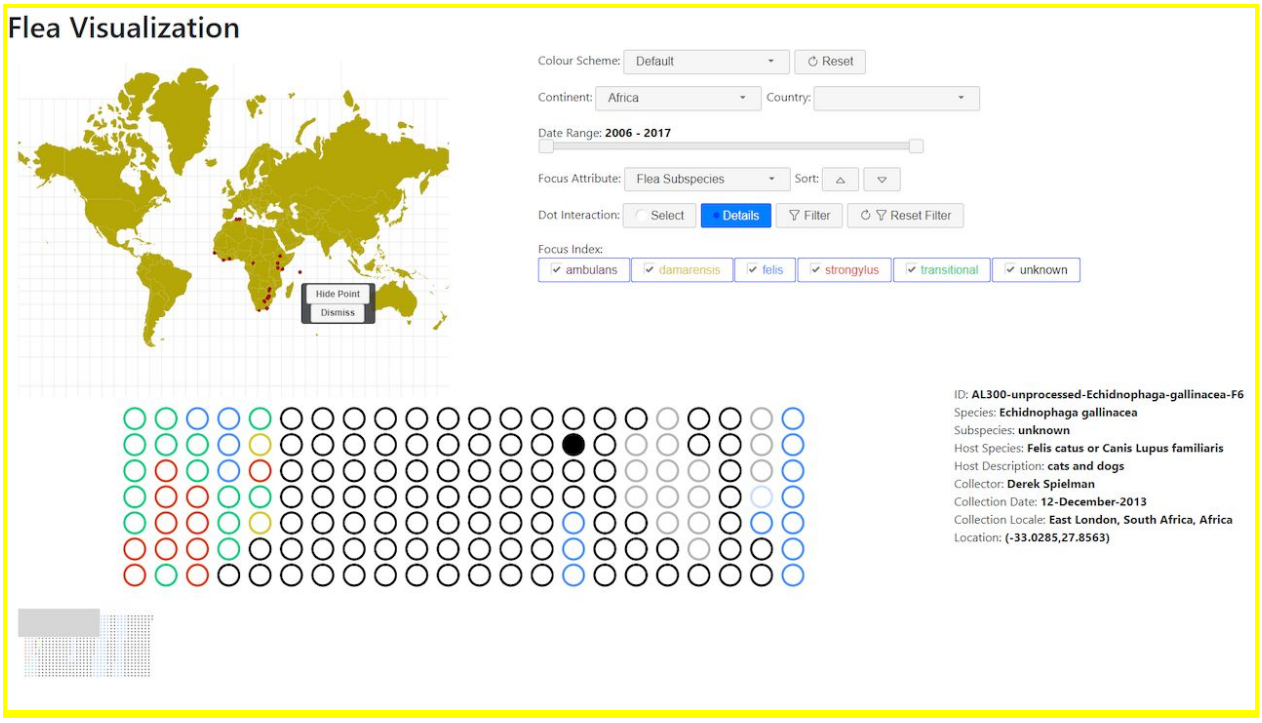
CPSC 436v Info. Visualization Process Log Writeup:

Flea Visualization v1.0-final - Gary and Friends
Team Members:

- Jiawei Tan, jiaweit
- Nick Birtch, nbirtch

Overview:

As an exploratory data analysis, we built a one page data visualization that allows scientists and interested parties to visually explore a dataset of flea specimens collected globally. Our visualization consists of a regional map view to which we map the flea specimens, and a dot view, which supports the data analysis and selection of the aforementioned specimens. Our views clearly show details about the fleas, their hosts and their locale, while also allowing users to explore different aspects of this data by filtering and re-ordering on different variables in order to discover and compare factors between collected specimens. This project can be considered a proof of concept visualization for larger, more comprehensive datasets of specimens akin to the Catalogue of Life datasets. (https://www.catalogueoflife.org/).



Data:

We have visualized a dataset of 5,128 flea specimens. Each flea (or sampling of fleas) has 32 associated variables. The variables our visualization focuses on describe the specimen characteristics (Sample #, Specimen #, Flea Species, Subspecies, Gender), the collection locality (Continent, Country, Region, Latitude, Longitude), and the host collection information (Host Species, Host Description, Collection Date). Further variables such as Collector Name, Clade, Cluster and Notes may be included in tooltips/informative sections but are not directly visualized.

