**Data Cleanup & Fusion**

First I loaded the cs171\_survey\_4years file into Google Refine and then exported it as a csv file so it would be in the same format as the cs171\_survey\_2013 file. Both files were then loaded into Google Refine again and the following clean up was done:

1. Combined the ‘Comfortable with Programming’ field with the ‘Programming skill’ field
2. Added the year 2013 to the year column for the 2013 survey data
3. Combined the ‘Degree’ and ‘Primary Concentration’ fields
4. Combined the ‘Secondary Concentration’ and ‘Secondary Concentration (if any)’ fields
5. Combined the ‘Languages’ and ‘What other languages do you know?’ fields
6. Combined the ‘Coding skills’ and the ‘Overall, how comfortable are you with programming?’ fields
7. Combined the ‘Programming Experience’ and the ‘How long have you been programming’ fields
8. Removed the ‘How comfortable are you with design’ column as it was not present in the previous years’ surveys.
9. Replaced occurrences of ‘Division of Continuing Education’ with DCE in the type field
10. Performed the following clean up on the student’s locations
    1. Separated out the 2013 data into individual columns for City, State and Country
    2. Normalized the Country name (i.e United States to USA)
    3. Normalized the State abbreviations (i.e. Mass to MA)
    4. General cleanup of some students comments
    5. Removed the town column as it wasn’t present in the previous years’ surveys