BUSA8090 - Session 2 2024

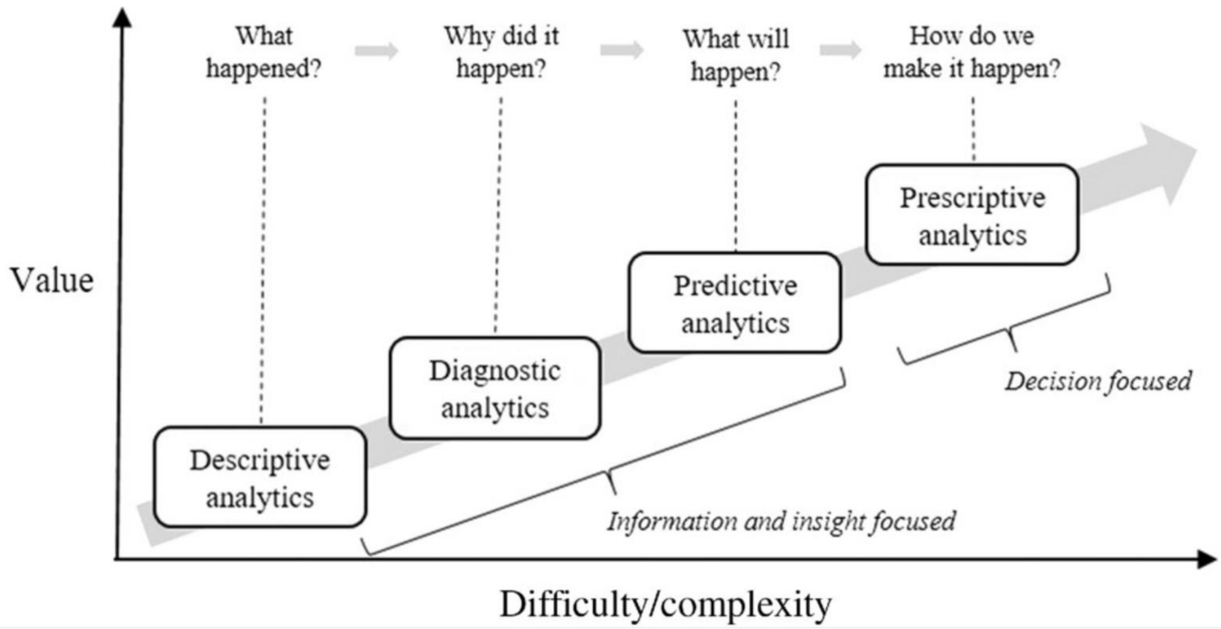
Assignment 2 - Part B Report

Author: Khuat Son Tra Nguyen (48144134)

Table of Contents

I/ Data Analytics Process and Governance

The data analytics process comprises four fundamental stages that form the backbone of effective data governance. These stages - descriptive, diagnostic, predictive, and prescriptive analytics - progress in both value and complexity (Houtmeyers et al., 2021) - see image 1. Descriptive analytics examine historical data, while diagnostic analytics investigates the causes behind past outcomes. Predictive analytics forecasts future trends, and prescriptive analytics provides actionable recommendations. Wolniak & Grebsk (2023) also identifies a fifth "real-time" stage following the descriptive phase, enhancing the analytics process's dynamism.



(Houtmeyers et al., 2021)

In large organizations, data governance has become crucial for managing data quality and value. It requires effective collaboration between business and IT departments to establish clear roles and responsibilities (Cheong & Chang, 2007). Organizations typically adopt either bottom-up or top-down approaches to data governance, depending on their specific context (B. Otto, 2011). The DAMA framework stands as a widely recognized data governance model that addresses key decision domains, including data principles, quality, metadata, access, and lifecycle management (Quinto, 2018).

Effective data governance significantly enhances data visualization capabilities. During the COVID-19 pandemic, Sandia National Laboratories demonstrated this by implementing a centralized data strategy to support visualization development for workforce protection (Harris et al., 2021). Data visualization transforms complex datasets into comprehensible visual representations, facilitating pattern identification and trend analysis (Bisht, 2024). When supported by robust governance frameworks, visualization tools can significantly improve decision-making processes and stakeholder communication.

II/ Ethical Principles in Data Visualization

Overall, ethical data visualization principles such as transparency, accuracy, and privacy — are crucial as organizations increasingly rely on visual data communication. Research shows deceptive visualization techniques lead to significant misinterpretation, even with explanatory text (O'Brien, 2017). These principles are especially vital for sensitive or decision-critical information.

