Visualization Dashboard for Space Object Classifications

Lauren Montion*
Northeastern University

Tara Joshi†
Northeastern University

Aiden Shymanski‡
Northeastern University

Shruti Kedharnath§
Northeastern University

ABSTRACT

There are so many points throughout space that it can be hard for scientists to work with the huge amounts of data representing so much information. The proposed visualization tool will allow these scientists and other researchers to easily visualize and understand the parts of space they want to explore. The tool allows the user to study celestial objects in a specific coordinate plane of the celestial sphere, display celestial objects of a desired class in a specific coordinate plane of the celestial sphere, study redshift for a group of points based on their location and class, and display the distribution of values for each type of the telescope bands (u, g, r, i, z). Performing all of these tasks will allow the user to gain access to specific information a lot easier than going through a huge dataset and finding it manually. Additionally, it will allow the user to be able to study the characteristics of specific points of interest instead of just reading those numbers from a dataset. These are all important tasks to support because with visual access to this information, anyone studying these points in space will be able to more accurately and confidently make assumptions about the characteristics of certain areas of space and perform more in-depth research after using the tool.

Index Terms: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 INTRODUCTION

It is important to understand how the tool will support each of the domain tasks listed above. First, the tool allows the user to study celestial objects in a specific coordinate plane of the celestial sphere. After using our tool users should have more insight into what characteristics points of a certain class have, or see what is similar between points of different classes that lie close together in space. Being able to understand these objects from different perspectives is the overarching goal of the tool. Specifically, the tool will display celestial objects of a desired class in a specific coordinate plane of the celestial sphere. Knowing this information is equivalent to knowing the latitude and longitude of a country on Earth. Instead of latitude and longitude, space objects use right-ascension (ra) and declination (dec). Knowing and plotting these points on a scatterplot will give the user insight into where in space certain space bodies are. Similar to plotting points on an x and y axis, points will be plotted by ra (x-axis) and dec (y-axis). The tool also allows the user to study redshift for a group of points based on their location. Redshift occurs when an object is radiating light or electromagnetic radiation that increases in wavelength, referring to the red end of the radiation spectrum. The tool provides insight into what the redshift is for a group of points of a class, or a group of points in a similar area of space. Following the general goal of the tool, this allows users to understand redshift in terms of the class of an object as well as its position in space. Lastly, the tool displays the distribution of values for each type of the telescope bands (u, g, r, i, z). Understanding what the band values are for space objects is crucial not only to understand how the photo was taken but what kind of light the object emits and shows visually to humans. Each band takes care of a different group of light, and so with our tool, the user will be able to see what the most common value is for each band and how that is distributed among all our data.

As mentioned above, supporting these tasks is crucial to the efficiency of research. As time goes on, the amount of space data will grow so large that humans will not be able to cover it all on their own. So, a tool programmed to analyze these huge amounts of information and display them so that they are easily understandable is crucial to keep up with our environment. For some perspective, note that we are using data from the Sloan Digital Sky Survey whose last publishing was in 2008. Since then there have been huge advances in database systems and it is projected that the amount of space data NASA will have alone will increase by six times what it is now. The importance of a tool like this that would help scientists study pinpointed aspects of overall space data cannot be stressed enough. As technology for finding data continues to advance, technology for visualizing it must advance at the same rate.

While the information our tool presents is user-friendly and can be used by anyone, it would be most helpful to experts in the astronomy field including: astronomers, astrophysicists, meteorologists, and research scientists. It is likely that people working in these fields would require the visualizations in our tool to help them solve problems in their research or to further understand certain areas of space.

As to the specific data our tool will be presenting, it will show the right-ascension value versus the declination value for a space object. This is very similar the x and y axes that are familiar to many. Right-ascension (ra) versus declination (dec) simply show where in the celestial coordinate plane the object is. We also will be showing redshift averages for each class of objects. Lastly, we will be displaying information about the five telescope bands used by the SDSS. These include the u, g, r, i, and z bands. Each holds a different kind of light that the space object possesses and can provide a user information as to how the picture was taken and what kind of light the body allows to be captured by a telescope.

