Analyzation and Visualization of Significant Earthquakes from 1965-2016

By Justin Yang

Why did I choose this dataset?

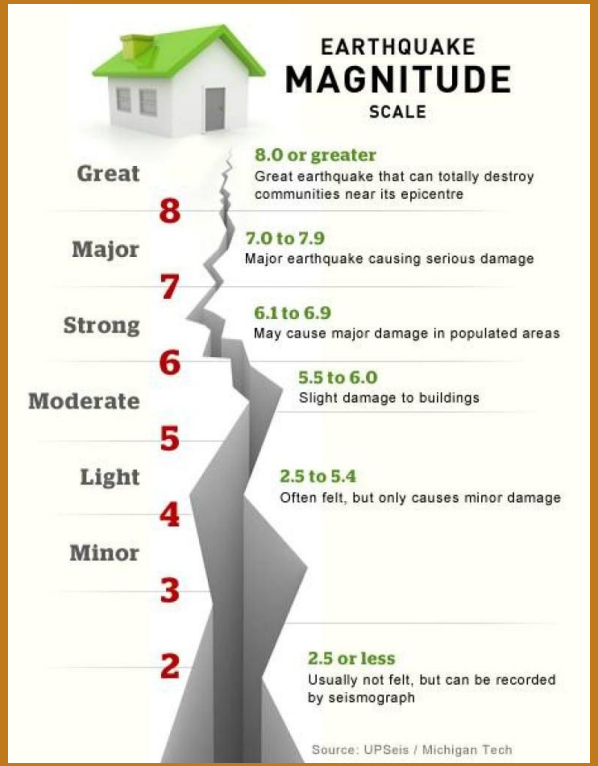
I chose ‘earthquakes.csv’ from Kaggle because I was curious about the trends of earthquakes happening over time, e.g. whether earthquakes generally got more violent over the years or whether they stayed relatively constant throughout the decades.

This data is important to locating earthquake hotspots and also a fundamental piece of information necessary for facilitating studies on Earth’s internal structure. This data allows scientists to better understand the epicenter and causes of earthquakes, in order to be more prepared to forecast an earthquake and set procedures for the public to limit the damage and loss of life caused by the earthquake.

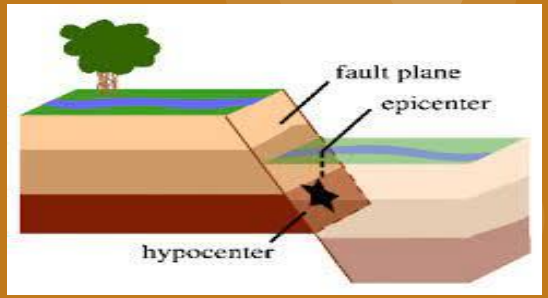
My Approach

I first wanted to understand what exactly the data in ‘earthquakes.csv’ is. Some of the columns had fields that I did not understand, such as Magnitude Type. My plan of action was to ‘clean’ up the dataset by removing columns that are unclear or unnecessary for my analysis. I excluded Nuclear Explosions from my dataset because I wanted to focus on Earthquakes specifically. Of course, all data is important but by limiting the types of data in the dataset, I am able to form more simple analysis without delving too deep. I used Python’s Jupyter Notebook, and specifically libraries such as pandas and matplotlib to create graphs and manipulate the data. With a cleaner, more simple dataset ‘cleaned_earthquakes.csv’, I was able to create other temporary datasets and form different graphs that depict the data.

