Earthquakes from 1900 to 2013

Data Visualization, E23

Group 14

Group members:
Christoffer Knudsgaard, cknud20@student.sdu.dk
Frederikke Lan, fjoer20@student.sdu.dk
Nida Basaran, nibas18@student.sdu.dk
Tesniem El-Merie, teelm19@student.sdu.dk

GitHub repository: https://github.com/telmerie/DataVisualization
SHA of last commit: a7db4a64c7622077596b783fd928b468886612d1
R-Shiny app: https://tesniem.shinyapps.io/datavisualization/

Abstract

This report delves into the analysis of earthquakes taking place from 1900 to 2013, and it aims to explore the evolution of earthquakes over the past century. The study focuses on a data set with 14 columns and 8318 rows from Kaggle made by "United States Geological Survey". It investigates various aspects, including variables that are used to measure earthquakes, the changing patterns in the occurrences of earthquakes, and potential correlations with climate changes. The data set was carefully selected to ensure a great amount of data for meaningful visualizations. The study explores key questions such as measurement variables for earthquakes, the significant evolution of earthquakes over the years, changes in earthquake magnitudes, global earthquake distribution, consistencies or variation among earthquakes, and the potential influence of climate.

The report presents various visualizations including bar charts, scatter plots, time series plots and a world map with the help of RStudio and libraries such as shiny, ggplot2, DT, mapview and leaflet. The chosen color palette ensures accessibility for both colorblind and non-colorblind individuals. These are all visible in the Shiny app that has been created which incorporates multiple tabs allowing users to explore the dataset, view visualizations and download the report. Findings suggest a rising trend in the number of earthquakes since the 20^{th} century despite the decrease in the mean magnitudes. The geographical distribution of the earthquakes show consistency, primarily occurring along the tectonic plate boundaries but there is no clear correlation between earthquakes and climate changes have been established.

Contribution

All members of the group have contributed an equal amount of hours to this project.

Contents

1. Background and Motivation

The purpose of this project is to examine how earthquakes have evolved over the past hundred years from 1900 to 2013. An earthquake is a shaking of the surface of the Earth and it is caused by movements in the outermost layer of the Earth, when two blocks of the earth slip past one another [NAS]. The blocks are called *tectonic plates*, and the edges of these plates are called *plate boundaries*. The tectonic plates divide the Earth's crust into distinct plates, and they are always slowly moving. Earthquakes are usually concentrated along the plate boundaries, which limits the occurrence of earthquakes to specific parts of the world [USG].

Denmark is placed in the middle of a tectonic plate called *the eurasian plate*, since it covers most of Europe, Russia and parts of Asia [EAR]. This means Denmark is not prone to Earthquakes, but countries like Japan and Indonesia, are prone to Earthquakes since they are placed near the plate boundaries [REL]. However, it is still interesting to examine earthquakes and the evolution of these, since the Earth is affected by around 55 earthquakes a day - that is 20.000 a year [NAJ].

In general, this project will concern how earthquakes have evolved over the past hundred years. This will be done by examining the questions from section 2. Project Objectives, and these questions will be answered with the visualizations in section 4. Visualization and Dashboard.

2. Project Objectives

The project group has elicited the following questions and sub-questions for the dataset:

1. Which variables are used to measure an earthquake?
2. Are there any significant evolution of earthquakes during the years?
    - **2.1.** Have the magnitudes of the earthquakes become higher throughout the years?
3. Which parts of the world have experienced the most earthquakes and has this changed throughout the years?
4. Are there any consistencies between earthquakes or do they all completely differ from each other throughout the years?
5. Does the climate affect the earthquakes?
    - **5.1.** Is there any correlation between the evolution of earthquakes and climate change?

3. Data

The following section contains information about the chosen data set for this project, which also includes descriptions of the relevant variables, number of records, and special features of the data set, including a description of the data processing.

