Introduction to Python 1

Chang Y. Chung

Office of Population Research

01/14/2014

Why Python

- ► Popular
- ► Easy to learn and use
- ► Open-source
- ► General-Purpose
- ► Multi-paradigm (procedureal, object-oriented, functional)
- ► Hettinger, "What makes Python Awesome?"[2] http://tinyurl.com/mndd4er
- ▶ Did I mention *popular*?

1/4

Organization

- ▶ First two hours for learning basics of Python language.
- ▶ Third hour for a guided tour of Python "ecosystem".

Running Python Read-Eval-Print Loop

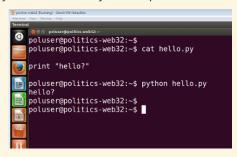
- ► For those who are using UNIX-like systems, including the Ubuntu image running on a Virtual Box.
- ▶ Type "python" at the terminal shell prompt.

```
| Principal Coach Minuse | Principal Coach Min
```

3/46

Running a Python Script File (.py)

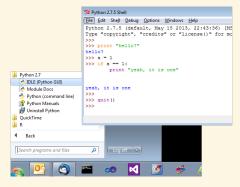
▶ Type "python" followed by the script file name.



4/4

Running Python REPL on Windows

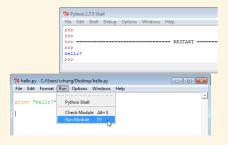
- ► For those who are using Windows sytem with Python (and IDLE) installed.
- ► Run the IDLE application.



5/46

Running a Python Script File (.py) on Windows

- ► File > New (or OPEN) brings up a script window.



6/46

Conceptual Hierarchy by Mark Lutz[3]

- ▶ Programs are composed of *modules*.
- ► Modules contain *statements*.

7/46

Conceptual Hierarchy by Mark Lutz[3]

- ▶ Programs are composed of *modules*.
- ▶ Modules contain *statements*.
- ► Statements contain expressions.
- ► Expressions create and process objects.

Script File (.py)

- ► A script file is a module.
- ► A script is a sequence of statements, delimited by a newline (or the end of line character).
- ▶ Python executes one statement at a time, from the top of a script file to the bottom.

8/46

Script File (.py)

- ▶ A script file is a module.
- ► A script is a sequence of statements, delimited by a newline (or the end of line character).
- ▶ Python executes one statement at a time, from the top of a script file to the bottom.
- ► Execution happens in namespaces (modules, classes, functions all have their own).
- ► Everything is runtime (even def, class, and import)

8/46

Executable Python Script File

► On a Unix-like system, we can change the mode of the script file and make it executable:

```
$$ $chmod +x hello.py
$$ $./hello.py
```

▶ Just add the first line with a hashbang (#!):

```
#!/usr/bin/python

# say hello to the world

def main():
    print "Hello, World!"

if __name__ == "__main__":
    main()
```

Comments

► Comments start with a hash (#) and end with a newline.

```
# this whole line is a comment

# add 10 integers.

total = 0
for i in range(10): #igoes 0, 1, 2, ..., 9

total += i  #shortcut for total = total + i

print "total=", total
# total = 45
```

10/46

Variables

▶ Variables are created when first assigned a value.

```
1 my_var = 3
2 answer_to_everything = 42
3
4 #also works are:
5 x = y = 0
6 a, b, c = 1, 2, 3
```

- ➤ Variable names start with a letter or an underscore(_) and can have letters, underscores, or digits.
- ▶ Variable names are case-sensitive.

11/46

Assignment Semantics According to David Godger[1]

➤ Variables in many *other languages* are a container that stores a value. 1

1 int a = 1;

Assignment Semantics According to David Godger[1]

- ➤ Variables in many *other languages* are a container that stores a value. ◀
- int a = 1;
- ► In Python, an assignment creates an object, and labels it with the variable. ✓
- 1 a = 1



12/46

Assignment Semantics According to David Godger[1]

- int a = 1;
- ► In Python, an assignment creates an object, and labels it with the variable. ✓
- a = 1
- ► If you assign another value to a, then the variable labels the new value (2).
- 1 a = 2



12/46

Assignment Semantics According to David Godger[1]

- int a = 1;
- ► In Python, an assignment creates an object, and labels it with the variable.
- a = 1
- ► If you assign another value to a, then the variable labels the new value (2).
- a = 2
- ▶ This is what happens if you assign a to a new variable b:
- 1 b = a



