Chapter 1 - Introduction to Java

Section 1.1 - Programming

A recipe consists of *instructions* that a chef executes, like adding eggs or stirring ingredients. Likewise, a *computer program* consists of instructions that a computer executes (or *runs*), like multiplying numbers or printing a number to a screen.

Figure 1.1.1: A program is like a recipe.



Bake chocolate chip cookies:

- Mix 1 stick of butter and 1 cup of sugar.
- Add egg and mix until combined.
- Stir in flour and chocolate.
- Bake at 350F for 8 minutes.



Ρ	Participation Activity 1.1.2: Instructions.	
Select tl	he instruction that achieves the desired goal.	
# Q	uestion	Your answer
Μ	Iake lemonade:Fill jug with water	Add salt
1	 Add lemon juice 	Add water
	• Stir	Add sugar
W	lash a car:	Rinse car with hose

2	 Fill bucket with soapy water Dip towel in bucket Wipe car with towel 	Add water to bucket Add sugar to bucket
	Wash hair:Rinse hair with water	Rinse hair with water
0	 While hair isn't squeaky clean, repeat: 	Apply shampoo to hair
5	 Work shampoo throughout hair Rinse hair with water 	Sing
	Compute the area of a triangle:	Multiply the previous answer by 2
4	 Determine the height Compute base times height 	Add 2 to the previous answer
	•	Divide the previous answer by 2

Section 1.2 - A first program

Below is a simple first Java program.



- The program consists of several lines of code. **Code** is the textual representation of a program. A **line** is a row of text.
- A program starts by executing a method called *main*. A method is a list of *statements* (see below). The various other items on main's line, as well as the line with the word "class", are described in later sections.
- "{" and "}" are called *braces*, denoting a list of statements. main's statements appear between braces.
- A *statement* is a program instruction. Each statement usually appears on its own line. Each program statement ends with a *semicolon* ";", like each English sentence ends with a period.
- The program ends when the *return;* statement executes.
- Each part of the program is described in later sections.

The following describes main's statements:

- Like a baker temporarily stores ingredients on a countertop, a program temporarily stores values in a memory. A *memory* is composed of numerous individual locations, each able to store a value. The statement int wage = 20 reserves a location in memory, names that location *wage*, and stores the value 20 in that location. A named location in memory, such as wage, is called a variable (because that value can vary).
- print statements print a program's output. println prints output followed by a new line.

Many code editors color certain words, as in the above program, to assist a human reader understand various words' roles.

A *compiler* is a tool that converts a program into low-level machine instructions (0s and 1s) understood by a particular computer. Because a programmer interacts extensively with a compiler, this material frequently refers to the compiler.



Participation Activity 1.2.3: Ba	asic program concepts.
Variable Statement	main Code Compiler Braces Line
Drag and drop above item	Textual representation of a program.
	Performs a specific action.
	A row of text.
	Delimits (surrounds) a list of statements.
	The starting place of a program.
	Represents a particular memory location.
	Converts a program into low-level machine instructions of a computer.
Reset	



Section 1.3 - Basic output

Printing of output to a screen is a common programming task. This section describes basic output; later sections have more details.

The **System.out.print** construct supports printing. Printing text is achieved via:

System.out.print("desired text");. Text in double quotes " " is known as a *string literal*. Multiple print statements continue printing on the same output line. System.out.println (note the ln at the end, short for "line"), starts a new line after the printed output.

```
Figure 1.3.1: Printing text and new lines.
   public class KeepCalm {
      public static void main (String [] args) {
         System.out.print("Keep calm");
         System.out.print("and");
                                        // Note: Does NOT print on new output
                                                                                 Keep ca
         System.out.print("carry on");
         return;
      }
   }
   public class KeepCalm {
      public static void main (String [] args) {
         System.out.println("Keep calm"); // println starts new line after out
                                                                                 Keep ca
         System.out.println("and");
                                                                                 and
         System.out.println("carry on"); // Usually finish output with new li
                                                                                 carry (
         return;
      }
   }
```

A common error is to use print when println should have been used, and vice-versa..

Printing a blank line is achieved by: System.out.println(""). Note that the string literal "" is empty.

A <u>common error</u> is to put single quotes around a string literal rather than double quotes, as in 'Keep calm', or to omit quotes entirely.

#	Question	Your answer
Which statement prints: Welcome!		<pre>System.out.print(Welcome!);</pre>
1		System.out.print "Welcome!";
		System.out.print("Welcome!")
	2 Which statement prints Hey followed by a new line?	System.out.print("Hey"ln);
2		System.out.println(Hey);
	System.out.println("Hey");	

P	Participation Activity 1.3.2: Basic text output.	1
End	each statement with a semicolon. Do not create a new	line unless instructed.
#	Question	Your answer
1	Type a statement that prints: Hello	
2	Type a statement that prints Hello and then starts a new output line.	

Printing the value of a variable is achieved via: System.out.print(variableName); Note the absence of quotes. println() could also be used.



Note that the programmer intentionally did *not* start a new output line after printing "Wage is: ", so that the wage variable's value would appear on that same line.

P	Participation Activity1.3.3: Basic variable output.	
#	Question	Your answer
	Given variable numCars = 9, which statement prints 9?	<pre>System.out.print("numCars");</pre>
1		System.out.print numCars;
		<pre>System.out.print(numCars);</pre>

P	Participation Activity1.3.4: Basic variable output.	
#	Question	Your answer
1	Type a statement that prints the value of numUsers (a variable). End statement with a semicolon.	

