INSTRUCTOR NOTES

▸ Insert Text Here

MATERIALS

▸ Insert Text Here

PRE-WORK

▸ Insert Text Here

DATA SCIENCE TOOLS

Chris Connell

PRE-WORK

PRE-WORK REVIEW

▸ Explain the difference between variance and bias

▸ Use descriptive stats to understand your data

DATA SCIENCE TOOLS

LET'S DISCUSS THE CURRENT LESSON OBEJCTIVES

‣ Understanding of Command line

‣ Build your own git repo

‣ Understand how a scraper works

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

DATA SCIENCE TOOLS
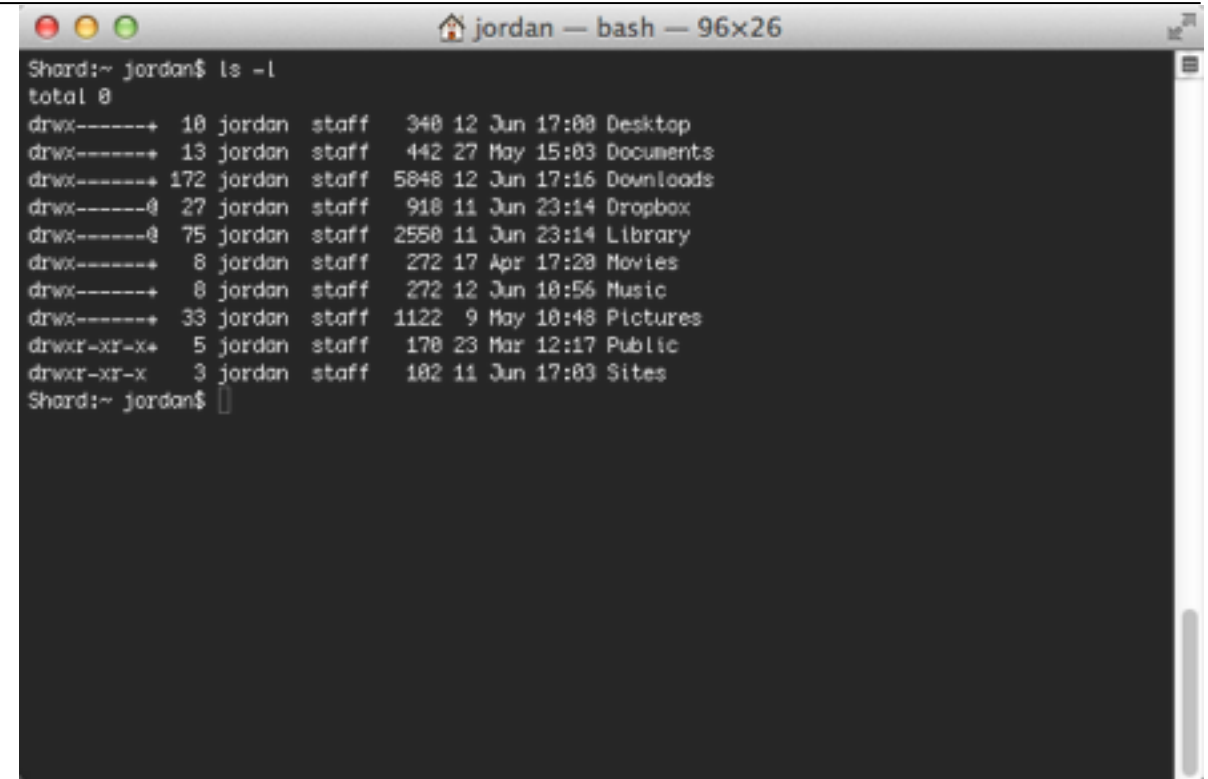
Outside World

Local Machine

Terminal/
Command Line

COMMAND LINE

COMMAND LINE

‣Most common commands.

