To Infinity and Beyond: Exploring Exoplanets Beyond Our Solar System

Nick Coury, Aidan Johansson, Frank Rodriguez, Ayushi Shirke*
Northeastern University

Abstract

In the past few years, astronomers are locating and discovering many new exoplanets, which are located beyond our solar system. These planets all have different sizes, masses, and types, and some are even predicted to sustain life! Despite these revolutionary findings, not much information is widely known about these exoplanets, which is mainly because they have obscure names – random strings of letters and numbers, making it very difficult to look them up and learn about them. Moreover, it is hard to visualize the sizes and distances of these celestial bodies from just a table, leading to a need for comprehensive yet informative data visualizations to present these findings not only for the public to stay-up to date, but also for scientists to explore different trends.

Keywords: Space discovery, Exoplanets.

GitHub Repository Link: https://github.com/DS4200-S22/final-project-exoplanet-visualization.

1 Introduction

Space has been charted and mapped by astronomers and scientists for thousands of years, resulting in a lot of knowledge and theories regarding our solar system and the astronomical objects within it. Only recently, with NASA's Kepler Telescope, have researchers been able to map out regions beyond our solar system resulting in the discovery of many new planets, called exoplanets, which have a variety of different properties such as mass, distance away from Earth, and stellar magnitude or brightness. While these properties are certainly interesting to learn about, it is also useful to determine if there are any correlations between different planets and their masses, radii, or planet-type. There could be hidden trends between these various attributes, which were previously unrecognized due to the lack of visual representation of these discoveries [2]. As a result, this visualization tool focuses on finding trends between exoplanet features, so that anyone looking to learn more about this research can clearly identify potential correlations, and key facts. We also want to emphasize that it is difficult to look up these exoplanets and learn about their various attributes, so our visualization presents the attributes of the exoplanets as well.

2 Related Works

Due to the primary data source having an inordinate number of tables pertaining to each exoplanet, we will use the data formatting techniques outlined in References [1] and [5] to create table templates in both D3 and HTML. By formatting

We referred to published visualizations as inspiration and found that the ones from References [3], and [4] were very good at charting positional data and orbits of the exoplanets. While at first, we were inspired to make a similar visualization that focuses on the spatial distance between Earth and the exoplanets, it soon became clear that this would become overwhelming to a user since the visualization would be too complicated for hundreds of exoplanets that it needs to support.

As a result, we concluded that we would want to focus our visualization on analyzing trends and comparisons among different exoplanets rather than their positions or orbits in space itself.

3 Use Case

The purpose of the proposed visualization tool is to present the findings and discoveries of different exoplanets beyond our solar system for improved human perception. If I consider myself to be an end-user, then I would hope that the visualization tool clearly showcases different trends between quantitative attributes of planets. I would also like to be able to learn more about each of the planets such as their masses, radii, year of discovery, etc. This field is lacking in informative visualizations, which contributes to how overwhelming it feels to keep track of the discoveries or even look-up new exoplanets since their names consist of seemingly arbitrary letters and numbers.

3.1 Specific Case

In a more specific case, this visualization could also be expanded to not only present information, but by grouping together exoplanets with similar features, scientists and researchers may be able to apply these quantitative values to answering more complicating research questions or performing more complex calculations. As a result, not only is it necessary to present these findings and group them by similarities, but also to apply their attributes to further research.

3.2 General Case

To summarize, this tool would firstly aid in the presentation of the discoveries of exoplanets, which directly contributes to educating any end-user and informing them of the recent discoveries. This visual presentation of the data could help display potential existing trends between some of the quantitative data, which could inspire further investigation and research. These will be based on the exact data (next section) that is used as well as exploring different attributes to determine which show an interesting trend, if at all.

4 Data

Primary Data Source:
https://exoplanets.nasa.gov/discovery/exoplanet-catalog/

Secondary Data Source: http://exoplanet.eu/catalog/

Our primary data source is from NASA and has records of exoplanet names, distance from Earth, mass, stellar magnitude, and discovery date. There is more data in the filter section such as the planet type, method of discovery, and planet system. By combining all the data across the filters, we have a rich data set describing the discoveries of exoplanets. The filters represent categorical data, while the data table has positional data (in the form of distance from Earth), and ordinal sequential data for the other three attributes (planet mass, stellar magnitude, and discovery date).

Our secondary data source is from a website called Exoplanet.eu but has robust data describing quantitative attributes of the exoplanets. Since this dataset has similar information to NASA's data, we will consult it only to reference the additional data not covered by NASA. This dataset contains information such as the radius, period, and eccentricity of the exoplanet, so this will be useful to analyze additional attributes between the two datasets.

5 Implementation Plan

5.1 Task Analysis (Refer to Appendix B)

Task ID#	Domain Task	Analyze Task (high-level)	Analyze Task (low-level)	Analytic Task (low-level, "query")
1	Display which attributes or features are most strongly correlated	Consume → Discover	Explore	Compare
2	Group together exoplanets with similar attributes	Consume → Discover	Browse	Identify
3	Display basic info of planets and allow in depth exploration of any of them	Consume → Present	Lookup	Summarize

The primary consumer of our visualization will be the exoplanet researcher although we are keeping other consumers in mind such as the researcher's associates and those without much scientific knowledge but may be interested in viewing screenshots of the visualization captured by the researchers or about the discoveries in general.

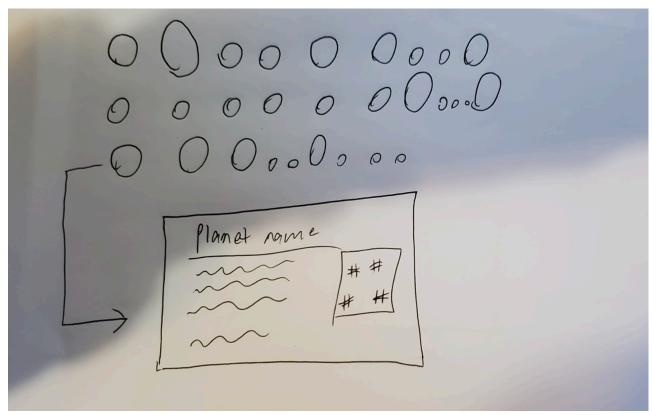
The main type of consumption for this visualization is "discover" due to the needs of the researcher. Our highest priority is to allow the researcher to view trends and commonalities between planets, something which is difficult to currently do. Finding data for any one individual planet is currently not much of a problem, so simply presenting the data which the researcher already has convenient access to isn't the main priority of the visualization. While enjoyment may be derived from the work that the researcher does, their main goal is to advance their work, and this tool is designed for that purpose.

5.2 Design Process

We will visualize NASA's data covering newly discovered exoplanets. To do this we will display which attributes or features of exoplanets are most strongly correlated, group together exoplanets with similar attributes, and display basic and in-depth information for any selected planet. HTML code will be responsible for modeling each section of our visualization (scatter plots, filter boxes, tooltips, and histograms) [5]. Importing the JavaScript Library D3 will allow our group to code all desired functionality including brushing and linking, selecting, and filtering [1]. Our CSS file will add visual understanding by adding style features such as color and font to each section of our visualization. By combining all three codebases, we will be able to effectively inform the user about newly discovered exoplanets. Once we have built a working visualization, we will begin to focus on our "nice to have's." The nice to have features include images of the exoplanet (or planet system if no such picture exists), zooming on our data plots, and dynamic filtering (filtering that applies immediately). These features were nice to have but not vital because they are not vital; they only make the visualization nicer to look at.

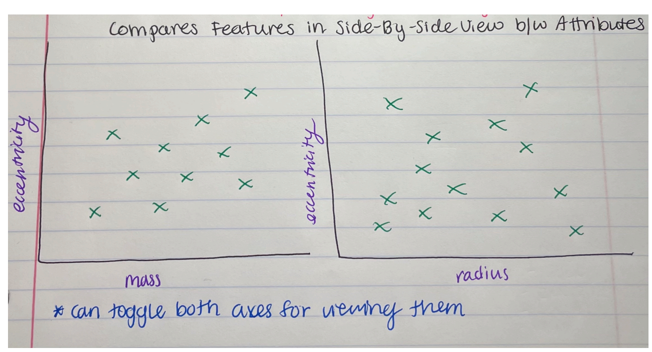
The implementation of this visualization does not require the use of any outside APIs or servers. We chose not to include either because it would make our implementation significantly more complex without adding any information for the user.

