

Wizualizacja danych w Python, semestr letni 2022

Anna Muranova

Ćwiczenie 10

```
https://www.jetbrains.com/help/pycharm/managing-dependencies.html#create-requirements
```

Wymagania są w pliku pod linkiem:

```
http://wmii.uwm.edu.pl/~muranova/WD/requirements.txt.
```

program104.py

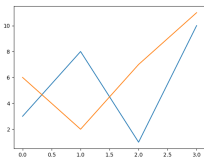
Pobierz plik z cenami jaj w Polsce w wybranych sieciach (źródła danych: <http://wmii.uwm.edu.pl/~piojas/wp-content/uploads/2019/04/jajka1.csv>, <http://www.dlahandlu.pl/koszyk/towar/10-jaj-najtansze,38.html>) - <http://wmii.uwm.edu.pl/~muranova/WD/jajka.csv>.

Zapisz plik w folderze projektu. Przeanalizuj następujący kod:

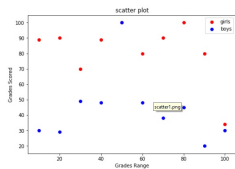
```
import pandas as pd

data = pd.read_csv('jajka.csv', sep=';', index_col=0,
encoding="cp1250")
data2 = data.stack()
data3 = data2.str.replace(',', ' ').astype('float')
srednia = data3.mean()
minCena = data3.min()
maxCena = data3.max()
print(data3[data3 == minCena])
print(data3[data3 == maxCena])
```

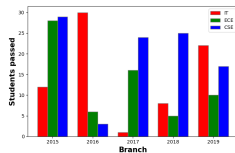
<https://www.w3resource.com/pandas/dataframe/dataframe-stack.php>



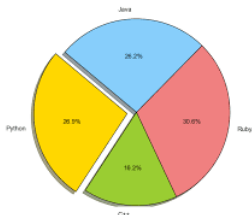
Rysunek: line



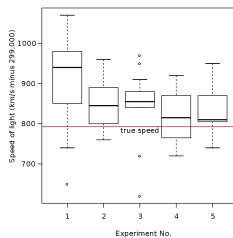
Rysunek: scatter/point



Rysunek: barplot



Rysunek: pie



Rysunek: boxplot

[https:](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.html)

[//pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.html](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.html)

Spośród przygotowanych plików csv z poprzednich ćwiczeń wybierz dwa, dla których wykonasz po 10 różnych operacji związanych z grupowaniem i filtrowaniem danych i zrobisz dwa wykresy.

https://prezydent20200628.pkw.gov.pl/prezydent20200628/pl/dane_w_arkuszach

[https://dane.gov.pl/pl/dataset/2582,
statystyki-zakazen-i-zgonow-z-powodu-covid-19-z-uw](https://dane.gov.pl/pl/dataset/2582,statystyki-zakazen-i-zgonow-z-powodu-covid-19-z-uw)

<https://github.com/mwaskom/seaborn-data>
penguins.csv

CC0-1.0 license

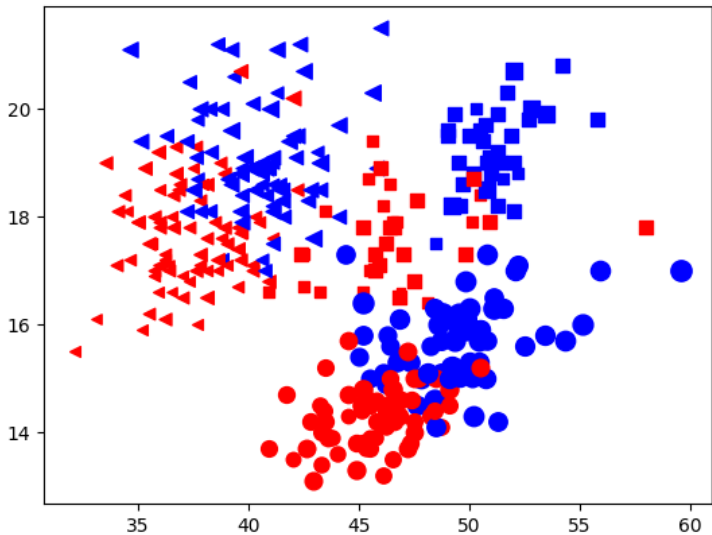
source: <https://github.com/allisonhorst/palmerpenguins>

<https://github.com/mwaskom/seaborn-data>

Na mojej stronie: <http://wmii.uwm.edu.pl/~muranova/WD/penguins.csv>.

Znajdź:

- średnia waga w każdej płcie, średnia waga z podziałem na płć i gatunek
- wszystkie wartości dla pingwinów z największej i najmniejszej wagami.
- ilość pingwinów z wyspy 'Torgersen' na każdej wyspie
- Przy pomocy <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.html> narysuj wykres słupkowy ilości pingwinów w zależności od wyspy: `data.groupby(['island']).size().plot.bar()`
Żeby pokazać wykres, podłącz `import matplotlib.pyplot as plt` i wykonaj polecenie `plt.show()`
- Narysuj wykres punktowy (scatter) zależności długości dzioba od szerokości.
- Zrób w poprzednim zadaniu, żeby kolor punktów zależał od płcie, rozmiar - od wagi/50, a kształt - od gatunku. Uwaga: Numpy <https://stackoverflow.com/questions/43482446/python-scatter-plot-different-colors-depending-on-value>
<https://stackoverflow.com/questions/47684652/how-to-customize-marker-colors-and-shapes-in-scatter-plot>



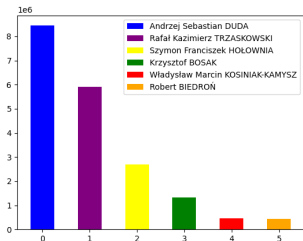
wyniki.csv

source: https://prezydent20200628.pkw.gov.pl/prezydent20200628/pl/dane_w_arkuszach

Wczytaj plik: <http://wmii.uwm.edu.pl/~muranova/WD/wyniki.csv>.

Znajdź:

- ilość ogólną otrzymanych kart do głosowania
- ile procentów ze wszystkich wyborców korzystali każdym sposobem (przez pełnomocnika, koperta,...)
- W której części kopert w każdym powiecie nie było oświadczenia (o osobistym i tajnym oddaniu głosu)
narysuj wykres punktowy tej zależności.
- Oblicz łączną sumy głosów dla każdego kandydata.
- narysuj podobny do następnego wykres osób z największymi liczbami głosów:



<https://stackoverflow.com/questions/57340415/matplotlib-bar-plot-add-legend-from-categories-dataframe-column>

Zrobić wykresy pudełkowy cen na jaja w zależności od sklepu i od miasta.
<http://wmii.uwm.edu.pl/~muranova/WD/jajka.csv>

