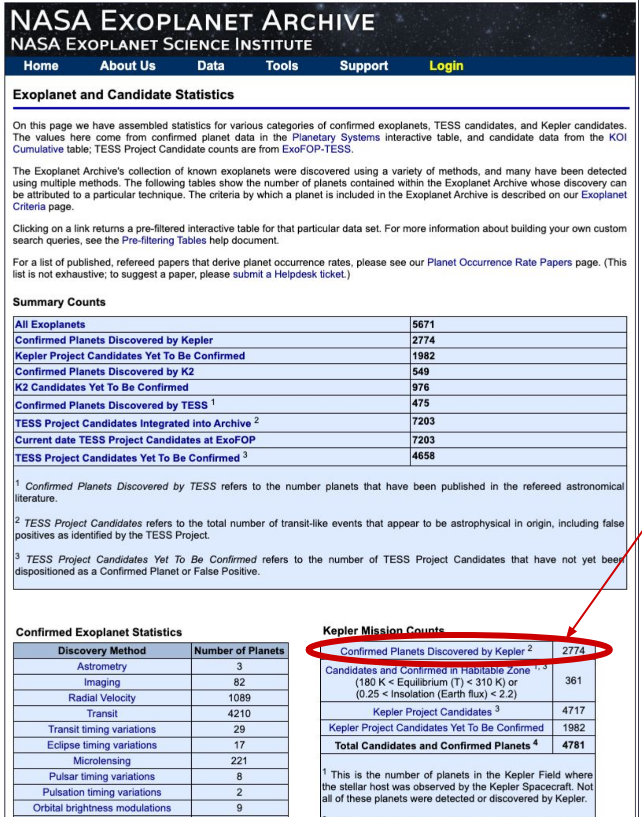
June 19th, 2024

Exoplanet Classification

1. Data Sources



NASA EXOPLANET ARCHIVE
NASA EXOPLANET SCIENCE INSTITUTE

Home About Us Data Tools Support Login

Exoplanet and Candidate Statistics

On this page we have assembled statistics for various categories of confirmed exoplanets, TESS candidates, and Kepler candidates. The values here come from confirmed planet data in the Planetary Systems interactive table, and candidate data from the KOI Cumulative table; TESS Project Candidate counts are from ExoFOP-TESS.

The Exoplanet Archive's collection of known exoplanets were discovered using a variety of methods, and many have been detected using multiple methods. The following tables show the number of planets contained within the Exoplanet Archive whose discovery can be attributed to a particular technique. The criteria by which a planet is included in the Exoplanet Archive is described on our Exoplanet Criteria page.

Clicking on a link returns a pre-filtered interactive table for that particular data set. For more information about building your own custom search queries, see the Pre-filtering Tables help document.

For a list of published, refereed papers that derive planet occurrence rates, please see our Planet Occurrence Rate Papers page. (This list is not exhaustive; to suggest a paper, please submit a Helpdesk ticket.)

Summary Counts

All Exoplanets	5671
Confirmed Planets Discovered by Kepler	2774
Kepler Project Candidates Yet To Be Confirmed	1982
Confirmed Planets Discovered by K2	549
K2 Candidates Yet To Be Confirmed	976
Confirmed Planets Discovered by TESS [1]	475
TESS Project Candidates Integrated into Archive [2]	7203
Current date TESS Project Candidates at ExoFOP	7203
TESS Project Candidates Yet To Be Confirmed [3]	4658

[1] *Confirmed Planets Discovered by TESS* refers to the number planets that have been published in the refereed astronomical literature.

[2] *TESS Project Candidates* refers to the total number of transit-like events that appear to be astrophysical in origin, including false positives as identified by the TESS Project.

[3] *TESS Project Candidates Yet To Be Confirmed* refers to the number of TESS Project Candidates that have not yet been dispositioned as a Confirmed Planet or False Positive.

