Alexandra Flores
Alyssa Difurio
Bekhem Horne
Shaunghnessy Robertson

Project 3 Group 3: Volcanoes

Data:

Our data, “The Volcanoes of Earth”, was sourced from the Kaggle datasets database in a CSV format (https://www.kaggle.com/datasets/deepcontractor/the-volcanoes-of-earth). This dataset needed to be cleaned to be used, first step was to ensure that the volcanoes populated in the correct hemisphere on the map, we first had to adjust the latitude and longitude positive and negative signs.

```
df.Latitude=df.Latitude.apply(lambda x: "+" + str(x) if "N" in str(x) else "-" + str(x))
df.Longitude=df.Longitude.apply(lambda x: "+" + str(x) if "E" in str(x) else "-" + str(x))
df.head()
```

Latitude and longitude were then converted into floats after the direction abbreviations and degree symbols were removed. The summit and elevation were in one column in the dataset, so we had to separate them into their own column and get rid of the meters and feet abbreviation. After that was finished, we could next fix misspellings, swap out duplicate names, remove null values, and remove unnecessary columns.

```
# Remove degree symbol from Latitude and Longitude
df['Latitude'] = df.Latitude.astype(str).str.replace('°N','')
df['Latitude'] = df.Latitude.astype(str).str.replace('°S' , '')
df['Longitude'] = df.Longitude.astype(str).str.replace('°E' , '')
df['Longitude'] = df.Longitude.astype(str).str.replace('°W' , '')
df.head()
```

Our Inspiration and Research Questions:

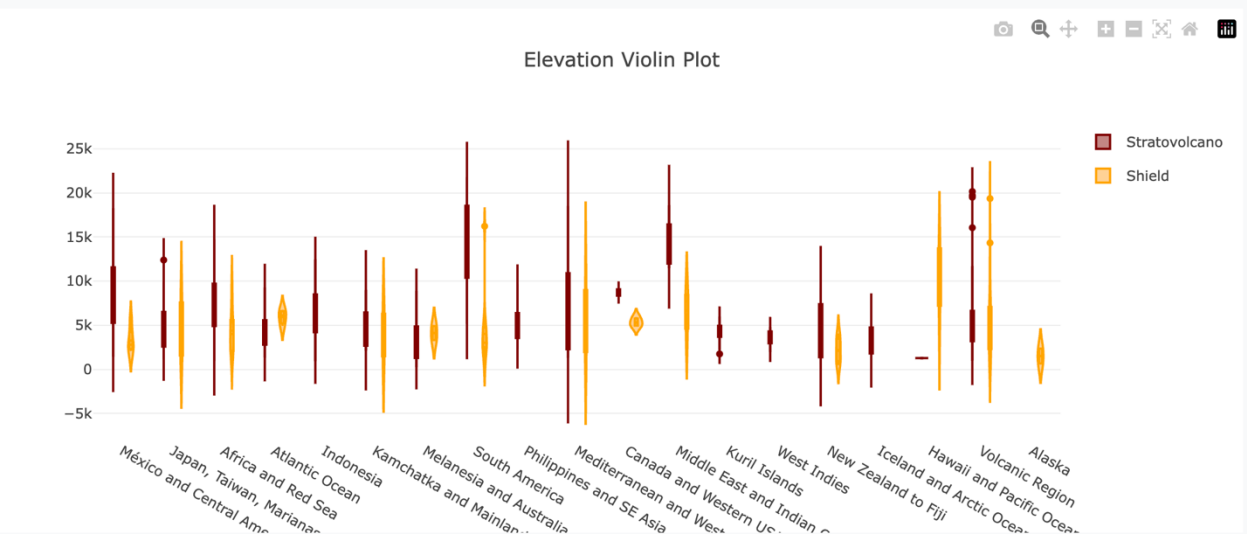
Due to a shared love of the Johnny Cash song "Ring of Fire," group 3 has expressed interest in working on our visualization project on volcanoes. We were all painfully and recently dealing with relationship issues, so the song's lines, "Love is a burning thing, and it makes a fiery ring, bound by wild desire, I fell into a ring of fire," resonated with us. As you may or may not know, The Ring of Fire, also referred to as the Circum-Pacific Belt, is a path along the Pacific Ocean characterized by active volcanoes and frequent earthquakes. Most of Earth’s volcanoes and earthquakes occur along the Ring of Fire due to the tectonic plates that move in that area.

