Measuring the Redshift of a Galaxy

ASTR 3800 Final Project

Author: Matthew Senseman

in collaboration with Dequan Foreman

Scientific Motivation for Study

In this study, we wanted to determine the redshift of a single galaxy the using Sloan Digital Sky Survey (SDSS) wavelength and flux data, evaluate the uncertainty in the found redshift measurement, and determine how this uncertainty scales with number of spectral lines used in the evaluation. Finding the redshift in a galaxy's spectrum is further used to evaluate the distance to the galaxy. Then, with the determined galaxy distance, the expansion rate of the universe can be calculated using Hubble's Law. It is beneficial for scientists trying to narrow down the uncertainty on Hubble's constant to have as many measurements of galaxy redshifts as possible. As more galaxy redshifts are determined from surveys like the SDSS, uncertainty in Hubble's constant should be reduced.

Description of the Data Used

The loaded in data from our SDSS file included measurements of the observed galaxy's flux and the observed wavelength of these flux values. The flux is the integrated galaxy light corresponding to the observed galaxy disc. Also used in this project was a data file containing the name and wavelength of potential emission lines that could be found in our galaxy flux spectrum.

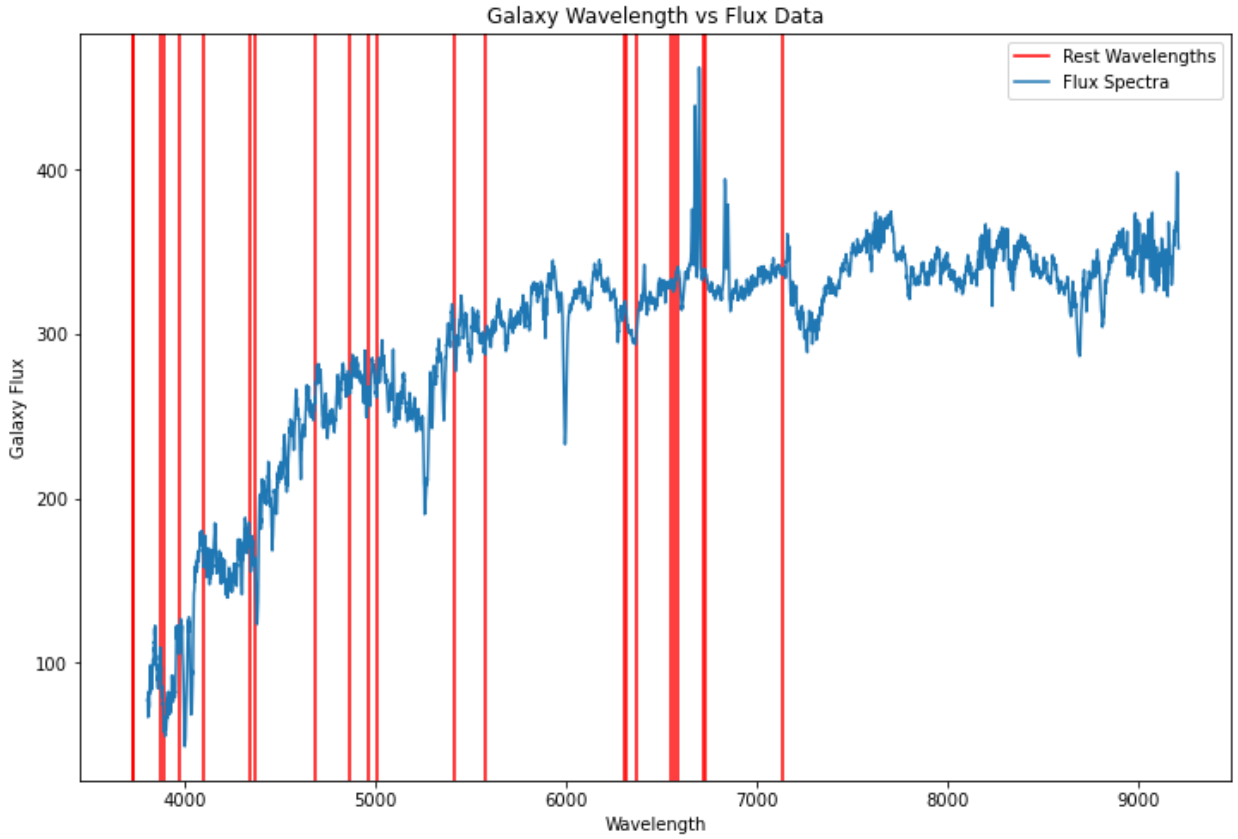

Fig 1. A line plot of the flux data vs wavelength, overplotted are the potential emission lines that could be found in the flux spectrum. Notice that several peaks in the flux spectrum are a little offset from the potential emission lines.

How the Data was Analyzed

The goal for using this data was to determine which of the peaks in the flux spectrum line up with the potential emission lines to determine the redshift of the galaxy. Initially, the polynomial trend in the flux spectrum (seen in figure 1) was removed such that the correlation of the spectrum with the emission lines could be seen more easily.

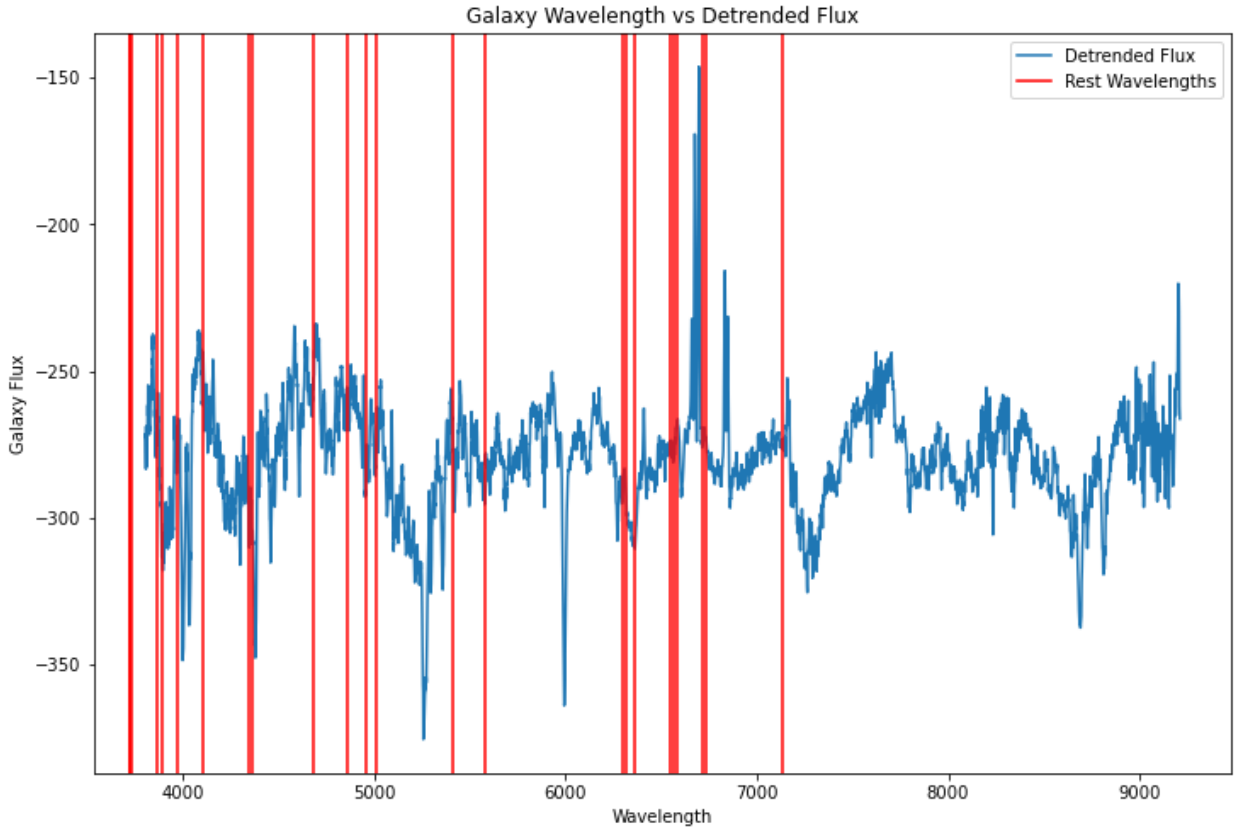

Fig 2. A line plot of the flux data vs wavelength, overplotted are the potential emission lines that could be found in the flux spectrum. This plot has the polynomial trend removed from the flux so peaks in the flux can be easily compared to the potential emission lines.

