An Analysis of the Relationship between Color and Morphology of Galaxies with Redshift

By Andrew Yu

Author Bio

Andrew Yu is a junior at Plano West Senior High in Plano, Texas. He loves astronomy and anthropography and hopes to study astronomy and astrophysics in college.

Abstract

The research topic derives from my deep curiosity in galaxy evolution. I am asking if there is a correlation between the distance to a galaxy (measured by its redshift) and its color, measured by u-v (ultraviolet magnitude subtracted by visible light magnitude) color index, for galaxies of various morphological types (spirals, ellipticals or irregular shapes). Astronomical data such as redshifts, morphological types, and color indices were queried from the database SIMBAD. Python was used to perform data analysis and to graph data. The best fit function was deployed for polynomial regression with different polynomial degrees to determine the best fitting trend. One-directional trending was not seen throughout, rather, a complex pattern roughly represented by $y = -0.027x^4 + 0.150x^3 + 0.053x^2 - 0.453x + 1.192$ was identified. The downward trend was present only in the redshift ranges $z < 1$ and $z > 3$. The study showed that for each morphological type, there is not a simple correlation between a galaxy's distance and its color index. However, a cross-morphological type comparison indicated irregular shaped galaxies are bluer than spirals, and spirals are bluer than ellipticals, which was the case throughout the entire redshift range.

Keywords: SIMBAD, Galaxy classification, Galaxy morphology, redshifts, color index, polynomial regression, python, star formation

Introduction

Galaxy morphology is used by astronomers to categorize galaxies based on their shapes. The most famous classification is called the Hubble sequence, visualized in the Hubble tuning fork, categorizing galaxies into three categories. There are spirals, which have a structure characterized by a disk of spiral arms, which have active star formation regions, and a bright, glowing bulge at the center without the necessary gas and dust to create new stars. Other galaxies, called ellipticals, are less structured: simply an elliptical halo of stars orbiting around a compact center. Ellipticals do not form stars in great quantities. In between the two is a category called lenticular galaxies, which are not purely spiral or elliptical but have similar characteristics to both. Finally, the last category is irregular galaxies, which as the name implies, do not fit in the first three categories. They usually have properties similar to the spiral arms of spiral galaxies, having high star formation rates (Buta et al., 2015).

Many researchers studied the Hubble Sequence with the observations of star formation rates (SFR) based on integrated light measurements in the ultraviolet (UV), far-infrared (FIR), or nebular recombination lines, also simply known as the color of the galaxies (Kennicutt, 1998). Color is important to a galaxy because it can help reveal the populations of stars in the galaxy and the galaxy's star formation rate. High-mass stars are brighter and bluer than low-mass stars, which are red and dim. Luminosity is proportional to mass to the 3½ power ($L \alpha M^{7/2}$), and since the amount of time a star shines is roughly its mass divided by luminosity, the higher the mass a star is, the shorter the amount of time it will live. Combined with the fact that blue stars are usually brighter than red stars, this means the blueness of a galaxy indicates the galaxy's star formation rate. Established astronomical research has indicated there is a strong correlation between color and galaxy morphology (Smethurst et al., 2021; Gusev et al., 2015). Spiral and irregular galaxies tend to be bluer than elliptical and lenticular galaxies (Skibba et al., 2009), meaning that spirals and irregulars undergo more star formation.

As we examine galaxies at very far distances, we look back into time, effectively allowing a glimpse of the earlier days of the universe through these distant galaxies. Data found on multiple distant galaxies enables us to study their evolution. A form of measuring distance, known as redshift, has become popular. Redshift is the lengthening of light waves due to the motion of the emitting galaxy away from the Earth because of the expansion of the universe, making its light redder than it would appear if it were not moving. It is calculated by dividing the observed wavelength by the known wavelength at certain critical emission lines, and then subtracting 1. For example, if a galaxy's hydrogen-alpha emission lines were observed to be triple the wavelength observed in our own galaxy, the redshift of that galaxy would be equal to 2. At very far distances, which this paper includes, the theory of special relativity must be used to accurately determine the recessional velocity, in the following formula: $z = \sqrt{((1+v/c)/(1-v/c))}$, where v is the velocity that the galaxy is moving away from Earth.

