Universidad Politécnica
de Madrid

**Escuela Técnica Superior de
Ingenieros Informáticos**

EIT Master School - Master in Digital Innovation

Data Visualization

Global Earthquake Trends and Interactive Analysis
-
Report
Group 11

Álvaro Honrubia Genilloud
José Antonio Ruiz Heredia
Xiya Sun
Joseph Tartivel

Madrid, 16 December 2024

Contents

1 Introduction **1**

2 Dataset **2**
2.0.1 Data description 2

3 Design abstraction **4**
3.1 Earthquake key characteristics change over region and time 5
3.1.1 Problem characterization 5
3.1.2 Data and task abstraction 5
3.1.3 Interaction and visual encoding 6
3.1.4 Algorithmic implementation 8
3.1.5 Results 8
3.2 Earthquake types frequency overtime 12
3.2.1 Problem characterization 12
3.2.2 Data and task abstraction 12
3.2.3 Interaction and visual encoding 13
3.2.4 Algorithmic implementation 15
3.2.5 Results 17
3.3 Distribution of key earthquake characteristics and geographic patterns 19
3.3.1 Problem characterization 19
3.3.2 Data and task abstraction 19
3.3.3 Interaction and visual encoding 20
3.3.4 Algorithmic implementation 20
3.3.5 Results 21
3.4 Coverage of seismic stations across regions and impact on detection and reporting 24
3.4.1 Problem characterization 24
3.4.2 Data and task abstraction 24
3.4.3 Interaction and visual encoding 24
3.4.4 Algorithmic implementation 27
3.4.5 Results 28

4 Instructions to run the App **32**
4.1 System Requirements 32
4.2 Data Requirements 32
4.3 Running the Application Locally 33
4.4 Access Via Web 33

1. Introduction

Earthquakes are one of the most devastating natural disasters, causing significant damage to infrastructure and posing serious risks to human lives. Despite advancements in seismic monitoring, understanding when and where earthquakes are likely to occur remains a major challenge for seismologists. These challenges highlight the importance of leveraging data-driven insights to improve earthquake prediction, risk assessment, and disaster preparedness.

As data scientists, our goal is to assist seismologists and geophysicists by providing an interactive tool that allows them to explore global earthquake patterns through data visualization and analysis. Using the Shiny framework in R, we aim to answer several key questions that can help uncover hidden patterns in seismic activity, such as:

- How do the key earthquake characteristics change across different regions?
- Are there temporal patterns in the types of earthquakes in terms of frequency?
- What is the distribution of key earthquake characteristics (such as magnitude and depth) across different regions, and what geographic patterns can be observed?
- How does the coverage of seismic stations (Gap) of each network vary across different regions, and what impact does this have on earthquake detection and reporting errors?

By answering these questions, we aim to provide valuable insights that can support risk assessment efforts and improve early warning systems for earthquake-prone areas. Our interactive Shiny application enables users to filter data by different variables according to the idiom, allowing for a more customized and dynamic exploration of seismic aspects.

In this project, we will apply various data visualization techniques to analyze earthquake data, with a focus on identifying temporal trends, geographic patterns, and magnitude distributions. These visualizations will help seismologists better understand the behavior of different earthquake types and contribute to better-informed decision-making in seismic risk management.

The application is published at:

https://bilk771.shinyapps.io/group11_final_project/

2. Dataset

To help experts answer the questions outlined in the problem characterization, we will conduct our analysis using global earthquake data from 2023. This dataset is sourced from Kaggle and can be accessed at the following link: Kaggle Dataset - Earthquakes 2023 Global.

This dataset contains valuable information on earthquake occurrences worldwide, including details about their location, magnitude, depth, and type. By leveraging this dataset, we aim to uncover patterns and provide meaningful insights that can assist seismologists in their research.

2.0.1 Data description

In this dataset we can find several variables, the table below provides brief descriptions of each attribute and its type.

Variable	Type	Description
time	quantitative	Timestamp of the earthquake event.
latitude	quantitative	Geographic coordinate specifying the north-south position.
longitude	quantitative	Geographic coordinate specifying the east-west position.
depth	quantitative	Depth of the earthquake in kilometers.
mag	quantitative	Magnitude of the earthquake.
magType	categorical	Type of magnitude measurement.
nst	quantitative	Number of seismic stations that reported the earthquake.
gap	quantitative	The gap between different seismic stations' coverage.
dmin	quantitative	Minimum distance to the earthquake epicenter for the nearest station.
rms	quantitative	Root Mean Square of the earthquake's amplitude spectrum.
net	categorical	Network reporting the earthquake.
id	categorical	Unique identifier for the earthquake event.
updated	categorical	Timestamp indicating when the earthquake information was last updated.
place	categorical	Location description of the earthquake.
type	categorical	Type of seismic event (e.g., earthquake, volcanic eruption).
horizontalError	quantitative	Horizontal error in location determination.

depthError	quantitative	Error in depth determination.
magError	quantitative	Error in magnitude determination.
magNst	quantitative	Number of seismic stations used to calculate the magnitude.
status	categorical	Status of the earthquake event (e.g., reviewed).
locationSource	categorical	Source reporting the earthquake location.
magSource	categorical	Source reporting the earthquake magnitude.

3. Design abstraction

In this section, we follow the four abstraction levels for each visualization idiom. We begin by characterizing the specific problem within the domain, explaining the variables used and defining the tasks. We then describe how the data is encoded and, finally, how it is implemented and the results we obtained. The objective is to use data visualization to provide valuable insights that can support decision-making and further analysis.

The interface is divided into five panels. The first one is the introduction and description of the work and the rest is divided into sections aimed to answer the four questions figured by the experts.

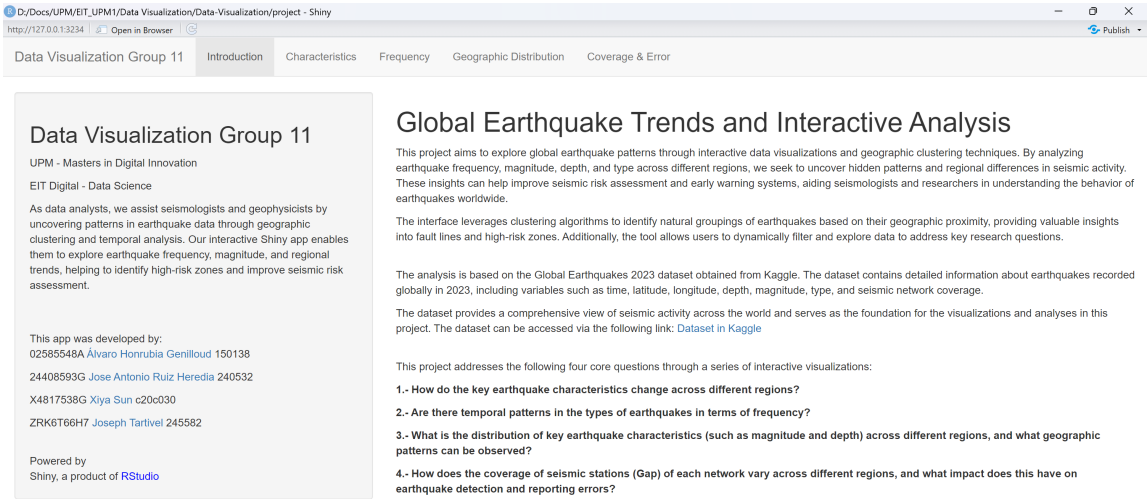


Figure 3.1: Main panel of the interface: Introduction

3.1 Earthquake key characteristics change over region and time

3.1.1 Problem characterization

Through the study of this problem, we search for both spatial and temporal patterns in the main characteristics of the earthquakes in our dataset. Concretely, we focus on magnitude and depth, which consist of the main measures to determine the impact of these catastrophes. The reason for studying them isolated by time or location is to obtain key insights into the evolution, trends, and possible correlations that could lead us to make valuable conclusions.

By implementing this analysis, we facilitate the prediction or estimation of future earthquakes, improving our understanding of their impact in specific locations of the globe or concrete months of the year. For instance, we can conclude patterns in the most relevant earthquakes in terms of magnitude based on the season of the year. Similarly, we can explore whether the depth of earthquakes is influenced by the region and, consequently, by the tectonic plate where they occur.

Therefore, we can summarize our problem in the following main questions:

- Is the depth of earthquakes influenced by geographical location or tectonic plate boundaries?
- Are there specific regions that experience consistently higher-magnitude earthquakes?
- Are there seasonal trends or patterns in the occurrence of high-magnitude earthquakes?

These questions aim to address the spatial and temporal dynamics of earthquakes, their underlying causes, and their potential for prediction and mitigation.

3.1.2 Data and task abstraction

To obtain the respective data abstractions of this question it is essential to identify key attributes. The primary focus is on earthquake events, which are characterized by magnitude, depth, location and timestamp. These attributes serve as the foundation for exploring both spatial and temporal patterns, with the data grouped by geographical regions (continents) and time periods (year-month pairs). Additionally, key metrics such as the average magnitude and depth for specific regions or time intervals are crucial for the study.

For the case of the task, it involves analyzing both spatial and temporal trends. Spatial analysis identifies continents with consistently high-magnitude or deep earthquakes, offering insights into geographical and tectonic influences. Temporal analysis looks for trends over time, such as seasonal variations in earthquake characteristics, to uncover patterns that may help predict future events. Additionally, comparing magnitude and depth across different regions and time periods helps identify correlations and external factors influencing earthquake behavior.

3.1.3 Interaction and visual encoding

In order to effectively analyze both spatial and temporal changes, we grouped the tasks into two different visualizations.

Firstly, we conducted an average analysis of magnitude and depth, visually represented using red and blue colors, respectively. As illustrated in Figure 3.2, the data was grouped either by continent or by year-month to facilitate meaningful comparisons.