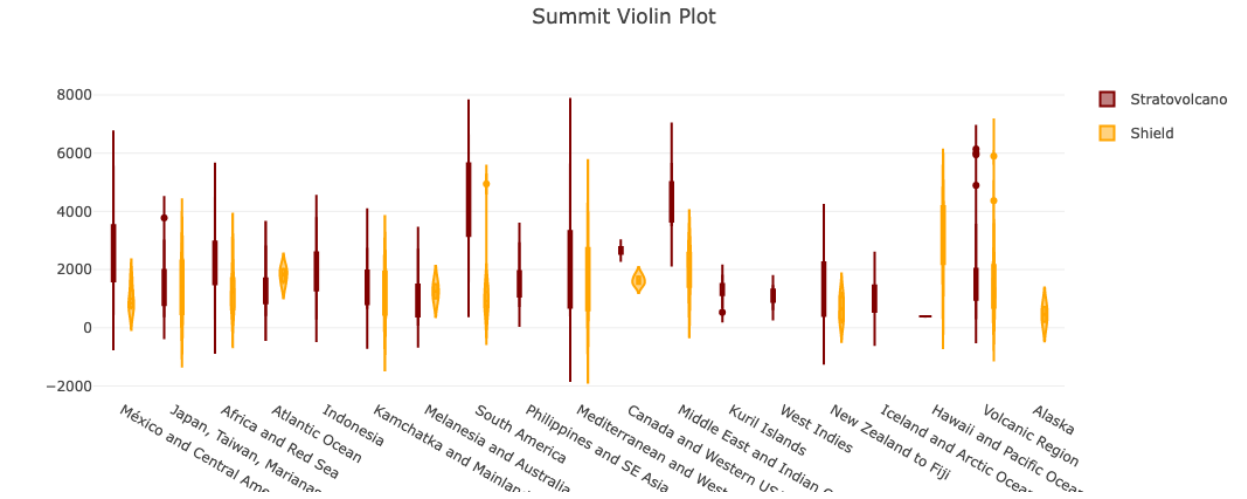
Our research questions are:

1. Does type affect the size of a volcano?
2. Which region and subregions have the most volcanoes?
3. Most common locations where volcanoes reside?
4. What does the summit and elevation look like when compared by region or volcano type?

Visualizations:

One of the visualizations that we have chosen is a violin plot to depict the different measurements of summit and elevation per region. We have also created drop down filters to be able to compare the summit of two different volcano types. We also created the same violin plot but to compare elevation of two different volcano types as well.

