The Relationship between Color and Morphology of Galaxies with Redshift

By Andrew Yu

AUTHOR BIO

Andrew Yu is a student at Jasper High School in Plano, Texas. This paper derives from the research at the High School Research Academy at the University of Texas at Austin over the course of the summer of 2022, and from science fair research in the school year afterwards. Andrew is an enthusiast in astronomy, math and computer science. He is the president of Teen TAS, a student organization part of the Texas Astronomical Society intended to stimulate interest and passion for student astronomers through group and self-led activities in astronomical observation, research and community outreach. Andrew is a qualifier and top 50 finisher of the National Astronomy Competition of USAAAO (US Astronomy and Astrophysics Olympiad) and a five-time qualifier for the AIME (American Invitational Mathematics Exam). He also enjoys sharing his knowledge in astronomy and math by regularly teaching younger students.

ABSTRACT

The research topic derives from the author's deep curiosity in galaxy evolution. He is 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 and calculated. 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 originally hypothesized 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, morphological types, 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).

Color

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 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 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 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 included 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.

	A	B	C	D	E	F	G	H	I	J
147	150	LEDA 747555	AGN	0.10378	18.83	17.699	17.3	S E 2011ApJ...727...83M		
148	151	EIS-DEEP CDI	EmG	0.4584	22.7651	21.764	20.2037	Sab D 2014ApJ...794..156R		
149	152	[HSN2014] 6	AGN	0.0874	21.25	19.64	19.11	S E 2011ApJ...727...83M		
150	153	2dFGRS TGS	EmG	0.0746	19.78	18.2213	17.419	Sab D 2014MNRAS.437.1698M		
151	154	2MASX J0332	EmG	0.07593	19.55	17.6891	16.891	S E 2011ApJ...727...83M		
152	155	ACS-GC 9004	QSO	0.454	22.0732	21.9232	21.497	Im D ~		
153	156	CXOCDFS J03	EmG	0.4345	23.3276	22.7276	21.53	S D 2014ApJ...794..156R		
154	157	NGC 1317	GiP	0.006438	12.2	11.78	11.02	S0a:r D ~		
155	158	NGC 1326	GiG	0.004584	11.69	11.43	10.54	SBa D ~		
156	159	NGC 1365	Sy1	0.005476	10.48	10.08	9.63	SB(s)b D 2013MNRAS.428.1927C		
157	160	NGC 1316	GiP	0.005911	9.81	9.15	8.53	E D 2014MNRAS.437..273K		
158	161	NGC 1393	GiG	0.007308	13.41	13	12.02	S0 D ~		
159	162	NGC 1383	GiG	0.006538	13.81	13.45	12.47	S0 D ~		
160	163	NGC 1400	EmG	0.001769	12.48	12.01	10.96	E1/S0 D 2014MNRAS.442.1003P		
161	164	NGC 1394	GiG	0.014126	14.28	13.66	12.81	S0 D ~		
162	165	NGC 1391	GiG	0.014457	14.79	14.25	13.34	Sa D ~		
163	166	NGC 1385	EmG	0.00501	11.28	11.5	10.94	SBc D ~		
164	167	NGC 1375	GiG	0.002442	13.5	13.02	12.4	SAB0 D 2013MNRAS.431.3060E		
165	168	NGC 1373	GiP	0.004573	14.44	14.03	13.26	Sa D ~		
166	169	NGC 1374	GiP	0.004443	12.46	11.97	11.08	E3 D 2013AJ....146...67B		
167	170	NGC 1389	GiP	0.003336	12.8	12.39	11.5	S0 D ~		
168	171	NGC 1381	GiP	0.005917	12.9	12.47	11.5	S0 D ~		
169	172	NGC 1382	GiP	0.006091	13.19	13.79	12.92	S0 D ~		
170	173	ESO 358-33	GiG	0.00524	13.63	13.36	12.41	S0a D ~		
171	174	NGC 1379	GiP	0.004563	12.17	12.04	10.91	E0 D 2014MNRAS.437..273K		
172	175	NGC 1387	GiP	0.004229	12.18	11.75	10.69	S0 D 2014MNRAS.437..273K		
173	176	NGC 1386	GiP	0.002905	12.42	12.16	11.23	Sa D ~		
174	177	NGC 1380	GiP	0.00616	11.32	10.94	9.93	S0 D ~		
175	178	NGC 1404	GiP	0.006498	11.53	10.69	10	E1 D 2013MNRAS.435.2274W		
176	179	NGC 1399	BiC	0.004755	11.05	9.74	9.59	E1 D 2013AJ....146..160R		
177	180	NGC 1482	GiP	0.006331	13.15	13.14	12.15	SA0+pec D 2013MNRAS.431.1823M		
178	181	ESO 483-13	G	0.003	14.91	14.23	13.61	9 D 2018MNRAS.479.4136K		
179	182	NGC 1592	GiP	0.003158	13.42	13.94	14.09	Ir D 1996MNRAS.278.1025L		
180	183	ESO 486-21	G	0.002776	14.89	14.41	14.03	9 D 2018MNRAS.479.4136K		
181	184	NGC 1744	LSB	0.002477	12.43	11.9	11.41	7 D 2018MNRAS.479.4136K		
182	185	NGC 1800	G	0.00272	13.51	13.19	12.7	9 D 2018MNRAS.479.4136K		
183	186	ESO 362-9	G	0.003097	13.66	13.05	12.67	dG/Ir D ~		

FIGURE 1: Raw data from SIMBAD

I filtered the data onto 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 calculation 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, so for example a galaxy with "visible" magnitude 14 and "ultraviolet" magnitude 16 would have a u-v color index of 2.

