

18. Searching a List

Topics:

Linear Search

Binary Search

Measuring Execution Time

The Divide and Conquer Framework

Search

Examples:

Is this song in that playlist?

Is this number in that phone book?

Is this name in that phone book?

Is this fingerprint in that archive of fingerprints?

Is this photo in that yearbook?

Linear Search

LinSearch: The Spec

```
def LinSearch(x,a) :  
    """ Returns an int k with the  
    property that a[k]==x is True.  
    If no such k exists, then  
    k==-1.  
  
    PreC: a is a nonempty list of  
    ints and x is an int.  
    """
```

Could also apply the same ideas for searching a list of strings.

Linear Search

0 1 2 3 4 5 6 7 8 9 10 11

a-→

86	73	43	35	23	45	42	62	15	25	51	35
----	----	----	----	----	----	----	----	----	----	----	----

k-→

0

x-→

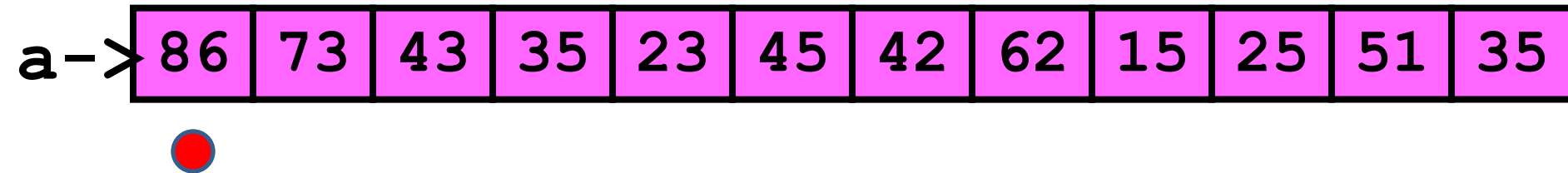
23

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]:  
            return k  
    return -1
```

Walk down the list looking for a match

Linear Search

0 1 2 3 4 5 6 7 8 9 10 11



k-→

0

x-→

23

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]:  
            return k  
    return -1
```

Walk down the list looking for a match

Linear Search

0 1 2 3 4 5 6 7 8 9 10 11

a-→

86	73	43	35	23	45	42	62	15	25	51	35
----	----	----	----	----	----	----	----	----	----	----	----



k-→

1

x-→

23

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]:  
            return k  
    return -1
```

Walk down the list looking for a match

Linear Search

0 1 2 3 4 5 6 7 8 9 10 11

a-→

86	73	43	35	23	45	42	62	15	25	51	35
----	----	----	----	----	----	----	----	----	----	----	----



k-→

2

x-→

23

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]:  
            return k  
    return -1
```

Walk down the list looking for a match

Linear Search

0 1 2 3 4 5 6 7 8 9 10 11

a-→

86	73	43	35	23	45	42	62	15	25	51	35
----	----	----	----	----	----	----	----	----	----	----	----



k-→

3

x-→

23

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]:  
            return k  
    return -1
```

Walk down the list looking for a match

Linear Search

0 1 2 3 4 5 6 7 8 9 10 11

a-→

86	73	43	35	23	45	42	62	15	25	51	35
----	----	----	----	----	----	----	----	----	----	----	----



k-→

4

x-→

23

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]: Yup  
            return k  
    return -1
```

Walk down the list looking for a match

Linear Search

0 1 2 3 4 5 6 7 8 9 10 11

a-→

86	73	43	35	23	45	42	62	15	25	51	35
----	----	----	----	----	----	----	----	----	----	----	----



k-→

4

x-→

23

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]:  
            return k  
    return -1
```

All done

Walk down the list looking for a match

Linear Search: No Match Case

0 1 2 3 4 5 6 7 8 9 10 11

a-→

86	73	43	35	23	45	42	62	15	25	51	35
----	----	----	----	----	----	----	----	----	----	----	----



k-→

11

x-→

7

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]: Nope  
        return k  
    return -1
```

