An Up-To-Date Exploratory Survey of 4434 Confirmed Exoplanets and an Insight into the Sensitivity of Detection Methods

By Donggeon Kim

Author Bio:

Donggeon Kim, born in Seoul, South Korea, is a senior at the Shanghai American School – Puxi Campus. He has an exceptional passion for astrophysics, particularly in exoplanets, and is intending to major in either Applied Mathematics, Astrophysics, or Aerospace Engineering. He first became interested in the universe after visiting an observatory with his family in grade 5. Now, he loves applying his math and physics skills to many different disciplines to solve real-world problems. He is also an avid communicator, striving to share his passion with his community through technology and education. Outside of academics, you can find him playing baseball, table tennis, solving Rubik's cube, and watching *Interstellar*.

Abstract

The recent development of detection methods has produced large metadata of exoplanets, which are stored and regularly updated on the NASA Exoplanet Archive. In this paper, classification and habitability schemes by Kopparapu, R. et al are incorporated to classify 4434 confirmed exoplanets into 5 different categories—Rocky, Super-Earths, Sub-Neptune, Sub-Jovian, and Jovian—and to compute their habitability. In addition, 20 graphs are plotted to analyze the trends of exoplanets' characterization, habitability, and detection methods sensitivity. This paper also upgrades a current NASA Exoplanet Archive catalog to be more comprehensive by appending classification and habitability flag columns. Moreover, a novel exoplanet classification scheme is proposed to utilize Weighted Average for more comprehensive metadata from futuristic telescopes such as the James Webb Space Telescope. Codes, datasets, plots, and the upgraded exoplanet catalog from this research are published on the GitHub page, for readers to repeat this research's analysis with newly updated exoplanet data.

Keywords: Physics and Astronomy; Planetary Science; Exoplanet; NASA Exoplanet Archive; Exoplanet Classification; Habitable Zone; Goldilocks Zone; Detection Method Sensitivity

Introduction

Our search for worlds beyond the Earth - the exoplanets, or the planets outside the Solar System - was sparked over 2000 years ago, when the Greek philosopher Epicurus (ca. 300 B.C.E.) asserted “There are infinite worlds both like and unlike this world of ours” [1]. However, it was only in 1992 that the first exoplanet was officially detected and confirmed [2]. Thus, the field of exoplanet research is in a relatively new era, but a rapid one: over the past two decades, with a help of advanced detection techniques, more than 4000 exoplanets have been confirmed along with more than 7000 “candidate” exoplanets [3]. As a side note, exoplanets are considered confirmed only once they are verified through additional observation using two other telescopes [4].

1. Classification of Exoplanets

The exponential detection of exoplanets signaled a growing need to classify exoplanets into certain categories to understand their diversity[5]. Exoplanets are typically classified into certain categories based on their characteristics, such as radius and mass, compared to our own known solar planets. For example, NASA categorized exoplanets into Terrestrial, Super Earth, Gas Giant, and Neptune-like [3]. While there is no one officially accepted classification scheme because of the complexity of exoplanets that cannot be described by just one mathematical model, this paper uses the classification scheme proposed by Dr. Ravi Kumar Kopparapu et al. in 2018 in their paper “Exoplanet Classification and Yield Estimates for Direct Imaging Missions”. This classification scheme prioritizes the size of exoplanets and the starlight flux on the planet as main factors on planet composition between 5 categories: Rocky, Super Earths, Sub-Neptune, sub-Jovian, and Jovian [5].

2. Habitability of Exoplanets

Another way to classify exoplanets is by determining the habitability of exoplanets: whether the exoplanet is within the habitable zone boundary. The habitable zone, also known as the Goldilocks zone, is the circular boundary of orbits around a star within which a planetary surface is not too cold and not too hot to support essential substances for life such as CO_2 and H_2O [6]. Again, there is no one officially accepted habitable zone scheme, but this paper applies the habitable zone scheme proposed by Dr. Ravi Kumar Kopparapu et al. in 2013 in their paper “Habitable Zones Around Main-Sequence Stars: New Estimates”, which takes into account the stellar luminosity computed by stellar effective temperature and stellar radius [6].

