Intro to Command Line

March, 2025

Table of Contents

# 1 About this Course

Fluency in programming and data science requires using computer software from the Command Line, a text-based way of controlling the computer. You will go on a guided under-the-hood tour behind the graphical interface we typically use: you will learn how to interact and manipulate files, folders, and software via the Command Line.

Targeted audience: Researchers who want to use scientific software launched from the command line, want to use a high-performance cluster computing environment, or want to use a cloud computing environment.

Commitment: A 1.5 hour workshop.

# 2 Command Line Workshop

## 2.1 What is the Command Line and why should I learn it?

Let’s think about the ways we can interact with a computer: keyboard and mouse, hand gestures on a smartphone, voice commands, AR/VR, etc. Most of these interactions are related to a Graphical User Interface (GUI), which centers on the interaction between a pointer and colorful windows and menus.

However, the way a computer interprets and executes instructions are based on text commands. Even graphical information, such as where the mouse is when it clicked a button, is converted into numbers and characters. That means to be an effective programmer and data scientist, we also need to learn how to interact with our computers in a text-based way. This text-based interaction is called the Command Line Interface (CLI).

Examples where we need to use the Command Line:

* Scalable manipulation of text, files, and folders: if we want to move all files that have the words “tax returns” to a new folder, it would probably not scale easily via a mouse, but it could be done in one command in the Command Line.
* Use of programming languages and scientific software tools often require Command Line knowledge: running Python and R programs, using Git, alignment and variant calling bioinformatics software. Although there are nice graphical user interfaces such as RStudio, Juypter Notebooks, and Galaxy, to have full flexibility of these languages you need to control them from the Command Line.
* Use to high performance computing systems and the cloud all require Command Line knowledge as they do not typically offer GUIs.

The most commonly used Command Line is the Shell. The Shell is a dynamic programming environment, similar to R and Python, but focuses on interacting with the operating system.

## 2.2 Getting Started

In this workshop, we are working on a virtual command line via Posit Cloud. After this workshop, if you want to use the command line locally from your computer, see the appendix.

To access the virtual command line on Posit Cloud, you can access the following project: <https://posit.cloud/content/9696853>

Once you are in, on the left hand side of the screen, click the “Terminal” tab.

You should see some text, ending with a blinking cursor, such as this:

/cloud/project$

This states that your current directory is /cloud/project. This current directory is important: in the Command Line, you interact with the computer from a directory, similar interacting with the computer using a file system window graphically.

Sometimes, the directory path is long so not the full path is printed. To see the path, type in the command pwd and hit enter.

/cloud/project$ pwd  
/cloud/project

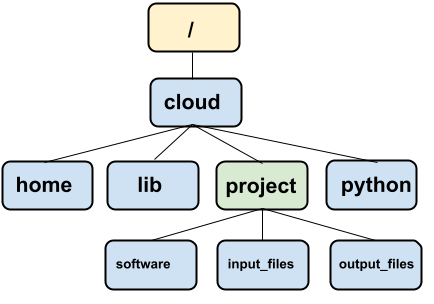
Unlike a GUI, the CLI does not provide immediate options to you to interact with. We have to know a learn a handful of vocabulary to interact with it well. But besides the vocabulary, we need to keep a mental model of a task we want to complete. In GUIs that that mental model is shown to us visually, such as a file browser.

We organizes our seminar by constructing several mental models and learning relevant commands related to each model.

## 2.3 Mental Model 1: Navigating a directory tree

On our computer, the **directory tree** organizes files and directories in an (upside down) tree-like structure. In each folder, there is a parental directory, and there can be files and directories within it. The root directory / does not have a parental directory and contains all directories and files in on the computer.

Below is a typical directory tree schematic, with the root directory being / and home directory being /home/runner. We refer to a particular directory by describing its **directory path** separated by / slashes.



Posit Cloud directory tree

To interact with files and programs in the file system in the GUI world, we moved up and down the directory tree via point and clicking on windows. In the CLI world we can use the following commands:

* pwd prints out our current directory.
* cd / changes directory to the root directory.
* cd /cloud/project/input\_files changes directory to the directory we started with.

So far, we have changed the directory by specifying the **absolute directory path**, which means specifying the directory from the root directory /. Often, we can specify a path by the **relative directory path**, which is a path *relative to our current directory*.

