BUSA8090:

Data and Visualisation for Business

Assignment 2

Global Earthquake Visualization

Quoc Phong Nguyen

Student ID: 48203971

TABLE OF CONTENTS

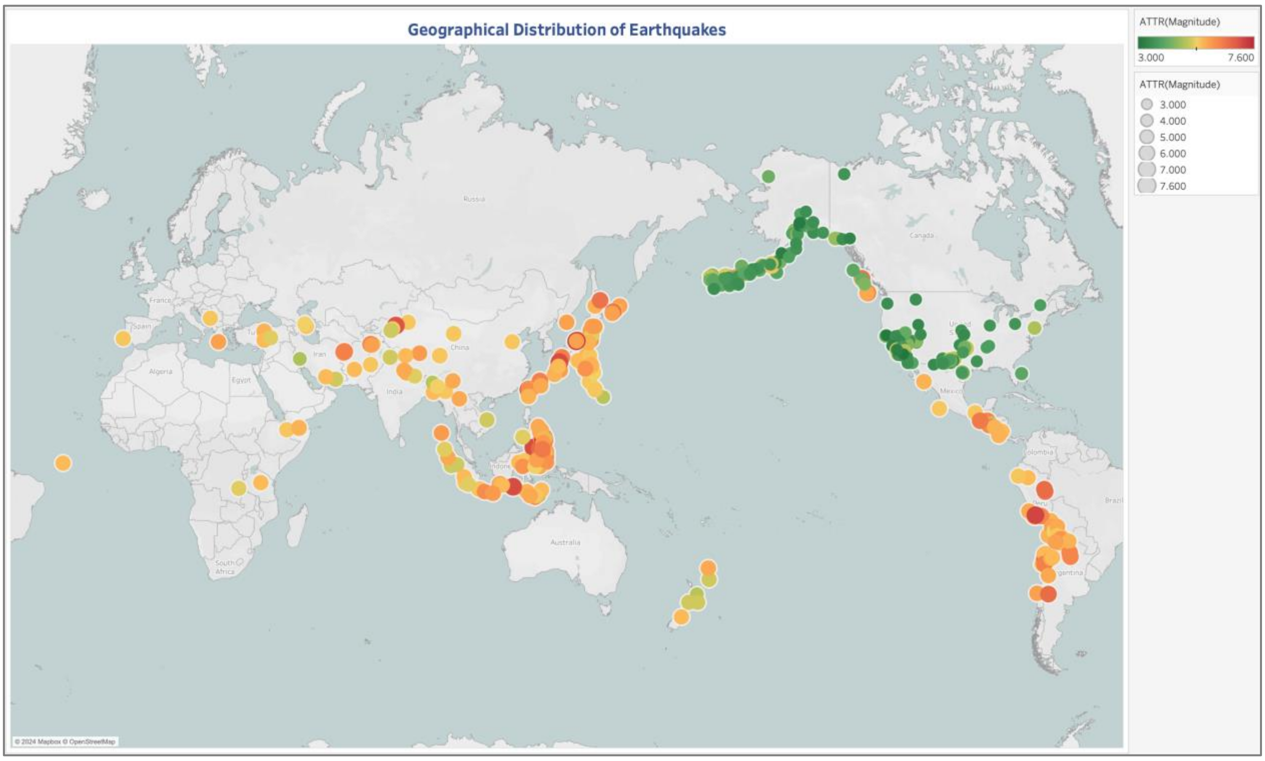
1. INTRODUCTION

Earthquakes pose significant threats to communities around the world, making it essential to have adequate emergency response and disaster management strategies in place. This report utilizes data visualization techniques with Tableau to analyze a comprehensive global earthquake dataset, focusing on critical parameters such as magnitude, depth, geographical distribution, and temporal patterns. By creating carefully designed visualizations, we aim to convert complex seismic data into actionable insights. These insights will empower emergency response directors to make informed decisions regarding resource allocation, risk assessment, and preparedness protocols. Our ultimate goal is to enhance community resilience and protect vulnerable populations through data-driven decision-making.

2. PART A: PRACTICAL WORK & DISCUSSION

Prior to visualization in Tableau, we conducted data exploration and preprocessing using Python programming language. This dataset is provided within the scope of the academic unit. The original version contains 1,137 records; however, after handling 337 duplicates, the final number of observations was reduced to 800. Upon checking for missing values, we found significant gaps in columns such as continent, country, and alert. For records with missing continent and country values, we developed a function to extract corresponding information from the placeOnly column before performing mapping with standardized and corrected values across the entire dataset. For the alert column, we analyzed the distribution of magnitude values and determined appropriate ranges to fill in corresponding alert levels. After completing the data exploration and cleaning process, the cleaned dataset was exported as an Excel file for visualization in Tableau. The following sections present our visualizations and provide detailed analysis of the insights and predictive implications derived from each figure.

