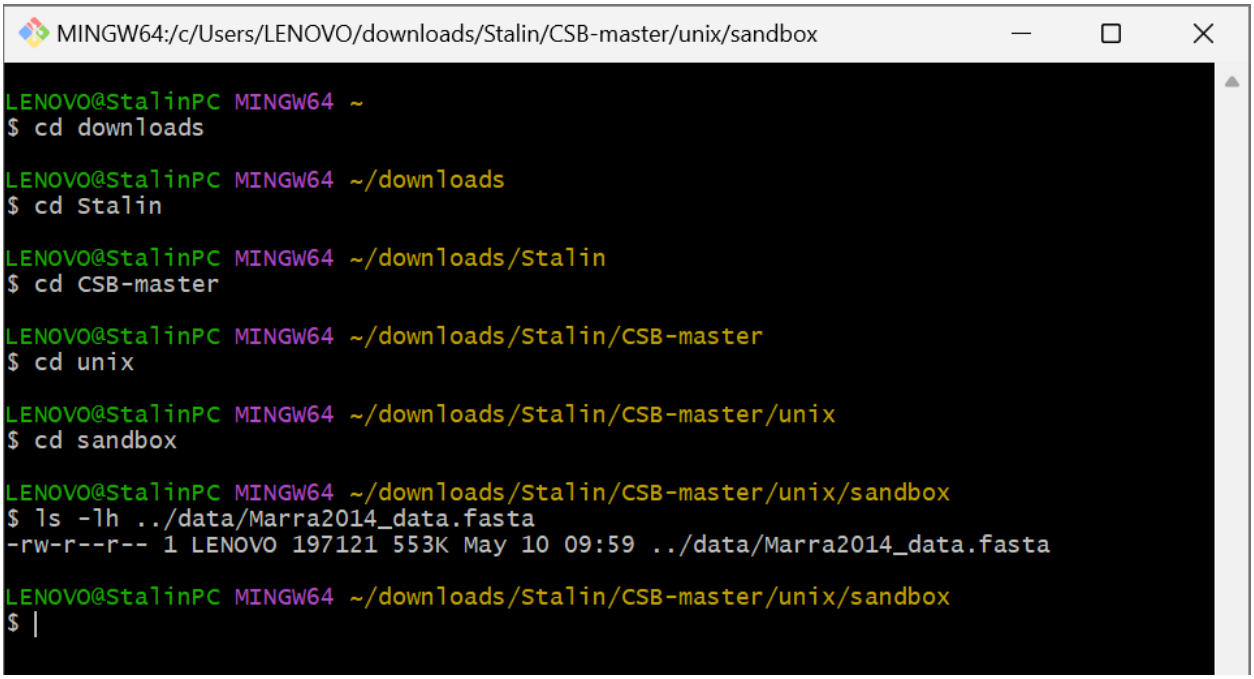
UNIVERSIDAD REGIONAL AMAZONICA IKIAM

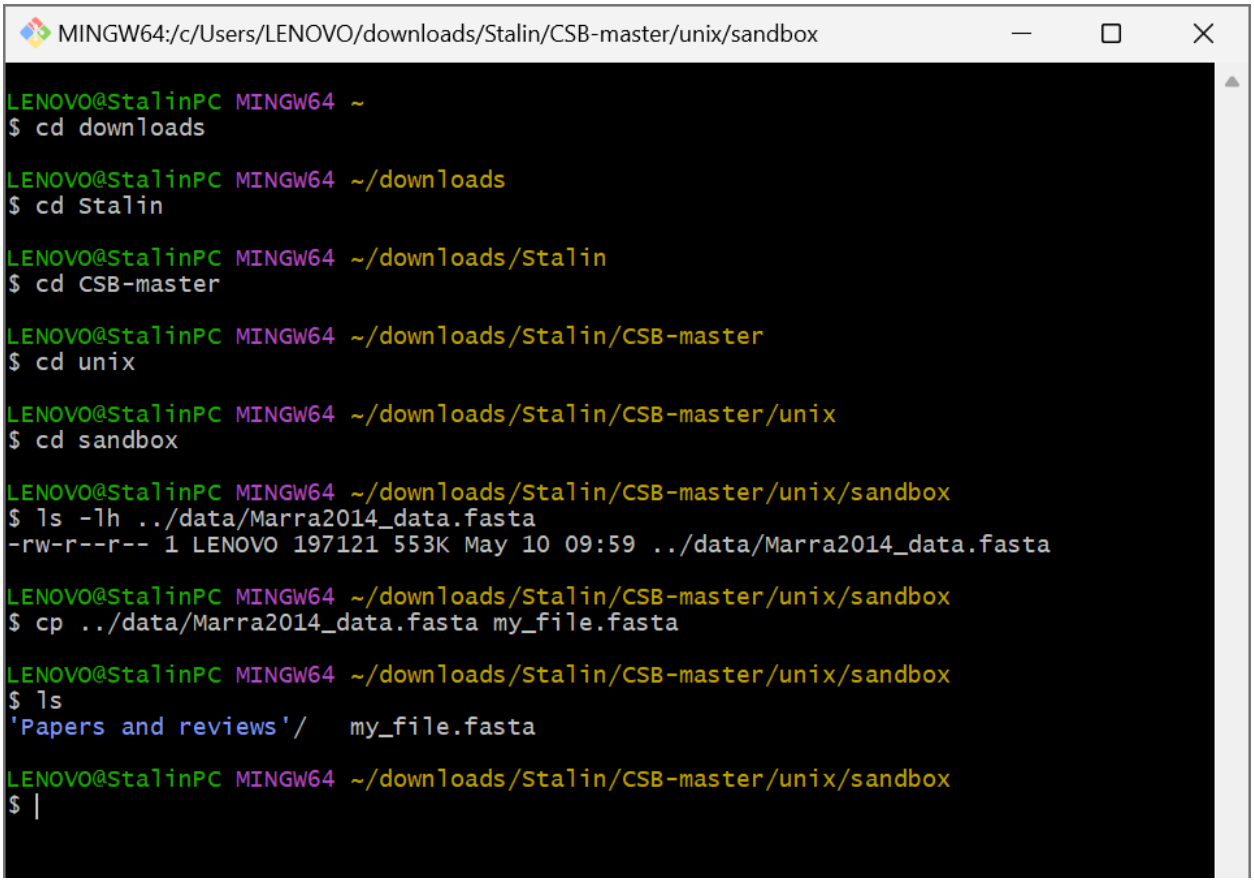
BIOINFORMATICA

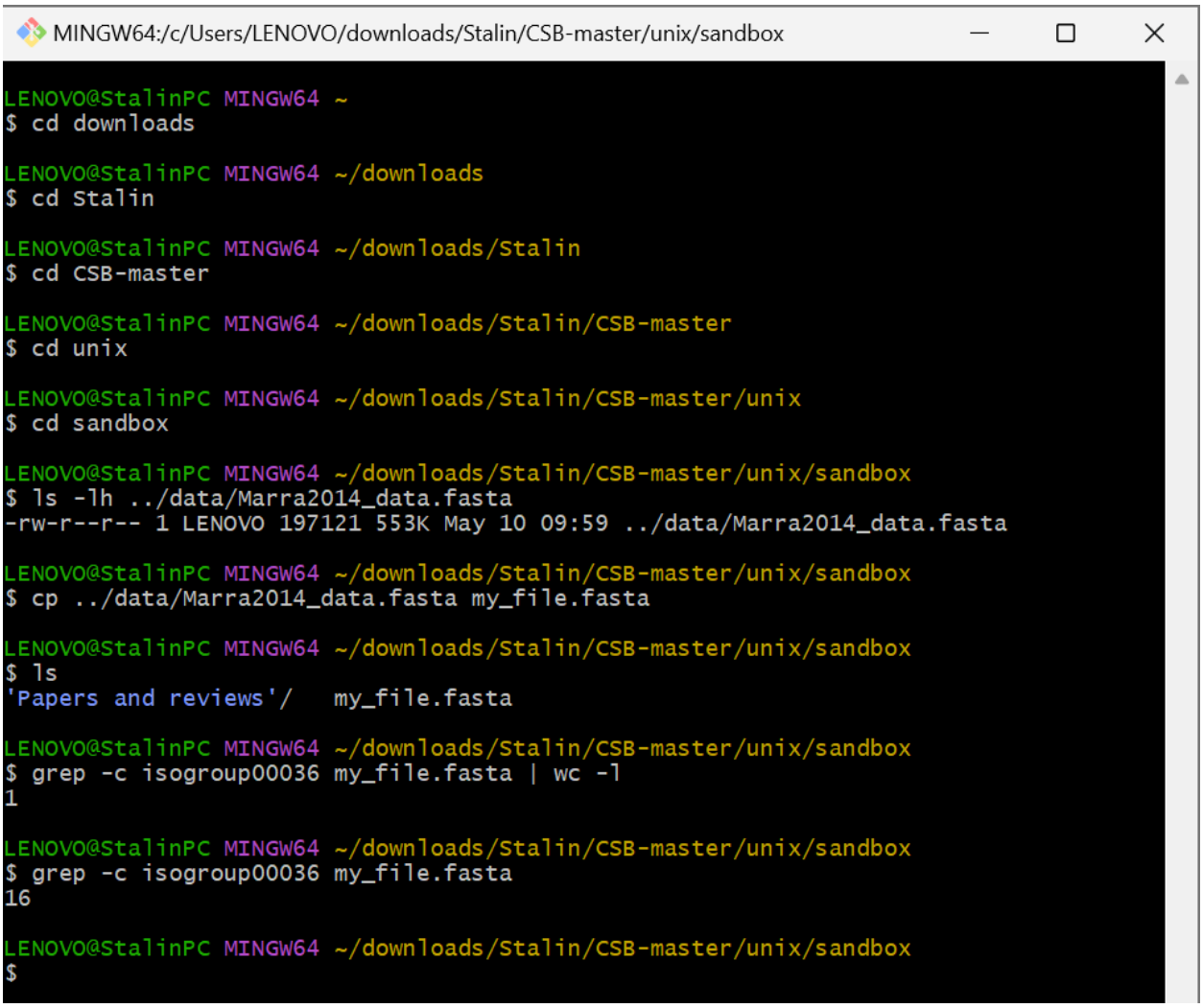
FECHA: 09/Mayo/2023

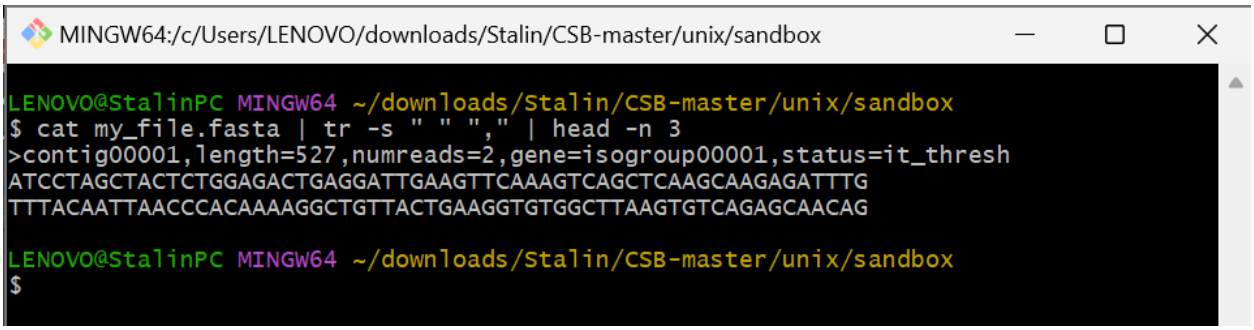
Nombre: Stalin Daniel Guaigua Puedma

Ejercicio 1.10.1




MINGW64:/c/Users/LENOVO/downloads/Stalin/CSB-master/unix/sandbox
LENOVO@StalinPC MINGW64 ~
$ cd downloads
LENOVO@StalinPC MINGW64 ~/downloads
$ cd Stalin
LENOVO@StalinPC MINGW64 ~/downloads/Stalin
$ cd CSB-master
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master
$ cd unix
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix
$ cd sandbox
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ ls -lh ../data/Marra2014_data.fasta
-rw-r--r-- 1 LENOVO 197121 553K May 10 09:59 ../data/Marra2014_data.fasta
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ cp ../data/Marra2014_data.fasta my_file.fasta
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ ls
'Papers and reviews'/   my_file.fasta
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$


MINGW64:/c/Users/LENOVO/downloads/Stalin/CSB-master/unix/sandbox
LENOVO@StalinPC MINGW64 ~
$ cd downloads
LENOVO@StalinPC MINGW64 ~/downloads
$ cd Stalin
LENOVO@StalinPC MINGW64 ~/downloads/Stalin
$ cd CSB-master
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master
$ cd unix
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix
$ cd sandbox
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ ls -lh ../data/Marra2014_data.fasta