Walk down the list looking for a match

Linear Search: No Match Case

0 1 2 3 4 5 6 7 8 9 10 11

a-→

86	73	43	35	23	45	42	62	15	25	51	35
----	----	----	----	----	----	----	----	----	----	----	----



x-→

7

```
def LinSearch(x,a):  
    for k in range(len(a)):  
        if x == a[k]:  
            return k  
    return -1
```

Yup

Return -1 if no match

Linear Search: While Implementation

```
def LinSearchW(x,a):  
    k=0  
    while k<len(a) and a[k]!=x:  
        k+=1  
    if k==len(a):  
        return -1  
    else:  
        return k
```

Binary Search

Now we assume that the list
to be searched is sorted
from little to big.

```
a = [10,20,40,60,90]
```

```
a = ['brown', 'dog', 'fox', 'lazy', 'quick', 'the']
```


Back to Using Phone Books

The Ithaca
phone book has
10,000+ entries.

The Manhattan phone book has 1,000,000+ entries. But it does not take 100 x longer to look something up. Why?

wide at **SuperPages.com**

195

Car

17 566-1282
26 Allen Ln Jewett 01938... 978 356-9960
Cartagena Leves
18 Javert St 02131... 617 323-7639
Cartagena Avith
9 Bancroft Court 02119... 617 442-9780
H Yd 02136... 617 361-5253
Cartagena 600 North St 02119... 617 241-0152
Lucilla 174 Harvard Cam 02139... 617 491-5621
M W 95 Rowe St 02133... 617 323-9713
Melvin 301 Green Cam 02139... 617 576-1061
Carte Nicholas
18 Appleton Court 02116... 617 695-9966
Cartegena O A Millard St 01118... 617 338-8219
Carten Thos Jr & Claire
1 Paradise Rd Mill 02186... 617 698-6163
Thomas & Kathleen
50 Thompson Ln Mill 02186... 617 696-6919
Carter A Rod 02133... 617 327-2257
A Roabury... 617 442-5230
A 31 Bethune Wy Roxbury 02129... 617 442-1219
A 260 Putnam Av Cambridge 02139... 617 492-4174
A M 255 Macthys Av 02115... 617 666-7153
Adams 361 Centre St Mill 02186... 617 698-0794
Alice 105 Kilmartock St 02215... 617 425-0193
Alice 45 Alameda Cambridge 02139... 617 945-2171
Andrew F 62 Vinal Av Som 02143... 617 625-7623
Carter Anne M 02118... 617 339-1022
Carter Athens
272 Newington Boston 02116... 617 536-6329
B E 68 Gledhill Ave Mill 02126... 617 296-6911
Carter Barbara M 02118... 617 698-6911
Tuffs-New England Medical Center 02111
Call... 617 636-0051
Carter Becky B 02114... 617 523-4368
Bernard J
112 Gladstone E Boston 02218... 617 567-3430
Bithiah 25 Midway Dr 02124... 617 298-8713
Blake 26 Mill Vernon St 02108... 617 367-9931
Carter Broadcasting Co
20 Park Pl 02114... 617 423-0210
Carter & Burgess Consultants Inc
23 East St Som 02141... 617 225-0200
Carter C 2000 Commwntv Av Bst 02135... 617 782-2118
C 228 Fayerham Ave Boston 02238... 617 569-1545
359 Harvard St 02115... 617 491-8822
C 610 Walk Hill Cam 02136... 617 296-6392
C & M 43 Burroughs Jan 02126... 617 242-9558

