Author – Eric Kong
Date – 3/8/2020
File Information – Graphical Analysis with Python Outputs and Answers

Earthquake Data
The data for this analysis was retrieved from the following URL:
https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php

On the right-hand side of the website, there is a section titled “Past 30 Days”, the option under that called “All Earthquakes” was selected. The resulting file is called “all_month.csv”.
The comma-separated values (CSV) file was downloaded on 2/27/2020 at 9:03 pm to the cloned assignment 07 directory.

Data Analysis
The following graphical analysis was conducted on the earthquake data by using a Python script:

- Histogram
- Kernel Density Estimate (KDE)
- Scatter Plots
- Normalized Cumulative Distribution
- Quantile Plot

numPy genfromtxt error
After attempting to use the genfromtxt function from the numPy module, it helped me isolate the error. The Python console reported back, “ValueError: Some errors were detected ! Line #84 (got column 22 columns instead of 23)”.
After looking at line 84 in the raw data and reading online, genfromtxt will not work if a column in the CSV data is missing data.

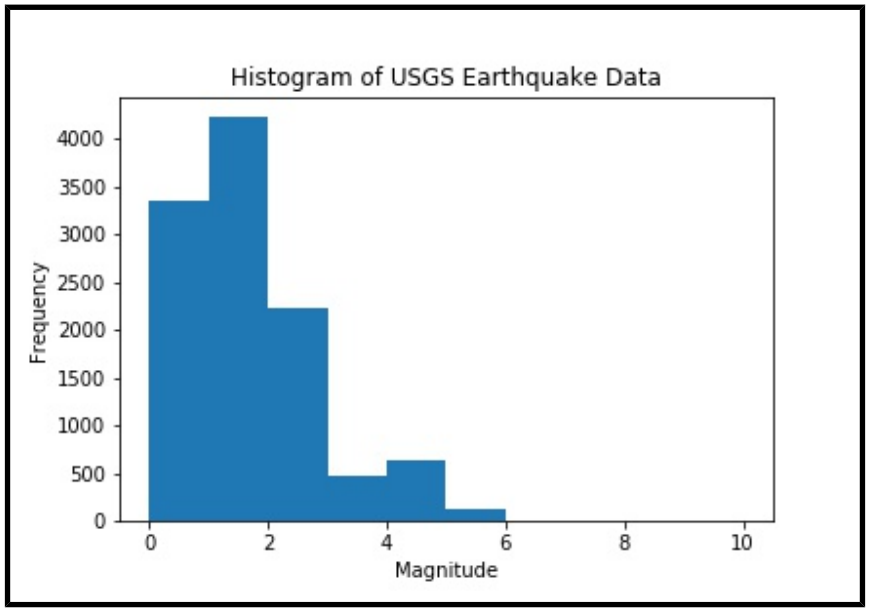

Image 1: Histogram of USGS Earthquake Data according to magnitude

How the histogram displays and how it is interpreted is dependent on the bin sizes and range. For example, a large number of earthquakes can have a magnitude below 0.5, but since the bin intervals are at 1, the subtlety of the decimal values is lost in the integer separation.
For plots of data, the data is the focus and needs to be shown. For this data, there is no earthquakes with magnitudes above 6. This will not be known unless the data is plotted first. Changing the plot range to match the data so it is displayed nicely is the correct thing to do after determining how the data actually sits on the range.

