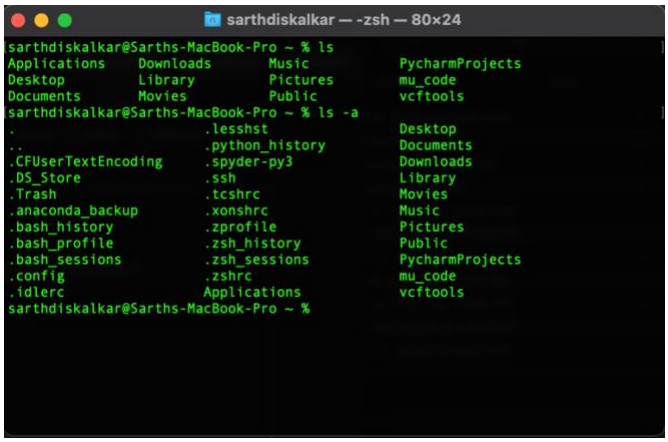
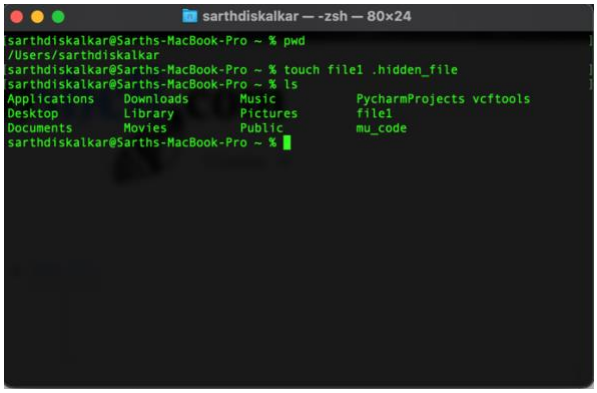
Exercises

1. Using documentation to explore functionality of ls
   1. List the files in your home directory



   2. Create two empty files in your home directory. One named “file1” and one named “.hidden_file” (note the dot in the second name)



   3. What is the size of file1? Show your working

```
sarthdiskalkar@Sarths-MacBook-Pro ~ % ls -lh
total 0
drwx------@  4 sarthdiskalkar  staff   128B Jul 19 18:04 Applications
drwx------+ 23 sarthdiskalkar  staff   736B Aug 28 02:13 Desktop
drwx------+ 31 sarthdiskalkar  staff   992B Aug 27 23:26 Documents
drwx------+ 83 sarthdiskalkar  staff   2.6K Aug 28 02:03 Downloads
drwx------@ 89 sarthdiskalkar  staff   2.8K Aug 21 15:10 Library
drwx------   4 sarthdiskalkar  staff   128B Nov  8  2022 Movies
drwx------+  6 sarthdiskalkar  staff   192B Mar 17 01:50 Music
drwx------+  5 sarthdiskalkar  staff   160B Apr  9 12:16 Pictures
drwxr-xr-x+  4 sarthdiskalkar  staff   128B Nov  8  2022 Public
drwxr-xr-x   5 sarthdiskalkar  staff   160B Aug 23 13:40 PycharmProjects
-rw-r--r--   1 sarthdiskalkar  staff     0B Aug 28 02:24 file1
drwxr-xr-x  10 sarthdiskalkar  staff   320B May 26 23:41 mu_code
drwxr-xr-x  13 sarthdiskalkar  staff   416B Aug 17 21:21 vcftools
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

file1 is 0 bytes.

4. What is the size of the ".hidden_file"?

