Final Project Report

Team : Venkata Dattatreya Reddy Allu, Aditya Gattu, Surendra Sai Kumar Mudipalli

1. Overview and goals of your project.

The primary objective of our project was to perform an in-depth analysis of earthquake data obtained from the United States Geological Survey (USGS) and create insightful visualizations to understand earthquake patterns and characteristics better. The project aimed to explore various attributes related to earthquakes, such as depth, magnitude, rms, horizontal error, and depth error, and study their relationships and impact on the overall earthquake occurrences and measurement accuracy.

By employing various visualization techniques, we sought to identify patterns and trends that could provide valuable insights into the distribution and frequency of earthquakes, their relationship with tectonic plate boundaries, and the factors influencing the accuracy and precision of earthquake measurements. Additionally, we aimed to investigate regional variations in earthquake activity, such as the frequency and magnitude of earthquakes in different parts of the world. By understanding these aspects, our project sought to contribute to the overall understanding of seismic events and their underlying causes, thereby aiding in the development of future research and mitigation strategies.

Furthermore, the project aimed to build our data visualization, analysis, and interpretation skills by using various software tools such as QGIS, Paraview, and Python libraries. Through this process, we hoped to gain valuable experience working with complex datasets, learning how to communicate our findings and insights effectively, and developing a deeper understanding of earthquake activity's underlying concepts and processes.

2. Background and related work. What books, papers, and websites did you learn from?

To gain a comprehensive understanding of the concepts and techniques related to earthquake analysis and visualization, we consulted a variety of sources, including research papers, books, and websites. Some of these sources include

USGS official website: The United States Geological Survey (USGS) website served as an essential resource for obtaining the earthquake dataset used in our project. The website also provided information about earthquake monitoring, reporting, and data interpretation, which proved invaluable in understanding the limitations and characteristics of the dataset.

Paraview and QGIS websites: The official websites of Paraview and QGIS provided us with tutorials and documentation on how to use these software tools for data visualization and mapping. These resources enabled us to create insightful visualizations and effectively explore different aspects of the earthquake dataset.

Online tutorials on data visualization tools: We learned how to use various Python libraries such as Matplotlib, Seaborn, and Pandas for data visualization and analysis by following online tutorials and guides. These resources allowed us to create a wide range of visualizations, including histograms, bar charts, heatmaps, and scatter plots, that helped us analyze and interpret the earthquake data. By consulting these resources, we were able to develop a solid foundation of knowledge and skills that enabled us to effectively analyze and visualize the earthquake dataset, identify patterns and trends, and draw meaningful conclusions about the factors influencing earthquake activity and measurement accuracy.

Envisat: We have used images of Earth taken by the Envisat, an Earth observation satellite operated by the European Space Agency (ESA) from 2002 to 2012.

Tools and Techniques: Some of the key tools and techniques used in earthquake data analysis include statistical modeling, machine learning, data visualization, and geographic information systems (GIS).

By consulting these resources, we were able to develop a solid foundation of knowledge and skills that enabled us to effectively analyze and visualize the earthquake dataset, identify patterns and trends, and draw meaningful conclusions about the factors influencing earthquake activity and measurement accuracy.

3. **Provide a description of your project. What data did you use? What questions did you answer? Describe any new questions that arose throughout the project.**

Our project aimed to visualize and analyze earthquake data to understand better the patterns, trends, and relationships between different earthquake-related attributes. We utilized data obtained from the USGS in CSV format, which contained various attributes related to specific earthquake events. The primary attributes we focused on for our visualization included depth, magnitude, rms, horizontal error, and depth error.

We used QGIS software to create the foundational layout of the world map and subsequently imported the data into the preview, applying a table-to-points filter. This enabled us to visualize the earthquakes based on their magnitudes and locations, using longitude and latitude for the x and y-axis and magnitude for the z-axis.

Throughout the project, we created multiple visualizations, including animations, histograms, bar charts, heat maps, and scatter plots to analyze the earthquake data. We sought to answer questions such as why earthquakes occur, why they tend to occur in specific regions, and how factors like depth and magnitude are related.

As we progressed through the project, new questions and interesting observations emerged. For instance, we discovered that some recorded events were not earthquakes but had significant impacts and were captured by observing stations, such as mining explosions, quarry blasts, and ice quakes. Additionally, we found that Alaska had the highest number of earthquakes among the countries, which could be attributed to two tectonic plates meeting in the region.

By visualizing and analyzing the data, we were able to gain insights into earthquake patterns, regional differences, and relationships between various attributes. Our project helped us understand the frequency of earthquakes, the distribution of earthquake depths and magnitudes, and the precision and accuracy of measurement errors in the dataset.

Throughout the project, we faced several challenges, such as visualizing the earthquake depths in Paraview and representing the depths within the Earth's surface as a global structure. However, these challenges presented opportunities for further exploration, and we were able to find innovative solutions to visualize and analyze the data more effectively.

4. **Discuss the implementation details of your project.**

We have divided the implementation into multiple phases:

1. Visualization of the data
2. Identification and analysis of insights gained from the visualizations
3. Data analysis and comparison with the visualizations.

The data obtained from the USGS is in CSV format, containing all attributes related to a particular earthquake. Out of all available data, we have selected the main attributes for visualization to be depth, magnitude, rms, horizontal error, and depth error.

The base setup of the visualization is illustrated in Figure 1, which serves as a foundation for our implementation.

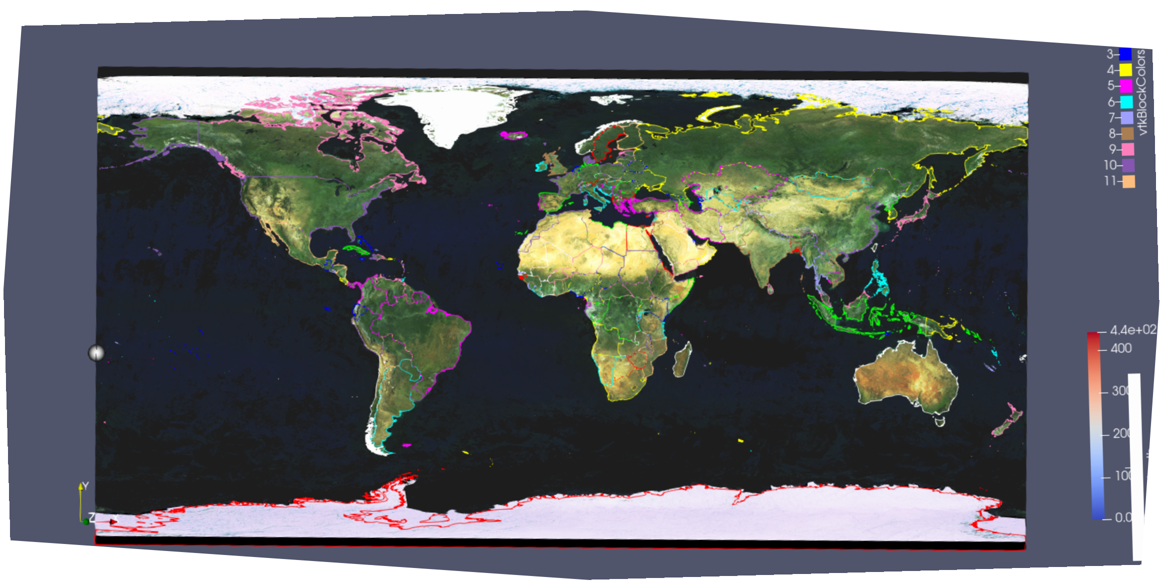

Figure 1: Base setup of our visualization

To create the foundational layout of the world map, we utilized the QGIS software. This tool is designed to generate shapefiles for maps for individual countries and the entire world map. With this software, we were able to produce an outline for all countries on the map to visualize. Furthermore, we were able to accurately plot the earthquakes on the map by referencing a world image from the Envisat directory.

After establishing the foundation, we imported the data into the preview and applied the table to the points filter. We used longitude and latitude for the x- and y-axis, respectively, and for initial representation, we utilized magnitude for the z-axis. This allowed us to visualize the earthquakes based on their magnitudes and locations.

