Please do your homework using an R script. Homework is NOT collected in this class. However, on the day it is due, you may be asked to share your screen and run some of your code when we discuss this homework in class. This will count towards your participation grade.

1. Make sure the tmData folder on your computer contains the appropriate files; see slide 9 of lecture 10.  
   1. Create a volatile corpus called co with the contents of all the .docx and .pdf files in the directory tmData; you should have five PlainTextDocument objects in the corpus. Inspect the contents of co.
   2. Put the variable desc in the corpus meta data with the value “Random Texts”. Then check to see if you have been successful.
   3. Show the content for every element of co. **Hint:** Check out the function lapply.
   4. Set the author part of the meta data of each document in co to some name (you may use your own). Then show the meta data for every element of co.
   5. Create a variable ptd that is equal to the fourth document in co. This variable should now be a PlainTextDocument with its content equal to a character vector of length 1, a string variable. Change this by splitting this string variable at each point where a new line is supposed to start (indicated by \n), and update the contents of ptd such that it has a character vector with an element for each piece instead. The content of ptd should be equal to a character vector of length five at this point; check this. Then find out the number of characters in each element of the content vector.
   6. Change the third element of the content of ptd, changing everything that says “I am” to “We are”, and every other occurrence of “I ” to “we ”.
   7. Change all letters in the content of ptd to lower case. Then change the first letter of each element of the content of ptd to upper case. **Hint:** Use the function toupper as well as some other functions we talked about in lecture 10.
   8. Create a term-document matrix for co, displaying only words that appear in all documents and have at least 2 characters. Then do that again, but this time also exclude stop words. Save that last matrix to a variable; you may call this variable mat.
   9. Add a column at the end of mat with the total frequency of the words across all documents in the corpus. Call this column “Total”.