Metadata File for ABE65100 Lab 07

Created March 20, 2020
by Miriam Stevens

This file summarizes the tasks conducted in Lab 07.
File “all_month.csv” was graphically analyzed using the script ”steve276_Program-07.py”.
All files related to Lab 07 can be found in the GitHub repository accessible at the following link:
https://github.com/Environmental-Informatics/07-graphical-analysis-with-python-steve276

**Source and format of input data**

Input data file name: all_month.csv
Downloaded at: 3/08/20 at 7:43pm EDT
File type: The input data is a CSV ASCII text file, size 2.1MB

The data was downloaded from the USGS Earthquake Hazards Program webpage for data in spreadsheet format: https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php

The Spreadsheet Format webpage is also accessible from the Real-time Notifications, Feeds, and Web Services page: https://earthquake.usgs.gov/earthquakes/feed/

The data includes earthquakes that happen worldwide during the last 30 days, updated every minute. There are twenty two types of data included. Full descriptions of the data types are available at the data source page.

**Types of analysis conducted by ‘steve276_Program-07.py’**

The script analyzes the location and characteristics of earthquakes recorded in the USGS dataset “all_month.csv”. The pandas library is used to read the input data into a DataFrame from which six plots are generated. The matplotlib library is used to generate the various plots. A histogram and KDE plot are used to show the distribution of earthquake magnitudes. A scatterplot is used to visualize where earthquakes occur. A cumulative density plot of magnitudes shows how likely it is for a certain size earthquake to occur. Another scatter plot reveals the relationship between earthquake depth and magnitude. The final plot uses the statsmodels module to create a quantile plot comparing the distribution of magnitudes to a normal distribution.

**genfromtxt()**

The pandas read_table() function is used instead of genfromtxt() to convert the data to a DataFrame. The numpy function genfromtxt() does not work with this dataset because the delimiter is not the same throughout the file.

**Figures generated in graphical analysis**

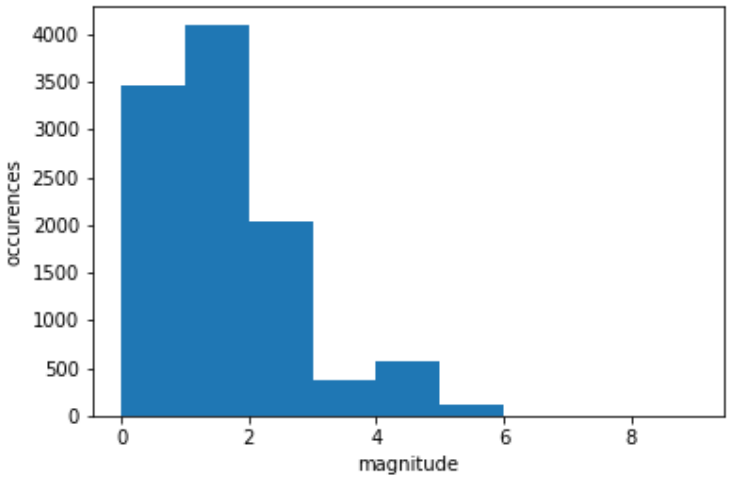


**Figure 1.** Histogram of Earthquake Magnitude.
More bins are auto-generated when the bounds of the range are closer to the range of the data, provided a specific number or size of bins is not defined. Having a greater number of bins reveals more details about the data. Selecting range [0,10] excludes the negative magnitude values. The histogram suggests the distribution is skewed to the right.

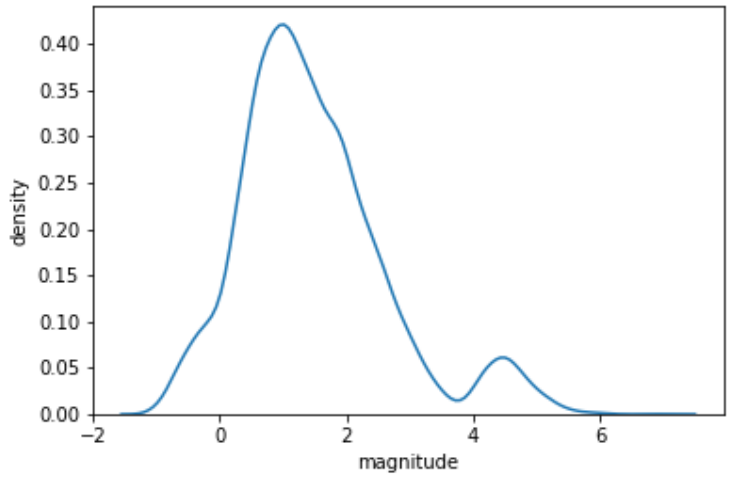


**Figure 2.** KDE of Earthquake Magnitude.
Kernel type is Gaussian and kernel width uses Scott's rule. The KDE plot shows a similar but smoother looking distribution than the histogram. A major difference between the two plots is that the histogram shows the number of occurrences on the y-axis, but the KDE shows the probability density of the x-axis values.

