# **BINF2111 - Introduction to Bioinformatics Computing**

**BASH 101 - Loops part duex** 



Richard Allen White III, PhD RAW Lab

Lecture 10 - Tuesday Sep 28<sup>th</sup>, 2021

# Learning Objectives

- Lab assignment clarifications
- Review quiz/bonus (SAM files)
- File conversions (awk/perl)
- Specific lines UNIX/BASH
- Review Bash for loops (C-style/Regular)
- Quiz 10

# Lab assignment clarifications

When I say write a bash script it on a question that means a script that runs in the terminal as an executable via

./script.sh

bash script.sh

OR

points will be removed if the scripts aren't attached in the future.

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:

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?

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?

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)
```

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
ATGCTAAGGCTATGTTGGCAACTGACTCCCTAG
```

How do I extract this sequence with grep? Answer:

>chr1\_geneA ATGCTAAGGCTATCTTGACAACTGACTGCCTAG >chr1\_geneB ATGCTAAGGCTATGTTGGCAACTGACTCCCTAG

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?

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

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

- 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.

- 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 ACGC 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
```

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

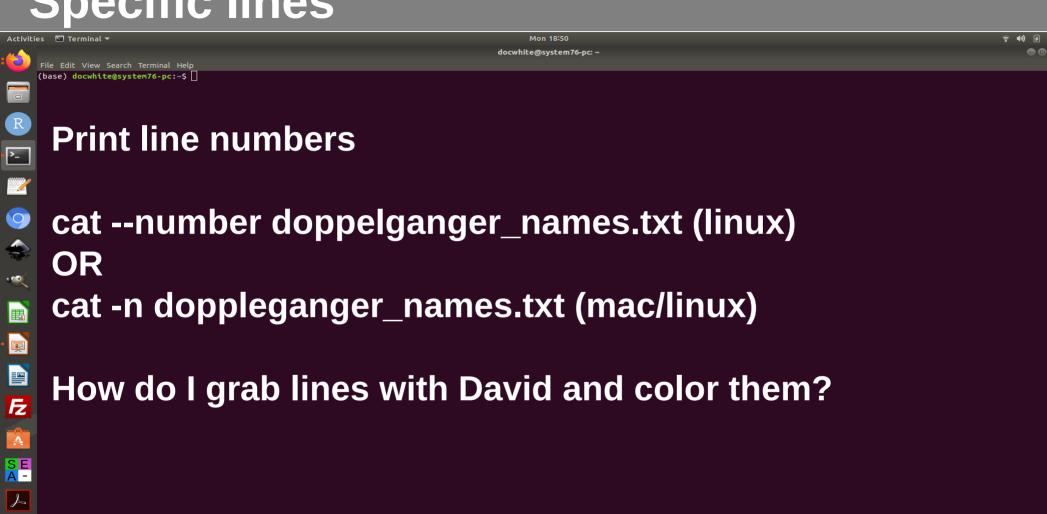
