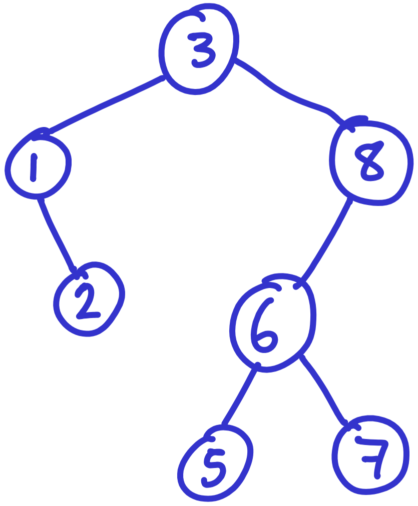
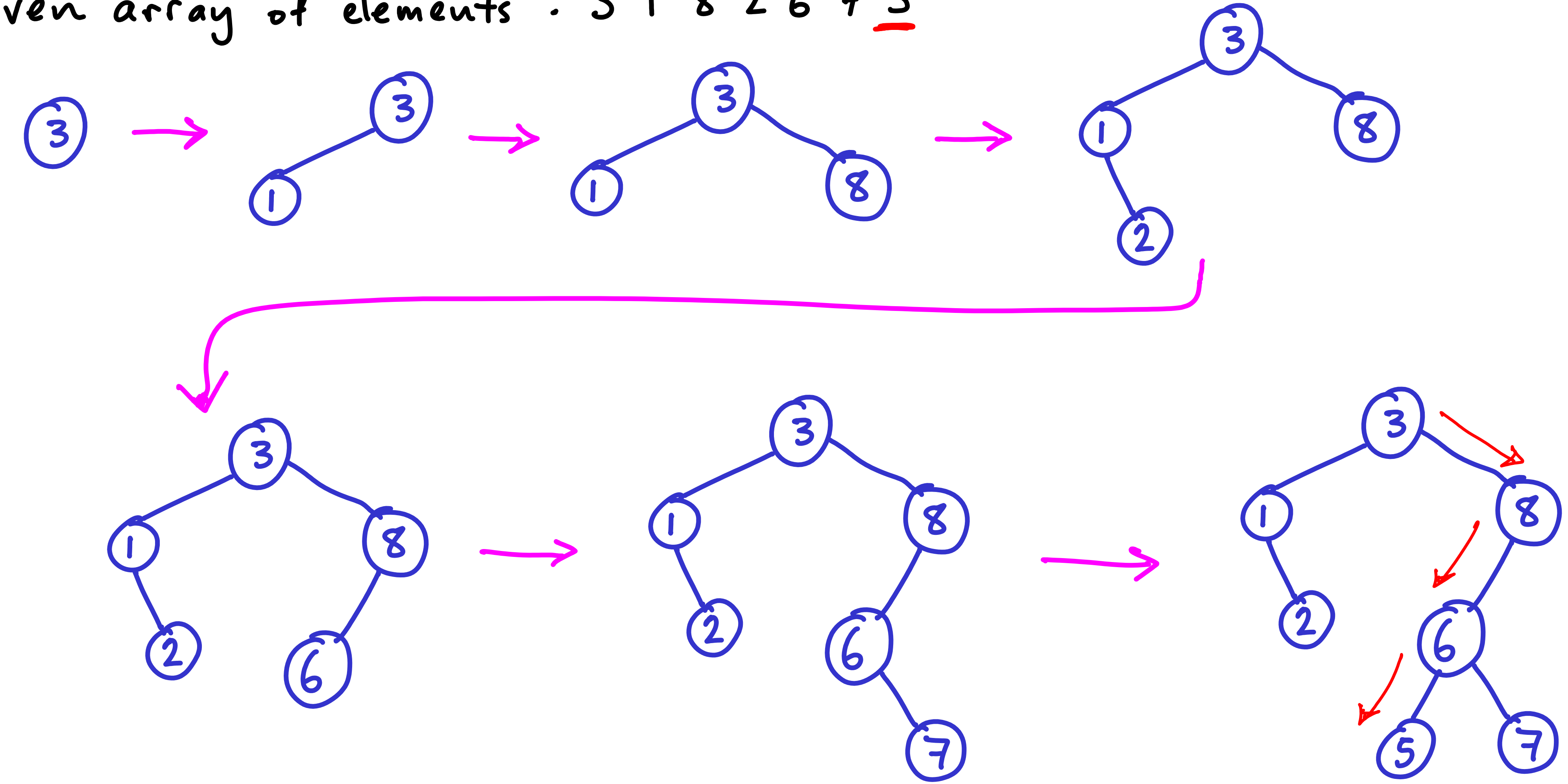
BINARY SEARCH TREES – BUILT RANDOMLY



