Richard von der Schmidt
NETID: rev5

Note:

I stored the raw data as image files since you cannot copy and paste the data from the GUI.

- Question 1: Hypothesis about how Compression Rate and Total Time depend on File Length and Alphabet size.
    - The best case for understanding how the alphabet size affects the compression rate is comparing the image file with a given text file.
    - **Characters in Image File**: 257
    - **Characters in Text File**: 67
    - **Image Compression Rate**: 6 %
    - **Text File Compression Rate**: % 43
    - *Explanation:*
        - The amount of space that you can save is directly related to the size of the alphabet. If we only need to express 67 characters, we don't need the extra space that the full character set provides. If we need to express over 257 characters, we have to compress a greater number of characters which means the rate of compression will go down.
        - This is an inverse relationship
        - As alphabet size increases, the compression rate decreases
        - As alphabet size decreases, the compression rate increases.
        - Yx = k
        - (Inverse proportionality)
    - The best case for understanding file size and total time to run the compression can be seen comparing the two text files.
    - **Size of Bigger File:** KJV, 4 million bytes
    - **Size of Smaller File:** Melville, 100K bytes
    - **Run Time for Bigger File:** 0.303 secs
    - **Run Time for Smaller File:** 0.009 secs
    - *Explanation:*
        - As the file size increases, the amount of time taken to compress increases.
        - As the file size decreases, the amount of time taken to compress decreases.
        - This is a directly proportional relationship
        - Y=kx
    - The size of the file will not affect the rate of compression, because in this instance the rate of compression refers to how much space is saved and not the amount of time it takes to compress the file. As

such, the only thing that affects the compression rate is the alphabet size.
    - A small alphabet does not guarantee a smaller file and as a result does not significantly affect the compression time.
- Question 2:
    - The text files compress more.
    - The binary files have a greater range of unique characters to express in order to render color (257 for .tif files)
    - The text files, need less than half of that (67 for Melville)
    - **The rate of compression for text:** % 34.50
    - **The rate of compression for binary:** % 20.9
- Question 3:
    - Very little extra compression is accomplished when compressing an already compressed file.
    - Huffman coding is not effective after the first compression because the compression rate is directly tied to the number of unique characters in a given file. After a file has been compressed, the resulting version may be able to find a few more unique characters to remove in the compressed file, but there likely will not be many since you have already reduced the entropy of the file itself.
    - **Space saved when doing compression**: % 0.07
- Question 4:
    - See code

- Cheers Efe! You're a bright young man and I am very grateful I have got to know you this year. All the best to you!