Analýza významných svetových katastrofických udalostí

Fakulta matematiky, fyziky a informatiky

Univerzity Komenského

TETE - správa

Názov: Analýza významných svetových katastrofických udalostí.

Autor: Tomáš Varga, Erik Božík, Tadeáš Kaminský, Erik Kolesár

Dátum: 15.5.2023

*Obsah*

[**Úvod 3**](#_heading=h.gjdgxs)

[**Načítanie dát 3**](#_heading=h.30j0zll)

[**Analýza vedeckých otázok 5**](#_heading=h.1fob9te)

[1. Otázka: Závisí úmrtnosť od magnitúdy zemetrasení? (Magnitúda a úmrtia). 5](#_heading=h.3znysh7)

[2. Otázka: Kde a kedy sa zemetrasenia diali/dejú? 11](#_heading=h.2et92p0)

[3. Otázka: Aké typy sopiek sú najviac výbušné. 20](#_heading=h.tyjcwt)

[4. Otázka: Závisí počet smrtí od výbuchov?. 25](#_heading=h.3dy6vkm)

[5. Otázka: Aké faktory zapríčiňujú tsunami. 32](#_heading=h.1t3h5sf)

[**Záver 38**](#_heading=h.1l325sq6nhux)

[**Zdroje 39**](#_heading=h.izyo2k9ubszo)

# 

# 

# 

# 

# 

# 

# 

# **Úvod**

Náš data set poskytoval informácie ohľadom katakliziem sveta od zemetrasení ku vulkanickým erupciam až po tsunami. Rozhodli sme sa preto venovať sa každej jednej kataklyzme. Analýzy pre jednotlivé pohromy sú individuálne keďže sme v rôznych datasetoch dostali rôzne informácie ohľadom danej kataklyzmi. Keďže nám boli poskytnuté informácie aj ohľadom polohy analýzu sme zameriavali aj na vykreslovanie do mapy.

# **Načítanie dát**

Načítanie dát do pandas DataFrame objektu bolo pre nás obzvlášť zložité (link na dáta spolu so všetkými použitými nástrojmi sú v sekcii zdroje). Dáta sme si totižto mohli stiahnuť len vo formáte .kmz. Tento formát je zazipovaný súbor, ktorý obsahuje hlavne .kml súbory. Použili sme 7zip na odzipovanie a dostali sme tak prístup k týmto .kml súborom. Jedná sa o formát, ktorý vie najlepšie využiť Google Earth na zobrazenie.

Najprv sme skúšali načítať tieto dáta pomocou funkcie z knižnice geopandas, s tým, že použijeme kml driver poskytovaný knižnicou fiona. Dáta sme síce načítali do dataframe-u, ale keď sme si ich skontrolovali pomocou aplikácie Google Earth pro tak sme si všimli, že sa ani zďaleka nenačítali všetky. Okrem tohto bol však problém aj s tým, že druhá súradnica v stĺpci geometry bola pre každý riadok nulová, čo nebolo správne nakoľko by sa všetky dáta načítali na rovníku. Driver, ktorý poskytuje knižnica fiona nefungoval dostačujúco na dané súbory.

Tento problém sme riešili vďaka konvertoru, ktorý je voľne dostupný na internete. Tento konvertor nám prerobil .kml súbory na .geojson. Geojson súbory s dátami sme si následne nalinkovali na github a pomocou url sme ich sťahovali do notebooku. Tieto dáta sa už načítali do dataframe-u bez žiadnej straty.

Všimli sme si však, že dáta sa načítali v troch stĺpcoch. Stĺpce name a geometry boli v poriadku využiteľné avšak stĺpec Description obsahoval pre každú katastrofu zvlášť tabuľku zakódovanú v html jazyku, ktorá popisuje jednotlivé katastrofy. Preto sme pre každý načítaný dataset museli tento stĺpec prerobiť na viacero stĺpcov. Ďalej sme sa v dátach venovali hlavne konvertovaniu údajov do iných dátových typov tak, aby bola vizualizácia jednoduchšia.

Pracovali sme so štyrmi dataframe-mami, z ktorých každý sa vyznačoval inými stĺpcami. Prvý dataset earthquakes obsahoval v stĺpci location informáciu o krajine a aj bližšie popísanú lokáciu (mesto, svetovú stranu atď.). Tento stĺpec sme pre niektoré vizualizácie rozdelili na dva.. Magnitude sme len prekonvertovali na float, s tým, že hodnoty Unknown sa zmenili na np.nan(). Formát v stĺpci deaths sme menili podľa toho, čo sme zrovna s dátami potrebovali robiť. Keď sme ich potrebovali vysortovať, tak sme ich zmenili na tuple a potom následne znova na pôvodný string. V stĺpci tsunami sme zmenili hodnoty Yes, No na True a False. Dátum (date) bol problematický, pretože pd.date\_time() mal určité obmedzenia. My sme však dátum používali pri vizualizácii zemetrasení, ktoré sa stali po roku 1900, čiže len tieto sme potrebovali prekonvertovať na pd.date\_time(). Stĺpec geometry sme vôbec nezmenili. Mali sme teda informáciu o lokácií, magnitúde, počtu úmrtí v rozsahu, dátume, polohe a o tom či dané zemetrasenie spôsobilo tsunami.

V ďalších dátach o výbuchoch sopiek (volEvents) bol problematický len stĺpec name, keďže v ňom bola obsiahnutá aj informácia o roku a aj o mieste. Tento stĺpec sme rozdelili na dva. Ostatné stĺpce sme menili analogicky rovnako ako pri earthquakes. Mali sme informáciu o štáte, morfológií, výbušného indexu VEI, počtu smrtí, názvu sopky, polohy, dátumu a o tom či daný výbuch spôsobil tsunami.

Tretie dáta neobsahovali žiadne číselné údaje. Jednalo sa hlavne o popisné informácie daného tsunami. Tieto stĺpce sme špeciálne neupravovali. Na prácu sme mali stĺpce o krajine, bližšej lokácií, príčine, validite tsunami, počte smrtí, dátumu, polohe.

Posledný dataset nám poskytol údaje o tom kde sa tsunami prejavilo. V tomto súbore sme konvertovali hlavne číselný údaj o výške (max\_height\_m) na float alebo np.nan(). Ostatné stĺpce obsahovali hlavne popisné informácie. Tento dataset obsahuje informácie o tom, v ktorej krajine sa dané tsunami prejavilo, údaj o bližšej lokácií, ako bol daný prejav nameraný, údaj o výške vlny, dátum a poloha.

Tieto štyri tabuľky sme menili rôznymi spôsobmi tak aby sme dokázali danú vizualizáciu vytvoriť.

K tomuto sme ešte použili údaje o krajinách, ktoré sme použili z databázy knižnice geopandas. Do tejto tabuľky ešte pridali stĺpec o hustote obyvateľstva, ktorý sme vypočítali pomocou údajov o rozlohe a populácii už obsiahnutých v tejto tabuľke.

Vo vizualizáciách sme pouźili aj dáta, ktoré nám sprístupnili údaje o mestách na svete aj s údajmi o populácií. Tieto dáta boli dosť obšírne.

Ako posledné sme použili ešte dáta o tektonických doskách. Tieto sme vôbec nezmenili. Použili sme ich iba na vizualizácií knižnicou Plotly, kvôli ktorej sme ich vôbec meniť nemuseli.

# 

# 

# 

# 

# 

# 

# **Analýza vedeckých otázok**

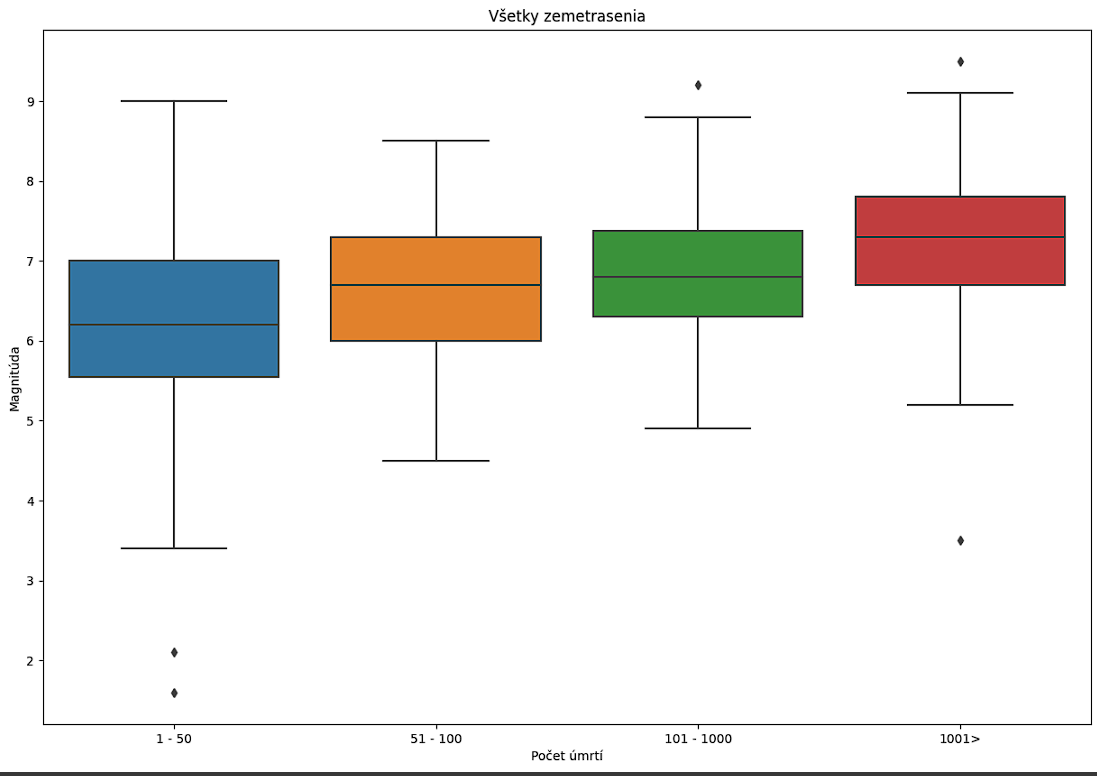
## 1. [Otázka: Závisí úmrtnosť od magnitúdy zemetrasení? (Magnitúda a úmrtia).](#_heading=h.3znysh7)

**Analýza**

Pátranie po odpovedi tejto otázky sme začali tým, že sme si na mapu nechali vykresliť všetky zemetrasenia a ich lokáciu. Tento graf nám nepomohol inak ako len na vizualizáciu všetkých zemetrasení v našom datasete.

Následne sme sa rozhodli dané dáta dať do box plotu, ktorý nám priblíži aké hodnoty magnitúdy v porovnaní s úmrtiami pre jednotlivé zemetrasenia máme.

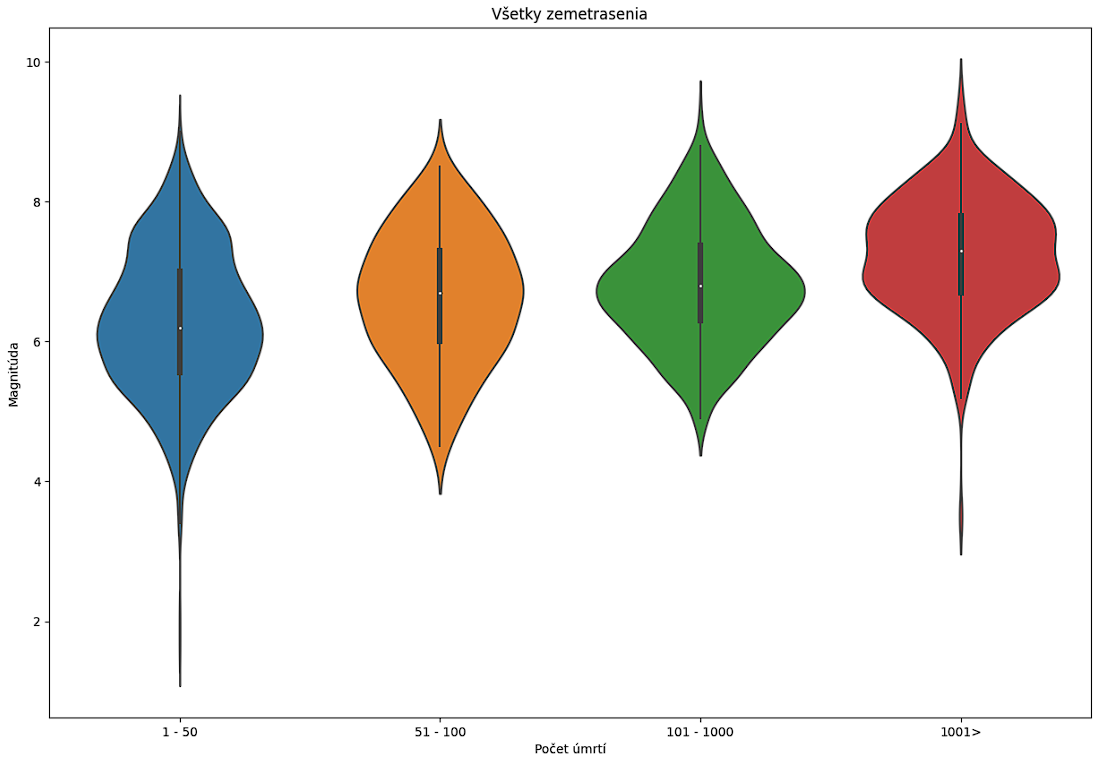
**BOX PLOT**

****

Na už vytvorenom box plote vidíme akúsi rastúcu priemernú výšku magnitúdy spolu s nárastom úmrtí. Následne pre ešte lepšiu vizualizáciu dát sme si dali vykresliť aj violin plot.

Violin plot má tiež celkom očakávaný tvar.

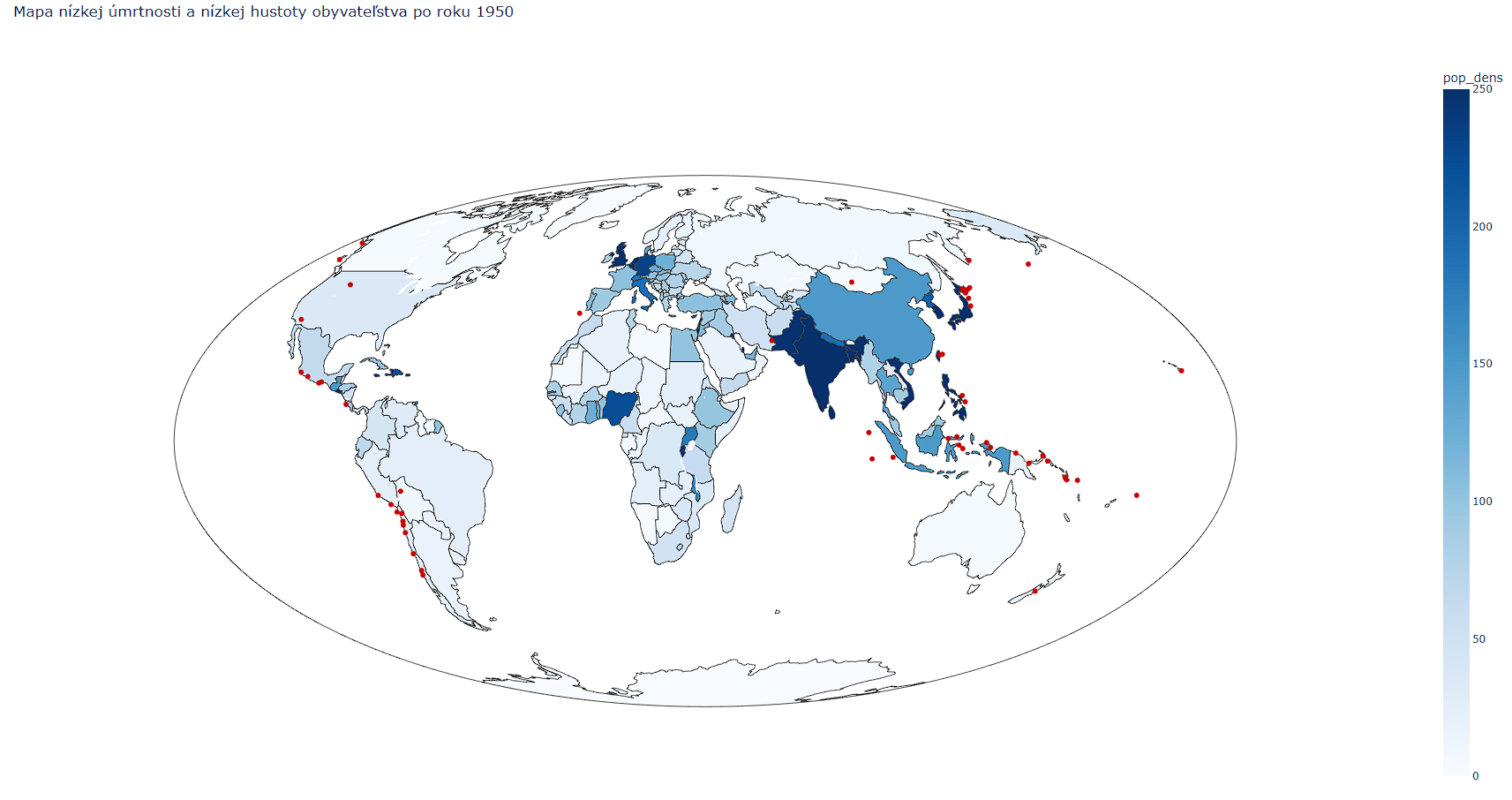
**VIOLIN PLOT**

****

Vidíme tu, že každá skupina úmrtí je zastúpená nejakou hodnotou magnitúdy najviac, čo je reprezentované tým, že je tam samotný plot najširší. V tomto grafe nás ale zaujala časť v ktorej boli reprezentované zemetrasenia o vysokej magnitude ale s malým počtom úmrtí, boli sme zvedavý ako je možné že niektoré zemetrasenia s takou veľkou magnitúdou majú tak málo úmrtí.

Vykreslili sme si zemetrasenia o magnitude vyššej ako 7.5 a úmrtiami 1-50 ľudí. Ďalšiu vec ktorú sme obmedzili boli roky v ktorých sa dané zemetrasenia stali. Toto rozhodnutie sme odôvodnili tým, že sa hustota populácie počas rokov rôzne menila a porovnávať staré dáta s novými by nám mohlo ovplyvniť odpoveď na našu otázku kvôli už spomenutému problému s rastúcou/klesajúcou hustotou populácie.

