# Metadata

**Author:** Tao Huang (huan1441@purdue.edu), Purdue University

**Created:** Feb 28, 2020 (Version 1.0, used in Python 3.6)

**Name of the script:** huan1441_program-07.py

**Purpose:** Script to read the dataset about all Earthquakes for the past 30 days in the current directory, and generate six figures for the graphical analysis.

**Input Data Description:**

Data Source: https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/all_month.csv

Download Time: 09:51 am, 02/29/2020

Data Type: The data file (“comma separated values” (CSV) ASCII text file) records all the earthquakes happened around the world for the past 30 days (from 01/30/2020 to 02/29/2020), which including the 22 fields of time, latitude (degree), longitude (degree), depth (km), mag, magType, nst, etc. The data types contain date, time, string, integer, float, etc. The reason why genfromtxt() in numpy does not work with the data file is that the dataset in this case contains different types of data. Thus, read_table() in pandas is employed for this work.
(For more information about the original data, please click the link below
https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php)

**The process of the script is as follows:**

-- Read the data file "all_month.csv" in the current driectory by using read_table() in pandas library.

-- Delete the NaN in the data, extract the corresponding data and generate the following six figures (.jpeg) shown below by using matplotlib library for further graphical analysis.

(1) A histogram of earthquake magnitude (10 bins with width of 1 and a range of 0 to 10).

(2) A KDE plot of earthquake magnitude (Gaussian kernel type with kernel width 0.5).

(3) A scatter plot of latitude vs. longitude for all earthquakes.

(4) A normalized cumulative distribution plot of earthquake depths.

(5) A scatter plot of earthquake magnitude vs. depth.

(6) A Normal Quantile-Quantile plot of the earthquake magnitude.

-- The detail descriptions for the six figures generated by the script are presented as follows.

**(1) A histogram of earthquake magnitude**

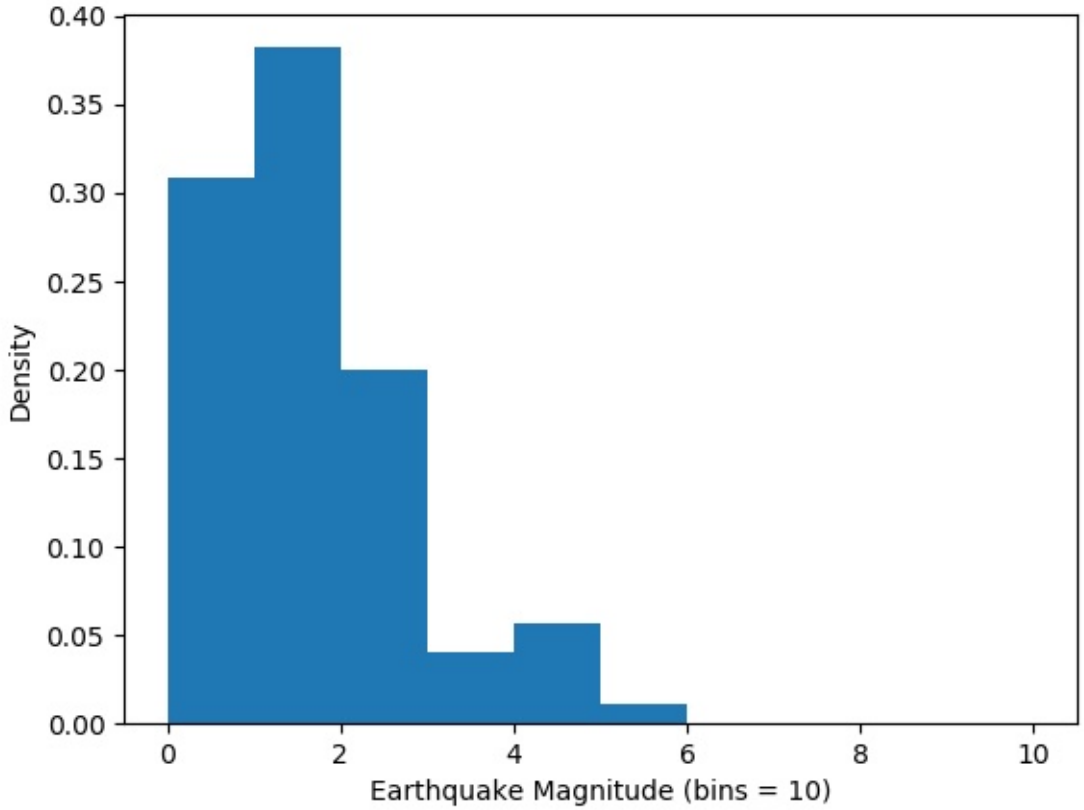


Figure 1. Histogram of Earthquake Magnitude

Figure 1 shows the density distribution of the earthquake magnitude. The density of magnitude from 1 to 2 is the highest (Density = 0.38), followed by the magnitude from 0 to 2 (Density = 0.31).