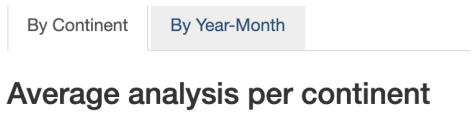
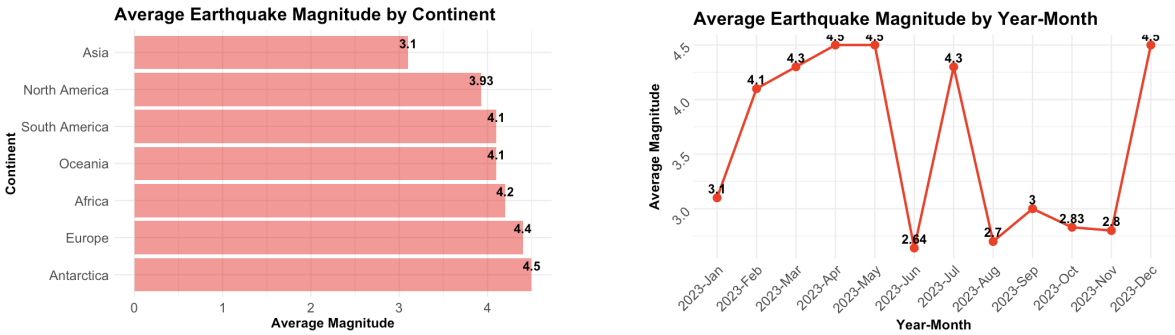


Figure 3.2: Spatial and temporal taps overview.

In the first case, we implemented a bar chart that allows us to present a comparison of the different regions while maintaining a clear, simple, and appealing appearance for the user. In the second case, we selected a line chart where each point corresponds to the average of a year-month within the selected time range. This representation shows a more intuitive and natural way to understand evolution in quantitative data with respect to a time range. Both marks can be visually compared in Figure 3.3.



Bar chart for categories

Line chart for dates

Figure 3.3: Comparison of bar and line charts.

Secondly, we chose to shift the focus by presenting a broader, more distributed view through a heatmap that uses color encoding to provide insights based on the selected characteristic. The map also includes a legend and allows for zooming and panning, enhancing the user's ability to explore the data presented. This visualization showed in Figure 3.4, allows us to contrast our conclusions obtained through average analysis by observing outstanding cases and specific examples that don't fit in our predictions.

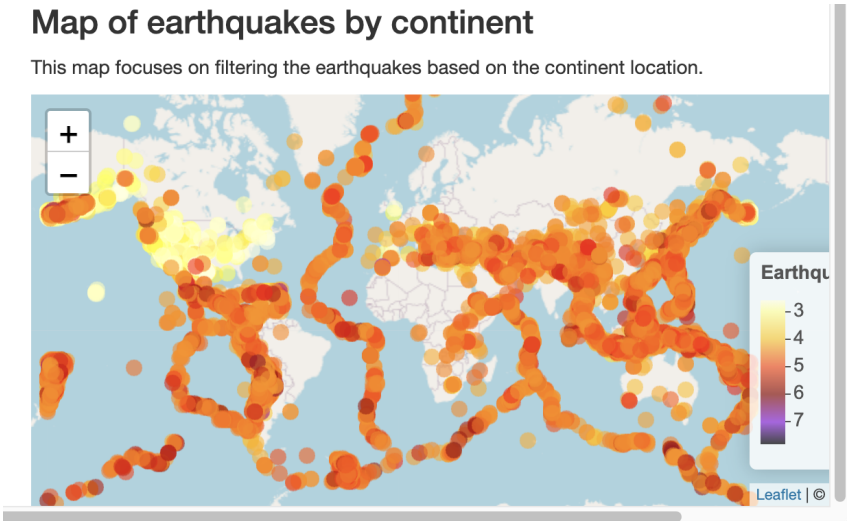


Figure 3.4: Heat map overview.

Additionally, in the sidebar, we provide the user with the necessary filtering tools to analyze these visualizations by changing the scope of the data shown.

- In Figure 3.5, we visualize a measurement selector that allows switching between magnitude and depth.

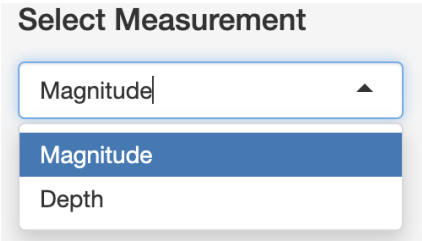


Figure 3.5: Measurement selector overview.

- If the user chooses the spatial view, a checkbox is provided as shown in Figure 3.6. It allows the user to select or deselect the continents required for the analysis.

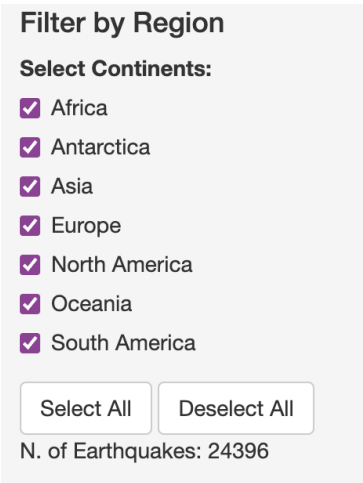


Figure 3.6: Checkbox selector overview.

- For the temporal tab, a time range input as shown in Figure 3.7, enables the user to specify the exact time frame for evaluating changes.

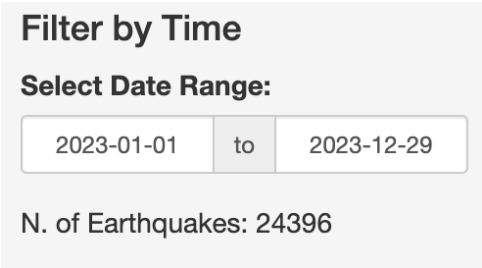


Figure 3.7: Time range selector overview.

3.1.4 Algorithmic implementation

The algorithmic implementation of this analysis employs several key features of R, primarily from the *Shiny*, *dplyr*, *ggplot2* and *leaflet* packages, to facilitate interactive data exploration. The *Shiny* framework provides the necessary infrastructure to enable dynamic filtering of earthquake data based on user inputs such as the selected continets or specified date range. The *observeEvent* function is implemented to trigger changes in the data when the user selects or deselects regions or date ranges, while *reactive* expressions ensure that the displayed data is automatically updated in real-time.

To process the data, we use the *dplyr* package for efficient data manipulation. It allows for grouping the data by continents or year-month and calculating the average earthquake magnitude or depth for each group. For visualization, we employ *ggplot2* to create bar and line charts that display these average values with dynamic labels. In addition to these static visualizations, the *leaflet* package is used to create interactive heat maps that visualize earthquake locations on a map. The map markers are color-coded according to earthquake magnitude or depth, providing a visual representation of how these variables vary spatially.

The algorithmic implementation can be summarized in the Table 3.1.

Library	**Main Usage**
Shiny	Interactive features like dynamic filtering and real-time data updates.
dplyr	Efficient data manipulation for grouping and calculating averages.
ggplot2	Visualization figures such as bar and line charts.
leaflet	Interactive heat maps.

Table 3.1: Summary of Libraries Used.

3.1.5 Results

After completing the analysis and visual implementation, we can now address the key questions outlined during the problem characterization phase. This section focus on answering these questions based on the patterns and trends observed in the data.

By examining the results, we can draw meaningful conclusions and create new knowledge about the impact of geographical location and seasonal variations on earthquakes magnitude and depth.

Is the depth of earthquakes influenced by geographical location or tectonic plate boundaries?

Firstly, we need to observe the data filtered by depth and region to match the specific requirements of this task.

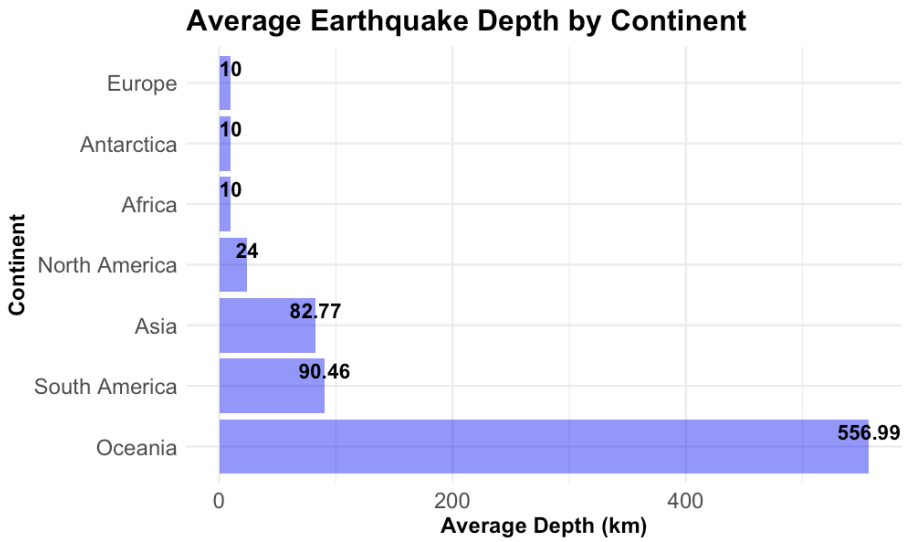


Figure 3.8: Bar chart of average depth by continent.

As observed in Figure 3.8, the average depth of earthquakes in Oceania is notably high, which may be correlated with the presence of the world's deepest oceanic points in that region. However, to better understand the conclusions obtained, we can compare the average estimations with specific individual cases in the heatmap.

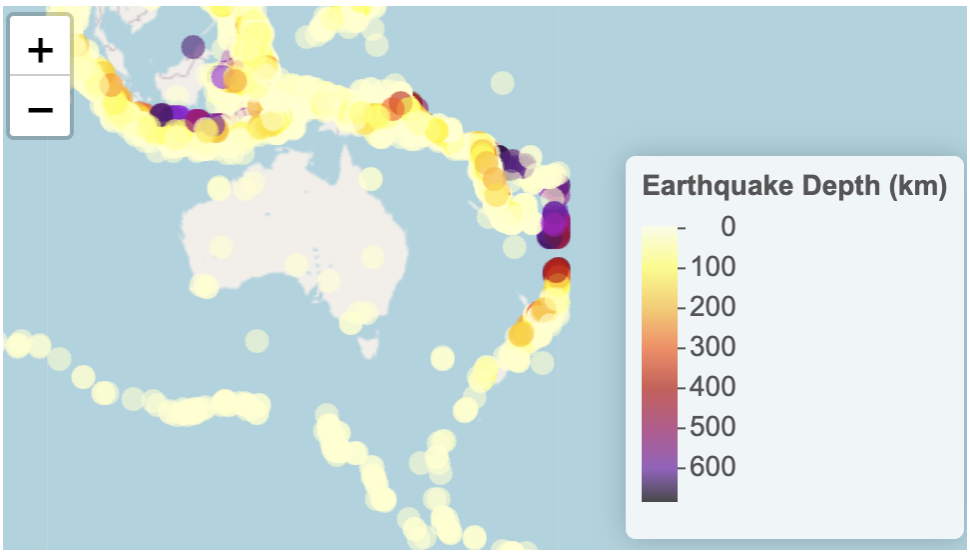


Figure 3.9: Heat map by eathquake depth in Oceania.

