STA 141B Final Report: The Change of Earthquakes overtime in North America and Asia

by Jeffrey Ugochukwu, Keerthan Panyala, Chan Woong Joo, and Eric Liu

Earthquakes have been around for many years and they have notably occurred in Asia and North America. Earthquakes serve a purpose of restructuring the topography to reform the tectonic plates that are stuck within the earth's layers. Even though earthquakes are still fairly common, we haven't heard of much information regarding major ones. Therefore, the core of this project is to answer how the severity of earthquakes in North America differs between that of the Asian countries of Indonesia, Japan, Papua New Guinea, and the Philippines which we will be referring to as **IJPP**.

Exploratory Data Analysis of North American Earthquakes

The first part of the analysis should focus on the exploratory data analysis that has occurred regarding North American earthquakes. We aim to investigate how many tsunamis have existed within an international scale created by earthquakes. The value will be provided below.

Number of tsunami warning due to earthquakes	
0	237

As we can see, there were 237 tsunamis that were caused by earthquakes.

Now we can shift our focus back to North America and we should discover what states tend to have the most earthquakes overall.

	number of quakes	state	country
0	19453	California	US
1	6222	Alaska	US
2	5302	Oklahoma	US
3	4786	Nevada	US
4	3449	Idaho	US
5	1511	Montana	US
6	1303	Yukon	CA
7	1054	Wyoming	US
8	960	Baja California	MX
9	887	Kansas	US

As we can see, these are the 10 states that contain the largest quantity of earthquakes and we can see that Kansas has the least amount of earthquakes at 887 earthquakes and California has the most at 19,453 earthquakes. This would mean California overall in North America is the most problematic state regarding their issues with earthquakes.

Another question that we want to answer is which month in North America had the highest magnitude produced from an earthquake within the past 5 years. This can be answered through the image below.

	max(mag)	month
0	6.6	01
1	5.8	02
2	6.5	03
3	5.5	04
4	6.5	05
5	6.3	06
6	7.1	07
7	6.4	08
8	7.1	09
9	6.8	10
10	5.7	11
11	6.6	12

As we can see, July and September had the highest magnitude from earthquakes within North America. This would mean that the most severe earthquakes that have occurred happened within those months, which could mean that the most damage has happened during those times.

Another question that can be answered would be what North American locations had a magnitude level of at least 6. The answer can be shown below.

	index	id	mag	place	time	tsunami	sig	rms	longitude	latitude	country	state	count_month	count_day
0	551	us20007z6r	6.6	164km W of Ferndale, California	2016-12-08 14:49:00	Yes	1120	1.11	-126.1936	40.4535	US	California	672	1
1	4168	us200062i1	6.3	106km SSW of San Patricio, Mexico	2016-06-07 10:51:00	No	611	1.00	-105.1731	18.3637	MX	Jalisco	557	1
2	7098	us10004g4l	6.6	215km SW of Tomatlan, Mexico	2016-01-21 18:06:00	Yes	673	0.98	-106.9337	18.8239	MX	Jalisco	771	1
3	9741	us2000ar20	7.1	1km E of Ayutla, Mexico	2017-09-19 18:14:00	No	1686	1.14	-98.4887	18.5499	MX	Puebla	1372	1
4	13845	us10008mgu	6.3	88km WNW of Skagway, Alaska	2017-05-01 14:18:00	Yes	975	0.85	-136.7042	59.8295	CA	Yukon	990	1

These are the top 5 North American locations that have a magnitude of at least 6 and it shows that the area with the highest magnitude would be the location, 1km E of Ayutla, Mexico with a magnitude level of 7.1. This would mean that this North American location would have the largest amount of damage resulting from an earthquake.