-rw-r--r-- 1 LENOVO 197121 553K May 10 09:59 ../data/Marra2014_data.fasta
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ cp ../data/Marra2014_data.fasta my_file.fasta
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ ls
'Papers and reviews'/   my_file.fasta
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ grep -c isogroup00036 my_file.fasta | wc -l
1
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ grep -c isogroup00036 my_file.fasta
16
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$


MINGW64:/c/Users/LENOVO/downloads/Stalin/CSB-master/unix/sandbox
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ cat my_file.fasta | tr -s " " "," | head -n 3
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG
TTTACAATTAACCCACAAAAGGCTGTTACTGAAGGTGTGGCTTAAGTGTCAGAGCAACAG
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$

```
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d "," -f 4 | sort | uniq | wc -l
955
```

MINGW64:/c/Users/LENOVO/downloads/Stalin/CSB-master/unix/sandbox

```
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d "," -f 1,3 | head -n 3
>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
```

```
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d "," -f 1,3 | sort -t "-" -k 2 -n
>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
>contig00005  length=580  numreads=12  gene=isogroup00001  status=it_thresh
>contig00006  length=3288  numreads=35  gene=isogroup00001  status=it_thresh
>contig00008  length=1119  numreads=10  gene=isogroup00001  status=it_thresh
>contig00010  length=202  numreads=4  gene=isogroup00001  status=it_thresh
>contig00011  length=5563  numreads=61  gene=isogroup00001  status=it_thresh
>contig00012  length=824  numreads=10  gene=isogroup00001  status=it_thresh
>contig00013  length=707  numreads=6  gene=isogroup00001  status=it_thresh
>contig00014  length=568  numreads=2  gene=isogroup00001  status=it_thresh
>contig00015  length=123  numreads=4  gene=isogroup00001  status=it_thresh
>contig00016  length=1721  numreads=22  gene=isogroup00001  status=it_thresh
>contig00017  length=788  numreads=15  gene=isogroup00001  status=it_thresh
>contig00018  length=1886  numreads=12  gene=isogroup00001  status=it_thresh
>contig00022  length=802  numreads=3  gene=isogroup00001  status=it_thresh
>contig00023  length=885  numreads=10  gene=isogroup00001  status=it_thresh
>contig00025  length=142  numreads=3  gene=isogroup00001  status=it_thresh
>contig00026  length=301  numreads=4  gene=isogroup00001  status=it_thresh
>contig00027  length=178  numreads=7  gene=isogroup00001  status=it_thresh
>contig00028  length=135  numreads=2  gene=isogroup00001  status=it_thresh
>contig00029  length=465  numreads=5  gene=isogroup00001  status=it_thresh
>contig00030  length=121  numreads=3  gene=isogroup00001  status=it_thresh
>contig00031  length=595  numreads=9  gene=isogroup00001  status=it_thresh
>contig00032  length=1202  numreads=5  gene=isogroup00001  status=it_thresh
