For my final project, I utilized several tools to perform a distant reading of a corpus of a few hundred novels by four authors of British boys’ adventure fiction. These authors included R.M. Ballantyne, G.A. Henty, W.H.G. Kingston, and H. Rider Haggard, all of whom wrote their fiction in the late nineteenth and early twentieth centuries. I had analyzed the works of these authors by performing close readings of their stories in my Junior I.S. Now, I hoped that a distant reading would tell me how their use of language could further inform readers about the role of imperialism, racism, and gender roles within their texts, and hopefully, by extension, the late-nineteenth century British imagination as a whole. I had aimed too high, however, and I was unable to make any claims about these topics without relying on overgeneralizations. When applying digital tools to an individual author, I had much more success in visualizing how language use can inform our reading of that author.

First, I cleaned up my texts after downloading them from Project Gutenberg and created my corpus. I then used the Stylo R package and Cytoscape to create a data web visualization of the corpus. The four authors clumped into four distinct groups, although there was some overlap between Kingston and Henty, and between Ballantyne and Haggard. I was drawn to these areas of overlap and I contemplated ways I could use Stylo or other digital tools to identify the reasons for the similarities between the novels. I was also curious as to whether or not identifying these similarities would add to the overall reading of these texts. In order to further analyze these overlaps, I installed Mallet, and performed topic modeling to identify common usages of words. Topic modeling of both the entire corpus and of the overlapping works of the authors provided me with little substance with which to work. I then attempted to use the Stylo “oppose” function to identify differences between the word usages in the texts. Again, the results were too general to be of any use. Stylo had shown me that similarities between the works of two authors did indeed exist, but I could find no way of doing anything additional with this information.

My problems resulted from the large size of my corpus and from the fact that I was trying to compare several individual authors. I therefore decided to use the same tools to look at the works of each author individually. I chose to focus on the works of R.M. Ballantyne, whose corpus contained 84 novels. I hoped to use text mining to demonstrate how the language and content of Ballantyne’s stories changed over the course of his life. I divided his work into two corpus, one containing his novels published between 1857 and 1874, and the other containing those published between 1875 and 1895. I then ran the Stylo oppose function to show the contrast in language use between the two sets. I found that the later set of novels contained a higher rate of usage for the word “God” and for the words, “father,” “mother,” and “daughter.” I then used Voyant to visually represent the increase in the usage of these words over time. Using both of these tools together, I made the hypothesis that Ballantyne’s fiction had more focus on religious and familial themes in the later years of his life. This hypothesis could be supported by an actual reading of the texts and could also inform one’s reading of the texts.

In the end, I am excited to have gained experience using the Stylo R package. Now that I am more familiar with how it works, I hope to use it as an aid in my future research.