Numbers

▶ Integers

```
1  x = 0
2  age = 20
3  size_of_household = 5

4
5  print type(age)
6  # <type 'int'>
7  # can handle arbitrarily large numbers
9  huge = 10 ** 100 + 1

Floating-point Numbers
1  g = 0.1
2  f = 6.67384

4  velocity = 1.  # it is the dot (.) that makes it a float
5  print velocity
6  #1.0
7  type(velocity)
```

13/46

Numeric Expressions

8 # <type 'float'>

▶ Most of the arithmetic operators behave as expected.

```
1 a = 10

2 b = 20

3 print a - (b ** 2) + 23

4 #-367

5

6 x = 2.0

7 print x / 0.1

8 #20.0
```

► Watch out for integer divisions. In Python 2, it *truncates* down to an integer.

```
1 print 10 / 3
2 #3 (in Python 2) 3.3333333333333335 (in Python 3)
3 print -10 / 3
4 #-4 (in Python 2) -3.3333333333335 (in Python 3)
5
6 #a solution: use floating point numbers
7 print 10.0 / 3.0
8 #3.33333333333333335 (in both Python 2 and 3)
```

14/46

String Literals

► A string is a sequence of characters.

```
# either double (") or single(') quotes for creating string literals.

name = "Changarilla Dingdong"

file_name = 'workshop.tex'

# triple-quoted string
starwars = """

A long time ago is a galaxy far, far away...

It is a period of civil war. Rebel
spaceships, striking from a hidden
base, have won their first victory

...

What is the last character of this string?

"""

last_char = starwars[-1]
print ord(last_char), ord("\n")

# 10 10
```

Working with strings

▶ Strings are immutable.

```
1  s = "abcde"
2  s[0] = "x" # trying to change the first char to "x"
3  #TypeError: 'str' object does not support item assignment
4  
5  t = "x" + s[1:] # creating a new string
6  print t
7  #xbcde
```

▶ Many functions and methods are available.

```
1  s = "abcde"
2  print s + s # concatenation
3  # abcdeabcde
4
5  print len(s)
6  # 5
7
8  print s.find("c") # index is 0-based. returns -1 if not found
9  # 2
```

16/46

String Manipulation

► A few more string methods.

17/46

String Manipulation

► A few more string methods.

► What are "methods"?

String Manipulation

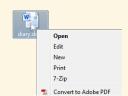
▶ A few more string methods.

- ▶ What are "methods"?

17/46

String Manipulation

► A few more string methods.



- ▶ What are "methods"?
 - $\,$ Functions which are a member of a type (or class).

 - 2, "abcde", diary.docx are an instance (or object) of the respective type.
 - ▷ Types have members: properties (data) and methods (functions).

17/46

Two Ways to Format

▶ format() method.

```
print "The answer is {0}".format(21)
#The answer is 21
print "The real answer is {0:6.4f}".format(21.2345678)
#The answer is 21.2346
```

► Formatting operator.

```
print "The answer is %d" % 21
#The answer is 21
print "The real answer is %6.4f" % 21.2345678
#The answer is 21.2346
```

Raw Strings

 Within a string literal, escape sequences start with a backslash (\)

```
1 a_string = 'It\'s a great day\nto learn \\Python\\.\n 1\t 2\t 3'
2 print a_string
3 #It's a great day
4 #to learn \Python\.
5 #1 2 3
```

► A raw string literal starts with the prefix r. In a raw string, the backslash (\) is not special.

```
import re

mathrightarrow import re

mathrightarrow import representation of the composition of the com
```

19/46

Unicode Strings

➤ You can create Unicode strings using the u prefix and "\u" escape sequences.

20/46

Unicode Strings

➤ You can create Unicode strings using the u prefix and "\u" escape sequences.

```
1 a_string = u"Euro \u20AC" #\u followed by 16-bit hex value xxxx print a_string, len(a_string)  
3 #Euro \in 6
```

- ▶ Unicode strings are sequences of *code points*.
- ➤ Code points are numbers, each representing a "character". e.g., U+0061 is 'Latin small letter a'.
- ▶ Unicode text strings are *encoded* into bytes. UTF-8 is one of many Unicode encodings, using one to four bytes to store a Unicode "character".
- ➤ Once you have Unicode strings in Python, all the string functions and properties work as expected.

Best Practices According to Thomas Wouters[5]

- ▶ Never mix unicode and bytecode [i.e. ordinary] strings.
- ▶ Decode bytecode strings on input.
- ▶ Encode unicode strings on output.
- ► Try automatic conversion (codecs.open())
- ► Pay attention to exceptions, UnicodeDecodeError
- ► An example

```
ustr = u"Euro \u20AC" #Euro €

#Python's default encoding codec is 'ascii'
ustr.encode()
#UnicodeEncodeError: 'ascii' codec can't encode characater 'u\u20ac'
utf_8 = ustr.encode("UTF-8") # encoding to UTF-8 works fine
#this takes 8 bytes. five one-byte's and one three-byte

#now we want to decode to ascii, ignoring non-ascii chars
print utf_8.decode("ascii", "ignore")
#Euro (no euro symbol character)
```

21/46

None

- ▶ Is a place-holder, like NULL in other languages.
- x = None
- ▶ Is a universal object, i.e., there is only one None.

```
print None is None
True
```

▶ Is evaluated as False.

```
1  X = None
2  if X:
3     print "this will never print"
```

▶ Is, however, distinct from others which are False.

```
print None is 0, None is False, None is []
#False, False, False
```

22/46

Core Data Types

- ▶ Basic core data types: int, float, and str.
- ▶ "Python is dynamically, but *strongly* typed."

```
1  n = 1.0
2  print n + "99"
3  #TypeError: unsupported operand types(s) for +: 'int' and 'str'
```

► Use int() or float() to go from str to numeric

```
print n + float("99")
# 100.0
```

▶ str() returns the string representation of the given object.

```
1  n = 1.0
2  print str(n) + "99"
3  #1.099
```

Operators

▶ Python supports following types of operators:

```
    ▷ Arithmetic (+ - * / % ** //)
    ▷ Comparison (==!=> < >=<=)</li>
    ▷ Assignment (= += -= *= /= %= **= //=)
    ▷ Logical (and or not)
    ▷ Bitwise (& | ^ ~ <<>>)
    ▷ Membership (in not in)
    ▷ Identity (is is not)
```

24/46

A Few Surprises

► Power operator (**) binds more tightly than unary operators on the left.

```
1 a = -2**2
2 print a
3 #-4
4 #solution: parenthesize the base
5 print (-2)**2
```

► Comparisons can be chained.

```
1  x < y <= z  #y is evaluated only once here
is equivalent to
1  x < y and y <= z</pre>
```

► Logical operators (and, or) short-circuit evaluation and return an *operand*.

```
print 3 or 2
# 3
```

25/46

Quiz

▶ Demographic and Health Surveys (DHS) Century Month Code (CMC)[4, p.5] provides an easy way working with year and month.

The CMC is an integer representing a month, taking the value of 1 in January 1900, 2 in February 1900, ..., 13 in January 1901, etc. The CMC in February 2011 is 1333.

What is the CMC for this month, i.e., January, 2014?

Quiz

▶ Demographic and Health Surveys (DHS) Century Month Code (CMC)[4, p.5] provides an easy way working with year and month.

The CMC is an integer representing a month, taking the value of 1 in January 1900, 2 in February 1900, ..., 13 in January 1901, etc. The CMC in February 2011 is 1333.

What is the CMC for this month, i.e., January, 2014?

► An answer:

```
1  year = 2014
2  month = 1
3  print "CMC", (year - 1900) * 12 + month
4  # CMC 1369
```

26/46

Quiz

▶ Demographic and Health Surveys (DHS) Century Month Code (CMC)[4, p.5] provides an easy way working with year and month.

The CMC is an integer representing a month, taking the value of 1 in January 1900, 2 in February 1900, ..., 13 in January 1901, etc. The CMC in February 2011 is 1333.

What is the CMC for this month, i.e., January, 2014?

► An answer:

```
1  year = 2014
2  month = 1
3  print "CMC", (year - 1900) * 12 + month
4  # CMC 1369
```

▶ What is the month (and year) of CMC 1000?

26/46

Quiz

- ► According to U.S. National Debt Clock, the current outstanding public debt, as of a day last month, was a big number:
- 1 debt = 17234623718339.96

Count how many times the digit 3 appears in the number. (Hint: create a string variable and use the count() method of the string type.)

Quiz

- ➤ According to U.S. National Debt Clock, the current outstanding public debt, as of a day last month, was a big number:
- 1 debt = 17234623718339.96

Count how many times the digit 3 appears in the number. (Hint: create a string variable and use the count() method of the string type.)

