Dataset

Our dataset is as followed:
https://www.kaggle.com/datasets/jahaidulislam/significant-earthquake-dataset-1900-2023
The dataset records relevant information about 37000 earthquakes from 09/10/1900 to 17/02/2023. It includes several columns that provide information about each earthquake, such as its title, magnitude, date and time, maximum reported intensity, e.g. All the information is available and little to no cleaning is required. The data is ready to be pre-processed and analyzed.

Problematic

Frame the general topic of your visualization and the main axis that you want to develop.

- What am I trying to show with my visualization?
    - We are trying to show the location, the amount and the magnitude of the earthquakes throughout time with our visualization: a proportional symbol map (depending on the magnitude) where their centers are the earthquake location with a slider representing time that the user could move to see earthquakes through time.

    - Our goal is to communicate precise information regarding significant earthquakes of the 21st century and for the audience to be able to explore the information from the data freely.

- Think of an overview for the project, your motivation, and the target audience
    - Our target audience can be scientists working on geoscience, as earthquake visualization would be helpful for them to grasp the overall intensity and frequency of a region of interest. Moreover, the visualization can shed light on some hidden patterns (on a broad scale) not observable when the data is examined too closely. For example, we could see what is the cause of the earthquake through time and see what tectonic plates are moving at every point in time.

    - The overview of our project would be to see the evolution of devastating earthquakes throughout the 21st century. Which will enable us to shed light on why some earthquakes happen by surprise. Earthquake visualization is an important tool for experts to discover patterns and to predict the next earthquakes which will enable them to take preventive measures to minimize the damages that an earthquake could cause.

Exploratory Data Analysis

Pre-processing of the data set you chose

- Show some basic statistics and get insights about the data

  Here is the data format for every attributes of our dataset

```
RangeIndex: 23412 entries, 0 to 23411
Data columns (total 21 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Date                        23412 non-null  object
 1   Time                        23412 non-null  object
 2   Latitude                    23412 non-null  float64
 3   Longitude                   23412 non-null  float64
 4   Type                        23412 non-null  object
 5   Depth                       23412 non-null  float64
 6   Depth Error                 4461 non-null   float64
 7   Depth Seismic Stations      7097 non-null   float64
 8   Magnitude                   23412 non-null  float64
 9   Magnitude Type              23409 non-null  object
 10  Magnitude Error             327 non-null    float64
 11  Magnitude Seismic Stations  2564 non-null   float64
 12  Azimuthal Gap               7299 non-null   float64
 13  Horizontal Distance         1604 non-null   float64
 14  Horizontal Error            1156 non-null   float64
 15  Root Mean Square            17352 non-null  float64
 16  ID                          23412 non-null  object
 17  Source                      23412 non-null  object
 18  Location Source             23412 non-null  object
 19  Magnitude Source            23412 non-null  object
 20  Status                      23412 non-null  object
dtypes: float64(12), object(9)
```

  As we can see there is no attribute country so we are going to have to do preprocessing to infer the country based on the latitude and longitude of the earthquake location.

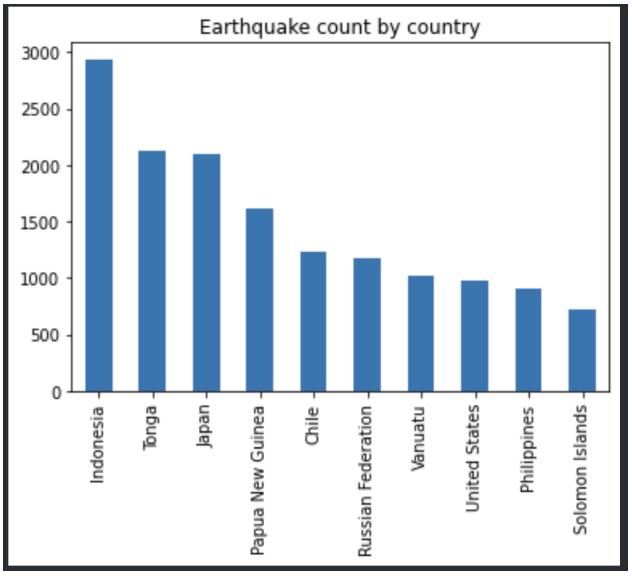
  Here we see the basic statistics per attributes:

	Latitude	Longitude	Depth	Depth Error	Depth Seismic Stations	Magnitude	Magnitude Error	Magnitude Seismic Stations	Azimuthal Gap	Horizontal Distance	Horizontal Error	Root Mean Square
count	23412.000000	23412.000000	23412.000000	4461.000000	7097.000000	23412.000000	327.000000	2564.000000	7299.000000	1604.000000	1156.000000	17352.000000
mean	1.679033	39.639961	70.767911	4.993115	275.364098	5.882531	0.071820	48.944618	44.163532	3.992660	7.662759	1.022784
std	30.113183	125.511959	122.651898	4.875184	162.141631	0.423066	0.051466	62.943106	32.141486	5.377262	10.430396	0.188545
min	-77.080000	-179.997000	-1.100000	0.000000	0.000000	5.500000	0.000000	0.000000	0.000000	0.004505	0.085000	0.000000
25%	-18.653000	-76.349750	14.522500	1.800000	146.000000	5.600000	0.046000	10.000000	24.100000	0.968750	5.300000	0.900000
50%	-3.568500	103.982000	33.000000	3.500000	255.000000	5.700000	0.059000	28.000000	36.000000	2.319500	6.700000	1.000000
75%	26.190750	145.026250	54.000000	6.300000	384.000000	6.000000	0.075500	66.000000	54.000000	4.724500	8.100000	1.130000
max	86.005000	179.998000	700.000000	91.295000	934.000000	9.100000	0.410000	821.000000	360.000000	37.874000	99.000000	3.440000

  We can also see that there are multiple NaN countries that we will have to handle with our visualization. After preprocessing we can see that they are 156 unique countries in our dataset over the 193 countries currently existing. We can also clearly see that the Indonesia is the country with the most earthquakes with a magnitude between 5.5 and 9.1

```
Indonesia              2938
Tonga                  2127
Japan                  2103
Papua New Guinea       1613
Chile                  1228
                        ...
Mozambique                1
Marshall Islands          1
Grenada                   1
Viet Nam                  1
Kenya                     1
```

