Exoplanet size in proportion to host star size & metallicity

Michael Bennett

2022-12-09

Hypothesis and Rationale

Mechanistic understanding of the planetary formation process is important for astronomers search of exoplanets. If the search is to be continued with progress, learning planetary size in relation to their host star is highly relevant. In addition, if planets are targeted with potential of having an atmosphere and life, knowing beforehand their star system chemical composition is essentially important.

Over 5,000 exoplanets have been cataloged to date (Brennan, 2022) and missions like Characterizing Exoplanet Satellite (CHEOPS) are currently tasked with the discovery of exoplanets (ESA, 2020). Many of these planets are in what is known as the habitable zone or the distance from the host star where liquid water can exist on the planet. Many are Jupiter sized and many are smaller than Earth (**Table 1**). Small planets are particularly the most frequent found and are mostly dense with iron rock compositions (Batalha, 2014).

Beginning in 1997 a trend correlating star metallicity and giant planets orbiting around them had caught the interest of research groups (Gonzalez, 2006). Metallicity is the abundance of elements heavier than hydrogen and helium in an object for example, a planet or a star. The element measured at the highest abundance in metallicity is iron. Chemical composition and mass are important through the life of the star as well as the planetary formation of the star system (Johnson, 2010). The likelihood of a star having a planet depends on its iron content (Johnson, 2010). Studies have found that giant planets increase in size with increasing star metallicity (Marcy, 2014). Other correlations are that increasing planetary core mass is associated with increasing star metallicity (Johnson, 2010).

Given these trends in stars and their orbiting planets I looked for evidence of host star mass and metallicity in relation to planetary radius with the hope of finding clues to planetary formation. Using this information astronomers looking to find life on exoplanets could be better equipped with an idea of what star systems to target. Stars with low metallicity will likely have planets with less dense iron rock cores and be the size of asteroids while stars with high metallicity will have planets Earth size and larger, dense iron cores and better potential for harboring life.

Investigating the idea of host star size, metallicity concentration and planetary properties, exoplanets were first found using doppler transit methods. While a planet transits in front of its host star our telescopes can see the eclipsing path of the planet with the star behind the transiting planet. As the light from the star passes through the planets surroundings, measurements can be made such as size, atmospheric composition, orbital distance from the star, and orbital period. With these measurements other measurements can be made for example, core and mass density of the planet, rotation, eccentricity, and others. Although, analysis was made on most of these variables just the variables relevant to this project are included here. There wasn't any fieldwork or lab work for this project as everything could be sourced from internet astronomical data websites.

Given this research, star metallicity, star mass, and planetary radius are correlated throughout the universe. Thus, the hypothesis are as follows, **H1** as host star radius, mass, and metallicity increases so does the planetary radius of planets orbiting around the star. **H2** as host star mass increases so does the planetary mass of planets orbiting around the star. Overall, it is expected that larger iron abundant stars will have larger planets orbiting around them.

Statistical Approach

Analysis

The Open Exoplanet Catalogue is a database of all discovered extra-solar planets. The database is licensed under an MIT license. Data analysis was completed with R statistical software. To investigate planet size with increasing star size and metallicity a linear model was built and tested in RStudio.

Constants and standards used in the database are first and foremost, stars with a higher metallicity than the Sun have a positive value, whereas those with a higher abundance of hydrogen than the Sun have a corresponding negative value. Metallicity is the amount of iron and hydrogen present, determined by analyzing absorption lines in a star spectrum, with the Sun being the relative comparative standard. The Suns radius is 695,700 km while its mass is $2 * 10^{30}$ kg, Jupiter's radius is 71,492 km while its mass is $1.9 * 10^{27}$ kg.

Given that planetary radius (response variable) was a non-normally distributed Gaussian (**Figure 1**) and the relationship between planetary radius and host star radius was linear, I used a linear regression model. Minor transformations were made to the data including the removal of all rows with NA values and filtering out extreme outliers. These transformations avoided bias because the stars and exoplanets are distributed in various random regions of space not selected in any uniform way whatsoever. Further, the linear model avoids pseudoreplication with use of a random population. This was tested with the rainbow test in the *lmtest* package ($p > 0.05$). The fit of a portion of the data was not significantly different than the fit of the whole database.

The package *ggplot2* was used for additional histograms and boxplots to view the medians of all variables (**Figures 2 & 3**). All the variables associated with stars have about the same variance while planetary mass has some larger variance than planetary radius. Although planetary radius looks normal in the histogram it turns out to be non-normal after testing. Shapiro-Wilk test for normality was used on the planetary radius variable rendering non-normality ($p < 0.05$). Planetary mass as well as star radius are skewed to the right, and star metallicity and temperature are skewed left. Using the *psych* package for scatter plot matrices (SPLOM), histograms, and correlations (**Figure 4**) was helpful for finding correlations between star mass, radius, and temperature. Although it appears planetary radius has a relatively high correlation with star mass, radius, and temperature it is later discovered that temperature has less to do with planetary radius than the other variables.