cat samplename.nomapping.sam | grep -v  $^@$  | awk '{print "@"\$1"\n"\$10"\n+\n"\$11}' > unmapped/samplename.fastq

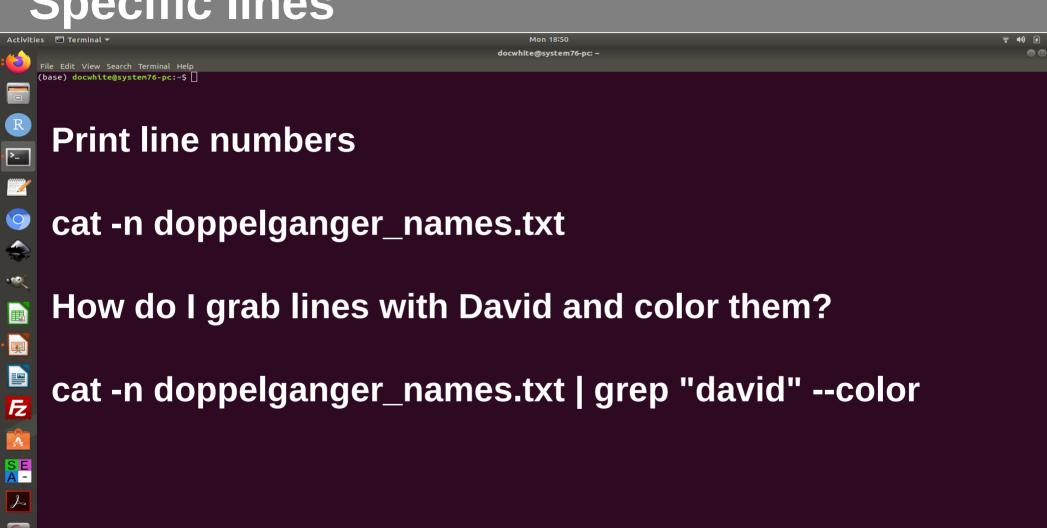
# Specific lines in UNIX/BASH

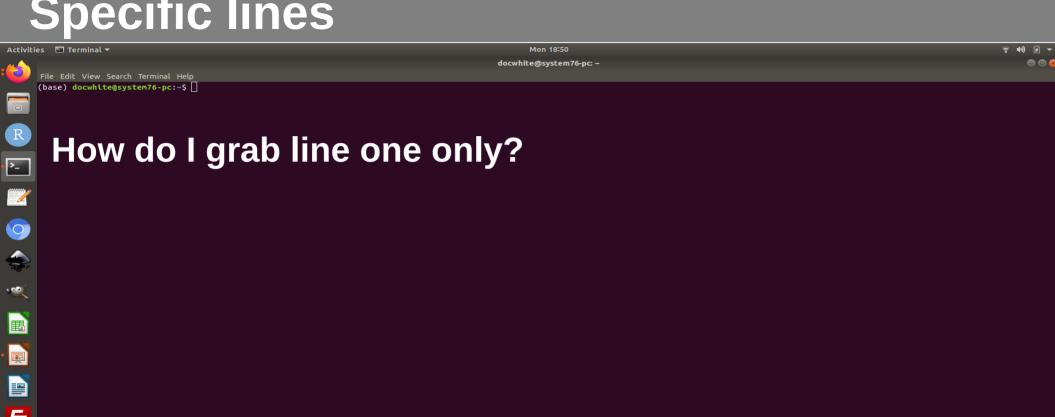
- How do I print line numbers?

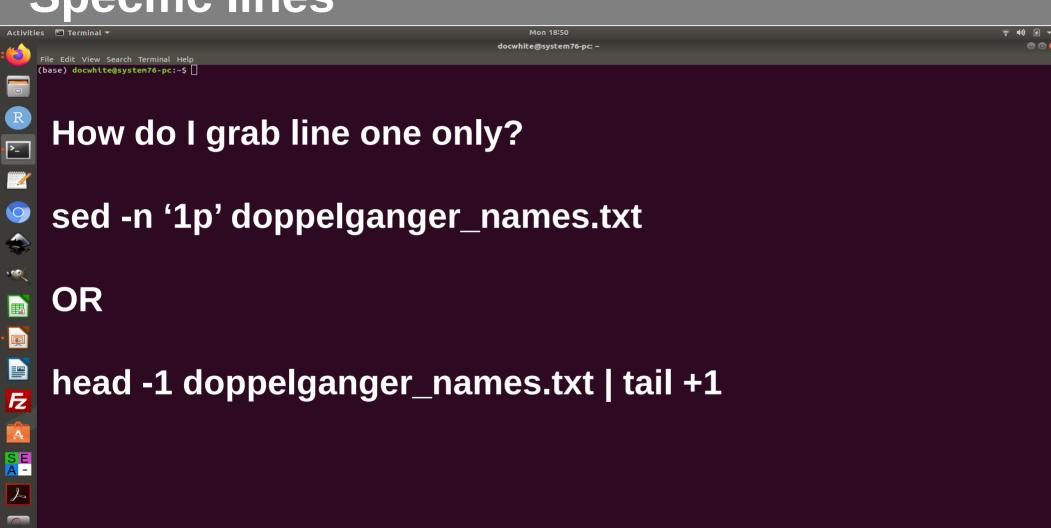
- What if you need specific lines in a file?

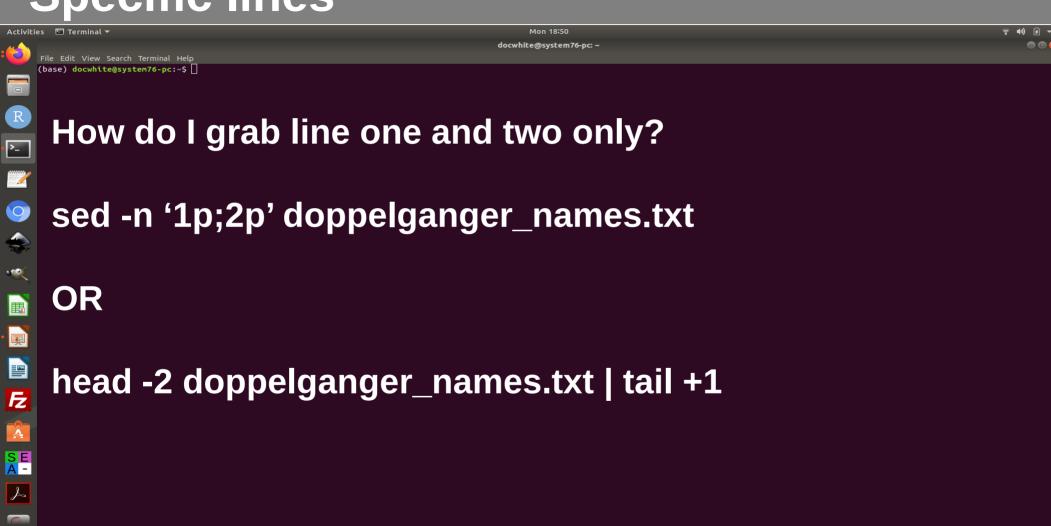
- What how do you do a range of specific lines?

