CS 6630 - Data Visualization

Process Book

Global Earthquake Visualization

https://github.com/Kaeleen/dataviscourse-pr-earthquakevis

By,

Kailing Fu, u1327581
Yu Song, u0857568

Contents

Background and Motivation

Earthquakes are one of the most devastating disasters in the world, and there is no efficient precaution to prevent the damage and loss from it. People have been trying to discover the underlying pattern of where and when an earthquake would happen based on the ANSS Comprehensive Earthquake Catalog. According to observation, ComCat contains an increasing number of earthquakes in recent years, which leads to a higher volume of research on the trends of earthquake's occurrence. The goal of our project "global earthquake visualization" is to explore the distribution and effects of earthquakes in the world since 1700, then analyze those data and list out the areas that are easier to be vulnerable to earthquakes. The NCEI/WDS Global Significant Earthquake Database was also used in our project as the data support.

There is a comprehensive searching system on National Centers for Environmental Information (NCEI)'s official website where users are able to define earthquake parameters, locations, magnitude and different dimensions of effects to filter the data records. The system doesn't contain any visualization stuff, it is more like a documented database, so we decided to implement a visualization interface to help users better understand the trends of earthquakes in recent years and the severe effects the hazard has caused.

Related Work

When we were exploring ideas for visualizing earthquake distribution, we did some research on different kinds of map views.

The “Earthquakes Since 1900 Visualizing location, Magnitude, and Quantity” visualization (http://www.princeton.edu/~efeibush/earthquakes/)
Has provided us the idea of how to visualize all the earthquakes on the map. We decide to show all earthquakes on the world map and add the interactions like zoom and move the map to have a better performance. We also designed the filters for users who just want the specified years and primaries.

The “OECD Regional Well-being” visualization (https://www.oecdregionalwellbeing.org/US49.html).
Which appeals in our midterm exam, has brought us insightful thoughts on how to make interaction between map points and info boxes. This visualization also provides us the ideas about how to set up the layout so that our visualization would be clearer and more acceptable for the users. We implemented this in our further designs.

Project Objectives

We are doing this project to answer the following questions:

1. When and where did the earthquake happen?
   We would like to implement a mapview to demonstrate where the earthquake happened. Specific information like time and coordinates will be displayed in an info box.

2. What's the difference between magnitudes?
   We would like to use different colors of the earthquake points to encode different magnitudes.

3. How many people suffered from the earthquake?
4. The total damage to the economy by the earthquake?
5. The damage to cities, especially houses?
   We will crunch the data and visually show the loss in each category (people/money) of earthquakes.

6. Which countries have the most possibilities to have earthquakes?
   We would like to have an integrated view to display the number of earthquakes that happened in each country. It will be helpful understanding the facts that which country is suffering most from earthquakes.

7. How does it easier to know which are the earthquakes that the users want to know most?
   We would like to distinguish earthquakes by different colors and also create certain filters which allow users to select their desired records.

8. Earthquakes cause Tsunamis sometimes, how do we know that?
   We would like to implement a checkbox which allows users to filter out earthquakes with tsunamis.

9. Which years have the most earthquakes, do they have patterns?

We will create a summary view to show the number of earthquakes that happened each year, which would drop insights on the trends of earthquakes worldwide.

Data

Data Citation

The data source is National Centers for Environmental Information:

https://www.ngdc.noaa.gov/hazel/view/hazards/earthquake/search

A rough description of the dataset is presented on the website as below:

''The Significant Earthquake Database contains information on destructive earthquakes from 2150 B.C. to the present that meet at least one of the following criteria: Moderate damage (approximately $1 million or more), 10 or more deaths, Magnitude 7.5 or greater, Modified Mercalli Intensity X or greater, or the earthquake generated a tsunami.''[1]

Data Example

Each row in the data represents a documented earthquake, attributes include:

ID: the unique id identifies each earthquake

Flag Tsunami: a flag marks if the earthquake generates Tsunami

Date: Year, Month, Day, Hour, Minute, Second

Earthquake Location: Country, State, Location Name, Latitude, Longitude, Region Code

Focal Depth: The depth of the earthquake is given in kilometers.

Magnitude: MS, MW, mb, ML, MFA, Unknown (different measurements of magnitude), we will use eq_primary which is the greatest value among these magnitudes.

Modified Mercalli Intensity (MMI): The Modified Mercalli Intensity (Int) is given in Roman Numerals (converted to numbers in the digital database)

Earthquake Effects:

Number of Deaths, Deaths Description

Number of Missing, Missing Description

Number of Injuries, Injuries Description

Damage Millions of Dollars, Damage Description

Number of Houses Destroyed, Houses Destroyed Description

Number of Houses Damaged, Houses Damaged Description

Total Effects (Earthquake and Tsunami, Volcano, etc.):

Total Number of Deaths, Total Deaths Description

Total Number of Missing, Total Missing Description

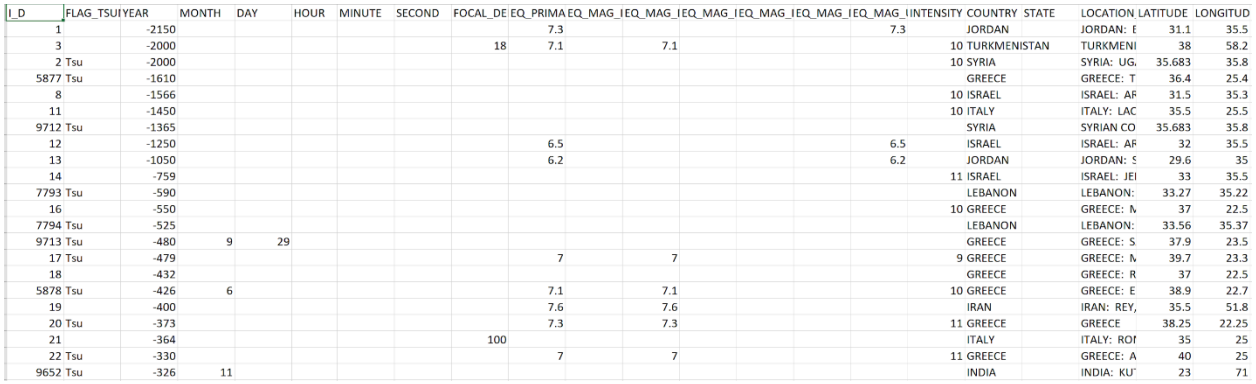
Total Number of Injuries, Total Injuries Description

Total Damage in Millions of Dollars, Total Damage Description

Total Number of Houses Destroyed, Total Houses Destroyed Description

Total Number of Houses Damaged, Total Houses Damaged Description

A screenshot of data is as below:



I_D	FLAG_TSU	YEAR	MONTH	DAY	HOUR	MINUTE	SECOND	FOCAL_DE	EQ_PRIMA	EQ_MAG_I	EQ_MAG_I	EQ_MAG_I	EQ_MAG_I	EQ_MAG_I	INTENSITY	COUNTRY	STATE	LOCATION	LATITUDE	LONGITUD
1		-2150							7.3					7.3		JORDAN		JORDAN: E	31.1	35.5
3		-2000						18	7.1	7.1					10	TURKMENISTAN		TURKMENI	38	58.2
2	Tsu	-2000													10	SYRIA		SYRIA: UG,	35.683	35.8
5877	Tsu	-1610														GREECE		GREECE: T	36.4	25.4
8		-1566													10	ISRAEL		ISRAEL: AF	31.5	35.3
11		-1450													10	ITALY		ITALY: LAC	35.5	25.5
9712	Tsu	-1365														SYRIA		SYRIAN CO	35.683	35.8
12		-1250							6.5					6.5		ISRAEL		ISRAEL: AF	32	35.5
13		-1050							6.2					6.2		JORDAN		JORDAN: S	29.6	35
14		-759													11	ISRAEL		ISRAEL: JEI	33	35.5
7793	Tsu	-590														LEBANON		LEBANON:	33.27	35.22
16		-550													10	GREECE		GREECE: M	37	22.5
7794	Tsu	-525														LEBANON		LEBANON:	33.56	35.37
9713	Tsu	-480	9	29												GREECE		GREECE: S	37.9	23.5
17	Tsu	-479							7	7					9	GREECE		GREECE: M	39.7	23.3
18		-432														GREECE		GREECE: R	37	22.5
5878	Tsu	-426	6						7.1	7.1					10	GREECE		GREECE: E	38.9	22.7
19		-400							7.6	7.6						IRAN		IRAN: REY,	35.5	51.8
20	Tsu	-373							7.3	7.3					11	GREECE		GREECE	38.25	22.25
21		-364						100								ITALY		ITALY: ROI	35	25
22	Tsu	-330							7	7					11	GREECE		GREECE: A	40	25
9652	Tsu	-326	11													INDIA		INDIA: KU	23	71

Data Processing

The dataset is open to the public and has been pre-formatted by columns so we do not need to do much modification. For the sake of more insights on recent earthquakes, we plan to take a subset of the year 1700 to 2020 from the original dataset. The original data has records since the year -2150, which is quite unnecessary for our research purpose, so we did clean up the data. Besides, there are unnecessary columns in the original dataset for our project, they will be removed as well.

