

wk2 assignment

September 28, 2020

```
[2]: longSeq =   
      ↪ ""ATGGAGAGCCTTGTTCCTTGGTGTCAACGAGAAAAACACAGTCCAACCTCAGTTTGCCTGTCCTTCAGGTT  
      AGAGACGTGCTAGTGCCTGGCTTCGGGACTCTGTGGAAGAGGCCCTATCGGAGGCACGTGAACACCTCAA  
      AAATGGCACTTGTGGTCTAGTAGAGCTGAAAAAGGCGTACTGCCCCAGCTTGAACAGCCCTATGTGTTT  
      ATTAAACGTTCTGATGCCTTAAGCACCAATCACGGCCACAAGGTCGTTGAGCTGGTTGCAGAAATGGACG  
      GCATTACGTACGGTCGTAGCGGTATAACACTGGGAGTACTCGTGCCACATGTGGGCGAAACCCCAATTGC  
      ATACCGCAATGTTCTTCTTCGTAAGAACGGTAATAAGGGAGCCGGTGGTCATAGCTATGGCATCGATCTA  
      AAGTCTTATGACTTAGGTGACGAGCTTGGCACTGATCCCATTGAAGATTATGAACAAAACCTGGAACACTA  
      AGCATGGCAGTGGTGCCTCCGTGAACTCACTCGTGAGCTCAATGGAGGTGCAGTCACTCGCTATGTCGA  
      CAACAATTTCTGTGGCCCAGATGGGTACCCTCTTGATTGCATCAAAGATTTTCTCGCACGCGCGGGCAAG  
      TCAATGTGCACTCTTTCGGAACAACTTGATTACATCGAGTCGAAGAGAGGTGTCTACTGCTGCCGTGACC  
      ATGAGCATGAAATTGCCTGGTTCACTGAGCGCTCTGATAAGAGCTACGAGCACCAGACACCCTTCGAAAT  
      TAAGAGTGCCAAGAAATTTGACACTTTCAAAGGGGAATGCCCAAAGTTTGTGTTTCTCTTAACTCAAAA  
      GTCAAAGTCATTCAACCACGTGTTGAAAAAGAAAAGACTGAGGGTTTCATGGGGCGTATACGCTCTGTGT  
      ACCCTGTTGCATCTCCACAGGAGTGAACAATATGCACCTGTCTACCTTGATGAAATGTAATCATTGCGA  
      TGAAGTTTCATGGCAGACGTGCGACTTTCTGAAAGCCACTTGTGAACATTGTGGCACTGAAAATTTAGTT  
      ATTGAAGGACCTACTACATGTGGGTACCTACCTACTAATGCTGTAGTGAAAATGCCATGTCCTGCCTGTC  
      AAGACCCAGAGATTGGACCTGAGCATAGTGTTCAGATTATCACAACCACTCAAAACATTGAAACTCGACT  
      CCGCAAGGGAGGTAGGACTAGATGTTTTGGAGGCTGTGTGTTTGCCTATGTTGGCTGCTATAATAAGCGT  
      GCCTACTGGGTTCTCTGCTAGTGCTGATATTGGCTCAGGCCATACTGGCATTACTGGTGACAATGTGG  
      AGACCTTGAATGAGGATCTCCTTGAGATACTGAGTCGTGAACGTGTTAACATTAAACATTGTTGGCGATTT  
      TCATTTGAATGAAGAGGTTGCCATCATTTTGGCATCTTTCTCTGCTTCTACAAGTGCCTTTATTGACACT  
      ATAAAGAGTCTTGATTACAAGTCTTCAAACCACTTGTGAGTCCTGCGTAACCTATAAAGTTACCAAGG  
      GAAAGCCCGTAAAAGGTGCTTGGAACATTGGACAACAGAGATCAGTTTAAACCACTGTGTGGTTTTCC  
      CTCACAGGCTGCTGGTGTTATCAGATCAATTTTTGCGCGCACACTTGATGCAGCAAACCACTCAATTCCT  
      GATTTGCAAAGAGCAGCTGTCACCATACTTGATGGTATTTCTGAACAGTCATTACGTCTTGTGACGCCA  
      TGGTTTATACTTCAGACCTGCTCACCAACAGTGTCAATTATTATGGCATATGTAACCTGGTGGTCTTGTACA  
      ACAGACTTCTCAGTGGTTGTCTAATCTTTGGGCACTACTGTTGAAAAACCTCAGGCCTATCTTTGAATGG  
      ATTGAGGCGAAACCTTAGTGCAGGAGTTGAATTTCTCAAGGATGCTTGGGAGATTCTCAAATTTCTCATT  
      CAGGTGTTTTTACATCGTCAAGGGTCAAATACAGGTTGCTTCAGATAACATCAAGGATTGTGTAAATG  
      CTTCAATTGATGTTGTTAAACAAGGCACTCGAAATGTGCATTGATCAAGTCACTATCGCTGGCGCAAAGTTG  
      CGATCACTCAACTTAGGTGAAGTCTTCATCGCTCAAAGCAAGGGACTTTACCGTCAGTGTATACGTGGCA  
      AGGAGCAGCTGCAACTACTCATGCCTCTTAAGGCACCAAAAGAAGTAACCTTTCTTGAAGGTGATTACAA  
      TGACACAGTACTTACCTCTGAGGAGGTTGTTCTCAAGAACGGTGAACCTCGAAGCACTCGAGACGCCCGTT  
      GATAGCTTCACAAATGGAGCTATCGTTGGCACACCAGTCTGTGTAAATGGCCTCATGCTCTTAGAGATTA  
      AGGACAAAGAACAATACTGCGCATTGTCTCCTGGTTTACTGGCTACAAACAATGTCTTTGCTTAAAGG  
      GGGTGCACCAATTAAAGGTGTAACCTTTGGAGAAGATACTGTTTGGGAAGTTCAAGGTTACAAGAATGTG
```

```
AGAATCACATTTGAGCTTGATGAACGTGTTGACAAAGTGCTTAATGAAAAGTGCTCTGTCTACACTGTTG
AATCCGGTACCGAAGTTACTGAGTTTGCATGTGTTGTAGCAGAGGCTGTTGTGAAGACTTTACAACCAGT
TTCTGATCTCCTTACCAACATGGGTATTGATCTTGATGAGTGGAGTGTAGCTACATTCTACTTATTTGAT
GATGCTGGTGAAGAAAACTTTTTCATCACGTATGTATTGTTTCCTTTTACCCTCCAGATGAGGAAGAAGAGG
ACGATGCAGAGTGTGAGGAAGAAGAAATTGATGAAACCTGTGAACATGAGTACGGTACAGAGGATGATTA
TCAAGGTCTCCCTCTGGAATTTGGTGCCTCAGCTGAAACAGTTCGAGTTGAGGAAGAAGAAGAGGAAGAC
TGGCTGGATGATACTACTGAGCAATCAGAGATTGAGCCAGAACCAGAACCTACACCTGAAGAACCAGTTA
ATCAGTTTACTGGTTATTTAAAACTTACTGACAATGTTGCCATTAAATGTGTTGACATCGTTAAGGAGGC"""
```

[3]: *#1. Define 'shorterSeq' as is in this lesson.*

```
shorterSeq = longSeq[2500:]
print(shorterSeq)
```

```
AGGTGTAACCTTTGGAGAAGATACTGTTTGGGAAGTTCAAGGTTACAAGAATGTG
AGAATCACATTTGAGCTTGATGAACGTGTTGACAAAGTGCTTAATGAAAAGTGCTCTGTCTACACTGTTG
AATCCGGTACCGAAGTTACTGAGTTTGCATGTGTTGTAGCAGAGGCTGTTGTGAAGACTTTACAACCAGT
TTCTGATCTCCTTACCAACATGGGTATTGATCTTGATGAGTGGAGTGTAGCTACATTCTACTTATTTGAT
GATGCTGGTGAAGAAAACTTTTTCATCACGTATGTATTGTTTCCTTTTACCCTCCAGATGAGGAAGAAGAGG
ACGATGCAGAGTGTGAGGAAGAAGAAATTGATGAAACCTGTGAACATGAGTACGGTACAGAGGATGATTA
TCAAGGTCTCCCTCTGGAATTTGGTGCCTCAGCTGAAACAGTTCGAGTTGAGGAAGAAGAAGAGGAAGAC
TGGCTGGATGATACTACTGAGCAATCAGAGATTGAGCCAGAACCAGAACCTACACCTGAAGAACCAGTTA
ATCAGTTTACTGGTTATTTAAAACTTACTGACAATGTTGCCATTAAATGTGTTGACATCGTTAAGGAGGC
```

[16]: *#2. Get it's length*

```
shortSeq_len = len(shorterSeq)
print(shortSeq_len)
```

623

[17]: *# 3. Add the counts of each nucleotide.*

```
sum_shorterSeq = shorterSeq.count('A') + shorterSeq.count('T') + shorterSeq.
    ↪count('C') + shorterSeq.count('G')
print(sum_shorterSeq)
```

615

[18]: *# 4. Compare the output for steps 2 and 3 to see if they are equal.*
NO

[12]: *#5. IMPLEMENT THE STRING METHOD THAT WILL REMOVE THE NEWLINES.*

```
strip_newline_seq = shorterSeq.split()
print(strip_newline_seq)
```

```
['AGGTGTAACCTTTGGAGAAGATACTGTTTGGGAAGTTCAAGGTTACAAGAATGTG',
'AGAATCACATTTGAGCTTGATGAACGTGTTGACAAAGTGCTTAATGAAAAGTGCTCTGTCTACACTGTTG',
'AATCCGGTACCGAAGTTACTGAGTTTGCATGTGTTGTAGCAGAGGCTGTTGTGAAGACTTTACAACCAGT',
'TTCTGATCTCCTTACCAACATGGGTATTGATCTTGATGAGTGGAGTGTAGCTACATTCTACTTATTTGAT',
'GATGCTGGTGAAGAAAACTTTTTCATCACGTATGTATTGTTTCCTTTTACCCTCCAGATGAGGAAGAAGAGG',
```

```
'ACGATGCAGAGTGTGAGGAAGAAGAAATTGATGAAACCTGTGAACATGAGTACGGTACAGAGGATGATTA',
'TCAAGGTCTCCCTCTGGAATTTGGTGCCTCAGCTGAAACAGTTCGAGTTGAGGAAGAAGAAGAGGAAGAC',
'TGGCTGGATGATACTACTGAGCAATCAGAGATTGAGCCAGAACCAGAACCTACACCTGAAGAACCAGTTA',
'ATCAGTTTACTGGTTATTTAAACTTACTGACAATGTTGCCATTAAATGTGTTGACATCGTTAAGGAGGC']
```

```
[15]: shorterSeq_join = "".join(strip_newline_seq)
```

```
[16]: print(shorterSeq_join)
```

```
AGGTGTAACCTTTGGAGAAGATACTGTTTGGGAAGTTCAAGGTTACAAGAATGTGAGAATCACATTTGAGCTTGATGAAC
GTGTTGACAAAGTGCTTAATGAAAAGTGCTCTGTCTACACTGTTGAATCCGGTACCGAAGTTACTGAGTTTGCATGTGTT
GTAGCAGAGGCTGTTGTGAAGACTTTACAACCAGTTTCTGATCTCCTTACCAACATGGGTATTGATCTTGATGAGTGGAG
TGTAGCTACATTCTACTTATTTGATGATGCTGGTGAAGAAAACCTTTTCATCACGTATGTATTGTTCTTTTACCCTCCAG
ATGAGGAAGAAGAGGACGATGCAGAGTGTGAGGAAGAAGAAATTGATGAAACCTGTGAACATGAGTACGGTACAGAGGAT
GATTATCAAGGTCTCCCTCTGGAATTTGGTGCCTCAGCTGAAACAGTTCGAGTTGAGGAAGAAGAAGAGGAAGACTGGCT
GGATGATACTACTGAGCAATCAGAGATTGAGCCAGAACCAGAACCTACACCTGAAGAACCAGTTAATCAGTTTACTGGTT
ATTTAAACTTACTGACAATGTTGCCATTAAATGTGTTGACATCGTTAAGGAGGC
```

```
[17]: len(shorterSeq_join)
```

```
[17]: 615
```

```
[20]: # 6. Step 3 and Step 5 have the SAME RESULT.
```