While extensive analysis on galaxy colors at all galaxy morphologies and redshifts have certainly been done, the big picture is still far from complete. For one, little is known about the internal workings of galaxies in the early era of the universe, and what caused such transitions of star formation to form. New discoveries, especially at further redshifts, could be made, allowing a clearer understanding of how our universe today came to be. Though some ideas about the process can be gathered from spectra (Kewley et al., 2019), galaxy evolution remains opaque to the observer.

The research question derives from the author's curiosity in galaxy evolution, asking if there is a correlation between the distance to a galaxy (measured by redshift) and its color for galaxies, and if this varies by morphology.

I asked the question: "Do galaxies become bluer at higher redshifts?", with my hypothesis being that when the distance (redshift) of a galaxy increases, the blue-red color index of the galaxy, adjusted for redshift, will also increase. As the redshift of a galaxy increases, its color index should decrease (becoming bluer), regardless of the morphological type. We expect that after a certain distance from Earth, galaxies of all morphological types will become bluer, because at some point in the past, they needed to be forming stars at a faster rate than now to account for all of the older stars today.

Materials And Methods

For the data in this research project, I used the Set of Identifications, Measurements and Bibliography for Astronomical Data (SIMBAD), a database of extragalactic objects. I chose to use SIMBAD to find the collection of galaxies because it includes data on morphological type, redshift, and color. In addition to SIMBAD, I considered using the Sloan Digital Sky Survey (SDSS), as it is one of the most thorough astronomical surveys conducted, and the NASA/IPAC Extragalactic Database (NED), one of the most comprehensive surveys for intergalactic objects. Eventually, I determined that the SIMBAD data was most accessible to begin addressing my question.

I queried SIMBAD with galaxy measurements of morphology, redshifts, and color index. I used the Hubble morphological scheme for galaxy classification of shape: elliptical, spiral/barred spiral, irregular, and then processed each classification separately. In total, 1603 galaxies were queried from the SIMBAD database and used in the analysis, which is only a tiny fraction of the total number of galaxies in the database, as most did not have the necessary data needed to calculate the u-v color index.

Table 1 displays partial data with header names ID, Galaxy Identifier, galaxy type, redshift, U/B/V color index value and MT (Morphological Type).

#	Galaxy Identifier	type	redshift	Mag U	Mag B	Mag V	MT
1	3C 65	rG	1.176	~	23	~	E D ~
2	Cowie 101	G	1.92	25	24.1525	24.82	I D ~
3	NAME SMM J123652+621354	G	1.355	~	22.4	22.3	I D ~
4	Cowie 57	G	1.145	23.33	23.4712	24.5	Sc/I: D 2000AJ....120. 2190V
5	SHARDS J123656.59+621252.8	G	1.233	~	24.3511	~	Sbc D ~
6	HNM 1215	G	1.84	27.9	26.9	26.4	Scd D ~
7	[EAD2001] HDFN J123653.43+621221.7	G	1.715	24.64	24.1405	24.67	I D ~
8	MODS deep 1627	G	1.24	25.2	25.24	25.22	I D ~
9	TKRS 8664	EmG	1.02	24.35	24.0528	24.24	Sap D 2000AJ....120. 2190V
10	GOODS J123648.64+621216.1	rG	1.0662	25.46	26.3837	25.8	Sbc D ~
11	[FLY99] 544	G	1.36	~	~	~	I D ~
12	PEARS n45983	EmG	1.06	23.84	24.0302	25.28	?p D 2000AJ....120. 2190V
13	[FLY99] 494	G	1.28	~	~	~	I D ~
14	TKRS 7589	G	1.04	26.4	25.6065	25.53	Scd D ~
15	MODS deep 813	G	1.12	26.5	26.5	26.8	I D ~
16	MODS deep 906	G	1.28	26.2	26.4	26.3	I D ~
17	HNM 937	G	1.92	26.3	25.84	26	I D ~
18	Cowie 100	G	1.0188	23.76	23.4996	25.24	I D 2000AJ....120. 2190V
19	[CBH2004] 189.166748+62.20214	G	1.0166	25.22	24.7294	24.18	Sbc D ~
20	SHARDS J123639.44+621211.8	EmG	1.18	26.8	26.3572	25.8	Sbc D ~
21	TKRS 7460	G	1.0499	24.07	24.3776	24.73	Sb: D 2000AJ....120. 2190V
22	MODS deep 2020	G	1.64	25.7	25.73	25.91	I D ~
23	MODS deep 1481	EmG	1.01	26.9	27	27	I D ~
---	---	---	---	---	---	---	---

