AST425 Final Report: Internal Kinematics of Galaxies in Relation to their Morphological Evolution Using MaNGA Survey

Juan Pablo Alfonzo

April 23 2021

Abstract

This study will focus on studying how internal kinematic trends of galaxies in the MaNGA survey relate to their morphological evolution. It will explore the the transition period between star forming galaxies (SFG) to quiescent galaxies (QG). This transition is a relatively fast one in comparison to a galaxy's life span, and hence not much is known about the drastic change in morphology that occurs when a galaxy's star forming rate (SFR) plummets. Galaxies in this transitional phase with rapidly decreasing SFR are known as green valley galaxies (GVG). The goal of this study is to see if there is a trend between kinematic parameters in SFGs, QGs, and GVGs, and whether we can notice any trends amongst these parameters across galaxies of the same type but of different masses. The study will use principal component analysis (PCA) to look for these trends in the data. The hope is having established these trends we can get a better idea of the physical phenomenon inside galaxies that lead to the observed morphological changes.

1 Introduction

The study of galaxy formation and their evolution through time is an area where much research has been done, as it is critical to getting a proper picture of our universe. Yet, it is an area of research where there is still much which we do not understand. The main way we can study how galaxies form and how they change in time is by studying their stellar populations. Not only how the stars happen to be spread throughout the galaxy, but also where we find star forming regions. The most important metric being the star forming rate (SFR), or how many stars the galaxy is making in a given amount of time. In fact one of the key ways we can tell the age of galaxy is by observing its star forming regions that may be scattered throughout the galaxy. By studying these star forming regions physical properties, such as their relative velocity compared to the galaxy or their flux, we can deduce information about the galaxy as a whole. Younger galaxies tend to have bigger and more active star forming regions compared to that of older galaxies, hence they have a higher SFR. These younger galaxies are often called star forming galaxies and late type galaxies (Spiral Galaxies), while the older galaxies are often called quiescent galaxies and early type galaxies (Elliptical Galaxies)[8]. The early/late naming distinction arises from the idea of early type stars (young and hot stars of spectral class O and B), and late type stars (old stars of cooler spectral types). The irony in this is that it is now evident that late type galaxies actually contain mostly early type stars, while early type galaxies actually contain mostly late type stars. To avoid this confusion in naming I will simply refer to late style galaxies as star forming galaxies (SFG), and to early type galaxies as quiescent galaxies (QG) from here on.

A question that is of specific interest to understanding how galaxies evolve in time is understanding the transition period between an SFG to a QG. We have seen that although this transition last a very short time, a few Gyr, it involves profound morphological and structural transformations of those galaxies, as well as dramatic ageing of their average stellar populations[8]. This rapid transition of a galaxy going from an SFG to a QG is known as quenching or rapid halting of the star-formation activity. Additionally, we see that this process happens in a relatively fast burst as current SFGs are observed to steadily decrease their SFR over a period time of time while showing only slightly older stellar populations on average. This means the aging of galaxies seems to be a smooth process until some threshold is reached in which rapid change begins to occur within the galaxy. Galaxies which are in this rapid transition phase between SFG and QG are often called green valley galaxies (GVG). There is a much lower observed amount of GVGs in comparison to SFGs and QGs. This is believed to be a consequence of how fast the transformation to SFG to QG occurs, and hence GVGs do not exist for very long in comparison to other galaxies[8].

The main aim of this research was to examine kinematic properties of GVGs and compare them to those of SFGs and QGs galaxies. This was done using information from the MaNGA (Mapping Nearby Galaxies at APO (Apache Point Observatory)) survey which gives spectrographic data on galaxies. Unlike previous Sloan Digital Sky Survey (SDSS) surveys which obtained spectra only at the centers of target galaxies, MaNGA enables spectral measurements across the face of each of 10,000 nearby galaxies thanks to 17 simultaneous "integral field units" (IFUs), each composed of tightly-packed arrays of optical fibers[1]. This allows for tracking of various important parameters within galaxies such as: two-dimensional maps of stellar velocity and velocity dispersion, mean stellar age and star formation history, stellar metallicity, element abundance ratio, stellar mass surface density, ionized gas velocity, ionized gas metallicity, star formation rate and dust extinction. The main parameters that I wish to compare amongst these galaxies are their stellar velocities and the velocities of gases that give us information about the star forming going on within the galaxy. In order to make these comparisons this study also focused on classifying each galaxy in the survey as a SFG, QG or GVG. The data release 15 (DR15) of MaNGA was used for this study[3].

2 Methods

This section of the paper aims to explain the methods used in this project. This section will offer a brief theoretical explanation of the statistical methods used in this project. It will additionally explain how the data was acquired/packaged before being fed into statistical algorithms. This is due to the fact that the data used in this project is not raw data from the telescope itself, but relies on a data reduction pipeline and an API that packages this data in a neat and useful manner.

2.1 Marvin API

The way the data from the MaNGA survey was accessed was through an interface called Marvin SDSS. Marvin is a complete ecosystem designed for overcoming the challenge of searching, accessing, and visualizing the MaNGA data. It consists of three components: a Web App, a Python package of Tools, and an API (application programming interface). Marvin combines these components to provide a seamless experience when using MaNGA Data[2]. Additionally Marvin does not only allow access to the raw .fits/image files from the telescope but gives access to: objects called "Marvin Maps" (maps for short) of a number of parameters, as well as global galaxy properties like mass or SFR. This information was obtained by running the galaxy spectra through a data reduction pipeline. This processed data was neatly packaged for access via the Marvin project. Figure 1 below shows the raw image taken from the telescope and an example of its corresponding Marvin map of the stellar velocity. The package additionally allows to extract the data for every pixel seen in the right hand side graph as a `numpy` array[4].

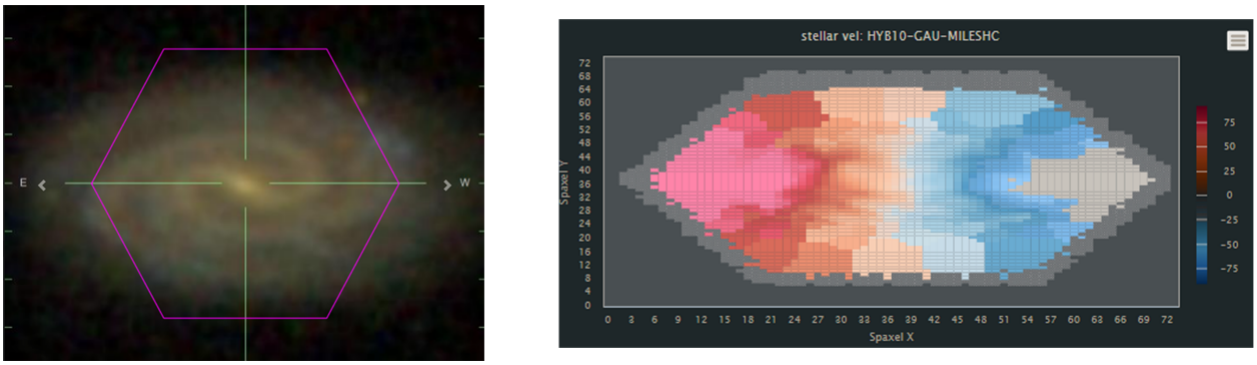

Figure 1: **Left:** Raw image of galaxy from telescope where each pixel in the hexagon corresponds to a microfiber tube collecting the spectra at that point/pixel. **Right:** The corresponding stellar velocity maps which shows the velocity of the stars in each pixel captured by the microfiber tubes

2.2 PCA

Since this project observed a large sample size of galaxies to look for trends PCA (Principle Component Analysis) algorithms were used to help make trends in this vast amount of data visible. PCA is a great statistical tool for

finding trends in large data sets such at this one (roughly 10^8 data points).

Briefly, PCA works by taking multiple variables and plotting them against each other. This in essence turns each variable at play into a dimension. What PCA then does is create a line of best fit using least squares, and ensures this line also passes through the origin of the graph. This line is called PC0 (Principle Component 0). From here it makes n (where n is the number of PC vectors which you are free to choose as long as n is at most equal to the number of dimensions in the data set) lines that are orthogonal to PC0, and also pass through the origin. These lines are known as the PC0-PCn lines. We can then graph this on what is known as a PCA graph, in which the axes of the graph are the PC lines and the projections of the data points indicate where to graph them on the PCA graph. We can consider this as a change of basis transformation in which the PC lines become the new basis vectors of the space. From here we can measure the variance of each line with respect to the data as follows:

$$\sigma_n = \frac{d_1^2 + d_2^2 + ... + d_k^2}{l - 1} \tag{1}$$

Where σ_n is the variance of a given PC line, $(d_1, .., d_k)$ are the distances of the projected points on the PC line to the origin, and l is the number of total data points

What this formula allows us to do is see is which PC is responsible for the most variance in the data. In other words which PC has the biggest impact on the data. This is extremely useful because it allows us to do dimension reduction on our data. This means if we have a data set with 4 dimensions, but we find 95% of the variance lies in two of the PC lines. Then we can simply use these two PC lines as axis and reduce our 4 dimensional data to 2 dimensions, making it much easier to work with and find trends within. Although this is the traditional way to use PCA, this study focuses more on the actual PC vectors the PCA returns. Using the generated vectors for different groups of galaxies we can see how certain trends between variables change across the different galaxy groups.

2.3 Bootstrapping

Bootstrapping in this study was used as a way to gauge the error of the PCA analysis. Bootstrapping entails re-sampling the data and then running the PCA algorithm on the re-sampled data. Re-sampling here means taking the original data set and randomly selecting data points from it to build a new data set of the same size. The new re-sampled data set then can include the same data point multiple times even if the original data set only contains it once as it randomly selects points until it reaches the desired number of data points. For this study the `resample` function from `sci-kit learn`[7] was used to resample the Marvin Maps of galaxies. The amount of times this re-sampling and PCA on the re-sampled data is called the number of reps, and one is free to choose this number. The larger the number of reps the more accurate of an error is expected as with any method that relies on random choice the more trials we have the closer we expect to get to a statistically constant value. In order to gauge an error for the purposes of this study the standard deviation of the returned PC vectors from each of the re-sampled data sets were calculated. This yielded an array which contains the error of each PC vector along each of the variables.

3 Results

3.1 Classification of Galaxies

In the following subsections the details of how the galaxy classification was done will be provided, as well as the final results of doing so.

3.1.1 Log SFR vs Log Mass Classification

The first accomplishment of this study was creating a log(SFR) vs log(M), with Hα flux as the depth parameter, using the MaNGA galaxies. This was done in order to create a way to differentiate/classify the galaxies as SFG or QG, following a method outlined in an astronomical survey [8]. A distinction line was drawn to separate the clusters visually as seen in Fig. 2. Galaxies within a certain distance to the line were defined as GVG, anything above this line was defined as SFGs and anything below was defined as QGs. This is illustrated in Fig. 3. The data to generate this figure was acquired using Marvin which allows for the global properties of galaxies such as mass and SFR to be accessed as `pandas` dataframes[6].

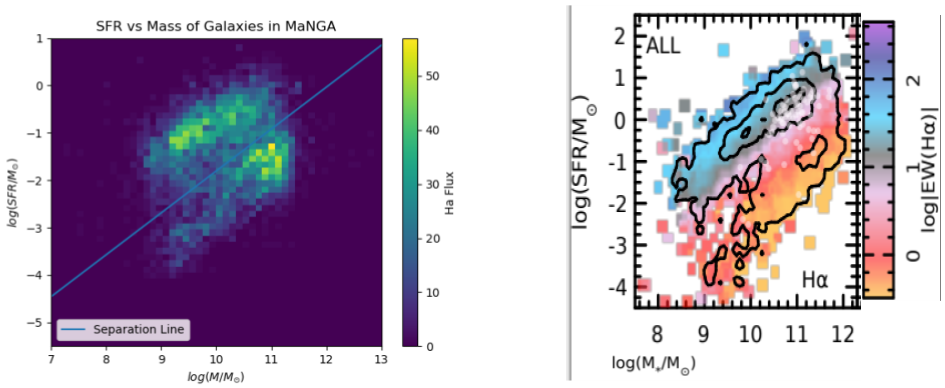

Figure 2: **Left**: My generated plot of SFR vs Mass using galaxy data from the MaNGA Survey. A distinction line is added visually to separate the two visually clear clusters. **Right**: A similar plot to the one on the left taken from an astronomical survey outlining the definition of the GVGs [8]

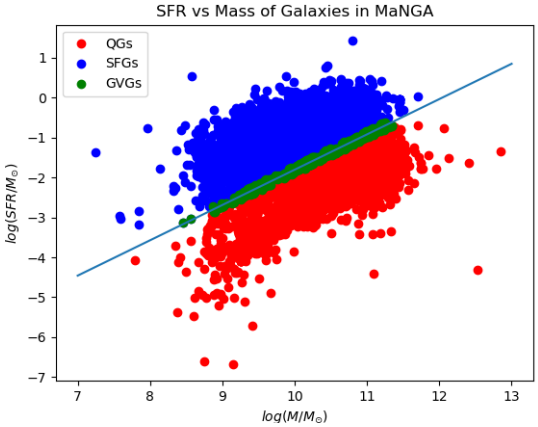

Figure 3: A 2D version of the plot on the left of Fig 1, where galaxies within a certain distance of the separation line have been defined as GVGs.

