

**NPTEL MOOC**

# **PROGRAMMING, DATA STRUCTURES AND ALGORITHMS IN PYTHON**

**Week 6, Lecture 1**

**Madhavan Mukund, Chennai Mathematical Institute**

**<http://www.cmi.ac.in/~madhavan>**



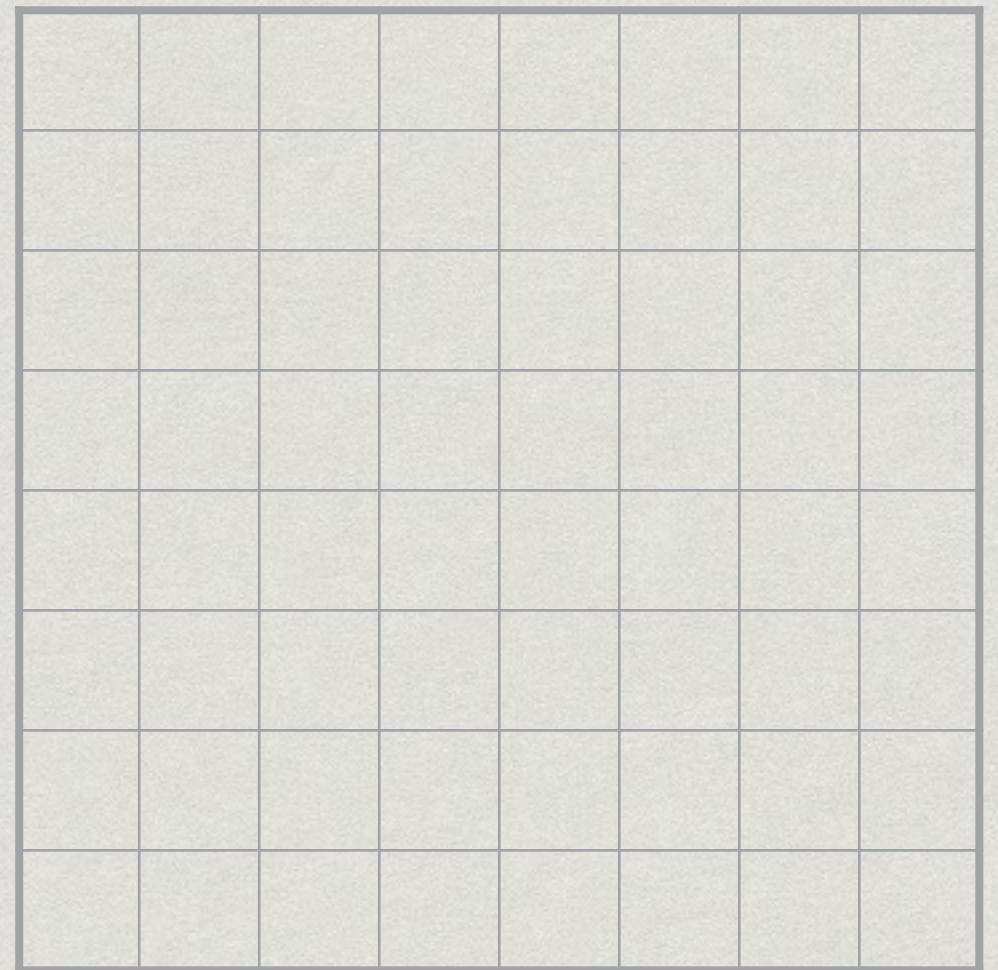
# Backtracking

- \* Systematically search for a solution
- \* Build the solution one step at a time
- \* If we hit a dead-end
  - \* Undo the last step
  - \* Try the next option



# Eight queens

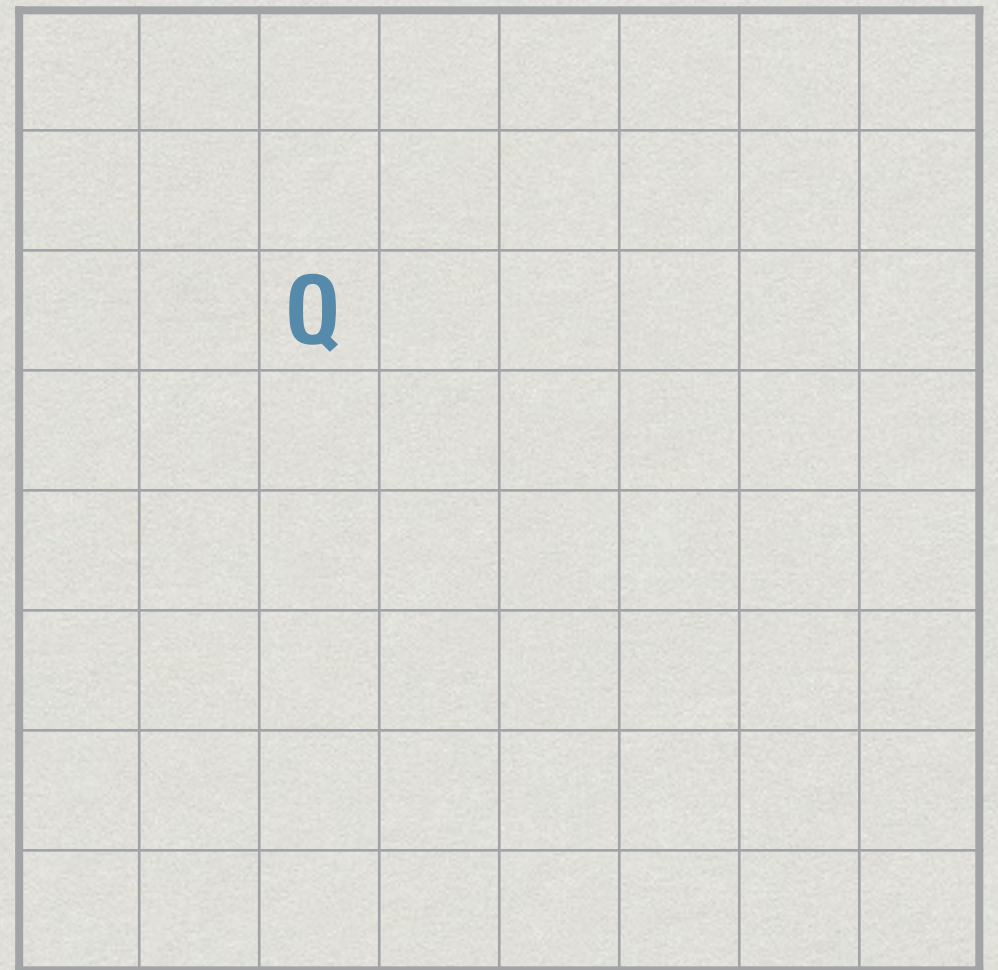
- \* Place 8 queens on a chess board so that none of them attack each other
- \* In chess, a queen can move any number of squares along a row column or diagonal





# Eight queens

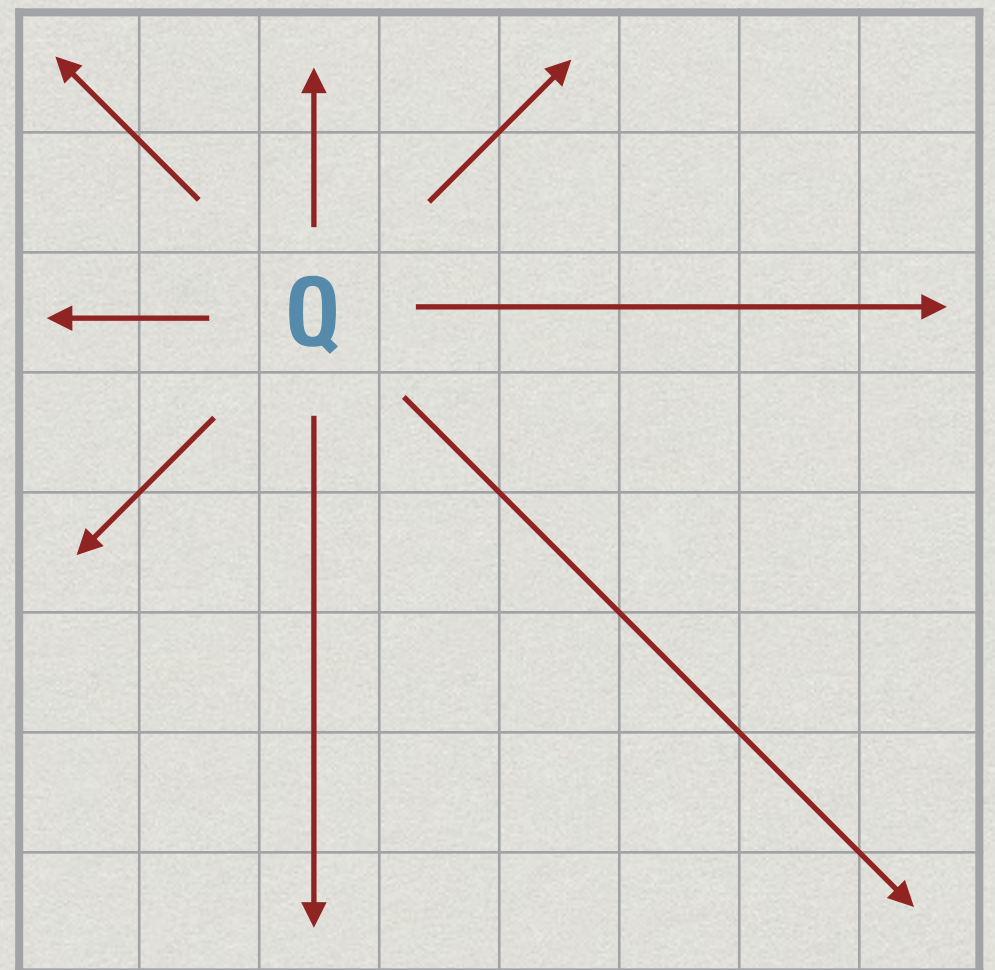
- \* Place 8 queens on a chess board so that none of them attack each other
- \* In chess, a queen can move any number of squares along a row column or diagonal





# Eight queens

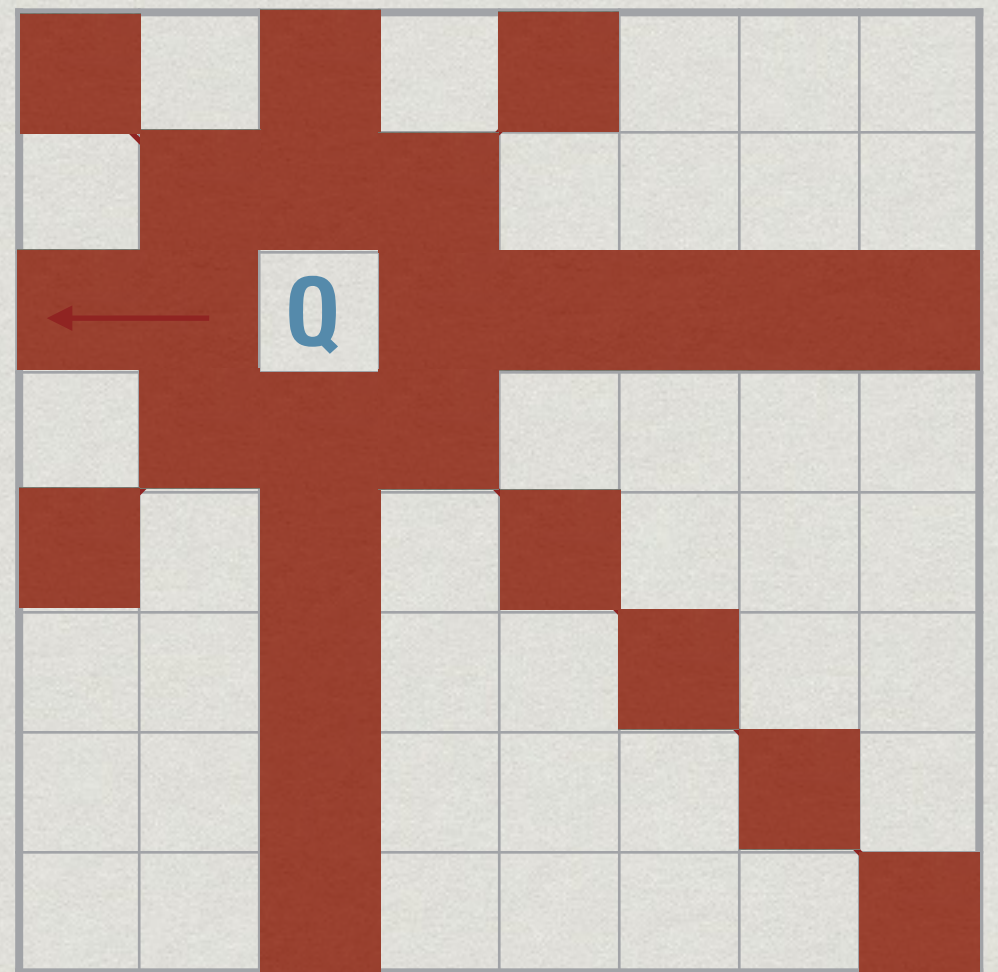
- \* Place 8 queens on a chess board so that none of them attack each other
- \* In chess, a queen can move any number of squares along a row column or diagonal





# Eight queens

- \* Place 8 queens on a chess board so that none of them attack each other
- \* In chess, a queen can move any number of squares along a row column or diagonal





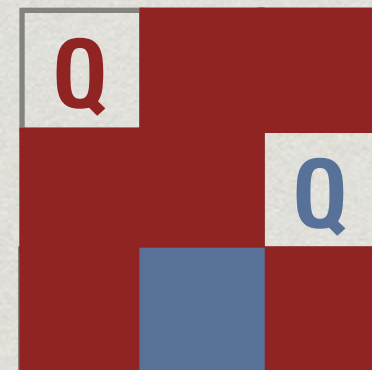
# N queens

- \* Place N queens on an N x N chess board so that none attack each other
- \* N = 2, 3 impossible



# N queens

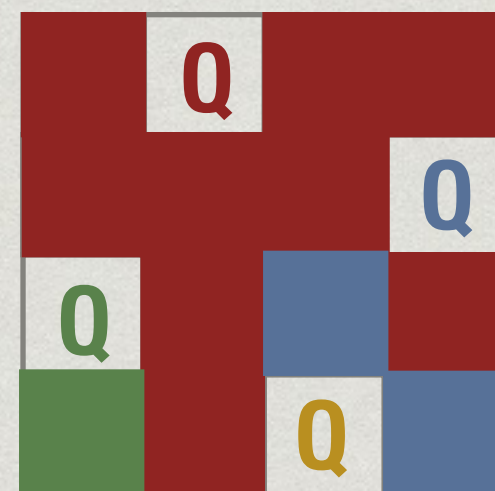
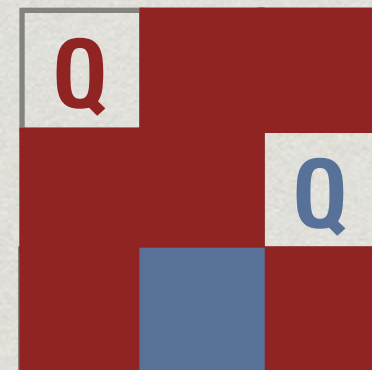
- \* Place N queens on an N x N chess board so that none attack each other
- \* N = 2, 3 impossible





# N queens

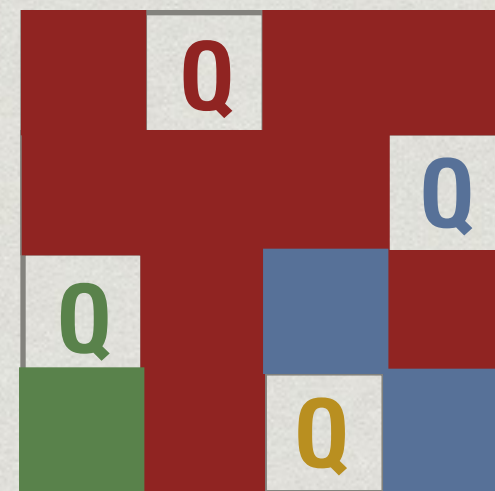
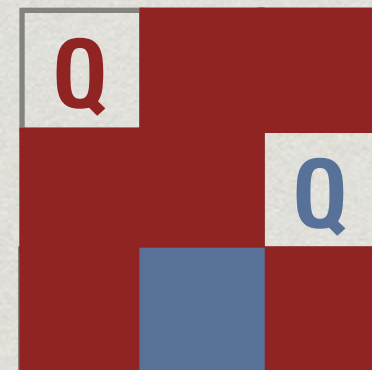
- \* Place N queens on an N x N chess board so that none attack each other
- \* N = 2, 3 impossible
- \* N = 4 is possible





# N queens

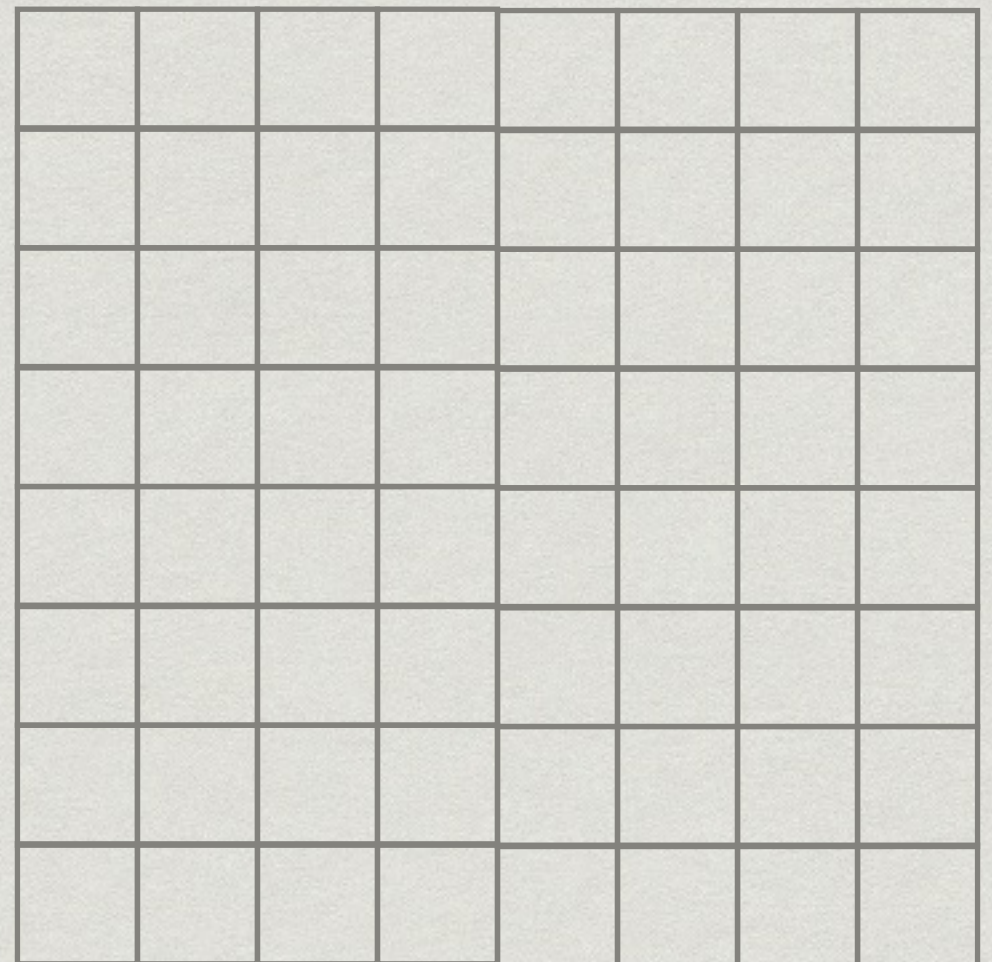
- \* Place N queens on an N x N chess board so that none attack each other
- \* N = 2, 3 impossible
- \* N = 4 is possible
- \* And all bigger N as well





# 8 queens

- \* Clearly, exactly one queen in each row, column
- \* Place queens row by row
- \* In each row, place a queen in the first available column







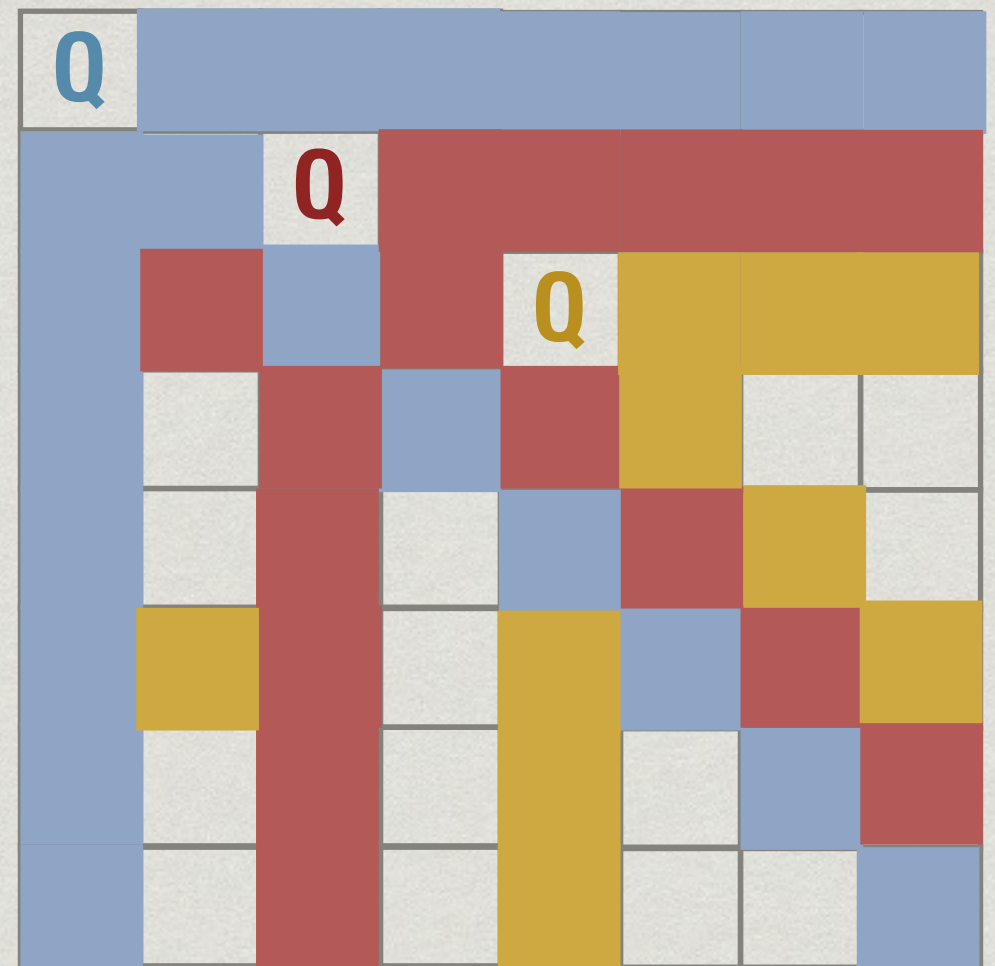






# 8 queens

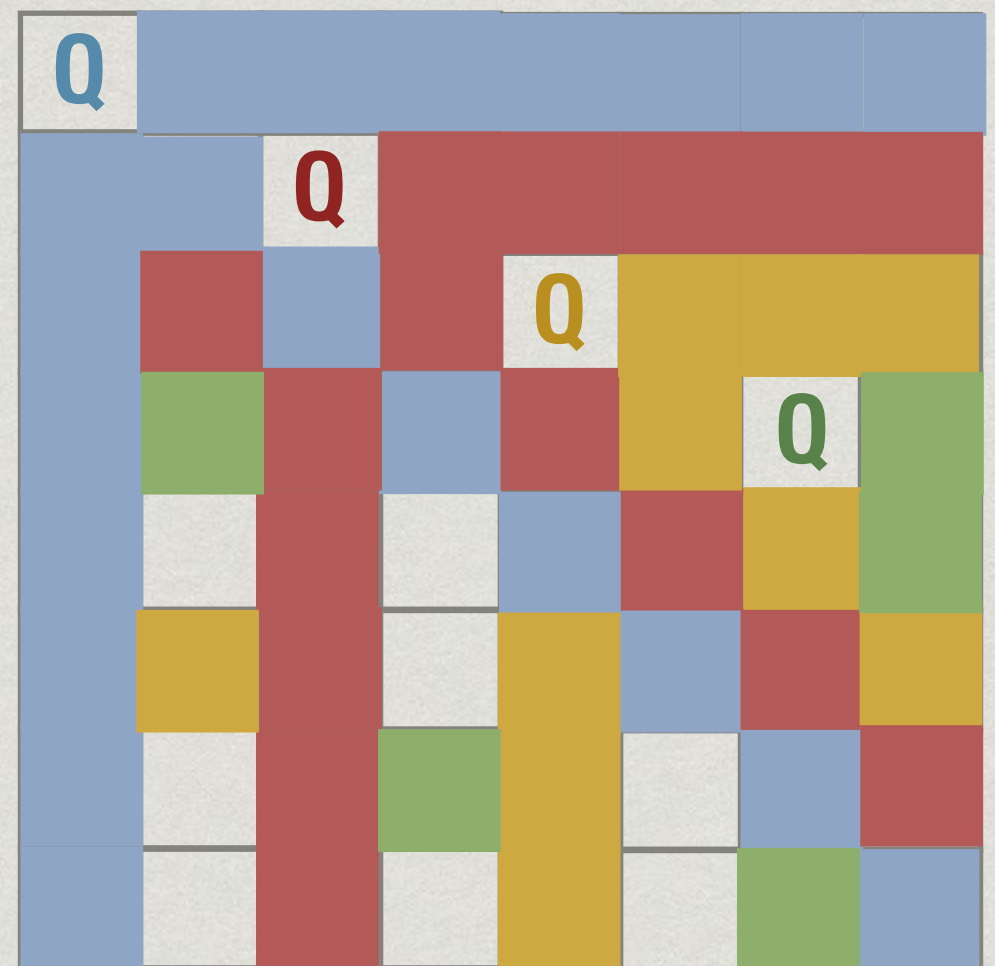
- \* Clearly, exactly one queen in each row, column
- \* Place queens row by row
- \* In each row, place a queen in the first available column