Insert n elements into a BST in the order that they're given.

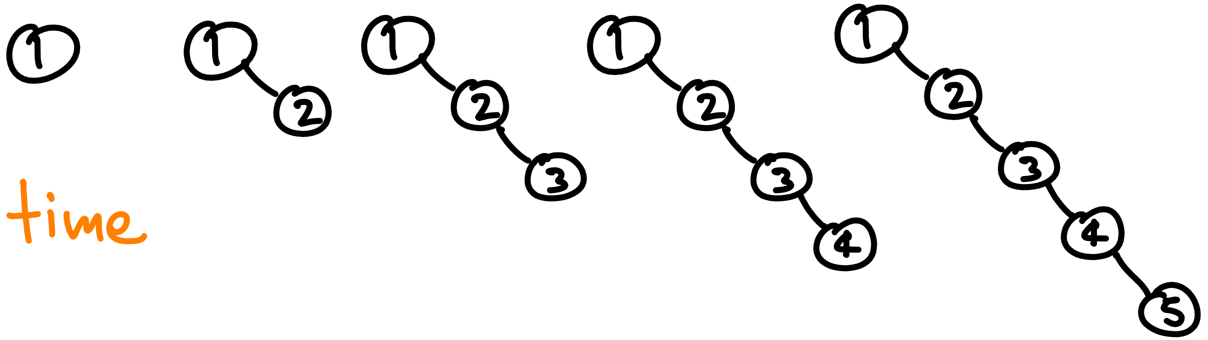
Given array of elements : 3 1 8 2 6 7 5



- What is the worst-case time complexity, and why?
- How unbalanced could the tree be?

↳ already sorted input, reverse-sorted, nearly sorted...

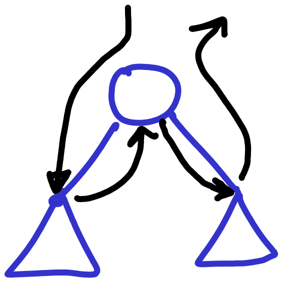
1 2 3 4 5 →



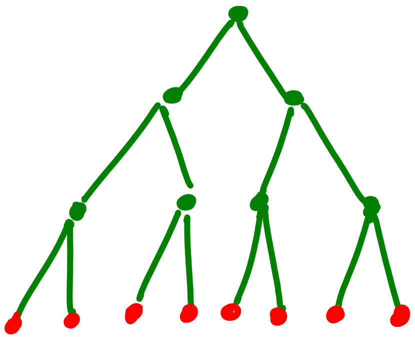
$O(n)$ depth, $O(n^2)$ time

- What would be ideal? → balanced partition every time
- Worst-case time complexity $= \Omega(?)$

Hint: after constructing the BST, what does an in-order traversal give?

