CS 6630 - Data Visualization

Process Book

Global Earthquake Visualization

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

By,

Kailing Fu, Yu Song

Background and Motivation

Earthquakes are one of the devastating disasters worldwide and there is no precaution from it. People have been trying to discover the underlying pattern of where and when an earthquake will 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 NCEI/WDS Global Significant Earthquake Database was used in our project to explore the distribution and effects of earthquakes worldwide since 1700.

There is already 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 for ideas for visualizing earthquake distribution, we did some research on different kinds of map views like http://www.princeton.edu/~efeibush/earthquakes/ and https://www.oecdregionalwellbeing.org/US49.html. The latter has brought us insightful thoughts on how to make interaction between map points and info boxes. 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?
2. What's the difference between primaries?
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?
6. Which countries have the most possibilities to have earthquakes?
7. What is the trend of areas that will have more earthquakes in future?

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 plan to clean up the data. Besides, there are unnecessary columns in the original dataset for our project, they will be removed as well.

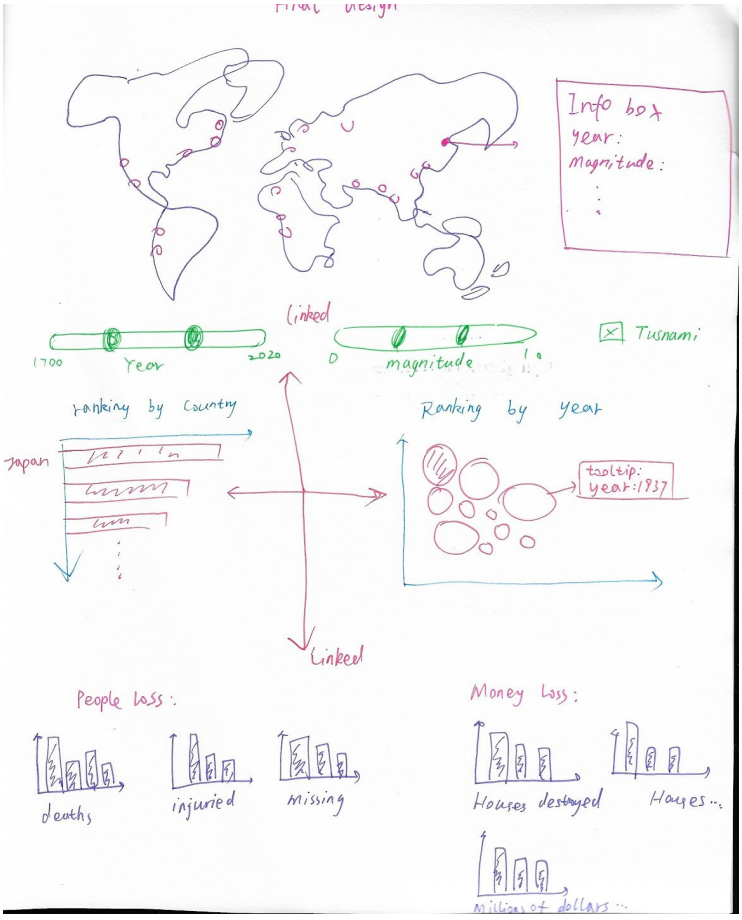
The remaining attributes after cleanup are i_d, flag_tsunami, year, eq_primary, country , latitude, longitude, region_code, total_deaths, total_deaths_description, total_missing, total_missing_description, total_injuries, total_injuries_description, total_damage_millions_dollars, total_damage_description,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

The data type is .csv.

Exploratory Data Analysis

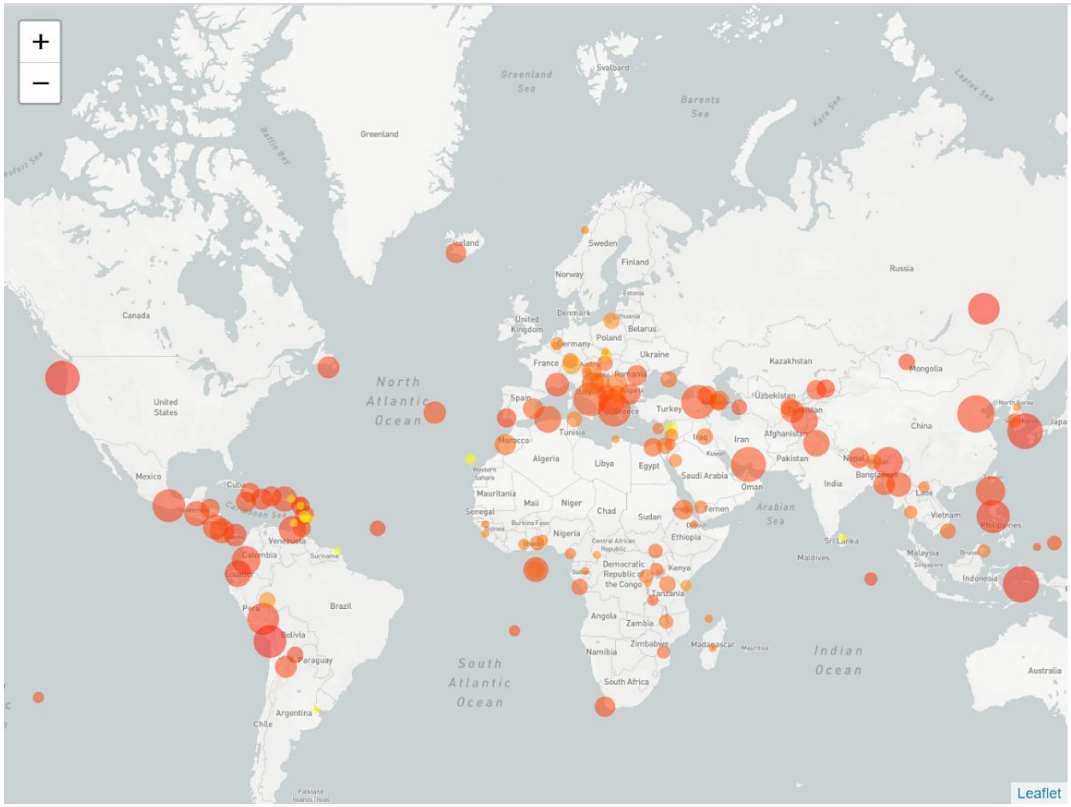
In the data processing step, we have filtered the data to a smaller subset. There are still many blank cells in our desired columns like total_missing, total_deaths and total_injuries. We are trying to figure out how to deal with such records. We will perform some exploratory data analysis through R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

Design Evolution



We are currently implementing our proposed design showing above. After placing a slider with over 300 year's range, we realized that it is not a decent way to do that. We are now switching to a user input to filter time range, the current design is to put filter stuff inside the table.js which will be grabbed out later for user purpose.

Implementation



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

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’s why we need exploratory data analysis in advance.