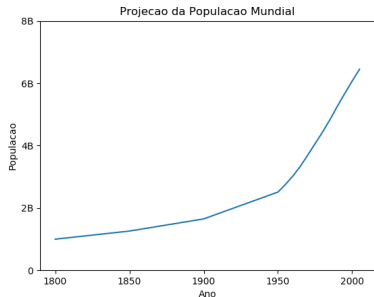
# Ticks (2)

```
1 plt.plot(anos, pops)
2 plt.xlabel('Ano')
3 plt.ylabel('Populacao')
4 plt.title('Projecao da Populacao
5           Mundial')
6 plt.yticks([0, 2, 4, 6, 8],
7            ['0', '2B', '4B', '6B', '8B'])
8 plt.show()
```



# Adicionando Dados Históricos

```
1 anos = [1800, 1850] + anos
2 pops = [1.0, 1.26] + pops
3 plt.plot(anos, pops)
4 plt.xlabel('Ano')
5 plt.ylabel('Populacao')
6 plt.title('Projecao da Populacao
7           Mundial')
8 plt.yticks([0, 2, 4, 6, 8],
9           ['0', '2B', '4B', '6B', '8B'])
10 plt.show()
```



# Antes x Depois

