

Sprawozdanie nr 1

Klaudia Janicka 262268 i Natalia Iwańska 262270

2022-12-17

1 Wstęp

Niniejszy raport stanowi analizę danych rzeczywistych dotyczących trzęsień ziemi o magnitudzie powyżej 6 na przestrzeni lat 1900-2013. Dane pochodzą ze strony **kaggle.com**. Zgodnie z informacjami zawartymi na podanej stronie źródłem danych jest United States Geological Survey.

Celem naszej analizy jest odpowiedzenie na pytanie jakie regiony są najbardziej narażone, zbadanie jakie czynniki mają wpływ na występowanie trzęsień ziemi oraz, czy poszczególne cechy trzęsienia są ze sobą powiązane.

1.1 Opis zmiennych

Typy danych występujących w poszczególnych kolumnach:

Tabela 1: Typy zmiennych dla danych zawartych w pliku.

	typ
X	integer
Date	character
Time	character
latitude	numeric
longitude	numeric
depth	numeric
mag	numeric
magType	character
nst	numeric
net	character
id	character
updated	character
place	character
type	character

Typy zawarte w tabeli 1 oznaczają odpowiednio:

integer - typ całkowity, który przyjmuje wartości całkowite,

numeric - typ zmiennoprzecinkowy, który przyjmuje wartości ułamkowe,

character - typ znakowy, przechowuje łańcuchy tekstowe.

Do analizy wykorzystujemy dane zawarte w kolumnach:

- `place` - tekstowy opis regionu geograficznego w pobliżu zdarzenia;
- `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` - magnituda, skala zdarzenia;
- `nst` - liczba stacji sejsmicznych użytych do określenia lokalizacji trzęsienia;
- `time` - czas wystąpienia trzęsienia;
- `depth` - głębokość zdarzenia w kilometrach;
- `date` - data zajścia zdarzenia;

gdzie jako zdarzenie rozumiemy wystąpienie trzęsienia ziemi.

Jako zmienne kategoryczne uznajemy zmienną `place`, która po transformacji wskazuje na kraj, w którym zdarzenie miało miejsce oraz `date`, która po transformacji oznacza miesiąc zajścia zdarzenia. Pozostałe zmienne to zmienne ciągłe.

