ABE651, Purdue University
Data downloaded, analyzed, and submitted by: Danielle Wagner
Context and Accessing Data

Data:
3/21/20 16:02 All data from the last 30 days downloaded & saved as a .csv file from USGS data that is updated every 30 minutes at https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php
Because data is taken from the USGS, locations reported are in the United States. Information reported includes 1ocation and time of observation, the earthquake depth, magnitude, and ID.

Data was imported into python using *pandas.read_csv()* as it was downloaded in .csv format. Some data is missing in some of the columns, and thus data separators become imperative in this case. In addition to these separations and empty slots, the data is not all in the same format; and for these reasons functions such as genfromtxt() would fail to properly import the information.

Explanation:
Graphical analyses were performed to understand the depth and magnitude data from spatial and statistical perspectives. Plots were double-checked in matlab but the following figures were generated in python. For the plots that did not turn out as expected, the matlab-made version is shown (Fig's 1 and 6). The code for created the figures is submitted along with this document in the GitHub folder '07-graphical-analysis-with-python-wagne216''.

Figures:

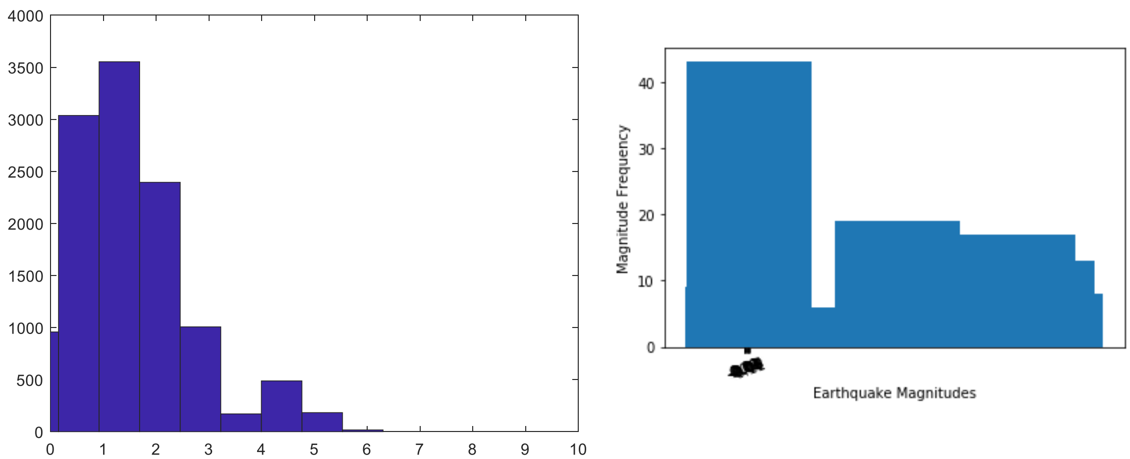

Fig. 1: a.) Python- and b). Matlab-created histograms of the earthquake magnitudes and the frequency of observation

The histogram in Figure 1a depicts the earthquake magnitudes, using matplotlib.pyplot (left). When choosing the data ranges, the histogram varied wildly as the shapes rearranged. This variation is due to the fact that the magnitudes are sorted into different bin categories and redistributed, thus there is not an unbiased way to show data using this discretized method. The matlab-created histogram (right) shows the same distribution but in a cleaner manner due to the equal bin widths.

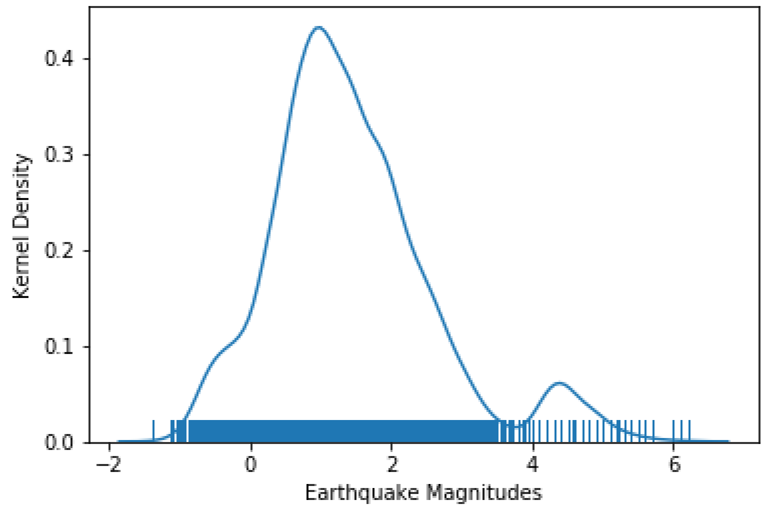

Fig. 2: Kernel Density (KDE) Plot of earthquake magnitudes using the distplot seaborn module.

