Les Warren
ABE 65100 Spring 2020
03/17/2020

Lab 7 Meta data (Meta data.pdf)

The input data set used for the lab was downloaded from the USGS Earthquake database found at <https://earthquake.usgs.gov/earthquakes/feed/>. The file was downloaded on 02/28/2020 at 1:10pm. The original file downloaded as a .csv can be found in the github repository.

The file downloaded represented all earthquakes detected on earth for the last 30 days up to the minute that you downloaded the file. The original file included 23 variables that varied in formats from strings, integers, date/time). To be able to input the file and read all of the variable formats, the pandas read_table() function was utilized. Genfromtxt() would not be suitable to use since it will not read the various data format types.

Once the input file was read into python, a script (program-07.py) was formed to generate six different figures. These figures were used to perform graphical analysis of the data set. The six figures included:
1) Histogram of earthquake magnitudes
2) KDE (Kernel Density Estimation) plot of earthquakes
3) Scatter plot of earthquake latitude/longitude
4) Normalized cumulative distribution plot of earthquake magnitude
5) Scatter plot of earthquake magnitude vs. depth
6) Q-Q plot of earthquake magnitude

Figures are presented below.

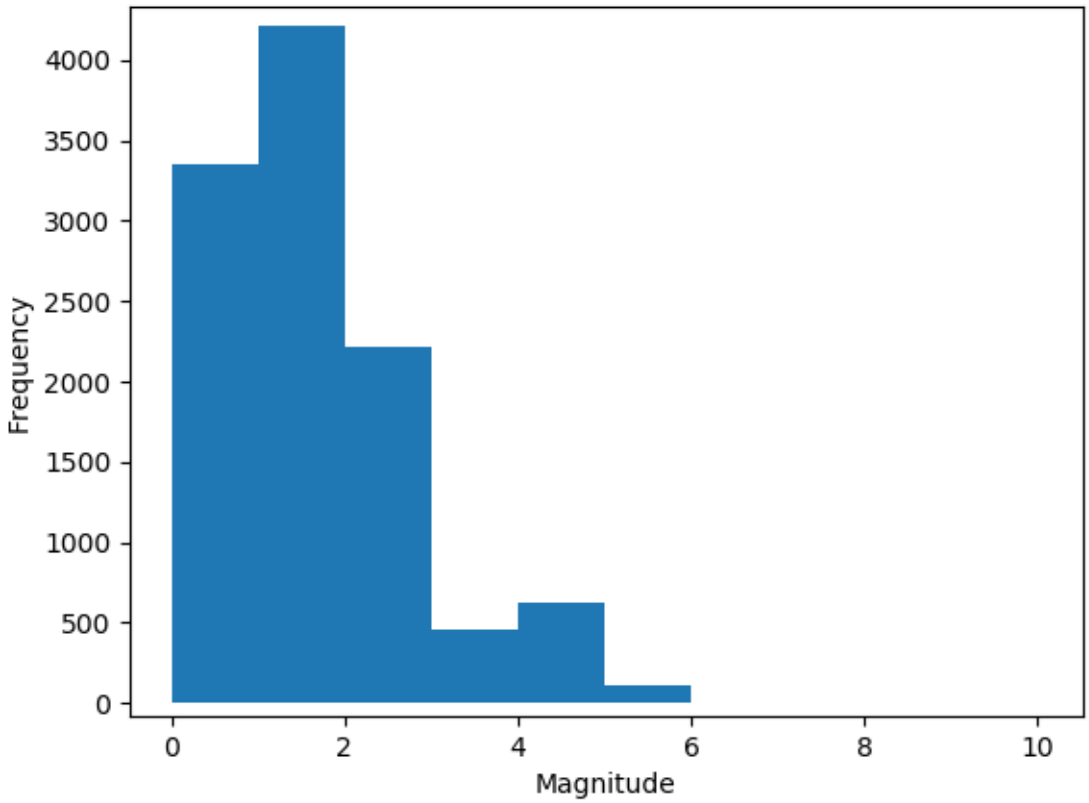

Figure 1. Histogram showing distribution of earthquake magnitudes. When changing the bin size, it can affect the visual distribution of the data. In this case, a bin size of one is used it shows majority of the magnitudes in the first two bins (0-2). Buy selecting a smaller or larger bin, it can drastically change the appearance of the data. This data set shows majority of magnitudes to be 2 or below and there are no earthquakes over a magnitude of 6. Another thing to notice is the data being skewed to the right side (2 and below), making it unnormalized.

