Background

The Exoplanet database as a whole is a list of all known exoplanets discovered as of June, 2018. Along with these planets are several other data for each planet such as jupiter mass, radius, stellar mass, eccentricity etc. I began by taking a subset from this large dataset that excluded every planet that lacked data for either Stellar Mass or Jupiter Mass, because these were the two main variables I was interested in investigating. An exoplanet's stellar mass is the size of the star that hosts the planet in relation to our sun. Therefore, an exoplanet with a stellar mass of 1 would orbit a star that is the exact same size as our sun. Jupiter Mass is the size of the planet itself in relation to Jupiter. Unfortunately, deciding to do so took me from roughly 4,000 planets all the way down to 1,464 planets. Within this subset, I split my the exoplanets into large and small mass planets (above and below the median) and large and small host stars (again, based on the median).

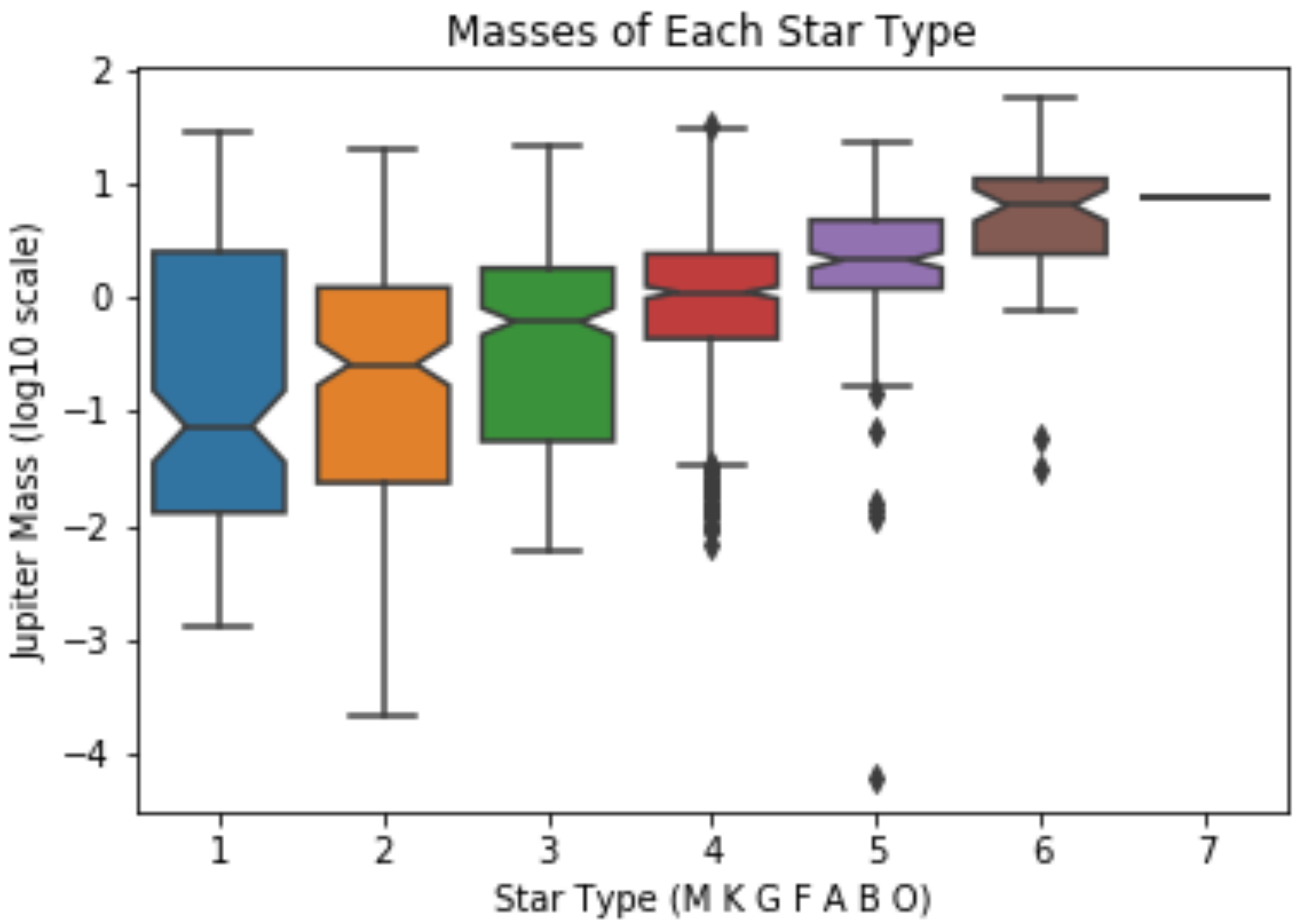
I was originally interested in the different methods of discovering these exoplanets because my knowledge on this subject was extremely limited. In particular, I was intrigued by the method of Microlensing because it was said to be the least biased of all the methods. I decided to attempt to test whether or not Microlensing should truly be considered unbiased by comparing the masses and radii of the planets discovered from this method to the masses and radii of planets in our "typical" solar system. However, if I was going to solely look at planets discovered through Microlensing, then I would have only been looking at roughly 40 planets. This was far too small of a dataframe to work with.