Confirmed Exoplanet Statistics

Discovery Method	Number of Planets
Astrometry	3
Imaging	82
Radial Velocity	1089
Transit	4210
Transit timing variations	29
Eclipse timing variations	17
Microlensing	221
Pulsar timing variations	8
Pulsation timing variations	2
Orbital brightness modulations	9

Kepler Mission Counts

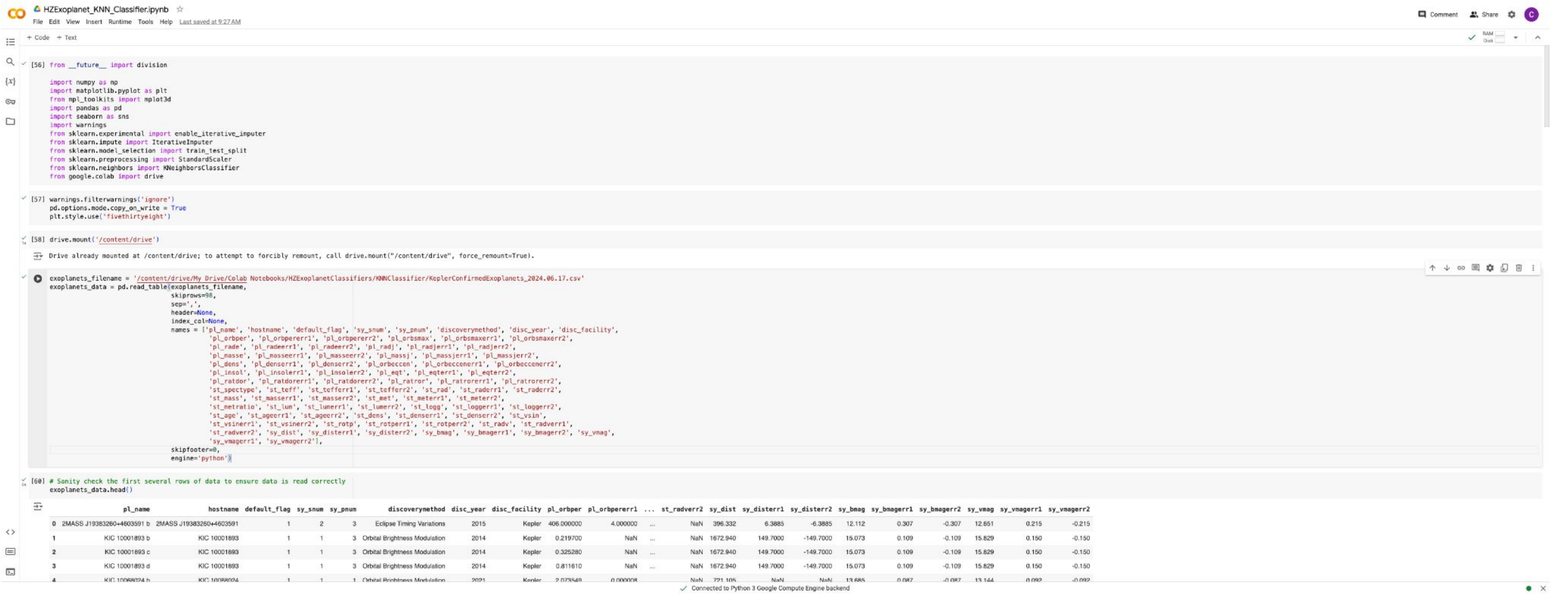
Confirmed Planets Discovered by Kepler [2]	2774
Candidates and Confirmed in Habitable Zone [1, 3] (180 K < Equilibrium (T) < 310 K) or (0.25 < Insolation (Earth flux) < 2.2)	361
Kepler Project Candidates [3]	4717
Kepler Project Candidates Yet To Be Confirmed	1982
Total Candidates and Confirmed Planets [4]	**4781**

[1] This is the number of planets in the Kepler Field where the stellar host was observed by the Kepler Spacecraft. Not all of these planets were detected or discovered by Kepler.

- Data from: [Confirmed Planets Discovered by Kepler]
    - 2774 confirmed

2. Environment

- Google Colab
- Jupyter Notebook



3. Exoplanet Type by Radius and Mass

```python
# Mapping from exoplanet type to the corresponding color in the plotting graph
exoplanet_type_colors = {
    'Miniterran': 'darkred',
    'Subterran': 'red',
    'Terran': 'blue',
    'Superterran': 'deepskyblue',
    'Neptunian': 'powderblue',
    'Jovian': 'wheat'
}

# Mapping from exoplanet type to the corresponding legend title in the plotting graph
exoplanet_type_labels = {
    'Miniterran': 'Miniterran',
    'Subterran': 'Subterran (Mars-Sized)',
    'Terran': 'Terran (Earth-Sized)',
    'Superterran': 'Superterran (Super-Earth/Mini-Neptunes)',
    'Neptunian': 'Neptunian (Neptune-Sized)',
    'Jovian': 'Jovian (Jupiter-Sized)'
}
```