2.1. Geographical Distribution of Earthquakes

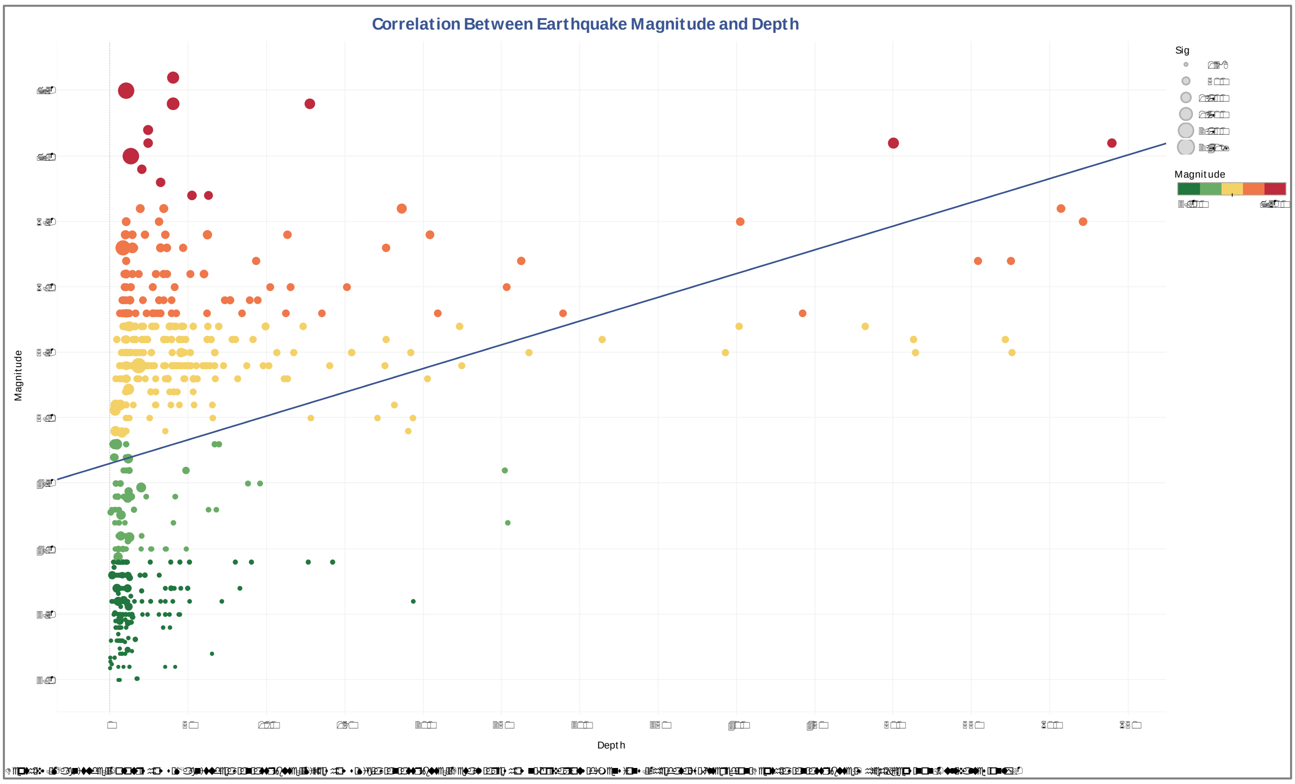


Using Tableau's mapping capabilities, we created an interactive geographical visualization by plotting earthquake events using latitude and longitude coordinates. We enhanced the visualization by encoding earthquake magnitude through both size and color attributes, with larger and darker-colored circles representing higher magnitude events. The map was further customized with detailed tooltips containing crucial information such as magnitude, country, continent, specific location, and date of occurrence. For optimal visualization, we implemented a magnitude-based color gradient ranging from 3.0 to 7.6 on the Richter scale.

The visualization reveals distinct patterns in global seismic activity, highlighting several significant earthquake zones. The Pacific Ring of Fire is clearly visible, with high concentrations of seismic activity along the western coast of South America, through Indonesia, and Japan. Notable clusters of high-magnitude earthquakes (shown in orange and red) are particularly evident in Indonesia, Japan, and Chile, while lower magnitude events (shown in green) are more prevalent along the North American west coast and Alaska.

