White Paper - NASA Exoplanet Archive Data Visualization

Final Project - Data 73000 - Introduction to Data Visualization

Betsy Rosalen

2023-05-22

Contents

The Questions

I was interested in visualizing the NASA Exoplanet Archive data because I have always had an intense interest in space exploration and astronomy. I wanted to know how many exoplanets have been found so far and what are the different ways we can classify them. For example, how many are terrestrial vs gaseous planets? How many are potentially habitable vs. not habitable? How many stars and planets do typical planetary systems have? As I explored the data other questions came to mind like, are there differences in the types of planets found by different discovery methods? What are those discovery methods and how do they work? Which observatories have found the most exoplanets? How many are in a circumbinary orbit (around two or more stars)? How do we determine whether a planet is potentially habitable or not? How far are the nearest and furthest exoplanets discovered so far in light years? What are 'average' or 'typical' exoplanets like? Or, in other words, what are the distributions of different types of exoplanets like? I tried to answer most, if not all, of these questions in the final visualization.

The Data

NASA Exoplanet Archive | 5,347 observations as of May 8th and growing...

I downloaded the complete Planetary Systems Composite Parameters table[1] as a CSV from the NASA Exoplanet Archive at exoplanetarchive.ipac.caltech.edu[2] as well as a data dictionary[3]. I thought that habitability would be one of the variables in the data set or at least that it would be easily calculable from the data, but after much searching, I realized that the calculations are much more complex than I anticipated and not as clearly defined as I anticipated either. Different calculation methods are debated among planetary scientists. So I found another reputable data set called The Habitable Exoplanets Catalog[4], which is a project of The Planetary Habitability Laboratory (PHL) at the University of Puerto Rico at Arecibo, that I was able to merge with the Exoplanets Archive. A limitation of using this data was that since the Catalog was last updated in January, it was missing exoplanets discovered in the months between January 2023 and May 2023 when I downloaded the Archive data.

Variables | 313 Variables Narrowed Down to 27

I narrowed the data down to 27 variables (out of 313 total) that I would potentially use in my project. Most, but not all, made it into the finished visualization. The remaining ones I hope to add to future iterations of the project.

Table 1: Selected Variables from the NASA Exoplanets Archive

Variable	Definition	Type	Description
planet_name	Planet Name	character	Planet name most commonly used in the literature
host_name	Host Name	character	Stellar name most commonly used in the literature
planet_letter	Planet Letter	character	Letter assigned to the planetary component of a planetary system
n_stars	Number of Stars	numeric	Number of stars in the planetary system
n_planets	Number of Planets	numeric	Number of planets in the planetary system
orbits_binary_system	Circumbinary Flag	numeric	Flag indicating whether the planet orbits a binary system (1=yes, 0=no)
discovery_method	Discovery Method	factor	Method by which the planet was first identified
discovery_year	Discovery Year	numeric	
disc_facility	Discovery Facility	factor	Name of facility of planet discovery observations
orbital_period	Orbital Period [days]	numeric	Time the planet takes to make a complete orbit around the host star or system
planet_Earth_radius	Planet Radius [Earth Radius]	numeric	Length of a line segment from the center of the planet to its surface, measured in units of radius of the Earth
planet_Jupiter_radius	Planet Radius [Jupiter Radius]	numeric	Length of a line segment from the center of the planet to its surface, measured in units of radius of Jupiter
planet_Earth_mass	Planet Mass or Mass*sin(i) [Earth Mass]	numeric	Amount of matter contained in the planet, measured in units of masses of the Earth
planet_Jupiter_mass	Planet Mass or Mass*sin(i) [Jupiter Mass]	numeric	Amount of matter contained in the planet, measured in units of masses of Jupiter
planet_Earth_density	Planet Density [g/cm**3]	numeric	Amount of mass per unit of volume of the planet
eccentricity	Eccentricity	numeric	Amount by which the orbit of the planet deviates from a perfect circle
ratio_planet_to_star_radius	Ratio of Planet to Stellar Radius	numeric	The planet radius divided by the stellar radius
star_spectral_type	Spectral Type	factor	Classification of the star based on their spectral characteristics following the Morgan-Keenan system
star_effective_temp	Stellar Effective Temperature [K]	numeric	Temperature of the star as modeled by a black body emitting the same total amount of electromagnetic radiation
star_radius_solar	Stellar Radius [Solar Radius]	numeric	Length of a line segment from the center of the star to its surface, measured in units of radius of the Sun
star_mass_solar	Stellar Mass [Solar mass]	numeric	Amount of matter contained in the star, measured in units of masses of the Sun
star_luminosity_log_solar	Stellar Luminosity [log(Solar)]	numeric	
star_age	Stellar Age [Gyr]	numeric	The age of the host star
dist_to_system_pc	Distance [pc]	numeric	Distance to the planetary system in units of parsecs
dist_to_system_ly	Distance [ly]	numeric	Distance to the planetary system in units of light years
star_class	Spectral Class	factor	Classification of the star based on their spectral characteristics