Programmers commonly try to use a single print statement for each line of output, by combining the printing of text, variable values, and new lines. The programmer simply separates the items with a + symbol. Such combining can improve program readability, because the program's code corresponds more closely to the program's printed output.

```
Figure 1.3.3: Printing multiple items using one print statement.

public class Salary {
    public static void main (String [] args) {
        int wage = 20;
        System.out.println("Wage is: " + wage); // The + separates ite
        Wage is: 20
        Goodbye.");
        return;
    }
}
```

A common error is to forget the + symbols between items, as in: System.out.println("Wage is: " wage);

P	Participation Activity1.3.5: Basic output.	
Indic	ate the actual output of each statement. Assume userAge is 22	
#	Question	Your answer
	<pre>System.out.print("You're " + userAge + " years.");</pre>	You're 22 years.
1		You're userAge years.
		No output; an error exists.
	<pre>System.out.print(userAge + "years is good.");</pre>	22 years is good.
2		22years is good.
		No output; an error exists.

A new line can also be output by inserting \n, known as a **newline character**, within a string literal. For example, printing "1\n2\n3" prints each number on its own output line. \n use is rare, but appears in some existing code so is mentioned here. \n consists of two characters, \ and n, but together are considered as one newline character. <u>Good practice</u> is to use println to print a newline, as println has some technical advantages not mentioned here.



Chall Activi	ty 1.3.1: Ge	enerate outpu	ut for given pr	ompt.
Start				
Type a singl Note: Each underline.	e statement that pr space is underlinec	oduces the follc I for clarity; you	wing output. should output a	space, not an
		u_v_w		
System.ou	ıt.print();
1	2	3	4	5
Check	Next			

C Challenge Activity	1.3.2: Enter t	he output.		
Start				
	Enter the out	put of the follov	ving program.	
<pre>public class ge public stati System.ou return; } }</pre>	neralOuput { c void main (Str: t.print("Sam is n	ing [] args) { nice.");	Sam is :	nice.
1	2	3	4	5
Check	Next			





Write a statement that prints the following on a single output line (without a newline):

3 2 1 Go!

Note: Whitespace (blank spaces / blank lines) matters; make sure your whitespace exactly matches t

```
1 public class OutputExample {
2     public static void main(String [] args) {
3     /* Your solution goes here */
5     return;
7     }
8 }
```





Section 1.4 - Basic input

Programs commonly require a user to enter input, such as typing a number, a name, etc. This section describes basic input; later sections have more details.

The following line (explained in a later section) at the top of a file enables a Java program to read input:



import java.util.Scanner;

Reading input is achieved by first creating a *scanner* object via the statement **scanner scnr = new Scanner(System.in)**, where System.in corresponds to keyboard input. Then, input integers can be read using hourlyWage = scnr.nextInt().

```
Figure 1.4.2: Reading user input
import java.util.Scanner; // Enables user input
public class Salary {
    public static void main(String [] args) {
        int hourlyWage = 0;
        int annualSalary = 0;
        Scanner scnr = new Scanner(System.in); // Setup to scan chars from System.in
        System.out.println("Enter hourly wage: ");
        hourlyWage = scnr.nextInt(); // Read next integer from scanner
        annualSalary = hourlyWage * 40 * 50;
        System.out.print("Salary is ");
        System.out.println(annualSalary);
        return;
        }
    }
```

P	Participation Activity 1.4.1: Basic input.	
#	Question	Your answer
	Assuming scnr already exists, which statement reads a user-entered number into variable numCars?	<pre>scnr.nextInt(numCars);</pre>
1		<pre>numCars = scnr.nextInt;</pre>
		<pre>numCars = scnr.nextInt();</pre>

P	Participation Activity 1.4.2: Basic input.	
#	Question	Your answer
1	Type a statement that reads a user-entered integer into variable numUsers. Assume scnr already exists.	







Section 1.5 - Comments and whitespace

A *comment* is text added to code by a programmer, intended to be read by humans to better understand the code, but ignored by the compiler. Two kinds of comments exist: a *single-line comment* uses the // symbols, and a *multi-line comment* uses the /* and */ symbols:

```
Construct 1.5.1: Comments.
// Single-line comment. The compiler ignores any text to the right, like ;, "Hi", //
/* Multi-line comment. The compiler ignores text until seeing the closing half of th
   so ignores ;, or (), or "Hi", or //, or /*, or num = num + 1, etc. Programmers us
   the opening and closing symbols and indent the comment text, but neither is manda
*/
```

The following program illustrates both comment types.

```
Figure 1.5.1: Comments example.
  import java.util.Scanner;
  /*
     This program calculates the amount of pasta to cook, given the
     number of people eating.
     Author: Mario Boyardee
     Date: March 9, 2014
  * /
  public class PastaCalculator {
     public static void main (String [] args) {
        Scanner scnr = new Scanner(System.in);
                             // Number of people that will be eating
        int numPeople = 0;
        int totalOuncesPasta = 0; // Total ounces of pasta to serve numPeople
        // Get number of people
        System.out.println("Enter number of people: ");
        numPeople = scnr.nextInt();
        // Calculate and print total ounces of pasta
        totalOuncesPasta = numPeople * 3; // Typical ounces per person
        System.out.println("Cook " + totalOuncesPasta + " ounces of pasta.");
        return:
     }
  }
```

Note that single-line comments commonly appear after a statement on the same line.

A multi-line comment is allowed on a single line, e.g., /* Typical ounces per person */. However, <u>good practice</u> is to use // for single-line comments, reserving /* */ for multi-line comments only. A multi-line comment is also known as a **block comment**.

Java supports a third type of comment, known as a JavaDoc comment, described in a later section.

Whitespace refers to blank spaces between items within a statement, and to blank lines between statements. A compiler ignores most whitespace.

