Gabby Hummel

Prof. Pockalny

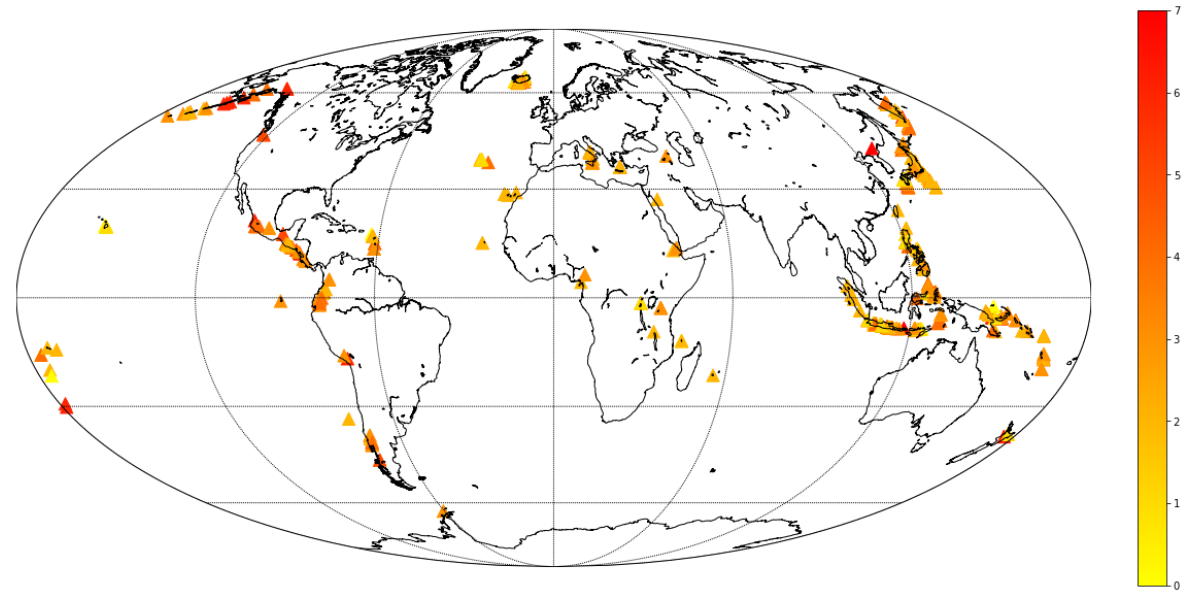
OCG 250

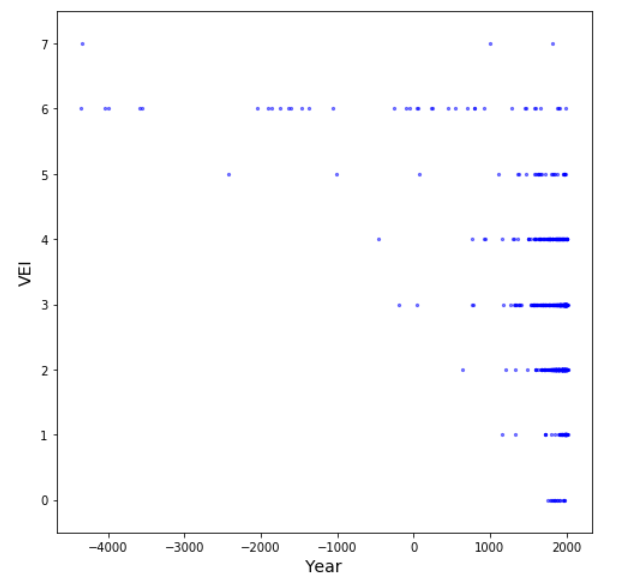
3 May 2018

Significant Volcanic Eruptions: A Statistical Analysis

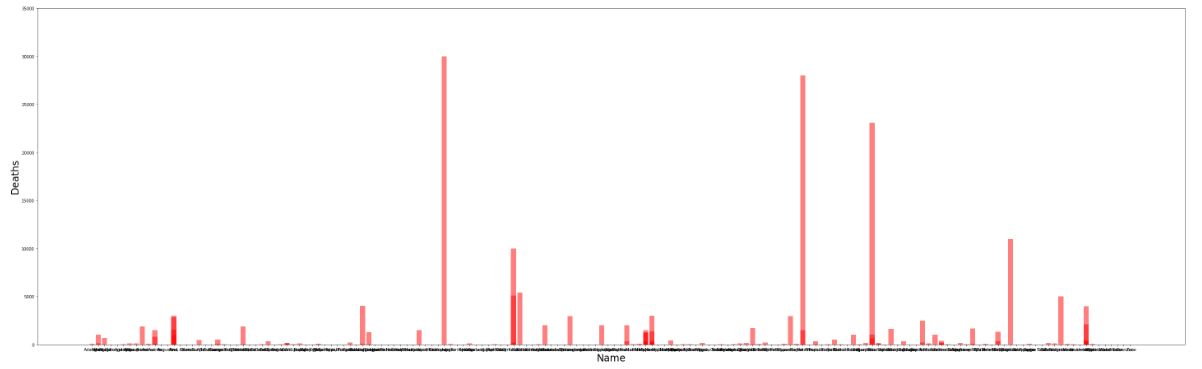
Statistical Analysis was performed using Python 3 through Jupyter Notebooks on The Significant Volcanic Eruption Database. This database was compiled by the National Oceanic and Atmospheric Administration (NOAA) as a global listing of significant volcanic eruptions throughout history. NOAA used the following criteria to determine if a volcanic eruption was “significant”: if the eruption caused fatalities, caused approximately $1 million in damage or more, had a VEI of 6 or larger, caused a tsunami, or was associated with a major earthquake. I primarily focused my statistical analysis on fatalities and VEI. Similarly, considering this database contained data from thousands of years ago, I was interested to discover how volcanoes have change over the years, and whether or not the changes I observed were true changes, or the result in a change in record keeping.

