Variables and Expressions

We often think of programs as tools to transform data, starting with some input, doing calculations, and resulting in some output. So the first thing we need is a way to store our data.

A data structure is a tool in a program for storing data. The most basic version of this is a variable; later we will see more complex versions, for instance arrays for storing lists of data, or objects for storing collections of values and actions that represent a component of a situation.

Data Values

Remember that data is anything stored for a program other than the instructions: starting values, input from users or files, intermediate values, and results are all data. Some may be permanent characteristics or attributes in the world of the program, others may be a temporary current state that changes as the program runs.

All data is stored as binary patterns and the same binary pattern might sometimes be interpreted as a number, a text symbol, a color, a sound, a coordinate, et cetera.

In a program we sometimes work with *literal values*, where we just write a price or name into a program. A program that only uses literals is very limited, it only works for those specific values and if we want it to do something different we have to re-write the whole program – if we used the literal 49.99 throughout our program whenever we needed to refer to the price of something, we have to update every mention of that in the program when the price goes up to 62.50.

At the hardware level, if each piece of data is stored on a row in Main Memory, we can use the Memory Address of the row to refer to the data instead of using the literal value. If the program is written this way, then we only have to change the value of that one location to change the way the whole program works.

This, however, requires thinking about hardware details and remembering a memory address, which is a number. Almost all high level languages should provide us with named variables instead.

Variables

Instead of trying to remember which line of MM we put which piece of data on, most high-level languages provide named variables, allowing us to associate a name with a memory location and use the name in the program. This will allow us to end up writing something more like

```
total = bonus + base_commission
```

which is certainly easier to read.

Pseudocode: Input, Output, and Comments

Before we move any farther, we need to add some basic tools to our pseudocode.

First we need a way to include notes to help understand the code. A comment is a note in the code that is there for programmers to read, but does not affect the way the program runs. In our pseudocode, we will use // to mean the rest of the line past the slashes is a comment, and /* and */ to mark off everything between the asterisks as comments, even if it crosses multiple lines:

```
// calculate the total for this sale
total = bonus + base_commission // using original bonus
```

In this example the parts highlighted in grey are comments and would be ignored by a compiler or interpreter. In some cases we might use comments to temporarily remove lines of a program, for instance if we want to see if a certain line is causing a problem, or we want to remove something that we might need again someday. This is called "commenting out" those lines:

```
//bonus = old_bonus * 10
total = bonus + base_commission
/*
print bonus
print base_comission
*/
print total // wait, we don't know "print" yet…
```

In this example, again the grey parts are currently comments and would not be run as part of the program

We also need ways to get data into and out of our programs – input and output. For output we will use the keyword `print` for printing.

In general, when getting input from a user, we should tell the user what we want them to give, which is called prompting them, we will use the keyword `prompt`, and then the words we want to say to the user in double quotes. We will also use + for concatenation if we need to put things next to each other in the same prompt or output

```
// print "what is the bonus" and put whatever
// value the user types into the bonus variable
bonus = prompt("what is the bonus?")
base_comission = prompt("commission?)
total = bonus + base_commission
```

```
// print the words "Total is " followed by
// the value in the total variable
print "Total is " + total
```

Note that when writing pseudocode to plan out the steps of a process, we might be less specific: we might not explicitly say how we want to prompt the user or how we want to format the output. We also sometimes need to specify other kinds of input and output. The point of pseudocode is to give a simple format for writing plans for code that all programmers can understand. See if you understand the following (if not, ask me!).

```
bonus = userInput
base_comission = databaseInput
total = bonus + base_commission
fileOutput("ledger.txt") = total
```

In general, for this course, we will tend to use the more specific version that looks more like real code, since we are mostly focusing on learning basic concepts.

It is also very common in pseudocode to mark the beginning and ending of a block of code using start/stop or begin/end. I will usually use open curly brace/ close curly brace, but if I am writing a short code example I will very often leave these off, which you may also do (until we start having code that has multiple blocks, when we will start to need this.

