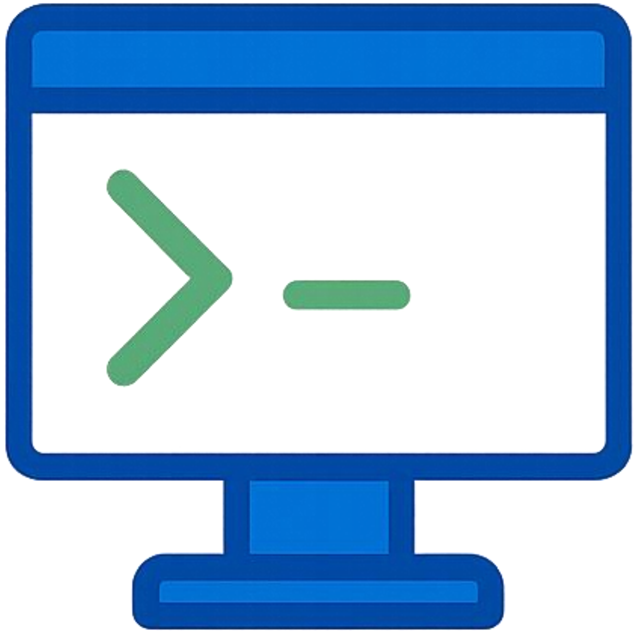
Introduction to BASH

A tour of command line navigation and 100 BASH commands

Kristine Lacek, MS
Bioinformatics Scientist
US CDC – Influenza Division



Disclaimer

Module Key

Lecture Content

Practical Content

Bonus intermediate content

Module Objectives

- Define fundamental terminal definitions, present **100 essential bash command encyclopedia**

Module Objectives

- Define fundamental terminal definitions, present **100 essential bash command encyclopedia**
- Trainees will be able to confidently move between directories, list files, and manage folder structures using core navigation commands

Module Objectives

- Define fundamental terminal definitions, present **100 essential bash command encyclopedia**
- Trainees will be able to confidently move between directories, list files, and manage folder structures using core navigation commands
- Trainees will develop proficiency in creating, copying, moving, renaming, deleting, and viewing files directly from the terminal

Module Objectives

- Define fundamental terminal definitions, present **100 essential bash command encyclopedia**
- Trainees will be able to confidently move between directories, list files, and manage folder structures using core navigation commands
- Trainees will develop proficiency in creating, copying, moving, renaming, deleting, and viewing files directly from the terminal
- Trainees will gain the ability to search, filter, and transform text files (samplesheets, fastas, fastqs) using key utilities such as grep, sed, and wc, enabling basic data handling and log analysis from the command line

Module Objectives

- Define fundamental terminal definitions, present **100 essential bash command encyclopedia**
- Trainees will be able to confidently move between directories, list files, and manage folder structures using core navigation commands
- Trainees will develop proficiency in creating, copying, moving, renaming, deleting, and viewing files directly from the terminal
- Trainees will gain the ability to search, filter, and transform text files (samplesheets, fastas, fastqs) using key utilities such as grep, sed, and wc, enabling basic data handling and log analysis from the command line
- Through practice with commands like sudo, chmod, ssh, curl, and wget, trainees will understand how to perform administrative tasks, set file permissions, and interact with remote systems or network resources.

Module Objectives

- Define fundamental terminal definitions, present **100 essential bash command encyclopedia**

- Enable you to confidently:
  - move between directories, list files, and manage folder structures
  - create, copy, move, rename, delete, and view files
  - search, filter, and transform text files
  - perform administrative tasks, set file permissions, and interact with remote systems

Practical component

- Module will alternate between lecture and practice
- Practical materials available at https://cdcgov.github.io/id-bioifx-workshop/

Command Line Navigation: Review

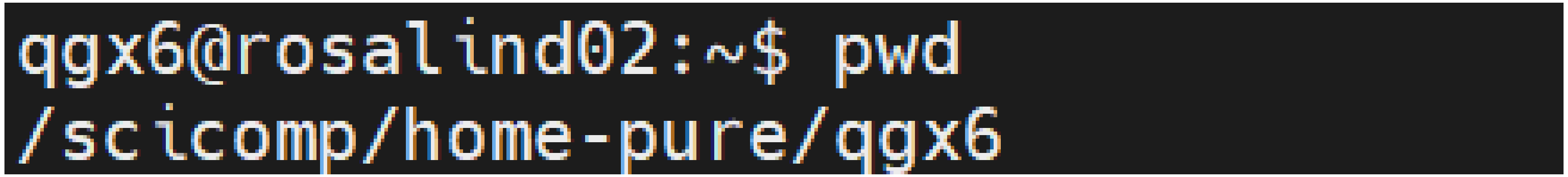
- Command prompt: text displayed by the shell indicating it is ready to accept a command

```
qgx6@rosalind02:~$ 
```

- Directory: folder

- Path: exact location of a file or directory in the filesystem

`pwd` : Print Working Directory = show where in the filesystem you are



```
qgx6@rosalind02:~$ pwd
/scicomp/home-pure/qgx6
```

Command Line Navigation: Review

- Flag (or option): A modifier added to a command (often starting with - or --) that changes how the command behaves

- *Defaults*: settings for the program if user provides no modifier

- *Argument*: Additional information provided to a command, such as a filename, directory name, or value to operate on.
  - Positional and/or *flag*

- *Flag/option*: A modifier added to a command (often starting with - or --) that changes how the command behaves
  - May or may not require their own *argument*

Command Line Navigation: Review

- **Positional arguments: Arguments defined by order**

```
cp file.txt backup.txt
```

  - file.txt is the 1st positional argument (source)
  - backup.txt is the 2nd positional argument (destination)
  - Switching the order changes the behavior

Command Line Navigation: Review

- **Positional arguments: Arguments defined by order**

```
cp file.txt backup.txt
```

  - file.txt is the 1st positional argument (source)
  - backup.txt is the 2nd positional argument (destination)
  - Switching the order changes the behavior

- **Binary (Boolean) flags: flag is either ON (present) or OFF (absent)**

```
ls -l
```

  - -l enables long listing mode, no value required

Command Line Navigation: Review

- **Positional arguments: Arguments defined by order**

```
cp file.txt backup.txt
```

  - file.txt is the 1st positional argument (source)
  - backup.txt is the 2nd positional argument (destination)
  - Switching the order changes the behavior

- **Binary (Boolean) flags: flag is either ON (present) or OFF (absent)**

```
ls -l
```

  - -l enables long listing mode, no value required

- **Argument (Value) flags: requires a value after it**

```
head –n 5 file.txt
```

- -n : flag
- 5: value

Command Line Navigation: Review

Type	Example	Takes Value?	Order Important?
Positional	cp a b	No (position defines meaning)	Yes
Binary Flag	ls -l	No	No
Argument Flag	-n 5	Yes	No

Manuals and help

- Command: Program
  - Difficult for anyone to remember exactly how to use every program on a system
  - There are programs that help you use programs!

  - man
  - help

Practical: Manual, flags, arguments

View the full manual for each of these commands by running “man <command>”
Get quick help with the flags for each command by running “<command> --help”

- ls
- sort
- head
- vim
- date
- time

1. Which of these is a text editor?
2. Which of these does *not* have a manual entry?

Command Line Navigation: moving in the filesystem

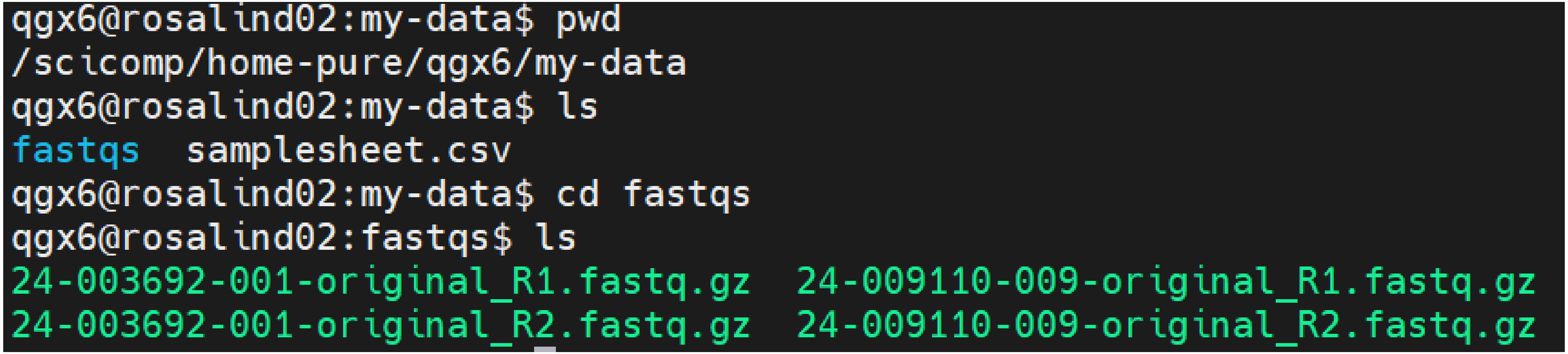
`ls` **: List directory contents**

`cd` **: Change directory**

`mkdir` **: Create directory**

`rmdir` **: Remove an empty directory**

`tree` : Show directory structure



```
qgx6@rosalind02:my-data$ pwd
/scicomp/home-pure/qgx6/my-data
qgx6@rosalind02:my-data$ ls
fastqs  samplesheet.csv
qgx6@rosalind02:my-data$ cd fastqs
qgx6@rosalind02:fastqs$ ls
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz
```

Command Line Navigation: moving in the filesystem

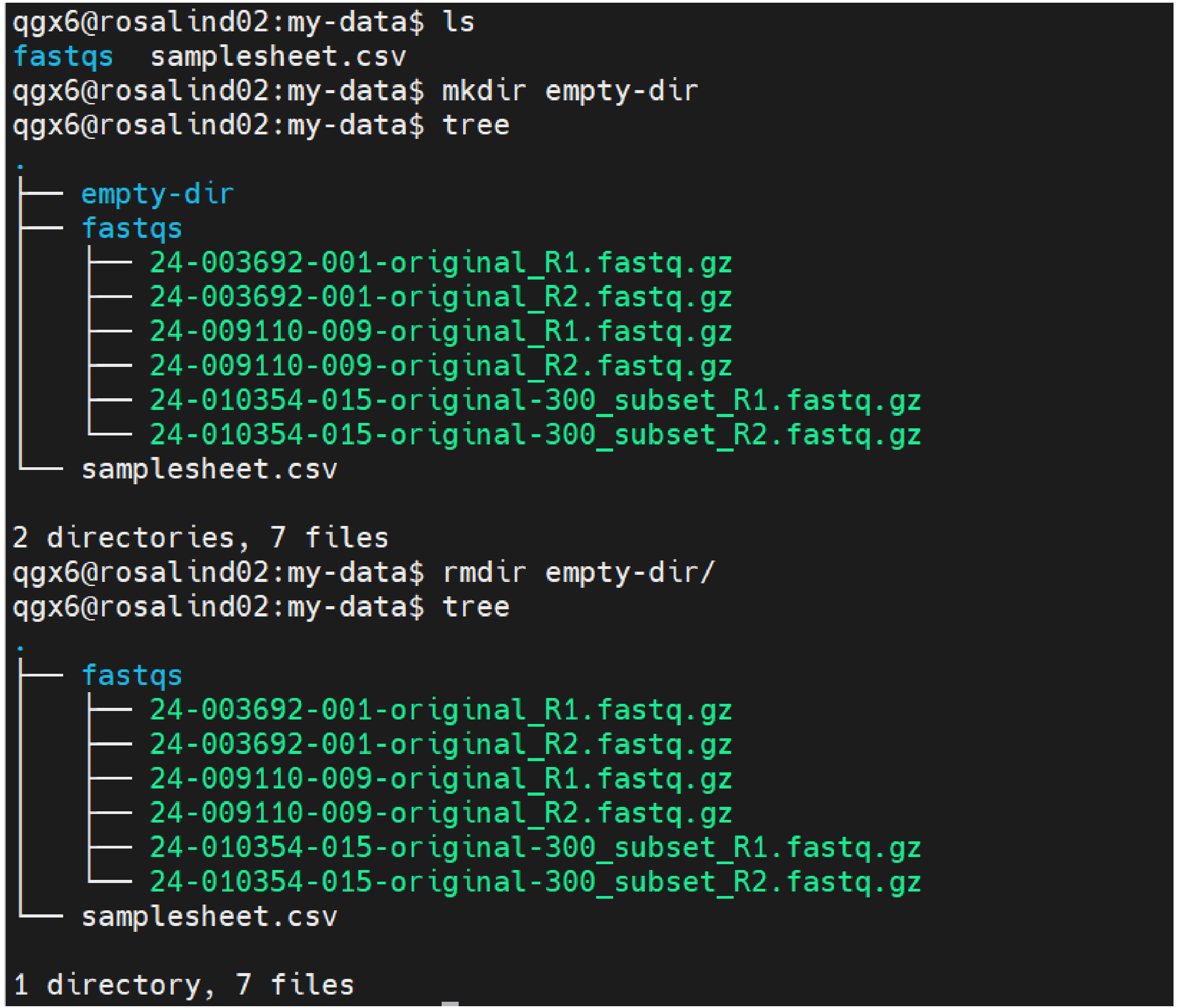
`ls` : List directory contents

`cd` : Change directory

`mkdir` : Create directory

`rmdir` : Remove an empty directory

`tree` : Show directory structure

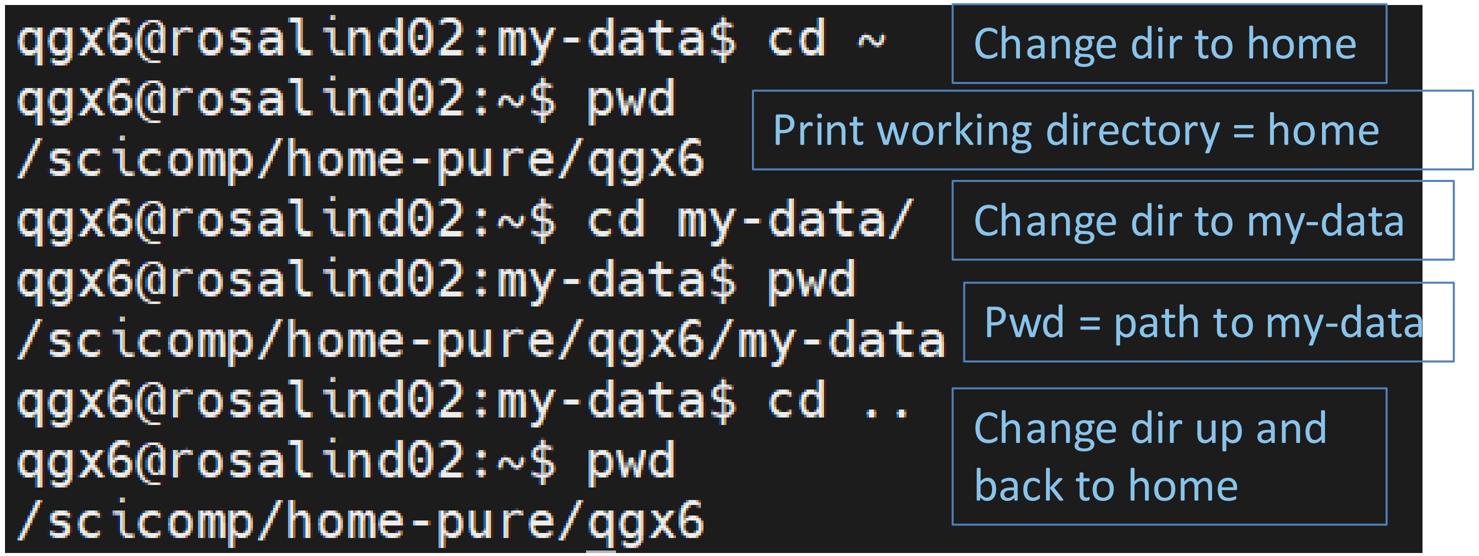


```
qgx6@rosalind02:my-data$ ls
fastqs  samplesheet.csv
qgx6@rosalind02:my-data$ mkdir empty-dir
qgx6@rosalind02:my-data$ tree
.
├── empty-dir
├── fastqs
│   ├── 24-003692-001-original_R1.fastq.gz
│   ├── 24-003692-001-original_R2.fastq.gz
│   ├── 24-009110-009-original_R1.fastq.gz
│   ├── 24-009110-009-original_R2.fastq.gz
│   ├── 24-010354-015-original-300_subset_R1.fastq.gz
│   └── 24-010354-015-original-300_subset_R2.fastq.gz
└── samplesheet.csv

2 directories, 7 files
qgx6@rosalind02:my-data$ rmdir empty-dir/
qgx6@rosalind02:my-data$ tree
.
├── fastqs
│   ├── 24-003692-001-original_R1.fastq.gz
│   ├── 24-003692-001-original_R2.fastq.gz
│   ├── 24-009110-009-original_R1.fastq.gz
│   ├── 24-009110-009-original_R2.fastq.gz
│   ├── 24-010354-015-original-300_subset_R1.fastq.gz
│   └── 24-010354-015-original-300_subset_R2.fastq.gz
└── samplesheet.csv

1 directory, 7 files
```

Command Line Navigation: moving in the filesystem

SPECIAL SYMBOLS

- `~` = home
- `.` = current location
- `..` = "back" or "up" a directory



```
qgx6@rosalind02:my-data$ cd ~
qgx6@rosalind02:~$ pwd
/scicomp/home-pure/qgx6
qgx6@rosalind02:~$ cd my-data/
qgx6@rosalind02:my-data$ pwd
/scicomp/home-pure/qgx6/my-data
qgx6@rosalind02:my-data$ cd ..
qgx6@rosalind02:~$ pwd
/scicomp/home-pure/qgx6
```

Change dir to home

Print working directory = home

Change dir to my-data

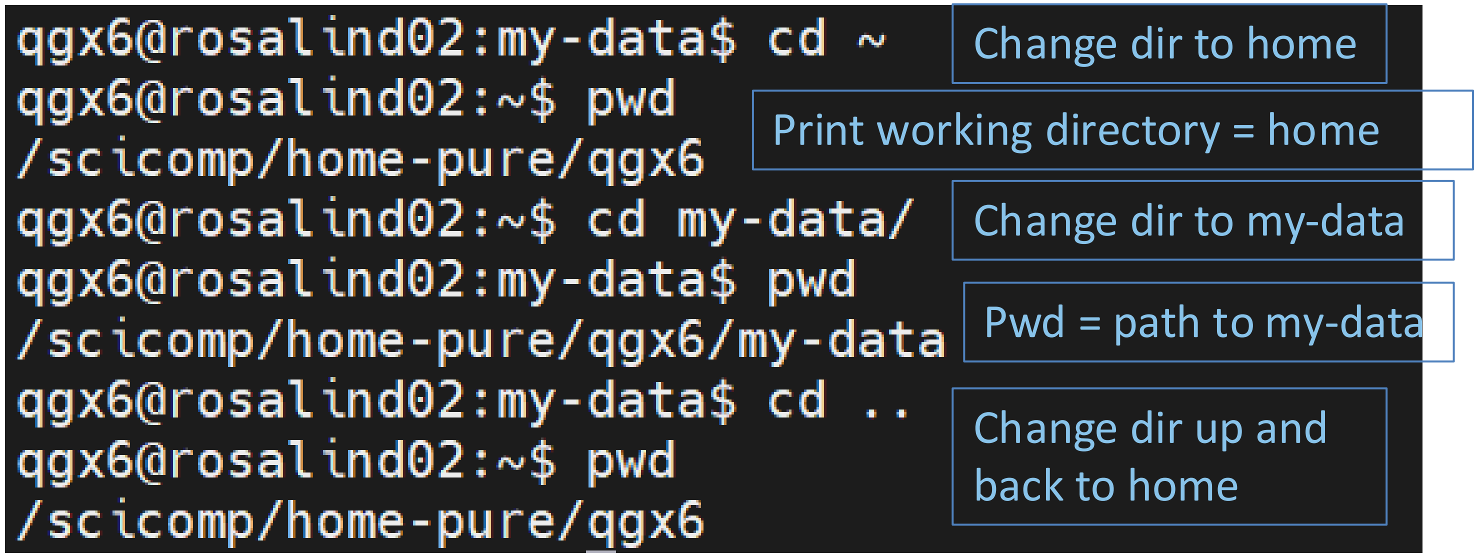
Pwd = path to my-data

Change dir up and back to home

Command Line Navigation: moving in the filesystem

SPECIAL SYMBOLS

- `~` = home
- `.` = current location
- `..` = “back” or “up” a directory



```
qgx6@rosalind02:my-data$ cd ~
qgx6@rosalind02:~$ pwd
/scicomp/home-pure/qgx6
qgx6@rosalind02:~$ cd my-data/
qgx6@rosalind02:my-data$ pwd
/scicomp/home-pure/qgx6/my-data
qgx6@rosalind02:my-data$ cd ..
qgx6@rosalind02:~$ pwd
/scicomp/home-pure/qgx6
```

Change dir to home

Print working directory = home

Change dir to my-data

Pwd = path to my-data

Change dir up and back to home

BASH Tips!

- Shell will try to “guess” what you’re typing next if it’s an existing filename or directory. Hit tab to complete the full name for faster navigation
- Shell remembers everything you run in your “history”, use up arrow to go back in history
- Use “cd –” to go to the last directory you were in; not necessarily relative to your cwd!

