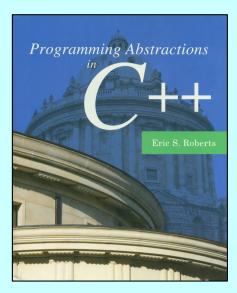
CHAPTER 12

Dynamic Memory Management

You have burdened your memory with exploded systems and useless names.

-Mary Shelley, Frankenstein, 1818



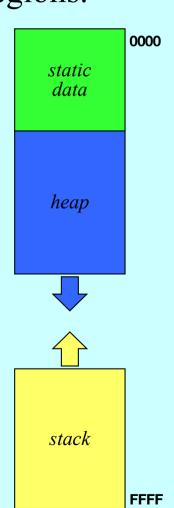
- 12.1 Dynamic allocation and the heap
- 12.2 Freeing memory
- 12.3 Heap-stack diagrams
- 12.4 Linked lists
- 12.5 Defining a CharStack class
- 12.6 Copying objects
- 12.7 The uses of const
- 12.8 Unit testing

Pointers

- In C++, pointers serve several purposes, of which the following are the most important:
 - Pointers allow you to refer to a large data structure in a compact way. Because a memory address typically fits in a few bytes of memory, this strategy offers considerable space savings when the data structures themselves are large. E.g., call by pointers.
 - Pointers make it possible to reserve new memory during program execution. In many applications, it is convenient to acquire new memory as the program runs and to refer to that memory using pointers, which is called dynamic allocation.
 - Pointers can be used to record relationships among data items.
 Data structures that use pointers to create connections between individual components are called *linked structures*. Programmers can indicate that one data item follows another in a conceptual sequence by including a pointer to the second item in the internal representation of the first.

The Allocation of Memory to Variables

- When you declare a variable in a program, C++ allocates space for that variable from one of several memory regions.
- One region of memory is reserved for program code and global variables/constants that persist throughout the lifetime of the program. This information is called *static data*.
- Each time you call a method, C++ allocates a new block of memory called a *stack frame* to hold its local variables. These stack frames come from a region of memory called the *stack*.
- It is also possible to allocate memory dynamically, as we will describe in Chapter 12. This space comes from a pool of memory called the *heap*.
- In classical architectures, the stack and heap grow toward each other to *maximize the available space*.



Dynamic Allocation

- C++ uses the **new** operator to allocate memory on the heap.
- You can allocate a single value (as opposed to an array) by writing new followed by the type name. Thus, to allocate space for a int on the heap, you would write:

```
int * pi = new int;
```

• You can allocate an array of values using the following form:

```
type * name = new type[size];
```

Thus, to allocate an array of 10000 integers, you would write:

```
int * arr = new int[10000];
```

• The delete operator frees memory previously allocated. For arrays, you need to include empty brackets, as in:

```
delete pi;

delete [] How does the compiler know how much memory to release?
```

Using delete

- Always pair a delete to a new, and a delete [] to a new [].
- No *size* needs to be specified for **delete** []. One of the approaches for compilers know how much memory to free is to allocate a little more memory and to store a count of the elements in a head segment (invisible to you) just before the first array element.
- delete ptr only frees the memory space pointed by ptr, but not the memory space occupied by ptr itself. ptr is still a live variable until it is released (depending on how it is declared), though it is dangling (i.e., does not point to a valid object of the appropriate type).
- To avoid dangling pointers, after deleting a pointer (freeing the memory space it points to), one can *nullify* a pointer by:

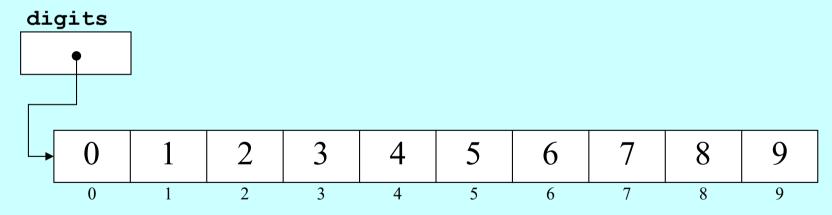
```
ptr = NULL; // nullptr since C++11
```

But there might be other pointers pointing to the same address that are not nullified.