- Different types of exoplanets:
    - **miniterran**:
        - *radius compared to Earth: 0.03 to 0.4 times.*
        - mass compared to Earth: 0.00001 to 0.1 times
    - **subterran:**
        - *radius compared to Earth: 0.4 to 0.8 times*
        - mass compared to Earth: 0.1 to 0.5 times
    - **terran**
        - *radius compared to Earth: 0.8 to 1.5 times*
        - mass compared to Earth: 0.5 to 5 times
    - **superterran**
        - *radius compared to Earth: 1.5 to 2.5 times*
        - mass compared to Earth: 5 to 10 times
    - **neptunian**
        - *radius compared to Earth: 2.5 to 6 times*
        - mass compared to Earth: 10 to 50 times
    - **jovian**
        - *radius compared to Earth:> 6 times*
        - mass compared to Earth:> 50 times

Habitable Zone

- A bunch of different ways to calculate based on different sources…

where A is the planetary albedo and σ is the Stefan–Boltzmann constant. The simplest habitability condition is $273\,\mathrm{K} < T_{eq} < 373\,\mathrm{K}$, with the low T_{eq} defining the outer boundary of the habitable zone (OHZ) and the high T_{eq} defining the inner boundary of the habitable zone (IHZ). The scalar quantity k is a correction factor that can be used to approximately incorporate the greenhouse effect of an assumed planetary atmosphere; see Ref. [18]. We adopt the Earth albedo of $A = 0.3$ and use $k = 1.13$, which reproduces the Earth surface temperature.

More realistic criteria exist in the literature. In this paper, we shall adopt two criteria obtained in a previous study [11] in which an effective solar flux is expressed in terms of

$$S_{eff} \equiv S/S_{\oplus} \tag{6}$$

where $S_{\oplus}$ is the current solar energy flux at the location of the Earth, as well as the temperature T of the host star. Note that S_{eff} is dimensionless. Following Ref. [11], we use two different ways to define HZ boundaries, one conservative, the other optimistic. The

source: [Exoplanets around Red Giants: Distribution and Habitability]

Habitable Zone

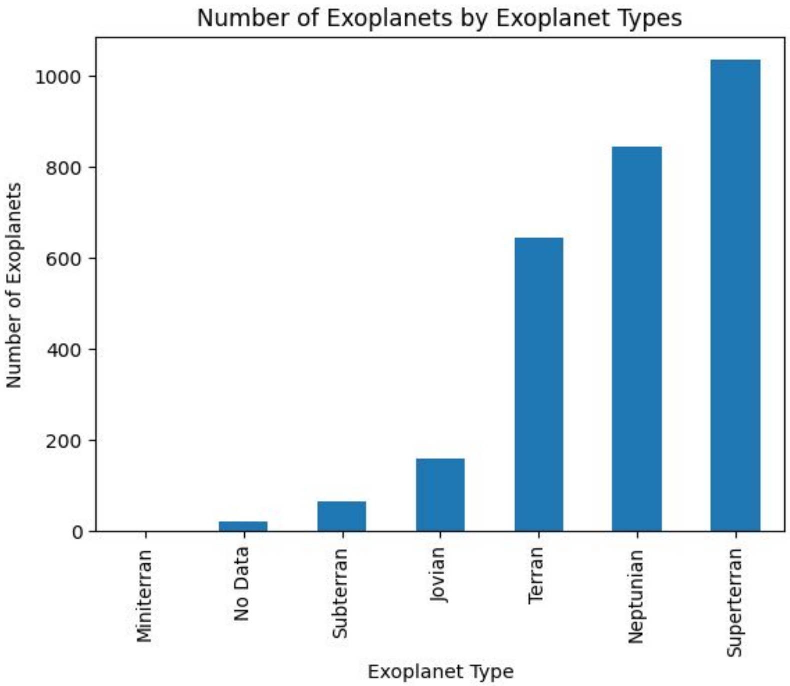
- This is the one I used for these graphs

