

Mathematica pozwala na importowanie danych z serwera Wolframu. Dane takie mogą dotyczyć, np. populacji miast, długości rzek, powierzchni państw ale także danych naukowych, np. masy jąder atomów czy budowy cząsteczek chemicznych. Dostęp do danych jest realizowany przez funkcje `CityData`, `CountryData`, `ElementData` i inne (częściowa lista jest dostępna w pliku pomocy pod adresem `guide/ScientificAndTechnicalData`).

## Zadanie 20

Proszę napisać funkcję `CityPopulation` zwracającą listę populacji miast spełniających kryterium analogiczne do pierwszego argumentu `CityData`. Zapytanie `CityData[All, "Population"]` nie zwraca oczekiwanego rezultatu bowiem nie wszystkie miasta mają zapisaną populację. Zamiast tego należy zmapować funkcję `CityData[#, "Population"]` na liście miast z `CityData[]` i wybrać odpowiednie pozycje np. przez `Select` lub `Cases`. Funkcja `CityPopulation[{All, "Poland", "Malopolskie"}]` powinna zwrócić listę populacji miast w Małopolsce.

Następnie, proszę utworzyć histogram populacji miast używając do tego funkcji `Histogram[dane, "Log"]` bądź poleceń `Tally` i `ListLogLinearPlot`.

Proszę kolejno napisać funkcję `FirstDigit`, która zwraca pierwszą cyfrę danej liczby całkowitej (przypominam o funkcji `IntegerDigits`). Proszę następnie wykonać histogram pierwszych cyfr populacji miast świata czyli `Histogram[FirstDigits/@CityPopulation[All]]`. Jaka cyfra występuje najczęściej? Proszę znormalizować histogram.

Proszę powtórzyć eksperyment z innymi danymi z bazy Wolframu, np. powierzchnią państw, długością rzek, populacją państw lub nawet masami pierwiastków bądź masami cząsteczek elementarnych. Do tego celu proszę zmienić funkcję `FirstDigit` tak by mogła przyjmować argumenty rzeczywiste i by podawała pierwszą znaczącą cyfrę argumentu nawet jeśli ten nie jest całkowity.

Rozkład ten jest nazywany rozkładem Benforda, i sprawdza się on najlepiej dla danych, których rozkład w skali logarytmicznej jest w dostatecznie szeroki, czyli danych rozciągniętych na bardzo różne skale (np. populacje miast zaczynają się od jednego mieszkańca a kończą na kilkudziesięciu milionach). Proszę nałożyć teoretyczny rozkład Benforda dany wzorem

$$P(d) = \log_{10} \left( 1 + \frac{1}{d} \right) \quad (1)$$

na znormalizowane wyniki eksperymentów. Rozkład Benforda jest inwariantny ze względu na skalowanie (musi tak być, ponieważ wartości liczbowe danych zależą od przyjętych jednostek, które są arbitralne - do mierzenia powierzchni państw mogliśmy użyć np. stóp kwadratowych). Proszę to sprawdzić przemnażając dane z jednego z eksperymentów przez dowolną liczbę. Dlatego też rozkład Benforda nie opisuje dobrze danych o ograniczonej skali, np. wzrostu człowieka.