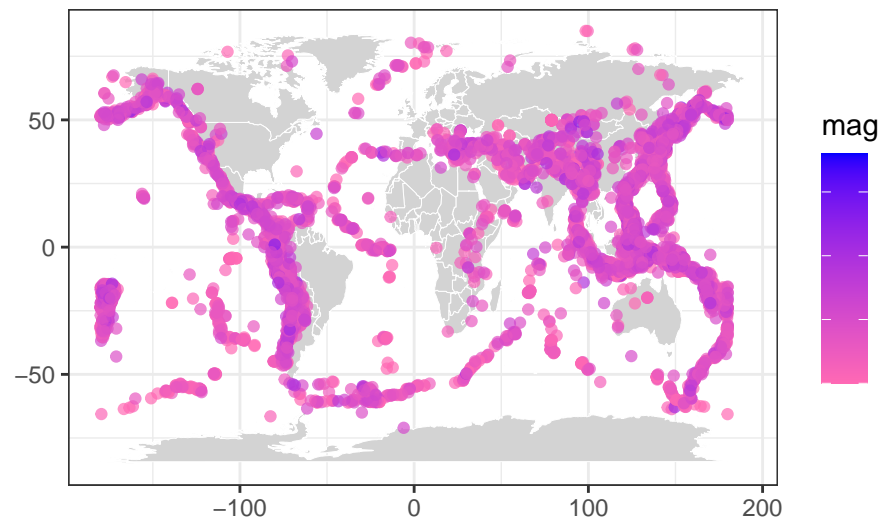
1.2 Obsługa błędów

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

2 Analiza danych

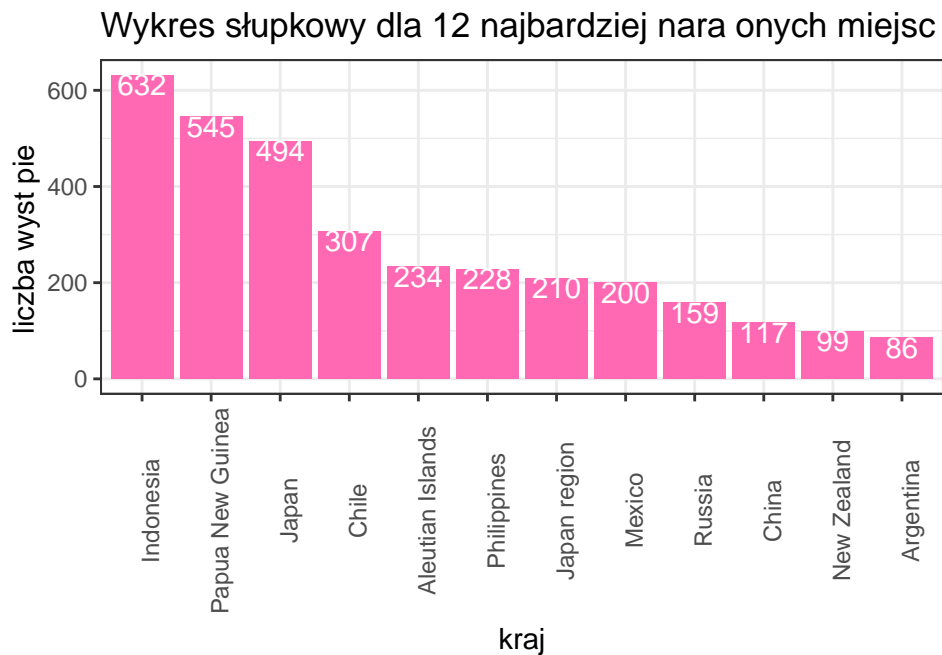
2.1 Miejsca najbardziej narażone na trzęsienia ziemi

W celu ustalenia regionów najbardziej narażonych na wystąpienie zdarzenia posłużymy się analizą graficzną.



Wykres 1: Mapa świata z naniesionymi miejscami trzęsień ziemi

Z powyższego wykresu 1 i posiadanej wiedzy z zakresu geografii, jesteśmy w stanie stwierdzić, że najbardziej narażone są między innymi Chile czy Japonia oraz inne kraje azjatyckie.

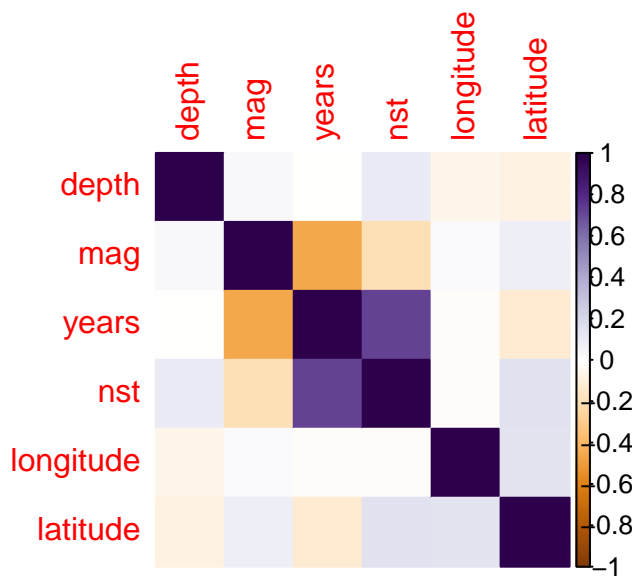


Wykres 2: Wykres słupkowy dla 12 najbardziej narażonych miejsc

Wykres 2 potwierdza nasze wcześniejsze przypuszczenia oparte o wykres 1.

