Elliptical Galaxy Classification through Palomar Observatory Sky Survey Data (POSS)

Tyler Jung and Aalyan Aamir

04/26/2024

Python Decal: Python for Physics and Astronomy Research

Introduction and Objectives:

Our objective for this project is to explore and analyze large amounts of data on galaxies from the Palomar Observatory. We are particularly interested in elliptical galaxies and decided to analyze the galaxies based on elliptic properties. In this project, we applied filters to reorganize our data, established categorization equations and conducted fitted models. The language we used was Python, the software we used was Jupyter Notebook, and the libraries involved were NumPy, Pandas, and Matplotlib.

Chosen Phenomenon and Data Source:

We utilized data generated from the POSS (Palomar Observatory Sky Survey). “The MCG database contains the ‘Morphological Catalogue of Galaxies,’ a compilation of information for approximately 34,000 galaxies found and examined on the Palomar Observatory Sky Survey (POSS). Individual identifiers are assigned for about 29,000 galaxies and information on the remaining 5,000 is present in the extensive notes of the published catalogs (Vorontsov-Velyaminov et al. 1962-1968). The original goal of the compilation was to be complete for galaxies brighter than magnitude 15.1, but the final catalog lists many objects considerably fainter.” From this source, we extracted a csv file containing all the data we needed to conduct our analysis on specifically elliptical galaxies.

Theory:

Equation 1: Ellipticity of Galaxies:

In order to determine the ellipticity of our galaxies in question, we must use the following equation:

$$E = 1 - \frac{b}{a}$$

where E is ellipticity, a is the major axis of our galaxy, and b is the minor axis of our galaxy.

Equation 2; Isophote 'n' values for Elliptic Classification:

Isophotes of elliptical galaxies are usually fitted by ellipses:

Major axis a; minor axis b, ellipticity $\epsilon = 1 - b/a$.

Types of ellipticals:

$$E_n \quad (n = 0, 1, \cdots, 7) \quad \text{if} \quad \frac{b}{a} = 1 - \frac{n}{10}.$$

In order to determine which E

$$\frac{b}{a} = 1 - \frac{n}{10}$$

Part 1: Dataset Filtration

Our first task was to download and filter the dataset at hand. We were working with data from the MCG database as noted above. The dataset included many parameters, including but not limited to: Name, Flag, Category, Right Ascension, Declinatio, Magnitude, Inner Major Axis (Diameter of the major axis of the bright inner port of the galaxy), Inner Minor Axis (Diameter of the minor axis of the bright inner port of the galaxy), Major Axis (Diameter of the major axis of the entire galaxy), Minor Axis (Diameter of the minor axis of the entire galaxy), Inclination, LII (galactic latitude of object), BII (galactic longitude of the object), and Classification.

In our initial download of the data, we obtained the Name, Major Axis, Minor Axis, magnitude, category number, Inner Major Axis, and Inner Minor Axis just to make sure that the data was downloading properly and generating correctly.

We then filtered out magnitude, category number, Inner Major Axis, and Inner Minor Axis to be left with just Name, Major Axis, and Minor Axis, as that is what our preliminary analysis was concerned.

	name	maj_axis	min_axis
0	MCG-05-27-002	8.0	3.0
1	MCG-05-32-043	7.0	4.0
2	MCG-05-32-055	6.0	2.0
3	MCG-05-32-016	0.0	0.0
4	MCG-05-27-004	6.0	5.0

