Part1: Intermezzo 1.4
// a. Find Dalziel csv file location and copy it to sandbox
```
(base) jiayiji@MacBook-Pro-106 data % cd ~/CSB/unix/sandbox
(base) jiayiji@MacBook-Pro-106 sandbox % find ../../ -type f -name "*Dalziel*.csv"
../../python/data/Dalziel2016_data.csv
(base) jiayiji@MacBook-Pro-106 sandbox % cp ../../python/data/Dalziel2016_data.csv
Dalziel2016_data_copy.csv
```
// b. Print the first few line. List all unique cities in its file name.
```
(base) jiayiji@MacBook-Pro-106 sandbox % head -n 5 Dalziel2016_data_copy.csv
biweek,year,loc,cases,pop
1,1906,BALTIMORE,NA,526822.1365
2,1906,BALTIMORE,NA,526995.246
3,1906,BALTIMORE,NA,527170.1981
4,1906,BALTIMORE,NA,527347.0136
(base) jiayiji@MacBook-Pro-106 sandbox % cut -d "," -f 3 Dalziel2016_data_copy.csv| tail
-n +2| uniq
BALTIMORE
BOSTON
BRIDGEPORT
BUFFALO
CHICAGO
CINCINNATI
CLEVELAND
COLUMBUS
DENVER
DETROIT
DULUTH
FALL RIVER
GRAND RAPIDS
HARTFORD
INDIANAPOLIS
KANSAS CITY
LOS ANGELES
MILWAUKEE
MINNEAPOLIS
NASHVILLE
NEW HAVEN
NEW ORLEANS
NEW YORK
NEWARK
PHILADELPHIA
PITTSBURGH
PROVIDENCE
READING.US
RICHMOND
ROCHESTER
SALT LAKE CITY
SAN FRANCISCO
SEATTLE
SPOKANE
SPRINGFIELD
ST LOUIS
TOLEDO
TRENTON
WASHINGTON
WORCESTER
(base) jiayiji@MacBook-Pro-106 sandbox % cut -d "," -f 3 Dalziel2016_data_copy.csv| tail
-n +2| wc -l
   44720
```
// Each city occurs 1118 times in the dataset.

// 2c. Find the largest number of Measles cases reported in Washington Dc.
```
(base) jiayiji@MacBook-Pro-106 sandbox % grep -w 'WASHINGTON' Dalziel2016_data_copy.csv |
cut -d "," -f 4 | sort -nr | head -n 1
1559
```
// 2d. Maximum number of reported cases in the data set is 13226, it occurs in NewYork
```
(base) jiayiji@MacBook-Pro-106 sandbox % cut -d "," -f 4 Dalziel2016_data_copy.csv| sort
-nr | head -n 1
13226
(base) jiayiji@MacBook-Pro-106 sandbox % grep '13226' -w Dalziel2016_data_copy.csv
7,1941,NEW YORK,13226,7507365.374
```

Part2
/ 2A: navigate to sandbox folder
```
(base) jiayiji@MacBook-Pro-106 ~ % cd ~/CSB/unix/sandbox
```
// 2B: making a HW1 directory under CSB/unix/sandbox
```
(base) jiayiji@MacBook-Pro-106 sandbox % mkdir HW1
```
// 2C: Navigates to HW1 directory
```
(base) jiayiji@MacBook-Pro-106 sandbox % cd HW1
```
// 2D: Create hw1_q3_file.txt
```
(base) jiayiji@MacBook-Pro-106 HW1 % touch hw1_q3_file.txt
```
// 2E: Open file in nano and typed "Homework1 Question3"
```
(base) jiayiji@MacBook-Pro-106 HW1 % nano hw1_q3_file.txt
```
// 2F: Print current working directory and append to file
```
(base) jiayiji@MacBook-Pro-106 HW1 % pwd >> hw1_q3_file.txt
```
// 2G: Print all content of the file in the terminal (content not shown here, see uploaded hw1_q3_file.txt)
```
(base) jiayiji@MacBook-Pro-106 HW1 % less hw1_q3_file.txt
```
// 2H: Check permission and appened the current permission to the file
```
(base) jiayiji@MacBook-Pro-106 HW1 % ls -l hw1_q3_file.txt
-rw-r--r--@ 1 jiayiji  staff  58 Sep  3 18:02 hw1_q3_file.txt
(base) jiayiji@MacBook-Pro-106 HW1 % ls -l hw1_q3_file.txt >> hw1_q3_file.txt
```
// 2I: Change the permission
```
(base) jiayiji@MacBook-Pro-106 HW1 % chmod u+rw,g+rw,o-rwx hw1_q3_file.txt
(base) jiayiji@MacBook-Pro-106 HW1 % ls -l hw1_q3_file.txt
-rw-rw----@ 1 jiayiji  staff  120 Sep  3 18:06 hw1_q3_file.txt
```
// 2J: Print the updated permission to hw1 file
```
(base) jiayiji@MacBook-Pro-106 HW1 % ls -l hw1_q3_file.txt >> hw1_q3_file.txt
```
// 2K: Select column 1-6 from Pacifici2013_data.csv and replace all ";" with ",". Then append the data to hw1 file.
```
(base) jiayiji@MacBook-Pro-106 data % cut -d ";" -f 1-6 Pacifici2013_data.csv| tr ";" ","
| tail -n +2 >> ../../unix/sandbox/HW1/hw1_q3_file.txt
```
// 2K1: Find the number of rows in the file (without header) and append to hw1 file.
```
(base) jiayiji@MacBook-Pro-106 data % tail -n +2 Pacifici2013_data.csv| wc -
l >> ../../unix/sandbox/HW1/hw1_q3_file.txt
```
// 2K2: Find the 5 largest values of TAXID and append to hw1 file.
```
(base) jiayiji@MacBook-Pro-106 data % cut -d ";" -f 1 Pacifici2013_data.csv | tail -n +2
| sort | uniq > sorted.txt
(base) jiayiji@MacBook-Pro-106 data % head -n 5 sorted.txt
9977
9978
9979
9987
9997
(base) jiayiji@MacBook-Pro-106 data % tail -n -5
sorted.txt >> ../../unix/sandbox/HW1/hw1_q3_file.txt
```

