EARTHQUAKES IN THE RING OF FIRE
(AND THE WORLD)

BLAKE WALSH, GABRIEL TREVINO, ANTONY BETT
{btwalsh, trevino, antonybett}@college.harvard.edu

FINAL PROJECT, CS171 (SPRING 2012)

[click for table of contents]

Group Members: Blake Walsh (btwalsh@college), Gabriel Trevino (trevino@college), and Antony Bett (antonybett@college)

FINAL PROJECT PROCESS BOOK - CS171

TABLE OF CONTENTS: LINKS TO COMPONENTS OF PROCESS BOOK

[this page intentionally left blank]

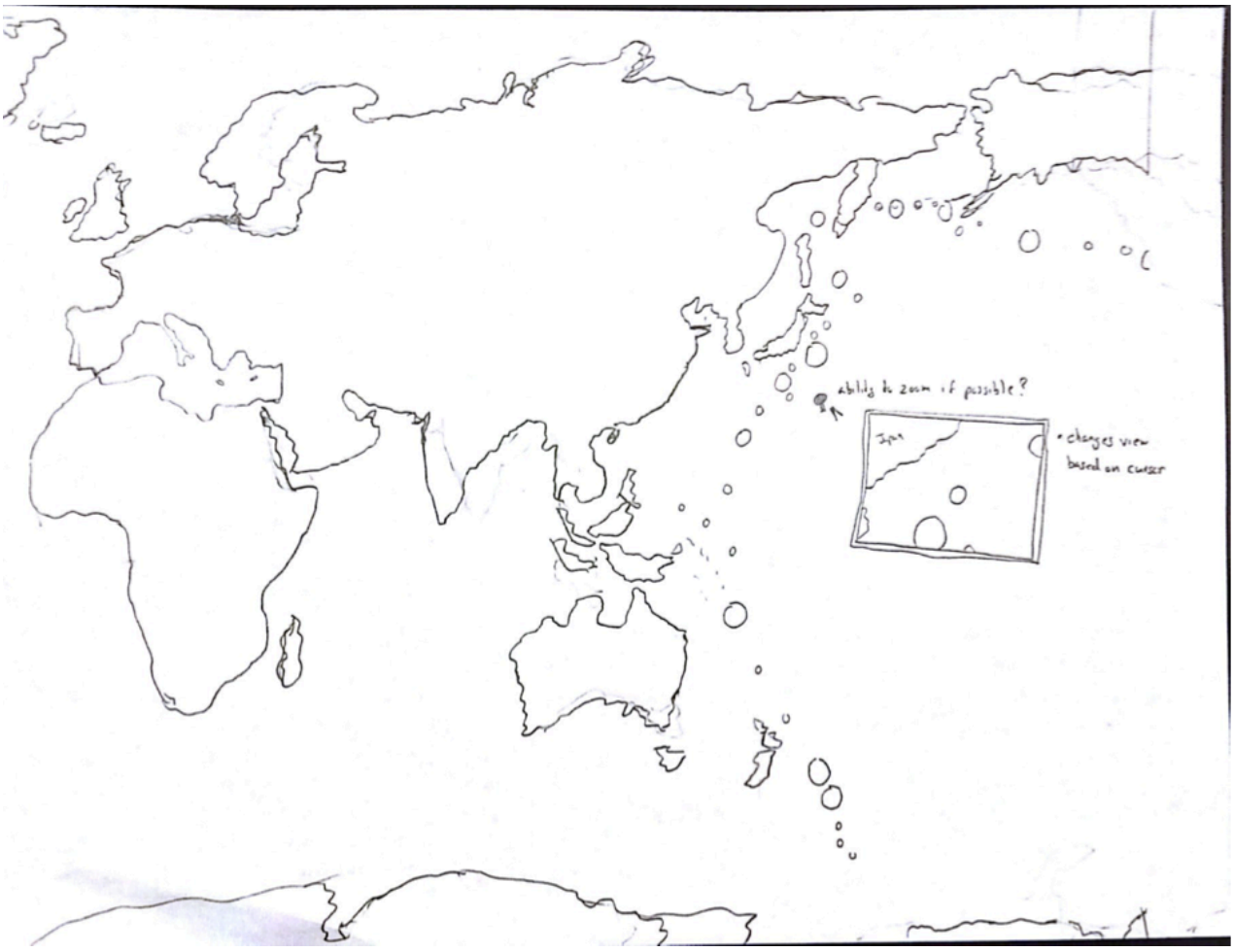
Part of Homework 5 (INITIAL PROJECT PROPOSAL):
[entry by btwalsh, 03.02]

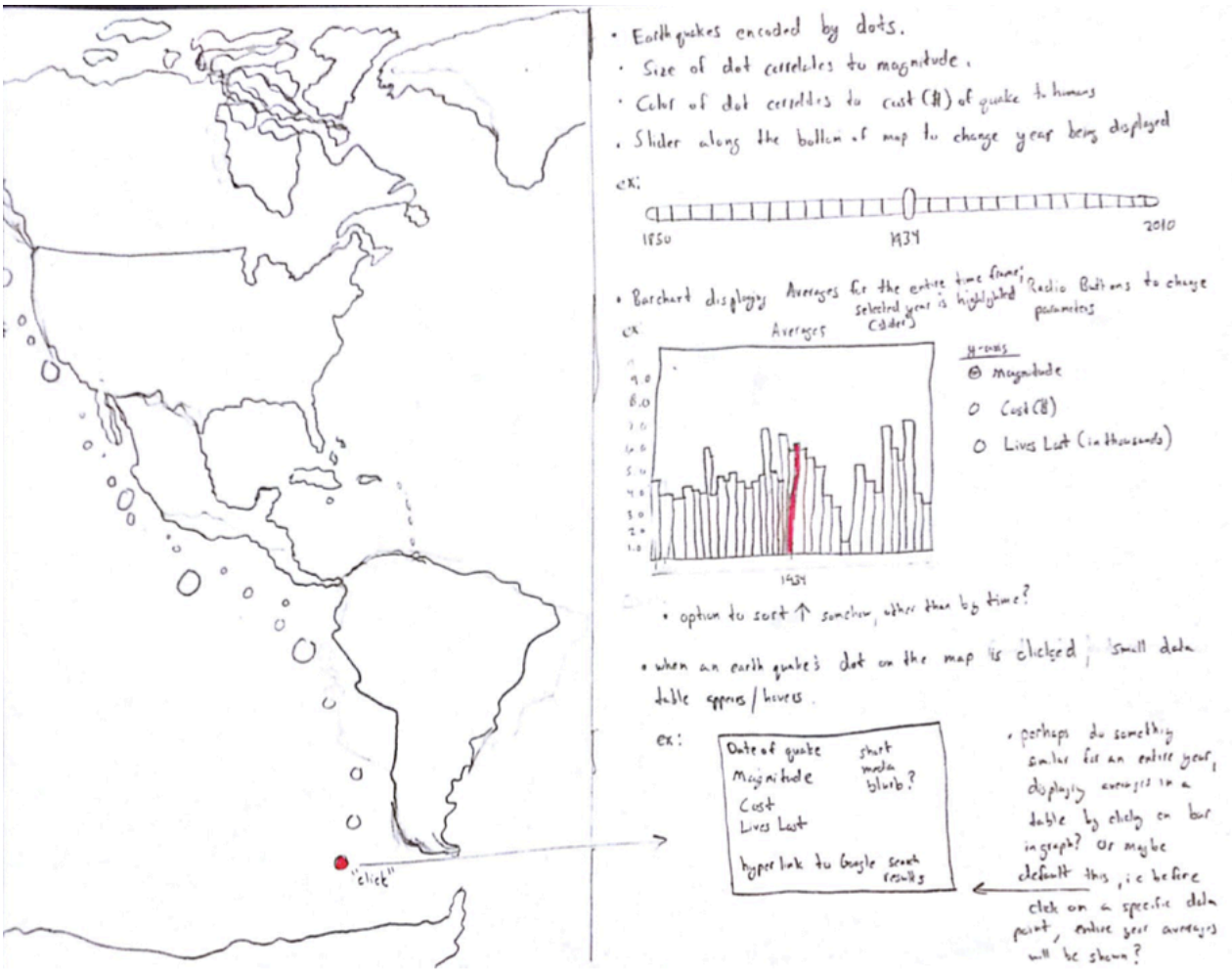
Start a Process Book as a Google doc with the format **project_title** and share it with your assigned final project TF. The TF will contact you soon and schedule a proposal feedback meeting next week, either Tu or Fr during lecture / lab time, or on Skype if you are not able to come to Cambridge. The choice of formatting is up to you, but it should be easily readable. Make sure the sharing settings of your process book are set to "Anybody With the Link" so that the TFs can easily share the document.

After your TF meeting next week, the first entry in your Project Process book should be your updated and final project proposal, addressing the same questions that were posed in the project proposal form:

- Title page with Project Title and Group Members (including emails)
    - DONE
- Background and Motivation: Discuss your motivations and reasons for choosing this project, especially any backgrounds or research interests that may have influenced your decision.
    - Earthquakes on the Ring of Fire (and those in general) have become increasingly popularized by globalization and the media, so naturally we have become more aware of their destructive nature and possible trends connecting the various events, etc... ; In the past decade or so, we've seen an apparent rise in the impact of natural disasters due to the increased media coverage. Never before have the cost of lives and dollars of these disasters been so apparent. However, we have been given the skewed impression that natural disasters of late have been far more severe than those of the more distant past. Now, is this actually true? We seek to answer this question by taking a significant quantity of data and visualizing it in an effective manner.
- Domain Goals & Audience: What domain goals would you like to focus on for your visualization? Who is the intended audience?
    - The intended audience for this visualization will be primarily the general population, though perhaps additionally for the more specific subsets of governments, insurance companies, etc. This can be used as an index level for preparedness for various regions for earthquakes and other natural disasters.
- Data: What data would you like to explore and/or convey in your visualization? Where and how do you intend to acquire that data?
    - Data to convey: Magnitude, intensity, and location of earthquakes in addition to economic damage caused, fatalities, injuries, and more. We will likely be acquiring our data from government/scientific websites for this project. Another area of interest is the amount of media coverage certain events received.
- Analysis Tasks: What low-level analysis tasks will your visualization enable? How will you aggregate or filter your data?
    - Value retrieval, comparison between relative values for one earthquake and another. In terms of filtering earthquakes, we should ignore earthquakes below a certain magnitude to avoid noise, and then give our time span a reasonable scope.
- Design Overview: What will your visualization ideally look like? Add pictures of your hand-drawn sketches of what it might look like.
    - Ideally, it will be a map of the world (or perhaps just the ring of fire) with a time slider on the bottom. Then, there will also be a second pane that acts as a bar

chart that compares the average for a particular variable along with the values for the current disaster being clicked on that can be toggled via a radio button of some sort. Below, you can see the original sketches of our visualization concept:





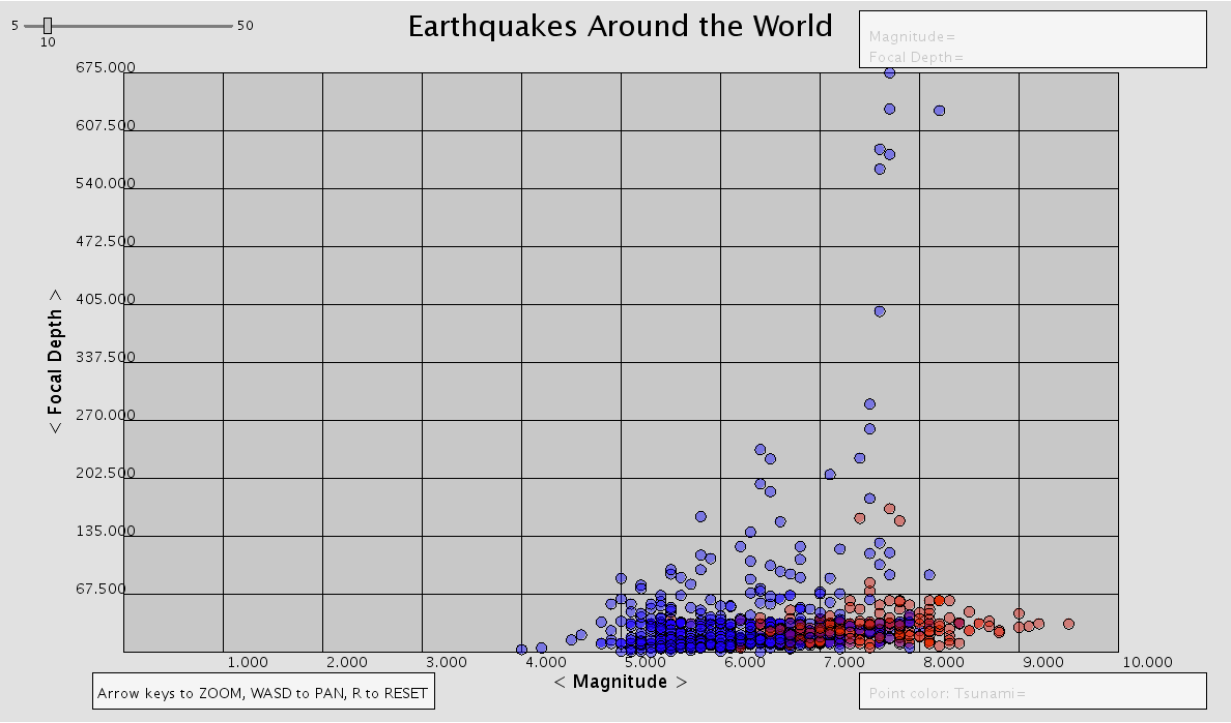
- Tools and Frameworks: What tools and frameworks do you plan to use for your implementation?
    - Processing, perhaps Python
- Team Experience: What is your team's level of experience or knowledge of the application area and the type of development / languages / toolkits you think will be required? If you have no experience this is fine! This section is to help us assess whether the scope of the project is reasonable for your team.
    - We all have programming backgrounds equivalent only to that which is covered in CS50. Additionally, Blake knows some C# and Java, and Gabe knows a fair amount of Java. ; "Blake: some Java, some Python, more C, some C#, a little Ocaml, CS50, MATLAB, etc; Gabriel: some C, some Python, lots more Java, CS50, MATLAB, etc; Antony: CS50 material, etc. Also, we have no experience in the application area, but we do have experience in the aforementioned languages.

- Planned Schedule: What is the planned schedule and division of tasks among the team members? Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members.
    - Given midterms and whatnot coming up soon around spring break, we likely

will not be making too terribly much progress before spring break. However, the research will be done by all of us. Based on the due date for the final project (4/20), we can easily map out exactly how we want things to progress when we return to campus. Nonetheless, we plan on having a working prototype not too long after spring break (perhaps a week or two). Given we intend on using Processing (which is basically Java) and Python if necessary, we may not have the most ideal array of experience for this sort of project, but we also have more than zero experience.

Homework 6 Extra Credit, Use of Earthquake Data for Processing Extra Credit
[entry by btwalsh, 03.09.2012]

For the extra credit for Homework 6, I examined a refined subset of our data to gain experience with refining our tentative data for our final project and actually using it in Processing. This was, more than anything, a test to see how sufficient my refining of our data has been thus far and to gain more exposure to our dataset, potential trends, and how I might want to go about dealing with the implementation of our project over the next month.



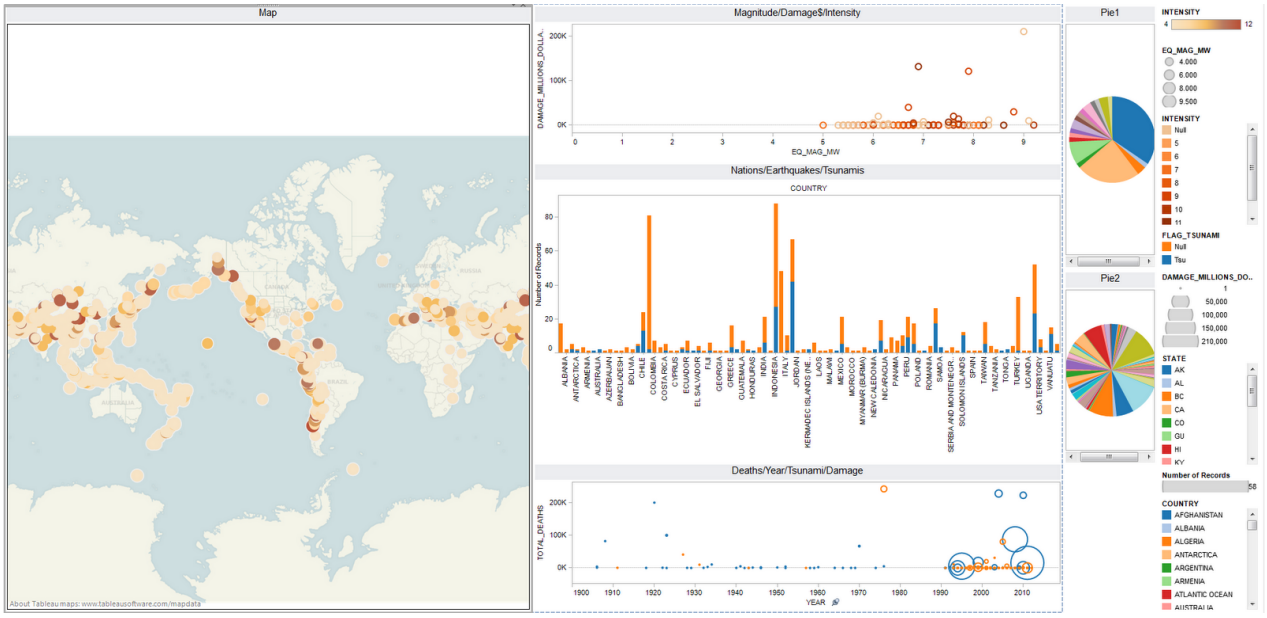
FEEDBACK ON INITIAL PROPOSAL FROM MEETING WITH TF: (first meeting w/ TF) [entry by btwalsh, 03.16]

On Friday, March 16th I met with Azalea regarding our final project thus far based on our process book, and here are my notes from our discussion:

- We seem to have a clear vision of where we are going with this project, so that is good
- For each new entry in the process book, perhaps a good idea to insert a page break. This is done in Google Docs by hitting ctrl-enter or command-enter, depending on your OS. The OS-independent route is to click “insert” above, then “page break”.
- After spring break, we have 3 homework assignments left, though each should not be as time consuming as, say, the Matplotlib assignment
- **Next Checkpoint**: March 30th Project Review, for which we should ideally have cleaned our data, filtered it, etc. Also, we should have some sort of prototype with the beginnings of some functionality prepared for this deadline. Critically look at data, little things to help with going forward
- **Following Checkpoint**: Meeting with TF the following week (approx. April 2-6)
- **Due date: FRIDAY, APRIL 20TH**
- During the weekend immediately following this deadline, we are to make a screencast video of some sort to provide a walkthrough of our visualization
- Final step of materials to produce: an HTML web page that contains our exported Processing file (really easy to do!) so the page can use our interactive visualization (provided an up-to-date JRE plugin) and perhaps contain an embedded copy of our video; this page will be hosted by CS171 so we need not worry about finding a server to host the files on
- Should probably use Processing for many reasons - better for user interactions, ease of exporting to an HTML file so a project can be shared on a website (which is what we will be ultimately doing)
- Website need not be fancy, but it is preferred that it not be too spartan
- Instead of one slider to select a year of earthquakes as indicated in our sketches, perhaps *two* sliders to select a *range* of times
- Be sure to check out Ben Fry’s book on *Visualizing Data* - a fantastic in-depth discussion of what can be done with Processing that is available for free and online through Hollis HERE.
- Be sure to keep updating the process book with great regularity in order to benefit the group staying on the same page and for the TFs to have a good understanding of the work we are doing
- **SEAS Project Fair on May 1st** for CS171, CS164, maybe CS51, and more
- Work on intuitive zooming in Processing for prototype? I [btwalsh] implemented zooming on previous problem sets in Processing, which we can initially use, though I do not think it’s as intuitive as it could be in terms of the front-end usage
- Consider to encode volcano-earthquake-tsunami relationships, how they correspond
- Acceptable (perhaps ideal!) to limit visualization to the Ring of Fire, need not cover the whole world

LOOKING AT OUR DATA IN TABLEAU, HW7 EXTRA CREDIT
[entry by btwalsh, 03.23]

For the extra credit portion of Homework 7 (Tableau - “Using your own data”) I decided to give our refined data another look using this framework. While we cannot of course use Tableau for our final project, it was nonetheless useful to gain an additional glimpse into our data at this point in time.



