How to Norming Processing

|  |  |
| --- | --- |
| Things to do! | File to clean up with |
| Spell check.   * consider changing commas to spaces |  |
| Put together files by subject/word.   * format should be txt * target response * space delimited * make sure that each line starts with a word you expect |  |
| Read those files to create files by word.   * make sure there are no extra words because the lines are off | Reader\_csv.pl |
| Process the word files for frequency. | Word counter folder, put in files folder  Run index.php |
| Combine words that are the same. | By hand ☹. |
| Clean out the common useless words.   * takes out descriptive verbs (are) * numbers * word frequencies of one * prepositions * pronouns | Run cleanupcrap.pl |
| Reduce the features across words.   * that are not used a certain percentage of the time * one frequency words (especially for smaller N – leave in originally because you need to combine them). | Reducefeatures.pl |