Hint: This figure is plot by 10 bins with width of 1 and a range of 0 to 10. The bin width and range would affect the display of the density distribution. An appropriate selection (neither too small nor too large for the bin width, and a reasonable range) would capture the main characteristics of the data distribution.

**(2) A KDE plot of earthquake magnitude**

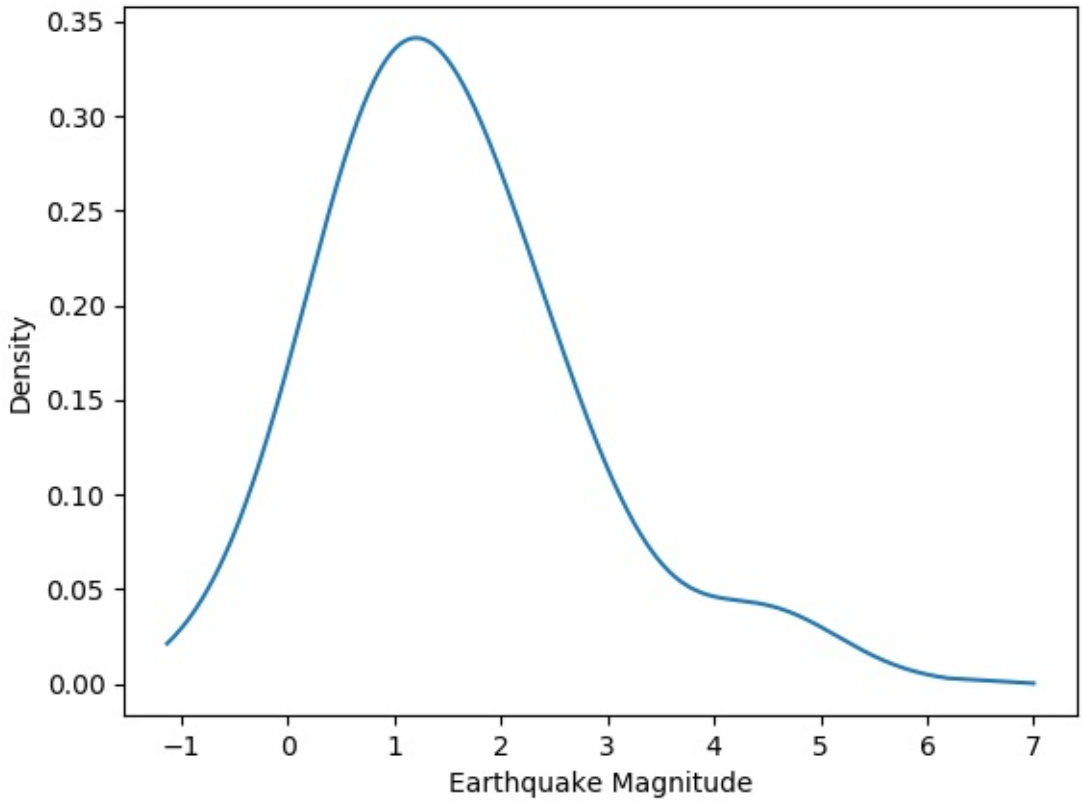


Figure 2. KDE Plot of Earthquake Magnitude.

Figure 2 shows the kernel density estimation for the earthquake magnitude by using the kernel type of Gaussian and the bandwidth of 0.5. Both Figure 1 and Figure 2 display the similar shape of the density distribution of the magnitude, and the highest density of the magnitude is located from 1 to 2. However, the KDE plot is smoother than the histogram, and would not lose the information for the original data.

Hint: The selection of kernel type and kernel width is important. In general, wider bandwidth reduces fine scale information, and narrower bandwidths preserve and accentuate fine scale information. The appropriate selection depends on several tries and the experience.

