

CSCI-1200 Data Structures — Spring 2016

Lecture 1 — Introduction to C++, STL, & Strings

Co-Instructors **email:** `ds_instructors@cs.rpi.edu`

Professor William Thompson
Amos Eaton(AE) 205 x6861
`thompw4@rpi.edu`

Professor Elsa Gonsiorowski
`gonsie@rpi.edu`

Today

- Discussion of Website & Syllabus:
<http://www.cs.rpi.edu/academics/courses/spring16/csci1200/>
- Getting Started in C++ & STL, C++ Syntax, STL Strings

1.1 Transitioning from Python to C++ (from CSCI-1100 Computer Science 1)

- Python is a great language to learn the power and flexibility of programming and computational problem solving. This semester we will work in C++ and study lower level programming concepts, focusing on details including efficiency and memory usage.
- Outside of this class, when working on large programming projects, you will find it is not uncommon to use a mix of programming languages and libraries. The individual advantages of Python and C++ (and Java, and Perl, and C, and UNIX bash scripts, and ...) can be combined into an elegant (or terrifyingly complex) masterpiece.
- Here are a few excellent references recommended to help you transition from Python to C++:
<http://cs.slu.edu/~goldwasser/publications/python2cpp.pdf>
<http://www4.wittenberg.edu/academics/mathcomp/shelburne/comp255/notes/Python2Cpp.pdf>

1.2 Compiled Languages vs. Interpreted Languages

- C/C++ is a *compiled language*, which means your code is processed (compiled & linked) to produce a low-level machine language executable that can be run on your specific hardware. You must re-compile & re-link after you edit any of the files – although a smart development environment or **Makefile** will figure out what portions need to be recompiled and save some time (especially on large programming projects with many lines of code and many files). Also, if you move your code to a different computer you will usually need to recompile. Generally the extra work of compilation produces an efficient and optimized executable that will run fast.
- In contrast, many newer languages including Python, Java, & Perl are *interpreted languages*, that favor incremental development where you can make changes to your code and immediately run all or some of your code without waiting for compilation. However, an interpreted program will often run slower than a compiled program.
- These days, the process of compilation is almost instantaneous for simple programs, and in this course we encourage you to follow the same incremental editing & frequent testing development strategy that is employed with interpreted languages.
- Finally, many interpreted languages have a Just-In-Time-Compiler (JIT) that can run an interpreted programming language and perform optimization on-the-fly resulting in program performance that rivals optimized compiled code. Thus, the differences between compiled and interpreted languages are somewhat blurry.
- You will practice the cycle of coding & compilation & testing during Lab 1. You are encouraged to try out different development environments (code editor & compiler) and quickly settle on one that allows you to be most productive. Ask the your lab TAs & mentors about their favorite programming environments! The course website includes many helpful links as well.
- As you see in today's handout, C++ has more required punctuation than Python, and the syntax is more restrictive. The compiler will proofread your code in detail and complain about any mistakes you make. Even long-time C++ programmers make mistakes in syntax, and with practice you will become familiar with the compiler's error messages and how to correct your code.

1.3 A Sample C++ Program: Find the Roots of a Quadratic Polynomial

```
#include <iostream>    // library for reading & writing from the console/keyboard
#include <cmath>        // library with the square root function & absolute value
#include <cstdlib>      // library with the exit function

// Returns true if the candidate root is indeed a root of the polynomial  $a*x*x + b*x + c = 0$ 
bool check_root(int a, int b, int c, float root) {
    // plug the value into the formula
    float check = a * root * root + b * root + c;
    // see if the absolute value is zero (within a small tolerance)
    if (fabs(check) > 0.0001) {
        std::cerr << "ERROR: " << root << " is not a root of this formula." << std::endl;
        return false;
    } else {
        return true;
    }
}

/* Use the quadratic formula to find the two real roots of polynomial. Returns
true if the roots are real, returns false if the roots are imaginary. If the roots
are real, they are returned through the reference parameters root_pos and root_neg. */
bool find_roots(int a, int b, int c, float &root_pos, float &root_neg) {
    // compute the quantity under the radical of the quadratic formula
    int radical = b*b - 4*a*c;
    // if the radical is negative, the roots are imaginary
    if (radical < 0) {
        std::cerr << "ERROR: Imaginary roots" << std::endl;
        return false;
    }
    float sqrt_radical = sqrt(radical);
    // compute the two roots
    root_pos = (-b + sqrt_radical) / float(2*a);
    root_neg = (-b - sqrt_radical) / float(2*a);
    return true;
}

int main() {
    // We will loop until we are given a polynomial with real roots
    while (true) {
        std::cout << "Enter 3 integer coefficients to a quadratic function:  $a*x*x + b*x + c = 0$ " << std::endl;
        int my_a, my_b, my_c;
        std::cin >> my_a >> my_b >> my_c;
        // create a place to store the roots
        float root_1, root_2;
        bool success = find_roots(my_a, my_b, my_c, root_1, root_2);
        // If the polynomial has imaginary roots, skip the rest of this loop and start over
        if (!success) continue;
        std::cout << "The roots are: " << root_1 << " and " << root_2 << std::endl;
        // Check our work...
        if (check_root(my_a, my_b, my_c, root_1) && check_root(my_a, my_b, my_c, root_2)) {
            // Verified roots, break out of the while loop
            break;
        } else {
            std::cerr << "ERROR: Unable to verify one or both roots." << std::endl;
            // if the program has an error, we choose to exit with a
            // non-zero error code
            exit(1);
        }
    }
    // by convention, main should return zero when the program finishes normally
    return 0;
}
```

