Computational Methods Assignment 1

ismisebrendan

December 23, 2023

Firstly, numpy and matplotlib.pyplot were imported. Then planets.csv was imported and processed as in the lab, saved to a list 'planets'.

1 Question A

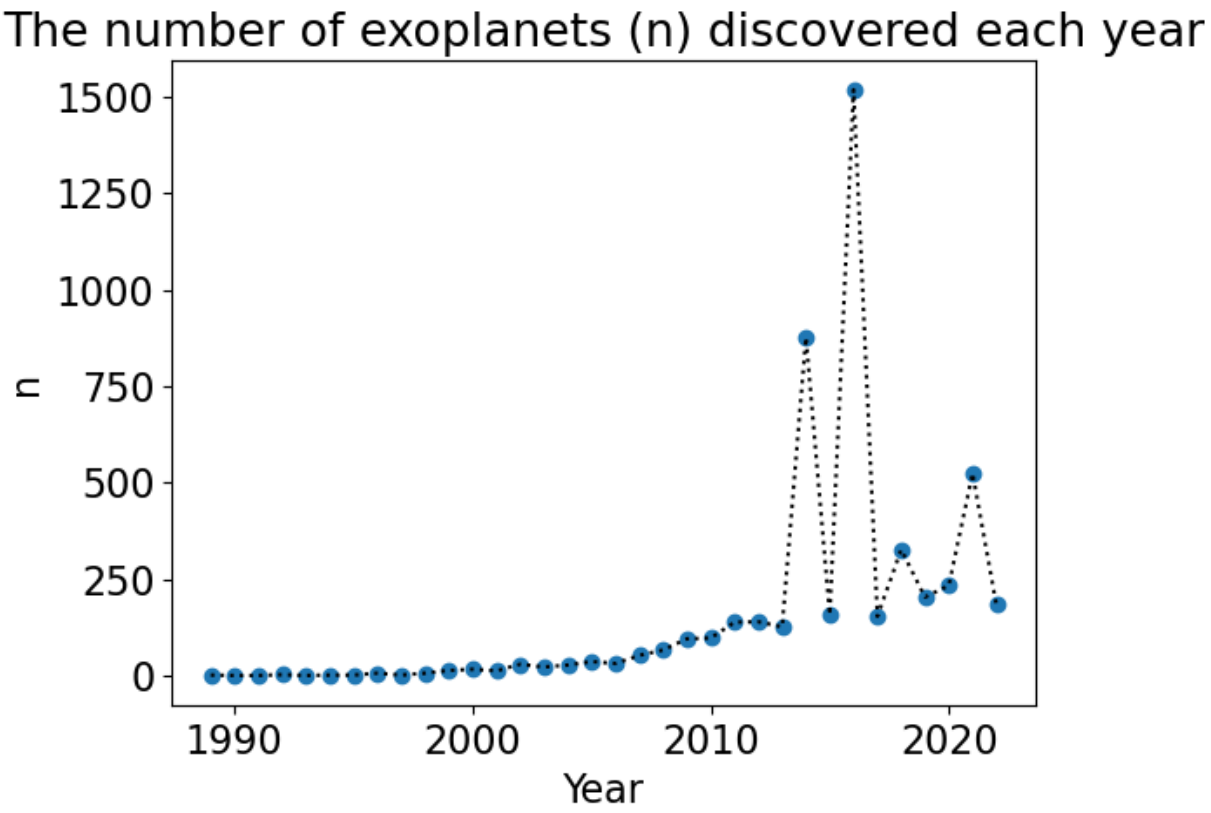

Figure 1: The number of exoplanets discovered each year from 1989 to 2022.

To find the number of planets discovered each year the "disc_year" column of the planets list was imported into the "years_list" variable, ignoring the initial column with the title. Then, using numpy an array was created from the first year of an exoplanet being discovered to the last year - this was necessary as there were some years, especially towards the start when there were no exoplanets discovered.

The number of times each year appears in "years_list" was counted, including the years which don't appear at all, using the code below. The count is stored in the "year_count" variable.

```
year_count = np.array([])
for i in years:
    year_count = np.append(year_count, np.count_nonzero(year_list == i))
```

Then the arrays "years_list" and "year_count" were plotted against each other, producing a scatter plot of the number of planets discovered per year, with a black dotted line connecting them, as without the line it was difficult to tell which data point came before or after some others (for me at least).

The number of planets discovered each year since 1989 is shown in the graph in figure 1.

2 Question B

The "pl_bmassj" and "st_mass" columns of the planets list were imported into the "masses" variable. The rows without a mass listed for either the planet or star were ignored as they would cause issues when converting to floats for the purposes of plotting.

This was converted from a list to a numpy array, and the first column, the planetary masses, was plotted on the y-axis and the second column, the stellar masses, was plotted on the x-axis. This graph is shown below in figure 2.

