

Input data:

Provenance: USGS (select all earthquakes for the past 30 days with link:

<https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php>)

This dataset was obtained on March 9th 2020.

Format: csv file, description of format of data inside the file can be found at the above link

Type of analysis:

Histograms were generated to check the distribution of events for different magnitude and compare the effect of different parameters on the visualization.

KDE (kernel density estimate) plots were generated to check the distribution and compare with the histogram generated above.

A locale figure was generated to check the pattern of occurrence.

A normalized cumulative distribution plot was generated to check the probability of occurrence for different magnitude.

A scatter plot was generated to check the relation between earthquake depth and magnitude.

Q-Q plots were generated to check if the data follow certain distribution.

Graphical analysis:

genfromtxt does not work as there are unequal number of columns, possibly from missing values.

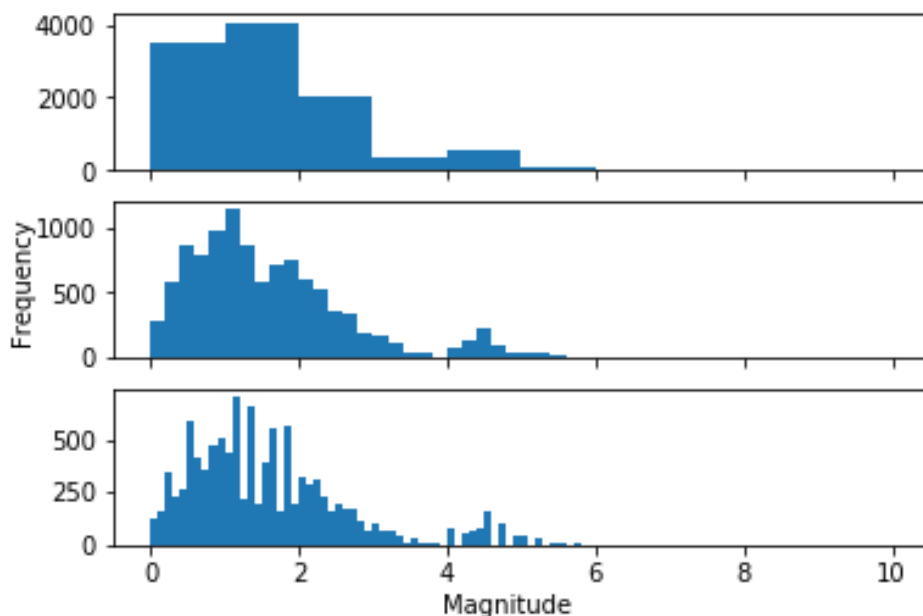


Fig 1. Effect of Bin Width Alteration

(Fig 1 include three plots to check the effect of bin width selection on visualization of distribution)

Change of bin width result in a smoother distribution, but there's a critical point where too much bins will render the plot useless as we can see a trend anymore and can make decision from it.

(A bin width of 1 is required, wouldn't that render the plot illegible? An attempt was attached at the very end)