► An answer:

```
sdebt = "17234623718339.96"
print sdebt.count("3")
# 4
```

27/46

Quiz

- ➤ According to U.S. National Debt Clock, the current outstanding public debt, as of a day last month, was a big number:
- 1 debt = 17234623718339.96

Count how many times the digit 3 appears in the number. (Hint: create a string variable and use the count() method of the string type.)

► An answer:

```
sdebt = "17234623718339.96"
print sdebt.count("3")

# 4
```

► (tricky) It feels rather silly to rewrite the value as a string. Can you think of a way to *convert* the number into a string?

27/46

Flow Control

- ► Conditional Execution (if)
- ► Iteration

▶ IF statement is used for conditional execution.

```
1  if x > 0:
2     print "x is positive"
3  else:
4     print "x is zero or negative"
```

► Only one suite (block of statements) under a True conditional expression is executed.

29/46

Compound statements

- ▶ if, while, and for are *compound* statements, which have one or more *clauses*. A *clause*, in turn, consists of a *header* that ends with a colon (:) and a *suite*.
- ▶ A suite, a block of statements, is identified by *indentation*.

```
1 a = 1
   if a > 0:
      desc = "a is positive"
                                  # these two lines form a suite
      print a, desc
                                 # (blank line ignored)
6 print "done"
                                  # This line being started "dedented"
                                  # signals the end of the block
      if a > 0:
                                 # indentation error
10 desc = "a is positive"
                                # indentation error
11
12 if a > 0:
desc = "a is positive" #OK
print a, desc #OK
print a, desc #OK
15 else:
16 print a
```

30/46

Compound statements

- ➤ The amount of indentation does not matter (as long as the same within a level). Four spaces (per level) and no tabs are the convention.
- ► Use editor's python mode, which prevents/converts <TAB> to (four) spaces.
- ► Why indentation? A good answer at: http://tinyurl.com/kxv9vts.
- ► For an empty suite, use pass statement, which does nothing.

WHILE

► Repeats a block of statements as long as the condition remains True.

```
1 total = 0
2 n = 0
3 while n < 10:
4     print n, total
5     n += 1
6     total += n
7 #00
8 #11
9 #23
10 #...
11 #945</pre>
```

- ▶ break terminates the loop immediately.
- ► continue skips the remainder of the block and goes back to the test condition at the top.

32/46

FOR

▶ for is used to iterate over a sequence.

33/46

FOR

► Another example.

```
numbers = range(5)
print numbers
| #[0, 1, 2, 3, 4]

for n in numbers:
    print n,
    if n % 2 == 0:
        print "is even"
    else:
        print "is odd"

# 0 is even
| # 2 is even
| # 3 is odd
| # 4 is even
```

File I/O

► The built-in function, open(), returns a file type object, unless there is an error opening the file.

```
in_file = open("yourfile.txt", "r") # for reading
out_file = open("myfile.txt", "w") # for writing
```

▶ Once we get the file type object, then use its methods.

```
# read all the contents from the input file
content = in_file.read()

# write out a line to the output file
out_file.write("hello?\n")

# close it when done
in_file.close()
out_file.close()
```

35/46

Reading a file one line at a time

▶ with ensures that the file is closed when done.

```
with open("lorem.txt", "r") as f:
for line in f:
print line
```

► Another example.

```
# creating a file and writing three lines
with open("small.txt", "w") as f:
    f. write("first\n")
    f. write("second\n")
    f. write("third")

with open("small.txt", "r") as f:
    for line in f:
    print line # line includes the newline char

# first
#
# second
# # third
```

36/46

Defining and Calling a Function

▶ Defined with def and called by name followed by ().

```
1 def the_answer():
2    return 42
3
4    print the_answer()
5    #42
```

► Argument(s) can be passed.

```
def shout(what):
    print what.upper() + "!"

shout("hello")

#HELLO!
```

► If the function returns nothing, then it returns None.

```
1    r = shout("hi")
2    # HI!
3    print r
4    # None
```

Another example

► CMC again

```
def cmc(year, month):
    ''' returns DHS Century Month Code ''' # doc string
    if year < 1900 or year > 2099:
        print "year out of range"
        return
    if month < 1 or month > 12:
        print "month out of range"
        return
    value = (year - 1900) * 12 + month
    return value

print cmc(2014, 1)
# 1369
print cmc(2014, 15)
# month out of range
# None
```

Local and Global Variables

- ► Within your function:
 - ▷ A new variable is *local*, and independent of the global var with the same name, if any.

```
1  x = 1  # global (or module)
2
3  def my_func():
4   x = 2  # local
5  my_func()
6  print x  # 1
```

- ⊳ Both local and global variables can be read.