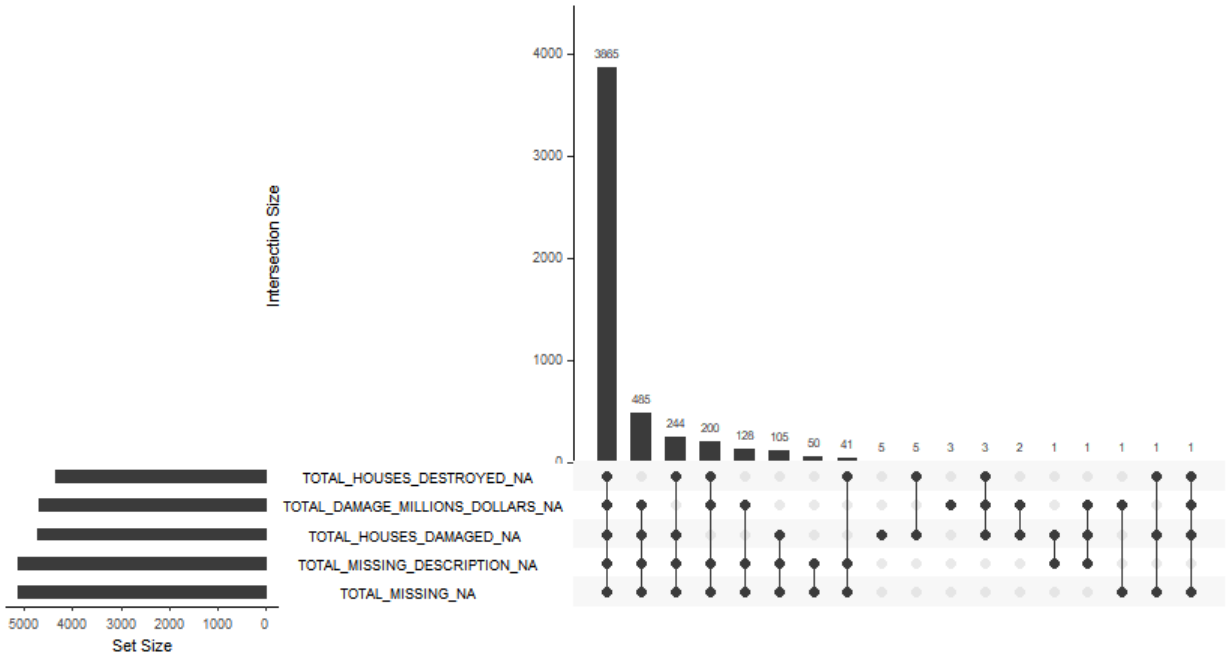
The remaining attributes after cleanup are i_d, flag_tsunami, year, eq_primary, country , latitude, longitude, region_code, total_deaths, total_deaths_description, total_missing, total_missing_description, total_injuries, total_injuries_description, total_damage_millions_dollars, total_damage_description, total_houses_destroyed, total_houses_destroyed_description, total_houses_damaged and total_houses_damaged_description

The data type is .csv.

Exploratory Data Analysis

In the data processing step, we have filtered the data to a smaller subset. There are still many blank cells in our desired columns like total_missing, total_deaths and total_injuries. We were trying to figure out how to deal with such records then R was used to perform the exploratory data analysis.

```
> eq %>%
+   select(everything()) %>%  # replace to your needs
+   summarise_all(funs(sum(is.na(.))))
  ID FLAG_TSUNAMI YEAR EQ_PRIMARY COUNTRY LATITUDE LONGITUDE REGION_CODE TOTAL_DEATHS
1  0            0    0       1131       0       38        33           1         3621
  TOTAL_DEATHS_DESCRIPTION TOTAL_MISSING TOTAL_MISSING_DESCRIPTION TOTAL_INJURIES
1                     3421          5121                      5120           3891
  TOTAL_INJURIES_DESCRIPTION TOTAL_DAMAGE_MILLIONS_DOLLARS TOTAL_DAMAGE_DESCRIPTION
1                       3715                          4689                     2239
  TOTAL_HOUSES_DESTROYED TOTAL_HOUSES_DESTROYED_DESCRIPTION TOTAL_HOUSES_DAMAGED
1                   4360                              3586                4718
  TOTAL_HOUSES_DAMAGED_DESCRIPTION
1                             4337
```



First, we explored the missing values in the dataset. From the first graph we could see that only 4 columns do not contain NAs, then we summarized the top 5 columns with most missing values through the second graph. We could tell that:

1. TOTAL_MISSING has the most missing values.
2. There are 3862 cases where TOTAL_HOUSES_DESTROYED, TOTAL_DAMAGE_MILLIONS_DOLLARS, TOTAL_HOUSES_DAMAGED,

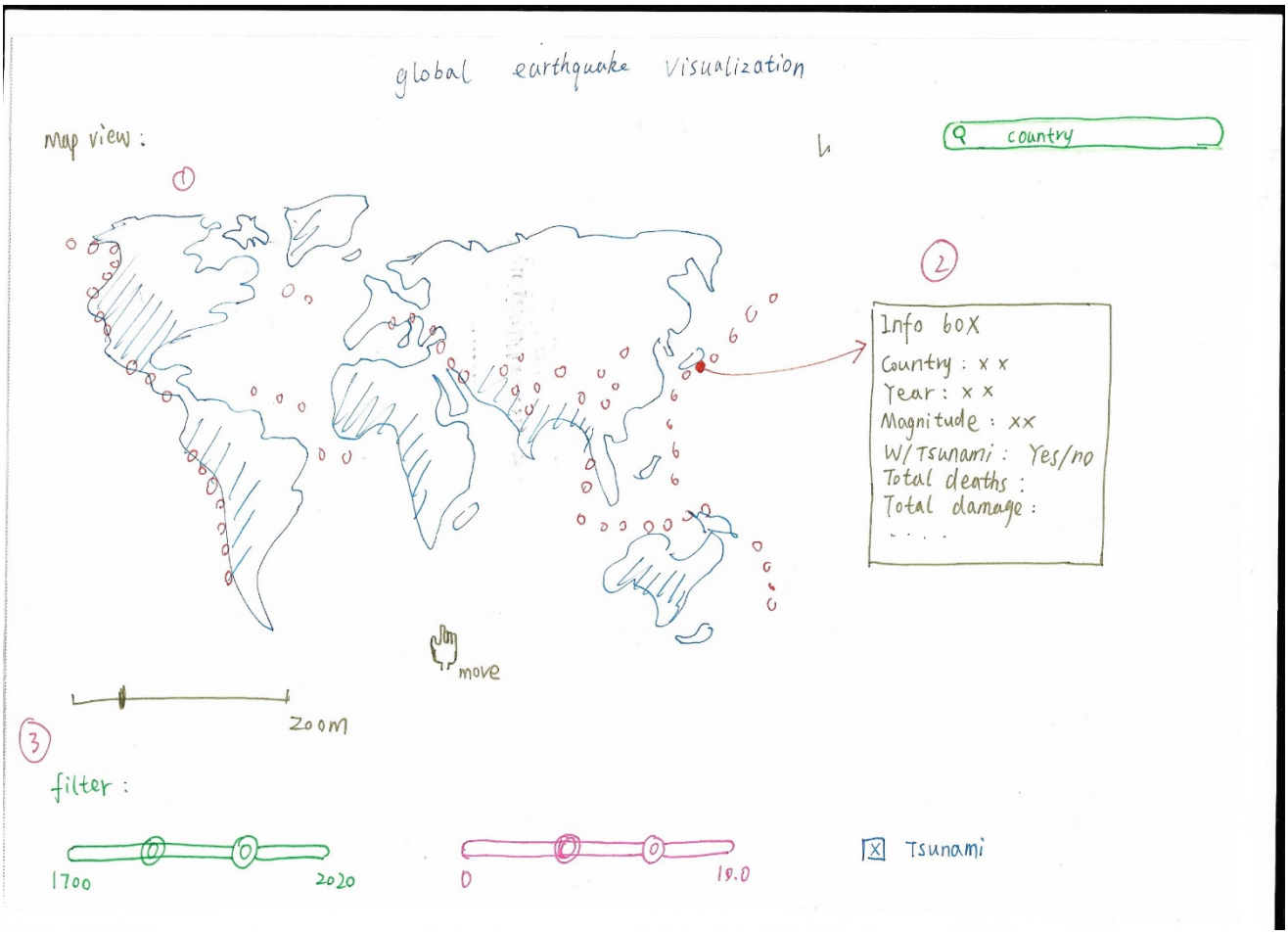
TOTAL_MISSING_DESCRIPTION and TOTAL_MISSING all have missing values together.