DATA DISCUSSION: perhaps we want to meet as a group in the near future to discuss this, or we can add comments and thoughts to the previous entry here:
[entry by antonybett, 03.27]

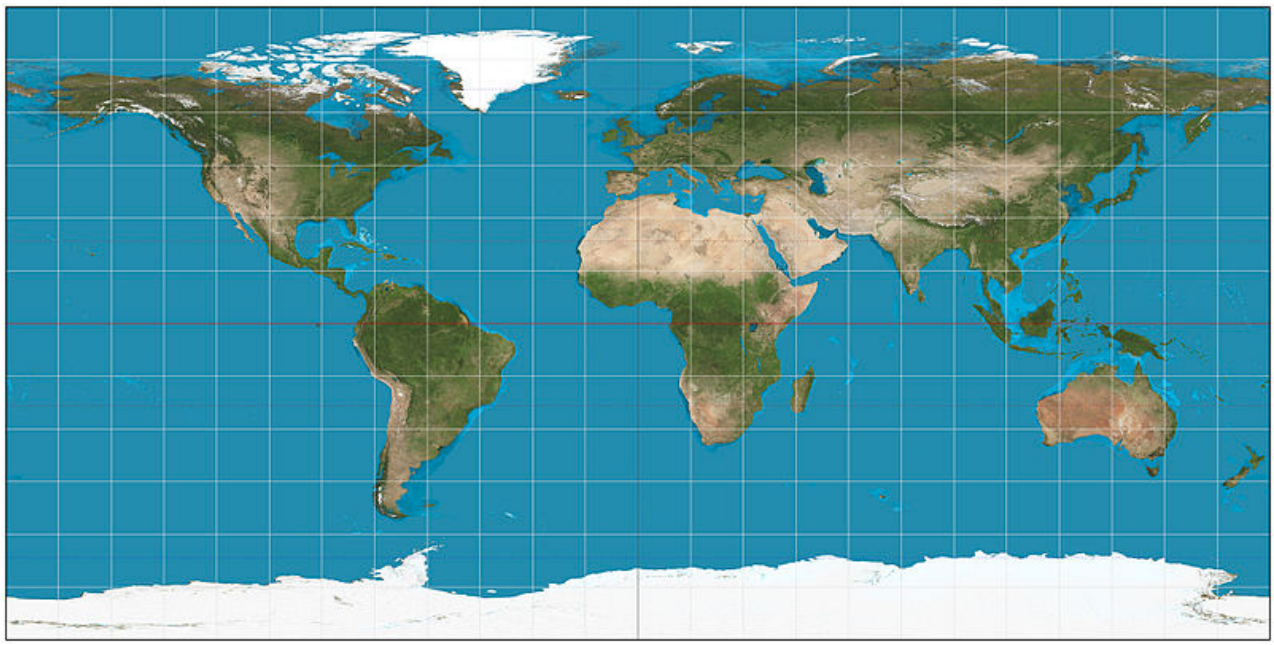
- **Data:**
  - http://www.ngdc.noaa.gov/nndc/struts/form?t=101650&s=1&d=1
    - Returns searches as tsv files,will make it very easy for us to acquire the data (we'll only need to clean it up.)
  - http://earthquake.usgs.gov/earthquakes/recenteqsww/
    - The second link provides earthquake data from the last 8-30 days. We can incorporate this to make the visualization up-to-date and more relevant in providing "real time" information on earthquake occurrences (I guess we have to scrape the data)

[updated by btwalsh, 03.29] +1

At least initially, we will be referring to the data available at the first of the two above links. The webform enables us to quickly filter our data down into a few different regions of the globe, types of earthquakes, whether or not tsunamis were generated and more. Of course, we plan to also use Google Refine to supplement the cleanliness of our datasets, despite the fact they are already reasonably well-organized and formatted to begin with.

March 30th PROJECT REVIEW
[entry by btwalsh, 03.30]
[updated with links by btwalsh, 04.01]

In addition to cleaning up our data with Google Refine for our purposes and making different versions of our data that may come in handy for various reasons down the road. The challenge at this point was the writing of code in Processing that would serve to map real world GPS coordinates to the pseudo-2D surface of the Earth which is of course the 3D surface of a sphere we are trying to represent in a two-dimensional grid of coordinates. The most appropriate map projection for this purpose is no doubt the *plate carrée* projection, or the Equirectangular Projection, which I determined given it seems to be the norm for GIS-pixel-mapping. There is apparent distortion of the poles due to this, though with a few formulas the code becomes much easier to work with. Here is what this projection looks like:



Going off the definitions of x and y coordinates for this projection, I was able to in relatively few lines of code accurately display several latitude and longitude pairs for cities in the NESW-hemispheres, after surfing through the Processing forums for inspiration - such as the discussion found here. Of course, the next step here is to automate this insertion of points from a TSV containing our data but, given that Processing deals with TSV files rather cleanly, this should not be all that difficult. See below for an image of the preliminary visualization I was able to generate with points plotted at Paris, Buenos Aires, Vientiane, Seattle, Barrow, Cape of Good Hope, Ecuador, and Turkey.

[note by gtrevino, 03.31]

A point of concern that needs to be addressed in the upcoming meeting with the TF is the view of this projection given the nature of our data. since the “Ring of Fire” encompasses what is currently the East and West sides of our projections, I wonder if it would not be better to use something more similar to our original sketches, for example with the Pacific Ocean in the center. The longitudinal span of the projection is fine, so btwalsh and I feel it would be best to just to manually edit the map and shift the coordinates as needed, but we will wait until the full group/TF meeting to implement any changes.

The current projection itself is clearly fine and way more than adequate for our prototype, though if it is thought to be necessary, we can attempt to make this modification for the later stages in our project.

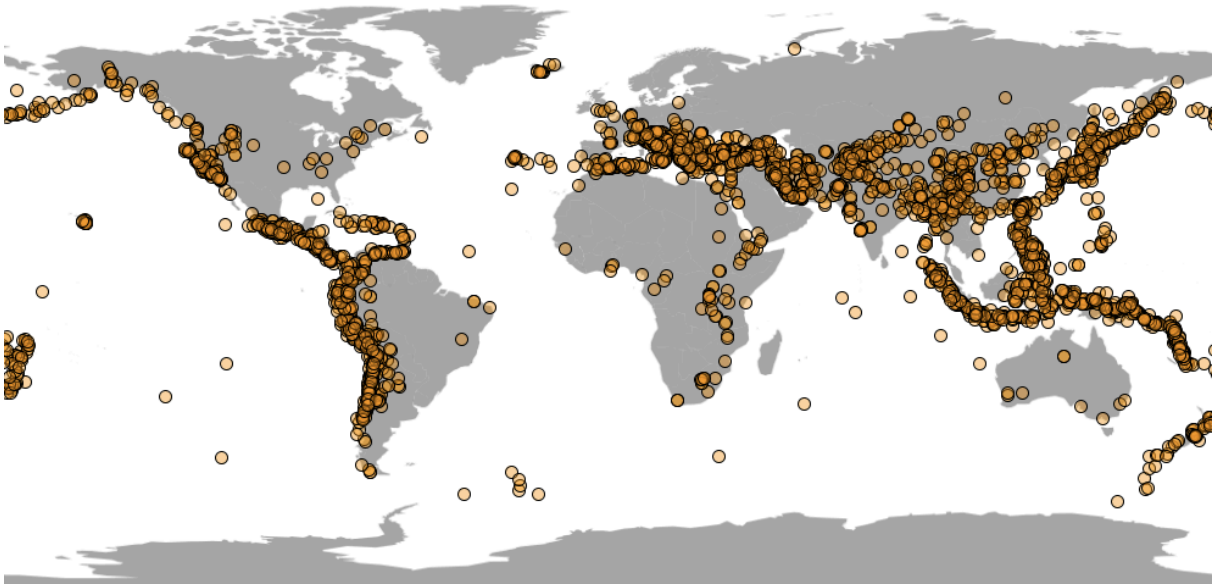
GIT Repository Now Set Up for Group Collaboration
[entry by btwalsh, 03.31]

Today, I set up a private git repository on bitbucket so it will be extremely easy for us to collaborate on code, even if we are not in the same location over the next couple of weeks. This way we can easily avoid the logistical nightmare of emailing different versions of the same code back and forth!

More Data Cleaning and Automation of Point Plotting in Prototype
[entry by btwalsh, 04.01]

Today, now that I have confirmed that my implementation of a few sample points is functional, I took our XLS file of earthquake data and began to clean it up some more with Google Refine specifically for the purpose of easily integrating it with our existing Processing code. In terms of the cleaning that was performed, I renamed the columns to read more sensibly, eliminated data entries that did not indicate a latititude and longitude for the location, and then did so for the rows of data that similarly did not actually report a magnitude.

Now, we have implemented a way to automatically map our latitudinal/longitudinal coordinates onto our grid of x-y pixels, as demonstrated below:



[note by gtrevino 04.01]

Thankfully we got this to work (after dealing with some null pointer exceptions), but we need to work on our data density, i.e. how to deal with multiple points overlapping, as this will become especially important as we implement **user interaction**. This will probably resolve itself a bit as we add filters by time and other characterstics of the data, but may still be an issue as we move forward and should be kept in constant consideration.

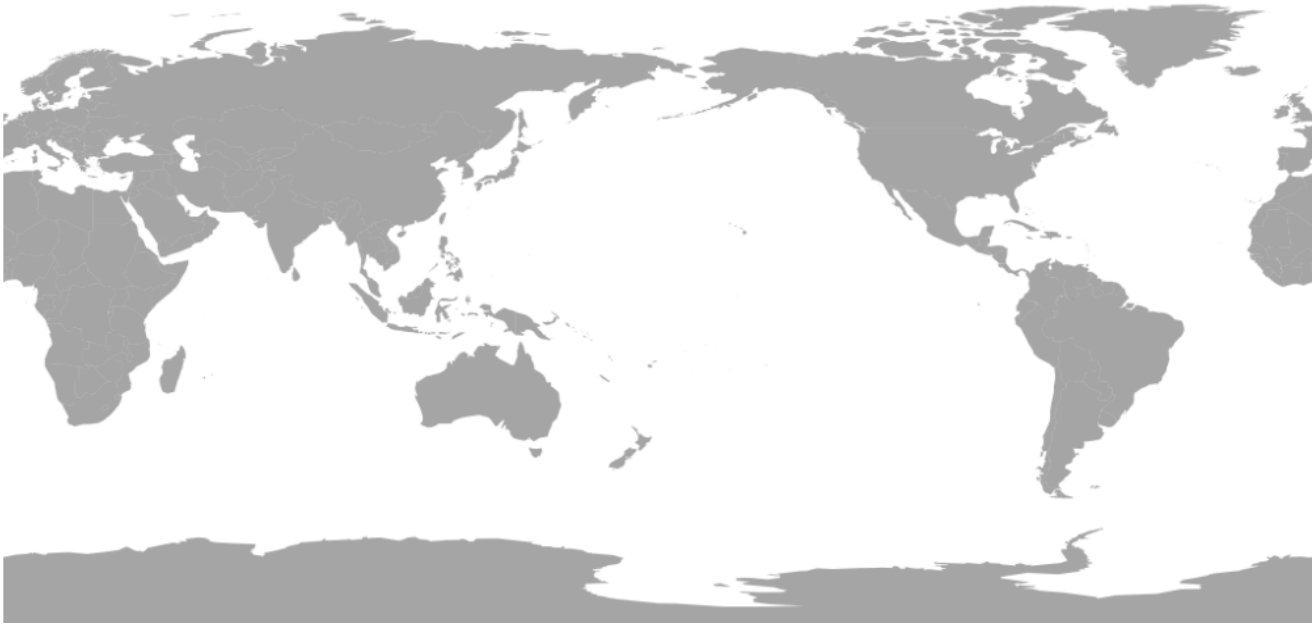
Additional Data Cleanup - Clustering
[entry by antonybett, 04.01 @ 11.25pm] +1

As **btwalsh** noted**,** most of the fields in the "Earthquake Location Name" column are not grouped into distinct clusters. For instance, fields that could easily be clustered into “"Albania", existed in different forms such as "Albania:Elbasan", "Albania:Elbasan,Peqin", “Albania: Korce, Starove, Pogradec, Salxhljas” or fields that belong to the “California” group existed as “California: Big Dear City” , “California: Central, Coalinga” etc. Inasmuch as these different groups provide specificity of the earthquake location, we felt that this was redundant given that the longitudes and latitudes will perform this same function. Using Google Refine’s text facet feature, I managed to transform the initial 2125 groups into 178 groups, each group representing a particular country or region. For locations in the US, I chose to remain with the State names since this will be more relevant and efficient in relaying the location. Certain fields contained generalized region names that could be further broken down into more distinct groups e.g “Balkans NW: Slovenia: Ilirska...” and “Balkans NW: Serbia: Krajevo...” could be broken down into the respective countries i.e Slovenia and Serbia.

In addition to maintaining uniformity in the clustering of the fields, grouping the fields into countries/region will prove useful once we start implementing search or filtering features based on country names.

Modifying Map Image to Suit Ring of Fire
[entry by gtrevino 04.02]

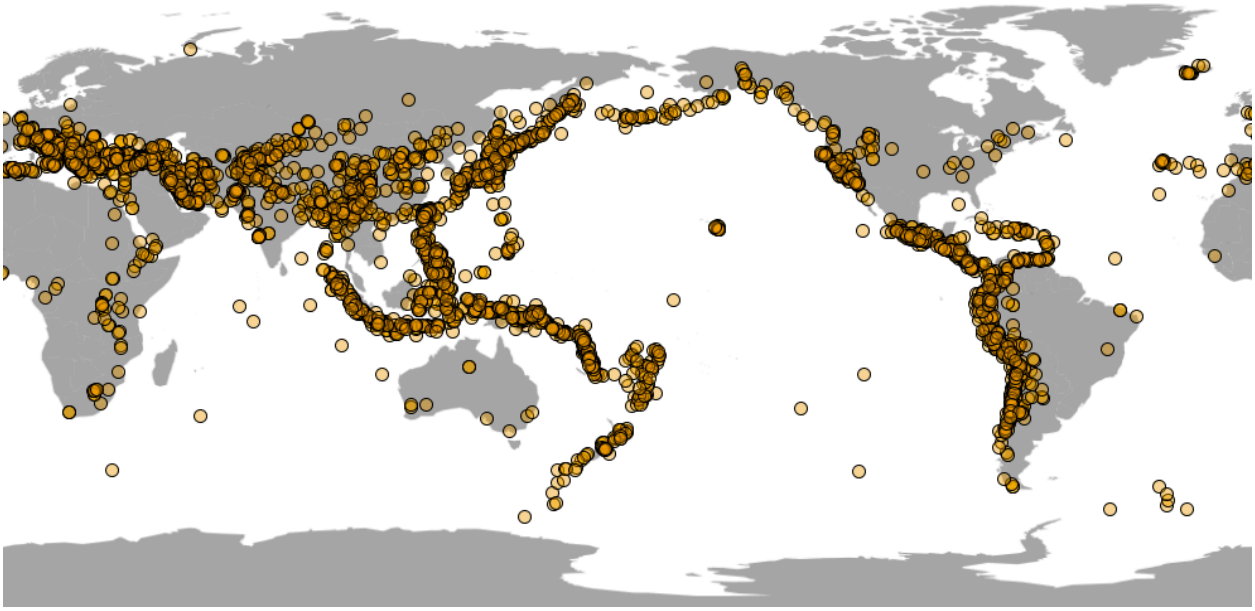
Finished editing the map per original sketches layout, i.e. centered Pacific Ocean and Ring of Fire. This should make our user interaction a lot more friendly, as well as allowing trends in our data to be perceived a bit easier, now that users have a “central” area of focus. It should be noted that in editing the map to this new configuration, it was discovered that the map was not wholly accurate with respect to land mass; that is, along the previously separated Bering Strait had a noticeable gap in land mass, equivalent to about 2x10 pixels of land. This gap was filled in to reflect the actual state of the area. This shouldn’t present any errors or difficulties as we move forward.



