

Metadata file

4/27/2020 by Charles Huang

all_month.csv is the input file in csv format, downloaded from the USGS earthquake hazards web site (<https://earthquake.usgs.gov/earthquakes/feed/>) on April 27th, 2020, which contains all earthquakes happened for the last 30 days. The data in the spreadsheet is separated by comma sign, and contains following columns that are used in the program:

longitude: Decimal degrees longitude. Negative values for western longitudes.

latitude: Decimal degrees latitude. Negative values for southern latitudes.

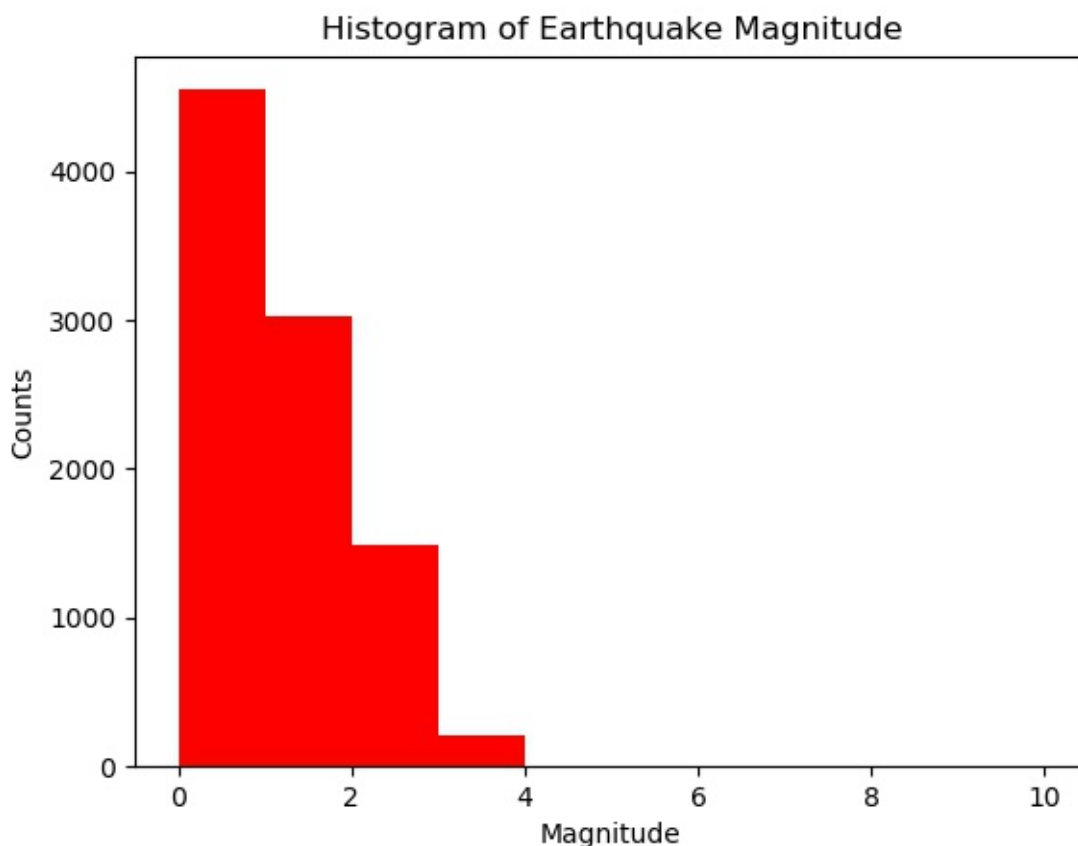
mag: Magnitude of the earthquake event.

depth: Depth of the event in kilometers.

Description of other variables can be found on the USGS website.