For Emergency Response and Disaster Management Director, this visualization serves as a crucial planning tool. Regions with concentrated high-magnitude earthquakes, particularly in Southeast Asia and South America, require robust emergency response infrastructure and resources. Directors should prioritize establishing rapid response protocols and maintaining well-equipped emergency facilities in these high-risk zones. Additionally, the visualization suggests that countries along the Pacific Ring of Fire should maintain heightened preparedness levels year-round and consider implementing more stringent building codes and safety measures.

2.2. Correlation Between Earthquake Magnitude and Depth

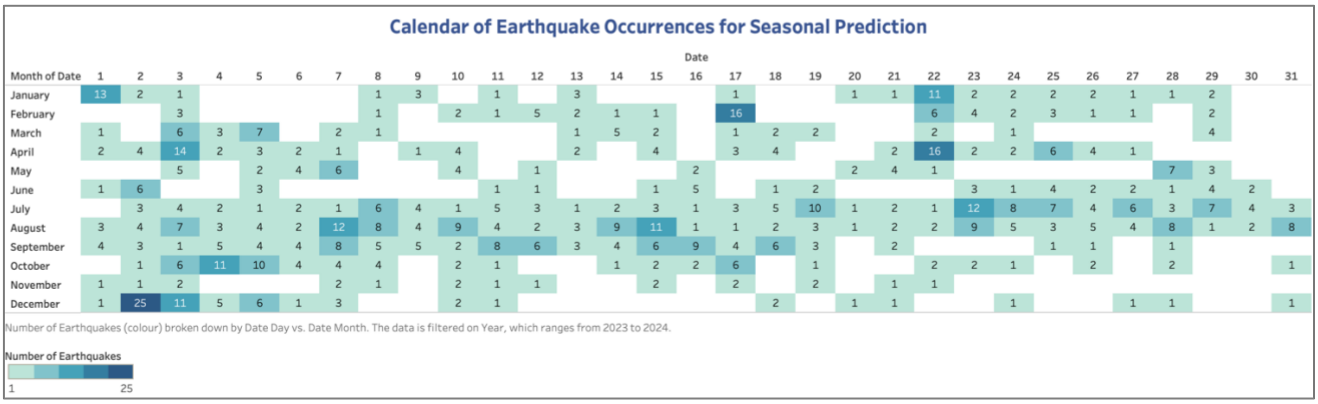


Next, we used Tableau to create a scatter plot representing earthquake depth on the x-axis and magnitude on the y-axis. Each earthquake event is represented by a circle, with the size indicating significance level and colours ranging from green to red representing different magnitude levels. To identify patterns, we added a trend line to visualize the correlation between depth and magnitude. The visualization includes detailed tooltips showing magnitude, depth, significance, alert level, location, and date information for each data point.

The scatter plot reveals a positive correlation between earthquake depth and magnitude, as indicated by the upward-sloping trend line. Most earthquakes cluster in the shallow depth range (0-100 km) with varying magnitudes from 3.0 to 7.6. Higher magnitude earthquakes (shown in red and orange) tend to occur at both shallow and deep depths, while lower magnitude events (shown in green) are predominantly found at shallower depths. Notably, there are fewer earthquakes recorded at extreme depths (>500 km), and those that do occur tend to have higher magnitudes.

For the Director, this correlation provides crucial insights for risk assessment and response planning. Shallow earthquakes, being more numerous and potentially destructive due to their proximity to the surface, require immediate response capabilities and robust early warning systems. The presence of high-magnitude deep earthquakes, though less frequent, suggests the need for specialized monitoring and response systems based on earthquake depth. The Directors should consider developing depth-specific response strategies, particularly for regions prone to both shallow and deep seismic activities.

