Exoplanets:

We Are Not Alone

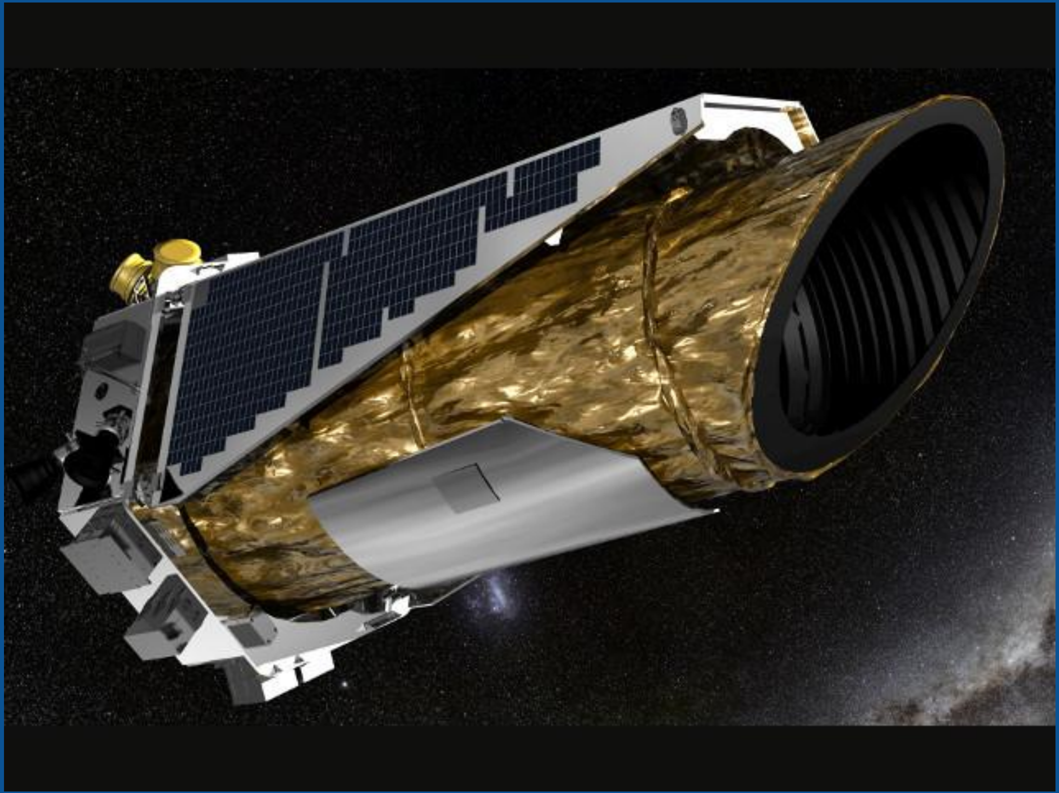
Kepler Space Telescope

Launched in 2009.

It was very successful at finding exoplanets, but failures in two of four reaction wheels crippled its extended mission in 2013.

Without three functioning wheels, the telescope could not be pointed accurately.

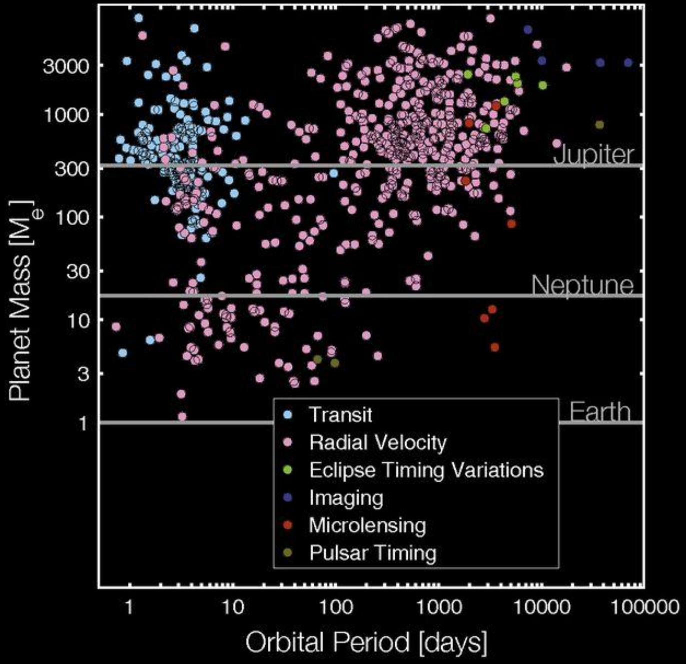
On October 30, 2018, NASA announced that the spacecraft was out of fuel and its mission was officially ended.