A few initial sketches include:



Sketch 6 from Appendix D

By: Frank Rodriguez



Sketch 10 from Appendix D

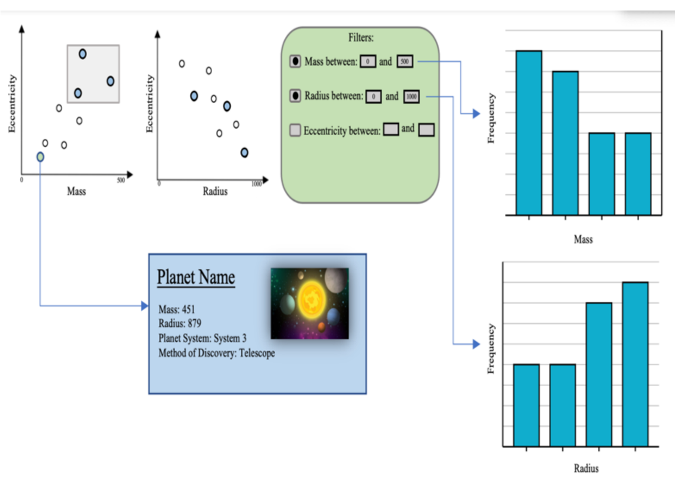
By: Ayushi Shirke

Sketch 12 from Appendix D

By: Nick Coury

In the first two sketches we include scatterplots as the visual encoding, where points (marks) represent exoplanets and colors and x and y positioning represent quantitative attributes. In the first sketch (Sketch 6), the interaction of a tooltip is shown to present the findings, which corresponds to Task ID #3 form section 5.1. In the second sketch the use of having a multi-view visualization is highlighted, and after discussing we decided to include an interaction as brushing and linking here. This will pertain to Task ID # 1 from section 5.1. The third sketch focuses on having a filtering aspect to not overwhelm the user as well as bar charts as the visual encoding to display how different exoplanets can be grouped together, which pertains to Task ID #2 from section 5.1.

After drawing inspiration from the sketches included in Appendix D, we decided upon this final digital sketch:

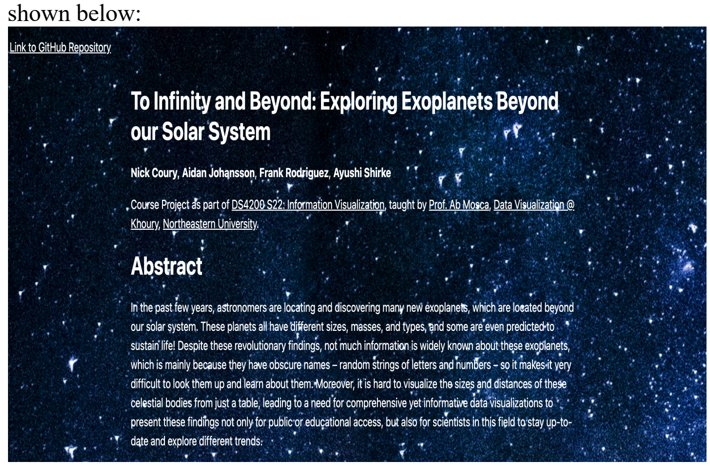


The digital sketch from Appendix E incorporates the three sketches above as they focus on each of the three Domain Tasks listed above. We did not change much from the original sketches in the final design; we simply combined each of the sketches and incorporated interaction in the form of a tooltip, brushing and linking, and filtering. Additionally, we did not alter the design after conducting the usability testing since the major changes to add were simply contextualizing the visualizations.

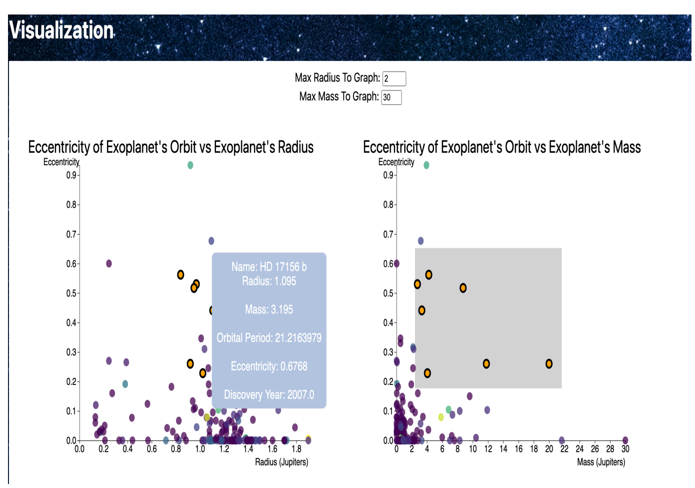
6 Visualization Design

Our final visualization tool is an HTML webpage using D3, a JavaScript library, to create visualizations that describe exoplanets. These will be based on our final sketch, and there will be interaction to display more detail for each planet, brush and link exoplanets, or filter by quantitative attributes. This is being done so we can provide a large scale, zoomed out viewpoint of the data initially so that it is easily perceived for any audience, and they can then find more information about a specific planet or data feature by exploring further. Practically speaking, this looks like having planets in a blob where they are represented by a shape, colored based on their discovery year; these shapes can also exist on a graph that displays some descriptive stats on a scatter plot. There will be two scatter plots in a side-by-side view so that different features can be compared at once, and even further through brushing and linking. Then, users can click on or hover over a certain exoplanet, and this will trigger a tooltip to pop up with more stats about the planet that are more in-depth but would be overwhelming to display at first. Additionally, will also have bar charts that can be filtered based on certain criteria, but we are not implementing these into separate graphs at the landing, so the dashboard does not become overwhelming.

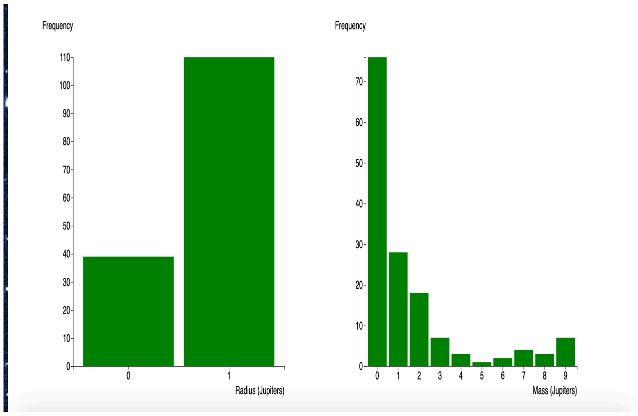
Introduction to the visualization on our webpage is shown below:



The scatterplots of eccentricity of orbit versus radius and eccentricity of orbit versus mass is shown in a side-by-side view below. The interactions are both brushing and linking between the scatterplots and a tooltip to present the attributes for each of the planets.



Below the scatterplots are two side-by-side bar charts that show the frequencies of exoplanets with the same radius or mass for the current filter which can be toggled in the image above. This allows for easy understanding of the distribution that exists and similarities between exoplanets that may occur.



This purpose of this visualization is to achieve the domain tasks. The visual encodings of scatterplots to display trends along with the bar charts to group similar exoplanets together, and the interactive tooltip, all together achieve the purpose. The reason that only two quantitative trends are visualizes is because the correlation matrix in Appendix B predicts that eccentricity is the most likely attribute to have a potential trend with any other quantitative data. Also, mass, radius, and eccentricity are very important attributes to perform further calculations such as determining the orbit of the planet or the acceleration due to gravity it experiences. Thus, this visualization design encompasses all the domain tasks and allows for easy user comprehension through interaction.

7 Discussion

As we obtained the data and used the visualization tool, we were able to gain a few insights about the trends that are displayed. The first is that as radius and mass increase, the eccentricity of the orbit seems to decrease. This means that as the planet becomes larger, the orbit starts to form a circular orbit. This is an interesting conclusion due to our visualization. We also found that by filtering, either some of the planets were very small and had a radius close to zero Jupiters (radius of Jupiter), or most of the planets were very big and had a radius close to that of Jupiter.

After completing the visualization tool, one major lesson we learned was the significance of carefully breaking down a large deliverable or project such as this into smaller and more manageable milestones to work more efficiently in the long run. We also deeply understand the design process and the significance of each step to make sure that the visualization remains relevant to the domain tasks and the process is improved iteratively.

8 Conclusion

In order to build this visualization, it was necessary to first understand the domain situation, which then translate to domain tasks for the visualization to achieve. Next, we obtained and cleaned relevant data and after exploring it, we created preliminary sketches to tackle each of the domain tasks. Finally, we combined the sketches that worked together into a final visualization that we encoded in HTML, D3 (a JavaScript library), and CSS. To verify our design, we conducted usability testing, and continued to make improvements to the visualization. We can conclude that after this whole process, our visualization does relate to and achieves the domain tasks we listed above. Overall, we achieved our goals and tasks outlined, but as this is an iterative process, we can continue to test it and improve it.