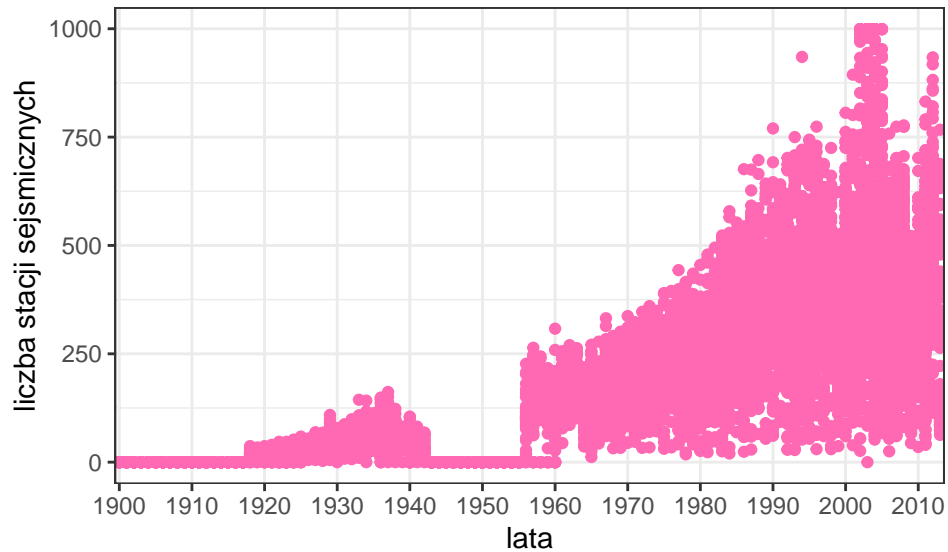
2.2 Macierz korelacji

W następnym kroku sprawdzimy zależności pomiędzy poszczególnymi zmiennymi. W tym celu tworzymy macierz korelacji 3.



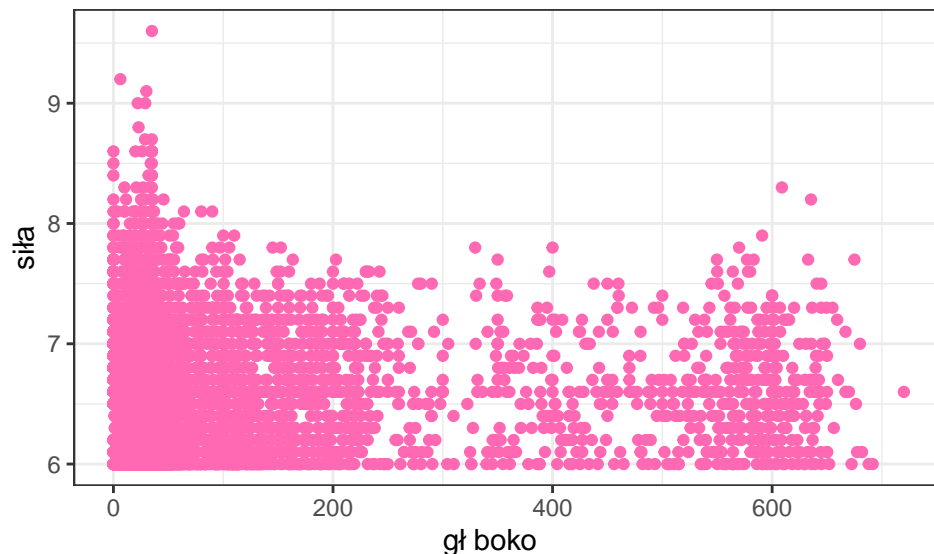
Wykres 3: Macierz korelacji

Jak widać na macierzy 3, najsilniejsza korelacja występuje pomiędzy kolumną *nst* oraz *years*, co więcej jest ona dodatnia i wynosi w zaokrągleniu 0.78 (korelacja Spearmana). Dzięki temu możemy stwierdzić, że w ciągu kolejnych lat powstawało coraz więcej stacji sejsmicznych, co możemy pokazać bardziej szczegółowo, tworząc poniższy wykres:



Wykres 4: Scatterplot zależności liczby stacji sejsmicznych od lat.

