**Problem:** How do I add features into the NLTK model or Word2Vec model where I can get separation between the clusters using the amputation type, or location, etc? I can’t seem to find good code examples to use. I liked the output from Word2Vec if it would have worked. Perhaps I need to use a k-means package but I’m still not sure how to tell the algorithms to consider my extra true false columns. I’ve read so many books, readmes and documentation I’ve become lost on how to get to where I want to go. If I could just get the inputs considered then I could play with optimization, bi-grams etc. But I’m lost on the basics now.