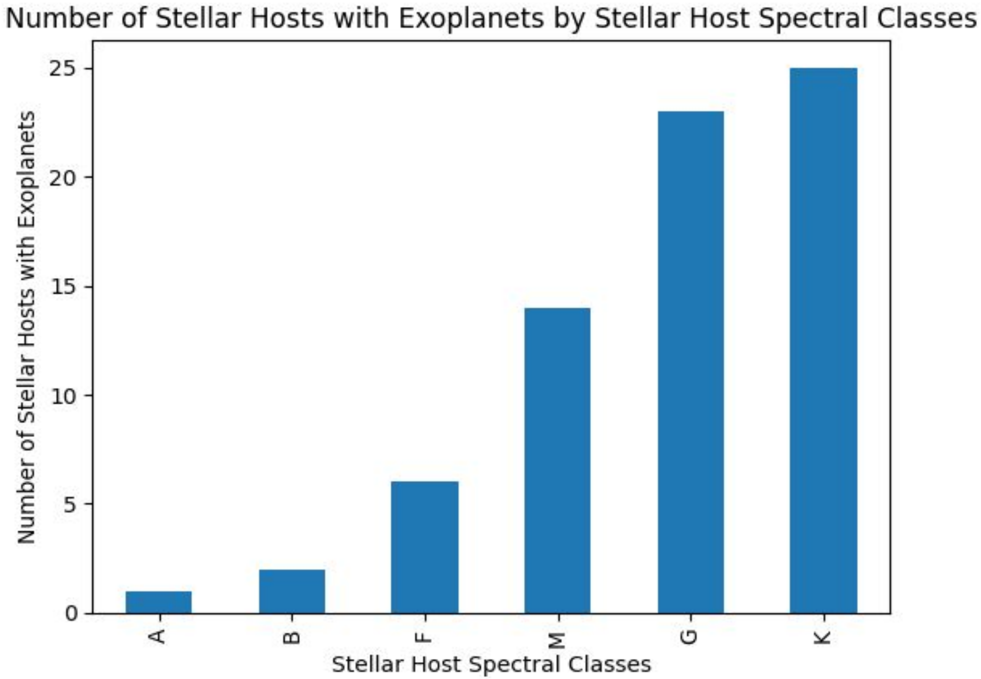
Kepler Mission Counts

Confirmed Planets Discovered by Kepler [2]	2774
Candidates and Confirmed in Habitable Zone [1, 3] (180 K < Equilibrium (T) < 310 K) or (0.25 < Insolation (Earth flux) < 2.2)	361
Kepler Project Candidates [3]	4717
Kepler Project Candidates Yet To Be Confirmed	1982
Total Candidates and Confirmed Planets [4]	**4781**

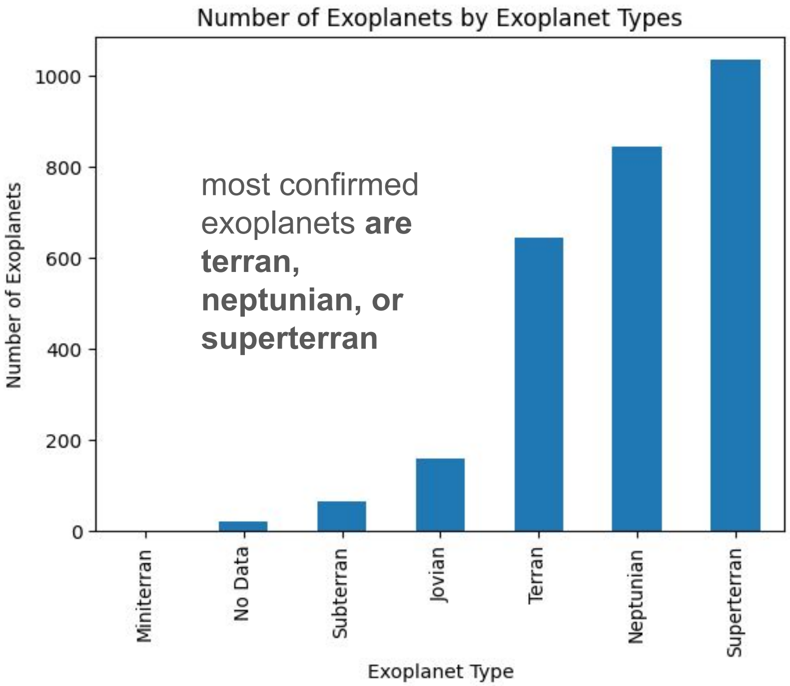
source: [Exoplanet and Candidate Statistics]()