```
sarthdiskalkar@Sarths-MacBook-Pro ~ % ls -alh
total 144
drwxr-x---+  35 sarthdiskalkar  staff   1.1K Aug 28 02:38 .
drwxr-xr-x    6 root            admin   192B Aug 23 23:37 ..
-r--------    1 sarthdiskalkar  staff     7B Nov  8  2022 .CFUserTextEncoding
-rw-r--r--@   1 sarthdiskalkar  staff    14K Aug 28 02:17 .DS_Store
drwx------+  28 sarthdiskalkar  staff   896B Aug 25 13:21 .Trash
drwxr-xr-x    4 sarthdiskalkar  staff   128B Aug 23 23:33 .anaconda_backup
-rw-------    1 sarthdiskalkar  staff    73B Aug 22 18:48 .bash_history
-rw-r--r--    1 sarthdiskalkar  staff   515B Aug 21 15:10 .bash_profile
drwx------    6 sarthdiskalkar  staff   192B Aug 22 18:48 .bash_sessions
drwxr-xr-x    3 sarthdiskalkar  staff    96B Aug 21 15:10 .config
-rw-r--r--    1 sarthdiskalkar  staff     0B Aug 28 02:24 .hidden_file
drwxr-xr-x    3 sarthdiskalkar  staff    96B Aug 23 23:35 .idlerc
-rw-------    1 sarthdiskalkar  staff    20B Aug 28 02:38 .lesshst
-rw-------    1 sarthdiskalkar  staff   125B Aug 23 23:33 .python_history
drwxr-xr-x   16 sarthdiskalkar  staff   512B Aug 23 23:27 .spyder-py3
drwx------    4 sarthdiskalkar  staff   128B Aug 25 13:15 .ssh
-rw-r--r--    1 sarthdiskalkar  staff   332B Aug 21 15:10 .tcshrc
-rw-r--r--    1 sarthdiskalkar  staff   703B Aug 21 15:10 .xonshrc
-rw-r--r--    1 sarthdiskalkar  staff   166B May  2 14:07 .zprofile
-rw-------    1 sarthdiskalkar  staff    18K Aug 28 01:59 .zsh_history
drwx------  192 sarthdiskalkar  staff   6.0K Aug 28 01:59 .zsh_sessions
-rw-r--r--    1 sarthdiskalkar  staff   514B Aug 21 15:10 .zshrc
drwx------@   4 sarthdiskalkar  staff   128B Jul 19 18:04 Applications
drwx------+  24 sarthdiskalkar  staff   768B Aug 28 02:34 Desktop
drwx------+  31 sarthdiskalkar  staff   992B Aug 27 23:26 Documents
drwx------+  83 sarthdiskalkar  staff   2.6K Aug 28 02:03 Downloads
drwx------@  89 sarthdiskalkar  staff   2.8K Aug 21 15:10 Library
drwx------    4 sarthdiskalkar  staff   128B Nov  8  2022 Movies
drwx------+   6 sarthdiskalkar  staff   192B Mar 17 01:50 Music
drwx------+   5 sarthdiskalkar  staff   160B Apr  9 12:16 Pictures
drwxr-xr-x+   4 sarthdiskalkar  staff   128B Nov  8  2022 Public
drwxr-xr-x    5 sarthdiskalkar  staff   160B Aug 23 13:40 PycharmProjects
-rw-r--r--    1 sarthdiskalkar  staff     0B Aug 28 02:24 file1
drwxr-xr-x   10 sarthdiskalkar  staff   320B May 26 23:41 mu_code
drwxr-xr-x   13 sarthdiskalkar  staff   416B Aug 17 21:21 vcftools
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

.hidden_file is 0 bytes.

5. List all the files in your home directory sorted with oldest first

```
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls -altr
total 144
drwxr-xr-x+   4 sarthdiskalkar  staff    128 Nov  8  2022 Public
-r--------    1 sarthdiskalkar  staff      7 Nov  8  2022 .CFUserTextEncoding
drwx------    4 sarthdiskalkar  staff    128 Nov  8  2022 Movies
drwx------+   6 sarthdiskalkar  staff    192 Mar 17 01:50 Music
drwx------+   5 sarthdiskalkar  staff    160 Apr  9 12:16 Pictures
-rw-r--r--    1 sarthdiskalkar  staff    166 May  2 14:07 .zprofile
drwxr-xr-x   10 sarthdiskalkar  staff    320 May 26 23:41 mu_code
drwx------@   4 sarthdiskalkar  staff    128 Jul 19 18:04 Applications
drwxr-xr-x   13 sarthdiskalkar  staff    416 Aug 17 21:21 vcftools
-rw-r--r--    1 sarthdiskalkar  staff    515 Aug 21 15:10 .bash_profile
-rw-r--r--    1 sarthdiskalkar  staff    514 Aug 21 15:10 .zshrc
drwxr-xr-x    3 sarthdiskalkar  staff     96 Aug 21 15:10 .config
-rw-r--r--    1 sarthdiskalkar  staff    703 Aug 21 15:10 .xonshrc
-rw-r--r--    1 sarthdiskalkar  staff    332 Aug 21 15:10 .tcshrc
drwx------@  89 sarthdiskalkar  staff   2848 Aug 21 15:10 Library
-rw-------    1 sarthdiskalkar  staff     73 Aug 22 18:48 .bash_history
drwx------    6 sarthdiskalkar  staff    192 Aug 22 18:48 .bash_sessions
drwxr-xr-x    5 sarthdiskalkar  staff    160 Aug 23 13:40 PycharmProjects
drwxr-xr-x   16 sarthdiskalkar  staff    512 Aug 23 23:27 .spyder-py3
drwxr-xr-x    4 sarthdiskalkar  staff    128 Aug 23 23:33 .anaconda_backup
-rw-------    1 sarthdiskalkar  staff    125 Aug 23 23:33 .python_history
drwxr-xr-x    3 sarthdiskalkar  staff     96 Aug 23 23:35 .idlerc
drwxr-xr-x    6 root            admin    192 Aug 23 23:37 ..
drwx------    4 sarthdiskalkar  staff    128 Aug 25 13:15 .ssh
drwx------+  28 sarthdiskalkar  staff    896 Aug 25 13:21 .Trash
drwx------+  31 sarthdiskalkar  staff    992 Aug 27 23:26 Documents
-rw-------    1 sarthdiskalkar  staff  18900 Aug 28 01:59 .zsh_history
drwx------  192 sarthdiskalkar  staff   6144 Aug 28 01:59 .zsh_sessions
drwx------+  83 sarthdiskalkar  staff   2656 Aug 28 02:03 Downloads
-rw-r--r--@   1 sarthdiskalkar  staff  14340 Aug 28 02:17 .DS_Store
-rw-r--r--    1 sarthdiskalkar  staff      0 Aug 28 02:24 file1
-rw-r--r--    1 sarthdiskalkar  staff      0 Aug 28 02:24 .hidden_file
drwx------+  24 sarthdiskalkar  staff    768 Aug 28 02:34 Desktop
-rw-------    1 sarthdiskalkar  staff     20 Aug 28 02:38 .lesshst
drwxr-x---+  35 sarthdiskalkar  staff   1120 Aug 28 02:38 .
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

2. Creating and viewing file contents using the terminal
   1. Add two lines of text to “file1”

