Part I: Coding assignment.

https://www.gpo.gov/fdsys/pkg/BILLS-111hr4173enr/html/BILLS-111hr4173enr.htm

Write a script that:

1. downloads this document and reads it into python
2. counts the number of words in the document
3. parses the file wherever 'SEC.' is found and saves the results to a new list (SEC. denotes a section of the bill)
4. counts the number of parsed elements in this new list
5. removes punctuation, stopwords, and words < 4 characters from each element in the list
6. identifies the top ten most frequent words among all of the elements in the list
7. removes one of these top ten words from all elements of the list (an example of removing a new stopword)
8. creates a dictionary where the 'keys' are two of the top 10 words, and the 'values' are the frequencies of those words within each element in the list.
9. pickles this dictionary as a .csv file

Submit your code and the .csv file to the dropbox!

Part II: Identifying and designing good text as data projects

Now that you are an expert on text as data methods, begin by reflecting on the strengths and limits of these methods. Give an example of how they can be used to shed new light on one or more existing research questions in your field. Give an example of a new research question that can be investigated because of these methods. Finally, new methods are often met with skepticism. How should researchers anticipate and address such concerns where text as data research findings are concerned?

Your response should be no longer than one single-spaced page.

.