David P. Stonehouse

University of Maryland University College

DATA 650 – Big Data Analytics

Fall 2016, Section 9040, Professor Gortcheva

Assignment :

## Introduction/Purpose

study the minutes from one of the Raleigh City Council committees.

## Experimental Design / Method

example, all of the documents start by mentioning who was present.

## Results

appendix has a listing of words that correlate with “trailwood” and “birch”, two of the top agenda item words.

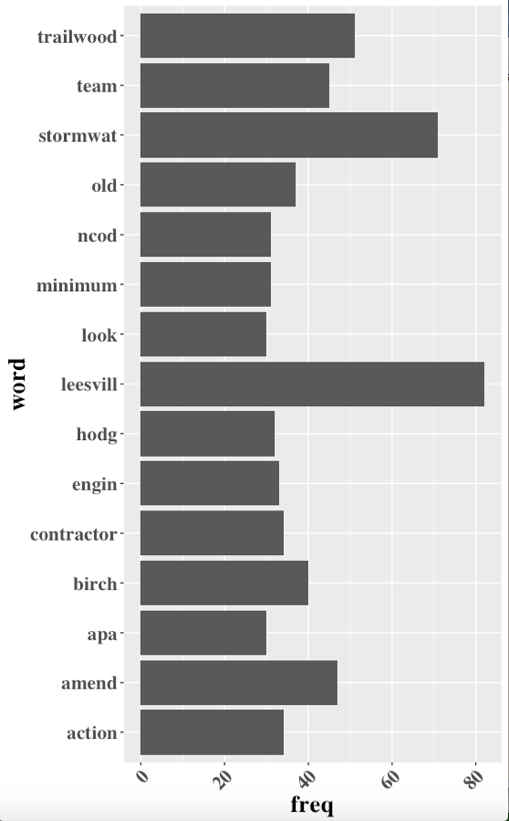
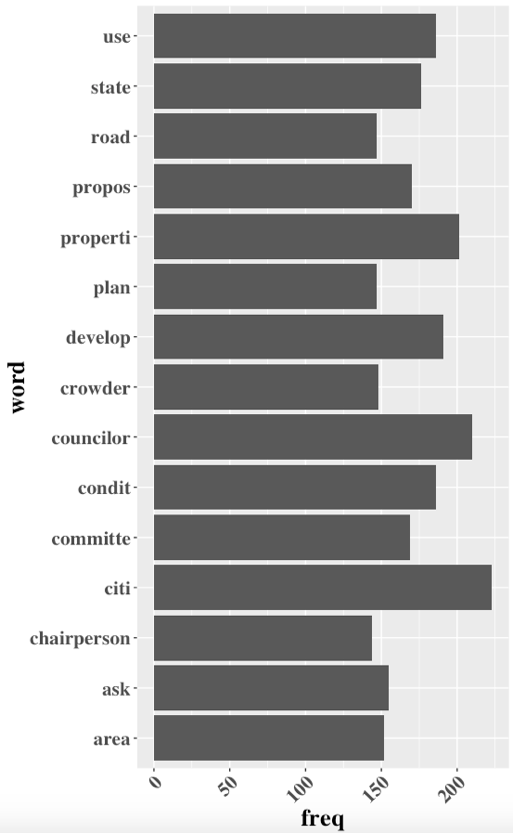


Figure 1. Term frequency plots: a) all terms b) after removing terms used in every document

## Limitations and Future Research

much easier to understand. E.g., “properti” is restored back to “properties”. However, changes related to version 3.3.1 of R made stem completion difficult to use.

## Conclusions

come up with topics even more targeted to mid-town communities of most interest to the readers.

## References

Ganesan, K., 2014. “All About Stop Words for Text Mining and Information Retrieval”. Retrieved from http://text-analytics101.rxnlp.com/2014/10/all-about-stop-words-for-text-mining.html

## Appendix: Supporting Information

**Selected logs from R code:**

**R source code:**

Now try with tf-idf

kfit <- DoKMeans(m.tf.idf.transpose2,8)