STATS 507
Data Analysis in Python

Lecture 2: Conditionals,
Recursion, Iteration and Strings

Boolean Expressions

Boolean expressions evaluate the truth/falsity of a statement

Python supplies a special Boolean type, `bool`

variable of type `bool` can be either `True` or `False`

```
1 type(True)
```

```
bool
```

```
1 type(False)
```

```
bool
```

Boolean Expressions

Comparison operators available in Python:

```
1 x == y # x is equal to y
2 x != y # x is not equal to y
3 x > y # x is strictly greater than y
4 x < y # x is strictly less than y
5 x >= y # x is greater than or equal to y
6 x <= y # x is less than or equal to y
```

Expressions involving comparison operators evaluate to a Boolean.

```
1 x = 10
2 y = 20
3 x == y
```

```
False
```

```
1 x != y
```

```
True
```

```
1 x < x
```

```
False
```

```
1 x <= x
```

```
True
```

Note: In true Pythonic style, one can compare many types, not just numbers. Most obviously, strings can be compared, with ordering given alphabetically.

Boolean Expressions

Can combine Boolean expressions into larger expressions via **logical operators**

In Python: `and`, `or` and `not`

```
1 x = 10
2 x < 20 and x > 0
```

```
True
```

```
1 x > 100 and x > 0
```

```
False
```

```
1 x > 100 or x > 0
```

```
True
```

```
1 not x > 0
```

```
False
```

Note: technically, any nonzero number or any nonempty string will evaluate to `True`, but you should avoid comparing anything that isn't Boolean.

```
1 1 and x > 0
```

```
True
```

```
1 0 and x > 0
```

```
0
```

```
1 'cat' and x > 0
```

```
True
```

```
1 '' and x > 0
```

```
''
```

Boolean Expressions: Example

Let's see Boolean expressions in action

```
1  def is_even(n):
2      # Returns a boolean.
3      # Returns True if and only if
4      # n is an even number.
5      return n % 2 == 0
```

Reminder: `x % y` returns the remainder when `x` is divided by `y`.

Note: in practice, we would want to include some extra code to check that `n` is actually a number, and to "fail gracefully" if it isn't, e.g., by throwing an error with a useful error message. More about this in future lectures.

```
1  is_even(0)
```

```
True
```

```
1  is_even(1)
```

```
False
```

```
1  is_even(8675309)
```

```
False
```

```
1  is_even(-3)
```

```
False
```

```
1  is_even(12)
```

```
True
```

Conditional Expressions

Sometimes we want to do different things depending on certain conditions

```
1 x = 10
2 if x > 0:
3     print 'x is bigger than 0'
4 if x > 1:
5     print 'x is bigger than 1'
6 if x > 100:
7     print 'x is bigger than 100'
8 if x < 100:
9     print 'x is less than 100'
```

```
x is bigger than 0
x is bigger than 1
x is less than 100
```

Conditional Expressions

Sometimes we want to do different things depending on certain conditions

```
1 x = 10
2 if x > 0:
3     print 'x is bigger than 0'
4 if x > 1:
5     print 'x is bigger than 1'
6 if x > 100:
7     print 'x is bigger than 100'
8 if x < 100:
9     print 'x is less than 100'
```

```
x is bigger than 0
x is bigger than 1
x is less than 100
```

This is an **if-statement**.

Conditional Expressions

Sometimes we want to do different things depending on certain conditions

```
1 x = 10
2 if x > 0:
3     print 'x is bigger than 0'
4 if x > 1:
5     print 'x is bigger than 1'
6 if x > 100:
7     print 'x is bigger than 100'
8 if x < 100:
9     print 'x is less than 100'
```

```
x is bigger than 0
x is bigger than 1
x is less than 100
```

This Boolean expression is called the **test condition**, or just the **condition**.

Conditional Expressions

Sometimes we want to do different things depending on certain conditions

```
1 x = 10
2 if x > 0:
3     print 'x is bigger than 0'
4 if x > 1:
5     print 'x is bigger than 1'
6 if x > 100:
7     print 'x is bigger than 100'
8 if x < 100:
9     print 'x is less than 100'
```

```
x is bigger than 0
x is bigger than 1
x is less than 100
```

If the condition evaluates to `True`, then Python runs the code in the body of the if-statement.

