

## TEAM SRV

Shemon Rawat

Rizul Sharma

Vaishnavi Raghavaraju

# PROJECT 2: The Herbarium!

## Documentation

### AIM

The purpose is to make a web application that will allow users to explore the Herbarium collection through interactive visualization. The motivation behind this project was to enable users to understand the herbarium data.

### DATA

This data was extracted from an online database, and without any modification, it contains 8,000 entries, most of which have a GPS coordinate for where the sample was isolated. The data is in CSV format and contains the following attributes. Some data attributes are sparsely populated with data value entries and did not record an entry for that data attribute. Also, there are sparsely populated data fields.

- **id**: a unique identifier for this entry in the database
- **Recorded by**: who collected the sample
- **Event date**: in the form M/D/Y
- **Year**
- **Month**
- **Day**
- **startDayOfYear**: how far into the year was this collected?
- **decimalLatitude**: gives latitude for location coordinates
- **decimalLongitude**: gives longitude for location coordinates
- **maximum elevation (m) / minimum elevation (m)**: which gives you a range of the elevation of the specimen
- **country**: Text fields data about in which country it was collected
- **state**: Text fields data about in which state it was collected
- **municipality**: Text fields data about in which municipality it was collected
- **locality**: Text fields data about in which locality it was collected
- **taxonId**: a unique identifier to find other specimens of the same taxonomic classification (such as the same species).

- **higherClassification**: an ordered list of the classification of this species
- **Habitat**- Unstructured text data giving information about where the sample was collected
- **Substrate**- Unstructured text data giving information about where was the sample growing?
- And one column for each taxonomic classification: **Kingdom, Phylum, Class, Order, Family, Genus, SubGenus, Species.**

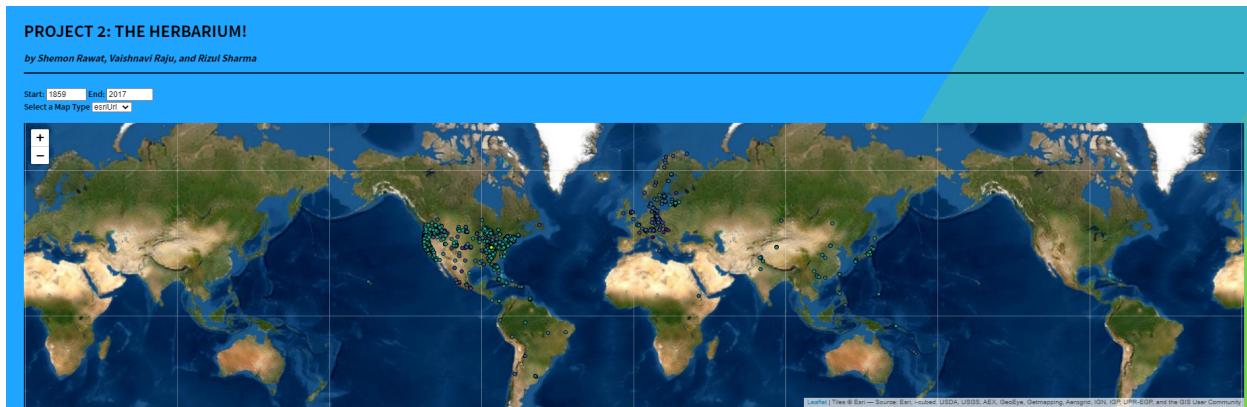
## VISUALIZATION COMPONENTS

### LEAFLET MAP

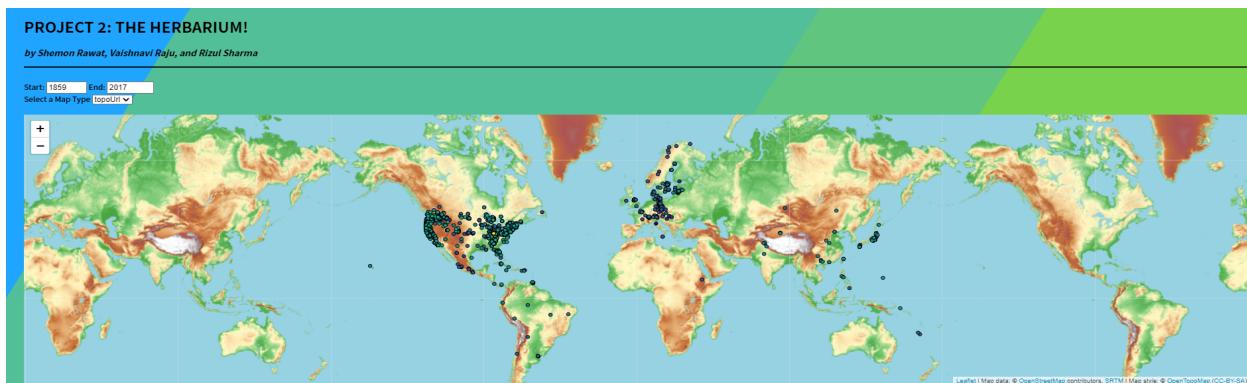
This component helps users to interact with the map with different terrain features to visualize where the sample was collected by year. To select the timeline, we have provided two components to select the start and the end year, and it will show where on the map the sample was collected between those years.

