

DATA SCIENCE TOOLS

DATA SCIENCE TOOLS

LEARNING OBJECTIVES

- Identify the data science toolkit
- Navigate Git and the Command Line

COURSE

PRE-WORK

PRE-WORK REVIEW

- Complete the General Assembly [Command Line Tutorial](#)
- Set up GitHub account

OPENING

DATA SCIENCE TOOLS

LET'S DISCUSS THE CURRENT LESSON OBEJCTIVES

- Identify the data science toolkit
- Navigate Git and the Command Line

INTRODUCTION

TOOLS OF THE TRADE

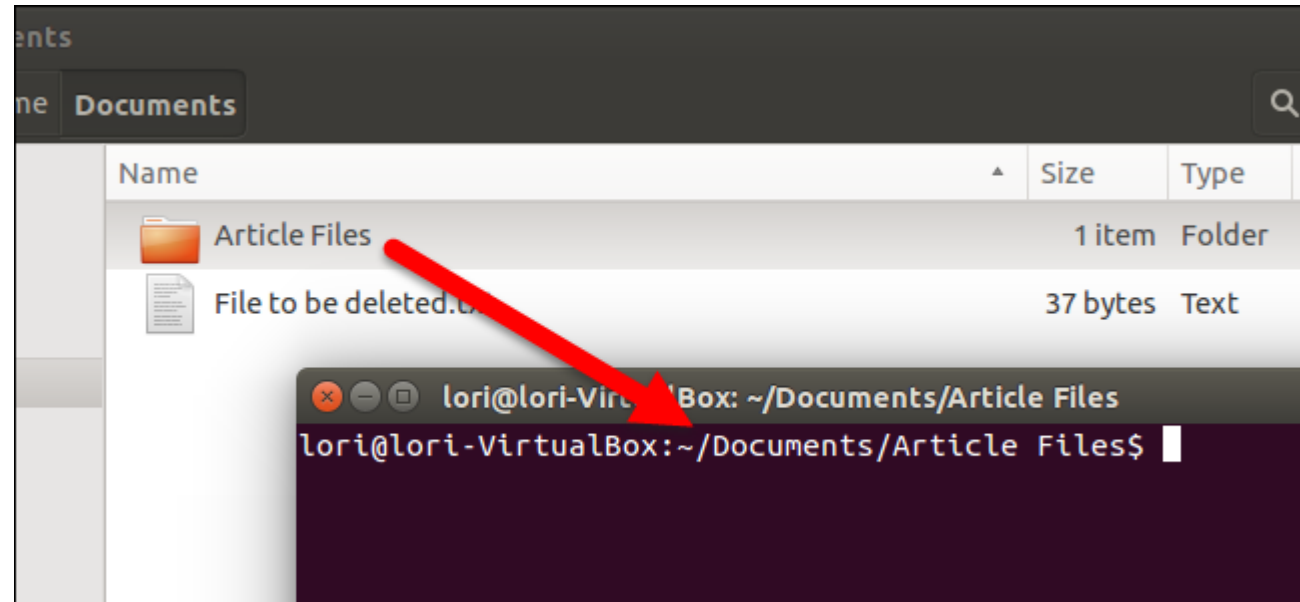
TOOLS OF THE TRADE

- Today we are going to review some of the tools we use in data science.
- We'll see how they fit into the wider programming environment.
- We'll start with the command line. This is your portal to your computer and the outside world.

LOCAL MACHINE

- ▶ On your local computer, you have a variety of tools at your disposal.

- ▶ Text editor
- ▶ Programs/tools
- ▶ Your files



- ▶ All of these can be accessed through the terminal or through a GUI (Graphical User Interface).
- ▶ You can navigate your files through the terminal or through Finder.

DATA SCIENCE TOOLS

Outside World
Local Machine

Terminal/
Command Line

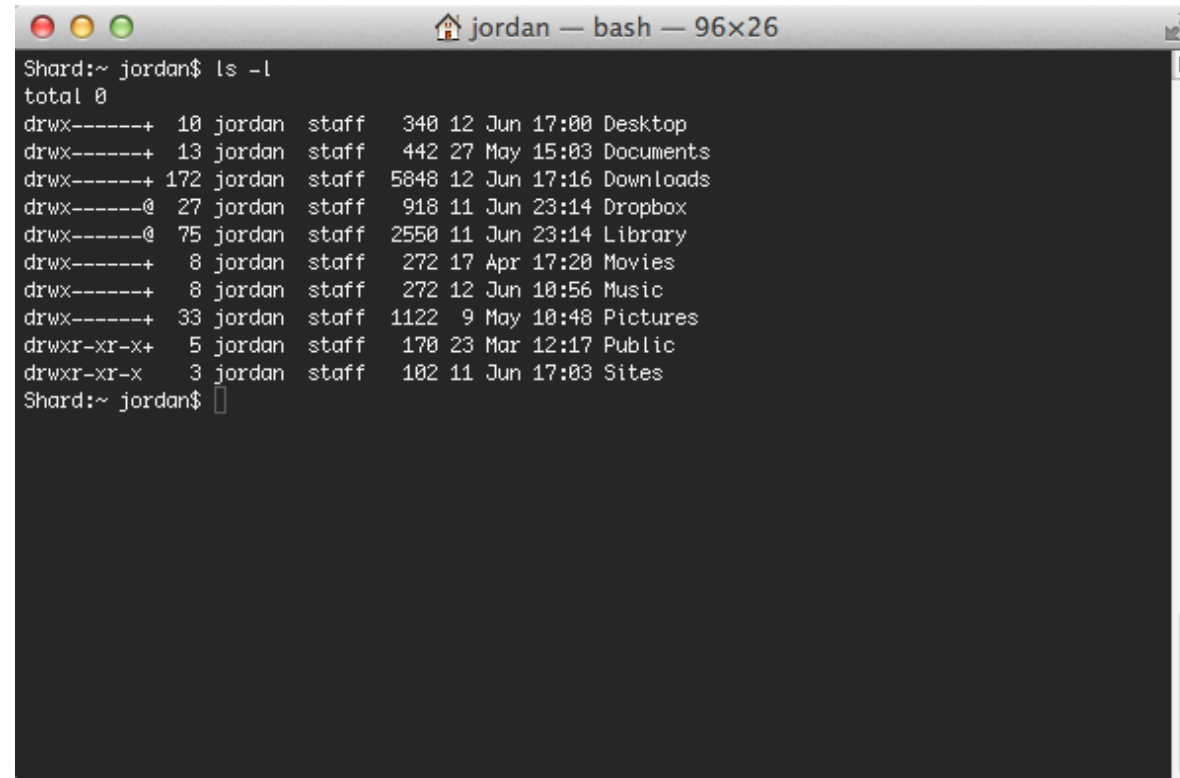
DEMO

COMMAND LINE

COMMAND LINE

- Let's walk through a few commands.

- cd
- pwd
- \$home
- mkdir
- open

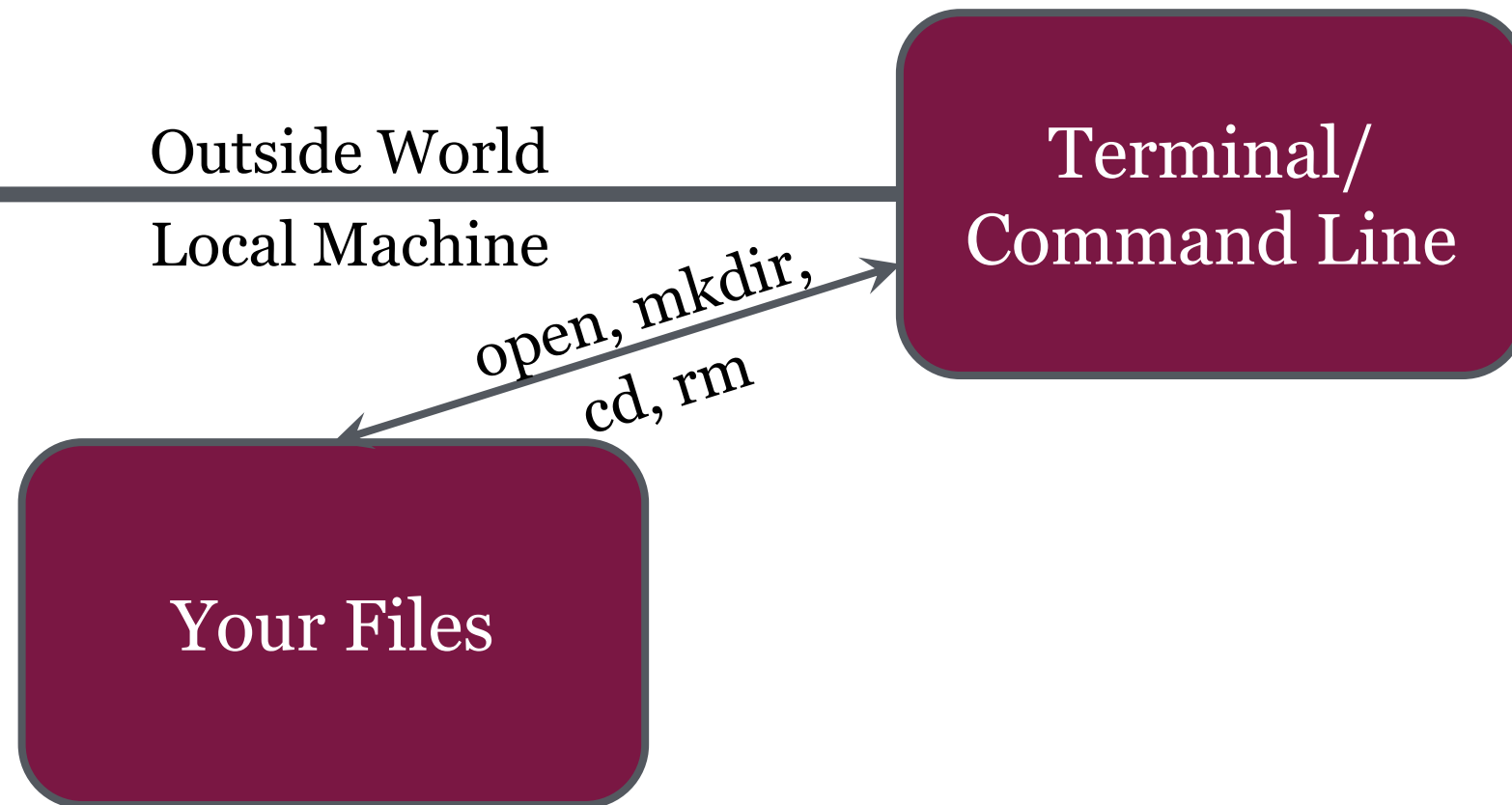


A terminal window titled 'jordan — bash — 96x26' showing the output of the 'ls -l' command. The output lists files and directories in the user's home directory with their permissions, owner, group, size, and modification date.

```
Shard:~ jordan$ ls -l
total 0
drwx-----+ 10 jordan  staff   340 12 Jun 17:00 Desktop
drwx-----+ 13 jordan  staff   442 27 May 15:03 Documents
drwx-----+ 172 jordan  staff  5848 12 Jun 17:16 Downloads
drwx-----@ 27 jordan  staff   918 11 Jun 23:14 Dropbox
drwx-----@ 75 jordan  staff  2550 11 Jun 23:14 Library
drwx-----+ 8 jordan   staff   272 17 Apr 17:20 Movies
drwx-----+ 8 jordan   staff   272 12 Jun 10:56 Music
drwx-----+ 33 jordan   staff  1122 9 May 10:48 Pictures
drwxr-xr-x+ 5 jordan   staff   170 23 Mar 12:17 Public
drwxr-xr-x 3 jordan   staff   102 11 Jun 17:03 Sites
Shard:~ jordan$
```

- We can access many tools with the terminal. Let's walk through a few.

DATA SCIENCE TOOLS



INTRODUCTION

TEXT EDITORS

TEXT EDITORS

- So far, we've used iPython Notebooks in place of a text editor.
- However, there are many options available

- eMacs

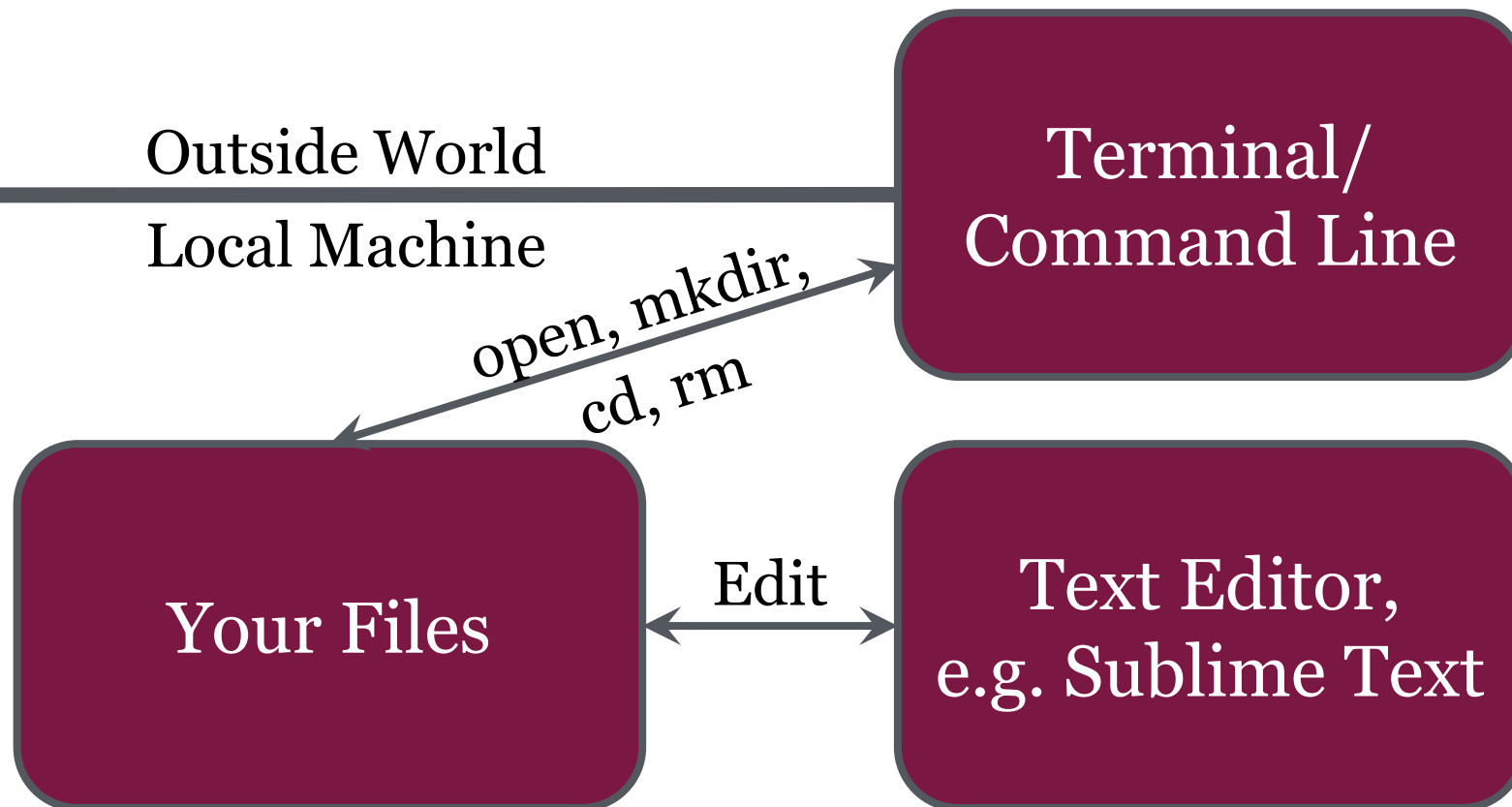
- Vim

- Sublime Text



- Let's see what Sublime Text look like with Python.

DATA SCIENCE TOOLS



INTRODUCTION

IPYTHON NOTEBOOK

INTRODUCTION

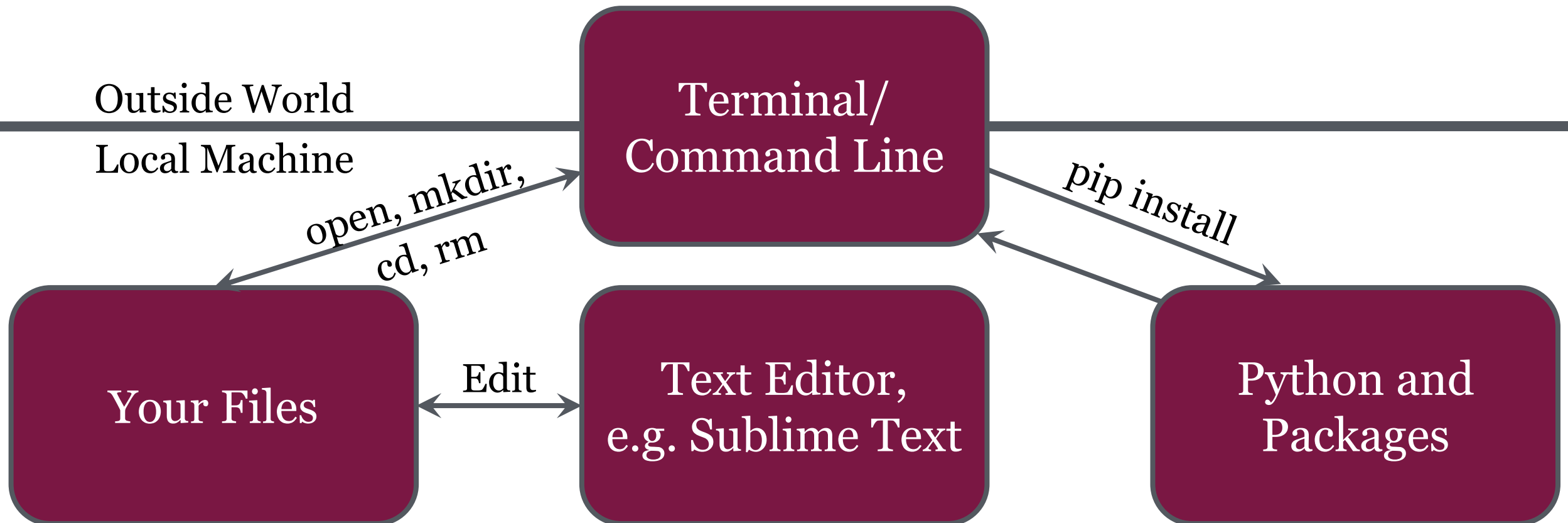
PYTHON PACKAGES

PYTHON PACKAGES

- The terminal allows us to run programs and reach out to the outside world.
- We can add programs and packages as needed.
- To add Python packages, we use a tool called *pip* a package manager which stands for “Pip Installs Python “
- Let's `pip install` a package with the command line. We'll install Beautiful Soup, a HTML/XML parsing package.

```
pip install beautifulsoup4
```

DATA SCIENCE TOOLS



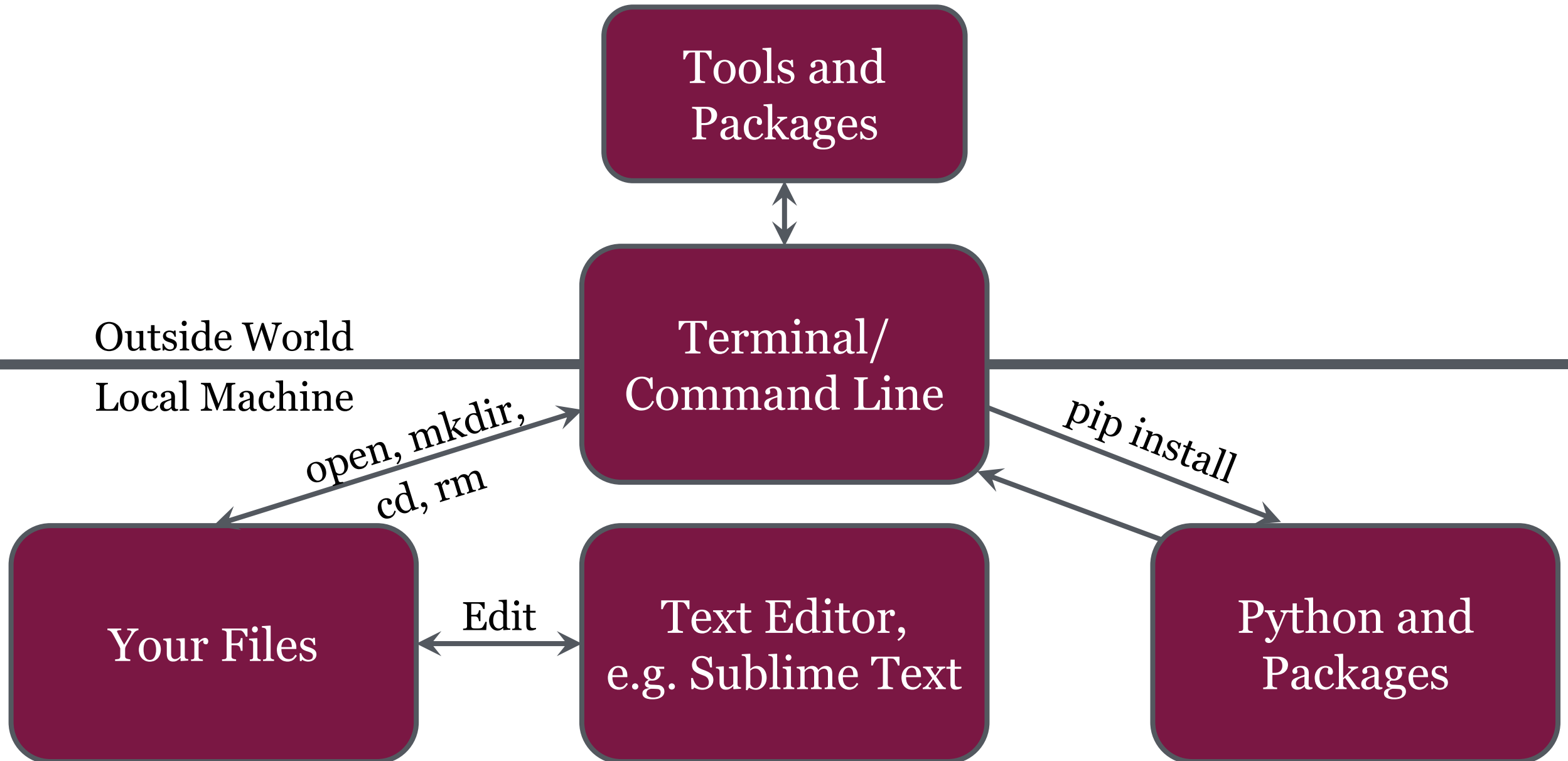
INTRODUCTION

THE OUTSIDE WORLD

THE OUTSIDE WORLD

- The command line also allows you to download and use other tools and packages.
- There are many tools for different purposes available in the outside world.

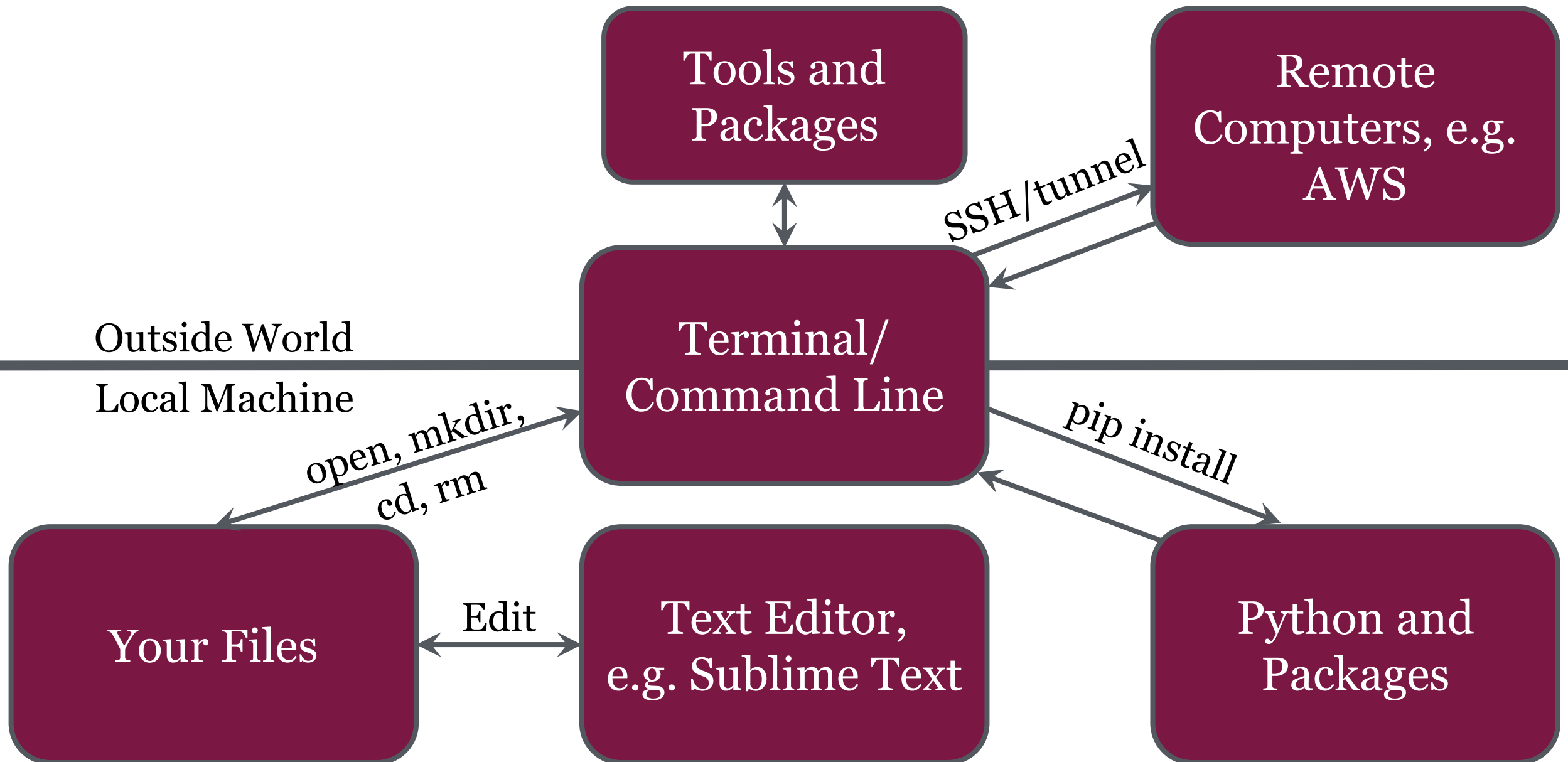
DATA SCIENCE TOOLS



THE OUTSIDE WORLD

- As we saw with pip, the command line can connect us to the outside world. This becomes more important for data.
- We may have Health Insurance Portability and Accountability Act (HIPAA) protected data. This means we can't leave this sensitive data on our local machine (i.e. laptop).
- We need to communicate with a *remote* machine (i.e. server) to access the data via command line.

DATA SCIENCE TOOLS



INTRODUCTION

GIT

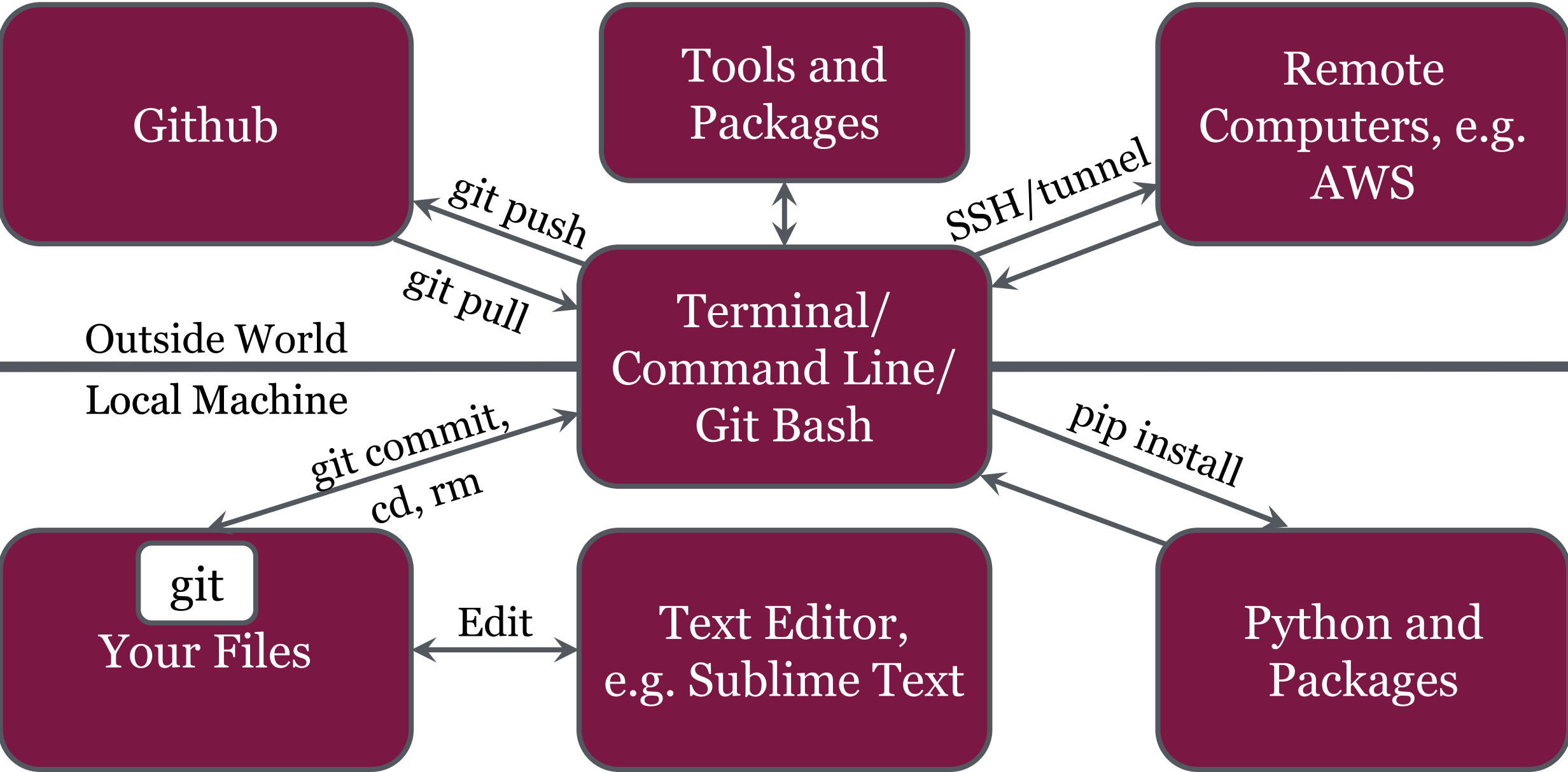
GIT

- Version control is necessary when working on complex projects.
- Git is a way of tracking changes we've made to our programs that allows us to go back in time to fix errors.
- Combined with Github, Git is a powerful tool for collaborating with colleagues. You can work on different aspects of projects simultaneously and merge the changes together seamlessly.
- There are many different ways to use these tools.

GIT

- Let's see an example of using Git and Github.
- There are three primary commands we'll use.
 - `git add`
 - `git commit`
 - `git push`
- When a colleague wants to implement our change, we may use the command `git pull`.

DATA SCIENCE TOOLS



ACTIVITY: KNOWLEDGE CHECK

ANSWER THE FOLLOWING QUESTIONS

1. What is a GUI?
2. What is the command line?
3. What are the big advantages of using the command line over a GUI?



EXERCISE

DELIVERABLE

Answers to the above questions

GUIDED PRACTICE

COMMAND LINE & GIT

Open: Your-Development-Environment.ipynb

CONCLUSION

TOPIC REVIEW

REVIEW

- What are some common data science tools?
- Why are these tools useful?
- Any other questions?

COURSE

BEFORE NEXT CLASS

LESSON

Q & A

LESSON

EXIT TICKET

DON'T FORGET TO FILL OUT YOUR EXIT TICKET