# 8 queens

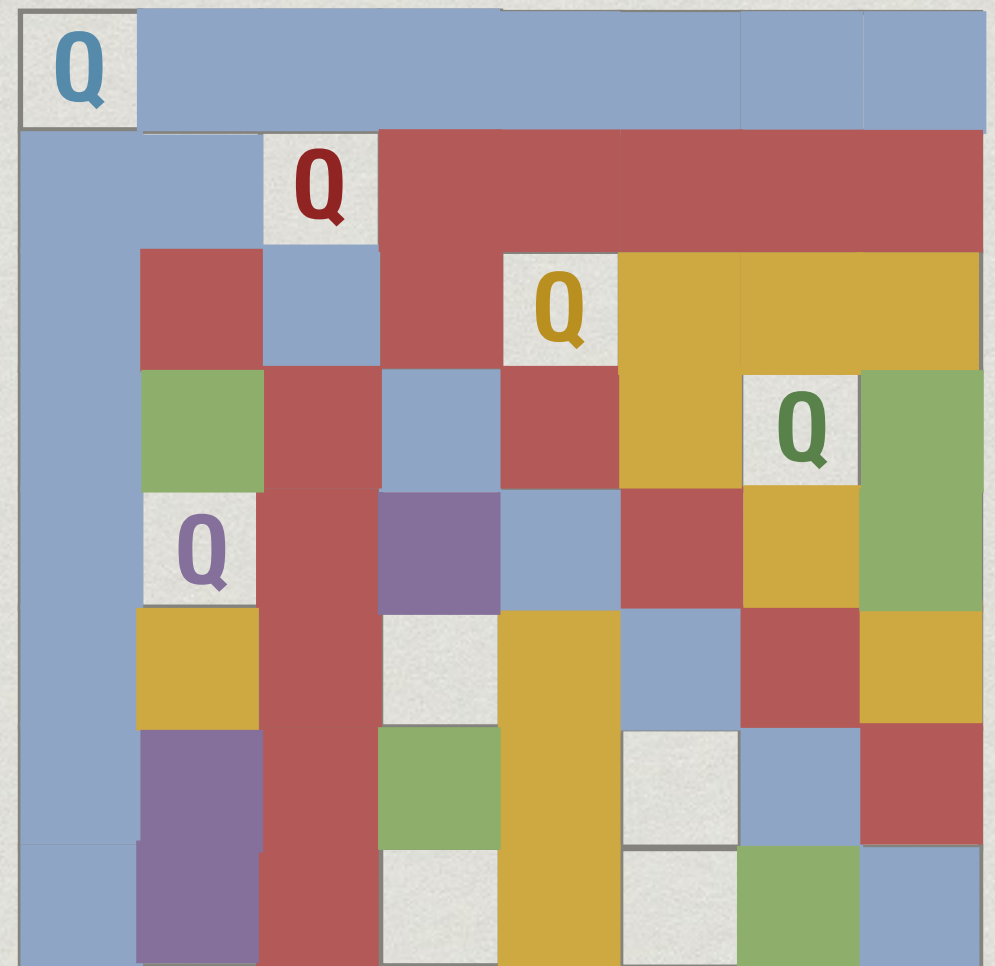
- \* Clearly, exactly one queen in each row, column
- \* Place queens row by row
- \* In each row, place a queen in the first available column





# 8 queens

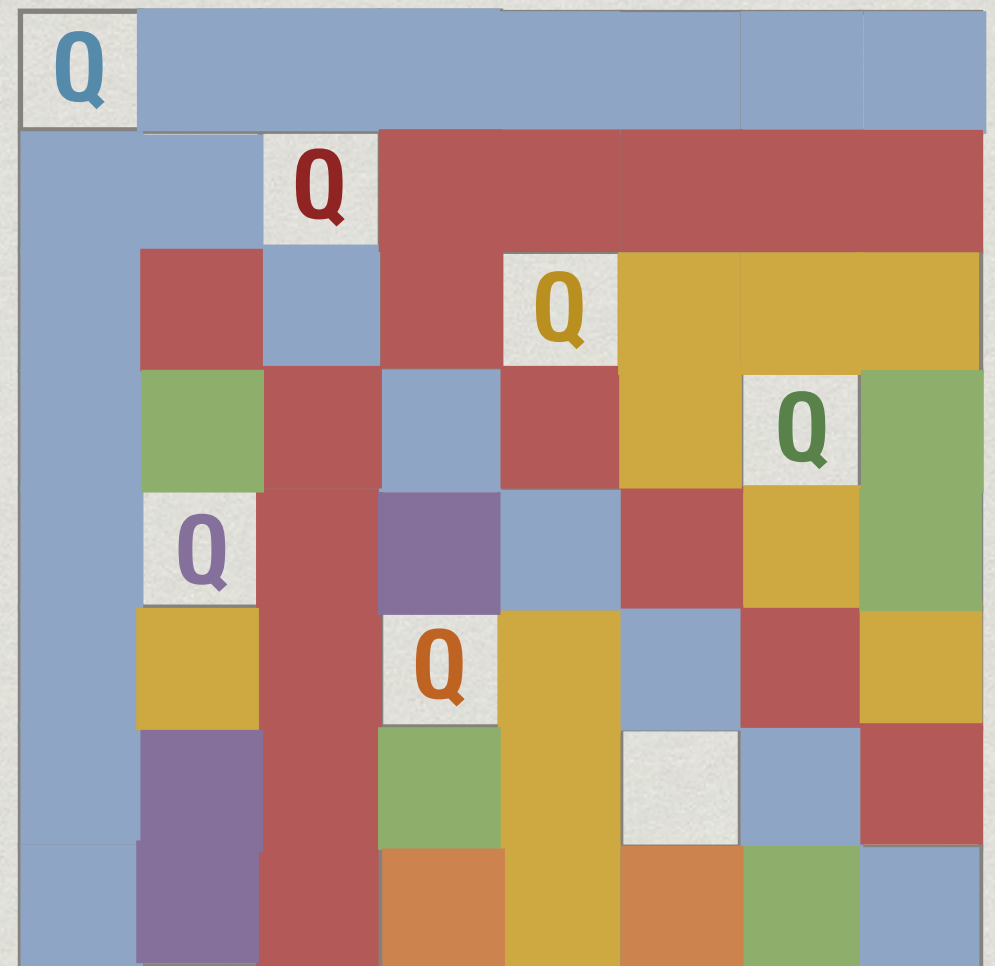
- \* Clearly, exactly one queen in each row, column
- \* Place queens row by row
- \* In each row, place a queen in the first available column





# 8 queens

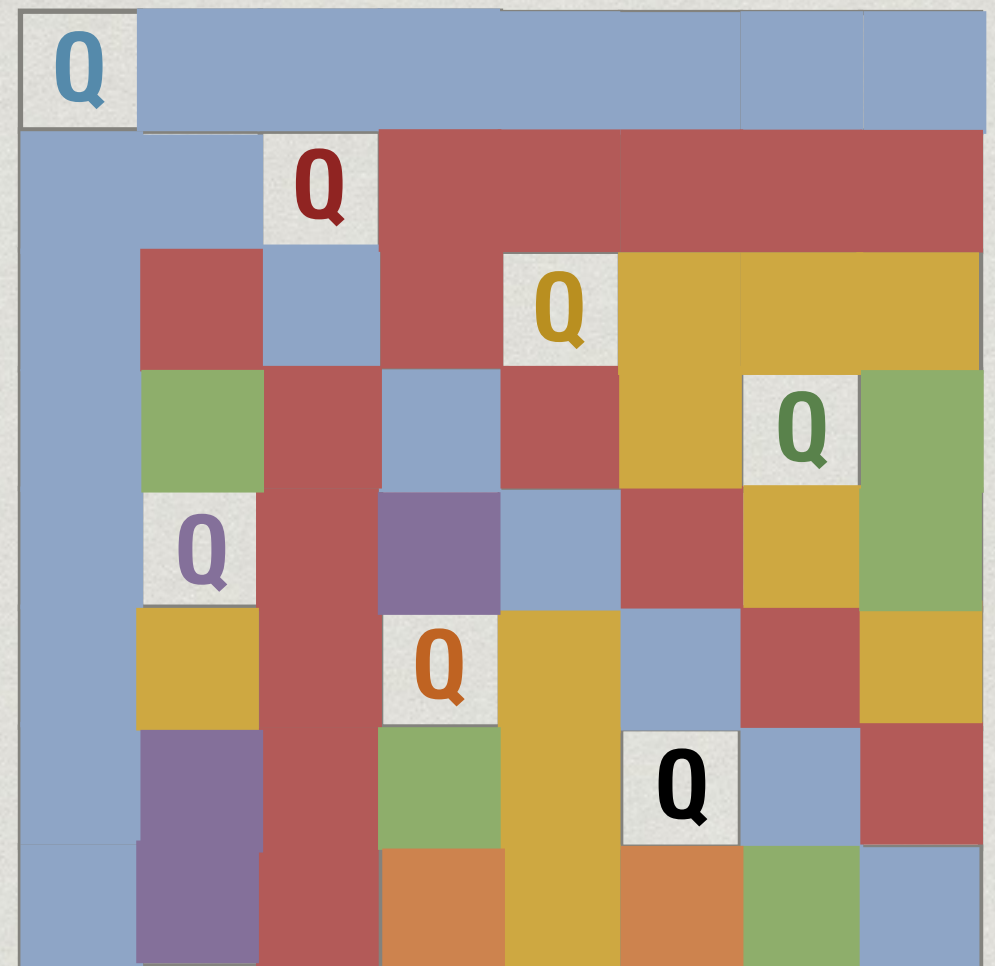
- \* Clearly, exactly one queen in each row, column
- \* Place queens row by row
- \* In each row, place a queen in the first available column





# 8 queens

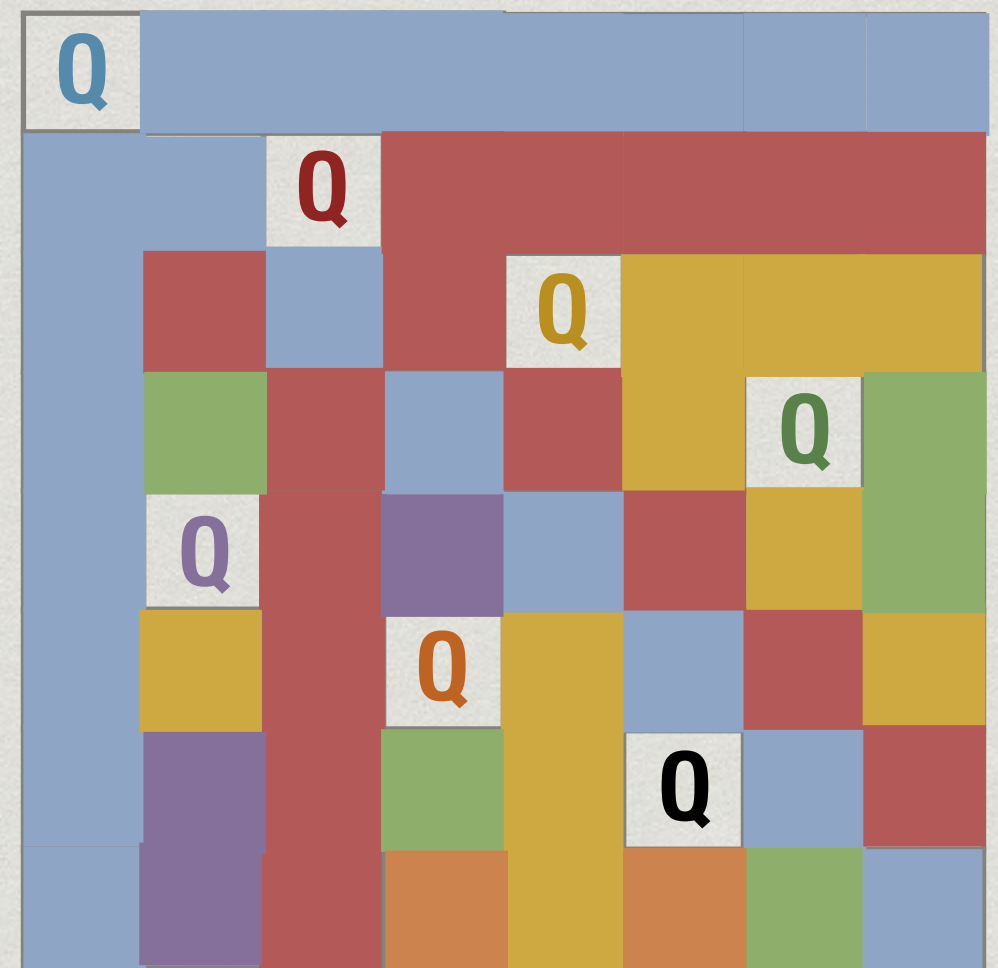
- \* Clearly, exactly one queen in each row, column
- \* Place queens row by row
- \* In each row, place a queen in the first available column





# 8 queens

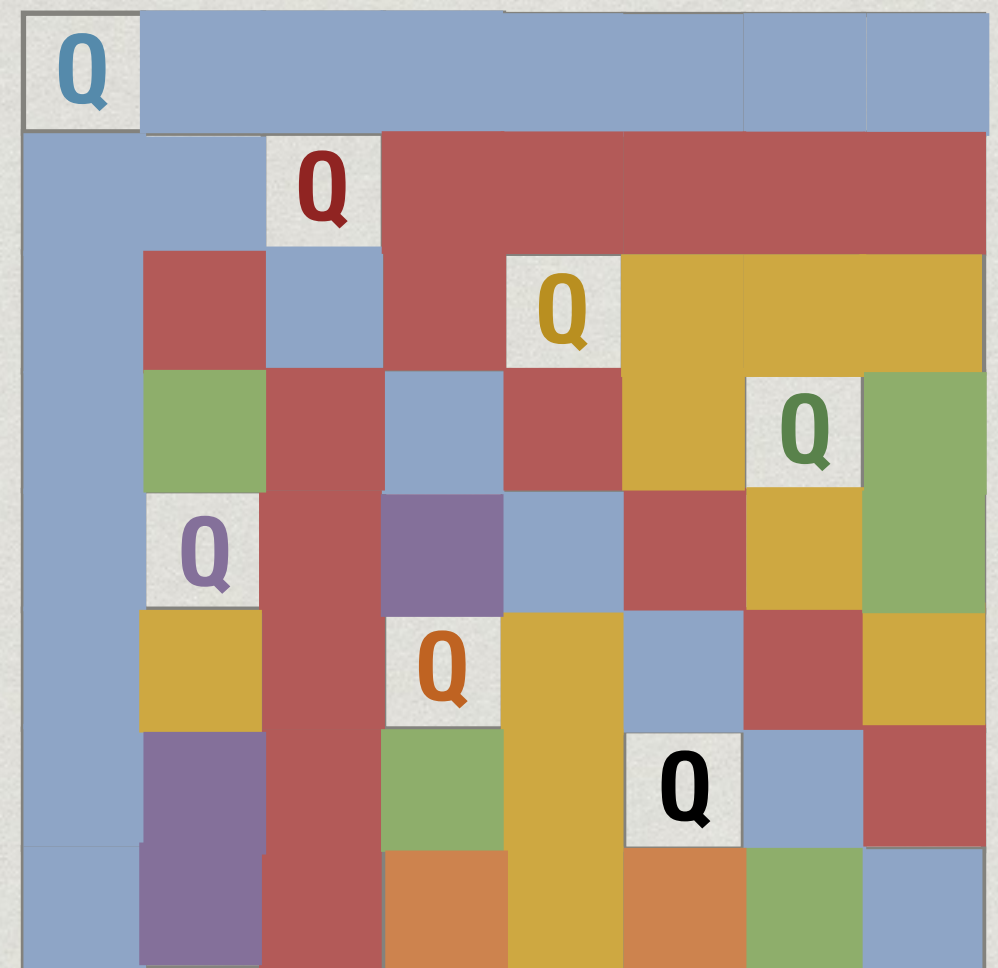
- \* Clearly, exactly one queen in each row, column
- \* Place queens row by row
- \* In each row, place a queen in the first available column
- \* Can't place a queen in the 8th row!





# 8 queens

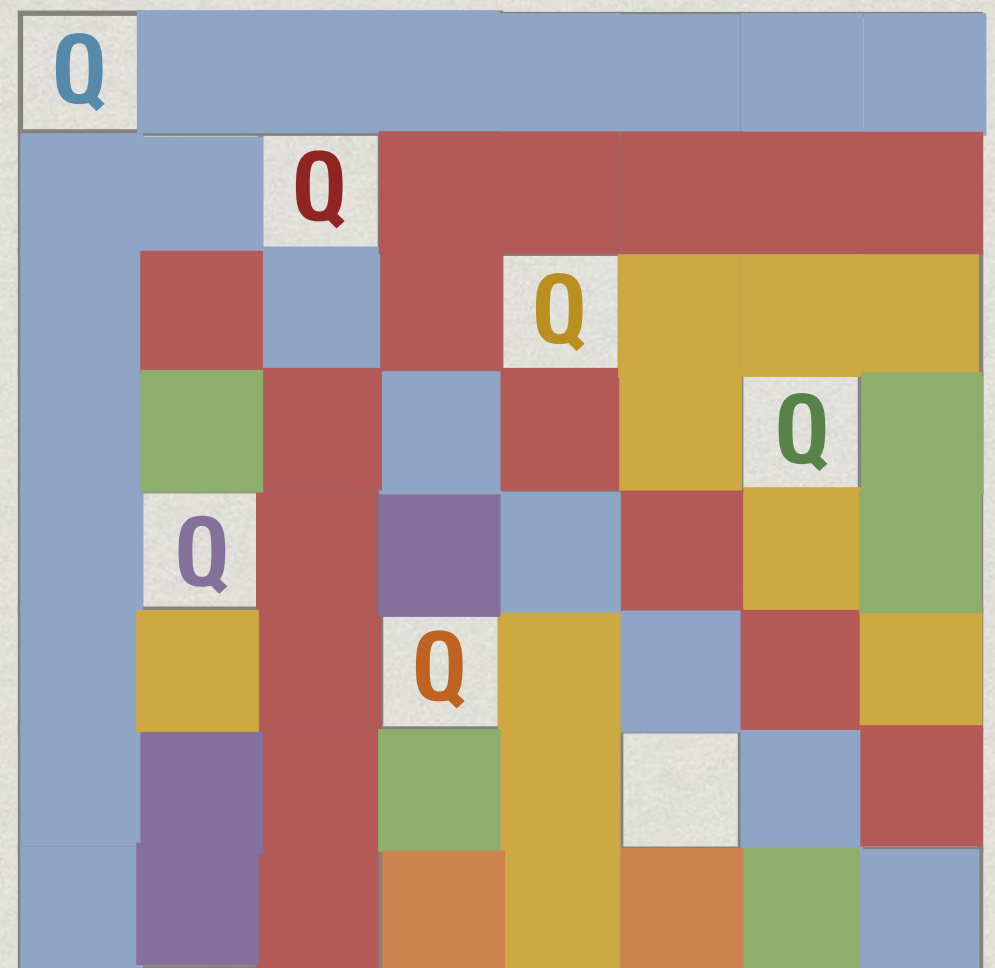
- \* Can't place the a queen in the 8th row!





# 8 queens

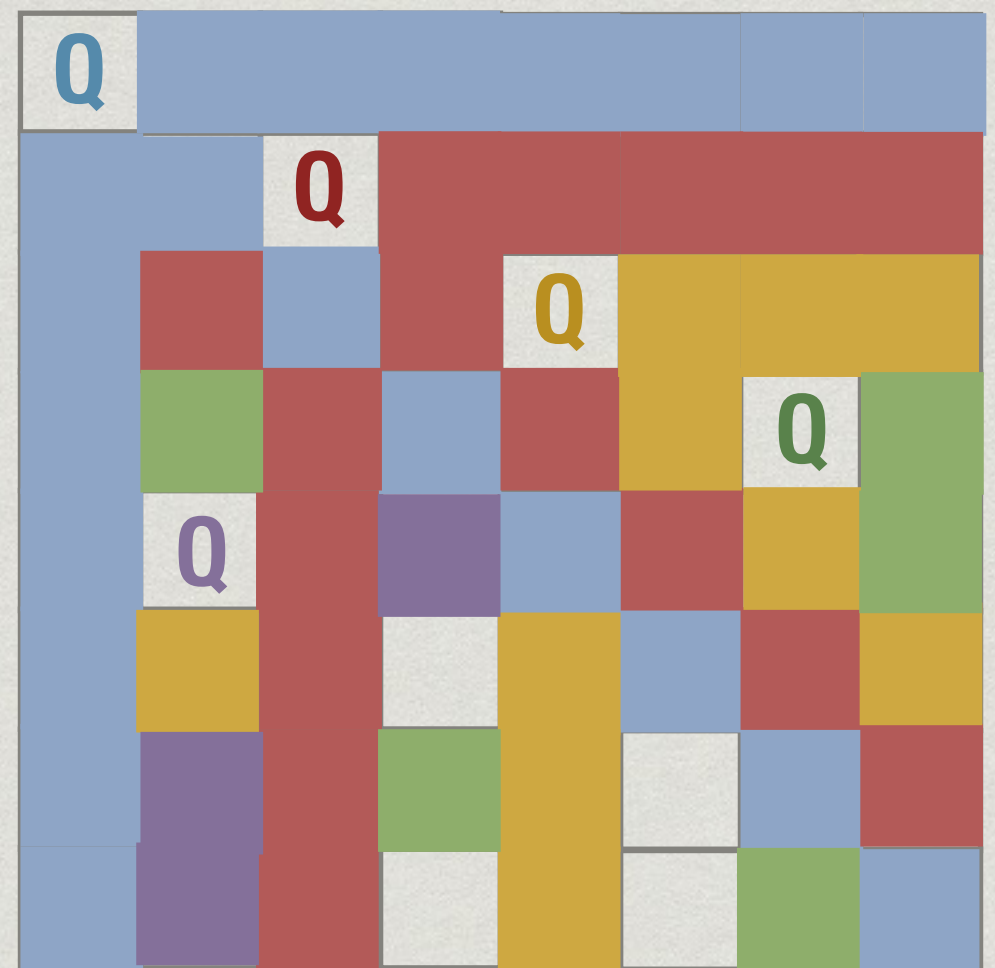
- \* Can't place the a queen in the 8th row!
- \* Undo 7th queen, no other choice





# 8 queens

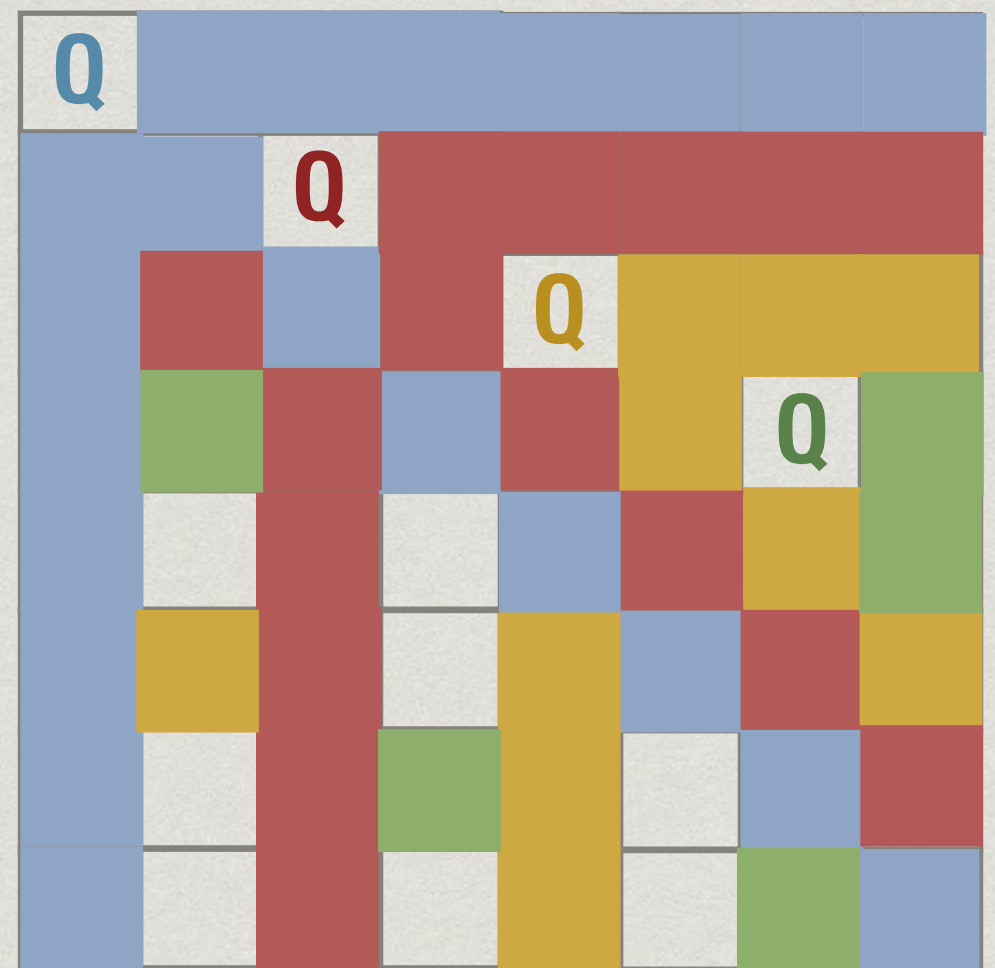
- \* Can't place the a queen in the 8th row!
- \* Undo 7th queen, no other choice
- \* Undo 6th queen, no other choice





# 8 queens

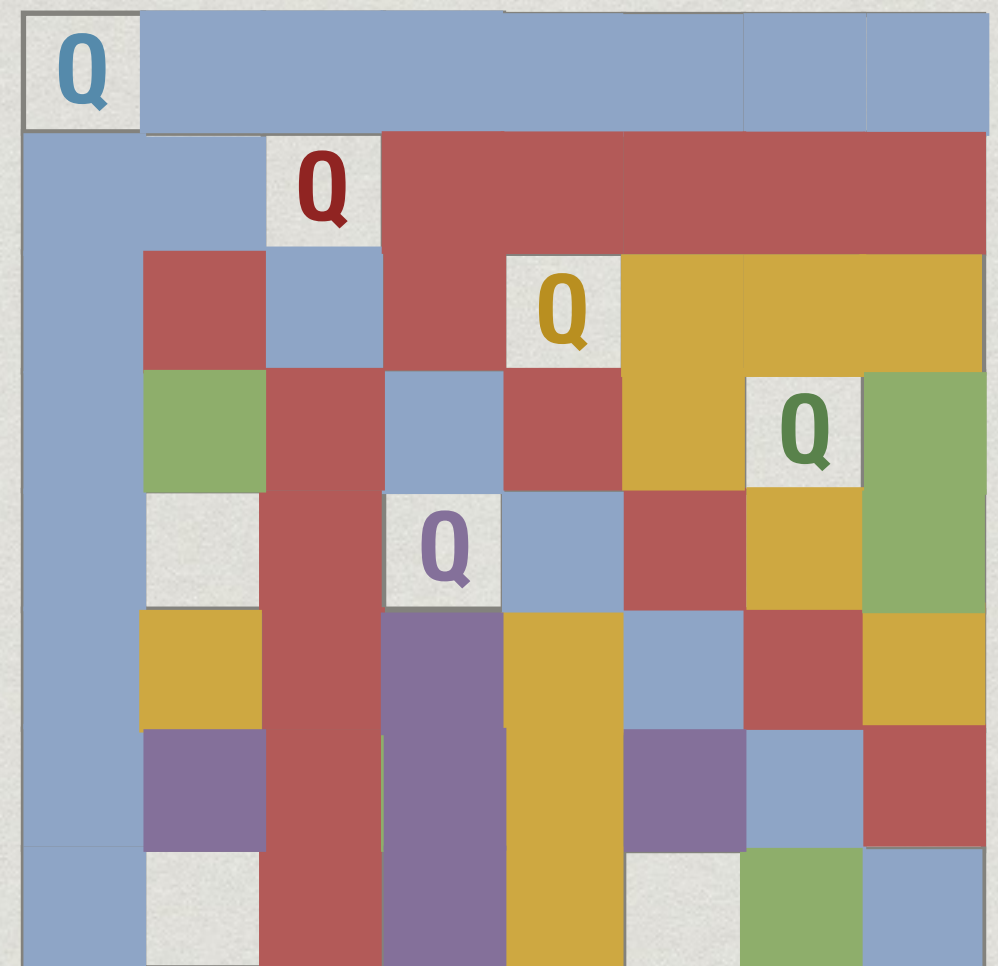
- \* Can't place the a queen in the 8th row!
- \* Undo 7th queen, no other choice
- \* Undo 6th queen, no other choice
- \* Undo 5th queen, try next





# 8 queens

- \* Can't place the a queen in the 8th row!
- \* Undo 7th queen, no other choice
- \* Undo 6th queen, no other choice
- \* Undo 5th queen, try next





# Backtracking

- \* Keep trying to extend the next solution
- \* If we cannot, undo previous move and try again
- \* Exhaustively search through all possibilities
- \* ... but systematically!



