Searching and Sorting

Part One
Recap from Last Time
double averageOf(const Vector<int>& vec) {
    double total = 0.0;

    for (int i = 0; i < vec.size(); i++) {
        total += vec[i];
    }

    return total / vec.size();
}

Assume any individual statement takes one unit of time to execute. If the input Vector has \( n \) elements, how many time units will this code take to run?
Assume any individual statement takes one unit of time to execute. If the input Vector has \( n \) elements, how many time units will this code take to run?

```cpp
double averageOf(const Vector<int>& vec) {
    double total = 0.0;
    for (int i = 0; i < vec.size(); i++) {
        total += vec[i];
    }
    return total / vec.size();
}
```
One possible answer: $3n + 4$. 

```cpp
double averageOf(const Vector<int>& vec) {
    double total = 0.0;
    for (int i = 0; i < vec.size(); i++) {
        total += vec[i];
    }
    return total / vec.size();
}
```
double averageOf(const Vector<int>& vec) {
    double total = 0.0;
    for (int i = 0; i < vec.size(); i++) {
        total += vec[i];
    }
    return total / vec.size();
}

One possible answer: $3n + 4$.
More useful answer: $O(n)$. 
void printStars(int n) {
    for (int i = 0; i < n; i++) {
        for (int j = 0; j < n; j++) {
            cout << '*' << endl;
        }
    }
}

Work Done: $O(n^2)$. 
How much time will it take for these functions to run, as a function of $n$?

```cpp
void beni(int n) {
    for (int i = 0; i < 2 * n; i++) {
        for (int j = 0; j < 5 * n; j++) {
            cout << '*' << endl;
        }
    }
}

void pando(int n) {
    for (int i = 0; i < 3 * n; i++) {
        cout << '*' << endl;
    }
    for (int i = 0; i < 8; i++) {
        cout << '*' << endl;
    }
}
```

$O(n^2)$

$O(n)$
New Stuff!
Sorting Algorithms
What is sorting?
One style of “sorting,” but not the one we’re thinking about...
<table>
<thead>
<tr>
<th>Time</th>
<th>Auto</th>
<th>Athlete</th>
<th>Nationality</th>
<th>Date</th>
<th>Venue</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:37.0</td>
<td>Anne</td>
<td>Smith Anne</td>
<td>United Kingdom</td>
<td>3 June 1967[8]</td>
<td>London</td>
</tr>
<tr>
<td>4:36.8</td>
<td>Maria</td>
<td>Gommers Maria</td>
<td>Netherlands</td>
<td>14 June 1969[8]</td>
<td>Leicester</td>
</tr>
<tr>
<td>4:35.3</td>
<td>Ellen</td>
<td>Tittel Ellen</td>
<td>West Germany</td>
<td>20 August 1971[8]</td>
<td>Sittard</td>
</tr>
<tr>
<td>4:29.5</td>
<td>Paola</td>
<td>Pigni Paola</td>
<td>Italy</td>
<td>8 August 1973[8]</td>
<td>Viareggio</td>
</tr>
<tr>
<td>4:17.44</td>
<td>Maricica</td>
<td>Puica Maricica</td>
<td>Romania</td>
<td>9 September 1982[8]</td>
<td>Rieti</td>
</tr>
<tr>
<td>4:15.61</td>
<td>Paula</td>
<td>Ivan Paula</td>
<td>Romania</td>
<td>10 July 1989[8]</td>
<td>Nice</td>
</tr>
<tr>
<td>4:12.56</td>
<td>Svetlana</td>
<td>Masterkova Svetlana</td>
<td>Russia</td>
<td>14 August 1996[8]</td>
<td>Zürich</td>
</tr>
<tr>
<td>4:12.33</td>
<td>Sifan</td>
<td>Hassan Sifan</td>
<td>Netherlands</td>
<td>12 July 2019</td>
<td>Monaco</td>
</tr>
</tbody>
</table>

**Problem:** Given a list of data points, sort those data points into ascending / descending order by some quantity.
Suppose we want to rearrange a sequence to put elements into ascending order.

