Assignment 7- Graphical Data Analysis with Python- Metadata

Source and Format of Input Data:

The earthquake data was downloaded from the USGS Earthquake Hazards Program website: https://earthquake.usgs.gov/earthquakes/feed/.
The file downloaded can be found using the link below and clicking *All earthquakes* in the *Past 30 Days* section on the right column of the page.
https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php
The file will be downloaded in csv format and the name of the input file used in this code is *all_month.csv*. This file contains the following fields:

time	latitude	longitude	depth	mag	magType	nst	gap	dmin	rms	net
id	updated	place	type	locationSource	magSource	horizontalError	depthError	magError	magNst	status

The descriptions of each of the columns can be found here:
https://earthquake.usgs.gov/data/comcat/data-eventterms.php#nst

The data that was used to develop the code was downloaded on Feb 28th, 2020, at 4.50pm and correct as of March 1st, 2020. This data is constantly updated based on real time collection of the earthquake data.

Types of analysis conducted:

The following is a list of analysis that was conducted on the data:

1. Visualization of magnitude of the earthquake using a histogram (with bin size=1)
2. Visualization of the Kernel Density Estimates (KDE) of the magnitude of the earthquakes (kernel width= 0.05)
3. Visualization of the location of the earthquakes using latitude and longitude
4. Cumulative Distribution (CDF) plot of the earthquake depths
5. Scatter plot of the Depth vs. Magnitude of earthquakes to search for a relationship
6. Q-Q plot of the earthquake magnitudes.

Discussion of Figures and Results:

Why will genfromtxt() not work with this dataset?
In this case genfromtxt() will not work as the file is in the csv format and it contains multiple data types.

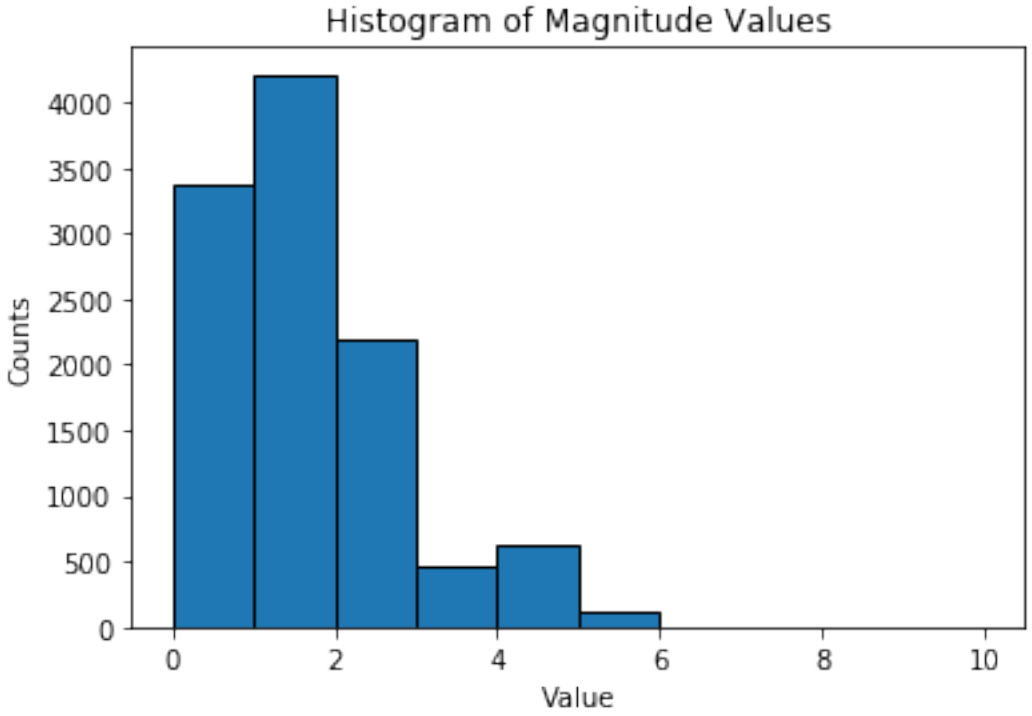

Figure 1: Changing the bin size changes the distribution of the histogram itself. Having many bins of a small size displays a more accurate distribution of the data. The histogram shows that the magnitude of earthquakes is more likely to be lower than 3 with the maximum number of earthquakes between the 1-2 range. There are no earthquakes larger than 6.