Figure 3.9 demonstrates that the average is influenced by specific locations where the depth attribute is significantly higher than in other areas within the same continent. Therefore, it is not a general pattern across the region, but rather correlated with particular locations where the depth is extremely higher than usual cases.

Are there specific regions that experience consistently higher-magnitude earthquakes?

In this case, we need to observe the data filtered by magnitude and grouped by continent to analyze possible trends.

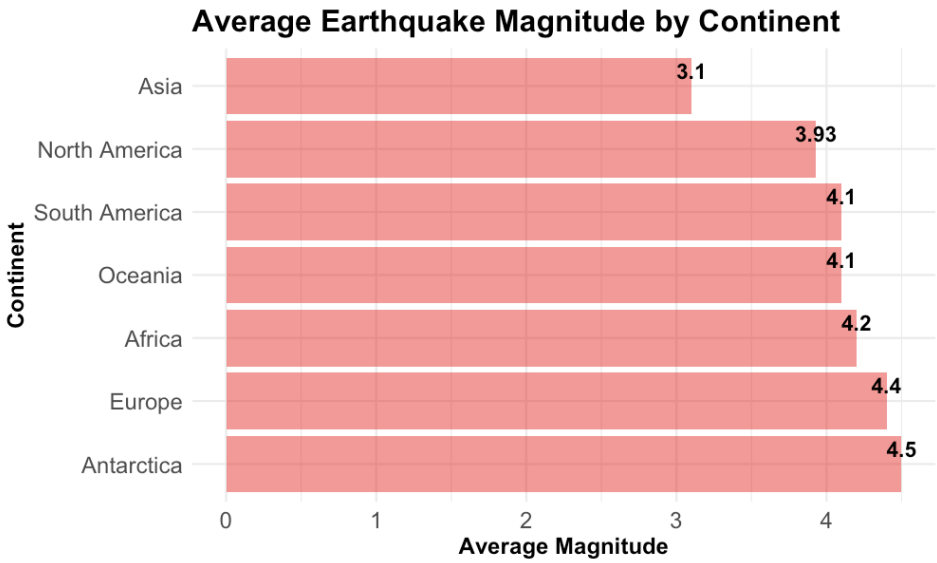
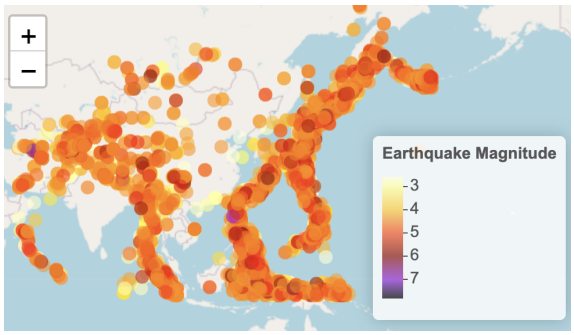
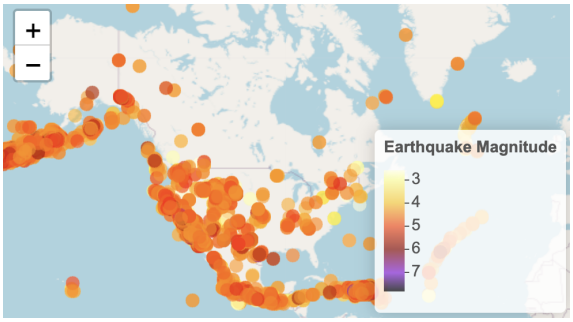


Figure 3.10: Bar chart of average magnitude by continent.

In Figure 3.10, we observe that average magnitudes are almost consistent across most continents, with the specific exception of Asia, which portraits s slightly lower average magnitude. This trend may be attributed to the region's comparatively lower tectonic plate activity, which could result in less seismic intensity than in other areas. On the contrary, it could be due to a higher frequency of lower-intensity earthquakes, which decrease the overall average magnitude.



Heatmap of Asia



Heatmap of North America

Figure 3.11: Comparison of heat maps of Asia and North America

In the Figure 3.11, we compare the heatmaps of Asia and North America. North America, being the second continent with the lowest magnitude, shows no significant differences when compared to Asia. However, the relatively lower average magnitude in Asia may be influenced by a higher frequency of smaller earthquakes (depicted by yellow points), which could be driving the overall downward trend in the data.

Are there seasonal trends or patterns in the occurrence of high-magnitude earthquakes?

The main challenge in answering this question is the limited scope of the dataset, which only includes data from 2023, preventing a broader temporal analysis. However, we can still examine the evolution of earthquake magnitudes over the months of this year.

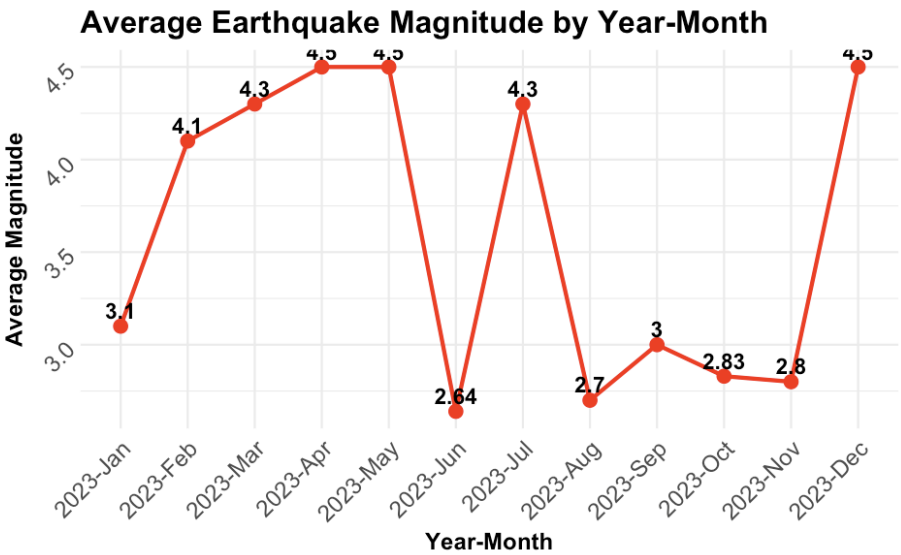


Figure 3.12: Line chart of average magnitude by year-month.

In Figure 3.12, the most significant patterns reveal higher magnitudes during the winter and spring seasons, with the exception of July. The months showing the highest values are April, May, and December, each with an average magnitude of 4.5 while the lowest are June and August at 2.64 and 2.7 respectively.

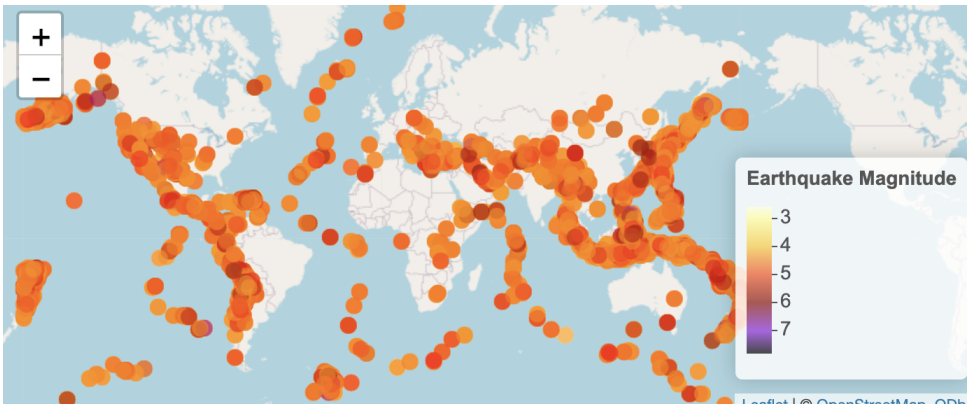


Figure 3.13: Heat map by earthquake magnitude in April-May 2023.

In Figure 3.13, we can observe that some of the highest magnitude measurements (represented in purple) occur during the months of April and May. This likely contributes to the increase in the average magnitude during the early months of the year.

3.2 Earthquake types frequency overtime

3.2.1 Problem characterization

The goal of this analysis is to investigate whether there are temporal patterns in the types of earthquakes by analyzing their frequency over time. Additionally, by applying a magnitude filter, we can examine how magnitude thresholds impact the frequency trends of different earthquake types. Understanding these patterns can help seismologists and geophysicists improve earthquake occurrence predictions and better understand the intensity of various types of seismic activity.

This analysis is useful for risk assessment and disaster preparedness, as different types of earthquakes have varying causes and impacts. For instance, tectonic earthquakes are caused by the movement of tectonic plates and are typically more frequent, whereas volcanic earthquakes are associated with volcanic activity and can signal potential eruptions.

The main questions addressed in this analysis are:

- Are there temporal trends in the frequency of different earthquake types?
- Does magnitude affect the frequency trends?
- Are there any seasonal or recurring patterns in the occurrence of various earthquake types?

By answering these questions, we aim to provide data-driven insights that can support further research on seismic patterns and contribute to improving early warning systems.

3.2.2 Data and task abstraction

To answer these questions, the relevant variables are time, type, and magnitude. Temporal trends require time as a core variable, while type and magnitude are essential to distinguish between different types of earthquakes and their intensities. However, these variables cannot be used directly without preprocessing.

First, the time variable must be converted to a date format to enable temporal grouping. The earthquake events are then grouped by type according to the dates we want to analyze. Finally, we apply a magnitude filter to select only earthquakes within a specific magnitude range and summarize the frequency for each earthquake type.

- **Magnitude Filter:** Allows users to filter earthquake cases according to the magnitude range they want to analyze. This helps focus on earthquakes of specific intensities.
- **Frequency Analysis:** Calculates the daily frequency of each type of earthquake based on the selected magnitude and time range. This task helps identify temporal trends in earthquake occurrences.