Carter C 24 Hillock Rd 02131... 617 327-1105
Faye & Ricky
377 Centre St Som 02116... 617 437-7331
Francis S 134 Temple W Rox 02332... 617 323-6781
Franklin & Anne
221 Mt Auburn Cam 02138... 617 354-0798
Fred 42 Haverdam dwn 02136... 617 524-3078
Fred 94 Hixfield Rd Mill 02186... 617 698-1543
G & R Verdun Dr 02124... 617 636-8906
G T 27 Franklin Av Som 02145... 617 623-7121
Gayle 25 Frontenac Dr 02124... 617 825-0322
Geo 5 115 Moss Hill Rd Cam 02130... 617 322-9228
Geo 5 125 Kuzma Rd 02114... 617 567-9548
Carter Halliday Associate
107 S Street Som 02111... 617 456-1689
Carter Harry F
26 Summng Bk W Rox 02132... 617 325-5465
Carter Hide Co
146 Summer St 02110... 617 542-7987
Carter Hilary 61 Harvey Cam 02136... 617 876-2750
Horace
200 Cambridge St Roxbury 02119... 617 442-5307
Howard Jr 26 Notre Dame Rox 02119... 617 445-5532
J Cam... 617 354-2688
J 15 Oatham Bk 02446... 617 732-7990
J 518 Harvard Rd 02446... 617 730-9493
J 75 Pine Wey Wey Roxbury 02132... 617 323-5574
Carter J Jacques M
1 Brookline Pl Bk 02446... 617 735-8787
Carter J M
1410 Columbia Rd Bst 02127... 617 464-1040
Carter J M Ornamental Ironwork
1000 Cambridge St 02114... 617 436-5353
Call... 617 436-5353
Carter J Veal Co
48 Newmarket Sq Rox 02118... 617 442-1775
Carter James
1100 Cambridge St Cam 02138... 617 492-1214
James 182 Fisher Av Roxbury 02120... 617 739-2193
James
37 Gold Star Rd Cambridge 02140... 617 876-0871
Jan 14 Roxbury Rd Mill 02126... 617 361-0773
John 14 Adams Rd Som 02146... 617 964-4035
Jeffrey 41 Warren Av Bst 02116... 617 426-5994
John 11 Mansfield Bk 02134... 617 987-2163
John 327 Summer St 02110... 617 423-4334
John 48 Westwind Rd Dor 02125... 617 282-1232
June O 329 A Summer Av Bst 02131... 617 734-6193
K 38 Broadview Av Dorchester 02124... 617 265-8456
K 17 Esmond Dorchester 02121... 617 282-1593

Carter Nella E
133 Randolph Av Bst 02115... 617 267-6483
Nichols S F
115 Raymond Av Mill 02186... 617 698-5307
Nick 21 Fairfield Bst 02116... 617 267-5222
Nick & Debby
166 Herrick Rd Newton 02459... 617 527-0490
Nicole M
38 Chickatawut Dr 02122... 617 822-1203
P 9 Crestwood Pk Rd 02121... 617 427-4754
P E 501 E South St Bst 02127... 617 268-4213
P 44 Hatching Brook 02121... 617 427-9170
P L 91 Byrner Jan 02130... 617 983-8692
Paul & Constance
114 Anawan Av Rox 02135... 617 325-2036
P 501 E South St Bst 02127... 617 268-4213
Paul M 27 Union Bk 02133... 617 787-2115
Carter Pile Driving Inc 17 Beaver Ct
Frankingham 02102... Wellesley Tello 781 235-8488
Carter Prudence
46 Franklin Waterway 02172... 617 393-3782
Prudence
46 Franklin Waterway 02172... 617 393-3782
Reginald
66 Brunswick Dorchester 02321... 617 541-2843
Renee & Andrew
10 Walnut Bst 02108... 617 720-3765
Carter Rice Dowd
Bulky Depot Publishing 163 Main Wilmington 01887
Tel Free-Dial 7 & Thru... 800 638-1671
Cust Svc-Industrial Prod 613 Main Wilmington
Tel Free-Dial 7 & Thru... 800 619-7447
Cust Svc-Printing 613 Main Wilmington
Tel Free-Dial 7 & Thru... 800 648-7447
Headquarters 613 Main Wilmington 01887
Ingalls Center 163 Main Wilmington 01887
Tel Free-Dial 7 & Thru... 800 638-1673
Carter Richard
1079 Commwntv Av Brighton 02125... 617 987-0836
Richard A 97 Mt Vernon Bst 02106... 617 566-7929
Richard A 170 Commwntv Av Bst 0211

Key Idea: Repeated Halving

To find Derek Jeter's number...

B = phone book

while (B is longer than 1 page):

1. P = middle page of B

2. Let Q be the first name on P

3. **if** 'Jeter' comes before Q:

 Rip away the 2nd half of B

else:

 Rip away the 1st half of B.

Scan remaining page P line-by-line for 'Jeter'

What Happens to Phone Book Length?