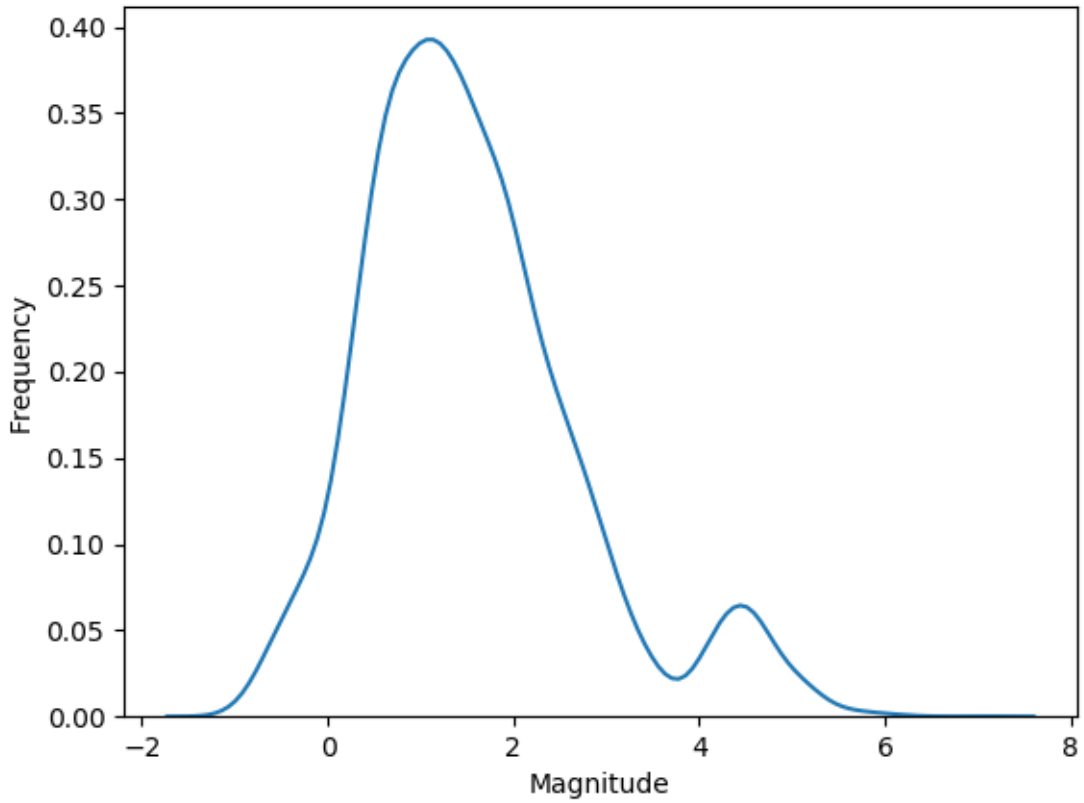

Figure 2. KDE (Kernel Density Estimation) plot of earthquake magnitude using a kernel width of 0.2 and a Gaussian kernel type. This KDE plot is very simular to the histogram (figure 1) as both show a large peak at 2. One difference though is that the KDE plot uses an estimation to form a probability whereas the histogram is plotting the actual data.

