Text (Assignment)

- Don’t do all of these (unless you want to), but be able to do:
- Get some text (50 or more “documents”, each at least 1000 words)

- Find 20 most common, least common words
- Remove stop words
- Find similarity matches (between texts)
  - (Should be based on word frequencies)
- Find Bigrams, trigrams

Text (Assignment)

- Frequency count characters
- Frequency count character pairs
- For a given word, how frequently does it appear? Next to other words?

Text

- Free text examples:

- https://www.csun.edu/science/ref/reference/public_domain_text.html
- https://www.gutenberg.org/
- Stop Words:
- https://www.ranks.nl/stopwords