3.1 Source of data set

The data set used for this project is called 'Earthquakes from 1900 - 2013' and it was found on the website Kaggle. The data has been collected by the 'United States Geological Survey' (USGS) and is posted by Varun Sai Kanuri on Kaggle. The data set consists of one CSV file, containing all the records that will be used for the visualizations. The data set has 14 columns and 8313 rows - and the size of the data set was essential when it was selected, since it was important to have enough data to visualize. Furthermore, it is important to clarify that the earthquakes in this data set have a magnitude of 6 or greater, which means smaller earthquakes with lower magnitudes are not a part of this data set. This information is relevant to have in mind, since it might have an impact on the visualizations, but this will be discussed further in section 6. Conclusion and Discussion.

Link to data set: Earthquakes from 1900 - 2013 [VAR]

3.2 Description of data set

The data set contains different variables that will be used to visualize the data. The variables used to answer the project objectives through visualization are listed in **table 1**.

Table 1: *Relevant variables from the data set*

Variable	**Description**
Date	Date when the earthquake occurred
Latitude	Decimal degrees latitude (negative values for southern latitudes)
Longitude	Decimal degrees longitude (negative values for western longitudes)
Depth	Depth of earthquake in kilometers
Mag	Magnitude of the earthquake (Richter scale)

The different variables from **table 1** are used for different visualizations and dashboards, and certain variables are used across different visualizations. The variable 'date' is used to map the amount of earthquakes every year, and also the mean of magnitude of the earthquakes per year. The 'longitude' and' latitude' are used to map the placement of the earthquakes. A variable that was provided which was filtered out was 'place'. The group chose to use the 'longitude' and 'latitude' instead since it was more precise. The depth variable is used to check for differences between the earthquakes throughout the years.

3.3 Data Processing

The data set used for this project was simple to work with, since the data did not need much cleanup before use. As earlier mentioned, the data set consisted of one CSV file that was imported into RStudio and used to produce the different visualizations and dashboards. After the data set was imported, all the rows with earthquakes from the year 2014 were removed from the data table.

Listing 1: *Import of data set and removal of rows*

```
#import data set
path <- dirname(rstudioapi::getSourceEditorContext()$path)
setwd(path)
data <- read.csv("Earthquakes.csv", header=TRUE, sep=",")

#remove all rows where the year is 2014
data <- subset(data, as.numeric(format(as.Date(Date), "%Y")) != 2014)
```

The code in **listing 1** shows how the data set was imported into RStudio, and how the rows with earthquakes from the year 2014 were removed from the data table. The reason for this removal was that the data set contained measured earthquakes with a magnitude of 6 or greater from the years 1900 to 2013. However, the data set also contained earthquakes from 2014, but only for the first three months of the year. Therefore, these rows were removed since it would make no sense to include these earthquakes in the dashboards and visualizations. It would be misleading, since this year would only contain data from the first three months of the year, and the number of earthquakes would not be comparable with the rest of the years from the data set.

Additionally, the magnitudes of earthquakes from the year 1978 are missing. Earthquakes have been registered this year, but the magnitudes are missing. This should of course be taken into account, when the dashboards and visualizations are made, but it is described further in section 5. Story and Results, where the missing magnitudes are relevant.

4. Visualization and Dashboard

The following section contains the general ideas for the visualization design, including descriptions of the designs and design choices. Furthermore, both must-have features and optional features for the visualizations are listed in this section.

4.1 Design

The design section contains information about how the data will be displayed, and this includes general ideas for the visualizations, choice of color palette, used tools and the Shiny app.

4.1.1 General ideas

In order to produce the different dashboards and visualizations, various considerations have been taken into account. This included additional columns that were added to the existing data table, and additional data frames that were produced in order to create the desired visualizations and dashboards. The new data frames were used to create various bar charts, scatter plots, and time series plots. The project group also considered creating a mapview, consisting of a map with earthquakes in the form of small dots, and maybe even an interactive map with a slider to change the year to see earthquakes from the different years. The final implementation is described further in section 5. Story and Results, where final dashboards and visualizations are also described.

