Violeta Lio King
2.11.2022

# Exploring Seismic Activity for the Past 30 Days

## Problem Statement

Earthquakes have long been unpredictable and it has long been the work of seismologists to link cause and effect of earthquakes, and a good starting point is exploring the relationships between earthquake magnitude and location, magnitude/depth/intensity, and patterns of frequency.

## Obtain

The United States Geological Survey (USGS) Earthquake Hazards program offers earthquake data on their website. For the purpose of answering all three questions, three separate data sets had to be obtained in order to capture all earthquakes that registered a magnitude of 2.5 and above. The first dataset captured smaller earthquakes of (magnitude 2.5+), the second captured larger earthquakes (magnitude 4.5+) and the third captured significant earthquakes. The period of record for all three datasets ranged over the last 30 days.

## Scrub

To prepare the data for analysis, the three datasets had to be merged. All three of them had the same parameters and number of columns. However, not all of the columns would be used for the analysis. The first challenge was to prepare the "Coordinates". This column contained the latitude, longitude, and depth of the earthquake all between brackets. This column was separated into three different columns to facilitate analysis. In addition, the time column had to be converted into a date. Its current format was recorded into milliseconds after the Unix epoch (Jan. 1st, 1970). Once converted, the column was split into a "Date" column and a "Time" column. The last major calculation was grouping the magnitudes by quartiles which would later be used for the map. The other columns of interest (Magnitude, ID, Intensity) were already in their ideal format so they were extracted as is. Once the final dataframe was created, it was then exported into a CSV file to be analyzed.

## Explore

The first exploration pertained to magnitude and location. The latitude and longitude were extracted from the scrubbed data and plotted on top of a world map. The "Quartile" column was used to group the data points according to the magnitude of the earthquake and a color map

was created in order to convey it visually (Fig. 1). The colors represent the quartiles ranging from the lowest (green) to the highest quartile (red). Then the tectonic plates were overlaid on top of the map in order to visually determine the relationship between earthquakes occurrence, location and magnitude. The map seems to imply that there is a strong relationship between the three.

