

[Notes from bird data analysis](#)

[Morning Day 2](#)

[Afternoon Day 2](#)

WINDOWS

(curses! Windows!)

This is for cmd.exe:

```
c:\anaconda\scripts\ipython notebook --pylab=inline
```

For GitBash, try typing this (first need to “cd ~/2013-04-az/notebooks”):

```
/c/anaconda/scripts/ipython notebook --pylab=inline
```

In **Windows & gitbash**, if your iPython notebook is not giving output (i.e. you shift-enter and nothing happens), try running through the Virtual Box rather than git bash OR, use cmd.exe (the windows command line rather than gitbash)

in VirtualBox, what do I need type in the cmd prompt to get ipython to work? the same
/c/anaconda/scripts/ipython notebook --pylab=inline? In Virtual Box, use the mac command
(below)

Mac OS X

```
~/anaconda/bin/ipython notebook --pylab=inline
```

Notes from bird data analysis

- look up list comprehension in python

```
%%file example-data.txt
```

```
0,0
```

```
1,2
```

```
2,4
```

```
3,8
```

```
4,16
```

```
5,32
```

```
6,64
```

%%file calc_gc.py

```
def calc_gc(dna):  
    gc = dna.count('G') + dna.count('C')  
    total = len(dna)  
    frac = gc / float(total)  
    return frac
```

```
print calc_gc('ATGC')
```

```
def test_1():  
    x = calc_gc('ATGC')  
    assert x == 0.5, x
```

```
def test_2():  
    x = calc_gc('AAAA')  
    assert x == 1.0, x
```

with uppercase

```
def calc_gc(dna):  
    dna = dna.upper()  
    gc = dna.count('G') + dna.count('C')  
    total = len(dna)  
    frac = gc / float(total)  
    return frac
```

pip install git+<https://github.com/ged-lab/screed.git>

In git bash,

/c/anaconda/scripts/pip install git+<https://github.com/ged-lab/screed.git>

On virtualbox: at command line,

```
sudo pip install git+https://github.com/ged-lab/screed.git
```

the password is 'swc

```
--
```

```
filename = '../data/25k.fq.gz'
```

```
for record in screed.open(filename):  
    print record.name  
    print record.sequence  
    break  
,
```

```
filename = '../data/25k.fq.gz'
```

```
n = 0
```

```
for record in screed.open(filename):  
    n += 1  
    if n > 10:  
        break  
    print record.name  
    print record.sequence
```

```
import calc_gc
```

```
gc_list = []  
for record in screed.open(filename):  
    seq = record.sequence  
    gc = calc_gc.calc_gc(seq)  
    gc_list.append(gc)
```

```
hist(gc_list)
```

Morning Day 2

```
birds_file = file('birds.csv','rU')
count = 0
for line in birds_file:
    print line
    if count > 20:
        break
    count +=1
```

C

<https://raw.githubusercontent.com/swcarpentry/2013-04-az/master/data/birds.csv>

In git-bash, you can run python using:

```
/c/anaconda/python
```

Over the break, please go to www.github.com and create an account. They're free!

```
birds_file = file('birds.csv','rU')
birds = {}
count = 0
birds_file.readline()
for line in birds_file:
    line_list = line.strip().split(',')
    bird_name = line_list[0]
    print bird_name
    if count > 20:
        break
    count += 1
```

```
=====
birds_file = file('birds.csv','rU')
birds = {}
count = 0
birds_file.readline()
for line in birds_file:
    line_list = line.strip().split(',')
    bird_name = line_list[0]
    birds[bird_name]=line_list[1:]
    if count > 20:
        break
    count += 1
```

```
for bird,data in birds.items():
    print bird, dat
```

Go to:

<https://github.com/swcarpentry/2013-04-az>=====

```
>>> L4
['0', '1', '0', '0', '0', '2', '2', '2', '7', '2', '7', '1']
>>> int('3')
3
>>> '3'+ '4'
'34'
>>> 3 + 4
7
>>> L5 = [int(x) for x in L4]
>>> L5
[0, 1, 0, 0, 0, 2, 2, 2, 7, 2, 7, 1]
>>> L4
['0', '1', '0', '0', '0', '2', '2', '2', '7', '2', '7', '1']
>>> L6 = []
>>> for x in L4:
... L6.append(int(x))
File "<stdin>", line 2
    L6.append(int(x))
    ^
IndentationError: expected an indented block
```

```
>>> for x in L4:
... L6.append(int(x))
...
>>> L6
[0, 1, 0, 0, 0, 2, 2, 2, 7, 2, 7, 1]
=====
```

```
birds_file = file('birds.csv','rU')
birds = {}
count = 0
birds_file.readline()
for line in birds_file:
    line_list = line.strip().split(',')
    bird_name = line_list[0]
    data = [int(x) for x in line_list[1:]]
    birds[bird_name]=data
```

```
if count > 20:
    break
count += 1
```

```
for bird,data in birds.items():
    print bird, data
```

```
=====
birds_file = file('birds.csv','rU')
birds = {}
count = 0
birds_file.readline()
for line in birds_file:
    line_list = line.strip().split(',')
    bird_name = line_list[0]
    data = [int(x) for x in line_list[1:]]
    birds[bird_name]=data
    if count > 20:
        break
    count += 1
```

```
for bird,data in birds.items():
    print bird, data
```

```
name = raw_input("Enter a bird: ")
print name
print birds[name]
```

```
year_str = raw_input("Enter a year: ")
year_int = int(year_str)
year_index = year_int - 2000
print birds[name][year_index]
```

```
=====
birds_file = file('birds.csv','rU')
birds = {}
count = 0
birds_file.readline()
for line in birds_file:
    line_list = line.strip().split(',')
    bird_name = line_list[0]
    data = [int(x) for x in line_list[1:]]
    birds[bird_name]=data
```

```

    if count > 20:
        break
    count += 1

for bird,data in birds.items():
    print bird, data

#name = raw_input("Enter a bird: ")
#print name
#print birds[name]

year_str = raw_input("Enter a year: ")
year_int = int(year_str)
year_index = year_int - 2000
#print birds[name][year_index]

for bird,data in birds.items():
    print "%30s %d"% (bird, data[year_index])

=====
# Final version

birds_file = file('birds.csv','rU')
birds = {}
count = 0
birds_file.readline()
for line in birds_file:
    line_list = line.strip().split(',')
    bird_name = line_list[0]
    data = [int(x) for x in line_list[1:]]
    birds[bird_name]=data
    if count > 20:
        break
    count += 1

for bird,data in birds.items():
    print bird, data

#name = raw_input("Enter a bird: ")
#print name
#print birds[name]

year_str = raw_input("Enter a year: ")

```

```

year_int = int(year_str)
year_index = year_int - 2000
#print birds[name][year_index]

for bird,data in birds.items():
    #print "%30s %d"% (bird, data[year_index])
    print "{:30s} {:d}".format(bird, data[year_index])

```

Afternoon Day 2

git, scripting in python and bash

Software carpentry github repo: <https://github.com/swcarpentry/2013-04-az/>

Original script:

```

import sys
import screed
4 filename = sys.argv[1]
5 print filename
6
7 for record in screed.open(filename):
8     print record
9     break
10

```

Activity:

1. read the sequence file and print out the first 10 DNA sequences

Helpful links (for this section in general)

- [git cheat sheet](#)
- [bash scripting guide](#)
- [screed documentation](#)