Galaxy Distribution Analysis

Laura F. Rivera Merida

May 6, 2024

Abstract

This project dives into an analysis of galaxy distribution, utilizing a unique dataset personally selected sourced from the Sloan Digital Sky Survey (SDSS). This dataset encompasses a diverse array of parameters, including object ID, type, celestial coordinates, magnitudes across multiple spectral bands (u, g, r, i, z), and pivotal redshift values. The primary objective entails meticulously computing the redshift distribution of galaxies, employing visualization techniques such as histograms or density plots. Through this examination, the aim is to unveil potential peaks or trends that may signify clusters or significant structures within the expansive cosmos. By leveraging the data from the SDSS, this analytical endeavor sheds light on the intricate spatial arrangement of galaxies and provides profound insights into the cosmic architecture and dynamic processes shaping the universe.

1 Introduction

The study of galaxies and their spatial arrangement within the cosmos is fundamental to contemporary astronomy, providing crucial insights into the evolution and organization of the universe. The Sloan Digital Sky Survey (SDSS), renowned for its extensive mapping of the celestial sphere, provides a rich dataset for investigating the distribution and characteristics of galaxies.

This project embarks on a comprehensive examination of galaxy distribution, drawing upon the vast data resources cataloged by the SDSS. Through analysis of parameters such as object type, coordinates, and magnitudes across diverse spectral bands, the objective is to unveil underlying patterns and structures inherent in the cosmic tapestry

Central to this analysis is the exploration of the redshift distribution of galaxies, a key metric indicative of their relative distance and velocity with respect to Earth. Leveraging sophisticated visualization methods such as histograms and density plots, I aim to elucidate the distribution of redshift values and identify potential clustering phenomena.

Table 1: Galaxy location information gathered from the SDSS

objid	specobjid	type	ra	dec	u	g	r	i	z	redshift
1237662301903192107	5.81432e+18	GALAXY	229.527	42.7441	17.4386	15.8068	15.2021	14.9146	14.5714	0.0407244
1237662301903192158	nan	GALAXY	229.565	42.727	24.3383	23.2861	21.4142	20.5625	20.4875	nan
1237662301903192222	nan	GALAXY	229.463	42.6944	21.1118	19.0846	18.1635	17.7624	17.4394	nan
1237662301903192523	nan	GALAXY	229.512	42.815	23.5101	22.0963	20.5471	19.9894	19.8391	nan
1237662301903192667	nan	GALAXY	229.522	42.7612	20.9239	19.0527	18.5264	18.258	18.154	nan
1237662301903192749	nan	GALAXY	229.52	42.708	22.28	22.1788	21.0745	20.5519	20.3794	nan
1237662301903192751	nan	GALAXY	229.542	42.7324	25.1248	22.4224	21.9458	21.4081	20.884	nan
1237662301903192762	nan	GALAXY	229.527	42.7069	25.6808	21.9166	20.6419	20.1563	19.4767	nan
1237662301903192764	nan	GALAXY	229.534	42.7126	23.3253	21.715	20.4104	19.8406	19.6534	nan
1237662301903192793	nan	GALAXY	229.562	42.722	24.2524	22.2399	20.5654	19.8204	19.3811	nan
1237662301903192801	nan	GALAXY	229.599	42.7585	22.8748	22.0585	20.684	20.1099	19.891	nan
1237662301903192806	nan	GALAXY	229.6	42.7543	23.5215	22.4736	21.1121	20.6727	20.0064	nan
1237662301903192822	nan	GALAXY	229.608	42.7553	23.7949	21.4124	20.2396	19.8335	19.6978	nan
1237662301903192827	nan	GALAXY	229.61	42.7577	23.6535	21.6882	19.9697	19.3104	19.0023	nan
1237662499465068599	nan	GALAXY	229.327	42.5637	24.3808	13.4247	13.3112	12.3842	22.9379	nan
1237662499465134188	nan	GALAXY	229.614	42.5951	10.8834	9.44388	9.12844	9.02917	9.89722	nan

