DATA 902 – Text Mining

**STUDENT**: Kim Lowell

**NARRATIVE**: I pulled postings from a UFO blog site. Because blogs are considerably longer than tweets (that are limited to 280(?) characters), I pulled about 700 blog postings.

**Github address**: https://github.com/KLowellNZ/UNHTextMine

Included in what is submitted are the following:

* Python code used to scrape UFO blog postings.
* Python code to convert the blog postings to a .csv file suitable for input into R’s text mining packages.
* R code to undertake the necessary analysis.
* The .csv file on which analysis was undertaken.
* This document showing screen shots of all required output.

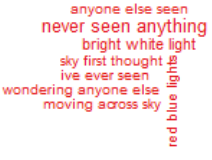
1. Download (712) unique blog postings.
2. Create word clouds:
   1. Unigram:



* 1. Bi-gram:



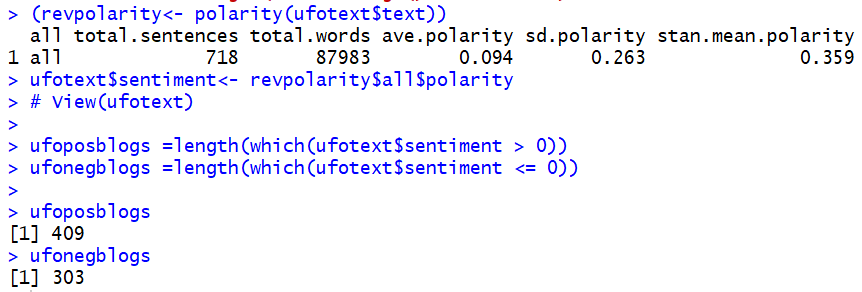
* 1. Tri-gram:



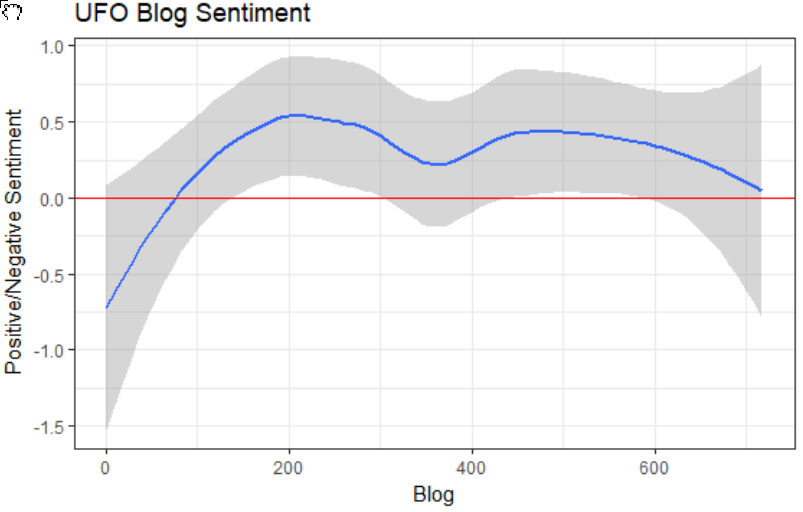
1. TF-IDF word cloud:

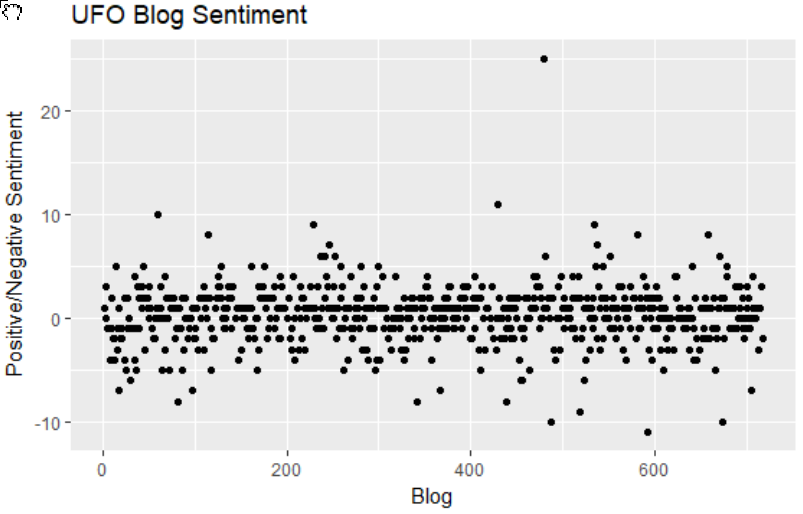


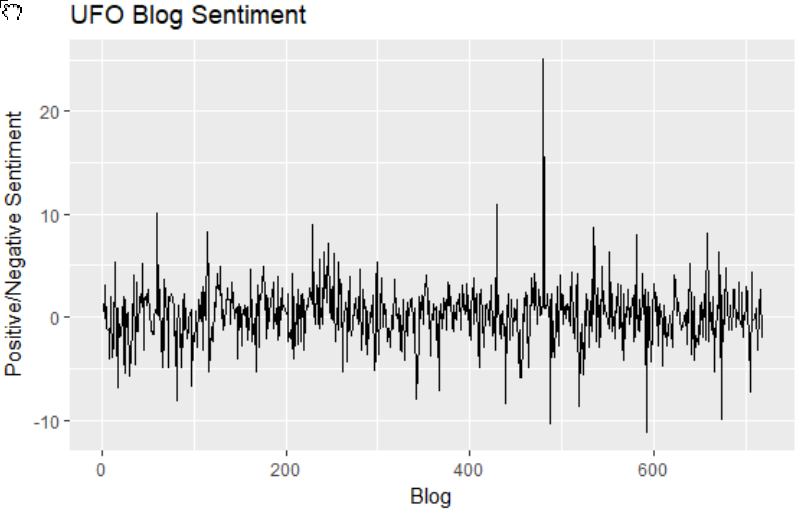
1. A. The number of tweets having positive/negative sentiment. (NOTE: Total words is about 88K.)



B. Various ways if plotting (positive or negative) sentiment.



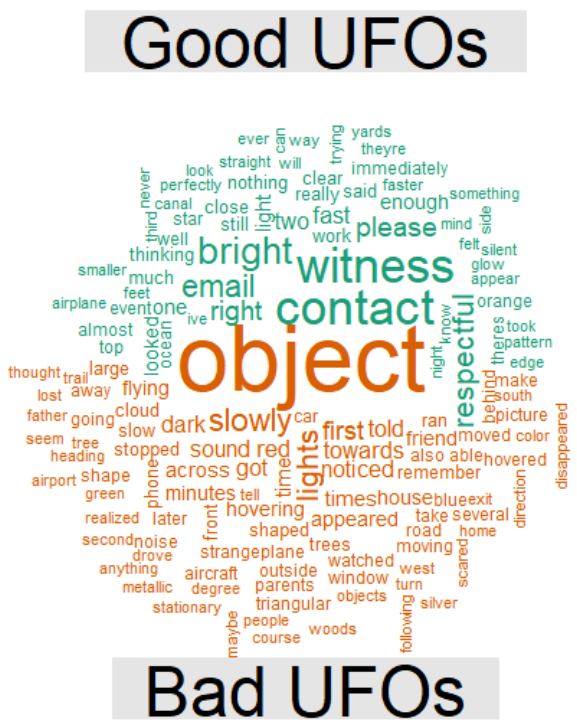




1. Corpus split into positive and negative blog postings.
   1. Commonality cloud:



* 1. Comparison cloud



1. Emotional radar chart