[Equation 1]

Planetary Radius = $\alpha + \beta_1$(Star Radius) + β_2(Star Mass) + β_3(Star Metallicity) + ε

[Equation 2]

Planetary Mass = $\alpha + \beta_1$(Star Mass) + ε

The model was built using the best independent variable first and then adding variables onto the model if significance allowed. The model was chosen after performing Akaike Information Criterion (AIC) on different models. Using the model with the lowest AIC value (**Table 3**) while including all important variables to test the hypothesis. To model planetary radius as a function of star size and metallicity to test the hypothesis, the linear model was analyzed with the *performance* package (**Figure 5**). Running a t-test of coefficients the p-values were all significant for this model. Attempts to add temperature to every possible model layout of important variables to this analysis always resulted in a non-significant p-value therefore temperature was not included in the model. Other models were made with interactions for example, star mass and star radius making the overall model less significant from observance of the p-values and higher AIC values.

Model Checks

Linear Model

The **H1** model coefficients and p values are shown in **Table 2**. Within the *lmtest* package, the test for autocorrelation called Durbin-Watson test shows a slight positive autocorrelation in the residuals with a DW value of 1.89 however, later tests prove the residuals to not be autocorrelated. Various diagnostic tests are made visual in the model (**Figure 5**). Of these tests from the *performance* package include an autocorrelation test determining the model not to be autocorrelated ($p > 0.05$). An outlier test finding no outliers. A collinearity test yielding low correlation. A heteroscedasticity test with an output of homoscedastic ($p > 0.05$) explaining the model has similar variance. A normality test providing that residuals appear as normally distributed (p = 0.055). A singularity test with a false result meaning the model is not overfitted.

After testing the model for **H1** the same tests were performed on **H2** and the result is that we fail to reject the null of **H2** (p > 0.05). **Table 3** shows the coefficiants and p values for **H2**. After the non significant result of **H2** I threw out the model and performed additional testing on the model for **H1**. Although this research did not find significant results, further testing on planetary mass and its correlation with host star mass can be planned in the future with an increasing exoplanet database.

Figure 5 Posterior Predictive Check plot shows the predictive model fits the observed data. The Linearity plot or residuals vs. fits plot within **Figure 5** shows that the residuals bounce randomly around the zero line (**Figure 5**). The Homogeneity of Variance plot shows the variance between groups is relatively even (**Figure 5**). The fourth plot in **Figure 5** is the Influential Observations showing no outliers that greatly affect the slope of the regression line. The Collinearity plot in **Figure 5** shows low correlation between the independent variable and the other variables. The last plot in **Figure 5** is the Normality of Residuals and shows the underlying residuals are normally distributed.

The response variable versus fitted plot (**Figure 6**) displays a 67% correlation. This suggests that there is a reasonable linear relationship. The variance of the error terms are equal because the residuals roughly form a horizontal band around the zero line. There are no outliers either because there are no individual residuals standing out from the residuals around the zero line. With the use of the model_performance() function (**Table 4**) an approximate R^2 value of 0.45 describes 45% of variation in planetary radius can be explained by the model (star radius, star mass, and star metallicity).

Results and Discussion

It can be associated that an increase of one solar radius of the star with an increase of 0.289 Jupiter radii of the planet adjusting or controlling for mass and metallicity of the star (**Figure 7**). It can also be associated that an increase of one solar mass of the star with an increase of 0.395 Jupiter radii of the planet adjusting for radius and metallicity of the star (**Figure 7**). As well as a decrease of one-unit solar metallicity of the star with a decrease of 0.2 Jupiter radii of the planet adjusting for mass and radius of the star.

The findings here are the result of over 25 years of monitoring stars. Soon new findings from different observing techniques will join these studies about star size and their orbiting planet size along with metallicity and luminosity. Another piece of interesting information is the correlation of star temperature and star size (**Figure 7**). As star mass and radius increase the temperature of the star increases. This is in accordance with Einstein's famous equation $E = mc^2$, increasing the mass of an object will increase the amount of energy of the object.

Planetary radius appears to significantly decrease with the decrease of star metallicity very slightly. Whereas planetary mass appears to increase with the increase of star metallicity slightly (**Figure 7**). This is because very large planets like Jupiter are known as gas giants and made mostly of hydrogen and helium. Thus, their core has little iron and if the host star has low iron then the planets around the star would also have low iron (Spiegel, 2014). Stars with higher metallicity will have planets with higher abundance of iron in their core orbiting around them (Spiegel, 2014). If Mercury an iron rock core planet was compared to Saturn a gas giant in regards to mass and radius the idea of this could be understood. The mass to radius ratio of Saturn is 9.4 while the mass to radius ratio of Mercury is 0.14. Saturn has low mass relative to its radius while Mercury is very massive relative to its radius.

If astronomers were to use the James Webb Space Telescope to search for planets, the information from this experiment would be helpful to their work. Future research is needed in the search for exoplanets harboring life and learning about the planetary formation process (Johnson, 2010). Research involving newly added exoplanets to the Open Exoplanet Catalog and selection of the database including the most relevant exoplanets. Further research could investigate only gas giants compared with their host stars or only small rock iron planets compared with their host stars.

Nevertheless, additional data analysis is needed for the planetary formation process. Whether it's the disk instability theory or the core accretion theory understanding of how planets form is crucial for the search for life in the universe.

