Metadata

Author: Mingda Lu
Time Created: March 06, 2020
Script: MingdaLu270_program_07.py
Description: The script was created to graphically analysis the earthquakes data for past 30 days.

1. Input Data

Data was downloaded from website operated by USGS Earthquake Hazards Program as all_month.csv on the same day the script was created (March 6, 2020). More data and information please refer to the following website:
https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php

Data was stored in a .csv file (comma separated value file) and contained multiple columns, namely: time, latitude, longitude, depth, mag, magType, nst, gap, dmin, rms, net, id, updated, place, type, horizontalError, depthError, magError, magNst, status, locationSource, and magSource. Only part of these columns were used for analysis.

2. Script Notes

Few packages were used for this script in order to complete the data analysis, such as pandas, matplotlib, numpy, scipy, statesmodels.api, and pylab. Pandas.read_table commend was used to read the data instead of genfromtxt() due to the diversity of the data types.

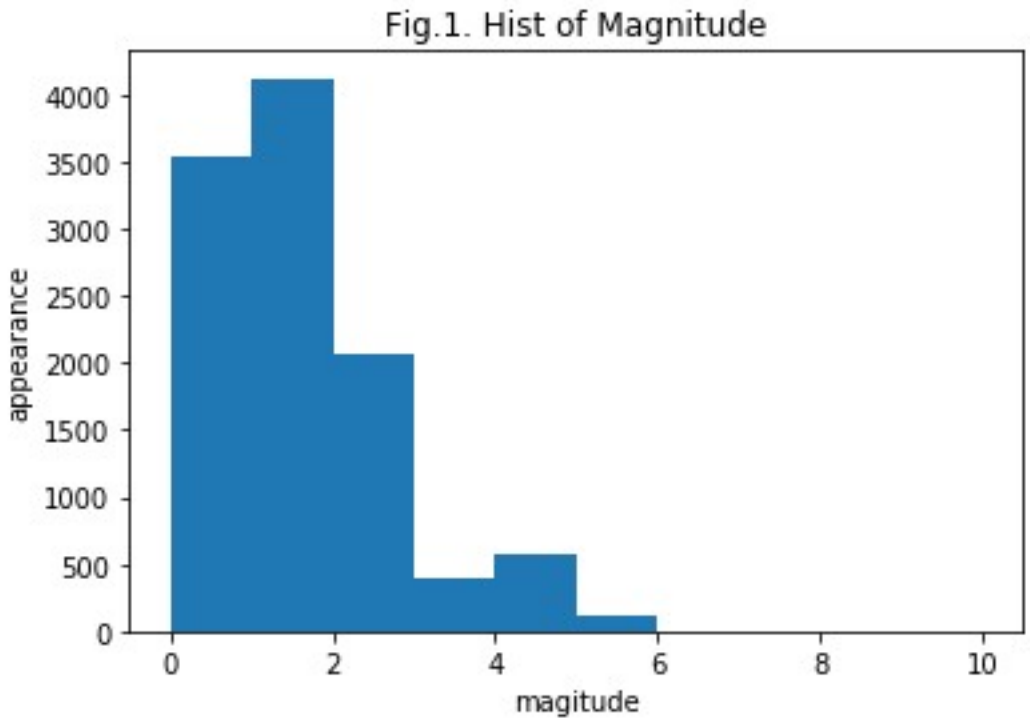
Six figures were generated for data analysis purpose, as:

1. Histogram of Earthquake Magnitude
2. KDE Plot for Earthquake Magnitude
3. Latitude & Longitude Scatter Plot
4. Normalized Cumulative Distribution Plot for Depth
5. Magnitude & Depth Scatter Plot
6. Quantile-Quantile Plot for Earthquake Magnitude

Each figure will be discussed in detail below.

