

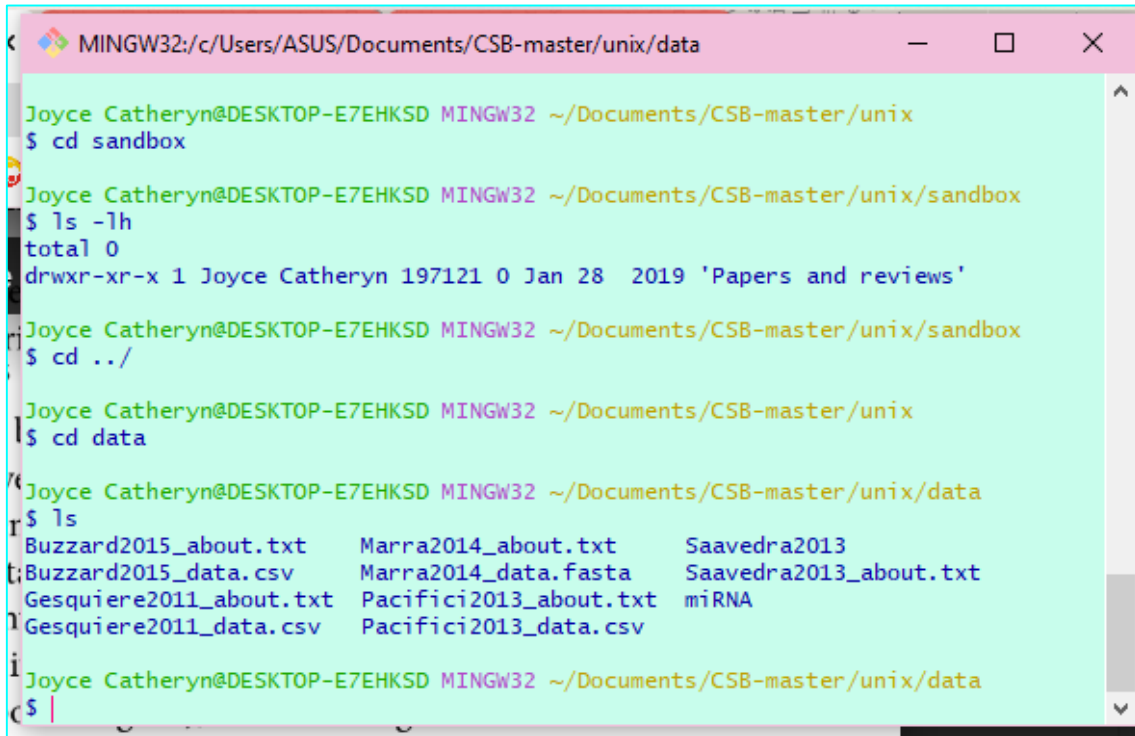
# Bioinformatics Assignment

## TC3

Name: Joyce Andrade

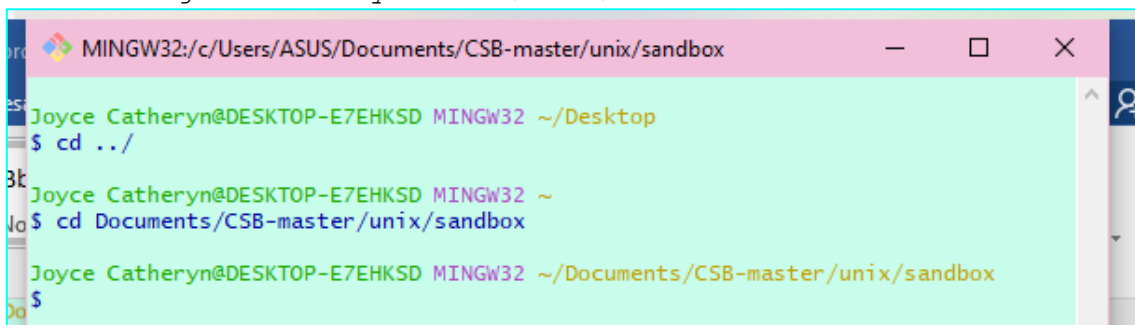
Screenshots. -

### 1.10.1 Next Generation Sequencing Data



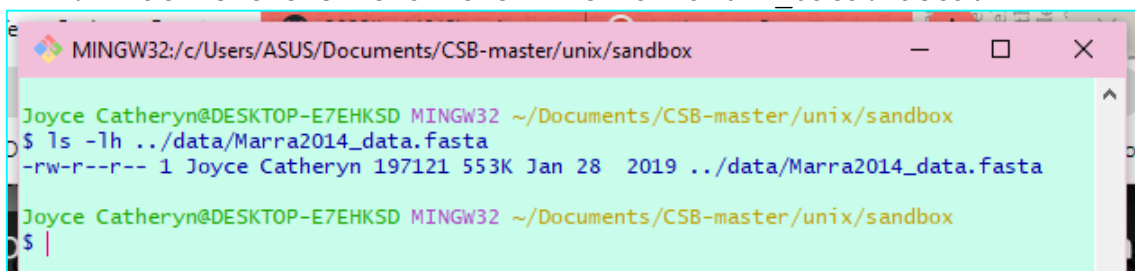
```
MINGW32:/c/Users/ASUS/Documents/CSB-master/unix/data
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix
$ cd sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ ls -lh
total 0
drwxr-xr-x 1 Joyce Catheryn 197121 0 Jan 28 2019 'Papers and reviews'
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ cd ../
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix
$ cd data
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/data
$ ls
Buzzard2015_about.txt  Marra2014_about.txt  Saavedra2013
Buzzard2015_data.csv  Marra2014_data.fasta  Saavedra2013_about.txt
Gesquiere2011_about.txt  Pacifici2013_about.txt  miRNA
Gesquiere2011_data.csv  Pacifici2013_data.csv
```

1. Change directory to CSB/unix/sandbox.



```
MINGW32:/c/Users/ASUS/Documents/CSB-master/unix/sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Desktop
$ cd ../
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~
$ cd Documents/CSB-master/unix/sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$
```

2. What is the size of the file Marra2014\_data.fasta?



```
MINGW32:/c/Users/ASUS/Documents/CSB-master/unix/sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ ls -lh ../data/Marra2014_data.fasta
-rw-r--r-- 1 Joyce Catheryn 197121 553K Jan 28 2019 ../data/Marra2014_data.fasta
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$
```

3. Create a copy of Marra2014\_data.fasta in the sandbox and name it my\_file.fasta.

