GINGER On-Site Training
Day 1: Intro to Linux

GINGER Program 2022

University of Cape Town

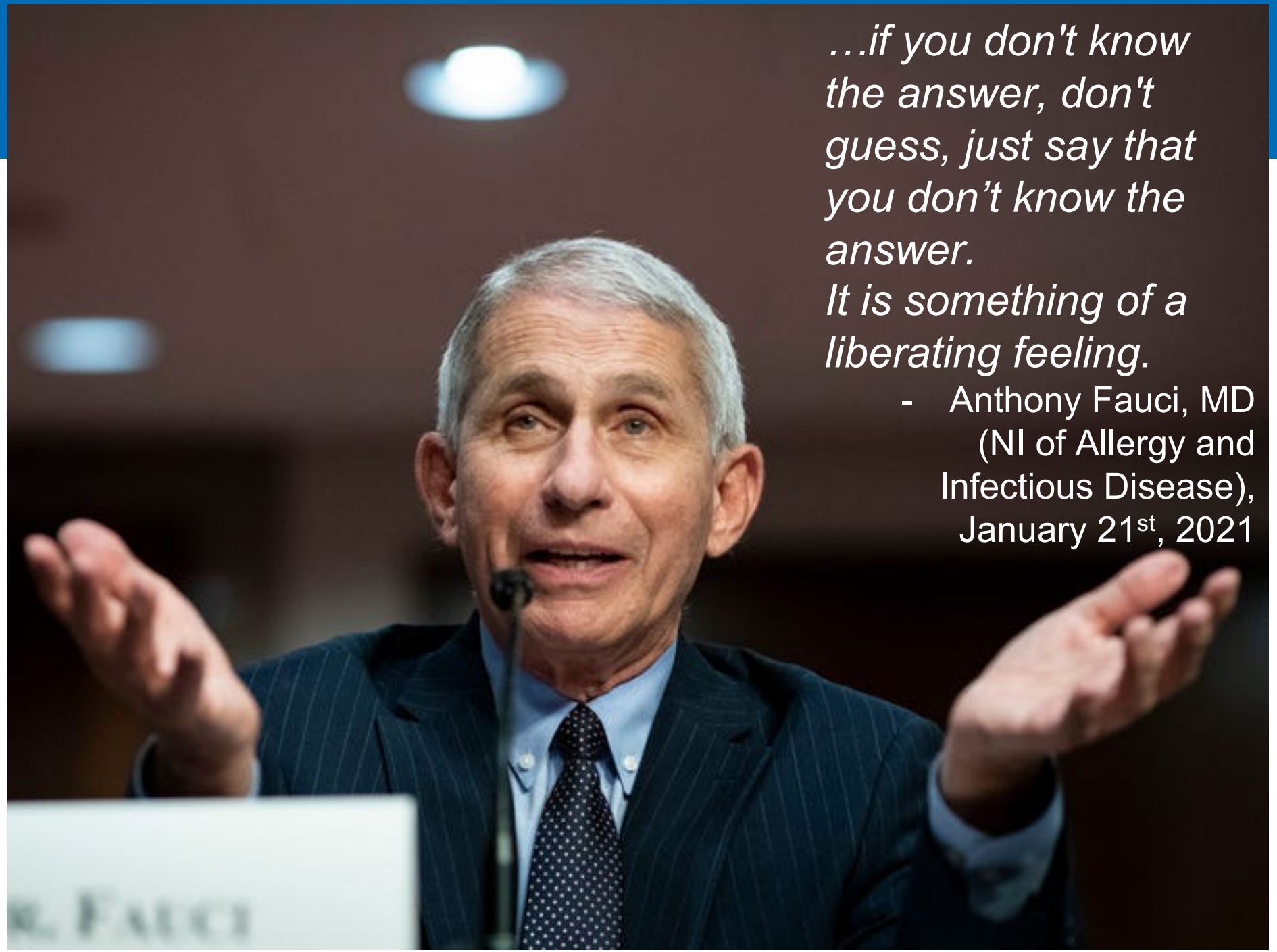
Teaching Fellows:

Kumar Veerapen, Ph.D.
Senior Expert I Data Science
Kumar.Veerapen@Novartis.com

Carla Marquez-Luna, Ph.D.
Postdoctoral Research Fellow
carlamarquezluna@gmail.com

	Monday, April 4	Tuesday, April 5	Wednesday, April 6	Thursday, April 7	Friday, April 8
9:00-10:30	**Training Welcome and Introduction** **9:00-9:30 - Training Overview** **9:30-10:00 - Professor Dan Stein Welcome** **10:00-11:00 - Begin Kampala Refresher**	**Plink Tutorial**	**Excursion to Robben Island (weather dependent)**	**9:00-10:00 am TBD** **10:00 am NeuroGAP Site Visit**	**Step-by-Step GWAS**
10:30-10:45	Tea Break	Tea Break			Tea Break
10:45-1:00	**Kampala Refresher continued**	**11:00-12:00 - Professor Colett Dandara** **12:00-1:00 - Intro to Plink**			**11:00 - Guest Lectures: Drs. Shareefa Dalvie and Nastassja Koen** **12:00-1:00 - Step-by-Step GWAS**
1:00-2:00	Lunch	Lunch	Lunch at the V&A Waterfront (weather dependent)	Lunch	Lunch
2:00-3:30	**Intro to UNIX** Fundamental Commands Genetic Data Formats and Conversion	**Plink Tutorial**	**Group Project Work**	**Step-by-Step GWAS**	**Step-by-Step GWAS**
3:30-3:45	Tea Break	Tea Break		Tea Break	Tea Break
3:45-5:00	**GINGER group projects intro**	**Plink Tutorial** **Step-by-Step GWAS**		**Step-by-Step GWAS**	**4:00-4:30 - Step by Step GWAS** **4:30-5:00 - Group Project Presentations**

GWAS all the time*


…if you don't know the answer, don't guess, just say that you don’t know the answer.
It is something of a liberating feeling.
- Anthony Fauci, MD (NI of Allergy and Infectious Disease), January 21st, 2021

Outline

- What is Linux?
- Why Linux?
- Exercises

Adapted from Konrad Karczewski

BROAD INSTITUTE

Word of Advice

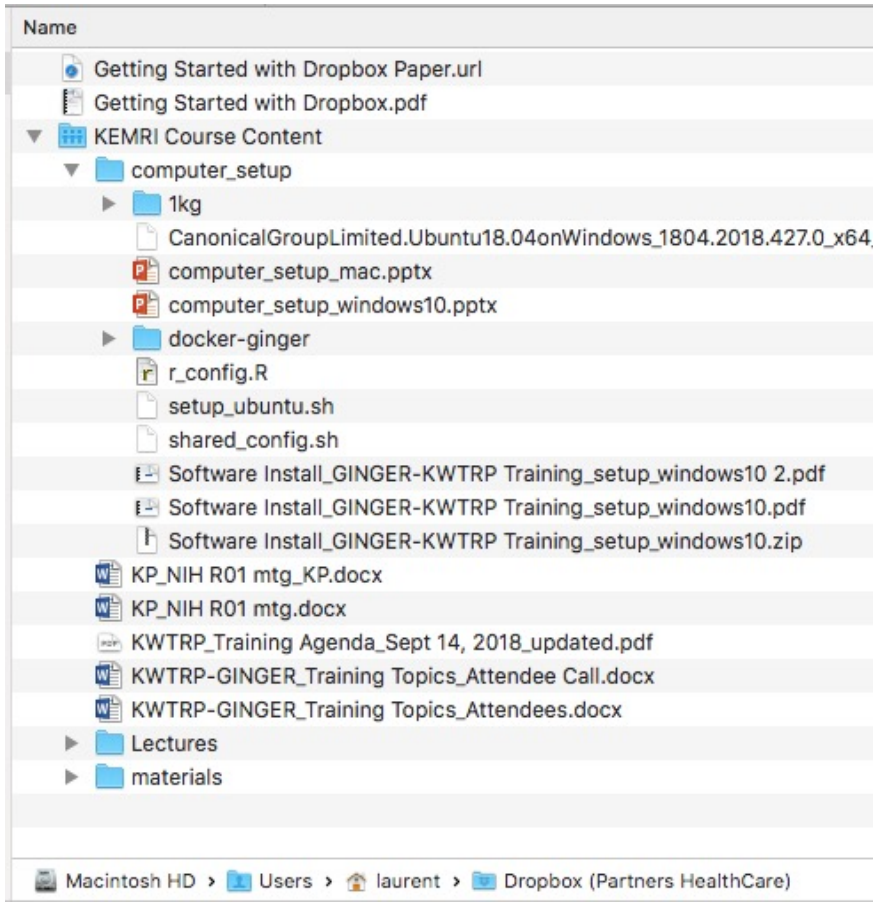
How do you code?

- There are many ways to code.
- Beginners, don't worry about how short your code is. As long as IT WORKS!
- Later, get someone who is more experienced to do code review with you.

