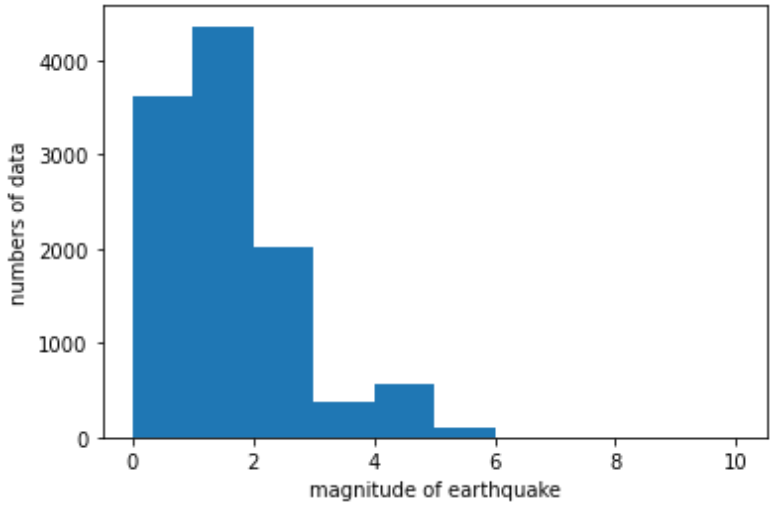
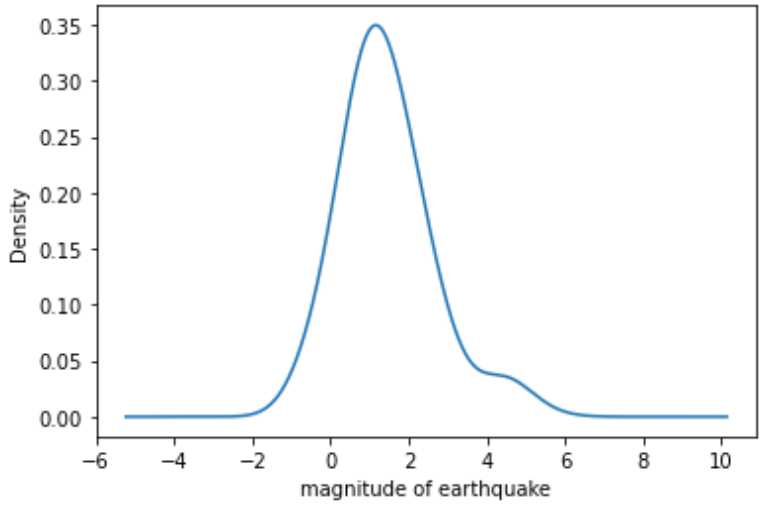
The earthquake data is downloaded on the following website:
https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php
The format of the input file is a csv file. The file contains earthquake data of past 30 days and is downloaded/accessed on 03/22/2020. Graphic analysis, such as Q-Q plot, KDE plot, and histogram, is applied to visualize the basic characteristic and quality of the earthquake data. The possible statistical distribution and the distribution of data points are discussed.

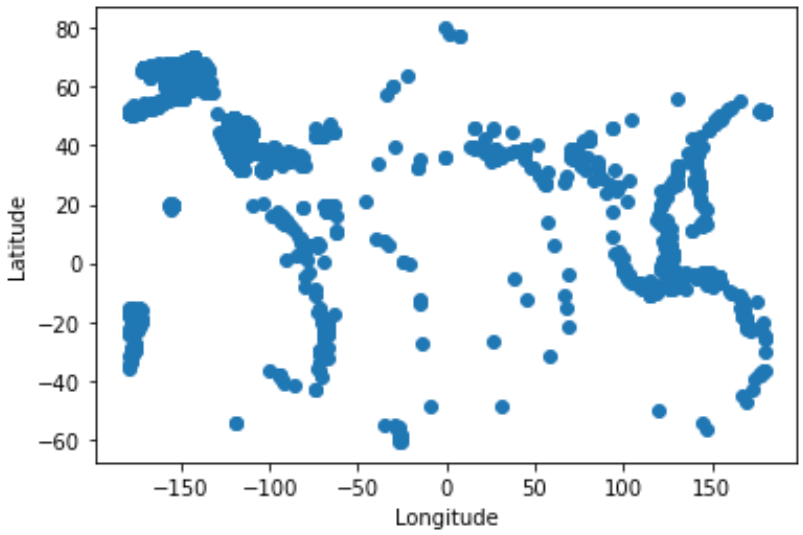


Plot 1 is the histogram of earthquake magnitude (from 0 to 1). The bin size and range would affect the shape of histogram. The numbers of data in a bin would change when the size of bins is changed. The distribution of this dataset seems to be slightly different from normal distribution, but the tail of the dataset seems not to be heavy enough to be a extreme value type 1 or 3 distribution. Therefore, we assume it to be a normal distribution in the following analysis (Q-Q plot).