>contig00033  length=540  numreads=27  gene=isogroup00001  status=it_thresh
>contig00034  length=1004  numreads=6  gene=isogroup00001  status=it_thresh
>contig00035  length=566  numreads=8  gene=isogroup00001  status=it_thresh
>contig00036  length=3571  numreads=8  gene=isogroup00001  status=it_thresh
>contig00037  length=1168  numreads=10  gene=isogroup00001  status=it_thresh
>contig00039  length=371  numreads=94  gene=isogroup00001  status=it_thresh
>contig00040  length=2505  numreads=15  gene=isogroup00001  status=it_thresh
>contig00042  length=620  numreads=11  gene=isogroup00001  status=it_thresh
>contig00044  length=1301  numreads=4  gene=isogroup00001  status=it_thresh
>contig00045  length=563  numreads=6  gene=isogroup00001  status=it_thresh
>contig00046  length=1046  numreads=44  gene=isogroup00001  status=it_thresh
>contig00048  length=211  numreads=4  gene=isogroup00001  status=it_thresh
>contig00049  length=740  numreads=3  gene=isogroup00001  status=it_thresh
```

```
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d "," -f 1,3 | sort -t "-" -k 2 -n | head -n 5
>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
>contig00005  length=580  numreads=12  gene=isogroup00001  status=it_thresh
```

```
LENOVO@StalinPC MINGW64 ~/downloads/Stalin/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d "," -f 1,3 | sort -t "-" -k 2 -n -r | head -n 1
>contig01385  length=1965  numreads=7  gene=isogroup00043  status=isotig
```