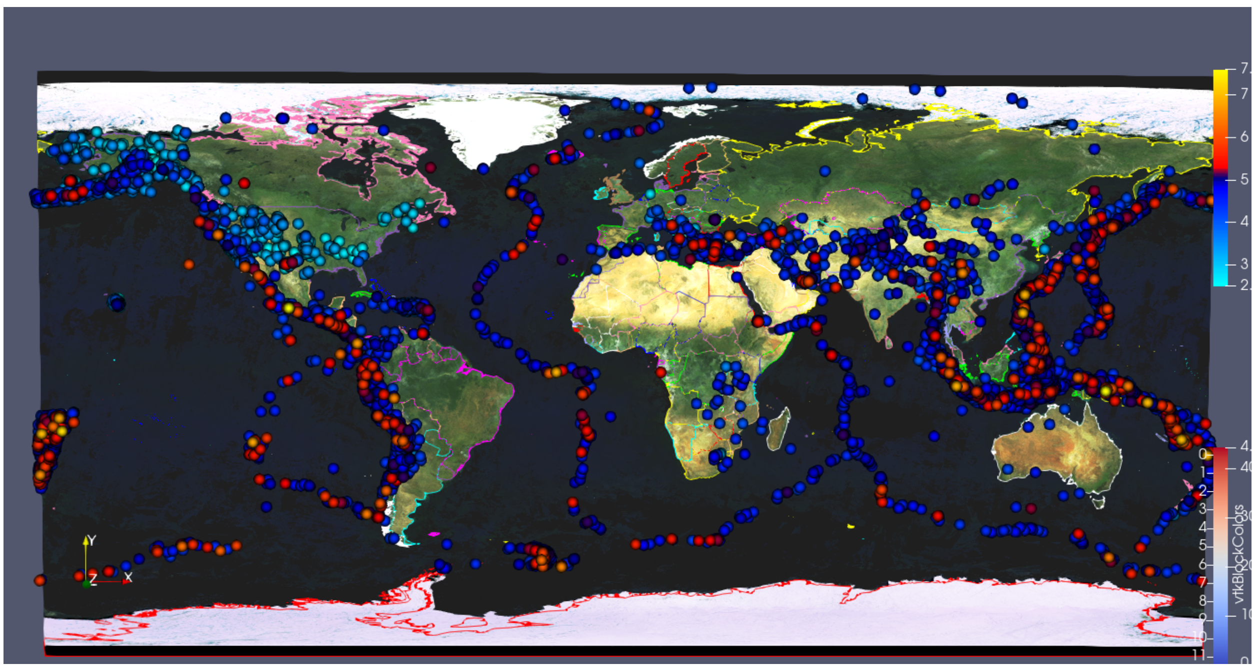

Figure 2: Representing earthquakes based on the magnitude

To aid in identifying the earthquakes based on their magnitudes, we applied a color map with hot and cold colors. The data points were represented using point Gaussian for improved visibility. The color map was set based on the magnitude, as illustrated in Figure 3. We mainly used colormap because we can easily differentiate the earthquakes with high magnitude represented by red(5+), and earthquakes with low magnitude are generally represented by blue, which are in the range of 1 to 3.5.

	Value	R	G	B
1	2.5	0	1	1
2	4.795	0	0	1
3	5.05	0	0	0.501961
4	5.305	1	0	0
5	7.6	1	1	0

Figure 3: color map division for magnitudes

We created an animation based on the magnitude with steps as the earthquake's magnitude.

Visualizing based on the depth of the earthquake:

Visualizing earthquakes based on depth in Paraview has presented a challenging task. To overcome this challenge, we researched general techniques and visualized the depths based on the latitude and longitude of the earthquake. We utilized a rainbow color scheme to visualize the earthquakes based on their locations.

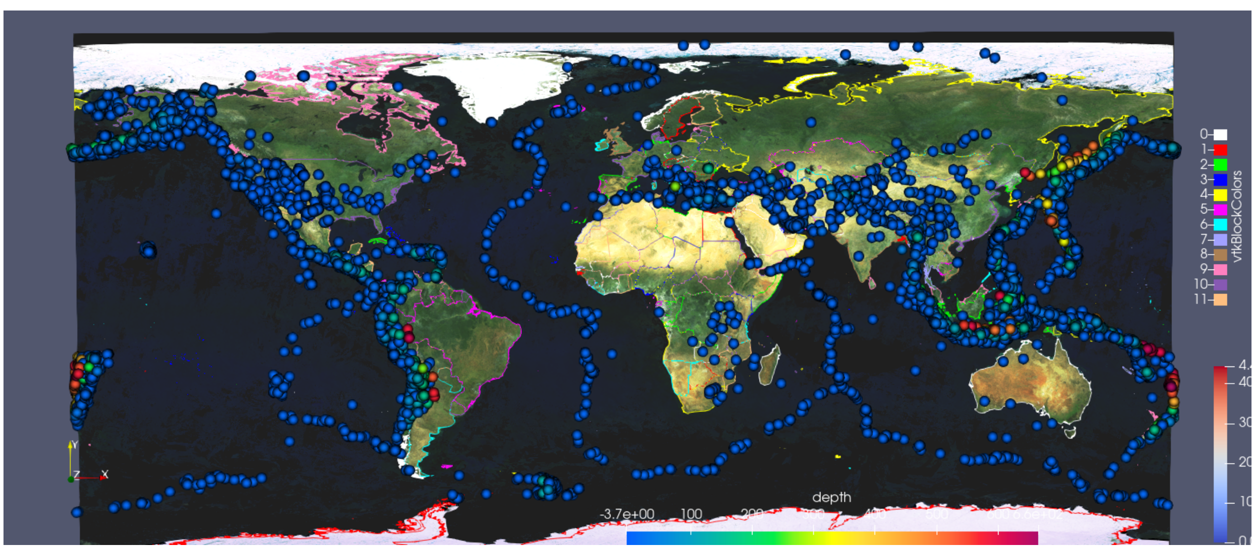

Figure 4: Representing earthquakes based on depth

Through analyzing the depth visualization, we have determined that a majority of earthquakes occurring at significant depths are concentrated in the Pacific Ocean near New Zealand.

To further enhance the visualization of earthquakes, we attempted to use a clip filter to isolate earthquakes that occurred within the United States. Specifically, we sought to determine the number of earthquakes that took place at significant depths between May 2022 and December 2022. Additionally, we aimed to represent the depths within the Earth's surface as a globe structure. However, this proved to be a challenging task that required a significant amount of effort. As such, we are currently representing the data in a flat source format.

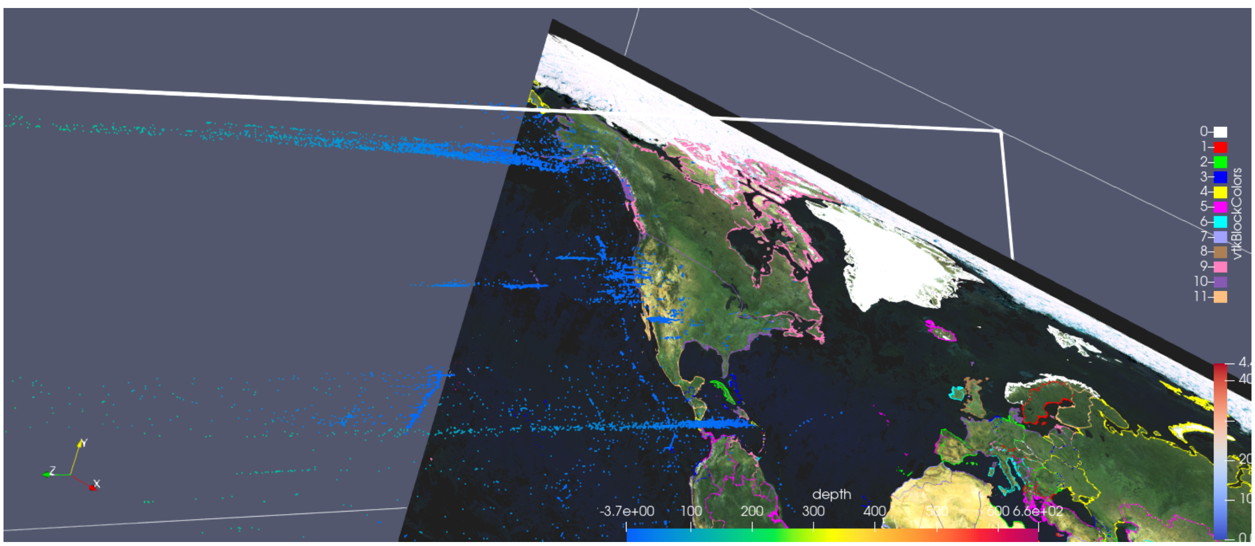

Figure 5: 3D view of the depth of earthquakes

Displayed below is a visualization of the entire world based on depths.

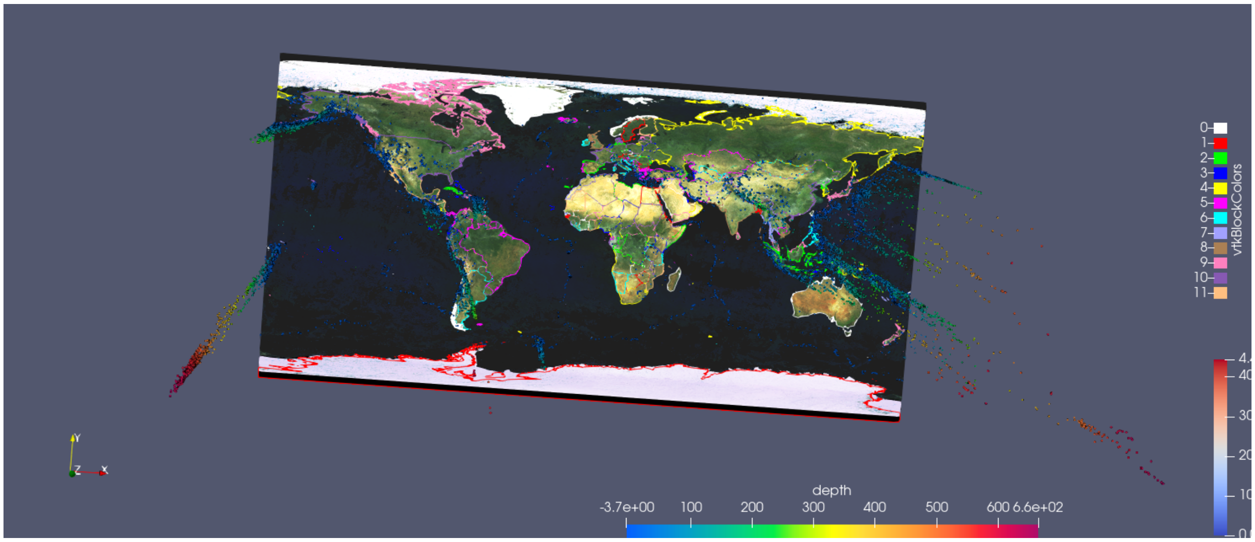

Figure 6: 3D view of earthquakes for the whole world.