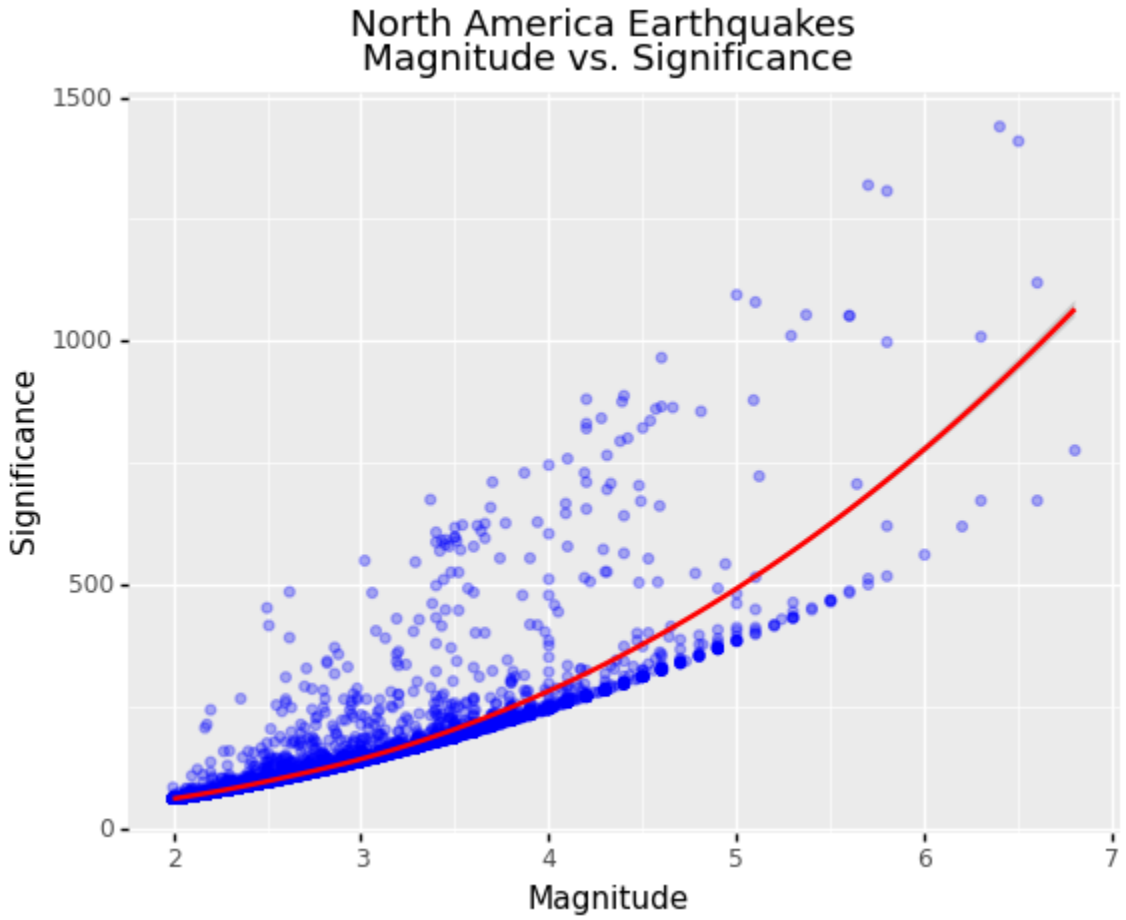
Building off of this topic, we can also figure out what North American locations have a high magnitude level and have created tsunamis from their earthquakes.

