

# Sprawozdanie nr 1

Klaudia Janicka 262268 i Natalia Iwańska 262270

2022-12-13

## 1. Wstęp

Niniejszy raport stanowi analizę danych rzeczywistych dotyczących trzęsień ziemi o skali wydarzenia powyżej 6 na przestrzeni lat 1900-2013. Dane pochodzą ze strony kaggle.com, gdzie źródłem danych jest United States Geological Survey.

Celem naszej analizy jest odpowiedzenie na pytanie jakie regiony są najbardziej narażone oraz zbadanie jakie czynniki mają wpływ na występowanie trzęsień ziemi.

### 1.1 Opis zmiennych

Do analizy wykorzystujemy dane zawarte w kolumnach:

- *place* - tekstowy opis regionu geograficznego w pobliżu wydarzenia;
- *latitude* - szerokość geograficzna podana w stopniach, przyjmuje wartości z przedziału  $[-90, 90]$ , gdzie wartości ujemne oznaczają południowe szerokości;
- *longitude* - długość geograficzna podana w stopniach, przyjmuje wartości z przedziału  $[-180, 180]$ , gdzie wartości ujemne dotyczą zachodnich długości;
- *mag* - skala wydarzenia;
- *nst* - liczba stacji sejsmicznych użytych do określenia lokalizacji trzęsienia;
- *time* - czas wystąpienia trzęsienia;
- *depth* - głębokość zdarzenia w kilometrach;

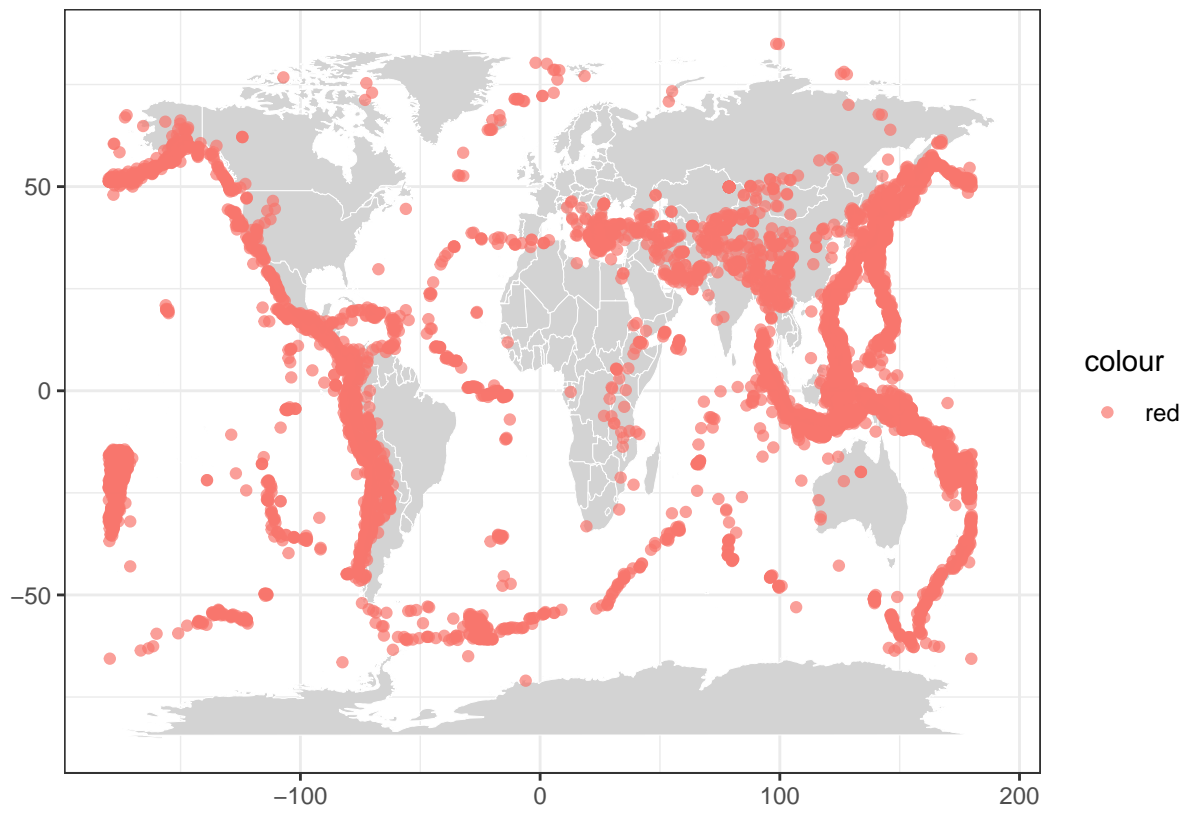
### 1.2 Obsługa błędów

Wiersze z brakami danych usuwamy za pomocą funkcji `drop_na()` z biblioteki `dplyr`.