Changing our code to deal with the new spatial coordinates of our data should be fairly straightforward.

Code Update to Match Revised Map
[entry by btwalsh, 04.02]

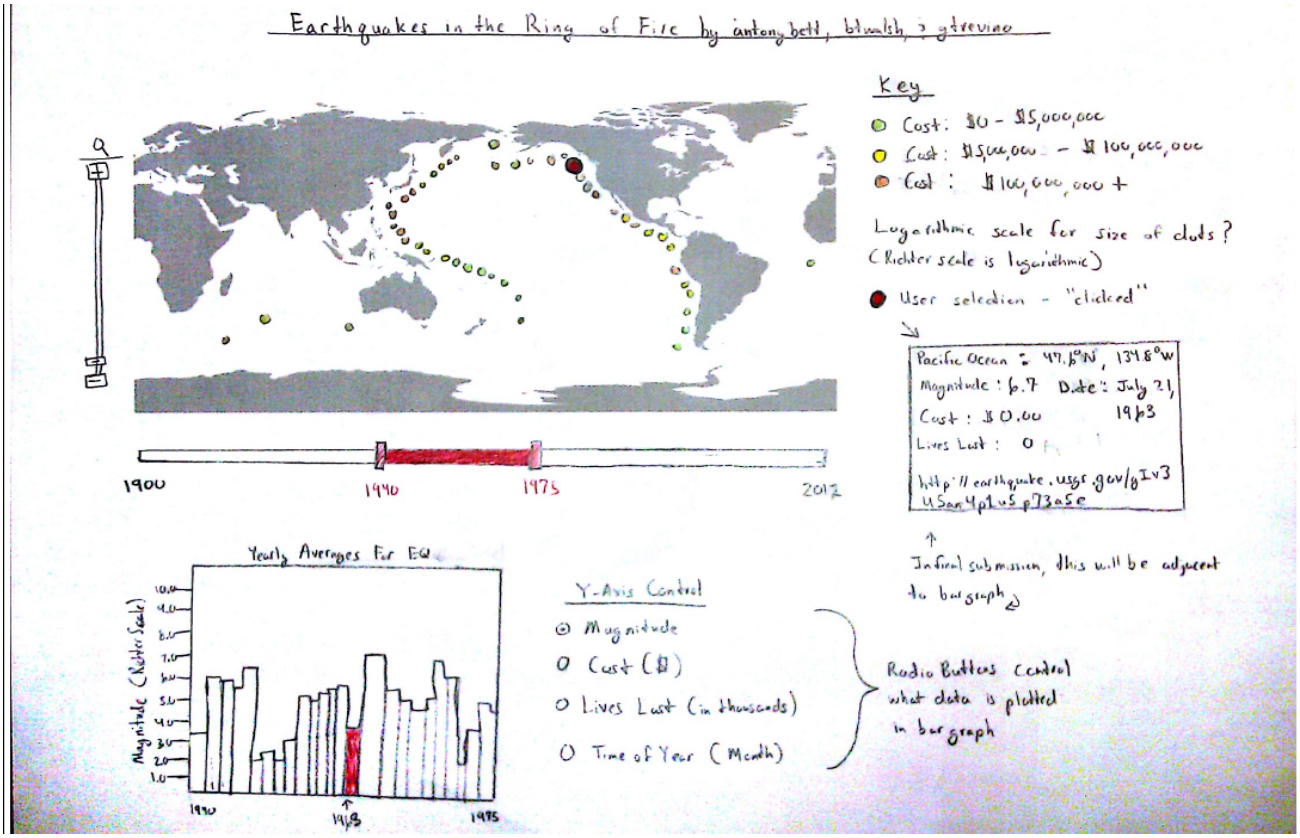
Given the newly revised map thanks to **gtrevino**, I have revisited the portion of my code that determines how latitude and longitudinal points are converted into pixel coordinates and added a rightward shift for values that were originally in the left half and a leftward shift (or a rightward wraparound depending on perspective) for points that originally fell on the right half of our map. Now we are nicely set up to move forward with implementing user interaction, additional panes of visualization, and other features in the coming weeks.



Updated Sketches For April 5th Meeting w/ TF
[entry by gtrevino 04.04]

With the large amounts of progress made since the original sketches and in preparation for the second meeting with our TF, Azalea, in the upcoming days, I have created an updated sketch of our final visualization. This new sketch encompasses the updated world map we have chosen, as well as implements the major critique from the first TF meeting, e.g. a slider that allows a range of years to be viewed. It should be noted that on the whole, per the notes from the initial TF meeting, the original sketches did not possess any further areas of necessary improvements. On that note, the rest of my updates to the sketches were things such as streamlining data tables, adding detail where necessary, adding a legend for greater clarity, and other things of that nature.

As we progress further into the creation of our visualization, it is to be determined the exact layout of our visualization. The layout presented in the sketches is more for the sake of presentation than feasible, as the world map will have to be fairly sizeable to be easily viewed by the user, but there is also a lot of necessary items that must be given space. My current recommendation is that we research pop-up/multiple windows based on user input, i.e. click on a data point: a small window with relevant information pops up/fades in.

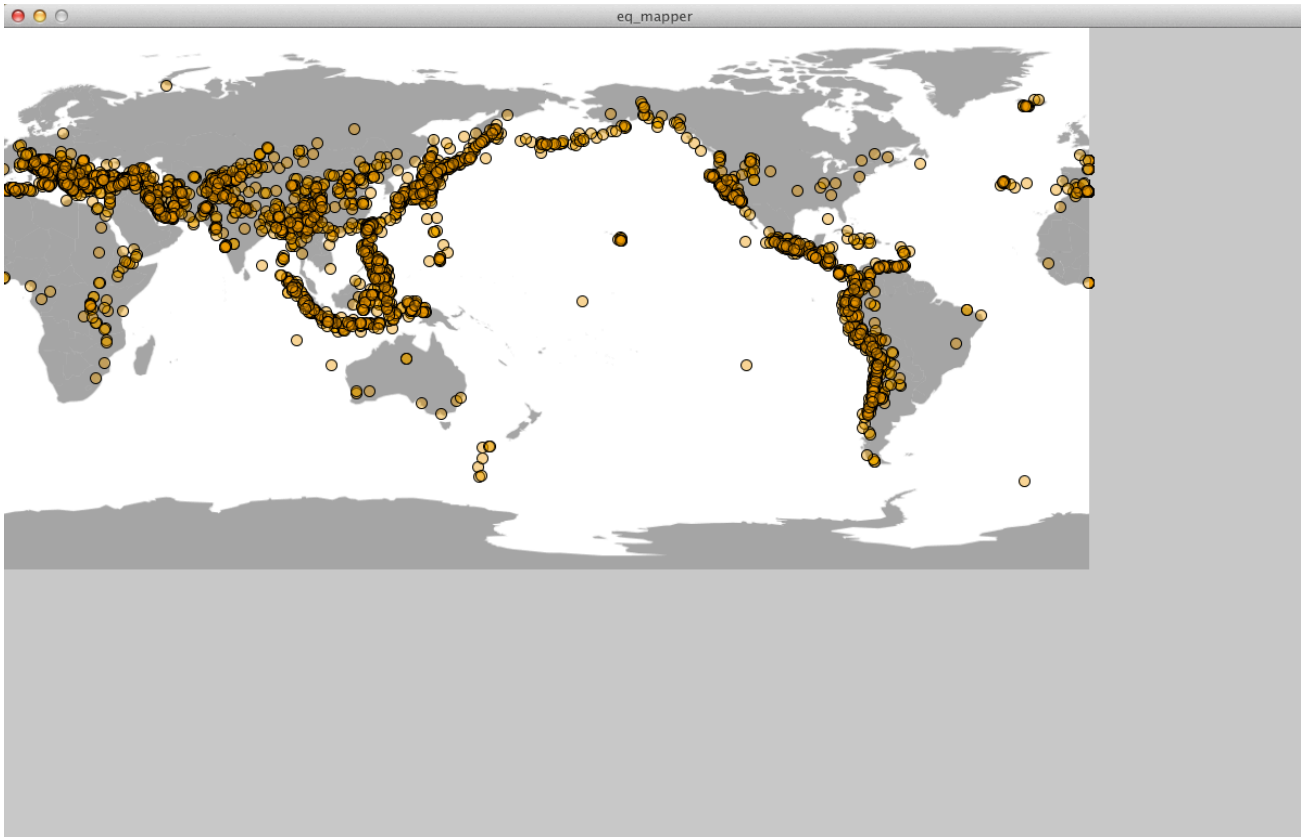


Feedback from TF (Azalea), Project Check-In
[entry by btwalsh, 04.05]

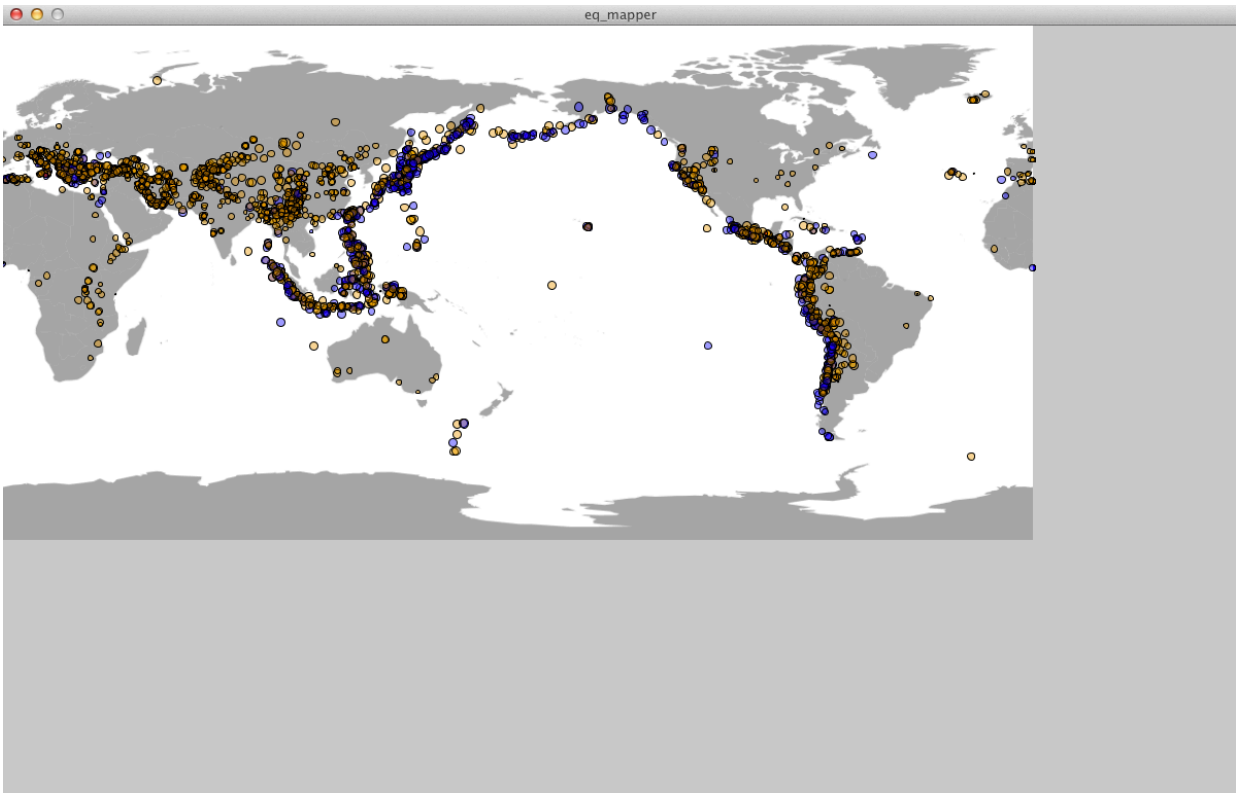
- Map looks good for background
- Excellent that we already have all our data collected and cleaned
- Sketches indicating the next steps for user interaction show a good sense of direction
- Certainly consider implementing a double slider
    - Start with a single slider, then think of ways to extend it
- Think of how to possibly implement zooming
    - As I did in the extra credit for a past Processing HW, zooming/panning
    - Perhaps the ScrollWheel event may be useful
- Think of additional panes to add to visualization
- Incorporate other data elements in visual encoding:
    - Earthquake magnitude (Richter scale) for size
    - Different colors for presence of Tsunamis and Volcanoes
    - Videos/Images of Earthquakes - explore possibility here
- Need to get all group members going with GIT
- (also this is hilarious: http://www.commitlogsfromlastnight.com/)
- (these notes have been added to the GIT repo as well)

Next Steps with Coding, Inclusion and Encoding of Additional Data
[entry by btwalsh, 04.08]

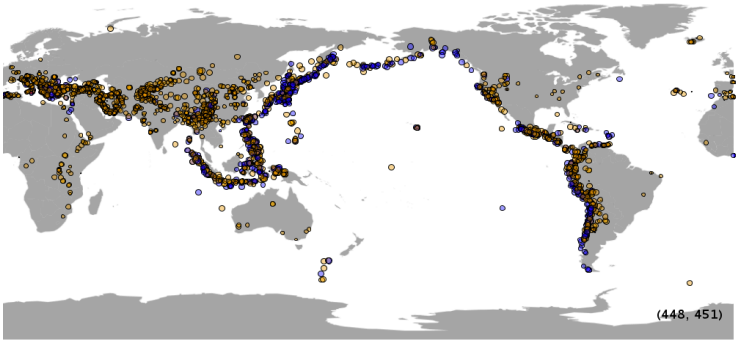
In order to move forward with our visualization, we need to be able to resize the map image within the Processing display all the while keeping our plotted points around the world proportionally in their proper locations. After spending some time with the code, I realized the well-factored out code I originally wrote made it incredibly easy to now have our Processing files output a display much like the following image:



The next step was to implement some visual encodings for additional aspects of our data, such as varying the color of the plotted points based on the presence or lack of tsunamis/volcanoes at the time and location of each earthquake and varying the sizes of the plotted points based on the measure of magnitude on the Richter scale. A preliminary result after finishing writing up some quick code to accomplish this looks like the following:

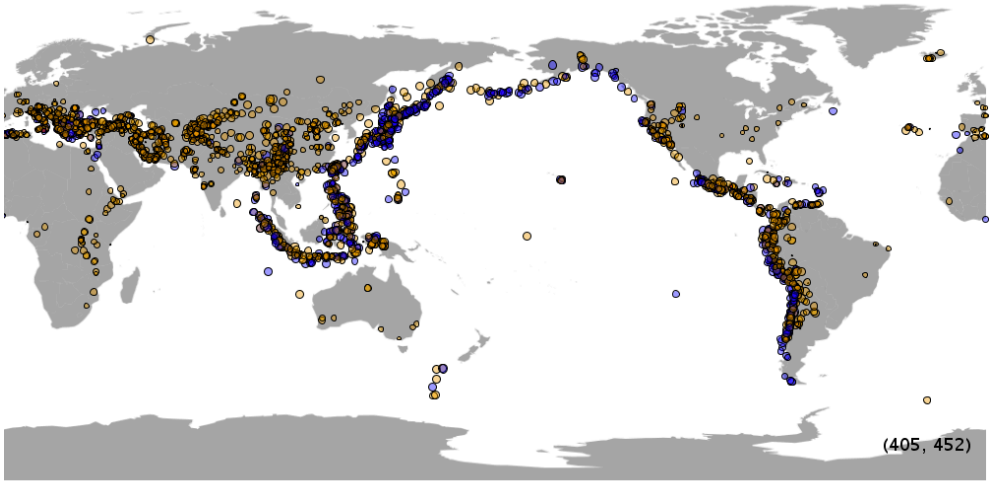


Of course, it is pretty clear that the density of our data is rather high, so it would be best if we were to implement some user interactivity that would enable additional filtering so that the users will be able to derive more meaning from our visualization. For debugging purposes, I thought it would be a good idea to have a real-time updated text field that shows us the pixel coordinates of the cursor. However, I ran into a problem with placing this text field on the Processing window and not over the map - the text field would continually overwrite itself rather than replace itself everytime it needed to be updated. The problem is documented in the below image:



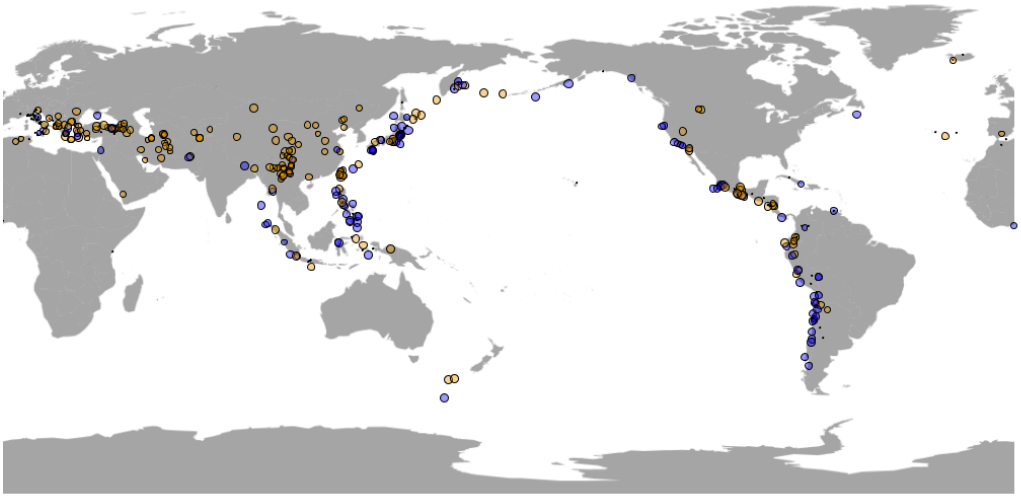
(■■■, ■■■)

Notice how the exact same text field I also instantiated over the map image does not have this same problem. After doing some brief research, I was not able to determine why this was happening, but I did have a good idea as to how to fix the problem. All I did was make a white PNG image of the same resolution as the Processing display and load that image in just prior to my loading of the map. And, sure enough, the problem was no more!



