COMP7507 Data visualisation and visual analysis

Group K: Fei Liu (3035460493), Xinwei Li (3035454949), Qingqing Yao (3035418622)

1. Data Preprocessing

1.1 Raw data and attributes selection

We used three types of data: Volcanoes information data from Kaggle [1], Volcanoes eruption data from Global Volcanism Program [2] and earthquake data signif.csv [3].

Volcanoes information data, which includes name, location and types of volcanoes active in the past 10,000 years,. Some attributes are irrelevant to us, hence, the attributes we used are as follows: volcano name, country, region, type, active evidence, latitude, longitude, elevation and tectonic setting. Tectonic is the process that controls the structure and properties of the Earth's crust and its evolution through time, and the attribute tectonic setting in our dataset indicates the tectonic zones and the crush depth.

"Volcanoes of the World" is a database describing the physical characteristics of volcanoes and their eruptions. For volcanoes eruption data, we extracted and analysed the volcanoes eruptions data from database. The features we mainly used are volcanoes names, latitude and longitude, eruption started year and VEI which indicates the level of the volcanoes eruption.

For the signif.csv file, which is loading the earthquake information, we need to eliminate all items that contain empty attribute value. Otherwise it will rise error during coding operation.

1.2 Modified Type Values

A number of type values has postfix strings, like '(s)', '(es)','(?)', which should be removed before using for analysis. Since, names with and without these postfixes, they indicate exactly the same type, but they will be divided into different types while shorting if not be modified.

B	C	D	E
Name	Country	Region	Type
West Eifel V	Germany	Mediterrane	Maar(s)
Chaine des I	France	Mediterrane	Lava dome(s)
Olot Volcani	Spain	Mediterrane	Pyroclastic cone(s)
Calatrava Vc	Spain	Mediterrane	Pyroclastic cone(s)
Larderello	Italy	Mediterrane	Explosion crater(s)
Vulsini	Italy	Mediterrane	Caldera
Colli Alban	Italy	Mediterrane	Caldera
Campi Flegr	Italy	Mediterrane	Caldera
Vesuvius	Italy	Mediterrane	Stratovolcano
Ischia	Italy	Mediterrane	Complex
Palinuro	Italy	Mediterrane	Submarine
Stromboli	Italy	Mediterrane	Stratovolcano
Panarea	Italy	Mediterrane	Stratovolcano
Lipari	Italy	Mediterrane	Stratovolcano(es)
Vulcano	Italy	Mediterrane	Stratovolcano(es)
Etna	Italy	Mediterrane	Stratovolcano(es)

from：vocanical_database.csv

2. Visualization

2.1 Globe visualization using Cesium

2.1.1 Introduction

CesiumJS is an opensource geospatial 3D mapping platform for creating visual globes, which could be efficiently used on the visualisation of the location of each volcanoes. This library provides varies features of building customized 3D global view including imagery and terrains.

2.1.2 Implementation

After download the latest Cesium release and tested that the browser was available for the Cesium, we setted up a web server for it. The web server could be executed after installed the Node.js and npm through the Terminal.

```
SylviadeMacBook-Pro:Cesium-1.44 sylviameow$ npm start

> cesium@1.44.0 start /Users/sylviameow/Documents/Cesium-1.44
> node server.js

Cesium development server running locally.  Connect to http://localhost:8080/
```

The usage of Cesium.js is similar to D3.js, which can be include directly in script. In order to use Cesium Viewer widget, which is base widget for building applications. It composites all of the standard Cesium widgets into one reusable package. The widget can always be extended by using mixins, which add functionality useful for a variety of applications. Its CSS was needed to be included. Then, created a viewer to start the globe visualisation.

2.1.3 Volcanoes data visualisation

Since there is a built method in the library which is a data source to process the GeoJson file. The original CSV data file was converted to GeoJson format and be loaded. According to the latitude and longitude of each piece of data, the location of the volcano can be represented as a pin.

At the very beginning, all the setting of the visual were the default, which is not intuitive.

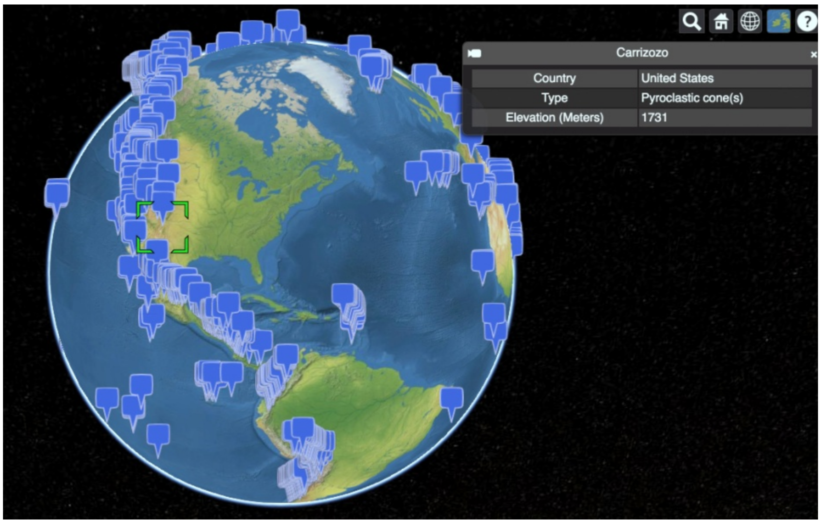


Figure 1.