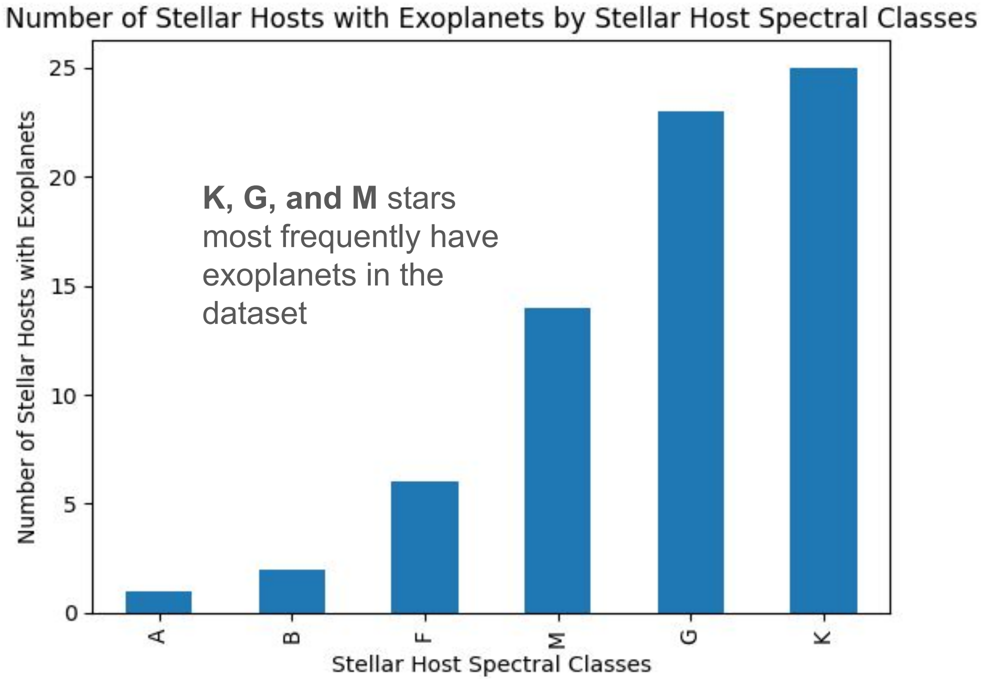
4. Graphs and Results





4. Graphs and Results





SPECIAL FEATURE: INTRODUCTION

Exoplanets

Adam S. Burrows[a,1] **and Geoffrey W. Marcy**[b]
[a]Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544; and [b]Department of Astronomy, University of California, Berkeley, CA 94720

We are in the midst of a revolution in our understanding of planets in the Universe. In the last 20 years, 200-times more planets have been discovered beyond our solar system than reside within it. Astronomers are determining their properties, are finding correlations between a star's type and its planet population, and have begun to probe exoplanetary atmospheres. One idea to emerge is that the planets of the solar system are not representative in type, mix, or orbit. Another is that comparative planetology has become a palpable reality. A third is that we are merely on the cusp of a transformation in our conceptions of planet birth and evolution. With all this progress, textbooks are being rewritten yearly. To communicate a sense of the pace and excitement of this new subject, we have assembled a collection of articles by some of the leading researchers in their subfields that summarizes the current state of play across the broadening spectrum of topics relevant to the study of exoplanets.

In "Spectra as windows into exoplanet atmospheres," Adam Burrows (1) provides a perspective on exoplanet theory and remote sensing via photometry and spectroscopy. He highlights the need for better spectra and the limitations—despite recent progress—in our current knowledge of exoplanet atmospheres. Importantly, Burrows emphasizes that the true function of the recent past of exoplanet research has been to train the next generation of scientists who will ensure that the foundations of comparative exoplanetolgy in the future will be robust.

Yoram Lithwick and Yanqin Wu (2) explore the organization of planetary orbits and systems resulting from chaos in their gravitational dynamics and mutual evolution. Eric Ford (3) also reviews ideas concerning the architectures of exoplanetary systems, and addresses some implications for their formation. In particular, Ford focuses on the creation of multiple planet systems in tight orbits, the excitation of orbital eccentricities, and the usefulness of transit timing variation caused by mutual gravitational interactions in providing constraints on the formation and orbital evolution of planetary systems with low-mass planets.

David Spiegel, Jonathan Fortney, and Christophe Sotin (4) delve into the physical structure of exoplanets. Emphasizing that the emerging data suggest that our solar system is in no way typical, these authors review the diversity of possible planet interior structures, from super-Jupiters to planets half the mass of the Earth. Along the way, they discuss the variety of material properties—from those of compressed solids to those of dilute gases—that determine the size and profiles of planets in the galaxy. The relevant equations of state are explored and the possible role of the radius-mass distribution function in informing our understanding of planet formation and evolution is investigated.

