

Introduction to python

Digital lab, 2017

Ludovic Dutoit

Today's main challenge



Learning a language in 3h:

Plan

- What is python?
- Play around
- Providing resources

***“Not everything is at
your level, but there is
something for all levels”***

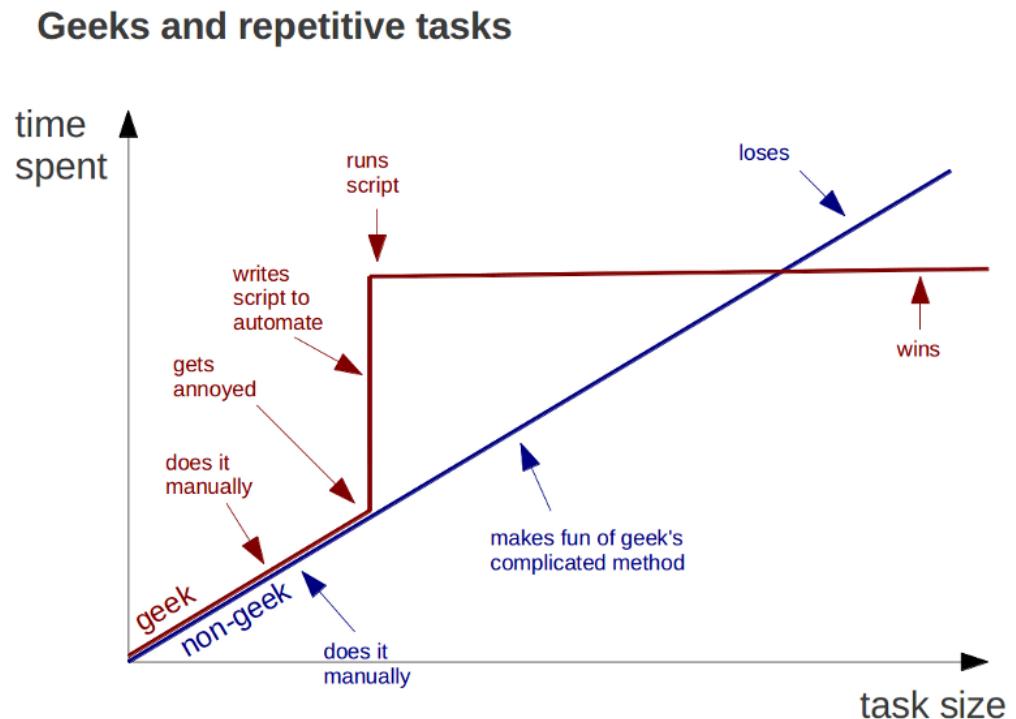
| Density | Total Larvae | Habitat type |
|----------------|---------------------|---------------------|
| 3 | 111 | A |
| 5 | 1322 | A |
| 11 | 123 | A |
| 23 | 123 | A |
| 43 | 342 | A |
| 2 | 123 | A |
| 31 | 143 | A |
| 4 | 54 | A |
| 3 | 123 | A |
| 5 | 4 | A |
| 11 | 1235 | B |
| 23 | 2 | B |
| 12 | 879 | B |
| 2 | 132 | B |
| 31 | 1 | B |
| 4 | 124 | B |
| 43 | 32 | B |
| 6 | 442 | B |
| 25 | 615 | B |
| 12 | 235 | B |



