CS-171 Final Project: Process Book



* Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

As an avid surfer, one member of our group had planned several surf trips before and was amazed by how hard it is to choose where to go.

When planning a surf trip, one needs to take into account several factors such as seasonal variation in tides, since surf spots tend to “go off” (surf slang for has good waves) only during certain periods of the year.

Depending on her surfing skill level, a surfer may also have a preference for a certain type of wave - whether fast and hollow “tubes” or “crumbly” soft waves – as well as wave height.

Surf stance also affects wave direction preferences, as it is much easier to surf front side, that is, facing the wave, than backside. This is why most “goofy” surfers (that surf with their left foot forward) will have a preference for left-hand breaks.

Novice surfers may also prefer to surf on beach-breaks rather than on reef-breaks (whether rocky or coral) since there is less chance of suffering scrapes and cuts on a sandy bottom.

This visualization aims to help surfers plan their next surf trip by allowing surfers to identify surf spots with their preferred characteristics that will be “going off” during their travel window.

* Related Work: Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

Inspiration for this visualization includes [Kayak’s Explorer](http://www.kayak.com/explore/), which allows an user to explore which destinations are within his budget and plan his trip accordingly.

Another source of inspiration is [Vilondo](http://www.vilondo.com/), a website that allows an user to search for villas in Indonesia by characteristics such as price, location and amenities such as swimming pool, air conditioning, etc.

* Questions: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

Our visualization allows a surfer to readily identify which surfing spots she should consider on her next surf trip according to travel window and desired characteristics. Say an intermediate surfer plans to travel during 2 weeks in May. This visualization provides a filtering tool for this surfer to identify to answer the following questions:

* Which surf spots have good surfing conditions in May?
* Which surf spots are appropriate for my surfing skill level?
* Which surf spots have wave characteristics that align with my preferences for wave size, power, direction and bottom?

Among the resulting filtered spots, a surfer can access further detailed information for a specific surf spot, including:

* What kind of surfing equipment will I need for this surf spot?
* Is this a crowded surf spot?
* Is this surf spot easily accessible?
* How consistent is the surf at this surf spot?
* What are the seasonal trends in air and water temperature?
* What is the current weather at this spot?
* What does this surf spot look like?
* Data: Source, scraping method, cleanup, etc.

The main data source is the <http://www.wannasurf.com/> website which lists surf spots around the world and their characteristics. We built a python web scraper to collect data from this source. The structure of the website is as follows:

1. <http://www.wannasurf.com/spot/index.html> contains a list of all countries with their associated links, if they have surf sports
2. These links give us country pages, which either contain a list of spots or a list of zones within the country and their associated links.
3. The zone pages contain either lists of spots or lists of subzones within the zones.
4. The subzone pages contain lists of spots.

We use the scraper to gather the links to all sites at each of these 3 levels, we classify them in order to identify whether a page contains links to surf spots, to a sub-level, or both. We then run a function (getspotdetails(spoturl)) on all pages containing links to spots in order to fetch all the necessary characteristics of each surfing spot.

The seasonal data is gathered at the subzone, zone or country level (at the prototype stage the seasonal data is fictional data).

The second data source is the wunderground API ([www.**w**underground.com/weather/api/‎](http://www.wunderground.com/weather/api/‎)), which we use to get weather forecasts for each one of the surf sport. At the prototype stage, this data was collected only for the first few spots. The rest is still dummy data.

Some of the surf spots on our source website do not have associated GPS information. In order to enrich the original dataset, we will in the final version geocode this spots using online geo-coding resources, such as the Google Maps API or Texas A&M geocoding tool.

Google images search API will allows us to included pictures of a selected surf spot.

* Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

The objective of our visualization was to provide a quick way for the users to explore surf spots and identify the ones that are the most appropriate for them. As such, it doesn’t really provide general insights applicable to all, but rather insights based on the preferences of the user. There must thus be two components to the visualization: 1) Filtering and 2) Detailed examination.

