
Metadata

Lab07: Graphical Analysis with python

Author: Alka Tiwari (tiwari13)

Github: <https://github.com/Environmental-Informatics/07-graphical-analysis-with-python-roccabye>

Source	Download at	Location
https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php	March 2, 2020 4:35 PM EST.	/home/tiwari13/ABE65100/07-graphical-analysis-with-python-roccabye/Input/Earthquake_30day.csv

Below are the fields included in the spreadsheet output:

time latitude longitude depth mag magType nst gap

dmin rms net id updated place type locationSource

magSource horizontalError depthError magError magNst status

It contains 11190 values (rows) and 22 variables (columns) enlisted above from February 01, 2020 starting 20:13:03.799Z to March 02, 2020 19:55:11.236Z.

The input file is in “csv” format and output figures are in “png” format.

Brief description of data visualization conducted in this exercise:

- (1) Histogram generation for magnitude of earthquake.
- (2) Kernel Density Estimates (KDE) plot for the earthquake magnitude.
- (3) Spatial Distribution of earthquake occurrence.
- (4) Normalized cumulative distribution plot for the earthquake depths.
- (5) Scatter plot of earthquake depth vs magnitude.
- (6) Quantile (Q-Q) plot of earthquake magnitude assuming normal distribution.

1. The numpyfunction genfromtxt() will not work with this data file (.csv). Why not?

Because the input file is in ".csv"(comma separated variables) format. Best option is to use "pandas" module and use "read_table()" or read_csv() and read_fwf() command for comma separated variables and fixed width files, respectively.

helpful link – (<http://pandas.pydata.org/pandas-docs/stable/10min.html>)

Data Visualization

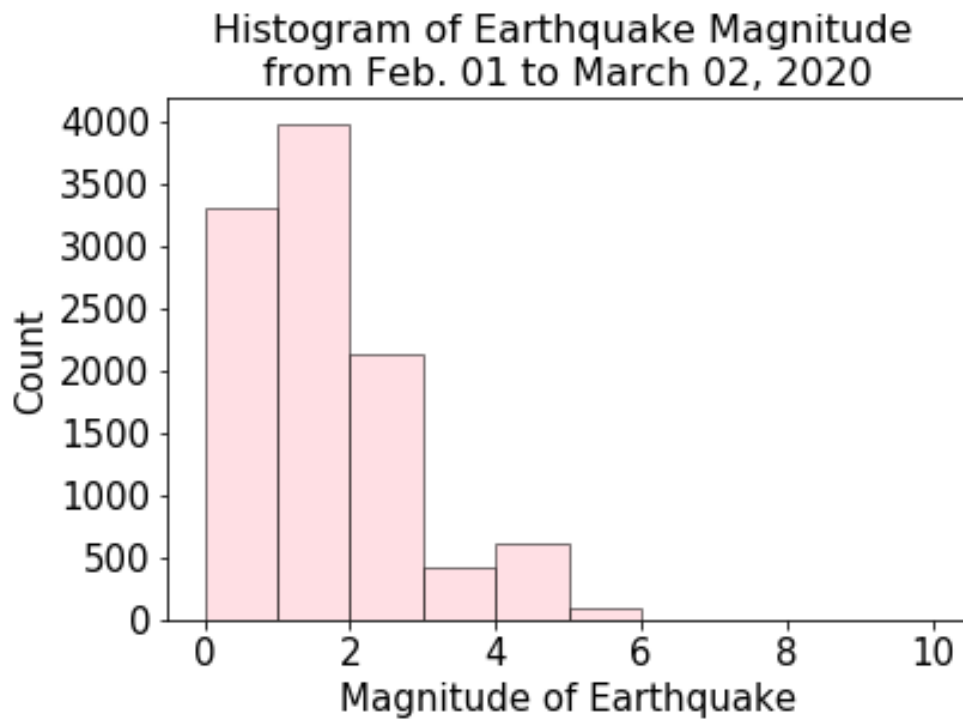


Figure 1: Histogram of earthquake magnitude, using a bin width of 1 and a range of 0 to 10. The bin range is chosen between 0 to 10, so it encompasses all the positive values of the magnitude. The bin size affects the accuracy of the distribution of data, smaller the bin size more accurate the distribution is. The histogram shows that lower magnitude i.e. magnitude between 0 to 3 has more frequency while higher magnitudes between 4 to 5 has lesser frequency.