3.1.2 Clustering Algorithms

Since the above classification method above was rather rudimentary, an attempt was made to find a more concrete way to cluster the galaxies in this plot using Python-based clustering algorithms. Three algorithms were used: DB scan, K-Means, and Spectral Clustering. They all differed in exactly how they went about clustering data, so all three were tried out on the data to see how they performed. In the end DB Scan was the best preforming, but none of them clustered in a satisfactory manner for the purposes of this study.

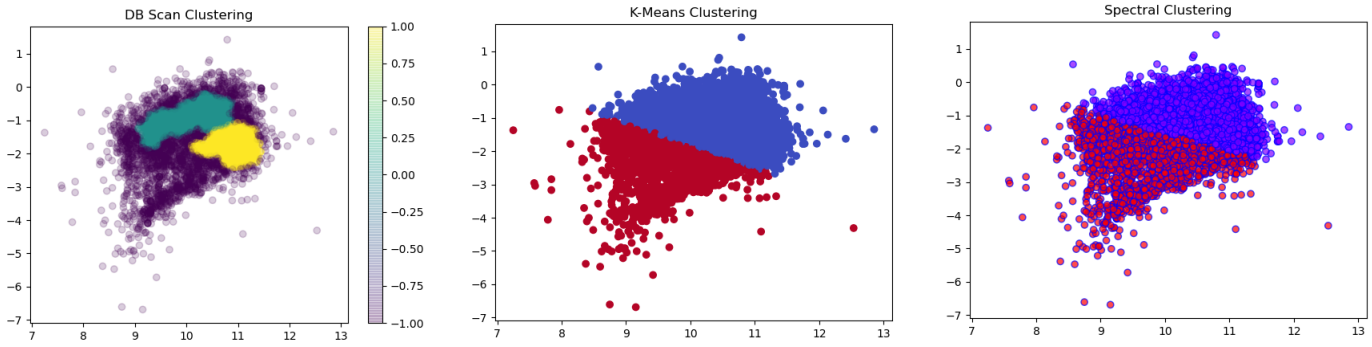

Figure 4: The three methods of clustering used and their results

3.1.3 Gaussian Definition of GVG

In order to get a better definition for GVGs a classification method was adopted from a recent paper by Angthopo et al. (2019) [5]. This process builds upon the results of section 3.1.1 of this paper. The process entails taking the data in Fig. 3, and binning the data into 5 bins based on mass and whether the galaxy is defined as SFG or QG. Then we construct normalized histograms of their $D_n(4000)$ spectra at 1 effective radius (a radius which contains half of the galaxies emitted light). These are then fitted with Gaussians. The point of intersection of the SFG and QG Gaussian in a given bin is taken as the mean of the GVG Gaussian and its standard deviation is taken to be half of the QG Gaussian's standard deviation.

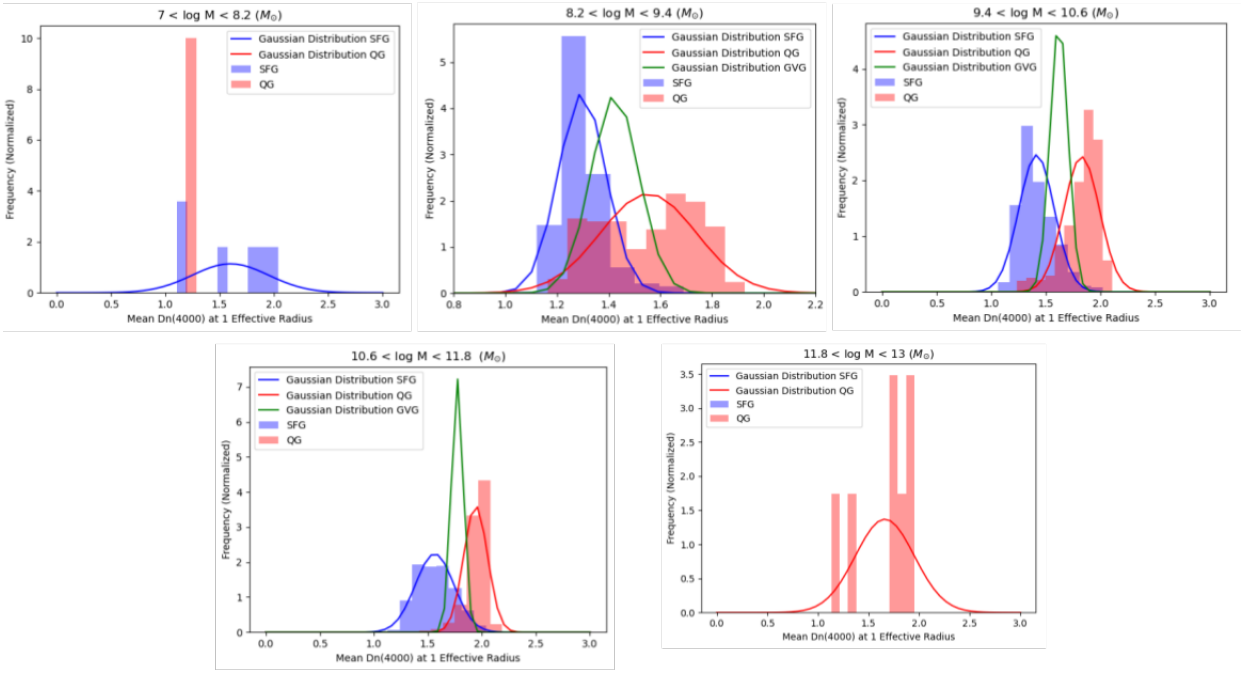

Figure 5: 676 of the 4832 (about 14%) of the galaxies in MaNGA were classified as GVG

