University of Liverpool

Computer Science Department

Research Methods in Computer Science - COMP516

NATURAL DISASTERS AND CASUALTIES: A DATA ANALYSIS

By:

GROUP-21

MAAZ AHMAD	**201667108**
ODIANOSEN MASADE	**201645608**
SADHVI PUGANOKAR	**201672582**
SINDHIYA SRIDHARAN	**201629979**
SUREKHA CHINTHALAPUDI	**201682363**
VAMSI MUDILA	**201668163**
YASHITA GAIKWAD	**201687343**

December 2022

Supervisors/Module Coordinator:

DR. OTHON MICHAIL

DR. FAWADA QAISER

PROJECT DESCRIPTION:

Natural hazards are defined as environmental phenomena that have the potential to impact societies and the human environment (National Risk Index, 2022) [1]. The word Natural hazards is analogous to Natural Disasters, and for the purposes of this report, we will be mostly using the word Natural Disasters. For clarity, the use of the term Natural disasters over Natural hazards in the definition above is to reduce any possible ambiguity that the definition might create as it is possible it gives leeway many other phenomena not considered in our scope. (Mizutori, 2020, as cited in IFRC, 2022) [2] opines that there exists a geopolitical context to Natural hazards that make them Natural disasters. This includes social exclusion, poverty, amongst other economic factors. According to the FEMA definition above, there are 18 environmental phenomena that are Natural hazards. They include avalanche, coastal flooding, cold wave, drought, earthquake, hail, heat wave, hurricane (tropical cyclone), ice storm, landslide, lightning, riverine flooding, strong wind, tornado, tsunami, volcanic activity, wildfire, winter weather. Other lesser considered phenomena are tornados and dust storms, which would make the entire spectrum 20. However, others, like the IFRC have considered a fewer number of phenomena as Natural hazards or disasters. All the above-mentioned Natural disasters happen across the different geographical zones of the earth with their severity ranging which have been codified by a variety of scholars and varying frequency of occurrence. In this project, we will be examining the changes in frequency of each disaster we have considered for analysis, we will also be studying the distinct populations affected by each of these disasters and the inherent effects on them.

AIMS & OBJECTIVES:

The purpose of this study is to analyze data on natural disasters and their impact on mortality, impacts and survival of the Earth's population aiming to examine the changes in frequency of each of the 9-disasters subject to different geographical locations and to study the population affected by each disaster by analysing the credible data from past 21 years (2000-2020). In today's technological development, knowing the future impacts of society is a continuous process. Data analysis tools have played a key role in assessing the long-term effects of natural disasters. There are many authorized agencies from which credible data can be gathered during and after a disaster. This study also focuses on analysing the impact on the top 10 natural disaster-prone countries and their impact on human population. The long-term consequences of a disaster affect people's lives, society and the atmosphere associated with them. This analysis will help to estimate the future effects of a disaster using historical data which helps in taking preventive measures to reduce its impact on global population.

KEY LITERATURE AND BACKGROUND READING:

Both industrialized and developing nations experience a variety of natural disasters that result in death, casualties, property destruction, and other costs. A tiny flood, tornado, landslide, lightning strike, or earthquake are examples of small-scale hazard occurrences that can produce relatively localized harm, injuring or killing a few people while destroying or causing significant damage to a small number of structures. Contrarily, massive occurrences like hurricanes and tropical cyclones, strong earthquakes, large volcanic eruptions, tsunamis, major floods, and drought can cause significant economic and social disruption due to both direct damage and indirect economic losses. They can also result in the death or injury of tens of thousands of people. The impact on the population of the natural disasters can be dependent on the preparedness of each society. While it is a secondary consideration, post disaster financial and economic challenges are a further issue to deal with in many populations. A recent example will be the economic impact of the COVID-19 pandemic on low and middle-income countries.

According to (Khetan et al, 2022) [3], the financial impact was greater in upper middle-income countries with odds ratio of 2.09 and greatest in lower middle-income countries with odds ratio of 16.88 when compared to high income countries. Also, one must consider the severity of the disaster.

Scope		Number of casualties		Geographic area affected
I	Small disaster	<10	or	<1 km^2
II	Medium disaster	10–100	or	1–10 km^2
III	Large disaster	100–1000	or	10–100 km^2
IV	Enormous disaster	1000–10,000	or	100–1000 km^2
V	Gargantuan disaster	>10,000	or	>1000 km^2

Table 1: Number of Causalities

Table 1 (Gad-El-Hak, 2009 as cited in Esfeh et al, 2016) [4] above shows the classification of disasters, dependent on the number of casualties or the geographical area affected. Another area of interest is the response to disaster; the consideration of the preparedness and early warning, impact and response, mitigation, risk, and vulnerability modelling It is believed that timely response can abate the effects of natural disasters as regards rate of casualties. In this area, data plays an important role, whether it is user generated from social media or publisher generated data, the use of data at all stages of response to natural disaster cannot be overemphasized. A recent example is the impact of social media in the response to flooding across Nigeria.