Command Line Navigation: moving in the filesystem

Relative paths vs absolute paths

- Absolute Path
    - Full path from the root directory (/)
    - Always starts with /
    - Works regardless of current location

  Example: /home/user/project/data/sample.fastq

- Relative Path
    - Path relative to your current directory
    - Does NOT start with /
    - Depends on where you are (pwd)

  Example: data/sample.fastq

Command Line Navigation Practical

1. Open your terminal
2. Change directory to your home directory (~)
3. List the contents of that directory
4. Make a directory within your home directory called 2026-BIFX-TRAINING
5. Cd to that new directory you just made
6. Using .. , cd back "up" to your home directory
7. Try tab-complete

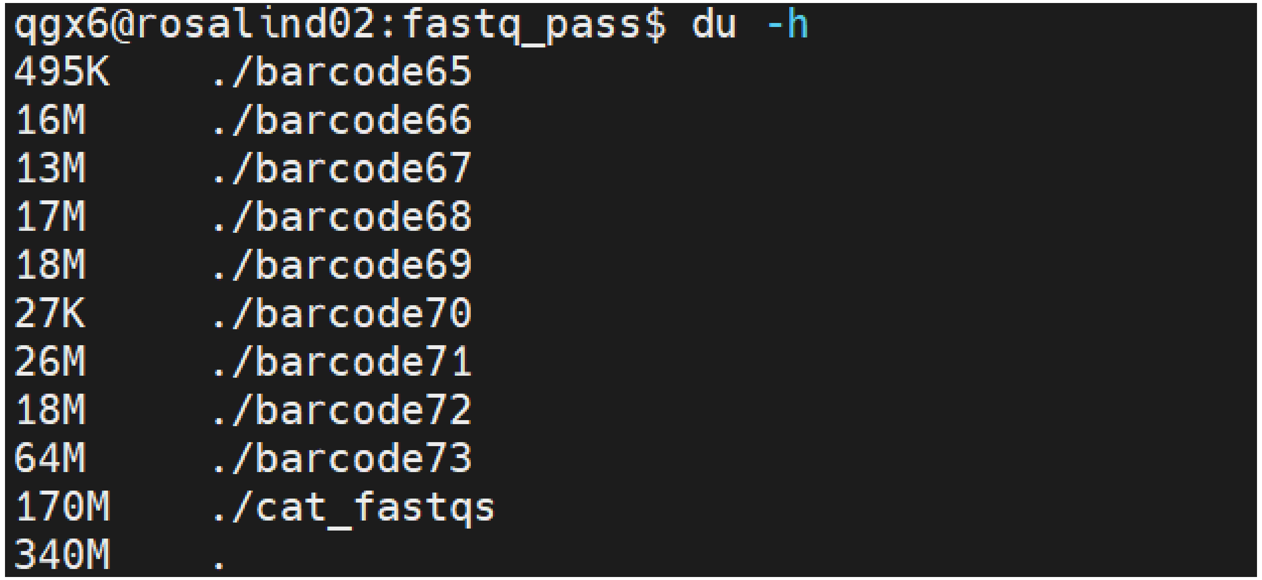
   Start the command "cd 2026-" then tab to complete the full directory name

8. Try using your history

   Using only the up arrow to move through your recent commands, re-run the command you ran for #2

Intermediate Directory Management commands

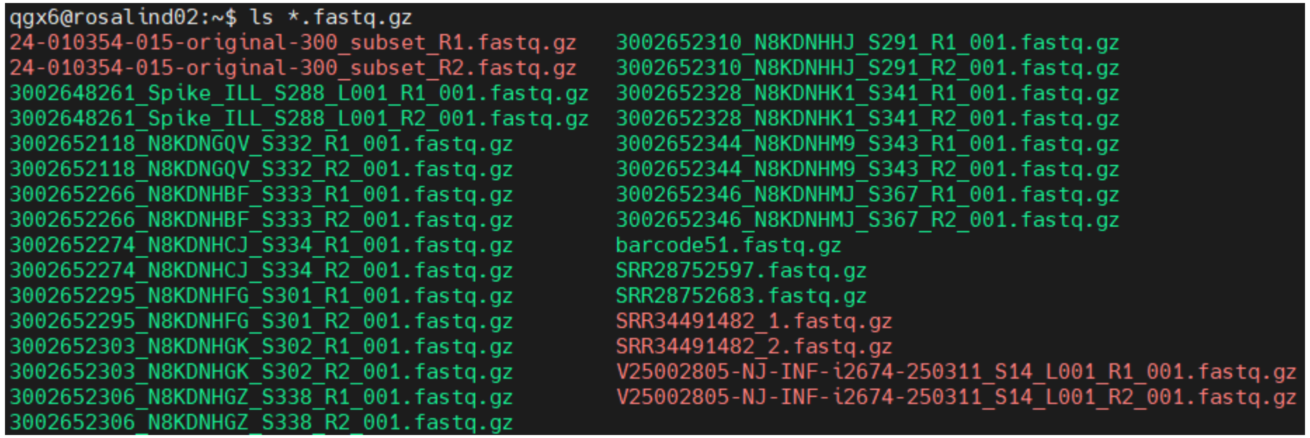
- `du` : Disk usage summary (-h flag suggested)
- `df` : Display disk space usage for all file-systems mounted by OS



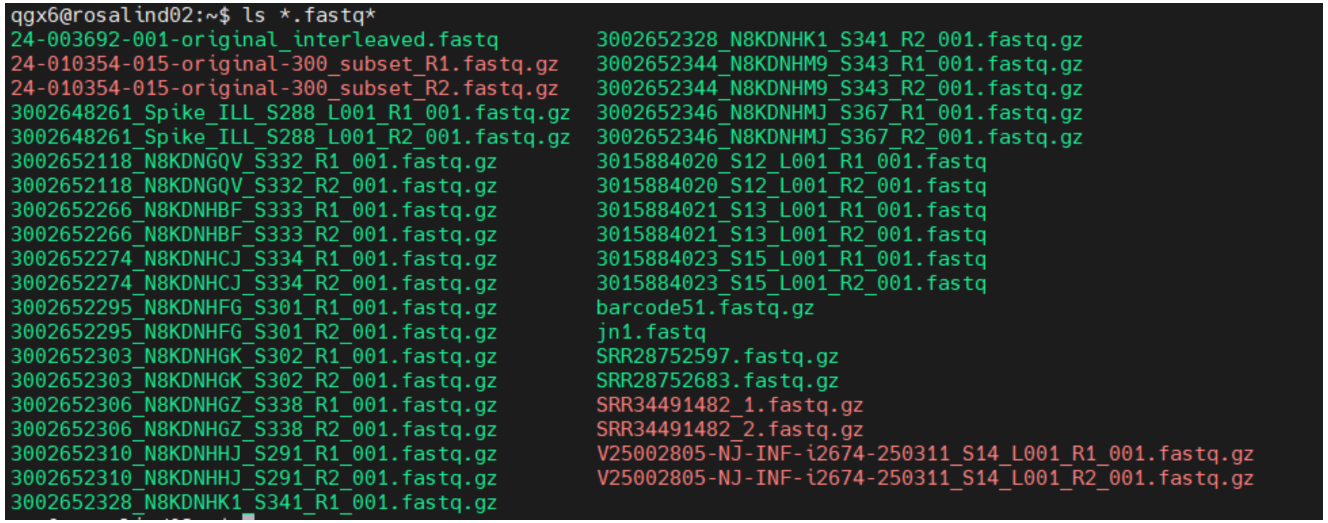
```
qgx6@rosalind02:fastq_pass$ du -h
495K    ./barcode65
16M     ./barcode66
13M     ./barcode67
17M     ./barcode68
18M     ./barcode69
27K     ./barcode70
26M     ./barcode71
18M     ./barcode72
64M     ./barcode73
170M    ./cat_fastqs
340M    .
```

Command Line Special Symbols & Shorthand

- Builtins include special symbols and words are **reserved**
    - Type `help` to see list
- Directories and pathing:
    - `~` = home
    - `.` = current location
    - `..` = “back” or “up” a directory
    - **`*` = glob, or any letter, words**
        - **Can represent nothing as well**



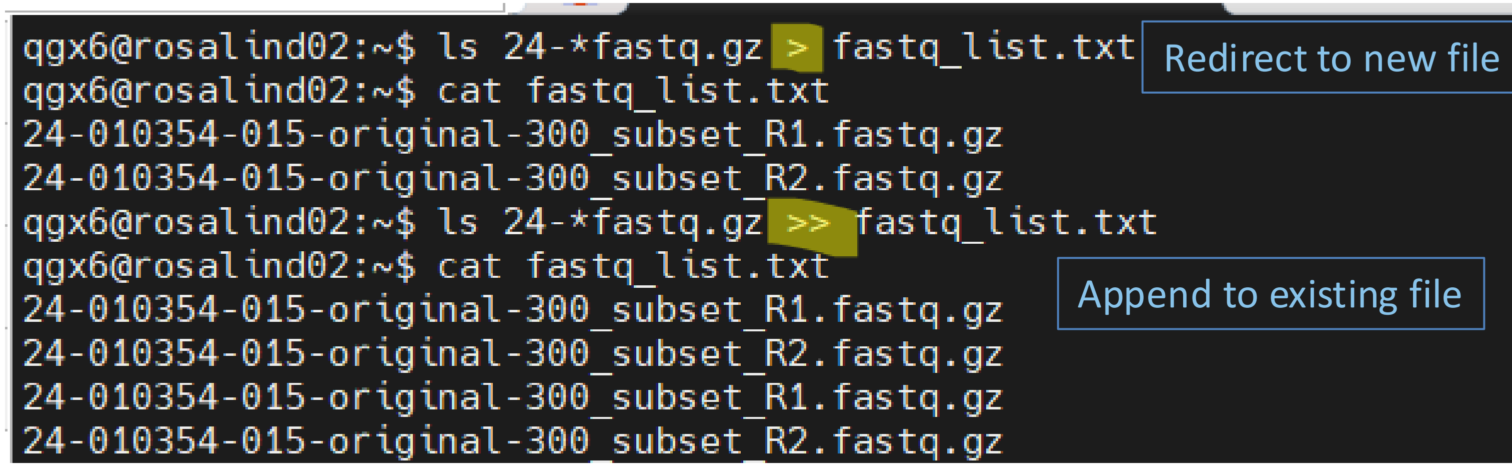
```
qgx6@rosalind02:~$ ls *.fastq.gz
24-010354-015-original-300_subset_R1.fastq.gz    3002652310_N8KDNHHJ_S291_R1_001.fastq.gz
24-010354-015-original-300_subset_R2.fastq.gz    3002652310_N8KDNHHJ_S291_R2_001.fastq.gz
3002648261_Spike_ILL_S288_L001_R1_001.fastq.gz  3002652328_N8KDNHK1_S341_R1_001.fastq.gz
3002648261_Spike_ILL_S288_L001_R2_001.fastq.gz  3002652328_N8KDNHK1_S341_R2_001.fastq.gz
3002652118_N8KDNGQV_S332_R1_001.fastq.gz        3002652344_N8KDNHM9_S343_R1_001.fastq.gz
3002652118_N8KDNGQV_S332_R2_001.fastq.gz        3002652344_N8KDNHM9_S343_R2_001.fastq.gz
3002652266_N8KDNHBF_S333_R1_001.fastq.gz        3002652346_N8KDNHMJ_S367_R1_001.fastq.gz
3002652266_N8KDNHBF_S333_R2_001.fastq.gz        3002652346_N8KDNHMJ_S367_R2_001.fastq.gz
3002652274_N8KDNHCJ_S334_R1_001.fastq.gz        barcode51.fastq.gz
3002652274_N8KDNHCJ_S334_R2_001.fastq.gz        SRR28752597.fastq.gz
3002652295_N8KDNHFG_S301_R1_001.fastq.gz        SRR28752683.fastq.gz
3002652295_N8KDNHFG_S301_R2_001.fastq.gz        SRR34491482_1.fastq.gz
3002652303_N8KDNHGK_S302_R1_001.fastq.gz        SRR34491482_2.fastq.gz
3002652303_N8KDNHGK_S302_R2_001.fastq.gz        V25002805-NJ-INF-i2674-250311_S14_L001_R1_001.fastq.gz
3002652306_N8KDNHGZ_S338_R1_001.fastq.gz        V25002805-NJ-INF-i2674-250311_S14_L001_R2_001.fastq.gz
3002652306_N8KDNHGZ_S338_R2_001.fastq.gz
```



```
qgx6@rosalind02:~$ ls *.fastq*
24-003692-001-original_interleaved.fastq        3002652328_N8KDNHK1_S341_R2_001.fastq.gz
24-010354-015-original-300_subset_R1.fastq.gz   3002652344_N8KDNHM9_S343_R1_001.fastq.gz
24-010354-015-original-300_subset_R2.fastq.gz   3002652344_N8KDNHM9_S343_R2_001.fastq.gz
3002648261_Spike_ILL_S288_L001_R1_001.fastq.gz  3002652346_N8KDNHMJ_S367_R1_001.fastq.gz
3002648261_Spike_ILL_S288_L001_R2_001.fastq.gz  3002652346_N8KDNHMJ_S367_R2_001.fastq.gz
3002652118_N8KDNGQV_S332_R1_001.fastq.gz        3015884020_S12_L001_R1_001.fastq
3002652118_N8KDNGQV_S332_R2_001.fastq.gz        3015884020_S12_L001_R2_001.fastq
3002652266_N8KDNHBF_S333_R1_001.fastq.gz        3015884021_S13_L001_R1_001.fastq
3002652266_N8KDNHBF_S333_R2_001.fastq.gz        3015884021_S13_L001_R2_001.fastq
3002652274_N8KDNHCJ_S334_R1_001.fastq.gz        3015884023_S15_L001_R1_001.fastq
3002652274_N8KDNHCJ_S334_R2_001.fastq.gz        3015884023_S15_L001_R2_001.fastq
3002652295_N8KDNHFG_S301_R1_001.fastq.gz        barcode51.fastq.gz
3002652295_N8KDNHFG_S301_R2_001.fastq.gz        jn1.fastq
3002652303_N8KDNHGK_S302_R1_001.fastq.gz        SRR28752597.fastq.gz
3002652303_N8KDNHGK_S302_R2_001.fastq.gz        SRR28752683.fastq.gz
3002652306_N8KDNHGZ_S338_R1_001.fastq.gz        SRR34491482_1.fastq.gz
3002652306_N8KDNHGZ_S338_R2_001.fastq.gz        SRR34491482_2.fastq.gz
3002652310_N8KDNHHJ_S291_R1_001.fastq.gz        V25002805-NJ-INF-i2674-250311_S14_L001_R1_001.fastq.gz
3002652310_N8KDNHHJ_S291_R2_001.fastq.gz        V25002805-NJ-INF-i2674-250311_S14_L001_R2_001.fastq.gz
3002652328_N8KDNHK1_S341_R1_001.fastq.gz
```

Command Line Special Symbols & Shorthand

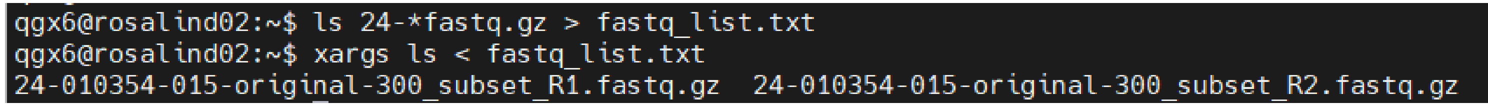
- Input and output *operators*:
  - `>` = send command output to a file (new or overwrite)
  - `>>` = append or add command output to a file
  - `<` = send multiple inputs to a command from a file
  - **`|` = Pipe, or send output from one command to input of another command**



```
qgx6@rosalind02:~$ ls 24-*fastq.gz > fastq_list.txt
qgx6@rosalind02:~$ cat fastq_list.txt
24-010354-015-original-300_subset_R1.fastq.gz
24-010354-015-original-300_subset_R2.fastq.gz
qgx6@rosalind02:~$ ls 24-*fastq.gz >> fastq_list.txt
qgx6@rosalind02:~$ cat fastq_list.txt
24-010354-015-original-300_subset_R1.fastq.gz
24-010354-015-original-300_subset_R2.fastq.gz
24-010354-015-original-300_subset_R1.fastq.gz
24-010354-015-original-300_subset_R2.fastq.gz
```

Redirect to new file

Append to existing file

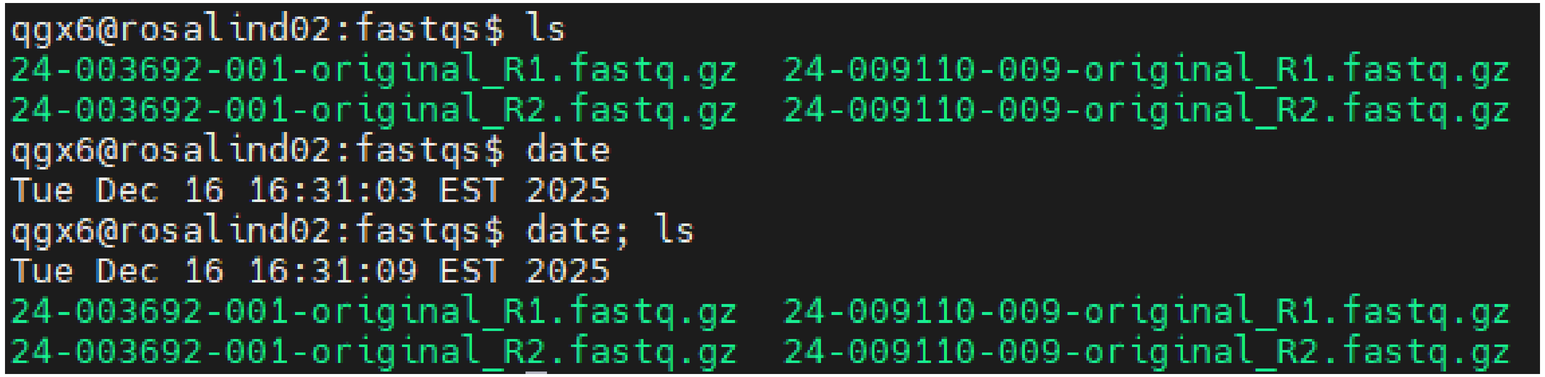


```
qgx6@rosalind02:~$ ls 24-*fastq.gz > fastq_list.txt
qgx6@rosalind02:~$ xargs ls < fastq_list.txt
24-010354-015-original-300_subset_R1.fastq.gz  24-010354-015-original-300_subset_R2.fastq.gz
```

We will work this into future examples

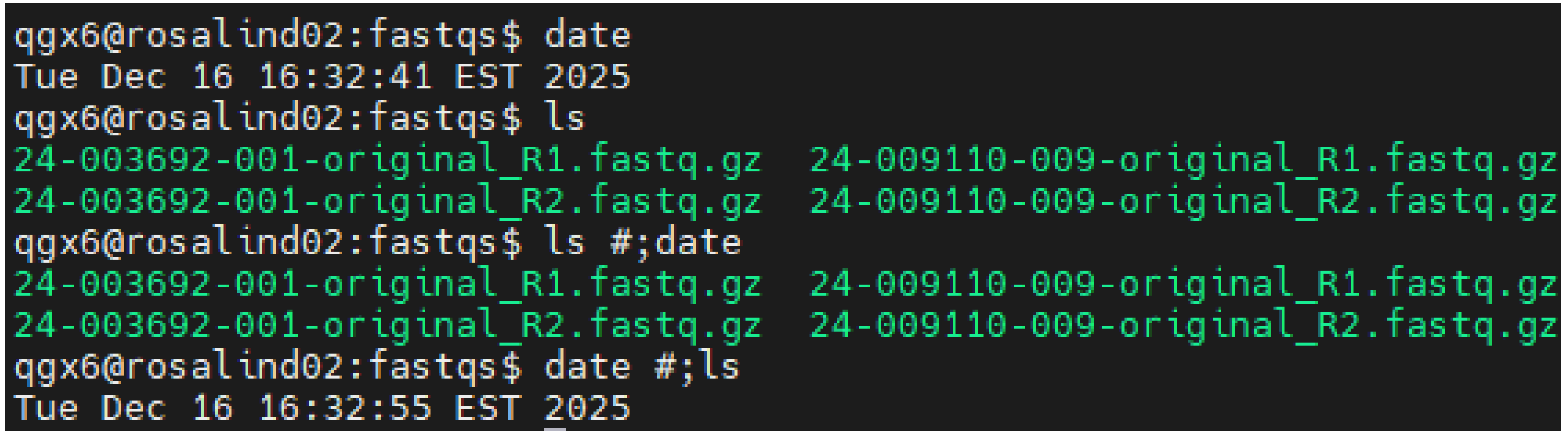
Command Line Special Symbols & Shorthand

- ; = separate multiple commands on a single line



```
qgx6@rosalind02:fastqs$ ls
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz
qgx6@rosalind02:fastqs$ date
Tue Dec 16 16:31:03 EST 2025
qgx6@rosalind02:fastqs$ date; ls
Tue Dec 16 16:31:09 EST 2025
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz
```

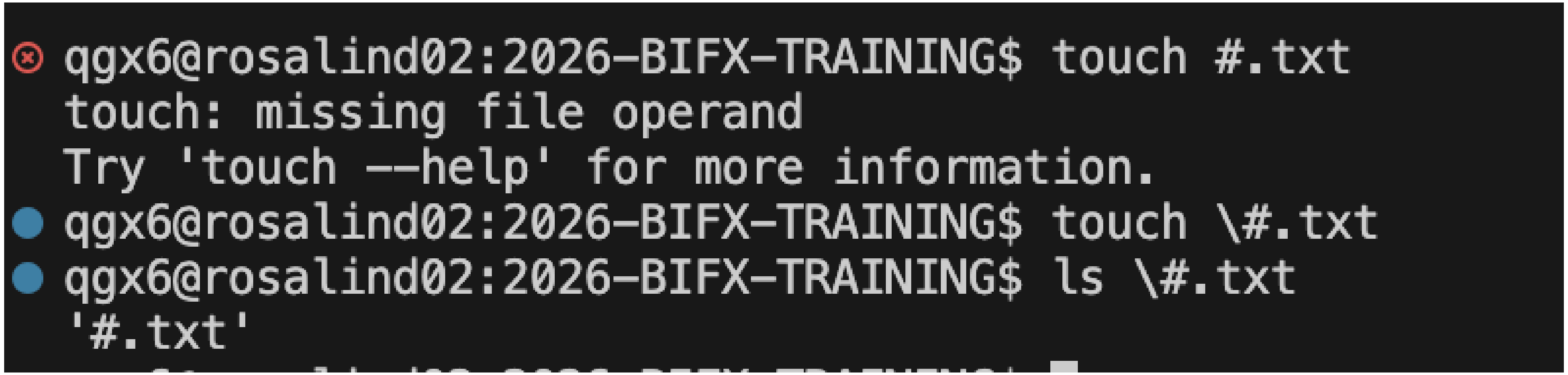
- # = comment, or ignore to the end of line



```
qgx6@rosalind02:fastqs$ date
Tue Dec 16 16:32:41 EST 2025
qgx6@rosalind02:fastqs$ ls
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz
qgx6@rosalind02:fastqs$ ls #;date
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz
qgx6@rosalind02:fastqs$ date #;ls
Tue Dec 16 16:32:55 EST 2025
```

Command Line Special Symbols & Shorthand

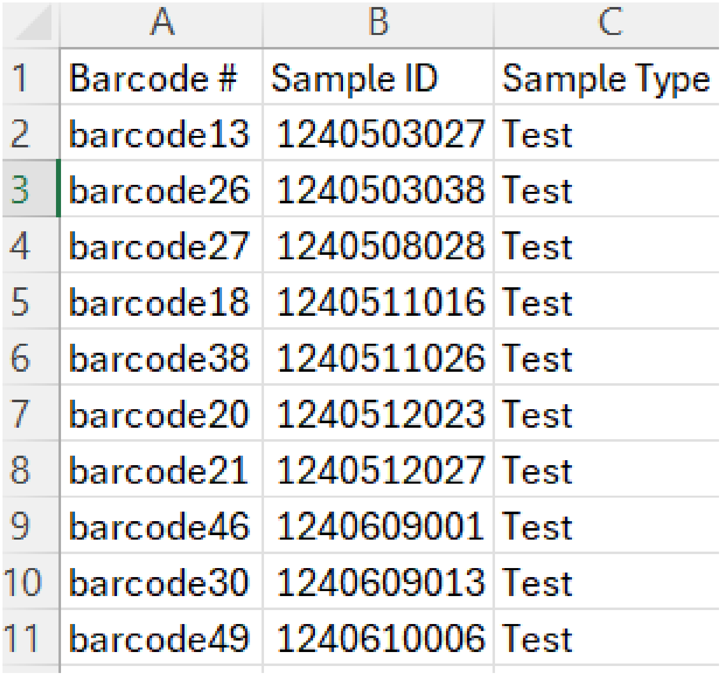
- Use \ to escape a character’s special behavior



```
⊗ qgx6@rosalind02:2026-BIFX-TRAINING$ touch #.txt
  touch: missing file operand
  Try 'touch --help' for more information.
● qgx6@rosalind02:2026-BIFX-TRAINING$ touch \#.txt
● qgx6@rosalind02:2026-BIFX-TRAINING$ ls \#.txt
  '#.txt'
```

Command Line Special Symbols & Shorthand

- \t = tab character in a textfile
- \n = newline character in a textfile



	A	B	C
1	Barcode #	Sample ID	Sample Type
2	barcode13	1240503027	Test
3	barcode26	1240503038	Test
4	barcode27	1240508028	Test
5	barcode18	1240511016	Test
6	barcode38	1240511026	Test
7	barcode20	1240512023	Test
8	barcode21	1240512027	Test
9	barcode46	1240609001	Test
10	barcode30	1240609013	Test
11	barcode49	1240610006	Test

Humans see this

```
Barcode #\tSample ID\tSample Type\n
barcode03\t1240503027\tTest\n
barcode26\t1240503038\tTest\n
barcode27\t1240508028\tTest\n
barcode18\t1240511016\tTest\n
barcode38\t1240511026\tTest\n
barcode20\t1240512023\tTest\n
barcode21\t1240512027\tTest\n
barcode46\t1240609001\tTest\n
barcode30\t1240609013\tTest\n
barcode49\t1240610006\tTest\n
barcode31\t1240610025\tTest\n
barcode74\t1240610040\tTest\n
barcode37\t1240612018\tTest\n
barcode44\t1240903032\tTest\n
barcode31\t1240904021\tTest\n
```

