Counting Inversions
Instructor: Anna Karlin

Outline of proof that algorithm for counting inversions is correct:

• Prove by induction on $k$, that if the algorithm correctly sorts and counts inversions in arrays of length $2^k$, then it correctly sorts and counts inversions in arrays of length $2^{k+1}$.

• Base case: $k = 0$. An array of length 1 is automatically sorted and has 0 inversions.

• Given an array of length $2^{k+1}$, let $L$ represent the left half and $R$ represent the right half. Each of these is an array of length $2^k$. By the inductive hypothesis, we can assume that applying our algorithm to these two half arrays yields two sorted subarrays (we will still call them $L$ and $R$), and correctly computes the inversions internal to each half. Thus, we only need to show that the merge step correctly sorts them (which I will skip – this is merge sort), and that it correctly counts the number of inversions between $L$ and $R$.

• For the latter, we make the following key claim:
  Consider the point at which the $t$ smallest elements from $L$ and $R$ have been added to the combined array $A$. Call these elements $S_t$. Our inductive claim is that during the Merge-And-Count step, we have already counted all $L - R$ inversions that involve at least one element from $S_t$.

Clearly this is true for $t = 0$. Suppose it is true for some larger $t$. When the $(t + 1)^{st}$ element, say $x$, is added to $A$, then all $L - R$ inversions that involve it and some element of $S_t$ have already been counted by the inductive hypothesis. Let $L_t$ be the left over elements of $L$ (not yet added to $A$) and let $R_t$ be the leftover elements of $R$. Thus, we only need to worry about inversions between $x$ and other elements of $L_t \cup R_t$. If $x \in L_t$, then no new inversions involving $x$ are created, since it is smaller than all remaining elements of $R_t$, and was to the left of them in the array prior to the merge step. If $x \in R_t$, then it is inverted relative to all elements remaining in $L_t$. But in this case, we add $|L_t|$ to our left-right inversion count. Thus, all inversions including $x$ are included in the total count, and the claim holds for $S_{t+1}$. 