Overview of Python

Python for Genomic Data Science

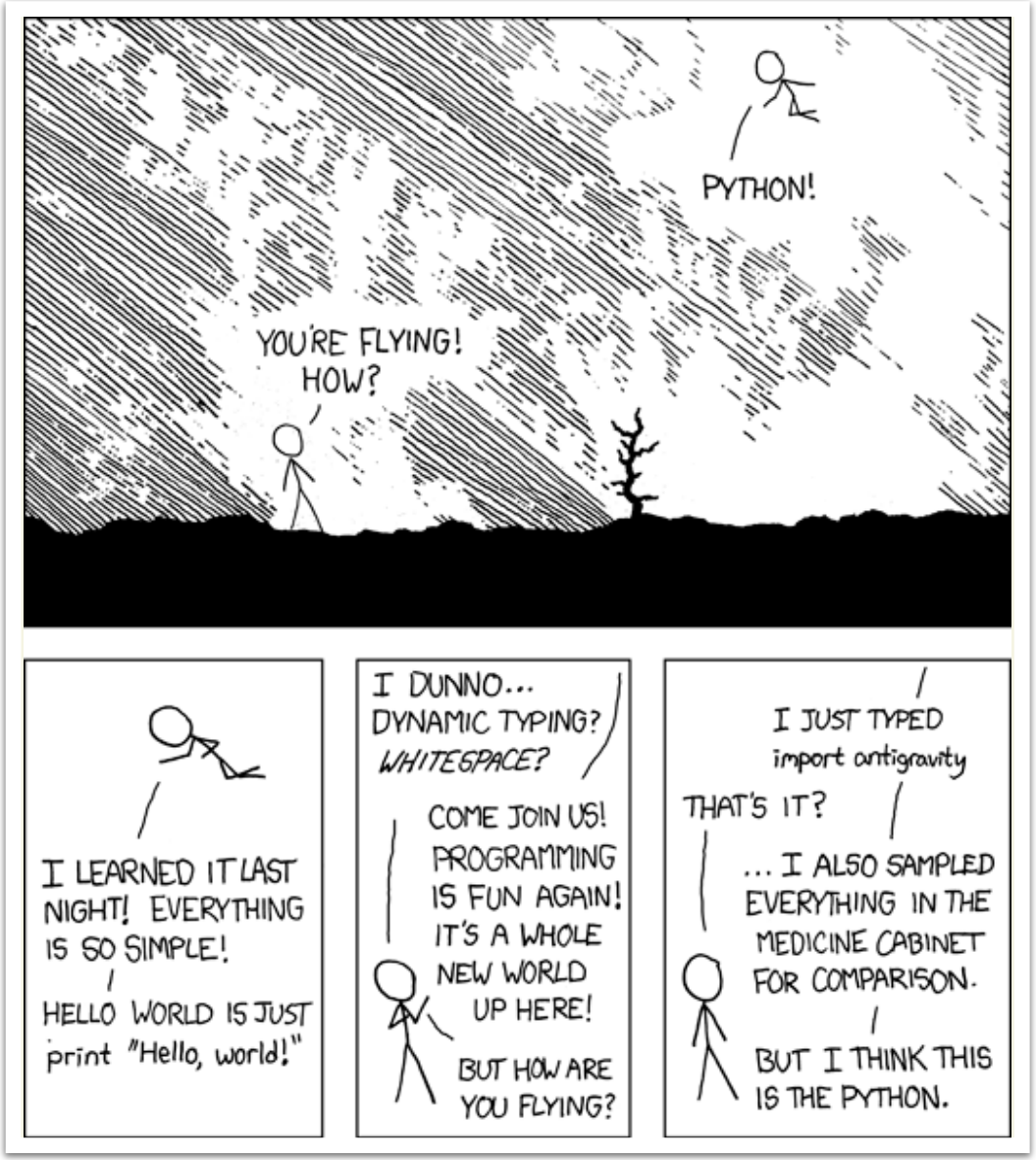
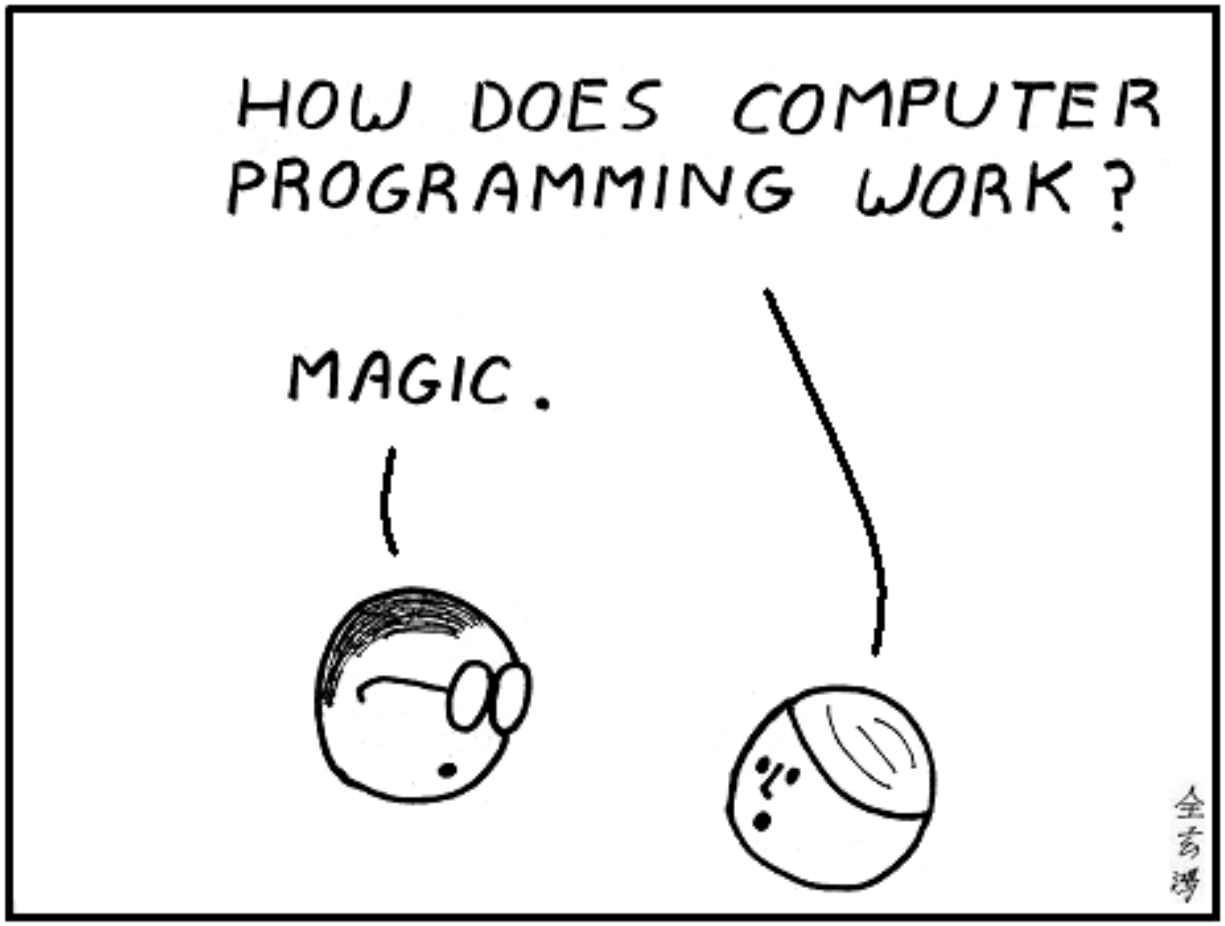


Image credit: XKCD, at http://xkcd.com/353/

Who is This Course For?

- No prior knowledge expected: intended for students with no experience in programming whatsoever.
- For you: oriented towards programming tasks for molecular biology.

Why Would You Want To Program?



Approaches to Learning How To Program

- Take a class like this that teaches you the basics
- Read a programming book
- Find a program written by somebody else and try to understand it
- Ask an expert!

Programming Strategies*

1. Identify the required inputs, such as data or specifications from the user.
2. Make an overall design for the program, including listing all the steps by which the program computes the output.
3. Decide what will be the output of the program: will the result be in a file, or displayed on the screen?
4. Refine the overall design by specifying more detail.
5. Write the program.

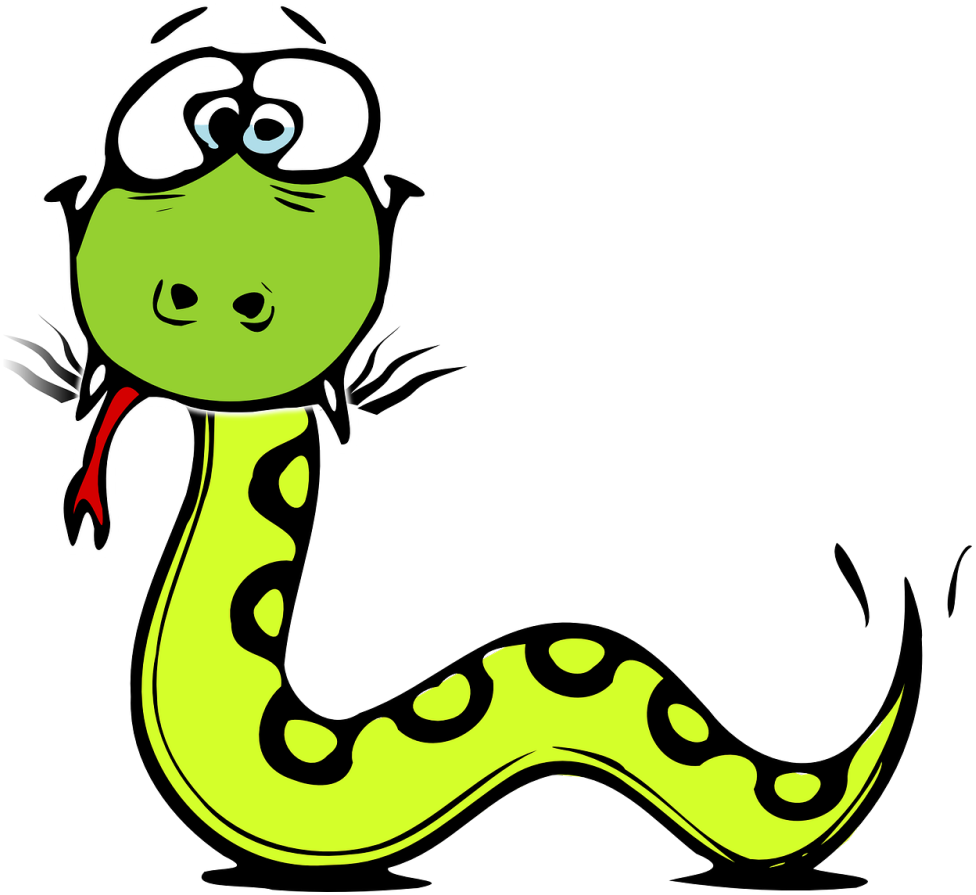
* Adapted from *Beginning Perl for Bioinformatics* by James Tisdall, Published by O'Reilly Media, Inc., 2001

