Tarea 3: TC3_ForLoop

Nombre: Ronald Rivera
Curso: G02
Docente: Moisés Gualapuro

1.10.1 Next Generation Sequencing Data

1. Change directory to CSB/unix/sandbox.

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                                zsh

ussenterprise@USSs-MacBook-Pro CSB % git checkout -b TC3_ForLoop  master
Switched to a new branch 'TC3_ForLoop'
ussenterprise@USSs-MacBook-Pro CSB % ls
LICENSE         data_wrangling  good_code       python          regex           sql
README.md       git             latex           r               scientific      unix
ussenterprise@USSs-MacBook-Pro CSB % mv unix/data unix/sandbox
ussenterprise@USSs-MacBook-Pro CSB % ls unix
installation    sandbox         solutions
ussenterprise@USSs-MacBook-Pro CSB %
No active issue   Sign in to Bitbucket   0 0   Git Graph   Quokka   Search Engine: Google   Screen Reader Optimized   Go Live   Spell
```

2. What is the size of the fileMarra2014_data.fasta? (553 Kilobytes)

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                                zsh - data

ussenterprise@USSs-MacBook-Pro data % ls
Buzzard2015_about.txt   Gesquiere2011_data.csv  Pacifici2013_about.txt  Saavedra2013_about.txt
Buzzard2015_data.csv    Marra2014_about.txt     Pacifici2013_data.csv   miRNA
Gesquiere2011_about.txt Marra2014_data.fasta    Saavedra2013
ussenterprise@USSs-MacBook-Pro data % ls -lh Marra2014_data.fasta
-rw-r--r--  1 ussenterprise  staff   553K May  9 20:34 Marra2014_data.fasta
ussenterprise@USSs-MacBook-Pro data %
active issue   Sign in to Bitbucket   0 0   Git Graph   Quokka   Search Engine: Google   Screen Reader Optimized   Go Live   Spell
```

3. Create a copy of Marra2014_data.fasta in the sandbox and name it my_file.fasta.

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                                zsh - data

ussenterprise@USSs-MacBook-Pro data % ls
Buzzard2015_about.txt   Gesquiere2011_data.csv  Pacifici2013_about.txt  Saavedra2013_about.txt
Buzzard2015_data.csv    Marra2014_about.txt     Pacifici2013_data.csv   miRNA
Gesquiere2011_about.txt Marra2014_data.fasta    Saavedra2013
ussenterprise@USSs-MacBook-Pro data % cp Marra2014_data.fasta my_file.fasta
ussenterprise@USSs-MacBook-Pro data % ls
Buzzard2015_about.txt   Gesquiere2011_data.csv  Pacifici2013_about.txt  Saavedra2013_about.txt
Buzzard2015_data.csv    Marra2014_about.txt     Pacifici2013_data.csv   miRNA
Gesquiere2011_about.txt Marra2014_data.fasta    Saavedra2013            my_file.fasta
ussenterprise@USSs-MacBook-Pro data %
No active issue   Sign in to Bitbucket   0 0   Git Graph   Quokka   Search Engine: Google   Screen Reader Optimized   Go Live   Spell
```

4. How many contigs are classified asisogroup00036? (16 configs)

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                                zsh - data
ussenterprise@USSs-MacBook-Pro data % grep -c isogroup00036 my_file.fasta
16
ussenterprise@USSs-MacBook-Pro data %
```

5. Replace the original "two-spaces" delimiter with a comma. (los 2 espacios son remplazados por la coma en el cat y el head me permite ver las 3 lineas y validar si en comando funciona)