The following animation provides a (simplified) demonstration of how a compiler processes code from left-to-right and line-by-line, finding each statement (and generating machine instructions using 0s and 1s), and ignoring comments.

D	Participation Activity	1.5.1: A compiler scans co left-to-right; whitespace is	ode line-by-line, mostly irrelevar
Star Syst	tem.out.pri	intln(wage * 40 * 50); /hr * 40 hrs/wk * 50 wks/year.	High-level
			piogram
Syst	tem.out.pri	<pre>intln("Enter integer: ");</pre>	myfile.java
Syst usei	rNum = scnr	<pre>intln("Enter integer: "); c.nextInt(); Compiler</pre>	myfile.java
Syst use 101: 001:	tem.out.pri rNum = scnr 10011011011 11010110111	<pre>intln("Enter integer: "); c.nextInt(); Compiler Ill Ill Ill Ill Ill Ill Ill Ill Ill Il</pre>	myfile.java Executable
Syst usei 101: 001: 111:	tem.out.pri rNum = scnr 10011011011 11010110111 00110110101 11000111101	<pre>intln("Enter integer: "); c.nextInt(); Compiler Ill Ill Ill Ill Ill Ill Ill Ill Ill Il</pre>	myfile.java Executable program myfile.class

Ρ	Participation Activity	1.5.2: Comments.			
Indicate which are valid code.					
#	# Question Your answer				

	// Get user input	Valid
1		Invalid
0	/* Get user input */	Valid
2		Invalid
0	<pre>/* Determine width and height, calculate volume, and return volume squared.</pre>	Valid
3	*/	Invalid
	// Print "Hello" to the screen //	Valid
4		Invalid
	<pre>// Print "Hello" Then print "Goodbye" And finally return.</pre>	Valid
5	//	Invalid
	/* * Author: Michelangelo * Date: 2014	Valid
6	* Address: 111 Main St, Pacific Ocean */	Invalid
	// numKids = 2; // Typical number	Valid
7		Invalid
	<pre>/* numKids = 2; // Typical number numCars = 5;</pre>	Valid
8	*/	Invalid
	1	

	9	<pre>/* numKids = 2; /* Typical number */ numCars = 5; */</pre>	Valid	
--	---	---	-------	--

The compiler ignores most whitespace. Thus, the following code is behaviorally equivalent to the above code, but terrible style (unless you are trying to get fired).

```
Figure 1.5.2: Bad use of whitespace.
import java.util.Scanner;
public class PastaCalculator {
    public static void main (String [] args) {
        Scanner scnr = new Scanner(System.in);int numPeople=0;int totalOuncesPasta=0;
        System.out.println("Enter number of people:");
        numPeople = scnr.nextInt(); totalOuncesPasta = numPeople * 3;
        System.out.println("Cook "+totalOuncesPasta+" ounces of pasta.");return;}}
```

In contrast, <u>good practice</u> is to deliberately and consistently use whitespace to make a program more readable. Blank lines separate conceptually distinct statements. Items may be aligned to reduce visual clutter. A single space before and after any operators like =, +, *, or << may make statements more readable. Each line is indented the same amount. *Programmers usually follow conventions defined by their company, team, instructor, etc.*

```
Figure 1.5.3: Good use of whitespace.
  import java.util.Scanner;
  public class WhiteSpaceEx {
     public static void main(String[] args) {
        Scanner scnr = new Scanner(System.in);
        int myFirstVar = 0; // Some programmers like to align the
        int yetAnotherVar = 0; // initial values. Not always possible.
        int thirdVar
                       = 0;
        // Above blank line separates variable definitions from the rest
        System.out.print("Enter a number: ");
        myFirstVar = scnr.nextInt();
        // Above blank line separates user input statements from the rest
        yetAnotherVar = myFirstVar; // Aligned = operators, and these aligned
                  = yetAnotherVar + 1; // comments yield less visual clutter.
        thirdVar
        // Also notice the single-space on left and right of + and =
        // (except when aligning the second = with the first =)
        System.out.println("Final value is " + thirdVar); // Single-space each side of
        return; // The above blank line separates the return from the rest
     }
  }
```

Section 1.6 - Errors and warnings

People make mistakes. Programmers thus make mistakes—lots of them. One kind of mistake, known as a *syntax error*, is to violate a programming language's rules on how symbols can be combined to create a program. An example is forgetting to end a statement with a semicolon.

Compilers are *extremely* picky. A compiler generates a message when encountering a syntax error. The following program is missing a semicolon after the first print statement.





Above, the 4 refers to the 4th line in the code.

u					
nd the syntax errors. Assume variable numDogs exists.					
Question		Your answer			
System.out.print(numDogs).		Error			
		No error			
	System.out.print("Dogs: " numDogs);	Error			
2		No error			
	<pre>system.out.print("Everyone wins.");</pre>	Error			
3		No error			
	<pre>System.out.print("Hello friends!);</pre>	Error			
ļ		No error			

5	<pre>System.out.print("Amy // Michael");</pre>	Error
5		No error
6	<pre>System.out.print(NumDogs);</pre>	Error
0		No error
7	<pre>int numCats = 3 System.out.print(numCats);</pre>	Error
1		No error
Q	<pre>System.print(numDogs);</pre>	Error
0		No error

-	Question
	import java.util.Scanner;
	<pre>public class CalcArea { public static void main(String[] args) { Scanner scnr = new Scanner(System.in); int triBase = 0; // Triangle base (cm) int triHeight = 0; // Triangle height (cm) int triArea = 0 // Triangle area (cm) System.out.print("Enter triangle base (cm): "); triBase = scnr.nextInt(); }</pre>
	System.out.print("Enter triangle height (cm): "); triHeight = nextInt();
1	// Calculate triangle area triArea = (triBase * triHeight) / 2;
	<pre>/* Print triangle base, height, area System.out.print("Triangle area = (") System.out.print(triBase); System.out.print(*); System.out.print("triHeight"); System.out.print(") / 2 = "); System.out.print(triArea); System.out.println(" cm^2");</pre>

Some compiler error messages are very precise, but some are less precise. Furthermore, many errors confuse a compiler, resulting in a misleading error message. *Misleading error messages are common. The message is like the compiler's "best guess" of what is really wrong.*



The compiler indicates a missing semicolon ';'. But the real error is the missing parentheses.

