Final Project

Blair Gemmer

CSCI 444 – Data Visualization

Fall 2011

Hypothesis

"Most earthquakes occur between the 36N and -36N lines of Latitude, with an exceptional number of occurrences between the -180W and -72W (Australia and the Americas); between -36W and 72W (Scandinavia, Europe, and Russian); and between 108W and 180W lines of Longitude (Japan, Asia, and Polynesia)."
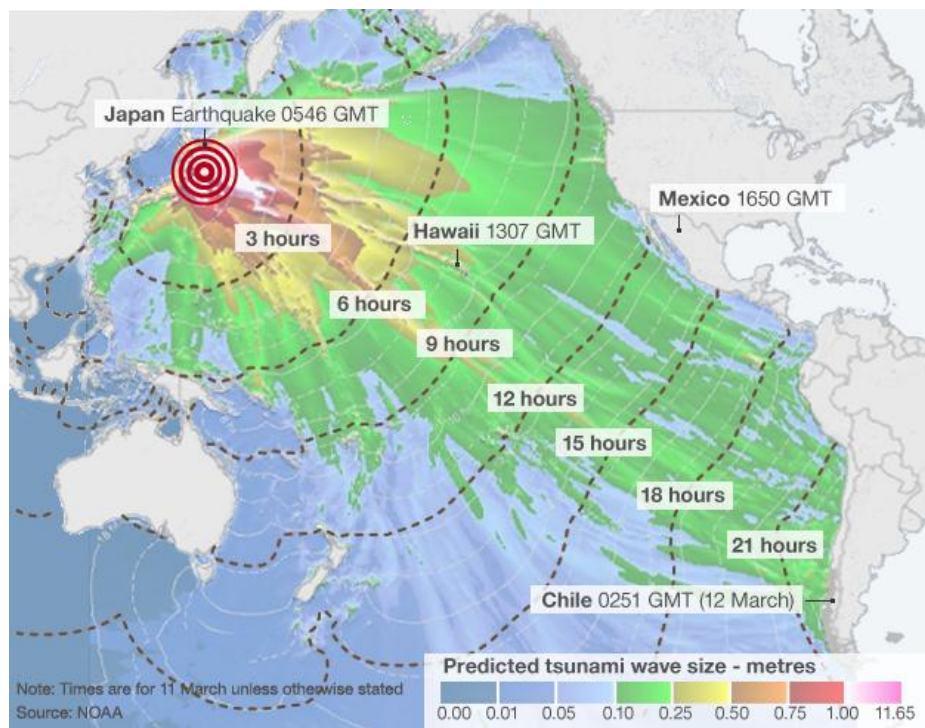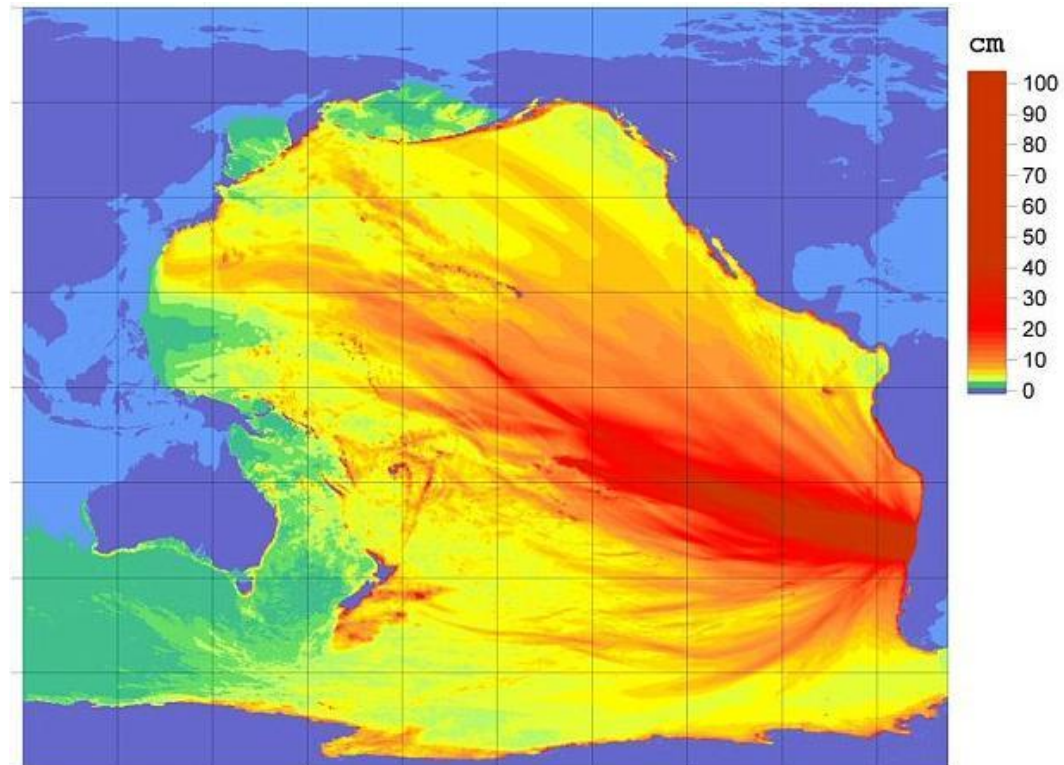
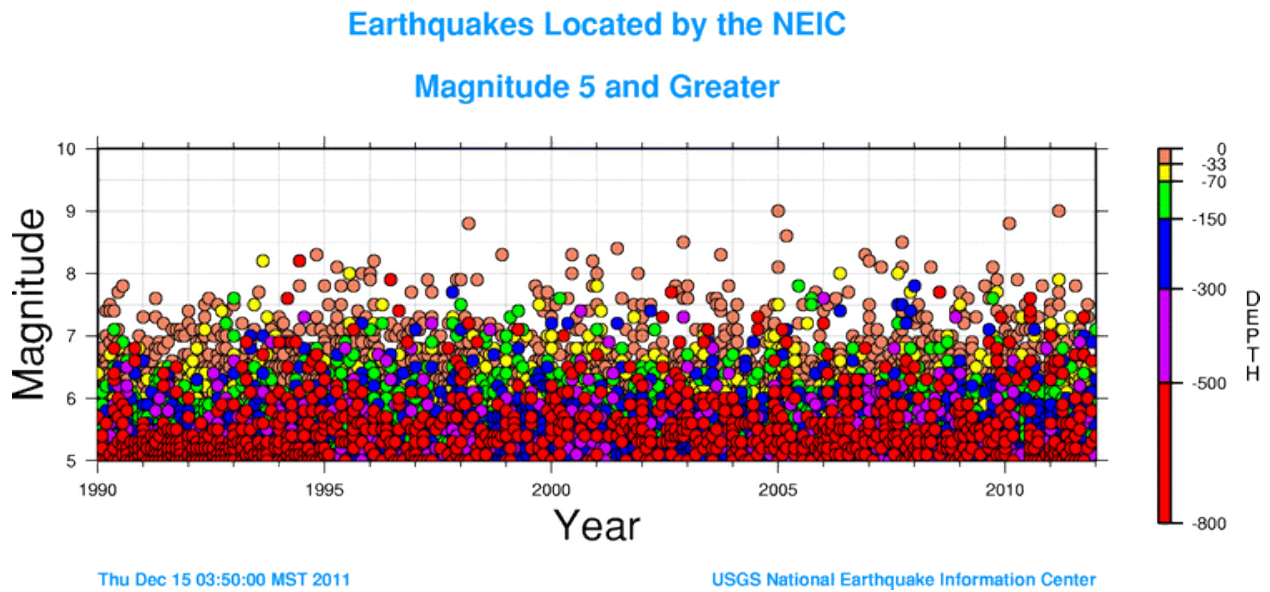## Good Visuals

Fig. 1

Fig. 2
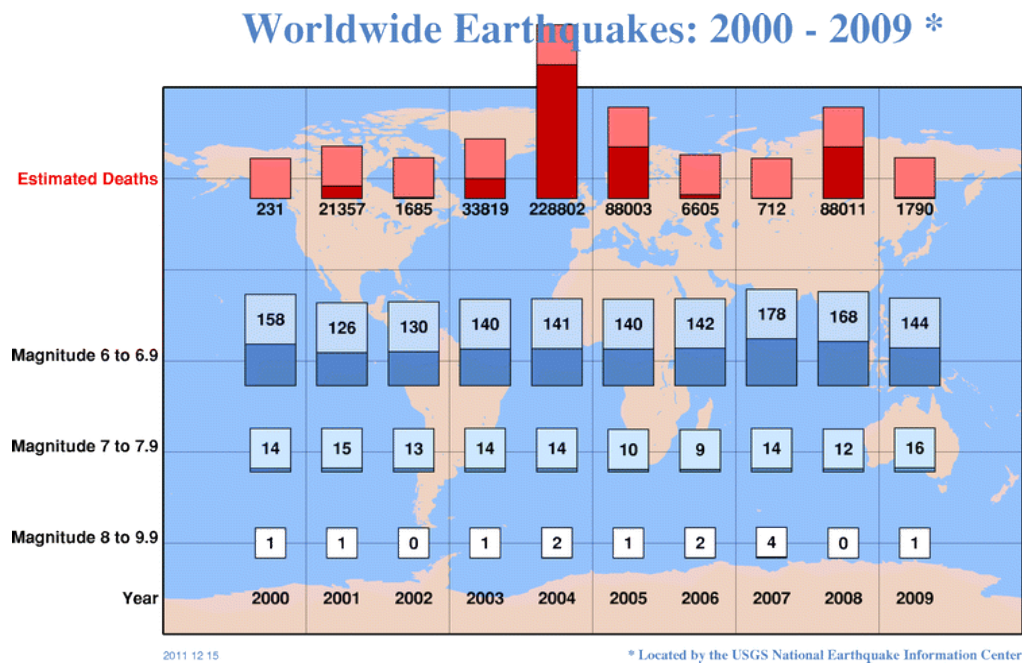
Bad Visuals

Fig. 3

Fig. 4

Visual Explanation

I found the good visuals (figs. 1-2) searching Google for "earthquake graph". I came upon a website called The Graph Bank [1] that had a few good earthquake graphs. I noticed that one of these had been used by a previous student, but opted to submit the other two that I considered good examples. The first (fig. 1) is the 2011 Japan (Sendai earthquake) tsunami predictions. The second (fig. 2) is the predicted height of the tsunami from the 2010 Chili earthquake.

I felt like these were good graphs because they limited the non-data ink by not including a bunch of garbage text. They included a good color map that was easily understood by the reader and showed exactly the information they intended to portray.

I got both of the bad visuals (figs. 3-4) ironically from the USGS National Earthquake Information Center website [2]. I felt like these were bad examples because the scatter plot (fig. 3) points overlapped in such a way that it is very difficult to read the graph and what its information is telling us. The bar chart (fig. 4) was bad for many reasons including non-data ink (the globe which holds no information), the random placement of the bar charts, and the bar for deaths running off the page. I also noticed that the bars have a darker portion which is unexplainable based on the graph information. This information may have best been portrayed using some sort of timeline plot or a simpler bar graph.