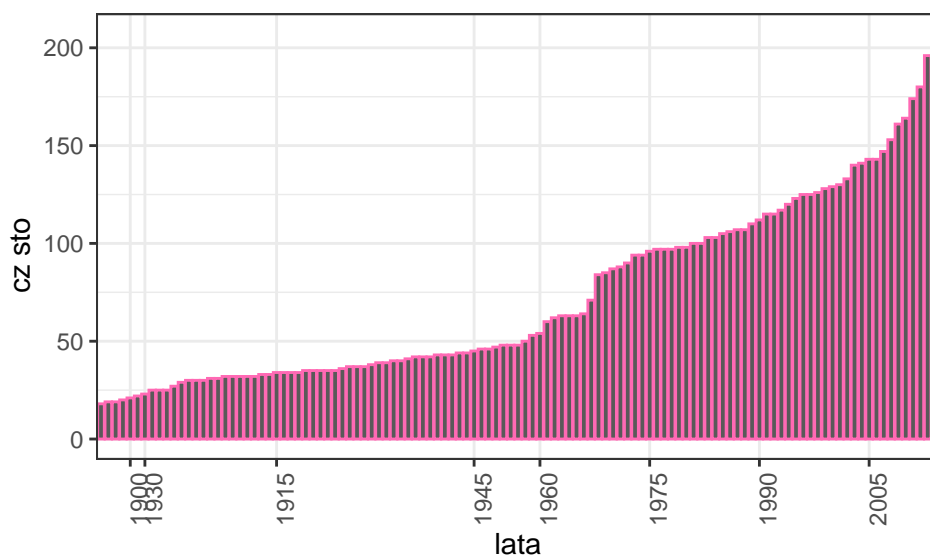
Zaskakiwać może fakt, że głębokość i siła trzęsienia nie są ze sobą silnie skorelowane, co możemy zobaczyć na wykresie poniżej:



Wykres 5: Scatterplot zależności siły trzęsienia od jego głębokości

Reszta kolumn nie jest ze sobą tak silnie skorelowana i nie jest tak zaskakująca jak zależności wyżej. Większość wartości oscyluje wokół zera, więc nie będziemy im się bliżej przyglądać.

2.3 Stosunek ilości trzęsień ziemi do kolejnych lat

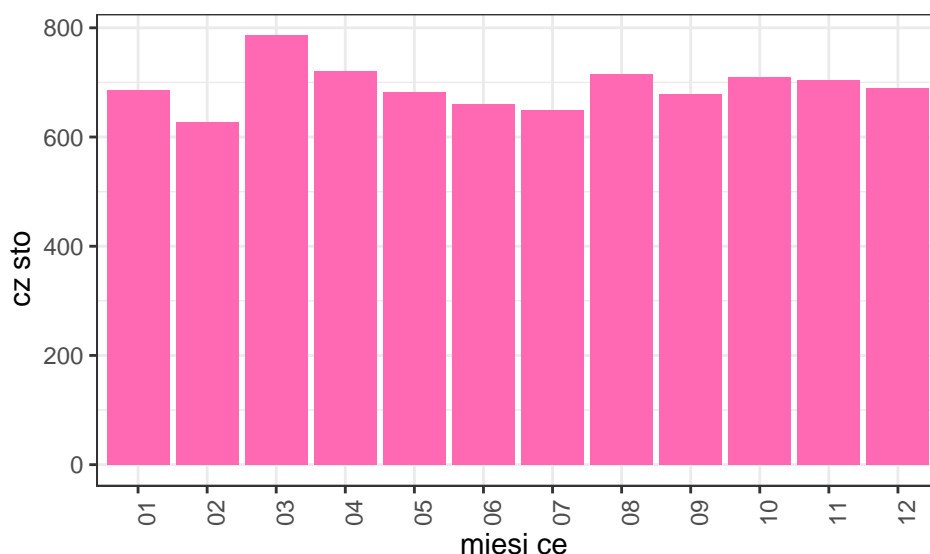


Wykres 6: Wykres kolumnowy ilości trzęsień ziemi w stosunku do lat

Na podstawie wykresu 6 można by przypuszczać, że wraz z upływem lat, przybywało trzęsień ziemi, jednak musimy wziąć pod uwagę fakt, że w tym samym czasie przybywało stacji sejsmicznych, które takie zdarzenia rejestrowały, co pokazaliśmy na wykresie 4. Zatem nie możemy jednoznacznie stwierdzić, czy liczba trzęsień ziemi zmieniała się w konkretny sposób.