Given that there are only a few missing values in LATITUDE and LONGITUDE column, almost all the earthquake points can be shown in our map. Instead of removing records with missing values, we decided to only display top 5 records for each loss category in the loss views.

Design Evolution & Implementation

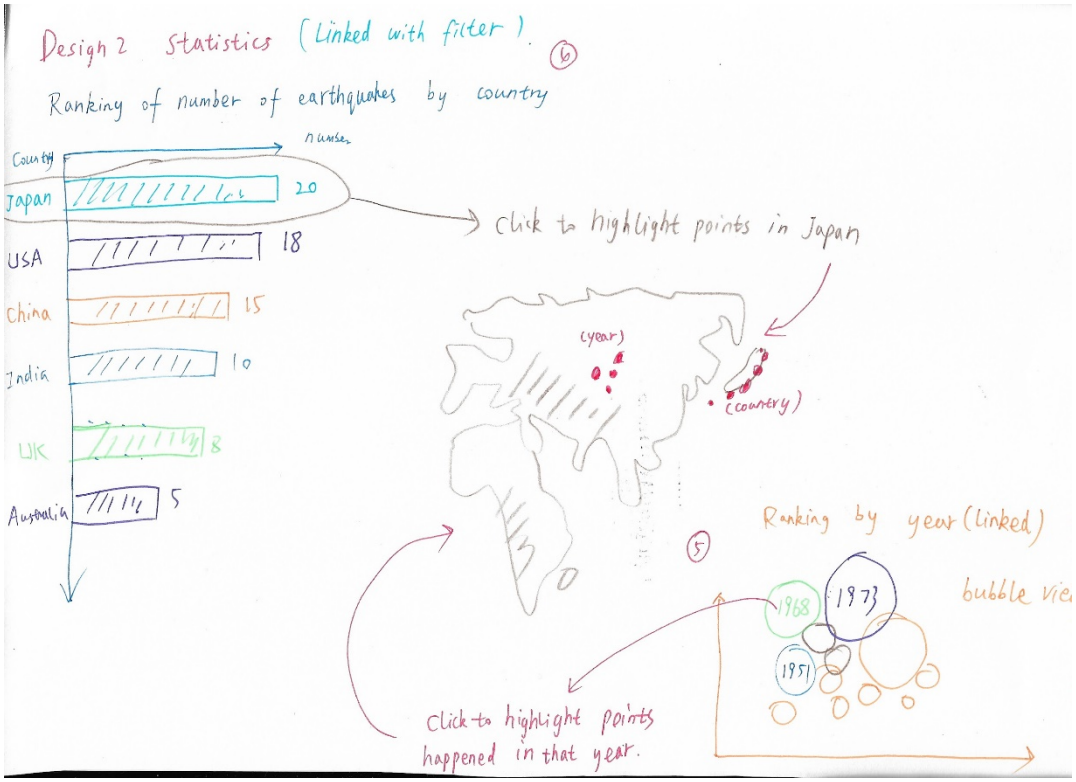
Original Design

Map view



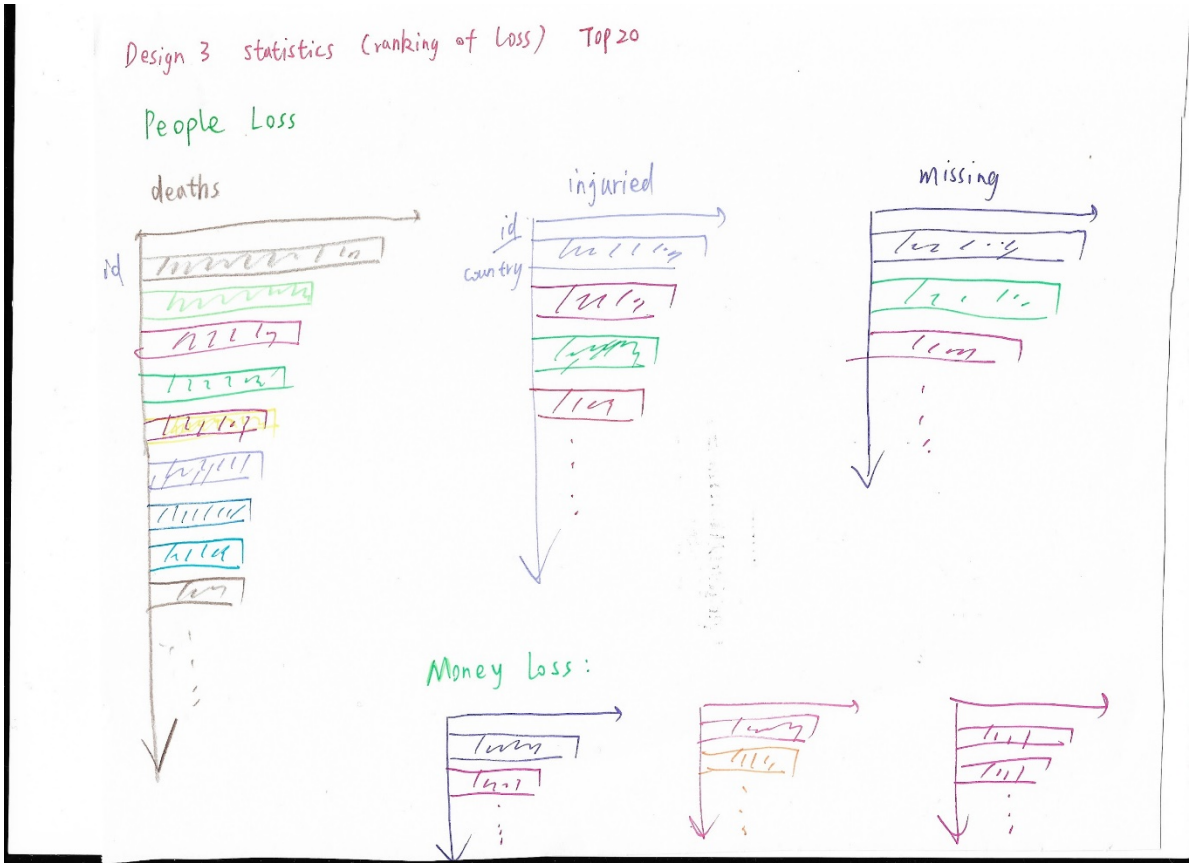
to show the earthquakes, with the search bar and selection filter, it is showing the earthquake by the year and primary. We can also see the details of each earthquake by clicking on it. The filters will be implemented as sliders. The info box will be put beside the map view.

