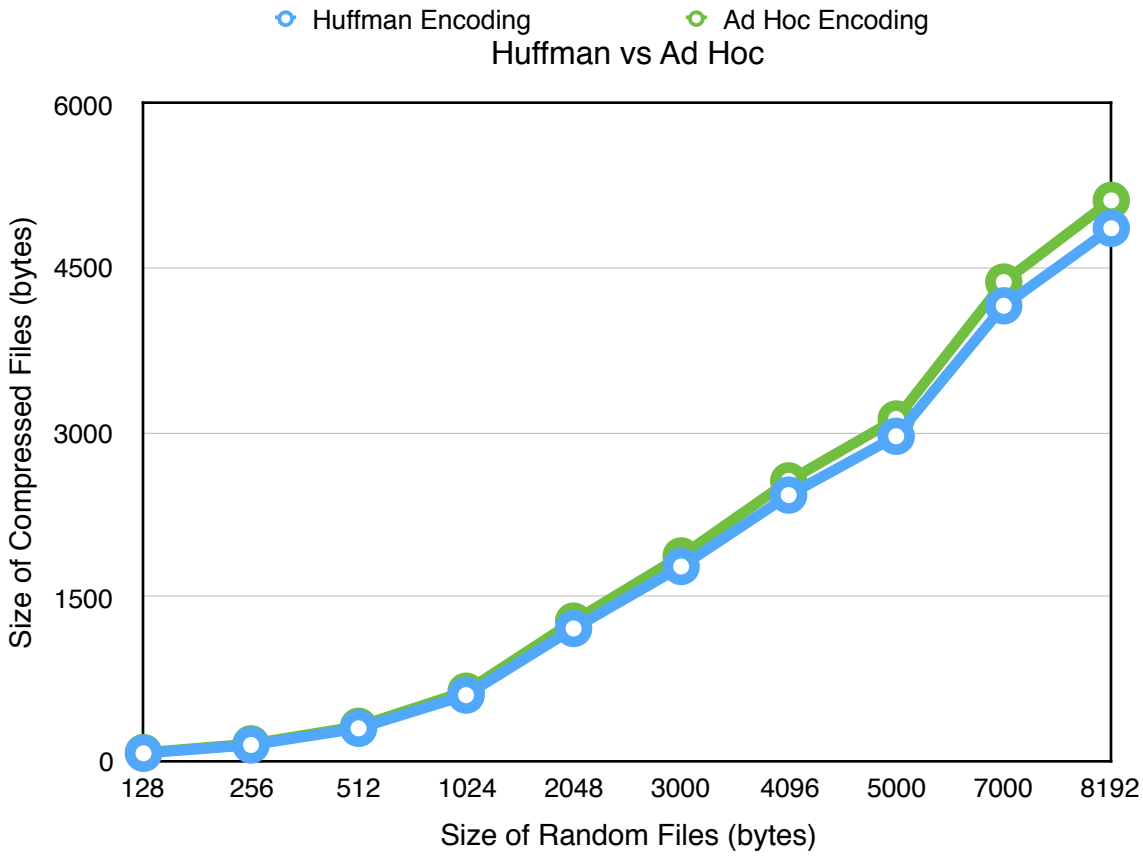
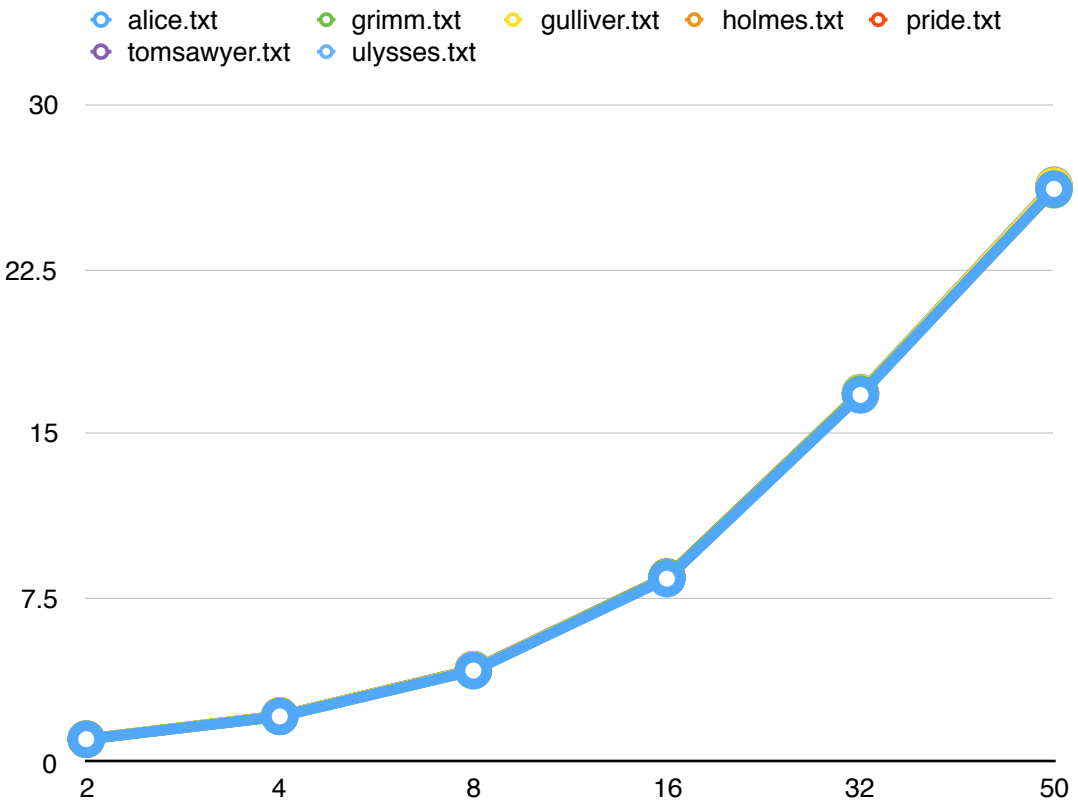
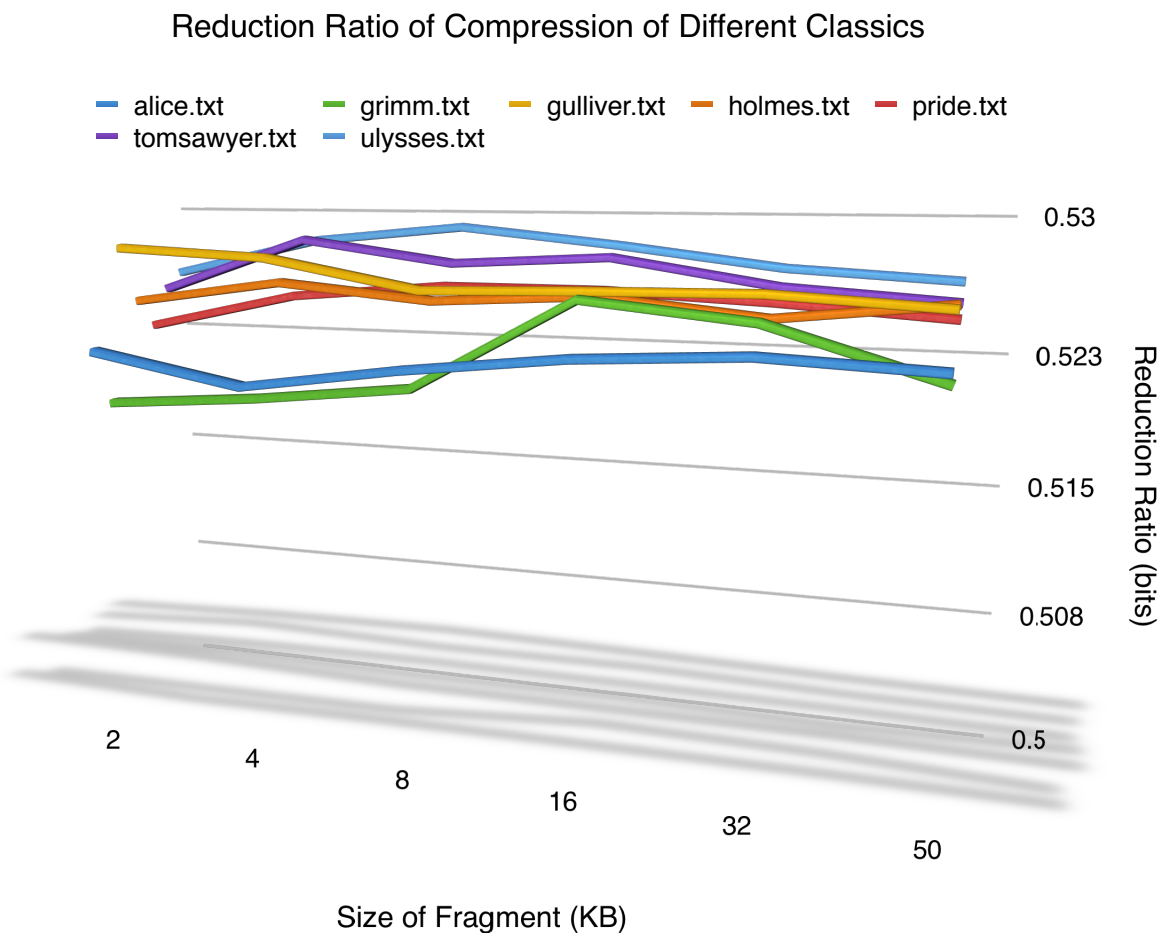
Size of Random Files (bytes)	Average Huffman Encoding Size (bytes); Average Savings Ratio (bits)	Ad Hoc Encoding Size (bytes); Savings Ratio (bits)
128	73.41; 0.5375	80; 0.6250
256	149.12; 0.5825	160; 0.6250
512	301.06; 0.5880	320; 0.6250
1024	604.36; 0.5902	640; 0.6250
2048	1212.55; 0.5921	1280; 0.6250
3000	1778.09; 0.5927	1875; 0.6250
4096	2430.10; 0.5933	2560; 0.6250
5000	2967.11; 0.5934	3125; 0.6250
7000	4157.89; 0.5940	4375; 0.6250
8192	4865.43; 0.5939	5120; 0.6250