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                                zsh - data
ussenterprise@USSs-MacBook-Pro data % cat my_file.fasta | tr -s ' ' ',' | head -n 3
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG
TTTACAATTAACCCACAAAAGGCTGTTACTGAAGGTGTGGCTTAAGTGTCAGAGCAACAG
ussenterprise@USSs-MacBook-Pro data %
```

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                                zsh - data
ussenterprise@USSs-MacBook-Pro data % cat my_file.fasta | tr -s ' ' ',' | head -n 3
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG
TTTACAATTAACCCACAAAAGGCTGTTACTGAAGGTGTGGCTTAAGTGTCAGAGCAACAG
ussenterprise@USSs-MacBook-Pro data % cat my_file.fasta | tr -s ' ' ',' > my_file.tmp
ussenterprise@USSs-MacBook-Pro data % ls
Buzzard2015_about.txt   Gesquiere2011_data.csv  Pacifici2013_about.txt  Saavedra2013_about.txt  my_file.tmp
Buzzard2015_data.csv    Marra2014_about.txt     Pacifici2013_data.csv   miRNA
Gesquiere2011_about.txt Marra2014_data.fasta    Saavedra2013            my_file.fasta
ussenterprise@USSs-MacBook-Pro data % cat my_file.tmp | head -n 3
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG
TTTACAATTAACCCACAAAAGGCTGTTACTGAAGGTGTGGCTTAAGTGTCAGAGCAACAG
ussenterprise@USSs-MacBook-Pro data %
```

6. How many unique isogroups are in the file? (43 grupos)

```
PROBLEMS 20   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                                zsh - data
ussenterprise@USSs-MacBook-Pro data % grep 'isogroup' my_file.tmp | cut -d ',' -f 4 | sort | uniq | wc -l
      43
ussenterprise@USSs-MacBook-Pro data %
```

7. Which contig has the highest number of reads (numreads)? How many reads does it have? (la configuracion 00302)

```
PROBLEMS 20   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                    zsh - data
ussenterprise@USSs-MacBook-Pro data % grep 'isogroup' my_file.tmp | cut -d ',' -f 4 | sort | uniq | wc -l
       43
ussenterprise@USSs-MacBook-Pro data % grep 'isogroup' my_file.tmp | cut -d ',' -f 4 | sort -t '=' -k 2 -n -r | head -n 1
gene=isogroup00043
ussenterprise@USSs-MacBook-Pro data % grep 'isogroup' my_file.tmp | cut -d ',' -f 1,3 | sort -t '=' -k 2 -n -r | head -n 1
>contig00302,numreads=3330
ussenterprise@USSs-MacBook-Pro data %
No active issue   Sign in to Bitbucket   ⊗ 0 ⚠ 0 ⓘ 20   Git Graph   Quokka   Spaces: 4   UTF-8   LF   Plain Text   Go Live   20 Spell   Prettier
```