Designing a Program

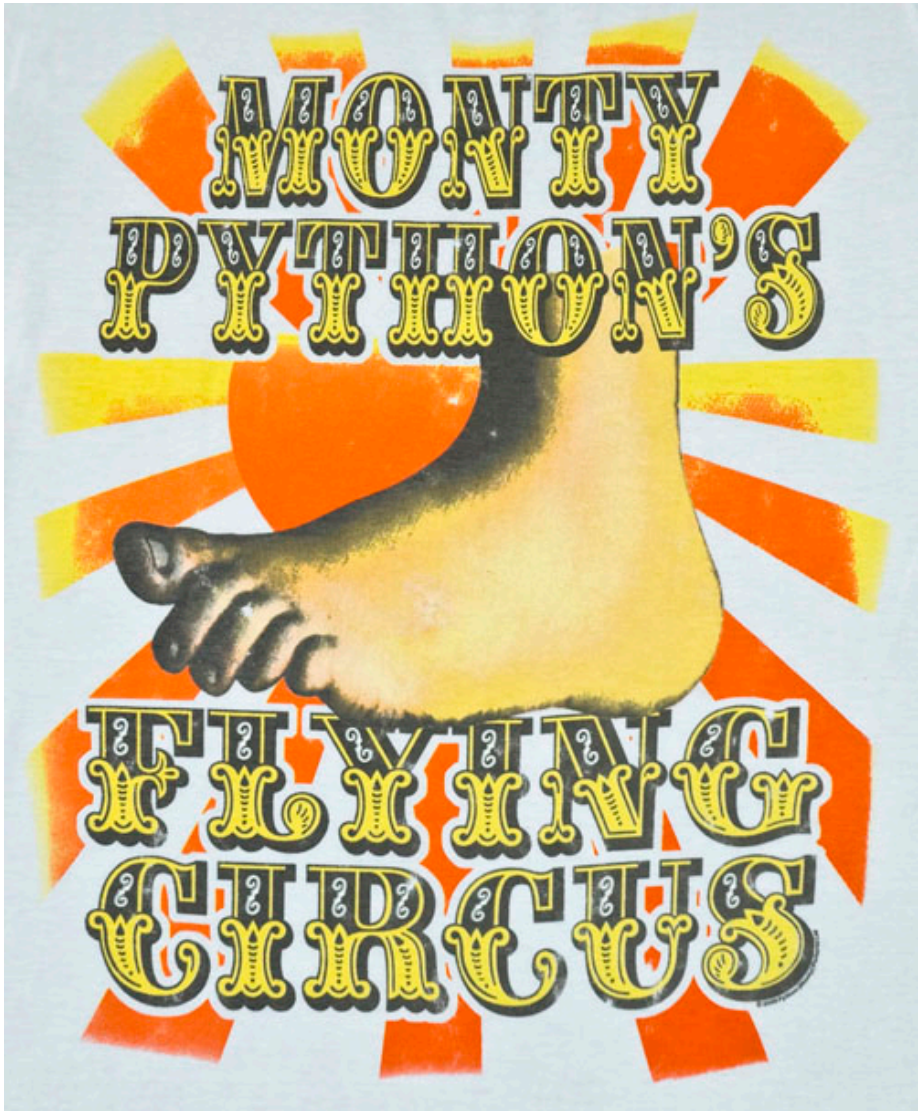
Write *pseudocode* for a program that computes the GC percentage composition of a DNA sequence:

- *read DNA sequence from user*
- *count the number of C's in DNA sequence*
- *count the number of G's in DNA sequence*
- *determine the length of the DNA sequence*
- *compute the GC%*
- *print GC%*

What is Python?



Python is not...



What is Python?

- Python is an easy to learn, powerful programming language
- It has efficient data structures and a simple but effective approach to object-oriented programming
- It is “interpreted”!
  - a compiled program is converted from source code into machine-language instructions by a compiler, and then run from its binary form
  - an interpreted program takes the code you wrote and converts it to binary “on the fly” each time you run it. This is slower.

History of Python

- Python was conceived and developed in the late 1980s and early 1990s by Guido van Rossum at the National Research Institute for Mathematics and Computer Science in the Netherlands.
- Python version 1.0 was released in January 1994.
- Python's core syntax and certain aspects of its philosophy are directly inherited from the language ABC, but Python is also influenced by Modula-3, C, C++, Perl, Java, the Unix shell, and other scripting languages.
- Python 2.0 was released on 16 October 2000.
- Python 3.0, a major, backwards-incompatible release, was released on 3 December 2008.