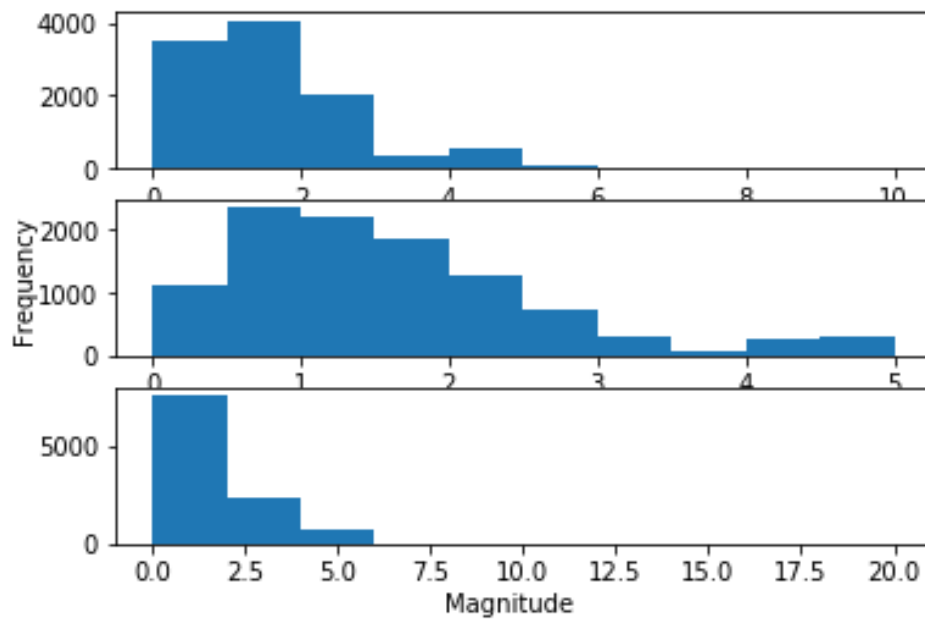


Fig 2. Effect of Range Alteration

(Fig 2 include three plots to check the effect of range selection on visualization of the distribution) smaller range result in a smoother distribution, but since range discard value outside of it, it should exercise with caution as not to lose data/information.

Looking at the required plot bin width 1 and range of 0 to 10, the distribution looks log-normal to me.

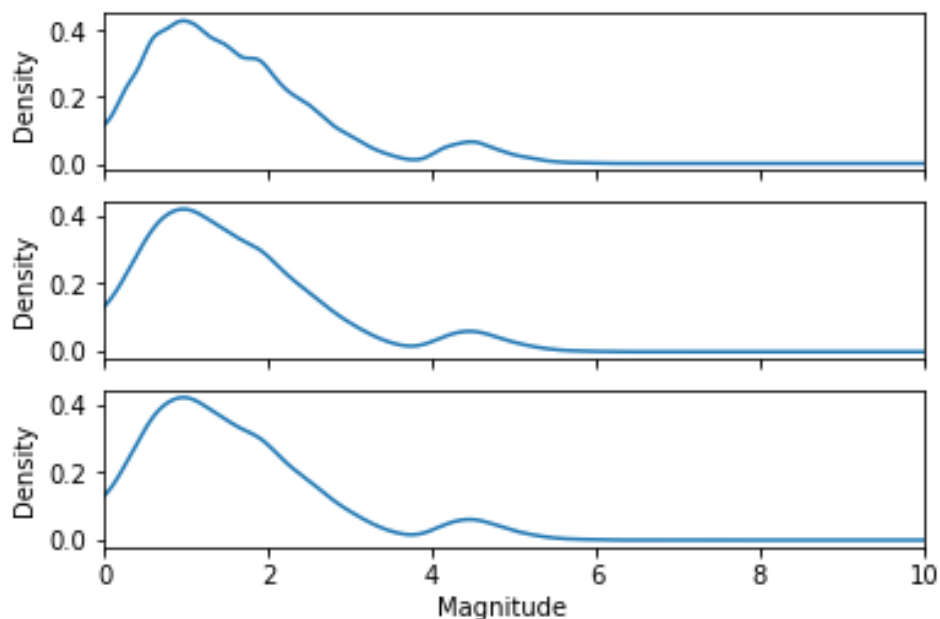


Fig 3. KDE Comparison

(Fig 3 include three plots to check the effect of bandwidth estimate method selection on visualization of distribution and compare to corresponding histogram)

Increase bw_method smooth distribution, 0.1 is the closest fit to the other two methods, width is left as default, a smaller width generates smoother result.

Naturally, KDE plot bear resemblance to histogram since the same data was used, similar distribution/shape can be observed, however, KDE is much smoother than histogram, at least when comparing the required plots (when the range was reduce to 5, their similarity converges).

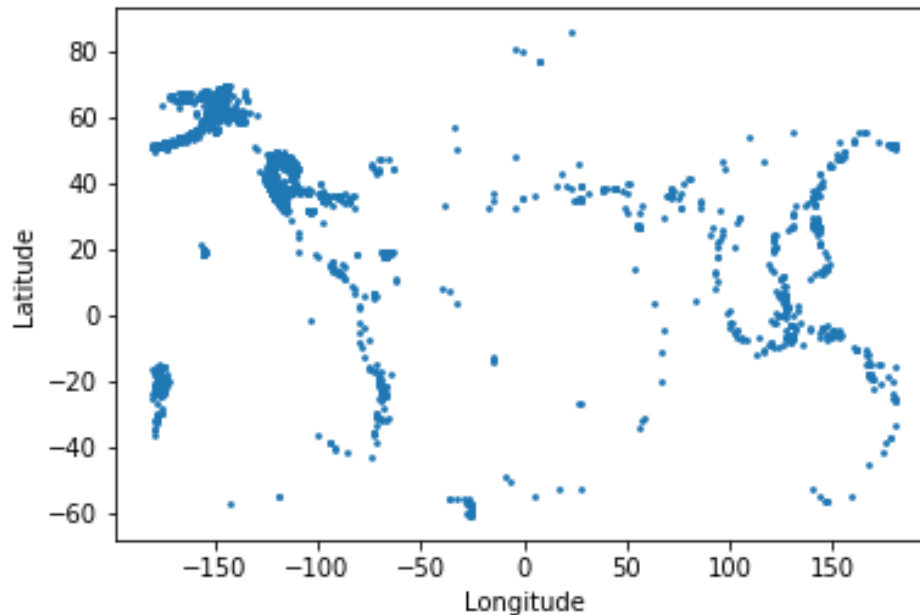


Fig 4. Earthquake Events Location

(Fig 4 shows the locale information of all earthquake happened in the last 30 days)

The placement of Long on x and Lat on y is due to Lat value changes along vertical axis and Long value changes along horizontal axis, just by looking at the map, it can be seen that most event occurs near shoreline where most of the Tectonic fault lines reside.

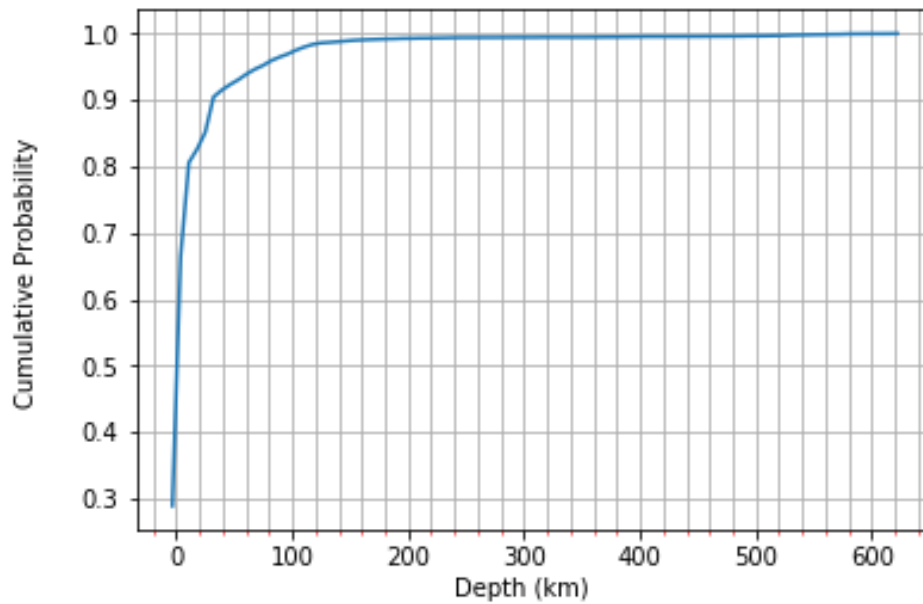


Fig 5. Normalized CDF of Earthquakes Depth

(Fig 5 shows the cumulative probability of event occurrence and shows most of the events occur are of low magnitude)

The plot (though very cluttered), shows 80% of the earthquake event happened has a depth lower than 10 km.

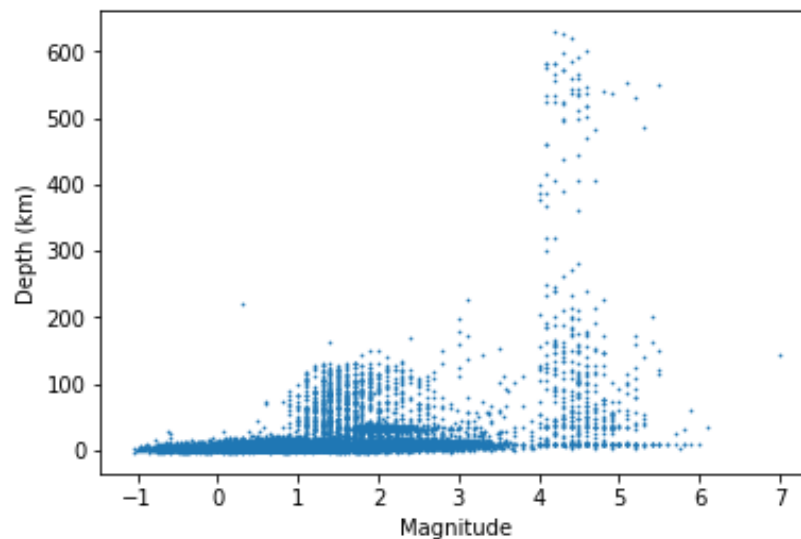


Fig 6. Relation between Earthquake Magnitude and Depth

(Fig 6 shows a possible positive correlation between magnitude and depth of earthquake)

Higher Magnitude has a positive relation to depth, especially for large value ($\text{mag} > 4$), however, there are much more occurrence of various magnitude associated with low depth.

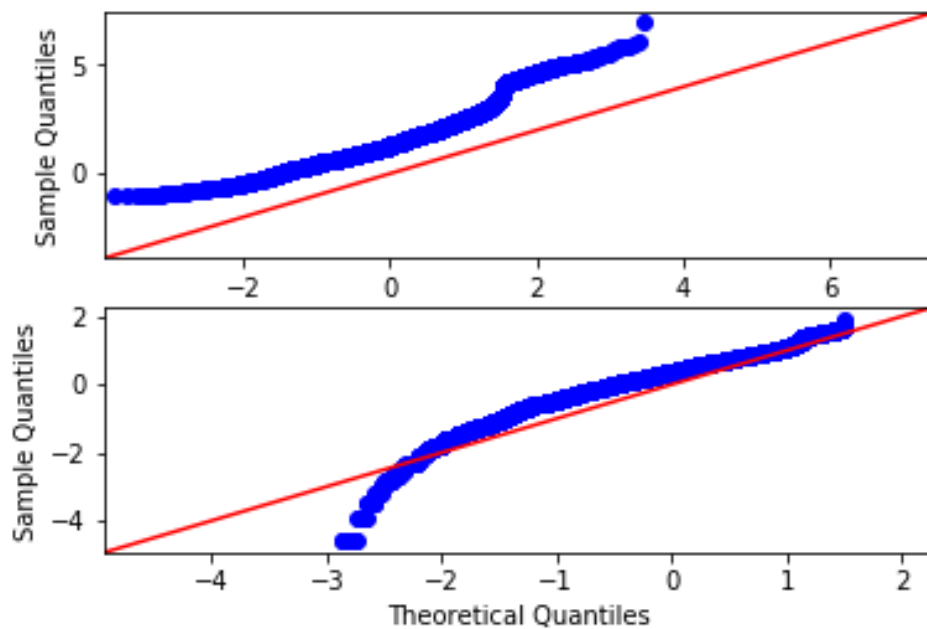


Fig 7. Q-Q Plots for Earthquake Magnitude

(Fig 7 checks if the earthquake events follow certain distribution)

Both normal and log-normal are tested , log-normal fit decently for larger values while normal fit ok overall with a shift of location parameter (mean), perhaps a Shapiro-Wilk normality test could be conducted to further answer this question.

Bin width of 1 in plot 3 of figure below:

