BINF2111 – Introduction to Bioinformatics Computing
BASH 101 – Loops part duex

Richard Allen White III, PhD
RAW Lab
Lecture 10 – Tuesday Sep 28th, 2021

Learning Objectives

- Review quiz

- Review bonus (SAM files)

- File conversions (awk/perl)

- Specific lines UNIX/BASH

- Review Bash for loops (C-style/Regular)

- Quiz 10

Write a one-liner that counts the number of times Steven is left of Jose?

more file.tsv
Steven Jose
Steven Jose
Steven Jose
Steven Jose

Answer:

Quiz 10 answers

Write a one-liner that counts the number of times Steven is left of Jose?

more file.tsv
Steven Jose
Steven Jose
Steven Jose
Steven Jose

Answer: egrep -o 'Steven.Jose' file.tsv | wc -l

Does awk '/[Ss]teven\tJose' file.tsv | wc -l work?

Quiz 10 answers

Write a one-liner that counts the number of times Steven is left of Jose?

more file.tsv
Steven Jose
Steven Jose
Steven Jose
Steven Jose

Answer: egrep -o 'Steven.Jose' file.tsv | wc -l

Does awk '/[Ss]teven\tJose' file.tsv | wc -l work?
NOPE

Write a one-liner that counts the number of times Steven is left of Jose?

more file.tsv
Steven Jose
Steven Jose
Steven Jose
Steven Jose

Answer: egrep -o 'Steven.Jose' file.tsv | wc -l

Does awk '/[Ss]teven\tJose' file.tsv | wc -l work?
awk '/[Ss]teven\tJose/' file.tsv | wc -l

Write a one-liner that counts the number of times Steven is left of Jose?

more file.tsv
Steven Jose
Steven Jose
Steven Jose
Steven Jose

Answer: egrep -o 'Steven.Jose' file.tsv | wc -l

Another way?

Write a one-liner that counts the number of times Steven is left of Jose?

more file.tsv
Steven Jose
Steven Jose
Steven Jose
Steven Jose

Answer: egrep -o 'Steven.Jose' file.tsv | wc -l

Another way?
egrep -o '[Ss]teven'$'\t''[Jj]ose' file.tsv | wc -l (" = ' ' as single quotes)

Quiz 10 answers

Provide a unique command to convert a tsv (tab) to csv (comma)?

cat file.tsv | tr -s ',' '\t' >file.csv
sed 's/,/\t/g' file.tsv >file.csv
cat file.tsv | tr -s '\t' ',' >file.csv
sed 's/,/**\t**/g' file.tsv >file.csv

Answer: egrep -o 'Steven.Jose' file.tsv | wc -l

Another way?
egrep -o '[Ss]teven'$'\t"[Jj]ose' file.tsv | wc -l (**" = ' '** as single quotes)

>chr1_geneA
ATGCTAAGGCTATCTTGACAACTGACTGCCTAG
>chr1_geneB
ATGCTAAGGCT**ATG**TTGGCAACTGACTCCCTAG

How do I extract this sequence with grep?
Answer:

>chr1_geneA

ATGCTAAGGCTATCTTGACAACTGACTGCCTAG

>chr1_geneB

ATGCTAAGGCT**ATG**TTGGCAACTGACTCCCTAG

How do I extract this sequence with grep?

Answer: grep 'ATG.*ATG' file.fna

What does this give me?

grep 'ATG|ATG' file.fna

Provide a unique command to convert a tsv (tab) to csv (comma)?

A) cat file.tsv | tr -s ',' '\t' >file.csv
B) sed 's/,/\t/g' file.tsv >file.csv
C) cat file.tsv | tr -s '\t' ',' >file.csv
D) sed 's/,/**\t**/g' file.tsv >file.csv

Provide a unique command to convert a tsv (tab) to csv (comma)?

A) cat file.tsv | tr -s ',' '\t' >file.csv
B) sed 's/,/\t/g' file.tsv >file.csv
C) cat file.tsv | tr -s '\t' ',' >file.csv
D) sed 's/,/**\t**/g' file.tsv >file.csv

Another way?

Quiz 10 answers

Provide a unique command to convert a tsv (tab) to csv (comma)?