Shell "sees" this

File Viewing and manipulation

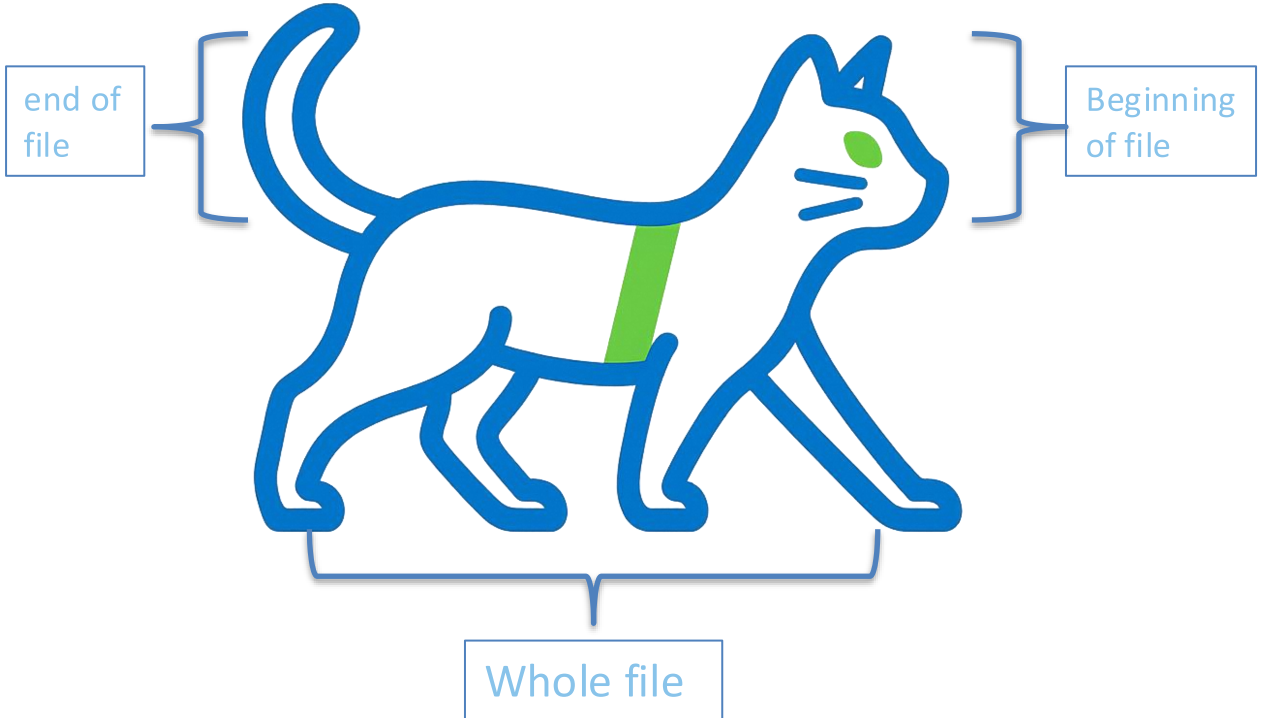
`cat` : Display file contents

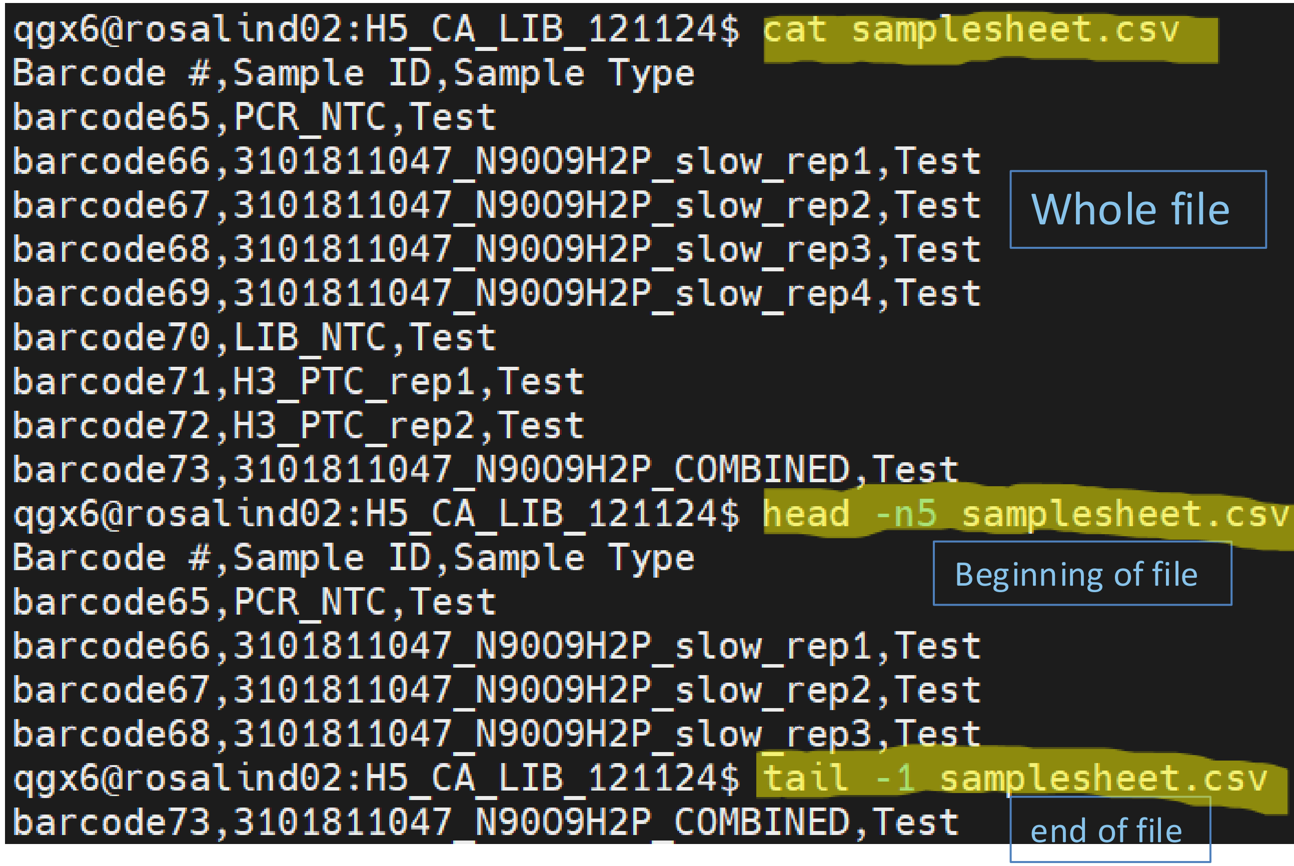
`head` : Show first 10 lines of file

- Change number of lines with –n flag or -#

`tail` : Show last lines 10 of file

- Change number of lines with –n flag -#





```
qgx6@rosalind02:H5_CA_LIB_121124$ cat samplesheet.csv
Barcode #,Sample ID,Sample Type
barcode65,PCR_NTC,Test
barcode66,3101811047_N9009H2P_slow_rep1,Test
barcode67,3101811047_N9009H2P_slow_rep2,Test
barcode68,3101811047_N9009H2P_slow_rep3,Test
barcode69,3101811047_N9009H2P_slow_rep4,Test
barcode70,LIB_NTC,Test
barcode71,H3_PTC_rep1,Test
barcode72,H3_PTC_rep2,Test
barcode73,3101811047_N9009H2P_COMBINED,Test
qgx6@rosalind02:H5_CA_LIB_121124$ head -n5 samplesheet.csv
Barcode #,Sample ID,Sample Type
barcode65,PCR_NTC,Test
barcode66,3101811047_N9009H2P_slow_rep1,Test
barcode67,3101811047_N9009H2P_slow_rep2,Test
barcode68,3101811047_N9009H2P_slow_rep3,Test
qgx6@rosalind02:H5_CA_LIB_121124$ tail -1 samplesheet.csv
barcode73,3101811047_N9009H2P_COMBINED,Test
```

File Viewing and manipulation

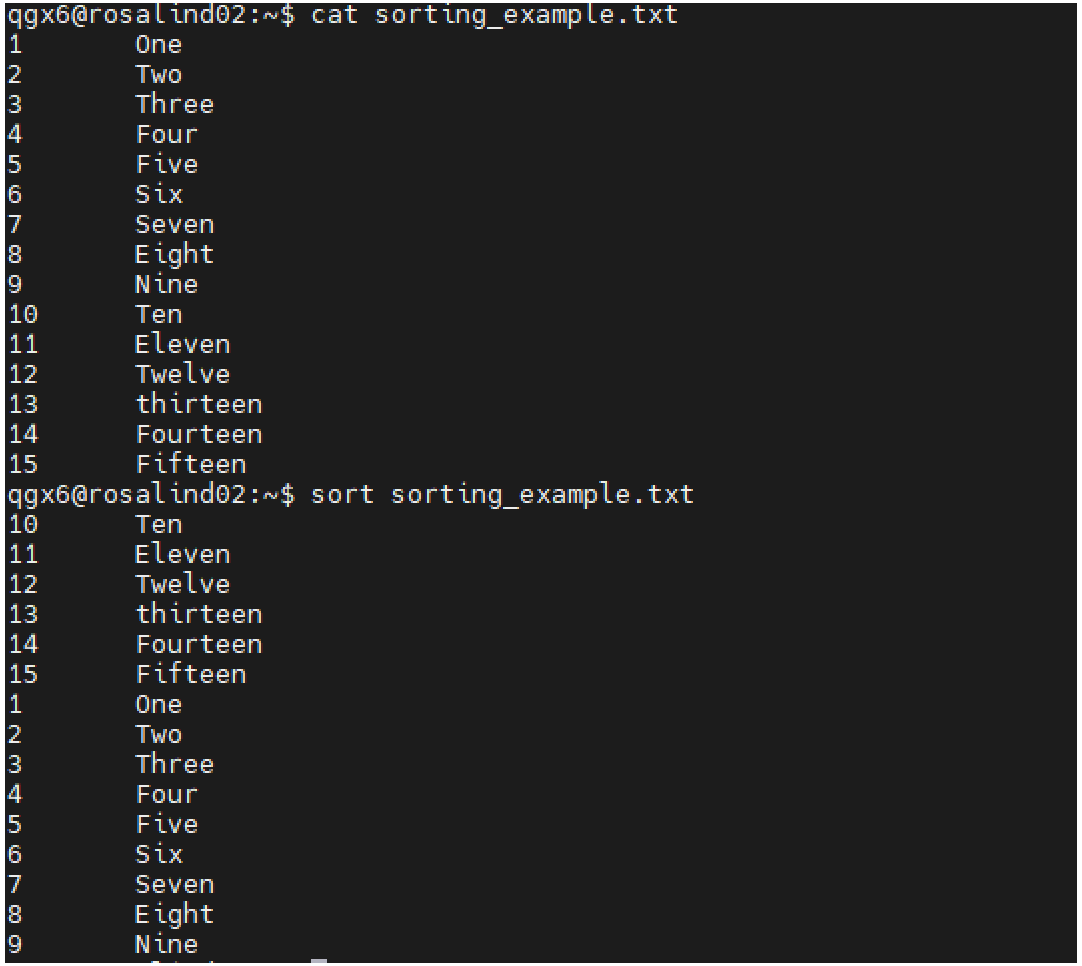
- `vim` **: Open text editor**

We will come back to this one

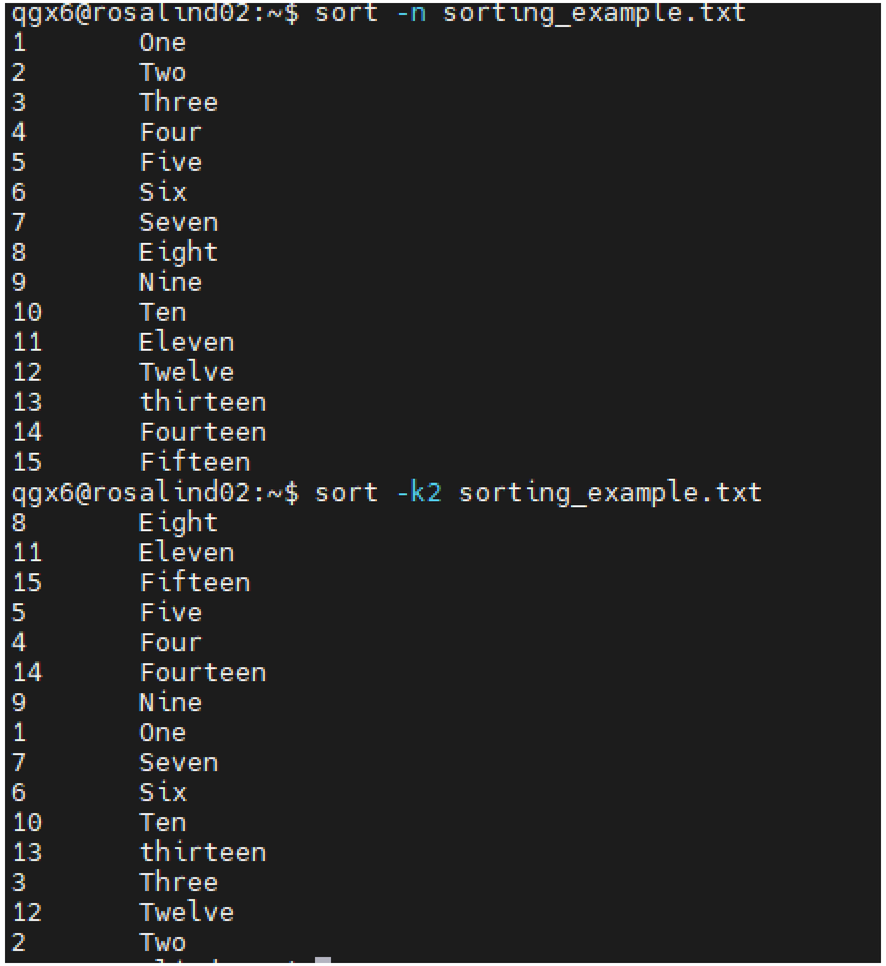
- `less` : View file content interactively
- `sort` : Sort lines in a file
  - Defaults to first column or field, sorting alphabetically
  - -n tells the command to sort numerically
  - -k 4 tells command to sort according to the fourth field
  - -d , tells command to divide fields according to ","

File Viewing and manipulation

- Sorting examples



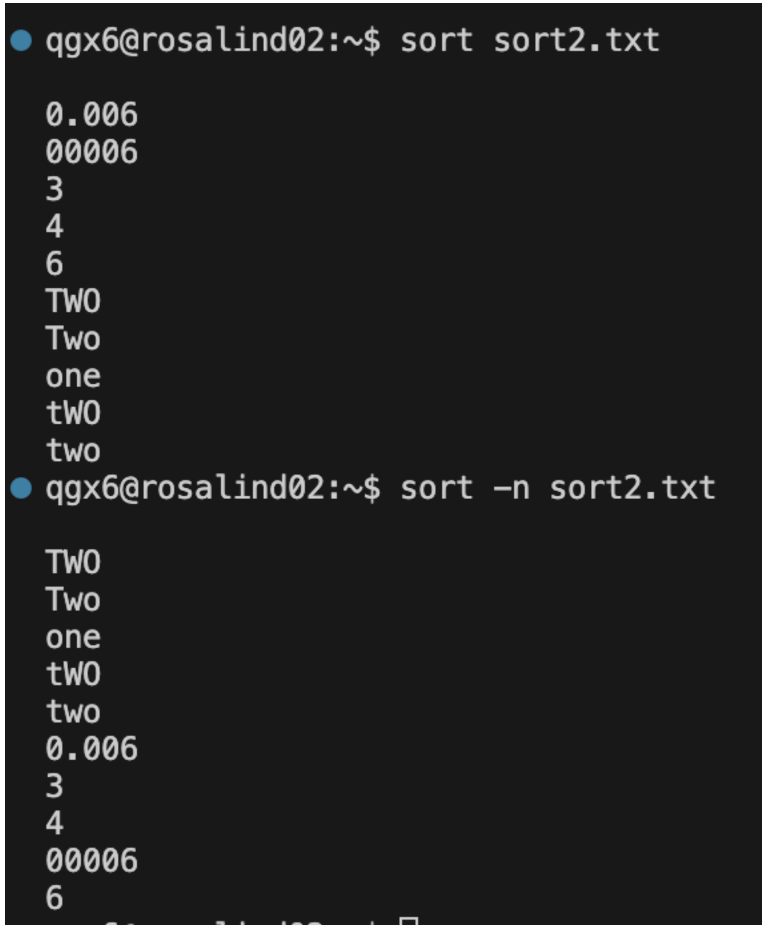
```
qgx6@rosalind02:~$ cat sorting_example.txt
1       One
2       Two
3       Three
4       Four
5       Five
6       Six
7       Seven
8       Eight
9       Nine
10      Ten
11      Eleven
12      Twelve
13      thirteen
14      Fourteen
15      Fifteen
qgx6@rosalind02:~$ sort sorting_example.txt
10      Ten
11      Eleven
12      Twelve
13      thirteen
14      Fourteen
15      Fifteen
1       One
2       Two
3       Three
4       Four
5       Five
6       Six
7       Seven
8       Eight
9       Nine
```



```
qgx6@rosalind02:~$ sort -n sorting_example.txt
1       One
2       Two
3       Three
4       Four
5       Five
6       Six
7       Seven
8       Eight
9       Nine
10      Ten
11      Eleven
12      Twelve
13      thirteen
14      Fourteen
15      Fifteen
qgx6@rosalind02:~$ sort -k2 sorting_example.txt
8       Eight
11      Eleven
15      Fifteen
5       Five
4       Four
14      Fourteen
9       Nine
1       One
7       Seven
6       Six
10      Ten
13      thirteen
3       Three
12      Twelve
2       Two
```

File Viewing and manipulation

- Sorting examples



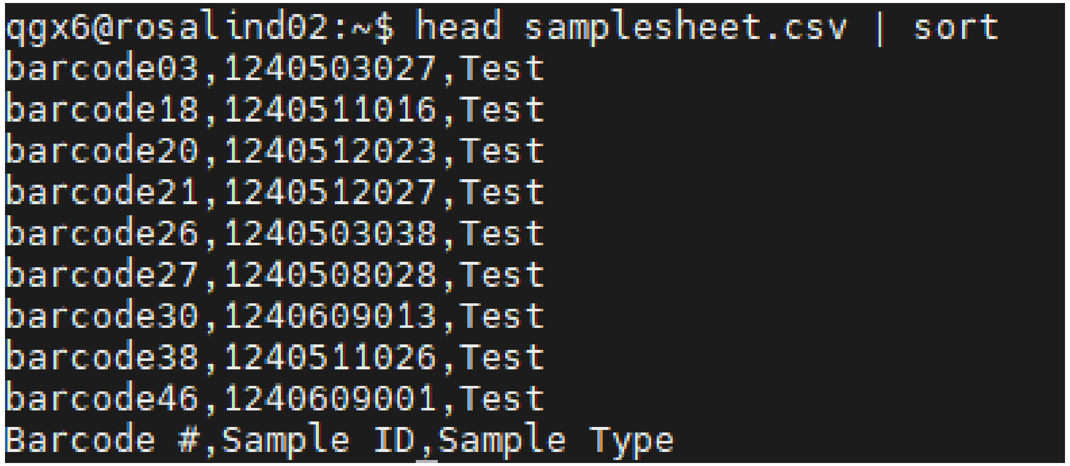
```
qgx6@rosalind02:~$ sort sort2.txt

0.006
00006
3
4
6
TWO
Two
one
tWO
two
qgx6@rosalind02:~$ sort -n sort2.txt

TWO
Two
one
tWO
two
0.006
3
4
00006
6
```

Piping Example

- **| = Pipe, or send output from one command to input of another command**
  - Caution: not all commands accept stdin

Sort the first 10 lines of a file

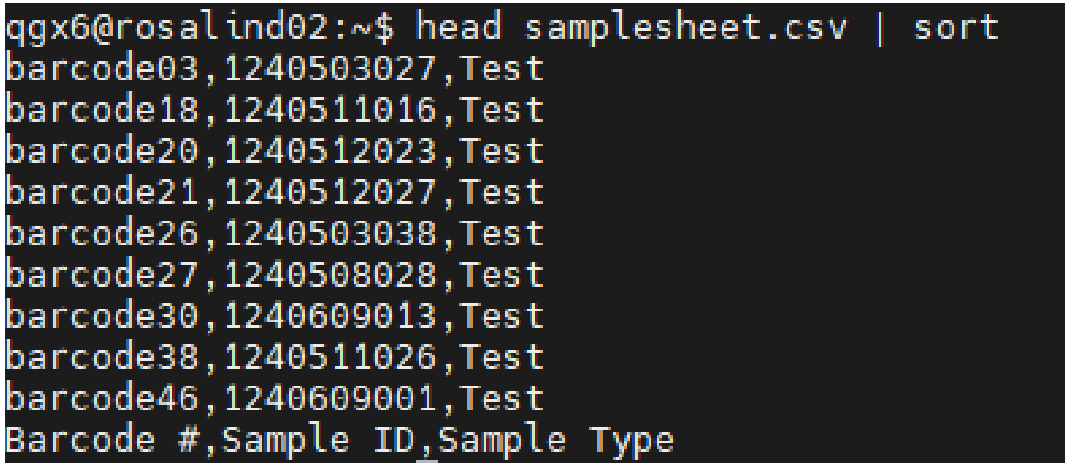
head samplesheet.csv | sort



```
qgx6@rosalind02:~$ head samplesheet.csv | sort
barcode03,1240503027,Test
barcode18,1240511016,Test
barcode20,1240512023,Test
barcode21,1240512027,Test
barcode26,1240503038,Test
barcode27,1240508028,Test
barcode30,1240609013,Test
barcode38,1240511026,Test
barcode46,1240609001,Test
Barcode #,Sample ID,Sample Type
```

Piping Example

- **| = Pipe, or send output from one command to input of another command**



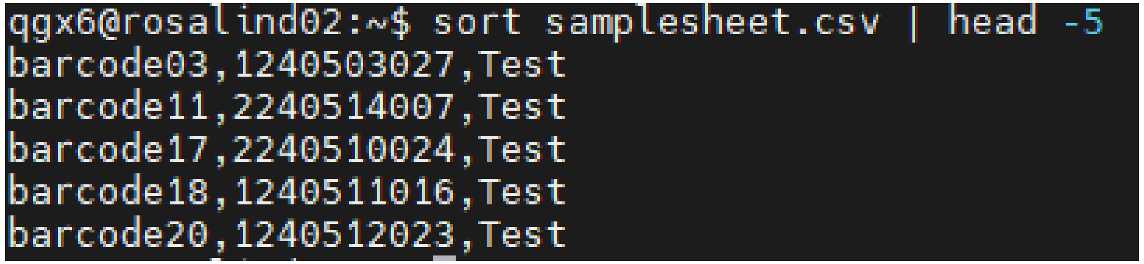
```
qgx6@rosalind02:~$ head samplesheet.csv | sort
barcode03,1240503027,Test
barcode18,1240511016,Test
barcode20,1240512023,Test
barcode21,1240512027,Test
barcode26,1240503038,Test
barcode27,1240508028,Test
barcode30,1240609013,Test
barcode38,1240511026,Test
barcode46,1240609001,Test
Barcode #,Sample ID,Sample Type
```

Sort the first 10 lines of a file

head samplesheet.csv | sort

Sort a file then output the first 5 entries alphabetically

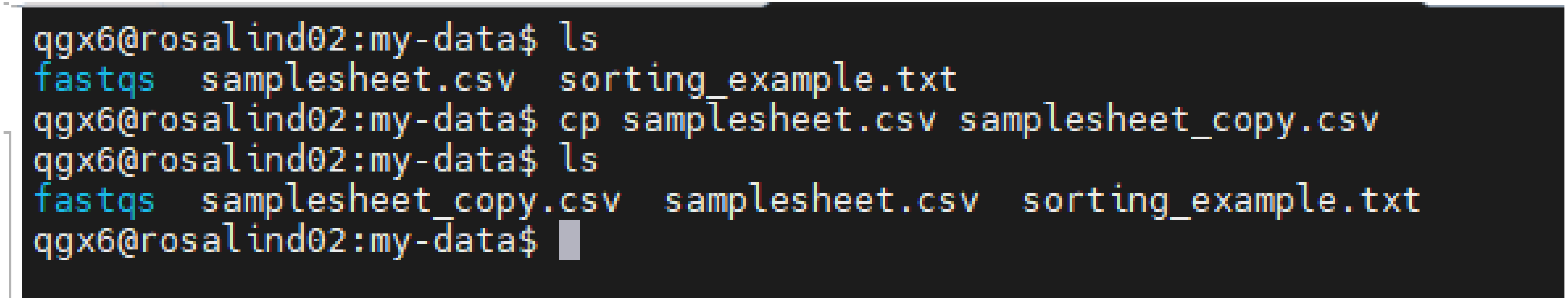
sort samplesheet.csv | head -5



```
qgx6@rosalind02:~$ sort samplesheet.csv | head -5
barcode03,1240503027,Test
barcode11,2240514007,Test
barcode17,2240510024,Test
barcode18,1240511016,Test
barcode20,1240512023,Test
```