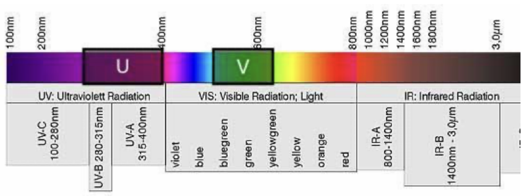

FIGURE 2: The rough passbands of the SIMBAD U and V color filters. Adapted from (Wavelength ranges of electromagnetic radiation)

The second calculation compared 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.

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 3: Polynomial coefficient calculation

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. In total, exactly 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.

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 about $z = 0.4$ 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 $z = 3$, despite the small sample size at this vast distance, 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, I suspect this is an artifact of the data rather than a true trend.

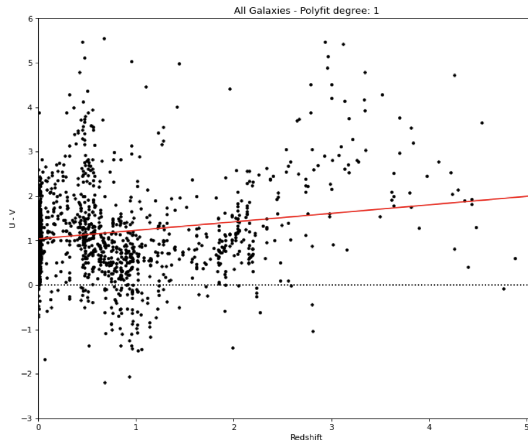

FIGURE 4: Polynomial regression of full data set, degree 1

Best fit line equation: $y = 0.192x + 1.036$

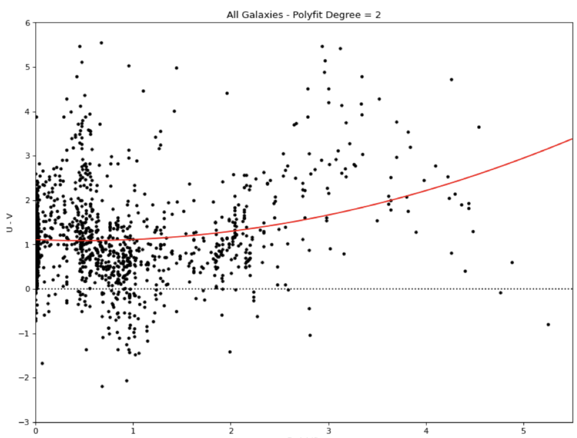

FIGURE 5: Polynomial regression of full data set, degree 2

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

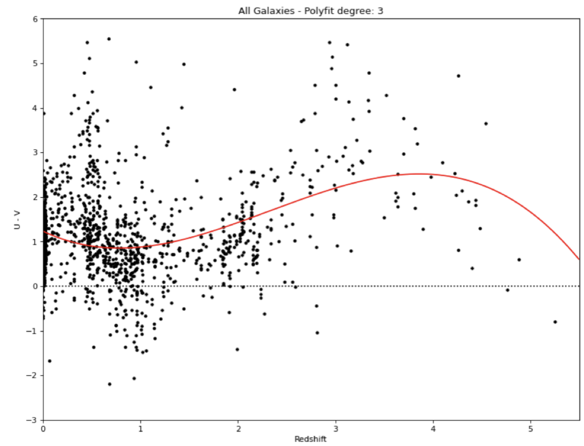

FIGURE 6: Polynomial regression of full data set, degree 3

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

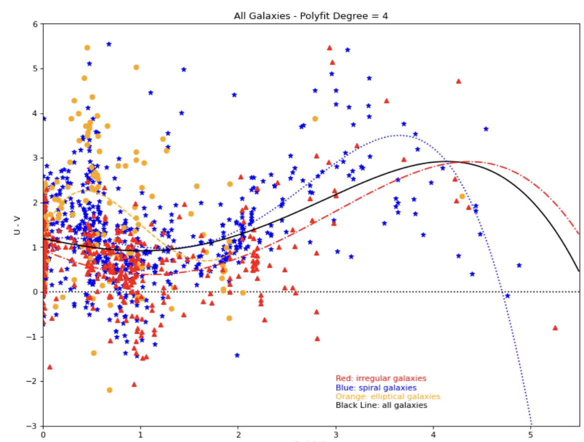

FIGURE 7: Polynomial regression, degree 4, separated by morphological type

DISCUSSION

The results are proven not as simple as a downward trending direction. Based on the above charts, 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 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 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. The lack of explanation for the findings is not conclusive, and I look forward to continued research.

Further research can be conducted 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.

REFERENCES

1. 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)
2. Kennicutt, R.C. Jr. (1998). Star Formation In Galaxies Along The Hubble Sequence. *Annual Review of Astronomy and Astrophysics*, 36, 189–231
3. 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
4. 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
5. Kewley, L., Nicholls, D., & Sutherland, R. (2019) Understanding Galaxy Evolution through Emission Lines, *Annual Review of Astronomy and Astrophysics*, 57, 511-570
6. 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
7. Zhang, F., & Deng, X., (2015) u-r color dependence of galaxy clustering in the main galaxy sample of SDSS DR10, *Astrophysics*, 58, 21-28
8. Wavelength ranges of electromagnetic radiation, accessed 5 March 2023, <https://light-measurement.com/wavelength-range>.