Metadata

Ankit Ghanghas
ABE 65100 Sp2020

Data Used for Analysis : USGS All Earthquakes for past 30 days

Basic Overview of Data: The dataset used for this study is a collection of location, magnitude, depth and other properties of all the earthquakes that happened across the globe in the past 30 days.

Source: https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/all_month.csv
Date of Access: 28th February 2020 16:38:00

Type of Data: The data is stored in a csv format and contains the following information as columns

-time, latitude, longitude, depth, mag, magType, nst, gap, dmin, rms, net, id, updated, place, type, locationSource, magSource, horizontalError, depthError, magError, magNst, status

Type of analysis:
The main outcome of the task is visualization of the data stored to get a better understanding of the data used. The different outcomes of the script are as follow
a) Histogram generation of magnitude of earthquakes.
b) Generating Kernel Density Estimates (KDE) plot for earthquake magnitude to get a better insight into its distribution.
c) A plot of earthquake across the latitude and longitude to assess the spatial distribution of earthquakes
d) Generating a Normalized Cumulative distribution plot of earthquake depths.
e) A scatter plot between earthquake magnitude and depth.
f) Q-Q plots of earthquake magnitude assuming Normal and Gumbel Distribution.

The numpy function genfromtxt() does not work with this dataset because the different columns have different number of values, possibly because of missing values and genfromtxt() requires all columns to have same number of values.

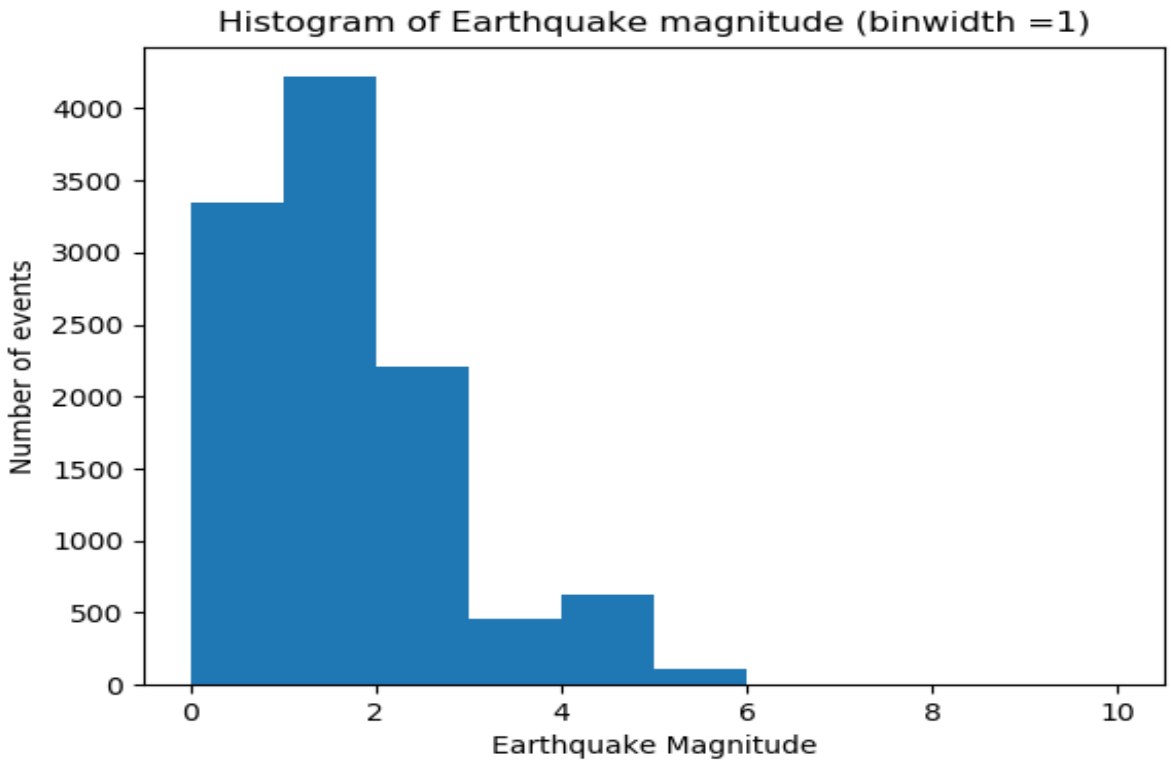

Fig.1 Histogram of Earthquake Magnitude using bin width of 1 and range of 0 to 10.

The bin width impacts the ability of the histogram to show regions of higher number. Too low bin width leads to poor group formation while too large bin width will average things and hide local regions of higher occurrence. The histogram tell us that most earthquakes had a magnitude between 0-3 on Richter scale, with earthquakes with magnitude 1-2 were most frequent.

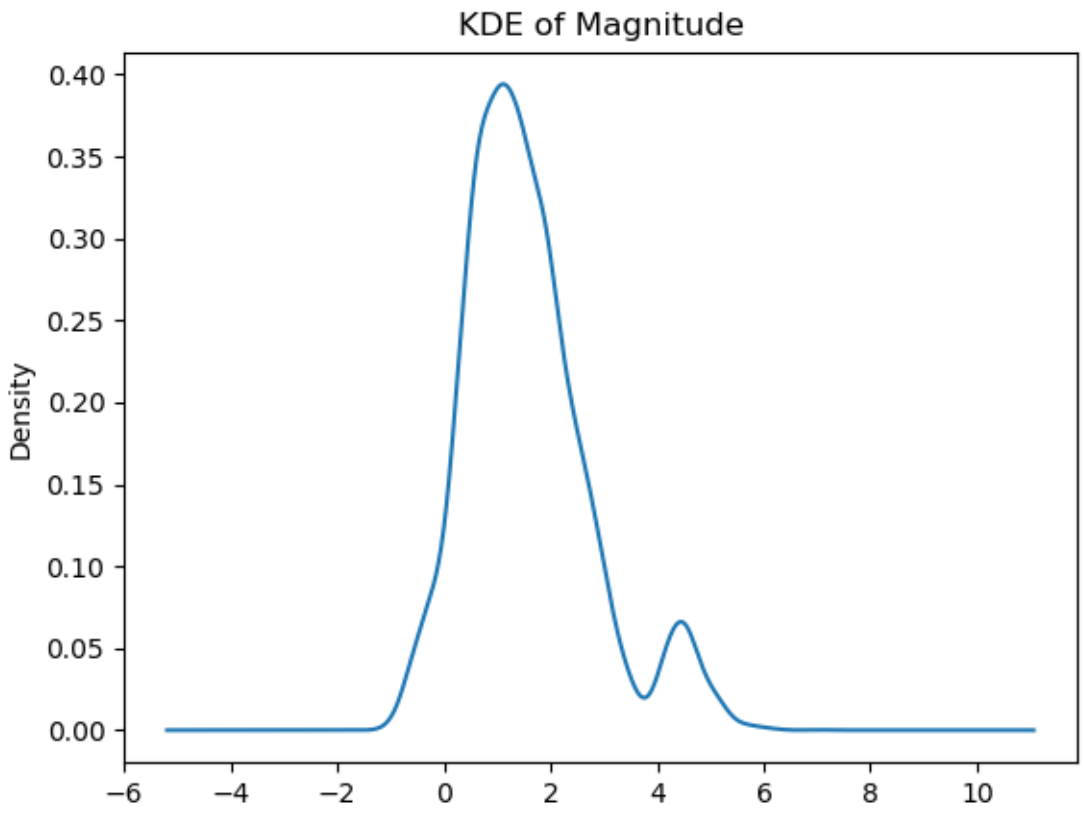

Fig 2. Kernel Density Estimate Plot of Earthquake Magnitudes. This KDE uses Gaussian kernel type and a kernel width of 0.2. The KDE plot is similar to the histogram and reflects the information given by the histogram that majority of earthquakes have a magnitude between 0-2. KDE is smother than histogram and information is not lost in the process of binning as is the case with histogram.

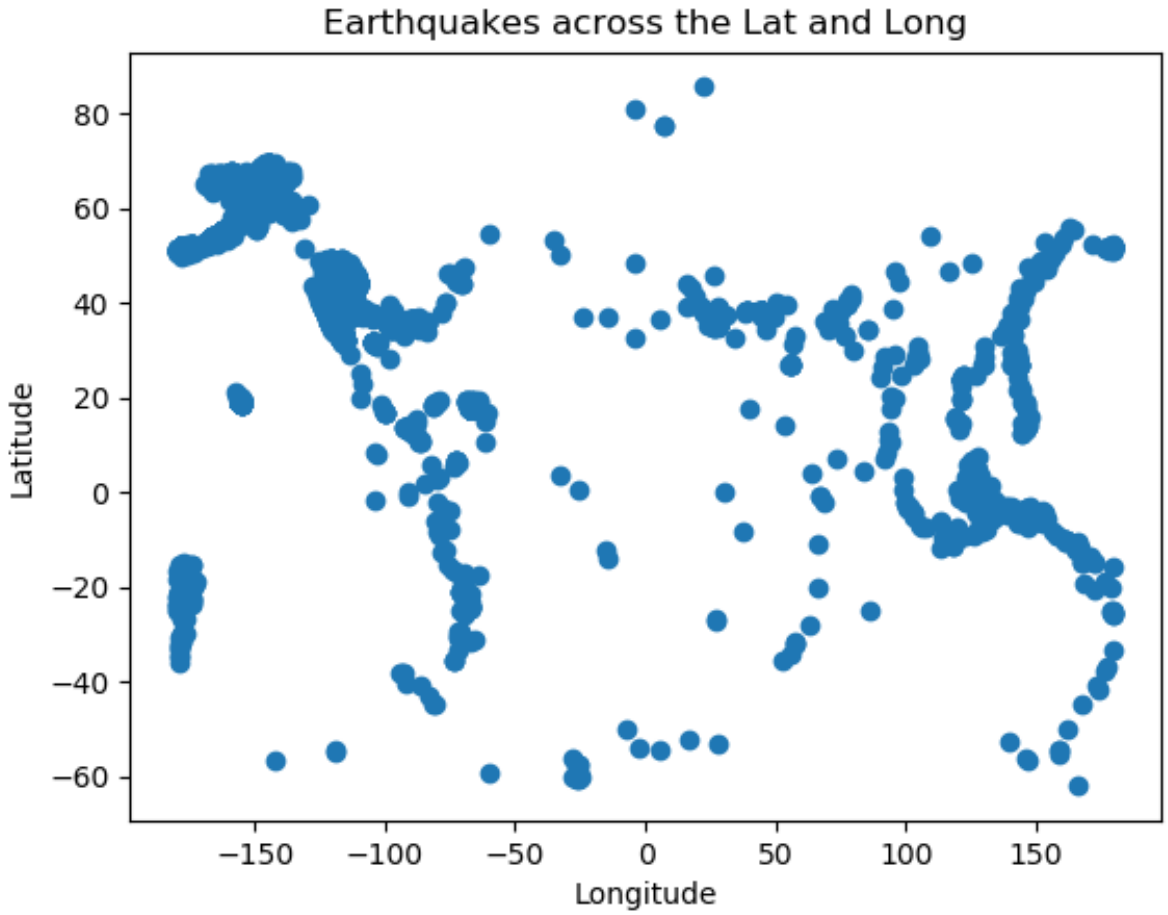

Fig3. Spatial Variation of Earthquakes across the globe. Longitude on x-axis and latitude on y-axis because longitudes are defined as lines running parallel to vertical plane(0 degree meridian) on earth while the latitudes are defines as the lines running parallel to the equator (horizontal plane). The give plot shows that most of the earthquakes occur along the tectonic plate boundaries.

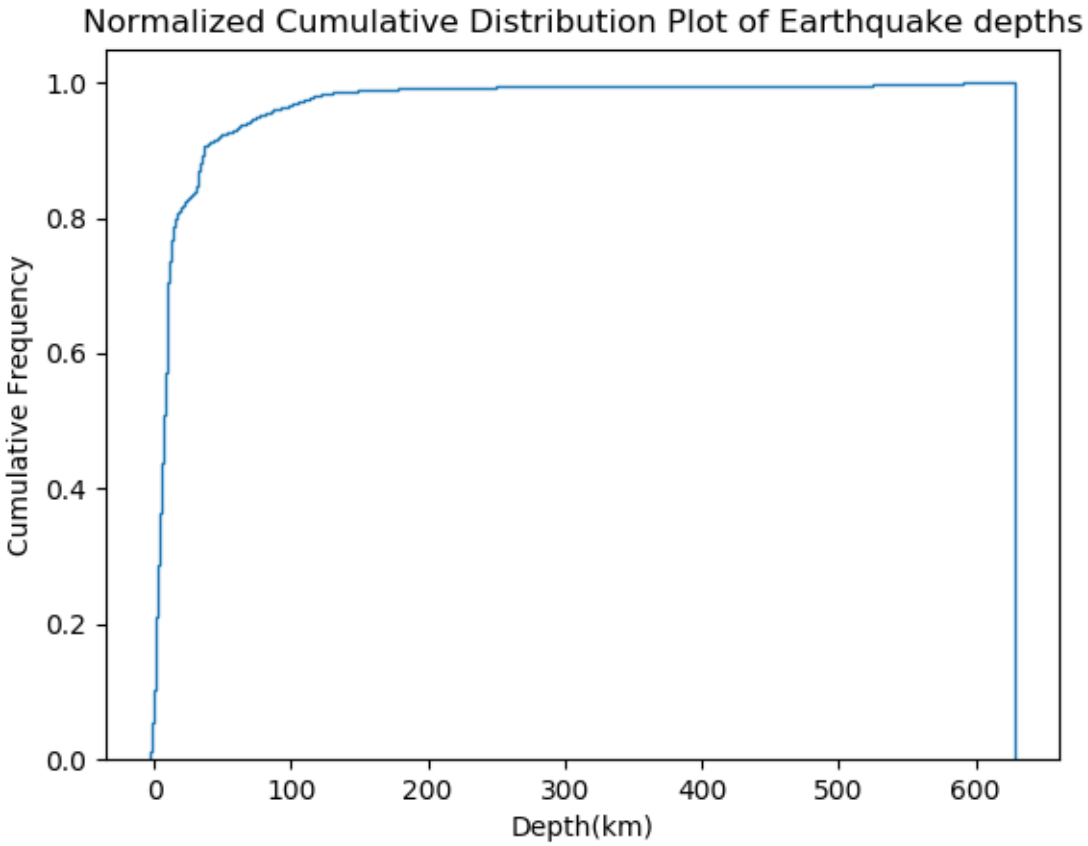

Fig 4. This plot shows a normalized CDF of earthquake depths. This plot indicates that nearly 97% of the earthquakes happened within 100 km depth from the surface.

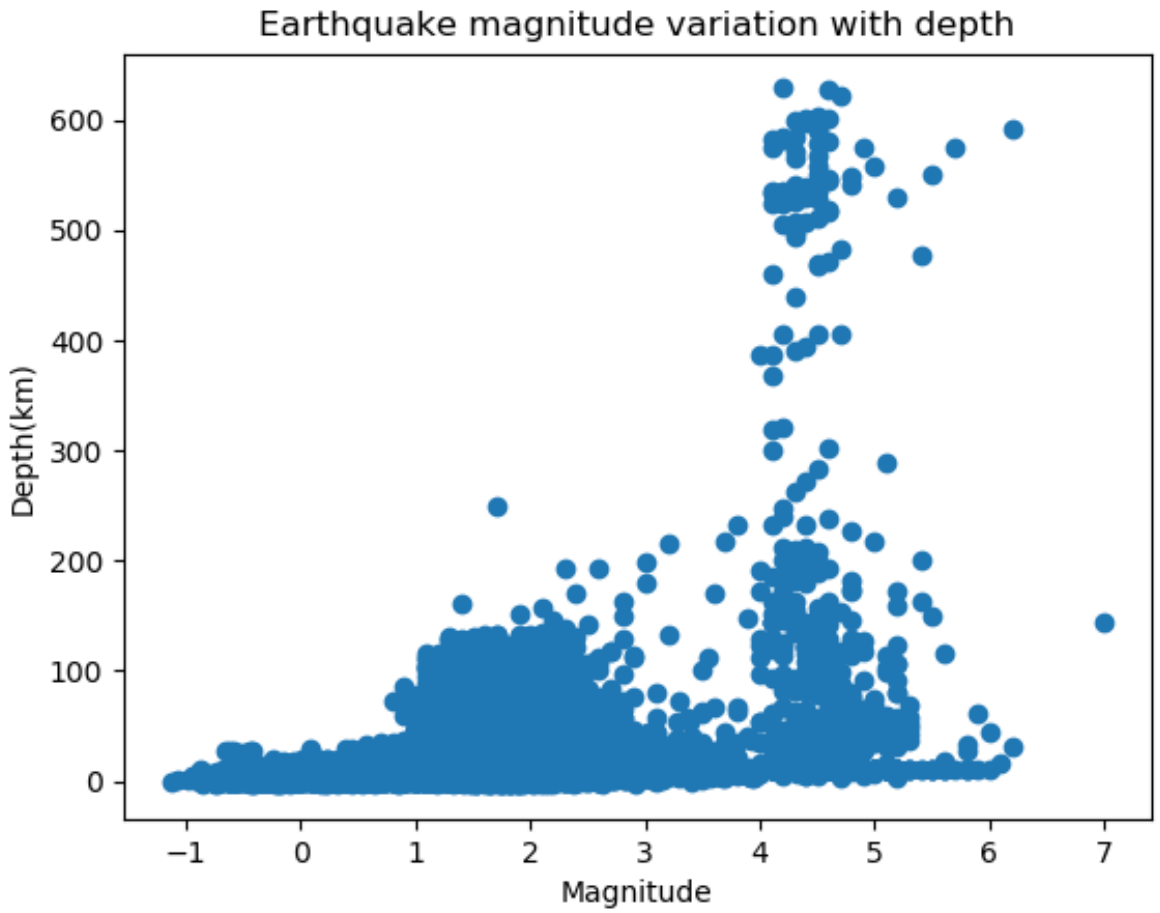

Fig 5. Scatter Plot of Earthquake Magnitude against Depth. This plot indicates that smaller magnitude earthquakes happen at a shallow depth, while higher magnitude earthquakes are likely to happen at even higher depths, even though they generally occur at lower depths.