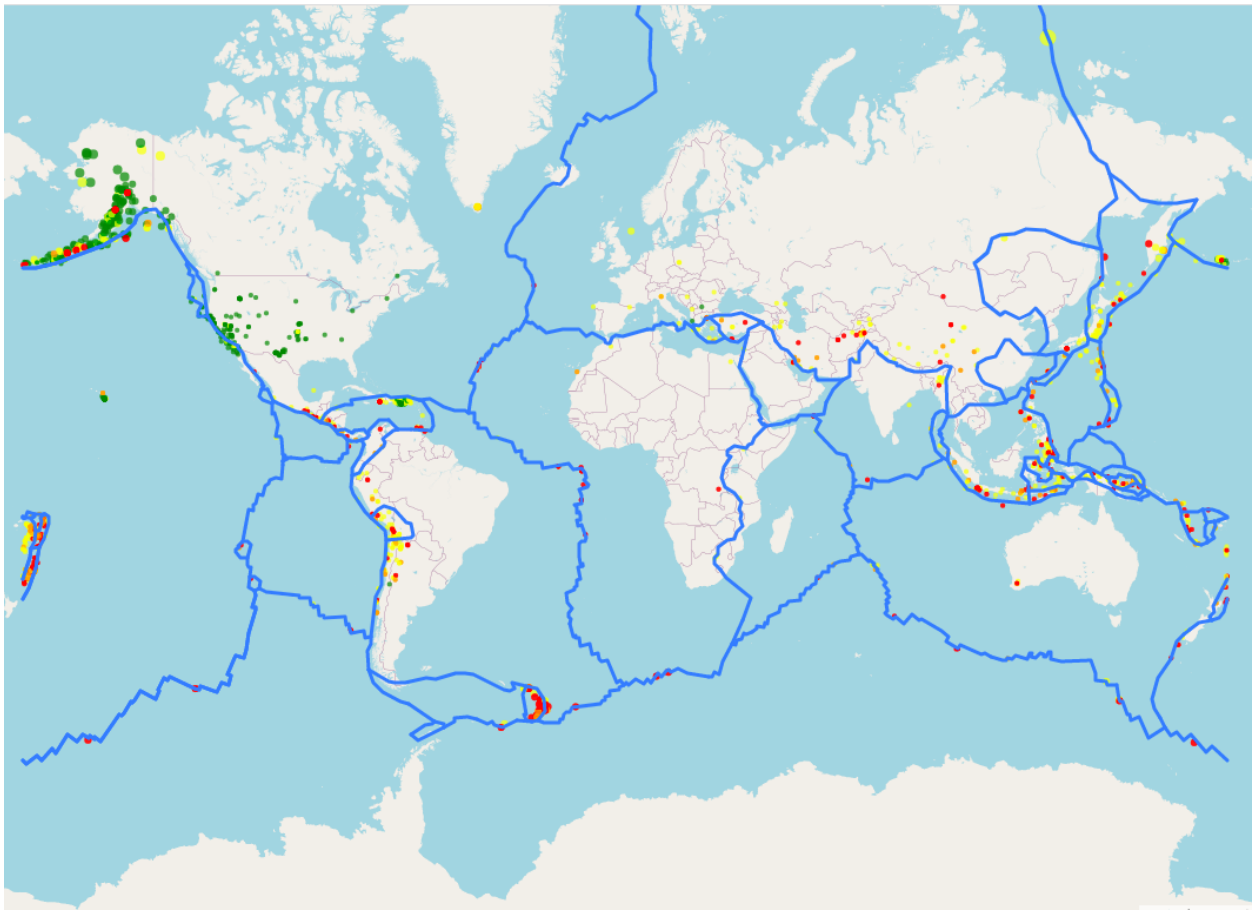

**Figure 1.** Earthquakes and their location. The colors represent the quartiles ranging from the lowest (green) to the highest quartile (red).

In order to obtain a better understanding of a potential empirical relationship between earthquake magnitude, intensity and the recorded depth, matrix plots were created using the pairplots() function in Python's Seabold library (Fig. 2). The plots between magnitude and intensity seem to indicate a positive and linear relationship. The plots between magnitude and depth suggest a weak positive and linear relationship. The histograms indicate that the data for magnitude and intensity follow a roughly normal distribution; the one for depth seems to be heavily skewed.