Suppose you want to access the folder input\_files.

* The **absolute directory path** specifies the directory from the root directory: /cloud/project/input\_files
* The **relative directory path** is a path *relative to our current directory*: if we are currently at /cloud/project, the relative path is input\_files .
* The symbol .. specifies the parent directory. If we are currently at /cloud/project/software, the relative path is ../input\_files. You go up one directory, then go down to input\_files.

To find out what files and folders are within your current directory, the ls command will list them for you:

/cloud/project$ ls  
input\_files output\_files project.Rproj software

It is very common to navigate a directory tree via iterations of ls and cd on relative paths.

### 2.3.1 Exercise: explore the /cloud/project folder.

Folders will not have an extension name, but files will, such as .txt or .fastq ones.

To examine a text file completely, use the cat command with the text file name, such as:

/cloud/project/input\_files$ cat metadata.txt  
metadata goes here.

To examine the first few lines of a text file, use the head command with the first argument to be the text file name. To examine the last few lines of a text file, use tail. To scroll through a text file, use less, which opens up a file, and you can use the arrow keys to browse. Then press q to quit.

Hint: to save typing, use the tab button twice to complete filepaths you don’t know existed.

If something looks out of control, hit Ctrl-C to stop a running program.

## 2.4 Mental Model 2: Treat text-based commands as functions

A **function** is a tool that you can use over and over again:

* It can take in inputs
* Does something
* and can give an output.

The commands you have been using, pwd, cd, ls, and cat can be viewed as functions:

* It can take in **input arguments**, **options**
* Does something
* and can give an output

For example, let’s look at the ways ls can be used:

* With an input argument
  + ls /cloud/project
* With options
  + ls -l
* With options and input argument
  + ls -l /cloud/project

### 2.4.1 How do we know what options and arguments to use for a command?

Often, there is a --help or -h option that tells you how to use the software:

ls --help

Online resources: <https://explainshell.com/>

### 2.4.2 Exercise: options for ls

In the project folder, try out a bunch of ways to list files and directories using various options of ls. Some questions to explore:

* Can you display all files? This would include files that start with ., which are considered “hidden”.
* Can you sort by time (last modified)?
* Can you show the long format? What does it show?
* Can you print out the entire project directory tree by using a recursive option?
* Can you show the size of each file?
* How can you use multiple options at once?

## 2.5 File manipulation

Here are some commands that allows you to create, move, copy, and delete files and folders. All of these commands have no output text.

* cp [from] [to] copies a file or folder from the [from] path to a [to] folder.
* mv [from] [to] moves a file or folder from the [from] path to a [to] folder.
* mkdir [folderPath] creates a new folder at the path specified by [folderPath].
* rm [path] deletes a file at [path]. rm -r [folder] deletes a folder and its subcontents. Cannot be undone unless there is a backup system (usually not on personal computers, but available on Fred Hutch’s computing cluster.)

There is a another special directory symbol that come up often in copying and moving. If you want to copy a file to the current directory, specify the working directory via . For example,

cp my\_folder/file.txt .

copies the file to the working directory.

Can you explain what I am doing below?

/cloud/project/input\_files$ mkdir cases  
/cloud/project/input\_files$ mkdir controls  
/cloud/project/input\_files$ cp sample1\_case.fastq cases  
/cloud/project/input\_files$ cp sample5\_control.fastq controls  
/cloud/project/input\_files$ cd control/  
/cloud/project/input\_files$ ls  
sample5\_control.fastq

### 2.5.1 Wildcards to access multiple files

How can I move multiple files at once based on some pattern, such as “case” or “control”?

The \* wildcard represents zero or more characters of any form. Therefore, \*case\* will specify all files that starts with anything, have the word case in the middle, and ends with anything.

/cloud/project/input\_files$ ls -l \*case\*  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample1\_case.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample2\_case.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample3\_case.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample4\_case.fastq

Similarly, if we want anything that ends in fastq,

/cloud/project/input\_files$ ls -l \*fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample1\_case.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample2\_case.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample3\_case.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample4\_case.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample5\_control.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample6\_control.fastq  
-rw-rw---- 1 r2244204 rstudio-user 1934812 Mar 20 23:24 sample7\_control.fastq

The input argument for ls is now a list of files.