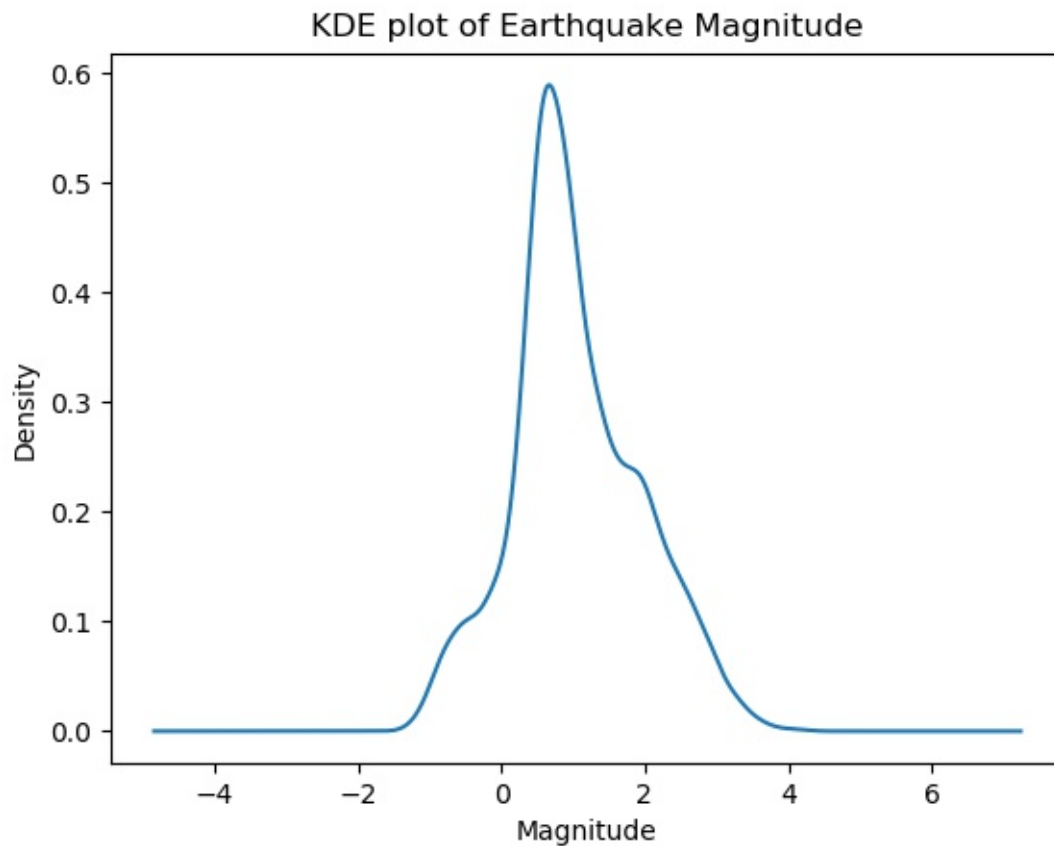
program-07.py is a script which uses python to read the input file, clean up the data by dropping missing value and generate plots for analysis. To read in the csv file, it uses `pd.read_csv()` instead of `genfromtxt()` because the csv file contains values in different format separated by comma, which `genfromtxt()` cannot handle. The script can 1. Generate a histogram of the earthquake magnitude; 2. Generate a KDE plot for the earthquake magnitude; 3. Generate a plot of latitude versus longitude for all earthquakes; 4. Generate a normalized cumulative distribution plot of earthquake depths; 5. Generate a scatter plot of depth vs. magnitude for all earthquakes; 6. Generate a Q-Q plot of the earthquake magnitudes.

Following are plots being generated for all_month.csv data:

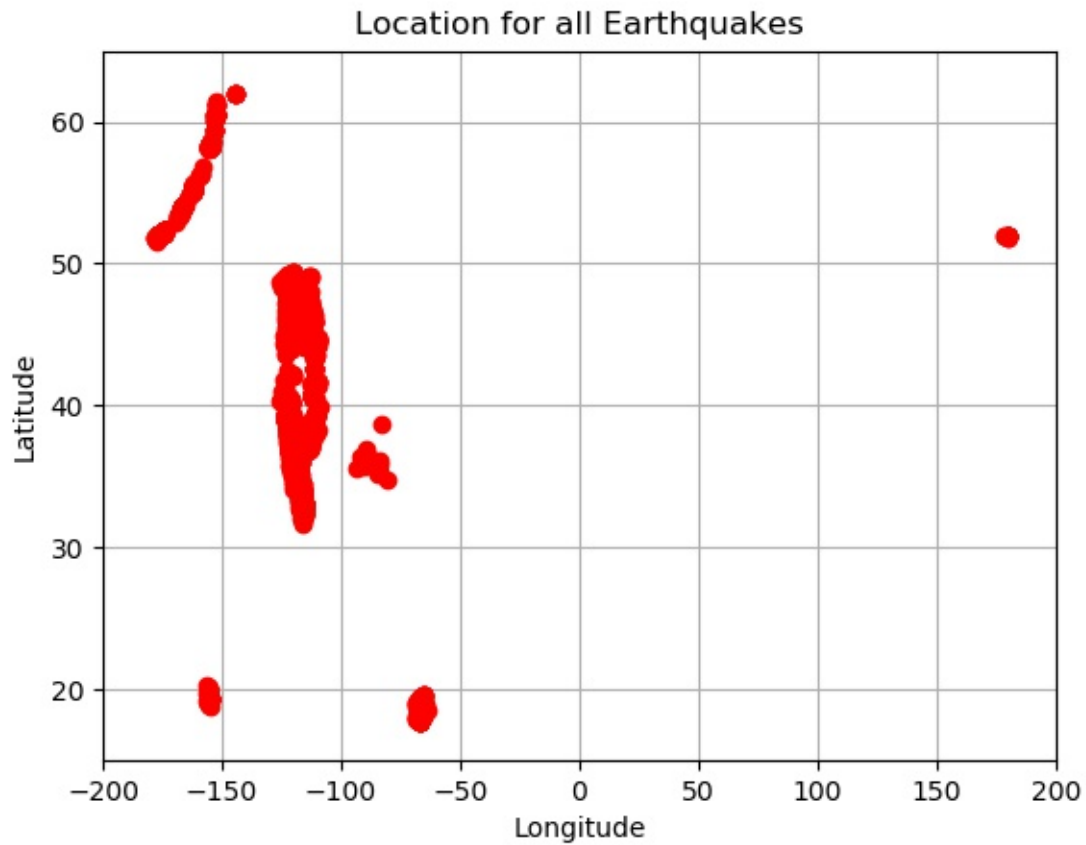


1. The histogram of earthquake magnitude skews to right and have no counts larger than magnitude = 4, meaning that only earthquake with small magnitude happened in the last 30 days, when most are

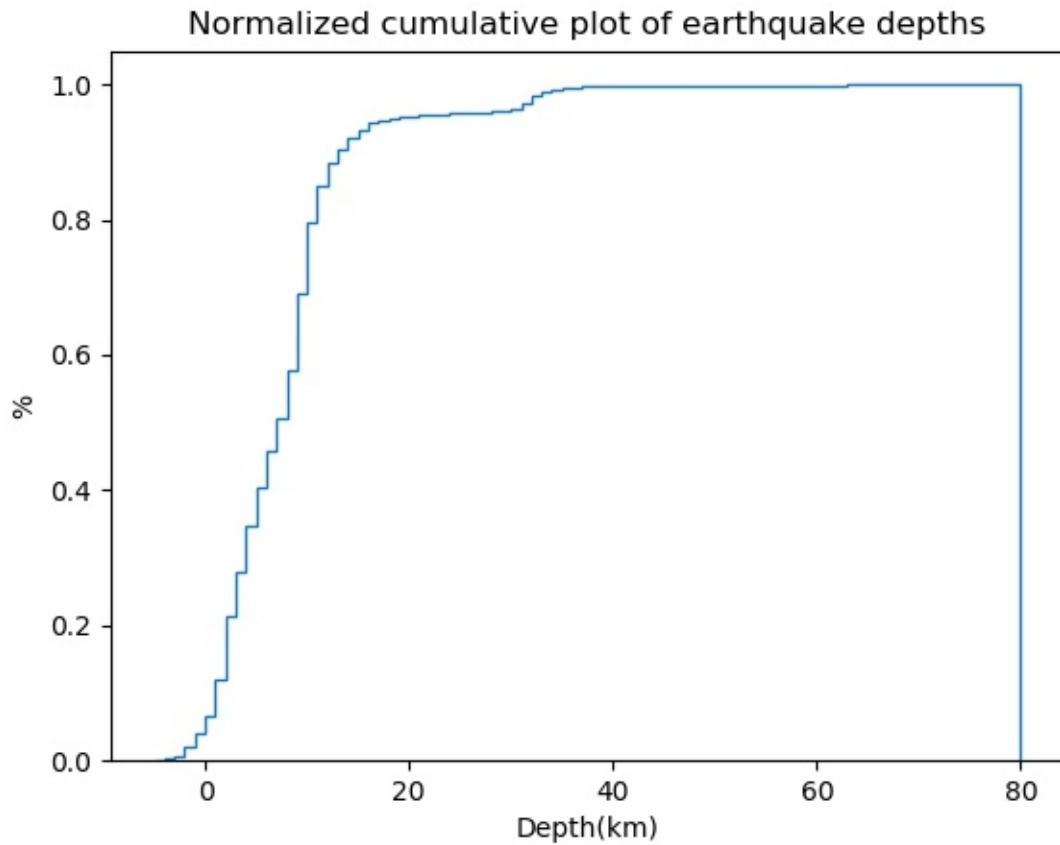
smaller than magnitude = 2. Since bin width is chosen as 1, some events (e.g. magnitude = 4.2) might be counted into Mg. = 4 in this case.



