

Homework 1 – Quanteda Getting Started

This homework is based on this introduction to quanteda:

<https://cran.r-project.org/web/packages/quanteda/vignettes/quickstart.html>

The starting code for this homework is here.:

<http://faculty.washington.edu/jwilker/559/2018/Homework1.R>

The more general quanteda reference (options are alphabetized) is here:

<https://cran.r-project.org/web/packages/quanteda/quanteda.pdf>

This assignment can be completed quickly, but that's not the point! Use it to learn more about R and quanteda. Test your understanding by doing things a little differently, or by looking at the quanteda documentation to understand the different options available etc. Explore some of the other quanteda tools that can be applied to a corpus.

1. Collect at least 5 text documents (.txt, .html, .csv, .json), place them in a folder on your desktop (if they are short, then collect more!).

[There are some .html files here if needed: <http://faculty.washington.edu/jwilker/559/FOMC/>]

2. Set your working directory to that folder

First check your current directory; reset it; and then check it again to make sure you are there

3. Use the `readtext()` command in quanteda to import your documents into a dataframe object

Check that it worked. If you get an error, try to decipher the cause (try google too!)

Look at the second document

4. Convert the data frame object to a quanteda corpus

Look at the 2-4th documents

5. Convert the corpus into a document frequency matrix *but do not remove stopwords*

View the top features

6. Repeat 5, *but remove stopwords*

7. Create a wordcloud from your dfm

8. Create a dfm that is based on a dictionary that you have created

9. Create a 10 topic LDA topic model using your first DFM; return the first 6 terms of each topic

10. Create a new corpus that chunks your original corpus into sentences.

Look at the 3rd through 5th sentences of the 3rd document

11. Write this last corpus to a text file and save it to the same folder where you put the original texts. Confirm that it is there and that it includes the correct content. Save your R file.