As of now, an improvement we can make in the future is to link images of each of the planets in the tooltip as well. Additionally, it might be beneficial to the user to have additional context with additional tooltips to present more information about the meaning of the unit Jupiters or what different eccentricity values mean. This would especially be helpful for end-users who may not be experts in the field.

8 Acknowledgments

Thank you to Professor Ab Mosca for all their help and support throughout this project!

References

[1] Agrawala, J. H. (2018). Converting Basic D3 Charts into Reusable Style Templates, vol.24, no. 03. *IEEE Transactions on Visualization & Computer Graphics*, 1274-1286.

[2] Brendan Crill, N. S.-G. (2018, March 12). *Key Technology Challenges for the Study of Exoplanets and the Search for Habitable Worlds*. Retrieved from arxiv: https://arxiv.org/pdf/1803.04457.pdf

[3] *Creating an interactive "exoplanets in orbit" visualization*. 2014, December 24). Retrieved from visualcinnamon: https://www.visualcinnamon.com/2014/12/creating-interactive-exoplanets-in/

[4] Jones, K. &. (2007). Visualizing perceived spatial data quality of 3D objects within virtual globes. *International Journal of Digital Earth*.

[5] Z. Peng, S. L. (2004). Extraction and Integration Information in HTML Tables. *The Fourth International Conference on Computer and Information Technology*, 315-320.

Appendix A: Group Purpose

This group is composed of data scientists, physicists, and computer scientists who are interested in outer space. The group was formed with the intention of visualizing exoplanet information released by NASA.

Group Goals

Our group has a goal of committing an equal effort among group members of 2-3 hours per week, per person on our project. We're aiming to achieve an A/A- in the course and are all planning to put in effort consistent with that goal. Tentatively, our goal is to start every meeting by checking in with one another, reviewing the work that we've all completed individually and together over the past week, and then talking about the work that we hope to get done in the future and discussing any topics that have come up.

Group Member Roles/Responsibilities

Research will be done by all group members. Each group member will handle one of the following roles: HTML, CSS, JavaScript, or project manager.

HTML:
- Responsible for the HTML code of the visualization
- Needs to work with the group members who are responsible for CSS and JS
- Needs to update the PM on any hang ups or bugs in code

CSS:
- Responsible for the CSS code of the visualization
- Needs to work with the group members who are responsible for HTML and JS
- Needs to update the PM on any hang ups or bugs in code

JS:
- Responsible for the JS code of the visualization
- Needs to work with the group members who are responsible for CSS and HTML
- Needs to update the PM on any hang ups or bugs in code

PM:
- Responsible for coordinating progress among all coders
- Responsible for communications outside of the group
- Needs to be able to help debug HTML, CSS, and JS

Ground Rules

We'll plan to text throughout the week to answer small questions and adjust meeting times if needed. Otherwise, we'll meet up weekly over Zoom on Tuesday's at 5pm by default, but at different dates/times, if need be, and sometimes in person if convenient for everybody. Discussions and decisions will be intentionally inclusive of everyone's opinions and preferences. We believe that dissenting views should be voiced within the group to be discussed as everyone deserves the chance to have their thoughts heard. We expect that everyone in the group will put in an equal amount of effort, with everyone leveraging their own skills to help in creating our final product. With that said, we do also understand that some weeks will be busier for some group members than others, so we're flexible with group members needing to step back some weeks and having other members fill in for some of their work in the group. To hold one another accountable, our greatest tool is communication, and we plan to use it greatly throughout the project. Letting one another know when we're falling behind in terms of project work and when we're doing well and overachieving, as well as voicing any other thoughts and concerns is integral for completing our project and respecting all group members.

Potential Barriers and Coping Strategies

Barriers:
1. Not being able to attend group meetings without communicating prior
2. Not devoting the same amount of time to the project as other group members

How to Handle Barriers:
- If a group member did not tell the others that they couldn't make it, ask the group member to communicate better next time. If the group member cannot make a meeting, they should continue working on their own at a time that works for them.
- If a group member is not contributing to the project, it is up to the rest of the group to decide how to encourage their participation while still meeting the deadlines. This could mean dividing up work, changing the work schedule, or contacting the professor (last resort).

Group Charter Amendments

After working together for several weeks, we've been consistent with following the group rules and expectations that we laid out during the beginning stage of the project. We anticipated that there would be times when group members were especially busy and would have to miss meetings, and this has been the case, but absences have been generally well communicated so that everyone else is aware ahead of time that they might have to pick up some of the work that's been left off. We've been consistently meeting on Tuesday evenings over Zoom to work collaboratively, and we've communicated to one another when we need to shift the time later or change the day to Wednesday on a given week.

One way that we could be more consistent with the rules laid out in our group charter would be to communicate more over text outside of our meetings, as most of our communication occurs during the Zoom meetings themselves. It might be to our benefit to discuss some of the ways we plan to delegate project work and general expectations for the week's project milestone before meeting together so that we can get right into doing the work instead of spending time going over the document and preparing to start on work.

There haven't been any major disagreements during the project, and everyone has been very respectful in voicing their opinions and respecting the opinions of others.

Positive things that we've noted from the group are as follows:

Aidan: "Everyone is very kind and easygoing. I find that meeting up to do work is fun and everyone is productive during the time we meet up. Ayushi has put a lot of work into going above and beyond to nicely format our group work and submit it after we're done meeting. Frank is great at data analysis and helped a lot with cleaning our initial data and formatting it in a usable way. Nick is very hardworking and took the lead on using his creativity to make a nice final digital sketch."

Ayushi: "This group has been so productive and friendly. We all collaborate and work together very well, and everyone is very supportive and willing to take on more work if needed. I have really enjoyed working with this team and feel more than comfortable asking questions or asking for help. I think Aidan has been great at keeping everything organized and managed in a shared Google Drive folder and keeping everyone on track. Frank has been super great coming up with creative ideas and thinking of solutions such as with how to clean our raw data. Nick has been very great at managing the upkeep and appearance of the working document and keeping everyone on track. Overall, everyone is working very hard on this project, communicating well, and is super fun to get-to-know and work with!"

Frank: "The group is great about both communicating their availability and being willing to make up their own individual work if they must miss. I never feel that because someone is unavailable during the group meeting time, they aren't willing to do their share of the work. I've especially appreciated Aidan's willingness to work outside of meeting times, Ayushi's great organization and work ethic, and Nick's consistency in creating great work and being available."

Nick: "I believe this group has worked well together and has stayed true to the original group charter. Each team member has performed their roles very well and it has led to greater synergy. I have really appreciated watching Frank's thought process when it came to sketching ideas on what to visualize. Ayushi has been a great coordinator and has a great data management hierarchy spanning between GradeScope, Canvas, and Google Drive. Thanks to Aidan we have had detailed notes of what we accomplished per meeting. This was particularly useful when I missed a meeting and he quickly caught me up to speed.

Appendix B: Data Exploration

Data Review

Data Types

Attribute Name	Data Type
Planet	Categorical
Mass	Ordinal Sequential
Radius	Ordinal Sequential
Period	Ordinal Sequential
Semi-Major Axis (a)	Ordinal Sequential
Orbital Eccentricity (e)	Ordinal Sequential
Orbital Inclination (i)	Ordinal Sequential
Angular Distance	Ordinal Sequential
Discovery	Ordinal Sequential

Data Review

	mass	radius	orbital_period	semi_major_axis	eccentricity	angular_distance	inclination	discovered
count	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000
mean	2.268748	1.088041	2227.406247	1.242185	0.078591	0.031985	84.144951	2008.872483
std	3.972018	0.395764	15418.958017	7.057402	0.144978	0.181281	11.531985	2.948404
min	0.014500	0.132000	0.736548	0.014110	0.000000	0.000015	25.000000	1995.000000
25%	0.503000	0.970000	2.788491	0.039300	0.000000	0.000106	84.960000	2008.000000
50%	0.940000	1.171000	3.662382	0.047400	0.020000	0.000195	87.070000	2010.000000
75%	2.244000	1.310000	5.113400	0.060600	0.087000	0.000441	88.658000	2011.000000
max	30.000000	1.900000	164250.000000	68.000000	0.933660	1.725888	90.760000	2012.000000

The data indicates that it would be good to visualize orbital period since it has the largest standard deviation. It also seems as though the medians and the means are relatively similar for most of the attributes, so this indicates there is little to no skew in the data.