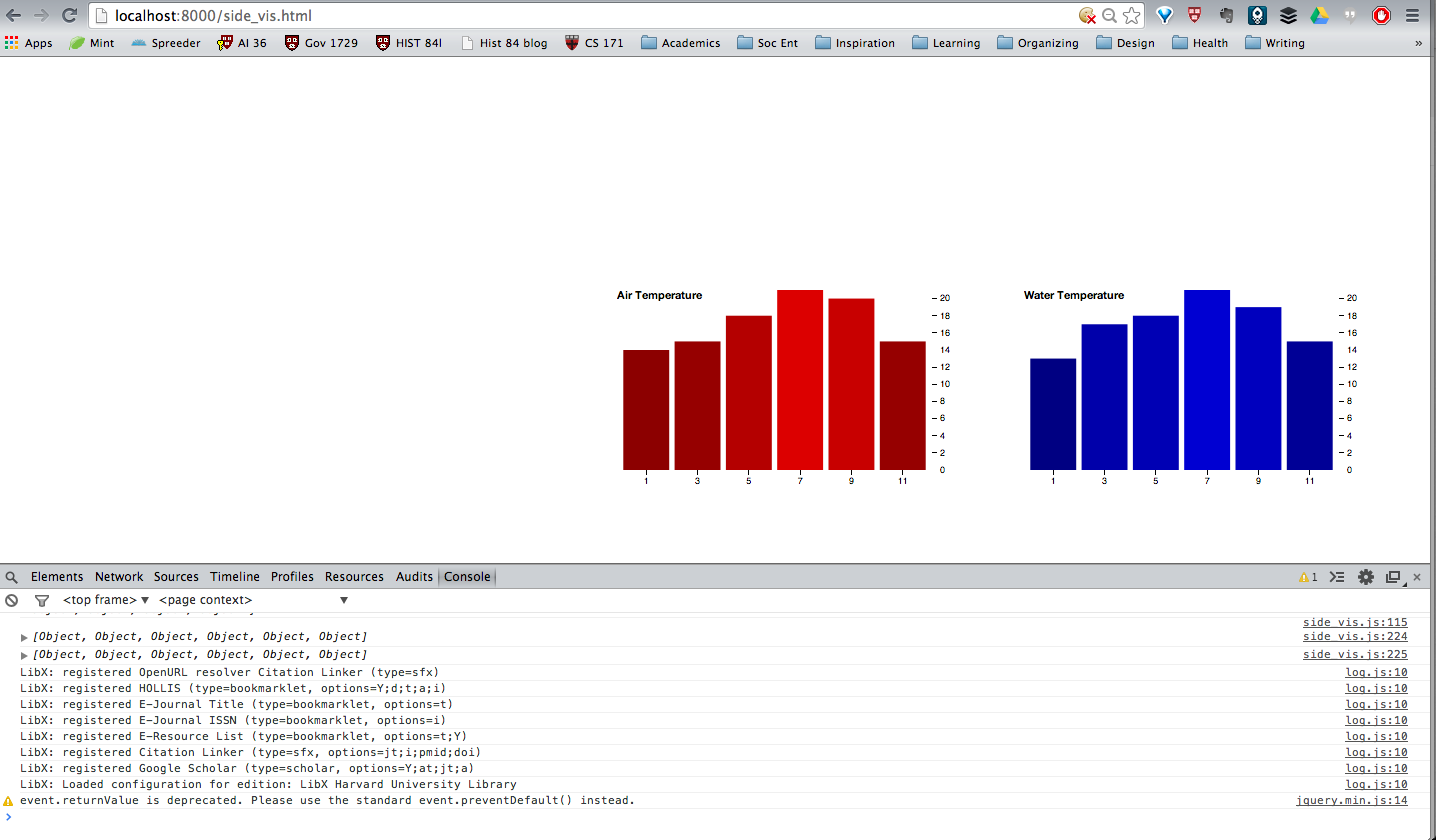
The filtering part must allow the user to exclude the majority of spots without the need to look at their detailed characteristics. Geography was seen as a good candidate, since it is likely that some (but not all) users would want to focus on a specific area of the globe. We decided that a zoomable map would give the user the most flexibility since we would not coerce the user to select countries or continents first, but instead gave the user the freedom to zoom in and zoom out however it pleased her/him. We then further imagined that the surf spots would appear as dots on the map, disappearing when they were filtered out. Other filters would need to capture the essential preferences of the surfer: wave height, wave power, direction, type of bottom (rock, sand, etc.) and experience level required. We initially thought of radio buttons for these filters, but then changed to checkboxes for the added flexibility of choosing several variable values.

The detailed examination part allows the surfer to get more detailed and visual information, such as the evolution of air and water temperature over the course of the year, walking distance, site description, etc. These are either more visual variables, or ones that are secondary in importance.

* Design Evolution: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course.

The core visualization is a map of surf visualizations, with side visualizations displaying information about each surf spot.

Our data is split up into quantitative and qualitative data. Because the quantitative data is relatively simple time series data of swell sizes, air and water temperature data, and ratings data of months, we decided on line graphs and bar graphs, which according to Tufte is the best way to represent low dimensional data. The air/water graph started two bar graphs (see below), but we thought comparison between the two was critical, so we converted it into a multi-line graph.



