Applied Statistical Programming - Text

3/9/2022

Write R code to answer the following questions. Write the code, and then show what the computer returns when that code is run. Thoroughly comment your solutions.

You have until the beginning of class 3/21 at 10:00am to complete the assignment below. You may use R, but not any online R documentation. Submit the Rmarkdown and the knitted PDF to Canvas. Have one group member submit the activity with all group members listed at the top.

`regex` practice

Write a `regex` expression that would be able to identify phone numbers with the following patterns.

- 1234567890

- 123 456 7890

- 123-456-7890

- +1 1234567890

```
# Your code
```

Text-as-Data

For this exercise, import the mayors data that we've been using.

- https://raw.githubusercontent.com/jmontgomery/jmontgomery.github.io/master/PDS/Datasets/Mayors.csv

Using string and/or `regex` commands, find how many mayors tweeted about (1) police and (2) Black Lives Matter.

You will need to define a list of synonymous terms related to law enforcement as well as Black Lives Matter. Your code comments should describe why you think the root you chose is sufficient to identify the aforementioned tweets. For example, "police", "policing", and "polices" are all relevant terms, but you would only subset based on the root "polic" and include a condition to exclude "policy" and "policies".

Once you have found the counts, plot the frequency distribution for each of (1) and (2). Be sure to properly label your figures.

```
# Your code
```