1. Transparency

Transparency serves as a fundamental ethical principle in data visualization, particularly crucial in building trust and credibility in news media (Kennedy et al., 2020). While it functions differently for various stakeholders - Linked Data increases machine transparency, visualization enhances human understanding (Degbelo, 2017) - transparency itself acts as a pro-ethical condition enabling other ethical practices rather than being a standalone principle (Turilli & Floridi, 2009). Organizations can demonstrate their ethical commitment by disclosing information about how ethical principles are embedded in their visualization design practices.

2. Accuracy

Accuracy in data visualization encompasses multiple aspects, from color integrity to editorial decisions. Research emphasizes the importance of accessible and accurate color techniques, including color-blind friendly palettes, to ensure inclusive data representation (Crameri & Hason, 2024). In journalistic contexts, ethical considerations must span the entire process from data acquisition to presentation (Diakopoulos, 2018). In addition, Dodd (2020) addresses the multiple ethical considerations in visual research methods, emphasizing the importance of integrity in data generation and analysis.

3. Privacy

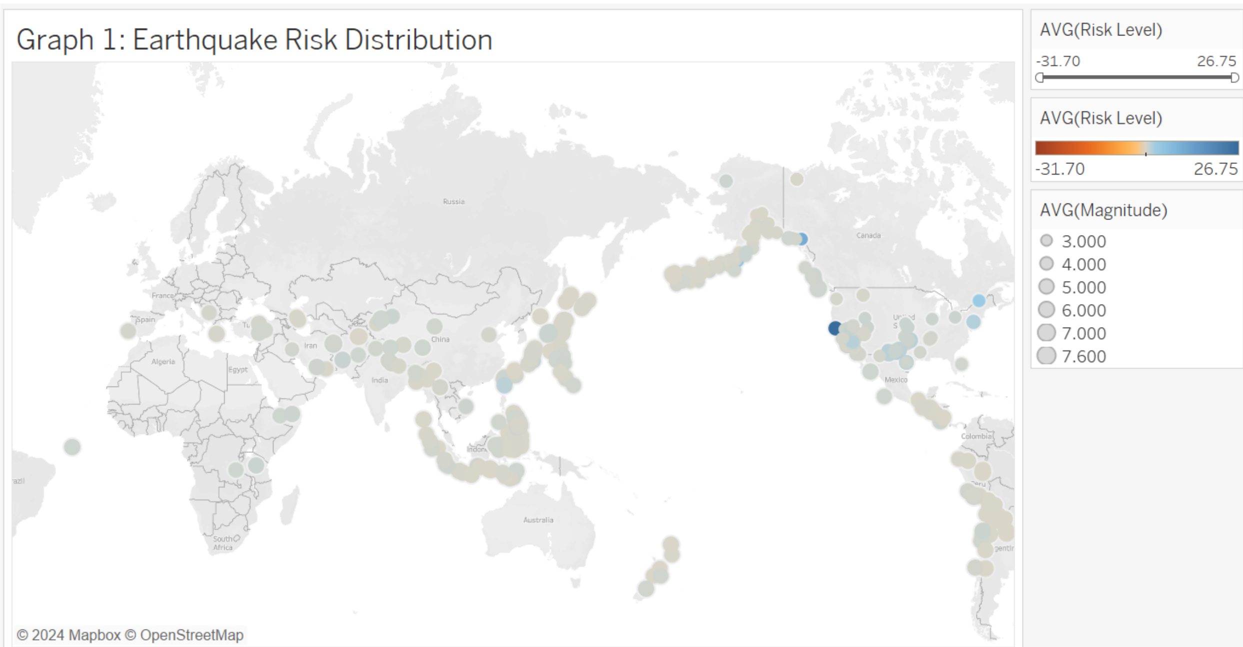
Privacy considerations in data visualization have become increasingly critical, particularly when handling sensitive information. Researchers have developed various privacy-preserving visualization techniques, including k-anonymization, centroid replacement, and stochastic noise addition (Avraam et al., 2021). These methods aim to maintain data utility while protecting individual confidentiality. Statistical agencies face the ongoing challenge of balancing data release with disclosure risk, necessitating methods that limit disclosure while maximizing information sharing (Fienberg, 2000).

Recent studies have confirmed that deceptive visualization techniques can lead to misinterpretation, even when accompanied by explanatory text (O'Brien, 2017). This finding underscores the importance of ethical visualization practices and proper education in the field. As the domain evolves, emerging trends include privacy-enhancing technologies, regulatory changes, and AI integration in data anonymization (Patel, 2024), emphasizing the growing need for responsible and ethical data visualization practices that protect individual privacy while maximizing data insights.