[1] "Planetary Systems Composite Data" (n.d.)
[2] "NASA Exoplanet Archive" (n.d.)
[3] "Planetary Systems and Planetary Systems Composite Parameters Data Column Definitions" (n.d.)
[4] "PHL @ UPR Arecibo - the Habitable Exoplanets Catalog" (n.d.)

Data Cleaning

Not much data cleaning or manipulation was necessary since the archive is a very well-structured and clean data set to begin with, but some manipulation was done in R just to: filter out the variables I would not be using, rename some variables to make them more easily recognizable names, convert some variables to factors, and convert the distance in parsecs to light years. See Appendix A

The Visualization

Design Concept

The design concept for the visualization was to use circles on a dark, starry, night-sky background as much as possible to suggest planets in space. I wanted the whole piece to evoke the *feeling* of being in space and of the vastness of space. Originally I hoped to make all the circles custom shapes that looked like planets, but due to limitations in Tableau, the bubble chart shapes could not be customized, so I used solid circles for those instead. Where I could use custom shapes I used pictures of imaginary planets that I downloaded from the internet then edited in Adobe Illustrator by using a circular clipping mask to make the background transparent, and adjusting the color balance to make them brighter and slightly less realistic. Feedback from an early draft of the project made it clear that if the planet shapes were too realistic they were interpreted as real planets rather then as symbolic shapes that represent relative values.

Figure 1: Imaginary Planets as Custom Shapes in Tableau

Fonts and Colors

I used fonts and colors that are also suggestive of outer space, early computer screens, and sci-fi. For most of the text I used Courier New[5], a monospaced, slab serif font, that was common in early computer systems and often used for coding[6] in a shade of green that is suggestive of 80's sci-fi[7], little green men[8], and the Matrix digital rain[9]. Some titles and additional supporting text used Tableau fonts in a medium gray to decrease the contrast against the mostly black background.

[5]Ali et al. (2022)
[6]"Courier (Typeface)" (2023)
[7]Schneider (n.d.)
[8]"Little Green Men" (2023)
[9]"Matrix Digital Rain" (2023), "Matrix Digital Rain" (n.d.)

Image 1: 80's Sci-fi Image 2: Little green men Image 3: Matrix digital rain

The resulting data dashboards were intended to evoke the image of planets in space while simultaneously being somewhat reminiscent of the data dashboards of sci-fi movie and TV spaceships.