1.10.2 Hormone Levels in Baboons

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

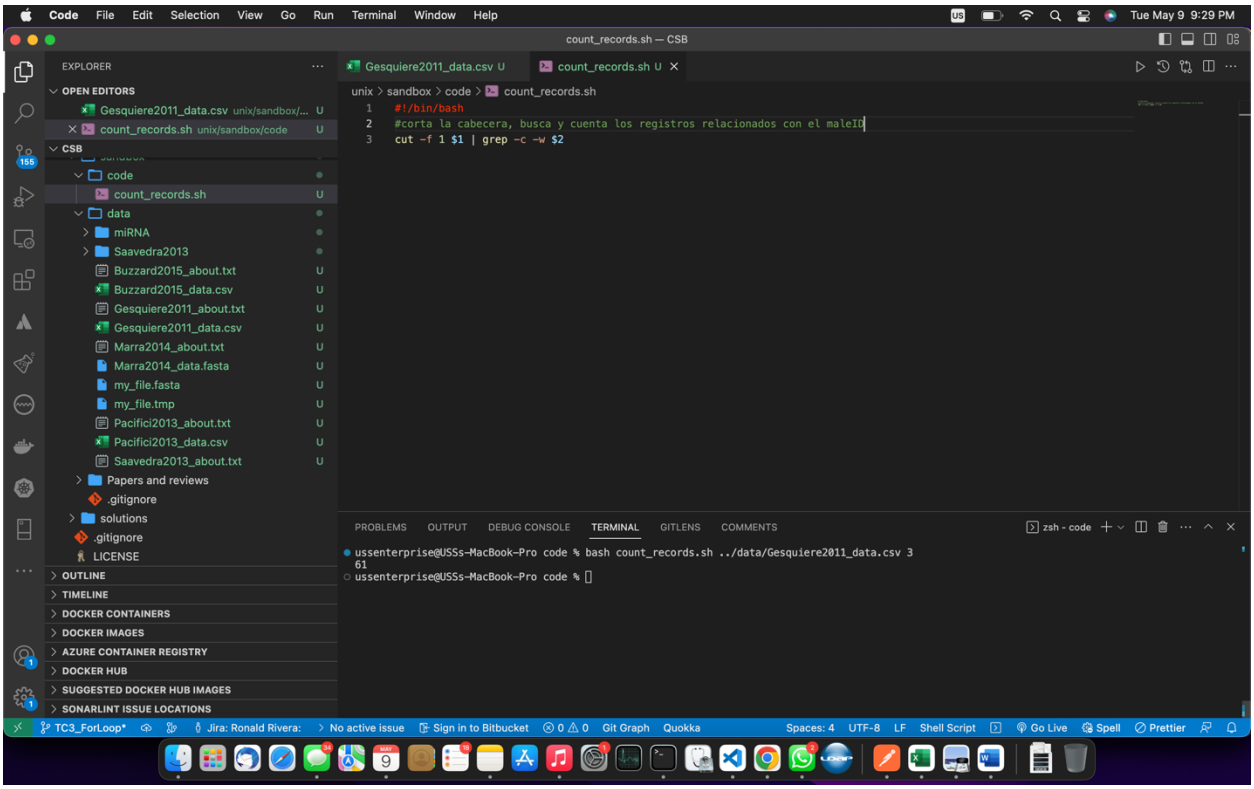
En el nivel 3 hay 61

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   GITLENS   COMMENTS                    zsh - data
