Project 1 report

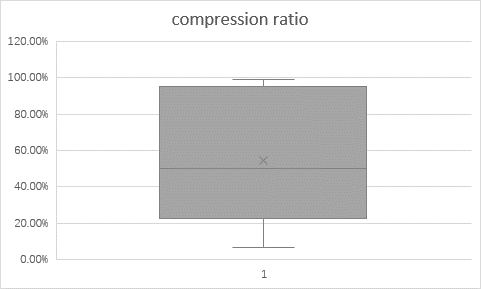
1. I use the sample input to do the experiment, it range from 1kb to 743kb and the line range from 3 line to around 10000 line

2. the compression ratio:

The avg compression ratio is: 12%

The min compression ratio is: 0%

The max compression ratio is: 99%



1. Encode run time (all done with clock function)

Insertion sort

Merge Sort

1. Decode run time (all done with clock function)

Insertion

Merge

1. If we encode 2,3…lines at a time, the compression ratio will improve because there will be more repeat chars at the same encode string, so if we encode 10000 line at a time, the max cluster char we will have is126 which is all the ascii chars we have. so, the compression ratio will increase a lot.
2. General conclusion
   1. The compression ratio is kind low right now snice we encode one line at time so there wont be much repeat chars
   2. If the file size getting bigger it will time more time to encode the file, when file size is small insertion sort runs faster than merge sort, but when file size getting bigger then merge sort should run faster than insertion sort.
   3. When more line encodes at same the compression ratio will goes up because there are more repeat chars;