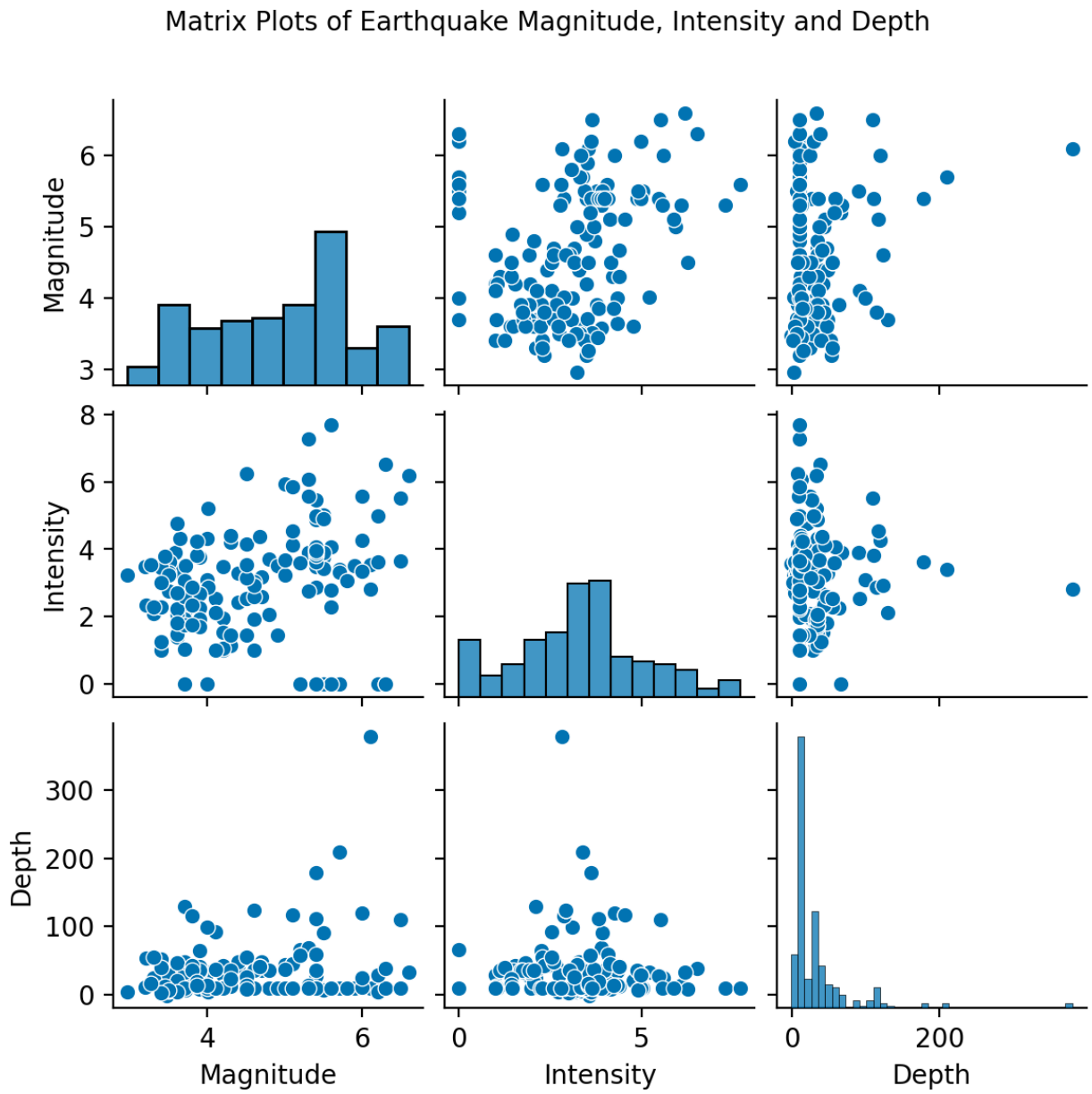


**Figure 2.** Pairplots between earthquake magnitude, intensity and depth. Generated using the Seabold library in Python.

Finally, the last part of the exploration wanted to determine if there is a pattern in seismic activity and when they occur. The first plot is a histogram of all the seismic events by the hour (Fig. 3). While there are some peaks and flows, there does not seem to be a discernible pattern. There does appear that many seismic events happen at 23 UTC. The second plot is another histogram but this time, the seismic events are sorted by the day in which they occur. The only

thing worth noting is that the seismic events were a little lower in the first 15 days but with an uptick in seismic events in the last 15 days.

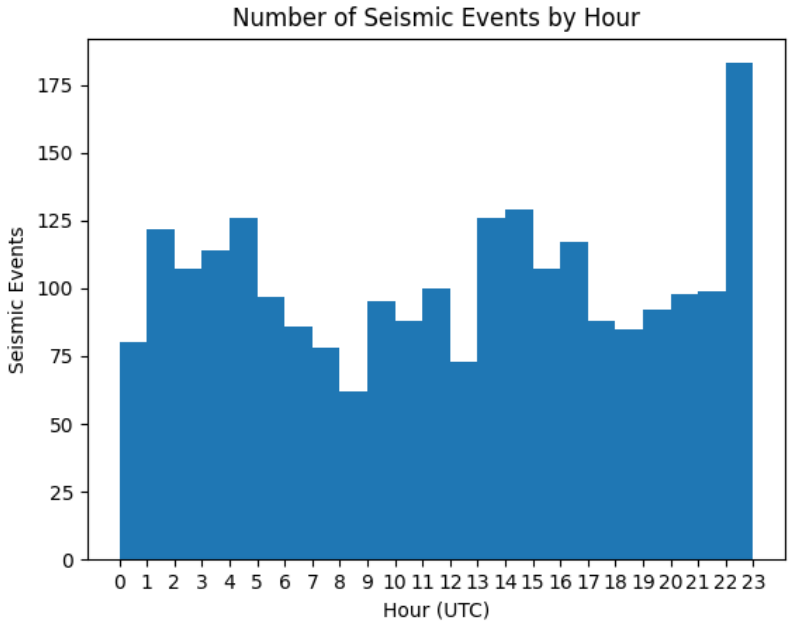


**Figure 3.** Histogram of all earthquakes over the past 30 days, broken down by hour.

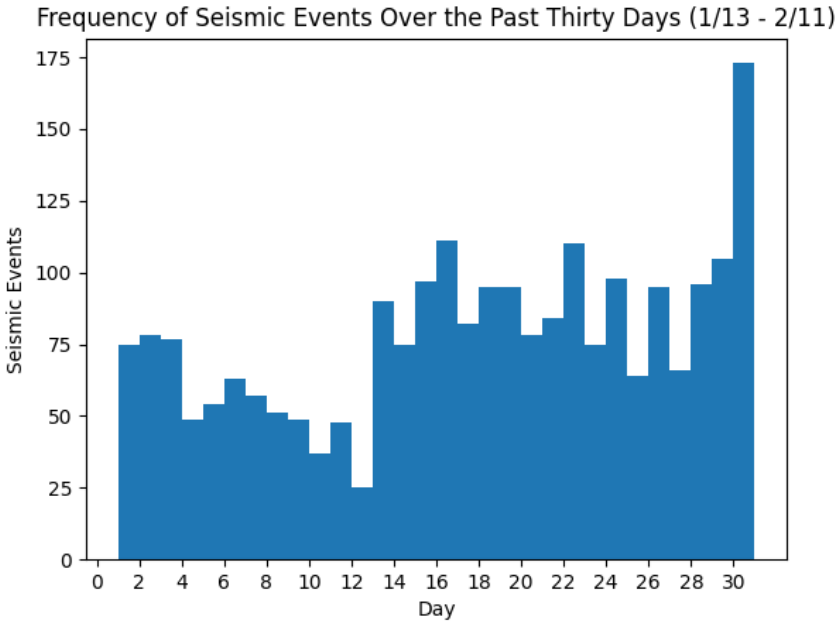

**Figure 4.** Histogram of all earthquakes over the past 30 days, broken down by day. Day 1 is January 13th and day 30 is February 11th.

