DATASCI200 Section 6 Project 2 – Exploratory Data Analysis

Title: Our Planet in Perspective: Exploring Exoplanet Discoveries

Audience: Board of Directors of local Science Museum

Team Members: Cedar Frost, Sammy Lee, Spencer Morris

GitHub Repository: https://github.com/UC-Berkeley-I-School/Project2_Frost_Lee_Morris

Contents

Motivation:

Since antiquity, humankind has been fascinated by the nighttime sky, and while the ancients clearly identified the 5 'wandering stars' (Mercury, Venus, Mars, Jupiter, and Saturn) as distinct from other stars, it was not until the 17th century that they were understood to be objects similar to Earth organized in a heliocentric configuration. Recent advances in astronomy have enabled scientists to begin characterizing planets around other stars (so-called 'exoplanets'), once again revolutionizing the context available for understanding our own planet. We envision expanding the Museum with a new interactive exhibit to showcase the wide variety of discovered exoplanets and to underscore just how singular the life-sustaining Earth is currently known to be. We think this exhibit will be of great interest to the public, who have enabled the funding to realize these data and who are abiding an increasing role in safeguarding the planet for future life. Here we present an initial exploration of the data as a prelude to an exciting exhibit that lies at the intersection of astronomy, geology, and climatology.

Background:

Here we review the concepts essential for appreciating information related to exoplanets.

Planets are compact objects that move in roughly elliptical orbits around at least one host star. We know from our own solar system that they can be extremely different from one another. (Recall that Mercury is a dense rocky world only slightly larger than the Moon that tightly orbits Sol (the Sun), whereas Neptune is a giant ball of gas and dust so distant from Sol that it was only discovered in 1846). Clearly, planets have a wide variety of properties, but the most important ones are mass (a measure of how much matter they contain), radius (which describes their geometric size), and period (a measure of their year, i.e. the time it takes them to make a single revolution around their host star).

The stars that planets orbit also come in a staggering diversity of forms which can differ greatly from Sol. And like planets, stars, too, have key properties that astronomers have been characterizing for 200 years, including mass, luminosity (or how 'bright' they would appear at a fixed distance), effective temperature (since stars can be much cooler or hotter than Sol, and metallicity (which is a measure of the abundance of heavy elements that comprise the star). The distances between stars are vastly greater than the distances within the solar system, and many attempts are made to determine how far away stars are from us. Finally, the age of stars is important; our understanding of fusion and cosmological development indicates that planets should only be possible around fairly young stars, since the heavy elements that form them were not abundant in the early universe.

Planets are tiny and dark relative to the stars they orbit, so observing them directly is difficult. When we detect exoplanets, most of the time we do so indirectly by observing stars and seeing how those stars change over time. Many techniques are available to characterize planets in this way. 'Radial velocity' measurements infer their mass by detecting periodic color changes in a star. "Transit photometry" measurements infer their radius by measuring periodic brightness changes in a star. Other techniques are available as well, and importantly, they all have limitations – they must therefore be combined to fully characterize the key properties of interest (eg. both mass and radius). 'Discoveries' are awarded for initial candidate detections, which are then validated using other techniques.

Disclaimer:

We caution that the available data (and therefore our analysis) is not representative of all planets, but rather the planets that have been detected thus far, which are inherently biased based on the available observational technologies. We expect, for example, that the population of small planets vastly exceeds large ones, but the available data are heavily skewed toward larger objects.

Thematic Questions:

- How have planet discoveries changed over time?
- How do exoplanet properties vary by discovery technique?
- What exoplanets discovered thus far are the most Earth-like?
- How do exoplanetary systems compare to the Sol system?

Dataset:

The primary dataset of interest is called the Planetary Systems Composite Parameters Planet Data, hosted by the NASA Exoplanet Archive. Each row corresponds to a single confirmed exoplanet (i.e. one that has been discovered and validated). The full dataset contains 84 columns, each corresponding to a property related to the discovery, star, planet, or metadata such as measurement error or attribution. The compiling organization has combined reported properties of a single exoplanet from various studies into a single curated dataset: where discrepancies of a measured property exist, the organization has an established process to choose a most appropriate measurement (and flag records as controversial in the case of unresolved disagreements). The dataset table was accessed on 4 April 2023 from here.

We restrict our analysis to the key astrophysical properties related to the exoplanet and its host star, as well as some critical discovery metadata. The following fields were defined as of primary interest:

Discovery Data

- **discoverymethod**: Observational technique by which planet was discovered
- **disc_year:** Year the planet was discovered
- **pl_controv_flag:** Flag indicating if the planet discovery is controversial

Planet Data

- **pl_name:** Planet's unique identifier
- **pl_orbper:** Orbital period [unit: Earth days]
- **pl_rade:** Radius [unit: relative to Earth's radius]
- **pl_bmasse:** Mass (minimum; best estimate) [unit: relative to Earth's mass]
- **pl_dens:** Density [unit: g/cm^3]
- **pl_nespec:** # of emission spectroscopy measurements [unit: count]
- **pl_ntranspec:** # of transit spectroscopy measurements [unit: count]

Star Data

- **hostname**: Star's unique identifier
- **sy_dist:** Distance [unit: parsecs]
- **st_mass:** Mass [unit: relative to Sun's mass]
- **st_spectype:** Spectral Type [unitless; Morgan-Keenann classification]
- **st_lum**: Luminosity [unit: relative to Sun's luminosity $\log_{10}$(Solar)]
- **st_teff:** Temperature [unit: Kelvin]
- **st_met:** Metallicity (proxy for system composition and age) [unit: dex]
- **st_age:** Age [unit: Gya]

Data Transformations:

Here we review the key transformations applied to clean and enrich the primary dataset.

