Lecture 1 Notes: Introduction

1 Compiled Languages and C++

1.1 Why Use a Language Like C++?

At its core, a computer is just a processor with some memory, capable of running tiny instructions like "store 5 in memory location 23459." Why would we express a program as a text file in a programming language, instead of writing processor instructions?

The advantages:

1. **Conciseness:** programming languages allow us to express common sequences of commands more concisely. C++ provides some especially powerful shorthands.

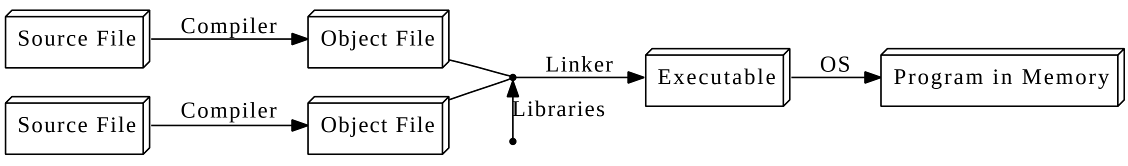
2. **Maintainability:** modifying code is easier when it entails just a few text edits, instead of rearranging hundreds of processor instructions. C++ is *object oriented* (more on that in Lectures 7-8), which further improves maintainability.

3. **Portability:** different processors make different instructions available. Programs written as text can be translated into instructions for many different processors; one of C++'s strengths is that it can be used to write programs for nearly any processor.

C++ is a *high-level* language: when you write a program in it, the shorthands are sufficiently expressive that you don't need to worry about the details of processor instructions. C++ does give access to some lower-level functionality than other languages (e.g. memory addresses).

1.2 The Compilation Process

A program goes from text files (or *source files*) to processor instructions as follows:

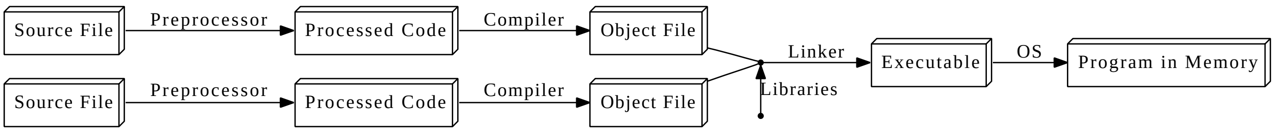


Object files are intermediate files that represent an incomplete copy of the program: each source file only expresses a piece of the program, so when it is compiled into an object file, the object file has some markers indicating which missing pieces it depends on. The linker takes those object files and the compiled libraries of predefined code that they rely on, fills in all the gaps, and spits out the final program, which can then be run by the operating system (OS).

The compiler and linker are just regular programs. The step in the compilation process in which the compiler reads the file is called *parsing*.

In C++, all these steps are performed ahead of time, before you start running a program. In some languages, they are done during the execution process, which takes time. This is one of the reasons C++ code runs far faster than code in many more recent languages.

C++ actually adds an extra step to the compilation process: the code is run through a *preprocessor*, which applies some modifications to the source code, before being fed to the compiler. Thus, the modified diagram is:



1.3 General Notes on C++

C++ is immensely popular, particularly for applications that require speed and/or access to some low-level features. It was created in 1979 by Bjarne Stroustrup, at first as a set of extensions to the C programming language. C++ extends C; our first few lectures will basically be on the C parts of the language.

Though you can write graphical programs in C++, it is much hairier and less portable than text-based (*console*) programs. We will be sticking to console programs in this course.

Everything in C++ is case sensitive: `someName` is not the same as `SomeName`.

2 Hello World

In the tradition of programmers everywhere, we'll use a "Hello, world!" program as an entry point into the basic features of C++.

2.1 The code

```
// A Hello World program
#include <iostream>

```

```
int main() {
  std::cout << "Hello, world!\n";

  return 0;
}
```

2.2 Tokens

Tokens are the minimals chunk of program that have meaning to the compiler – the smallest meaningful symbols in the language. Our code displays all 6 kinds of tokens, though the usual use of operators is not present here:

Token type	Description/Purpose	Examples
Keywords	Words with special meaning to the compiler	`int, double, for, auto`
Identifiers	Names of things that are not built into the language	`cout, std, x, myFunction`
Literals	Basic constant values whose value is specified directly in the source code	`"Hello, world!", 24.3, 0, 'c'`
Operators	Mathematical or logical operations	`+, -, &&, %, <<`
Punctuation/Separators	Punctuation defining the structure of a program	`{ } ( ) , ;`
Whitespace	Spaces of various sorts; ignored by the compiler	Spaces, tabs, newlines, comments

2.3 Line-By-Line Explanation

1. `//` indicates that everything following it until the end of the line is a comment: it is ignored by the compiler. Another way to write a comment is to put it between `/*` and `*/` (e.g. `x = 1 + /*sneaky comment here*/ 1;`). A comment of this form may span multiple lines. Comments exist to explain non-obvious things going on in the code. Use them: document your code well!

2. Lines beginning with `#` are preprocessor commands, which usually change what code is actually being compiled. `#include` tells the preprocessor to dump in the contents of another file, here the `iostream` file, which defines the procedures for input/output.

4. `int main() {...}` defines the code that should execute when the program starts up. The curly braces represent grouping of multiple commands into a *block*. More about this syntax in the next few lectures.

5.
   - `cout << :` This is the syntax for outputting some piece of text to the screen. We'll discuss how it works in Lecture 9.
   - **Namespaces:** In C++, identifiers can be defined within a context – sort of a directory of names – called a *namespace*. When we want to access an identifier defined in a namespace, we tell the compiler to look for it in that namespace using the *scope resolution operator* (`::`). Here, we're telling the compiler to look for `cout` in the `std` namespace, in which many standard C++ identifiers are defined.

     A cleaner alternative is to add the following line below line 2:

     ```
     using namespace std;
     ```

     This line tells the compiler that it should look in the `std` namespace for any identifier we haven't defined. If we do this, we can omit the `std::` prefix when writing `cout`. This is the recommended practice.
   - **Strings:** A sequence of characters such as `Hello, world` is known as a *string*. A string that is specified explicitly in a program is a *string literal*.
   - **Escape sequences:** The `\n` indicates a *newline* character. It is an example of an *escape sequence* – a symbol used to represent a special character in a text literal. Here are all the C++ escape sequences which you can include in strings:

Escape Sequence	Represented Character
`\a`	System bell (beep sound)
`\b`	Backspace
`\f`	Formfeed (page break)
`\n`	Newline (line break)
`\r`	"Carriage return" (returns cursor to start of line)
`\t`	Tab
`\\`	Backslash
`\'`	Single quote character
`\"`	Double quote character
`\`*some integer x*	The character represented by x

7. `return 0` indicates that the program should tell the operating system it has completed successfully. This syntax will be explained in the context of functions; for now, just include it as the last line in the `main` block.

Note that every statement ends with a semicolon (except preprocessor commands and blocks using {}). Forgetting these semicolons is a common mistake among new C++ programmers.

3 Basic Language Features

So far our program doesn't do very much. Let's tweak it in various ways to demonstrate some more interesting constructs.

3.1 Values and Statements

First, a few definitions:

- A *statement* is a unit of code that does something – a basic building block of a program.
- An *expression* is a statement that has a *value* – for instance, a number, a string, the sum of two numbers, etc. `4 + 2`, `x - 1`, and `"Hello, world!\n"` are all expressions.

Not every statement is an expression. It makes no sense to talk about the value of an `#include` statement, for instance.

3.2 Operators

We can perform arithmetic calculations with *operators*. Operators act on expressions to form a new expression. For example, we could replace `"Hello, world!\n"` with `(4 + 2) / 3`, which would cause the program to print the number 2. In this case, the `+` operator acts on the expressions `4` and `2` (its *operands*).

Operator types:

- **Mathematical:** `+`, `-`, `*`, `/`, and parentheses have their usual mathematical meanings, including using `-` for negation. `%` (the *modulus* operator) takes the remainder of two numbers: `6 % 5` evaluates to 1.
- **Logical:** used for "and," "or," and so on. More on those in the next lecture.
- **Bitwise:** used to manipulate the binary representations of numbers. We will not focus on these.

3.3 Data Types

Every expression has a type – a formal description of what kind of data its value is. For instance, `0` is an integer, `3.142` is a *floating-point* (decimal) number, and `"Hello, world!\n"` is a *string* value (a sequence of characters). Data of different types take a different amounts of memory to store. Here are the built-in datatypes we will use most often:

Type Names	**Description**	**Size**	**Range**
`char`	Single text character or small integer. Indicated with single quotes (`'a'`, `'3'`).	1 byte	signed: -128 to 127 unsigned: 0 to 255
`int`	Larger integer.	4 bytes	signed: -2147483648 to 2147483647 unsigned: 0 to 4294967295
`bool`	Boolean (true/false). Indicated with the keywords `true` and `false`.	1 byte	Just `true` (1) or `false` (0).
`double`	"Doubly" precise floating point number.	8 bytes	+/- 1.7e +/- 308 (15 digits)

Notes on this table:

- A *signed* integer is one that can represent a negative number; an *unsigned* integer will never be interpreted as negative, so it can represent a wider range of positive numbers. Most compilers assume signed if unspecified.
- There are actually 3 integer types: `short`, `int`, and `long`, in non-decreasing order of size (`int` is usually a synonym for one of the other two). You generally don't need to worry about which kind to use unless you're worried about memory usage or you're using really huge numbers. The same goes for the 3 floating point types, `float`, `double`, and `long double`, which are in non-decreasing order of precision (there is usually some imprecision in representing real numbers on a computer).
- The sizes/ranges for each type are not fully standardized; those shown above are the ones used on most 32-bit computers.

An operation can only be performed on compatible types. You can add `34` and `3`, but you can't take the remainder of an integer and a floating-point number.

An operator also normally produces a value of the same type as its operands; thus, `1 / 4` evaluates to `0` because with two integer operands, `/` truncates the result to an integer. To get 0.25, you'd need to write something like `1 / 4.0`.

A text string, for reasons we will learn in Lecture 5, has the type `char *`.

4 Variables

We might want to give a value a name so we can refer to it later. We do this using *variables*. A variable is a named location in memory.

For example, say we wanted to use the value `4 + 2` multiple times. We might call it `x` and use it as follows:

```
1 #include <iostream>
2 using namespace std;
3
4 int main() {
5   int x;
6   x = 4 + 2;
7   cout << x / 3 << ' ' << x * 2;
8
9   return 0;
10 }
```

(Note how we can print a sequence of values by "chaining" the `<<` symbol.)

The name of a variable is an identifier token. Identifiers may contain numbers, letters, and underscores (`_`), and may not start with a number.

Line 5 is the *declaration* of the variable `x`. We must tell the compiler what type `x` will be so that it knows how much memory to reserve for it and what kinds of operations may be performed on it.

Line 6 is the *initialization* of `x`, where we specify an initial value for it. This introduces a new operator: `=`, the assignment operator. We can also change the value of `x` later on in the code using this operator.

We could replace lines 5 and 6 with a single statement that does both declaration and initialization:

```
int x = 4 + 2;
```

This form of declaration/initialization is cleaner, so it is to be preferred.

5 Input

Now that we know how to give names to values, we can have the user of the program input values. This is demonstrated in line 6 below:

```
1 #include <iostream>
2 using namespace std;
3
4 int main() {
5   int x;
6   cin >> x;
7
8   cout << x / 3 << ' ' << x * 2;
9
10   return 0;
11 }
```

Just as `cout <<` is the syntax for outputting values, `cin >>` (line 6) is the syntax for inputting values.

Memory trick: if you have trouble remembering which way the angle brackets go for `cout` and `cin`, think of them as arrows pointing in the direction of data flow. `cin` represents the terminal, with data flowing from it to your variables; `cout` likewise represents the terminal, and your data flows to it.

6 Debugging

There are two kinds of errors you'll run into when writing C++ programs: *compilation errors* and *runtime errors*. Compilation errors are problems raised by the compiler, generally resulting from violations of the syntax rules or misuse of types. These are often caused by typos and the like. Runtime errors are problems that you only spot when you run the program: you did specify a legal program, but it doesn't do what you wanted it to. These are usually more tricky to catch, since the compiler won't tell you about them.

6.096 Introduction to C++ — January 5, 2011

Massachusetts Institute of Technology — John Marrero

Lecture 2 Notes: Flow of Control

1 Motivation

Normally, a program executes statements from first to last. The first statement is executed, then the second, then the third, and so on, until the program reaches its end and terminates. A computer program likely wouldn't be very useful if it ran the same sequence of statements every time it was run. It would be nice to be able to change which statements ran and when, depending on the circumstances. For example, if a program checks a file for the number of times a certain word appears, it should be able to give the correct count no matter what file and word are given to it. Or, a computer game should move the player's character around when the player wants. We need to be able to alter the order in which a program's statements are executed, the *control flow*.

2 Control Structures

Control structures are portions of program code that contain statements within them and, depending on the circumstances, execute these statements in a certain way. There are typically two kinds: *conditionals* and *loops.*

2.1 Conditionals

In order for a program to change its behavior depending on the input, there must a way to test that input. Conditionals allow the program to check the values of variables and to execute (or not execute) certain statements. C++ has *if* and *switch-case* conditional structures.

2.1.1 Operators

Conditionals use two kinds of special operators: *relational* and *logical.* These are used to determine whether some condition is true or false.

The relational operators are used to test a relation between two expressions:

Operator	**Meaning**
>	Greater than
>=	Greater than or equal to
<	Less than
<=	Less than or equal to
==	Equal to
!=	Not equal to

They work the same as the arithmetic operators (e.g., a > b) but return a Boolean value of either `true` or `false`, indicating whether the relation tested for holds. (An expression that returns this kind of value is called a Boolean expression.) For example, if the variables `x` and y have been set to 6 and 2, respectively, then `x > y` returns `true`. Similarly, `x < 5` returns `false`.

The logical operators are often used to combine relational expressions into more complicated Boolean expressions:

Operator	Meaning
`&&`	and
`||`	or
`!`	not

The operators return `true` or `false`, according to the rules of logic:

a	b	a && b
true	true	true
true	false	false
false	true	false
false	false	false

a	b	a || b
true	true	true
true	false	true
false	true	true
false	false	false

The `!` operator is a unary operator, taking only one argument and negating its value:

a	!a
true	false
false	true

Examples using logical operators (assume x = 6 and y = 2):

```
!(x > 2) → false
(x > y) && (y > 0) → true
(x < y) && (y > 0) → false
(x < y) || (y > 0) → true
```

Of course, Boolean variables can be used directly in these expressions, since they hold `true` and `false` values. In fact, any kind of value can be used in a Boolean expression due to a quirk C++ has: `false` is represented by a value of `0` and anything that is not `0` is `true`. So, `"Hello, world!"` is true, `2` is true, and any `int` variable holding a non-zero value is true. This means `!x` returns `false` and `x && y` returns `true`!

2.1.2 *if, if-else* and *else if*

The *if* conditional has the form:

```
if(condition)
{
	statement1
	statement2
	…
}
```

The condition is some expression whose value is being tested. If the condition resolves to a value of `true`, then the statements are executed before the program continues on. Otherwise, the statements are ignored. If there is only one statement, the curly braces may be omitted, giving the form:

```
if(condition)
	statement
```

The *if-else* form is used to decide between two sequences of statements referred to as *blocks*:

```
if(condition)
{
	statementA1
	statementA2
	…
}
else
{
	statementB1
	statementB2
	…
}
```

If the condition is met, the block corresponding to the `if` is executed. Otherwise, the block corresponding to the `else` is executed. Because the condition is either satisfied or not, one of the blocks in an if-else *must* execute. If there is only one statement for any of the blocks, the curly braces for that block may be omitted:

```
if(condition)
	statementA1
else
	statementB1
```

The *else if* is used to decide between two or more blocks based on *multiple* conditions:

```
if(condition1)
{
	statementA1
	statementA2
	…
}
else if(condition2)
{
	statementB1
	statementB2
	…
}
```

If `condition1` is met, the block corresponding to the `if` is executed. If not, then *only if* `condition2` is met is the block corresponding to the `else if` executed. There may be more than one `else if`, each with its own condition. Once a block whose condition was met is executed, any `else if`s after it are ignored. Therefore, in an *if-else-if* structure, either one or no block is executed.

An `else` may be added to the end of an if-else-if. If none of the previous conditions are met, the `else` block is executed. In this structure, one of the blocks *must* execute, as in a normal if-else.

Here is an example using these control structures:

```
1  #include <iostream>
2  using namespace std;
3
4  int main() {
5     int x = 6;
6     int y = 2;
7
8     if(x > y)
9           cout << “x is greater than y\n”;
10    else if(y > x)
11          cout << “y is greater than x\n”;
12    else
13          cout << “x and y are equal\n”;
14
15    return 0;
16 }
```

The output of this program is `x is greater than y`. If we replace lines 5 and 6 with

```
int x = 2;
int y = 6;
```

then the output is `y is greater than x`. If we replace the lines with

```
int x = 2;
int y = 2;
```

then the output is `x and y are equal`.

2.1.3 *switch-case*

The *switch-case* is another conditional structure that may or may not execute certain statements. However, the switch-case has peculiar syntax and behavior:

```
switch(expression)
{
      case constant1:
            statementA1
            statementA2
            ...
            break;
      case constant2:
            statementB1
            statementB2
            ...
            break;
      ...
      default:
            statementZ1
            statementZ2
            ...
}
```

The `switch` evaluates `expression` and, if `expression` is equal to `constant1`, then the statements beneath `case constant 1:` are executed until a `break` is encountered. If `expression` is not equal to `constant1`, then it is compared to `constant2`. If these are equal, then the statements beneath `case constant 2:` are executed until a `break` is encountered. If not, then the same process repeats for each of the constants, in turn. If none of the constants match, then the statements beneath `default:` are executed.

Due to the peculiar behavior of `switch-case`s, curly braces are not necessary for cases where there is more than one statement (but they *are* necessary to enclose the entire `switch-case`). `switch-case`s generally have `if-else` equivalents but can often be a cleaner way of expressing the same behavior.

Here is an example using `switch-case`:

```
1  #include <iostream>
2  using namespace std;
3
4  int main() {
5     int x = 6;
6
7     switch(x) {
8           case 1:
9                 cout << “x is 1\n”;
10                break;
11          case 2:
12          case 3:
13                cout << "x is 2 or 3";
14                break;
15          default:
16                cout << "x is not 1, 2, or 3";
17    }
18
19    return 0;
20 }
```

This program will print `x is not 1, 2, or 3`. If we replace line 5 with `int x = 2;` then the program will print `x is 2 or 3.`

2.2 Loops

Conditionals execute certain statements if certain conditions are met; loops execute certain statements *while* certain conditions are met. C++ has three kinds of loops: *while, do-while,* and *for.*

2.2.1 *while* and *do-while*

The *while* loop has a form similar to the if conditional:

```
while(condition)
{
      statement1
      statement2
      …
}
```

As long as condition holds, the block of statements will be repeatedly executed. If there is only one statement, the curly braces may be omitted. Here is an example:

```
#include <iostream>
using namespace std;

int main() {
   int x = 0;

   while(x < 10)
         x = x + 1;
```

```

    cout << "x is " << x << "\n";

    return 0;
}
```

This program will print `x is 10`.

The *do-while* loop is a variation that guarantees the block of statements will be executed *at least once:*

```
do
{
	statement1
	statement2
	…
}
while(condition);
```

The block of statements is executed and then, if the condition holds, the program returns to the top of the block. Curly braces are *always* required. Also note the semicolon after the `while` condition.

2.2.2 *for*

The *for* loop works like the while loop but with some change in syntax:

```
for(initialization; condition; incrementation)
{
	statement1
	statement2
	…
}
```

The for loop is designed to allow a counter variable that is initialized at the beginning of the loop and incremented (or decremented) on each iteration of the loop. Curly braces may be omitted if there is only one statement. Here is an example:

```
#include <iostream>
using namespace std;

int main() {

    for(int x = 0; x < 10; x = x + 1)
        cout << x << "\n";

    return 0;
}
```

This program will print out the values `0` through `9`, each on its own line.

If the counter variable is already defined, there is no need to define a new one in the initialization portion of the for loop. Therefore, it is valid to have the following:

```
#include <iostream>
using namespace std;

int main() {

    int x = 0;
    for(; x < 10; x = x + 1)
          cout << x << "\n”;

    return 0;
}
```

Note that the first semicolon inside the for loop's parentheses is still required.

A for loop can be expressed as a while loop and vice-versa. Recalling that a for loop has the form

```
for(initialization; condition; incrementation)
{
      statement1
      statement2
      …
}
```

we can write an equivalent while loop as

```
initialization
while(condition)
{
      statement1
      statement2
      …
      incrementation
}
```

Using our example above,

```
#include <iostream>
using namespace std;

int main() {

    for(int x = 0; x < 10; x = x + 1)
          cout << x << "\n”;

    return 0;
}
```

is converted to

```
#include <iostream>
using namespace std;

int main() {

    int x = 0;
    while(x < 10) {
          cout << x << "\n”;
          x = x + 1;
    }

    return 0;
}
```

The incrementation step can technically be anywhere inside the statement block, but it is good practice to place it as the last step, particularly if the previous statements use the current value of the counter variable.

2.3 Nested Control Structures

It is possible to place ifs inside of ifs and loops inside of loops by simply placing these structures inside the statement blocks. This allows for more complicated program behavior.

Here is an example using nesting if conditionals:

```
#include <iostream>
using namespace std;

int main() {
   int x = 6;
   int y = 0;

   if(x > y) {
         cout << “x is greater than y\n”;
         if(x == 6)
               cout << “x is equal to 6\n”;
         else
               cout << “x is not equalt to 6\n”;
   } else
         cout << “x is not greater than y\n”;

   return 0;
}
```

This program will print `x is greater than y` on one line and then `x is equal to 6` on the next line.

Here is an example using nested loops:

```
#include <iostream>
using namespace std;

int main() {
   for(int x = 0; x < 4; x = x + 1) {
         for(int y = 0; y < 4; y = y + 1)
               cout << y;
         cout << “\n”;
   }

   return 0;
}
```

This program will print four lines of `0123`.

6.096 Lecture 3: Functions

How to reuse code

Geza Kovacs

```cpp
#include <iostream>
using namespace std;

int main() {
  int threeExpFour = 1;
  for (int i = 0; i < 4; i = i + 1) {
    threeExpFour = threeExpFour * 3;
  }
  cout << "3^4 is " << threeExpFour << endl;
  return 0;
}
```

Copy-paste coding

```
#include <iostream>
using namespace std;

int main() {
  int threeExpFour = 1;
  for (int i = 0; i < 4; i = i + 1) {
    threeExpFour = threeExpFour * 3;
  }
  cout << "3^4 is " << threeExpFour << endl;
  int sixExpFive = 1;
  for (int i = 0; i < 5; i = i + 1) {
    sixExpFive = sixExpFive * 6;
  }
  cout << "6^5 is " << sixExpFive << endl;
  return 0;
}
```

Copy-paste coding (bad)

```cpp
#include <iostream>
using namespace std;

int main() {
  int threeExpFour = 1;
  for (int i = 0; i < 4; i = i + 1) {
    threeExpFour = threeExpFour * 3;
  }
  cout << "3^4 is " << threeExpFour << endl;
  int sixExpFive = 1;
  for (int i = 0; i < 5; i = i + 1) {
    sixExpFive = sixExpFive * 6;
  }
  cout << "6^5 is " << sixExpFive << endl;
  int twelveExpTen = 1;
  for (int i = 0; i < 10; i = i + 1) {
    twelveExpTen = twelveExpTen * 12;
  }
  cout << "12^10 is " << twelveExpTen << endl;
  return 0;
}
```

With a function

```
#include <iostream>
using namespace std;

// some code which raises an arbitrary integer
// to an arbitrary power

int main() {
  int threeExpFour = raiseToPower(3, 4);
  cout << "3^4 is " << threeExpFour << endl;
  return 0;
}
```

With a function

```
#include <iostream>
using namespace std;

// some code which raises an arbitrary integer
// to an arbitrary power

int main() {
  int threeExpFour = raiseToPower(3, 4);
  cout << "3^4 is " << threeExpFour << endl;
  int sixExpFive = raiseToPower(6, 5);
  cout << "6^5 is " << sixExpFive << endl;
  return 0;
}
```

With a function

```
#include <iostream>
using namespace std;

// some code which raises an arbitrary integer
// to an arbitrary power

int main() {
  int threeExpFour = raiseToPower(3, 4);
  cout << "3^4 is " << threeExpFour << endl;
  int sixExpFive = raiseToPower(6, 5);
  cout << "6^5 is " << sixExpFive << endl;
  int twelveExpTen = raiseToPower(12, 10);
  cout << "12^10 is " << twelveExpTen << endl;
  return 0;
}
```

Why define your own functions?