Brief Overview of Earthquakes

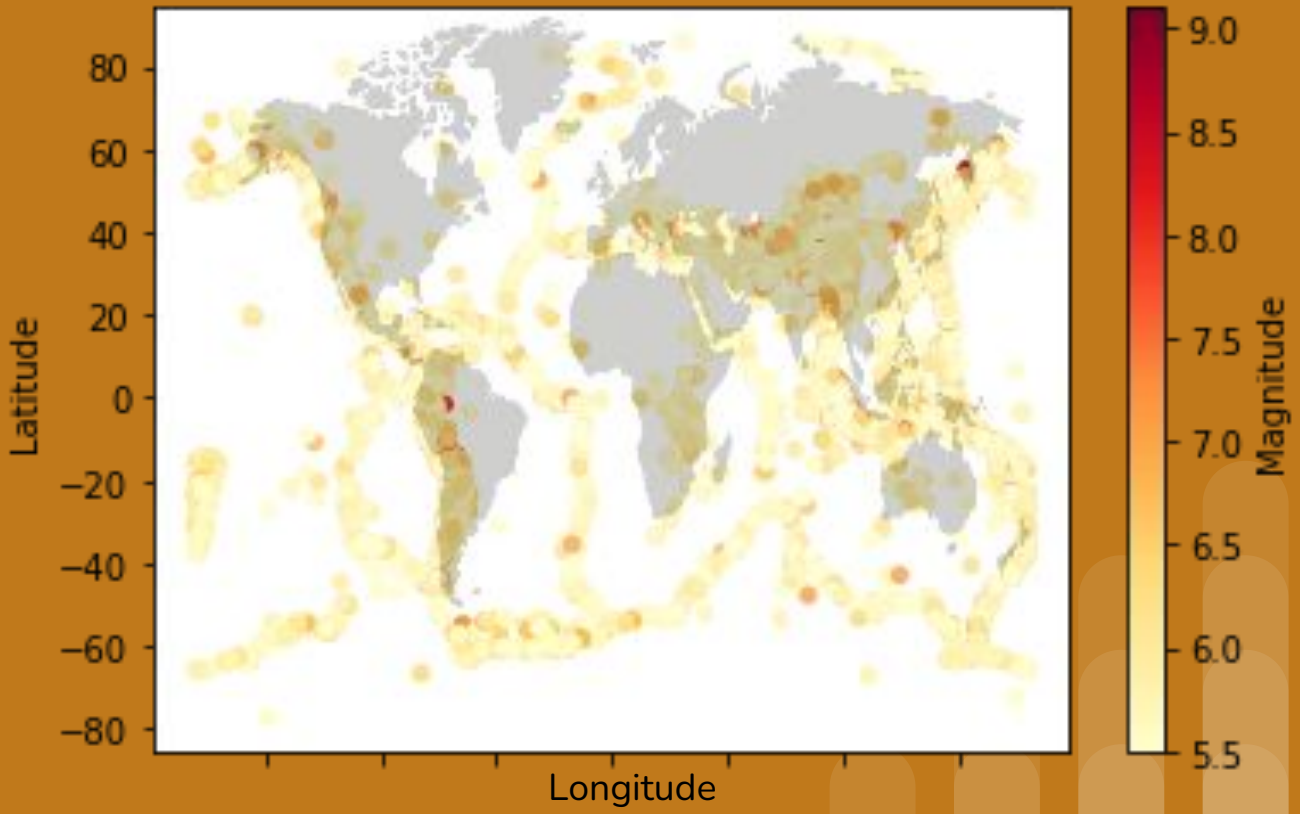


Quote USGS:
“Earthquake is a term used to describe both sudden slip on a fault, and the resulting ground shaking and radiated seismic energy caused by the slip, or by volcanic or magmatic activity, or other sudden stress changes in the earth.”
Earthquakes can inflict massive damage and understanding earthquakes can help prevent the loss of lives

