Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Quiz 1: Getting Started

LATEST SUBMISSION GRADE
100%

1. The `en_US.blogs.txt` file is how many megabytes? **1 / 1 point**

- (•) 200
- ( ) 100
- ( ) 250
- ( ) 150

Correct
Do `ls -alh` in the `Coursera-Swiftkey/final/en_US` directory.

2. The `en_US.twitter.txt` has how many lines of text? **1 / 1 point**

- (•) Over 2 million
- ( ) Around 1 million
- ( ) Around 2 hundred thousand
- ( ) Around 5 hundred thousand

Correct
Do `wc -l en_US.twitter.txt` at the prompt (or git bash on windows) or `length(readLines("en_US.twitter.txt"))` in R

3. What is the length of the longest line seen in any of the three en_US data sets? **1 / 1 point**

- ( ) Over 11 thousand in the blogs data set
- (•) Over 40 thousand in the blogs data set
- ( ) Over 11 thousand in the news data set
- ( ) Over 40 thousand in the news data set

Correct
Again a simple wc command suffices `wc -L *.txt` inthe directory with the three files. Note, we had a small discrepancy between doing thin in R versus WC.

4. In the en_US twitter data set, if you divide the number of lines where the word "love" (all lowercase) occurs by the number of lines the word "hate" (all lowercase) occurs, about what do you get? **1 / 1 point**

- ( ) 0.5
- (•) 4
- ( ) 2
- ( ) 0.25

Correct
`grep "love" en_US.twitter` $wc - l|$ and

`grep "hate" en_US.twitter` $wc - l|$ gives you the counts. Then you could divide in whatever. If you never want to leave the console, you can use bc (not present on gitbash in windows). You could also read into R (`readLines`) and use character search.

This worked on gitbash

`love=$(grep "love" en_US.twitter.txt` $wc - l)|$ then

`hate=$(grep "hate" en_US.twitter.txt` $wc - l)|$ then

`let m=love/hate` then

`echo $m`

5. The one tweet in the en_US twitter data set that matches the word "biostats" says what? **1 / 1 point**

- ○ They just enrolled in a biostat program
- ◉ They haven't studied for their biostats exam
- ○ It's a tweet about Jeff Leek from one of his students in class
- ○ They need biostats help on their project

✓ **Correct**

`grep -i "biostat" en_US.twitter.txt` (note the -i doesn't matter since there's only one line ignoring case).

6. How many tweets have the exact characters "A computer once beat me at chess, but it was no match for me at kickboxing". (I.e. the line matches those characters exactly.) **1 / 1 point**

- ◉ 3
- ○ 0
- ○ 2
- ○ 1

✓ **Correct**

`grep -x "A computer once beat me at chess, but it was no match for me at kickboxing" en_US.twitter.txt` $wc -l$