Distplot takes the histogram a step further by calculating appropriate bin sizes to show the variable's (magnitude) probability density. The bandwidth used to create the KDE in Figure 2 was automatic. A Gaussian distribution was used by setting "kde = true". The density plot is similar to the histogram in Fig. 1 in that it conveys the general distribution of magnitudes and how often they were observed. In both Fig's 1 and 2, the data distributions have a higher, more pronounced peak where there are lower values, and a smaller peak at the higher values. They are different, however, as the KDE plot is more exact because it displays a continuous curve rather than sorting the data into bins. The barcode-like line at the bottom is an optional feature, called a rugplot, that shows the data spread.

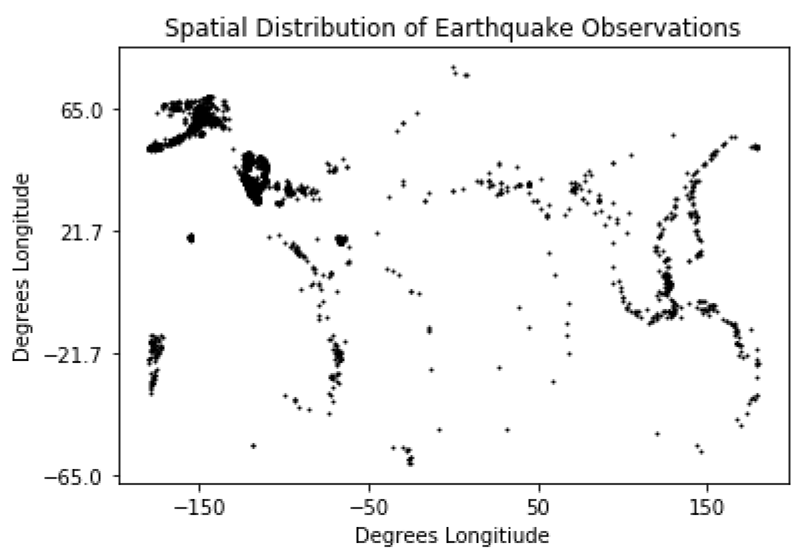

Fig. 3. A scatterplot is used to display the locations of the 30-day earthquake observations.

In Figure 3, the longitude is represented on the x-axis, as it is the East-West direction and thus more intuitive to think of it as left-to-right; and for the same reason latitude (North-South) is on the y-axis. As expected with earthquakes, there is a clear spatial arrangement to the observations. If this map were overlapped with a map of the world, one would see that the left side of the plots (around x = -150 deg) is the west coast of the western hemisphere, and that the right side (~x =150 deg) is the east coast of the eastern hemisphere, which is essentially the ring of fire.

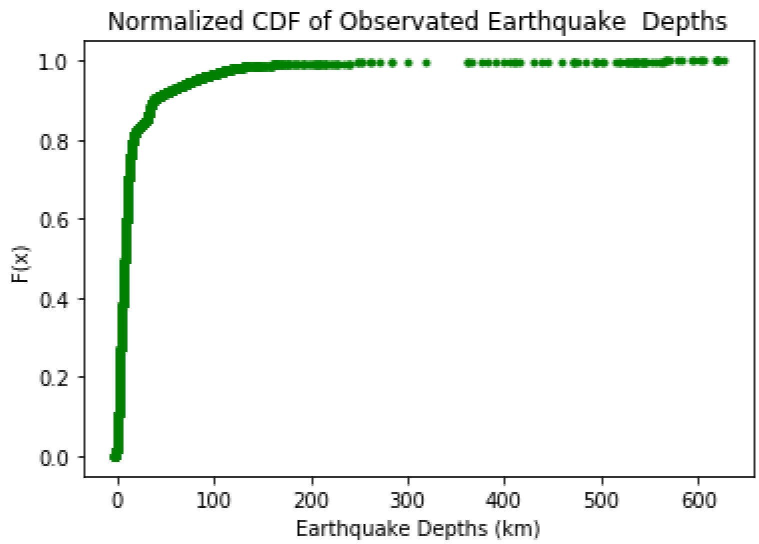

Fig. 4. Normalized CDF of observed earthquake depths

Fig. 4 is a plot of the normalized cumulative distribution function (CDF) of the earthquake depths. The depth data was first sorted in order from lowest to highest, and then plotted vs. its index number. The index number was divided by the total number of datapoints in order to normalize the data. Buy plotting the sorted data as a proportion of the total, the CDF shows the proportion of measured depths that are less than a certain value. In this case, most depths are more shallow than 100 units. This information can quickly be inferred by observing that F(x) progresses rapidly toward 1 in the beginning of the plot, and almost peaks very early on. The '.' marker was chosen in this plot rather than a line, to show an extra level of information; around x=300, there are less measurements. This choice is not as useful for overlapping measurements.

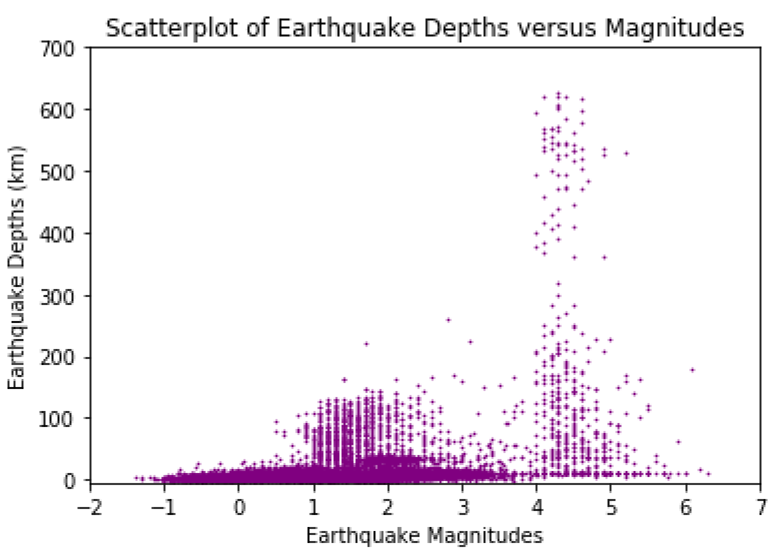

Fig. 5. Observed earthquake depths vs magnitudes

Fig. 5 visualizes the relationships between the earthquake depths and magnitudes in a scatterplot. From the plot, there is not one specific trend, however there are a few notable patterns. For magnitudes less than 5, the depth does not usually exceed 200 meters. Magnitude levels above 4 can reach depths of around 650 meters, which are not seen in lower-intensity earthquakes. It is interesting that higher-intensity earthquakes can begin at almost any depth, and that those with intensities less than 4 are not measured below 200 meters for this dataset.

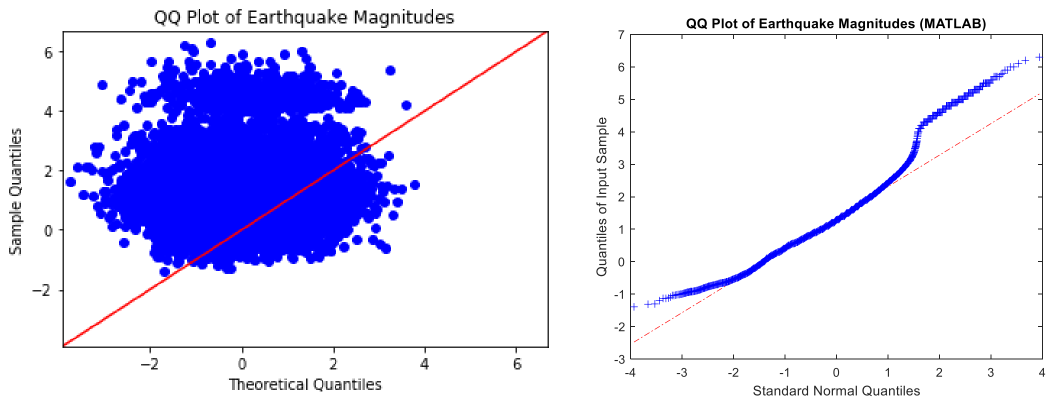

Fig. 6. Quantile-quantile (QQ) plots of the same magnitude data with a.) an attempt in Python and b.) the expected plot generated in matlab.

The QQ plot in Fig. 6a. does not show any obvious relationship when compared to normal quantile data, however, there is a very close relationship when viewed in Fig. 6 b. The 45 degree red line in Fig 6. a. represents what the blue dots would align with if there were a normal relationship. Because the earthquakes are global, it is not unreasonable that there is not a normal relationship amongst them.