


# COMP 482 / ELEC 420

## Order Statistics

# Your To-Do List

- Read [CLRS] 9.
- Assignment 3.

# What are Order Statistics?

Selecting $i^{th}$-ranked item from a collection.

- First: $i = 1$
- Last: $i = n$
- Median(s): $i = \left\lfloor \frac{n}{2} \right\rfloor, \left\lceil \frac{n}{2} \right\rceil$

# Order Statistics Overview

Assume collection is unordered, otherwise trivial.

Can sort first – $O(n \lg n)$, but can do better – $\Theta(n)$.

# Order Statistics Overview

Algorithms for i=1 & i=n are easy.

What are they?

?

?

Scan data, keeping track of smallest/largest element seen so far.

# Order Statistics Overview

How can we modify Quicksort to obtain expected-case $\Theta(n)$?

?

?

Pivot, partition, but recur only on one set of data. No join.

# Order Statistics

We'll use this idea.

But, by guaranteeing a good split, can get worst-case $\Theta(n)$.

Warning: Non-obvious & unintuitive algorithm ahead!

Blum, Floyd, Pratt, Rivest, Tarjan (1973)

# Order Statistics: Algorithm

Select(A,n,i):

Divide input into $\lceil n/5 \rceil$ groups of size 5.

/* Partition on median-of-medians */
medians = array of each group's median.
pivot = Select(medians, $\lceil n/5 \rceil$ , $\lceil n/10 \rceil$)
L,G = partition(A, pivot)

/* Find $i^{th}$ element in L, pivot, or G */
k = # of lesser elements + 1
If i=k, return pivot
If i<k, return Select(L, k-1, i)
If i>k, return Select(G, n-k, i-k)

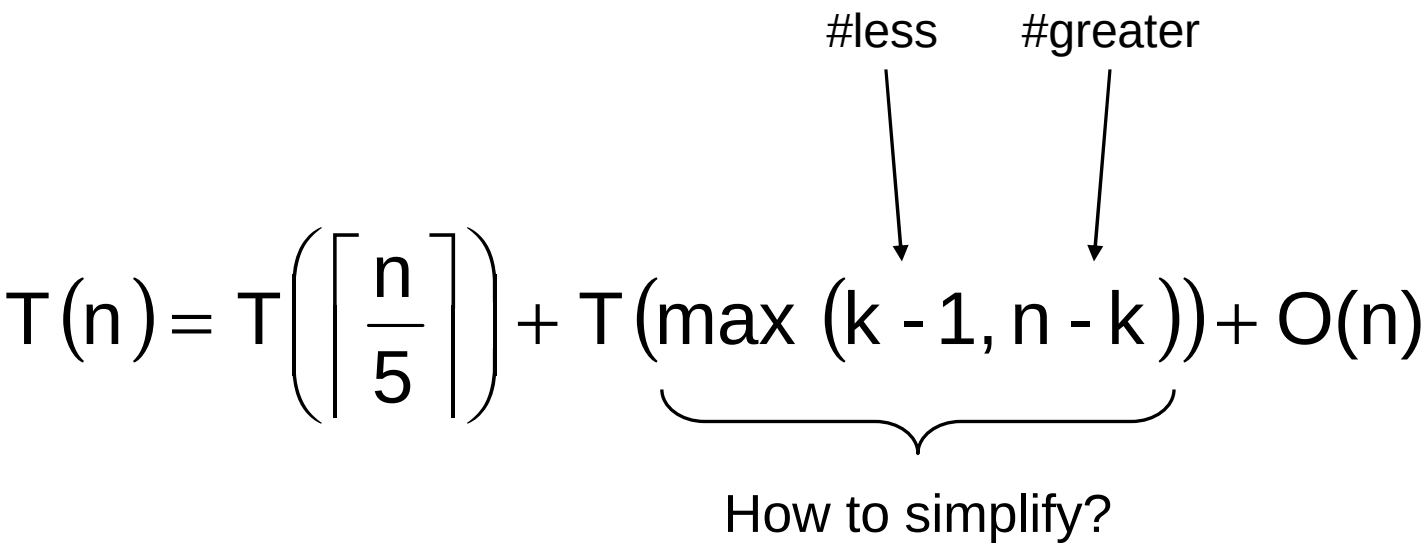
T(n)

O(n)
O(n)
T($\lceil n/5 \rceil$)
} All this to find a good split.

O(n)

O(1)
O(1)
T(k-1)
T(n-k)
} Only one done.

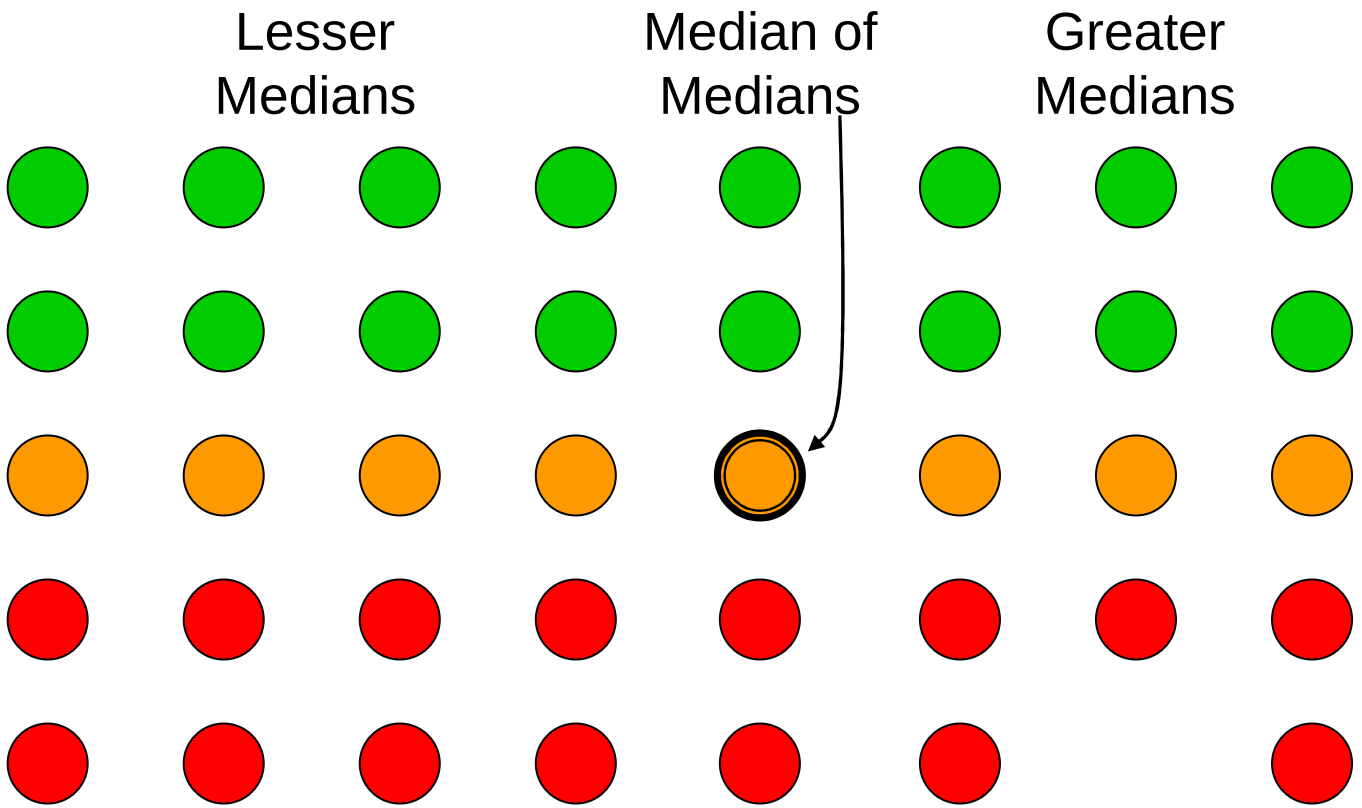
# Order Statistics: Analysis



# Order Statistics: Analysis


One group of 5 elements.