2.3. Calendar of Earthquake Occurrences for Seasonal Prediction

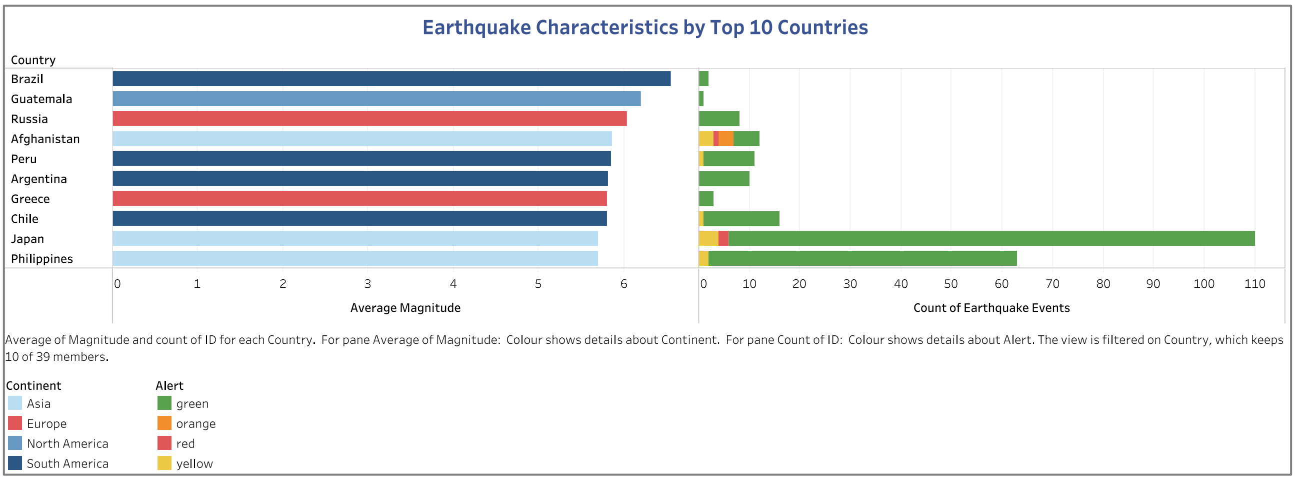


The calendar heat map emerges from a thoughtful combination of day-month dimensions and earthquake frequency data. Each cell in this calendar-style visualization represents a specific day of the year, with colour intensity indicating the number of recorded earthquakes. We selected square-shaped markers for optimal grid presentation, while the sequential blue colour palette provides an intuitive visual hierarchy of seismic activity levels.

Examining the visualization reveals fascinating temporal distributions throughout the year. The late summer months, particularly August and September, indicate intense seismic activity, with several days recording 8-12 earthquakes (shown in darker blue). A notable spike appears on December 2nd with 25 recorded events, representing an extreme pattern. The data also suggests a subtle pattern of increased activity during mid-month periods across multiple months, particularly around the 15th-16th days.

The Director can utilize these findings for predicting earthquake events seasonally. The clear identification of high-activity seasons allows for strategic contingency plans, suggesting increased preparedness during August and September, as well as vigilance during mid-month periods. The anomaly observed on December 2nd requires special attention; however, the overall seasonal patterns should inform the timing of emergency drills, maintenance schedules, and staff deployment. It is advisable for the Director to maintain heightened readiness during historically active periods while ensuring flexible resource allocation to address the inherent unpredictability of seismic events.