While handling the earthquake types in the dataset, we discovered that it includes several distinct types of seismic events. These types can be broadly grouped into three categories: natural earthquakes (e.g., earthquake, volcanic eruption); human-made earthquakes (e.g., mining explosions, quarry blasts); others (e.g., ice quakes, landslides).

```
earthquake         explosion         ice quake         landslide  mining explosion      quarry blast volcanic eruption
     26428                 3                26                 3               167                 2                 13
```

Figure 3.14: Earthquake types

Out of the total 26,642 cases in the dataset, 26,428 cases were classified as "earthquake" (see figure 3.14). However, to allow for a more detailed analysis, we decided to retain each type individually rather than grouping them into broader categories. This approach gives users the flexibility to select and analyze specific types of earthquakes based on their interests.

3.2.3 Interaction and visual encoding

To analyze temporal trends in earthquake frequency, we decided to use a line chart, as it is well-suited for identifying patterns over time. Initially, we considered using a multi-line chart with dual axes, where the left y-axis would represent the frequency of earthquakes and the right y-axis would represent the average magnitude. However, after testing, we opted for a single-line chart focused on frequency because the average magnitude did not provide useful insights for our specific research questions. Additionally, displaying too many lines for each earthquake type made the plot more complex and harder to interpret. Therefore, we decided to use magnitude as a filtering option instead of plotting it as a separate axis.

Each earthquake type is represented by a different colored line, with the legend helping users distinguish between the types. To enhance user interaction, we implemented the following interactive features:

- **Date Range Filter:** Users can select the date range they want to analyze, with the data shown at a daily resolution. Since our dataset only covers one year, aggregating the data by month or year would not provide meaningful insights. Therefore, a daily time unit is the most appropriate for observing trends.
- **Magnitude Range Filter:** Users can filter the earthquake records by magnitude, allowing them to focus on earthquakes of specific intensities. We chose to implement this feature using a slider bar, as it provides a more intuitive experience for users; when the slider increases, the magnitude values grow proportionally, making it easier to understand.
- **Earthquake Type Selection:** Users can select specific earthquake types using checkboxes. This interactive feature displays all available earthquake types in the dataset, enabling users to customize their analysis based on their interests.

Note: According to the magnitude distribution (Figure 3.15), most earthquakes recorded in the dataset have magnitudes between 2.6 and 5.0. Therefore, selecting a higher magnitude range will result in fewer earthquakes being displayed, making it difficult to observe meaningful patterns.

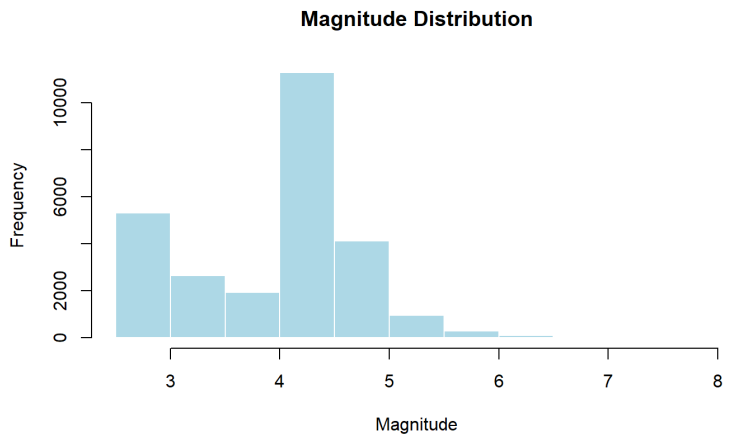


Figure 3.15: Magnitude distribution

This line chart (Figure 3.16) displays the frequency of earthquakes over time based on the selected date range, magnitude range, and earthquake types. The chart allows users to observe temporal trends in seismic activity and identify periods of higher or lower earthquake occurrences. The interactive filters enable users to dynamically adjust the time period, magnitude, and types of earthquakes they want to explore.

It is important to note that, even though all earthquake types are selected in the filters, not all types may appear on the plot. This is because some types, such as "quarry blast", may not have any recorded occurrences that meet the selected date and magnitude criteria. As a result, those types are not represented on the line chart for certain filter combinations.

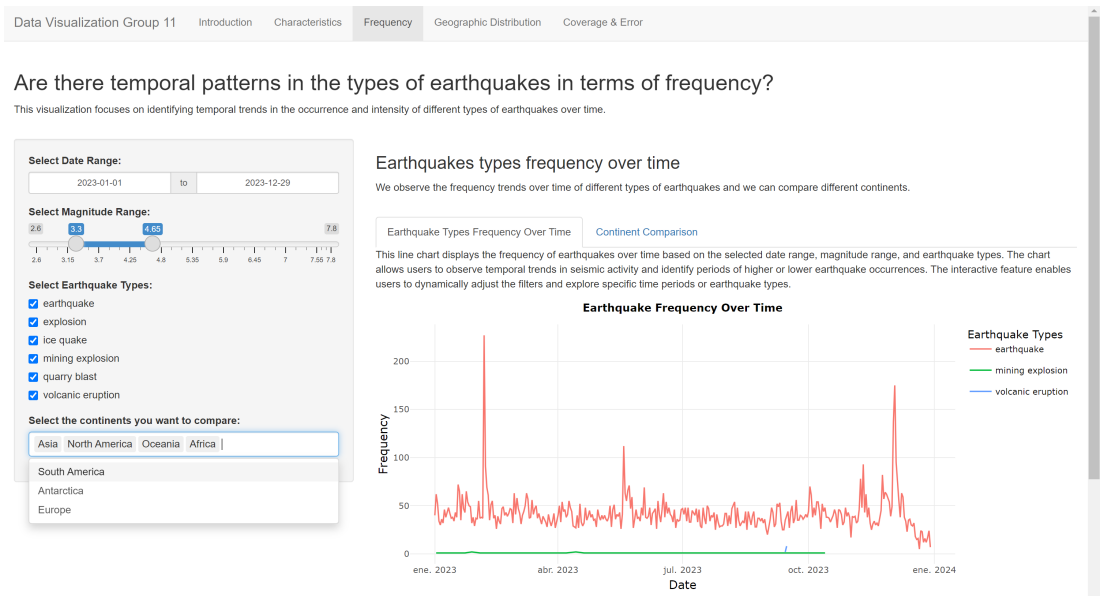


Figure 3.16: Earthquakes types frequency over time panel- Line chart

Additionally, to provide a more visual summary of the data, we included a bar chart (Figure 3.17) that shows the total number of earthquakes for each type within the selected time range. The bar chart complements the line chart by giving a clearer understanding of which earthquake types are more prevalent during the analyzed

timeframe. Users can use this chart to quickly identify the most common types of earthquakes based on their frequency.

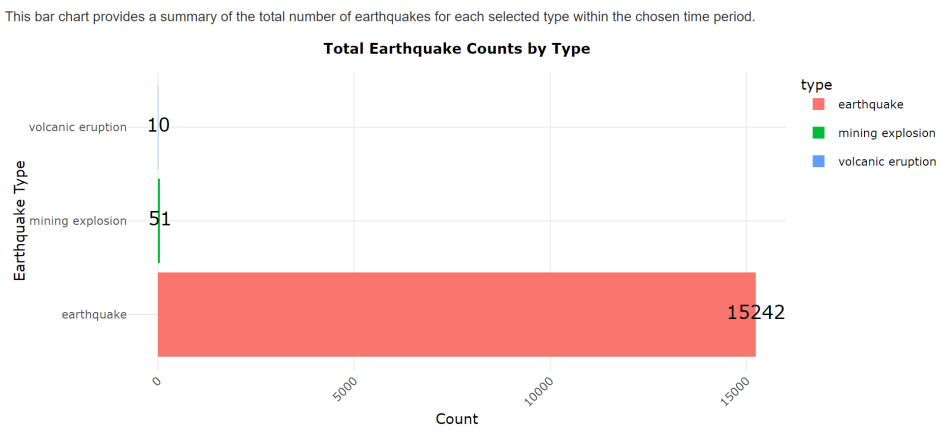


Figure 3.17: Earthquakes types frequency over time panel - Bar chart

For deeper analysis, users can also compare trends across different continents because we analysed the characteristics according to the continents. This feature is synchronized with the main analysis, meaning that the date range, magnitude filter, and earthquake type filters apply to both the general line chart and the regional comparison (Figure 3.18). This allows users to explore regional differences in seismic activity and observe whether certain continents experience more frequent or more intense earthquakes over time.

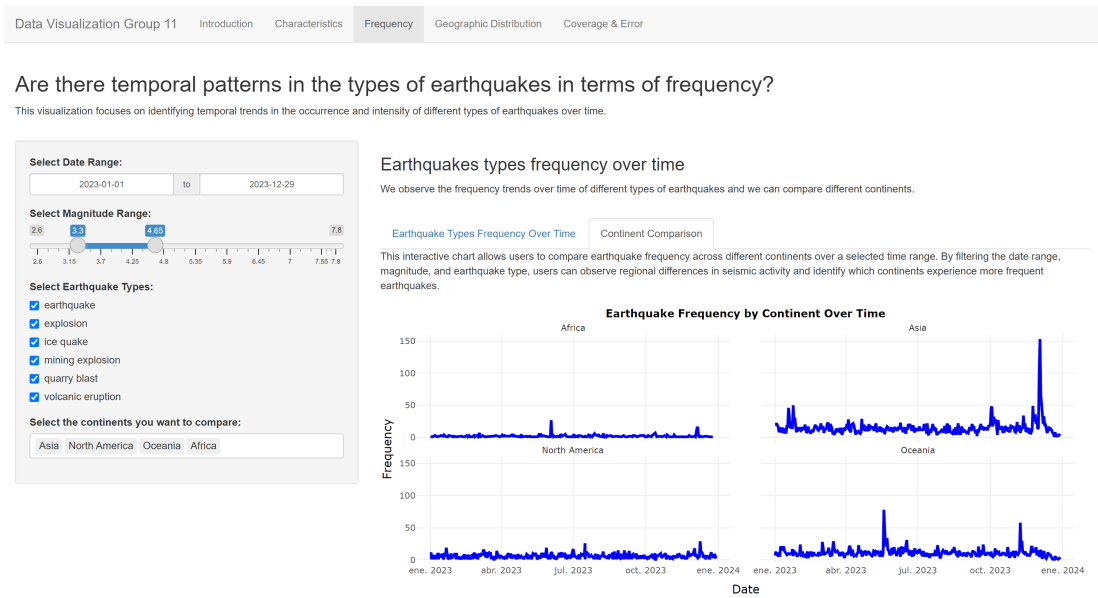


Figure 3.18: Earthquakes types frequency over time panel - Comparison

3.2.4 Algorithmic implementation

The charts used in our analysis were created using *ggplot2* for visualization and *plotly* to make them interactive and user-friendly. By integrating plotly, we enhanced the user experience by allowing users to interact with the charts in various ways:

- Users can hover over points on the line chart to obtain detailed tooltip

information, including the date, frequency, and average magnitude of the earthquakes at that point.

