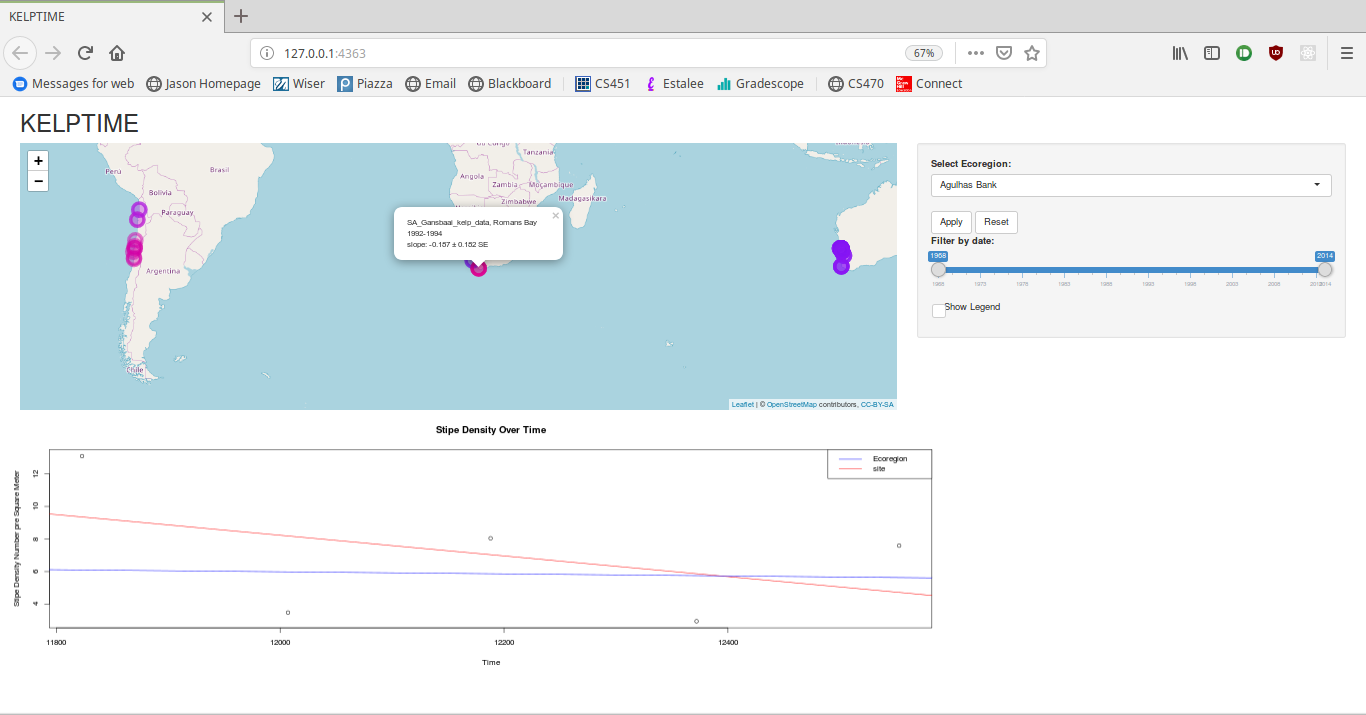
Kelp Time

A web application to better visualize massive collections of

data regarding kelp over decades of experiments

Nathaniel Namenyi

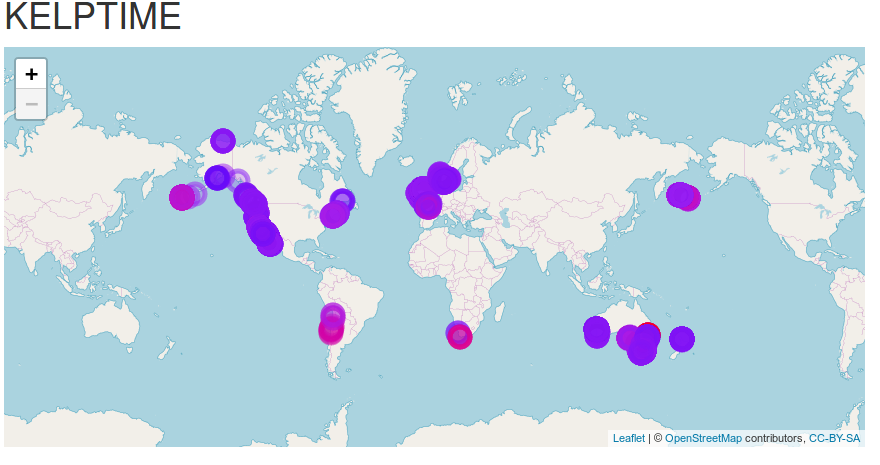
Jason Griffith

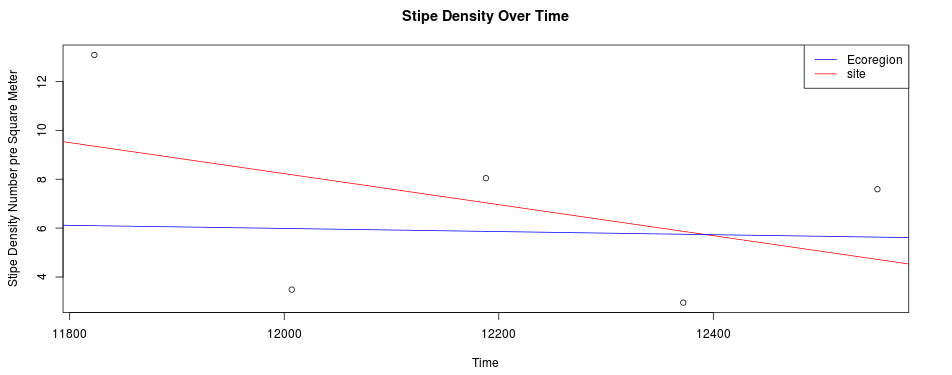
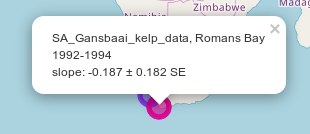
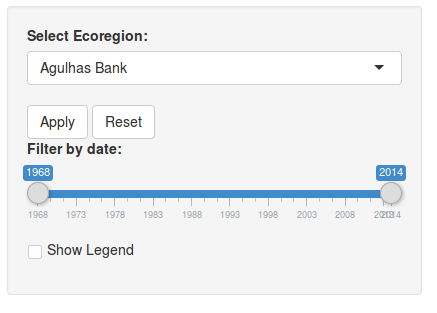


Introduction

Our project’s goal is to develop a web application in Shiny, leaflet, and R. This is being made in conjunction with Dr. Byrnes and uses his data gathered regarding his research in marine biology. The goal is to create a user interactive way to view the expansive dataset that may lead to new insights on data interpretation, or give viewers away to see some of the experiment data without having to know R. The primary window will be a map displaying the location the data points were taken, and potentially a heat map and allows the user to choose what parameters create the heat densities. There will also be an optional secondary window that will graph the data as either: points, lines, bar, etc based on what points the user has selected in the map window.

