**Topics we’d like to cover with our module...**

**Examining one document**

* What terms are important to this document?
  + Word frequencies
    - Bar charts, scatterplots, line graphs, word clouds
* Can we identify different themes?
  + Clustering (topic modeling)
  + What are the differences between the clusters?
    - Bar graphs to show different terms within each topic, boxplots to show how documents were assigned to certain topics
* Does the document have emotion/attitude? Does it change over the course of the document?
  + Sentiment analysis
    - Ex: How “positive” is the NYT’s article about the State of the Union address?
    - Breaking down a novel by chapters → can you see where the high and low points of the book occurred?
    - Bar charts (many different types work well here depending on what you’re comparing), word clouds

**Comparing multiple documents (a corpus)**

* Looking for similarities and differences between documents
  + Using tf-idf to explore relationships between documents (whether that be emails, different parts of a dataset, articles, books, etc) and to explore relevance to certain keywords → bar charts
  + Measures of correlation → bar charts, network graphs
  + Sentiment analysis
  + Do they use different words to talk about the same broad topic? Do they talk about completely different topics?
  + Compare word usage between documents or twitter users by calculating log odds ratios (was a word that was used more likely to have been said by person A or person B?)