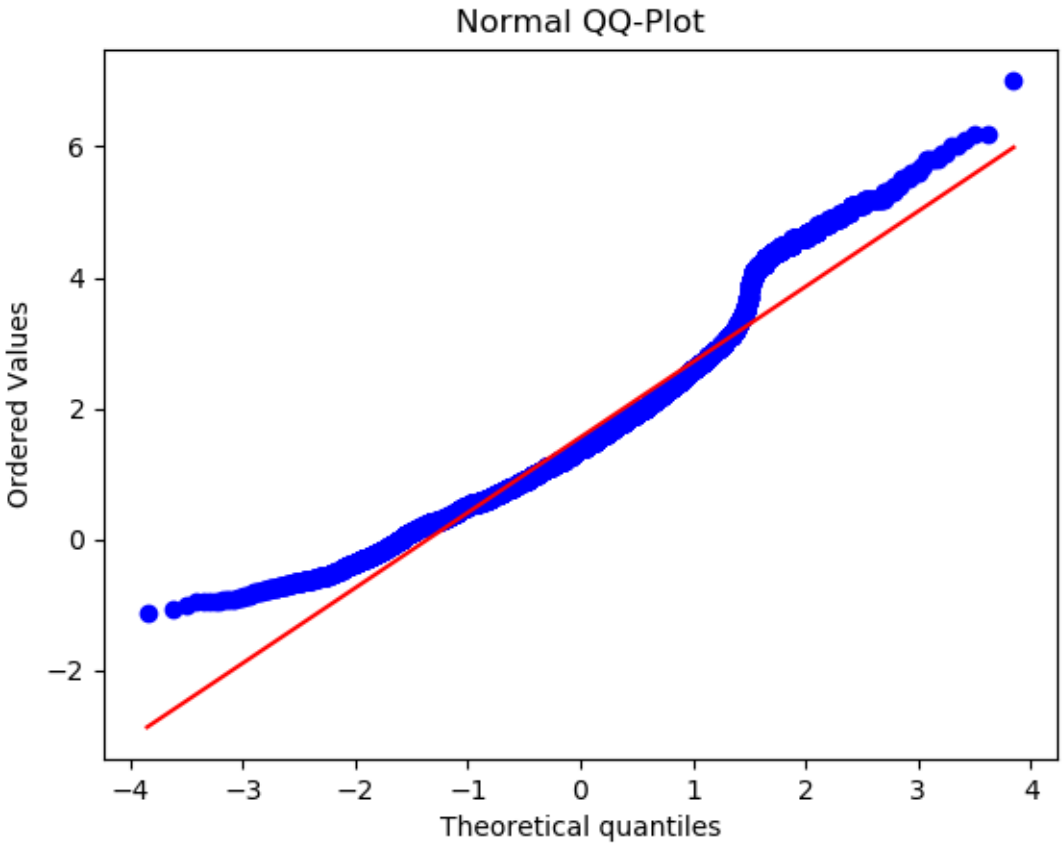

Fig 6. This is a qq-plot of the earthquake magnitude assuming it follows a normal distribution. As one can infer from the plot the since a large portion of blue dots do not align with the red line so the earthquake magnitude does not follow a normal distribution.

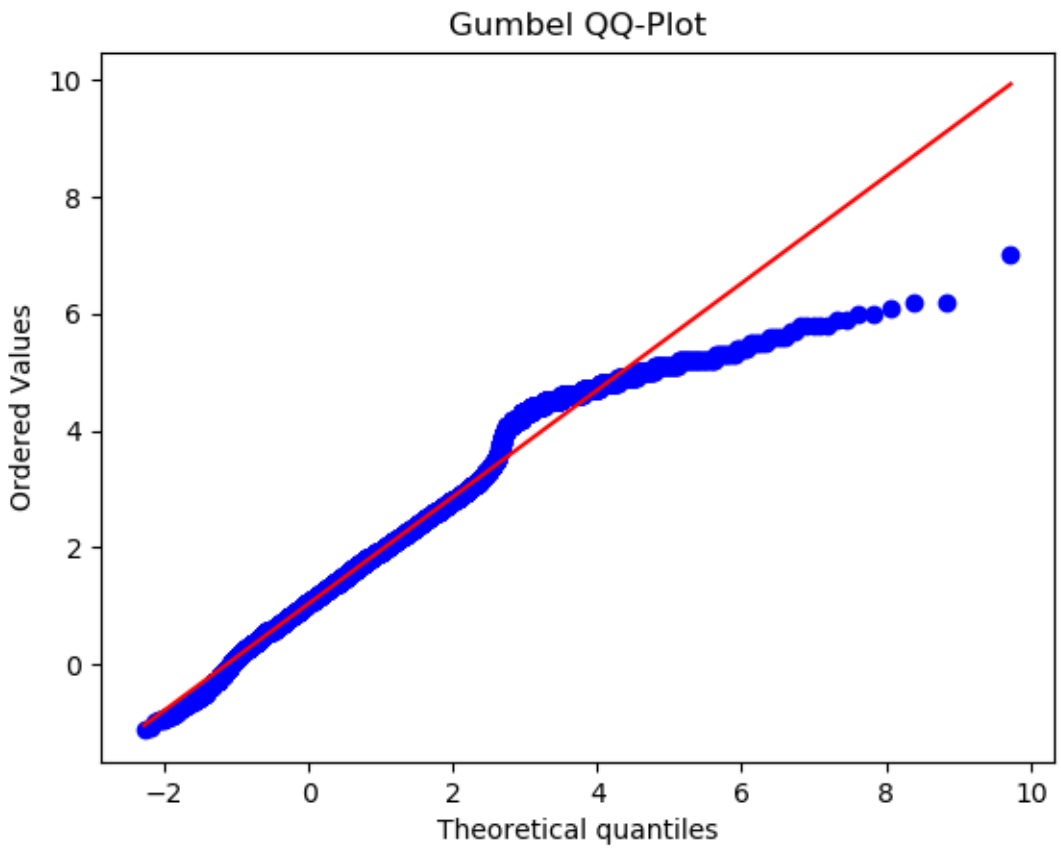

Fig 7. This is a qq-plot of earthquake magnitude assuming Gumbel distribution. This is a better fit for the data than normal distribution.