Exercise: Dynamic Arrays

• Write a method createIndexArray(n) that returns an integer array of size n in which each element is initialized to its index. As an example, calling

should result in the following configuration:



How would you free the memory allocated by this call?

Exercise: Dynamic Arrays

```
/*
 * Function: createIndexArray
 * Usage: int *array = createIndexArray(n);
 * Returns a new dynamic integer array of length n
 * in which every element is set to its own index.
 */
int* createIndexArray(int n) {
                                       If we change dynamic
   int* array = new int[n];
                                       array to automatic:
   for (int i = 0; i < n; i++) {
                                       int array[n];
      array[i] = i;
                                       what happens?
   return array;
int main()
   int *digits = createIndexArray(10) :
                                               If we change 10 to 1, do we
   delete [] digits;
                                               still need [] after delete?
                                               Yes.
```

Garbage collection

- The big challenge in working with dynamic memory allocation is freeing the heap memory you allocate. Programs that fail to do so have what computer scientists call *memory leaks*.
- In Python and Java, objects are created on the heap and are automatically reclaimed by a *garbage collector* when those objects are no longer accessible. This discipline makes memory management very easy and convenient (for you).
- Garbage collection frees the programmer from manually dealing with memory deallocation. As a result, certain categories of bugs are eliminated or substantially reduced:
 - Certain kinds of memory leaks;
 - Dangling pointer bugs;
 - Double free bugs;
 - Efficient implementations of persistent data structures, etc.

```
char *p1 = new char;
char *p2 = new char;
p2 = p1;
delete p1;
p1 = NULL;
delete p2;
```

C++ Memory Management

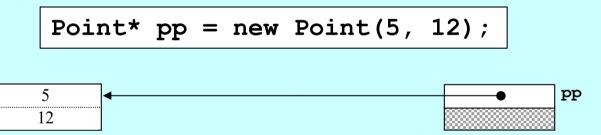
- Garbage collection, however, is not always efficient:
 - Consuming additional resources;
 - Performance impacts;
 - Possible stalls in program execution;
 - Incompatibility with manual resource management, etc.
- For performance, garbage collection is not adopted in C++:
 - Objects can be allocated either on the stack or in the heap.
 - Programmers must manage heap memory allocation explicitly.
- In object-oriented programming, it is nearly impossible for clients to remember exactly what heap storage is currently active when using objects. Fortunately, the designers of C++ made it possible to free heap memory allocated by objects (clients) when those objects *go out of scope* on the stack.
- In well-designed C++ programs, (the implementer of) each class takes responsibility for its own heap storage.

Allocating a Point Object

• The usual way to allocate a **Point** object is to declare it as a local variable on the stack, as follows:

Point pt(3, 4); FFFC 4

• It is, however, also possible to allocate a **Point** object on the heap using the following code:



• However, because of the data protection mechanism, it is nearly impossible for clients to know exactly what heap storage is currently active when deleting objects. For instance, some of the data fields might have been using the heap memory dynamically allocated by the member methods.

Destructors

- In C++, class definitions often include a *destructor*, which specifies how to free the storage used to represent an instance of that class.
- The prototype for a destructor has no return type and uses the name of the class preceded by a tilde (~). The destructor must not take any arguments.
- C++ calls the destructor automatically whenever a variable of a particular class is released. For stack objects, this happens when the function returns. The effect of this rule is that a C++ program that declares its objects as local variables on the stack will automatically reclaim those variables.
- If you instead allocate space for an object in the heap using new, you must explicitly free that object by calling delete. Calling delete also automatically invokes the destructor.
- If new is used in constructors, delete should most probably be used in destructors too.

Heap-Stack Diagrams

- It is easier to understand how C++ works if you have a good mental model of its use of memory. One of the most useful models is a *heap-stack diagram*, which shows the heap on the left and the stack on the right, separated by a dotted line.
- Whenever your program uses new, you need to add a block of memory to the heap side of the diagram. That block must be large enough to store the entire value you're allocating. If the value is a struct or an object type, that block must include space for all the members inside that structure.
- Whenever your program calls a method, you need to create a new stack frame by adding a block of memory to the stack side. For method calls, you need to add enough space to store the local variables for the method, again with some overhead information that tracks what the program is doing. When a method returns, C++ reclaims the memory in its frame.