ussenterprise@USSs-MacBook-Pro data % cut -f 1  Gesquiere2011_data.csv | grep -c -w 3
61
ussenterprise@USSs-MacBook-Pro data %
active issue   Sign in to Bitbucket   ⊗ 0 ⚠ 0   Git Graph   CSVLint   Query   Align   Rainbow OFF   Quokka   LF   TSV   Go Live   Spell   Prettier
```

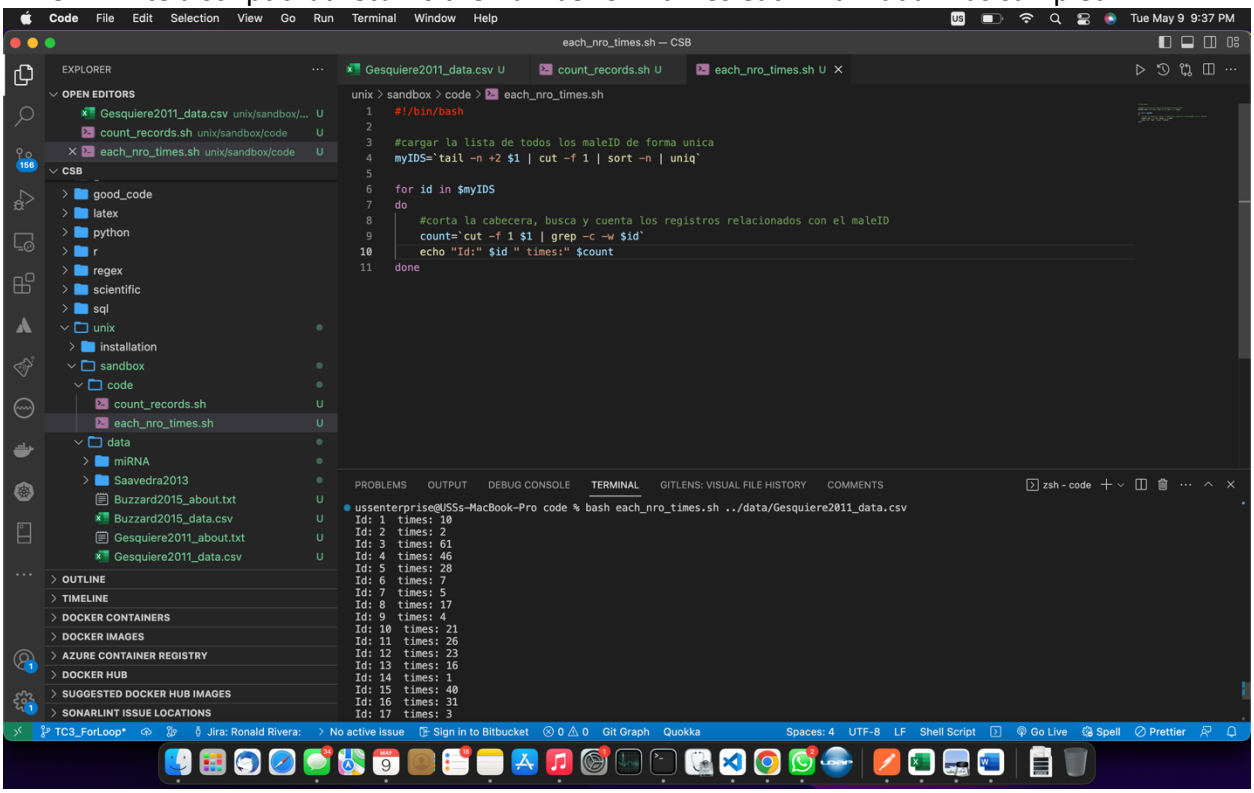
En el nivel 27 hay 5