Rank view



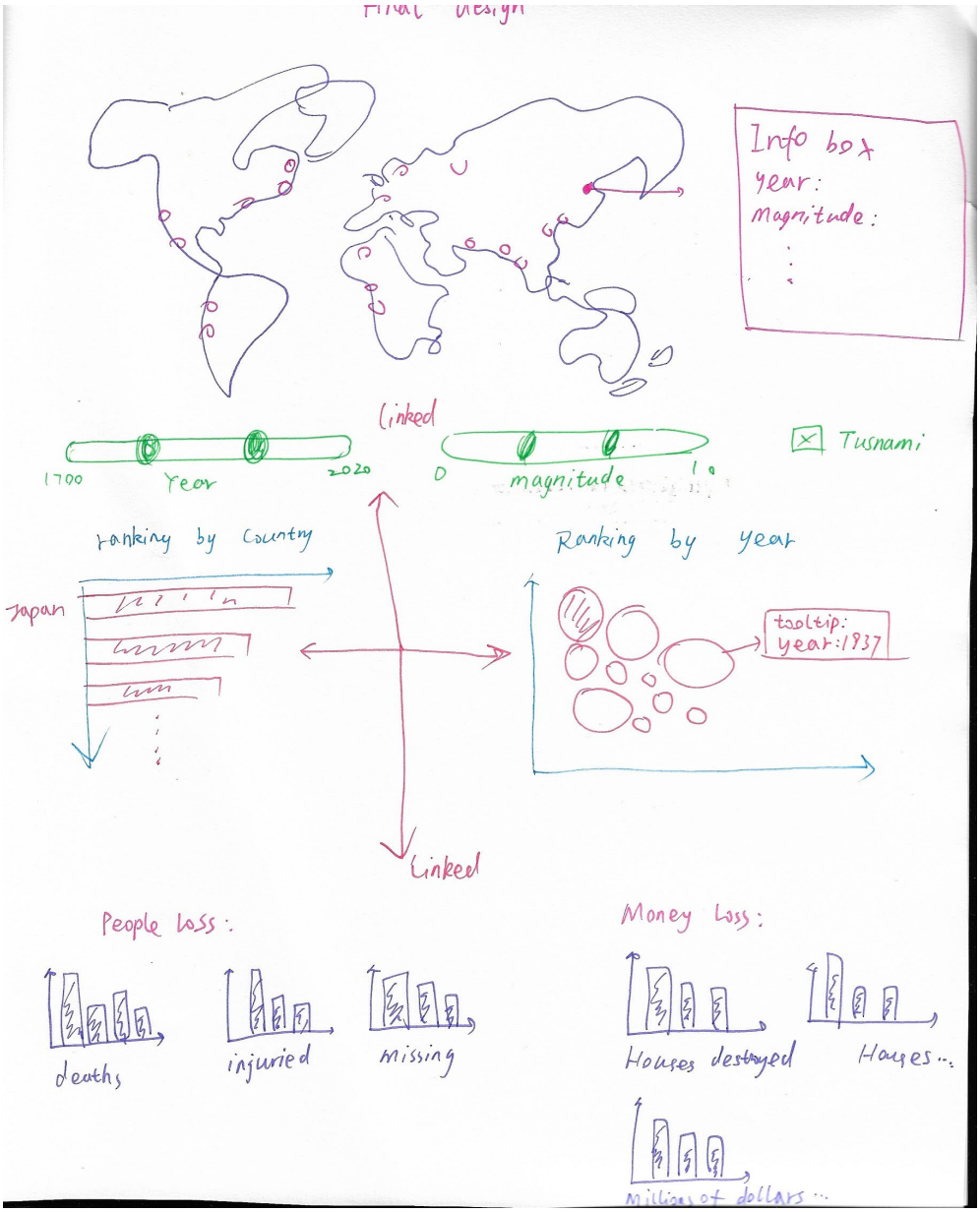
We planned to show the view of rank with bar charts showing the top 10 countries with the most earthquakes, and you can click each country to highlight the earthquakes which happened in this country at the world map. We also designed a bubble chart to show the numbers of earthquakes that happened each year. The two integrated views are linked with that map. While the map view is updated, these two views will be updated correspondingly.

Damage view

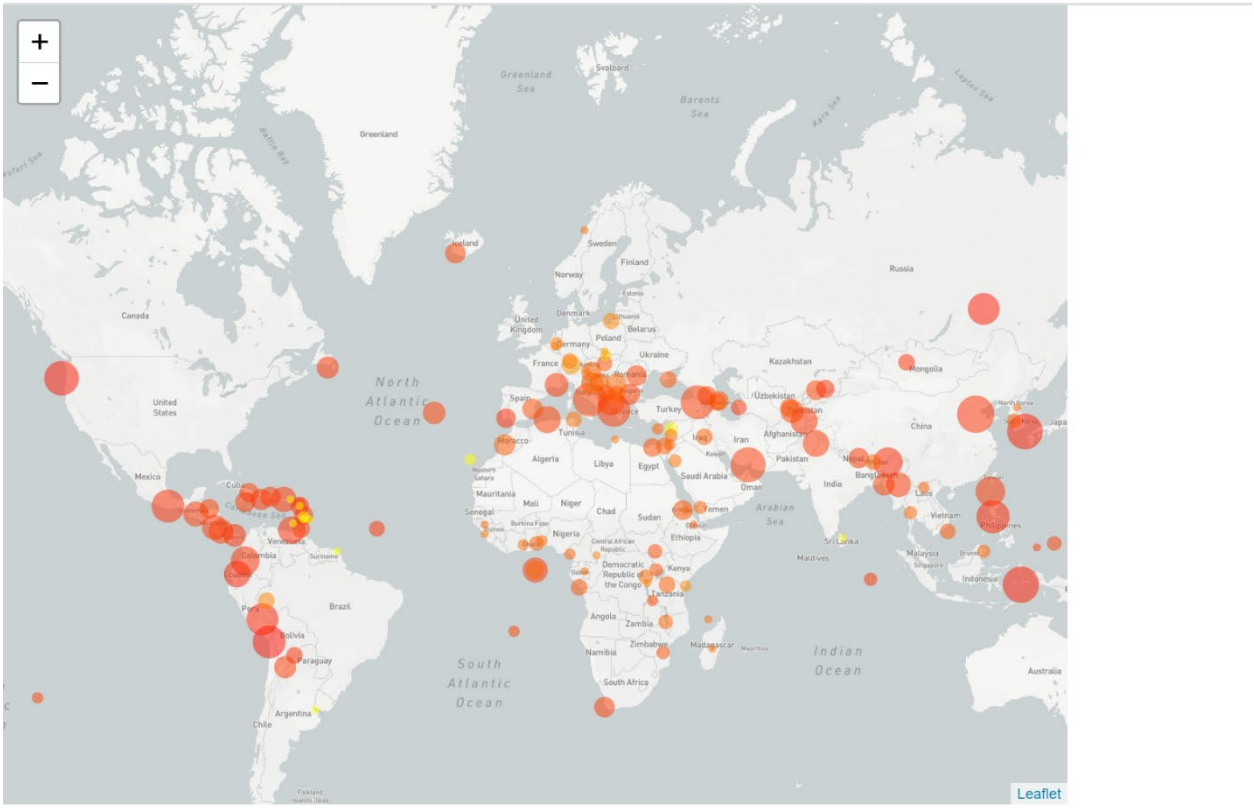


In the view of showing damage, we would like to create individual views for each loss category. The loss will be grouped by country and then sorted. The bars will be clickable and linked with the map view. If you click on a country, all earthquakes belonging to that country will be highlighted.

Milestone Changes



Our milestone design is shown above. After placing a slider with over 300 year's range, we realized that it is not a decent way to do that. We switched to a user input to filter time range to put filter stuff inside the table.js which would be grabbed out later for user purpose.