Conditional Expressions

Sometimes we want to do different things depending on certain conditions

```
1 x = 10
2 if x > 0:
3     print 'x is bigger than 0'
4 if x > 1:
5     print 'x is bigger than 1'
6 if x > 100:
7     print 'x is bigger than 100'
8 if x < 100:
9     print 'x is less than 100'
```

```
x is bigger than 0
x is bigger than 1
x is less than 100
```

If the condition evaluates to `False`, then Python skips the body and continues running code starting at the end of the if-statement.

Conditional Expressions

Sometimes we want to do different things depending on certain conditions

```
1 x = 10
2 if x > 0:
3     print 'x is bigger than 0'
4 if x > 1:
5     print 'x is bigger than 1'
6 if x > 100:
7     print 'x is bigger than 100'
8 if x < 100:
9     print 'x is less than 100'
```

```
x is bigger than 0
x is bigger than 1
x is less than 100
```

Note: the body of a conditional statement can have any number of lines in it, but it must have at least one line. To do nothing, use the `pass` keyword.

```
1 y = 20
2 if y > 0:
3     pass # TODO: handle positive numbers!
4 if y < 100:
5     print 'y is less than 100'
```

```
y is less than 100
```

Conditional Expressions

More complicated logic can be handled with **chained conditionals**

```
1 def pos_neg_or_zero(x):
2     if x < 0:
3         print 'That is negative'
4     elif x == 0:
5         print 'That is zero.'
6     else:
7         print 'That is positive'
8 pos_neg_or_zero(1)
```

```
That is positive
```

```
1 pos_neg_or_zero(0)
2 pos_neg_or_zero(-100)
3 pos_neg_or_zero(20)
```

```
That is zero.
That is negative
That is positive
```

Conditional Expressions

More complicated logic can be handled with **chained conditionals**

```
1 def pos_neg_or_zero(x):
2     if x < 0:
3         print 'That is negative'
4     elif x == 0:
5         print 'That is zero.'
6     else:
7         print 'That is positive'
8 pos_neg_or_zero(1)
```

```
That is positive
```


```
1 pos_neg_or_zero(0)
2 pos_neg_or_zero(-100)
3 pos_neg_or_zero(20)
```

```
That is zero.
That is negative
That is positive
```

Conditional Expressions

More complicated logic can be handled with **chained conditionals**

```
1 def pos_neg_or_zero(x):
2     if x < 0:
3         print 'That is negative'
4     elif x == 0:
5         print 'That is zero.'
6     else:
7         print 'That is positive'
8 pos_neg_or_zero(1)
```

```
That is positive
```

If this expression evaluates to `True`...

```
1 pos_neg_or_zero(0)
2 pos_neg_or_zero(-100)
3 pos_neg_or_zero(20)
```

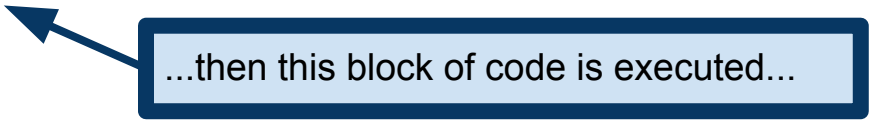
```
That is zero.
That is negative
That is positive
```

Conditional Expressions

More complicated logic can be handled with **chained conditionals**

```
1 def pos_neg_or_zero(x):
2     if x < 0:
3         print 'That is negative'
4     elif x == 0:
5         print 'That is zero.'
6     else:
7         print 'That is positive'
8 pos_neg_or_zero(1)
```

```
That is positive
```




```
1 pos_neg_or_zero(0)
2 pos_neg_or_zero(-100)
3 pos_neg_or_zero(20)
```

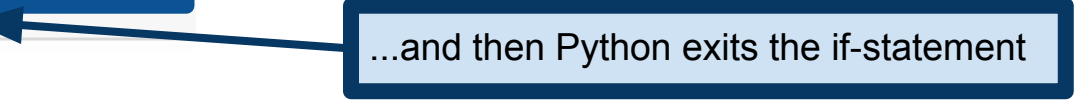
```
That is zero.
That is negative
That is positive
```