4.1.2 Color palette

Since the visualizations are going to answer the questions from section 2. Project Objectives, it is important to use the right color palette to ensure everybody is able to distinguish between the different colors of the dashboards. Different colors are used in the dashboards, since it is a crucial tool to visualize quantitative information [ROB]. When choosing different colors, it is important to take into account that people see colors differently, actually about 1 in 20 people are colorblind in some way. The most common type of colorblindness is red-green color deficiency, where people cannot see the shades of red and green the same way as people with normal color perception do [TOM].

To ensure the visualizations and dashboards are accessible for both colorblind and non-colorblind people, the project group has decided to use the color palette that is illustrated on **figure 1**. This specific color palette has been chosen, because it consists of colors that are accessible for both colorblind and non-colorblind people.

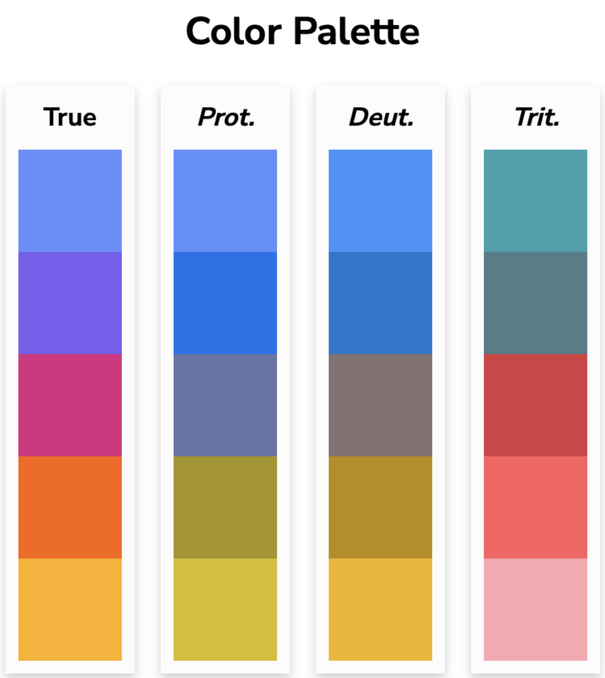

Figure 1: *Color palette (IBM Design Library)*

Figure 1 illustrates four different columns; The left outermost column is called 'True' and it illustrates how people with normal or nearly normal color vision will see the colors. The other three columns are called '*Prot.*', '*Deut.*' and '*Trit.*' which stands for protanopia, deuteranopia and tritanopia respectively [DAV]. These columns illustrate how a person with one of these color deficiencies would see the 'True' colors, respectively. As the figure illustrates, by using the colors in the column 'True' this means persons with any kind of color deficiency would still be able to distinguish between the different colors in the visualization or dashboard.

4.1.3 Tools

Different dashboards and visualizations have been made in order to answer the questions from section 2. Project Objectives, and in this context various *R Libraries* have been used. Libraries such as; shiny, ggplot2, DT, mapview, leaflet and more have been used to create the different visualizations and dashboards. DT has been used to display the data table in the Shiny application, where the library shiny has been used to build the interactive web application. Mapview and leaflet have been used in order to create a map overview with earthquakes, while ggplot2 has been used to generate all other plots.

4.1.4 Shiny App

The Shiny application is implemented in R and deployed to shinyapps.io, where it is accessible. The application consists of different tabs, and it is possible to navigate between these different tabs that contain the data table, different plots, a mapview, and the report. The application has these tabs to form a general view of the different questions from section 2.

Project Objectives, since the tabs are divided into these questions. One of the tabs contains the entire data set used for this project, where it is also possible to search in the data set, and another tab contains the project report that is also downloadable.

4.2 Must-have Features

The must-have features of the visualizations in the Shiny application are listed in **table 2**. These features are considered to be essential for the application, some of them were given in the 'Project Delivery Guidelines' [PAR] and others are elicited by the project group.

Table 2: *Must-have features*

ID	Description
MHF1	The project must include at least three types of graphs
MHF2	The project must include at least one animated graph
MHF3	There must be a total of at least eight graphs
MHF4	The report must include a link to the dashboard with visualizations
MHF5	The dashboard with visualizations must include an option to download the report

4.3 Optional Features

The optional features of the Shiny application are listed in **table 3**, and these features are considered to be optional since they are nice to have, but it is not critical if they are omitted.

Table 3: *Optional features*

ID	Description
OF1	The dashboard with visualizations could include an option for reading the report
OF2	The Shiny application could include an option to see the entire data set
OF3	The Shiny application could include an option to download the entire data set

5. Story and Results

The following section contains descriptions of the dashboards and visualizations, produced to answer the questions from section 2. Project Objectives, including answers to these questions.

5.1 Variables used to measure an earthquake

To answer the question about which variables that are used to measure an earthquake, different papers and websites have been studied.

5.1.1 Which variables are used to measure an earthquake?

Different variables are used to measure an earthquake, but the most significant variable is the magnitude. The magnitude of an earthquake is determined by using the logarithm of the height of the largest seismic wave that is calibrated to a scale by a seismograph [JOH]. The height of an earthquake is also called the *depth* of an earthquake, and the depth can range from 0 to 700 kilometers, since this is the depth of the three outermost layers of the Earth. The depth of an earthquake can be divided into three zones: Shallow earthquakes, 0 to 70 kilometers deep; intermediate earthquakes, 70 to 300 kilometers deep; and deep earthquakes, 300 to 700 kilometers deep [EAH].

5.2 Evolution of earthquakes

In order to answer the question about the evolution of earthquakes, including the sub-question about how the magnitudes have evolved, different visualizations have been made. The evolution factors that are valuable to visualize are about the amount, the magnitude, and the placement of earthquakes.

5.2.1 Are there any significant evolution of earthquakes during the years?

One of the questions about the data set was to find out if there was a significant evolution of earthquakes during the years. To answer this question, the group decided to create a bar chart showing the number of earthquakes as a function of time. The visualization shows the number of earthquakes each year from 1900 to 2013, and it is clear to see that certain years have more earthquakes than others. However, it is possible to see a trend where the number of earthquakes have increased in the last 100 years. The numbers were relatively similar from 1900 to 1963, but after 1963 the number of earthquakes increased significantly. This means that there has been a rising trend in the number of earthquakes in the last 50 years or so, and it is only possible to guess why the number has risen that much.

Maybe, the number of earthquakes have actually increased because of human activities, such as mining and oil extraction, or natural variability, such as geological process and seismic activity. The rising number could also be caused by improved monitoring methods and better technology, which means both detecting and recording earthquakes have become better over time. What speaks against this is that all the earthquakes in the data set have a magnitude of 6 or greater, and this means they have been in the categories ‘strong’, ‘major’, or ‘great’ [JOH]. Earthquakes in these categories would cause at least moderate damage in populated areas, which means it would have been difficult to not notice these earthquakes, unless they appeared in the oceans [JOH].

5.2.2 Have the magnitudes of the earthquakes become higher throughout the years?

A way to measure an earthquake is by magnitude, where the Richter scale is used. This scale is a quantitative measure of the magnitude of an earthquake [JOH]. To answer the question, different visualizations were made to show the mean of magnitudes of each year, from 1900 to 2013. The scatter plot, time series plot, and bar chart show clearly that the mean of magnitudes have been just under 7 from 1900 to 1963, but after 1963 the mean drops to around 6.5 and this is illustrated with the two different colors in the bar chart. Both the time series plot and the bar chart show clearly that the bar and line are empty at the year 1978, caused by missing data from that specific year. Earthquakes have been registered, but the mean of the magnitudes have not been registered, and therefore the bar and line are empty. However, the empty bar and line have not been omitted since the visual representation of the different years was essential, and it was preferred to show that there was no data for the year 1978.