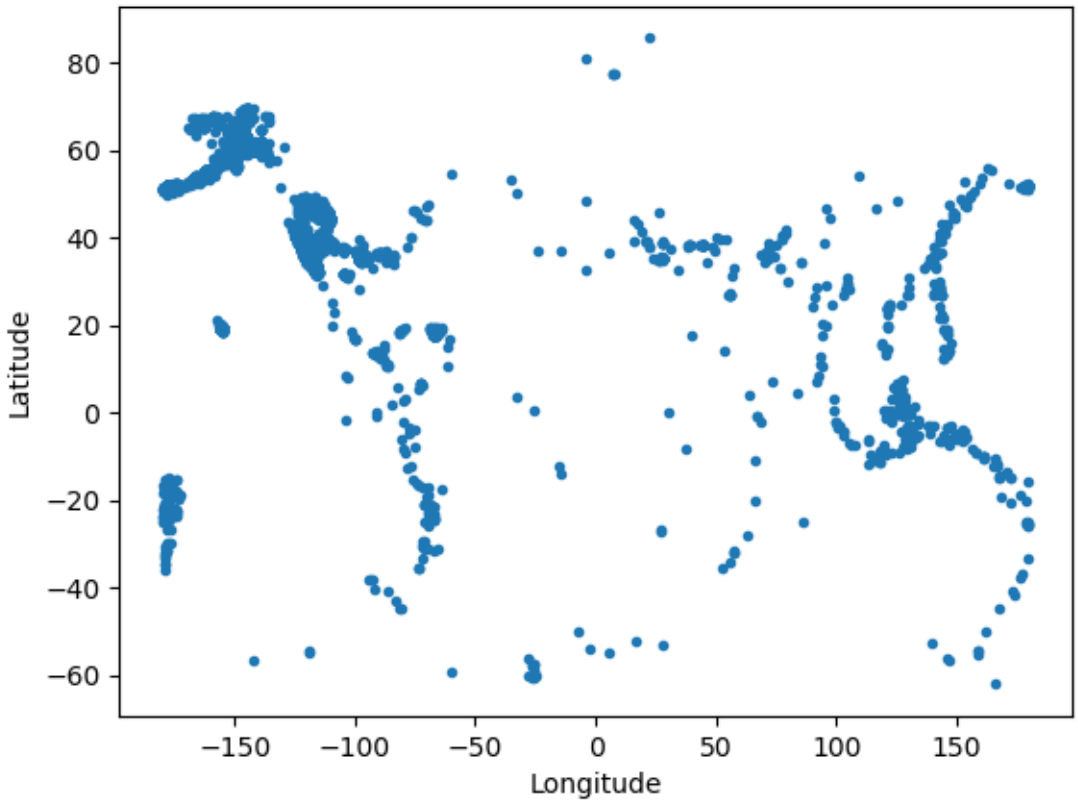

Figure 3. Scatter plot of the earthquake locations using longitude on the x-axis and latitude on the y-axis. By putting the respective variables on the axis's used, it is just like plotting a point on a map. If you look close enough, you can see a faint distinction to a world map.

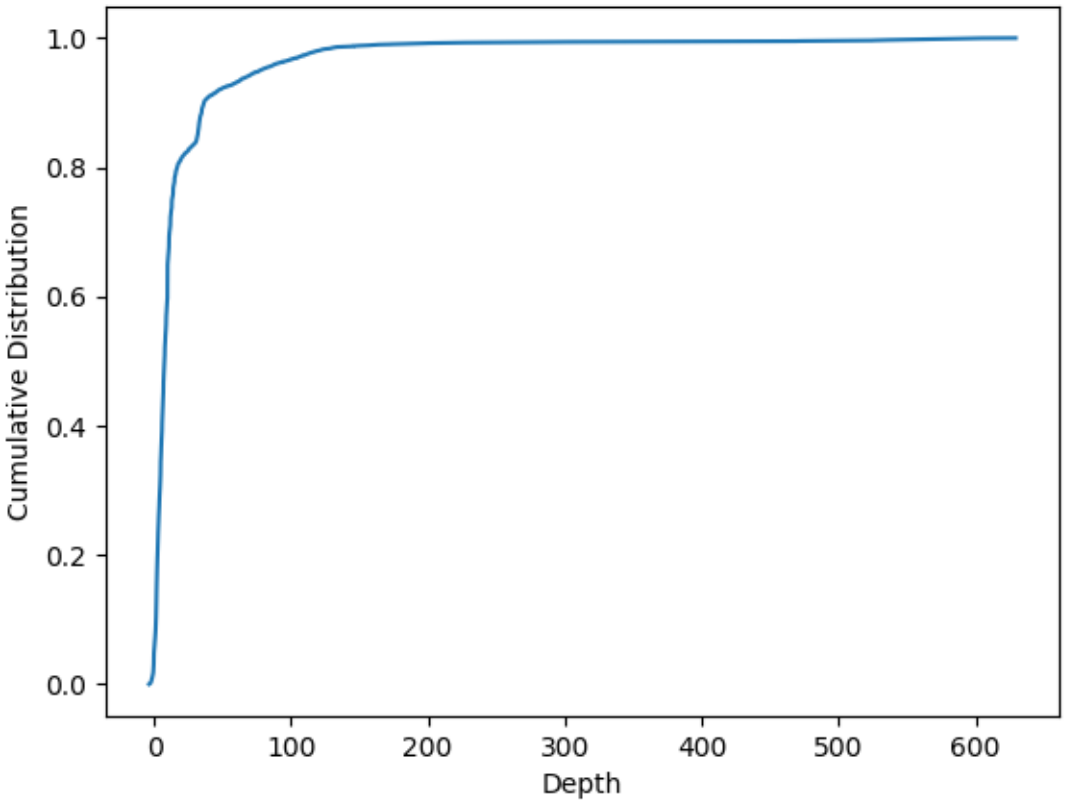

Figure 4. Normalized cumulative distribution of earthquakes based on depth. From this figure, you can see majority of earthquakes occurred withing 100km of the earth's surface. Earthquakes are deeper depths are much less frequent.

