1 - Write a script to compare the similarity of FOMC speeches (or you are welcome to use another data source!). This will entail:

* Reading in the meetings transcripts
* Parsing by speech
* (stopwords, stemming etc are options – benefits less obvious for text reuse)
* Writing a swalign function
* Calling that function and applying it to the file list(s)
* Investigating whether you learn anything from the speeches that produce the highest similarity scores

2 - Test the argument that the Federal Reserve’s response to the financial crisis was hindered by a preoccupation with inflation.

-For starters, produce a graph of attention to inflation over time (using quarterly meetings).

Do not assume that the only relevant word is ‘inflation,’ and do not assume that raw word counts are the best measure of attention.

-Then use an NLP approach to test whether there was “growing concern” at the Fed about the mortgage crisis and when this occurred. Overlay these results on the inflation graph.

The most obvious approach would be to create a dictionary of terms (unigrams, bigrams) related to the mortgage crisis. But you could also consider an approach that captures the amount of discussion focusing on the mortgage crisis (as opposed to keyword frequencies), or even the ‘anxiousness’ of the discussion etc.