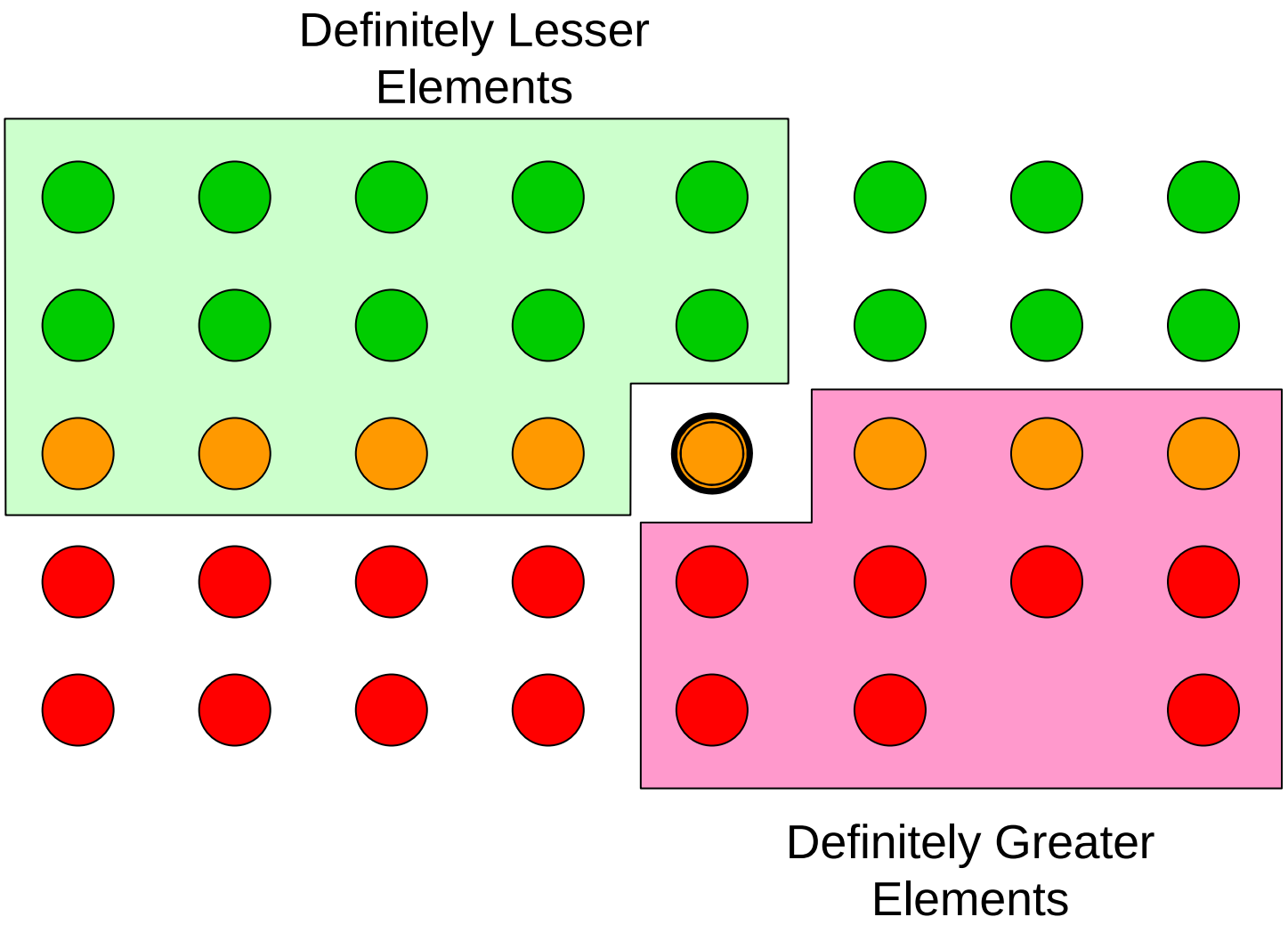
# Order Statistics: Analysis