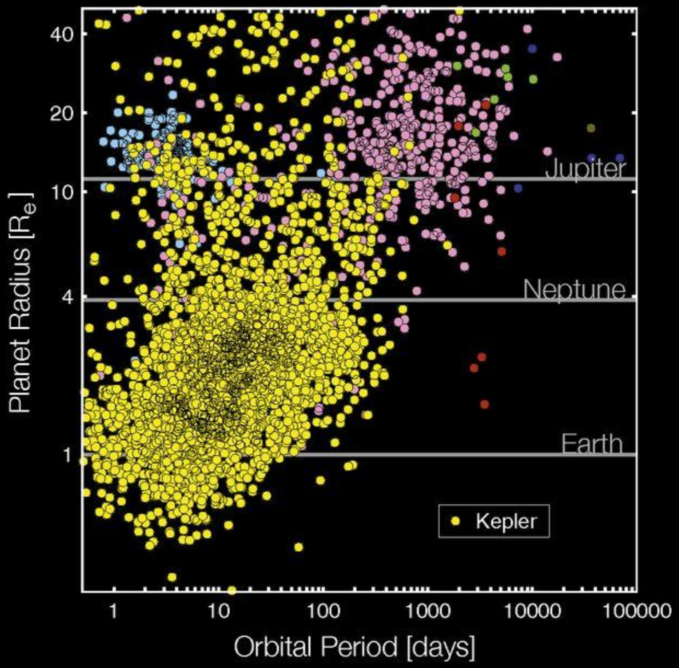


Successful Searching

The dots on the left show exoplanets found with the color indicating the method used.

Kepler's exoplanets are the yellow dots on the right.





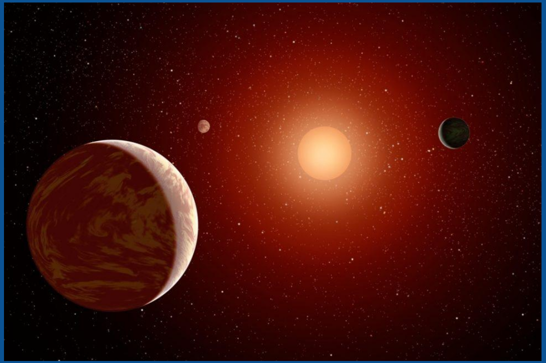
Finding Exoplanet Candidates

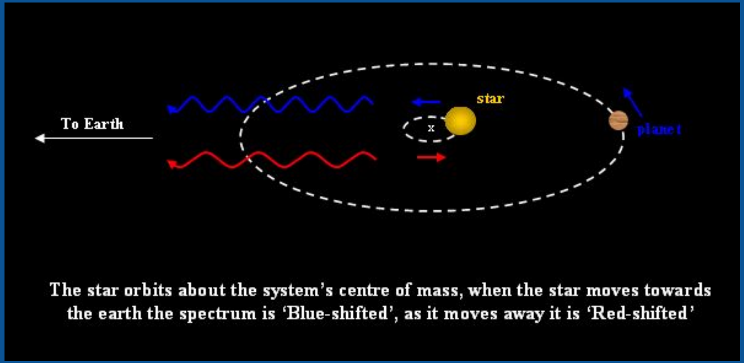
Several methods are available to directly or indirectly find exoplanets. The project's "readme" file has the technical terms and simple definitions. Essentially, they fall into three categories:

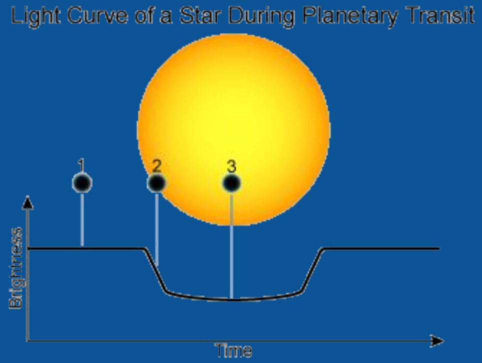
Blocking a bright object to see a nearby dim object.

Wobbles in position or color. Doppler shift.

Changes in a star's brightness. This is a Transit.