Having opined that the casualties and mortalities are a function of economic indices and other factors, available data points to the fact that certain natural disasters have been more severe than others. Historically, droughts have affected a large space of demography ranging from the Indian subcontinent to Africa. Floods have also been a menace, with the incidence and severity increasing across the world. According to (CRED, 2015, Doocy et al., 2013, Gu et al., 2020, Lee et al., 2018) [5], flooding has been responsible for more than half a million deaths over the past 30 years. Earthquakes are another major natural disaster with severity. They even are a source of secondary natural disasters. The tectonic movements can cause major marine landslides which also lead to tsunamis. (Li et al, 2022) [6] suggest that they are the biggest threat to coast infrastructure and cities. Other phenomena of grave severity worthy of mention include volcanic activity, storms, extreme temperatures, landslides, and wildfires.

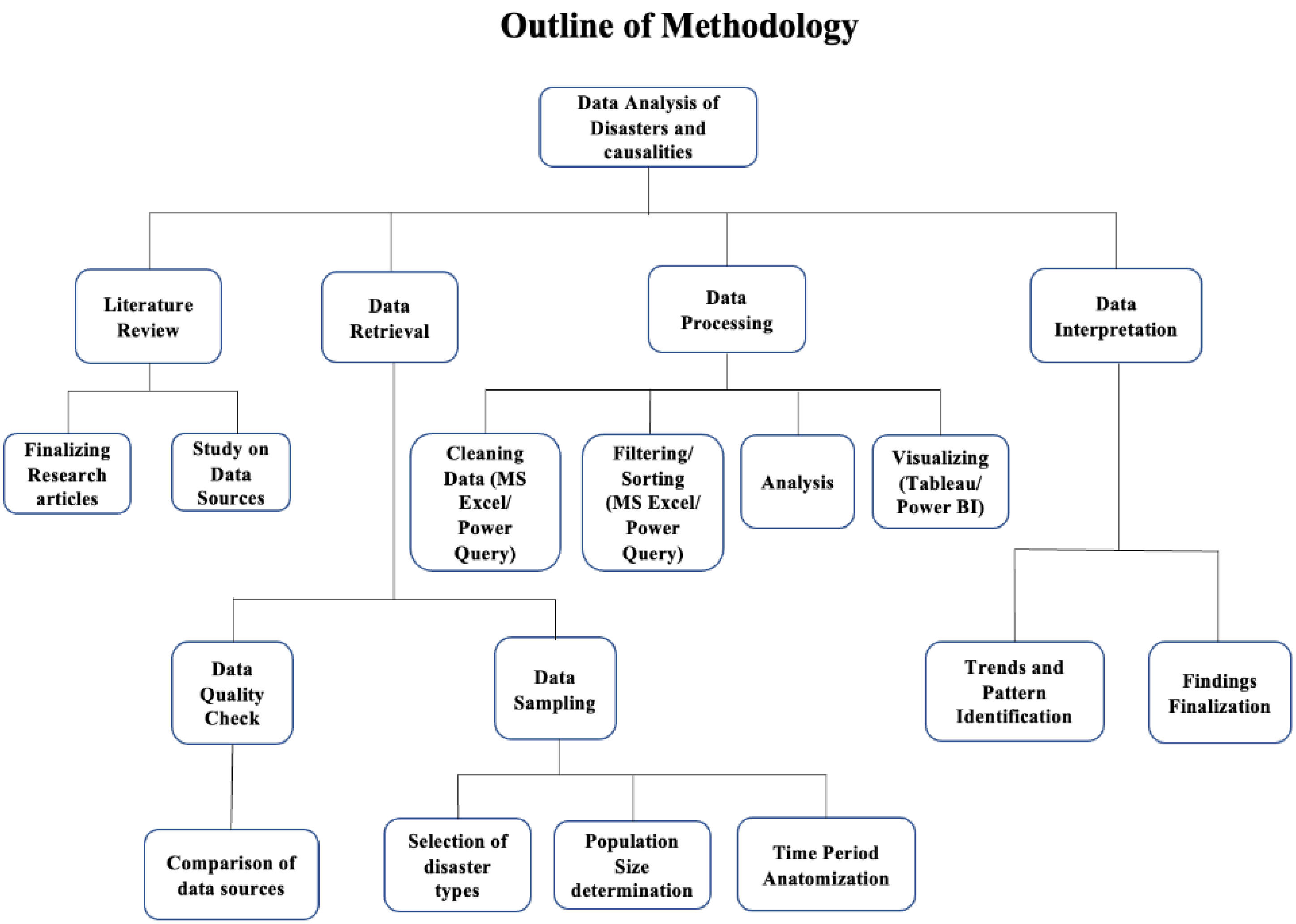

Figure 1: Outline of Methodology

DEVELOPMENT AND IMPLEMENTATION SUMMARY:

According to the aim of the research work, an outline of methodologies involved is mentioned in the above block diagram.

Data: The data regarding natural disasters where culled from the EM-DAT, The International Disaster Database (Guoqiang Shen, 2019) [7]. This database consists of freshly updated forecast regarding a huge set of disasters and their impacts over many countries for different timespans (EM-DAT, 2015) [8]. The geographical coverage of this database is global-country and regional level. To exclude inconsistency and improve quality of research, data from the period of 2000-2020 were taken into consideration.

Data Retrieval: With the help of review of literature, different databases like SHELDUS and EM-DAT were compared, and the data related to the research work was retrieved from EM-DAT alone because of its non-redundancy. Two datasets were taken into consideration those were global population dataset and natural disasters dataset.

