# Process Book

**Liam Moynihan - u1262004, liam@cs.utah.edu**

**Daniel Malmuth - u0746311, daniel.malmuth@utah.edu**

## Overview and Motivation

The main motivation of the project was to set the framework for better understanding the methods used to discover exoplanets. If we are able to draw correlations between discovery methods and exoplanet attributes, we can get a better idea of the strengths and weaknesses of the different discovery methods. This will help to inform future decisions on which methods should be used when creating new discovery facilities, as well as to provide an intuition for all of those interested in the field.

New data is coming in every day on the discovery of these exoplanets, so it is becoming more and more important to be able to narrow down on which methods are best able to capture which kinds of exoplanets.

While our final visualization gave us the ability to study the methods used to discover exoplanets, the framework proved to be much more general than that. For example, it is just as easy to get an answer to the question “how does the distance of the different exoplanets discovered vary over time?” by using this visualization.

## Related Work

We were both inspired by the examples of parallel coordinate plots that were shown in Lecture 15. They were a new (to us) and intuitive way to display tabular data. They

require a number of interaction features to be useful, but the nature of a final project allows us to take the time to implement them fully.

## Questions

- Which scientific methods are most commonly used to discover exoplanets?
- How do the discovery methods vary in terms of capabilities? Are some better able to discover far away planets? or small planets?
- Are different facilities more successful than others? Do they specialize in specific discovery methods? Do they have different limitations in terms of planet size, distance, etc?
- How has all of the above changed over time?

As we implemented our visualization to cover these questions, we found that there was very little emphasis on individual exoplanet data. While we were much more interested in general trends of the set of exoplanets, we still wanted to implement individual exoplanet details for anyone who may be interested in a particular planet's attributes, so we added details on hover/click in the scatterplot.

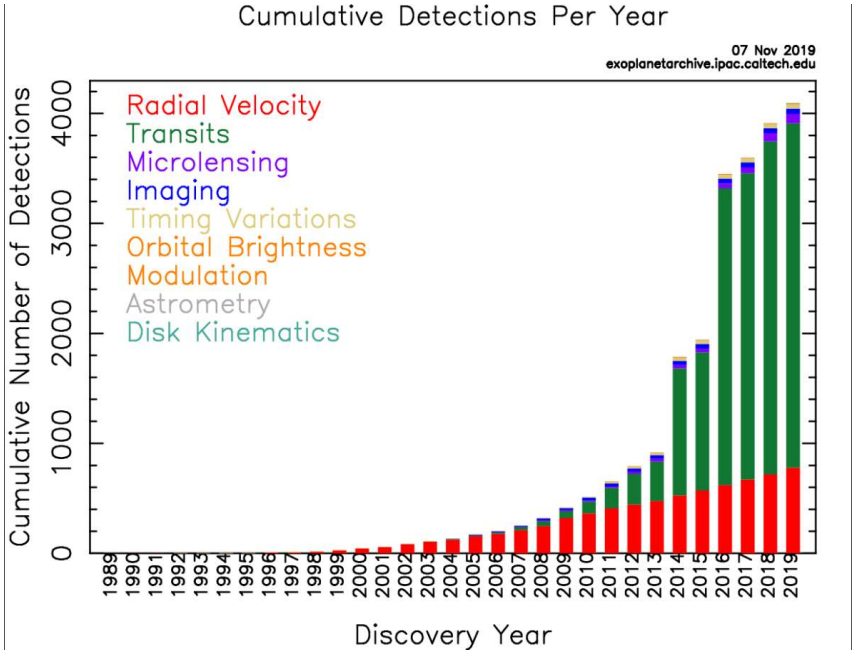
## Data

- The data will come from the NASA Exoplanets archive, Confirmed Planets table. https://exoplanetarchive.ipac.caltech.edu/. We only used a select number of the columns, since we wanted our dropdown to be manageable to navigate.
- We converted the .csv file containing the data to a .json object and extracted the columns that we cared to have
- We did some supplementary research on the different discovery methods in order to give a brief explanation of them all
- Many rows are missing specific data, such as mass, radius, and distance. We will create the option to filter these out depending on how useful they are for the visualization. Many values also have errors associated with them. We decided to not implement this in order to avoid issues with readability.
- We considered to develop the star-map plots described below, where we would need to do calculations to convert Right Ascension and Declination data into

pixel positions, however we were already somewhat limited in vertical space of the visualization and decided to not implement this.