Firstly, the style of the pin can be changed. Using Maki icon (which was downloaded from the internet), to add a volcano symbol on each pin. There is also a built method: markerSymbol in the load GeoJSONDataSource to modify the default symbol.

Secondly, the types of the volcano can be used to sort the colour of the volcano label. Since there is no built-in function to do that, and it is hard to realize through scripting, the original file were separated into single files sorted by volcano types. Then, it is easy to load different data source and modify the marker setting of each.

At last, the imagery of the globe is not informative because it is not clear to show the countries or regions around the world. After comparison on all the types of imagery that the website provided, a national geographical world map was used. For each pin of the volcano, it is simple to click on it and check the detailed information of the volcanoes. The final version of this globe visualisation is as Figure1.1.

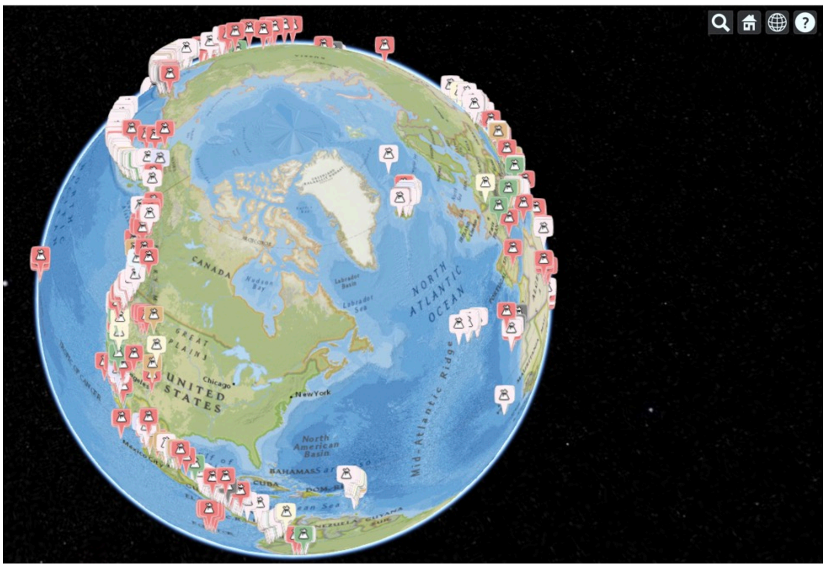


Figure 1.1

2.1.4 Volcanoes data visualisation functions

- Camera movement

Using mouse to drag the globe to rotate to specific positions. Using wheel to zoom in/out to a place.

- Place search

Entering a name to find a place.

- Home view

Click the button to back to default point of view.

- View mode to 2D and 2.5D (Hausdorff dimension)

View as 2D

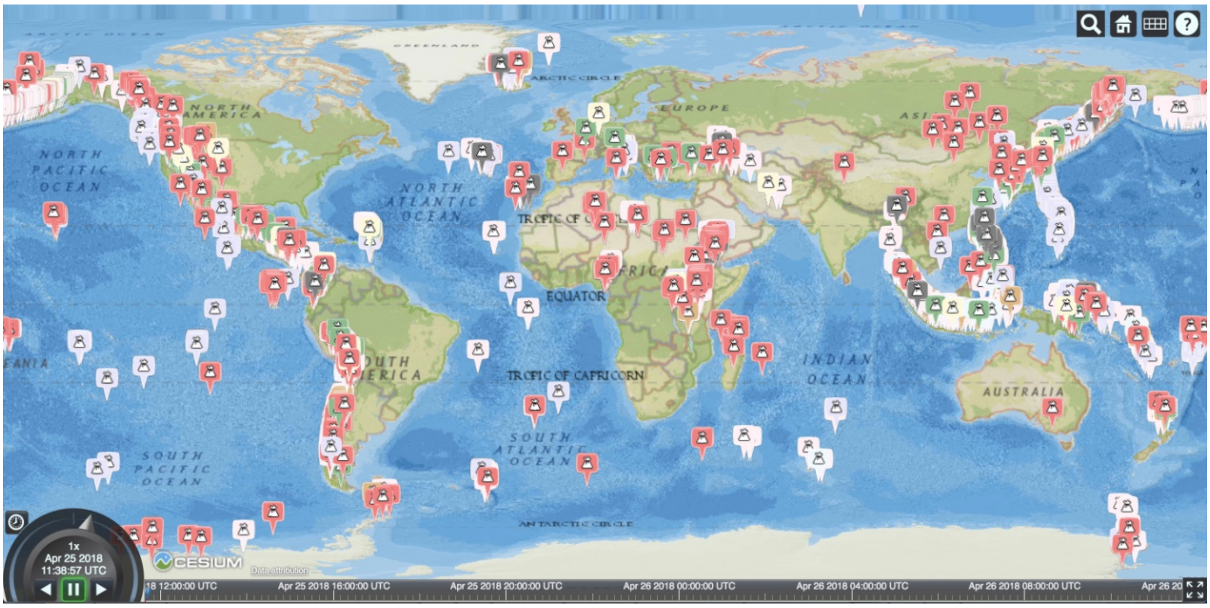


Figure 1.2

View as 2.5D

Figure 1.3

- Help button

Clicking to show controlling help.

2.2 Visualization using Tableau

2.2.1 Visualization about top 30 unrest volcanoes.

Since there the amount of the volcanoes in the dataset is over 1,500, it is necessary to distinguish the active volcanoes among all. Different volcanoes are represented by different colours. It is obvious that count of the eruption time of the most active volcanoes is about three times than the 30th. In conlcution, most of the volcanoes around the world were not that active.