To visualize the relationship between depth and magnitude more effectively, we combined depth visualization with an array of magnitudes to color map. By doing so, we are able to observe both depth and magnitude simultaneously.

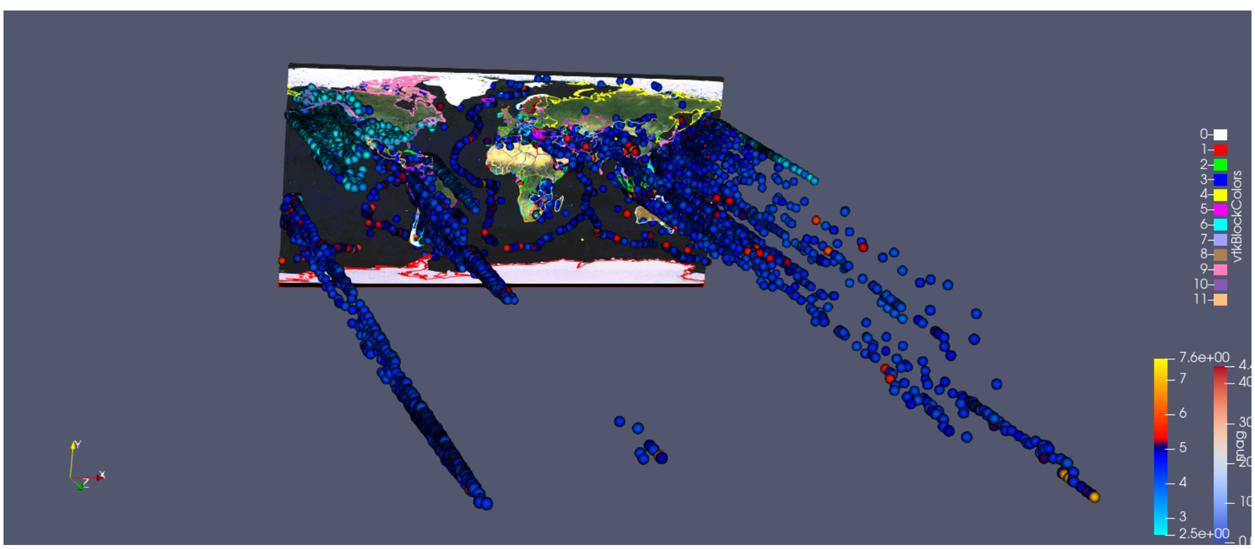

Figure 7: Representing both depth and magnitude simultaneously.

These visualizations prompted us to question why earthquakes occur and why they tend to occur in specific regions. To investigate this further, we conducted research on earthquakes and consulted relevant sources (cited at the end of the report). Our research revealed that earthquakes are most likely to occur where tectonic plates meet. We utilized a point Gaussian blur type to visualize better these areas where tectonic plates meet. This allowed us to observe these regions more clearly and identify where the tectonic plates are located.

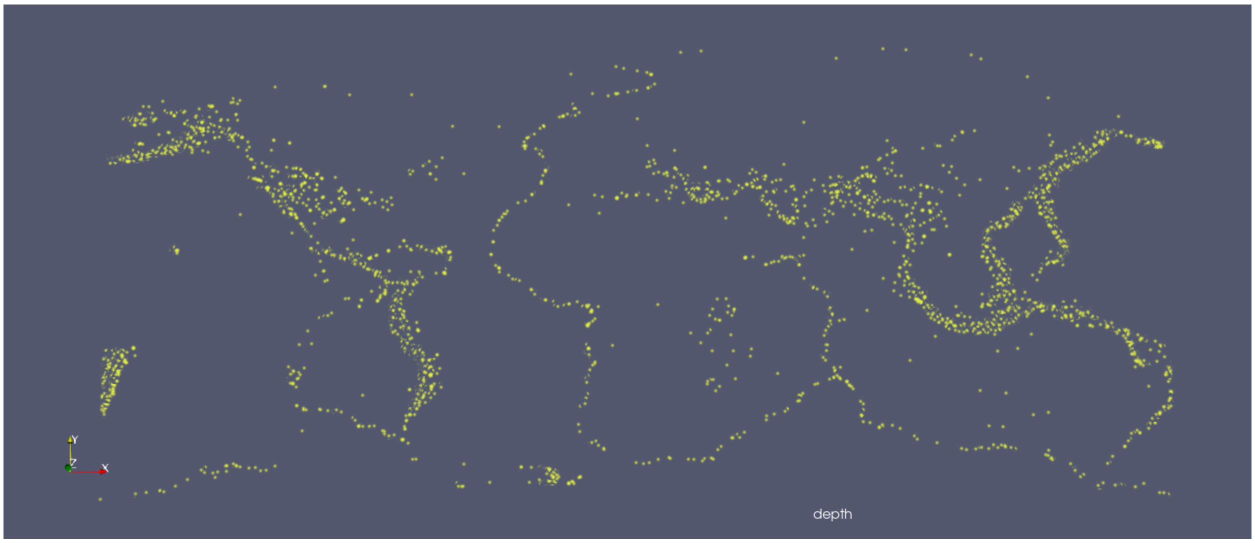

Figure 8: Tectonic Layout of Earth

Comparing the tectonic plates to an actual map of the world provides us with a different perspective of the Earth's surface.

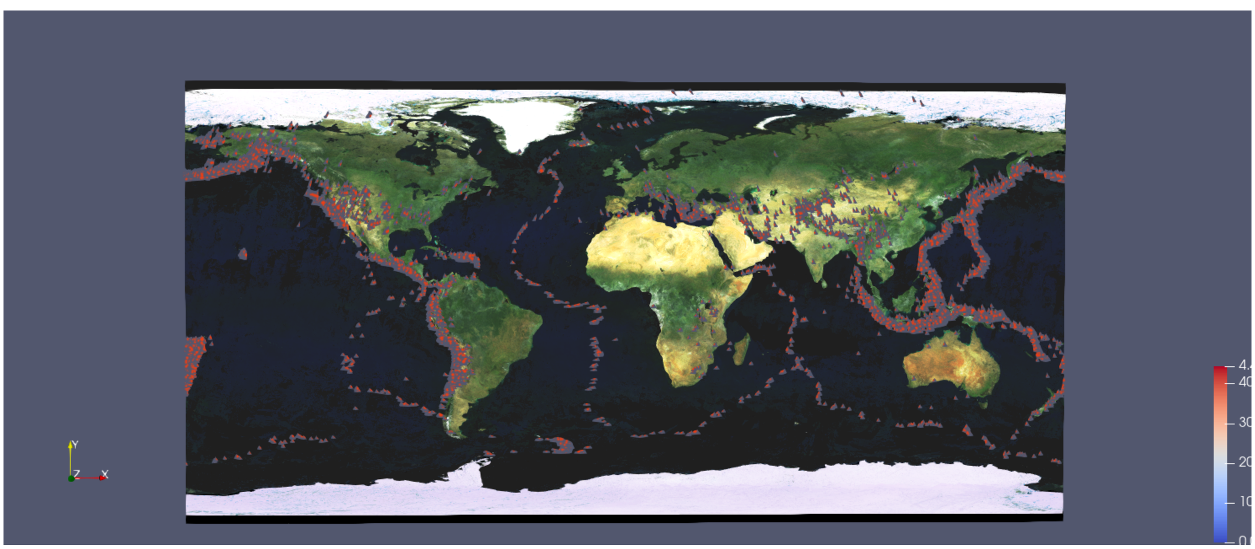

Figure 9: Comparing tectonic layouts with an actual earth map.

Additionally, we attempted to create visualizations using Python to gain further insights into the data we gathered. After conducting our initial analysis, we discovered that earthquakes with magnitudes ranging from 4 to 5 were the most frequently occurring. We presented this information in a bar chart.

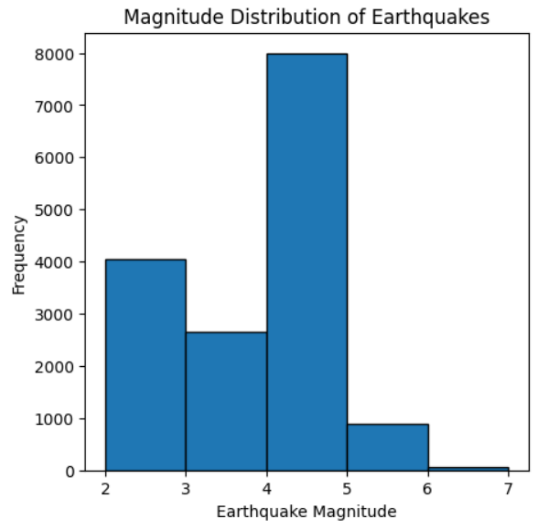

Figure 10: Magnitude distribution of Earthquakes

The above histogram mainly shows the distribution of earthquake magnitudes, which helps us to understand the frequency of different magnitudes. We can observe that earthquakes having a magnitude in the range of [4-5] have the highest frequency of occurring.