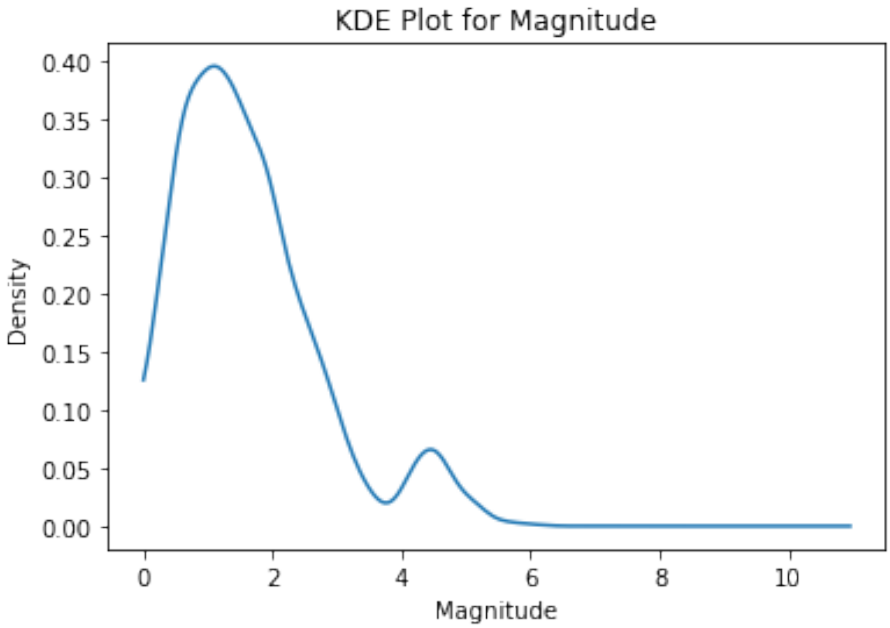

Figure 2: Kernel type: Gaussian; Kernel width: 0.05. The KDE plot is similar to the histogram as both show the density of the magnitude values fairly well. The KDE plot shows more detail than the histogram as it has a very small width which allows the plot to display the density of exact values.

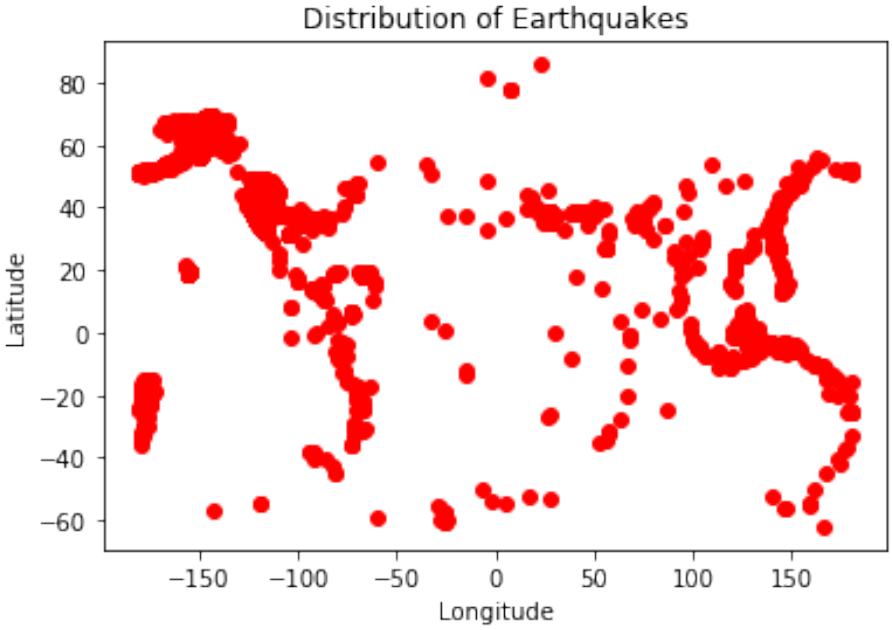

Figure 3: Earthquakes mostly occur around the ring of fire (Longitudes<-60 and Longitudes>100) due to the constantly moving plate boundaries in this region. Other regions of earthquakes are also generally around plate boundaries. Longitude is plotted on the x axis as they are the vertical lines that vary in the E-W direction which can be measured on the x axis. Latitudes are the horizontal lines that vary in the N-S direction which is measured on the y axis.

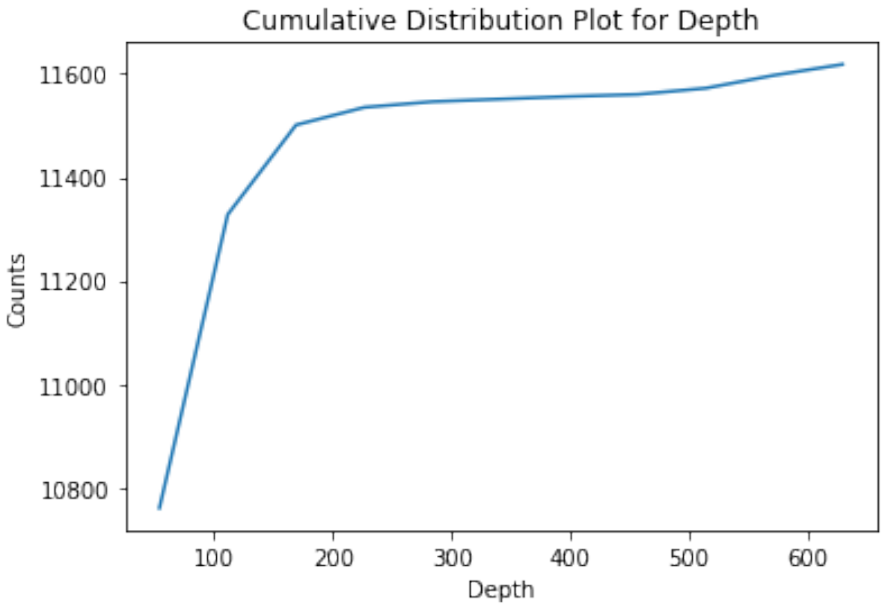

Figure 4: This plot shows that the depths in the range of 0-200 have the highest frequency as the slope in this region of the graph is the steepest while all greater values have considerably lower frequencies due to the relatively flat line.