There is one more added feature to this, we gave the user the option to select the map type as well. Users can select the map type for four different terrains as shown in the pictures below.

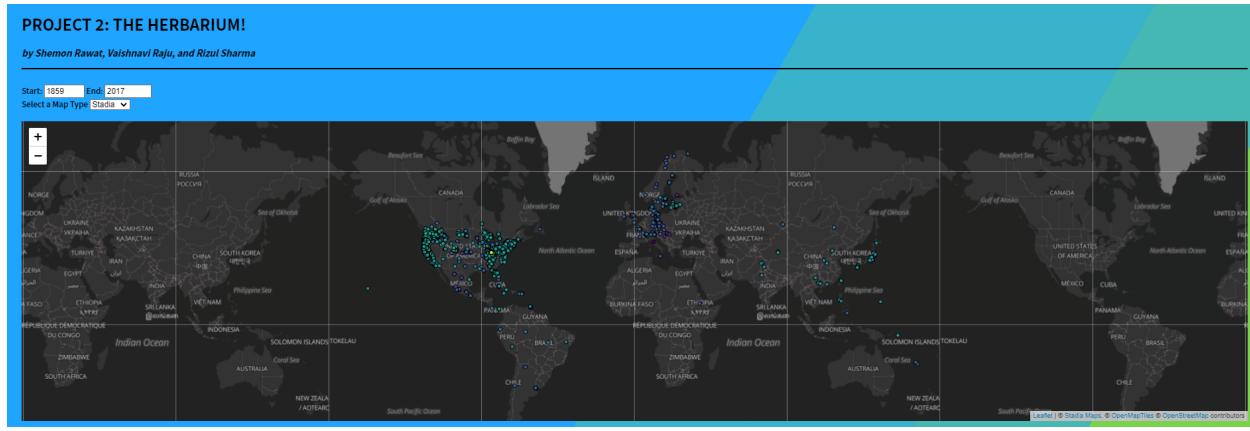
The first terrain type is esriUrl.



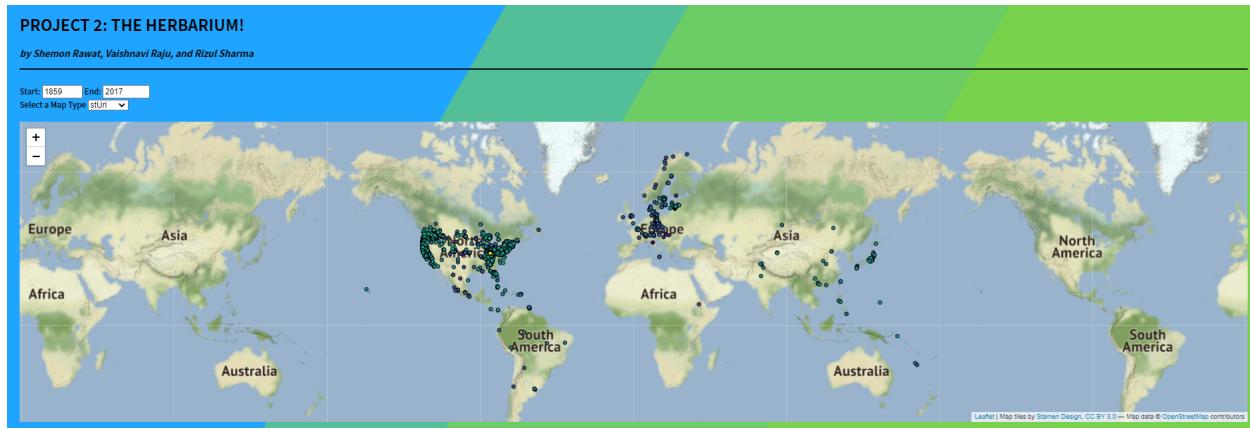
The second terrain type is topoUrl.