Part 2: Preliminary Classification and Binnings

After filtering our data, we began the preliminary classification of our galaxies. Using equation 1, we used the major and minor axis to determine E, the ellipticity of the galaxies, generally ranging from 0 - 1. Then we created a simple function filtering the E from 0 to 0.3 into the less elliptical category, E from 0.3 to 0.5 into the average elliptical category, and E from 0.5 to 0.7 into the Less Elliptical category, and anything else into the undefined category.

```
            name  maj_axis  min_axis  ellipticity ellipticity_category
0   MCG-05-27-002       8.0       3.0     0.625000      More Elliptical
1   MCG-05-32-043       7.0       4.0     0.428571   Average Elliptical
2   MCG-05-32-055       6.0       2.0     0.666667      More Elliptical
3   MCG-05-32-016       0.0       0.0          NaN            Undefined
4   MCG-05-27-004       6.0       5.0     0.166667      Less Elliptical
..            ...       ...       ...          ...                  ...
95  MCG-05-48-012      12.0       4.0     0.666667      More Elliptical
96  MCG-05-47-013      10.0       2.0     0.800000            Undefined
97  MCG-05-16-015      14.0       8.0     0.428571   Average Elliptical
98  MCG-05-25-002      10.0       7.0     0.300000   Average Elliptical
99  MCG-05-09-017       0.0       0.0          NaN            Undefined

[100 rows x 5 columns]
```

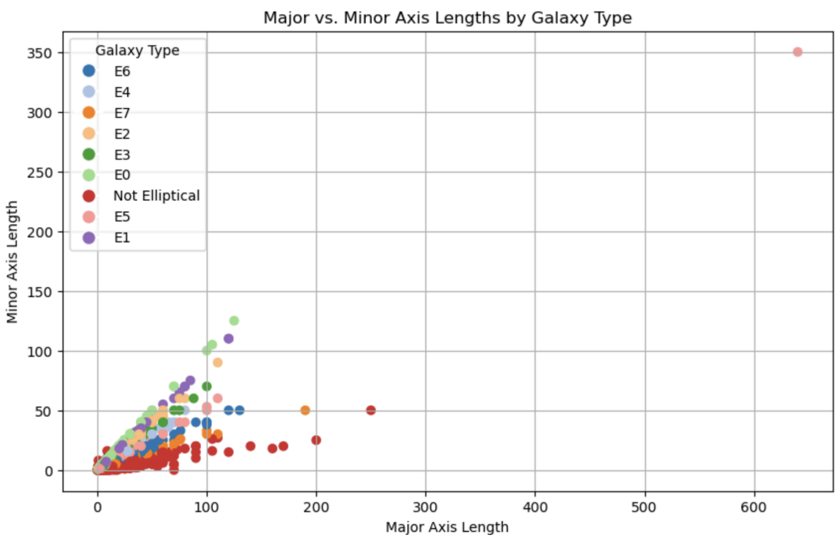
Part 3: Galaxy Type Classification (E0-E7)

The next major part of our categorization is using the equation that categorizes ellipticity to find values for n based on minor and major axis values. E0 is least elongated and E7 is most elongated. However, due to the quotient of minor and major axis values not being an integer, we decided to round to the near whole number so we could categorize into each E(n) bin. We also filtered out all NaNs as some of our data did not have Major or Minor axis data, resulting in NaNs for some ellipticities. This removed 1,605 galaxies, with our new dataset having 18,395 galaxies. Through our code, we were able to create a new column in our dataset that labels each galaxy with its E(n) category.