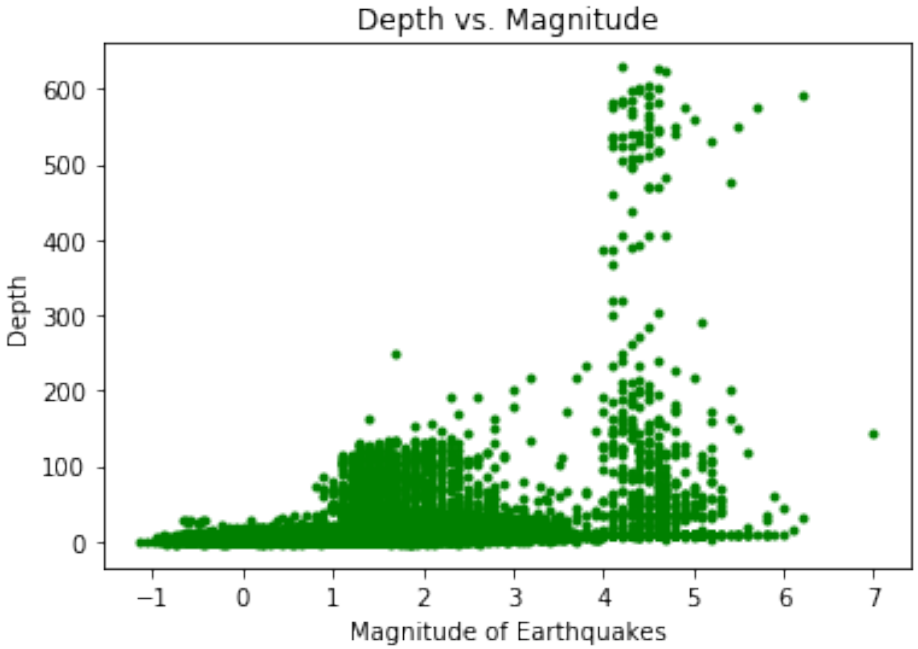

Figure 5: Generally, the lower magnitude earthquakes also have a lower depth (less than 200m) and the higher magnitude earthquakes have a higher depth (up to 650), however there doesn't seem to be a linear relationship between these points.

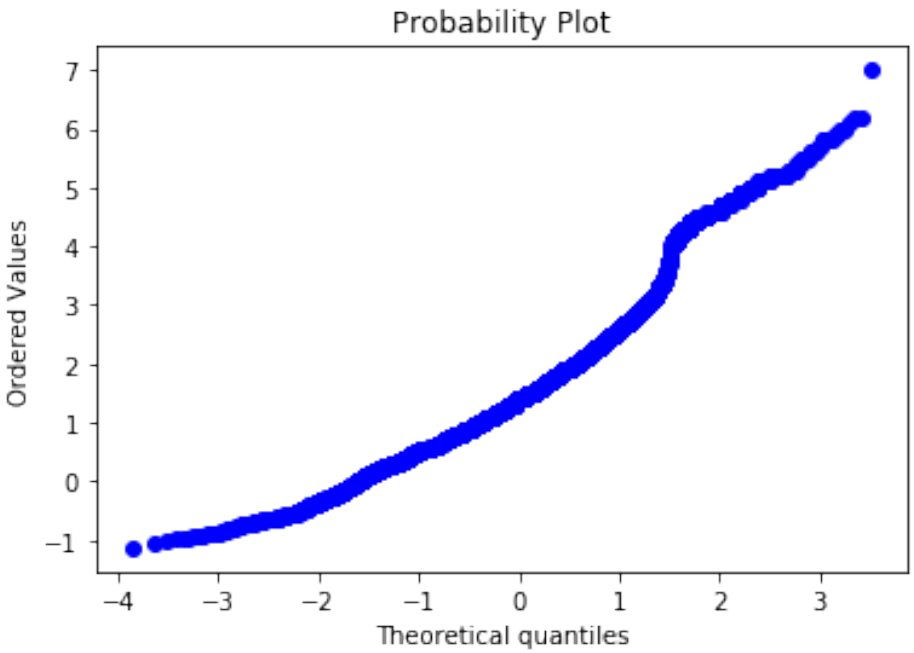

Figure 6: The probability plot assumes normal distribution and this dataset complies fairly well with the distribution.