GNU/Linux: Introduction and Administration

3h Session for the course on "Data Science. Applications to Biology and Medicine with Python and R", at IL3 - University of Barcelona[1]. April 2rd, 2025. 16:00h-19:00h.

https://seeds4c.org/LinuxDataScience25[2]

Presentation Slides

Video recording (from a previous edition, a few years ago)

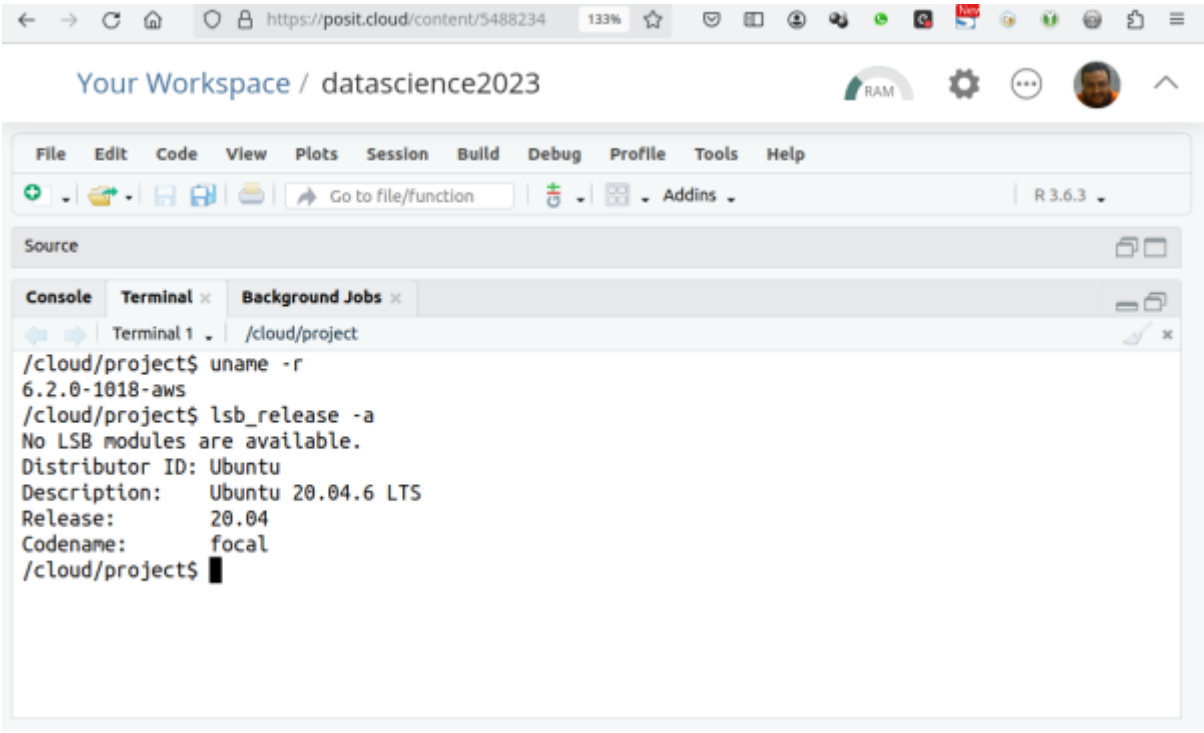
SLIDES IN PDF:
https://www.slideshare.net/slideshow/gnu-linux-introduction-and-administration-aed1/277441886[3]

Hands-on Exercise

Source data derived from data obtained from here:
https://analisi.transparenciacatalunya.cat/en/Medi-Ambient/Dades-meteorol-giques-de-la-XEMA/nzvn-apee[4]

Steps:

1. **PART A**: Enter the GNU/Linux machine.
   Choose one option from the following 3 options below:
   1. Sign up at https://posit.cloud/plans/free[5] to get a free account. Connect to posit.cloud and use the terminal window from the RStudio server there.



OR

.

2. Import a recent **iso** file form the Lubuntu GNU Linux distribution[6] (latest Long Term Support version, 24.04 LTS as of this writing) into the VirtualBox program[7] in your own computer.
    - ISO File:

      ```
      https://cdimage.ubuntu.com/lubuntu/releases/noble/release/lubuntu-24.04.2-desktop-amd64.iso
      ```

      .

      (Or alternatively, import an older but customized version of Lubuntu GNU Linux through, through importing the **.ova** file provided below for VirtualBox (explained within the session notes). OVA file: http://cloud.seeds4c.org/lubuntu_1804_64bit_v03.ova)

      .

      Keep in mind that it will take some time to download the iso (3.1 Gb) or ova file (7.6Gb), and also to import it to your Virtual Box (5-10 minutes or more),

.

2. **PART B**: Fetch and subset data

   Obtain a subset of columns and rows from a dataset, using Linux simple commands in a terminal (using shell commands, not R nor Python in this case),

    1. Copy the source data file (**data_smc.csv.bz2** from the usb disk provided by the course

professor), or from here for instance: http://cloud.seeds4c.org/data_smc.csv.bz2[8] (50Mb file, 10.000.000 rows csv file, bz2 compressed)

Open a Linux terminal in your home folder /home/datascience/

```
cd /home/userNN/ # just in case, change directory to your home folder
wget http://cloud.seeds4c.org/data_smc.csv.bz2 # fetch the file from the internet
```

2. Uncompress (`bunzip2 file.bz2 -k`) and show (with `cat file`), or use +-bzcat file.bz2 -k+- to send to standard output (stdout) on-the-fly while keeping the source compressed file (-k)

```
bunzip2 data_smc.csv.bz2 -k
```

3. filter (keep) the first 100 rows (with `head -n100 file`)
4. save as new file: `file.csv`

Oneliner with the previous commands piped one after the other in the same line

```
bzcat data_smc.csv.bz2 -k | head -n100 > file_all.csv
```

5. filter out one column, for instance, remove column 7 (variable _), with `cut`

```
cut --complement -d',' -f7 file_all.csv > file.csv
```

6. save in zip

```
zip file.csv.zip file.csv
```

7. Change permissions so that only your user can read and write it

```
chmod 600 file.csv.zip
```

.

3. **PART C**: Your turn
   - Creativity, Exploration...
   - Doubts?

That should be it. Done!

Feel free to test more Linux commands in the linux terminal window from your positcloud space, or from the Linux you have imported in VirtualBox.

Additional info

If you want to keep practising and learning, beyond this course session, you can do so for instance here:

1. https://davidadrian.cc/definitive-data-scientist-setup/[9]

Alias names for this page:
GNULinuxOS25 | LinuxDataScience25

---

[1] https://www.il3.ub.edu

[2] https://seeds4c.org/LinuxDataScience25

[3] https://www.slideshare.net/slideshow/gnu-linux-introduction-and-administration-aed1/277441886

[4] https://analisi.transparenciacatalunya.cat/en/Medi-Ambient/Dades-meteorol-giques-de-la-XEMA/nzvn-apee

[5] https://posit.cloud/plans/free

[6] https://cdimage.ubuntu.com/lubuntu/releases/noble/release/

[7] https://www.virtualbox.org/wiki/Downloads

[8] http://cloud.seeds4c.org/data_smc.csv.bz2

[9] https://davidadrian.cc/definitive-data-scientist-setup/