39/46

Quiz

➤ Write a function that returns Body Mass Index (BMI) of an adult given weight in kilograms and height in meters. (Hint: BMI = weight(kg) / (height(m) squared). For instance, if a person is 70kg and 1.80m, then BMI is about 21.6.)

Quiz

- ➤ Write a function that returns Body Mass Index (BMI) of an adult given weight in kilograms and height in meters. (Hint: BMI = weight(kg) / (height(m) squared). For instance, if a person is 70kg and 1.80m, then BMI is about 21.6.)
- An answer.

```
def bmi(kg, m):
return float(kg) / (m ** 2)
```

40/46

Quiz

- ➤ Write a function that returns Body Mass Index (BMI) of an adult given weight in kilograms and height in meters. (Hint: BMI = weight(kg) / (height(m) squared). For instance, if a person is 70kg and 1.80m, then BMI is about 21.6.)
- ► An answer.

```
def bmi(kg, m):
return float(kg) / (m ** 2)
```

➤ Re-write the bmi function so that it accepts height in feet and inches, and the weight in pounds. (Hint. Make pound, foot, and inch arguments. Convert them into local variables, kg and m, before calculating bmi to return.)

40/46

Importing a module

- ▶ import reads in a module, runs it (top-to-bottom) to create the module object.
- ► Via the module object, you get access to its variables, functions, classes, ...
- ► We've already seen an example of importing a standard regular expression module:

```
import re

mathrightarrow imports referred in the imported referred re
```

Another example

► There are many standard modules that come already installed, and be ready to be imported.

```
import math

s s = math.sqrt(4.0)
print "4.0 squared is {0:.2f}".format(s)
#4.0 squared is 2.00
```

► You can selectively import as well.

```
from math import sqrt #importsqrt() alone
print sqrt(9.0) # no "math."
# 3.0
```

➤ You can import your own Python script file (.py) the same way. The default import path includes the current working directory and its sub-directories.

```
1 import hello #suppose that hello.py defines a main() function
2
3 hello.main()
4 #"hello, World!"
```

42/46

Quiz

► Write a function such that, given a BMI value, returns the BMI category as a string. Recall the categories are:

```
Underweight less than 18.5
Normal weight 18.5 upto but not including 25
Overweight 25 upto but not including 30
Obesity 30 or greater
```

For instance, the function should return a string "Normal weight", when it is called with an argument of, say 20. (Hint: use conditional statements i.e., if ... elif ...)

► Print out a BMI table showing several lines of a pair: a BMI value and its category. BMI value may start at 15 and go up by 3 up to 36. (Hint: use a loop)

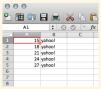
43/46

Quiz (cont.)

► Create a comma-separated values (.csv) file of the BMI table. (Hint: You may start with below, and modify it as neccessary.)

```
import csv

with open("test.csv", "wb") as f:
    my_writer = csv.writer(f)
    i = 15
    while i < 30:
        my_writer.writerow([i, "yahoo!"])
    i += 3</pre>
```



44/46

Summary

- ▶ Using Python interactively or by running a script.
- ► Comments.
- ▶ Variables and assignment semantics.
- ► Core data types (int, float, str, None).
- ▶ Operators.
- ► Conditionals and Looping.
- ▶ Defining and using functions.
- ► Basic File I/O.
- ▶ Importing a module.

References



Godger, D.

Code like a pythonista: Idiomatic python.

http://tinyurl.com/2cv9kg.



Hettinger, R.

What makes Python Awesome?

http://tinyurl.com/mndd4er.



Learning Python, fifth ed.

O'Reilly Media, Sebastopol, CA, 2013.



MEASURE DHS Plus.

Description of the Dempgraphic and Health Surveys Individual Recode Data File, version 1.0 ed., March 2008.



Wouters, T.

Advanced python: (or understanding python) google tech talks. feb 21, 2007. http://www.youtube.com/watch?v=u0zdG3lwcB4.