Python Fundamentals 1

Assignment PF1

Assignment Background:

You have sent off for some sequencing data. You have received the information, but it's not formatted how you want it:

```
seq1 = "SEQ001_ATCGATCGTAGCTAGCTAGCTA"
seq2 = "ATCG_CGTAGCTAG_CTAGCTA"

sequences = [
    "SEQ002_GCTAGCTAGCTAGCTAGCTAGT",
    "SEQ003_TAGCTAGCTAGCTAGCTAGCTA",
    "SEQ004_CTAGCTAGCTAGCTAGCTAGCT",
    "SEQ005_ATCGATCGATCGATCGATCGAT"
]
```

Later you receive a list of patient blood types for your clinical trial:

```
group_a_blood_types = ['A+', 'B+', 'O-', 'A+', 'AB+', 'O-', 'B+', 'A+', 'O+']
group_b_blood_types = ['O+', 'A-', 'B+', 'O+', 'A+', 'AB-', 'O-', 'A-']
```

Assignment Questions:

Strings

1. You have been given two sequences that should be 1, however there was an error when it was returned to you. You need to clean the data and join them together.
   - Remove the sequence ID from `seq1`, the characters before and including the `_`

   - `seq2` is missing some information, replace the underscores with `G`

   - Join the two sequences together, with `seq1` first

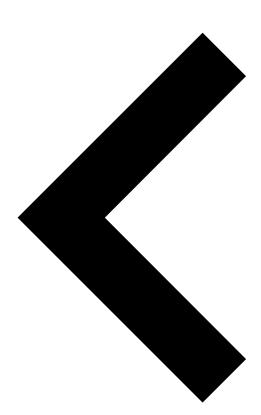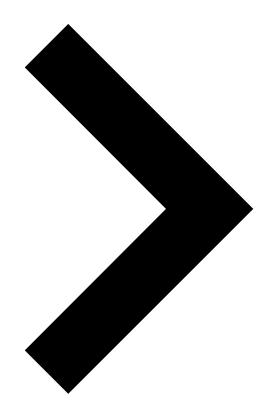2. Count how often `AG` appears in the string you created

Lists and Tuples

1. Convert your sequence to a list
   - Isolate and print the 5th item in the list

   - Isolate and print the 3rd to last item in the list

2. Count how many times "G" appears in the list you created in **1**

3. Add `seq1` to the start of the list called `sequences`.
   - You only need the first 4 items in the list, create a slice that achieves this

   - You decide you want to remove the last item instead, achieve this through a list method

4. Convert one of the new lists you created into a Tuple
   - Use indexing to access the 2nd sequence

   - Repeat the actions in question 1 of strings, i.e.:

- - Remove the sequence ID from the string, the characters before and including the _

  - - Add the given string to the end string; "AGTCGG"

  - Print the tuple to demonstrate the changes

Sets

1. Create sets called unique_group_a and unique_group_b to identify how many different blood types are present in each group.
   - Find all blood types present across both groups

   - Find blood types common to both groups

   - Find blood types that appear in only one group, not both (symmetric difference)