III/ Part A Visualization’s Interpretation

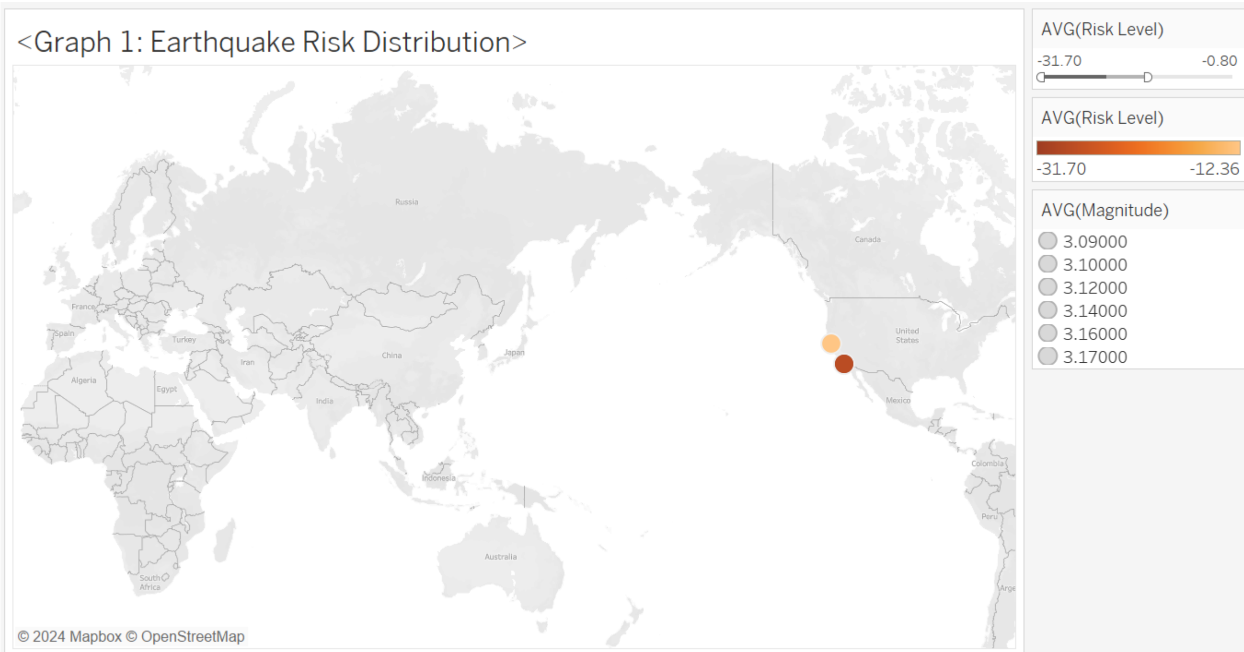
Insight 1: Risk Distribution

The visualization effectively displays global earthquake risk patterns by employing a color-coded system that factors in both magnitude and depth through the Risk Level formula (Magnitude × 1/Depth). This mapping approach proves particularly valuable for Emergency Response and Disaster Management directors in prioritizing resource allocation and response strategies.



Graph 1: Earthquake Risk Distribution

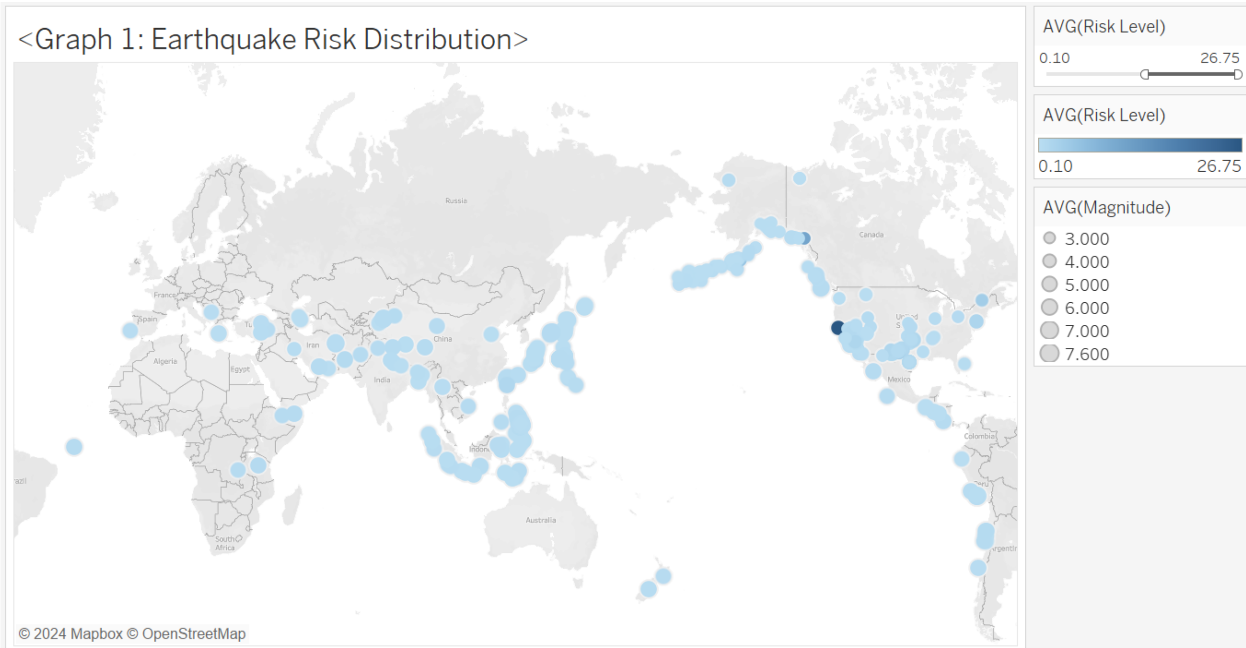
The most critical areas, depicted in orange/brown tones, are concentrated in seismically active regions such as **Western United States (particularly California and Texas)**, parts of Alaska, and selected zones in Mexico. These locations experience extremely shallow earthquakes that pose the highest risk due to their proximity to the surface, demanding immediate emergency response protocols and heightened preparedness measures.