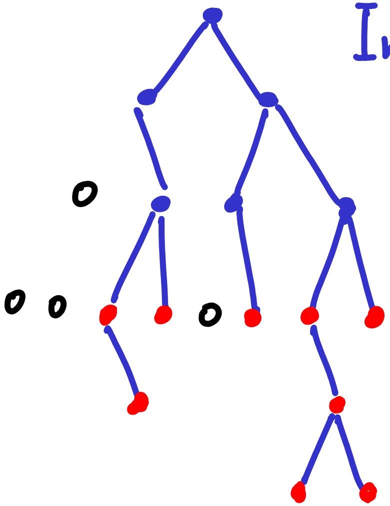


Constructing a BST essentially sorts data so the sorting lower bound applies $\Big\}$ worst case $\Omega(n\log n)$



In a balanced tree, $\sim\frac{n}{2}$ nodes have depth $\sim\log n$

so they take $\Omega(n\log n)$ time to insert.



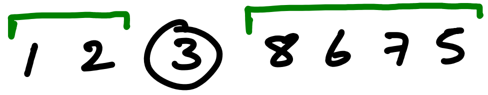
In any other tree we can find $\geq\frac{n}{2}$ nodes with depth $\geq\log n$

$\hookrightarrow$ Every algorithm requires $\Omega(n\log n)$ time to construct any BST

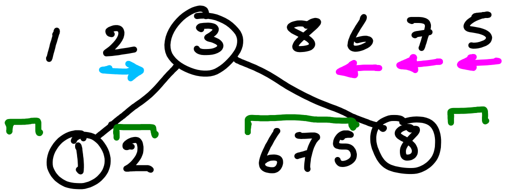
Stable quicksort

3 1 8 2 6 7 5

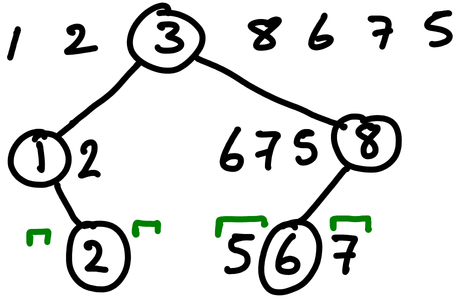
- use first elt to partition



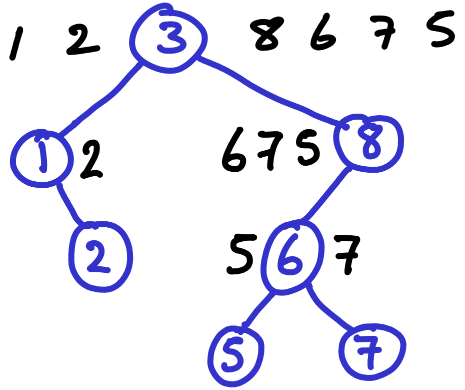
- repeat on each side



- 3rd round



- 4th round



same tree as BST

quicksort round 1: compare all elts to 3

BST 3 = root; eventually all elts pass through.

quicksort: partitions into 2 groups <3 & >3
each is independent

BST same

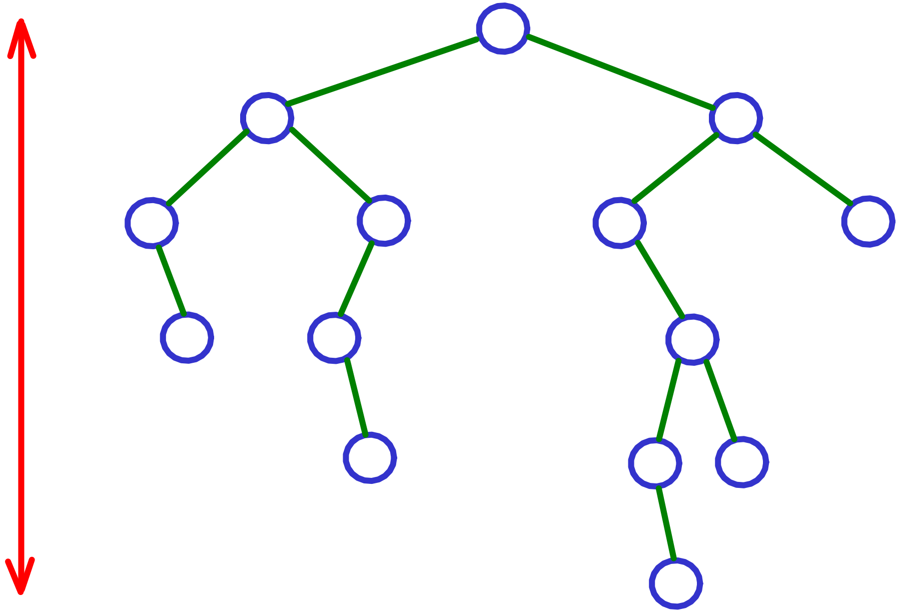
exactly same comparisons but in different order

