DSC 540 Final Project

Name: Katie Briggs   
Title: Earthquakes  
  
**Milestone 1**

**Introduction**

Earthquakes happen every day and we may not feel them until their magnitude becomes greater than 3. When I lived in Alaska from 2006 to 2009, we had on average 13 earthquakes a day. However, most of them were between 2 and 4 on the Richter Scale. I was so worried about earthquake damage to my home that I began visiting U.S. Geological Survey’s website everyday and this is where I started my fascination with earthquakes. For this term project I will gather information from three different sources of data regarding earthquakes around the world.

**Research questions**

Questions that I will be asking myself to find insights into earthquakes:

1. What percentage of the world earthquakes happen in the United States?
2. Do earthquakes happen more frequently around active volcanic areas?
3. At what magnitude and depth does a Tsunami occur when an earthquake happens.
4. What trends, if any, are among weather cycles when earthquakes happen?
5. Is it possible to look for patterns to know when and where the next large earthquake will happen?

**Data**

* **Data source 1:** API from the U.S. Geological Survey (USGS). This data set is located on <https://earthquake.usgs.gov/fdsnws/event/1/> I was able to pull data from the site and came up with 22 columns and well over 1000 rows. The variable are:
  + Time– data and timestamp of earthquake.
  + Depth – depth of the earthquake.
  + Mag – The magnitude of the earthquake.
  + MagType – ML or MD
  + Id – ID for each earthquake.
  + Type – What cause the seimeic activity. Quarry blast, fracking, earthquake
  + Location – Where the earthquake occurred.
  + Latitude
  + Longitude
  + Nst
  + Gap
  + Dmin
  + Rms
  + Net
* **Data source 2:** CSV data file of Significant Global Earthquake available on Kaggle at <https://www.kaggle.com/mohitkr05/global-significant-earthquake-database-from-2150bc>. This file contains 47 columns, which have the same variables that the USGS website does. I will be adding:
  + Tsunami – Did one occur because of an earthquake. Recoding to 0 for no and 1 for yes.
* **Data source 3:** Website, <https://www.ngdc.noaa.gov/nndc/struts/form?t=101650&s=1&d=1>, National Geophysical Data Center / World Data Service (NGDC/WDS). This website has the same columns as Data source 1 and 2. On this source I will be adding:
  + Volcanic Eruption – Did one occur because of an earthquake. Recoding to 0 for no and 1 for yes.

**Approach**

I believe that I will have to create relationships between these data sets to complete all milestones. I would like to merge the data together and avoid having duplicates. How to get there, I am hoping this class will give me the tools to get through every milestone.

Reference/Resource:

<https://www.kaggle.com/mohitkr05/global-significant-earthquake-database-from-2150bc>

<https://earthquake.usgs.gov/fdsnws/event/1/>

<https://www.ngdc.noaa.gov/nndc/struts/form?t=101650&s=1&d=1>,