Disasters in Context: World Disaster Events Mapping and Analysis

Samuel Harris

The George Washington University

DATS 4001-Data Science Capstone

Dr. Edwin Lo

May 6, 2024

1. Abstract

To reduce the impact of unexpected disaster events, it is imperative to prepare accordingly for disasters as if they are going to occur. This project seeks to help disaster management officials mitigate disasters' impact by providing visualizations and a risk score that highlight historical disaster trends. The project enhances current disaster event data by placing disaster events in the context of the development level of the country where the disaster occurred. By combining development indicators with disaster data, this analysis provides a useful way to view disaster types comparatively by country. The main results include temporal and spatial trends, mapping by development score, and mapping by risk score. Overall, this report gives a detailed overview of world disasters by development and risk scores. It could be further expanded in the future with a more narrow analysis of certain disaster types.

2. Introduction

- **Background**

Ranging from industrial explosions to volcanic eruptions, disaster events can be the result of human error or occur naturally. Over the past century, the world has seen a precipitous decline in high death toll disasters. This decline is mostly due to technological advancements allowing for better response and recovery efforts. Still, since 2000 over 1.7 million deaths have occurred from disaster events (Centre for Research on the Epidemiology of Disasters [CRED], 2024). With accelerating climate change, disaster event severity only stands to increase (UN, 2023). The relatively high death toll in the twenty-first century highlights a fundamental concern about disasters. Most disasters are unpredictable and thus unpreventable. Disasters like earthquakes and floods will always occur, and no one can accurately predict the date and location of these disasters well in advance. However, despite disasters being almost impossible to entirely prevent due to their inherent randomness, disasters' impact on populations can be mitigated through effective response, recovery, and preparedness efforts. Leaders often overlook preparing for disasters that may not happen, but adequate preparation is crucial to reduce human suffering when a disaster occurs unexpectedly.

- **Objective**

The objective of this project is to combine world disaster event data with World Bank country indicator data to place historical disaster events in a given country's developmental context. Understanding historical disasters by their type, impact, and development level is essential to gain information about how different disasters affect different countries. From this information, officials can set goals for disaster management and adequately assess disaster risk comparatively by country. The goal of this analysis is to mitigate disasters' impact by increasing awareness about disasters and providing relevant resources to disaster management officials.

3. Datasets

- **EM DAT Data**

The first dataset used in this project is the Emergency Events Database (EM-DAT) by the Centre for Research on the Epidemiology of Disasters. EM-DAT contains tabular data with 15,573 rows of disaster events around the world from 2000 onward. The disaster events included in EM DAT are technological or natural disasters meeting the criteria of having more than 10 deaths, more than 100 affected, or a call for international assistance. Technological disasters are defined as unintentional accidents and do not include acts of war or terrorism (CRED, 2023a). There are 46 columns in the dataset ranging from disaster type, country, location name, start year and month, latitude and longitude, magnitude, total damage, number of affected, and total deaths. The dataset is comprehensive in providing basic information about each world disaster event.

The generality of EM DAT sometimes comes at the cost of the quality of the data. Biases relevant to the dataset relate to time, thresholds, geography, and accounting. Time biases include differences in reporting disaster events over time. For this analysis, only disaster events since 2000 were used to reduce time bias. Threshold biases relate to news organizations reporting disaster impact variables higher than their true value to get more attention. Geographic biases concern the differences in reporting between countries for disaster events. Finally, accounting biases occur through differences in the reporting of certain impact variables. For example, the impact variable of costs associated with a disaster event is often missing in Africa due to the lack of insurance coverage (CRED, 2023b). Despite the biases mentioned here, EM DAT is still one of the best sources for world disaster event data. The analysis in the rest of the report acknowledges and attempts to mitigate the biases discussed here as much as possible.

- **World Bank Data**

The World Bank's World Development Indicators Database contains relevant indicator statistics by country since 2000. Development indicators used in this analysis include GDP per capita, infant mortality rate, life expectancy, electricity percentage, and tonnes of CO_2 emitted (World Bank, 2024). These indicators were chosen because they give a relevant overview of a given country's development while also being widely available for almost every country since 2000. Combining the World Bank data with the disaster data places each disaster event in the context of the given country's economic, technological, and health conditions.

4. Methods

- **Merging the Datasets**

The foundation of this project was merging EM DAT with the World Bank development indicator data. The process was relatively simple but important since this merging places disaster events in a given country's developmental context. The merging was done by country and year and resulted in each disaster event in EM DAT being associated with the relevant country development indicators. Notable countries that did not have World Bank indicator statistics included Taiwan, Puerto Rico, and North Korea. These countries were excluded from any analysis involving development. The combination of the datasets allowed for the calculation of the scores discussed next.

- **Development Score**

Development score captures a country's overall development in a given year compared to the rest of the world. By merging the EM DAT and World Bank development data, values for a country's GDP per capita, infant mortality rate, life expectancy, electricity percentage, and tonnes of CO2 emitted were applied to each disaster event. Next, by calculating the decile for each variable and adding them together, an overall development score was created for each country and year on a scale from 5-50. Principal component analysis was able to confirm initial intuition on how to add the variables together into one composite indicator for development (infant mortality was inverted for the sum). The equation is as follows, where D is the decile function:

$$\begin{aligned}\text{Development Score} = {} & D(\text{Life Expectancy}) + D(\text{Infant Mortality, decreasing}) \\ & + D(\text{Carbon Dioxide Emitted}) + D(\text{Electricity Percentage}) \\ & + D(\text{GDP Per Capita})\end{aligned}$$

The development score was then binned into 3 categories: least developed, developing, and developed. These categories were created by the bins 5-20, 20-35, and 35-50 respectively. Binning into these categories is meant to provide an easy interpretation of development score for the reader.

