

# Kurs rozszerzony języka Python

## Prezentacja danych

Marcin Młotkowski

14 grudnia 2021

# Plan wykładu

1 Wprowadzenie

2 matplotlib

- Wprowadzenie

3 Analiza danych

4 Coś na czasie

# Plan wykładu

## 1 Wprowadzenie

## 2 matplotlib

- Wprowadzenie

## 3 Analiza danych

## 4 Coś na czasie

# Wstęp

Analiza, przetwarzanie i wizualizacja danych

# Pakiety

- NumPy
- SciPy
- matplotlib
- Pandas

# Narzędzia

IPython

Jupyter Notebook

# Plan wykładu

1 Wprowadzenie

2 **matplotlib**

- Wprowadzenie

3 Analiza danych

4 Coś na czasie

# Co to jest

Narzędzie do rysowania wykresów. Bardzo różnych.

# Prosty przykład: parabola

```
import matplotlib.pyplot as plt

xs = [x for x in range(-10, 11, 1) ]
ys = [ x*x for x in xs ]

plt.plot(xs, ys)
plt.show()
```

## Inny przykład: funkcja parametryczna

$$x(t) = \alpha * \cos(t) - \beta * \cos(\gamma * t)$$

$$y(t) = \alpha * \sin(t) - \beta * \sin(\gamma * t)$$

gdzie  $t \in [0, 2 * \beta * \pi]$ .

## Inny przykład: funkcja parametryczna

$$x(t) = \alpha * \cos(t) - \beta * \cos(\gamma * t)$$
$$y(t) = \alpha * \sin(t) - \beta * \sin(\gamma * t)$$

gdzie  $t \in [0, 2 * \beta * \pi]$ .

```
import matplotlib.pyplot as plt
ts = frange(0, 3*2*math.pi, 0.01)

xs = [ 9*math.cos(t) - 3*math.cos(7*t) for t in ts ]
ys = [ 9*math.sin(t) - 3*math.sin(7*t) for t in ts ]

plt.plot(xs, ys)
plt.show()
```

## Inny przykład: histogram

```
xs = np.random.randint(1,10, size=1000)
```

```
plt.hist(xs, bins=10)  
plt.show()
```

# Porównanie implementacji: ciąg Fibonacciego

```
xs = [ n for n in range(1, 10) ]  
  
ys1 = [ timeit.timeit(lambda : fib1(i), number=1000) for i in xs]  
ys2 = [ timeit.timeit(lambda : fib2(i), number=1000) for i in xs]  
ys3 = [ timeit.timeit(lambda : fib3(i), number=1000) for i in xs]  
  
plt.plot(xs, ys1, marker='x')  
plt.plot(xs, ys2, marker='o')  
plt.plot(xs, ys3, marker='*')  
  
plt.legend(['rekurencja', 'iteracja', "macierzowy"])  
  
plt.show()
```

# Plan wykładu

## 1 Wprowadzenie

## 2 matplotlib

- Wprowadzenie

## 3 Analiza danych

## 4 Coś na czasie

# pandas

*panel data*

```
import pandas as pd
```

# Jakie dane

Różne, jakoś uporządkowane: csv, json, etc.

# Podstawowe typy danych

## Series

Seria danych, być może z etykietami (indeksami):

```
s = pd.Series(np.random.randn(5), index=['a', 'b', 'c', 'd', 'e'])
```

# Podstawowe typy danych

## Series

Seria danych, być może z etykietami (indeksami):

```
s = pd.Series(np.random.randn(5), index=['a', 'b', 'c', 'd', 'e'])
```

## DataFrame

Dwuwymiarowa tablica, gdzie kolumny mają różne typy:

poniedziałek	1.61
wtorek	2.71
środa	3.14

# Dane meteo

Instytut Meteorologii i Gospodarki Wodnej  
<https://danepubliczne.imgw.pl/>

# Dane meteo

Instytut Meteorologii i Gospodarki Wodnej  
<https://danepubliczne.imgw.pl/>

Dane o opadach i stanie wód z 2010 roku: pliki csv.

# Wczytanie danych

```
import pandas as pd
opady = pd.read_csv('o_m_2010.csv', encoding='iso8859-2',
                     header=None, names=nag_opady)
hydro = pd.read_csv("mies_2010.csv", encoding="iso8859-2",
                     header=None, names=nagl_hydro)
```

# Selekcja danych: Głogów

```
opady_gl = opady.loc[opady['Nazwa'] == 'GŁOGÓW']
hydro_gl = hydro.loc[(hydro['Nazwa'] == 'GŁOGÓW') &
                      (hydro['Ekstremum'] == 3)]
```

# Konwersja typów

Sprawdzenie typów kolumn:

```
dane.dtypes
```

Często typy kolumn są poprawnie rozpoznawane.

Można wskazać typy kolumn:

```
pd.read_csv("plik.csv", dtype={"kolumna2" : 'int'})
```

# Wyzwania

Kłopoty z konwersją:

- wstępnie sformatowane dane: "19 234";
- bogactwo zapisu daty.

```
strconv = lambda x : str(x.replace(' ', ''))  
pd.read_csv("plik.csv", converters={ "kolumna2" : strconv })
```

# Wykresy

```
import matplotlib.pyplot as plt

fig = plt.figure()
ax1 = fig.add_subplot(211)
opady_gl.plot(x='Mies', y='Opad maks', ax=ax1)
opady_gl.plot(x='Mies', y='Suma mies', ax=ax1)

ax2 = fig.add_subplot(212)
hydro_gl.plot(x='Mies kal', y='przepływ', ax=ax2)
plt.show()
```

# Posortowanie danych hydrologicznych

```
hydro_gl = hydro_gl.sort_values(by=['Mies kal'])
```

# Plan wykładu

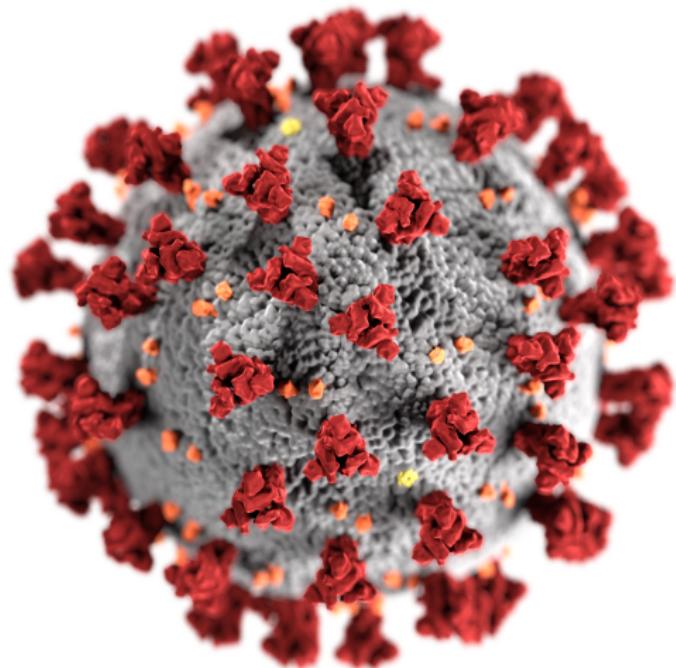
1 Wprowadzenie

2 matplotlib

- Wprowadzenie

3 Analiza danych

4 Coś na czasie



Źródło: [wikimedia.org](https://commons.wikimedia.org)

# Źródło danych

<https://www.gov.pl/web/koronawirus/wykaz-zarazen-koronawirus>  
zakładka: dane do pobrania