2 RELATED WORK

It is crucial to understand the factors that help classify galaxies in space so the tool can be useful for scientists looking for galaxies, quasars, or stars with certain characteristics. Greene, Anderson, and Marinelli used SDSS and MaNGA data to classify E+A galaxies using characteristics such as color, spectral shape, absorption lines, redshift, and more. The work presented in this paper gives us insight on what factors scientists look into when trying to classify objects in. We will include these characteristics in our visualization by using checkboxes and brushing so the user can narrow down the points they are looking at, and locate where the points they desire are in space.

Similarly, Clarke, Scaife, Greenhalgh used machine learning to classify objects as general galaxies, quasars, and stars using SDSS data without spectra. The authors trained a random forest machine learning model using photometry to return the classification of different objects as well as the classification probability per class.

*e-mail:montion.l@northeastern.edu
†e-mail:joshi.tar@northeastern.edu
‡e-mail:shymanski.a@northeastern.edu
§e-mail:kedharnath.s@northeastern.edu

The work presented in the paper will inform our visualization as it gives insights into the different factors that go into classifying a galaxy, quasar, or star, such as emission lines, bands, or redshift. We will use these factors to create meaningful visualizations that contain characteristics that will be valuable to the user's research.

3 USE CASE

Astrophysicists and astronomers use data from releases like the SDSS (Sloan Digital Sky Survey) to classify areas of space based on certain characteristics of points like redshift, location, and declination. Examples of these points include stars, galaxies, and quasars. As to who will be using this tool, it would likely be taken advantage of by researchers that are studying certain areas of space. Astronomers and astrophysicists are two huge career fields that require a full understanding of space data. This tool would also be useful for students getting into the field of astronomy. It is simple to understand but can be used for complex problems, making it a great learning tool.

Suppose an astronomer is looking to find clusters of stars with specific numerical attributes, but it is difficult to make sense of all the data from the survey as there are many points in the survey that include more than just star data. The astronomer can use the visualization tool to organize their thoughts and get a better sense of what they're looking to explore.

The tool starts with a checkbox option where the scientist can check which types of data points they want to look at: stars, galaxies, quasars, or any combination of the three. All the points of selected classes are then shown on a scatterplot which is colored by the class of that object. The astronomer wants to know more about these points besides their class. So, they brush over certain points in the scatterplot and a linked bar graph appears to the right. This bar graph contains the average redshift value of the classes of objects that were brushed. Additionally, 5 histograms will show the distribution of band value for each of the 5 telescope bands. This will give information as to what type of light the band was absorbing when the picture of the celestial body was taken.

This tool will make the astronomer's research easier as they now know where to look for certain space objects based on the area in space it is contained in and certain physical attributes it has. These physical attributes include the redshift of those objects and the type of light the star shows based on telescope pass-band system used by the SDSS as denoted by the 5 band distributions. Suppose this astronomer is interested in E+A (post-starbust galaxies), which have low brightness and redshift. They can brush over points in the scatterplot and now they can see where these galaxies might are and what their redshift values are, which will be helpful when continuing their research.

4 DATA

We found our dataset on Kaggle where it was posted by Lennart Grosser, a machine learning engineer from Berlin, Germany. The dataset contains 10,000 rows of observations, each having 17 numerical columns and one class column and is a subset of the SDSS dataset from their 14th release. The SDSS provides us with the information needed to learn and discover more about our universe. This exact data table was created by querying the CasJobs database and joining the photometric and spectral data tables published by the SDSS. The dataset we will be working with can be found at `https://www.kaggle.com/datasets/lucidlenn/sloan-digital-sky-survey?resource=download`.

Since the collected data simply contains observations about the universe, there are not any apparent ethical issues that arise. When it comes to biases in the data, the subset of the SDSS dataset provided on Kaggle may not have been fully representative of the true distribution of observations across the universe that are shown in the original SDSS dataset. Since we do not know if the subset of data was chosen at random, or handpicked, we do not now how widely distributed the values are for the true scale, but it seemed to provide us with sufficiently distributed data.