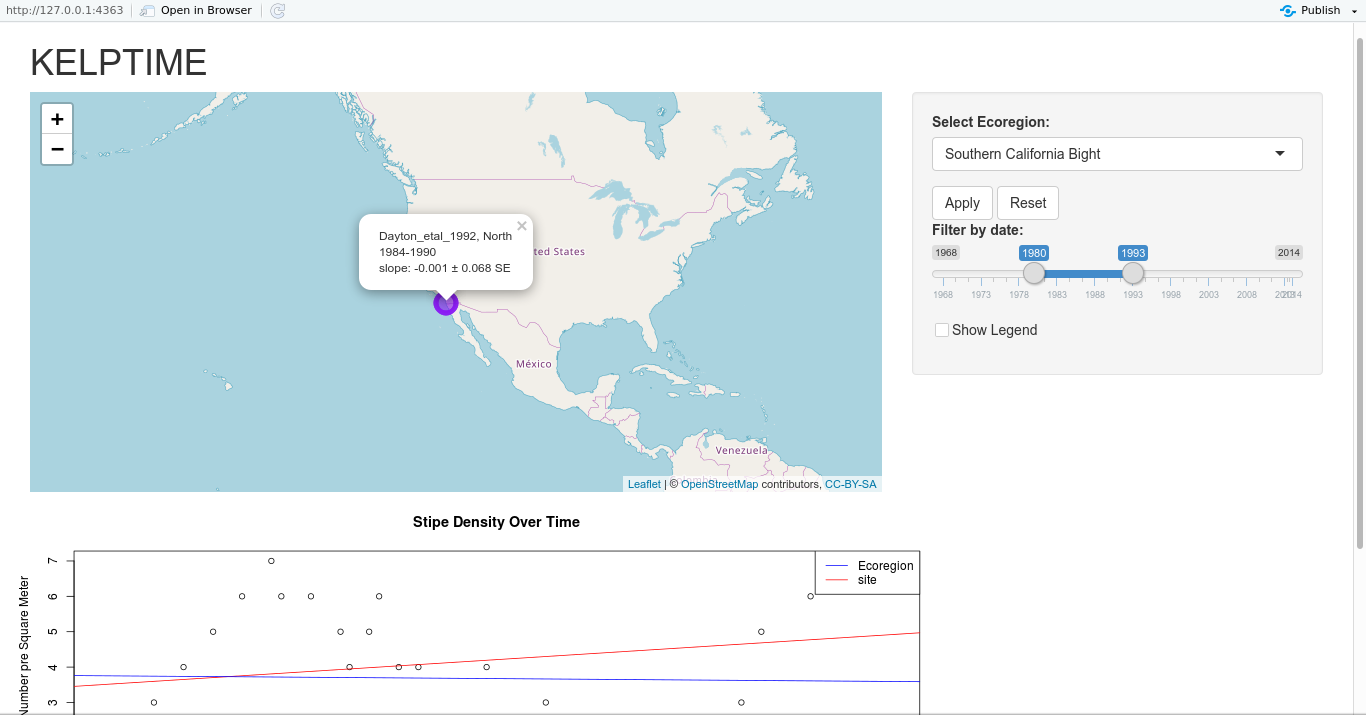
Current Progress

The application is still a work in progress with numerous features that can be implemented over time. Currently, what is displayed to the user, is primarily a map that can be explored by the user, displaying the locations of the experiments. Clicking on each point will display the name of the experiment, the location, the date range, and the slope (which to this point we are not sure how the slope is obtained; it is retrieved from from one of the datasets and calculated by one of the marine biologists) with standard error. Clicking on a point also brings up a graph below the map displaying time along the y-axis and and the stipe density per square meter along the x-axis, id displays the line of best fit for the points, and displays the line of best fit for the entire ecoregion to compare them. There are many points that do not contain data regarding the stipe density, and if one of those points are clicked, the the graph displaysa default graph and alerts the user that there is no relevant data. To the side of the map, there is a box containing the options the user can manipulate to change what points are plotted on the map. There is a dropdown that allows the user to select an ecoregion and, once they click apply, the map will only display points from that ecoregion. To display all of the points again, the user can click the reset button. There is a double ended slider in which the user can change the date ranges of points that are displayed, and also displays the current fate ranges being used and the available ranges. To change the date range, the user can drag the left side to the minimum date they want displayed of the start of experiments, and then can drag the right side to change the end date of the end of the experiment. The user can also change the minimum and maximum dates then drag the center of the slider to change the focus of the selected range. Finally, at the bottom of the options box, there is a check box the user can click to display a legend to show how the colors of the points relate to their slope. Currently, the legend will be removed if any of the display options are changed.

Data

The dataset we are using in the project is taken from Dr. Byrnes and his collaboration with other marine biologists and their studies on kelp over the course of many years. They have also incorporated many open source datasets into theirs giving us a wide range of data from as far back as the 19th century. The data can be found on Github:

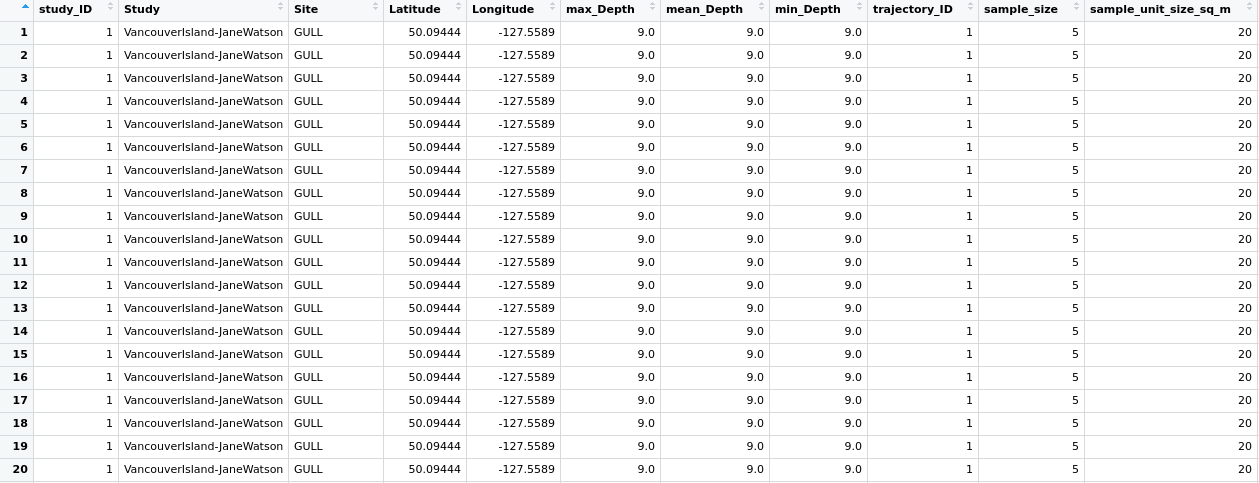
<https://github.com/kelpecosystems/global_kelp_time_series>