Sometimes the compiler error message refers to a line that is actually many lines past where the error actually occurred. Not finding an error at the specified line, the programmer should look to previous lines.

Stmt Stmt Stmt-with-e Stmt Stmt Stmt	error Compiler processing	Programmer should examine line 5 and earlier lines.
	Stmt Stmt Stmt-with-e Stmt Stmt Stmt	Stmt Stmt-with-error Stmt Stmt Stmt Compiler processing Stmt

	Ρ	Participation Activity1.6.4: Error messages.	
	#	Question	Your answer
	4	When a compiler says that an error exists on line 5, that line must have an error.	e True
1	I		False
	0	If a compiler says that an error exists on line 90, the actual error may be on line 91, 92, etc.	True False
2	2		
	0	If a compiler generates a specific message like "missing semicolon", then a semicolon must be missing somewhere,	True
	3	though maybe from an earlier line.	False

Some errors create an upsettingly long list of error messages. <u>Good practice</u> is to focus on fixing just the first error reported by the compiler, and then re-compiling. The remaining error messages may be real, but more commonly are due to the compiler's confusion caused by the first error and are thus irrelevant.





<u>Good practice</u>, especially for new programmers, is to compile after writing only a few lines of code, rather than writing tens of lines and then compiling. New programmers commonly write tens of lines before compiling, which may result in an overwhelming number of compilation errors and warnings.



Because a syntax error is detected by the compiler, a syntax error is known as a type of *compile-time error*.

New programmers commonly complain: "The program compiled perfectly but isn't working." Successfully compiling means the program doesn't have compile-time errors, but the program may have other kinds of errors. A *logic error* is an error that occurs while a program runs, also called a *runtime error* or *bug*. For example, a programmer might mean to type numBeans * numJars but accidentally types numBeans + numJars (+ instead of *). The program would compile, but would not run as intended.

```
Figure 1.6.4: Logic errors.

public class BeansInJars {
    public static void main (String [] args) {
        int numBeans = 500;
        int numJars = 3;
        int totalBeans = 0;
        System.out.print(numBeans + " beans in ");
        System.out.print(numJars + " jars yields ");
        totalBeans = numBeans + numJars; // Oops, used + instead of *
        System.out.println(totalBeans + " total");
        return;
    }
}
```

```
Participation
                   1.6.7: Fix the bug.
       Activity
Click run to compile and execute, and note the incorrect program output. Fix the bug in the
program.
    1
                                                                      Run
    2 public class BeansInJars {
          public static void main (String [] args) {
    3
    4
    5
             int numBeans = 500;
    6
             int numJars = 3;
    7
             int totalBeans = 0;
    8
    9
             System.out.print(numBeans + " beans in ");
             System.out.print(numJars + " jars yields ");
   10
   11
             totalBeans = numBeans * numJars;
   12
             System.out.println("totalBeans" + " total");
   13
   14
             return;
   15
          }
   16 }
   17
```

Figure 1.6.5: First bug.

The term "bug" to describe a runtime error was popularized when in 1947 engineers discovered their program on a Harvard University Mark II computer was not working because a moth was stuck in one of the relays (a type of mechanical switch). They taped the bug into their engineering log book, still preserved today (The moth).



A compiler will sometimes report a *warning*, which doesn't stop the compiler from creating an executable program, but indicates a possible logic error. For example, some compilers will report a warning like "Warning, dividing by 0 is not defined" if encountering code like:

totalItems = numItems / 0 (running that program does result in a runtime error). Even though the compiler may create an executable program, <u>good practice</u> is to write programs that compile without warnings. In fact, many programmers recommend the <u>good practice</u> of configuring compilers to print even more warnings. For example, javac can be run as <u>javac -Xlint yourfile.java</u>.

Ρ	Participation Activity1.6.8: Compiler warnings.		
#	Question	Your answer	
-	A compiler warning by default will prevent a program from being created.	True	
I		False	
0	Generally, a programmer should not ignore warnings.	True	
2		False	
0	A compiler's default settings cause most warnings to be reported during compilation.	True	
3		False	



1.6.1: Basic syntax errors.

Type the statements. Then, correct the one syntax error in each statement. Hints: Statements end in double quotes.

```
System.out.printl("Predictions are hard.");
System.out.print("Especially ');
System.out.println("about the future.").
System.out.println("Num is: " - userNum);
```

```
import java.util.Scanner;
 1
 2
 3 public class Errors {
 4
      public static void main(String [] args) {
 5
         int userNum = 5;
 6
 7
         /* Your solution goes here */
8
9
         return;
10
      }
11 }
```

Run

```
Challenge
                 1.6.2: More syntax errors.
Retype the statements, correcting the syntax errors.
System.out.println("Num: " + songnum);
System.out.println(int songNum);
System.out.println(songNum " songs");
   1 import java.util.Scanner;
   2
   3
     public class Errors {
   4
        public static void main (String [] args) {
   5
           int songNum = 5;
   6
   7
           /* Your solution goes here */
   8
   9
           return;
  10
         }
  11 }
        Run
```

Section 1.7 - Computers and programs



Just as knowing how a car works "under-the-hood" has benefits to a car owner, knowing how a computer works under-the-hood has benefits to a programmer. This section provides a very brief introduction.

