**Programming Exercises Part III: Name: Ryan Kyger**

This week we went over:

Opening and writing to files (part of Chapter 10) and Regular Expressions. All materials are on the github repository.

**Part I: Short Answer**

1. *In this statement: InFileName = open(InFileName,‘r’) what does the ‘r’ indicate?*

‘r’ tells python to read the file

1. *What does this command do? Line.strip(‘\n’)why would you use it when printing out lines to the screen?*

This command will remove new line characters at the end of line.

You would use this when printing out new lines to a screen, because otherwise two line breaks will be applied when the print() function is used.

1. *How would you split a line from a csv file into a list?*

new\_list = line.split(‘,’)

1. *What are regular expressions? When would you use them?*

They are sequences of characters that are part of a search pattern.

You would use them to find and replace characters in a file.

1. *What does it mean that regular expressions are greedy?*

It means that regular expressions will match as much as they can.

1. *What would this match (^AB\S+) in the following line: ABCDEF GHIJK*

match: *ABCDEF*

1. *What would this match: (\S+$) in this line: Green eggs and ham.*

match: ham.

1. *What would this replacement look like:*
   1. *Line: Allen,Julie,Assistant-Professor*
   2. *Search statement: ^(\S+),(\S+),(\S+)$*
   3. *Replacement: \2 \1 is an \3*

replacement: Julie Allen is an Assistant-Professor

**Part II: Practice**

\*\* for bonus points put your scripts and pseudocode up on github and send me the link!

1. *As part of this exercise please practice in python. Practice print statements, adding input from the command line and making and editing variables of different types. Add in practicing if statements, for and while loops and dictionaries and opening files reading through them and writing.*
2. *Write the pseudocode to do the following. Open up the file “Bloom\_etal\_2018\_Reduced\_Dataset”.* ***Read through the file and print out the taxon name and their diadromous status.******Add up all of the log body sizes and print out the total log body size for all the individuals in the file.***

Pseudocode:

1. Open and read “Bloom\_etal\_2018\_Reduced\_Dataset.csv”

2. Loop trough each line

3. Print taxon name and diadromous status

4. Add the log body size to a list

5. Sum all items in list to get the total log body size

6. Print the total log body size

1. *Write the script to do number 2. You can use your pseudocode to do the documentation. Remember documentation should be at the beginning – explaining the script and then throughout explaining what the steps do. This is for future you!*
   1. *Looking for more practice? Create a tally of how many are diadromous vs. not and print out those counts. –not required.*

Script name: bloom\_data\_script.py

1. *Regular Expression Practice in BBedit or Sublime text. Open the file ‘regex.practice1.fasta’ – use regular expressions to replace all the names in the name row with the words ‘Homo\_sapens:’ followed by the numeric string at the beginning (e.g. 6573\_0:002008). Make sure it is in fasta format in other words the name lines starts with > and there are no spaces. Type your regular expression search and replacement statements here, and paste the new file contents.*

Search regex: >(\d+\_\d:\S+).+

Replacement regex: >Homo\_sapiens:$1

File contents:

>Homo\_sapiens:6573\_0:002008

MMHAVRRACCGTLKATKCGFSVPTATHSAVRAVPSFFTTRDYAAEAAKATSTISGKVVAVIGAVVDVQFDDDLPPILNALEVENRSPRLILEVAQHLGENTVRTIAMDGTEGLVRGTVCQDIGSPIRIPVGPATLGRIMNVIGEPIDERGPFNTKHFASIHAEAPDFVEMSVTQEILETGIKVVDLLAPYAKGGKIGLFGGAGVGKTVLIMELINNVAKAHGGYSVFAGVGERTREGNDLYHEMITSKVISLTDDSSKVSLVYGQMNEPPGARARVALTGLTVAEYFRDQEGQDVLLFIDNIFRFTQAGSEVSALLGRIPSAVGYQPTLATDMGTMQERITTTRKGSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSRGISELGIYPAVDPLDSISRILDANVVGEEHYVVARAVQKVLQDYKSLQDIIAILGMDELSEDDKMTVIRARKMQKFLSQPFQVAEVFTGQEGKYVPLTETIKGFKDILAGKYDHLPEVAFYLVGSIEEVVAKADRLAEEQSS