1 N00001 555 . C T 941.40 . AC=21;AF=0.228;AN=92;BaseQRankSum=0.431;ClippingRankSum=0.00;DP=222;ExcessHet=0.000;FS=0.000;InbreedingCoeff=0.5692;MLEAC=19;MLEAF=0.207;MQ=29.01;MQRankSum=0.00;QD=20.03;ReadPosRankSum=0.00;SOR=3.611 GT:AD:DP:GQ:PGT:PID:PL 1/1:0,2:2:6:...:55,6,0 1/1:0,2:2:6:...:58,6,0 1/1:0,2:2:6:...:58,6,0 0/0:3,0:3:9:...:0,9,84 0/0:7,0:7:21:...:0,21,194 0/0:2,0:2:6:...:0,6,54 0/0:9,0:9:27:...:0,27,250 0/0:9,0:9:27:...:0,27,250 ./:2,0:2:...:0,0,0 0/0:5,0:5:15:...:0,15,137 0/0:5,0:5:15:...:0,15,135 0/0:8,0:8:24:...:0,24,221 0/0:6,0:6:18:1:...:0,18,165 0/1:3,5:8:94:0|1:555_C_T:112,0,94 1/1:0,7:7:21:...:198,21,0 0/0:4,0:4:12:...:0,12,108 0/1:1,2:3:36:...:0,49,0,36 0/0:4,0:4:12:...:0,12,109 0/0:4,0:4:12:...:0,12,111 0/0:7,0:7:21:...:0,21,194 0/0:3,0:3:9:...:0,9,83 0/0:6,0:6:18:...:0,18,161 0/0:6,0:6:18:...:0,18,167 0/0:8,0:8:24:...:0,24,225 0/0:9,0:9:24:...:0,24,360 0/0:2,0:2:6:...:0,6,53 1/1:0,8:8:24:...:228,24,0 0/0:1,0:1:3:...:0,3,27 0/0:14,0:14:39:...:0,39,585 1/1:0,2:2:6:...:57,6,0 1/1:0,4:4:12:...:120,12,0 0/0:2,0:2:6:...:0,6,58 0/0:3,0:3:9:...:0,9,82 0/0:5,0:5:15:...:0,15,141 0/1:2,2:4:44:...:44,0,78 0/0:5,0:5:15:...:0,15,139 0/0:4,0:4:12:...:0,12,98 0/0:5,0:5:15:...:0,15,139 0/0:6,0:6:18:...:0,18,166 0/0:1,0:1:3:...:0,3,28 0/0:5,0:5:15:...:0,15,138 1/1:0,3:3:9:...:85,9,0 0/0:3,0:3:0:3:0:...:0,31 0/0:3,0:3:9:...:0,9,85 1/1:0,2:2:6:...:58,6,0 0/0:6,0:6:15:...:0,15,225 2 N00001 564 . C T 2389.72 . AC=34;AF=0.370;AN=92;BaseQRankSum=0.444;ClippingRankSum=0.00;DP=255;ExcessHet=1.4812;FS=4.461;InbreedingCoeff=-0.0155;MLEAC=39;MLEAF=0.424;MQ=30.27;MQRankSum=0.00;QD=14.75;ReadPosRankSum=-4.500e-02;SOR=2.647 GT:AD:DP:GQ:PGT:PID:PL 0/0:13,0:3:0:...:0,45 0/0:2,0:2:6:...:0,6,57 0/0:6,0:6:18:...:0,18,145 1/1:0,4:4:12:...:115,12,0 1/1:0,7:7:21:...:201,21,0 ./:2,0:2:...:0,0,0 0/1:5,4:9:88:...:88,0,110 0/1:4,5:9:84:...:117,0,84 0/0:4,0:4:12:...:0,12,99 0/0:5,0:5:15:...:0,15,137 0/0:6,0:6:0:...:0,0,113 0/1:4,5:9:84:...:118,0,84 0/1:6,3:9:59:...:59,0,140 0/1:5,2:7:36:1|0:555_C_T:36,0,204 0/0:8,0:8:24:...:0,24,227 1/1:0,3:3:9:...:85,9,0 0/1:2,2:4:45:...:45,0,78 0/1:1,2:3:18:...:47,0,18 0/0:5,0:5:15:...:0,15,139 0/1:4,3:7:60:...:60,0,91 0/1:3,2:5:42:...:42,0,71 0/0:6,0:6:18:...:0,18,161 0/1:4,2:6:37:...:37,0,94 