A Simple Method of Producing Images of SDSS Spectra in a Free Spreadsheet Program

JULIA FALCOME

1 Department of Physics and Astronomy, Georgia State University, Atlanta, GA 30303, USA

1. ABSTRACT

Using Google Sheets, I develop a method to easily reproduce thousands of images of SDSS spectra so that they may be studied in only a fraction of the time it would otherwise take. This method may be helpful in projects requiring large samples of SDSS objects with spectra, and is described in a step-by-step manner so that it is accessible to everyone.

2. MOTIVATION

As of the fourth generation of the Sloan Digital Sky Survey (SDSS-IV), Data Release 15 (DR15) (Blanton et al. 2017; Aguado et al. 2018), over 2 million galaxies have been observed with single-fiber spectra which visualize flux levels over a wide range of wavelengths. These spectra are invaluable in projects involving the observation of properties like broad-line emission or characteristics of specific galaxy morphologies. However, when projects require one to look through thousands of spectra to obtain a sufficient sample size, the process of clicking thousands of different links to arrive at the images of desired spectra can be extremely time-consuming. There have certainly been attempts to automate morphological classifications (e.g. Sreejith et al. 2017; de Diego et al. 2020), but others have shown that human observation of spectra may yield more accurate results, especially when classifications rely on noticing subtle spectral characteristics that are difficult for automated systems to spot (Greene et al. 2020). In this research note, I detail a method to easily and quickly reproduce images of SDSS spectra in Google Sheets, which allows for thousands of spectra to be observed by the human eye in only a fraction of the time it would normally take, by centralizing all of the necessary data in one place instead of over a multitude of webpages. This method was developed when my research group was tasked with looking through more than 15,000 spectra to identify E+A galaxies (Liu et al. 2020, in preparation; Dressler & Gunn 1983) in numerous large fields. This technique reduced the time to complete the assignment by at least a factor of 10.

3. METHODOLOGY

I have written a Github post explaining the steps in this section steps in greater detail for anyone who might find it useful. As mentioned, the application I find best suited for this method is Google Sheets, which allows for the extraction of images from websites with the correct use of its library of functions. The only necessary information for each galaxy is its spectral ID (hereafter called the SpecID), RA, and declination, all of which can be obtained when sending a data query through the SDSS SkyServer Search Form and checking the applicable boxes. The RA and decl will only be necessary in the event of troubleshooting (Section 4).

Use Figure 1 as a reference where the first, second, and third columns contain each galaxy’s RA, dec, and SpecID respectively. Then, if the fourth column is to display an image of the galaxy’s spectrum, that column’s cell for a given row should contain the following text:

=IMAGE(CONCATENATE("http://skyserver.sdss.org/dr15/en/get/SpecById.ashx?id=",[SpecID]),1)

in which [SpecID] should point to the cell containing that galaxy’s SpecID. This command uses CONCATENATE to combine the SpecID with an otherwise static URL to produce a link of the spectrum’s image. The IMAGE function grabs the image of the spectrum from that URL and inserts it into the cell as shown in Figure 1. The second parameter in the IMAGE function (in this example, 1) relates to the different options in displaying the image which is detailed in the function’s description.

This produces the image of the spectrum in the desired cell, and can then be applied to all other rows.
4. TROUBLESHOOTING

Since this method makes use of the functions available in Google Sheets, it is not applicable to other similar applications such as Excel. Even in Google Sheets, this method produces null results approximately 15 percent of the time, as the SpecID does not always correspond with the galaxy with which it is listed. In such cases, the correct SpecID can be obtained using one of the following solutions:

**Problem:** After inputting the command above, I receive an error message.
**Solution:** First, make sure that the [SpecID] text is deleted and replaced with the correct cell containing the SpecID. Additionally, depending on how the command is copied and pasted, there may be inconsistencies in how the quotation marks are represented. Make sure that all quotation marks are of the Unicode number U+022(“) and not the left and right double quotation marks U+201C/U+201D (“ ”).

**Problem:** After inputting the command above, the cell reads No image exists for this region.
**Solution:** When this happens, it means the SpecID which was received from the SkyServer query is incorrect. To remedy this problem, one can either obtain it manually by searching through SDSS using the RA and dec, or one can copy the following command into the cell containing that galaxy’s SpecID:

```excel
=MIDB(INDEX(IMPORTHTML(CONCATENATE("http://skyserver.sdss.org/dr15/en/tools/explore/Summary.aspx?ra=",[ra],&"dec=",[dec]), "table",28),1),14,19)
```

Here, [ra] and [dec] are the cells containing the galaxy’s RA and dec, respectively. This command makes use of several other functions, which have the following purposes: CONCATENATE once again creates a URL from inserted values of the galaxy’s RA and dec; IMPORTHTML imports the data from this URL into the sheet in the form of a table (which I have specified with the parameter "table"). The SpecID is typically on row 28 of this table, so I choose to isolate that row; INDEX further helps to isolate the SpecID; and lastly MIDB is used to pinpoint the exact location of the desired characters. The SpecID is the 14th character in the row and runs 19 digits long, hence the final two numbers. Although this command stretches over two lines in this note, it should read as one continuous line in the program, so it will be necessary to delete any line breaks once the command is pasted into the program.

**Problem:** I typed in the command from the solution above to obtain the SpecID, but the cell for the spectrum is now completely blank.
**Solution:** Change 28 to 29 in the command above, so it reads as such:

```excel
=MIDB(INDEX(IMPORTHTML(CONCATENATE("http://skyserver.sdss.org/dr15/en/tools/explore/Summary.aspx?ra=",[ra],&"dec=",[dec]), "table",29),1),14,19)
```

The problem is caused by flags on some galaxies warning of possibly unreliable photometry. The inclusion of this text causes the placement of all other measurements below it to shift by a line.

5. ACKNOWLEDGEMENTS

I thank Charles Liu for his valuable discussions and input. This work was supported by the Alfred P. Sloan Foundation via the SDSS-IV Faculty and Student Team (FaST) initiative, ARC Agreement SSP483, and by NSF grant AST-1460860 to the CUNY College of Staten Island.

REFERENCES

Aguado, D. S., Ahumada, R., Almeida, A., et al. 2018, The Fifteenth Data Release of the Sloan Digital Sky Surveys: First Release of MaNGA Derived Quantities, Data Visualization Tools and Stellar Library. https://arxiv.org/abs/1812.02759

Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28, doi: 10.3847/1538-3881/aa7567

Bundy, K., Bershady, M. A., Law, D. R., et al. 2014, The Astrophysical Journal, 798, 7, doi: 10.1088/0004-637x/798/1/7

de Diego, J. A., Nadolny, J., Bongiovanni, Á., et al. 2020, A&A, 638, A134, doi: 10.1051/0004-6361/202037697

Dressler, A., & Gunn, J. E. 1983, ApJ, 270, 7, doi: 10.1086/161093
Greene, O. A., Anderson, M. R., Marinelli, M., Holley-Bockelmann, K., & Liu, C. 2020, ApJ, submitted

Sreejith, S., Pereverzyev Jr, S., Kelvin, L. S., et al. 2017, Monthly Notices of the Royal Astronomical Society, 474, 5232, doi: 10.1093/mnras/stx2976
**Figure 1.** Above: An example table on Google Sheets containing the commands necessary to produce the desired images. Whereas the first two rows have the correct SpecID, the third row does not and commands are entered into the cell as described in Section 3. Bottom: The resulting table after the commands are entered, showing images of spectra inside the cells corresponding to each SpecID.