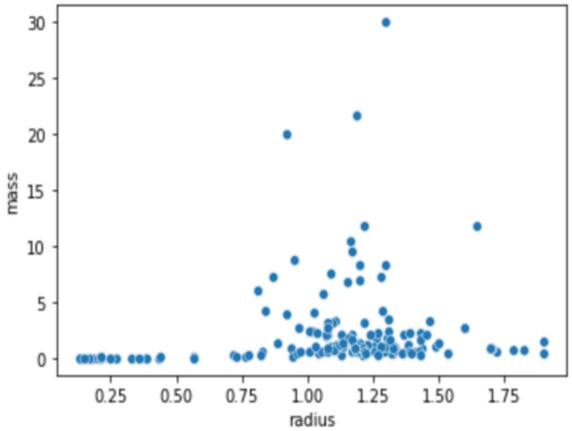
Issues with Data

- The first issues we noticed were that there were many unpopulated columns and discrepancies in the data collection.
- There are some columns with very few values, so it will be harder to observe trends for those attributes.
- Some of the columns are very hyper specific so the attribute may not be relevant to study in our case.
- A potential bias is that the data collected is from NASA, which is a public governmentally funded company, meaning that if private or international researchers have discovered other exoplanets they may not be included in this source.

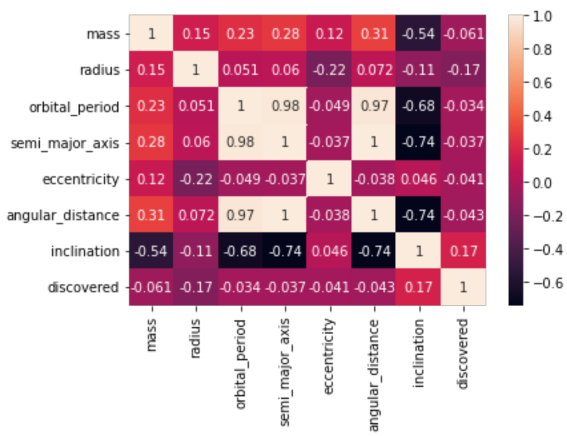
Insights

After having processed the data, a few trends were identified. There is a decreasing trend among orbital eccentricity and orbital period. We also discovered that there is a bell-curve distribution between mass and radius approximately centered at 1.71. Surprisingly, there was not a strong correlation between the semi-major axis and the angular distance. Our group focused on cleaning the data to remove any null values and get rid of any discrepancies. Due to most of our attributes being ordinal, we do not predict having many bar charts – something we did not expect.

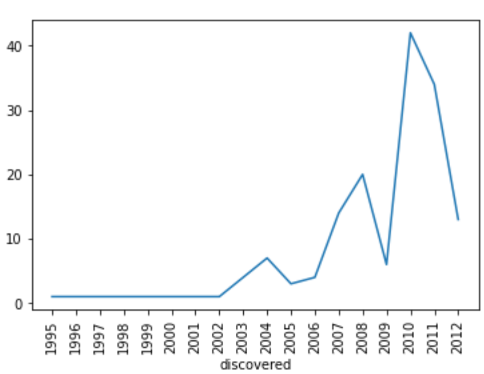
Screenshots



Above is a bell-curve distribution between mass and radius. Both attributes are measured relative to Jupiter. This is a scatterplot visual encoding and uses points as marks, as well as x and y positioning for channels.



Above is a correlation matrix visual encoding between exoplanet attributes and their frequency. It uses area as marks, as well as color for channels. We discovered that there is significant correlation between inclination, mass, orbital period, semi-major axis, and angular distance. In particular, the inclination seems to be most strongly correlated with the other listed attributes.



We found this line plot visual encoding to be interesting because it shows which years had significant initiatives, advancements, and focuses on space exploration.

Data Snippet

	mass	radius	orbital_period	semi_major_axis	eccentricity	angular_distance	inclination	discovered
count	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000	149.000000
mean	2.268748	1.088041	2227.406247	1.242185	0.078591	0.031985	84.144951	2008.872483
std	3.972018	0.395764	15418.958017	7.057402	0.144978	0.181281	11.531985	2.948404
min	0.014500	0.132000	0.736548	0.014110	0.000000	0.000015	25.000000	1995.000000
25%	0.503000	0.970000	2.788491	0.039300	0.000000	0.000106	84.960000	2008.000000
50%	0.940000	1.171000	3.662382	0.047400	0.020000	0.000195	87.070000	2010.000000
75%	2.244000	1.310000	5.113400	0.060600	0.087000	0.000441	88.658000	2011.000000
max	30.000000	1.900000	164250.000000	68.000000	0.933660	1.725888	90.760000	2012.000000

Above is a small portion of the data that is stored in a csv file. This has been cleaned to remove any null values or missing value discrepancies.

Task Analysis

Task ID#	Domain Task	Analyze Task (high-level)	Analyze Task (low-level)	Analytic Task (low-level, "query")
1	Display which attributes or features are most strongly correlated	Consume → Discover	Explore	Compare
2	Group together exoplanets with similar attributes	Consume → Discover	Browse	Identify
3	Display basic info of planets and allow in depth exploration of any of them	Consume → Present	Lookup	Summarize

Final Review of Task Analysis

The primary consumer of our visualization will be the exoplanet researcher with whom we interviewed, although we are keeping other consumers in mind such as the researcher's associates and those without much scientific knowledge who may view screenshots of the visualization captured by the researchers.

The main type of consumption for this visualization is "discover" due to the needs of the researcher. Our highest priority is to allow the researcher to view trends and commonalities between planets, something which is difficult to currently do. Finding data for any one individual planet is currently not much of a problem, so simply presenting the data which the researcher already has convenient access to isn't the main priority of the visualization. While enjoyment may be derived from the work that the researcher does, their main goal is to advance their work, and this tool is designed for that purpose. In the end, discovery of trends and groups is both the main intent for the visualization and the main benefit that the researcher could derive from a visualization.

Appendix C: Interview

End-User Persona

The persona is that of a researcher on space that has access to all the data that we have but does not have prior experience with visualizations or the techniques used to generate them. The researcher spends a lot of time each day making decisions about where to focus attention and resources for further research and studies, along with examining space objects/planets deeper. The visualization could help with this greatly, as it could expose patterns that take up a lot of the work in describing patterns in space objects and their attributes. Currently, these patterns need to be discovered by just eyeballing them, and are generally not found at all, so a lot of extra time and resources go into trying to discover basic relationships. A visualization that displays these correlations or relationships clearly would enable the researcher to spend their time on more advanced things and see trends between object attributes that may be difficult to discover at first.

Interview Script

1. How familiar are you with exoplanets?
    a. What information about exoplanets do you find most valuable?
    b. Is location or planet stats more interesting information to you?
    c. If we modeled the location of exoplanets, how important is each planet's stats to you?
    d. Would size be interesting to you to visualize?
    e. Do you ever compare exoplanets in your work?
        i. If you do compare exoplanets, how do you do it?
        ii. If you don't, how would you do it? (mass, size, temp, etc.)
2. Why exactly do you need a visualization tool?
    a. Would a visualization be helpful in allowing you to understand where in space the exoplanets are?
    b. Do you need to be able to better compare and contrast different exoplanets and their features in a way that's difficult without a visualization tool?
    c. Do you need the ability to visually group exoplanets by their different features and interact with individual exoplanets within a given group?
    d. Would a visualization tool be helpful in allowing you to share your findings with other researchers?
    e. Would a screenshot of your visualization tool ideally allow you to share your data with those that understand little about the realm of exoplanets?
3. What specific actions do you want to perform?
    a. Is it useful to have filters that can be toggled (mass, size, temp, etc.)?
    b. How do you want to switch between exoplanet information?
    c. Do you want one big visualization, which you can drill down into for more information? Or, do you want one overall visualization with sub-visualizations for each exoplanet?
    d. How do you want to reset the visualization (click a button)?

