

Plot Those Earthquakes!

David Franklin, TheProgrammersCabin.com, Litchfield, NH

ABSTRACT

When major earthquakes occur, whether it is in Japan, New Zealand or our own shores, information hits the news wires in minutes. But those that are reported are only the earthquakes where there is measurable damage or big events, however many go being unreported.

The United States Geological Survey (USGS) records earthquakes throughout the world on a near real-time basis. There are many tools out there that will plot these events, however, using SAS, not only can we take this data and plot it, giving an indication of its size, but it is also possible to take this data and do predictive analysis on it, something that SAS is very good at.

This paper will first look at where and how you get the earthquake data from the USGS, and then follow this up with a SAS program that is able to read this file and plot the location of earthquakes on a map, indicating their size, and also do rough calculations for time and distance between two events. Along the way we will learn about some of the maps available in SAS that can be used to plot the data, longitude and latitude coordinates, and how the data can be subset using this coordinate system, i.e. look at earthquakes specific to an area.

INTRODUCTION

Earthquakes are a natural event on our planet and occur when there is a sudden release of energy in the Earth's crust that creates seismic waves. There are a number of organizations around the world that record information such as time, location and magnitude, one the largest being the USGS (U.S. Geological Survey). The USGS themselves supply this information in a near real-time feed that can be used by SAS to plot these events.

THE DATA

The USGS has a number of datasets that are available for public use – these can be found at <http://earthquake.usgs.gov/earthquakes/feed/>. The format of these datasets are in the form of:

- ATOM datasets
- CSV files
- GeoJSON

For the purposes of this paper, CSV files will be used.

Another consideration is the length of time that you may want to have in your analysis, e.g. within the past hour, past day, past 7 days or past 30 days. Within these datasets are yet another subset where magnitude may be done for you. For the purposes of this paper we are going to look at the CSV dataset for earthquakes in the past 30 days.

THE CSV FILE

The CSV file has the following variables:

- Src -- source
- EqId – an ID value for the earthquake
- Version – as more information is known about an event, the data may be updated, hence updated versions
- Datetime – date/time of the event (GMT)
- Lat -- latitude
- Lon -- longitude
- Magnitude – size of event, usually in the R
- Depth – estimate of depth of event in kilometers
- NST -- Number of seismic stations which reported P- and S-arrival times for the event
- Region – a general note of the geographic region where the event occurred

An example of the output is given below:

```
Src,EqId,Version,Datetime,Lat,Lon,Magnitude,Depth,NST,Region
us,b000buag?, "Friday, August 10, 2012 18:37:44
UTC", 52.6954, -167.4686, 6.2, 18.96, ?, "98km ESE of Nikolski, Alaska"
```

```
ak,10532503,?, "Friday, August 10, 2012 18:17:43 UTC",51.6187,177.2736,3.2,45.8,?, "58km  
SSW of Kiska Volcano, Alaska"  
ci,15190521,?, "Friday, August 10, 2012 16:37:33  
UTC",32.2158,-115.2908,2.7,41.4,?, "11km W of Alberto Oviedo Mota, Mexico"
```

The following is the code that will load the file into a SAS dataset:

```
filename mydata url 'http://earthquake.usgs.gov/earthquakes/feed/csv/2.5/month';  
data quake0;  
  length Src $10  
         EqId $20  
         Version $10  
         Datetime $50  
         Lat 8  
         Lon 8  
         Magnitude 8  
         Depth 8  
         NST $20  
         Region $100;  
  infile mydata dsd dlm="," firstobs=2;  
  input  Src $ EqId $ Version $ Datetime $ Lat Lon Magnitude Depth NST $ Region $;  
run;
```

Note that the FILENAME statement gets the data directly from the USGS web page.

The date and time can be extract from the DATETIME variable using the following code:

```
date=input(scan(dt,2,' ')||' '||scan(dt,3,' ')||' '||scan(dt,4,' '),ANYDTDTTE32.);  
time=input(scan(dt,5,' '),HHMMSS8.);
```

THE WORLD MAP

Anyone who has tried to use SAS map datasets know that the projections that SAS has are not good, even after SAS introduced the GALL procedure in SAS 9.1.3. The main problem with the SAS provided World Maps is that there is a known issue where Deprojected longitude and latitude values might not line up with map longitude and latitude coordinates, refer to SAS Usage Note 40034. This was later confirmed when the initial plot came up with wildly incorrect placement of known earthquakes.

