Volcanoes and Earthquakes

Group - 17

Sahiti Krishna Addada
Akhil Kanike
Raj Karra

Course Name: Statistical Data Mining-1
Professor: Prof. E Bruce Pitman

Contents

1. Introduction

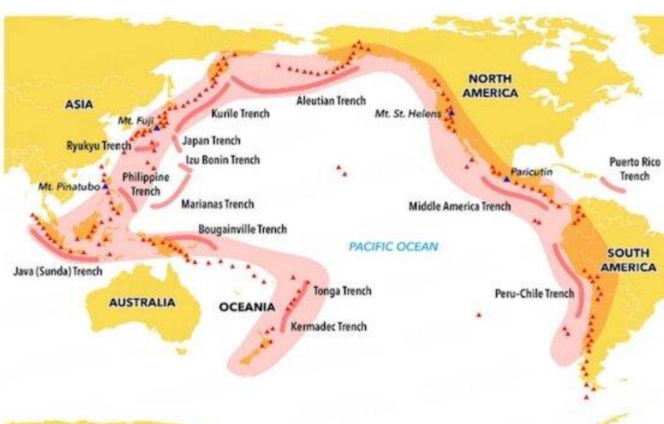
Earthquakes often precede volcano eruptions, sometimes only by hours and other times by several months. While most seismic activity related to volcanoes is not necessarily followed by an eruption, almost all eruptions are preceded or accompanied by earthquake activity at shallow depths in the vicinity of the volcano.

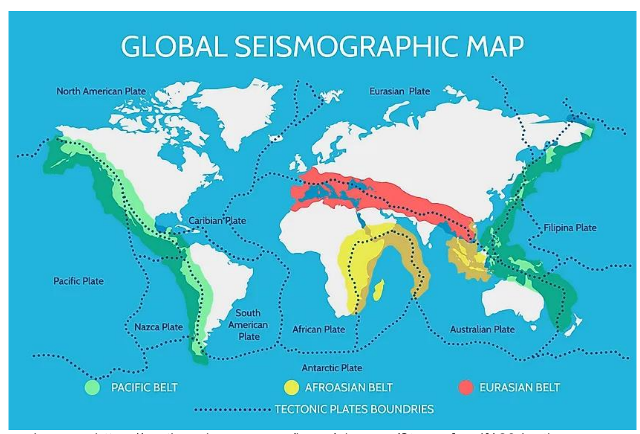
Volcanic earthquakes are triggered by the explosive eruption of a volcano. Due to the fact that not all volcanoes erupt violently, these earthquakes are very rare. Volcanic earthquakes may be felt within an area of 10 to 20 miles around the epicenter, whereas a tectonic earthquake has no limits.

Volcanically-caused long period earthquakes are produced by vibrations generated by the movement of magma or other fluids within the volcano. Pressure within the system increases and the surrounding rock fails, creating small earthquakes.

Volcano Facts:

1. Chile has the second most active string of volcanoes.
2. Ojos del Salado on the Chilean and Argentinian border is the highest active volcano in the world.
3. Japan alone accounts for nearly 1/10 of all the world's active volcanoes.
4. The Ring of Fire extends from New Zealand clockwise in an almost circular arc through: Indonesia, Philippines, Japan, Papua New Guinea, Kamchatka Peninsula of Russia, Aleutian Island of Alaska (USA), Western coasts of North, Central, and South America and Also, the interior includes the Galapagos Islands and the Hawaiian Islands





Imgsrc : https://earthquake.usgs.gov/learn/glossary/?term=focal%20depth

Earthquake Facts:

1. The world's greatest earthquake belt, the circum-Pacific seismic belt, is found along the rim of the Pacific Ocean, where about 81 percent of our planet's largest earthquakes occur. It has earned the nickname "Ring of Fire".
2. The Alpide earthquake belt extends from Java to Sumatra through the Himalayas, the Mediterranean, and out into the Atlantic. This belt accounts for about 17 percent of the world's largest earthquakes

3. The third prominent belt follows the submerged mid-Atlantic Ridge. The ridge marks where two tectonic plates are spreading apart (a divergent plate boundary).

Let's Get Started!

2. Problem Statement

In the interest of earthquakes and volcanoes the question is to determine the circumstances under which an earthquake is generated by volcanic activity. Looking forward to find out!

3. Dataset

'Significant Earthquakes' from National Centers for Environmental Information.

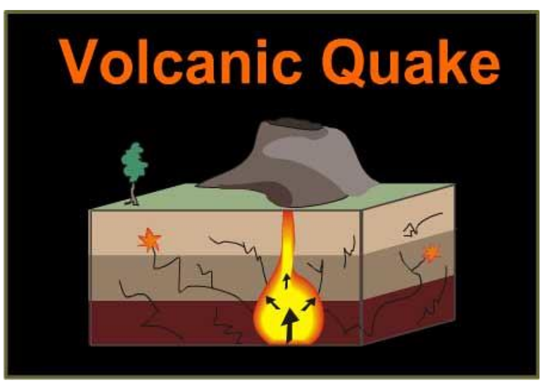
Link:

https://www.ngdc.noaa.gov/hazel/view/hazards/earthquake/search

Parameters Used:

1. earthquakes from years 01 to 2020
2. Of magnitude from 2.0 to 9.9.

4. Feature Analysis

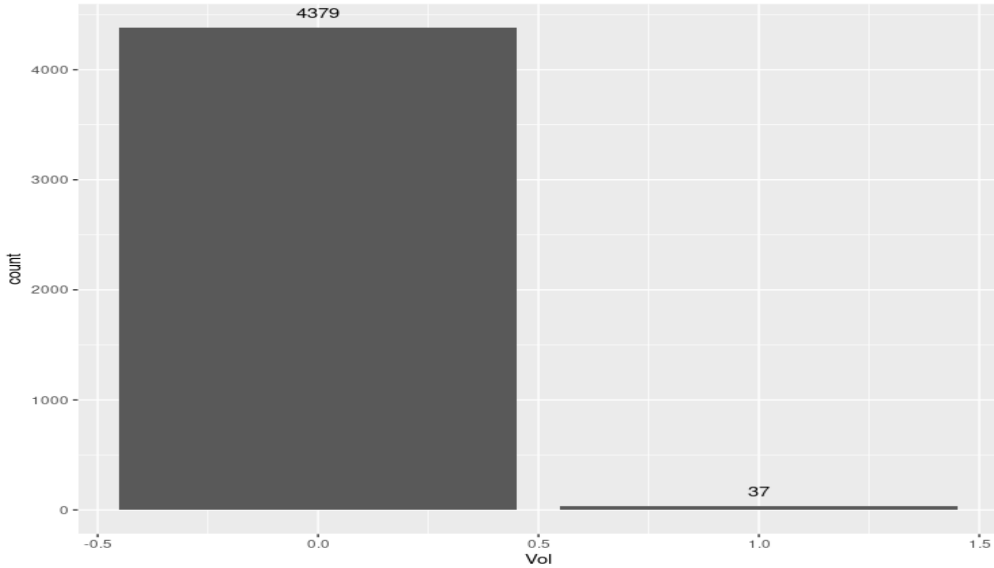


a. Vol

This feature tells if the earthquake was associated with a volcanic eruption. If we have VEI index, it implies the volcanic activity is associated. If null, no volcanic activity is associated

In the dataset, Value

0 = earthquake was not associated with a volcanic eruption.
1 = earthquake was associated with a volcanic eruption.



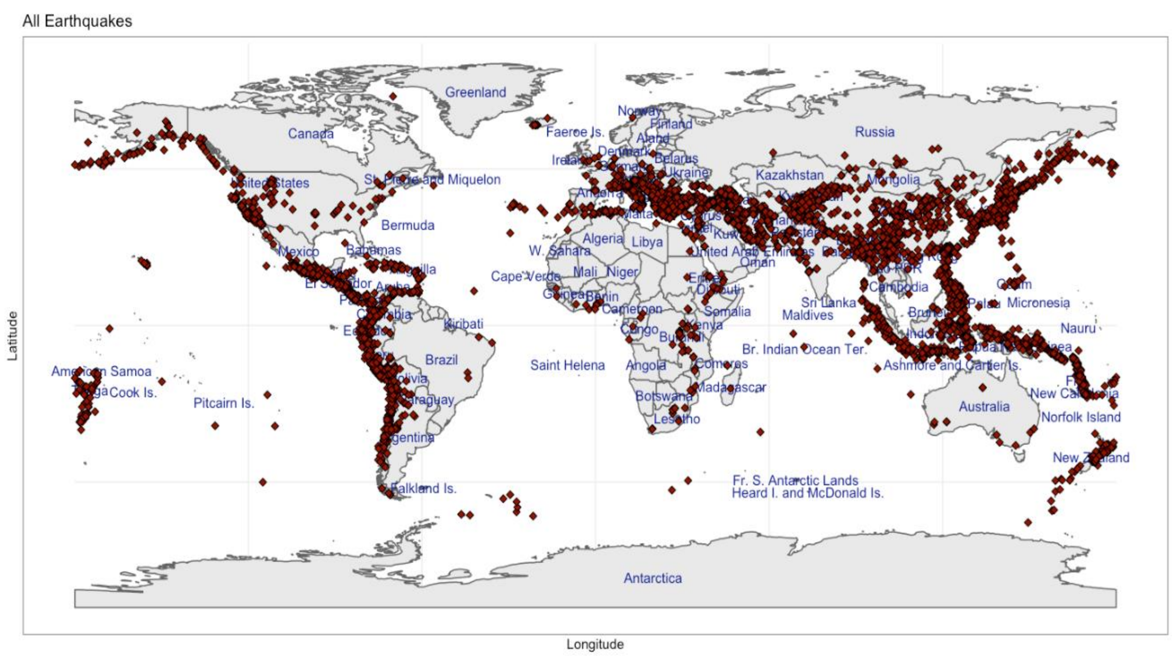
The dataset contains 4416 records and Out of which 4379 records are not associated with volcanic activities and 37 records are associated with volcanic activities.

