EEB 177 Lecture 11

Tuesday Feb 18th, 2020

Topics

- functions
- debugging

Office Hours

Today 1045-1145

Download this data file

use `curl`

```
curl -L -o fishdata.csv  http://bit.ly/fishlength
```

Write a block of code that will open the file "fish_data.csv"

1. Extract records for the family Labridae
2. Calculate the average length of the family
3. Write the Labridae records to a new csv file.
4. (Bonus) Print the length of the largest species in each family

Functions

Functions are blocks of code that do something useful. You have already used many built-in functions of strings (**str.lower()**, **file.write()**).

Functions take **arguments**. You define functions in python with the **def** keyword.

```
def stringAnalyzer(ss):
    length = len(ss)
    print("This string has {} characters".format(length))

tt = stringAnalyzer("hello")
This string has 5 characters
```

This function expects a string, counts the characters in it, and prints the total length. No value is returned in this case.

Returning values

More commonly your function will do something to the arguments you provide and return the modified values.

Challenge

Write a function that will return the AT content of a DNA sequence.

```
def get_at_content(dna):
length = len(dna)
    a_count = dna.count('A')
    t_count = dna.count('T')
    at_content = (a_count + t_count) / length
    return at_content
```

Variable scope

Variables created within functions cannot be used outside of that function.

```
def get_at_content(dna):
    length = len(dna)
    a_count = dna.count('A')
    t_count = dna.count('T')
    at_content = (a_count + t_count) / length
    return at_content

print(a_count)
---------------------------------------------------------------
NameError                                 Traceback (most r
```

Improving functions

One advantage to functions is that they allow you edit and improve your code in a modular way. Here we will modify our code to avoid long decimals and an incorrect AT content when lowercase values are passed to the function **For the these examples we want our function to work correctly on the following four lines.**

```
#we want our function to properly execute these lines
my_at_content = get_at_content("ATGCGCGATCGATCGAATCG")
print(str(my_at_content))
print(get_at_content("ATGCATGCAACTGTAGC"))
print(get_at_content("aactgtagctagctagcagcgta"))
```

unimproved version

```
def get_at_content(dna):
    length = len(dna)
    a_count = dna.count('A')
    t_count = dna.count('T')
    at_content = (a_count + t_count) / length
    return at_content
```

this gives some problems

```
 print(get_at_content("ATGCATGCAACTGTAGC"))
0.529411764706
 print(get_at_content("aactgtagctagctagcagcgta"))
0.0
```

Improved version

```
def get_at_content(dna):
    length = len(dna)
    a_count = dna.upper().count('A')
    t_count = dna.upper().count('T')
    at_content = (a_count + t_count) / length
    return round(at_content, 2)

my_at_content = get_at_content("ATGCGCGATCGATCGAATCG")
print(str(my_at_content))
print(get_at_content("ATGCATGCAACTGTAGC"))
print(get_at_content("aactgtagctagctagcagcgta"))
```

Better! Here we have added upper() and round() methods to format values before they are returned. What if we wanted the number of decimal places to be an argument in our function so that the user could specify the decimal length?

Modify your function to do this Does it work?

Here is one way to control the number of significant figures

```
def get_at_content(dna, sig_figs):
    length = len(dna)
    a_count = dna.upper().count('A')
    t_count = dna.upper().count('T')
    at_content = (a_count + t_count) / length
    return round(at_content, sig_figs)
```

Encapsulation

OK you have made changes to your function to improve formatting. What you may not have noticed is that you did not need to modify (much) the rest of the program outside the function. This demonstrates the idea of **encapsulation** which really just mean breaking up a big program in to smaller parts that can be managed in a straightforward way.
Keep this idea in mind as you start to consider you final projects!