When people in the 1800s began using electricity for lights and machines, they created switches to turn objects on and off. A *switch* controls whether or not electricity flows through a wire. In the early 1900s, people created special switches that could be controlled electronically, rather than by a person moving the switch up or down. In an electronically-controlled switch, a positive voltage at the control input allows electricity to flow, while a zero voltage prevents the flow. Such switches were useful, for example, in routing telephone calls. Engineers soon realized they could use electronically-controlled switches to perform simple calculations. The engineers treated a positive voltage as a "1" and a zero voltage as a "0". Os and 1s are known as *bits* (*binary digits*). They built connections of switches, known as *circuits*, to perform calculations such as multiplying two numbers.





These circuits became increasingly complex, leading to the first electronic computers in the 1930s and 1940s, consisting of about ten thousand electronic switches and typically occupying entire rooms as in the above figure. Early computers performed thousands of calculations per second, such as calculating tables of ballistic trajectories.

To support different calculations, circuits called **processors** were created to process (aka *execute*) a list of desired calculations, each calculation called **instruction**. The instructions were specified by configuring external switches, as in the figure on the left. Processors used to take up entire rooms, but today fit on a chip about the size of a postage stamp, containing millions or even billions of switches.



Instructions are stored in a memory. A *memory* is a circuit that can store 0s and 1s in each of a series of thousands of addressed locations, like a series of addressed mailboxes that each can store an envelope (the 0s and 1s). Instructions operate on data, which is also stored in memory locations as 0s and 1s.



Thus, a computer is basically a processor interacting with a memory, as depicted in the following example. In the example, a computer's processor executes program instructions stored in memory, also using the memory to store temporary results. The example program converts an hourly wage

(\$20/hr) into an annual salary by multiplying by 40 (hours/week) and then by 50 (weeks/year), outputting the final result to the screen.



The arrangement is akin to a chef (processor) who executes instructions of a recipe (program), each instruction modifying ingredients (data), with the recipe and ingredients kept on a nearby counter (memory).

Below are some sample types of instructions that a processor might be able to execute, where X, Y, Z, and num are each an integer.

ole 1.7.1: Sam	ple processor instructions.
Add X, #num, Y	Adds data in memory location X to the number <i>num</i> , storing result in location Y
Sub X, #num, Y	Subtracts <i>num</i> from data in location X, storing result in location Y
Mul X, #num, Y	Multiplies data in location X by <i>num</i> , storing result in location Y
Div X, #num, Y	Divides data in location X by <i>num</i> , storing result in location Y
Jmp Z	Tells the processor that the next instruction to execute is in memory location Z

For example, the instruction "Mul 97, #9, 98" would multiply the data in memory location 97 by the number 9, storing the result into memory location 98. So if the data in location 97 were 20, then the instruction would multiply 20 by 9, storing the result 180 into location 98. That instruction would actually be stored in memory as 0s and 1s, such as "011 1100001 001001 1100010" where 011 specifies a multiply instruction, and 1100001, 001001, and 1100010 represent 97, 9, and 98 (as described previously). The following animation illustrates the storage of instructions and data in memory for a program that computes $F = (9^*C)/5 + 32$, where C is memory location 97 and F is memory location 99.



The programmer-created sequence of instructions is called a *program*, *application*, or just *app*.

When powered on, the processor starts by executing the instruction at location 0, then location 1, then location 2, etc. The above program performs the calculation over and over again. If location 97 is connected to external switches and location 99 to external lights, then a computer user (like the women in the above picture) could set the switches to represent a particular Celsius number, and the computer would automatically output the Fahrenheit number using the lights.



Participation Activity 1.7.5: Computer basics.				
#	Question	Your answer		
	A bit can only have the value of 0 or 1.	True		
1		False		
0	Switches have gotten larger over the years.	True		
2		False		
0	A memory stores bits.	True		
3		False		
	The computer inside a modern smartphone would have been huge 30 years ago.	True		
4		False		
_	A processor executes instructions like, Add 200, #9, 201, represented as 0s and 1s.	True		
5		False		

In the 1940s, programmers originally wrote each instruction using 0s and 1s, such as "001 1100001 001001 1100010". Instructions represented as 0s and 1s are known as *machine instructions*, and a sequence of machine instructions together form an *executable program* (sometimes just called an *executable*). Because 0s and 1s are hard to comprehend, programmers soon created programs called *assemblers* to automatically translate human readable instructions, such as "Mul 97, #9, 98", known as *assembly* language instructions, into machine instructions. The assembler program thus helped programmers write more complex programs.

In the 1960s and 1970s, programmers created **high-level languages** to support programming using formulas or algorithms, so a programmer could write a formula like: F = (9 / 5) * C + 32. Early high-level languages included *FORTRAN* (for "Formula Translator") or *ALGOL* (for "Algorithmic Language") languages, which were more closely related to how humans thought than were machine or assembly instructions.

To support high-level languages, programmers created *compilers*, which are programs that automatically translate high-level language programs into executable programs.

```
Participation
                1.7.6: Program compilation and execution.
    Activity
Start
MainProgram()
begin
                                               High level
                                                program
  hourlyWage = GetInput("Enter wage: ");
                                             myfile.txt
  Print("Salary is: ");
  Print(hourlyWage * 40 * 50);
end
. . .
                                                Executable
011 1100001 001001 1100010
                                                 program
100 1100010 000101 1100010
                                               myfile.exe
001 1100010 100000 1100011
. . .
                                             Not human readable
                    V
> myfile.exe
Enter wage: 20
Salary is: 40000
>
```

Using the above approach, an executable can only run on a particular processor type (like an x86 processor); to run a program on multiple processor types, the programmer must have the compiler generate multiple executables. Some newer high-level languages like Java use an approach that allows the same executable to run on different processor types. The approach involves having the compiler generate an executable using machine instructions of a "virtual" processor, such an executable sometimes called **bytecode**. Then, the real processor runs a program, sometimes called a **virtual machine**, that executes the instructions in the bytecode. Such an approach may yield slower program execution, but has the advantage of portable executables.