Table 1: Raw data (total 1603 galaxies with sufficient u-v index data) from SIMBAD. I filtered the data into galaxies, and then within galaxies, looking for magnitude in near-ultraviolet (300-400 nm) and visible light roughly corresponding to yellow-green (500-600 nm), abbreviated u and v; the measure of redshift, abbreviated as z; and morphological type. SIMBAD collected this data through color filters letting the aforementioned bands of a certain wavelength through. I categorized morphological types into spirals, ellipticals, and irregulars.

The first step was to find the u-v color: this color index measures how much brighter the yellow-green visible light is than the near ultraviolet light, and has been used to determine "redness" previously, such as in (Zhang and Deng, 2015). This method of color index subtracts the "visible" magnitude from the "ultraviolet" magnitude. For example, a galaxy with "visible" magnitude 14 and "ultraviolet" magnitude 16 would have a u-v color index of 2. Figure 1 shows the rough passbands of the SIMBAD database U and V color filters, the two magnitudes used to calculate u-v color index.

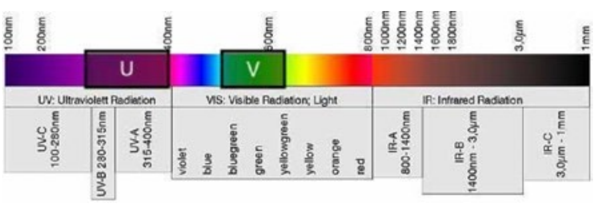

FIGURE 1: The rough passbands of the SIMBAD U and V color filters. Adapted from (Wavelength ranges of electromagnetic radiation)

The second step was to compare the u-v index against the redshift for a given galaxy. As redshift is a function of distance, I used redshift as a stand-in for distance on the x-axis, so in essence, I graphed color by distance. I then broke the data into spiral, elliptical, and irregular galaxies to compare the color by distance relationship for each galaxy type.

I conducted data analysis in Python 3 using Jupyter Notebooks instead of an IDE to run the code, as running chunks of code and visualizing the results right away is only possible with the former. I used the Python library Pandas for my manipulation of the data file and Matplotlib to make my histograms and scatter plots. I also used the NumPy library to handle data analysis.

The built-in best fit function in NumPy was used for linear/polynomial regression. Different polynomial degrees were applied to determine the best fit by finding the least square. Matplotlib was used for graphing. All polynomials referenced use my specific x = redshift and y = u-v color index graph axes. Figure 2 shows how a best fit curve is calculated. A polynomial is chosen which minimizes the sum of the squared deviations in the y-direction between the calculated value of the polynomial across all data points.

The solution minimizes the squared error

$$E = \sum_{j=0}^{k} |p(x_j) - y_j|^2$$

in the equations:

```
x[0]**n * p[0] + ... + x[0] * p[n-1] + p[n] = y[0]
x[1]**n * p[0] + ... + x[1] * p[n-1] + p[n] = y[1]
...
x[k]**n * p[0] + ... + x[k] * p[n-1] + p[n] = y[k]
```

FIGURE 2: Polynomial coefficient calculation

As shown in Figure 3, I first used polynomial regression with degree of 1 for all 1603 galaxies (drawn as dots in the chart) with X axis as redshift and Y axis as U-V index for each galaxy. The red line shows a fit equation y = 0.192x + 1.036. Clearly, a 1st degree linear polynomial does not reveal the correlation between red shift and u-v index. To investigate further, 2nd and 3rd degree polynomials were used to analyze with the same data, and results are shown in Figure 4 and Figure 5. With the increase of polynomial degree, some trends and correlations are shown more clearly.