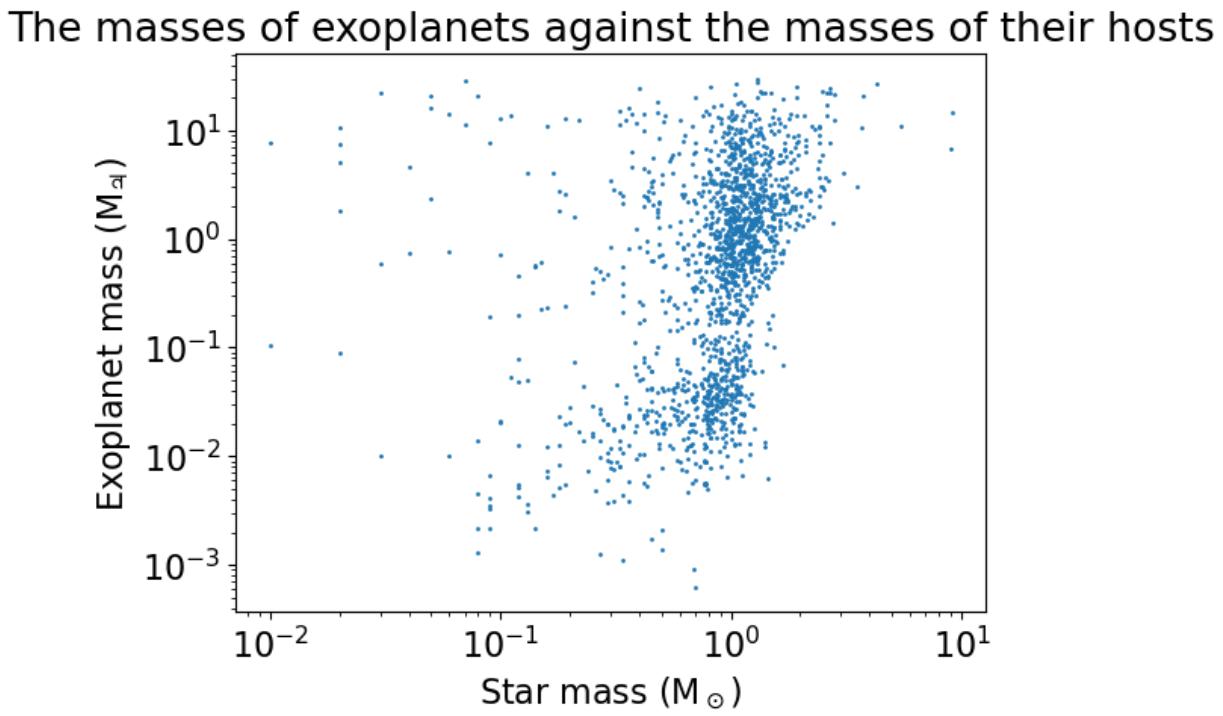

Figure 2: The mass of each exoplanet discovered against the mass of its host star (in the cases in which both masses were known).

As can be seen from the figure, there doesn't appear to be any correlation between the mass of a planet and the mass of its host star. The only interesting observations I can make of this data are that there is a large proportion of the stars have masses between approximately 0.8-2 $M_\odot$, and that there appears to be a gap in the data, showing very few planets with a mass of approximately 0.14 $M_♃$. I don't know if this is simply an artefact of the data, or an actual feature of the mass distribution of planets.

3 Question C

An array of the facilities and the years in which they discovered each planet was created by importing the "disc_facility" and "disc_year" columns of the planets list and saving them as the numpy array "fac_array".

Using numpy's "unique" function the number of times each facility discovered an exoplanet was counted and stored in "facs" and "planet_count", as shown in the code snippet here.

```
# Count occurrences
fac_count = np.array(np.unique(fac_array[:,0], return_counts=True)).T
facs, planet_count = fac_count[:,0], np.array([fac_count[:,1]])
```

Meanwhile "fac_array" was sorted according to the year of discovery using numpy's argsort. This was then looped over to find the year of the facility's first discovery. This was done by looping over all the unique facilities, as stored in "facs" and storing the facility and the first year in which it appears on the same line of the sorted list. The code for this is shown below.

```
# Sort by year
inds_year = np.argsort(fac_array[:,1].astype(int))
fac_sorted_years = np.array(fac_array)[inds_year]

# Loop over each facility to find year of each discovery
first_discovery = np.empty_like(fac_count)
pos = 0
for i in facs:
    # Find the first occurrence of each facility and get the year
    for j in range(len(fac_array)):
        if i == fac_sorted_years[:,0][j]:
            first_discovery[pos] = [fac_sorted_years[:,0][j],int(fac_sorted_years[:,1][j])]
            break
        j+=1
    pos += 1
```

This array was then sorted by the facility name as well, as "planet_count" was also sorted by facility. The number of planets as in "planet_count" was inserted in between the facility and year of first discovery in "first_discovery" to produce the array "fac_planets_first".

This was then sorted by the number of planets which were discovered in each facility, and the top 10 were taken. This finally was sorted by the year of first discovery and outputted as shown in table

Table 1: The ten facilities (or groups thereof) which discovered the most exoplanets, ordered chronologically from their first discovery.

Facility	Number of planets discovered	Year of first discovery
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