Participation Activity 1.7.7: P	Programs.
Assembly language	Compiler Machine instruction Application
Drag and drop above item	Translates a high-level language program into low-level machine instructions.
	Another word for program.
	A series of 0s and 1s, stored in memory, that tells a processor to carry out a particular operation like a multiplication.
	Human-readable processor instructions; an assembler translates to machine instructions (0s and 1s).
Reset	

Note (mostly for instructors): Why introduce machine-level instructions in a high-level language book? Because a basic understanding of how a computer executes programs can help students master high-level language programming. The concept of sequential execution (one instruction at a time) can be clearly made with machine instructions. Even more importantly, the concept of each instruction operating on data in memory can be clearly demonstrated. Knowing these concepts can help students understand the idea of assignment (x = x + 1) as distinct from equality, why x = y; y = x does not perform a swap, what a pointer or variable address is, and much more.

Section 1.8 - Computer tour

The term *computer* has changed meaning over the years. The term originally referred to a person that performed computations by hand, akin to an accountant ("We need to hire a computer.") In the 1940s/1950s, the term began to refer to large machines like in the earlier photo. In the 1970s/1980s, the term expanded to also refer to smaller home/office computers known as personal computers or PCs ("personal" because the computer wasn't shared among multiple users like the large ones) and to portable/laptop computers. In the 2000s/2010s, the term may also cover other computing devices like pads, book readers, and smart phones. The term computer even refers to computing devices embedded inside other electronic devices such as medical equipment, automobiles, aircraft, consumer electronics, military systems, etc.

In the early days of computing, the physical equipment was prone to failures. As equipment became more stable and as programs became larger, the term "software" became popular to distinguish a computer's programs from the "hardware" on which they ran.

A computer typically consists of several components (see animation below):

- *Input/output devices*: A *screen* (or monitor) displays items to a user. The above examples displayed textual items, but today's computers display graphical items too. A *keyboard* allows a user to provide input to the computer, typically accompanied by a *mouse* for graphical displays. Keyboards and mice are increasingly being replaced by *touchscreens*. Other devices provide additional input and output means, such as microphones, speakers, printers, and USB interfaces. I/O devices are commonly called *peripherals*.
- **Storage**: A **disk** (aka hard drive) stores files and other data, such as program files, song/movie files, or office documents. Disks are *non-volatile*, meaning they maintain their contents even when powered off. They do so by orienting magnetic particles in a 0 or 1 position. The disk spins under a head that pulses electricity at just the right times to orient specific particles (you can sometimes hear the disk spin and the head clicking as the head moves). New *flash* storage devices store 0s and 1s in a non-volatile memory rather than disk, by tunneling electrons into special circuits on the memory's chip, and removing them with a "flash" of electricity that draws the electrons back out.
- *Memory*: *RAM* (random-access memory) temporarily holds data read from storage, and is designed such that any address can be accessed much faster than disk, in just a few clock ticks (see below) rather than hundreds of ticks. The "random access" term comes from being able to access any memory location quickly and in arbitrary order, without having to spin a disk to get a proper location under a head. RAM is costlier per bit than disk, due to RAM's higher speed. RAM chips typically appear on a printed-circuit board along with a processor chip. RAM is volatile, losing its contents when powered off. Memory size is typically listed in bits, or in bytes where a

byte is 8 bits. Common sizes involve megabytes (million bytes), gigabytes (billion bytes), or terabytes (trillion bytes).

- **Processor**: The **processor** runs the computer's programs, reading and executing instructions from memory, performing operations, and reading/writing data from/to memory. When powered on, the processor starts executing the program whose first instruction is (typically) at memory location 0. That program is commonly called the BIOS (basic input/output system), which sets up the computer's basic peripherals. The processor then begins executing a program called an *operating system (OS)*. The **operating system** allows a user to run other programs and which interfaces with the many other peripherals. Processors are also called *CPUs* (central processing unit) or *microprocessors* (a term introduced when processors began fitting on a single chip, the "micro" suggesting small). Because speed is so important, a processor may contain a small amount of RAM on its own chip, called **cache** memory, accessible in one clock tick rather than several, for maintaining a copy of the most-used instructions/data.
- **Clock**: A processor's instructions execute at a rate governed by the processor's **clock**, which ticks at a specific frequency. Processors have clocks that tick at rates such as 1 MHz (1 million ticks/second) for an inexpensive processor (\$1) like those found in a microwave oven or washing machine, to 1 GHz (1 billion ticks/second) for costlier (\$10-\$100) processors like those found in mobile phones and desktop computers. Executing about 1 instruction per clock tick, processors thus execute millions or billions of instructions per second.

Computers typically run multiple programs simultaneously, such as a web browser, an office application, a photo editing program, etc. The operating system actually runs a little of program A, then a little of program B, etc., switching between programs thousands of times a second.