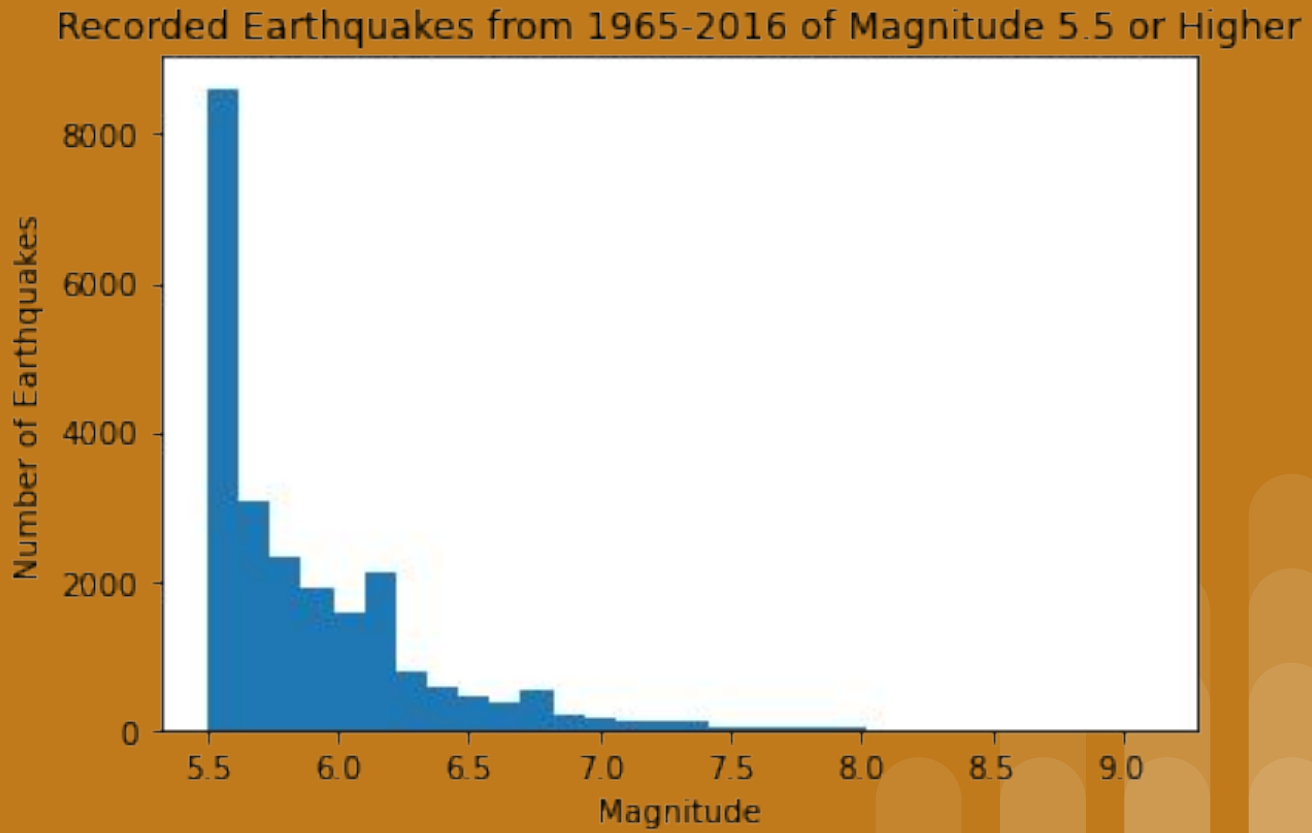


Scatter Plot of Earthquakes (magnitude 5.5+) Locations on World Map from 1965-2016

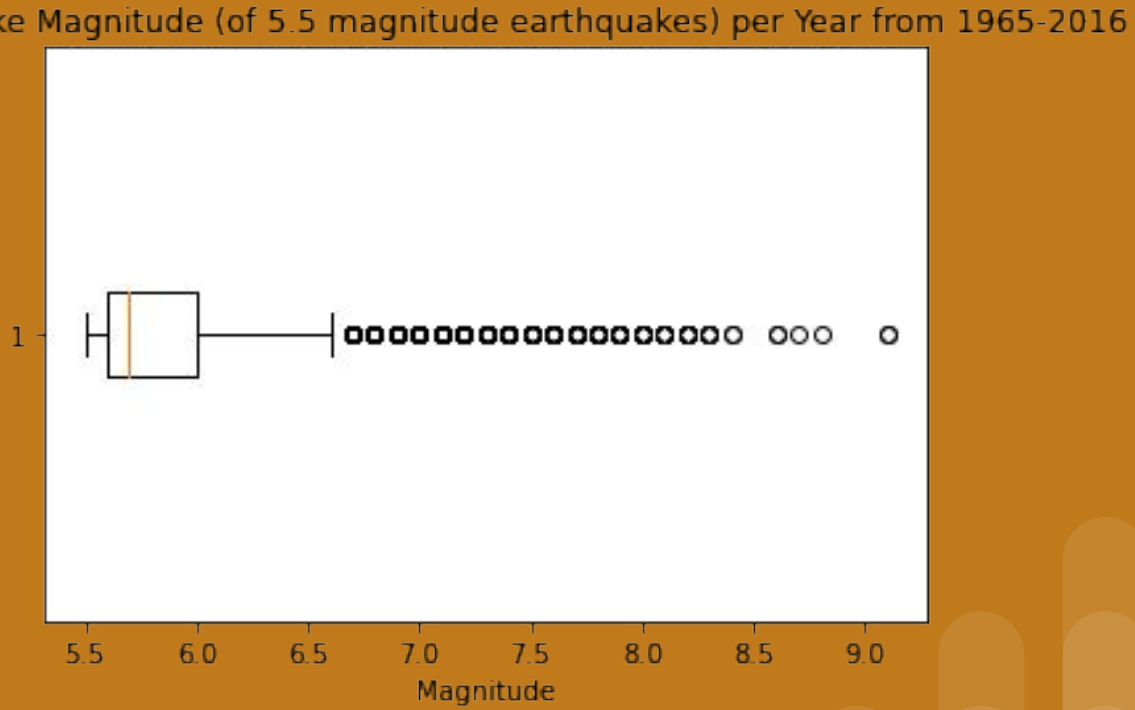
In this diagram, we can see the concentrated traces of earthquakes. These concentrated traces are Earth's fault lines. Through this diagram, we can see that the fault lines are what causes most earthquakes to occur. The Earth's crust is prone to shifting around these areas and thus generating earthquakes with greater magnitude.



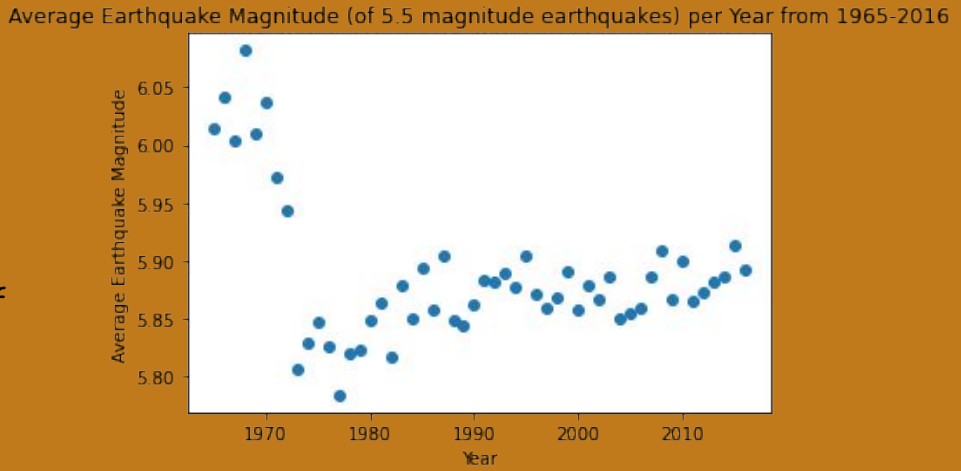
The histogram on the right displays the spread of magnitude of earthquakes from 1965-2016. We can see that the histogram is skewed right, whereas most earthquakes tend to have a magnitude between 5.5 - 6 on the richter scale. Only a few percent of earthquakes reached higher scales of magnitude

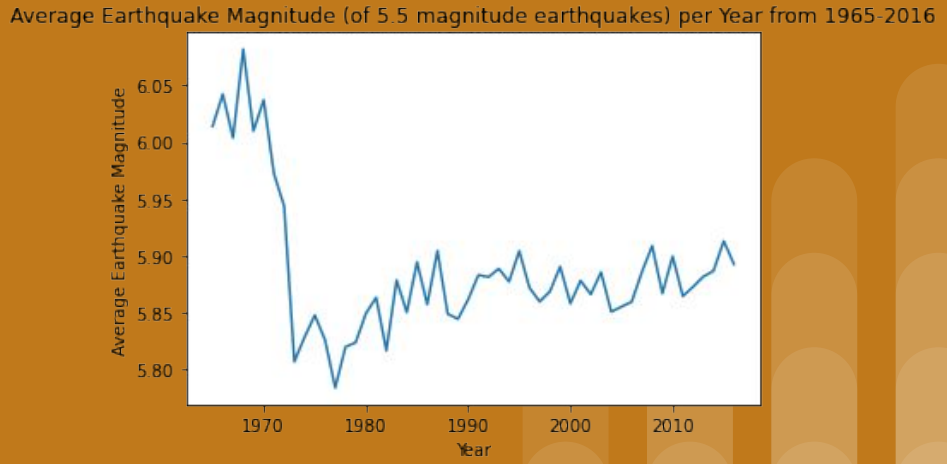


Here is another box plot visualization of the distribution of earthquakes based on their magnitude. Overall, 75% of recorded earthquakes 5.5 magnitude or higher lie between magnitudes of 5.5 to 6. There are few outliers that lie greater than 6.7 magnitude. This distribution is important to understand that although most earthquakes are within a certain magnitude range, building structure and procedures in inhabited areas should consider the possibility of a much greater earthquake to occur, especially in earthquake-prone areas.