What are some strategies we could use?

How do those strategies compare?

Is there a “best” strategy?
An Initial Idea: *Selection Sort*
An Initial Idea: Selection Sort

4 1 2 7 6
An Initial Idea: *Selection Sort*

The smallest element should go in front.
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*

This element is in the right place now.

The remaining elements are in no particular order.
An Initial Idea: *Selection Sort*
An Initial Idea: Selection Sort
An Initial Idea: *Selection Sort*

The smallest element of the remaining elements goes at the front of the remaining elements.
An Initial Idea: **Selection Sort**
An Initial Idea: Selection Sort
An Initial Idea: *Selection Sort*

These elements are in the right place now.

The remaining elements are in no particular order.
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*

The smallest of these remaining elements goes at the front of the remaining elements.
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*

These elements are in the right place now.

The remaining elements are in no particular order.
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*

![Diagram of Selection Sort](image)
An Initial Idea: *Selection Sort*

The smallest of these elements needs to go at the front of this group of elements.
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*

These elements are in the right place now.

The remaining elements are in no particular order.
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*

The smallest element from this group needs to go at the front of the group.

The smallest element from this group needs to go at the front of the group.
An Initial Idea: *Selection Sort*
An Initial Idea: *Selection Sort*

These elements are in the right place now.
Selection Sort

- Find the smallest element and move it to the first position.
- Find the smallest element of what’s left and move it to the second position.
- Find the smallest element of what’s left and move it to the third position.
- Find the smallest element of what’s left and move it to the fourth position.
- (etc.)
/**
 * Sorts the specified vector using the selection sort algorithm.
 */
void selectionSort(Vector<int>& elems) {
    for (int index = 0; index < elems.size(); index++) {
        int smallestIndex = indexOfSmallest(elems, index);
        swap(elems[index], elems[smallestIndex]);
    }
}

/**
 * Given a vector and a starting point, returns the index of the
 * smallest element in that vector at or after the starting point.
 */
int indexOfSmallest(const Vector<int>& elems, int startPoint) {
    int smallestIndex = startPoint;
    for (int i = startPoint + 1; i < elems.size(); i++) {
        if (elems[i] < elems[smallestIndex]) {
            smallestIndex = i;
        }
    }
    return smallestIndex;
}
How fast is selection sort?
How fast is selection sort?
How fast is selection sort?
How fast is selection sort?

{ 09, 69, 20, 16, 46, 10, 29, 90, 67, 18, 53, 20, 38, 20, 46 }
Finding the element that goes in position 0 requires us to scan all $n$ elements.

How fast is selection sort?
Finding the element that goes in position 0 requires us to scan all \( n \) elements.

How fast is selection sort?
Finding the element that goes in position 0 requires us to scan all $n$ elements.

How fast is selection sort?
Finding the element that goes in position 0 requires us to scan all $n$ elements.

How fast is selection sort?
Finding the element that goes in position 0 requires us to scan all $n$ elements.
Finding the element that goes in position 0 requires us to scan all $n$ elements.

Finding the element that goes in position 1 requires us to scan $n - 1$ elements.
Finding the element that goes in position 0 requires us to scan all \( n \) elements.

Finding the element that goes in position 1 requires us to scan \( n - 1 \) elements.

How fast is selection sort?
Finding the element that goes in position 0 requires us to scan all \( n \) elements.
Finding the element that goes in position 1 requires us to scan \( n - 1 \) elements.

\[
\begin{align*}
\text{Number of elements scanned: } & \quad n + (n - 1) + (n - 2) + \ldots + 2 + 1 \\
\end{align*}
\]
Finding the element that goes in position 0 requires us to scan all \( n \) elements.

