Worldwide Volcanoes Analysis and Impact

— on economic development and natural disasters

Overview

Natural disasters, such as volcano eruptions, are catastrophic and extreme weather events. Although it can occur anytime around the world, some regions are more vulnerable and sensitive to specific kinds of disasters than others. Our group would like to use these datasets to tell a data science story about the eruption of volcanoes and the impact of volcanic activities on economics and other natural disasters. It has been known that natural disasters can heavily destroy people's lives and livelihoods for example the recent news of volcanic eruption happened in Hawaii; the occured losses might even be the major obstacles to human's sustainable development. Therefore, we are going to analyze the activities of volcanoes along with a given timeline, visualizing the influence of different geographical locations on volcano movements, the relationships between volcano types and various geo features, and its impact on economics.

Detailed description of the data and data pre-processing (click 'Dataset #' for URL)

Dataset 1: This dataset is about worldwide volcanic eruptions activities. It contains 1343 volcanoes ranging by alphabetical order. Since this dataset contains geo data about the 'region', 'subregion', 'longitude', 'latitude', 'elevation' of the volcano, the geomap visualization can be generated to display the area with frequent volcano activities. Additionally, text data about 'Primary Volcano Types' and 'Volcano Name' here would be used to further discover the type of volcano which has the most frequent eruptions, and any specific relationship between volcanoes and subregion's location.

Dataset 2: The dataset shows the information about the number of volcanoes found in 77 different counties, which provides geospatial information and can help us have a worldview of the volcanoes activities. We will be visualizing the relationship between the frequency of historical active records and current eruptions for Holocene Volcanoes based on 5 variables that describe countries, number of volcanic activities since Holocene/ 1800CE/ 1950CE, and current eruptions. From these variables, we can discover the activity trend of the top 5 volcanoes, then get connections with other datasets.

Dataset 3: This dataset contains 48 volcanoes which were in continuing eruption status as of 27 January 2022. It includes two columns called 'Eruption Start Date' and 'Eruption Stop Date', thus we can do some time series analysis. It also includes the country names, which provide some geospatial information. And for the column 'WVAR (rollover for report)' , it can act as a label. And we can do supervised analysis by using the country/volcano names in order to predict which volcano activities are more likely to be reported.

Dataset 4: Detailed Description: This dataset contains 36 columns which describe the properties of the volcano, as well as the relevant economic and human impacts from the volcanic eruptions, including number of death, number of missing, number of injuries, damage in million dollars, number of houses destroyed, etc. This project aims to separate the volcanoes by regions, by plotting the boxplot, comparing the number of deaths, injuries and damages in million dollars.

Pre-processing: The original data here should be preprocessed and cleaned up before the further analysis on both quantitative data and qualitative data. Besides, there are some overlap columns in some datasets, such as the 'Country' column in both 'Countries with Holocene Volcanoes' and 'Holocene Volcano List', the 'Currently Erupting' column in 'Current Eruptions' dataset and the 'Volcano Name' column in 'Holocene Volcano List'. So we need to join multiple datasets to make the datasets much clearer for the following analysis. For the data pre-processing and cleaning, we would use R scripts, and for data

visualization, we would utilize both R and Python to plot basic statistical graphs and use Tableau to draw geomap figures.

