

# Analiza trzęsień ziemi na przestrzeni lat 1900-2013

Klaudia Janicka 262268 i Natalia Iwańska 262270

2022-12-18

## 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.

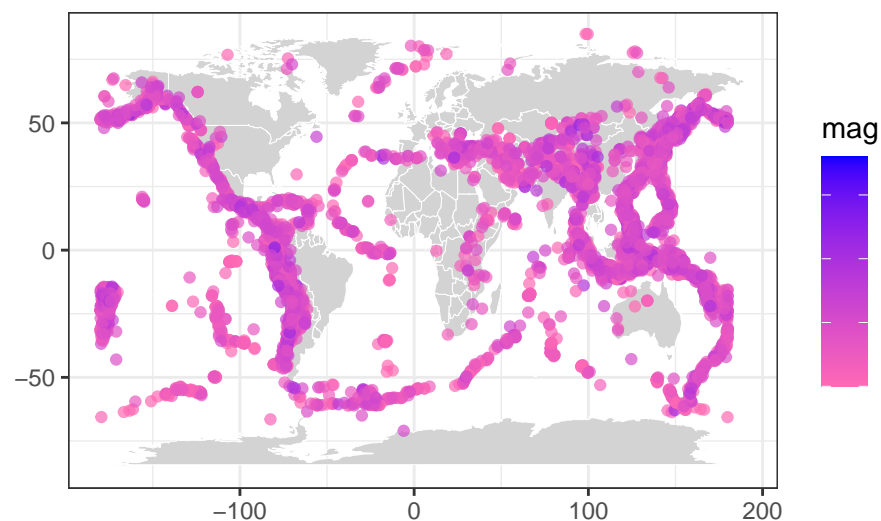
## 1.2 Obsługa błędów

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

# 2 Analiza danych

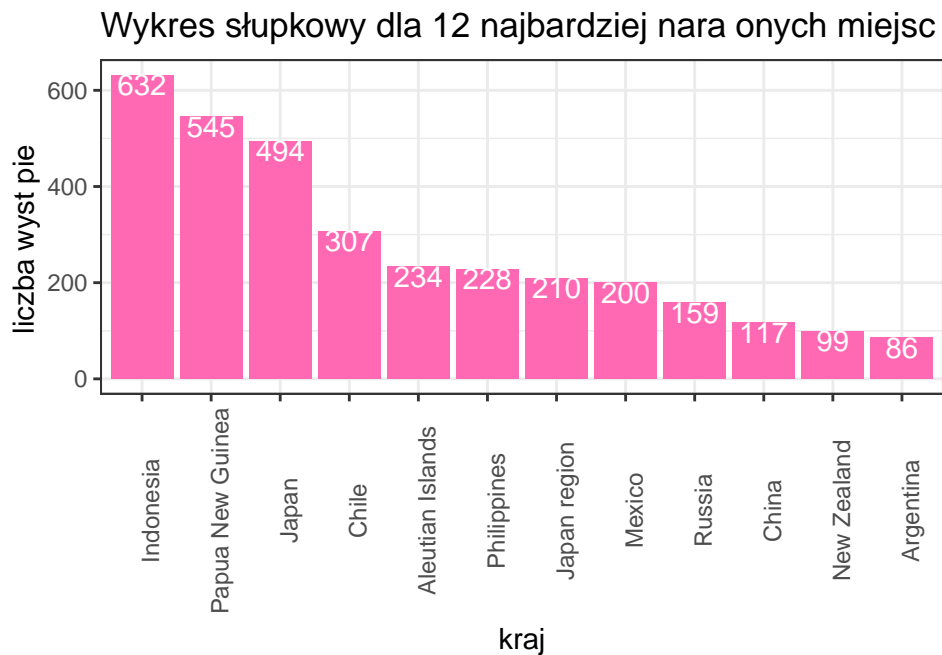
## 2.1 Miejsca najbardziej narażone na trzęsienia ziemi

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.

