Data Science Competition 2017: Challenge Question

Title:	NUS Data Science Challenge 2017
Release Date:	11 March 2017
Submission Deadline:	19 March 2017 12:00nn

Your team will receive another email containing an invitation to edit a google drive folder named after your team number. All program codes are to be submitted with a deck of presentation slides to the google drive before the submission deadline.

Mission: You are to explore the given outer space to discover the number of exo-planets and the details of the planets.

Prologue: (Episode IV: A New Hope)
"It is a period of civil war. Rebel spaceships, striking from a hidden base, have won their first victory against the evil Galactic Empire. During the battle, Rebel spies managed to steal secret plans to the Empire's ultimate weapon, the DEATH STAR, an armored space station with enough power to destroy an entire planet.

Aim: Dear comrade, the Jedi team has worked hard to identify an area in the space that is highly likely to contain the death star. However, due to the limit of our technology, we are only able to capture the light intensity of the stars in that area. The information such as the number of exo-planets, the period of each exo-planets and most importantly identifying the death star still remain as a mystery to us. We are very concern with this matter, as it would mean the end of the world. Hence, your job is to help us to identify the number of exo-planets given from the data set and other information that you found from the data and report to your superior about your findings.

Things that you can do with the data:

1. Identify the number of exo-planets in the system.
2. Find the radius and the period for the exo-planets that you have found.
3. Construct a model that fit the data. This model will need to explain the light variations caused as the planet orbits about the stars.
4. Optimized the model with any optimization method.
5. Derive parameters for the exo-planet and star from their best-fit model. If possible, derive uncertainties/ confidence intervals for those parameters.

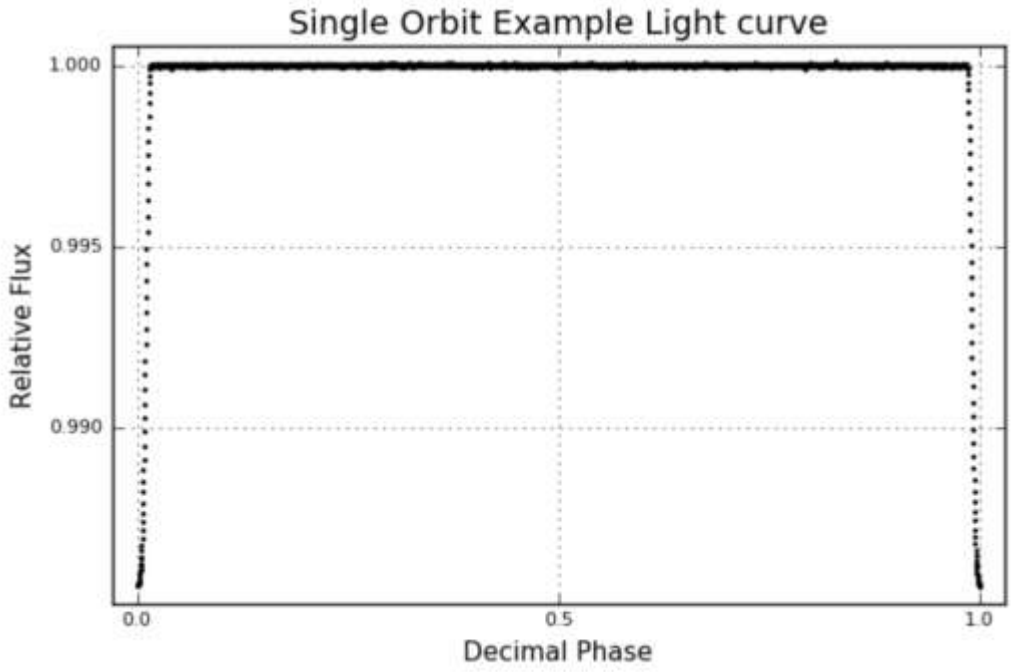
Data sets:

dataset_1.txt	Consists of a single planet making a single orbit
new_dataset_1.txt	Consists of a single planet making multiple orbits
dataset_2.txt	Consists of multiple planets making multiple orbits.

An example of the data you would see:

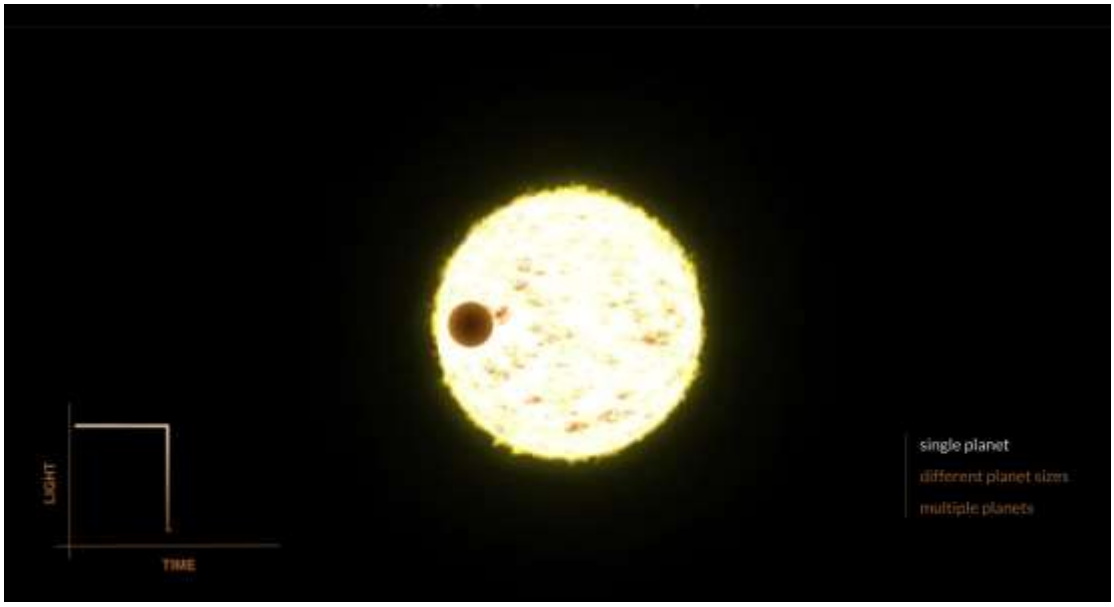
	phase	flux
1	0.0001193	0.9856705
2	0.0003668	0.9856353
3	0.0006201	0.9856503
4	0.0008731	0.9856394
5	0.0011276	0.9856864
6	0.0013819	0.9856918

Time series plot of a single orbit:

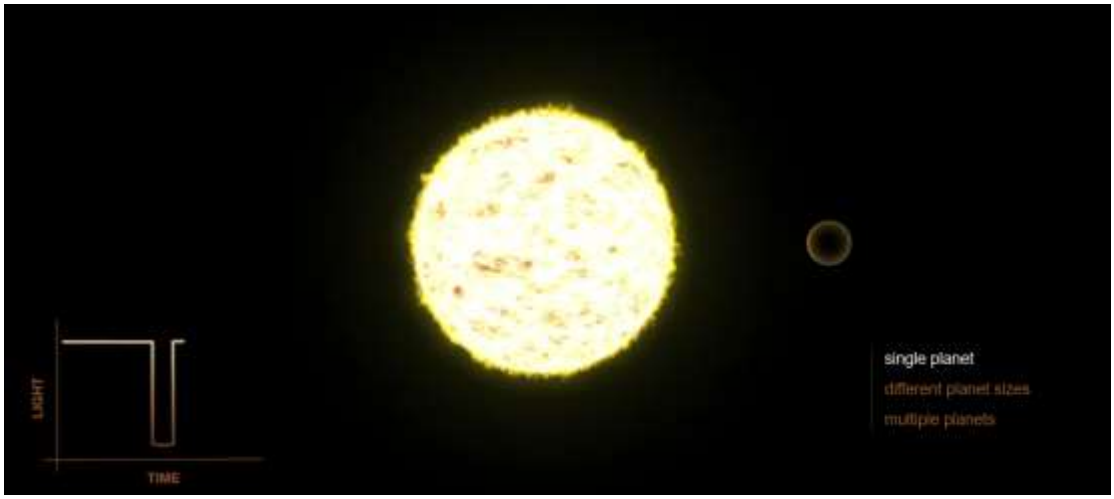


From the plot, we could observe that the intensity measured by the flux is normalized at 1 and a dimming effect is observed when the planet orbits in front of the stars and intercept the stars causing the satellite to record at a lower intensity. Hence from this information, we could derive the period and the dimming tells us the radii of that planet.

For the details to the planet of the system, you can refer to https://exoplanets.nasa.gov/interactable/11/. (Under the transit link)



As the exo-planet cross and orbit in front of the stars, we could observe the dimming effect of the light intensity since some of the light is blocked up by the exo-planet.



The duration of the dimming could give us some insight to the speed and the size of that planet.

You are also encouraged to research on other relevance materials on your own for your reference to solving the problem.

Data Science Competition 2017: Challenge Question

You are graded based on:

Grading criteria	Description(s)	Percentage
Analysis	• The completeness of the analysis (i.e. Introduction, Data analysis, Assumptions made, Model/Parameters validity, Conclusion) • The accuracy of your answer (i.e. The number of exo-planets, Period/Radius of the exo-planets, Estimated parameters, Best fitted models) • Efficiency and the methods used for the analysis. • Reproducibility of your work. • Choice of data used with permission by the organizing team. (i.e. you can dissect the data given up to your preference)	40%
Presentation	• Clear explanations of your methods and the approach to the problems. • Presentation is well-organized • Presentation is within the limit of 10 minutes per team (Round 2). • Presentation is well assisted with diagrams and tables.	40%
Others (Bonus)	• Identify the uncertainty and confidence intervals of your parameters and models. • Identify other information in the data given if possible.	20%