Exercise: Heap-Stack Diagrams

```
int main() {

void nonsense(int list[], Point pt, double & total) {

   Point *pptr = new Point;

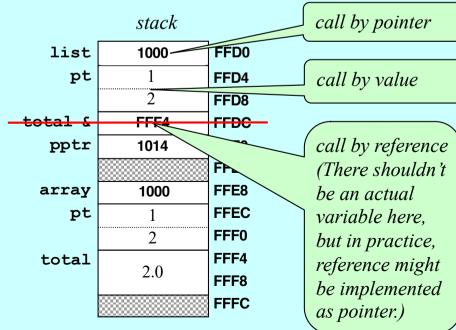
   list[1] = pt.x;

   total += pt.y;
}
```

Freed by the OS when the process terminates. But you should try to delete them consciously. Otherwise there is no point using the heap. heap

???
1 1000
1004
1008
??? 100C
??? 1010
??? 1014
1018

Forgotten memory



Dynamic Allocation in C

• Supported in <cstdlib> (stdlib.h)

| Dynamic memory management | |
|---------------------------|--|
| calloc | Allocate and zero-initialize array (function) |
| free | Deallocate memory block (function) |
| malloc | Allocate memory block (function) |
| realloc | Reallocate memory block (function) |

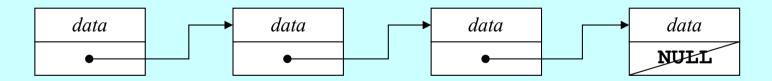
Things that you only need to know that you don't know. I don't need you to memorize it for the exam.

Pointers

- In C++, pointers serve several purposes, of which the following are the most important:
 - Pointers allow you to refer to a large data structure in a compact way. Because a memory address typically fits in a few bytes of memory, this strategy offers considerable space savings when the data structures themselves are large. E.g., call by pointers.
 - Pointers make it possible to reserve new memory during program execution. In many applications, it is convenient to acquire new memory as the program runs and to refer to that memory using pointers, which is called *dynamic allocation*.
 - Pointers can be used to record relationships among data items.
 Data structures that use pointers to create connections between individual components are called *linked structures*. Programmers can indicate that one data item follows another in a conceptual sequence by including a pointer to the second item in the internal representation of the first.

Linking Objects Together

- Pointers are important in programming because they make it possible to represent the relationship among objects by linking them together in various ways.
- The simplest example of a linked structure (which appears first in this chapter and is used throughout the later chapters) is called a *linked list*, in which each object in a sequence contains a reference to the one that follows it:



- C++ marks the end of linked list using the constant **NULL**, which signifies a pointer that does not have an actual target.
- In diagrams, the **NULL** value marking the end of a list is often indicated by drawing a diagonal line across the box.

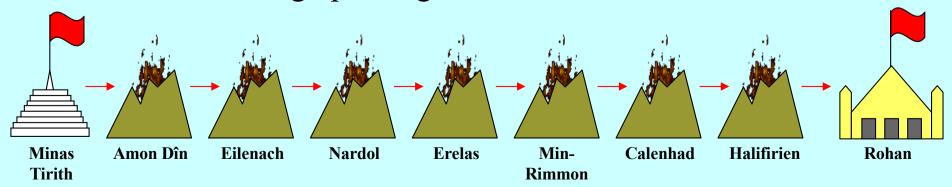
The Beacons of Gondor

For answer Gandalf cried aloud to his horse. "On, Shadowfax! We must hasten. Time is short. See! The beacons of Gondor are alight, calling for aid. War is kindled. See, there is the fire on Amon Dîn, and flame on Eilenach; and there they go speeding west: Nardol, Erelas, Min-Rimmon, Calenhad, and the Halifirien on the borders of Rohan."

