Heidi Lischer
University of Zurich

URPP Evolution in Action

16 March 2015

Exercises Python basics Tutorial

Download the files:

https://www.dropbox.com/s/2d3hmp3g3xb46ox/Python.zip?dl=0

Work locally on your laptop

The following instructions have been tested under Python 2.7.3:

1. Copy the zipped data to your computer (Ubuntu):
   ```
   wget https://www.dropbox.com/s/2d3hmp3g3xb46ox/Python.zip?dl=0
   ```

2. Unzip the data:
   ```
   unzip Python.zip
   ```

3. Start Python by typing `python` in the terminal

4. Or work in an IDE (integrated development environment):
   a. Download and install Spyder:
      ```
      sudo apt-get install spyder
      ```
   b. Start spyder by typing `spyder` in the terminal

Exercise 1: writing for-loops

a) Write some lines of code that will calculate the sum 1+2+3+...+300.
The idea is to create a variable that will store the current value of the sum. Set it to zero, then use a for-loop that will run through the numbers 1, 2, 3, ... and add each of these to the current sum. After the for-loop, output the value of the sum.

b) Write some lines of code that will calculate 1 + 1/1! + 1/2! + … 1/10! (3! = factorial = 3*2*1 = 6). Display the sum after each term is added.

c) Write some lines of code that lists all numbers from 1 to 100 that are divisible by 2 and 3

Exercise 2: writing while-loops

Write some lines of code that will figure out how many terms in the sum 1+2+3+... it requires for the sum to exceed one million. The idea is to create a variable that will store the current value of the sum and another variable that keeps track of what number you are adding to the sum.

Exercise 3: list

The number of birds banded at a series of sampling sites has been counted by your field crew and entered into the following list:

```
data = [['A1', 28], ['B1', 99], ['D3', 55], ['B2', 27], ['B4', 25],
['B5', 9], ['C3', 36], ['A7', 30], ['B6', 38], ['A3', 1], ['B3', 36],
['B7', 21], ['C1', 122], ['A4', 0], ['C2', 87], ['A6', 22], ['C4', 3],
['D1', 0], ['A2', 32], ['D2', 5], ['A8', 19], ['B8', 12], ['D4', 62],
['D5', 98], ['A5', 10], ['D6', 32]]
```

The first item in each sublist is an alphanumeric code for the site and the second value is the number of birds banded. Cut and paste the list into your assignment and then write some lines of code that answer the following questions:

a) How many sites are there?

b) How many birds were counted at the 7th site?

c) How many birds were counted at the second last site?

d) What is the total number of birds counted across all sites?

e) What is the average number of birds seen on a site?

f) What is the total number of birds counted on sites with codes beginning with C? (don't just identify this sites by eye, use the function `string.startswith(prefix)` to check if the string starts with the prefix. Example:

```
a = 'bla'
a.startswith('b')
true
```

)

g) Sort the list by sampling site.

h) One of your field assistant made a mistake and entered at sampling site B1 99 birds instead of 145. Correct the entry in the list.

i) You went back into the field and counted birds at four additional sites: C5=11, E1=34, E2=26 and E3=68. Add these sites to the list and make sure that the site C5 is entered after the site C4.

Exercise 4: input/output

The file simple.fa contains sequence data in FASTA format.

a) Calculate the average line length of the file (make sure that you don’t count newline characters).
b) Calculate the average length of the sequences in the file (use the function `string.startswith(prefix)` to check if the string starts with the prefix)
c) Rename the sequences to read_1, read_2, read_3, … and save it into a new file.
d) Read the data into a list ([[name1,seq1],[name2,seq2],…]). Then remove all reads shorter than 60bp. Afterwards, save a file with read names and a file with sequence length.

Solutions:

You can find solutions to all exercises within the downloaded folder (solutions.py).

Sources:

- http://www-math.bgsu.edu/~zirbel/programming/
- http://www.programmingforbiologists.org/