0/0:9,0:9:0:...:0,0,146 1/1:0,9:9:27:...:237,27,0 0/0:2,0:2:6:...:0,6,53 0/0:10,0:10:30:...:0,30,277 0/0:1,0:1:3:...:0,3,27 0/0:1:7,8:15:99:...:183,0,152 0/0:3,0:3:0:...:0,28 0/0:4,0:4:12:...:0,12,111 0/1:2,3:5:43:...:72,43 1/1:0,7:7:21:...:201,21,0 0/1:2,3:5:42:...:69,0,42 0/1:2,2:4:43:...:43,0,78 1/1:0,6:6:18:...:172,18,0 1/1:0,4:4:12:...:115,12,0 0/0:5,0:5:12:...:0,12,180 0/1:3,4:7:64:...:75,0,64 0/0:1,0:1:3:...:0,3,28 0/1:2,3:5:41:...:68,0,41 0/0:4,0:4:0:...:0,0,53 0/1:1,2:3:20:...:47,0,20 0/1:1,3:4:17:...:75,0,17 0/0:3,0:3:9:...:0,9,77 0/0:4,0:4:0:...:0,0,58 1/1:0,6:6:18:...:169,18,0 3 N00001 636 . A T 506.60 . AC=4;AF=0.043;AN=94;BaseQRankSum=0.00;ClippingRankSum=0.00;DP=482;ExcessHet=3.2983;FS=0.000;InbreedingCoeff=-0.0427;MLEAC=4;MLEAF=0.043;MQ=32.87;MQRankSum=1.38;QD=14.90;ReadPosRankSum=0.319;SOR=1.329 GT:AD:DP:GQ:PL 0/0:7,0:7:21:0,21,192 0/0:6,0:6:18:0,18,166 0/0:11,0:11:33:0,33,316 0/0:8,0:8:24:0,24,224 0/0:13,0:13:36:0,36,540 0/0:5,0:5:15:0,15,140 0/0:20,0:20:57:0,57,855 0/1:7,7:14:99:162,0,163 0/0:8,0:8:24:0,24,231 0/0:6,0:6:18:0,18,169 0/0:15,0:15:45:0,45,413 0/0:16,0:16:48:0,48,446 0/0:16,0:16:48:0,48,447 0/0:7,0:7:18:0,18,270 0/0:17,0:17:51:0,51,459 0/0:3,0:3:9:0,9,81 0/0:7,0:7:21:0,21,196 0/0:8,0:8:24:0,24,207 0/0:8,0:8:24:0,24,227 0/0:11,0:11:30:0,30,450 0/1:2,11:13:19:279,0,19 0/0:12,0:12:36:0,36,327 0/0:10,0:10:30:0,30,290 0/0:10,0:10:30:0,30,280 0/0:10,0:10:27:0,27,405 0/0:4,0:4:12:0,12,110 0/0:17,0:17:51:0,51,460 0/1:3,1:4:16:16,0,54 0/0:24,0:24:72:0,72,667 0/0:6,0:6:18:0,18,170 0/0:16,0:16:48:0,48,461 0/0:7,0:7:21:0,21,195 0/0:13,0:13:39:0,39,367 0/0:7,0:7:18:0,18,270 0/0:6,0:6:18:0,18,171 0/0:11,0:11:33:0,33,313 0/0:5,0:5:15:0,15,140 0/0:8,0:8:24:0,24,223 0/0:12,0:12:36:0,36,336 0/0:7,0:7:21:0,21,198 0/0:10,0:10:30:0,30,284 0/0:13,0:13:39:0,39,364 0/0:11,0:11:33:0,33,312 0/0:9,0:9:27:0,27,254 4 N00001 648 . A G 6378.34 . AC=51;AF=0.543;AN=94;BaseQRankSum=-9.570e-01;ClippingRankSum=0.00;DP=521;ExcessHet=1.4574;FS=0.000;InbreedingCoeff=0.0661;MLEAC=52;MLEAF=0.553;MQ=32.36;MQRankSum=0.114;QD=14.97;ReadPosRankSum=-1.910e-01;SOR=1.474 GT:AD:DP:GQ:PL 0/0:8,0:8:24:222,24,0 1/1:0,14:14:42:371,42,0 0/1:4,4:8:88:88,0,88 0/1:11,9:20:99:192,0,246 0/1:7,7:14:99:155,0,155 0/0:11,0:11:30:0,30,450 0/0:6,0:6:18:0,18,170 