```
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
 Applications    Downloads       Music           PycharmProjects vcftools
 Desktop         Library         Pictures        file1
 Documents       Movies          Public          mu_code
[sarthdiskalkar@Sarths-MacBook-Pro ~ % echo 'Hello! My name is Sarth.' > file1
[sarthdiskalkar@Sarths-MacBook-Pro ~ % echo 'I am in BIOL 7200.' >> file1
 sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

   2. View the contents of “file1” in your terminal

```
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
 Applications    Downloads       Music           PycharmProjects vcftools
 Desktop         Library         Pictures        file1
 Documents       Movies          Public          mu_code
[sarthdiskalkar@Sarths-MacBook-Pro ~ % echo 'Hello! My name is Sarth.' > file1
[sarthdiskalkar@Sarths-MacBook-Pro ~ % echo 'I am in BIOL 7200.' >> file1
[sarthdiskalkar@Sarths-MacBook-Pro ~ % cat file1
 Hello! My name is Sarth.
 I am in BIOL 7200.
 sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

3. Copying and removing files
   1. Use cp to copy “file1” to “file1_copy.txt”

```
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
Applications    Downloads       Music           PycharmProjects vcftools
Desktop         Library         Pictures        file1
Documents       Movies          Public          mu_code
[sarthdiskalkar@Sarths-MacBook-Pro ~ % cp file1 file1_copy.txt
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
Applications    Downloads       Music           PycharmProjects mu_code
Desktop         Library         Pictures        file1           vcftools
Documents       Movies          Public          file1_copy.txt
```

2. Has the addition of ".txt" to the file name changed how the file contents look?

```
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
Applications    Downloads       Music           PycharmProjects mu_code
Desktop         Library         Pictures        file1           vcftools
Documents       Movies          Public          file1_copy.txt
[sarthdiskalkar@Sarths-MacBook-Pro ~ % cat file1_copy.txt
Hello! My name is Sarth.
I am in BIOL 7200.
[sarthdiskalkar@Sarths-MacBook-Pro ~ % cat file1
Hello! My name is Sarth.
I am in BIOL 7200.
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

The addition of .txt has **not** changed how the file contents look.

Are file extensions significant in Unix systems?

File extensions are not significant to Unix systems. I think the content of the file matters more than the file extension.

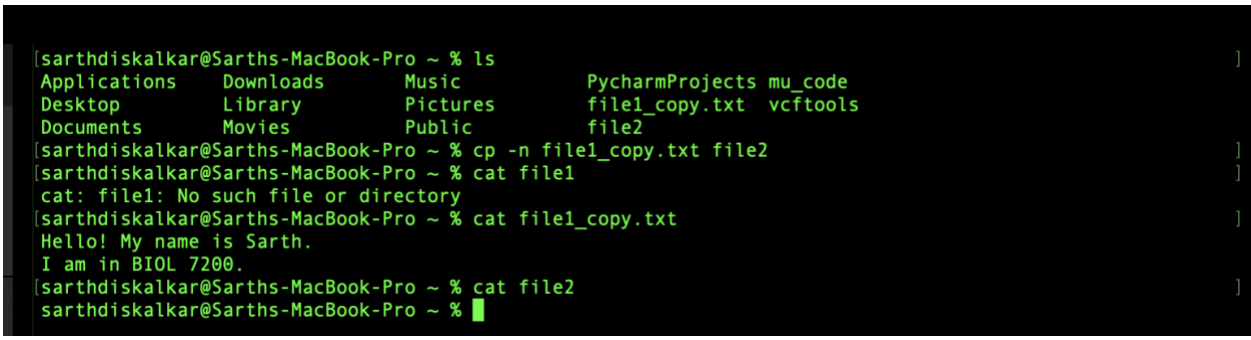
3. Use rm to remove "file1"

```
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
Applications    Downloads       Music           PycharmProjects mu_code
Desktop         Library         Pictures        file1           vcftools
Documents       Movies          Public          file1_copy.txt
[sarthdiskalkar@Sarths-MacBook-Pro ~ % rm file1
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
Applications    Downloads       Music           PycharmProjects vcftools
Desktop         Library         Pictures        file1_copy.txt
Documents       Movies          Public          mu_code
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

4. Create an empty file named "file2"

