BCB 5300
Homework 2

1. Sorting by reversals: Can you find an example where the Breakpoint Reversal Sort produces more than 3 times as many reversals as the optimal solution?

2. DNA molecules are not always line segments - simple organisms have circular DNA molecules as a genome. (Visualize a sequences of values written along the perimeter of a circle.). Two such sequences are considered equivalent if you can rotate one of the circles and get the same sequence as written on the other. Devise an algorithm to sort a circular genome by reversals - in other words, given a circular set of input numbers, find a way to transform it to the identity circular permutation. What is your solution's approximation guarantee?

3. Consider two strings v and w, of length n and m respectively, and their longest common subsequence $LCS(v, w)$ and their edit distance $Edit(v, w)$, but where only insertions and deletions (and no substitutes) are allowed. Prove that $Edit(v, w) = n + m - 2LCS(v, w)$. (Hence, edit distance and LCS are connected!)

4. A shuffle of two strings X and Y is formed by interspersing the characters into a new string, keeping the characters of X and Y in the same order.

   For example, the string BANANAANANAS is a shuffle of the strings BANANA and ANANAS in several different ways. Similarly, the strings PRODGYRNAMAMMIINCG and DYPRONGARMAMMICING are both shuffles of DYNAMIC and PROGRAMMING.

   Given three strings $A[1..n]$, $B[1..n]$, and $C[1..m + n]$, describe and analyze an algorithm to determine whether C is a shuffle of A and B.