Joshua Tellier
2/28/2020

Lab 7 Metadata File

The input data for this assignment was downloaded from "https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php" on 2/28/2020 at 1:20 PM. The file represents all earthquakes detected globally within the past 30 days up to the minute. Information on the variables included (23 variables in total) can be found at the original data source location (the link above). Data is presented as a .csv file, which is comma-separated ASCII text. As this file contains multiple data types (strings, integers, date/time, floats, etc.), numpy's genfromtxt() will not work with it because that program cannot recognize multiple data types in a row. Instead, we use pandas read_table().

Using a python script, program-07.py, I have generated 6 figures for graphical analysis of the enclosed data. This script was designed to conduct graphical analysis on several aspects of the data. It produces:
- A histogram of earthquake magnitudes for all earthquakes in the dataset.
- A Kernel Density Estimation plot for all earthquakes in the dataset.
- A scatter plot of earthquake locations based on lat-long coordinates.
- A normalized cumulative distribution plot of earthquake magnitude.
- A scatter plot of earthquake magnitude vs. depth.
- A Q-Q plot of earthquake magnitude.

These six figures are included and captioned with relevant information and interesting notes below.

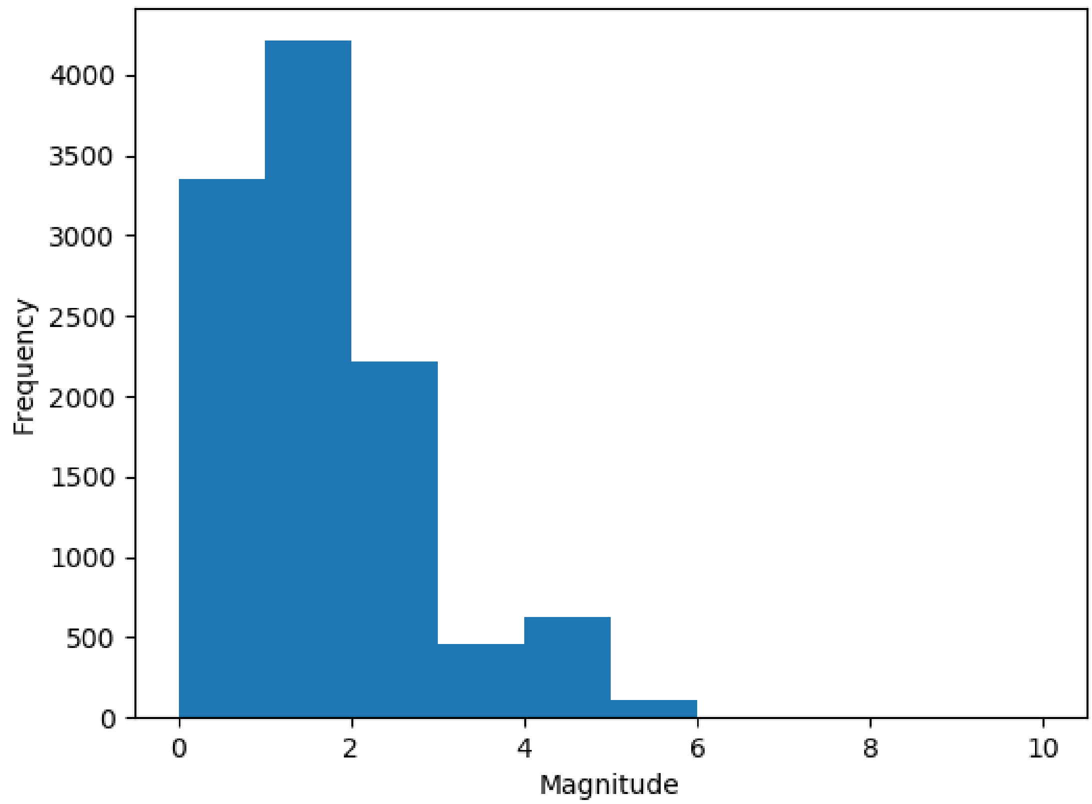

Figure 1: Histogram frequency distribution of earthquake magnitudes. Changing bin size affects the apparent visual distribution of the data by affecting the frequency count of individual bins. With a bin size of 1, we see that the majority of earthquakes in the data set have a frequency between 0 and 2. There is a second, much smaller peak between 4 and 5. No earthquakes with a magnitude greater than 6 were present in this data set. The shape of this histogram may also indicate that this particular variable is not normally distributed. It looks to be skewed right.

