

Exercises

(The following exercises require that you lightly research a few things that have not been taught in class)

1. The file `directory.txt` contains 3 records of persons, their email and their age. Using only `awk`:
 - a. print the first line of each record.
Take in account that each record is separated by an empty line
 - b. print the first line of each record affixed with an incremental number
 - c. calculate the age average of the 3 persons
 - d. print only each name (without the email)
 - e. print only emails (without the name)
This is more difficult. You will have to use `awk` twice in a pipe (unless you think of another solution)

PS: there is a difference between the newlines between Windows and Linux which can sometimes cause problems. If you encounter such a problem, please check <https://kb.iu.edu/d/acux>

2. In the GTF file previously given to you, fields 4,5 represent the point where the nucleotide sequence of the gene starts and ends, respectively. By subtracting those 2 you can calculate the length of each gene.
Print the length of each gene affixed with an incremental number.
If you wish to remove first lines that are not records, use `NR` (`awk 'NR>5 {.....}'`)