We first confirmed that there was only one row per planet by checking for duplicated rows, then dropped all but the key columns. Given that mass and radius are the essential characteristics of an exoplanet, all records were dropped where either of these variables was unavailable (reducing the number of records from an initial 5322 to 5281). Then, missing values for planetary density were filled based on the available data.

An initial evaluation of the descriptive statistics for key astrophysical properties showed that several records contained extreme outliers, which may arise from large measurement errors, or otherwise do not contribute to our study's main themes. Therefore, to restrict our analysis to the highest-quality available data, we identified and excluded outliers using the following approaches:

- All records flagged as 'controversial' were dropped, reducing the dataset length to 5258.
- The median orbital period was of order 10e1, but the mean and standard deviation were of order 10e4 and 10e6, respectively, with a maximum value of order 10e8. Knowing that astronomers have been observing exoplanets for only a few decades, and given our understanding that robust discoveries rely on periodic (i.e. multiple repeating) indications of orbital periods, we think keeping long-period planets we chose to exclude objects with periods greater than 200 years (which is already well beyond the period of the most-distance planet in the Sol system). The number of remaining records was reduced to 5251.
- A descriptive analysis of the planet masses We excluded objects with masses more than 3200 times the Earth, as such objects are, by definition, brown dwarfs, and therefore considered as distinct objects more like small stars than planets. We also removed one

object with a much larger radius than the others. Following these drops, the number of remaining planets was 5084.

- A descriptive analysis of planet densities (in units of g/cm^3) yielded a mean of 3.8, a median of 2.5, and a maximum of 1290. We find densities greater than lead to be impossible, given the current understanding of elemental abundances and planetary formation, and therefore excluded such objects, reducing the number of records to 4940.
- We excluded objects whose host stars were recorded as older than 13.8 Gy (the age of the universe), reducing the number of records to 4879.
- Planets orbiting stars with temperatures greater than 8000K were dropped, since such stars are normally short-lived hot giants, reducing the number of records to 4860.
- During data exploration, we identified and added a few additional properties to enrich the analysis.
- The original dataset provided the distance of each exoplanet's host star in units of parsecs, which is the standard unit used within the astronomical community for astronomical distances. The public, however, is more likely to be familiar with light-years (ly) as a unit, and so a new column was added using the standard conversion between parsecs and ly.
- A new column was added to capture the mean distance between an exoplanet and its host star, in Astronomical Units (AU). These values were estimated using Kepler's Law and assuming star-centered circular orbits.

<u>Observational techniques over time</u>

Given the extraordinary technical hurdles required to detect planets around distant stars, we are interested in how exoplanet discoveries can inform us of trends in observational technologies. Which methods were developed first, and are they still widely used? What are the newest methods? One might assume that ground-based methods are pre-dominant, since they cut the financial and technical constraints considerably, but is that reflected in the data? Here we explore how exoplanet discovery counts have varied in time and across the various techniques.

4860 exoplanets are included in our dataset, and we are interested in the discovery that kicked off this body of knowledge. The first reported discovery was made in 1992. That year includes two exoplanet discoveries (figure 1), both around the same star and detected using the pulsar timing technique. This discovery is clearly significant since it was the first time that planets were confirmed outside the Sol system. Since the host is not a normal star but rather a pulsar (which is a remnant of a dead star), this is a strong indication that planets are widespread, even around highly unusual stellar objects.

disc_year	hostname	pl_name	discoverymethod
1992	PSR B1257+12	PSR B1257+12 c	Pulsar Timing
1992	PSR B1257+12	PSR B1257+12 d	Pulsar Timing

Figure 1. First discovered planets

Taking a look at cumulative discovery counts by year since this initial discovery, we see that additional discoveries accelerated until around 2014, where we see a brief explosion in the mid-2010's, followed by a return to the previous exponential trend.