0/1:7,7:14:99:152,0,149 0/1:7,11:18:99:246,0,142 0/1:0,6:16:99:122,0,212 1/1:0,4:12:111,12,0 0/1:12,9:21:99:183,0,247 1/1:0,6:6:18:162,18,0 1/1:0,7:7:21:196,21,0 0/1:3,6:9:55:118,0,55 0/0:10,0:10:30:0,30,276 0/1:5,6:11:99:114,0,109 0/1:11,2:13:18:18,0,266 0/0:12,0:12:36:0,36,327 0/1:8,4:12:59:59,0,187 0/1:6,5:11:99:107,0,139 1/1:0,9:9:27:249,27,0 0/0:5,0:5:0:0,0,89 0/1:9,11:20:99:208,0,194 0/0:6,0:6:18:0,18,170 0/1:11,15:26:99:338,0,238 1/1:0,8:8:24:221,24,0 1/1:1,14:15:32:15,0 0/1:3,5:8:62:109,0,62 1/1:0,14:14:42:385,42,0 0/1:4,2:6:37:37,0,95 1/1:0,6:6:18:165,18,0 1/1:0,10:10:30:276,30,0 1/1:0,5:5:15:140,15,0 0/0:8,0:8:24:0,24,222 0/1:9,4:13:71:71,0,218 0/0:10,0:10:30:0,30,272 0/1:3,5:8:62:113,0,62 0/1:8,5:13:97:97,0,189 1/1:0,13:13:39:357,39,0 0/1:4,8:12:79:183,0,79 0/0:8,0:8:13:0,13,189 0/1:7,5:12:99:104,0,164 1/1:0,15:15:45:45,45,0 5 N00001 833 . T C 1408.78 . AC=5;AF=0.053;AN=94;BaseQRankSum=-2.020e-01;ClippingRankSum=0.00;DP=1114;ExcessHet=3.4957;FS=0.000;InbreedingCoeff=-0.0552;MLEAC=5;MLEAF=0.053;MQ=41.48;MQRankSum=0.831;QD=12.81;ReadPosRankSum=-2.690e-01;SOR=0.733 GT:AD:DP:GQ:PL 0/0:12,0:12:35:0,35,262 0/0:17,0:17:51:0,51,485 0/1:3,17:20:26:387,0,26 0/0:15,0:15:45:0,45,427 0/0:32,0:32:90:0,90,1350 0/0:19,0:19:57:0,57,526 0/0:27,0:27:69:0,69,1035 0/0:20,0:20:60:0,60,583 0/0:29,0:29:84:0,84,1260 0/0:12,0:12:36:0,36,339 0/0:32,0:32:87:0,87,1305 0/0:30,0:30:90:0,90,834 0/0:34,0:34:99:0,99,935 0/0:25,0:25:75:0,75,621 0/0:35,0:35:99:0,99,1485 0/0:27,0:27:80:0,80,685 0/0:24,0:24:63:0,63,945 0/0:23,0:23:63:0,63,945 0/0:24,0:24:72:0,72,657 0/0:27,0:27:81:0,81,759 0/0:32,0:32:90:0,90,901 0/0:21,0:21:60:0,60,586 0/0:30,0:30:90:0,90,834 0/0:30,0:30:90:0,90,845 0/0:28,0:28:81:0,81,711 0/0:19,0:19:57:0,57,550 0/1:13,14:27:99:292,0,289 0/0:19,0:19:57:0,57,519 0/0:34,0:34:99:0,99,101,895 0/0:16,0:16:23:0,23,392 0/0:28,0:28:81:0,81,738 0/0:22,0:22:63:0,63,596 0/0:13,0:13:39:0,39,368 0/0:27,0:27:81:0,81,752 0/0:15,0:15:42:0,42,630 0/0:24,0:24:69:0,69,1035 0/0:15,0:15:45:0,45,416 0/0:27,0:27:81:0,81,746 0/0:23,0:23:69:0,69,627 0/0:26,0:26:75:0,75,1125 0/0:27,0:27:81:0,81,739 0/1:8,15:23:99:347,0,152 0/0:17,0:17:48:0,48,720 0/1:10,12:22:99:266,0,211 0/0:20,0:20:60:0,60,512 0/1:9,9:18:29:100,200,184 0/0:27,0:27:81:0,81,712

Why programming

- After some size, hand is impossible!
- Reproducible
- Adaptable
- Tools available



Why python?

Why python?

Matlab

Java

Ruby

Perl

Fortran

C

R

C++

Why python?

+

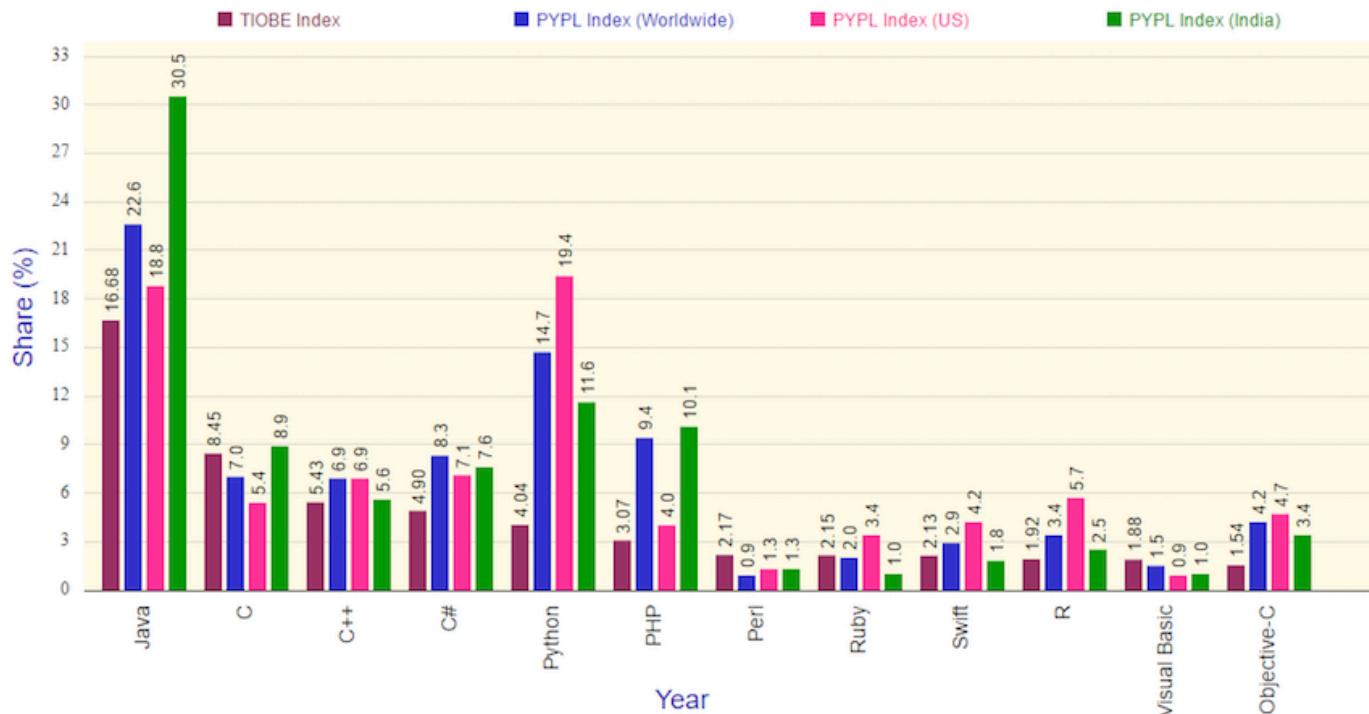
1. Modern
2. Widely used (in Science too!)
3. High level
4. Highly transferable skill
5. Free

Bonus: considered good for learning

-

1. Learning a new language is in investment
2. You might not need to master all languages!
3. Field tradition

Top Computer Languages (Feb 2017)



TIOBE Index

| Feb 2017 | Feb 2016 | Change | Programming language | Ratings | Change |
|----------|----------|--------|----------------------|----------|---------|
| 1 | 1 | | Java | 16.676 % | -4.47 % |
| 2 | 2 | | C | 8.445 % | -7.15 % |
| 3 | 3 | | C++ | 5.429 % | -1.48 % |
| 4 | 4 | | C# | 4.902 % | +0.50 % |
| 5 | 5 | | Python | 4.043 % | -0.14 % |
| 6 | 6 | | PHP | 3.072 % | +0.30 % |
| 7 | 9 | ↑ | JavaScript | 2.872 % | +0.67 % |
| 8 | 7 | ↓ | Visual Basic .NET | 2.824 % | +0.37 % |
| 9 | 10 | ↑ | Delphi/Object Pascal | 2.479 % | +0.32 % |
| 10 | 8 | ↓ | Perl | 2.171 % | -0.08 % |

