Code Structure

/thesis-research-2.0

|-- data

|-- raw

|-- spectroscopy

|-- photometry

|-- processed

|-- spectro\_photo

|-- etg\_selected

|-- foundation

|-- fp\_sample

|-- pv\_sample

|-- smoothing

|-- bulk\_flow

|-- utils

|-- cosmoFunc.py

|-- calc\_kcor.py

|-- src

|-- preprocessing.py

|-- select\_etg.py

|-- fit\_fp.py

|--

|-- README.md

1. data
   1. raw
      1. spectroscopy

* LAMOST DR7: directly obtained from Khaled.
* 6dFGS: using table 2 and 8
* SDSS: pake query di bawah:

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)

preprocessing.ipynb

PART 1: menghitung variabel-variabel relevan

* ngambil data dari raw\_data/spectrophotometry 🡪 udah kombinasi data spektroskopi dan fotometri. Step untuk mendapatkan data ini lupa ngapain aja, tapi kemungkinan:
* Ubah format file jadi csv (dari fits atau ascii)
* Join data spektroskopi dan fotometri (2MASS)
* Join dengan data grup dan cluster (buat dapetin redshift)
* Bubuhin ‘SDSS’ dan ‘LAMOST’ ke objID SDSS dan obsid LAMOST
* Untuk galaksi SDSS dan LAMOST, apply kriteria seleksi dari John:
  + fit\_ok\_j == ‘OK’
  + log\_r\_h\_model\_j > 0.
  + red\_chi\_j <= 2.
* Hitung PSF-corrected radii:
  + Data SDSS dan LAMOST:
    - Hitung besar koreksi: delta\_r\_j = 10^log\_r\_h\_smodel\_j – 10^log\_r\_h\_model\_j
    - Hitung radius terkoreksi: 10^log\_r\_h\_app\_j – delta\_r\_j
  + Data 6DFGS:
    - Radius terkoreksi langsung dihitung dari 10^Jlogr
* Hitung redshift dalam CMB frame untuk setiap galaksi menggunakan code dari Khaled, kemudian gunakan data redshift grup atau cluster jika tersedia (perlu join individual survey dengan data grup).
* Apply koreksi apertur terhadap dispersi kecepatan:
  + Pake rumus empiris buat konversi J band ke R band
  + Hitung koreksi apertur menggunakan nilai apertur masing-masing survei:
    - LAMOST: apertur 1.65”, pake kolom veldisp dan veldisp\_err
    - SDSS: apertur 1.5”, pake kolom sigmaStars dan sigmaStarsErr
    - 6dFGS: apertur 3.35”, pake kolom Vd dan e\_Vd
* Hitung ekstingsi Galaksi di pita JHK menggunakan dustmaps (masukin ke utils): input ra dec dalam degree (cek!)
* Hitung K-corrections (masukin ke utils)
  + Perlu input redshift heliosentris, warna J2-H2 dan J2-Ks2.

PART 2: MENURUNKAN r, s, I (mungkin codenya dipisah aja)

* Generate redshift-distance lookup table (masukin ke utils/cosmoFunc)
* Hitung r:
  + Hitung estimasi jarak dari redshift cmb cluster
  + Hitung radius menggunakan circularized apparent radii: radius sudut dikali akar rasio b/a
  + Hitung menggunakan rumus r (konversi ke jarak diameter sudut)
* Hitung i:
  + Hitung menggunakan rumus i yang panjang itu (yang ada koreksi macem2), tapi koreksi evolusi diabaikan
* Hitung s: tinggal log10 aja, tapi untuk SDSS ada nilai negatif dan outlier ekstrim (nilainya sangat tinggi), jadi waktu itu diseleksi kuantil < 0.995
* Save data yang sudah ada rsi 🡪 masuk ke first\_step

PART 3: KALIBRASI DISPERSI KECEPATAN

PART 3.1: TENTUKAN GALAKSI YANG ADA PENGUKURAN BERULANG

* Untuk mendapatkan kalibrasi yang lebih baik, tidak dilakukan seleksi morfologi terlebih dahulu agar menggunakan data sebanyak mungkin
* Langkah pertama: ambil data dari folder 1st\_step: 2MASS\_id, masing2 id survei, dispersi kecepatan dan errornya
* Join ketiga survei untuk mendapatkan galaksi yang overlap: 6df-lamost, 6df-sdss, dan sdss-lamost. Kemudian join lagi untuk mendapatkan list galaksi yang ada di ketiga survei.
* Buang galaksi yang ada 3 repeat measurements dari list 2 itu, kemudian semuanya concat dan join dengan data dispersi kecepatan
* Hasilnya disave agar bisa generate grafik dan melakukan fine tuning

PART 3.2: CARI ERROR SCALING (HANYA UNTUK SDSS DAN LAMOST) 🡪 sepertinya perlu dirombak, pelajari lagi algoritmanya

* Buang galaksi yang eror dan offsetnya sangat tinggi
* Pertama, hitung scaling eror lamost: tebakan awal 1.0
* kemudian hitung statistik epsilon
* Hitung scaling yang baru
* Repeat until convergent (di code sekarang kriteria stopping hanya Nmax)
* Print nilainya

PART 3.3: HITUNG OFFSET DENGAN ALGORITMA WEGNER

* Sepertinya tinggal copas
* Print nilainya
* Inspeksi distribusi chi setelah apply scaling dan offset
* Apply kalibrasi dan simpan ke kolom baru, save hasilnya

PART 4: APPLY SELECTION CRITERIA

* Seleksi:
  + Upper redshift limit z\_cmb\_group < 16120
  + Lower redshift limit z\_cmb\_group > 3000
  + Magnitude limit < 13.65
  + Veldisp limit > log(112) + offset yang diperoleh
* Hitung galaksi yang diexclude di tiap seleksi
* Lakukan yang sama untuk galaksi low redshift -> diexclude dari fitting FP karena scatternya besar, tapi untuk perhitungan PV dimasukkan kembali karena harusnya sangat akurat.

PART 5: TURUNKAN EROR FOTOMETRI (kenapa ini gak sebelum selection criteria? find out why)

* Sama seperti kalibrasi veldisp, kali ini pilih galaksi yang unik saja (karena fotometrinya harusnya sama persis).
* Kolom yang diperlukan: 2MASS id, j\_m\_ext, extinction\_j, i\_j, i\_h, i\_k. Semua diconcat, terus drop duplicates 2MASS id
* Hitung warna, kalkulasi variansi tiap bin magnitudo, plot
* Fitting terhadap piecewise linear: konstan di m<m\_0, linear setelahnya. Perlu automasi buang outlier
* Hitung eror fotometri i dan r: eror r = 0.5 eror i

PART 6: SELECT ETG

* 6dFGS: pake sampel terbaru Khaled
* SDSS: pake sampel Cullan
* LAMOST: pake seleksi visual dari John

PART 7: SELEKSI KOLOM FINAL UNTUK FP

* Kolom yang diperlukan untuk fitting FP: ['2MASS', ‘survei id’, 'ra', 'dec', 'z\_hel', 'z\_cmb', 'z\_cmb\_group', 'j\_m\_ext', 'extinction\_j', 'kcor\_j', 'r\_j', 'er\_j', 's\_corr', 'es\_corr', 'i\_j', 'ei\_j']
* Combine data SDSS dan LAMOST