# Coding the solution

- \* How do we represent the board?
- \*  $n \times n$  grid, number rows and columns from 0 to  $n-1$ 
  - \* `board[i][j] == 1` indicates queen at  $(i, j)$
  - \* `board[i][j] == 0` indicates no queen
- \* We know there is only one queen per row
- \* Single list `board` of length  $n$  with entries 0 to  $n-1$ 
  - \* `board[i] == j` : queen in row  $i$ , column  $j$ , i.e.  $(i, j)$



# Overall structure

```
def placequeen(i,board): # Trying row i
    for each c such that (i,c) is available:
        place queen at (i,c) and update board
        if i == n-1:
            return(True) # Last queen has been placed
        else:
            extendsoln = placequeen(i+1,board)
            if extendsoln:
                return(True) # This solution extends fully
            else:
                undo this move and update board
    else:
        return(False) # Row i failed
```



# Updating the board

- \* Our 1-D and 2-D representations keep track of the queens
- \* Need an efficient way to compute which squares are free to place the next queen
- \*  $n \times n$  **attack** grid
  - \* **attack**[i][j] == 1 if (i, j) is attacked by a queen
  - \* **attack**[i][j] == 0 if (i, j) is currently available
- \* How do we undo the effect of placing a queen?
  - \* Which **attack**[i][j] should be reset to 0?



# Updating the board

- \* Queens are added row by row
- \* Number the queens 0 to  $n-1$
- \* Record earliest queen that attacks each square
  - \*  $\text{attack}[i][j] == k$  if  $(i, j)$  was first attacked by queen  $k$
  - \*  $\text{attack}[i][j] == -1$  if  $(i, j)$  is free
- \* Remove queen  $k$  — reset  $\text{attack}[i][j] == k$  to  $-1$ 
  - \* All other squares still attacked by earlier queens



# Updating the board

- \* **attack** requires  $n^2$  space
  - \* Each update only requires  $O(n)$  time
  - \* Only need to scan row, column, two diagonals
- \* Can we improve our representation to use only  $O(n)$  space?



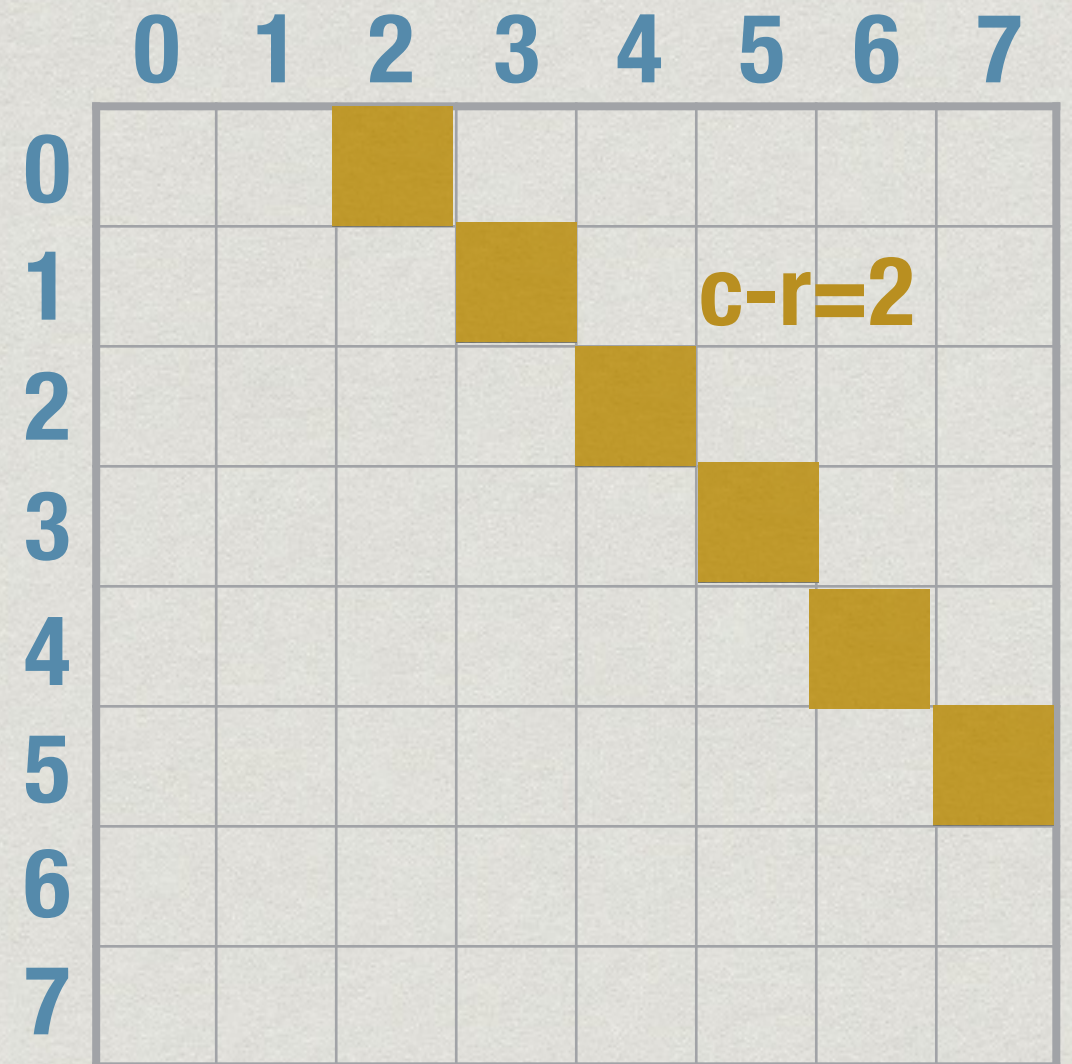
# A better representation

- \* How many queens attack row  $i$ ?
- \* How many queens attack row  $j$ ?
- \* An individual square  $(i,j)$  is attacked by upto 4 queens
  - \* Queen on row  $i$  and on column  $j$
  - \* One queen on each diagonal through  $(i,j)$



# Numbering diagonals

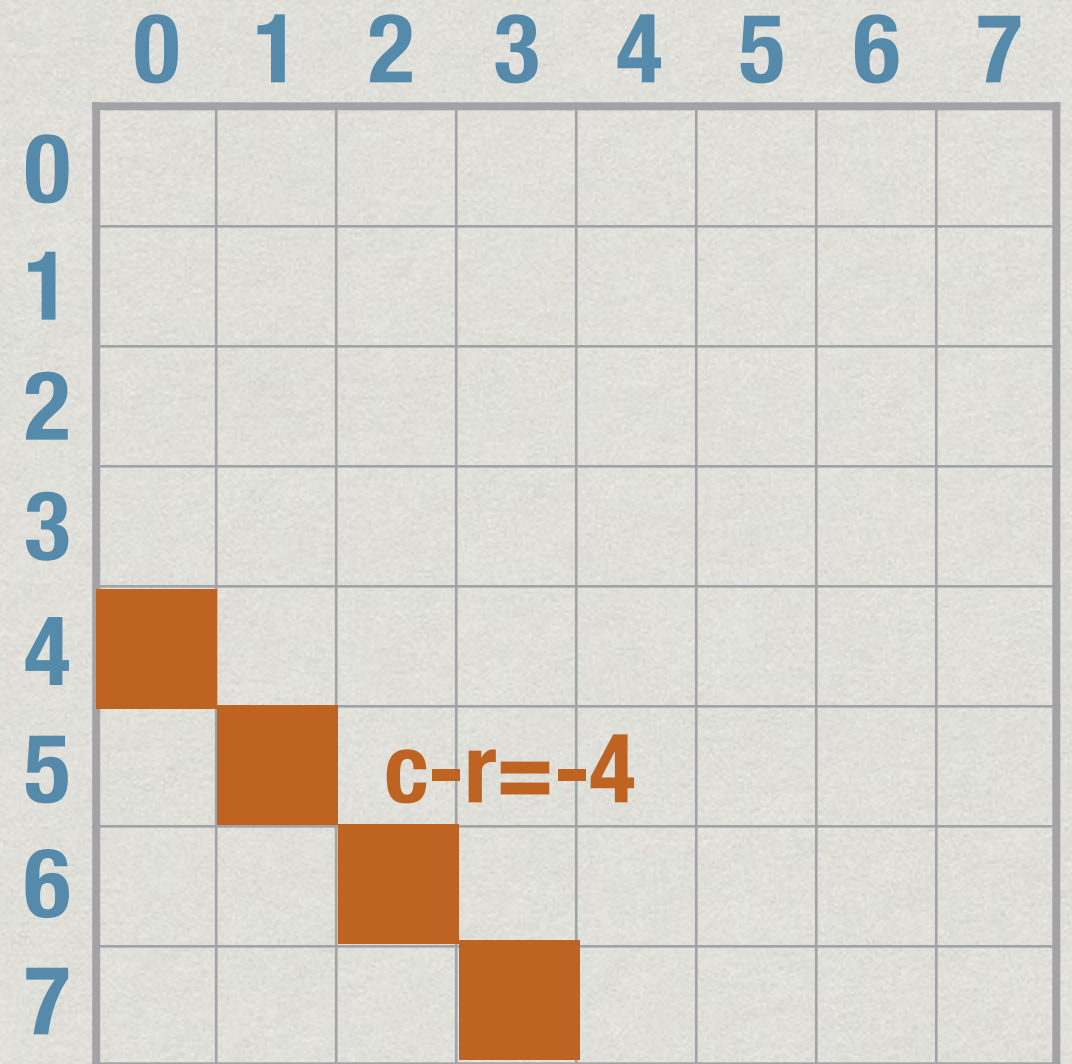
- \* Decreasing diagonal:  
column - row is invariant





# Numbering diagonals

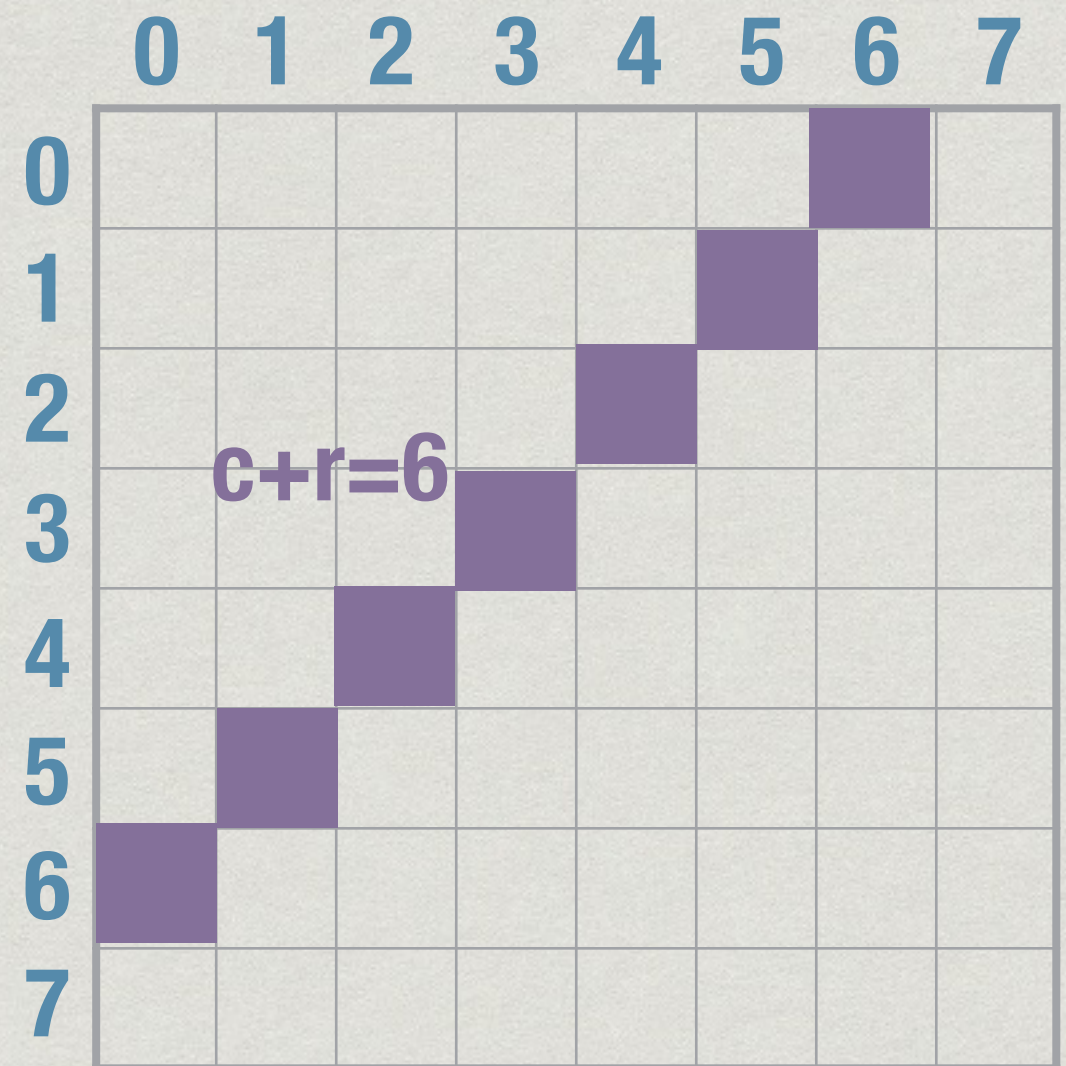
- \* Decreasing diagonal:  
column - row is invariant





# Numbering diagonals

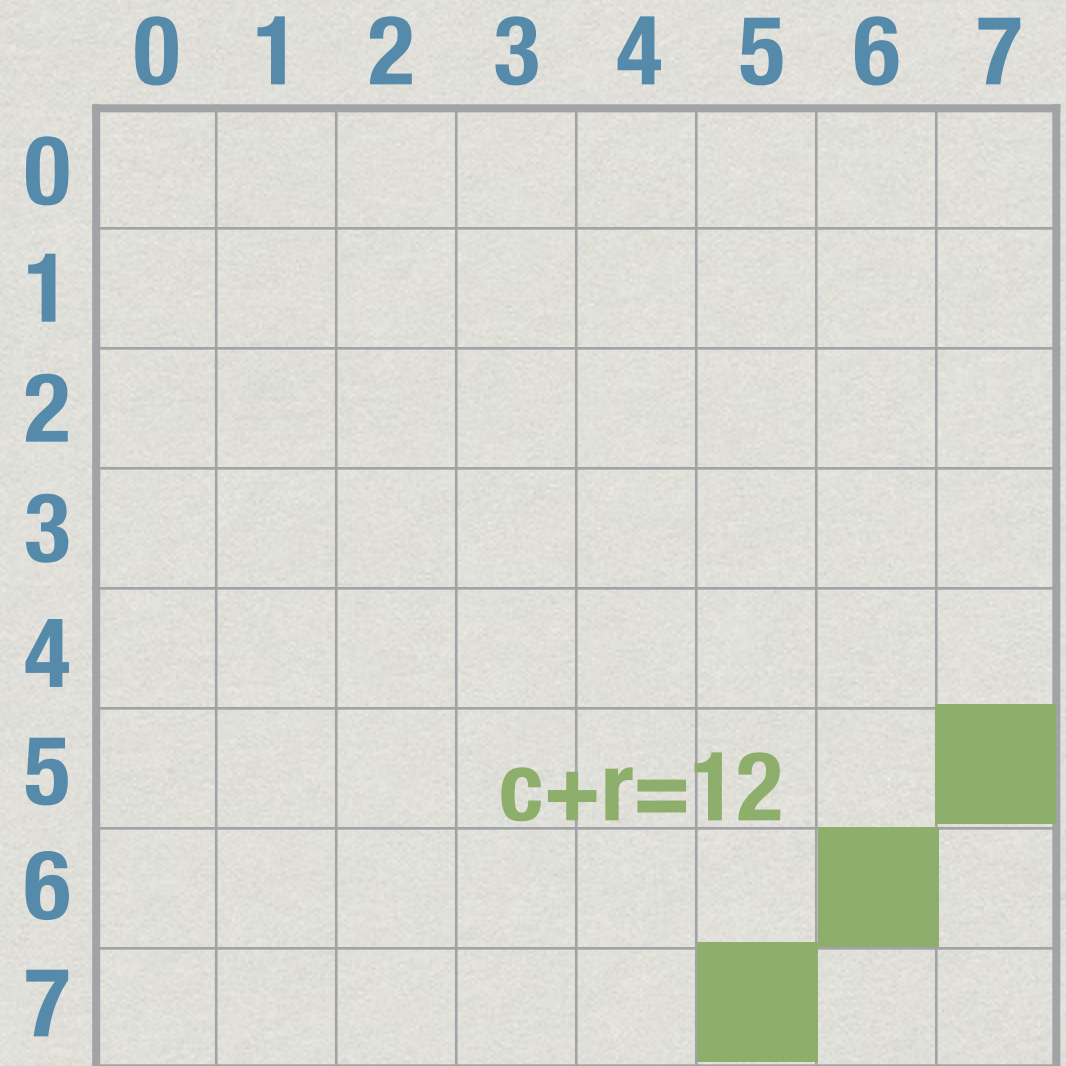
- \* Decreasing diagonal:  
column - row is invariant
- \* Increasing diagonal:  
column + row is invariant





# Numbering diagonals

- \* Decreasing diagonal:  
column - row is invariant
- \* Increasing diagonal:  
column + row is invariant





# Numbering diagonals

- \* Decreasing diagonal:  
column - row is invariant
- \* Increasing diagonal:  
column + row is invariant
- \* (i,j) is attacked if
  - \* row i is attacked
  - \* column j is attacked
  - \* diagonal  $j-i$  is attacked
  - \* diagonal  $j+i$  is attacked

	0	1	2	3	4	5	6	7
0								
1								
2								
3								
4								
5								
6								
7								

$c+r=12$



# $O(n)$ representation

- \*  $\text{row}[i] == 1$  if row  $i$  is attacked,  $0 \dots N-1$
- \*  $\text{col}[i] == 1$  if column  $i$  is attacked,  $0 \dots N-1$
- \*  $\text{NWtoSE}[i] == 1$  if NW to SE diagonal  $i$  is attacked,  $-(N-1)$  to  $(N-1)$
- \*  $\text{SWtoNE}[i] == 1$  if SW to NE diagonal  $i$  is attacked,  $0$  to  $2(N-1)$



# Updating the board

- \*  $(i, j)$  is free if  
$$\text{row}[i] == \text{col}[j] == \text{NWtoSE}[j-i] == \text{SWtoNE}[j+i] == 0$$
- \* Add queen at  $(i, j)$   
$$\begin{aligned} \text{board}[i] &= j \\ (\text{row}[i], \text{col}[j], \text{NWtoSE}[j-i], \text{SWtoNE}[j+i]) &= \\ &\quad (1, 1, 1, 1) \end{aligned}$$
- \* Remove queen at  $(i, j)$   
$$\begin{aligned} \text{board}[i] &= -1 \\ (\text{row}[i], \text{col}[j], \text{NWtoSE}[j-i], \text{SWtoNE}[j+i]) &= \\ &\quad (0, 0, 0, 0) \end{aligned}$$



# Implementation details

- \* Maintain `board` as nested dictionary
  - \* `board['queen'][i] = j` : Queen located at  $(i, j)$
  - \* `board['row'][i] = 1` : Row  $i$  attacked
  - \* `board['col'][i] = 1` : Column  $i$  attacked
  - \* `board['nwtose'][i] = 1` : NWtoSW diagonal  $i$  attacked
  - \* `board['swtone'][i] = 1` : SWtoNE diagonal  $i$  attacked



# Overall structure

```
def placequeen(i,board): # Trying row i
    for each c such that (i,c) is available:
        place queen at (i,c) and update board
        if i == n-1:
            return(True) # Last queen has been placed
        else:
            extendsoln = placequeen(i+1,board)
            if extendsoln:
                return(True) # This solution extends fully
            else:
                undo this move and update board
    else:
        return(False) # Row i failed
```



# All solutions?

```
def placequeen(i,board): # Try row i
    for each c such that (i,c) is available:
        place queen at (i,c) and update board
        if i == n-1:
            record solution # Last queen placed
        else:
            extendsoln = placequeen(i+1,board)
        undo this move and update board
```



**NPTEL MOOC**

# **PROGRAMMING, DATA STRUCTURES AND ALGORITHMS IN PYTHON**

**Week 6, Lecture 2**

**Madhavan Mukund, Chennai Mathematical Institute**

**<http://www.cmi.ac.in/~madhavan>**



# Recall 8 queens

```
def placequeen(i,board): # Trying row i
    for each c such that (i,c) is available:
        place queen at (i,c) and update board
        if i == n-1:
            return(True) # Last queen has been placed
        else:
            extendsoln = placequeen(i+1,board)
            if extendsoln:
                return(True) # This solution extends fully
            else:
                undo this move and update board
    else:
        return(False) # Row i failed
```



# Global variables

- \* Can we avoid passing `board` explicitly to each function?
- \* Can we have a single `global` copy of `board` that all functions can update?



# Scope of name

- \* Scope of name is the portion of code where it is available to read and update
- \* By default, in Python, scope is local to functions
  - \* But actually, only if we update the name inside the function



# Two examples

```
def f():  
    y = x  
    print(y)
```

```
x = 7  
f()
```

**Fine!**



# Two examples

```
def f():  
    y = x  
    print(y)
```

```
x = 7  
f()
```

**Fine!**

```
def f():  
    y = x  
    print(y)  
    x = 22
```

```
x = 7  
f()
```

**Error!**



# Two examples

```
def f():  
    y = x  
    print(y)
```

```
x = 7  
f()
```

**Fine!**

```
def f():  
    y = x  
    print(y)  
    x = 22
```

```
x = 7  
f()
```

**Error!**

- \* If `x` is not found in `f()`, Python looks at enclosing function for **global** `x`
- \* If `x` is updated in `f()`, it becomes a **local** name!



# Global variables

- \* Actually, this applies only to immutable values
- \* Global names that point to mutable values can be updated within a function

```
def f():  
    y = x[0]  
    print(y)  
    x[0] = 22
```

```
x = [7]  
f()
```

**Fine!**



# Global immutable values

- \* What if we want a global integer
- \* Count the number of times a function is called
- \* Declare a name to be `global`

```
def f():  
    global x  
    y = x  
    print(y)  
    x = 22
```

```
x = 7  
f()  
print(x)
```



# Global immutable values

- \* What if we want a global integer
- \* Count the number of times a function is called
- \* Declare a name to be `global`

```
def f():  
    global x  
    y = x  
    print(y)  
    x = 22
```

```
x = 7  
f()  
print(x)
```

**22**



# Nest function definitions

- \* Can define local “helper” functions
- \* `g()` and `h()` are only visible to `f()`
- \* Cannot be called directly from outside

```
def f():  
    def g(a):  
        return(a+1)  
  
    def h(b):  
        return(2*b)  
  
    global x  
    y = g(x) + h(x)  
    print(y)  
    x = 22  
  
x = 7  
f()
```



# Nest function definitions

- \* If we look up `x`, `y` inside `g()` or `h()` it will first look in `f()`, then outside
- \* Can also declare names global inside `g()`, `h()`
- \* Intermediate scope declaration: `nonlocal`
- \* See Python documentation

```
def f():  
    def g(a):  
        return(a+1)  
  
    def h(b):  
        return(2*b)  
  
    global x  
    y = g(x) + h(x)  
    print(y)  
    x = 22  
  
x = 7  
f()
```



# Summary

- \* Python names are looked up inside-out from within functions
- \* Updating a name with immutable value creates a local copy of that name
  - \* Can update global names with mutable values
- \* Use `global` definition to update immutable values
- \* Can nest helper function — hidden to the outside



**NPTEL MOOC**

# **PROGRAMMING, DATA STRUCTURES AND ALGORITHMS IN PYTHON**

**Week 6, Lecture 3**

**Madhavan Mukund, Chennai Mathematical Institute**

**<http://www.cmi.ac.in/~madhavan>**



# Backtracking

- \* Systematically search for a solution
- \* Build the solution one step at a time
- \* If we hit a dead-end
  - \* Undo the last step
  - \* Try the next option



# Generating permutations

- \* Often useful when we need to try out all possibilities
  - \* Each potential columnwise placement of N queens is a permutation of  $\{0, 1, \dots, N-1\}$
- \* Given a permutation, generate the next one
- \* For instance, what is the next sequence formed from  $\{a, b, \dots, m\}$  , in dictionary order after

d c h b a e g l k o n m j i



# Generating permutations

- \* Smallest permutation — all elements in ascending order

a b c d e f g h i j k l m

- \* Largest permutation — all elements in descending order

m l k j i h g f e d c b a

- \* Next permutation — find shortest suffix that can be incremented

- \* Or longest suffix that cannot be incremented



# Next permutation

- \* Longest suffix that cannot be incremented
- \* Already in descending order

d c h b a e g l k o n m j i



# Next permutation

- \* Longest suffix that cannot be incremented
  - \* Already in descending order

d c h b a e g l k o n m j i

- \* The suffix starting one position earlier can be incremented



# Next permutation

- \* Longest suffix that cannot be incremented
  - \* Already in descending order

d c h b a e g l k o n m j i

- \* The suffix starting one position earlier can be incremented
  - \* Replace **k** by next largest letter to its right, **m**
  - \* Rearrange **k o n j i** in ascending order

d c h b a e g l m i j k n o



# Implementation

- \* From the right, identify first decreasing position

d c h b a e g l **k** o n m j i



- \* Swap that value with its next larger letter to its right

d c h b a e g l **m** o n **k** j i



- \* Finding next larger letter is similar to insert
- \* Reverse the increasing suffix

d c h b a e g l m **i j k n o**



**NPTEL MOOC**

# **PROGRAMMING, DATA STRUCTURES AND ALGORITHMS IN PYTHON**

**Week 6, Lecture 4**

**Madhavan Mukund, Chennai Mathematical Institute**

**<http://www.cmi.ac.in/~madhavan>**



# Data structures

- \* *Algorithms + Data Structures = Programs*  
Niklaus Wirth
- \* Arrays/lists — sequences of values
- \* Dictionaries — key-value pairs
- \* Python also has sets as a built in datatype



# Sets in Python

- \* List with braces, duplicates automatically removed

```
colours = {'red', 'black', 'red', 'green'}
```

```
>>> print(colours)
{'black', 'red', 'green'}
```

- \* Create an empty set

```
colours = set()
```

- \* Note, not `colours = {}` — empty dictionary!