- Readability: sqrt(5) is clearer than copy-pasting in an algorithm to compute the square root
- Maintainability: To change the algorithm, just change the function (vs changing it everywhere you ever used it)
- Code reuse: Lets other people use algorithms you've implemented

Function Declaration Syntax

Function name

```
int raiseToPower(int base, int exponent)
{
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}
```

Function Declaration Syntax

Return type

```
int raiseToPower(int base, int exponent)
{
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}
```

Function Declaration Syntax

Argument 1

```
int raiseToPower(int base, int exponent)
{
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}
```

- Argument order matters:
  - raiseToPower(2,3) is 2^3=8
  - raiseToPower(3,2) is 3^2=9

Function Declaration Syntax

Argument 2

```
int raiseToPower(int base, int exponent)
{
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}
```

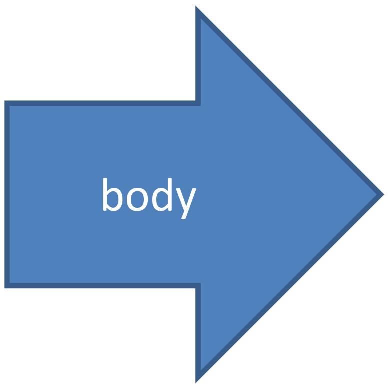
- Argument order matters:
  - raiseToPower(2,3) is 2^3=8
  - raiseToPower(3,2) is 3^2=9

Function Declaration Syntax

signature

```
int raiseToPower(int base, int exponent)
{
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}
```

Function Declaration Syntax




```
int raiseToPower(int base, int exponent)
{
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}
```

Function Declaration Syntax

```
int raiseToPower(int base, int exponent)
{
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}
```

Return statement

Function declaration

```
#include <iostream>
using namespace std;

int raiseToPower(int base, int exponent) {
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}

int main() {
  int threeExpFour = raiseToPower(3, 4);
  cout << "3^4 is " << threeExpFour << endl;
  return 0;
}
```

Function invocation

Returning a value

- Up to one value may be returned; it must be the same type as the return type.

```
int foo()
{
  return "hello"; // error
}
```

```
char* foo()
{
  return "hello"; // ok
}
```

Returning a value

- Up to one value may be returned; it must be the same type as the return type.
- If no values are returned, give the function a **void** return type

```
void printNumber(int num) {
  cout << "number is " << num << endl;
}

int main() {
  printNumber(4); // number is 4
  return 0;
}
```

Returning a value

- Up to one value may be returned; it must be the same type as the return type.
- If no values are returned, give the function a **void** return type
  - Note that you cannot declare a variable of type void

```
int main() {
  void x; // ERROR
  return 0;
}
```

Returning a value

- Return statements don’t necessarily need to be at the end.
- Function returns as soon as a return statement is executed.

```
void printNumberIfEven(int num) {
  if (num % 2 == 1) {
    cout << "odd number" << endl;
    return;
  }
  cout << "even number; number is " << num << endl;
}

int main() {
  int x = 4;
  printNumberIfEven(x);
  // even number; number is 3
  int y = 5;
  printNumberIfEven(y);
  // odd number
}
```

Argument Type Matters

```cpp
void printOnNewLine(int x)
{
    cout << x << endl;
}
```

- `printOnNewLine(3)` works
- `printOnNewLine("hello")` will not compile

Argument Type Matters

```
void printOnNewLine(char *x)
{
    cout << x << endl;
}
```

- `printOnNewLine(3)` will not compile
- `printOnNewLine("hello")` works

Argument Type Matters

```
void printOnNewLine(int x)
{
    cout << x << endl;
}

void printOnNewLine(char *x)
{
    cout << x << endl;
}
```

- `printOnNewLine(3)` works
- `printOnNewLine("hello")` also works

Function Overloading

```
void printOnNewLine(int x)
{
    cout << "Integer: " << x << endl;
}

void printOnNewLine(char *x)
{
    cout << "String: " << x << endl;
}
```

- Many functions with the same name, but different arguments
- The function called is the one whose arguments match the invocation

Function Overloading

```
void printOnNewLine(int x)
{
    cout << "Integer: " << x << endl;
}

void printOnNewLine(char *x)
{
    cout << "String: " << x << endl;
}
```

- printOnNewLine(3) prints “Integer: 3”
- printOnNewLine(“hello”) prints “String: hello”

Function Overloading

```
void printOnNewLine(int x)
{
    cout << "1 Integer: " << x << endl;
}

void printOnNewLine(int x, int y)
{
    cout << "2 Integers: " << x << " and " << y << endl;
}
```

- printOnNewLine(3) prints “1 Integer: 3”
- printOnNewLine(2, 3) prints “2 Integers: 2 and 3”

- Function declarations need to occur before invocations

```
int foo()
{
    return bar()*2; // ERROR - bar hasn’t been declared yet
}

int bar()
{
    return 3;
}
```

- Function declarations need to occur before invocations
  - Solution 1: reorder function declarations

```
int bar()
{
    return 3;
}

int foo()
{
    return bar()*2; // ok
}
```

- Function declarations need to occur before invocations
  - Solution 1: reorder function declarations
  - Solution 2: use a function prototype; informs the compiler you'll implement it later

function prototype

```
int bar();

int foo()
{
    return bar()*2; // ok
}

int bar()
{
    return 3;
}
```

- Function prototypes should match the signature of the method, though argument names don't matter

```
int square(int);
```

function prototype

```
int cube(int x)
{
    return x*square(x);
}

int square(int x)
{
    return x*x;
}
```

- Function prototypes should match the signature of the method, though argument names don't matter

```
int square(int x);
```

function prototype

```
int cube(int x)
{
    return x*square(x);
}

int square(int x)
{
    return x*x;
}
```

- Function prototypes should match the signature of the method, though argument names don’t matter

```
int square(int z);
```

function prototype

```
int cube(int x)
{
    return x*square(x);
}

int square(int x)
{
    return x*x;
}
```

- Function prototypes are generally put into separate header files
  - Separates specification of the function from its implementation

```
// myLib.h - header
// contains prototypes

int square(int);
int cube (int);
```

```
// myLib.cpp - implementation
#include "myLib.h"

int cube(int x)
{
    return x*square(x);
}

int square(int x)
{
    return x*x;
}
```

Recursion

- Functions can call themselves.
- fib(n) = fib(n-1) + fib(n-2) can be easily expressed via a recursive implementation

```
int fibonacci(int n) {
  if (n == 0 || n == 1) {
    return 1;
  } else {
    return fibonacci(n-2) + fibonacci(n-1);
  }
}
```

Recursion

- Functions can call themselves.
- fib(n) = fib(n-1) + fib(n-2) can be easily expressed via a recursive implementation

base case

```
int fibonacci(int n) {
  if (n == 0 || n == 1) {
    return 1;
  } else {
    return fibonacci(n-2) + fibonacci(n-1);
  }
}
```

Recursion

- Functions can call themselves.
- fib(n) = fib(n-1) + fib(n-2) can be easily expressed via a recursive implementation

recursive step

```
int fibonacci(int n) {
  if (n == 0 || n == 1) {
    return 1;
  } else {
    return fibonacci(n-2) + fibonacci(n-1);
  }
}
```

Global Variables

- How many times is function foo() called? Use a global variable to determine this.
  - Can be accessed from any function

```
int numCalls = 0;

void foo() {
  ++numCalls;
}

int main() {
  foo(); foo(); foo();
  cout << numCalls << endl; // 3
}
```

Global variable

Scope

- Scope: where a variable was declared, determines where it can be accessed from

```
int numCalls = 0;

int raiseToPower(int base, int exponent) {
  numCalls = numCalls + 1;
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}

int max(int num1, int num2) {
  numCalls = numCalls + 1;
  int result;
  if (num1 > num2) {
    result = num1;
  }
  else {
    result = num2;
  }
  return result;
}
```

Scope

- Scope: where a variable was declared, determines where it can be accessed from
- numCalls has global scope – can be accessed from any function

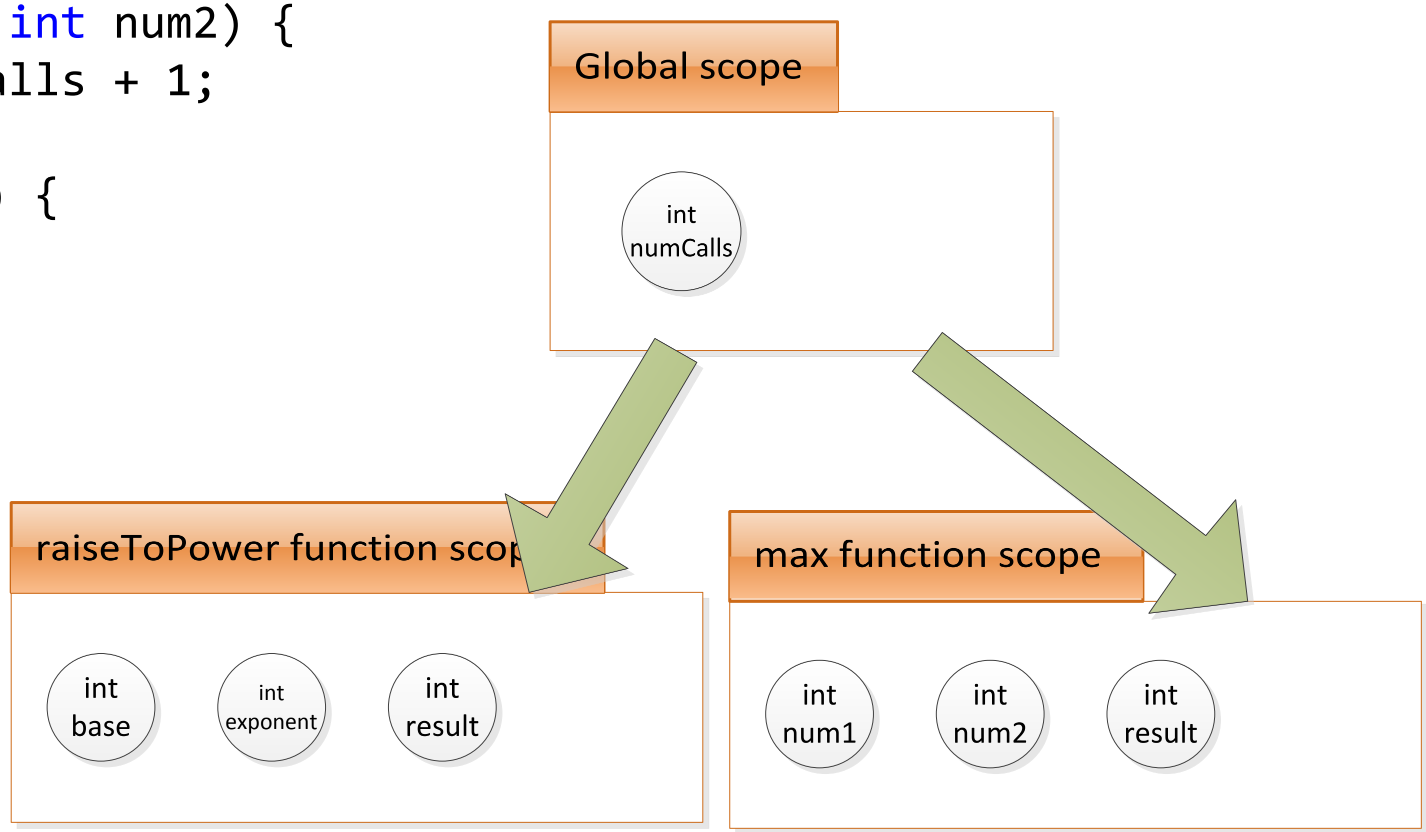
```
int numCalls = 0;

int raiseToPower(int base, int exponent) {
  numCalls = numCalls + 1;
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}

int max(int num1, int num2) {
  numCalls = numCalls + 1;
  int result;
  if (num1 > num2) {
    result = num1;
  }
  else {
    result = num2;
  }
  return result;
}
```

Scope

- Scope: where a variable was declared, determines where it can be accessed from
- numCalls has global scope – can be accessed from any function
- result has function scope – each function can have its own separate variable named result

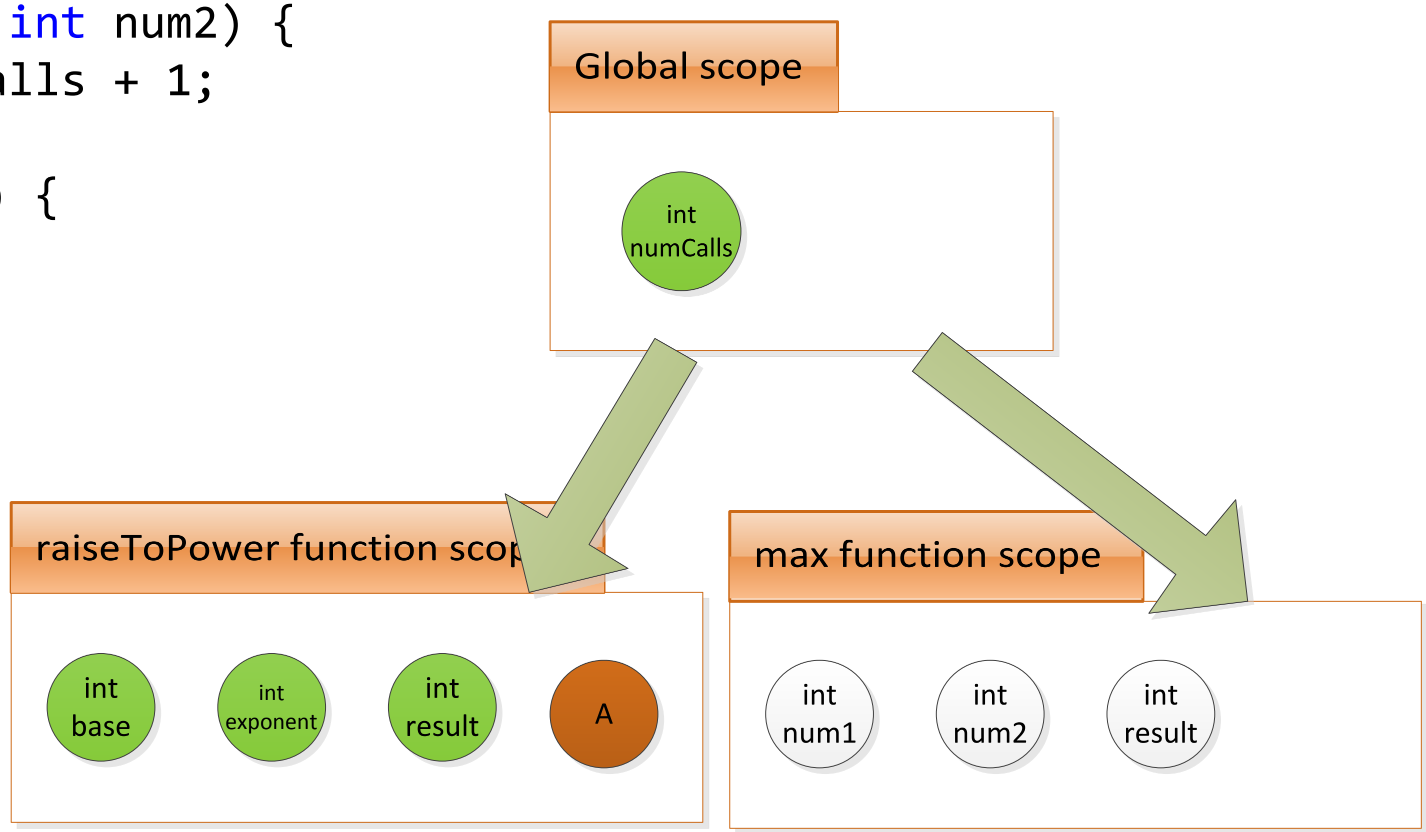
```
int numCalls = 0;

int raiseToPower(int base, int exponent) {
  numCalls = numCalls + 1;
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  return result;
}

int max(int num1, int num2) {
  numCalls = numCalls + 1;
  int result;
  if (num1 > num2) {
    result = num1;
  }
  else {
    result = num2;
  }
  return result;
}
```

```
int numCalls = 0;
int raiseToPower(int base, int exponent) {
  numCalls = numCalls + 1;
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  // A
  return result;
}
int max(int num1, int num2) {
  numCalls = numCalls + 1;
  int result;
  if (num1 > num2) {
    result = num1;
  }
  else {
    result = num2;
  }
  // B
  return result;
}
```



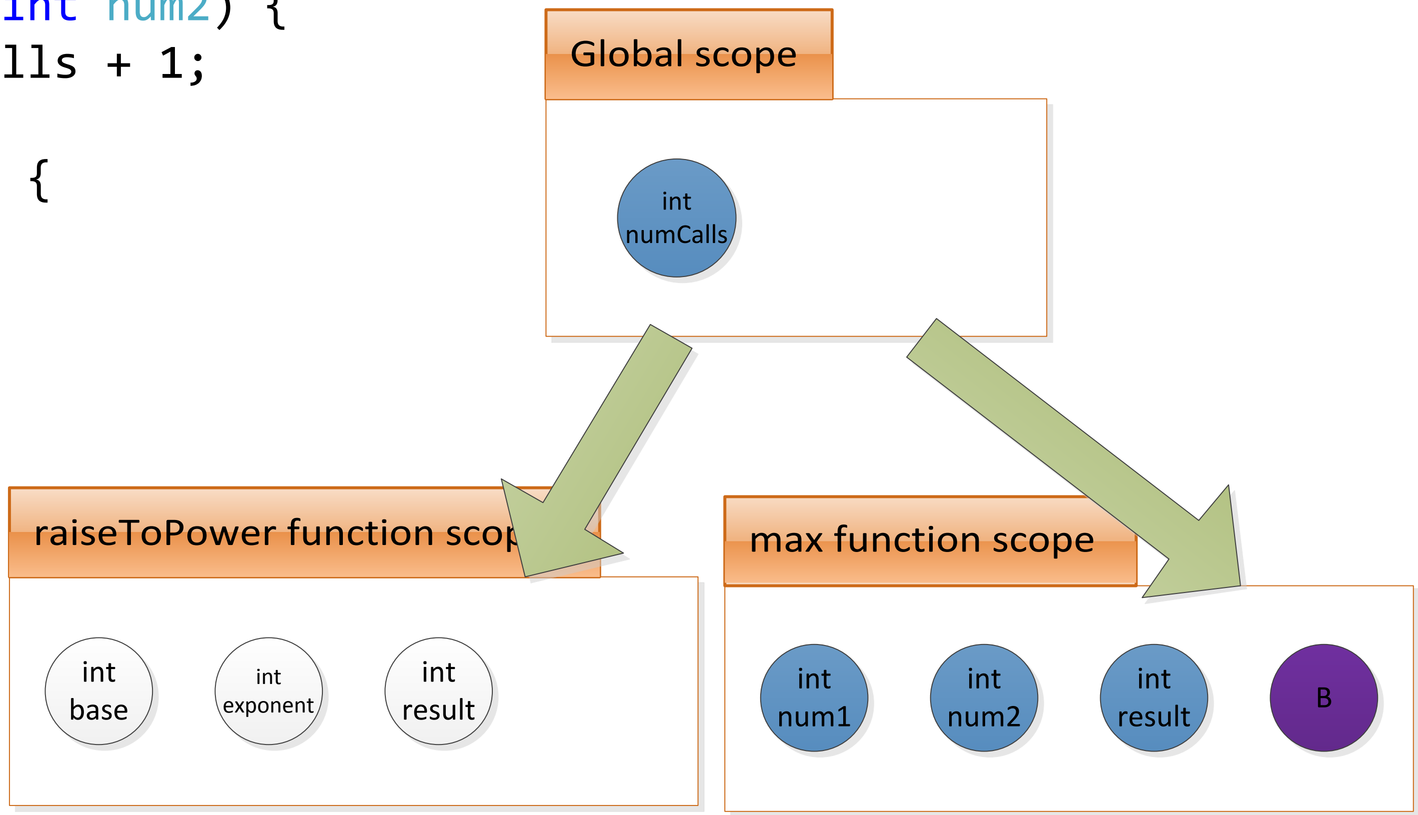
```
int numCalls = 0;
int raiseToPower(int base, int exponent) {
  numCalls = numCalls + 1;
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  // A
  return result;
}
int max(int num1, int num2) {
  numCalls = numCalls + 1;
  int result;
  if (num1 > num2) {
    result = num1;
  }
  else {
    result = num2;
  }
  // B
  return result;
}
```

- At A, variables marked in green are in scope

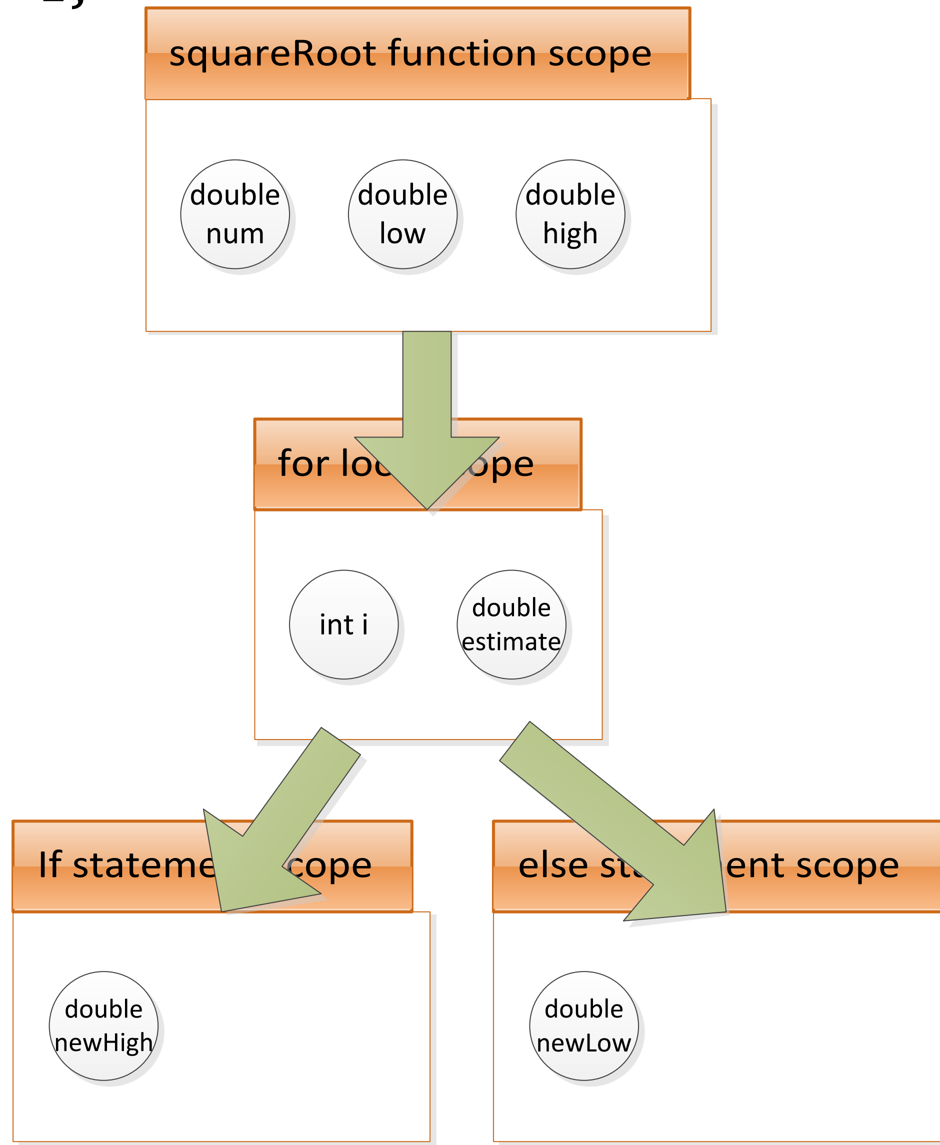


```
int numCalls = 0;
int raiseToPower(int base, int exponent) {
  numCalls = numCalls + 1;
  int result = 1;
  for (int i = 0; i < exponent; i = i + 1) {
    result = result * base;
  }
  // A
  return result;
}
int max(int num1, int num2) {
  numCalls = numCalls + 1;
  int result;
  if (num1 > num2) {
    result = num1;
  }
  else {
    result = num2;
  }
  // B
  return result;
}
```