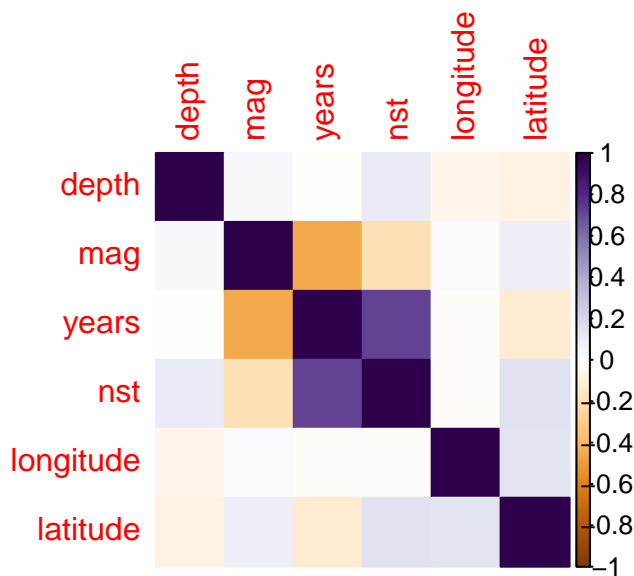


Wykres 2: Wykres słupkowy dla 12 najbardziej narażonych miejsc

Wykres 2 potwierdza nasze wcześniejsze przypuszczenia oparte o wykres 1.