# Sets in Python

- \* Set membership

```
>>> 'black' in colours  
True
```

- \* Convert a list into a set

```
>>> numbers = set([0,1,3,2,1,4])  
>>> print(numbers)  
{0, 1, 2, 3, 4}
```

```
>>> letters = set('banana')  
>>> print(letters)  
{ 'a', 'n', 'b' }
```



# Set operations

```
odd = set([1,3,5,7,9,11])  
prime = set([2,3,5,7,11])
```

- \* Union

```
odd | prime → {1, 2, 3, 5, 7, 9, 11}
```

- \* Intersection

```
odd & prime → {3, 11, 5, 7}
```

- \* Set difference

```
odd - prime → {1, 9}
```

- \* Exclusive or

```
odd ^ prime → {1, 2, 9}
```



# Stacks

- \* Stack is a last-in, first-out list
  - \* `push(s,x)` — add `x` to stack `s`
  - \* `pop(s)` — return most recently added element
- \* Maintain stack as list, push and pop from the right
  - \* `push(s,x)` is `s.append(x)`
  - \* `s.pop()` — Python built-in, returns last element



# Stacks

- \* Stacks are natural to keep track of recursive function calls
- \* In 8 queens, use a stack to keep track of queens added
  - \* Push the latest queen onto the stack
  - \* To backtrack, pop the last queen added



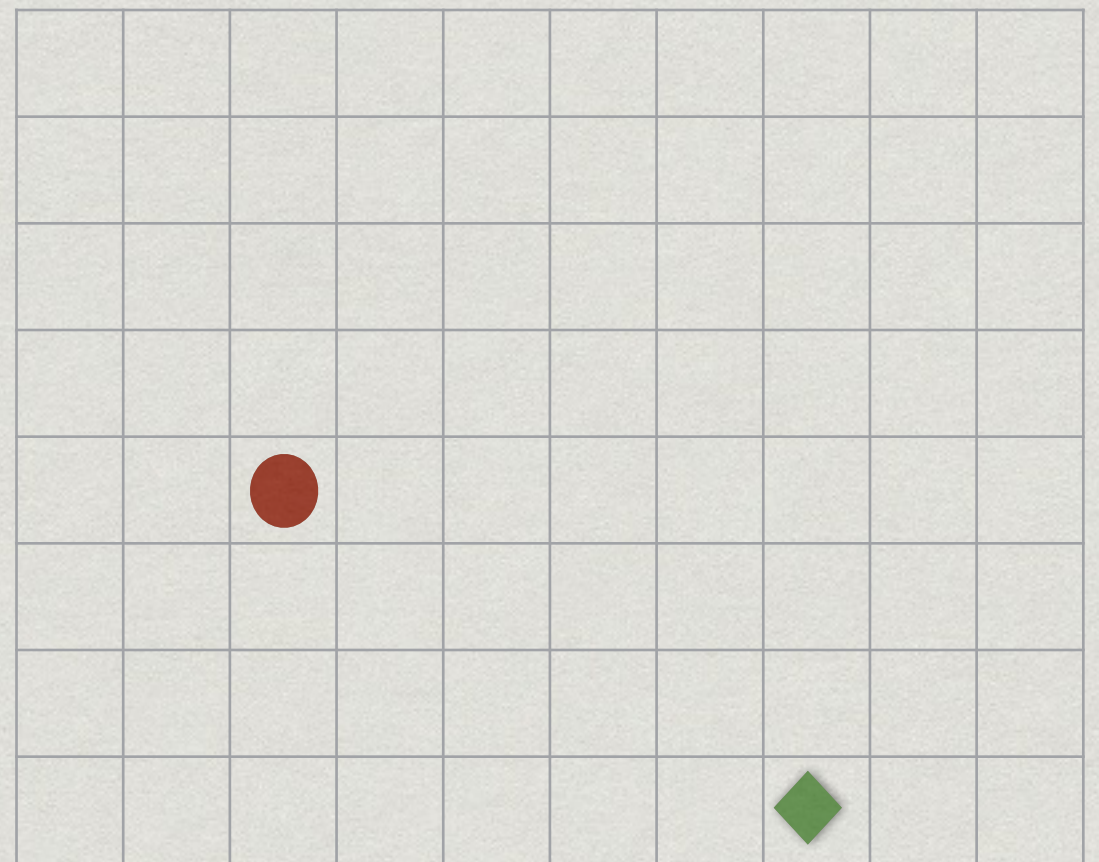
# Queues

- \* First-in, first-out sequences
  - \* `addq(q, x)` — adds `x` to rear of queue `q`
  - \* `removeq(q)` — removes element at head of `q`
- \* Using Python lists, left is rear, right is front
  - \* `addq(q, x)` is `q.insert(0, x)`
    - \* `l.insert(j, x)`, insert `x` before position `j`
  - \* `removeq(q)` is `q.pop()`




# Systematic exploration

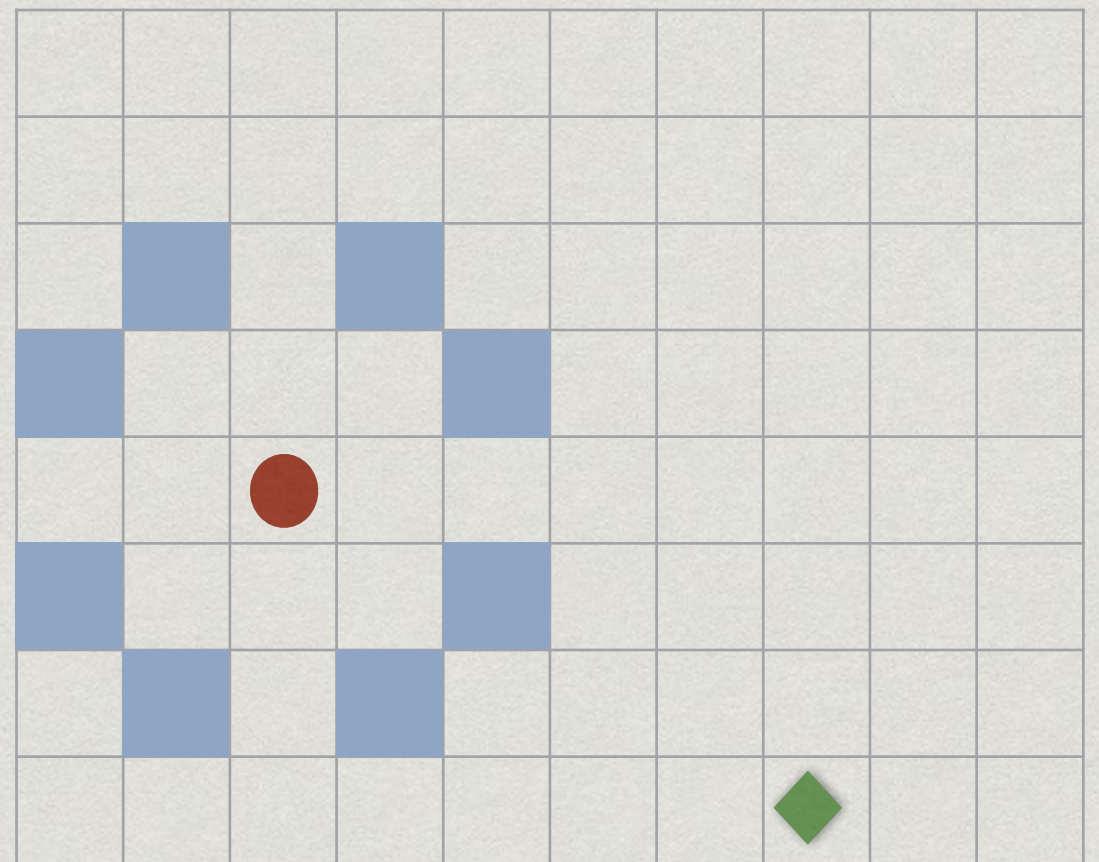
- \* Rectangular  $m \times n$  grid
- \* Chess knight starts at  $(sx, sy)$ 
  - \* Usual knight moves
- \* Can it reach a target square  $(tx, ty)$ ? ◆





# Systematic exploration

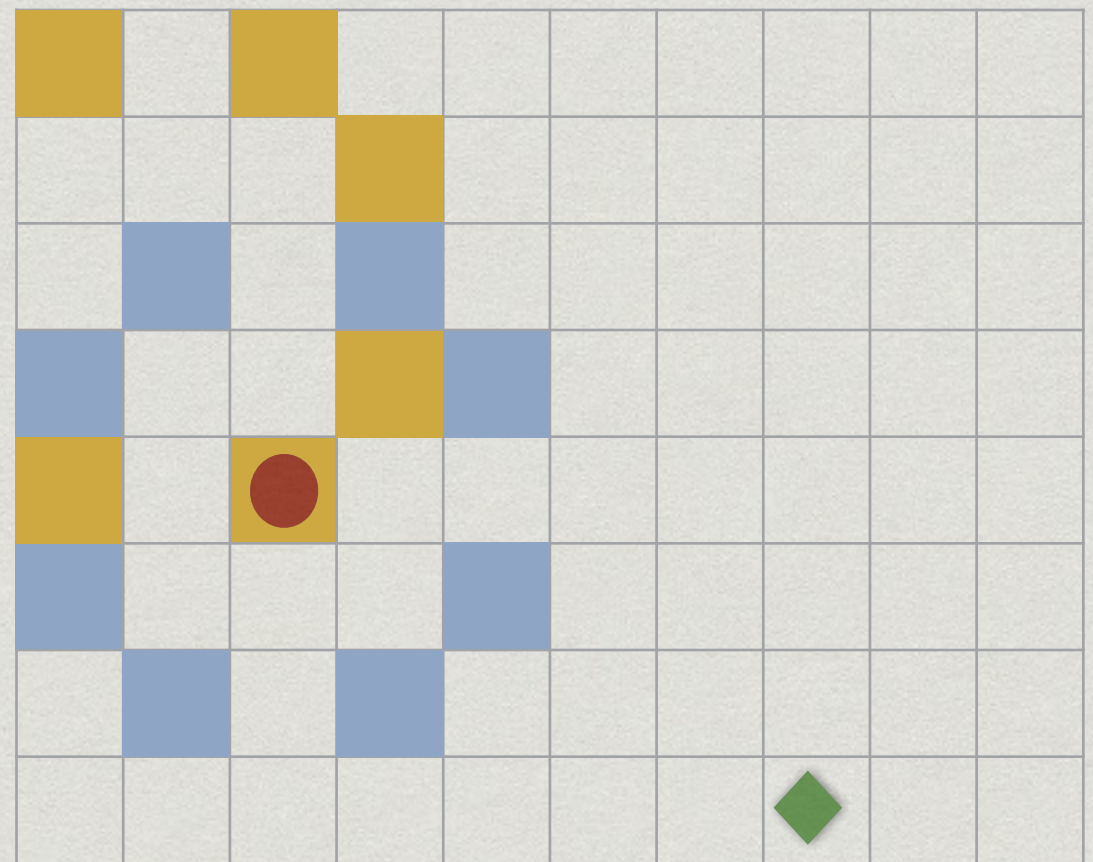
- \* Rectangular  $m \times n$  grid
- \* Chess knight starts at  $(sx, sy)$
- \* Usual knight moves
- \* Can it reach a target square  $(tx, ty)$ ? 





# Systematic exploration

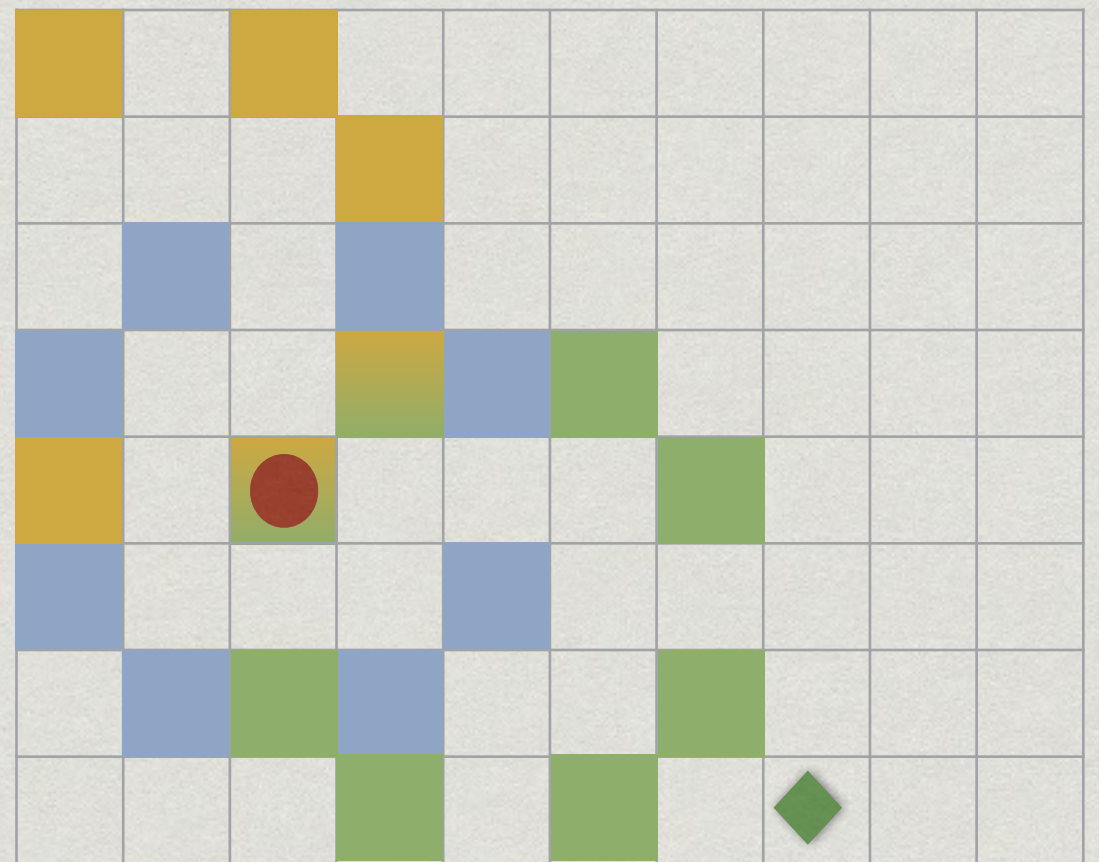
- \* Rectangular  $m \times n$  grid
- \* Chess knight starts at  $(sx, sy)$
- \* Usual knight moves
- \* Can it reach a target square  $(tx, ty)$ ? ◆





# Systematic exploration

- \* Rectangular  $m \times n$  grid
- \* Chess knight starts at  $(sx, sy)$
- \* Usual knight moves
- \* Can it reach a target square  $(tx, ty)$ ? ◆





# Systematic exploration

- \* X1 — all squares reachable in one move from (sx,sy)
- \* X2 — all squares reachable from X1 in one move
- \* . . .
- \* Don't explore an already marked square
- \* When do we stop?
  - \* If we reach target square
  - \* What if target is not reachable?



# Systematic exploration

- \* Maintain a queue  $Q$  of cells to be explored
- \* Initially  $Q$  contains only start node  $(sx, sy)$ 
  - \* Remove  $(ax, ay)$  from head of queue
  - \* Mark all squares reachable in one step from  $(ax, ay)$
  - \* Add all newly marked squares to the queue
- \* When the queue is empty, we have finished



# Systematic exploration

```
def explore((sx,sy),(tx,ty)):
    marked = [[0 for i in range(n)]
               for j in range(m)]
    marked[sx][sy] = 1
    queue = [(sx,sy)]
    while queue != []:
        (ax,ay) = queue.pop()
        for (nx,ny) in neighbours((ax,ay)):
            if !marked[nx][ny]:
                marked[nx][ny] = 1
                queue.insert(0,(nx,ny))
    return(marked[tx][ty])
```



# Example

**src = (0,1)**      **tgt = (1,1)**



- \* This is an example of **breadth first search**



# Summary

- \* Data structures are ways of organising information that allow efficient processing in certain contexts
- \* Python has a built-in implementation of sets
- \* Stacks are useful to keep track of recursive computations
- \* Queues are useful for breadth-first exploration



**NPTEL MOOC**

# **PROGRAMMING, DATA STRUCTURES AND ALGORITHMS IN PYTHON**

**Week 6, Lecture 5**

**Madhavan Mukund, Chennai Mathematical Institute**

**<http://www.cmi.ac.in/~madhavan>**



# Job scheduler

- \* A job scheduler maintains a list of pending jobs with their priorities.
- \* When the processor is free, the scheduler picks out the job with maximum priority in the list and schedules it.
- \* New jobs may join the list at any time.
- \* How should the scheduler maintain the list of pending jobs and their priorities?



# Priority queue

- \* Need to maintain a list of jobs with priorities to optimise the following operations
  - \* `delete_max()`
    - \* Identify and remove job with highest priority
    - \* Need not be unique
  - \* `insert()`
    - \* Add a new job to the list



# Linear structures

- \* Unsorted list

- \* `insert()` takes  $O(1)$  time

- \* `delete_max()` takes  $O(n)$  time

- \* Sorted list

- \* `delete_max()` takes  $O(1)$  time

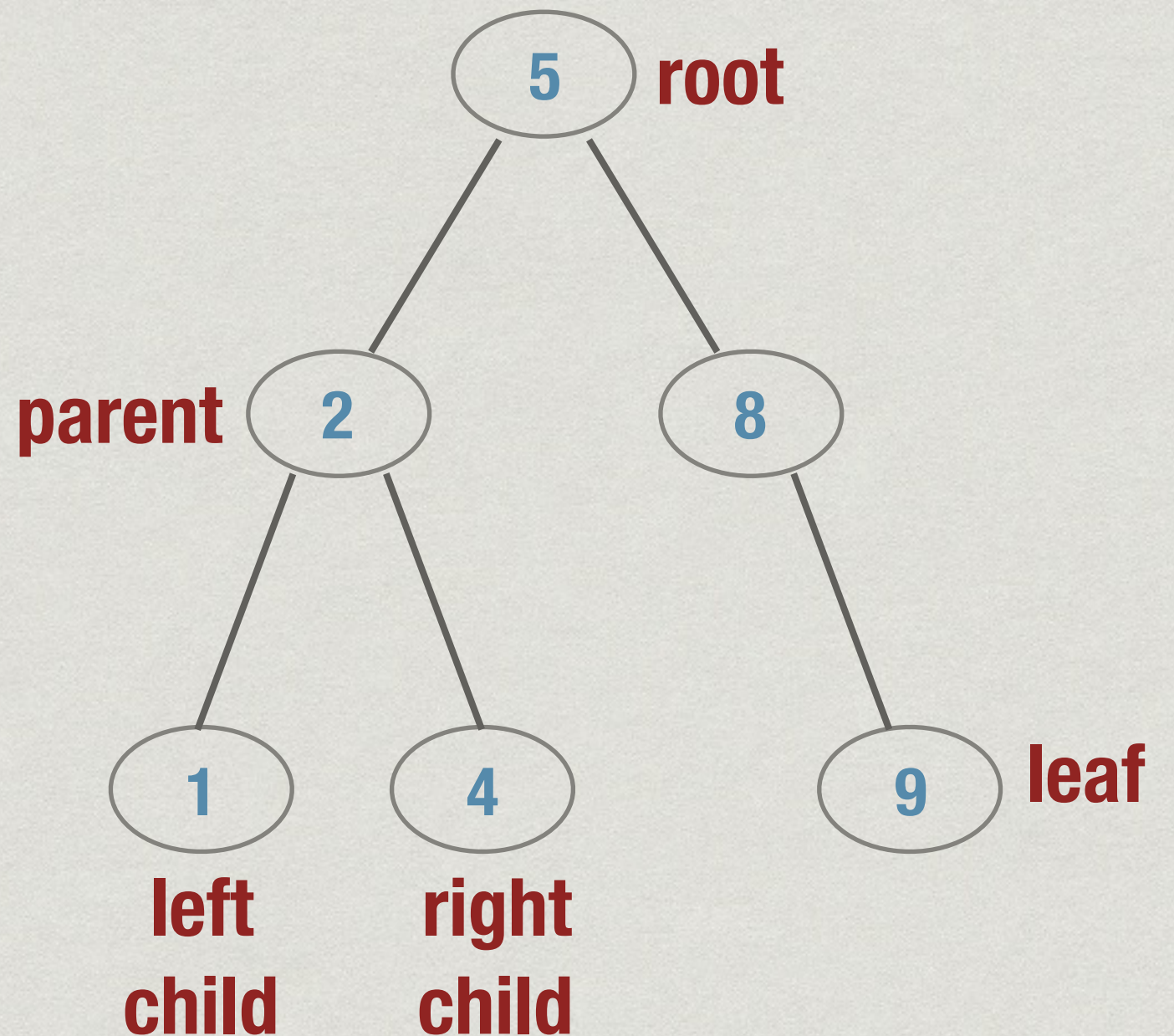
- \* `insert()` takes  $O(n)$  time

- \* Processing a sequence of  $n$  jobs requires  $O(n^2)$  time



# Binary tree

- \* Two dimensional structure
- \* At each node
  - \* Value
  - \* Link to parent, left child, right child





# Priority queues as trees

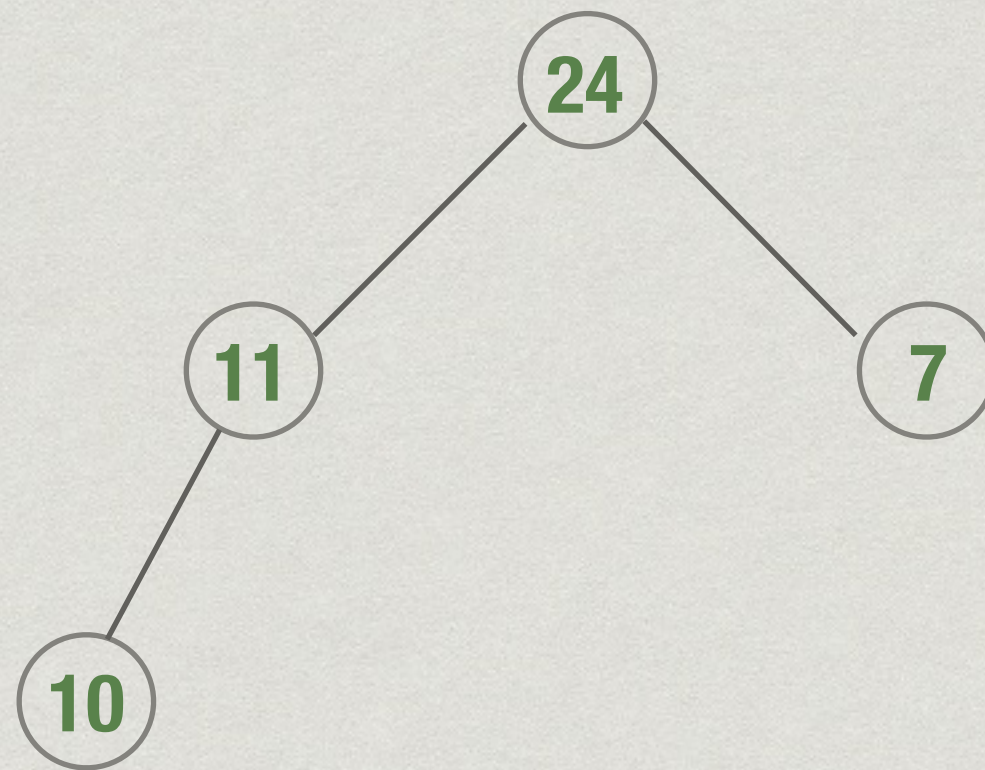
- \* Maintain a special kind of binary tree called a **heap**
  - \* **Balanced**:  $N$  node tree has height  $\log N$
- \* Both `insert()` and `delete_max()` take  $O(\log N)$ 
  - \* Processing  $N$  jobs takes time  $O(N \log N)$
- \* Truly flexible, need not fix upper bound for  $N$  in advance