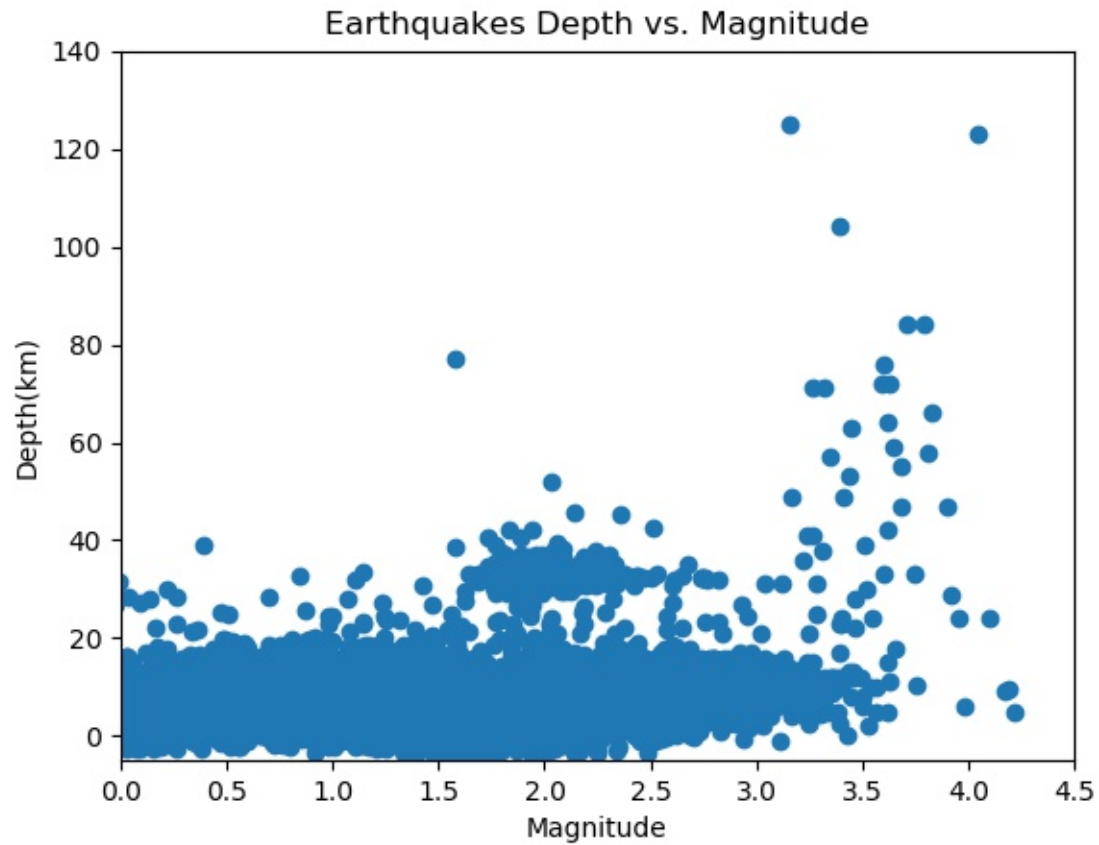
2. The KDE plot of earthquake magnitude is generated with kernel type and width being "scott" as default. The plot shows that almost 60% of the reported earthquakes are around Mg. = 1, which provides clear percentage of the whole population at each magnitude instead of numbers of events in the histogram.