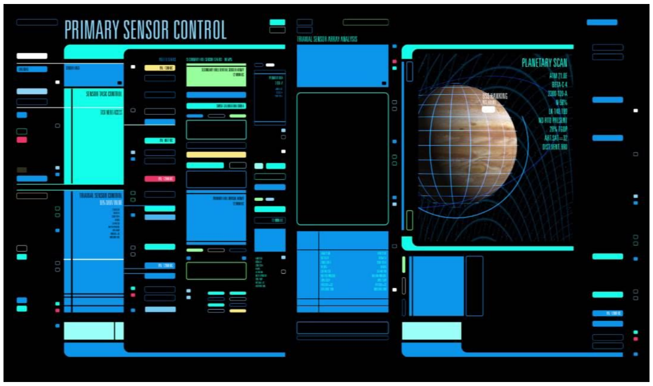

Image 4: Sci-fi Data Dashboard

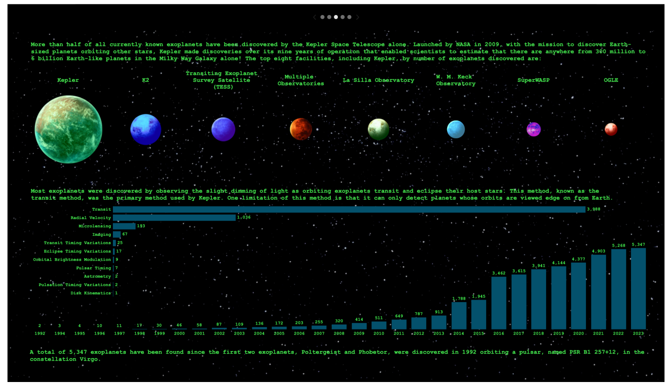

Image 5: Exoplanets Screenshot

The Hook

Feedback on an early draft of the project lead me to add much more explanatory text than I had originally intended. It was clear, albeit surprising to me, that most people didn't know what exoplanets were let alone why we should care about them. As a result of this feedback, I added two introductory pages to the Tableau story, to hook the viewer in and explain what exoplanets are and why they matter. These pages ask fundamental human questions that everyone can relate to, so while I believe they succeed in making the visual aesthetic of the final project much clearer and stronger, they also make it much more engaging.

Both of the introductory pages were intended to evoke the *feeling* of being in the vast emptiness of outer space. I chose to center on the first page a well-known quote by Carl Sagan[10] in large bold text that addresses not only how vast the universe is, but how likely it is that life exists elsewhere in the universe beyond Earth. I chose to put nothing else on this page in order to highlight the message and hook the reader in to the project in the hope that it would make them want to read more. The second page elaborates more on the possibility of extraterrestrial life, by asking the question, "Are we alone?" which is also visually represented by an image of the Earth floating alone in the center of the screen on a dark star-speckled background. The green text explains why the search for exoplanets is essential research if we want to answer this question and invites the reader to continue to the next page to learn more about exoplanets. These two pages are meant to prime the reader for the following three dashboard pages.

Image 6: Page 1 Carl Sagan Quote

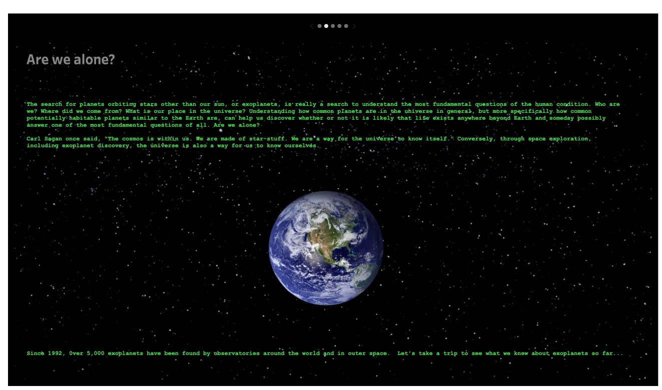

Image 7: Page 2 "Are we Alone?"

[10] "A Quote from Contact" (n.d.)

Interactivity and Tooltips

I wanted the dashboards to have some interactive features to allow the viewer to engage with them as if they really were spaceship control boards to reinforce the sci-fi connection that was the inspiration for much of the design. I wanted users to be able to learn from them the same way sci-fi characters gain new information by interacting with their computer terminals. All three of the dashboards after the first two introductory pages have tooltips that give more in-depth information when you hover over parts of the visualizations. These tooltips include: explanations for the different discovery methods, details about all 63 of the exoplanets which are potentially habitable, and information about how the physical characteristics of exoplanets influence whether or not it is believed to be habitable, as well as other details. Significant effort was made to research the information in the tooltips to ensure accuracy, and sources are sometimes cited within the tooltip. Finally, some filters (called "Controls") were added to the last page to allow the users to select different parameters to dynamically change the visualization and compare different characteristics of exoplanets to the Earth in order to gain a better understanding of what makes a planet habitable.

Image 8: Potentially Habitable Exoplanet Tooltip

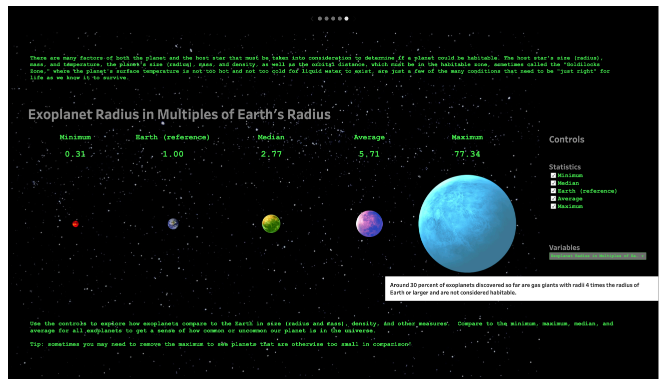

Image 9: Habitability Tooltip and Dashboard with Controls

References

"A Quote from Contact." n.d. Accessed May 22, 2023. https://www.goodreads.com/quotes/282727-the-universe-is-a-pretty-big-place-if-it-s-just.