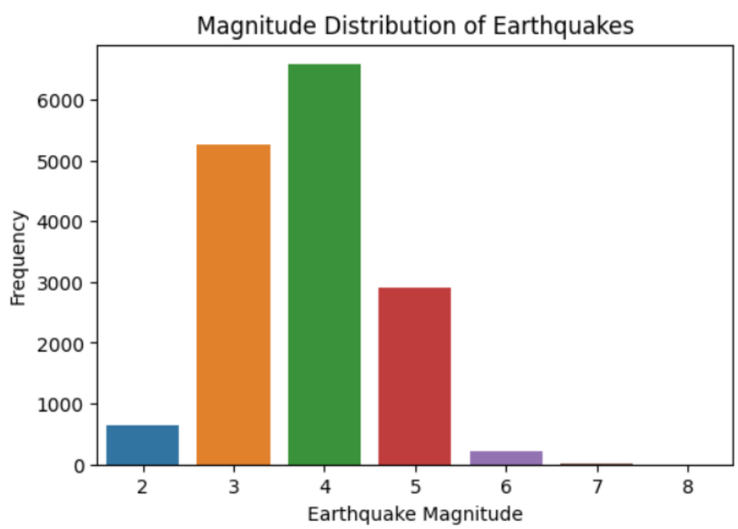

Figure 11: Rounded the earthquake magnitudes to the nearest integer and create a bar chart using Seaborn showing the distribution of earthquake magnitudes.

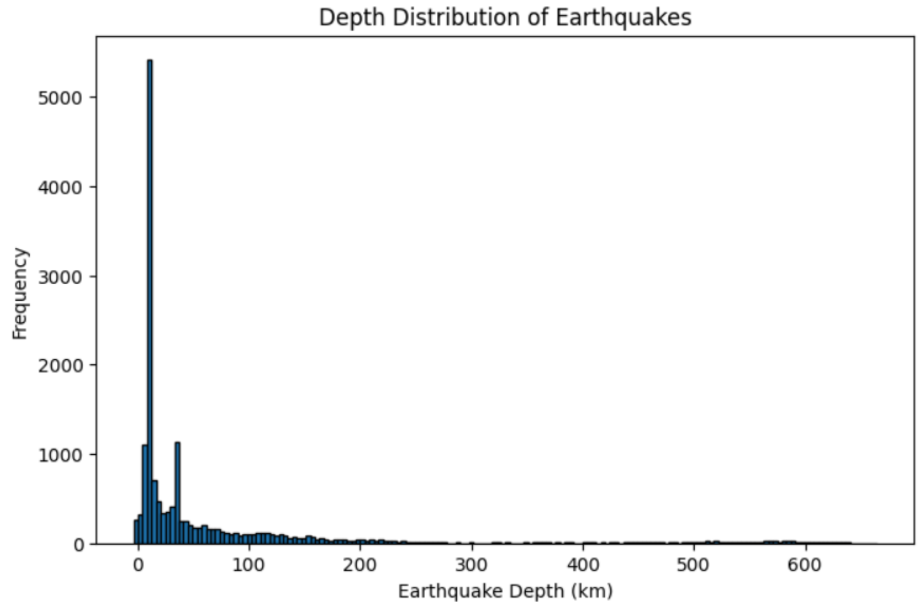

Figure 12: Depth Distribution of Earthquakes

The above histogram provides insights into the distribution of earthquake depths, allowing us to observe the frequency of earthquakes occurring at different depths. By analyzing this histogram, we can identify patterns, trends, or potential anomalies in the dataset, which may have implications for our understanding of the seismic activity, i.e., we can easily say that earthquake depth >100km is very rarely observed.

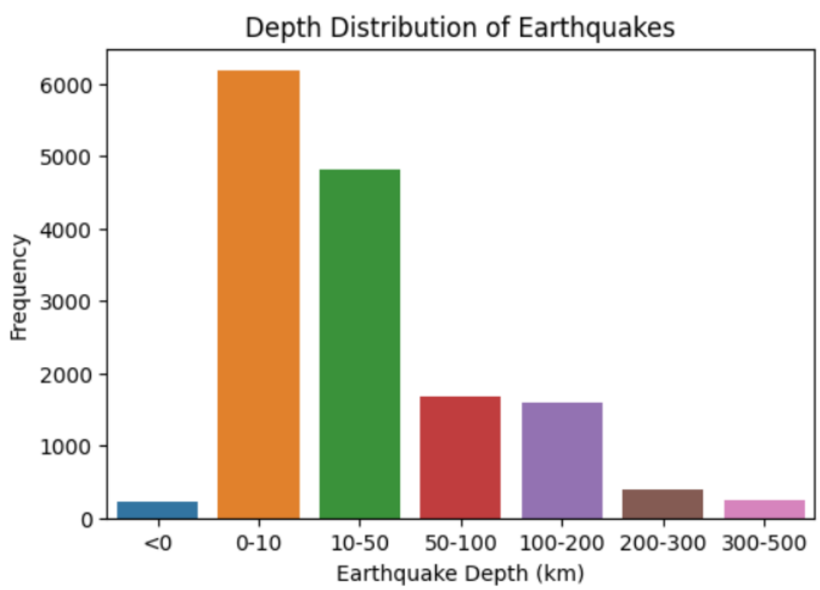

Figure 13: Bar Chart of Earthquake Depth Distribution

In the above visualization of the earthquake depth data in the form of a bar chart, the depths are first categorized into bins, enabling a clearer comparison of the frequency of earthquakes within specific depth ranges. This visualization complements the histogram by providing a different perspective on the data, which may highlight additional patterns or trends that were not apparent in the first figure. Together, these two visualizations offer a comprehensive view of the depth distribution of earthquakes in the dataset, contributing to a deeper understanding of the nature of seismic events.

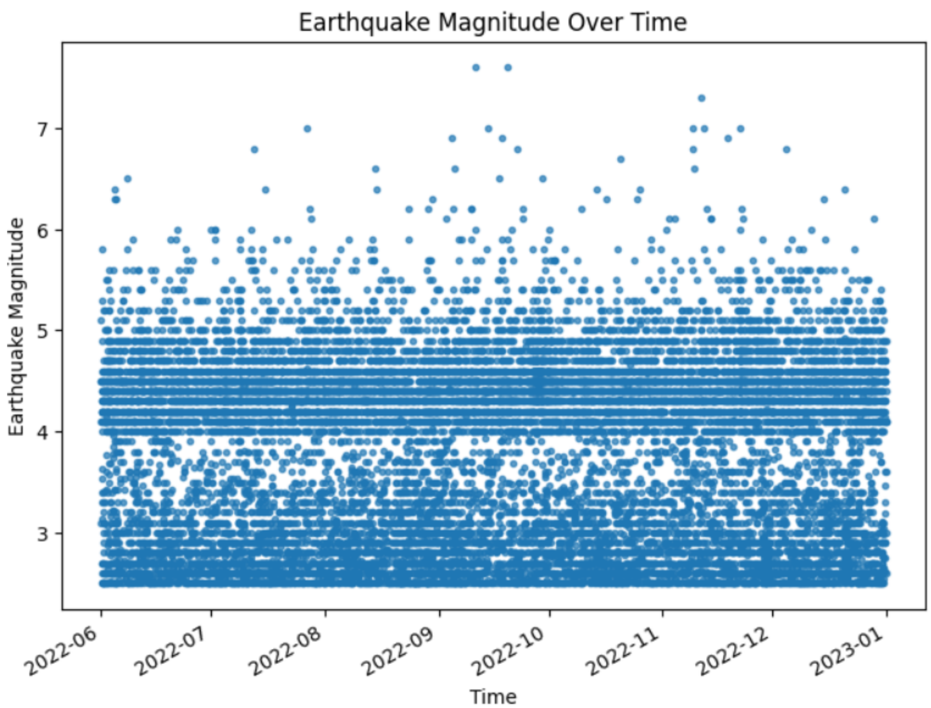

Figure 14: Time series of Earthquake magnitudes

This Fig 14 represents a time-series plot of earthquake magnitudes, showing the change in magnitudes over time. So, In order to obtain the above figure, we converted the 'time' column to a DateTime object, set it as the index, and plotted a time series of earthquake magnitudes.

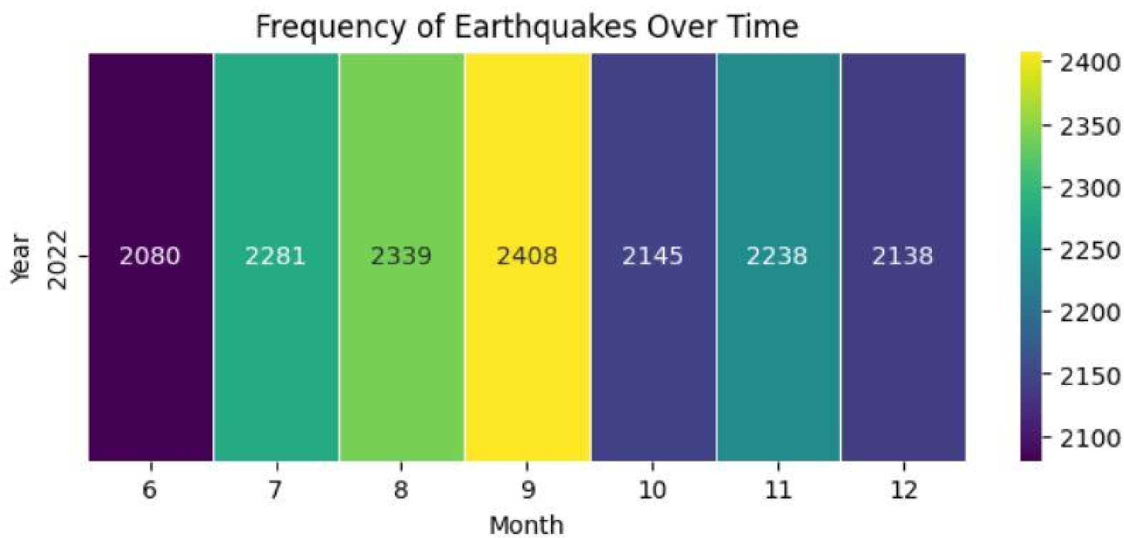

Figure 15: Frequency of Earthquakes Over Time

The above Fig 15 represents Earthquake Frequency by Month heatmap. Here we tried to visualize the frequency of earthquakes per month, revealing any potential seasonal patterns or variation which helps in understanding the distribution of earthquake occurrences throughout the year, i.e., Here, since we have taken the past six months' dataset, we were able to visualize the recurring monthly trends mostly we see that frequency of earthquakes is approx. 2000 every month.

