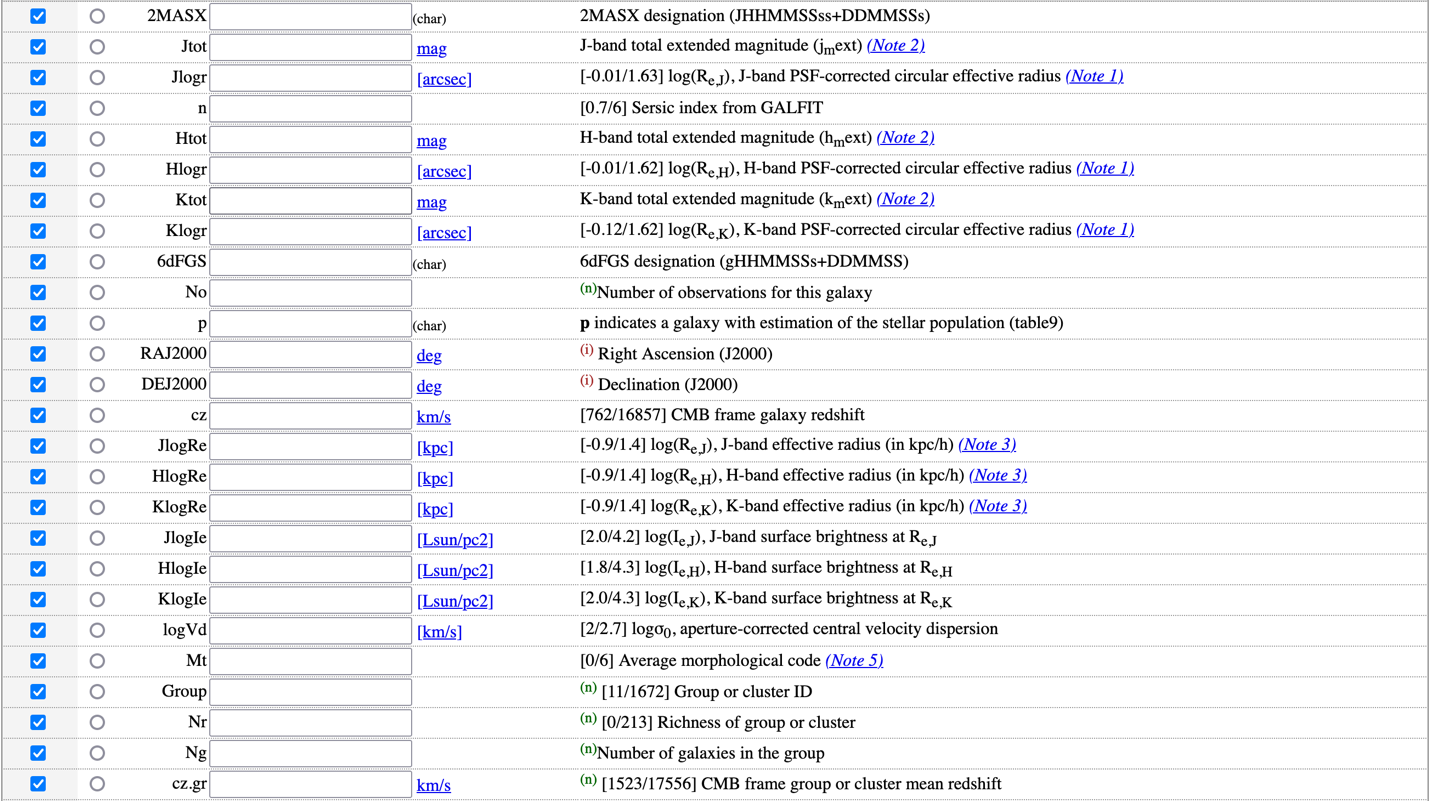
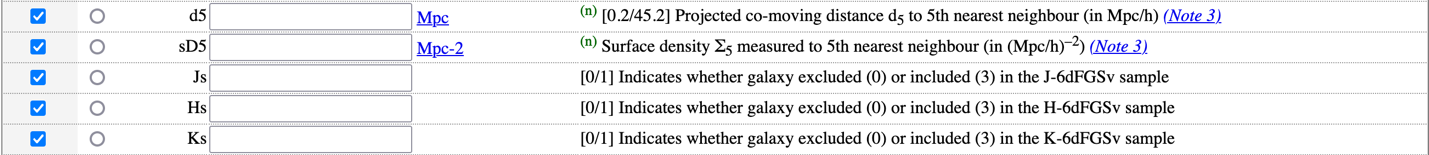
**WHAT I HAVE BEEN DOING**