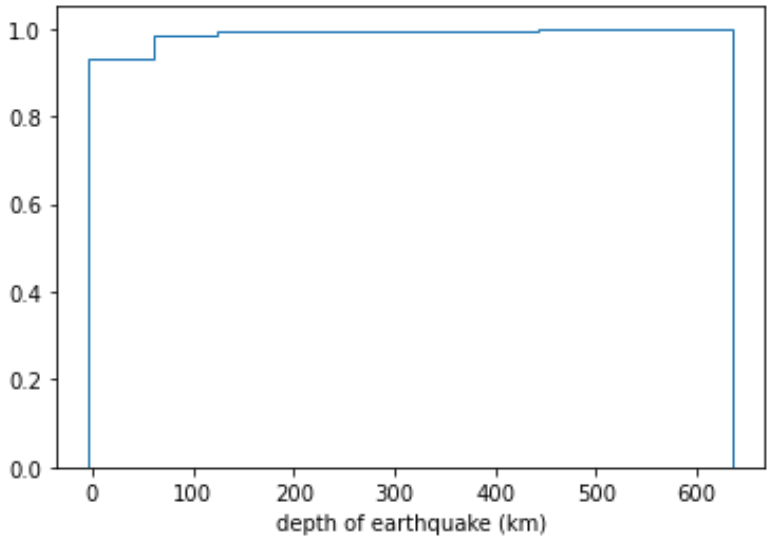


Plot 2 is the KDE plot of the same data. The kernel of KDE chosen in this plot is Gaussian kernel with size = 0.5. The difference between KDE and histogram is the way to count density or numbers. Histogram counts every within the interval and creates a

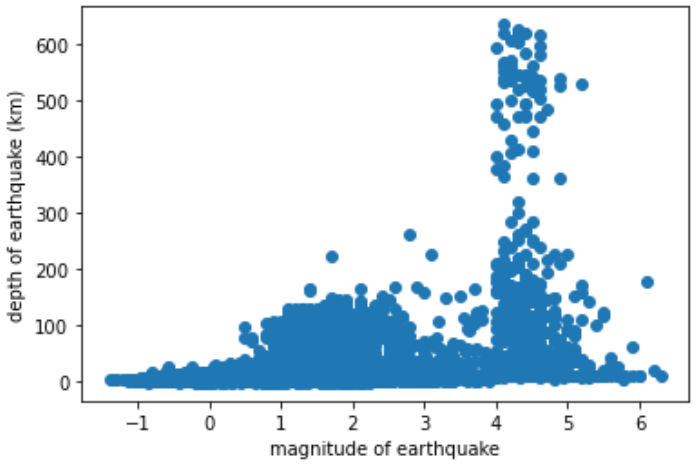
discrete pattern plot. On the contrary, KDE plot generates density with a kernel window and smooths the original dataset into a continuous plot. The similarity between these two plots is they are dealing with the same problems: showing how the data distributed. Both plots provide some aspects of the distribution of data.



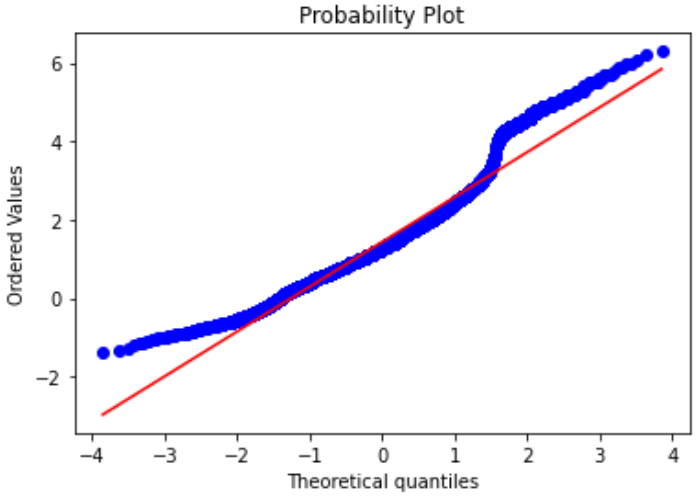
Plot 3 is the scatter plot of latitude vs longitude of earthquakes. The occurrence of earthquake within 30 days seems to be several belts and some small dots. The reason why there would be some points gathered as a continuous belt on the plot is that earthquakes are caused by the plate movement and happen at the intersection of two plates. This plot shows the location of seismic belt. The Pacific Ring of Fire is obvious in this plot if we try to connect the right most part with the left most part.



Plot 4 is the normalized cumulative distribution plot of earthquake depths. The normalized cumulative distribution plot of earthquake depths shows over 90% earthquake happens within 100 km. The deepest depth of earthquake is over 600 km. This event could be the outlier of dataset.



Plot 5 is the scatter plot of earthquake magnitude with depth. Most of the depth of earthquake is lower than 200. No significant trend or periodicity. There is an obvious cluster of magnitude of earthquake ranging from one to three. When magnitude of earthquake is larger (from 4 to 6), there are two clusters. One is shallow with depth less than 300. The other cluster is at the top right corner which include the deepest earthquakes of the whole record.



Plot 6 is the Q-Q plot of earthquake magnitudes. We compare normal distribution with our dataset. The earthquake magnitude seems not fit well with normal distribution. Tow tails of the dataset cannot fit red line. The tails are heavy and not similar to the tails of normal distribution.