These visualizations are interesting, particularly if the previous visualization with the number of earthquakes is taken into account. The previous visualization illustrated that the number of earthquakes was increasing, but these visualizations with the mean of magnitudes indicate that the mean is decreasing. Overall, the data set indicates that the world is experiencing more earthquakes, but the magnitude of the earthquakes are lower.

5.3 Geographical placing of earthquakes

To answer the question about the geographical placing of earthquakes, a dashboard has been made. The dashboard and question are both described further in this section.

5.3.1 Which parts of the world have experienced the most earthquakes and has this changed throughout the years?

By creating a world map and plotting all earthquakes with the correct longitude and latitude and filtering this by its year we get an overview of their placements which can be seen in the application tab “World Map”.

By analyzing the world map for each year a clear pattern can be found. Despite the number of earthquakes changing from year to year, their placements seem to be consistent.
Most of them are placed by the western border of North America and South America, the eastern and southern border of Asia, the Northern border of Australia and north of New Zealand. As 80% of all earthquakes occur when tectonic plates are pushed together [CAS] it will seem as a predictable pattern as each of these occurrences are placed by the border between two tectonic plates.

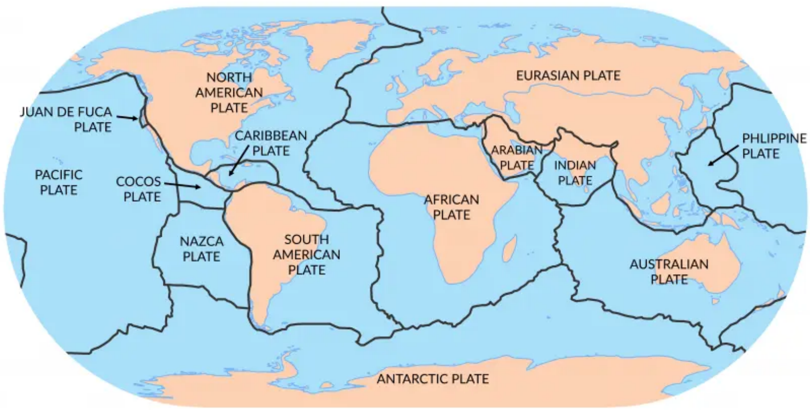

Figure 2: *Illustration of 7 major tectonic plates [EAE]*

As tectonic plates only move a few cm each year [EAE] it will be many years before the placement of earthquakes will have a significant change.

5.4 Consistency between earthquakes

To answer the question about consistency between earthquakes, different visualizations have been made.

5.4.1 Are there any consistencies between earthquakes or do they all completely differ from each other throughout the years?

To gain insight into the consistency of the earthquakes, two bar charts were created and a scatterplot. One of the bar charts is shown in chapter 5.2.2. It demonstrates the mean magnitude of the earthquakes measured from 1900 to 2013. The other bar chart shows the mean depth of the earthquakes from 1900 to 2013.

As mentioned previously in chapter 5.2.2, the bar chart of the mean magnitude has not fluctuated much throughout the years, except in 1963 when it dropped below 6.5. This shows that there is a uniform pattern in the mean magnitude compared to the mean depth. The other bar chart does not show a uniform distribution, nor does it show a clear linear correlation. It fluctuates a lot more than the bar chart showing the mean magnitude throughout the years.
Creating these two bar charts is not enough to comprehend the relationship between these two variables. To provide valuable understanding about the relationship between these two variables a scatterplot was created. The x-axis shows the mean magnitude whereas the y-axis displays the mean depth. The scatter plot illustrates clusters but these two clusters do not show any pattern.