When we mark out blocks of code, the lines in each block should be indented one tab farther

```
{// a properly formatted block of code
     bonus = userInput
     base_comission = databaseInput
     total = bonus + base_commission
     fileOutput("ledger.txt") = total
}
```

Variable Declaration and Assignment

Declaration is the term for introducing a new variable into our program. When this happens, we are telling the language that we want to have space for a variable with that name, and it will take care of reserving that row (or rows) in MM and associating the name we choose with that MM address. The language will handle knowing what the MM address is, so we don't have to.

Some languages do allow the programmer to access the actual MM address associated with a variable as a number and even do operations on that value while other languages hide this information. Having that access means we have more control over the specifics of what our

program is doing at the hardware level, but it also means we have more power to do something that will mess things up[1].

Important Principle: **When a language designer chooses what tools to include in a programming language, there is a tradeoff between power and ease of use.** A language that provides more firepower also makes it easier to blow off your own foot.

One way in our pseudocode to declare a variable is just to put the keyword var in front of the name we want to use for the variable. So by saying

```
var bonus
```

we have told the language that from now on in our program, bonus will be the name of a variable.

Assignment is giving a variable a value. We use a single equals sign to indicate assignment, and assignment will always be from right to left: the value on the right is being put into the variable on the left. So if we say

```
bonus = 70
```

we mean that the value 70 is being stored in the variable called bonus. Under the hood, this means that the language looks up what row of main memory the bonus variable represents, and puts the binary pattern for 70 into that row in MM.

If we know the value of the bonus when we create the variable, we could do declaration and assignment all in one line:

```
var bonus = 70
```

A variable lets us associate a name with a specific address in memory. Whenever we use that name in our code, it will (when the program is translated to run later) stand for the value in that memory location.

[1] Old-school programmers using languages with this kind of access sometimes did tricks like having one program leave data at a certain memory address and having another program simply read from that memory address, as a way of very efficiently moving data from one program to another, instead of having the first program save to a file on the hard drive and the second program read from that file. There are situations where resources such as space on a device or time to complete a task are very limited, where such approaches might really help. However, this relies on knowing exactly which MM addresses a program would be using every time, that no other programs would use or change those addresses. It also means that a small mistake might have one program overwrite the wrong data, or even an instruction in another program.

Variable Names

It is very important to get in the habit of giving your variables names that will help you read and write your code.

What does this code do? Is it correct?

```
a = b - ((c * b) + d + e)
```

Single letter variable names make sense if you are doing abstract math, or need a temporary variable to hold an intermediate value briefly while doing calculations. But if you think you are saving time by making them short to type, you probably aren't considering how much extra time you're going to spend trying to remember what b represents.

The problem isn't just the length, here's an even worse version:

```
var1 = var2 - ((var3 * var2) + var4 + var5)
```

Here is a more readable version. Now it is more clear what is going on, but if we're checking our math, we might still be unsure whether we're using the variables correctly

```
home = pay - ((tax * pay) + insure + retire)
```

And now a version that has longer variables, and makes it clear that the tax variable is a percentage, so it does need to be multiplied, while the insurance is a flat fee and does not.

```
take_home = paycheck - ((tax_percent * paycheck) +
     insurance_fee + retirement_contribution)
```

You will develop your own style for variable naming that is the best compromise between names that have enough information to make it clear what is going on and names that are short enough to easily read and type. Adding comments can help when variable names are bad, but it would always be better to fix the variable names so comments can explain more complicated issues.

Some new programmers start off too sloppy about variable names, using names like xyz and fklsjhu, which means their code is impossible to read once they've forgotten what those represented. Other new programmers freeze up every time they have to create a variable name and overthink it, which slows down their coding. For the moment, try for one-word names, and stick two together if you get stuck.

Programming languages do not analyze your variable names for meaning. If you create a variable called tax_percent but you store the overall tax amount in it, the language will not give you an error.