```
Command executed 1 min ago   G CONSOLE   TERMINAL   GITLENS   COMMENTS                    zsh - data
ussenterprise@USSs-MacBook-Pro data % cut -f 1  Gesquiere2011_data.csv | grep -c -w 3
61
ussenterprise@USSs-MacBook-Pro data % cut -f 1  Gesquiere2011_data.csv | grep -c -w 27
5
ussenterprise@USSs-MacBook-Pro data %
active issue   Sign in to Bitbucket   ⊗ 0 ⚠ 0   Git Graph   CSVLint   Query   Align   Rainbow OFF   Quokka   LF   TSV   Go Live   Spell   Prettier
```

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.

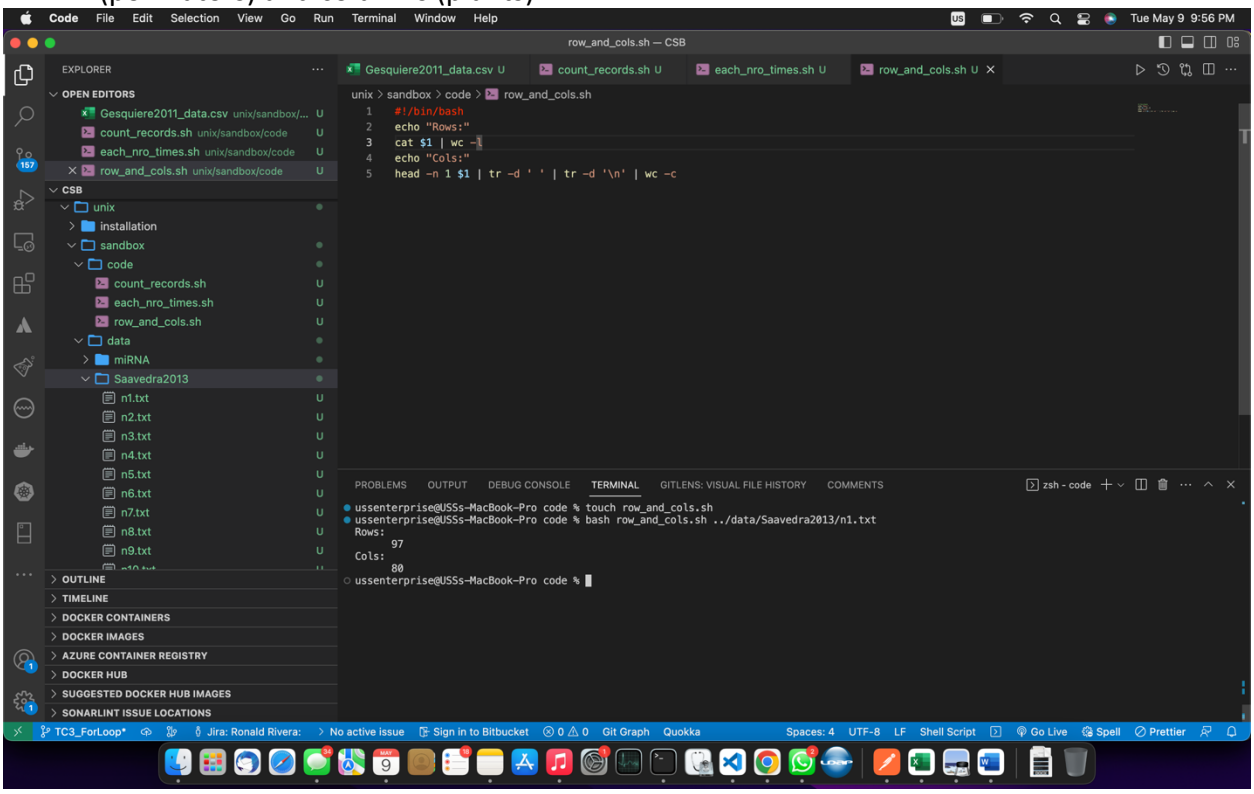


3. Write a script that returns the number of # times each individual was sampled.

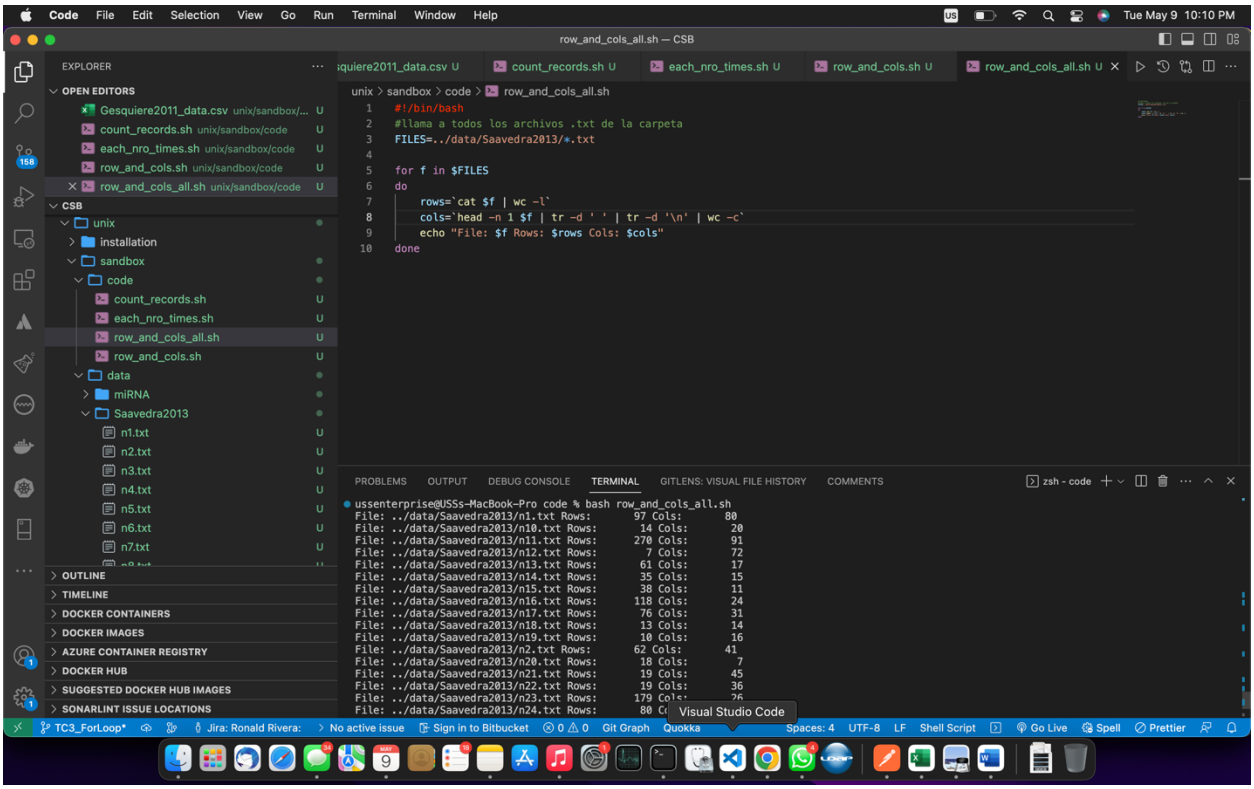


1.10.3 Plant–Pollinator Networks

