1. War and Peace, Leo Tolstoy (I think); I think it was a good piece. Very well written. I’ve been interested in reading it but never got it.

Step 2

1. Removing the unwanted characters produces n-grams of punctuation marks which is useless
2. Stop words are words that do not provide much information besides those that are used for grammatical rules, like the, that, and
3. Removing stop words allows us to extract more important terms. The extracted tokens do appear to be of better value.

Step 3

1. 2-grams: Total tokens are 170, 159 ngrams; upper lip has highest frequency, and always,case is lowerst frequency.
2. Health appeared quite a lot which is not surprising since the text talks about the health of the older lady.