Finding the element that goes in position 1 requires us to scan \( n - 1 \) elements.

\[
\text{Number of elements scanned: } n + (n-1) + (n-2) + \ldots + 2 + 1
\]
Finding the element that goes in position 0 requires us to scan all \( n \) elements.

Finding the element that goes in position 1 requires us to scan \( n - 1 \) elements.

How fast is selection sort?
How fast is selection sort?

Finding the element that goes in position 0 requires us to scan all $n$ elements.

Finding the element that goes in position 1 requires us to scan $n - 1$ elements.
Finding the element that goes in position 0 requires us to scan all \( n \) elements.

Finding the element that goes in position 1 requires us to scan \( n - 1 \) elements.

Finding the element that goes in position 2 requires us to scan \( n - 2 \) elements.

How fast is selection sort?
Finding the element that goes in position 0 requires us to scan all \( n \) elements.

Finding the element that goes in position 1 requires us to scan \( n - 1 \) elements.

Finding the element that goes in position 2 requires us to scan \( n - 2 \) elements.

... 

Number of elements scanned:

\[
 n + (n-1) + (n-2) + ... + 2 + 1.
\]
\[ n + (n-1) + \ldots + 2 + 1 = \frac{n(n+1)}{2} \]
The Complexity of Selection Sort

\[ O(n (n + 1) / 2) \]
\[ = O(n (n + 1)) \]
\[ = O(n^2 + n) \]
\[ = O(n^2) \]

So selection sort runs in time \( O(n^2) \).
Our theory predicts that the runtime of selection sort is $O(n^2)$.

Does that match what we see in practice?

What should we expect to see when we look at a runtime plot?
Another Sorting Algorithm
Our Next Idea: Insertion Sort
Our Next Idea: **Insertion Sort**

7  2  4  1  6
Our Next Idea: *Insertion Sort*

This sequence in blue, taken in isolation, is in sorted order.

This sequence in gray is in no particular order.
Our Next Idea: Insertion Sort

Swap this element back until it’s in the proper place in the blue sequence.
Our Next Idea: Insertion Sort
Our Next Idea: *Insertion Sort*
Our Next Idea: **Insertion Sort**

This sequence in blue, taken in isolation, is in sorted order.

This sequence in gray is in no particular order.
Our Next Idea: **Insertion Sort**

Swap this element back until it’s in the proper place in the blue sequence.
Our Next Idea: *Insertion Sort*
Our Next Idea: *Insertion Sort*
Our Next Idea: *Insertion Sort*
Our Next Idea: **Insertion Sort**

This sequence in blue, taken in isolation, is in sorted order.

This sequence in gray is in no particular order.
Our Next Idea: *Insertion Sort*

Swap this element back until it’s in the proper place in the blue sequence.
Our Next Idea: **Insertion Sort**

2 4 7 1 6
Our Next Idea: **Insertion Sort**
Our Next Idea: **Insertion Sort**
Our Next Idea: **Insertion Sort**
Our Next Idea: **Insertion Sort**
Our Next Idea: *Insertion Sort*
Our Next Idea: Insertion Sort

This sequence in blue, taken in isolation, is in sorted order.

This sequence in gray is in no particular order.
Our Next Idea: **Insertion Sort**

Swap this element back until it's in the proper place in the blue sequence.
Our Next Idea: **Insertion Sort**
Our Next Idea: **Insertion Sort**
Our Next Idea: **Insertion Sort**

This sequence in blue, taken in isolation, is in sorted order.

There are no more gray elements, so the sequence is sorted!
Insertion Sort

• Repeatedly *insert* an element into a sorted sequence at the front of the array.