The visualization reveals a substantial "moderate risk" zone, shown in neutral/ gray tones (around 0.00), encompassing the Pacific Ring of Fire, including the Indonesian archipelago, Philippines, Japan, and New Zealand. This zone extends to the Mediterranean Basin (Turkey, Greece, and Italy) and Central Asian regions (Iran and Afghanistan). While these areas experience shallow earthquakes, their slightly deeper occurrence provides marginally more response time compared to surface events.

Of particular interest for strategic planning are the blue-coded areas, indicating deeper seismic activity in regions such as the Peru-Chile border, Bolivia, the Afghanistan-Pakistan border

(Hindu Kush region), and the Fiji-Tonga region in the South Pacific. While these areas experience frequent seismic activity, their greater depth typically results in reduced surface impact, allowing for more measured response strategies.

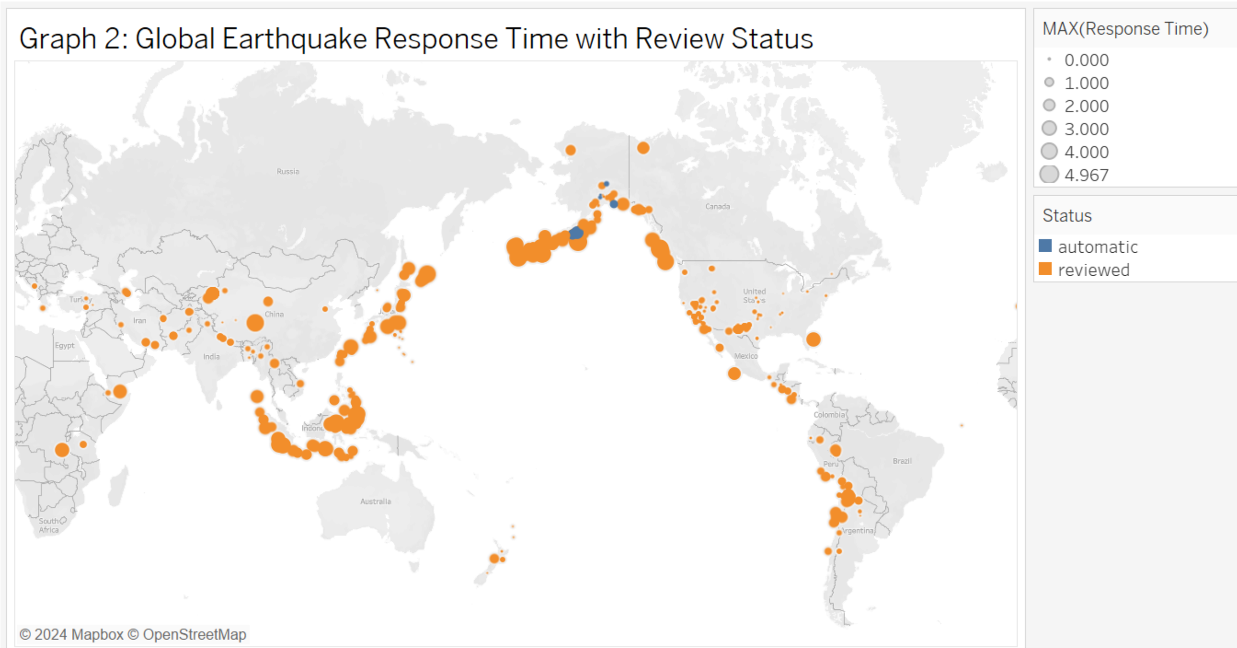


Recommendations for Emergency Management:

1. Prioritize resource deployment to high-risk orange/brown zones
2. Develop graduated response protocols based on risk levels
3. Establish regional response centers in strategic locations near high-risk areas
4. Plan preventive measures and infrastructure reinforcement in areas of consistent seismic activity

This data-driven approach to risk assessment and visualization provides emergency response directors with a clear framework for decision-making, resource allocation, and strategic planning in earthquake preparedness and response operations.

Insight 2: Global Earthquake Response Time with Tsunami



This visualization maps global earthquake response times (calculated as Distance to Nearest City/60 km/h, shown by circle size ranging from 0 to 4.967 hours) and tsunami occurrence (indicated by blue shading intensity from 0 to 4 events), providing crucial insights for emergency management planning. Overall, the map highlights the relationship between response accessibility and tsunami risk zones.

Response Time Distribution. The visualization reveals varying emergency response times globally, ranging from immediate (0 hours) to nearly 5 hours in remote locations. The Pacific Ring of Fire demonstrates particularly concentrated activity, with response times varying significantly based on proximity to urban centers. Coastal regions show moderate response times, complicated by offshore access challenges.

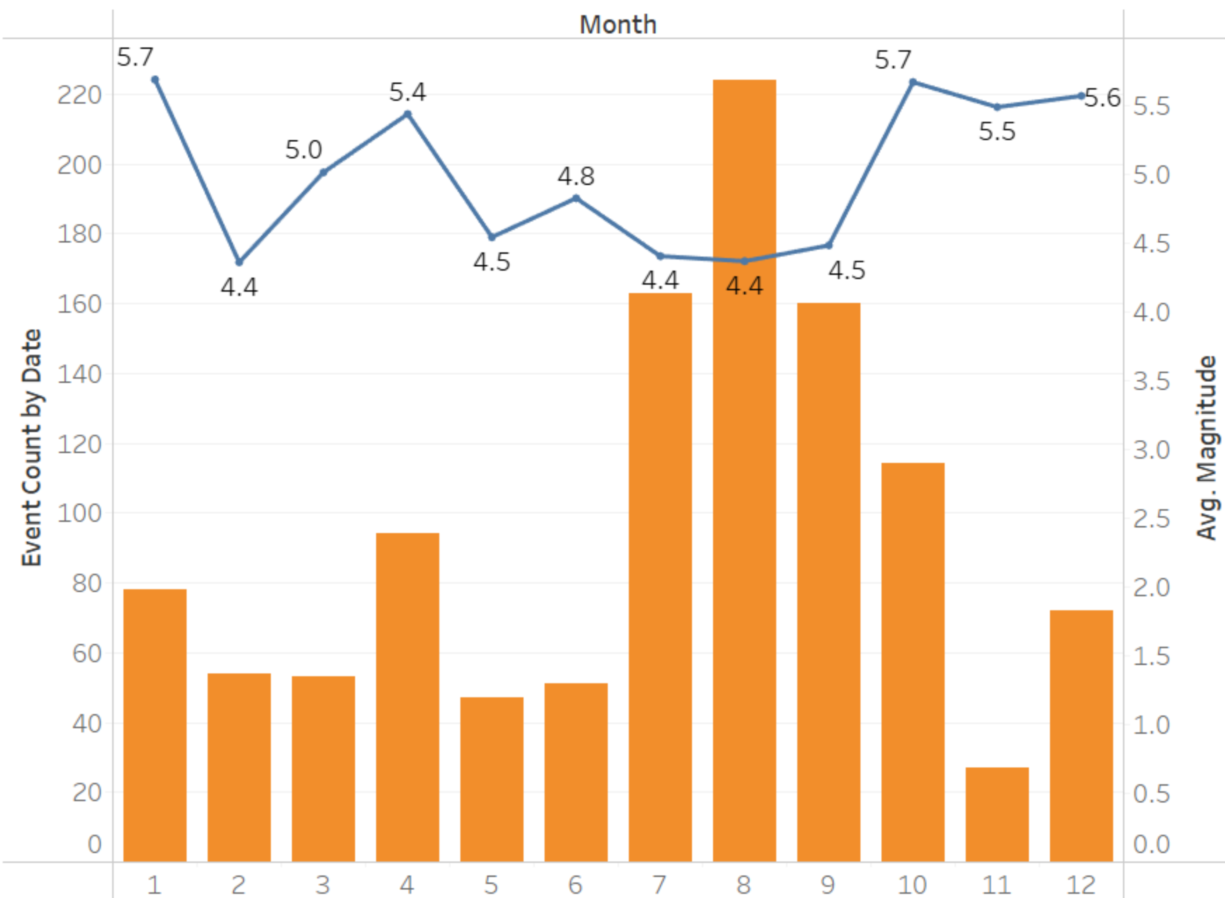
Tsunami Correlation and Critical Zones. The data reveals a crucial pattern where regions with higher tsunami frequency (indicated by darker blue shading) often coincide with longer response times, particularly along the Alaskan coastline and parts of East Asia. This correlation is especially concerning as these areas require rapid response capabilities due to the compound risk of earthquakes and subsequent tsunamis. The visualization shows tsunami events are most frequent in areas with response times ranging from 2-4 hours, creating a critical window for emergency actions.

