

Topic 3 Homework

An Introduction to Science Using the Sloan Digital Sky Survey

1: Introduction.

The goal of this assignment is to help you get acquainted with the SDSS Sky Server, as well as learn the basics of the database language SQL which has broad applications, beyond its use in SkyServer.

The homework has 3 parts. In the first you use a simple web-interface to extract galaxy data, and reproduce some of the basic color-concentration-absolute magnitude relations we met in class. In the second you go through the SDSS SQL Tutorial to learn how to write more complex SQL queries. Finally, you write your own SQL to extract data on elliptical and S0 galaxies and explore the various 2-D and 3-D parameter relations: the Faber Jackson and Fundamental Plane relations.

The primary home-page for SDSS is: <http://cas.sdss.org/dr7/en/>, and the home-page for Skyserver and how to use it is: <http://cas.sdss.org/dr7/en/tools/>

2: Simple Web-based Extraction of Galaxy Image Data.

Go to the SDSS Imaging Query form located at

<http://cas.sdss.org/astro/en/tools/search/IQS.asp>

With relatively little effort, one can extract a wide range of information from the SDSS databases by using this simple web interface. After the data are downloaded, one can then examine it more carefully.

Moving down the form: initially request just 50 objects (output rows) and choose “html” output. When you are ready for your main run, reset this to 10,000 objects and .csv format (the file will be downloaded to your computer as “comma separated variable” format which you can easily read into your plotting/analysis programs).

In the section on Parameters to Return: use the “BEST” imaging, and scroll through the imaging parameters, using Ctrl-mouse (Apple-mouse on a mac) to select several parameters to be returned. The “Imaging” title is a web-link to a brief description of all the parameters (scroll through them all, and see if you recognize any mentioned in class). For our purposes, choose Petrosian magnitudes (petroMag_) and extinction (extinction_) for the g,r, and i bands. For the r-band, also select the r magnitude error (petroMagErr_r), and two parameters that allow you to create a concentration index (petroR50_r and petroR90_r, the radii containing 50% and 90% of the Petrosian flux). Over on the spectroscopy side: choose redshift (z) and its confidence (zconf).

For Position Constraints: choose a rectangle somewhere in the middle of the SDSS survey (e.g. RA 180 to 185 and DEC 30 to 35).

For Imaging Constraints: select the Petro magnitude option, and specify a range in r magnitude:

choosing 12 to 19 is sufficient to yield reliable data. Under object type, select “Extended Sources”. For Quality, select a minimum of 2 (good).

Don’t worry about the flags at the bottom – they allow you to include or exclude objects based on certain flags set during processing.

More general information about the Imaging Query Form (IQF) and Spectroscopic Query Form (SQF) is given here: http://cas.sdss.org/astro/en/help/docs/QS_UserGuide.asp

Submit the Request, which should return quite quickly. The html arrives on your screen, while the .csv arrives in your “download” folder. When you are satisfied by both the html and .csv format returns, then increase the number of requested objects from 50 to 10,000.

At this point, you need to use your favorite programming/graphing environment to read the .csv file and create three x - y plots:

- (1) A simple color-color plot of (g-r) vs (r-i);
- (2) A galaxy “type” vs color (g-r) plot, using concentration = petroR90/petroR50 to roughly follow the galaxy type: spirals or ellipticals;
- (3) Absolute magnitude, M_r , vs concentration.

These mirror some of the plots we discussed in class.

You need to apply quality control to the data, so only include data for which the r magnitude error is less than, say, 0.05. For the third plot (only), there obviously needs to be a redshift, and its confidence, zconf, should be greater than, say, 0.95. Remember to first apply the extinction corrections to each magnitude, and don’t worry about a cosmologically accurate calculation of the absolute magnitude – a simple Hubble Law with $H_0 = 72$ km/s/Mpc will do. Despite your quality controls, some spurious data will still enter your plots as wildly discrepant points – so make sure you choose your axes ranges intelligently, to focus only on the astrophysically relevant data.

With so many objects, plotting single points isn’t so useful. So divide the plot into, say, 50×50 bins. Ascertain how many objects are in each bin, and graph the array as a gray-scale with black for the most densely populated parts (or chose your own way to show the distribution of data points – using contours, for example).

Submit your final graphs and copies of the (commented) programs used to analyze and plot the data, and briefly explain in astrophysical terms the distribution of data within each graph.

3: Using SQL to Access the SDSS Database

You may have noticed from the html return from the previous exercise that the submission form you constructed was itself translated into an SQL script, and it was this script that was submitted to the SDSS server. It is worth learning how to use SQL because it is a widely used database language with many applications beyond its use in SDSS.

This next part of the homework is fairly straightforward, and involves you going through the SQL

Tutorial on the SDSS website:

<http://cas.sdss.org/astro/en/help/howto/search/>

It is probably worth going through the whole thing and trying the examples. By the end, you'll be able to do relatively simple SQL searches, and be familiar with some of the primary SDSS tables and parameters.

Rather than specify explicitly what next to look at, here are some other helpful links that you should probably check out:

http://cas.sdss.org/astro/en/help/docs/sql_help.asp Second level Tutorial.

<http://cas.sdss.org/astro/en/help/cooking/> Frequently used tasks

<http://cas.sdss.org/astro/en/help/docs/realquery.asp> Examples, starting with simple ones, but later ones can be very sophisticated.