File Viewing and manipulation

- `cp` **: Copy files or directories** from first arg (source) to second arg (destination)
  - Copying directories typically needs -R flag
- `mv` : Move or rename files
- `rm` : Remove files or directories
- `touch` : Create empty file



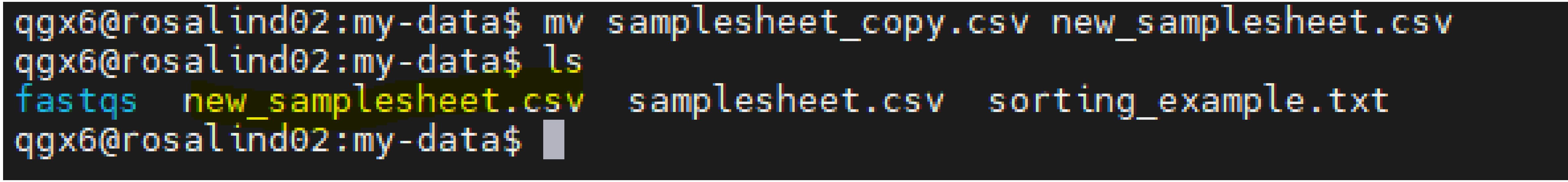
```
qgx6@rosalind02:my-data$ ls
fastqs  samplesheet.csv  sorting_example.txt
qgx6@rosalind02:my-data$ cp samplesheet.csv samplesheet_copy.csv
qgx6@rosalind02:my-data$ ls
fastqs  samplesheet_copy.csv  samplesheet.csv  sorting_example.txt
qgx6@rosalind02:my-data$ 
```

File Viewing and manipulation

`cp` **: Copy files or directories** from first arg (source) to second arg (destination)

  - Copying directories typically needs -R flag

- `mv` : Move or rename files
- `rm` : Remove files or directories
- `touch` : Create empty file



```
qgx6@rosalind02:my-data$ mv samplesheet_copy.csv new_samplesheet.csv
qgx6@rosalind02:my-data$ ls
fastqs  new_samplesheet.csv  samplesheet.csv  sorting_example.txt
qgx6@rosalind02:my-data$ 
```

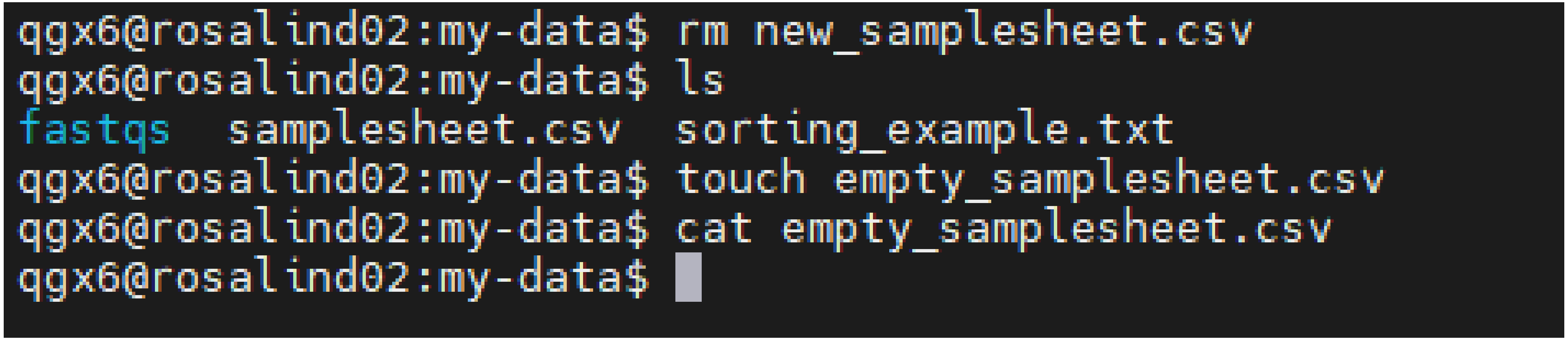
File Viewing and manipulation

`cp` **: Copy files or directories** from first arg (source) to second arg (destination)

- Copying directories typically needs -R flag

- `mv` : Move or rename files
- `rm` : Remove files or directories -- **CAUTION: This is *permanent deletion***

- `touch` : Create empty file

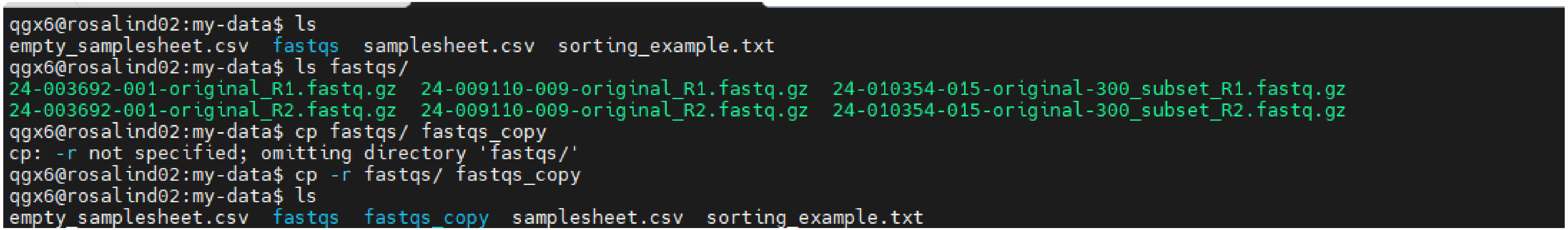


```
qgx6@rosalind02:my-data$ rm new_samplesheet.csv
qgx6@rosalind02:my-data$ ls
fastqs  samplesheet.csv  sorting_example.txt
qgx6@rosalind02:my-data$ touch empty_samplesheet.csv
qgx6@rosalind02:my-data$ cat empty_samplesheet.csv
qgx6@rosalind02:my-data$ 
```

File Viewing and manipulation

- **cp : Copy files or directories** from first arg (source) to second arg (destination)
- **mv : Move or rename files**
- **rm : Remove files or directories**
- touch : Create empty file

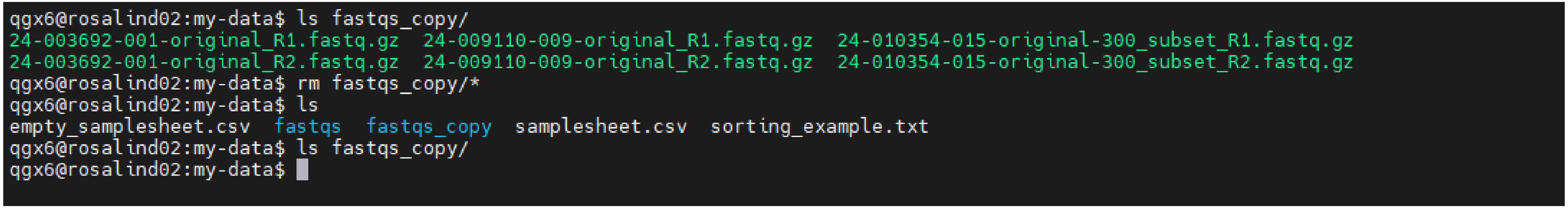
- Copying and removing a directory: need recursive flag



```
qgx6@rosalind02:my-data$ ls
empty_samplesheet.csv  fastqs  samplesheet.csv  sorting_example.txt
qgx6@rosalind02:my-data$ ls fastqs/
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz  24-010354-015-original-300_subset_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz  24-010354-015-original-300_subset_R2.fastq.gz
qgx6@rosalind02:my-data$ cp fastqs/ fastqs_copy
cp: -r not specified; omitting directory 'fastqs/'
qgx6@rosalind02:my-data$ cp -r fastqs/ fastqs_copy
qgx6@rosalind02:my-data$ ls
empty_samplesheet.csv  fastqs  fastqs_copy  samplesheet.csv  sorting_example.txt
```

File Viewing and manipulation

- **cp : Copy files or directories** from first arg (source) to second arg (destination)
- **mv : Move or rename files**
- **rm : Remove files or directories**
- touch : Create empty file

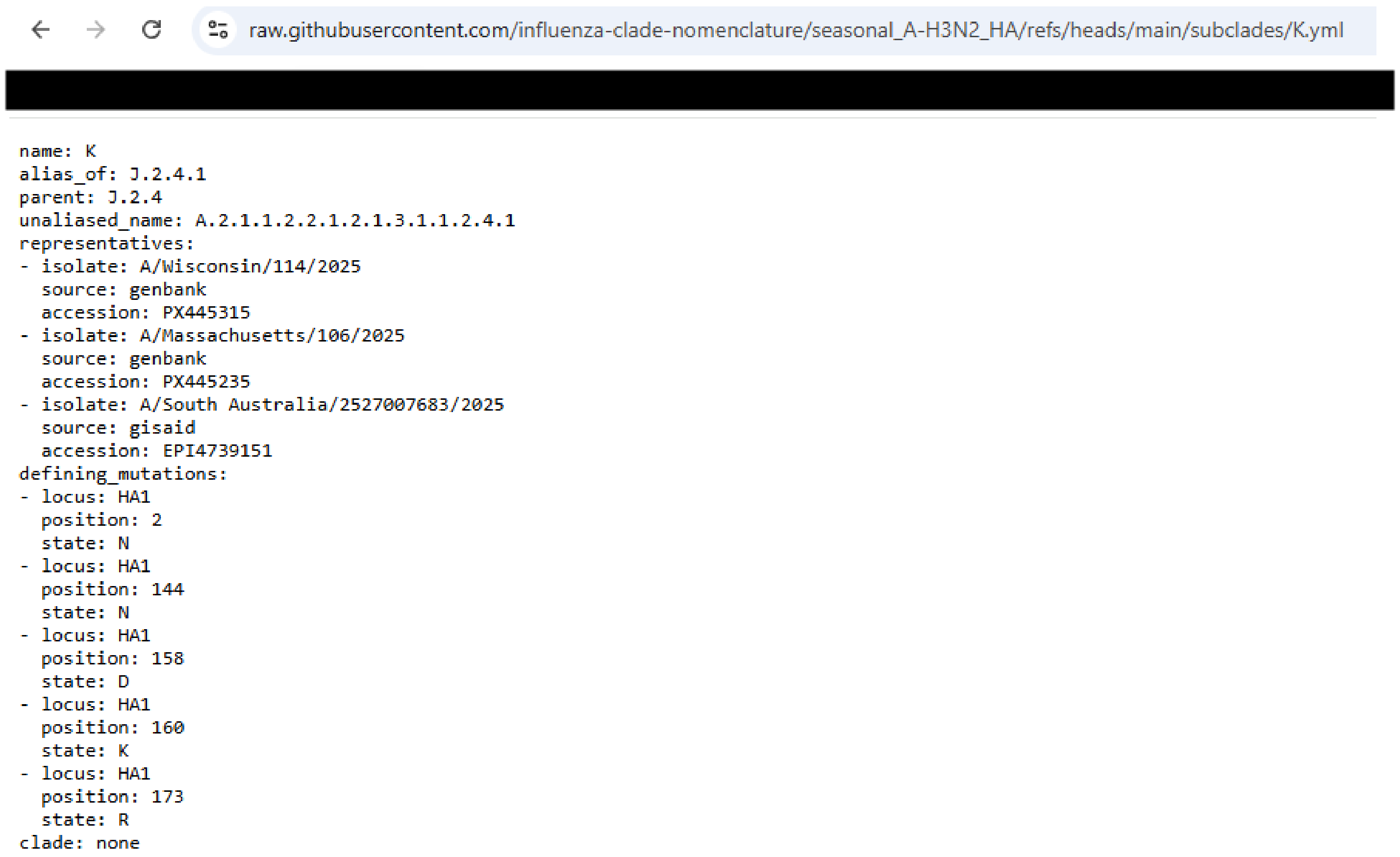


```
qgx6@rosalind02:my-data$ ls fastqs_copy/
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz  24-010354-015-original-300_subset_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz  24-010354-015-original-300_subset_R2.fastq.gz
qgx6@rosalind02:my-data$ rm fastqs_copy/*
qgx6@rosalind02:my-data$ ls
empty_samplesheet.csv  fastqs  fastqs_copy  samplesheet.csv  sorting_example.txt
qgx6@rosalind02:my-data$ ls fastqs_copy/
qgx6@rosalind02:my-data$ 
```

To remove a directory AND everything in it, use recursive flag:
rm –r directory

Network and Downloads part 1

- **wget : Download files from the web**
- **curl : Transfer data to/from URLs**



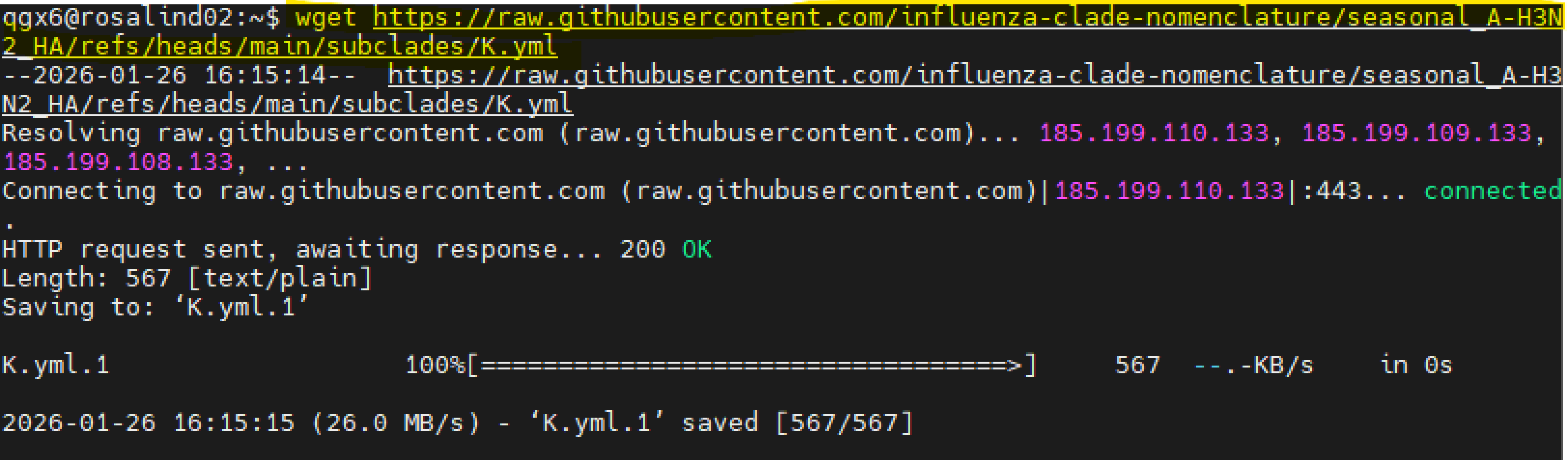
raw.githubusercontent.com/influenza-clade-nomenclature/seasonal_A-H3N2_HA/refs/heads/main/subclades/K.yml

[redacted]

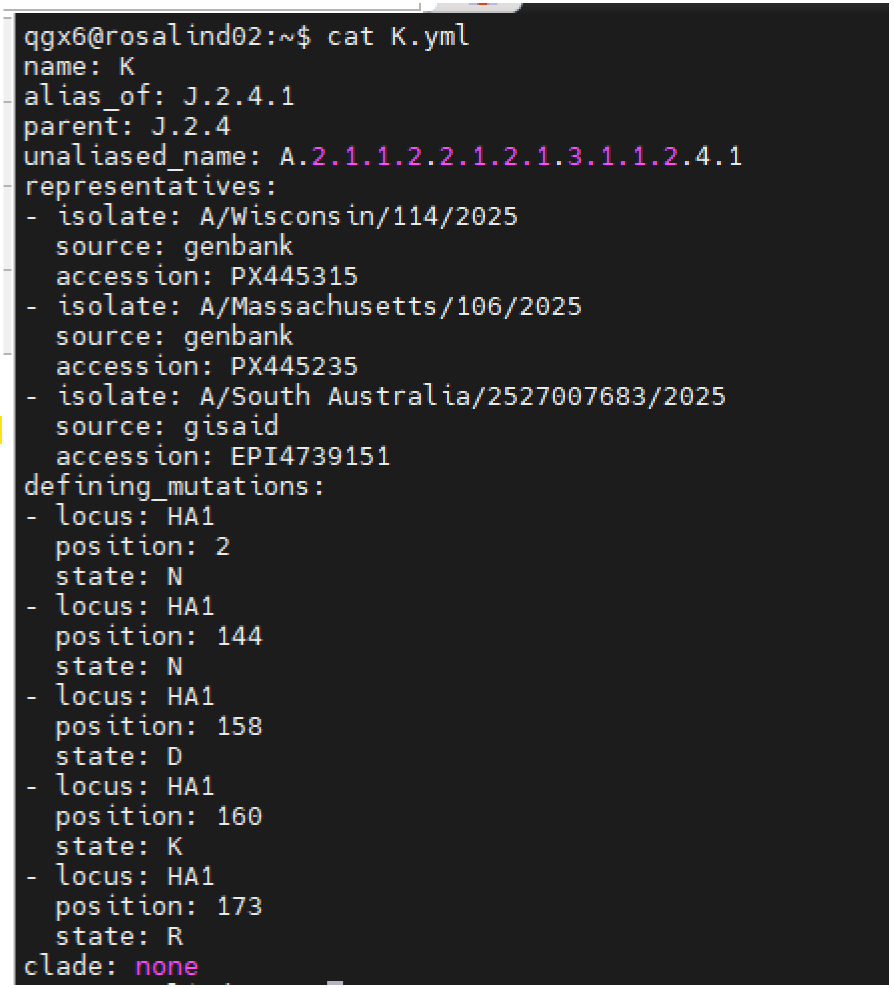
```
name: K
alias_of: J.2.4.1
parent: J.2.4
unaliased_name: A.2.1.1.2.2.1.2.1.3.1.1.2.4.1
representatives:
- isolate: A/Wisconsin/114/2025
  source: genbank
  accession: PX445315
- isolate: A/Massachusetts/106/2025
  source: genbank
  accession: PX445235
- isolate: A/South Australia/2527007683/2025
  source: gisaid
  accession: EPI4739151
defining_mutations:
- locus: HA1
  position: 2
  state: N
- locus: HA1
  position: 144
  state: N
- locus: HA1
  position: 158
  state: D
- locus: HA1
  position: 160
  state: K
- locus: HA1
  position: 173
  state: R
clade: none
```

Network and Downloads part 1

- **wget : Download files from the web**
- **curl : Transfer data to/from URLs**
  - wget will pull the file, curl will pull the contents
  - Use a redirect (> or >>) with curl



```
qgx6@rosalind02:~$ wget https://raw.githubusercontent.com/influenza-clade-nomenclature/seasonal_A-H3N
2_HA/refs/heads/main/subclades/K.yml
--2026-01-26 16:15:14--  https://raw.githubusercontent.com/influenza-clade-nomenclature/seasonal_A-H3
N2_HA/refs/heads/main/subclades/K.yml
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.110.133, 185.199.109.133,
185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.110.133|:443... connected
.
HTTP request sent, awaiting response... 200 OK
Length: 567 [text/plain]
Saving to: ‘K.yml.1’

K.yml.1                   100%[===================================>]     567  --.-KB/s    in 0s

2026-01-26 16:15:15 (26.0 MB/s) - ‘K.yml.1’ saved [567/567]
```



```
qgx6@rosalind02:~$ cat K.yml
name: K
alias_of: J.2.4.1
parent: J.2.4
unaliased_name: A.2.1.1.2.2.1.2.1.3.1.1.2.4.1
representatives:
- isolate: A/Wisconsin/114/2025
  source: genbank
  accession: PX445315
- isolate: A/Massachusetts/106/2025
  source: genbank
  accession: PX445235
- isolate: A/South Australia/2527007683/2025
  source: gisaid
  accession: EPI4739151
defining_mutations:
- locus: HA1
  position: 2
  state: N
- locus: HA1
  position: 144
  state: N
- locus: HA1
  position: 158
  state: D
- locus: HA1
  position: 160
  state: K
- locus: HA1
  position: 173
  state: R
clade: none
```

Network and Downloads Practical

- **wget : Download files from the web**
- **curl : Transfer data to/from URLs**

- **Download fastqs and samplesheet from our website for the next practical**

```
wget https://raw.githubusercontent.com/CDCgov/id-bioifx-workshop/refs/heads/main/practical/samplesheet.csv
```

```
curl https://raw.githubusercontent.com/CDCgov/id-bioifx-workshop/refs/heads/main/practical/samplesheet.csv > samplesheet.csv
```

File viewing and Manipulation Practical

1. Ensure your samplesheet.csv on the system pulled properly with which command?
2. Copy your samplesheet to make samplesheet_copy.csv
3. Rename your samplesheet_copy.csv to samplesheet2.csv using which command?
4. Output the last 20 lines of your samplesheet, and using redirect, write them to another file called samplesheet_subset.csv
5. Using `head` and then piping into `tail`, output only the 8th line of samplesheet.csv
6. Sort the samplesheet according to sample ID, numerically, then redirect to a file called samplesheet_sorted.csv

File viewing and Manipulation Practical

7. Cat samplesheet_sorted.csv to *append to* samplesheet2.csv. Cat samplesheet2.csv, it should look twice as long!

8. Cat samplesheet_sorted.csv to *overwrite* samplehseet2.csv. Cat samplesheet2.csv, it should look just like samplesheet_sorted.csv
9. Rename samplesheet2 back to samplesheet_copy.csv

10. Using rm and glob (*) remove *both* samplesheet_copy.csv and samplesheet_sorted.csv without removing samplesheet.csv

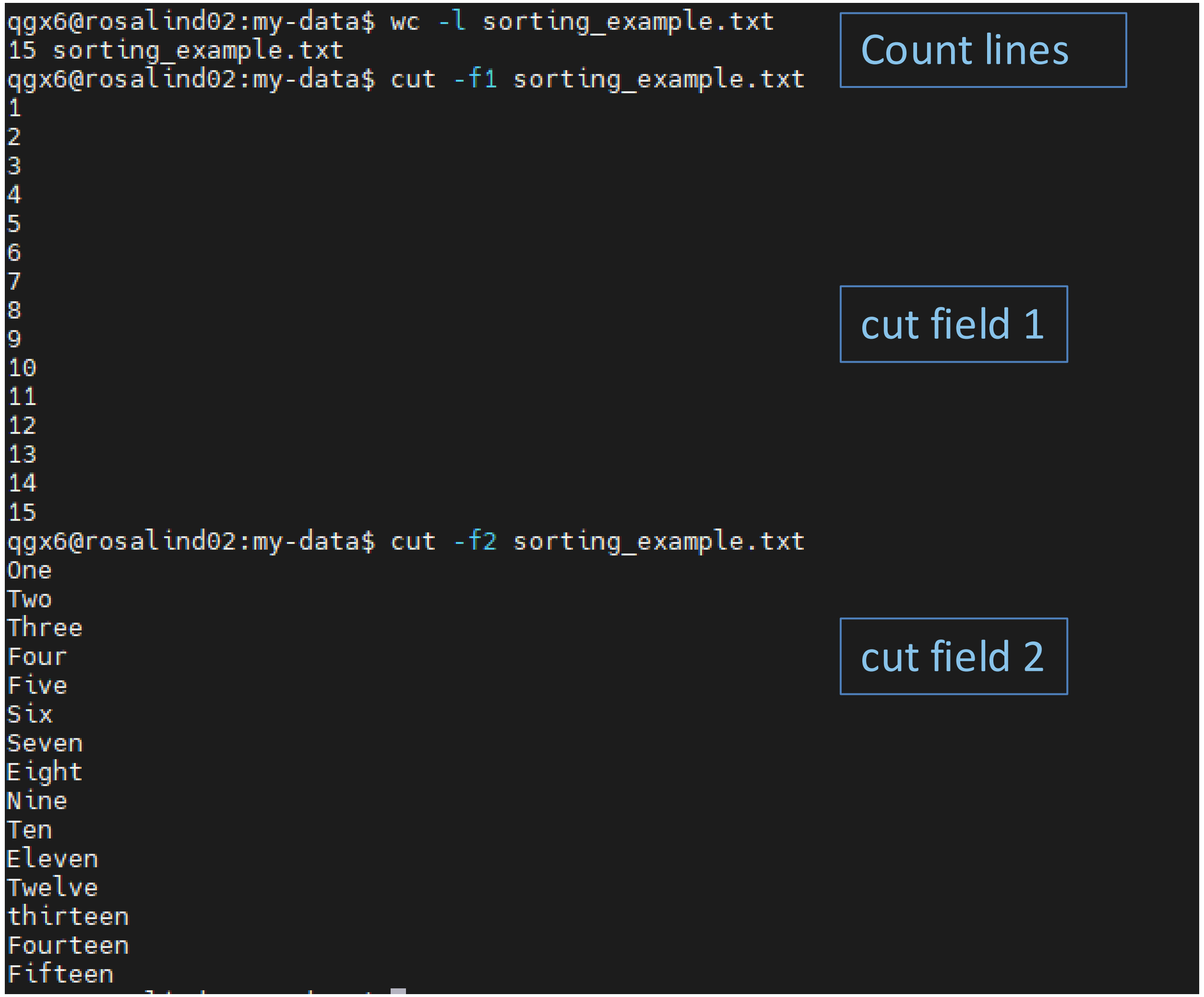
File viewing and Manipulation Practical

Searching and Text Processing

`wc` : **Count lines, words, and bytes**

`cut` : **Remove sections from lines**

`uniq`: Report or filter duplicate lines



```
qgx6@rosalind02:my-data$ wc -l sorting_example.txt
15 sorting_example.txt
qgx6@rosalind02:my-data$ cut -f1 sorting_example.txt
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
qgx6@rosalind02:my-data$ cut -f2 sorting_example.txt
One
Two
Three
Four
Five
Six
Seven
Eight
Nine
Ten
Eleven
Twelve
thirteen
Fourteen
Fifteen
```