A) cat file.tsv | tr -s ',' '\t' >file.csv
B) sed 's/,/\t/g' file.tsv >file.csv
C) cat file.tsv | tr -s '\t' ',' >file.csv
D) sed 's/,/**\t**/g' file.tsv >file.csv

Another way?
sed 's/\t/,/g' file.tsv >file.csv

The file conversation problem

Provide unique commands to convert a tsv to csv?

awk '{**gsub**("\t", ","); print}' file.tsv

Try the name_game.tsv and tsv file.

Also, it works similar to sed and similar syntax to tr (translate):
awk '{gsub("original", "replace"); print}' file.txt
OR
awk '{gsub(/original/, "replace"); print}' file.txt

Another way?

The file conversation problem

Provide unique commands to convert a tsv to csv?

 perl -pi -e 's/\t/,/g' name_game.tsv

Try the name_game.tsv and csv file.

Also, it works just like sed try this!

sed 's/\t/,/' name_game.tsv

OR

perl -pi -e 's/\t/,/' name_game.tsv

Bonus 8

- Use grep to convert sam file to a fastq file?

Bonus 8

- Use grep to convert sam file to a fastq file?

What is a SAM file?

The **Sequence Alignment/Map (SAM)** is a file format to save alignment information of short reads mapped against reference sequences. It usually starts with a header section followed by alignment information as tab separated lines for each read.

Bonus 8

- Use grep to convert sam file to a fastq file?

Header section
@HD   VN:1.3   SO:coordinate

Tab-delimited read alignment information lines
readID43GYAX15:7:1:1202:19894/1   256   contig43   613960   1   65M   *   0   0
CCAGCGCGAACGAAATCCGCATGCGTCTGGTCGTTGCACGGAACGGCGGCGGTGTGATGC
ACGGC   EDDEEDEE=EE?DE??DDDBADEBEFFFDBEFFEBCBC=?BEEEE@=:?::?7?:8-6?7?
@??#   AS:i:0   XS:i:0 XN:i:0 XM:i:0 XO:i:0 XG:i:0 NM:i:0 MD:Z:65 YT:Z:UU

Bonus 8

- Use grep to convert sam file to a fastq file?
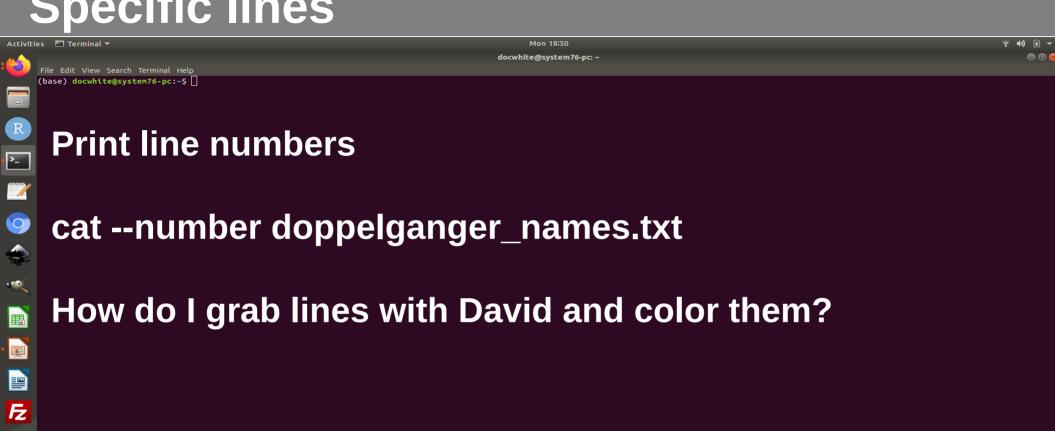
cat samplename.nomapping.sam | grep -v ^@ | awk '{print "@"$1"\n"$10"\n+\n"$11}' > unmapped/samplename.fastq

Specific lines in UNIX/BASH

- How do I print line numbers?

- What if you need specific lines in a file?

- What how do you do a range of specific lines?

Specific lines

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

Print line numbers

cat --number doppelganger_names.txt

How do I grab lines with David and color them?

Specific lines

```
(base) docwhite@system76-pc:~$ ▯
```

Print line numbers

cat --number doppelganger_names.txt

How do I grab lines with David and color them?

cat --number doppelganger_names.txt | grep "david" --color

Specific lines

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

How do I grab line one only?

Specific lines

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$

How do I grab line one only?

sed -n '1p' doppelganger_names.txt

OR

head -1 doppelganger_names.txt | tail +1

Specific lines

How do print the line number 1 only and grab david with color?