In order to find which peaks in the flux spectrum lined up with a potential emission line, a cross-correlation function was applied. The cross-correlation function needs two functions to compare where they line up best. But because the potential emission lines were a one dimensional array, a second dimension of flux was added to create an array of the same length as the galaxy flux spectrum.

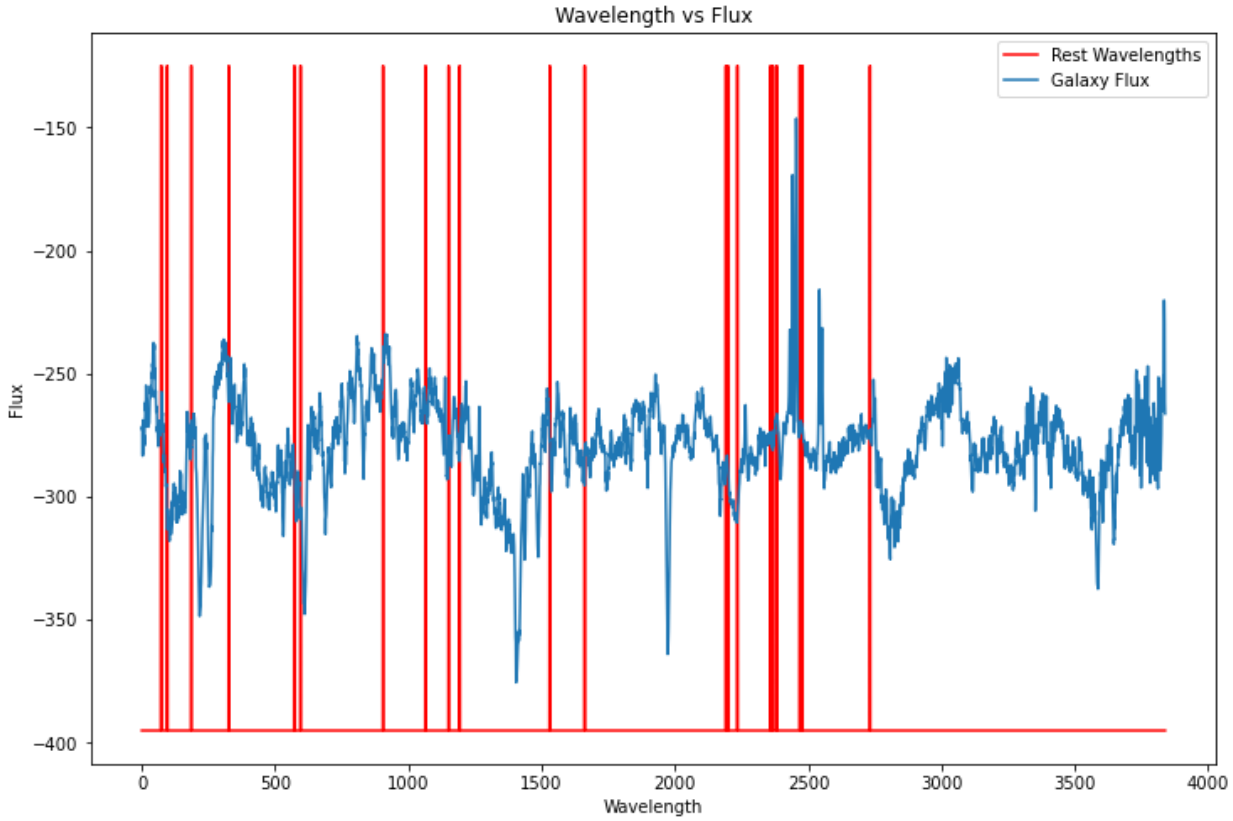

Fig 3. Two line plots of the flux spectrum data vs wavelength and the potential emission lines now as a line plot. The potential emission line plot has a peak flux of -125 at the desired wavelengths and the rest of the array has flux at -395.