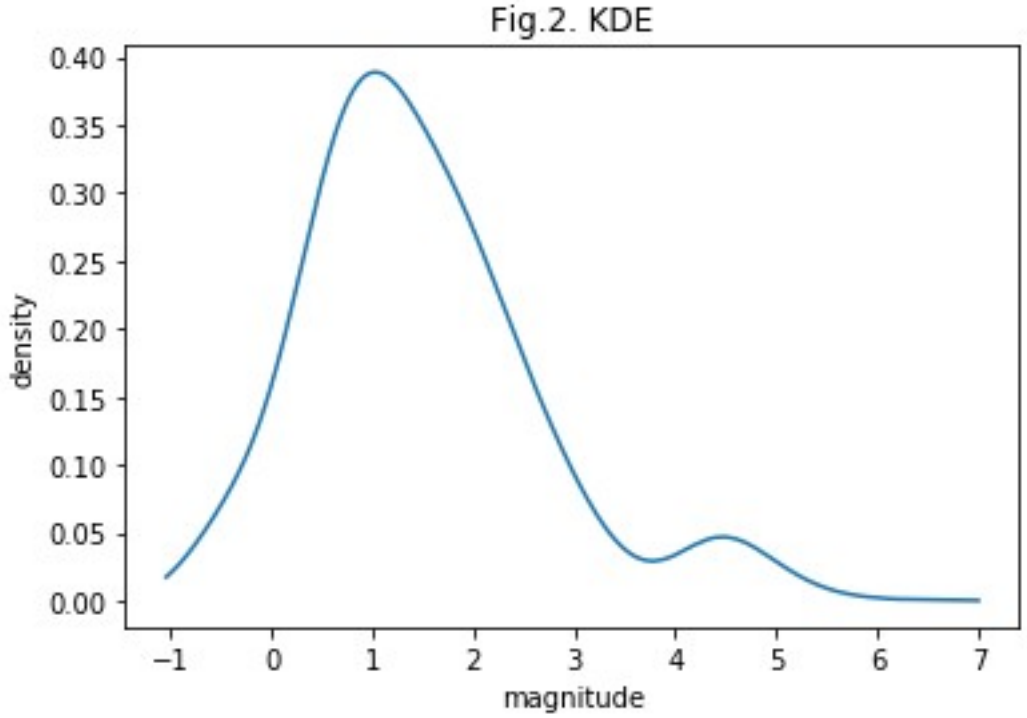
3. Graphical Data Analysis

1. Histogram of Earthquake Magnitude



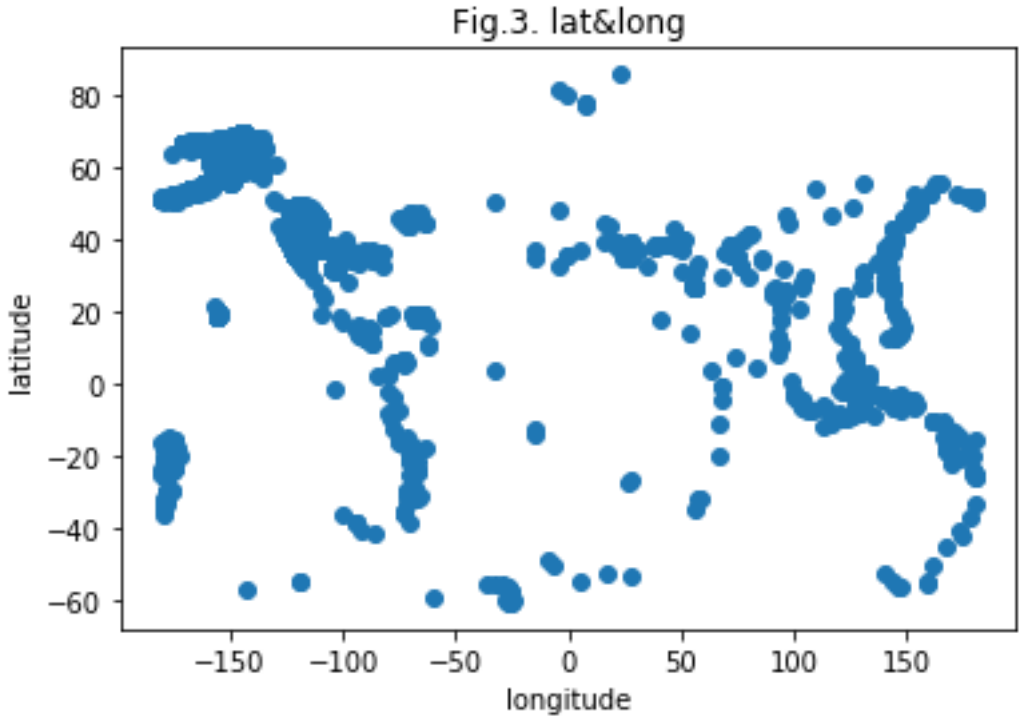
10 bins with width of 1 and a range of 0 to 10 was selected for this graph. Width and range change will change the shape of the graph. The graph will be more accurate and analyzable if the width increases and range decreases. In this case, with the given width and range, the data distribution is somehow suggested by the histogram as a normal distribution.