• To *insert* an element, swap it backwards until either
  
  • (1) it’s at least as big as the element before it in the sequence, or
  
  • (2) it’s at the front of the array.
/**
 * Sorts the specified vector using insertion sort.
 *
 * @param v The vector to sort.
 */

void insertionSort(Vector<int>& v) {
    for (int i = 0; i < v.size(); i++) {
        /* Scan backwards until either (1) there is no
         * preceding element or the preceding element is
         * no bigger than us.
         */
        for (int j = i - 1; j >= 0; j--) {
            if (v[j] <= v[j + 1]) break;
            /* Swap this element back one step. */
            swap(v[j], v[j + 1]);
        }
    }
}
How Fast is Insertion Sort?
How Fast is Insertion Sort?

1  2  4  6  7
How Fast is Insertion Sort?
How Fast is Insertion Sort?
How Fast is Insertion Sort?

1  2  4  6  7

The diagram shows the sequence of steps in the insertion sort algorithm for a list of integers.
How Fast is Insertion Sort?

1 2 4 6 7
How Fast is Insertion Sort?
How Fast is Insertion Sort?
How Fast is Insertion Sort?
How Fast is Insertion Sort?
How Fast is Insertion Sort?
How Fast is Insertion Sort?

1

2

4

6

7
How Fast is Insertion Sort?
How Fast is Insertion Sort?

1 2 4 6 7
How Fast is Insertion Sort?

Work done: $O(n)$
How Fast is Insertion Sort?

7  6  4  2  1
How Fast is Insertion Sort?

7 6 4 2 1
How Fast is Insertion Sort?

7
6
4
2
1
How Fast is Insertion Sort?

7 6 4 2 1
How Fast is Insertion Sort?

6 7 4 2 1
How Fast is Insertion Sort?
How Fast is Insertion Sort?

6 7 4 2 1
How Fast is Insertion Sort?

6 7 4 2 1
How Fast is Insertion Sort?

6
4
7
2
1
How Fast is Insertion Sort?

6 4 7 2 1
How Fast is Insertion Sort?

4 6 7 2 1
How Fast is Insertion Sort?

4 6 7 2 1
How Fast is Insertion Sort?

4 6 7 2 1
How Fast is Insertion Sort?

4 6 7 2 1
How Fast is Insertion Sort?

4 6 2 7 1
How Fast is Insertion Sort?
How Fast is Insertion Sort?

4  2  6  7  1
How Fast is Insertion Sort?

4  2  6  7  1
How Fast is Insertion Sort?
How Fast is Insertion Sort?

2 4 6 7 1
How Fast is Insertion Sort?

2 4 6 7 1
How Fast is Insertion Sort?

2 4 6 7 1
How Fast is Insertion Sort?

2
4
6
1
7
How Fast is Insertion Sort?

![Diagram showing the process of insertion sort with elements 2, 4, 6, 1, and 7.](image)
How Fast is Insertion Sort?

2 4 1 6 7
How Fast is Insertion Sort?
How Fast is Insertion Sort?

```
2
1
4
6
7
```
How Fast is Insertion Sort?

2 1 4 6 7
How Fast is Insertion Sort?
How Fast is Insertion Sort?
How Fast is Insertion Sort?

Work Done: \[ 1 + 2 + 3 + \ldots + n-1 \]
\[ = O(n^2) \]
Three Analyses

- **Worst-Case Analysis**
  - What's the *worst* possible runtime for the algorithm?
  - Useful for “sleeping well at night.”

- **Best-Case Analysis**
  - What's the *best* possible runtime for the algorithm?
  - Useful to see if the algorithm performs well in some cases.

- **Average-Case Analysis**
  - What's the *average* runtime for the algorithm?
  - Far beyond the scope of this class; take CS109, CS161, or CS265 for more information!
The Complexity of Insertion Sort

• In the best case (the array is sorted), insertion takes time $O(n)$.

• In the worst case (the array is reverse-sorted), insertion sort takes time $O(n^2)$.

• **Fun fact:** Insertion sorting an array of random values takes, on average, $O(n^2)$ time.
  