- At B, variables marked in blue are in scope



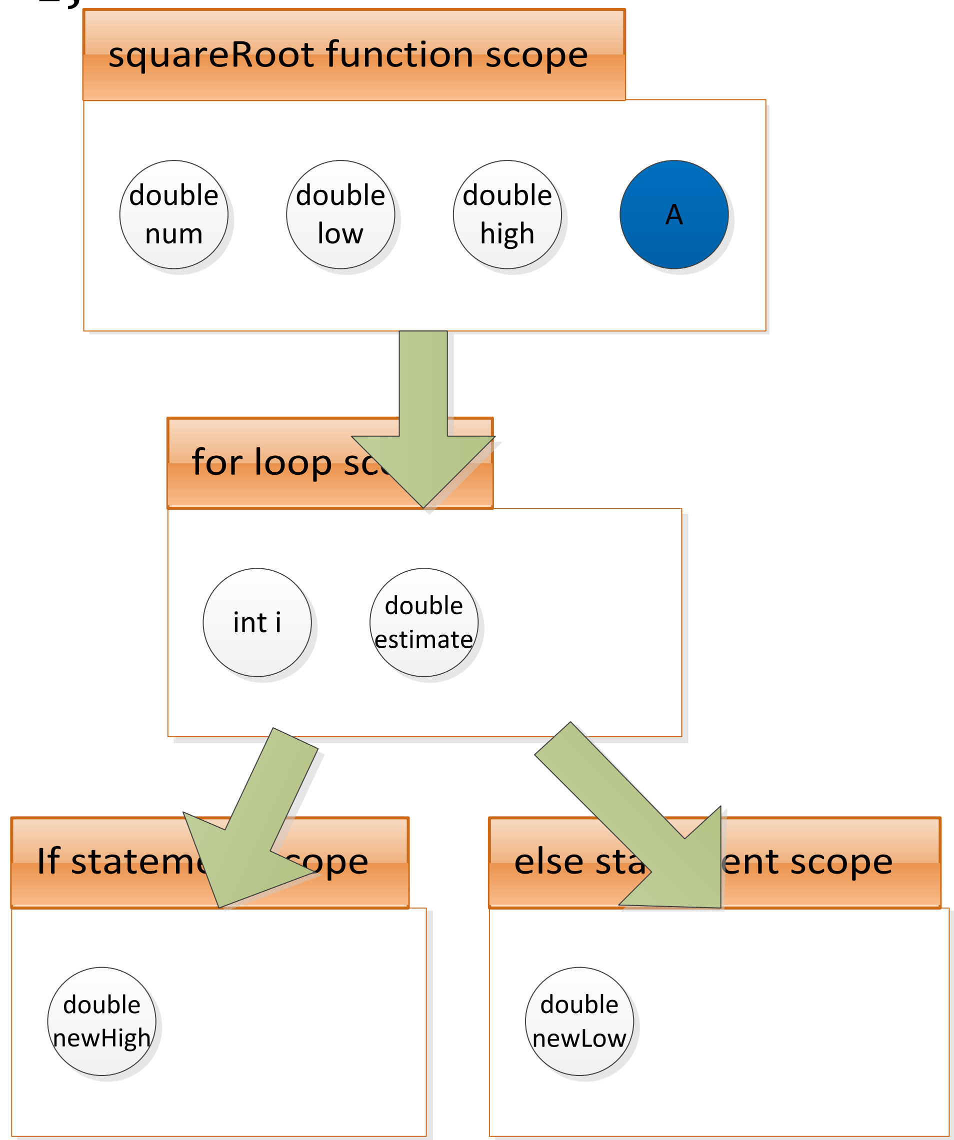
```
double squareRoot(double num) {
  double low = 1.0;
  double high = num;
  for (int i = 0; i < 30; i = i + 1) {
    double estimate = (high + low) / 2;
    if (estimate*estimate > num) {
      double newHigh = estimate;
      high = newHigh;
    } else {
      double newLow = estimate;
      low = newLow;
    }
  }
  return (high + low) / 2;
}
```



- Loops and if/else statements also have their own scopes
  - Loop counters are in the same scope as the body of the for loop

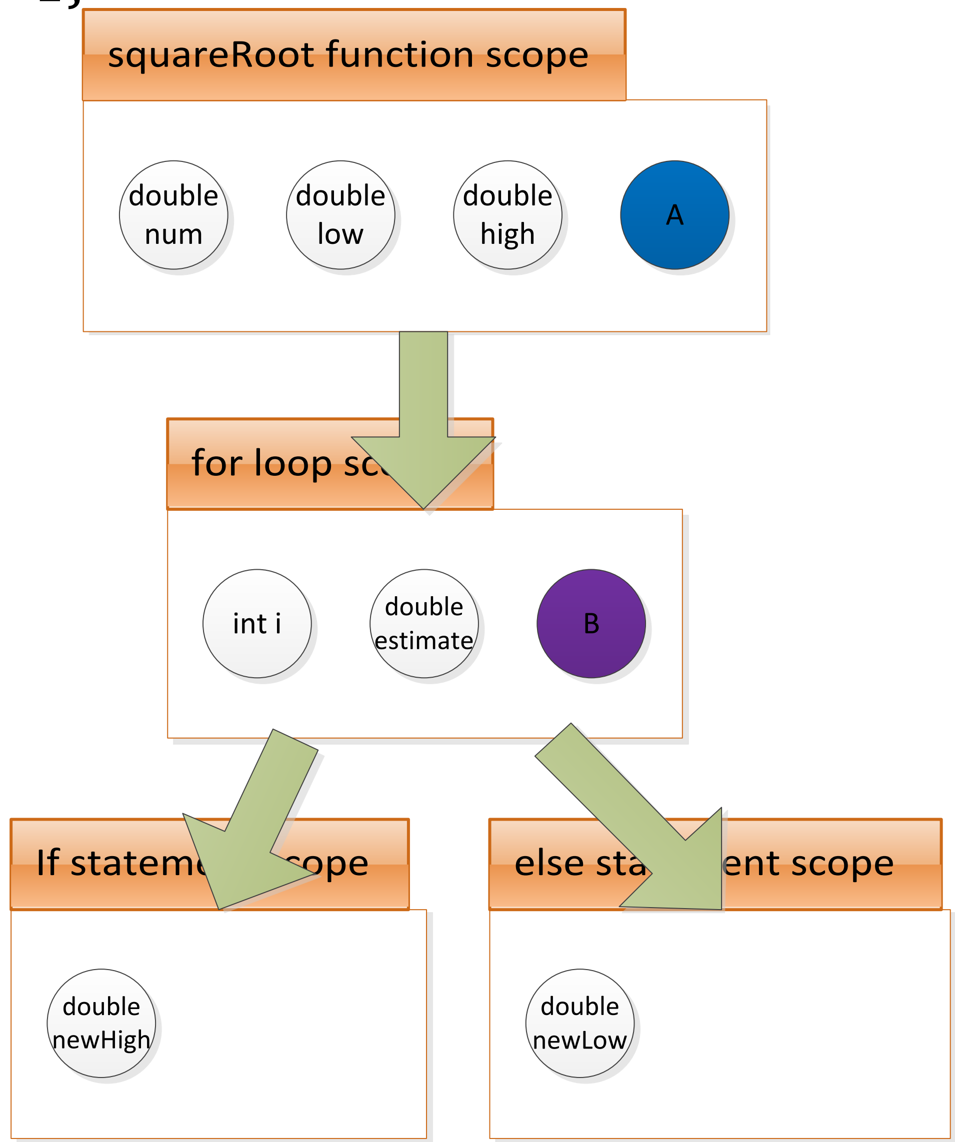
```
double squareRoot(double num) {
  double low = 1.0;
  double high = num;
  for (int i = 0; i < 30; i = i + 1) {
    double estimate = (high + low) / 2;
    if (estimate*estimate > num) {
      double newHigh = estimate;
      high = newHigh;
    } else {
      double newLow = estimate;
      low = newLow;
    }
  }
  // A
  return estimate; // ERROR
}
```

- Cannot access variables that are out of scope

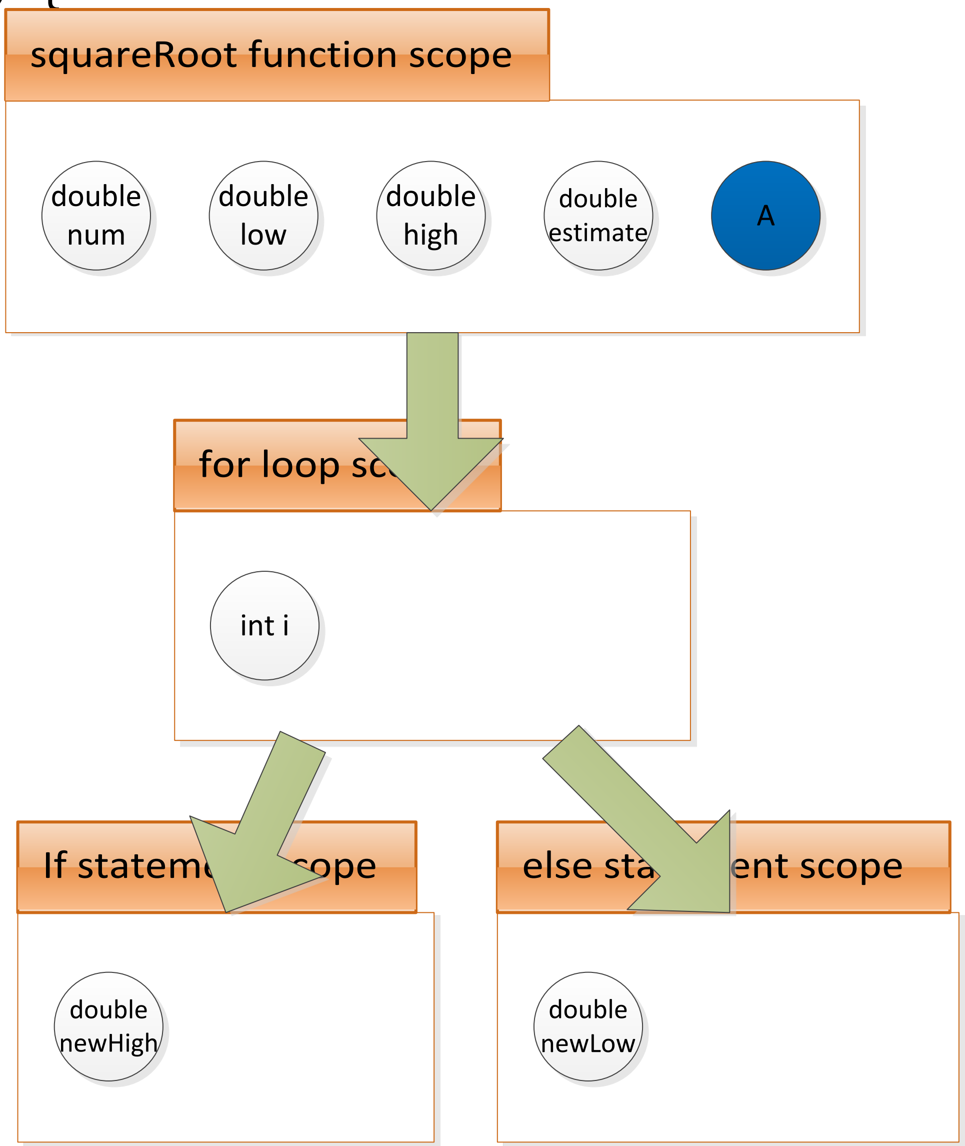


```
double squareRoot(double num) {
  double low = 1.0;
  double high = num;
  for (int i = 0; i < 30; i = i + 1) {
    double estimate = (high + low) / 2;
    if (estimate*estimate > num) {
      double newHigh = estimate;
      high = newHigh;
    } else {
      double newLow = estimate;
      low = newLow;
    }
    if (i == 29)
      return estimate; // B
  }
  return -1; // A
}
```

- Cannot access variables that are out of scope
- Solution 1: move the code



```
double squareRoot(double num) {
  double low = 1.0;
  double high = num;
  double estimate;
  for (int i = 0; i < 30; i = i + 1) {
    estimate = (high + low) / 2;
    if (estimate*estimate > num) {
      double newHigh = estimate;
      high = newHigh;
    } else {
      double newLow = estimate;
      low = newLow;
    }
  }
  return estimate; // A
}
```



- Cannot access variables that are out of scope
- Solution 2: declare the variable in a higher scope

Pass by value vs by reference

- So far we've been passing everything by value – makes a copy of the variable; changes to the variable within the function don't occur outside the function

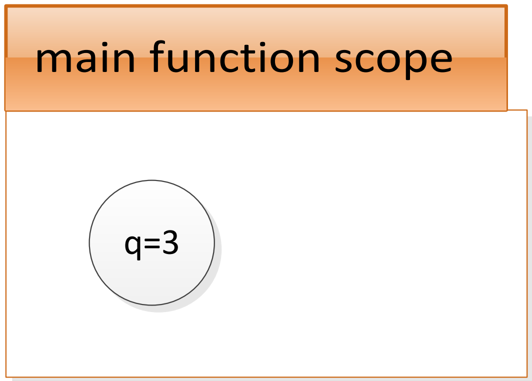
```
// pass-by-value
void increment(int a) {
  a = a + 1;
  cout << "a in increment " << a << endl;
}

int main() {
  int q = 3;
  increment(q); // does nothing
  cout << "q in main " << q << endl;
}
```

Output

a in increment 4

q in main **3**

Pass by value vs by reference



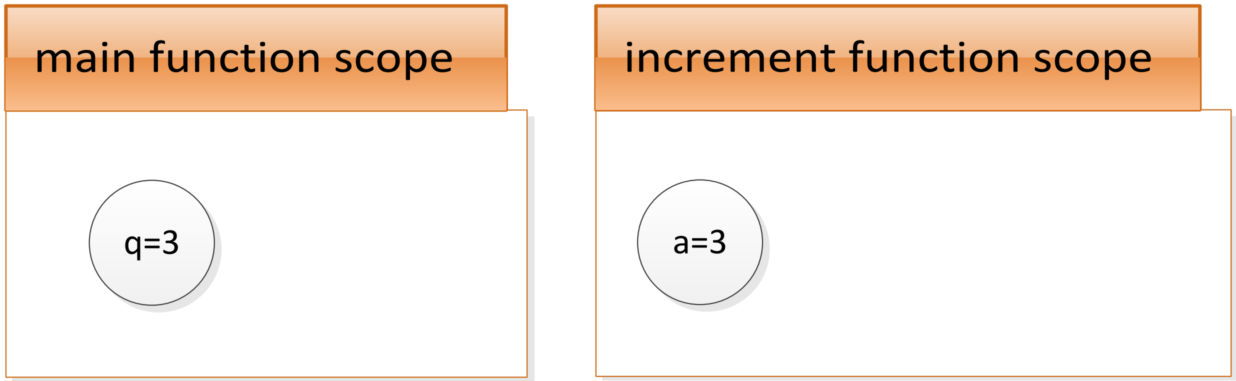

```
// pass-by-value
void increment(int a) {
  a = a + 1;
  cout << "a in increment " << a << endl;
}

int main() {
  int q = 3; // HERE
  increment(q); // does nothing
  cout << "q in main " << q << endl;
}
```

Output

a in increment 4

q in main **3**

Pass by value vs by reference



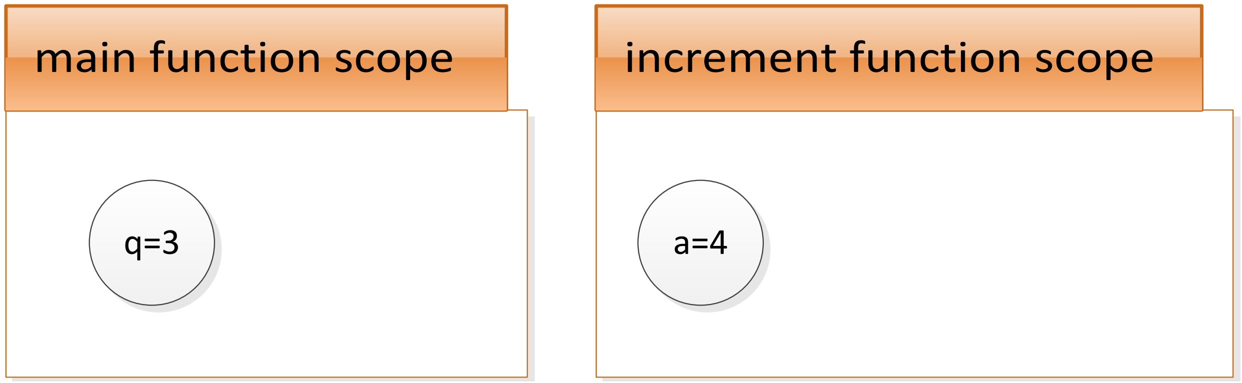

```
// pass-by-value
void increment(int a) { // HERE
  a = a + 1;
  cout << "a in increment " << a << endl;
}

int main() {
  int q = 3;
  increment(q); // does nothing
  cout << "q in main " << q << endl;
}
```

Output

a in increment 4

q in main **3**

Pass by value vs by reference




```
// pass-by-value
void increment(int a) {
  a = a + 1; // HERE
  cout << "a in increment " << a << endl;
}

int main() {
  int q = 3;
  increment(q); // does nothing
  cout << "q in main " << q << endl;
}
```

Output

a in increment 4

q in main **3**

Pass by value vs by reference

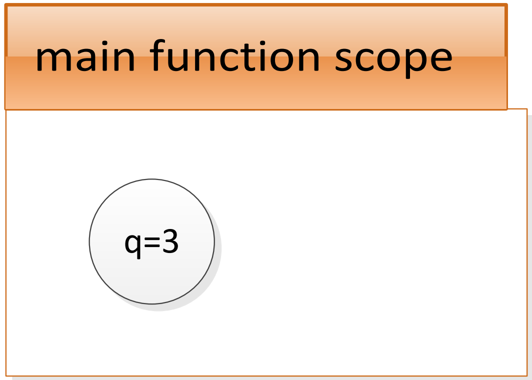
- If you want to modify the original variable as opposed to making a copy, pass the variable by reference (**int &a** instead of **int a**)

```
// pass-by-value
void increment(int &a) {
  a = a + 1;
  cout << "a in increment " << a << endl;
}

int main() {
  int q = 3;
  increment(q); // works
  cout << "q in main " << q << endl;
}
```

Output

a in increment 4
q in main **4**

Pass by value vs by reference



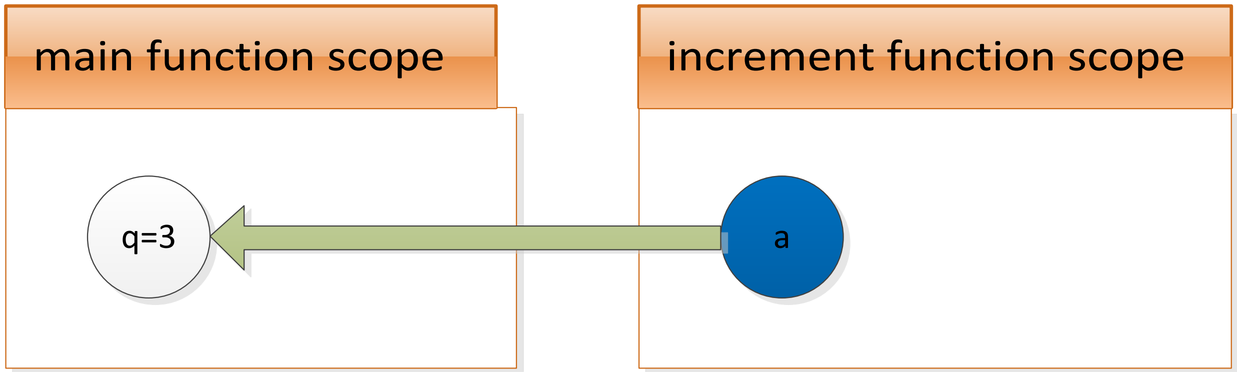

```
// pass-by-value
void increment(int &a) {
  a = a + 1;
  cout << "a in increment " << a << endl;
}

int main() {
  int q = 3; // HERE
  increment(q); // works
  cout << "q in main " << q << endl;
}
```

Output

a in increment 4

q in main **4**

Pass by value vs by reference



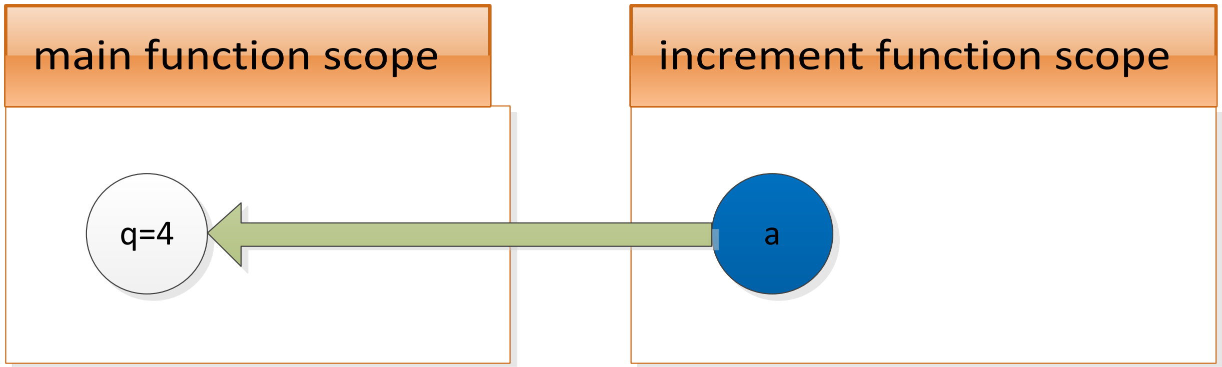

```
// pass-by-value
void increment(int &a) { // HERE
  a = a + 1;
  cout << "a in increment " << a << endl;
}

int main() {
  int q = 3;
  increment(q); // works
  cout << "q in main " << q << endl;
}
```