2.4 Badanie wpływu pory roku na liczbę trzęsień ziemi

Kolejnym aspektem jaki poddamy analizie jest wpływ pór roku na ilość zdarzeń.

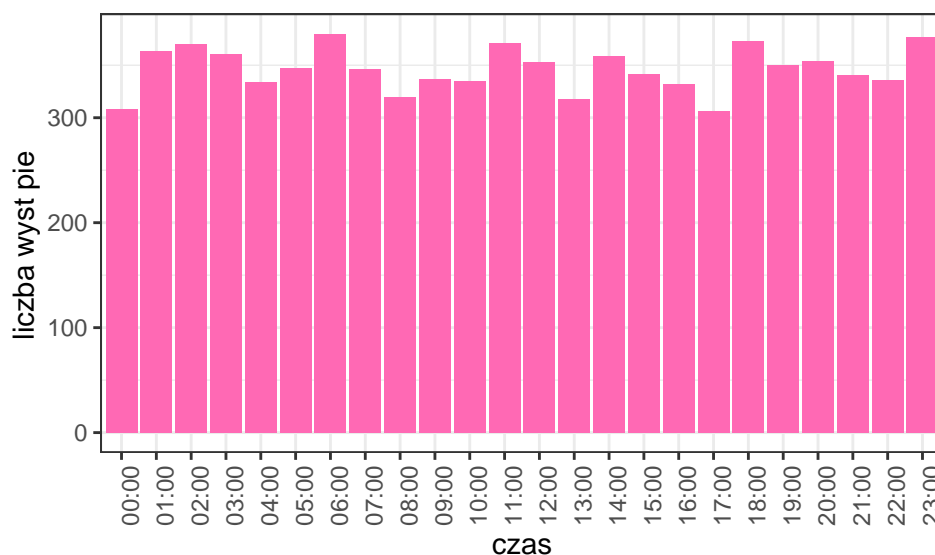


Wykres 7: Wykres kolumnowy ilości trzęsień ziemi w stosunku do miesiąca, w którym wystąpiły

Na powyższym wykresie możemy zauważyć, że słupki dla wszystkich miesięcy są dość podobnej wysokości. Nie ma żadnych mocno odstających liczb, zatem wnioskujemy, że pora roku nie ma wpływu na to, czy trzęsienie ziemi wystąpi, czy nie.

2.5 Badanie Zależności między porą dnia, a liczbą trzęsień ziemi

W celu zbadania owej zależności równie posłużymy się wykresem słupkowym. Na potrzeby zrobienia wykresu, zaokrąglone, do najbliższej całkowitej godziny, zostały zmienne z kolumny **Time**.



Wykres 8: Wykres kolumnowy ilości trzęsień ziemi w stosunku do godziny, w której wystąpiły? XD

Tak jak przy badaniu wpływu pory roku, na wykresie 8 możemy zauważyć, że wartości znów są do siebie zbliżone. Dodatkowo wartość współczynnika korelacji Spearmana dla tych danych wynosi w zaokrągleniu 0.01, co jest wartością bliską zera, a zatem pora dnia również nie ma wpływu na występowanie trzęsień ziemi.

3 Wnioski

Przechodząc do podsumowania naszej pracy przeprowadzona przez nas analiza dla danych na temat trzęsień ziemi z lat 1900-2013 doprowadziła nas do następujących wniosków. Po pierwsze odpowiadając na pytanie badawcze postawione na początku pracy możemy stwierdzić, że najbardziej narażone na występowanie trzęsień ziemi były regiony azjatyckie, Oceania jak również zachodnie wybrzeża Ameryki Południowej oraz Ameryki Północnej (1). Stąd wniosek, że to właśnie te regiony są najbardziej narażone. Jeśli jednak spojrzymy na konkretne państwa to zgodnie z wykresem 2 w czołowej trójce plasują się takie państwa jak Indonezja, Papua-Nowa Gwinea i Japonia. Idąc dalej...