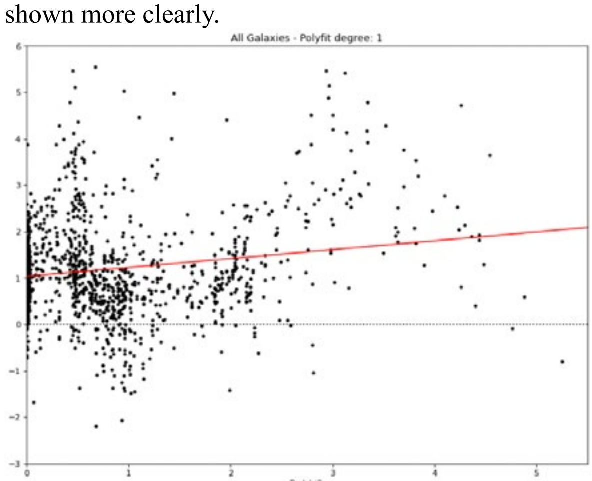

FIGURE 3: Polynomial regression of full galaxy data set with degree 1

Best fit line equation: y = 0.192x + 1.036

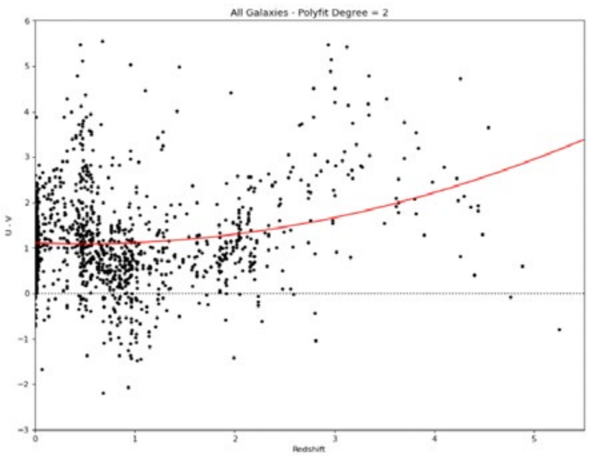

FIGURE 4: Polynomial regression of full galaxy data set, degree 2

Best fit equation: $0.091x^2 - 0.089x + 1.109$

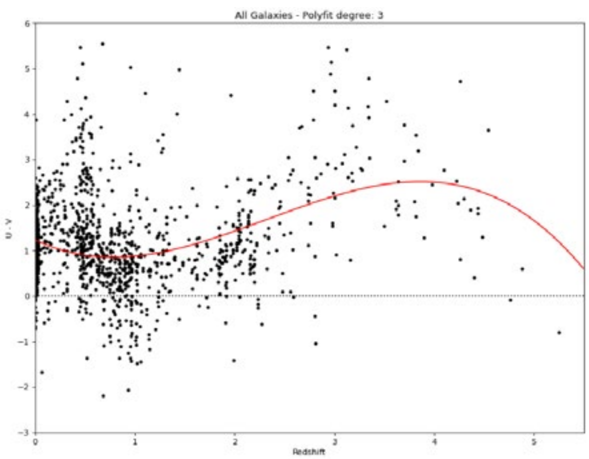

FIGURE 5: Polynomial regression of full galaxy data set, degree 3

Best fit equation: $-0.114x^3 + 0.798x^2 - 1.043x + 1.235$

Results

The data consisted mostly of low redshift (z < 1) galaxies, however, there were enough galaxies for a useful analysis up until z = 3, with sparse data after that. Figure 6 contains the final results which best reveal the correlations between U-V index, red shift and galaxy morphology types . In this chart, I categorize galaxies with different morphological types with different color and shape. Red triangles represent irregular galaxies, blue stars represent spiral galaxies, and orange dots represent elliptical galaxies. A 4th degree polynomial was drawn with the best fit function $y = -0.027x^4 + 0.150x^3 + 0.053x^2 - 0.453x + 1.192$, shown as the black line. The same degree was used to draw fit lines for the individual morphological types: a red semi-dashed line for irregular galaxies, a blue dotted line for spiral galaxies, and an orange dashed line for elliptical galaxies.

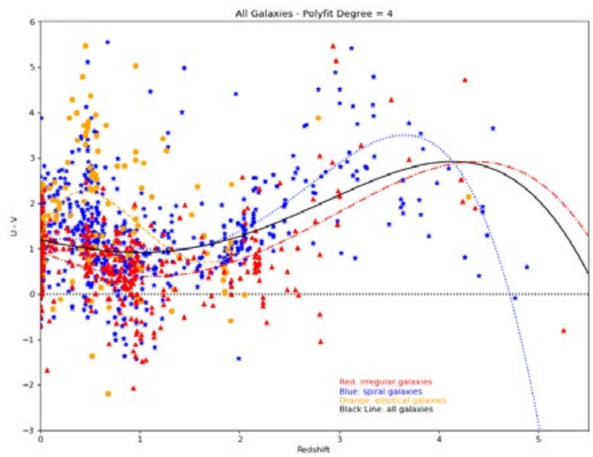

FIGURE 6: Polynomial regression for all galaxies with degree 4, separated by morphological type

I found several interesting yet inconclusive trends in the data.

Between z = 0 and about z = 0.2, there is an upward trend in the u-v value, which shows an increased reddening as galaxies get farther away. Elliptical galaxies also appear to be redder than spirals and ellipticals at this range.

Between z = 0.4 and z = 0.8, there is a vague downward trend in the u-v color index, showing that these galaxies become less red with increasing distance.

Between z = 1 and z = 3, the trend reverses, and u-v color index once again shows a positive trend, so galaxies once again become redder with increasing distance.

