Avery Field
Milestone 3

Since last week, I first downloaded more transcripts of monologues and wrote more code to preprocess my day. I added more stop words to be eliminated and removed thinks like punctuation and newline characters. After that I wrote more code to organize the data into a token list and put in the LDA algorithm that makes the actual topic models. Once I was able to get this running for one txt file, I wrote a loop that would go through all of the txt files and organize the data as needed to create the topic model.

The problem I'm facing right now is that my topic models are not giving me very interesting information. Since the LDA creates topics by finding frequently used words, the words going into my topics do not provide a lot of information. These are words like "play, go, walk, no, etc." One of the things I have been doing was adding these words to my stop word list, but I want to look into better ways to solve this problem. Another thing I want to do this coming week is use pyLDAvis to create visual analyses of my topic models.