```
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
Applications    Downloads       Music           PycharmProjects vcftools
Desktop         Library         Pictures        file1_copy.txt
Documents       Movies          Public          mu_code
[sarthdiskalkar@Sarths-MacBook-Pro ~ % touch file2
[sarthdiskalkar@Sarths-MacBook-Pro ~ % ls
Applications    Downloads       Music           PycharmProjects mu_code
Desktop         Library         Pictures        file1_copy.txt  vcftools
Documents       Movies          Public          file2
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

5. Run the command cp -n file1_copy.txt file2.



Does “file2” now contain the same contents as “file1_copy.txt”? Explain.

file2 does not contain the same contents as file1_copy.txt. -n option in the cp command stands for no clobber and does not allow files that already exist to be overwritten by the cp command. Because file2 already exists, the content of file1_copy.txt is not copied into file2.

4. Using documentation to explore useful commands. State the command and options you could use to perform the following tasks:
    1. Create a directory structure “./a/b/c” in a single command (i.e., create a directory and any missing parent directories)

    2. Check if a file has Windows or Unix line endings

You use the file command.

```
[sarthdiskalkar@Sarths-MacBook-Pro ~ % file file1_copy.txt
 file1_copy.txt: ASCII text
[sarthdiskalkar@Sarths-MacBook-Pro ~ % file Applications
 Applications: directory
[sarthdiskalkar@Sarths-MacBook-Pro ~ % file file2
 file2: empty
[sarthdiskalkar@Sarths-MacBook-Pro ~ % file Documents
 Documents: directory
[sarthdiskalkar@Sarths-MacBook-Pro ~ % file --version file1_copy.txt
 file-5.41
 magic file from /usr/share/file/magic
[sarthdiskalkar@Sarths-MacBook-Pro ~ % file file1_copy.txt
 file1_copy.txt: ASCII text
 sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

3. Copy files but only replace existing files if they are older than the source file

Cp -u doesn't work. So used rsync with -u option

```
sarthdiskalkar@Sarths-MacBook-Pro ~ % cp -u file1_copy.txt file2
cp: illegal option -- u
usage: cp [-R [-H | -L | -P]] [-fi | -n] [-aclpsvXx] source_file target_file
       cp [-R [-H | -L | -P]] [-fi | -n] [-aclpsvXx] source_file ... target_directory
sarthdiskalkar@Sarths-MacBook-Pro ~ % rsync -u file1_copy.txt file2
sarthdiskalkar@Sarths-MacBook-Pro ~ % cat file1_copy.txt
Hello! My name is Sarth.
I am in BIOL 7200.
sarthdiskalkar@Sarths-MacBook-Pro ~ % cat file2
sarthdiskalkar@Sarths-MacBook-Pro ~ % rsync -u file2 file1_copy.txt
sarthdiskalkar@Sarths-MacBook-Pro ~ % cat file1_copy.txt
sarthdiskalkar@Sarths-MacBook-Pro ~ % cat file2
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

4. Check if whitespace characters in a file are tabs or spaces

```
sarthdiskalkar@Sarths-MacBook-Pro ~ % od -c file1_copy.txt
0000000    H   e   l   l   o       m   y       n   a   m   e       i   s
0000020        s   a   r   t   h  \n   i   m       h   a   v   i   n   g
0000040        f   u   n       w   i   t   h       t   h   i   s  \n
0000057
sarthdiskalkar@Sarths-MacBook-Pro ~ % echo '\t im tired' >> file1_copy.txt
sarthdiskalkar@Sarths-MacBook-Pro ~ % cat file1_copy.txt
Hello my name is sarth
im having fun with this
        im tired
sarthdiskalkar@Sarths-MacBook-Pro ~ % od -c file1_copy.txt
0000000    H   e   l   l   o       m   y       n   a   m   e       i   s
0000020        s   a   r   t   h  \n   i   m       h   a   v   i   n   g
0000040        f   u   n       w   i   t   h       t   h   i   s  \n  \t
0000060        i   m       t   i   r   e   d  \n
0000072
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

od -c file1_copy.txt

5. View the last 5 commands you issued

