

**Nombre:** Michael Steewart Hurtado Campoverde

1.10.1

1. Change directory to CSB/unix/sandbox.

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02
$ cd CSB/

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB (master)
$ ls
LICENSE      data_wrangling/  good_code/  python/  regex/      sql/
README.md    git/             latex/      r/       scientific/  unix/

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB (master)
$ cd unix

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix (master)
$ cd sandbox/
```

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

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ ls -lh ../data/Marra2014_data.fasta
-rw-r--r-- 1 CORE I5 197121 563K Apr 19 17:02 ../data/Marra2014_data.fasta
```

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

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ cp ../data/Marra2014_data.fasta my_file.fasta

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ ls
'Papers and reviews'/  contador.sh  g02.sh  michael.txt  rutas.txt
buzzard.sh             contador2.sh hola.sh  my_file.fasta  txt
buzzart.sh             familia.txt  michael. pikachu.txt   vagotes.gbi
```

4. How many contigs are classified as isogroup00036?

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ grep -c isogroup00036 my_file.fasta
16
```

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

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ cat my_file.fasta | tr -s ' ' ',' | head -n 3
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG
TTTACAATTAACCCACAAAAGGCTGTTACTGAAGGTGTGGCTTAAGTGTCAAGCAACAG
```

6. How many unique isogroups are in the file?

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ cat my_file.fasta | cut -d ',' -f 4 | head -n 2
>contig00001 length=527 numreads=2 gene=isogroup00001 status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG
```

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

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ 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
```

## 1.10.2

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

3

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ cut -f 1 my_file.fasta | grep -c -w 3
29
```

27

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ cut -f 1 my_file.fasta | grep -c -w 27
9
```

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.

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ nano 2id.sh
```

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ cat 2id.sh
read -p "Ingrese el nombre del archivo: " archivo
read -p "Ingrese el número de ID a buscar: " id
registro=$(grep -n "$id" "$archivo" | cut -d: -f1)
if [ -n "$registro" ]; then
    echo "El número de registro del individuo con el ID $id es: $registro"
else
    echo "No se encontró ningún registro con el ID $id en el archivo $archivo"
fi
```

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ bash 2id.sh
Ingrese el nombre del archivo: my_file.fasta
Ingrese el número de ID a buscar: 1
El número de registro del individuo con el ID 1 es: 1
11
22
33
39
50
106
126
131
225
240
253
264
268
298
313
346
```

3. [Advanced]17 Write a script that returns the number of times each individual was sampled.

```

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ nano pre23.sh

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ cat pre23.sh
echo "Ingrese el nombre del archivo:"
read archivo
echo "Ingrese el número de ID:"
read id
registros=$(tail -n +2 "$archivo" | cut -d ',' -f 1 | sort -n | uniq)
for registro in $registros
do
    num_registros=$(grep -c "^$registro," "$archivo")
    echo "ID: $registro - Registros: $num_registros"
done

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/sandbox (master)
$ bash pre23.sh
Ingrese el nombre del archivo:
my_file.fasta
Ingrese el número de ID:
1
ID: >contig00002 - Registros: 0
ID: length=551 - Registros: 0
ID: numreads=8 - Registros: 0
ID: gene=isogroup00001 - Registros: 0
    - Registros: 0hresh
ID: >contig00003 - Registros: 0
ID: length=541 - Registros: 0
ID: numreads=2 - Registros: 0
ID: gene=isogroup00001 - Registros: 0
    - Registros: 0hresh
ID: >contig00004 - Registros: 0
ID: length=291 - Registros: 0
ID: numreads=3 - Registros: 0
ID: gene=isogroup00001 - Registros: 0
    - Registros: 0hresh
ID: >contig00005 - Registros: 0
ID: length=580 - Registros: 0
ID: numreads=12 - Registros: 0
ID: gene=isogroup00001 - Registros: 0

```

### 1.10.3

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

```

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ cat pre3.sh
echo "Ingrese el nombre del archivo:"
read archivo
num_filas=$(wc -l < "$archivo")
echo "El número de filas en el archivo es: $num_filas"
echo "Ingrese el nombre del archivo:"
read archivo
primera_fila=$(head -n 1 "$archivo")
sin_espacios=$(echo "$primera_fila" | tr -d '[:space:]')
num_columnas=$(echo -n "$sin_espacios" | wc -c)
echo "El número de columnas en la primera fila del archivo es: $num_columnas"

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ bash pre3.sh
Ingrese el nombre del archivo:
n1.txt
El número de filas en el archivo es: 97
Ingrese el nombre del archivo:
n1.txt
El número de columnas en la primera fila del archivo es: 80

```

## 2. [Advanced]18 Write a script that prints the numbers of rows and columns for each network:

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ nano pre32.sh

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ cat pre32.sh
echo "Ingrese el nombre del archivo:"
read archivo
num_filas=$(wc -l < "$archivo")
echo "El número de filas en el archivo es: $num_filas"
for ((i=1; i<=num_filas; i++))
do
    fila=$(head -n $i "$archivo" | tail -n 1)
    sin_espacios=$(echo "$fila" | tr -d '[:space:]')
    num_columnas=$(echo -n "$sin_espacios" | wc -c)
    echo "Número de columnas en la fila $i: $num_columnas"
done

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ bash pre32.sh
Ingrese el nombre del archivo:
n.txt
pre32.sh: line 3: n.txt: No such file or directory
El número de filas en el archivo es:

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ bash pre32.sh
Ingrese el nombre del archivo:
n5.txt
El número de filas en el archivo es: 21
Número de columnas en la fila 1: 7
```

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

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ nano pre33.sh

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ cat pre33.sh
echo "Ingrese el nombre del archivo:"
read archivo
num_filas=$(wc -l < "$archivo")
echo "El número de filas en el archivo es: $num_filas"
max_filas=0
max_columnas=0
for ((i=1; i<=num_filas; i++))
do
    fila=$(head -n $i "$archivo" | tail -n 1)
    sin_espacios=$(echo "$fila" | tr -d '[:space:]')
    num_columnas=$(echo -n "$sin_espacios" | wc -c)
    if ((num_columnas > max_columnas))
    then
        max_columnas=$num_columnas
    fi
    if ((i > max_filas))
    then
        max_filas=$i
    fi
done
echo "El número máximo de filas es: $max_filas"
echo "El número máximo de columnas es: $max_columnas"

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data/Saavedra2013 (master)
$ bash pre33.sh
Ingrese el nombre del archivo:
n5.txt
El número de filas en el archivo es: 21
El número máximo de filas es: 21
El número máximo de columnas es: 7
```

### 1.10.4 Data Explorer

- the corresponding column name;
- the number of distinct values in the column;
- the minimum value;
- the maximum value.

```
CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data (master)
$ nano pre4.sh

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data (master)
$ cat pre4.sh
cut -d ',' -f 7 Buzzard2015_data.csv | head -n 1
cut -d ',' -f 7 Buzzard2015_data.csv | tail -n +2 | sort | uniq | wc -l
cut -d ',' -f 7 Buzzard2015_data.csv | tail -n +2 | sort -n | head -n 1
cut -d ',' -f 7 ../data/Buzzard2015_data.csv | tail -n +2 | sort -n | tail -n 1

CORE I5@MICHAELHURTADO08 MINGW64 ~/Documents/gbi6g02/CSB/unix/data (master)
$ bash pre4.sh
biomass
285
1.048466198
14897.29471
```