Count lines

cut field 1

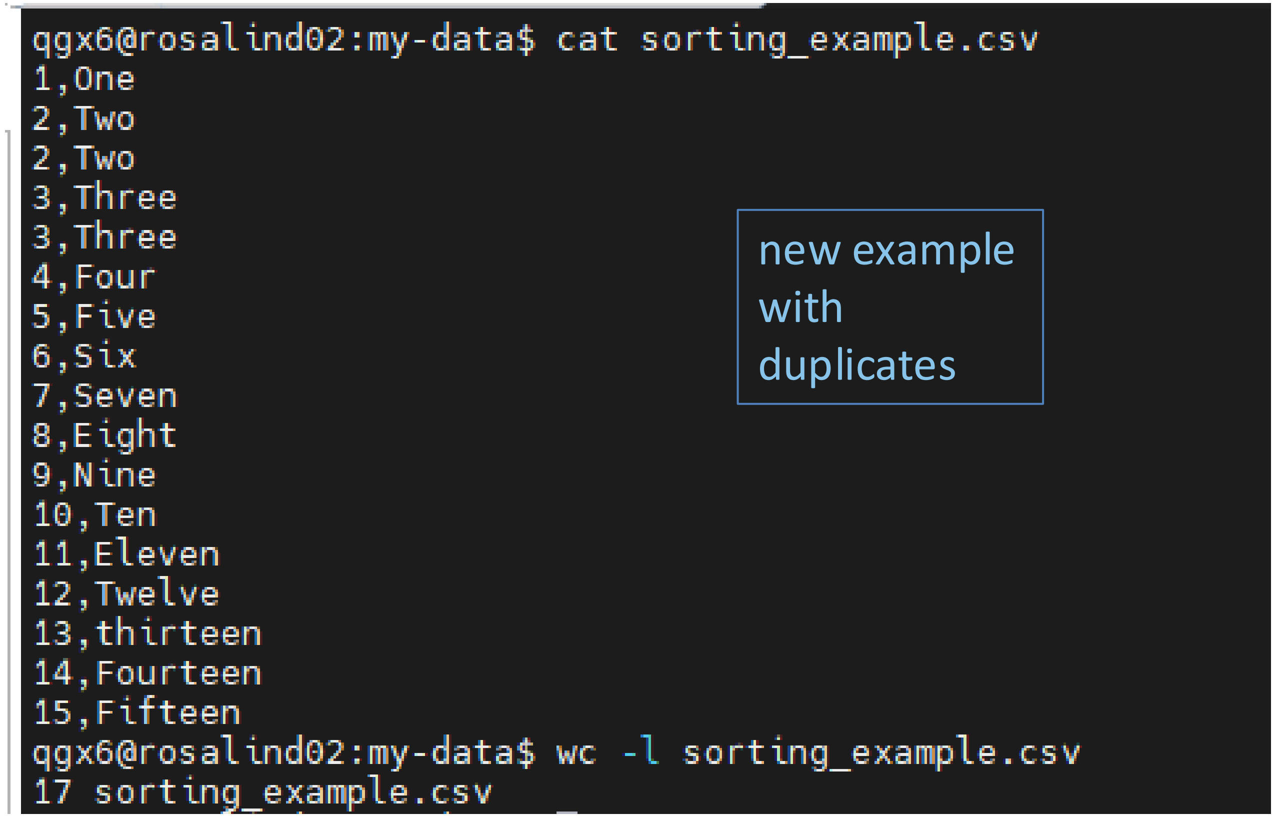
cut field 2

Searching and Text Processing

`wc` : Count lines, words, and bytes

`cut` : Remove sections from lines

`uniq`: Report or filter duplicate lines



```
qgx6@rosalind02:my-data$ cat sorting_example.csv
1,One
2,Two
2,Two
3,Three
3,Three
4,Four
5,Five
6,Six
7,Seven
8,Eight
9,Nine
10,Ten
11,Eleven
12,Twelve
13,thirteen
14,Fourteen
15,Fifteen
qgx6@rosalind02:my-data$ wc -l sorting_example.csv
17 sorting_example.csv
```

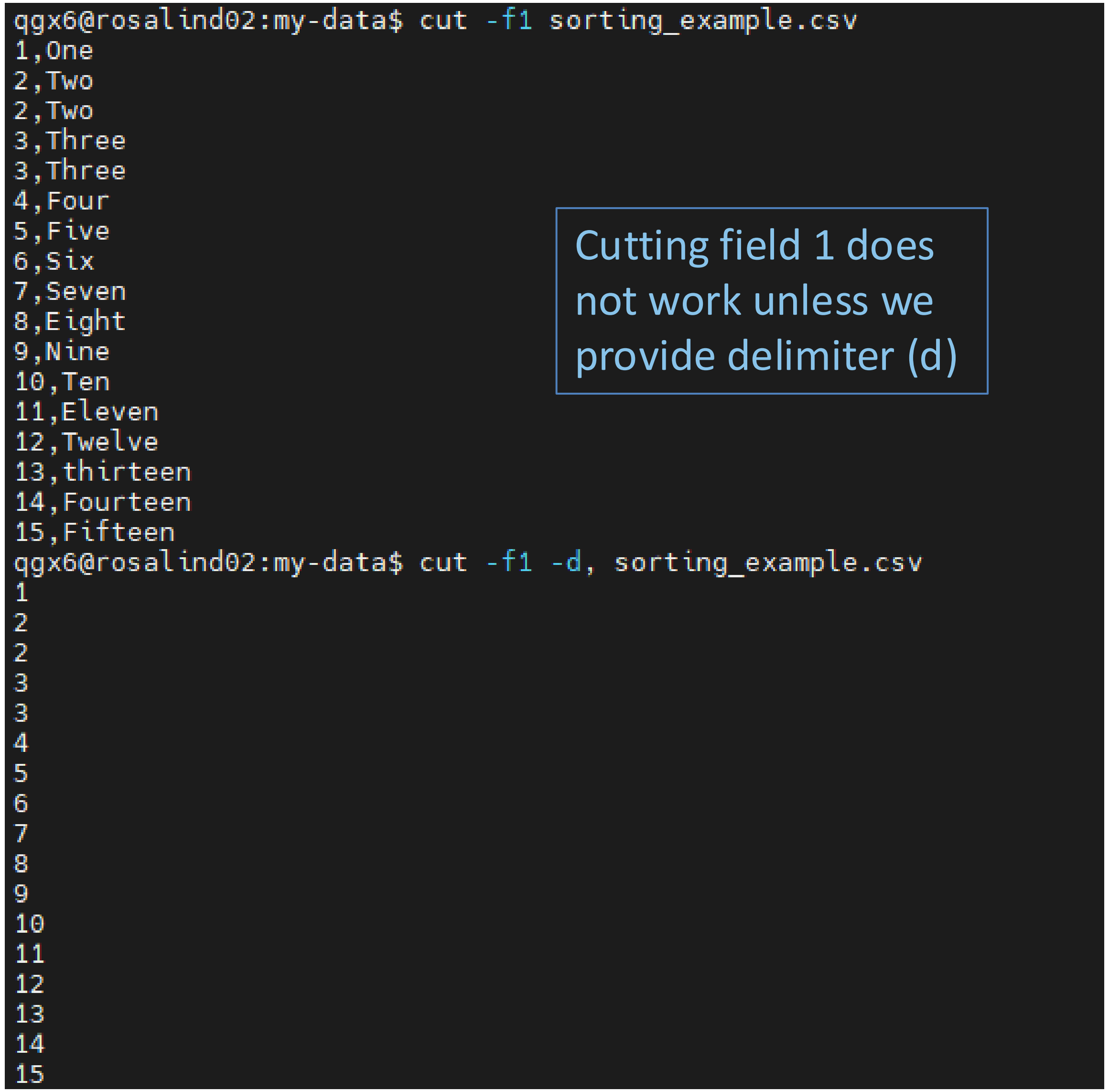
new example with duplicates

Searching and Text Processing

`wc` : Count lines, words, and bytes

`cut` : Remove sections from lines

`uniq`: Report or filter duplicate lines



```
qgx6@rosalind02:my-data$ cut -f1 sorting_example.csv
1,One
2,Two
2,Two
3,Three
3,Three
4,Four
5,Five
6,Six
7,Seven
8,Eight
9,Nine
10,Ten
11,Eleven
12,Twelve
13,thirteen
14,Fourteen
15,Fifteen
qgx6@rosalind02:my-data$ cut -f1 -d, sorting_example.csv
1
2
2
3
3
4
5
6
7
8
9
10
11
12
13
14
15
```

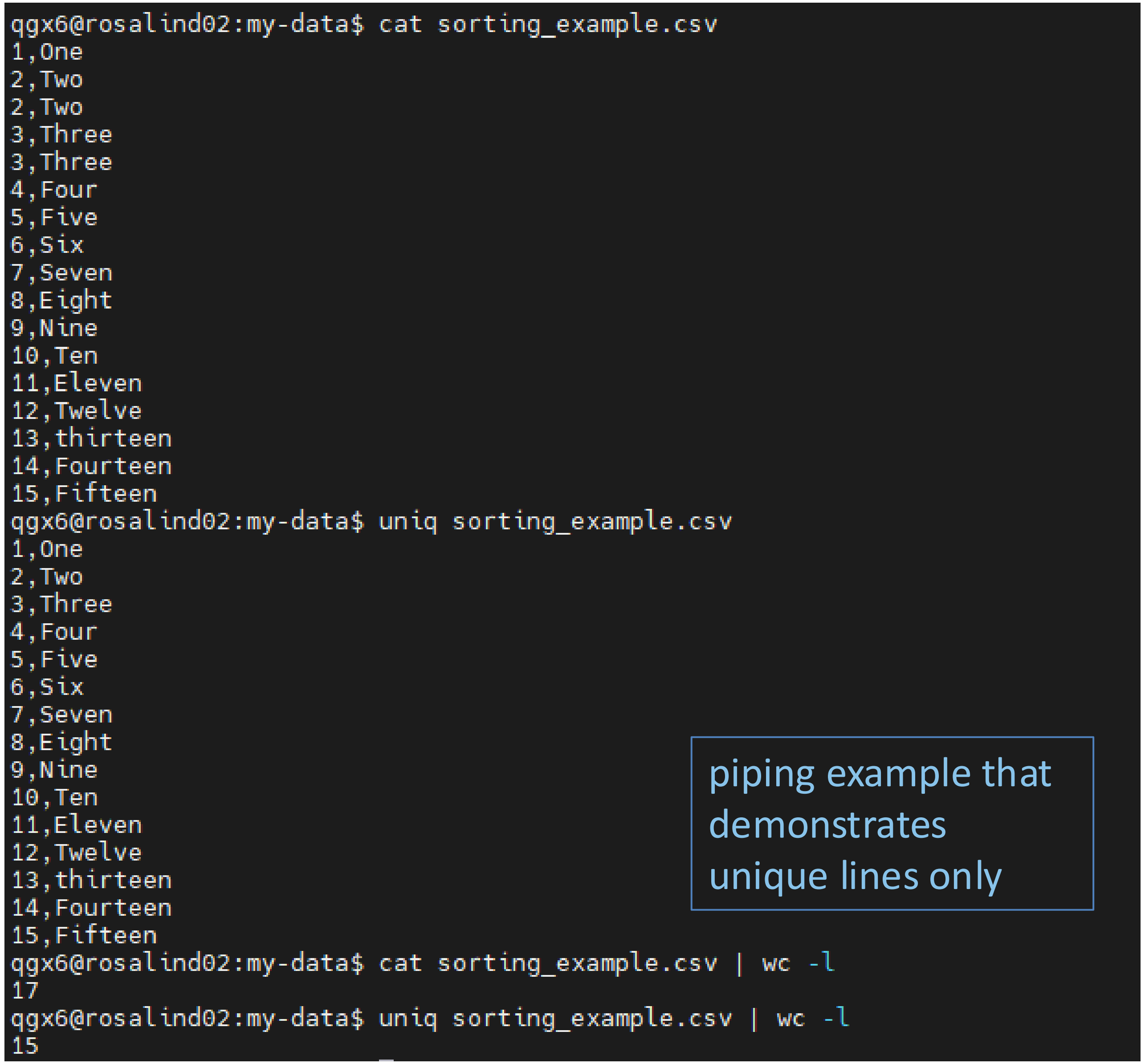
Cutting field 1 does not work unless we provide delimiter (d)

Searching and Text Processing

`wc` **: Count lines, words, and bytes**

`cut` **: Remove sections from lines**

`uniq`: Report or filter duplicate lines



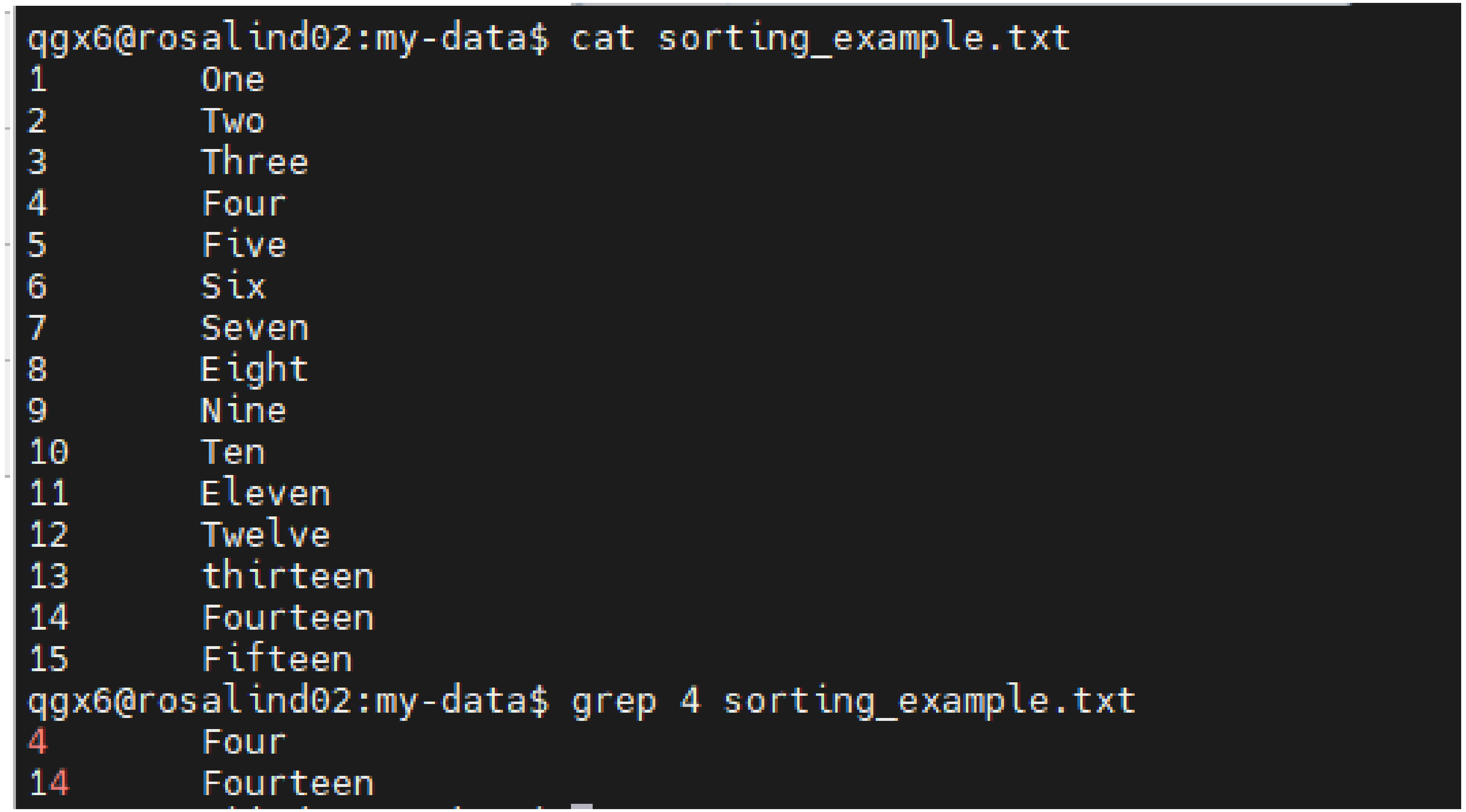
```
qgx6@rosalind02:my-data$ cat sorting_example.csv
1,One
2,Two
2,Two
3,Three
3,Three
4,Four
5,Five
6,Six
7,Seven
8,Eight
9,Nine
10,Ten
11,Eleven
12,Twelve
13,thirteen
14,Fourteen
15,Fifteen
qgx6@rosalind02:my-data$ uniq sorting_example.csv
1,One
2,Two
3,Three
4,Four
5,Five
6,Six
7,Seven
8,Eight
9,Nine
10,Ten
11,Eleven
12,Twelve
13,thirteen
14,Fourteen
15,Fifteen
qgx6@rosalind02:my-data$ cat sorting_example.csv | wc -l
17
qgx6@rosalind02:my-data$ uniq sorting_example.csv | wc -l
15
```

piping example that demonstrates unique lines only

Searching and Text Processing

`grep`: **Search for text in files**

`sed` : Stream editor for find/replace

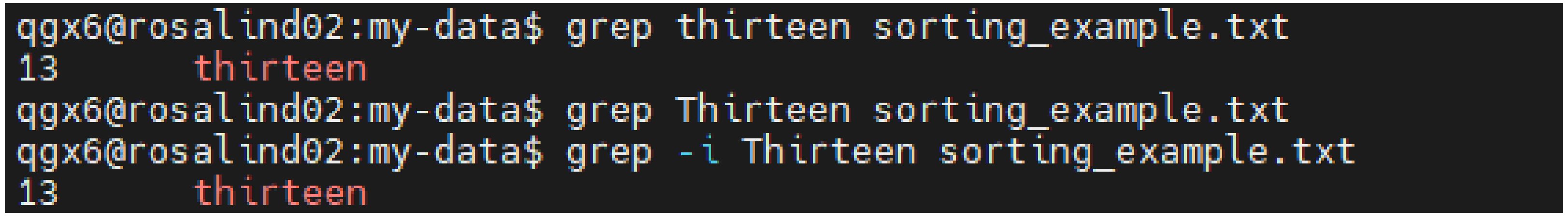


```
qgx6@rosalind02:my-data$ cat sorting_example.txt
1       One
2       Two
3       Three
4       Four
5       Five
6       Six
7       Seven
8       Eight
9       Nine
10      Ten
11      Eleven
12      Twelve
13      thirteen
14      Fourteen
15      Fifteen
qgx6@rosalind02:my-data$ grep 4 sorting_example.txt
4       Four
14      Fourteen
```

Searching and Text Processing

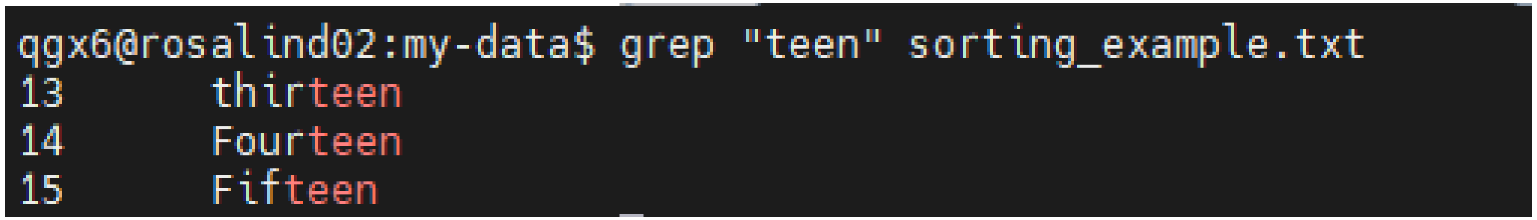
`grep`: **Search for text in files**

`sed` : Stream editor for find/replace



```
qgx6@rosalind02:my-data$ grep thirteen sorting_example.txt
13      thirteen
qgx6@rosalind02:my-data$ grep Thirteen sorting_example.txt
qgx6@rosalind02:my-data$ grep -i Thirteen sorting_example.txt
13      thirteen
```

grep is case sensitive unless you use -i



```
qgx6@rosalind02:my-data$ grep "teen" sorting_example.txt
13      thirteen
14      Fourteen
15      Fifteen
```

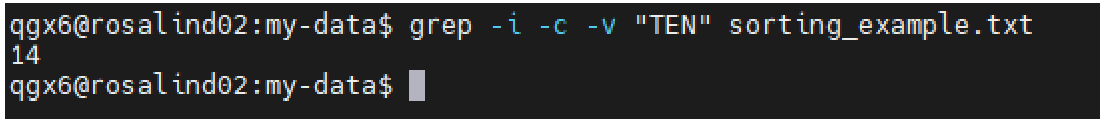
Any text is searchable, does not have to match entire word

Searching and Text Processing

grep — Most useful flags

- **-v** — *Invert match*
  - Shows lines that **do NOT** match the pattern
- **-i** — *Ignore case*
  - Matches text regardless of capitalization
- **-r** — *Recursive search*
  - Searches all files in a directory tree
- **-n** — *Show line numbers*
  - Displays line numbers for each match
- **-c** — *Count matches*
  - Outputs the number of matching lines

Remember: Flags can be combined! grep -i –v –c will ignore case AND invert match and count matches



```
qgx6@rosalind02:my-data$ grep -i -c -v "TEN" sorting_example.txt
14
qgx6@rosalind02:my-data$
```

Searching and Text Processing

grep : Search for text in files

sed : Stream editor for find/replace

sed -s "s/find/replace/g" <file>

- Sends replaced line to output stream, can pipe (|) or redirect (>, >>)
- “g” = “global”, replaces _all_ matches

