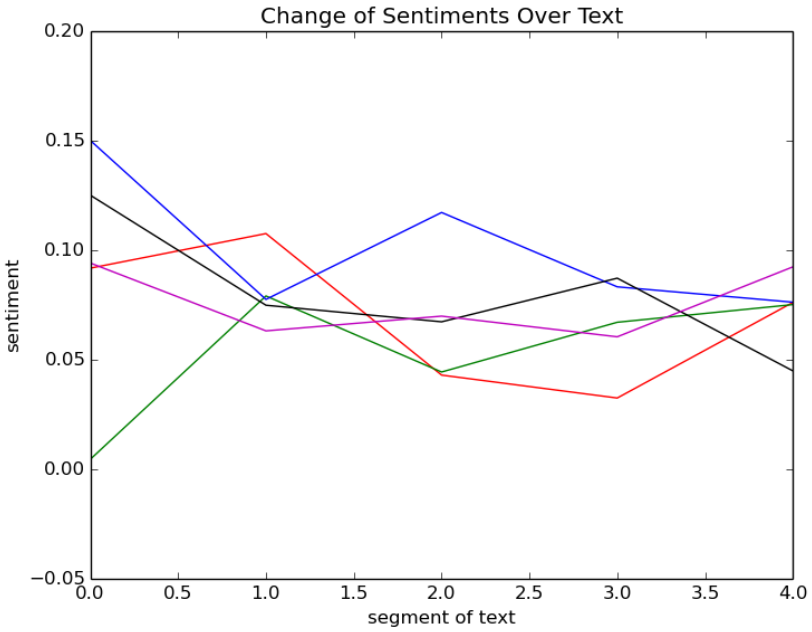
Lucy Wilcox
Mini-Project 3 Reflection

Overview:
I analyzed the sentiment of different gothic fiction books, namely *The Picture of Dorian Grey, The Yellow Wallpaper, Frankenstein, Turn of the Screw,* and *Dracula* over the course of the text. I used a Project Gutenberg mirror to get the books. I used a sentiment() from pattern to analyze sections of the text. I also used matplotlib.pyplot to plot the data. I hoped to see if most books shared a similar sentiment path of the course of the book.

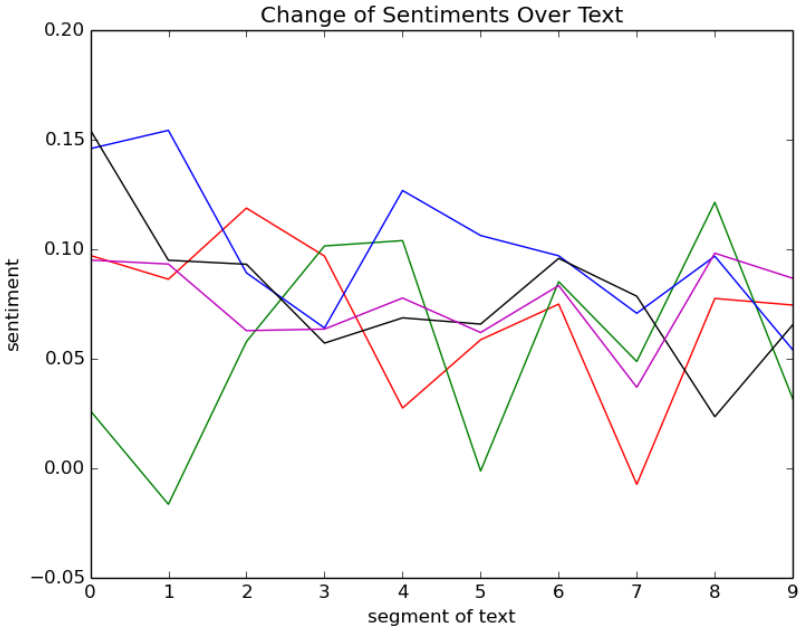
Implementation:
I have three function being called from main. One is get_books(), another is all_sentiments(books, r), and the third is plot_sentiments(sentiments, r). The heart of my code is accessed from all_sentiments which runs through all of my other functions (except for the ones called in main). All sentiments loops through each book and goes through all of my functions to eventually determine the sentiment. I decided to store most data in lists because I did not want to use a dictionary because word order is important and I don't need any of the functionalities dictionaries provide.