Conclusion:

The expected time complexity of building a BST on random data is the same as for Quicksort: O(nlogn)

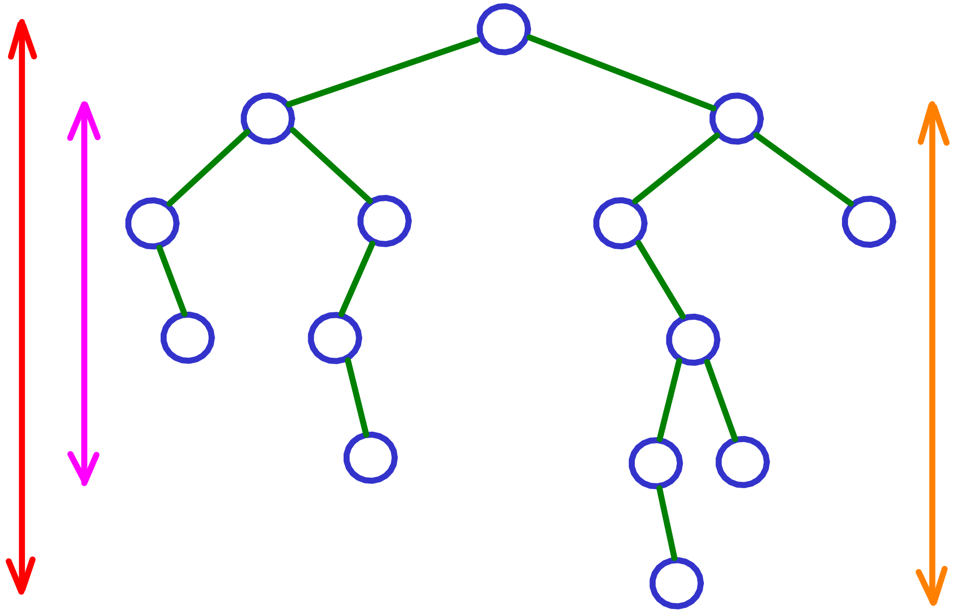
Unfortunately this doesn't imply anything useful about the expected depth of a random BST.

In particular it doesn't imply expected O(logn) depth.
There exist trees that have much greater depth, but that can be constructed in O(nlogn) time.



$H(n) = ?$

Expected height of a randomly constructed BST with n elements



$$H(n) = 1 + \max\{H(k), H(n-k-1)\}$$

for some random k $(0 \le k \le n-1)$

with more rigorous analysis we can get $\sim 3\log n$

If $\frac{n}{4} \le k \le \frac{3n}{4}$

50%

$$\hookrightarrow H(n) \le 1 + H\left(\frac{3n}{4}\right)$$

else $H(n) \le 1 + H(n-1)$

50%

$$< 1 + H(n)$$

Expected height will be:

$$H(n) \le 1 + \frac{1}{2} H\left(\frac{3n}{4}\right) + \frac{1}{2} H(n)$$

$$\frac{1}{2} H(n) \le 1 + \frac{1}{2} H\left(\frac{3n}{4}\right)$$

$$H(n) \le 2 + H\left(\frac{3n}{4}\right) = O(\log n)$$