1973-05-20
14:20:33
51.701N, -176.684W

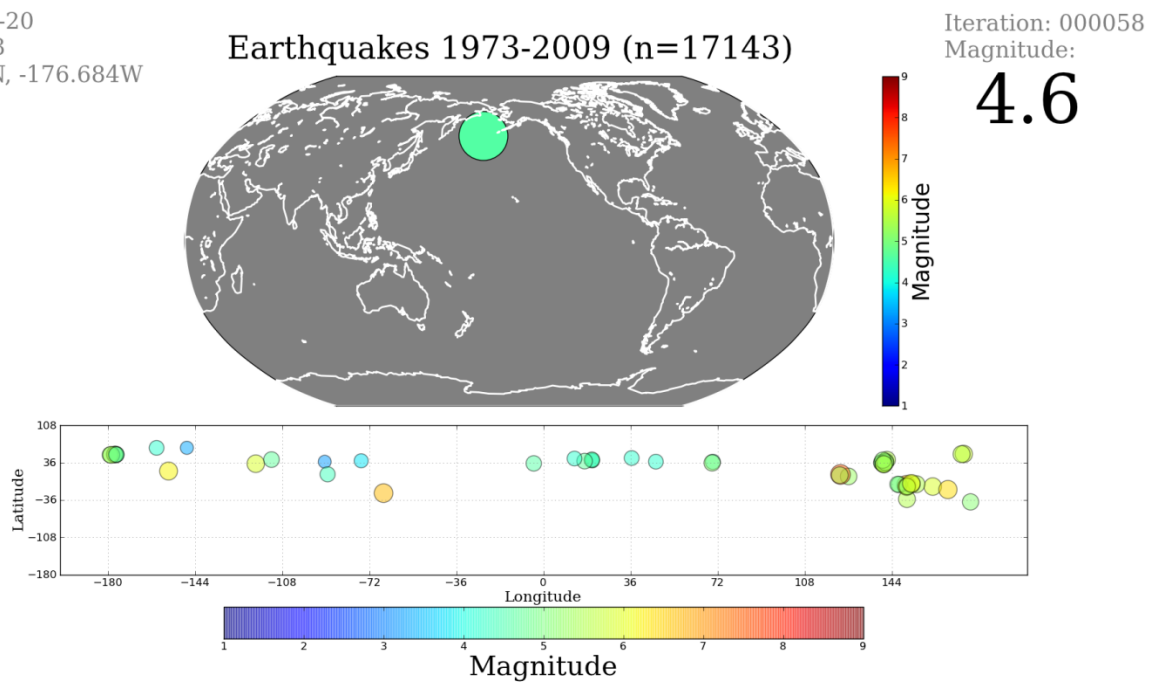## Earthquakes 1973-2009 (n=17143)

Iteration: 000058
Magnitude:

4.6

Fig. 5

1986-07-28
20:29:02
-33.282N, -71.934W

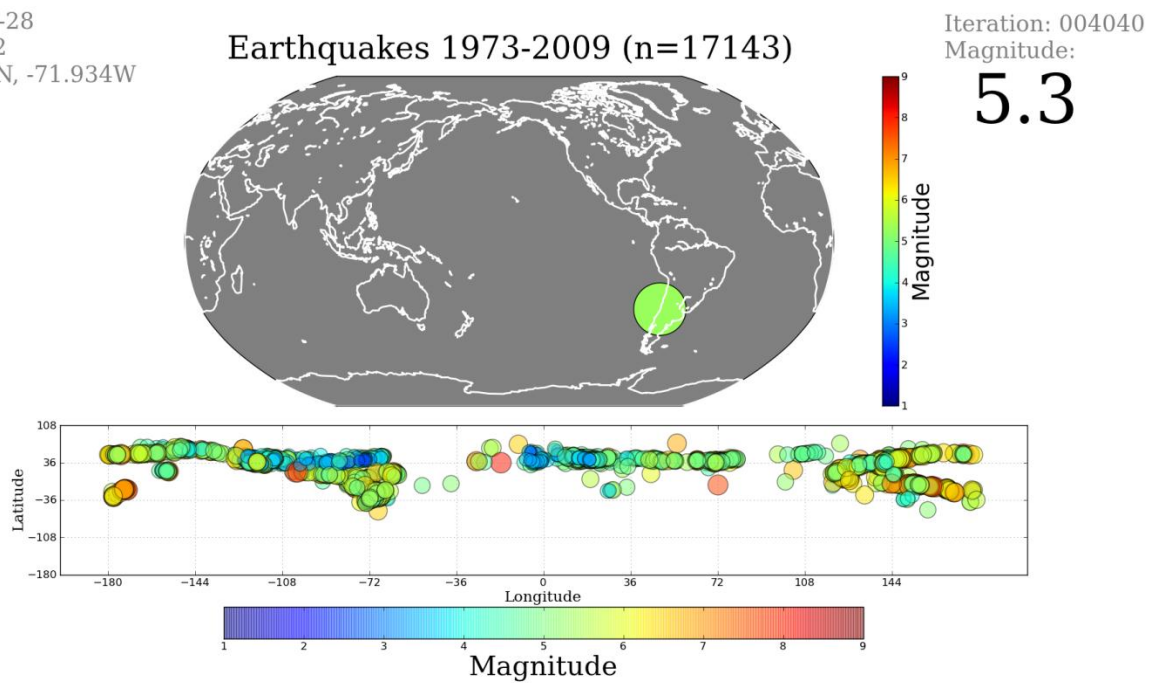## Earthquakes 1973-2009 (n=17143)