To solve this issue, shoreline co-ordinates around the world are available from the National Oceanic and Atmospheric Administration (NOAA), National Geophysical Data Center at <http://www.ngdc.noaa.gov/mgg/coast/>, a screen-shot of which is below with the values used to extract:

Coastline Extractor

[Use Java Map for lat/lon](#)**Geographic Range of Coastline to Extract**
(decimal degrees, west and south are negative)

upper latitude

westernmost longitude easternmost longitude

lower latitude

Note: Please enter the geographic range first, then choose criteria below.

If you select a range that covers more than a few degrees, be prepared to wait while the data are extracted and (optionally) the coastline image is created.

Coastline database: **Compression method for extracted ASCII data:** None GNU GZIP UNIX Compressed ZIP**Coast Format options:** GMT ASCII Arc/Info Ungenerate Matlab Splus Mapgen**Coast Preview options:** No preview GMT PlotExtract the Coast:

Note that the lowest resolution of shoreline was selected and has an output similar to that given below and saved in a file which was called WORLD.DAT:

```

>
0.19244    5.669954
-0.114995  5.515063
-0.391921  5.360171
-0.657114  5.195893
-0.887104  5.151303
-1.262597  5.062123
-1.417488  5.008146
-1.560645  4.874377
-1.736658  4.775809
-1.936139  4.68663
-2.198985  4.787544
-2.4102    4.874377
-2.5956    4.930701
-2.771612  4.951822
>
-2.771612  4.951822
-2.860792  4.963556
-3.003949  4.97529

```

Between each pair of '>' is the co-ordinate data that is the same as the variable ID on the SAS maps dataset.

Loading the map data used the following code:

```

data world;
  retain id 0;
  infile 'c:\temp\world.dat' length=len;
  input text $varying200. len ;
  if text='>' then id+1;
  else do;
    lon=input(scan(text,1,'09'x),best.);
    lat=input(scan(text,2,'09'x),best.);
    output;
  end;
  drop text;
run;

```

Now that the program has a world map that is usable with latitude and longitude co-ordinates, the next task was the plot.

NOW THE FUN STUFF, THE PLOT ...

Now that we have our earthquake and map data above, the next thing is to do the plot.

Common approaches to mapping data in SAS is to convert degrees to radians using either

```
radians = atan(1)/45 * degrees;
```

or

```

pi=constant('pi');
radians = pi /180 * degrees;

```

However, this is not needed since as long as you are using one unit of measure, SAS will deal with either degrees or radians – for the purposes of this exercise I will use degrees as both the Earthquake and World Map data use degrees.

Since we are going to do some graphics, lets set up the GOPTIONS that are going to be used:

```
goptions reset=all cback=white border htitle=12pt htext=10pt;
```

Now some setup of the data where subsets are possible. This is a little unorthodox but lets look at the following dataset:

```

data quake1 world1;
  set quake0 (in=a) world0 (in=b);
  /*Can subset data here by Longitude and Latitude here, as well as date and
  magnitude,
  However be aware that the date and magnitude values are only in the QUAKE0
  dataset*/
  x=lon;
  y=lat;
  if a then output quake1;
  if b then output world1;
run;

```

The dataset effectively brings in the earthquake and mapping data, places the longitude (variable LON) into a variable X, and also places the latitude (variable LAT) into a variable Y – PROC GMAP uses these X and Y variables.

It is also in this step that the quake data can be subset by a set of longitude, latitude, date and magnitude ranges if needed – explanation and examples of how to subset will follow.

Now that we have our subsets of map (dataset WORLD1) and earthquake (dataset QUAKE1), our data is just about ready for mapping. First, though we are going to setup the earthquake data into a form where the plots are going to be annotations on the map by using the following SAS code:

```

data quake2;
  length color $ 8 text $ 20 style $ 25;
  set quake1;
  retain xsys ysys '2' function 'label' when 'a';
  size=1;

```

```

style='special'; text='J';
select;
  when (Magnitude <= 2.5) color='red';
  when (Magnitude > 2.5 and Magnitude <= 5) color='blue';
  when (Magnitude > 5 and Magnitude <= 6.5) color='purple';
  otherwise color='green';
end;
output;
run;

```

Note that in our maps are going to use a dot, from the SAS FONT library, to signify where an earthquake occurred and we are going to use color to give an indication of size.