3.2 PCA of Kinematic Galaxy Maps

A function was designed in Python with the use of the Marvin API[2] in order to extract the galaxy maps from MaNGA and apply PCA to them. The function is designed to take in a list of galaxies, extract relevant maps, and preform PCA on them. For the purposes of this study the galaxies were divided into the 12 groups seen above (i.e by mass bin and by type). The function pulled all 23 maps of kinematic variables (22 various ionized gas velocities and the stellar velocities) available in Marvin for each galaxy in the group, flattened the arrays, and combined them into one big array that was fed into the PCA algorithm from `sci-kit learn`[7]. Below is a sample of a PCA profile plot that the function generated for one group of galaxies (Fig. 6). Along a given line of the same colour (same PC vector) variables with similar y-values are correlated to each other. Variables with vastly different y-values are anti-correlated to each other. Additionally, the function generates a scree plot (Fig. 7 below) which allows us to gauge how much of the variance in the data each PC vector accounts for.

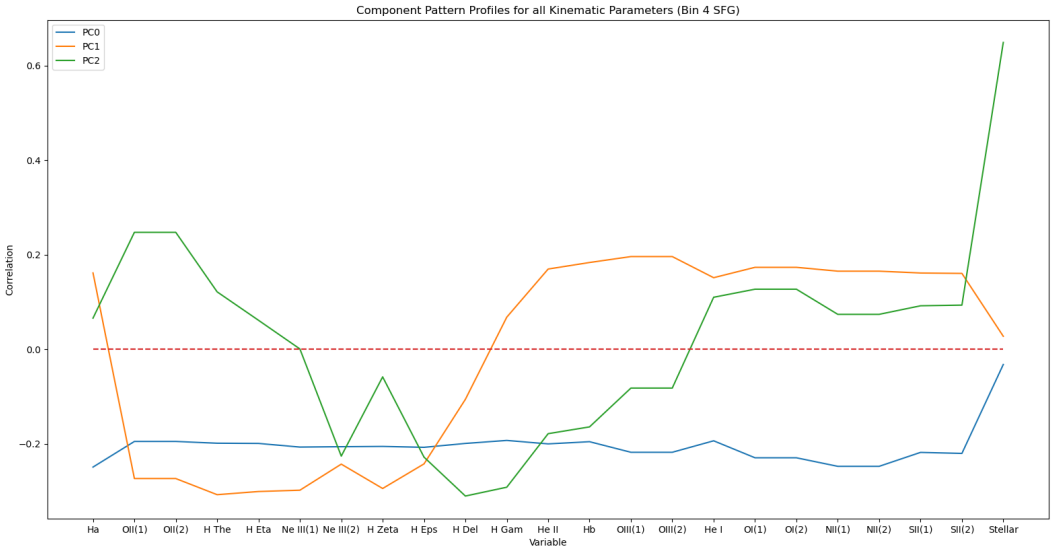

Figure 6: PCA Profile Plot for Bin 4 SFG

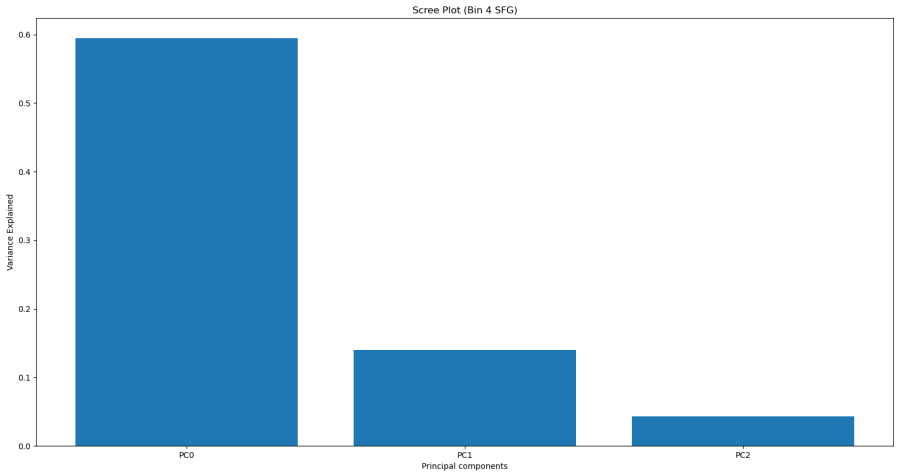

Figure 7: PCA Scree Plot for Bin 4 SFG

In order to compare trends in kinematic variables across all galaxy groups the following array of plots in Fig. 8 were generated. What each cell of the array represents is the difference between two given variables value along each PC vector for each of the galaxy groups. Each cell's x axis tells us the PC vector we are looking at, while the y axis tells us the difference of the variable's value along that PC vector. In order to compare all 23 variables across all galaxy groups the array in Fig. 8 was created. The rows and columns list all the variables, and the cells tell us information about the correlation of two variables which correspond to its row and column index. For example going to the cell with the row marked "Stellar" and the column marked "Hα" tells us about the correlation between stellar velocities and the Hα velocities across all PC vectors. Each galaxy group is also given a unique colour across all cells to represent it, these are outlined in the legend. Note by how this array is constructed it is anti-symmetric and only the top or bottom of half of the array is needed to analyze trends. This due to the fact that we do not care about the sign of the y values of the plots, only the absolute value of them.