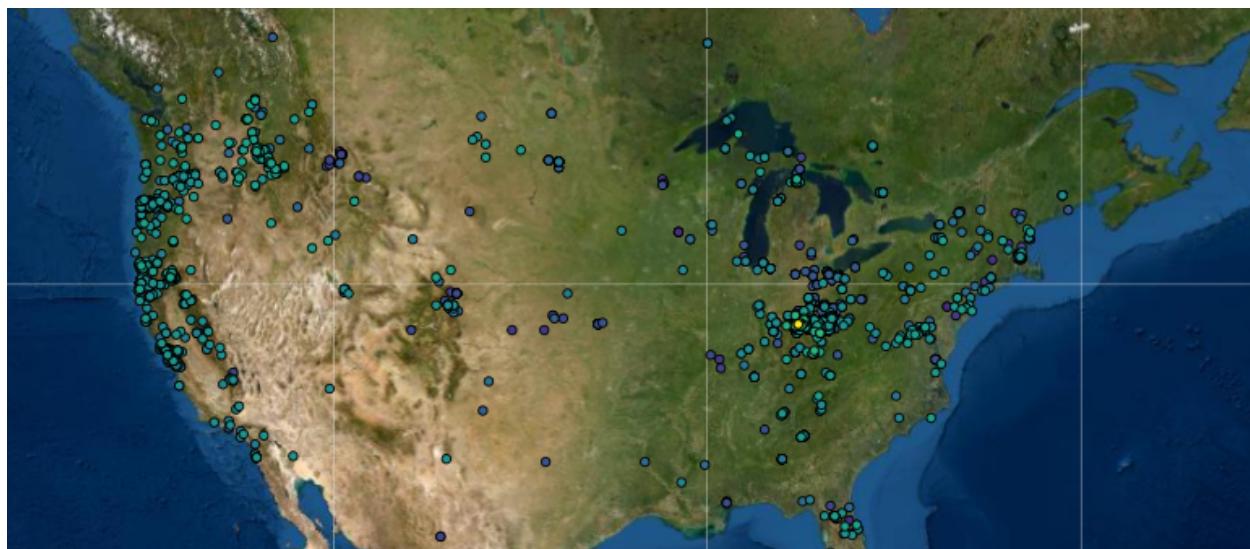
The third terrain type is Stadia.



The fourth terrain type is stUrl.



Additionally, if you click on the dots on the map, it will direct you towards the details about that particular sample that you just clicked on.



Details Map Comments Linked Resources

[Share 0](#) [Tweet](#)

 University of Cincinnati, Margaret H. Fulford Herbarium - Fungi (CINC)

**Catalog #:** CINC-F-0007610  
**Occurrence ID (GUID):** 34621d83-cf89-4b2f-93c8-8de766c3f4b5  
**Taxon:** *Calvatia craniiformis* (Schwein.) Fr. ex De Toni  
**Identification Qualifier:** Calvatia craniiforms  
**Family:** Agaricaceae  
**Collector:** Tokarski, R.  
**Number:** s.n.  
**Date:** 1964-10-17  
**Verbatim Date:** October 17, 1964  
**Locality:** United States, Ohio, Scioto, Shawnee State Forest, Shawnee State Forest 38.74287 -83.204008 +-2289m.  
**Substrate:** on soil

**Specimen Images**

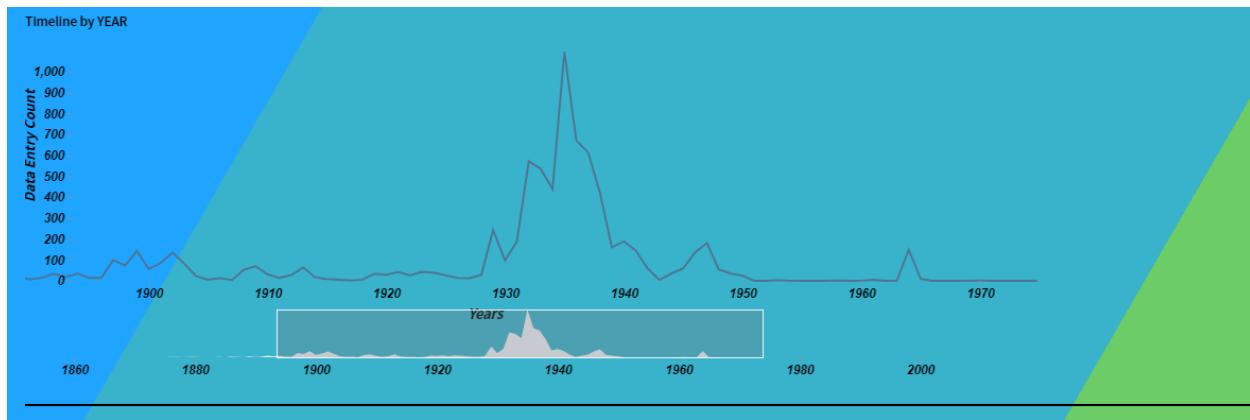


[Open Medium Image](#) [Open Large Image](#)