Part3
1.10.1
// 1. Change directory to sandbox
(base) jiayiji@MacBook-Pro-106 ~ % cd ~/CSB/unix/sandbox
// 2. Size of file is 553K
(base) jiayiji@MacBook-Pro-106 sandbox % ls -lh my_file.fasta
-rw-r--r-- 1 jiayiji staff   553K Sep 4 18:17 my_file.fasta
// 3. Create a copy of fasta file and rename it.
(base) jiayiji@MacBook-Pro-106 sandbox % cp ../../unix/data/Marra2014_data.fasta
my_file.fasta
// 4. 16 Contigs are classified as isogroup00036
(base) jiayiji@MacBook-Pro-106 sandbox % grep -c isogroup00036 my_file.fasta
16
(base) jiayiji@MacBook-Pro-106 sandbox % grep isogroup00036 my_file.fasta | wc -l
16
// 5. Replace "two-space" delimiter with comma. (directly overwriting the file did not
work out for me, so I created a tmp file and moved it later)
(base) jiayiji@MacBook-Pro-106 sandbox % cat my_file.fasta | tr -s ' ' ',' > my_file.tmp
(base) jiayiji@MacBook-Pro-106 sandbox % mv my_file.tmp my_file.fasta
// 6. There are 43 unique isogroups
(base) jiayiji@MacBook-Pro-106 sandbox % grep '>' my_file.fasta | cut -d ',' -f 4 | sort
| uniq | wc -l
43
// 7. contig00302 has the highest number of reads, which is 3330.
// Use -t '=' to split, then -k 2 to select the second item, and finally -n -r to sort
numerically in reverse order
(base) jiayiji@MacBook-Pro-106 sandbox % grep '>' my_file.fasta | cut -d ',' -f 1,3 |
sort -t '=' -k 2 -n -r | head -n 1
>contig00302,numreads=3330

1.10.2
// 1. Individual #3 appeared 61 times, individual #27 appeared 27 times
(base) jiayiji@MacBook-Pro-106 data % cut -f 1 Gesquiere2011_data.csv| grep -c -w 3
61
(base) jiayiji@MacBook-Pro-106 data % cut -f 1 Gesquiere2011_data.csv| grep -c -w 27
5
// 2. Script written in nano that count number of record for each ID
(base) jiayiji@MacBook-Pro-106 data % touch count_baboons.sh
(base) jiayiji@MacBook-Pro-106 data % nano count_baboons.sh

```
    // script (the script is not written in terminal but through nano)
        #!/bin/bash
        cut -f 1 $1 | grep -c -w $2
    // end
```

(base) jiayiji@MacBook-Pro-106 data % touch count_all_baboons.sh
(base) jiayiji@MacBook-Pro-106 data % nano count_all_baboons.sh

```
    // script (the script is not written in terminal but through nano)
        #!/bin/bash
        myIDS=`tail -n +2 ../data/Gesquiere2011_data.csv | cut -f 1 | sort -n | uniq`
        for id in $myIDS
        do
            mycounts=`bash count_baboons.sh ../data/Gesquiere2011_data.csv $id`
            echo "ID:" $id "counts:" $mycounts
        done
    // end
```

(base) jiayiji@MacBook-Pro-106 data % bash count_all_baboons.sh
ID: 1 counts: 10
ID: 2 counts: 2
... (total 127 outputs)

1.10.3
// 1. determining the number of rows and cols of all files in a directory
```
(base) jiayiji@MacBook-Pro-106 data % touch netsize.sh
(base) jiayiji@MacBook-Pro-106 data % touch netsize_all.sh
(base) jiayiji@MacBook-Pro-106 data % nano netsize.sh
```

// script

```
#!/bin/bash
echo "Filename:"
echo $1
echo "Number of rows:"
cat $1 | wc -l
echo "Number of columns:"
ead -n 1 $1 | tr -d ' ' | tr -d '\n' | wc -c
```

// end

```
(base) jiayiji@MacBook-Pro-106 data % nano netsize_all.sh
```

// script

```
#!/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
```

// end

```
(base) jiayiji@MacBook-Pro-106 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
... (total 59 outputs)
```