**(3) A scatter plot of latitude vs. longitude for all earthquakes**

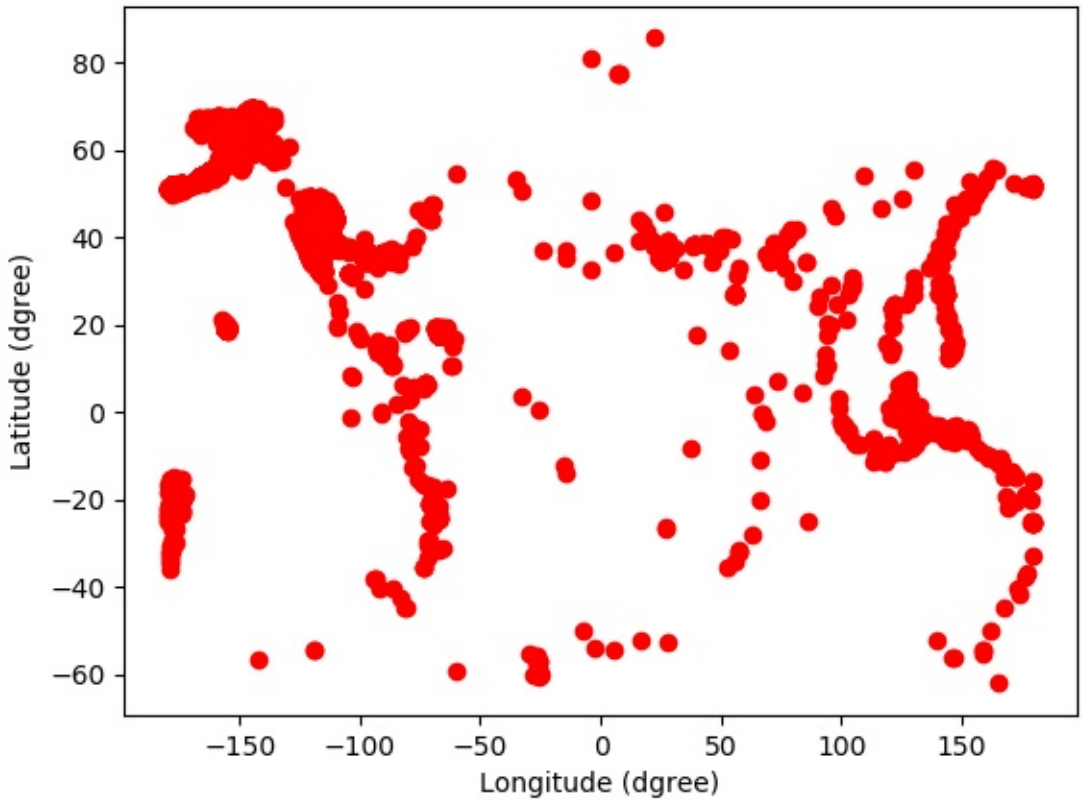


Figure 3. Scatter Plot of Latitude vs. Longitude for All Earthquakes.

Figure 3 shows the scatter distribution of the all the earthquakes happened over the past 30 days in the world. The high density of scatters shows that the Pacific Ring of Fire (120°E ~ 150°E and 90°W ~ 130°W) is the most geologically active region of Earth.