2. KDE Plot for Earthquake Magnitude



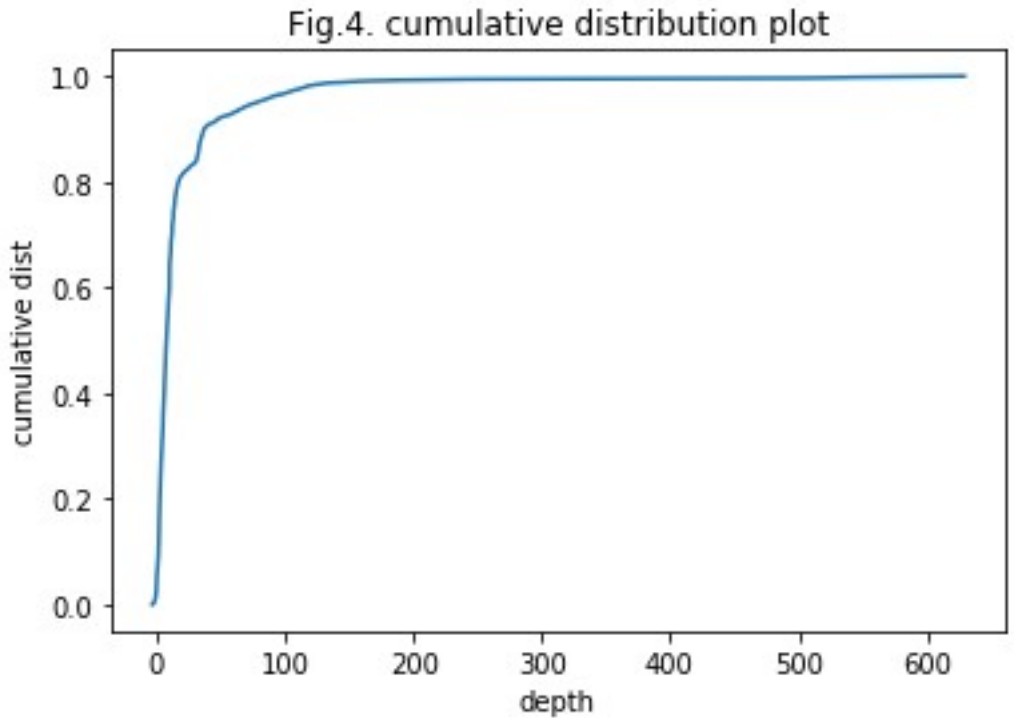
The KDE plot is affected significantly by the kernel type and width. In this case, Gaussian was selected, and the width was selected as 0.3. Bandwidth is commonly chosen proportional to (scale)/n^(1/5). Small bandwidths can be used for finer scale information analysis. KDE uses distances to known samples in order to assign probabilities. Which all points should not have a zero probability. However, histogram is a simple and fast way to visualize the appearances of data for each range. Appearance could be zero.

3. Latitude & Longitude Scatter Plot



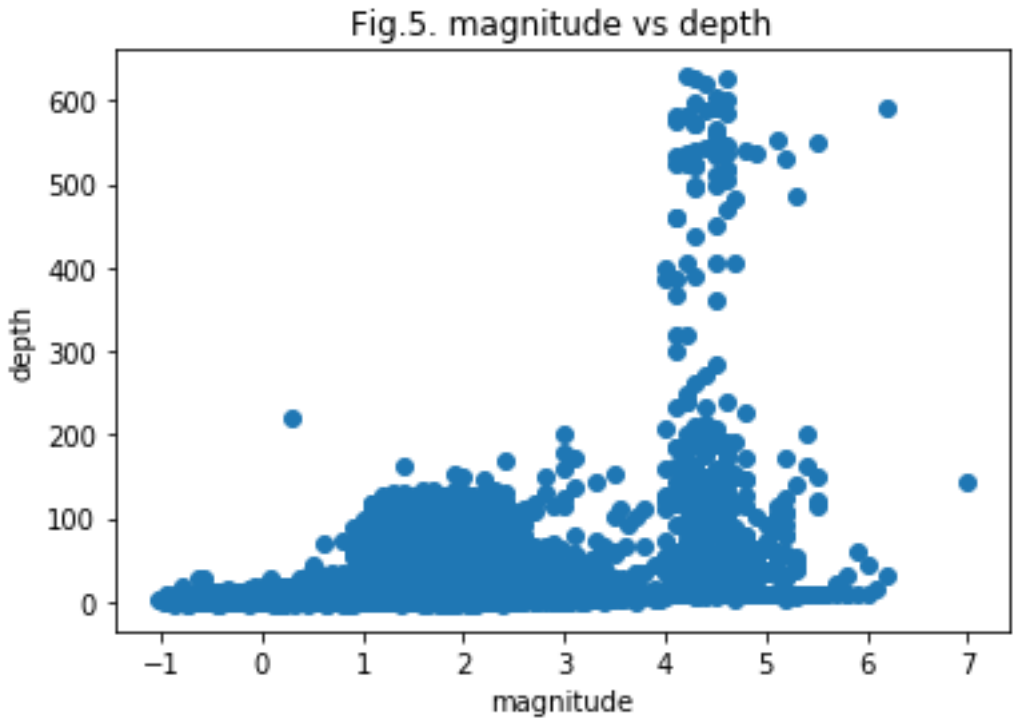
Longitude was put on x and latitude on y because the map was setup this way. Based on the distribution of these points, obviously earthquake happens more often around pacific plate.