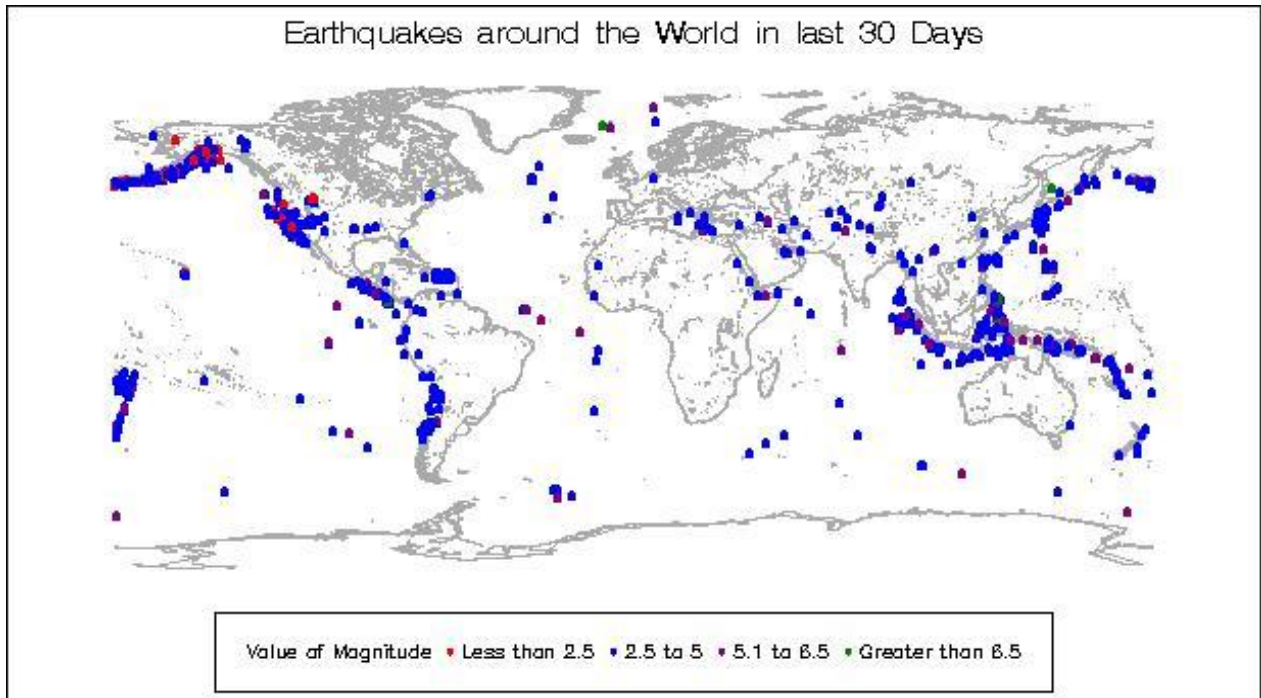
The last puzzle to our piece – set up a pattern, add a title and footer, then plot the map:

```

pattern1 v=me r=99;
title 'Earthquakes around the World in last 30 Days';
footnotel box=1
  f='SIMPLEX' c=black h=0.8 'Value of Magnitude '
  f=special c=red h=0.8 'J' f='SIMPLEX' c=black h=0.8 ' Less than 2.5 '
  f=special c=blue h=0.8 'J' f='SIMPLEX' c=black h=0.8 ' 2.5 to 5 '
  f=special c=purple h=0.8 'J' f='SIMPLEX' c=black h=0.8 ' 5.1 to 6.5 '
  f=special c=green h=0.8 'J' f='SIMPLEX' c=black h=0.8 ' Greater than 6.5';
proc gmap data=world1 map=world1;
  id cont id;
  choro id / anno=quake2 nolegend coutline=grayaa;
run;
quit;

```

An example of a map, using no subsets of the data, is shown below:



Data can also be subset easily by magnitude so that the output only produces those we are interested in.