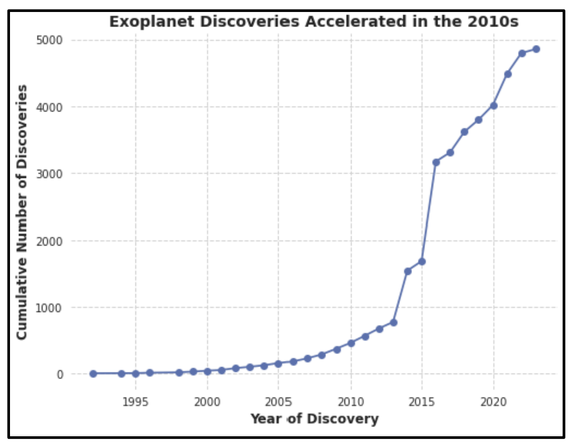

Figure 2. Discoveries over time

By grouping discoveries into 5-year increments, we clearly see that the bulk of discoveries were made during the 2010's.

But which techniques are responsible for these discoveries? While the initial discovery was based on pulsar timing, that seems to be an unlikely origin for the ~5000 exoplanets detected thus far, given how rare pulsars are. A look at the total number of discoveries by technique shows that, of the 11 detection methods, 3 of them (transit, radial velocity, and microlensing) account for over 99% of all exoplanets in the dataset.

Exoplanet Discoveries Accelerated in the 2010s

Period	Count	Cumulative	% of Total
1990-1994	4.0	4.0	0.1
1995-1999	36.0	40.0	0.7
2000-2004	115.0	155.0	2.4
2005-2009	303.0	458.0	6.2
2010-2014	1230.0	1688.0	25.3
2015-2019	2329.0	4017.0	47.9
2020-2025	843.0	4860.0	17.3

Figure 3. Discovery counts at half-decade increments

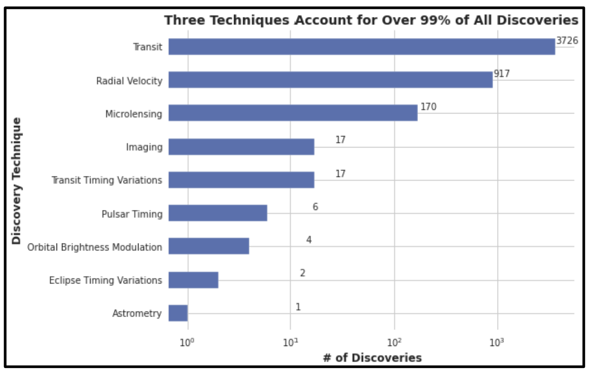

Figure 4. Discovery counts by technique

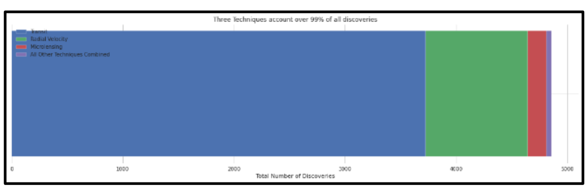

Figure 5. Breakdown of discoveries by technique

If we look again at the total discovery counts by year and group by discovery technique, we observe that the transit method accounts for the highly productive period in the mid-2010's. These discoveries appear to be related to the Kepler mission, which was a space-based observatory that operated for a few years before its mission ended.

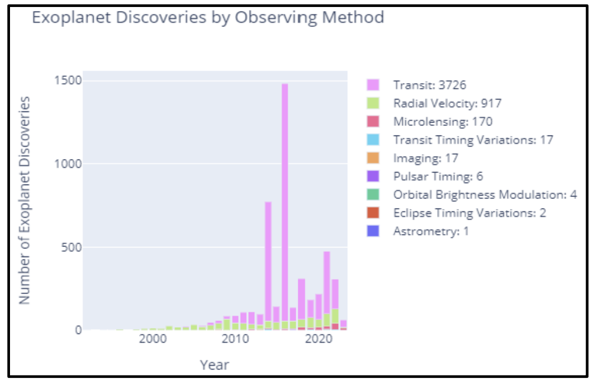

Figure 6. Discoveries by technique over time

An evaluation of discoveries attributable to the top-3 techniques shows that the radial velocity method has been a strong contributor since the 1990's and accounts for much of the underlying growth rate in discoveries. Microlensing appears to have been proven in the 2000's, and each decade shows a large increase in discoveries. Transit discoveries are again shown to comprise the bulk of discoveries. And the various 7 other techniques contribute only a handful of planets each – clearly these are demonstration methods that may be of scientific interest but are not especially currently practical at detecting large numbers of exoplanets.

Exoplanet Discoveries by Observing Method

Discovery Method	1990's	2000's	2010's	2020's
Microlensing	0	10	73	87
Others	3	5	27	12
Radial Velocity	24	268	414	211
Transit	0	60	2915	751

Figure 7. Decadal summary of the main techniques

Exoplanet Characteristics

A planet's period provides critical insight into the gravitational relationship with its host star (a short period, for example, is indicative of a tightly-bound orbit). The 8 planets of our own solar system have orbital periods that range from 88-68000 days (all times in equivalent Earth days).

An evaluation detected exoplanets shows that they have orbital periods of wide ranges, up to and exceeding 68000 days. The histogram below shows the frequency of objects with orbital periods up to 150 days, and clearly demonstrates that most detected exoplanets have rather short periods of tens of days.