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

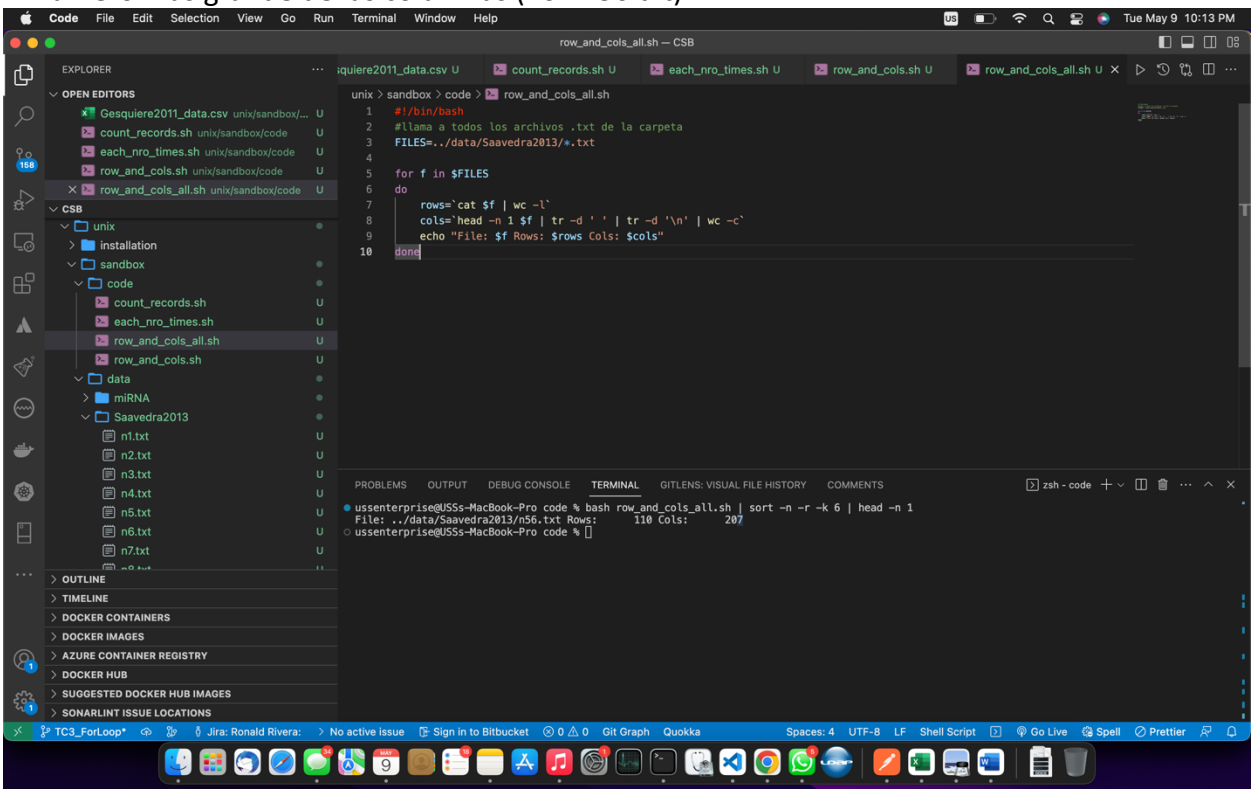


2. Write a script that prints the number of rows and columns for each network

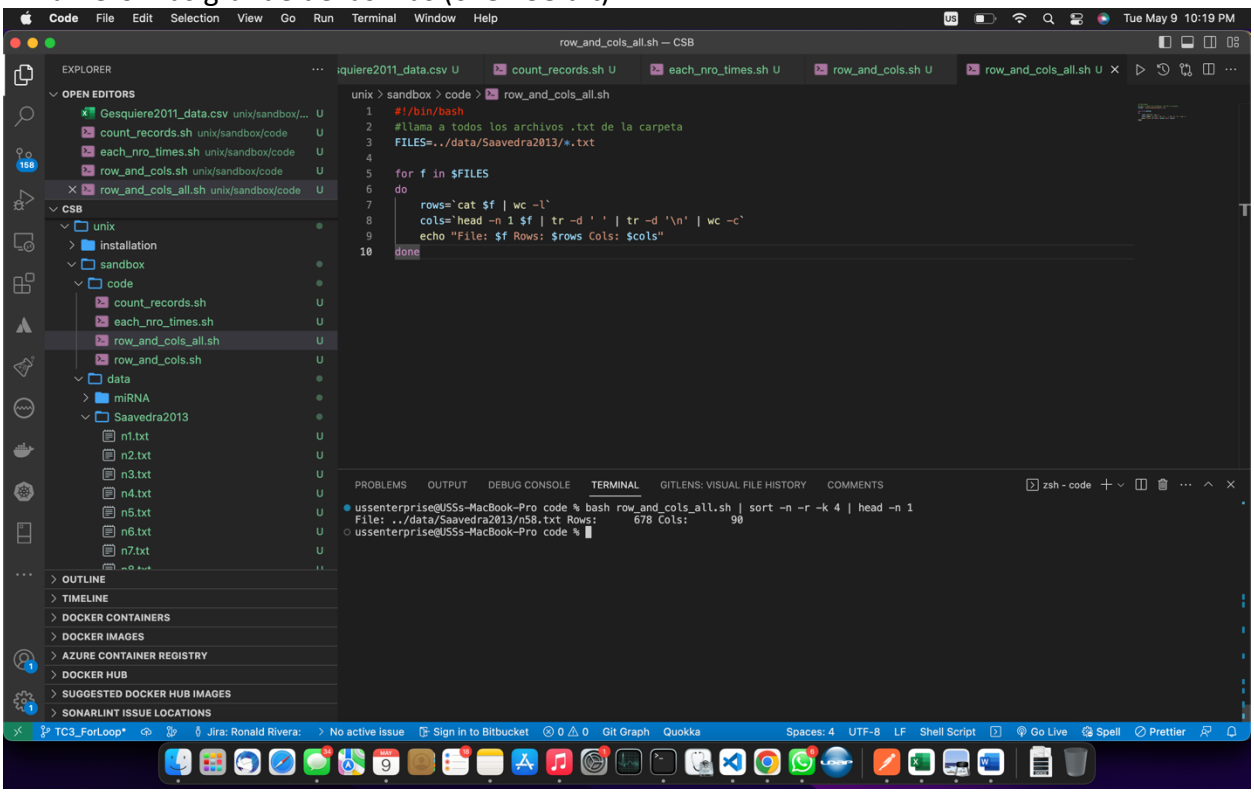


3. Which is the network with the largest number of rows? Which the one with the largest number of columns?

El número más grande de las columnas (207 n56.txt)