Interview Notes

1. I am familiar with exoplanets, as I research them in my daily work. That said, I examine most of them individually
    a. I need to understand the basics about an exoplanet before I can understand other things about it.
    b. Space is hard to conceptualize spatially, as we don't really know where (or if) it starts and ends, so more concrete planet stats are more important
    c. Each planet's stats are important, but it would be more interesting to see how they compare to each other
    d. Yes
    e. I compare them semi frequently, though this usually happens through examining them side by side and in-depth, which takes a lot of time
        i. Answered above
        ii. It's generally done through comparing size, temperature, weight, etc.
2. To make sharing my research easier.
    a. A visualization would be helpful in allowing me to understand where exoplanets are if it were designed accurately, but given that they move around a lot, this could end up being more confusing than no visualization at all if it is inaccurate
    b. It would be helpful to compare large numerical values like size and correlations between quantitative variables that can only realistically be done with computers. It would be very helpful to have a quick, high level view so I can get into more important and challenging research.
    c. This would help me in not repeating research over nearly identical or very similar planets, so grouping them together could be useful.
    d. If the tool was made to be understandable by anyone, I think it would be helpful to share. Otherwise, I think I would have to explain what I'm showing and that's too much effort.
    e. Yes, I think being able to have all the necessary information in a single screen would make sharing results really easy.
3. Really it comes down to ease of comparison.
    a. I think that given the large number of exoplanets out there, it would be nice to be able to quickly change the way my view is filtered by different variables through some sort of toggle.
    b. I think being able to see a subsection of planets would be more helpful than seeing one planet and its info
    c. Overall, I want to be able to share one screen. If you can give multiple visualizations on one screen, then you should. If the one screen is overly complicated, I think it makes more sense to have just one visualization.
    d. I think being able to instantly reset the visualization to its start state would make switching between comparisons much easier. Maybe if the filters are toggleable, there could be a button to uncheck all filters?

Interview Results

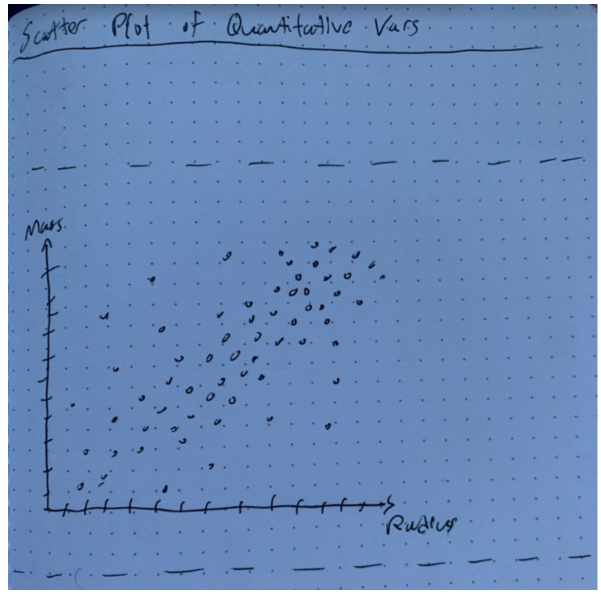
I am familiar with exoplanets, as I research them in my daily work. That said, I examine most of them individually. In terms of a data visualization, I need to understand the basics about an exoplanet before I can understand other things about it. This is because space is hard to conceptualize spatially, as we don't really know where (or if) it starts and ends, so concrete planet stats are more important. Each planet's stats are important, but it would be more interesting to see how they compare to each other.

Currently, I compare exoplanets semi-frequently, though this usually happens through examining them side-by-side and in-depth, which takes a lot of time because I need to compare aspects such as size, temperature, and weight. As of now, a visualization would be helpful in allowing me to understand where exoplanets are if it were designed accurately, but given that they move around a lot, this could end up being more confusing than no visualization at all if it is inaccurate. For this reason, it would be useful to have a feature to instantly reset the visualization to its start date.

It would be helpful to compare large numerical values like size and correlations between quantitative variables that can only realistically be done with computers. It would be even more helpful to have a quick, high-level view so I can get into more important and challenging research. This would help me in not repeating research over nearly identical or very similar planets, so grouping them together could be useful. As a result, it would be nice to be able to quickly change the way my view is filtered by different variables through some sort of toggle (which could also be undone). This way, analyzing a subsection of planets would be more helpful than only focusing on one planet at a time.

The tool should be made to be understandable by anyone, so I think it would be helpful to share. To do this, I think being able to have all the necessary information in a single screen would be ideal; otherwise, I think I would have to explain what I'm showing and that's too much effort.

Appendix D: Design Sketches

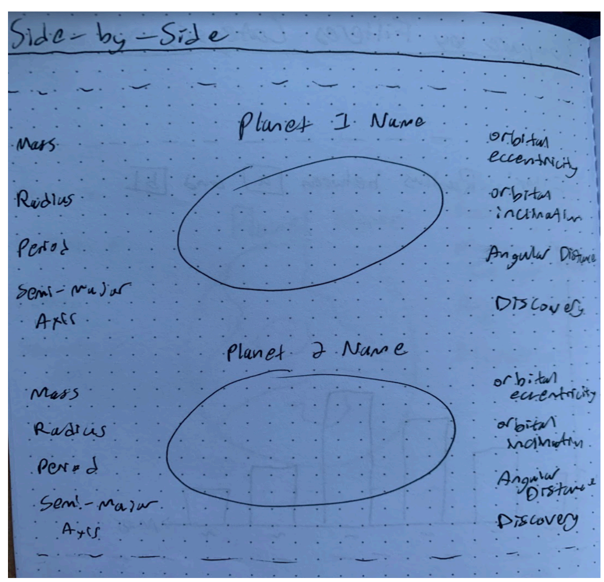


Sketch 1:

By Aidan Johansson

The points are shown to be circular to represent the generally round nature of exoplanets. Their size is consistent to allow for accurate comparison and visualization of trends within the data.

This visualization mainly address task #1: "Display which attributes or features are most strongly correlated" by plotting two attributes against one another.

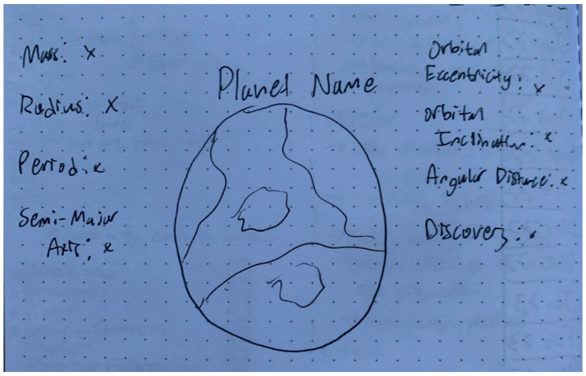


Sketch 2:

By: Aidan Johansson

The circular shapes representing exoplanets are intentionally large and round to give a visually appealing planetary image. The information relevant to specific planets are smaller and off to the side, as to not be too aggressive and in the face of the user, yet still easy to find and view.

Due to the two planets and their data being placed one beside the other, this sketch addresses both task #1: “Display which attributes or features are most strongly correlated” and task #3: “Display basic info of planets and allow in depth exploration of any of them.”

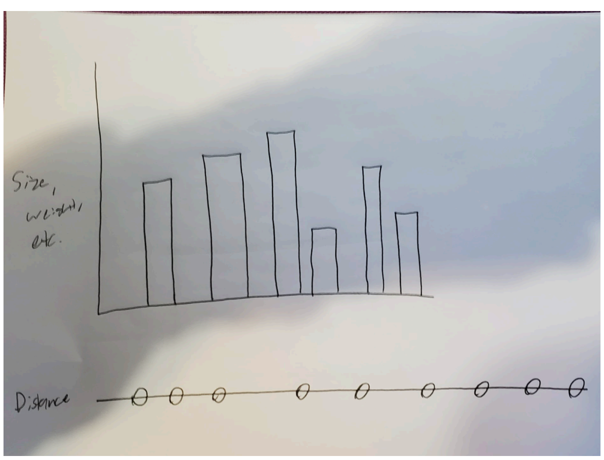


Sketch 3:

By: Aidan Johansson

The circular shape representing the exoplanet is intentionally large and round to give a visually appealing planetary image. The information relevant to the specific exoplanet is smaller and off to the side, as to not be too aggressive and in the face of the user, yet still easy to find and view.

This sketch address task #3: “Display basic info of planets and allow in depth exploration of any of them” due to the main advantage of the view being to look deeply at one planet, without any other planets immediately present.

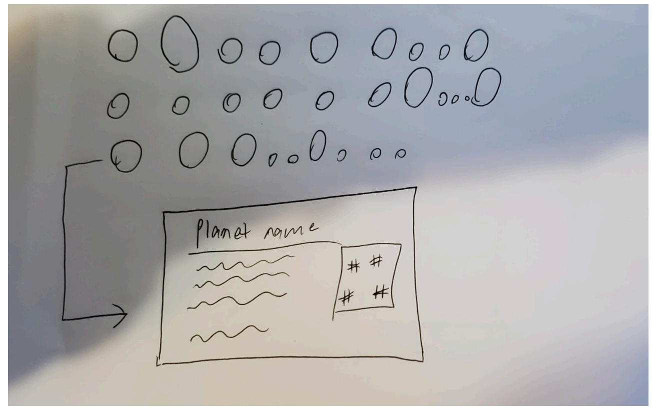


Sketch 5:

By: Frank Rodriguez

The first sketch is to show planets’ main features in different visualizations, i.e., displaying their size or weight (or both in separate charts) in a bar chart or similar encoding, and then using distance or location metrics to show approximate spacing of how far they are in another section.