b. Latitude and Longitude

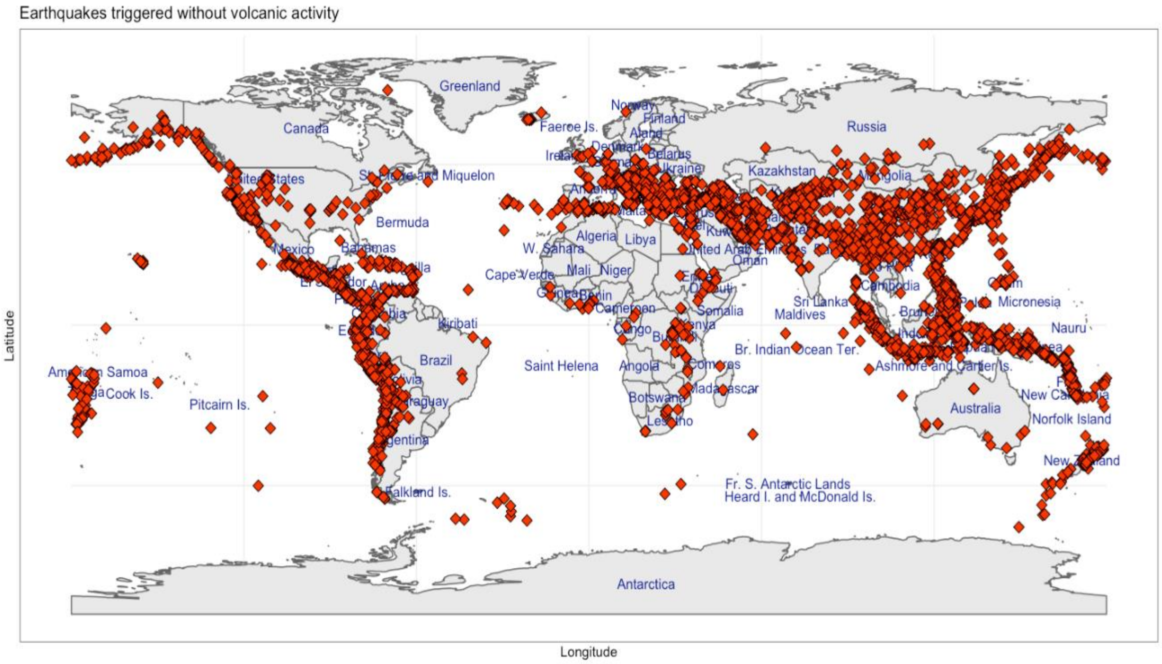
Latitude and longitude are the location of the earthquake on Earth.

Plots:

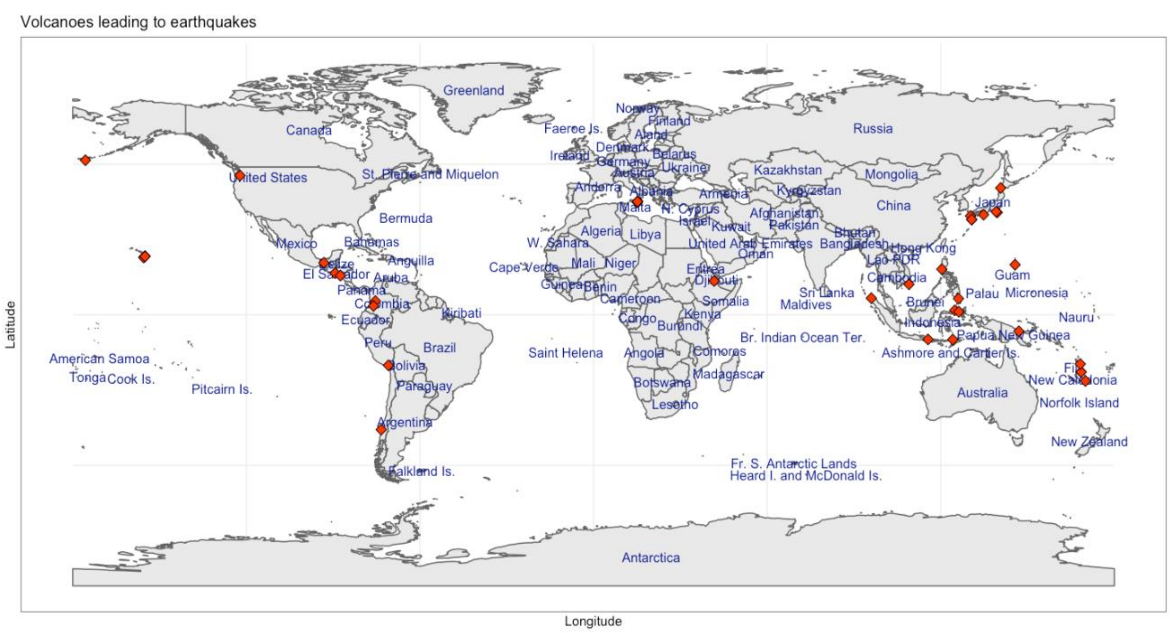
i. Geographical Plot of all Earthquakes in the dataset based on Latitude and Longitude.



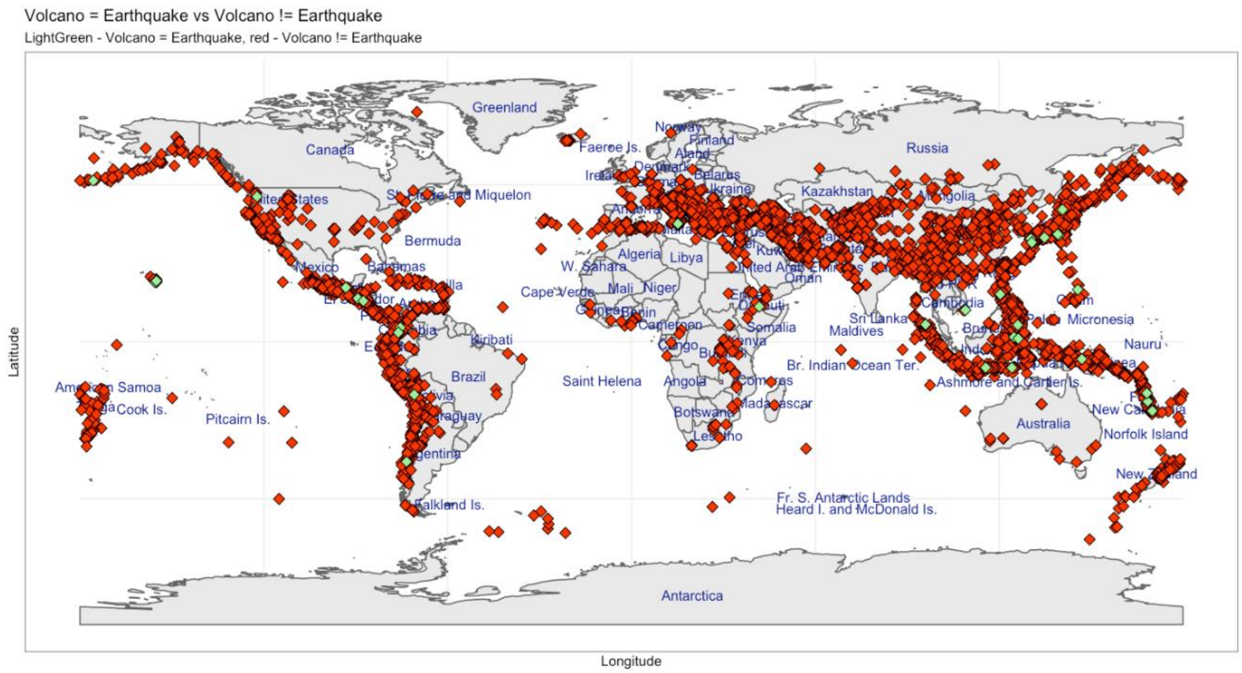
ii. Geographical Plot of all Earthquakes not associated with volcanic activity in the dataset based on Latitude and Longitude.



iii. Geographical Plot of all Earthquakes associated with volcanic activity in the dataset based on Latitude and Longitude.