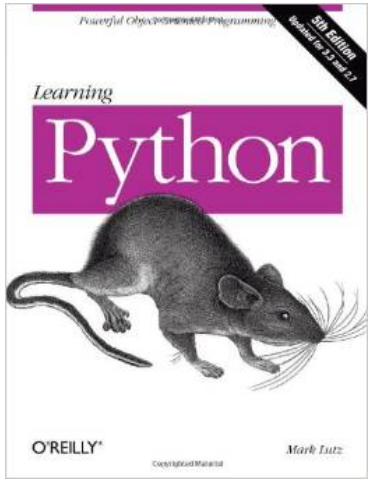
Python Features

- It is simple to use: Python syntax is clearly defined which makes it easily readable. Programs written in Python are typically much shorter than equivalent C, C++, or Java programs.
- It is interactive: you can write and test your programs directly from a terminal window.
- It has a large standard library: Python's library of built-in functions offers a wide range of programs that are already written for you.
- It is portable: Python runs on many Unix variants, on Mac OS, and on Windows.
- It is extensible: you can use it as an extension language for applications that need a programmable interface.
- It is scalable: you can use it for very small or very large programs.

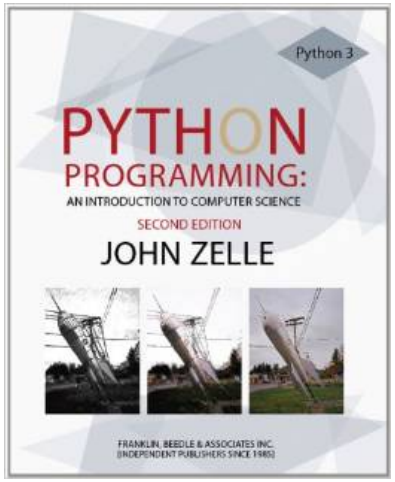
Some Python Resources

- http://www.python.org : the official Python site
- Introduction to Programming using Python -http://www.pasteur.fr/formation/infobio/python/ : a programming course for biologists at the Pasteur Institute
- Beginning Python for Bioinformatics -http://www.onlamp.com/pub/a/python/2002/10/17/biopython.html : a tutorial by Patrick O'Brien
- LearnPython.org : an interactive Python tutorial.
- Think Python: How to Think Like a Computer Scientist - http://www.greenteapress.com/thinkpython/ : a free online Python book by Allen B. Downey

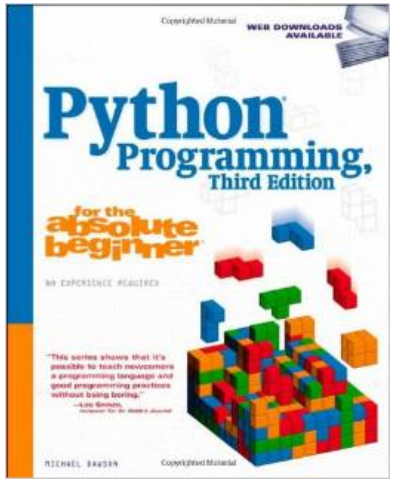
Some Useful Python Books



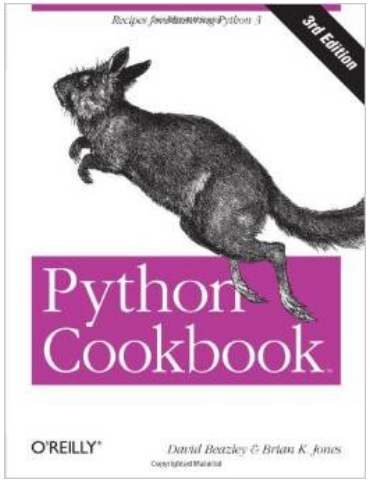
Learning Python
by Mark Lutz



Python Programming
by John Zelle



Python Programming for the Absolute Beginner
by Michael Dawson



Python Cookbook
by David Beazley and Brian K. Jones

Getting Python

The most up-to-date and current source code, binaries, documentation, news, etc. is available at the official website of Python:

http://www.python.org/download/

Running the Python Interpreter

```
$ python
Python 2.7.2 (default, Oct 11 2012, 20:14:37)
[GCC 4.2.1 Compatible Apple Clang 4.0 (tags/Apple/
clang-418.0.60)] on darwin
Type "help", "copyright", "credits" or "license" for more
information.
>>>
```