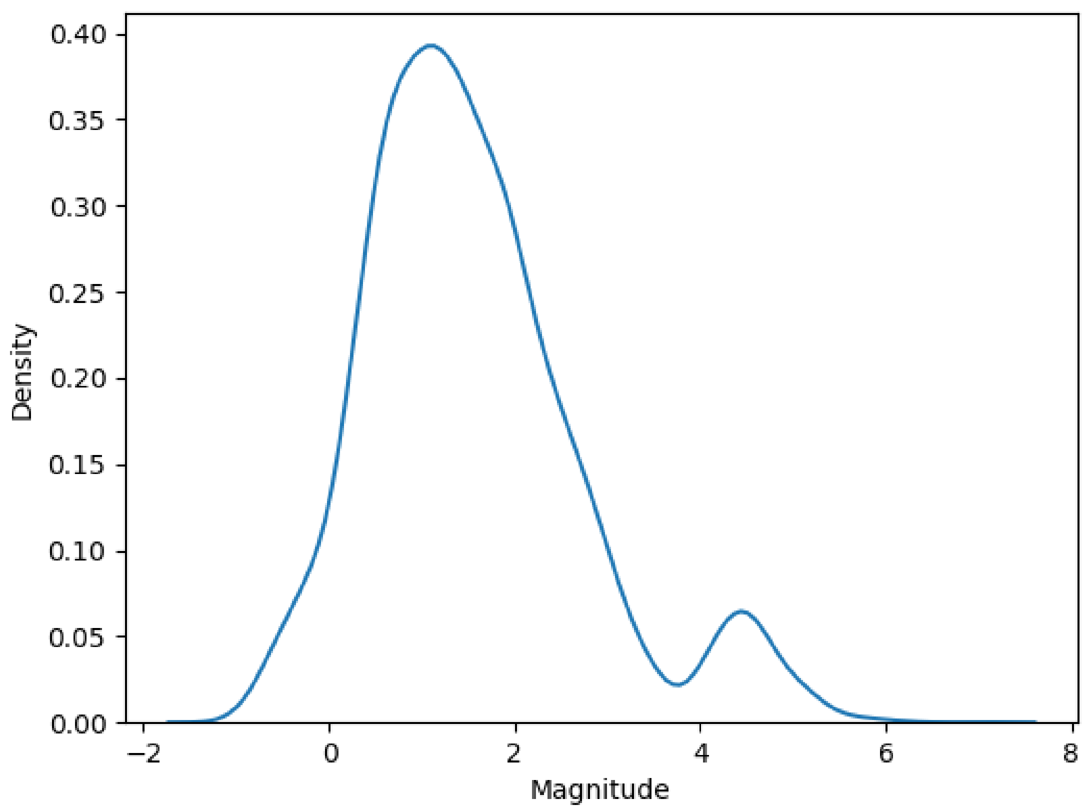

Figure 2: A kernel density estimation plot of earthquake magnitude using a gaussian kernel type and kernel width of 0.2. Both the histogram and this KDE plot show a peak in frequency around 1 or 2, and a second smaller peak just over 4. This plot provides a smoother visual representation, and we lose less data than with the rigid bins of the histogram. However, whereas the histogram only uses real data values, this KDE plot may lead us to some erroneous interpretations because it shows some probability outside of the range of possibilities. For example, this KDE plot extends beyond the points of real data on the x-axis..