—J. R. R. Tolkien, The Return of the King, 1955



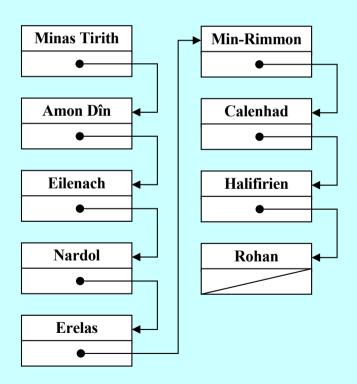
In a scene that was brilliantly captured in Peter Jackson's film adaptation of *The Return of the King*, Rohan is alerted to the danger to Gondor by a succession of signal fires moving from mountain top to mountain top. This scene is a perfect illustration of the idea of message passing in a linked list.



Message Passing in Linked Structures

To represent this message-passing image, you might use a definition such as the one shown on the right.

You can then initialize a chain of **Tower** structures, like this:



Calling signal on the first tower sends a message down the chain.

```
struct Tower {
   string name; /* The name of this tower
                                                 */
   Tower *link; /* Pointer to the next tower */
};
/*
  Function: createTower(name, link);
 * Creates a new Tower with the specified values.
 */
Tower *createTower(string name, Tower *link) {
   Tower *tp = new Tower;
   tp->name = name; >
                             Does the new tower go
   tp->link = link;
   return tp;
                             to the head or tail of the
                             linked list of old towers?
 * Function: signal(start);
 * Generates a signal beginning at start.
void signal(Tower *start) {
   if (start != NULL) {
      cout << "Lighting " << start->name << endl;</pre>
      signal(start->link);
```

Example: The CharStack Class

- Write classes that use dynamic allocation
- CharStack: a stack of characters

CharStack cstk;

Initializes an empty stack.

cstk.size()

Returns the number of characters pushed onto the stack.

cstk.isEmpty()

Returns true if the stack is empty.

cstk.clear()

Deletes all characters from the stack.

cstk.push(ch)

Pushes a new character onto the stack.

cstk.pop()

Removes and returns the top character from the stack.

The charstack.h Interface

```
* File: charstack.h
 * This interface defines the CharStack class.
 */
#ifndef charstack h
#define charstack h
class CharStack {
public:
 * CharStack constructor and destructor
 * The constructor initializes an empty stack. The destructor
 * is responsible for freeing heap storage.
 */
   CharStack();
   ~CharStack();
```

The charstack.h Interface

```
/*
 * Methods: size, isEmpty, clear, push, pop
 * These methods work exactly as they do for the Stack class.
 * The peek method is deleted here to save space.
 */
   int size();
  bool isEmpty();
   void clear();
   void push(char ch);
   char pop();
#include "charstackpriv.h"
#endif
```

The charstackpriv.h File

```
* File: charstackpriv.h
* This file contains the private data for the CharStack class.
*/
private:
/* Instance variables */
  char *array; /* Dynamic array of characters */
  int capacity; /* Allocated size of that array
  /* Private function prototypes */
  void expandCapacity();
```



arstack.cpp Implementation

```
tack.cpp
              mplements the CharStack class.
 */
#include "charstack.h"
#include "error.h"
using namespace std;
 * Constant: INITIAL CAPACITY
 * This constant defines the initial allocated size of the dynamic
 * array used to hold the elements. If the stack grows beyond its
 * capacity, the implementation doubles the allocated size.
 */
const int INITIAL CAPACITY = 10;
```

The charstack.cpp Implementation

```
/*
 * Implementation notes: constructor and destructor
 * The constructor allocates dynamic array storage to hold the
 * stack elements. The destructor must free these elements
 */
CharStack::CharStack() {
   capacity = INITIAL CAPACITY;
   array = new char[capacity];
   count = 0;
CharStack::~CharStack() {
   delete[] array;
```

The charstack.cpp Implementation

```
int CharStack::size() {
   return count;
bool CharStack::isEmpty() {
   return count == 0;
void CharStack::clear() {
   count = 0;
void CharStack::push(char ch) {
   if (count == capacity) expandCapacity();
   array[count++] = ch;
char CharStack::pop() {
   if (isEmpty()) error("pop: Attempting to pop an empty stack");
   return array[--count];
```

