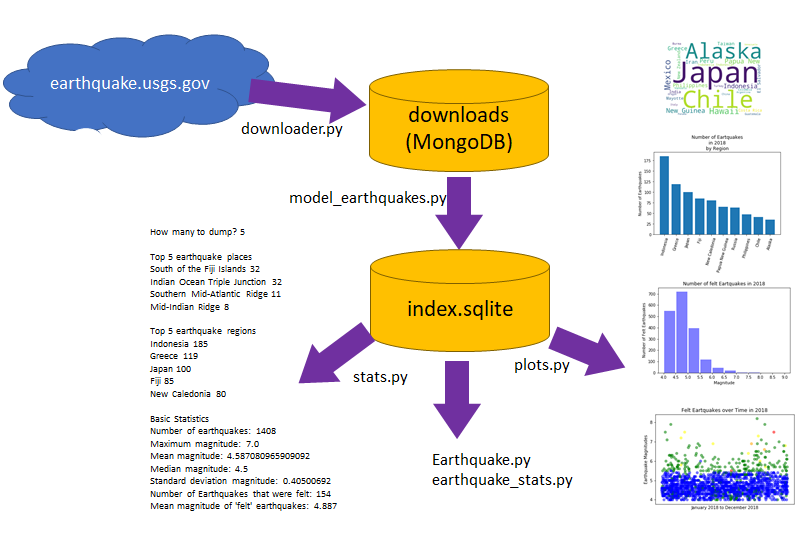
**Assignment: Data Analysis and Visualization**

* This assignment is the third part a a project where you use earthquake data from http://earthquake.usgs.gov/ to understand earthquake trends.
* In this part of the assignment you will run  an analysis to calculate some basic statistics about the data.
* You will then create at least three different types of visualizations of the earthquake data you have retrieved and processed and save some plots. For example:
  + Either a word cloud to or a bar chart to visualize the frequency distribution by region.
  + A histogram of the magnitudes of "felt" earthquakes in the data set.
  + A scatter plot to show how the data is changing over time.
* You will be developing **stats.py** and **plots.py**for this part of the assignment.

**Project Structure:**



**stats.py**

The first, simplest data analysis is to do a simple analysis of the data to answer the following questions:

* How many earthquakes are in the data set?
* What "place" has the most earthquakes?
* What "region" has the most earthquakes?
* What was the "biggest" earthquake int he data set?
* What is the mean earthquake magnitude in the data set?
* What is the median earthquake magnitude in the data set?
* What is the standard deviation of the earthquake magnitudes in the data set?
* What are the same basic statistics for earthquakes filtered by either felt, tsunami or depth?  
  e.g felt > 0, tsunami > 0, depth < 50 or depth<100.

**Code:**

# -\*- coding: utf-8 -\*-

"""

Python program to calculate basic statistics about a sample of earthquakes

Created on Sat Dec 15 10:00:02 2018

@author: dawng

"""  
  
# add necessary import statements

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# ADD input validation to make sure an integer between 1 and 20 is provided  
# Program should repeat request for a number OR exit if invalid input is provided  
howmany = int(input("How many earthquake locations to show? "))

# connect to the output database and name it index.sqlite   
conn = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
# forces database to return strings for TEXT attributes   
conn.text\_factory = str   
# get the cursor for the connection   
cur = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# run select query on database to get data for statistics

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# set up variables to hold database data  
# will be counting earthquakes by place/region  
# will just be storing data about magnitude, felt and tsunami   
placecounts = dict()

regioncounts = dict()

mag = []

felt = []  
tsunami = []  
depth = []  
for quake in cur :  
 # Get data from cursor and add to lists or dictionary  
 # Dictionaries will hold counts, Lists will append data  
 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Print out top earthquake places  
print('')

print('Top',howmany,'earthquake places')

x = sorted(placecounts, key=placecounts.get, reverse=True)

for k in x[:howmany]:

print(k, placecounts[k])

if placecounts[k] < 10 : break

# Print out top earthquake regions  
print('\nTop',howmany,'earthquake regions')   
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

print("\nBasic Statistics")