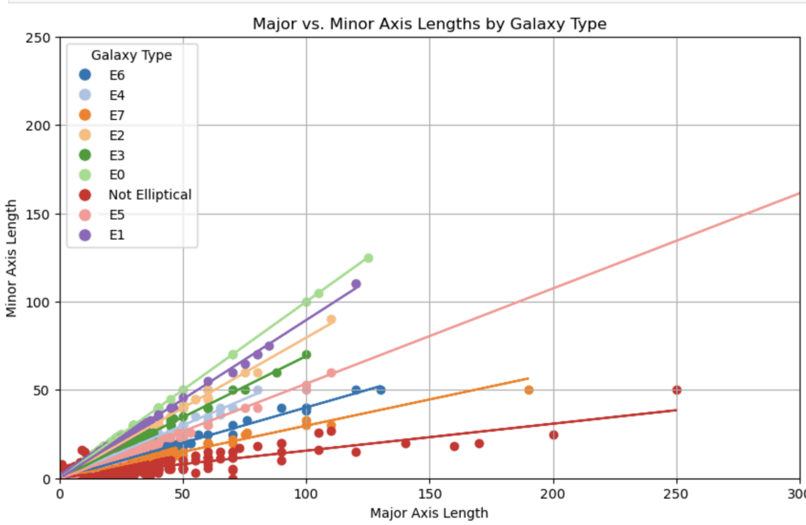
	name	maj_axis	min_axis	ellipticity	ellipticity_category	ellipticity_label
0	MCG-05-27-002	8.0	3.0	0.625000	More Elliptical	E6
1	MCG-05-32-043	7.0	4.0	0.428571	Average Elliptical	E4
2	MCG-05-32-055	6.0	2.0	0.666667	More Elliptical	E7
4	MCG-05-27-004	6.0	5.0	0.166667	Less Elliptical	E2
5	MCG-05-32-053	7.0	2.0	0.714286	Undefined	E7
...	...	...	...	...	...	...
19994	MCG+07-30-065	8.0	1.5	0.812500	Undefined	Not Elliptical
19995	MCG+07-29-049	7.0	4.0	0.428571	Average Elliptical	E4
19996	MCG+07-22-076	6.0	4.0	0.333333	Average Elliptical	E3
19997	MCG+07-21-027	2.5	1.3	0.480000	Average Elliptical	E5
19999	MCG+07-22-053	8.0	7.0	0.125000	Less Elliptical	E1

18395 rows × 6 columns

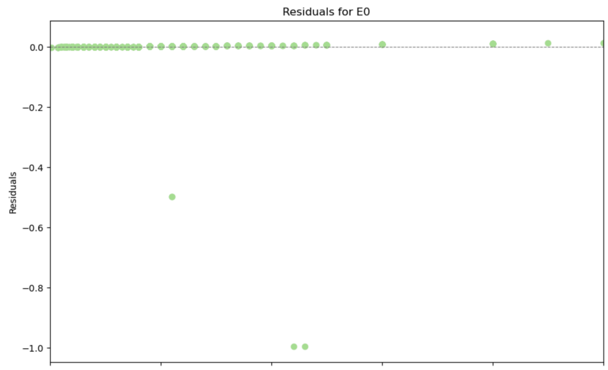
Part 4: Major-Minor Axis Plotting

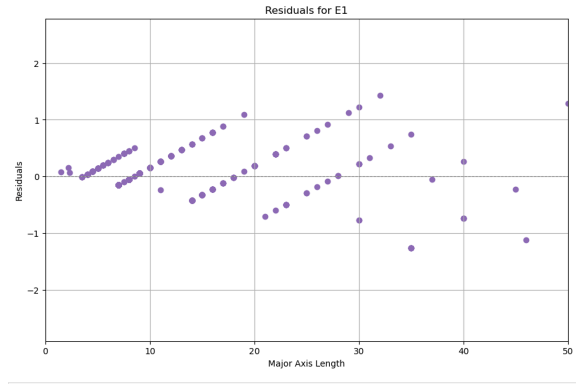


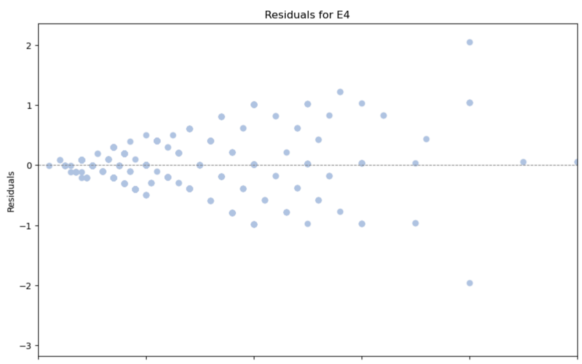
This is our data in a major-minor axis scatter plot. A general linear trend can be seen for each E(n) category, but for more accuracy, we decided zoom in on the data as there is only one outlier in the E5 classification level, and then used apply polynomial regression model.