Output

a in increment 4

q in main **4**

Pass by value vs by reference




```
// pass-by-value
void increment(int &a) {
  a = a + 1; // HERE
  cout << "a in increment " << a << endl;
}

int main() {
  int q = 3;
  increment(q); // works
  cout << "q in main " << q << endl;
}
```

Output

a in increment 4

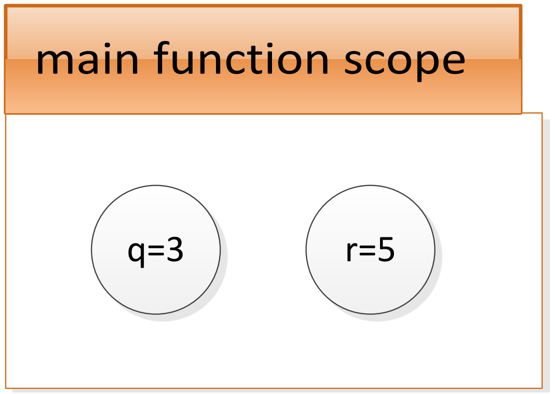
q in main **4**

Implementing Swap

```
void swap(int &a, int &b) {
  int t = a;
  a = b;
  b = t;
}

int main() {
  int q = 3;
  int r = 5;
  swap(q, r);
  cout << "q " << q << endl; // q 5
  cout << "r " << r << endl; // r 3
}
```

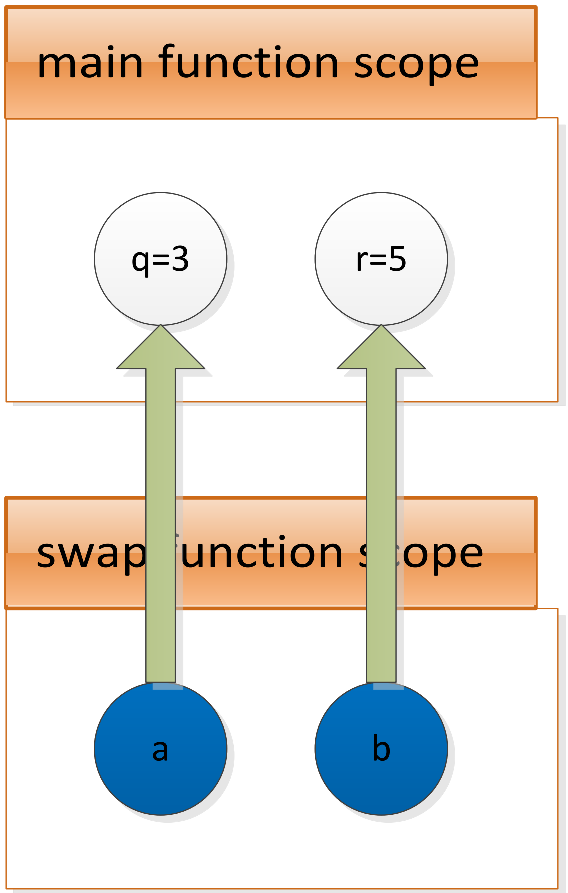
Implementing Swap

```
void swap(int &a, int &b) {
  int t = a;
  a = b;
  b = t;
}

int main() {
  int q = 3;
  int r = 5; // HERE
  swap(q, r);
  cout << "q " << q << endl; // q 5
  cout << "r " << r << endl; // r 3
}
```



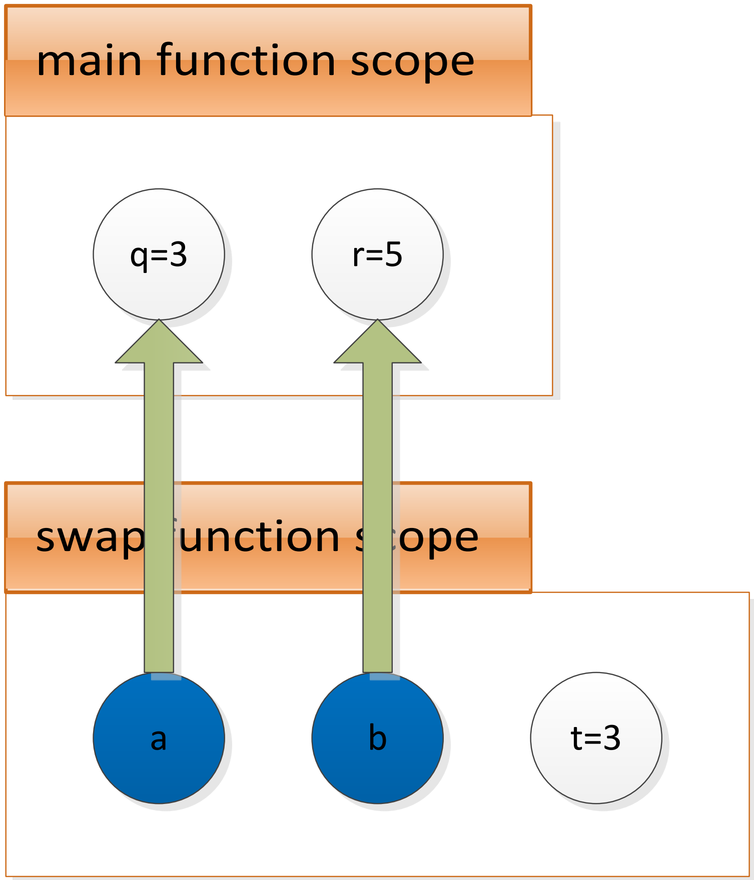
Implementing Swap

```
void swap(int &a, int &b) { // HERE
  int t = a;
  a = b;
  b = t;
}

int main() {
  int q = 3;
  int r = 5;
  swap(q, r);
  cout << "q " << q << endl; // q 5
  cout << "r " << r << endl; // r 3
}
```



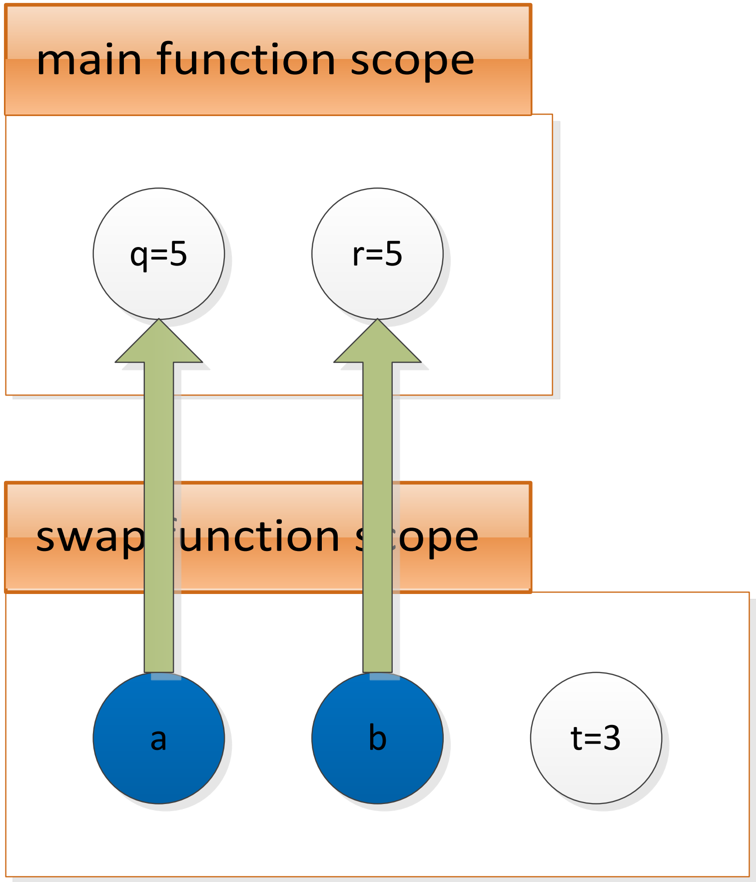
Implementing Swap

```
void swap(int &a, int &b) {
  int t = a; // HERE
  a = b;
  b = t;
}

int main() {
  int q = 3;
  int r = 5;
  swap(q, r);
  cout << "q " << q << endl; // q 5
  cout << "r " << r << endl; // r 3
}
```



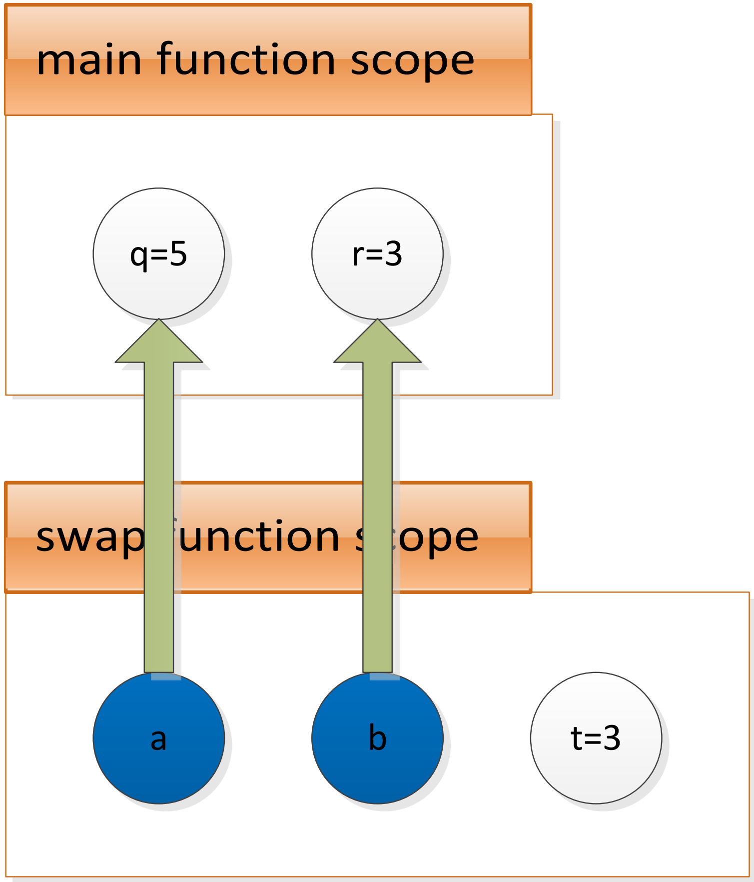
Implementing Swap

```
void swap(int &a, int &b) {
  int t = a;
  a = b; // HERE
  b = t;
}

int main() {
  int q = 3;
  int r = 5;
  swap(q, r);
  cout << "q " << q << endl; // q 5
  cout << "r " << r << endl; // r 3
}
```



Implementing Swap

```
void swap(int &a, int &b) {
  int t = a;
  a = b;
  b = t; // HERE
}

int main() {
  int q = 3;
  int r = 5;
  swap(q, r);
  cout << "q " << q << endl; // q 5
  cout << "r " << r << endl; // r 3
}
```



Returning multiple values

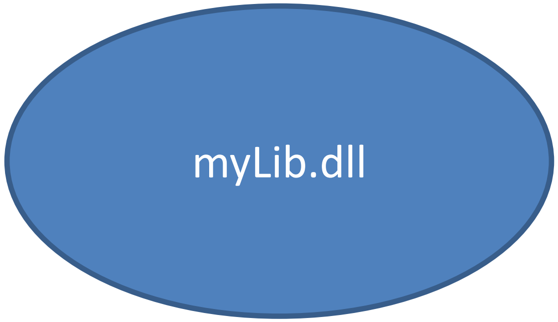
- The return statement only allows you to return 1 value. Passing output variables by reference overcomes this limitation.

```
int divide(int numerator, int denominator, int &remainder) {
  remainder = numerator % denominator;
  return numerator / denominator;
}

int main() {
  int num = 14;
  int den = 4;
  int rem;
  int result = divide(num, den, rem);
  cout << result << "*" << den << "+" << rem << "=" << num << endl;
  // 3*4+2=12
}
```

Libraries

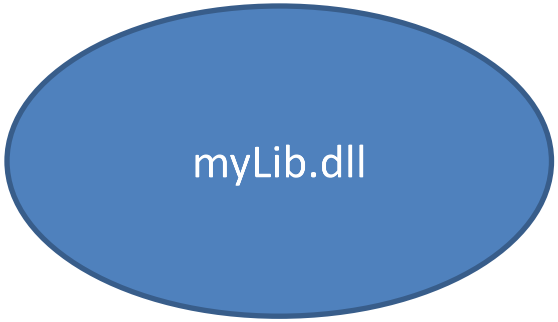
- Libraries are generally distributed as the header file containing the prototypes, and a binary .dll/.so file containing the (compiled) implementation
  - Don’t need to share your .cpp code

```
// myLib.h – header
// contains prototypes
double squareRoot(double num);
```



- Library user only needs to know the function prototypes (in the header file), not the implementation source code (in the .cpp file)
  - The **Linker** (part of the compiler) takes care of locating the implementation of functions in the .dll file at compile time

```
// myLib.h – header
// contains prototypes
double squareRoot(double num);
```




```
// libraryUser.cpp – some other guy’s code
#include "myLib.h"

double fourthRoot(double num) {
  return squareRoot(squareRoot(num));
}
```

Final Notes

- You don’t actually need to implement raiseToPower and squareRoot yourself; cmath (part of the standard library) contains functions **pow** and **sqrt**

```
#include <cmath>

double fourthRoot(double num) {
  return sqrt(sqrt(num));
}
```

6.096 Introduction to C++ January 10, 2011
Massachusetts Institute of Technology John Marrero

Lecture 4 Notes: Arrays and Strings

1 Arrays

So far we have used variables to store values in memory for later reuse. We now explore a means to store *multiple* values together as one unit, the *array.*

An array is a fixed number of *elements* of the same type stored sequentially in memory. Therefore, an integer array holds some number of integers, a character array holds some number of characters, and so on. The size of the array is referred to as its *dimension.* To declare an array in C++, we write the following:

```
type arrayName[dimension];
```

To declare an integer array named `arr` of four elements, we write `int arr[4];`

The elements of an array can be accessed by using an *index* into the array. Arrays in C++ are zero-indexed, so the first element has an index of 0. So, to access the third element in arr, we write `arr[2];` The value returned can then be used just like any other integer.

Like normal variables, the elements of an array must be initialized before they can be used; otherwise we will almost certainly get unexpected results in our program. There are several ways to initialize the array. One way is to declare the array and then initialize some or all of the elements:

```
int arr[4];

arr[0] = 6;
arr[1] = 0;
arr[2] = 9;
arr[3] = 6;
```

Another way is to initialize some or all of the values at the time of declaration:

```
int arr[4] = { 6, 0, 9, 6 };
```

Sometimes it is more convenient to leave out the size of the array and let the compiler determine the array's size for us, based on how many elements we give it:

```
int arr[] = { 6, 0, 9, 6, 2, 0, 1, 1 };
```

Here, the compiler will create an integer array of dimension 8.

The array can also be initialized with values that are not known beforehand:

```
#include <iostream>
using namespace std;

int main() {

   int arr[4];
   cout << “Please enter 4 integers:” << endl;

   for(int i = 0; i < 4; i++)
         cin >> arr[i];

```

```
    cout << "Values in array are now:";

    for(int i = 0; i < 4; i++)
          cout << " " << arr[i];

    cout << endl;

    return 0;
}
```

Note that when accessing an array the index given must be a positive integer from 0 to n-1, where n is the dimension of the array. The index itself may be directly provided, derived from a variable, or computed from an expression:

```
arr[5];
arr[i];
arr[i+3];
```

Arrays can also be passed as arguments to functions. When declaring the function, simply specify the array as a parameter, without a dimension. The array can then be used as normal within the function. For example:

```
#include <iostream>
using namespace std;

int sum(const int array[], const int length) {
    long sum = 0;
    for(int i = 0; i < length; sum += array[i++]);
    return sum;
}

int main() {
    int arr[] = {1, 2, 3, 4, 5, 6, 7};
    cout << "Sum: " << sum(arr, 7) << endl;
    return 0;
}
```

The function `sum` takes a constant integer array and a constant integer length as its arguments and adds up `length` elements in the array. It then returns the sum, and the program prints out `Sum: 28`.

It is important to note that arrays are *passed by reference* and so any changes made to the array within the function will be observed in the calling scope.

C++ also supports the creation of multidimensional arrays, through the addition of more than one set of brackets. Thus, a two-dimensional array may be created by the following:

```
type arrayName[dimension1][dimension2];
```

The array will have *dimension1* x *dimension2* elements of the same type and can be thought of as an array of arrays. The first index indicates which of *dimension1* subarrays to access, and then the second index accesses one of *dimension2* elements within that subarray. Initialization and access thus work similarly to the one-dimensional case:

```
#include <iostream>
using namespace std;

int main() {
    int twoDimArray[2][4];
    twoDimArray[0][0] = 6;
    twoDimArray[0][1] = 0;
    twoDimArray[0][2] = 9;
```

```
    twoDimArray[0][3] = 6;
    twoDimArray[1][0] = 2;
    twoDimArray[1][1] = 0;
    twoDimArray[1][2] = 1;
    twoDimArray[1][3] = 1;

    for(int i = 0; i < 2; i++)
        for(int j = 0; j < 4; j++)
            cout << twoDimArray[i][j];

    cout << endl;
    return 0;
}
```

The array can also be initialized at declaration in the following ways:

```
int twoDimArray[2][4] = { 6, 0, 9, 6, 2, 0, 1, 1 };

int twoDimArray[2][4] = { { 6, 0, 9, 6 } , { 2, 0, 1, 1 } };
```

Note that dimensions must *always* be provided when initializing multidimensional arrays, as it is otherwise impossible for the compiler to determine what the intended element partitioning is. For the same reason, when multidimensional arrays are specified as arguments to functions, all dimensions but the first *must* be provided (the first dimension is optional), as in the following:

```
int aFunction(int arr[][4]) { … }
```

Multidimensional arrays are merely an abstraction for programmers, as all of the elements in the array are sequential in memory. Declaring `int arr[2][4];` is the same thing as declaring `int arr[8];`.

2 Strings

String literals such as `"Hello, world!"` are actually represented by C++ as a sequence of characters in memory. In other words, a string is simply a character array and can be manipulated as such.

Consider the following program:

```
#include <iostream>
using namespace std;

int main() {
    char helloworld[] = {  'H', 'e', 'l', 'l', 'o', ',', ' ',
                           'w', 'o', 'r', 'l', 'd', '!', '\0' };

    cout << helloworld << endl;

    return 0;
}
```

This program prints `Hello, world!` Note that the character array `helloworld` ends with a special character known as the *null character*. This character is used to indicate the end of the string.

Character arrays can also be initialized using string literals. In this case, no null character is needed, as the compiler will automatically insert one:

```
char helloworld[] = "Hello, world!";
```

The individual characters in a string can be manipulated either directly by the programmer or by using special functions provided by the C/C++ libraries. These can be included in a program through the use of the `#include` directive. Of particular note are the following:

- cctype (ctype.h): character handling
- cstdio (stdio.h): input/output operations
- cstdlib (stdlib.h): general utilities
- cstring (string.h): string manipulation

Here is an example to illustrate the cctype library:

```
1   #include <iostream>
2   #include <cctype>
3   using namespace std;
4
5   int main() {
6      char messyString[] = "t6H0I9s6.iS.999a9.STRING";
7
8      char current = messyString[0];
9      for(int i = 0; current != '\0'; current = messyString[++i]) {
10         if(isalpha(current))
11            cout << (char)(isupper(current) ? tolower(current) : current);
12         else if(ispunct(current))
13            cout << ' ';
14     }
15
16     cout << endl;
17     return 0;
18  }
```

This example uses the `isalpha`, `isupper`, `ispunct`, and `tolower` functions from the cctype library. The `is`- functions check whether a given character is an alphabetic character, an uppercase letter, or a punctuation character, respectively. These functions return a Boolean value of either `true` or `false`. The `tolower` function converts a given character to lowercase.

The for loop beginning at line 9 takes each successive character from `messyString` until it reaches the null character. On each iteration, if the current character is alphabetic and uppercase, it is converted to lowercase and then displayed. If it is already lowercase it is simply displayed. If the character is a punctuation mark, a space is displayed. All other characters are ignored. The resulting output is `this is a string`. For now, ignore the `(char)` on line 11; we will cover that in a later lecture.

Here is an example to illustrate the cstring library:

```
#include <iostream>
#include <cstring>
using namespace std;

int main() {
    char fragment1[] = "I'm a s";
    char fragment2[] = "tring!";
    char fragment3[20];
    char finalString[20] = "";

    strcpy(fragment3, fragment1);
    strcat(finalString, fragment3);
    strcat(finalString, fragment2);

    cout << finalString;
    return 0;
}
```

This example creates and initializes two strings, `fragment1` and `fragment2`. `fragment3` is declared but not initialized. `finalString` is partially initialized (with just the null character).

`fragment1` is copied into `fragment3` using `strcpy`, in effect initializing `fragment3` to `I'm a s`. `strcat` is then used to concatenate `fragment3` onto `finalString` (the function overwrites the existing null character), thereby giving `finalString` the same contents as `fragment3`. Then `strcat` is used again to concatenate `fragment2` onto `finalString`. `finalString` is displayed, giving `I'm a string!`.

You are encouraged to read the documentation on these and any other libraries of interest to learn what they can do and how to use a particular function properly. (One source is http://www.cplusplus.com/reference/.)

Lecture 5 Notes: Pointers

1 Background

1.1 Variables and Memory

When you declare a variable, the computer associates the variable name with a particular location in memory and stores a value there.

When you refer to the variable by name in your code, the computer must take two steps:

1. Look up the address that the variable name corresponds to
2. Go to that location in memory and retrieve or set the value it contains

C++ allows us to perform either one of these steps independently on a variable with the `&` and `*` operators:

1. `&x` evaluates to the address of `x` in memory.
2. `*( &x )` takes the address of `x` and *dereferences* it – it retrieves the value at that location in memory. `*( &x )` thus evaluates to the same thing as `x`.

1.2 Motivating Pointers

Memory addresses, or *pointers*, allow us to manipulate data much more flexibly; manipulating the memory addresses of data can be more efficient than manipulating the data itself. Just a taste of what we'll be able to do with pointers:

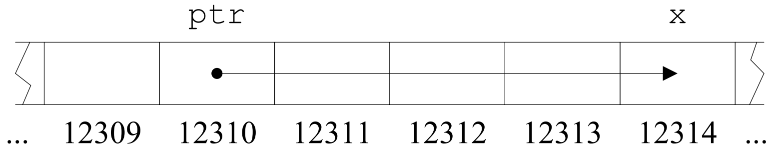
- More flexible pass-by-reference
- Manipulate complex data structures efficiently, even if their data is scattered in different memory locations
- Use polymorphism – calling functions on data without knowing exactly what kind of data it is (more on this in Lectures 7-8)

2 Pointers and their Behavior

2.1 The Nature of Pointers

Pointers are just variables storing integers – but those integers happen to be memory addresses, usually addresses of other variables. A pointer that stores the address of some variable `x` is said to *point to* `x`. We can access the value of `x` by dereferencing the pointer.

As with arrays, it is often helpful to visualize pointers by using a row of adjacent cells to represent memory locations, as below. Each cell represents 1 block of memory. The dot-arrow notation indicates that `ptr` "points to" `x` – that is, the value stored in `ptr` is 12314, `x`'s memory address.



2.2 Pointer Syntax/Usage

2.2.1 Declaring Pointers

To declare a pointer variable named `ptr` that points to an integer variable named `x`:

```
int *ptr = &x;
```

`int *ptr` declares the pointer to an integer value, which we are initializing to the address of `x`.

We can have pointers to values of any type. The general scheme for declaring pointers is:

```
data_type *pointer_name; // Add "= initial_value" if applicable
```

`pointer_name` is then a variable of type `data_type *` – a "pointer to a `data_type` value."