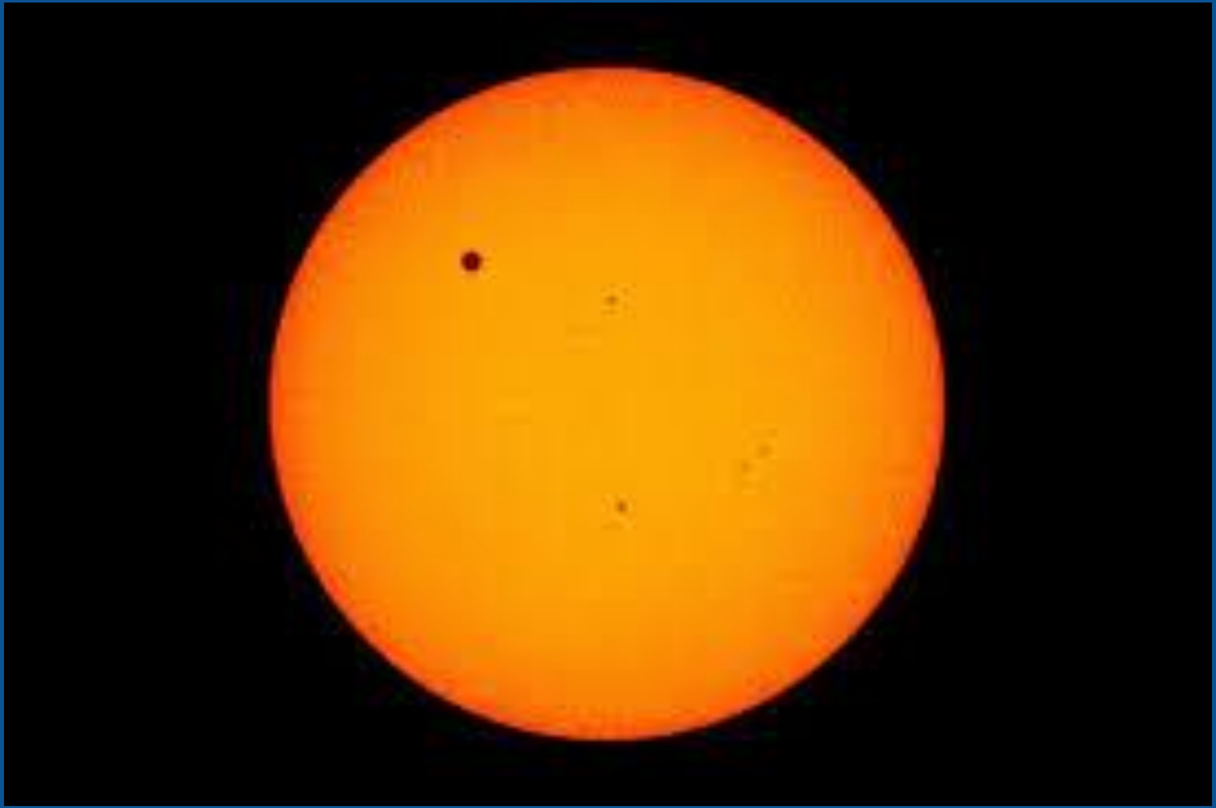


Transits are hard to see

Kepler used "transit" for most of the exoplanets it found.

Venus is shown in transit across the Sun. It dimmed the Sun's light by .1%. The next Venus transit visible from Earth is over 200 years from now.

A single transit may not indicate an exoplanet.



Looking at public data

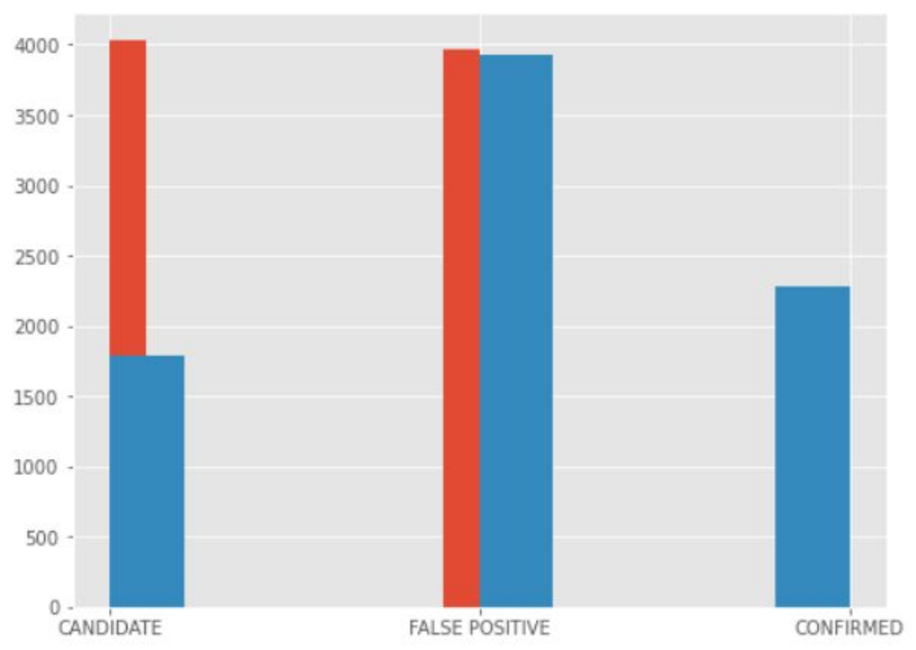
The summary data from the Kepler Space Telescope is available at kaggle.com.