Then, using both arrays, the cross-correlation function was performed. The index of maximum cross-correlation was then used to find the shift needed to best line up the peaks with the potential emission lines. This shift is taken as the global $\Delta\lambda$, and was primarily necessary so that we could visually see which peaks in the galaxy spectrum were close to rest emission lines.

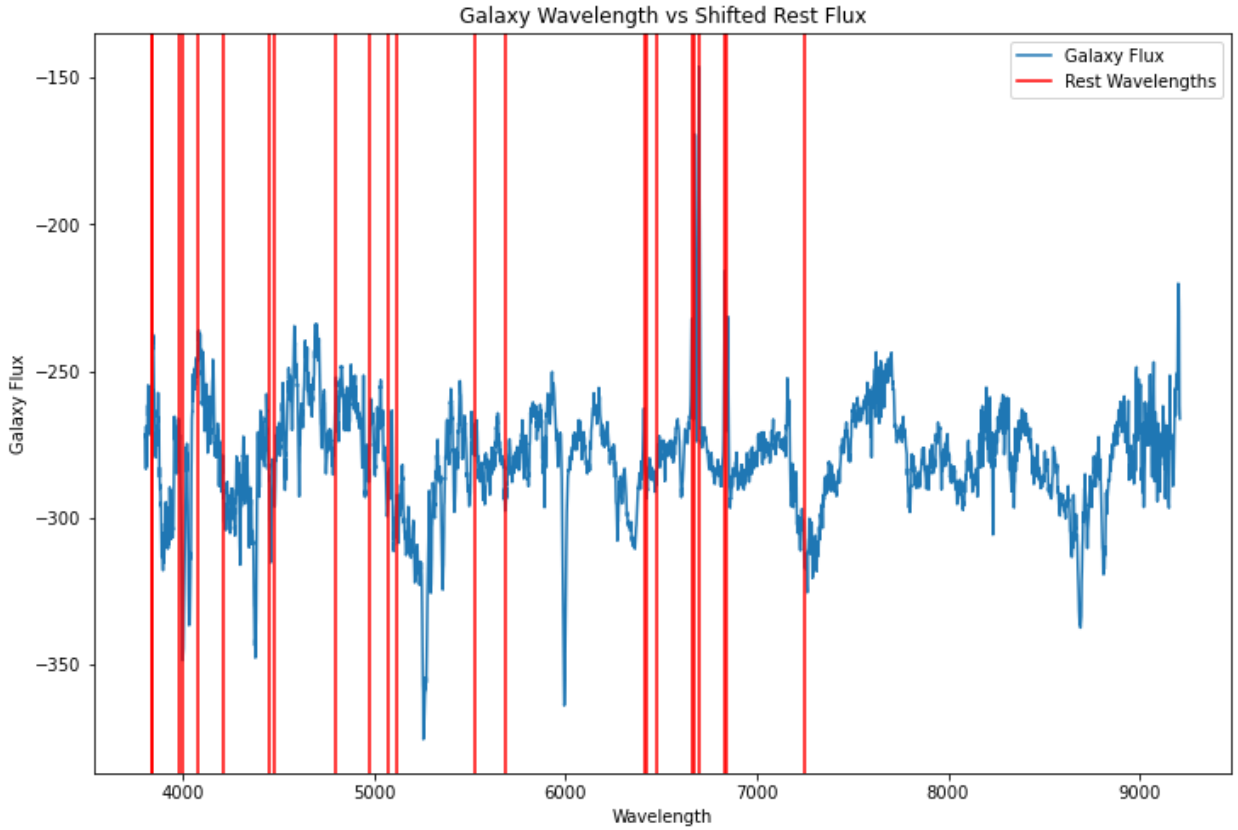

Fig 4. Similar to figure 2, this line plot of the flux spectrum data vs wavelength and the potential emission lines. But this figure has the potential emission lines shifted to the right by the calculated Δλ.