2.4. Earthquake Characteristics by Top 10 Countries

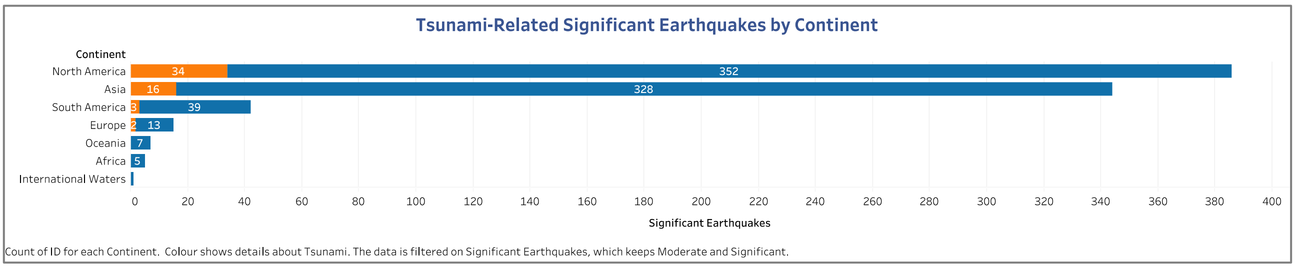


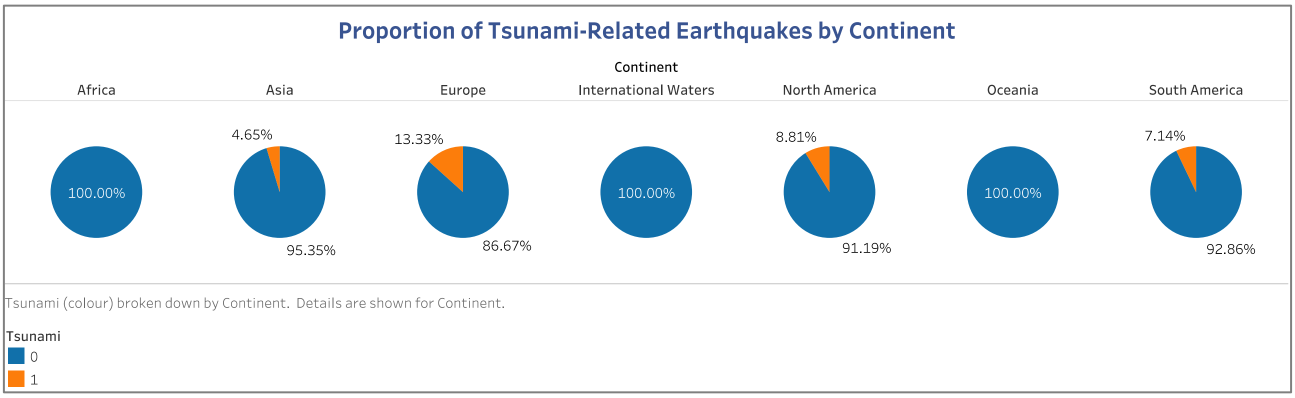
We created combined visualizations to examine global seismic activity through two distinct lenses. By examining the top 10 earthquake-prone nations, we have uncovered a relationship between seismic intensity and frequency. On the left side of our visualization, average earthquake magnitudes are displayed using continent-coded bars. Conversely, the right side presents a more detailed view through stacked bars that illustrate earthquake frequencies across different alert levels, ranging from precautionary green to urgent red. This combination of metrics transforms raw seismic data into clear, actionable insights, providing a comprehensive overview of each country's unique earthquake profile.

Our visualizations indicate that Brazil has the highest average earthquake magnitude, closely followed by Guatemala and Russia, with each recording an average magnitude of around 6.0. This leading group showcases geographical diversity, encompassing South America, Central America, and Europe. Although Japan and the Philippines also appear in the top 10, they exhibit lower average magnitudes but significantly higher event frequencies. Japan, in particular, stands out with over 100 recorded events, most of which are classified as green alerts.

These insights reshape our understanding of global earthquake risk patterns. Countries such as Brazil, Guatemala, and Russia need to be prepared for high-magnitude seismic events, even though they experience them less frequently. In contrast, Japan illustrates the effectiveness of comprehensive monitoring systems that can detect and classify a wide range of seismic activities. The varying distribution of alert levels across different regions highlights the necessity for tailored response strategies, from focusing on structural resilience in areas prone to high-magnitude events to developing advanced early warning systems in regions with frequent seismic activity.

