File organization: NLP in DR Crunchr folder

* **Follow-up and Tag**: has all the scripts for followup and hits (most up to date is **FollowUp 1\_27.ipynb)**
  + **Input** has the input excels for just the followup and hit scripts
  + **Output** has the output csvs from followup & hit)
* **Geocoding** has the scripts for geocoding addresses (for fips comparison and address standardization)
  + **Input** has the input files for both fips and address standardization
    - Pataddressdim.csv is for address standardization and all others are for FIPS
  + **Output** has the output from address standardization
  + **Graphs** just has some graphs of the histogram of levenshtein matches
  + All scripts are title Geocoding Script and those followed immediately by a date are for address standardization and the others are for FIPs.
    - Most up to date address standardization is **Geocoding Script 3\_20\_ratio\_zip.ipynb** and fips is **Geocoding Script FIPS 3\_14.ipynb**)
* **Eye Scan and other scripts:** misc scripts used to test
* **Data Cleaning BCVA:** scripts used for getting the best corrected vision acuity for each eye

Wilson desktop folder:

* Text files: holds txt files with old script code
* **SAS Geocoding Files:** 
  + **SAS inputs**: contains all the files from geocoding scripts and that is input to SAS
  + **SAS outputs:** contains all the files output from SAS that have new FIPs
  + **Geooder.**sas: the SAS code run on SAS inputs
* **Followup & Hit Outputs:** contains outputs of followup & hit scripts
* **Wilson Annotations:** annotations of first 1000 notes (before running scripts on them)
* **Status Reports:** followup & hit script reports of accuracy and other statistical benchmarks
* **CXC notes:** Dr. Cai’s annotations of the notes
* **Original Notes:** unannotated notes split into 2 sets of 1000 prob list and 1000 prog notes
* **All notes 12.31:** the full set of notes