# CSCI 572 Homework 3 Report
# Visual Search and Interaction with NSF and NASA Polar Datasets

## Setting up our Dashboard

We built our server using Twitter Bootstrap components on the front end and using our own Python server on the back end. Python has a built-in http.server library which we used to implement the back end. We use a web form on the front end to send requests to our back end server. Using these requests, we send a request to our SOLR server and then use the response to generate the data for the desired D3 visualization. Then we pass this data back to the front end via an AJAX call and then use the D3 rendering code there to display the visualizations.

## How easy to use was D3?

D3 is a combination of scripting language and the SVG framework. D3 is not very intuitive. So in order to understand how to use it, we read the D3 paper and saw the 3 hour long video tutorials on lynda.com.

Initially we thought that D3 would be very easy and similar to JavaScript. As it turned out D3 is very similar to JavaScript but there are certain new techniques in D3 which are not easy to follow. For example in D3 we can select HTML elements before they are created and then later create them, this is definitely very counter-intuitive. Also the way to give data is via a function named ‘data’ and ‘enter’.

One of the best part about D3 is that they have an fantastic collection of examples and most of them have supporting HTML code to go with that. Therefore, the biggest challenge is to identify which D3 visualization goes best with our data and which data fields are suited to be visualized. For example, something like a bar graph is best suited to show time series.

The next step is to understand the code and make the necessary changes to parameters (or the fields) which we want to display.

The biggest crux of the D3 visualization is transforming the data which we get from querying Solr into the required format. Different D3 visualizations require different input files such as tsv,csv, json etc.

## Overview of D3 Visualizations

We created 6 different programs to generate D3 visualizations. The visualizations were bar graph, pie chart, donut, bubble chart, force directed graph, and map. We have submitted

python files which take in a Solr JSON response and produce the appropriate JSON, csv, or tsv to produce the corresponding D3 visualization.

### How easy to use was Banana?

Banana has powerful dashboard capabilities and works with both time series and non-time series data which is stored in Apache Solr. It also has panels that leverage D3.js. Moreover, it provides many additional capabilities like heatmaps, range facets, panel specific filters, global parameters, and visualization of "group-by" style queries. All of this helps in creating a rich and flexible UI which in turn enables users to rapidly develop end-to-end applications that leverage the power of Apache Solr.

Banana is an AngularJS app so in order to understand how to use it well we had to learn AngularJS and hence we saw tutorial videos on Lynda.com. Banana can be run in any web server and hence we decided to host it on Solr's web server. Installing Banana was quite easy. We had to just follow the simple steps from its GitHub page. After installing it we found it a bit difficult to configure it because the settings were not very intuitive. But after trying around the various configurations we were finally able to understand the powerful capabilities of Banana. Watching the tutorial videos of AngularJS helped us a lot.

After understanding the various configurations available in Banana it was easy to configure it to our needs. We used Banana to create a powerful analytics dashboard with various visualizations such as Bar Chart, Pie Chart, Tag Cloud, etc and an interactive search interface to the Polar data from the Solr index. Thus helping us answer challenging scientific questions in a very effective way.

### How easy was it to use FacetView?

FacetView is a pure javascript frontend for search indices like Solr and ElasticSearch. It's been developed as a jQuery plugin and uses the latest twitter bootstrap. It helps the user to embed a faceted browse front end into any web page.

However, in the latest version of FacetView support for Solr has been removed. So in order to add it back we have to a write an alternative to the elasticsearchquery() with a solrsearchquery(), and making sure the returned resultset can be parsed out properly. This was the most challenging part in getting FacetView to work. We had to understand the code for ElasticSearch and try to replicate the functionality for Solr. In doing so we had to modify a lot of existing code because a lot of existing code was made specifically for ElasticSearch and not for Solr.

Finally, after we were able to get FacetView to work we checked the various configurations available. Then we checked how to use it for helping us in answering the scientific questions.

We were easily able to add facets which helped in drilling drown the results to get more specific results from the resultset in a very efficient way.

We felt that if FacetView fully supported Solr we would have been able to integrate it easily.

**What was the hardest part, loading data, or visualization?**

For Banana we felt the visualization was a bit challenging because we had to understand how to visualize the data in the most intuitive way. Also the data loading for Banana was the easiest which required almost no pre-processing or code modification on our end.

For FacetView we felt the hardest part was getting it to work with Solr because it had ended support for Solr. Also the data loading was a bit challenging because we had to modify the format in which the data was needed to be displayed at the front end.

For D3, the visualization would have been the most difficult part, if we didn't have the examples on the D3 website to consult for help. So for us, loading the D3 was actually the more difficult part. For example, the bubble chart required a nesting structure and a mechanism for assigning sizes to the bubbles. We have to perform computations on the Solr JSON response, and for many D3 visualizations, these computations can be pretty complicated to implement.

**Describe in detail how your map answers the challenge questions presented in Task #3. Specify how to use the map to visualize the results.**

The map which we have used is present in the Geo Projections gallery. The visualization is called d3.geo.equirectangular. We had to get the world map json file from the mbostock server. This file contents the polygons and boundaries for all the countries of the world.

The next part was to display the latitude and longitudes onto the map. These latlongs were extracted from CLAVIN and indexed in Solr. They are dynamically queried from Solr and displayed on the Map. The map displays the queries such as how many regions are present in the data which we have. The regions are shown as red filled circles. Each result has a list of lat-longs, which we extract using a custom script and this put in csv file. The D3 loads from this and displays on the map.

Note: As seen in the video, the map visualization plots a lot places which are present in the Europe and US or in China and even some in India. This is due to the nature of how CLAVIN works. For a given document CLAVIN will provide a list of lat-longs and these will get mapped to the world map. Now CLAVIN also gives lat-longs for any places it finds mentioned in the document. Since a lot of these documents are commissioned by US government or academic institutions, those places get tagged. Nevertheless we do get plots for arctic regions such as Russia, Alaska, Antarctica, Canada, Sweden, Greenland, etc.

**Queries in the video:**
We were given 4 challenge questions in the assignment write up. We have based our queries on those and have thus covered all the 4 challenge questions. We believe that our queries are more detailed and thus provide better result.

Below are the nine queries to demonstrate answers.

- ❏ Are there any changes in migratory patterns of animals or birds due to rise in sea level?
- ❏ What natural resources are present in the Arctic Region?
- ❏ What military operations are currently being conducted in Antarctica?
- ❏ How do scientists monitor the Arctic wildlife?
- ❏ What oil exploration is happening in polar regions?
- ❏ Have sea ice changes led to new international shipping routes?
- ❏ What minerals can be found in Alaska?
- ❏ What programs are present for endangered species?
- ❏ How do polar species adapt to changing environmental conditions?
- ❏ How many regions of interest are represented by the data you collected?

Identify geographic “regions” as e.g., Circumpolar Arctic region, Antarctica and the Southern Ocean.