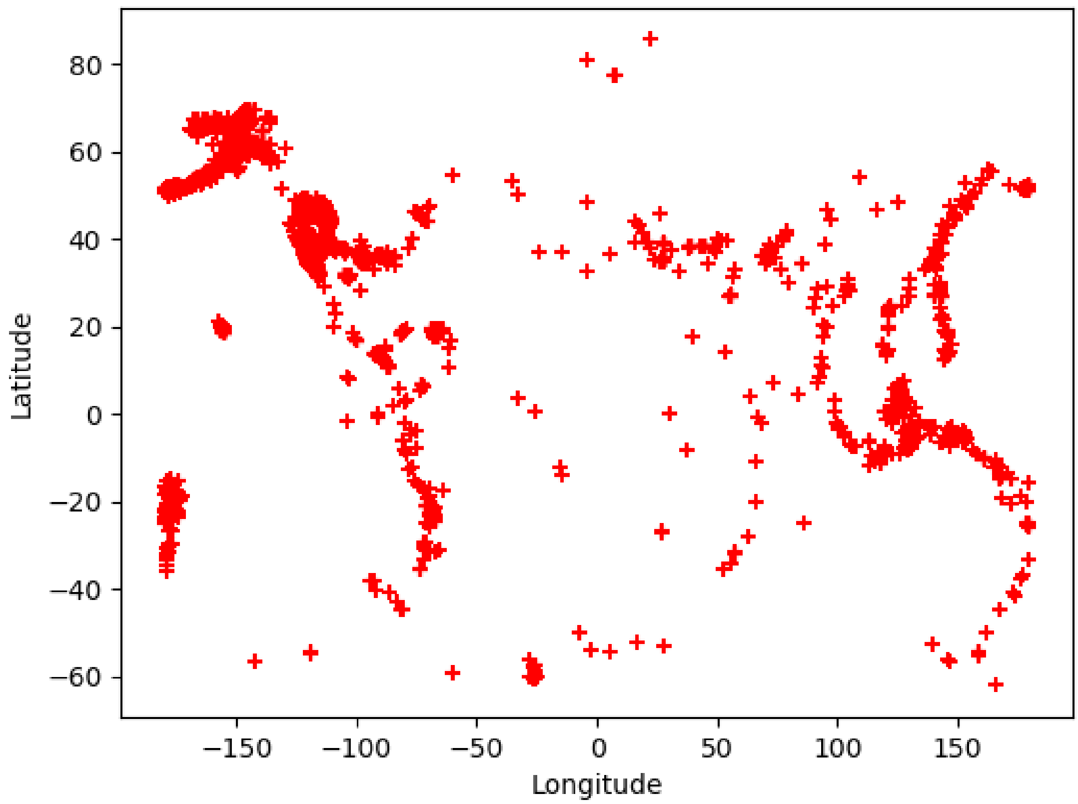

Figure 3: Scatter plot of earthquake locations using latitude(xaxis) and longitude (y-axis). Putting the variables on these respective axes mirrors that which is seen in real life, so we get a better sense of where these earthquakes are on the globe. It is even possible to see the outline of some countries/continents based on this scatter plot alone. Those areas of the world which are visible in this plot correspond to areas of high geological activity. For example, the northwestern coast of the United States and Alaska are clearly visible, as is the western coast of South America, Southeast Asia, and some mid-Pacific fault lines.

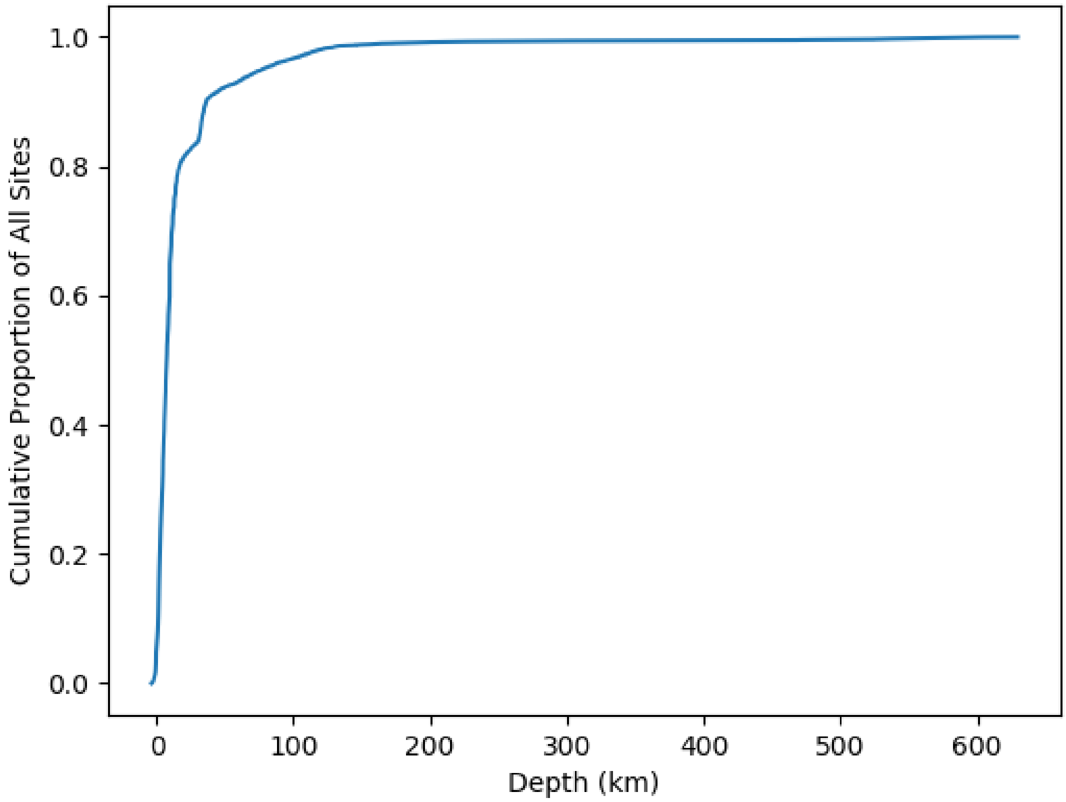

Figure 4: Normalized cumulative distribution plot for earthquakes based on depth. From this figure, we can see that the vast majority of earthquakes (about 90-95%) in this data set occurred within 100 km of the Earth's surface. From there, we see that the cumulative distribution slowly continues growing up to 1.0, but over a much wider interval. This indicates that earthquakes below 100 km depth are sparsely found and highly variable in their depth.

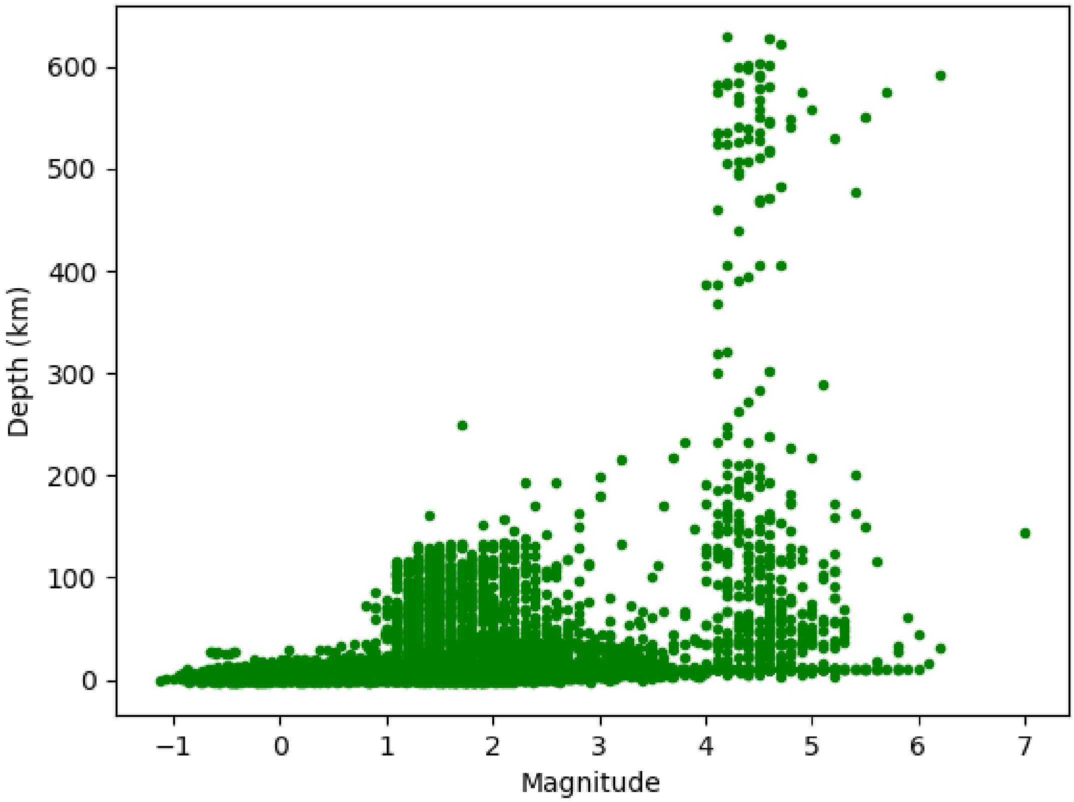

Figure 5: Scatter plot of earthquake magnitude (x-axis) vs. depth (y-axis) for the USGS 30-day data set. From this figure, we can see that there is some weak correlation between earthquake magnitude and depth. The strongest relationship seen is that weaker earthquakes (< magnitude 4) occur shallower in the Earth, whereas earthquakes of greater magnitudes can occur anywhere along the subsurface depth. The majority of earthquakes occur in the shallow depths, which backs up the conclusion that we drew from the cumulative distribution plot.

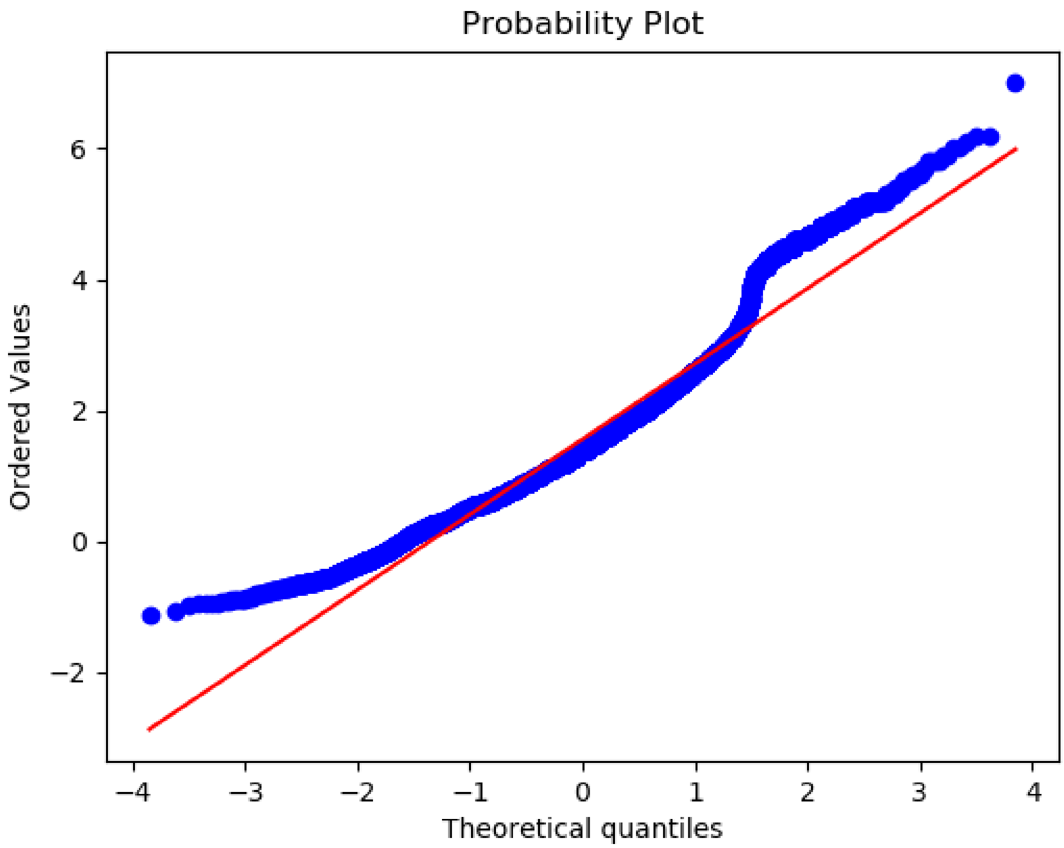

Figure 6: Q-Q plot of earthquake magnitude. The data does not follow the red normalized line and so we can clearly say that this variable is not normally distributed. This figure backs up the conclusion that we drew from the first histogram. The Q-Q plot assumes normal distribution of the data. Our chosen variable clearly violates that assumption, and the result is reflected in this figure.