To begin our data cleaning process, we first created a subset of the dataset that has 450 rows to work with. The small amount of data we used is only because using more rows would have caused the browser to crash. We then inspected the data types of the 18 attributes to confirm they were all correct, and they were. Since we did not need to utilize all 18 columns in the dataframe, we dropped the unnecessary columns. The columns we ended up keeping included: objid, ra, dec, u, g, r, i, z, specobjid, class, redshift, and color. After, we counted the amount of null values in each column, and found that there were none. Next, we took a look at the class column to make sure there were only three values used (star, galaxy, and quasar), to which we found was implemented correctly. The last step of our data cleaning process was to remove all rows that contained one or more outlier to have more concise data. After completing these steps, our data was ready to be used.

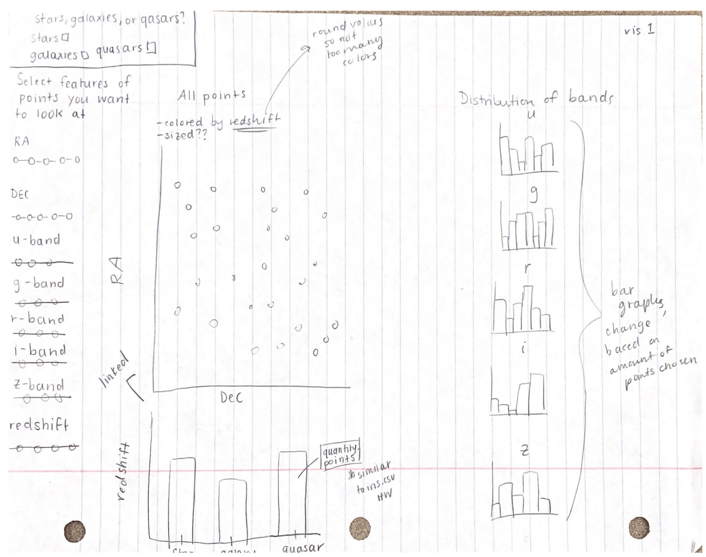
5 DESIGN PROCESS

Multiple stages of sketching were necessary in order to reach a suitable visualization tool to fulfill the needs of a user. Three rough sketches were created and a fourth, final sketch which more properly outlines what our visualization tool looks like to users.

5.1 Rough Sketches

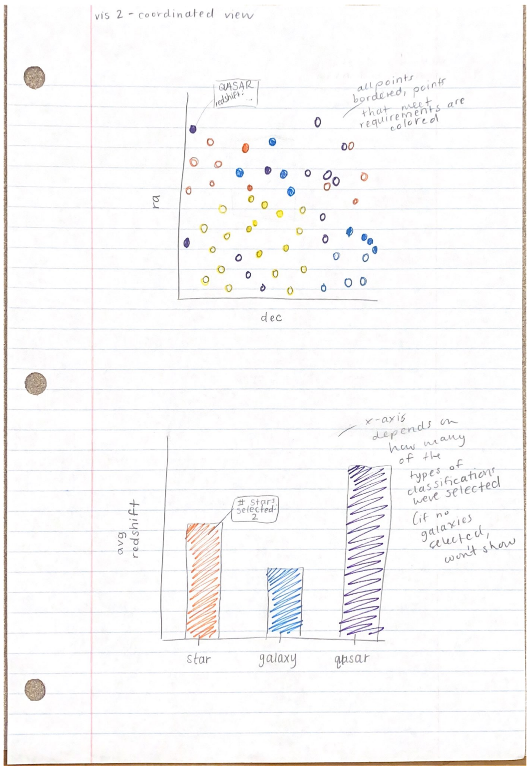
We made three rough sketches and one final sketch before implementing our tool using programming. Below we provide descriptions on each of these sketches and how they influenced the final sketch.