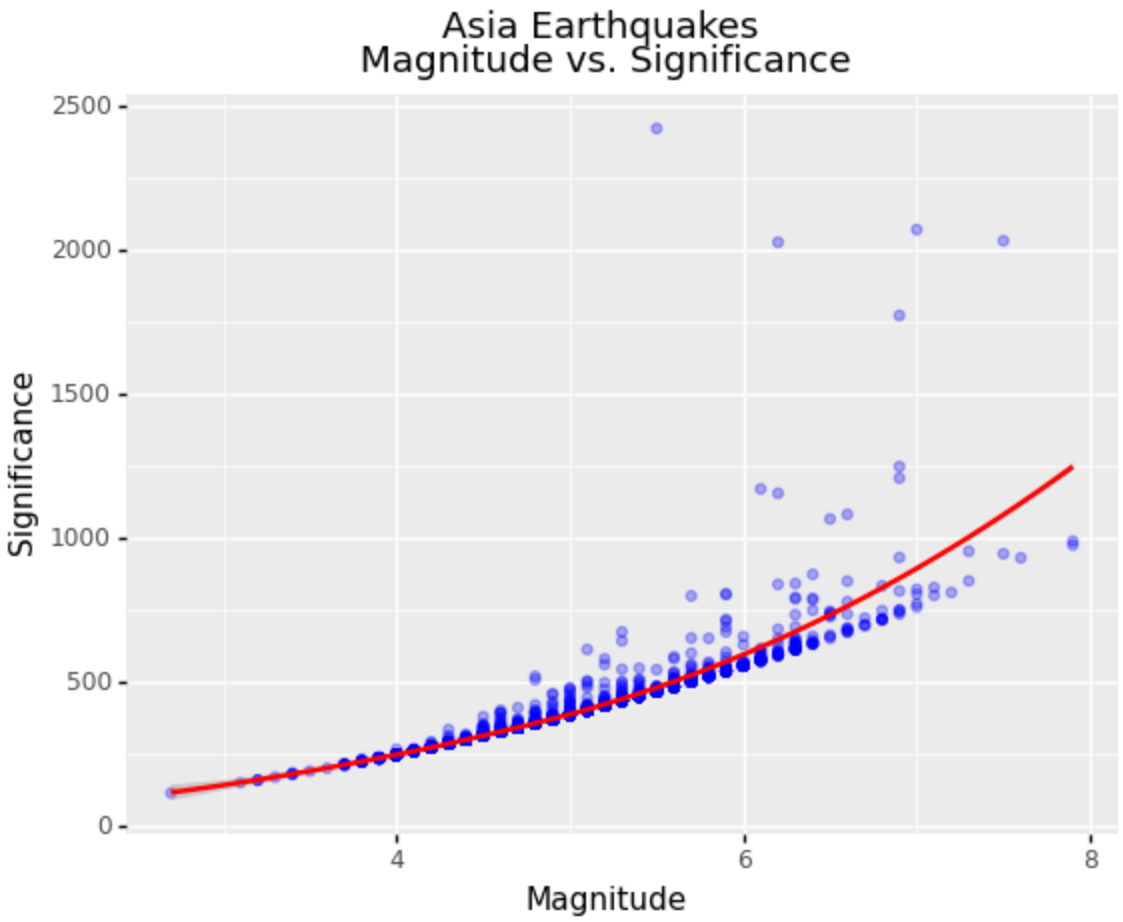
	index	id	mag	place	time	tsunami	sig	rms	longitude	latitude	country	state	count_month	count_day
0	551	us20007z6r	6.6	164km W of Ferndale, California	2016-12-08 14:49:00	Yes	1120	1.11	-126.1936	40.4535	US	California	672	1
1	7098	us10004g4l	6.6	215km SW of Tomatlan, Mexico	2016-01-21 18:06:00	Yes	673	0.98	-106.9337	18.8239	MX	Jalisco	771	1
2	13845	us10008mgu	6.3	88km WNW of Skagway, Alaska	2017-05-01 14:18:00	Yes	975	0.85	-136.7042	59.8295	CA	Yukon	990	1
3	13894	us10008mel	6.2	88km WNW of Skagway, Alaska	2017-05-01 12:31:00	Yes	1035	0.96	-136.7114	59.8209	CA	Yukon	990	1
4	15721	us10007s20	6.0	79km ESE of Resolute, Canada	2017-01-08 23:47:00	Yes	557	0.78	-92.4156	74.3859	CA	Nunavut	565	1
5	17290	us1000hfgv	6.8	202km SW of Port Hardy, Canada	2018-10-22 06:16:00	Yes	776	1.09	-129.2890	49.3346	CA	British Columbia	803	1
6	17291	us1000hfgn	6.5	214km SW of Port Hardy, Canada	2018-10-22 05:39:00	Yes	669	0.94	-129.4124	49.2586	CA	British Columbia	803	1
7	19296	us1000gf8u	6.2	272km WNW of Bandon, Oregon	2018-08-22 09:31:00	Yes	680	0.73	-127.7166	43.5637	US	Oregon	2088	1
8	20549	ak018aasz24z	6.0	73km SSW of Kaktovik, Alaska	2018-08-12 21:15:00	Yes	565	0.86	-144.3388	69.5227	US	Alaska	2088	1
9	20691	ak018aap2cqu	6.4	90km SW of Kaktovik, Alaska	2018-08-12 14:58:00	Yes	686	1.31	-145.2910	69.5762	US	Alaska	2088	1

As we can see, these are the top 10 North American locations that satisfy both conditions and the area with the highest magnitude from an earthquake would be located at 202 km SW of Port Hardy, Canada with a level of 6.8. This means that this area has the largest amount of earthquake damage alongside with water damage given that it had a tsunami created by the earthquake.

North America vs IJPP

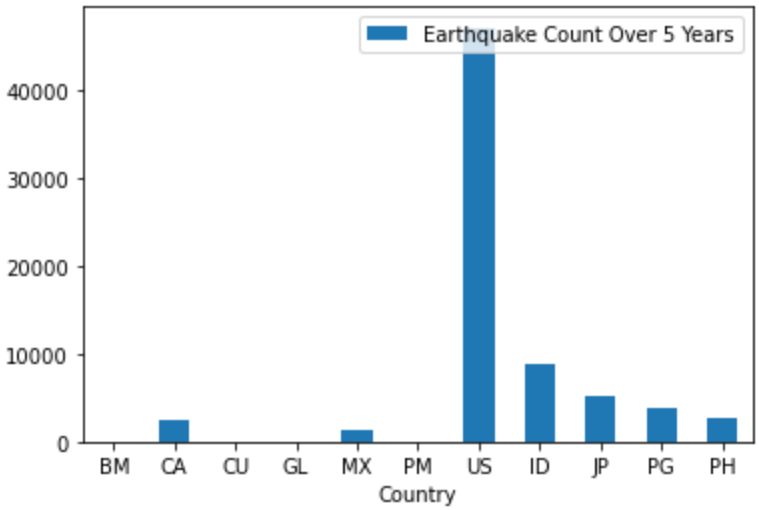
Now that we have done an in depth analysis regarding earthquake information in North America, we can now compare the differences of earthquakes that have happened in both North America and IJPP. The first step would be to compare the significance of the level of magnitudes from the earthquakes created in both continents. To demonstrate this, we created two separate regression models that can show where the data points of the magnitudes fit their lines of best fit