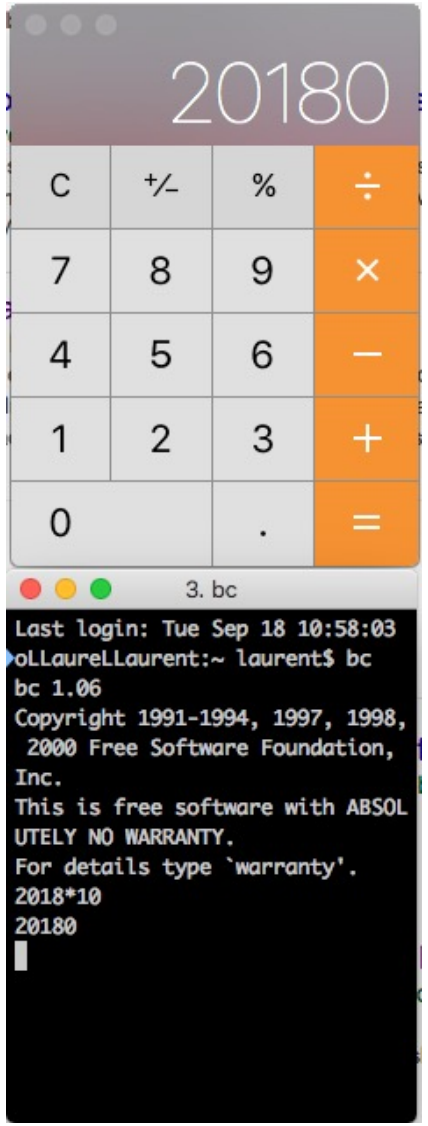
Computing basics

Programs

Data



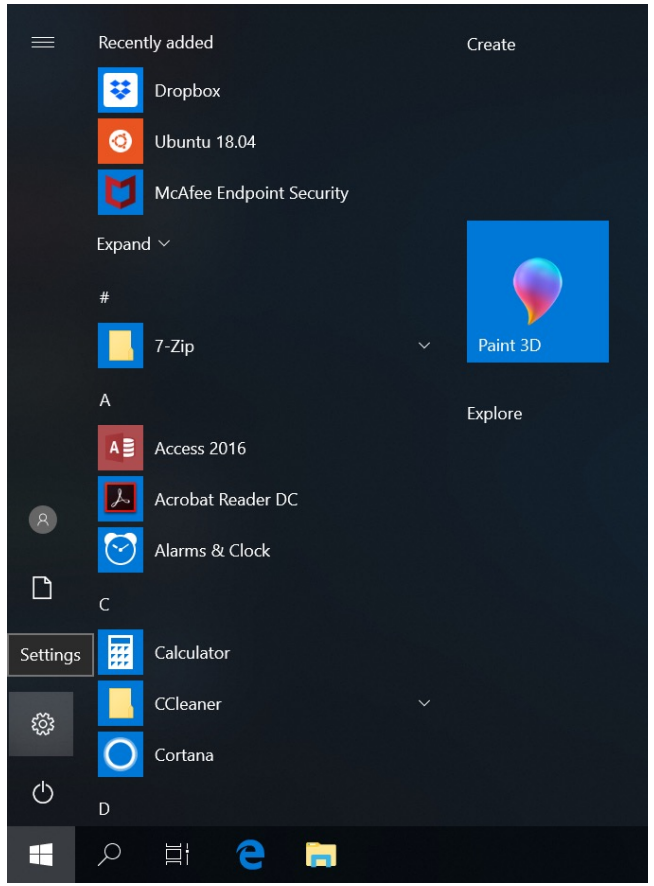
Interface



Launching a Unix Shell

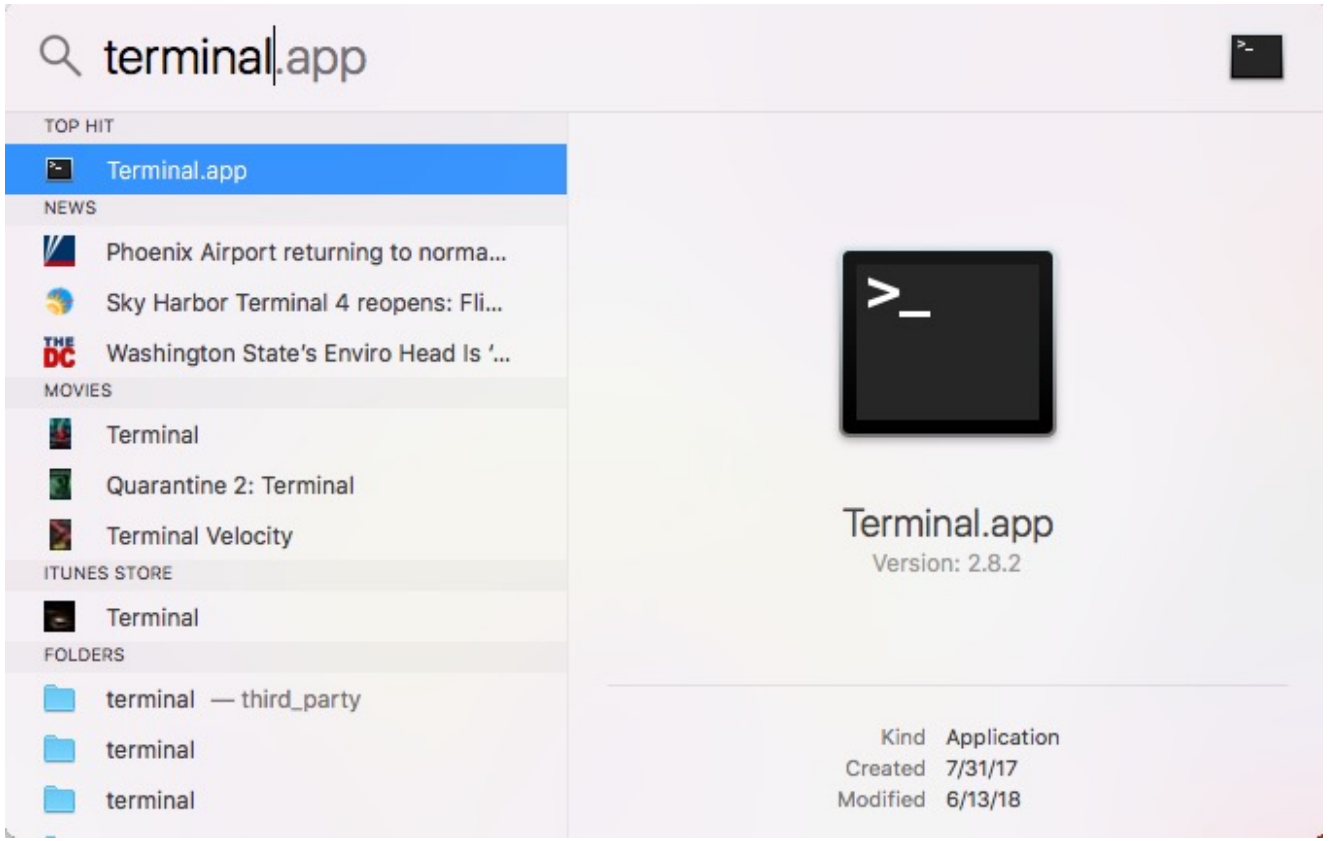
Windows 10

- Open the “Ubuntu” App

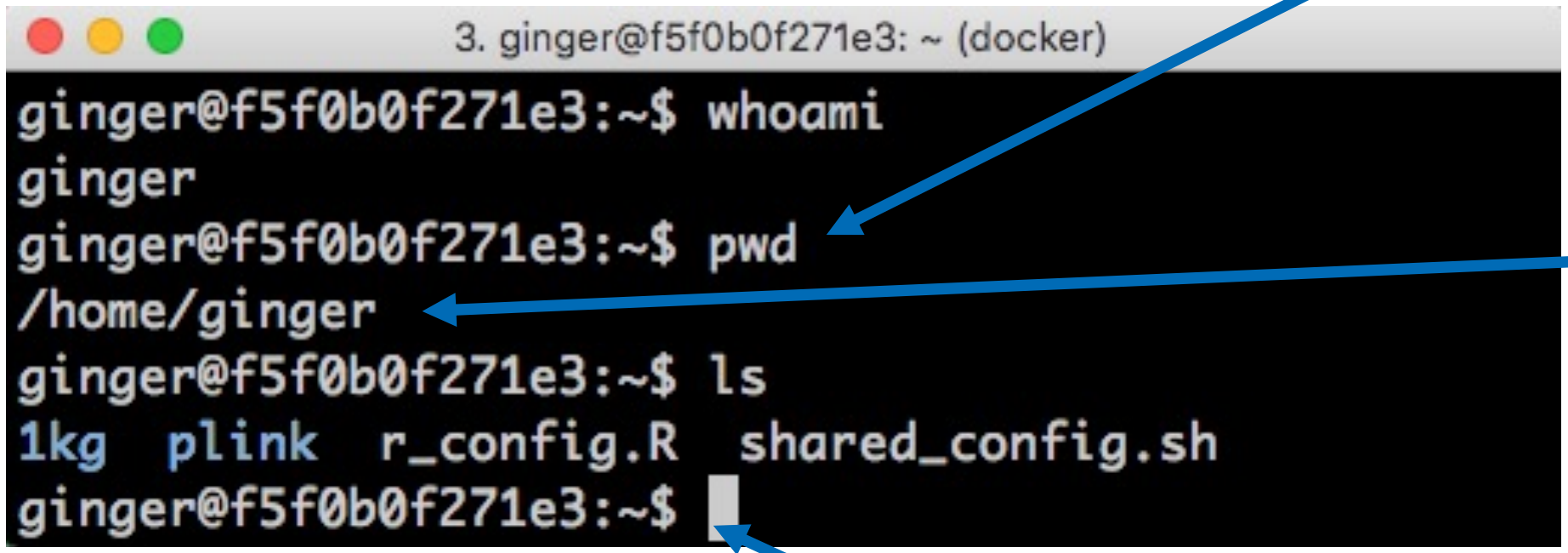


Mac OS

- Press ⌘ + Space and then type “terminal”



What does a Shell look like?



A command previously execute
"pwd"

Result of
"pwd"
command

Prompt – waiting for the next command

How do you know what to type in the Shell?

Unix/Linux Command Reference

FOSSwire.com

File Commands

- **ls** – directory listing
- **ls -al** – formatted listing with hidden files
- **cd *dir*** - change directory to *dir*
- **cd** – change to home
- **pwd** – show current directory
- **mkdir *dir*** – create a directory *dir*
- **rm *file*** – delete *file*
- **rm -r *dir*** – delete directory *dir*
- **rm -f *file*** – force remove *file*
- **rm -rf *dir*** – force remove directory *dir* *
- **cp *file1 file2*** – copy *file1* to *file2*
- **cp -r *dir1 dir2*** – copy *dir1* to *dir2*; create *dir2* if it doesn't exist
- **mv *file1 file2*** – rename or move *file1* to *file2* if *file2* is an existing directory, moves *file1* into directory *file2*
- **ln -s *file link*** – create symbolic link *link* to *file*
- **touch *file*** – create or update *file*
- **cat > *file*** – places standard input into *file*
- **more *file*** – output the contents of *file*
- **head *file*** – output the first 10 lines of *file*
- **tail *file*** – output the last 10 lines of *file*
- **tail -f *file*** – output the contents of *file* as it grows, starting with the last 10 lines

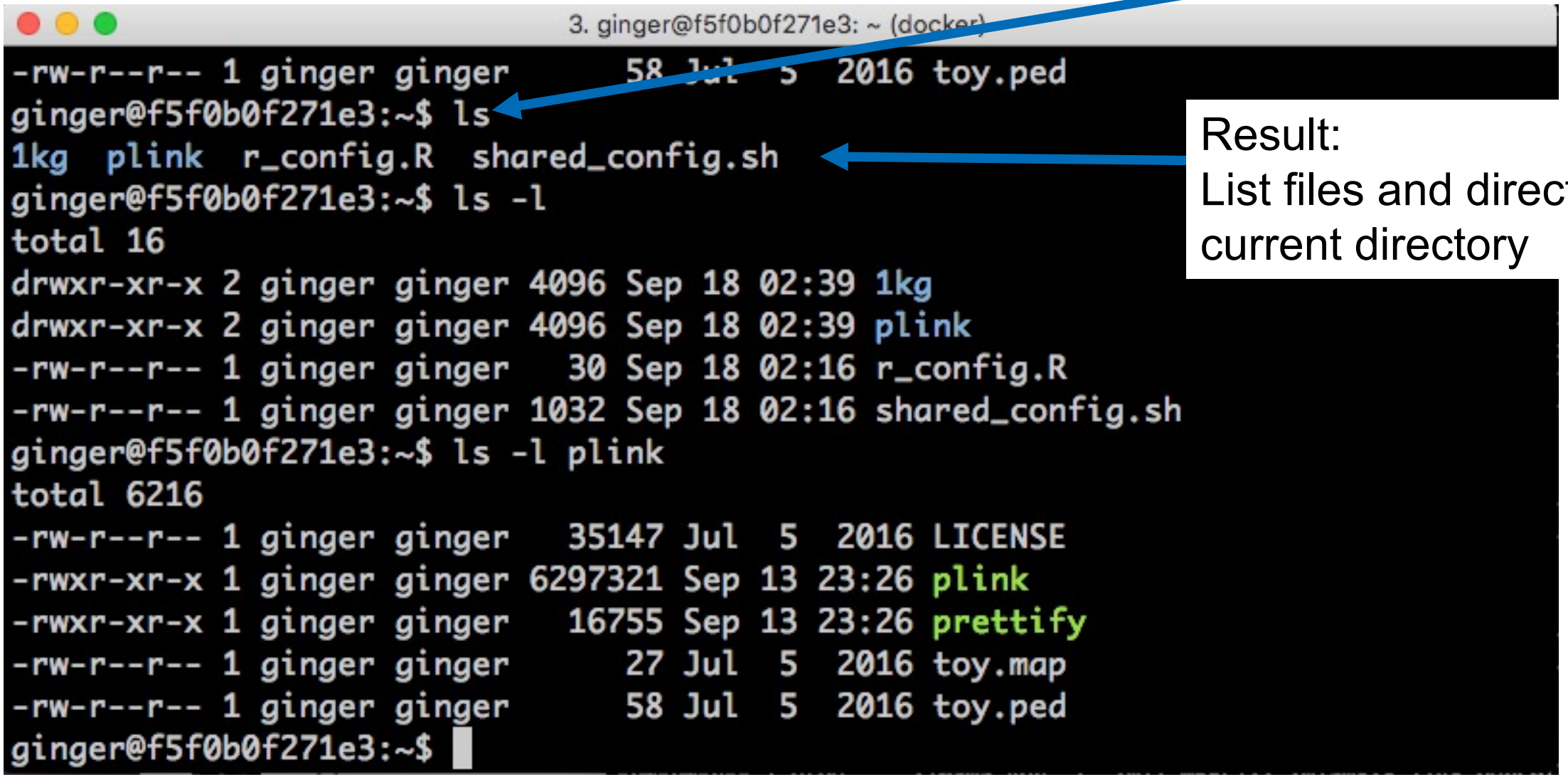
System Info

- **date** – show the current date and time
- **cal** – show this month's calendar
- **uptime** – show current uptime
- **w** – display who is online
- **whoami** – who you are logged in as
- **finger *user*** – display information about *user*
- **uname -a** – show kernel information
- **cat /proc/cpuinfo** – cpu information
- **cat /proc/meminfo** – memory information
- **man *command*** – show the manual for *command*
- **df** – show disk usage
- **du** – show directory space usage
- **free** – show memory and swap usage
- **whereis *app*** – show possible locations of *app*
- **which *app*** – show which *app* will be run by default