Looking at figure 4, I identified 7 peaks that appeared to have the highest potential as being emission lines in the spectrum when compared to the shifted rest wavelengths. It was then necessary to determine the true values of these 7 peak wavelengths. Using these values of the 7 peak wavelengths along with the values of the corresponding 7 rest wavelengths, the true Δλ values were determined. Δλ values went as the absolute value of the difference between the rest wavelength and the wavelength value of the peak in the galaxy flux.

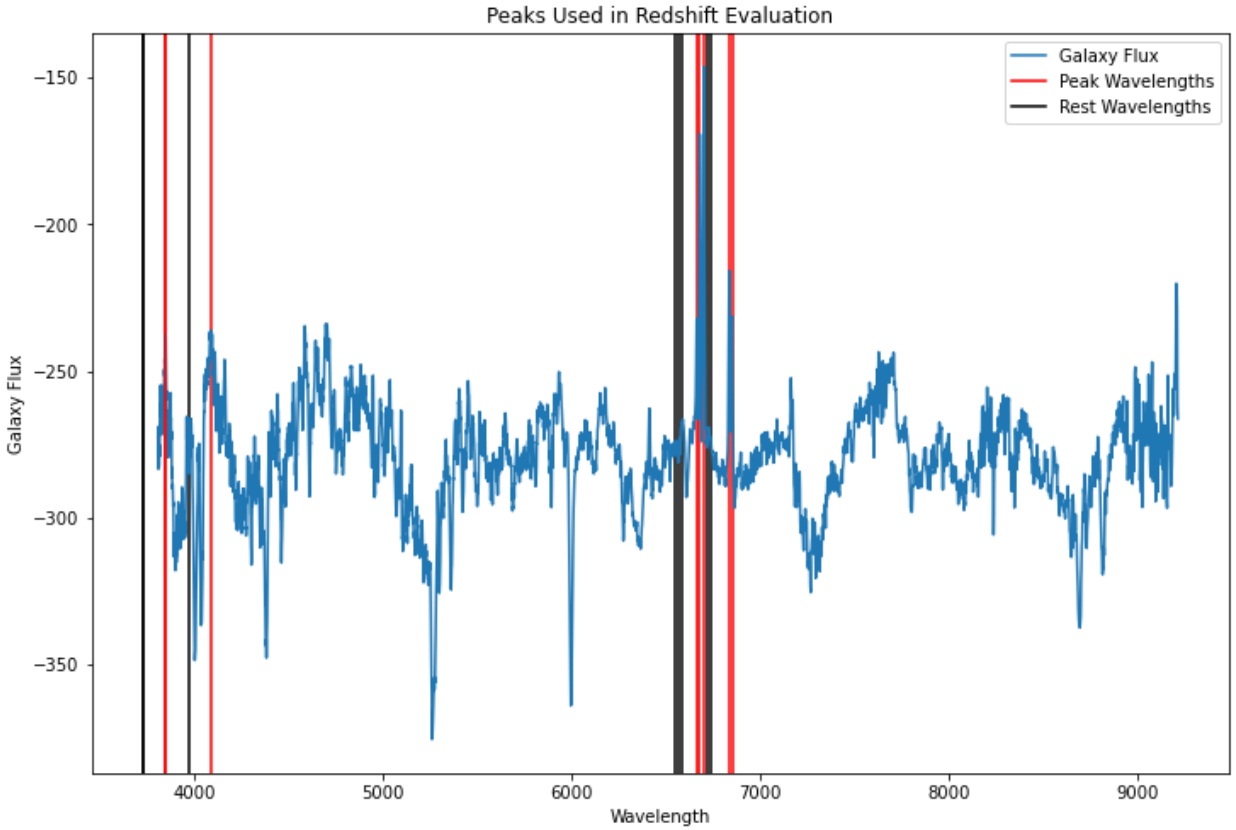

Fig 5. A line plot of the galaxy flux vs wavelength, with vertical lines in black showing the rest wavelength emission lines and the red vertical lines the peaks in the galaxy spectrum. The distance between each of the black and red lines equate to a corresponding Δλ.

