

Bi 410/510: Introduction to Programming for Biologists

Spring 2018

Project 4: Lists and Loops

To complete this project you need to write five small Python programs, each similar to the GPA calculator and other programs used to introduce `for` loops in lectures. Put all five files in a single folder named `lists`, compress the folder, and submit the resulting zip file to Canvas.

```
$ zip -r lists.zip lists
```

Note: The examples for each program show the results of calculations. You do not have to print your results in the same format – we'll be looking for correct results and good style but not the same exact output format.

Mean String Length

Write a program named `mean_length.py` that will print the average (mean) length of the strings passed to the program on the command line.

In this first example, there are 15 letters in 4 names, so the mean is 3.75:

```
$ python mean_length.py Luke Han Leia Kylo
average length: 3.75
```

Here there are 24 letters in 5 names for a mean of 4.8:

```
$ python mean_length.py Luke Han Leia Kylo Chewbacca
average length: 4.8
```

Fibonacci Sequence

Write a program named `fibonacci.py` that will generate the numbers in the Fibonacci sequence. Pass the length of the sequence you want to create on the command line.

Examples:

```
$ python fibonacci.py 7
[1, 1, 2, 3, 5, 8, 13]
```

```
$ python fibonacci.py 30
[1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ... 317811, 514229, 832040]
```

```
$ python fibonacci.py 1
error: sequence length must be 2 or longer
```

DNA Similarity

Write a program named `similarity.py` that computes the similarity of two DNA sequences. Check to make sure the sequences are the same length, then compute the similarity as the percentage of locations that are the same in each sequence. If the sequences are identical similarity is 100%, if they differ at every location they are 0% similar.

Examples:

```
$ python similarity.py GATTACA GTTTACA
85.71%
```

```
$ python similarity.py GATTACA CTAATGT
0.00%
```

```
$ python similarity.py GATTACA ATT
sequences must be the same length
```

Disease Model

Suppose researchers have analyzed data from the spread of a particular disease and have fit the data to the following model, which predicts the number of people in a small population who will have the disease on day t :

$$p(t) = \frac{20t^3 - t^4}{1000}$$

Write a program named `disease.py` that will compute $p(t)$ for 20 days (*i.e.* for $1 \leq t \leq 20$) and print the day at which the most people are infected and the number of people infected on that day. There are no command line options, just evaluate the model 20 times and find the maximum.

```
$ python disease.py
Max = 16.875 at day 15
```

Codons

Write a program named `codons.py` that will print the codons (3-letter substrings) in a DNA sequence specified on the command line:

```
$ python codons.py ATGCTCAAGTAG
ATG
CTC
AAG
TAG
```

If the length of the sequence is not a multiple of 3 the last codon will have less than 3 letters:

```
$ python codons.py GATTACA
GAT
TAC
A
```