- Users can click on the chart to isolate specific points of interest, showed in Figure 3.19.
- Users can double-click on the legend to isolate one trace.
- The charts also allow users to zoom in and zoom out to explore specific time periods in greater detail.
- Additional actions, such as resetting the view or downloading the plot as an image, are available through the plotly toolbar.

The implementation of these interactive features makes the charts dynamic and engaging, providing users with the flexibility to explore the data according to their needs. As shown in Figure , users can customize their analysis by applying filters for date range, magnitude, and earthquake types.

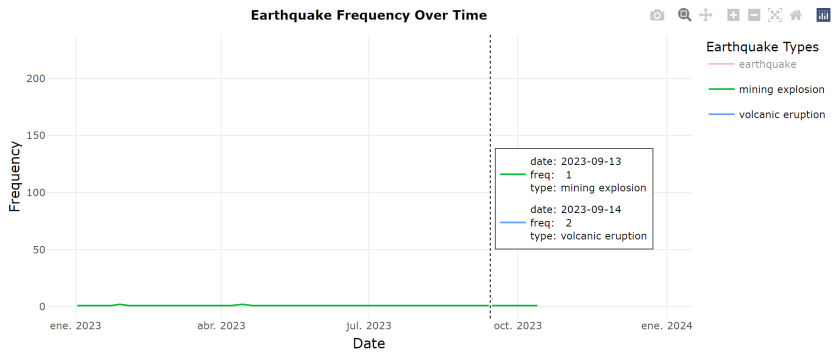


Figure 3.19: Plotly tools: Red line asolated

The algorithmic implementation begins with *Data Preprocessing*, where the time variable is converted to date format, missing values are removed, and the earthquake types are standardized using the `dplyr`, `lubridate`, and `stringr` packages. This ensures that the data is clean and ready for analysis.

Next, the magnitude filtering step allows users to apply a magnitude filter to select earthquakes within the chosen range. This is achieved using the `filter()` function, applied to various reactive objects such as *filtered_data*, *type_counts*, and *regionByContinents*. This filtering enables users to focus on specific earthquake intensities.

The third step, frequency calculation, involves grouping the data by date and earthquake type, followed by calculating the daily frequency using the `group_by()` and `summarise()` functions. This step provides the foundation for visualizing the number of earthquakes over time for different types.

In the chart creation step, the processed data is used to create a line chart representing the earthquake frequency over time and a bar chart to display the total count of each earthquake type. These charts are generated using the `ggplot2` package. Finally, The interactive charts include *line_chart*, *bar_chart*, and *region_chart*, providing users with a comprehensive and engaging experience when analyzing earthquake trends.

3.2.5 Results

The following table summarises the design abstraction process for the Earthquake Frequency Over Time visualization.

Category	Description
Idiom	Line Chart
What	Frequency of earthquakes over time for different earthquake types based on user-selected filters, including date range, magnitude, and type of earthquake.
How	Using ggplot2 for the static plot and plotly for interactivity.
Why	To observe temporal patterns in earthquake occurrences.

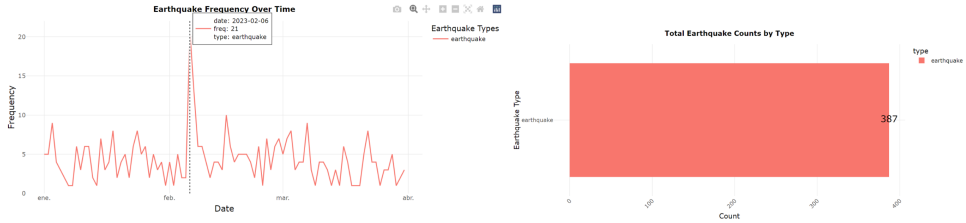


Figure 3.20: Frequency from 2023-01-01 to 2023-03-31 and 3.0 $\leq$ magnitude $\leq$ 6.2

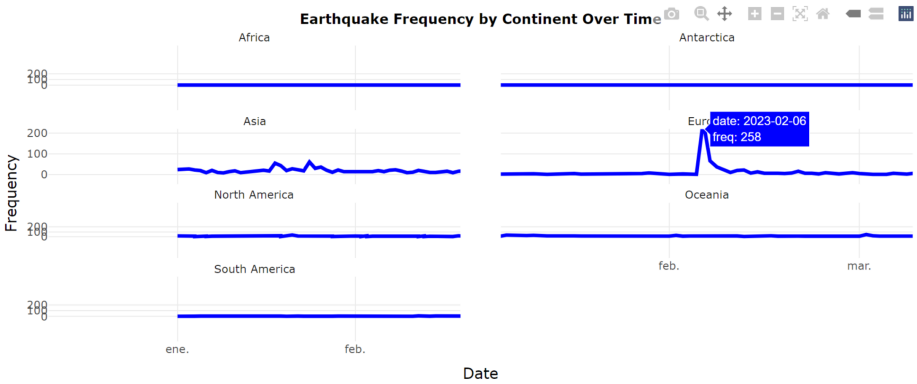


Figure 3.21: Comparison from 2023-01-01 to 2023-03-31 and 3.0 $\leq$ magnitude $\leq$ 6.2

For the selected time range of January 1, 2023, to March 31, 2023, and considering all earthquake types with a magnitude between 3.0 and 6.2 (Figures 3.20, 3.21), we answer the following questions.

Are there temporal trends in the frequency of different earthquake types?

During this period, the frequency of earthquakes (red line) shows no consistent temporal pattern across the entire timeframe. However, there is a clear spike in February, suggesting a series of significant earthquakes that greatly impacted the frequency during that month.

When cross-referencing this spike with the continental analysis, we see that 258 earthquakes occurred in Europe during this period (Figure 3.21), which explains the

sudden increase. This spike is likely due to a single large seismic event or a cluster of aftershocks following a major earthquake.

After the February peak, the frequency of earthquakes returns to a more stable level, fluctuating between 20 to 50 earthquakes per day.

The frequency of mining explosions (blue line) remains low and stable throughout the period. There are no significant spikes or temporal trends observed. This suggests that mining explosions are human-controlled events that occur regularly and predictably, without the variability seen in natural seismic activity.

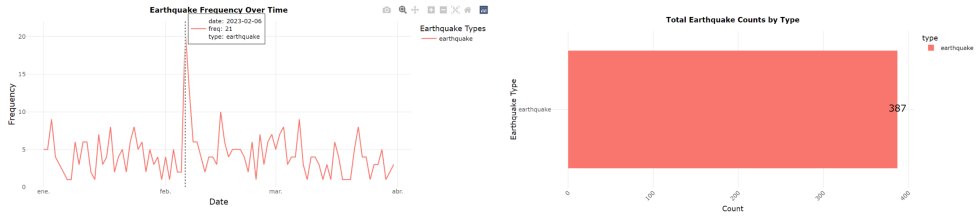


Figure 3.22: Frequency from 2023-01-01 to 2023-03-31 and $5.0 \leq$ magnitude ≤ 6.2

Does magnitude affect the frequency trends?

Yes, magnitude significantly impacts frequency trends. With a magnitude filter of 5.0, we observe 387 cases (Figure 3.22), compared to 5,434 cases with a lower threshold (3.0 to 6.2) (Figure 3.20).

This shows that most earthquakes have lower magnitudes, and higher-magnitude earthquakes are much less frequent. Increasing the magnitude filter reduces the number of cases displayed, revealing that lower-magnitude earthquakes dominate daily seismic activity.

Are there any seasonal or recurring patterns in the occurrence of various earthquake types?

Based on the selected time range and filters, no clear seasonal or recurring patterns are observed in the occurrence of earthquakes. The frequency fluctuates irregularly, with occasional spikes, but these do not appear to follow a predictable cycle or seasonal trend.

3.3 Distribution of key earthquake characteristics and geographic patterns

3.3.1 Problem characterization

Understanding the spatial distribution of earthquake characteristics is important for seismologists and disaster management teams. This analysis focuses on visualizing and analyzing how key earthquake parameters, specifically magnitude and depth, vary across different geographical regions. By mapping these characteristics, we can identify potential seismic hotspots and better understand the relationship between earthquake properties and geographical locations.

The **key questions** this analysis aims to address are:

- Are there specific geographic regions that consistently experience deeper earthquakes?
- Are there correlations between earthquake depth and magnitude in specific areas?
- What patterns emerge when examining clusters of seismic activity?

This visualization is particularly valuable for risk assessment and emergency preparedness, as it helps identify areas that might be prone to specific types of seismic events.

3.3.2 Data and task abstraction

The primary **variables** used in this analysis are:

- Latitude and longitude: Geographic coordinates for mapping.
- Magnitude: The strength of the earthquake.
- Depth: The depth at which the earthquake occurred.
- Time: The timestamp of the seismic event.
- Additional details: Location description, type, and other metadata for detailed information.

The **main tasks** involved in this analysis include:

- Spatial mapping: Plotting earthquakes on a geographic map.
- Visual encoding: Representing magnitude through point size and depth through color intensity.
- Filtering: Allowing users to focus on specific ranges of magnitude, depth, and time.
- Interactive exploration: Enabling users to access detailed information about specific events.

3.3.3 Interaction and visual encoding

We implemented an interactive map visualization to explore the geographic distribution of earthquakes globally. The visualization uses circle markers to represent individual earthquakes, with point size corresponding to magnitude, larger markers indicate stronger earthquakes. Additionally, the color intensity of the markers (shades of brown) represents the depth of the earthquake, where darker shades indicate greater depths. This dual encoding of size and color allows users to quickly grasp both the intensity and depth of seismic events.

To address the challenge of displaying large numbers of earthquakes in a confined space, we incorporated clustering functionality. Clusters provide a visual summary of the earthquake count in specific areas, reducing visual clutter while maintaining a clear understanding of regional earthquake density.