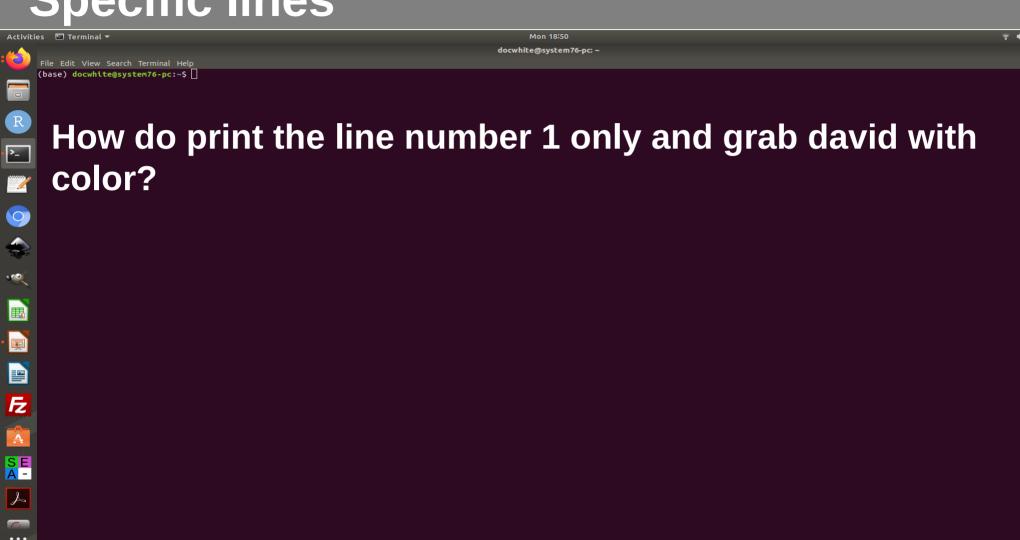


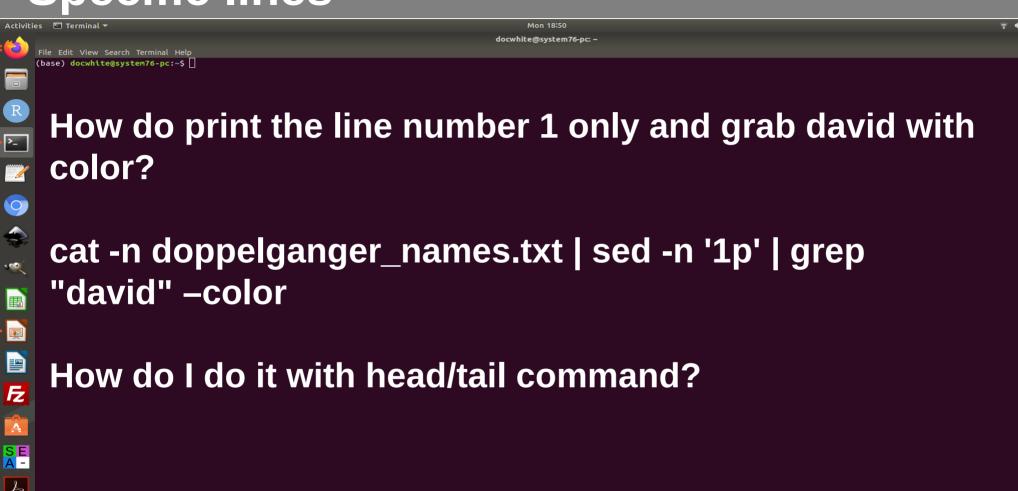


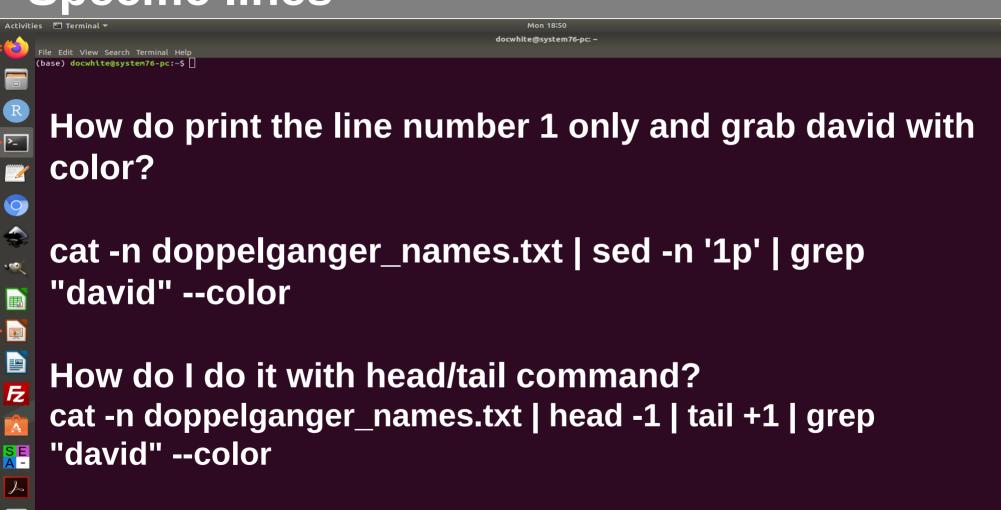


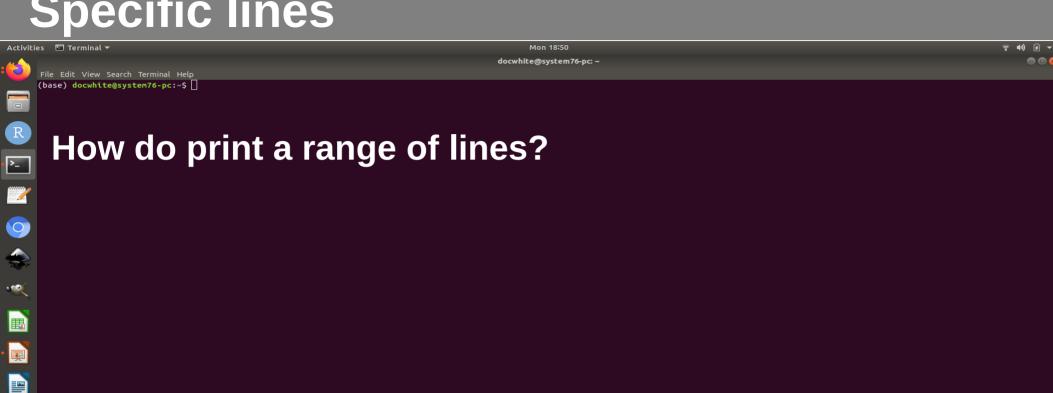


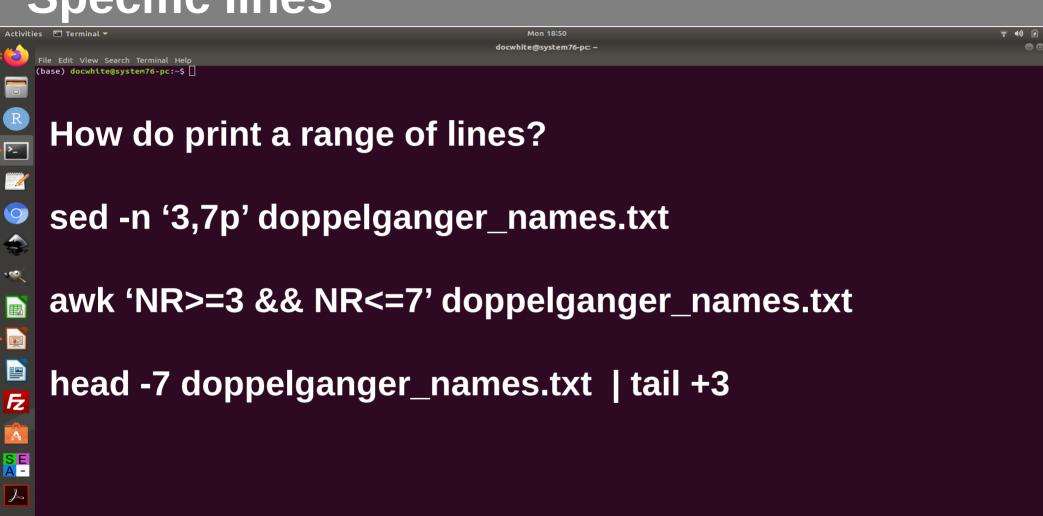


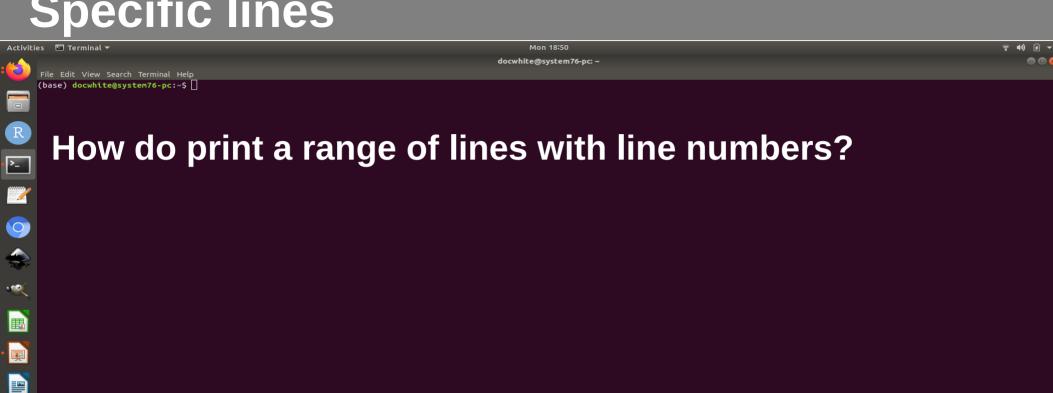


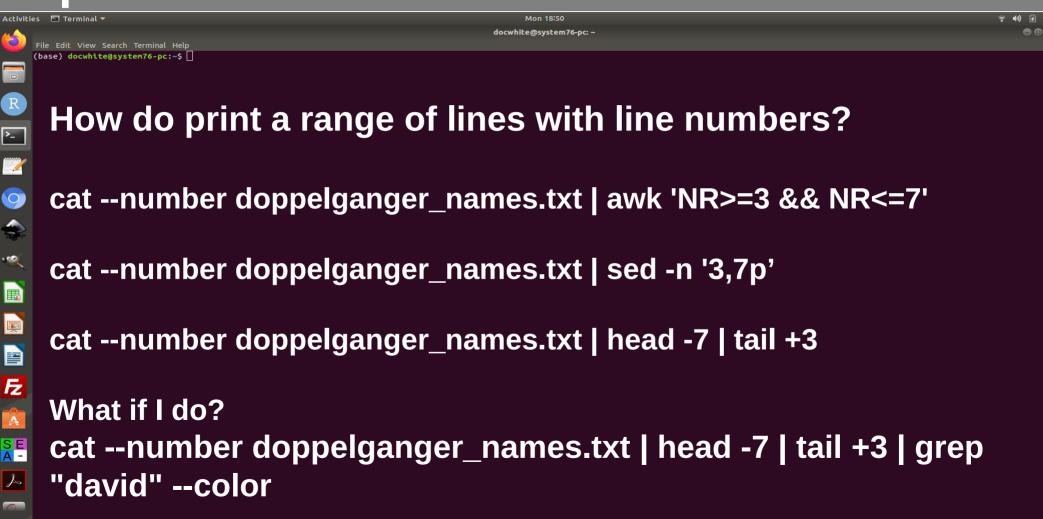


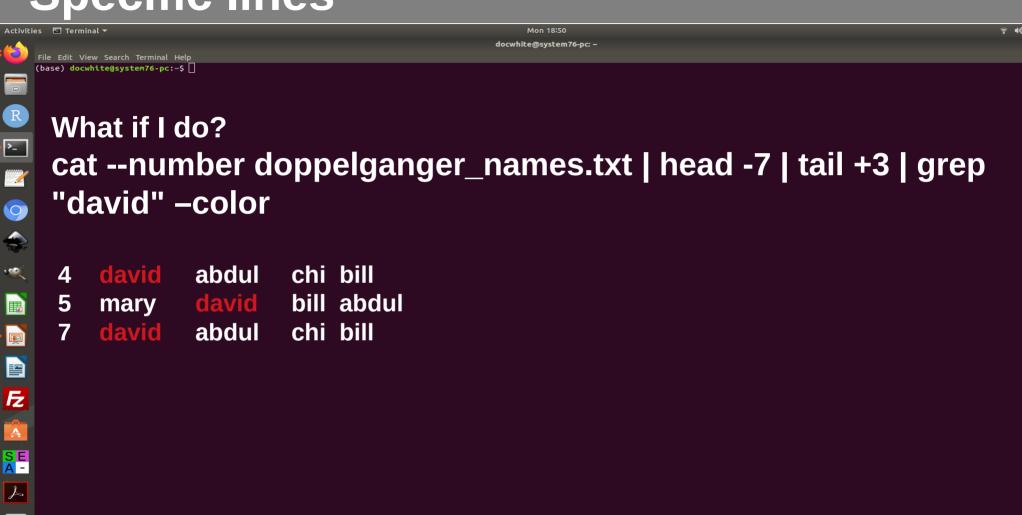






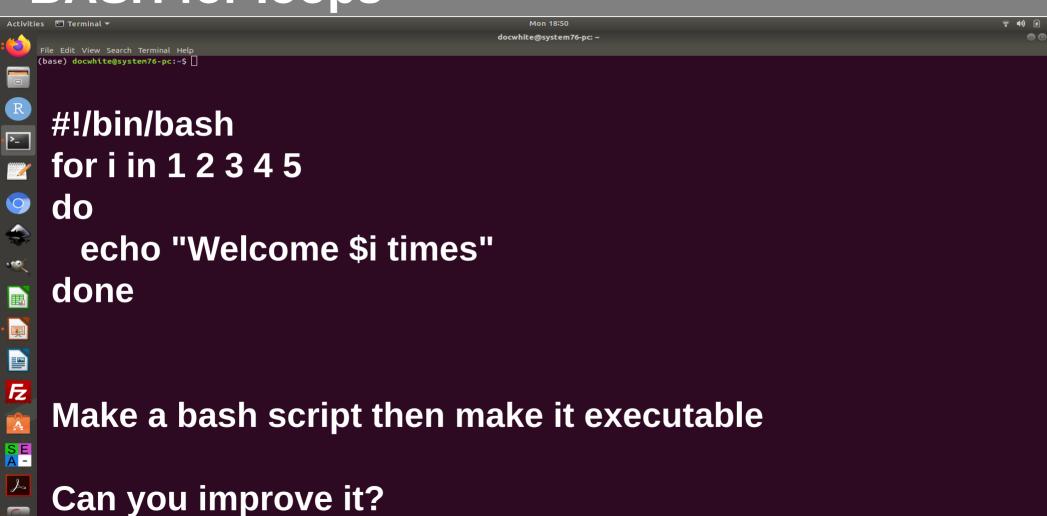