To determine the error in my Δλ measurements, a gaussian was fitted to each of the 7 peaks used. Then the mean of the fitted gaussians was used as the estimated peak of the gaussian. The absolute value of the difference between the gaussian mean and the true peak value was then taken as the error in the Δλ value.

```
Δλ for Peak 1: 116.0472 +/- 1.7570
Δλ for Peak 2: 113.0338 +/- 0.9575
Δλ for Peak 3: 110.4940 +/- 1.5304
Δλ for Peak 4: 111.1503 +/- 1.5385
Δλ for Peak 5: 112.0050 +/- 1.5445
Δλ for Peak 6: 114.5199 +/- 0.7885
Δλ for Peak 7: 115.8700 +/- 0.7918
```

Fig 6. The displayed output of the calculated Δλ and its uncertainty for each peak used in the above calculations. We see that the Δλ value varies for each, therefore a global Δλ is not accurate when determining a galaxy's redshift.

Finally, using the data processed and collected above, the redshift (Δλ/λ) could be determined for each individual peak used. The values for λ were taken as the given rest wavelengths and the error for each redshift found in Δλ was then propagated through the equation.

```
Redshift for Peak 1: 0.0311 +/- 0.000471
Redshift for Peak 2: 0.0285 +/- 0.000241
Redshift for Peak 3: 0.0169 +/- 0.000234
Redshift for Peak 4: 0.0169 +/- 0.000234
Redshift for Peak 5: 0.0170 +/- 0.000235
Redshift for Peak 6: 0.0170 +/- 0.000117
Redshift for Peak 7: 0.0172 +/- 0.000118
```

Fig 7. The displayed output of the calculated redshift (Δλ/λ) and its uncertainty for each peak used in the above calculations. Notice that at smaller wavelengths, such as Peak 1 and 2, the Δλ/λ is a significantly larger value.

The average of these redshifts for the galaxy was calculated to be **0.0207 +/- 0.000098**. But the actual variance derived from just the sum of the Δλ/λ array values was found to be 0.000034.

Conclusions of the Data Analysis

It can be drawn from the data analysis above that the value for redshift of this galaxy is reasonably accurate. The 7 peaks that were used to find the Δλ/λ values were ones I deemed to be the most obvious and useful for further analysis in the project, and determination of the average redshift. But in order to increase the certainty in the value of galaxy redshift, I found another peak in the spectrum and used the same data analysis method above to determine its redshift. The new average of the redshifts for the galaxy using the extra peak was determined to be **0.0202 +/- 0.000091**, which is a difference in redshift of ~0.0005. The new variance of the redshifts array then became 0.000031, a small reduction in variance. Still, the addition of more peaks/lines would more than likely lead to a reduction in uncertainty of the galaxy's average redshift.

There was a degree of systematic error introduced into the data, primarily in how I personally analyzed the data. One of the potential error sources I introduced was the manual selection of 'good' peaks in the galaxy spectrum. I may have included peaks that were not actually emission sources or I didn't include peaks that were emission sources.

Another source of error was my choice for measuring uncertainty in Δλ not being accurate when compared with the found variance value of the redshift array. I used the difference in the mean of the fitted gaussian and the real value of the peak being gaussian fitted. These final values for my found uncertainty (seen in figure 7) turned out to be an overestimate of uncertainty when compared with the variance value. Although that gives me more confidence in knowing that the way we analyzed the data may have been more accurate and didn't include very much error.

Also of note to look at is the trend found in the redshift measurements as they correlate with wavelength. Because the values found for Δλ are relatively close together, and did not have a discernible trend in them, shorter wavelengths in the spectrum develop higher redshifts.