Is this data sufficient to reproduce the determinations?

The Null Hypothesis is that the information is insufficient.

Orange indicates the preliminary classification.

Blue is the final classification.

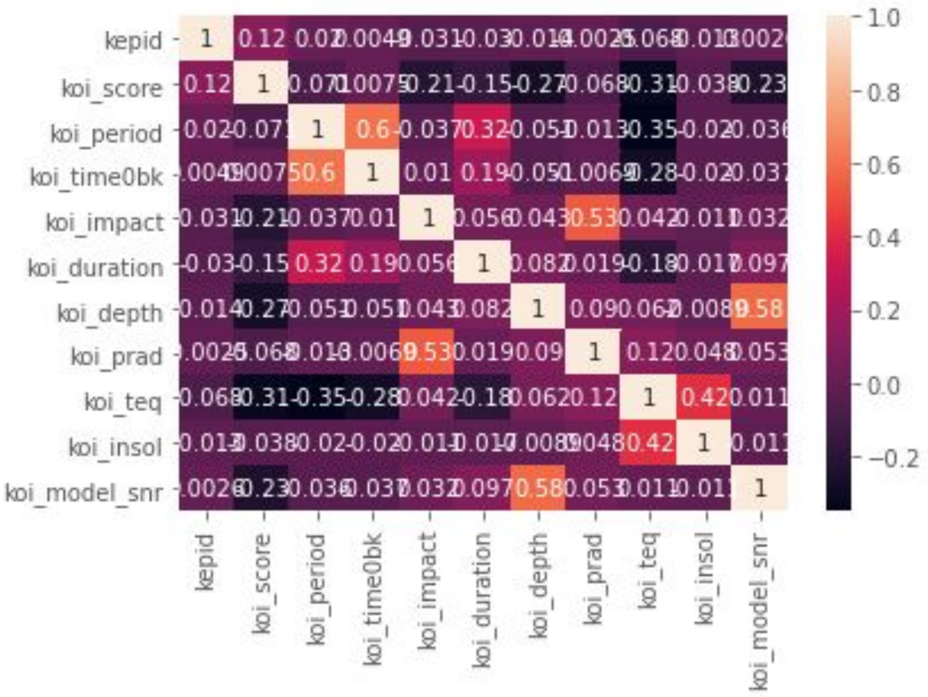


Sorting it out

For each koi (Kepler Object of Interest) value, many data values are generated.

This subset shows correlation between some values. The diagonal is 1 because everything matches itself. Darker boxes show less connection.

Without going deep into the values, data includes: temperature, transit time, radius, inclination (tilt from our view), and more.