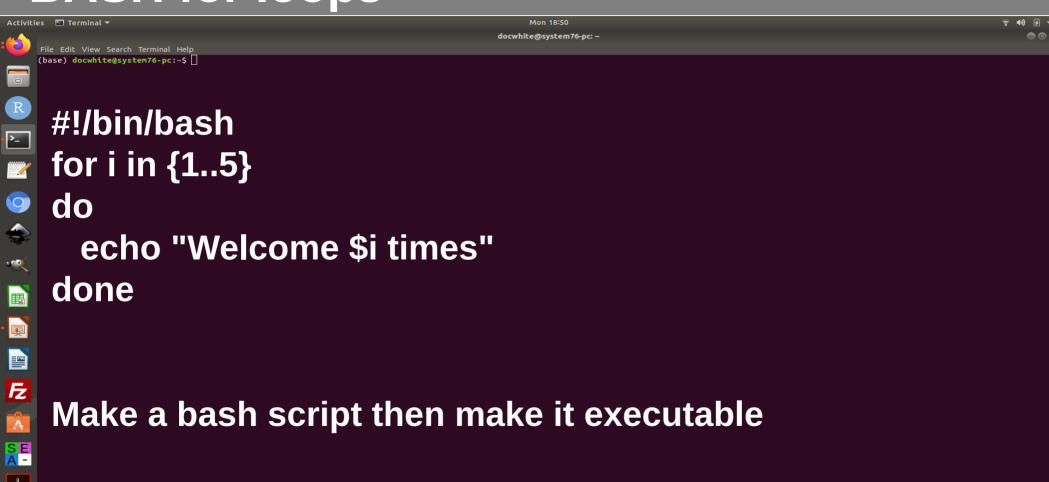


# for i in file.\*;do command \$i done

# **BASH** for loops



## **BASH for loops**

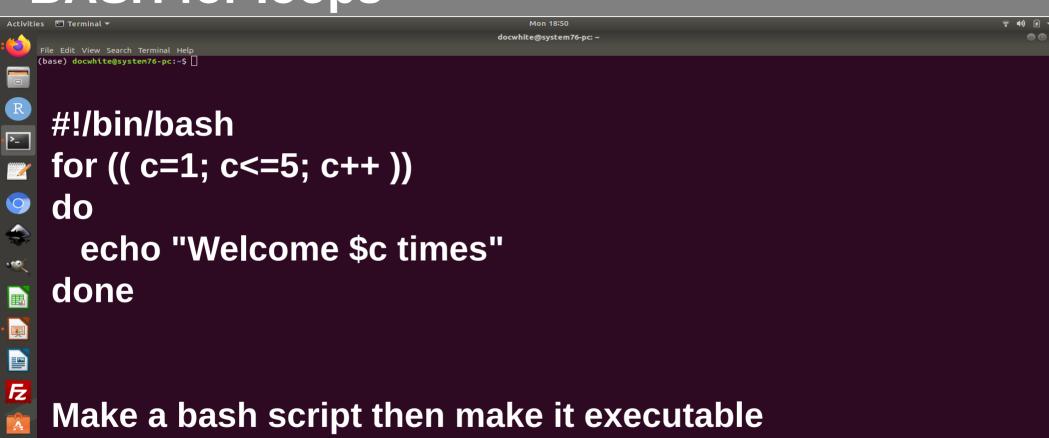


# BASH - for loop (C-style)

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

## **BASH for loops**

SE A-



# Quiz 10

- On canvas now

- 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?