Original: 3000 pages

After 1 rip: 1500 pages

After 2 rips: 750 pages

After 3 rips: 375 pages

After 4 rips: 188 pages

After 5 rips: 94 pages

After 12 rips: 1 page

Binary Search

The idea of repeatedly halving the size of the "search space" is the main idea behind the method of binary search.

An item in a sorted array of length n can be located with approximately $\log_2 n$ comparisons.

$$\log_2 8 = 3 \quad \log_2 64 = 7 \quad \log_2 2^{**k} = k$$

What is $\log_2(n)$?

n	$\text{ceil}(\log_2(n))$
10	4
100	7
1000	10
10000	14
100000	17
1000000	20

BinSearch: The Spec

```
def BinSearch(x,a) :  
    """ Returns an int k with the  
    property that a[k]==x is True.  
    If no such k exists, then  
    k==-1.  
  
    PreC: a is a nonempty list of  
    ints that is sorted from smallest  
    to largest. x is an int.  
    """
```

Example:
Does this List have an Element
With Value Equal to 70?

0 1 2 3 4 5 6 7 8 9 10 11

12	15	33	35	42	45	51	62	73	75	86	98
----	----	----	----	----	----	----	----	----	----	----	----

Let's Look For x in a

0 1 2 3 4 5 6 7 8 9 10 11

$a \rightarrow$

12	15	33	35	42	45	51	62	73	75	86	98
----	----	----	----	----	----	----	----	----	----	----	----



L:

0

$a[\text{Mid}] \leq x$????

Mid:

5

R:

11

x :

70

$\text{Mid} = (\text{L} + \text{R}) / 2$

The Midpoint Computations

L	R	$(L+R) / 2$

0	11	5
2	6	4
1	100	50

Let's Look For x in a

0 1 2 3 4 5 6 7 8 9 10 11

$a \rightarrow$

12	15	33	35	42	45	51	62	73	75	86	98
----	----	----	----	----	----	----	----	----	----	----	----



L:

0

$a[\text{Mid}] \leq x$????

Mid:

5

R:

11

x :

70

Let's Look For x in a

0 1 2 3 4 5 6 7 8 9 10 11

$a \rightarrow$

12	15	33	35	42	45	51	62	73	75	86	98
----	----	----	----	----	----	----	----	----	----	----	----



L:

0

Mid:

5

R:

11

$a[\text{Mid}] \leq x$

x :

70

Yes!

So throw away
The "left half"

Let's Look For x in a

0 1 2 3 4 5 6 7 8 9 10 11

a →

12	15	33	35	42	45	51	62	73	75	86	98
----	----	----	----	----	----	----	----	----	----	----	----



L:

0

Mid:

5

R:

11

x:

70

$a[\text{Mid}] \leq x$

Yes!

So throw away
The "left half"

Let's Look For x in a

0 1 2 3 4 5 6 7 8 9 10 11

a →

12	15	33	35	42	45	51	62	73	75	86	98
----	----	----	----	----	----	----	----	----	----	----	----



L:

0

$a[\text{Mid}] \leq x$

Mid:

5

Revise L and Mid

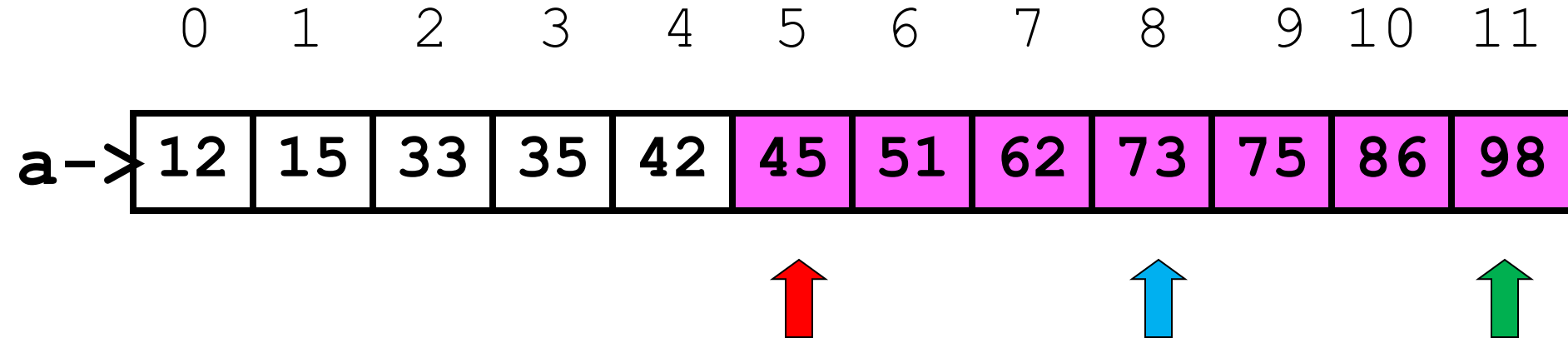
R:

11

x:

70

Let's Look For x in a



L: 5

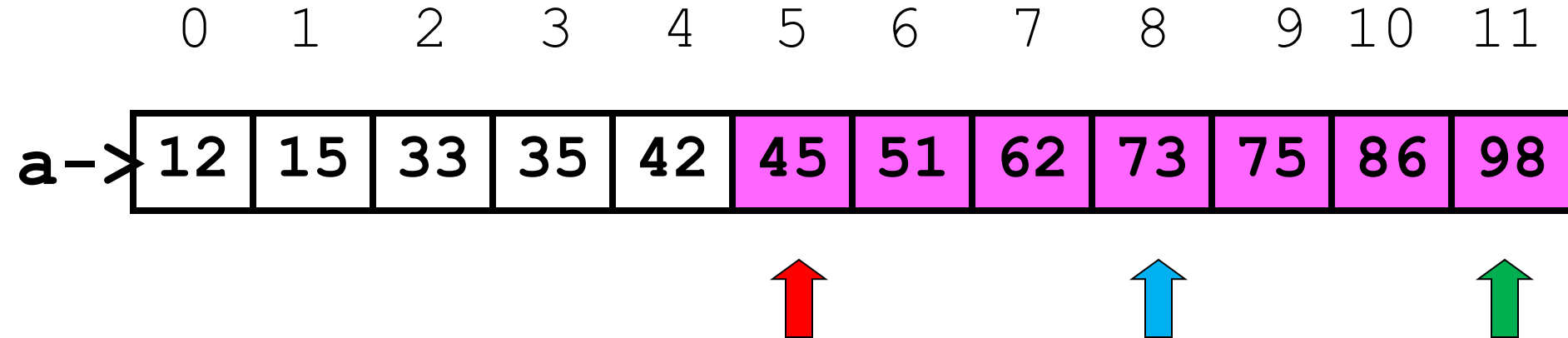
Mid: 8

R: 11

x: 70

a[Mid] <= x ???

Let's Look For x in a



L:

5

$a[\text{Mid}] \leq x$

Mid:

8

No

R:

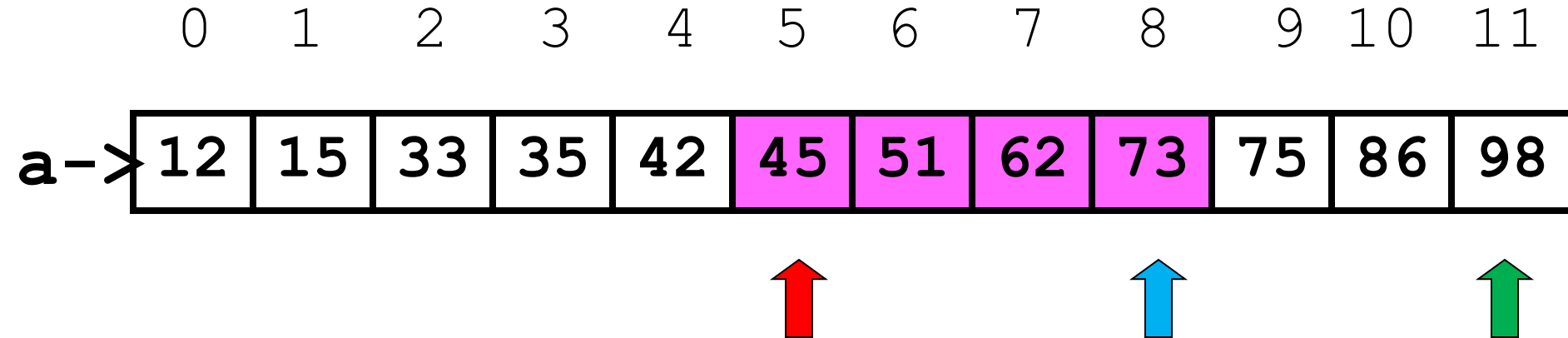
11

x :