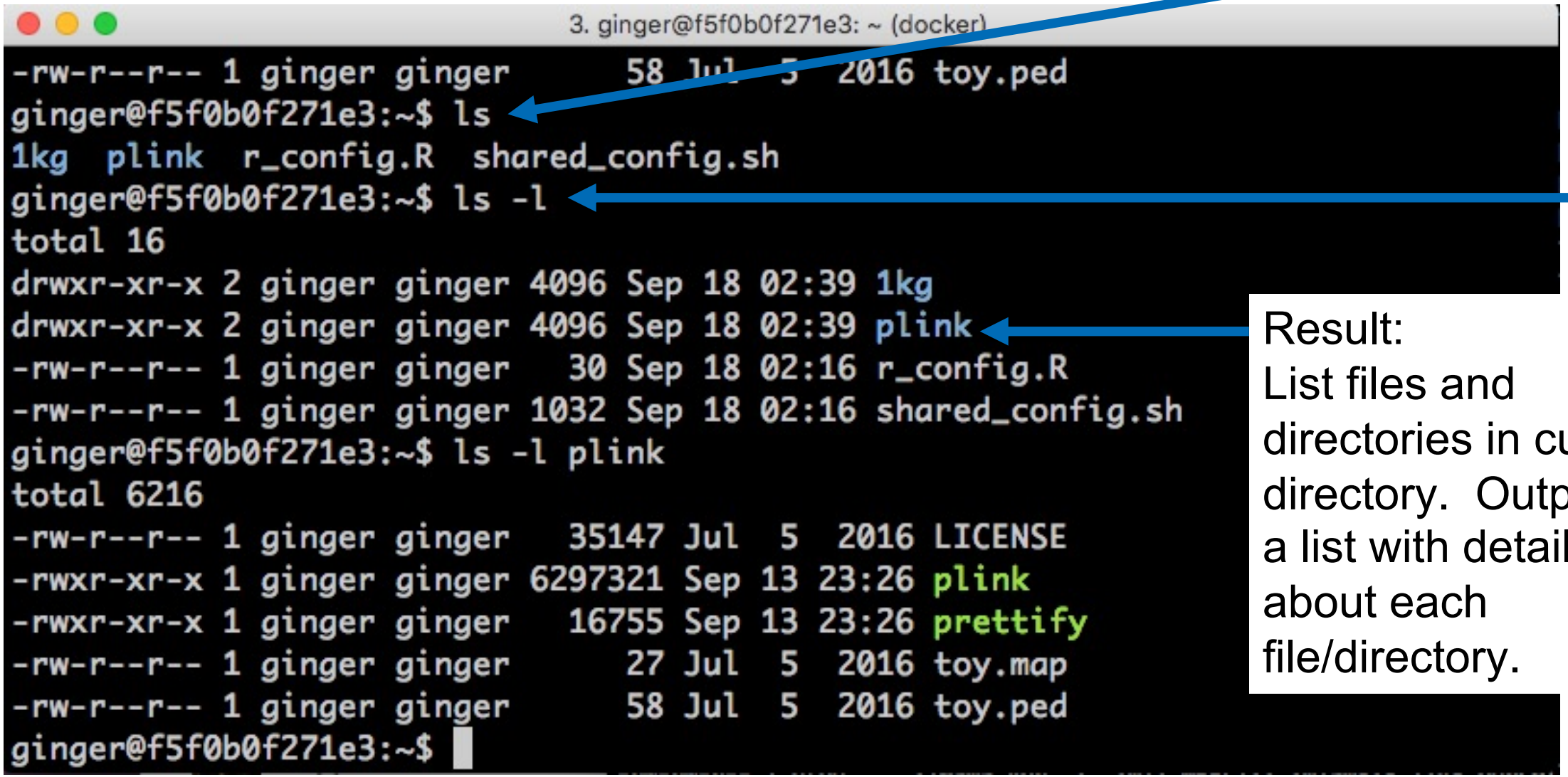
Compression

- **tar cf *file.tar files*** – create a tar named *file.tar* containing *files*
- **tar xf *file.tar*** – extract the files from *file.tar*
- **tar czf *file.tar.gz files*** – create a tar with Gzip compression
- **tar xzf *file.tar.gz*** – extract a tar using Gzip
- **tar cjf *file.tar.bz2*** – create a tar with Bzip2 compression

Anatomy of a Shell command



Anatomy of a Shell command

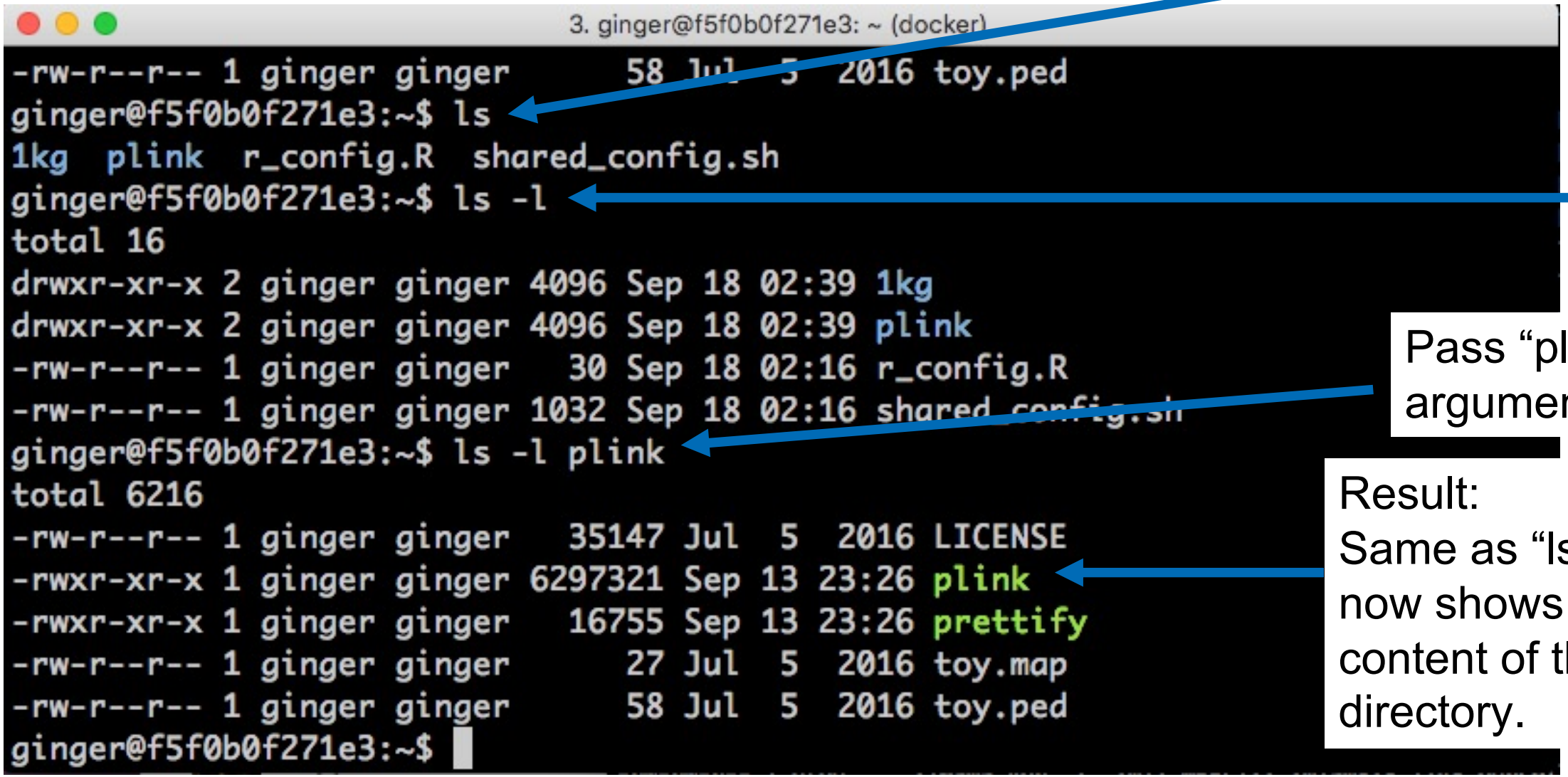


"ls" command

"-l" flag

Result:
List files and directories in current directory. Output is a list with details about each file/directory.

Anatomy of a Shell command



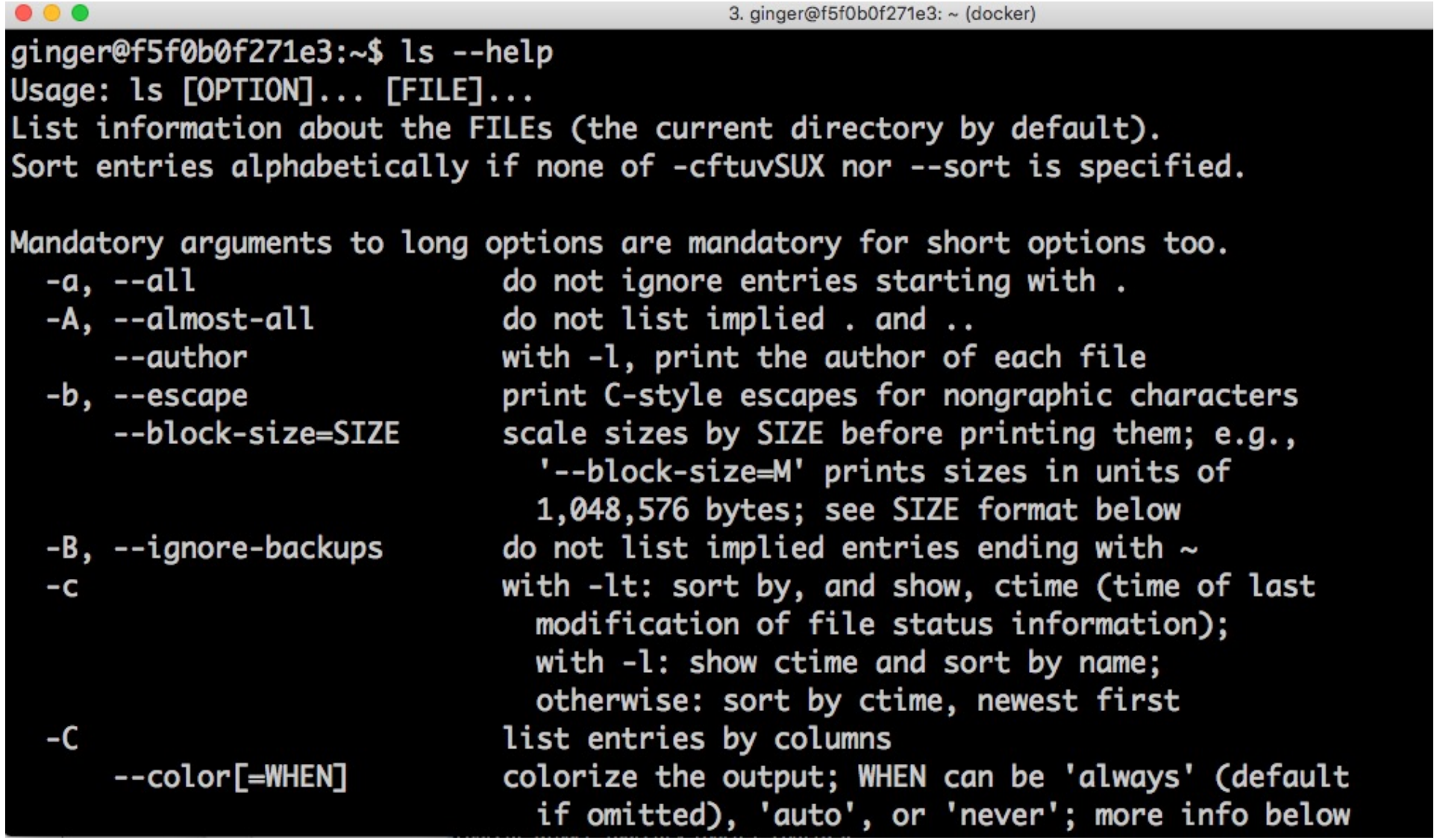
"ls" command

"-l" flag

Pass "plink" as argument to "ls"

Result:
Same as "ls –l", but now shows the content of the "plink" directory.