**(4) A normalized cumulative distribution plot of earthquake depths**

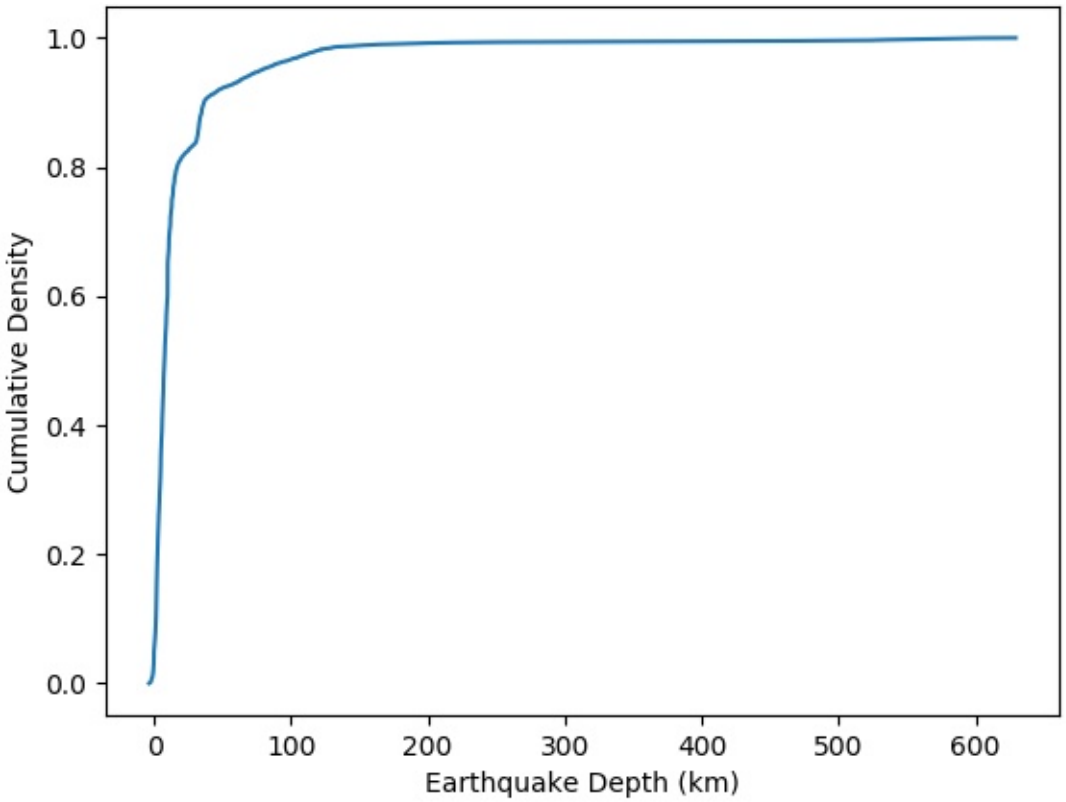


Figure 4. Normalized Cumulative Distribution Plot of Earthquake Depths.

Figure 4 shows that most earthquake (about 90%) over the past 30 days happened within the depth of 100 km. The steep slope of the curve within 100 km indicated that the depth interval is small and the depth distribution is dense. While the depth is deeper than 100 km, the number of earthquakes becomes less (less than 10%) and the depth interval becomes wider.

**(5) A scatter plot of earthquake magnitude vs. depth**

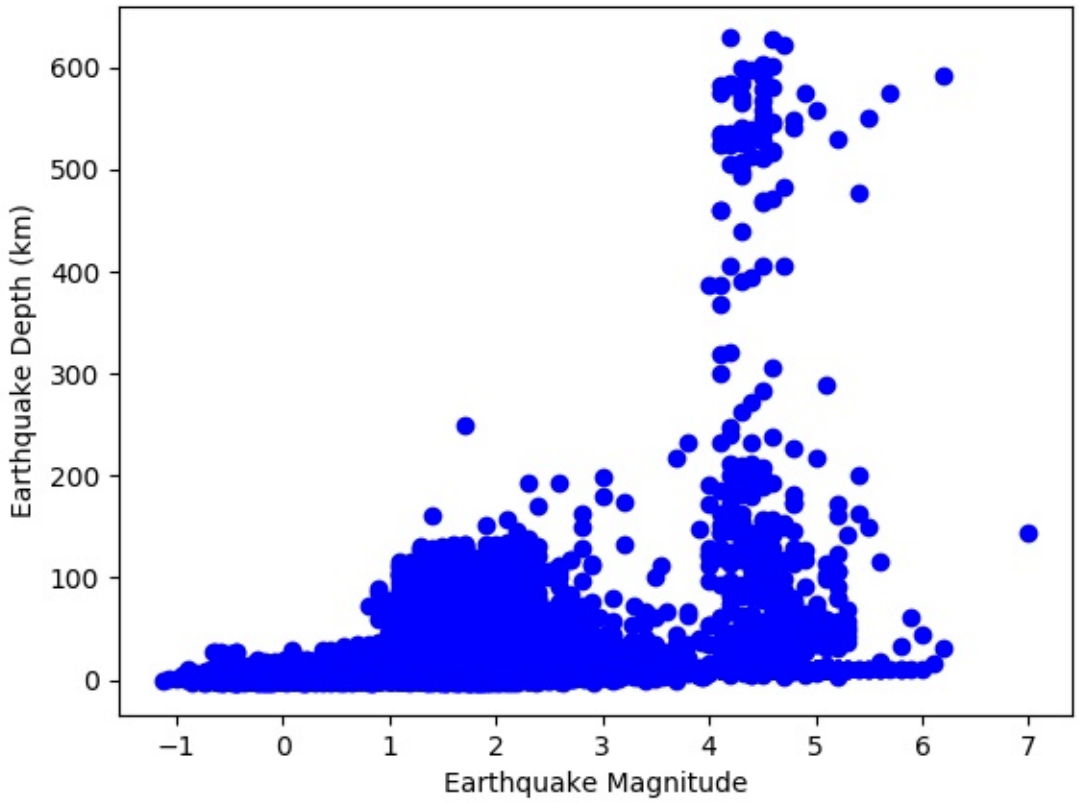


Figure 5. Scatter Plot of Earthquake Magnitude vs. Depth.