However, through experimenting with Microlensing planets, I discovered the stellar mass of the exoplanets and what this meant. I found a new interest in exploring this variable, especially the size of the host star's relationship with the planet. It seemed to me that a larger host star should correspond to a larger exoplanet. Investigating this claim would prove to be the question of my entire project.

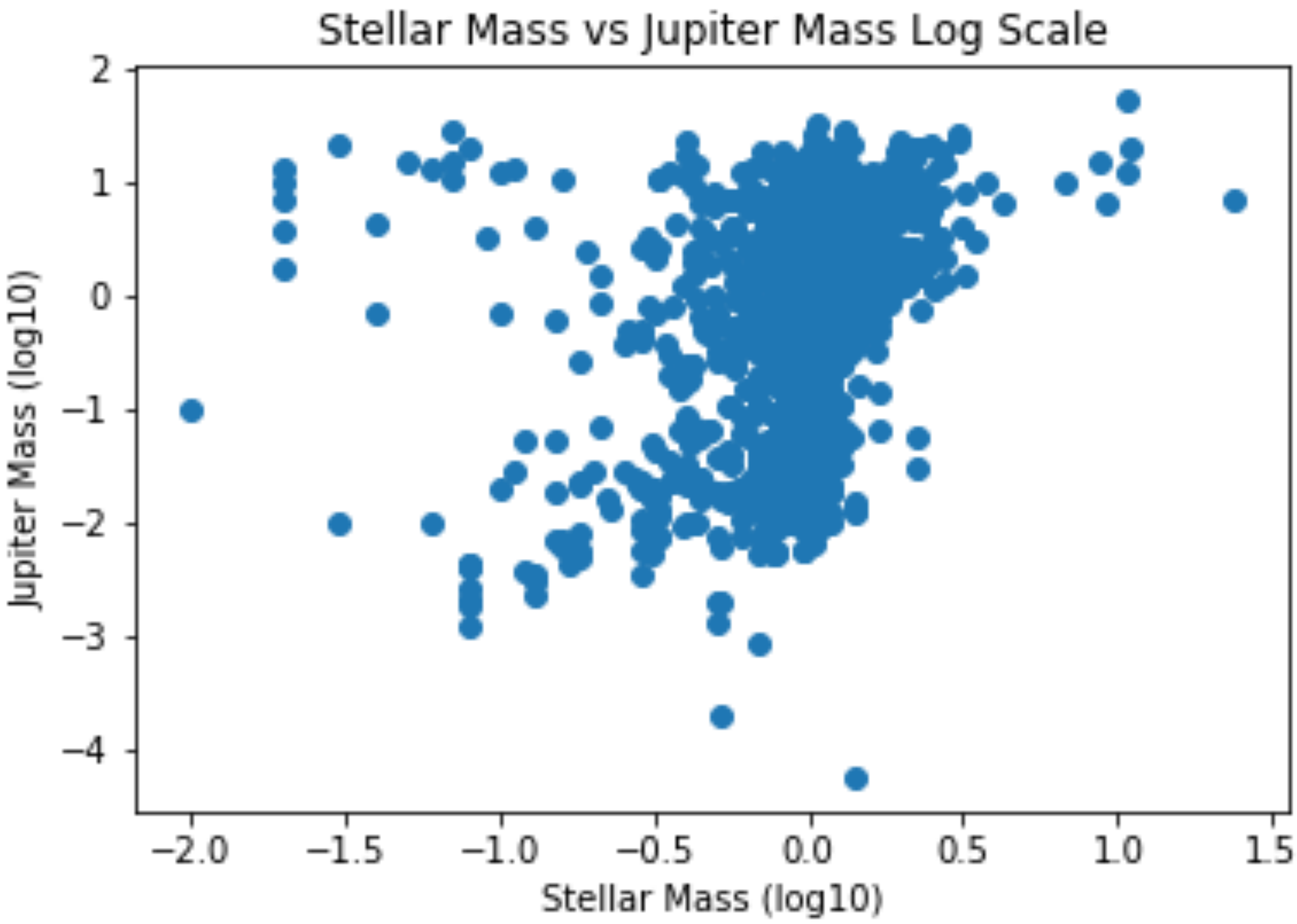
Procedure

To begin, I rid the original dataset of any exoplanets containing "NaN" values for either stellar mass or jupiter mass so that I could run correlation tests and make plots. This step provided a pretty heavy blow to the dataset, leaving me with only 1,464 planets out of the original 3,924 planets. I then assigned each star a type. To do this, I created a "type" column in the dataset that looked at each planet's