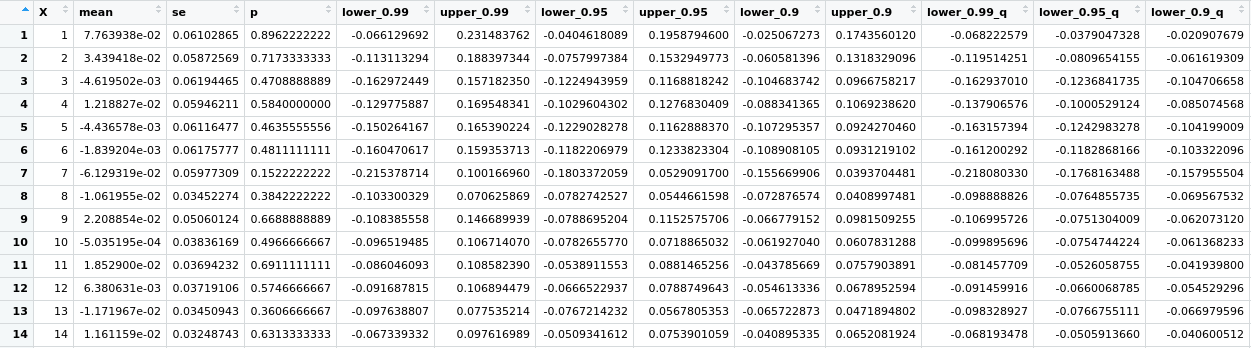
with filters applied

Data:

Raw data:



Organized by site:



From this data, we are considering a few different datasets, particularly the [CleanDataWithRegions.csv](https://github.com/kelpecosystems/global_kelp_time_series/blob/master/03_derived_data/CleanDataWithRegions.csv) which has the data nicely formatted making it easier to work with for our purposes. Plus, the addition of regions gives us another method to partition the data which Dr. Byrnes has suggested would be useful.

Our codebase is also open source, which can be viewed at <https://github.com/nnamenyi10/KelpTime>.

Planned Features

Despite many features implemented, there are still many features that we want to implement of think should be implemented, some of which were previously suggested but time constraints and feasibility have prevented current implementation:

* Changing the way the data is displayed on the map, from the current slope oriented dots, to larger groups that split once zoomed on, to changing the data that determines the color of the dot.
* Select multiple points on the map to be displayed overlain on the same graph for comparisons.
* A filter based on which measurement was used.
* Letting the user select what data is displayed over time (would alleviate the issue with points not containing relevant data)

Division of Labor:

Nate:

Set up the basic R Shiny app to use leaflet, explored options related to modifying a leaflet map in place (leaflet proxies) and programmatically filter data based on ecoregion. Making a separate data frame for each ecoregion would have been a pain, so I had to learn how to take data from the drop down in the UI and use it in the server behind the scenes, and make sure that the data being used was also filtered using the date slider.

Jason:

My main tasks were setting up the date filter, which I did by taking the data input on the date slider, and used a reactive function to redraw the map after filtering by both new start and end dates. Secondly, and most challengingly, I added a the scatter-plot with trend-lines. This was complicated due to the way shiny handles clicking on the map. It does so by returning specific information about what is clicked and not related to anything connected to it. Shiny returns the latitude and longitude, but with slightly different precision that the dataset we are using. As such, I had to reduce the precision of both, provide a margin of error, and find the point that best fits the location. Once the point is found, I have to traverse the raw data and find every point at that location. Once all the points are gathered, it was fairly simple in R to display them as a scatter-plot. To get the entire region, I save the name of the first points ecoregion, then use that to find all points in the data that have that ecoregion. Unfortunately, the redrawing of the map because of the date filter, the legend would cease displaying and would have to be clicked twice to display. The only solution I could find in a timely manner was to uncheck the tickbox every update. While this does not solve the disappearing legend issue, it does at least make the checkbox match what is displayed.