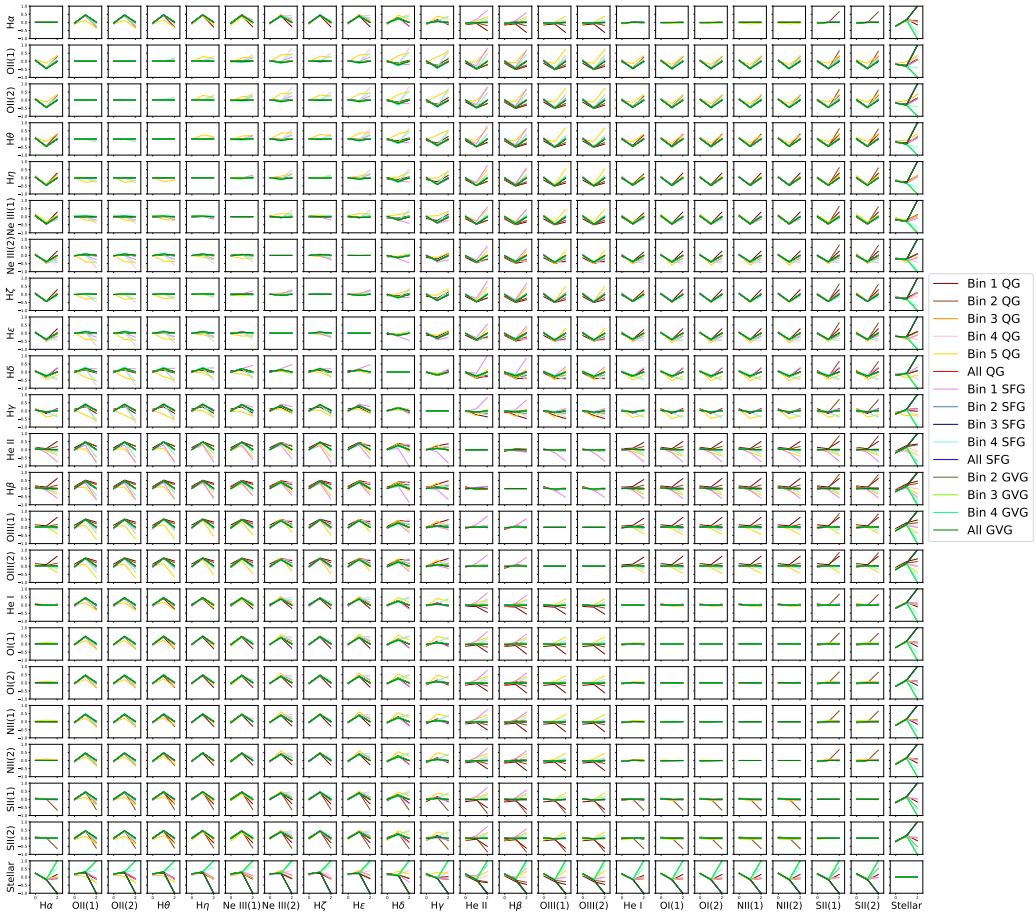

Figure 8: Correlation of kinematic galactic parameters across all galaxy groups. Note this image is of high resolution and can be zoomed into to look at individual cells more closely. To see what each elements/variables corresponding Angstrom break in the spectra is refer to table 2 in the appendix

4 Discussion

4.1 Classification of Galaxies

Firstly I will discuss the results of the log SFR vs log mass classification method. It can be noted from the left side of Fig. 2 that it is clear two distinct groups/clusters of galaxies are seen in this plot. They are clearly pronounced by their high Hα flux. This is the standard way to classify galaxies as either SFG or QG[8]. This is because it allows us to clearly see which galaxies are producing stars by the SFR, and the mass acts a way to normalize the SFR to the size/number of stars in the galaxy. What this system fails to do is give us a proper definition for GVGs. This is due to the fact that we know the GVGs must be those galaxies closest to the lines, as they are in the in-between phase of SFGs and QGs, but this method does not give us a concrete way to figure out what an appropriate distance from the line is. Thus, in an effort to define this middle ground the second classification method of clustering algorithms was used.

As seen in Fig. 4 and previously discussed in section 2.1 of this paper, K-means and Spectral clustering did not do very well to cluster the data. While both methods recognized two major groups to cluster together, they failed to locate the clusters in their proper locations. DB scan was more successful as it properly clustered the SFG group together, and heavier concentration of QGs that can be seen in the left hand side of Fig. 2. DB scan however came short in two major ways. It failed to cluster in the lower mass QGs with the higher mass once, and most importantly it did define a GVG cluster between the SFG and QG clusters. For these reasons the clustering algorithms were simply used a verification check of the clusters created using the log SFR vs log mass plot, and a Gaussian distribution method was adopted in other to classify GVGs.

The method outlined in Angthopo et al. is very similar to what is presented in section 3.1.3 of this report, with a key difference being how the data was binned and that they applied this technique to a different galaxy survey. In their paper Angthopo et al. binned their data by stellar velocity dispersion[5], while in this study the galaxies were binned by mass. This was done since what we were interested in studying in this paper was kinematic trends, including stellar ones, and thus binning it by velocity dispersion would just give us redundant data. Additionally, we wished to compare how binning by mass would compare to binning by velocity dispersion. The use of building histograms based on the $D_n(4000)$ of the galaxy (this is the 4000 $\mathring{A}$ spectra break) comes from the fact that we observe deep drops/spikes here for older galaxies. This is due to the fact that many heavy metals, which are more abundantly found in older galaxies, tend to create this sharp dive in the spectra. This then is a good measure to identify QGs from SFGs, or in our case to verify the that the results of the first classification method are sound. Furthermore, we could fit two Gaussians to these plots as they show bi-modal distributions (as seen in Fig. 5), and using the overlap of the Gaussians we could define the GVGs as outlined in section 3.1.3 of this paper. The results of the classification are as follows. Of the 4832 galaxies in the surveys that had Marvin Maps, 2013 (about 42%) were classified as QGs, 2143 (about 44%) were classified as SFGs, and 676 (about 14%) were classified as GVGs. These numbers are about what we expected to them to be with QGs and SFGs being about equally common while GVGs are a lot more rare for low red-shift surveys[8].

4.2 PCA of Kinematic Galaxy Maps

4.2.1 Correlated Variables Across all Galaxy Groups

The first major step taken after the plot in Fig. 8 was generated was finding variables that were perfectly correlated with one and another. This allowed us to represent multiple variables with just one as we knew whatever trend occurred in one of the correlated variables must apply across all the other variables it is correlated with. In Fig. 8 two perfectly correlated variables would appear as perfectly flat lines, this is to say no difference in any of the 3 PC vectors in any of the galaxy groups. This is both a step towards a further conclusion in the study, and a significant result of the study in itself. This is because it allows us to remove these trends and focus on whats difference across galaxy groups, while also allowing us to gauge which variables have the correlated velocities across all MaNGA galaxies. Variables that were found to be perfectly correlated are shown in the table below, where the left hand variable is the variable chosen to represent the group of variables that its correlated with which are on the right hand side of the table.