The map includes several interactive features to enhance usability. Users can zoom and pan to explore different regions and click on individual markers to view detailed information about specific earthquakes, such as the date, location, magnitude, and depth. Additionally, users can apply range filters to customize the data displayed:

- Magnitude Slider: Adjusts the minimum and maximum magnitude range to filter earthquakes by intensity.
- Depth Slider: Allows users to select specific depth ranges to analyze shallow or deep earthquakes.
- Time Range Picker: Enables users to filter earthquakes by a specific time period, making it easier to analyze temporal patterns.

The choice of a brown color scheme to represent depth was intentional, as it reflects the natural association with earth layers, making the visualization intuitive. Varying point sizes for magnitude further enhances the user experience by providing a clear visual representation of earthquake strength. The clustering feature helps manage visual complexity, especially when viewing large geographic areas or extended time periods.

Overall, this interactive map allows users to identify geographic hotspots of seismic activity, observe patterns in earthquake depth and magnitude, and customize their analysis through dynamic filtering options. The tool is designed to help seismologists and analysts gain valuable insights into earthquake distribution and risk assessment across different regions.

3.3.4 Algorithmic implementation

The implementation leverages the Leaflet library within R's Shiny framework, ensuring a balance between performance and user experience. The key components of the map are carefully designed to provide users with a clear and intuitive way to interpret earthquake data.

The color encoding for depth visualization utilizes a custom gradient ranging from light brown (#edbb99) to dark brown (#873600), normalized to handle depths from 0 to 700 kilometers. This intuitive design ensures deeper earthquakes appear in darker shades, aiding quick interpretation of depth. The color palette effectively enhances visual clarity while accommodating the entire depth range.

The data filtering system leverages reactive programming for efficient updates and

simultaneous filtering across multiple parameters, including magnitude, depth, and time period. Each earthquake point is enriched with a depth-based color and a popup displaying formatted time, location, magnitude, and depth details. This comprehensive filtering approach ensures seamless exploration and analysis.

The map configuration is designed for optimal usability with boundaries and zoom constraints. The base map offers a minimum zoom level of 2 and a maximum of 5, with a default view centered at longitude 0° and latitude 20°. To prevent excessive scrolling, strict map boundaries are enforced. Additionally, a depth legend is positioned at the bottom right corner to provide users with a clear reference.

Two visualization modes are available to cater to different needs: individual circles and clustered markers. Individual circles, with a fixed radius of 5 units, 60

Interactive updates are managed efficiently through Leaflet's proxy mechanism. When filters are modified, the system clears existing markers, applies the new filtered data, and updates clustering status if enabled. These updates occur without reloading the entire map, ensuring a smooth user experience and maintaining the visualization's interactivity. This implementation ensures smooth performance even with large datasets and provides users with a responsive interface for exploring earthquake patterns. The careful balance of visual elements and technical constraints allows for both detailed analysis and broad pattern recognition.

3.3.5 Results

Once the analysis and visual implementation is completed, we can answer the main questions settled during the problem characterization.

Are there specific geographic regions that consistently experience deeper earthquakes?

Figure 3.23: View of the application map with depth superior to 350 km and magnitude superior to 4.

Upon filtering for earthquakes deeper than 350 km and with a significant magnitude (greater than 4), three distinct geographical regions emerged as areas with a high

concentration of deep earthquakes. These regions include:

- **East Asia & Oceania:** This is the largest concentration of deep earthquakes globally, with two prominent lines of activity likely following the boundaries of tectonic plates.
- **Western South America:** A few very deep earthquakes were observed along the Andes mountain range, indicating a relationship between the region's tectonic activity and seismic depth.
- **Southern Italy:** Several deep earthquakes were noted near the coast of Italy, suggesting a localized seismic pattern in the Mediterranean region.

Are there correlations between earthquake depth and magnitude in specific areas?

When analyzing earthquakes of all depths with significant magnitudes (greater than 4), it becomes clear that deeper earthquakes account for only a small portion of such events. Additionally, these deeper earthquakes do not consistently exhibit higher magnitudes, suggesting a weak or negligible correlation between depth and magnitude across the observed regions.

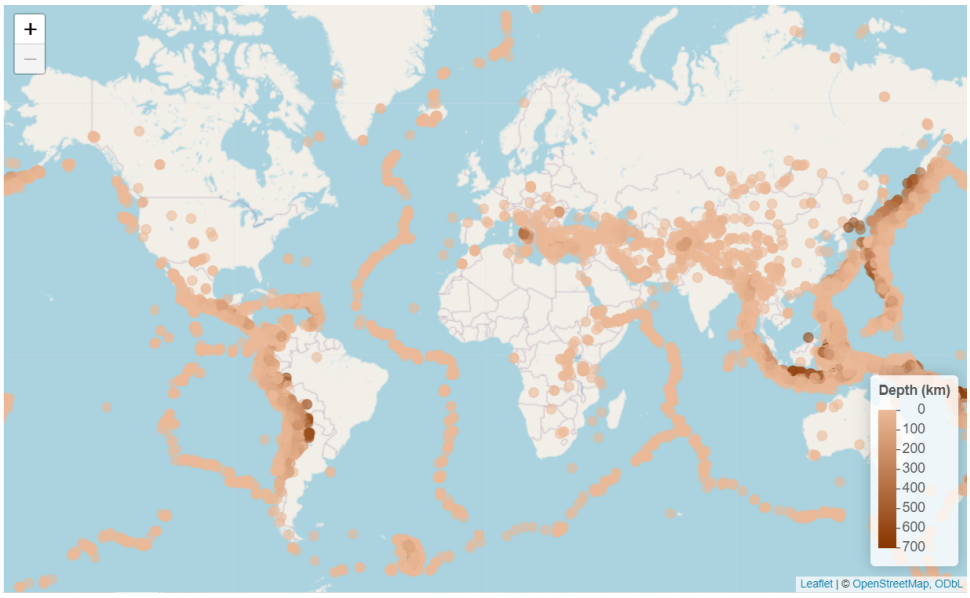


Figure 3.24: View of the application map with depth superior to 350 km and magnitude superior to 6.

However, focusing on "dangerous" earthquakes (those with magnitudes greater than 6) reveals notable regional patterns. South East Asia and South West America emerge as hotspots for these powerful events. While these regions experience both shallow and deep earthquakes, the highest magnitudes are more frequently associated with shallower depths, with deeper earthquakes occasionally contributing to destructive events. This highlights that while depth can play a role, it is not the primary driver of higher magnitudes in these regions.

What patterns emerge when examining clusters of seismic activity?

From the previous views, it became clear that the majority of earthquakes align with tectonic plate boundaries, forming noticeable lines of seismic activity around the

globe. This clustering behavior is a defining feature of earthquake distribution. Notably, regions such as North America, Europe, and Africa are relatively unaffected by “dangerous” earthquakes, indicating their lower seismic activity in comparison to other parts of the world. Interestingly, the deadliest earthquakes in 2023, which struck Syria, Turkey, and Afghanistan, were located outside the most seismically active zones, suggesting that while these regions may not be the most active tectonically, they remain vulnerable to devastating seismic events, likely due to factors such as local geological conditions and population density.

3.4 Coverage of seismic stations across regions and impact on detection and reporting

3.4.1 Problem characterization

Through the study of this problem, our aim is to identify both spatial and temporal patterns in the coverage of seismic stations and its impact on earthquake detection and reporting. Specifically, we focus on parameters such as the gap between seismic stations (gap), the minimum distance to the earthquake epicenter (dmin), the number of seismic stations (Nst) that report each earthquake and various types of detection errors (horizontal, depth, magnitude). These measures are crucial in determining the quality of reported earthquake data and how they influence the reliability and accuracy of recorded data.

Questions addressed in this analysis:

- How does the coverage of seismic stations vary across different regions?
- What is the impact of seismic station coverage on the accuracy of earthquake detection (horizontal error, depth error, magnitude error)?
- Are there specific regions where seismic station coverage is insufficient, leading to higher detection errors?

By answering these questions, we aim to gain insight into possible correlations that can lead us to make valuable conclusions about improving earthquake monitoring systems.

3.4.2 Data and task abstraction

To answer these questions, the relevant variables are Dmin, Gap, Nst, and the detection errors (horizontal, depth, magnitude). Spatial trends require Dmin and Gap as core variables, while Nst and detection errors are essential to understand the reliability of earthquake detection.

Primary tasks for this analysis:

- **Data Preprocessing:** Determine the continent by the description of the earthquake location (place) and, in case of missing or invalid data, convert latitude and longitude to determine the continent, and handle missing or invalid data.
- **Coverage Analysis:** Calculate the average Gap, Dmin and Nst for each continent and network.
- **Error Analysis:** Calculate the average horizontal error, depth error, and magnitude error for each continent and network.

3.4.3 Interaction and visual encoding

To analyze the impact of seismic station coverage on detection and reporting accuracy, we decided to use a bubble chart and a heatmap, as they are appropriate for representing multi-dimensional data. The bubble chart effectively visualizes the relationship between Gap, Dmin, and Nst, while the heatmap highlights detection errors across different regions and networks. We added detailed hover information for both the bubble chart and the heatmap, providing immediate access to detailed

data without cluttering the visualizations, quickly gaining insights into specific data points.

Bubble Chart: Represents the coverage of seismic stations networks across regions. Each continent is represented by a different color in the bubble chart, with bubble sizes indicating the mean number of seismic stations (Nst).

