Metadata

1. Data Source

This data was downloaded from USGS earthquake hazards web site as ‘csv’ format by creator -Tianle Xu. This file named ‘all_month.csv’ was created at 3/3/2020.

The data contains some information about earthquakes happened 30 days past till the day when it was downloaded from the website. It provides when, where, the earthquakes occurred and how strong they are, the depth of where they are from and so on. However, in this analysis, only a few columns will be useful.

2. Analysis process

Pandas’ read_table() commend was used to read csv file instead of genfromtext() commend due to the different data types inside the file.

a. A histogram of earthquake magnitude would be generated to study the happening probability of earthquakes.

b. A KDE plot for the same data would be made to study the kernel density estimation.

c. Plotting latitude versus longitude for all earthquakes to find out the law of earthquake occurrence

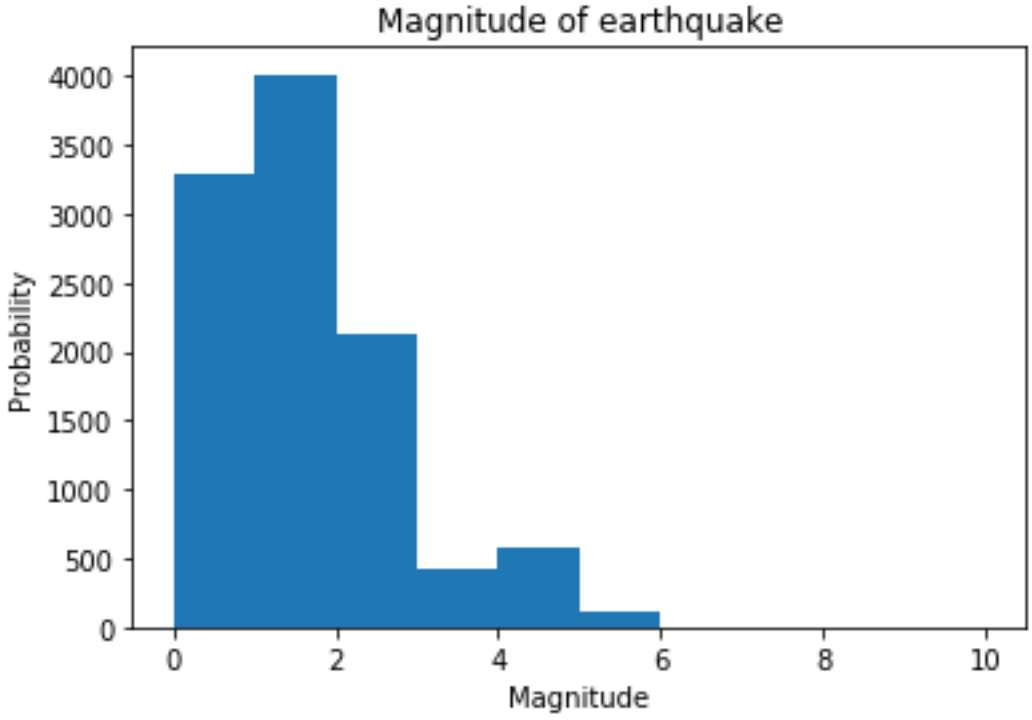
d. Generating a normalized cumulative distribution plot of earthquake depths to analyze its cumulative distribution

e. A scatter plot of earthquake magnitude with depth would be plot to analyze the relationship of magnitude and depth of earthquakes.

f. Q-Q plot of the earthquake magnitudes

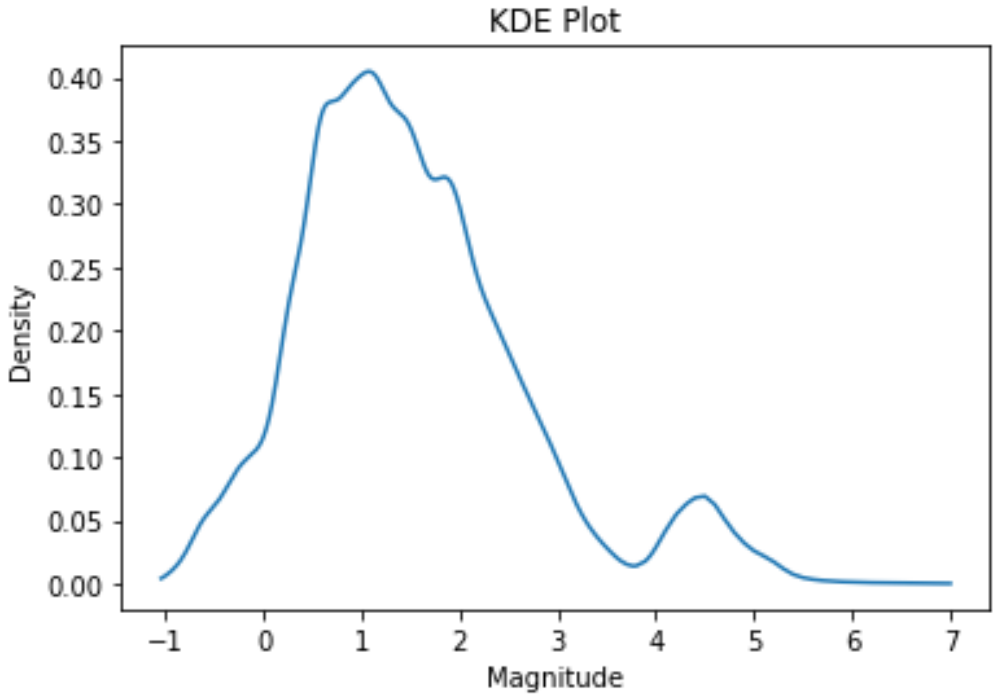
3. Graphical Analysis

a.



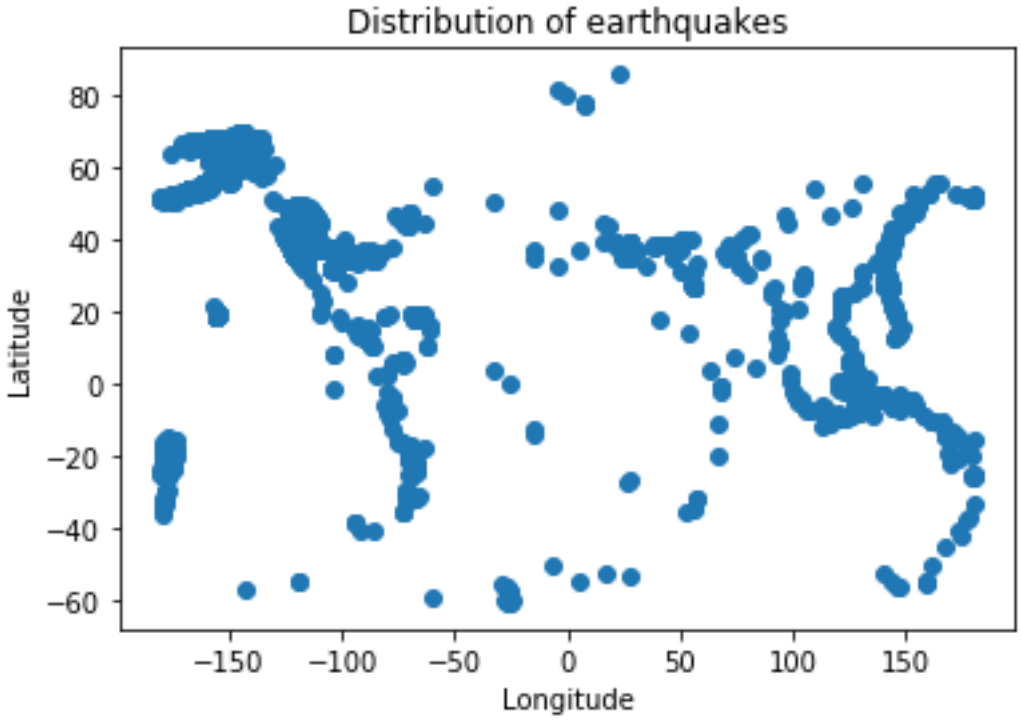
The width of the bins is 1 and the range is 1 to 10. When the width decreases, the plot would be more smooth and accurate and with larger range, the plot would contain more magnitude data. The histogram suggests that earthquakes with bigger magnitude is less possible to appear compared with smaller earthquakes.

b.



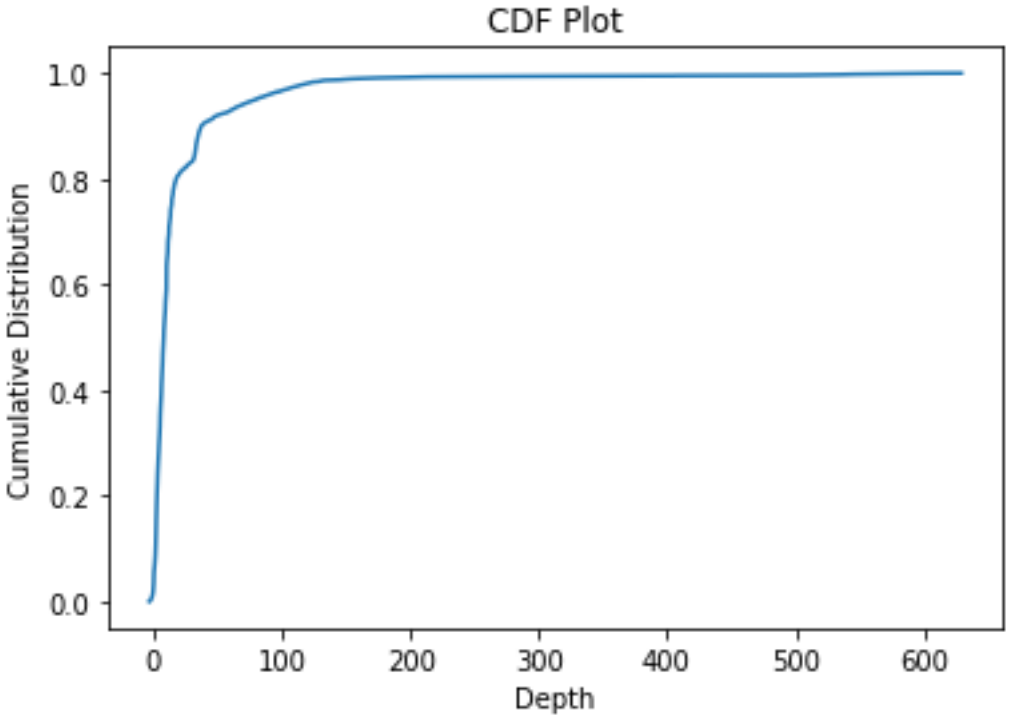
The kernel type is Gaussian and width is 0.1. KDE and histogram both can give people a direct distribution of the probability while the KDE plot is more smooth, which means every sample responses to a certain probability, but in histogram only a coarse value would be found.

c.



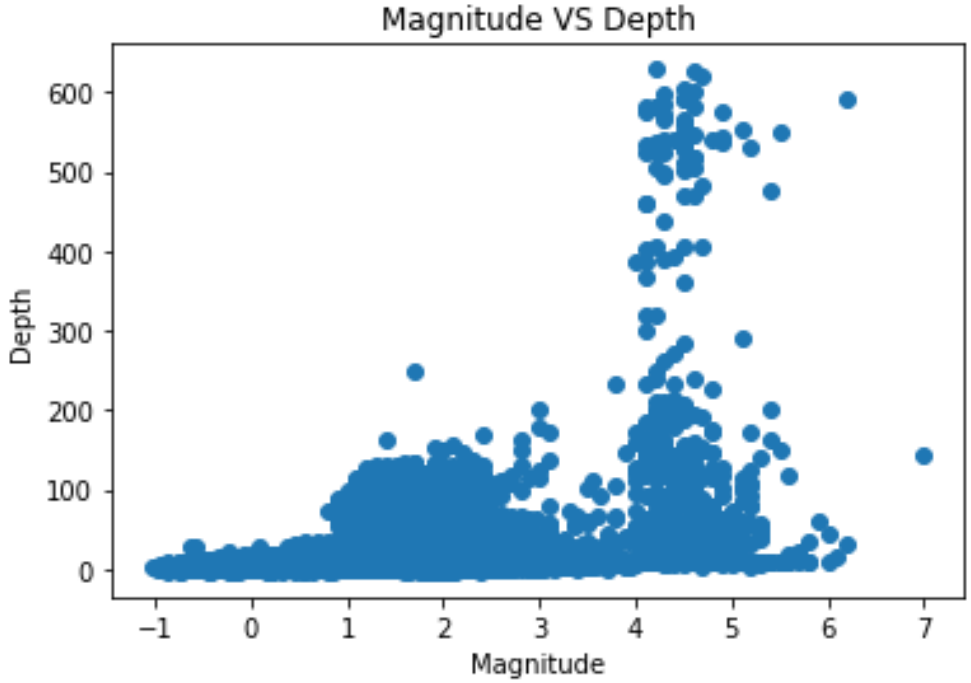
The reason why longitude is x-axis and latitude is y-axis is because it is what all maps' coordinate systems are. From the plot of longitude and latitude, it is obvious that most points located at the intersection of continental plates.