Specific lines

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

How do print the line number 1 only and grab david with color?

cat --number doppelganger_names.txt | sed -n '1p' | grep "david" –color

How do I do it with head/tail command?

Specific lines

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

How do print the line number 1 only and grab david with color?

cat --number doppelganger_names.txt | sed -n '1p' | grep "david" –color

How do I do it with head/tail command?
cat --number doppelganger_names.txt | head -1 | tail +1 | grep "david" --color

Specific lines

How do print a range of lines?

Specific lines

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

How do print a range of lines?

sed -n '3,7p' doppelganger_names.txt

awk 'NR>=3 && NR<=7' doppelganger_names.txt

head -7 doppelganger_names.txt | tail +3

Specific lines

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$

How do print a range of lines with line numbers?

Specific lines

docwhite@system76-pc: ~

File   Edit   View   Search   Terminal   Help

(base) docwhite@system76-pc:~$ ☐

How do print a range of lines with line numbers?

cat --number doppelganger_names.txt | awk 'NR>=3 && NR<=7'

cat --number doppelganger_names.txt | sed -n '3,7p'

cat --number doppelganger_names.txt | head -7 | tail +3

What if I do?
cat --number doppelganger_names.txt | head -7 | tail +3 | grep "david" --color

Specific lines

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

What if I do?

cat --number doppelganger_names.txt | head -7 | tail +3 | grep "david" –color

```
4   david    abdul    chi  bill
5   mary     david    bill abdul
7   david    abdul    chi  bill
```
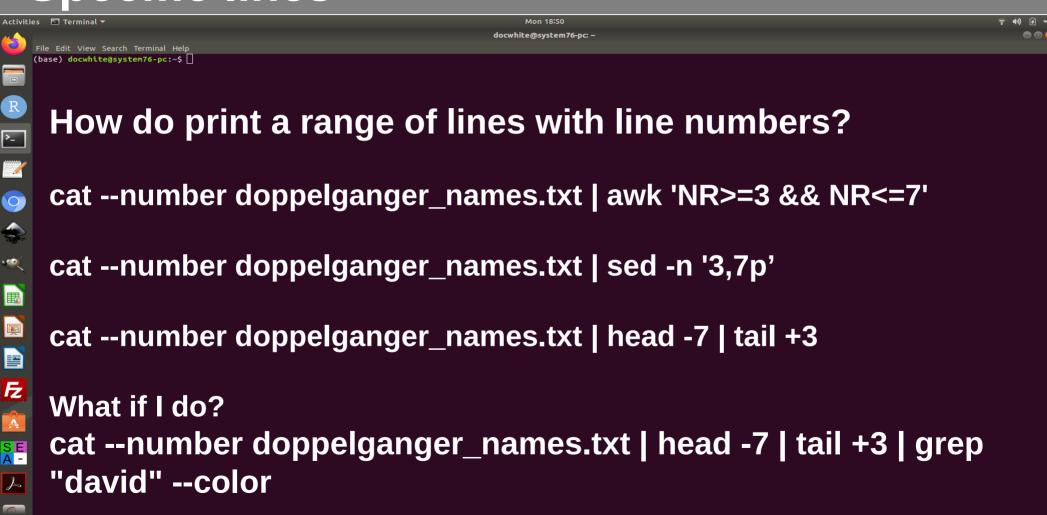
```
for i in file.*;do
command $i
done
```

BASH for loops

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

```
#!/bin/bash
for i in 1 2 3 4 5
do
   echo "Welcome $i times"
done
```

Make a bash script then make it executable

Can you improve it?

BASH for loops

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

```
#!/bin/bash
for i in {1..5}
do
   echo "Welcome $i times"
done
```

Make a bash script then make it executable

```
for ((i = 0 ; i < 100 ; i++)); do
  command $i
done
```

BASH for loops

docwhite@system76-pc: ~

File Edit View Search Terminal Help

(base) docwhite@system76-pc:~$ ▯

```bash
#!/bin/bash
for (( c=1; c<=5; c++ ))
do
   echo "Welcome $c times"
done
```

Make a bash script then make it executable

Quiz 10

- On canvas now

Bonus 9

- Write a bash script that prints the working directory, counts all the sequences within a fasta files within the working directory, and prints the first five lines of the file into std_out.txt?