# Print out basic statistics like number of earthquakes, mean magnitude, median, standard deviation etc.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
# Convert the "felt/tsunami/depth" list to a NumPy Array

npfilter = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
# Create a NumPy array of Booleans where felt or tsunami" > 0 or depth < 50  
filterB = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
# Create a NumPy array of magnitudes for all of the filtered earthquakes

magF = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
# Print out basic statistics like number of felt earthquakes, mean felt earthquake ...   
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**plots.py**

* **plots.py** will be VERY similar to **stats.py**.
* You would start with **stats.py**, save as **plots.py**
* Add an import for pyplot
* Get data from the database and add it to dictionaries and/or lists
* Create at least three different types of plots using the ideas below:
  + Create a dictionary for earthquake regions but instead of printing out the regions, plot the top "howmany" regions, show them on a bar chart and/or a WordCloud
  + Create a histogram and/or scatter plot for "felt" earthquake magnitudes.
  + Create a histogram and/or scatter plot for all earthquake magnitudes
  + Create a histogram and/or scatter plot for "tsunami waning" earthquake magnitudes.
  + Create a histogram and/or scatter plot for felt earthquake depths.
  + Create any of the above histograms and/or scatter plots for a specific region (especially useful if you are tracking earthquake activity around a specific major quake)
* Create a title, xlabel and ylabel for each of your plots (except for a wordcloud)
* Call savefig for each of your plots and save as a png.

**Submitting Your Work**

**Please Upload Your Submission:**

* Submit either:
  + A screen shot of you running the stats.py program to compute basic histogram data and statistics on the messages you have retrieved.
  + A saved .png file each of the plots you made with **plots.py**.
* Or
  + Jupyter Notebook .ipynb file(s) that includes the code and shows the output statistics and plots.
* AND a zip file containing all of the .py, .ipynb and .sqlite files you created/used as a part of the assignment.
* **Data Visualization**

| Data Visualization | | |
| --- | --- | --- |
| **Criteria** | **Ratings** | **Pts** |
| This criterion is linked to a Learning OutcomeThe updated stats.py file that displays the top regions AND places where earthquakes have occurred - with appropriate error checking on the input number. | |  |  |  | | --- | --- | --- | | **12.0 pts**  **Full Marks** | **4.0 pts**  **Mostly correct but missing some required code or formatting.** | **0.0 pts**  **No Marks** | | 12.0 pts |
| This criterion is linked to a Learning OutcomeThe updated stats.py file that computes statistics for all earthquakes AND statistics for earthquakes filtered by either felt, tsunami or depth | |  |  |  | | --- | --- | --- | | **13.0 pts**  **Full Marks** | **7.0 pts**  **Missing some important program elements.** | **0.0 pts**  **No Marks** | | 13.0 pts |
| This criterion is linked to a Learning OutcomeA screen shot/.ipynb file showing the output of you stats.py program. | |  |  |  | | --- | --- | --- | | **15.0 pts**  **Full Marks** | **8.0 pts**  **Runs but missing some important program elements.** | **0.0 pts**  **No Marks** | | 15.0 pts |
| This criterion is linked to a Learning Outcomeplots.py edited to produce at least 3 different types of visualizations. | |  |  |  | | --- | --- | --- | | **10.0 pts**  **Full Marks** | **5.0 pts**  **one or more plots are submitted but they are not different types of visualizations** | **0.0 pts**  **No Marks** | | 10.0 pts |
| This criterion is linked to a Learning OutcomeVisualizations set correct data, titles and axis labels and are saved correctly. | |  |  |  | | --- | --- | --- | | **10.0 pts**  **Full Marks** | **5.0 pts**  **Visualizations missing some data, titles or axis labels or were not saved correctly.** | **0.0 pts**  **No Marks** | | 10.0 pts |
| This criterion is linked to a Learning OutcomeA screen shot, downloaded png, or .ipynb file showing 3 different types of visualizations based on the earthquake data you retrieved. | |  |  |  | | --- | --- | --- | | **20.0 pts**  **Full Marks** | **10.0 pts**  **One or two plots submitted** | **0.0 pts**  **No Marks** | | 20.0 pts |
| Total Points: 80.0 | | |