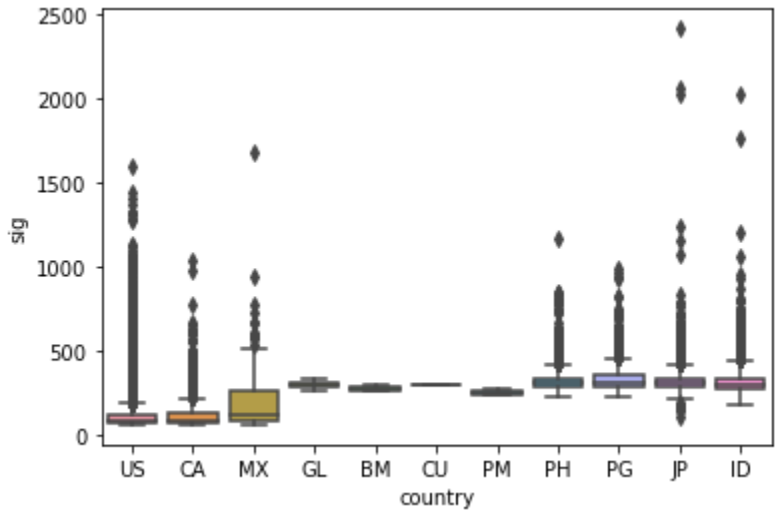
Judging from the regression plots, the one similarity that both plots share is that the magnitude levels between 2 to 4 was fairly close to the line of best fit. The major difference between both plots is that the North American plot shows that a fair portion of the data points throughout the entire graph and past the magnitude level of 4 were further away from the line of best fit (making them outliers) whereas in the Asia plot, you don't really see outliers until a magnitude level past 5. This means that overall that the relationship between magnitude and significance level is stronger for earthquakes in IJPP countries than North American Countries.

Another thing that we can look at would be what countries from both contents have the most earthquakes overall. This can be shown within the bar plot below.



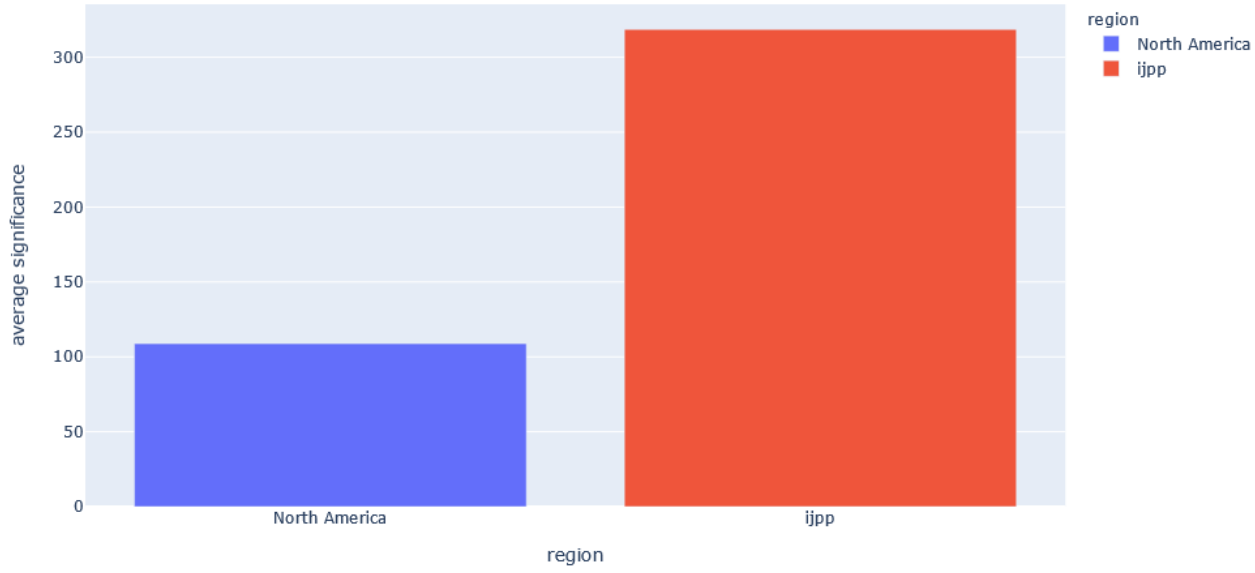
As we can see in the barplot, the most amount of earthquakes within the past 5 years have occurred in the US with an amount of over 40,000. This would mean that North America overall has the largest quantity of earthquakes compared to IJPP within that time period.