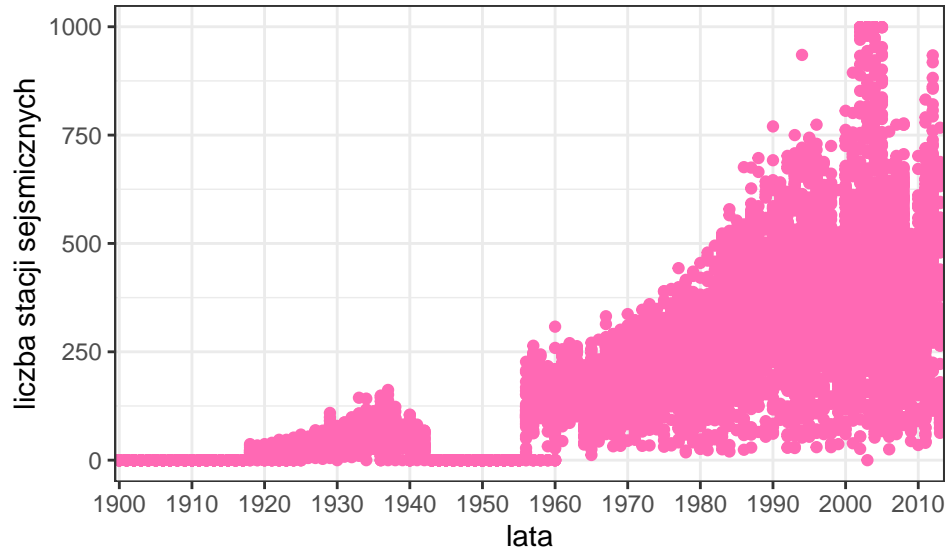
## 2.2 Macierz korelacji

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



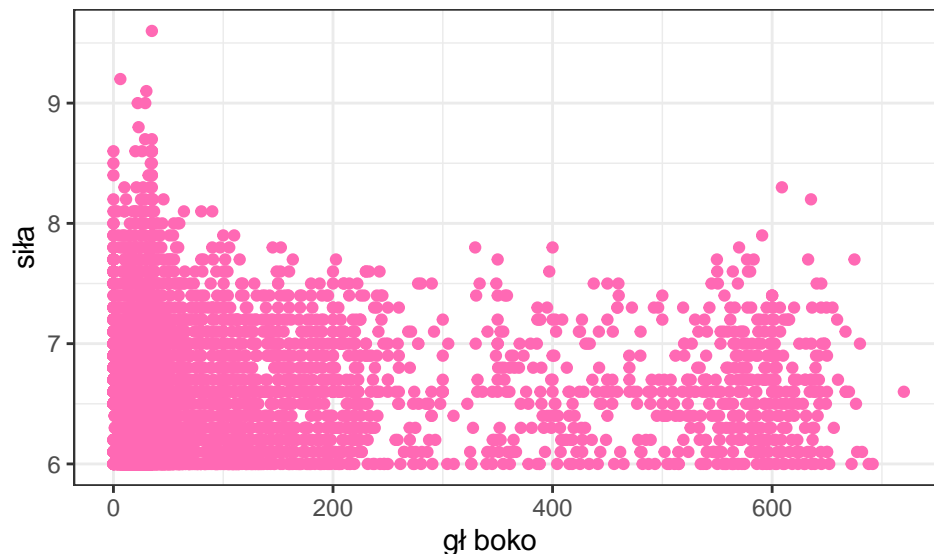
Wykres 3: Macierz korelacji

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, współczynnik korelacji Spearmana dla tych wielkości wynosi bowiem zaledwie 0.08. Dodatkowo fakt ten możemy zaobserwować na wykresie 5 zamieszczonym poniżej.



Wykres 5: Scatterplot zależności siły trzęsienia od jego głębokości

Nie widzimy tu żadnego trendu, który mogły wskazywać na zależność między danymi. (Czy to ma sens?)

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 Średnia magnituda a miejsce wystąpienia trzęsienia (?czy to ma sens?)

W tabeli 2 zawartej poniżej sprawdzamy średnią siłę trzęsienia dla każdego miejsca, w którym ono wystąpiło.

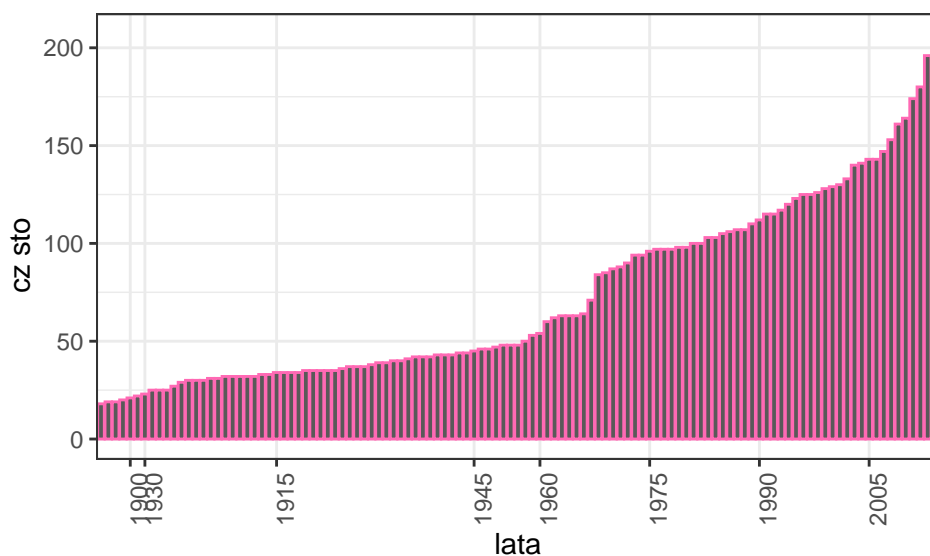
Tabela 2: Wartość średniej siły dla poszczególnych miejsc.

miejsce	średnia magnituda
Afghanistan	6.607229
Alaska	6.447273
Aleutian Islands	6.561539
Argentina	6.547674
Argentina region	7.000000
Australia	6.366667
Baja California	6.500000
Barbados	6.500000
Bolivia	6.611111
Brazil	6.693750
California	6.450000
Canada	6.554167
Canada region	6.428571
Chile	6.648208
China	6.682051
Christmas Island	6.700000
Colombia	6.800000
Congo-Rwanda-Uganda	6.200000
Congo-Uganda	7.000000
Dem. Republic of the Congo-Tanzania	6.600000
Dominican Republic	7.000000
Ecuador region	6.800000
Federated States of Micronesia	6.618750
Fiji	6.050000
French Polynesia region	6.000000
Greece	6.471429
Guatemala	6.300000
Hawaii	6.622222
Honduras	6.000000
India	6.700000
India region	6.417308
Indonesia	6.519810
Iran	6.733333
Italy	6.800000
Japan	6.602024
Japan region	6.604286
Kazakhstan	6.100000
Kiribati region	6.800000
Leeward Islands	6.611111
Los Lagos	7.000000
Magallanes	7.800000
Malaysia	6.200000
Mexico	6.716500