Literature Cited

Batalha, N. M. (2014). Exploring exoplanet populations with NASA's Kepler Mission. *Proceedings of the National Academy of Sciences of the United States of America*, 111(35), 12647–12654. http://www.jstor.org/stable/43043386

Brennan, P., (2022). How many exoplanets are there?. *Exoplanet Exploration*. https://exoplanets.nasa.gov/faq/6/how-many-exoplanets-are-there/

Dong, S., Xie, J.-W., Zhou, J.-L., Zheng, Z., & Luo, A. (2018). LAMOST telescope reveals that Neptunian cousins of hot Jupiters are mostly single offspring of stars that are rich in heavy elements. *Proceedings of the National Academy of Sciences of the United States of America*, 115(2), 266–271. https://www.jstor.org/stable/26506238

ESA. (2020). Exoplanet Mission Timeline. *Exoplanets*. https://sci.esa.int/web/exoplanets/-/60649-exoplanet-mission-timeline

Gonzalez, G. (2006). The Chemical Compositions of Stars with Planets: A Review. *Publications of the Astronomical Society of the Pacific*, 118(849), 1494–1505. https://doi.org/10.1086/509792

Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. (2010). Giant Planet Occurrence in the Stellar Mass-Metallicity Plane. *Publications of the Astronomical Society of the Pacific*, 122(894), 905–915. https://doi.org/10.1086/655775

Laughlin, G. P. (2006). Extrasolar Planetary Systems: Observations of distant worlds are beginning to reveal how panetary systems form and evolve. *American Scientist*, 94(5), 420–429. http://www.jstor.org/stable/27858834

Marcy, G. W., Weiss, L. M., Petigura, E. A., Isaacson, H., Howard, A. W., & Buchhave, L. A. (2014). Occurrence and core-envelope structure of 1-4× Earth-size planets around Sun-like stars. *Proceedings of the National Academy of Sciences of the United States of America*, 111(35), 12655–12660. http://www.jstor.org/stable/43043387

McBride, J., Graham, J. R., Macintosh, B., Beckwith, S. V. W., Marois, C., Poyneer, L. A., & Wiktorowicz, S. J. (2011). Experimental Design for the Gemini Planet Imager. *Publications of the Astronomical Society of the Pacific*, 123(904), 692–708. https://doi.org/10.1086/660733

Santos, N. C., Benz, W., & Mayor, M. (2005). Extrasolar Planets: Constraints for Planet Formation Models. *Science*, 310(5746), 251–255. http://www.jstor.org/stable/3842644

Spiegel, D. S., Fortney, J. J., & Sotin, C. (2014). Structure of exoplanets. *Proceedings of the National Academy of Sciences of the United States of America*, 111(35), 12622–12627. http://www.jstor.org/stable/43043382

Table 1: Constants

Constant Name	Value
Solar radius	$695,700$ km
Jupiter equatorial radius	$71,492$ km
Earth equatorial radius	$6,378$ km
Solar mass	$1.99 * 10^{30}$ kg
Jupiter mass	$1.90 * 10^{27}$ kg

These astronomical constants are used as measurements for other objects in the universe. For example, solar radius is the radius of the Sun and can explain the radius of other stars in the way of another star being two solar radii or four solar radii. The same goes for Jupiter radius and exoplanets. Five Jupiter radii would be the equivalant of an exoplanets radius of Jupiters radius times five.