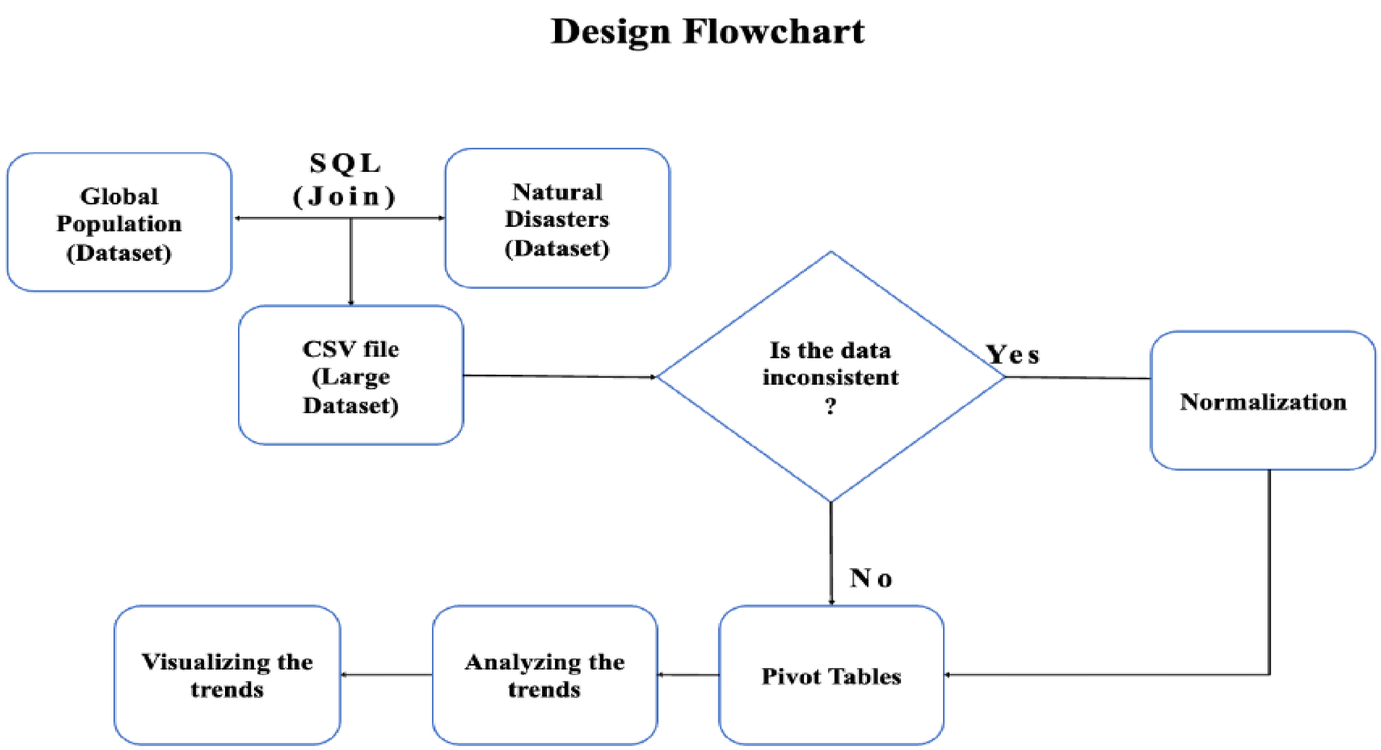

Figure 2: Design Flowchart

Data Processing: Data cleaning was performed over both the datasets retrieved, due to the presence of incorrect and duplicate data, preventing erratic outcomes of this research work (Erhard Rahm, 2000) [9]. The same were joined using SQL Server via the Join operation for the further investigation. The cleaning of the data was performed in MS Excel/ Power Query.

DATA SOURCES:

We gathered every single one the data used in our project from the following datasets: EM-DAT (2015), and (NGDC/WDS) [10]. These every single one currently is available in the public domain in addition to accessing in addition to processing of the data currently exists available to everyone. Just as our project currently exists extensively based on the quality in addition to credibility of the data, we have put a lot of thought into it. It is available to everyone, so all of us don't need any kind of permission to use it in addition to all of us imported it directly from the metadata of the websites.

All of us have made sure not to include any kind of personal data in our analysis. For long-term trends in natural disasters, all of us know that object over there reporting in addition to recording of events this day currently exists much more advanced in addition to complete than in the past. We have considered the data from the turn of this millennium to the present (2000-2020). Credibility, quality, in addition to the six dimensions of data (accuracy, completeness, consistency, uniqueness, timeliness, in addition to validity) are of utmost importance for good analysis. We have performed regression analysis to observe if there currently are any correlations to establish between the multi-variables.

EVALUATION:

According to the aims and objectives, we have observed the trend analysis comparing various factors and have provided the observations below:

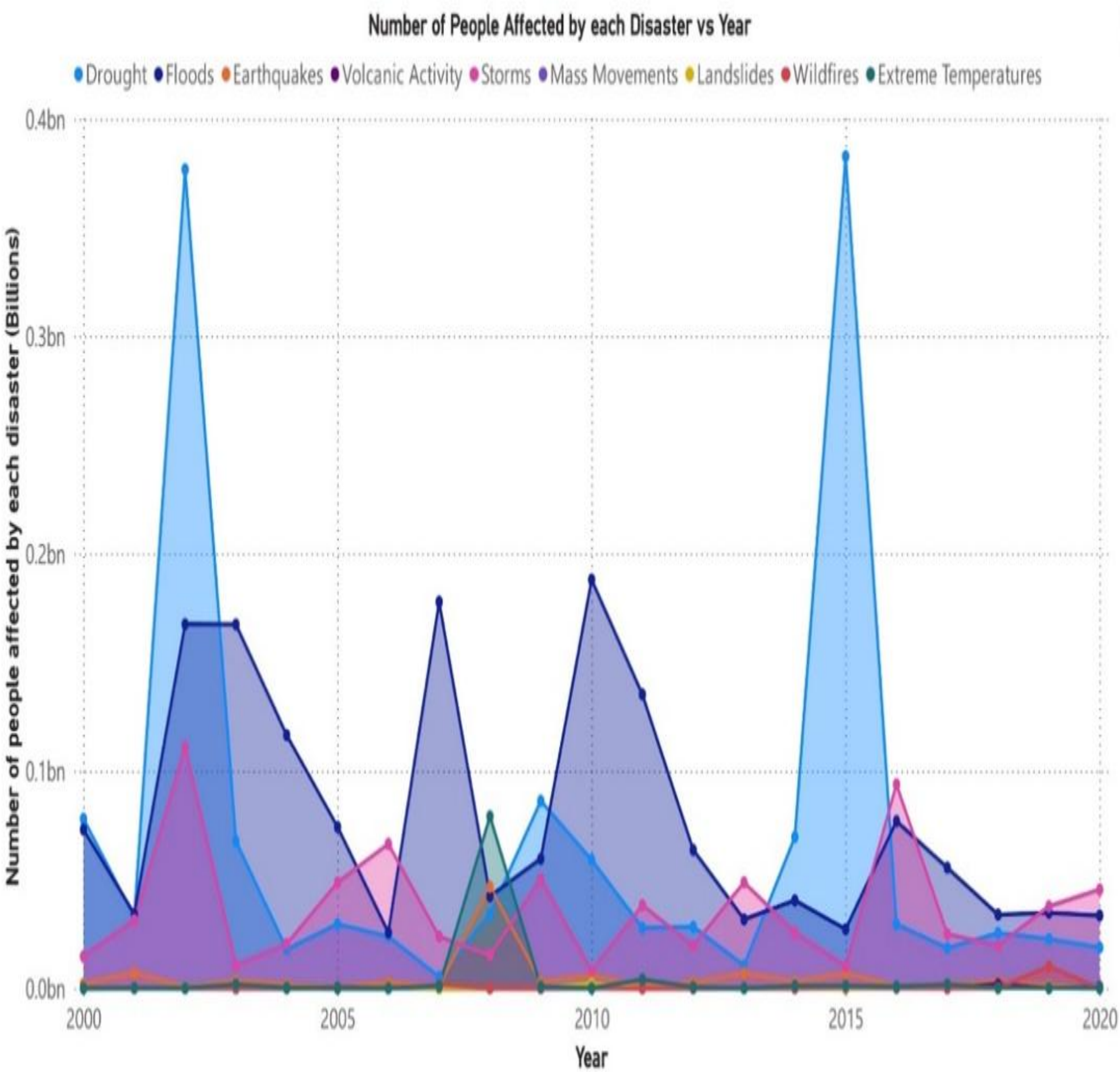

Figure 3: Trend analysis of Number of People Affected by each Disaster for the Years

In Figure 3, that drought has the highest value for the number of people affected with the value reaching up to 382985115 (382million) in the year 2015 and second highest value for up to 376956935 (376 million) for the year 2002. On the contradictory extreme temperature had the lowest value amongst all the natural disasters with its value as low as zero for the year 2006. Storms and extreme temperatures had the lowest value for the timeline of two decades but for the year 2008 we see that the trends change, and these disasters affected maximum number of people in comparison to other calamities. After drought floods affects the highest number for people with values reaching as hight as 188 million for the year 2010 whereas floods affected lowest people in the year 2006.

As concluded from the figure 3, that drought had the highest value for the number of people affected one might assume that the number of people injured by drought would also be large, but this is not the case depicted by figure 4 below.

Instead, the figure 4 shows that the number of people injured by drought is the least to the minimum value as low as zero except for the year 2019 which is 32 people in total which is almost negligible in comparison with other values. Year 2004 has the highest value for total number of people injured which is near 1.8 million which was caused by extreme temperature. Additionally, Earthquake showed the highest value in terms of number of people injured for most of the years over the time frame of two decades.