70

So throw away the
"right half"

Let's Look For $x = 70$



L:

5

Mid:

8

R:

11

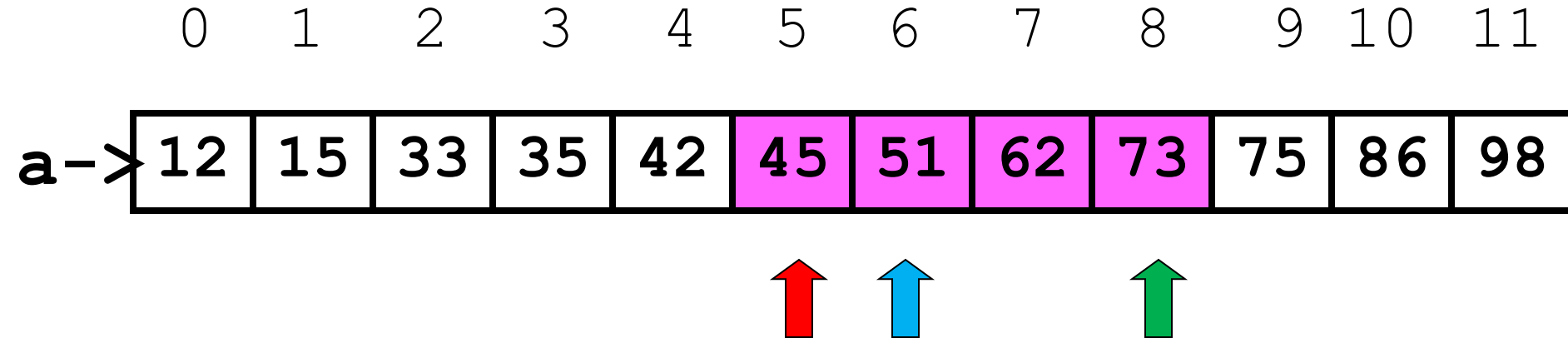
x:

70

$a[\text{Mid}] \leq x$

Revise R and Mid

Let's Look For $x = 70$



L:

5

$a[\text{Mid}] \leq x$

Mid:

6

Revise R and Mid

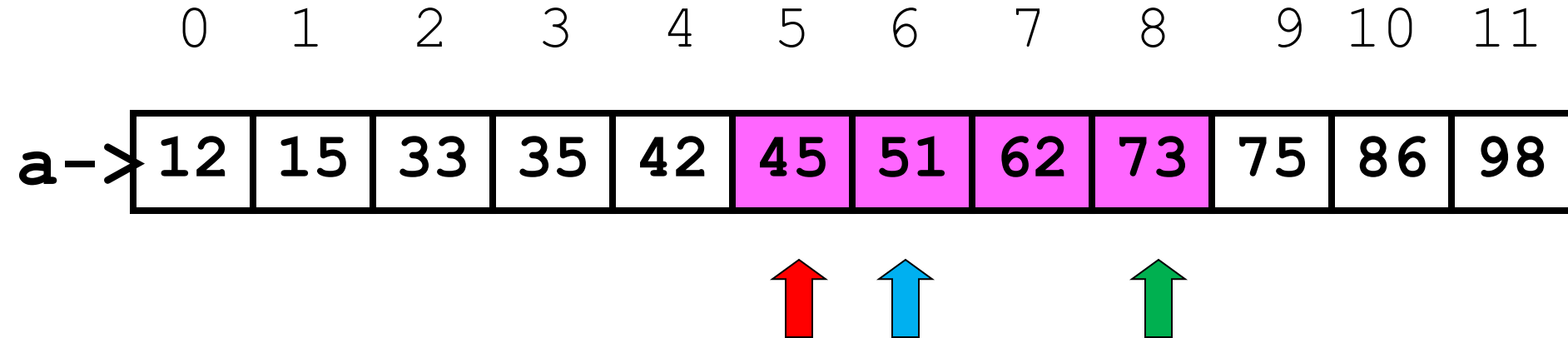
R:

8

x:

70

Let's Look For x in a



L:

5

a[Mid] <= x ????