Christopher McKay (5) reviews the requirements and limits for life and habitability in the context of exoplanets. Using the variety of life on Earth as a guide, McKay emphasizes that temperature is a key and that low light levels, high UV fluxes, and low oxygen levels need not be existential problems for various possible forms of alien life, although the biological availability of nitrogen may be an issue.

Sara Seager (6) reviews the likely means by which life could be inferred on rocky exoplanets—the identification of atmospheric biosignatures in the planet's spectrum—and the many obstacles to the robust identification of those biomolecules. Seager summarizes lessons learned to date through the study of scores of giant and sub-Neptune exoplanet atmospheres and the inherent limitations of remote sensing from the Earth, and posits a way forward to reliably ascertain the presence of life beyond the solar system. Along the way, Seager suggests an expansion of the concept of habitable zone. Continuing on this theme, James Kasting et al. (7) investigate the exact thermodynamic and radiative conditions necessary to maintain liquid water. These authors then go on to argue that liquid water is necessary for carbon-based, photosynthetic life to proliferate on a rocky planet to a degree sufficient to modify its atmosphere enough to enable spectroscopic detection by us. Kasting et al. also argue that conservative habitable-zone definitions should be used for designing future space-based telescopes, but that optimistic definitions will be necessary to interpret data from such missions. Interestingly, Kasting et al. go on to estimate that the frequency of Earth-like planets around stars smaller than the Sun might be as high as 40–50%.

Natalie Batalha (8) summarizes the demographics of planetary systems that have emerged from the Kepler space mission. With more than 3,500 planet candidates and a reliability rate of 85–90%, Batalha presents the revealed distributions in planet radius and orbital period as a function of primary star. Designed to determine the prevalence of planets like the Earth, such distributions and statistics are the major products of the Kepler mission. Batalha also reviews the role of the methods of transit timing, radial-velocity measurement, and statistical validation in determining reliable planet occurrence rates, and shows that planets—in particular small planets and multiple planet systems—abound in our galaxy.

Geoffrey Marcy et al. (9) continue this theme to show that planets one- to four-times the radius of Earth, a population missing in the solar system, are common. These authors reiterate that the smallest planets are the most common and determine distributions in planet radius and orbital period. Marcy et al. supplement the Kepler data for radii with mass measurements from radial velocity techniques to reveal roughly two populations of exoplanets smaller than four-times the radius of the Earth: those with the lowest masses, whose radii imply thin atmospheres, and those with larger radii that can be explained only by the presence of thick gaseous envelopes. Although noisy, this rough systematic must reflect modes of planet formation and subsequent envelope evaporation. Marcy et al. note that rocky planets occur around stars having a range of heavy element abundances, whereas gas giant exoplanets are more common around stars rich in heavy elements. This finding leads the authors to speculate that the fast formation of rocky cores in protoplanetary disks enriched in heavy elements might

Author contributions: A.S.B. and G.W.M. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: burrows@astro.princeton.edu.

This content downloaded from
174.165.1.15 on Wed, 19 Jun 2024 19:32:49 +00:00
All use subject to https://about.jstor.org/terms