(405, 452)

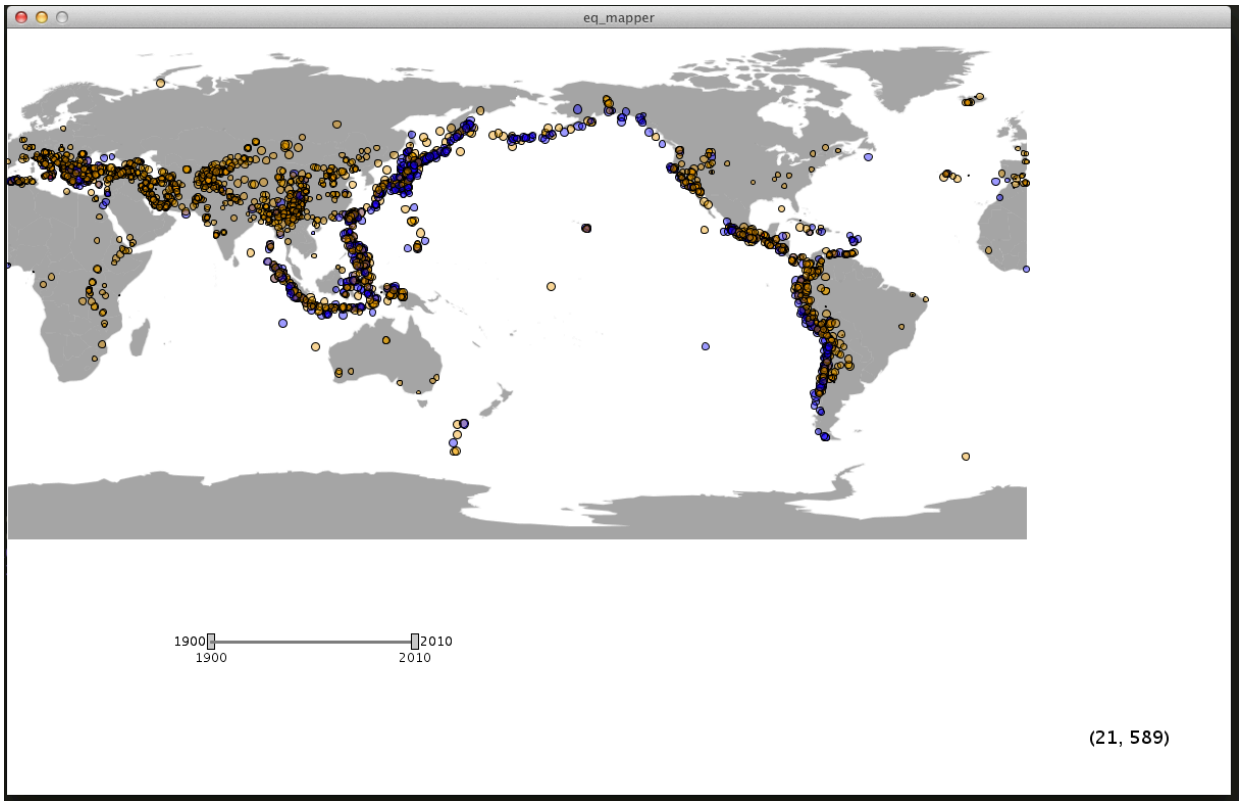
Now that I have resolved the issue with presenting debugging text regarding cursor location, I could then begin the implementation of some user interaction and filtering. At this stage in the visualization creation process, I elected to implement sliders as upper and lower bounds on the years of earthquake data to be displayed so that various ranges of time can be plotted and singled out from each other. As of now, the bottom slider indicates the lower bound on years to be plotted while the upper slider determines the upper bound on years to be plotted. If the lower bound slider oversteps the upper bound slider (which should not be possible), my current implementation simply displays only the earthquakes that occurred in the year of the location of the upper bound slider.



Of course, some of the next steps with respect to this project include improving the look and feel of the visualization, adding additional panes of graphs (perhaps bar graphs and other ways of encoding and looking at our information), and exploring the ControlP5 library for Processing.

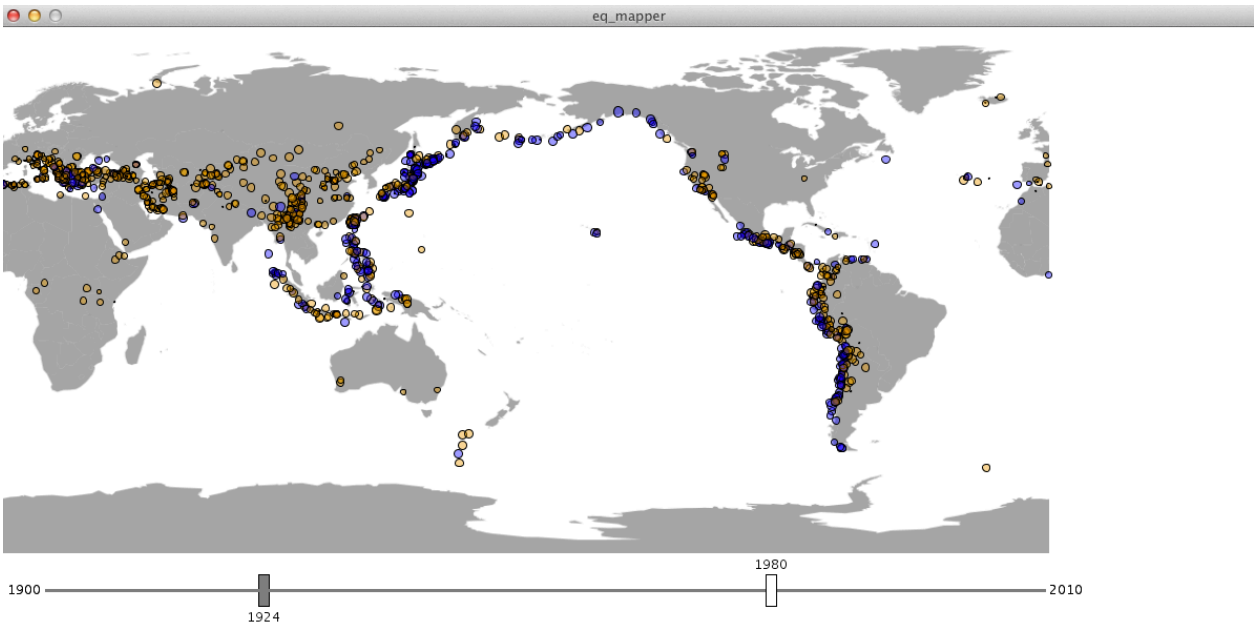
More Work with Visualization, Sliders, Visual Encodings, etc
[entry by btwalsh + gtrevino, 04.09]

Tonight we had yet another group meeting to discuss the next steps with our project. The first possibility we explored was a implementing a single slider by superimposing the two sliders **btwalsh** had already created.



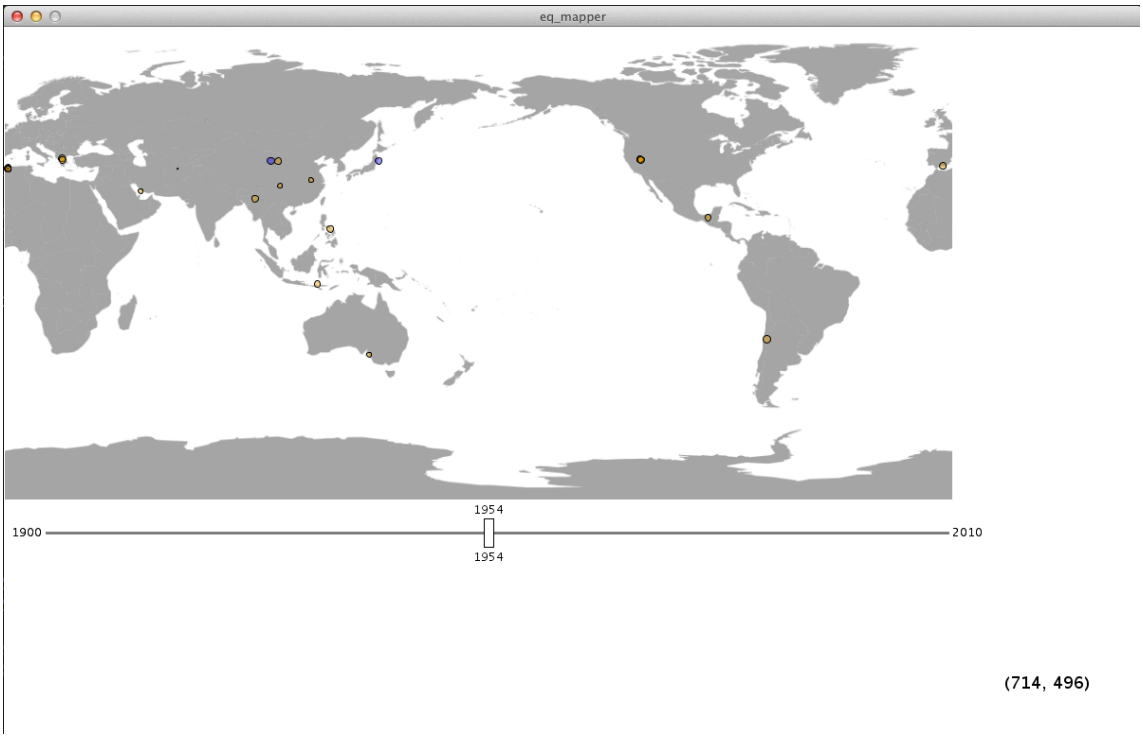
The results were marginally successful, but we had issues with differentiating between which slider the user intended on clicking. However, this was quickly resolved as Processing allows us to easily differentiate between left and right clicks. It was then simply a matter of differentiating between the left and right click for these two slider handles to ensure the correct slider would be selected as the user intended. The left click corresponds with the lower bound slider (which is also physically to the left of the other slider bar), and the right click corresponds with the upper bound slider by design (which is physically to the right of the lower bound slider bar).

In order to focus on a specific year, users can both drag the sliders individually as normal using left and right click and drag, or click the left and right mouse buttons simultaneously (or simply in succession without moving the cursor) to have both sliders focus in on the same year. It was then simply a matter of differentiating the sliders by color such that the slider responsible for lower bound (gray) and the slider responsible for the upper bound (white) are clearly distinguishable. Talk about some smooth user interaction!



(598, 352)

Sliders in same location (single year, triggered by simultaneous left and right click):



(714, 496)

It should be noted that we have edited the slider functions (in particular, the slider.pde class file)

such that it is physically impossible for the user to drag the lower bound slider beyond the upper bound slider, so as to avoid an impossible ‘range’ of years where the lower bound oversteps the upper bound. When the upper bound slider is moved is dragged below the lower bound slider, the lower bound slider moves with it, focusing in on one year, as shown above.

Divided Tasks, Week of April 9th and Onward
[entry by btwalsh, 04.09]

Moving forward, we have several main tasks that need to be addressed and that have been assigned to each of us, as the deadline is fast approaching:

1. Creating a secondary data file containing averages of each earthquakes various data things, for use in bar graph. **[antonybett]**
2. Implementing and refining the bar chart (see below). **[gtrevino]**
3. Implementing hover and click concerning data points, such that all the relevant information about each data point is presented in a readable table. [**btwalsh]**
4. Cleaning up, organizing visualization space, adding legend. **[btwalsh+ALL]**

Supplementary tasks (to be done after initial 4/20 deadline):

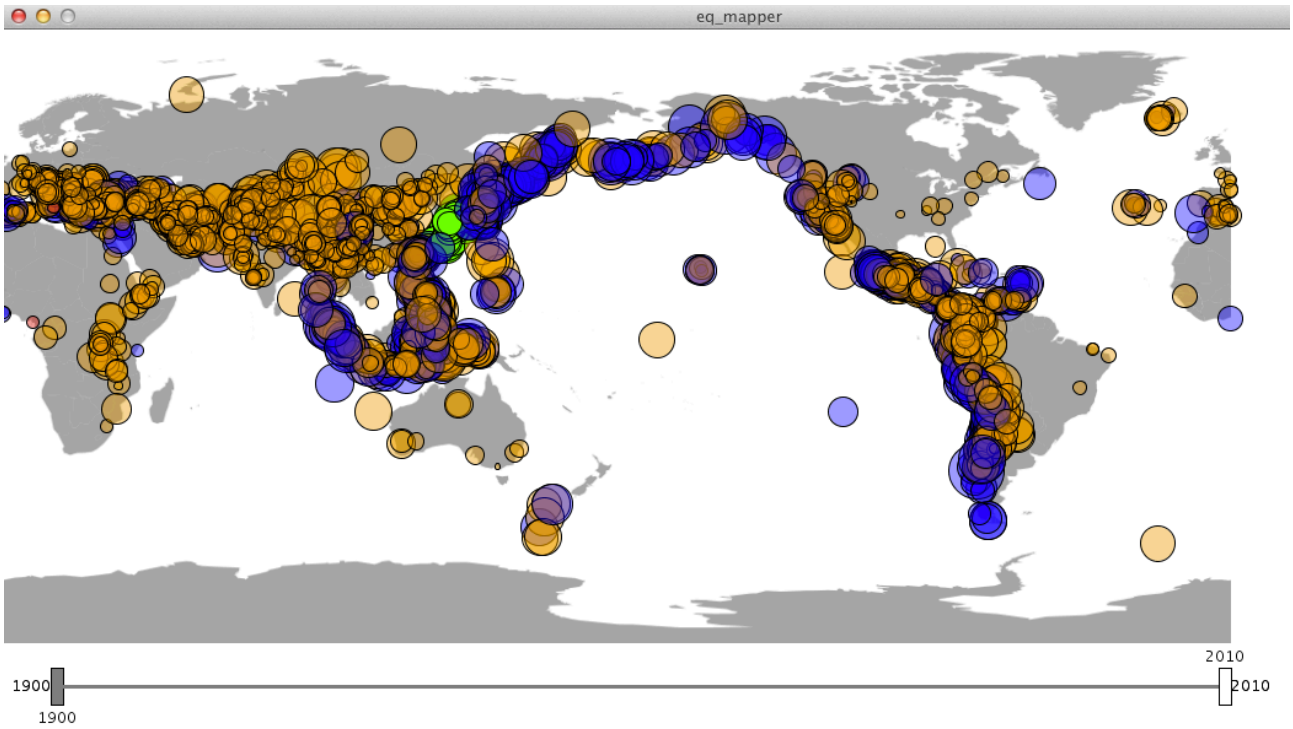
1. Screencast (~2min) for presentations (due Sunday, 4/22, to be done over that weekend)
2. Export to html/website (due Friday, 4/27, to be done over the week leading up to that day - will not be too difficult given the convenient “Export Applet” ability in Processing)

For Bar Chart (more data work):

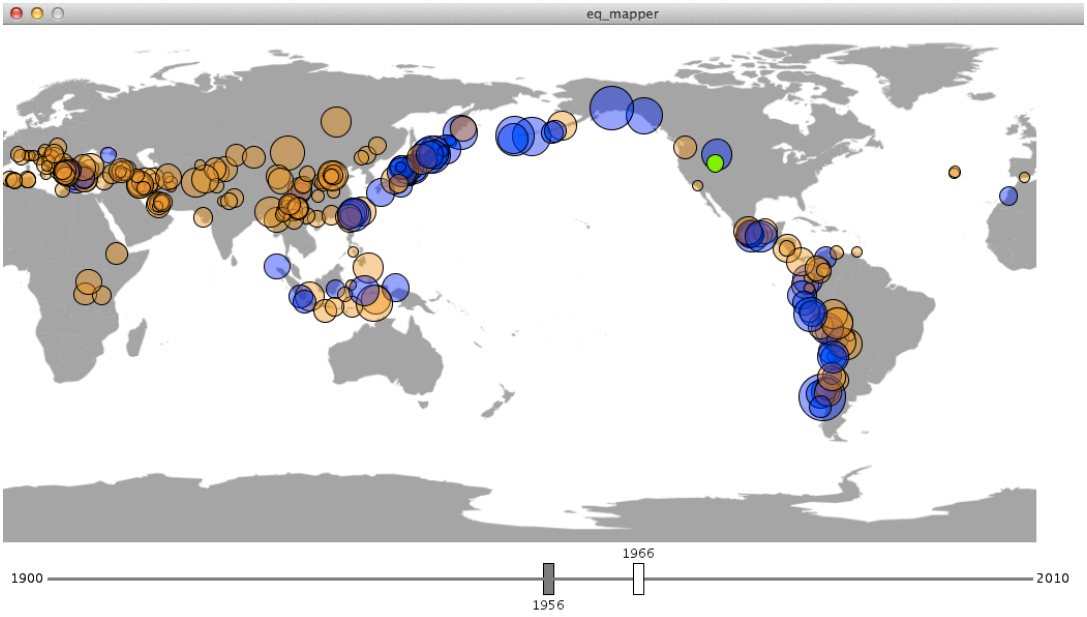
- One entry per year
- For each year:
    - Average magnitude for that year
    - Average Cost ($) for that year
    - Average Lives Lost for that year
    - Time of year average, perhaps?
        - Divided into seasons? May not be worth it, though potentially worth exploring.
    - Number of photos available as data, may be worth looking at

Implementation of Hover Effects and Work Towards More User Interaction
[entry by btwalsh, 04.10]

