Karoll Quijano - Kquijano
ABE 651 - Environmental Informatics
Assignment 07 – Graphical Analysis with Python

Metadata

Source: https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php

Download date and time: 03/08/2020 4:20 pm. The dataset includes all events recorded for the previous 30 days.

The dataset 'all_month.csv' includes the following information:

- Time: UTC time when the event occurred and reported in milliseconds
- Latitude, longitude: Decimal degrees where negative values correspond for southern latitudes and western longitudes.
- Depth, magnitude: Depth of the event in kilometers and magnitude.
- The dataset also includes: total number of seismic stations used to determine earthquake location (nst); the largest azimuthal gap between azimuthally adjacent stations (gap); horizontal distance from the epicenter to the nearest station (dmin); the root-mean-square (RMS) travel time residual, in sec, using all weightst; the ID of a data contributor(id); type of seismic event (type); horizontal location error, in km (horizontalError); the depth error, in km (depthError); uncertainty of reported magnitude of the event (magError); total number of seismic stations used to calculate the magnitude for this earthquake (magNst); status is either automatic or reviewed (status); The network that originally authored the reported location of this event (locationSource), and Network that originally authored the reported magnitude for this event (magSource)

Script program-07.py:

- The script uses 'all_month.csv' file as input and uses the variables latitude, longitude, depth, and magnitude of all earthquakes to perform a data analysis
- First, the CSV file is imported as a data frame using Pandas, followed by the removal of rows with empty values. The tool used is read_csv() instead of genfromtxt() due to the different data types in the CSV.
- The following graphs are shown as results: histogram of earthquake magnitude (Figure 1.), KDE plot of earthquake magnitude (Figure 2.), Plot latitude versus longitude for all earthquake (Figure 3.), normalized cumulative distribution plot of earthquake depths (Figure 4.), scatter plot of earthquake magnitude Vs. depth (Figure5.), and Q-Q plot of the earthquake magnitudes (Figure 6.).

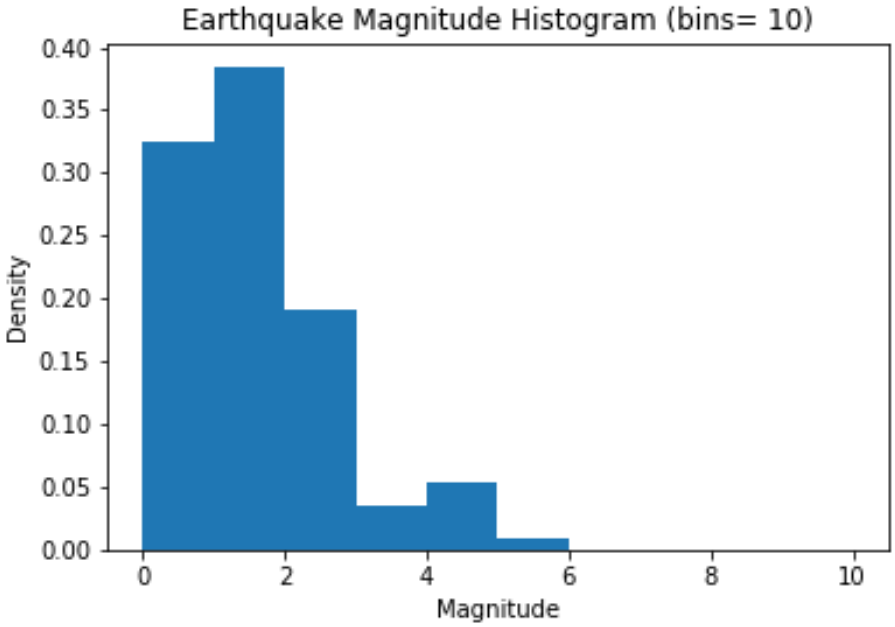

Figure 1. Histogram of earthquake magnitude (bin= 1, range 0-10). The magnitude with the highest density is 1-2, followed by 0-1, and 2-3. The histogram suggests a left-skewed, bimodal distribution of the data. The selection of the bin size and range affect the histogram distribution to show.

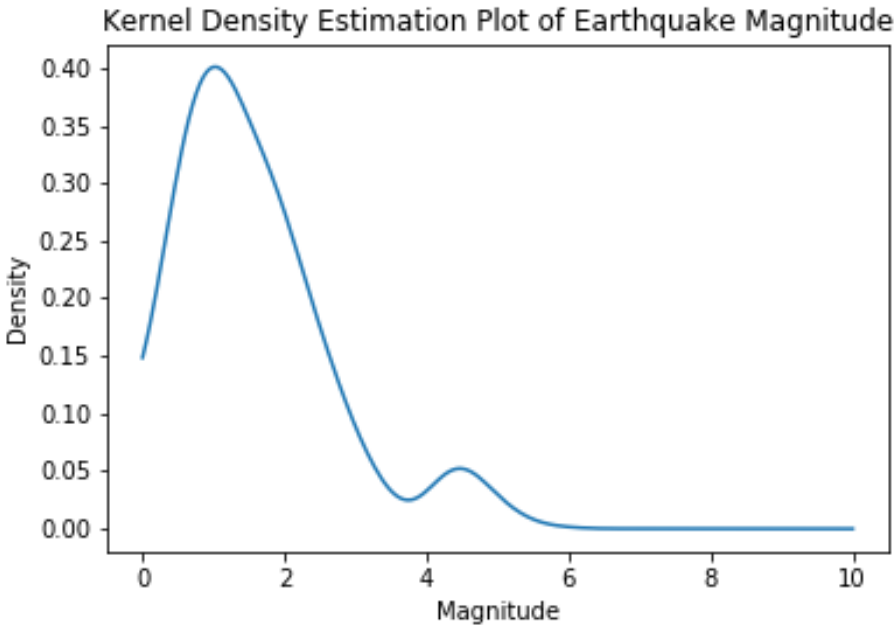

Figure 2. Kernel Density Estimation (KDE) plot of earthquake magnitude. KDE shows a smooth line estimating the probability density distribution of the earthquake magnitudes. KDE was built with a Gaussian kernel and kernel width of 0.25. As with the histogram, the bandwidth can affects the shape of the curve and could hide the bimodal distribution by smoothening too much the curve.

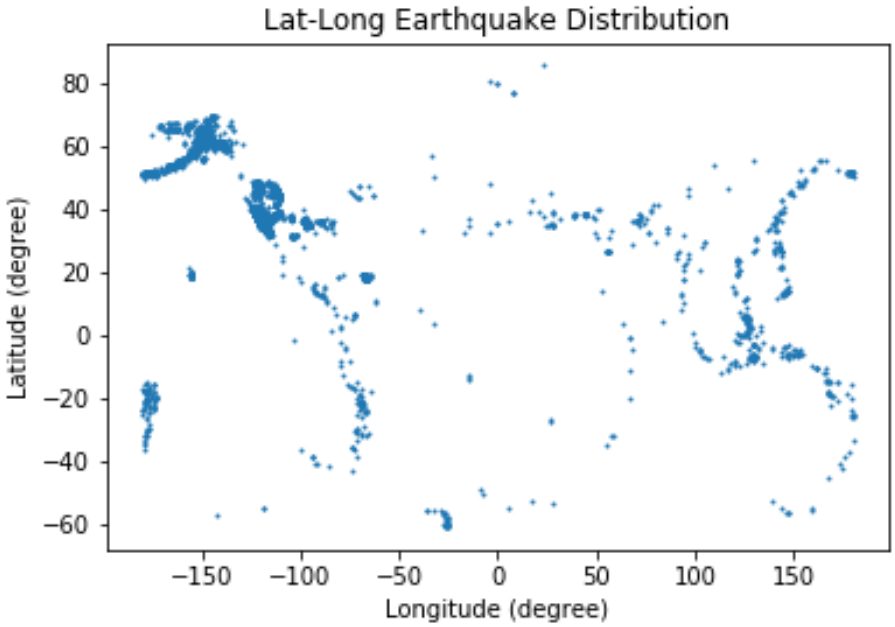

Figure 3: Plot latitude versus longitude for all earthquakes. This scatters plot of the distribution of events reported, shows a higher concentration of events on the west side of the American continent. The longitude is set on the x-axis and latitude on the y-axis representing the geographic coordinate system.

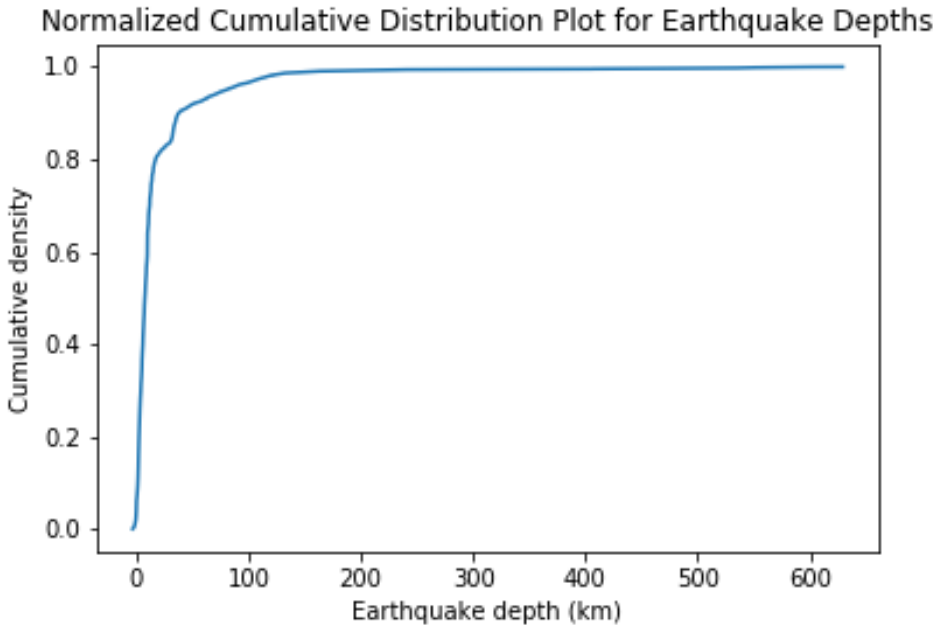

Figure 4: Normalized cumulative distribution plot of earthquake depths. About 95% of the earthquakes in the past 30 days occurred with a depth less than 150 km, and the 5% left reported depths with a range of 150-600 km.

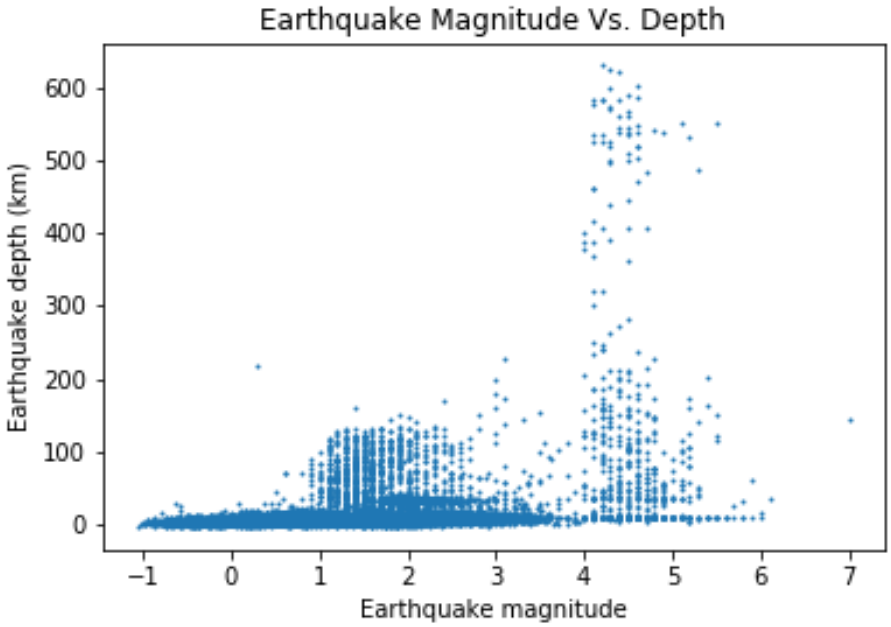

Figure 5: Scatter plot of earthquake magnitude Vs. depth. From this plot is appropriate to say that earthquakes with magnitudes less than 4, usually present depths not larger than 150 km. also, earthquakes with less than 50 km depth have more occurrence than on a magnitude range of -1 – 3.5.

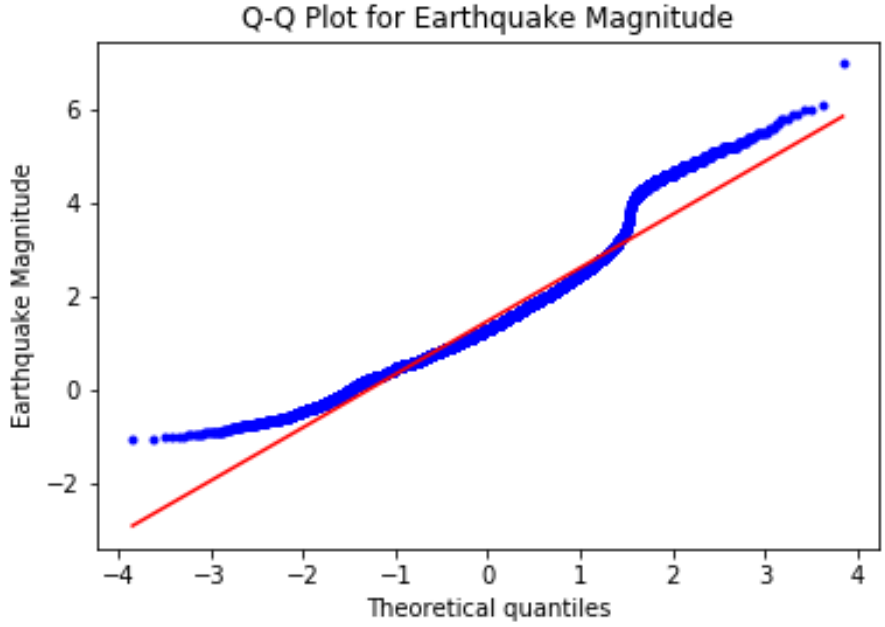

Figure 6: Q-Q plot of the earthquake magnitudes. This plot confirms that the data does not follow a normal distribution, as it separates from the red line (that simulates the ideal normal distribution) at the beginning and the end of the data.