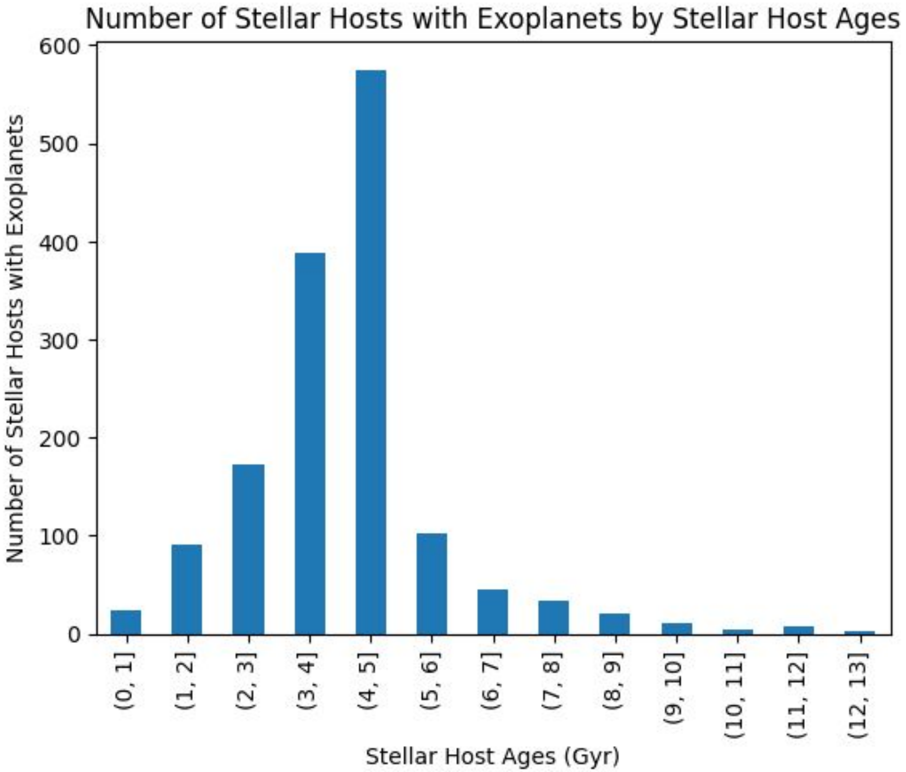
Geoffrey Marcy et al. (9) continue this theme to show that planets one- to four-times the radius of Earth, a population missing in the solar system, are common. These

matches up with my findings!!

source:

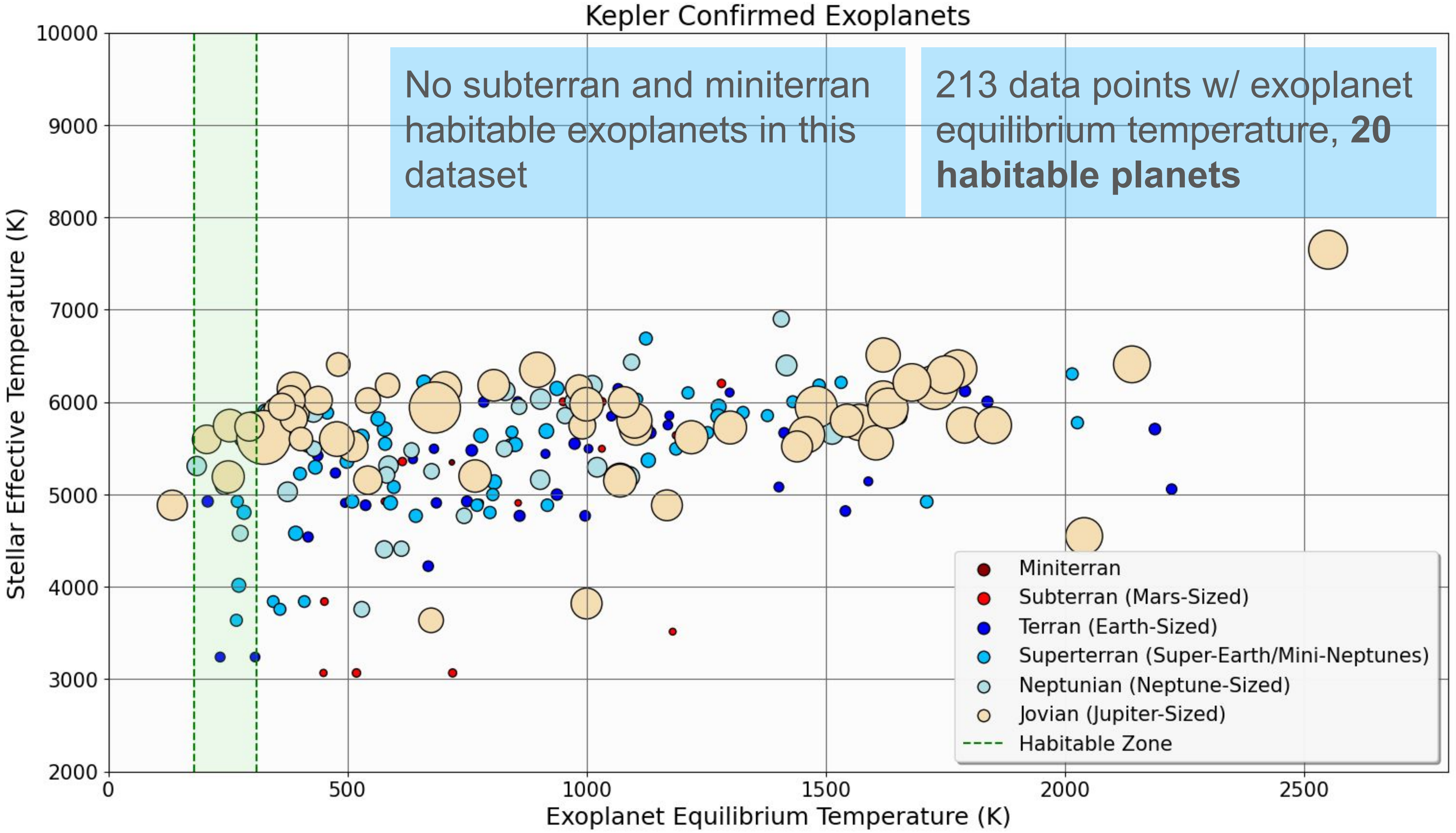
Burrows, Adam S., and Geoffrey W. Marcy. "INTRODUCTION: Exoplanets." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 35, 2014, pp. 12599–600. *JSTOR*, http://www.jstor.org/stable/43043378. Accessed 19 June 2024.