We move it:

/cloud/project/input\_files$ cp \*case\* cases/

We can use \* to specify *all* files in a directory. We delete all files in cases (cannot be easily undone!!)

/cloud/project/input\_files$ cd cases/  
/cloud/project/input\_files/cases$ ls  
sample1\_case.fastq sample2\_case.fastq sample3\_case.fastq sample4\_case.fastq  
/cloud/project/input\_files/cases$ rm \*

## 2.6 Applying what we just learned: running bioinformatics software

A popular DNA sequence alignment tool is called “BWA”, and here we present a toy version of it to simulate what it like to use a bioinformatics command line tool. It takes in an unaligned sequence file and a reference genome, and then pretends to align the unaligned sequence to the reference genome.

/cloud/project/input\_files$ cd ../software/  
  
/cloud/project/software$ python aligner.py --help  
usage: aligner.py [-h] --reference REFERENCE --input INPUT  
  
optional arguments:  
 -h, --help show this help message and exit  
 --reference REFERENCE, -r REFERENCE  
 Reference genome file  
 --input INPUT, -i INPUT  
 Input fastq file

You should use a fastq file from the input\_files folder and the reference genome file from the input\_files/reference\_genome folder.

If it works, you will get a wall of text ending with something like:

Aligned sequence 9996 : ATATTAAATTCTAATATCTGGATCCTTTTGTAGTTCATGAGCGTGATGATTGGGTGTTTCACGCATGTGTGTGCAA  
Aligned sequence 9997 : ATTTTCTCTTTTTATTTATTATTTATTGGTTTCATAGAGCTGAGTGGAGTACGAATGTCTTAACTCTGATATCATT  
Aligned sequence 9998 : TGGTCTCCATTCACACCAGCCTCTTGTCACTGCTCTGTTCTTGTCTCTGGCTTAGAGCTACTTCCTGCTATGGTCC  
Aligned sequence 9999 : ATTCTTCCACCTCTGAAAAACCCCTACCCTCACAGACACTCTAGCCCACAGTCCAACCAGTCCCCAGCCCTCTCTG  
Aligned sequence 10000 : ATCTGAGGGGACGAGAGGGTAAGATGATTGATGGAGGGGAAATCCACAGAGCCTCAGGCACCAAATATGTAGCAAA

Nice job!

## 2.7 Redirects and Pipes

The python aligner.py command seems pretty messy - it just dumps an aligned sequencing file in your command line console - what do you do with it? If you look at its manual, there is no option or argument to specify an output file name, which is a bit not user-friendly at first.

However, in the CLI world, this is a perfectly normal way to output a result from a program. We simply need to **redirect** that output to a file ourselves. The output of ./bwa mem and the output of ls can be redirected to a file, via the > operation:

python aligner.py --reference [reference genome fasta file] --input [unaligned sequences fastq file] > [output file]

Remember when we looked at the output of the ls command, an emphasis is placed that the files and directories listed are considered the output of the program. We can redirect that output into a file too:

ls > ls\_output.txt

Another reason why in the CLI world the output is often dumped to the console is that it gives the user an option to **pipe** it to the next program that takes the first program’s output as input, using the symbol |. This allows the user to chain together several programs together, each performing a modular task. For instance, if we want to align the sequences, and then look at the first few aligned sequences, we can pipe the output of bwa mem with tail:

python aligner.py --reference [reference genome fasta file] --input [unaligned sequences fastq file] | head

and then save it:

python aligner.py --reference [reference genome fasta file] --input [unaligned sequences fastq file] | head > [output file]

This encourages the development of modular, flexible programs that can be connected together via pipes and saved via redirecting.

## 2.8 Appendix: Starting the command line locally on your computer:

### 2.8.1 For Mac users

Open up “Terminal” Application in Applications/Utilities.

### 2.8.2 For Windows users

* Open Windows Powershell.
* Type in: wsl --install -d ubuntu, and hit enter.
* You will might be asked to enter a new username and password. You can use the same as you have for your computer.
* A shell terminal should show up. If it doesn’t show up, look in your search bar for “Ubuntu on Windows”, and open it.