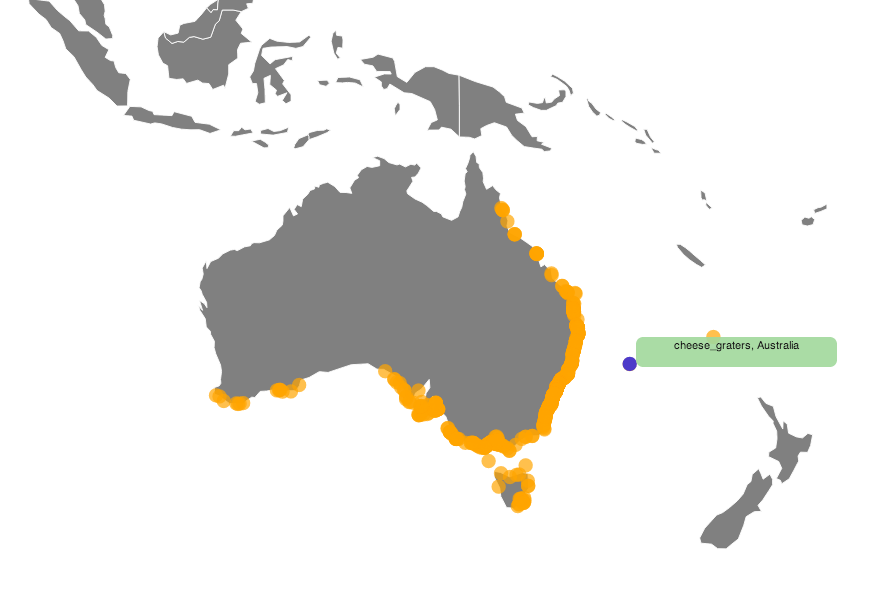
Our data is rich in qualitative dimensions, such as ideal transportation to the site, type of surf gear, bottom of the wave, etc. Because these are primarily ordinal data with few dimensions, we decided to implement filters to filter out the data based on user specifications. We are still working on a dashboard view to add the qualitative dimensions below the map.

* Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

Currently, we have 3 buckets of interactive visualizations: (1) map, (2) qualitative data filters, and (3) quantitative graphs.

**Map**

The intent is to visualize the location of various surf spots around the world, and allow filtering by various dimensions. We currently have zoom by click and zoom by scroll, tooltip, and the cursor tracks the surf spot you are currently mousing over. Here is a zoomed screenshot of surf spots in Australia:



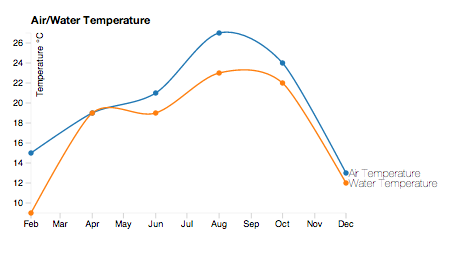
**Qualitative filters**

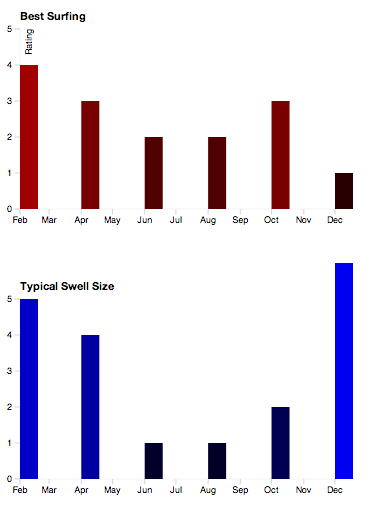
We are working on filters that allow a user to filter various qualitative dimensions on the map. Currently, the map only filters by skill level of surfer (see below), but we are also working on regional filters, wave characteristics filters (size, power, bottom, direction), and travel window.



**Quantitative graphs**

There are 3 quantitative graphs: (1) a multi-line graph representing air temperature and water temperature by month, (2) a bar graph representing ratings of best surfing by month, and (3) a bar graph representing ratings of swell size by month. Currently, these have minimal interactivity.





* Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

By visualizing our data, we realized that it’s very low dimensional, so it might be more useful to consider visualizations that engage in more comparison, such as a scatterplot that includes all of the surf spots in a region. We thought we did a reasonably good job of answering our questions, but we are still missing a portion that adds the qualitative data.

**Areas for improvement**

**1. Map**

First, we want to allow filtering by specific geographic region, which will zoom to the specific region and only show surf spots for that region.

Second, we want to size the circles on the map by different filters the user inputs. To do this, we will create an index variable that is a one/zero for each factor the user checks, and we will loop through the data sizing the surf spots based on how many of these criteria they meet. This is a way to keep all of the surf spots on the map but making it more clear to the user which ones are preferable.

**2. Qualitative data**

We have yet to add visualizations of the qualitative data, which we will likely add in a description box about the surf site, or in a recommendation type feed.

**3. Quantitative graphs**

We need to make it clear that the “best wave graph” is ordinal and not cardinal, by splitting up the bar graphs into small, discrete boxes with spaces in between them. We could also combine the swell size & best wave graphs into a multi-bar graph to save space.

With the air and water temperature multi-line graph, we could add more interactivity, such as hovering over the dots and seeing a tooltip of the temperature number.

As an additional feature, we are considering adding a scatterplot visualization such as [this example](http://nvd3.org/examples/scatter.html) on the NVD3 site, that visualizes multiple surf spots in a region by two dimensions, such as wave length and wave power. This would be especially useful in geographic regions dense with surf spots, such as Australia.