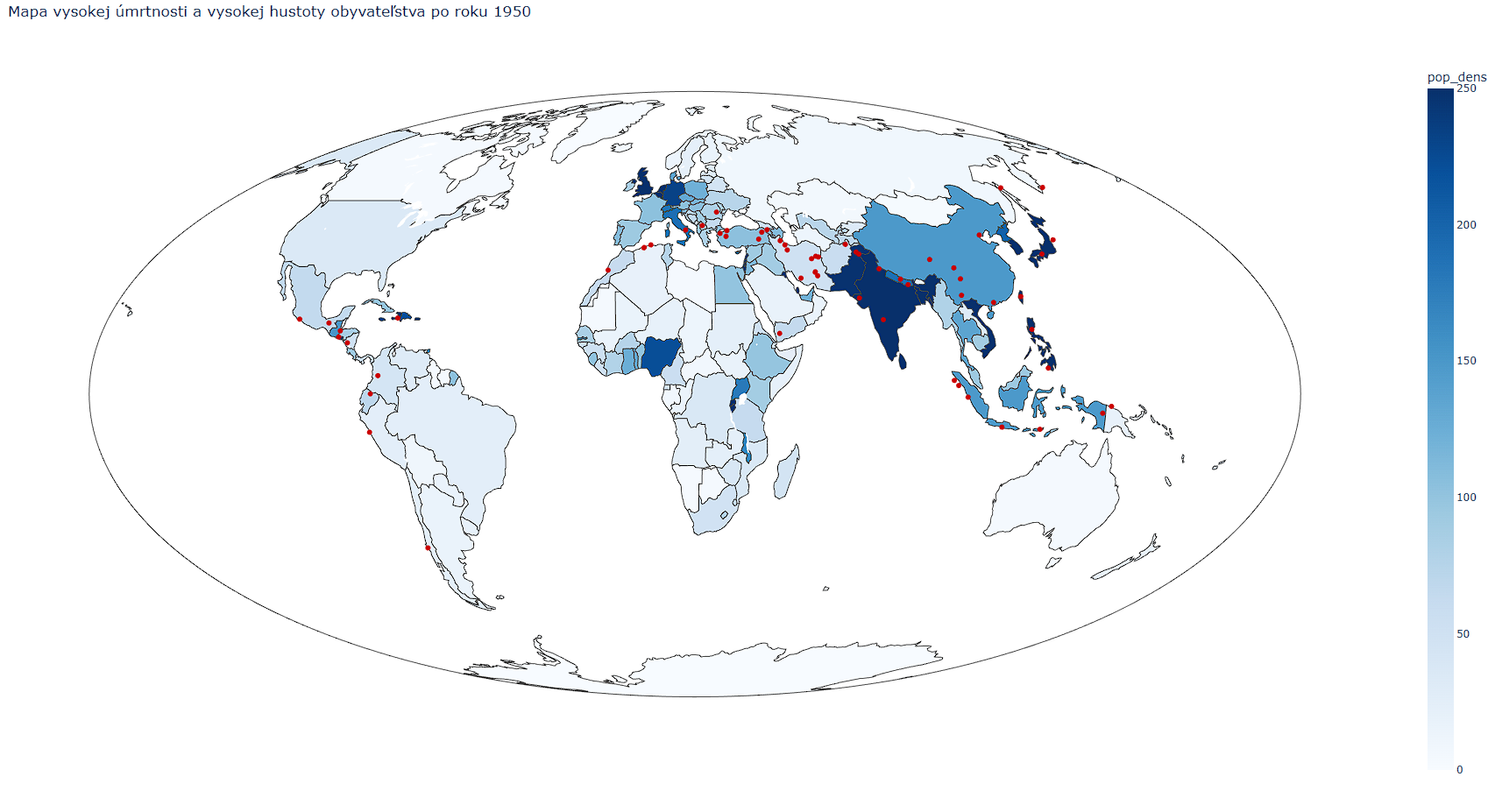
**Mapa nízkej úmrtnosti vzhľadom na hustotu obyvateľstva po roku 1950**



Očakávali sme, že sa tieto zemetrasenia budú orientovať niekde v okolí málo zahustených krajín. Táto teória nebola úplne pravda . Po vykreslení grafu s mapou a hustotou populácie pre danú krajinu vyznačenou tmavšou a tmavšou farbou sme zistili, že veľké množstvo týchto zemetrasení sa vyskytuje aj pri krajinách so skoro najväčšou hustotou populácie pre celý svet. Po bližšom vyšetrení grafu sme zistili, že prevažná väčšina zemetrasení pri krajinách s veľkou hustotou populácie malo epicentrum mimo pevniny, čo teda asi znamená, že epicentrum zemetrasenia nezasiahlo práve veľké mesto ale časť oceánu pri danej krajine z čoho dedukujeme, že zemetrasenia mimo krajiny nemusia mať až taký veľký dopad na pevninu. Samozrejme na tomto grafe vidno aj zemetrasenia, s veľkou magnitúdou ale malými úmrtiami, ktoré sa diali na pevnine. Tieto zemetrasenia sa vyskytovali väčšinou v krajinách s menšou hustotou populácie. Z toho sa dá vyvodiť, že dané zemetrasenia predpokladane malo epicentrum mimo mest v danej krajine. Netreba však zabúdať, že niektoré tieto na prvý pohlaď “oceánske” zemetrasenia môžu mať epicentrum na ostrove alebo blízko ostrova ktorý nevidno, kvôli nedokonalosti vykreslovania grafu. Väčšinou tieto “oceánske” zemetrasenia boli tiež iba v blízkosti daných “neviditelnych” ostrovov.

Nato aby sme tieto naše vyvodenia potvrdili vykreslili sme si další graf ktorý oproti předešlému grafu zobrazuje zemetrasenia s najväčšou možnou úmrtnosťou, čo je viac ako 1001 ľudí. Obmedzenie na roky použijeme aj v tomto grafe.

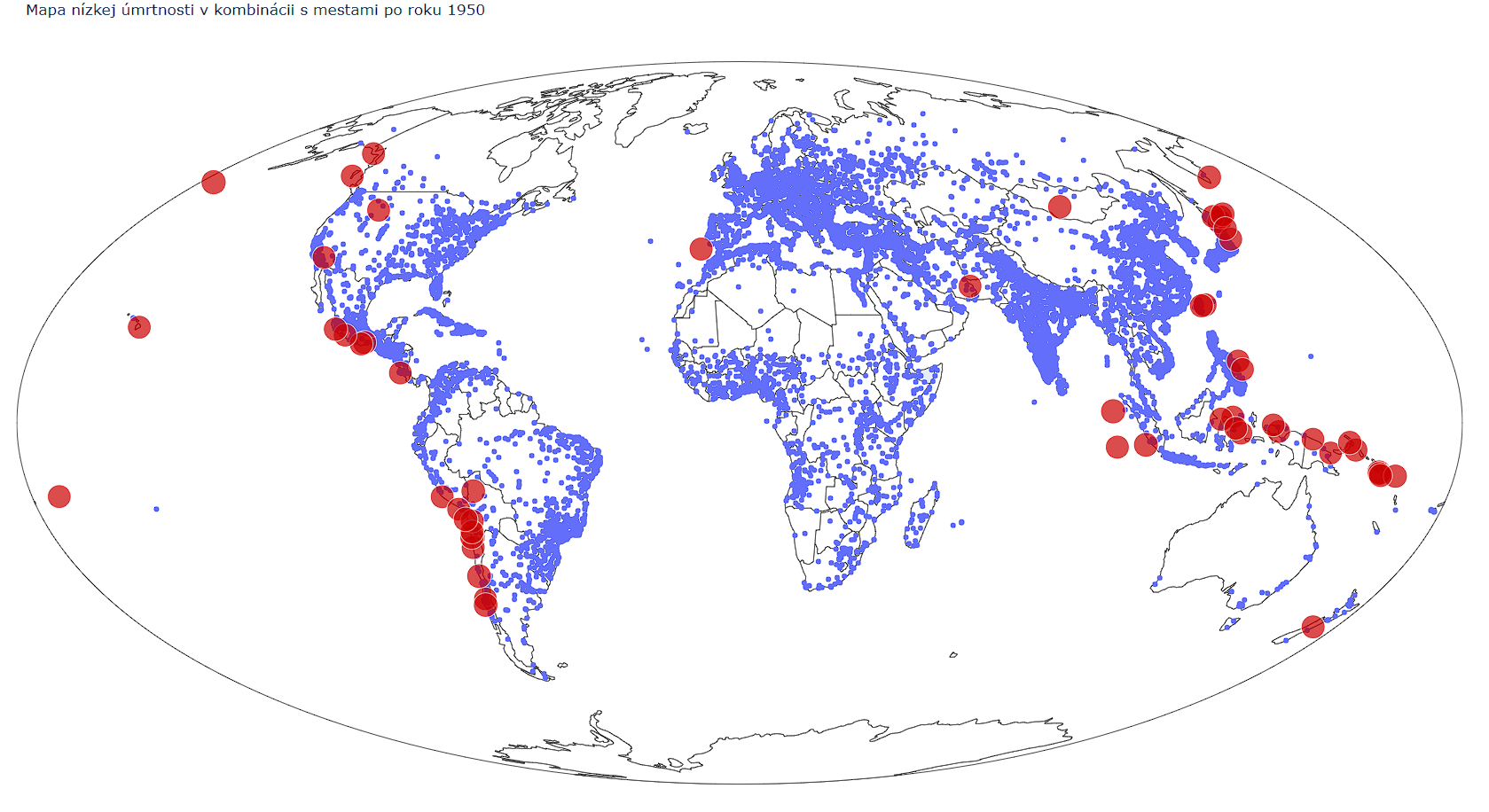
**Mapa vysokej úmrtnosti vzhľadom na hustotu obyvateľstva po roku 1950**



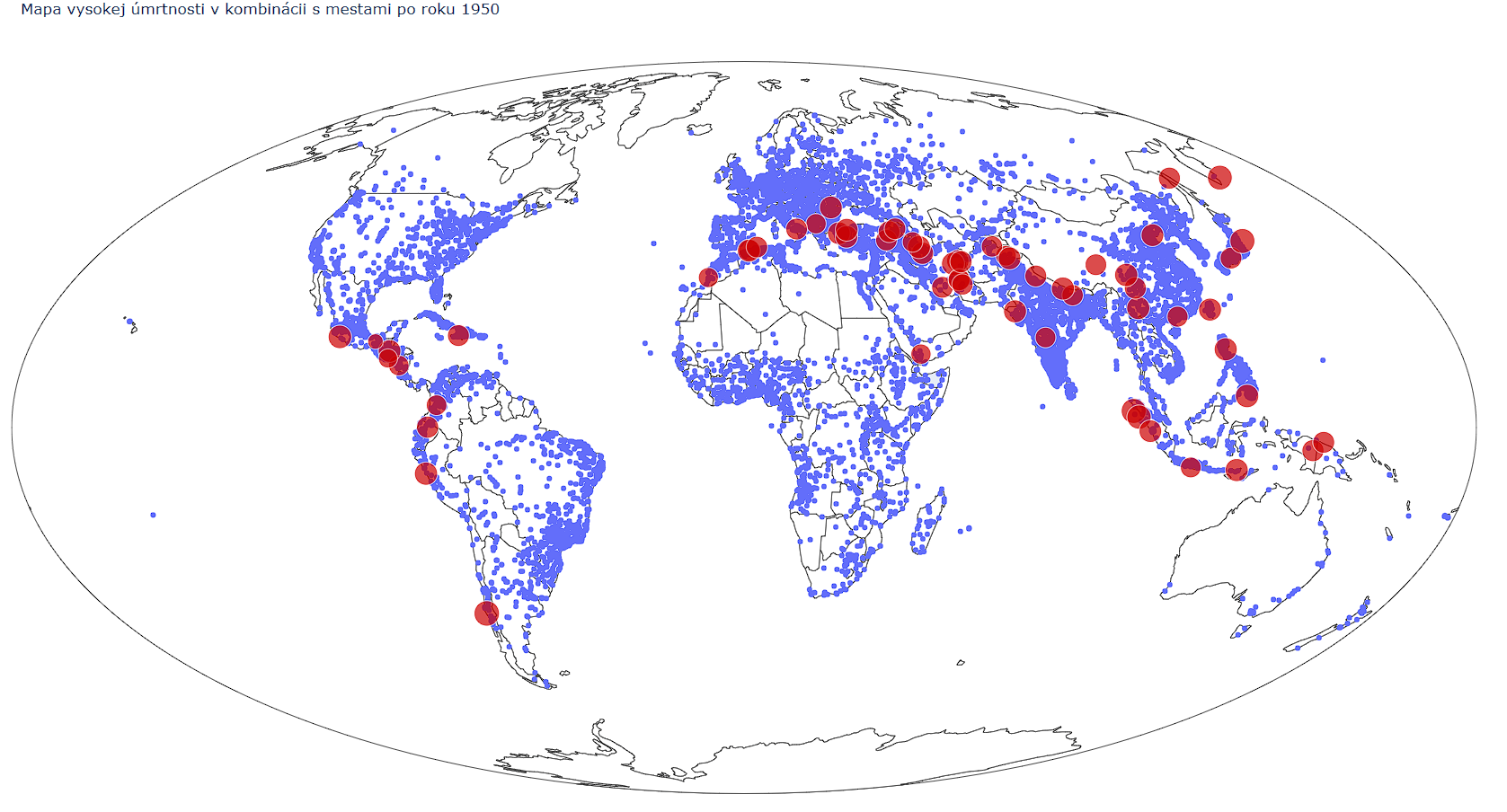
Tu môžeme pozorovať, že tieto zemetrasenia mali epicentrá prevažne vo vnútrozemí. To môže znamenať, že sa epicentrá týchto zemetrasení s veľkou úmrtnosťou orientovali v okolí veľkých miest alebo v samotých mestách krajín s veľkou hustotou zaľudnenia. Na tomto grafe si môžeme dokonca všimnúť, že skoro žiadne so zemetrasení sa nenachádza v mori/oceáne, čo nám potvrdzuje myšlienku z predošlého grafu, že zemetrasenia v oceánoch nemusia mať až taký veľký vplyv na súš. Pár epicentier však pozorujeme rovnakých ako pri zemetraseniach s vysokou magnitúdou ale malou úmrtnosťou. Pri týchto sa môžeme domnievať, že sa jednalo o iné vplyvy, ktoré tento jav spôsobili.

Po následnom vykreslení si máp so zemetraseniami s veľkou úmrtnosťou a malou úmrtnosťou so zvýraznenými mestami s populáciou väčšou ako 50 tisíc, môžeme pozorovať akési potvrdenie našej hypotézy.

**Mapa nízkej úmrtnosti v kombinácii s mestami po roku 1950**



**Mapa vysokej úmrtnosti v kombinácii s mestami po roku 1950**



Zemetrasenia s väčšou úmrtnosťou sú v okolí mest s viac ako 50 tis.obyvateľmi a pre nízku úmrtnosť práve naopak no nie je to vždy pravidlom.

**Záver**

Odpoveď na našu otázku by mohla znieť nasledovne:

Zá

vislosť úmrtnosti od magnitúdy nezáleží až tak moc od veľkosti magnitúdy ako skôr od lokality daného epicentra. Po našej analýze by sme mohli povedať teda, že zemetrasenia s epicentrami mimo pevniny neprispievajú až tak k úmrtnosti zapríčinenou zemetrasením bez ohľadu na ich veľkosť, ako naopak zemetrasenia kľudne aj s nižšou magnitúdou ale na pevnine. Dalo by sa teda povedať že úmrtnosť záleží od lokality zemetrasenia v závislosti od toho aká veľká je magnitúda.

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

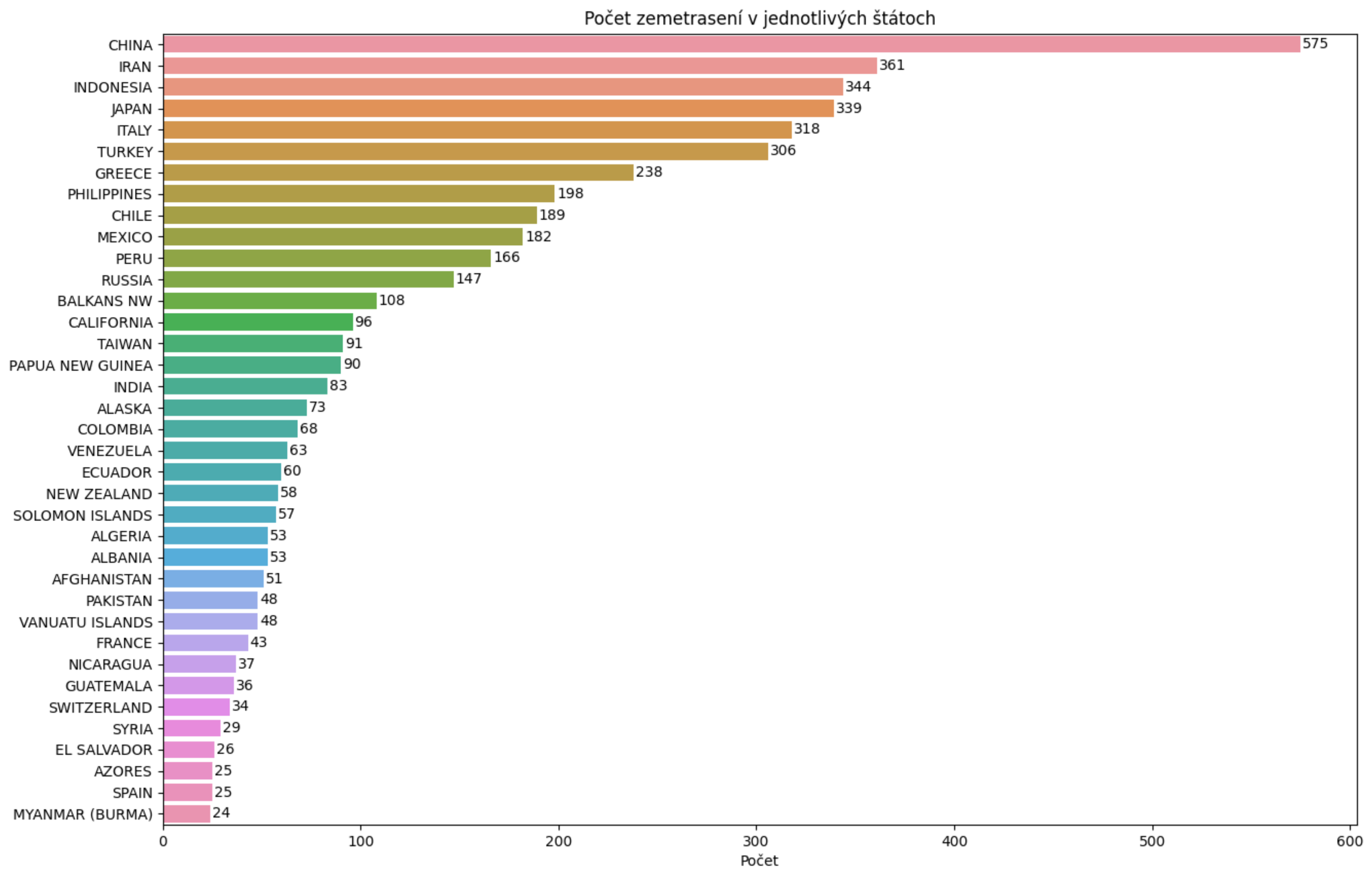
## 

## 2. [Otázka: Kde a kedy sa zemetrasenia diali/dejú?](#_heading=h.2et92p0)

**Analýza**

Po analýze predošlej otázky nás zaujalo kde sa vlastne samotné zemetrasenia dejú. V predošlej analýze sme pracovali s vykreslovaním dát na mapu no s jednou podmienkou, že  dané vykreslované zemetrasenia sa museli udiať po roku 1950. Preto nás zaujalo aký štát je celkovým vedúcim v zemetraseniach na jeho území za celý čas, v respektíve zo všetkých dát s ktorými pracujeme.

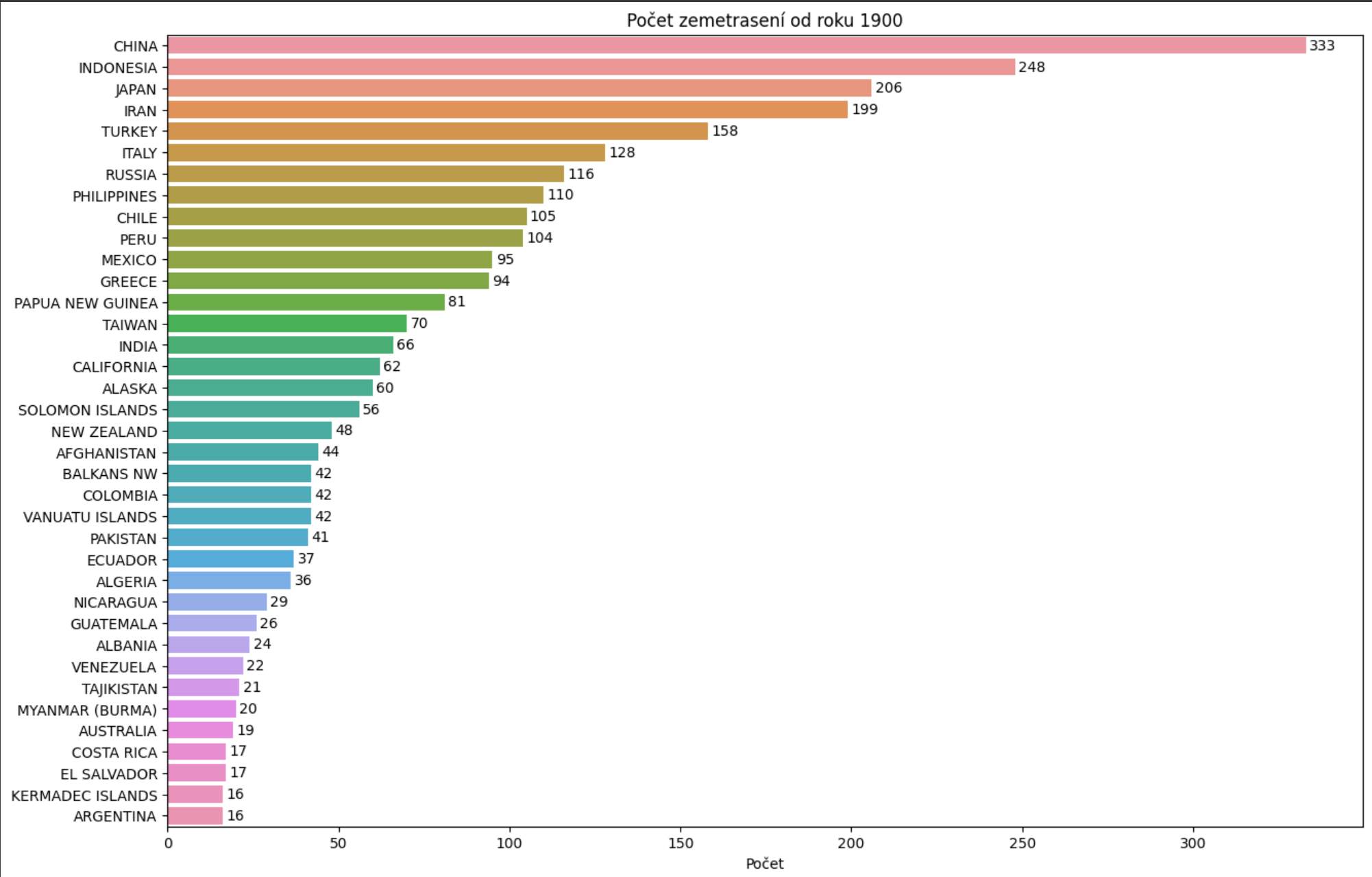
**Barplot celého datasetu**

****

Po vykreslení jednoduchého bar plotu sme zistili, že vedúcou krajinou v počte zemetrasení na jeho území je Čína s približne 575 zemetraseniami. Následne nás zaujalo, či sa rebríček lídrov krajín v zemetraseniach na ich území zmenil v nejakom rozumnom rozpätí rokov, s ktorým by sme vedeli aj ďalej pracovať.

Vykreslili sme si ďalší bar plot ktorý teda reprezentoval lídrov krajín v zemetrasenia na ich území od roku 1900.

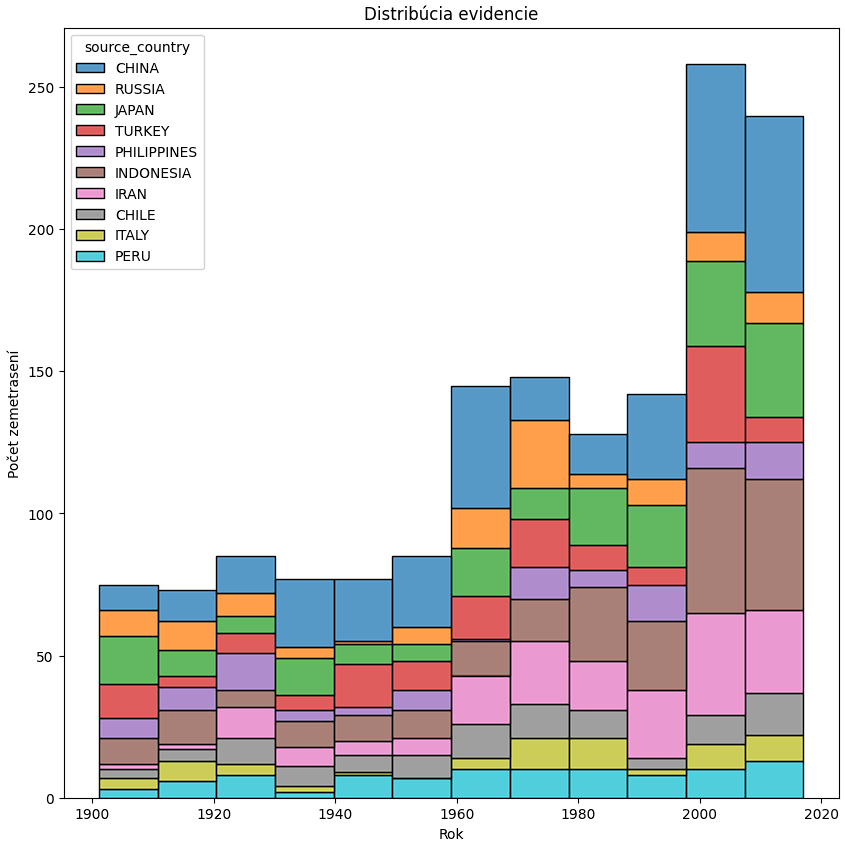
**Barplot po roku 1900**

****

Ako si môžeme všimnúť líder rebríčku sa nezmenil no jeho hodnota klesla zo 575 na približne 333, čo teda indikuje, že sa od roku 2150 pred kristom do roku 1900 stalo na území Číny udialo 242 zemetrasení. Toto zistenie je veľmi divné, no samozrejme môžeme predpokladať že vedci a teda aj náš dataset nie sú schopný zaznamenať každé jedno zemetrasenie ktoré sa stalo a nie je o ňom nejaký záznam z histórie preto tvrdiť, že sa v Číne od roku 2150 pred kristom do 1900  stalo iba 242 zemetrasení je totálny nezmysel a človek si nemôže myslieť že sa prevažná väčšina zemetrasení v Číne stala za posledných 123 rokov. Zároveň si môžeme všimnúť výmeny pozícii v top 4 krajinách a to teda napríklad Irán sa posunul z 2. pozície na 4.. Čo teda znamená, že posledných 123 rokov sa stalo viac zemetrasení na území Indonézie a Japonska ako na území Iránu, no za celý čas má Irán viac zaznamenaných zemetrasení ako Japonsko a Indonézia, Samozrejme tvrdiť toto s úplnou istotou nie je možné z dôvodu ktorý sme poznamenali s Čínou.

Ďalej sme si dali vykresliť nárast zemetrasení vzhľadom na roky v jednotlivých top 10 štátoch od roku 1900.

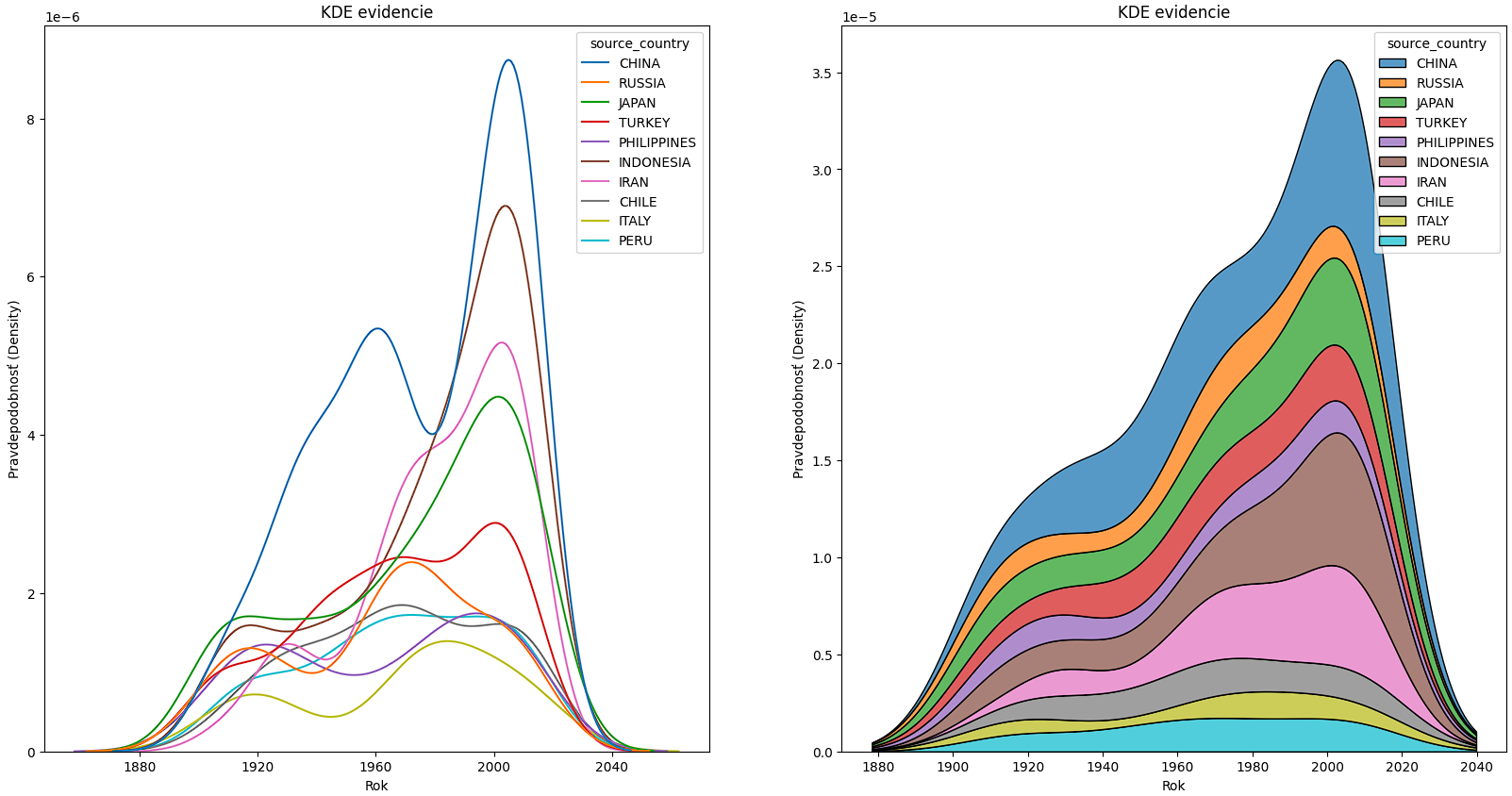
**Distribúcia evidencie**



Tu môžeme pozorovať akýsi nárast počtu evidencii zemetrasení s nárastom rokov pre jednotlivé krajiny. Toto s najväčšou pravdepodobnosťou vzniklo nie na základe toho že by sa začalo diať omnoho viac zemetrasení ale na základe vývoja nových spôsobov a prístrojov na presnejšie meranie a predikovanie zemetrasení.

Na základe tohto zistenia sme teda vykreslili KDE evidenciu top 10 štátov vzhľadom na zemetrasenie. KDE evidencia je len taká vyhladená reprezentácia predošlého histplotu.

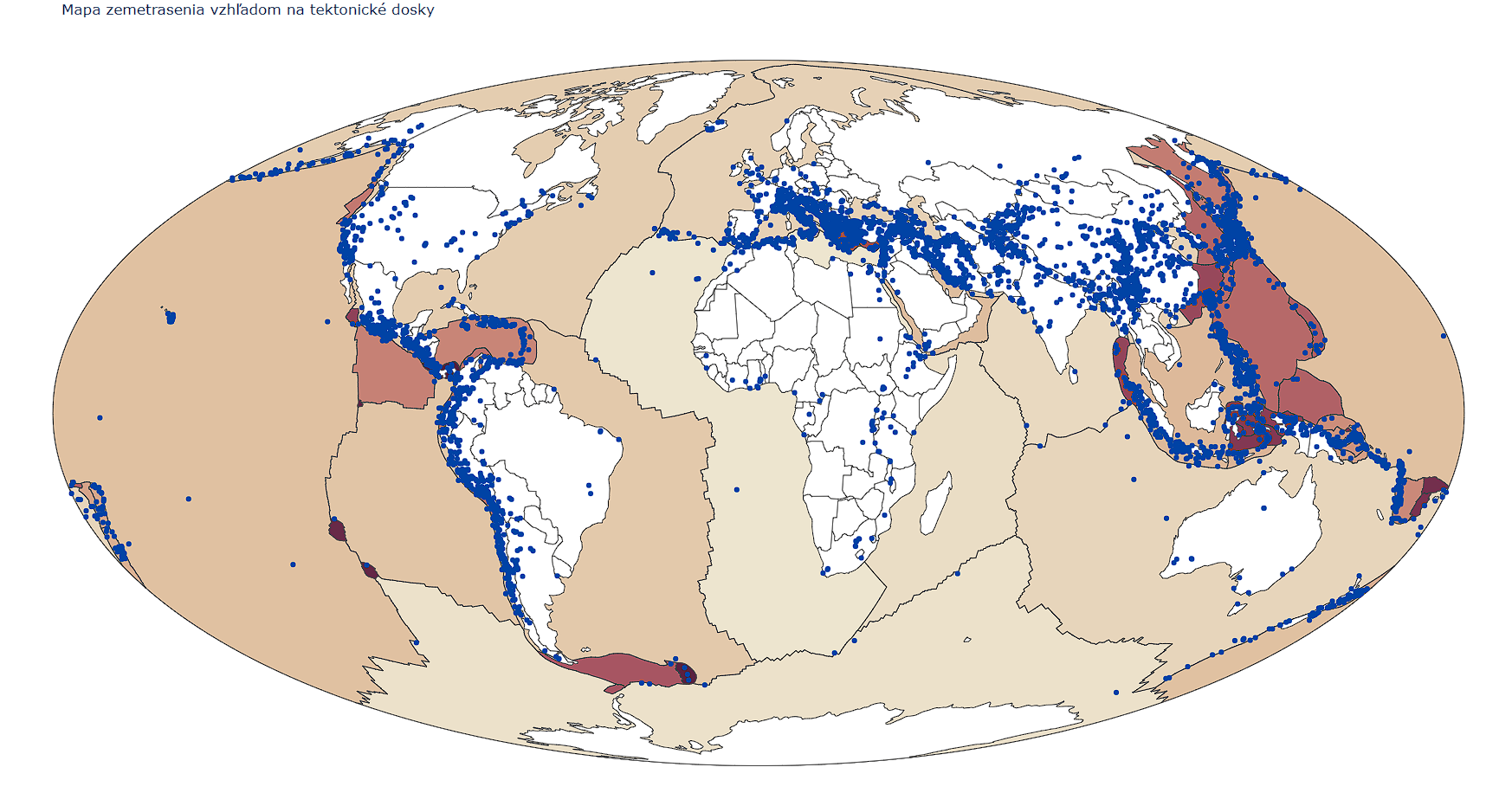
**KDE evidenica**



Tu si môžeme všimnúť viacero zaujímavostí. Prvá zaujímavá informácia je že sa tu ukazuje už z predošlo vykreslených grafov zmena rebríčku pre dané štáty, čo je znázornené tým že jedna čiara ma zrazu Y hodnotu väčšiu ako druhá. Druhá vec, pri pozorovaní grafu sme si všimli zlomy pri viacerých štátoch v určitom období. Toto obdobie bolo okolo roku 1940 čo je obdobie 2. svetovej vojny. Domnievame sa teda, že evidencia zemetrasení môže byť ovplyvnená aj konkrétnou politickou, ekonomickou alebo napríklad pandemickou situáciou sveta, kedy vlastne daná krajina nemá čas ani prostriedky na to aby sa zameriavala na meranie zemetrasení.

Následné nám napadlo, či lokalita danej krajiny nemá niečo spoločné s počtom zemetrasení. Vykreslili sme dané zemetrasenia na mapu na ktorej sú zobrazené tektonické platne.

