Homework 4 - Basic Data Mining

Don't forget to `svn up` *before you issue any other commands in SVN—this is to guard against you changing a document that someone else is working on in the same directory*[1].

Don't forget to `svn ci` *(with* `-m` *comments) frequently as you work. This allows others to see how your work progressed, and it automatically backs your work up as you produce it so that you're less likely to lose any of your work and/or so that you can revert to earlier versions of your work.*

Remember to comment your code carefully with your initials before every comment (as in `; ADM I just wrote a Python comment`*). Remember to provide an informative header for* **every** *function that you write. Also provide a README file to inform people how to run your code.*

An astronomer wants to study the GALEX (UV) properties of SDSS optical point sources that are detected in the radio by the FIRST survey (*remember from the lecture notes that files containing GALEX, SDSS and FIRST information are all stored locally*). As an initial test, she chooses to limit her study to a specific region of the sky—two overlapping circular regions of $\theta = 2^\circ$ in radius, centered at $(\alpha, \delta) = (163^\circ, 50^\circ)$ and $(\alpha, \delta) = (167^\circ, 50^\circ)$.

1. Determine which FIRST sources lie in the astronomer's survey and write them out to a FITS file.

2. Use my `sdss_sweep_data_index.py` code to determine the SDSS `PSFFLUXES` and GALEX `NUV` and `FUV` fluxes for *primary*[2] point sources in the SDSS that are within $1''$ of the FIRST sources in the astronomer's survey. Note that *primary* SDSS sources in the sweep files can be specified as follows (if the rec array containing your SDSS sweep information is called `objs`):

   ```
   >>> primaryflag = 2**8
   >>> w = np.where( (objs["RESOLVE_STATUS"] & primaryflag) != 0)
   >>> objs = objs[w]
   ```

   Retain the SDSS `RA` and `DEC` for these sources to help answer question 4, below. Note that only $\sim$5–10% of the FIRST sources will match an SDSS primary point source.

3. Determine which of the FIRST sources in the astronomer's survey is brightest (has the largest flux) in GALEX `FUV`. Let us refer to this source as $FUV1$. Plot the 7 fluxes (5 SDSS and 2 GALEX) of $FUV1$ as a function of wavelength[3].

4. Use the *DR8 SDSS Navigate Tool* linked from the syllabus, and the SDSS `RA` and `DEC` of $FUV1$, to retrieve the image of $FUV1$ from the SDSS database. By clicking `Explore` you should be able to see the spectrum and identifying information of $FUV1$. As part of your code for question 3 of this homework, print out a few brief comments indicating why, given how it was selected in the astronomer's survey, the spectrum and identification of $FUV1$ is as might be expected.

---

[1] this shouldn't be a big deal unless we're working collaboratively, but you should get into the habit *now*

[2] "primary" in this context means the best observation in the SDSS, given that the SDSS scanned some parts of the sky multiple times

[3] `PSFLUX` represents SDSS $ugriz$ bands at 3543Å, 4770Å, 6231Å, 7625Å and 9134Å, respectively. GALEX `FUV` and `NUV` are at about 1600Å and 2400Å, respectively