- **Risk Score**

Risk score captures a country's degree of risk of a certain disaster type based on historical disaster events in this country and the country's development. The score is directly calculated from historical data in both the EM DAT and World Bank datasets. The calculation itself is modeled after the Risk Index used by FEMA for counties in the US (FEMA, 2024). The formula for FEMA's index is as follows:

$$\text{Risk Index} = \frac{\text{Expected Annual Loss} \times \text{Social Vulnerability}}{\text{Community Resilience}}$$

The calculation for risk score in this project is slightly different due to the variables that are available for countries. Specifically, the calculation is additive and not multiplicative because of the skewed distribution of some of the variables. The risk score calculation is as follows:

$$\text{Risk Score} = .5 \times \text{Impact} + .25 \times \text{Inverted Development Score} + .25 \times \text{Frequency}$$

The variables used in this calculation are aggregate numbers grouped by country and disaster type. “Impact” represents the mean number of total deaths and total affected, “Inverted Development Score” is the previous score discussed inverted, and “Frequency” is the count of disaster events relative to other countries. All these variables have been normalized to be on a 0-100 scale. A higher weight of .5 is given to “Impact” to reflect the importance of deaths from a disaster event. Furthermore, “Inverted Development Score” and “Frequency” contribute equally to the score with a weight of .25 each.

Ultimately, the variables presented in this project’s risk score seek to model the variables of “Expected Annual Loss” and “Social Vulnerability” in FEMA’s Risk Index. The risk score could likely be improved with better data, and the weights could also be adjusted based on how to classify importance among variables. Still, the risk score used in this project offers a useful comparative statistic for a disaster type’s degree of risk among countries.

5. Results

- **Spatial and Temporal Trends**

The first main results of this project come from the spatial and temporal trends of disasters in EM DAT since 2000. Understanding these trends provides a baseline knowledge of disaster events before analyzing them by development.

Figure 1 provides an overview of disaster deaths since 2000 by region. The years 2004, 2008, and 2010 stand out as high death toll years. These deadly years turn out to be the result of outlier disaster events like the 2004 Boxing Day Tsunami, the 2008 Cyclone Nargis, and the 2010 Haiti Earthquake. The figure shows that these major outlier death events significantly impact overall death toll numbers. Figure 2 provides an interactive map to view total deaths by country with a year slider.

To avoid outliers, Figure 3 uses box and violin plots to view the typical value of impact variables in geographic regions. The violin plot of costs per disaster event grouped by region shows that the Americas and Europe have the most costs in a typical disaster event. This result is influenced by the bias in EM DAT whereby insured countries are more likely to report exact disaster cost statistics. Thus, the result is disregarded. The figure also provides total affected and total deaths box plots. Asia and Africa represent the regions with the most deaths in a median disaster event demonstrating a clear spatial trend. In Figure 4, the spatial death trends among median disasters are further broken down by subregion. The figure highlights Southern Asia, Sub-Saharan Africa, and Latin America and the Caribbean as subregions with high median death tolls.

Next, Figure 5 groups disaster counts by month and disaster type to provide an understanding of disasters by season. From the figure, floods and storms are the most common disaster types occurring more in the northern summer months of July and August. Disasters overall are very consistent by month with no disaster type standing out in a particular month. Finally, Figure 6 describes disasters by their duration in time. This stands to be useful to see if most disasters quickly happen, like earthquakes and tornadoes, or more slowly evolve, like droughts and epidemics. The figure displays that most disasters start and end in one day. Short, high-impact events are common for disasters in general.

- **World Bank Indicator Analysis**

The results of merging disaster event data with the World Bank indicator data are discussed in this section. Disaster events are assessed by development score and countries are further assigned a risk score by disaster type.

In Figure 7, the top 20 highest death toll disasters are mapped and grouped by development type. Using geocoding, the map shows the exact location where the disaster event occurred. From the map, numerous disease outbreaks in Africa have led to many deaths. This is in contrast to other developed countries where most of the top disasters by death are from heat waves. Heat waves are often under-reported in less developed countries which is likely the reason why heat waves in places like Africa are not observed in the data. Moreover, Figure 8 is a line graph showing disaster deaths over time grouped by development type. Again, outliers can be observed in the graph, but this time it is clear that the major outliers all occurred in the least developed countries.

Moving on, Figure 9 provides an interesting comparison between the distribution of disaster types between development levels. Floods are more significant in less developed countries, but the visualization overall seems to support the idea that disasters occur at roughly the same rates no matter the development type. This makes some intuitive sense given the somewhat random nature of disasters. The last figure concerning development, Figure 10, shows the median deaths in a disaster event by development score. The figure displays a clear downward trend for deaths in a median disaster event as the development score increases. So, less developed countries have higher median death tolls in addition to outlier death events.

Finally, Figure 11 is the risk score map which shows the highest disaster type by risk score in every country. The map mainly answers two questions: What disaster type has the highest risk in a given country, and how does the risk compare with other countries? Thus, the map is useful to locate basic information about the worst disaster types in a given country in addition to comparison between countries. Furthermore, Table 1 shows the top 20 risk scores overall. Countries like China and India have multiple disaster types with high risk scores which is a factor that is not directly displayed in the map. Ultimately, earthquakes as a disaster type seem to dominate risk in many countries. This is likely due to the magnitude of earthquakes allowing them to be high-impact events.

6. Conclusion

- **Limitations**

Limitations in this project concern the datasets and methods. For EM DAT, the dataset's generality can hold back some deeper patterns about specific disaster types. The data is also gathered from wildly different sources which can lead to the biases previously mentioned in the datasets section. To expand on the analysis in this project, it is recommended to dive deeper into specific disaster types and perhaps find a better dataset that better suits a certain disaster type's data characteristics.

Furthermore, the methods in this project could likely be improved. Specifically, other development score calculations could be applied to disaster events that may produce more accurate development measures. Additionally, the risk score calculation could be improved with a new statistic measuring community resilience. The risk score in this project mostly takes into account only impact and social vulnerability, a resilience statistic would improve the score significantly.

- **Implications and Future**

The analysis and mapping of disasters in this project demonstrated many trends in world disaster events. The first obvious conclusion of this project is the presence of outlier deadly disaster events. These events have a high impact, often causing many times the deaths of all total disaster events in other years. The outliers also mostly occur in less developed countries. Another conclusion is the spatial trends in the data. Specifically, disasters in Africa and Asia have a higher overall median death toll in disaster events. Temporal trends also highlight that disasters are usually one-day events. By looking at disasters grouped by development, it was determined that disaster types occur at roughly the same ratios in each development type. However, despite similar ratios, the median death toll was shown to be at least three times higher in the least developed countries compared to the developed countries. Finally, this project demonstrated the value of calculating a risk score for countries. The score allows for both the identification of risky disaster types in countries and the comparison between countries.

Ultimately, this project was completed to mitigate the impact of disasters by helping officials in disaster management. The visualizations and risk scores presented in this analysis will hopefully be used to identify countries that require more assistance with disaster management efforts. In addition, this project will help spread awareness about the importance of disaster preparedness. While this project helps mitigate disasters' impact, the lack of current research being done to understand past disasters and their trends is concerning. As climate change results in more intense disasters, more research must be done to understand past disasters so that humanity can be adequately prepared for unexpected future disasters.

7. Figures and Tables

Figure 1[1]

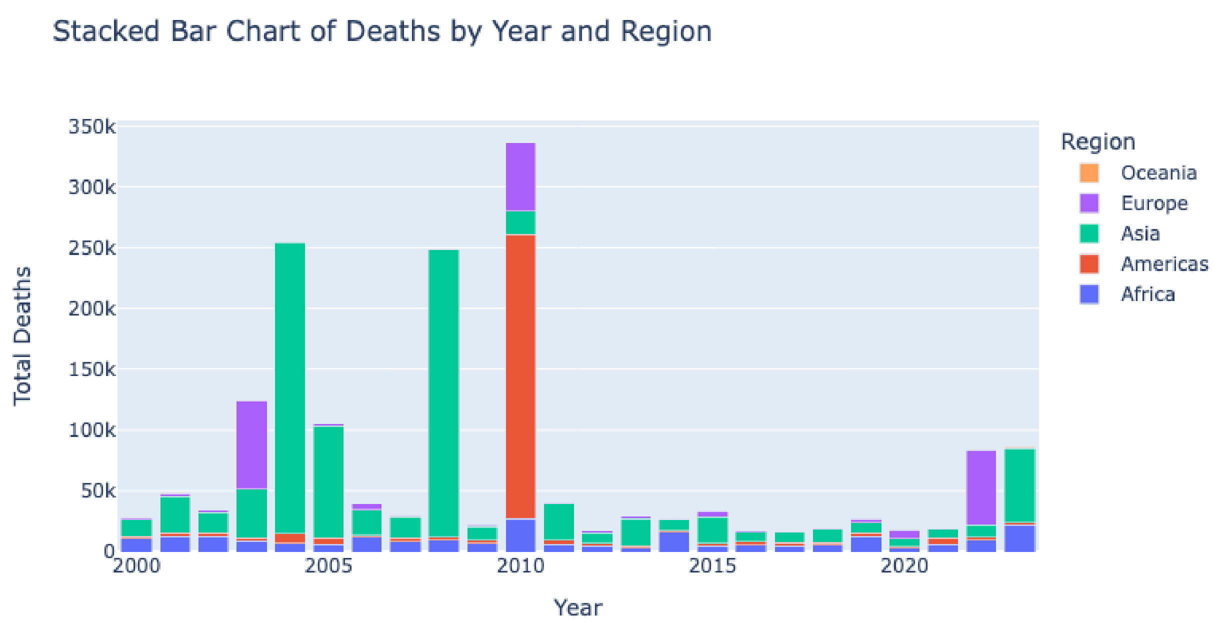
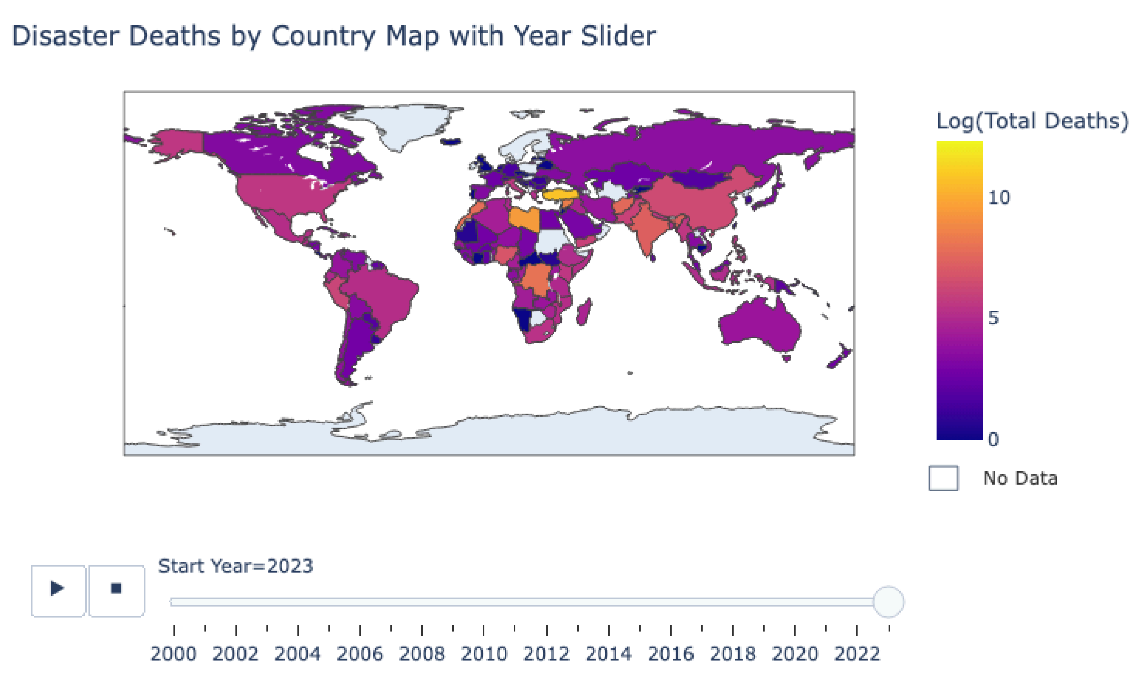


Figure 2



[1] Please note that all visualizations can be found at **https://disastersanalysis.streamlit.app**

Figure 3

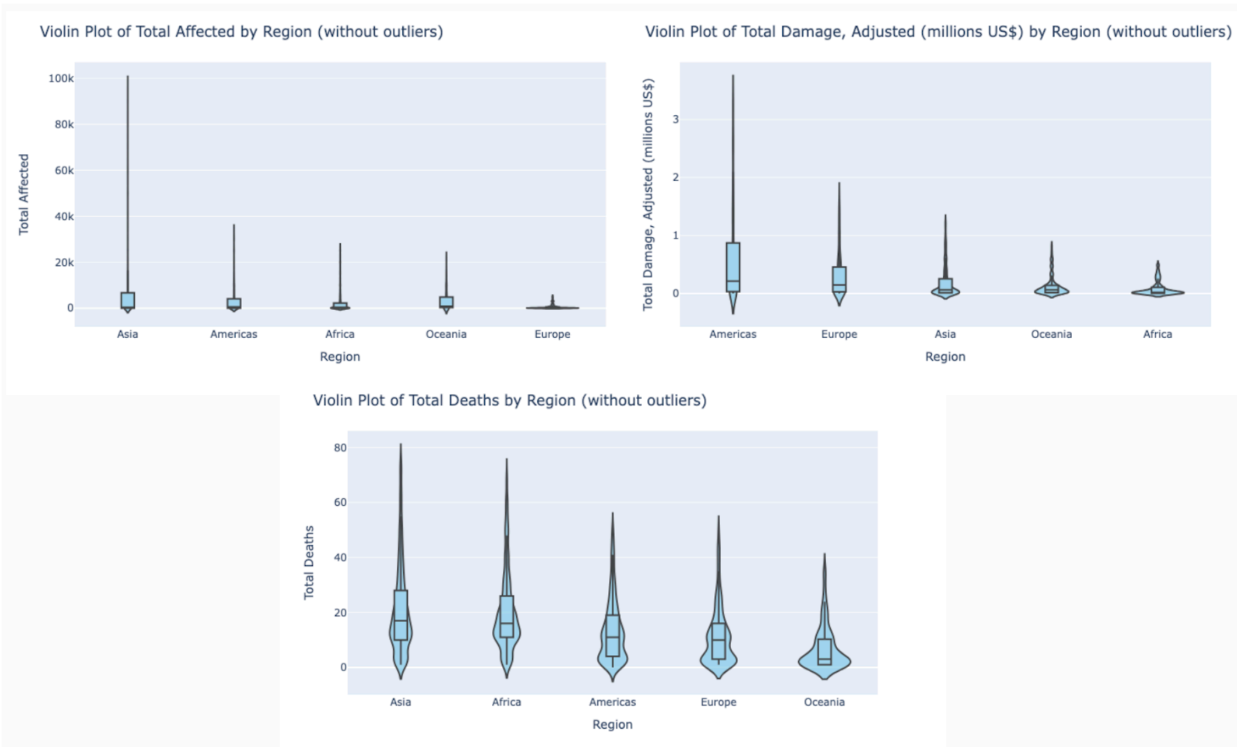


Figure 4

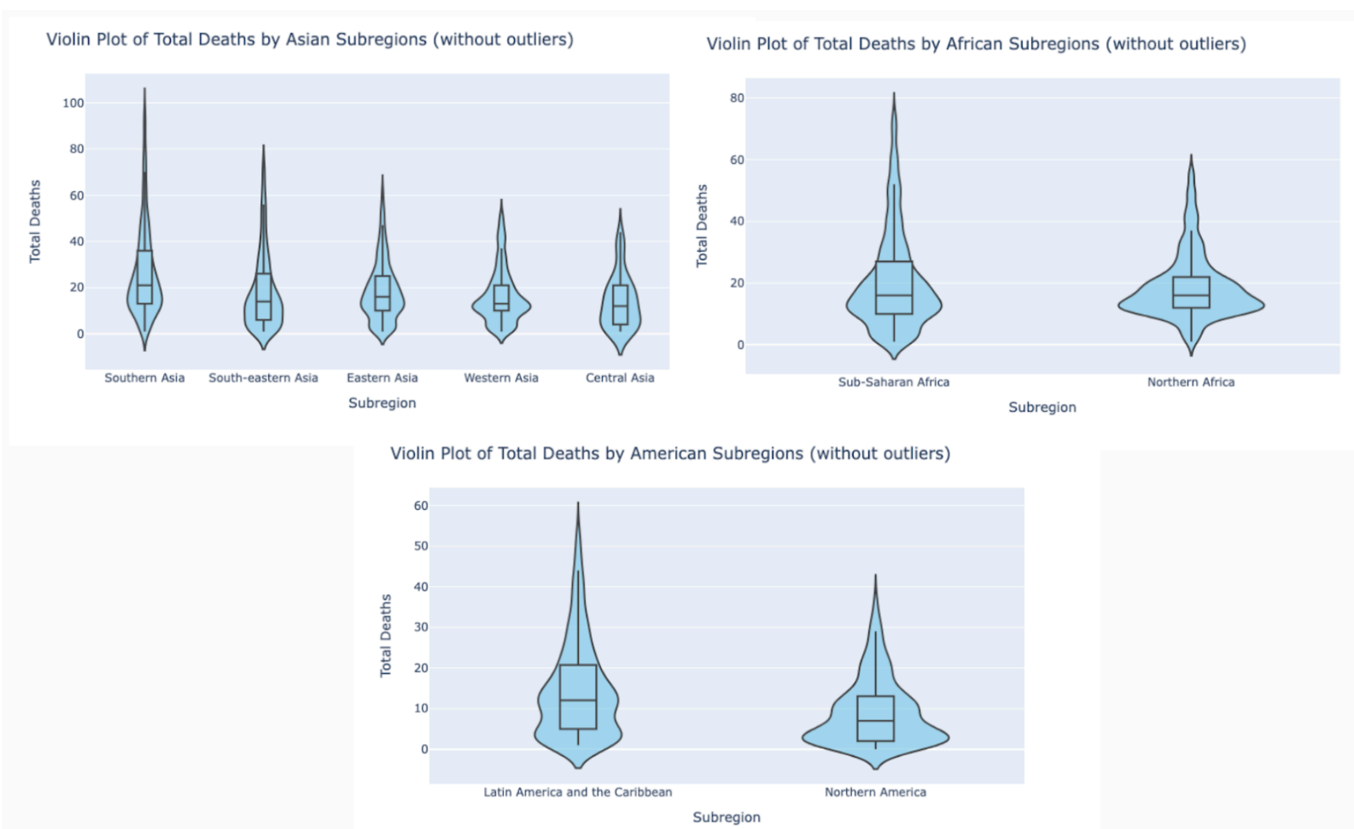


Figure 5

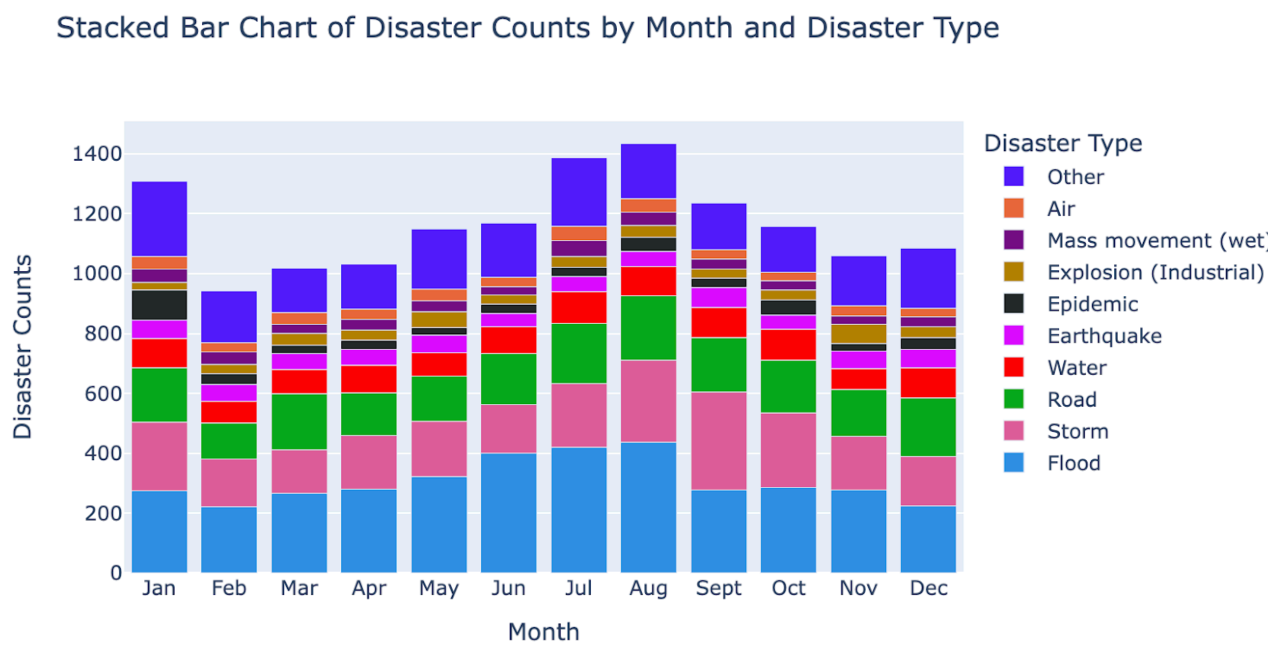


Figure 6

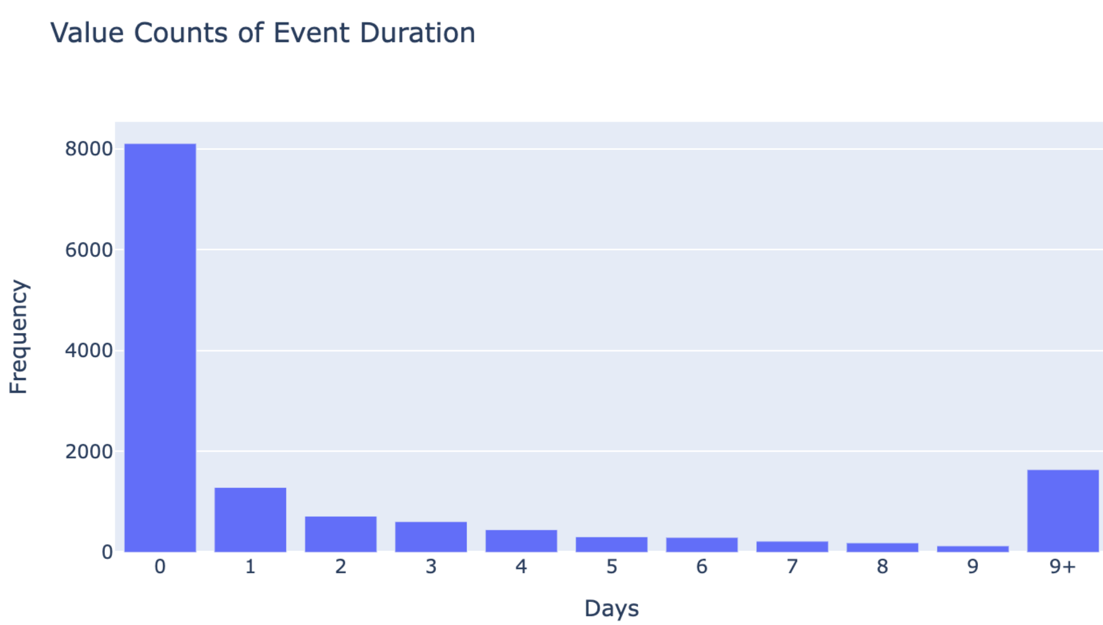


Figure 7

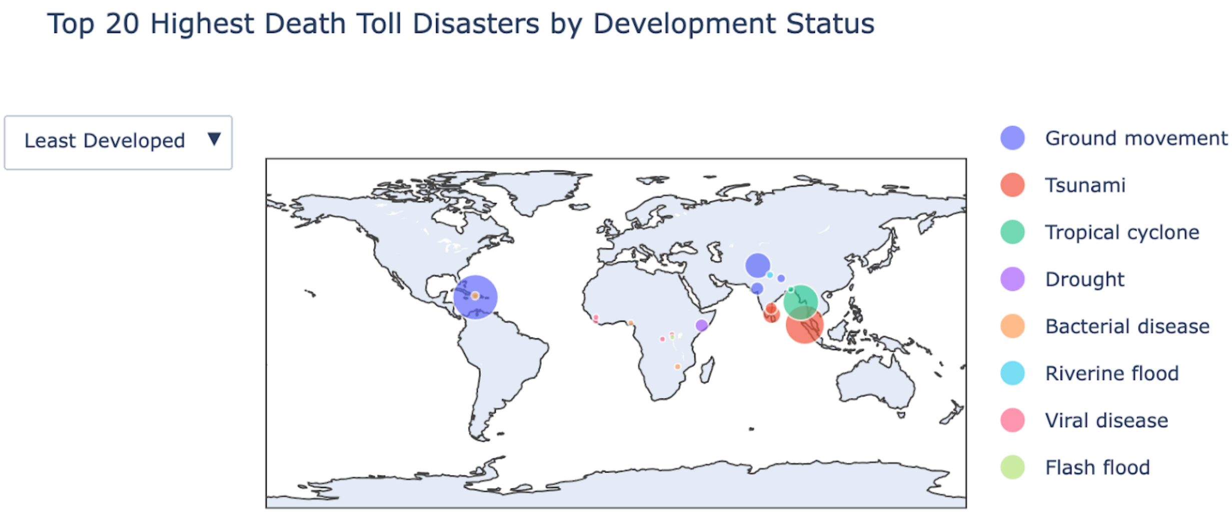


Figure 8

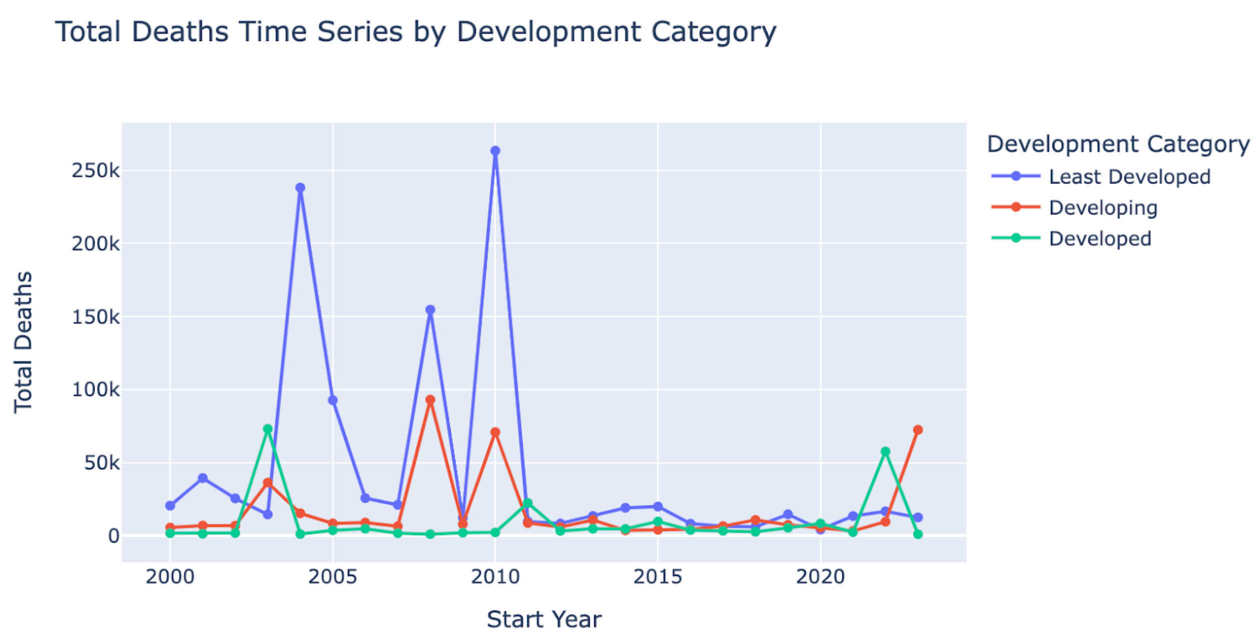


Figure 9

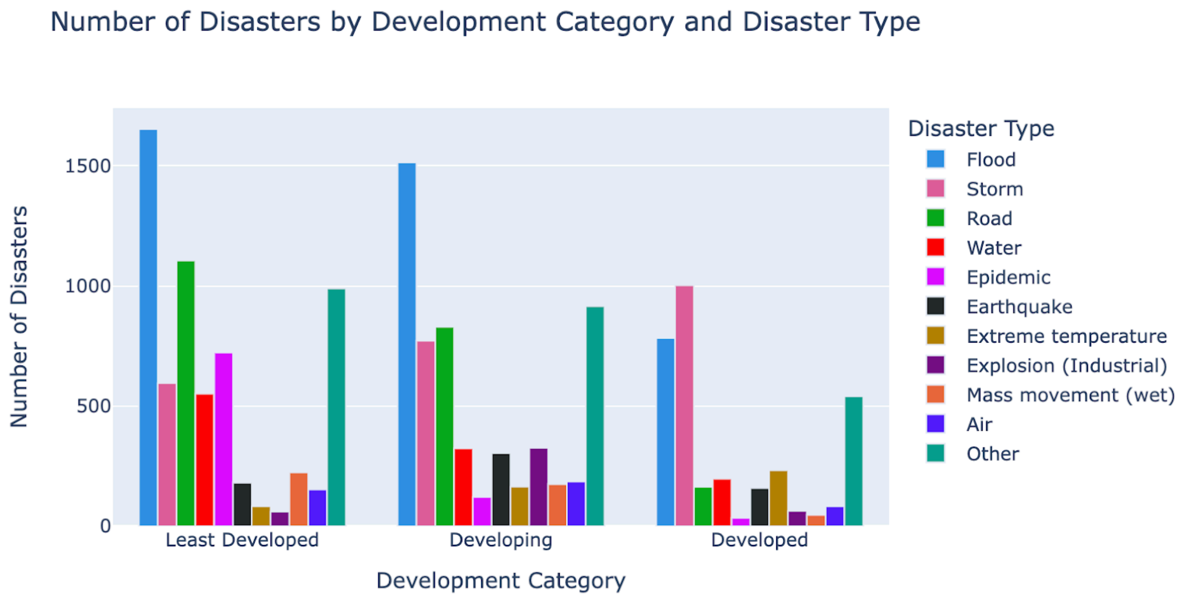


Figure 10

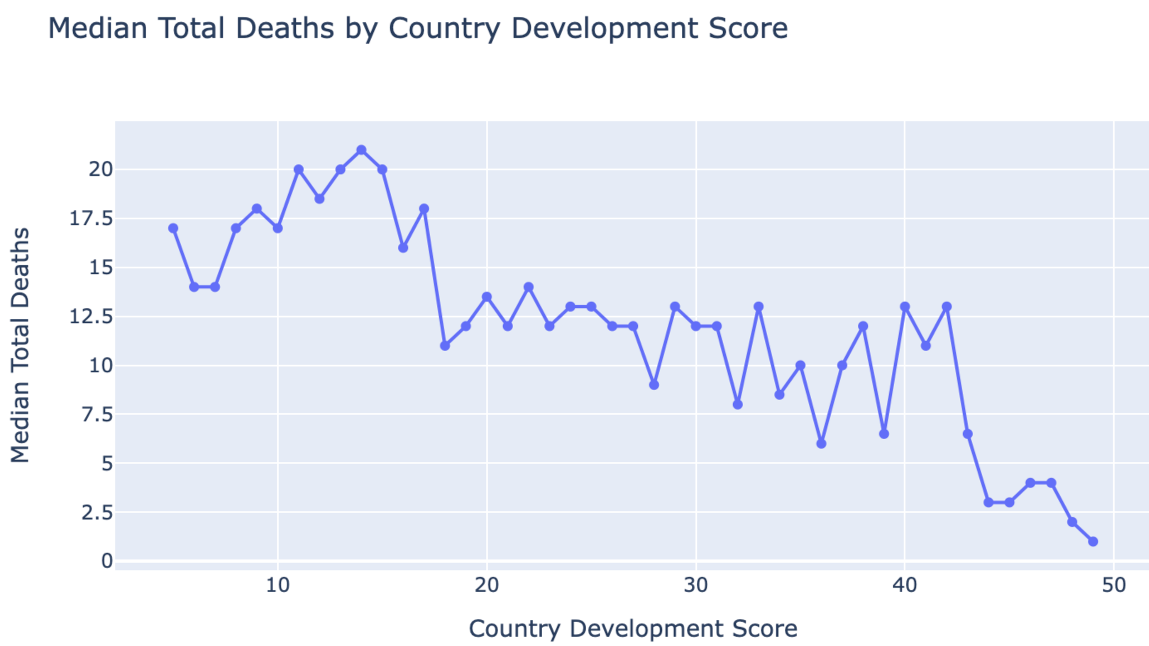


Figure 11

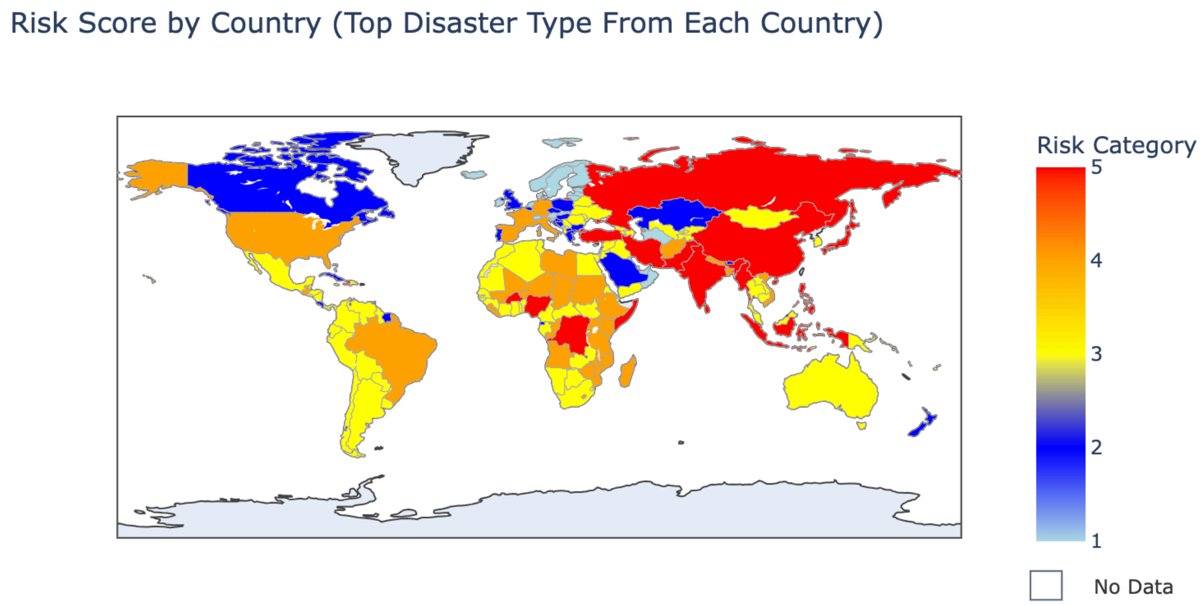


Table 1: Top 20 Risk Scores[2]

Country	ISO	Disaster Type	scaled_relative_frequency	Impact	lack_of_resilence_scaled	Risk Score	Category
India	IND	Flood	89.400922	100.00000	73.146371	90.636823	Very High
Democratic Republic of the Congo	COD	Epidemic	100.000000	76.26625	98.691260	87.805940	Very High
Somalia	SOM	Drought	40.000000	100.00000	98.787879	84.696970	Very High
China	CHN	Earthquake	100.000000	100.00000	36.176046	84.044012	Very High
Indonesia	IDN	Earthquake	69.369369	100.00000	63.041178	83.102637	Very High
Philippines	PHL	Storm	52.421652	100.00000	66.456490	79.719536	Very High
Nigeria	NGA	Epidemic	76.190476	74.78875	87.400088	78.292016	Very High
Pakistan	PAK	Earthquake	16.216216	100.00000	81.448857	74.416268	Very High
Haiti	HTI	Earthquake	3.603604	100.00000	87.031600	72.658801	Very High
Iran (Islamic Republic of)	IRN	Earthquake	49.549550	100.00000	40.029697	72.394812	Very High
Myanmar	MMR	Storm	2.849003	100.00000	84.132275	71.745319	Very High
China	CHN	Flood	100.000000	73.92250	36.176046	71.005262	Very High
India	IND	Earthquake	9.909910	100.00000	73.146371	70.764070	Very High
Türkiye	TUR	Earthquake	27.927928	100.00000	35.021854	65.737446	Very High
Sri Lanka	LKA	Earthquake	0.900901	100.00000	55.601355	64.125564	Very High
China	CHN	Explosion (Industrial)	100.000000	55.51375	36.176046	61.800887	Very High