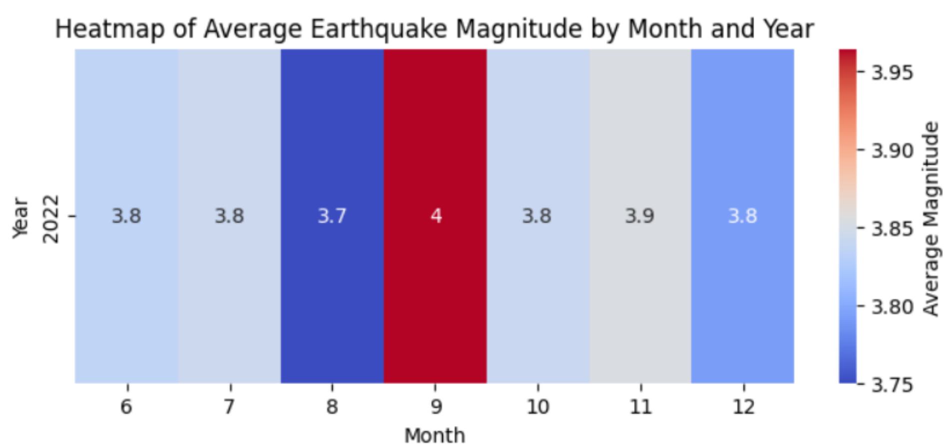

Figure 16: Heatmap of Avg Earthquake Magnitude by Month and Year

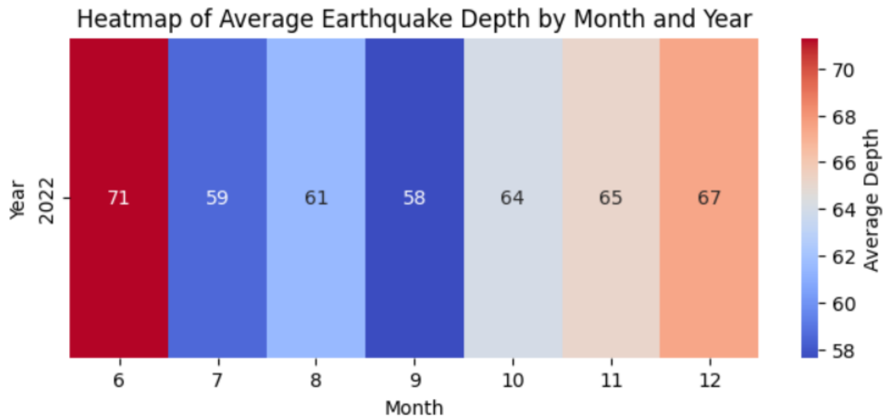

Figure 17: Heatmap of Average Earthquake Depth by Month and Year

The above Fig 16 mainly represents the heatmap of the average earthquake magnitude over time, and Fig 17 depicts a heatmap of the average earthquake depth over time, grouped by month and year, which highlights trends, patterns, or potential seasonal variations in earthquake magnitudes and depth over time respectively, offering a more granular view of the temporal behavior of seismic events. So, we can easily say that avg magnitude of earthquakes occurring is near 4, and the depth is in the range of (65-70 km).

	region	mean_magnitude	frequency	mean_depth
155	Unknown	4.339916	3344	95.636640
2	Alaska	3.015080	2931	47.873664
62	Indonesia	4.504821	892	74.603869
115	Puerto Rico	2.977153	850	23.761539
55	Hawaii	2.862064	533	18.668767
..	...	...	...	...
83	Mississippi	2.520000	1	11.760000
79	Martinique	4.500000	1	62.227000
1	Alabama	2.660000	1	13.140000
77	Malawi	4.000000	1	10.000000
78	Malta	4.500000	1	10.000000

[156 rows x 4 columns]

Figure 18: Regional analysis of the earthquake dataset

The above Fig 18 represents the regional analysis of the earthquake dataset where we tried to extract the region from the 'place' attribute, grouping the data by region, and calculating the mean magnitude, mean depth, and frequency (count) of earthquakes in each region.

Thus, results provide insights into regional differences in earthquake activity, highlighting variations in the average magnitude, depth, and frequency of earthquakes across different regions.

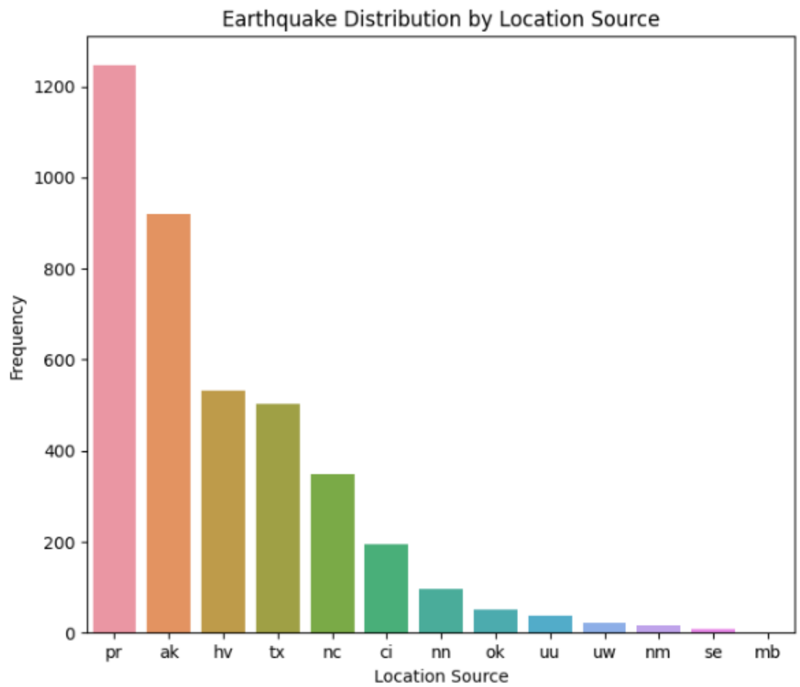

Figure 19: Regional analysis by Location source of the earthquake dataset

In the above Fig 19, we tried to analyze Location source analysis where we can examine the distribution of earthquakes by the source (locationSource) attribute. This can help you understand potential biases in the data collection process or identify differences between various reporting organizations and calculates the frequency of earthquakes for each locationSource here are pr, ak, hv, tx, nc, ok, uw, nm, se, mb

The location sources above are the list of abbreviations that represent U.S. states or territories and seismic networks. Here's an interpretation of the abbreviations:

Pr: Puerto Rico
Ak: Alaska
Hv: Hawaii
Tx: Texas
Nc: North Carolina
Ok: Oklahoma
Uw: University of Washington (Pacific Northwest Seismic Network)

Nm: New Mexico
Se: Southeastern U.S. Seismic Network
Mb: Montana Bureau of Mines and Geology (Earthquake Studies Office)

Thus, these geographical locations or seismic networks where earthquake occurrences have been recorded or monitored.

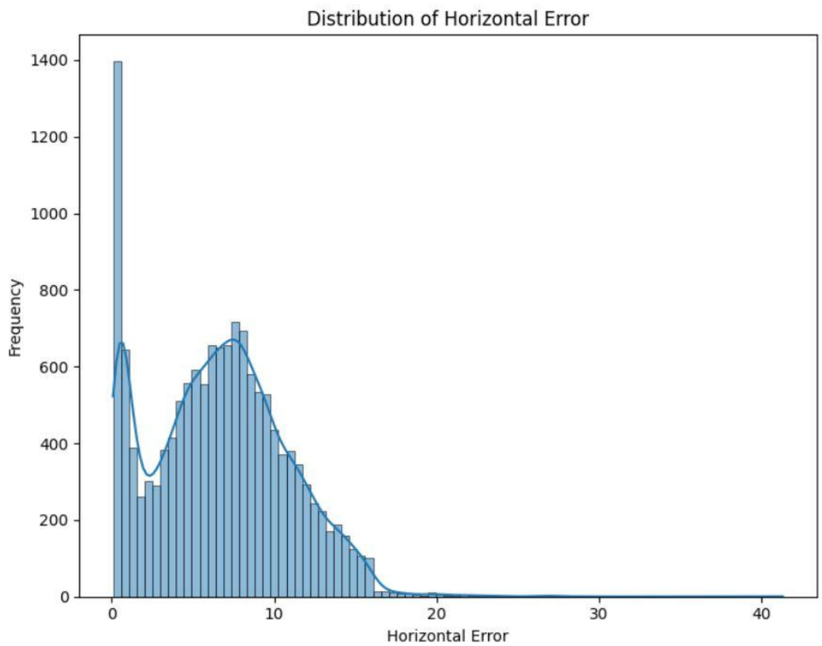

Figure 20: Horizontal Error Distributions in Earthquake Dataset

The above Fig 20 Graph represents the distributions of measurement errors for horizontal positions of earthquake events. It mainly illustrates the distribution of errors in determining the horizontal location (latitude and longitude) of earthquake events. A lower horizontal error value indicates a more accurate estimation of the event's location in the horizontal plane, which helps to understand the accuracy and precision of horizontal position measurements in the dataset.