**Mapa zemetrasenia vzhľadom na tektonické dosky**



Všimli sme si, že najviac zemetrasení sa nachádza na prelomoch 2 platní, no nie je to pravidlom. Zároveň po pátraní na internete sme zistili prečo je možné že Čína má tak veľa zemetrasení. Jedným z indikátorov je to, že v oblasti Číny sa nachádza veľa platní, čo znamená výskyt viac možných dosiek čo sa môžu navzájom trieť. Ďalšie zistenie bolo to, že v tejto oblasti sa dosky takzvane zbiehajú, čo znamená, že sa narážajú čo tiež môže mať väčší dopad na výskyt zemetrasenia ako to keď sa od seba dosky rozchádzajú, čo dokonca u nás na grafe vidno. Juhoamerická doska (index 7) a Africká doska (index 0) sa navzájom rozbiehajú. Na našom grafe vidno, že v tomto okolí sa nevyskytuje skoro žiadne zemetrasenie.

Následne sme pre lepšiu reprezentáciu vykreslili graf s znázornenými zemetraseniami vzhľadom na tektonické dosky len pre top 10 krajín rebríčku.

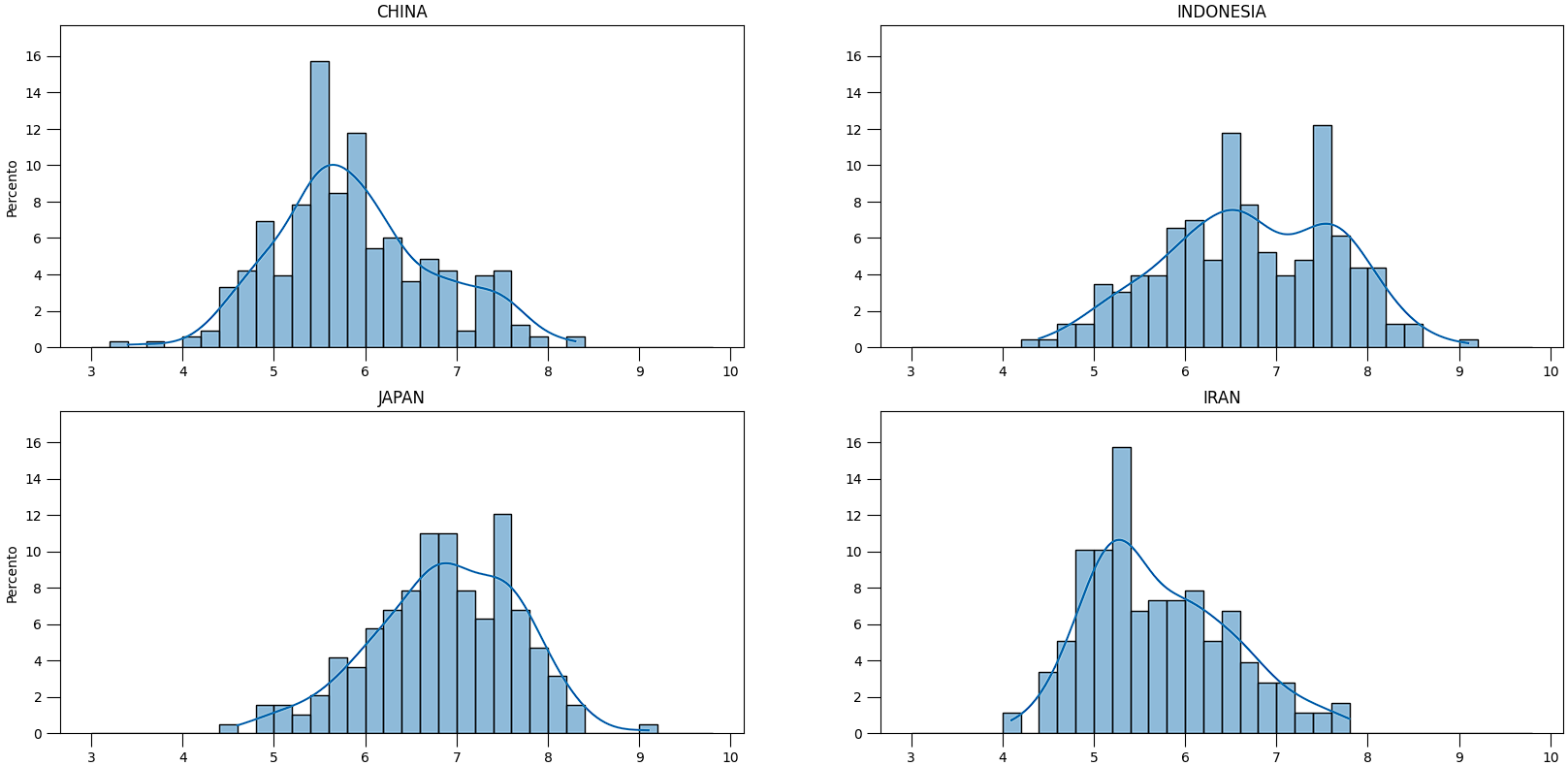
**TOP 10 krajín v počte zemetrasení v porovnaní s tektonickými doskami**

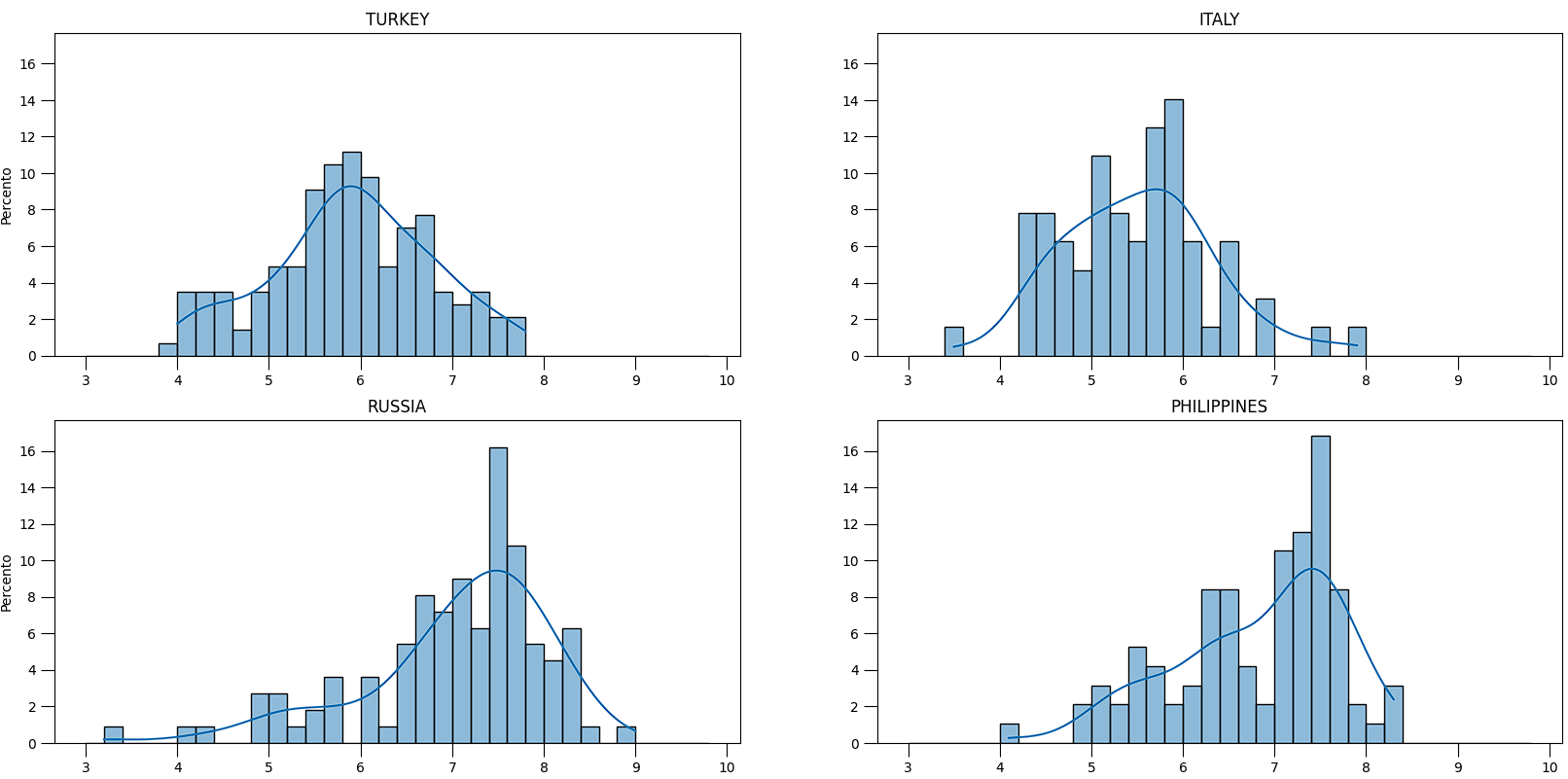


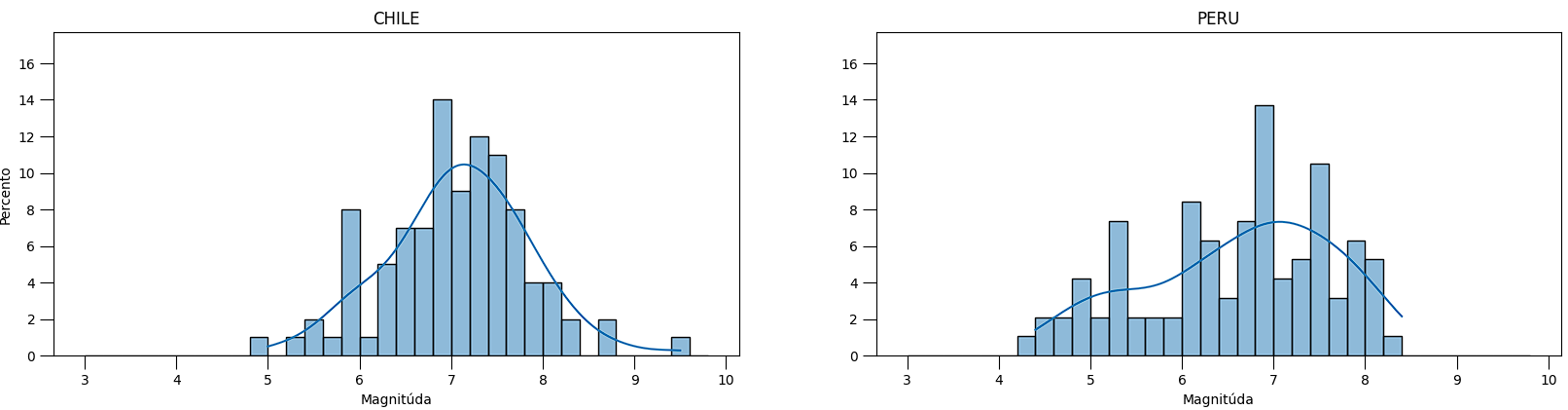
Toto sme spravili len pre lepšiu reprezentáciu a orientáciu zemetrasení v top 10 krajinách aby sme sa teda lepšie vedeli orientovať o ktorých zemetraseniach sa rozprávame v porovnaní so všetkými zemetraseniami.

Následne nás zaujalo v akom počte sa pre top 10 krajín vyskytlo zemetrasenie s danou magnitúdou.

**Zastúpenie magnitúdy v TOP 10 krajinách**



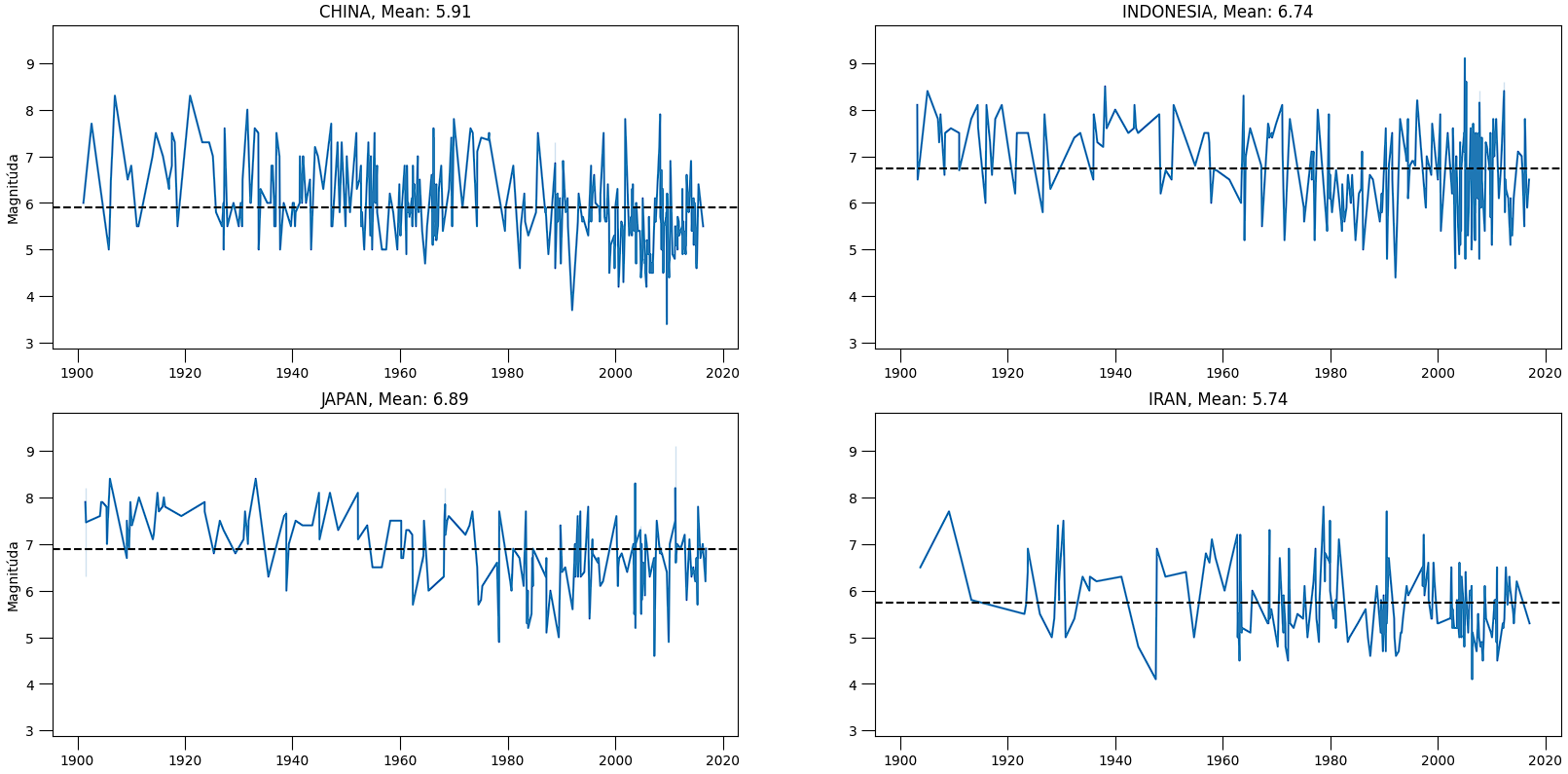


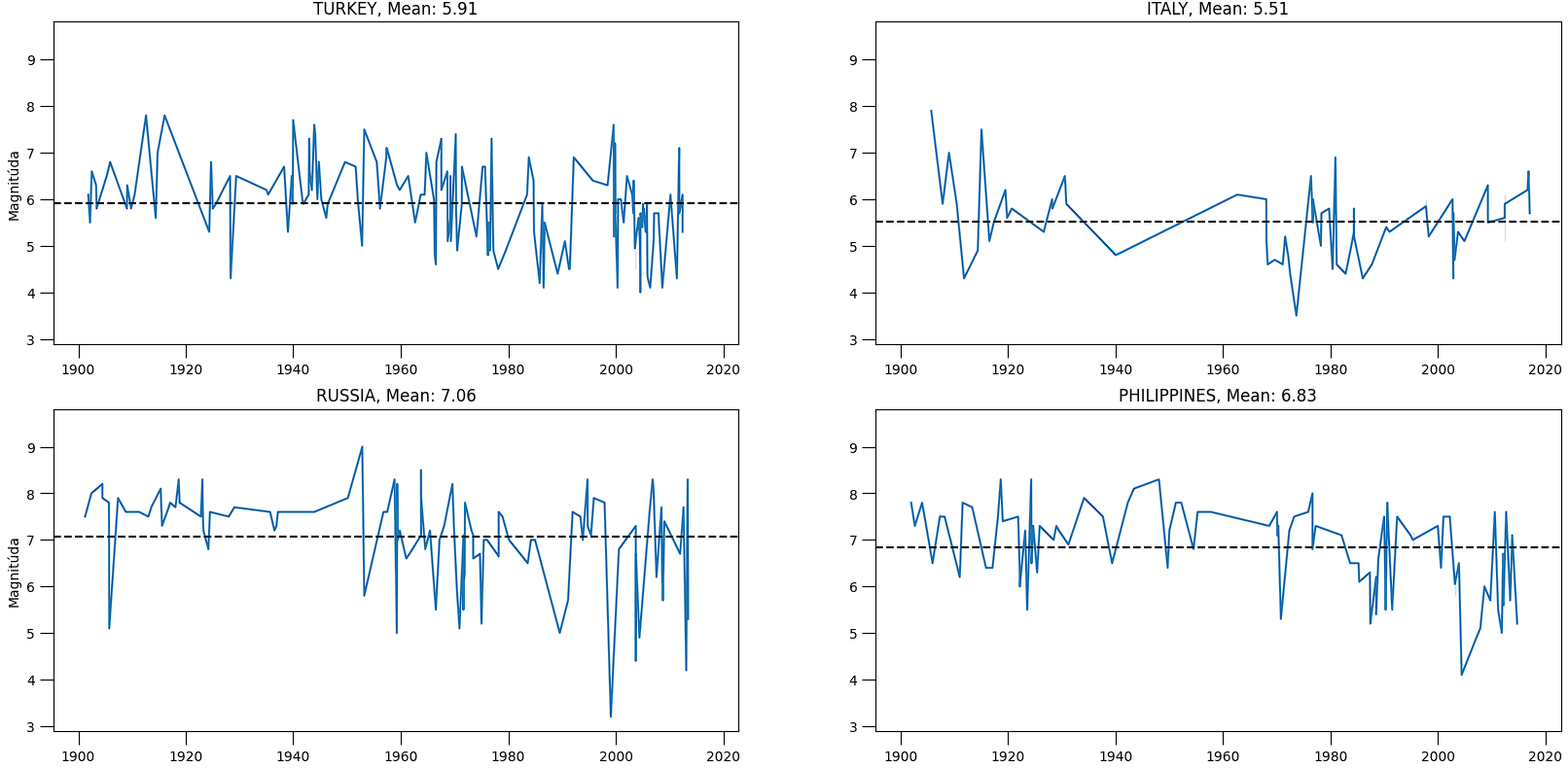


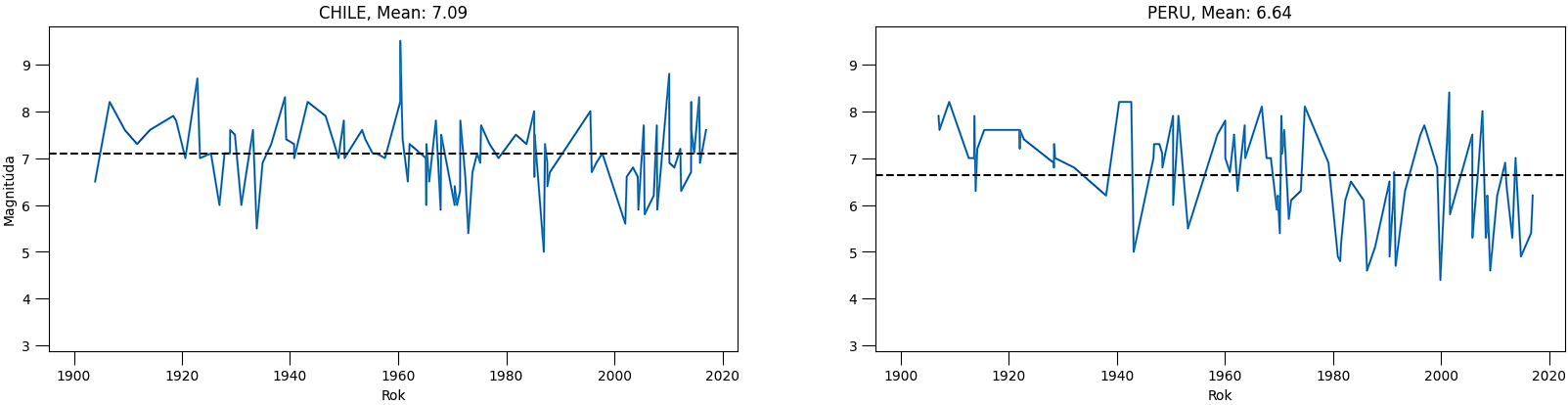
Tu môžeme pozorovať totálnu nezávislosť medzi krajinami. Každá krajina má iný počet zastúpení počtu zemetrasení s danou magnitúdou. No práve toto zistenie nás informuje o tom, v akej krajine sa prejavujú zemetrasenia s akou magnitúdou. Čiže vieme povedať v akej krajine sa udiali zemetrasenia s najhoršou magnitúdou. Samozrejme si treba uvedomiť, že sme tieto hodnoty reprezentovali v percentách čiže hovoríme, koľko percent zemetrasení malo akú magnitúdu, teda rovnako veľký bar nemusí reprezentovať rovnaký počet zemetrasení s danou magnitúdou.