## Exploratory Data Analysis

The NASA exoplanet archive includes a number of plots with the data, which is useful to better understand the data before writing our own code. For example:



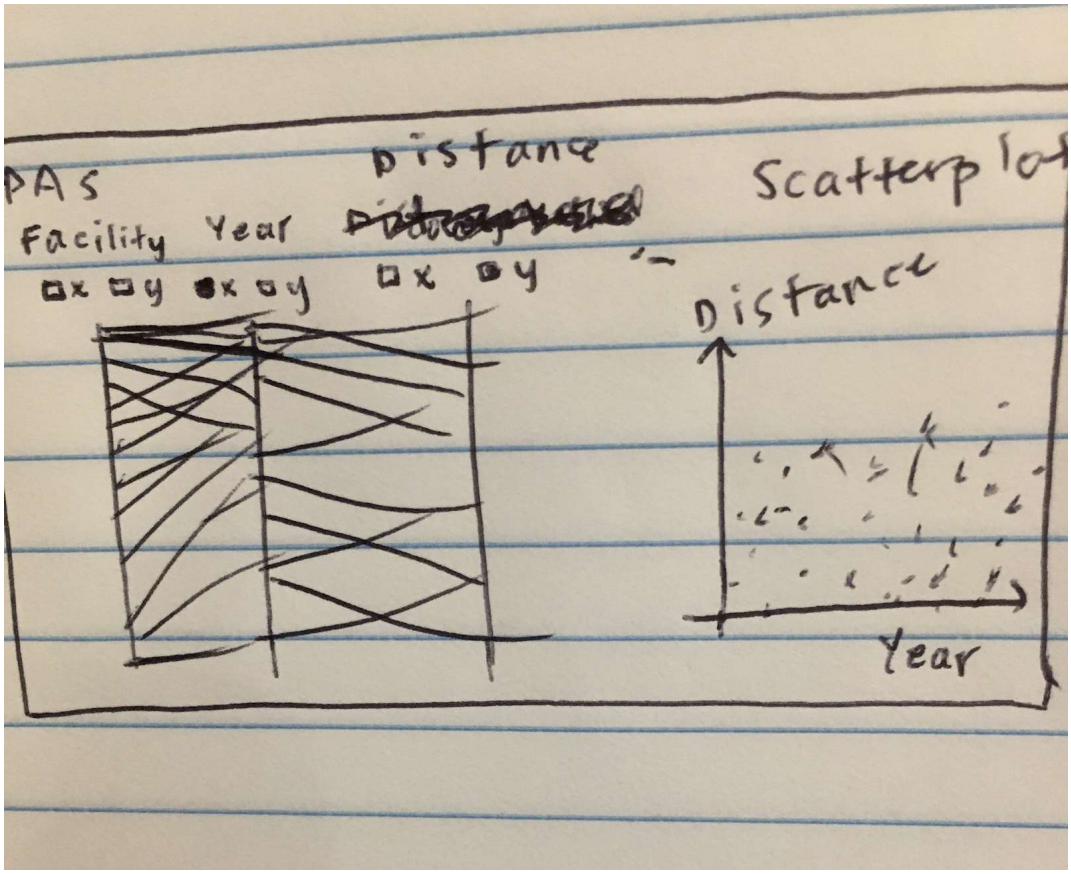

The Exoplanet archive also allows us to pick axes to generate custom scatterplots. Control over the visualization beyond that is very limited, but it is still useful for exploratory purposes. Together, these tools allowed us to identify which questions would be interesting to ask. They also showed some of the challenges that we would have to overcome with our designs. Namely, much of the data can be densely packed, and we may need interactive features (filtering, brushing etc) to make the visualizations useful.