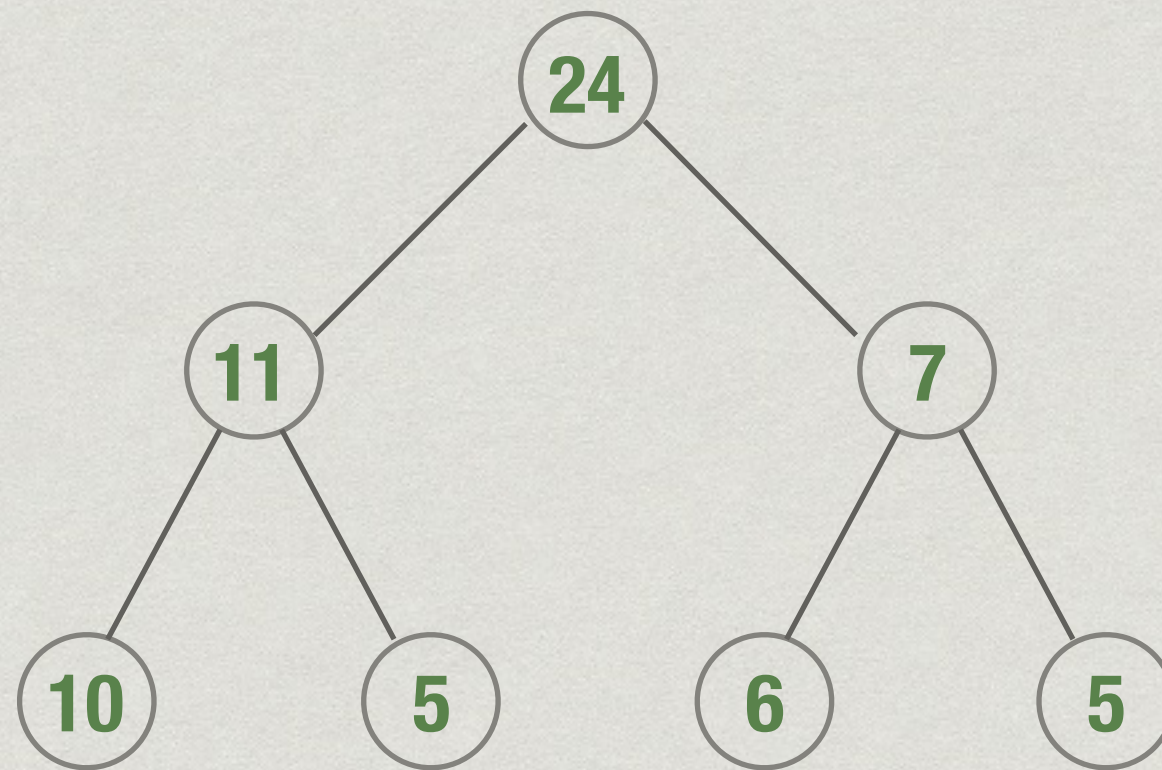


# Examples





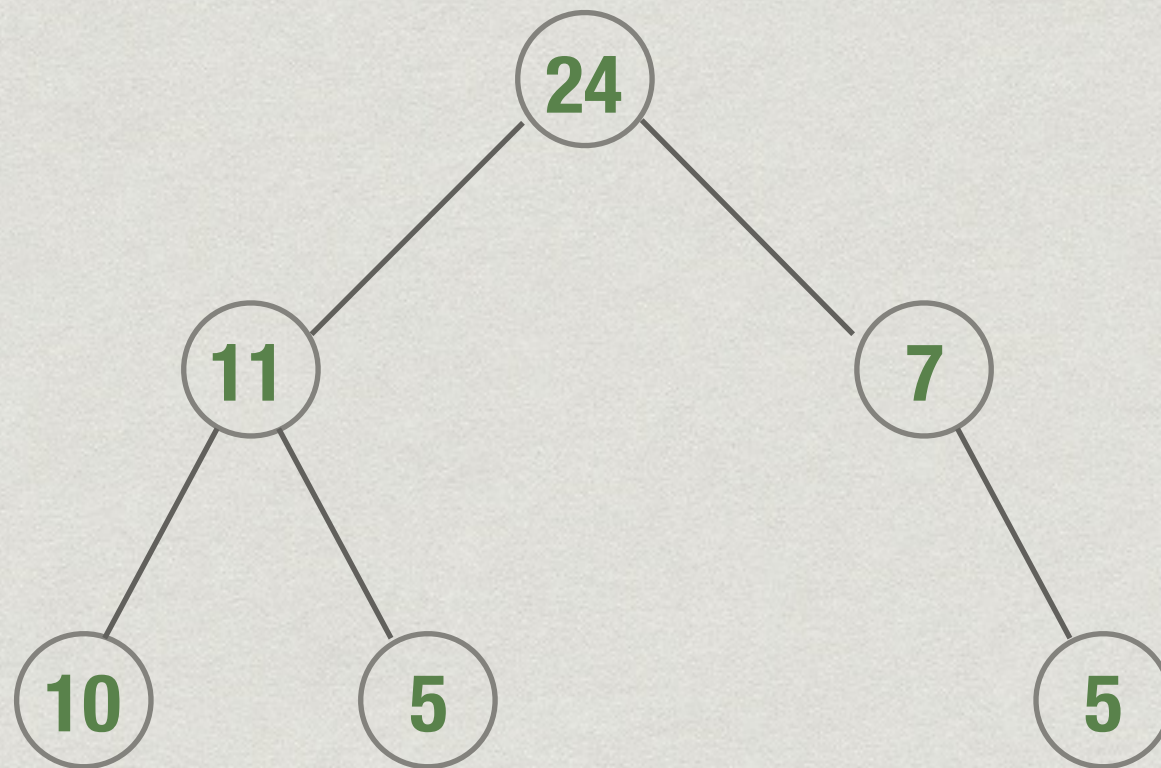
# Examples





# Non-examples

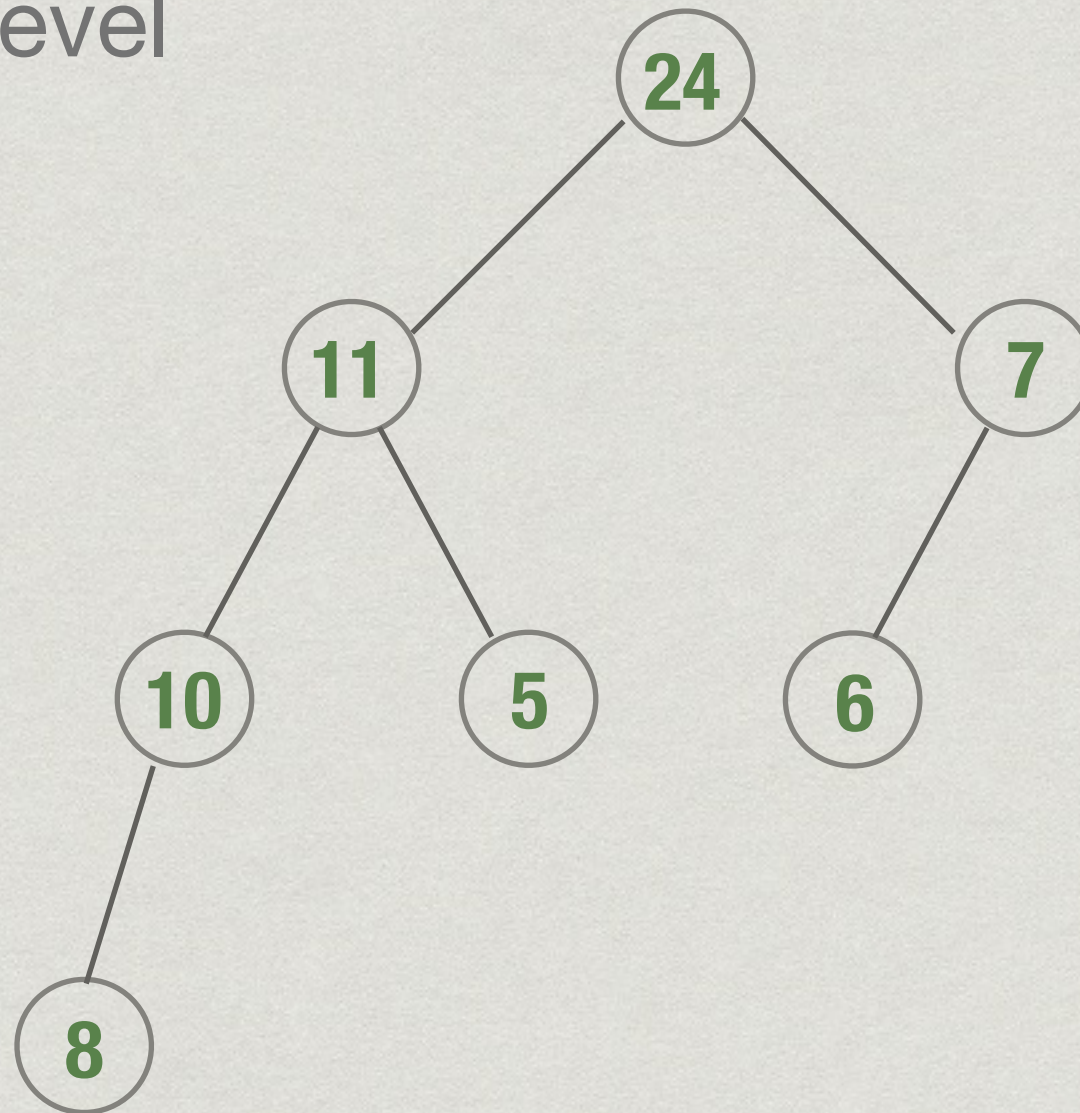
- \* No “holes” allowed





# Non-examples

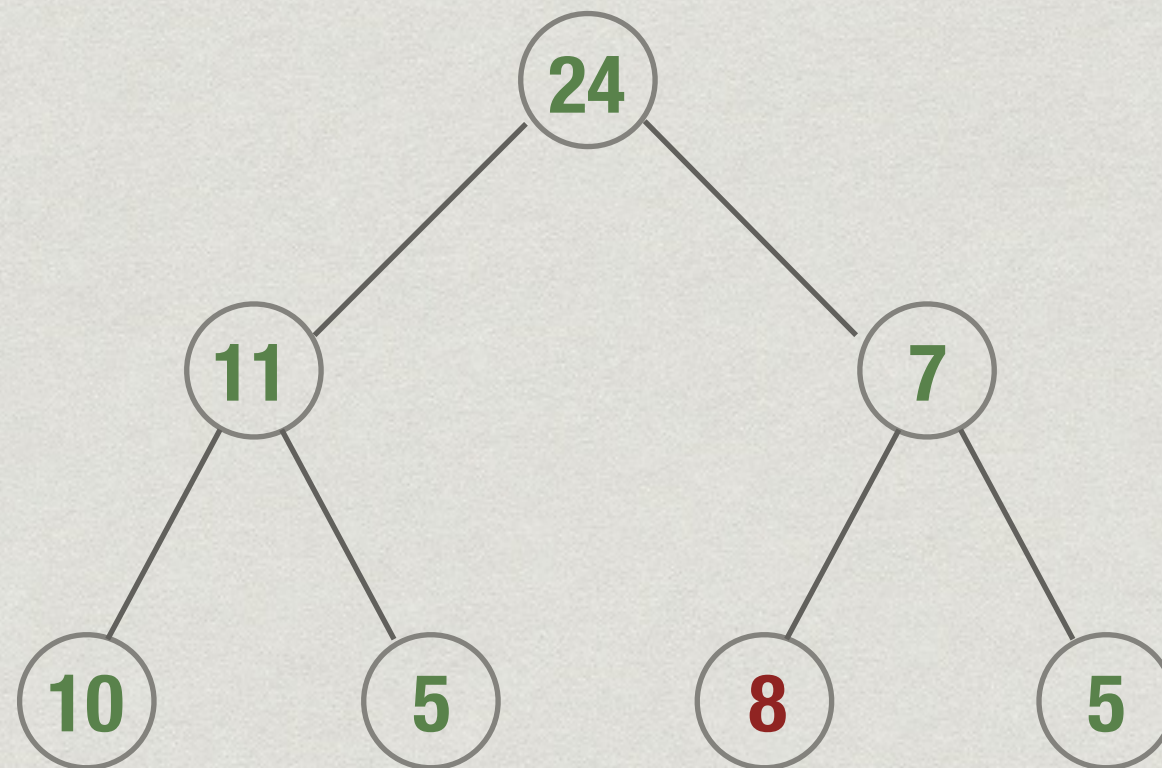
- \* Can't leave a level incomplete





# Non-examples

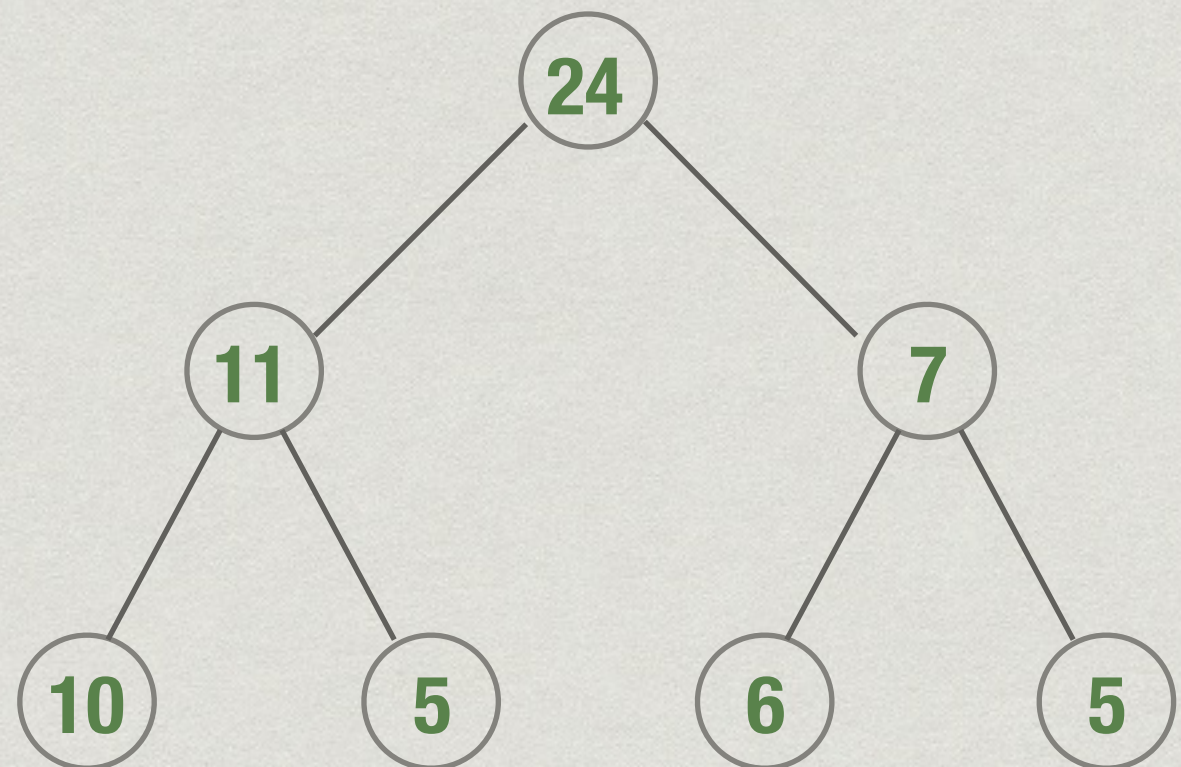
- \* Violates heap property





# insert()

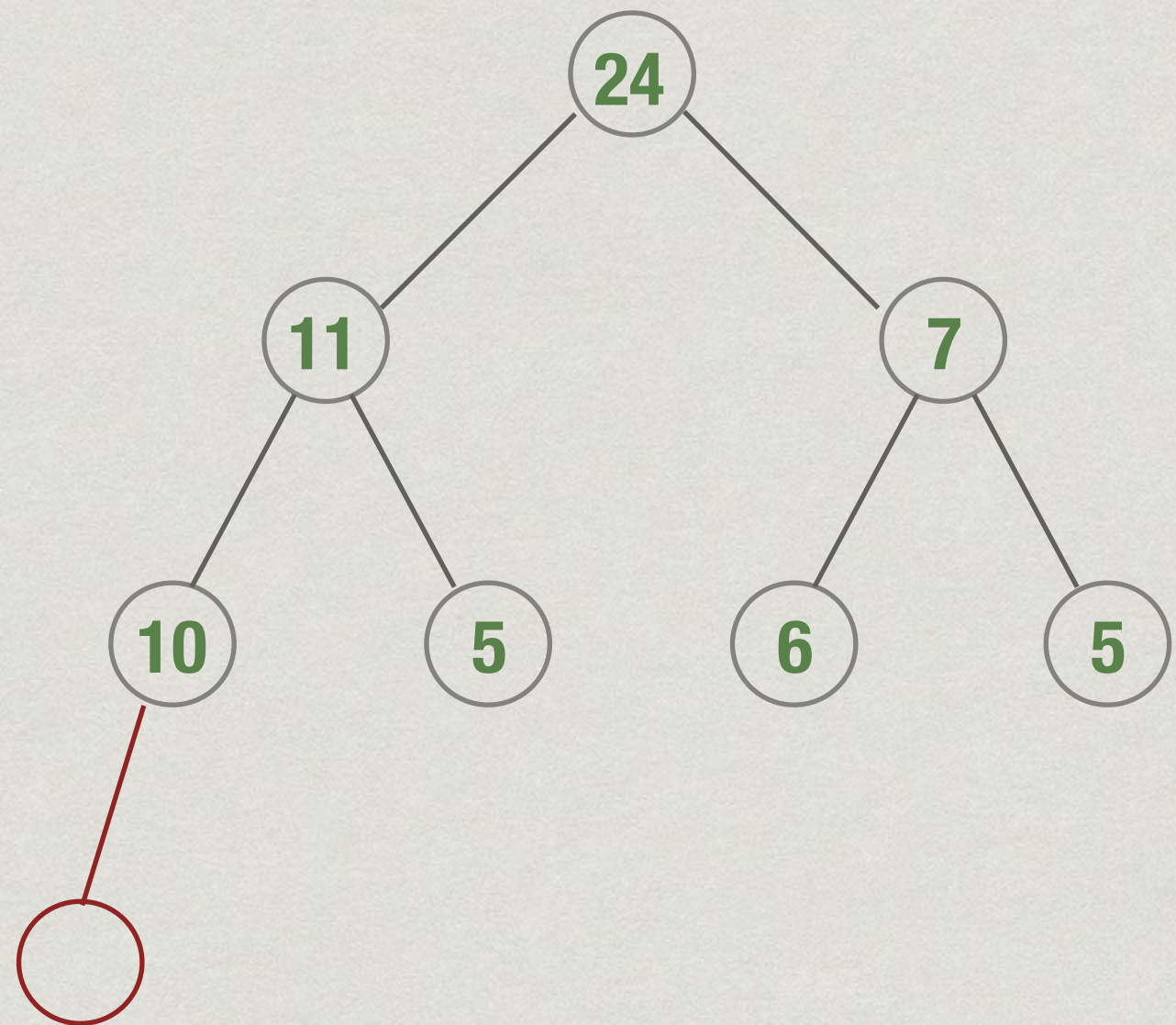
- \* insert 12
- \* Position of new node is fixed by structure
- \* Restore heap property along the path to the root





# insert()

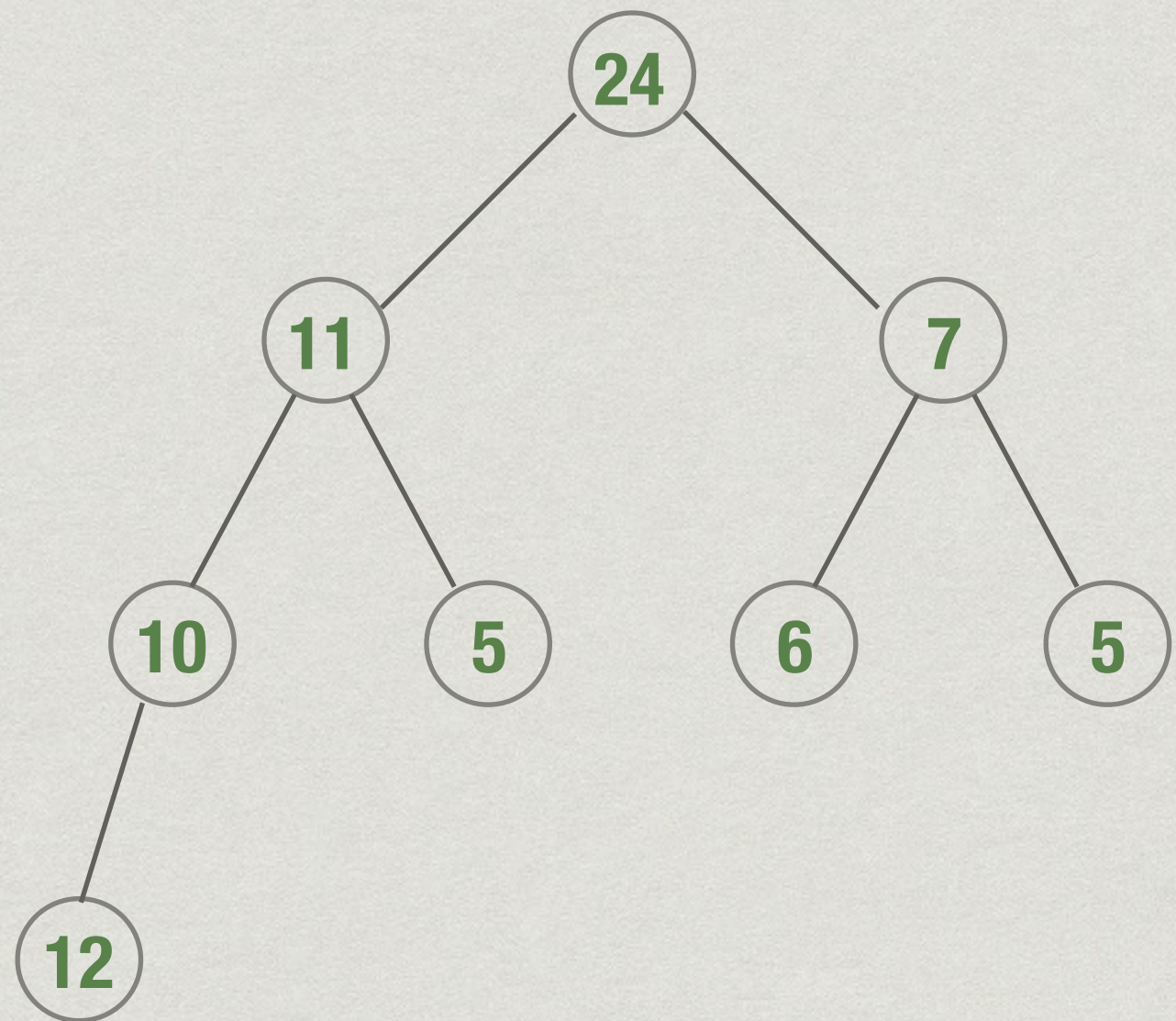
- \* insert 12
- \* Position of new node is fixed by structure
- \* Restore heap property along the path to the root





# insert()

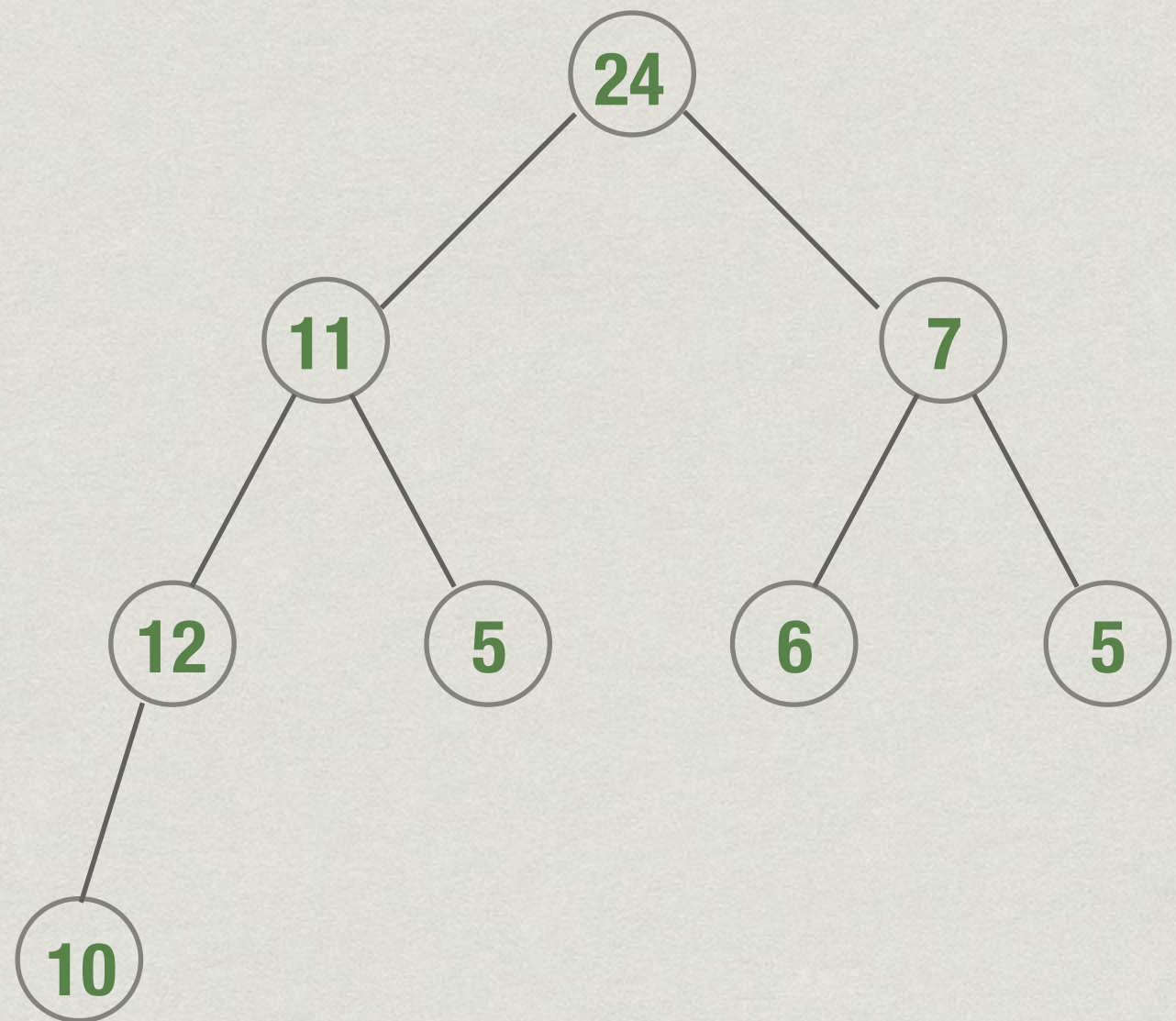
- \* insert 12
- \* Position of new node is fixed by structure
- \* Restore heap property along the path to the root