## 2.9 Appendix: Special symbols for directories

* . the current directory.
* .. the parent directory.
* / the root directory.
* ~ the home directory.

## 2.10 Appendix: Using a text editor in CLI

A commonly used task in CLI is to edit text files. nano, vim, and emacs are the three most popular ones, in increasing learning curve but also complexity in the tasks you can perform. The easiest one to use, nano is not installed by default in Posit Cloud. If it was installed on your computer, here is how you would use it:

Pick a file of interest, and run nano [filename]. You will see a new screen of the filename’s contents. You can move around via the letter keys and make edits as needed. On the bottom of the screen are commands you use to manage the file, such as saving, opening, and quitting. The ^ symbol refers to the Control key on your keyboard. To quit, hit Ctrl-x, and you may have to hit y to confirm. You should get comfortable using a text editor in CLI as it is a common task.

## 2.11 Appendix: Why do we run some programs using ./?

Sometimes, you run a piece of software within a folder via ./my\_program. Why is that?

When you type in a command, such as ls in the command line, the command line looks to see whether the command belongs to a list of commands it knows how to run. This list of approved commands is stored in an environmental variable. We can see it via:

andrew@MGQQR2YQRT9 ~ % echo $PATH  
/opt/homebrew/bin:/opt/homebrew/sbin:/usr/local/bin

In each folder (using : to distinguish folders apart) are the programs to run commands such as ls. If we want to run a program that is not in the $PATH environmental variable, there could be security issues. For instance, perhaps someone created a program called ls that does something bad to your computer, and you run it, thinking that you are running the program ls you are familiar with located in the $PATH environment variable. Or someone creating a malicious program called sl that takes advantages of typos.

To protect that, to run programs not in $PATH, we use: ./program\_name. This is a short-form of referring to the current directory ./ and running program\_name.

For more info, see this [post](https://unix.stackexchange.com/questions/4430/why-do-we-use-dot-slash-to-execute-a-file-in-linux-unix).

# About the Authors

These credits are based on our [course contributors table guidelines](https://www.ottrproject.org/more_features.html#giving-credits-to-contributors).

| Credits | Names |
| --- | --- |
| **Pedagogy** |  |
| Lead Content Instructor(s) | Chris Lo |
| Lecturer | Chris Lo |
| Content Author(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | If any other authors besides lead instructor |
| Content Contributor(s) (include section name/link in parentheses) - make new line if more than one section involved | Wrote less than a chapter |
| Content Editor(s)/Reviewer(s) | Checked your content |
| Content Director(s) | Helped guide the content direction |
| Content Consultants (include chapter name/link in parentheses or word “General”) - make new line if more than one chapter involved | Gave high level advice on content |
| Acknowledgments | Gave small assistance to content but not to the level of consulting |
| **Production** |  |
| Content Publisher(s) | Helped with publishing platform |
| Content Publishing Reviewer(s) | Reviewed overall content and aesthetics on publishing platform |
| **Technical** |  |
| Course Publishing Engineer(s) | Helped with the code for the technical aspects related to the specific course generation |
| Template Publishing Engineers | [Candace Savonen](https://www.cansavvy.com/), [Carrie Wright](https://carriewright11.github.io/), [Ava Hoffman](https://www.avahoffman.com/) |
| Publishing Maintenance Engineer | [Candace Savonen](https://www.cansavvy.com/) |
| Technical Publishing Stylists | [Carrie Wright](https://carriewright11.github.io/), [Ava Hoffman](https://www.avahoffman.com/), [Candace Savonen](https://www.cansavvy.com/) |
| Package Developers ([ottrpal](https://github.com/jhudsl/ottrpal)) [Candace Savonen](https://www.cansavvy.com/), [John Muschelli](https://johnmuschelli.com/), [Carrie Wright](https://carriewright11.github.io/) |  |
| **Art and Design** |  |
| Illustrator(s) | Created graphics for the course |
| Figure Artist(s) | Created figures/plots for course |
| Videographer(s) | Filmed videos |
| Videography Editor(s) | Edited film |
| Audiographer(s) | Recorded audio |
| Audiography Editor(s) | Edited audio recordings |
| **Funding** |  |
| Funder(s) | Institution/individual who funded course including grant number |
| Funding Staff | Staff members who help with funding |

