

## Python 7 Problem Set

1. Create a custom function to format a string of DNA to be a series of lines with a max length of 60.
  - INPUT: a string of DNA
  - RETURNS: a string of DNA with lines no more than 60 nucleotides long

```
1 INPUT :
2 DNA=" GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTTCTCAAAAGTCTAGAGCCAC
3
4 OUTPUT :
5 GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTT
6 CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTT
7 GCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGCCA
8 GACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCCCC
9 TCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCCTGAAAACAACGTTCTGTC
10 CCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAATG
11 GTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCCGT
12 GGCCCCTGCACCAGCAGCTCCTACACGGCGGCCCTGCACCAGCCCCCTCCTGGCCCCT
13 GTCATCTTCT
```

2. Modify your function so that it does not matter if the DNA string does or does not have newlines.

```
1 INPUT :
2 GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTTCTCAAAAGTCTAGAGCCACC
3 GTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACG
4 CTTCCCTGGATTGGCAGCCAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCCCC
5 TCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCCTGAAAACAACGTTCTGTCCCCCTTGCCGTCCCAAGCAA
6 TGGATGATTTGATGCTGTCCCCGGACGATATTGAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATG
7 CCAGAGGCTGCTCCCCCGTGGCCCCTGCACCAGCAGCTCCTACACGGCGGCCCTGCACCAGCCCCCTCCTGGCCCCT
8 GTCATCTTCT
9
10
11 OUTPUT :
12 GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTT
13 CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTT
14 GCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGCCA
15 GACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCCCC
16 TCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCCTGAAAACAACGTTCTGTC
17 CCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAATG
18 GTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCCGT
19 GGCCCCTGCACCAGCAGCTCCTACACGGCGGCCCTGCACCAGCCCCCTCCTGGCCCCT
20 GTCATCTTCT
```

3. Modify your function so that it takes two arguments, the DNA string and the max length of each line.

```

1 INPUT :
2 LEN=80
3 DNA=" GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTTCTCAAAAGTCTAGAGCCAC
4
5 OUTPUT :
6 GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTTCTCAAAAGTCTAGAGCCACC
7 GTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACG
8 CTTCCCTGGATTGGCAGCCAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCCCC
9 TCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCCTGAAAACAACGTTCTGTCCCCCTTGCCGTCCCAAGCAA
10 TGGATGATTTGATGCTGTCCCCGGACGATATTGAACAATGGTTCAGTGAAGACCCAGGTCCAGATGAAGCTCCAGAATG
11 CCAGAGGCTGCTCCCCCGTGGCCCCTGCACCAGCAGCTCCTACACGGCGGCCCTGCACCAGCCCCCTCCTGGCCCCT
12 GTCATCTTCT

```

4. Modify your script so that it can take two commandline arguments:

- 1) FASTA file name
- 2) Max length of each line

The script should reformat every sequence in the file to the specified max line length. Make sure your output is in proper FASTA format.