4. Graphs and Results



No subterran and miniterran habitable exoplanets in this dataset

213 data points w/ exoplanet equilibrium temperature, **20 habitable planets**



5. Looking into a ML-based Approach

- Implement a simple machine learning model
    - Can help find patterns better

KNN Classifier Machine Learning Model

- Used the SKLearner ML library's **KNeighborsClassifier** – to train a KNN classifier based on training data from the Kepler confirmed exoplanets
- Training data generation
  - Label data -

    ```
    [223] exoplanets_data.loc[((~np.isnan(exoplanets_data['pl_eqt'])) & (exoplanets_data['pl_eqt'] > 180) & (exoplanets_data['pl_eqt'] < 310)), 'label'] = 1
          exoplanets_data.loc[((~np.isnan(exoplanets_data['pl_eqt'])) & (exoplanets_data['pl_eqt'] <= 180) | (exoplanets_data['pl_eqt'] >= 310)), 'label'] = 0
    ```
  - Clean data – drop irrelevant features (e.g. pl_name) or features with too many missing values, with 24 remaining
  - Challenges in the training data:
    - small size (only 213 data samples with labels)
    - many missing data field values – solved by simple imputation (taking the mean)

      ```
      training_data = training_data.fillna(training_data.mean())
      ```
  - Standard scalar – scale features to standardized values
- Train the model
  - Split the training data: 70% for training, 30% for testing

KNN Classifier Machine Learning Model (Results)

```
KNN Classifier Accuracy:  0.921875

KNN Classifier Classification Report :
              precision    recall  f1-score   support

         0.0       0.93      0.98      0.96        58
         1.0       0.67      0.33      0.44         6
```

Model performs decently to predict non-HZ exoplanets with precision 0.93 and recall 0.98.

However, the model performs pretty badly in predicting HZ exoplanets with precision 0.67 and recall 0.33.

KNN Classifier Machine Learning Model (Results)

Discussion:

- Not enough training data (only 213 data samples for both training and testing).
- Data is heavily skewed to non-HZ planets, which could be a cause as to why the model performs decently on predicting non-HZ but badly when predicting HZ.
    - This may be solvable by techniques such as downsampling and/or oversampling.
- But I don’t think that there is enough training data for an accurate model in regards to predicting HZ exoplanets.
- However, we have enough (7203) candidates on the TESS project candidates.

Next Steps: Training a Model to Predict Exoplanets

- Use the TESS Project Candidates (7203 data samples)
- Enough training data with positive (PC) and negative (FP) samples!!

TESS Object of Interest	TESS Input Catalog ID	TFOPWG Dispostion (CP | FP | KP | PC)
TOI-1000.01	TIC 50365310	FP
TOI-1001.01	TIC 88863718	PC
TOI-1002.01	TIC 124709665	FP
TOI-1003.01	TIC 106997505	FP
TOI-1004.01	TIC 238597883	FP
TOI-1005.01	TIC 169904935	FP
TOI-1006.01	TIC 156115721	FP
TOI-1007.01	TIC 65212867	PC
TOI-1008.01	TIC 440801822	FP
TOI-1009.01	TIC 107782586	PC
TOI-101.01	TIC 231663901	KP
TOI-1010.01	TIC 139853601	FP
TOI-1011.01	TIC 114018671	PC