Another question that can be answered would be the distribution of the skewness of the significance of earthquakes within countries in North America and IJPP. This boxplot demonstrates the amount of skewness that exists in various countries.



As we can see from the boxplots, the majority of them are skewed right, which means that many of them have a high significance level. Asia does tend to have the most skewed significance level distributions in general since countries such as the Philippines, Papua New Guinea, Japan, and Indonesia make up most of the skewed distributions regarding significance levels of the earthquakes.

Furthermore, if we take a look at the bar chart for average significance between the two regions. We see that the IJPP countries have earthquakes with much higher average significance when compared to North american earthquakes. This is mostly due to the higher population density of the IJPP countries when compared to the North American region.



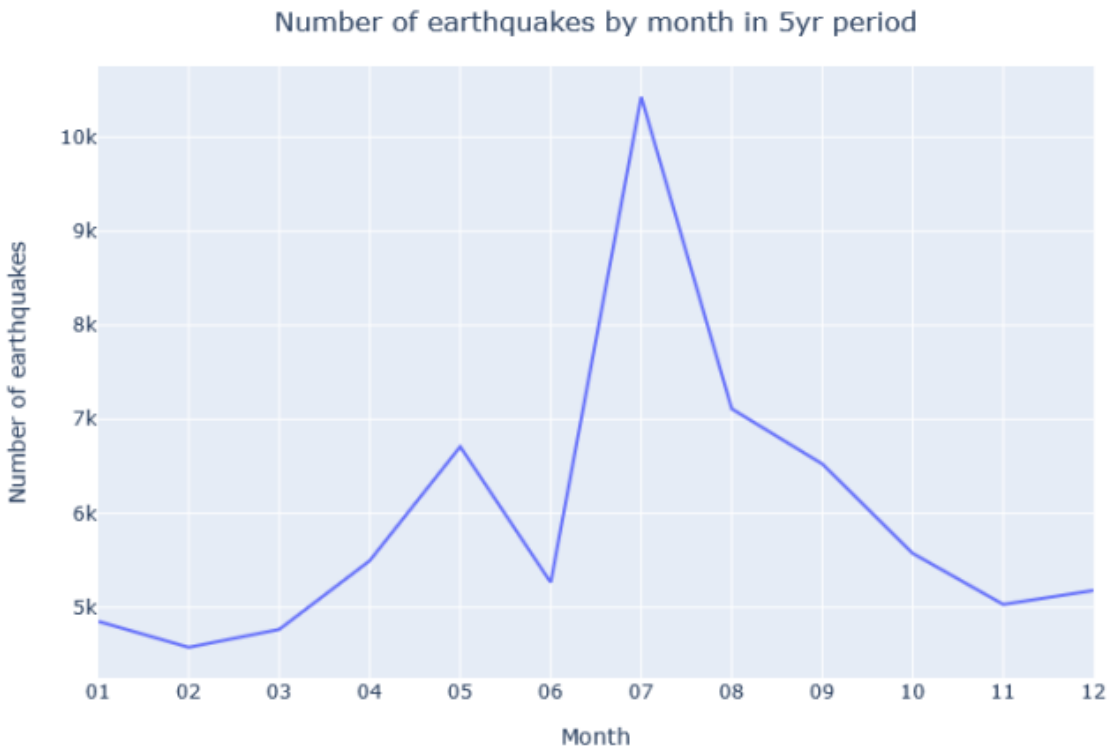
Building off of the topic of significance levels, we can look into what is the average significance level per magnitude range to measure the level of danger. The following image will explain this further.

	Mag	Avg Significance
0	2-3	88.148469
1	3-4	180.739878
2	4-5	303.680918
3	5-6	434.442659
4	6-7	677.512563
5	>7	1030.028269

As we can see here, as the magnitude level increases, the more significant the effects of the earthquake are. The most significant earthquakes have a magnitude greater than 7 at a value of 1030.028269 (making them the most dangerous) and the least significant earthquakes have a magnitude level ranging from 2 to 3 (making them the least dangerous).

