1. Say A and B are the two databases and $A(i)$, $B(i)$ are i^{th} smallest elements of A, B.

First, let us compare the medians of the two databases. Let k be $\lceil \frac{1}{2}n \rceil$, then $A(k)$ and $B(k)$ are the medians of the two databases. Suppose $A(k) < B(k)$ (the case when $A(k) > B(k)$ would be the same with interchange of the role of A and B). Then one can see that $B(k)$ is greater than the first k elements of A. Also $B(k)$ is always greater than the first $k-1$ elements of B. Therefore $B(k)$ is at least $2k^{th}$ element in the combine databases. Since $2k \geq n$, all elements that are greater than $B(k)$ are greater than the median and we can eliminate the second part of the B database. Let B' be the half of B (i.e., the first k elements of B).

Similarly, the first $\lfloor \frac{1}{2}n \rfloor$ elements of A are less than $B(k)$, and thus, are less than the last $n-k+1$ elements of B. Also they are less than the last $\lceil \frac{1}{2}n \rceil$ elements of A. So, they are less than at least $n-k+1+\lceil \frac{1}{2}n \rceil = n+1$ elements of the combine database. It means that they are less than the median and we can eliminate them as well. Let A' be the remaining parts of A (i.e., the $\left[\lfloor \frac{1}{2}n \rfloor + 1; n\right]$ segment of A).

Now we eliminate $\lfloor \frac{1}{2}n \rfloor$ elements that are less than the median, and the same number of elements that are greater than median. It is clear that the median of the remaining elements is the same as the median of the original set of elements. We can find a median in the remaining set using recursion for A' and B'. Note that we can't delete elements from the databases. However, we can access i^{th} smallest elements of A' and B': the i^{th} smallest elements of A' is $i + \lfloor \frac{1}{2}n \rfloor^{th}$ smallest elements of A, and the i^{th} smallest elements of B' is i^{th} smallest elements of B.

Formally, the algorithm is the following. We write recursive function `median(n,a,b)` that takes integers n, a and b and find the median of the union of the two segments $A[a+1; a+n]$ and $B[b+1; b+n]$.

```
median(n, a, b)
   if n=1 then return min(A(a+k), B(b+k)) // base case
   k=⌈n/2⌉
   if A(a+k)<B(b+k)
     then return median (k, a+⌊n/2⌋ ,b)
     else return median (k, a, b+⌊n/2⌋)
```

To find median in the whole set of elements we evaluate `median(n,0,0)`.

Let $Q(n)$ be the number of queries asked by our algorithm to evaluate `median(n,a,b)`. Then it is clear that $Q(n) = Q(\lceil \frac{1}{2}n \rceil) + 2$. Therefore $Q(n) = 2\lceil \log n \rceil$.

A final note. In order to prove this algorithm correct, note that it is not enough to prove simply that, in the recursive call, the median remains in the set of numbers considered; one must prove the stronger statement that the median value in the recursive call will in fact be the same as the median value in the original call. Also, the algorithm cannot invoke the recursive call by simply saying, "Delete half of each database." The only way in which the algorithm can interact with the database is to pass queries to it; and so a conceptual "deletion" must in fact be implemented by keeping track of a particular interval under consideration in each database.

2. We'll define a recursive divide-and-conquer algorithm ALG which takes a sequence of distinct numbers $a_1, \ldots, a_n$ and returns N and $a'_1, \ldots, a'_n$ where

- N is the number of significant inversions
- $a'_1, \ldots, a'_n$ is same sequence sorted in the increasing order

ALG is similar to the algorithm from the chapter that computes the number of inversions. The difference is that in the 'conquer' step we merge twice: first we merge $b_1, \ldots, b_k$ with $b_{k+1}, \ldots, b_n$ just for sorting, and then we merge $b_1, \ldots, b_k$ with $2b_{k+1}, \ldots, 2b_n$ for counting significant inversions.

Let's define ALG formally. For $n = 1$ ALG just returns $N = 0$ and $\{a_1\}$ for the sequence. For $n > 1$ ALG does the following:

- let $k = \lfloor n/2 \rfloor$.
- call $\mathsf{ALG}(a'_1, \ldots, a'_k)$. Say it returns N_1 and $b_1, \ldots, b_k$.
- call $\mathsf{ALG}(a'_{k+1}, \ldots, a'_n)$. Say it returns N_2 and $b_{k+1}, \ldots, b_n$.
- compute the number N_3 of significant inversions (a_i, a_j) where $i \leq k < j$.
- return $N = N_1 + N_2 + N_3$ and $a'_1, \ldots, a'_n = \mathsf{MERGE}(b_1, \ldots, b_k;\ b_{k+1}, \ldots, b_n)$

MERGE can be implemented in $O(n)$ time. According to the discussion in the book, it remains to find a way to compute N_3 in $O(n)$ time. We implement a variant of merge-count of $b_1, \ldots, b_k$ and $2\,b_{k+1}, \ldots, 2\,b_n$ as follows.

- Initialize counters: $i \leftarrow k$, $j \leftarrow n$, $N_3 \leftarrow 0$.
- If $b_i \leq 2b_j$ then
  - if $j > k + 1$ decrease j by 1.
  - if $j = k + 1$ return N_3.
- If $b_i > 2b_j$ then increase N_3 by $j - k$. Then
  - if $i > 1$ decrease i by 1.
  - if $i = 1$ return N_3.

Explanation For every i we count the number of significant inversions between b_i and all b_j's. If $b_i \leq 2b_j$ then there are no significant inversions between b_i and any b_m s.t. $m \geq j$, so we decrease j. If $b_i > 2b_j$ then $b_i > 2b_m$ for all m s.t. $k < m \leq j$. In other words, we have detected $j - k$ significant inversions involving b_i. So we increase N_3 by $j - k$. Finally, when we are down to $i = 1$ and have counted significant inversions involving b_1, there are no more significant inversions to be detected.

3. We give two solutions for this problem. The first solution is a divide and conquer algorithm, which is easier to think of. The second solution is a clever linear time algorithm.

Via divide and conquer: Let $e_1, \ldots, e_n$ denote the equivalence classes of the cards: cards i and j are equivalent if $e_i = e_j$. What we are looking for is a value x so that more than $n/2$ of the indices have $e_i = x$.

Divide the set of cards into two roughly equal piles: a set of $\lfloor n/2 \rfloor$ cards and a second set for the remaining $\lceil n/2 \rceil$ cards. We will recursively run the algorithm on the two sides, and will assume that if the algorithm finds an equivalence class containing more than half of the cards, then it returns a sample card in the equivalence class.

Note that if there are more than $n/2$ cards that are equivalent in the whole set, say have equivalence class x, than at least one of the two sides will have more than half the cards also equivalent to x. So at least one of the two recursive calls will return a card that has equivalence class x.

The reverse of this statement is not true: there can be a majority of equivalent cards in one side, without that equivalence class having more than $n/2$ cards overall (as it was only a majority on one side). So if a majority card is returned on either side we must test this card against all other cards.

```
If |S| = 1 return the one card
If |S| = 2
    test if the two cards are equivalent
    return either card if they are equivalent
Let S_1 be the set of the first ⌊n/2⌋ cards
Let S_2 be the set of the remaining cards
Call the algorithm recursively for S_1.
If a card is returned
    then test this against all other cards
If no card with majority equivalence has yet been found
    then call the algorithm recursively for S_2.
    If a card is returned
        then test this against all other cards
Return a card from the majority equivalence class if one is found
```

The correctness of the algorithm follows from the observation above: that if there is a majority equivalence class, than this must be a majority equivalence class for at least one of the two sides.

To analyze the running time, let $T(n)$ denote the maximum number of tests the algorithm does for any set of n cards. The algorithm has two recursive calls, and does at most $2n$ tests outside of the recursive calls. So we get the following recurrence (assuming n is divisible by 2):

$$T(n) \leq 2T(n/2) + 2n.$$

As we have seen in the chapter, this recurrence implies that $T(n) = O(n \log n)$.

In linear time: Pair up all cards, and test all pairs for equivalence. If n was odd, one card is unmatched. For each pair that is not equivalent, discard both cards. For pairs that are equivalent, keep one of the two. Keep also the unmatched card, if n is odd. We can call this subroutine ELIMINATE.

The observation that leads to the linear time algorithm is as follows. If there is an equivalence class with more then $n/2$ cards, than the same equivalence class must also have more than half of the cards after calling ELIMINATE. This is true, as when we discard both cards in a pair, then at most one of them can be from the majority equivalence class. One call to ELIMINATE on a set of n cards takes $n/2$ tests, and as a result, we have only $\leq \lceil n/2 \rceil$ cards left. When we are down to a single card, then its equivalence is the only candidate for having a majority. We test this card against all others to check if its equivalence class has more than $n/2$ elements.

This method takes $n/2 + n/4 + \ldots$ tests for all the eliminates, plus $n - 1$ tests for the final counting, for a total of less than $2n$ tests.

6. For simplicity, we will say u *is smaller than* v, or $u \prec v$, if $x_u < x_v$. We will extend this to sets: if S is a set of nodes, we say $u \prec S$ if u has a smaller value than any node in S.

The algorithm is the following. We begin at the root r of the tree, and see if r is smaller than its two children. If so, the root is a local minimum. Otherwise, we move to any smaller child and iterate.

The algorithm terminates when either (1) we reach a node v that is smaller than both its children, or (2) we reach a leaf w. In the former case, we return v; in the latter case, we return w.

The algorithm performs $O(d) = O(\log n)$ probes of the tree; we must now argue that the returned value is a local minimum. If the root r is returned, then it is a local minimum as explained above. If we terminate in case (1), v is a local minimum because v is smaller than its parent (since it was chosen in the previous iteration) and its two children (since we terminated). If we terminate in case (2), w is a local minimum because w is smaller than its parent (again since it was chosen in the previous iteration).

7. Let B denote the set of nodes on the *border* of the grid G — i.e. the outermost rows and columns. Say that G has *Property* $(*)$ if it contains a node $v \notin B$ that is adjacent to a node in B and satisfies $v \prec B$. Note that in a grid G with Property $(*)$, the *global minimum* does not occur on the border B (since the global minimum is no larger than v, which is smaller than B) — hence G has at least one local minimum that does not occur on the border. We call such a local minimum an *internal local minimum*

We now describe a recursive algorithm that takes a grid satisfying Property $(*)$ and returns an internal local minimum, using $O(n)$ probes. At the end, we will describe how this can be easily converted into a solution for the overall problem.

Thus, let G satisfy Property $(*)$, and let $v \notin B$ be adjacent to a node in B and smaller than all nodes in B. Let C denote the union of the nodes in the middle row and middle column of G, not counting the nodes on the border. Let $S = B \cup C$; deleting S from G divides up G into four sub-grids. Finally, let T be all nodes adjacent to S.

Using $O(n)$ probes, we find the node $u \in S \cup T$ of minimum value. We know that $u \notin B$, since $v \in S \cup T$ and $v \prec B$. Thus, we have two cases. If $u \in C$, then u is an internal local minimum, since all of the neighbors of u are in $S \cup T$, and u is smaller than all of them. Otherwise, $u \in T$. Let G' be the sub-grid containing u, together with the portions of S that border it. Now, G' satisfies Property $(*)$, since u is adjacent to the border of G' and is smaller than all nodes on the border of G'. Thus, G' has an internal local minimum, which is also an internal local minimum of G. We call our algorithm recursively on G' to find such an internal local minimum.

If $T(n)$ denotes the number of probes needed by the algorithm to find an internal local minimum in an $n \times n$ grid, we have the recurrence $T(n) = O(n) + T(n/2)$, which solves to $T(n) = O(n)$.

Finally, we convert this into an algorithm to find a local minimum (not necessarily internal) of a grid G. Using $O(n)$ probes, we find the node v on the border B of minimum value. If v is a corner node, it is a local minimum and we're done. Otherwise, v has a unique neighbor u not on B. If $v \prec u$, then v is a local minimum and again we're done. Otherwise, G satisfies Property $(*)$ (since u is smaller than every node on B), and we call the above algorithm.