# insert()

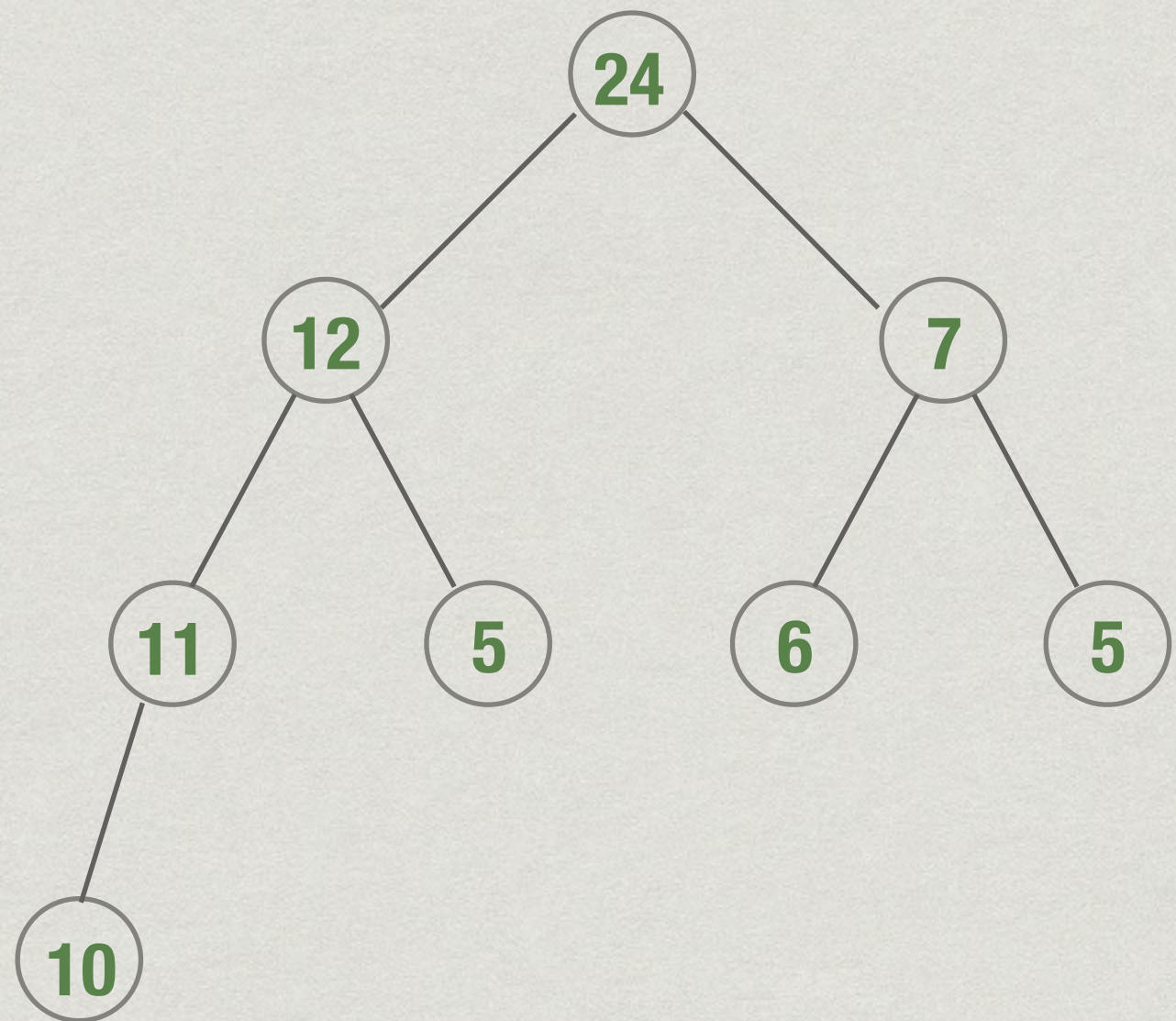
- \* insert 12
- \* Position of new node is fixed by structure
- \* Restore heap property along the path to the root





# insert()

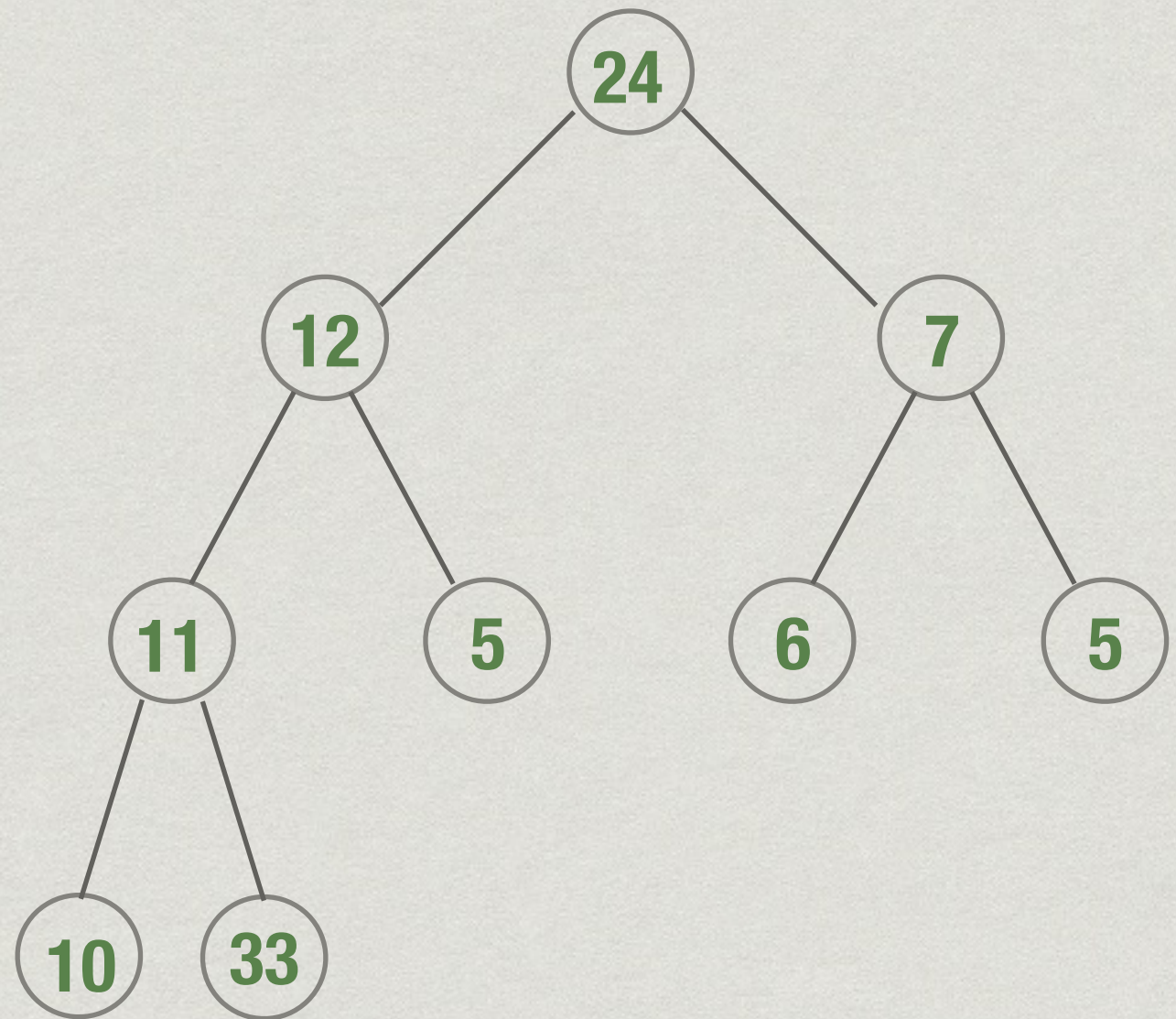
- \* insert 12
- \* Position of new node is fixed by structure
- \* Restore heap property along the path to the root





# insert()

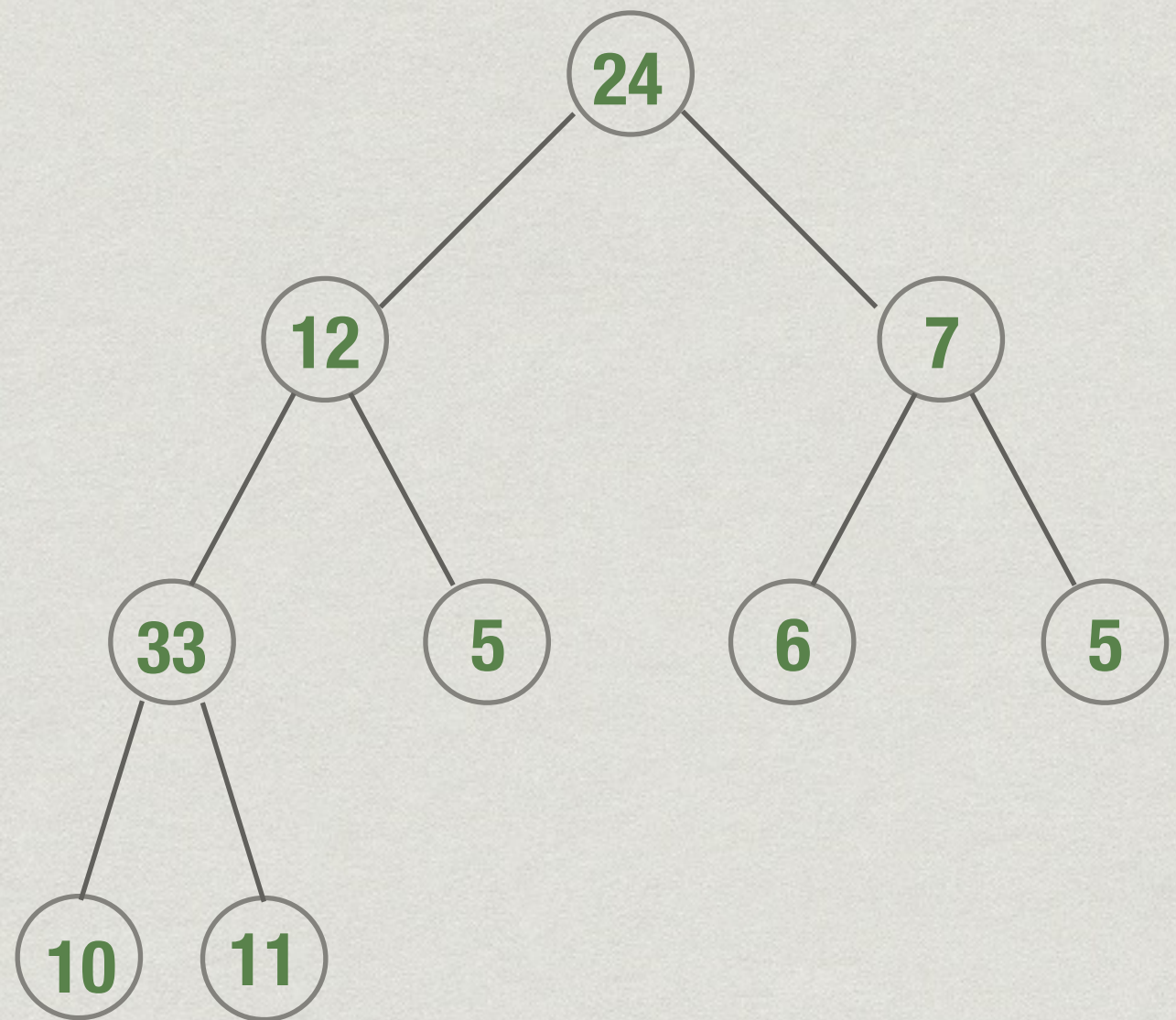
\* insert 33





# insert()

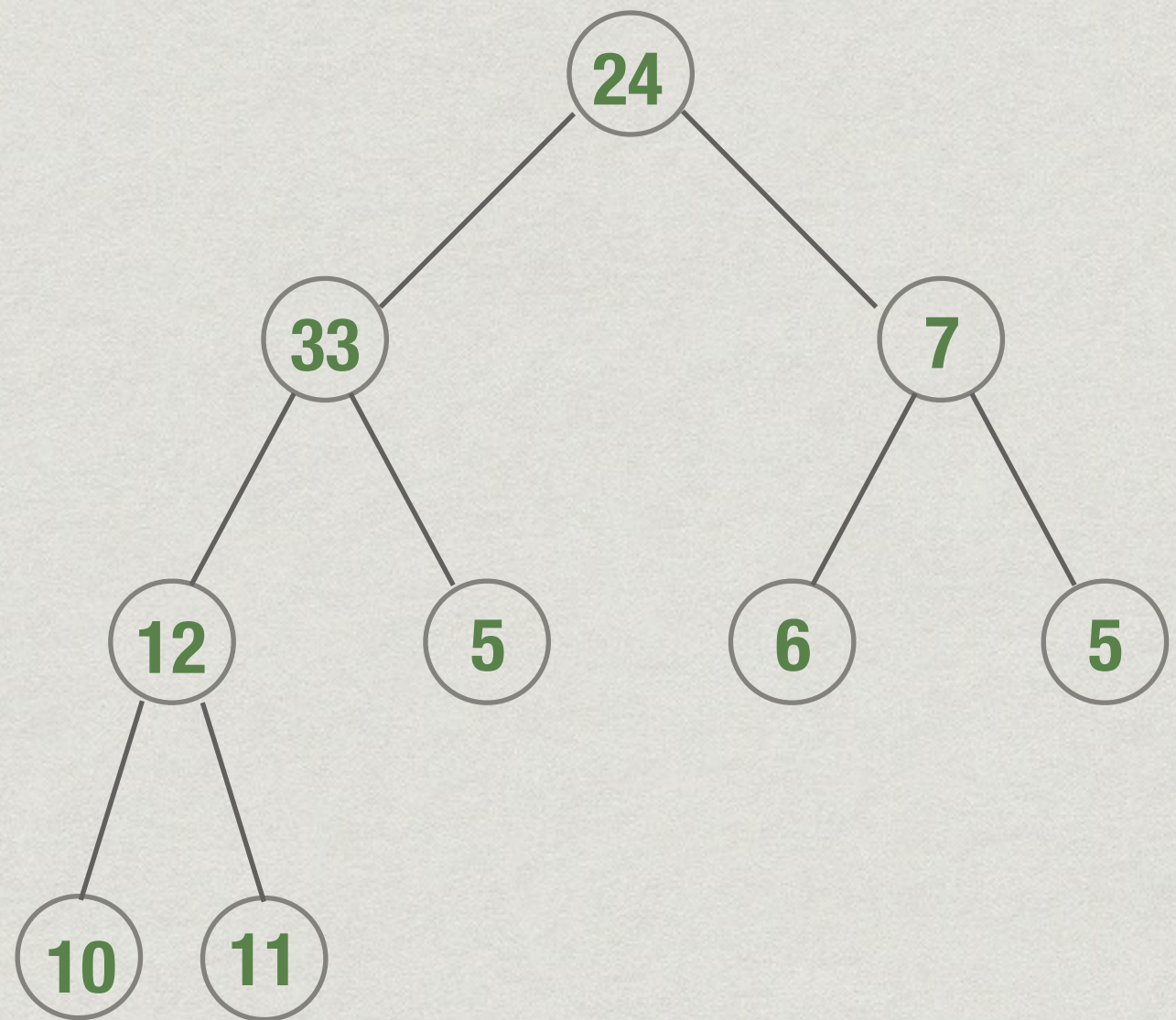
\* insert 33





# insert()

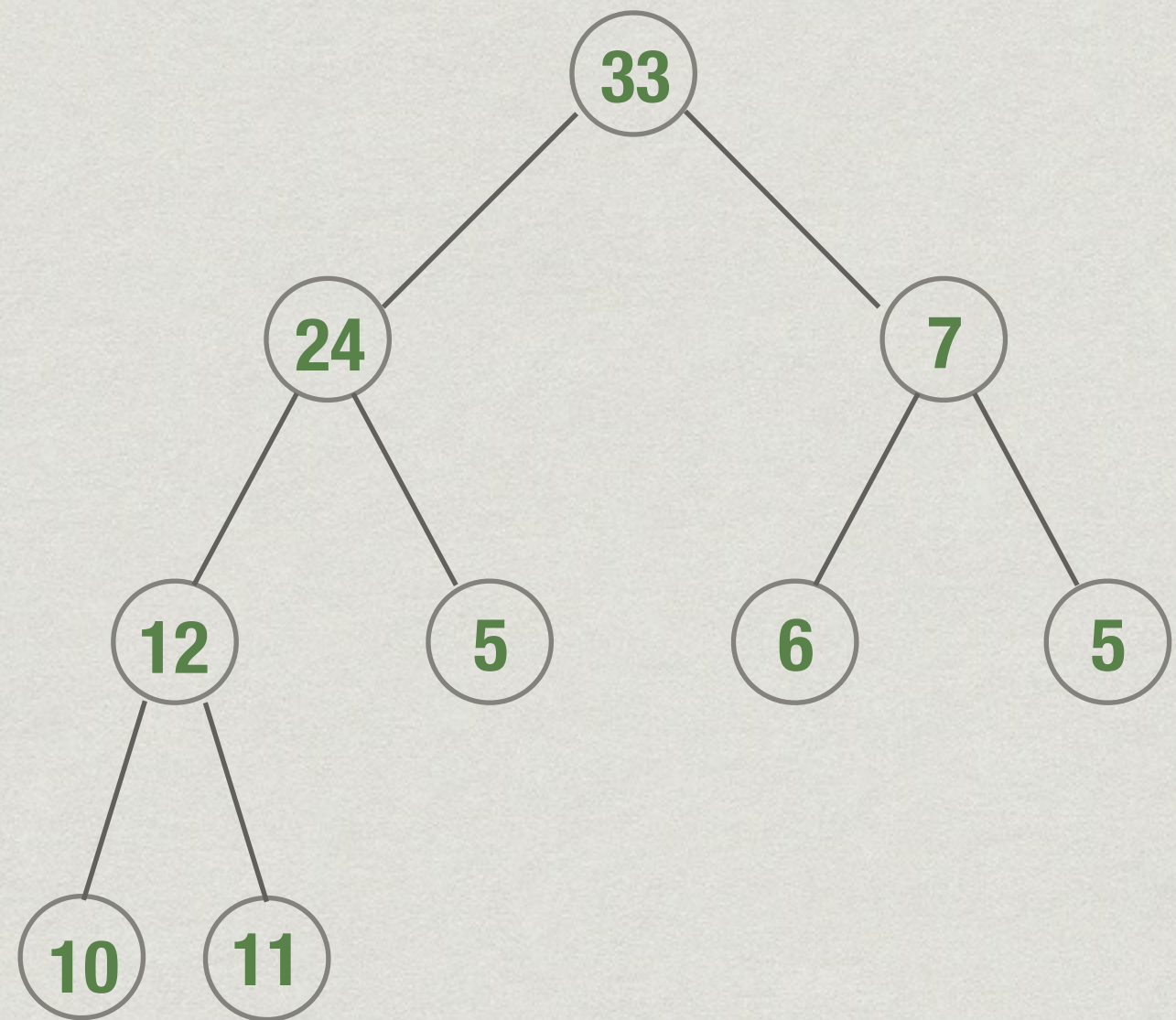
\* insert 33





# insert()

\* insert 33





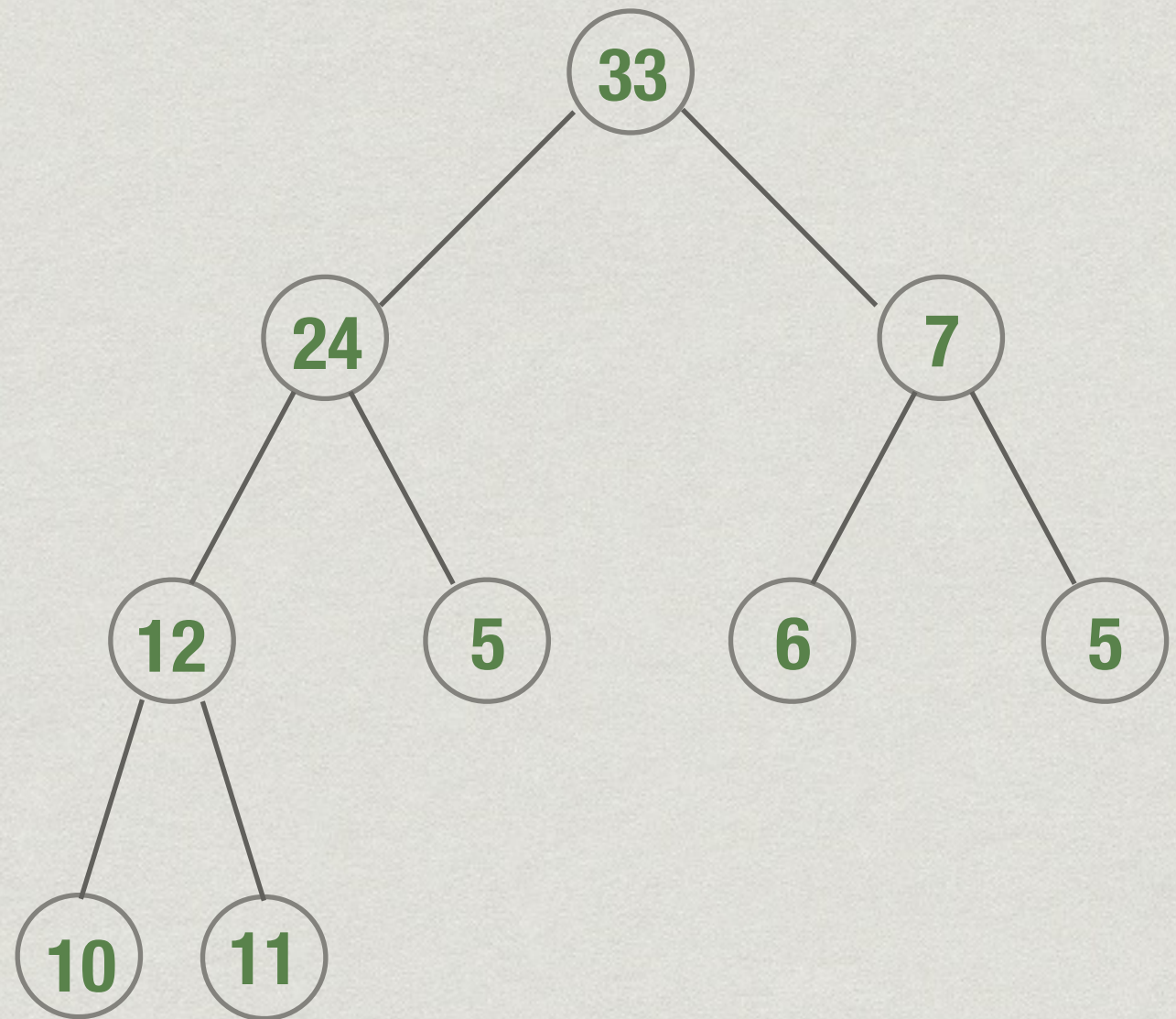
# Complexity of insert()

- \* Need to walk up from the leaf to the root
  - \* Height of the tree
- \* Number of nodes at level 0,1,...,i is  $2^0, 2^1, \dots, 2^i$
- \* K levels filled :  $2^0 + 2^1 + \dots + 2^{k-1} = 2^k - 1$  nodes
- \* N nodes : number of levels at most  $\log N + 1$
- \* insert() takes time  $O(\log N)$



# delete\_max()

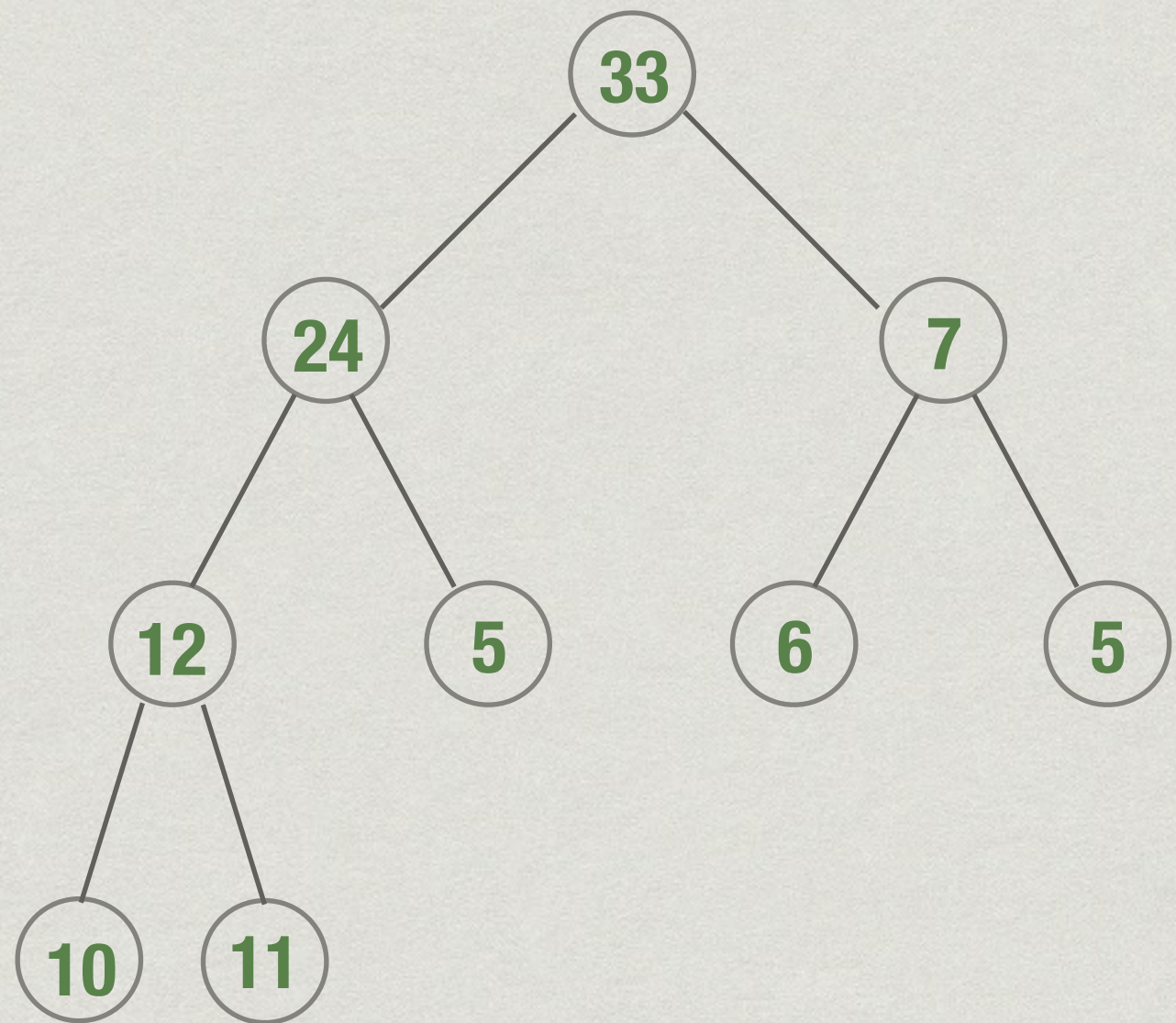
- \* Maximum value is always at the root
- \* From heap property, by induction
- \* How do we remove this value efficiently?