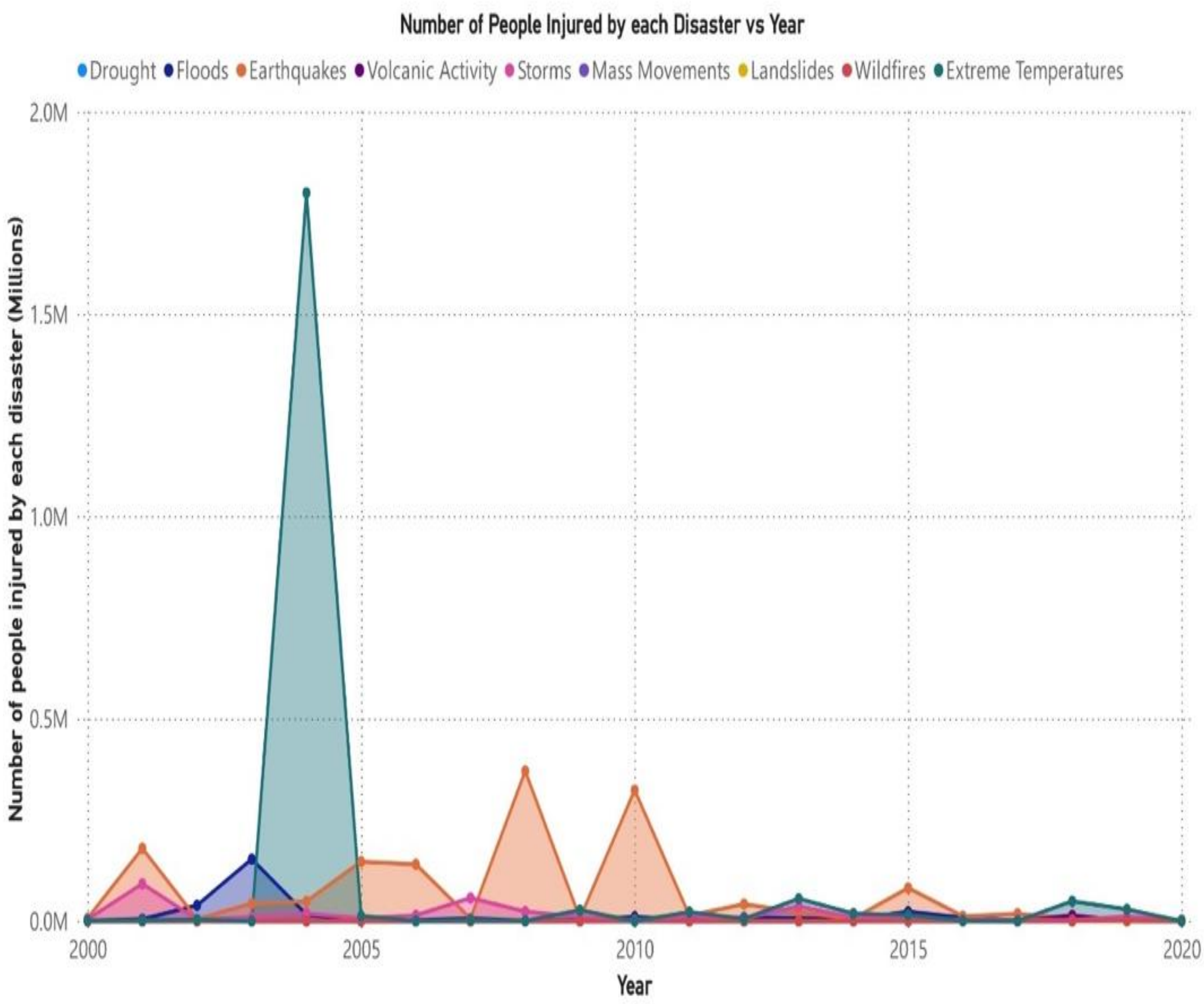

Figure 4: Trend Analysis of Number of People Injured by each Disaster for the Years