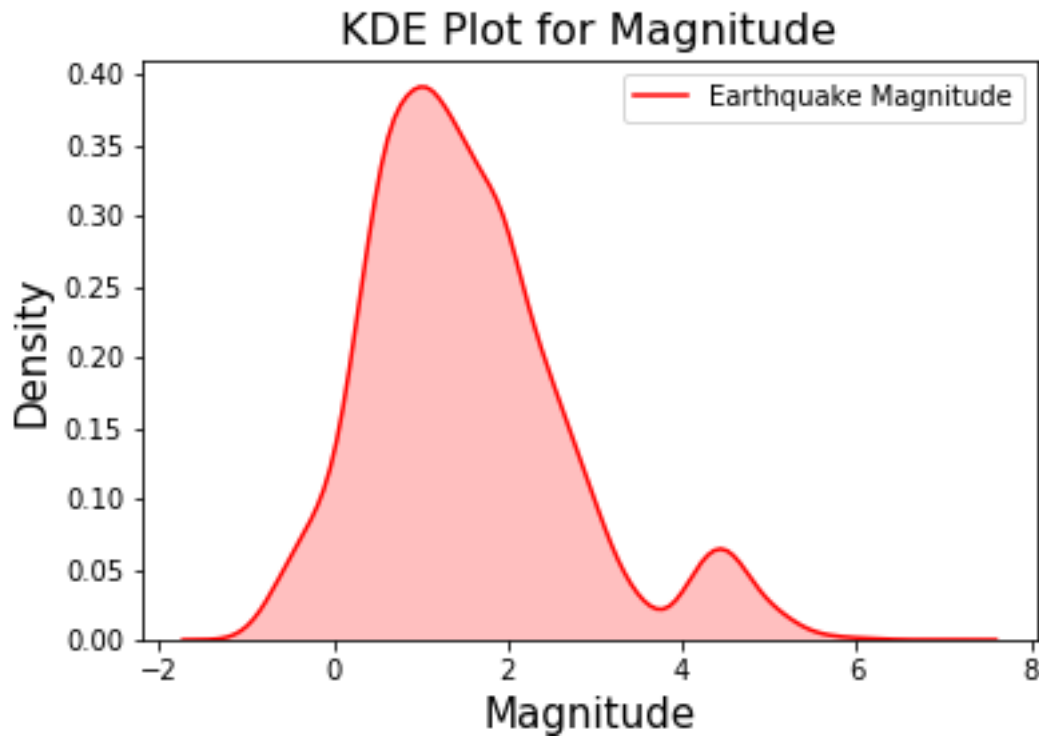


Figure 2: Kernel Density (KDE) Plot of earthquake magnitude. The selected kernel type is Gaussian and kernel width of 0.2. The KDE plot gives better visualization of the data. The plot shows that there is highest density of data near magnitude of 1 while there is one more peak near magnitude 5, i.e. higher magnitudes have lesser density which is similar to histogram analysis that the higher magnitudes have lesser frequency.

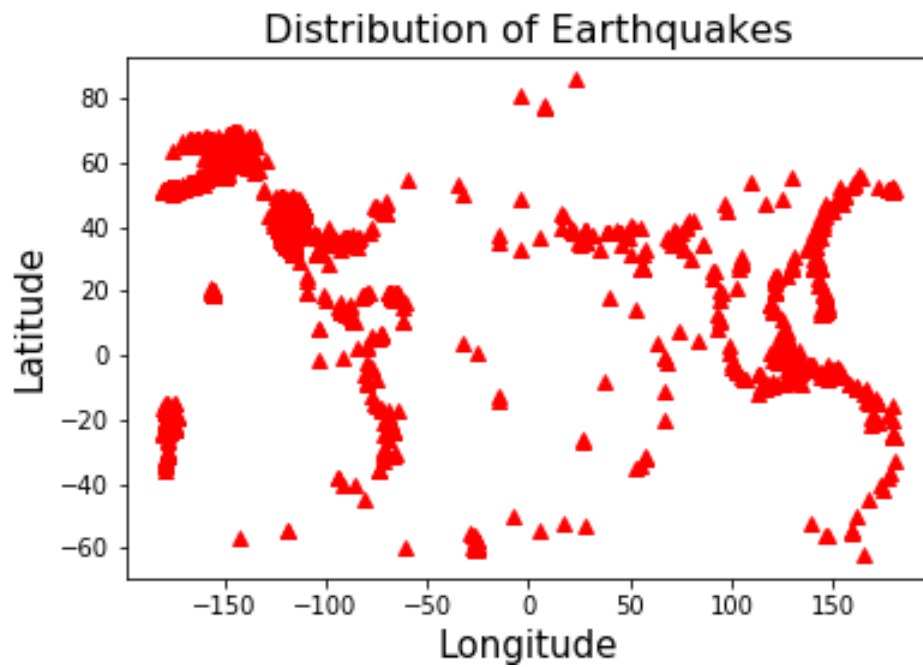


Figure 3: Earthquake occurrence over the globe with longitude on x-axis as it varies horizontally and latitude on y-axis as it varies vertically in the global map. It is evident that there are fewer earthquakes within the latitude range between -50 deg. to 100 deg. Maximum occurrences are in the region with latitude <-50 and >100 deg.

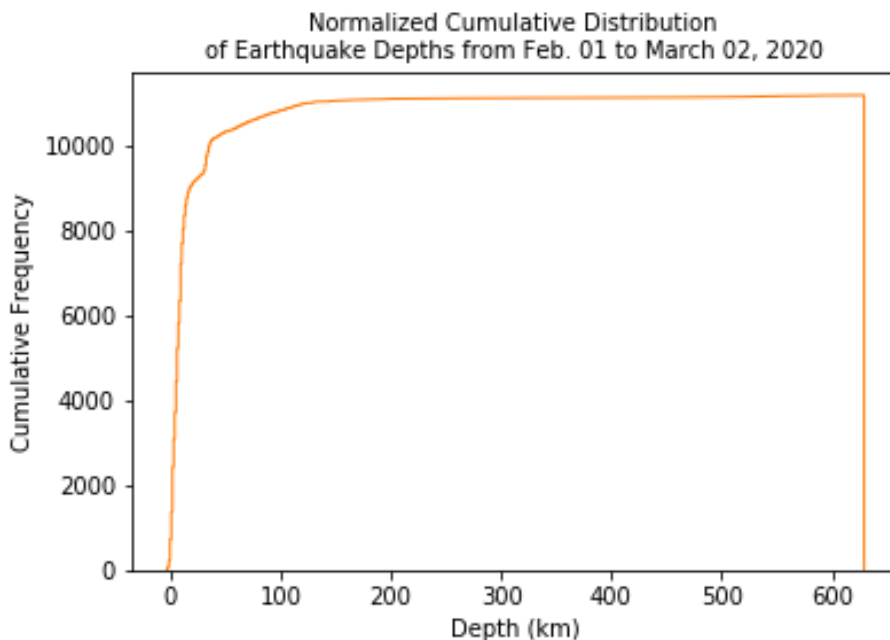


Figure 4: Cumulative distribution plot for earthquake depth. For depth between 0 to 150 km the slope is positive while after depth of 150 km the distribution is flat with almost zero slope, with bin size of 600.

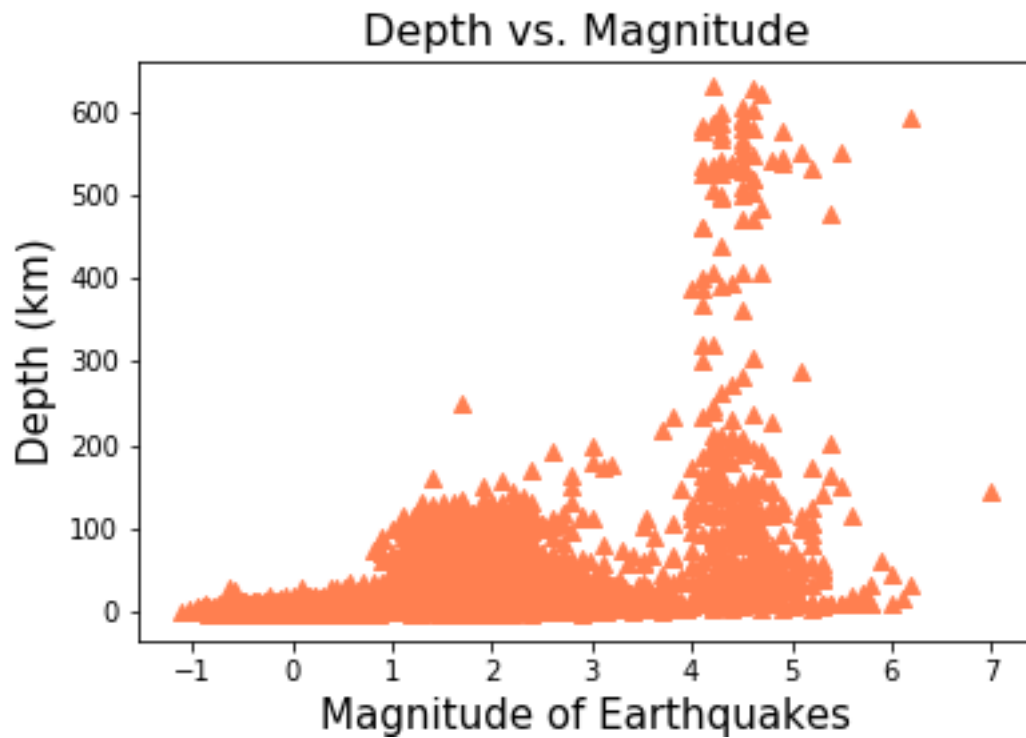


Figure 5: Scatter plot of earthquake magnitude with depth (km). The plot shows that the magnitude of 4 to 5 has depth ranging from 0 to 600 km i.e. the probability of magnitude 4 to 5 to occur with all the depths is possible. While the lower magnitude has higher probability to occur at a depth lower than 200 km. Because of the cluster of the data a suitable relationship cannot be strictly exhibited but on average it shows a linear relation.

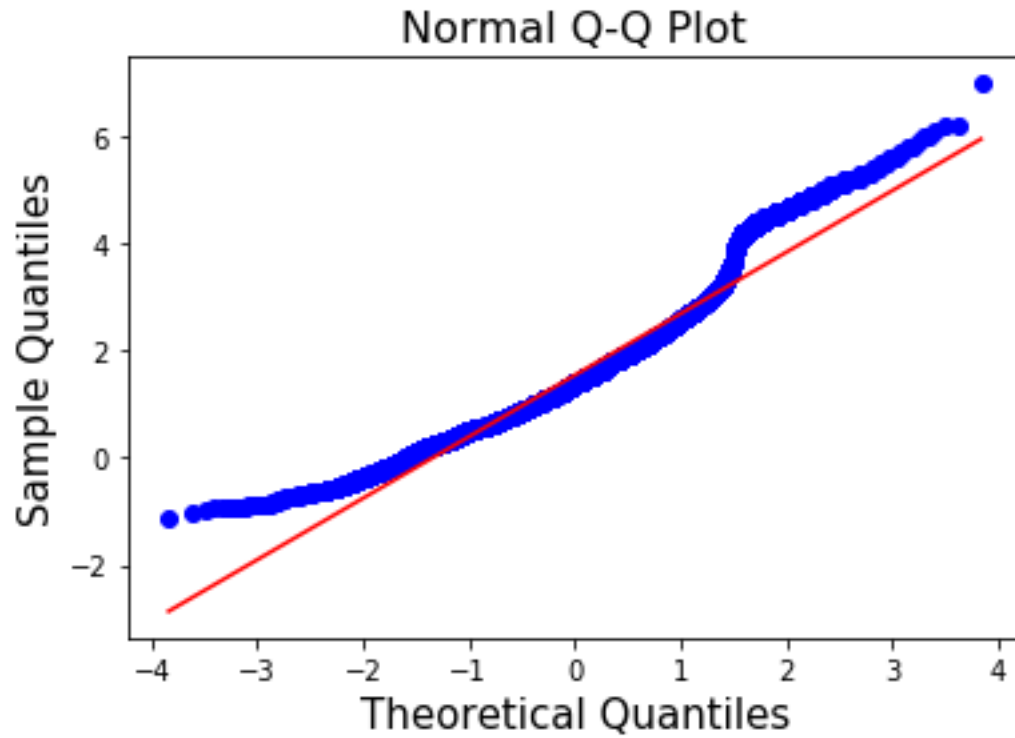


Figure 6: Q-Q plot of earthquake magnitude considering normal distribution. The red line shows the points of normal distribution, but the magnitudes do not fall around the red line hence assuming normal distribution is not fair overall.