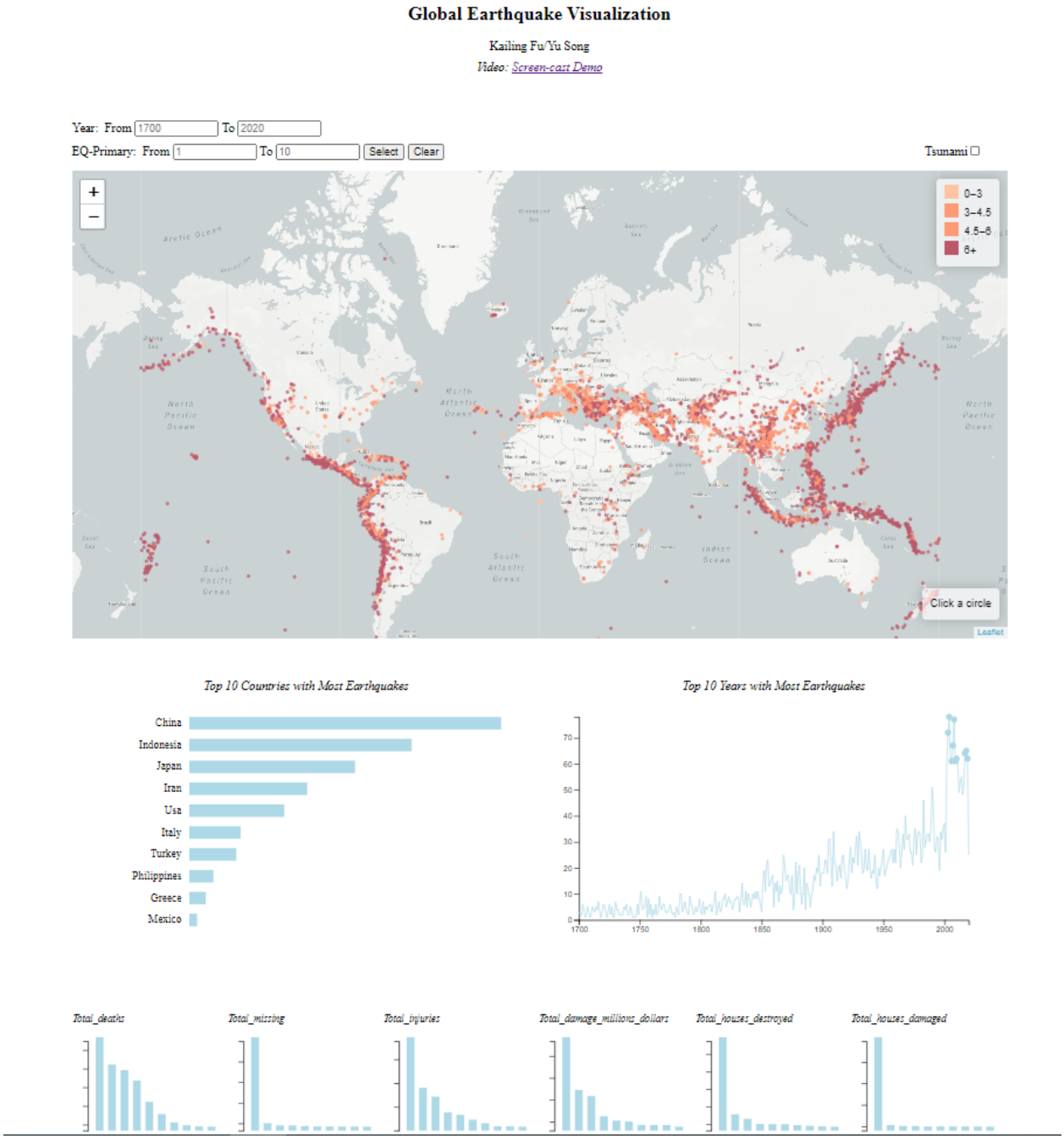
Year	Deaths/Missing/Injuries	Houses Destroyed/Damaged	Total Damage (M)
1703			
1783			
1783			
1783			
1846			
1857			
1905			
1908			116
1915			60
1919			
1928			
1930			
1962			
1968			320

Above is our milestone1 real design. When we were using Leaflet to deploy the current map view, we noticed that it would take a long time to load data points if a record is represented as a point in the map. So, in milestone one, we used a country-point map view instead, we grouped

the data by country and assigned circles of different sizes which encode different numbers of earthquakes that happen in that country. Below the map view is a bar chart showing all earthquakes and effects for the selected country. As you could see that there are many blank data points, that is why we need exploratory data analysis in advance.

Final version

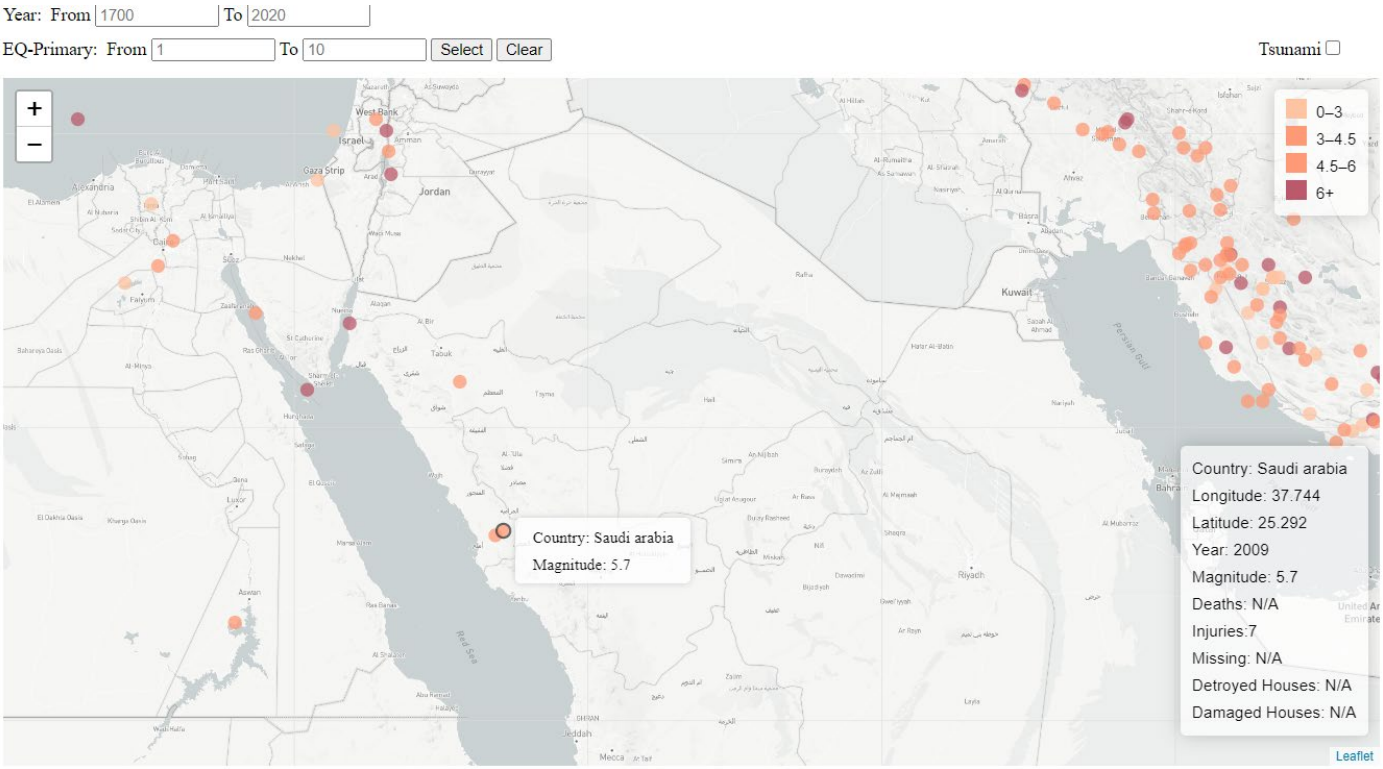
Overview



This is our overview of the final version. Different from the original design and milestone changes, we kept the color simple for a better viewing experience, and we added a view to show

the total earthquake numbers in our selected year range, and showing the top 10 years with the most earthquakes by showing the dot on the line chart.