```
sarthdiskalkar@Sarths-MacBook-Pro ~ % history -n 5 | tail -n 5
mkdir -p ./a/b/c
file file1_copy.txt
rsync -u file1_copy.txt file2
od -c file1_copy.txt
man history
sarthdiskalkar@Sarths-MacBook-Pro ~ %
```

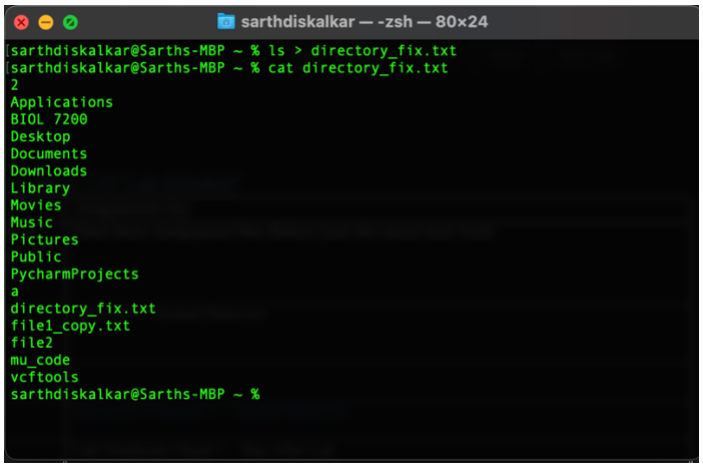
5. Provide a glob or extended glob pattern that would match and not match the sets of filenames in the table below. Give 1 pattern for each row of the table

#	Match these strings	Don't match these strings
1	README.txt, data.tsv, figure.tiff	Homework.pdf, data_to_analyze/, doc.rtf
2	SRR124515, ERR123252, SRR3161371316	PRR161356 LRR124636 error.txt
3	File.txt, another.pdf	temp.csv, data.csv
4	sample_reads_1.fastq, sample_reads_2.fastq, SRR1352235_1.fq, SRR1352235_2.fq	sample_assembly.fasta, SRR1352235_assembly.fasta, sample_feats.bed, SRR1352235_feats.bed, longreads.fastq
5	Samples/a/assembly.fasta, Samples/b/assembly.fasta	assembly.fasta, Samples/assembly.fasta