5.1.1 Sketch 1



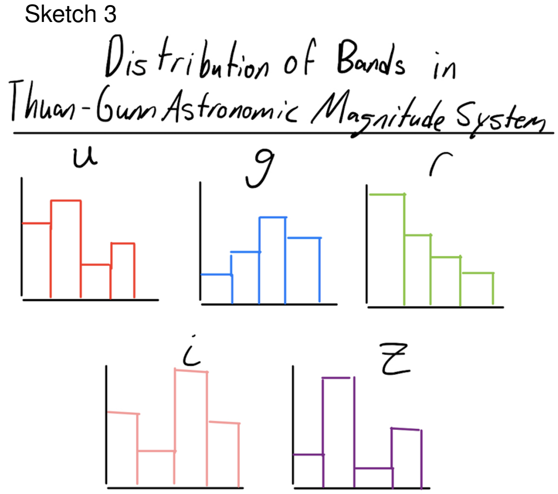
Sketch 1 is an overview of everything we wanted the tool to do. It focuses on what components will be included and how. This includes figuring out that the user can select up to three types of objects to look at, can use sliders to select designated values, and what the visualizations will actually look like and how they will be interactive. Also, the coordinated view was determined in this sketch. This helped influenced the final sketch because it provided a foundation for where we wanted to place each component and what visualizations we wanted to show to the user.

5.1.2 Sketch 2



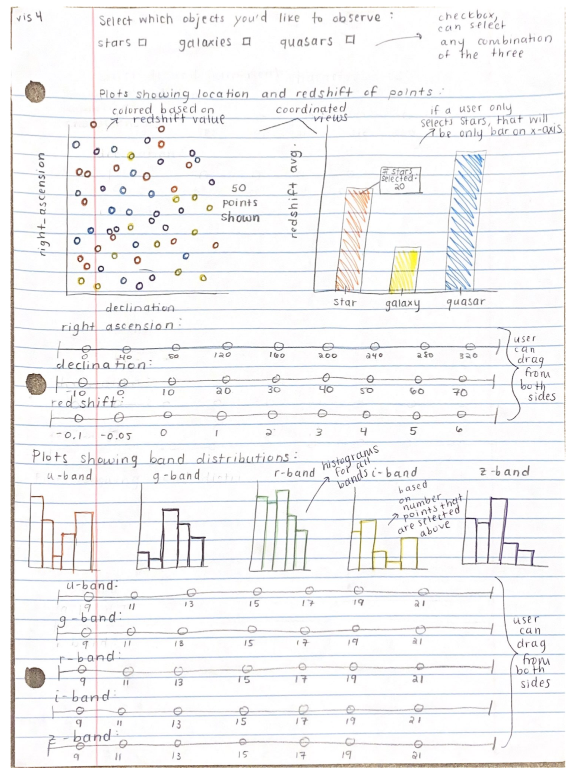
Sketch 2 focuses on the linking we planned to implement. The top shows a scatterplot which is colored by the class. We planned to implement a tooltip so a user could hover over a point to see its class and other numerical characteristics it contained. It is also set up so that the points are bordered, but once they are brushed they fill in completely. The bottom graph is a bar graph which is dependent upon the scatterplot. When the user brushes over points in the scatterplot, those points appear in the bar graph which shows the average redshift of the points, again where the bars are colored by class. We had also planned to use a tooltip for the bar graph to show how many points of that class were being used.

5.1.3 Sketch 3



Sketch 3, the last rough sketch, shows the 5 histograms that would show the distribution of the band lengths for u, g, r, i, and z bands. They are colored with a categorical colormap, and are labeled based on the band they represent.

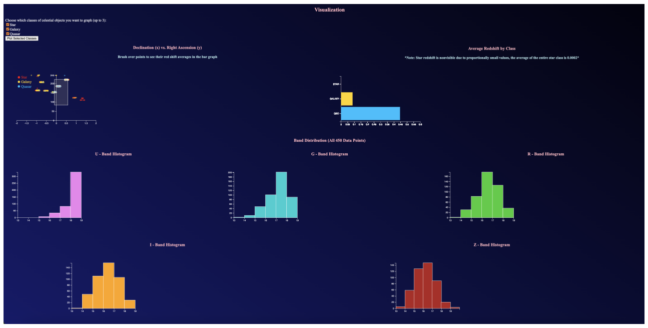
5.1.4 Final Sketch



Using the three rough sketches, we made a final sketch. We used to first sketch but just switched the orientation, so it goes from top to bottom rather than left to right. It still starts with the user choosing desired classes. We also kept the organization shown in the second sketch, which has the linked scatter and bar graphs implemented right under the checkboxes. The only change from the sketch is that the points start out fully colored in and receive a border when brushed over. Because we wanted the user to be able to change physical attributes, we wrote in sliders so that the graphs above would change accordingly based on those values. The third rough sketch we also implemented, as can be seen by the five categorically colored histograms shown below. These also contain sliders for each band that would allow the user to filter values they want to look at. We also put a tooltip for the histograms so the user could see the band value they were looking at and how many points fell in each bin.