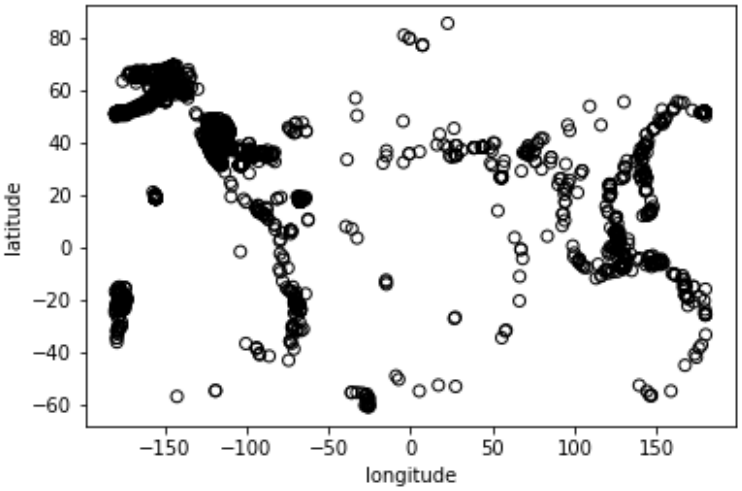


**Figure 3.** Latitude vs. Longitude Scatter Plot.
The points are clustered around the locations of crustal plates. Putting longitude on the x-axis allows the data to be viewed in the same orientation as most maps.

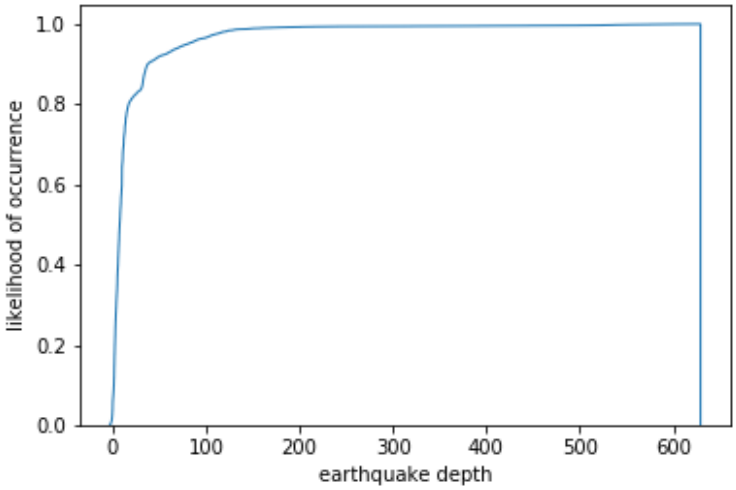


**Figure 4.** CDF of Earthquake Depth.
The plot indicates that over 80% of earthquake depths are less than 50km.

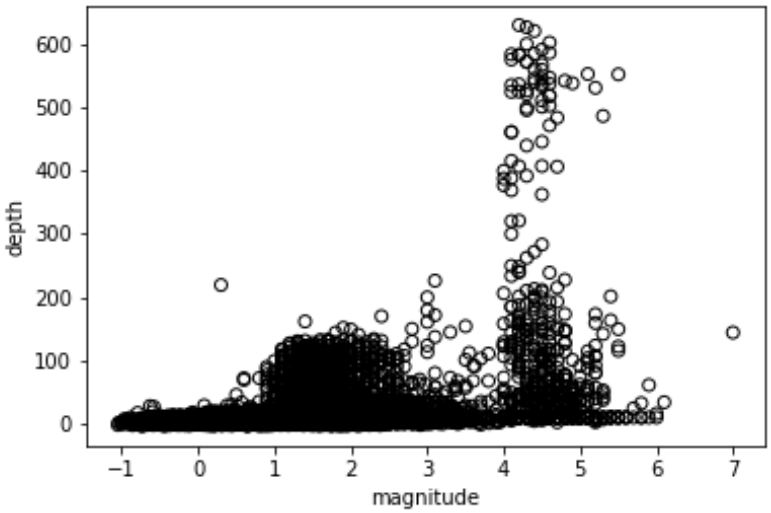


**Figure 5.** Depth vs. Magnitude Scatter Plot.
The magnitude of earthquakes is commonly between 1-3 or 4-5. The depth of earthquakes at these common magnitudes is concentrated between 50 and 150 km. Additionally, earthquakes with a magnitude between 4-5 exhibit a wide range of depths, concentrated between 50-150km or 500-600km. There are also many shallow earthquakes of all magnitudes.

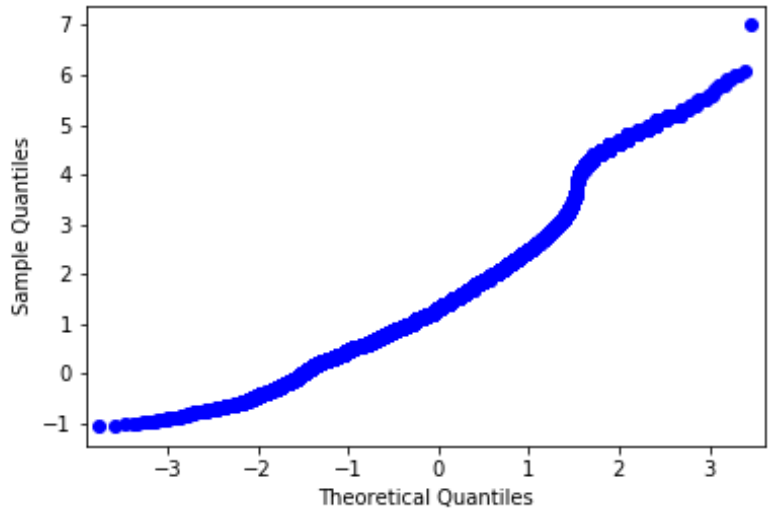


**Figure 6.** Quantile Plot of Earthquake Magnitudes.
The plot assumes a normal distribution and the data does not comply with this distribution.