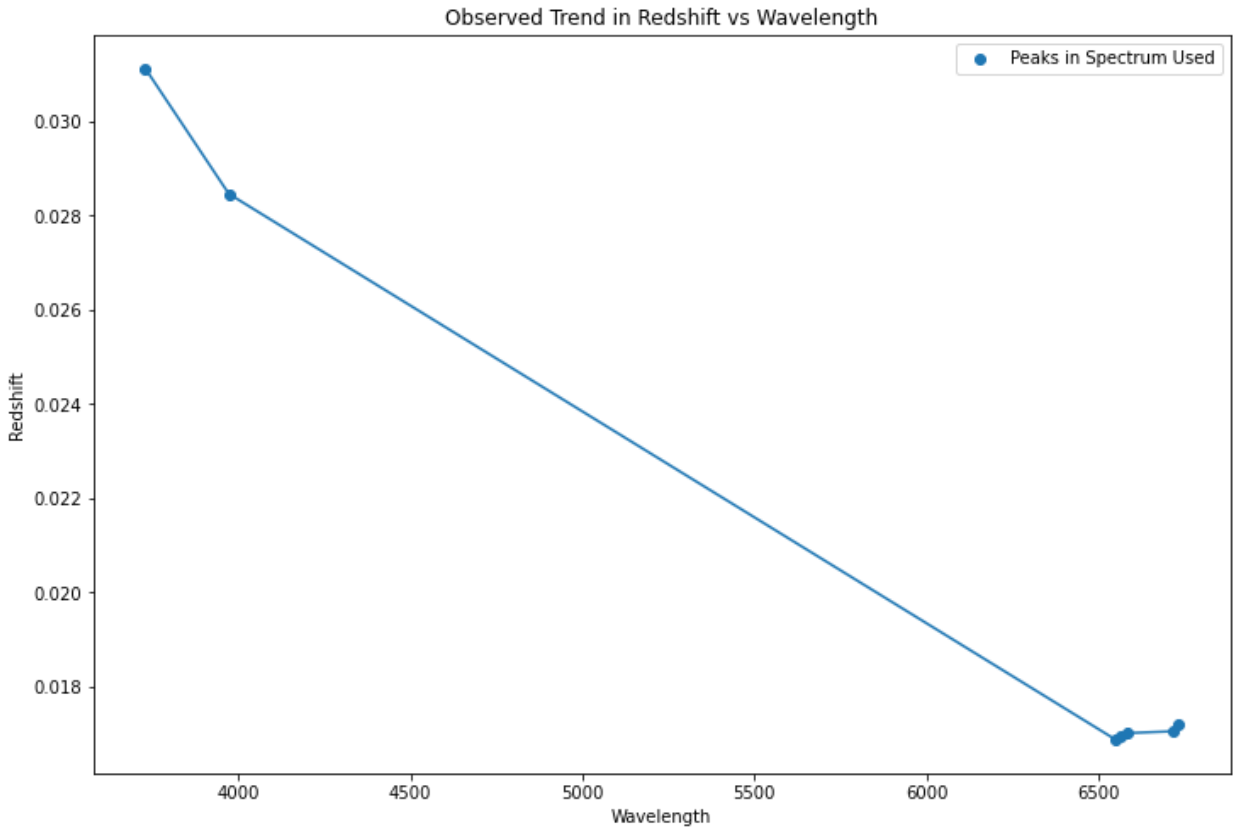

Fig 8. A combination scatter/line plot showing the trend in redshift vs wavelength of the spectrum. As we can see, there is a trend with Δλ/λ going down with higher wavelength.

Suggestions for Reduction of Uncertainty

As stated in the analysis conclusion section using more spectral lines/peaks in the galaxy spectra would provide a reduction in the redshift's uncertainty value. I also did not use any of the large dips in the galaxy spectrum, which could potentially correspond to absorption lines in the galaxy spectrum. Adding the absorption lines into the calculation of the redshift would definitely reduce the uncertainty value, just as if we added more emission lines.

Comment on Project

I personally didn't like the lack of directions given for the project. I was constantly having to ask for help in order to complete the code, which in turn delayed my report on the project. Also could have used more time in class to ask these questions instead of doing a lecture.