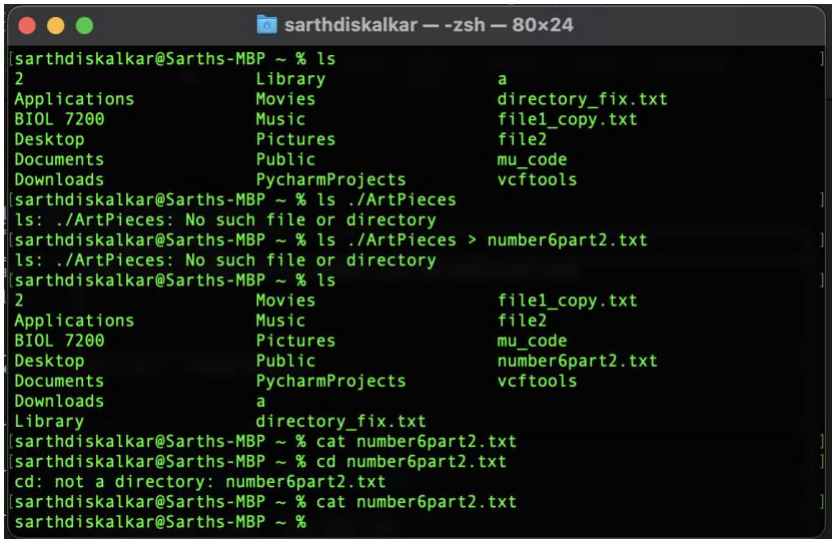
ls at the beginning of each glob pattern to list the files

1. *.{txt,tsv,tiff}
2. @(SRR*|ERR*)
3. {File,another}.{txt,pdf} or !*.csv
4. {sample_reads_[12].fastq,SRR1352235_[12].fq}
5. Samples/[ab]/assembly.fasta

6. Redirecting Outputs
    1. Pick a command that produces stdout, run it, and direct its stdout to a file

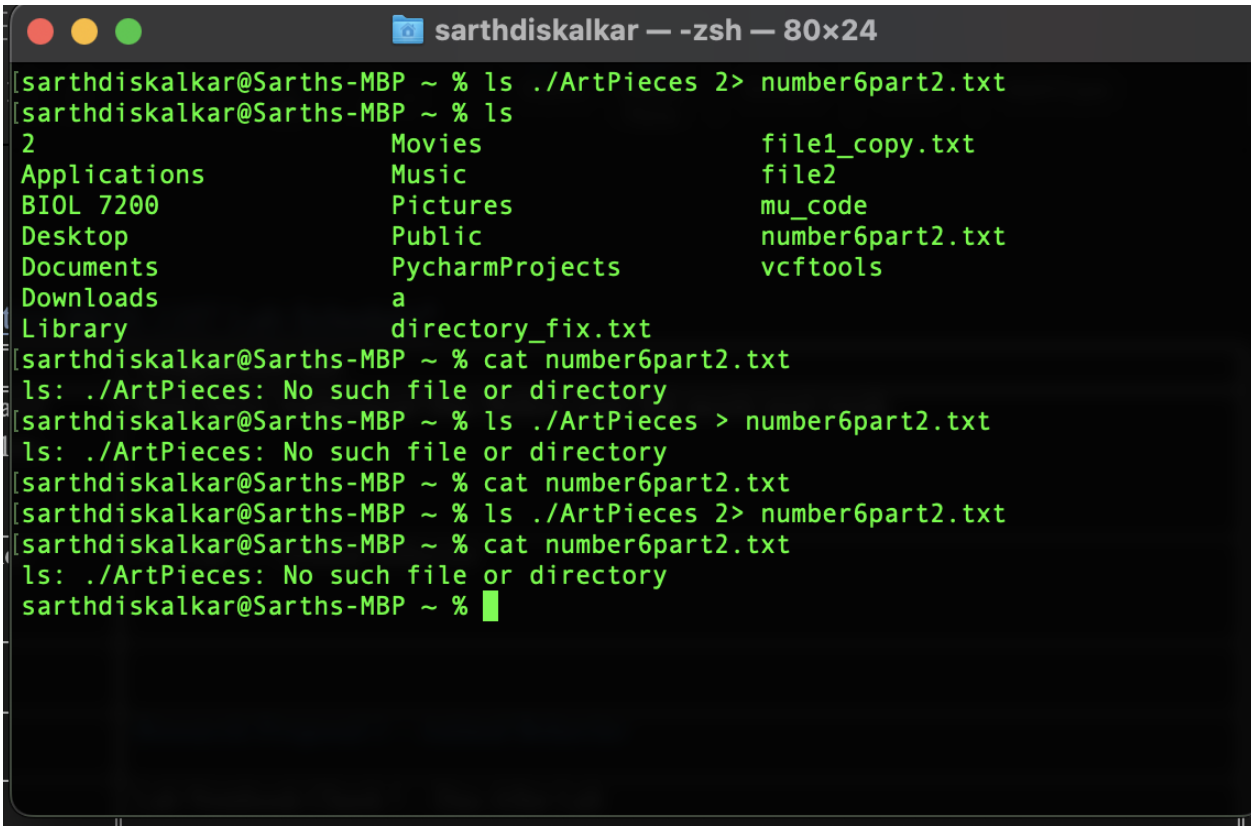


2. Is a path that does not exist in your current directory. Which output stream does the message you see come from?



It comes from the stderr, standard error stream.

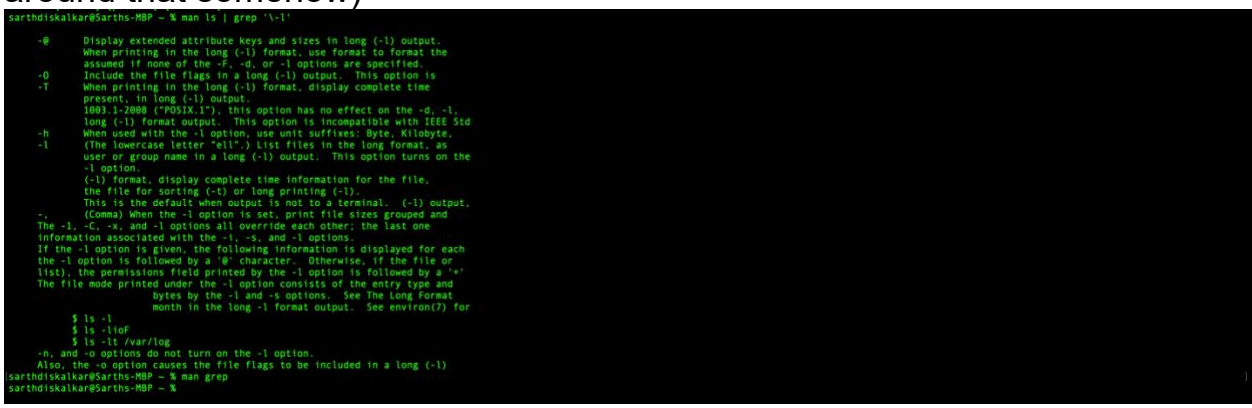
3. Rerun the command from step 2, but now direct the output to a file



4. ls both a non-existent path and "./" (i.e., provide two positional inputs). direct the stdout to one file and the stderr to another file.