d.



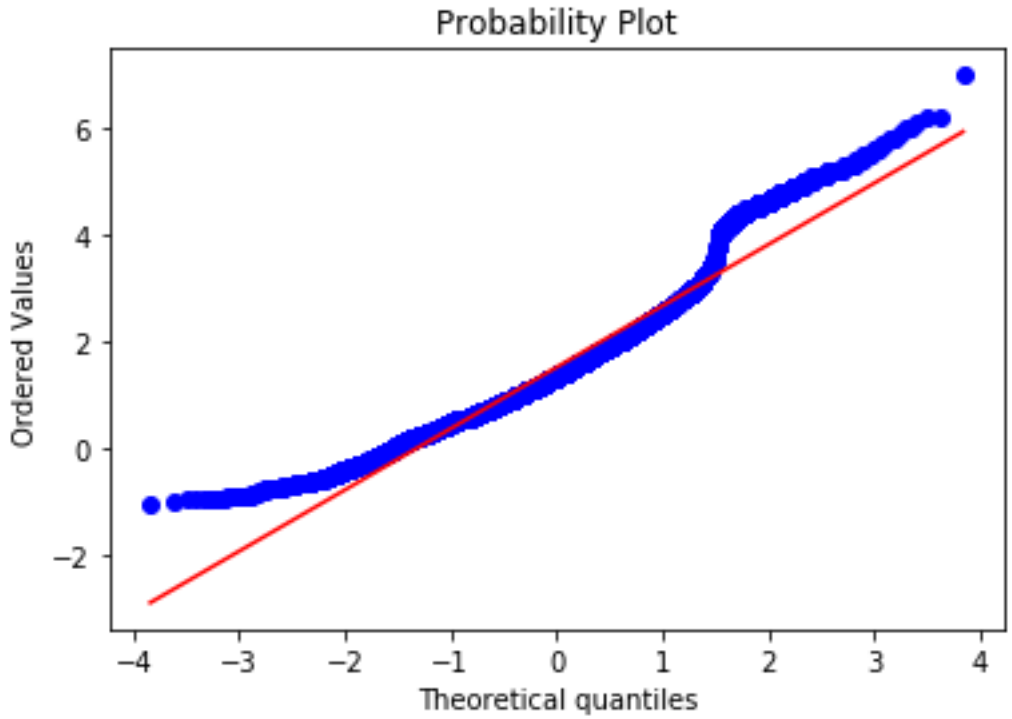
Only a few earthquakes' depth can be bigger than 200. Most of their depth is concentrated in 0 to 50. And from the slope of the distribute line, it turns out that most earthquakes has a depth with 0 to 50 since the slope from 0 to 30 increases quickly.

e.



With smaller magnitude, less earthquakes have deep depth while when the magnitude is close to 4 and bigger , the depth of earthquakes is much deeper. It reveals that the earthquakes have deeper depth always also more possibly to have bigger magnitude.

f.



The distribution used is Gaussian. From the plot, it is clearly to see that the data almost fit well with the 45 degree line, so the statistical distribution that plot indicates is normal distribution. Most data comply with the distribution.