3. Detection Methods of Exoplanets

Since exoplanets are small, and host stars are so bright that they outshine their planets, it is generally difficult to directly detect exoplanets. In light of this, there have been many approaches to indirectly detect exoplanets by searching for characteristics of host stars, which are easier to detect. As of now, there are largely 5 detection methods: transit method, radial velocity, microlensing, direct imaging, and pulsar timing. The transit method detects the tiny dips in light when the exoplanet crosses the host star and blocks the starlight in the direction to the Earth [1]. The radial velocity method detects the Doppler’s effect of a host star’s changes in radial velocity caused by the gravitational pulls between the star and the planet. Keep in mind that the star orbits around the center of its planetary system, not just staying in its center. If the host star moves toward the direction of the Earth, the wavelengths of starlight are squeezed, emitting blue-shifted light signal, and if it moves away, the wavelengths of starlight are stretched, emitting red-shifted light signal [1]. The microlensing method is used when a foreground star happens to pass a more distant background star. As the foreground star passes the background star, the background star’s brightness will increase due to the Gravitational Lensing effect. If the foreground star happens to host a planet, the planet will also act as a Gravitational lens, resulting in a unique peak in the background star’s brightness [1]. Direct imaging uses infrared wavelengths to directly observe planets [1]. Lastly, pulsar timing detects exoplanets around pulsars, which emit an intense electromagnetic radiation on a regular rate. The slight regular variations in the timing of the pulses indicate that pulsar orbits around the center of mass of a system with one or more planets, suggesting the existence of the exoplanets [1].

4. Outline of the Paper

This paper focuses on analyzing the demographics of currently confirmed 4434 exoplanets based on classification and habitability schemes, as well as an insight of the sensitivity of each detection method. Along with this analysis, I made two contributions to the exoplanet community. While the exoplanets catalog provides extensive information about each exoplanet and its host star, the catalog does not include the habitability and the classification category of each exoplanet. Employing a highly respected habitable zone scheme (cited 1102 times) and classification scheme proposed by the same author, I updated the current exoplanet catalog with the habitability flag column and classification column. The updated catalog will provide more comprehensive information and trends of currently confirmed exoplanets.

This paper is divided into 6 parts: *introduction, materials and method, plots, discussion, caveat and future research, and conclusion*. I also shared my code, plots, dataset, and the updated exoplanet catalog on my GitHub page (https://github.com/SteveHawKim03/exoplanet-analysis) so that readers can repeat my analysis whenever new confirmed exoplanets are added to the NASA Exoplanet Archive catalog.

Materials and Methods

1. Exoplanet Catalog

This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program [7]. There were 4434 confirmed exoplanets as of July 12th 2021.

While there are more than 100 variables to each exoplanet, my research focuses on 10 variables. These 10 variables are explained in more detail in the table below [8]:

I. Variables and Descriptions

Variable Name	Unit	Description
Detection Method	N/A	Method by which the planet was first identified.
Orbital Period	Days	Time the planet takes to make a complete orbit around the host star or system.
Orbit Semi-Major Axis	AU	The longest radius of an elliptic orbit of the planet. Used to represent the separation between the host star and the exoplanet.
Planet Radius	Earth Radius	Length of a line segment from the center of the planet to its surface, measured in units of radius of the Earth.
Planet Mass	Earth Mass	Best planet mass measurement/approximation in units of masses of Earth
Insolation Flux	Earth Flux	Flux of solar radiation per unit of horizontal area for a planet. Another way to give the equilibrium temperature, which is the temperature of the planet as modeled by a black body heated only by its host star.
Stellar Effective Temperature	Kelvins	Temperature of the star as modeled by a black body emitting the same total amount of electromagnetic radiation.
Stellar Radius	Solar Radius	Length of a line segment from the center of the star to its surface, measured in units of radius of the Sun.
RA	Degree	Right Ascension - east and west of the celestial equator - of the planetary system.
Dec	Degree	Declination - north and south of the celestial equator - of the planetary system.

2. Python Packages

For my analysis on these data, I used Python along with four packages: Pandas, NumPy, Matplotlib, and Astropy [9][10]. I used Pandas to read and write csv files of the exoplanets catalog, NumPy to generate arrays of data and compute mathematical operations, Matplotlib to plot graphs, and Astropy to bring astrophysical constants and analyze the plots.

3. Habitability Scheme

As mentioned in the introduction, I used the habitable zone scheme proposed by Dr. Ravi Kumar Kopparapu et al. in their paper "Habitable Zones Around Main-Sequence Stars: New Estimates" [6]. The scheme first calculates the habitable zone stellar fluxes (S_{eff}), measured in Kelvins [1], reaching the top of the atmosphere of an Earth-like planet in relation to the stellar effective temperature (T_{eff}), measured in Kelvins:

where T. = T_{eff} − 5780 K and the coefficients , *a, b, c*, and *d* are as follows:

II. Habitable Zone Constants

Constant	Recent Venus	Runaway Greenhouse	Moist Greenhouse	Maximum Greenhouse	Early Mars
	1.7753	1.0512	1.0140	0.3438	0.3179
a	1.4316×10^{-4}	1.3242×10^{-4}	8.1774×10^{-5}	5.8942×10^{-5}	5.4513×10^{-5}
b	2.9875×10^{-9}	1.5418×10^{-8}	1.7063×10^{-9}	1.6558×10^{-9}	1.5313×10^{-9}
c	-7.5702×10^{-12}	-7.9895×10^{-12}	-4.3241×10^{-12}	-3.0045×10^{-12}	-2.7786×10^{-12}
d	-1.1635×10^{-15}	-1.8328×10^{-15}	-6.6462×10^{-16}	-5.2983×10^{-16}	-4.8997×10^{-16}

While this scheme proposes two types of definition, the narrower 'conservative habitable zone' and wider 'optimistic habitable zone', I chose to use the wider 'optimistic habitable zone' definition because the wider definition of the habitable zone is more comprehensive as it entails the potential factors of water and CO2 clouds on the planet [11]. The wider optimistic habitable zone definition is bounded by the 'Recent Venus' and 'Early Mars' limits whereas the narrower conservative habitable zone definition is bounded by the 'Moist Greenhouse' and 'Maximum Greenhouse' limits. Thus, we will only need data in the 'Recent Venus' and 'Early Mars' columns.

Once we calculate the habitable zone stellar fluxes (S_{eff}), we can calculate the corresponding HZ distance limits (d) by using the relation [2] where $L/L_\odot$ is the luminosity of the star compared to the Sun, which can be calculated by [3] where R is the star's radius, $R_\odot$ is the Sun's radius, equal to 695700 km, and $T_{eff\odot}$ is the temperature of the Sun, equal to 5778 K [12].

Using the HZ distance limits (d) obtained with coefficients that correspond to 'Recent Venus' and 'Early Mars', we can calculate the boundary of optimistic habitable zones. If the exoplanet's separation (distance) from the star, defined by the orbital semi-major axis, is within this habitable-zone boundary, then that exoplanet is classified as habitable with a Boolean value of True. If not, meaning that the exoplanet's semi-major axis is either greater or smaller than the maximum limit or the minimum limit of the boundary, respectively, the exoplanet is classified as not habitable with a Boolean value of False.

4. Exoplanet Classification Scheme

The exoplanet classification scheme proposed by Dr. Ravi Kumar Kopparapu et al. in their paper "Exoplanet Classification and Yield Estimates for Direct Imaging Missions" follows a more simple relation, based on chemical species' condensation sequences in planetary atmospheres. Their classification scheme can be summarized by the table below [5]:

III. Exoplanet Classification Scheme

Planet Type (Stellar Flux Range) [Earth Flux]	Planet Radius [Earth Radius]
Hot rocky (182 - 1.0)	0.5-1.0
Warm rocky (1.0 - 0.28)	0.5-1.0
Cold rocky (0.28 - 0.0035)	0.5-1.0
Hot super-Earths (187 - 1.12)	1.0-1.75
Warm super-Earths (1.12 - 0.30)	1.0-1.75
Cold super-Earths (0.30 - 0.0030)	1.0-1.75
Hot sub-Neptune (188 - 1.15)	1.75-3.5
Warm sub-Neptune (1.15 - 0.32)	1.75-3.5
Cold sub-Neptune (0.32 - 0.0030)	1.75-3.5
Hot sub-Jovian (220 - 1.65)	3.5-6.0
Warm sub-Jovian (1.65 - 0.45)	3.5-6.0
Cold sub-Jovian (0.45 - 0.0030)	3.5-6.0
Hot Jovian (220 - 1.65)	6.0-14.3
Warm Jovian (1.65 - 0.40)	6.0-14.3
Cold Jovian (0.40 - 0.0025)	6.0-14.3

Results

Using the Python packages, variables, and schemes mentioned in the previous section, I analyzed the data and produced two types of graphs - bar graph and scatter plot. Bar graphs show the frequency of different demographical categories, and scatter plots show the trends that the data of 4434 confirmed exoplanets follow. I categorized each plot with a distinct alphanumerical index, as shown below. The plots are summarized in the appendix.

A. Bar Graph

1. Classification Frequency (*A-1.1*)
    - Rocky Classification Frequency (*A-2.2*)
    - Super Earths Classification Frequency (*A-.3*)
    - Sub-Neptune Classification Frequency (*A-3.4*)
    - Sub-Jovian Classification Frequency (*A-3.5*)
    - Jovian Classification Frequency (*A-3.6*)

2. Habitability Frequency (*A-2.1*)
   - Habitable Classification Frequency (*A-2.2*)
   - Habitable Classification Percentage Frequency (*A-2.3*)
   - Detection Method Frequency (*A-3*)