>Homo\_sapiens:6573\_0:002229

MMSFMMKKKRFKFHVNFCIEELSSVPYVSGILFAKIRLLDGGNFAGQSERQDITNNCVKWKAPFSFQCKMTANANTGSLDECKCRISIRRELKGGKTYQKLGYKDVDLAEYAGSGCQTKKFLLEGYDSKHRQDNSVIEVTVEMSLVSGDPVFKVPQKSKSVFYHLPGEMMDVQEQEARCIEDCSEGSLASNSSGFDSLPRKDRPTILPVENPPEFEISHDFEKSHSRNNSYASQHSKSSTGYVSLSHSRQSSIGNENPTHTRFLNVFYLSPSAGSALINDLIKGDRRKKFEDSMKERRMDSTRVDATGVVDELFNSADLLSSREESSGLQLLVDKDGTTALR

>Homo\_sapiens:6573\_0:0018f1

MLLRTQIVGVEFYDAVLNNELQLLKRLVTQHKIDLNAKFIEVRKKNHADLCPIHLASYRGYTYMLQYLIESKCDVNQTTTTLRRTALHFAVLRHKMACMLLLIAAGAQLDAKDTFSNSPCHYAADDGYCQMLDVLIRRGVNVDTLDITSKTPLMKAVRNNKTDAVLRLLRAQCNINITDRNSDMALHYAARNGCADVIDILISAGSQIDVQNYWGRTPLMETVCYNHKDAVARLLKASCDLNRREFKTGDTALHIAIKRNYTEVVGLLLAAGSRHDIYNHQGETAAYDAVVSNKVEIIRLMVIHNCDPEQPGKYFSDGVYKSLFQIAAEKGHFDICRLLASFGYVDFRARGYIYMNYIPPRLVTEQEDVVQWLRRKMQTPTSLQLLCRKCIRKRLGYKIGNTVECLPIPRSLKEFVLIKDMDEEEEKHLR

>Homo\_sapiens:6573\_0:0055ae

MTSVTARRNNQEYVVPSSTSQEVQPSPLALLAATCSKIGAPPPTPTEEGNNPTGGTQVRVIGQGSGEGAVAPTWIQLANGAIVDASGKQQTIGIPIQAAGAQPQLITSVIPQFQTVNVDGQEAIFIPSSGVGGGQTILAGNQQQVYTTPTGQIIRQGLPTGSMIPNMGYNLAGNMVNIGGNVVSLAGVQGLQNRPHGVVQAVQMPQQMQQMSNLIQIPVSTSNGQSTYQTIQLPSMQGFPMAFPQGQQATPQLQALSVNTSGVGQGSAGQITTVGNQQVTQVPQPQLIEIAAATTVKSQSDGSKPSTPQSQPTIASQATTTPVLQSSQQNNAGAIMTAGSFSATPGTPNILSQMPTLVPIGTNFINTATGQVIPMSQAMGTNGQNVIVASMTQQSGSSSTTTTSATSTVASTQQQPPSLVQAIPQQITVQALPSQNFANLQFAGQNQVIAQNPWLSALNVANIRAPSMQTIQVPNLQSLQNIQGIQAVQNVQGLQGFQITPQGQLIATGNVQNMGAVTLTPSGGLTVANAVGNNTQQQQQPSPQQFTATIQNVATQLTTGQTGQTLAGTPQIIAAGQQIQQDPNDPTKWQVVQTSQGAAVTLSPSQVSQQGVSVLGEQSVTSNRRLRRVACTCPNCTDNEGRTGENKKKQHICHIPGCKKVYGKTSHLRAHLRWHTGERPFVCNWMFCGKRFTRSDELQRHKRTHTGEKRFECKECKKKFMRSDHLSKHRKTHFNKKAGPNLDHSGNGEEDEGEMIEGGTMVMIECDDDGTEDGAEASFEVTDDDIDVQ