Each programming language will have its own restrictions on how variables can be named. Variable names cannot be the same as a keyword that is part of the language, so we could not have a variable whose name is var, because we are already using var to indicate declaration, similarly we can't use something that's already a literal value, so we couldn't use 70 as the name of the variable. In most languages, variables cannot have spaces in their names, so if we want to use multiple words in the name we might put a dash or underscore between the parts, or else put the parts next to each other but capitalize the later parts, called *camel casing*. some examples

```
var number_of_puppies
var cats-per-square-ironing-board
var priceOfBeans
```

For our pseudocode, we will allow variable names to contain only: upper-case and lower-case letters, digits, underscores, and dashes, and nothing else. That way, if you see anything else, you know it cannot be part of a variable name. Most variables should use mostly lower-case letters and generally start with a lower-case.

Assignment

Assignment means putting a value into a variable. That value could be a literal value like 7 or "cat", or a value currently stored in another variable, or the result of an operation like adding.

```
var cats = 17
var dogs = cats // since cats is 17, dogs is now also 17
var animals = cats + dogs // animals is the result of
     adding, so 34
```

In our pseudocode, as in most languages, assignment always goes from right to left. We refer to the Left Hand Side (LHS) and Right Hand Side (RHS) of an assignment. The LHS must be a variable, it doesn't make sense to say

```
6 = how_many_pies
```

because that would mean we are trying to change the value of 6!

The RHS of an assignment can be any expression that, when fully evaluated, comes down to a single value.

Setting equality through assignment makes a one-time copy of a value, it does not link the variables forever. If I set one variable equal to another, and then change the first one, that does not go back and change the value of the second:

```
var cats = 17
var dogs = cats // copies value so dogs is 17
```

```
cats = 18 // cats has changed, but dogs is still 17
```

Variable Types

Remember that we re-use the same binary patterns for many purposes, since there are a limited number of possible combinations of ons and offs of a given length. So the same binary pattern might be used in one situation to store a whole number, in another situation as a fractional number, in another place as a letter of the alphabet, etc. In some cases we might need longer binary patterns to store special values, so we'd combine patterns on multiple rows of MM. So it would be helpful to have a way to tell the program how to interpret a pattern and how many rows in MM to use.

In most languages, variables have type. Each variable can only store one type of information and this tells how to interpret the binary pattern stored at that variable's location. The most common types of variable are integers (whole numbers) doubles (real numbers, which can have a decimal part), characters (individual symbols for text, enclosed in single quotes) and Booleans (true or false). For our pseudocode, we will also include strings, which are sequences of characters, enclosed in double quotes.

keyword	type	description	example values
int	integer	whole number	0, 1, 2, -3779
double	real	can have decimal place	0.001, -87.02
char	text symbol	letters, numbers, punctuation	'a', 'Z', '1', '.', ' '[2]
boolean	boolean	true/false yes/no	true, false
string	text	words, strings of text symbols	"a", "ZZ", "hi bye", " "

Instead of using var, to declare a typed variable, we put the type before the variable name, and then we don't say the type again thereafter.

```
int num_puppies = 3
double cost_of_puppies = 59.99
char puppy_letter = "p"
boolean can_puppy_fly = true
string puppy_name = "Mr. Fluffenstuff"
```

[2] Technically, these curved quotes 'x' and "x" are different from straight quotes 'x' and "x" (which are actually foot and inch markings), they are stored in different binary patterns. If you type in a word processor like Word, it will often automatically curl your quotes to make the text more attractive and readable. But in general, programming languages only use straight quotes. If you copy curly quotes into real code you will probably get a syntax error, but for our pseudocode it does not matter.

Having typed variables makes it easier for the language to know how to interpret binary patterns, but more importantly, it makes writing code easier by limiting possible values we can give to a variable. We said above that num_puppies is an integer; we want to have a whole number of puppies, having 0.6 of a puppy is probably not good! But if later I forget what I'm doing, and try to say

num_puppies = 0.6

I will get a syntax error.