miejsce	średnia magnituda
Mindanao	6.833333
New Caledonia	6.441667
New Zealand	6.485859
New Zealand region	6.621429
Nicaragua	6.350000
Northern Mariana Islands	6.541463
Nunavut	6.250000
offshore Azerbaijan	6.650000
offshore Turkmenistan	6.333333
Pakistan	7.250000
Papua New Guinea	6.527339
Peru	6.340000
Philippines	6.631140
Portugal	6.500000
Prince Edward Islands	6.100000
Puerto Rico	6.850000
Russia	6.666038
Russia region	6.494444
Solomon Islands	6.420000
South Sandwich Islands	6.750000
Sulawesi	6.678049
Taiwan	6.250000
Tanzania	7.600000
Tonga	6.342857
Trinidad and Tobago	6.550000
Uganda	6.100000
Vanuatu	6.280000
Venezuela	6.428571
Washington	6.600000
Windward Islands	7.050000
Wyoming	7.300000

Ponadto wyliczając korelację Spearmana pomiędzy średnią magnitudą a ilością wystąpień zdarzenia w danym miejscu otrzymujemy wynik -0.04, co wskazuje na ujemną korelację między tymi wartościami. Jednak jego wartość jest bardzo bliska zera co pozwala nam przypuszczać, że między danymi nie ma związku liniowego. (idk, ale tak myślę)

## 2.4 Stosunek ilości trzęsień ziemi do kolejnych lat

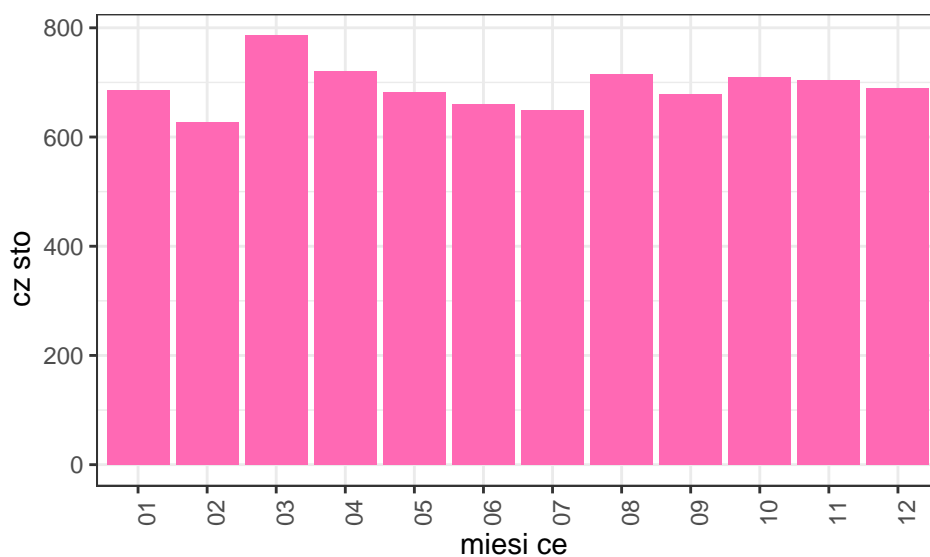


Wykres 6: Wykres kolumnowy ilości trzęsień ziemi w stosunku do lat

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.