PYPL Index (Worldwide)

| Feb 2017 | Change | Programming language | Share | Trends |
|----------|--------|----------------------|--------|--------|
| 1 | | Java | 22.6 % | -1.3 % |
| 2 | | Python | 14.7 % | +2.8 % |
| 3 | | PHP | 9.4 % | -1.2 % |
| 4 | | C# | 8.3 % | -0.3 % |
| 5 | ↑↑ | Javascript | 7.7 % | +0.4 % |
| 6 | | C | 7.0 % | -0.2 % |
| 7 | ↓↓ | C++ | 6.9 % | -0.6 % |
| 8 | | Objective-C | 4.2 % | -0.6 % |
| 9 | ↑ | R | 3.4 % | +0.4 % |
| 10 | ↓ | Swift | 2.9 % | +0.1 % |

Duck typed

```
#In C
```

```
int x;  
x = 4;
```

```
#In Python
```

```
x = 4
```

“If it walks like a duck if it quacks like a duck, let’s just assume it is a a duck”

```
# whatever is after # is not read,  
it is a comment  
  
x = 4 # x is an integer  
y = "4" # y is a string or  
z = 4. # is a float  
test = True # boolean: True/False
```

Some more

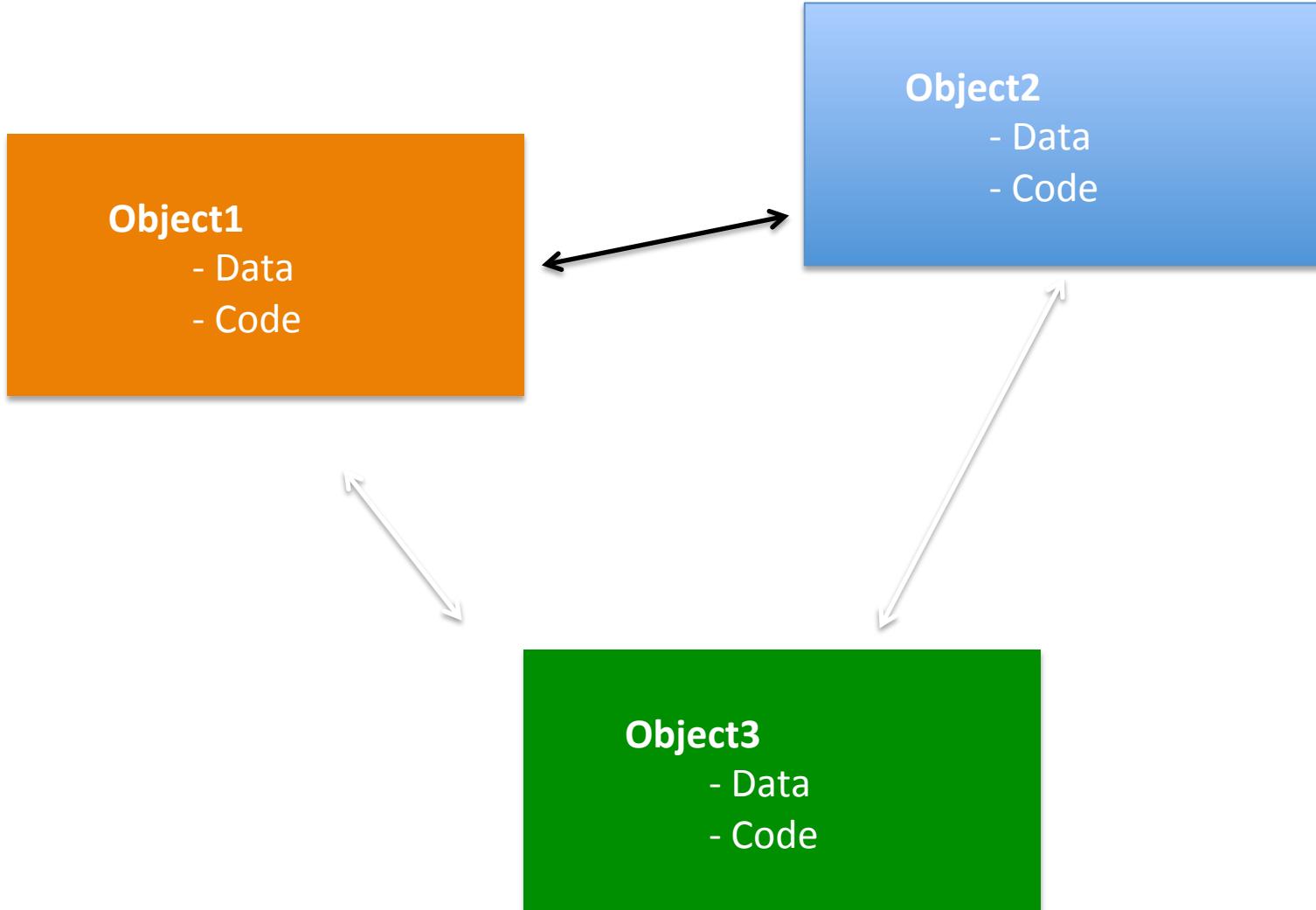
```
list_a = [4,'3'] # list
list_b = [[2,3],[1]] # list of list

set_d = set([1,2])

dict_e = {"days": ["Mon","Tue"], month :
["Jan"] }
```

Procedural programming

- Procedural programming focuses on the series of actions.
- It contains data and actions all mixed together.
- Very linear!
- C, Fortran



Indentation in python

```
if a == 2 and len(b) == 3: #check
    tab print a,b
else:
    print a

for i in range(10):
    tab print i,i*2,i**2,i.bit_length()
```

Objects

```
a = "awekjbcdba" # string
type(a) # print out the type of a, a is a
dir(a) #string
help(a.count()) # print help for the count
method
a.count("b") # count the b
a.split("j") # split a in a list
```

Indentation in python

```
def is_even(x): #create your own funct  
    return x%2 == 0
```

```
is_even(2)
```

```
#module  
import os # import the module os  
help(os) # list possibilities of os  
os.listdir() # function listdir of os
```

Today's goal

You play around and familiarise yourself with python

It might take a tiny bit of time to get everyone up and running

BONUS: You try to see if you can apply it to your own needs.

If you have never programmed

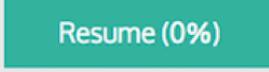
codecademy

Upgrade to Pro Learn

Dashboard

▼ Courses

 Learn Python
Currently on: Python Syntax.

 Resume (0%)
Reset Progress

Learn Python 0%
Learn the Command Line 0%

Codecademy.com



You need:

1. Windows Command Prompt

It will read python code if you type:
ipython

A screenshot of a Microsoft Windows XP Command Prompt window titled 'cmd.exe'. The window shows the following text:

```
c:\> C:\WINDOWS\System32\cmd.exe
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\Gay>
```

The cursor is located at the bottom left of the window.

2. A Text editor. You can open scripts there.
For example *wordpad*. All the scripts are
Textfiles that can be ran bit by bit or at once.
Today we want to run them bit by bit.

A screenshot of a Microsoft WordPad window titled 'NaiveBayes.py - WordPad'. The window shows a Python script named 'NaiveBayes.py'. The code is as follows:

```
import os
import sys

#####
# Execute
#####

def Execute (cmd):
    print "Exec: " + cmd
    os.system(cmd + " >> run.log")

#####
# Main
#####

# Check the number of command line arguments
if (len(sys.argv) != 2):
    print "Usage: NaiveBayes.py <basename>"
    sys.exit(1)
base = sys.argv[1]

# split the data file into training and test cases\
Execute("ITSC_Sample -c class -i " + base + ".arff -o trn.arff -t tst.arff -p 0.66")

# Train the classifier
Execute("ITSC_NaiveBayesTrain -c class -i trn.arff -b bayes.txt")

# Apply the classifier
Execute("ITSC_NaiveBayesApply -c class -i trn.arff -b bayes.txt -o res_trn.arff")
Execute("ITSC_NaiveBayesApply -c class -i tst.arff -b bayes.txt -o res_tst.arff")

# Evaluate the results
Execute("ITSC_Accuracy -c class -t res_trn.arff -v trn.arff -o acc_trn.txt")
Execute("ITSC_Accuracy -c class -t res_tst.arff -v tst.arff -o acc_tst.txt")
```