1. Obtain the raw data:
   * 6dFGS (there are two data sources, I have not inspected whether they are different):
     + First source is what I actually used throughout my thesis, Campbell’s table on Vizier obtained [here](https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=+J%2FMNRAS%2F443%2F1231%2Ftable2&-from=nav&-nav=cat%3AJ%2FMNRAS%2F443%2F1231%26tab%3A%7BJ%2FMNRAS%2F443%2F1231%2FFPsample%7D%26key%3Asource%3DJ%2FMNRAS%2F443%2F1231%2Ftable2%26HTTPPRM%3A%26) . I use two tables, first is FPsample table (FP quantities) with the following schema:





second table is table2 (velocity dispersion) with the following schema:

A screenshot of a computer

Description automatically generated

* + - Second source is from Campbell et al. (2014) on MNRAS, the files campbell\_table2.ascii (velocity dispersion data), campbell\_table4.ascii (NIR photometry) and campbell\_table8.ascii (derived FP data) obtained [here](https://academic.oup.com/mnras/article/443/2/1231/1060810) (supplementary data).
  + SDSS: obtained from SDSS DR14 via the following CasJobs query:

SELECT s.specobjid, p.objID, s.ra, s.dec, s.plate, s.instrument, s.mjd, s.fiberid, p.devMag\_u, p.devRad\_u, devAB\_u, p.devMag\_r, p.devRad\_r, devAB\_r, s.z,s.zErr,s.veldisp,s.veldispErr, em.sigmaStars, em.sigmaStarsErr

into mydb.SDSS\_spectro

From SpecObjAll as s

JOIN emissionLinesPort em ON (em.specObjID = s.specobjid)

JOIN PhotoObjAll p ON (p.specObjID = s.SpecObjID)

WHERE

(s.sdssPrimary = 1)

AND (s.z <= 0.1)

* + LAMOST: the file lamost\_DR7\_VDcat\_20200825.fits obtained from Khaled.

1. Obtain the sky coordinates from all the raw galaxies. I use the get\_coordinates.py script to fetch the (ra, dec) and store them to data/preprocessed/sky\_coord/<survey>.ascii file (IPAC format). I had to multiply 6dFGS ra by 15 to get the ra in degrees (it was given in hour), while for SDSS and LAMOST I obtained the ra and dec directly.

I set the cone search radius to be 1 arcsecond and checked the One to One Match so that I can simply merge the resulting dataframes later. These are the fields that I queried:

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

I saved the results at data/raw/2mass/<survey>\_tmass.csv. Also backed up the files in Google Drive (cannot push them to GitHub as they are too large).

1. Merge the original spectroscopy data with all others in the following order:
   * 2MASS (from the response above): this file has the same number of rows as the original spectroscopy data, and I checked that their orders are the same, so I merged them simply by using their indices. Then I dropped all rows where ‘designation’ is null (no 2MASS counterpart).
   * John’s radii measurements: merge them by using the 2MASS id as primary key. Actually the number of galaxies are different (this time SDSS sample has much more), and not sure why.
   * For SDSS and LAMOST: cluster measurements from Tempel et al. (SDSS DR8 groups and clusters) done in two steps:
     + First, I performed sky crossmatch for the individual galaxies
     + Second, I join the individual galaxy with the group and cluster table to obtain the cluster redshift, member, and richness