Assignment 2- Pandas

In this assignment we will use pandas to examine earthquake data.

Start by importing pandas, numpy and matplotlib.

I saved you some time by pre-downloading some data in .csv format from the [USGS Earthquakes Database](https://earthquake.usgs.gov/earthquakes/search/). It is located at:

[**http://www.ldeo.columbia.edu/~rpa/usgs\_earthquakes\_2014.csv**](http://www.ldeo.columbia.edu/~rpa/usgs_earthquakes_2014.csv)

You don't even need to download it. You can open it directly with Pandas.

1) Use Pandas' read\_csv function directly on this url to open it as a DataFrame

(Don't use any special options). Display the first few rows and the DataFrame info.

You should have seen that the dates were not automatically parsed into datetime types.

2) Re-read the data in such a way that all date columns are identified as dates and the earthquake id is used as the index

Verify that this worked using the head and info functions.

3) Use describe to get the basic statistics of all the columns

Note the highest and lowest magnitude of earthquakes in the databse.

4) Use sort\_values to get the top 20 earthquakes by magnitude

Examine the structure of the place column. The country information seems to be in there. How would you get it out?

5) Extract the country using Pandas [text data functions](https://pandas.pydata.org/pandas-docs/stable/text.html)

Add it as a new column to the dataframe. (Is it really just country? No, some rows have the name of a US state.)

6) Find the 10 states / countries with the highest *number* of earthquakes

7) Find the top 10 states / countries where the strongest and weakest earthquakes occured

It looks like US states are being treated differently from foreign countries. We would like to fix that.

How can we tell if a name is a US state name? Python has a package for that: <https://pypi.python.org/pypi/us!>

This is a good time to try installing a new package using pip. Pip is the original python package manager that predates conda. Basically conda is more oriented towards data science while pip is more general purpose. There are lots more packages on pip than on conda. You can read a [comparision](https://www.quora.com/What-is-the-comparison-among-conda-vs-pip-vs-anaconda) of these two utilities if you want to know more.

8) Install the us package using pip, either directly from the notebook or the command line

The shell command is pip install us.

9) Import the us package to verify your installation works

10) Read the [us documentation](https://pypi.python.org/pypi/us) to figure out how to create a list of state names (all upper case)

11) Write a function to check whether a string is a US state name.

This function should not be case sensitive. It should also strip any whitespace out of the test string.

12) Use Pandas' [apply](https://pandas.pydata.org/pandas-docs/stable/generated/pandas.Series.apply.html) function to apply this to each row of the country name series

You should produce a new boolean series. (Hint: use dropna() to drop rows that have null values before calling apply.)

13) reindex this boolean series to match the dataframe's index

Fill the null values with False using .fillna().

14) Now re-assign the country column in the DataFrame to USA if the row is a state.

Also add the state name as a new column.

15) Now redo the country count and minimum magnitdue using the corrected data

16) Create a filtered dataset that only has earthquakes of magnitude 4 or larger

17) Analyze the distribution of the Earthquake magnitudes in the filtered distribution

Make a histogram of the Earthquake count versus magnitude. Make sure to use a Logarithmic scale. What sort of relationship do you see?

fig, ax = plt.subplots() df\_filt.hist('mag', bins=20, ax=ax) ax.set\_yscale('log')

18) Visualize the locations of earthquakes by making a scatterplot of their latitude and longitude.

Use the filtered data. Color it by magnitude. Make it pretty

19) Do the same for depth[¶](https://rabernat.github.io/research_computing/assignment-7-pandas.html#19)-Do-the-same-for-depth)