Table 1	
Representative Variable	Correlated Variable(s)
Hα	He I, OI(1), OI(2), NII(1), NII(2)
Hϵ	Ne III(2)
OII(1)	OII(2)
SII(1)	SII(2)
OIII(1)	OIII(2)
Hβ	He II

Table 1: Variables which are perfectly correlated across all galaxy groups

Removing these redundant variables from Fig. 8 we get the following reduced plot:

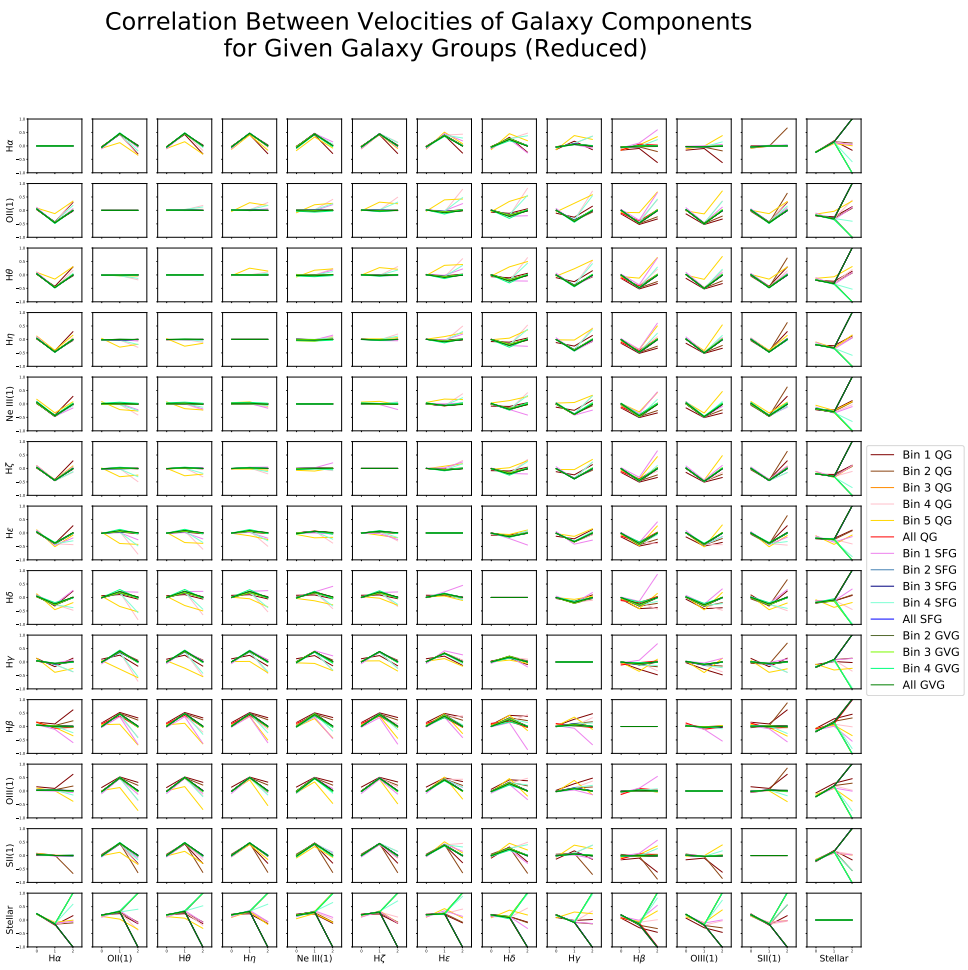

Figure 9: Reduced version of Fig. 8 taking into the account the information in table 1. As with Fig. 8 this is a high resolution image and can be zoomed into to examine details

4.2.2 Discrepancy of Stellar Velocities

Looking at the bottom row of Fig. 9 we see that the stellar velocities are very distinct from the velocities of any of the ionized gasses. We can see this by how disperse the difference across PC vectors for the different groups are. The main hypothesis to explain this trend is the fact that stars move very differently in the galaxy compared to gasses. They are more elastic in nature and less prone to thermodynamic effects from the surrounding environment when compared to free gasses. These thermodynamic effects can cause the gas to contract/expand among other things leading to a very different velocity profile to the relatively more stable stars. Another key thing we observe among the stellar velocity trends is how much Bin 2 and Bin 3 GVGs differ. This then tells us that something in the mass of the GVGs is affecting the way stars are moving within them, which could provide insight into the morphological differences between these two groups. Additionally, we see Bin 4 SFG being the sole non-GVG group to have a very large variance from the other non-GVG groups.

4.2.3 SII(1) and Hα Variation in Bin 2 QG

There is a very unique and interesting trend along PC2 of the Bin 2 QG group with respect to the velocities of Hα and SII(1). This of course extends to all variables that are correlated with any of these two as outlined in table 1. The point of interest for this plot is how uniform and otherwise perfectly flat this plot is with the sole exception of Bin 2 QG as seen on the left hand side of Fig. 10 below, which is a cell from Fig. 9 above. This would suggest

that there is some unique physical phenomenon occurring within Bin 2 QG galaxies that involves these gasses that no other galaxy has.

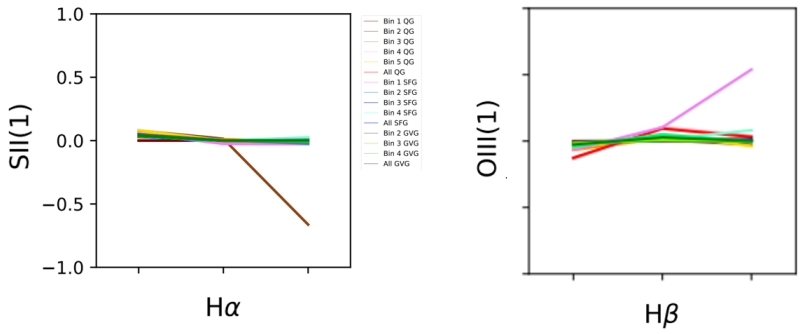

Figure 10: **Left:** Bin 2 QG galaxies acting different from other groups along PC2. **Right:** Bin 1 SFG acting differently than other groups along PC2

4.2.4 Bin 1 SFG PC2 Diverging for OIII(1) and Hβ

This trend is very similar to the described in section 4.2.3. This trend can be seen on the right hand side of Fig. 10 above, which is a cell from Fig. 9. In this case however, we have Bin 1 SFG as the sole outlier from the family of variables (from table 1) represented by the correlation of OIII(1) and Hβ.We are then in a similar fashion as in the previous section lead to conclude that there is some unique physical phenomenon occurring within Bin 1 SFG galaxies that involves these gasses that no other galaxy has.

