

## 8.2 Files – Tutorial

---

At the end of this tutorial you should be able to:

- View files in the command line
  - Output lines at the beginning or end of a file
  - View help pages
  - Redirect output to files
  - Pipe commands
  - Use `cut` and `paste` to rearrange columns into new files
  - Edit files with the `sed` command
- 

### *How to complete this tutorial*

- Go through each question in order and complete any tasks that are described in the question.
  - As you complete the questions, mark your answer to each question.
  - Questions will be either:
    - o multiple-choice questions that require you to provide either a single answer or to select multiple answers
    - o questions that require a short text answer
  - Open the associated quiz on Quercus and enter your answers to each question to verify that you completed the tutorial questions correctly.
  - Alternatively, open the Quercus quiz when you start the tutorial and verify your answers as you complete the tutorial. **Note that there may be some information that is in this file that is not in the Quercus quiz!**
  - The answers will be released at the end of the week.
- 

### *Before you begin:*

- Open a new terminal session from your JupyterHub (New > Terminal)
- Set the PWD to `/home/jovyan/Week.8/8.2.Files/Tutorial.8.2`
- Verify that the directory contains a file called `clinvar_data.csv`

### **Data Sources:**

ClinVar Data was downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/clinvar/>)

## 8.2.1: Viewing Files

---

### Question 1

Which command can you use to output the first 5 lines of `clinvar_data.csv`?

### Question 2

Explore the `less` command. Test out the `less` command by running the command: `less clinvar_data.csv`. After you run the command, press the down arrow key a few times, press the up arrow key a few times. Press the letter Q (this quits the `less` command).

Which of the following describes what the `less` command does?

- a. Reduces the file size so that the file takes up less space
- b. Removes all double letters and replaces it with one letter so that the file takes up less space
- c. Uses all the numbers in the file to calculate a Linear Estimate of the Sum of Squares
- d. Allows you to view and scroll through the contents of a file

### Question 3

Write a command to output only the last line of `clinvar_data.csv`.

What does your output look like?

- a. 16,3757288,glaucoma,CREBBP,Likely\_pathogenic
- b. 7,117479129,cystic.fibrosis,CFTR,Benign
- c. 19,38519292,glaucoma,RYR1,Uncertain
- d. 6,160716688,cystic.fibrosis,PLG,Other

### Question 4 (SELECT ALL THAT APPLY)

Which of the following commands will output the last 5 bytes of `clinvar_data.csv`?

(Hint: check the help page `tail` of the command.)

- a. `tail -n 5 clinvar_data.csv`
- b. `tail --bytes=5 clinvar_data.csv`
- c. `tail --b 5 clinvar_data.csv`
- d. `tail -c 5 clinvar_data.csv`

## 8.2.2: Redirection

---

### Question 5

Execute a command to add the first line of the `clinvar_data.csv` to a file named `clinvar_glaucoma_data.csv`.

What command did you use?

### Question 6

Verify that your new file `clinvar_glaucoma_data.csv` contains only the header line from `clinvar_data.csv`. The last 52 lines of `clinvar_data.csv` contain the glaucoma data. Execute a command to add all the glaucoma data from `clinvar_data.csv` to `clinvar_glaucoma_data.csv` without overwriting the header line that is already in the file.

What command did you use?

### Question 7

Execute a command to output the last 4 bytes in the list of files in the PWD.

What is the result?

- a. pdf
- b. .pdf
- c. `clinvar_glaucoma_data.csv` `clinvar_glaucoma_data.csv`  
`Tutorial.8.2.pdf`
- d. `Tutorial.8.2.pdf`

### 8.2.3: Editing Files

---

#### Question 8

Use `sed` to replace all the commas in `clinvar_glaucoma_data.csv` with tabs (use the escape character for tabs: `\t`). Redirect the output to a file called `clinvar_glaucoma_data.txt`. Use one of the file viewing commands to check that your `sed` command worked.

What command did you use?

- a. `sed 's/,/\t/' clinvar_glaucoma_data.csv > clinvar_glaucoma_data.txt`
- b. `sed 'li \t' clinvar_glaucoma_data.csv > clinvar_glaucoma_data.txt`
- c. `sed ',d' clinvar_glaucoma_data.csv > clinvar_glaucoma_data.txt`
- d. `sed 's/,/\t/g' clinvar_glaucoma_data.csv > clinvar_glaucoma_data.txt`

#### Question 9

Extract columns 1 & 2 (chromosome and position) from `clinvar_glaucoma_data.txt`. and save it to a file called `clinvar_glaucoma_location.txt`.

What command did you use?

#### Question 10

Extract the 5<sup>th</sup> column (clinical.significance) from `clinvar_glaucoma_data.txt` and save it to a file called `clinvar_glaucoma_significance.txt`.

What command did you use?

#### Question 11

Create a new file called `clinvar_glaucoma_significance_and_location.txt` using `clinvar_glaucoma_location.txt` and `clinvar_glaucoma_significance.txt`. The significance should be the first column in the new file.

Fill in the command below to match the command you used (do not include what is already there!).

\_\_\_\_\_ > `clinvar_glaucoma_significance_and_location.txt`

### Question 12

The `cut` command can also be used with the `-c` option to return a set or range of characters. For example, to return the 2<sup>nd</sup> to 10<sup>th</sup> characters on each line of `clinvar_glaucoma_data.csv` we can write:

```
cut -c 2-10 clinvar_glaucoma_data.csv
```

View the help page for the `cut` command. Write a command to output the characters from the 20<sup>th</sup> character to the end of the line, for the first 8 lines of `clinvar_glaucoma_data.csv`. (Use short versions of options.)