This is mainly for task #3, as it is made to explore different attributes of the planets in greater detail with different encodings.

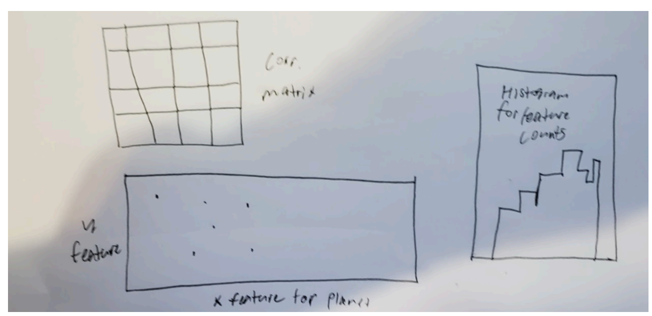


Sketch 6 (Favorite 2)

By: Frank Rodriguez

This visualization uses circles to represent individual exoplanets in their approximate location in a specific region. They can each be selected to view more detailed information and other attributes. There can be additional encodings in these tooltips, which would have marks such as points for scatter plots or areas if they are bar charts to show relationships between significant attributes through channels of positioning in x and y.

Since the planets are grouped together by region, this directly contributes to task #2: “Group together exoplanets with similar attributes”. Once they are individually selected, they can be analyzed in more detail, which directly contributes to task#3.

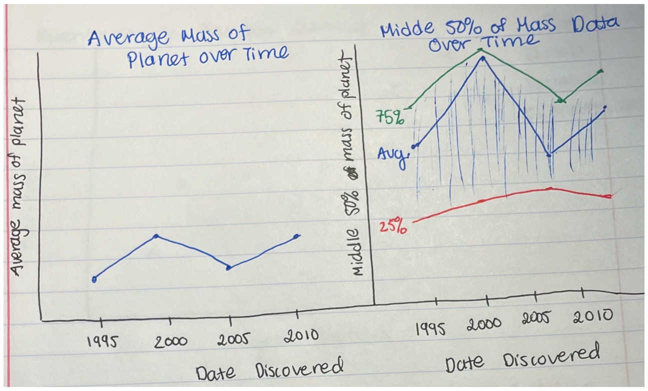


Sketch 7

By: Frank Rodriguez

These visualizations are based on the relationships shown in the correlation matrix and which attributes seem to have a more significant relationship. They are then encoded into a scatter plot which has points for marks and x and y positioning for channels. There is an additional encoding for a histogram of feature counts for each feature that is plotted in the scatter plot to better understand if any skews present because of the sample size.

Since the correlation matrix drives this features that are plotted in this visualization, these encodings directly address task#1: “Display which attributes or features are most strongly correlated”.

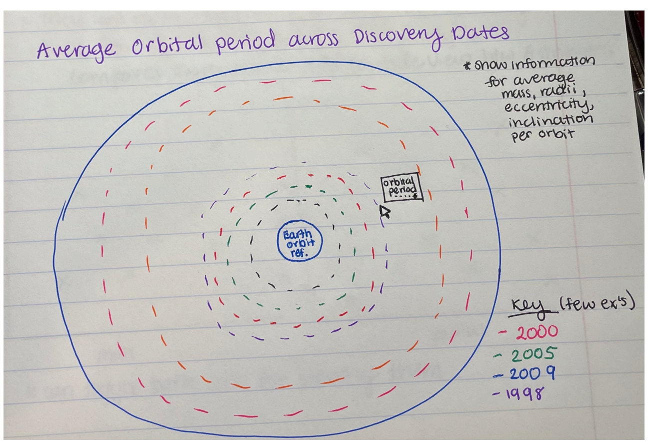


Sketch 8

By: Ayushi Shirke

In this visualization, the marks are points and lines, and the channels are x and y positioning as well as color. The purpose is to show how an average quantitative attribute (i.e. mass) changes over time, and then to also analyze how the middle 50% of the data for this feature changes over time. This is to zoom out and analyze how data near the average also changes.

Since this would be done across all the planetary data this visualization directly contributes to task#1: “Display which attributes or features are most strongly correlated” because there may be a stronger relationship with different attributes over time.

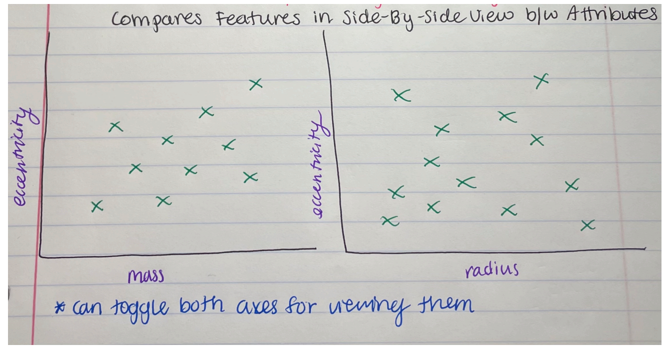


Sketch 9

By: Ayushi Shirke

In this visualization, the marks are lines, and the channels are colors and dashed lines. The idea is to explore the average orbital period over different discovery dates, and if an orbital period is selected, then additional information about its features can be analyzed.

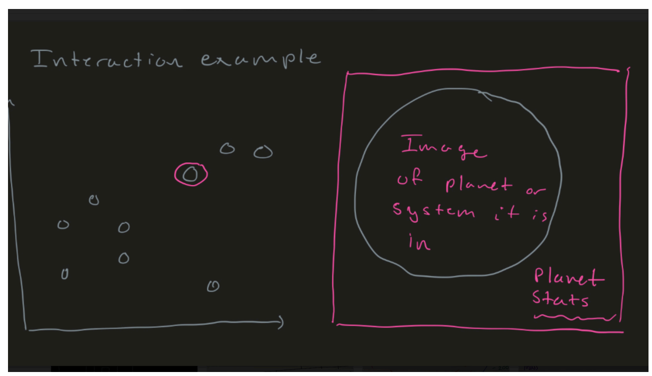
For this reason, this would directly contribute towards task#2: “Group together exoplanets with similar attributes”. This is because they will be grouped by orbital periods, which may have other implications such as similar eccentricities, and can be explored more in the tooltip.



Sketch 10 (Favorite 3)

By: Ayushi Shirke

This side-by-side view has points as marks and x and y positioning for channels. The scatter plots can be toggled to display different quantitative attributes next to each other for easy comparison. This directly contributes to task#1 to determine which attributes are more strongly correlated than others.



Sketch 11

By: Nick Coury

The circles on the plot will be highlighted as they are selected. Only one planet can be clicked and highlighted at a time, but it creates a linked visualization showing in-depth information about the planet. The primary channel of this visualization would be color showing the link between dots and in-depth information.

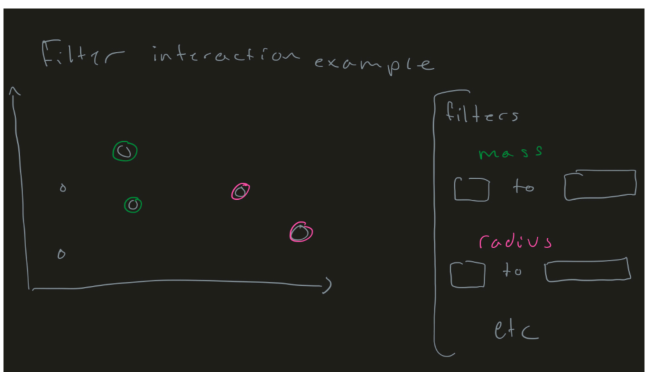
This interaction example addresses task #3: “Display basic info of planets and allow in depth exploration of any of them”.

Sketch 12 (Favorite 1)

By: Nick Coury

The rectangular bars representing the count of exoplanets found with a variable in each range are intended to allow the user to easily visualize the data through a particular lens of their choosing. Their height is proportional to the count of planets found within a given range that fit the filter, and their width is kept constant across all bars to allow for consistent comparison.

Due to the histograms use of binning and the result of a group of exoplanets with similarities, the sketch addresses both task #2: “Group together exoplanets with similar attributes”, and task #3.



Sketch 13

By: Nick Coury

The circles on the plot represent exoplanets and their relationships with each other. This visualization would have n number of highlights where n is the number of filters selected. The primary marks would be the circles on the plot. The primary channels would be the colors that link the filter ranges to the exoplanets on the plot.