The charstack.cpp Implementation

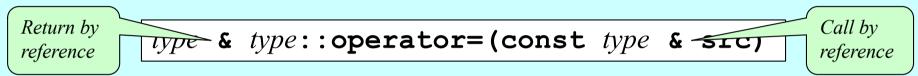
```
/*
 * Implementation notes: expandCapacity
 * This private method doubles the capacity of the elements array
 * whenever it runs out of space. To do so, it must copy the
 * pointer to the old array, allocate a new array with twice the
 * capacity, copy the characters from the old array to the new
 * one, and finally free the old storage.
 */
void CharStack::expandCapacity() {
   char *oldArray = array;
   capacity *= 2;
   array = new char[capacity];
   for (int i = 0; i < count; i++) {
     array[i] = oldArray[i];
   delete[] oldArray;
```

Copying Objects

- There is one remaining issue, from previous chapters, about creating new abstract classes that is extremely important in practice, which is how such objects behave if you copy them.
- When you are defining a new abstract data type in C++, you typically need to define two methods to ensure that copies are handled correctly:
 - The operator operator=, which takes care of assignments
 - A copy constructor, which takes care of by-value parameters
- These methods have well-defined signatures and structures, and the easiest thing to do is simply to copy the sample code on the next few slides, adapting it as necessary to account for the specific instance variables that need to be copied in the underlying representation.

Copying Objects

- The process of copying an object is controlled by two standard methods, each of which has a specific prototype.
- The *assignment operator* has the following form:



• The prototype for the *copy constructor* looks like this:

```
Constructors have the same name as the class (type).
```

- Call/return by value might cause unnecessary calls to the copy constructor. Constant call by reference protects the source.
- An assignment operator can return anything it wants, but the standard C and C++ assignment operators return a reference to the left-hand operand (e.g., for primitive types). This allows you to chain assignments together. Unless you have a really good reason, you want to follow this convention.

Assignment and Copy Constructors

- If you don't provide assignment operator and copy constructor yourself, C++ provides them for you.
- Unfortunately, the default behavior of C++ is to copy only the top-level fields in an object, which means that all dynamically allocated memory is shared between the original and the copy.
- This default behavior, which is called *shallow copying*, violates the semantics of data structures such as the collection classes. The collection classes in C++ are defined so that copying one collection to another creates an entirely new copy of the collection. What you need to implement instead is *deep copying*, which copies the data in the dynamically allocated memory as well.
- The simplest strategy for ensuring that clients don't violate the integrity of data structures through assignment is to prevent the client from copying an object altogether by making the assignment operator and copy constructor private methods. (Remember streams?)

Shallow vs. Deep Copying

• Suppose that you have an existing **Vector**<int> with three elements as shown in the diagram to the right.

| 1000 | 10 |
|------|----|
| | 20 |
| | 30 |
| | • |
| | • |
| | • |

| 1000 | elements |
|------|----------|
| 100 | capacity |
| 3 | count |

• A *shallow copy* allocates new fields for the object itself and copies the information from the original. Unfortunately, the dynamic array is copied as an address, not the data.

| 1000 | elements |
|------|----------|
| 100 | capacity |
| 3 | count |

• A *deep copy* also copies the contents of the dynamic array and therefore creates two independent structures

| 2000 | 10 | |
|------|----|--|
| | 20 | |
| | 30 | |
| | • | |
| | • | |

| 2000 | elements |
|------|----------|
| 100 | capacity |
| 3 | count |

Implementing Deep Copy Semantics

- This becomes an issue especially when you implement a class that uses dynamic memory.
- The crux of the problem is that copying an abstract data object typically needs to copy the underlying data (*deep copying*) and not just the fields directly accessible in the object.
- Unfortunately, the default interpretation in C++ is to copy only the top-level fields (*shallow copying*), which can lead to serious errors if you are expecting deep copying.
- Therefore, it is important to specify how the object can be properly copied in your implementation.