Variable:	**Type:**	**Cardinality/Range:**
Sample #	Categorical	944
Specimen #	Categorical	686
Flea Species	Categorical	14
Subspecies	Categorical	6
Gender	Categorical	3
Continent	Categorical	6
Country	Categorical	57
Region	Categorical	218
Latitude	Quantitative	-46.0988 , 58.3776
Longitude	Quantitative	-179.9813, 177.4356
Host Species	Categorical	14
Host Description	Categorical	40
Collection Date	Quantitative	Jan 1st 2006, March 2nd 2017
Collector Name	Categorical	81
Clade	Categorical	8
Cluster	Categorical	4
Notes	Categorical	37

Data Source:

Dataset: http://dx.doi.org/10.17632/2f3hchym9v.1

Attribution: Lawrence, Andrea (2019), “Global flea collection specimen details:

Supplementary Table 1”, Mendeley Data, v1.

License: https://creativecommons.org/licenses/by/4.0/

Data Preprocessing:

Originally, some fleas in our dataset did not have an ID, so in order to distinguish individual fleas we combined the "sample no#", "specimen no#" and "flea species" as a default flea ID if not provided.

As well, the dataset only shows the number of female and male fleas per flea ID (e.g. flea CTENO370-18 contains 1 female and 1 male). In order to represent each individual flea, we created new fleas based on these gender counts, and we added a new property "gender" to the flea data. We set the newly created flea IDs to be a combination of original ID (or default flea ID), gender and an indexed count (e.g. CTENO370-18-M1 as male 1, CTENO370-18-F1 as female 1).

We also had to make some small adjustments for names of regions in our json files and dates among other fields in the dataset that did not conform to standards.

Goals And Tasks:

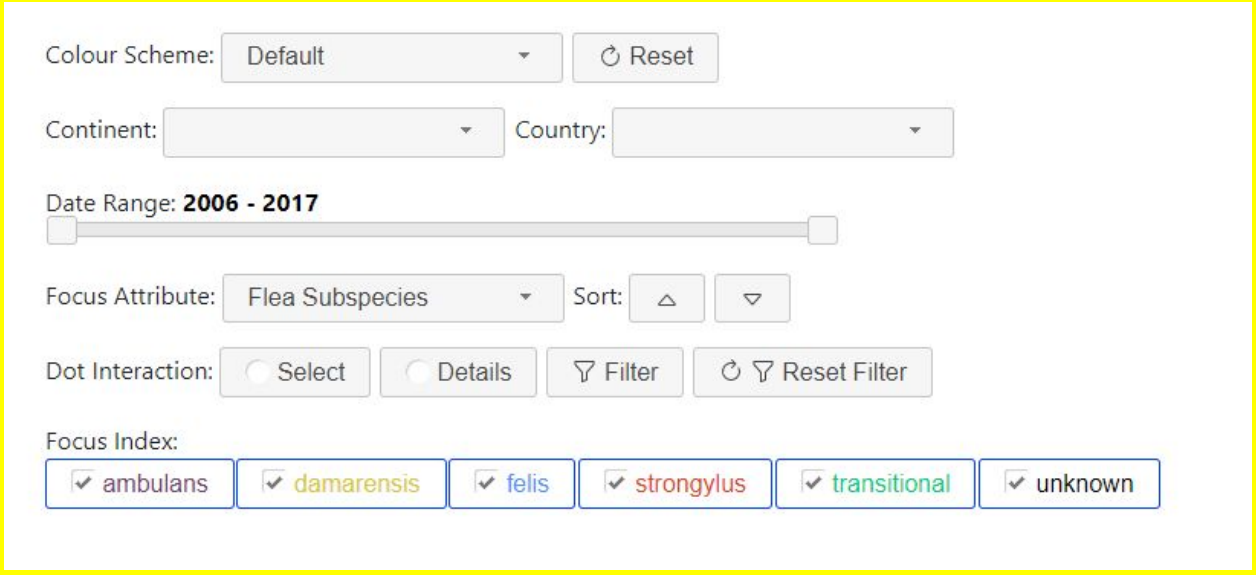
A user is able to [explore features] a dataset that contains flea specimen records, [browse distribution] how different flea species are distributed in different habitats and hosts, and [compare trends] the differences and trends between flea species including gender, host and locale. A user can also [discover distribution/extremes] the dominant flea species on the current continent or per type of host. Ultimately, we hope the user [enjoys] themselves.

Visualization:

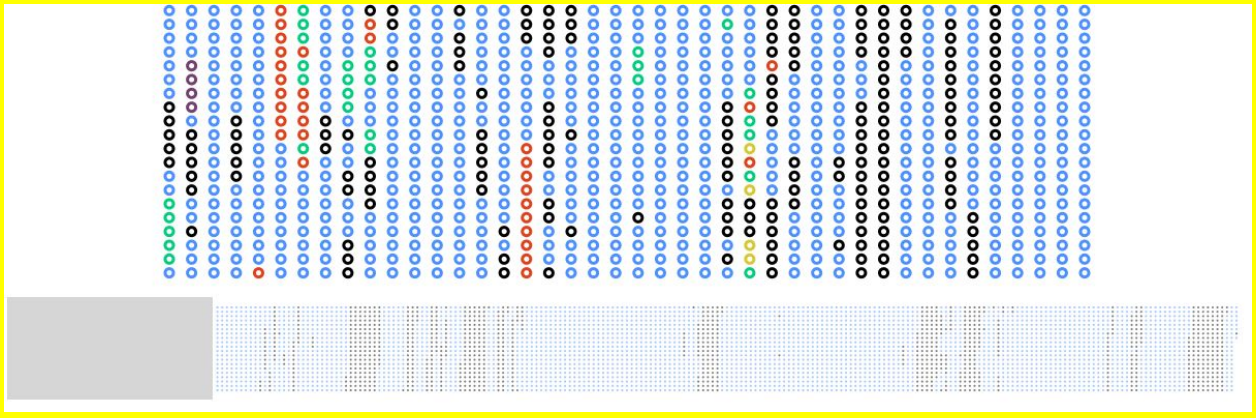
On the top left hand side of the page, we have the regional map visualization which shows the flea distribution across the world at different regional levels (continent, country). Individual flea

data points are visually encoded as small circle point marks while geographical regions are encoded as area marks.

On the right hand side of the page we have various UI widgets that support bidirectional interactions and filters both the map and dot visualizations as described below.



At the bottom of the page, we have the dot visualization that allows the user to sort, filter and individually select data points by an attribute of focus. Individual flea data points are visually encoded as hollow circle point marks in the dot visualization. Dot colour encodes the attribute of focus values, Dot luminance/highlight encodes the filter selection and the filled center encodes the current focus of the information box.



On the bottom right hand side of the page below the UI widgets, we display relevant record information of the currently selected map point or dot in an information box.

Interactions:

- The default state: When no filtering has occurred, all map points will be shown and all dots selected.
- When a continent or country on the map is clicked on, the map will recenter and zoom on the selection area, and the map and dot visualization will reflect the region filtering.
- Hovering on continents, countries or points on the map will display a related tooltip.
- Clicking on a map point displays an interactive tooltip that allows the user to individually hide map points and deselect their related dots.
- Clicking on a dot using the detail interaction, the related data record will be shown in the information box.
- Clicking on a dot using the select interaction, the individual dot will be selected or deselected in the current selection.

- The user will be able to scroll dot visualization left and right with a subset view of data points.
- The user will be able to select and change the attribute of focus via dropdown which updates the dot visualization and its index.
- Clicking checkboxes next to index entries will allow the user to quickly filter the current selection of data points on particular values.