>Homo\_sapiens:6573\_0:00082a

MELKLHSPAGAEAVSYTWPLSNSDGKEGANEIIETIRWVSEDFPELETALDNNILQDYDVRSFESMKQLCDKYNRTIDSILQLWKGTSRPQKIKERPSTGLLKHILQQCYDHAVKDPERLNQYEPFSPEVYGETSFDLVDQMIKSINFSETDYFIDLGSGVGQVVLQVAAATSCKMCYGIEKAEYPAKYAEDMEKQFKRWMKWYGKKFSNYQLDKGDFLCDEMVEKINNATVIFVNNFAFGPQVDHQLKLRFANMKEGAKIVSSKAFCPLNFRITDRNLSDIGTIMHVAELSPLSGAVSWTGKAFTYFVHTIDRTLLEKYFLRMKNPKLKDDEDVRRDRRGRPIHLREKHTCCSGHGDSKPHEELKNGALKKKIRALEVNKNEMPALNVIKDHSYQAARNLDFDSASNASTATSMSGTAEEIAQVMGGPTTRRQWSECLSKPAPEVASPEHPPTFREESDENAVVVDIPKKMTRTRAAKTASKRLRNGQLRKEPHTAVGEVSSTPPVTGRKKSNLGSRARANKNKEKEKVLGLDSLNLLHTHTLMSTAASPESNDKWNDRSMVSMTNAFFKPSTQAQTVSSLETEPALQRLLDLFRHQYNKFLAYMQTPDYKSNLLLQIDQEREKQSCLKSRVAYLEKQIYGLQKDSVGQLKARMSEINFVSQLGIDTDCPADFLTKAKSIVQEHKELEGQMGTLKGQVASLEADRRKLVEVQQQVLLEKTGGINGKRNGFHKPTVQENVLKEINNSYSQKKNLITKVQELEVNLNDLQQHKLARLDDDIKGDDSPSQDSGISLTPTVPPPQSPDEAATTDMSGKPLGLYFQSTTDKSQDQTNVSADTKQIKIVTGAISGFRQTSPCNNLLDHKPSAKDSGTLGVSASSHENHSSIKSSPYMSSSLSSSSSPPPPRKPGDGYTPGALSIRILLEASGEQVRQRNRGEIASRLAGYKSGVVKDLHQDSTGKLEYERIKKEAKPAHLIDQVVHMALSSQPTDSNQGNVLVTHNKKVKLSNGVKSSAKGVALESNRPTLPVSIPLDQLQKEQLRQKILNRTGPSKHEAGDVTSQVSKQSSKSDHDAGFYSPISRPSSRGSTTESADEPGCDVTASHSYSPATASNSPRGPGYSSSRHRTDSKAKTPARVPVNNHFVENMSTQTKSNNLLHGRGQRVHTDQQKFNLAEALLQMSEVRSNKPVSPRQALPSVPTKCGDSSKSPKMRYTLESYQAADRKRLKRKKTEILENDNKKMALSSNAVSYTPSHGVLDSMLGHTSGISSFSSPDSGMSTPHVQSPLFPTARPKLKSHYNLQNSSAKDSPSKNWQAQISSGFDALVALASSELDKNREIRRKSLEGQSPSPRKSQVIPTKQHVSLRDSLRGRDNDKRRGPKTPPGSPPRSKRGPRTPPGSPHSQKKGVKGRRNNSYSSSRSRSCSLSRTSSHTQSSRSSSRSSCYSSSRSRSRSRSSSSSRSSGSGRHRNTKISLKKASETVNKPVVTHPVINSNIKPVSQQYSGNNSTSVPKQPAPLVSNAVTGTVMNNTNYCVLNNGTSFGGGMMLMTNNQLHTGTTTTGISQRPMTNQMPIIPSSIPSNLVVLSFTPQNNSMNNQILGGPTAALQTSTITQQPAPQVPNLSEQLKQCAKQPPPPPPNNDNVQKGLTLQPTAQNIHYTNPGPVFVSTAPNFSLPDLSRPPPNMALKPPNSSLPGSGGGTVVATPTGQLQMCAGGAGQPATPVFTINHRPPSITPTPIRPRSDLPPRPMTTFNERPFTPGNVGNIPDFTRPPVNLQPRLPSPAKGQHQQQGPRPQQNQPANFRGPIQIRPAISNPAGPRLDASATRFQGYATLQPTQQTSNIRLGMFANQSGPRQQFNGPGQQNWQM

>Homo\_sapiens:6573\_0:0047a8

MGNGMNKIIPGLYVGNFRDSKDPEQLEKYKITHILSIHDNAKKILEDKEYLCIVASDTPEQTLTKFFPQCIDFIHEARLKGGNVIVHCIAGVSRSVTVTAAYIMTVTSLGWRDAVNAIRGARSCANPNFGFQRQLQQFESDLVEKERKRIKEKYDPSPYNDEMECRALLTAFKKFVLYDDPKDKDDGLYPLPPNAYKPDPRDKSPPTDSRKEDKSKKTGYSKYAEKDAIQKKLS