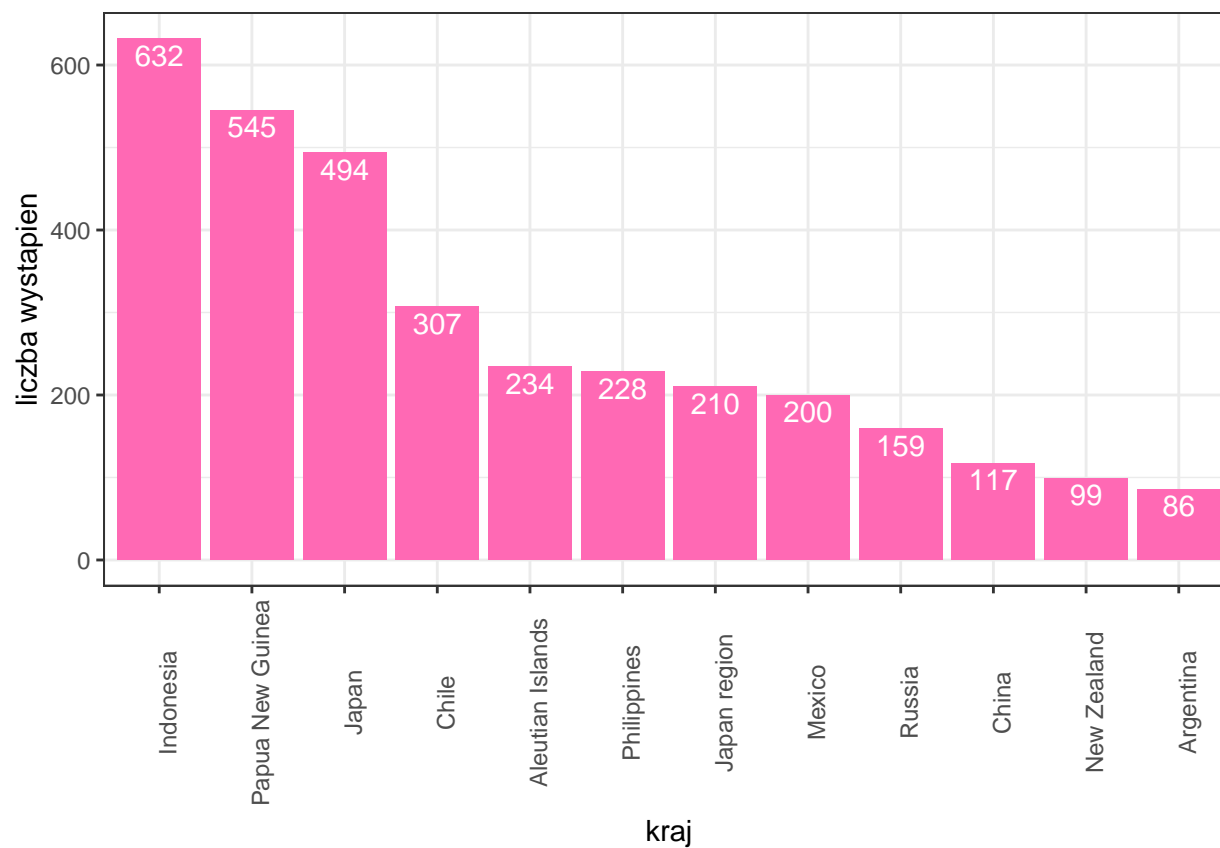
## 2. Analiza danych

- Wczytanie danych, nadanie odpowiednich etykiet, typów oraz wartości, obsługa braków danych

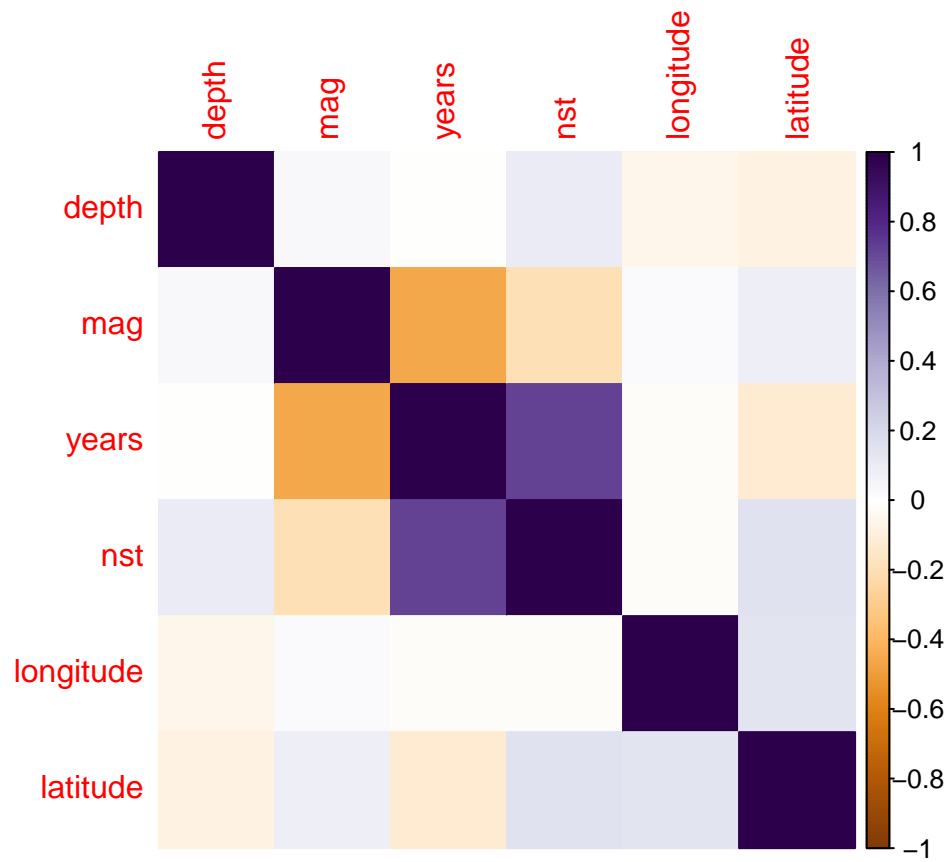
## 2.1 Mapa świata z naniesionymi miejscami trzęsień ziemi



## 2.2 Wykres słupkowy ilości wystąpień zdarzenia dla 12 najbardziej narażonych miejsc (?)



### 2.3 Macierz korelacji



### 3. Wnioski