Map view

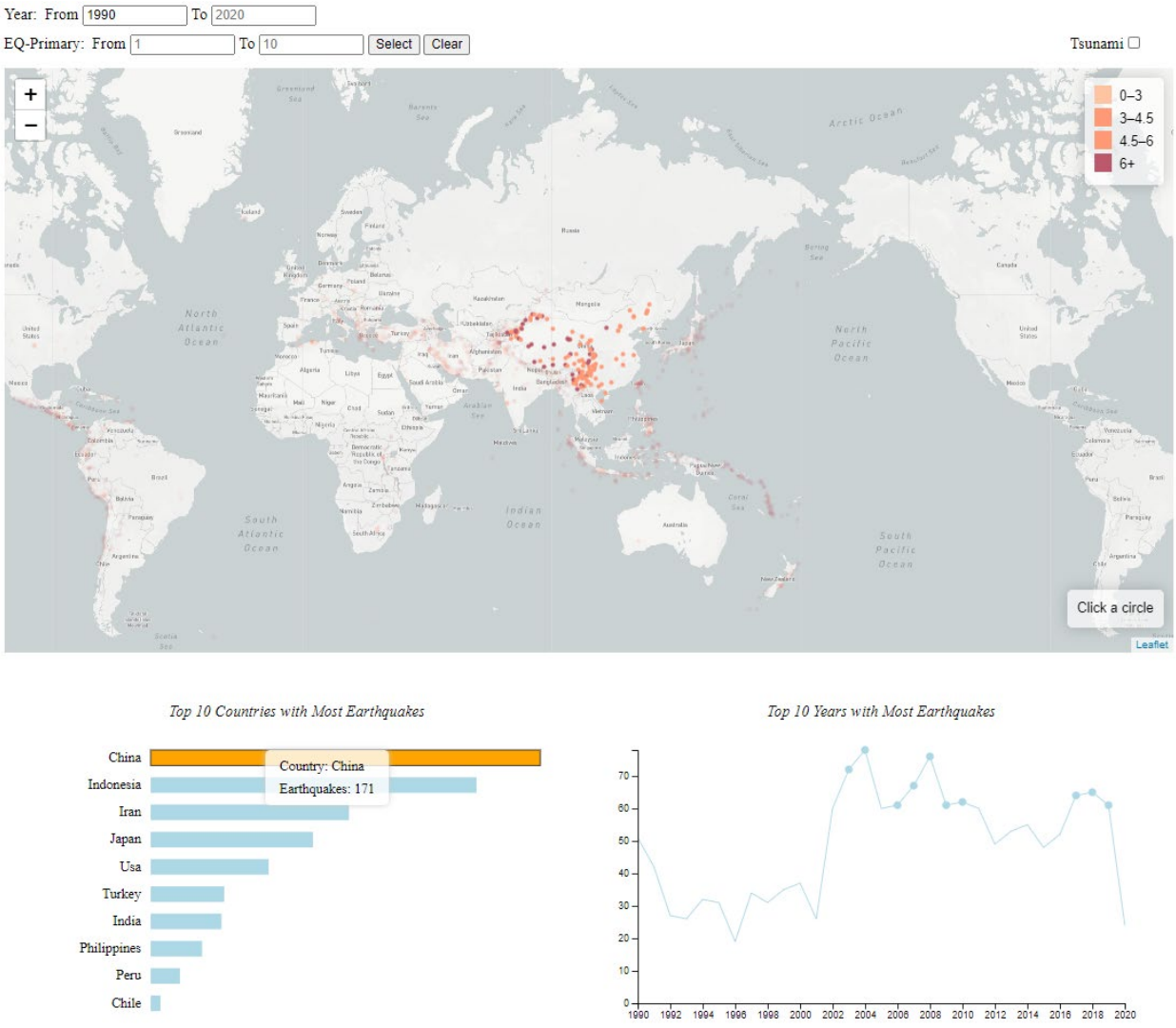


In the final map view, we have a world map with the zoom and move functionalities. All the earthquakes are shown on the map. Each earthquake has different colors to represent the magnitude, and the tooltips to show the country and exact magnitude of this earthquake. We clicked; the information of that unique earthquake will appear at the bottom right of the map view.

Different from the original design, we changed the search bar to the user-input boxes for accuracy and project performance: the slider bar is lagging and cannot navigate to the exact year (we have 320 years in total) in most cases.

The filter of tsunami has been designed to a checkbox as well, when clicked, there will be only the earthquakes that caused tsunamis left on the map.

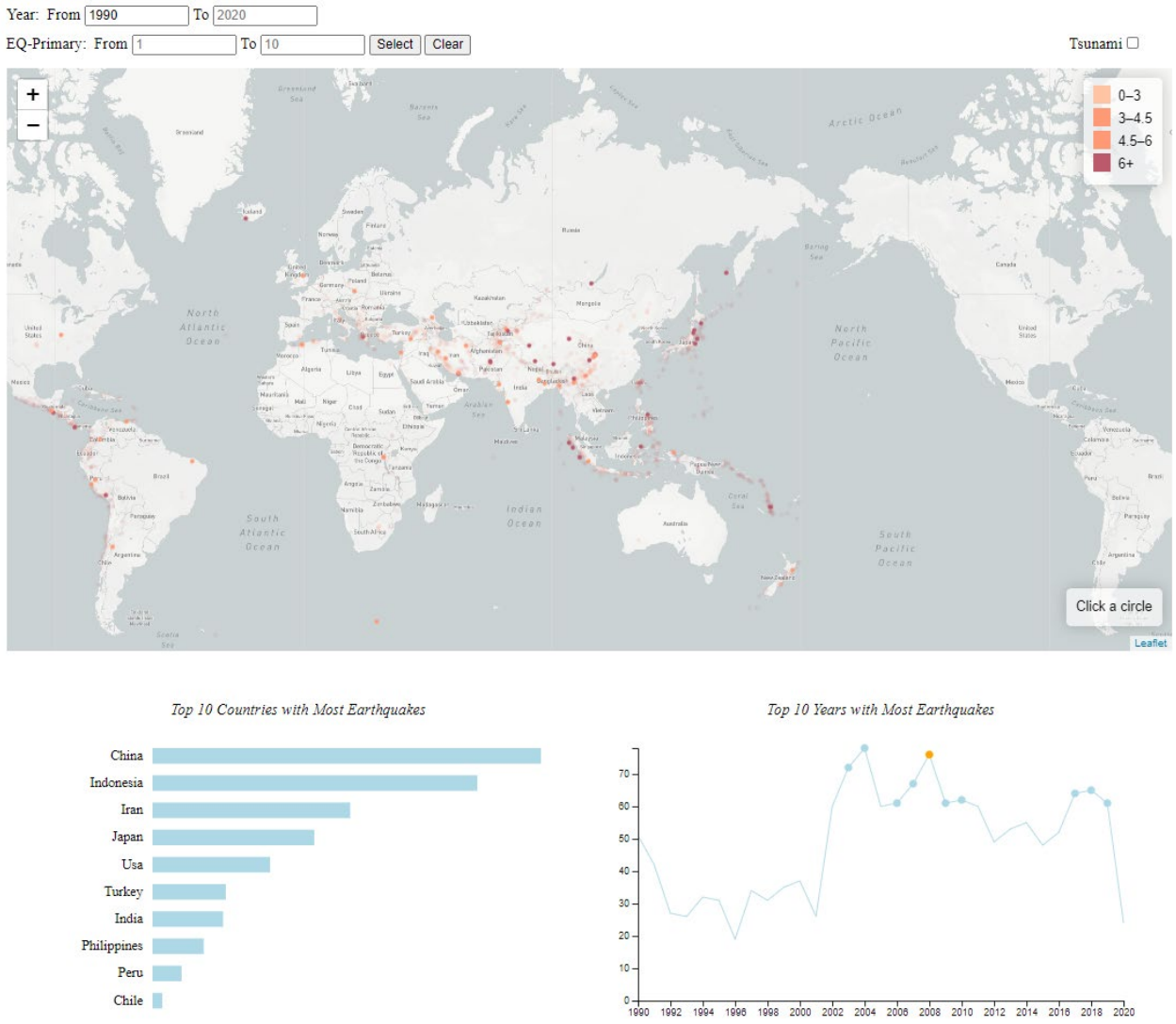
Country ranking & Year ranking



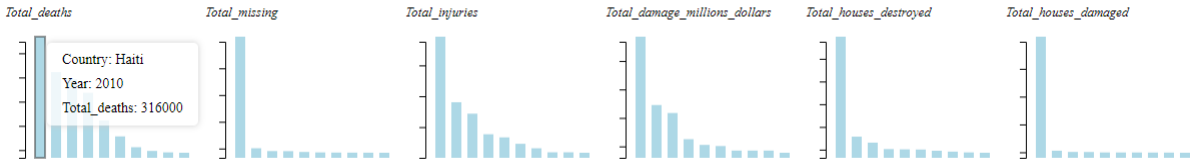
In our ranking for countries, we put the top 10 countries with the most earthquakes in a selected range and rank them in a bar chart. Each bar is selectable, and when clicked, all the earthquakes in that country within the selected range would be highlighted for a better understanding of how many earthquakes happened in this country. The country bar itself has tooltips showing the country name and exact number of earthquakes as well.

On the bottom right is the year ranking. This view shows a line chart of the number of earthquakes in each year within the selected range. For the top 10 years with most earthquakes,