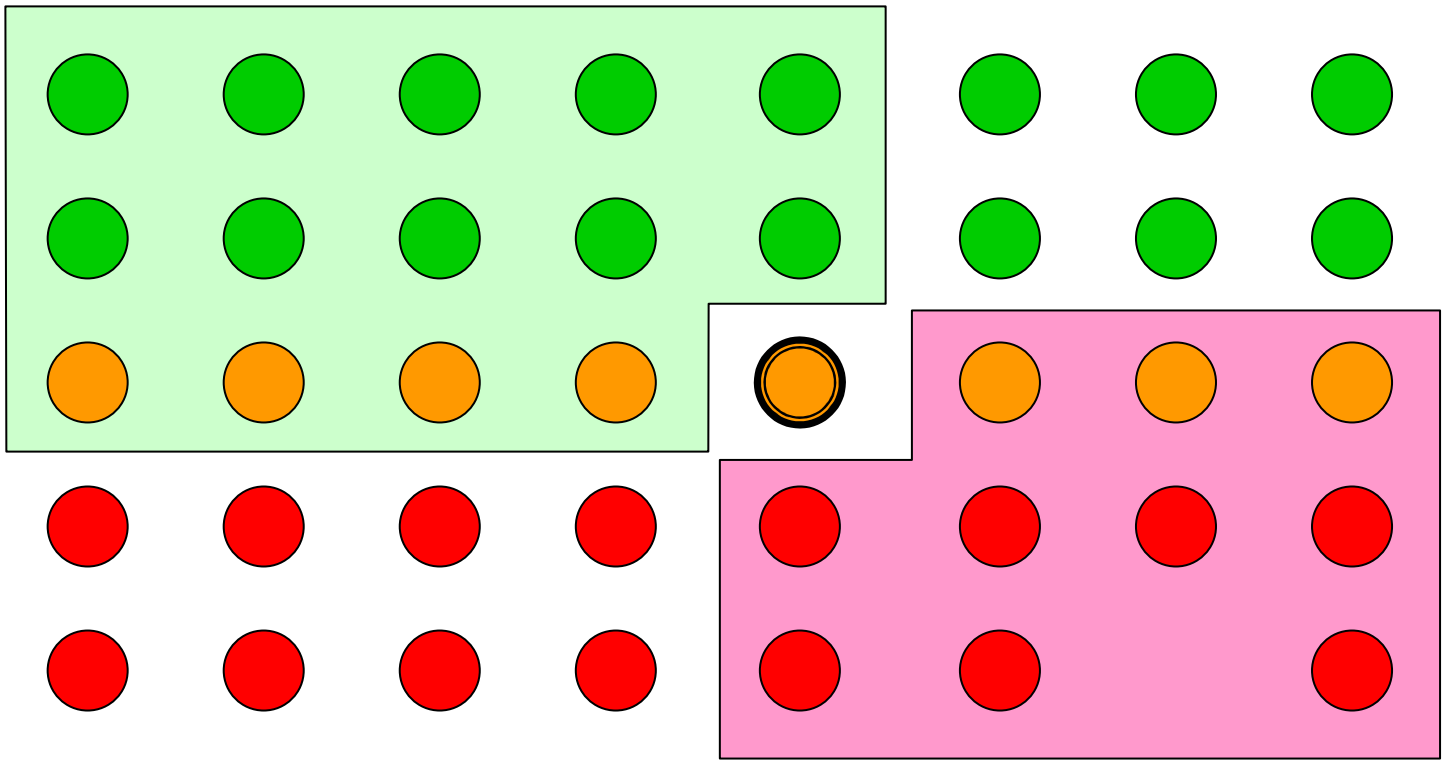

All groups of 5 elements.

(And at most one smaller group.)