```
MINGW32:/c/Users/ASUS/Documents/CSB-master/unix/sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ cp ../data/Marra2014_data.fasta my_file.fasta

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$
```

4. How many contigs are classified as isogroup00036?

```
MINGW32:/c/Users/ASUS/Documents/CSB-master/unix/sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ grep -c isogroup00036 my_file.fasta
16

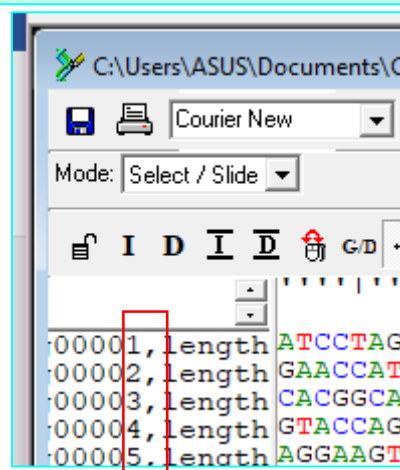
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$
```

5. Replace the original "two-spaces" delimiter with a comma.

```
MINGW32:/c/Users/ASUS/Documents/CSB-master/unix/sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ cat my_file.fasta | tr -s " " "," > my_file.tmp

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ mv my_file.tmp my_file.fasta

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$
```



6. How many unique isogroups are in the file?

```
MINGW32:/c/Users/ASUS/Documents/CSB-master/unix/sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | head -n 4
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
>contig00002,length=551,numreads=8,gene=isogroup00001,status=it_thresh
>contig00003,length=541,numreads=2,gene=isogroup00001,status=it_thresh
>contig00004,length=291,numreads=3,gene=isogroup00001,status=it_thresh

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 4 | head -n 4
gene=isogroup00001
gene=isogroup00001
gene=isogroup00001
gene=isogroup00001

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 4 | sort | uniq | wc -l
43

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$
```

