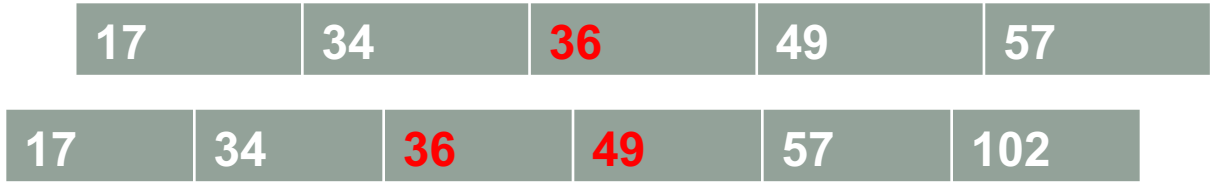
MEDIANS AND ORDER STATISTICS

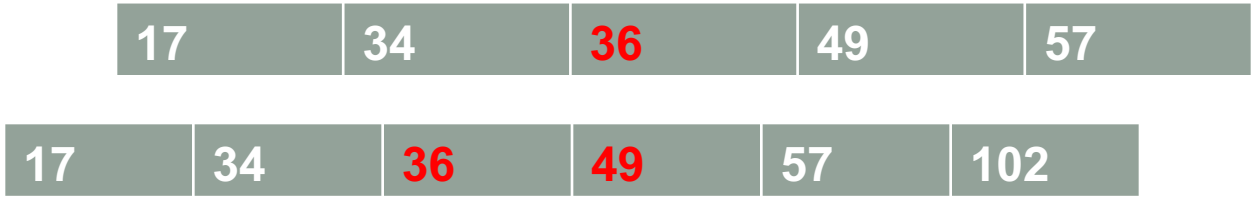
CS340

What is a median?



- The **median** of a set of numbers is the middle value, such that half the numbers are less than the median and half are greater.
- The median of an odd-length sorted array is at index (n+1)/2.
- An even-length sorted array has 2 medians, at n/2 and n/2 + 1
- How can we find the median of a sorted array?

What is an order statistic?



- In a set of n distinct numbers, (i-1) of the numbers are less than the ith order statistic.
- The **selection problem** is the problem of selecting an order statistic.
- For example, the 2nd order statistic of the above arrays is 34. (i-1=1 numbers are less than 34).

Minimum and Maximum

17	34	36	49	57	102

```
MINIMUM(A)
1  min = A[1]
2  for i = 2 to A.length
3       if min > A[i]
4            min = A[i]
5  return min
```

What is the time complexity?
Is this the best we can do?

Simultaneous Min and Max

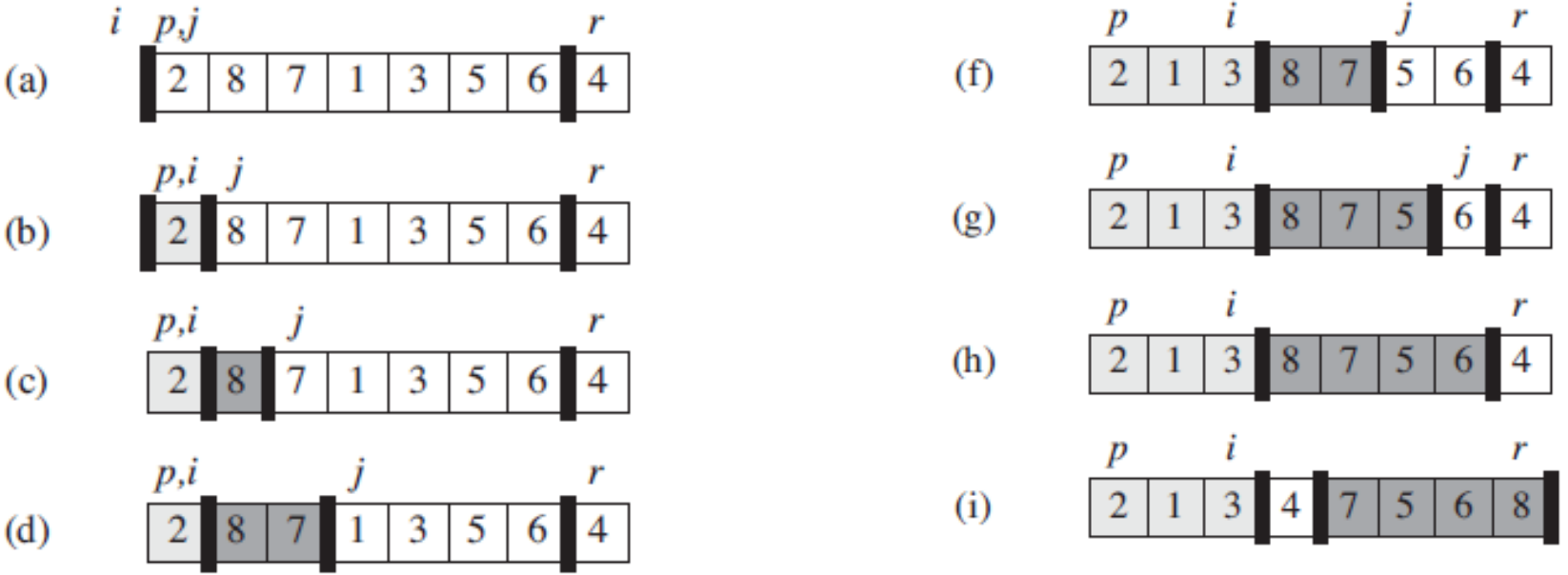
```
MINIMUM(A)
1  min = A[1]
2  for i = 2 to A.length
3       if min > A[i]
4            min = A[i]
5  return min
```

17	34	36	49	57	102

- Can find the minimum using n-1 comparisons, and the max using n-1 comparisons = 2n-2 comparisons.
- Can do better: pick elements in pairs and compare greater item to max and smaller item to min. 3 comparisons for every 2 items = 3n/2-2 comparisons.

Selecting in O(n) time

- Remember Quicksort Parition?



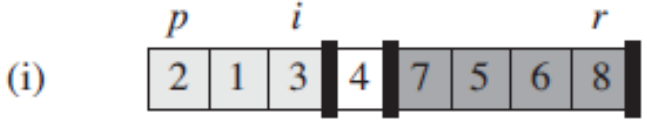
- In this example, Partition found the 4th smallest value

Selecting in O(n) time

- Here we found the 4th smallest value
- What if we are looking for the 2nd smallest value?
- We can throw away half the data, and partition on the range p..i

What is the time complexity of partitioning?

Selecting in O(n) time




```
RANDOMIZED-SELECT(A, p, r, i)
1  if p == r
2       return A[p]
3  q = RANDOMIZED-PARTITION(A, p, r)
4  k = q − p + 1
5  if i == k            // the pivot value is the answer
6       return A[q]
7  elseif i < k
8       return RANDOMIZED-SELECT(A, p, q − 1, i)
9  else return RANDOMIZED-SELECT(A, q + 1, r, i − k)
```

What is the expected time complexity?