# Upper Case Problem Instructions

**Problem 2 15 points**

**Introduction**

The hypothetical business scenario related to this problem is that your boss has given you a file that she needs to have analyzed. The file contains a header record and each of the other records contains a social media post.

Note: the posts in this test file are random phrases from the internet. The verbiage and my putting some words in uppercase are not intended to convey anything in particular.

Your boss wants you to create an R script that will read the file and create a new file based on data in the file she gave you.

**Instructions**

**Part 1**

Create a function named **get\_uc\_words**. This function will have one argument named **post**.

The function must return a character vector that contains the uppercase words present in the post argument. If the post contains no uppercase words the function will return a vector of length zero.

The function must use a regular expression that is written so that single letter words like “I” and “A” will not be considered to be an uppercase word.

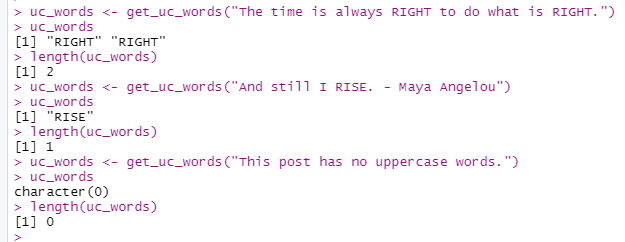
The function must use a regular expression to evaluate the data passed into it via the post argument.

Hint:

Consider using the following regular expression: \\b[A-Z]{2,}\\b

Consider using the str\_extract\_all and unlist functions.

After you write the function, consider unit testing it with code as shown below:



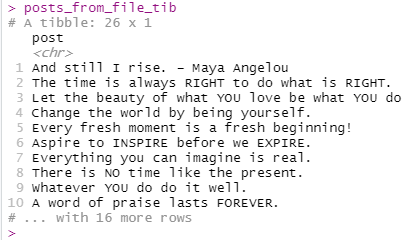
If you want to learn more about the general concepts of unit testing, consider reviewing this [web site](https://searchsoftwarequality.techtarget.com/definition/unit-testing#:~:text=Unit%20testing%20is%20a%20software,developers%20and%20sometimes%20QA%20staff.).

**Part 2**

For this part and the remainder of this problem, the file named **social\_media\_posts.csv** will be referred to as the input file. You should have previously downloaded this file from Blackboard.

Read the input file to create a tibble named **posts\_from\_file\_tib**.

Print **posts\_from\_file\_tib** to the console and the console must look like this:

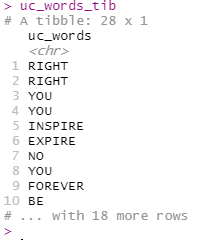
****

**Part 3**

Process data in **posts\_from\_file\_tib** to create a tibble named **uc\_words\_tib**. This tibble must contain the uppercase words found in the input file.

Your solution for this part must use the get\_uc\_words and lapply functions.

Print **uc\_words\_tib** to the console and the console must look like this:



**Part 4**

Write out a file named **upper\_case\_posts.csv**. This file will contain the data in **uc\_words\_tib**. There must be no column header record in the output file.

You may visually compare your output file to the file named **instructor\_upper\_case\_posts.csv** to verify your script creates the output file as expected.