Recommendations for Emergency Management:

1. Prioritize establishing rapid response units in coastal areas where high tsunami risk overlaps with extended response times
2. Develop specialized tsunami early warning protocols for regions showing darker blue shading, particularly in Alaska and the Western Pacific
3. Create dedicated response strategies for areas where tsunami risk coincides with response times exceeding 3 hours
4. Position emergency resources strategically along coastlines with high tsunami frequency to reduce current response delays
5. Implement enhanced monitoring systems in regions where tsunami risk and extended response times overlap

Insight 3 - Monthly Patterns

Graph 3: Monthly Earthquake Frequency and Magnitude



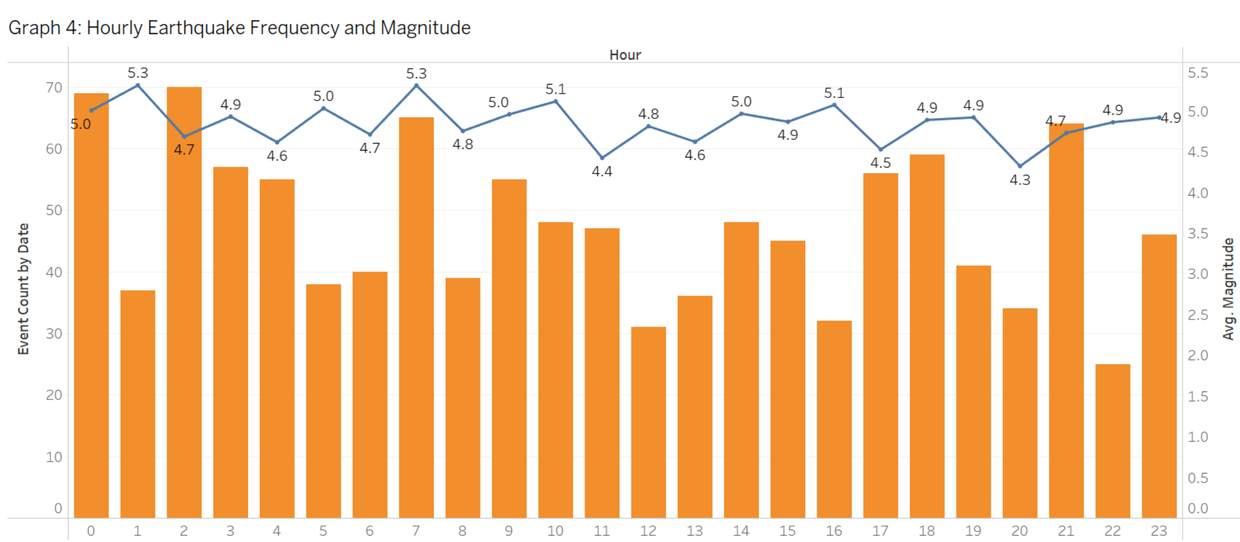
Event Frequency Pattern. Monthly earthquake frequency shows distinct seasonal variations. Peak activity occurs mid-year (July-September), with August recording the highest at 224 events. The first half of the year maintains relatively low activity (40-90 events), while year-end (October-December) shows moderate levels, dropping to 27 events in November. Total annual events: 1,137.

Magnitude Trends. Average magnitudes demonstrate an inverse relationship with frequency. Lower-frequency periods often correspond with higher magnitudes (5.4-5.7). July-September, despite high frequency, shows lower average magnitudes (4.4-4.5). Year-end months maintain consistently high magnitudes (5.5-5.6), requiring sustained high-level response capabilities.