2.2.2 Using Pointer Values

Once a pointer is declared, we can dereference it with the `*` operator to access its value:

```
cout << *ptr; // Prints the value pointed to by ptr,
              // which in the above example would be x's value
```

We can use deferenced pointers as l-values:

```
*ptr = 5; // Sets the value of x
```

Without the * operator, the identifier `x` refers to the pointer itself, not the value it points to:

```
cout << ptr; // Outputs the memory address of x in base 16
```

Just like any other data type, we can pass pointers as arguments to functions. The same way we'd say `void func(int x) {...}`, we can say `void func(int *x){...}`. Here is an example of using pointers to square a number in a similar fashion to pass-by-reference:

```
1 void squareByPtr( int *numPtr ) {
2   *numPtr = *numPtr * *numPtr;
3 }
4
5 int main() {
6   int x = 5;
7   squareByPtr(&x);
8   cout << x; // Prints 25
9 }
```

Note the varied uses of the * operator on line 2.

2.2.3 `const` Pointers

There are two places the `const` keyword can be placed within a pointer variable declaration. This is because there are two different variables whose values you might want to forbid changing: the pointer itself and the value it points to.

```
const int *ptr;
```

declares a changeable pointer to a constant integer. The integer value cannot be changed through this pointer, but the pointer may be changed to point to a different constant integer.

```
int * const ptr;
```

declares a constant pointer to changeable integer data. The integer value can be changed through this pointer, but the pointer may not be changed to point to a different constant integer.

```
const int * const ptr;
```

forbids changing either the address `ptr` contains or the value it points to.

2.3 Null, Uninitialized, and Deallocated Pointers

Some pointers do not point to valid data; dereferencing such a pointer is a runtime error. Any pointer set to 0 is called a *null pointer*, and since there is no memory location 0, it is an invalid pointer. One should generally check whether a pointer is null before dereferencing it. Pointers are often set to 0 to signal that they are not currently valid.

Dereferencing pointers to data that has been erased from memory also usually causes runtime errors. Example:

```
int *myFunc() {
  int phantom = 4;
  return &phantom;
}
```

`phantom` is deallocated when `myFunc` exits, so the pointer the function returns is invalid.

As with any other variable, the value of a pointer is undefined until it is initialized, so it may be invalid.

3 References

When we write `void f(int &x) {...}` and call `f(y)`, the reference variable `x` becomes another name – an *alias* – for the value of `y` in memory. We can declare a reference variable locally, as well:

```
int y;
int &x = y; // Makes x a reference to, or alias of, y
```

After these declarations, changing `x` will change `y` and vice versa, because they are two names for the same thing.

References are just pointers that are dereferenced every time they are used. Just like pointers, you can pass them around, return them, set other references to them, etc. The only differences between using pointers and using references are:

- References are sort of pre-dereferenced – you do not dereference them explicitly.
- You cannot change the location to which a reference points, whereas you can change the location to which a pointer points. Because of this, references must always be initialized when they are declared.
- When writing the value that you want to make a reference to, you do not put an `&` before it to take its address, whereas you do need to do this for pointers.

3.1 The Many Faces of * and &

The usage of the * and & operators with pointers/references can be confusing. The * operator is used in two different ways:

1. When declaring a pointer, * is placed before the variable name to indicate that the variable being declared is a pointer – say, a pointer to an `int` or `char`, not an `int` or `char` value.
2. When using a pointer that has been set to point to some value, * is placed before the pointer name to dereference it – to access or set the value it points to.

A similar distinction exists for &, which can be used either

1. to indicate a reference data type (as in `int &x;`), or
2. to take the address of a variable (as in `int *ptr = &x;`).

4 Pointers and Arrays

The name of an array is actually a pointer to the first element in the array. Writing `myArray[3]` tells the compiler to return the element that is 3 away from the starting element of `myArray`.

This explains why arrays are always passed by reference: passing an array is really passing a pointer.

This also explains why array indices start at 0: the first element of an array is the element that is 0 away from the start of the array.

4.1 Pointer Arithmetic

Pointer arithmetic is a way of using subtraction and addition of pointers to move around between locations in memory, typically between array elements. Adding an integer `n` to a pointer produces a new pointer pointing to `n` positions further down in memory.

4.1.1 Pointer Step Size

Take the following code snippet:

```
long arr[] = {6, 0, 9, 6};
long *ptr = arr;
ptr++;
```

```
4 long *ptr2 = arr + 3;
```

When we add 1 to `ptr` in line 3, we don't just want to move to the next byte in memory, since each array element takes up multiple bytes; we want to move to the next element in the array. The C++ compiler automatically takes care of this, using the appropriate step size for adding to and subtracting from pointers. Thus, line 3 moves `ptr` to point to the second element of the array.

Similarly, we can add/subtract two pointers: `ptr2 - ptr` gives the number of array elements between `ptr2` and `ptr` (2). All addition and subtraction operations on pointers use the appropriate step size.

4.1.2 Array Access Notations

Because of the interchangeability of pointers and array names, *array-subscript notation* (the form `myArray[3]`) can be used with pointers as well as arrays. When used with pointers, it is referred to as *pointer-subscript notation*.

An alternative is *pointer-offset notation*, in which you explicitly add your offset to the pointer and dereference the resulting address. For instance, an alternate and functionally identical way to express `myArray[3]` is `*(myArray + 3)`.

4.2 `char *` Strings

You should now be able to see why the type of a string value is `char *`: a string is actually an array of characters. When you set a `char *` to a string, you are really setting a pointer to point to the first character in the array that holds the string.

You cannot modify string literals; to do so is either a syntax error or a runtime error, depending on how you try to do it. (String literals are loaded into read-only program memory at program startup.) You can, however, modify the contents of an array of characters. Consider the following example:

```
char courseName1[] = {'6', '.', '0', '9', '6', '\0' };
char *courseName2 = "6.096";
```

Attempting to modify one of the elements `courseName1` is permitted, but attempting to modify one of the characters in `courseName2` will generate a runtime error, causing the program to crash.

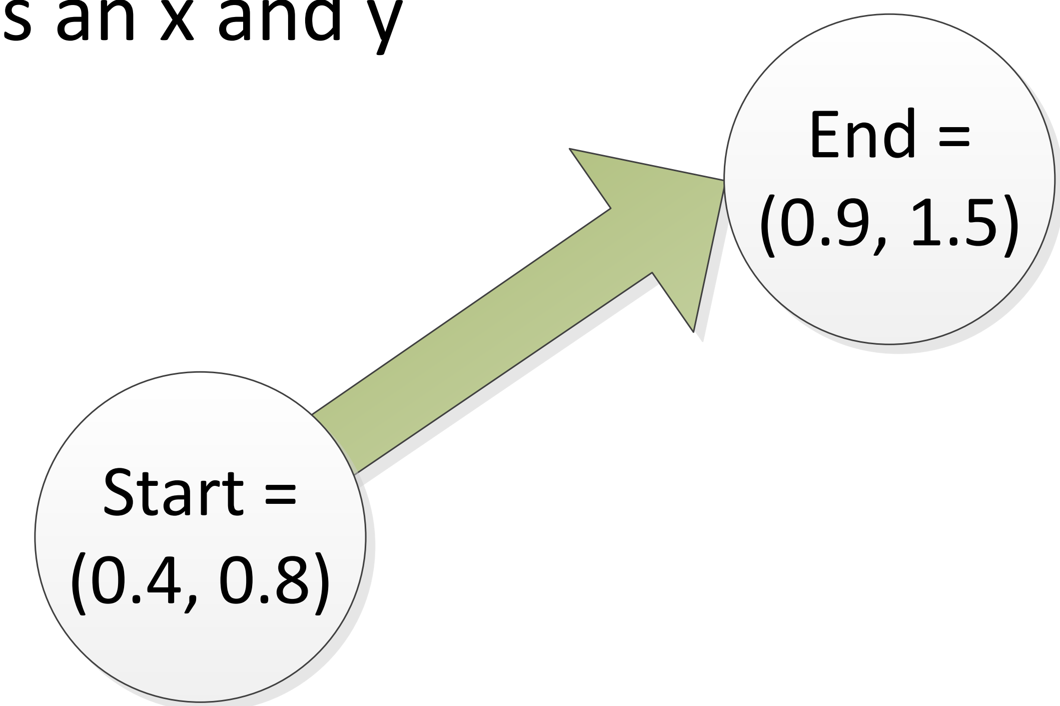
6.096 Lecture 6:
User-defined Datatypes

classes and structs

Geza Kovacs

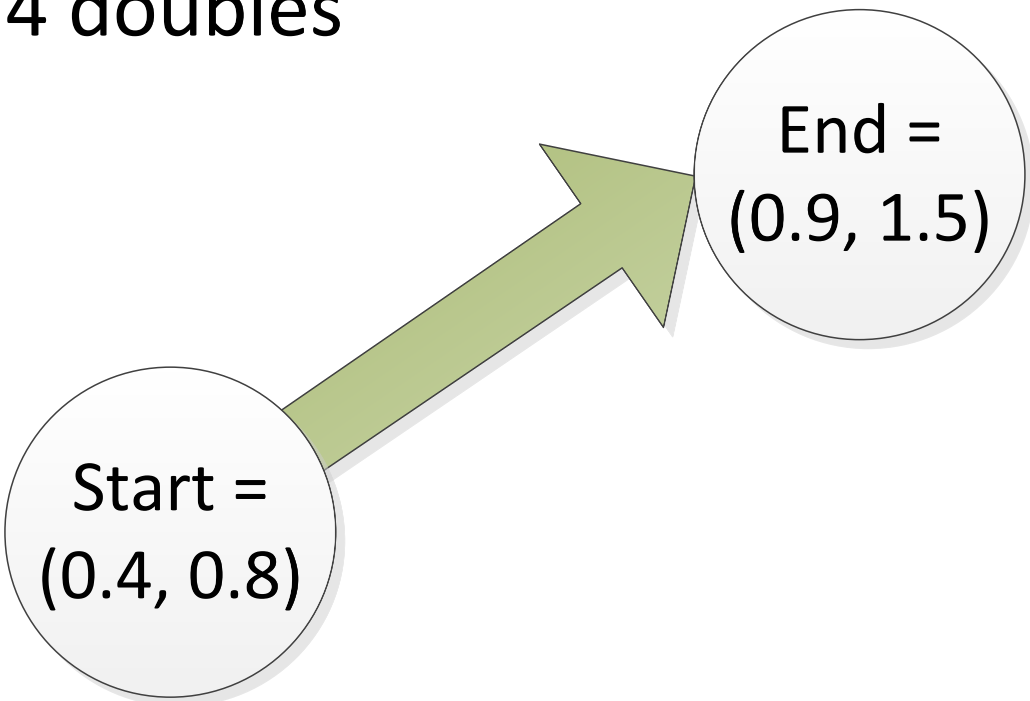
Representing a (Geometric) Vector

- In the context of geometry, a vector consists of 2 points: a start and a finish
- Each point itself has an x and y coordinate

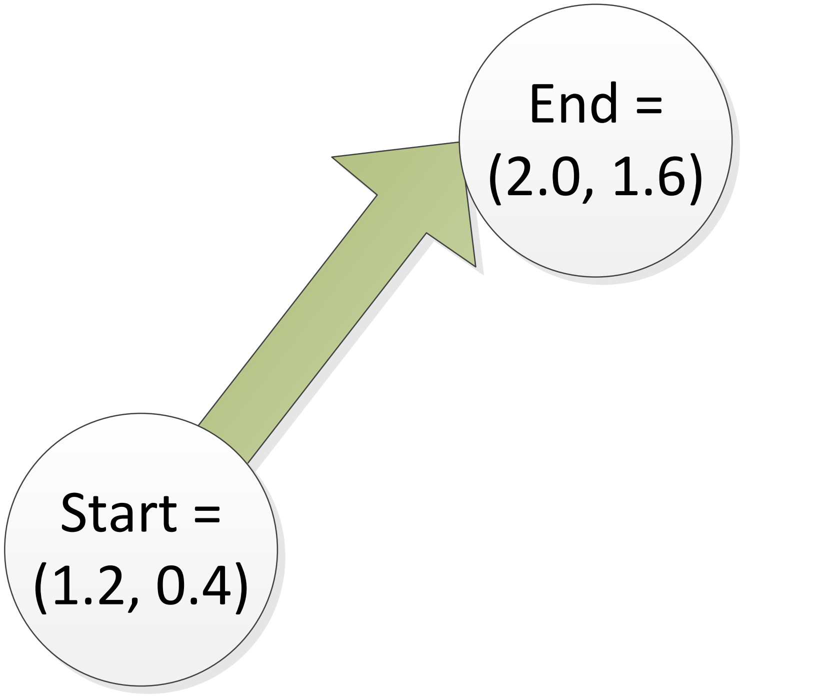


Representing a (Geometric) Vector

- Our representation so far? Use 4 doubles (startx, starty, endx, endy)
- We need to pass all 4 doubles to functions



```
int main() {
  double xStart = 1.2;
  double xEnd = 2.0;
  double yStart = 0.4;
  double yEnd = 1.6;
}
```



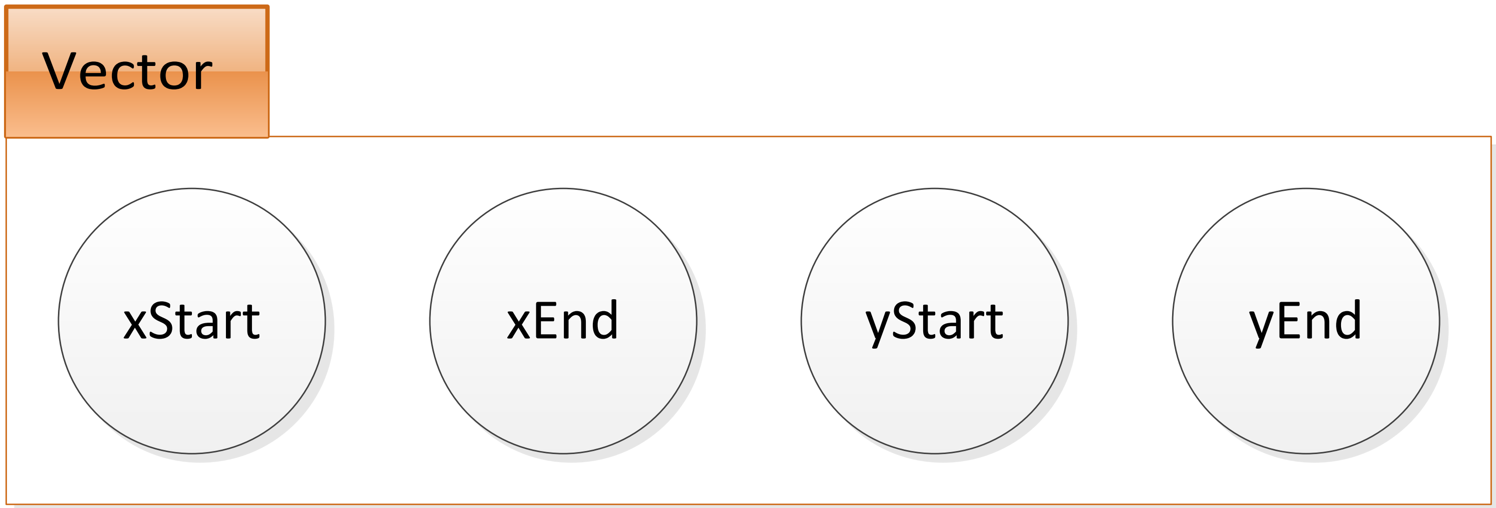
```cpp
void printVector(double x0, double x1, double y0, double y1) {
  cout << "(" << x0 << "," << y0 << ") -> ("
       << x1 << "," << y1 << ")" << endl;
}

int main() {
  double xStart = 1.2;
  double xEnd = 2.0;
  double yStart = 0.4;
  double yEnd = 1.6;
  printVector(xStart, xEnd, yStart, yEnd);
  // (1.2,2.0) -> (0.4,1.6)
}
```

```cpp
void offsetVector(double &x0, double &x1, double &y0, double &y1,
                  double offsetX, double offsetY) {
  x0 += offsetX;
  x1 += offsetX;
  y0 += offsetY;
  y1 += offsetY;
}

void printVector(double x0, double x1, double y0, double y1) {
  cout << "(" << x0 << "," << y0 << ") -> ("
       << x1 << "," << y1 << ")" << endl;
}

int main() {
  double xStart = 1.2;
  double xEnd = 2.0;
  double yStart = 0.4;
  double yEnd = 1.6;
  offsetVector(xStart, xEnd, yStart, yEnd, 1.0, 1.5);
  printVector(xStart, xEnd, yStart, yEnd);
  // (2.2,1.9) -> (3.8,4.3)
}
```

Many variables being passed to functions

class

- A user-defined datatype which groups together related pieces of information



class definition syntax

name

```
class Vector {
public:
  double xStart;
  double xEnd;
  double yStart;
  double yEnd;
};
```

- This indicates that the new datatype we're defining is called Vector

class definition syntax

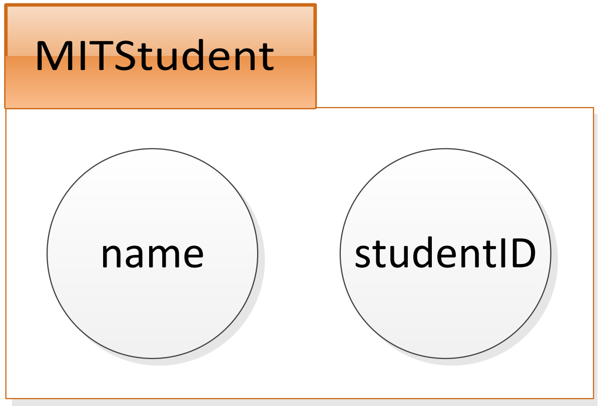
```
class Vector {
public:
  double xStart;
  double xEnd;
  double yStart;
  double yEnd;
};
```

fields

- **Fields** indicate what related pieces of information our datatype consists of
  - Another word for field is **members**

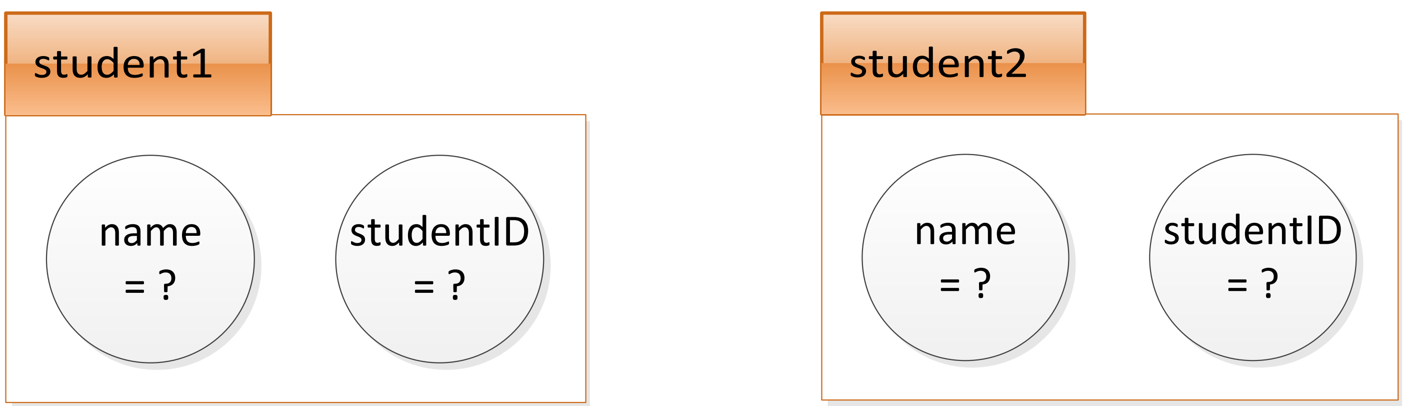
Fields can have different types

```
class MITStudent {
public:
  char *name;
  int studentID;
};
```



Instances

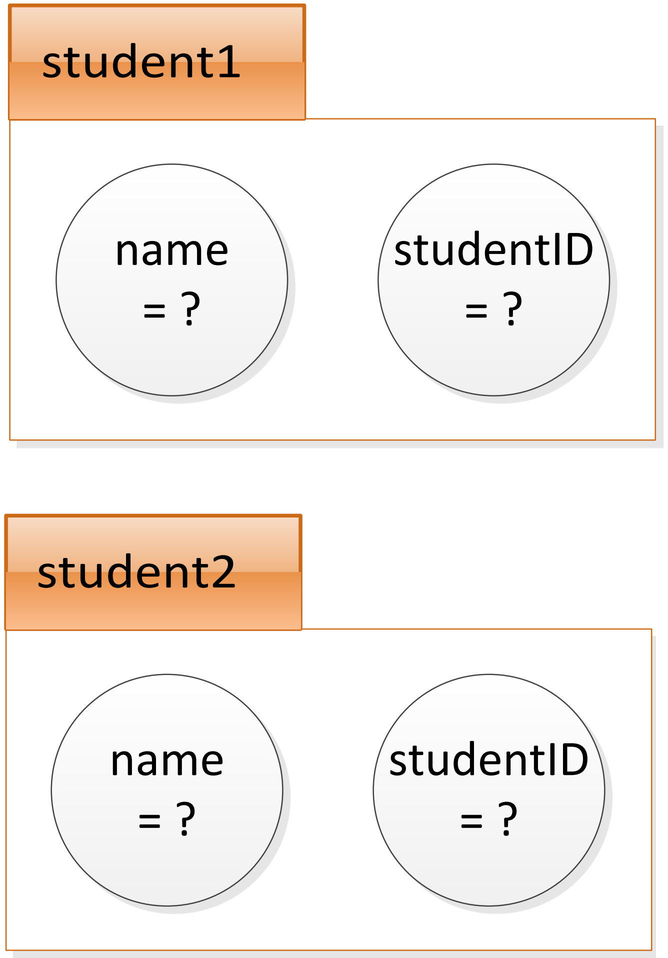
- An instance is an occurrence of a class. Different instances can have their own set of values in their fields.
- If you wanted to represent 2 different students (who can have different names and IDs), you would use 2 instances of MITStudent



Declaring an Instance

- Defines 2 instances of MITStudent: one called student1, the other called student2

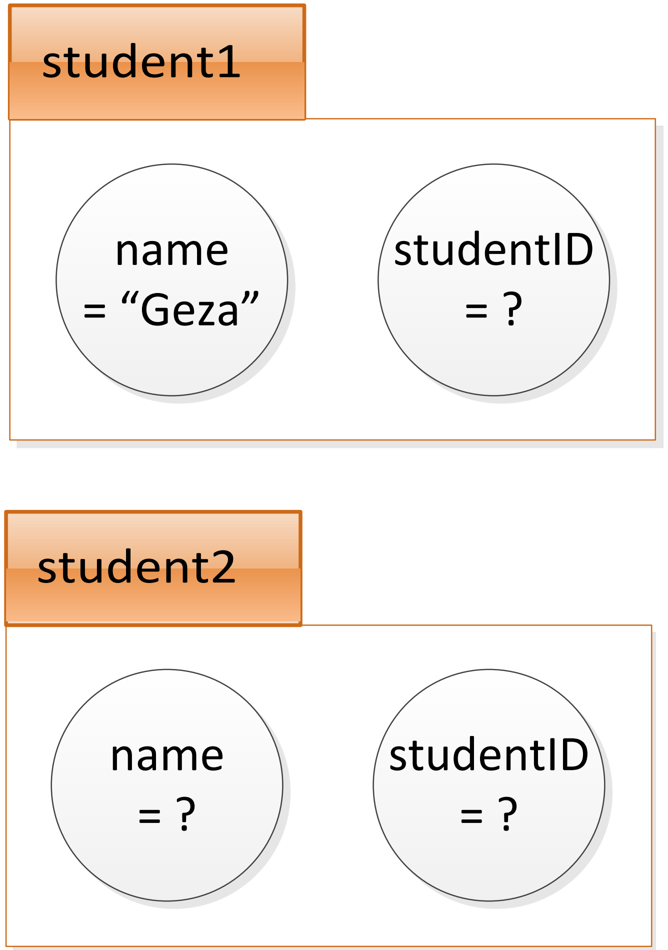
```
class MITStudent {
public:
  char *name;
  int studentID;
};

int main() {
  MITStudent student1;
  MITStudent student2;
}
```