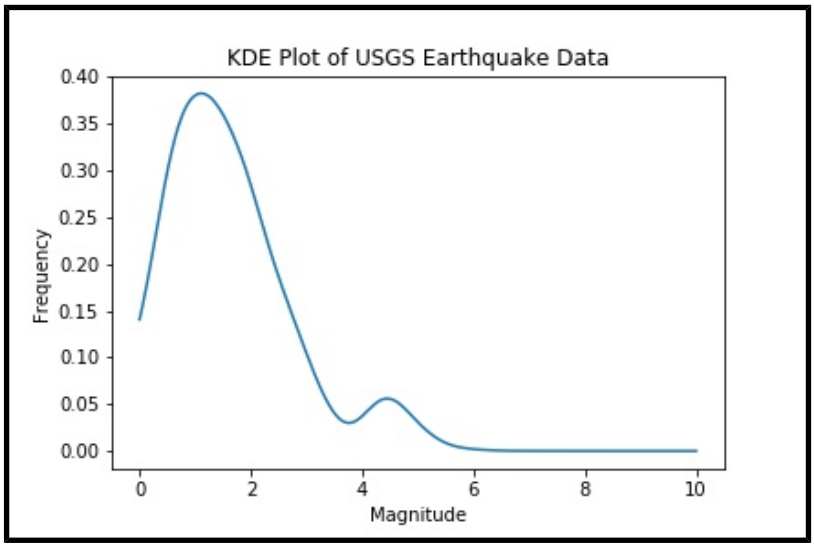

Image 2: KDE Plot of USGS Earthquake Data according to magnitude
The kernel type is Gaussian and the kernel width is 0.25. The scipy Gaussian KDE works for both uni-variate and multi-variate data. This method has the possibility of oversmoothing bi-modal and multi-modal distributions. For a better comparison of the KDE and Histogram, the same range for the x-axis is used. Both plots show two distinct peaks near the magnitudes of 1 and 5. There is a positive skew because the frequency of high magnitude earthquakes is low. The separation of the bins is slightly lost with the smoothing of the KDE, especially in the 0-1 interval.

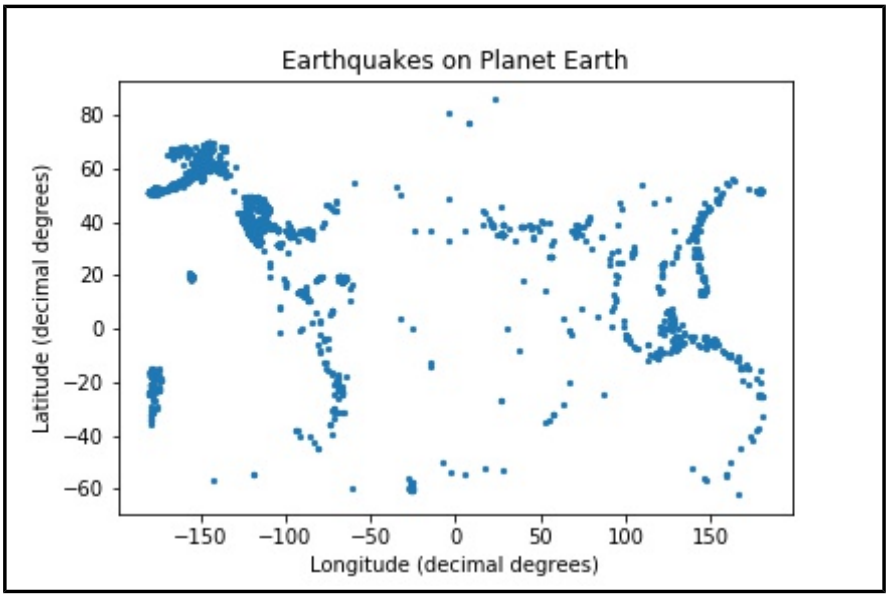

Image 3: Longitude and latitude of all earthquake data as a scatter plot
The data is plotted with longitude on the x-axis and latitude on the y-axis because that is how the global coordinate system for the Earth works. Based on the distribution of these points, it is easy to see the Ring of Fire. The Ring of Fire is an earthquake heavy area in the Pacific Ocean.

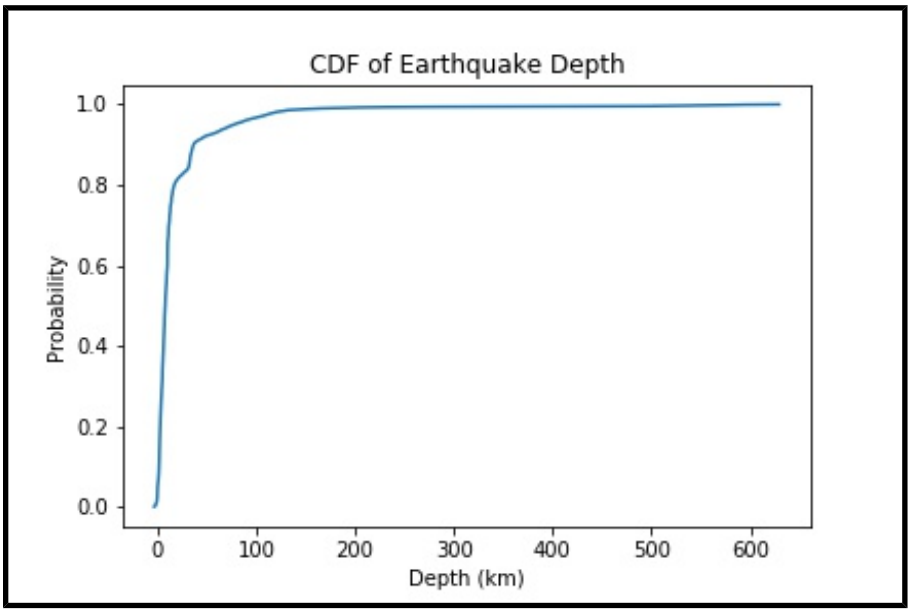

Image 4: Normalized cumulative distribution function of earthquake depth
The results in image 4 indicate that most earthquakes occur at a depth that is less than 100 km. The slope of the CDF line increases very quickly on the left side of the plot. This means that the probability of an earthquake with low depth is very high.