Recommendations for Emergency Management:

1. Scale up resources and staffing July-September for high-frequency response
2. Maintain advanced response capabilities October-December for higher-magnitude events
3. Implement seasonal staffing patterns aligned with frequency peaks
4. Develop dual response protocols:
    - High-volume, lower-magnitude events (mid-year)
    - Lower-frequency, high-magnitude events (year-end)
5. Schedule major maintenance and training during low-activity periods (May-June)

Insight 4 - Hourly Patterns



Event Frequency Pattern. Earthquake events show distinct 24-hour distribution with three notable peaks: early morning (midnight-2AM, ~70 events), mid-morning (7AM, 65 events), and evening (6PM-9PM, 55-65 events). The lowest activity occurs during early afternoon (12PM-2PM, 30-35 events). This pattern suggests a cyclical nature of seismic events, with heightened activity during off-peak hours.

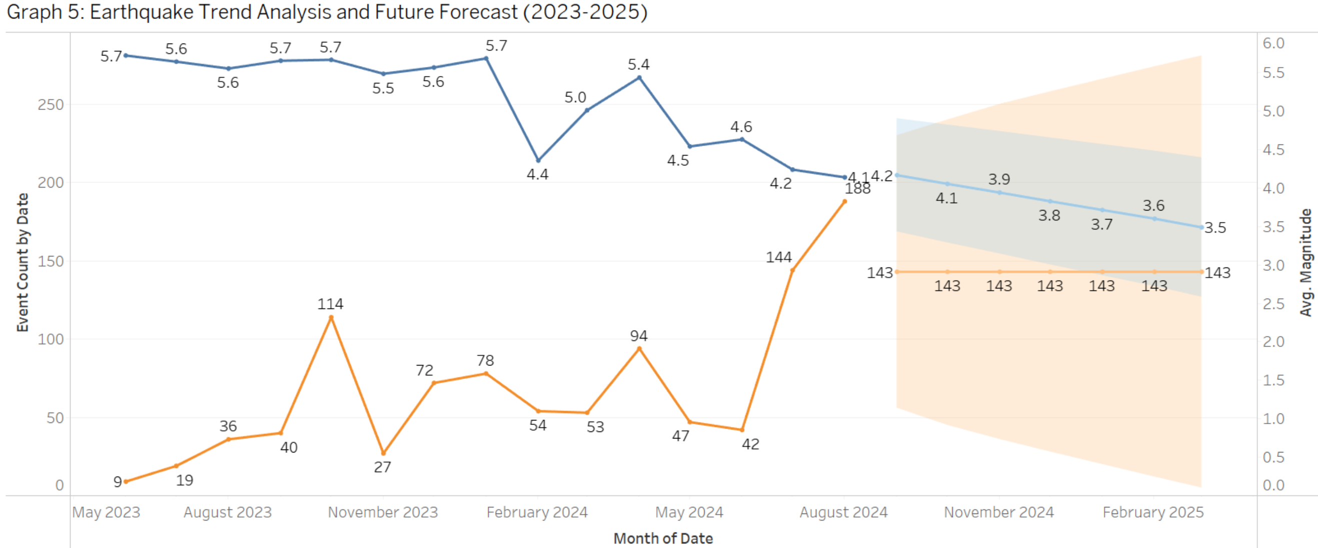
Magnitude Trends. Average magnitudes fluctuate between 4.3 and 5.3, showing key spikes at 1AM (5.3), 7AM (5.3), and 10AM (5.1). Notably, periods of lower frequency often coincide with higher magnitudes, particularly during morning hours. Evening periods (6PM-11PM) maintain consistent moderate magnitudes around 4.7-4.9.

Recommendations for Emergency Management:

1. Implement three-tiered staffing model:
    - Peak staffing: midnight-2AM and 6PM-9PM (high frequency)
    - Enhanced staffing: 7AM-10AM (high magnitude)
    - Standard staffing: 12PM-5PM (lower activity)
2. Position specialized response teams during high-magnitude morning hours
3. Maintain surge capacity during early morning peaks
4. Schedule shift changes outside peak activity periods
5. Concentrate training and maintenance during 12PM-2PM window

Insight 5 - Earthquake Trend Analysis and Future Forecast (2023-2025)

Graph 5: Earthquake Trend Analysis and Future Forecast (2023-2025)



Event Frequency Analysis. Historical earthquake frequency shows significant volatility, starting from 9 events (May 2023) and escalating to peak at 188 events (August 2024). The data reveals a clear upward trend in frequency, with notable spikes in November 2023 (114 events) and May 2024 (94 events). Forecasts through February 2025 predict a stabilization at 143 events per period, though with widening uncertainty margins.

Magnitude Trend Analysis. Average earthquake magnitude demonstrates a concerning downward trend. From a stable period of 5.5-5.7 magnitude through late 2023, values declined sharply to 4.4 in early 2024, briefly recovered to 5.4, then continued declining. Projections indicate further decrease from 4.1 to 3.5 by February 2025.

Recommendations for Emergency Management:

1. Restructure response protocols to handle higher-frequency, lower-magnitude events
2. Maintain resources for consistent response to ~140 events per period
3. Develop scalable deployment strategies to address the increasing frequency
4. Consider redistributing resources from high-magnitude response capabilities to multiple simultaneous lower-magnitude event management
5. Implement monthly forecast reviews to adjust resource allocation based on actual trends

References

Avraam, D., Wilson, R., Butters, O., Burton, T., Nicolaides, C., Jones, E., Boyd, A., & Burton, P. (2021). Privacy preserving data visualizations. *EPJ Data Science*, *10*(1). https://doi.org/10.1140/epjds/s13688-020-00257-4

Bisht, D. (2024). The impact of data visualization technique on decision making in business analytics Vaibhav,. *Indian Scientific Journal of Research in Engineering and Management*, *08*(04), 1–5. https://doi.org/10.55041/ijsrem32049

Cheong, L. K., & Chang, V. (2007). The Need for Data Governance: A Case Study. *ACIS*. https://www.semanticscholar.org/paper/The-Need-for-Data-Governance%3A-A-Case-Study-Cheong-Chang/682212e08fd1711b955dd40b515294ca724529c4

Crameri, F., & Hason, S. (2024). Navigating color integrity in data visualization. *Patterns*, *5*(5), 100972. https://doi.org/10.1016/j.patter.2024.100972

Degbelo, A. (2017). Linked data and visualization: Two sides of the transparency coin. *UrbanGIS'17: Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*. https://doi.org/10.1145/3152178.3152191

Diakopoulos, N. (2018). Ethics in Data-Driven Visual Storytelling. In N. Henry Riche, S. Carpendale, N. Diakopoulos, & C. Hurter (Eds.), *Data-Driven Storytelling* (pp. 233–248). A K Peters/CRC Press. https://doi.org/10.1201/9781315281575-10

Dodd, S. (2020). *Ethics and Integrity in Visual Research Methods*. Emerald Publishing Limited.

Fienberg, S. E. (2000). *Confidentiality and data protection through disclosure limitation: Evolving principles and technical advances*. https://api.semanticscholar.org/CorpusID:166955253

Harris, R., Jones, T., Flores, M., & Bustamante, D. (2021). COVID-19, An Exercise in Data Governance at Sandia National Laboratories. In *OSTI OAI (U.S. Department of Energy Office of Scientific and Technical Information)*. Office of Scientific and Technical Information. https://doi.org/10.2172/1808095

Houtmeyers, K. C., Jaspers, A., & Figueiredo, P. (2021). Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics. *International Journal of Sports Physiology and Performance*, *1*(aop), 1–5. https://doi.org/10.1123/ijspp.2020-0958

Kennedy, H., Weber, W., & Engebretsen, M. (2020). 11. Data visualization and transparency in the news. In M. Engebretsen & H. Kennedy (Eds.), *Data Visualization in Society* (pp. 169–186). Amsterdam University Press. https://doi.org/10.1515/9789048543137-015

O'Brien, S., & Laure, C. (2017). Interpretations of Data in Ethical vs. Unethical Data Visualizations. In *ProQuest Dissertations and Theses* (No. 10681043, Arizona State University; p. 83). ProQuest Dissertations & Theses Global. https://simsrad.net.ocs.mq.edu.au/login?url=https://www.proquest.com/dissertations-theses/interpretations-data-ethical-vs-unethical/docview/1976992451/se-2

Otto, B. (2011). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. *Communications of the Association for Information Systems*, *29*. https://doi.org/10.17705/1cais.02903

Patel, P. (2024). Advancing Data Visualization: Integrating Privacy - Preserving Techniques in the Era of Big Data. *International Journal of Science and Research (IJSR)*, *13*(2), 1228–1232. https://doi.org/10.21275/sr24215084635

Quinto, B. (2018). Big Data Governance and Management. In *NextGeneration Big Data: A Practical Guide to Apache Kudu, Impala, and Spark* (pp. 495–506). Apress. https://doi.org/10.1007/9781484231470_11

Turilli, M., & Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, *11*(2), 105–112. https://doi.org/10.1007/s10676-009-9187-9

Wolniak, R., & Grebski, W. (2023). The five stages of business analytics. *Zeszyty Naukowe*, *2023*(178). https://doi.org/10.29119/1641-3466.2023.178.41