5. Use grep to find the help message entry for the -l option of ls (hint: "-" is a special character interpreted by bash so you need to get

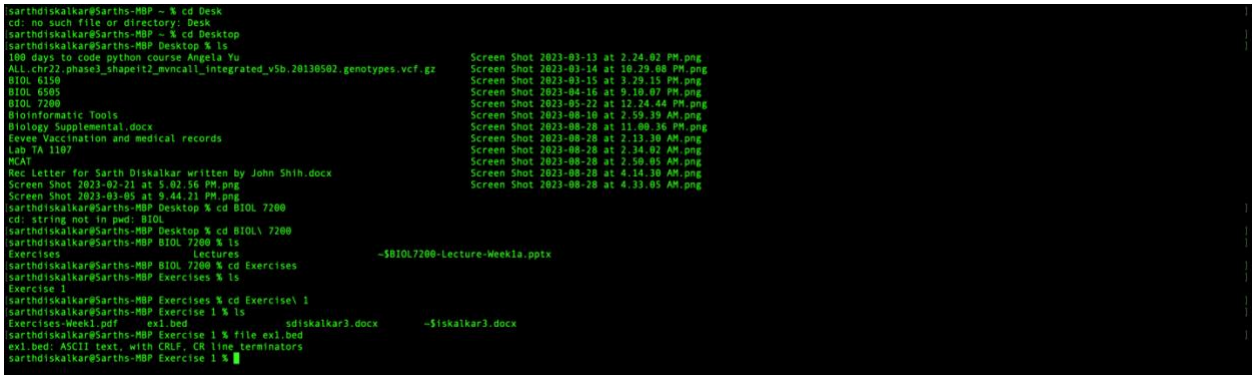
around that somehow)



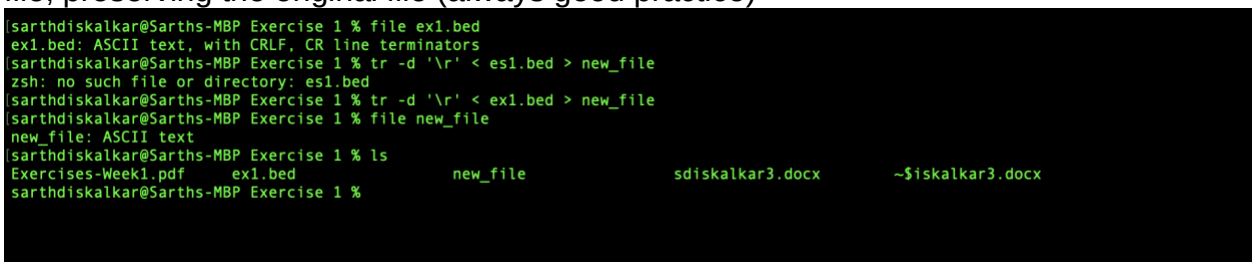
6. How many commands are there in your “/bin” dir?

36 commands

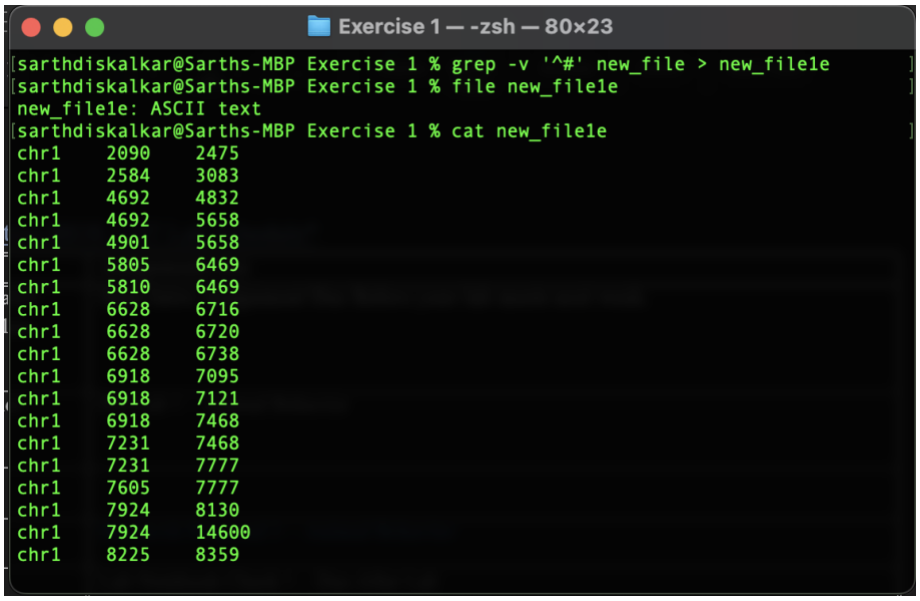
7. Data cleaning. Bioinformaticians often have to work with data generated by others. Perform the following operations to tidy the data in file “ex1.bed” provided on Canvas
    1. Check if the file uses windows line endings instead of unix line endings



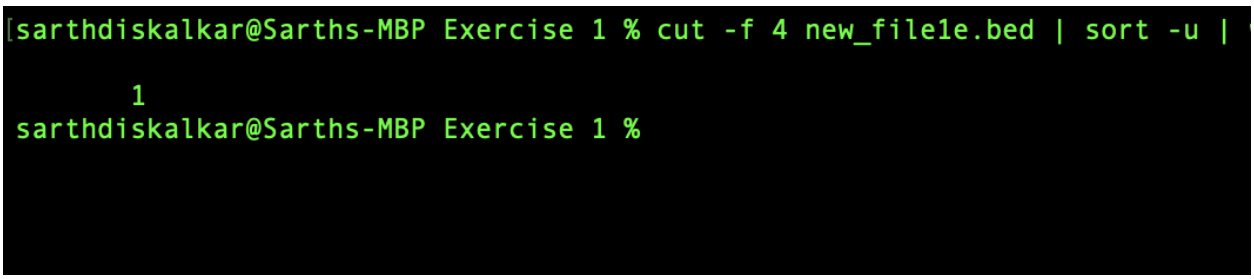
    2. Remove the windows line endings and output the new version to a new file, preserving the original file (always good practice)



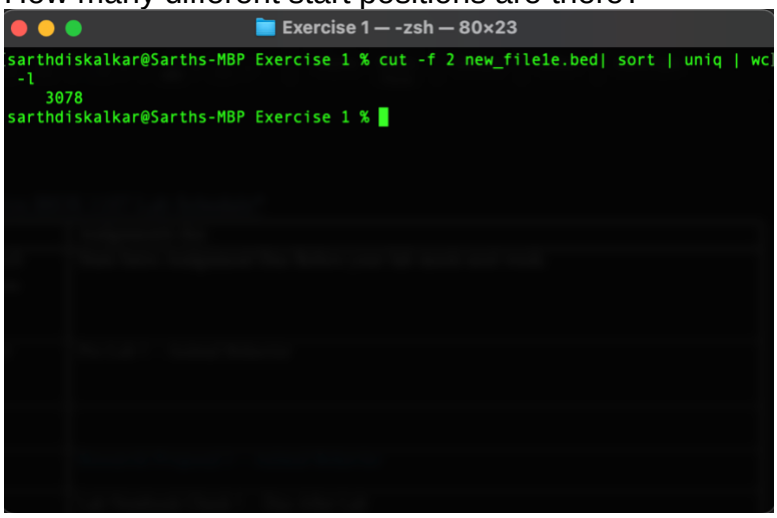
3. Remove the header lines starting with “#” and output the new version to a new file