4.2.5 Bin 5 QG Trends

Bin 5 QGs present various unique trends across a variety of variables. Of special note for these relationships is that even the PC0 and PC1 for the Bin 5 QG group seem to diverge from the others. This leads us to conclude that whatever is causing these different internal kinematic trend must be something really fundamental to the heaviest of QG galaxies in the survey. These trends can be seen in Fig. 11 below which are just specific sections of Fig. 9. This group is relatively small compared to the other groups having only 7 galaxies in it. This could point that perhaps the variation has something to do with a small sample size rather than some actual physical property of the galaxies. This is however not the case and will be further explored in section 4.3 of this paper which will discuss the results of the bootstrapping as a measure of error for the PCA analysis.

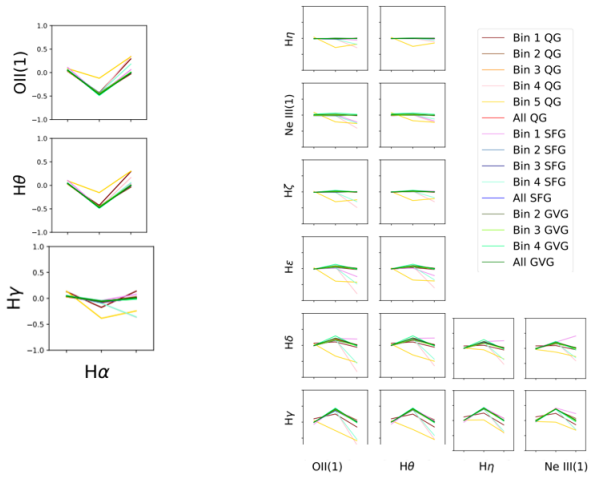

Figure 11: Variables across which the Bin 5 QG group shows behaviour different from other groups

4.2.6 Consistent Anti-Correlation with Hα Across All Groups

For the family of variables associated with Hα, and those with the following set of variables (Hη, Ne III (1), H ζ, Hϵ, Hδ) we see an interesting trend across all galaxy groups. We see that along PC1 for all groups this set of variables have a slight (-0.5 magnitude) anti-correlation with Hα. This leads us to conclude that this specific variation with respect to these variables is expressing some constant physical trend all galaxies seem to follow (at least to a close approximation, the fact that there is slight deviations in the groups points to perhaps even more being at play). Another interesting trend we see in these variables is that Bin 5 QG and Bin 1 QG both deviate more than other groups in PC0 and PC2. These groups are the heaviest and lightest QG groups respectively, and so perhaps this is pointing to some trend associated with the mass of the galaxies. All these trends can be seen in Fig. 12 below, which again is just a selection of cells from Fig. 9.

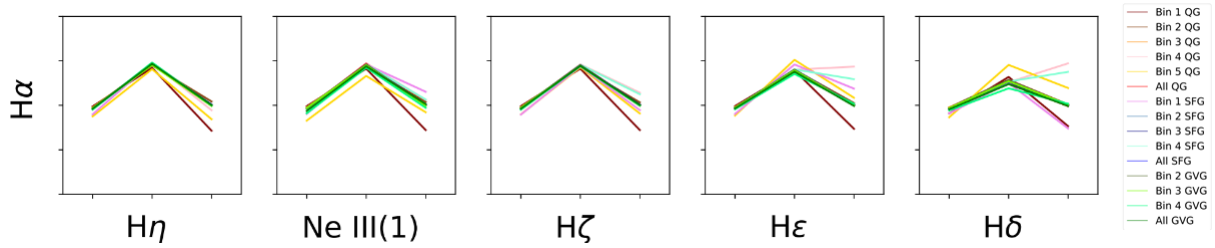

Figure 12: Consistent Anti-Correlation with Hα Across All Groups along PC1

4.2.7 Bin 1,4 SFG and Bin 4,5 QG Diverging From other Groups across PC2

4 specific galaxy groups (Bin 1,4 SFG and Bin 4,5 QG) seem to follow the same trend as every other galaxy for these variables until we get to PC2 for the specific variable combinations seen in Fig. 13 (a subset of Fig. 9) below. Especially for the left hand side of the figure we get massive differences in correlation in PC2 despite being very close across PC0 and PC1. This then points to something of interest happening at very low mass galaxies (Bin 1 SFG) and very high mass galaxies (Bin 4 SFG and Bin 4,5 QG).

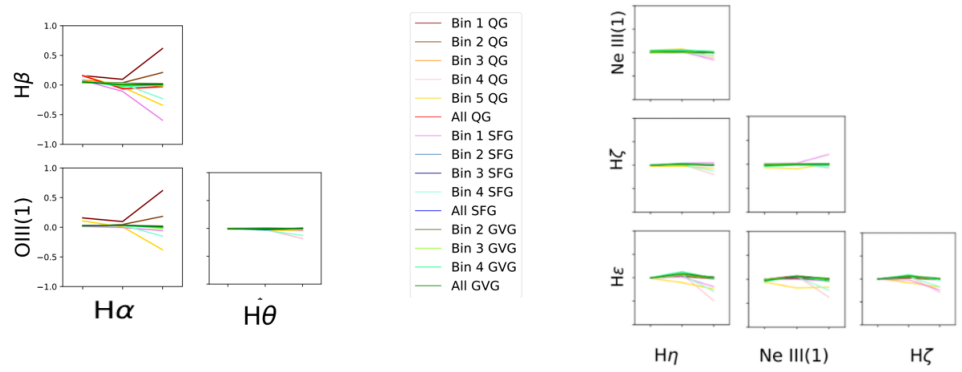

Figure 13: PC2 Divergence for Bin 1,4 SFG and Bin 4,5 QG

4.2.8 Variance in Different QG Groups

For the family of variables associated with Hβ and OIII(1) we see interesting trends when compared to SII(1). We have the usual very low/high mass galaxy groups (Bin 1,4 SFG and Bin 5 QG) being outliers in these plots. Again pointing that something of interest is occurring in galaxies at the edges of the mass spectrum. Additionally, we spot a very odd trend of Bin 1,2 QG completely anti-correlated relationships to those in Bin 5 QG. So, we see in this case that not only do low and high mass galaxies group behave very differently from the other groups but that furthermore they act completely opposite to each other. This again then seems to be a trend heavily tied to the mass of the galaxies as we see divergence even in PC0 for the two ends of the mass spectrum of QGs. These trends can be seen in Fig. 14 below, which as always are sections of Fig. 9.