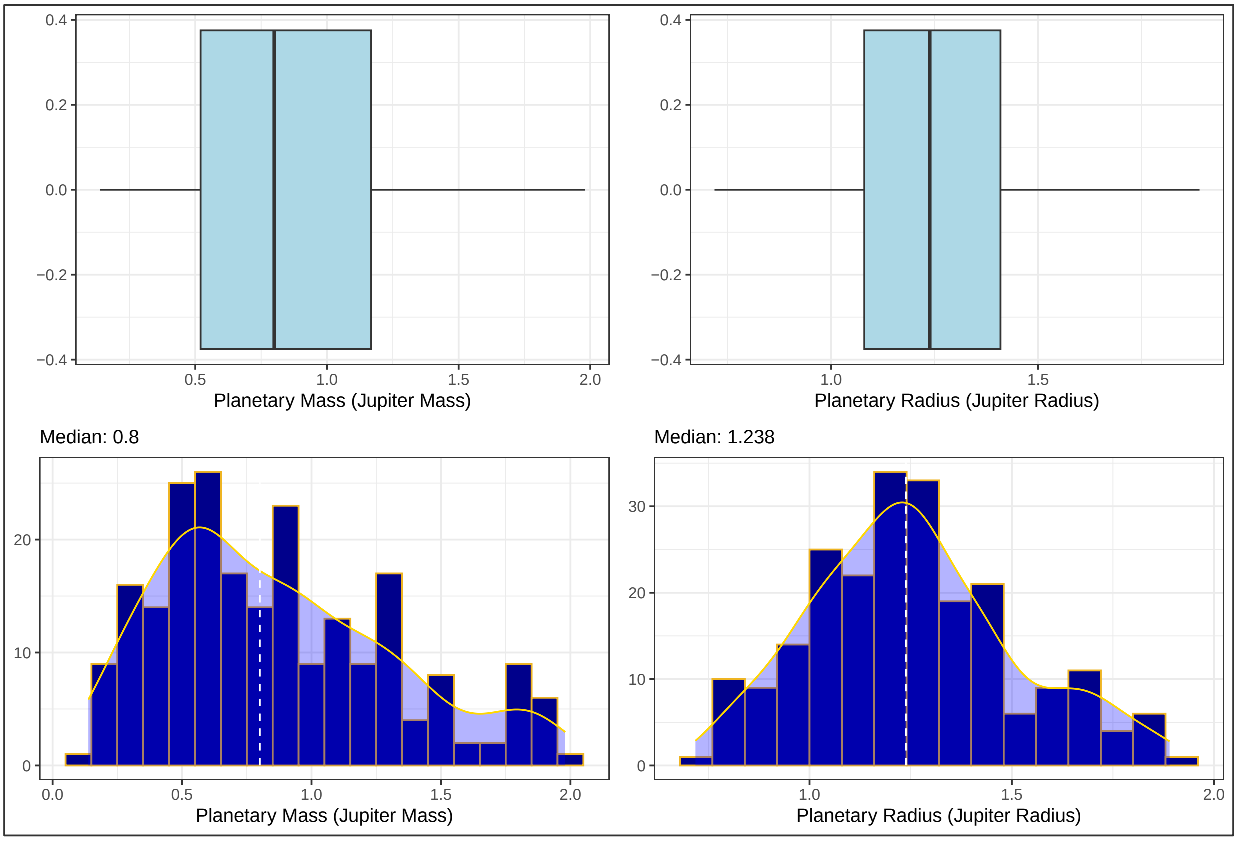


Figure 1

Although all four plots show the median of the planetary variables, the top plots are boxplots and the bottom are histograms. There is a density curve Gaussian line on the histograms. Calculating the binwidth with the equation: binwidth = max(variable) - min(variable) / sqrt(n observations data).

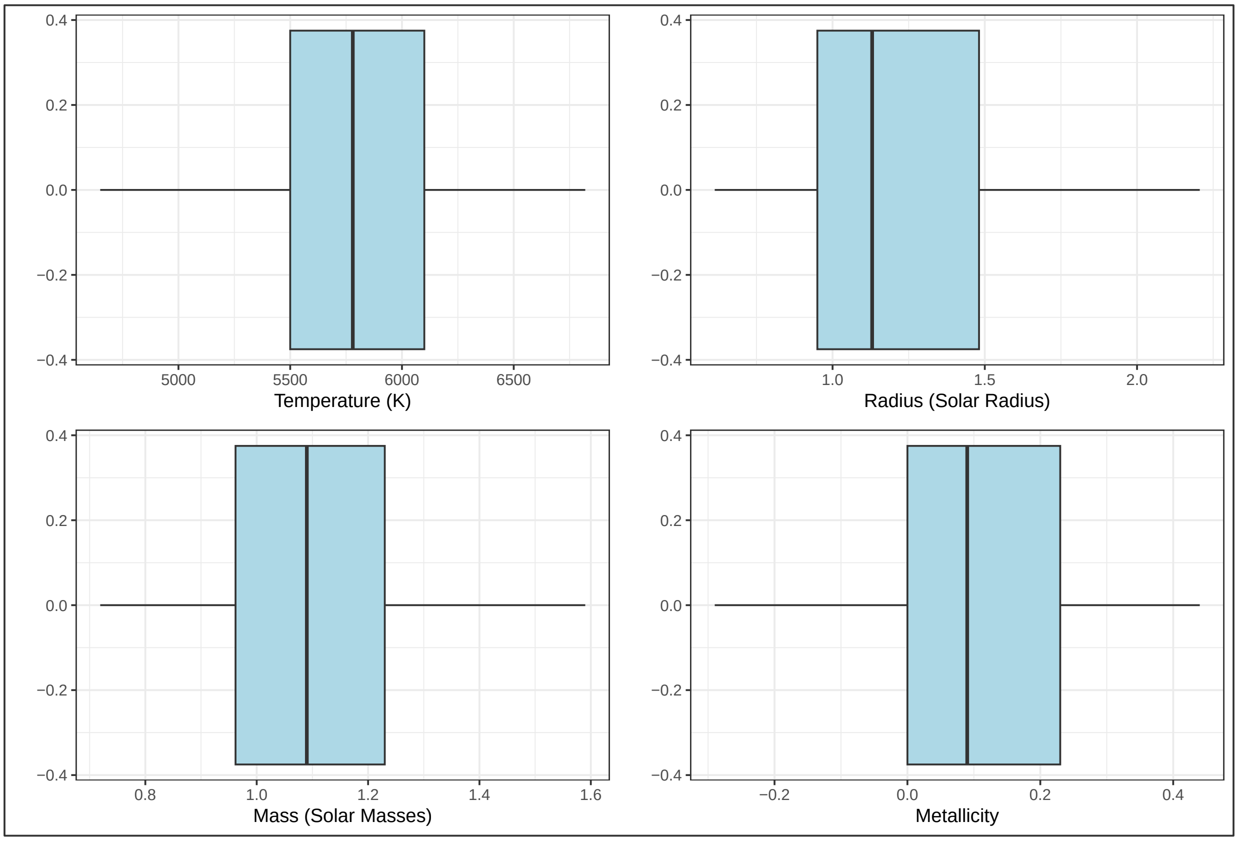


Figure 2

Boxplots showing the medians of the four star variables. Outliers were filtered out so the whiskers are without outlier points on either side. The median values are printed on **Figure 3**.