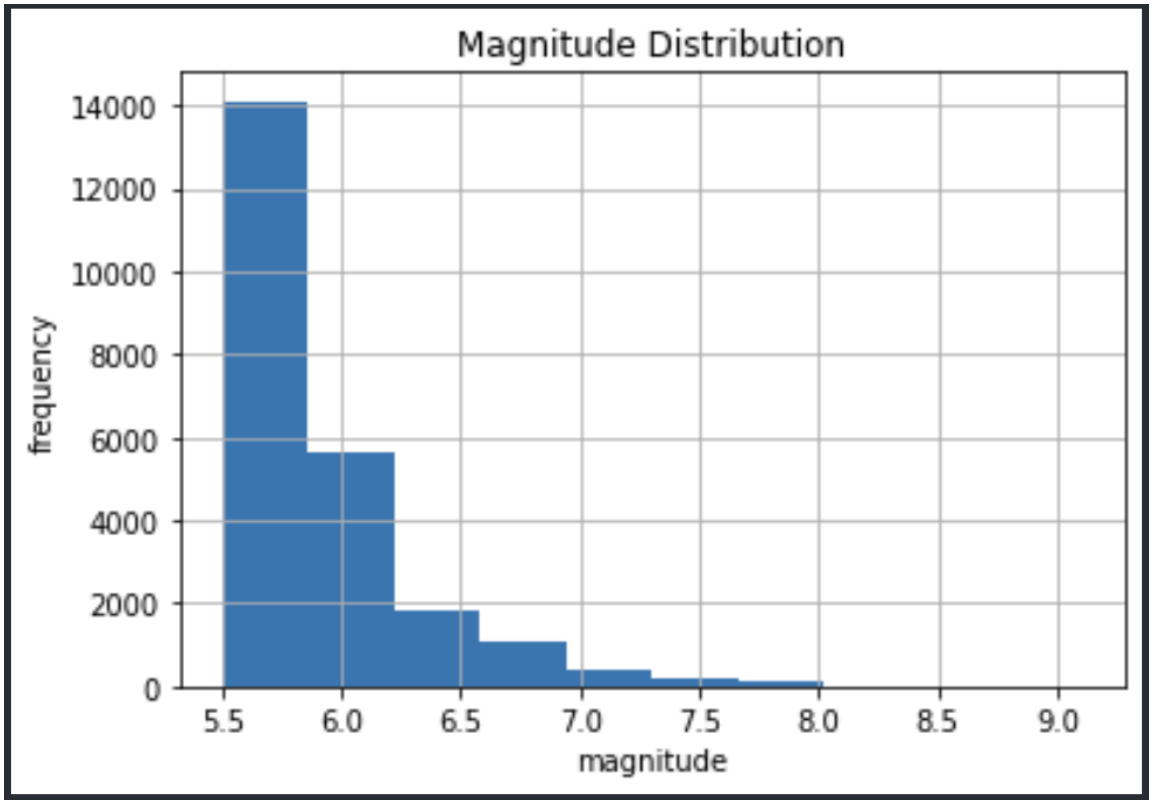
We can also see that the number of earthquakes of our dataset follows a distinct distribution of the number of earthquakes per country



We can also see the basics statistics of the number of earthquakes by countries

```
count     156.000000
mean      150.076923
std       401.325940
min         1.000000
25%         3.000000
50%        13.000000
75%        95.500000
max      2938.000000
Name: Country, dtype: float64
```

We can also see a clear pattern in the magnitude distribution. The 5.5-6.0 magnitude earthquakes are the most frequent.



Related work

- What others have already done with the data?

  This data set is a subset of data registered at The National Earthquake Information Center (NEIC), which determines the location and size of all significant earthquakes that occur worldwide. Our data set is used mainly by academics for data analysis and data visualization, specifically maps.

  In general, in terms of data visualization, a lot of designs have presented the data through a simple static map, where each point represents an earthquake based on latitude and longitude of its impact. Some designs show the magnitude of the earthquake through points of different sizes, others will use colors.

- Why is your approach original?

  We want to create an interactive experience where the user can see the most significant earthquakes of the 21st century. Our approach is original because we want to make a design that won't only use a map as the main visual support. We want our project to be a simple refreshing design that can better communicate the information of the dataset than what we see in current design.

- What source of inspiration do you take? Visualizations that you found on other websites or magazines (might be unrelated to your data).

We looked at different designs for inspiration, mostly at the ones that are using location and time, because that is partly what our dataset is based on.

One that inspired us is GlobalView: Climate Change in Perspective - Bloomberg View. This website guides the user through a story with a powerful interactive design. They use multiple types of data visualization such as tables and graphs in an elegant way. Also, the message they leave the user with is really powerful.

Finally, the website Wikiverse : Wikiverse: a galactic reimagining of Wikipedia which maps all the entries of wikipedia to all the ones that are mentioned in it, is another great design that inspires us. It shows that even with great amounts of data it is possible to make a design that is both pretty and usable. We really liked the visual effects of this data visualization, and that the user can use the mouse or the keyboard to interact with the design. The user experience is an aspect we shouldn't forget during the design process of our design.

- In case you are using a dataset that you have already explored in another context (ML or ADA course, semester project...), you are required to share the report of that work to outline the differences with the submission for this class.