Pri zistovaní týchto informácii nás zároveň zaujalo ako sa tieto informácie vyvíjali počas rokov 1900 až 2020 a priemer ohraničený týmito rokmi.

**Vývin magnitúdy v čase**

****

****

****

Tu si môžeme všimnúť, že každá krajina má svoj vlastný nezávislý vývoj počas rokov, nemôžeme povedať, že by sa nejaké 2 grafy podobali. To teda potvrdzuje, že každá krajina je nezávislá v počte zemetrasení a teda aj vo vývoji počas rokov čo sa týka zemetrasení.

**Záver**

Odpoveď na našu otázku kde a kedy sa zemetrasenia dejú by sme mohli formulovať nasledovne:

Výskyt zemetrasení vzhľadom na štáty záleží hlavne na tom, kde sa daný štát nachádza vzhľadom na tektonické dosky. Zároveň po vypátraní informácii z viacerých zdrojov sme zistili, že pohyb daných dosiek dosť závisí na tom či vyvolávajú viac zemetrasení alebo nie. Čiže by sa dalo povedať, že na polohe záleží. Taktiež treba podotknúť, že náhly nárast v počte zemetrasení za posledných 120 rokov môže byť zapríčinený tým, že sa zlepšili metódy merania zemetrasení a teda hovoriť o momentálnych top štátoch ako o lídroch všetkých čias je špekulatívne ale s predpokladom, že za týchto 120 rokov s dobrými prostriedkami merania stále mali najväčší nárast sa dá teda povedať, že by mohli byť aj celkovými lídrami keď ešte neboli tak dobré meracie prostriedky. Čo sa týka času nárastu meraní zemetrasení, všimli sme si rastúci trend skoro pre každú jednu krajinu. Pri našej analýze sme ale skonštatovali že pri týchto rastúcich trendoch vznikli nejaké výkyvy. Zhodnotili sme, že tieto výkyvy v tomto rastúcom trende mohli byť ovplyvnené nie tým že sa zmenil pohyb tektonických platní ale naopak meranie mohlo byť ovplyvnené nejakým vonkajším podnetom ako napríklad vojnou alebo niečim podobným, kde v týchto prípadoch štát nesustredil úsilie na meranie zemetrasení ale na niečo iné. Zároveň ako už sme spomenuli s časom prišli aj lepšie meracie prostriedky čiže hovoriť o tom, že sa nárast zemetrasení v počte zapríčinil časom je hlúposť. Zemetrasenia tu vždy boli len vedci neboli schopný ich zachytiť v takom množstve ako sme schopný dnes. Zároveň sa možno akési historické záznamy o zemetraseniach nemuseli zachovať. Čiže by sa dalo povedať len to, že čas ako taký je celkom irelevantný faktor čo sa týka pravdivosti dát ako takých. Teda nemôžeme porovnávať informácie spred 1000 rokov s momentálnymi dátami, pretože by výsledky boli totálne nezmyselné. Čiže povedať, že vtedy a tam sa zemetrasenia v histórii diali s istotou povedať nemôžeme no čo sa týka súčasnosti sme schopný povedať kde a aké zemetrasenia sa môže stať.

## 

## 

## 

## 

## 

## 

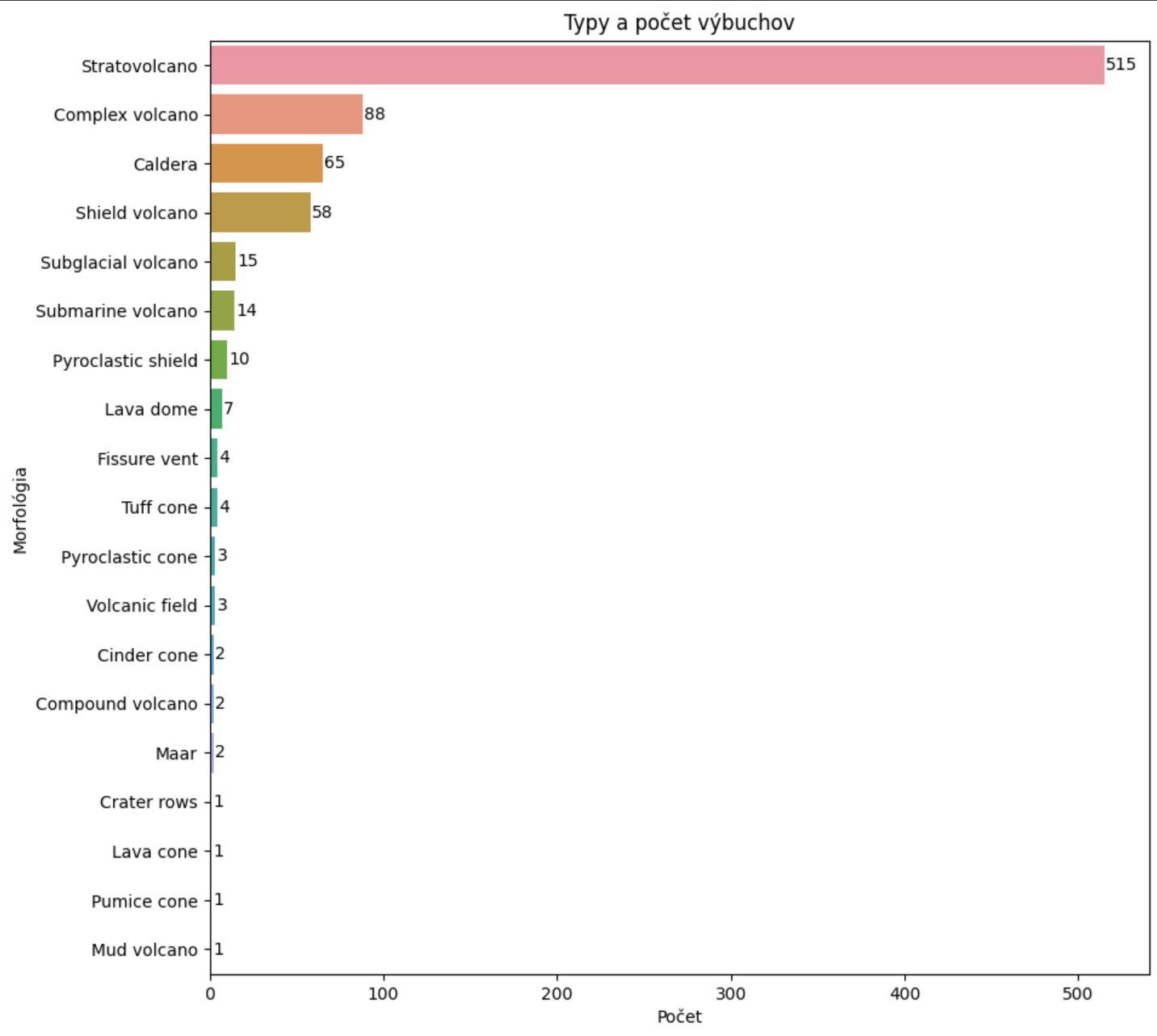
## 

## 3. [Otázka: Aké typy sopiek sú najviac výbušné.](#_heading=h.tyjcwt)

**Analýza**

Začali sme tým, že sme si náš data set dali vykresliť do bar plotu na základe typu danej sopky.

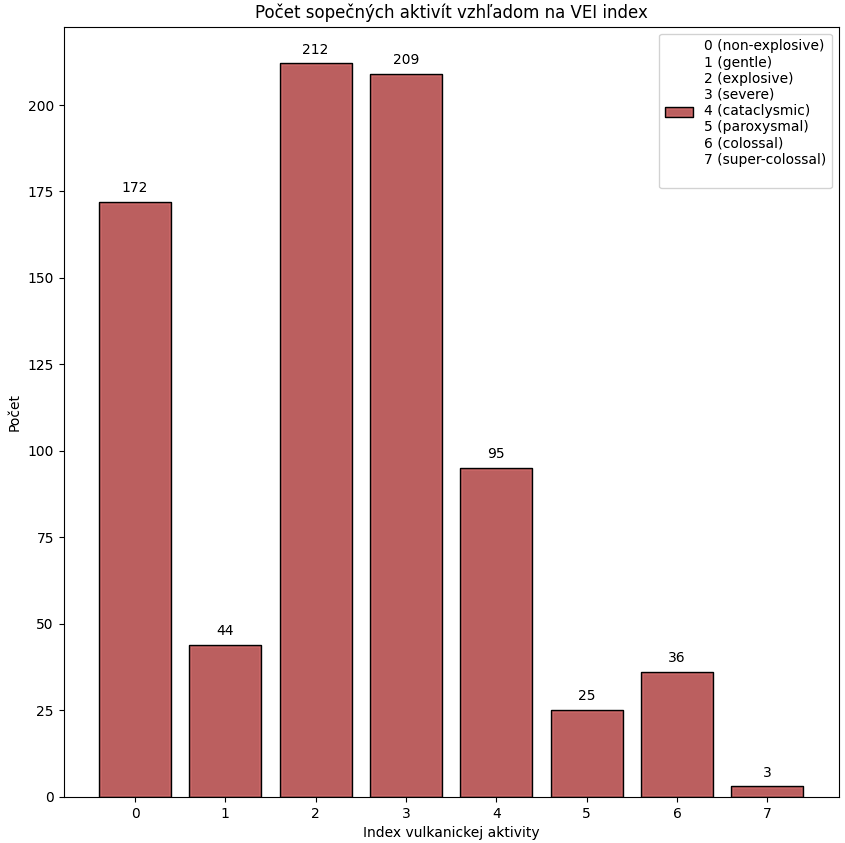
**Barplot**



Zistili sme, že prevažná väčšina typu sopiek je [Stratovolcanoveho](https://en.wikipedia.org/wiki/Stratovolcano) typu následne [Complex Volcano](https://en.wikipedia.org/wiki/Complex_volcano) no s veľmi porovnateľným skokom v jeho počte.  Veľmi malé zastúpenie malo napríklad [Mud volcano](https://en.wikipedia.org/wiki/Mud_volcano), [Pumice cone](https://www.britannica.com/science/pumice-cone), [Lava cone](https://en.wikipedia.org/wiki/Lava_cone) atď. čo priamo vidno z grafu.

Následne nás zaujala distribúcia sopiek vzhľadom na VEI index. Na to aby sme sa mohli touto otázkou zaoberať, musíme pochopiť čo [VEI](https://en.wikipedia.org/wiki/Volcanic_explosivity_index) index je. [VEI](https://en.wikipedia.org/wiki/Volcanic_explosivity_index) index je skratkou pre Volcanic Explosivity Index. Čiže je to index vulkanickej aktivity pre klasifikaciu intenzity sopečnej činnosti, založený na kvantitativnom prístupe, jeho škála sa pohybuje na škále od 0 až po 7. Takže inak povedané naša otázka sa zaoberá tým ako moc ovplyvnená je vulkanická aktivita na základe iných faktorov.

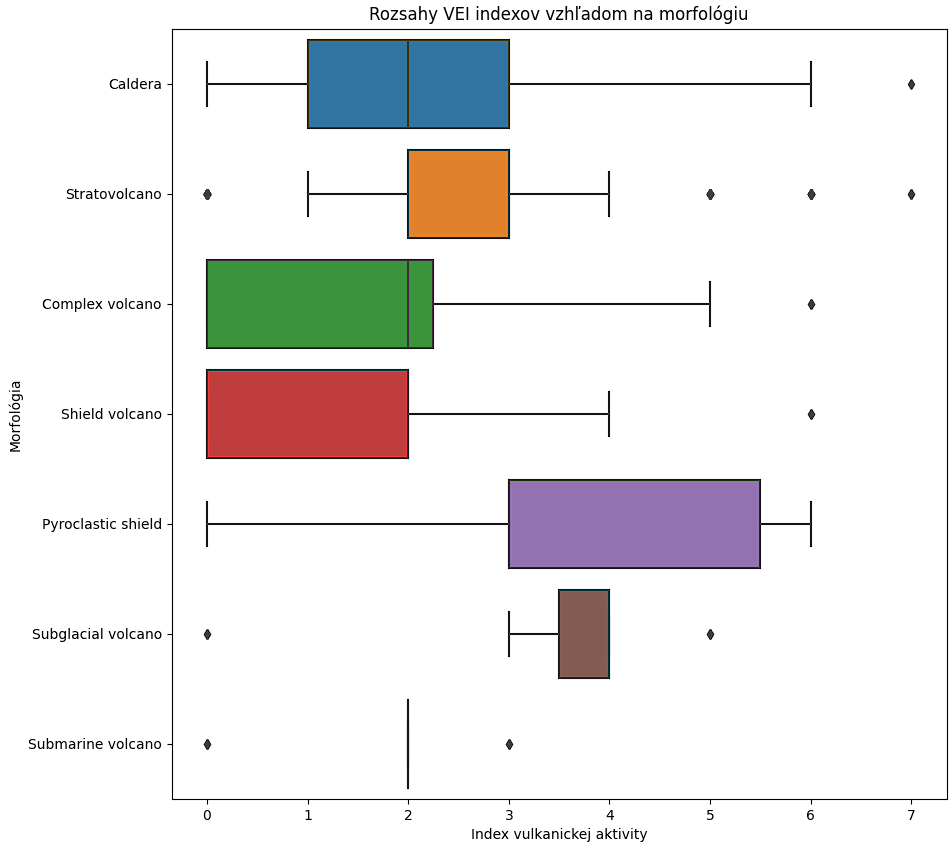
**Počet sopečných aktivít vzhľadom na VEI index**



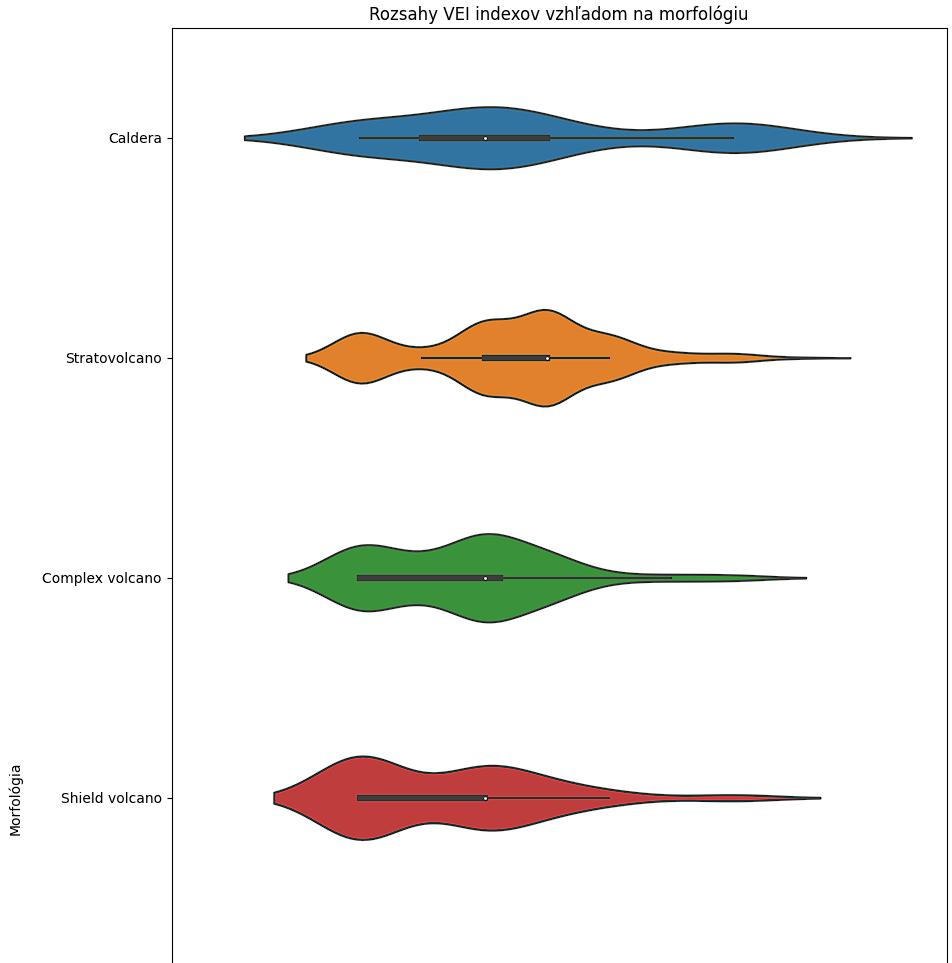
Môžeme si všimnúť, že najväčšie zastúpenie má hodnota VEI indexu 2. Zároveň môžme z pozorovania dedukovať, že hocijaká lineárna závislosť je úplne vylúčená, keďže rozdelenie grafu je totálne náhodné.

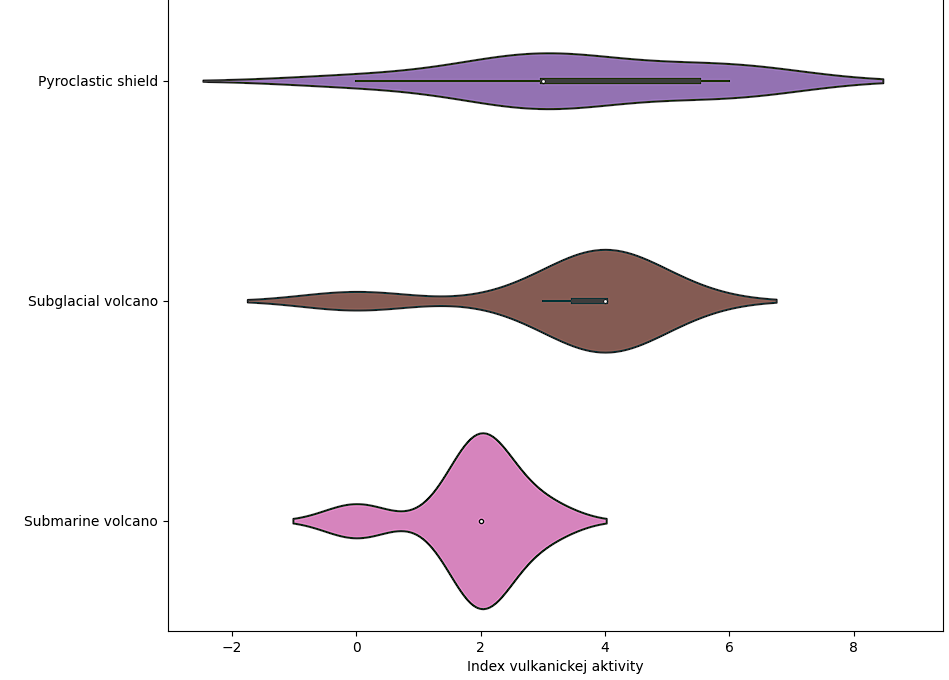
Následne sme sa teda rozhodli nakresliť Violin a Box plot aby sme si vedeli zvizualizovať dané typy výbuchov a ich rozptyl vo VEI indexe. Rozhodli sme sa z tejto časti vizualizacie odstranit prvky s poctom zastúpení menším ako 10.

**Boxplot**



**Violin plot**





Vidíme veľmi nepravidelný pattern. Tu môžeme odôvodniť naše rozhodnutie prečo sme si vybrali iba typy výbuchov s počtom výskytu viac ako 10. Pretože pri vizualizácii cez box plot by mali veľmi divné vyzobrazenia, mohli by byť reprezentované čiarkami a nie samotnými “štvorcami” tak ako sa napríklad náhodou stalo pri [Submarine Volcano](https://en.wikipedia.org/wiki/Submarine_volcano). Zároveň nemajú dostatočný počet výskytov aby sa na nich dalo urobiť nejaké pozorovanie. Ďalej môžeme povedať, že vieme povedať s akým VEI indexom sa môžeme stretnúť pri určitých typoch výbuchov. Napríklad výbuchoch typu [Caldera](https://en.wikipedia.org/wiki/Caldera) alebo [Pyroclastic shield](https://en.wikipedia.org/wiki/Pyroclastic_shield) si môžeme všimnúť, že rozsah VEI indexov pokrýva skoro celú možnú škálu. Čo sa týka violin plotu môžme vidieť ako sú dáta rozdistribuované a približne ako sú kategoricky rozmiestnené

**Záver**

Táto analýza nám moc vzájomných vzťahov nedala ale objasnila nám pár faktov ohľadom sopiek. Analýza nás prinútila zistiť čo je VEI index a pochopiť tomu čo dané hodnoty znamenajú. Ďalej nás oboznámila s  najrozšírenejším typom aktívnych sopiek ktorý teda je Stratovolcano s veľmi veľkým rozdielom v počte od druhej pozície čo je Complex volcano. Z toho môžeme dedukovať,  že Stratovolcano je najaktívnejší typ sopky čo teda znamená, že bývať pri Stratovolcano nemusí byť zrovna dobrý nápad. Následne nám táto analýza poukázala na to, že rozdelenie VEI indexu je náhodné nie je možné si tu všimnúť nejaký rastúci alebo klesajúci trend ktorý by mohol naznačovať že výbuchy s nižším Vei indexom sú viac pravdepodobné na výskyt ako výbuchy s väčším VEI indexom. Následne nám táto analýza ukázala, že aký typ sopky má aký VEI index najviac zastúpený tu by sa dalo povedať, že rozdelenie sopiek je veľmi rozdielne. Patterny sa neopakujú no treba povedať, že na to aby sme niečo takéto ako je teda napríklad opakovaný pattern mohli pozorovať museli by sme mať dataset ktorý má aspoň trochu podobný počet zastúpení pre každý jeden typ sopky. Toto pri našom datasete nie je. Povedať, že by pri rovnakom počte sopiek v každom type nastal nejaký opakovaný pattern je teda asi hlúposť keďže presne pre toto to dané rozdelenie sopiek existuje. Kvôli ich danej špecifikácii. Čo sa teda prejavuje v rôznom počte výbuchov, rôznej explozivite s tým aj teda možnosť rôzneho zastúpenia VEI indexu.

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 4. [Otázka: Závisí počet smrtí od výbuchov?.](#_heading=h.3dy6vkm)

**Analýza pre správu**

Túto analýzu sme sa rozhodli  spraviť na základe predošlej analýzy. Pri predošlej analýze sme sa skôr dozvedeli informácie ohľadom typov sopiek ako takých. No nedozvedeli sme sa  mnoho informácii ako napríklad počet úmrtí pre dané sopky. Či tieto úmrtia sú ovplyvnené už spomínaným VEI indexom, či dokonca záleží od pozície výbuchu. Pre začiatok analýzy sme sa rozhodli vykresliť hist plot ktorý ukazuje percentuálne zastúpenie rôznych VEI indexov v závislosti od počtu úmrtí dokopy. Čiže náš histplot ukazoval aké veľké percentuálne zastúpenie mal VEI index 0 v príspevku napríklad 0 úmrtí.

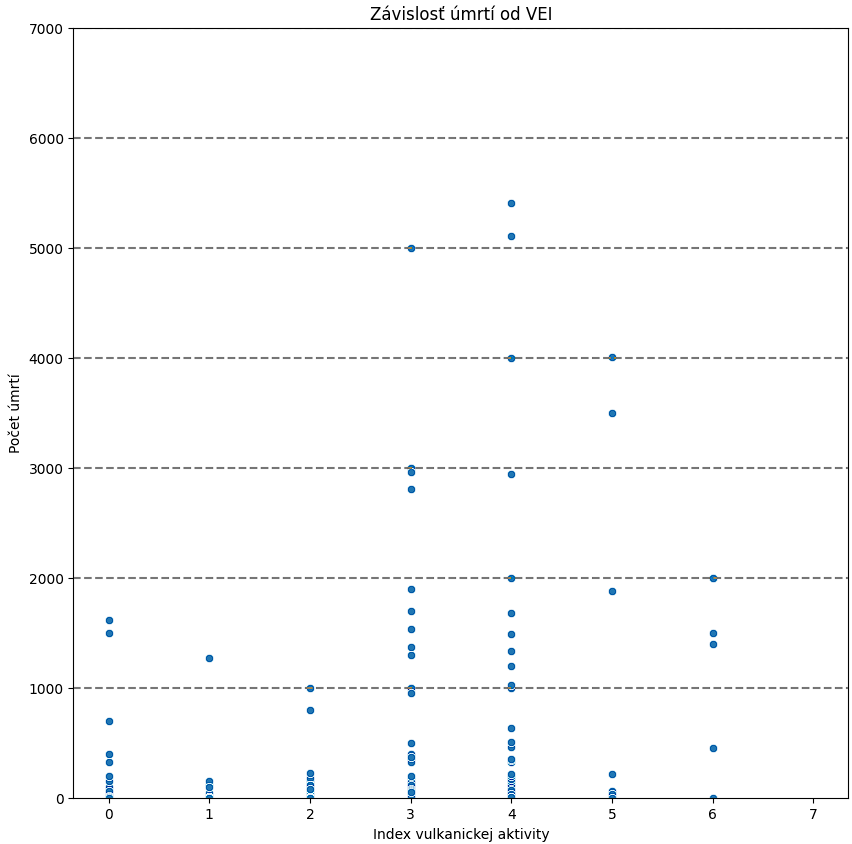
**Histplot**



Môžeme si tu všimnúť, že obrovské zastúpenie tu má úmrtnosť 0, teda v bare 10 su zastúpenia aj viac ako 10 úmrtí ako vidieť toto zastúpenie je mizivé. Čo teda znamená že mnoho sopiek nezabije veľa ľudí. Treba tu poznamenať, že v kontajneri od 9 do 10 sa nachádzajú smrte aj viac ako 10 čiže všetky väčšie ako 10

Následne nás teda zaujalo aké sa samotné výbuchy sopiek prejavili na úmrtiach.

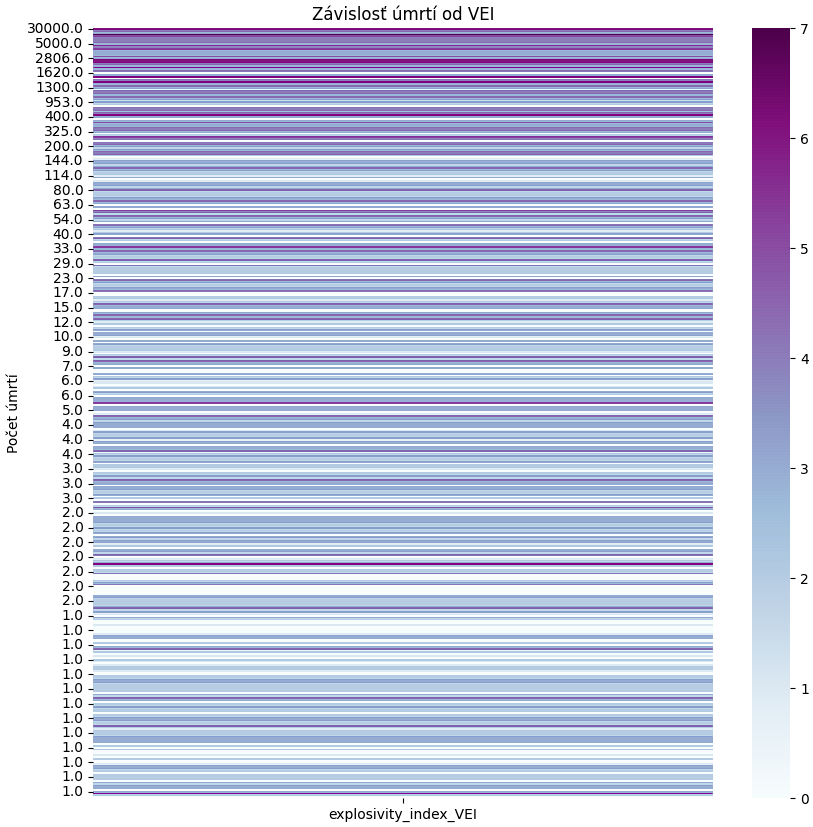
**Scatterplot**



Po zvizualizovaní jednoduchým scatterplotom a upravením limít aby bol graf trochu viac čitateľnejší sme si všimli, že daný graf ma malo grúp a teda zdá sa, že samotný VEI index priamo nesúvisí s úmrtnosťou. Vieme toto tvrdenie potvrdiť až pri vyšších hodnotách kde si môžeme všimnúť aspoň trochu nárast rozptylu hodnôt.  No táto reprezentácia nebola moc dobrá.

Pre lepšiu reprezentáciu dát sme sa rozhodli spraviť heat mapu.

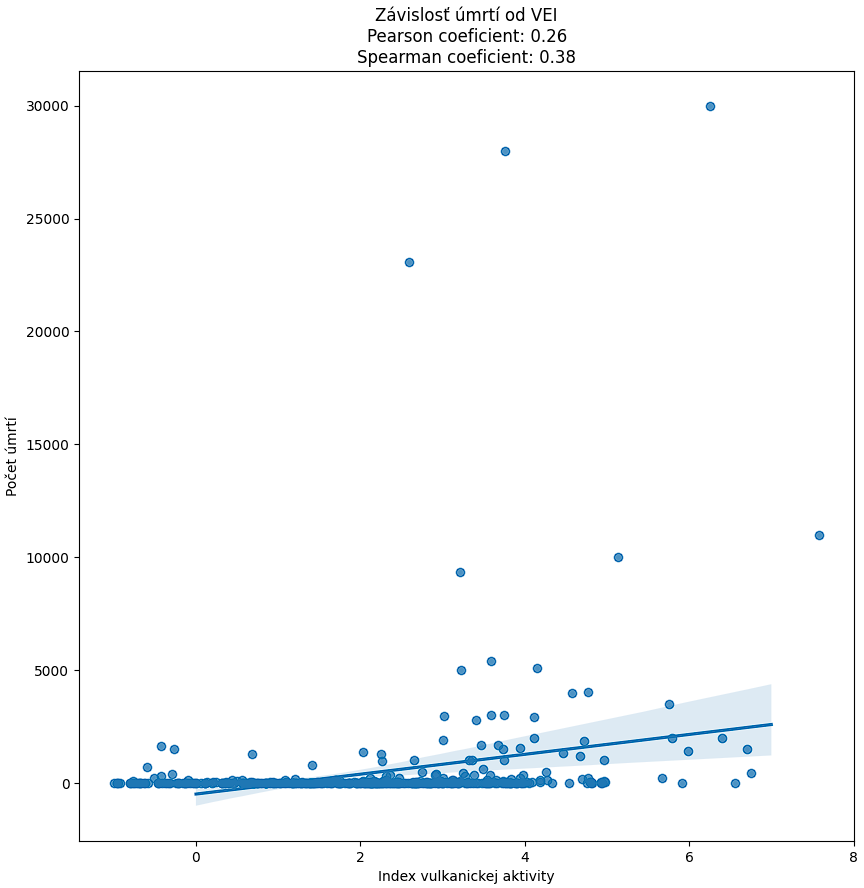
**Heatmapa**



Tu už trochu lepšie vidno že so stúpajúcim VEI indexom sa zväčšuje úmrtnosť ale nie je to pravidlom. Vo vyššej časti grafu prevažuje trochu tmavšia farba ale nedá sa s jednoznaćnosťou povedať, že je tam strašne poznamenatelný rozdiel.

Pre 100% istotu sme sa rozhodli na dané dáta použiť Pearsonov a Spearmanov koeficient.

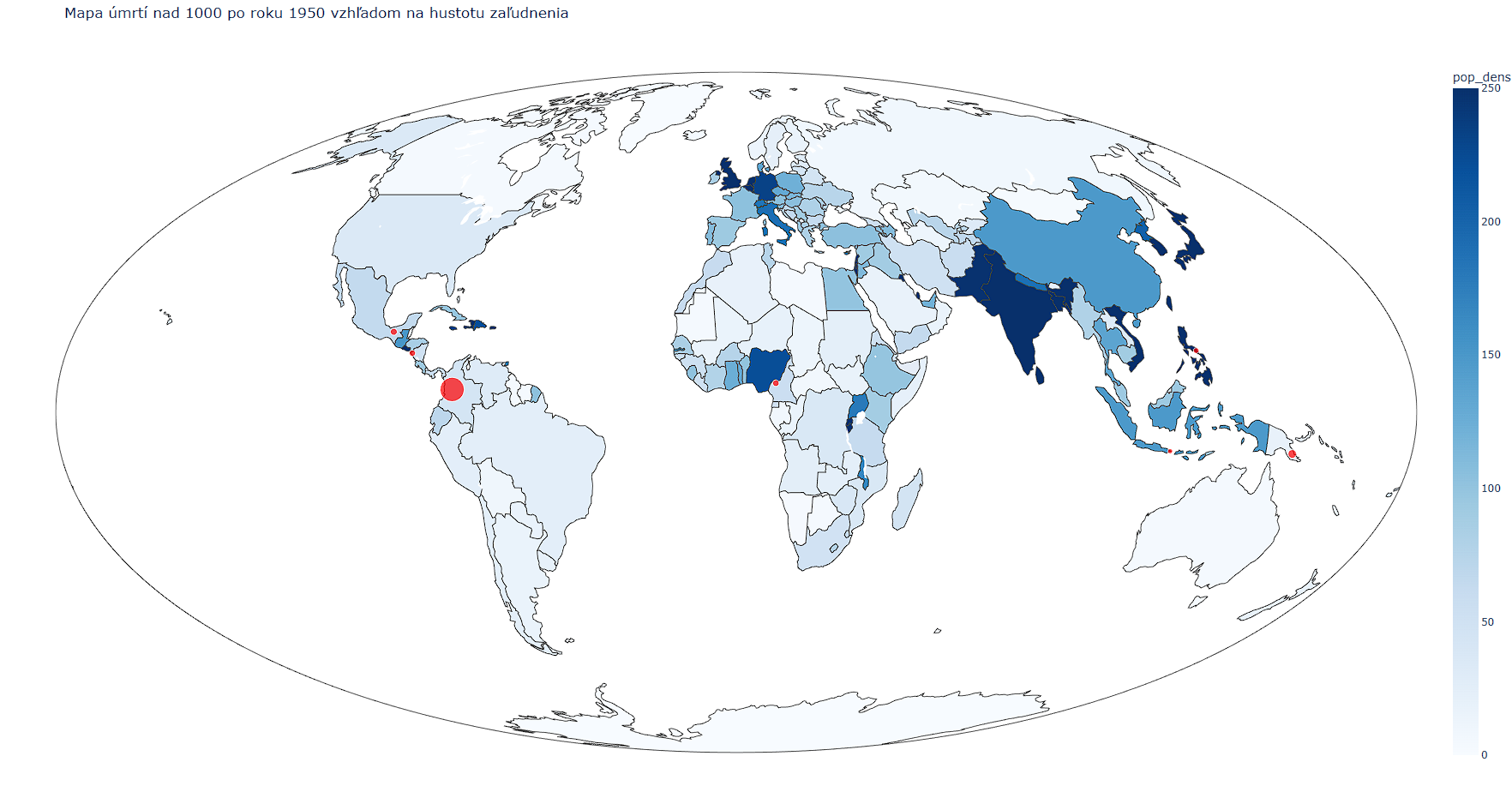
**Pearsonov a Spearmanov koeficient**

****

Zistili sme, že dané hodnoty vyšli pre Pearsona: 0.26 a Spearmana: 0.38. Výpovedná hodnota týchto hodnôt je taká že už teda  vieme asi s istotou povedať, že VEI index spolu s úmrtnosťou teda nemá nič spoločne, inak povedané nemajú spolu žiadnu závislosť, teda hodnoty koeficientov nie su dostatočne veľké na to aby sme s istotou mohli povedať, že úmrtia priamo závisia od VEI indexu.