The marks would directly address task #2: “Group together exoplanets with similar attributes”, and the aspect of linking addresses task #3.

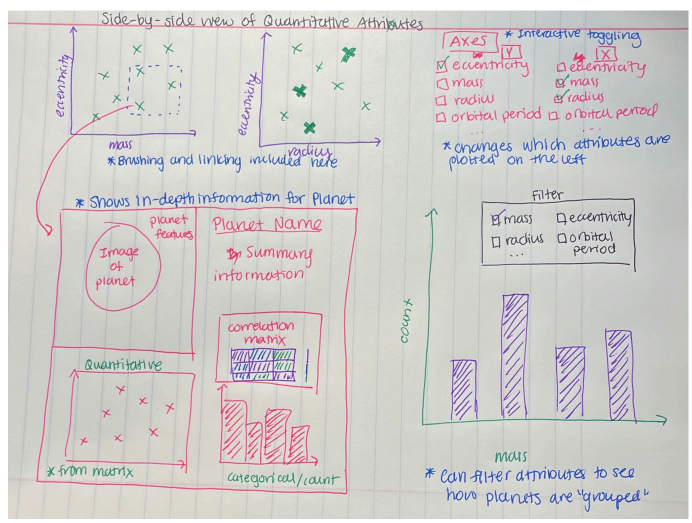
Favorite Sketches Reasoning

Our three favorite sketches were chosen due to their interactive nature and their ability to cater specifically to the user’s needs at a given time. In Nick’s first sketch, this occurs through a custom filter which allows the user to select a column to visualize, and a range of values to explore with a custom minimum and maximum value. The data is then binned and displayed as a histogram, allowing the user to not only see the number of planets that exist within that filtered range, but also to see the distribution of that factor within the range. This is effective because it allows the user to easily explore patterns in the data and get an idea of where they would like to look more closely with other visualization methods we supply.

In Frank’s second sketch, when viewing a scatter plot, the user is then able to select a given planet and view its data. When interviewing our end user, they voiced a specific need to not only see correlations between data, but also to easily pluck specific planets from those graphs and explore them individually in more detail. By creating graphs like scatter plots and then allowing a pop-up on mouse-click, the user can get a better idea of the actual exoplanets that make up the larger visualization they’re using.

Finally, in Ayushi’s third sketch, scatter plots are utilized again, but in a different manner from the previous favorite. In this sketch, 3+ variables can be explored at the same time and graphed next to one another. Utilizing brushing and linking, the end user is not only able to visualize multiple different correlations at once, they’re also able to highlight a given selection of points in one of the scatter plots and view where those individual exoplanets occur in the other scatter plot. This fulfills the need to explore correlations between data and gives the end user far more power in the patterns they’re able to discover using our visualization.

Coordinated View Visualization Tool



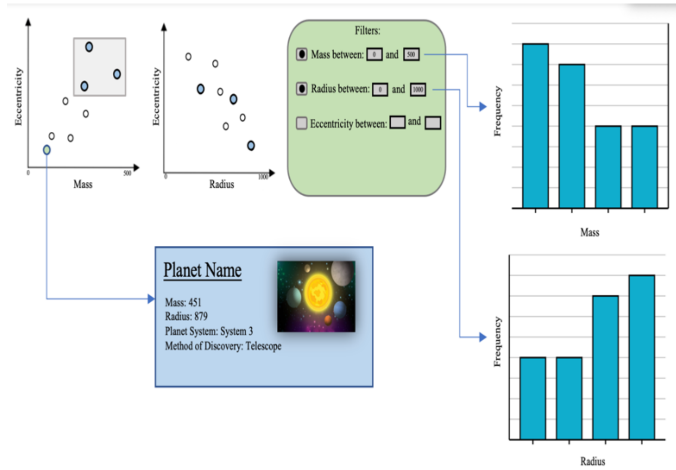
In this interactive tool, the different encodings are meant to address the different tasks in the Task Analysis section above. Starting at the top, the first view is to plot different quantitative attributes side-by-side to easily compare them. These quantitative attributes can be chosen by the toggling system on the right, and there is also brushing and linking to select a region of planets in one graph and highlight them in the other (decided on highlighting because the same planets may be in different locations on the other graph).

On this same graph, one planet can be selected to see a very detailed pop-up about its information as summary statistics, correlation matrix for its attributes, a corresponding scatter plot for the most correlated relationship for the planet, and/or a histogram of a categorical feature can be analyzed. It would be nice to include an image of this planet as well.

The last encoding is to have a bar chart that contains information about all the planets that can be filtered by different attributes. This can be used to show how different planets are “grouped” together for different attributes and what similarities they share.

We decided on these encodings to have different marks (points in the scatter plots, and areas in the bar charts) and channels (colors to differentiate the charts, opacities for brushing and linking, and x and y positioning to show relationships) to convey the relationships between the attributes in the most effective way.

Appendix E: Digital Sketch



We envision a user interacting with our visualization in different ways depending on what their specific goal is. If the user intends to look at correlations between different variables, then they can use the scatter plots by selecting an x and y-axis from a drop-down selection representing a quantitative value and then the plot will update to display the two variables graphed against one another. Then, if the user wishes to visualize where planets lie on two different scatter plots, they can go through the same process of choosing an x and y-axis, creating a second scatter plot next to the first one. Using a brush tool, they can highlight a selection from either plot, highlighting the points in both plots. Additionally, if the user is curious about a given point, especially an outlier planet or one that has an interesting placement relative to the two scatter plots juxtaposed, then the user can click on that point and see the planet’s information pop up.

If the user wishes to instead view the distribution of one variable, they can use the histogram tool with filters to narrow their search to a given range for a given variable. For example, if the user wants to see a distribution of mass for planets with a mass of x where 10 <= x <= 50, then they could select [mass] with minimum value [10] and maximum value [50] and then view a histogram with bins of values between 10 and 50 that show how many planets within each bin fit that filtered range. Then, they can utilize the scatter plots to visualize mass in that range and view individual planets they’re interested in. These specific functions of the visualization allow for the user’s main needs to be fulfilled; they can view exoplanets with similar traits, visualize the distributions of traits, and look at specific planets that they find interesting with ease. Thus, all three tasks in the Task Analysis section will be accounted for.

Appendix F: Usability Testing

Preparation

Our visualization tool is one used to examine exoplanets and their different features. The tool is meant to help people learn about the different patterns and details of exoplanets while being easy and intuitive to use. Our project has two main graphs with different exoplanets represented as circles and they can be selected to provide additional detailed information for the given exoplanet(s). We also have descriptive graphs on the side that can be filtered to only present the data gathered from exoplanets that meet a certain condition dictated by the user's filtering parameters.

Task 1: Describe an exoplanet in as much detail as possible

a. We are testing this to make sure that people can easily see the detail from each exoplanet and that the visualization easily encourages retainment.
b. We will measure this by counting how many of the details that subjects are able to retain and how accurate they are. We will additionally ask for any qualitative observations as feedback.

Task 2: Which mass has the highest number of counts between 3 and 6 Jupiters.

a. This is being tested to determine both ease of use of the interaction as well as how intuitive it is to understand the information that is being displayed.
b. We will measure this by asking the users to filter the mass values between 3 and 6 Jupiters and then determining if they correctly identify which mass has the greatest frequency. We will also ask for any qualitative feedback.

Task 3: How many exoplanets have both a radius between 0.8 and 1.0 Jupiters and a mass between 4 and 5 Jupiters?

a. We are doing this to see how easy it is for users to interact with the scatterplots and analyze the relationship shown by the brushing and linking implementation.
b. We will measure this task by how long it takes people to correctly determine the number exoplanets in the given ranges. We will also ask for any qualitative feedback with this interaction.

Results

As a result of the usability test, it became clear that the visualization needs more context. The attributes of eccentricity, radius, and mass need units, and may not be apparent to everyone, so having a description for each would be helpful. We observed that the users found the brushing and linking to be intuitive, and work as expected. But overall, there was a lack of color, context, and cohesiveness, which we will continue to implement as we complete the visualization encoding.

Summary of Task 1: Describe an exoplanet in as much detail as possible

a. This task is pertained to the implementation of the tooltip, which would allow for easy viewing of select attributes for each planet. It does shed light on the fact that these attributes may not be commonly known, such as the meaning of eccentricity.
b. The result of the metrics of speed and accuracy presented to us that it was possible to correctly understand different attributes of exoplanets in a short amount of time.
    - The actual result of the test is: One exoplanet had a radius of around 0.55, eccentricity of around 0.125, and a mass of around 0.25 Jupiters.
c. Yes, these results indicate that is necessary to have more description and context for the different attributes and visual encodings.