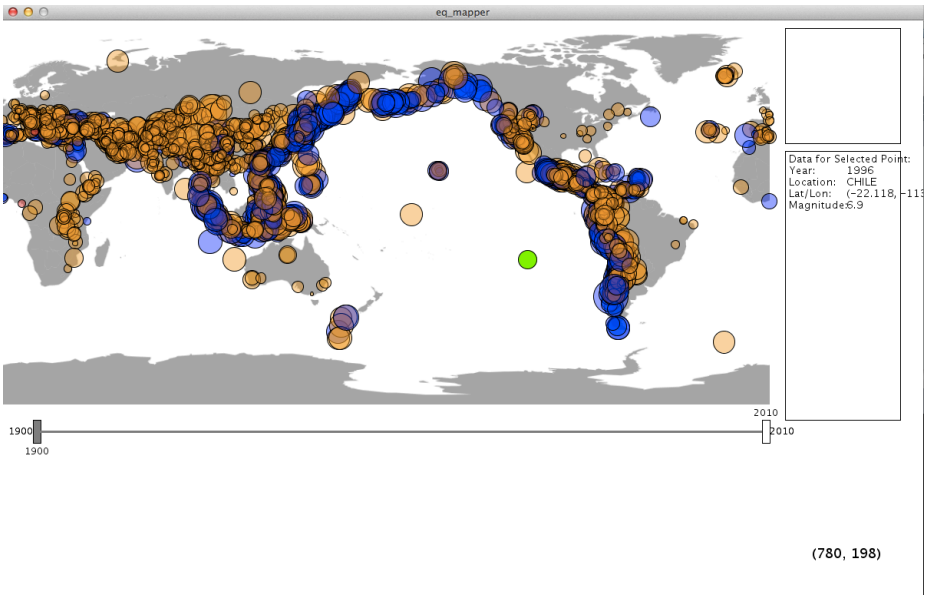
Now that we are adding more user interaction to our visualization, there are a few things that need to be considered. The first task I set out with this evening was to implement some sort of hover (or clicking) effect when the user intends to single out a particular point in the visualization. I revisited the code written for a past problem set (the basketball data visualization wherein we were to write our own hover effect and implement some other forms of user interaction) and adapted it for use here in our project. Initially, I ran into some problems with multiple points being selected via the hover effect, as the below screenshot indicates:



The way to fix this, of course, was to have some sort of global boolean variable that would only permit one plotted point to be selected at a time. This was easy enough, though as a group we decided a better design choice would be to have this be a “click” interaction instead of a “hover” interaction. So, I added the event listener for a mouse click that essentially waits for the user to click the mouse, and upon the firing of this event, then the program checks (as it did in the previous hover-based implementation of this interaction) if the click occurred on a plotted point. If so, that point is highlighted until another point is selected with a contrasting green color to easily differentiate it from the surrounding points of various other colors. Of course, it becomes difficult to click on the point one intends to if the data is really dense (as the above image indicates), and this a perfectly good example of the usefulness of our implemented year-filtering sliders at the bottom of the map. This way, the user can filter out the years that do not interest them and narrow down the sheer number of datapoints that are plotted. See the below image for a demonstration of this:



Next comes a placeholder for the legend and a text region to show detailed information about points selected by the user. Of course, some cleanup here needs to be done, visually. For now, I have kept the Processing window resolution reasonably low given my partners' machines do not have as high a screen resolution as my own (1920x1080 versus 1366x768 and 1200x800). When they conclude their duties to the group on this project, we can increase the resolution to better display all of the information we would like in their entirety and in an uncramped manner. See the below image for what the project looks like at this point in time:



Further Data Generation

[entry by antonybett, 04.11]

At this point in the design process, we are moving into adding more user interaction that will empower the users to delve deeper into the visualization in search of more information on the earthquakes. We plan on integrating a bar graph that will offer a quick comparison of the selected earthquake and its relationship to other earthquakes within a given year. In this regard, there was the need to compute yearly averages for the earthquake magnitudes, number of deaths and economic cost that will enable the user to interpret a selected data point and how it stands when compared to other occurrences within that specific year.

In computing the averages, we had two approaches at our disposal. One approach was to compute the averages using Processing, in such a way that calculations would be carried out everytime the user selected a datapoint. This would mean looping through the database and selecting the entries of interest as per the selected data point. However, given the large size of the dataset, we decided against this approach since it posed the danger of slowing down the visualization. We opted to carry out the computations externally and have the averages stored on a secondary file. This second approach was more reliable as it would eliminate the weight of the computations, enabling us to utilize Processing solely for the visualization.

I managed to compute the averages for earthquake magnitudes, deaths and damages for each year using Excel, utilizing its function feature. In addition, I computed the total number of photos of the earthquakes in a given year which we hope to use as measure of the media coverage.

Code Commenting, Rewriting, and Other Cleanup
[entry by btwalsh, 04.11]

I realized I needed to comb back through our project code, given the potential for redundant code that can be factored out, and the code also desperately needed more thorough comments in some areas so that my group members would not have problems reading and understanding my implementation of the various parts of our visualization.

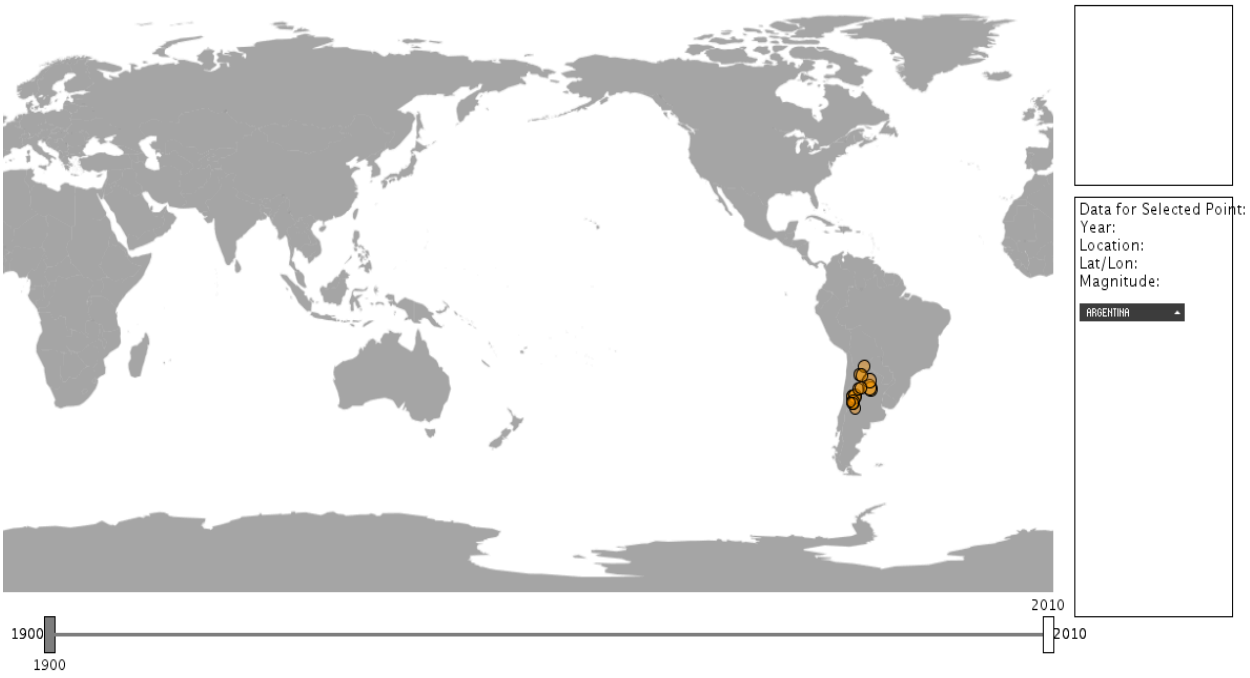
With my coding thus far essentially complete, I am now waiting for **antonybett** to complete his data work with a secondary data file we plan on using for an additional to component to our visualization and for **gtrevino** to implement a bar chart visualization to add to the bottom of what we currently have (as per our sketches of prototypes a few weeks ago). As for my (**btwalsh**) work, I have done absolutely all I can for this project at this point in time and cannot continue until the other group members finish their respective parts, given my next duty is the cleaning up of, and organization of the visualization space, as well as completing the map legend.

Dropdown List and Pie Chart
[entry by antonybett, 04.16]

Utilizing the ControlP5 library, I managed to add a drop-down list to the visualization that will

enable the user to filter earthquake events based on a selected country. The drop- down list integrates well with the slider, in such a way that users are able to scroll/run through the occurrences within that specific country- this feature will become more apparent when we include a play button to the slider, since plotted points will be constrained within a given border, decreasing the scope of view while directing focus on the occurrences within the country of interest. Below is a snapshot - the position of the drop-down list may be altered as we add more features.

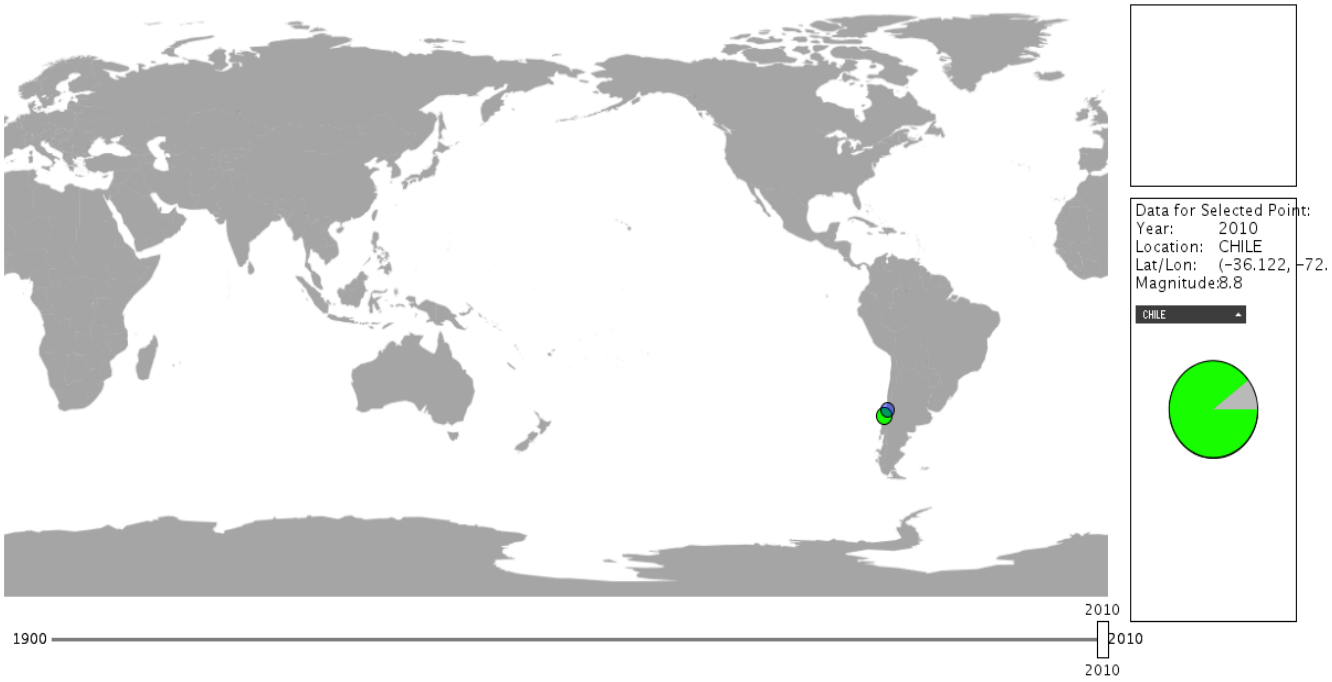
The drop-down list has been ordered alphabetically, making it easier for the user to scroll down to the country of interest.



In addition, since we are now able to filter the data in terms of the affected countries, we can draw comparisons between the selected earthquake event and other earthquake occurrences within that specific country. I added a pie chart to relay a sense of how disastrous the selected

earthquake was based on the damages caused, as compared to the total economic damages suffered by that specific country between 1900-2010. A second pie-chart could be added to depict the comparison of a selected earthquake event with other occurrences within that same year.

The pie-chart, as mentioned in lecture, has its own limitations hence it calls for prudence if we actually end up utilising it. The bar chart will thus be of higher priority since it will be more effective in depicting the trends over the years.



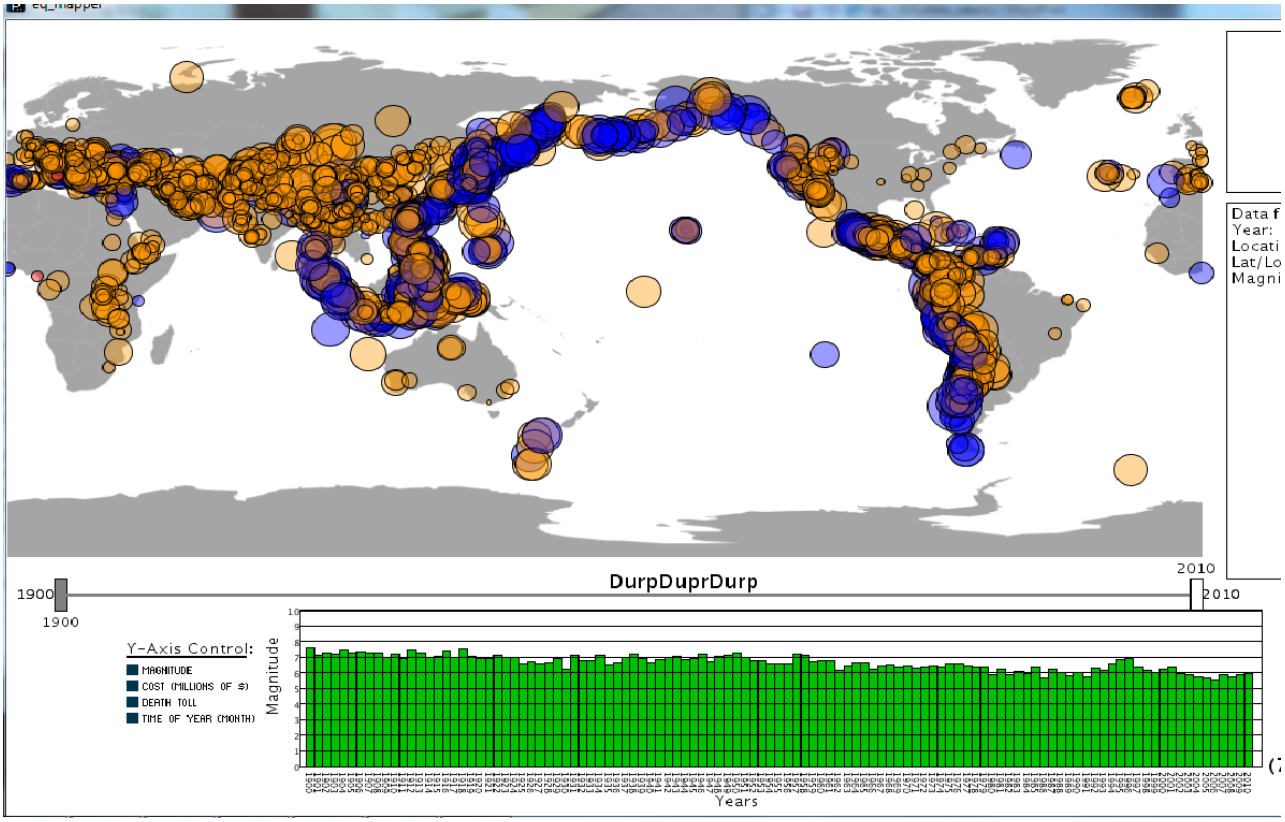
Degrees: 322.95908

(867, 580)

Prototype Bargraph
[entry by gtrevino, 04.16]

To provide the user with a greater ability to see trends in the data across different years, I created a bar graph for our visualization. Using a secondary data file created by antonybett in which the averages and totals of the primary data file were calculated per year, the bar graph would provide the user with insight into the average magnitude of earthquakes, average cost of the earthquakes (in millions of dollars), total deaths caused, and number of earthquakes recorded, for every year for which we have data. It was also decided that we would continue forward with our initial idea of using radio buttons in order to control the bar graph, such that the various types of data graphed along the y-axis could be changed by the user.

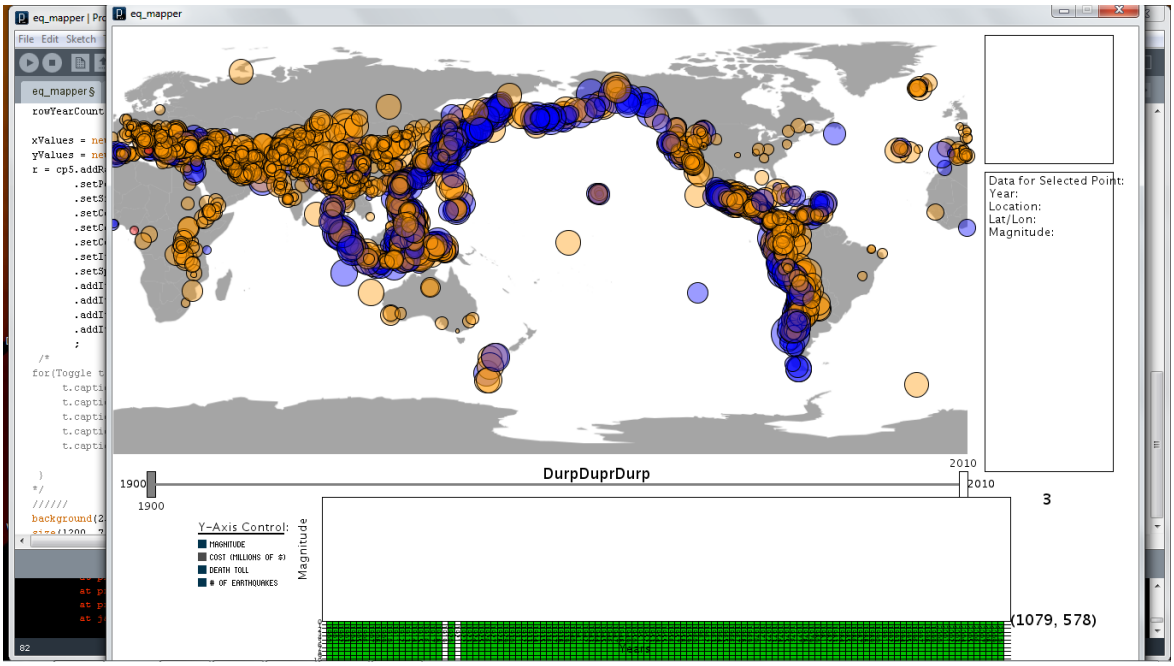
The initial prototype of the bar graph, i.e. the first version that could successfully read in the data file and accurately scale and draw the bars, is shown below:

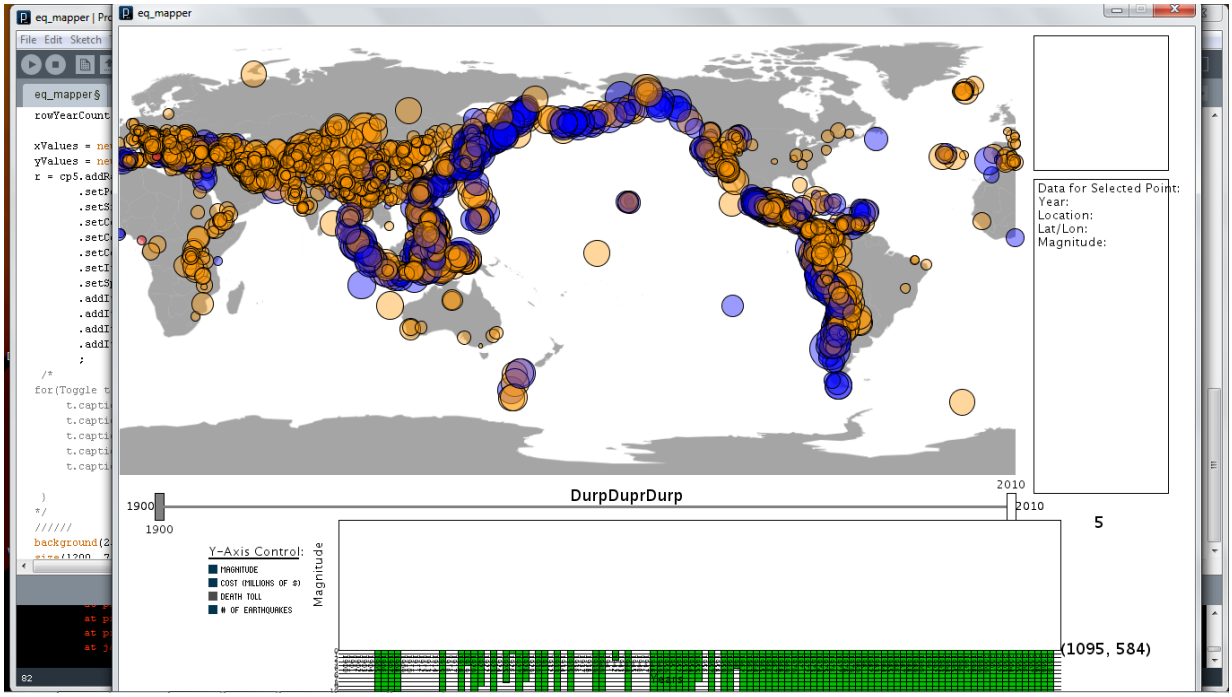


Do note that this is not our final color scheme, nor placement of these features. We also have the initial radio buttons, though they currently lack functionality.

