GLOBAL EARTHQUAKES

Data Visualization - Nova IMS

Master in Data Science and Advanced Analytics

FELIX GABER | 20221385
MAFALDA PAÇO | 20220619
MARTA DINIS | 20220611
PATRÍCIA MORAIS | 20220638

1. Introduction

Earthquakes have been a significant geological phenomenon for centuries, causing mass destruction and still huge mortality. The main inspiration for this project emerged from the pertinent issue of earthquake impact and aims to identify the patterns and trends that lead to seismic activity across the globe. We wish to draw attention, in particular, to the needless harm caused by human activity-induced earthquakes, which could be avoided with willpower and better understanding and divulgation of this issue.

It's important to note that, while earthquake data has been monitored for many years, this recording has not been as frequent, accurate, and detailed as it is now. Therefore, analyzing this data to identify patterns and trends can be challenging. To overcome this challenge, we combined earthquake data over the past centuries with additional data on tectonic plates and human activity, the main identified seismic activity causes. Using Plotly and Dash, we created a platform that allows users to explore earthquake data from various angles and gain insights into patterns and trends in seismic activity.

Our goal is to make the data accessible and easy to understand, even for those without a background in geology or seismology. Through this project, we hope to demonstrate the power of data visualization tools like Plotly and Dash to communicate complex information simply and promote a better understanding of important issues like earthquake activity. We hope that the platform we created will inspire curiosity, promote awareness, and encourage action to mitigate the impact of earthquakes on communities around the world.

With this dashboard, users will be able to interact with various graphs through dropdowns, where they can select and explore information on one or more desired countries, linked graphs, radio items components, in which users can select an option from a specific set, clicking on specific parts of the plot to unlock more information, or just simply hovering the graphs themselves to obtain the primary information they provide.

2. Data and Methods

We downloaded the earthquake data from Opendatasoft, a French company that provides data-sharing software. Our initial earthquake dataset contains information on destructive earthquakes from 2150 B.C. to the present, that meet at least one of the following criteria: moderate damage created (at least $1 million), 10 or more deaths, a magnitude of at least 7.5, modified Mercalli intensity X or greater, or if the earthquake generated a tsunami.

For our human activity data, we used The Human-induced earthquake database (HiQuake) where we have information about the industrial projects proposed to have induced or triggered earthquakes.

Finally, for the tectonic plate information we downloaded a dataset from Kaggle that contains the latitude and longitude data that completely encloses the 56 tectonic plates.

2.1. Preprocessing

Since we had several metrics for magnitude scale we decided to look for a way to convert the ones we had more records on to the same scale, which was M_s and M_w. Using the statistical findings on (Kadirioglu & Kartal, 2016) we converted the records on M_s (surface-wave magnitude) scale to M_w (moment magnitude) using the following relationship:

$$M_w = 0.5716 \times M_s + 2.4980$$

To use this conversion we had to eliminate all records with M_s magnitude smaller than 4 (only 9 records were eliminated). In the end, we only kept the records that had magnitude measured either in M_s or M_w all converted to the same units (M_w).

Seeing as our data had records related to very old dates that we don't have much or credible information on we decided only to consider records starting from the year 1900. This data is also more relevant to visualize and relate to what we see nowadays. We also removed all other columns that weren't of interest.

Furthermore, it was important for our visualizations to delete all records with null values on the variable 'Focal Depth'.

2.2. Dataset Description

Our main dataset has the following variables:

Variable	Description
ID Earthquake	Identification of the earthquake
Year	Year of the earthquake
Month	Month of the earthquake
Day	Day of the earthquake
Focal Depth	Depth of the earthquake given in Km
Magnitude	Moment magnitude of the earthquake
Country	Name of the Country
Country ISO Code	3-digit ISO code of the Country
State	Name of the state from United States, Canada and Australia
Total Effects : Deaths	Total number of deaths caused by the earthquake
Total Effects : Missing	Total number of missing people due to the earthquake
Total Effects : Injuries	Total number of injured people due to the earthquake
Total Effects : Damages in million Dollars	Total damages in million Dollars
Total Effects : Houses Destroyed	Total number of Houses destroyed by the earthquake
Total Effects : Houses Damaged	Total number of Houses damaged by the earthquake
Latitude	Latitude of the earthquake
Longitude	Longitude of the earthquake

Table 1 Variable description for earthquake dataset

From the Human-induced earthquake database we used the following variables:

Variable	Description
Country	Name of the Country
Earthquake cause (main class)	Industrial project's main class that caused the earthquake
Number of recorded earthquakes	Number of recorded earthquakes
Observed maximum magnitude (Mmax)	Maximum observed magnitude of the earthquake by cause
Date of Mmax (yyyy/mm/dd)	Date of the observed earthquake with the magnitude

Table 2 Variable description for Human-induced earthquake dataset

2.2. Technical aspects

The data was extracted on 21/03/2023 from the Opendatasoft and Inducedearthquakes websites and it can be accessed in the GitHub repository through this link. To access the dashboard follow this Link to Dash Visualization on Render. The project was implemented using Plotly and Dash.

2.3. Visualization and Interaction Choices

It is possible to get detailed information about the earthquakes by hovering over the markers on our Global Earthquake Overview visualization, such as the country, year, magnitude, and Focal Depth (km) of the earthquake. For the Globally Recorded Earthquakes, hovering will give the user information about the amount of earthquakes and the year.

The hover template for the Magnitude by Earthquakes graph gives information on the country, year, magnitude, and focal depth of the specific seismic occurrence. For the line charts, the hover template shows the country, year, and count. Finally, for the histogram, hovering over the marks will show information on the year and the number of earthquakes recorded.

