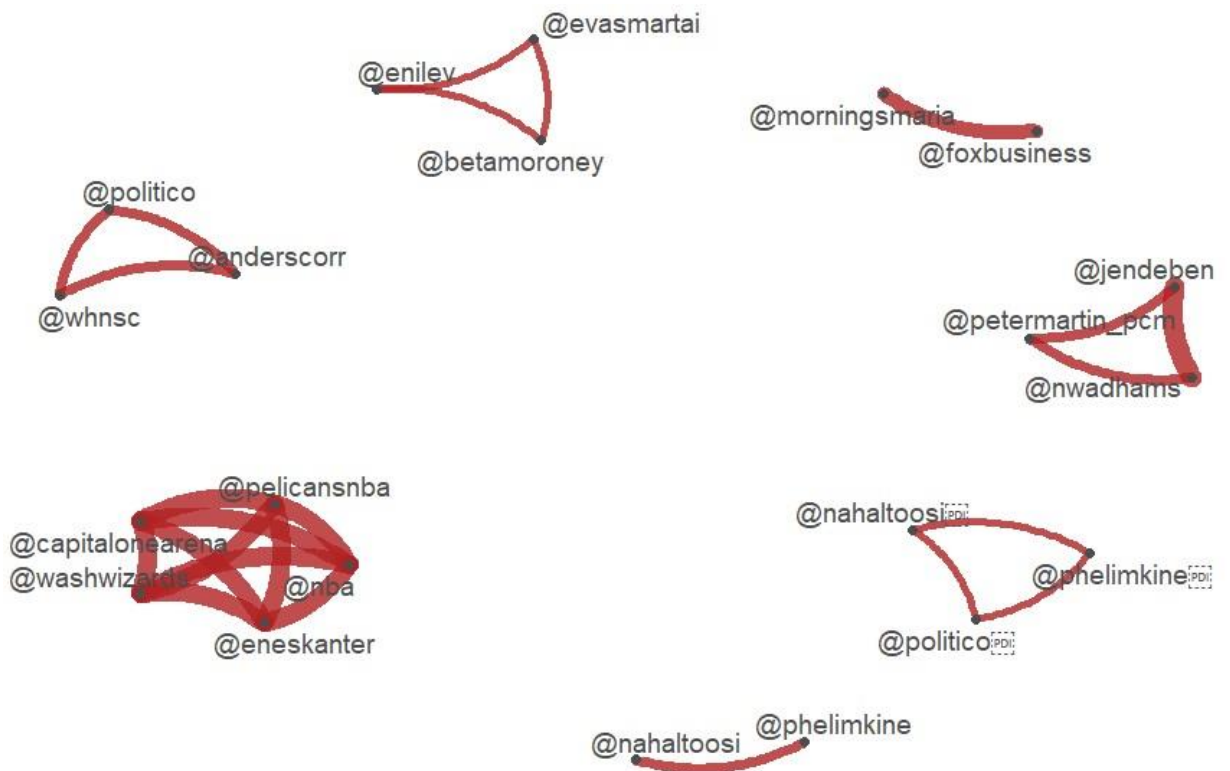


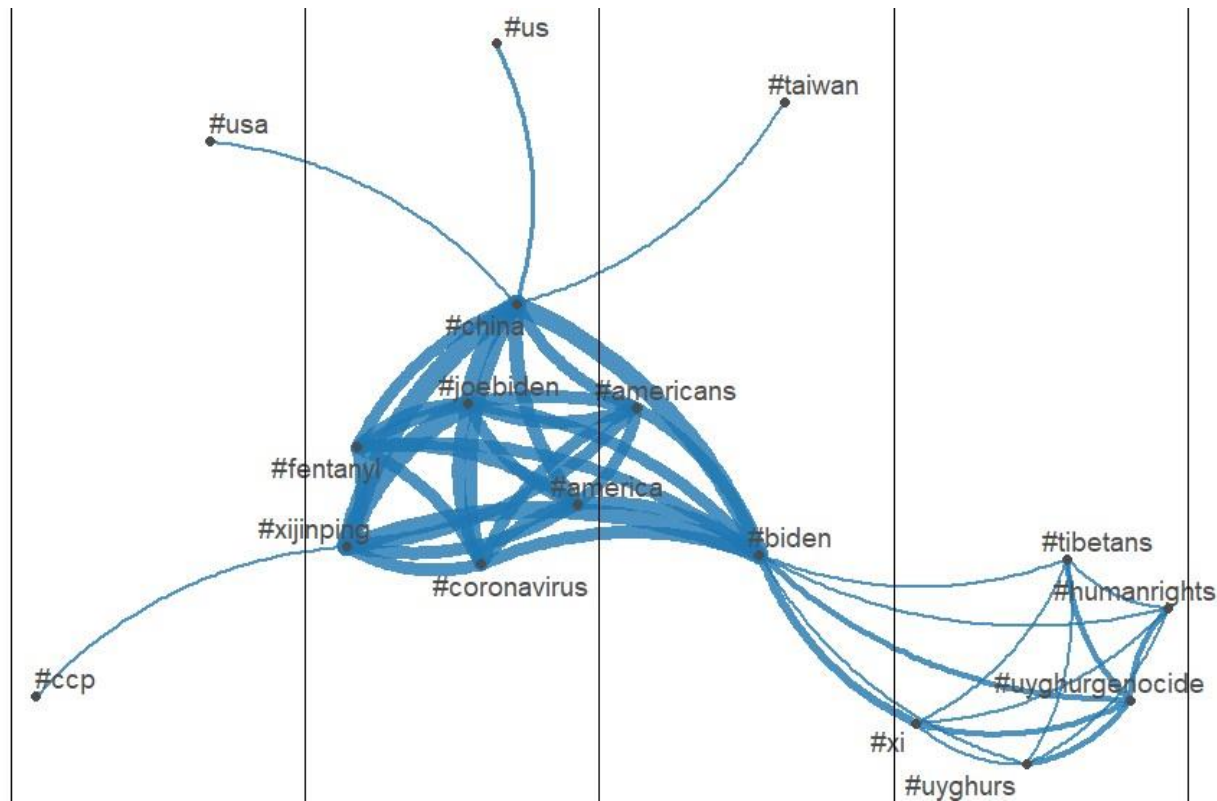
Assignment 3: Quanteda

3. Analyze

a. Biden-Xi summit data

- a. The following data is Twitter text data scraped into R code. The data contains multiple observations, and each of them are based on articles posted to Twitter involving Xi Jinping and Joe Biden. The data points are also organized by quote and retweet. Along with the article titles, the names of the users tweeting about those articles are displayed as well. This dataset is very large, as it contains well over 1,000 observations. This dataset overall describes relations between Joe Biden and Xi Jinping, and contains articles comparing their actions involving one another, as well as international relations.

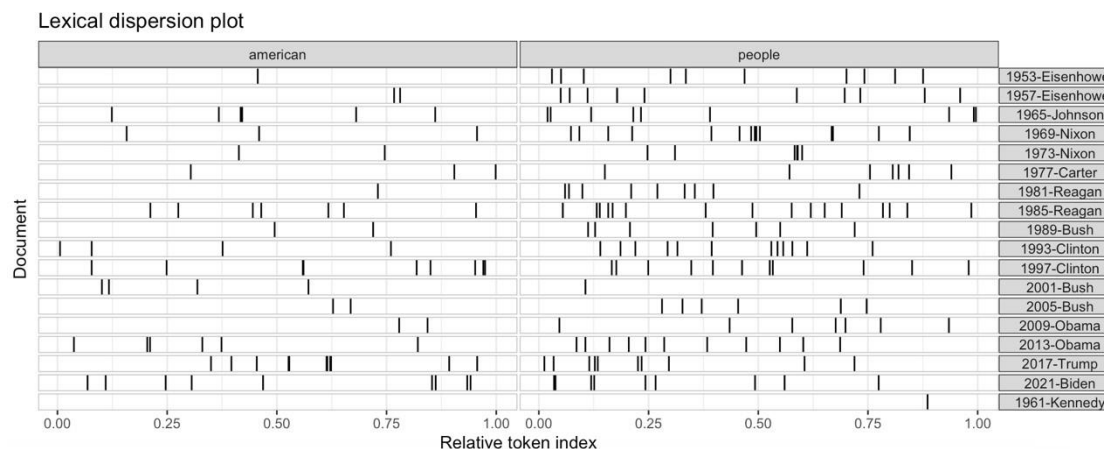




b. US Presidential Inaugural Speeches

a. This dataset contains observations regarding the inaugural speeches of US presidents going all the way back to President Eisenhower and ends at President trump. The graphical displays demonstrate the frequency of which certain words were used in the speeches, such as “people,” “American,” and “communist.”

i. It makes sense that the word “communist” would be used more during the Nixon and Reagan eras, as the Red Scare was at a very high point during those times. What is interesting, though, is how little “American” has been mentioned in the dataset, outside of recent presidents. “American” seems to have been used more by the most recent presidents than later presidents.



- c. “Wordfish” describes the prediction of the positioning of one-dimensional documents. Wordfish does this by simply observing the number of times words are used. Typically used in R, one can organize data based on the frequency of the words used, as was done before with the inaugural speech data.