<http://cas.sdss.org/astro/en/help/browser/browser.asp> Schema Browser: descriptions of all datasets and their parameters.

<http://cas.sdss.org/astro/en/tools/search/sql.asp> SQL Query Form for SDSS: this is the place to construct your SQL and then submit it to Skyserver.

4: Practice Example: the Faber Jackson and Fundamental Plane Relations for Ellipticals.

Use the SDSS database to explore two fundamental relations found for Elliptical and S0 galaxies: the Faber-Jackson relation, and its parent, the Fundamental Plane (we'll meet both of these in Topic 7 on elliptical galaxies). The Faber-Jackson relation is between absolute magnitude and stellar velocity dispersion, and the Fundamental plane involves effective (half-light) radius, surface brightness and stellar velocity dispersion.

Use the main SQL submission page:

<http://cas.sdss.org/astro/en/tools/search/sql.asp>

Select a sample of elliptical galaxies which have measured absorption line velocity dispersions (veldisp), and measured photometry. You'll need two datasets: "galaxy" (a "view" subset of photoObj) for the photometric data, and specObj (a "view" subset of specObjAll) for the spectroscopic data. You'll need to output the following parameters for later off-line analysis: objID, petroMag_r, extinction_r, petroR50, z (redshift) and velDisp.

Using SQL, select top 10 and return to html with the following criteria:

- a) Select only elliptical and S0 galaxies, by using: concentration (petroR90/R50) > 2.6 and extinction corrected red colors: g-r > 1.0 (these criteria should be evident from part 2 above).
- b) Galaxies with r magnitudes (petroMag_r) in the range 0 – 19 and r magnitude errors (petroMagErr_r) less than 0.05 mag

- c) Galaxies with measured redshift $z < 0.35$ that are reliable, i.e. $z_{\text{conf}} > 0.95$
- d) Galaxies with measured stellar velocity dispersion ($v_{\text{disp}} > 30$ km/s) and error ($v_{\text{dispErr}}/v_{\text{disp}}$) less than 20%.
- e) **Don't** constrain the region of sky searched, since this slows things down greatly.

You will need to combine the photometric and spectroscopic datasets, so that the data from each is matched to the other. This is possible using something like: $s.\text{bestobjID} = p.\text{objID}$ as one of the conditions after WHERE (where s is alias for “specObj” and p is alias for “galaxy”). Alternatively, you can use a “JOIN” structure, as explained in the second level Tutorial (see above link).

When you have a working SQL, check a few objects at random by copy/pasting the object ID into the navigate tool to confirm they look like elliptical or S0 galaxies, and have red absorption line dominated spectra (or use the “thumbs” tool at: <http://cas.sdss.org/astro/en/tools/chart/list.asp> to look at several). Then increase the number of lines returned to 10,000 and direct the output to .csv file.

Write programs to do the following with the returned .csv file:

1) Recreate the 2-D Faber Jackson x - y plot, between absolute magnitude, M_r (y -axis) and Log velocity dispersion, $\text{Log } \sigma$ (x -axis). Choose sensible ranges for the axes, and use the gray-scale density plot technique you used in part 2 above. Use a simple Hubble law to calculate M_r from the petrosian apparent magnitude (petroMag_r), the local extinction (extinction_r), and redshift (z). Roughly, what is the gradient of the relation, i.e. what is n in the relation $L \propto \sigma^n$? (Remember, 1 dex in L is 2.5 mags).

2) Consider the 3-D x,y,z volume given by $\text{Log } R_e$ (in kpc; use the approximation $R_e \approx \text{petroR50}_r$), $\langle \mu_e \rangle$ (mean surface brightness within R_e , in magnitudes per square arcsecond), and $\text{Log } \sigma$ (in km/s). Notice, we're using subscript “e” to mean “effective radius” which is the half-light radius. So evaluate $\langle \mu_e \rangle$ using the petrosian magnitude and the area enclosed within petroR50_r . Create all three 2-D “projections” of the data, again using the gray-scale graphs you've been using before:

(a) The three simple projections onto the primary axes are:

(i) The x - y projection of $\text{Log } R_e$ vs $\langle \mu_e \rangle$ (This is called “The Kormendy Relation”)

(ii) The x - z projection of $\langle \mu_e \rangle$ vs $\text{Log } \sigma$

(iii) The y - z projection of $\text{Log } R_e$ vs $\text{Log } \sigma$

Compare the scatter on each of these relations. Which is the “strongest” correlation, i.e. which has the least scatter? Try to visualize how the data fill the 3-D space, using the projections to help.

(b) The data actually fall on a tilted plane (called “The Fundamental Plane”) within this x,y,z volume. To look at the plane “edge on”, plot $\text{Log } R_e$ (x -axis) against $0.36 \times \langle \mu_e \rangle + 1.4 \times \text{Log } \sigma$ (y -axis). The scatter on this relation should be even less than the scatter on the best relation from part (a). Is this true for your plots? It should be!

For this homework, please print out and hand in:

A) Your SQL file(s)

B) All plots, with axes labeled and titles given.

C) The programs you wrote to analyze the .csv output files and make the plots.

D) Answers to any of the (simple) questions asked above.