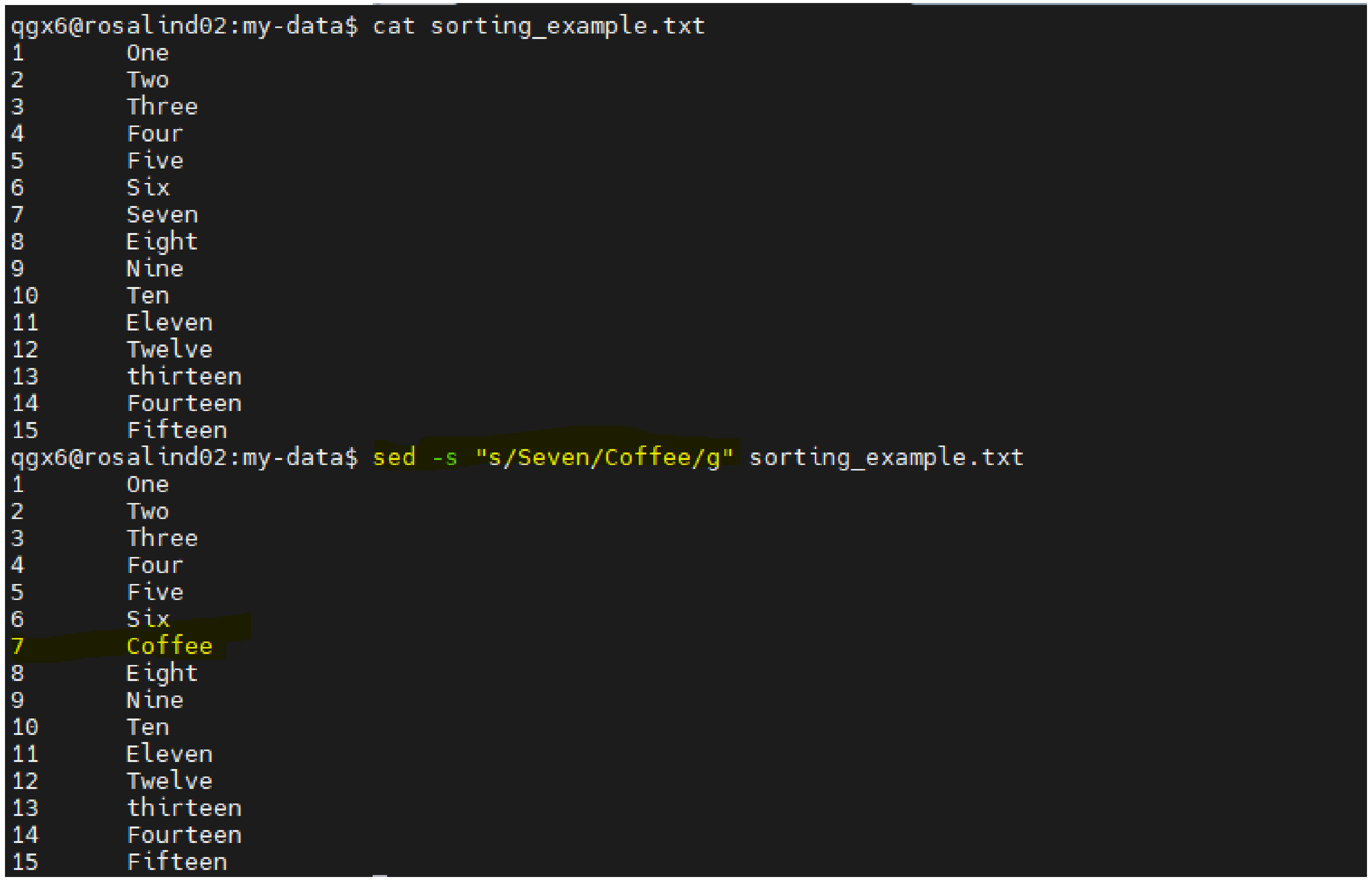
sed -i "s/find/replace/g" <file>

- Replaces within your file
- Be careful! Not easy to undo if you make a mistake
- TIP: add `-i .bak` copies original file to original.bak first

Searching and Text Processing

grep : Search for text in files

sed : Stream editor for find/replace



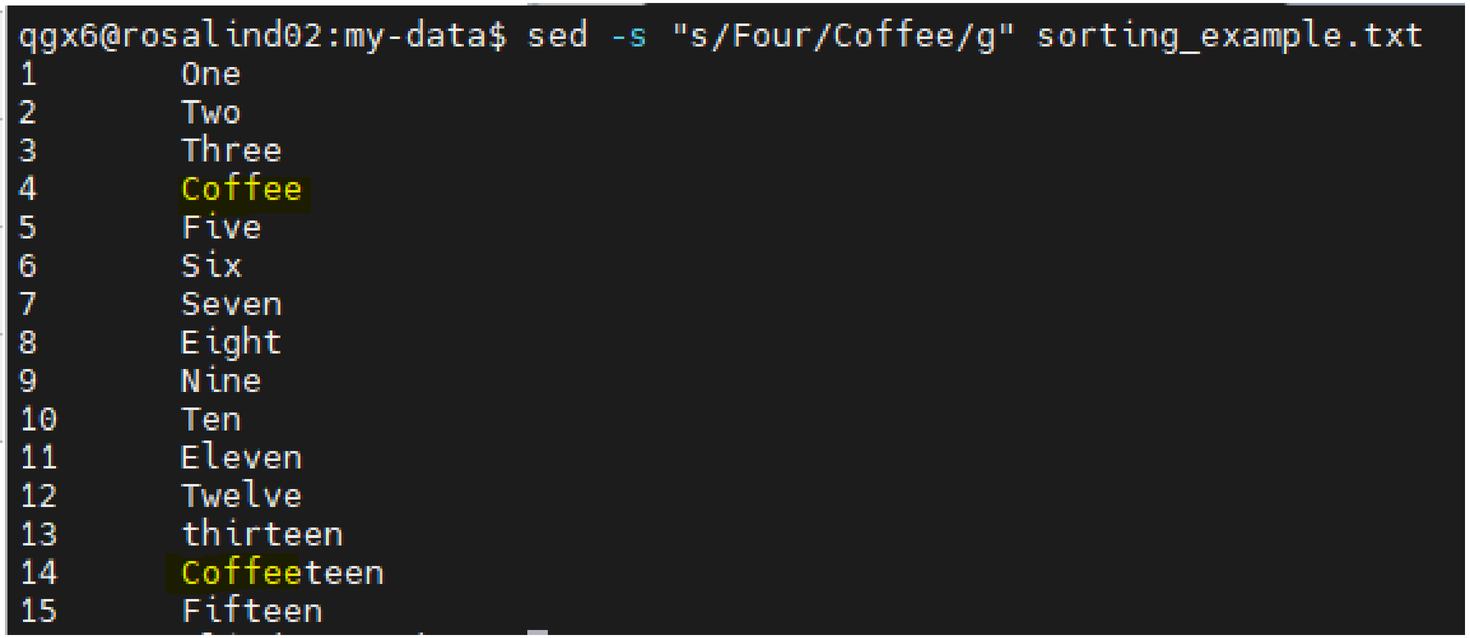
```
qgx6@rosalind02:my-data$ cat sorting_example.txt
1       One
2       Two
3       Three
4       Four
5       Five
6       Six
7       Seven
8       Eight
9       Nine
10      Ten
11      Eleven
12      Twelve
13      thirteen
14      Fourteen
15      Fifteen
qgx6@rosalind02:my-data$ sed -s "s/Seven/Coffee/g" sorting_example.txt
1       One
2       Two
3       Three
4       Four
5       Five
6       Six
7       Coffee
8       Eight
9       Nine
10      Ten
11      Eleven
12      Twelve
13      thirteen
14      Fourteen
15      Fifteen
```

Searching and Text Processing

grep : Search for text in files

sed : Stream editor for find/replace

sed is very powerful and it is easy to make inadvertent changes



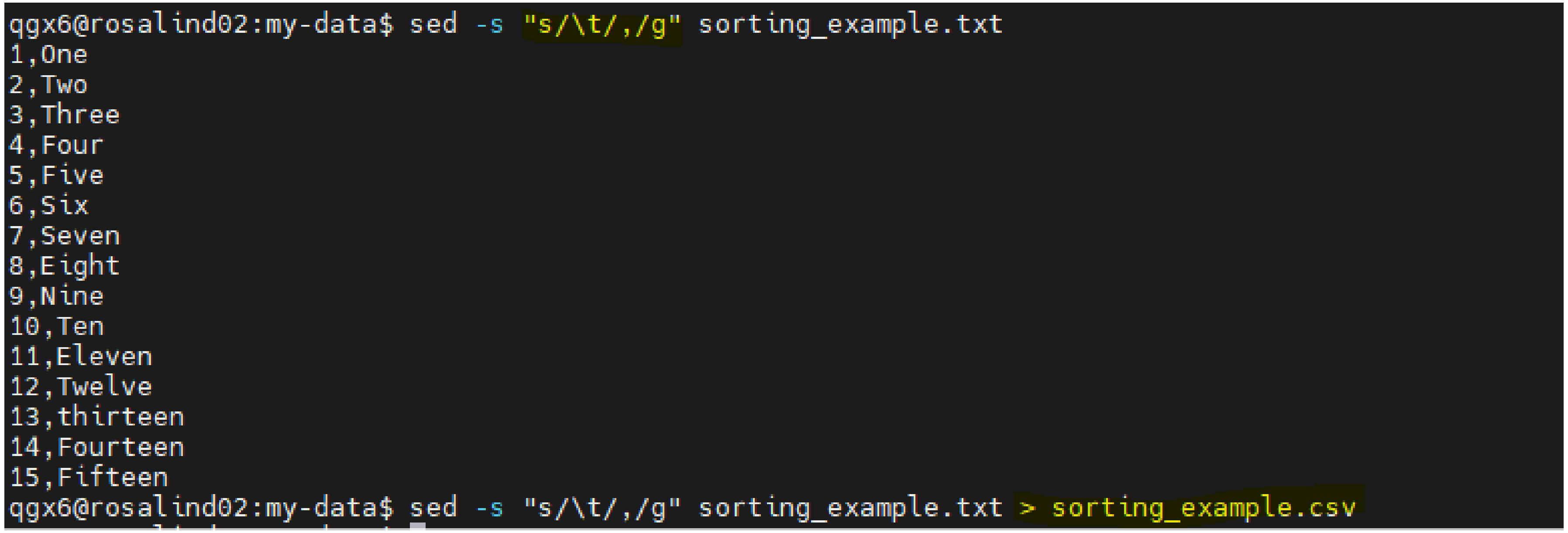
```
qgx6@rosalind02:my-data$ sed -s "s/Four/Coffee/g" sorting_example.txt
1       One
2       Two
3       Three
4       Coffee
5       Five
6       Six
7       Seven
8       Eight
9       Nine
10      Ten
11      Eleven
12      Twelve
13      thirteen
14      Coffeeteen
15      Fifteen
```

Searching and Text Processing

grep : Search for text in files

sed : Stream editor for find/replace

Special character example: tsv to csv



```
qgx6@rosalind02:my-data$ sed -s "s/\t/,/g" sorting_example.txt
1,One
2,Two
3,Three
4,Four
5,Five
6,Six
7,Seven
8,Eight
9,Nine
10,Ten
11,Eleven
12,Twelve
13,thirteen
14,Fourteen
15,Fifteen
qgx6@rosalind02:my-data$ sed -s "s/\t/,/g" sorting_example.txt > sorting_example.csv
```

Searching and Text Processing

grep : Search for text in files

sed : Stream editor for find/replace

```
sed -s "s/find/replace/g”
```

What if the string we want to find or replace has a / in it?

You don’t have to use / to pass your sed command

```
sed -s "s%find%replace%g”
```

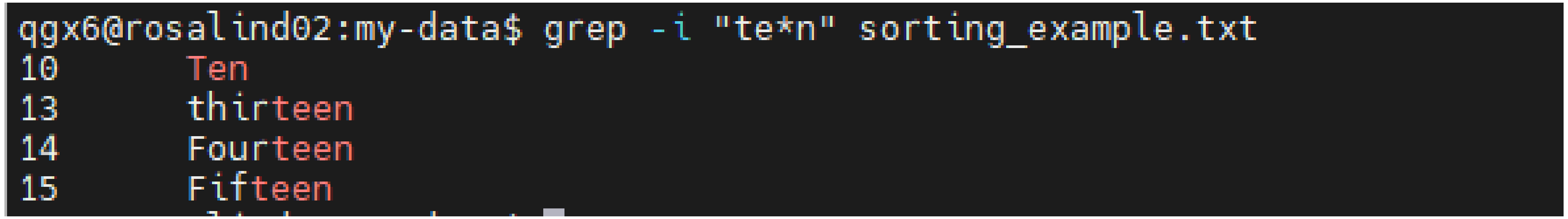
```
sed -s "s&find&replace&g”
```

Searching and Text Processing

Regex

Regular expression: pattern matching

- Used by grep, sed, other text editors
- Used to find and extract text in strings and files
- Most important symbols:
  - . (any character)
  - * (previous character zero or more times)
  - \+ (previous character one or more times)
  - [] (character sets)

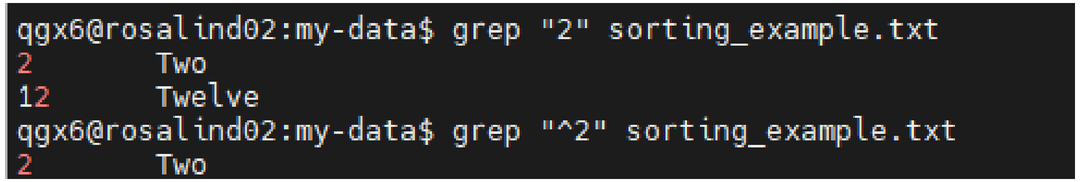


```
qgx6@rosalind02:my-data$ grep -i "te*n" sorting_example.txt
10      Ten
13      thirteen
14      Fourteen
15      Fifteen
```

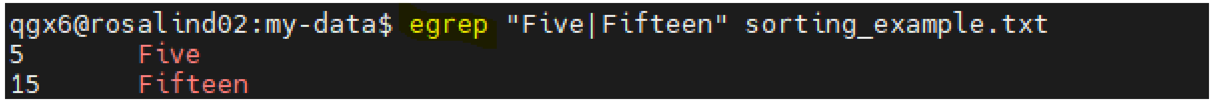
Searching and Text Processing

Intermediate regex

- **egrep or grep –e**
- **Anchors control position**
    - ^ matches start of line, $ matches end
    - Example: ^ERROR (lines that start with “ERROR”)
- **Character classes & ranges**
    - [abc] matches a, b, or c; [0-9] matches digits
    - Example: user[0-9]+ (user1, user42)
- **Quantifiers control repetition**
    - * (0 or more), + (1 or more), {n} (exact count)
    - Example: .{8,} (at least 8 characters)
- **Alternation (OR)**
    - | matches one pattern or another
    - Example: error|fail|critical
- **Escaping special characters**
    - **\. matches a literal dot**, * a literal asterisk
    - Example: \d+\.\d+ (decimal numbers)



```
qgx6@rosalind02:my-data$ grep "2" sorting_example.txt
2       Two
12      Twelve
qgx6@rosalind02:my-data$ grep "^2" sorting_example.txt
2       Two
```

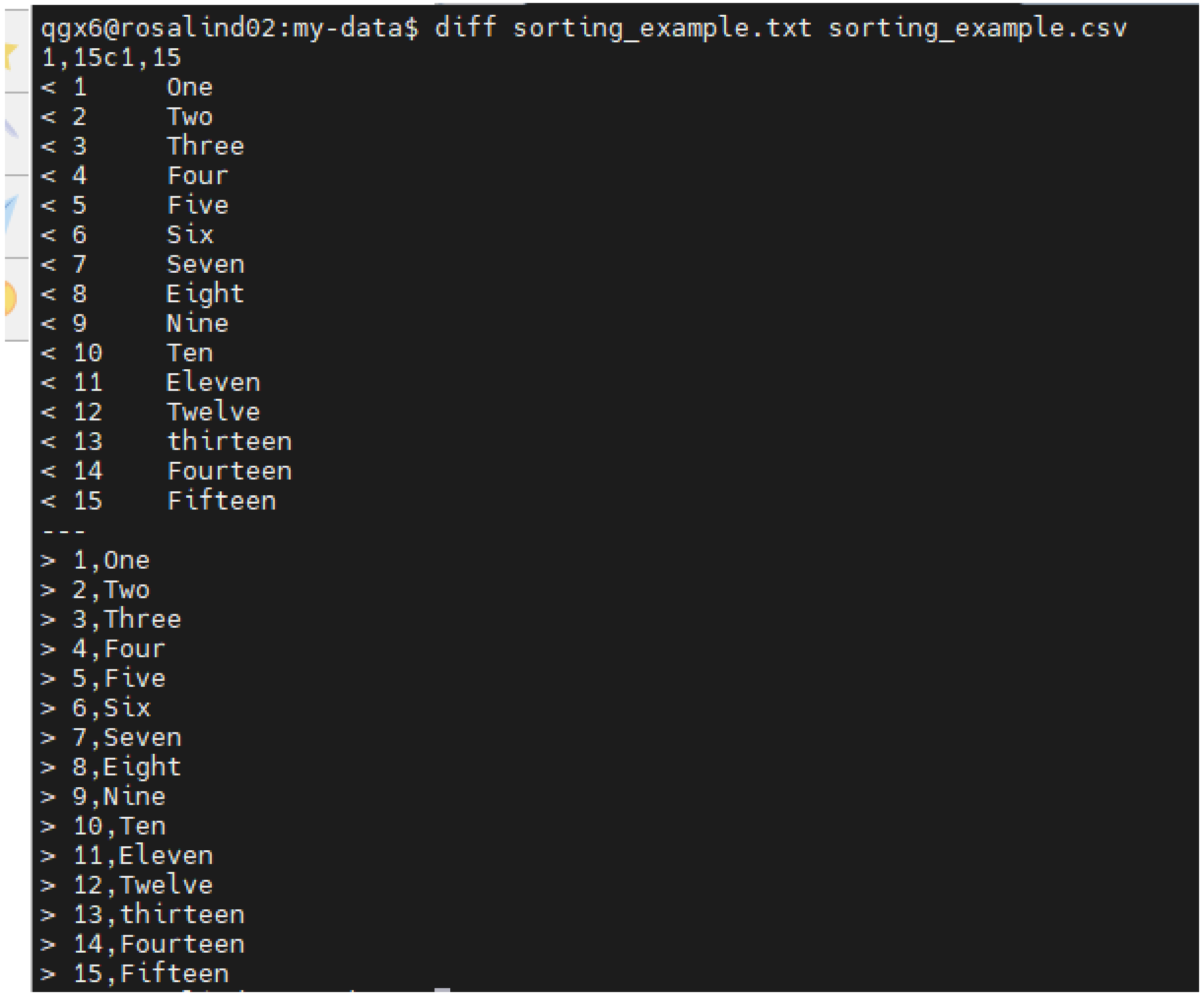


```
qgx6@rosalind02:my-data$ egrep "Five|Fifteen" sorting_example.txt
5       Five
15      Fifteen
```

www.regex101.com is a great resource for checking and testing regex

Intermediate Text Processing

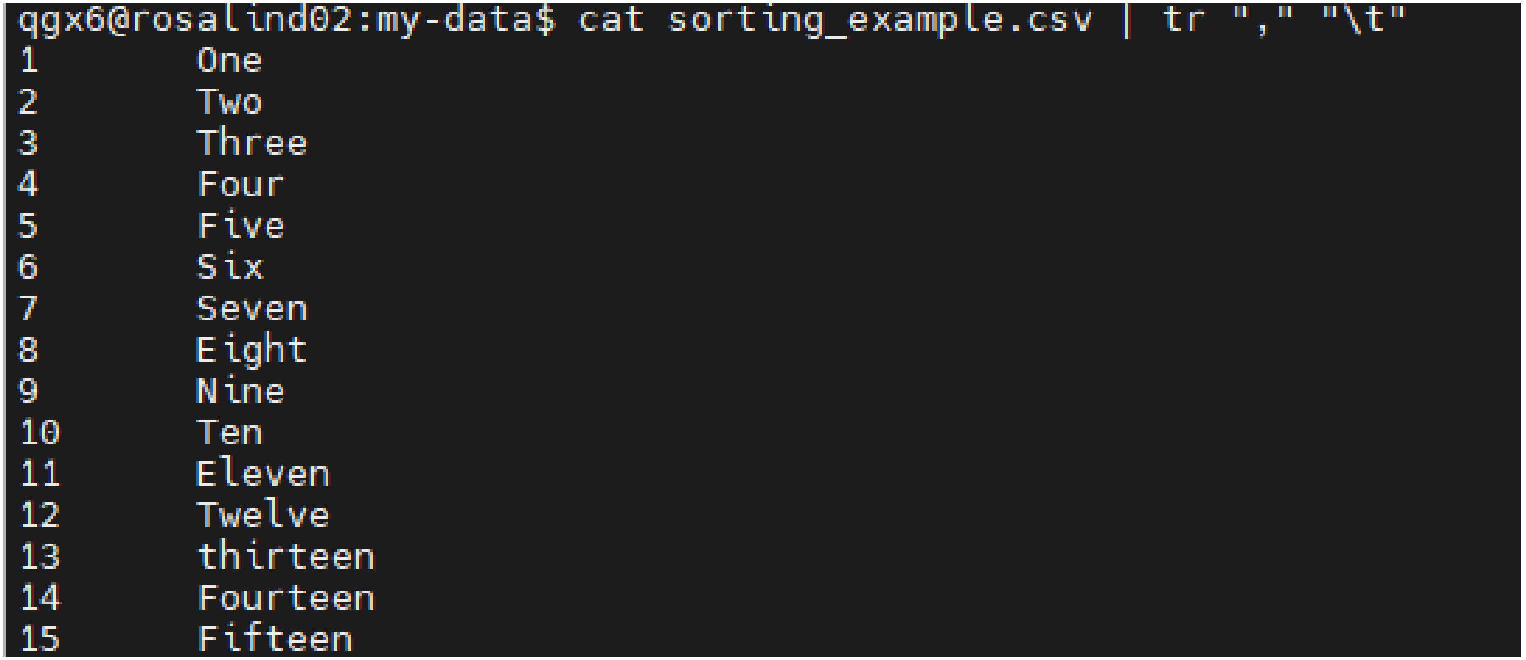
- diff : Compare files line by line
- comm : Compare sorted files
- tr : Translate or delete characters
- xargs : Build and execute command lines



```
qgx6@rosalind02:my-data$ diff sorting_example.txt sorting_example.csv
1,15c1,15
< 1      One
< 2      Two
< 3      Three
< 4      Four
< 5      Five
< 6      Six
< 7      Seven
< 8      Eight
< 9      Nine
< 10     Ten
< 11     Eleven
< 12     Twelve
< 13     thirteen
< 14     Fourteen
< 15     Fifteen
---
> 1,One
> 2,Two
> 3,Three
> 4,Four
> 5,Five
> 6,Six
> 7,Seven
> 8,Eight
> 9,Nine
> 10,Ten
> 11,Eleven
> 12,Twelve
> 13,thirteen
> 14,Fourteen
> 15,Fifteen
```

Intermediate Text Processing

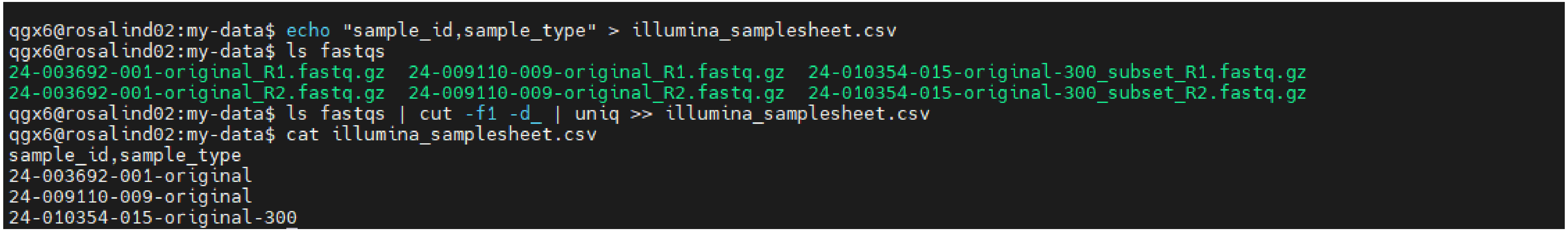
- diff : Compare files line by line
- comm : Compare sorted files
- tr : Translate or delete characters
- xargs : Build and execute command lines



```
qgx6@rosalind02:my-data$ cat sorting_example.csv | tr "," "\t"
1       One
2       Two
3       Three
4       Four
5       Five
6       Six
7       Seven
8       Eight
9       Nine
10      Ten
11      Eleven
12      Twelve
13      thirteen
14      Fourteen
15      Fifteen
```

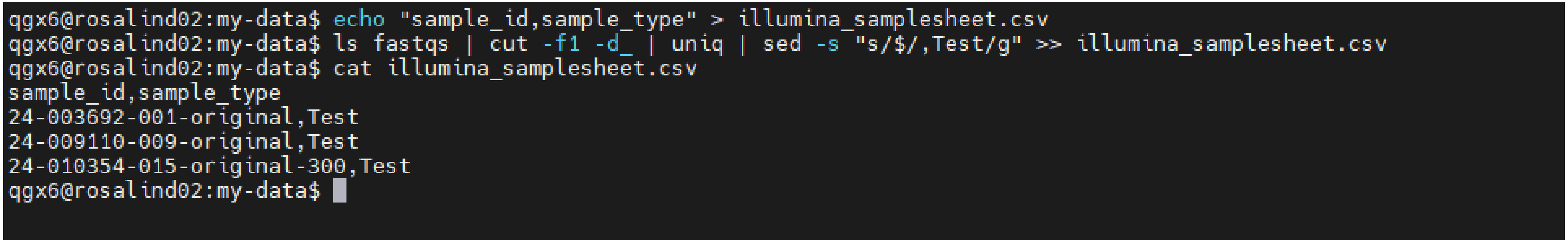
Intermediate Text Processing

- Putting it all together: real world example



```
qgx6@rosalind02:my-data$ echo "sample_id,sample_type" > illumina_samplesheet.csv
qgx6@rosalind02:my-data$ ls fastqs
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz  24-010354-015-original-300_subset_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz  24-010354-015-original-300_subset_R2.fastq.gz
qgx6@rosalind02:my-data$ ls fastqs | cut -f1 -d_ | uniq >> illumina_samplesheet.csv
qgx6@rosalind02:my-data$ cat illumina_samplesheet.csv
sample_id,sample_type
24-003692-001-original
24-009110-009-original
24-010354-015-original-300
```

- Need “sample_type” column, replace (sed) end of line ($) with ,Test



```
qgx6@rosalind02:my-data$ echo "sample_id,sample_type" > illumina_samplesheet.csv
qgx6@rosalind02:my-data$ ls fastqs | cut -f1 -d_ | uniq | sed -s "s/$/,Test/g" >> illumina_samplesheet.csv
qgx6@rosalind02:my-data$ cat illumina_samplesheet.csv
sample_id,sample_type
24-003692-001-original,Test
24-009110-009-original,Test
24-010354-015-original-300,Test
qgx6@rosalind02:my-data$
```

Searching and Text Processing Practical

1. Using wc on samplesheet.csv, how many lines are there?
2. Using wc on samplesheet.csv, how many characters are there?
3. Using cut on samplesheet.csv, output only the Barcode # field.
4. Make a file called sample_list.txt by cutting and redirecting unique sample IDs from your fastqs

Bonus: using tr and sort on samplesheet.csv, output the highest barcode number only (no “barcode”)

Searching and Text Processing Practical

1. Curl or wget https://raw.githubusercontent.com/CDCgov/id-bioifx-workshop/refs/heads/main/practical/influenza.fasta
2. Using grep with a flag or piping into `wc`, how many times does the following sequence appear in your influenza.fasta?