2.5. Significant Earthquakes and Tsunami Occurrences by Continent



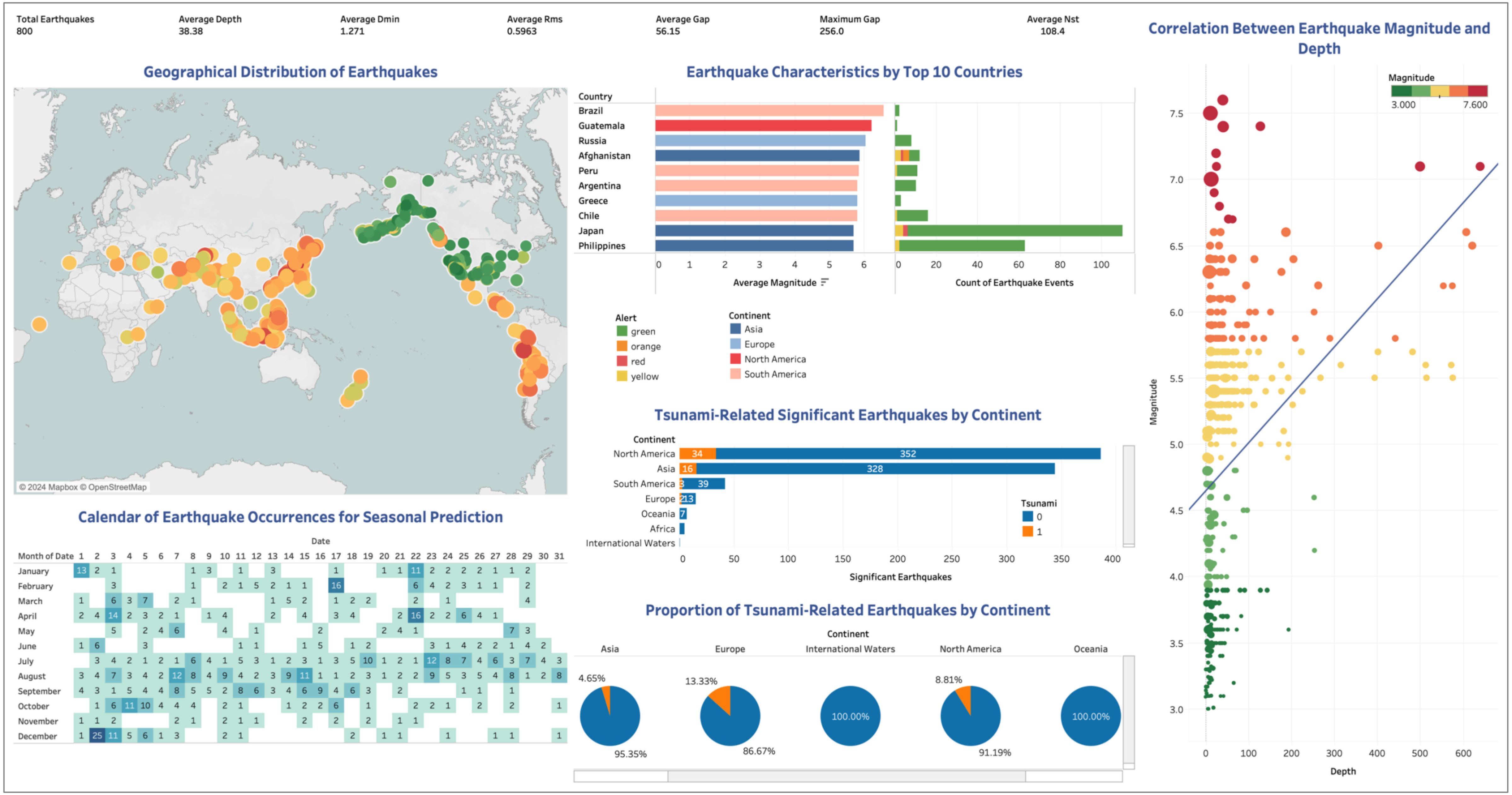


We continued using a dual visualisation approach to explore the relationship between significant earthquakes and tsunami occurrences across continents. The stacked bar chart presents the total number of significant earthquakes, using colour coding to distinguish between tsunami and non-tsunami events. Additionally, we included a series of pie charts that depict the proportion of tsunami-generating earthquakes for each continent, providing both quantitative data and percentage-based perspectives.

The figures reveal North America and Asia as epicenters of seismic activity, collectively accounting for around 700 significant earthquakes. Notably, Europe has the highest rate of tsunami generation, with 13.33% of its earthquakes resulting in tsunamis, despite a lower overall number of seismic events. In contrast, although Asia experiences many more earthquakes, its tsunami generation rate is relatively low at 4.65%. This suggests that factors such as earthquake magnitude and location may play a more crucial role in tsunami generation than the frequency of earthquakes.

For the Director, these patterns highlight key focus areas for emergency preparedness. Europe's coastal regions require strong tsunami response measures despite their overall lower seismic activity. In contrast, North America and Asia authorities should maintain robust general earthquake response systems with additional coastal safeguards. We recommend implementing region-specific early warning systems that consider the varying rates of tsunami generation, especially in areas where the risk of tsunamis is disproportionately high compared to the overall frequency of earthquakes.

2.6. Dashboard of Global Earthquake Data



3. PART B: DATA GOVERNANCE AND ETHICS

3.1. Data Analytics Process in Data Governance

In today's data-driven business environment, data governance has become a critical component of organizational success. Data governance encompasses the principles and practices that ensure high-quality data management throughout its lifecycle, enabling organizations to handle data consistently while supporting business outcomes (Brous et al., 2016; Idemudia et al., 2024). The necessity for robust data governance has intensified with the emergence of advanced technologies such as Machine Learning and AI, which heavily rely on data quality (Janssen et al., 2020; N. Gudivada et al., 2017). Furthermore, successful digital transformation initiatives require well-governed data assets, making it imperative for organizations to implement effective data governance frameworks that align with their business objectives and operational models (Dingre, 2023).