## Design Evolution

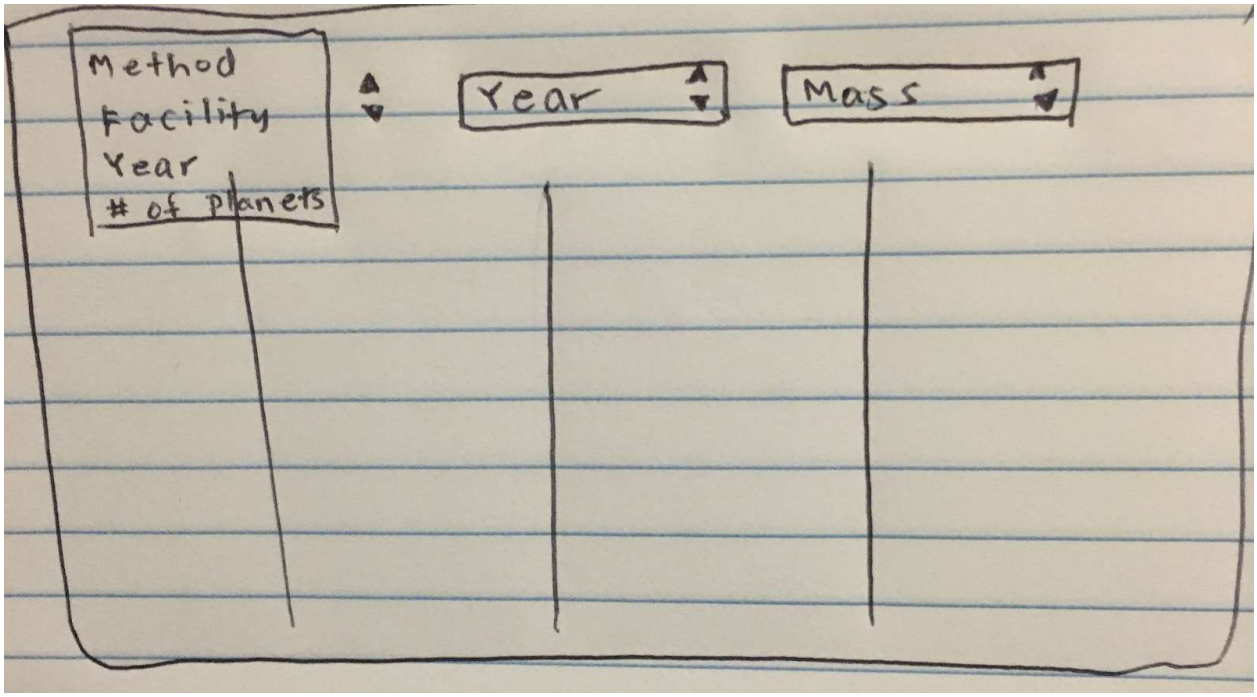
These were the initial points we covered concerning which visualizations we were implementing and their respective purposes at the point of the project proposal:

- Scatterplot to visualize two attributes of particular interest
- Parallel axes to get a broader view of correlation between different attributes with the ability to swap the order of the axes
- Brushes for both to filter the large amount of data
- Storytelling to give an introduction to different methods and planet attributes (including units used)
- So far, very little deviation from the proposal. We have initially chosen to make fewer columns available in the parallel axes plot and may restrict the different choices to not include everything available in the data set