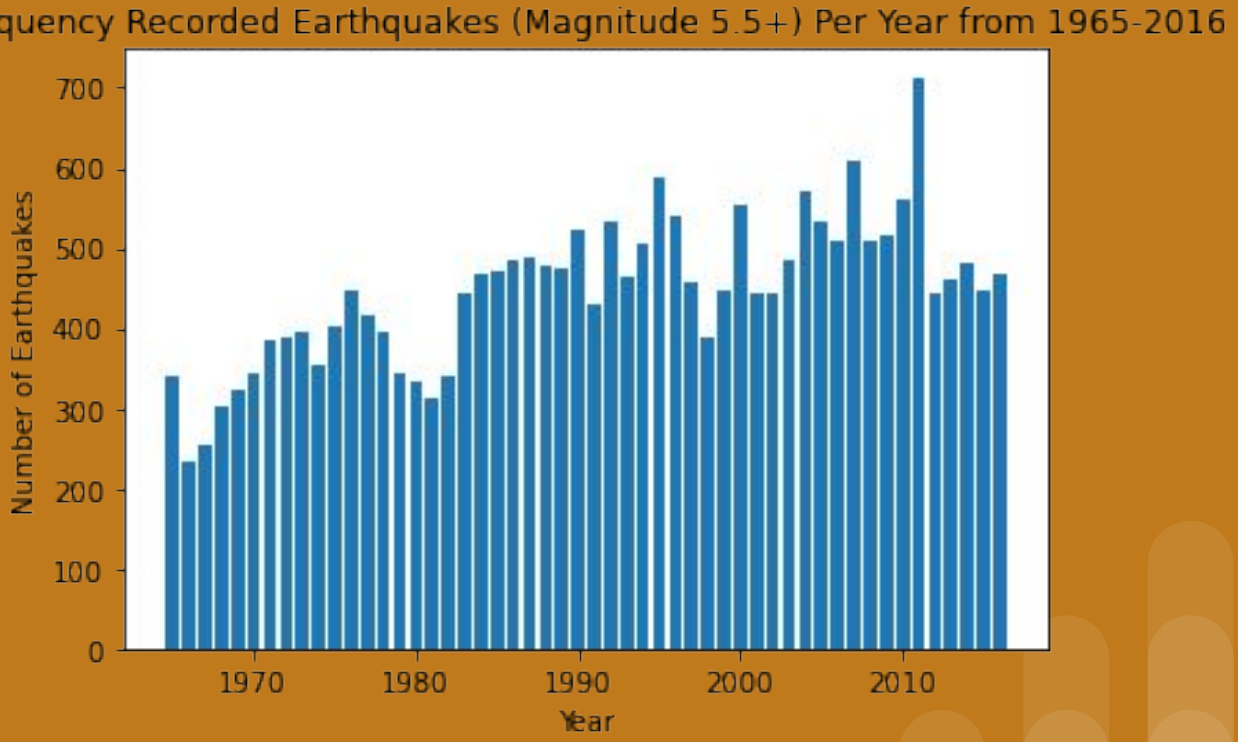


This scatter plot and line graph demonstrates the average earthquake magnitude over the years. From looking, earthquakes from 1965-1972 had a greater intensity than any of the other years. This may have been because of errors in extrapolation of data due to inaccuracies in measuring instruments back then. The graphs show that from 1973 to 2016, the average magnitude of earthquakes have been in an upward trend. Another inference could be that seismic instruments have been improving throughout the years and thus more earthquakes are recorded, even though there might have been relatively the same amount of earthquakes as 1900, for example.





This bar graph demonstrates an upward trend in the frequency of earthquakes (magnitude 5.5+) that occurs per year. This may indicate that an increase in human activities such as mining, oil welling, and groundwater extraction may have been a probable cause of the upward trend. However, it still may have been an improvement in seismic instruments and research that have shown an increase in recorded earthquakes.



Conclusion

I would definitely dive into other programming languages too and not just Python. I would research more into ways to manipulate and display data. I would also group the longitude and latitude coordinates to specific areas to better picture the frequency of earthquakes in areas such as Japan or Mexico.

Some things I couldn't quite pull off is overlaying the world map over the coordinate scatter plot. Learning from scratch, much of the time was spent experimenting with different commands and solving through trial and error.

I had a fairly fun experience learning about python's libraries pandas and matplotlib. My takeaway is that data can be manipulated and utilized in many different ways using simple commands with powerful, integrated software.