Conditional Expressions

More complicated logic can be handled with **chained conditionals**

```
1 def pos_neg_or_zero(x):
2     if x < 0:
3         print 'That is negative'
4     elif x == 0:
5         print 'That is zero.'
6     else:
7         print 'That is positive'
8 pos_neg_or_zero(1)
```

```
That is positive
```




```
1 pos_neg_or_zero(0)
2 pos_neg_or_zero(-100)
3 pos_neg_or_zero(20)
```

```
That is zero.
That is negative
That is positive
```

Conditional Expressions

More complicated logic can be handled with **chained conditionals**

```
1 def pos_neg_or_zero(x):
2     if x < 0:
3         print 'That is negative'
4     elif x == 0:
5         print 'That is zero.'
6     else:
7         print 'That is positive'
8 pos_neg_or_zero(1)
```

```
That is positive
```

If this expression evaluates to `False`...

```
1 pos_neg_or_zero(0)
2 pos_neg_or_zero(-100)
3 pos_neg_or_zero(20)
```

```
That is zero.
That is negative
That is positive
```

Conditional Expressions

More complicated logic can be handled with **chained conditionals**

```
1 def pos_neg_or_zero(x):
2     if x < 0:
3         print 'That is negative'
4     elif x == 0:
5         print 'That is zero.'
6     else:
7         print 'That is positive'
8 pos_neg_or_zero(1)
```

```
That is positive
```

```
1 pos_neg_or_zero(0)
2 pos_neg_or_zero(-100)
3 pos_neg_or_zero(20)
```

```
That is zero.
That is negative
That is positive
```

Note: `elif` is short for **else if**.

...then we go to the condition. If this condition fails, we go to the next condition, etc.

Conditional Expressions

More complicated logic can be handled with **chained conditionals**

```
1 def pos_neg_or_zero(x):
2     if x < 0:
3         print 'That is negative'
4     elif x == 0:
5         print 'That is zero.'
6     else:
7         print 'That is positive'
8 pos_neg_or_zero(1)
```

```
That is positive
```

If all the other tests fail, we execute the block in the `else` part of the statement.

```
1 pos_neg_or_zero(0)
2 pos_neg_or_zero(-100)
3 pos_neg_or_zero(20)
```

```
That is zero.
That is negative
That is positive
```

Conditional Expressions

Conditionals can also be nested

```
if x==y:
    print 'x is equal to y'
else:
    if x > y:
        print 'x is greater than y'
    else:
        print 'y is greater than x'
```

This if-statement...

Conditional Expressions

Conditionals can also be nested

```
if x==y:
    print 'x is equal to y'
else:
    if x > y:
        print 'x is greater than y'
    else:
        print 'y is greater than x'
```

This if-statement...

...contains another if-statement.

Conditional Expressions

Often, a nested conditional can be simplified

When this is possible, I recommend it for the sake of your sanity, because debugging complicated nested conditionals is tricky!

These two if-statements are equivalent, in that they do the same thing!

```
1 if x > 0:
2     if x < 10:
3         print 'x is a positive single-digit number.'
```

But the second one is (arguably) preferable, as it is simpler to read.

```
1 if 0 < x and x < 10:
2     print 'x is a positive single-digit number.'
```

Recursion

A function is a allowed to call itself, in what is termed **recursion**

```
1  def countdown(n):
2      if n <= 0:
3          print'We have lift off!'
4      else:
5          print n
6          countdown(n-1)
```

Countdown calls itself!

```
1  countdown(10)
```

```
10
9
8
7
6
5
4
3
2
1
We have lift off!
```

But the key is that each time it calls itself, it is passing an argument with its value decreased by 1, so eventually, `n <= 0` is true.