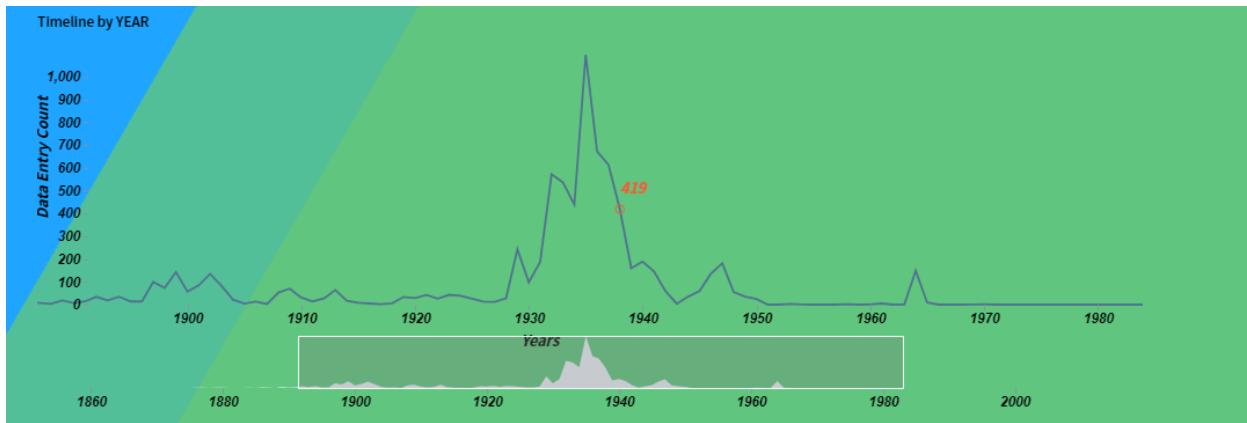
**Usage Rights:** [CC0 1.0 \(Public-domain\)](#)  
**Record ID:** 34621d83-cf89-4b2f-93c8-8de766c3f4b5

## LINE CHART: TIMELINE BY YEAR

This component enables users to see how many samples were collected in the respective years and what was the rate of collection between the years. This is possible because of the brushing linked with the line chart as shown in the picture below.

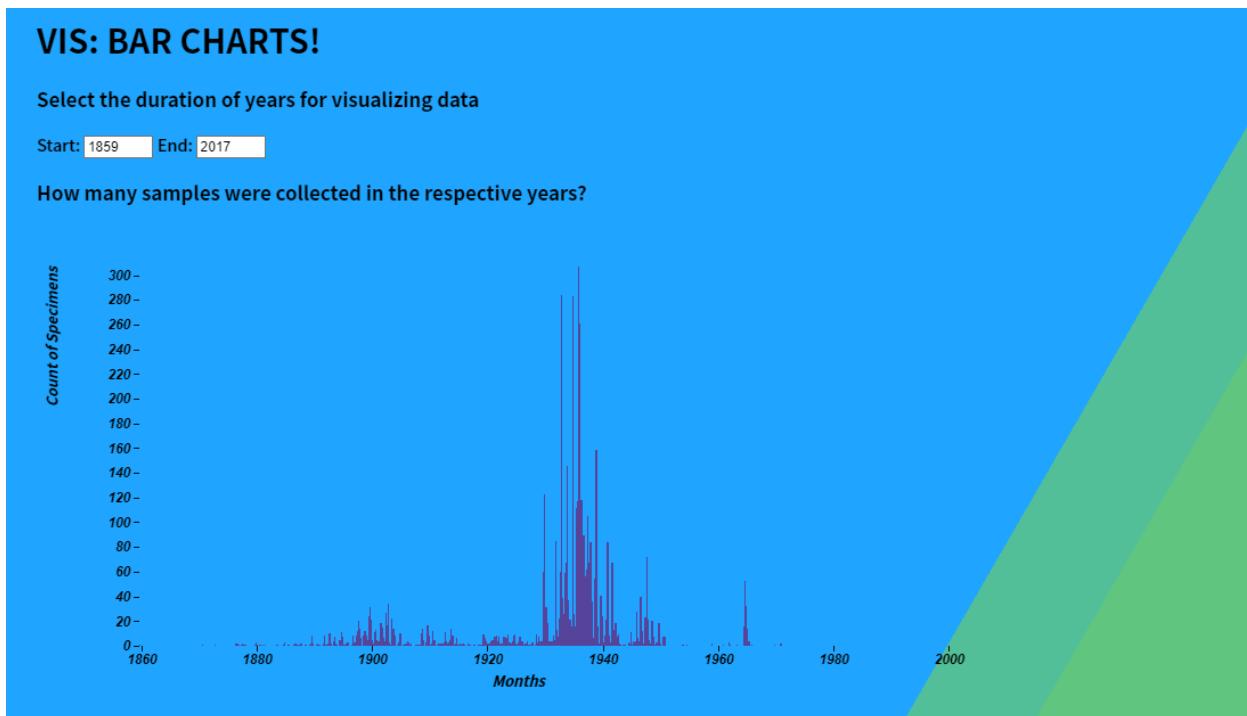


Also, we have provided the tooltip for the user to see the exact number of counts when he/she/they hovers over the line in the line chart.