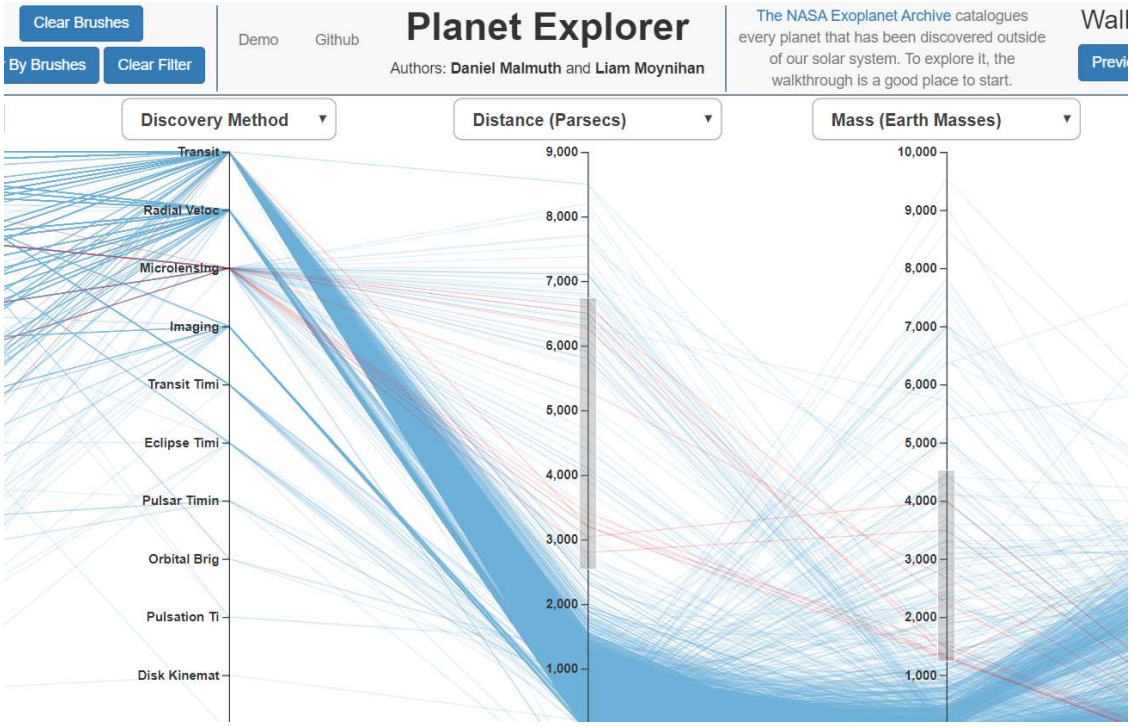
We had planned on a visualization looking roughly like this:

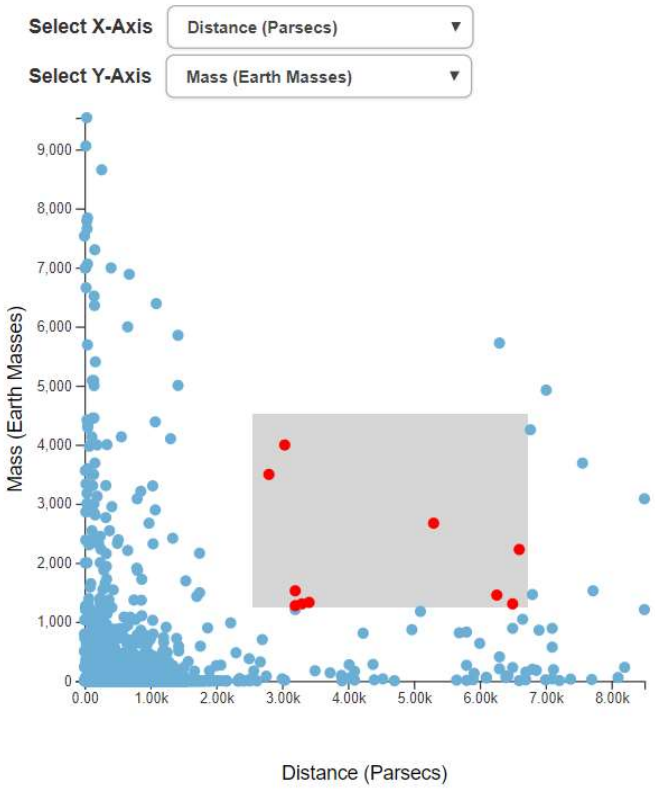


In this case, the x- and y-axes of the scatterplot were selected via the checkboxes in the parallel axes plot. This, however, had many obstacles. We wanted to later implement the ability to change the axes to something that wasn't already present in the parallel axes headers. We decided that we should change the headers to be dropdowns, like this:

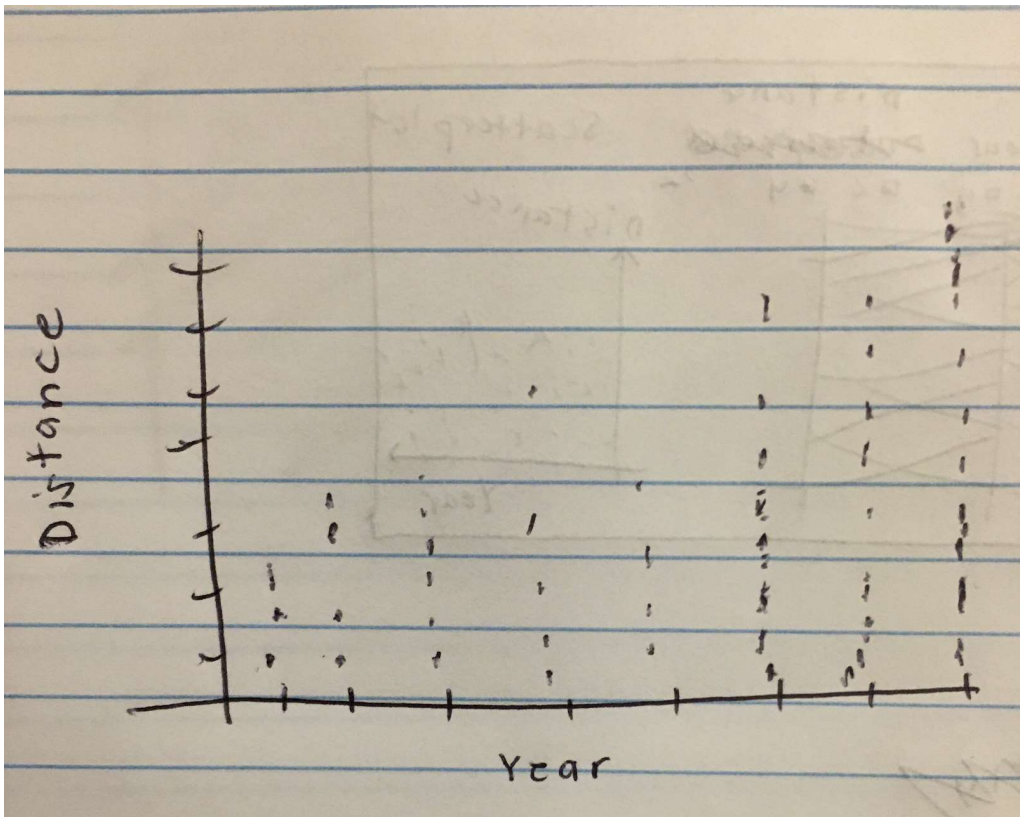



At this point, we also removed the checkboxes because we wanted to have the ability to look at data in the scatterplot that wasn't necessarily on the parallel axes. That being said, we also wanted to implement a connection between the parallel axes and the scatterplot, so we connected the ranges covered in the different brushes. For example, the two brushes selected below correlate to the following in the scatterplot:

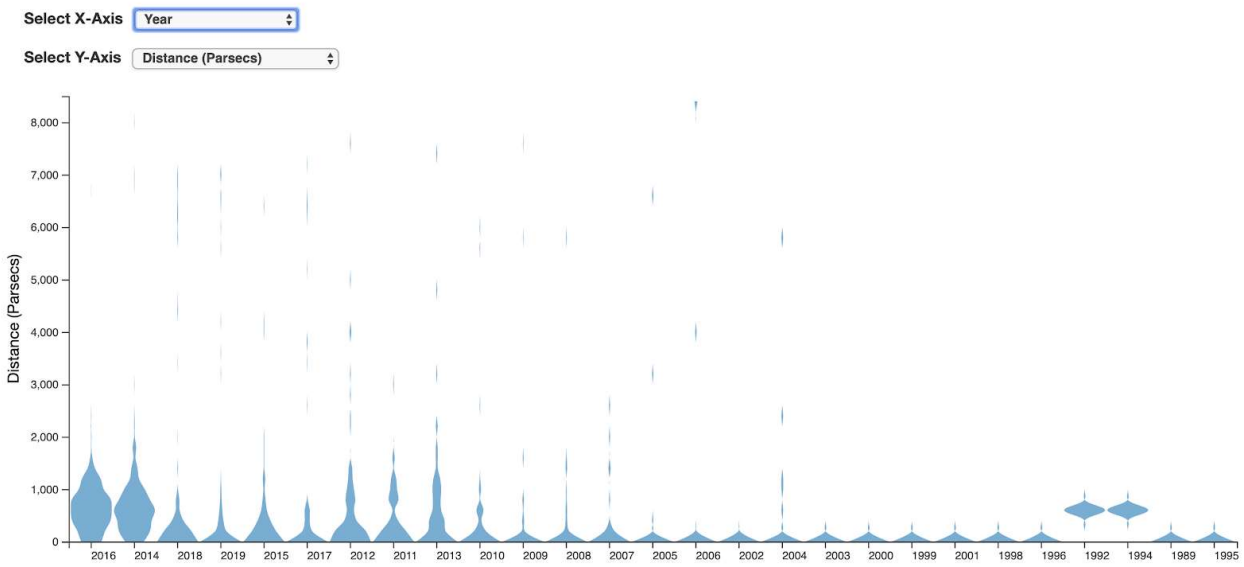

Clear Brushes
By Brushes
Clear Filter
Demo
Github
Planet Explorer
Authors: Daniel Malmuth and Liam Moynihan
The NASA Exoplanet Archive catalogues every planet that has been discovered outside of our solar system. To explore it, the walkthrough is a good place to start.
Wall
Previ
Discovery Method
Distance (Parsecs)
Mass (Earth Masses)
Transit
Radial Veloc
Microlensing
Imaging
Transit Timi
Eclipse Timi
Pulsar Timin
Orbital Brig
Pulsation Ti
Disk Kinemat
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000


Select X-Axis
Distance (Parsecs)
Select Y-Axis
Mass (Earth Masses)
Mass (Earth Masses)
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
0.00
1.00k
2.00k
3.00k
4.00k
5.00k
6.00k
7.00k
8.00k
Distance (Parsecs)