Figure 5 shows that relationship between the earthquake magnitude and the depth. The earthquakes with magnitude less than 4 happened mostly in the depth less than 150 km, and the depth range for the earthquakes with magnitude from 4 to 7 varies from depth from 0 to 630 km. The number of shallow earthquakes (depth < 100 km) is much more than that of deep earthquake (depth > 100 km).

**(6) A Normal Quantile-Quantile plot of the earthquake magnitude**

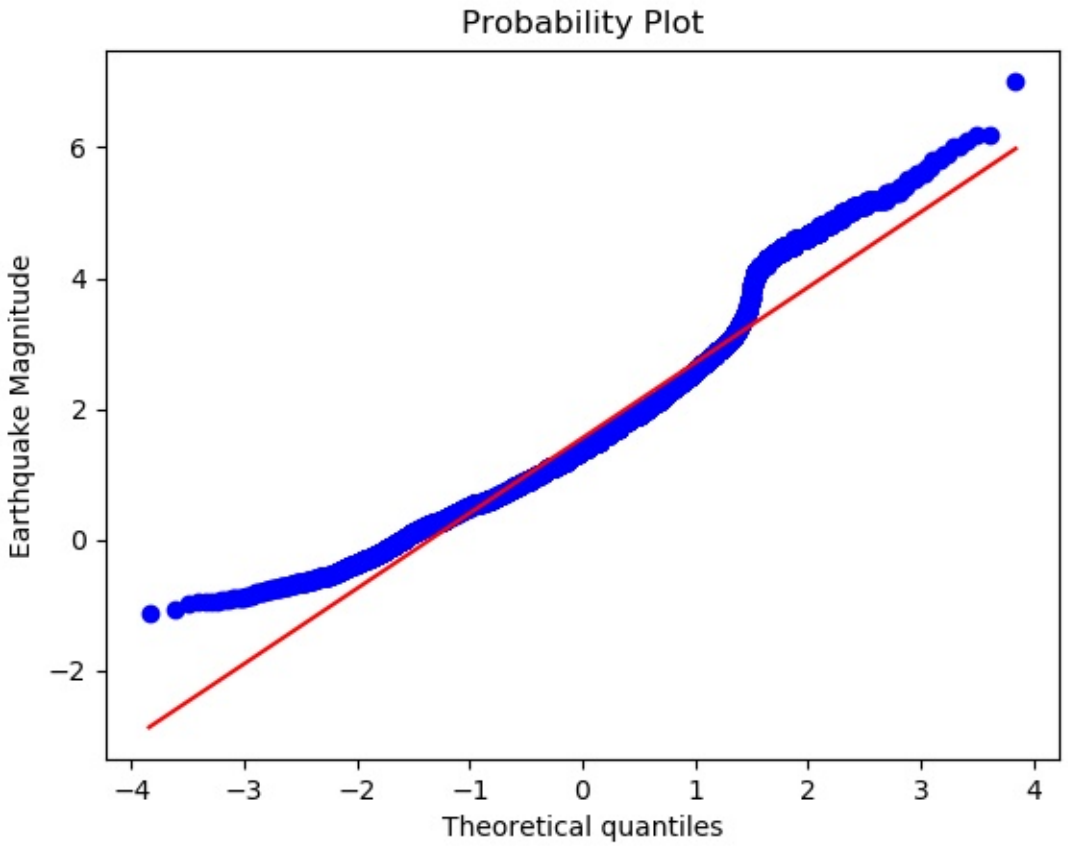


Figure 6. Normal Q-Q Plot of Earthquake Magnitude.

Figure 6 indicates the extent that the earthquake magnitude follows the normal distribution, and the red straight line is the theoretical normal distribution line. The figure shows that although there are some evident deviations at the extremes, the points from -1 to 1 in the theoretical quantiles are relatively close to the theoretical line, which means the distribution of the earthquake magnitude within this range could be approximated by the normal distribution.