For Help, press F1

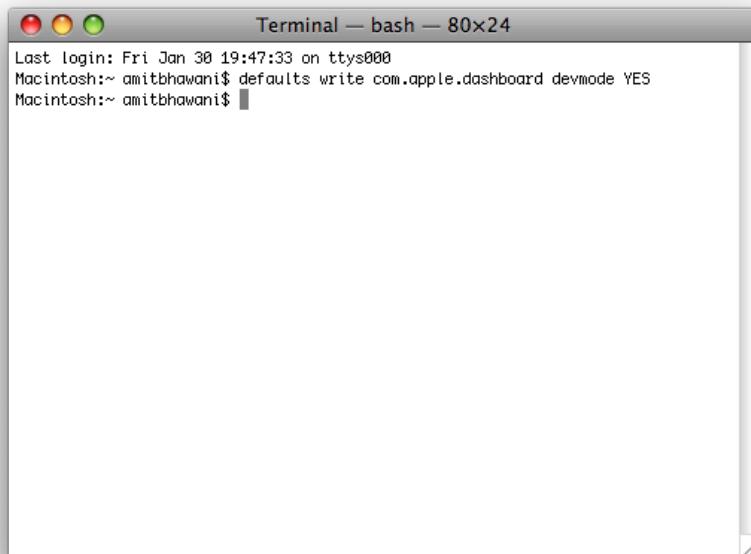


You need:



1. Terminal (in Applications/Utilities)

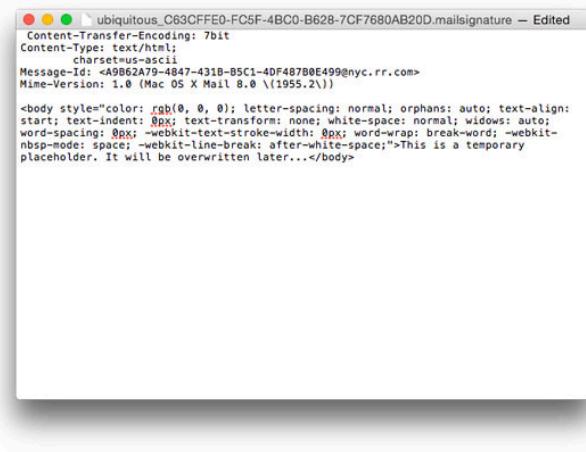
It will read python code if you type:
ipython



A screenshot of a Mac OS X Terminal window titled "Terminal — bash — 80x24". The window shows the following command being run and its output:

```
Last login: Fri Jan 30 19:47:33 on ttys000
Macintosh:~ amitbhawani$ defaults write com.apple.dashboard devmode YES
Macintosh:~ amitbhawani$
```

2. A Text editor. You can open scripts there.
For example *wordpad*. All the scripts are
Textfiles that can be ran bit by bit or at once.
Today we want to run them bit by bit.



A screenshot of a Mac OS X TextEdit window titled "ubiquitous_C63CFFE0-FC5F-4BC0-B628-7CF7680AB20D.mailsignature — Edited". The window contains the following HTML code:

```
Content-Transfer-Encoding: 7bit
Content-Type: text/html;
charset=us-ascii
Message-Id: <A9B62A79-4B47-431B-B5C1-40F487B0E499@nyc.rr.com>
Mime-Version: 1.0 (Mac OS X Mail 8.0 \1955.2\))

<body style="color: #000; font-family: sans-serif; margin: 0; padding: 0; font-size: 1em; line-height: 1.4; letter-spacing: 0.02em; word-spacing: 0.05em; -webkit-text-stroke-width: 0px; word-wrap: break-word; -webkit-nbsp-mode: space; -webkit-line-break: after-white-space;">This is a temporary
placeholder. It will be overwritten later...</body>
```

<http://tinyurl.com/k9ore9o>

- GettingStarted.pdf
 - How to open python on your own computer
- AdvancedBasics.py
 - data structures and python basics
- Summary_fasta.py
 - Output a brief summary from an input file
- data_and_plotting.py
 - process data using python and easy plotting
- resources.txt
 - Some links to a few nice resources