First each book starts as a huge string, which I split into a list of words to search for "***" which determines the distinction between the book and Project Gutenberg header. I chose this method instead of searching through lines to find the header because it also removes the footer and because it was easy to measure the length of my book by the length of the string of words. Within each book, the r segments (r can be changed in main) and their sentiments are passed as a list of strings or values, I chose to do this because then later on they would be easy to graph, and are in a clear format, as opposed to having a lot of values to organize later on. I choose to plot all my books in different colors on one plot to see if there was any correlation.

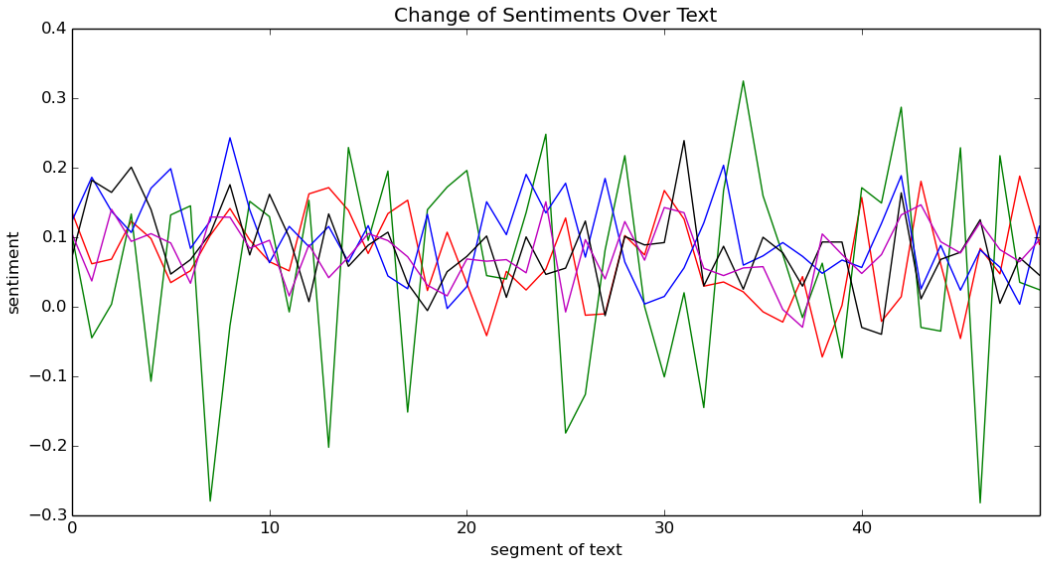
Results:
I found that (no suprise) gothic fiction is not too positive. For the following graphs:
Red = The Picture of Dorian Grey
Green = The Yellow Wallpaper
Blue = Frankenstein
Black = Turn of the Screw
Magenta = Dracula



This graph to the left shows the sentiment when the text is split into five sections. You will see that there is not an overarching trend, except that most books seem to oscillate, instead of purely going up or down.



When the books are split into 10 sections the oscillation trend still holds and becomes more pronounced. Also it appears that most (4 out of 5) books drop in sentiment at the very end of the novel.



When the novel is split into 50 sections the same chaos persists. One thing to note here is that *The Yellow Wallpaper* change in sentiment more dramatically. This is likely because it is a short story, so each block of text being analyzed is smaller and thus more polarized.

Reflection:
I started early, which was helpful and went to ninjas. I tried to layout my functions before implementing them, which I think was helpful. Sometimes I realized that I needed to add a new function or split one up though. I could have made my program more general at the beginning instead of going back and generalizing it. Learning to plot data was really helpful. I was able to print parts of the lists to make sure each function was working appropriately and made sure I did this before going on. I learned about how to structure a program when you are building it up from nothing and will be able to improve that skill going forward.