2 Data Selection

To ensure the inclusivity and diversity of the dataset, I employed a different approach during the data selection process. Navigating through the SDSS website's data interaction section provided me with a wealth of options to explore. Leveraging the "Navigate" tool, I sampled galaxies across different regions of the celestial sphere. By embracing randomness in my selection process, I aimed to capture a broad spectrum of galactic properties and characteristics. As I curated this virtual "notebook" of galaxies, I continuously assessed the diversity and representativeness of the sampled data. This iterative approach allowed me to refine my selection criteria and ensure the inclusion of a balanced assortment of galaxies spanning various types and magnitudes. Upon reaching a total of 15 galaxies in my "notebook," I extracted the corresponding dataset, thus securing a robust foundation for the subsequent analysis. The information obtained from the website is displayed in 1

3 Data Description

The dataset procured from the Sloan Digital Sky Survey (SDSS) encapsulates an array of astronomical information, offering a comprehensive snapshot of galaxies scattered across the area of the universe this data originated from. Each entry in the dataset comprises several attributes cataloged to delineate the nature of these cosmic entities. Fundamental parameters such as object identification (objid) and spectroscopic object identification (specobjid) serve as unique identifiers, facilitating precise referencing and analysis. Moreover, the dataset encompasses crucial details regarding the type of object under scrutiny, ranging from galaxies to other celestial phenomena. The spatial coordinates of each galaxy, denoted by right ascension (ra) and declination (dec), provide vital positional information essential for mapping the spatial distribution of galaxies within the universe. Additionally, the dataset furnishes photometric measurements across multiple spectral bands (u, g, r, i, z), offering insights into the intrinsic luminosity and spectral characteristics of galaxies. Notably, the inclusion of redshift

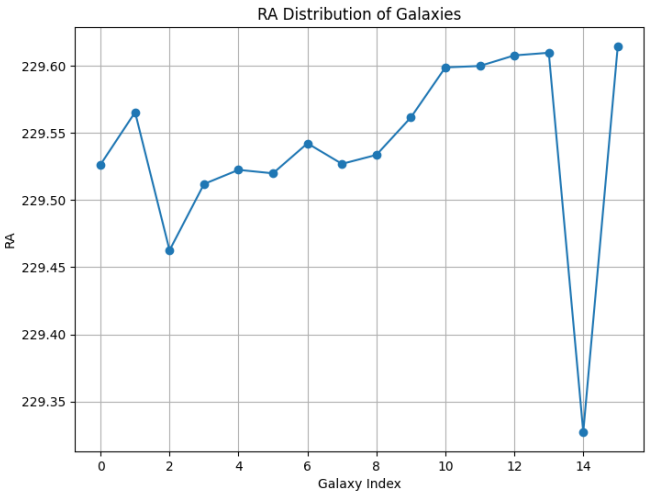

Figure 1: Each data point on the plot represents a galaxy from the dataset. The position of each data point is determined by its RA coordinate (vertical axis) and its corresponding Galaxy Index value (horizontal axis).

values further augments the dataset's utility, enabling astronomers to gauge the recession velocities and distances of galaxies with remarkable precision.

4 Visualization Techniques

The visualization process began with the preparation of data stored in text files, primarily sourced from astronomical observations. Leveraging the command-line interface, raw data was accessed and manipulated using various text processing utilities. Through the terminal, data cleansing and formatting operations were conducted to ensure compatibility with subsequent analysis steps. Following this initial preprocessing phase, Python scripts were employed to orchestrate the visualization pipeline. Leveraging Python's extensive ecosystem of libraries, including but not limited to NumPy, Pandas, and Matplotlib, the data was ingested, transformed, and ultimately rendered into insightful visualizations. Python scripts executed in the terminal environment orchestrated the entire process seamlessly, from data ingestion to the generation of rich graphical representations. This integrated approach facilitated the creation of dynamic and informative visualizations, empowering astronomers to explore and analyze complex astronomical phenomena with precision and clarity.

5 Results

The statistical analysis of the galaxy data reveals intriguing insights into the characteristics of celestial objects within the dataset. The mean u magnitude, representing the brightness in ultraviolet light, is calculated to be 22.32. This value indicates the average intensity of this wavelength across the observed galaxies. Meanwhile, the median r magnitude, which measures the brightness in the red wavelength, stands at 20.48. This statistic provides a robust measure of the central tendency of the data, suggesting that the median brightness in the red spectrum is notably lower than the mean brightness in the ultraviolet spectrum. However, it's important to note that the standard deviation of the redshift is reported as "nan," indicating that there might be missing or undefined values in this attribute. Further investigation into the redshift data is warranted to better understand the distribution and variability of galactic velocities within the dataset.

References