Keyword arguments

If you define a function in the typical way, the order that you pass the arguments is critical.

```
In [62]: def outOfOrder(number, pet):
    ....:         print "I own {} {}s".format(number, pet)
    ....:
In [63]: outOfOrder(3, "dog")
I own 3 dogs
In [64]: outOfOrder("dog", 3)
I own dog 3s
```

Keyword arguments, wherein the name of the argument followed by an "=" followed by the value, can help avoid these kinds of errors.

```
outOfOrder(number=3, pet="dog")
#same as
outOfOrder(pet="dog", number=3)
```

Default values

It is often useful to pass a default value to a function that can be overridden.

```
def get_at_content(dna, sig_figs=2):
    length = len(dna)
    a_count = dna.upper().count('A')
    t_count = dna.upper().count('T')
    at_content = (a_count + t_count) / length
    return round(at_content, sig_figs)

get_at_content("ATCGTGACTCG")
get_at_content("ATCGTGACTCG", 3)
get_at_content("ATCGTGACTCG", sig_figs=4)
```

Now the value will be written to two decimal places unless otherwise specified.

Testing functions

It is important that your code is behaving the way you expect it will. One way to do this is to invoke functions with arguments that should produce a known value. You can use the **assert** function to test if your code passes this test.

```
assert get_at_content("ATGC") == 0.5 #this should work
assert get_at_content("ATGCNNNNNNNNNN") == 0.5 #what about
```

Modify your get_at_content() program so that you function will pass an assert with "Ns"

Improved code

```
def get_at_content(dna, sig_figs=2):
    dna = dna.replace('N', '')
    length = len(dna)
    a_count = dna.upper().count('A')
    t_count = dna.upper().count('T')
    at_content = (a_count + t_count) / length
    return round(at_content, sig_figs)
```

Now the function passes the assert test.

It is a good idea to include a series of assert tests for functions you creates. You can easily comment the assert statements out as needed.

```
assert get_at_content("A") == 1
assert get_at_content("G") == 0
assert get_at_content("ATGC") == 0.5
assert get_at_content("AGG") == 0.33
assert get_at_content("AGG", 1) == 0.3
assert get_at_content("AGG", 5) == 0.33333
```

Debugging

the implementation of the python debugger (**pdb**)Ipython provides another way of searching out bugs. Lets examine it. Start with this function.

```
def createabug(x):
    y = x**4
    z = 0.
    y = y/z
    return y

createabug(5)
```

What happens?

Use your text editor to save this file as *debugme.py* in your classwork directory, Load and execute your function:

```
In [1]: %run debugme.py
...
----> 4      y = y/z
      5      return y
      6

ZeroDivisionError: float division by zero #doesn't work...
```

Now start the debugger in ipython and run again:

```
In [72]: %pdb
Automatic pdb calling has been turned ON
In [73]: %run debugme.py
...
ZeroDivisionError: float division by zero
> /home/vagrant/scripts/debugme.py(4)createabug()
      3     z = 0.
----> 4     y = y/z
      5     return y

ipdb> # <---- NOTICE THIS!
```

Now we are in the debugger shell!

pdb commands

Within the debugger we can move around in our code and examine variables to see what is happening.

- **n** move to the next line.
- **ENTER** repeat the previous command.
- **s** "step" into function or procedure (i.e., continue the debugging inside the function, as opposed to simply run it). p x print variable x.
- **c** continue until next break-point.
- **q** quit
- **l** print the code surrounding the current position. r continue until the end of the function.

try out pdb

```
ipdb> p x
25
ipdb> p y
390625
ipdb> pz
*** NameError: name 'pz' is not defined
ipdb> px
*** NameError: name 'px' is not defined
ipdb> p x
25
ipdb> p y
390625
ipdb> p z
0.0
ipdb> p y/z
*** ZeroDivisionError: ZeroDivisionError('float division by
ipdb> q
In [74]: pdb
Automatic pdb calling has been turned OFF
```

Automatically launch pdb

If you want to start pdb within a longer script, place this line at the point you want to examine:

```
import pdb;
pdb.set_trace()
```