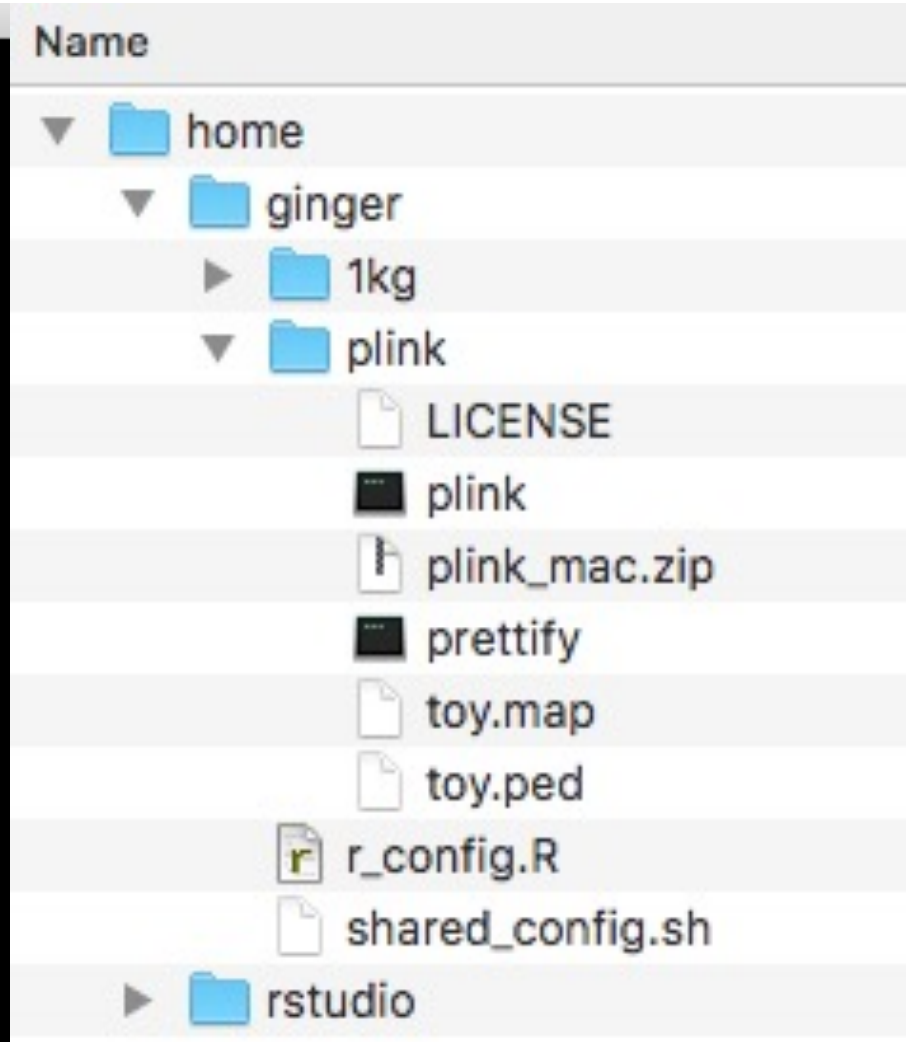
Getting help



Interacting with the file system

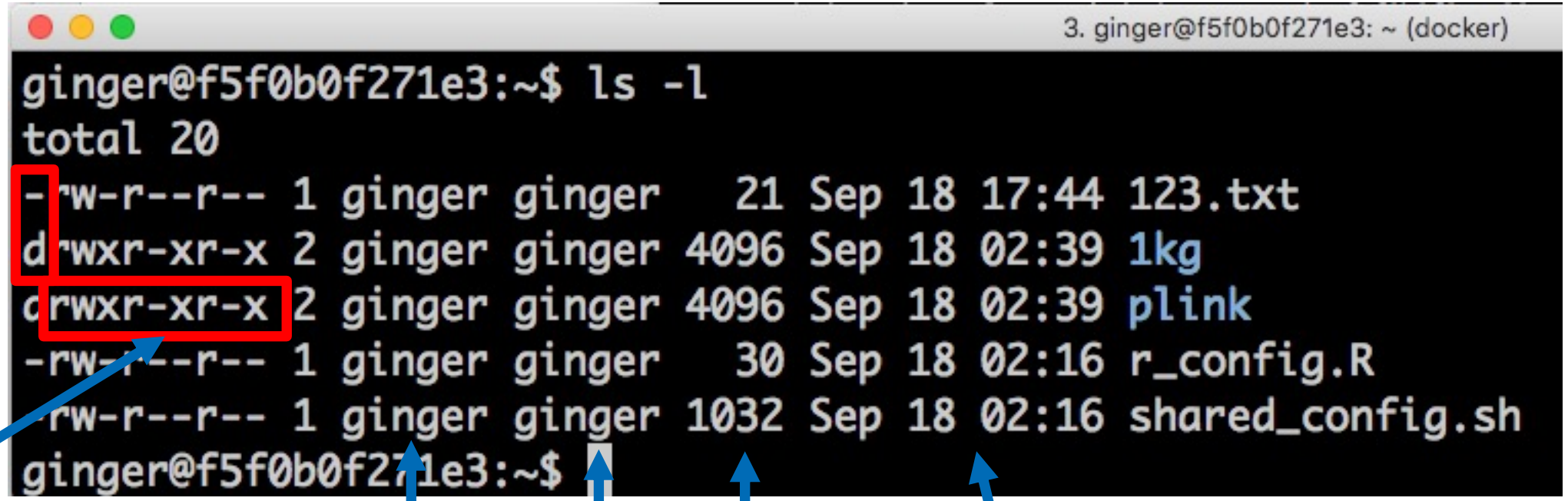
Moving around the filesystem

```
3. ginger@f5f0b0f271e3: ~ (docker)
ginger@f5f0b0f271e3:~$
ginger@f5f0b0f271e3:~$ pwd
/home/ginger
ginger@f5f0b0f271e3:~$ ls
1kg  plink  r_config.R  shared_config.sh
ginger@f5f0b0f271e3:~$ cd plink
ginger@f5f0b0f271e3:~/plink$ pwd
/home/ginger/plink
ginger@f5f0b0f271e3:~/plink$ cd ..
ginger@f5f0b0f271e3:~$ pwd
/home/ginger
ginger@f5f0b0f271e3:~$ cd /home
ginger@f5f0b0f271e3:/home$ ls
ginger  rstudio
ginger@f5f0b0f271e3:/home$ cd ~
ginger@f5f0b0f271e3:~$ pwd
/home/ginger
ginger@f5f0b0f271e3:~$ cd .
ginger@f5f0b0f271e3:~$ pwd
/home/ginger
ginger@f5f0b0f271e3:~$
```



More about the filesystem and ls -l

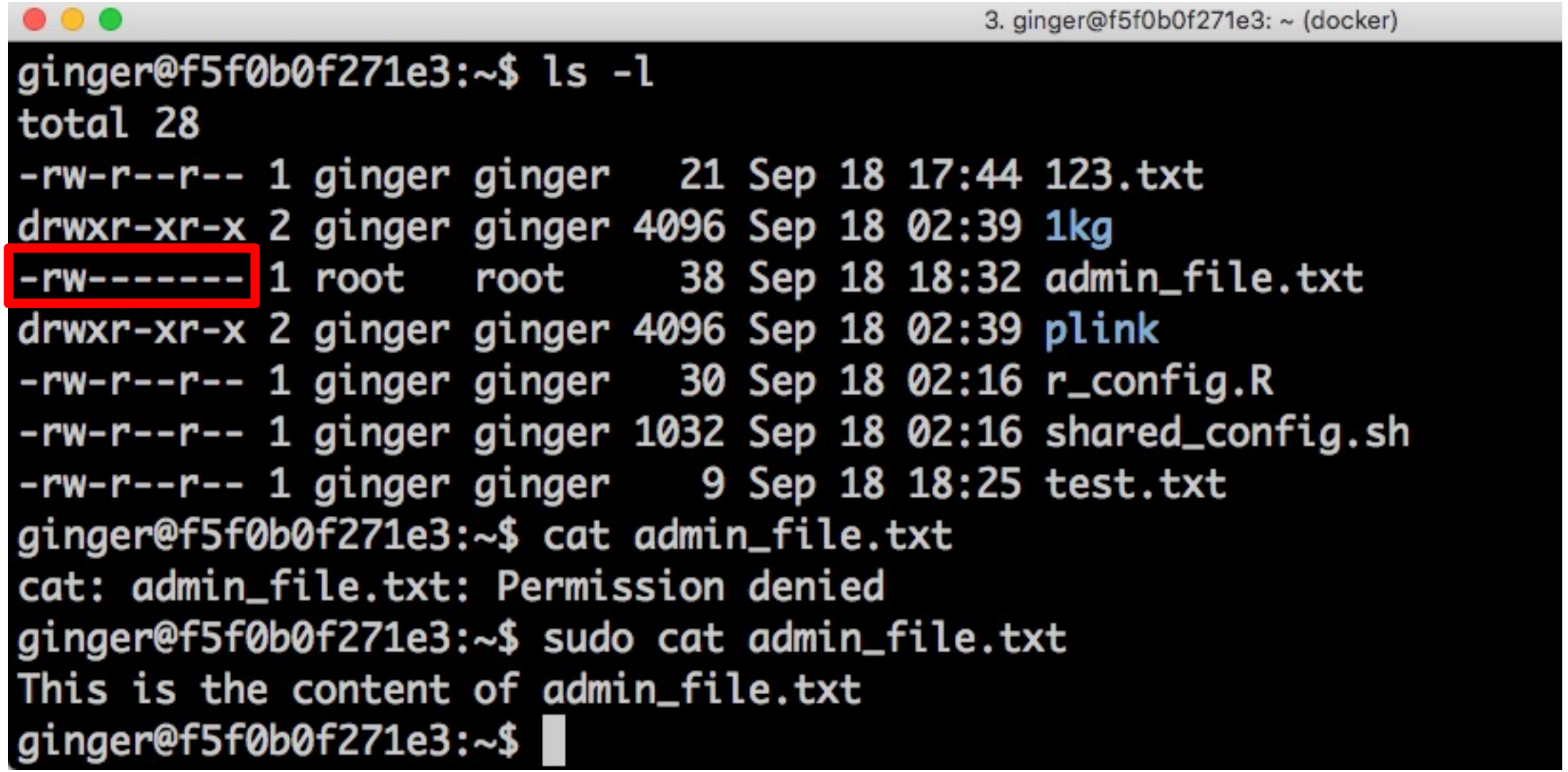
- = file
d = directory



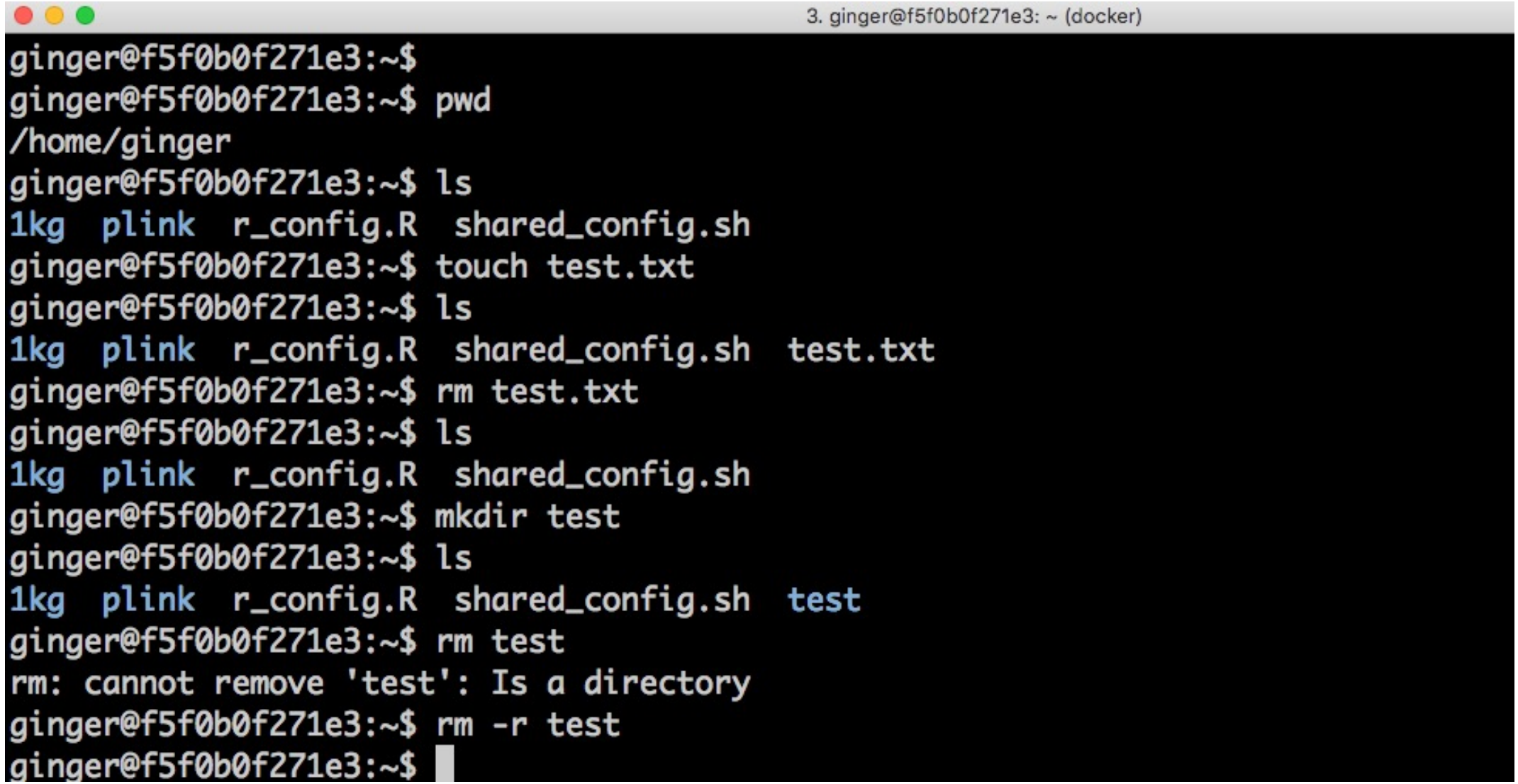
Permissions:
rwx r-x r-x

owner group size modification date

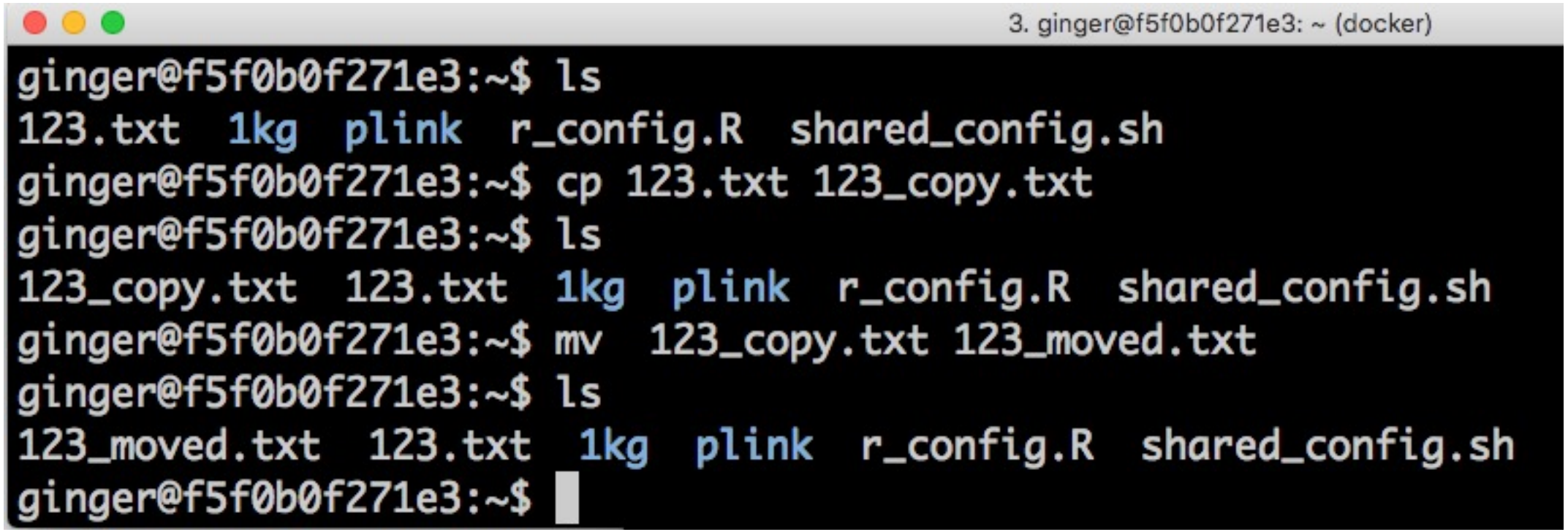
Sudo: becoming root (admin) for a command



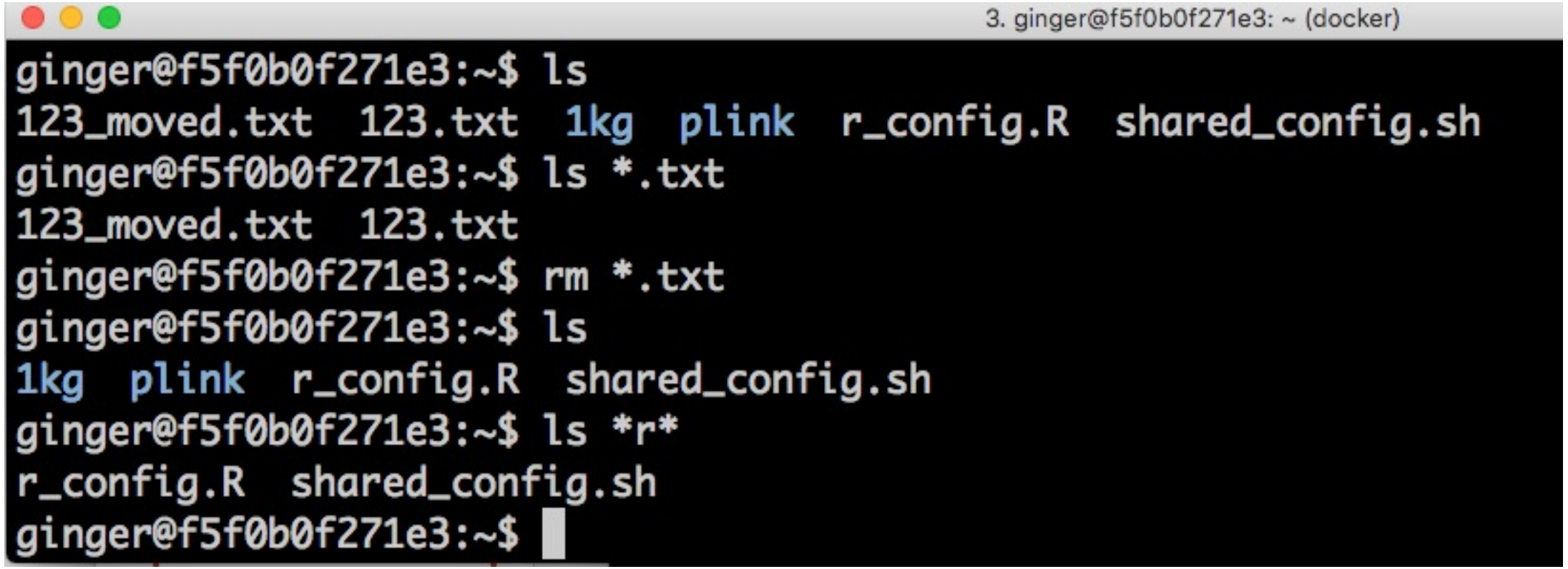
Creating / deleting files and folders



Copying and moving files

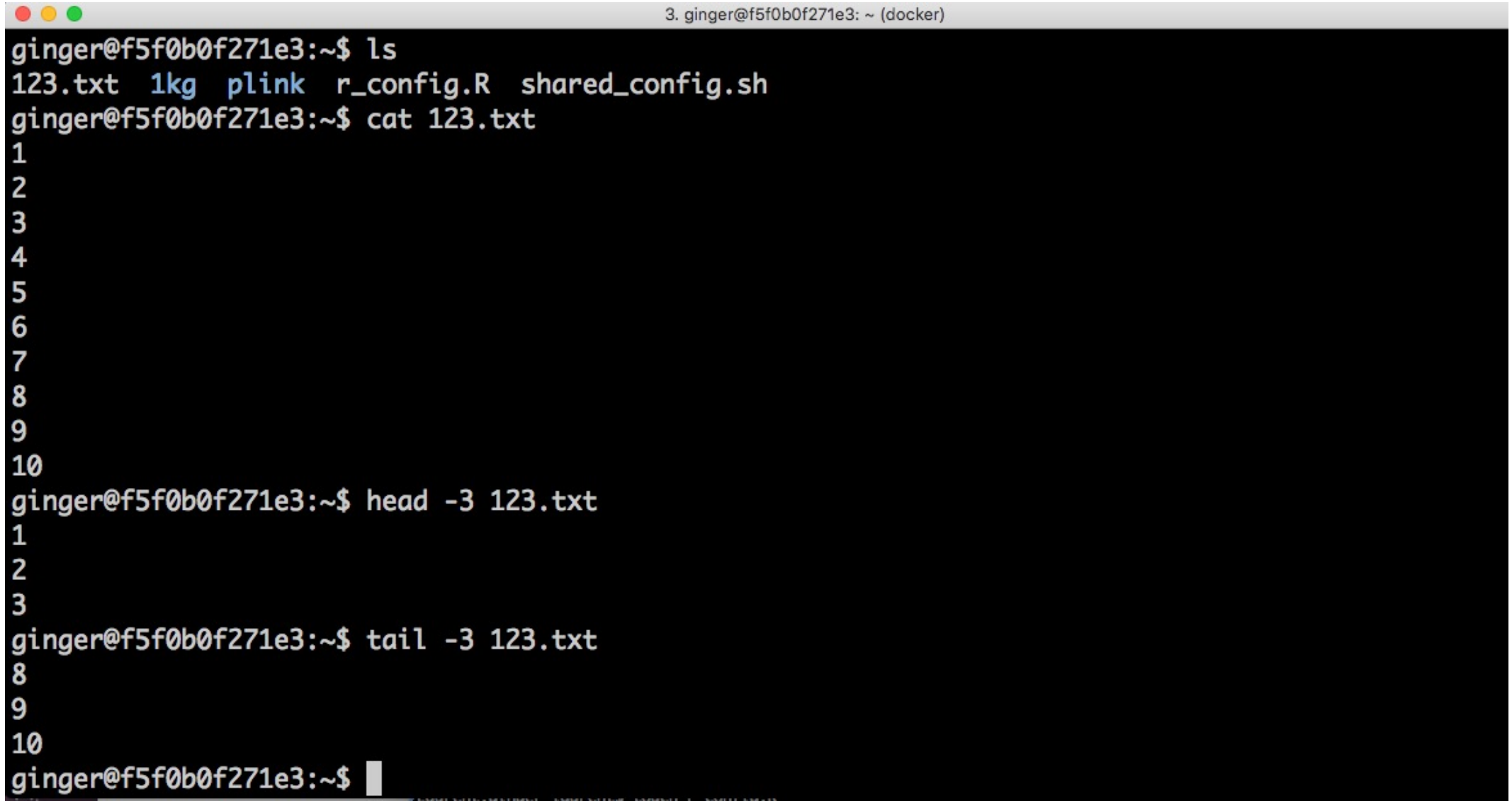


Wildcards (*)


3. ginger@f5f0b0f271e3: ~ (docker)
ginger@f5f0b0f271e3:~$ ls
123_moved.txt 123.txt 1kg plink r_config.R shared_config.sh
ginger@f5f0b0f271e3:~$ ls *.txt
123_moved.txt 123.txt
ginger@f5f0b0f271e3:~$ rm *.txt
ginger@f5f0b0f271e3:~$ ls
1kg plink r_config.R shared_config.sh
ginger@f5f0b0f271e3:~$ ls *r*
r_config.R shared_config.sh
ginger@f5f0b0f271e3:~$

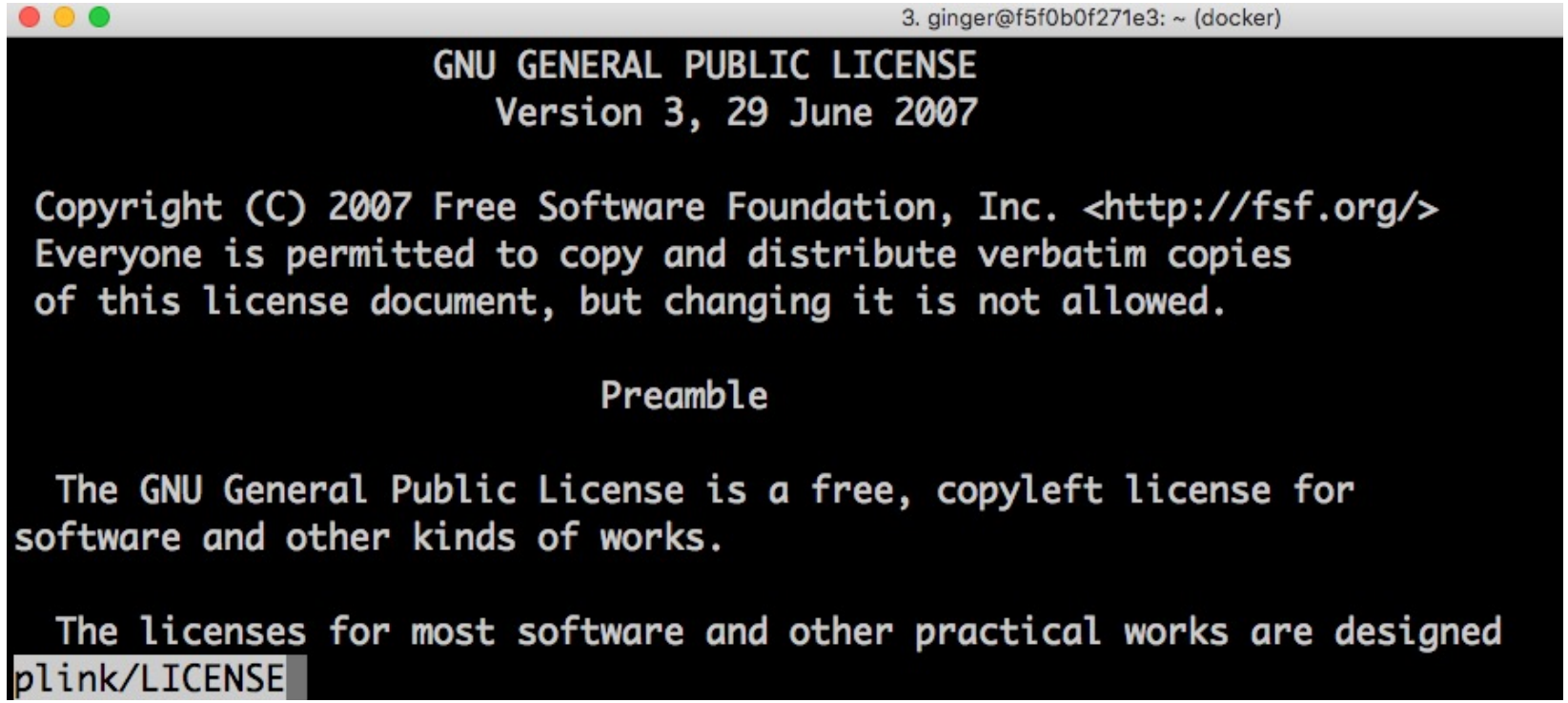
Interacting with text files

Viewing the content of a text file



Reading longer text files: less

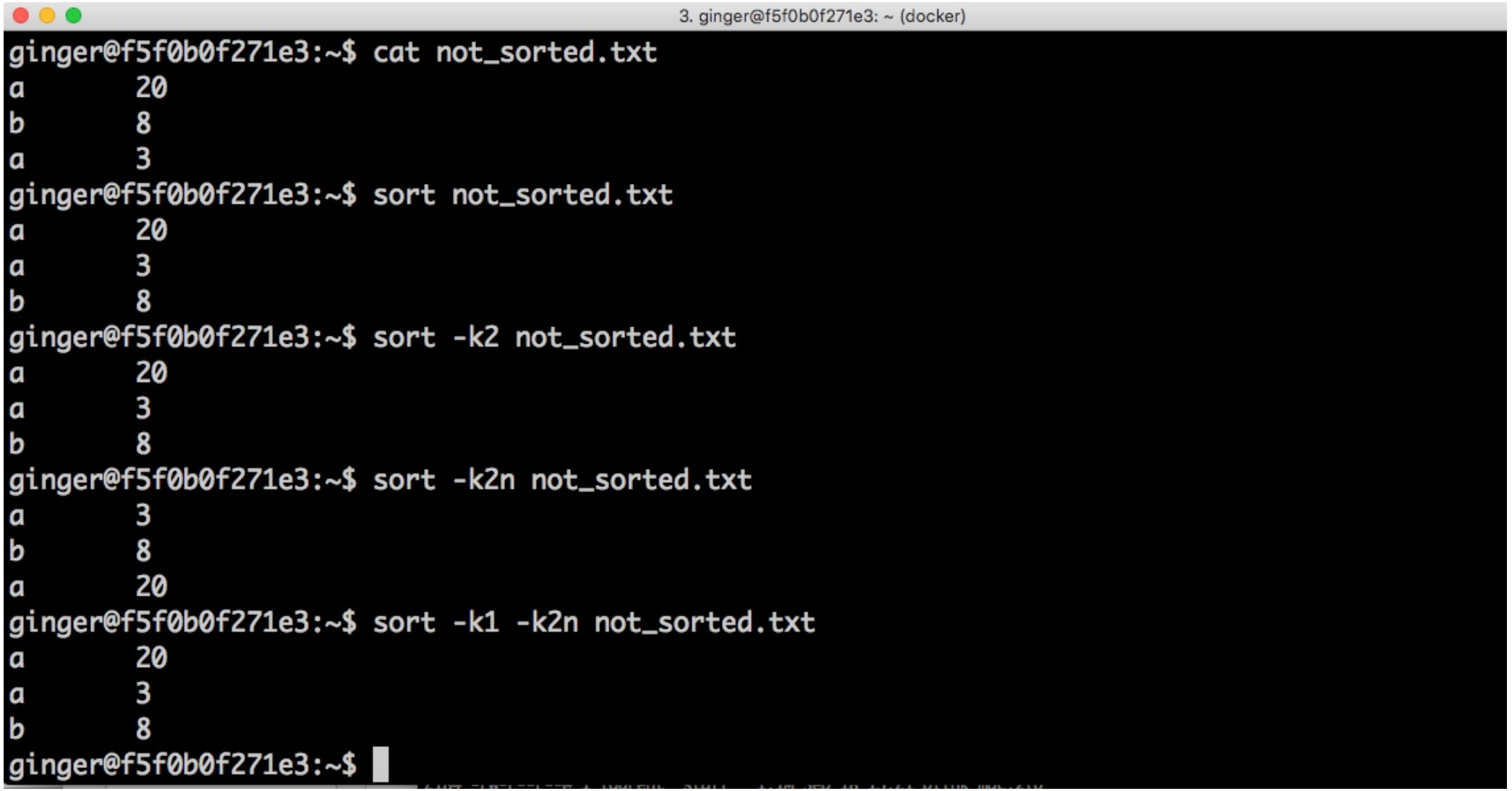
```
ginger@f5f0b0f271e3:~$ ls plink/
LICENSE  plink  prettify  toy.map  toy.ped
ginger@f5f0b0f271e3:~$ less plink/LICENSE
```



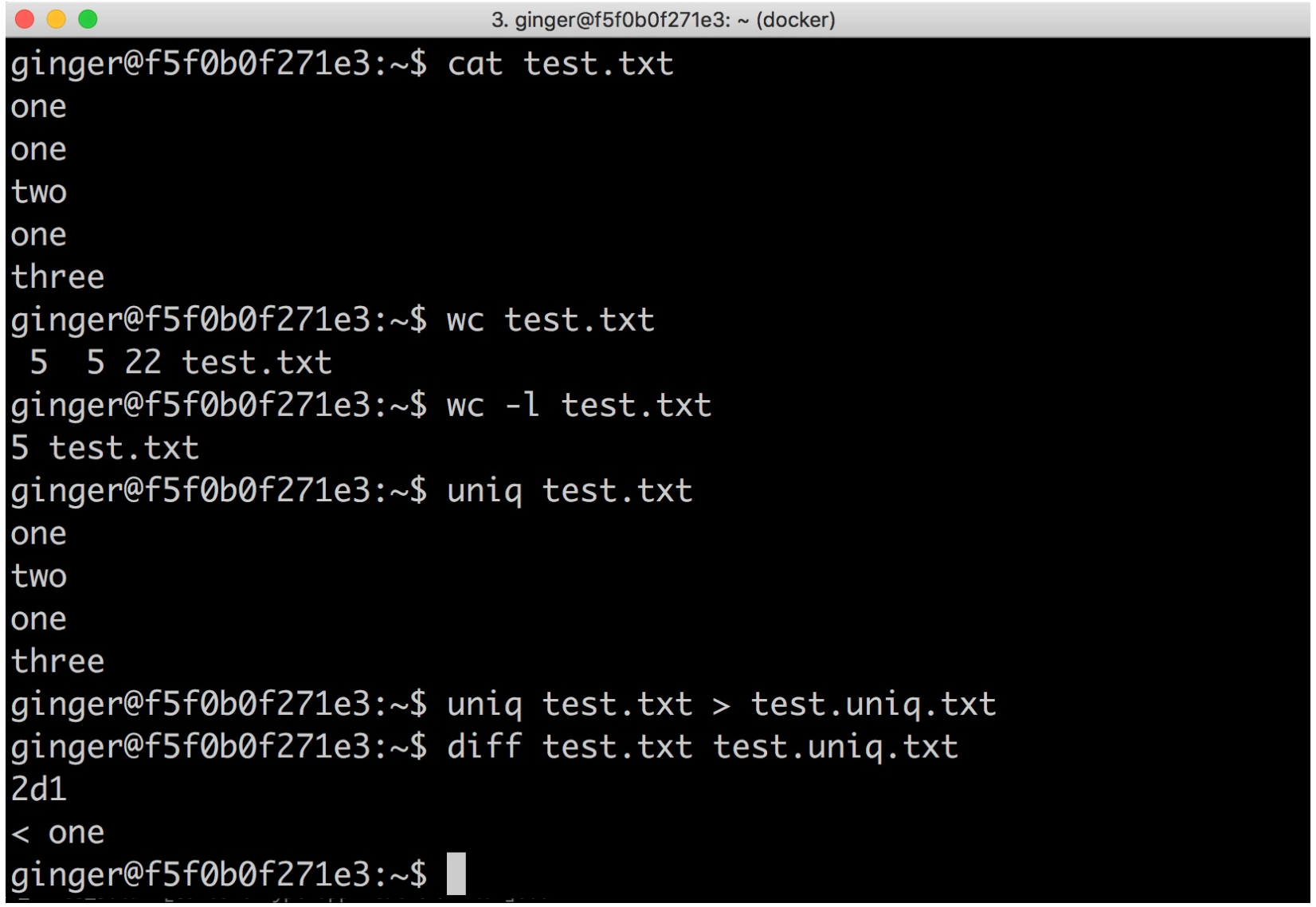
Commands:

- q: quit
- f: forward 1 page
- b back 1 page
- ↑↓→←: move by 1 character in the direction of the arrow
- /text: search for “text”
- -S <enter>: chop long lines
- 100g: Go to line 100

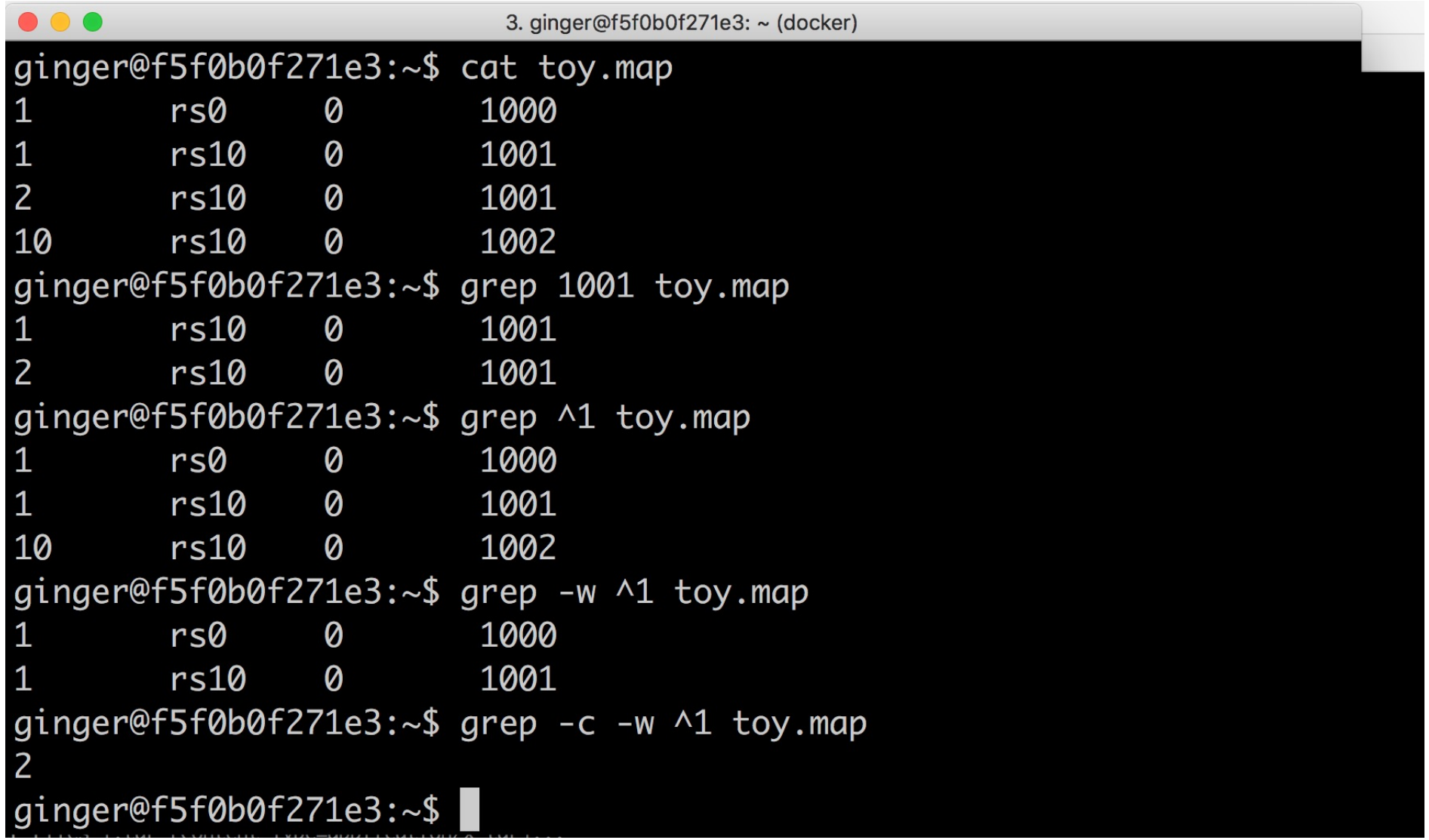
Sorting text files: sort



More text utilities: wc, uniq and diff



Searching a text file: grep (basic)



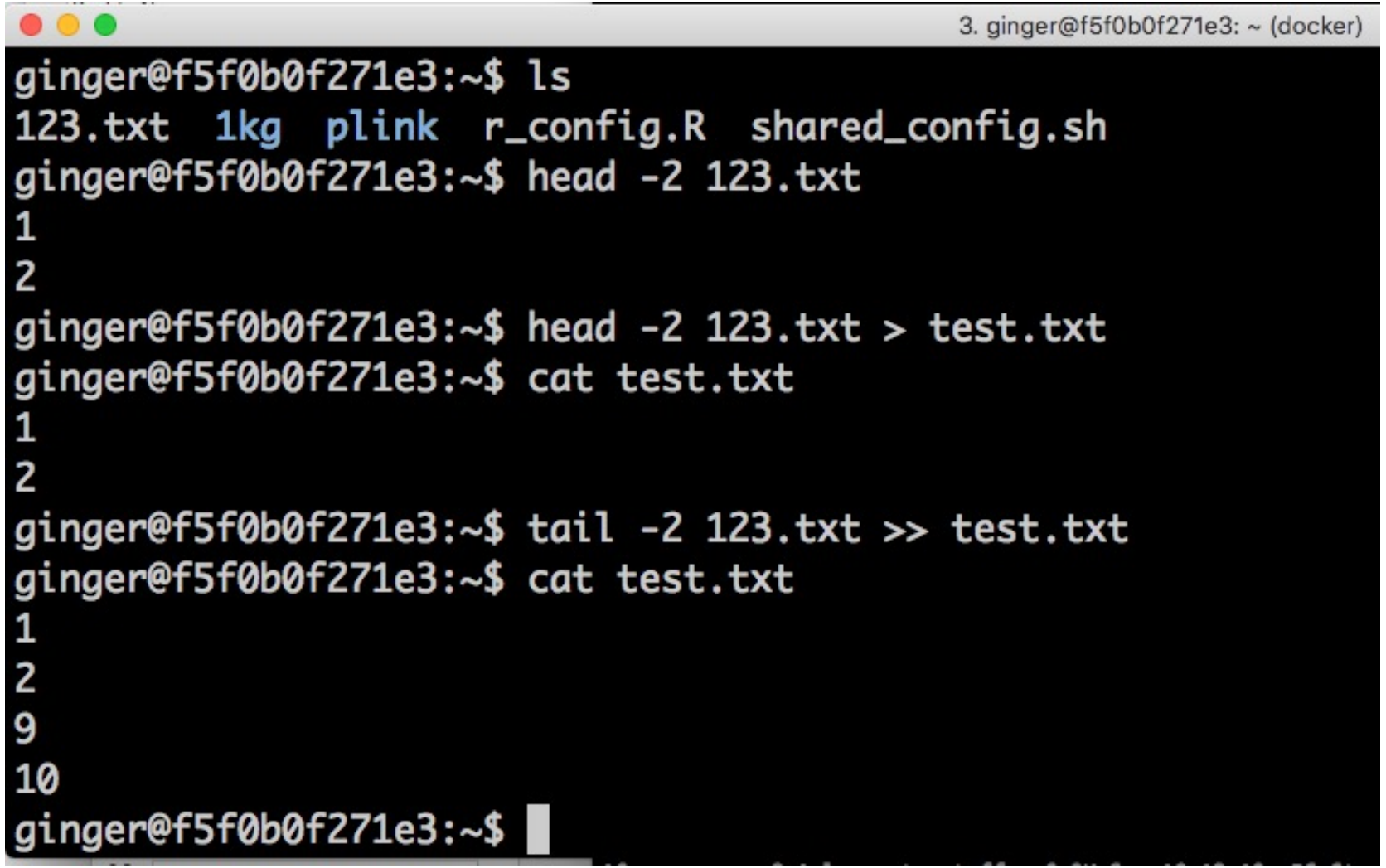
Looking for a file within a subdirectory: find

- Go into a directory and try:
  - find .

- Maybe you know what the name is:
  - find . -name “myfile.txt”

- Maybe you just know the extension:
  - find . -name “*.R”

Redirecting output to a text file



Looking back at your old commands

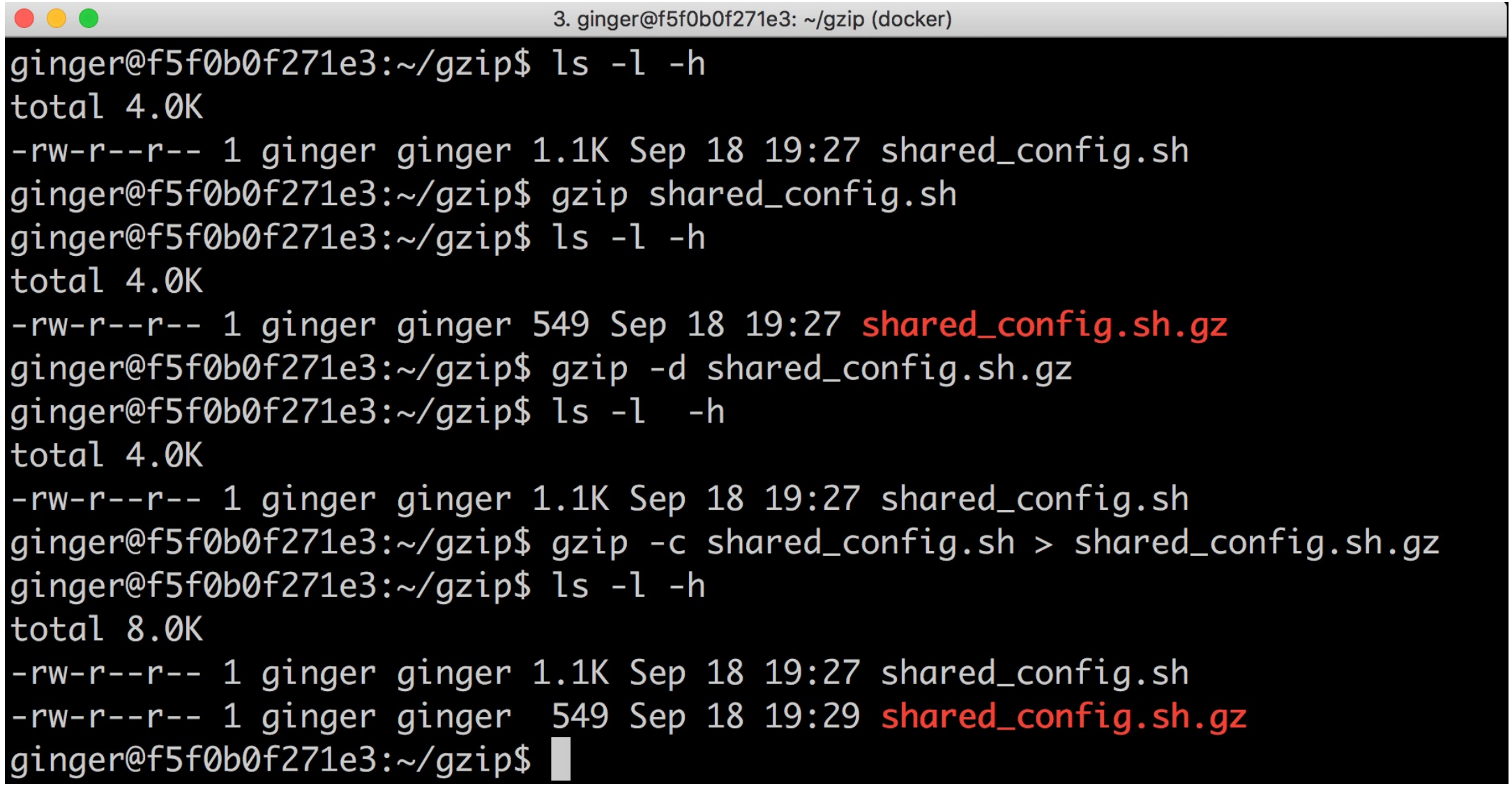
- Run the command: history

Compressing text data

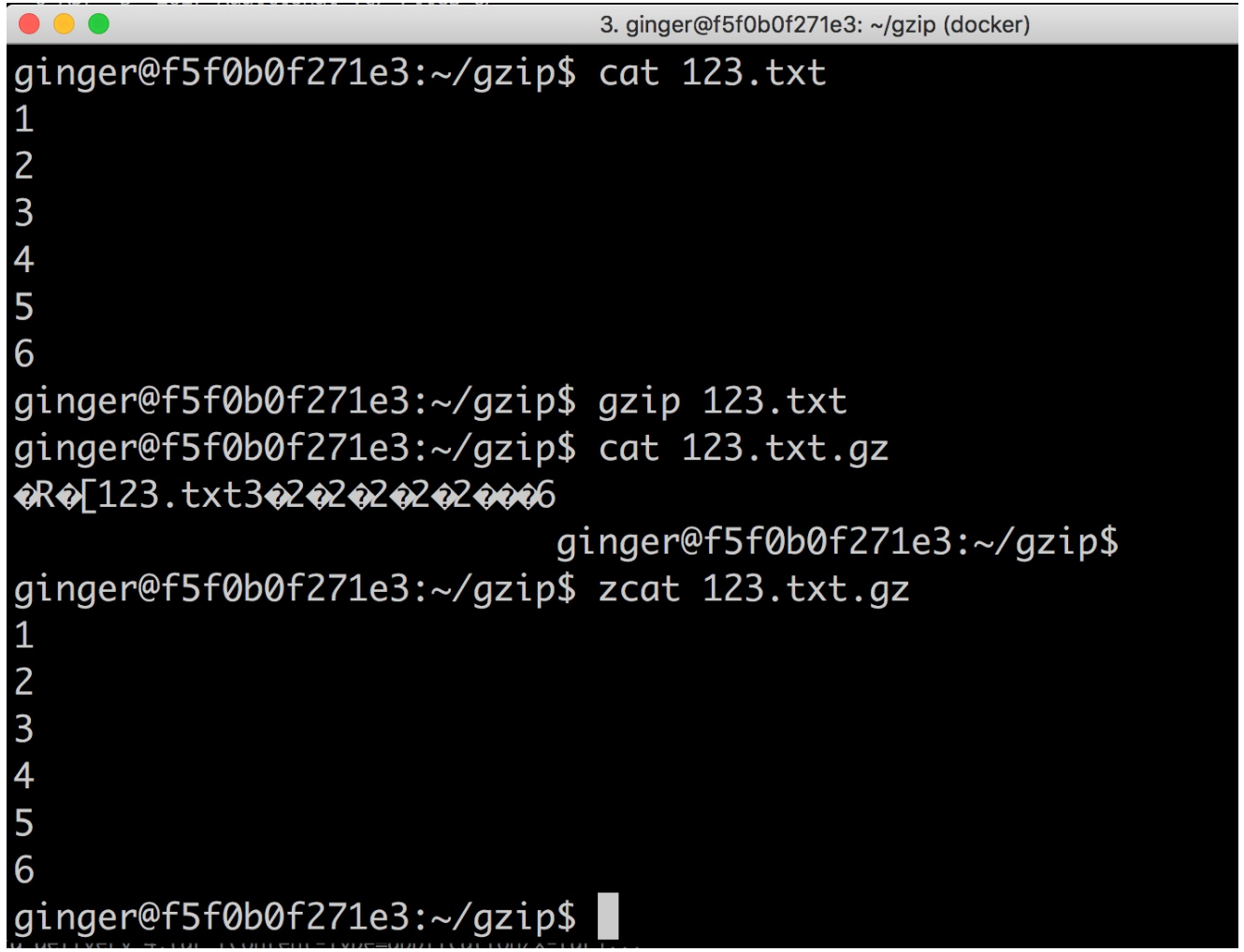
gzip everything!

- There are many algorithms for file compression. gzip is the most popular in the unix world
- Achieves compression of 1/10th to ½ file size
- Many languages can read gzip files natively – and often faster!
  - Python: gzip.open(“file.gz”)
  - R: read.table(‘file.gz’)

Compressing text files with gzip



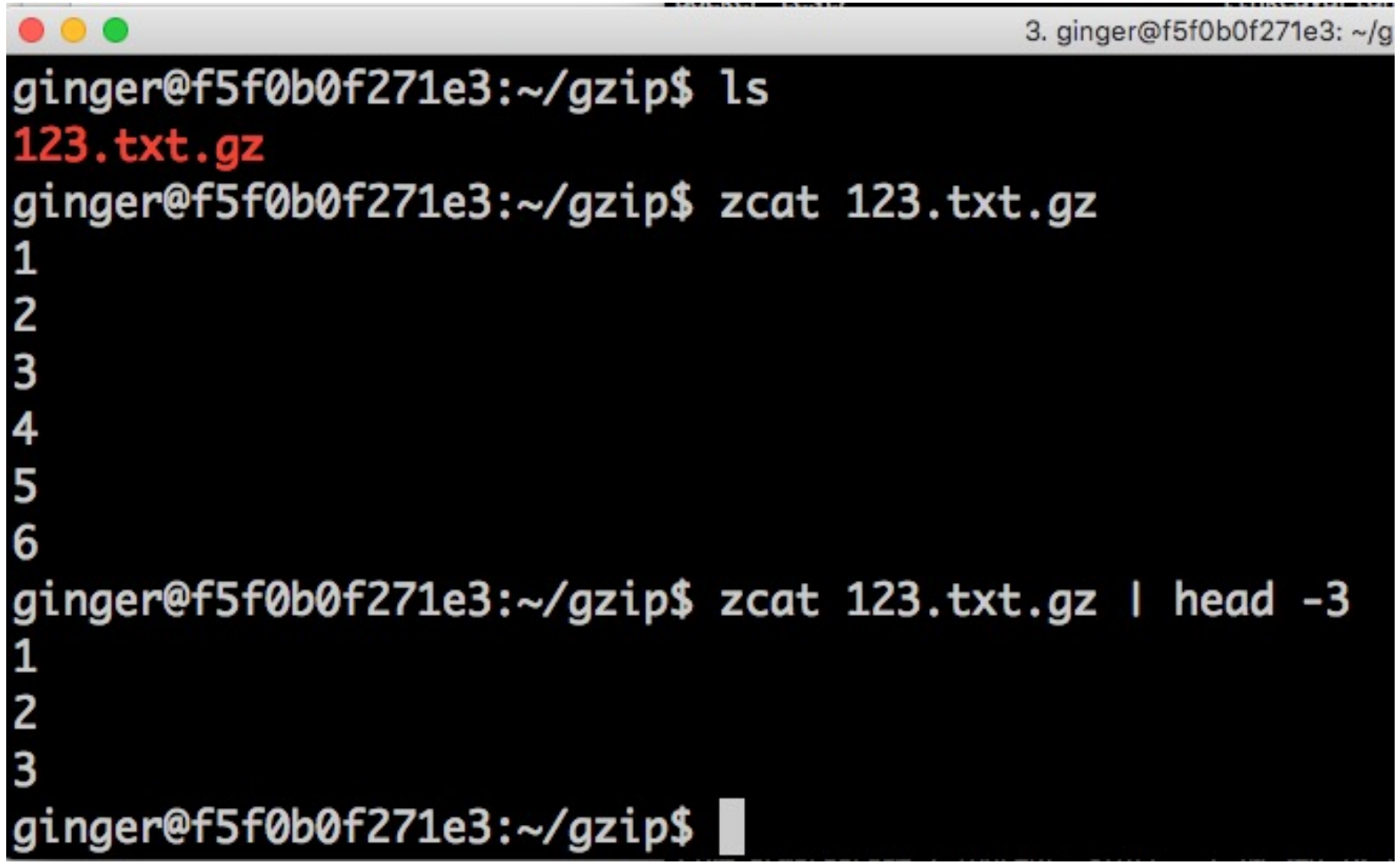
Working with gzipped files



- cat -> zcat (gzcat on MacOS)
- grep -> zgrep
- less -> zless
- For other tools, we need pipes!

Pipes

Pipe example: display 3 first lines of gzipped



A pipe (“|”) joins two commands together by taking the output (stdout) of the left command and sending it as input (stdin) to the right command.

You can chain as many commands as you like with pipes

```
3. ginger@f5f0b0f271e3: ~ (docker)
ginger@f5f0b0f271e3:~$ cat toy.map
1       rs0     0       1000
1       rs0     0       1000
1       rs10    0       1001
2       rs10    0       1001
10      rs10    0       1002
1       rs0     0       1000
ginger@f5f0b0f271e3:~$ sort -k1n toy.map | uniq | wc -l
4
ginger@f5f0b0f271e3:~$ grep -w rs10 toy.map | sort -k1n | uniq | gzip -c > result.txt.gz
ginger@f5f0b0f271e3:~$ zcat result.txt.gz
1       rs10    0       1001
2       rs10    0       1001
10      rs10    0       1002
ginger@f5f0b0f271e3:~$
```

Pipes to find old commands

- Look through your old commands
  - history | less

- Find an old command you wrote that you knew had the word “find” in it
  - history | grep find

Disconnecting from a session to reconnect later

- The most common tools are: screen and tmux

- screen starts a new screen
- Detach the window by holding control, and typing: ad
- Then you can safely log off
- Re-attach the window using: screen -r

Running large computations

- On a cluster:
  - qsub: Sun Grid Engine
  - bsub: LSF
- On the cloud:
  - dsub: Closest to a grid system like q/bsub. Some light setup required
  - Cromwell/WDL: requires a server to set up and bulky json/wdl files to run
  - Hail Batch (coming soon!): will be the easiest but not yet available

Tasks for today wrt Linux

1. Ensure gcloud CLI is installed

2. Check your cloud access
    1. using `gcloud auth list`
        1. Then config accordingly
    2. 1.1) fails, try out `gcloud auth login`
3. What if you are lazy to point and click? We could use gsutil. Let's find out what is in this bucket: `gs://neurogap_phenos_genos/` (hint something to do with `ls` or listing files)

4. Remember that VM we created? How do you, using your own command line? Ask Kumar
    1. ssh into it
    2. pausing vs stopping
    3. upload vs download files to VM
5. Download the exercise in pdf and do the exercises on your VM

Going further

- https://swcarpentry.github.io/shell-novice/
- http://swcarpentry.github.io/shell-extras/

MASSACHUSETTS GENERAL HOSPITAL

HARVARD MEDICAL SCHOOL

BROAD INSTITUTE

GINGER On-Site Training
Day 1: Intro to Linux
QUESTIONS? ☺

GINGER Program 2022

University of Cape Town

Teaching Fellows:

Kumar Veerapen, Ph.D.
Senior Expert I Data Science
Kumar.Veerapen@Novartis.com

Carla Marquez-Luna, Ph.D.
Postdoctoral Research Fellow
carlamarquezluna@gmail.com