Accessing Fields

- To access fields of instances, use variable.fieldName

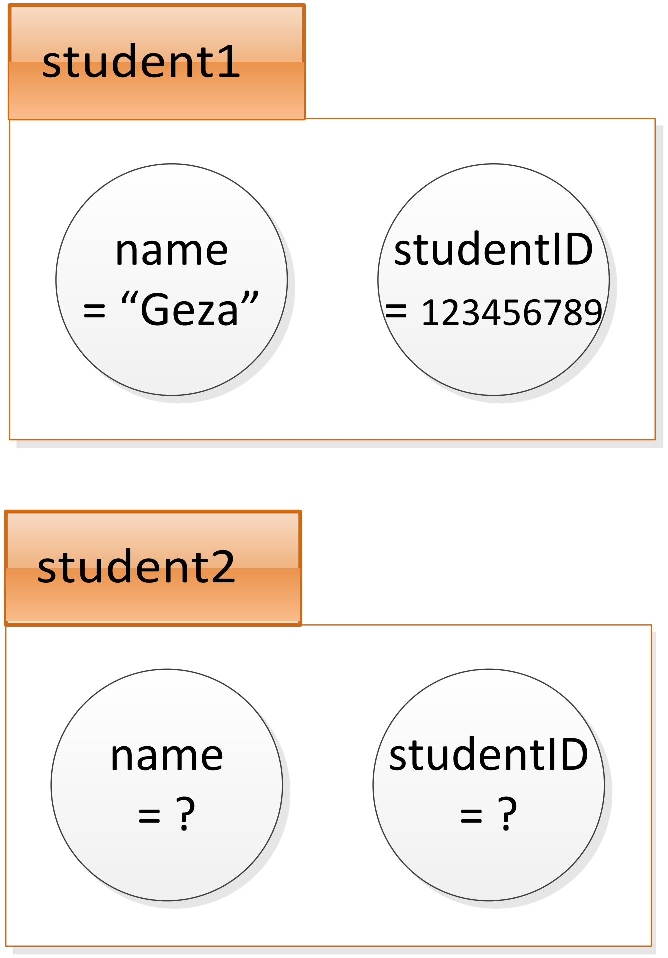
```
class MITStudent {
public:
  char *name;
  int studentID;
};

int main() {
  MITStudent student1;
  MITStudent student2;
  student1.name = "Geza";
}
```



Accessing Fields

- To access fields of instances, use variable.fieldName

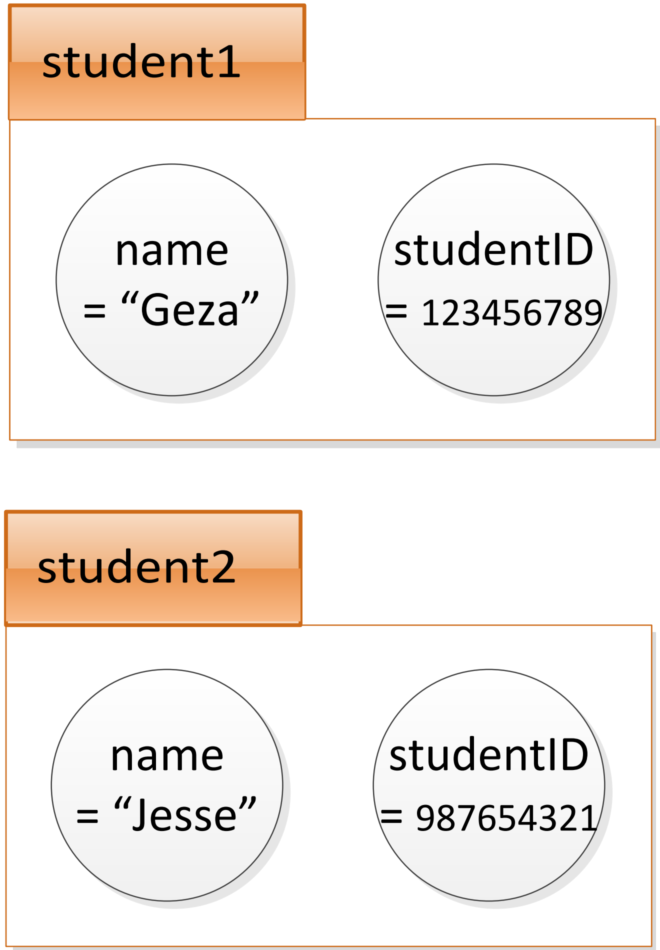
```
class MITStudent {
public:
  char *name;
  int studentID;
};

int main() {
  MITStudent student1;
  MITStudent student2;
  student1.name = "Geza";
  student1.studentID = 123456789;
}
```



Accessing Fields

- To access fields of instances, use variable.fieldName

```
class MITStudent {
public:
  char *name;
  int studentID;
};

int main() {
  MITStudent student1;
  MITStudent student2;
  student1.name = "Geza";
  student1.studentID = 123456789;
  student2.name = "Jesse";
  student2.studentID = 987654321;
}
```

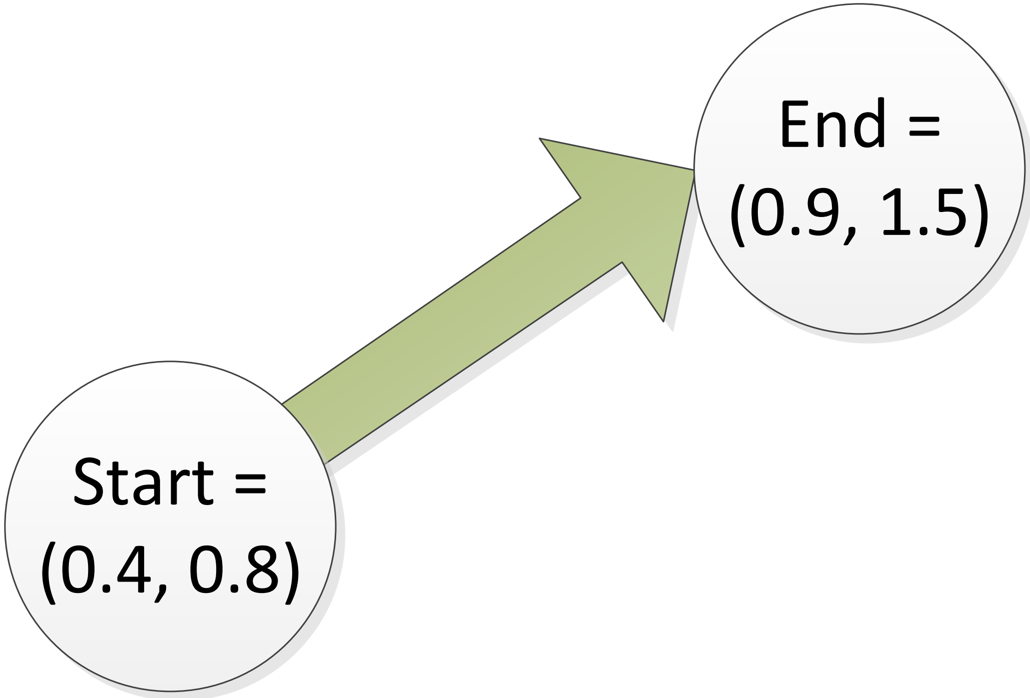


Accessing Fields

- To access fields of instances, use variable.fieldName

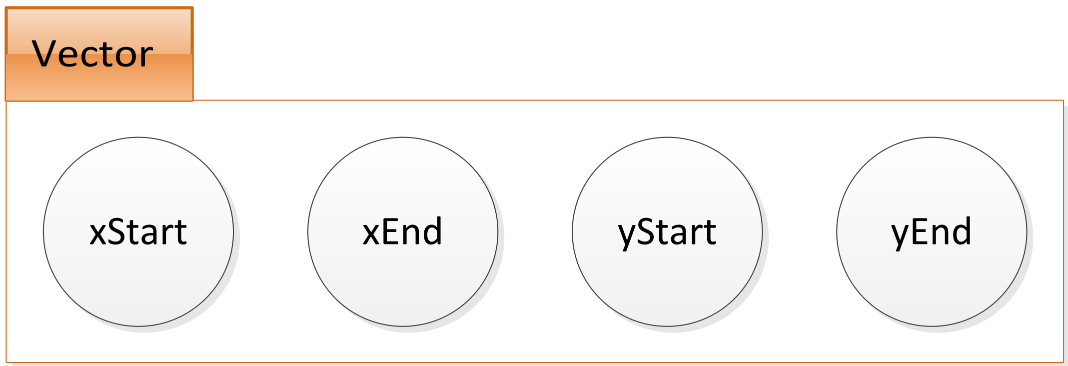
```
class MITStudent {
public:
  char *name;
  int studentID;
};

int main() {
  MITStudent student1;
  MITStudent student2;
  student1.name = "Geza";
  student1.studentID = 123456789;
  student2.name = "Jesse";
  student2.studentID = 987654321;
  cout << "student1 name is" << student1.name << endl;
  cout << "student1 id is" << student1.studentID << endl;
  cout << "student2 name is" << student2.name << endl;
  cout << "student2 id is" << student2.studentID << endl;
```

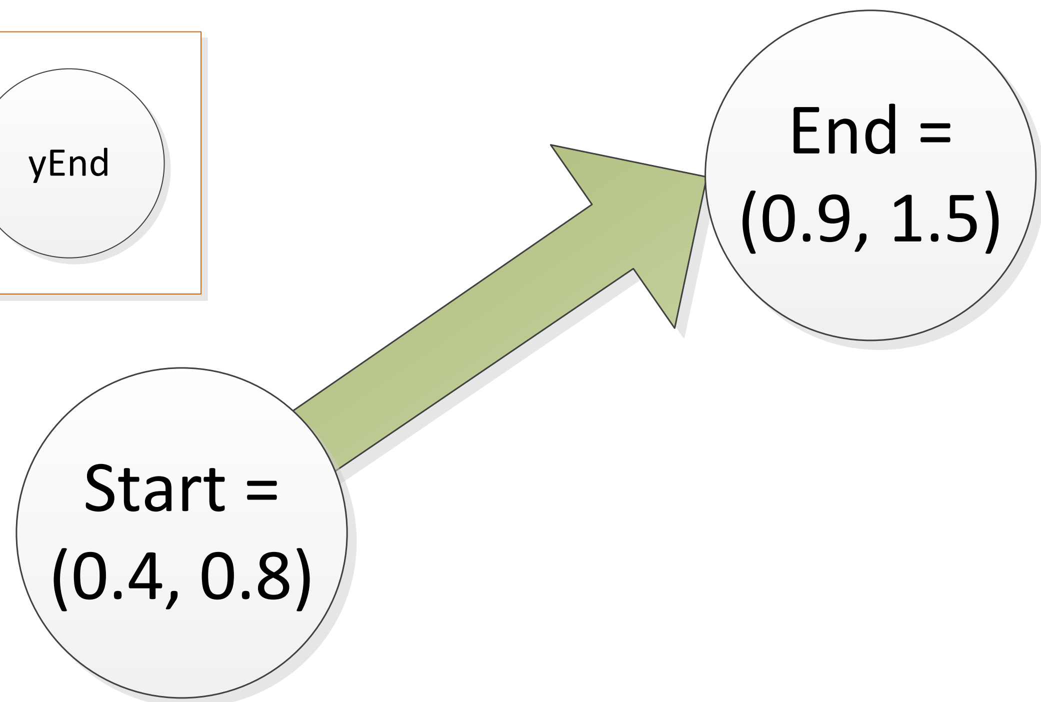
- A point consists of an x and y coordinate
- A vector consists of 2 points: a start and a finish



```
class Vector {
public:
  double xStart;
  double xEnd;
  double yStart;
  double yEnd;
};
```

- A point consists of an x and y coordinate
- A vector consists of 2 points: a start and a finish

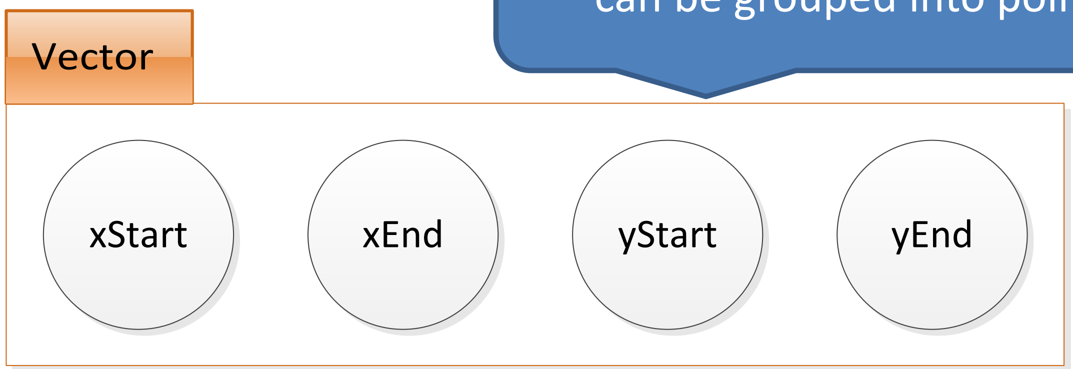


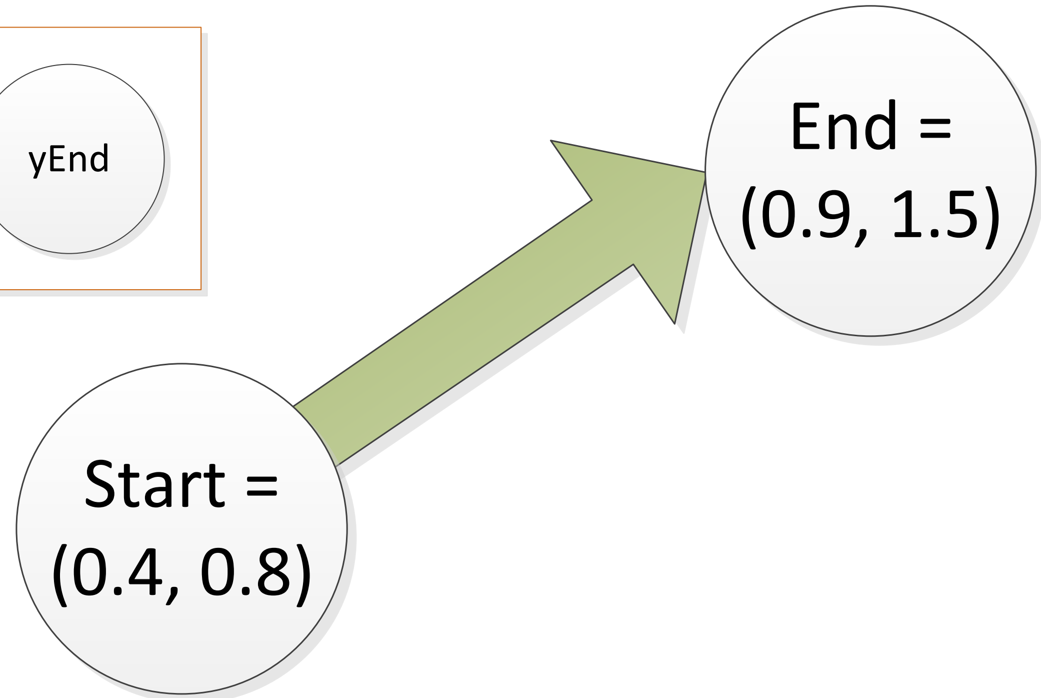


- A point consists of an x and y coordinate
- A vector consists of 2 points: a start and a finish

```
class Vector {
public:
  double xStart;
  double xEnd;
  double yStart;
  double yEnd;
};
```

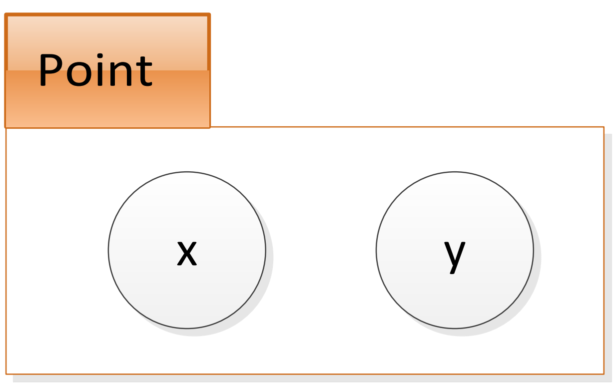
Doesn't show that coordinates can be grouped into points

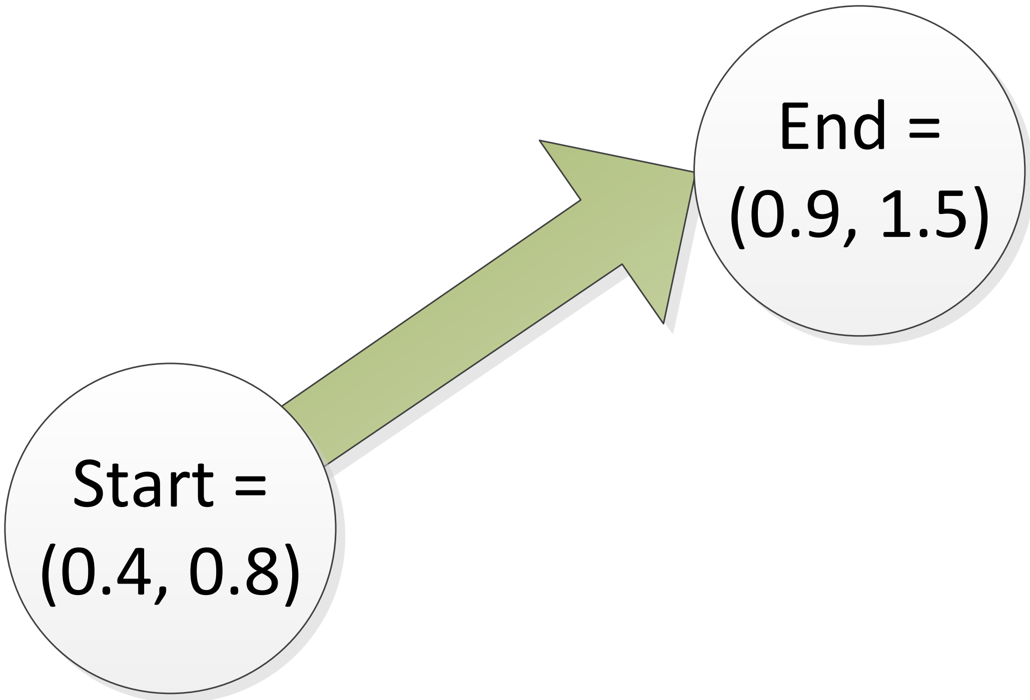




```
class Point {
public:
  double x;
  double y;
};
```

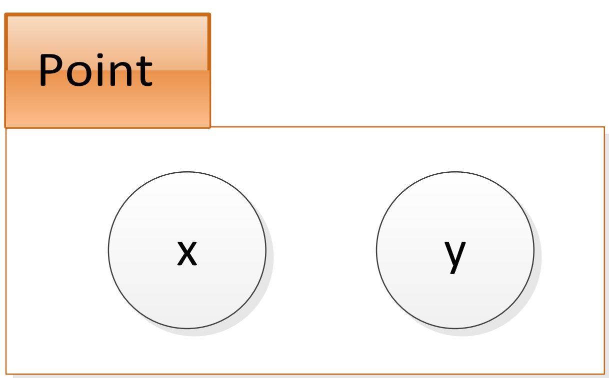
- **A point consists of an x and y coordinate**
- A vector consists of 2 points: a start and a finish

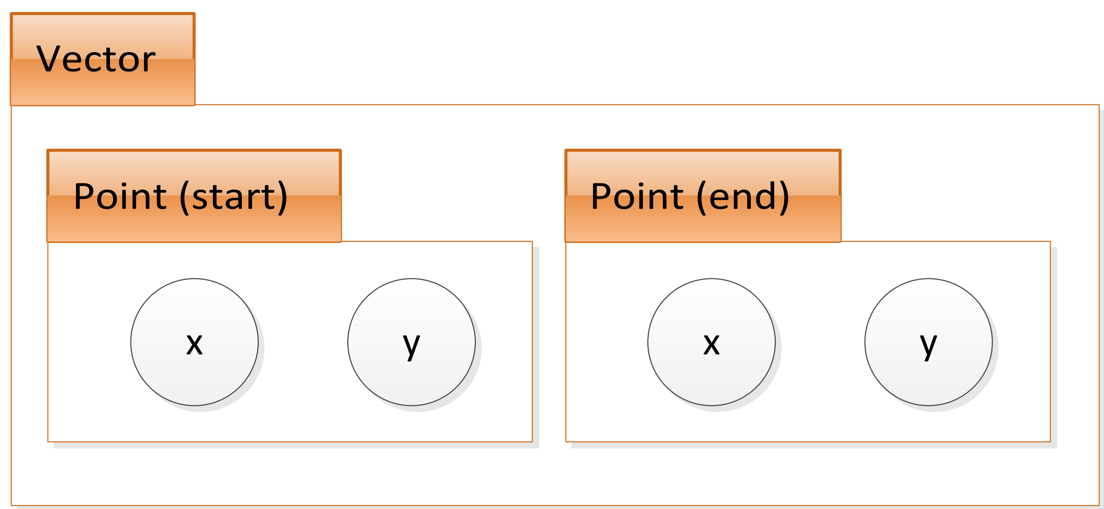




```
class Point {
public:
  double x;
  double y;
};
```

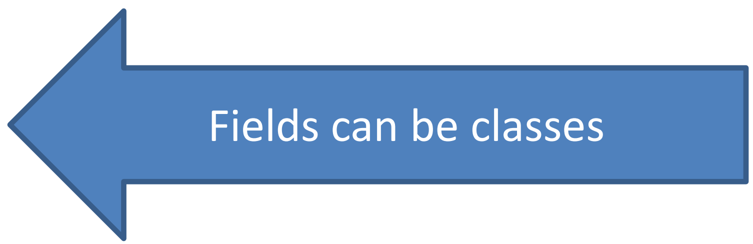
- A point consists of an x and y coordinate
- **A vector consists of 2 points: a start and a finish**