Mid:

6

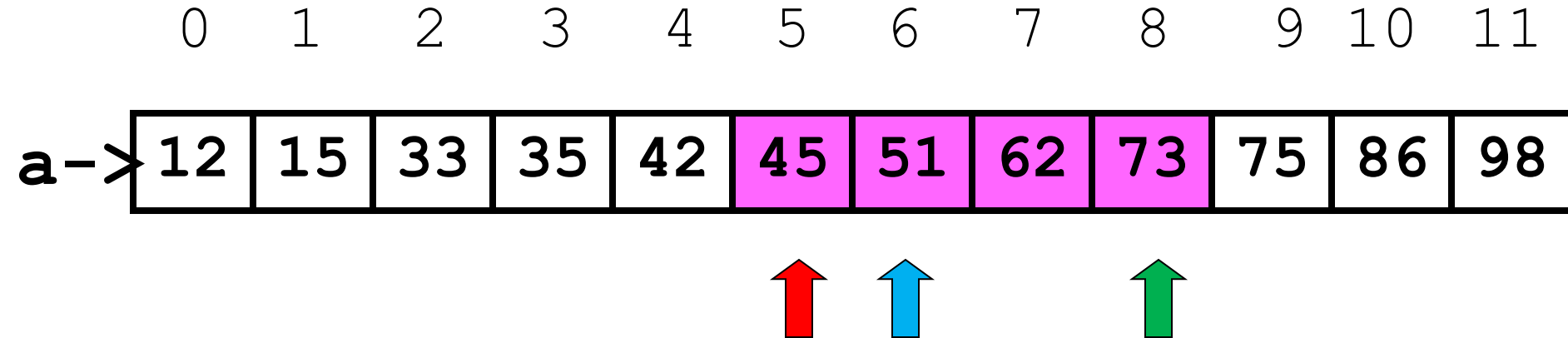
R:

8

x:

70

Let's Look For x in a



L:

5

$a[\text{Mid}] \leq x$

Mid:

6

Yes

R:

8

x:

70

Throw away the
Left half

Let's Look For x in a

0 1 2 3 4 5 6 7 8 9 10 11

a-→

12	15	33	35	42	45	51	62	73	75	86	98
----	----	----	----	----	----	----	----	----	----	----	----



L:

6

$a[\text{Mid}] \leq x$

Mid:

7

Yes

R:

8

x:

70

Let's Look For x in a

0 1 2 3 4 5 6 7 8 9 10 11

a-→

12	15	33	35	42	45	51	62	73	75	86	98
----	----	----	----	----	----	----	----	----	----	----	----



L:

6

$a[\text{Mid}] \leq x$

Mid:

7

Throw away the
left half

R:

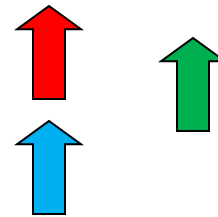
8

x:

70

Let's Look For x in a

	0	1	2	3	4	5	6	7	8	9	10	11
a →	12	15	33	35	42	45	51	62	73	75	86	98



L: 6

Mid: 7

R: 8

x: 70

Done! At this point we just compare x with $a[L]$ and $a[L+1]$.

What We Just Did

```
L = 0
R = len(a) - 1
while R - L > 1:
    # a[L] <= x <= a[R]
    Mid = (L + R) / 2
    if x <= a[mid]:
        R = Mid
    else:
        L = Mid
```

A Loop
Invariant

Note that $a[L] \leq x \leq a[R]$ remains True throughout the loop

What We Just Did

```
L = 0
R = len(a) - 1
while R - L > 1:
    # a[L] <= x <= a[R]
    Mid = (L + R) / 2
    if x <= a[mid]:
        R = Mid
    else:
        L = Mid
```

What is the situation when the loop terminates?