Combined Geographical Plot: Green => Earthquake associated with Volcanic Activity
Red => Earthquake not associated with Volcanic Activity

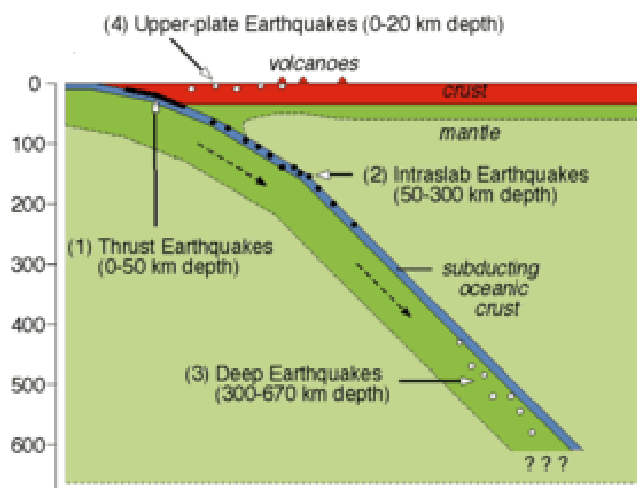


Key Observations:

1. Based on dataset, most of the earthquakes are happening on the islands around Pacific Ocean.
2. Earthquakes associated with Volcanic Activity are observed highly in islands around Japan and Indonesia.

b. Focal Depth

The depth of the earthquake, given in kilometers.

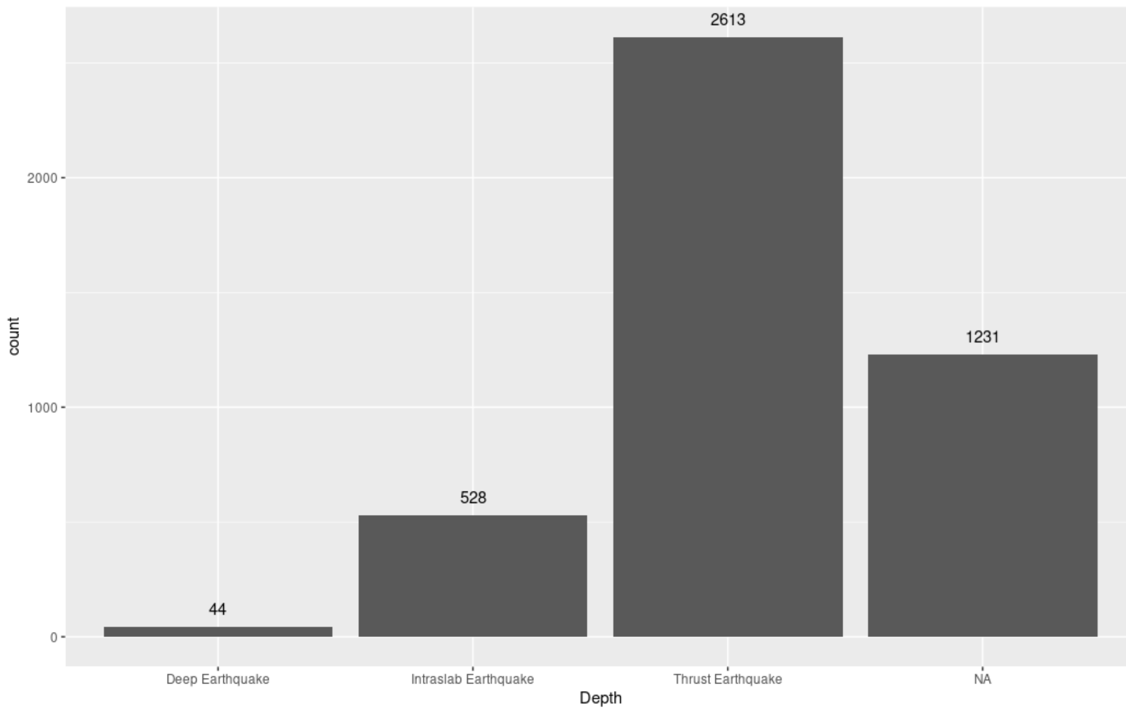


Imgsrc : https://earthquake.usgs.gov/learn/glossary/?term=focal%20depth

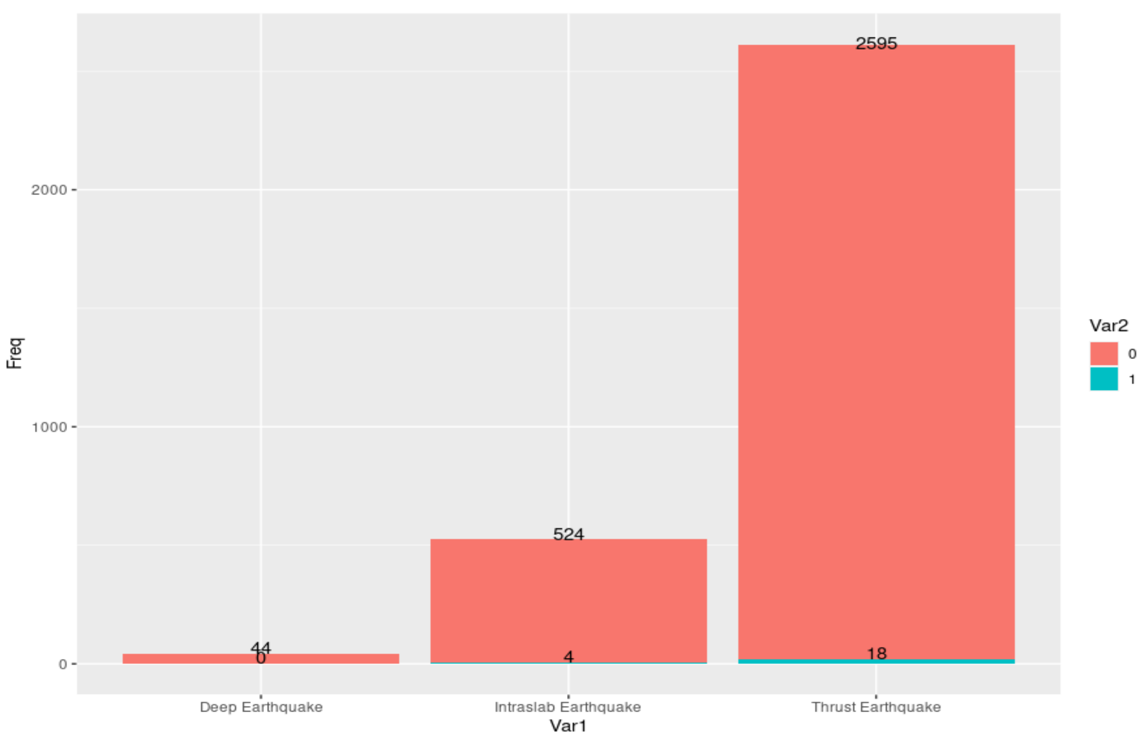
Earthquakes occurring at a depth of less than 50 km are classified as **Thrust Earthquakes**, while those with a focal depth between 50 km and 300 km are **Intra-slab**

Earthquakes. In subduction zones, where older and colder oceanic crust sinks under another tectonic plate, **Deep Earthquakes** may occur at much greater depths in the mantle, ranging from 300 to 670 km.

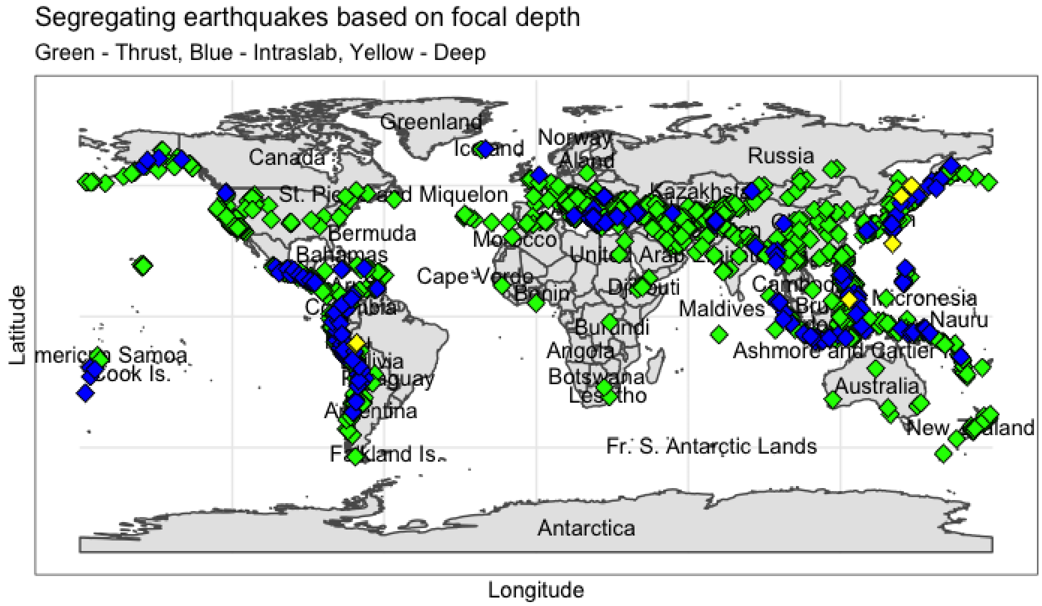
Plots:



Dataset contains 44 Deep Earthquakes, 528 Intra-slab Earthquakes, 2613 Thrust Earthquakes. Also, there are 1231 records without mention on Depth of Earthquake.



