

Problem 6

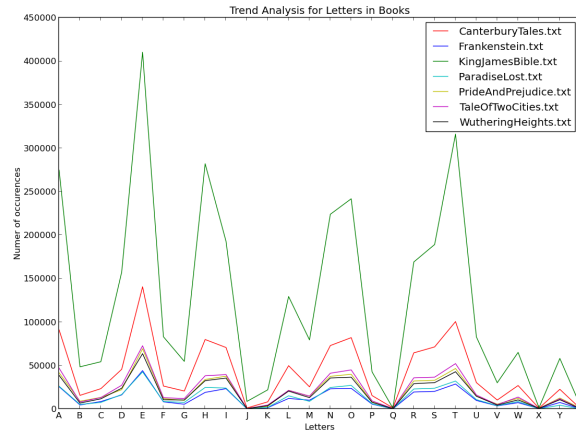


Figure 1: Line plot showing the number of occurrences of each letter in different books.

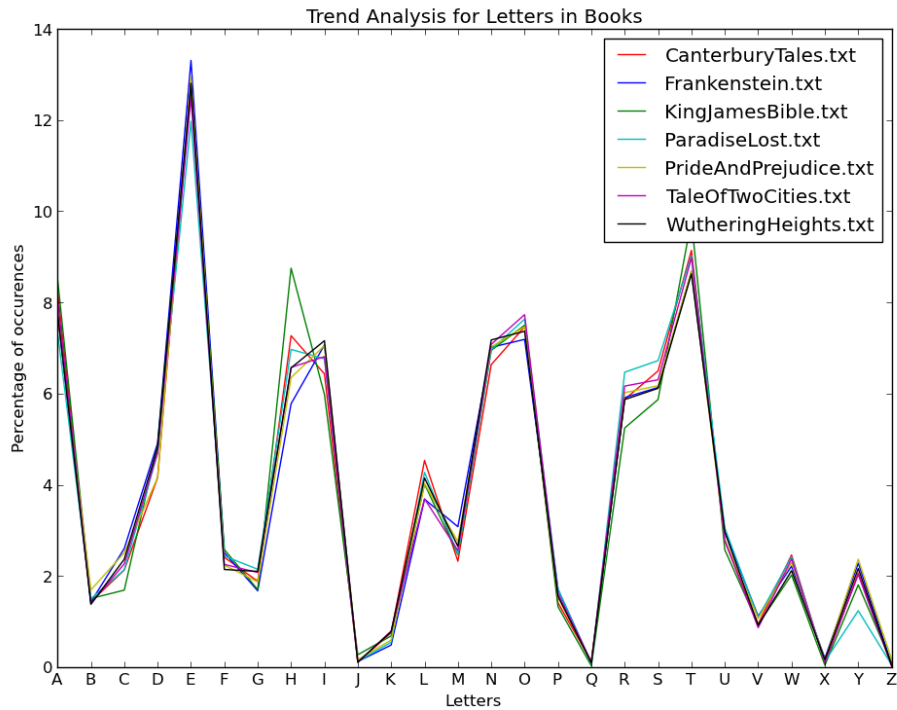


Figure 2: Line plot showing the ratio of the number of each letter to the total number of letters in different books.

The trends are easier to see on the normalized plot in Figure 2. We notice that most letters have similar occurrence percentages in different books. Most notable differences are noted in C, H, I, R, S, T, and Y. E is the most used letter. Letters E, H, and T are the top three most used letter, since they are the letter in

'THE'. J, Q, X, and Z are the lest used letters, since there aren't many words that contain these letters. Overall, it's very onteresting how the trends are very similar across a wide variety of genres.