What We Just Did

```
L = 0
R = len(a) - 1
while R - L > 1:
    # a[L] <= x <= a[R]
    Mid = (L + R) / 2
    if x <= a[mid]:
        R = Mid
    else:
        L = Mid
```

$R - L \leq 1$ implies $R = L + 1$

After the Loop Ends



This is True: $a[L] \leq x \leq a[L+1]$

After the Loop Ends



```
if x==a[L]:  
    return L  
elif x==a[L+1]:  
    return L+1  
else:  
    return -1
```

Measuring Execution Time

We now have two ways to search a list:

`LinSearch(x, a)`
`BinSearch(x, a)`

Intuition: BinSearch much faster.

Can we quantify this with a "stop watch"?

The `timeit` Module

This module can be used to time how long it takes to execute a chunk of code.

Typical chunk = some function of interest.

This is called benchmarking.

Benchmarking

Let's benchmark `LinSearch(x, a)` and `BinSearch(x, a)`.

Compare how long it takes when `len(a)` equals 1000, 10000, 100000, and 1000000.

Our intuition tells us that as `len(a)` increases, `BinSearch` will be dramatically faster.

BinSearch vs LinSearch

n	tBin	tLin	tLinW
1000	0.0007	0.0064	0.0119
10000	0.0009	0.0668	0.1203
100000	0.0011	0.8296	1.2082
1000000	0.0015	17.7388	13.9341

tBin = time for BinSearch

tLin = time for LinSearch (for loop version)

tLinW = time for LinSearch (while-loop version)

BinSearch vs LinSearch

n	tLin/tBin
1000	9
10000	74
100000	754
1000000	7095

Reporting ratios is more illuminating since we do not really care about the time units in this informal comparison

Using the `timeit` Module

We show how this module was use to get the results on the previous slides.

Our `LinSearch` vs `BinSearch` example is very typical: is one function faster than another?

A Benchmarking Framework

```
from timeit import *
```

```
S = """
```

Set-up code

```
"""
```

```
B = """
```

Code to Benchmark

```
"""
```

```
p = 10; m = 100
```

```
t = min(Timer(B, setup=S) . repeat(p, m) )
```

Yes, these are doc strings.

The Set-Up and Bench Codes

```
from random import randint as randi
from ShowSearch import BinSearch
n = 10000
s = [randi(0,10*n) for i in range(n)]
s.sort()
x = s[n/2]
```

```
k=BinSearch(x,s)
```

The set-up code is run once.

It is not timed.

It just sets up the code to be timed.

A Benchmarking Framework

```
from timeit import *
```

```
S = """
```

Set-up code

```
"""
```

```
B = """
```

Code to Benchmark

```
"""
```

```
p = 10; m = 100
```

```
t = min(Timer(B, setup=S).repeat(p, m))
```

An “experiment” consists of running the blue code m times.

The stopwatch will time how long it takes to do one experiment

Larger values necessary if the blue code executes very quickly

A Benchmarking Framework

```
from timeit import *
```

```
S = """
```

Set-up code

```
"""
```

```
B = """
```

Code to Benchmark

```
"""
```

```
p = 10; m = 100
```

```
t = min(Timer(B, setup=S) . repeat(p, m) )
```

Timer returns a length-p list. Each element is the stopwatch time for 1 experiment

This helps control for other stuff that may be running on your computer.

A Benchmarking Framework

```
from timeit import *
```

```
S = """
```

Set-up code

```
"""
```

```
B = """
```

Code to Benchmark

```
"""
```

```
p = 10; m = 100
```

```
t = min(Timer(B, setup=S) . repeat(p, m) )
```

In general, it is best to take the minimum as the most reliable. The benchmark time is assigned to t

This helps control for other stuff that may be running on your computer.

Why Benchmarking is Important

Confirms/refutes what our intuition might say about efficiency.

Makes us sensitive to the various issues that affect efficiency.

Steers us away from simplistic comparisons of different methods that can be used on the same problem.