Computer simulation: Computational Methods, assignment 1

Name: Rohan Kadam

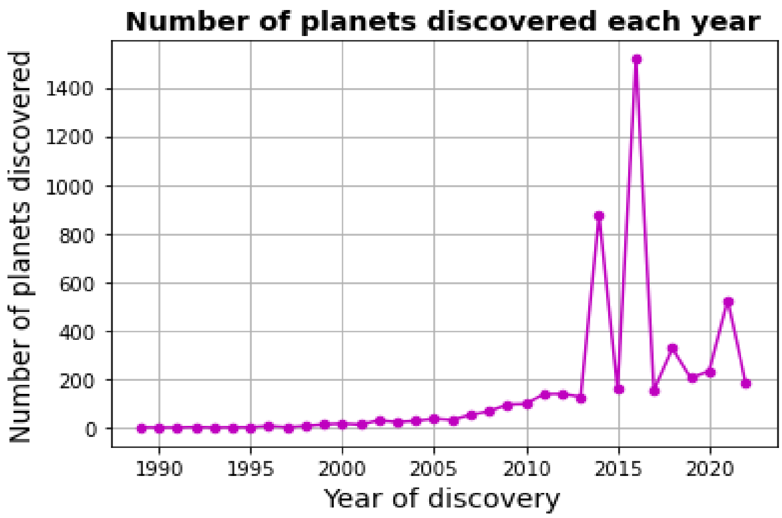
Student number: 20334092

The premise of this assignment is to download and extrapolate the data from the NASA exoplanet archive. This data set provides us with information ranging from but not limited to a planets: name, discovery year, stellar mass, and facility of discovery. We download the package and open the data file on Jupyter notebook using the function f = open('planets.csv'). Initially the data is not in a format that is easily accessible to read and manipulate. Using pythons in built functions we strip *([string.strip])* all lines starting with # and split each line of the file into a list. The data is now ready to use for our upcoming problems. Two packages were used in this assignment, [matplotlib.pyplot] and [numpy]. Sample pseudocode of how I completed the assignment is provided below.

(a)

In part 'a' we are trying to discover how many planets are discovered each year, and to make a graph ('number of planets discovered' against 'year of discovery') to denote this information. One method to ensure we have included all planets in our data set is to detail what the youngest and oldest planet discovered is, and creating a list which will yield all planets in between these two years. To do this we use pythons maximum and minimum functions ([max = '-', min = '-']) to yield an integer value ([int.]).

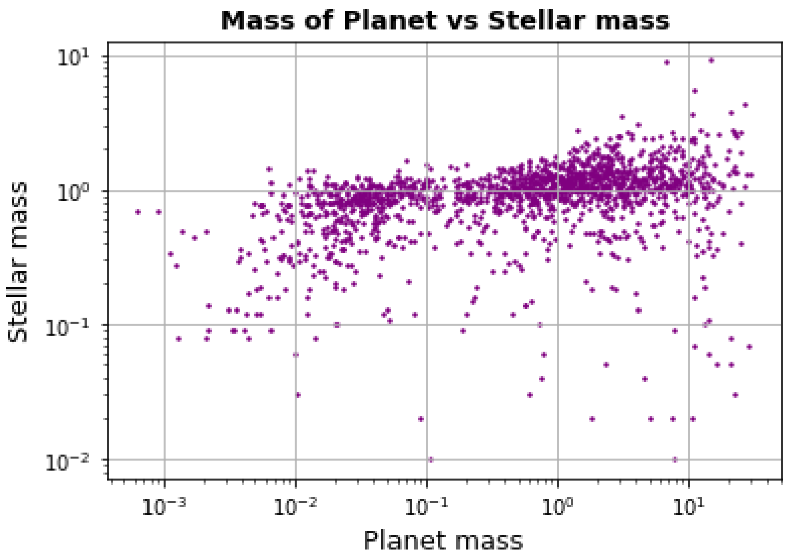
We create a range function ([for i in range]) to include all years that range in between the oldest and youngest year of discovery. We need to make sure that the graph details the values in a chronological manner ranging from oldest year of discovery to youngest year. To do this we sort the variable for year of discovery (in this case yr) using pythons sort function *([yr.sort]).* We also create a function which returns the 'number of years/years that a planet was discovered in'. We plot a graph by using a function from the matplotlib.pyplot library *([plt.plot]).* The graph is labelled and manipulated for clear viewing using functions from the matplotlib.pyplot library. The graph is shown below.



The intervals on the x-axis were experimented with in the view of how to get the most aesthetic while most informative graph. As we can see from the graph, not in every year there was a planet discovered. The year in which the most planets discovered was 2017-2018. From 1887-1998, almost no planets were discovered. As expected with tools improving and developing as time 'moves forward' the number of planets discovered increases.

(b)

In this task we are required to investigate the correlation between the mass of a planet and the mass of its respective star. For example in our own galaxy the mass of the earth is about 3 millionths (0.0003%) the mass of the sun while the mass of Jupiter is about 9 hundredths (0.09%). A log-log scale is applied to the graph. This is to decrease the level of skew in the graph, i.e when a couple of points may be much larger/lower than the range where the rest of the values lie. A log-log scale is also applied when you are trying to display a large range of values in a compact, precise, and easy to read manner (in this case masses of stars and planets). To manipulate the data we organized at the start to yield the desired result we need to make a list which includes all the relevant data (in this case stellar mass and planet mass). We create a variable for each and assign the variables as a float. For example *(planetmass = [float(entry[5]) for entry in planets[1:] if entry[5] != '' and entry[6] != '']).* This also done for each planets corresponding stellar mass. A scatter plot is made using the function from matplotlib.pyplot *[plt.scatter]* and a log-log scale is applied (plt.xscale and plt.yscale). The graph is labelled and manipulated for clear viewing in the same manner as previously carried out in task (a).



In this graph a log-log graph was plotted to ensure we can see all the results clearly without data points being 'squeezed' in. It is used when the data you are displaying is much more or much less than the other data you are displaying i.e. there is a large spread of data. A very large difference/spread from all the data is expected when we are dealing with planet and stellar mass (as the difference in mass' can be so high). Here we have plotted the stellar mass of a star against the mass of the respective orbiting planet. We can see a majority of the planets have close to the same stellar mass. The planets with the highest mass and lowest stellar mass can be seen on the bottom right. Similarly the planets with the lowest mass but highest stellar mass can be seen near the upper left of the scatter plot. These are not that common, potentially because planets with such a low relative mass may not even be considered as planets (as we have seen with pluto in our galaxy). We can conclude there is a strong correlation between the mass of the star and its relative orbiting planet, with the higher the mass of the planet the higher the mass of the star.

(c)

We wish to extract many pieces of information for this task and print them. First we need to find the number of discoveries each facility has made and when they made their first discovery. To do this we first create a list of all the unique facilities, like so: [facilities = *{entry[3] for entry in planets[1:] if entry[3] != ''}]*
[facilities = list(facilities)], and create a function to tell us when they each made their first discovery. Then we create a list in such a manner, facility name ---> number of discoveries --> year of first discovery. We sort this list in chronological order like so:
[Top10 = [main_list[i] for i in range(10)]]
[Top10.sort(key = lambda x: x[2])]. We print this list and print the facility that discovered the most planets, its name and how many planets it discovered. The results are shown below.

Name of facility	Number of discoveries made	Year that the first discovery was made in
Multiple Observatories	191	1996
W. M. Keck Observatory	184	1998
La Silla Observatory	270	1999
OGLE	81	2002
HATNet	67	2006
SuperWASP	113	2007
Kepler	2708	2009
HATSouth	73	2012
K2	537	2014
Transiting Exoplanet Survey Satellite (TESS)	253	2018

```
The facility that has discovered the most planets is Kepler, with 2708 discovered planets
```

As we can see the data is displayed in the order of when which facility made their first discovery. To find out over how long a time period the facility had been discovering planets we subtract the year that the first discovery was made in from 2022 (assuming they are still attempting to discover planets, i.e. Kepler has been discovering planets for 13 years.) From the list we can clearly see that Kepler has discovered the most planets overall.