**Model**

For the first and third part of the analysis, no model was chosen since it was a visual analysis of the data. However, for the second part of the analysis, a multiple linear regression was used. The exploratory variables were depth and intensity while the response variable was the magnitude of the earthquakes. The intercept was 4.401 while the coefficients for intensity and depth were 0.125 and 0.003, respectively.

In order to validate the model, additional tests were performed. The first was to calculate the coefficient of determination (R-squared). This was easily achieved by using the statsmodel library function OLS( ). The R-squared value was low (0.0778) indicating that there is no linear relationship between the parameters.

**Interpret**

For the first part of the analysis, it is clear that there is a relationship between the location of earthquakes and their occurrence. Most of them over the last 30 days occurred right along the tectonic plates. Interestingly enough, there did appear to be a difference between the location of the seismic events and their severity. Along Alaska and California, for example, most of the earthquakes were in the lowest quartile. Along South America and Japan, the magnitudes were higher since the earthquakes were mostly yellow dots and red. This could imply that there are some tectonic plates which hold a lot more energy than others or that the particular fault line tends to have more friction.

While initially promising, the model was not a good fit. Due to the low R-squared coefficient, there does not appear to be a linear relationship between the parameters. This could be because of a few factors. The first is that there is a relationship, but it might not be linear (despite the initial appearance of the pair plots). The second is probably the most likely, however, and that is that the data might not be recorded correctly. Upon further investigation, the USGS website indicated that the measurements for depth will sometimes default to certain values (1 meter) if certain conditions are met.

The third part of the analysis yielded no patterns of frequency whether it be by hours or days. Both plots showed an unusually high amount of seismic activity towards the right end of the plots. This can easily be explained by quality control and/or when the report was entered. For the hourly plot, it could be that the reports are not all entered until the end of the day. For the daily plot, the USGS has had more time to do quality control data on earlier reports than the most recent one. This also makes logical sense since earthquakes are spontaneous events in which the tectonic plates displace and release energy.

## Samples

{"type":"FeatureCollection","metadata":{"generated":1644731767000,"url":"https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/2.5_month.geojson","title":"USGS Magnitude 2.5+ Earthquakes, Past Month","status":200,"api":"1.10.3","count":1722},"features":[{"type":"Feature","properties":{"mag":3.5,"place":"56 km S of Whites City, New Mexico","time":1644729112128,"updated":1644729886089,"tz":null,"url":"https://earthquake.usgs.gov/earthquakes/eventpage/tx2022dayy","detail":"https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/tx2022dayy.geojson","felt":null,"cdi":null,"mmi":null,"alert":null,"status":"reviewed","tsunami":0,"sig":188,"net":"tx","code":"2022dayy","ids":",us7000gkfh,tx2022dayy,","sources":",us,tx,","types":",origin,phase-data,","nst":15,"dmin":0.09846342032,"rms":0.3,"gap":60,"magType":"ml","type":"earthquake","title":"M 3.5 - 56 km S of Whites City, New Mexico"},"geometry":{"type":"Point","coordinates":[-104.3803075,31.66774561,7.185595703]},"id":"tx2022dayy"}, {"type":"Feature","properties":{"mag":3.6,"place":"55 km S of Whites City, New Mexico","time":1644729028334,"updated":1644730661360,"tz":null,"url":"https://earthquake.usgs.gov/earthquakes/eventpage/tx2022dayw","detail":"https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/tx2022dayw.geojson","felt":null,"cdi":null,"mmi":4.293,"alert":null,"status":"reviewed","tsunami":0,"sig":199,"net":"tx","code":"2022dayw","ids":",us7000gkfg,tx2022dayw,","sources":",us,tx,","types":",origin,phase-data,shakemap,","nst":16,"dmin":0.09956655322,"rms":0.3,"gap":58,"magType":"ml","type":"earthquake","title":"M 3.6 - 55 km S of Whites City, New Mexico"},"geometry":{"type":"Point","coordinates":[-104.3816916,31.67210295,7.828295898]},"id":"tx2022dayw"}

## Sources

https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/2.5_month.geojson
https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/4.5_month.geojson
https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/significant_month.geojson