# Order Statistics: Analysis



# Order Statistics: Analysis 1



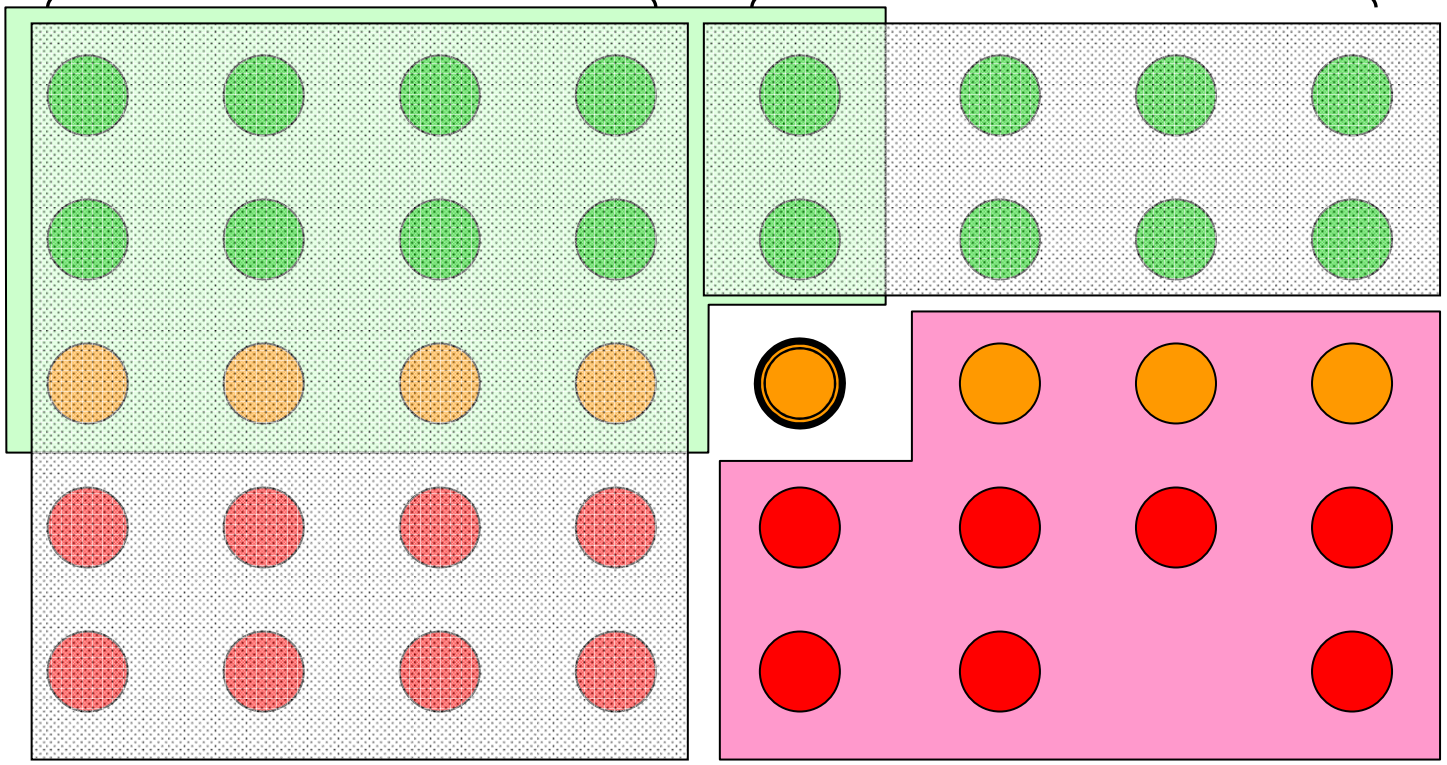
Must recur on all elements outside one of these boxes.

How many?

# Order Statistics: Analysis 1

$\lfloor \lceil n/5 \rceil / 2 \rfloor$ full groups of 5

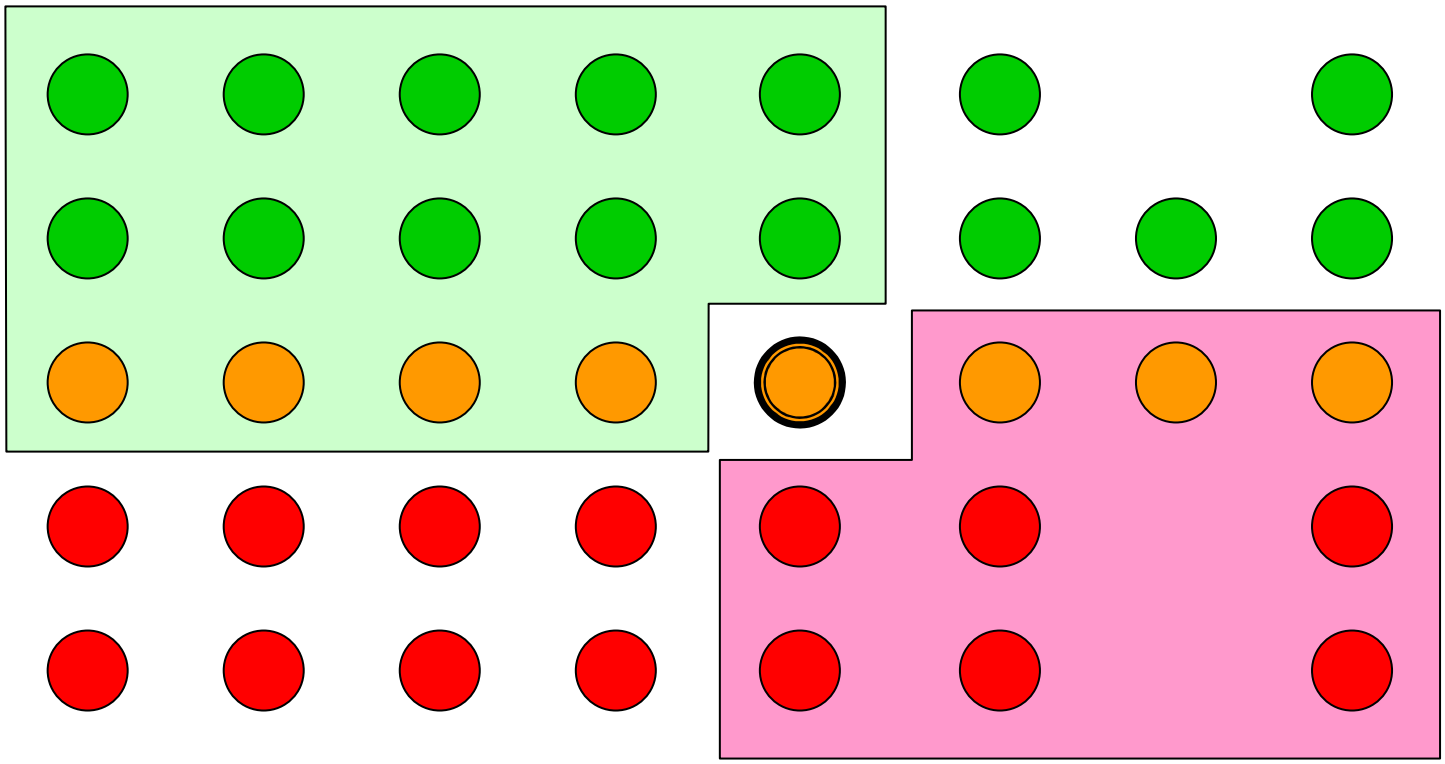
$\lceil \lceil n/5 \rceil / 2 \rceil$ partial groups of 2



Count elements outside smaller box.

At most $5\left\lfloor \left\lceil \frac{n}{5} \right\rceil / 2 \right\rfloor + 2\left\lceil \left\lceil \frac{n}{5} \right\rceil / 2 \right\rceil \le \frac{7n}{10} + 7$

# Order Statistics: Analysis 2

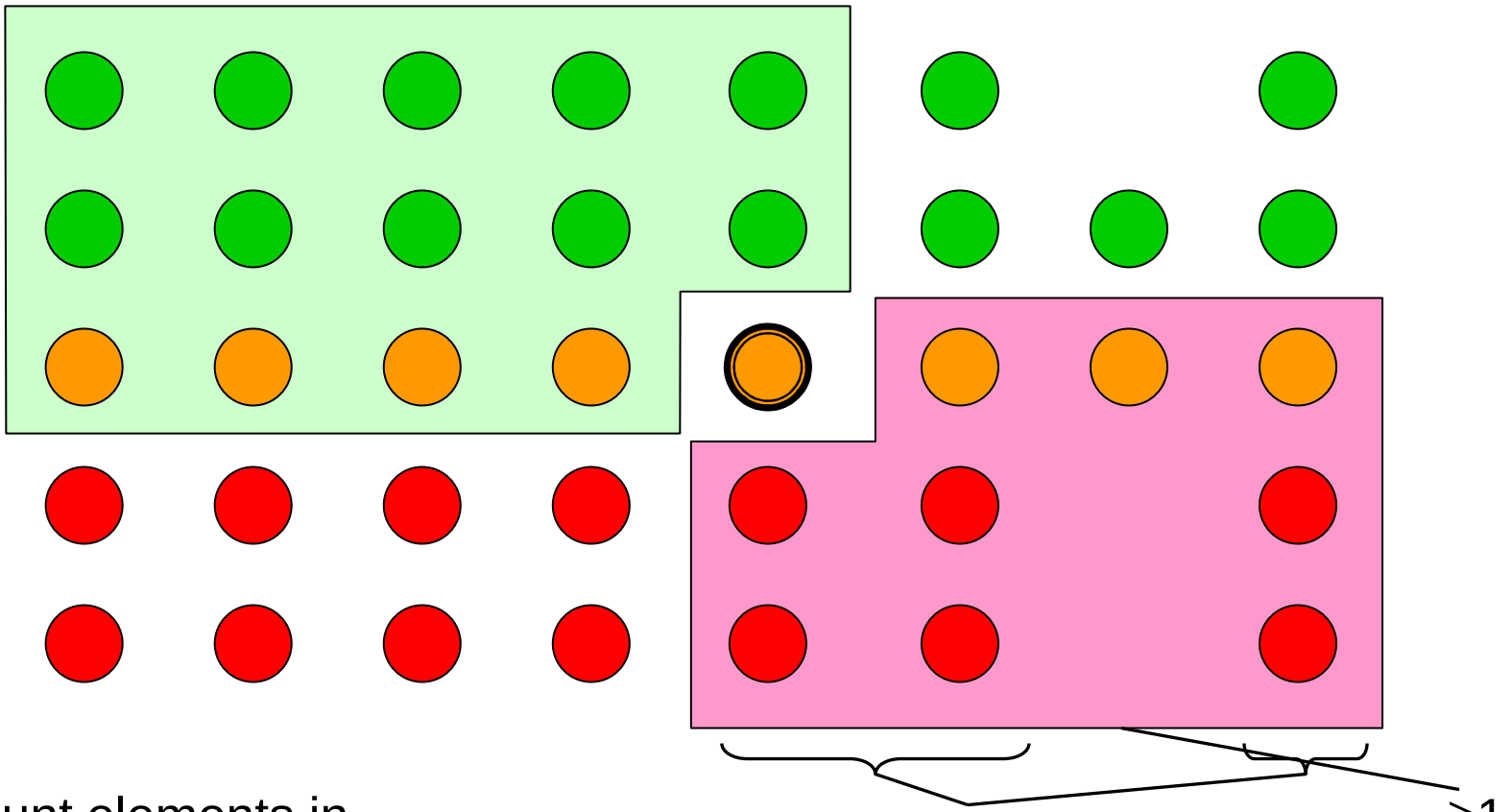


Equivalently, must recur on all elements not inside one of these boxes.

How many?

# Order Statistics: Analysis 2

At most $n - \left( 3\left( \left\lceil \left\lceil \frac{n}{5} \right\rceil / 2 \right\rceil - 1 \right) + 1 \right) \leq \frac{7n}{10} + 2$



Count elements in smaller box & pivot.

$\left\lceil \left\lceil n/5 \right\rceil / 2 \right\rceil - 1$ groups of 3

≥1

# Order Statistics: Analysis

$$T(n) = T\left(\left\lceil \frac{n}{5} \right\rceil\right) + T\left(\frac{7n}{10} + 2\right) + O(n)$$

A very unusual recurrence. How to solve?

?

?

# Order Statistics: Analysis

Substitution: Prove $\exists c, n_0 > 0, T(n) \le c \cdot n, \forall n \ge n_0$

$$T(n) \le c\left\lceil \frac{n}{5} \right\rceil + c\left(\frac{7n}{10} + 2\right) + kn$$

$$\le c\left(\frac{n}{5} + 1\right) + c\left(\frac{7n}{10} + 2\right) + kn$$ Overestimate ceiling

$$= \frac{9}{10}cn + 3c + kn$$ Algebra

$$\le cn$$

when $0 \le \frac{1}{10}cn - 3c - kn$

$\forall$ k, can find $c, n_0$ such that this holds $\forall n \ge n_0$.

# Order Statistics

? Why groups of 5? ?

Sum of two recurrence sizes must be < 1.

Grouping by 5 is smallest size that works.