

Assignment 7

Due Wednesday, October 14 before midnight (California time)

On Blackboard, you can find two files: “beforerrain.bracken” and “afterrain.bracken”. The data in the first file is the sample from LA River before rain on a certain day, and the second file is the sample from the same river few days later after rain.

These files are outputs from the Bracken program.

1. Write an R program that takes as input one of these files and a threshold, and returns all “names” (1st column) and “fractions” (7th column) of those rows where the fraction is greater than the threshold. Run the program for each file with threshold 0.01 [2pt].
2. Write an R program that takes as input both of these files and a number n , and returns the “names” and “fractions” (the fractions in both files) for the n names with the greatest absolute difference in fractions between the two files. Note: some names might be present in one file but absent (not even listed) in the other file. The fraction for the name not listed in the file is zero. Run the program with the number n equal to 10 [3pt].
3. Let $\{r_i\}$ be the “new_est_reads” numbers (6th column) in one of the files. Define $\{p_i = r_i / \sum r_j\}$. The Shannon diversity for the file is defined as $-\sum p_i \ln(p_i)$. Write an R program to compute the Shannon diversity. Run this program for both files and tell us what you find [3pt].

Turn in the code for the aforementioned R functions and the answers into one file in Jupyter Notebook format (.ipynb). Use the “Turnitin” link on Blackboard/Assignments/Assignment 7 to submit this file.