

Wizualizacja danych

Dr Anna Muranova
Semestr letni 2024, UWM w Olsztynie

11. Zajęcie (10.05.2024)

Plik z wymaganiami:

<http://wmii.uwm.edu.pl/~muranova/WDI2024/requirements.txt>

Upewnij się, że w projekcie masz podpięta biblioteki pandas, numpy oraz matplotlib.

Ćwiczenie 1 (program111.py). Dla pliku penguins.csv

source: <https://github.com/mwaskom/seaborn-data/blob/master/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 każdego gatunku na każdej wyspie
- ilość pingwinów gatunku 'Adelie' 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.

Żeby pokazać wykres, podłącz

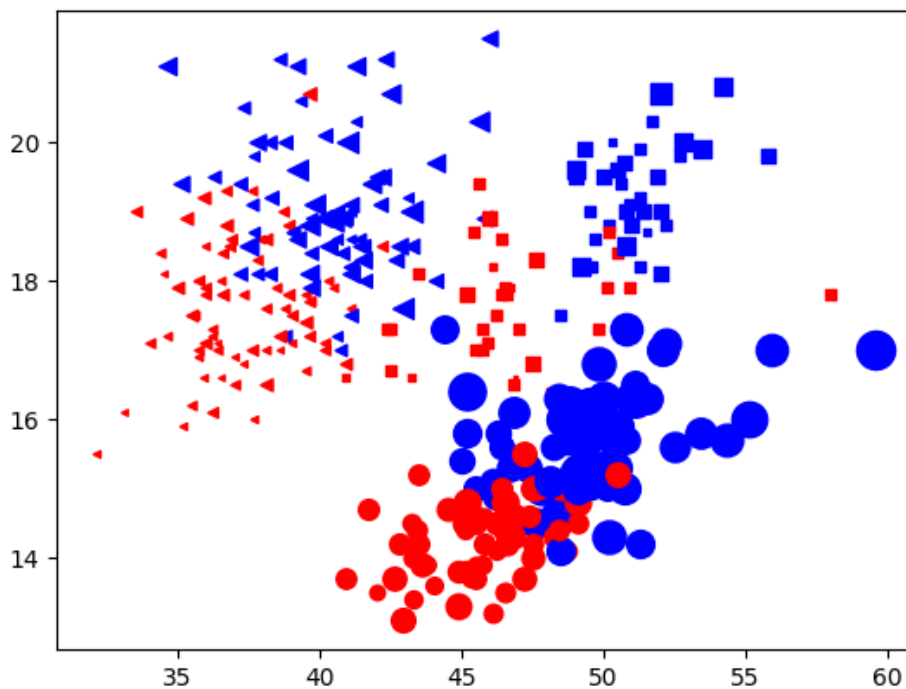
```
import matplotlib.pyplot as plt i wykonaj polecenie plt.show()
```

- Narysuj wykres punktowy (scatter) zależności szerokości dzioba od długości.
- Zrób w poprzednim zadaniu, żeby kolor punktów zależał od płcie, rozmiar - od wagi ($rozmiar = waga^5 / 2000$), *a kształt - od gatunku.

Uwaga: Numpy

<https://stackoverflow.com/questions/43482446/python-scatter-plot-different-col>

<https://stackoverflow.com/questions/47684652/how-to-customize-marker-colors-an>



Ćwiczenie 2 (program112.py). Biblioteka **seaborn**

Dokumentacja:

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

Baza danych 'tips':

<https://github.com/mwaskom/seaborn-data/blob/master/tips.csv>

Poćwicz. Żeby pokazać wykres, podłącz

```
import matplotlib.pyplot as plt i wykonaj polecenie plt.show()
```

Znowu pingwiny: Narysuj wykres punktowy zależności szerokości dzioba od długości. Zrób żeby kolor punktów zależał od płcie, rozmiar - od wagi, a kształt - od gatunku. Użyj polecenia:

```
penguins = sns.load_dataset("penguins") zamiast ładowania pliku.
```

Ćwiczenie 3 (program113.py). Irysy: <https://github.com/mwaskom/seaborn-data/blob/master/iris.csv>

- Narysuj 2 wykresy zależności punktowej (gatunki zaznacz kolorami) oraz 3 wykresy `catplot` różnych wartości od gatunków na danych z pliku.
- Narysuj `pairplot`. Jakie dwie cechy są najlepiej, żeby rozróżnić gatunki?