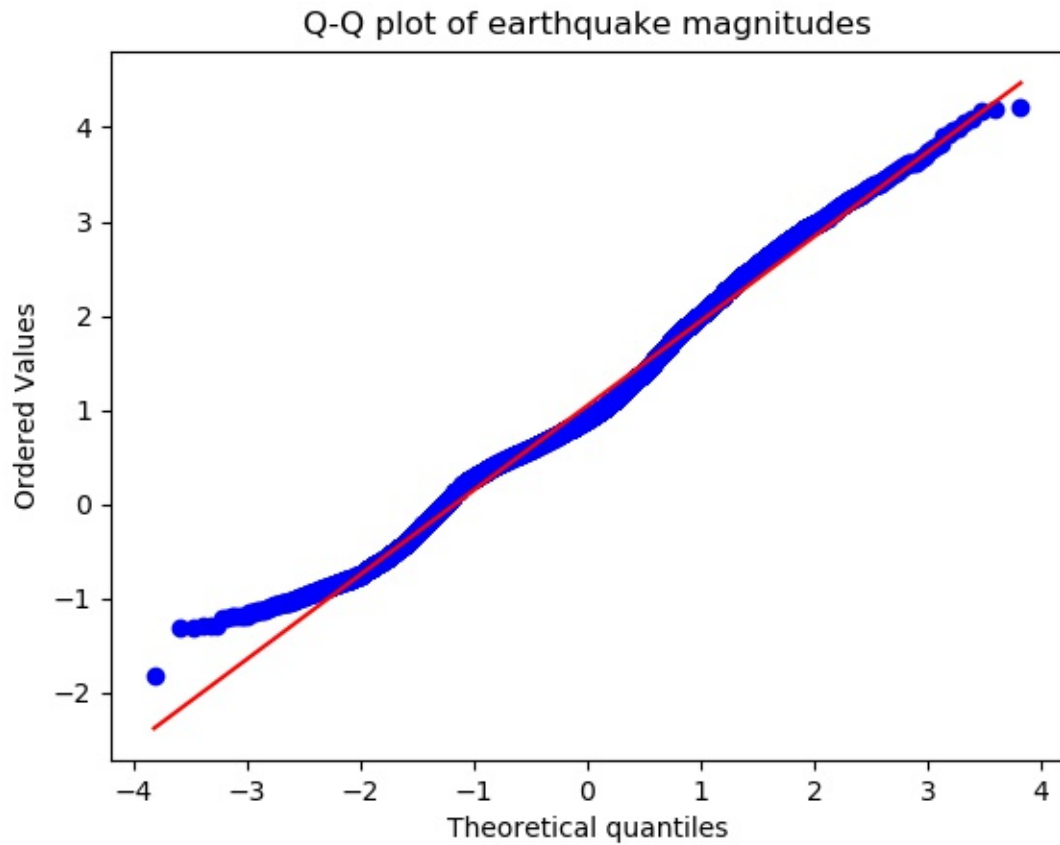
3. Data is plotted as longitude on the x-axis and latitude on the y-axis to match the global map. According to the plot, most earthquake events are distributed around (16-62N, 50-180 W), when a great portion of them cluster around (30-50N,105-128W), and some outliers are close to (50N, 180 E).



4.The normalized cumulative distribution plot of earthquake depths shows that about 80% of the earthquake happened above 10km depth, and 97% happened above 30km depth. It then starts to end at around 63km depth with a few more outliers all the way up to 140 km.



5. The scatter plot of earthquake depth vs. magnitude shows clustering of data at low magnitude and shallow depth. It seems like shallow depth generally have earthquakes with smaller magnitude, when those with larger magnitude (> 3) can happen in shallow layer but also more likely in the greater depth.



6. The Q-Q plot of the earthquake magnitudes is assuming the earthquake magnitudes should be normal distributed. According to the plot, data points are close to the straight line, which means that they are normal distributed