Ali, Basit, Lesli Schauf, Peter Constable, Nicholas Hart, Mike Jacobs, and v-nihar. 2022. "Courier New Font Family - Typography." March 30, 2022. https://learn.microsoft.com/en-us/typography/font-list/courier-new.

"Courier (Typeface)." 2023. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Courier_(typeface)&oldid=1150744603.

"Little Green Men." 2023. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Little_green_men&oldid=1152269129.

"Matrix Digital Rain." 2023. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Matrix_digital_rain&oldid=1142485270.

"Matrix Digital Rain." n.d. Accessed May 21, 2023. https://rezmason.github.io/matrix/.

"NASA Exoplanet Archive." n.d. Accessed May 8, 2023. https://exoplanetarchive.ipac.caltech.edu/index.html.

"PHL @ UPR Arecibo - the Habitable Exoplanets Catalog." n.d. Accessed May 13, 2023. https://phl.upr.edu/projects/habitable-exoplanets-catalog.

"Planetary Systems and Planetary Systems Composite Parameters Data Column Definitions." n.d. Accessed May 8, 2023. https://exoplanetarchive.ipac.caltech.edu/docs/API_PS_columns.html.

"Planetary Systems Composite Data." n.d. Accessed May 8, 2023. https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=PSCompPars.

Schneider, Bernd. n.d. "Ex Astris Scientia - Galleries - Earth and Federation Interfaces and Script." Accessed May 22, 2023. https://www.ex-astris-scientia.org/gallery/displays1.htm#phoenix.

Appendix A - R Code

```
col_names <- read_csv("PSCompPars_2023.05.08_06.44.02.csv", skip = 2, n_max = 312)
exo_raw <- read_csv("PSCompPars_2023.05.08_06.44.02.csv", skip = 316)
habitable_lu <- read_excel  ("Habitable_Planets.xlsx")

exo_raw <- exo_raw %>%
  mutate(across(c(discoverymethod, disc_locale, disc_facility, disc_telescope,
                 disc_instrument, st_spectype), factor),
         dist_to_system_ly = sy_dist * 3.262,
         star_class = factor(substr(st_spectype, 1, 1))) %>%
  rename(c('planet_name' = pl_name,
           'planet_letter' = pl_letter,
           'host_name' = hostname,
           'discovery_method' = discoverymethod,
           'discovery_year' = disc_year,
           'n_stars' = sy_snum,
           'n_planets' = sy_pnum,
           'orbits_binary_system' = cb_flag,
           'orbital_period' = pl_orbper,
           'planet_Earth_radius' = pl_rade,
           'planet_Earth_mass' = pl_bmasse,
           'planet_Earth_density' = pl_dens,
           'planet_Jupiter_radius' = pl_radj,
           'planet_Jupiter_mass' = pl_bmassj,
           'eccentricity' = pl_orbeccen,
           'ratio_planet_to_star_radius'  = pl_ratror,
           'star_spectral_type' = st_spectype,
           'star_luminosity_log_solar' = st_lum,
           'star_age' = st_age,
           'star_effective_temp' = st_teff,
           'star_radius_solar' = st_rad,
           'star_mass_solar' = st_mass,
           'dist_to_system_pc' = sy_dist
           ))

exo <- exo_raw %>%
  left_join(habitable_lu, by = c('planet_name' = 'Planet Name Cleaned')) %>%
  select(planet_name, habitable, planet_letter, host_name, orbits_binary_system,
         discovery_method, discovery_year, disc_facility,
         n_stars, n_planets,
         orbital_period,
         planet_Earth_radius, planet_Earth_mass, planet_Earth_density,
         planet_Jupiter_radius, planet_Jupiter_mass, eccentricity,
         ratio_planet_to_star_radius, star_spectral_type, star_class,
         star_luminosity_log_solar, star_age, star_effective_temp,
         star_radius_solar, star_mass_solar,
         dist_to_system_pc, dist_to_system_ly
         )
```