

Hannah Fowler  
Software Design Spring '15  
February 25  
Project Writeup: Mini Project 3

## Project Overview:

For this project, I acquired and analyzed texts from Project Gutenberg. Interested in both word frequency and computational art, I decided to make a combination of the two. One thing that inspired me was a word cloud, with the size of a word dependent on its frequency in a text. My end goal was to generate an image that implemented the idea of a word's placement/characteristics dependent on its "importance" in a corpus.

## Implementation:

My first step in creating the script was to remove the punctuation from a text. If this did not take place, then a word with a period would be mistaken as a different word than itself (e.g., 'text' != 'text.'). Next, I had to analyze the frequency of a word. Within this, I made a list of tuples of words and frequencies. I decided to do this to make each element simpler to access. I also refined my list of words to only the top 100 words, but this could easily be changed to a smaller or larger number.

My final function is where a lot of design decisions had to be made. But it is also very flexible as well. Currently, I have it generate an image where the most frequent word is the largest and positioned at the bottom most right. (The least frequent is smallest and at upper left.) Moreover, my image is in black white, which I personally think is more visually appealing. However, the font, color, placement of word, and size ratio can be personalized.

## Results

In doing text analysis, the most frequent words were common English stop words, which would have a frequency somewhere in the thousands, and more unique words, despite being frequent in a text, had a frequency in the hundreds. For a trial run, I simply used a list of three words, with “Hi” being the least frequent and “Eva” being the most frequent. The following was the result:

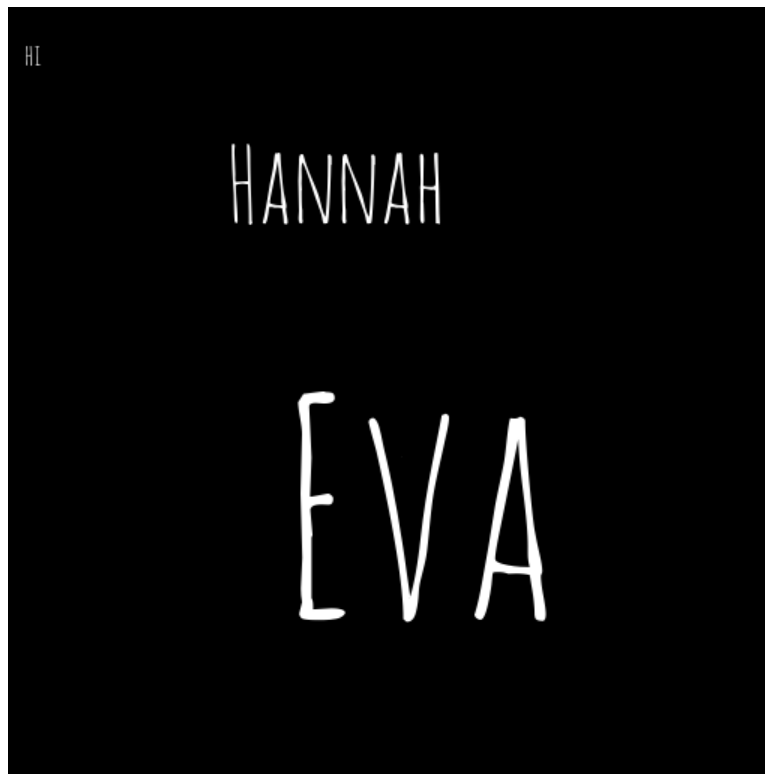


Figure 1: Image generated from trial

When using a text from a book, my code generated the following image:

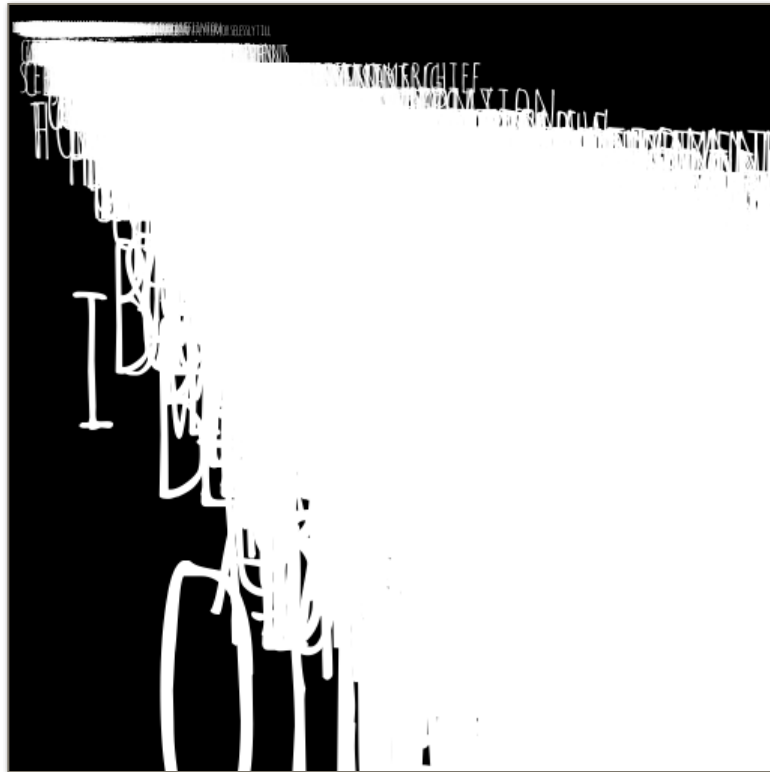


Figure 2: Image generated from *Wuthering Heights* text

## Reflection

This project forced me to learn about the PIL package and its modules. Sometimes when looping through dictionaries or lists of tuples, it was difficult to access specific elements of those types. For future iterations I would want to prevent words from overlapping, but I do not know PIL modules well enough; then I could generate my own word cloud, implementing different fonts and colors. I enjoyed the freedom of this project, however, the time crunch made exploring difficult.