4. Normalized Cumulative Distribution Plot for Depth



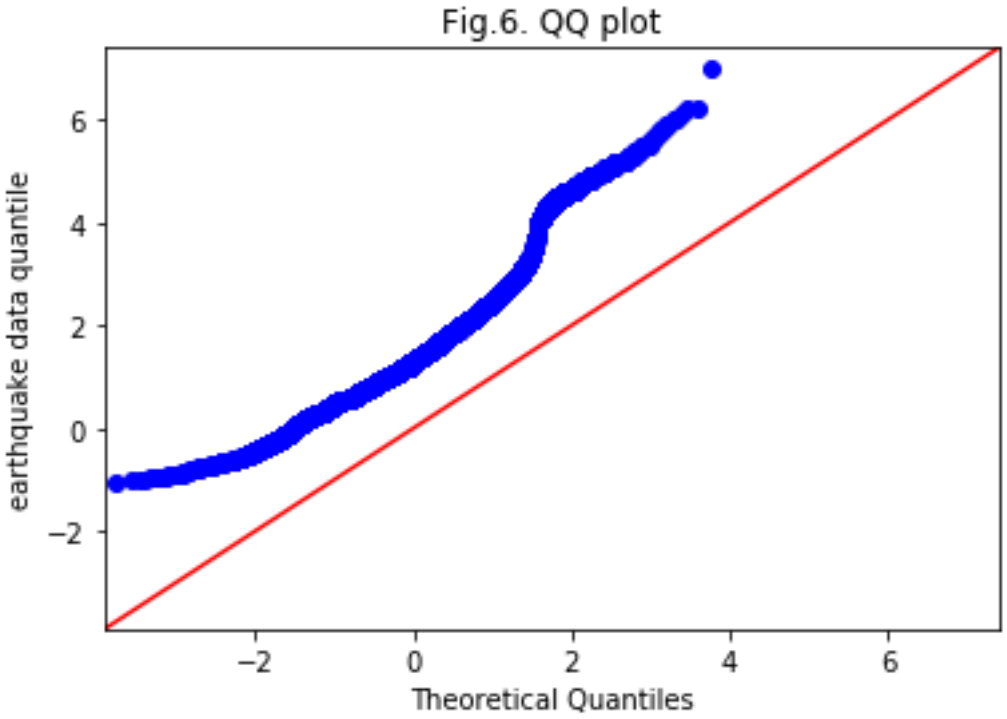
The slope can indicate the frequency of earthquake. Base on the figure, the slope between 0 to 30 is high, which the line in this range has a rapid rise. Which indicates that most earthquakes happened in past 30 days had a depth between 0 to 30.

5. Magnitude & Depth Scatter Plot



Based on this figure, earthquakes which had magnitude less then 4 had depth lest than 200 km, and for earthquake magnitude larger than 4, depths could be very large. So the relationship here is, the larger the depth, the larger probability to have a earthquake with large magnitude.

6. Quantile-Quantile Plot for Earthquake Magnitude



Normal distribution was selected as default in this case. A 45 degree line was drawn to see the fit of the data to normal distribution. The data does not lie on the line, which indicates that the population distribution is not normal distribution.