1.4 Some Basic C++ Syntax

- Comments are indicated using `//` for single line comments and `/*` and `*/` for multi-line comments.
- `#include` asks the compiler for parts of the standard library and other code that we wish to use (e.g. the input/output stream function `std::cout`).
- `int main()` is a necessary component of all C++ programs; it returns a value (integer in this case) **and** it may have parameters.
- `{ }`: the curly braces indicate to C++ to treat *everything* between them as a unit.

1.5 The C++ Standard Library, a.k.a. “STL”

- The standard library contains types and functions that are important extensions to the core C++ language. We will use the standard library to such a great extent that it will feel like part of the C++ core language. `std` is a *namespace* that contains the standard library.
- I/O streams are the first component of the standard library that we see. `std::cout` (“console output”) and `std::endl` (“end line”) are defined in the standard library header file, `iostream`

1.6 A few notes on C++ vs. Java

- In Java, everything is an object and everything “inherits” from `java.lang.Object`. In C++, functions can exist outside of classes. In particular, the `main` function is never part of a class.
- Source code file organization in C++ does not need to be related to class organization as it does in Java. On the other hand, creating one C++ class (when we get to classes) per file is the *preferred* organization, with the *main* function in a separate file on its own or with a few helper functions.

1.7 Variables and Types

- A *variable* is an object with a name. A name is C++ identifier such as “a”, “root_1”, or “success”.
- An *object* is computer memory that has a type. A type (e.g., `int`, `float`, and `bool`) is a memory structure and a set of operations.
- For example, `float` is a type and each `float` variable is assigned to 4 bytes of memory, and this memory is formatted according IEEE floating point standards for what represents the exponent and mantissa. There are many operations defined on floats, including addition, subtraction, printing to the screen, etc.
- In C++ and Java the programmer must specify the data type when a new variable is declared. The C++ compiler enforces type checking (a.k.a. *static typing*). In contrast, the programmer does not specify the type of variables in Python and Perl. These languages are *dynamically-typed* — the interpreter will deduce the data type at runtime.

1.8 Expressions, Assignments and Statements

Consider the *statement*: `root_pos = (-b + sqrt_radical) / float(2*a);`

- The calculation on the right hand side of the `=` is an expression. You should review the definition of C++ arithmetic expressions and operator precedence from any reference textbook. The rules are pretty much the same in C++ and Java and Python.
- The value of this expression is assigned to the memory location of the float variable `root_pos`. Note also that if all expression values are type `int` we need a *cast* from `int` to `float` to prevent the truncation of integer division.
- The `float(2*a)` expression casts the integer value `2*a` to the proper float representation.

1.9 Conditionals and IF statements

- The general form of an if-else statement is

```
if (conditional-expression)
    statement;
else
    statement;
```

- Each **statement** may be a single statement, such as the `cout` statement above, a structured statement, or a compound statement delimited by `{...}`.

1.10 Functions and Arguments

- Functions are used to:
 - Break code up into modules for ease of programming and testing, and for ease of reading by other people (never, ever, under-estimate the importance of this!).
 - Create code that is reusable at several places in one program and by several programs.
- Each function has a sequence of parameters and a return type. The function *prototype* below has a return type of `bool` and five parameters.

```
bool find_roots(int a, int b, int c, float &root_pos, float &root_neg);
```

- The order and types of the parameters in the calling function (the main function in this example) must match the order and types of the parameters in the function prototype.