5.5 Climate changes and earthquakes

In order to answer the questions about earthquakes and the climate, different papers and websites have been studied, including a plot of earthquakes each season.

5.5.1 Does the climate affect earthquakes?

No concrete evidence has been found confirming a correlation between earthquakes and climate change. However, evidence supporting the idea that *the weather* and *global warming* can affect seismic activity has been found, but experts cannot confirm that the climate is a factor for earthquakes. According to NASA, many earthquakes are caused by large downpours, such as heavy rain or snow, but they are deemed insignificant to the point that the scale is so low humans do not feel it [KOR]. According to the U.S. Geological Survey, the only correlation between earthquakes and the weather is large changes in atmospheric pressure, usually caused by large storms like hurricanes, which can cause a *slow earthquake*. They also say that potentially a large low-pressure change could trigger a damaging earthquake, but "the numbers are small and are not statistically significant" [ALA].

5.5.2 Is there any correlation between the evolution of earthquakes and climate changes?

Potentially, a volcanic eruption could happen if an earthquake appears nearby. This usually happens when a *tectonic plate* shifts near a volcano, which results in an earthquake and a potential volcanic eruption. The release of toxic gasses from the volcano, such as sulfur dioxide and carbon dioxide will contribute to the greenhouse effect, resulting in warming the Earth's atmosphere [BIK]. The movement of *tectonic plates* causing the earthquake can potentially form mountains or cause them to grow. Possibly, this will result in a change of circulation of air around the planet. Large mountain chains influence the pathways of the air, and changes in the mountains can redirect the airways [BGS].

To visualize if there is a correlation between the evolution of earthquakes and climate changes, a pie chart has been made. The visualization shows the number of earthquakes each season since the seasons are impacted by the climate [USD]. The number of earthquakes each season is very similar; winter has the fewest number of earthquakes with 24 % and spring has the most with 26,3 %. However, it is very important to notice that the seasons vary around the world - when it is winter in Denmark, it is summer in Australia. The pie chart for this assignment has been made, based on the seasons seen from the perspective of Denmark, but this also means it can be difficult to determine which season has the most earthquakes. All things considered, it is difficult to say there is a correlation between the evolution of earthquakes and climate change.

6. Conclusion and Discussion

In summary, this report has provided a comprehensive analysis of earthquakes from 1900 to 2013, with a focus on exploring their evolution. The study successfully addresses the project objectives that were created and described in section 2. Project Objectives. These objectives have been created either through research or through the visualizations in the RShiny application, built on feature requirements created by the group. The most significant factors when speaking about earthquakes are magnitude and number of earthquakes, and even with a rising number of earthquakes the mean of the magnitude were decreasing.

The geographical patterns of earthquakes were visualized by creating a world map in the RShiny application, with the earthquakes plotted on their position using longitude and latitude. This creates a visual representation where it is clear that the earthquakes have a pattern related to the earth's tectonic plates. To get a clearer representation a boxplot showing the number of earthquakes in every continent, were under development. To complete this task every longitude and latitude needed to be correlated to a continent. This was proven difficult when earthquakes often occur in the water between continents. The topic of climate change being affected or affecting earthquakes is a non-conclusive topic. The research showed that there was, no significant impact on the climate and the same for the climate affecting earthquakes.

The RShiny application fulfills most of the requirements created. The application has eight diagrams with three different types, where one is animated. The option to view the report and download it through the application is also possible. The OF3 requirement, about being able to download the data set is not available. It is important to note, that the requirement was a "could", and does not affect the overall performance of the application.