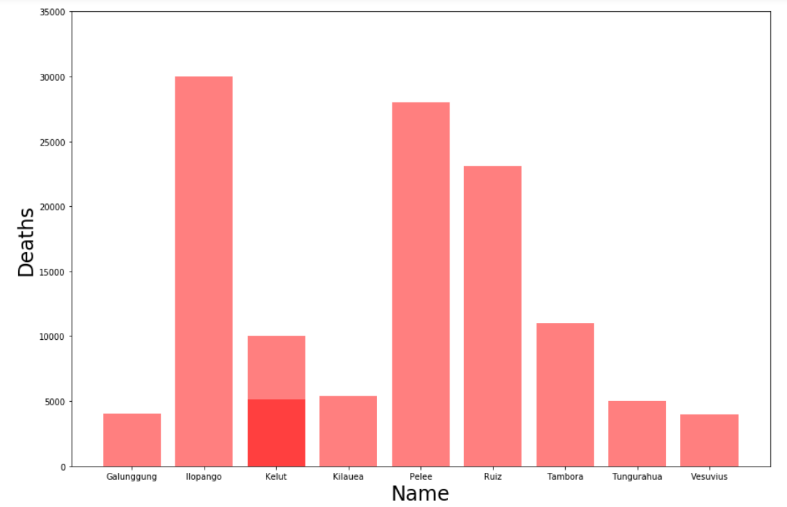
I began my statistical analysis, after importing the necessary software and the database csv file, by defining variables. I defined variables for latitude, longitude, vei, deaths, year, and name. I also created numpy arrays out of longitude, latitude, vei, year, and deaths variables. Then, I create my first visualization. The goal of this visualization was to plot the location of every volcano that is mentioned in the database on a world map. Each volcano is represented by a triangle, and is colored on a scale according to the VEI. By this scale, the brighter red triangles represent volcanos with a VEI of 7 (the greatest VEI in this database) and the yellow triangles represent a VEI of 0. This was done by plotting a scatter plot using the numpy array vei variable over a world map. This visualization can be seen below:

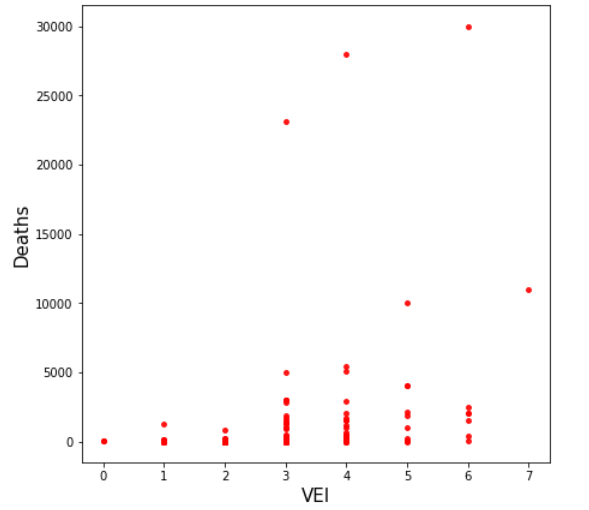


Next, I plotted a scatter plot to visualize VEI over the years. I did this to understand whether or not volcanic eruptions have increased in VEI overtime. My interpretation of this data is not so much that volcanic eruptions have been decreasing in VEI overtime, as the visualization may immediately suggest, but that record keeping has improved overtime. The earliest volcanos recorded in this database are of higher VEIs, likely because the higher VEI volcanic eruptions were the ones about which data survived long enough for modern scientists to know the eruption occurred. This visualization can be seen to the left.

Next, I was curious to see the number of deaths caused by volcanos visualized. I did this first by creating a bar graph with the x-value being the names of the volcanos, and the y-value being the number of deaths. I used pick\_range to limit the data being read into this code to only the volcanos that caused fatalities. My code resulted in this graph:



Because so many volcanos were listed on this graph, making it hard to read and understand, I decided to limit the graph to only the top ten most fatal volcanos. I did this again by using pick\_range. This graph can be seen to the right.

I then created a scatter plot to visualize the relationship between fatalities and VEI. This was done with the intention of understanding better how VEI is related to the number of fatalities caused by the eruption. I went into this with the hypothesis that a greater VEI would result in a greater number of deaths. However, I found that the greatest number of recorded deaths was a 6, with a 7 having considerably less fatalities. Similarly, only one of the eruptions recorded at a 7 had resulted in fatalities, according to this dataset. Also, eruptions classified at both a 3 and a 4 had considerably more fatalities than the one eruption classified at a 7. My conclusion is that fatalities is likely related not only to VEI, but also to location. Volcanos that are located in highly populated areas are much more likely to result in a greater number of deaths, than an every greater eruption located in a remote area. Similarly, I would hypothesize that volcanos that do not erupt often are more likely to have large fatalities, as populations living near the volcano are not prepared for or expecting the eruption. However, this hypothesis is not entirely supported by the data, as Kelut has erupted multiple times, and in fact is listed twice in the top ten most deadly volcanic eruptions.

Finally, using the savefig function, I created 8 graphs measuring VEI over the years, to determine the relationship between explosivity and year. The savefig function saved each graph to my computer, which I then imported into a movie using powerpoint.

Below, the code I wrote to analyze and visualize this data can be found.