WHAT IS LONGITUDE AND LATITUDE, SUBSETTING DATA BY LOCATION, AND CALCULATING DISTANCE

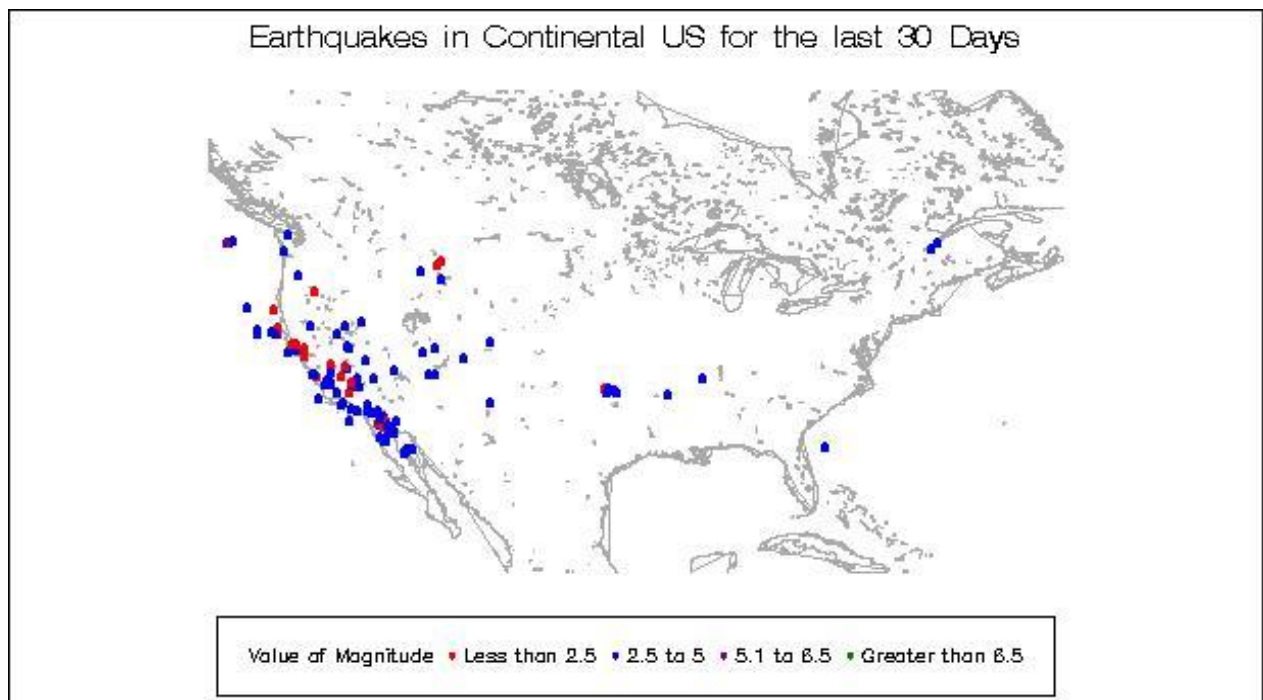
Longitude and Latitude is a common geographic coordinate system that enables every location on the Earth to be specified by a set of numbers and often a suffix of (E)ast and (W)est for Longitude, and (N)orth and (S)outh for

Latitude. It is also important to note that North and East values have positive values while South and West have negative values.

In the data from the USGS, the Longitude and Latitude co-ordinate system is used, and it is with this information that we can subset data by location. It is therefore necessary to know the Longitude and Latitude of the area in question in order to subset the data – this information can be found either by doing a simple Google search or going to specific web sites such as <http://www.findlatitudeandlongitude.com/>.

It is important to know that while there is a large number of earthquakes across the world, there may not be many in the region that you may want to subset the data by so it is best to subset the data by a large region.

As an example, if we were to look at Southern Canada and the Continental US, the rectangle in question is Latitude of 20° to 60° and Longitude of -130° to -60°. It is important to note that Latitude and Longitude must be in decimal degrees to match that used by the USGS. Also note that the data that is subset forms a rectangle and does not follow any recognized border.