Functional Bargraph and Radio Button Control
[entry by gtrevino, 04.18]

Using the ControlP5 library, we then implemented radio buttons as a user control for the y-axis of the bar graph. Although utilizing ControlP5 and the radio buttons were not completely painless, I was eventually able to get them working such that they responded to user input (either clicking on the radio buttons or using the number pad, #1-4), thereby changing both the value being graphed on the y-axis and tick marks and sizes of the bars as necessary. However, while the radio buttons were able to change all of these things as specified, it was then required that we correct the graph code such that everything scaled correctly. The initial errors we had with proper scaling can be seen below:



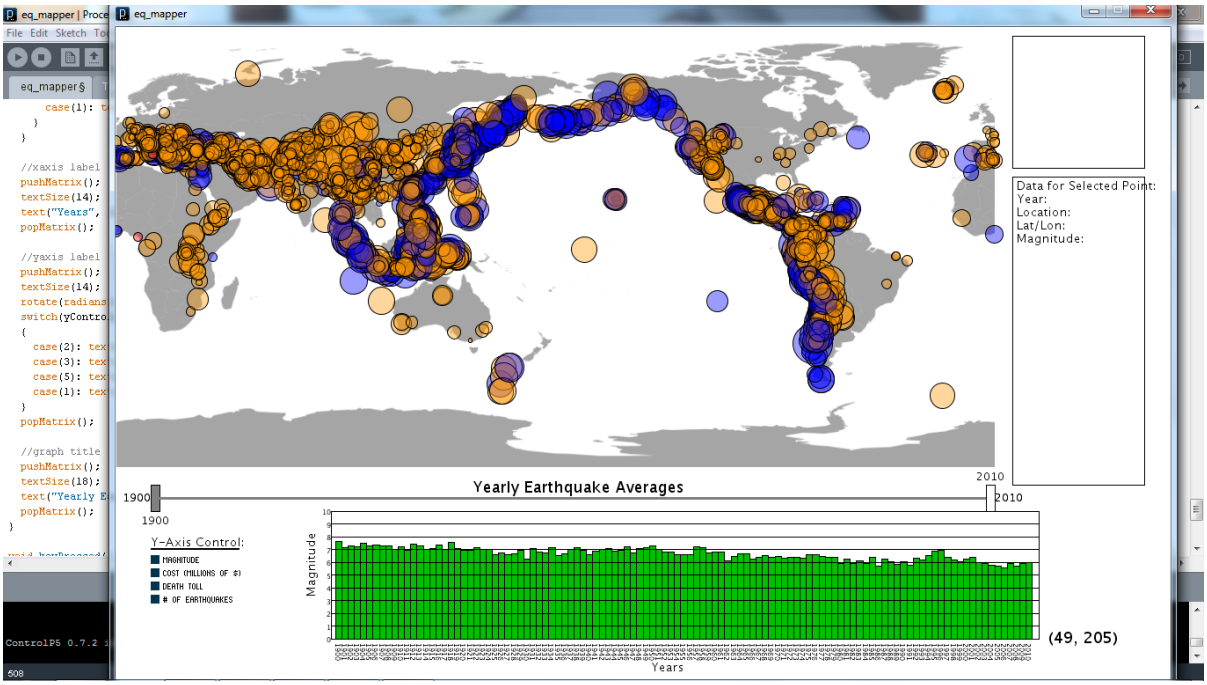


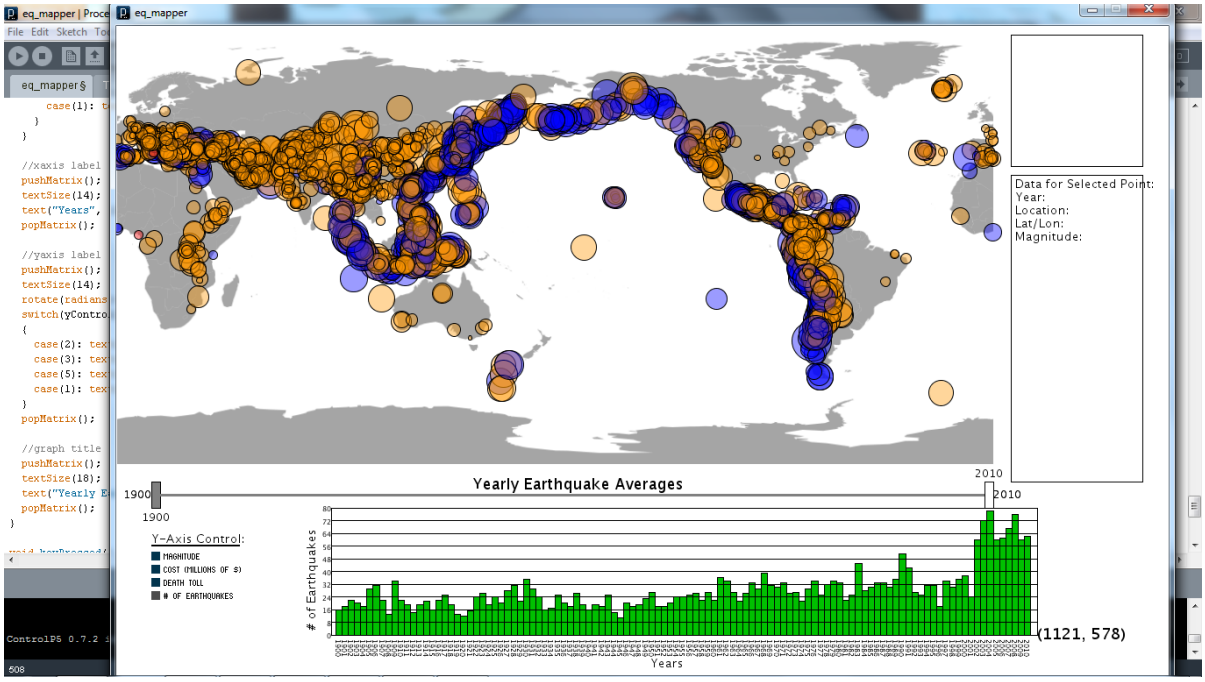
As is readily obvious, although the y-values were changing appropriately, the rest of the graph was yet unable to properly scale the different types of data such that everything would be clean and neat, as well as properly labeled.

Nonetheless, the ability of the radio buttons to result in noticeable change based on user input was a very big step in finishing this feature of our visualization.

The biggest challenge in correcting our scaling of each y-variable was avoiding the hard-coding of certain variables so as to ensure our visualization remained dynamic, while still implementing a clean fix to our issues. In particular, we wanted our bar graph area to remain a consistent size, while still being able to accurately graph the different data ranges (some varying as little as ~2.5, other varying as much as 300,000) for each y-variable.

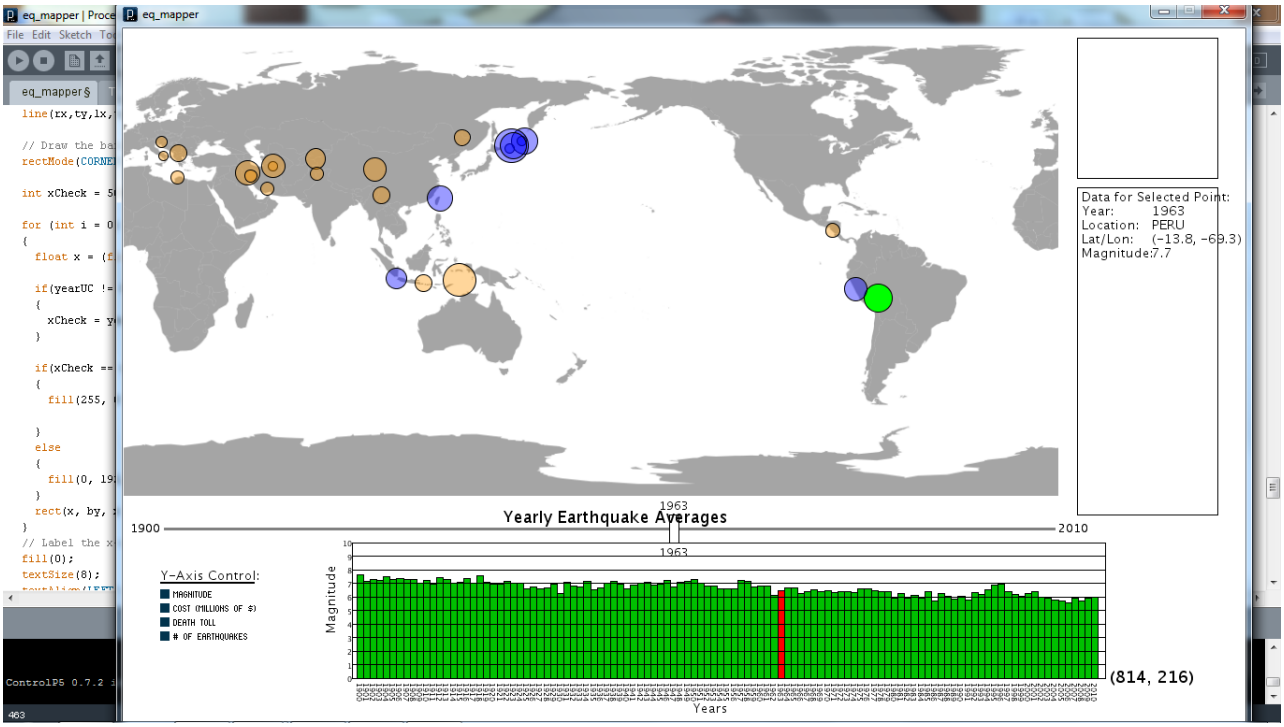
Examples of the corrected feature, with scaling bar graphs and radio button interaction can be seen below:

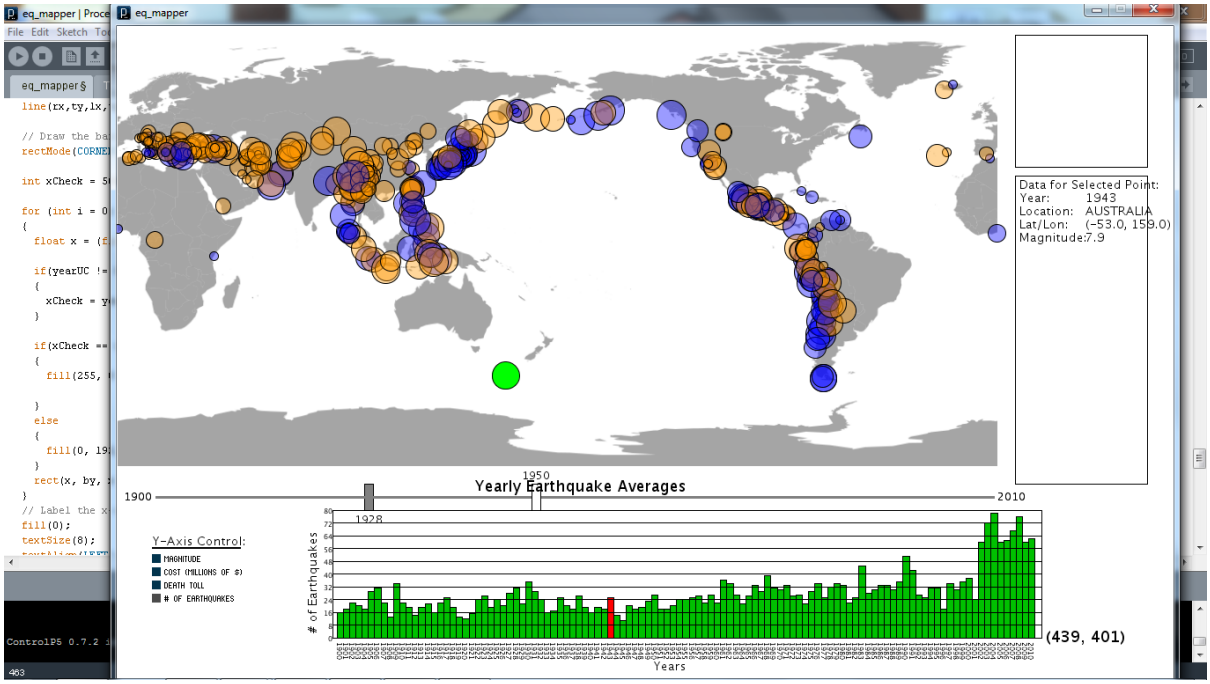

eq_mapper
Data for Selected Point:
Year:
Location:
Lat/Lon:
Magnitude:
Yearly Earthquake Averages
1900
2010
Y-Axis Control:
MAGNITUDE
COST (MILLIONS OF $)
DEATH TOLL
OF EARTHQUAKES
Magnitude
Years
(49, 205)

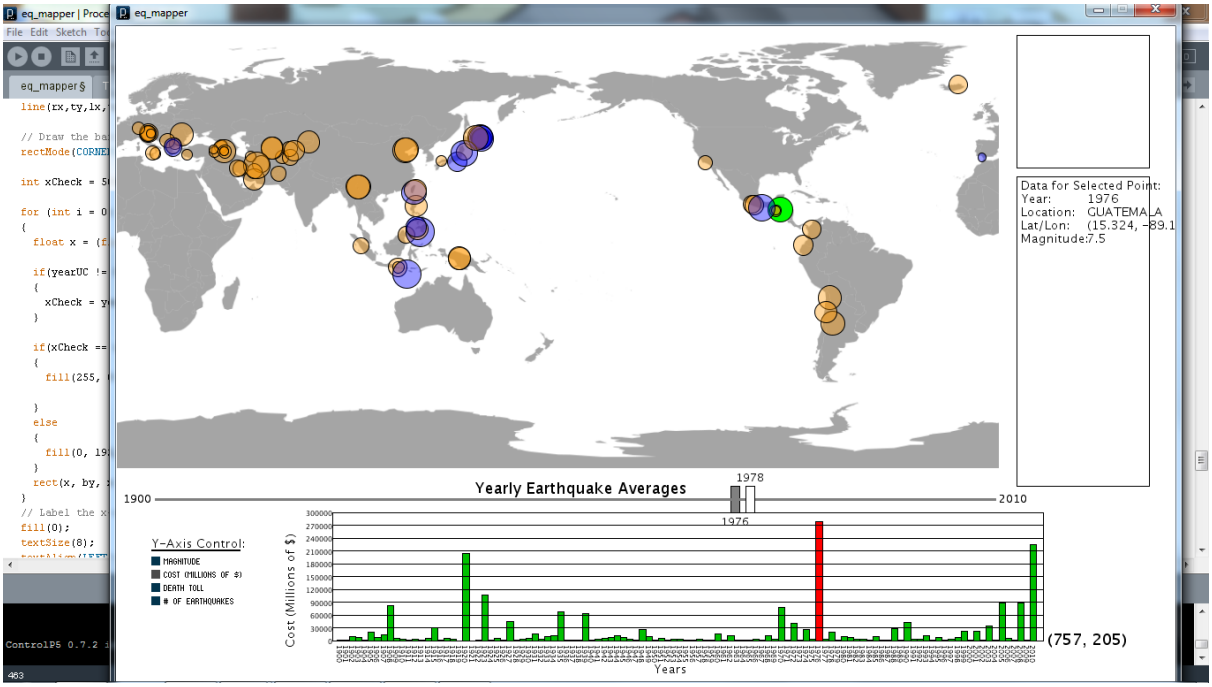

eq_mapper
Data for Selected Point:
Year:
Location:
Lat/Lon:
Magnitude:
Yearly Earthquake Averages
1900
2010
Y-Axis Control:
MAGNITUDE
COST (MILLIONS OF $)
DEATH TOLL
OF EARTHQUAKES
of Earthquakes
Years
(1121, 578)

Bargraph Compatibility with Data Point Selection
[entry by gtrevino, 04.19]