With a small change, we can make it so that `countdown(1)` encounters an **infinite recursion**, in which it repeatedly calls itself.

```
1 def countdown(n):
2     if n <= 0:
3         print'We have lift off!'
4     else:
5         print n
6         countdown(n)
```

```
1 countdown(10)
```

```
---------------------------------------------------------------------------
RuntimeError                              Traceback (most recent call last)
<ipython-input-163-a972007fb272> in <module>()
----> 1 countdown(10)

<ipython-input-162-33965ef63097> in countdown(n)
      4     else:
      5         print n
----> 6         countdown(n)

... last 1 frames repeated, from the frame below ...

<ipython-input-162-33965ef63097> in countdown(n)
      4     else:
      5         print n
----> 6         countdown(n)

RuntimeError: maximum recursion depth exceeded
```

Repeated actions: Iteration

Recursion is the first tool we've seen for performing repeated operations

But there are better tools for the job: `while` and `for` loops.

```
1 def countdown(n):
2     while n>0:
3         print n
4         n = n-1
5     print 'We have lift off!'
```

```
1 countdown(10)
```

```
10
9
8
7
6
5
4
3
2
1
We have lift off!
```

Repeated actions: Iteration

Recursion is the first tool we've seen for performing repeated operations

But there are better tools for the job: `while` and `for` loops.

```
1 def countdown(n):
2     while n>0:
3         print n
4         n = n-1
5     print 'we have lift off!'
```

This block specifies a while-loop. So long as the condition is true, Python will run the code in the body of the loop, checking the condition again at the end of each time through.

Repeated actions: Iteration

Recursion is the first tool we've seen for performing repeated operations

But there are better tools for the job: `while` and `for` loops.

```
1 def countdown(n):
2     while n>0:
3         print n
4         n = n-1
5     print 'We have lift off!'
```

Warning: Once again, there is a danger of creating an **infinite loop**. If, for example, `n` never gets updated, then when we call `countdown(10)` , the condition `n>0` will always evaluate to `True`, and we will never exit the while-loop.

```
1 countdown(10)
```

```
10
9
8
7
6
5
4
3
2
1
We have lift off!
```

Repeated actions: Iteration

```
1 def collatz(n):
2     while n!=1:
3         print(n)
4         if n % 2 == 0:
5             n = n/2
6         else:
7             n = 3*n+1
```

```
1 collatz(20)
```

```
20
10
5
16
8
4
2
```

One always wants to try and ensure that a while loop will (eventually) terminate, but it's not always so easy to know!
https://en.wikipedia.org/wiki/Collatz_conjecture

"Mathematics may not be ready for such problems."
Paul Erdős

Repeated actions: Iteration

We can also terminate a while-loop using the `break` keyword

```
1 a = 4
2 x = 3.5
3 epsilon = 10**-6
4 while True:
5     print(x)
6     y = (x + a/x)/2
7     if abs(x-y) < epsilon:
8         break
9     x=y # update to our new estimate
```

```
3.5
2.32142857143
2.02225274725
2.00012243394
2.00000000375
```

The `break` keyword terminates the current loop when it is called.

Newton-Raphson method:
https://en.wikipedia.org/wiki/Newton's_method

Repeated actions: Iteration

We can also terminate a while-loop using the `break` keyword

```
1  a = 4
2  x = 3.5
3  epsilon = 10**-6
4  while True:
5      print(x)
6      x = (x + a/x)/2
7      if abs(x-y) < epsilon:
8          break
9      y = x # update to our new estimate
```

```
3.5
2.32142857143
2.02225274725
2.00012243394
2.00000000375
```

Notice that we're not testing for equality here. That's because testing for equality between pairs of floats is dangerous. When I write `x=1/3`, for example, the value of `x` is actually only an approximation to the number 1/3.

Newton-Raphson method:
https://en.wikipedia.org/wiki/Newton's_method

Strings in Python

Strings are sequences of characters

Python sequences are 0-indexed. The index counts the offset from the beginning of the sequence. So the first letter is the 0-th character of the string.

Note: in some languages, there's a difference between a character and a string of length 1. That is, the character 'g' and the string "g" are different data types. In Python, no such difference exists. A character is just a one-character string.

```
1 animal = 'goat'
2 letter = animal[1]
3 letter
```

```
'o'
```

```
1 animal[0]
```

```
'g'
```

```
1 animal[1]
```

```
'o'
```

```
1 animal[2]
```

```
'a'
```

```
1 animal[3]
```

```
't'
```

Strings in Python

Strings are **sequences** of characters

All Python sequences include a **length** attribute, which is the number of elements in the sequence.

```
1 len(animal)
```

```
4
```

If we try to access an element of the sequence that doesn't exist, we get an error.

```
1 animal[4]
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
<ipython-input-8-71de68f745e5> in <module>()
----> 1 animal[4]

IndexError: string index out of range
```

We can also index into a sequence counting from the end.

```
1 animal[-1]
```

```
't'
```

Strings in Python

```
1 i=0
2 while i<len(animal):
3     print animal[i]
4     i=i+1
```

```
g
o
a
t
```

We can index into a sequence using an index variable.