- X-axis: Mean Minimum Distance - Dmin (km)
- Y-axis: Mean Gap (km)
- Bubble Size: Mean Number of Stations - Nst
- Color: Continent

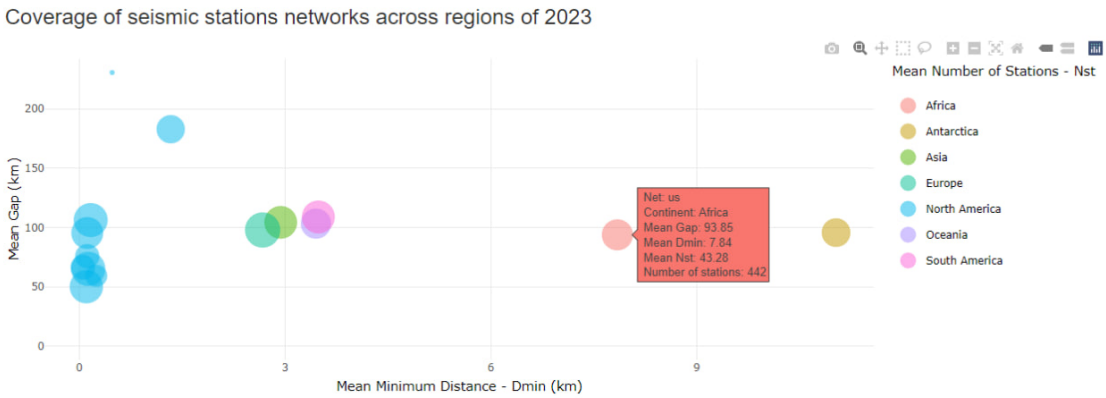


Figure 3.25: Bubble chart overview

Heatmap: Encodes the average detection errors (horizontal, depth, magnitude) for each network across different continents. By using color gradients, the heatmap effectively communicates the magnitude of errors, where lighter shades represent higher errors and darker shades represent lower errors.

- X-axis: Continents
- Y-axis: Nets
- Fill: Selected error type (Horizontal, Depth, Magnitude)

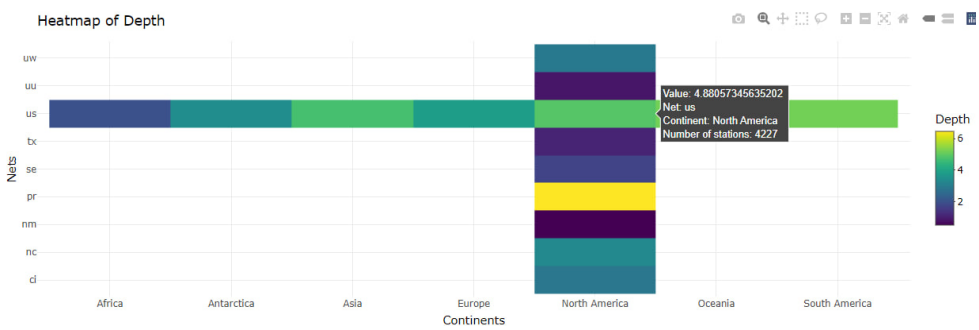


Figure 3.26: Heatmap overview

Interactive Features: The bubble chart and heatmap are synchronized with the filters, ensuring that changes made in the selection of continents, networks, error types, or Gap and Dmin ranges are reflected in both visualizations. This

synchronization allows users to dynamically explore the relationship between seismic station coverage and detection accuracy, providing valuable insights into regions that may require improved monitoring systems.

- Continent Selection: Users can select specific continents to focus on. This is done using checkbox inputs.

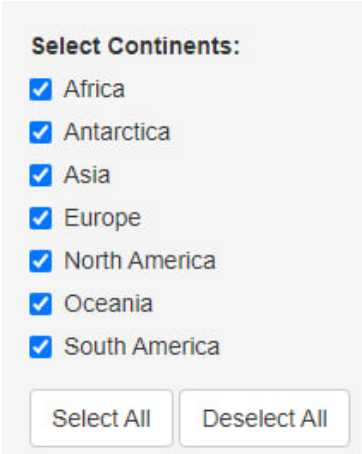


Figure 3.27: Continent selector overview

- Network Selection: Users can select specific networks to focus on. This is also done using checkbox inputs.

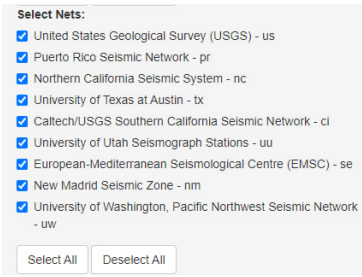


Figure 3.28: Network selector overview

- Error Type Selection: Users can choose the type of error (Horizontal, Depth, Magnitude) to visualize in the heatmap. This is provided as a dropdown input.

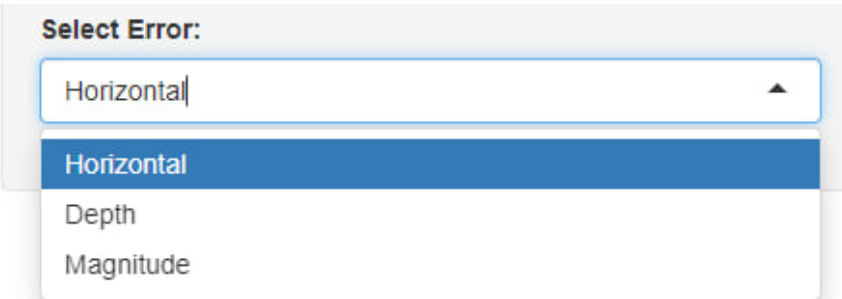


Figure 3.29: Error selector overview

- Gap and Dmin Range Filters: Users can filter the earthquake records by Gap and Dmin, allowing them to focus on specific ranges. This is implemented using slider bars for an intuitive user experience.

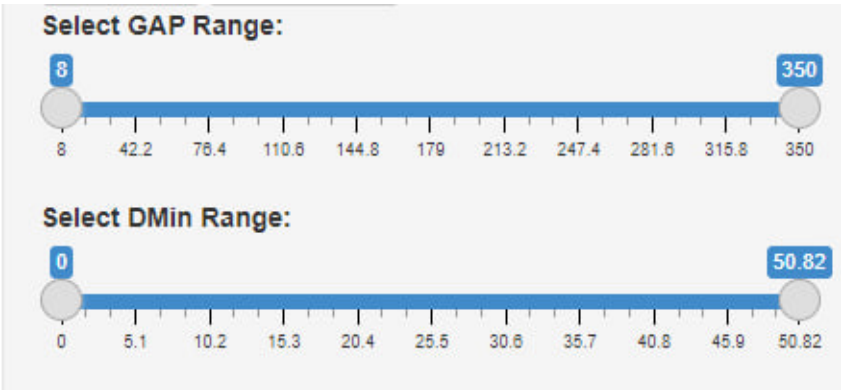


Figure 3.30: Ranges (Dmin, Gap) selector overview

This interactive bubble chart and heatmap enable users to dynamically explore the relationship between seismic station coverage and detection accuracy, providing valuable insights into regions that may require improved monitoring systems.

3.4.4 Algorithmic implementation

The algorithmic implementation of this analysis utilizes several powerful features of R, primarily from the Shiny, dplyr, ggplot2, plotly, countrycode and viridis packages, to enable interactive data exploration. The Shiny framework sets up the infrastructure for dynamic filtering of earthquake data based on user inputs, such as selected continents or networks or specified ranges of the observed data. The observeEvent function is used to trigger updates when users select or deselect regions or date ranges, while reactive expressions ensure that the displayed data is updated in real-time.

To handle data processing, the dplyr package is employed for efficient data manipulation, allowing for grouping by continent with countrycode or network and calculating averages for Gap, Dmin, Nst, and detection errors. For visualization purposes, ggplot2 and plotly is utilized to create bubble charts and heatmaps colored with viridis that dynamically display these averages with labels.

The algorithmic implementation can be summarized in the Table 3.3.

Library	**Main Usage**
Shiny	Interactive features like dynamic filtering and real-time data updates.
dplyr	Efficient data manipulation for grouping and calculating averages.
ggplot2	Visualization figures such as bubblechart and heatmaps.
plotly	Interactive charts, hover information (labels).
countrycode	Determination of the continent
viridis	Color for heatmap

Table 3.3: Summary of Libraries Used.

3.4.5 Results

Upon concluding the analysis and visual implementation, we are now in a position to address the critical questions identified during the problem characterization phase. This section delves into answering these questions by analyzing the patterns and trends observed in the data.

How does the coverage of seismic stations vary across different regions?

By examining the bubble chart, we can observe the variation in seismic station coverage across different regions. Notably, the network "United States Geological Survey (USGS)",the label "us", accounts for approximately 80% of the dataset and is the only network present across all continents. Which limit us in the comparison of networks globally to only the USGS and we can only compare networks on the north America region.

Network USGS: For the USGS network, the mean Gap across continents is around 100 km, except in North America, where it is approximately 180 km. In most regions, Dmin is around 2.5-3.5 km, with a mean Nst between 40 and 64 stations. Exceptions include:

- North America: Dmin of 1.3 km with a mean Nst of 34 stations.
- Africa: Dmin of 7.8 km with a mean Nst of 43 stations.
- Antarctica: Dmin of 11 km with a mean Nst of 35 stations.

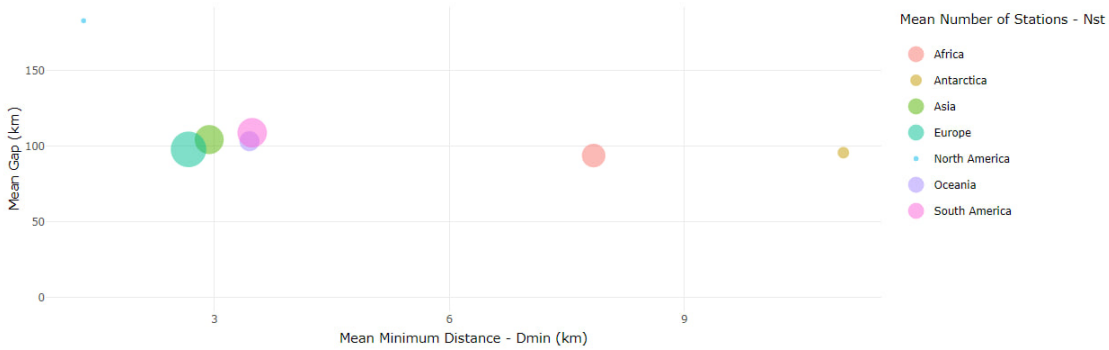


Figure 3.31: Network USGS coverage

Other Networks in North America: The remaining networks (excluding "us"):

All the stations excepting "Puerto Rico Seismic Network (PRSN)", the label "pr", are around the mean Gap range from 50 to 106 km, within the mean Dmin range from 0.05 to 0.25 km and with a mean Nst between 22 and 57 stations. The exception, the PRSN network has a Gap of 230 km, Dmin of 13.98 km, and a mean Nst of 14 stations.