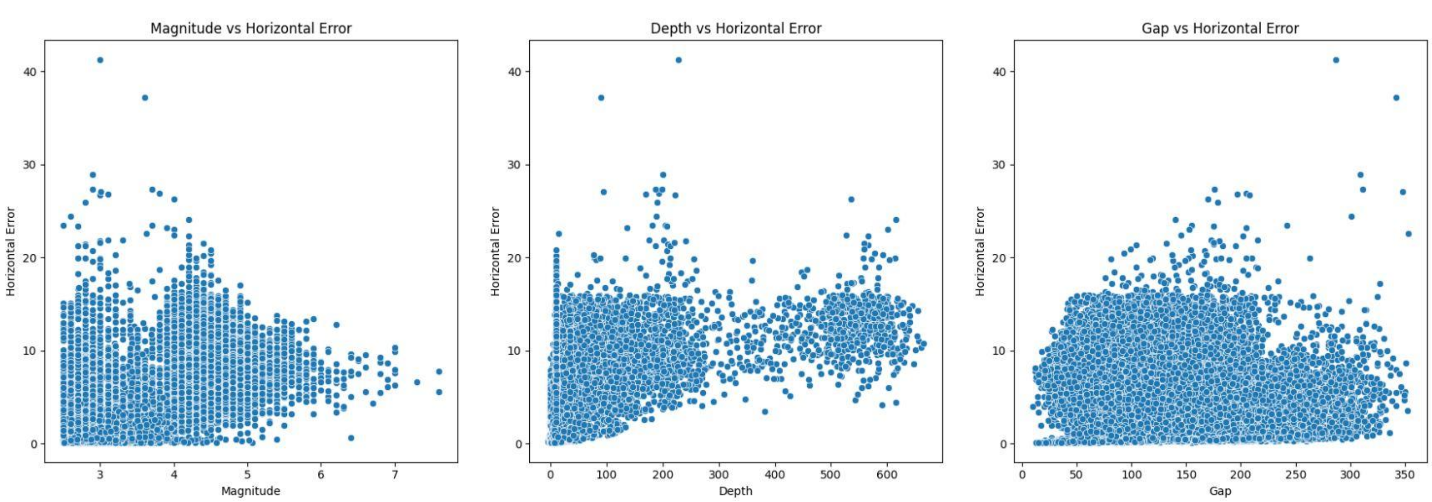

Fig 21: Horizontal Error Relationships: Magnitude(left), Depth(center) & Gap(right)

Horizontal Error Relationships:

- **Magnitude vs. Horizontal Error:** This shows the relationship between earthquake magnitude and horizontal error, representing the uncertainty in the earthquake's horizontal location (latitude and longitude).
- **Depth vs. Horizontal Error:** Displays the relationship between the depth of the earthquake beneath the Earth's surface and horizontal error.
- **Gap vs. Horizontal Error:** Demonstrates the relationship between the gap (the largest angle between seismic stations observing the earthquake) and the horizontal error.

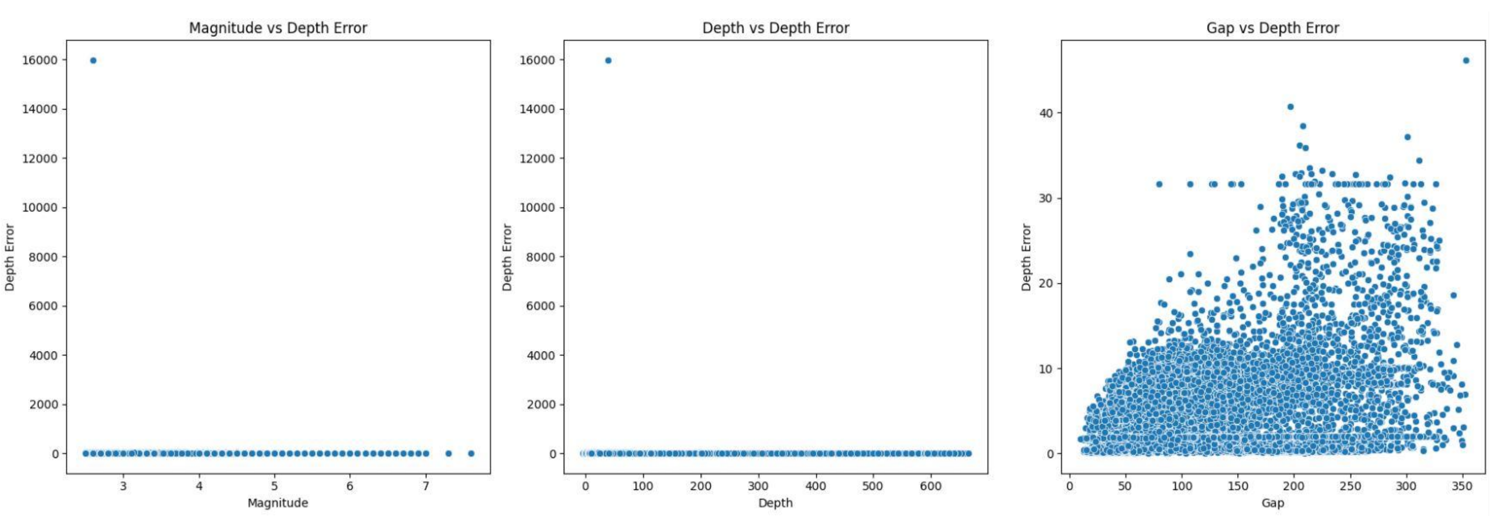

Fig 22: Depth Error Relationships: Magnitude(left), Depth(center), and Gap(right)

Depth Error Relationships:

- **Magnitude vs. Depth Error:** This shows the relationship between earthquake magnitude and depth error, which represents the uncertainty in estimating the depth of the earthquake.
- **Depth vs. Depth Error:** Displays the relationship between the depth of the earthquake and its depth error.
- **Gap vs. Depth Error:** Demonstrates the relationship between the gap and the depth error.

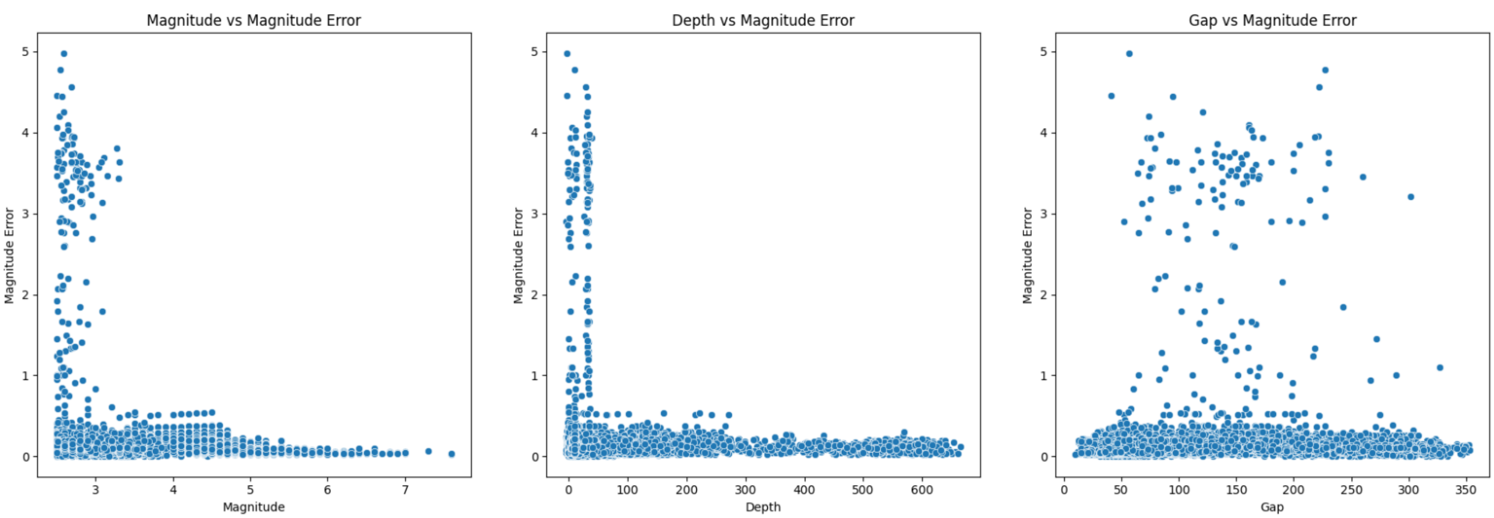

Fig 23: Magnitude Error Relationships: Magnitude(left), Depth(center), and Gap(right)

Magnitude Error Relationships:

- **Magnitude vs. Magnitude Error:** This shows the relationship between earthquake magnitude and magnitude error, which represents the uncertainty in estimating the earthquake's magnitude.
- **Depth vs. Magnitude Error:** Displays the relationship between the depth of the earthquake and its magnitude error**.**
- **Gap vs. Magnitude Error:** Demonstrates the relationship between the gap and the magnitude error.