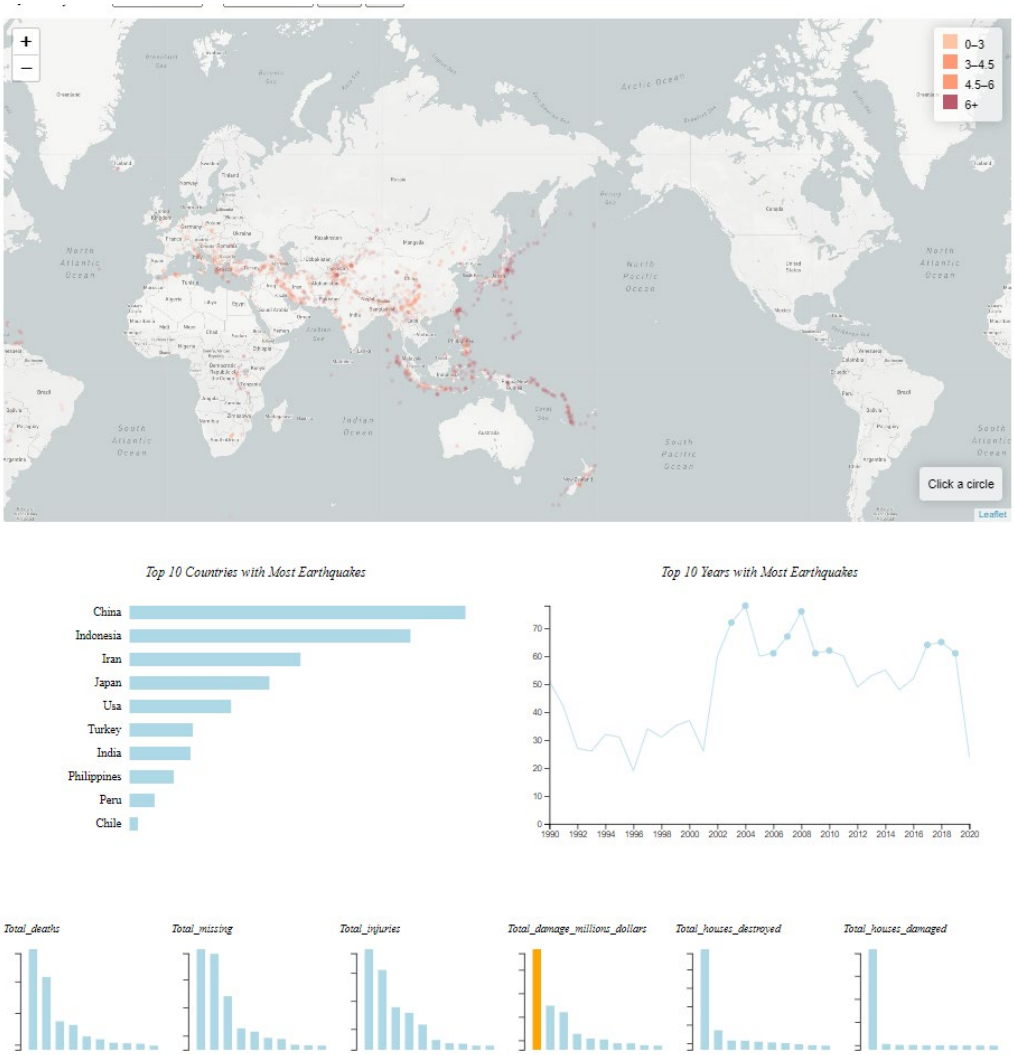
we did a circle for each of them. The circles are also interacting with the world map, it highlights the earthquakes in the year of the circle when clicked (shows below).



Loss counting



The view of loss counting has 6 bar charts showing the top 10 earthquakes that caused most damage. The charts are showing total death, total missing, total injuries, total damage(millions of dollars), total house destroyed and total house damaged information. Each bar has a tooltip to show the country name, Earthquake year and the damage it made.



The interaction of loss counting is the highlight, everytime a bar is selected, the only one earthquake will be highlighted on the map.

Evaluation

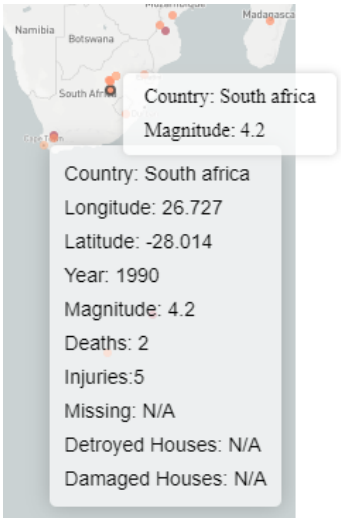
Our visualization helps answering the questions discussed in the section ‘project objectives’.

- When and where did the earthquake happen?

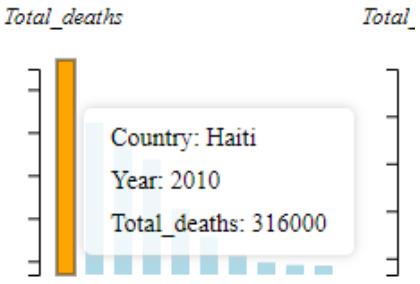
  This becomes the main issue of our filter and info box part in our project.

  For the year selection, we did the input box for year range to filter the data we need.

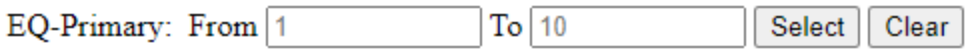
  

  We have signed up the information for every earthquake, make them clickable and provide an info box at the right bottom corner of our map to show the detailed information of each earthquake.

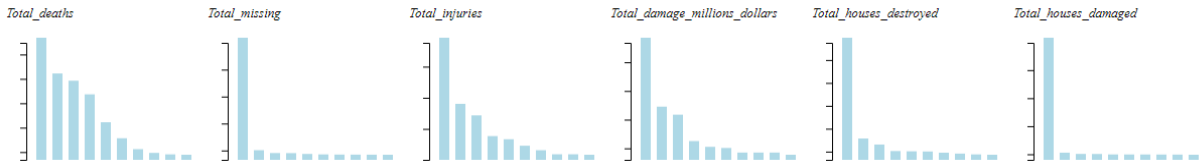
  We did the highlight the earthquakes in that selected year on the mapview for top 10 damages when choose the bar

- What is the difference between magnitude?



We transformed the answer to the filter system in our program: users can input the EQ-Primary level to negative the magnitude of earthquakes they want.

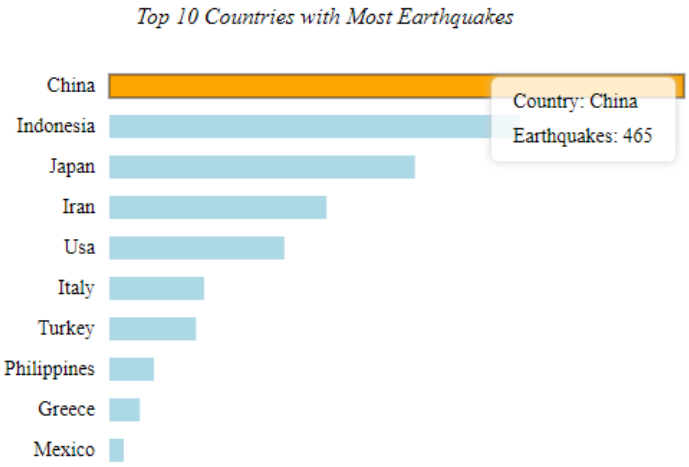
- How many people suffered from the earthquake?
- The total damage to the economy by the earthquake?
- The damage to cities, especially houses?



Answers to question 3, 4 and 5 have become our view of our top 10 count view of total death, total missing, total injuries, total damage(millions of dollars), total house destroyed and total house damage.

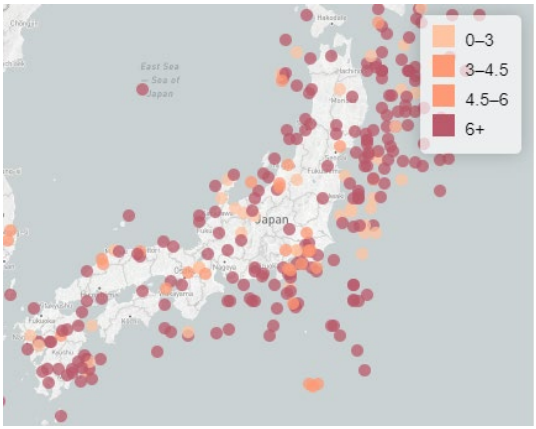
With the highlight of a significant earthquake on the map. In this way we can have a better idea of how earthquakes influence our human culture.

- Which countries have the most possibilities to have earthquakes?



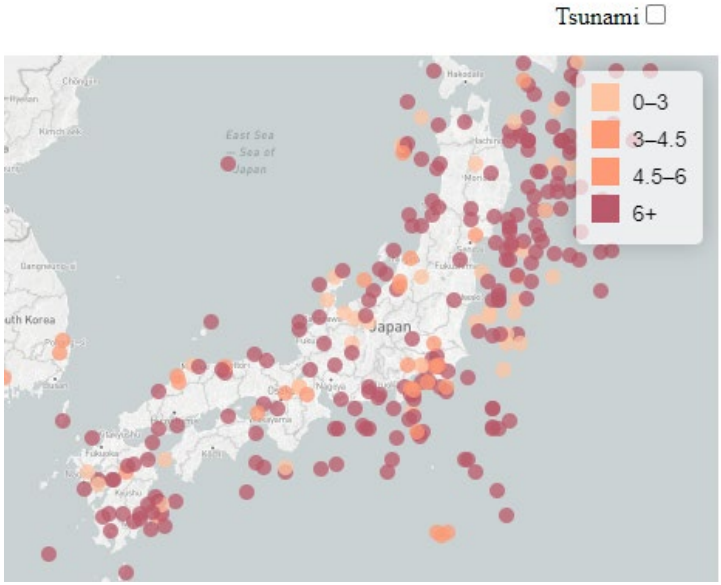
This question became our ranking of top 10 countries that have the most earthquakes, the bar chart can be interacted with the highlights on map as well.