To optimize data governance through analytics, organizations need to implement a structured analytical process that comprises four interconnected stages. The process begins with **data discovering and defining**, where organizations identify and understand the data required for their specific objectives (Alhassan et al., 2016). This initial stage involves exploring available data sources, determining data requirements, and establishing clear definitions for data elements. This foundational step ensures that subsequent analytical processes are built on a clear understanding of data needs and their alignment with business goals.

Following discovery, the **data quality management** stage focuses on consolidating and harmonizing data from various sources. This phase involves implementing rigorous cleaning protocols, establishing standardized formats, and creating efficient integration workflows. The emphasis on data quality management at this stage ensures consistency and reliability across all data platforms, forming a solid foundation for analytical processes (Fan & Geerts, 2012).

The third stage addresses **data protection and compliance**, a critical component in modern data governance. According to Cavoukian and Jolly (2018), it is necessary for companies to implement robust security measures, ensure compliance with privacy regulations, and maintain data integrity. This stage is particularly crucial as organizations handle increasingly sensitive data and face stricter regulatory requirements, making it essential to establish comprehensive data protection frameworks while maintaining data usability for analysis.

The fourth stage concentrates on **data stewardship and continuous monitoring**, which ensures the long-term sustainability of data governance efforts. the stage of data stewardship and continuous monitoring implements systematic measurement frameworks to evaluate data quality metrics, monitor compliance with established governance policies, and track the effectiveness of data management practices (Dingre, 2023). The research results of Peng et al. (2015) also point out that effective data stewardship requires clear accountability structures and regular assessment of data lifecycle processes to maintain data integrity and value over time. This final stage creates a feedback loop that enables organizations to continuously refine their data governance practices while maintaining transparency and auditability of their data assets.

This systematic approach to data governance has been effectively implemented by leading organizations, such as JPMorgan Chase and Uber (Kanerika, 2024; Poonkuzhale, n.d.). JPMorgan Chase, one of the largest financial institutions in the world, has established a comprehensive data governance framework that focuses on three key aspects: enhancing data quality, ensuring security, and maintaining regulatory compliance (Jain et al., 2024). Their governance structure aligns data management practices with stringent regulatory requirements, which enables accurate financial reporting and improves decision-making processes. This robust approach has resulted in better risk management and increased customer trust through secure and reliable financial services.

In contrast, Uber has developed a flexible governance framework to manage data across its operations in over 70 countries, each with unique regulatory requirements (Atlan, 2023). Their approach employs a core platform for centralized data privacy and security management while also allowing for customizations and plugins to adapt governance policies based on the origin of each dataset. This enables Uber to maintain global data collection while ensuring compliance with local and regional laws across its diverse operational landscape.

The implementation of such robust governance frameworks significantly enhances data visualization capabilities through several key mechanisms. When organizations adopt comprehensive governance frameworks, they ensure data consistency and reliability, which are essential for creating accurate visual representations. This standardization of data formats and quality controls allows visualization tools to process and display information more effectively, minimizing errors and inconsistencies in the outputs. Additionally, proper governance ensures that data sources are well-documented and verified, enabling visualization teams to quickly access and understand the data they are working with. This leads to more efficient and accurate production of visual analytics. In short, this structured approach to data governance results in more trustworthy and actionable insights for decision-makers.

3.2. Ethics in Data Visualization

When it comes to data visualization, adhering to ethical principles is essential for maintaining integrity and ensuring the responsible presentation of information. This is especially important in the context of earthquake data visualization, as our visual outputs can directly influence emergency response and disaster management decisions. The ethical handling of such critical data not only affects the quality of decision-making but also impacts public safety and resource allocation. In this project, we focused on three fundamental ethical principles that guided our approach to data representation and analysis.

- **Accuracy of Data, Information, and Knowledge**: We prioritized data accuracy through thorough pre-processing, which included removing 337 duplicate records and systematically addressing missing values. Our alert level assignments were based on careful analyses of magnitude distributions to ensure an accurate representation of earthquake severity. This commitment to accuracy is crucial, as misleading visualizations could result in misguided emergency response decisions.
- **Accessibility of Data and Information**: Our visualizations were designed with accessibility in mind, enabling the Director of Emergency Response and Disaster Management to easily interpret and utilize the information. We implemented clear labelling, intuitive colour schemes, and appropriate scaling to make the figures comprehensible for various levels of data literacy. This approach supports effective decision-making by making complex earthquake data accessible to diverse audiences, from technical experts to general stakeholders.
- **Transparency of Data, Information, and Knowledge**: We maintained ethical standards of transparency by properly acknowledging that this dataset was provided within our academic unit scope. All data transformation processes were thoroughly documented, detailing the transition from the initial 1,137 records to the final cleaned dataset of 800 records. In cases where data was imputed or transformed, we provided appropriate context and explanations to prevent misinterpretation. We ensured transparency through clear documentation of our methodology, including our rationale for visualization choices and how they were created. This allows stakeholders to understand the origin of the data, how it was processed, and how conclusions were derived, which is essential for maintaining data integrity and supporting informed decision-making in emergency response scenarios.