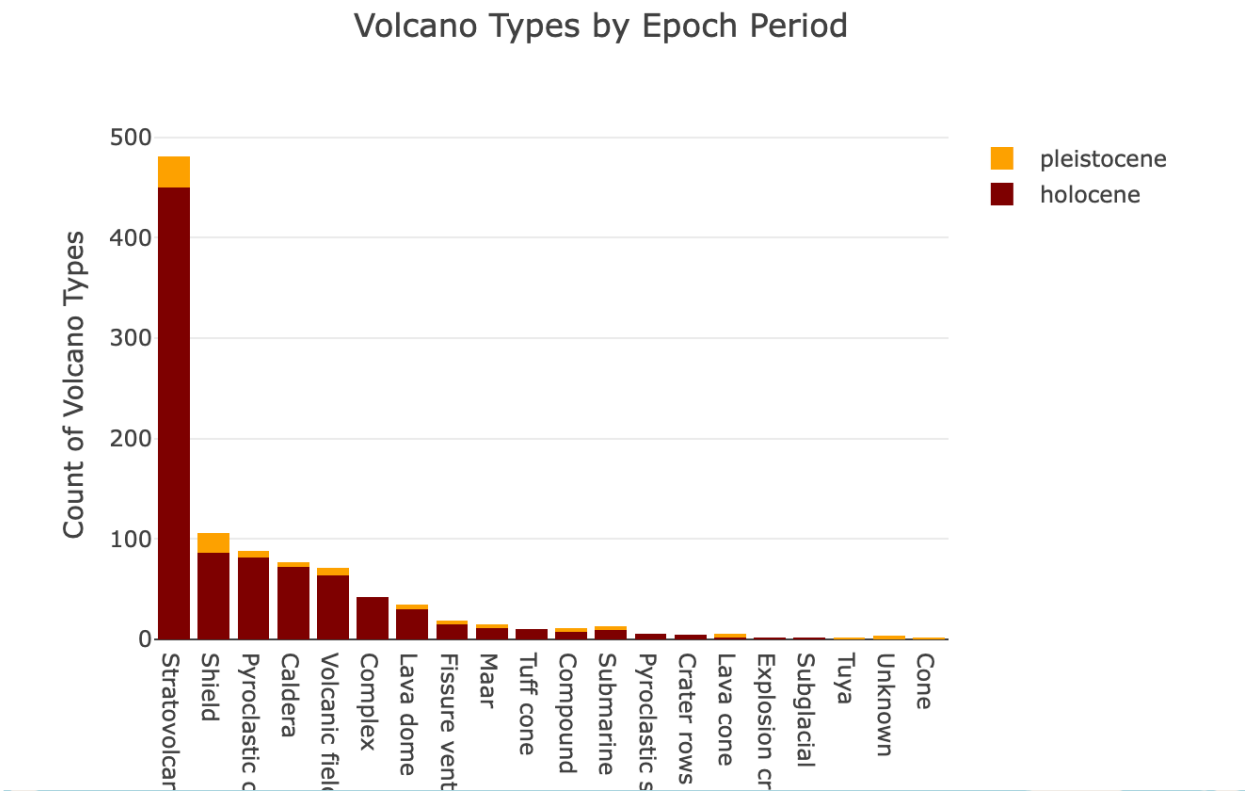




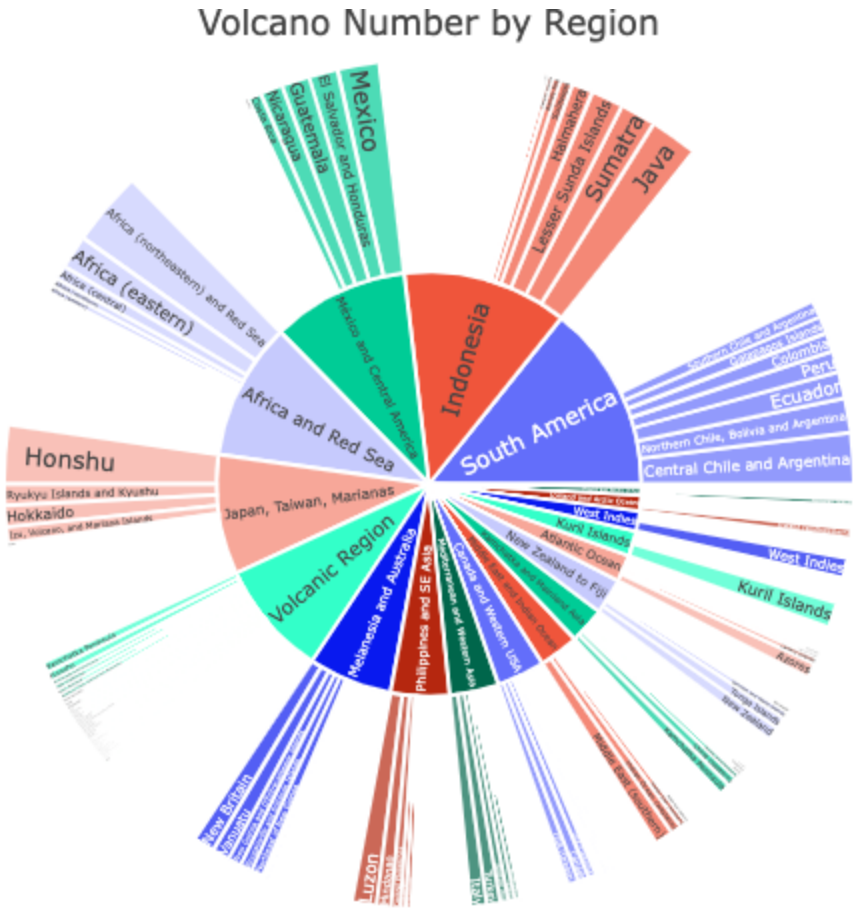
The visualizations turned out as the images above and we saw that although the summit and elevation numbers differed, the shape of the violin plots remained the same as you compared two different volcano types.