1.11 Value Parameters and Reference Parameters

- What’s with the `&` symbol on the 4th and 5th parameters in the `find_roots` function prototype?
- Note that when we call this function, we haven’t yet stored anything in those two root variables.


```
float root_1, root_2;
bool success = find_roots(my_a, my_b, my_c, root_1, root_2);
```
- The first three parameters to this function are *value parameters*.
 - These are essentially local variables (in the function) whose initial values are *copies* of the values of the corresponding argument in the function call.
 - Thus, the value of `my_a` from the main function is used to initialize `a` in function `find_roots`.
 - Changes to value parameters within the called function do NOT change the corresponding argument in the calling function.
- The final two parameters are *reference parameters*, as indicated by the `&`.
 - Reference parameters are just aliases for their corresponding arguments. No new objects are created.
 - As a result, changes to reference parameters are changes to the corresponding variables (arguments) in the calling function.
- In general, the “Rules of Thumb” for using value and reference parameters:
 - When a function (e.g., `check_root`) needs to provide just one simple result, make that result the return value of the function and pass other parameters by value.
 - When a function needs to provide more than one result (e.g., `find_roots`, these results should be returned using multiple reference parameters.
- We’ll see more examples of the importance of value vs. reference parameters as the semester continues.

1.12 for & while Loops

- Here is the basic form of a `for` loop:


```
for (expr1; expr2; expr3)
    statement;
```

 - `expr1` is the initial expression executed at the start before the loop iterations begin;
 - `expr2` is the test applied before the beginning of each loop iteration, the loop ends when this expression evaluates to `false` or `0`;
 - `expr3` is evaluated at the very end of each iteration;
 - `statement` is the “loop body”
- Here is the basic form of a `while` loop:


```
while (expr)
    statement;
```

`expr` is checked before entering the loop and after each iteration. If `expr` ever evaluates the false the loop is finished.

1.13 C-style Arrays

- An array is a fixed-length, consecutive sequence of objects all of the same type. The following declares an array with space for 15 double values. Note the spots in the array are currently *uninitialized*.

```
double a[15];
```

- The values are accessed through subscripting operations. The following code assigns the value 3.14159 to location `i=5` of the array. Here `i` is the *subscript* or *index*.

```
int i = 5;
a[i] = 3.14159;
```

- In C/C++, array indexing starts at 0.
- Arrays are fixed size, and each array knows *NOTHING* about its own size. The programmer must keep track of the size of each array. (Note: C++ STL has generalization of C-style arrays, called *vectors*, which do not have these restrictions. More on this in Lecture 2!)

1.14 Python Strings vs. C chars vs. C-style Strings vs. C++ STL Strings

- Strings in Python are immutable, and there is no difference between a string and a char in Python. Thus, `'a'` and `"a"` are both strings in Python, not individual characters. In C++ & Java, single quotes create a character type (exactly one character) and double quotes create a string of 0, 1, 2, or more characters.
- A “C-style” string is an array of `chars` that ends with the special char `'\0'`. C-style strings (`char*` or `char[]`) can be edited, and there are a number of helper functions to help with common operations. However...
- The “C++-style” STL `string` type has a wider array of operations and functions, which are more convenient and more powerful.

1.15 About STL String Objects

- A `string` is an object type defined in the standard library to contain a sequence of characters.
- The string type, like all types (including `int`, `double`, `char`, `float`), defines an interface, which includes construction (initialization), operations, functions (methods), and even other types(!).
- When an object is created, a special function is run called a “constructor”, whose job it is to initialize the object. There are several ways of constructing string objects:
 - By default to create an empty string: `std::string my_string_var;`
 - With a specified number of instances of a single char: `std::string my_string_var2(10, ' ');`
 - From another string: `std::string my_string_var3(my_string_var2);`
- The notation `my_string_var.size()` is a call to a function `size` that is defined as a **member function** of the `string` class. There is an equivalent member function called `length`.
- Input to string objects through streams (e.g. reading from the keyboard or a file) includes the following steps:
 1. The computer inputs and discards white-space characters, one at a time, until a non-white-space character is found.
 2. A sequence of non-white-space characters is input and stored in the string. This overwrites anything that was already in the string.
 3. Reading stops either at the end of the input or upon reaching the next white-space character (without reading it in).
- The (overloaded) operator `'+'` is defined on strings. It concatenates two strings to create a third string, without changing either of the original two strings.
- The assignment operation `'='` on strings overwrites the current contents of the string.
- The individual characters of a string can be accessed using the subscript operator `[]` (similar to arrays).
 - Subscript 0 corresponds to the first character.
 - For example, given `std::string a = "Susan";`
Then `a[0] == 'S'` and `a[1] == 'u'` and `a[4] == 'n'`.