GGGGCGGGG (4G 1C 4G)

3. Using `grep` with the –A flag (--`help` if you don't recall what that does), output all the A_HA_H1 sequences to a file called H1_HA.fasta

H1_HA.fasta should look like:

```
qgx6@rosalind02:2026-BIFX-TRAINING$ head H1_HA.fasta
>3004125289_N8KIRZ8D_v1 | A_HA_H1
ATGAAGGCAATACTAGTGGTTATGCTGTATACATTTACAACCGCAAATGCAGACACATTATGTATAGGTTATCATGCGAACAATTCAACAGACACTGTGGACACAGTACTAGAAAAGAATGTAACAGTAACACACTCTGTCAATCTTCTAGAAGACAAGCATAACGGAAAACTATGCAAACT
AAGAGGGGTAGCCCCATTGCATTTGGGTCAATGTAACATTGCTGGCTGGATCCTGGGAAATCCAGAGTGTGAATCACTATCCACAGCAAGGTCATGGTCCTACATTGTGGAAACATCTAATTCAGACAATGGAACGTGTTACCCAGGAGATTTCATCAATTATGAGGAGCTAAGAGAGCAAT
TGAGCTCAGTGTCATCATTTGAAAAGTTTGAAATATTCCCCAAGGCAAGTTCATGGCCTAATCATGACTCAGACAATGGTGTAACGGCAGCATGTTCTCACGATGGAGCAAGAAGCTTCTACAAAAACTTGATATGGCTGGTTAAAAAAGGAAAATCGTACCCAAAGATCAACCAAACCTAC
ATTAATGACAAAGGGAAAGAAGTCCTCGTGCTGTGGGGCATTCACCATCCACCCACTATTACTGACCAAGAAAGTCTCTATCAGAATGCAGATGCGTATGTTTTTGTGGGGACATCAAGATACAGCAAGAAGTTCAAGCCGGAAATAGCAGCAAGACCCAAAGTGAGGGATCAAGCAGGGAG
AATGAACTATTACTGGACACTAGTAGAACCGGGAGACAAAATAACATTCGAAGCAACTGGTAATCTAGTGGCACCGAGGTATGCATTCACAATGGAAAAAGAAGCTGGATCTGGTATCATCATTTCAGATACACCAGTCCACGATTGCAATGCAACTTGTCAGACACCCAAGGGTGCTATAA
ACACCAGCCTCCCATTTCAAAATGTACATCCGATCACGATTGGGGAATGTCCAAAGTATGTAAGAAGCACAAAACTGAGGCTGGCCACAGGATTGAGGAATGTCCCGTCTATTCAATCTAGAGGCCTATTCGGGGCCATTGCTGGCTTCATCGAAGGGGGGTGGACAGGAATGGTTGATGGA
TGGTACGGTTATCACCATCAAAATGATCAGGGATCAGGATATGCAGCCGATCTGAAGAGCACACAAAATGCCGTTGATAAGATTACCAACAAAGTAAATTCTGTTATTGAAAAGATGAATACACAGTTCACAGCAGTTGGTAAAGAGTTCAACCACCTTGAAAAAAGAATAGAAAATCTAAA
TAAAAAGGTTGATGATGGTTTCCTGGACGTTTGGACTTACAATGCCGAACTGCTGGTTCTACTGGAAAATGAAAGAACTTTGGACTATCACGATTCAAATGTGAAGAACTTGTATGAAAAAGTAAGACACCAGTTAAAAAACAATGCCAAGGAAACTGGAAACGGCTGCTTTGAATTTTACC
ACAAATGCGACAACACATGCATGGAAAGTGTCAAGAATGGAACTTATGACTACCCAAAATACTCAGAGGAAGCAAAATTGAACAGAGAAAAAATAGATGGAGTAAAGCTGGACTCAACAAGGATTTACCAGATTTTGGCAATCTATTCAACTGCTGCCAGTTCATTGGTACTGGTAGTCTCC
CTGGGGGCAATCAGCTTCTGGATGTGCTCTAATGGGTCTCTACAGTGCAGAATATGTATTTAA
--
>3004125325_N8KIRZDD_v1 | A_HA_H1
ATGAAGGCAATACTAGTGGTTATGCTGTATACATTTACAACCGCAAATGCAGACACATTATGTATAGGTTATCATGCGAACAATTCAACAGACACTGTGGACACAGTACTAGAAAAGAATGTAACAGTAACACACTCTGTCAATCTTCTAGAAGACAAGCATAACGGAAAACTATGCAAACT
AAGAGGGGTAGCCCCACTGCATTTGGGTCAATGTAACATTGCTGGCTGGATCCTGGGAAATCCAGAGTGTGAATCACTATCCACAGCAAGGTCATGGTCCTACATTGTGGAAACATCTAATTCAGACAATGGAACGTGTTACCCAGGAGATTTCATCAATTATGAGGAGCTAAGAGAGCAAT
TGAGCTCAGTGTCATCATTTGAAAAGTTTGAAATATTCCCCAAGGCAAGTTCATGGCCTAATCATGACTCGGACAACGGTGTAACGGCAGCATGTTCTCACGATGGAGCAAGAAGCTTCTACAAAAACTTGATATGGCTGGTTAAAAAAGGAAAATCGTACCCAAAGATCAACCAAACCTAC
ATTAATGACAAAGGGAAAGAAGTCCTCGTGCTGTGGGGCATTCACCATCCACCCACTATTACTGACCAAGAAAGTCTCTATCAGAATGCAGATGCGTATGTTTTTGTGGGGACATCAAGATACAGCAAGAAGTTTAAGCCGGAAATAGCAGCAAGACCCAAAGTGAGGGATCAAGCAGGGAG
AATGAACTATTACTGGACACTAGTAGAACCGGGAGACAAAATAACATTCGAAGCAACTGGTAATCTAGTGGCACCGAGGTATGCATTCACAATGGAAAAAGAAGCTGGATCTGGTATCATCATTTCAGATACACCAGTCCACGATTGCAATGCAACTTGTCAGACACCCAAGGGTGCTATAA
ACACCAGCCTCCCATTTCAAAATGTACATCCGATCACGATTGGGGAATGTCCAAAGTATGTAAGAAGCACAAAACTGAGGCTGGCCACAGGATTGAGGAATGTCCCGTCTATTCAATCTAGAGGCCTATTCGGGGCCATTGCTGGCTTCATCGAAGGGGGGTGGACAGGAATGGTTGATGGA
TGGTACGGTTATCACCATCAAAATGATCAGGGATCAGGATATGCAGCCGATCTGAAGAGCACACAAAATGCCGTTGATAAGATTACCAACAAAGTAAATTCTGTTATTGAAAAGATGAATACACAGTTCACAGCAGTTGGTAAAGAGTTCAACCACCTTGAAAAAAGAATAGAAAATCTAAA
TAAAAAGGTTGATGATGGTTTCCTGGACGTTTGGACTTACAATGCCGAACTGCTGGTTCTACTGGAAAATGAAAGAACTTTGGACTATCACGATTCAAATGTGAAGAACTTGTATGAAAAAGTAAGACACCAGTTAAAAAACAATGCCAAGGAAACTGGAAACGGCTGCTTTGAATTTTACC
ACAAATGCGACAACACATGCATGGAAAGTGTCAAGAATGGAACTTATGACTACCCAAAATACTCAGAGGAAGCAAAATTGAACAGAGAAAAAATAGATGGAGTAAAGCTGGACTCAACAAGGATTTACCAGATTTTGGCAATCTATTCAACTGCTGCCAGTTCATTGGTACTGGTAGTCTCC
CTGGGGGCAATCAGCTTCTGGATGTGCTCTAATGGGTCTCTACAGTGCAGAATATGTATTTAA
--
```

Searching and Text Processing Practical

Now, we will use `sed` to turn our HA_H1.fasta headers into strain names

sed -s "s/find/replace/g”

>3004125441_N8KIRZTI_v1 | A_HA_H1 → >A/Country/ 3004125441_N8KIRZTI_v1 /2026

Intermediate question: can you do this all in one command using multiple sed commands with pipe | ?

Part 1: use sed –s to remove the lines with two dashes from grep output, redirect to HA_H1_1.fasta

Find: -- Replace: nothing

Part 2: Use sed –s to replace “>” with “>A/Country/” in HA_H1_1.fasta, redirect to HA_H1_2.fasta

Since this has a /, you will need to use another character (%, #) for your sed

Searching and Text Processing Practical

Now, we will use `sed` to turn our HA_H1.fasta headers into strain names

sed -s "s/find/replace/g”

>3004125441_N8KIRZTI_v1 | A_HA_H1 → >A/Country/ 3004125441_N8KIRZTI_v1 /2026

Part 3: Use sed –s to replace “ | A_HA_H1” with “/2026” in HA_H1_2.fasta, redirect to HA_H1_3.fasta
Since this has a /, you will need to use another character (%, #) for your sed

Part 4: Renane HA_H1_3.fasta to be HA_H1_final.fasta

Intermediate solution

Now, we will use `sed` to turn our HA_H1.fasta headers into strain names

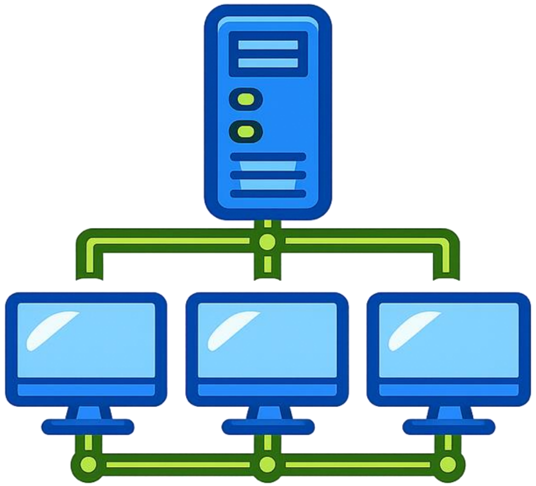
sed -s "s/find/replace/g”

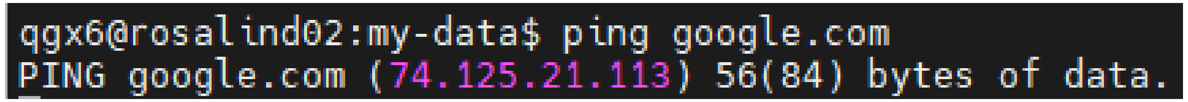
>3004125441_N8KIRZTI_v1 | A_HA_H1 → >A/Country/ 3004125441_N8KIRZTI_v1 /2026

```
sed -s "s/--//g" H1_HA.fasta | sed -s "s#>#>A/Country/#g" | sed -s "s% | A_HA_H1%/2026%g"
> H1_HA_final.fasta
```

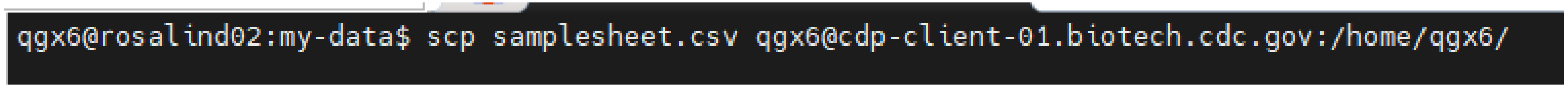
Networking and Downloads part 2

- ping : Check network connection
- ssh : Connect to remote server
- scp : Copy files via SSH from one server to another
- host : DNS lookup utility
- ftp : File transfer protocol client





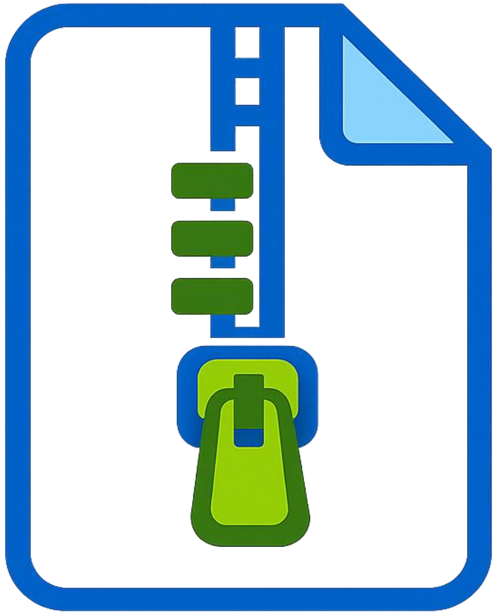
```
qgx6@rosalind02:my-data$ ping google.com
PING google.com (74.125.21.113) 56(84) bytes of data.
```



```
qgx6@rosalind02:my-data$ scp samplesheet.csv qgx6@cdp-client-01.biotech.cdc.gov:/home/qgx6/
```

Compression and archives

- **NGS generates very large files**
  - Raw data (e.g., FASTQ, BAM) can quickly reach GB–TB scale
- **Compressed files are faster to transfer**
  - Reduced size improves download, upload, and sharing speed
- **Best practice for long-term storage**
  - Saves disk space and lowers storage and backup costs



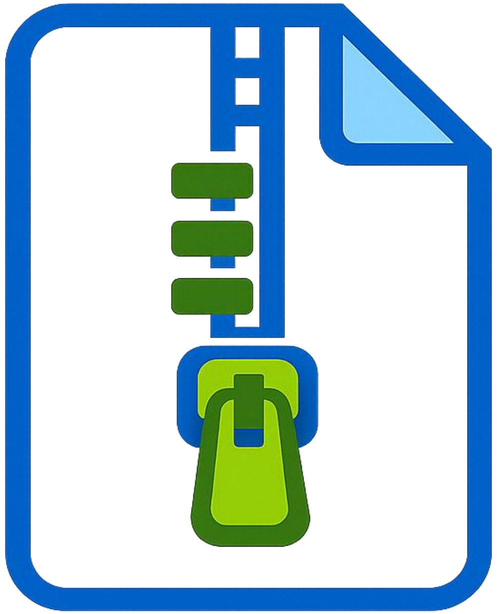
Compression and archives

- **Tarballs (`.tar`)**
  - Bundles multiple files into a single archive (no compression by itself)
- **Gzip (`.gz`)**
  - Common for FASTQ files; fast compression/decompression
- **Zip archives (`.zip`)**
  - Widely supported; combines archiving and compression

Compressed tar archives (`.tar.gz`)

•Archive multiple files *and* compress them in one step

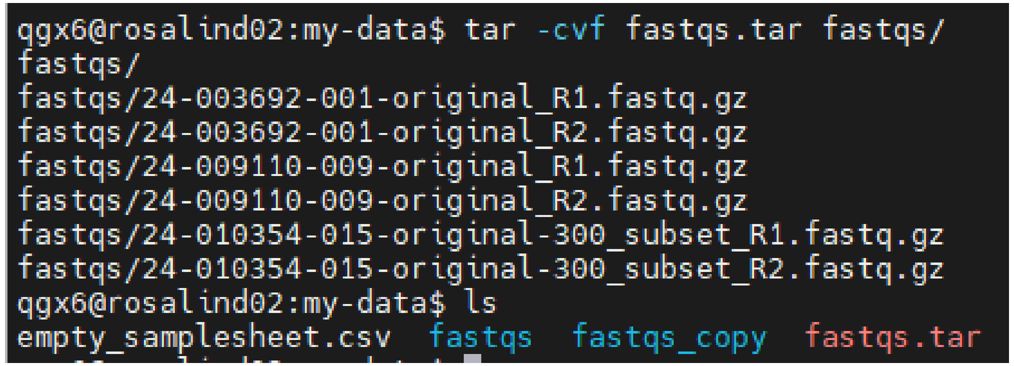
•Common for distributing sequencing datasets



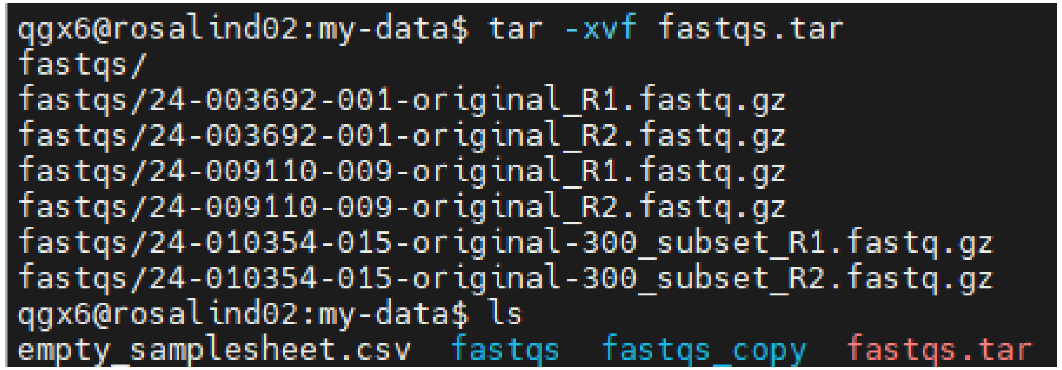
Compression and archives

- **tar –c : Create archives**
- gzip : Compress files
- zip : Create zip archives
- zcat : View compressed files

- **tar –x : Extract archives**
- gunzip : Decompress gzip files
- unzip : Extract zip archives



```
qgx6@rosalind02:my-data$ tar -cvf fastqs.tar fastqs/
fastqs/
fastqs/24-003692-001-original_R1.fastq.gz
fastqs/24-003692-001-original_R2.fastq.gz
fastqs/24-009110-009-original_R1.fastq.gz
fastqs/24-009110-009-original_R2.fastq.gz
fastqs/24-010354-015-original-300_subset_R1.fastq.gz
fastqs/24-010354-015-original-300_subset_R2.fastq.gz
qgx6@rosalind02:my-data$ ls
empty_samplesheet.csv  fastqs  fastqs_copy  fastqs.tar
```

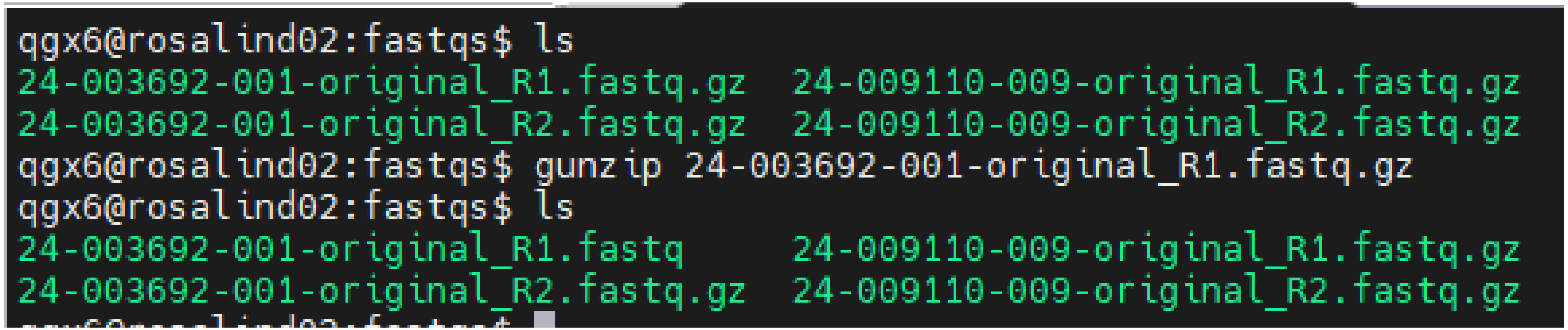


```
qgx6@rosalind02:my-data$ tar -xvf fastqs.tar
fastqs/
fastqs/24-003692-001-original_R1.fastq.gz
fastqs/24-003692-001-original_R2.fastq.gz
fastqs/24-009110-009-original_R1.fastq.gz
fastqs/24-009110-009-original_R2.fastq.gz
fastqs/24-010354-015-original-300_subset_R1.fastq.gz
fastqs/24-010354-015-original-300_subset_R2.fastq.gz
qgx6@rosalind02:my-data$ ls
empty_samplesheet.csv  fastqs  fastqs_copy  fastqs.tar
```

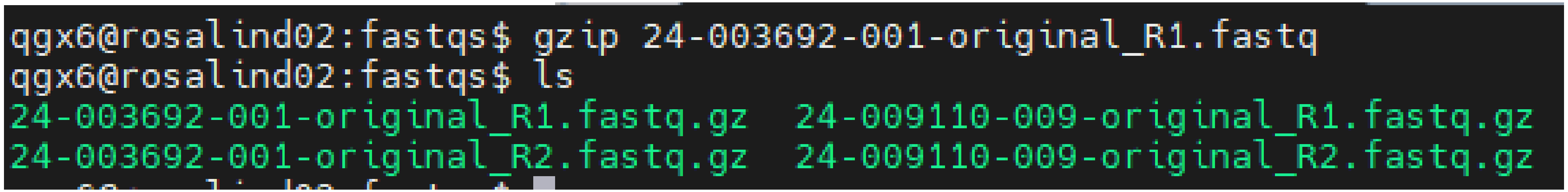
Compression and archives

- tar : Create/extract archives
- **gzip : Compress files (pigz for many many files)**
- zip : Create zip archives
- **zcat : View compressed files**