After z = 3, the color index appears to trend downwards again, meaning that galaxies this far get bluer with increasing distance. Because of the small number of data points in this range magnifying any inaccuracies, I suspect this is an artifact of the data rather than a true trend.

Discussion

The results are proven not as simple as a downward trending direction as a function of distance. Based on the results, one-directional trending was not seen throughout, rather, a complex pattern roughly represented by $y = -0.027x^4 + 0.150x^3 + 0.053x^2 - 0.453x + 1.192$ was identified.

At low redshift ($z < 0.2$), the upward trend shows that galaxies do get redder with distance, which is likely due to the effects of redshift itself altering the u-v color index value.

Between $z = 0.4$ and $z = 0.8$, the downward trend aligns with my initial hypothesis, but it is not clear why this specific bound has a downward trend and not before or after it.

The upward trend resumes between $z = 1$ and $z = 3$, which may again be due to redshift or some other unexplained factor.

After $z = 3$, if the downward trend is truly valid and not due to a data collection artifact, this may also align with my hypothesis about the early universe having younger and bluer stars.

The study showed that for each morphological type, there is not a simple one directional correlation between a galaxy's distance and its color index. However, cross morphological type comparison indicated irregular shaped galaxies are bluer than spirals, and spirals are bluer than ellipticals.

With several trends present in the data, the hypothesis was proven to be not entirely correct. While I was expecting mostly graphs trending in one direction throughout, the redshifts displayed a complex pattern roughly represented by the polynomial previously given. The downwards trend that I was expecting was present only in the redshift ranges $0.2 < z < 1$ and $z > 3$. This was a trend that held up across all three morphologies, though data for ellipticals were so sparse past $z = 2$ that accurate conclusions cannot be drawn for them. Interestingly, a similar trend can be seen in the reference papers, so further investigation will be required to confirm reasons for these patterns.