# delete\_max()

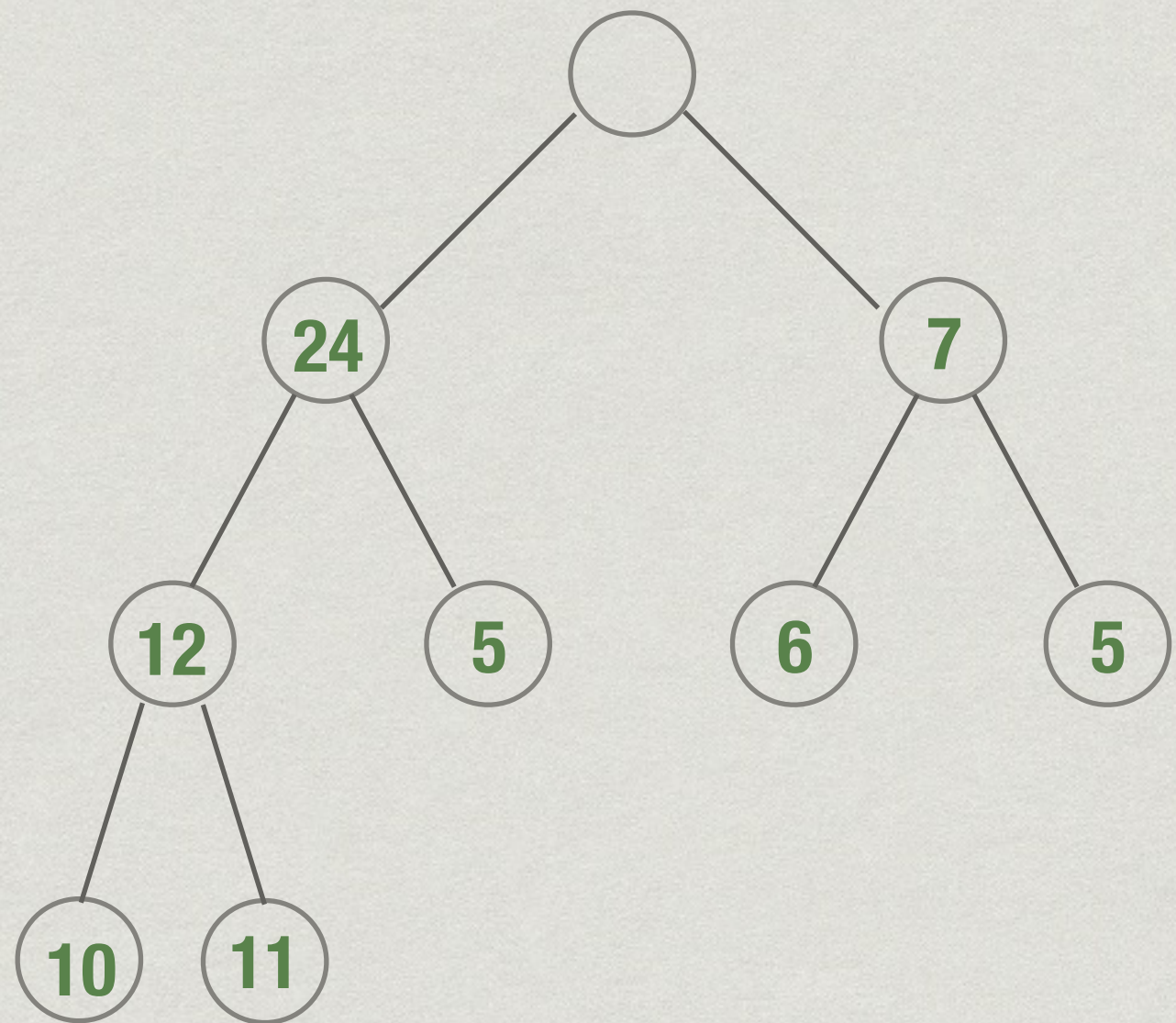
- \* Removing maximum value creates a “hole” at the root
- \* Reducing one value requires deleting last node
- \* Move “homeless” value to root





# delete\_max()

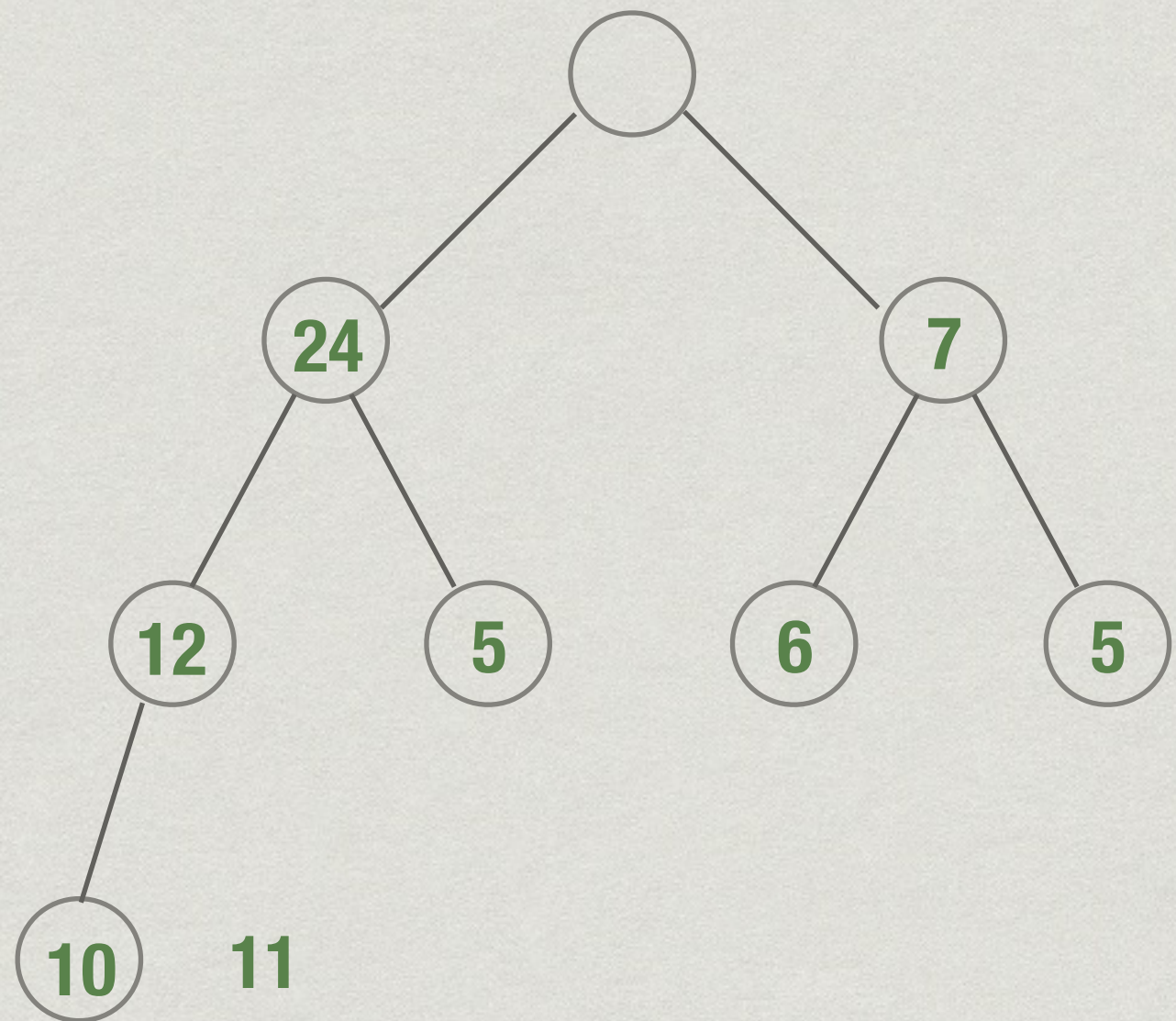
- \* Removing maximum value creates a “hole” at the root
- \* Reducing one value requires deleting last node
- \* Move “homeless” value to root





# delete\_max()

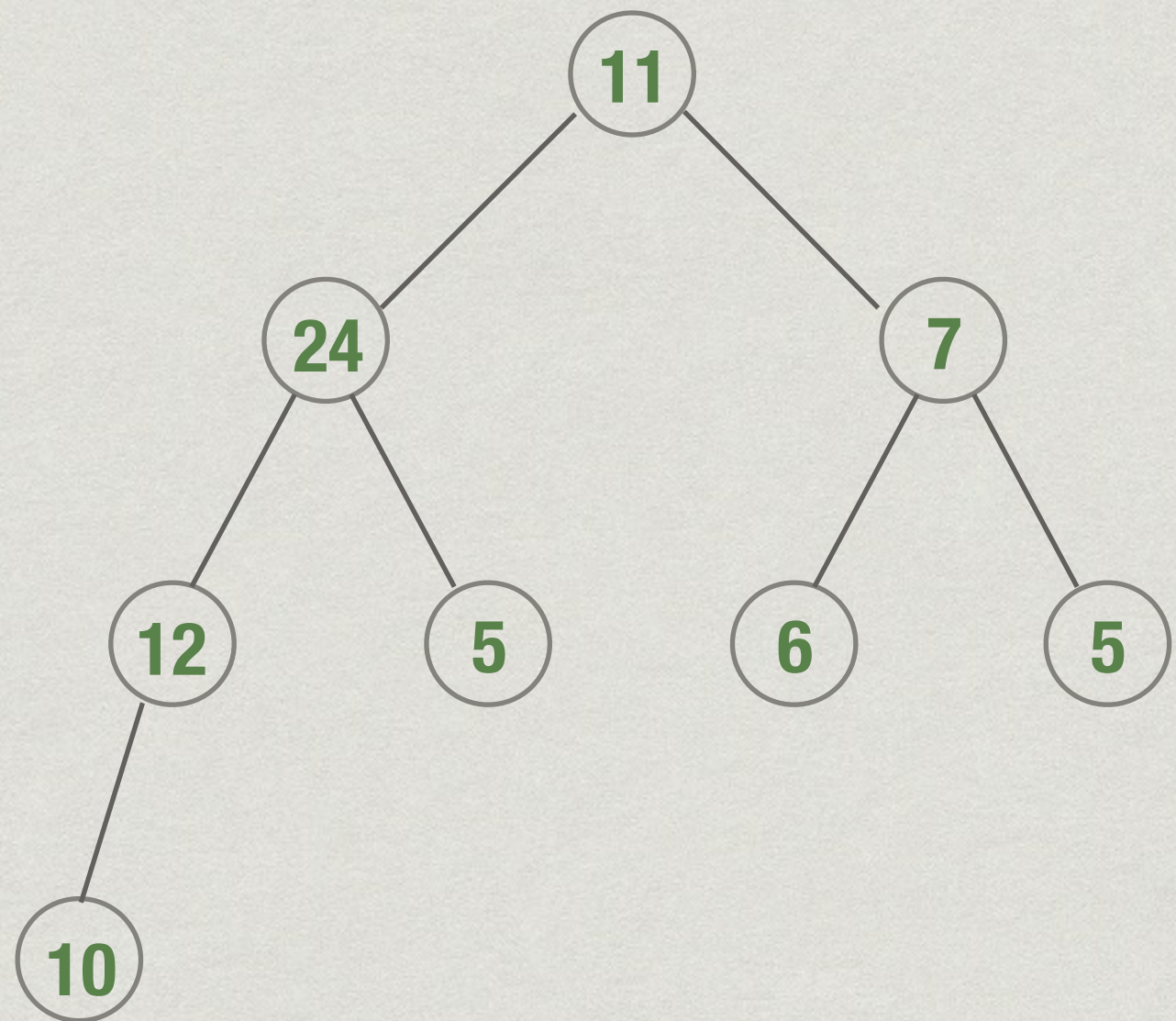
- \* Removing maximum value creates a “hole” at the root
- \* Reducing one value requires deleting last node
- \* Move “homeless” value to root





# delete\_max()

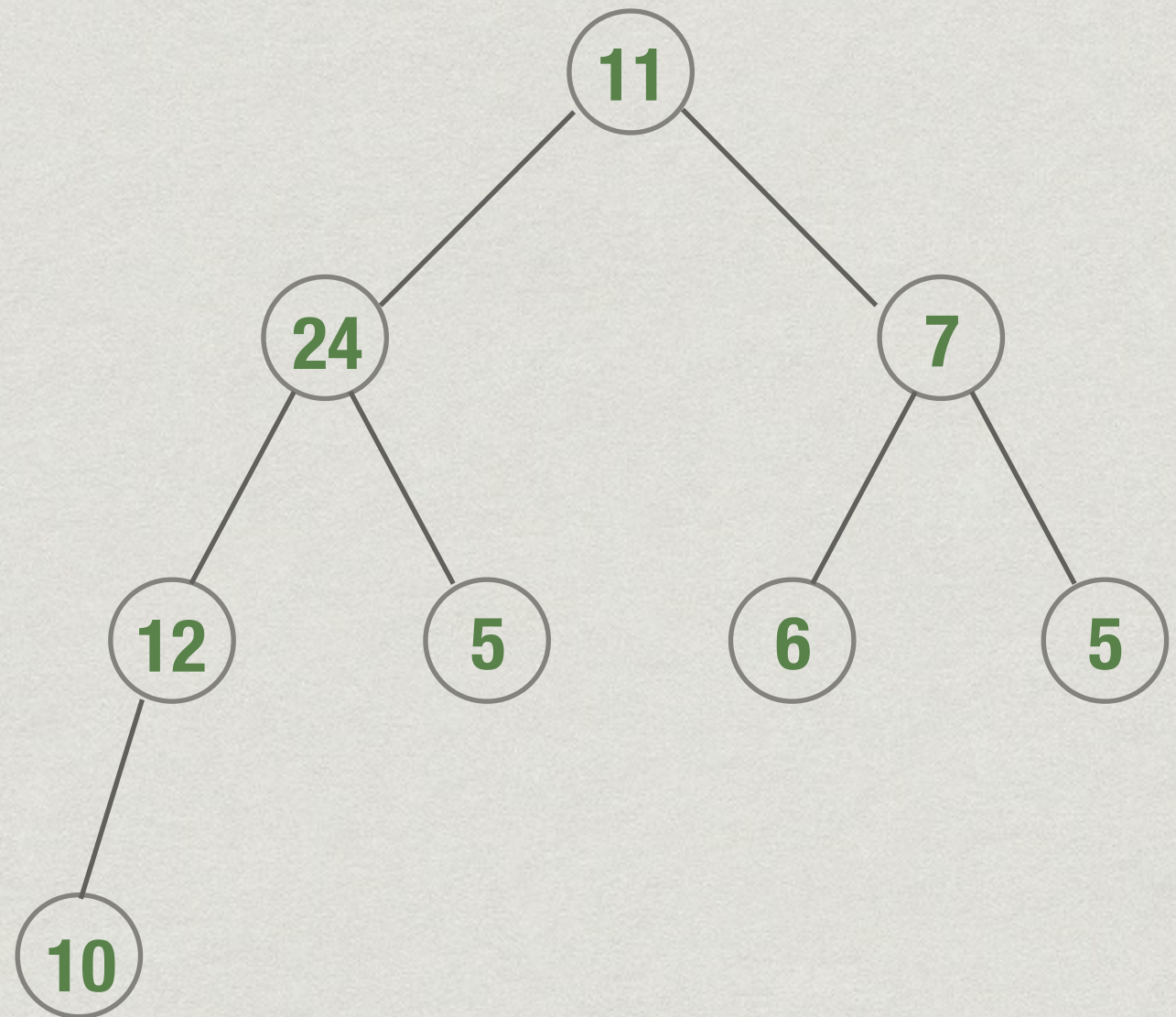
- \* Removing maximum value creates a “hole” at the root
- \* Reducing one value requires deleting last node
- \* Move “homeless” value to root





# delete\_max()

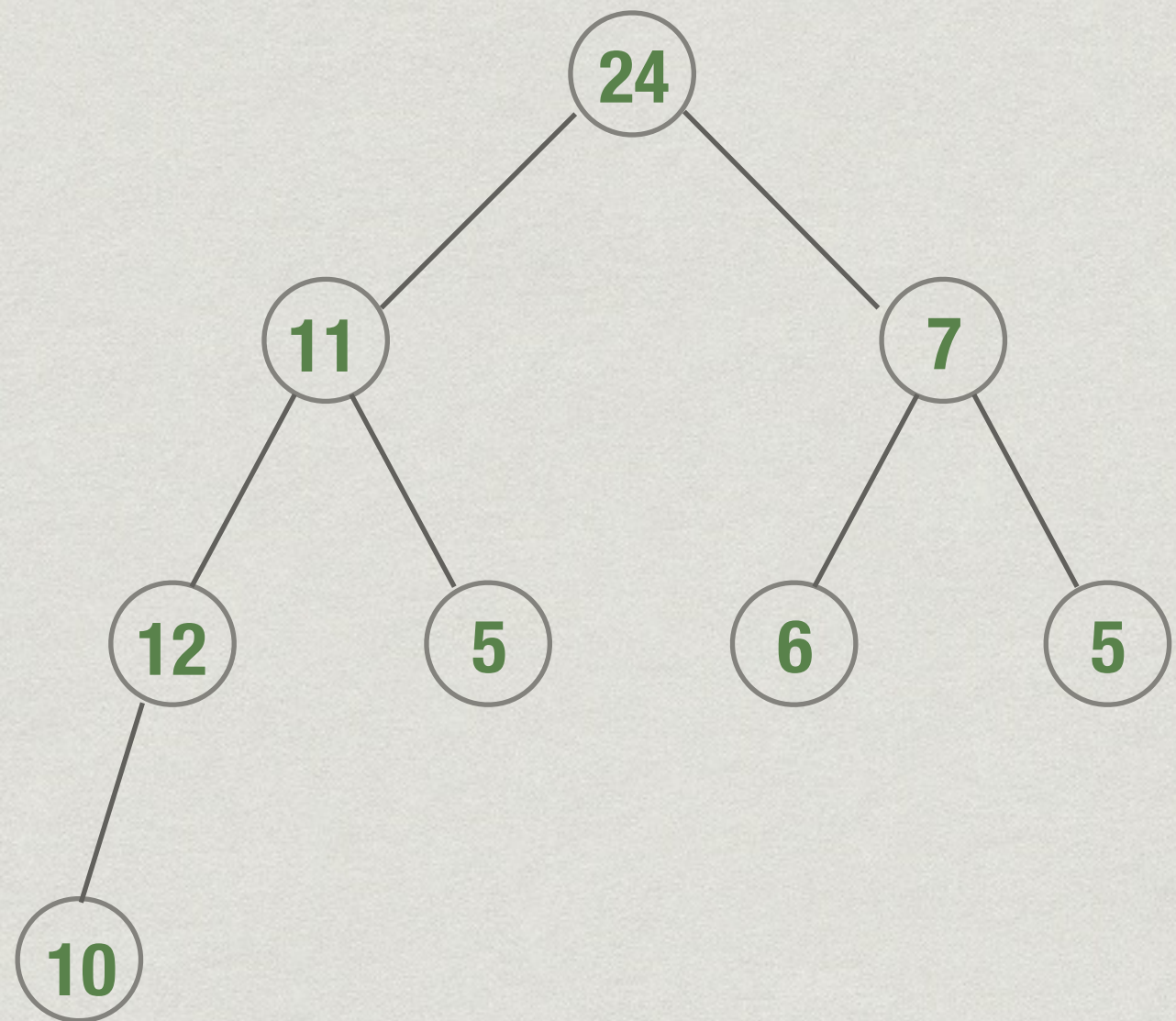
- \* Now restore the heap property from root downwards
- \* Swap with largest child
- \* Will follow a single path from root to leaf





# delete\_max()

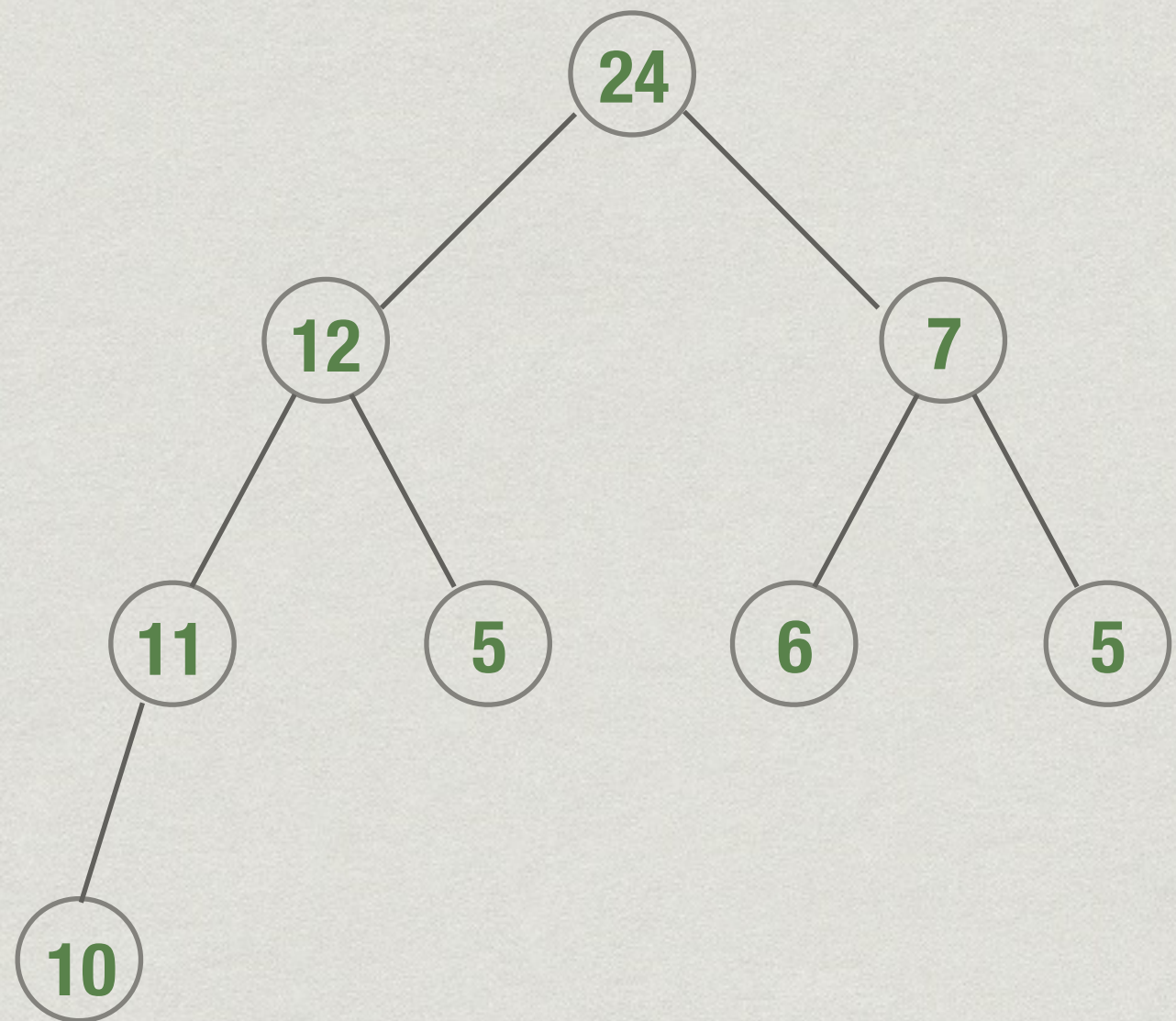
- \* Now restore the heap property from root downwards
- \* Swap with largest child
- \* Will follow a single path from root to leaf





# delete\_max()

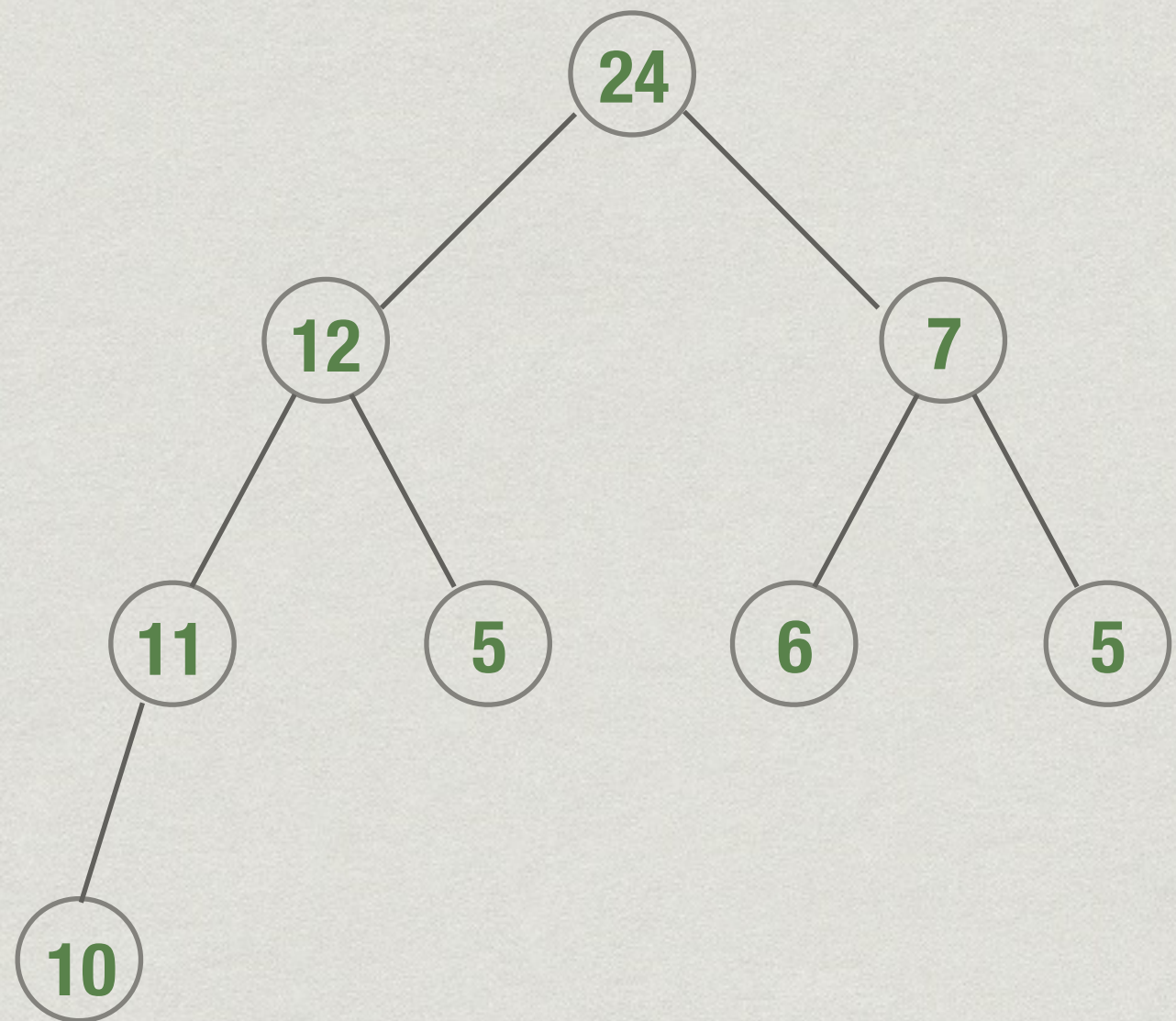
- \* Now restore the heap property from root downwards
- \* Swap with largest child
- \* Will follow a single path from root to leaf