The x-axis represents the number of eruptions, and y-axis represents corresponding volcanoes' name. The visulization is presented the top 30 volcanoes with the most number of eruptions from BC to 2018. For example, the top one is Etna volcano, and it has 241 times of eruption.

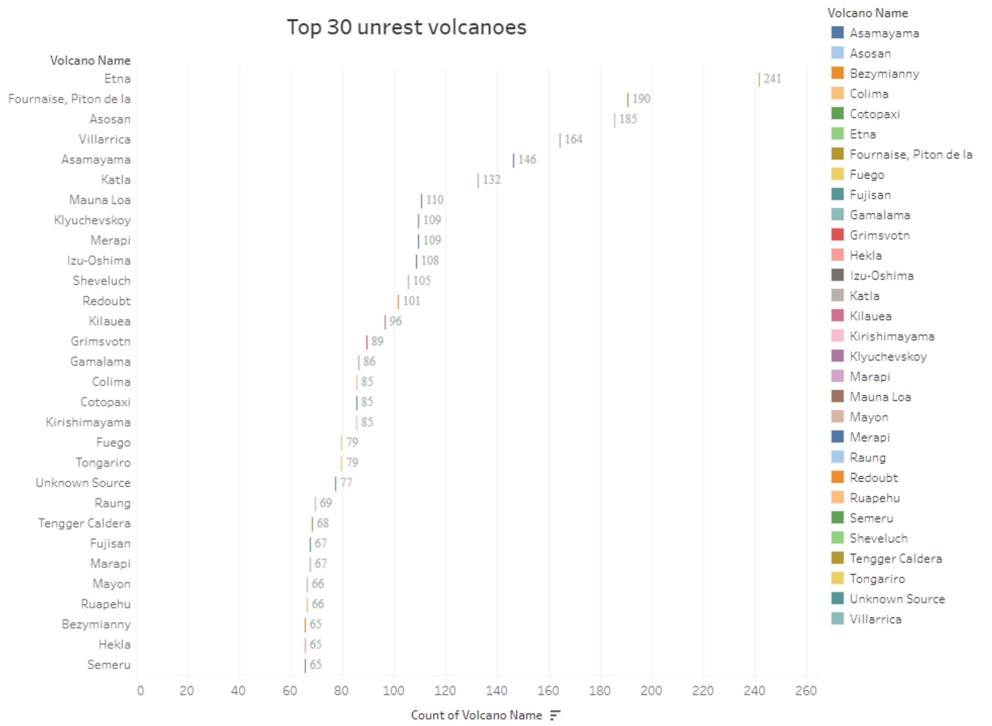


Figure 2.

2.2.2 Top 5 unrest volcanoes with VEI.

From the above visualization, we can get the top 5 unrest volcanoes which are Etna, Fournaise/Piton de la, Asosanm and Villarrica. We know about these 5 volcanoes erupt very frequently, and then we also want to know the detialed intensity of their eruptions. This visualization shows the intensity of every eruptions for top 5 unrest volcanoes..

The x-axis represents the start year with duration of 500 years, and y-axis represents top 5 name of unrest vocalnoes with VEI intensity levels. One vertical line represent one time eruption with x year and y VEI.

The depth of color represents the intensity of the volcanic eruption, is that, the darker the color, the larger the VEI's value, and the eruption is more intensity. We can see that from light to dark representing VEI from 0 to 5. The color of 0 and unknown VEI have the same level of color. Moreover, 0 has lightest color, and 5 has deepest color.

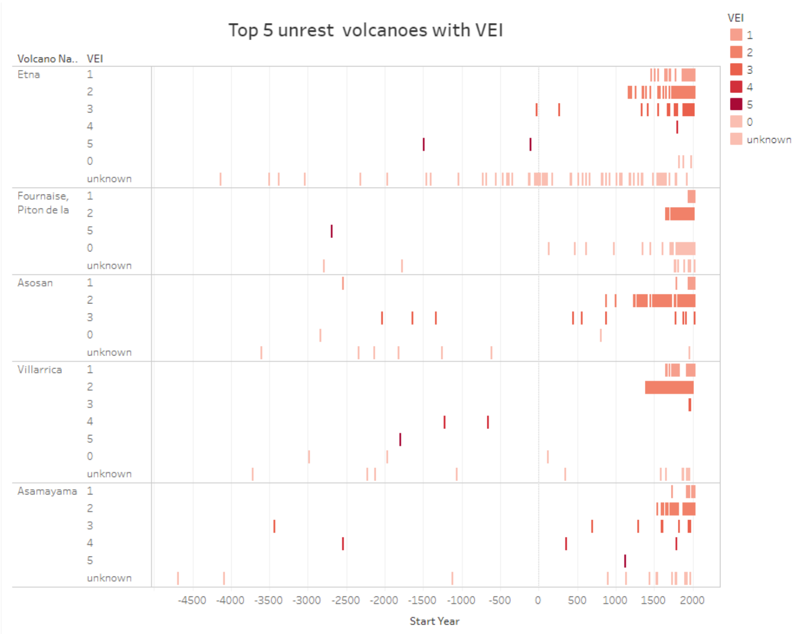


Figure 3.

The most eruptions' in intensity VEI is 2 or 3, and for Fournaise/Piton de la vocalno, even though it has many times for eruption, the intensity is not very strong, and has many eruptions with 0 VEI intensity.

2.2.3 Relationship between tectonic setting and geographical area.

We use the tableau 'symbol maps' to point the volcanoes. We firslty prepocess the original vocanical data's attribute 'Tectonic Setting', oringally, it has many types of tectonic setting: Intraplate / Continental Crust (>25 km), Intraplate / Intermediate Crust (15-25 km), Intraplate / Oceanic Crust (< 15 km), Rift Zone / Continental Crust (>25 km), Rift Zone / Intermediate Crust (15-25 km), Rift Zone / Oceanic Crust (< 15 km), Subduction Zone / Continental Crust (>25 km), Subduction Zone / Crust Thickness Unknown, Subduction Zone / Intermediate Crust (15-25 km), Subduction Zone / Oceanic Crust (< 15 km), Unknown, Blank, and then we split it to get four types: Intraplate, Rift Zone, Subduction Zone, unknown, and only show three types and one full visual of these three types.

We assign volcanoes as points on the map, and want to find whether the location of volcanoes has corrrealtion with the geographical area. If put the mouse on the point, the tectonic setting type name and the volcano's name will be presented as tooltip.

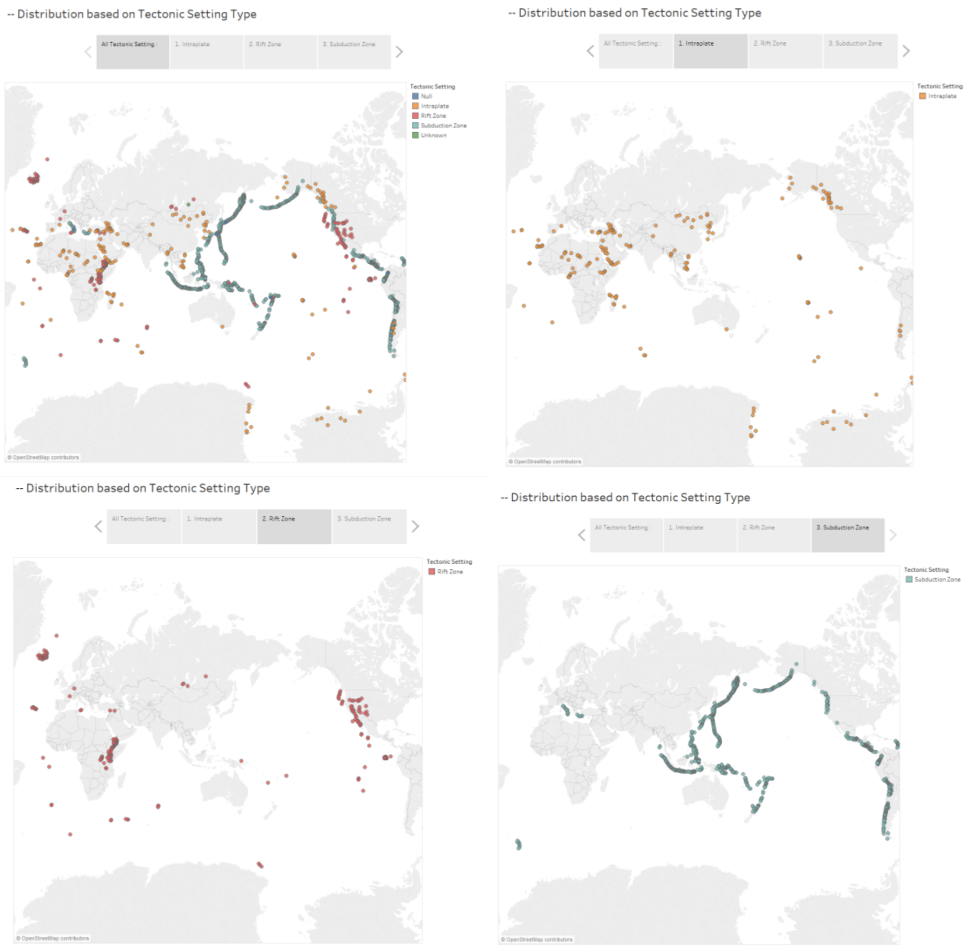


Figure 4.

In these visualizations, we find that for the intraplate tectonic setting type the volcano is mostly located inside the plate. And for the rift zone, there are many volcanoes loicated in Afiaca Great Rift Valley.

In addition, for the Subduction Zone type of tectonic setting, the volcano is located in the coastal zone, where is the junction of the ocean plate and the continental plate. Especially, they are moslty located in the Pacific Ocean, where the junction between the Pacific plate and continental plates.

The new inside from this visualization is that the formation of volcanoes may be related to the activities and movements among plates, such as plate impacting and extrusion. These activities conclude ocean plate and continental plates, continental and continental plates.

2.2.4 Visualization for Volcanoes Type

The dataset we used in this task is the same as 2.2.5, which is ‘vocanical_database(inUsed).csv’. We want to get most common types of volcanoes. Firstly,we sorted the volcanoe type according to each type sum. Precentage for each type is shown. And then we used the tableau ‘horizontal bars’ to present the top 20 most common volcanic types. Different colors indicats different regions.

We can get the information from this visualization that Stratovolcano is the most common type, Shield is the second and the next is Submarine.

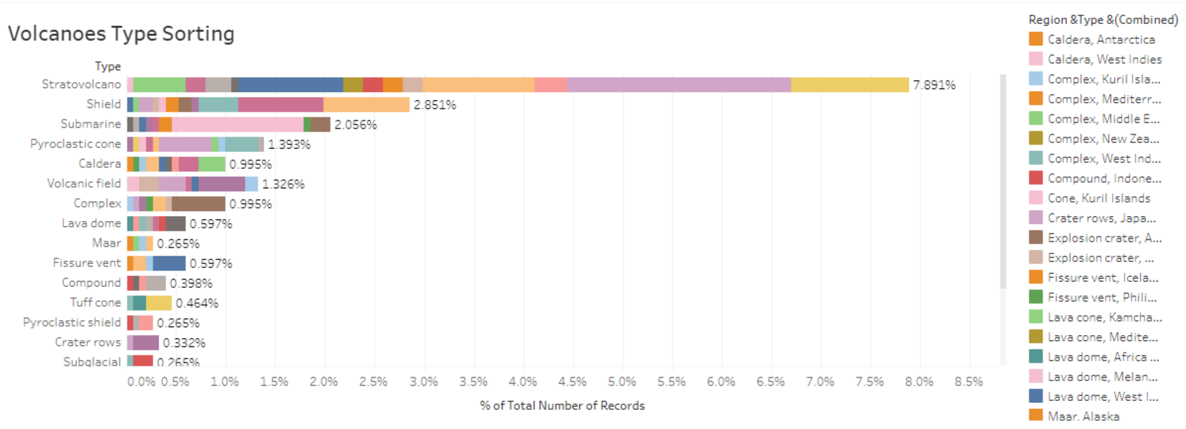
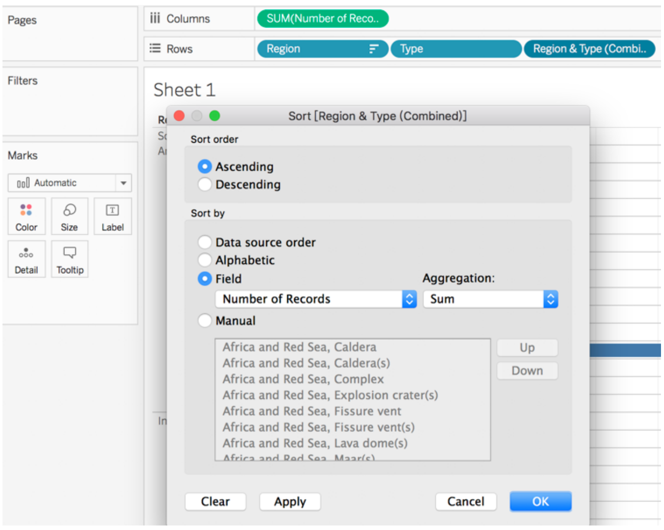


Figure 5.

2.2.5 Detail information between volcanoes type and regions.

For getting the sorting result, we made some settings shown below. We created a combine field containing fields Region and Type, so that we could short the region corresponding to each volcano type. We shorted this field according to the sum number of records, by ascending.



The above Figure 5 is about the most common type of volcanoe, and the next we want to get the detail about which region contributes the most to the proportion in top 20 common volcanic types. For every type of vocalno, sort the region with numbers of vocalnoes from most to least. In addtion, from the visualization:

i. for Stratovolcano: South America contains most of this kind of type, and the next is Indonesia, Alaska, and Japan,Taiwan,Marianas.

ii. for the second most common type Shield: top three regions are Kamchatka and Mainland, Africa and Red Sea, and South America.

iii.for Submarine, Japan,Taiwan,Marianas region contains the most.

iv. The fourth type Pyroclastic cone: Africa and Red Sea, and South America.

v. Caldera: Indonesia, Japan,Taiwan,Marianas.

Inside this visualization, we can get the information that these regions: South America, Africa and Red Sea, Japan,Taiwan,Marianas contains many volcanoes and also the types of volcanoes are complex and diverse.

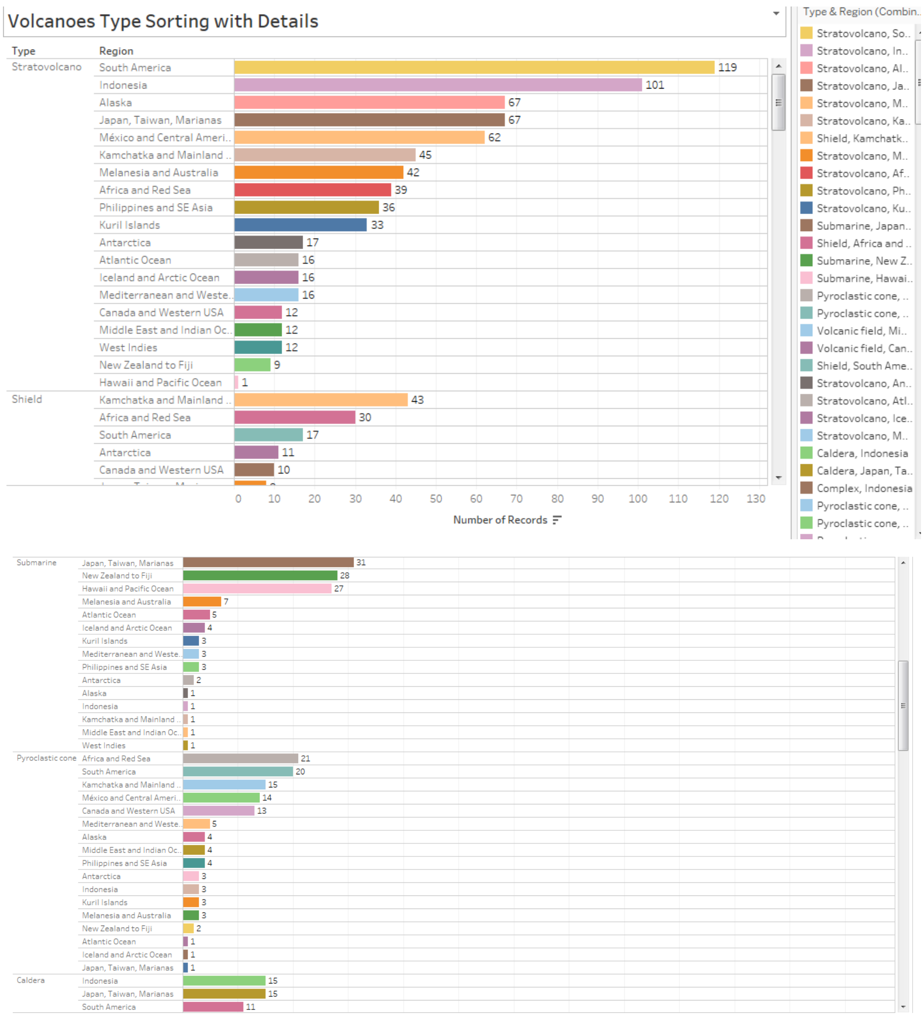


Figure 6.

2.2.6 Statistics of volcanoes in each region.

After discovering the relationship between the volcanoes locations and the tectonic setting, the relationship between the region and the amount of volcanoes should also be explored.

Different colors indicate different regions, and the size of circles represent the number of volcanoes in this region. If there are more volcanoes, then the size of circle is larger. These circles are labeled with related regions' names, and put the mouse on the circle, the number of vocalno and it's region's name are shown.

Hence, the visualization shows that the South America contains the most number of volcanoes with largest size of circle. Then, according to the above image: Figure 4. Relationship between tectonic setting and geographical area, we could get the insight that the reason might be that most of this area located at both the ring of fire and Mid-ocean ridge, these two rings are area with high density distribution of volcanoes.

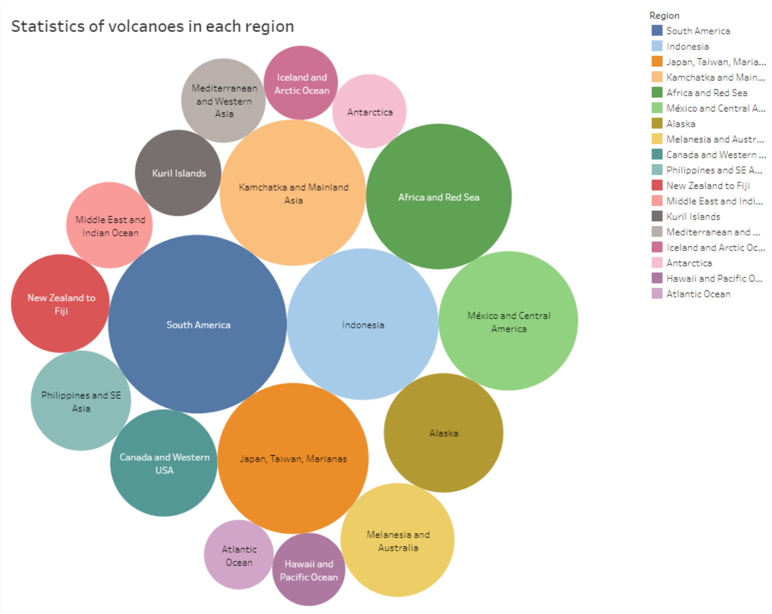
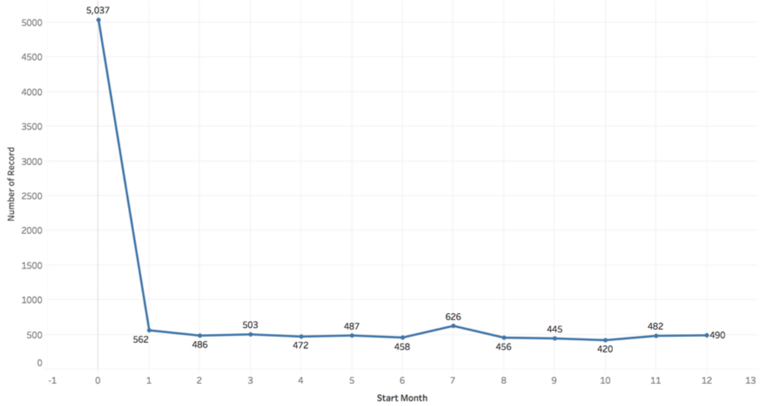


Figure 7.

2.2.7 The number of volcanoes eruptions distribution along with years.

We use tableau "lines (continuous)" to visualize it. Firstly, we want to know whether there are some relations bewtween the vocalno eruption and the month (season), and we get this visualization:



In raw data, the start month has 0; then we ignore it and only care about Jauary to December. Next, we find there is no relaitonship, because the curve is very flat, which means The distribution of the number of eruptions per month is more uniform. Instead, we change the start month to start year, and then we want to see the trend of volcanic eruptions along with years.

The x-axis represents the start year, and y-axis represents the number of records. The data is about all volcanoes eruption from BC to 2018. This visualizaiton is to present the number of volcanoes eruptions along with years, and The numbers on the line are the number of records in corresponding years. In addtion, when put the mouse on the line, the tip label with start year and number of records will be showed. Negative year means it is in BC, for example, -4050 means that 4050 BC.

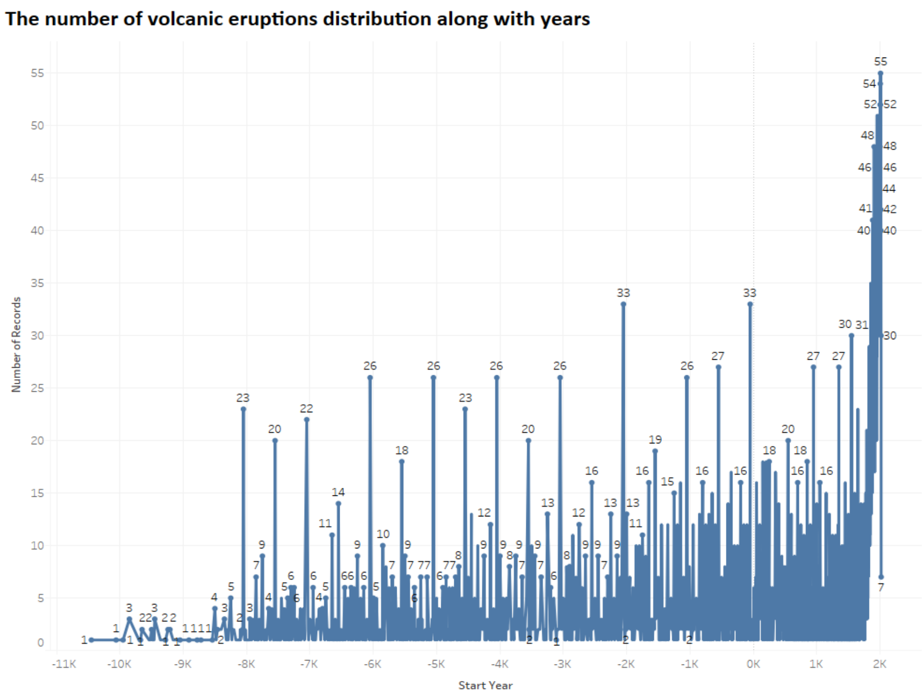


Figure 8.

The information we can get from the visualization is that from -8050, -7050, -6050, … , -2050, … , -50, the number of records separetely are 23, 22, 26, 26, 26, 26, 33, 26, 33, then from the BC, every 1000 years the number of volcanic eruptions increases, and that is, the probability of volcanic eruption is much larger in every 1000 years. Futhermore, after the AD, records are 27, 27, 30, 31 in year 950, 1350, 1550, 1878, we can see that the deuration is shortened, and volcanic eruptions are more frequent, expecially from the 1500 to 2017.

The reason why volcanic eruptions are more frequent, especially active in recent years, one is that may be the data in volcano eruptions in BC is analyzed by experts, and still has many eruptions in BC didn't know and record. Another reason is that, environmental changes in recent years, and have increased the probability of extreme weather.

2.2.8 Relationship between volcano types and tectonic setting

The attribute tectonic setting is splited into two subattributes: tectonic zone and crust depth. For tectonic zone, it contains three main type: Intraplate, Rift Zone, and Subduction Zone. For crust depth, there are also three main type: crust depth < 15, crust depth 15-25, crust depth > 25.

According to the volcanoes type's count and crust depth in each tectonic zone, it is obvious that no matter for which types of volcanoes, most of them are located at the crust depth over 25km. On the other hand, most of the volcanoes located at the subduction zone, which are sites of convective downwelling of Earth's lithosphere (the crust plus the top non-convecting portion of the upper mantle). The type of Stratovolcano are mostly located there as well.

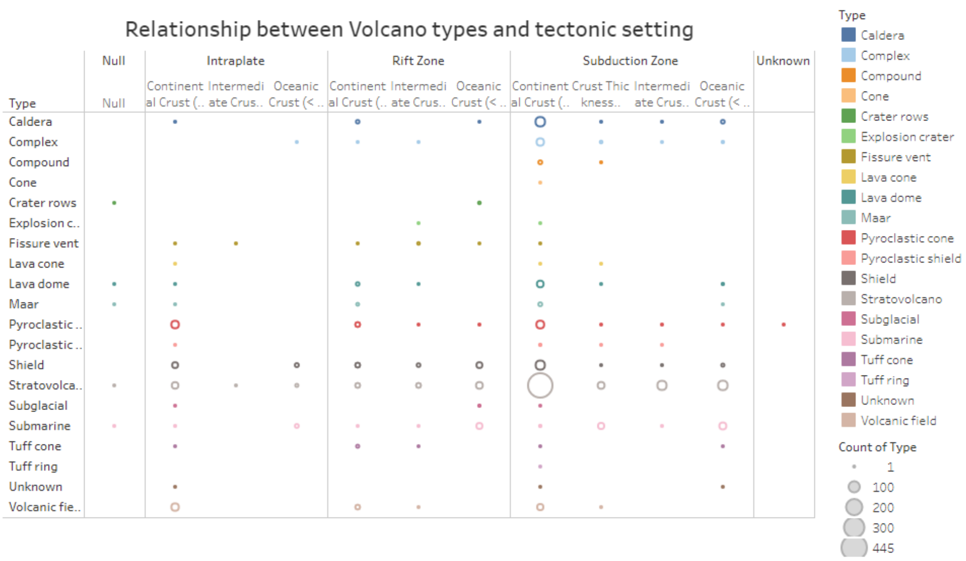


Figure 9.

2.3 Geographical visualization using Python

2.3.1 Implementation

We have used these libaraies to implement the plotting on earth map in python:

```
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
import numpy as np
import imageio
import csv
import pandas as pd
```

2.3.2 Two databases for visualization

Databases ‘vocanical_database(inUsed).csv’ and ‘earthQuake.csv’ are ultilized. ‘earthQuake.csv’ file contains global significant earthquake data from 2150BC to present. We have generated an earth map model, plotted the coordinate points with 2 colors representing 2 datasets on the map, produced mapping images and converted the images into gif file.

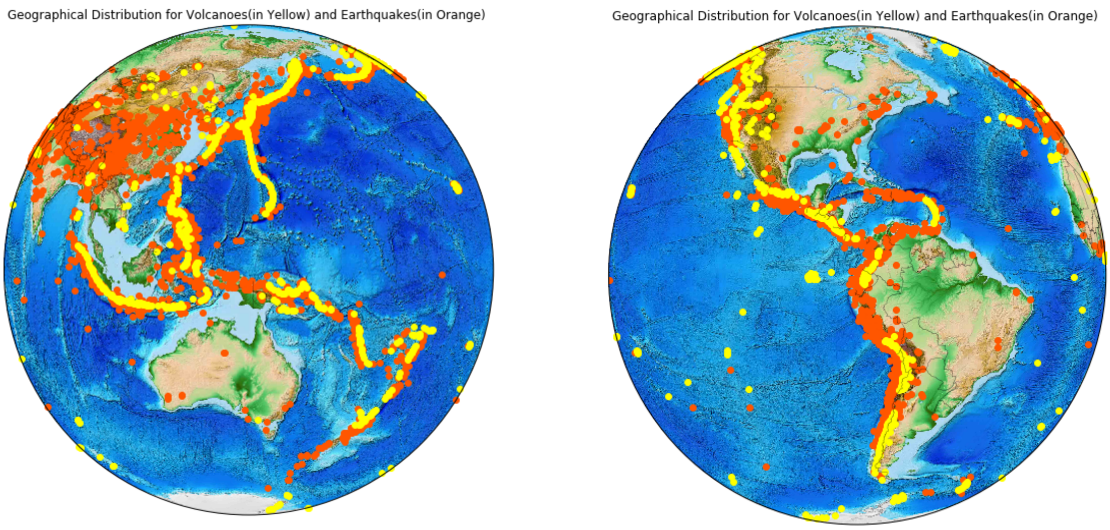


Figure 10.

2.3.2 Data visualization and analysis

Orange points represent earthquakes’ spots, while yellow ones represent volcanoes’ spots. We think the colors in the visualization look comfortable and are convenient for observation since they are opponent colors. In addition, plotting geographical data on a map is more intuitional than doing that on a chart or a graph.

The purpose we choose these two datasets to visualize is that we want to know whether there is a relationship of them, and then find out some insights. Fortunately, there are some clues on the map showing they are likely relevant. We can find that places volcanoes exist have high possibility that earthquakes had happened or will happen. There are critical similarities, though there are obviously differentials as well, for instance, a part of earthquake spots is located in inner mainland, while a majority of volcanoes is located by and off mainland—seldom inner mainland. For example, overlaps of yellow and orange points shape in paths, like banded lines. We know that sometimes earth plate movements are likely to cause earthquake and/or volcanic eruptions. Therefore, we hypothesis some of these overlapping paths may be the earth plate boundaries.

3. Difficulties and Limitation

3.1 Cesium globe visualization

It is initially designed as an automatic rotating globe, with interaction of the mouse dragging and clicking. However, after searched online, there is no built-in method to be used directly, it might be possible to make the camera rotating to the center of the globe, but it is difficult to compute the coordination through current information, the other way is keeping input a mouse moving parameter, and it would be complicated to modify the source code if we tried to achieve that.

3.2 Geographical visualization using Python

We want to draw the earth plate boundaries on the map also, so that it can proof our assumption. However, we have not done that, because there is no such function provided by Matplotlib library in Python. And we have not found out relevant data that can be applied on our map model.

4. Distribution

Fei Liu:

Website creating,

Report wringting, combination and modification,

2.2.3 Relationship between tectonic setting and geographical area.

2.2.7 The number of volcanoes eruptions distribution along with years.

Xinwei Li:

Report wringting,

2.1 Globe visualization using Cesium

2.2.1 Visualization about top 30 unrest volcanoes.

2.2.2 Top 5 unrest volcanoes with VEI.

2.2.6 Statistics of volcanoes in each region.

2.2.8 Relationship between volcano types and tectonic setting

Video making

Qingqing Yao:

Report wringting,

2.2.4 Visualization for Volcanoes Type.

2.2.5 Detail information between volcanoes type and regions.

2.3 Geographical visualization using Python

5. References

[1] The Smithsonian Institution, Volcanic eruption in the Holocene period. [Online] Available at: https://www.kaggle.com/smithsonian/volcanic-eruptions/data

[2] The Smithsonian Institution, http://volcano.si.edu/search_eruption.cfm

[3] National Geophysical Data Center / World Data Service (NGDC/WDS): Significant Earthquake Database. National Geophysical Data Center, NOAA. doi:10.7289/V5TD9V7K