4. CONCLUSION

In conclusion, the report has utilized comprehensive data visualization techniques to transform complex seismic data into actionable insights for emergency response and disaster management. By employing five visualizations using Tableau, we have highlighted crucial patterns in global earthquake activity, including geographical distributions along the Pacific Ring of Fire, seasonal patterns, and correlations between earthquake characteristics. Our analysis has yielded valuable predictive insights that can enhance decision-making in emergency response plans.

The visualization work is built upon a foundation in data governance principles and ethical considerations. By adhering to the four stages of the data analytics process and maintaining strict ethical standards, such as accuracy, accessibility, and transparency, we have ensured the reliability and integrity of our visual analytics.

The insights derived from this analysis can significantly improve emergency preparedness and response strategies. Key recommendations include:

1. Prioritizing resource allocation in high-risk zones identified through geographical analysis.
2. Implementing depth-specific response strategies based on correlations with magnitude.
3. Adjusting preparedness levels according to seasonal patterns.
4. Developing region-specific guidelines that take into account unique seismic profiles of countries and continents.
5. Establishing specialized tsunami response measures in high-risk areas.

5. REFERENCE

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: an analysis of the literature. *Journal of Decision System*, *25*(sup1), 64–75. https://doi.org/10.1080/12460125.2016.1187397

Atlan. (2023, May 25). 5 Data Governance Examples: Case Studies from Airbnb, Uber, and More. *Atlan*. https://atlan.com/data-governance-examples/#:~:text=Modern%20data%20culture-,How%20Uber%20set%20up%20a%20flexible%20approach%20to%20data%20governance,data%20privacy%20and%20security%20centrally.

Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles. *Lecture Notes in Computer Science*, 115–125. https://doi.org/10.1007/978-3-319-44421-5_9

Cavoukian, A., & Jolly, N. (2018). Embedding privacy and security to gain a competitive advantage. *Journal of Data Protection & Privacy*, *1*(4). https://doi.org/10.69554/ephc9497

Dingre, S. S. (2023). Exploration of data governance frameworks, roles, and metrics for success. *Journal of Artificial Intelligence & Cloud Computing*, 1–3. https://doi.org/10.47363/jaicc/2023(2)195

Fan, W., & Geerts, F. (2012). Foundations of Data quality Management. *Synthesis Lectures on Data Management*, *4*(5), 1–217. https://doi.org/10.2200/s00439ed1v01y201207dtm030

Idemudia, N. C., Ige, N. a. B., Adebayo, N. V. I., & Eyieyien, N. O. G. (2024). Enhancing data quality through comprehensive governance: Methodologies, tools, and continuous improvement techniques. *Computer Science & IT Research Journal*, *5*(7), 1680–1694. https://doi.org/10.51594/csitrj.v5i7.1352

Jain, A., Person, G., & Conroy, P. (2024, April 16). *How JPMorgan Chase built a data mesh architecture to drive significant value to enhance their enterprise data platform*. Amazon Web Services. https://aws.amazon.com/blogs/big-data/how-jpmorgan-chase-built-a-data-mesh-architecture-to-drive-significant-value-to-enhance-their-enterprise-data-platform/

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Kanerika. (2024, July 19). Data Governance Examples: How Top Companies Manage Their Data. *Kanerika*. https://kanerika.com/blogs/data-governance-examples/

N. Gudivada, V., Apon, A., & Ding, J. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *International Journal on Advances in Software*, *10*(1 & 2). https://personales.upv.es/thinkmind/dl/journals/soft/soft_v10_n12_2017/soft_v10_n12_2017_1.pdf

Peng, G., Privette, J. L., Kearns, E. J., Ritchey, N. A., & Ansari, S. (2015). A unified framework for measuring stewardship practices applied to digital environmental datasets. *Data Science Journal*, *13*(0), 231–252. https://doi.org/10.2481/dsj.14-049

Poonkuzhale, K. (n.d.). *Data Governance in action: Real examples of analytics success*. Performix. https://www.performixbiz.com/blog/data-governance-in-action-real-examples-of-analytics-success