Our final design changed a bit from the final sketch. The most substantial change is that no sliders were implemented. The reason for this is mainly the time constraint for the project. However, we most likely would have still decided to remove the sliders for the scatter and bar graphs. This is because we wanted the user to be able to focus solely on the location and redshift of the points and not get confused as to what attributes they were applying. With more time we still would have implemented the 5 band sliders for the histograms so the user would have been able to filter data in another way. Further, we did not implement a tooltip for the scatter or bar graphs. When all 450 data points show on the scatterplot, a lot of them are so close together a tooltip would not have helped the user in seeing what exact point they were looking at. For the bar graph, when all points are brushed, the star average redshift value is so low relative to galaxies and quasars that it won't show on the graph. A tooltip again would not have helped the user so we did not implement it. Lastly, after usability testing we learned we needed to provide further instruction on our webpage so the user could see the full effects of the tool. We did this by telling the user they could brush over points, letting them know the star bar wouldn't show because it's average redshift value was so low, and for the histograms making sure the user knows the distributions are out of 450 points. Besides these changes, the overall layout of the tool remained the same, and we decided to keep the implementation of the tooltip for the histograms.

5.2 Final Design



The final design shown above puts all the components of our partial drawings together. The top of the tool allows the user to filter the data based on what types of objects they want to look at using checkboxes and a button, and from there the tool begins visualization. Once the desired space object classes are chosen, a scatterplot appears underneath, plotting the location of those points. When the user brushes over a certain area, a bar graph forms to the right and gives the average redshift value of the selected classes. The bar graph will change according to what points the user brushes over. Underneath are 5 histograms, one for each telescope band. They show the distribution of all 450 data points within these band values.

Suppose the astronomer from our use case wants to research stars and quasars, and examine their different distributions in space, as well as a specific attribute like redshift. The astronomer can click ”Star” and ”Quasar” on the checkboxes by the scatterplot, and click ”Plot selected classes”. They can now examine the distribution of stars and quasars in space from the Sloan Digital Sky Survey to observe what regions each class is more likely to show up in and where clusters appear. The astronomer is interested in a specific cluster of stars and quasars with a certain right-ascension and declination, so they brush over it, creating a redshift average barplot to the right of the scatterplot. They can observe the average redshifts by class (separated by star and quasar in this case), and even brush over only one point at a time to get individual redshift values. The astronomer can now observe how the redshift changes based on region of space from the scatterplot, and how the redshift of each individual class of celestial object changes too. They can also compare the redshift averages of both objects against each other. For more information about telescope band distribution, the astronomer looks at the histograms at the bottom of the visualization to better understand the data they are observing.

6 Discussion

Overall, the visualization tool solves the overarching domain problem. With it, scientists are able to filter through a subset of SDSS data and choose which classes of objects they want to look at without having to filter through the original data themselves. They can also gain additional understanding about the physical characteristics of these points including their redshift averages and the distribution of their telescope band values. Some limitations that took away from the tool's potentiality include the fact we were only able to use 450 data points so as not to crash the browser, so the graphs aren't as populated as they could be. Because of this, the star redshift values are not even visible on the bar graph. Resolving that issue would have allowed a tooltip to be implemented for the bar graph so that users can visualize the numeric value of each bar. Additionally, the histograms stand alone and if they were able to also filter the data it might provide more insight for the user. Future improvements include creating sliders for the 5 telescope bands so the histograms could be made based on what values the user wants to see in the bands. Additionally for the histograms we would want to add how many of each class of object are in the histograms so the user could leverage that information along with the information given by the graphs above. Lastly, a zoom feature on the scatterplot would have helped the user differentiate more between points, and then a tooltip would have shown the class of the object and other physical data that we did not visualize. With these changes the user could get more insight out of the data that could be used for research.