Iteration: 004040
Magnitude:

5.3

Fig. 6

2011-03-11
05:46:00
38.322N, 142.369W

Worst Earthquakes Recorded 1902-2011 (n=87)

Iteration: 000087
Magnitude:

9.0

Magnitude

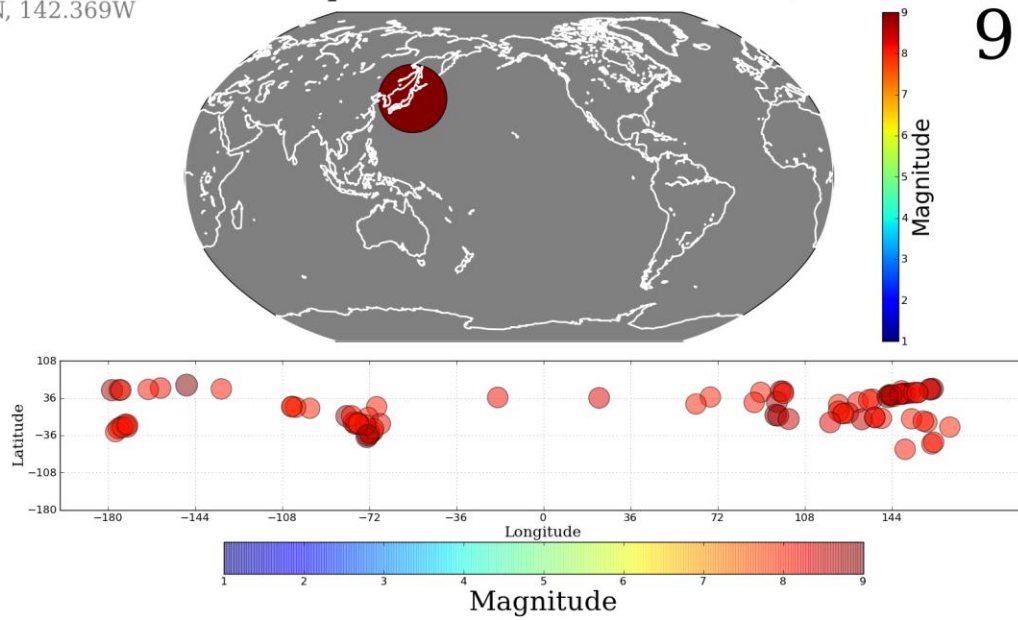Latitude

Longitude

Magnitude

Fig. 7

Visual Explanation

My original intentions for this project were to create a movie of all the earthquakes between 1973 and 2009 and create a sound file for each quake to synthesize a song utilizing the magnitude information from that quake. Unfortunately, I ran into problems with recording audio in Python. I had no problems recording audio and no problems creating audio, but creating a chunk of frequencies and then writing those frequencies to a .wav file turned out to be a gigantic hassle, coupled with the hassle of editing the sound files and putting them into a movie turned it into a much larger task than I was willing to tackle in the last few weeks of class.

I turned my talents, instead, to the art of creating a good video visual utilizing the magnitude data from my earthquake dataset. I grabbed the earthquake dataset from the USGS website [3], after doing a search for 1973-2009 worldwide earthquake data. I was able to find a total of 17143 earthquakes after refining my search a little bit. Unfortunately, I got quite a few magnitudes without any values, so I had to filter through the missing data when parsing the CSV file.

After I plotted the first few earthquakes and got a color map that I liked working, I worked on formatting the annotation and playing with different parameters [4]. I had to format the .png file names and date time to maintain preceding zeros. I formatted the coordinates and date time to be less obstructive to the visual. I also formatted the magnitude to be more noticeable in the visual. After all the annotation and titles were complete, I adjusted the font-family to "serif".