References

[ALA] *Alan Buis (NASA's Jet Propulsion Laboratory)*
Can Climate Affect Earthquakes, Or Are the Connections Shaky?
URL: https://climate.nasa.gov/news/2926/can-climate-affect-earthquakes-or-are-the-connections-shaky/
(accessed: 15. November 2023)

[BGS] *Discovering Geology - Climate change (British Geological survey)*
What causes the Earth's climate to Change?
URL: https://www.bgs.ac.uk/discovering-geology/climate-change/what-causes-the-earths-climate-to-change/#:~:text=The%20movement%20of%20the%20plates,to%20cooler%20regions%20by%20mountains.
(accessed: 15. November 2023)

[BIK] *Bikash Sadukhan et. al*
Investigating the relationship between earthquake occurrences and climate change using RNN-based deep learning approach
URL: https://link.springer.com/article/10.1007/s12517-021-09229-y
(accessed: 15. November 2023)

[DAV] *David Nichols*
We see colors differently
URL: https://davidmathlogic.com/colorblind/#%23648FFF-%23785EF0-%23DC267F-%23FE6100-%23FFB000
(accessed: 2. November 2023)

[CAS] *California Academy of Sciences*
Plate Boundaries: Divergent, Convergent, and Transform
URL:
https://www.calacademy.org/explore-science/plate-boundaries-divergent-convergent-and-transform#:~:text=About%2080%25%20of%20earthquakes%20occur,crumple%20and%20are%20pushed%20up.
(accessed: 22. November 2023)

[EAE] *Earth How*
7 Major Tectonic Plates: The World's Largest Plate Tectonics
URL: https://earthhow.com/7-major-tectonic-plates/
(accessed: 22. November 2023)

[EAH] *Earthquake Hazards Program*
Determining the Depth of an Earthquake
URL: https://www.usgs.gov/programs/earthquake-hazards/determining-depth-earthquake
(accessed: 14. November 2023)

[EAR] *Earth How*
7 Major Tectonic Plates: The World's Largest Plate Tectonics
URL: https://earthhow.com/7-major-tectonic-plates/
(accessed: 26. October 2023)

[JOH] *John P. Rafferty*
Richter scale
URL: https://www.britannica.com/science/Richter-scale
(accessed: 13. November 2023)

[KOR] *Kori Williams (Green Matter)*
Climate Change and Earthquakes Are Connected – Here's How
URL:https://www.greenmatters.com/weather-and-global-warming/does-global-warming-cause-earthquakes
(accessed: 15. November 2023)

[NAJ] *NASA Jet Propulsion Laboratory*
Understanding Earthquakes
URL: https://www.jpl.nasa.gov/topics/-earthquakes
(accessed: 26. October 2023)

[NAS] *NASA Science*
What is an earthquake?
URL: https://spaceplace.nasa.gov/earthquakes/en/
(accessed: 26. October 2023)

[PAR] *Parisa Niloofar*
Project Delivery Guidelines
URL: ProjectDeliveryInstructions.pdf
(accessed: 29. October 2023)

[REL] *Relating*
Top countries where earthquakes most often occur
URL: https://realting.com/news/which-countries-are-most-prone-to-earthquakes
(accessed: 26. October 2023)

[ROB] *Robert Simmon*
Use of Color in Data Visualization
URL:
https://earthobservatory.nasa.gov/resources/blogs/intro_to_color_for_visualization.pdf
(accessed: 2. November 2023)

[TOM] *Tom Mangan*
Red-Green Color Blindness
URL: https://www.allaboutvision.com/conditions/color-blindness/red-green-color-blindness/
(accessed: 2. November 2023)

[USD] USDA Climate Hubs
Growing Season in a changing Climate
URL: https://www.climatehubs.usda.gov/growing-seasons-changing-climate
(accessed: 17. December 2023)

[USG] *USGS*
The science of Earthquakes
URL: https://www.usgs.gov/programs/earthquake-hazards/science-earthquakes
(accessed: 26. October 2023)

[VAR] *Varun Sai Kanuri (USGS)*
Earthquakes from 1900 - 2013
URL: https://www.kaggle.com/datasets/varunsaikanuri/earthquakes-from-1900-2013/data
(accessed: 29. October 2023)