7. Which contig has the highest number of reads (numreads)? How many reads does it have?

```
MINGW32:/c/Users/ASUS/Documents/CSB-master/unix/sandbox
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 1,3 | head -n 4
>contig00001,numreads=2
>contig00002,numreads=8
>contig00003,numreads=2
>contig00004,numreads=3

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 1,3 | sort -t '=' -n -k 2 | head -n 4
>contig00089,numreads=1
>contig00176,numreads=1
>contig00210,numreads=1
>contig00001,numreads=2

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 1,3 | sort -t '=' -n -k 2 -r | head -n 1
>contig00302,numreads=3330

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/unix/sandbox
$ |
```

Push/Git-hub

```

MINGW32/c/Users/ASUS/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04
GNU nano 5.8                               Clase04_if.sh
## Exercise 1.10.1 ##
## 1. Change directory to CSB/unix/sandbox##
echo "1. Change directory to CSB/unix/sandbox"
cd ../../../../unix/sandbox/
## 2. What is the size of file Marra2014_data.fasta ##
echo "2. What is the size of file Marra2014_data.fasta"
ls -lh ../data/Marra2014_data.fasta
## 3. Create a copy of Marra2014_data.fasta in the sandbox and name it my_file.fasta##
echo "3. Create a copy of Marra2014_data.fasta in the sandbox and name it my_file.fasta"
cp ../data/Marra2014_data.fasta my_file.fasta
## 4. How many contigs are classified as isogroup00036? ##
echo "4. How many contigs are classified as isogroup00036? "
grep -c isogroup00036 my_file.fasta
## 5. Replace the original "two-spaces" delimiter with a comma ##
echo "5. Replace the original "two-spaces" delimiter with a comma "
cat my_file.fasta | tr -s " " "," > my_file.tmp
mv my_file.tmp my_file.fasta
## 6. How many unique isogroups are in the file?
echo "6. How many unique isogroups are in the file?"
grep '>' my_file.fasta | head -n 4
grep '>' my_file.fasta | cut -d ',' -f 4 | head -n 4
grep '>' my_file.fasta | cut -d ',' -f 4 | sort | uniq | wc -l
## 7. Which contig has the highest number of reads (numreads)? How many reads does it have? ##
echo "7. Which contig has the highest number of reads (numreads)? How many reads does it have? "
grep '>' my_file.fasta | cut -d ',' -f 1,3 | head -n 4
grep '>' my_file.fasta | cut -d ',' -f 1,3 | sort -t '=' -n -k 2 | head -n 4
grep '>' my_file.fasta | cut -d ',' -f 1,3 | sort -t '=' -n -k 2 -r | head -n 1
^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   M-U Undo
^X Exit      ^R Read File  ^_ Replace    ^U Paste      ^J Justify    ^/_ Go To Line M-E Redo

```

```

MINGW32/c/Users/ASUS/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2
022II_gbi6/Clase04 (main)
$ nano Clase04_if.sh

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git add Clase04_if.sh
warning: in the working copy of 'Clase04/Clase04_if.sh', LF will be replaced by CRLF the next time Git touches it

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git commit -m "Tarea de Clasesita 004"
[main 594c4ca] Tarea de Clasesita 004
1 file changed, 7 insertions(+)

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git push origin
Enumerating objects: 7, done.
Counting objects: 100% (7/7), done.
Delta compression using up to 2 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (4/4), 482 bytes | 120.00 KiB/s, done.
Total 4 (delta 2), reused 0 (delta 0), pack-reused 0
remote: Resolving deltas: 100% (2/2), completed with 2 local objects.

```



.sh a .csv

```

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ bash Clase04_if.sh > Clase04_if.csv

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)

```

	Clase04_if	14/11/2022 13:31	Archivo de valores separados por comas de Microsoft Excel	2 KB
	Clase04_if	14/11/2022 13:30	Shell Script	2 KB

### 1.10.2 Hormone Levels in Baboons

1. How many times were the levels of individuals 3 and 27 recorded?

```

MINGW32/c/Users/ASUS/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04
GNU nano 5.8 Clase04_while.sh
## Exercise 1.10.2 Hormone levels in Baboons

##1. How many times were the levels of individuals 3 and 27 recorded?##
echo "1. How many times were the levels of individuals 3 and 27 recorded?"
cd ../../../../
cd unix/data/
ls
head -n 3 Gesquiere2011_data.csv
cut -f 1 Gesquiere2011_data.csv | head -n 3
echo "Male ID:3"
cut -f 1 Gesquiere2011_data.csv | grep -c -w 3
echo "Male ID:27"
cut -f 1 Gesquiere2011_data.csv | grep -c -w 27

```

2. Write a script taking as input the file name and the ID of the individual, and returning the number of records for that ID.

```

##2. Write a script taking as input the file name and the ID of the individual, and returning the number of records for that ID.
echo "2. Write a script taking as input the file name and the ID of the individual, and returning the number of records for that ID."
echo "Male ID:3"

bash ../solutions/count_baboons.sh Gesquiere2011_data.csv 3
echo "Male ID:27"
bash ../solutions/count_baboons.sh Gesquiere2011_data.csv 27
## cut -f 1 $1 | grep -c -w $2##
## $1 filename and $2 ID##

```

3.[Advanced] Write a script that returns the number of times each individual was sampled.

```

##3. [Advanced] Write a script that returns the number of times each individual was sampled.##
echo "3. [Advanced] Write a script that returns the number of times each individual was sampled."
echo "times each individual was sampled"
tail -n +2 Gesquiere2011_data.csv | cut -f 1 | sort -n | uniq
echo "BUCLE"
myIDS='tail -n +2 Gesquiere2011_data.csv | cut -f 1 | sort -n | uniq`
for id in $myIDS
do
    mycounts='bash ../solutions/count_baboons.sh Gesquiere2011_data.csv $id`
    echo "ID:" $id "counts:" $mycounts
done

```

Push/Git-Hub



```

MINGW32/c/Users/ASUS/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ nano Clase04_while.sh
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git add Clase04_while.sh
warning: in the working copy of 'Clase04/Clase04_while.sh', LF will be replaced by CRLF the next time Git touches it
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git commit -m "Tarea de Clasesita 004-2"
[main 82f9b7b] Tarea de Clasesita 004-2
1 file changed, 37 insertions(+)
create mode 100644 Clase04/Clase04_while.sh
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git push origin
Enumerating objects: 6, done.
Counting objects: 100% (6/6), done.
Delta compression using up to 2 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (4/4), 901 bytes | 75.00 KiB/s, done.
Total 4 (delta 1), reused 0 (delta 0), pack-reused 0
remote: Resolving deltas: 100% (1/1), completed with 1 local object.
To https://github.com/Joycita/2022II_gbi6.git
d64a7c7..82f9b7b main -> main

```

.sh a .csv

```
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ bash Clase04_while.sh > Clase04_while.csv
```

	Clase04_while	14/11/2022 16:58	Archivo de valores separados por comas de Microsoft Excel	4 KB
	Clase04_while	14/11/2022 16:55	Shell Script	2 KB

### 1.10.3 Plant-Pollinator Networks

1. Write a script that takes one of these files and determines the number of rows (pollinators) and columns (plants). Note that columns are separated by spaces and that there is a space at the end of each line. Your script should return.

```
MINGW32:/c/Users/ASUS/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04
GNU nano 5.8 Clase04_until.sh
## Ejercicio 1.10.3 Plant-Pollinator Networks ##
##1.write a script that takes one of these files and determines the number of rows (pollinators) and columns (plants)##
echo "1. Write a script that takes one of these files and determines the number of rows (pollinators) and columns (plants)"
cd ../../../../unix/data

echo "Para n1.txt"
cat Saavedra2013/n1.txt | wc -l
head -n 1 Saavedra2013/n1.txt | tr -d ' ' | tr -d '\n' | wc -c

echo "Para n59.txt"
cat Saavedra2013/n59.txt | wc -l
head -n 1 Saavedra2013/n59.txt | tr -d ' ' | tr -d '\n' | wc -c
```

2. [Advanced]Write a script that prints the numbers of rows and columns for each network:

```
##2.[Advanced]Write a script that prints the numbers of rows and columns for each network##
echo "[Advanced]write a script that prints the numbers of rows and columns for each network:"

for z in Saavedra2013/*.txt;
do
    Filitas=`cat $z | wc -l`;
    echo "Archivos:" $z "Filitas:" $Filitas;
done

for z in Saavedra2013/*.txt;
do
    Columnitas=`head -n 1 $z | tr -d ' ' | tr -d '\n' | wc -c`;
    echo "Archivos:" $z "Columnitas:" $Columnitas;
done
```

3. Which file has the largest number of rows? Which has the largest number of columns?

```
##3.Which file has the largest number of rows? Which has the largest number of columns?##
echo "which file has the largest number of rows? Which has the largest number of columns?"

echo "Mayor número de columnas"
bash Todaslascositas.sh | sort -n -r -k 3 | head -n 1

echo "Mayor número de filas"
bash Todaslascositas.sh | sort -n -r -k 2 | head -n 1
```

Push/Git-Hub

```
echo "Archivos:" $? "Filtros:" $Filtros;

MINGW32:/c/Users/ASUS/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6...
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git commit -m "Tarea de la Clasesita 004-3"



Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git add Clase04_until.sh

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git pull
Already up to date.

Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ git push origin
Enumerating objects: 21, done.
Counting objects: 100% (20/20), done.
Delta compression using up to 2 threads
Compressing objects: 100% (15/15), done.
Writing objects: 100% (15/15), 1.36 KiB | 231.00 KiB/s, done.
Total 15 (delta 11), reused 0 (delta 0), pack-reused 0
remote: Resolving deltas: 100% (11/11), completed with 3 local objects.
To https://github.com/Joycita/2022II_gbi6.git
485ecae..d0a7a0a main -> main
```

.sh a .csv

```
Joyce Catheryn@DESKTOP-E7EHKSD MINGW32 ~/Documents/CSB-master/git/sandbox/gbi6/2022II_gbi6/Clase04 (main)
$ bash Clase04_until.sh > Clase04_until.csv
```

	Clase04_until	15/11/2022 9:14	Archivo de valores separados por comas de Microsoft Excel	6 KB
	Clase04_until	15/11/2022 8:52	Shell Script	2 KB

## Commits

File Edit View Repository Branch Help

Current repository 2022II\_gbi6

Current branch main

Fetch origin Last fetched 2 minutes ago

Changes

History

Select branch to compare...

Merge branch 'main' of https://github.com/Joycita/2022II\_gbi6

Mi primer cambio

Tarea de Clase 4

Tarea de Clasesita 004-3

Tarea de Clasesita 004-3

Tarea de Clasesita 004-3

Tarea de Clasesita 004-3

Tarea de Clasesita 004-3

Tarea de Clasesita 004-3

Tarea de Clasesita 004-3

Tarea de Clasesita 004-2

Tarea de Clasesita 004-3

Tarea de Clasesita 004-3

Joycita 105255 1 changed file +10 -8

Clase04/Clase04\_until.sh

@@ -1,20 +1,22 @@

1 ## Ejercicio 1.10.3 Plant-Pollinator Networks ##

2 ##1. Write a script that takes one of these files and determines the number of rows (pollinators) and columns (plants)##

3 echo "1. Write a script that takes one of these files and determines the number of rows (pollinators) and columns (plants)"

4 - cd ../../../../unix/data/Saavedra2013

5 - ls -l

6 - bash n5.txt

7 - cat n1.txt | wc -l

8 - cat n59.txt | wc -l

9

10 echo -e "\n"

11

12 - head -n 1 n1.txt | tr -d ' ' | tr -d '\n' | wc -c

13 - head -n 1 n1.txt | tr -d ' ' | tr -d '\n' | wc -c

14 + cd ../../../../unix/data/Saavedra2013

15 + wc -l Saavedra2013/n1.txt

16 +

17 + cat Saavedra2013/n1.txt | wc -l

18 + cat Saavedra2013/n59.txt | wc -l

19

20 echo -e "\n"

21 + head -n 1 Saavedra2013/n1.txt

22 + head -n 1 Saavedra2013/n59.txt

23

24 + head -n 1 Saavedra2013/n1.txt | tr -d ' ' | tr -d '\n' | wc -c

25 + head -n 1 Saavedra2013/n59.txt | tr -d ' ' | tr -d '\n' | wc -c