- Strings define a special type `string::size_type`, which is the type returned by the string function `size()` (and `length()`).
 - The `::` notation means that `size_type` is defined within the scope of the `string` type.
 - `string::size_type` is generally equivalent to `unsigned int`.
 - You may see have compiler warnings and potential compatibility problems if you compare an `int` variable to `a.size()`.

This seems like a lot to remember. Do I need to memorize this? Where can I find all the details on `string` objects?

1.16 C++ vs. Java

- Standard C++ library `std::string` objects behave like a combination of Java `String` and `StringBuffer` objects. If you aren't sure of how a `std::string` member function (or operator) will behave, check its semantics or try it on small examples (or both, which is preferable).
- Java objects must be created using `new`, as in:

```
String name = new String("Chris");
```

This is not necessary in C++. The C++ (approximate) equivalent to this example is:

```
std::string name("Chris");
```

Note: There is a `new` operator in C++ and its behavior is somewhat similar to the `new` operation in Java. We will study it in a couple weeks.

1.17 Problem: Writing a Name Along a Diagonal

- Let's study a simple program to read in a name using `std::cin` and then output a fancier version to `std::cout`, written along a diagonal inside a box of asterisks. Here's how the program should behave:

```
What is your first name? Bob
```

```
*****
*      *
* B    *
*  o   *
*   b  *
*      *
*****
```

- There are two main difficulties:
 - Making sure that we can put the characters in the right places on the right lines.
 - Getting the asterisks in the right positions and getting the right number of blanks on each line.

```
#include <iostream>
#include <string>

int main() {
    std::cout << "What is your first name? ";
    std::string first;
    std::cin >> first;
    const std::string star_line(first.size()+4, '*');
    std::string middle_line = "*" + std::string(first.size()+2, ' ') + "*";
    std::cout << '\n' << star_line << '\n' << middle_line << std::endl;
    // Output the interior of the greeting, one line at a time.
    for (unsigned int i = 0; i < first.size(); ++i) {
        // Create the output line by overwriting a single character from the
        // first name in location i+2. After printing it restore the blank.
        middle_line[ i+2 ] = first[i];
        std::cout << middle_line << '\n';
        middle_line[ i+2 ] = ' ';
    }
    std::cout << middle_line << '\n' << star_line << std::endl;
    return 0;
}
```

CSCI-1200 Data Structures — Spring 2016

Collaboration Policy & Academic Integrity

iClicker Lecture exercises

Responses to iClicker lecture exercises will be used to earn incentives for the Data Structures course. Discussion of collaborative iClicker lecture exercises with those seated around you is encouraged. However, if we find anyone using an iClicker that is registered to another individual or using more than one iClicker, we will confiscate all iClickers involved and report the incident to the Dean of Students.

Academic Integrity for Exams

All exams for this course will be completed individually. Copying, communicating, or using disallowed materials during an exam is cheating, of course. Students caught cheating on an exam will receive an F in the course and will be reported to the Dean of Students for further disciplinary action.

Collaboration Policy for Programming Labs

Collaboration is encouraged during the weekly programming labs. Students are allowed to talk through and assist each other with these programming exercises. Students may ask for help from each other, the graduate lab TA, and undergraduate programming mentors. But each student must write up and debug their own lab solutions on their own laptop and be prepared to present and discuss this work with the TA to receive credit for each checkpoint.

As a general guideline, students may look over each other's shoulders at their labmate's laptop screen during lab — this is the best way to learn about IDEs, code development strategies, testing, and debugging. However, looking should not lead to line-by-line copying. Furthermore, each student should retain control of their own keyboard. While being assisted by a classmate or a TA, the student should remain fully engaged on problem solving and ask plenty of questions. Finally, other than the specific files provided by the instructor, electronic files or file excerpts should not be shared or copied (by email, text, Dropbox, or any other means).