B. Scatter Plot

1. Exoplanet Characterization Analysis
   - *Orbit Semi-Major Axis vs Insolation Radius vs Stellar Effective Temperature (B-1.1)*
   - *(Orbit Semi-Major Axis)^3 vs (Orbital Period)^2 (B-1.2)*
   - *Mass Radius Ratio vs Orbit Semi-Major Axis vs Classification (B-1.3)*
   - *Mass Radius Ratio vs Flux vs Classification (B-1.4)*
   - *Planet Mass vs Planet Radius vs Classification (B-1.5)*

2. Habitability Analysis
   - *Planet Mass vs Planet Radius vs Habitability (B-2.1)*
   - *Mass Radius Ratio vs Flux vs Habitability (B-2.2)*

3. Detection Method Analysis
   - *Orbit Period vs Planet Radius vs Detection Method (B-3.1)*
   - *Orbit Period vs Planet Radius vs Detection Method (B-3.2)*
   - *Skymap of Confirmed Exoplanets (RA vs Dec vs Detection Method) (B-3.3)*

Materials and Methods

With the following plots, I surveyed currently confirmed exoplanets in three different perspectives: characterization and classification, detection method, habitability of confirmed exoplanets.

There are four variables that characterize exoplanets (orbital period, orbit semi-major axis, planet radius, and planet mass), three variables that characterize the host star (insolation flux, stellar effective temperature, and stellar radius), and two variables that indicate the location of the exoplanet (right ascension and declination).

1. Exoplanet Characterization Analysis

First, I analyzed the characterization and classification of currently confirmed exoplanets with bar graphs *A-1.1 to A-1.6* and scatter plots *B-1.1 to B-1.5*. The bar graph A-1.1 shows the number of categories that each exoplanet is assigned to based on the classification scheme [5]. According to this graph, Sub-Neptunes are the most confirmed exoplanet type with 1554, exoplanets followed by Jovians with 1202, Super Earths with 944, Sub-Jovians with 328, and Rocky with 164. This is quite similar to NASA's manual classification, which has classified exoplanets into 1497 Neptune-like, 1403 Gas Giant, 1364 Super Earths, 165 terrestrial, and 5 unknown [3]. Unlike typical exoplanet classifications, [5] this classification scheme also proposed classification by flux to further classify into 'hot', 'warm', and 'cold' exoplanets of each 5 category. The trend that more bigger exoplanets are frequent in the dataset shows the sensitivity of our current telescopes on bigger exoplanets. Since not all exoplanets have flux data, there are exoplanets that cannot be further classified, marked as 'Flux N/A' in the graphs. A huge number of 'Flux N/A' shows the limitations of our current telescopes. *Figures A-1.2 to A-1.6* all conform to the agreement that hot exoplanets dominate all of rocky, super earth, sub-neptune, sub-jovian, and jovian exoplanets by far. Since hot exoplanets are likely to reside closer to the host star, this trend implies that our current telescopes are more keen and sensitive to detecting exoplanets orbiting closer to the host star.

The scatter plots can show a more unique trend for the characterization of exoplanets. *Figure B-1.1* shows the correlation between flux, orbital semi-major axis, and stellar effective temperature. This figure is plotted in the idea that stellar flux on the planet is dependent on the temperature of the host star and the separation between host star and the planet. Specifically, the insolation flux should be greater with greater stellar effective temperature and shorter semi-major axis, as the insolation flux is computed by the inverse square law [13] [4] where a is the semi-major axis and the luminosity $L/L_\odot$ is computed by equation (3), which is proportionally related to the stellar effective temperature by a power of 4. *Figure B-1.1* plot clearly confirms this equation as exoplanets at the top left (higher stellar effective temperature and shorter orbit semi-major axis) are marked by higher insolation flux (color yellow).

Figure B-1.2 shows the correlation between (Orbit Semi-Major Axis)^3 and (Orbital Period)^2 to test Kepler's third law on the actual confirmed exoplanet dataset. The Kepler's third law states that (Orbit Semi-Major Axis)^3 is directly proportional to (Orbital Period)^2 by the equation [14] where *G* is Newton's Gravitational Constant, *P* is the orbital period, *M* is the mass of the star, *m* is the mass of the planet, and a is the orbital radius (orbit semi-major axis). *Figure B-1.2* confirms this law by showing a linear trendline with a R2 value of 0.9967, which is very close to 1. Based on this observation, we can conclude that the trend from the orbit semi-major axis can also tell the trend from the orbital period, and vice versa. One outlier from this plot at around 17 (Orbit Semi-Major Axis)^3 and around 108 (Orbital Period)^2 is CFBDSIR J145829+101343 b and may be worth a future investigation.

Figure B-1.3 test the classification of exoplanets on two of the most important factors of exoplanet characteristics - planet mass and planet radius [15]. As the three discrete, yet connected, straight lines suggest, there is indeed a clear, linear, proportional correlation between planet mass and planet radius [16]: the greater the planet mass is, the longer the planet radius is. Rocky planets (blue) correlate with lower planet mass and planet radius, followed in order of super earth (red), sub neptune (yellow), sub jovian (green), and jovian (purple). Also, note that at around planet mass of 10^2, the positive straight line stops, and the negative line starts. This plot also convinces us that there could be more a refined classification scheme based on exoplanet mass, instead of that on radius that [5] uses, or the ratio between mass and radius [17] [18].

Based on the idea about planet mass radius ratio, I plotted *figures B-1.4* and *B-1.5* to test classification in relation to insolation flux and orbit semi-major axis. The mass radius ratio at the x-axis clearly proves that exoplanets can be classified into a certain range of mass radius ratio. Rocky exoplanets, clustered on the leftist side, can be categorized with lower mass radius ratio, whereas Jovian exoplanets, clustered on the far right side, can be categorized with higher mass radius ratio. Super Earth, sub-Neptune, and sub-Jovian exoplanets can also be classified with mass radius ratio - from lower to higher mass radius ratio. While there is no large trend with insolation flux, the exoplanets classified with larger mass radius ratio tend to have greater separation (orbit semi-major axis) from the host star.

2. Exoplanet Habitability Analysis

Next, I analyzed the habitability of currently confirmed exoplanets with bar graphs *A-2.1*, *A-2.2, A-2.3* and scatter plots *B-2.1* and *B-2.2*. Based on the habitable zone scheme [6], I calculated the number of exoplanets within the habitable zone and made a bar graph *A-2.1*. The bar graph *A-2* shows that only around 5 percent of currently confirmed exoplanets (231 out of 4434) are habitable according to the scheme proposed in "Habitable Zones Around Main-Sequence Stars: New Estimates". Then, I classified these habitable exoplanets into rocky, super earth, sub neptune, sub jovian, and jovian exoplanets based on the classification scheme [5] and generated bar graphs *A-2.2* and *A-2.3*. These two graphs show that jovian is the most common habitable exoplanet classification and that rocky is the least common habitable exoplanet classification not only by number but also by percentage. I plotted *B-2.1* and *B-2.2* to show trends within habitable exoplanets. It turns out that the habitable exoplanets almost perfectly follow the mass radius ratio lines, identified in *figure B-1.3* and have relatively lower insolation flux, from 0.1 to 10 Earth Flux. This range is closer to the Earth's own flux, 1 Earth Flux, indicating that the right amount of flux is a significant factor of habitability of the exoplanets.

3. Exoplanet Detection Method Analysis

Finally, I analyzed the detection method with bar graph *A-3* and *B-1 to B-3* plots. First, bar graph *A-3* shows that transit method detected the most exoplanets, thanks to transit missions like Kepler Space Telescope and Transiting Exoplanet Survey Satellite (TESS), followed by radial velocity, microlensing, imaging, and pulsar timing.

On the other hand, plots *B-3.1* to *B-3.3* show the sensitivity of the detection methods. Plot *B-3.1* and *B-3.2* show a large cluster of transit method plots (orange) concentrated at a relatively small planet mass (1 ~ 30 Earth Mass), planet radius (1~5 Earth Radius) and orbital period (1 ~ 10^2 Earth Days). This conforms with the sensitivity of transit method; since transit method detects the change in starlight blocked by exoplanets, and exoplanets orbiting closer to the host star are more likely to block the starlight by greater area, it is more likely to detect exoplanets with smaller orbit semi-major axis. Since orbit semi-major axis is related to orbital period by Kepler's third law, as the plot *B-1.2* shows, transit method is more sensitive to exoplanets with relatively shorter orbital periods [1]. However, this intuition fails for the cluster of shorter planet-radius exoplanets: the transit method should be more sensitive to exoplanets with greater radius because bigger exoplanets will block more of the starlight. This suggests that more smaller exoplanets reside closer to the host star. The trend with planet mass directly follows the trend with planet radius by our observation of mass radius ratio from *figure B-1.3*. Another notable trend is the cluster of radial velocity plots (blue) concentrated at a relatively longer planet radius (10 to 15 Earth Radius) and heavier planet mass (102 to 104 Earth Mass). This shows the sensitivity of radial velocity exoplanets on heavier exoplanets, as radial velocity is detecting the gravitational pull between host star and the planet and the gravitational pull between host star and planet proportionally depends on their mass [1]. The trend with planet radius directly follows the trend with planet radius by our observation of mass radius ratio from *figure B-1.3*. Lastly, we can note that the direct imaging plots (green) are concentrated at higher planet radius (12 to 15 Earth Radius) and longer orbital period (10^4 to 10^5 Earth Days). This is because our current direct imaging method is not sensitive enough: it is limited to exoplanets around nearby stars with very large radius and longer separation from stars [1]. Trends for other detection methods are not that clear due to lack of exoplanets confirmed by those methods.

Plot *B-3.3*, which is a skymap of all the confirmed exoplanets, shows a large concentration of yellow-colored exoplanets around 300 degree RA and 50 degree DEC, which coincides with the Kepler Space transit Telescope field of view [19], as well as a large concentration of red-colored exoplanets around 275 degree RA and -30 degree DEC. On the other hand, radial velocity and imaging seem to detect exoplanets regardless of their location.

Caveat and Future Research

Since the study of exoplanets is a relatively new field, there exist some caveats and limitations in this paper. The three most notable caveats are the limitation of data, disregard for the uncertainties, and the complex nature of exoplanets and their habitability. Along with the identification of these caveats, I also propose some future works that can remedy these caveats.

1. Limitation of Data

Since the search for exoplanets has been underway for only around two decades, the current datasets are greatly limited in three ways. Firstly, most of the 4434 confirmed exoplanets are in a relatively small, concentrated region of the Milky Way galaxy because that is as far as current telescopes have been able to probe. These 4434 exoplanets represent less than a 0.000004434 percent of the planets within our Milky Way galaxy as it has been shown that there are at least 100 billion exoplanets in the Milky Way galaxy [20]. Therefore, the survey of demographics in this paper only applies to these confirmed 4434 exoplanets, so trends and analysis found in the plots may be completely irrelevant as more exoplanets are detected and confirmed. Secondly, some of the data were missing for some exoplanets. For example, more than 1600 exoplanets missed their insolation flux data, so only around 2800 exoplanets were further classified into "Hot", "Warm", and "Cold". In order to obtain more complete demographics of exoplanets, future missions can be undertaken to find missing data of currently confirmed exoplanets by detecting them using different detection methods or by detecting them with more powerful, overarching telescopes such as the James Webb Telescope, which will be launched in November this year [21]. Moreover, future research can be done by manipulating more variables than the 10 variables I used, as well as with data of candidate exoplanets, which include almost twice as many exoplanets (7,472) as the confirmed exoplanets [3]. Lastly, some of the data may be underestimated because most indirect detection methods, especially radial velocity, are heavily dependent on the orientation of the planetary system. These orientation-dependent detection methods only show the component of the velocity in the observer's direction, leading to the underestimation of data. Indeed, the mass from the NASA Exoplanet Archive catalog show the minimum value, and NASA acknowledges this caveat [22] [23]. Future work can be done to improve the high-contrast direct imaging technology to find more accurate exoplanet data because direct imaging is independent of the orientation of the planetary systems [24].

2. Disregard for Uncertainties

Another caveat of my research is that I ignored the uncertainties when plotting graphs in order to simplify the plots. Since the main goal of this paper is to survey the general demographical trends of currently confirmed planets, uncertainties are negligible, especially because uncertainties are really small compared to the actual values. However, future surveys may be done to include these uncertainties to provide a more complete representation of the data.

3. Complex Nature of Exoplanet and Habitability

The last caveat to note is the complexity of exoplanets and habitability. There is no one official exoplanet classification scheme and habitability scheme. Instead, each exoplanet has to be individually examined to find out the information about its classification and habitability. Although this paper follows two of the most respected schemes, they may not represent exoplanets' actual classification and habitability. In particular, the definition of current habitable zones is made in respect to the condition of Earth. However, different creatures may live in conditions different from Earth such as the dependency on CO^2 and H_2O [25]. Moreover, different planetary systems, such as pulsar [26] and binary [27], may require a completely different habitable zone scheme due to their completely different environments.

Therefore, while this paper employs habitable zone schemes cited by more than 1000 other papers, and exoplanets deemed habitable by proposed habitability schemes may be worth further probes, it is by no means the perfect formula to determine if a planet hosts life or not. Future research can be done to apply different classification and habitable zone schemes based on the data and code presented in my paper.

Conclusion

The NASA Exoplanet Archive is an effective open catalog that provides data of both confirmed and candidate exoplanets. However, there are some limitations in this catalog such as lacking information about rough classification and habitability of the exoplanet. Throughout this research, I have analyzed 10 variables of 4434 confirmed exoplanets from the NASA Exoplanet Archive by generating 20 different plots and updated the catalog by including classification and habitability flags. I analyzed the data in three different ways, by examining the characterization and classification of exoplanets, habitability of exoplanets, and sensitivity of exoplanet detection methods.