# delete\_max()

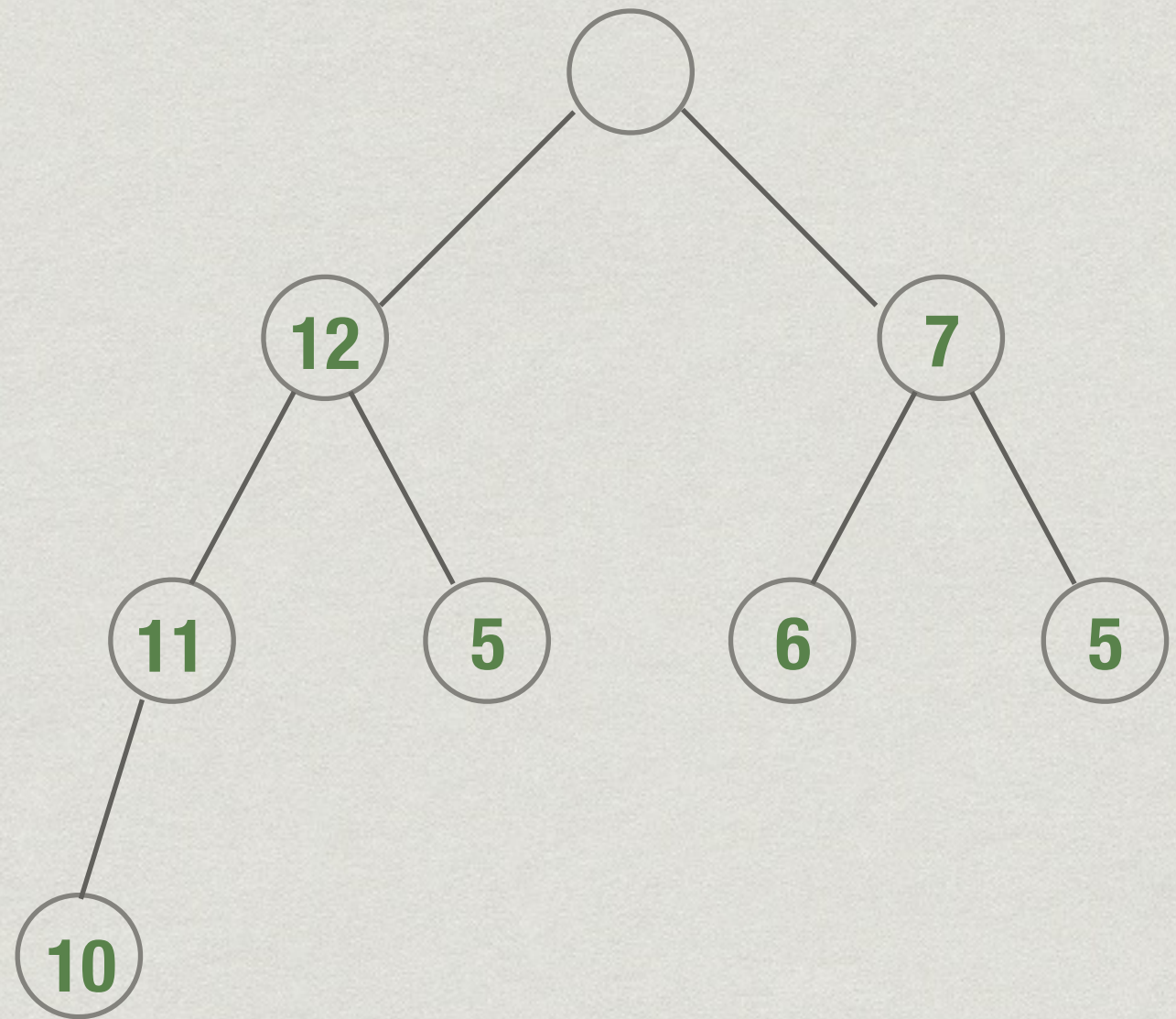
- \* Will follow a single path from root to leaf
- \* Cost proportional to height of tree
- \*  $O(\log N)$





# delete\_max()

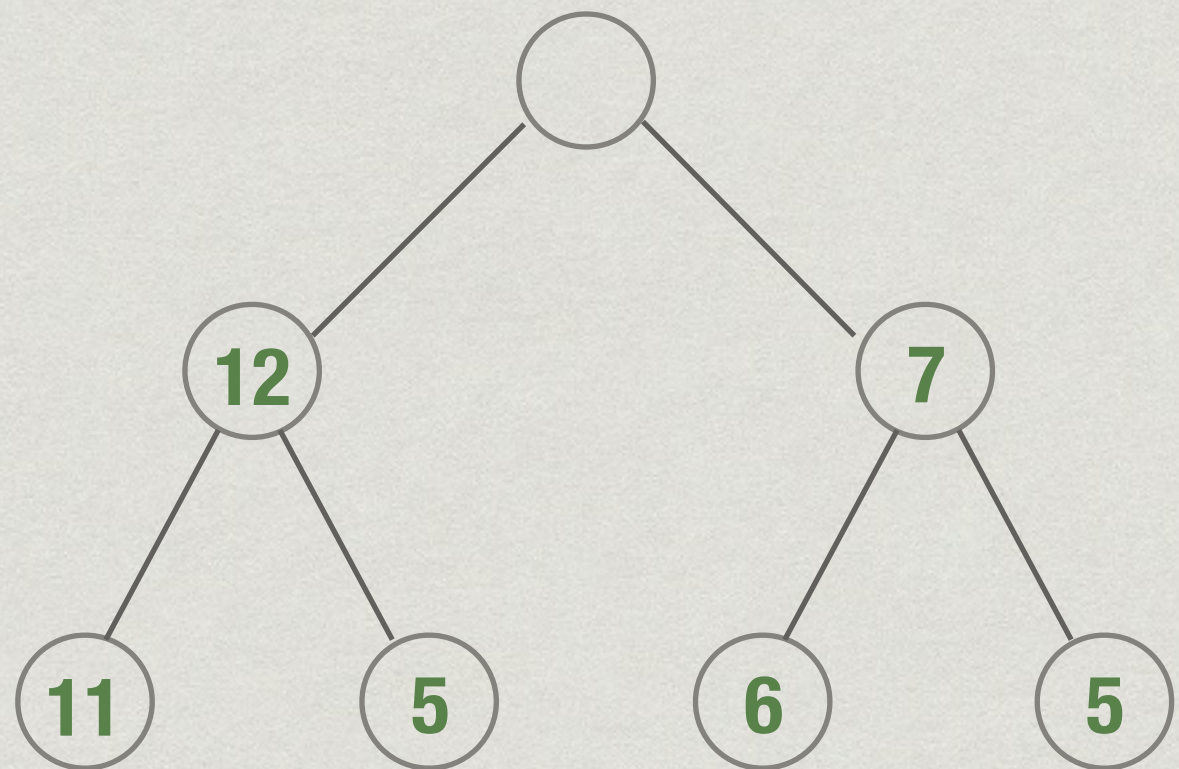
- \* Will follow a single path from root to leaf
- \* Cost proportional to height of tree
- \*  $O(\log N)$





# delete\_max()

- \* Will follow a single path from root to leaf
- \* Cost proportional to height of tree
- \*  $O(\log N)$

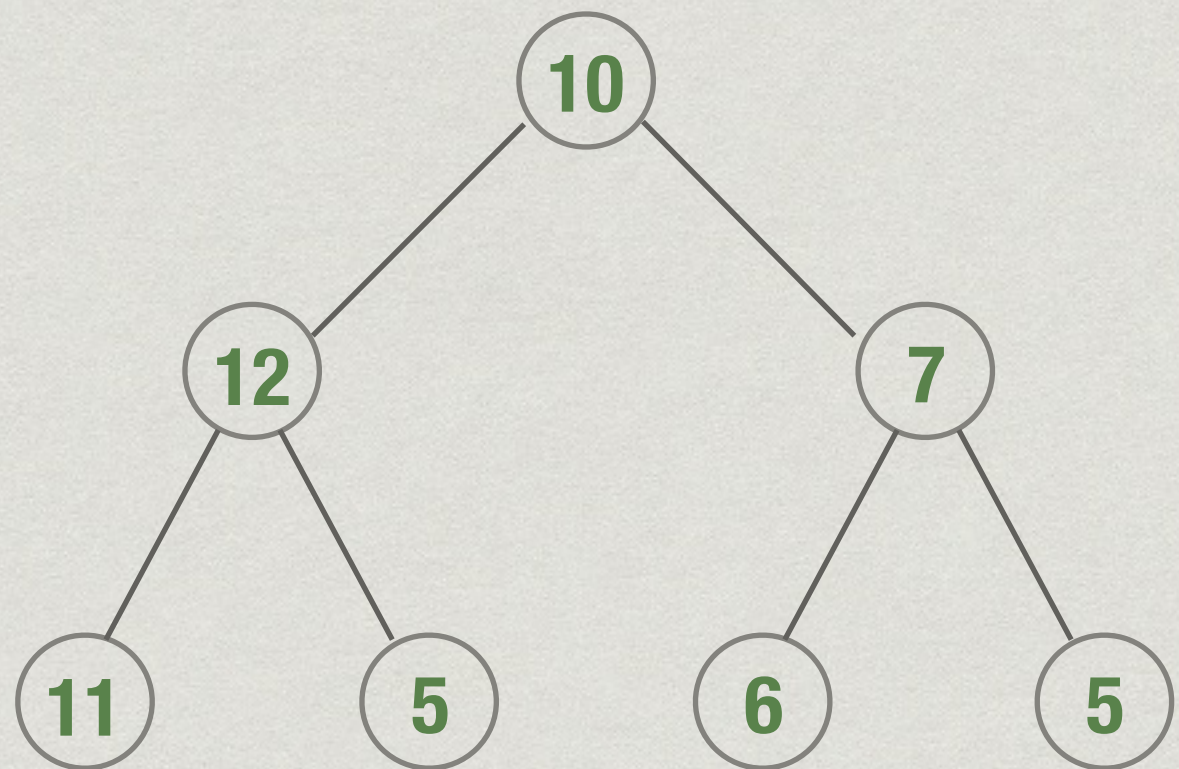


10



# delete\_max()

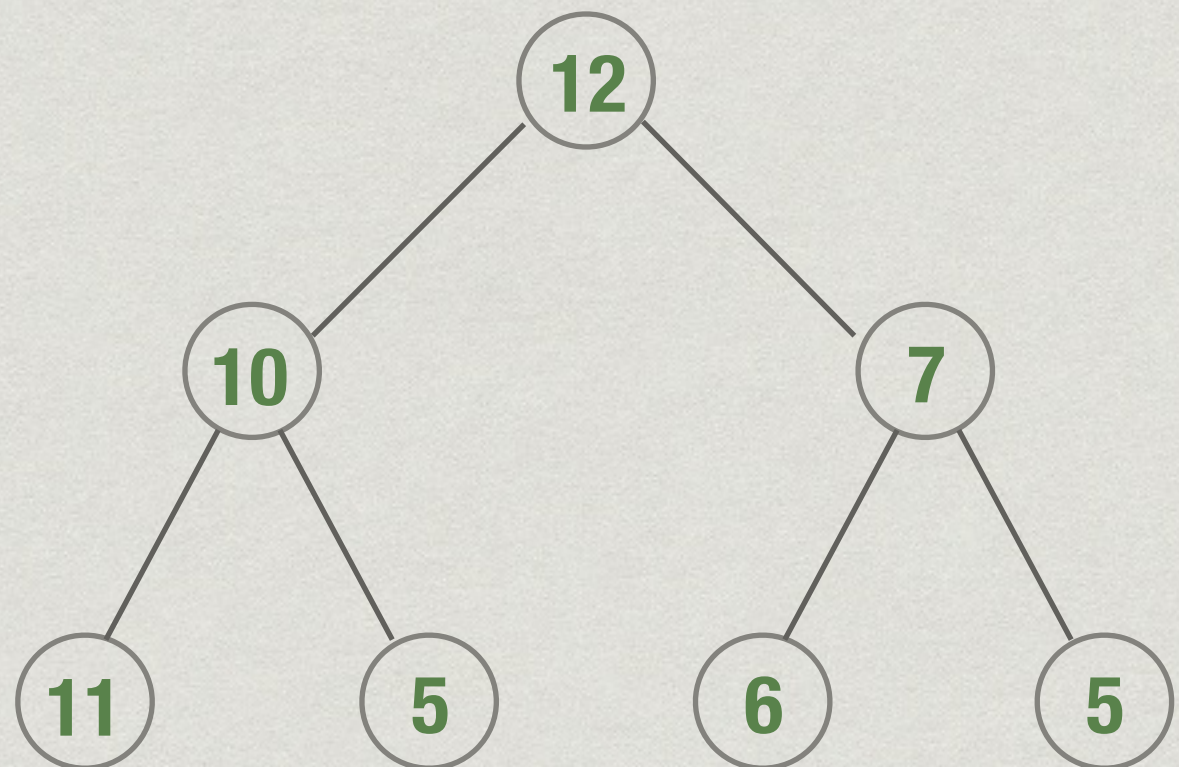
- \* Will follow a single path from root to leaf
- \* Cost proportional to height of tree
- \*  $O(\log N)$





# delete\_max()

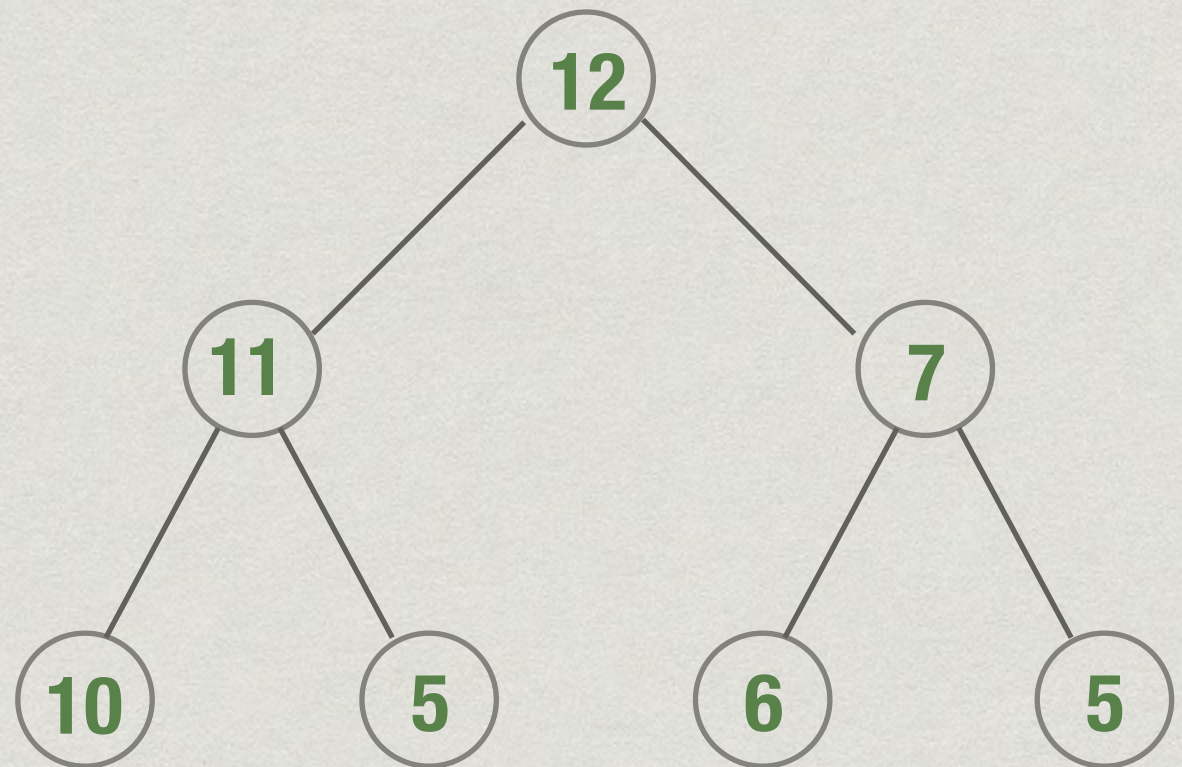
- \* Will follow a single path from root to leaf
- \* Cost proportional to height of tree
- \*  $O(\log N)$





# delete\_max()

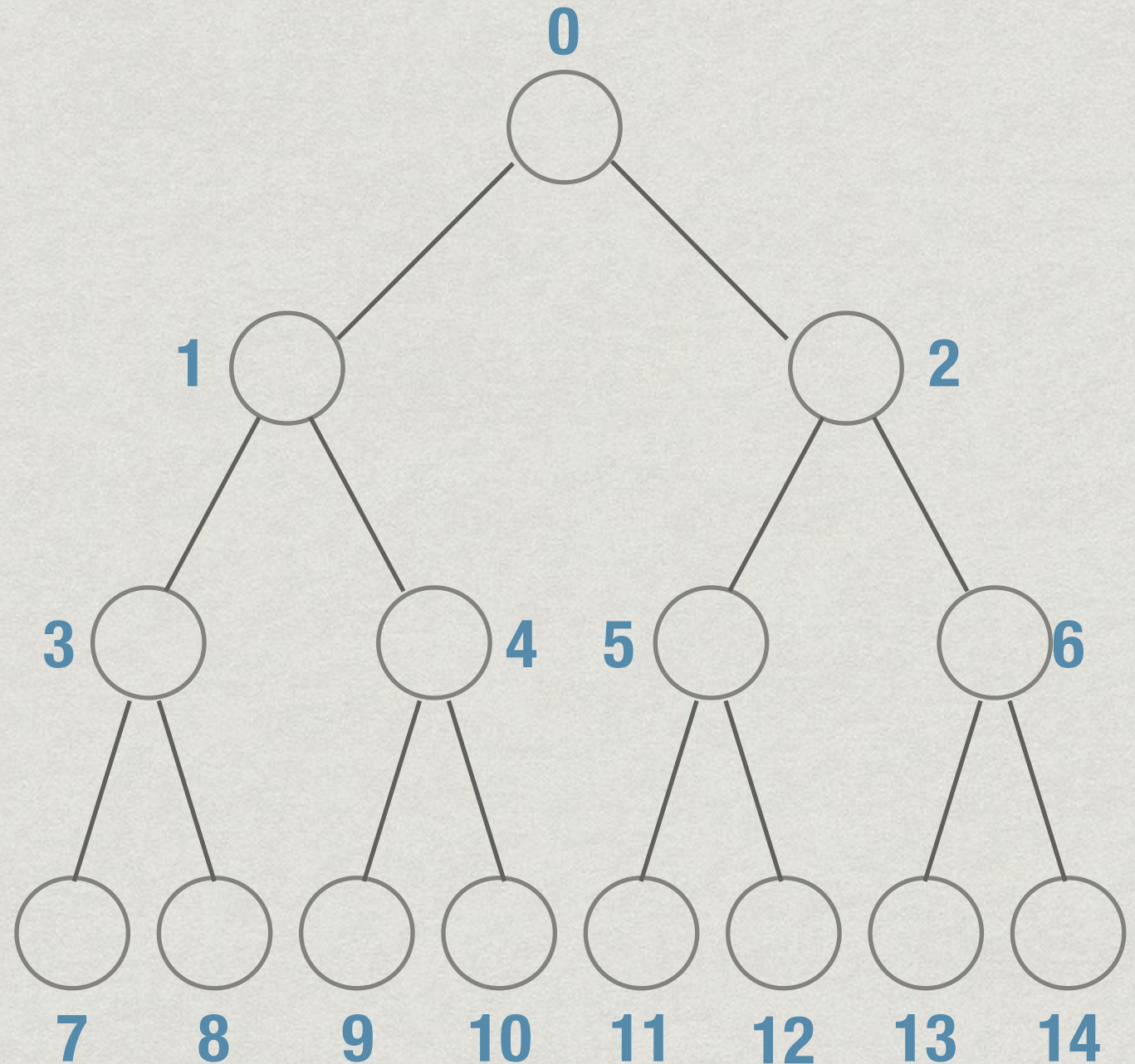
- \* Will follow a single path from root to leaf
- \* Cost proportional to height of tree
- \*  $O(\log N)$





# Implementing using arrays

- \* Number the nodes left to right, level by level
- \* Represent as an array  $H[0..N-1]$
- \* **Children** of  $H[i]$  are at  $H[2i+1]$ ,  $H[2i+2]$
- \* **Parent** of  $H[j]$  is at  $H[\text{floor}((j-1)/2)]$  for  $j > 0$





# Building a heap, `heapify()`

- \* Given a list of values  $[x_1, x_2, \dots, x_N]$ , build a heap
- \* Naive strategy
  - \* Start with an empty heap
  - \* Insert each  $x_j$
  - \* Overall  $O(N \log N)$



# Better heapify( )

- \* Set up the array as  $[x_1, x_2, \dots, x_N]$ 
  - \* Leaf nodes trivially satisfy heap property
  - \* Second half of array is already a valid heap
- \* Assume leaf nodes are at level  $k$ 
  - \* For each node at level  $k-1, k-2, \dots, 0$ , fix heap property
  - \* As we go up, the number of steps per node goes up by 1, but the number of nodes per level is halved
  - \* Cost turns out to be  $O(N)$  overall



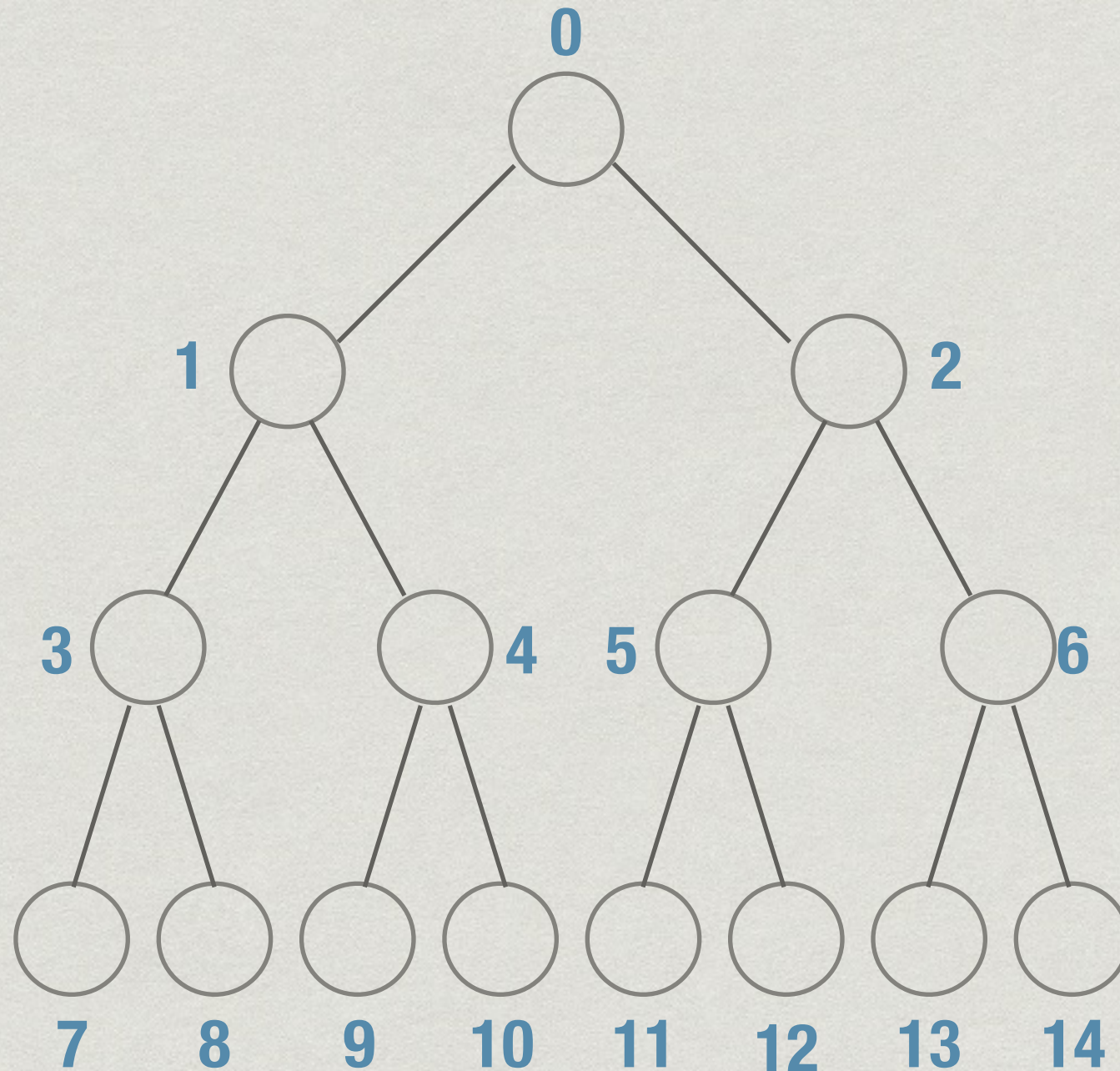
# Better heapify()

**1 node,  
height 3 repair**

**2 nodes,  
height 2 repair**

**4 nodes,  
height 1 repair**

**$N/2$  nodes  
already satisfy  
heap property**





# Heap sort

- \* Start with an unordered list
- \* Build a heap —  $O(n)$
- \* Call `delete_max()`  $n$  times to extract elements in descending order —  $O(n \log n)$
- \* After each `delete_max()`, heap shrinks by 1
  - \* Store maximum value at the end of current heap
  - \* In place  $O(n \log n)$  sort



# Summary

- \* Heaps are a tree implementation of priority queues
  - \* `insert()` and `delete_max()` are both  $O(\log N)$
  - \* `heapify()` builds a heap in  $O(N)$
  - \* Tree can be manipulated easily using an array
- \* Can invert the heap condition
  - \* Each node is **smaller** than its children
  - \* **Min-heap**, for `insert()`, `delete_min()`