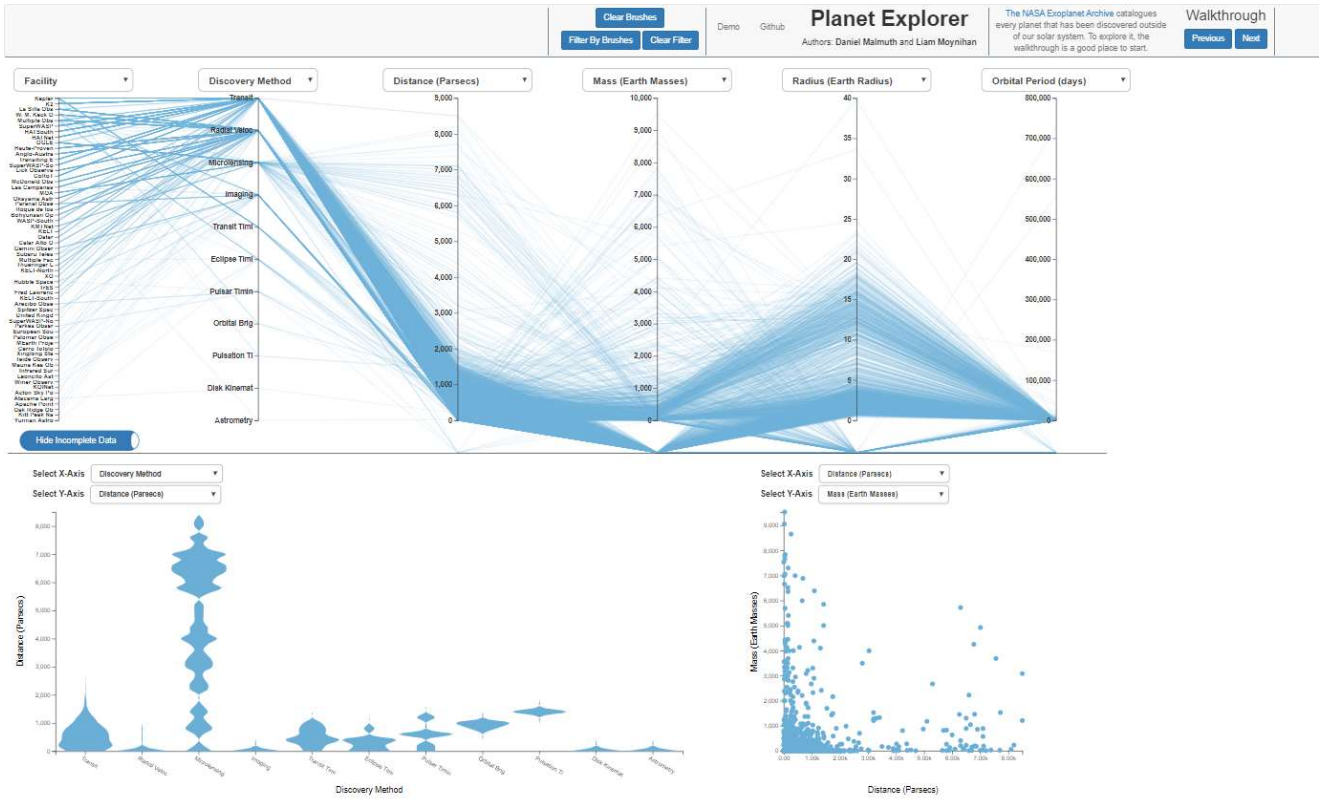
The next issue we ran into was whenever the x- or y-axes were using a discrete selection from the dropdown (e.g. year discovered) or a categorical selection (e.g. discovery method), this would cause the scatterplot to stack points in a very small number of bins, like this:




This causes issues with seeing the density of the points for certain y values. We then decided to add a violin plot in order to properly visualize these selections.

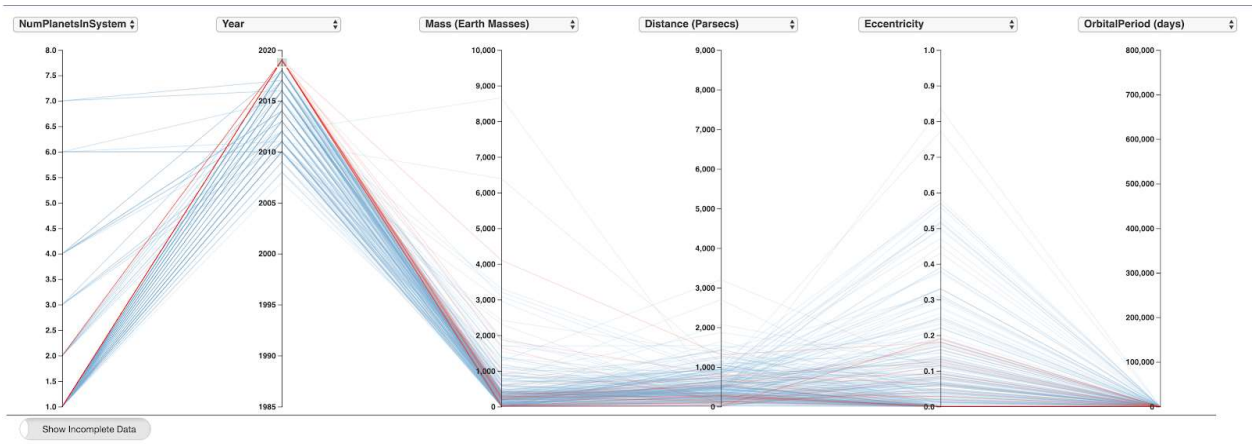



This gave us the final comprehensive visualization:



## Implementation

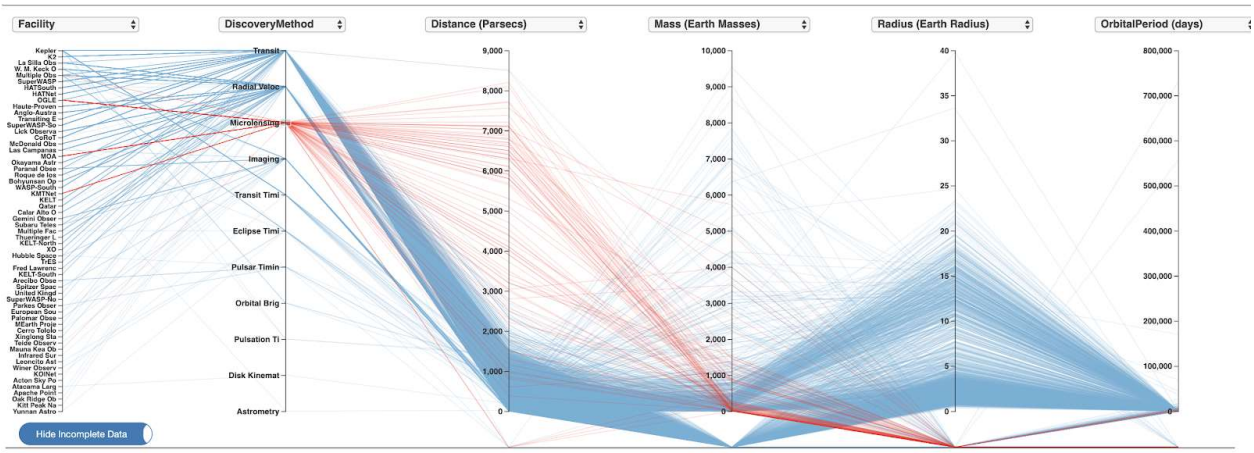
The main intent of the interaction was to both filter and to reorganize. The brushes were the main tool used to filter in the parallel axes and scatterplot, and the dropdowns acted as the filter for each of the visualizations. Our goal was to build a very general framework that could easily show someone the answer to any questions concerning some subset of the data. For example, if one were curious about the shared attributes between exoplanets that were discovered in the last year, this is extremely easy to implement:




Additionally, if someone cared more about eccentricity than distance when comparing planets of a certain year, the drop down makes this very easy. All known details of each planet are also encoded within clicking on any of the points within the scatterplot.

## Evaluation

We learned many interesting pieces of data from our visualization that we had not originally imagined. For example, we discovered that the microlensing discovery method does not give us data on the radius of the exoplanet, as indicated by this image:




We see that all of the lines go down to “incomplete data” at the Radius axis. These kinds of insights are much more interesting to discover through experimentation.

One implementation that we feel might help the plot is to add in a table with data on the different planets that links to the parallel axes. This would give a bit more data about the individual planets and would give a direct way to highlight a particular planet of interest.

Another point we discussed was to add a hover option to the parallel axes that shows how many points are currently being selected, either with or without brushes. This helps give a better idea of the density of the selection.

To make the visualization a bit more visual appealing, we could have been a bit more alluring with the choice of styles. While the visualizations are cohesive, they are rather bare-boned and could use a bit of styling.

Our visualization works very well with what we were hoping to accomplish: a general framework for exploring NASA exoplanets that gives details on demand.