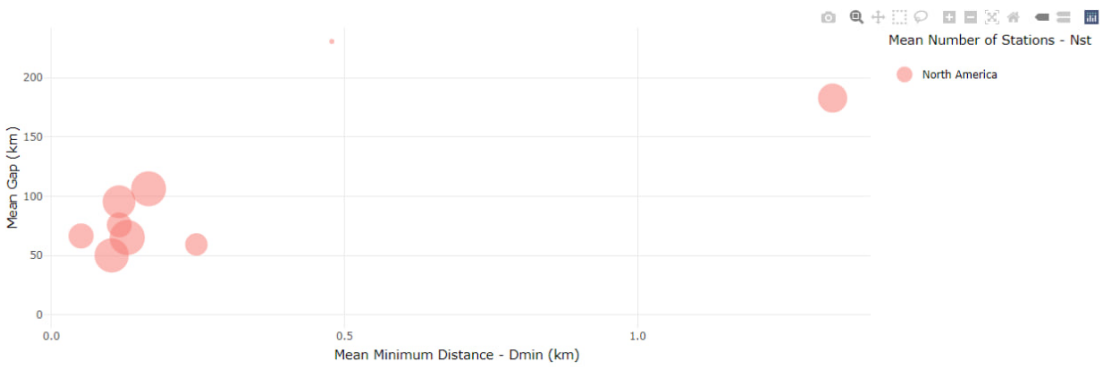


Figure 3.32: Other networks coverage in North America

In conclusion, we observe that the PRSN and USGS networks are more scattered than the rest, especially the PRSN network, which has a low count of Nst, for his higher Gap and Dmin values. The USGS stations are usually farther from the epicenter than other networks, possibly indicating a need for denser coverage, particularly Africa, and Antarctica.

What is the impact of seismic station coverage on the accuracy of earthquake detection (horizontal error, depth error, magnitude error)?

We will analyze each error separately:

- **Horizontal Error:** We observe that most errors, excluding PRSN and USGS, are around 0.25 km and 0.91 km. The PRSN network has an average horizontal error of around 2 km, while the USGS network shows varying values with North America at a minimum of 5.5 km and the highest errors observed in Antarctica (11.3 km), Oceania (10.1 km), and Africa (9.5 km), with the rest around 7.5 km.

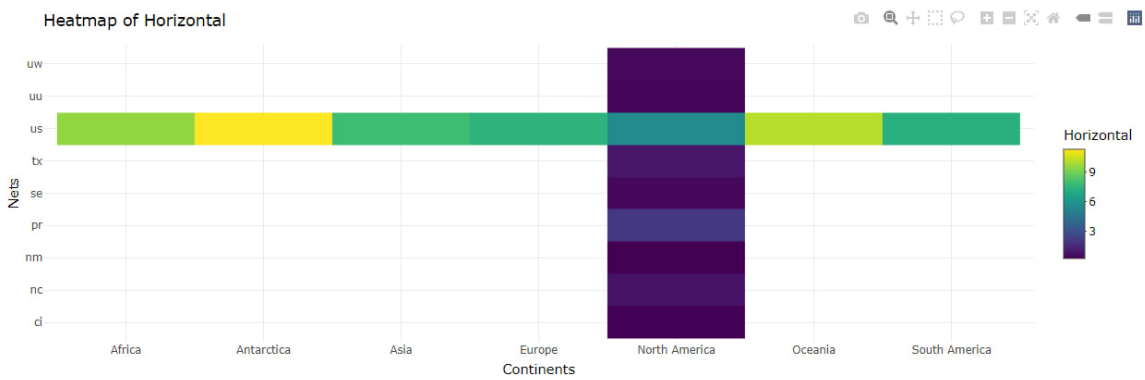


Figure 3.33: Horizontal Error

- **Depth Error:** Excluding USGS and PRSN, most regions have depth errors between 0.8 km and 3.3 km. The PRSN network has a depth error of around 6.4 km, while the USGS network shows varying values with North America at a minimum of 4.8 km and the highest errors observed in Oceania and South America (5.2 km), and the lowest in Africa (2 km).

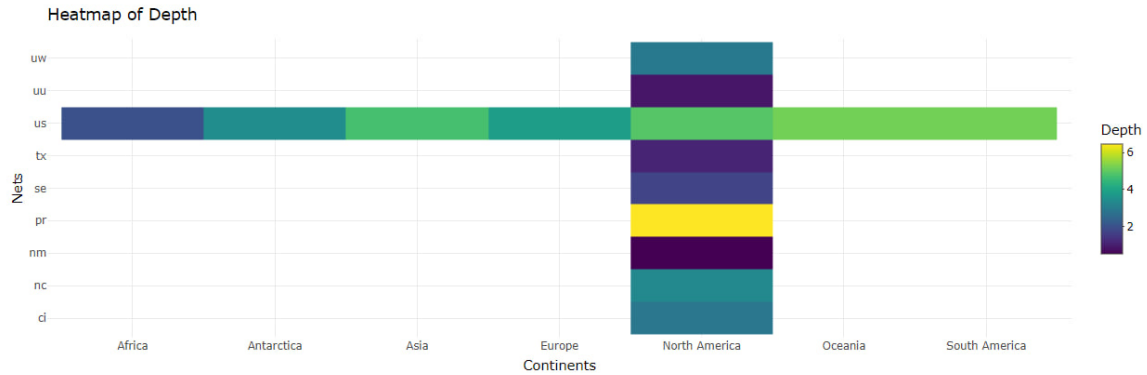


Figure 3.34: Depth Error

- **Magnitude Error:** The mean magnitude error is generally small, ranging from 0.08 to 0.18, suggesting that magnitude errors are not significantly impacted by seismic station coverage.

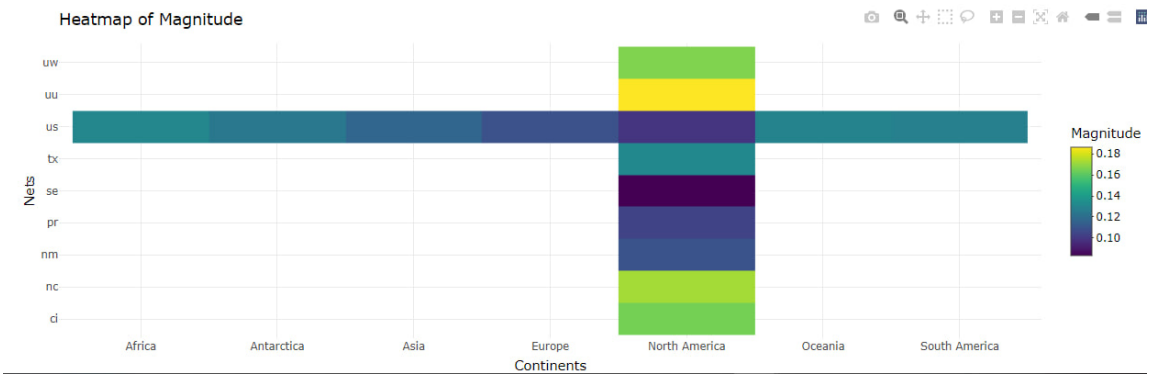


Figure 3.35: Magnitude Error

From these results and those of the previous question we could elucidate there is a relation between horizontal and depth errors and Dmin, Gap and Nst values, a behavior resembling to: "Regions with larger gaps, dmins and fewer seismic stations tend to have higher detection errors".Due to the distribution of the data and his complexity, it is difficult to assign a weight to each variable on their contribution it would seem that gap and nst are the main one and then dmin. Further analysis is required to obtain more conclusive results.

Are there specific regions where seismic station coverage is insufficient, leading to higher detection errors?

Even with all these uncertainties, we have found some regions and networks where seismic station coverage appears to be insufficient, leading to higher detection errors:

- **North America:**
  - Network USGS: With a mean Gap of 180 km and a Dmin of 1.3 km. It shows some of the highest horizontal errors (minimum 5.5 km) and significant depth errors (4.8 km) on the region. This suggests that despite having a substantial number of stations, there might be issues related to station placement or other factors impacting detection accuracy.

  - Other Networks: The PRSN network showing high Gap (230 km) and Dmin (13.98 km) values, and higher horizontal (2 km) and depth (6.4 km) errors than rest on the region. This indicates that the PRSN network's scattered coverage contributes to higher detection errors. Surely affected by his low Nst (14 stations).

- **Other regions:** The USGS network shows a high horizontal error and depth error across different regions even with his high number of station and mean Nst. The main region to improve would be Antarctica and Oceania, followed by Africa and South America.

4. Instructions to run the App

4.1 System Requirements

- **RStudio** (recommended for easier package management and running)
- **R** (version 4.0.0 or higher). To run this Shiny application, the following R packages must be installed:
  - shiny: For building interactive web applications.
  - dplyr: For efficient data manipulation.
  - ggplot2: For creating static visualizations.
  - plotly: For adding interactivity to the charts.
  - leaflet: For creating interactive map visualizations.
  - lubridate: For handling and manipulating date and time.
  - stringr: For string manipulation.
  - countrycode: For converting country names to continents.

```
install.packages(c("shiny", "dplyr", "ggplot2", "plotly", "leaflet", "lubridate", "stringr", "countrycode"))
```

4.2 Data Requirements

- Download the earthquakes dataset from Kaggle: Kaggle Dataset - Earthquakes 2023 Global.
- Place the downloaded file in the same directory as the R script and rename it as "earthquakes.csv".

Ensure the project folder (*DV_g11_earthquakes*) contains the following files in the structure below:

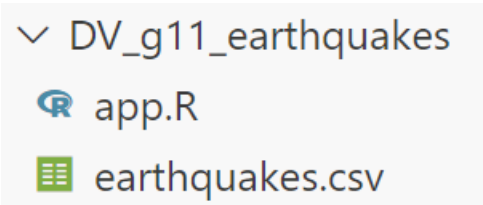


Figure 4.1: Folder structure

4.3 Running the Application Locally

To run the Shiny app locally in RStudio:

1. Download the project folder containing all files.
2. Install the required packages by running the following command in R.
3. Set your working directory to the folder containing your app files.
4. Run the Shiny app using the following command *shiny::runApp*.

```
install.packages(c("shiny", "dplyr", "ggplot2", "plotly", "leaflet", "lubridate", "stringr", "countrycode"))
setwd("path/to/DV_g11_earthquakes")
shiny::runApp()
```

4.4 Access Via Web

The application is available in https://www.shinyapps.io. (click this link)