- tar : Create/extract archives
- **gunzip : Decompress gzip files**
- unzip : Extract zip archives



```
qgx6@rosalind02:fastqs$ ls
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz
qgx6@rosalind02:fastqs$ gunzip 24-003692-001-original_R1.fastq.gz
qgx6@rosalind02:fastqs$ ls
24-003692-001-original_R1.fastq     24-009110-009-original_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz
```

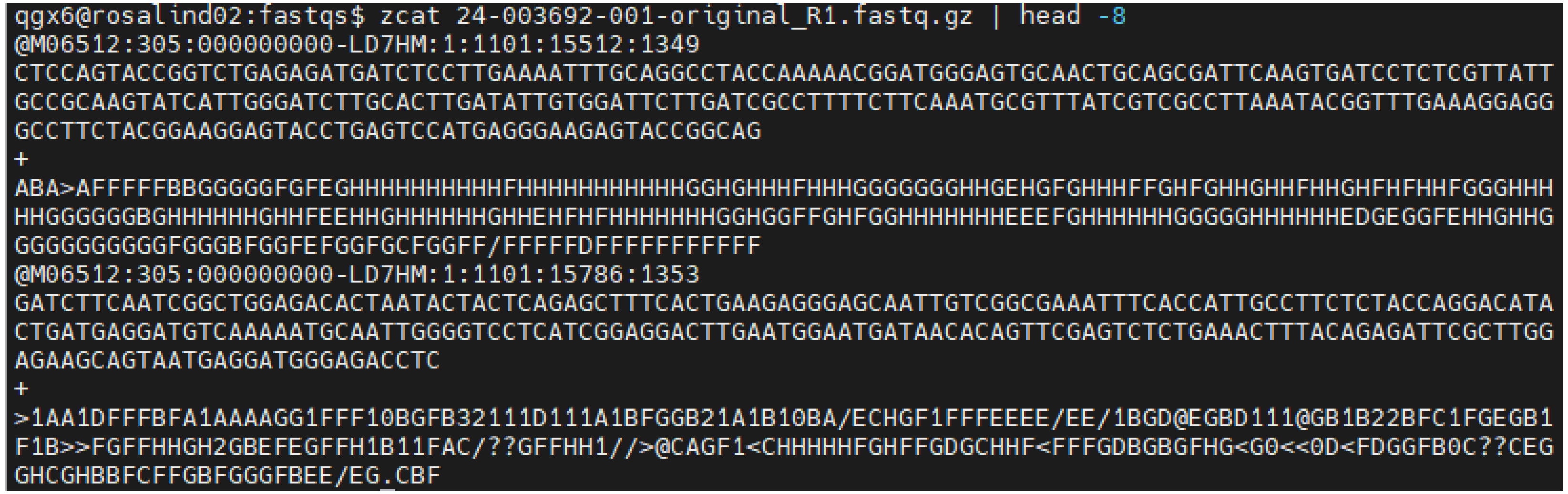


```
qgx6@rosalind02:fastqs$ gzip 24-003692-001-original_R1.fastq
qgx6@rosalind02:fastqs$ ls
24-003692-001-original_R1.fastq.gz  24-009110-009-original_R1.fastq.gz
24-003692-001-original_R2.fastq.gz  24-009110-009-original_R2.fastq.gz
```

Compression and archives

- tar : Create/extract archives
- **gzip : Compress files**
- zip : Create zip archives
- **zcat : View compressed files**

- tar : Create/extract archives
- **gunzip : Decompress gzip files**
- unzip : Extract zip archives

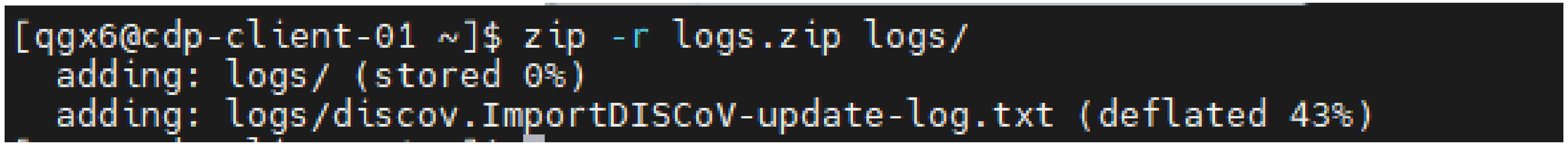


```
qgx6@rosalind02:fastqs$ zcat 24-003692-001-original_R1.fastq.gz | head -8
@M06512:305:000000000-LD7HM:1:1101:15512:1349
CTCCAGTACCGGTCTGAGAGATGATCTCCTTGAAAATTTGCAGGCCTACCAAAAACGGATGGGAGTGCAACTGCAGCGATTCAAGTGATCCTCTCGTTATT
GCCGCAAGTATCATTGGGATCTTGCACTTGATATTGTGGATTCTTGATCGCCTTTTCTTCAAATGCGTTTATCGTCGCCTTAAATACGGTTTGAAAGGAGG
GCCTTCTACGGAAGGAGTACCTGAGTCCATGAGGGAAGAGTACCGGCAG
+
ABA>AFFFFFBBGGGGGFGFEGHHHHHHHHHHHFHHHHHHHHHHHGGHGHHHFHHHGGGGGGGHHGEHGFGHHHFFGHFGHHGHHFHHGHFHFHHFGGGHHH
HHGGGGGGBGHHHHHHGHHFEEHHGHHHHHHGHHEHFHFHHHHHHHGGHGGFFGHFGGHHHHHHHHEEEFGHHHHHHGGGGGHHHHHHEDGEGGFEHHGHHG
GGGGGGGGGGFGGGBFGGFEFGGFGCFGGFF/FFFFFDFFFFFFFFFFF
@M06512:305:000000000-LD7HM:1:1101:15786:1353
GATCTTCAATCGGCTGGAGACACTAATACTACTCAGAGCTTTCACTGAAGAGGGAGCAATTGTCGGCGAAATTTCACCATTGCCTTCTCTACCAGGACATA
CTGATGAGGATGTCAAAAATGCAATTGGGGTCCTCATCGGAGGACTTGAATGGAATGATAACACAGTTCGAGTCTCTGAAACTTTACAGAGATTCGCTTGG
AGAAGCAGTAATGAGGATGGGAGACCTC
+
>1AA1DFFFBFA1AAAAGG1FFF10BGFB32111D111A1BFGGB21A1B10BA/ECHGF1FFFEEEE/EE/1BGD@EGBD111@GB1B22BFC1FGEGB1
F1B>>FGFFHHGH2GBEFEGFFH1B11FAC/??GFFHH1//>@CAGF1<CHHHHHFGHFFGDGCHHF<FFFGDBGBGFHG<G0<<0D<FDGGFB0C??CEG
GHCGHBBFCFFGBFGGGFBEE/EG.CBF
```

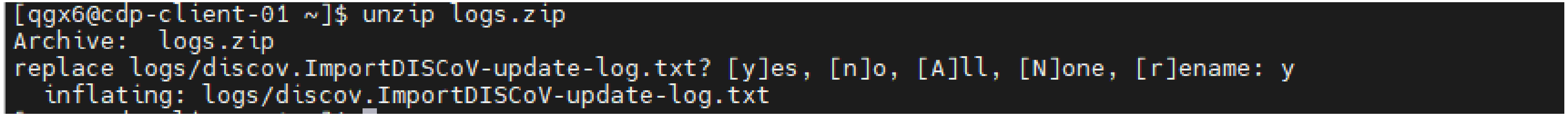
Compression and archives

- tar : Create/extract archives
- gzip : Compress files
- **zip : Create zip archives**
- zcat : View compressed files

- tar : Create/extract archives
- gunzip : Decompress gzip files
- **unzip : Extract zip archives**



```
[qgx6@cdp-client-01 ~]$ zip -r logs.zip logs/
  adding: logs/ (stored 0%)
  adding: logs/discov.ImportDISCoV-update-log.txt (deflated 43%)
```



```
[qgx6@cdp-client-01 ~]$ unzip logs.zip
Archive:  logs.zip
replace logs/discov.ImportDISCoV-update-log.txt? [y]es, [n]o, [A]ll, [N]one, [r]ename: y
  inflating: logs/discov.ImportDISCoV-update-log.txt
```

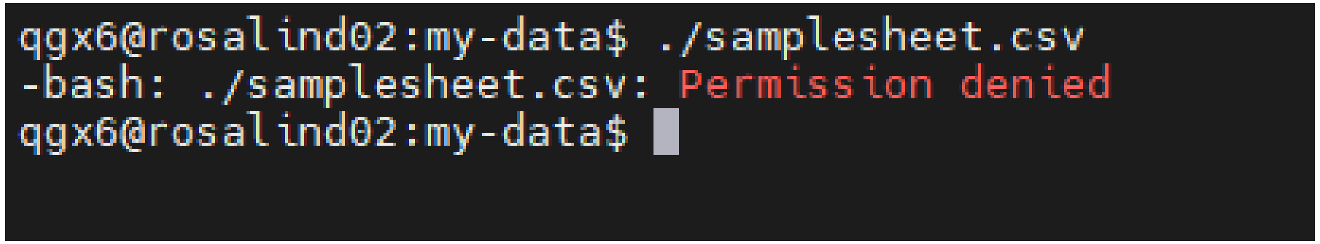
Compression and Archives Practical

wget/curl, uncompress, and re-compress the following files
2012704893_273083_v1_H3_PCR_original_R1_001.fastq.gz
2012704893_273083_v1_H3_PCR_original_R2_001.fastq.gz

Use ls –lah (h makes file sizes **H**uman-readable) to display the file sizes before and after compression

Permissions and Ownership

- **Ownership defines control**
  - Each file has a user (owner) and group assigned to it
- **Permissions control access**
  - Read (r), write (w), and execute (x) permissions
- **Three permission levels**
  - Apply separately to owner, group, and others

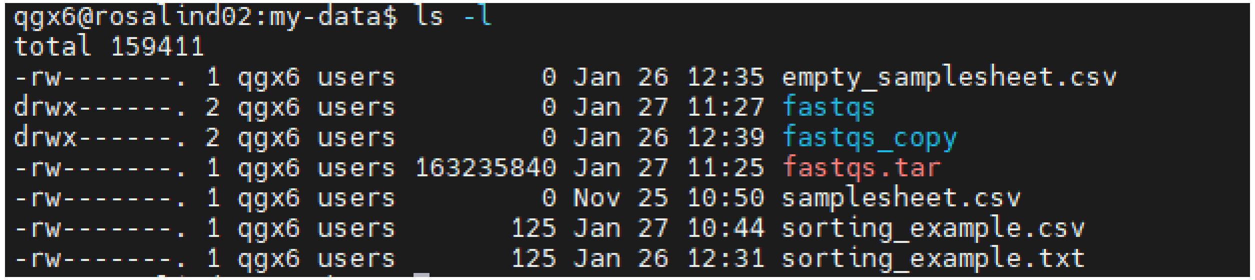


```
qgx6@rosalind02:my-data$ ./samplesheet.csv
-bash: ./samplesheet.csv: Permission denied
qgx6@rosalind02:my-data$ █
```

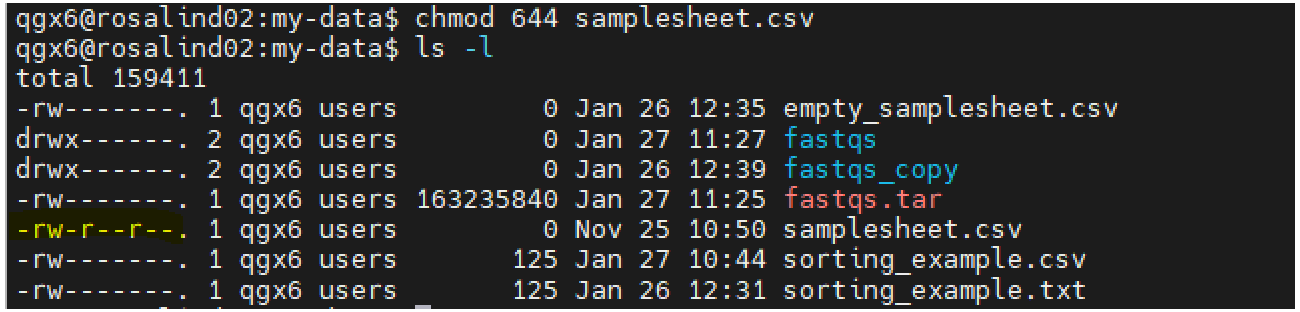
Permissions and Ownership

Permissions are managed with standard commands

- `ls -l` : view
  - Owner, group, others
  - -rwx
  - This shows owner (qgx6) can read and write every file, group (users) and others have no permissions (---)
- `chmod` change permissions
  - Letters or numbers `r = 4` (read) • `w = 2` (write) • `x = 1` (execute)
  - Add values to set permissions
    - `7 = 4+2+1 (rwx)`
    - `6 = 4+2 (rw-)`
    - `5 = 4+1 (r-x)`
    - `-rw-r--r--` → 644
    - Owner: read/write, Group & others: read
- `chown` change ownership



```
qgx6@rosalind02:my-data$ ls -l
total 159411
-rw-------. 1 qgx6 users         0 Jan 26 12:35 empty_samplesheet.csv
drwx------. 2 qgx6 users         0 Jan 27 11:27 fastqs
drwx------. 2 qgx6 users         0 Jan 26 12:39 fastqs_copy
-rw-------. 1 qgx6 users 163235840 Jan 27 11:25 fastqs.tar
-rw-------. 1 qgx6 users         0 Nov 25 10:50 samplesheet.csv
-rw-------. 1 qgx6 users       125 Jan 27 10:44 sorting_example.csv
-rw-------. 1 qgx6 users       125 Jan 26 12:31 sorting_example.txt
```



```
qgx6@rosalind02:my-data$ chmod 644 samplesheet.csv
qgx6@rosalind02:my-data$ ls -l
total 159411
-rw-------. 1 qgx6 users         0 Jan 26 12:35 empty_samplesheet.csv
drwx------. 2 qgx6 users         0 Jan 27 11:27 fastqs
drwx------. 2 qgx6 users         0 Jan 26 12:39 fastqs_copy
-rw-------. 1 qgx6 users 163235840 Jan 27 11:25 fastqs.tar
-rw-r--r--. 1 qgx6 users         0 Nov 25 10:50 samplesheet.csv
-rw-------. 1 qgx6 users       125 Jan 27 10:44 sorting_example.csv
-rw-------. 1 qgx6 users       125 Jan 26 12:31 sorting_example.txt
```

Intermediate Permissions and Ownership

- chgrp : Change group ownership
- **sudo : Run command as superuser**
- umask : Set default file permissions
- id : Display user identity
- whoami : Show current user
- groups : Show group memberships
- passwd : Change user password
- su : Switch user

Permissions and ownership practical

Using wget or curl, pull down permissions.txt at
https://raw.githubusercontent.com/CDCgov/id-bioifx-workshop/refs/heads/main/practical/permissions.txt

1. Cat permissions.txt
2. Using chmod, remove read permissions for owner on permissions.txt
3. Cat permissions.txt – what does it say now?
4. Using chmod again, add read permissions back to permissions.txt

Using the up arrow to load your last wget or curl command, pull down permissions.sh by changing the command's final .txt to .sh

System info & Process management

top : Display running processes (htop)

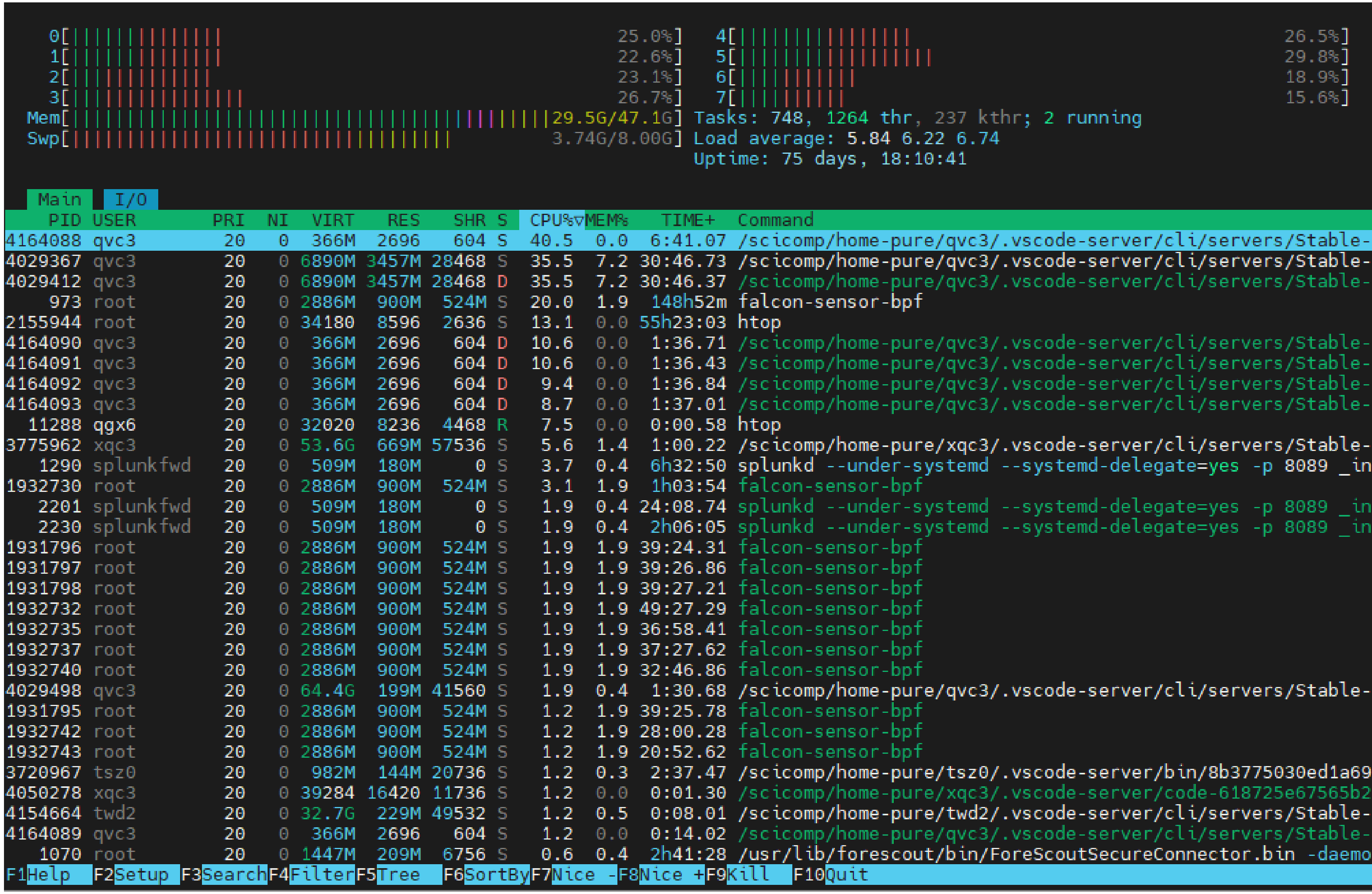
ps : Report current processes

kill : Terminate processes

pkill : Kill by process name

jobs : List active jobs

Ever need to cancel what you just ran? Ctrl + c



```
    0[||||||||||||||                                  25.0%]   4[||||||||||||||||                                  26.5%]
    1[||||||||||||||                                  22.6%]   5[||||||||||||||||||||                              29.8%]
    2[|||||||||||||                                   23.1%]   6[||||||||||||                                      18.9%]
    3[||||||||||||||||                                26.7%]   7[|||||||||||                                       15.6%]
  Mem[||||||||||||||||||||||||||||||||||||||||||||||||||29.5G/47.1G] Tasks: 748, 1264 thr, 237 kthr; 2 running
  Swp[||||||||||||||||||||||||||||||||||||||||||||         3.74G/8.00G] Load average: 5.84 6.22 6.74
                                                                      Uptime: 75 days, 18:10:41

  Main  I/O
    PID USER       PRI  NI  VIRT   RES   SHR S  CPU%▽MEM%   TIME+  Command
4164088 qvc3        20   0  366M  2696   604 S  40.5  0.0  6:41.07 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
4029367 qvc3        20   0 6890M 3457M 28468 S  35.5  7.2 30:46.73 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
4029412 qvc3        20   0 6890M 3457M 28468 D  35.5  7.2 30:46.37 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
    973 root        20   0 2886M  900M  524M S  20.0  1.9  148h52m falcon-sensor-bpf
2155944 root        20   0 34180  8596  2636 S  13.1  0.0 55h23:03 htop
4164090 qvc3        20   0  366M  2696   604 D  10.6  0.0  1:36.71 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
4164091 qvc3        20   0  366M  2696   604 D  10.6  0.0  1:36.43 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
4164092 qvc3        20   0  366M  2696   604 D   9.4  0.0  1:36.84 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
4164093 qvc3        20   0  366M  2696   604 D   8.7  0.0  1:37.01 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
  11288 qgx6        20   0 32020  8236  4468 R   7.5  0.0  0:00.58 htop
3775962 xqc3        20   0 53.6G  669M 57536 S   5.6  1.4  1:00.22 /scicomp/home-pure/xqc3/.vscode-server/cli/servers/Stable-
   1290 splunkfwd   20   0  509M  180M     0 S   3.7  0.4  6h32:50 splunkd --under-systemd --systemd-delegate=yes -p 8089 _in
1932730 root        20   0 2886M  900M  524M S   3.1  1.9  1h03:54 falcon-sensor-bpf
   2201 splunkfwd   20   0  509M  180M     0 S   1.9  0.4 24:08.74 splunkd --under-systemd --systemd-delegate=yes -p 8089 _in
   2230 splunkfwd   20   0  509M  180M     0 S   1.9  0.4  2h06:05 splunkd --under-systemd --systemd-delegate=yes -p 8089 _in
1931796 root        20   0 2886M  900M  524M S   1.9  1.9 39:24.31 falcon-sensor-bpf
1931797 root        20   0 2886M  900M  524M S   1.9  1.9 39:26.86 falcon-sensor-bpf
1931798 root        20   0 2886M  900M  524M S   1.9  1.9 39:27.21 falcon-sensor-bpf
1932732 root        20   0 2886M  900M  524M S   1.9  1.9 49:27.29 falcon-sensor-bpf
1932735 root        20   0 2886M  900M  524M S   1.9  1.9 36:58.41 falcon-sensor-bpf
1932737 root        20   0 2886M  900M  524M S   1.9  1.9 37:27.62 falcon-sensor-bpf
1932740 root        20   0 2886M  900M  524M S   1.9  1.9 32:46.86 falcon-sensor-bpf
4029498 qvc3        20   0 64.4G  199M 41560 S   1.9  0.4  1:30.68 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
1931795 root        20   0 2886M  900M  524M S   1.2  1.9 39:25.78 falcon-sensor-bpf
1932742 root        20   0 2886M  900M  524M S   1.2  1.9 28:00.28 falcon-sensor-bpf
1932743 root        20   0 2886M  900M  524M S   1.2  1.9 20:52.62 falcon-sensor-bpf
3720967 tsz0        20   0  982M  144M 20736 S   1.2  0.3  2:37.47 /scicomp/home-pure/tsz0/.vscode-server/bin/8b3775030ed1a69
4050278 xqc3        20   0 39284 16420 11736 S   1.2  0.0  0:01.30 /scicomp/home-pure/xqc3/.vscode-server/code-618725e67565b2
4154664 twd2        20   0 32.7G  229M 49532 S   1.2  0.5  0:08.01 /scicomp/home-pure/twd2/.vscode-server/cli/servers/Stable-
4164089 qvc3        20   0  366M  2696   604 S   1.2  0.0  0:14.02 /scicomp/home-pure/qvc3/.vscode-server/cli/servers/Stable-
   1070 root        20   0 1447M  209M  6756 S   0.6  0.4  2h41:28 /usr/lib/forescout/bin/ForeScoutSecureConnector.bin -daemo
F1Help  F2Setup F3SearchF4FilterF5Tree  F6SortByF7Nice -F8Nice +F9Kill  F10Quit
```

Questions?