Hovering over any of the elements of the sunburst will show information on the number of occurrences and the main cause or country, if it's the parent or the child, respectively. It's possible to interact with the line chart by selecting any of the causes present on the legend to omit it from the graph, or by simply hovering over the lines to show the cause, year, and number of earthquakes on the hover template.

3. Results and Discussion

3.1. Data encoding

For the first visualization we decided, using geographic data, to do a Scattergeo with an overview of the earthquakes where color and size encoding represents the magnitude of each earthquake (the bigger the marker and the lighter the color the bigger the magnitude). With this map, we are also using positional encoding providing the map's location of the earthquake.

The Globally Recorded Earthquakes visualization maps the count of earthquakes to the height of each bar, while the width represents the year.

The Magnitude of Earthquakes by Country section includes 5 different graphs relating to the impact of seismic activity:

1. In the first one, we can see how the earthquakes of specific countries are distributed through the years. To be consistent with the previous visualization, we used color to represent the magnitude (the lighter the color the bigger the magnitude), size encoding (the bigger the marker, the bigger the magnitude), and positional encoding to convey the time distribution of the earthquakes.
2. The 3 graphs below are linked to this first one - the country in which we hover over in the first graph is the one setting the information in these subsequent three. These are all line charts that represent the cumulative damages in millions of dollars, the cumulative house damage, and the deaths of the selected country. Therefore, the marks used were lines and the channel used to control their appearance was the spatial position on a common scale, which related to the years on the x-axis and the main feature described before on the y-axis.
3. The graph on the right side is a histogram describing the number of recorded seismic activities through the years. Encoding was done through spatial positioning and length, which represent the year and the number of recorded earthquakes, respectively.

The next and last section of our dashboard is composed of two more graphs regarding human activity-induced earthquakes. The one on the left is a sunburst that contains information about the human-induced earthquakes' main cause and the countries where there were most incidents related to those causes. We use as the main mark the area of each cause and country inside each cause and used as channels the colors in a gradient of the same hue (the more saturated the colors the bigger the frequency of the incidents). The angle mark was also utilized further to enhance the importance of the number of occurrences - the bigger the angle, the higher the number of incidents for that main cause or country.

The graph on the right is a line chart with the distribution of human activity-induced earthquakes caused over time. The number of earthquakes for each specific cause is represented through the years using a line as a mark. The color of the line is used as a categorical channel that distinguishes each cause. The position is also a channel that allows the representation over time on the x-axis and the number of recorded earthquakes on the y-axis.

3.2. Data filtering and interactivity

We implemented radio buttons on our Global Earthquake Overview visualization, so the user can choose whether or not to see the tectonic plates as well as the earthquakes.

Our Magnitude of Earthquakes by Country visualization allows the user to select the countries they wish to see using a dropdown, having the top 5 countries already chosen. Furthermore, this first graph is linked to the three line charts below through the hover feature - the information on the line charts changes according to the country hovered over on the first graph.

We also have a slider implemented on our first visualizations of Humanity as Cause that lets the user decide what percentage of space each graph is taking up. The user is also able to focus on any of the causes of human-induced earthquakes in the corresponding visualization by clicking on it.

4. Conclusion

4.1. Accomplishments

We worked based on three main premises: simplicity, effectiveness, and usefulness.

- Simplicity: we use a palette of colors that match our intention, but also the theme of the dashboard.
- Effectiveness: the sidebar does not move, so at any given moment, the viewer can change the parameters as desired without moving up and down on the dashboard.
- Usefulness: as mentioned earlier, through our choice of graphs, the focus was first to give an overview of the topic, then present the details.

Throughout this project, we have been able to create a simple, beautiful, and user-friendly platform that successfully conveys the importance of this geological phenomenon and the urgency to completely cease any human activity that might trigger these shattering events.

4.2. Limitations

While developing our project, we encountered some limitations, mostly regarding both the datasets in use, but mostly the one concerning human activity-induced earthquakes.

Firstly, we struggled with the various units of measurement used for magnitude. There were several types of metrics but only one for each earthquake which complicated the use of this information in actual plots. Furthermore, the relationship between each of these measurements was unclear and came with certain conditions that implied losing some data.

Another problem we encountered was the high percentage of missing values, especially in the second dataset. This dataset was also very unstructured, as each attribute would sometimes be recorded as a string, datetime, or integer for different observations. To be able to use all of the available data, a very detailed and time-consuming preprocessing of the dataset would be necessary.

Finally, some attributes would be interesting to have in the chosen datasets, and that could have resulted in very insightful graphs that we, unfortunately, weren’t able to create.

4.3. Future work

Unfortunately, there wasn’t a dataset available that combined both naturally induced earthquakes as well as those caused by human activity. In future work, it would be relevant to develop such a dataset to gain more insights into their comparison. For example, it would be interesting to extract information related to the damage that resulted from human activity-induced earthquakes, be it calculated in monetary value, amount of deaths, destruction, etc. Moreover, a comparison between the location, magnitude, and damage caused by these two types of earthquakes could be made, hopefully highlighting the urgency to cease these activities and prevent further unnecessary harm and destruction.

References

Cover page photo by Jose Antonio Gallego Vázquez, 2021

Kadirioglu, F. T., & Kartal, R. F. (2016). The new empirical magnitude conversion relations using an improved earthquake catalog for Turkey and its near vicinity (1900–2012). Turkish Journal of Earth Sciences, 25, 300–310. (https://doi.org/10.3906/yer-1511-7)

Data products for this study were accessed through The Human-Induced Earthquake Database (HiQuake). (www.inducedearthquakes.org)