Thus, these scatter plots help to understand the correlations between different errors and the key parameters of earthquake events, giving various insights into the precision and accuracy of the measurements in the dataset.

Additionally, we found that Alaska has the highest number of earthquakes among the countries. This could be attributed to two tectonic plates meeting in the region. Indonesia also has the second-highest number of earthquakes.

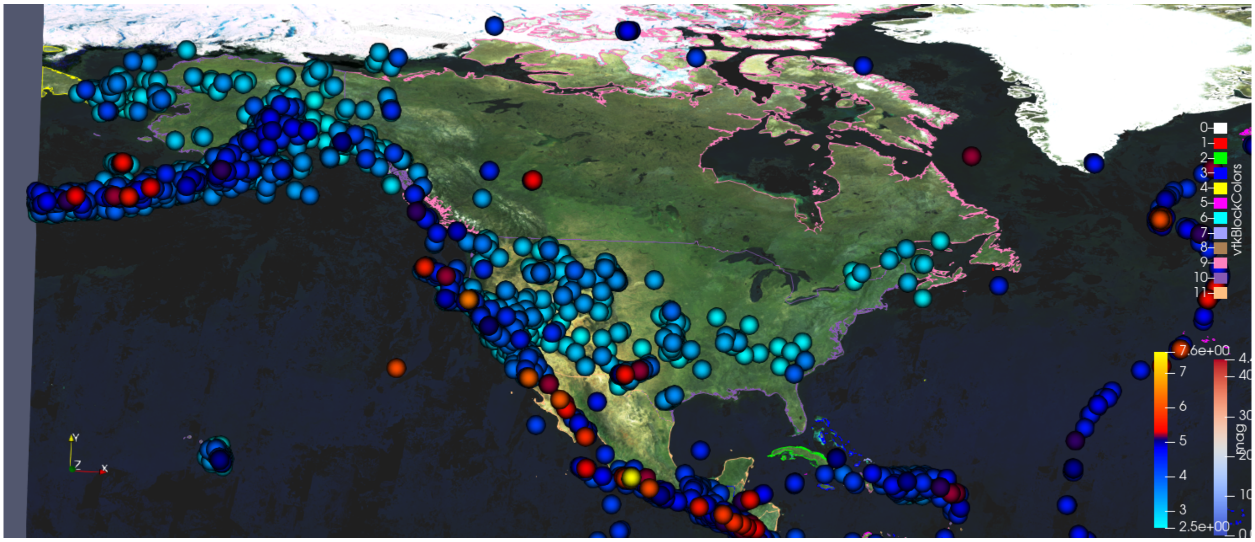

Figure 24: Earthquakes near the USA

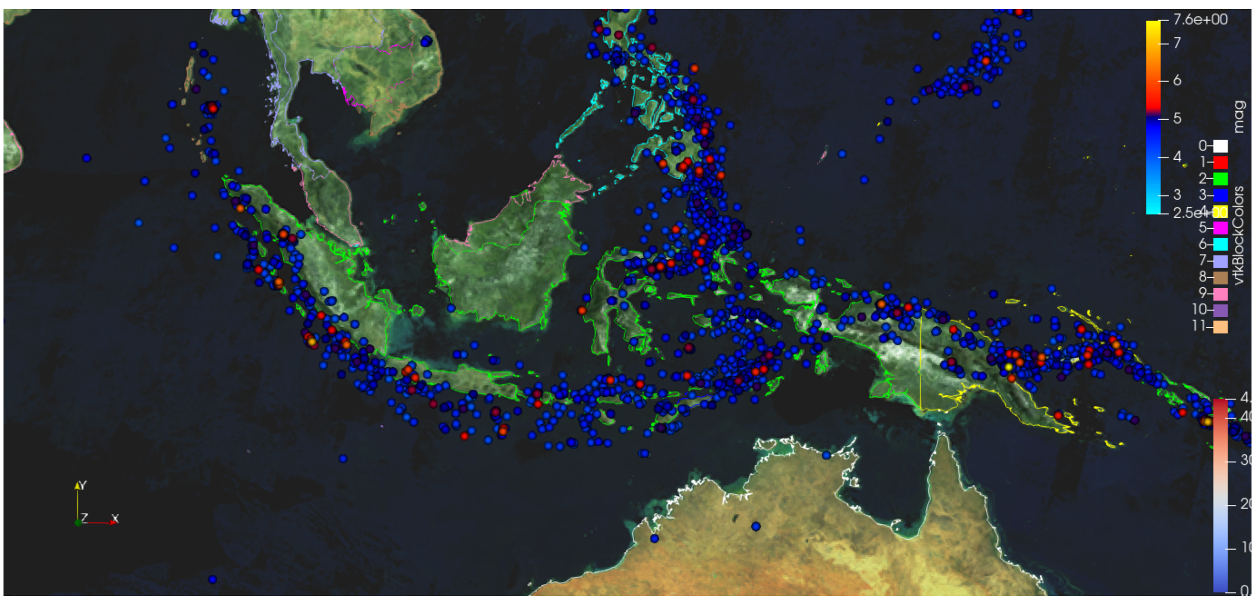

Figure 25: Earthquakes near Indonesia

Interesting observations:

While visualizing the data, we found that some recorded events were not earthquakes but had a significant impact and were captured by Seismometers. These included huge mining explosions, quarry blasts, and, interestingly, two ice quakes.