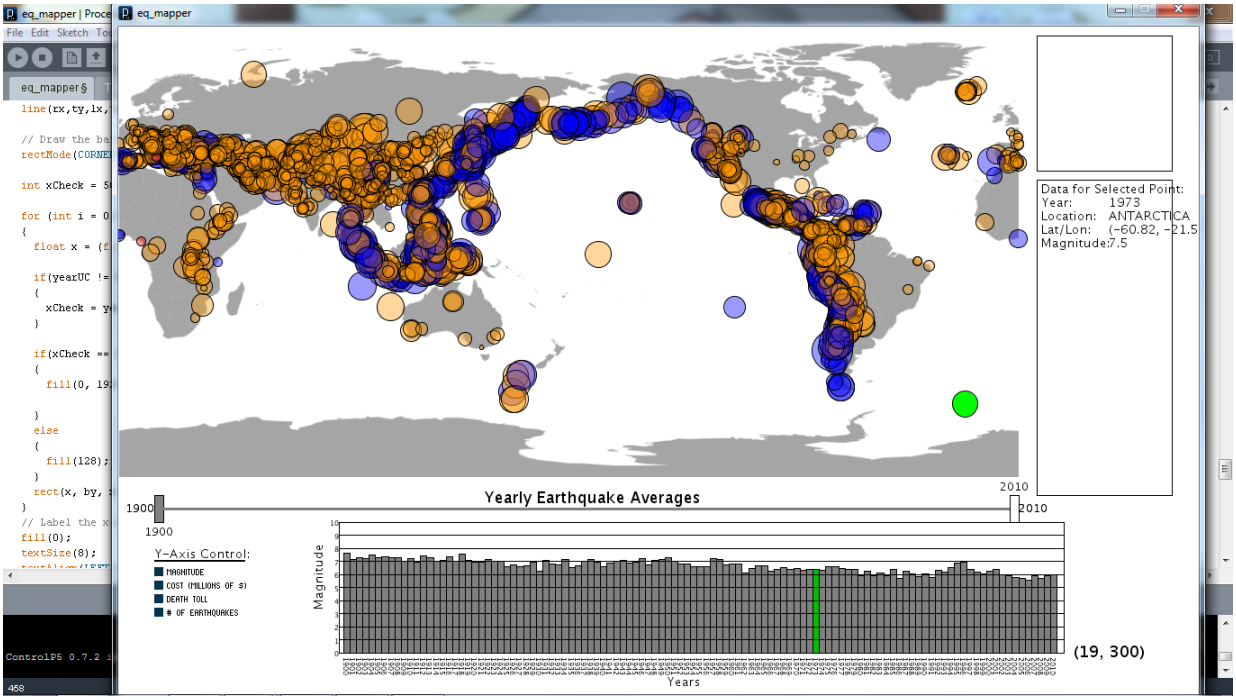
Now that our bar graph had full radio button functionality, there remained only one major feature of the graph to implement: the compatibility of the graph with the user interaction involving the mapped data points. Currently, when a user clicks on an particular data point, all the relevant information about that point is displayed in a table to the right of the map. To increase usability and comprehension, we thought that it would be ideal if, when a user clicks on a data point, the corresponding bar of the year of that data point was highlighted, such that the user could compare the information in the data table to the averages displayed on the graph. This feature can be seen below:




eq_mapper
Data for Selected Point:
Year: 1943
Location: AUSTRALIA
Lat/Lon: (-53.0, 159.0)
Magnitude:7.9
Yearly Earthquake Averages
1900
1950
2010
1928
Y-Axis Control:
MAGNITUDE
COST (MILLIONS OF $)
DEATH TOLL
OF EARTHQUAKES
of Earthquakes
Years
(439, 401)


eq_mapper
Data for Selected Point:
Year: 1976
Location: GUATEMALA
Lat/Lon: (15.324, -89.1
Magnitude:7.5
Yearly Earthquake Averages
1900
1978
1976
2010
Y-Axis Control:
MAGNITUDE
COST (MILLIONS OF $)
DEATH TOLL
OF EARTHQUAKES
Cost (Millions of $)
Years
(757, 205)

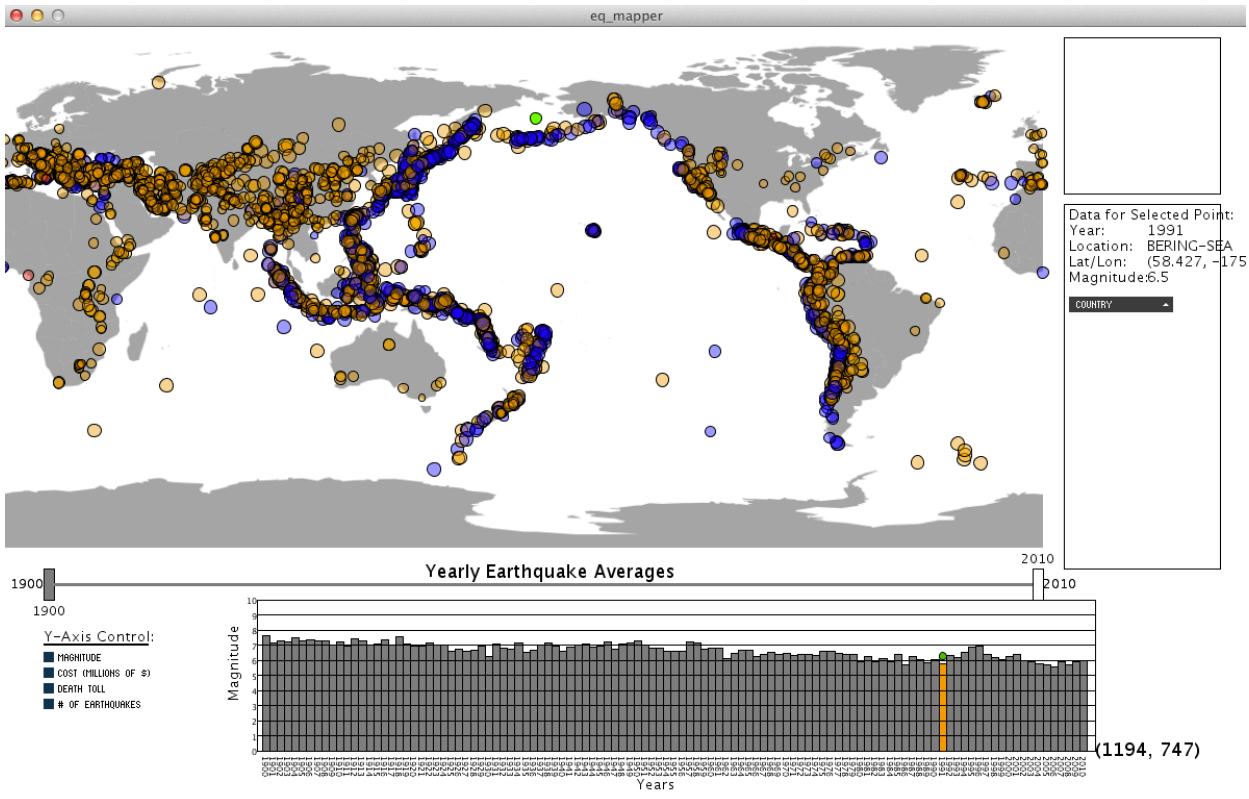
Finally, to make our color choices consistent and make it obvious that the year highlighted in the bar graph corresponded to the data point selected, we changed the bar graph as follows:



Final Touches to Bar Graph: Awesome Features
[entry by gtrevino 04.19]

Since we had sufficient time left before we had to hand in our already awesome project, we decided to add some extra cool features to our bar graph.

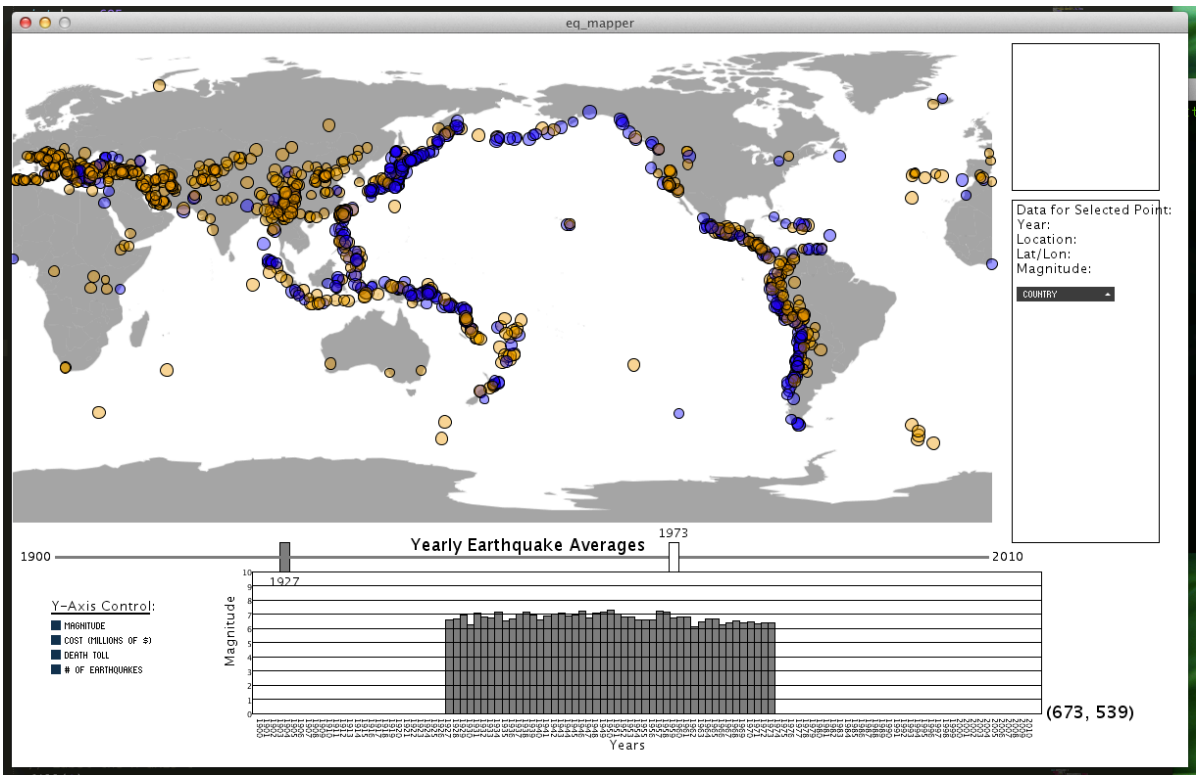
First, we thought that while the highlighting of the bar corresponding to the data point selected was nice, that it would be even better if the corresponding data about that data point was also graphed. That way, the user would be able to see where that particular data point measured up against both its year's average, and every other year's average. The implementation of that is as follows:



Note that we had to once again change our color coding to increase usability and intuitiveness.

Furthermore, we thought that it would be excellent if the bar graph reflected the range of years as selected by the user per the slider. That way, the user could view the data as densely (or sparsely) as they wanted.

An example of the implementation of this feature is below:



In this way, we could differentiate to the user clearly the range of years and data they had chosen, without having to add any excess color changes or highlighting, resulting in an overall cleaner user interface.

Slider: Play Functionality
[entry by antonybett, 04.19]

To make the slider more interactive and more user friendly, we decided to add a play functionality to the slider. This will enable the user to run through the earthquake occurrences by clicking on a play button or pressing a key. The play feature utilizes the global yearMin and yearMax variables together with the frame count to run through the years. The play functionality supports user selection for both the start year and the end year, giving the user more control over what range of years to play through. The space bar acts as the play button, and the user can pause by pressing the backspace key.

An interesting feature is the integration of the play functionality with the bar chart, in such a way that the year on the slider coincides with the active bar on the bar chart, with the movement of the active bar at par with the slider movement.

Final Visual Touches and Tweaks, Legend, More Information
[entry by btwalsh, 04.19]

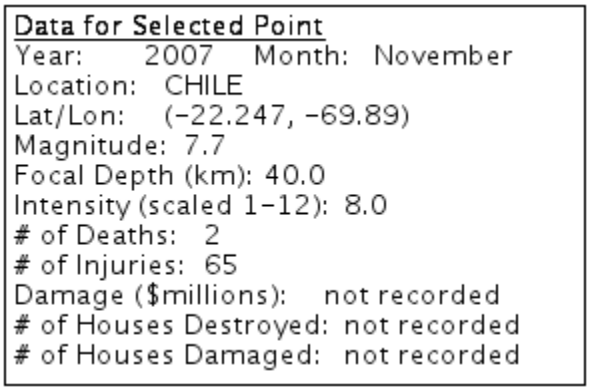
At this point in time, essentially all of our coding is now complete, and the goals for the final night of working on this project as a group include, first and foremost, merging the additional work of antonybett and gtrevino to avoid conflicts, and then ultimately merging their files with my code in our git repository. Of course, this also meant spending some time with the git mergetool to resolve conflicts amongst the various versions and edits of certain PDE files.

Once this merging was complete, we were left with our visualization looking essentially like it does in the screenshots in the previous entries in this process book by gtrevino, the only difference being the additional ControlP5 presence of a dropdown menu to select countries to provide additional filtering on our dataset.

Then came the need for beautifying our merged code, in terms of the code and comments themselves in addition to the appearance of the visualization that was produced when the project was compiled. First, I combed through our code, line-by-line, to improve existing comments, add new ones where necessary, and to remove redundant or simply unnecessary code. Obviously, these changes made little functional difference in our project's appearance on the end-user side of things, though my revisions here make a world of difference for anyone who wishes to read or otherwise sift through our thousand or so lines of code.

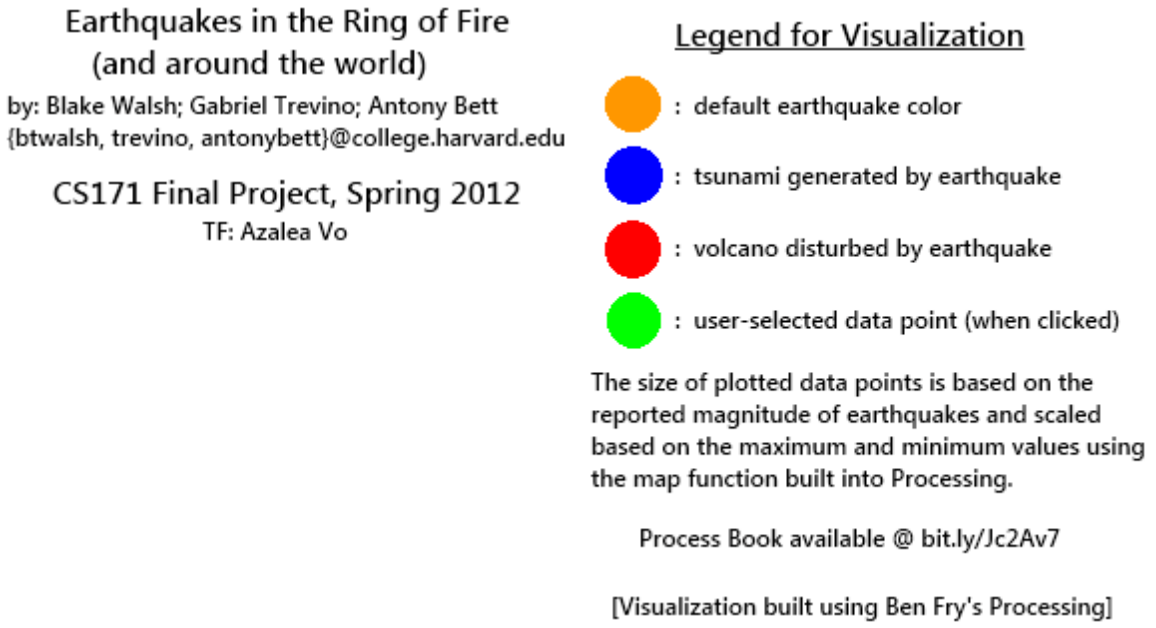
In terms of visual changes, the first major change involved lining up the below bar chart (originally worked on by gtrevino) with the time slider, which makes perfect sense given the x-axis of the bar chart also pertains to year. Dealing with the rotation of text here via pushMatrix(), rotate(), and popMatrix() provided to be slightly less than intuitive, though ultimately did not take that much time to adjust thanks to gtrevino's expertise with aforementioned functions. We then moved the "Y-Axis Control" radio buttons from their original location on the left of the bar graph to the bottom-right side of our visualization.

Next, I updated the text table on the right column that responds to a user-selected point by displaying detailed information (thanks to some well-thought out output-formatting!) from our original data file about that specific point in a manner as below:



Aside from also making antonybett's dropdown menu for nation selection mesh more nicely with our visualization interface, I set out to make a nice looking legend and title. I decided it would be easiest to create these as images, and simply display them in the appropriate

locations in Processing, given neither the title nor the legend would need to change as one uses our visualization. I made quick image mockups for a title and legend for our visualization in Paint.NET, per the sketches by gtrevino, and they appear as follows:



To add a few final touches for uniformity in our visualizations appearance, I also added a thin black border around the map of plotted points, which we have been lacking until now.

Also, as required according to the CS171 website, I have composed a README.txt file that sits in the root directory of the submitted ZIP file of this project to explain in depth its file hierarchy and how compilation and running works.

FINAL PRODUCT
[entry by btwalsh/gtrevino/antonybett, 04.20]

Note: the entirety of this portion of our process book can be found above in greater, chronological detail, and this section serves to be an efficient assemblage and otherwise summary of how our work has progressed over the duration of this project.

Description of Problem
[entry by btwalsh, 04.20]

As discussed in the initial planning stages of this project, the intended audience for this visualization is primarily the general population, though perhaps additionally for the more specific subsets of governments, insurance companies, and more. The data conveyed in the project, acquired from the NGDC-NOAA (National Geophysical Data Center - National Oceanic and Atmospheric Association), include magnitude, intensity, and location of earthquakes in addition to economic damage caused, fatalities, injuries, and more. The analysis tasks for this project include value retrieval, comparison between relative values for one earthquake and another. In terms of filtering earthquakes, we should ignore earthquakes below a certain magnitude to avoid noise, and then give our time span a reasonable scope.

Related Work and Inspirations; Problem Abstraction
[entry by gtrevino, 04.19]

With the increasing reach of global media, now more than ever is there an increased awareness concerning the calamities and disasters present in our natural world. In recent years especially, there have been several natural disasters that have rocked the public consciousness. While we originally began our project with the intention of mapping all kinds of natural disaster data, we realized that given the wealth of information available, it would be optimal to focus on one type of disaster.

