We now want to figure out which words appear most often in the headlines. We'll be developing another script, called count.py to accomplish this. We'll need to import our load\_data function from read.py into count.py so we can use it.

You'll recall that if you have a folder with two files, read.py and count.py, you can use the function load\_data in read.py from count.py by writing the following code in count.py:



import read

df = read.load\_data()

Instructions

Writing the script for this will require a series of steps:

* Make a file called count.py, using the file browser, or the command line.
* Import load\_data from read.py, and call the function to read in the dataset.
* The order in which you do the below two steps is up to you, but it's suggested to first combine all the headlines (you can use a *for* loop for this, among other methods), and then split everything into words.
  + Combine all of the headlines together into one long string. You'll want to leave a space between each headline when you combine them. [Here's](http://stackoverflow.com/questions/4435169/good-way-to-append-to-a-string) a good reference on joining strings.
  + Figure out how to split the long string into words. Each headline is a string, such as Anticlimax As Motivation Killer. Combining that with Swype acquired by Nuance for 100 million would look like Anticlimax As Motivation Killer Swype acquired by Nuance for 100 million. Adding more headlines would make a longer string. You'll need to figure out a way to split the long string, so you end up with a list of words. The documentation for [str](https://docs.python.org/3/library/stdtypes.html" \l "textseq" \t "_blank) might help here.
* You might want to think about lowercasing each word, so Hello and hello aren't treated as different words when you do a count.
* Find a way to count up how many times each word occurs in the list. The [Counter](https://docs.python.org/3/library/collections.html#collections.Counter) class might help you.
* Add code to print the 100 words that occur the most in your data.