Samantha Eppinger
Project 3 Reflection

I used the website Project Gutenberg to download full texts of books. The books I selected were collected using the “Random Book” search link, and selected in pairs according to gender (male and female). I used pickle to save the files, as a way to ensure that I would not be kicked off the page due to downloading too many items. I then analyzed the texts for the most frequently used words (by passing the text through functions I wrote), and by using a sentiment analyzer to gather all the information I wanted to collect from the text.

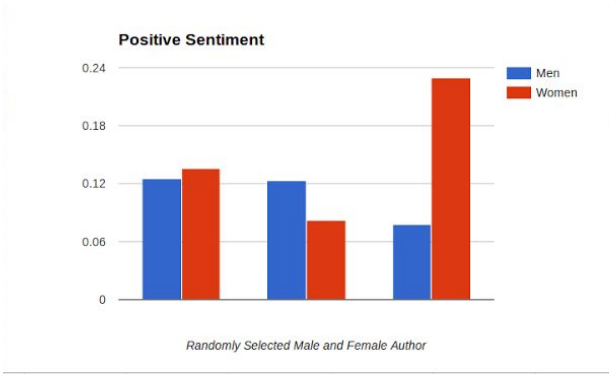
My code is broken into three main sections. The first section of my code is what I dubbed, “the pickling section”. In this section I downloaded the randomly selected books from Project Gutenberg, “pickled” it (saved a file of it), and then also opened the text. While the first two operations of this section are only necessary once, the third operation is how the rest of my code obtains access to the input. The next section of my code was where I spent most of my time. This section parses through the given text using various functions and loops to break the text into words and then ranks the most used words (excluding the most common words). This section returns a list of the top ten most commonly used words in a piece of text.

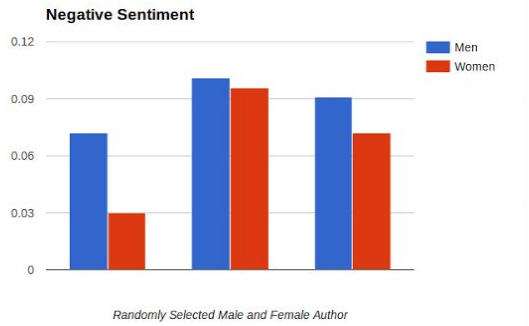
My third section was where most of the organization outside of functions for my code happened. I made a list of all randomly selected text’s authors and called the first two sections on them. I then printed out the list of the most commonly used words. I also imported a package called Sentiment Intensity Analyzer because I thought the results would make my results more interesting and accurate.

I found that 5:6 authors (with a 50% gender split) had male pronouns or nouns in their top 6 words. Approximately 60% of the collective male author’s top words were male pronouns or nouns, while approximately 15% of the collective female author’s top words were male pronouns or nouns. On the other hand 3:6 authors (with a 50% gender split) had female pronouns or nouns in their top 6 words. Approximately 8% of the collective male author’s top words were female pronouns or nouns, while approximately 15% of the collective female author’s top words were female pronouns or nouns. Below is a table of the randomly selected 6 authors and their top 6 words (in order from most used to least).

Titus	Maria	Alfred	Helen	Walter	Rebecca
own	Brown	been	tell	man	heart
man	Claus	WIlliam	Helen	door	ladyship
father	House	king's	came	no	your
then	Little	many	didn't	Anne	now
son	Santa	same	hands	Peony	could
He	again	other	silence	came	He

In addition my analysis also revealed that women ⅔ of the time have a higher amount of positive sentiment in their work. The mean of positive sentiment for women was .149 while the mean for men was .108. In addition my analysis also revealed that men always had a higher amount of negative sentiment in their work. The mean of negative sentiment for women was .066 while the mean for men was .088. Below are two graphs that visually represent my results.





I enjoyed working on this project. I thought the way I approached this project optimized my learning and reinforced/strengthened skills that I have been learning this semester. For instance, I sectioned my code into a hoard of functions, wrote doc tests, and opted to implement “loops”. Since I wasn’t comfortable with doc tests before this project, I had a clear plan for what I wanted to test when writing my code. Since I was focusing on improving these skills my code is not optimized and runs slow. My next step would be to make my code run faster. This would be easy to implement because I now have a better understanding for how my code is structured (and a collection of future recommendations from Ninjas). In addition I had to make some edits to my project idea during the final stages because I realized what I wanted my program to do was too easy, and it wouldn’t be a satisfying end to a project. Though I am happy I changed courses my code could have been better structured if I had my final idea as my original plan. I think it would have been really useful to have talked to a peer and explained my project idea. If I had gone through what I wanted my basic functions to be, and their outputs to be I think the project could have been easier because a peer could have caught one of my silly mistakes.