Sometimes people seeing types for the first time feel like they are making programming harder. Now there are more errors you can get! But actually types are like guard-rails that take some of the burden of remembering variable types off of the programmer; you choose the type for a variable when you declare it, and after that the language will keep track of what values are allowed and not allowed. The error you might get isn't a new burden you have to take on, it is a reminder that you already chose what kinds of values were appropriate for this variable.

If you now need a place to store some other type, the answer isn't to try to stuff it into an existing variable. Just make a new variable of the type you need! There may be cases where you are programming in a situation where you have to limit the number of variables you declare, but for the most part, new programmers tend to be too stingy about declaring variables, so for right now, think of variables as cheap tools; you can make as many as you want.

To summarize, in our pseudocode: Declaring a variable in our pseudocode requires putting the type of the variable first, then the variable name. We will use a single equals sign for assignment, and assignment can be done on the same line as declaration. Once a variable is declared, we do not put the type in front of the variable again. If doing an example that requires variables not to be strongly typed, we will use var in place of the type, but this will not be used for most code.

Weakly Typed, Dynamically Typed, and Casting

The above explains how type works in most languages, which are *strongly typed*. Other languages, however, may take different approaches.

In a *dynamically typed* language, it *is* syntactically legal to assign a variable a value that does not match the variable's type. When the program runs, the value will be changed into the correct type for the variable. The process of changing a value from one type to another is called *casting*. Casting will at least always involve changing the binary pattern used to store the value into the system used for the variable's type. Weakly typed languages do a lot of *implicit casting* – the casting happens automatically without the programmer having to deliberately write code to make it happen.

For instance, in a dynamically typed language, we would be allowed to say

```
string str = 10
```

But what value would str end up with when the program runs? The designers of the language would have to have determined how integer values are converted into strings. It could be that it simply becomes the string a human would use to type that integer value, "10". It could be that the binary to store that number would be reinterpreted as a character to be stored in the string; the number 10 can be stored as 00001010 in binary, and this happens to be the same as one common binary pattern for storing "A".

Strongly typed languages do usually include some implicit casting. If we are able to print numeric values, for instance, those values will probably be cast to string types, because humans don't usually like to read binary patterns. So if we say

```
print 10
```

We would expect the number 10 to be implicitly cast to the string "10" so it could be printed in a way users understand, instead of seeing the binary pattern 00001010 for the value 10 printed.

The other most common implicit cast is from int to double. If we say

```
double d = 17
```

It may not even register that we are using different types. For humans, 17 and 17.0 are obviously equivalent. But it is unlikely that under the hood, the binary pattern used to store 17 in the system for integers and the binary pattern used to store 17.0 in the system for real numbers would be the same binary pattern. Casting has to happen to convert from one to the other. Most strongly typed languages would do implicit casting here so that d1 ends up with the value 17.0.

A strongly typed language would not let us go in the other direction, however. If we try to put a real number into an int like this

```
int num = 8.9
```

This would be a syntax error, because there is a decimal part to 8.9 that could not be preserved if we store it under the integer system. The strongly typed system put s up guardrails to avoid losing some of our data.

Strongly typed languages do often provide tools for deliberately changing the type of a value by *explicit casting.* In our psuedocode, we can explicitly cast by putting the type we want inside parentheses in front of the value whose type we want to change.

```
int num = (int)8.9 // become the integer value 8
```

Note that casting is not a mathematical operation, although in this case it does the same as taking the floor function of the number. Rounding is a mathematical operation that takes a real input and results in an integer output (in this case 8.9 would round to 9) , but that isn't what we are doing here, we are just throwing away the part of the data that won't fit into our representation system.

Weakly typed languages could be said to respect the types of their variables and just do constant implicit casting. *Dynamically typed* languages change the type of a variable on the fly whenever it is assigned. So in a dynamically typed language, if we said

```
catval = "Fluffy"
print catval
catval = 3
print catval
catval = 8.2
print catval
```

Then the same variable would first be a string, then an int, then a double. Note that this is going to take more work for the language – when we print a value for catval, we have to go retrieve the binary pattern stored in the catval location in MM, and this value has to be interpreted appropriately, which means the language will have to keep track of whether it should evaluate the variable under the system for strings or for ints or for real numbers. Never having to think about the type of a variable can be very convenient and for some kinds of tasks it frees us up from thinking about all those implementation details like what format we are using for our binary patterns. But it does take off the guardrails that would stop us from using inappropriate values and means that it becomes our responsibility to check types ourselves.