  • Curious why? Come talk to me after class!
How do selection sort and insertion sort compare against one another?
Building a Better Sorting Algorithm
\[ n \]

\[ 2n \]
Thinking About $O(n^2)$

$T(n)$

$T(\frac{1}{2}n)$

$T(\frac{1}{2}n)$
Thinking About $O(n^2)$
Thinking About $O(n^2)$

$$2 \cdot \frac{1}{4}T(n) = \frac{1}{2}T(n)$$
With an $O(n^2)$-time sorting algorithm, it takes twice as long to sort the whole array as it does to split the array in half and sort each half.

Can we exploit this?
The Key Insight: **Merge**
The Key Insight: *Merge*
The Key Insight: \textit{Merge}
The Key Insight: **Merge**
The Key Insight: *Merge*
The Key Insight: *Merge*
The Key Insight: Merge
The Key Insight: Merge
The Key Insight: **Merge**
The Key Insight: *Merge*
The Key Insight: \textbf{Merge}
The Key Insight: Merge
The Key Insight: Merge
The Key Insight: *Merge*
The Key Insight: \textit{Merge}

Each step makes a single comparison and reduces the number of elements by one.

If there are $n$ total elements, this algorithm runs in time $O(n)$. 
The Key Insight: **Merge**

- The *merge* algorithm takes in two sorted lists and combines them into a single sorted list.
- While both lists are nonempty, compare their first elements. Remove the smaller element and append it to the output.
- Once one list is empty, add all elements from the other list to the output.
- We’ll leave the code for this as an Exercise to the Reader.
“Split Sort”

1. Split the input in half.

2. Insertion sort each half.

3. Merge the halves back together.
“Split Sort”

1. Split the input in half.
2. Insertion sort each half.
1. Split the input in half.
2. Insertion sort each half.
3. Merge the halves back together.

“Split Sort”
void splitSort(Vector<int>& v) {
    /* Split the vector in half */
    int half = v.size() / 2;
    Vector<int> left  = v.subList(0, half);
    Vector<int> right = v.subList(half);

    /* Sort each half. */
    insertionSort(left);
    insertionSort(right);

    /* Merge them back together. */
    v = merge(left, right);
}
void splitSort(Vector<int>& v) {
    /* Split the vector in half */
    int half = v.size() / 2;
    Vector<int> left  = v.subList(0, half);
    Vector<int> right = v.subList(half);

    /* Sort each half. */
    insertionSort(left);
    insertionSort(right);

    /* Merge them back together. */
    v = merge(left, right);
}
void splitSort(Vector<int>& v) {
    /* Split the vector in half */
    int half = v.size() / 2;
    Vector<int> left = v.subList(0, half);
    Vector<int> right = v.subList(half);

    /* Sort each half. */
    insertionSort(left);
    insertionSort(right);

    /* Merge them back together. */
    v = merge(left, right);
}

Takes O(n) time, since we copy all n elements into new Vectors.

Takes O(n^2) time, but about half as much as what we did before.
void splitSort(Vector<int>& v) {
    /* Split the vector in half */
    int half = v.size() / 2;
    Vector<int> left  = v.subList(0, half);
    Vector<int> right = v.subList(half);

    /* Sort each half. */
    insertionSort(left);
    insertionSort(right);

    /* Merge them back together. */
    v = merge(left, right);
}
void splitSort(Vector<int>& v) {
    /* Split the vector in half */
    int half = v.size() / 2;
    Vector<int> left  = v.subList(0, half);
    Vector<int> right = v.subList(half);

    /* Sort each half. */
    insertionSort(left);
    insertionSort(right);

    /* Merge them back together. */
    v = merge(left, right);
}

Prediction: This should still take time \(O(n^2)\), but be about twice as fast as insertion sort.
Next Time

- **Mergesort**
  - A beautiful, elegant sorting algorithm.
- **Analyzing Mergesort**
  - An unusual runtime analysis.
- **Hybrid Sorting Algorithms**
  - Improving on mergesort.
- **Binary Search**
  - Finding things fast!