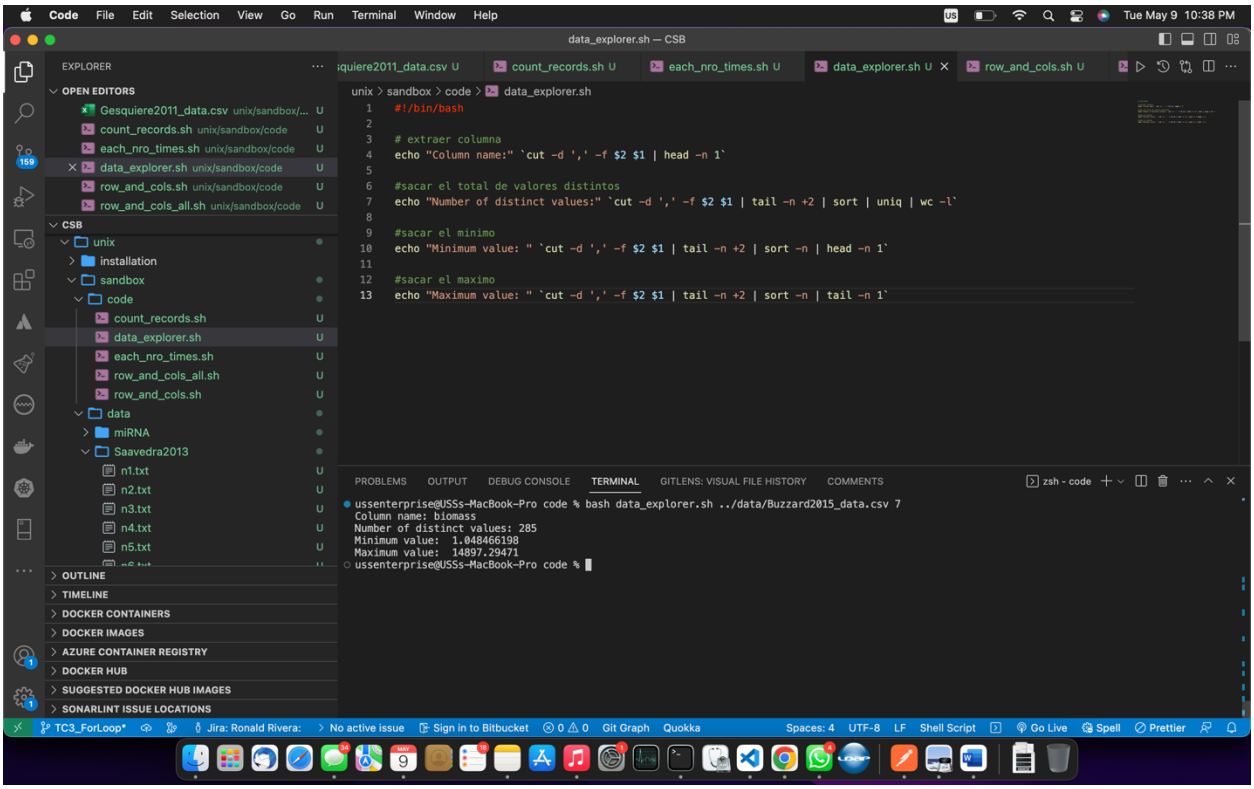
El número más grande de las filas (678 n58.txt)



1.10.4 Data Explorer

1. Write a script that, for a given csv file and column number, prints:

- Column name
- Number of distinct values
- Minimum value
- Maximum value



Scripts subidos:

https://github.com/ronaldsoft/CSB/tree/TC3_ForLoop/unix/sandbox/code

Rama: TC3_ForLoop