Ejercicio 1.10.2

MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
  GNU nano 7.2                    Ejercicio1_10_2.sh
#!/bin/bash

#¿Cuántas veces fueron los niveles de los individuos 3 y 27 grabados?

#Para los individuos 3

cut -f 1 Gesquiere2011_data.csv | grep -w 3 | grep -c 3

#Para los individuos 27

cut -f 1 Gesquiere2011_data.csv | grep -w 27 | grep -c 27
```

MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ nano Ejercicio1_10_2.sh

LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ nano Ejercicio1_10_2.sh

LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ cat Ejercicio1_10_2.sh
#!/bin/bash

#¿Cuántas veces fueron los niveles de los individuos 3 y 27 grabados?

#Para los individuos 3

cut -f 1 Gesquiere2011_data.csv | grep -w 3 | grep -c 3

#Para los individuos 27

cut -f 1 Gesquiere2011_data.csv | grep -w 27 | grep -c 27
```

```
LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ bash Ejercicio1_10_2.sh
61
5
```

MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
  GNU nano 7.2                    Ejercicio1_10_2.sh                    Modified
#!/bin/bash

#¿Cuántas veces fueron los niveles de los individuos 3 y 27 grabados?

#Para los individuos 3

cut -f 1 Gesquiere2011_data.csv | grep -w 3 | grep -c 3

#Para los individuos 27

cut -f 1 Gesquiere2011_data.csv | grep -w 27 | grep -c 27

#Luego se crea un vector para los IDS del documento

vdownloads=`tail -n +2 Gesquiere2001_data.csv | cut -f 1 | uniq `

for x in $vdownloads
do
ids=`bash Conteo_Ejercicio1_10_2.sh Gesquiere2011_data.csv $x`
echo "ID:" $x "conteo:" $ids
done
```

MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
  GNU nano 7.2                  Conteo_Ejercicio1_10_2.sh                 Modified
#!/bin/bash

#Codigo con sus variables

cut -f 1 $1 | grep -c -w $2
```

MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ bash Ejercicio1_10_2.sh
61
5
tail: cannot open 'Gesquiere2001_data.csv' for reading: No such file or directory

LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$
```

Ejercicio 1.10.3

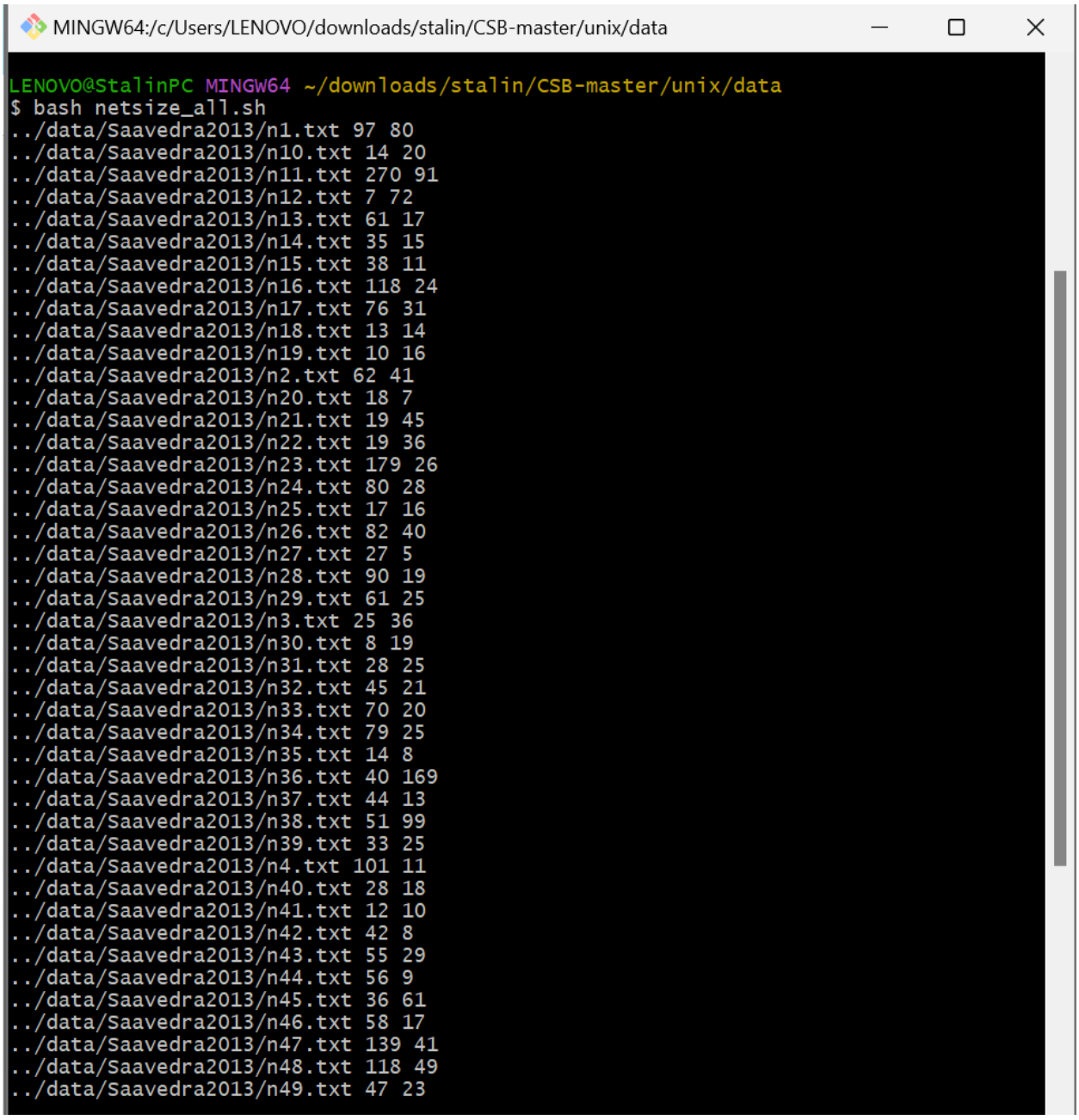
MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
  GNU nano 7.2                          netsize.sh                          Modified
#!/bin/bash

echo "Filename:"
echo $1
echo "Number of rows:"
cat $1 |wc -l
echo "Number of columns:"
head -n 1 $1 | tr -d ' ' | tr -d '\n' | wc -c
```

```
LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ bash netsize.sh Saavedra2013/n1.txt
Filename:
Saavedra2013/n1.txt
Number of rows:
97
Number of columns:
80
```

MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
  GNU nano 7.2                        netsize_all.sh                        Modified
#!/bin/bash

FILES=../data/Saavedra2013/*.txt
for f in $FILES
do
        myrow=`cat $f | wc -l`
        mycol=`head -n 1 $f | tr -d ' ' | tr -d '\n' | wc -c`
        echo $f $myrow $mycol
done
```


MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data
LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ bash netsize_all.sh
../data/Saavedra2013/n1.txt 97 80
../data/Saavedra2013/n10.txt 14 20
../data/Saavedra2013/n11.txt 270 91
../data/Saavedra2013/n12.txt 7 72
../data/Saavedra2013/n13.txt 61 17
../data/Saavedra2013/n14.txt 35 15
../data/Saavedra2013/n15.txt 38 11
../data/Saavedra2013/n16.txt 118 24
../data/Saavedra2013/n17.txt 76 31
../data/Saavedra2013/n18.txt 13 14
../data/Saavedra2013/n19.txt 10 16
../data/Saavedra2013/n2.txt 62 41
../data/Saavedra2013/n20.txt 18 7
../data/Saavedra2013/n21.txt 19 45
../data/Saavedra2013/n22.txt 19 36
../data/Saavedra2013/n23.txt 179 26
../data/Saavedra2013/n24.txt 80 28
../data/Saavedra2013/n25.txt 17 16
../data/Saavedra2013/n26.txt 82 40
../data/Saavedra2013/n27.txt 27 5
../data/Saavedra2013/n28.txt 90 19
../data/Saavedra2013/n29.txt 61 25
../data/Saavedra2013/n3.txt 25 36
../data/Saavedra2013/n30.txt 8 19
../data/Saavedra2013/n31.txt 28 25
../data/Saavedra2013/n32.txt 45 21
../data/Saavedra2013/n33.txt 70 20
../data/Saavedra2013/n34.txt 79 25
../data/Saavedra2013/n35.txt 14 8
../data/Saavedra2013/n36.txt 40 169
../data/Saavedra2013/n37.txt 44 13
../data/Saavedra2013/n38.txt 51 99
../data/Saavedra2013/n39.txt 33 25
../data/Saavedra2013/n4.txt 101 11
../data/Saavedra2013/n40.txt 28 18
../data/Saavedra2013/n41.txt 12 10
../data/Saavedra2013/n42.txt 42 8
../data/Saavedra2013/n43.txt 55 29
../data/Saavedra2013/n44.txt 56 9
../data/Saavedra2013/n45.txt 36 61
../data/Saavedra2013/n46.txt 58 17
../data/Saavedra2013/n47.txt 139 41
../data/Saavedra2013/n48.txt 118 49
../data/Saavedra2013/n49.txt 47 23

Ejercicio 1.10.4

MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
  GNU nano 7.2                          explore.sh                          Modified
#!/bin/bash

# $1 is the file name
# $2 is the column of interest

echo "Column name"
cut -d ',' -f $2 $1 | head -n 1
echo "Number of distinct values:"
cut -d ',' -f $2 $1 | tail -n +2 | sort |uniq | wc -l
echo "Minium value:"
cut -d ',' -f $2 $1 | tail -n +2 | sort -n | head -n 1
echo "Maxium value:"
cut -d ',' -f $2 $1 | tail -n +2 | sort -n | tail -n 1
```

MINGW64:/c/Users/LENOVO/downloads/stalin/CSB-master/unix/data

```
LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ nano explore.sh

LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ bash explore.sh Buzzard2015_data.csv 7
Column name
biomass
Number of distinct values:
285
Minium value:
1.048466198
Maxium value:
14897.29471
```

```
LENOVO@StalinPC MINGW64 ~/downloads/stalin/CSB-master/unix/data
$ bash explore.sh Buzzard2015_data.csv 9
Column name
SLA
Number of distinct values:
144
Minium value:

Maxium value:
369.6
```