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value  
## version R version 4.3.2 (2023-10-31)  
## os Ubuntu 22.04.4 LTS  
## system x86\_64, linux-gnu  
## ui X11  
## language (EN)  
## collate en\_US.UTF-8  
## ctype en\_US.UTF-8  
## tz Etc/UTC  
## date 2025-03-20  
## pandoc 3.1.1 @ /usr/local/bin/ (via rmarkdown)  
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date (UTC) lib source  
## bookdown 0.41 2024-10-16 [1] CRAN (R 4.3.2)  
## cachem 1.0.8 2023-05-01 [1] RSPM (R 4.3.0)  
## cli 3.6.2 2023-12-11 [1] RSPM (R 4.3.0)  
## devtools 2.4.5 2022-10-11 [1] RSPM (R 4.3.0)  
## digest 0.6.34 2024-01-11 [1] RSPM (R 4.3.0)  
## ellipsis 0.3.2 2021-04-29 [1] RSPM (R 4.3.0)  
## evaluate 0.23 2023-11-01 [1] RSPM (R 4.3.0)  
## fastmap 1.1.1 2023-02-24 [1] RSPM (R 4.3.0)  
## fs 1.6.3 2023-07-20 [1] RSPM (R 4.3.0)  
## glue 1.7.0 2024-01-09 [1] RSPM (R 4.3.0)  
## htmltools 0.5.7 2023-11-03 [1] RSPM (R 4.3.0)  
## htmlwidgets 1.6.4 2023-12-06 [1] RSPM (R 4.3.0)  
## httpuv 1.6.14 2024-01-26 [1] RSPM (R 4.3.0)  
## knitr 1.48 2024-07-07 [1] CRAN (R 4.3.2)  
## later 1.3.2 2023-12-06 [1] RSPM (R 4.3.0)  
## lifecycle 1.0.4 2023-11-07 [1] RSPM (R 4.3.0)  
## magrittr 2.0.3 2022-03-30 [1] RSPM (R 4.3.0)  
## memoise 2.0.1 2021-11-26 [1] RSPM (R 4.3.0)  
## mime 0.12 2021-09-28 [1] RSPM (R 4.3.0)  
## miniUI 0.1.1.1 2018-05-18 [1] RSPM (R 4.3.0)  
## pkgbuild 1.4.3 2023-12-10 [1] RSPM (R 4.3.0)  
## pkgload 1.3.4 2024-01-16 [1] RSPM (R 4.3.0)  
## profvis 0.3.8 2023-05-02 [1] RSPM (R 4.3.0)  
## promises 1.2.1 2023-08-10 [1] RSPM (R 4.3.0)  
## purrr 1.0.2 2023-08-10 [1] RSPM (R 4.3.0)  
## R6 2.5.1 2021-08-19 [1] RSPM (R 4.3.0)  
## Rcpp 1.0.12 2024-01-09 [1] RSPM (R 4.3.0)  
## remotes 2.4.2.1 2023-07-18 [1] RSPM (R 4.3.0)  
## rlang 1.1.4 2024-06-04 [1] CRAN (R 4.3.2)  
## rmarkdown 2.25 2023-09-18 [1] RSPM (R 4.3.0)  
## sessioninfo 1.2.2 2021-12-06 [1] RSPM (R 4.3.0)  
## shiny 1.8.0 2023-11-17 [1] RSPM (R 4.3.0)  
## stringi 1.8.3 2023-12-11 [1] RSPM (R 4.3.0)  
## stringr 1.5.1 2023-11-14 [1] RSPM (R 4.3.0)  
## urlchecker 1.0.1 2021-11-30 [1] RSPM (R 4.3.0)  
## usethis 2.2.3 2024-02-19 [1] RSPM (R 4.3.0)  
## vctrs 0.6.5 2023-12-01 [1] RSPM (R 4.3.0)  
## xfun 0.48 2024-10-03 [1] CRAN (R 4.3.2)  
## xtable 1.8-4 2019-04-21 [1] RSPM (R 4.3.0)  
## yaml 2.3.8 2023-12-11 [1] RSPM (R 4.3.0)  
##   
## [1] /usr/local/lib/R/site-library  
## [2] /usr/local/lib/R/library  
##   
## ──────────────────────────────────────────────────────────────────────────────

# 3 References