7 Conclusion

The goal of this project was to create an interactive dashboard for scientists and astrophysicists to use to sort through space data. We focused in on the Sloan Digital Sky Survey's (SDSS) data because they are widely known throughout the community of scientists and astronomers and their data is crucial for important research. The need for this tool is so user's could search through this huge database in a much more efficient way and gain more insight from the data than they would filtering through a dataset themselves. The importance of our tool is not only that it gives numerical information about space objects, but it visualizes them so that the user can gain a full understanding of what they are looking at. The tool will leverage information about the space object's physical characteristics and visualize them in multiple ways to give a different lens to look at the data through. Because it is interactive it is very easy for the user to change what they want to look at and understand more about all space objects which include stars, galaxies, and quasars in the SDSS. We all contributed to this project and made sure it met standards we set from the beginning. Aiden made the tooltip for the histograms, corrected the axis values for the bar chart, and added checkboxes that the user intially uses. Lauren initializes the scatterplot, created the histograms and worked with Tara on the creation of the first draft bar graph. Tara finalized the scatterplot with correct axes and created the bar graph so that it would change based on the points the user brushes over in the scatterplot. Shruti worked on extracting the data from the button and passing that into each of the the visualizations, making sure the scatter plot doesn't layer with each click of the button, and made sure brushing worked for the scatterplot. We all contributed to the aesthetics of the tool and worked together to debug individual components of the visualization.

8 Appendix

8.1 Data Abstraction

In this dataset, each row represents an observation in space taken by the SDSS. Each observation is detailed by a class column, classifying each observation as a star, galaxy, or quasar, and ten other columns describing the astronomical coordinates, astronomic magnitude, redshift, and more.

- Attribute Types:
  - Right Ascension (ra): Ordered (Quantitative)
  - Declination (dec): Ordered (Quantitative)
  - Thuan-Gunn Astronomic Magnitude System (u, g, r, i, z): Ordered (Quantitative)
  - Class: Categorical
  - Redshift: Ordered (Quantitative)

8.2 Domain Task Abstractions

For each domain task the visualization tool will perform for the user, a task abstraction is given. This involves a breakdown of the high, medium, and low levels of each task.

- Domain Task 1: Study celestial objects in a specific coordinate plane of the celestial sphere.
  - High: Analyze, Consume, Present
  - Middle: Search, Locate
  - Low: Query, Identify
  - Targets: Spatial data, shape
- Domain Task 2: Display celestial objects of desired class in a specific coordinate plane of the celestial sphere.
  - High: Analyze, Consume, Present
  - Middle: Search, Lookup
  - Low: Query, Identify
  - Targets: Attributes, many, similarity
- Domain Task 3: Study redshift for a group of points based on a specified red-shift range.
  - High: Analyze, Consume, Present
  - Middle: Search, Lookup
  - Low: Query, Identify
  - Targets: Attributes, one, distribution
- Domain Task 4: Locate celestial objects that are within a specific range for each type of telescope band (u, g, r, i, z).
  - High: Analyze, Consume, Present
  - Middle: Search, Lookup
  - Low: Query, Compare
  - Targets: Attributes, many, similarity

References

[1] Clarke, A. O., Scaife, A. M., Greenhalgh, R., Griguta, V. (2020). Identifying galaxies, quasars, and stars with Machine Learning: A new catalogue of classifications for 111 million SDSS sources without spectra. Astronomy Astrophysics, 639. https://doi.org/10.1051/0004-6361/201936770

[2] Greene, O. A., Anderson, M. R., Marinelli, M., Holley-Bockelmann, K., Campbell, L. E., Liu, C. T. (2021). Refining the E + a galaxy: A spatially resolved spectrophotometric sample of nearby post-starburst systems in SDSS-IV Manga (MPL-5). The Astrophysical Journal, 910(2), 162. https://doi.org/10.3847/1538-4357/abe4d1

[3] LennartGrosser. (2018, September 20). Sloan Digital Sky Survey DR14. Kaggle. Retrieved March 14, 2023, from https://www.kaggle.com/datasets/lucidlenn/sloan-digital-sky-survey

[4] Sloan Digital Sky Survey Dr14. SDSS SkyServer DR14. (n.d.). Retrieved March 14, 2023, from https://skyserver.sdss.org/dr14/en/home.aspx