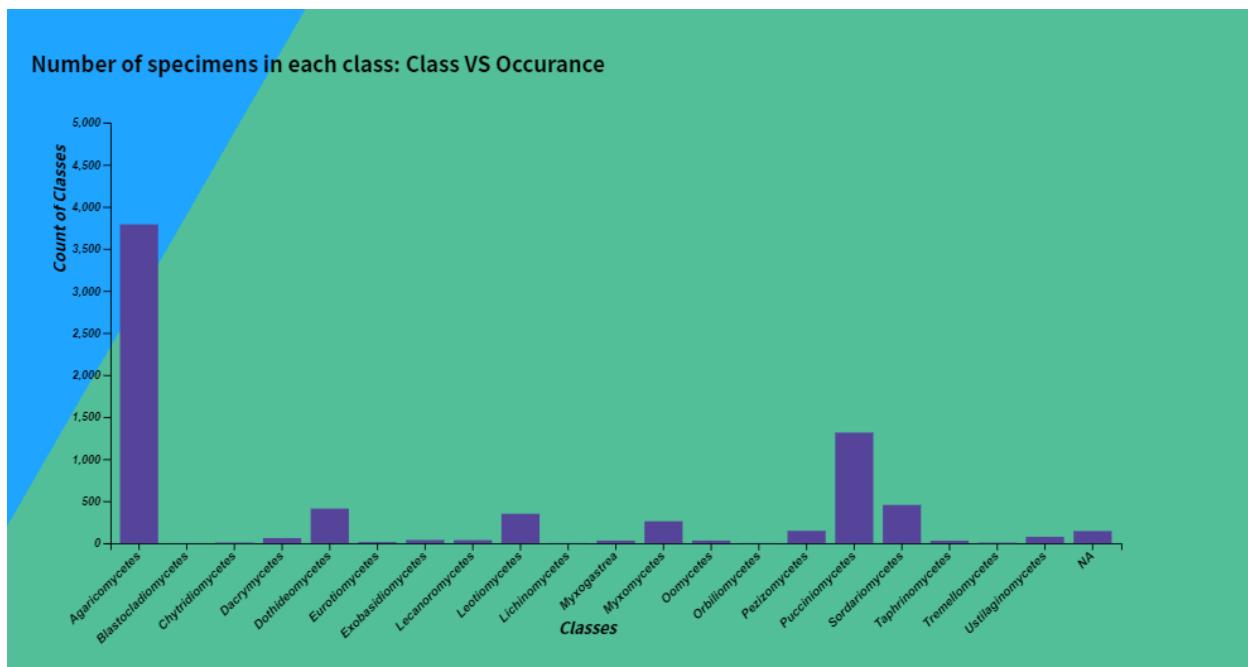
## BAR CHART 1: How many samples were collected in the respective years?

This bar chart enables users to see how many samples were collected in the months of the respective years. The process is to select the timeline by choosing the start and end year and the x-axis will shift accordingly. Users can see the different bars between the year labels to determine how the collection progressed with the passing year.



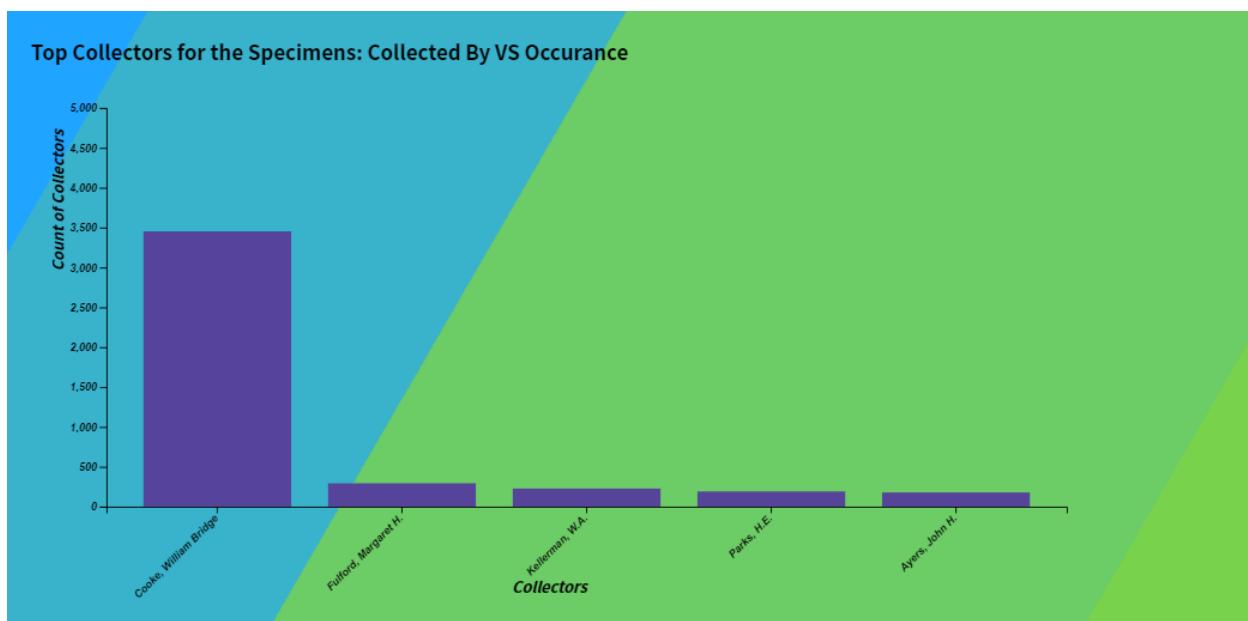
## BAR CHART 2: Number of specimens in each class