to Implement Deep Copying

n notes: copy constructor and assignment operator

```
make it possible to pass a CharStack by value or
                 arStack to another.
CharStack::CharStack(const CharStack & src) {
   deepCopy(src);
CharStack & CharStack::operator=(const CharStack & src) {
   if (this != &src) {
      delete[] array;
      deepCopy(src);
   return *this;
void CharStack::deepCopy(const CharStack & src) {
   array = new char[src.count]; capacity (a typo in the textbook, at least in my copy)
   for (int i = 0; i < src.count; i++) {</pre>
      array[i] = src.array[i];
   count = src.count;
   capacity = src.capacity;
```



Unit Testing

- One of the most important responsibilities you have as a programmer is to test your code as thoroughly as you can. Although testing can never guarantee the absence of errors, adopting a deliberate testing methodology helps you to find at least some of the errors in your code.
- Whenever you write a module, it is good practice to create a test program that checks the correctness of that module in isolation from the rest of the code. Such tests are called *unit tests*.
- The text uses the assert macro from the <cassert> library to implement the unit-testing strategy. If the conditional expression in the assert macro is true, execution continues normally. If the expression is false, the assert macro prints a message identifying the source of the error and exits from the program.

FIGURE 12-10 Unit test for the CharStack class

```
/*
 * File: CharStackUnitTest.cpp
 * This file contains a unit test of the CharStack class that uses the
 * C++ assert macro to test that each operation performs as it should.
#include <iostream>
#include <cassert>
#include "charstack.h"
using namespace std;
int main() {
                                             /* Declare an empty CharStack
   CharStack cstk;
                                             /* Make sure its size is 0
   assert(cstk.size() == 0);
                                                                             */
                                             /* And that isEmpty is true
                                                                             */
   assert(cstk.isEmpty());
   cstk.push('A');
                                             /* Push the character 'A'
                                                                             */
                                             /* The stack is now not empty
   assert(!cstk.isEmpty());
                                             /* And has size 1
   assert(cstk.size() == 1);
                                             /* Check that peek returns 'A' */
   assert(cstk.peek() == 'A');
                                             /* Push the character 'B'
   cstk.push('B');
   assert(cstk.peek() == 'B');
                                             /* Make sure peek returns it
                                             /* And that the size is now 2
   assert(cstk.size() == 2);
                                                                             */
                                             /* Pop and test for the 'B'
                                                                             */
   assert(cstk.pop() == 'B');
   assert(cstk.size() == 1);
                                             /* Recheck the size
                                                                             */
   assert(cstk.peek() == 'A');
                                             /* And make sure 'A' is on top */
   cstk.push('C');
                                             /* Test a push after a pop
   assert(cstk.size() == 2);
                                             /* Make sure size is correct
                                                                             */
   assert(cstk.pop() == 'C');
                                             /* And that pop returns a 'C'
   assert(cstk.peek() == 'A');
                                             /* The 'A' is now back on top
   assert(cstk.pop() == 'A');
                                             /* Pop and test for the 'A'
   assert(cstk.size() == 0);
                                             /* And make sure size is 0
                                                                             */
   for (char ch = 'A'; ch <= 'Z'; ch++) {
                                             /* Push the entire alphabet
                                                                             */
      cstk.push(ch);
                                                  one character at a time
                                                                             */
                                                  to test stack expansion
                                                                             */
                                             /* Make sure the size is 26
                                                                             */
   assert(cstk.size() == 26);
   for (char ch = 'Z'; ch >= 'A'; ch--) {
                                             /* Pop the characters in
                                                                             */
      assert(cstk.pop() == ch);
                                                reverse order to make
                                                                             */
                                                                             */
                                                  sure they're all there
   assert(cstk.isEmpty());
                                             /* Ensure the stack is empty
                                                                             */
   for (char ch = 'A'; ch <= 'Z'; ch++) {
                                             /* Push the alphabet again to
      cstk.push(ch);
                                                test that it works after
                                                                             */
                                                  expansion
                                             /* Check size is again 26
                                                                             */
   assert(cstk.size() == 26);
                                             /* Check the clear method
   cstk.clear();
   assert(cstk.size() == 0);
                                             /* And check if stack is empty */
   cstk.clear();
                                             /* Test clear with empty stack */
   assert(cstk.size() == 0);
   cout << "CharStack unit test succeeded" << endl;</pre>
   return 0;
```