host star's stellar mass. The given type aligned with the different classes of stars: M, K, G, F, A, B, O. Stars in the M class are the smallest, with solar masses from 0.08 to 0.45, and stars in the O class are the largest, with solar masses greater than 16. However, for graphical reasons, I named the classes 1-7 instead of M-O. This was to ensure the correct order of star types when I created my boxplot. My first plot took each star type and created a boxplot with notches representing 95 percent confidence intervals around their medians. I decided to use a boxplot particularly because of the way they present the median.

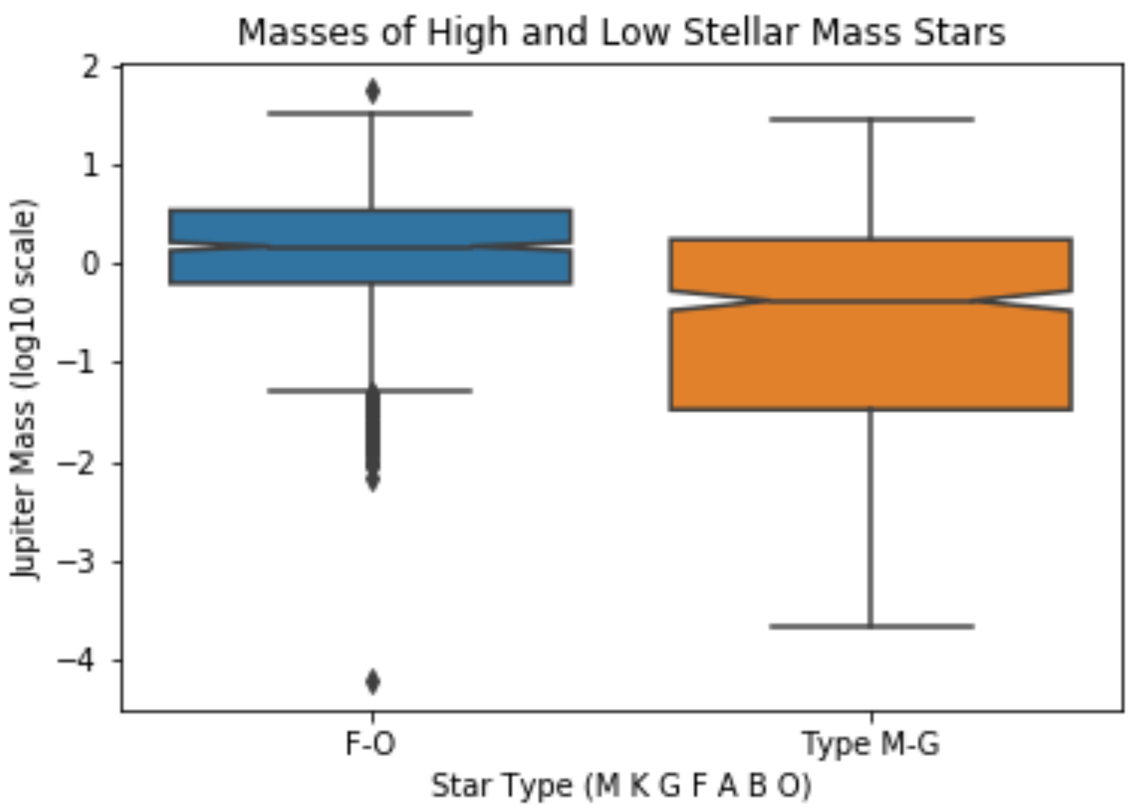
Several of the different types of stars had significant outliers when it comes to planet jupiter mass, I didn’t want these outliers to affect the datum that represented the whole star type.