After computers were first invented and occupied entire rooms, engineers created smaller switches called *transistors*, which in 1958 were integrated onto a single chip called an *integrated circuit* or IC. Engineers continued to find ways to make smaller transistors, leading to what is known as *Moore's Law*: The doubling of IC capacity roughly every 18 months, which continues today.^{ML} By 1971, Intel produced the first single-IC processor named the 4004, called a *microprocessor* ("micro" suggesting small), having 2300 transistors. New more-powerful microprocessors appeared every few years, and by 2012, a single IC had several *billion* transistors containing multiple processors (each called a *core*).

Participation Activity 1.8.2: F	Programs.
Moore's Law Clock	Operating system RAM Disk Cache
Drag and drop above item	Manages programs and interfaces with peripherals.
	Non-volatile storage with slower access.
	Volatile storage with faster access usually located off processor chip.
	Relatively-small volatile storage with fastest access located on processor chip.
	The frequency a processor executes instructions.
	The doubling of IC capacity roughly every 18 months.
Reset	

A side-note: A common way to make a PC faster is to add more RAM. A processor spends much of its time moving instructions/data between memory and storage, because not all of a program's instructions/data may fit in memory—akin to a chef that spends most of his/her time walking back and forth between a stove and pantry. Just as adding a larger table next to the stove allows more ingredients to be kept close by, a larger memory allows more instructions/data to be kept close to the processor. Moore's Law results in RAM being cheaper a few years after buying a PC, so adding RAM to a several-year-old PC can yield good speedups for little cost.

Exploring further:

- Video: Where's the disk/memory/processor in a desktop computer (20 sec).
- Link: What's inside a computer (HowStuffWorks.com).

- Video: How memory works (1:49)
- Video: Adding RAM (2:30)
- "How Microprocessors Work" from howstuffworks.com.

(*ML) Moore actually said every 2 years. And the actual trend has varied from 18 months. The key is that doubling occurs roughly every couple years, causing enormous improvements over time. Wikipedia: Moore's Law.

Section 1.9 - Language history

In 1978, Brian Kernighan and Dennis Ritchie at AT&T Bell Labs (which used computers extensively for automatic phone call routing) published a book describing a new high-level language with the simple name **C**, being named after another language called B (whose name came from a language called BCPL). C became the dominant programming language in the 1980s and 1990s.

In 1985, Bjarne Stroustrup published a book describing a C-based language called C++, adding constructs to support a style of programming known as object-oriented programming, along with other improvements. The unusual ++ part of the name comes from ++ being an operator in C that increases a number, so the name C++ suggests an increase or improvement over C.

In 1991, James Gosling and a team of engineers at Sun Microsystems (acquired by Oracle in 2010) began developing the *Java* language with the intent of creating a language whose executable could be ported to different processors, with the first public release in 1995.

The language had a C/C++ style and for portability reasons was designed to execute on a virtual machine. Java was initially intended to be run on consumer appliances like interactive TVs. Web browsers like Netscape Navigator began providing support for running Java programs within the browser, bringing much attention to the language. The Java language continues to evolve for the programming of traditional computers and consumer devices. Java should not be confused with JavaScript, which is an entirely different language used for developing web applications that was named similarly.

An December 2015 survey ranking language by their usage (lines of code written) yielded the following:

Language	Usage by percentage
Java	21%
С	17%
C++	6%
Python	5%
C#	4%
PHP	3%
Visual Basic .NET	2%
Javascript	2%
Perl	2.2%
Ruby	2%
Assembly language	1%



1.9.1: C/C++ history.

#	Question	Your answer
1	In what year was the first C book published?	
2	In what year was the first C++ book published?	

Ρ	Participation Activity 1.9.2: Java history.	
#	Question	Your answer
1	When was the first public release of Java?	

Section 1.10 - Problem solving

A chef may write a new recipe in English, but inventing a delicious new recipe involves more than just knowing English. Similarly, writing a good program is about much more than just knowing a programming language. Much of programming is about **problem solving**: Creating a methodical solution to a given task.

The following are real-life problem-solving situations encountered by one of this material's authors.

Example 1.10.1: Solving a (non-programming) problem: Matching socks.

A person stated a dislike for matching socks after doing laundry, indicating there were three kinds of socks. A friend suggested just putting the socks in a drawer, and finding a matching pair each morning. The person said that finding a matching pair could take forever: After pulling out a first sock, then pulling out a second, placing back, and repeating until the second sock matches the first, could go on for many times (5, 10, or more).

The friend provided a better solution approach: Pull out a first sock, then pull out a second, and repeat (without placing back) until a pair matches. In the worst case, the fourth sock will match one of the first three.

Participation
Activity

1.10.1: Matching socks solution approach.

Three sock types A, B, and C exist in a drawer.

#	Question	Your answer
4	If sock type A is pulled first, sock type B second, and sock type C third, the fourth sock type must match one of A, B, or	True
1	0.	False
	If socks are pulled one at a time and kept until a match is found, at least four pulls are necessary.	True
2		False
3	If socks are pulled two at a time and put back if not matching, and the process repeated until the two pulled socks match, the maximum number of pulls is 4.	True
		False

Participation Activity

1.10.2: Greeting people problem.

An organizer of a 64-person meeting wants to start by having every person individually greet every other person for 30 seconds. Indicate whether the proposed solution achieves the goal, without using excessive time. Before answering, think of a possible solution approach for this seemingly simple problem.

	#	Question	Your answer
F p	Form an inner circle of 32, and an outer circle of 32, with people matched up. Every 30 seconds, have the outer circle shift loft one position	Yes	
	I		No
	2	Pair everyone randomly. Every 30 seconds, tell everyone to find someone new to greet. Do this 63 times.	Yes
	2		No
	Q	Have everyone form a line. Then have everyone greet the person behind them.	Yes
	0		No
		Have everyone form a line. Have the first person greet the other 63 people for 30 seconds each. Then have the second person greet each other person for 30 seconds each (skipping anyone already met). And so on.	Yes
	4		No
		Form two lines of 32 each, with attendees matched up. Every 30 seconds, have one line shift left one position (with the person on the left end wrapping to right). Once the person that started on the left is back on the left, then have each line split into two matched lines, and repeat until each line has just 1 person.	Yes
ξ	5		No

Example 1.10.2: Solving a (non-programming) problem: Sorting (true story).