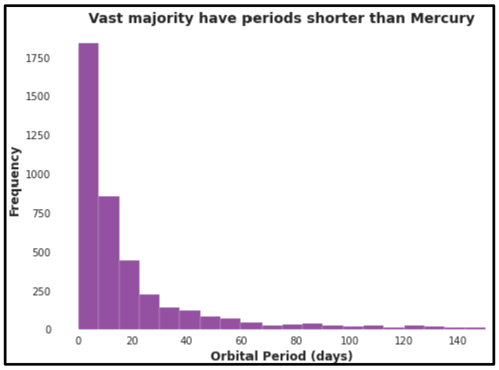

Figure 8. Orbital period frequencies

Using Kepler's Law to estimate the orbital radii for exoplanets, we propose the following categorization for describing exoplanet distances from their hosts: 'tight' (0-0.5 AU, which would encompass Mercury in the Sol system), 'inner' (0.5-1.55 AU, encompassing Venus, Earth, and Mars), 'middle' (1.55-10 AU, encompassing Saturn and Jupiter), and 'outer' (encompassing Uranus and Neptune). Based on this convention, we can see that fewer than 400 of the detected exoplanets exist at ranges comparable to the inner Sol system.

Tight	4023
Inner	334
Middle	294
Outer	28

Figure 9. Discovery counts by orbital radius category

We consider how orbital period measurements have varied over time and technique. Below we see a step-change in mean orbital period beginning with detections in the 2010's, dropping from ~hundreds of days down to ~tens of days. An examination of the period distributions by technique clearly shows that the transit method, which we know accounts for a disproportionate proportion of overall detections since the 2010's, is particularly sensitive to this regime. This makes sense, given that we know transit detections involve a periodic change in star brightness (hence requiring multiple periods to account for an observation), and that the Kepler mission lasted only a few years. We find that transit discoveries do not span the full range of periods detected by other techniques such as radial velocity, which further reinforces the transit method's bias to low-period objects. Based on these data, the microlensing technique appears particularly suitable for detecting long-period exoplanets.

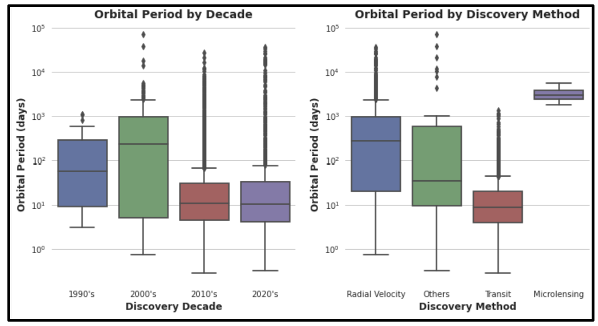

Figure 10. Period by decade and technique.

We can clearly see that discovery methods grouped as 'Others' outside the top-3 represent the greatest range of orbital periods, which is not surprising as these are likely representing champion demonstrations.

We next consider the radii of detected exoplanets. By investigating the distribution of radii, we see that two planet sizes are evident, one that is between the size of Earth and Neptune, and another that is larger than Jupiter.

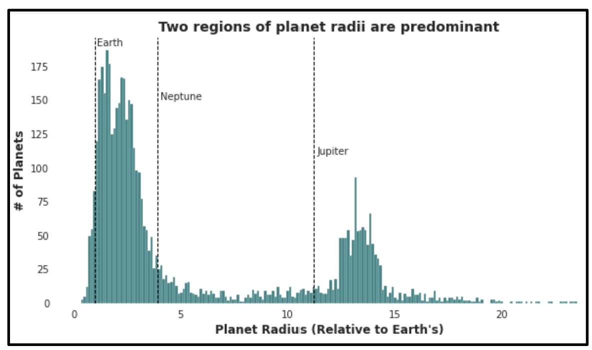

Figure 11. Radius frequencies.

This is quite surprising, considering there are no planets of these sizes within our own solar system. An evaluation of planet size by observational technique is therefore warranted, and we see that the transit method is skewed toward detecting smaller planets, whereas the radial velocity method is biased toward larger planets. Interestingly, while the bimodal distribution may be a consequence of measurement bias (since there the astrophysical models do not account for it), we do find it noteworthy that all 3 of the major detection methods demonstrate a low-abundance region between 5 and 10 times the Earth's radius.

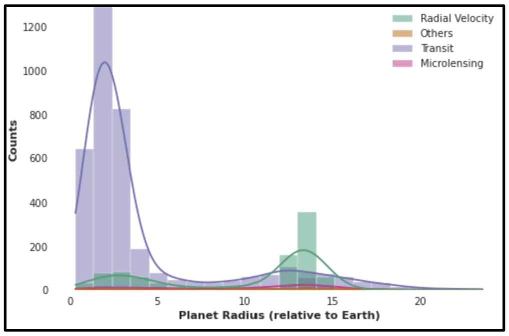

Figure 12. Radius frequencies by technique

Based on the frequency distribution of planet radii, and in analogy to our own solar system, we propose the following categorization for describing exoplanet radii: ‘tiny’ (0-0.5 times the Earth’s radius, which would include Mercury), ‘small’ (0.5-2 times Earth’s radius, which includes Venus, Earth, and Mars), ‘medium’ (2-6 times the Earth’s radius, which includes Uranus and Neptune), and ‘large’ (more than 6 times Earth’s radius, which includes Saturn and Jupiter). Based on this convention, nearly 1500 small planets comparable to Earth have been discovered.