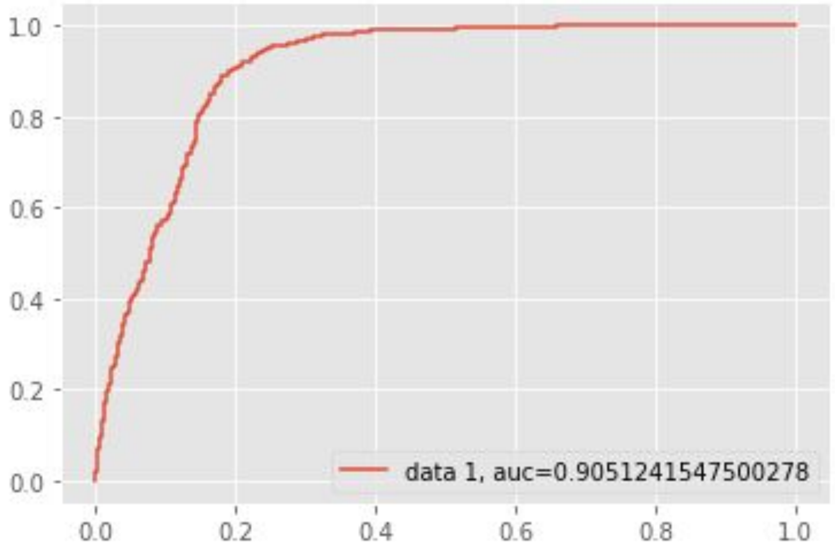
Logistic Regression

Logistic Regression analyzes multiple data columns to try to reproduce a known result.

The diagram on the right shows auc (area under curve). More is better.

The values below are numerical tallies. 645 and 396 are the correct predictions from the model. 131 and 69 are the wrong answers.

```
array([[645, 131],
       [ 69, 396]])
```



Trying to predict

Decision Tree and Random Forest were used to craft the input data to match the conclusions. The methods were adjusted to improve the results. For each value, higher is better. One is perfect.

Accuracy is the overall likeliness that the determination is correct.
Precision is the likeliness that a determination of an exoplanet is correct.
Recall is the likeliness that a determination of not an exoplanet is correct.

In round numbers, the Random Forest will correctly identify an exoplanet 7 times out of 8 predictions.

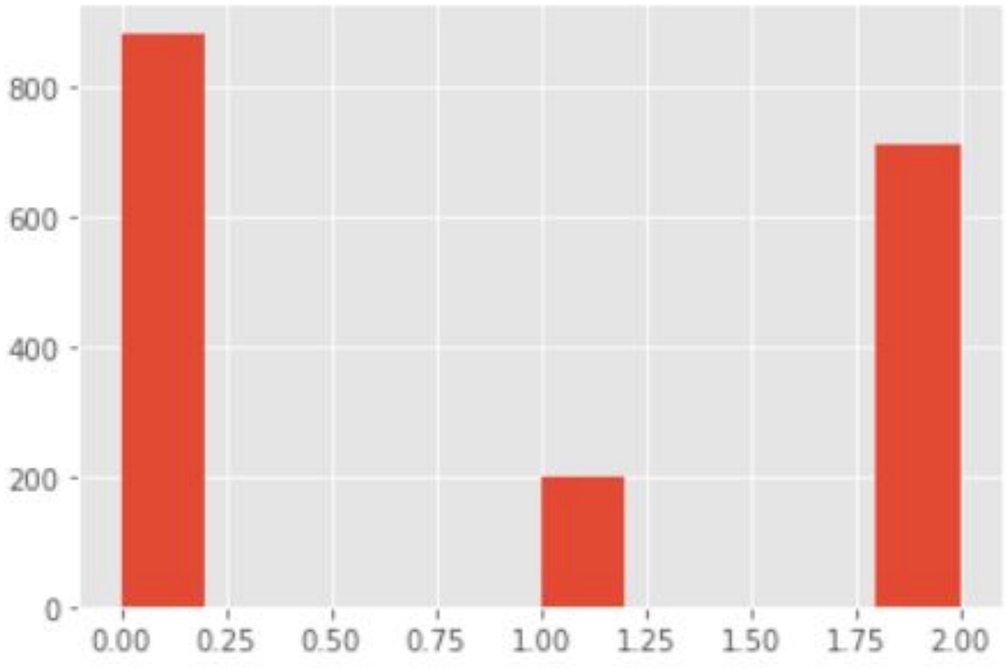
```
Accuracy:  0.9331184528605962
Precision: 0.8804780876494024
Recall:    0.9505376344086022
```

Other data

There are 1792 “candidate” objects in the data. They are neither confirmed exoplanets nor rejected.

The entries were processed through the decision tree and random forest models.

882 results conclude not exoplanets.
711 results conclude exoplanets.
199 results have the models draw opposite conclusions.



Conclusions

- The data usefulness is limited by having summary data.
- The model gives insight into the likeliness that many objects of interest are unconfirmed exoplanets.
- The precision of determining if an object of interest is an exoplanet is too low to be used by astronomers.
- The Null Hypothesis is supported by the model being incorrect 1 time out of 8.

Next steps

- More data will be coming. TESS (Transiting Exoplanet Survey Satellite) is currently gathering data.
- Examining older data sets can add value.
- The detection method of the object of interest was not included in this data set. It could be a factor is the ability to predict.

Questions ?

Thank you