After creating the basemap plot, I decided to add another subplot to show the scatter of the magnitude distributions over the coordinates. This also served to help me come up with a hypothesis, as I wasn't able to see much of a conclusion from the original visual. I also had to figure out how to adjust the subplot [5] to get everything exactly where I wanted it in the visual. I ended up using subplot2grid [6] to create a table for my subplots. I finished up the scatter plot by using an alpha channel on the points so they can overlap and show some sort of color variation based on region. I included a key to show the different regions as I figured out where they existed on the globe. The key shows a rough overlapping of the areas I denoted on the globe. This is only a rough idea of what's going on in the scatter plot, but it helps to illustrate exactly what my visual is showing.

The total rendering time of all of the 17143 individual plots was approximately 7 hours. I wrote a piece of the script to catch if it had already rendered some .png file so it wouldn't repeat itself in case the machine shut off during rendering. This helped save a ton of time if I needed to stop the script and rerun it.

I decided to create a short video of the worst earthquakes recorded between 1902 and 2011 as well, to strengthen my argument for where the most/worst earthquakes are located (fig. 7).
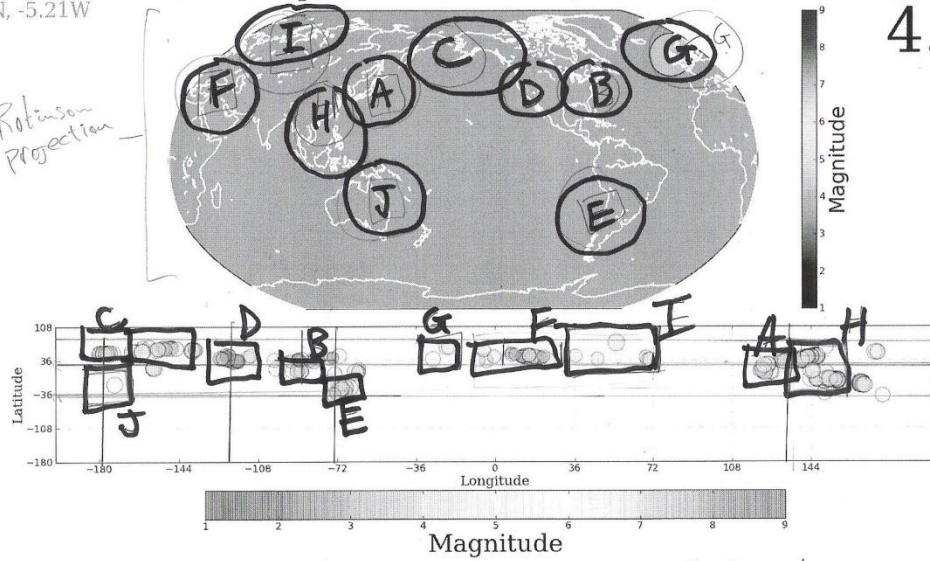
Subplots

Serif

Key

1974-08-10
12:49:38
57.147N, -5.21W

Iteration: 000339
Magnitude:

4.3

Earthquakes 1973-2009 (n=17143)

Robinson Projection

Magnitude

Longitude

Magnitude

A. Japan
B. USA
C. Alaska/Russian Border

D. California
E. South America
F. Europe/Africa

G. Scandinavia
H. Asia/polynesia
I. Russia
J. Australia

Sources

[1] The Graph Bank, "A repository for interesting graphs"
16 Dec. 2011
http://www.graphbank.org/index.html

[2] USGS National Earthquake Information Center, "Earthquake Facts and Statistics Graphs"
16 Dec. 2011
http://earthquake.usgs.gov/earthquakes/eqarchives/year/graphs.php

[3] USGS National Earthquake Information Center, "Global Earthquake Search"
16 Dec. 2011
http://earthquake.usgs.gov/earthquakes/eqarchives/epic/epic_global.php

[4] The Matplotlib API, "pyplot.annotate"
16 Dec. 2011
http://matplotlib.sourceforge.net/api/pyplot_api.html#matplotlib.pyplot.annotate

[5] The Matplotlib API, "pyplot.subplot"
16 Dec. 2011
http://matplotlib.sourceforge.net/api/pyplot_api.html#matplotlib.pyplot.subplot

[6] The Matplotlib API, "Customizing Location of Subplot Using GridSpec"
16 Dec. 2011
http://matplotlib.sourceforge.net/users/gridspec.html