I employed four different open Python packages – namely, Pandas, NumPy, Matplotlib, and Astropy – in reading, writing, analyzing data and generating plots. The analysis methodology is implementation-friendly, as all the codes, plots, updated catalogs, and dataset are publicly uploaded on my GitHub page for readers to repeat my analysis whenever new confirmed exoplanets are added on the NASA Exoplanet Archive.

The main take-aways from the analysis of the dataset include the confirmation of flux-stellar temperature-separation relation, Kepler's third law, mass radius ratio and correlation, habitability on 'hot' and Jovian classified exoplanets, and sensitivity of transit, radial velocity, and direct imaging detection methods.

As more exoplanets are detected with more extensive missions, such as the James Webb Space Telescope and the Roman Telescope [28], many more exoplanets with more diverse environments will be detected, getting humans closer to the goal of finding other planetary life. Therefore, there will be more need for sophisticated, comprehensive habitability and classification schemes. I end this paper by proposing a possible exoplanet classification scheme that uses a weighted average method for future purposes, when there are more advanced, comprehensive telescopes to obtain more complete exoplanet metadata. Although some scientists argue that classification should be based on easily detectable characteristics of exoplanets and based on the fewest possible criteria [29], I noticed that there are many different important factors and correlations in categorizing exoplanets, such as planet mass, planet radius, insolation flux, and semi-major axis. Some combination of weighted average of these variables may lead to a more sophisticated classification scheme because the world of exoplanets is so complex that they cannot be categorized just by a single variable; variables that are deemed more important in the classification of exoplanets may take greater weight than do other variables. Although current detection methods are quite limited to what types of variables they can accurately detect, such as the limitations of accurate planet mass data for transit method and the limitation of accurate planet radius data for radial velocity method [14], more advanced future telescopes will allow us to find more comprehensive data of exoplanets and to obtain more complete knowledge on the world of exoplanets.

Acknowledgments

I thank Dr. Dan Sirbu, a Research Scientist and Engineer at BAERI, NASA Ames Research Center, for being my mentor on this project. The completion of this project would not have been possible without his guidance, review, and feedback.

I also acknowledge that this research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

References

1. Seager, S. (Ed.). (2011). Exoplanets. Tucson, AZ: The University of Arizona Press.

2. National Aeronautics and Space Administration. (2019, June 13). Historic Timeline. Exoplanet Exploration: Planets Beyond Our Solar System. Retrieved July 11, 2021, from https://exoplanets.nasa.gov/alien-worlds/historic-timeline/#first-exoplanets-discovered

3. National Aeronautics and Space Administration. (2021, July 17). *Discoveries Dashboard*. Exoplanet Exploration: Planets Beyond Our Solar System. https://exoplanets.nasa.gov/discovery/discoveries-dashboard/

4. National Aeronautics and Space Administration. (2021, April 28). *In Depth | Confirmed vs. Candidate*. Exoplanet Exploration: Planets Beyond Our Solar System. https://exoplanets.nasa.gov/what-is-an-exoplanet/in-depth/

5. Kopparapu, R. et al. (2018). Exoplanet Classification and Yield Estimates for Direct Imaging Missions. The Astrophysical Journal, 856(2). 10.3847/1538-4357/aab205

6. Kopparapu, R. et al. (2013). Habitable Zones Around Main-Sequence Stars: New Estimates. The Astrophysical Journal, 765(2). 10.1088/0004-637X/765/2/131

7. NASA Exoplanet Archive. (2021). Planetary Systems Composite Data [Data set]. Retrieved from https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=PSCompPars. 10.26133/NEA13

8. NASA Exoplanet Archive. (2020, August 4). Extended Planet Data Table Data Column Definitions. NASA Exoplanet Archive. Retrieved July 11, 2021, from https://exoplanetarchive.ipac.caltech.edu/docs/API_exomultpars_columns.html

9. Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Günther, H. M., . . . Streicher, O. (2013). Astropy: A community Python package for astronomy. Astronomy & Astrophysics, 558(33). https://doi.org/10.1051/0004-6361/201322068

10. Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., Ardelean, C., . . . Zabalza, V. (2018). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package.