Tiny	4
Small	1476
Medium	1912
Large	1468

Figure 13. Discovery counts by planet size category

A closer look at how planet radii vary across time and technique shows that the 2010’s, because of the abundance of transit measurements, saw a significant reduction in mean values for this property. Interestingly, the transit method demonstrated the ability to make detections across the entire range of planet radii, with good coverage of exoplanet sizes.

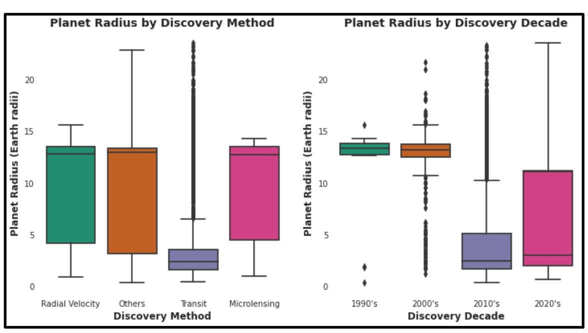

Figure 14. Radius by decade and technique.

Now we turn our focus to the masses of detected exoplanetary bodies. Masses as great as 3000 times the mass of the Earth are included in the dataset, but the vast majority of exoplanets have masses less than 10 times the Earth’s mass. The following histograms demonstrate what appears to be an exponential decay in the frequency of planets as mass increases, although upon closer inspection the number of detected planets diminishes at less than 3 times the Earth’s mass. We can conclude that planets with masses similar to Earth are detectable using current technologies, but barely.

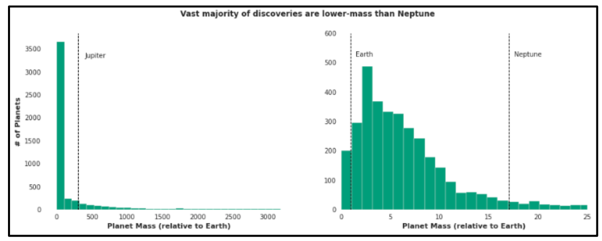

Figure 15. Mass frequencies.

We adopt a convention of categorizing planets by mass, with 6 categories in total: ‘sub-terrestrial’ (0-0.75 times the Earth’s mass, which includes Mercury and Mars), ‘terrestrial’ (0.75-2 times Earth’s mass, which includes Earth and Venus), ‘super-terrestrial’ (2-10 times Earth’s mass), ‘sub-giant’ (10-80 times Earth’s mass (which includes Uranus and Neptune), ‘giant’ (80-400 times Earth’s mass, which includes Saturn and Jupiter), and ‘super-giant’ (more than 400 times Earth’s mass. Based on this convention, the plurality of detected exoplanets falls in the ‘super-terrestrial’ category. These planets are often referred to as ‘Super Earths’ in the literature and have no analogue in our own solar system.

	count	mean	std	min	25%	50%	75%	max
Sub-Terrestrial	138	0.455338	0.187113	0.02	0.327	0.443	0.615	0.749
Terrestrial	304	1.42904	0.357701	0.76	1.1175	1.42	1.76	1.98
Super-Terrestrial	2334	5.29243	2.21116	2	3.3075	5.1	7.08	9.96
Sub-Giant	825	25.2495	18.2121	10	12.4	16.8	31.1	79.4575
Giant	579	218.682	85.674	80	145.1	210	282.868	397.288
Super-Giant	680	1179.78	756.475	400.4	594.263	872.5	1633.76	3178.3

Figure 16. Mass statistics by mass category.

Furthermore, the counts of exoplanets across mass and orbital radius categories, we find that all six mass categories are well-represented in tight orbits, but only very massive planets are seen far from their hosts. This is perhaps not surprising, since the star-planet interactions will necessarily make only the most massive planets detectable in the outer regions.

	count	unique	top	freq
Tight	4023	6	Super-Terrestrial	2258
Inner	334	4	Super-Giant	144
Middle	294	4	Super-Giant	207
Outer	28	2	Super-Giant	27

Figure 17. Mas category counts by orbital radius

A comparison of mass statistics over time and measurement technique again shows that the 2010's saw an appreciable decline in detected masses, which is largely attributable to transit discoveries. We observe that the lowest-mass planets were detected in the 1990's, which were the first planets discovered using the pulsar timing technique.