...but there's a better way to perform this operation...

Iterations and traversals: for-loops

```
1 for c in animal:
2     print(c)
```

```
g
o
a
t
```

```
1 i=0
2 while i<len(animal):
3     print animal[i]
4     i=i+1
```

```
g
o
a
t
```

For-loop provides a more concise way to express the pattern on the right.

Selecting subsequences: slices

A segment of a Python sequence is called a **slice**

```
1 s = "And now for something completely different"
2 s[0:7]
```

```
'And now'
```

```
1 s[12:21]
```

```
'something'
```

`string[m:n]` picks out the `m`-th character to the `n`-th character, including the `m`-th character, but **not** including the `n`-th character.

Selecting subsequences: slices

A segment of a Python sequence is called a **slice**

```
1 s = "And now for something completely different"
2 s[:7]
```

```
'And now'
```

`string[:m]` picks out the subsequence starting at 0 through the `(m-1)`-th character.

```
1 s[22:]
```

```
'completely different'
```

`string[m:]` picks out the subsequence starting at the `m`-th character through the end of the sequence.

```
1 s[-20:-10]
```

```
'completely'
```

Slices also work with negative indexing.

```
1 s[-20:]
```

```
'completely different'
```

Selecting subsequences: slices

`string[:]` picks out the entire string.

```
1 s = "And now for something completely different"
2 s[:]
```

```
'And now for something completely different'
```

`string[x:x]` picks out the `x`-th through `x`-th letters, not including the `x`-th, so this gets the **empty string.**

```
1 s[2:2]
```

```
''
```

Selecting subsequences: slices

`string[:]` picks out the entire string.

```
1 s = "And now for something completely different"
2 s[:]
```

```
'And now for something completely different'
```

`string[x:x]` picks out the `x`-th through `x`-th letters, not including the `x`-th, so this gets the **empty string.**

```
1 s[2:2]
```

```
''
```

The empty string is a string just like any other, but it contains **no letters** and has length 0.

Important concept: immutability

What if I want to change a letter in my string?

```
1 mystr = 'goat'
2 mystr[0] = 'b'
```

Try and assign a different string to a subsequence of a string.

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-27-cf531ebc1ce4> in <module>()
      1 mystr = 'goat'
----> 2 mystr[0] = 'b'

TypeError: 'str' object does not support item assignment
```

We get an error because strings are **immutable**.
We can't change the value of an *existing* string.

Important concept: immutability

What if I want to change a letter in my string?

```
1 mystr = 'goat'
2 mystr = 'b'+mystr[1:]
3 mystr
```

```
'boat'
```

This avoids the error we saw before because it changes the value of the variable `mystr`, rather than trying to change the contents of a string.

Example: string traversal

```
1 def count(word, letter):
2     cnt = 0
3     for c in word:
4         if c==letter:
5             cnt = cnt+1
6     return cnt
```

```
1 count('banana', 'a')
```

```
3
```

```
1 count('banana', 'z')
```

```
0
```

The function `count` makes use of a common pattern, often called a **traversal**. We examine each element of a sequence (i.e., a string), taking some action for each element.

The variable `cnt` keeps a tally of how many times we have seen letter in the string word, so far. We call such a variable a **counter** or an **accumulator**.

Python string methods

Python strings provide a number of built-in operations, called **methods**

```
1 mystr = 'goat'
2 mystr.upper()
```

```
'GOAT'
```

```
1 'aBcDeFg'.lower()
```

```
'abcdefg'
```

`str.upper()` makes all letters in `str` upper case. `str.lower()` is analogous.

```
1 'banana'.find('na')
```

```
2
```

`str.find(sub)` finds the index of the first location of the string sub in str.