Analysis of the Number of Earthquakes that have happened Over Time

Now that we have gone through the difference of the effects of the earthquakes in both Asia and North America, we can now focus on analyzing the data regarding the number of earthquakes that have occurred within a certain period of time. The first thing we want to analyze is what month tends to have the largest amount of earthquakes within the past 5 years. This can be given within the time series plot below.



As we can see within the time series model, the most earthquakes happened in July with a value of over 10,000 in total. This means that in the past 5 years, the majority of earthquakes happened in July.

Adding on to this topic, we can use another time series model to answer what year had the largest amount of earthquakes within the 5 year span.

Hypothesis testing:

The plot of the number of earthquakes by month illustrates a huge leap in the month of July. To investigate this eccentricity, we conducted a hypothesis test. The hypothesis are as follows:

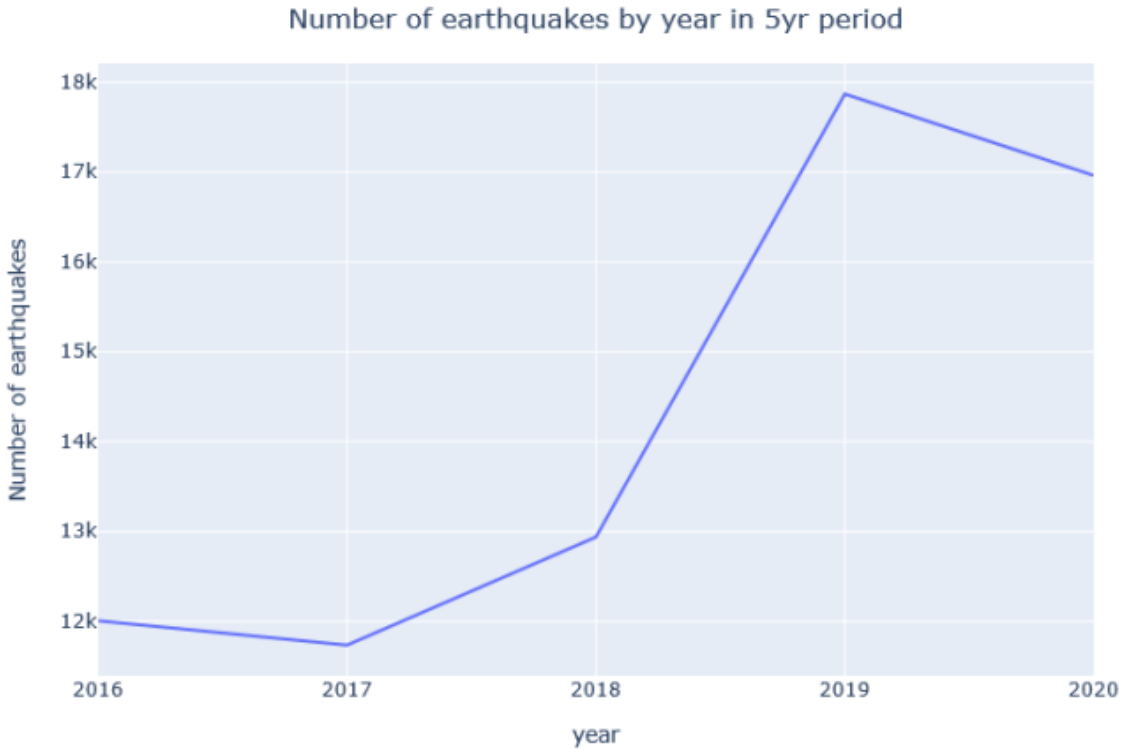
Null: average number of earthquakes for the rest of the year is equal to the number of earthquakes in July

Alternative : Average number of earthquakes for the rest of year is different from the number of earthquakes in July.

North America - (statistic=-9.33360462547868, pvalue=4.1695984771341865e-09)

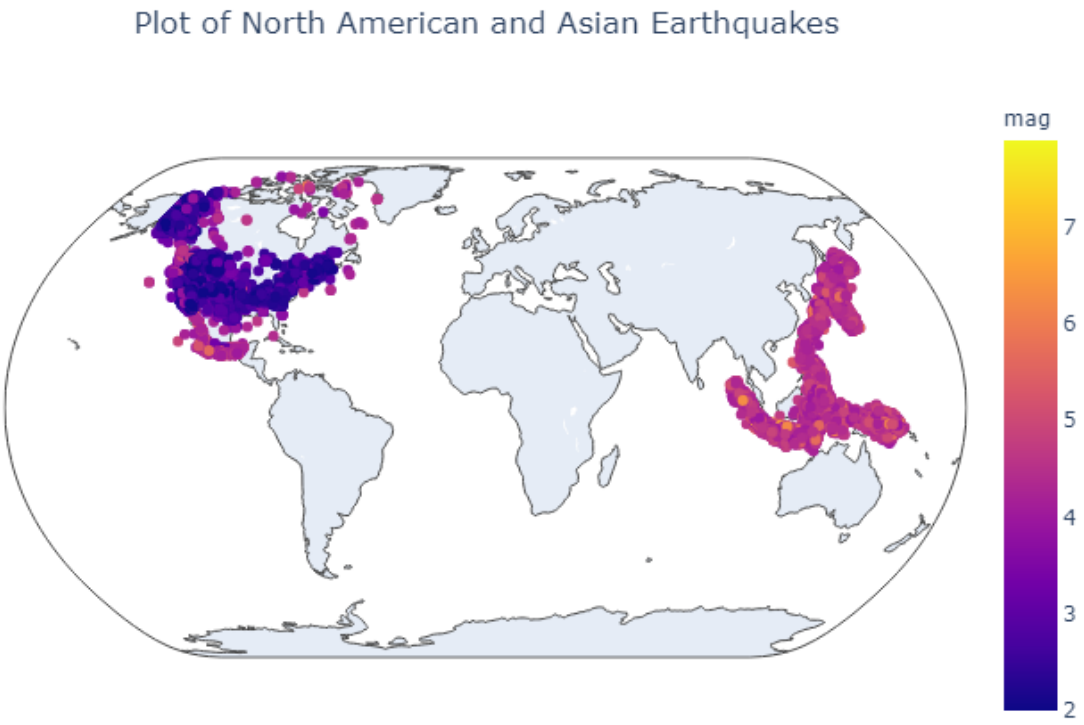
Asia - (statistic=-3.576466086066434, pvalue=0.001684300303631888)

At a significance level of 0.05, we are able to reject the null hypothesis and say that the average number of earthquakes for the rest of the year is different from the number of earthquakes in July for both North America and Asian since our P-values are very small.



As we can see in this plot, the majority of earthquakes have occurred in 2019 with it being a bit under 18,000 in total and it only slightly decreases to 17,000 by 2020. This means that most of the earthquakes that have happened within the 5 year span happened towards the end of the 2010's and the start of the current 2020's.

Another question that we can try to answer would be between North America and Asia, which continent had the most dangerous earthquakes overall. We can answer this through the plotly visual that is created below.

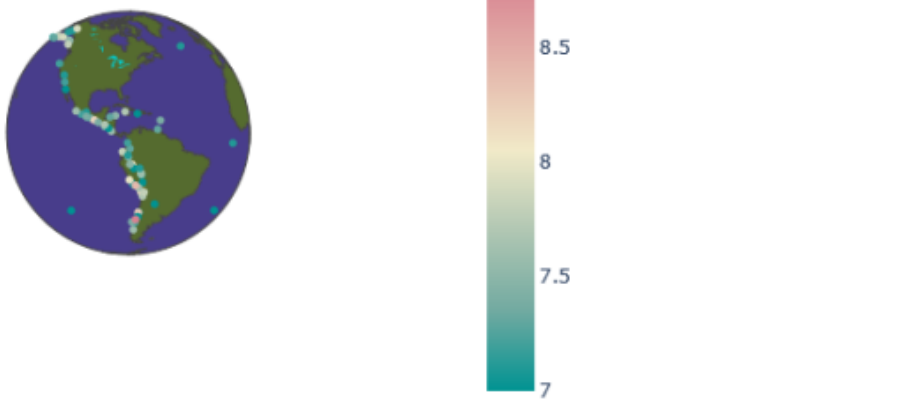


To explain the graph, the following colored bar gives an intensity level of the magnitude that the earthquakes were produced from parts of each continent. The darker the color, the lower the magnitude level; the brighter the color, the higher the magnitude level. As we can see, the

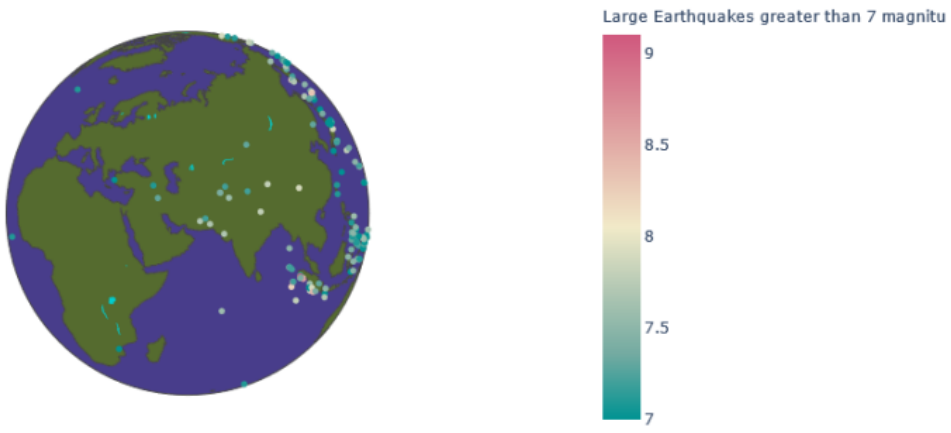
Southeast Asian areas had the lightest areas regarding magnitude compared to North America that is much darker in color. Most of the colors are purple to orange so this means the magnitude levels within Southeast Asia were between 5 and 7. This means that Southeast Asia overall had the most dangerous earthquakes that exist in the dataset, which would give them more devastating results compared to North America.

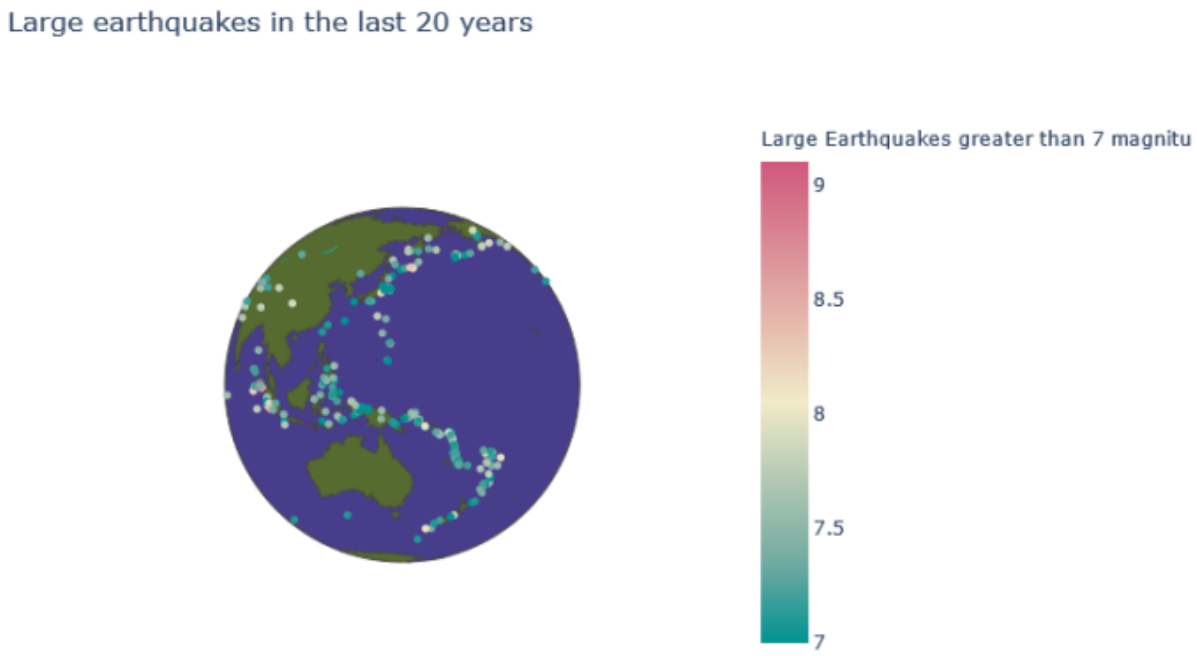
Adding on towards the topic of magnitude levels, we can also add on to the global areas that had the highest magnitude levels in the entire dataset (magnitude greater than 7). This can be given within the global plots produced by plotly.

Large earthquakes in the last 20 years



Large earthquakes in the last 20 years





To explain the graphs for a bit, the bar gives an intensity level of the largest earthquakes that exist within the dataset. The colored bar gives the intensity level and as the colors approach teal, it's the least intense; as the colors approach crimson, it's the most intense. As we can see, the majority of the largely intense earthquakes occurred in Asia since the data points are blue to light blue (making them between 7 to 8 in magnitude) and more of them occurred in that continent. This further proves that Asia had the most dangerous earthquakes overall.

Conclusion

IJJP regions have much higher magnitude and significance compared to North American Region, possibly due to the reason that they are located near the tectonic plates. IJJP regions have higher population density, which contributes to higher significance levels. Throughout our analysis, the month July has a higher number of earthquakes compared to the other months. This is probably due to a selection bias since earthquakes aren't affected by seasons or surface conditions. There has been an increase in the number of earthquakes in the last 5 years. However, we expect to have further data analysis as for this phenomenon. At this point, we are concluding that there has been significant development in technology and the number of measurement devices which then results in an increase.