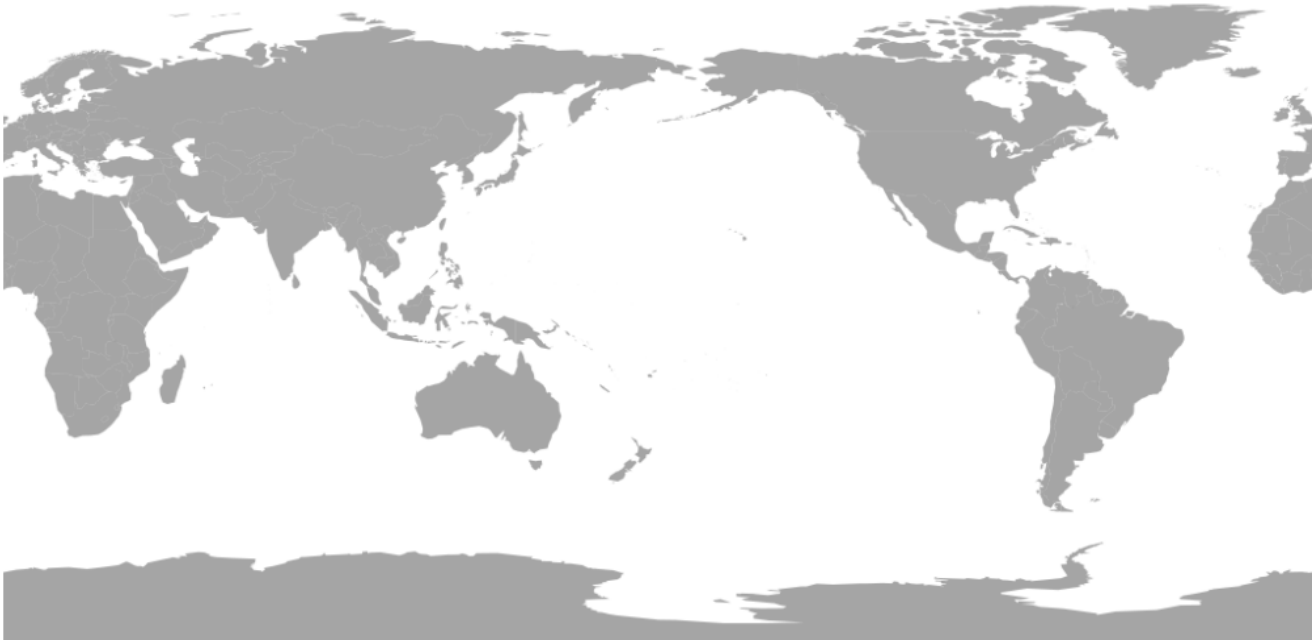
Each of our group members has experienced an earthquake firsthand; furthermore, earthquakes have recently shown themselves to be monstrous disasters such as the recent earthquakes in Japan and Haiti. Thus, it made perfect sense for us to focus our visualization on earthquakes, a natural disaster with tons of data behind it and a lot of relevance in the public sphere. Ultimately, it is this relevance that is at the heart of our decision, as we wanted our visualization to not only be a mere final project, but a tool for data analysis that the general public would find interesting, inventive, and informative.

Design Evolution
[entry by antonybett 04.19]

Map
When we set out to work on the project, we envisioned a graphical visualization that employed the use of a map. The map is the most effective and efficient interface given that we were working with data that had to be geographically represented in terms of latitudes and longitudes,

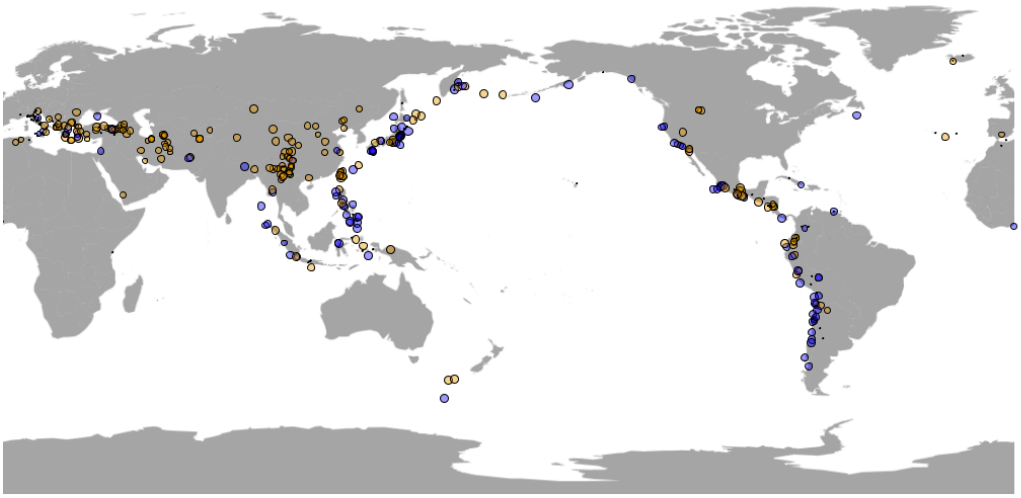
hence the need to make the most out of the visual power of the map in accurately displaying this spatial distribution. Initially, we had in mind a map of the Ring of Fire - paying more attention to this region since, as the most affected region, it acted as the main motivation for the project, the central region of focus, and also due to the fact that there would be an abundance of data to work with. As we progressed, we made a slight alteration to the map layout; instead of solely representing the Ring of Fire, we decided to increase the scope of view to include a map of the world. Nevertheless, we maintained the Ring of Fire as the central region of focus.

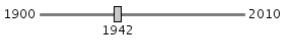
To represent this region accurately as the central area of focus, we needed a slight modification in the projection of the map, so that the Pacific Ocean and Ring of Fire are visually centered on the map. This was accomplished using a helper function ("WorldMap"), that translates the latitudes and longitudes from a 3 dimensional world map and maps the respective values into a 2 dimensional map layout. In addition, the "WorldMap" class file takes into account the warping that usually occurs at points near the poles by converting latitudinal and longitudinal coordinates to x-y pixel coordinates for an equirectangular projection, ensuring that such points are accurately plotted on the map. This compensation for the warping goes a long way in maintaining the graphical integrity of the visualization. Below is a snapshot of the final map layout, showing the map's adoption of the Pacific Ocean and Ring of Fire as the central region of focus:



Slider

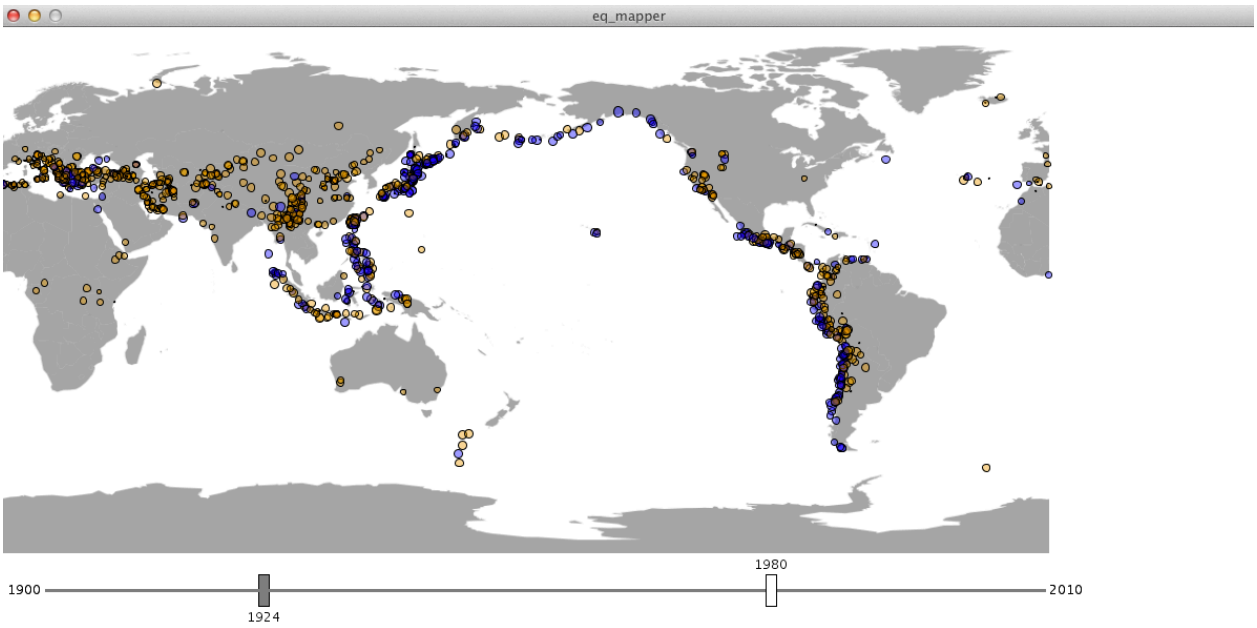
The slider has gone through a number of modifications. In the earlier stages, the slider was implemented as two distinct sliders: the upper slider for the upper bound of the range while the bottom slider to indicate the lower bound. Below is a snapshot of the initial slider structure, showing the implementation of both the upper and bottom sliders:





(529, 556)

Once the functionality of the slider was successfully implemented, we did away with this rudimentary structure and replaced the two single-sliders with a single double-slider. This was achieved by superimposing the two sliders in such a way that there was singularity both in their structural appearance and even more importantly, in their functionality - the lower bound, for example, cannot overlap the upper bound, either physically or in value. Below is a snapshot of the double-slider:



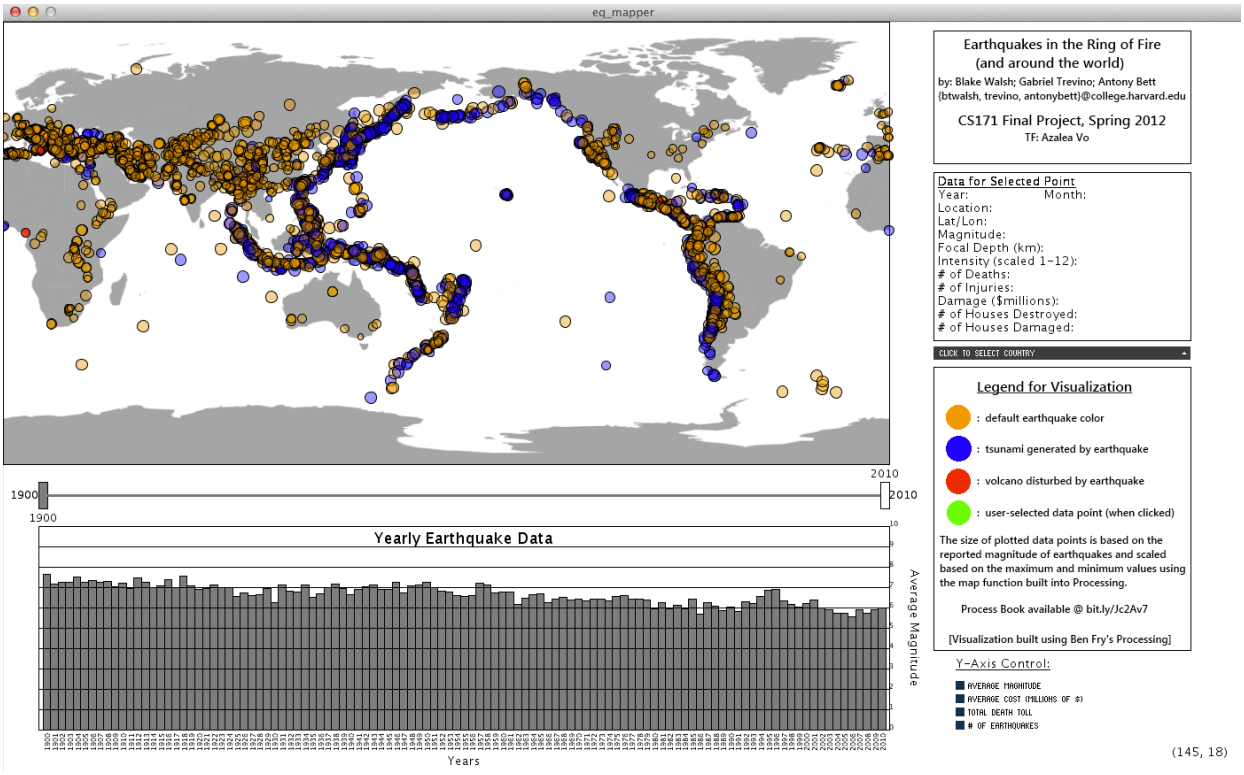
(598, 352)

The next transition in the slider was the incorporation of a play functionality, where the user presses the spacebar to move through time as the earthquake events unravel on the map; the backspace key is used to pause. In addition, the play functionality of the slider integrates uniformly with the bar chart, such that the active bar on the bar chart is set to move in unison with the the current year running on the slider.

In general, having a solid sketch at the beginning of the design process proved extremely helpful in guiding how we went about the transition from the various design modifications.

Final Design
[entry by antonybett, 04.20]

Born out of our initial sketch, the final design layout represents work that strives to meet all the design principles.



C.R.A.P. Principles

In choosing the map projection, we settled on this particular map since its white background meant that it would Contrast well with most of the colors for the plotted points. The placement of the map and the bar chart ensures that there's Alignment in their positioning. In addition, the length of the slider was set in such a way as to run through the entire width of the map, keeping the slider in alignment with both the map and the chart. On the left hand side, there's perfect Alignment in the positioning of the text boxes. We decided to place the slider between the map and the bar chart since, as much as it affects both of them, the central positioning preserves its Proximity to both of these elements.

Graphical Integrity

The initial stages of the design process had great bearings in determining how successful the visualization would be in maintaining graphical integrity. To start off, we had a function that ensured that the X-Y mapping of the points on the map reflected accurately the latitudinal and longitudinal positioning of these points.

Visual Encodings

- Ellipse Size - For the plotted points, the size of the ellipse encodes for the magnitude of the earthquakes. Initially, we envisioned giving the user the ability to change the criteria for the ellipse size / selecting the field to encode. However, after working with the data for sometime, it came to our realisation that most of the fields other than magnitude had most of their

attributes missing / with null values or the range of the values was extensive enough that some points would be insignificant when mapped. We therefore selected to use magnitude to encode for the size of the ellipse.

- Color of Plotted Points - In addition to earthquake events, the plotted points also represent occurrences where volcanoes or tsunamis have been generated by or related to an earthquake. We used different dolor encodings for the ellipses in order to differentiate among the three possibilities, namely sole earthquakes, earthquake in tandem with volcanoes and earthquake-generated tsunamis. Red represents earthquake generated volcanoes, blue represents earthquake generated tsunamis while orange represents sole earthquake events. When a user clicks on a point, the selected point is represented by green color. The bar chart with the year coinciding with the user selection on the map is further highlighted on the chart. An ellipse is plotted on the bar to depict how the selected point on the chart relates to other to the total or average for that particular year. Both the ellipse on the map and the ellipse on the chart are encoded using green color; this consistency in the use of the green color augments the interpretability of the chart and the selected point.
- Scale: We adopted a consistent scale for the bar chart that ensures that there's no distortion in the representation of the data. Given that we were plotting more than 100 values on the X-axis, it was deemed fit to rotate the X-axis labels. This not only maintains the aesthetics of the chart, but also amounts to good design since it preserves the vertical alignment of the chart with the map.

Analysis of Data and Conclusions
[entry by btwalsh, 04.20]

As indicated elsewhere in this process book, much can be gathered from this visualization, particularly in its user interactivity. Based on the types of data filtering done by the user, a variety of conclusions can be drawn depending on whether the data is narrowed by years, countries, region, any characteristics based on the bar graph, or more based on the decisions made by the user. But just from glancing at our visualization without interacting with it, there is not too terribly much that can be concluded given the sheer quantity of data being represented - namely, all recorded earthquakes that have occurred anywhere on the globe in the past 110 years!

Of course, one of the more obvious conclusions that can be drawn is the "Ring of Fire" surely does exist, given the high amount of earthquake, volcano, and tsunami activity that takes place on the tectonic plates that form the approximate perimeter of the Pacific Ocean. This was of course the original inspiration for this project, though it is incredibly edifying to see this concept come into fruition with the success of our visualization.

It should be noted, however, that it initially appeared based on our visualization that more relatively larger earthquakes took place in the early 1900s compared to those that occurred more recently, though further thought on why this might be the case leads us to conclude that this is likely the residual result of the fact that the instrumentation with which earthquakes were measured back then was no doubt less precise, which may lead to decreased reliability in our dataset as one looks further back in time.

Team Evaluation
[entry by gtrevino, 04.20]

Blake Walsh: Over the course of the project, Blake acted as the initial programmer and scribe to the process book. For the first half of this project, the entirety of our code existed on Blake's computer, as he was responsible for most of the initial developments, such as mapping of the data points, the creation of our prototype slider, and much of the groundwork code that our later features would rely on. In terms of major program features, Blake was also responsible for the user 'click' interaction with the data points. Beyond the code, he was responsible for organizational breakthroughs such as the slick look of the process book, the selection of the equirectangular map projection, and the ever important establishment of our project Git Repository. To top things off, Blake also created the data table showcasing specific data point information, and was responsible for our lovely legend and title.

Gabriel Trevino: Initially, Gabriel was responsible for the initial sketches that would come to act as the blueprints for our project. Together, he and Blake were able to establish such features as the creation of the double-slider, and the perfected plotting and scaling of data points on the map. His major contribution was during the 'second' half of the project: by following Blake's establishment of the vital groundwork for the project, Gabriel was responsible for the entirety of the bar graph and its various features. Those features, summarized in earlier sections of this Process book, stand as key aspects of the project's overarching theme of education through interaction, providing a vital layer of data presentation and usability.

Antony Bett: For the initial half of our project, Antony was responsible for our data acquisition and clean-up. Beginning with just under 2200 bins of data in our unmodified, raw data file, Antony was able to reduce this down to a mere 180 bins in our final implementation. In the second half of our project, Antony was responsible for the creation of our secondary data files concerning year and country, as well as the creation of two very key features for our final project: the drop down menu to allow filtering by country, as well as the final glorious touch that is the "play slider" functionality which allows users to view each year in the currently selected range individually by cycling through them.

THE END - THANKS FOR READING!