Summary of Task 2: Which mass has the highest number of counts between 3 and 6 Jupiters.

a. Using the visualization to answer this question, results in the need for having units everywhere on the graphs, so the users understand more context with what each visualization is showing.
b. The metric of accuracy helped us determine if it would be easy to correctly determine the values from the second task.
c. Yes, I think for better user comprehension it would be better to include units and improve this aspect of the visualization.

Summary of Task 3: How many exoplanets have both a radius between 0.8 and 1.0 Jupiters and a mass between 4 and 5 Jupiters?

a. This task shows the positive impact of the tooltip in the visualization. But the users did indicate that the bar graphs are a little ambiguous and hard to understand, so we will work on adding context here.
b. The direct result of this task is that there is one planet within those ranges, and the accuracy metric of collecting these results is important to understand if users can brush and link across both scatterplots.
c. Yes, I think it would be helpful to improve the bar charts and add more context.

As a direct result of the usability tests, we will improve our visualization by adding more units and context in every way we can. We will also improve filtering with the bar charts and include a description of each of the visual encodings for additional user comprehension. Some additional changes we as a group want to implement including adding an interesting background to the webpage, adding a color scheme, as well as images of each of the exoplanets as a potential embellishment.

Appendix G: Reflections

Aidan: “I feel that I did particularly well in helping with giving the project direction in the earlier stages through things like the group charter, potential barriers, and strategies. I also feel that I did well in attending group meetings consistently and supplying sketches that helped to inspire aspects of our final design. If I could do the project again, I would put more effort into becoming more familiar with D3 so that I could provide greater support for the group in implementing the visualization. I also think it would have been interesting to try to find additional datasets which could be joined with our main dataset to provide more columns to analyze. Finally, I would make a goal of finding a dataset including the pictures of exoplanets so that tooltips could potentially include a picture of the exoplanet they describe. I was surprised to learn about the importance of responsiveness and interactivity in a successful visualization. I didn’t originally realize how important it is to have a dynamic view where users can change the dimensions they’re looking at or the range of values, but I realize now that it makes a massive difference.”

Ayushi: “I really enjoyed this project! This group is so friendly, so it was very easy to be productive and still have fun. I think we did a great job communicating whenever we could or could not meet, splitting up the work so that every aspect of the milestones would get done, and listening to each other’s ideas. I think I would want to do even more research given the opportunity to do this process again to determine other characteristics to measure trends between. Additionally, I think it would be better to try and find a researcher or a potential end-user that can use this visualization. Lastly, another improvement could be to have more in-depth preliminary sketches that are more creative since this may in turn impact the overall creativity of the design. The most valuable aspect about design I have learned is how intuitive the process is since it had previously seemed daunting and intimidating to do, but after completing this process, it feels much more manageable to do.”

Frank: “I learned a lot about the artistic visuals that can work well with code, and I think I learned (and applied) this in videomaking and considering how our project is viewed by an external user. Secondly, I thought I was good at the data collection and exploration, and it made it easier to find where we wanted to go with the project. Finally, I think I did well from a planning stage of making and helping with larger-scale plans about how we wanted our project to look. In the future, I'd like to delve deeper into making visualizations like the ones on our page in less of a template-based style. I'd also like to get to know the HTML a bit better personally to lay things out more clearly and dynamically. Thirdly, I would try to put more information in than what we have, though of course, some of this is with regards to time (like uploading pictures). The most valuable lesson I learned is how scalability works, in that it seems easy to make visualizations like the ones on our page from scratch for two points, but it could intuitively be hard for a lot. Seeing how we can use JavaScript to dynamically change our page and graphs is interesting to me.”

Nick: “I feel that I was able to decompose our domain tasks into code. I did this in three main ways including: comparing eccentricity versus mass and/or radius, histogram binning of data, and filtering bounds of a given plot. If I were to do this project again, I would like to add more customizability. I think it would be interesting to add coloring to convey size or other metrics. The most valuable takeaway from this project was how approachable large visualizations can be. The idea of taking data that interests me and transforming it into something that is approachable by all is no longer scary!”

Appendix H: Presentation Feedback

Project Name	Specific aspect of the visualization tool or presentation that you liked the most	Follow-up question that you have for the group
Basketball Shot Percentages	I liked how they broke down a shot percentage visual into what type of play benefits a given team. I thought this was an interesting way to analyze which players to trade and release.	Will you eventually include a basketball court visualization?
Wordle	I liked the dataset explanation, and it shows the initiative you guys took to do it yourself!	How would you adapt this project to show commonness compared to the wordle dictionary instead of the actual dictionary?
Bechdel Test	I appreciated the background information you guys provided. I think it made your visualization easier to understand!	Is there any way to incorporate more tests (besides Bechdel) to get a more in-depth score?
Baseball Opening Day Salary vs. Success	I like the comparison versus success because the salary cap is so important in other professional sports.	Could you guys implement a filter to show if a team has a few huge salaries versus a lot of medium salaries? In my mind, this is equivalent to a team having one or two superstar players with huge salaries versus a team that pays most of their players the same
Learn	I really liked that they talked with an end-user in the area! Shows a lot of initiative and how their tool can help make real improvements.	How do you keep the data up to date? Could you include a breakdown of what curriculum is required to be considered as “has cs?”
Plastic pollution	I liked how the visual encodings are directly based on the different domain tasks! Very structured visualization format.	Could you extend the coastal population vs. mismanaged waste visual to include color coordination with the country’s income?
Basketball Salaries	I really like how in-depth the iterative process is shown for the design sketches. Also, there is good background information, and I like the data collection process!	Could you color coordinate each data point to show a heatmap? It would be interesting to see multiple plots, one per position, with a heatmap coloring.
Visualizing Congress Stock Trading	I really like how the visualization has a lot of contexts such as colors for the parties and representation of the region the congressperson represents. Makes it easier for user-understanding and adds additional features to compare.	Could you add a view/ toggleable filter to show which political party is making trades (as opposed to a top congress member's traders)?
Covid-19 Dashboard	I like the design process and the filtering interactivity shown to make it easier to zoom in on the data	Do you need to anonymize the data? Is there a potential HIPAA violation? Could you work with Northeastern to use the anonymous testing data they have collected?
Expected Performance in the EPL	I really like the interactive features on the charts as well as the logos in the scatterplot! Everything looks very polished and neat	If you could add other soccer leagues to the visualization, which would you choose and why?

Project Name	Specific aspect of the visualization tool or presentation that you liked the most	Follow-up question that you have for the group
Books	I liked how you guys mentioned that there was work behind the scenes that didn't make it into the final visualization	Out of the 87 "good books" you were able to use, how do you plan to expand on that dataset?
MBTA	I liked the geolocation aspect for brushing and linking, not many groups did that, and I really liked it!	Could you eventually add a map to the geolocation visual?
Recipes	I really liked that you could filter based on diet requirements!	Could you add an aspect where users can input the name of recipes? Could this become a repository for many different recipes for each type of dish?
Where should I live?	I think the heatmap is important because it opens the door to looking continentally for a place to live (as opposed to locally)	Would it be possible to implement a way to filter cities based on specific criteria that users want? What features does the overall quality of a city represent?
Crypto	I really like the approach of teaching instead of visualizing. I think this project's visualization promotes financial literacy and deeper learning of crypto!	Would you want to apply this visualization to a human-subject experiment to determine if it does help users learn better?
Citi Bike Optimization	I liked how you didn't focus on a Boston-centric dataset. It was also impressive to hear how your group planned on leveraging the visualization for business development was very interesting.	Would you consider including specific routes from different start and end stations, and ways this could be optimized?
Co-op'd	It seems like this was an extended project that was adapted for this class (which I really like). I appreciate the fact that this is a highly iterative project for you guys.	Do you think this could be extended to employers as well to have better insights about previous co-op experiences at their company?
Affordability of diets around the world	I think the topic choice is very important! Choosing a topic that actively needs a larger public understanding is smart and makes your visualization invaluable.	How do vegan or vegetarian diets skew the trends because vegan or plant-based diets tend to be more expensive?
Food Dessert Visualization	I enjoyed the level of specificity offered in your visualization. I think having food access leveraged against racial makeup creates a space for further visualizations and explanation.	What other demographics could you measure in addition to race for food desserts?
Bar hop	This is a fun topic! I liked how you broke down the task of "where to go" into easy-to-understand visuals.	What additional features could you implement in the future? Maybe pricing or cover fees?