1000 name tags were printed and sorted by first name into a stack. A person wishes to sort the tags by last name. Two approaches to solving the problem are:

- Solution approach 1: For each tag, insert that tag into the proper location in a new last-name sorted stack.
- Solution approach 2: For each tag, place the tag into one of 26 sub-stacks, one for last names starting with A, one for B, etc. Then, for each sub-stack's tags (like the A stack), insert that tag into the proper location of a last-name sorted stack for that letter. Finally combine the stacks in order (A's stack on top, then B's stack, etc.)

Solution approach 1 will be very hard; finding the correct insertion location in the new sorted stack will take time once that stack has about 100 or more items. Solution approach 2 is faster, because initially dividing into the 26 stacks is easy, and then each stack is relatively small so easier to do the insertions.

In fact, sorting is a common problem in programming, and the above sorting approach is similar to a well-known sorting approach (radix sort).



1.10.3: Sorting name tags.

1000 name tags are to be sorted by last name by first placing tags into 26 unsorted sub-stacks (for A's, B's, etc.), then sorting each sub-stack.

#	Question	Your answer
	If last names are equally distributed among the alphabet, what is the largest number of name tags in any one sub- stack?	1
1		39
		1000
	Suppose the time to place an item into one of the 26 sub- stacks is 1 second. How many seconds is required to place all 1000 name tags onto a sub-stack?	26 sec
2		1000 sec
		26000 sec
	When sorting each sub-stack, suppose the time to insert a name tag into the appropriate location of a sorted N-item sub-stack is N * 0.1 sec. If the largest sub-stack is 50 tags, what is the longest time to insert a tag?	5 sec
3		50 sec
4	Suppose the time to insert a name tag into an N-item stack is N * 0.1 sec. How many seconds are required to insert a name tag into the appropriate location of a 500 item stack?	5 sec
4		50 sec

A programmer usually should carefully create a solution approach *before* writing a program. Like English being used to describe a recipe, the programming language is just a description of a solution approach to a problem; creating a good solution should be done first.

Section 1.11 - Java example: Salary calculation program

This material has a series of sections providing increasingly larger program examples. The examples apply concepts from earlier sections. Each example is in a web-based programming environment so that code may be executed. Each example also suggests modifications, to encourage further understanding of the example. Commonly, the "solution" to those modifications can be found in the series' next example.

This section contains a very basic example for starters; the examples increase in size and complexity in later sections.

```
Participation
                   1.11.1: Modify salary calculation.
       Activity
The following program calculates yearly and monthly salary given an hourly wage. The program
assumes a work-hours-per-week of 40 and work-weeks-per-year of 50.
      1. Insert the correct number in the code below to print a monthly salary. Then run the
         program.
   Reset
    1 public class Salary {
    2
    3
         public static void main (String [] args) {
    4
            int hourlyWage = 20;
    5
            System.out.print("Annual salary is: ");
    6
    7
            System.out.println(hourlyWage * 40 * 50);
    8
    9
            System.out.print("Monthly salary is: ");
            System.out.println((hourlyWage * 40 * 50) / 1);
   10
            // FIXME: The above is wrong. Change the 1 so the statement prints monthly sal
   11
   12
   13
            return;
   14
         }
   15 }
   16
   Run
```

Section 1.12 - Java example: Married-couple names

Participation Activity

1.12.1: Married-couple names.

Pat Smith and Kelly Jones are engaged. What are possible last name combinations for the married couple (listing Pat first)?

- 1. Run the program below to see three possible married-couple names.
- 2. Extend the program to print the two hyphenated last name options (Smith-Jones, and Jones-Smith). Run the program again.

Reset

```
1 import java.util.Scanner;
   2
   3 public class ShowMarriedNames {
   4
   5
        public static void main(String[] args) {
            Scanner scnr = new Scanner(System.in);
   6
            String firstName1 = "";
   7
   8
            String lastName1 = "";
   9
            String firstName2 = "";
            String lastName2 = "";
  10
  11
            System.out.println("What is the first person's first name?");
  12
  13
            firstName1 = scnr.nextLine();
            System.out.println("What is the first person's last name?");
  14
  15
            lastName1 = scnr.nextLine();
  16
            System.out.println("What is the second person's first name?");
  17
            firstName2 = scnr.nextLine();
  18
  19
            Svstem.out.println("What is the second person's last name?"):
Pat
Smith
Kelly
  Run
```

```
Participation
                   1.12.2: Married-couple names (solution).
       Activity
A solution to the above problem follows:
   Reset
    1 import java.util.Scanner;
    2
    3 public class ShowMarriedNames_Solution {
    4
    5
         public static void main(String[] args) {
    6
            Scanner scnr = new Scanner(System.in);
    7
            String firstName1 = "";
            String lastName1 = "";
    8
            String firstName2 = "";
    9
   10
            String lastName2 = "";
   11
   12
            System.out.println("What is the first person's first name?");
   13
            firstName1 = scnr.nextLine();
            System.out.println("What is the first person's last name?");
   14
   15
            lastName1 = scnr.nextLine();
   16
   17
            System.out.println("What is the second person's first name?");
   18
            firstName2 = scnr.nextLine();
   19
            System.out.println("What is the second person's last name?"):
Pat
Smith
Kelly
   Run
```