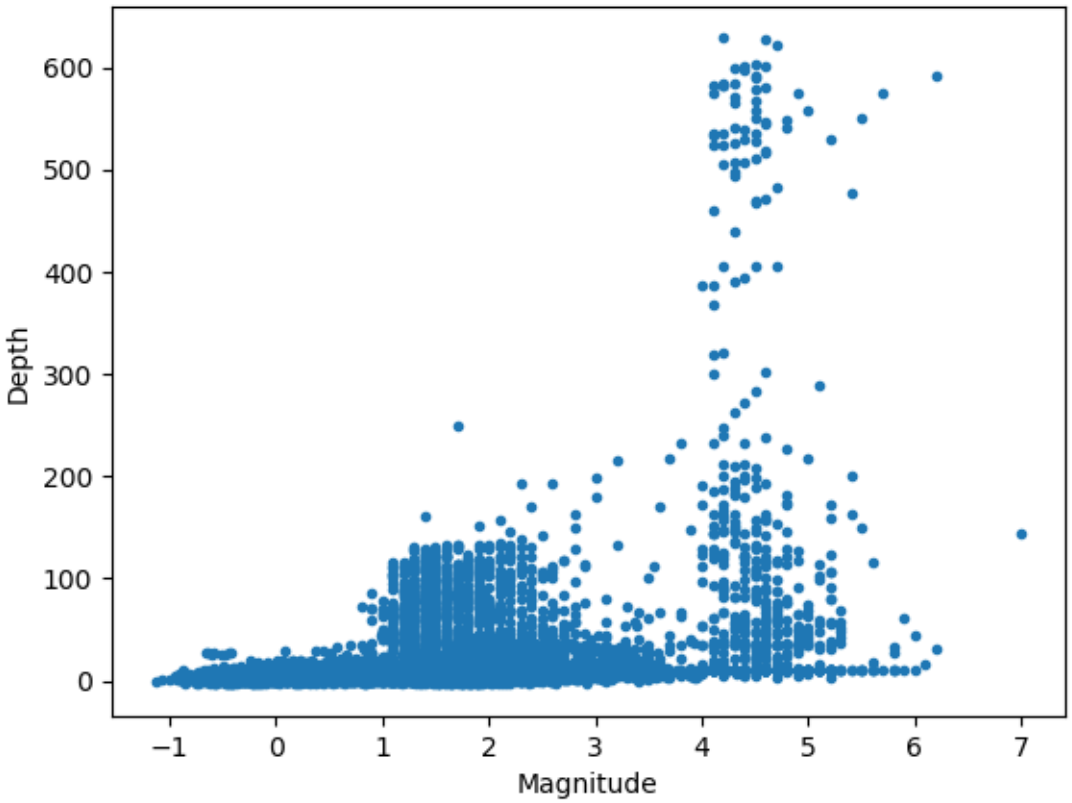

Figure 5. Scatter plot of earthquake magnitude vs. earthquake depth. Looking at this figure, shallower earthquakes occur at much less magnitudes, whereas earthquakes with deeper depths have potential for much greater magnitudes.

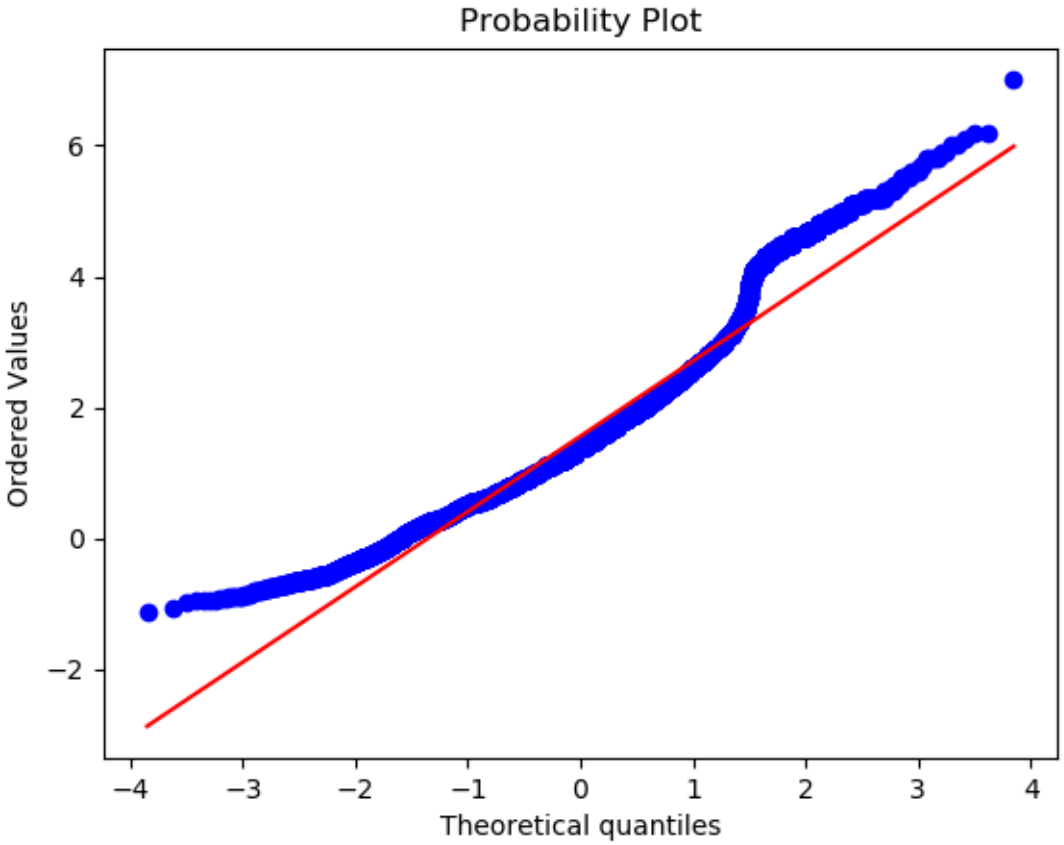

Figure 6. Q-Q plot of earthquake magnitude. From this plot, you can tell that the data is not normally distributed as it does not follow the red line. The Q-Q plot assumes normal distribution and this variable does not follow that assumption. This is much like the histogram in figure 1.