Na základe predošlého zistenia sme sa pokúsili zistiť či náhodou tieto úmrtia priamo nesúvisia s lokalitou výbuchov, teda s hustotou obyvateľstva v krajine výskytu, s daným VEI indexom. Tak ako v 1. analýze sme sa rozhodli použiť dáta od roku 1950 kvôli rastu/poklesu populácie.

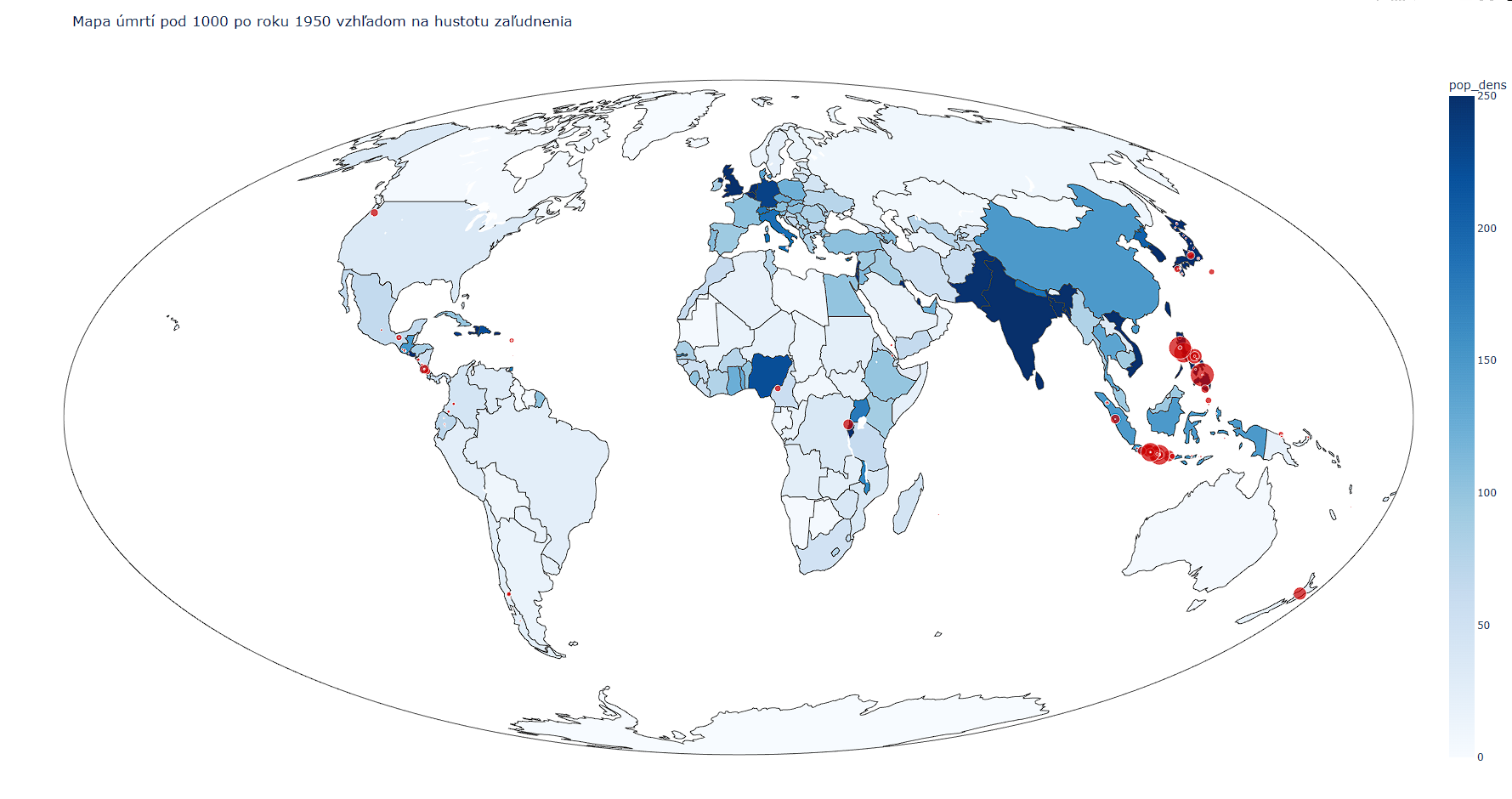
**Mapa úmrtí nad 1000 po roku 1950 vzhľadom na hustotou zaľudnenia**



Po vykreslení bodov, ktoré maju velkosť v závislosti od počtu úmrtí, na mapu sme zistili, že by sa toto tvrdenie s istotou nedalo tvrdiť keďže výbuchy s veľkým počtom úmrtí nenastávali v krajinách s veľkou hustotou populácie čo bol náš logický predpoklad. Dokonca sa nám tu potvrdzuje tvrdenie že ťažko odhadnúť počet úmrtí vzhľadom na VEI index keďže, Stratovolcano ktorý vybuchol v Nikaragui s VEI indexom 0 zabil 1620 ľudí a naopak [Maar](https://en.wikipedia.org/wiki/Maar) vulkán ktorý vybuchol v Kamerone s VEI indexom 3 zabil 1700 ľudí. Samozrejme typ vulkanickej erupcie mení podmienky a teda s tým aj veľkosť počtu úmrtí.

Na potvrdenie hypotézy si pre istotu vykreslíme aj erupcie s počtom úmrtí pod 1000 ľudí.

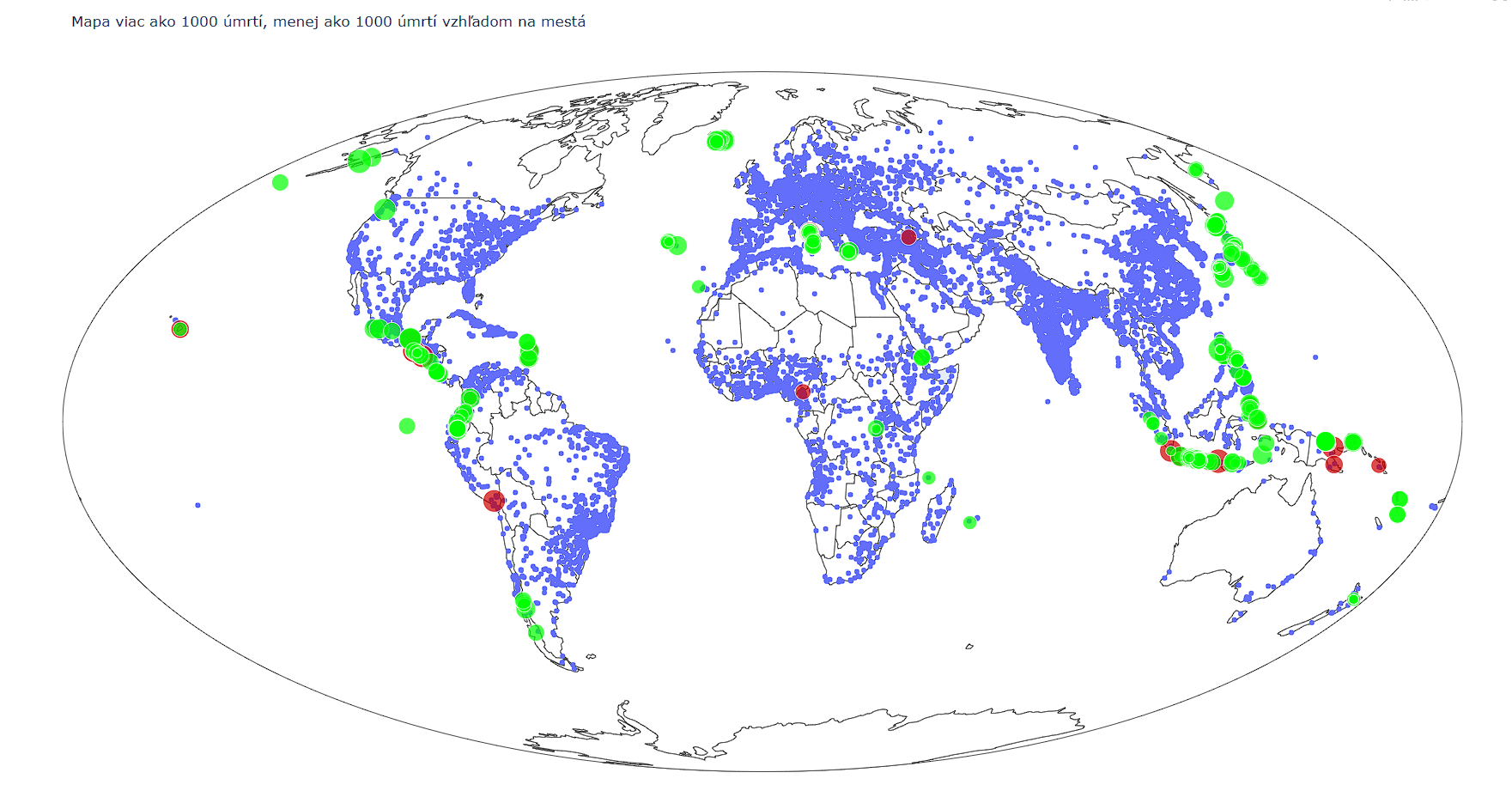
**Mapa úmrtí pod 1000 po roku 1950 vzhľadom na hustotou zaľudnenia**



Môžeme pozorovať, že sa na 2 rôznych mapách vyskytli rovnaké sopky no s diametrálne rozličným počtom úmrtí. Na mape vidno, že sa veľa výbuchov vyskytuje aj v miestach s vysokou hodnotou hustoty populácie ako môžeme vidieť napríklad vo Filipínach a v Japonsku.

Pre rozvitie tejto analýzy sme sa rozhodli teda na jednu mapu spolu znázorniť výbuchy s úmrtiami vyššími ako 1000 a zároveň s nižšími ako 1000 spolu so znázornenými miestami s populáciou väčšou ako 50000 čo by teda mohlo upresniť zhromaždenie ľudí ako len zovšeobecniť hustotu populácie pre celú krajinu.

**Mapa viac ako 1000 úmrtí, menej ako 1000 úmrtí vzhľadom na mestá**

Všimneme si teda, že tu teda pozorujeme prekrytie výbuchov s viac ako 1000 úmrtiami a menej ako 1000 úmrtiami. Toto teda môže naznačovať, že na lokalite až tak nezáleží keďže jedna sopka je schopná zabiť menej ľudí ale aj viac ako 1000. Zároveň nemôžeme s istotou povedať, že sa výbuchy s menším počtom úmrtí nachádzajú ďaleko miest alebo teda naopak výbuchy s veľkým počtom úmrtí pri mestách. Čiže dokopy povedať źe lokalita erupcie závisí od počtu mŕtvych spolu s VEI indexom je hlúposť. Po zanalyzovaní predošlého grafu a všimnutí si, že veľa erupcii je na jednom mieste čo indikuje aktivitu jednej sopky nás zaujala otázka ktorá sopka všetkých čias bola najaktívnejšia. Zistili sme že sopka Merapi v Indonézii je najaktívnejšia. Taktiež nás zaujalo prečo tomu tak je. Zistili sme, že to je kvôli tomu, že pod [Mount Merapi](https://www.twinkl.sk/teaching-wiki/mount-merapi) tečie odosť hustejšia láva ako zvykla 100,000 rokov pred tým čo zapríčiňuje viac výbuchov.

**Záver**

Odpoveď na otázku by mohla znieť nasledovne:

Úmrtia vzhľadom na sopečné erupcie sú veľmi nepredvídateľné. VEI index ako taký neurčuje počet smrtí. Toto tvrdenie sme zistili na základe viacerých faktorov ako napríklad veľký rozptyl v jednotlivých grupach pre daný VEI indexoch, Heat mapa neukazovala jednoznačné stmavnutie s nárastom smrtí, Pearson a Spearman koeficienty vyšli veľmi malé čísla. Či úmrtia závisia s lokalitou sopky je tiež veľmi diskutabilné. Toto môžeme tvrdiť na základe toho, že mnoho sopiek v rovnakej lokalite je schopných spôsobiť rôzny počet úmrtí. Tu asi treba poznamenať, že istým spôsobom môžu počet úmrtí ovplyvniť mnohé faktory ktoré nemajú nič spoločné so sopkami ako takými. Ako napríklad bohatosť krajiny, s tým spojené možnosti výskumu a pozorovania aktívnych sopiek a teda pred pripravenosť na nasledujúce výbuchy. Ďalej napríklad to ako moc je sopka aktívna. V našom pozorovaní sme zistili napríklad to, že jedna sopka vybuchla za rok 3 krát prvýkrát zabila 28 tisíc ľudí s VEI indexom menším ako keď vybuchla neskôr v tom istom roku s menším poćtom úmrtí ale s väčším VEI indexom, pretože sa po prvom výbuchu ľudia začali báť a odsťahovali sa alebo im sopka zničila domov a nemali . Dá sa povedať, že na lokalite môže záležať ale nemusí nevieme to s istotou povedať. Teda či daná sopka pri rôznom VEI indexe s rôznou lokalitou zapríčiní veľa alebo málo úmrtí je skoro s istotou náhodné.

## 

## 

## 

## 

## 

## 

## 

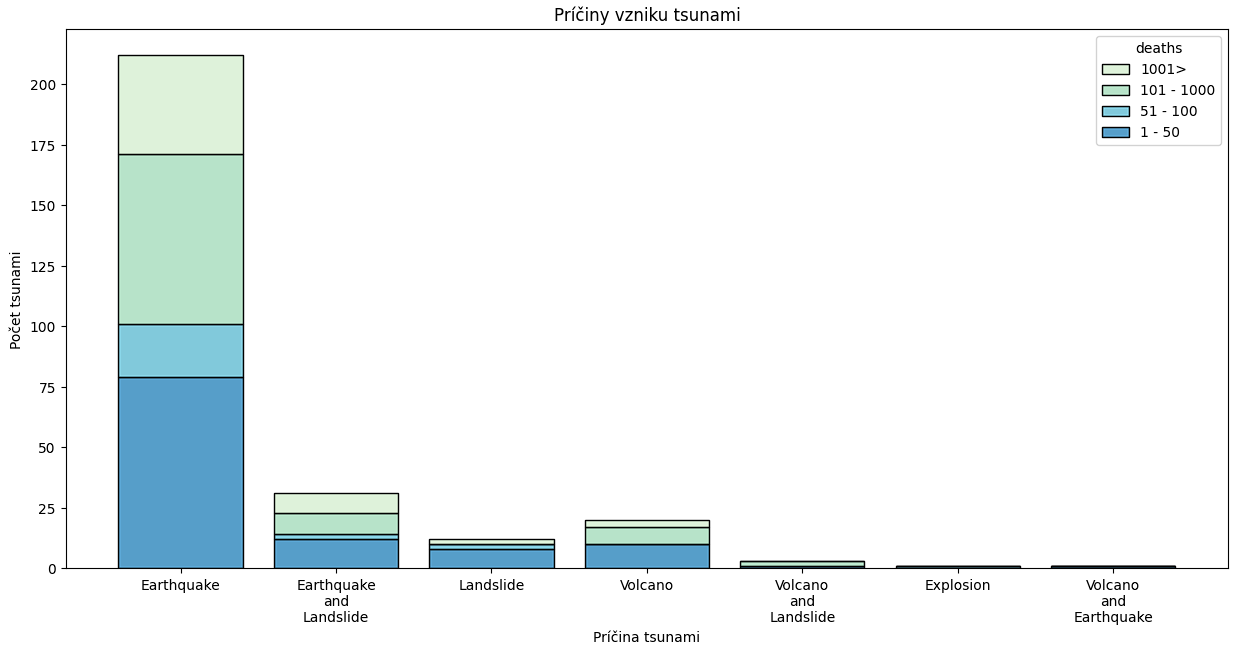
## 

## 5. [Otázka: Aké faktory zapríčiňujú tsunami.](#_heading=h.1t3h5sf)

**Analýza:**

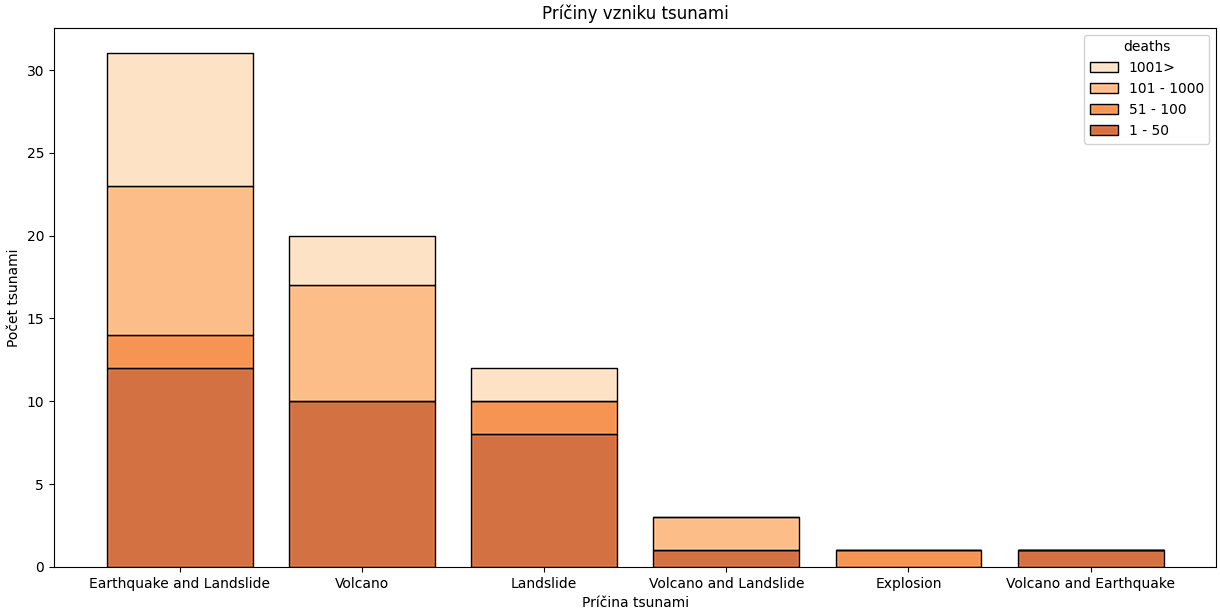
Túto analýzu sme sa rozhodli začať tým, že sme si položili otázku čím sú samotné tsunami zapríčinené a ako veľa úmrtí táto daná príčina tsunami spôsobila. Toto sme si zvizualizovali pomocou obyčajného stacked histplotu.