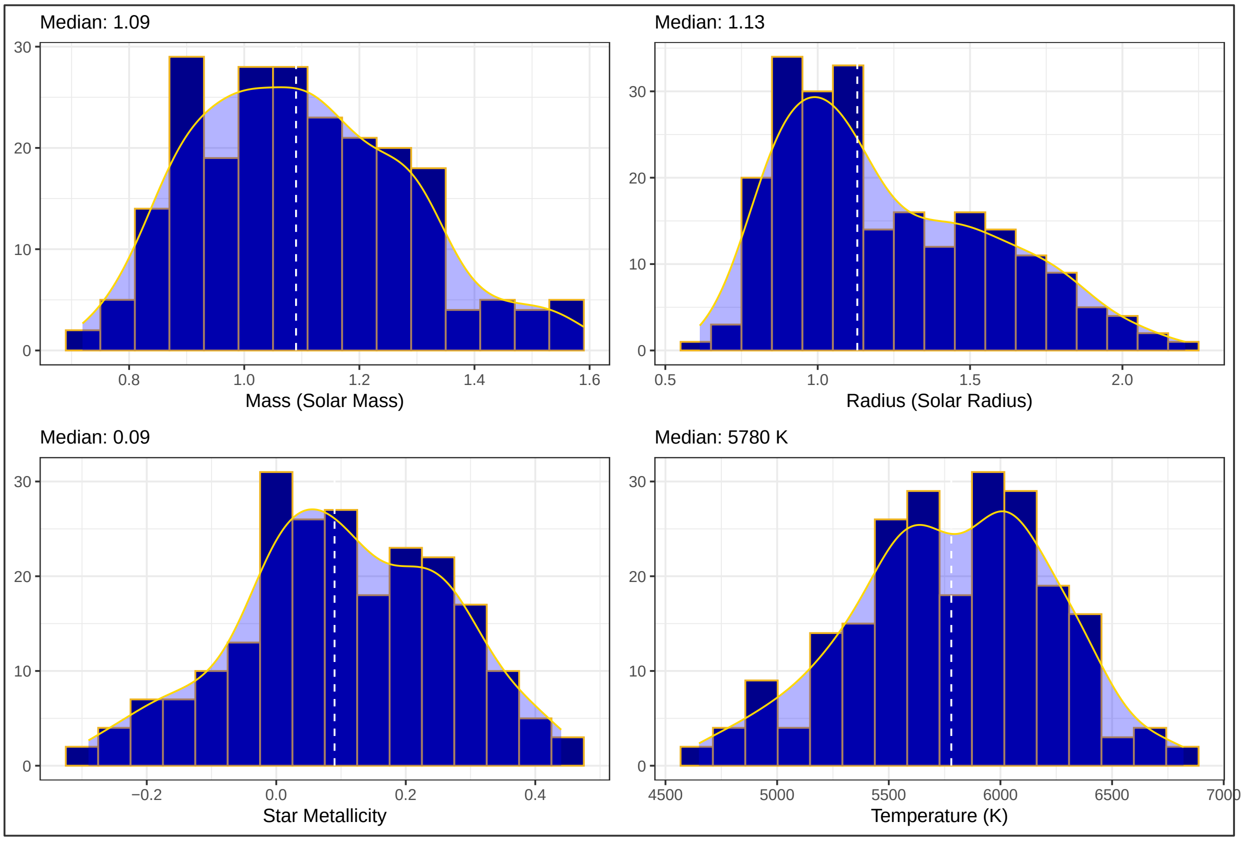
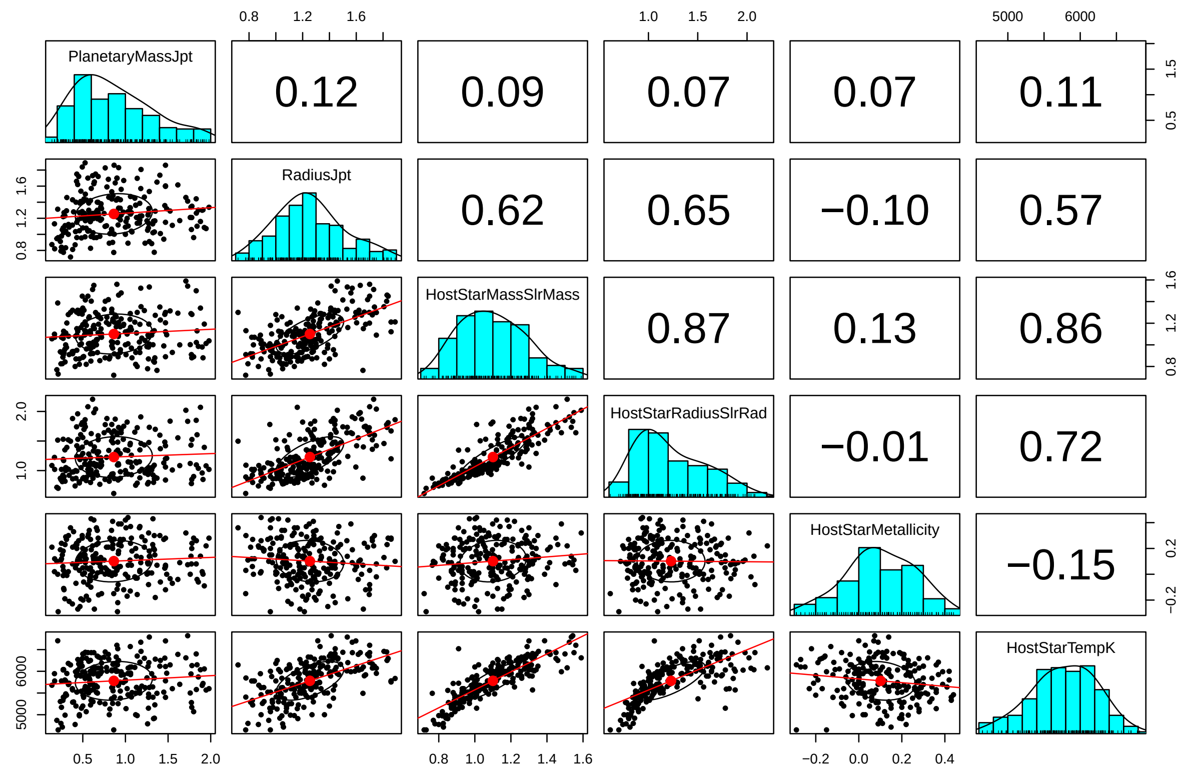


Figure 3

Histograms of star variables mass, radius, metallicity, and temperature. Medians given and shown as a dotted line. Gaussian density curves placed over the bars and bins calculated from the binwidth = max(variable) - min(variable) / sqrt(n observations data) equation.

Figure 4: Correlation Matrix



Matrix of Bivariate scatter plots in the lower left portion, histograms with variable names on the diagonal, and the Pearson correlation above the diagonal. Looking at the matrix it appears that star mass and planet radius have a positive relationship. Also, star mass and star radius are positively related. Star metallicity and star temperature are somewhat negatively related. The highest correlation between variables is 87% correlation for star radius and star mass however, this is not surprising because it is expected that a high mass star will also be large in size.

Table 2: **H1** Model Output

	Estimate	Std. Err	t value	p value
Intercept	0.49	0.09	5.55	< 0.001
Star Radius	0.29	0.08	3.80	< 0.001
Star Mass	0.39	0.14	2.76	< 0.010
Star Metallicity	-0.20	0.08	-2.45	0.015

The first hypothesis states that as host star radius, mass, and metallicity increases so does the planetary radius of planets orbiting around the star. The table shows the output of the linear model. All values are significant for the t test ($p < 0.05$). Looking at the estimate column it can be determined that 1 Jupiter radius increase in the exoplanet will result in a 0.29 solar radius increase, 0.39 increase in mass, and a decrease of 0.2 solar metallicity for the host star. Metallicity is measured from the metallicity of the Sun being the standard therefore, if the exoplanet increases 1 radius then the metallicity will decrease 0.2 the metallicity of the Sun. This decrease does not support the hypothesis for star metallicity and planet radius being positively related however, later it is shown that planetary mass and star metallicity are slightly positively related. The standard error column is the average amount that the estimate varies from the actual value. That being said, the lower the number the better. Here the numbers are very low and show little error in the estimates. The t-values are literally the estimates divided by the standard errors. The high magnitudes of the t-values are a sign that the coefficients are going to be statistically significant.

Table 3: **H2** Model Output

	Estimate	Std. Err	t value	p value
Intercept	0.61	0.18	3.35	< 0.001
Star Mass	0.23	0.16	1.41	> 0.050

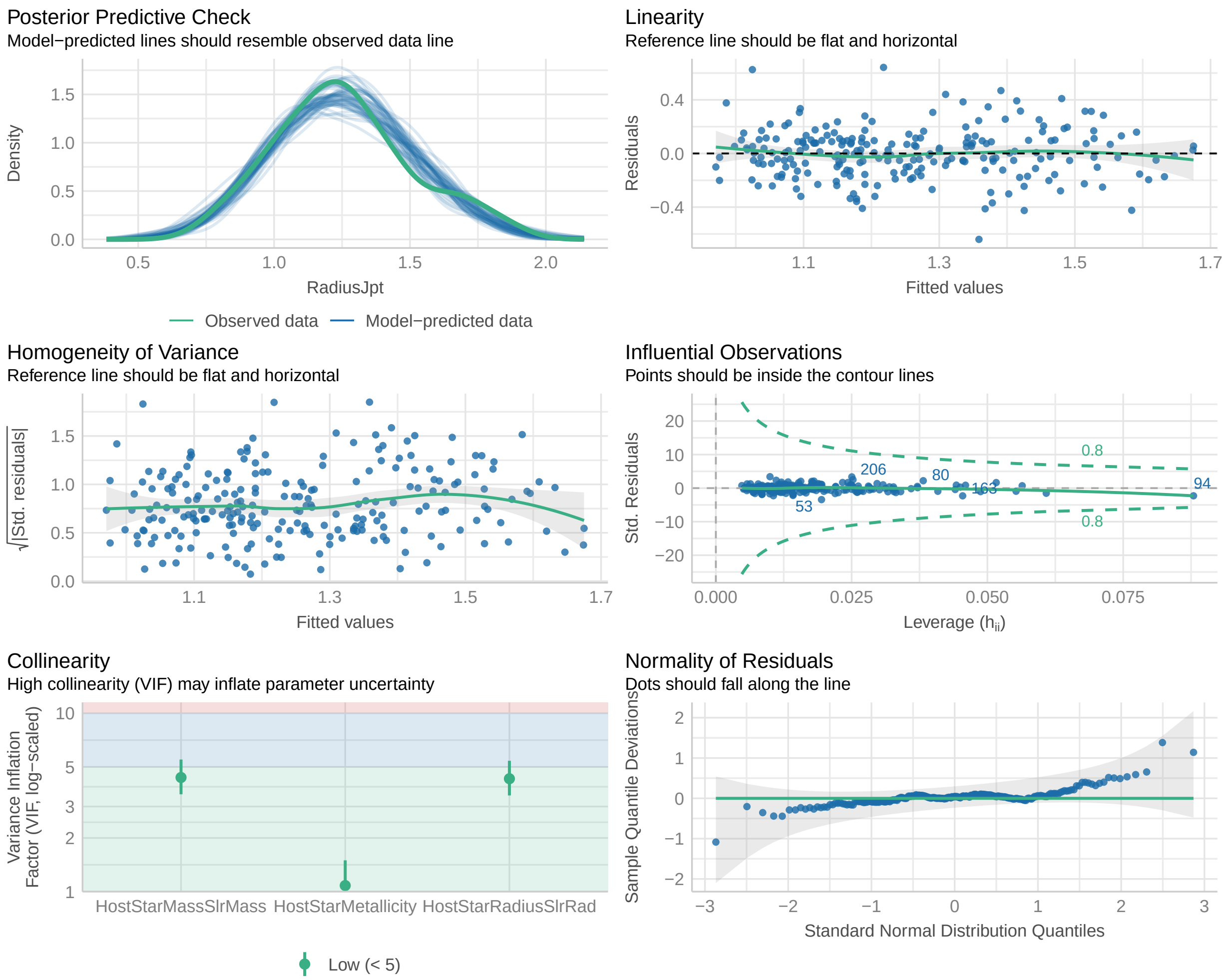
The second hypothesis states that as host star mass increases so does the planetary mass of planets orbiting around the star. This hypothesis appears to be rejected from the t test here ($p > 0.05$).