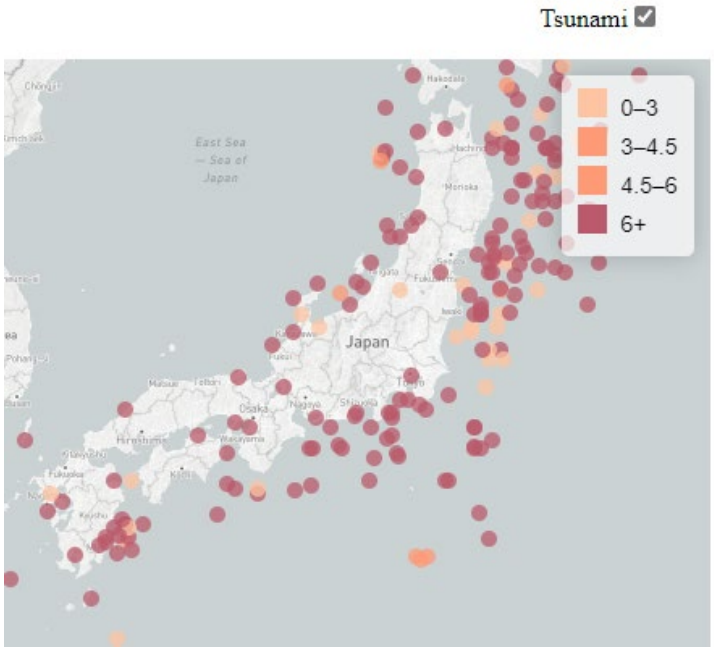
- How does it easier to know which are the earthquakes that the users want to know most?



We designed our map view with showing the earthquakes with different levels of red to show the different magnitudes, so that the users can easily pick the earthquake they want to know about.

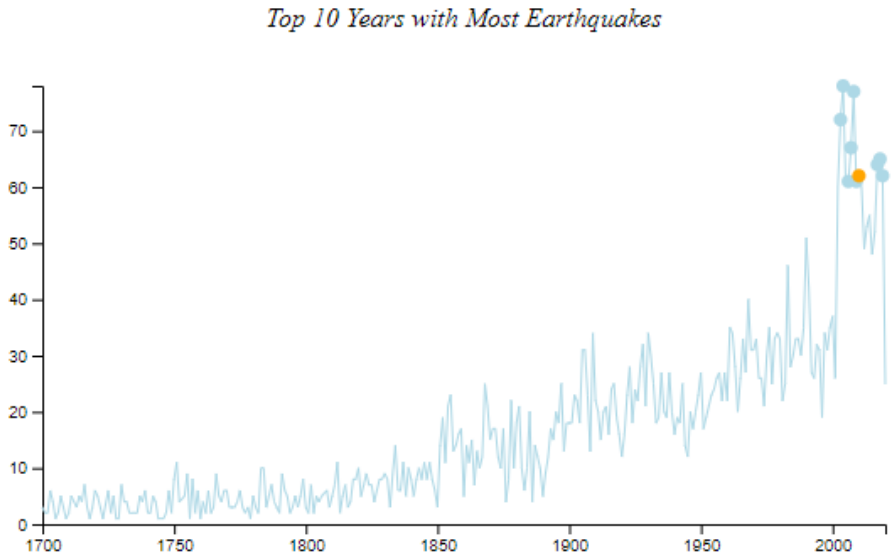
- Earthquakes cause Tsunamis sometimes, how do we know that?

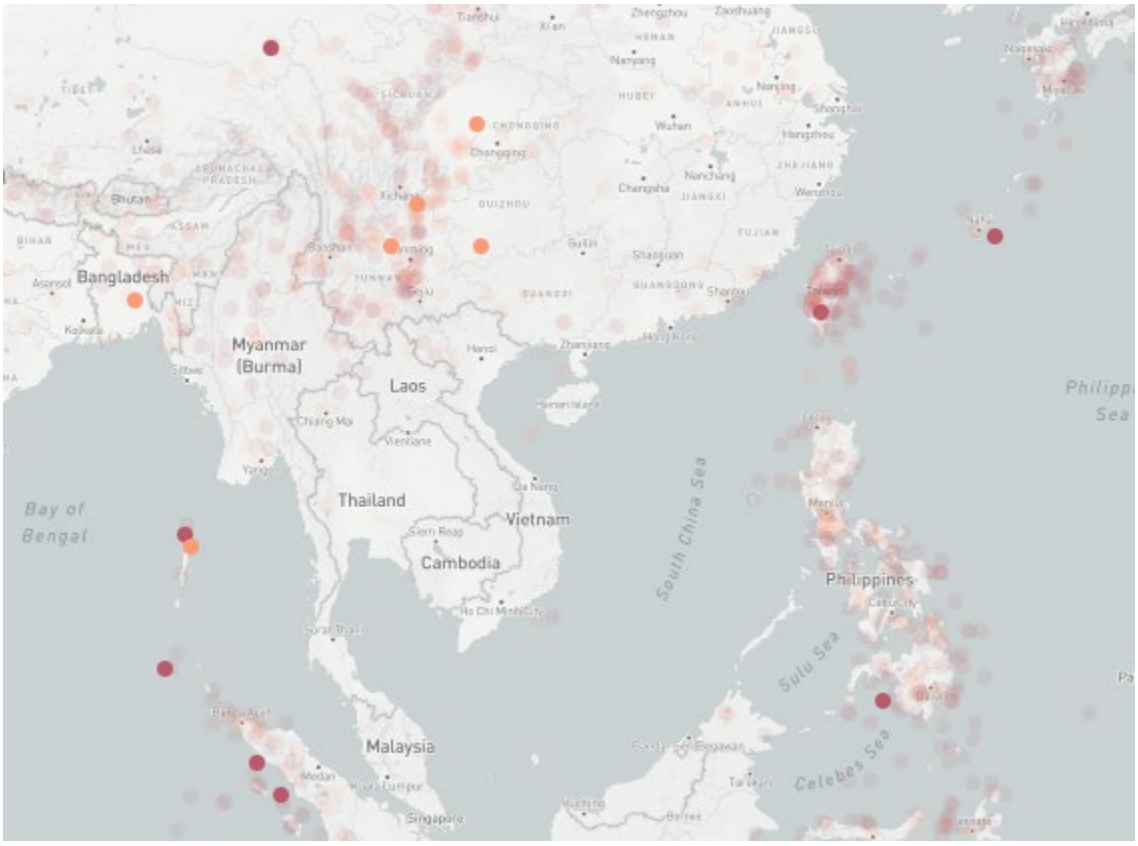




We actually added the filter to show the earthquakes which cause Tsunamis, and it is more direct for the users to see how much side effects the earthquakes can make, and which areas are getting more dangerous when the earthquake happens.

- Which years have the most earthquakes, do they have patterns?





This becomes our view of the top 10 years with the most earthquake view, and shows the numbers of all earthquakes in each of our selected years, with the circles representing the largest 10 years. Choose on the circle just highlights earthquakes in that year on the map view. This helps us to understand the patterns of regularity for the frequency of earthquakes.

Our project works in a decent way and all the functionalities are performed well.
The visualization shows all the information about the earthquakes all over the world from 1700 to 2020. It provides choice for the users to zoom in/out and move the map, and helps the users to target the earthquakes in specific years and magnitudes (and also with/without the cause of tsunami) they want to know. Nevertheless, the visualization provides multiple ranking charts for countries who have the top 10 number of earthquakes, top 10 earthquake years and top 10 most serious earthquakes(loss of human, money and houses) in the user-choice range of year and magnitudes. We did pretty much decent interactions with the map view and analysis view as well.

One of the inconveniences of our project is that our data analysis is limited to top 10, because too many ranking items may cause the whole visualization to become chaotic. The helpful way to improve the ranking range could be: let the users choose how many items they want to show in the visualization.

The other improvement could be not enough freedom to choose the area. Currently we have the choice to highlight earthquakes in a top 10 country listed in our analysis view, but we cannot insert the country name, and we cannot choose earthquakes in a continent, we may add the functionalities to do such operations in future improvements.