The Astronomical Journal, 156(3). https://doi.org/10.3847/1538-3881/aabc4f

11. Planetary Habitability Laboratory. (2013, March 24). Summary of the Limits of the New Habitable Zone. Planetary Habitability Laboratory @ UPR Arecibo. https://phl.upr.edu/library/notes/summarylimitsofthenewhabitablezone

12. Szyk, B. (2018, January 31). Luminosity Calculator. Omni Calculator. https://www.omnicalculator.com/physics/luminosity

13. NASA Exoplanet Archive. (2016, November 2). *How Predicted Observables for Exoplanets are Calculated*. https://exoplanetarchive.ipac.caltech.edu/docs/poet_calculations.html#insol_flux

14. Simon Fraser University. (n.d.). *Calculating exoplanet properties - President's Dream Colloquium - Simon Fraser University*. Retrieved July 19, 2021, from https://www.sfu.ca/colloquium/PDC_Top/astrobiology/discovering-exoplanets/calculating-exoplanet-properties.html

15. Ulmer-Moll, S., Santos, N. C., Figueira, P., Brinchmann, J., & Faria, J. P. (2019). Beyond the exoplanet mass-radius relation. *Astronomy & Astrophysics*, 630, A135. https://doi.org/10.1051/0004-6361/201936049

16. Sotin, C., Grasset, O., & Mocquet, A. (2007). Mass–radius curve for extrasolar Earth-like planets and ocean planets. *Icarus, 191*(1), 337–351. https://doi.org/10.1016/j.icarus.2007.04.006

17. Seager, S., Kuchner, M., Hier□Majumder, C. A., & Militzer, B. (2007). Mass□Radius Relationships for Solid Exoplanets. *The Astrophysical Journal*, 669(2), 1279–1297. https://doi.org/10.1086/521346

18. Swift, D. C., Eggert, J. H., Hicks, D. G., Hamel, S., Caspersen, K., Schwegler, E., Collins, G. W., Nettelmann, N., & Ackland, G. J. (2011). MASS-RADIUS RELATIONSHIPS FOR EXOPLANETS. The Astrophysical Journal, 744(1), 59. https://doi.org/10.1088/0004-637x/744/1/59

19. Batalha, N. M., Borucki, W. J., Koch, D. G., Bryson, S. T., Haas, M., Brown, T. M., Caldwell, D. A., Hall, J. R., Gilliland, R. L., Latham, D. W., Meibom, S., & Monet, D. G. (2010). SELECTION, PRIORITIZATION, AND CHARACTERISTICS OF KEPLER TARGET STARS. *The Astrophysical Journal*, 713(2), L109–L114. https://doi.org/10.1088/2041-8205/713/2/l109

20. National Aeronautics and Space Administration. (2013, June 7). Billions and Billions of Planets. NASA. https://www.nasa.gov/mission_pages/kepler/news/kepler20130103.html.

21. Kalirai, J. (2018). Scientific discovery with the James Webb Space Telescope. *Contemporary Physics*, 59(3), 251–290. https://doi.org/10.1080/00107514.2018.1467648

22. Otegi, J. F., Bouchy, F., & Helled, R. (2020). Revisited mass-radius relations for exoplanets below 120 M⊕. *Astronomy & Astrophysics*, 634. https://doi.org/10.1051/0004-6361/201936482

23. Changeat, Q., Keyte, L., Waldmann, I. P., & Tinetti, G. (2020). Impact of Planetary Mass Uncertainties on Exoplanet Atmospheric Retrievals. *The Astrophysical Journal*, 896(2). https://doi.org/10.3847/1538-4357/ab8f8b

24. Jovanovic, N., Guyon, O., N'Diaye, M., Galicher, R., Sirbu, D., Kenworthy, M. A., Ygouf, M., Baudoz, P., Kühn, J., Huby, E., Wilby, M. J., Por, E. H., Haffert, S. Y., Keller, C. U., Snik, F., Miller, K. L., Wallace, J. K., Beaulieu, M., Cady, E., . . . Riggs, A. J. E. (2018). Review of high-contrast imaging systems for current and future ground-based and space-based telescopes: Part II. Common path wavefront sensing/control and coherent differential imaging. Adaptive Optics Systems VI, 107031. https://doi.org/10.1117/12.2314260

25. Ramirez, R. (2018). A More Comprehensive Habitable Zone for Finding Life on Other Planets. Geosciences, 8(8), 280. https://doi.org/10.3390/geosciences8080280

26. Patruno, A., & Kama, M. (2017). Neutron star planets: Atmospheric processes and irradiation. *Astronomy & Astrophysics*, 608. https://doi.org/10.1051/0004-6361/201731102

27. Whitmire, D. P., Matese, J. J., Criswell, L., & Mikkola, S. (1998). Habitable Planet Formation in Binary Star Systems. Icarus, 132(1), 196–203. *https://doi.org/10.1006/icar.1998.5900*

28. National Aeronautics and Space Administration. (2020, November 10). Mission. Exoplanet Exploration: Planets Beyond Our Solar System. https://exoplanets.nasa.gov/discovery/missions

29. Stern, S. A., & Levison, H. F. (2002). Regarding the Criteria for Planethood and Proposed Planetary Classification Schemes. *Highlights of Astronomy*, 12, 205–213. https://doi.org/10.1017/s1539299600013289