Lab #0: Introduction to ADQL and Gaia Data
Due 4pm, February 4, 2019

Astronomy Data Lab AY128/256 (UC Berkeley, 2019)

January 21, 2019

1 Introduction & Logistics for Your Lab Submission

First of all: welcome to the **Astronomy Data Lab Class**! This first lab (indexed to zero of course!) serves as a gentle introduction to *Gaia* data and ADQL, a scripting language used to query databases. It is shorter than the three primary labs you'll be working on in this course, but it should provide a rough idea of the sort of work we expect you to do in Astro 128/256. We hope it will be useful to you in preparing for Lab #1. For this lab, the expectations for undergrads and grad students are the same.

You will turn in an IPython/Jupyter notebook that includes both discussion and all the code and ADQL queries needed to reproduce your results. You should explain (in words) what your code and queries are doing and how you made any non-obvious judgement calls. Your code should be legible and adequately commented (e.g., give your variables descriptive names; avoid long very long lines, etc.) You can use the PEP 8 style guide as a general guideline to help you make consistent looking Python code. Avoid hard-coded variables, "magic numbers", etc.

You can write your discussion and answers to any questions in Markdown, using the built-in LaTeX support for equations. If you prefer, you are also welcome to write up your final lab report as a separate LaTeX document (but either way, you need to turn in a notebook with your code). **You are welcome to consult with other students in the class, but all code, write-ups, and analysis must be your own.**

You have several options for where to write and test your lab notebooks. If you have a working Python 3 installation[1] and know how to create notebooks locally, that should work fine. Alternatively, with a CalNet ID you can use Berkeley's Datahub (http://datahub.berkeley.edu) or Google Colaboratory (http://colab.research.google.com) to host and run notebooks that can be accessed from anywhere with an internet connection (even without a local Python installation).

Most of the packages needed for the labs are already available on Datahub and can be imported immediately. If you want to install new packages, this is also straightforward (you can execute bash commands from within the notebook environment). If you would like access to a computer in the undergraduate data lab, please let us know and we will have an account created for you.

2 Preamble

a) Make yourself an account on the Gaia Archive by going to https://gea.esac.esa.int/archive and clicking "SIGN IN" and "Register new user" in the upper right-hand corner. You don't need an account to query the Gaia catalogs, but having an account will allow you to save your previous queries and to upload and query your own catalogs, which will be useful later on.

[1] We will insist that you use Python 3 for this class. Python 2 support in Jupyter, numpy, scipy, etc. will be depreciated officially by the end of this year!

The Gaia Archive hosts most of the catalogs we will need for this lab and Lab #1. To see which catalogs are available, click "search", and click on the "Advanced (ADQL)" tab on the upper left. Available catalog are listed on the left side of the page in nested drop-down menus.

The Gaia Archive is a good place to test out your ADQL queries before integrating them into your notebook. If your queries have errors, the error messages in the Gaia Archive will help you identify them.

b) Install ASTROQUERY in your Python environment[2]. The ASTROQUERY.UTILS.TAP.CORE.TAPPLUS utility will allow you to combine ADQL queries with Python code.

c) Familiarize yourself with the Gaia mission. Here are some papers and web resources for doing so, ordered roughly by their usefulness to you for this assignment and Lab #1. You are not required or expected to read these in full, but some familiarity with the data products will be extremely very helpful.

1. Babusiaux et al. 2018 (about construction of color-magnitude diagrams with Gaia; arxiv: 1804.09378)

2. Brown et al. 2018 (general summary of the 2nd Gaia data release; arxiv: 1804.09365).

3. Summary of the Gaia DR2 catalogs and the meanings of all columns: https://www.cosmos.esa.int/documents/29201/1645651/GDR2_DataModel_draft.pdf

4. Introduction to ADQL, the database language used to query the catalogs (this is quite similar to SQL): https://gea.esac.esa.int/archive-help/index.html

5. Lindegrin et al. 2018 (summary of Gaia DR2 astrometry; arxiv: 1804.09366).

3 Now the Fun: The Assignment

d) After much ado, here's the meat of the assignment. Construct color – absolute magnitude diagrams for the following star clusters: The Hyades (a young, nearby open cluster), M67 (an old, more distant open cluster), and NGC 6397 (a globular cluster, at an even larger distance). If you've looked at Babusiaux et al. 2018, you should have an idea of what this means and how to accomplish it. Some hints are below. The data you need can be found in the GAIADR2.GAIA_SOURCE CATALOG.

You will need to identify stars that are actually members of each cluster, and to distinguish them from other stars that are in the same part of the sky but not physically associated with each cluster. Gaia astrometry (parallaxes and proper motions) will be useful for this.

Section 3 of Babusaix et al. provides some guidelines for distinguishing cluster members from background stars. You don't need to follow their procedure exactly, and you can make use of any known properties of these clusters in the literature, but your selection should involve some sort of cuts in position, parallax, and proper motion.

One option is to use a cluster-finding finding algorithm to identify likely cluster members (e.g. based on the number of nearby neighbors they have in phase space). If you know how to implement such a search, it is likely the most principled option. For this assignment, it's also fine to do something more kludgy; e.g. circular or rectangular selections in position and proper motion.

[2]For those using Datahub, astroquery, should have already been installed.

You will want to remove objects with unreliable photometry and astrometry. Read through sections 2 and 3 of Babusiaux et al. 2018, and use quality cuts similar to those they recommend to filter out bad sources. You will probably want to use less stringent cuts for objects that are relatively far away (and thus have lower signal-to-noise ratios) than for those that are nearby. Discuss what sorts of problems lead to bad astrometry and how the cuts you implement remove suspect objects.

Along with the final color-magnitude diagrams for cluster members, show (at least for one cluster) some diagnostic plots that illustrate the selection of cluster members in phase space. For example, try plotting the proper motion in RA and Dec of potential cluster members based on a selection in position and parallax only. Is there a clear clump of stars in proper motion space that corresponds to the cluster? Try plotting the color-magnitude diagram for potential cluster members with and without cuts on proper motion. Do the proper motion cuts make the color-magnitude diagram cleaner? Also, show how your astrometric quality cuts clean up the final color-magnitude diagram (i.e., compare the CMDs you obtain with and without quality cuts).

e) Once you have clean color-magnitude diagrams for the three clusters, overplot some synthetic photometry in Gaia bands for theoretical isochrones. You can use MIST models (http://waps.cfa.harvard.edu/MIST/interp_isos.html), and you can use estimates of the ages and metallicities of the three clusters from Babusiaux et al. 2018, or elsewhere in the literature. Identify the various phases of stellar evolution in each cluster. Are there are stars that are likely to be in binary (or higher order) systems? Finally, comment on any discrepancies between the theoretical models and the data.

4 Next Steps

This Lab is due at 4pm on Feb 4th, 2019. Details of the submission process will be given later. The first official meeting/lecture will be on Jan 28 from 4–7 pm in 131 Campbell. A big part of this time will be for interactive discussion of your progress on Lab #0. Be prepared to discuss your progress on Lab #0. Josh and Dan will be hosting a meet and greet on Friday Jan 25 from 1–2:30pm in 355 Campbell.