The Uses of const

- The const keyword has many distinct purposes in C++. This text uses it in the following three contexts:
 - 1. Constant definitions. Adding the keyword const to a variable definition tells the compiler to disallow subsequent assignments to that variable, thereby making it constant. const double PI = 3.14159265358979323846; static const int INITIAL_CAPACITY = 10;
 - 2. Constant call by reference. Adding const to the declaration of a reference parameter signifies that the function will not change the value of that parameter. This guarantee allows the compiler to share the contents of a data structure without allowing methods to change it. void deepCopy (const CharStack & src);
 - 3. Constant methods. Adding const after the parameter list of a method guarantees that the method will not change the object.
 int CharStack::size() const
- Classes that use the const specification for all appropriate parameters and methods are said to be const-correct.

Class Constants

```
#include <iostream>
using namespace std;
class A {
public:
   const int c1:
   const int c2;
                         You can only use
                         the initializer
   A();
                         list to initialize
   A(int x, int y);
                         a class constant.
};
A::A():c1(3),c2(4){}
A::A(int x, int y):c1(x),c2(y) {}
int main()
   A a;
   A b(5, 6);
   cout << a.c1 << a.c2 << endl;</pre>
   cout << b.c1 << b.c2 << endl;
```

```
#include <iostream>
using namespace std;
class A {
public:
   const int c1 = 1; // Since C++11
   const int c2 = 2; // Since C++11
};
int main()
   A a, b;
   cout << a.c1 << a.c2 << endl;</pre>
   cout << b.c1 << b.c2 << endl:
   cout << (&a.c1 == &b.c1) << endl;
   cout \ll (\&a.c2 = \&b.c2) \ll end1;
```

They have different addresses even though they share the same constant value, and you can even re-initialize them using initializer list in the constructor functions.

Static Class Members

```
#include <iostream>
using namespace std;
class A {
public:
   static const int c = 1;
   static int i;
};
int A::i = 2;
int main()
   A a, b;
   cout << a.c << a.i << endl;</pre>
   cout << b.i << endl;</pre>
   cout << (&a.c == &b.c) << endl;
   cout << (&a.i == &b.i) << endl;
   a.i = 3;
   cout << b.i << endl;</pre>
```

Now they will have the same address and the same value.

The random.cpp Implementation

```
Implementation notes: initRandomSeed
   The initRandomSeed function declares a static variable that
   keeps track of whether the seed has been initialized.
 * first time initRandomSeed is called, initialized is false,
   so the seed is set to the current time.
 */
void initRandomSeed() {
   static bool initialized = false;
    if (!ini alized) {
       srand()
                      ime (NULL)));
       initializ
             The lifetime of static variables begins the first time the program flow
             encounters the declaration and it ends at program termination. The
             compiler allocates only one copy of initialized, which is initialized
             exactly once, and then shared by all calls to initRandomSeed. This
             ensures that the initialization step must be performed once and only once.
```

The Uses of static

| Applied to | Meaning | |
|---------------------------|--|--|
| a local | The variable is "permanent", in the sense that it is initialized only | |
| variable | once and retains its value from one function call to the next. It is like | |
| variable | having a global variable with local scope. | |
| | Since a global constant has internal linkage by Things that you only need | |
| a global | variables, which have external linkage by defito know that you don't | |
| constant | available for use in the file in which it is defir know. I don't need you to | |
| | effect, static const). memorize it for the exam. | |
| a global variable or a | The scope of the function, or the yar atole, is limited to the file in | |
| | which it is defined. Without a static qualifier, any free function or | |
| free function | global variable in a file has the extern qualifier by default, making it | |
| nee function | visible from other files in the compilation unit. | |
| a member | There is only one such variable for the class, no matter how many | |
| variable of a | objects of the class are created. In other words, it turns the member | |
| class. | variable from an "instance variable" into a "class variable". | |
| a member | The function may access only static members of the class. That is, it | |
| function of a | may not access any instance members (since there may not be any, if | |
| class. | no objects of the class have been created). | |

The End