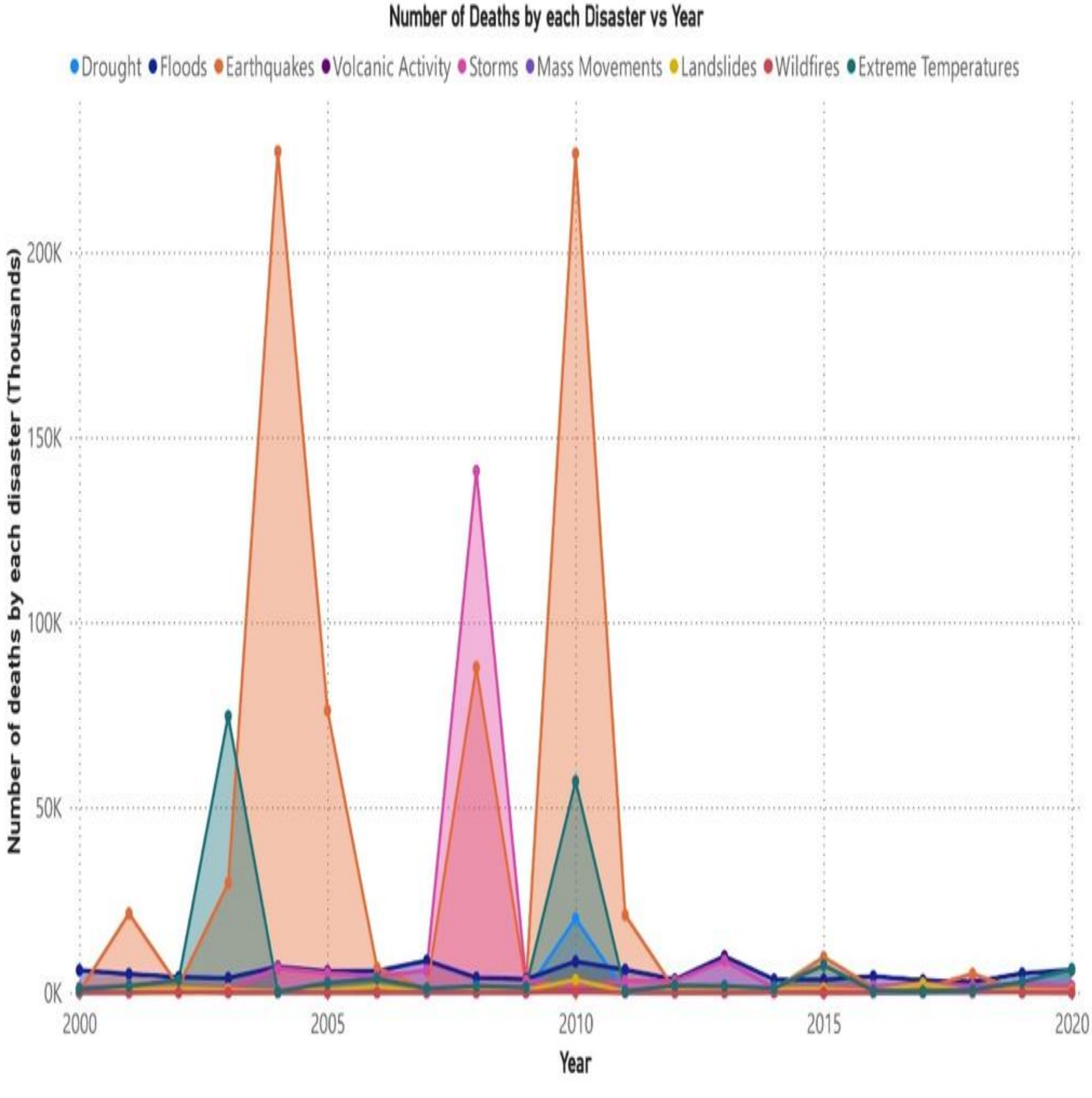

Figure 5: Trend Analysis of Number of Deaths by each Disaster for the Years

Coming to figure 5 above, which represents the trends for the number of people died due to different disasters it is observed that these values were higher in lakhs for the early years and the number of people died dripped to thousands in the recent years. In the years 2000 to 2010 we see a lot of fluctuations in the values from different calamities with the highest deaths caused by earthquake for the year 2010 and 2004 with values corresponding to 2.2 lakhs for both the years. After drought storm has the second highest value of around 1.4 lakhs peopled died in the year 2008. After 2010 till 2020 the deaths caused by all the natural disasters varied in the range of thousands which if very less in comparison with the previous years.

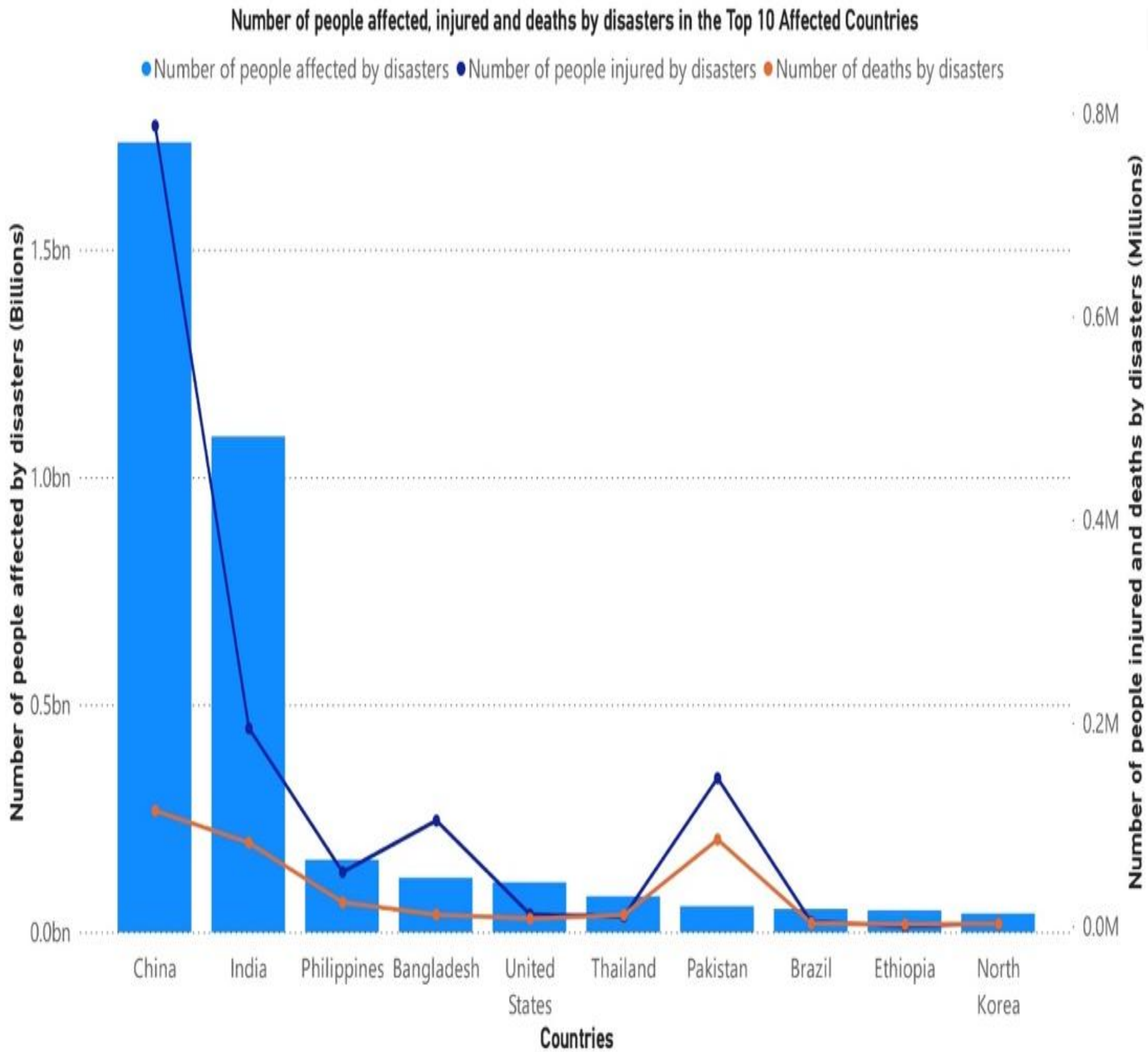

Figure 6: Trend Analysis and comparison between Number of People Affected, Injured and Deaths by Disasters in the top 10 Affected Countries

According to figure 6, which shows the graphical representation for the overall people affected, injured, and died by various disasters in the top 10 affected countries it is observed that China has the heights value for all the categories which values crossing 1.7 billion for people affected whereas North Korea being the least. Usually, the number of people affected is greater than the number of people injured for all the top ten countries except for China, Pakistan, and Bangladesh where we can see that the number of people injured outrages the total people affected when considered in terms of ratio. Number of people affected by disasters and total injured by disasters are positively correlated. China accounted for 49.80% of Number of people affected by disasters. India stands on the 2nd highest where the value for people affected is slightly greater than 1 billion out of which 1.9 lakhs were injured and nearly 82 thousand people died. It is observed that Bangladesh and United States has almost the same value for affected people, but the number of people injured is much higher in Bangladesh when compared to US with values accounting roughly around 1 lakh and 11 thousand, respectively.

ETHICAL CONSIDERATIONS:

Natural disaster research focuses on the effects of disasters on people and social structures. Natural disaster planning and response requires evidence to guide decision makers. This need for evidence provides an ethical imperative for conducting acoustic hazard research. However, we also turn our attention to cross-cultural cooperation and the vulnerability of participants resulting from the experiment. However, disaster research involves several distinctive factors, such as the degree of devastation affecting participants and the urgency involved in initiating research projects. Our analysis considered authoritative data obtained from open and published sources. We do not consider racism or geographic discrimination. Our analysis is based on reliable data collected from reliable sources and does not contain information about individuals or groups.

PROJECT PLAN:

Activity		**Effort**	**Duration**
1) Review of Literature (M1)	Finalising research articles/ Data Sources	2 weeks	3 weeks
2) Data Retrieval		1 week	1 week
3) Data Processing	Cleaning Data	1 week	1 week
	Filtering/Sorting	1 week	1 week
	Analysis	1 week	1 week
4) Data Interpretation	Trends and Pattern Identification	1 week	1 week
	Findings Finalisation	1 week	1 week
5) Report/Presentation (M2)		1 week	1 week
Total		**9 weeks**	**10 weeks**

Table 2: PROJECT PLAN – WORK BREAKDOWN

MILESTONES

M1	Completion of Literature review
M2	Completion of report / Project

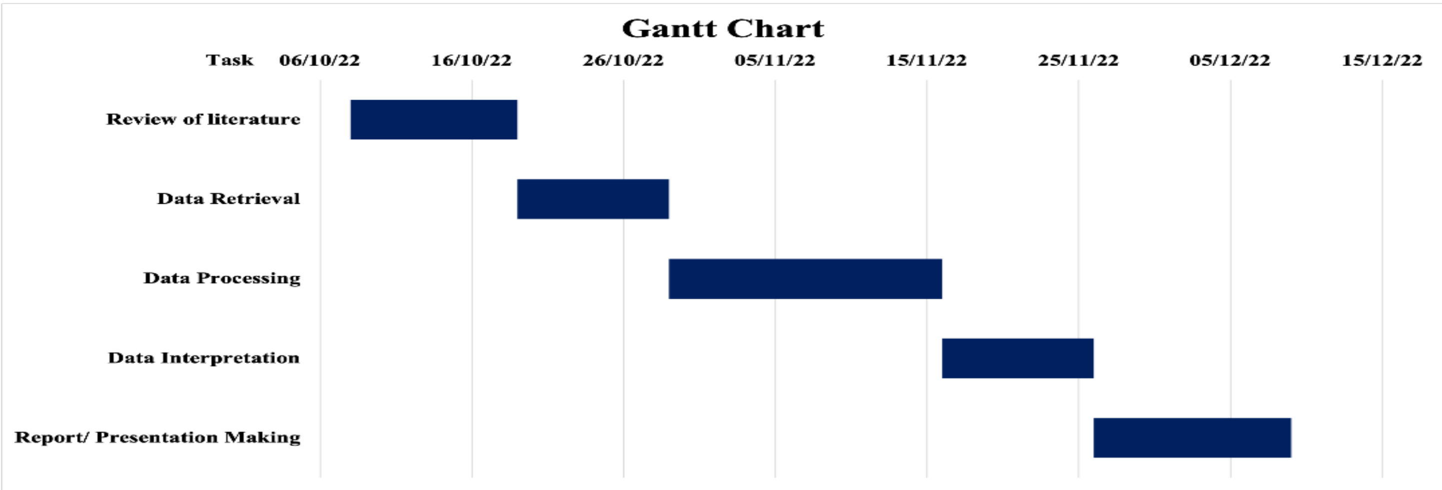

Figure 7: Project Plan – Gantt Chart

The Gantt chart shows the overall timeline for the project with an average timespan of 10 days for consecutive processes of the project.

The first part of the process is Review of Literature starting from 6/10/22 to 16/10/22. This part of the project includes background reading and collecting various sources of information and aligning it to the project domain. The second part of the process is Data Retrieval. This part of the process includes extraction of required data from two large datasets in a csv file.

The third part of the process is Data Processing starting from 26/10/22 to 15/11/22. The fourth part of the process is Data Interpretation starting from 15/11/22 to 25/11/22. This part of the process identifies trends and patterns from processed data and visualises the final trends from data through Tableau. The fifth part of the process is Report/ Presentation making starting from 25/11/22 to 5/12/22. This part of the process showcases the conclusions and milestones achieved from the project.

Data interpretation entirely depends on the Data processing part of the project as after using all the process methods, analysation and generation of trends can be achieved correctly. If there is any inconsistent data present, it will showcase different trends/patterns which may not align within the targeted findings.

RISK & CONTINGENCY PLAN:

Risks	Contingencies	Likelihood	impact
Hardware Failure	Backup to cloud	Med (2)	10
Software Failure	Use of existing software and formulas instead of developing if ourselves	Med (2)	6
Running out of time	Split the project among everyone	High (3)	6
Programming Tools	Used the tools which we know the best	Low (1)	1
Data integrity	Perform elementary tests	Low (1)	3

Table 3: Risk and contingency

* Impact = Likelihood X Consequence

Risk Likelihood	Score
Low	1
Medium	2
High	3

Table 4: Likelihood

Risk Consequence	Score
Very Low	1
Low	2
Medium	3
High	4
Very High	5

Table 5: Consequences

- RAG grading was used in our project:
  Red Risks with impact greater than 10 - Critical risks
  Amber Risks with impact between 6 and 10 - Deserves some attention
  Green Risks with impact smaller than 6 - Can be ignored

REFERENCES:

1. FEMA. (n.d.). Natural Hazards | National Risk Index. Retrieved from hazards.fema.gov website: https://hazards.fema.gov/nri/natural-hazards

2. What is a disaster? | IFRC. (n.d.). Retrieved from www.ifrc.org website: https://www.ifrc.org/our-work/disasters-climate-and-crises/what-disaster

3. Khetan, A. K., Yusuf, S., Lopez-Jaramillo, P., Szuba, A., Orlandini, A., Mat-Nasir, N., … Mohan, D. (2022). Variations in the financial impact of the COVID-19 pandemic across 5 continents: A cross-sectional, individual level analysis. EClinicalMedicine, 44. https://doi.org/10.1016/j.eclinm.2022.101284

4. Ansari Esfeh, M., Caldera, H. J., Heshami, S., Moshahedi, N., & Wirasinghe, S. C. (2016). The severity of earthquake events – statistical analysis and classification. International Journal of Urban Sciences, 20(sup1), 4–24. https://doi.org/10.1080/12265934.2016.1138876

5. Feng, S., Liu, J., Zhang, Y., Beck, H. E., & Gu, X. (2020). Global changes in floods and possible mechanisms. Washington: American Geophysical Union. Retrieved from https://doi.org/10.1002/essoar.10501527.1

6. Special report of the intergovernmental panel on climate change managing the risks of extreme events and disasters to advance climate change adaptation. (n.d.). Retrieved from https://www.ipcc.ch/site/assets/uploads/2018/03/SREX_Full_Report-1.pdf

7. Shen, G., & Hwang, S. N. (2019). Spatial–Temporal snapshots of global natural disaster impacts Revealed from EM-DAT for 1900-2015. Geomatics, Natural Hazards and Risk, 10(1), 912–934. https://doi.org/10.1080/19475705.2018.1552630

8. EM-DAT | The international disasters database. (2015). Retrieved from Emdat.be website: https://www.emdat.be/

9. (PDF) Data Cleaning: Problems and Current Approaches. (n.d.). Retrieved from ResearchGate website: https://www.researchgate.net/publication/220282831_Data_Cleaning_Problems_and_Current_Approaches

10. ncei.info@noaa.gov. (2019). NOAA National Centers for Environmental Information (NCEI). Retrieved from Noaa.gov website: https://www.ngdc.noaa.gov/

11. Jones, B. G., & Tomazevic, M. (1982). Social and economic aspects of earthquakes. In Proceedings of the Third International Conference: The Social and Economic Aspects of Earthquakes and Planning to Mitigate their Impacts. Yugoslavia. Institute for Testing and Research in Materials and Structures; US Cornell University. Program in Urban and Regional Studies.

12. Beer, Parkhill, D. F., Enslow, P. H., & Andrews, C. (1984). So this is 1984: some personal views by governors of the International Council for Computer Communication. North-Holland.

13. de Bruycker, M., Greco, D., Annino, I., Stazi, M. A., de Ruggiero, N., Triassi, M., … Lechat, M. F. (1983). The 1980 earthquake in southern Italy: rescue of trapped victims and mortality. Bulletin of the World Health Organization, 61(6), 1021–1025. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2536241/

14. New Horizons in Health Promotion: Emergency Relief and Disaster Preparedness, Medical Society Lectures. (1980). WHO Dialogue.

15. Bryant, E. (2004). Natural Hazards (2nd ed.). Cambridge: Cambridge University Press. Retrieved from https://doi-org.liverpool.idm.oclc.org/10.1017/CBO9780511811845