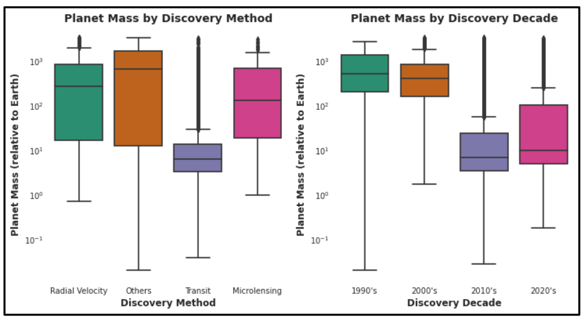

Figure 18. Mass by decade and technique.

Planetary density is believed to be an important contributor to the likelihood of the development of life, since high densities imply the presence of heavy metals which increase the probability of catalyzing complex organic reactions. A density histogram shows a general trend of higher frequencies at lower densities, although we find a sharp drop-off in planet density at around 6 g/cm3. There is no obvious physical reason why such a discontinuity would exist in nature, so we suspect this edge to be an artifact of the models used to infer planet radii (given that radial velocity measurements directly characterize mass and radii must be estimated).

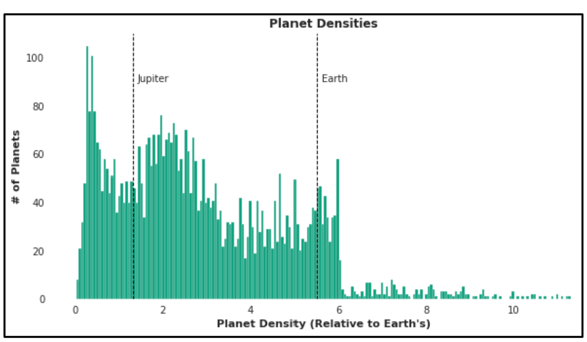

Figure 19. Density frequencies

We see further evidence for such modeling artifacts when we evaluate a scatter plot of object masses and radii. As seen in the scatter plot, there is a clear indication of underlying linear trends that are largely attributable to objects discovered using the radial velocity method. Nevertheless, we do see that the model used to estimate planet radii detected by radial velocity seems fairly descriptive of trends seen in exoplanets discovered with the transit method (for which both properties can be directly measured).

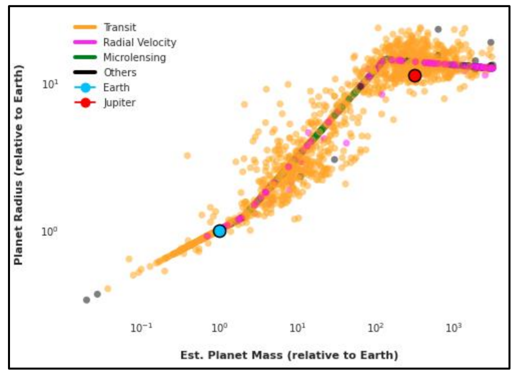

Figure 20. Discoveries by mass and radius

<u>Stellar characteristics</u>

The dataset contains a rich variety of stellar properties including mass, spectral type, luminosity, temperature, metallicity, and age. These data represent a fascinating opportunity for comparison with large star catalogues. Furthermore, a star's luminosity and temperature, given assumptions, can be used to estimate that star's so-called habitable zone, which is the range of distances at which planetary bodies of various atmospheric compositions can retain liquid water. In principle, such estimates can be used to identify exoplanets with the potential for hosting water oceans, a likely necessity to life as we know it.

While our initial exploration of the data did include a brief foray into these variables, we restrict our current discussion of stars to their distance from Sol. Here, we find that exoplanets have been detected orbiting stars over 8,000 ly (for context, the galaxy is believed to be roughly 90,000 ly). However, a frequency distribution clearly shows that far more of the stars are within a few hundred ly. Given that photon counts drop off with the square of distance, and therefore nearer stars present larger signals for all observational techniques, this trend is not surprising.

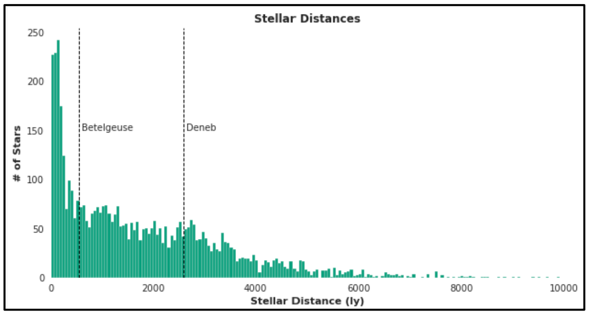

Figure 21. Distance frequencies

Earth-like exoplanets

We consider the number of planets co-inhabiting a given solar system and find that the vast majority of detected exoplanets are companionless. We expect most systems to have many planets, like our own, and expect future discoveries to flatten this trend. The infamous system around the star TRAPPIST-1 is comprised of 7 detected planets.