UI Widgets:

- Two dropdowns that update the map on selected Continent or Country, filtering the data.
- A colour scheme dropdown that allows users to change the map and dot visualization encoding between default and colour blind friendly.
- A reset button that returns the entire page to the default state.
- A range slider that allows users to filter the data by date.
- A sort button with ascending/descending options that resorts the dot visualization data points on the attribute of focus.
- A radio button widget that allows user to change the dot visualization between selection (left click to select/deselect) and details (display info box) interactions
- A filter button that filters both the map and dot visualization on the current selection.
- A reset filter button that returns only the map and dot visualization filter selection to the default state.

Rationale:

Some form of map visualization was an obvious choice for our flea data, as there were many locale based attributes describing the regions and subregions in which the flea specimens were gathered, including latitude and longitude which ties directly into our mapping processes. Sticking to continents and countries as our filterable regions made the most sense for general user knowledge and usability.

Our dot visualization was our innovative component where we really wanted to improve and expand on how users engage with our data beyond reading a data record by providing various functionalities such as the subset dot view, the various filter actions and an attribute of focus. The goal was to guide the user in essentially querying our data via the tools given with the dot visualization.

The focus of our design choices were to allow the user to drill down on the dataset as quickly and intuitively as possible while conveying useful and interesting information. Visually-oriented

users will be serviced by being able to select regions on the map and interacting with the dot visualization to filter on and discover their areas of interest while more data-oriented individuals will be able to more directly find the information they need using the dropdowns, sliders and index filters.

The data point filtering through region or dot selection, dropdown and the date range slider (among other widgets) as well as the map zoom and subset dot view all serve to aid browsing and exploring the distribution of fleas relevant to the user by restricting their view.

The two visualizations in conjunction, the information box and the colour coded indexing in the dot visualization specifically allow the user to easily discover and compare the differences and trends of various attributes between fleas.

Reflection:

Our visualization goal has been consistent since our initial proposal, in wanting to build a one page data visualization that allows scientists and interested parties to visually explore a dataset of flea specimens collected globally. We wanted our views to clearly show details about the fleas, their hosts and their locale, while also allowing users to explore different aspects of the data by filtering and re-ordering on different variables in order to discover and compare factors between collected specimens.

Our first proposal, while realistically quite within the realm of D3 possibility, was lacking the expected level of technical difficulty for the project. We had to reconsider our design choices and perhaps even our dataset, but we ultimately decided to move forward with the flea visualization. We revised our designs and attempted to flesh out an innovative component in our dot view and looked into implementing interesting functionality and bidirectional connections with the map view to support our project.

With the plethora of online resources and support from our in class instruction and TAs throughout the project, we were able to implement what we wanted or make informed alternative design choices, even if we couldn't figure it out at first. By M2, we were happy with our new direction and progress, and most importantly we were starting to come to a better understanding of the scope and difficulty of different aspects of our views and were able to better prioritize our development efforts accordingly.

While our initial technical goals were not challenging enough, having expanded upon our initial horizons we may have been too lofty in our expectations and how we hoped to handle d3

animations. Ideally, we would have liked to implement smoother, visually appealing map and dot visualization transitions on our interactions, but we ran into issues with unreliable animations, freezes and performance with the dot view specifically.

We continued forward with the animations as best as we could by simplifying them and testing out various animation methods and fixes to ensure reliability, but we had to scale back our expectations on the level of polish and user experience we were hoping for.

Reflecting on our final project now, although there were problems along the way, it was a fun and engaging exercise and we are proud of our finished product. If we were to make another project, we would focus more of our time on the initial design and project planning, putting as much effort and care into how each view and interaction should look and feel beforehand as well as researching and discovering new ideas and methodologies to incorporate. Creating a functional UI mockup as early as possible would also be beneficial. With a clearer grasp of our target visualization early, it would be easier to fulfill and reach for goals from there.

Team Assessment:

Nick primarily worked on the map visualization which included designing the map layers and the connections therein, adding zoom, pan and transition behaviour and connecting the GeoJSON data to the dataset. Jiawei focused on the dot visualization, implementing sort, selection and filter while handling the performance, animation and scrollability of the dot subset view. We both contributed to our design choices, setting up the page layout and styling, implementing widget interactions, connecting the visualizations and widgets bidirectionally and testing. Both of us communicated our progress and expectations throughout and delivered what was expected of one another.