## 2.5 Badanie wpływu pory roku na liczbę trzęsień ziemi

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

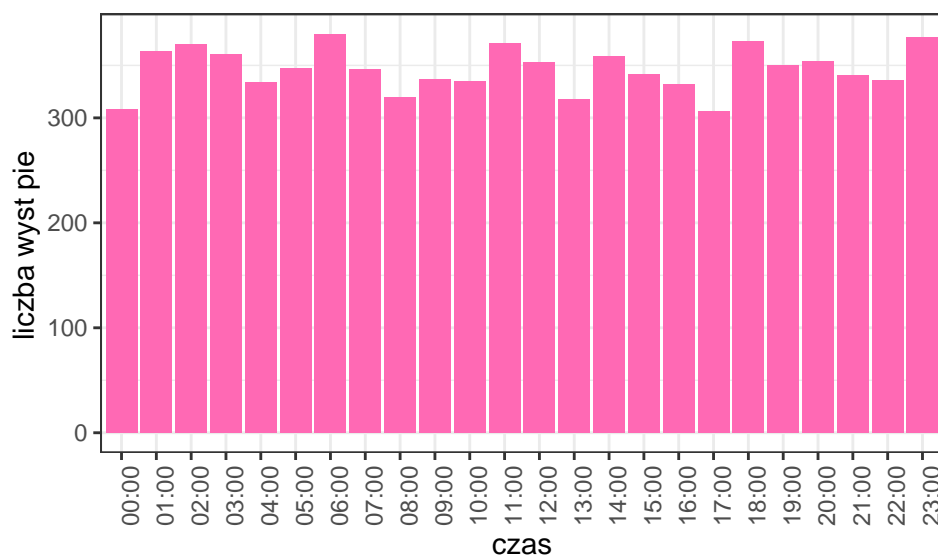


Wykres 7: Wykres kolumnowy ilości trzęsień ziemi w stosunku do miesiąca, w którym wystąpiły

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.

## 2.6 Badanie zależności między porą dnia, a liczbą trzęsień ziemi

W celu zbadania owej zależności również 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?

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 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 kraje jak Indonezja, Papua-Nowa Gwinea i Japonia. Warto tutaj zaznaczyć, że wyniki te mogą być delikatnie zakłamane (*wtf... brzmi dziko*) ze względu na ilość stacji sejsmicznych użytych do zlokalizowania zdarzenia. Zgodnie z wykresem 4 z biegiem lat ośrodków takich przybywało, co pozwalało dokładniej określić położenie niż w latach wcześniejszych. Jednak nie uważamy, aby miało to znaczący wpływ na otrzymane wyniki. (Czy ja właśnie sobie zaprzeczyłam? Być może XD)

Idąc dalej, zgodnie z macierzą korelacji 3 jak i późniejszą analizą nie widzimy znaczących korelacji w danych. Co może zaskakiwać głębokość odnotowanego zdarzenia nie jest znacząco zależna od jego siły, można powiedzieć, że jest to nieintuicyjny wynik. (*włączył mi się polski, ok?*)



Badając średnią siłę trzęsienia (tabela 2) i porównując ją dla 12 najbardziej narażonych miejsc (wykres 2) możemy stwierdzić, że średnie wartości są do siebie zbliżone. Dodatkowo korelacja dla wszystkich danych jest bliska zera (Tylko co nam to daje?)

Kolejne analizy nie wykazały żadnych zależności związanych z czasem czy też porą roku, zatem możemy stąd wywnioskować, że czynniki te nie mają wpływu na badane zjawisko.

Reasumując nasze rozważania trzęsienia ziemi są zjawiskiem nieprzewidywalnym i zależnym w głównej mierze od położenia geograficznego. Można przypuszczać, że zaczęły występować częściej niż miało to miejsce w poprzednich latach, jednak nie możemy tego bezpośrednio stwierdzić z uwagi na liczbę stacji badawczych, których liczba powiększała się, a co za tym idzie obszar pomiarów również ulegał zwiększeniu. Warto też zaznaczyć, że dane poddane analizie opisywały trzęsienia o magnitudzie powyżej 6, czyli nasze wnioski nie muszą być prawdziwe w ogólnym przypadku. (O ile w ogóle są XD)