2.11.16

* Many strings as values – this is a problem; value corresponds to a label even though it is stored as an integer
* This might require a lot of shell script
* Data is in wide format; most relevant viz tools use long format
* Learn about pipes – like method chaining
* Read about miniConda and conda environments
* Do the csvkit tutorial - <https://csvkit.readthedocs.org/en/0.9.1/>
* Use the conda environment from now on
* Created an environment called (csv) for conda - this is not a location; it’s an environment that gives access to the stuff installed
  + To activate this environment, use:
  + $ source activate csv
  + To deactivate this environment, use:
  + $ source deactivate
* Note – Learn to use ‘sed’ or ‘awk’
* Read Lynn’s articles
* Need to learn to use sed/grep to remove NULLs
* Need to do something about PrivacySurpressed (this shows up as a string and messes up numeric columns)
* ‘Replicable’ data processing pipeline
* Between codebook and CSV, the columns numbers might not match it. We need to figure out what the numbering scheme is
* Trifacta Wrangler – original file was too big
  + Learn how to unpivot or denormalize data

2/16/16

* source activate csv
* conda install pandas (also installs numpy)
* conda install jupyter
* Wrote bash script `sedscript.bash` for cleaning out Nulls and PrivacySuppressed values
* Need to learn `seaborn` for viz