The en\_US.blogs.txt file is how many megabytes?

* 200

The en\_US.twitter.txt has how many lines of text?

* Over 2 million

What is the length of the longest line seen in any of the three en\_US data sets?

* Over 40 thousand in the blogs data set

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?

* 4

The one tweet in the en\_US twitter data set that matches the word "biostats" says what?

* They haven't studied for their biostats exam

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.)

* 3