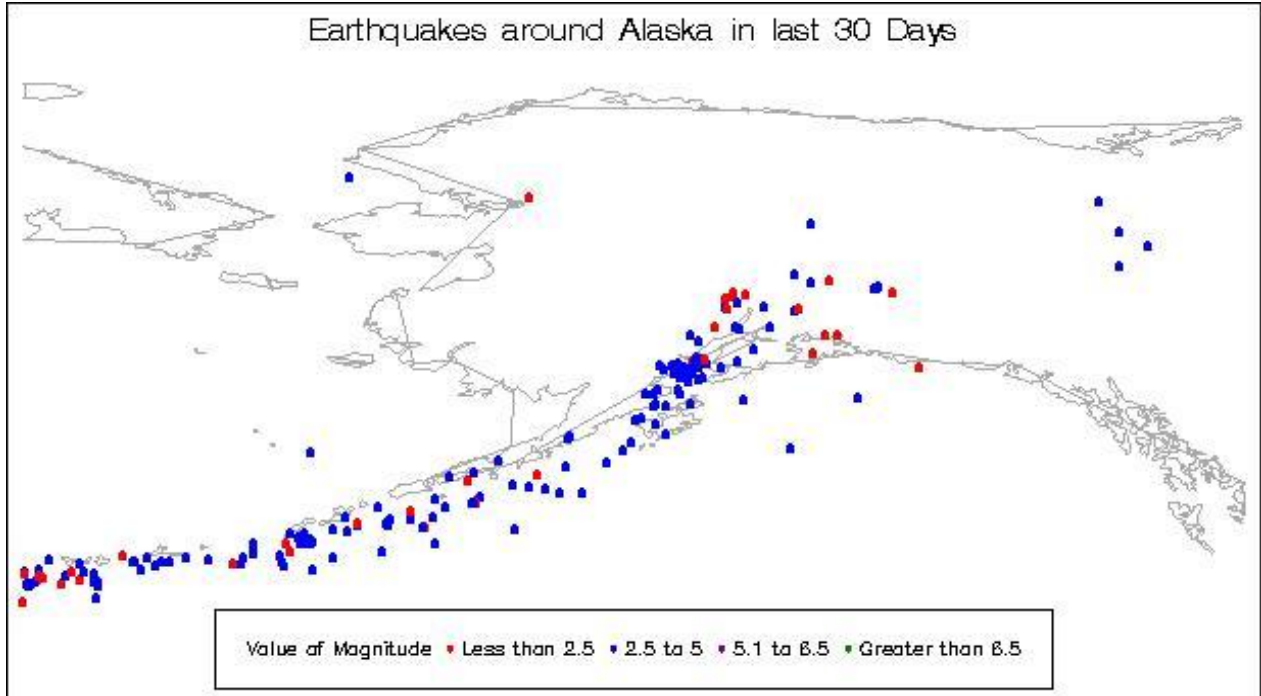


To do the actual subset of data the following two lines of code in the dataset that creates the QUAKE1 and WORLD1 datasets were added:

```
if lat>20 and lat<60;  
if lon>-130 and lon<-60;
```

Another example, where there is a history of a number of earthquakes, is Alaska, using the condition

```
if lat>40 and lat<90;  
if lon<=-130 and lon>=-180;
```



You will notice in this example that the contours of the shoreline are not exact – this is a fault of the data coming in from NOAA which appears to have at least two layers of data in the dataset, one being a very basic polygram approach and then a second layer which is more detailed. The WORLD0 dataset can be modified to correct this but this example shows the issues that may be encountered if you use 3rd party data for your maps. Unfortunately, the SAS maps, while more accurate, do suffer from incorrect longitude and latitude co-ordinates as discussed earlier.

Although no analysis in this paper of calculation distance is done, there is the following calculation that may be useful for doing this task:

$$d = 60 * \text{ARCOS} (\text{SIN}(l_1) * \text{SIN}(l_2) + \text{COS}(l_1) * \text{COS}(l_2) * \text{COS}(g_2 - g_1))$$

where

- l_1 = latitude at the first point (degrees)
- l_2 = latitude at the second point (degrees)
- g_1 = longitude at the first point (degrees)
- g_2 = longitude at the second point (degrees)
- d = computed distance (nautical miles)

The formula shown above is the Great Circle Distance calculation (the easiest) that treats the earth as a sphere and does assume that 1 minute of arc is 1 nautical mile. To convert the distance from nautical miles to kilometers, multiply the result d by 1.852. Similarly to convert the distance from nautical miles to standard miles multiply the result by 1.150779.

In SAS version 9.1.3, the procedure GEOCODE (calculates a longitude and latitude for a given address) and the function GEODIST (calculates a distance from one point to another point) were introduced in the SAS/GIS product but their description and usage is beyond the scope of this paper – refer to the relevant SAS documentation for details.

CONCLUSION

Using SAS and the data from the USGS, analysis and plotting of earthquakes is a highly powerful combination that works well.

REFERENCES

SAS Institute Inc. 2006. SAS/Graph® 9.1.3 Guide, Second Edition, Volumes 1 and 2. Cary, NC: SAS Institute Inc.

SAS Usage Note 40034: Deprojected longitude and latitude values might not line up with map longitude and latitude coordinates <http://support.sas.com/kb/40/034.html>

USGS Earthquake Data, <http://earthquake.usgs.gov/earthquakes/feed/>

Longitude and Longitude <http://www.findlatitudeandlongitude.com/>

National Oceanic and Atmospheric Administration (NOAA), National Geophysical Data Center <http://www.ngdc.noaa.gov/mgg/coast/>

Great Circle Distance formula http://en.wikipedia.org/wiki/Great-circle_distance

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name:	David Franklin
Enterprise:	TheProgrammersCabin.com
Work Phone:	603-275-6809
Email:	dfranklin@TheProgrammersCabin.com
Web:	TheProgrammersCabin.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.