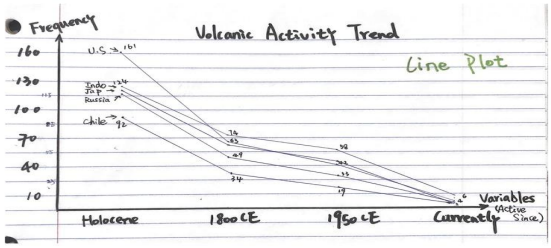
Usage scenario & tasks

15th Jan, 2022, the Georgetown University student Lisa Chen was shocked by the news of a volcanic eruption that happened in Tonga. She was very worried about the situation there, since one of her best friends Fangu Fukofuka is from Tonga. Lisa texted to Fangu right after she saw the news in the morning. 6 hours passed, Fangu still did not text back. And the whole world had trouble connecting with Tonga as well. Lisa desperately wanted to know how devastating the effect of the volcanic eruption could be, and how possibly it can be related to other natural disasters, including earthquakes, tsunami, etc. She opened the website "VolcaKnows", she will see dataset of the worldwide distribution of volcanoes, which have attributes of city, country, volcano type and time of eruption, along with a map plot of the dataset. The second part of the website shows the datasets with various properties and economic and human impacts of volcanoes. Lisa could filter the dataset by location, and see how the historical impacts of volcanic eruptions like in the area near Tonga. Besides that, there is another dataset containing information about earthquakes that happened worldwide from the year 1965 to 2016, which can be plotted on the same map with the volcanic eruption to see if they are related to each other. As well as how frequently earthquakes happen in the close area of volcanic eruption in a given time window.

Description of visualization & initial sketch

Visualization for dataset 1: This is a map plot showing the top 15 subregions that had the most frequent volcano eruptions during the Holocene period. Each of the sites marked by volcano labels is an area of high volcano activities.

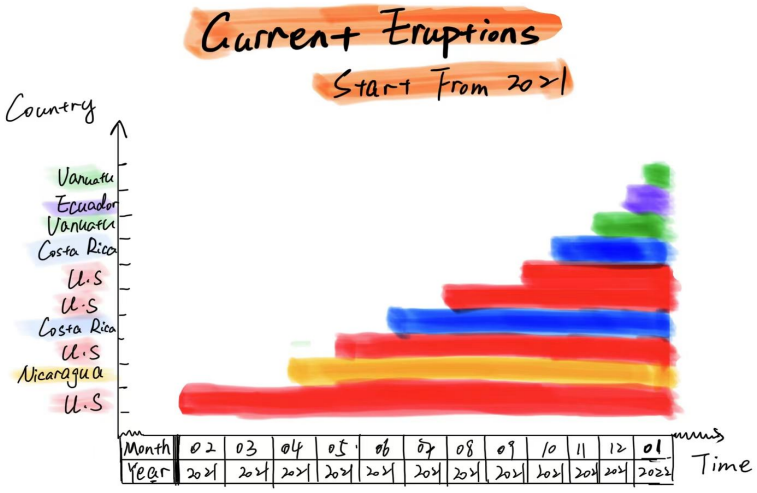
Viz #1: Geomap



Viz #2: Line Plot

Visualization for dataset 2: This is a traditional line plot setting 'Period Variables' as x axis and 'Frequency' as y axis. This line plot shows the trend of top 5 frequent volcanic activities, and can provide insight for future analysis on the relationship between the frequency of current eruptions and historical time record for different types of volcanoes. The clear view of this plot makes it special, and it can provide sufficient geospatial information as well as time data altogether.

Visualization for dataset 3: This is a barplot made from dataset 4. The x-axis is Time and divided by month; The y-axis is Country and split by color (red for U.S, yellow for Nicaragua, blue for Costa Rica, green for Vanuatu and purple for Ecuador). Considering the limited space of our sketch, we only choose the current eruptions starting from 2021. We can conclude from this sketch that, U.S had the most active volcanoes in 2021, and then Costa Rica and Nicaragua.



Viz #3: Bar Plot

Work breakdown and schedule

Milestones	Hours will take	Completion Date
Data gathering and project introduction	7	02/20/2022
Create problems we want to solve	2	02/27/2022
Data cleaning for dataset 1/2/3	2	03/06/2022
Data cleaning for dataset 4 (we might need to gather more data here) and visualization	4	03/13/2022
Merge dataset 1 and dataset 2 to df1 then do basic visualization	6	03/20/2022
Merge dataset 2 and dataset 3 to df2, then do basic visualization	5	03/27/2022
Create map plots and line plots based on df1	6	04/03/2022
Create bar plots and box plots based on dataset4	6	04/10/2022
Create a novel visualization type based on df2	2	04/17/2022
Use html+CSS to build the webpages	12	04/24/2022
Conclusions and Infographic	4	4/28/2022