8. Summarizing real data using bash commands. The following questions relate to the cleaned version of the “ex1.bed” file you generated above. BED format is a commonly used format for storing the location of features in an assembly. The provided file includes the three mandatory columns of a bed file: Sequence ID, start, and stop positions. Using bash commands answer the following commands about these data
    1. How many sequence IDs are present in the file?

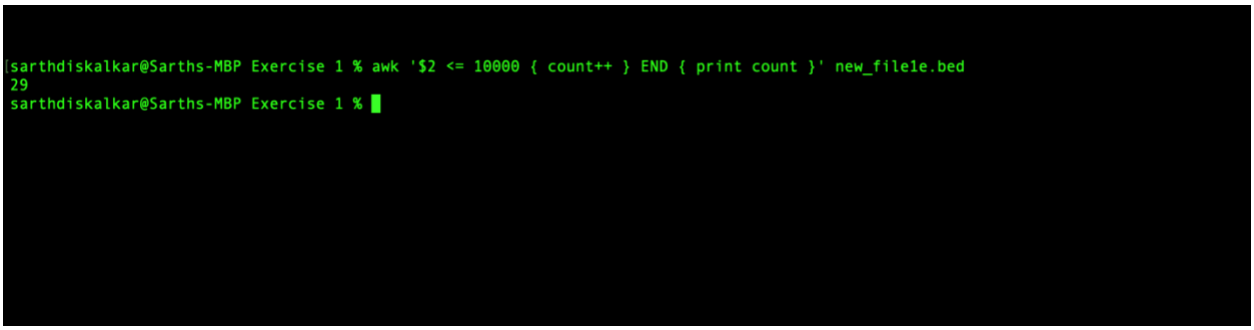


2. How many different start positions are there?

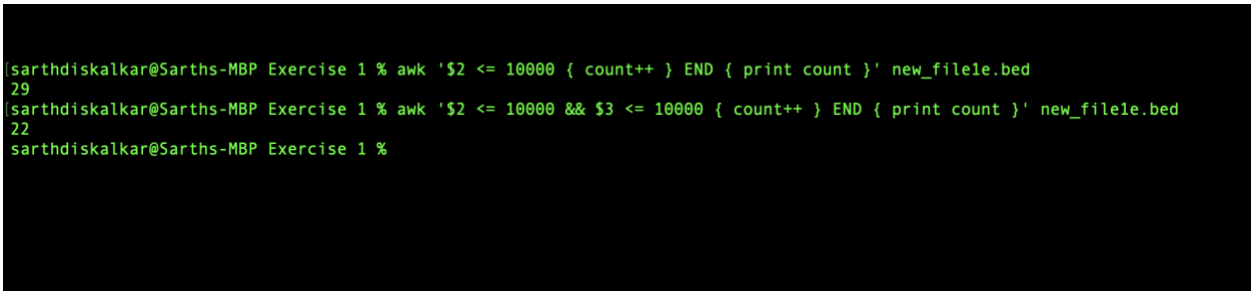


3. What is the highest number of features starting at the same start position?

4. How many features start in the first 10Kb of the sequence?



5. How many features start and end in the first 10Kb of the sequence?



EXTRA CREDIT (5 points)

6. Which feature is the largest? Show your work