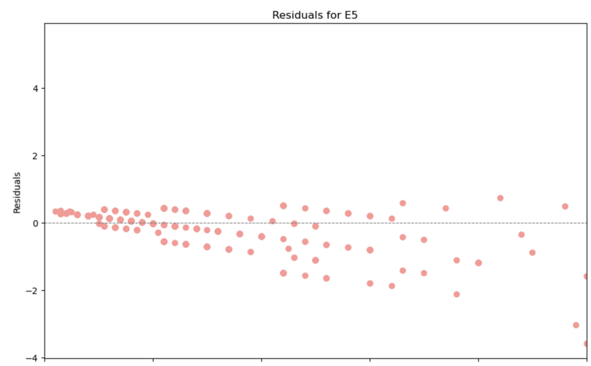


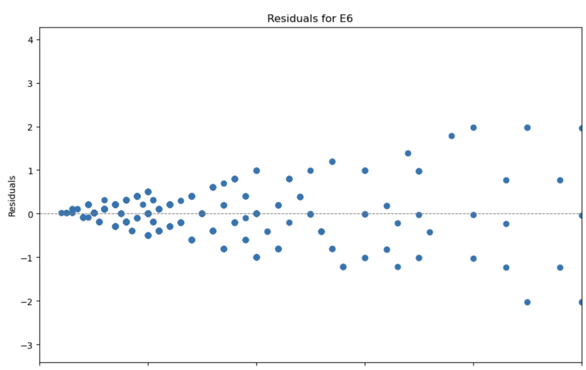
These are our data points now with the fit model visualized onto the graph. Our model seems to be fairly accurate. To confirm that our model is accurate, we also created a residuals plot for each E(n).

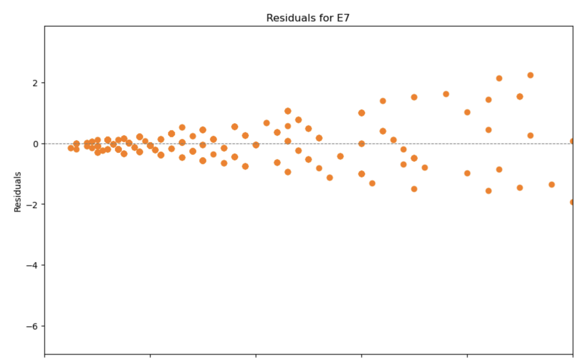

Residuals for E0
Residuals
0.0
−0.2
−0.4
−0.6
−0.8
−1.0


Residuals for E1
Residuals
Major Axis Length
2
1
0
−1
−2
0
10
20
30
40
50


Residuals for E4
Residuals
2
1
0
−1
−2
−3


Residuals for E5
Residuals
4
2
0
−2
−4


Residuals for E6
Residuals
4
3
2
1
0
−1
−2
−3


Residuals for E7
Residuals
2
0
−2
−4
−6

Our residuals tend to expand as predicted major axis increases indicates heteroscedasticity, which suggests that the variance of the residuals is not constant across the range of the major axis length. In other words, the prediction error of the model varies depending on the size of the galaxy; it is not equally accurate for galaxies of all sizes. However, the general randomness is still high and therefore our model is fairly accurate.

Final Conclusion

Our project was able to demonstrate the use of isophotic parameters in the classification of elliptical galaxies. By leveraging equations for ellipticity and isophote ‘n’ values, we were able to categorize galaxies into distinct groups with considerable precision and that closely the actual data. Our analysis supports the validity of our theoretical models, and the successful application of polynomial regression to our major-minor axis plotting further validates the validity of our classification framework.

This project establishes a solid groundwork for the advancement of astronomical classification, particularly in the realm of elliptical galaxies. Future research can build upon this in a few promising directions. Exploring more sophisticated statistical models could refine our understanding of the underlying distributions of these galactic bodies. Incorporating additional datasets, especially from newer surveys or different wavelengths, could provide a more comprehensive view of the characteristics that influence galaxy morphology. Additionally, considering various astrophysical factors, such as environmental influences and evolutionary histories or location, could significantly enhance our ability to interpret the development and classification of elliptical galaxies.