

Kurs rozszerzony języka Python

matplotlib, pandas

Marcin Młotkowski

27 listopada 2019

Plan wykładu

- 1 Obrazy w matplotlib
- 2 Analiza danych

Plan wykładu

1 Obrazy w matplotlib

2 Analiza danych

Przetwarzanie

W wersji podstawowej `matplotlib.image` przetwarza tylko obrazy w formacie PNG.

Przetwarzanie

W wersji podstawowej `matplotlib.image` przetwarza tylko obrazy w formacie PNG.

Do obsługi innych formatów konieczne jest zainstalowanie dodatkowych pakietów (`pillow`).

Zaczynamy

```
import matplotlib.image as mpimg
import matplotlib.pyplot as plt

img = mpimg.imread('stary_budynek.png')
plt.imshow(img)
plt.show()
```

Czym jest obrazek i jak z tego korzystać

```
print(img)
```

Redukcja jednej składowej koloru

```
img[:, :, 0] = 0
```


Podbicie czerwonego

```
red = np.where(img[:, :, 0] > 0.7, 1., img[:, :, 0])  
img[:, :, 0] = red
```

Losowe obrazki

```
img = np.zeros((512, 512, 3))

for i in range(img.shape[0]):
    for j in range(img.shape[1]):
        img[i, j] = np.random.random(3)

imgplot = plt.imshow(img)
plt.show()
```

Plan wykładu

1 Obrazy w matplotlib

2 Analiza danych

Kłopot z numpy

Tydzień temu

```
import numpy as np  
dane = np.loadtxt('dane.csv', delimiter=',', usecols=(5,7))
```

Tablice z numpy są jednego typu.

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=nagl_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)]
```

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'])
```