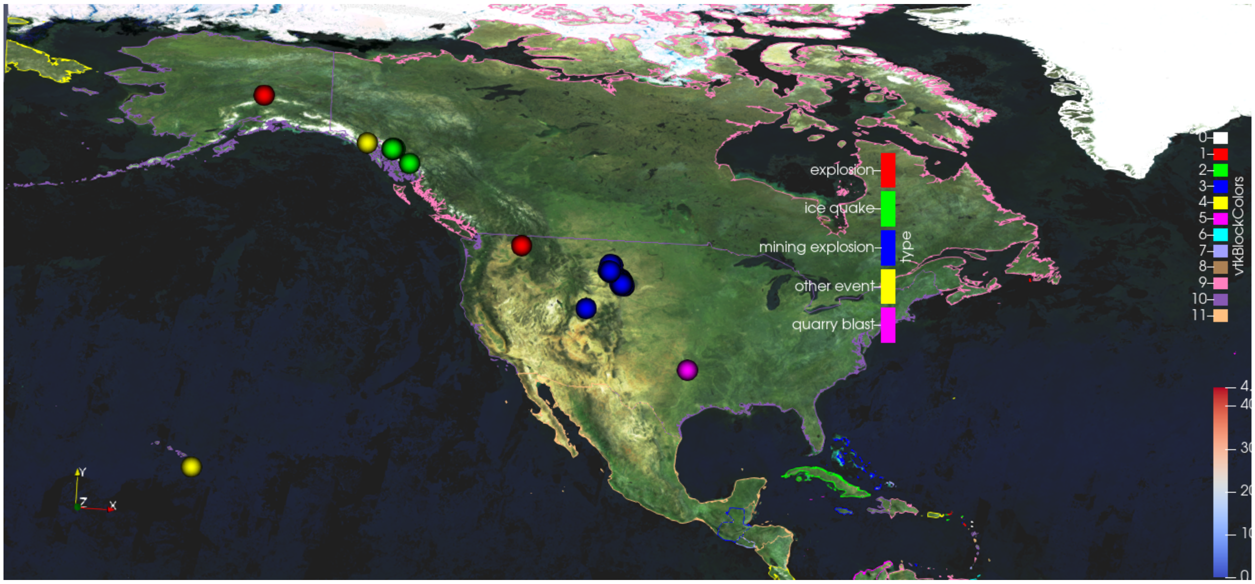

Figure 26: Other recorded events than earthquakes

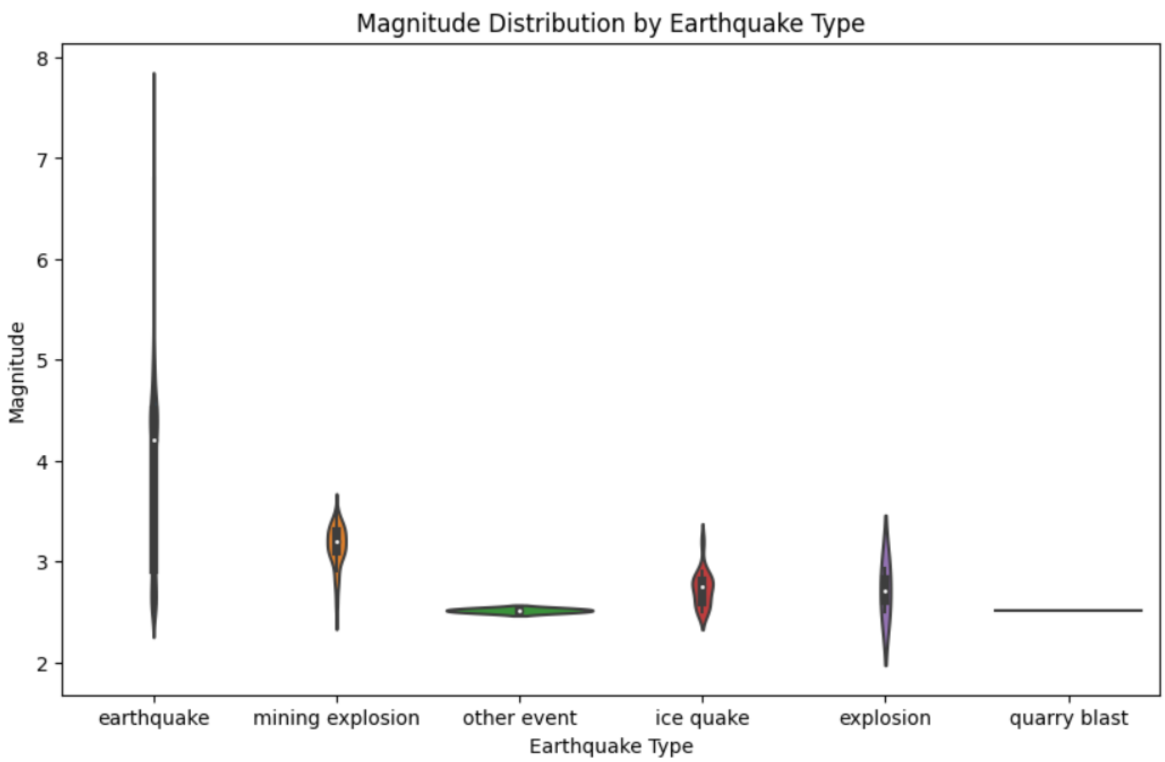

Figure 27: Magnitude Distribution by Earthquake Type

The above violin plot compares the magnitude distribution for each earthquake type, helping you understand how the magnitude /distributions differ between various earthquake types. We have chosen a violin plot for this data visualization because it combines aspects of a box plot and a kernel density plot. It is used to display the distribution of a continuous variable across different categories or groups, which is best suited for Magnitude Distribution by Earthquake Type. The plot displays a series of vertical "violins," each representing one category or group. Each violin is symmetrical and shows the estimated probability density of the data at different values, with the violin's width indicating the data's density.

Based on all the data we had, we created two animations, one for magnitude and another for visualizing earthquakes over time, The latter proved to be a difficult animation as trying to create direct animation is it needs to be visualized via time steps. So we took a small portion of the dataset and visualized it with time transition by manually adding each data point to the animation.

Another animation is about the magnitude; we have created an animation based on the magnitude of the earthquakes. You can find these animations in the submission we made.

5. **Outline what you learned from doing this project.**

   Throughout the course of this project, we gained valuable insights and knowledge in multiple areas. Firstly, we learned how to effectively use data visualization tools such as QGIS, Paraview, and Python to create visually compelling and informative representations of the earthquake data. This process helped us understand the importance of choosing the right visualization methods and techniques to communicate complex data effectively.

   Secondly, we gained a deeper understanding of the factors influencing earthquake occurrences, such as tectonic plate interactions and the geographical distribution of seismic events. We also discovered that earthquake magnitudes and depths have specific patterns and trends, which led us to explore the relationship between these attributes and measurement errors.

   We also learned the importance of data preprocessing, cleaning, and handling large datasets efficiently. This experience has strengthened our ability to work with complex data and identify potential biases and limitations in data collection processes.

   Moreover, the project allowed us to develop our critical thinking while we encountered challenges along the way, such as difficulties in visualizing depth information and representing data in a three-dimensional globe structure.

Overcoming these challenges required to research, experimentation, and perseverance, which ultimately contributed to our growth as data analysts.

Lastly, we learned about the potential applications of our findings in real-world scenarios, such as informing disaster preparedness and response strategies. This aspect of the project highlighted the broader implications of our work and the value of data-driven decision-making in addressing complex societal issues.

In summary, this project provided us with a wealth of learning experiences and opportunities for growth. We expanded our technical skills in data visualization and analysis, deepened our understanding of earthquake phenomena, and developed our critical thinking and problem-solving abilities. These lessons will undoubtedly prove valuable in our future endeavors and contribute to our ongoing development as data analysts and researchers.

6. **Evaluate your project: how successful do you think it was? What are the strengths and weaknesses of your project?**

Evaluating the success of our project, I would argue that it was largely successful in achieving its objectives. We were able to visualize and analyze earthquake data effectively.. This was achieved through the use of various visualization techniques and tools, such as QGIS, Paraview, and Python, which allowed us to represent the data in diverse and informative ways.

One of the strengths of our project lies in the methodical approach we employed, dividing the implementation into multiple phases. This ensured that we first focused on creating accurate and compelling visualizations before moving on to the analysis and interpretation of the results. Our use of multiple visualization tools also allowed us to compare the effectiveness of different methods and gain insights from various perspectives.

Additionally, our project has identified specific patterns and trends in the data, such as the concentration of deep earthquakes in the Pacific Ocean near New Zealand and the high frequency of earthquakes with magnitudes ranging from 4 to 5. These findings have contributed to a deeper understanding of seismic events and can potentially inform disaster preparedness and response strategies.

However, our project also had some weaknesses. One challenge we faced was visualizing earthquake depths in a three-dimensional globe structure, which proved difficult. Despite our efforts, we were unable to achieve this goal within the scope of the project timeline. This limitation may have restricted our ability to understand the spatial distribution of earthquake depths fully.

Another weakness was our reliance on a single data source, the USGS, which may have introduced biases or limitations in our analysis. A more comprehensive evaluation of earthquake data would ideally involve the use of multiple data sources to ensure a more complete and accurate representation of seismic events.

In conclusion, our project successfully achieved its objectives, with some notable strengths and weaknesses. Overall, we were able to visualize and analyze earthquake data effectively, leading to valuable insights that can inform future research and real-world applications.

7. **Provide additional comments useful in evaluating your project.**

In evaluating our project, it is essential to consider both the successes and challenges encountered during the process. One of the major achievements was our ability to create informative and visually appealing visualizations of earthquake data. These visualizations effectively communicated the patterns and trends within the data and facilitated a deeper understanding of earthquake occurrences.

Furthermore, our exploration of the relationship between various attributes and measurement errors provided us with valuable insights into potential biases in the data collection process. This understanding allowed us to consider the limitations of the dataset and the implications for our conclusions.
However, we also faced some challenges during the project. For instance, we encountered difficulties in handling the large dataset and needed to find efficient ways to process and filter the data. Additionally, determining the most suitable visualization techniques to represent specific aspects of the data required experimentation and critical thinking.

Another area for potential improvement is the incorporation of additional data sources. We believe we could find other sources for data to visualize and visualize better in depth. Moreover, if we were able to fit all our work into a single dashboard with a live data feed, it could be used as a comprehensible dashboard to know information about earthquakes.

In conclusion, our project effectively demonstrated the power of data visualization and analysis in understanding complex phenomena such as earthquakes. The insights gained during this project can be applied to understand earthquakes much better and can be used as material for the explanation. The lessons learned can provide a strong foundation for ongoing growth and improvement in the areas of data visualization and analysis.

References:

1. United States Geological Survey (USGS). (n.d.). Earthquake Catalog. Retrieved from https://earthquake.usgs.gov/earthquakes/search/

2. QGIS Development Team. (2021). QGIS Geographic Information System. Open Source Geospatial Foundation Project. https://www.qgis.org/en/site/index.html

3. ParaView. (n.d.). The ParaView Guide. https://www.paraview.org/paraview-guide/

4. European Space Agency. (n.d.). Envisat picture galleries. Retrieved April 19, 2023, from https://www.esa.int/Applications/Observing_the_Earth/Envisat/Envisat_picture_galleries

5. Google. (n.d.). Earthquakes. Google Maps Platform Documentation. Retrieved April 19, 2023, from https://developers.google.com/maps/documentation/javascript/earthquakes

6. Feibush, E. (n.d.). Earthquakes. Retrieved April 19, 2023, from http://www.princeton.edu/~efeibush/earthquakes/

7. ResearchGate. (2012). 3D visualization of earthquake clusters may shed light on how small stress-relieving quakes lead to big ones. [Visualization]. Retrieved April 19, 2023, from https://www.researchgate.net/figure/D-visualization-of-earthquake-clusters-may-shed-light-on-how-small-stress-relieving_fig1_225174720

8. Bird, P. (2003). An updated digital model of plate boundaries. Geochemistry, Geophysics, Geosystems, 4(3), 1027. https://doi.org/10.1029/2001GC000252

9. **Python references :**

- Matplotlib : https://matplotlib.org/stable/users/index.html
- Matplotlib Tutorials : https://matplotlib.org/stable/tutorials/index.html
- Seaborn Official documentation: https://seaborn.pydata.org/
- Seaborn Tutorials: https://seaborn.pydata.org/tutorial.html
- Official Pandas documentation: https://pandas.pydata.org/pandas-docs/stable/index.html
- Pandas User Guide: https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html

Files Included:

1. finalPVSM - A copy of our PVSM file for visualization purposes.
2. query(1)(1) - A copy of our comprehensive dataset.
3. Earthquake_data - Dataset utilized for data analysis.
4. Animation 1 - Visualization of earthquake magnitudes.
5. Animation 2 - Visualization of earthquakes over time.
6. VIZ_Final_Project_Work.ipynb - Python worksheet for the project.
7. Other State Files - (depth.pvsm ,mag.pvsm,rms.pvsm - For Visualization of depth, magnitude and rms for the earthquakes respectively)