Table 4: Model Performance

AIC	BIC	R^2	RMSE	Sigma
-105.06	-87.98	0.45	0.19	0.19

These values are for the best model as the other models were discarded because their lack of quality. Although other models were tested in this study for example interaction between star mass and star radius, the addition of star temperature, and interaction between star mass and star metallicity their AIC values were all higher than this model and made the p values skyrocket above the alpha value of 0.05 for all the explanatory variables. The R^2 value is positive and generally acceptable however, less than 70% making it have limited strength for explaining the results of the hypothesis. The root mean square error and the standard error are both low here showing some hope for the hypothesis. The goodness of fit comparison or the Bayesian Information Criterion is very low and therefore gives evidence of a good model.

Figure 5: Diagnostics



Posterior Predictive Check plot shows the predictive model fits the observed data in a relatively acceptable parallel. The Linearity plot or residuals vs. fits plot shows that the residuals bounce randomly around the zero line. Although the line isn't totally straight, the Homogeneity of Variance plot shows the variance between groups is even. Influential Observations shows no outliers that greatly affect the slope of the regression line. The Collinearity plot shows low correlation between the independent variable and the other variables. Normality of Residuals shows the underlying residuals are normally distributed.

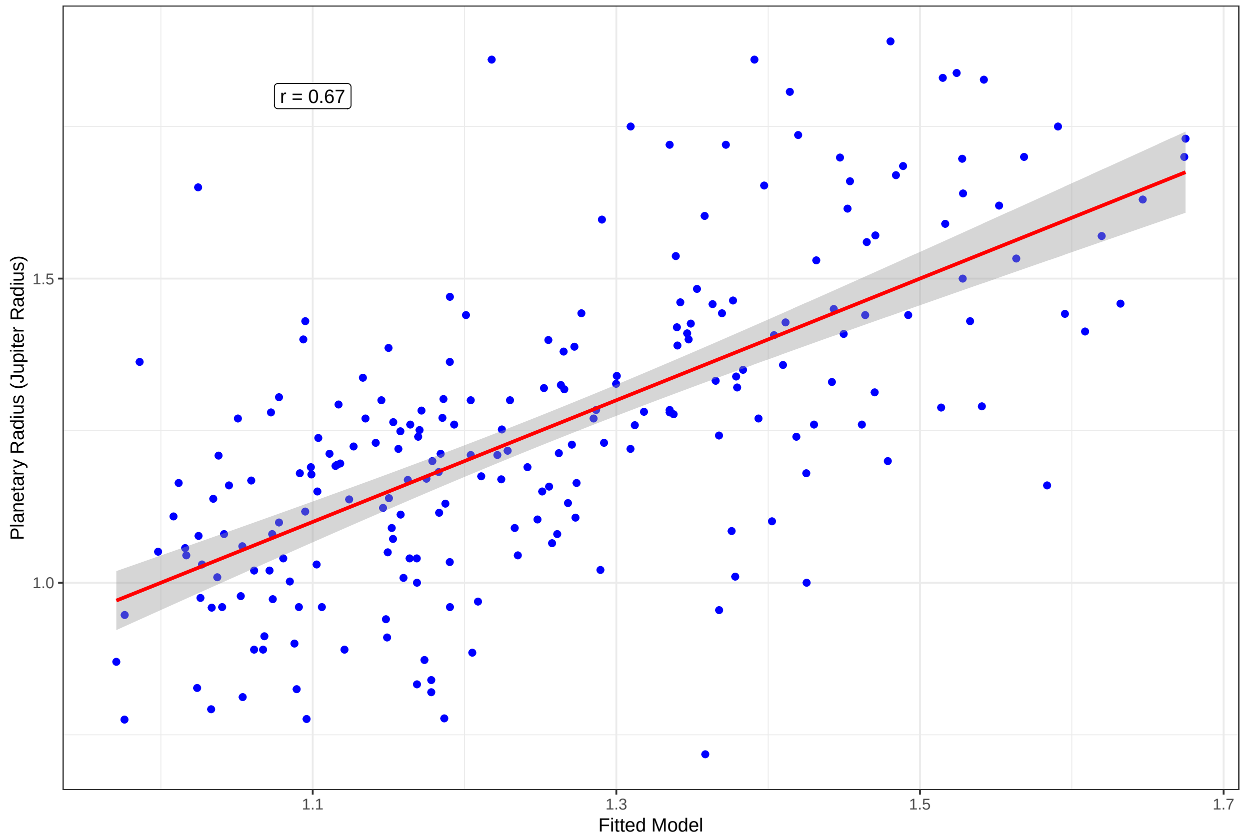


Figure 6

The correlation coefficient value of 0.67 gives good fit for the fitted model and the dependent variable. The 95 percent confidence interval around the predicted line is minimal giving strong evidence for the predictions.

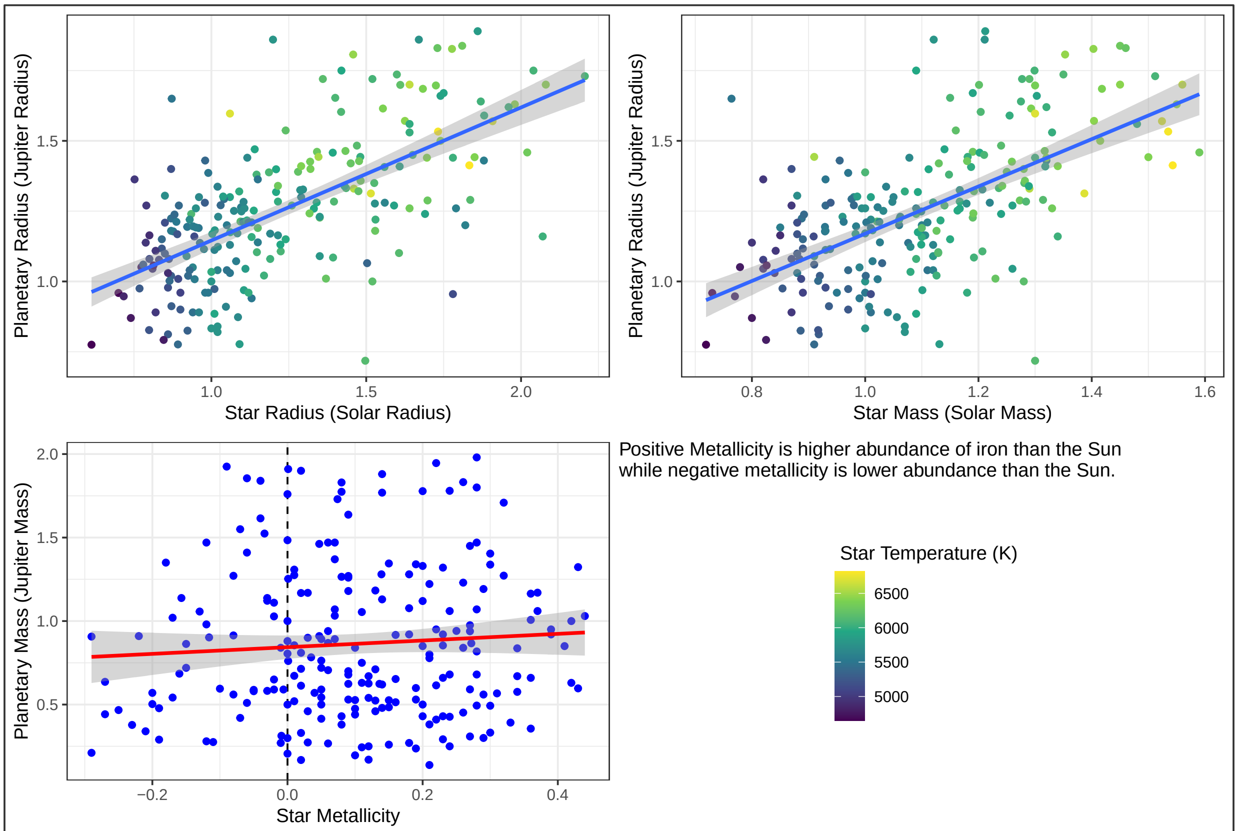


Figure 7

The top scatter plots are the dependent variable (planetary radius) with star radius and mass. Star temperature is colored with very clear evidence of increasing star size correlated with increasing temperature. The temperature cannot be assumed to be correlated with planetary radius from these plots because the temperature is only following the characteristics of the star. Star metallicity shows a slightly positive relationship with planetary mass however, the statistics were not analyzed for planet mass and its host star metallicity in this project. Further research is needed for this with inclusion of newly discovered planets recently added to the Open Exoplanet Catalogue dataframe.