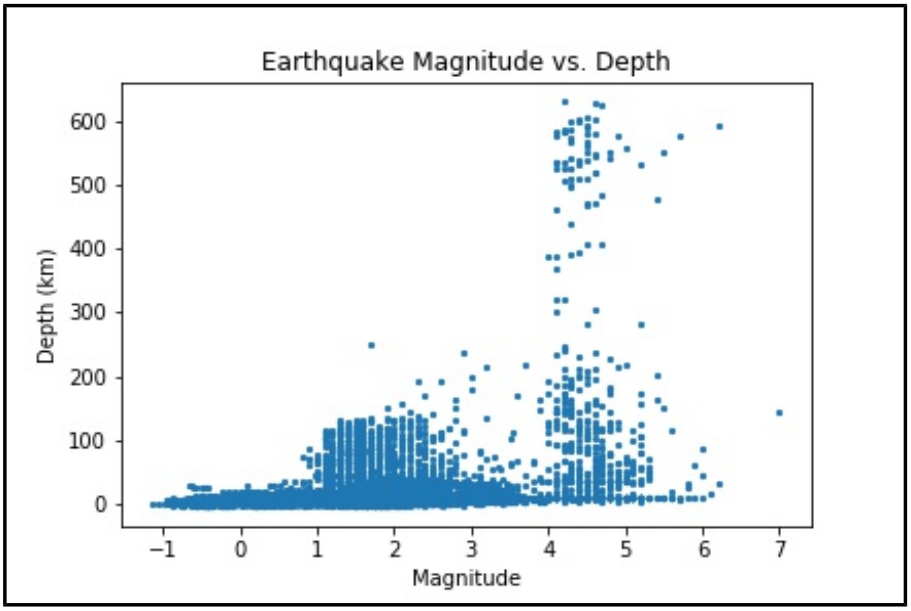

Image 5: Scatterplot of earthquake magnitude versus earthquake depth
Based on image 5, the relationship between magnitude and depth seems to be positive. Meaning that more powerful earthquakes occur at deeper locations. This can be seen in the peak near the 4 to 6

magnitude region. There is a large cluster of earthquakes that occur at very low depths from 0 to 4, the markers form a thick concentration in this area.

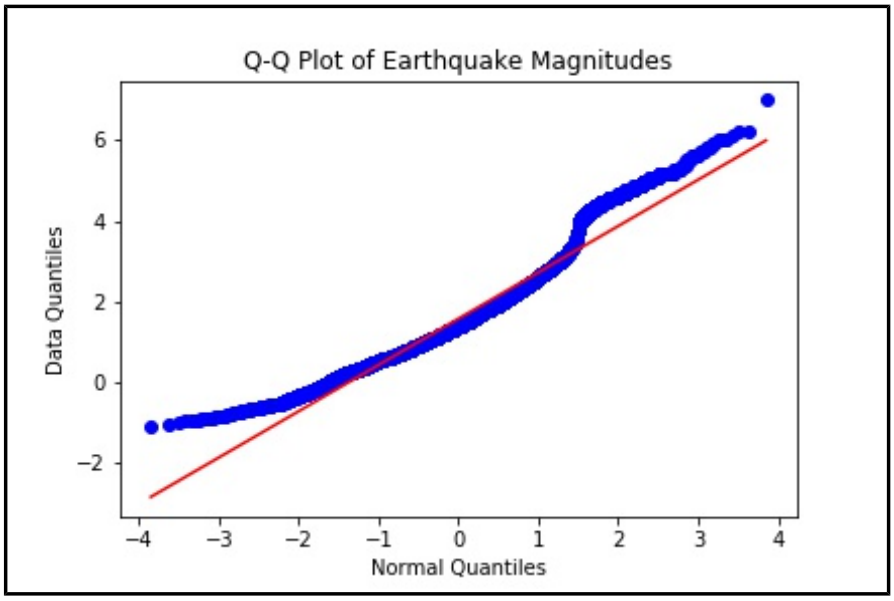

Image 6: Quantile plot of earthquake magnitudes
The normal distribution was used along the x-axis to compare it with the earthquake data along the y-axis. The red line shows a perfect relation between the data's quantiles and the quantiles formed by a normal distribution. The data complies with the normal distribution for a small region near negative 1 and positive 1. The blue circles are markers from the USGS data and it does not align well at the tails of the data or the center at zero.