We can see 22 Earthquakes associated with volcanic activity with 18 of Thrust Earthquakes and 4 of Intra-slab Earthquakes. And 1216 of unspecified ones are not associated with Volcanoes and 15 are associated with Volcanic Activity. Following is the geographical plot for different types of earthquakes based on focal depth.

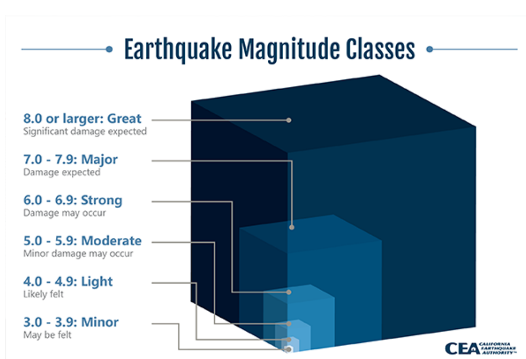


Key Observations

Observed that Volcanic Activities could not generate Deep earthquakes and if it generates, it could be more possibly be Thrust Earthquake

c. Magnitude

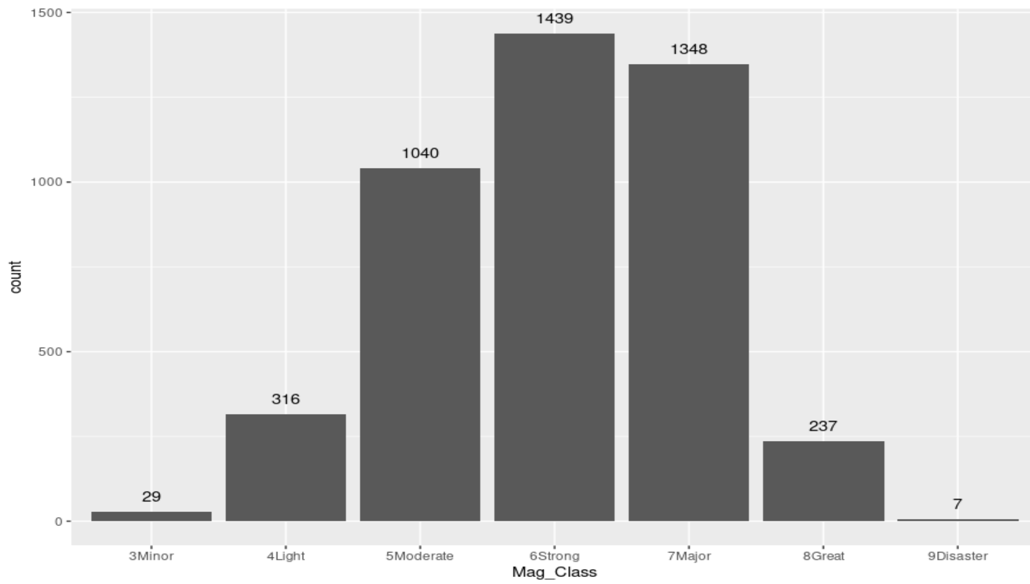
Magnitude measures the energy released at the source of the earthquake. Magnitude is determined from measurements on seismographs. Higher the Magnitude, more the energy released at the source of earthquake.



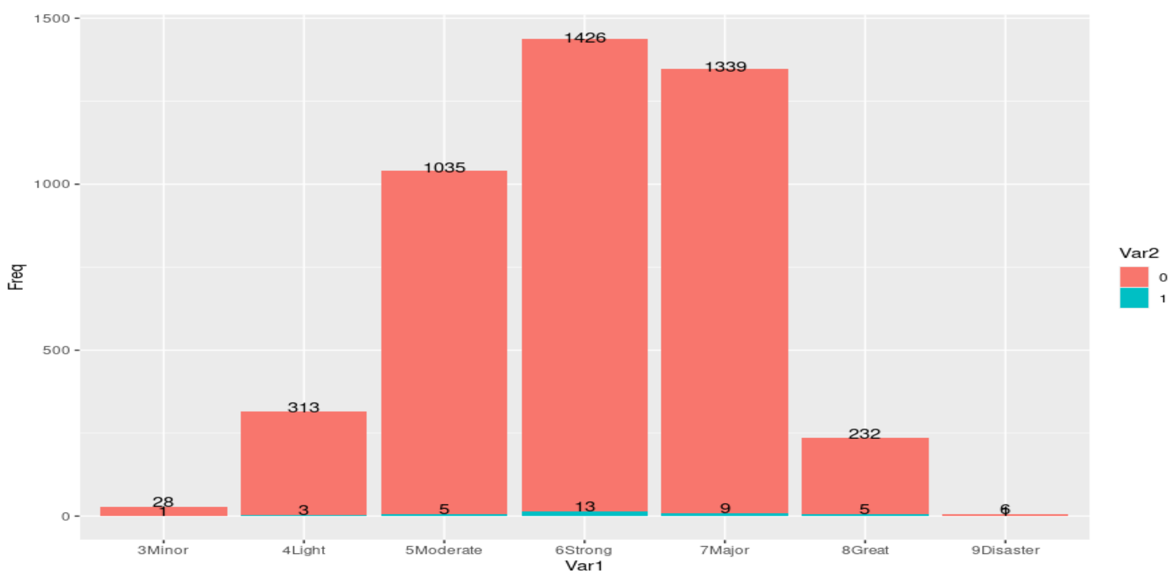
\Imgsrc: https://www.earthquakeauthority.com/Blog/2020/Earthquake-Measurements-Magnitude-vs-Intensity

Plot:

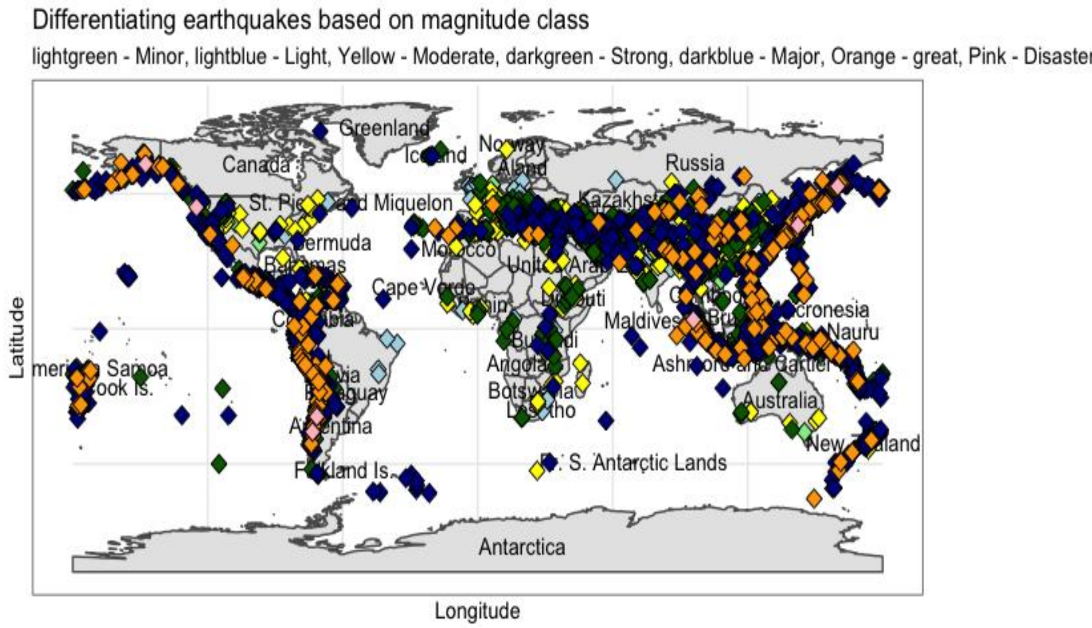
The plot talks about the count of earthquakes with certain magnitudes. After Scaling the magnitudes with the categories as represented in the picture implies following.



There are great number of earthquakes with magnitude between 5 and 8.



Following is the geographical plot of different magnitudes of the earthquakes.



Key Observations

1. Volcanic associated Earthquakes are greatly hit the magnitudes more than 5.
2. More often they generate Strong and Major Earthquakes.
3. Out of all Minor and Disaster Earthquakes, one of each earthquake are associated with Volcanic Activity.
4. There is high chance of earthquake with Great Magnitude being of associated with Volcanic Activity.

