

Text and regular expressions

06 March, 2023

Write the R code to answer the following questions. You have until the beginning of next class to answer all of the questions below and commit to GitHub. **It's okay if you want to do this in a .R script. Because the data is so large, the code might run slowly, and you might not want to knit.**

Overview

We will continue using the 'tweets' data from class containing mayor Tweets. As a reminder, these are data shared with me. Please do not use beyond class without inquiring with me further, and do not post publicly.

Question 1

- Subset down to all of the tweets from Mayor Lyda Krewson (STL Mayor 2017 – 2021)
- Find the mean number of words in her tweets. Find the total number of unique words.

Question 2

- Come up with your own approach for searching for a dictionary of words to find Tweets relevant to some concept, such as policing, schooling, climate change, gun violence, etc. It's up to you. This isn't a dissertation project, so just try to look for a concept using our new regex skills. It won't be perfect measurement, and that's totally fine.
- Then, for each mayor in the dataset (N=572), what number of tweets match your criteria? Or, you might look at the proportion that match instead. Plot a histogram of the results using the `hist` function.