Linear Selection

Select(S,k)

Choose a "good pivot" $x \varepsilon S$

Partition S into

$S_1 = \{y \ \varepsilon S \mid y < x\}$

$S_2 = \{y \ \varepsilon S \mid y > x\}$

If $\|S_1\| > k$ then return Select(S_1,k)

Else if $\|S_1\| = k$ then return x

Else return Select(S_2, $\|S\|-\|S_1\|-1$)

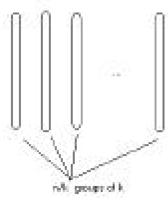
What is a good pivot?

We say $x \varepsilon S$ is a good pivot if its rank is between n/c and (c-1)n/c for some constant c>1.

If we always choose a good pivot we get $\Theta(n)$ running time.

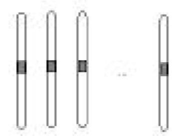
Find a good pivot

- Divide the input into groups of *k*.



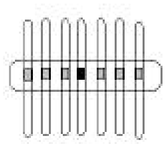
Find a good pivot

- Divide the input into groups of *k*.
- Sort each group and mark its median.



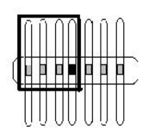
Find a good pivot

- Divide the input into groups of *k*.
- Sort each group and mark its median.
- Sort the medians and mark the median of the medians.

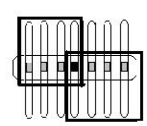


Find a good pivot

- Elements in upper left quadrant are smaller than median of medians.

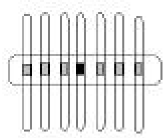


- Elements in lower right quadrant are larger than median of medians.



Median of medians

- Median of medians is a good pivot provided k satisfies the following:



CS140 11-7

BUT

- Finding the good pivot requires a recursive call to Select
- We hadn't counted on this …

CS140 11-8

New Analysis

1. Divide the input into groups of 5. Find the median of each group.
2. Find the median of the medians.
3. Partition the input around the median of medians.
4. Recurse on appropriate set of the partition.

1. O(1) time per group, O(n) groups ➔ O(n)
2. T(n/5)
3. O(n)
4. T(3n/4)

CS140 11-9

Linear selection

$T(n) = T(n/5) + T(3n/4) + O(n) = \Theta(n)$

BUT BE CAREFUL OF DETAILS!

CS140 11-10