```
1 'goat'.startswith('go')
```

```
True
```

`str.startswith(sub)` returns `True` if and only if `str` starts with `sub`.

Python string methods

Python strings provide a number of built-in operations, called **methods**

```
1 mystr = 'goat'
2 mystr.upper()
```

```
'GOAT'
```

```
1 'aBcDeFg'.lower()
```

```
'abcdefg'
```

```
1 'banana'.find('na')
```

```
2
```

```
1 'goat'.startswith('go')
```

```
True
```

This `variable.method()` notation is called **dot notation**, and it is ubiquitous in Python (and many other languages).

A **method** is like a function, but it is provided by an **object**. We'll learn much more about this later in the semester, but for now, it suffices to know that some data types provide what *look* like functions (they take arguments and return values), and we call these function-like things **methods**.

Many more Python string methods:
https://docs.python.org/3/library/stdtypes.html#string-methods

Optional arguments: `str.find()`

```
1 'banana'.find('na')
```

```
2
```

```
1 'banana'.find('na', 3)
```

```
4
```

```
1 'banana'.find('na', 3, 4)
```

```
-1
```

```
1 'banana'.find('na', 3, 6)
```

```
4
```

The `str.find()` method takes **optional arguments**, which specify where in the string to start looking for a match, and the last index to consider for a match.

Find first occurrence of `'na'`, starting from index `3`.

Find first occurrence of `'na'`, starting from index `3`, and nowhere past `4`.

The documentation writes this method as `str.find(sub[, start[, end]])`. Square brackets indicate optional arguments. In this case, brackets also indicate that with two arguments, the second one will be interpreted as the `start` argument.
https://docs.python.org/3/library/stdtypes.html#string-methods

Searching sequences: the `in` keyword

```
1 'a' in 'banana'
```

```
True
```

```
1 'z' in 'banana'
```

```
False
```

```
1 'ban' in 'banana'
```

```
True
```

```
1 'anan' in 'banana'
```

```
True
```

```
1 'zoo' in 'banana'
```

```
False
```

`x in y` returns `True` if `x` occurs as a substring of `y`, and `False` otherwise.

Importantly, we can check for a whole substring, making this very similar to `str.find()`.

The `in` keyword applies more generally to check whether an object is contained in a sequence. We'll see more examples of this in the future, but for now, we only need to worry about strings.

String Comparison

Sometimes we want to check if two strings are equal

```
1 'cat' == 'cat'
```

```
True
```

```
1 'cat' == 'dog'
```

```
False
```

```
1 'dog' == 'doge'
```

```
False
```

Use the equality operator (==), just like for comparing numbers.

Strings have to match exactly. Substring is not enough!

String Comparison

Sometimes we want to check if two strings are equal

```
1 'cat' == 'cat'
```

```
True
```

```
1 'cat' == 'dog'
```

```
False
```

```
1 'dog' == 'doge'
```

```
False
```

Use the equality operator (==), just like for comparing numbers.

Strings have to match exactly. Substring is not enough!

If we can compare strings with equality, we should be able to compare them with inequalities, too...

String Comparison

We can also compare words under alphabetical ordering

```
1 'cat' < 'dog'
```

```
True
```

```
1 'cat' >= 'dog'
```

```
False
```

```
1 'dog' < 'doge'
```

```
True
```

```
1 '' < 'goat'
```

```
True
```

```
1 '1' < 'a'
```

```
True
```

Words earlier in the dictionary are “smaller” than words later in the dictionary.

The empty string `''` comes first in the ordering.

Strings including numbers, symbols, etc. are also ordered.

String Comparison

Important: upper case and lower case letters ordered differently!

```
1 'Cat' == 'cat'
```

```
False
```

```
1 'cat' > 'Cat'
```

```
True
```

Upper case letters are ordered before lower case letters.

For more information:
https://docs.python.org/3/library/stdtypes.html#comparisons

For **much** more information:
https://docs.python.org/3/library/operator.html?highlight=equallity

Python Lists

Strings in Python are “sequences of characters”

But what if I want a sequence of something else?

- A vector would be naturally represented as a sequence of numbers
- A class roster might be represented as a sequence of strings

Python lists are sequences whose values can be of any data type

- We call these list entries the **elements** of the list