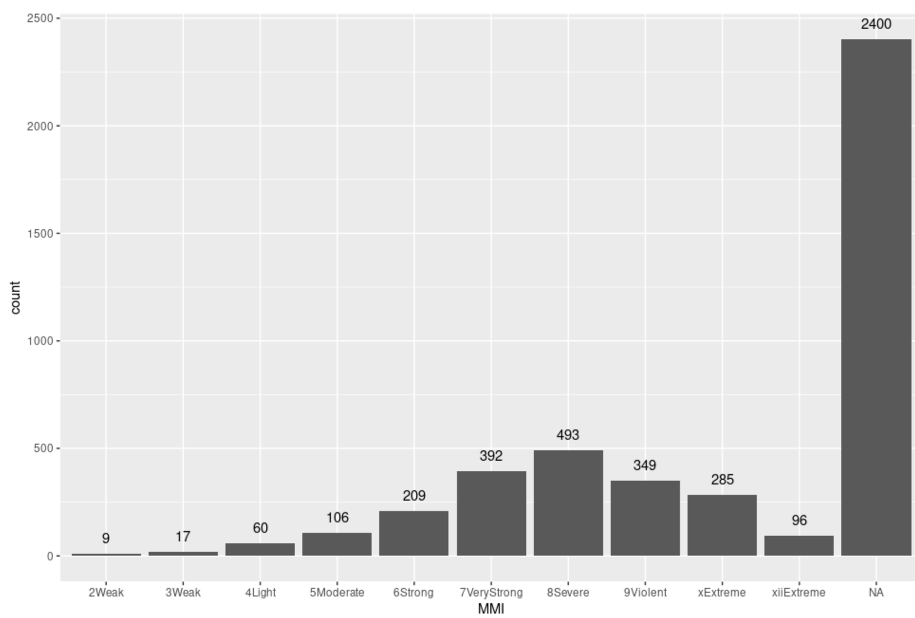
c. Modified Intensity Index

The measurement explains the severity of earthquake shaking and its effects on people and their environment. The lesser degrees of the MMI scale generally describe the manner in which the earthquake is felt by people. The greater numbers of the scale are based on observed structural damage.

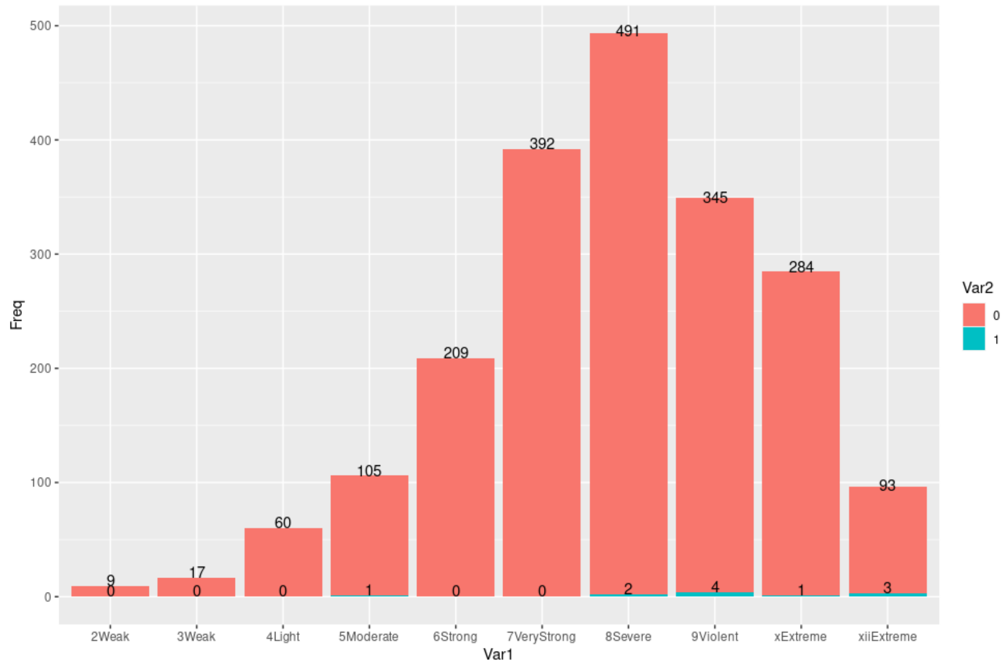
Scale level	Ground conditions
I. Not felt	Not felt except by very few under especially favorable conditions.
II. Weak	Felt only by a few people at rest, especially on upper floors of buildings.
III. Weak	Felt quite noticeably by people indoors, especially on upper floors of buildings: Many people do not recognize it as an earthquake. Standing motor cars may rock slightly. Vibrations are similar to the passing of a truck, with duration estimated.
IV. Light	Felt indoors by many, outdoors by few during the day: At night, some are awakened. Dishes, windows, and doors are disturbed; walls make cracking sounds. Sensations are like a heavy truck striking a building. Standing motor cars are rocked noticeably.
V. Moderate	Felt by nearly everyone; many awakened: Some dishes and windows are broken. Unstable objects are overturned. Pendulum clocks may stop.
VI. Strong	Felt by all, and many are frightened. Some heavy furniture is moved; a few instances of fallen plaster occur. Damage is slight.
VII. Very strong	Damage is negligible in buildings of good design and construction; but slight to moderate in well-built ordinary structures; damage is considerable in poorly built or badly designed structures; some chimneys are broken.
VIII. Severe	Damage slight in specially designed structures; considerable damage in ordinary substantial buildings with partial collapse. Damage great in poorly built structures. Fall of chimneys, factory stacks, columns, monuments, walls. Heavy furniture overturned. Examples include: 1925 Charlevoix–Kamouraska earthquake [17] and the 2000 Nicaragua earthquake.[18]
IX. Violent	Damage is considerable in specially designed structures; well-designed frame structures are thrown out of plumb. Damage is great in substantial buildings, with partial collapse. Buildings are shifted off foundations. Liquefaction occurs. Examples include: 2004 Indian Ocean earthquake and tsunami[19] and the 2011 Tōhoku earthquake and tsunami[20]
X. Extreme	Some well-built wooden structures are destroyed; most masonry and frame structures are destroyed with foundations. Rails are bent. Examples include 1939 Chillán earthquake[21] and the 1960 Agadir earthquake.[22]
XI. Extreme	Few, if any, (masonry) structures remain standing. Bridges are destroyed. Broad fissures erupt in the ground. Underground pipelines are rendered completely out of service. Earth slumps and land slips in soft ground. Rails are bent greatly. Examples include 1819 Rann of Kutch earthquake,[23] 1964 Alaska earthquake[24] and the 1976 Tangshan earthquake.[25]
XII. Extreme	Damage is total. Waves are seen on ground surfaces. Lines of sight and level are distorted. Objects are thrown upward into the air. Examples include 1960 Valdivia earthquake [26] and the 1920 Haiyuan earthquake.[27]

Imgsrc : https://en.wikipedia.org/wiki/Modified_Mercalli_intensity_scale

Plots:

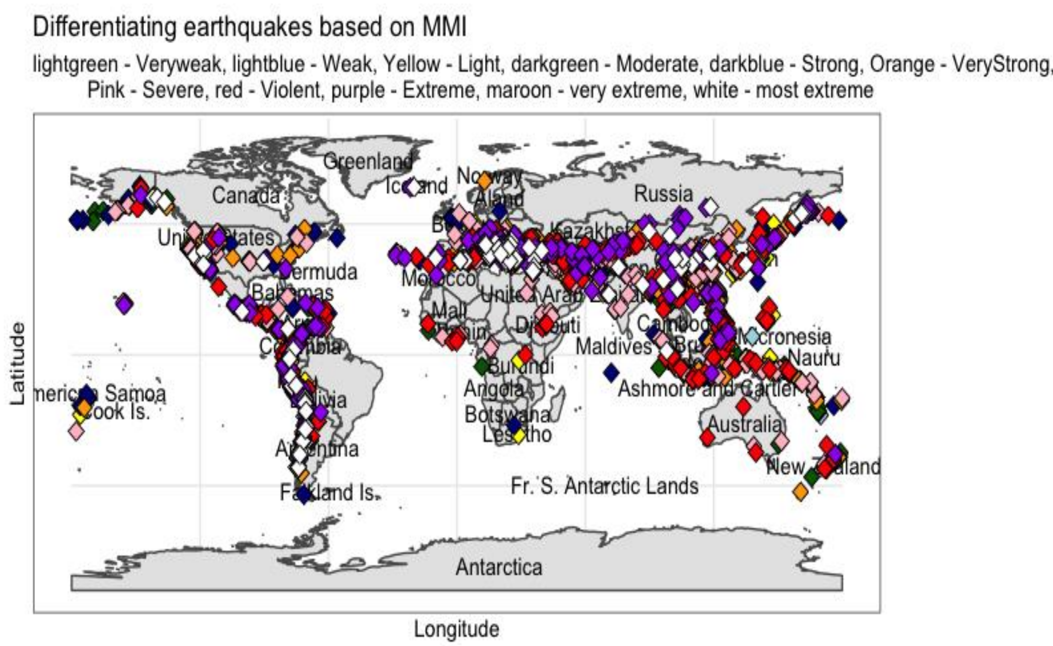


Earthquakes in the dataset are felt by people (Very Strong) in high number and caused damage to the buildings (Severe and above). Less number of earthquakes have impacted less.



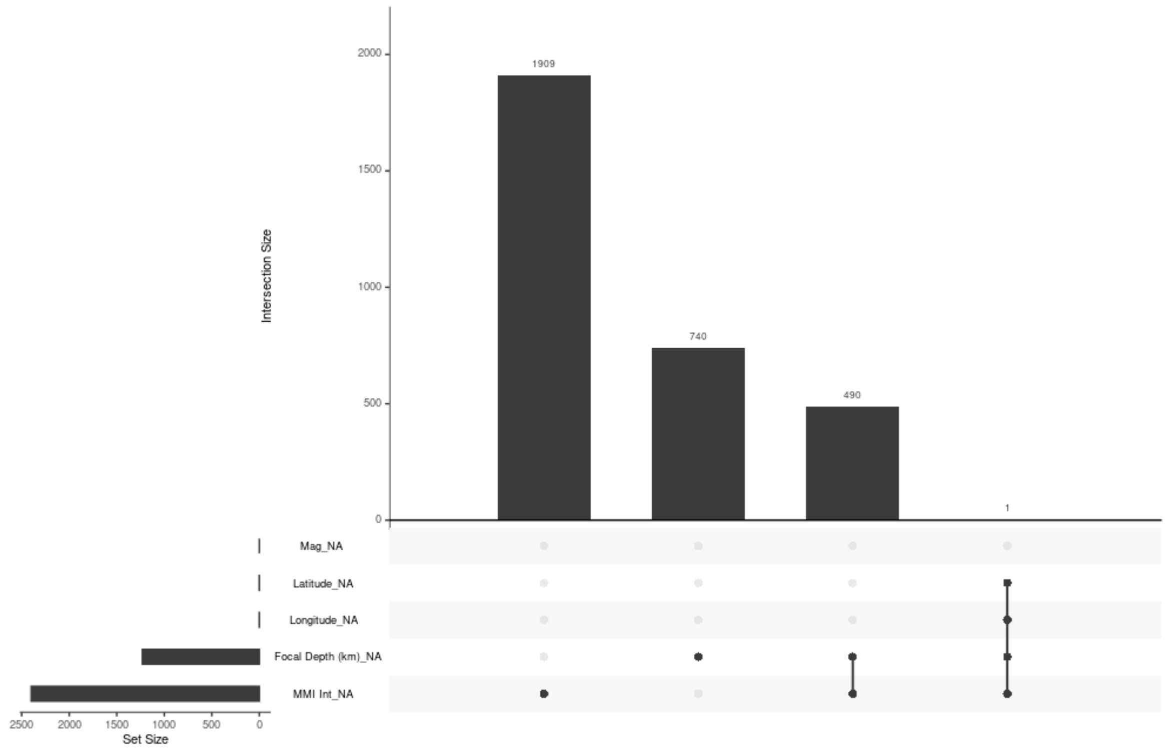
Other than 1 Moderate Earthquake in the valued ones, other earthquakes associated with Volcanic Activity impacted the people and buildings a lot.

Following is the geographical plot for different MMIs.

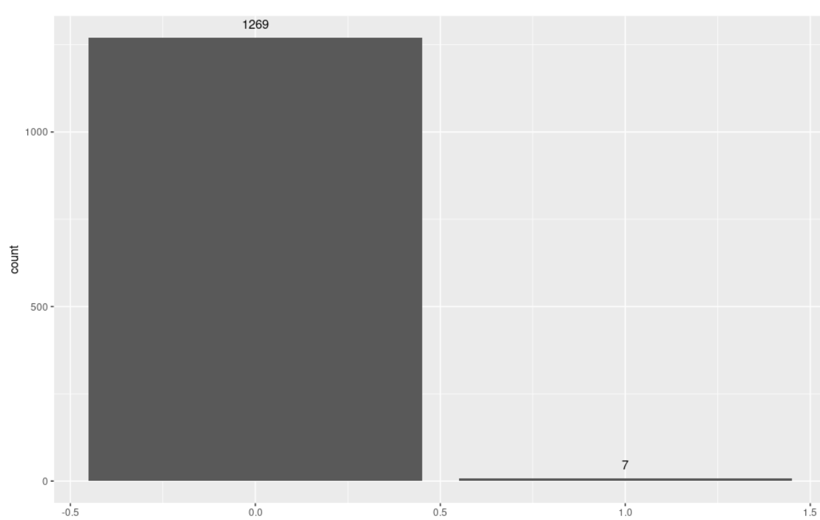


5. Data Balancing

Let's check the missing values in the dataset and balance the data before we perform bi-variate analysis. And also try to under sample the dataset since it is highly imbalanced.



As there are many missing values of features, 'Focal Depth (km)' and 'MMI Int' in the dataset. We try to read only the features that have no nulls in the record. As we cannot feed any inappropriate random values into the dataset that could give fallible parameters during fitting.

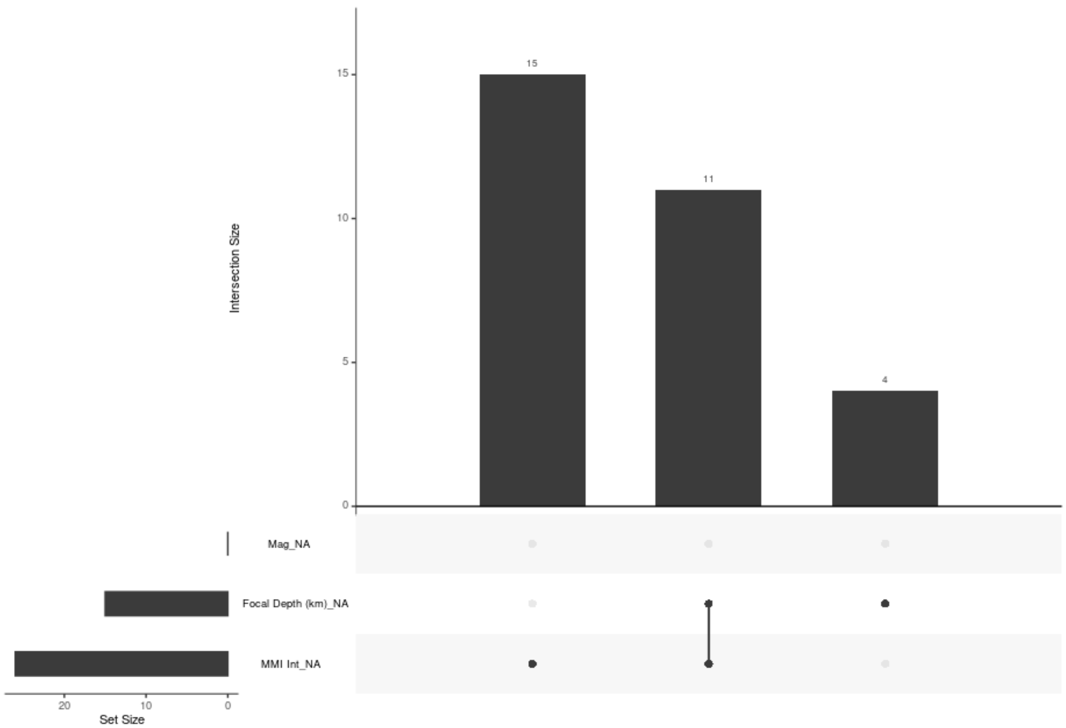


Records left after omission

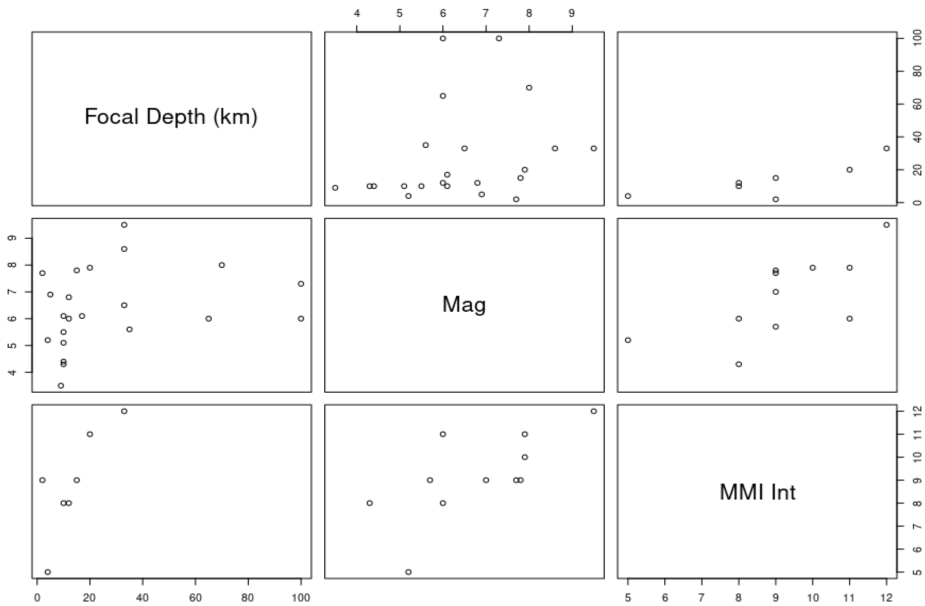
ALERT:
Omitting the record if any feature in the record has a null is removing the records that have volcanic activity associated (Vol=1) to the earthquakes in great number. But we have good number of records to read at for earthquakes that have not associated with the volcanic activity (Vol=0)

So now we put back the records that have volcanic activity associated (Vol =1) and will try to populate the missing data by fitting the missing values based on other features. i.e., we populate the missing values of MMI Intensity based on Magnitude finding the relation and doing a fit.

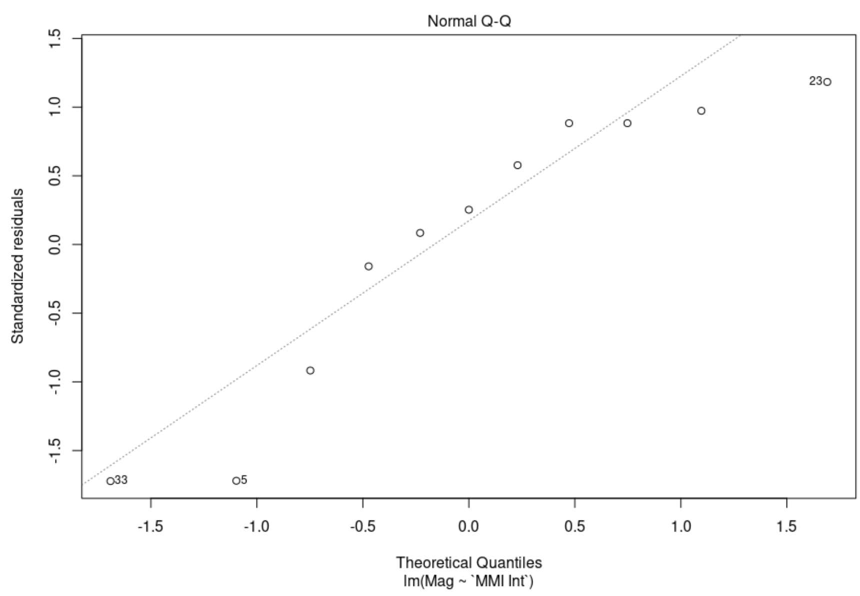
Now we have 1269 records as Vol = 0 and 7 records as Vol=1 without any null values in its other features. Now we got to populate null values in other features for the 30 Vol=1 record.



Checking the scatter matrix to find the relation of missing value features with non-missing one i.e., Magnitude.



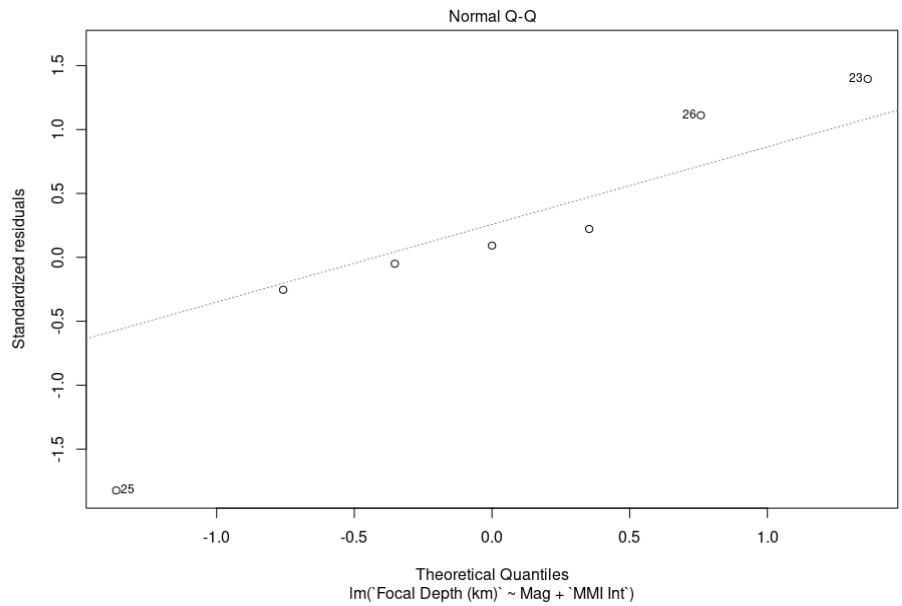
Observed that they both have linear Relationship, so we now try to fit the other columns based on the Magnitude.
By doing a linear fit on MMI using Magnitude, since they both are related on the existing records of Vol=1, QQ Plot looks like



The parameters of fit are:

Intercept	`MMI Int`
3.38	0.852

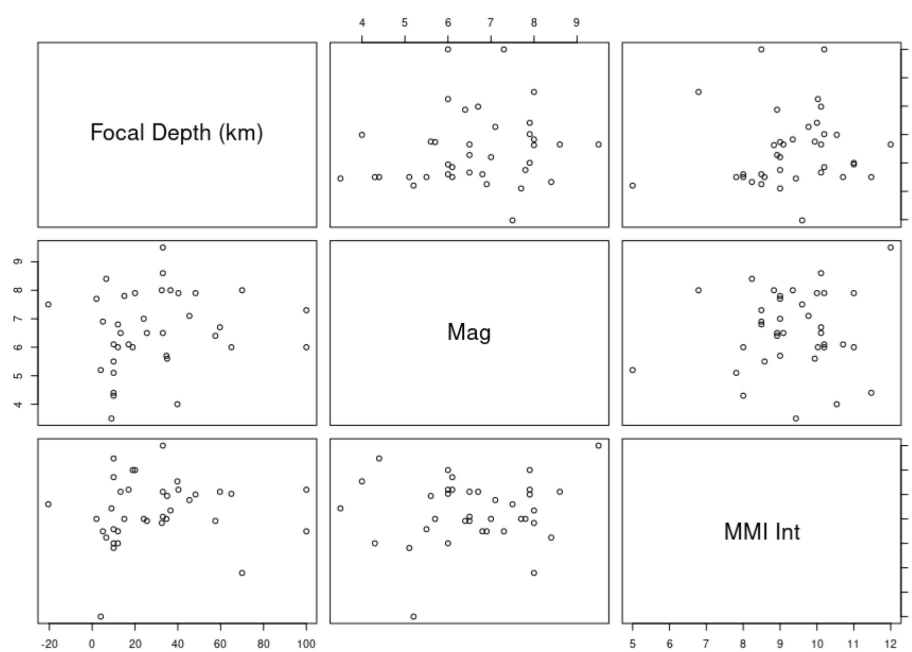
By doing a linear fit on Focal Depth using MMI and Magnitude, QQ Plot looks like



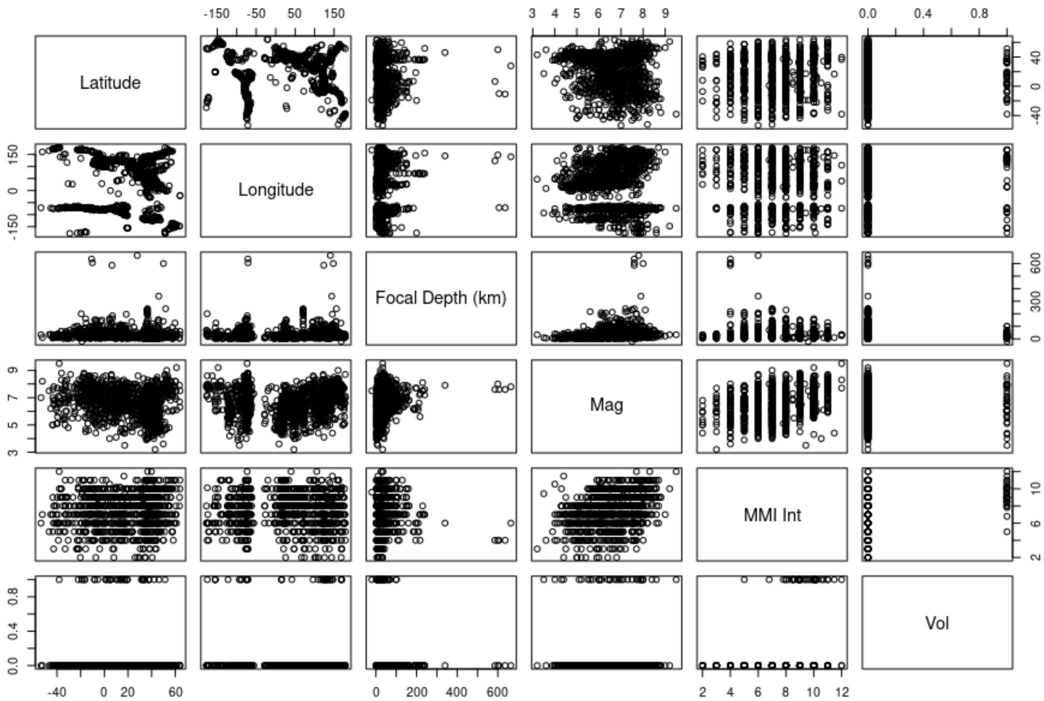
The parameters of fit are

(Intercept)	\`MMI Int\`	Mag
-20.951	-5.579	14.810

Scatter matrix of records with Vol=1 after loading the missing values using linear fit.



Visualizing Whole Scatter Matrix (Vol=1 and Vol=0) to find relations



Observations:

Mag and MMI Int are positively correlated.

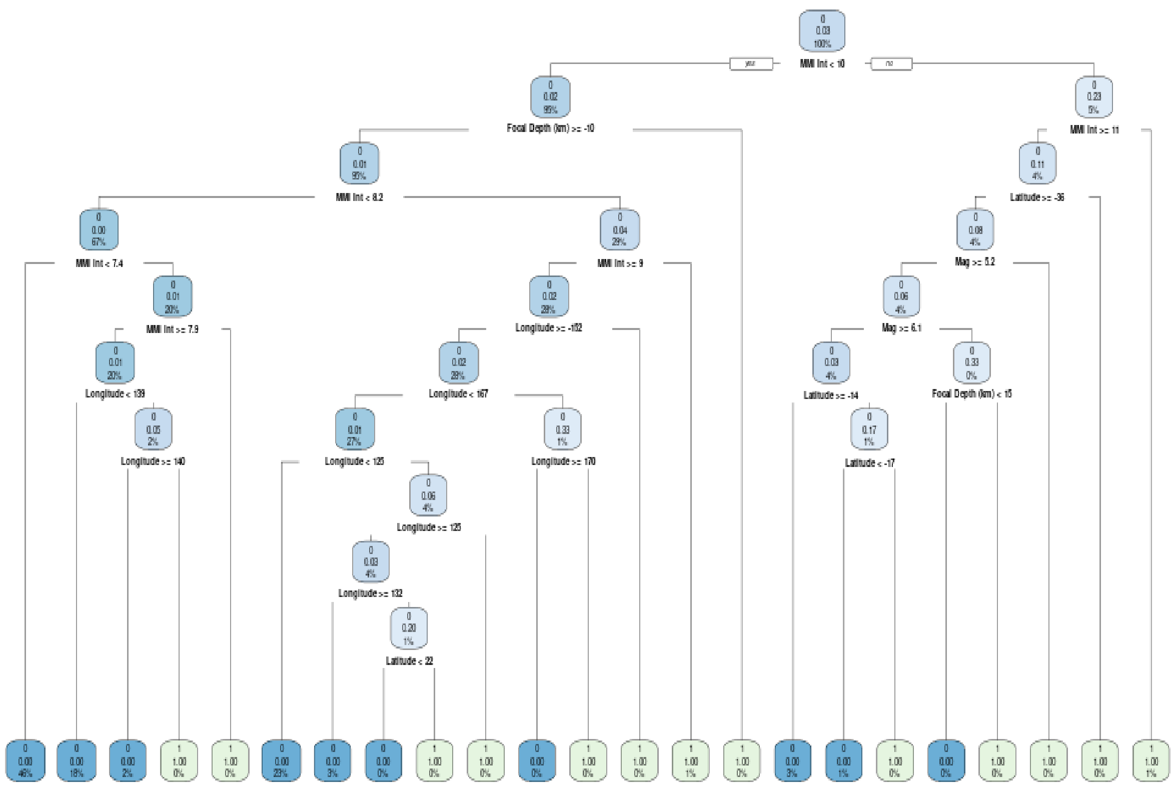
6. Modelling

Following is the distribution of the data in the 0-class and 1-class. The data is split into training & testing set in the ratio **70:30**. The test set is normalized using the mean and variance of the training set. Here, positive class = 1 (Volcanic Earthquakes) and negative class = 0 (Non-Volcanic Earthquakes)

Non-Volcanic Earthquakes	Volcanic earthquakes
1269	37

1. Decision Trees

Following is the decision tree obtained over the entire data.



	True Class (0)	True Class (1)
Predicted Class (0)	378	3
Predicted Class (1)	0	11

Accuracy	Specificity	Sensitivity	Balanced-Accuracy
0.9923	0.7857	1.0000	0.8929

Key Observations:
- Along with the accuracy, the metrics like specificity, sensitivity and balanced accuracy are good enough.
- Misclassification rate is less when compared to other ensemble models.

2. Random Forest

Random Forests are ensemble of random decision trees.

	True Class (0)	True Class (1)
Predicted Class (0)	378	13
Predicted Class (1)	0	1

Accuracy	Specificity	Sensitivity	Balanced-Accuracy
0.9668	1.000000	0.071429	0.535714

Key Observations:
- Though the overall accuracy is decent enough, the sensitivity is very low due to the high misclassification rate. Hence, this model cannot be generalized for both the classes.

3. Logistic Regression with SMOTE:

It is evident that volcanic earthquakes are very sparse, which leads to an imbalance in the classes. To overcome the imbalance, Synthetic Minority Oversampling Technique (SMOTE) before we fit in the data into non-tree machine learning algorithms.

SMOTE is an oversampling technique where the synthetic samples are generated for the minority class. This algorithm helps to overcome the overfitting problem posed by random oversampling. It focuses on the feature space to generate new instances with the help of interpolation between the positive instances that lie together (considering the neighbours)

	True Class (0)	True Class (1)
Predicted Class (0)	275	4
Predicted Class (1)	103	10

Accuracy	Specificity	Sensitivity	Balanced-Accuracy
0.727	0.72751	0.71429	0.72090

Key Observations:

- The overall metrics of the log-reg model are good when SMOTE is used.

4. Naïve Bayes with SMOTE:

Naive Bayes algorithm assumes that all the features are independent of each other. But we know that few features are correlated, hence affecting the accuracy.

	True Class (0)	True Class (1)
Predicted Class (0)	244	5
Predicted Class (1)	134	9

Accuracy	Specificity	Sensitivity	Balanced-Accuracy
0.6454	0.64550	0.64286	0.64418

Key Observations:

- Naive bayes assumes that the predictors are independent of each other, but we have observed that Mag and MMI Int are positively correlated. Due to this, the overall performance metrics of this model are not good enough, even after using SMOTE.

5. K Nearest Neighbors with SMOTE:

On training the model for different values of K (from 5 to 25), considering the training error, for K = 23, the model produced the best accuracy for the test data.

	True Class (0)	True Class (1)
Predicted Class (0)	315	4
Predicted Class (1)	63	10

Accuracy	Specificity	Sensitivity	Balanced-Accuracy
0.8291	0.83333	0.71429	0.77381

Key Observations:

- KNN (for K = 23) produced good results when SMOTE is used. The model can be generalized for data belonging to both the classes.

7. Conclusion

Following are the circumstances under which volcanic-earthquakes might occur.

- MMI > 10 and MMI < 11 => Vol 1
- MMI > 10 and MMI >= 11 and Latitude < -36 => Vol 1
- MMI > 10 and MMI >= 11 and Latitude >= -36 and Mag <5.2 => Vol1
- MMI > 10 and MMI >= 11 and Latitude <= -36 and Mag >=5.2 and Mag<6.1 and Focal Depth>=5=> Vol 1
- MMI > 10 and MMI >= 11 and Latitude <= -36 and Mag >=5.2 and Mag>=6.1 and Latitude<-14 and Latitude>=-17 => Vol 1
- MMI<10 and FocalDepth<-10 => Vol1
- MMI<10 and FocalDepth>-10 and MMI>=8.2 and MMI <9 => Vol1
- MMI<10 and FocalDepth>-10 and MMI>=8.2 and MMI >=9 and Longitude<-152 => Vol1
- MMI<10 and FocalDepth>-10 and MMI>=8.2 and MMI >=9 and Longitude>=152 and Longitude>=167 and Longitude<170 => Vol1
- MMI<10 and FocalDepth>-10 and MMI>=8.2 and MMI >=9 and Longitude>=152 and Longitude<167 and Longitude>=125 and Longitude<125 => Vol1
- MMI<10 and FocalDepth>-10 and MMI>=8.2 and MMI >=9 and Longitude>=152 and Longitude<167 and Longitude>=125 and Longitude>=125 and Longitude<132 and Latitude>=22=> Vol1
- MMI<10 and FocalDepth>-10 and MMI<8.2 and MMI>=7.4 and MMI<7.9 => Vol1
- MMI<10 and FocalDepth>-10 and MMI<8.2 and MMI>=7.4 and MMI>=7.9 and Longitude >=139 and Longitude<140 => Vol1

8. References

https://en.wikipedia.org/wiki/Volcano_tectonic_earthquake#:~:text=A%20volcano%20tectonic%20earthquake%20is,rock%20to%20break%20or%20move.

https://earthquakeandvolcano.weebly.com/volcanic-earthquakes.html

https://www.r-spatial.org/r/2018/10/25/ggplot2-sf.html

https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/

9. Contribution

Akhil, Rajyavardhan and Sahiti Krishna conceived of this project. Akhil and Rajyavardhan wrote the introduction. Data preprocessing was done by all. Data Analysis was done by Akhil and Rajyavardhan. Generating models and performance evaluation was done by Akhil and Sahiti. The conclusion was written by Rajyavardhan and Sahiti. All authors read and revised the final report.

****