India	IND	Extreme temperature	100.000000	36.14750	73.146371	61.360343	Very High
Pakistan	PAK	Flood	38.709677	61.97500	81.448857	61.027134	Very High
Russian Federation	RUS	Extreme temperature	38.235294	83.33025	33.574879	59.617668	Very High
Indonesia	IDN	Flood	90.322581	41.26750	63.041178	58.974690	Very High

[2] Please note that 'lack_of_resilence_scaled' is 'inverted development score.'

References

Centre for Research on the Epidemiology of Disasters (CRED). (2024). *EM Dat - The International Disaster Database*. EM DAT. https://www.emdat.be/

Centre for Research on the Epidemiology of Disasters (CRED). (2023a, April 13). *Entry criteria*. EM DAT Documentation. https://doc.emdat.be/docs/protocols/entry-criteria/

Centre for Research on the Epidemiology of Disasters (CRED). (2023b, April 13). *Specific biases*. EM DAT Documentation. https://doc.emdat.be/docs/known-issues-and-limitations/specific-biases/

FEMA. (2024). *Determining risk*. Determining Risk | National Risk Index. https://hazards.fema.gov/nri/determining-risk

UN. (2023, May 12). *Our impact*. UNDRR. https://www.undrr.org/our-work/our-impact

World Bank. (2024). *World development indicators*. WDI - Home. https://datatopics.worldbank.org/world-development-indicators/