In order to further investigate this relarionship, I went to a simpler plot design. I created a scatter plot with stellar mass as the x-axis and jupiter mass as the y-axis. Due to a large variety in values and significant outliers, especially in the stellar mass, I put the graph into log-log scale. This plot did not seem to display much of a relationship at all, let alone a strong, linear one. However, I decided to run correlation coefficient tests to be sure. I assumed that the most appropriate of these tests would be the Spearman correlation coefficient test due to the nature of my variables’ distributions. Neither of them seemed linear or normal, so the Pearson test was not an option, and the variables were not categorical so I didn’t feel as though the Kendall Tau test would suffice either. Therefore, the Spearman test was the one I decided to use for my test. The results from this test were honestly slightly disappointing, however. The correlation coefficient was 0.411 with a p-value of 8.78e-61.



I then decided to simply split the data into high stellar mass stars and low stellar mass stars to see if the two had separate medians. This boxplot provided evidence that planets with smaller host stars may have smaller masses than planets with larger host stars.



Analysis/Discussion

The first boxplot, while nicer to look at and perhaps suggestive of a relationship, wasn't truly as accurate of an indicator as I had hoped. The overlapping confidence intervals of each star type proved that we can't conclusively claim each population of star types to differ from one another.

M	-0.904	1.052	
K	-0.112	0.6379	no
G	0.300	0.936	no
F	0.748	1.500	no
A	1.604	2.676	yes
B	3.987	9.003	yes
O	7.52	7.52	no

This table shows each star type's 95 percent confidence interval about its median log10 jupiter mass and whether or not it significantly different from the class below it. Given the fact that only two of the types could be considered to be statistically different, it is unclear whether the positive relationship between stellar mass and planetary jupiter mass is concrete. Additionally, the correlation coefficient of 0.411 suggests a slight possibility of a relationship between stellar mass and planetary jupiter mass, but it is clear that many other variables affect the mass of an exoplanet more directly than the solar mass of its host star.

However, after dividing the exoplanets into only two groups, large and small host stars, the new confidence intervals about their median did not overlap. Suggesting that the larger host stars do have larger stars than the smaller host stars, or we can at least be 95 percent confident of this. While it may not be a pretty, linear relationship with a perfect relationship and correlation coefficient anywhere near a perfect one, it is within reason to say that the population of exoplanets with host stars that are larger than 1.04 solar masses typically have larger jupiter masses than the planets with host stars that have solar masses smaller than 1.04.

Small Host Stars (<1.04)	0.163	0.689	
Large Host Stars(>1.04(	1.142	1.878	yes

Towards the end, I also was interested in seeing if certain discovery methods had any biases towards the host star of the planet that they detect. In order to do this, I replotted the scatter plot, but designated a color for each of the four scatterplots. From this plot, microlensing

appears to have a tendency to detect planets with host stars of smaller solar masses, however the other methods seem to be fairly consistent.