This bar chart enables the user to see the number of specimens in each class for the chosen timeline. The process is the same, the user selects the start and the end year and the graph will show the number of samples collected classified into different classes.



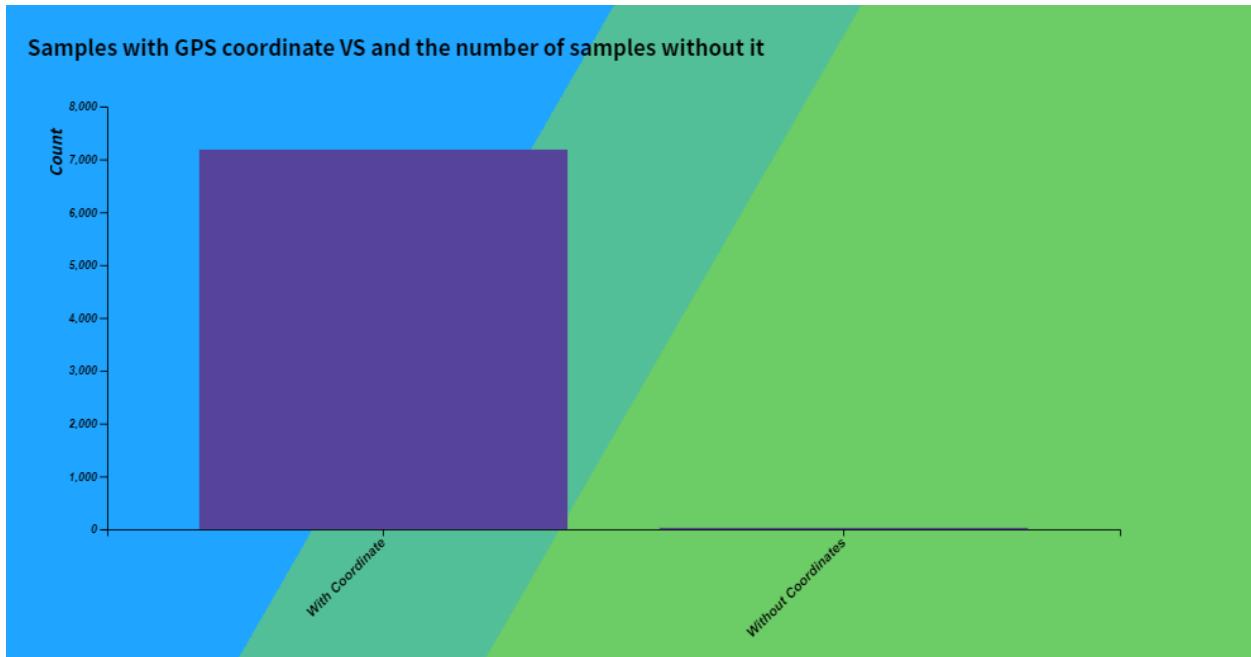
### BAR CHART 3: Top Collectors for the Specimens

This bar chart enables users to see the top collectors of the samples for the chosen timeline. The process is the same, the user selects the start and the end year and the graph will show the top collectors for that timeline.



#### **BAR CHART 4: Samples with GPS coordinate VS and the number of samples without it**

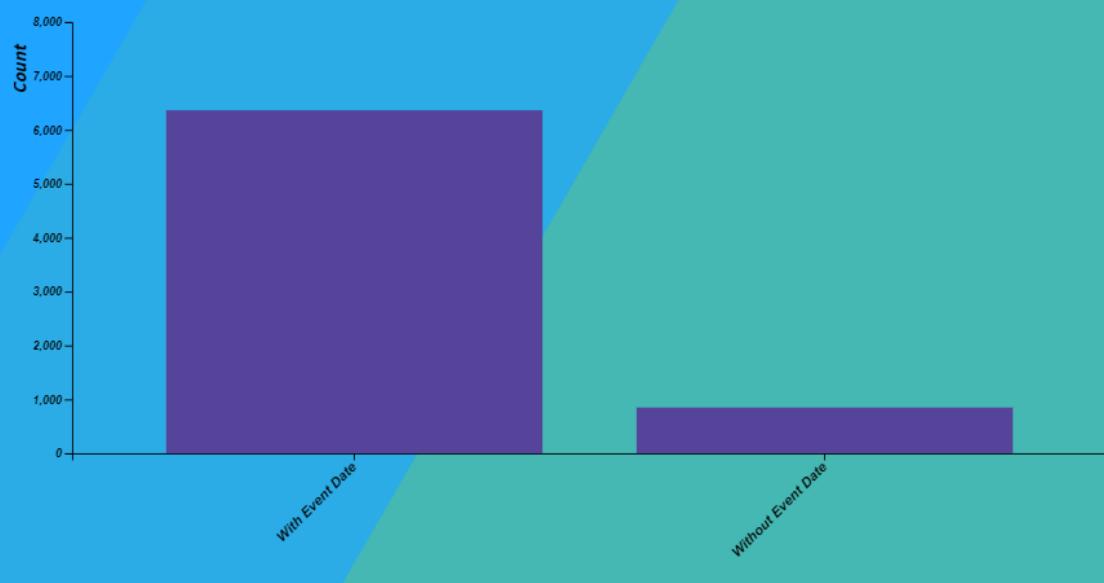
This bar chart enables users to compare the number of samples with GPS coordinates with the samples without GPS coordinates. This is a static graph.



#### **BAR CHART 5: The number of specimens total, the number with an event date, and the number without**

This bar chart enables users to compare the total number of samples with an event date with the total number of samples without an event date. This is also a static graph.

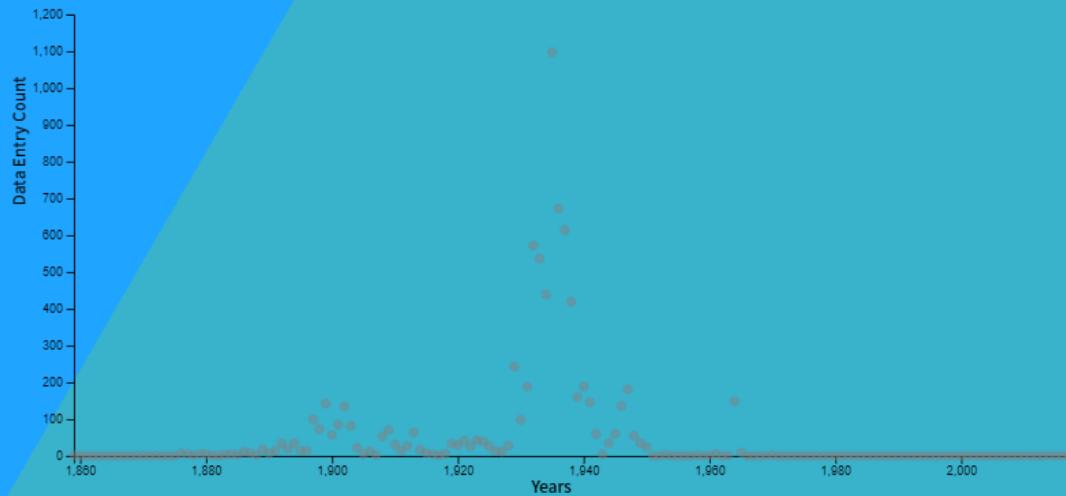
The number of specimens total, the number with an event date, and the number without



### SCATTERPLOT: Data entry done during respective years

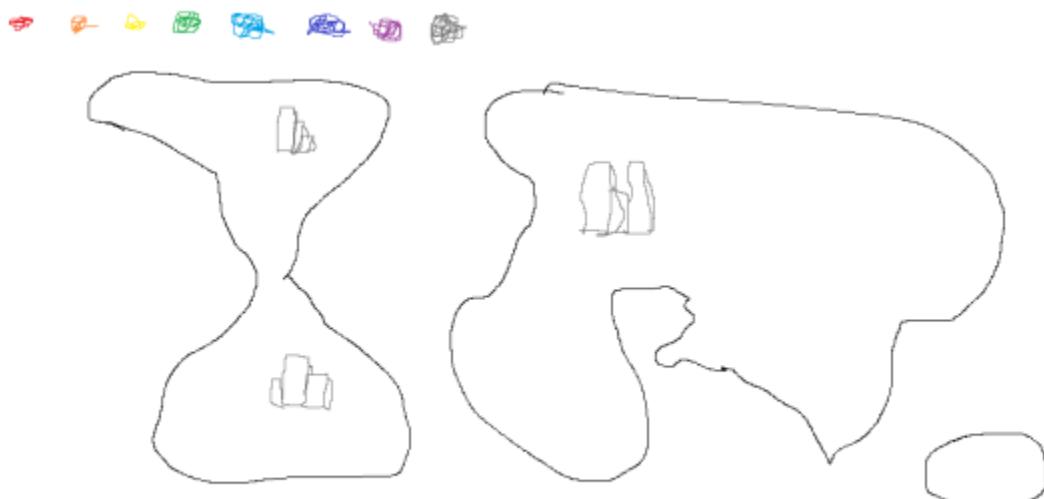
This component helps the user to determine the data entry done during the respective years.

Data entry done during respective years



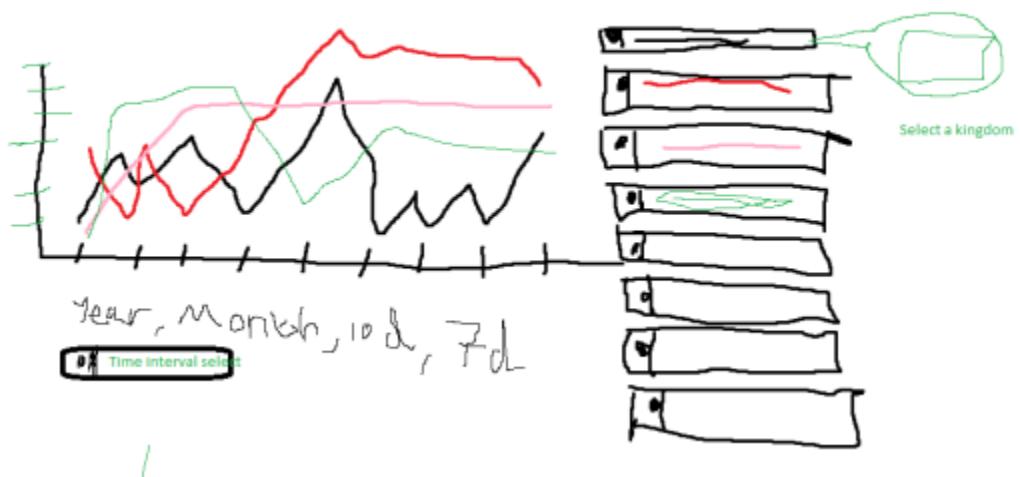
## DESIGN SKETCHES

### Sketch 1:



Sketch 1 shows a map with 3D histograms on the map. That was our initial idea. But then we switched to a map with different terrain features visualizing where the sample was collected by year.

### Sketch 2:



Sketch 2 shows a line chart that will have options so display data by the kingdom. The more options you select more lines to appear. You can also select different time intervals. We switched from this to a dynamic line chart with a brushing component linked with the timeline.

### Sketch 3:



Sketch 3 Shows a simple line chart that displays all the data collected over the year. We added bar charts to this.

## DISCOVERY

This visualization app provides features to easily visualize the data and determine the finding this data can provide. In other words, if the visualization is good, it adds priceless value to the data. With that approach, we also found interesting facts about this data.

- Most of the data entry happened in 1935.
- From the map and considering the full timeline (1859 - 2017), most of the samples collected were from the North American and the European regions.
- Most samples were collected in October of 1935.
- Among the total number of samples, most belong to the class *Agaricomycetes*.
- William Bridge Cooke collected the most number of samples amounting to nearly 3,500 in the count.
- There is a fairly less number of samples that lack the data for GPS coordinates and an event date.

## PROCESS

### Tools Used:

- Javascript
- HTML
- CSS
- D3 (.js library)

- R (for data pre-processing)

### Structure:

We have six folders, one index.html and one README.txt inside the main project folder. The structure of the application is as follows:

- brush - Folder containing the data and the HTML page for the scatterplot component.
- css - Folder containing the two CSS files, one is leaflet.css for styling the map, and the other is style.css for the overall app design.
- data - Folder containing the datasets.
- Fungi\_UC\_Herbarium - Folder containing the bar chart components and CSS styles for them.
- images - Folder containing the images used in the project.
- js - Folder containing the js files for visualization components.
- Index.html - First index HTML page for the app.
- README.txt - 'Read me' file containing the team information and notes.

### Source Code:

You can access the code [here](#).

### DEMO

You can access the demo video [here](#).

### TEAM CONTRIBUTION

1. Shemon - Worked on the leaflet map and its features, added brushing and other components to the application including scatterplot and dynamic timeline. Integrated the separate components into a single project before the final design.
2. Rizul - Worked on the data pre-processing and made sure it was clean and the attributes are ready to be directly used in the application. Fixed the timeline line chart. Also, worked on the styling, animation, debugging and final design.
3. Vaishnavi - Worked on the section for the Bar Charts & Histograms, and added dynamic features to them. Also, worked toward fixing the axes and labels.

All three contributed equally towards the completion of this project and were equally responsible for the decision making in the structure, and design choices for the application.

Documentation is done by Rizul Sharma.