Both the Huffman encoding and the ad hoc encoding compress the randomized files significantly. However, the Huffman encoding compresses the files in a more effective manner, on average. With the average savings ratio for the Huffman encoding for most of the tested sizes being around 0.59, the Huffman encoding improves the savings ratio for the ad hoc encoding by about .035, on average. Because the text files were generated randomly, it is likely that the occurrences of the letters were decently evenly distributed. Thus, Huffman encoding is likely not as advantageous for random text files, as Huffman encoding takes advantage of specific letters occurring more frequently than others. However, we can see here that Huffman encoding still does perform better than ad hoc encoding because of some of the possible lack of variation, or repetition, in the random text files.

Study 2:





	Reduction Ratio for size 2 KB fragment	Reduction Ratio for size 4 KB fragment	Reduction Ratio for size 8 KB fragment	Reduction Ratio for size 16 KB fragment	Reduction Ratio for size 32 KB fragment	Reduction Ratio for size 50 KB fragment
alice.txt	0.5223	0.5207	0.522	0.523	0.5235	0.5231
grimm.txt	0.5191	0.5198	0.5208	0.5259	0.525	0.5223
gulliver.txt	0.528	0.5276	0.526	0.5262	0.5263	0.5258

	Reduction Ratio for size 2 KB fragment	Reduction Ratio for size 4 KB fragment	Reduction Ratio for size 8 KB fragment	Reduction Ratio for size 16 KB fragment	Reduction Ratio for size 32 KB fragment	Reduction Ratio for size 50 KB fragment
holmes.txt	0.5248	0.5261	0.5253	0.5258	0.5249	0.5259
pride.txt	0.5232	0.5252	0.526	0.526	0.5256	0.525
tomsawyer.txt	0.5253	0.5284	0.5272	0.5277	0.5263	0.5257
ulysses.txt	0.5262	0.5283	0.5292	0.5283	0.5272	0.5267

For Study 2, the reduction ratios for each of the texts were pretty similar. Of the files that were compressed, grimm.txt and alice.txt were the files that were the best encoded by Huffman compression, and ulysses.txt and tomsawyer.txt were the files that were the worst encoded by Huffman compression. Thus, it seems that Lewis Carrol and Brothers Grimm produce prose that is better encoded using Huffman codes. This likely means that these authors use diction that repeats certain letters more often than the other authors. This can be deemed a measure of the predictability of the authors' writing, or how consistent the authors' prose is. These authors likely use words or phrases that lead to a greater repetition in certain characters, making the texts more easily compressible.