**Histplot**



Tu vidíme že obrovitánsky počet zapríčinenia tsunami má zemetrasenie a naopak strašne malý má volcanoes and earthquakes čo je myslené tak, že vulkán spôsobil zemetrasenie čo spôsobilo tsunami. Zároveň si môžeme všimnúť, že v pomerne veľké percentuálne zastúpenie v rámci skoro všetkých jednotlivých príčin ma úmrtie 1-50. No nevieme toto povedať s istotou pre málo zastúpené hodnoty kvôli ich malému počtu a teda vykresleniu malého “obdĺžnika”. Rozhodli sme sa teda “useknúť” časť s iba Earthquakes ktorá rozťahuje graf od výšky.

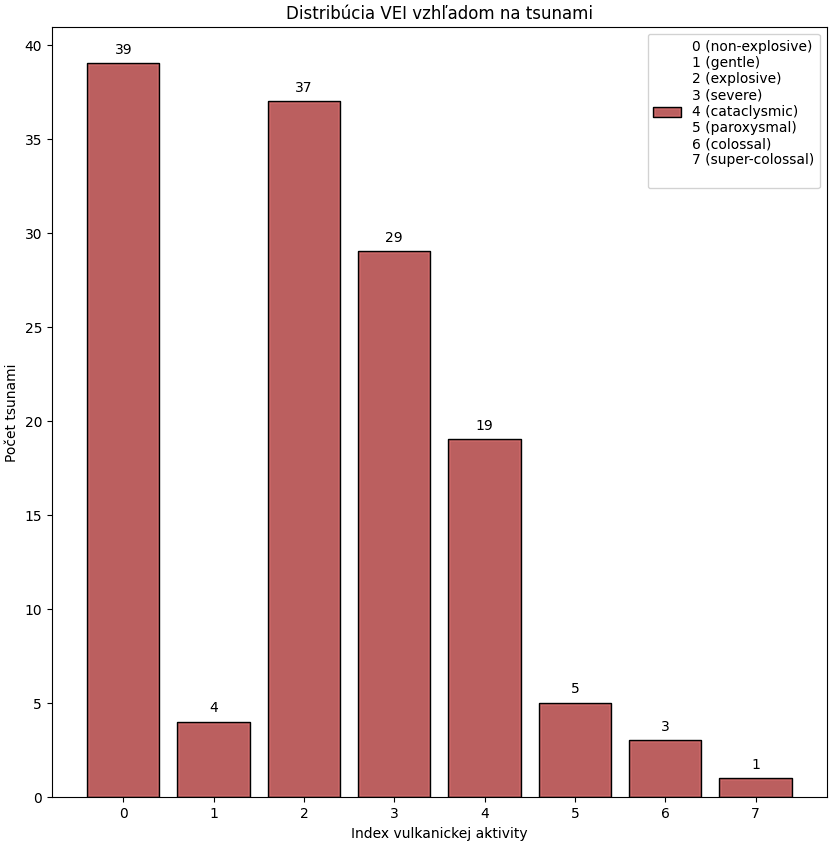
**Histplot useknutý**



Tu teda vidíme že toto prevažujúce percentuálne zastúpenie 1-50 pre každú grupu ostalo okrem Explosion a Volcano and Landslide. Keď sa na toto pozorovanie pozriem vo všeobecnosti s tým, že porovnávame samotné rôzne príčiny môžeme povedať že sa líšia v počte samotných výskytov a počte úmrtí. Na základe týchto zistení sme sa rozhodli teda zistiť pre 2 najzastúpenejšie príčiny spôsobujúce Tsunami a sme oboznámení s jej meracou jednotkou, teda vieme ich nejakým spôsobom merať, zistiť či ich meracia jednotka závisí s výškou tsunami.

Rozhodli sme sa začať vulkánmi. Naša počiatočná otázka čo sa týka vulkánov bola aký minimálny VEI index vie vyvolať tsunami.

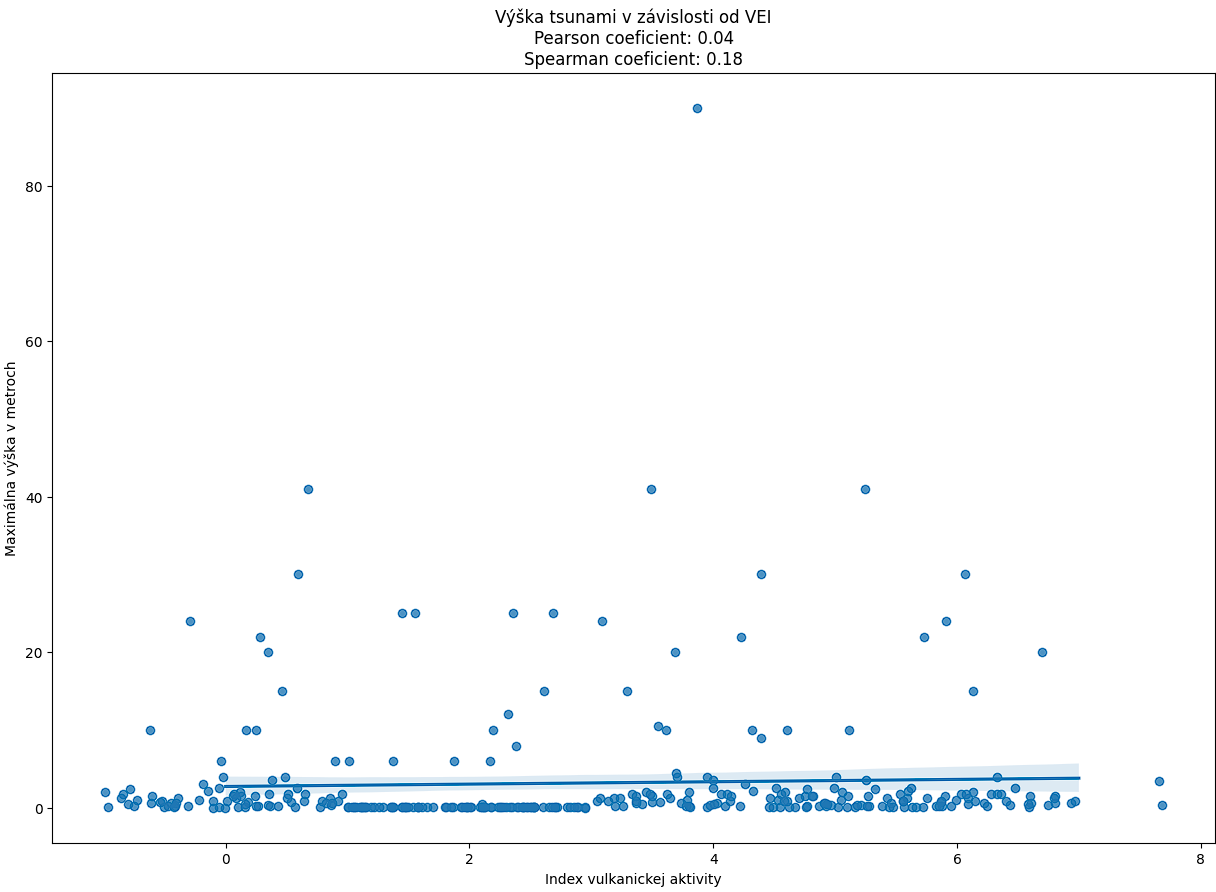
**Distribúcia VEI vzhľadom na Tsunami**



Po vizualizácii sme si všimli veľmi zaujímavú vec a to takú, že na to aby výbuch hocijakého typu vyvolal tsunami netreba vôbec veľký VEI index čo teda znamená, že aj non-explosive výbuchy môžu spôsobiť tsunami, dokonca v našom datasete spôsobili najväčší počet tsunami. Možno by sa tu dala pozorovať nejaká pomaly klesajúca závislosť no VEI index 1 to celé kazí pretože ma úplne “zlú” hodnotu na to aby sme mohli povedať, že tam nejaké lineárne klesanie je preto sa k tomu nevyjadrujeme.

Po tomto zistení sme sa rozhodli overiť toto tvrdenie.

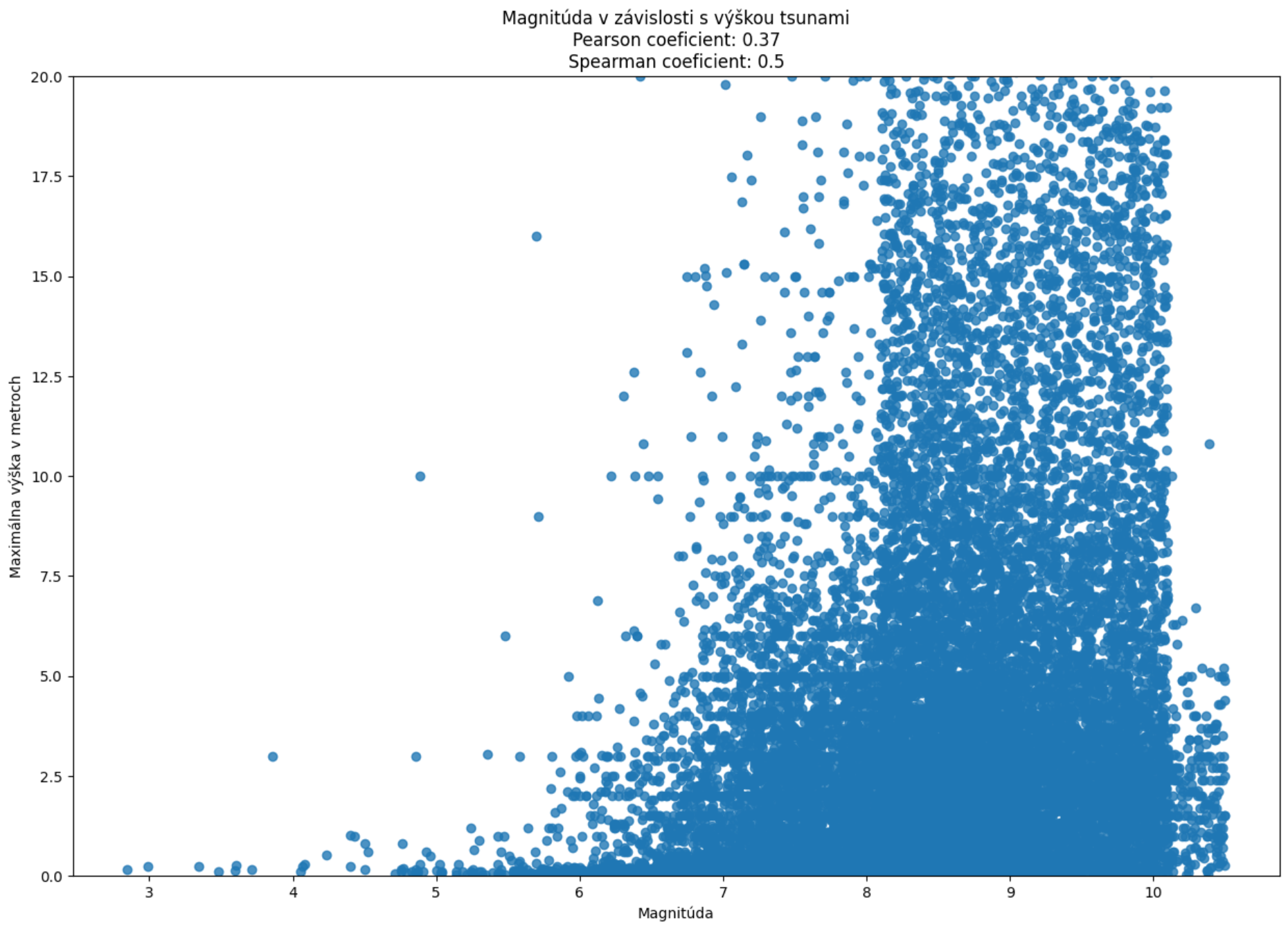
**Pearson a Spearman závislosť VEI indexu od výšky tsunami**



Po úprave a následnej vizualizácii dát pomocou Pearsonovho a Spearmanovho koeficientu sme zistili, že tu absolutne žiadna závislosť nie je. Toto sme odôvodnil tým,že hodnoty oboch koeficientov vyšli skoro 0.

Následne sme sa teda rozhodli pozrieť sa na závislosť magnitúdy a výšky tsunami. Rovno sme sa rozhodli teda, že zistíme Pearsonov a Spearmanov koeficient.

**Pearson a Spearman závislosť magnitúdy od výšky tsunami**



Tieto hodnoty vyšli Pearson: 0.37 Spearman: 0.50. Tento graf je celkom chaotický no zaujímali nás hlavne hodnoty koeficientov. Hodnoty sú nedostatočne veľké na to aby sme s istotou vedeli povedať, že závislosť výšky tsunami súvisí s magnitúdou zemetrasenia. Tu treba poznamenať jednu vec a to teda, že náš dataset reprezentuje jedno zemetrasenie viacerými tsunami. Čiže jedno zemetrasenie vyvolá viacero tsunami s rôznymi výškami. Čo teda napovedá že by závislosť magnitúdy a veľkosti tsunami by mala byť nezmyselná lebo ako by sme mohli predikovať závislosť ked jedno zemetrasenie magnitúdy 3 je schopné vyvolať tsunami výšky napr 10 metrov a aj 0.1 metra.

**Záver**

Táto analýza nám dala nasledovné informácie:

Zistili sme že najväčšia príčina vzniku tsunami sú zemetrasenia. Zistili sme hodnoty úmrtí pre jednotlivé príčiny tsunami, zistili sme leadera v tomto rebríčku. Zároveň sme sa dozvedeli, že percentuálne zastúpenie pre jednotlivé príčiny je každé iné. Ďalej sme pre pre 2. najviac zastúpený merateľný typ príčiny zistili jeho závislosti voči tsunami. Tu sme zistili, že VEI index pri sopkách nemá žiadne ovplyvnenie vzhľadom na výšku tsunami. Následne sme sa obrátili na 1. najviac zastúpený typ príčiny tsunami a to teda zemetrasenia. Po analyzovaní sme zistili, že aj zemetrasenia nemajú žiaden vplyv na výšku tsunami. Toto rozhodnutie sme spravili na základe Spearmanovho a Pearsonovho koeficientu. Tu treba poznamenať jednu vec a to teda, že náš dataset reprezentuje jedno zemetrasenie viacerými tsunami. Čiže jedno zemetrasenie vyvolá viacero tsunami s rôznymi výškami. Čo teda napovedá že by závislosť magnitúdy a veľkosti tsunami mala byť nezmyselná lebo ako by sme mohli predikovať závislosť ked jedno zemetrasenie magnitúdy 3 je schopné vyvolať tsunami výšky napr 10 metrov ale aj 0.1 metra. Teda na potrebnú presnú analýzu by trebalo brať do úvahy, že akým spôsobom sa tsunami vyvíja v závislosti od vzdialenosti.

# 

# **Záver**

Počas skúmania analytických otázok sa nám podarilo potvrdiť rôzne všeobecne známe tvrdenia ohľadom kataklyzmických javov. Zistili sme napríklad to, že magnitúda zemetrasení má väčší potenciál na vyššiu úmrtnosti. Ďalej sme potvrdili tvrdenie, že sa zemetrasenia orientujú v okolí zlomov tektonických dosiek, a viacero ďalších tvrdení, ktoré sme ešte pred vyhotovením analýz očakávali.

Čo sa týka rozvitia analýz všetkých katastrof, tak nami vyšetrované analýzy sa týkali hlave celého sveta. Takéto analyzovanie je veľmi všeobecné na pozorovanie prípadných trendov/závislostí. Preto by bolo rozumnejšie alebo teda výpovednejšie, zamerať sa napríklad na jednotnú oblasť/krajinu. Takéto pozorovania vykonané na jednotnej oblasti/krajine by nám mohli poskytnúť lepšie pozorovania ako pre celý svet. Ak by sme však mali pokračovať v našich analýzach ďalej tak by sme skúsili tieto údaje porovnať s dátami iných relevantných datasetov. Napríklad porovnať dáta všetkých sopiek na svete s tým ktoré sú aktívne. Poprípade pozorovať koncentráciu sopečných aktivít vzhľadom na svet. Ďalej by sme mohli rozvinúť analýzu tsunami a zistiť či tsunami zapríčinené kataklizmou má rozdielnu výšku vzhľadom na to akú vzdialenosť prejde od zdroju ku pevnine. Prípadne rozvitie analýzy ktoré sme mohli spraviť by bolo či lokalita zemetrasení nie je spojená s lokalitou erupcií sopiek.

Okrem všeobecne známych tvrdeniach o kataklizmických udalostiach sme sa dozvedeli aj nové zaujímavé zistenia. Napríklad porovnanie priemernej magnitúdy zemetrasení v Taliansku a Čile. Vôbec sme netušili, že v Taliansku sa vyskytujú o dosť miernejšie zemetrasenia oproti Čile. Taktiež sme netušili, že tento jav môže byť spôsobený práve lokáciou vzhľadom na tektonické dosky (Čile celým územím opisuje zlom, kdežto Taliansko sa nachádza v zlome iba Sicíliou). Pri analýze zemetrasení nám vôbec nenapadlo, že evidencia zemetrasení môže byť ovplyvnená politickou, ekonomickou alebo pandemickou situáciou krajiny. Ďalšie zaujímavé zistenie bolo, že sme sa dozvedeli, že vulkanické aktivity s VEI indexom 0 teda non-explosive zapríčinili najviac tsunami.

Analýza a práca na tomto projekte bola zaujímavá, mnoho externých zdrojov sme nepoužili no z každej jednej analýzy sme boli schopný overiť alebo vyvrátiť nami na začiatku položené otázky. Zistili sme mnoho nových vzťahov a teda aj naopak vyvrátili množstvo predpokladatelných závislostí.

# 

# **Zdroje**

Hlavné dáta o významných kataklizmických udalostiach <https://data.noaa.gov/metaview/page?xml=NOAA/NESDIS/NGDC/MGG/Hazards/iso/xml/G012153.xml&view=getDataView>

Konvertor kml na geojson

<https://anyconv.com/kml-to-geojson-converter/>

Dáta o tektonických doskách

<https://github.com/fraxen/tectonicplates>

Dáta o mestách

<https://simplemaps.com/data/world-cities>