1	2977
2	511
3	159
4	62
5	15
6	9
7	1

Figure 22. Planet counts by system multiplicity

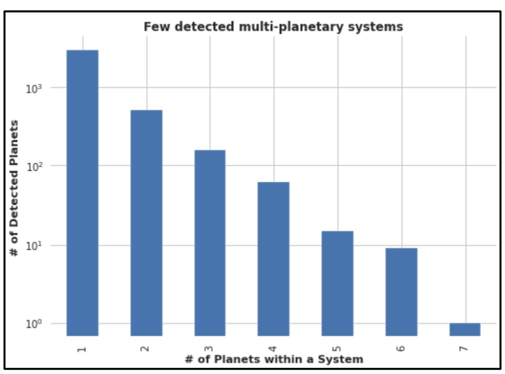

Figure 23. Planet counts by system multiplicity

Our own solar system has a distinctive differentiation, with smaller planets orbiting more closely to the Sun than the giants. But does this relationship generalize to observed exoplanets? Based solely on the planet mass and radius categories previously proposed, we find generally good agreement with this trend. Notably, we find that the dataset contains 303 'small' planets with 'terrestrial' masses.

bmasse_cat	Tiny	Small	Medium	Large
Sub-Terrestrial	4	133	1	0
Terrestrial	0	303	1	0
Super-Terrestrial	0	1037	1292	5
Sub-Giant	0	3	617	205
Giant	0	0	1	578
Super-Giant	0	0	0	680

Figure 24. Counts by mass and radius categories

What exoplanets discovered thus far are the most Earth-like? We can identify the most Earth-like planets detected by further constraining their mass and radii of to between 75% and 125% the mass and radius of Earth. When we do so, we can evaluate the number of them within 50 ly, of which there are 8. All of them are somewhat larger than Earth, and most are within 15 ly. Interestingly, one of them orbits Proxima Centauri (the star nearest to the Sun), and another is within the TRAPPIST-1 system.

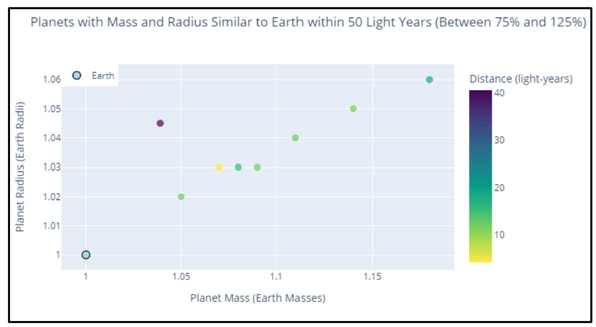

Figure 25. Nearest Earth-like planets by mass & radius

And how do these exoplanetary systems compare to the Sol system? The following figure portrays the distances of planets within these systems from their host stars, sized according to their masses, with red planets indicating they are Earth-like.

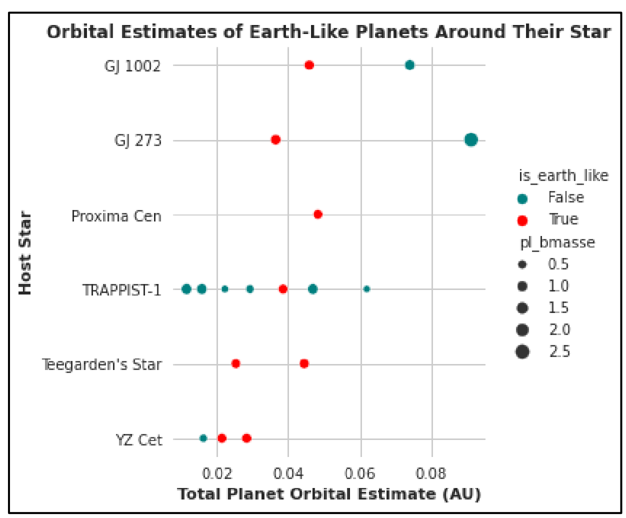

Figure 26. Notable nearby systems

We note that the orbital radii seem quite small, at less than 10% of he Earth-Sun distance, and we thus re-scale the above figure to include the Earth-Sun distances for scale. We note that all planets within these systems are in much tighter orbits than even Mercury.

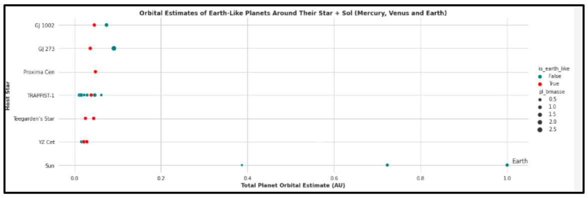

Figure 26. Notable nearby systems (comp. to Sol)

A closer examination of these host stars indicates all of them are low-temperature (less than 3,500K) red-dwarf (of spectral type M) stars. Such stars are considerably smaller and cooler than Sol, and therefore these planets may reside within their star's habitable zone. Nevertheless, even if such worlds could sustain a liquid ocean, they'd be quite different dynamically since their proximity to their hosts suggests they are tidally locked.

Frontier Characteristics

Spectroscopy measurements represent the cutting edge of exoplanet characterization. These measurements cannot be used to make an exoplanet discovery, but rather involve collecting additional data on confirmed exoplanets. Such data are indicative of a planet's atmospheric composition, which can be used to infer the presence of chemicals indicative of biological processes.