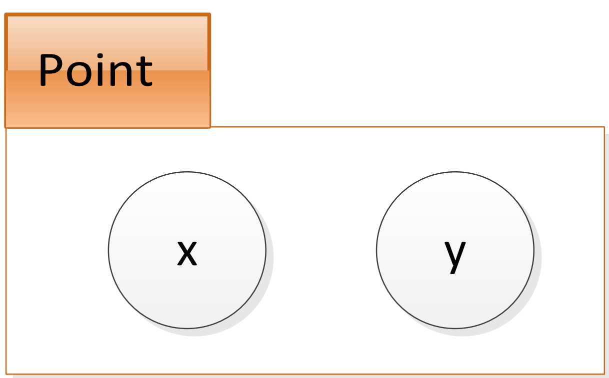


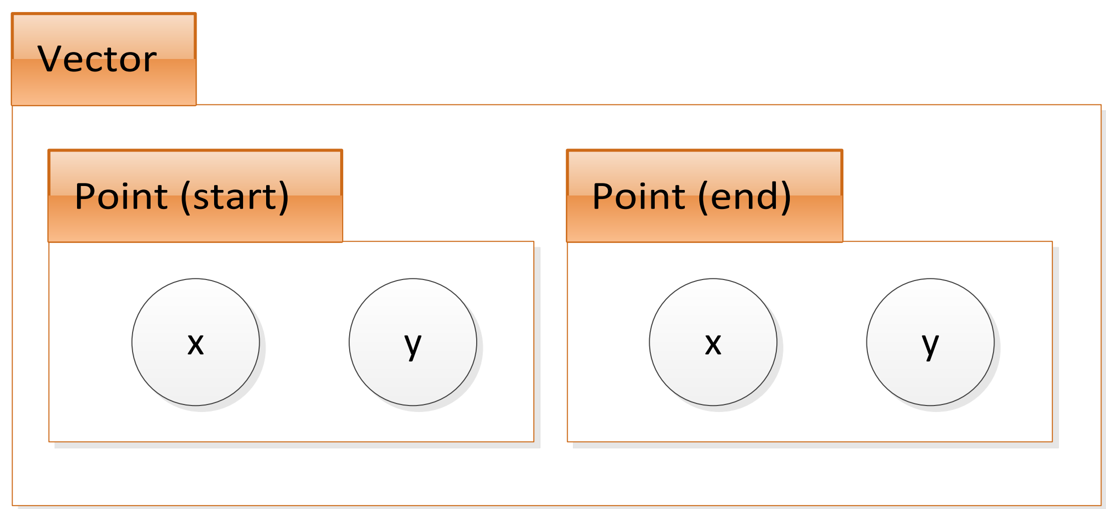


- A point consists of an x and y coordinate
- **A vector consists of 2 points: a start and a finish**

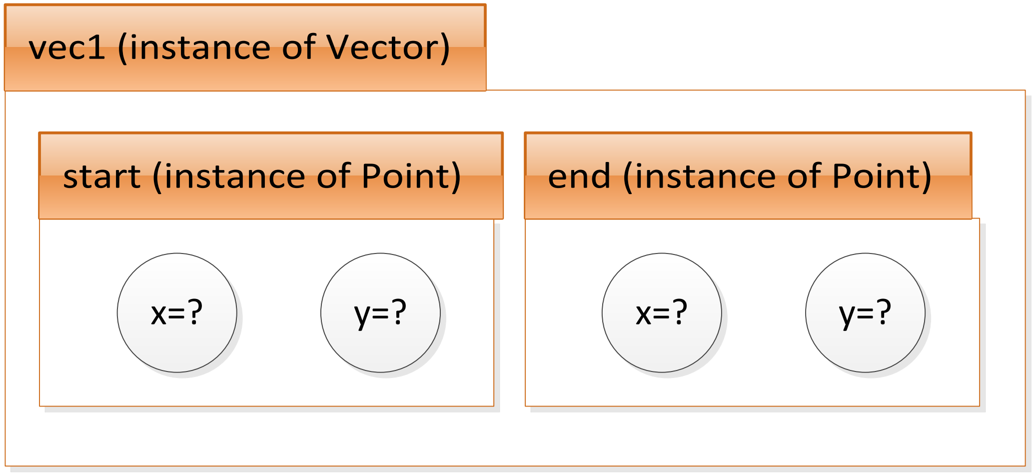
```
class Point {
public:
  double x;
  double y;
};

class Vector {
public:
  Point start;
  Point end;
};
```



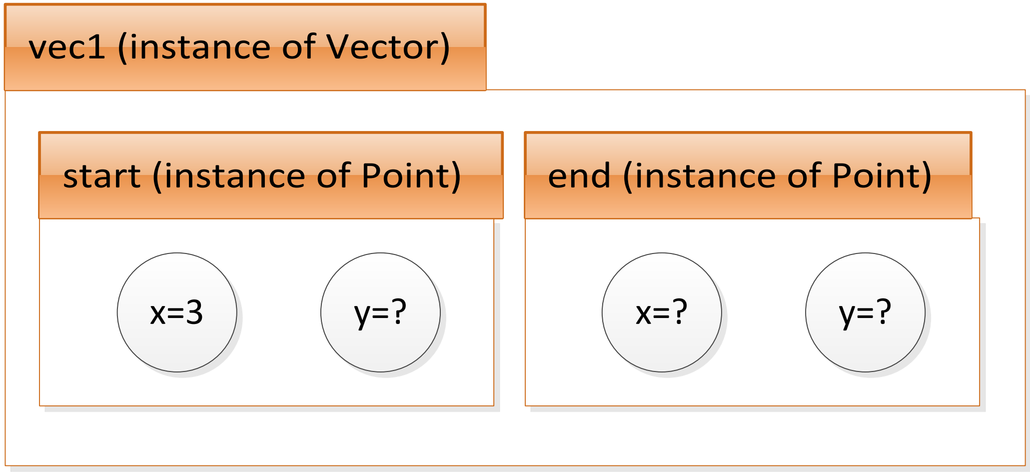




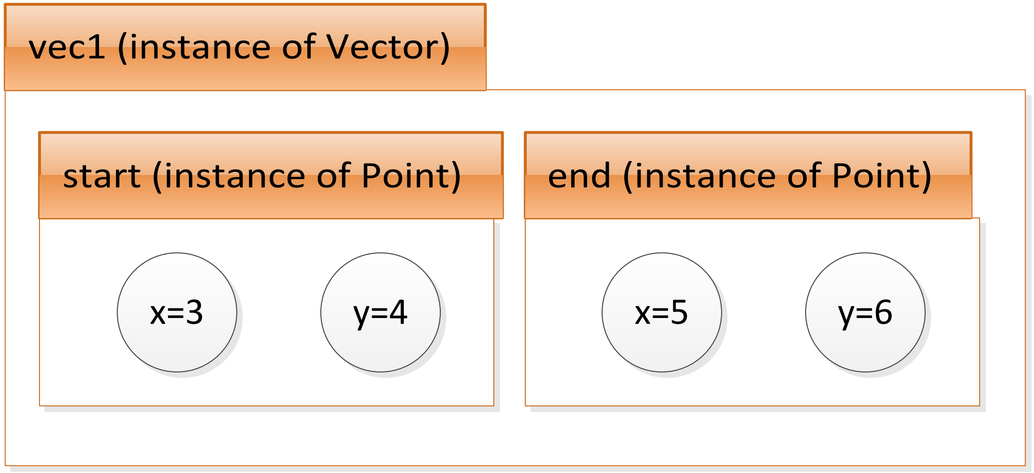
```cpp
class Point {
public:
  double x, y;
};

class Vector {
public:
  Point start, end;
};

int main() {
  Vector vec1;
}
```



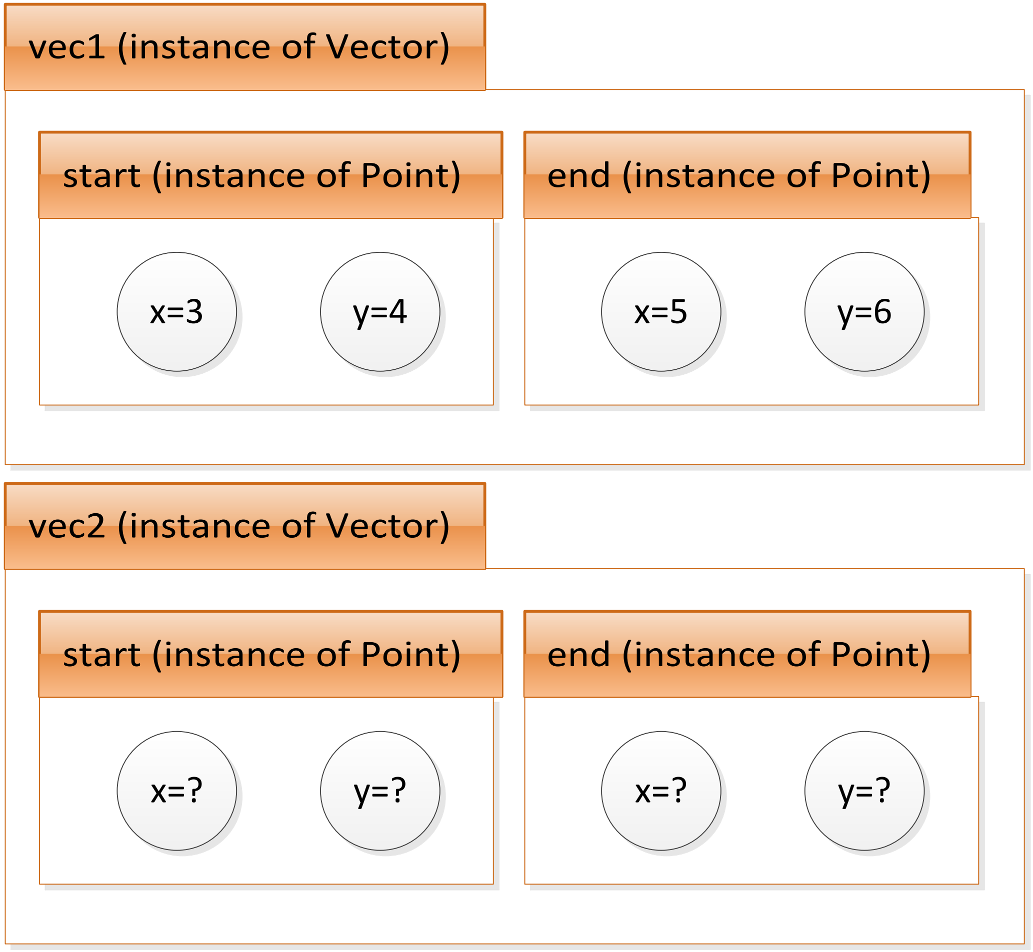
```
class Point {
public:
  double x, y;
};

class Vector {
public:
  Point start, end;
};

int main() {
  Vector vec1;
  vec1.start.x = 3.0;
}
```



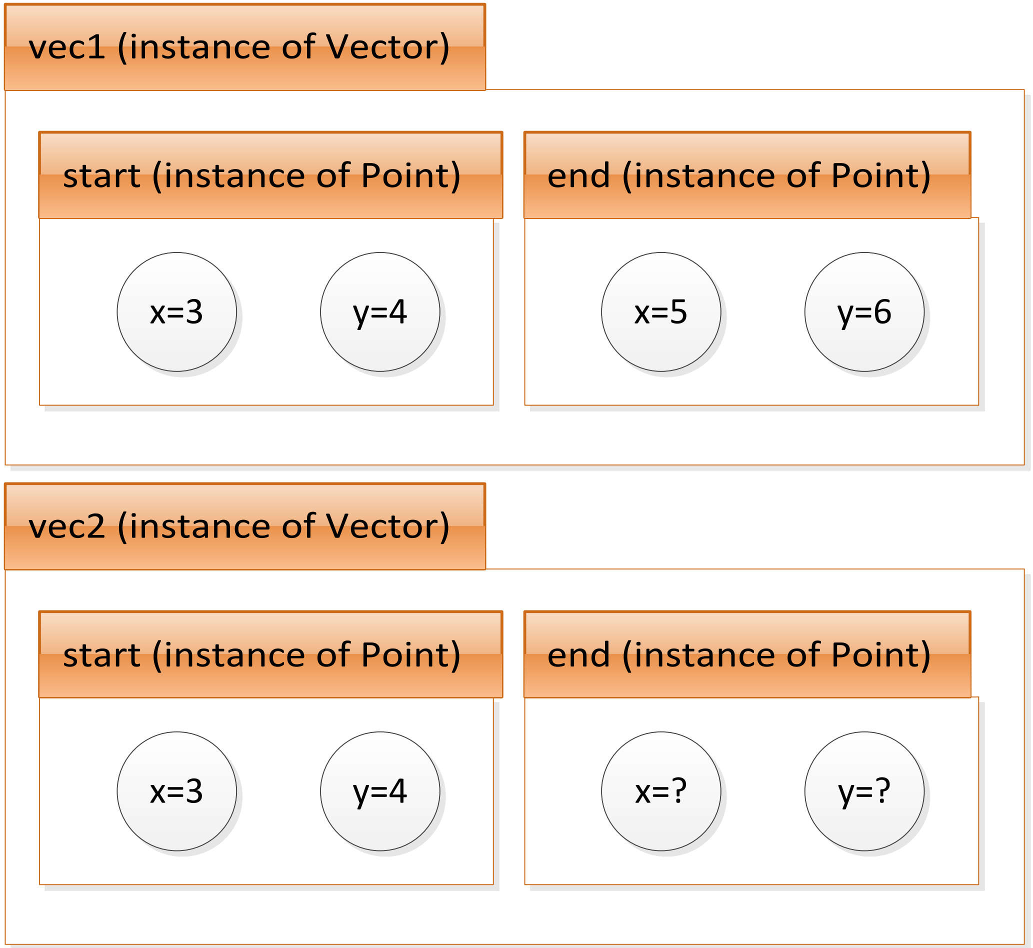
```
class Point {
public:
  double x, y;
};

class Vector {
public:
  Point start, end;
};

int main() {
  Vector vec1;
  vec1.start.x = 3.0;
  vec1.start.y = 4.0;
  vec1.end.x = 5.0;
  vec1.end.y = 6.0;
}
```



```cpp
class Point {
public:
  double x, y;
};

class Vector {
public:
  Point start, end;
};

int main() {
  Vector vec1;
  vec1.start.x = 3.0;
  vec1.start.y = 4.0;
  vec1.end.x = 5.0;
  vec1.end.y = 6.0;
  Vector vec2;
}
```

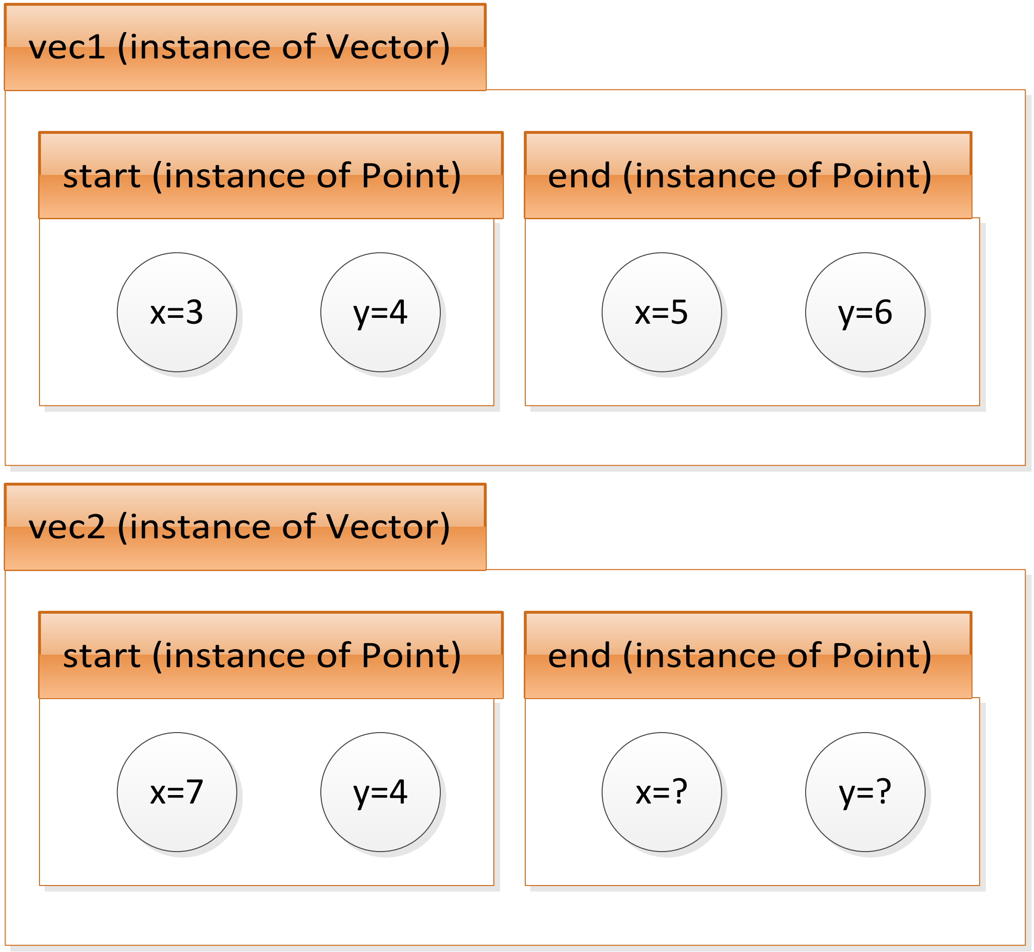


```
class Point {
public:
  double x, y;
};

class Vector {
public:
  Point start, end;
};

int main() {
  Vector vec1;
  vec1.start.x = 3.0;
  vec1.start.y = 4.0;
  vec1.end.x = 5.0;
  vec1.end.y = 6.0;
  Vector vec2;
  vec2.start = vec1.start;
}
```



- Assigning one instance to another copies all fields

```cpp
class Point {
public:
  double x, y;
};

class Vector {
public:
  Point start, end;
};

int main() {
  Vector vec1;
  vec1.start.x = 3.0;
  vec1.start.y = 4.0;
  vec1.end.x = 5.0;
  vec1.end.y = 6.0;
  Vector vec2;
  vec2.start = vec1.start;
  vec2.start.x = 7.0;
}
```



- Assigning one instance to another copies all fields

Passing classes to functions

- Passing by value passes a copy of the class instance to the function; changes aren't preserved

```
class Point { public: double x, y; };

void offsetPoint(Point p, double x, double y) { // does nothing
  p.x += x;
  p.y += y;
}

int main() {
  Point p;
  p.x = 3.0;
  p.y = 4.0;
  offsetPoint(p, 1.0, 2.0); // does nothing
  cout << "(" << p.x << "," << p.y << ")"; // (3.0,4.0)
}
```

Passing classes to functions

- When a class instance is passed by reference, changes are reflected in the original

```
class Point { public: double x, y; };

void offsetPoint(Point &p, double x, double y) { // works
  p.x += x;
  p.y += y;
}

int main() {
  Point p;
  p.x = 3.0;
  p.y = 4.0;
  offsetPoint(p, 1.0, 2.0); // works
  cout << "(" << p.x << "," << p.y << ")"; // (4.0,6.0)
}
```

Passed by reference

```
class Point {
  public: double x, y;
};
```

Point class, with fields x and y

```cpp
class Point {
  public: double x, y;
};
class Vector {
  public: Point start, end;
};
```

Fields can be classes

```
class Point {
  public: double x, y;
};
class Vector {
  public: Point start, end;
};
```

```
int main() {
  Vector vec;
}
```

vec is an instance of Vector

```
class Point {
  public: double x, y;
};
class Vector {
  public: Point start, end;
};
```

```
int main() {
  Vector vec;
  vec.start.x = 1.2;
}
```

Accessing fields

```
class Point {
  public: double x, y;
};
class Vector {
  public: Point start, end;
};
```

```
int main() {
  Vector vec;
  vec.start.x = 1.2; vec.end.x = 2.0; vec.start.y = 0.4; vec.end.y = 1.6;
}
```

```
class Point {
  public: double x, y;
};
class Vector {
  public: Point start, end;
};




void printVector(Vector v) {
  cout << "(" << v.start.x << "," << v.start.y << ") -> (" << v.end.x <<
"," << v.end.y << ")" << endl;
}

int main() {
  Vector vec;
  vec.start.x = 1.2; vec.end.x = 2.0; vec.start.y = 0.4; vec.end.y = 1.6;
  printVector(vec); // (1.2,0.4) -> (2.0,1.6)
}
```

classes can be passed to functions

```
class Point {
  public: double x, y;
};
class Vector {
  public: Point start, end;
};
```

Can pass to value if you don’t need to modify the class

```
void printVector(Vector v) {
  cout << "(" << v.start.x << "," << v.start.y << ") -> (" << v.end.x <<
"," << v.end.y << ")" << endl;
}

int main() {
  Vector vec;
  vec.start.x = 1.2; vec.end.x = 2.0; vec.start.y = 0.4; vec.end.y = 1.6;
  printVector(vec); // (1.2,0.4) -> (2.0,1.6)
}
```

```cpp
class Point {
  public: double x, y;
};
class Vector {
  public: Point start, end;
};
```

Pass classes by reference if they need to be modified

```cpp
void offsetVector(Vector &v, double offsetX, double offsetY) {
  v.start.x += offsetX;
  v.end.x += offsetX;
  v.start.y += offsetY;
  v.end.y += offsetY;
}
void printVector(Vector v) {
  cout << "(" << v.start.x << "," << v.start.y << ") -> (" << v.end.x <<
"," << v.end.y << ")" << endl;
}

int main() {
  Vector vec;
  vec.start.x = 1.2; vec.end.x = 2.0; vec.start.y = 0.4; vec.end.y = 1.6;
  offsetVector(vec,  1.0, 1.5);
  printVector(vec); // (2.2,1.9) -> (3.8,4.3)
}
```

- Observe how some functions are closely associated with a particular class

```
void offsetVector(Vector &v, double offsetX, double offsetY);
void printVector(Vector v);
```

```
int main() {
  Vector vec;
  vec.start.x = 1.2; vec.end.x = 2.0;
  vec.start.y = 0.4; vec.end.y = 1.6;
  offsetVector(vec,  1.0, 1.5);
  printVector(vec);
}
```

- Observe how some functions are closely associated with a particular class
- **Methods**: functions which are part of a class

```
Vector vec;
vec.start.x = 1.2; vec.end.x = 2.0;
vec.start.y = 0.4; vec.end.y = 1.6;
vec.print();
```

Method name

- Observe how some functions are closely associated with a particular class
- **Methods**: functions which are part of a class
  - Implicitly pass the current instance

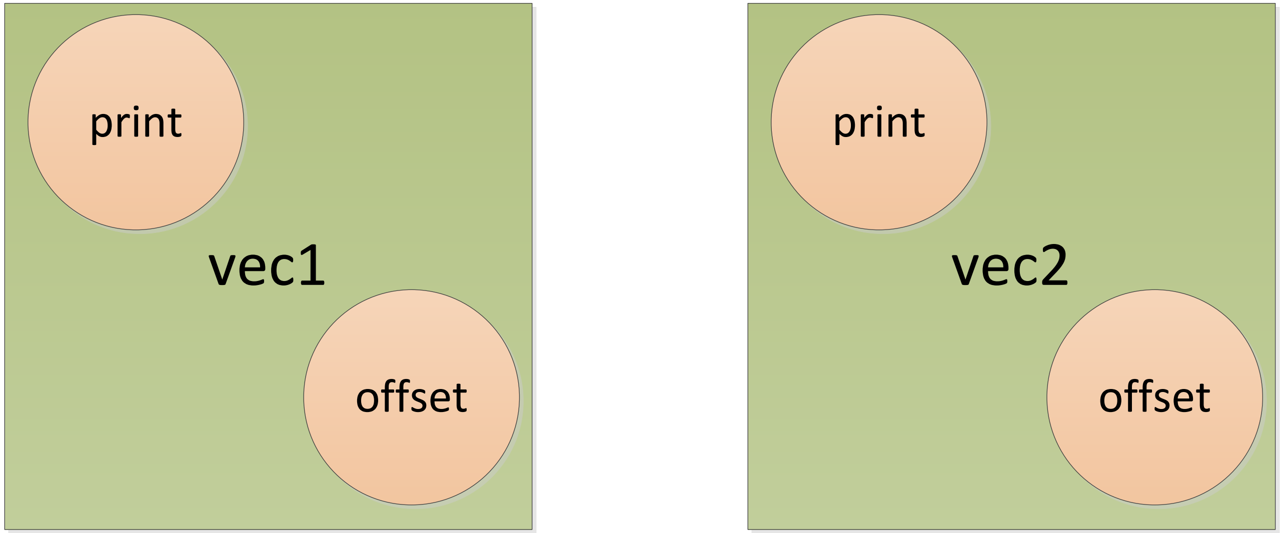
```
Vector vec;
vec.start.x = 1.2; vec.end.x = 2.0;
vec.start.y = 0.4; vec.end.y = 1.6;
vec.print();
```

Object instance

- Observe how some functions are closely associated with a particular class
- **Methods**: functions which are part of a class
  - Implicitly pass the current instance

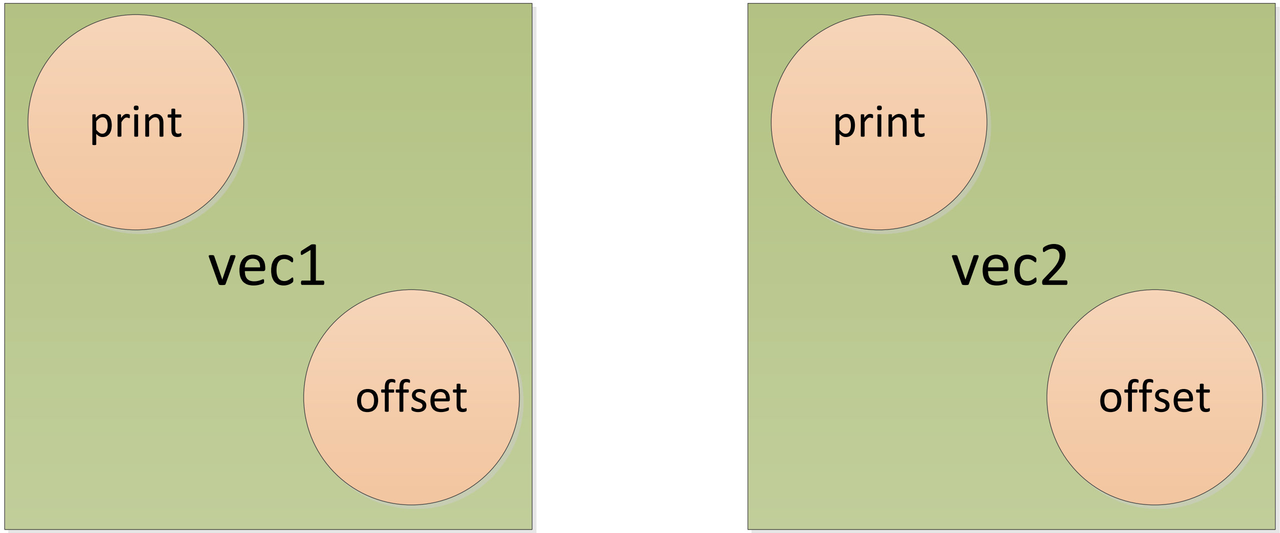
```
Vector vec;
vec.start.x = 1.2; vec.end.x = 2.0;
vec.start.y = 0.4; vec.end.y = 1.6;
vec.print();
vec.offset(1.0, 1.5);
```

Arguments can be passed to methods



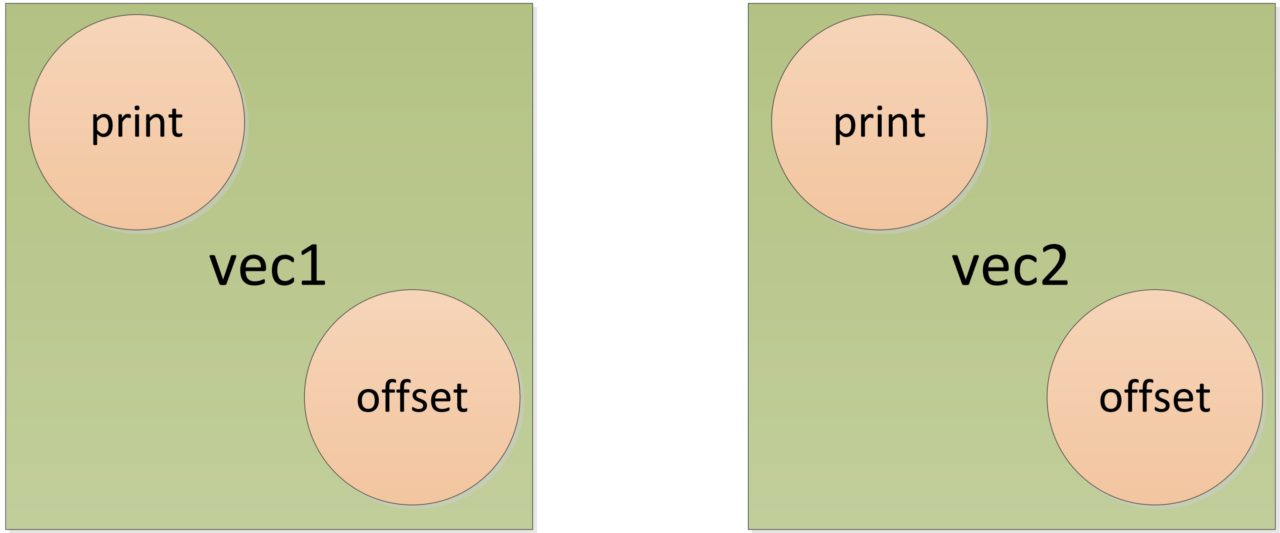

```
Vector vec1;
Vector vec2;
// initialize vec1 and vec2
vec1.print();
```

- Analogy: Methods are “buttons” on each box (instance), which do things when pressed




```
Vector vec1;
Vector vec2;
// initialize vec1 and vec2
vec1.print();
```

Which box’s button was pressed?




```
Vector vec1;
Vector vec2;
// initialize vec1 and vec2
vec1.print();
```

Which button was pressed?

```cpp
class Vector {
public:
  Point start;
  Point end;

  void offset(double offsetX, double offsetY) {
    start.x += offsetX;
    end.x += offsetX;
    start.y += offsetY;
    end.y += offsetY;
  }
  void print() {
    cout << "(" << start.x << "," << start.y << ") -> (" << end.x << "," << end.y << ")" << endl;
  }
};
```

methods

```cpp
class Vector {
public:
  Point start;
  Point end;

  void offset(double offsetX, double offsetY) {
    start.x += offsetX;
    end.x += offsetX;
    start.y += offsetY;
    end.y += offsetY;
  }
  void print() {
    cout << "(" << start.x << "," << start.y << ") -> (" << end.x << "," << end.y << ")" << endl;
  }
};
```

Fields can be accessed in a method

```cpp
class Vector {
public:
  Point start, end;

  void offset(double offsetX, double offsetY) {
    start.offset(offsetX, offsetY);
    end.offset(offsetX, offsetY);
  }
  void print() {
    start.print();
    cout << " -> ";
    end.print();
    cout << endl;
  }
};
```

methods of fields can be called

```cpp
class Point {
public:
  double x, y;
  void offset(double offsetX, double offsetY) {
    x += offsetX; y += offsetY;
  }
  void print() {
    cout << "(" << x << "," << y << ")";
  }
};
```

Implementing Methods Separately

- Recall that function prototypes allowed us to declare that functions will be implemented later
- This can be done analogously for class methods

```
// vector.h - header file
class Point {
public:
  double x, y;
  void offset(double offsetX, double offsetY);
  void print();
};

class Vector {
public:
  Point start, end;
  void offset(double offsetX, double offsetY);
  void print();
};
```

```cpp
#include "vector.h"
// vector.cpp - method implementation
void Point::offset(double offsetX, double offsetY) {
  x += offsetX; y += offsetY;
}
void Point::print() {
  cout << "(" << x << "," << y << ")";
}
void Vector::offset(double offsetX, double offsetY) {
  start.offset(offsetX, offsetY);
  end.offset(offsetX, offsetY);
}
void Vector::print() {
  start.print();
  cout << " -> ";
  end.print();
  cout << endl;
}
```

:: indicates which class' method is being implemented

- Manually initializing your fields can get tedious
- Can we initialize them when we create an instance?

```
Vector vec;
vec.start.x = 0.0;
vec.start.y = 0.0;
vec.end.x = 0.0;
vec.end.y = 0.0;
```

```
Point p;
p.x = 0.0;
p.y = 0.0;
```

Constructors

- Method that is called when an instance is created

```
class Point {
public:
  double x, y;
  Point() {
    x = 0.0; y = 0.0; cout << "Point instance created" << endl;
  }
};

int main() {
  Point p; // Point instance created
  // p.x is 0.0, p.y is 0.0
}
```

Constructors

- Can accept parameters

```
class Point {
public:
  double x, y;
  Point(double nx, double ny) {
    x = nx; y = ny; cout << "2-parameter constructor" << endl;
  }
};

int main() {
  Point p(2.0, 3.0); // 2-parameter constructor
  // p.x is 2.0, p.y is 3.0
}
```

Constructors

- Can have multiple constructors

```
class Point {
public:
  double x, y;
  Point() {
    x = 0.0; y = 0.0; cout << "default constructor" << endl;
  }
  Point(double nx, double ny) {
    x = nx; y = ny; cout << "2-parameter constructor" << endl;
  }
};

int main() {
  Point p; // default constructor
  // p.x is 0.0, p.y is 0.0)
  Point q(2.0, 3.0); // 2-parameter constructor
  // q.x is 2.0, q.y is 3.0)
}
```

- Recall that assigning one class instance to another copies all fields (default **copy constructor**)

```
class Point {
public:
  double x, y;
  Point() {
    x = 0.0; y = 0.0; cout << "default constructor" << endl;
  }
  Point(double nx, double ny) {
    x = nx; y = ny; cout << "2-parameter constructor" << endl;
  }
};

int main() {
  Point q(1.0, 2.0); // 2-parameter constructor
  Point r = q;
  // r.x is 1.0, r.y is 2.0)
}
```

Invoking the copy constructor

- You can define your own copy constructor

```
class Point {
public:
  double x, y;
  Point(double nx, double ny) {
    x = nx; y = ny; cout << "2-parameter constructor" << endl;
  }
  Point(Point &o) {
    x = o.x; y = o.y; cout << "custom copy constructor" << endl;
  }
};

int main() {
  Point q(1.0, 2.0); // 2-parameter constructor
  Point r = q; // custom copy constructor
  // r.x is 1, r.y is 2
}
```

- Why make a copy constructor? Assigning all fields (default copy constructor) may not be what you want

```
class MITStudent {
public:
  int studentID;
  char *name;
  MITStudent() {
    studentID = 0;
    name = "";
  }
};
```

```
int main() {
  MITStudent student1;
  student1.studentID = 98;
  char n[] = "foo";
  student1.name = n;
  MITStudent student2 = student1;
  student2.name[0] = 'b';
  cout << student1.name; // boo
}
```

By changing student 2's name, we changed student 1's name as well

- Why make a copy constructor? Assigning all fields (default copy constructor) may not be what you want

```
class MITStudent {
public:
  int studentID;
  char *name;
  MITStudent() {
    studentID = 0;
    name = "";
  }
  MITStudent(MITStudent &o) {
    studentID = o.studentID;
    name = strdup(o.name);
  }
};
```

```
int main() {
  MITStudent student1;
  student1.studentID = 98;
  char n[] = "foo";
  student1.name = n;
  MITStudent student2 = student1;
  student2.name[0] = 'b';
  cout << student1.name; // foo
}
```

Changing student 2’s name doesn’t effect student 1’s name

Access Modifiers

- Define where your fields/methods can be accessed from

Access Modifier

```
class Point {
public:
  double x, y;

  Point(double nx, double ny) {
    x = nx; y = ny;
  }
};
```

Access Modifiers

- public: can be accessed from anywhere

```
class Point {
public:
  double x, y;

  Point(double nx, double ny) {
    x = nx; y = ny;
  }
};

int main() {
  Point p(2.0,3.0);
  p.x = 5.0; // allowed
}
```

Access Modifiers

- private: can only be accessed within the class

```
class Point {
private:
  double x, y;

public:
  Point(double nx, double ny) {
    x = nx; y = ny;
  }
};

int main() {
  Point p(2.0,3.0);
  p.x = 5.0; // not allowed
}
```

Access Modifiers

- Use getters to allow read-only access to private fields

```
class Point {
private:
  double x, y;

public:
  Point(double nx, double ny) {
    x = nx; y = ny;
  }
  double getX() { return x; }
  double getY() { return y; }
};

int main() {
  Point p(2.0,3.0);
  cout << p.getX() << endl; // allowed
}
```

Default Access Modifiers

- class: private by default

```cpp
class Point {
  double x, y;
};
```

Equivalent to

```cpp
class Point {
private:
  double x, y;
};
```

Structs

- Structs are a carry-over from the C; in C++, classes are generally used
- In C++, they're essentially the same as classes, except structs' default access modifier is public

```
class Point {
public:
  double x;
  double y;
};
```

```
struct Point {

  double x;
  double y;
};
```

Default Access Modifiers

- struct: public by default
- class: private by default

```
struct Point {
  double x, y;
};
```

Equivalent to

```
struct Point {
public:
  double x, y;
};
```

```
class Point {
  double x, y;
};
```

Equivalent to

```
class Point {
private:
  double x, y;
};
```

Lecture 7 Notes: Object-Oriented Programming (OOP) and Inheritance

We've already seen how to define composite datatypes using classes. Now we'll take a step back and consider the programming philosophy underlying classes, known as *object-oriented programming* (OOP).

1 The Basic Ideas of OOP

Classic "procedural" programming languages before C++ (such as C) often focused on the question "What should the program do next?" The way you structure a program in these languages is:

1. Split it up into a set of tasks and subtasks
2. Make functions for the tasks
3. Instruct the computer to perform them in sequence

With large amounts of data and/or large numbers of tasks, this makes for complex and unmaintainable programs.

Consider the task of modeling the operation of a car. Such a program would have lots of separate variables storing information on various car parts, and there'd be no way to group together all the code that relates to, say, the wheels. It's hard to keep all these variables and the connections between all the functions in mind.

To manage this complexity, it's nicer to package up self-sufficient, modular pieces of code. People think of the world in terms of interacting *objects*: we'd talk about interactions between the steering wheel, the pedals, the wheels, etc. OOP allows programmers to pack away details into neat, self-contained boxes (objects) so that they can think of the objects more abstractly and focus on the interactions between them.

There are lots of definitions for OOP, but 3 primary features of it are:

- **Encapsulation:** grouping related data and functions together as objects and defining an *interface* to those objects
- **Inheritance:** allowing code to be reused between related types
- **Polymorphism:** allowing a value to be one of several types, and determining at runtime which functions to call on it based on its type

Let's see how each of these plays out in C++.

2 Encapsulation

Encapsulation just refers to packaging related stuff together. We've already seen how to package up data and the operations it supports in C++: with classes.

If someone hands us a class, we do not need to know how it actually works to use it; all we need to know about is its public methods/data – its *interface*. This is often compared to operating a car: when you drive, you don't care how the steering wheel makes the wheels turn; you just care that the interface the car presents (the steering wheel) allows you to accomplish your goal. If you remember the analogy from Lecture 6 about objects being boxes with buttons you can push, you can also think of the interface of a class as the set of buttons each instance of that class makes available. Interfaces abstract away the details of how all the operations are actually performed, allowing the programmer to focus on how objects will use each other's interfaces – how they interact.

This is why C++ makes you specify `public` and `private` access specifiers: by default, it assumes that the things you define in a class are internal details which someone using your code should not have to worry about. The practice of hiding away these details from client code is called "data hiding," or making your class a "black box."

One way to think about what happens in an object-oriented program is that we define what objects exist and what each one knows, and then the objects send messages to each other (by calling each other's methods) to exchange information and tell each other what to do.

3 Inheritance

Inheritance allows us to define hierarchies of related classes.

Imagine we're writing an inventory program for vehicles, including cars and trucks. We could write one class for representing cars and an unrelated one for representing trucks, but we'd have to duplicate the functionality that all vehicles have in common. Instead, C++ allows us to specify the common code in a `Vehicle` class, and then specify that the `Car` and `Truck` classes share this code.

The `Vehicle` class will be much the same as what we've seen before:

```
class Vehicle {
protected:
    string license;
    int year;
```

```
5
6 public:
7     Vehicle(const string &myLicense, const int myYear)
8         : license(myLicense), year(myYear) {}
9     const string getDesc() const
10        {return license + " from " + stringify(year);}
11    const string &getLicense() const {return license;}
12    const int getYear() const {return year;}
13 };
```

A few notes on this code, by line:

2. The standard `string` class is described in Section 1 of PS3; see there for details. Recall that `string`s can be appended to each other with the `+` operator.
3. `protected` is largely equivalent to `private`. We'll discuss the differences shortly.
8. This line demonstrates *member initializer syntax*. When defining a constructor, you sometimes want to initialize certain members, particularly `const` members, even before the constructor body. You simply put a colon before the function body, followed by a comma-separated list of items of the form `dataMember(initialValue)`.
10. This line assumes the existence of some function `stringify` for converting numbers to `string`s.

Now we want to specify that `Car` will inherit the `Vehicle` code, but with some additions. This is accomplished in line 1 below:

```
1 class Car : public Vehicle { // Makes Car inherit from Vehicle
2     string style;
3
4 public:
5     Car(const string &myLicense, const int myYear, const string
          &myStyle)
6         : Vehicle(myLicense, myYear), style(myStyle) {}
7     const string &getStyle() {return style;}
8 };
```

Now class `Car` has all the data members and methods of `Vehicle`, as well as a `style` data member and a `getStyle` method.

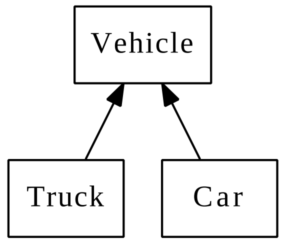
Class `Car` *inherits from* class `Vehicle`. This is equivalent to saying that `Car` is a *derived class*, while `Vehicle` is its *base class*. You may also hear the terms *subclass* and *superclass* instead.

Notes on the code:

1. Don't worry for now about why we stuck the `public` keyword in there.

6. Note how we use member initializer syntax to call the base-class constructor. We need to have a complete `Vehicle` object constructed before we construct the components added in the `Car`. If you do not explicitly call a base-class constructor using this syntax, the default base-class constructor will be called.

Similarly, we could make a `Truck` class that inherits from `Vehicle` and shares its code. This would give a *class hierarchy* like the following:



Class hierarchies are generally drawn with arrows pointing from derived classes to base classes.

3.1 Is-a vs. Has-a

There are two ways we could describe some class `A` as depending on some other class `B`:

1. Every `A` object *has a* `B` object. For instance, every `Vehicle` *has a* `string` object (called `license`).

2. Every instance of `A` *is a* `B` instance. For instance, every `Car` *is a* `Vehicle`, as well.

Inheritance allows us to define "is-a" relationships, but it should not be used to implement "has-a" relationships. It would be a design error to make `Vehicle` inherit from `string` because every `Vehicle` has a license; a `Vehicle` is not a `string`. "Has-a" relationships should be implemented by declaring data members, not by inheritance.

3.2 Overriding Methods

We might want to generate the description for `Car`s in a different way from generic `Vehicle`s. To accomplish this, we can simply redefine the `getDesc` method in `Car`, as below. Then, when we call `getDesc` on a `Car` object, it will use the redefined function. Redefining in this manner is called *overriding* the function.

```
class Car : public Vehicle { // Makes Car inherit from Vehicle
    string style;

```

```
public:
    Car(const string &myLicense, const int myYear, const string &myStyle)
        : Vehicle(myLicense, myYear), style(myStyle) {}
    const string getDesc() // Overriding this member function
        {return stringify(year) + ' ' + style + ": " + license;}
    const string &getStyle() {return style;}
};
```

3.2.1 Programming by Difference

In defining derived classes, we only need to specify what's different about them from their base classes. This powerful technique is called *programming by difference*.

Inheritance allows only overriding methods and adding new members and methods. We cannot remove functionality that was present in the base class.

3.3 Access Modifiers and Inheritance

If we'd declared `year` and `license` as `private` in `Vehicle`, we wouldn't be able to access them even from a derived class like `Car`. To allow derived classes but not outside code to access data members and member functions, we must declare them as `protected`.

The `public` keyword used in specifying a base class (e.g., `class Car : public Vehicle {...}`) gives a limit for the visibility of the inherited methods in the derived class. Normally you should just use `public` here, which means that inherited methods declared as `public` are still `public` in the derived class. Specifying `protected` would make inherited methods, even those declared `public`, have at most `protected` visibility. For a full table of the effects of different inheritance access specifiers, see
http://en.wikibooks.org/wiki/C++_Programming/Classes/Inheritance.

4 Polymorphism

Polymorphism means "many shapes." It refers to the ability of one object to have many types. If we have a function that expects a `Vehicle` object, we can safely pass it a `Car` object, because every `Car` is also a `Vehicle`. Likewise for references and pointers: anywhere you can use a `Vehicle *`, you can use a `Car *`.

4.1 `virtual` Functions

There is still a problem. Take the following example:

```
1 Car c("VANITY", 2003);
2 Vehicle *vPtr = &c;
3 cout << vPtr->getDesc();
```

(The `->` notation on line 3 just dereferences and gets a member. `ptr->member` is equivalent to `(*ptr).member`.)

Because `vPtr` is declared as a `Vehicle *`, this will call the `Vehicle` version of `getDesc`, even though the object pointed to is actually a `Car`. Usually we'd want the program to select the correct function at runtime based on which kind of object is pointed to. We can get this behavior by adding the keyword `virtual` before the method definition:

```
class Vehicle {
    ...
    virtual const string getDesc() {...}
};
```

With this definition, the code above would correctly select the `Car` version of `getDesc`.

Selecting the correct function at runtime is called *dynamic dispatch*. This matches the whole OOP idea – we're sending a message to the object and letting it figure out for itself what actions that message actually means it should take.

Because references are implicitly using pointers, the same issues apply to references:

```
Car c("VANITY", 2003);
Vehicle &v = c;
cout << v.getDesc();
```

This will only call the `Car` version of `getDesc` if `getDesc` is declared as `virtual`.

Once a method is declared `virtual` in some class `C`, it is virtual in every derived class of `C`, even if not explicitly declared as such. However, it is a good idea to declare it as `virtual` in the derived classes anyway for clarity.

4.2 Pure `virtual` Functions

Arguably, there is no reasonable way to define `getDesc` for a generic `Vehicle` – only derived classes really need a definition of it, since there is no such thing as a generic vehicle that isn't also a car, truck, or the like. Still, we do want to require every derived class of `Vehicle` to have this function.

We can omit the definition of `getDesc` from `Vehicle` by making the function *pure virtual* via the following odd syntax:

```
class Vehicle {
    ...
    virtual const string getDesc() = 0; // Pure virtual
};
```

The `= 0` indicates that no definition will be given. This implies that one can no longer create an instance of `Vehicle`; one can only create instances of `Car`s, `Truck`s, and other derived classes which do implement the `getDesc` method. `Vehicle` is then an *abstract class* – one which defines only an interface, but doesn't actually implement it, and therefore cannot be instantiated.

5 Multiple Inheritance

Unlike many object-oriented languages, C++ allows a class to have multiple base classes:

```
class Car : public Vehicle, public InsuredItem {
    ...
};
```

This specifies that `Car` should have all the members of both the `Vehicle` and the `InsuredItem` classes.

Multiple inheritance is tricky and potentially dangerous:

- If both `Vehicle` and `InsuredItem` define a member `x`, you must remember to disambiguate which one you're referring to by saying `Vehicle::x` or `InsuredItem::x`.
- If both `Vehicle` and `InsuredItem` inherited from the same base class, you'd end up with two instances of the base class within each `Car` (a "dreaded diamond" class hierarchy). There are ways to solve this problem, but it can get messy.

In general, avoid multiple inheritance unless you know exactly what you're doing.