Initialization

The first time a variable is assigned a value is called initialization. But what happens if we try to use a variable's value before it is initialized?

```
int cats = 17
int dogs
int animals = dogs + cats // what value???
```

This will depend on the language you are working in. In some languages, this is simply a syntax error. For our pseudocode, this is what we will assume.

In some languages, all variable types have default values, with numbers generally defaulting to 0, Booleans to false, and other types to the special value `null`.

In some languages, however, the value of uninitialized variable is whatever binary pattern was sitting in the row in MM that the variable was allocated when it was declared. This "garbage value" could be a value from some other program that was using this part of memory earlier. It is almost never a good or useful value, so in these languages it is very important to remember to initialize your variables!

Constants

Sometimes we have a value in a program that will not change, so it is not "vary-able" but it is still a good idea to put it in a variable, since this means we don't have to remember the literal value and can use a name instead, and also this means that if this value ever is to change, we only have to change the line with the assignment, not everywhere the value appears in the program.

We call these *constants* and use the keyword const for them. We often give them all-caps names so they stand out.

```
const double BONUS = 8.99
BONUS = 7.88 // error, cannot change
base_comission = databaseInput
total = BONUS + base_commission
fileOutput = total
```

Expressions

For algebraic operations, we will use

+ for addition, - for subtraction, * for multiplication, and / for division. Remember that multiplication and division have high precedence and addition and subtraction have low precedence. Within a level of precedence we work left to right, and we can adjust order of operations by adding parentheses.

We will also use + for concatenation for strings and when printing, and assume it has the same precedence as for addition.

Later we will see other operators, for example relational operators such as > and < and logical operators like && and ||. In our pseudocode, operators generally are infix (remember this means they go between the operands).

Not all operators are defined on all types in all languages. In most languages it doesn't make sense to multiply a string or divide a boolean.

Expressions are evaluated by first replacing any variables by their values, and then working through all operators, in order of precedence (remember to go left to right at the same precedence level).

Changing values

Variables often are changed over the course of a program and most of the time, a variable's new value will be related to its old value, so we will often see statements where the same variable is on both sides of an assignment.

```
double sale = 8.00
double coupon = 1.50
sale = sale – coupon
print "Final amount: $" + sale
```

In the bold line, sale is on both sides of the equals sign. ***This does not mean you have to "solve" the equation!*** Remember that the RHS and LHS of an assignment are treated differently. We find the value of the RHS, and then change the variable on the LHS to hold that value. Also remember that when we evaluate an expression, we replace any variables by their values. So the bold line would be evaluated in the following steps

```
1. sale = sale – coupon
2. sale = 9.00 – 1.50 // replace variables by values
3. sale = 7.50 // do the operation
```

So now sale has a new value, which was based on the previous value. In speech, we would say that we "decreased sale by coupon" or just "subtracted coupon from sale." When we describe it this way, new programmers sometimes think they should write

```
double sale = 8.00
double coupon = 1.50
sale – coupon
```

But that doesn't work! Remember that we should never be doing a calculation without storing the result. We need to put the result of that subtraction back into a variable, and in this case, it should go back into sale, replacing the previous value.

Make sure you understand why the following descriptions in comments would be written as the code below each:

```
//increase price by 5
price = price + 5

// halve price
price = price / 2
```

```
// triple price
price = price * 3

//decrease price by rebate
price = price - rebate
```

Increasing by 1 is also called incrementing, and decreasing by 1 is called decrementing. These operations happen so often that we have special operators to write them more briefly

```
//increment price
price = price + 1
// increment operator
price++  // means same thing

//decrement price
price = price - 1
// decrement operator
price--  // means same thing
```