We find that 114 planets have undergone spectroscopy measurements. Typically, a single spectroscopy measurement is insufficient to generate a spectrum well-resolved enough to make inferences about an exoplanet's atmosphere, and so we are interested in the number of spectra collected for each planet. We find that most planets characterized by this technique have only a fewer than 50 spectra measurements. Depending on the resolution of the data and the compounds of interest, 50 spectra may be sufficient to resolve key organics such as carbon dioxide and water.

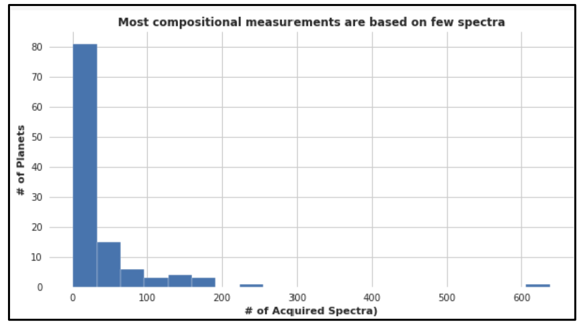

Figure 27. Spectroscopy measurement frequencies

Conclusion

We have only explored a small subset of the available data, but in doing so have demonstrated it contains a rich new context for understanding our own solar system. The transit detection technique exploited for a few years in the 2010's generated a great deal of the discoveries thus far, and has demonstrated excellent range sensitivity for mass, radius, and period. It is furthermore well-suited for detecting small terrestrial planets like the Earth, particularly those at short orbital distances around the common red dwarf stars.

Clearly, the data are in their infancy. We can see that there are biases in the data in the form of detection limits, modeling artifacts, and insufficient sampling. We look forward to improvements in data quality, and imagine further enrichment by incorporating supplemental datasets (such as stellar catalogues and comprehensive Sol planetary data).

But based on the available data it is evident that our solar system's configuration is somewhat abnormal, in that it lacks super-terrestrial planets which appear to be abundant. And crucially, we have learned that Earth-like planets are even rarer than we previously understood: indeed, of the nearly 5000 planets discovered, none sit at comparable distances around comparable stars and are of comparable size and mass – all of which are critical for life as we know it.

Contrary to popular science fiction, we believe that, as new discoveries are made in the coming years, and those planets are characterized with ever-more sophisticated methods, the bounty of data will reinforce the realization that Earth is special, perhaps truly unique in the galaxy. Such knowledge belongs to our community, and an interactive exhibit at the Museum would be a fantastic medium for transmitting this exciting and humbling knowledge.

References

- Bashi, D.; Helled, R.; Zucker, S. “A Quantitative Comparison of Exoplanet Catalogs”. Geosciences 8. 9(2018): 325.
- Natalie M. Batalha. "Exploring exoplanet populations with NASA’s Kepler Mission". Proceedings of the National Academy of Sciences 111. 35(2014): 12647-12654.
- Andrew N. Youdin. "THE EXOPLANET CENSUS: A GENERAL METHOD APPLIED TO KEPLER". The Astrophysical Journal 742. 1(2011): 38.
- Jacob L. Bean, et al. "A Statistical Comparative Planetology Approach to the Hunt for Habitable Exoplanets and Life Beyond the Solar System". The Astrophysical Journal Letters 841. 2(2017): L24.
- Zhu, Wei et al. "Exoplanet Statistics and Theoretical Implications". Annual Review of Astronomy and Astrophysics 59. 1(2021): 291-336.
- Emsenhuber, A., Mordasini, C. & Burn, R. “Planetary population synthesis and the emergence of four classes of planetary system architectures”. Eur. Phys. J. Plus 138, 181 (2023).
- Colby Ostberg, et al. "The Demographics of Terrestrial Planets in the Venus Zone". The Astronomical Journal 165. 4(2023): 168.
- Rob J. Spaargaren, et al. "Plausible constraints on the range of bulk terrestrial exoplanet compositions in the Solar neighbourhood." (2022).
- Sarah R N McIntyre, et al. “A rocky exoplanet classification method and its application to calculating surface pressure and surface temperature”. Monthly Notices of the Royal Astronomical Society 519. 4(2023): 6210–6221.
- S. Marchi, . "Extrasolar Planet Taxonomy: A New Statistical Approach". The Astrophysical Journal 666. 1(2007): 475.
- Udry, S. et al. "Statistical properties of exoplanets - I. The period distribution: Constraints for the migration scenario". A&A 407. 1(2003): 369-376.
- Eggenberger, A. et al. "Statistical properties of exoplanets - III. Planet properties and stellar multiplicity". A&A 417. 1(2004): 353-360.
- Halbwachs, J. L. et al. "Statistical properties of exoplanets - IV. The period-eccentricityrelations of exoplanets and of binary stars". A&A 431. 3(2005): 1129-1137.
- Udry, Stéphane et al. "Statistical Properties of Exoplanets". Annual Review of Astronomy and Astrophysics 45. 1(2007): 397-439.