In our data we were also given the epoch period of each volcano, and there are two different epoch periods, Pleistocene and Holocene. The Pleistocene is the geological epoch that lasted from about 2,580,000 to 11,700 years ago, spanning the Earth's most recent period of repeated glaciations. The Holocene is the current geological epoch. It began approximately

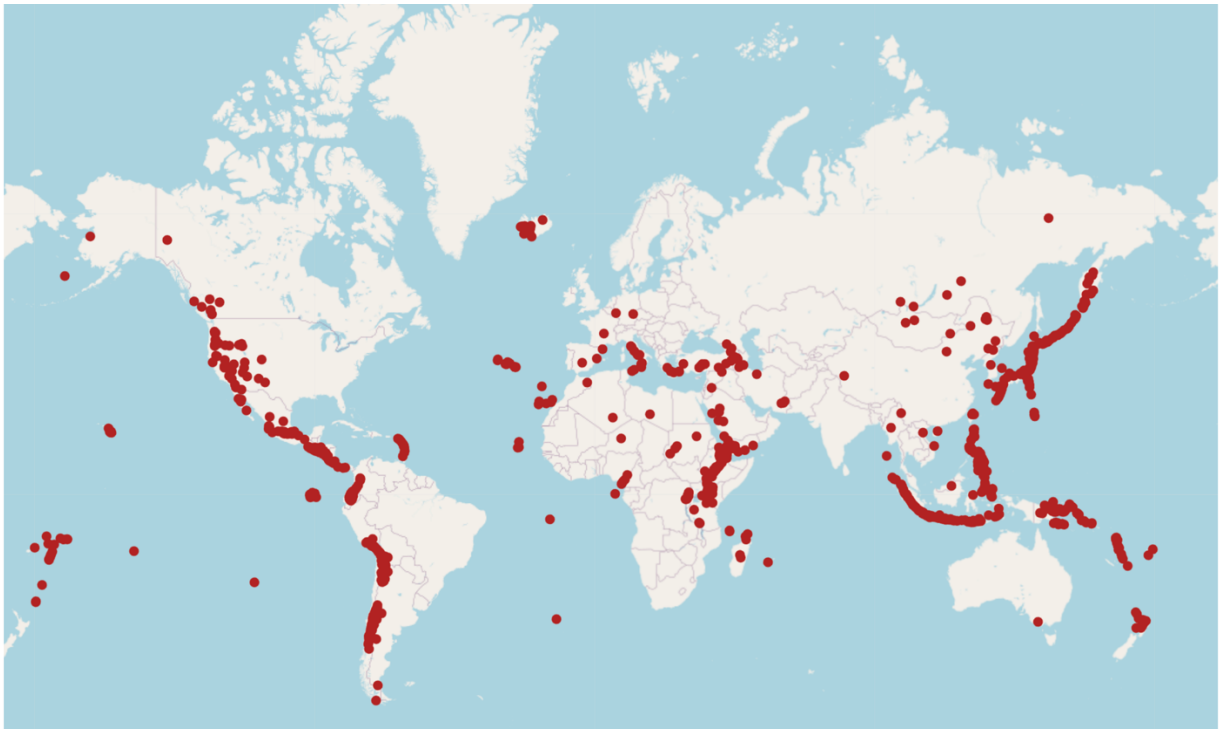
11,650 cal years Before Present, after the Last Glacial Period, which concluded with the Holocene glacial retreat. The Holocene and the preceding Pleistocene together form the Quaternary period. We created a stacked bar chart that shows the number of volcanoes for each volcano type for each epoch period. We were also able to create a filter for this chart that sorts the data based on region. Overall, it was clear to see that the stratovolcano is the most common volcano type.



Another visualization we chose was a sunburst chart which shows the number of volcanoes in each region and sub region. The chart below answered our research question "Which region and subregions have the most volcanoes?". As we can observe, the South American and Indonesian region contains the largest number of volcanoes, water-based regions have the least number of volcanoes, and the remaining regions have on average, less than 30 volcanoes.



We were also able to build a map that shows the location of each volcano throughout the world.



This map is also interactive as we can filter the data by region. This map not only shows the location and the name of each volcano but also the period in time where it lasted erupted.

Conclusion:

Our data and visualizations allowed us to see a geographic spread of our data. Each data set answered our questions differently. The first data set showed us the significant difference between the number of volcanoes by type via a bar graph. Our next set showed us the spread of volcanoes by region, it allowed us to see the disparity of volcano locations based on region and subregion. Our third data set was visualized in a map; using the latitude/longitude of the volcanoes we were able to create a map that had each point plotted with their last active date. And our final data set was shown through a violin plot, where we were able to look at the comparison of elevation and summits by type and region. Overall, our data was able to provide conclusive statements that answered our questions, however only answered questions very specific to the physical aspects of volcanoes.

Further Work and Limitations:

The data set we chose was clean and neat, but we did see that we had a lack of data that we could use. With the data that we did have we did not use the population data regarding the population surrounding a certain radius. Given more time we could've used that data potentially on our map to show the number of people that are affected by the presence of each volcano. A limitation that we did face was that the last eruption date we were given was not an actual date but a period in time so we could not use it to create another visualization.