Finally, my initial hypothesis on irregulars being bluer than spirals, and spirals being bluer than ellipticals, turned out to be correct. My findings fit into the conclusion on correlation between color and galaxy morphology by other researchers using similar methods (Smethurst et al., 2021; Gusev et al., 2015). A similar search found spiral and irregular galaxies to be bluer than elliptical and lenticular galaxies (Skibba et al., 2009).

My findings showed a few preliminary relationships, namely the positive trend at $z < 0.2$ and $1 < z < 3$, and the negative correlation at $0.4 < z < 0.8$, but further investigation is needed to prove these trends. Further research could compare more specific types of galaxies (comparing different types of spirals, ellipticals, or irregular galaxies in addition to lenticular galaxies as opposed to just the three main types) and a broader range of color indices (including blue and infrared filters, in addition to the ultraviolet and visible filters considered here).

Finally, I would welcome a dataset from a different source (such as SDSS or NED) with a greater quantity of galaxies with very high redshifts, as some researchers did on other databases (Smethurst et al., 2021; Zhang and Deng, 2015). The lack of explanation for the findings is not conclusive, and I look forward to continued research.

Conclusion

To answer my question on "Do galaxies become bluer at higher redshifts (as increasing distance from us)", with my hypothesis being that after a certain distance from Earth, galaxies of all morphological types will become bluer, the result shows my hypothesis is not as simple as a downward trend direction. More research can be conducted by analyzing more astronomical data, especially with high redshift galaxies. Furthermore, future research can be expanded using other methods as such:

Use different color indices like u - g or g - v to determine if there are any differences in trends.

Use machine learning libraries like Pytorch to predict properties of galaxies through already identified trends.

Use software to simulate evolution of galaxies themselves and watch if it matches up to the data collected.

Possibly, extend color indices to other areas of the electromagnetic spectrum (gamma rays, X-rays, infrared), which could tell us additional information about galaxy evolution.

Acknowledgments

I would like to thank Dr. Shyamal Mitra for his mentorship and guidance at the University of Texas High School Research Academy. I would like to thank my research partner, William Coury, for contributing a large portion to the data collection and programming. I would like to thank the peers and mentors of the University of Texas High School Research Academy for the help and providing excellent resources.

References

Buta, R.J., Sheth, K., Athanassoula, E., Bosma, A., … & Madore, B.F. (2015) A Classical Morphological Analysis Of Galaxies In The Spitzer Survey Of Stellar Structure In Galaxies, The Astrophysical Journal Supplement Series, 217(32)

Gusev, A.S., Guslyakova, S.A., Novikova, A.P., … & Ezhkova, O.V. (2015) Studies of the Stellar Populations of Galaxies Using Two-Color Diagrams, Astronomy Reports, 59, 899-919

Kennicutt, R.C. Jr. (1998). Star Formation In Galaxies Along The Hubble Sequence. Annual Review of Astronomy and Astrophysics, 36, 189–231

Kewley, L., Nicholls, D., & Sutherland, R. (2019) Understanding Galaxy Evolution through Emission Lines, Annual Review of Astronomy and Astrophysics, 57, 511-570

Skibba, R., Bamford, S., Nichol, R., … & Vandenberg, J. (2009) Galaxy Zoo: disentangling the environmental dependence of morphology and colour, Monthly Notices of the Royal Astronomical Society, 399, 966-692

Smethurst, R.J., Masters, K.L., … & Walmsley, M. (2021) Quantifying the Poor Purity and Completeness of Morphological Samples Selected by Galaxy Colour, Monthly Notices of the Royal Astronomical Society, 510 (3), 4126–4133

Wavelength ranges of electromagnetic radiation, accessed 5 March 2023, <https://light-measurement.com/wavelength-range>

Zhang, F., & Deng, X., (2015) u-r color dependence of galaxy clustering in the main galaxy sample of SDSS DR10, Astrophysics, 58, 21-28