Homework Collaboration Policy

Academic integrity is a complicated issue for individual programming assignments, but one we take very seriously. Students naturally want to work together, and it is clear they learn a great deal by doing so. Getting help is often the best way to interpret error messages and find bugs, even for experienced programmers. Furthermore, in-depth discussions about problem solving, algorithms, and code efficiency are invaluable and make us all better software engineers. In response to this, the following rules will be enforced for programming assignments:

- Students may read through the homework assignment together and discuss what is asked by the assignment, examples of program input & expected output, the overall approach to tackling the assignment, possible high level algorithms to solve the problem, and recent concepts from lecture that might be helpful in the implementation.
- Students are not allowed to work together in writing code or pseudocode. Detailed algorithms and implementation must be done individually. Students may not discuss homework code in detail (line-by-line or loop-by-loop) while it is being written or afterwards. In general, students should not look at each other's computer screen (or hand-written or printed assignment design notes) while working on homework. As a guideline, if an algorithm is too complex to describe orally (without dictating line-by-line), then sharing that algorithm is disallowed by the homework collaboration policy.
- Students are allowed to ask each other for help in interpreting error messages and in discussing strategies for testing and finding bugs. First, ask for help orally, by describing the symptoms of the problem. For each homework, many students will run into similar problems and after hearing a general description of a problem, another student might have suggestions for what to try to further diagnose or fix the issue. If that doesn't work, and if the compiler error message or flawed output is particularly lengthy, it is okay to ask another student to briefly look at the computer screen to see the details of the error message and the corresponding line of code. Please see a TA during office hours if a more in-depth examination of the code is necessary.
- Students may not share or copy code or pseudocode. Homework files or file excerpts should never be shared electronically (by email, text, LMS, Dropbox, etc.). Homework solution files from previous years

(either instructor or student solutions) should not be used in any way. Students must not leave their code (either electronic or printed) in publicly-accessible areas. Students may not share computers in any way when there is an assignment pending. Each student is responsible for securing their homework materials using all reasonable precautions. These precautions include: Students should password lock the screen when they step away from their computer. Homework files should only be stored on private accounts/computers with strong passwords. Homework notes and printouts should be stored in a locked drawer/room.

- Students may not show their code or pseudocode to other students as a means of helping them. Well-meaning homework help or tutoring can turn into a violation of the homework collaboration policy when stressed with time constraints from other courses and responsibilities. Sometimes good students who feel sorry for struggling students are tempted to provide them with “just a peek” at their code. Such “peeks” often turn into extensive copying, despite prior claims of good intentions.
- Students may not receive detailed help on their assignment code or pseudocode from individuals outside the course. This restriction includes tutors, students from prior terms, friends and family members, internet resources, etc.
- All collaborators (classmates, TAs, ALAC tutors, upperclassmen, students/instructor via LMS, etc.), and all of the resources (books, online reference material, etc.) consulted in completing this assignment must be listed in the README.txt file submitted with the assignment.

These rules are in place for each homework assignment and extends two days after the submission deadline.

Homework Plagiarism Detection and Academic Dishonesty Penalty

We use an automatic code comparison tool to help spot homework assignments that have been submitted in violation of these rules. The tool takes all assignments from all sections and all prior terms and compares them, highlighting regions of the code that are similar. The plagiarism tool looks at core code structure and is not fooled by variable and function name changes or addition of comments and whitespace.

The instructor checks flagged pairs of assignments very carefully, to determine which students may have violated the rules of collaboration and academic integrity on programming assignments. When it is believed that an incident of academic dishonesty has occurred, the involved students are contacted and a meeting is scheduled. All students caught cheating on a programming assignment (both the copier and the provider) will be punished. For undergraduate students, the standard punishment for the first offense is a 0 on the assignment and a full letter grade reduction on the final semester grade. Students whose violations are more flagrant will receive a higher penalty. Undergraduate students caught a second time will receive an immediate F in the course, regardless of circumstances. Each incident will be reported to the Dean of Students.

Graduate students found to be in violation of the academic integrity policy for homework assignments on the first offense will receive an F in the course and will be reported both to the Dean of Students and to the chair of their home department with the strong advisement that they be ineligible to serve as a teaching assistant for any course at RPI.

Academic Dishonesty in the Student Handbook

Refer to the *The Rensselaer Handbook of Student Rights and Responsibilities* for further discussion of academic dishonesty. Note that: “Students found in violation of the academic dishonesty policy are prohibited from dropping the course in order to avoid the academic penalty.”

Number of Students Found in Violation of the Policy

Historically, 5-10% of students are found to be in violation of the academic dishonesty policy each semester. Many of these students immediately admit to falling behind with the coursework and violating one or more of the rules above and if it is a minor first-time offense may receive a reduced penalty.

Read this document in its entirety. If you have any questions, contact the instructor or the TAs immediately. Sign this form and give it to your TA during your first lab section.

Name:

Section #:

Signature:

Date: