Lecture 11 - More practice with control structures

Practice with control structures

Challenge: Use a `for loop` and `if-else` statement to calculate the sum of wages for males and females in `wages.csv`.

Quick review on reading files

What are commands we've used to read data into `Python` or `R`?

Quick review on reading files

What are ways we've used to read data into `Python` or `R`?

`numpy.loadtxt()` works for tabular data of the same type

`read.table()` loads tabular data as dataframe

`pandas.read_csv()` loads tabular data as dataframe

`scan()` loads data as a vector; each line is an element

we can also directly open a file and access its contents with `open()`

Challenges

1. Create a script that reads DNA sequence records from a fasta file and generates a summary table with the following columns: *sequenceID*, *sequenceLength*, *percentGC*, and *meltingTemp*. The *sequenceID* is provided in the sequence record line (starts with "$>$") for each sequence, the *sequenceLength* is the total number of base pairs in each sequence, the *percentGC* is the precent of bases that are a G or C, and *meltingTemp* indicates the melting temperature of a sequence (assuming it is doublestranded). The calculation for *meltingTemp* is 4 x (number of G or C bases)+2 x (number of A or T bases), but this calculation is only reliable for sequences with 14 or fewer base pairs. Please report -9999 for the *meltingTemp* when the sequence is too long to generate a reliable estimate.
"Lecture11.fasta" is available on Sakai and should be analyzed with your script.

Challenges

2. Create a script that reads a series of integers from a text file ("findRuns.txt" available on Sakai) and reports the index for the element at the beginning of runs of repeated values and the length of the runs. For example, in the vector 0, 1, 2, 2, 3, 4 your script should store and/or return the values 3 (because the repeated 2's begin in element 3 of the vector) and 2 because there are two 2's in a row. Make sure your solution uses a for loop and at least one if-else statement.