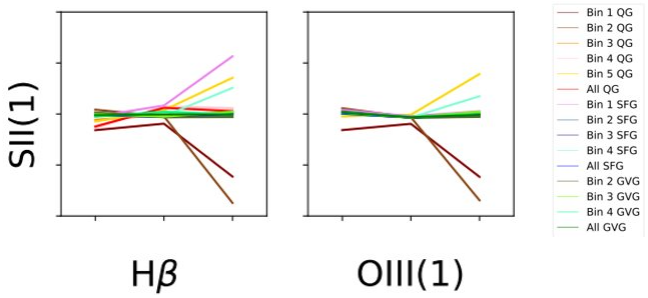

Figure 14: Variance in low/high mass galaxy groups, especially noticeable across low/high mass QG groups

4.3 Gauging PCA Error with Bootstrapping

Bootstrapping was done on every one on the PCA run on each of the galaxy groups as outlined in section 2.3 of this paper. This method allowed us to place error bars on the trends seen above to see if it could perhaps be noise that accounts for any of the trends. It was found that the error for any given variable along any given PC line was never more than 0.3, and the average error of all variables across all PC vectors was found to be 0.18. These then means the errors in the PCA method itself were relatively small, given all the trends discussed discussed had magnitudes of over 0.5. Meaning at most we had a 36% error. Another key point of information given by the bootstrap was that it was observed that the size of the error did not change with the different sized galaxy groups. This then indicates that the odd behaviours of the very low/high mass galaxy groups are pointing to something of physical significance in the galaxies, and are not a result of a smaller sample size.

5 Conclusion and Possible Points of Continuation

In closing, this study had 3 major conclusion. The first conclusion was being able to classify all MaNGA galaxies as either SFG, QG or GVG using the methods outlined in Angthopo et. al [5]. The size of each group is around what we theoretically would expect and so this result also serves to strengthen the new definition of GVG presented in Angthopo et. al. This conclusion can be seen in Fig. 5 and was discussed in section 4.1 of this paper. The second major conclusion of this study was finding which ionized gasses in galaxies had perfectly correlated velocity profiles using PCA. This conclusion can be seen by observing the perfectly flat lines in Fig. 8 and are discussed in detail in section 4.2.1 of this paper. The third and final conclusion of this report was finding unique velocity profiles of certain ionized gases for specific galaxy groups using PCA. These individual conclusions can be seen in sections 4.2.2-4.2.8 of this report. The validity of the last two conclusions were checked by using bootstrapping as way to gauge the error in the PCA that leads to these conclusions. This was discussed in section 4.3 of this paper.

The scope of this study ended at finding correlations of ionized gasses/stars in galaxies, however the conclusions of this study can serve as a basis for multiple new studies. The first conclusion of galaxy classification could be used to study the morphological appearance of the galaxies in each of the 3 groups. These results could then be compared to what we expect these galaxies to look like in theory, or comparing them to other galaxy surveys whose galaxies have already been classified. Similarly, the approach used to classify galaxies here could be tried on other surveys and the results could be compared. Results such as do the galaxies across the different surveys have the same morphology, rotation speed, etc. As for the two results of internal velocities of different variables within the galaxy each one of them could serve to launch a study in its own right. Investigating the physical phenomenon(s) that are the cause(s) of certain galaxy groups having different velocity profiles would be a very worthwhile endeavour. One could imagine perhaps coming up with a theoretical model of the galaxy to describe the trends we observe in their velocity profiles, then making a simulation of these theoretical models and comparing what comes out of that to the observed data. For example using the results of section 4.2.2 one could check if there is a clear morphological difference in the GVGs of the different mass bins, and if this trend in stellar velocities versus gas velocities is somehow linked to it. These sorts of analysis can be done on every trend discussed in section 4.2 of this paper. These sorts of studies into the trends found in this study could yield an abundance of knowledge about the nature of galaxy evolution and how their internal kinematic profiles contribute to their morphological evolution.

References

[1] Mapping Nearby Galaxies at APO (MaNGA)
https://www.sdss.org/surveys/manga/.

[2] Brian Cherinka, Brett H. Andrews, José Sánchez-Gallego, Joel Brownstein, María Argudo-Fernández, Michael Blanton, Kevin Bundy, Amy Jones, Karen Masters, David R. Law, Kate Rowlands, Anne-Marie Weijmans, Kyle Westfall, and Renbin Yan. Marvin: A Tool Kit for Streamlined Access and Visualization of the SDSS-IV MaNGA Data Set. , 158(2):74, Aug 2019.

[3] D. S. Aguado et al. The Fifteenth Data Release of the Sloan Digital Sky Surveys: First Release of MaNGA-derived Quantities, Data Visualization Tools, and Stellar Library. *The Astrophysical Journal*, 240, 2019.

[4] Charles R. Harris, K. Jarrod Millman, St'efan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fern'andez del R'ıo, Mark Wiebe, Pearu Peterson, Pierre G'erard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[5] Ignacio Ferreras James Angthopo and Joseph Silk. Exploring a new definition of the green valley and its implications. *Monthly Notices of the Royal Astronomical Society*, 488, 2019.

[6] The pandas development team. pandas-dev/pandas: Pandas 1.2.4, April 2021.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[8] Sebastian F. Sanchez. Spatially-Resolved Spectroscopic Properties of Low-Redshift Star-Forming Galaxies. *Annual Review of Astronomy and Astrophysics*, 58, 2020.

6 Appendix- Reference Spectra Break for Each Gas

Table 2	
Variable	Angstrom Break ($\mathring{A}$)
Hα	6564
OII(1)	3737
OII(2)	3729
Hθ	3798
Hη	3836
Ne III(1)	3869
Ne III(2)	3968
Hϵ	3971
Hδ	4102
Hγ	4341
He II	4687
OIII(1)	4960
OIII(2)	5008
He I	5877
OI(1)	6302
OI(2)	6365
NII(1)	6549
NII(2)	6585
SII(1)	6718
SII(2)	6732
Stellar	N/A

Table 2: Angstrom break of all gasses whose velocities were studied in this report