Mitch Bouma
Milestone 1 Project 2

For this project, I first plan to count the number of unique characters used and the frequencies of each one. I will create an array of 1-127 (all ascii characters) and parse the file, reading one character at a time and incrementing array[that character] by one until the EOF. I will then make a linked list of all the characters we did find, and put it in ascending order based off their frequencies. It will be from lowest to highest frequency while excluding the zeros. Next, I will make each one of the nodes in this linked list into a node of a separate binary tree. It will then continue to combine the two lowest frequency trees so that the 2 nodes are replaced by one root node with children as what the two original nodes were. This new parent node will have a frequency value of its two children combined. This process will continue until only one parent node is left. Next, I will build a 2-dimensional "codebook" by going through the tree to each leaf and writing the process to get there (each 0 for left and 1 for right) as well as the letter at that leaf. The top row in the array will be the letter, and the bottom will be the sequence of 0's and 1's. Finally, it will have to save the header which is just the frequencies of all the characters. To write the compressed file, it will loop through the input file and use the codebook to write the bit sequences for each character. The file will have a .huff extension added.

For the unhuff section, I will recreate the binary tree in a similar fashion to the above section. It will then go through the given bit sequences and follow the tree until it gets to a letter. Once it does, it will print that letter and start back at the top of the tree. It will do this until it has gone through the entire bit sequence. It will be printing to a normal file with a .unhuff extension added on.