    ‣cd

    ‣pwd

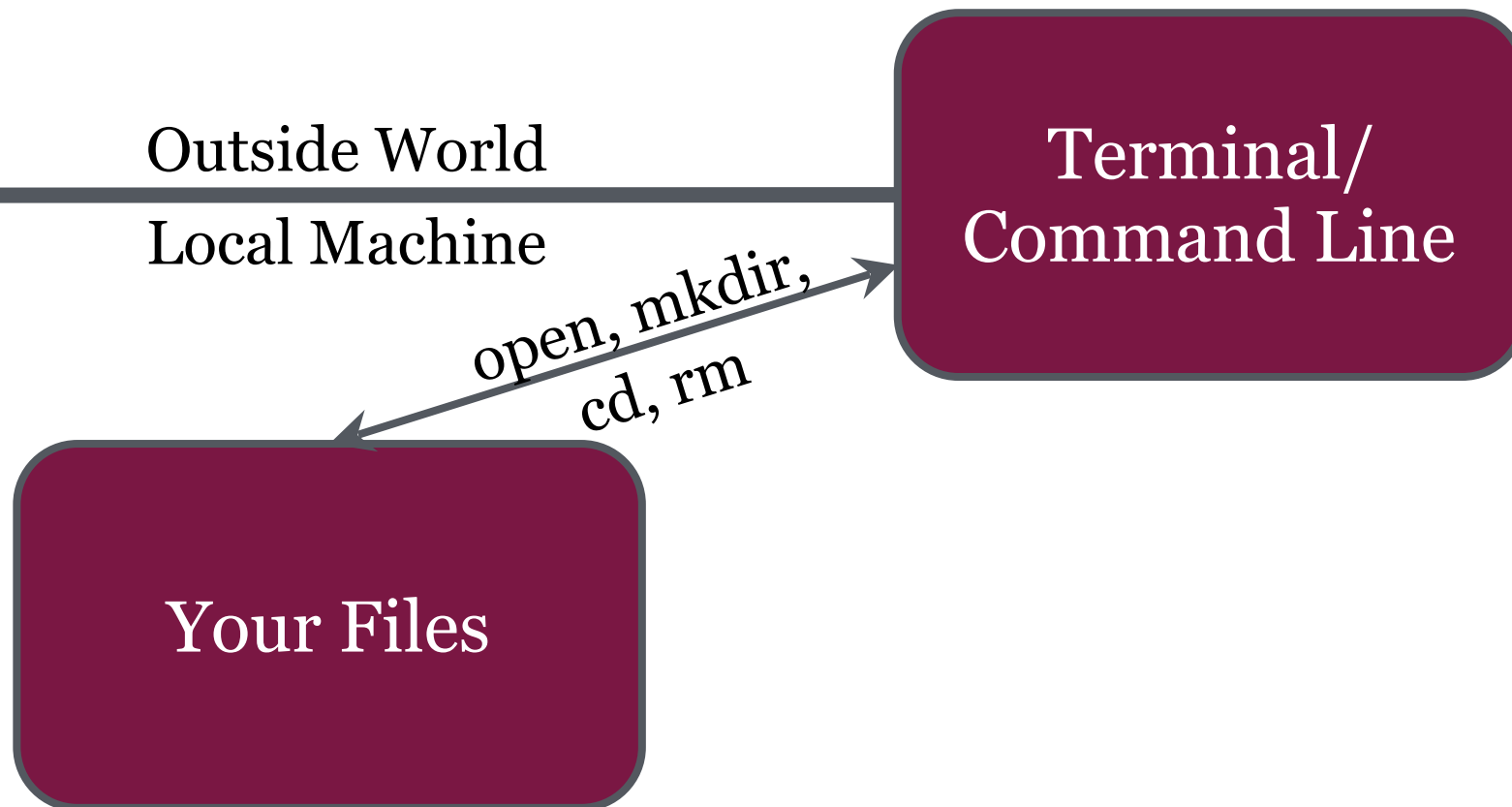   ‣mkdir

    ‣rm

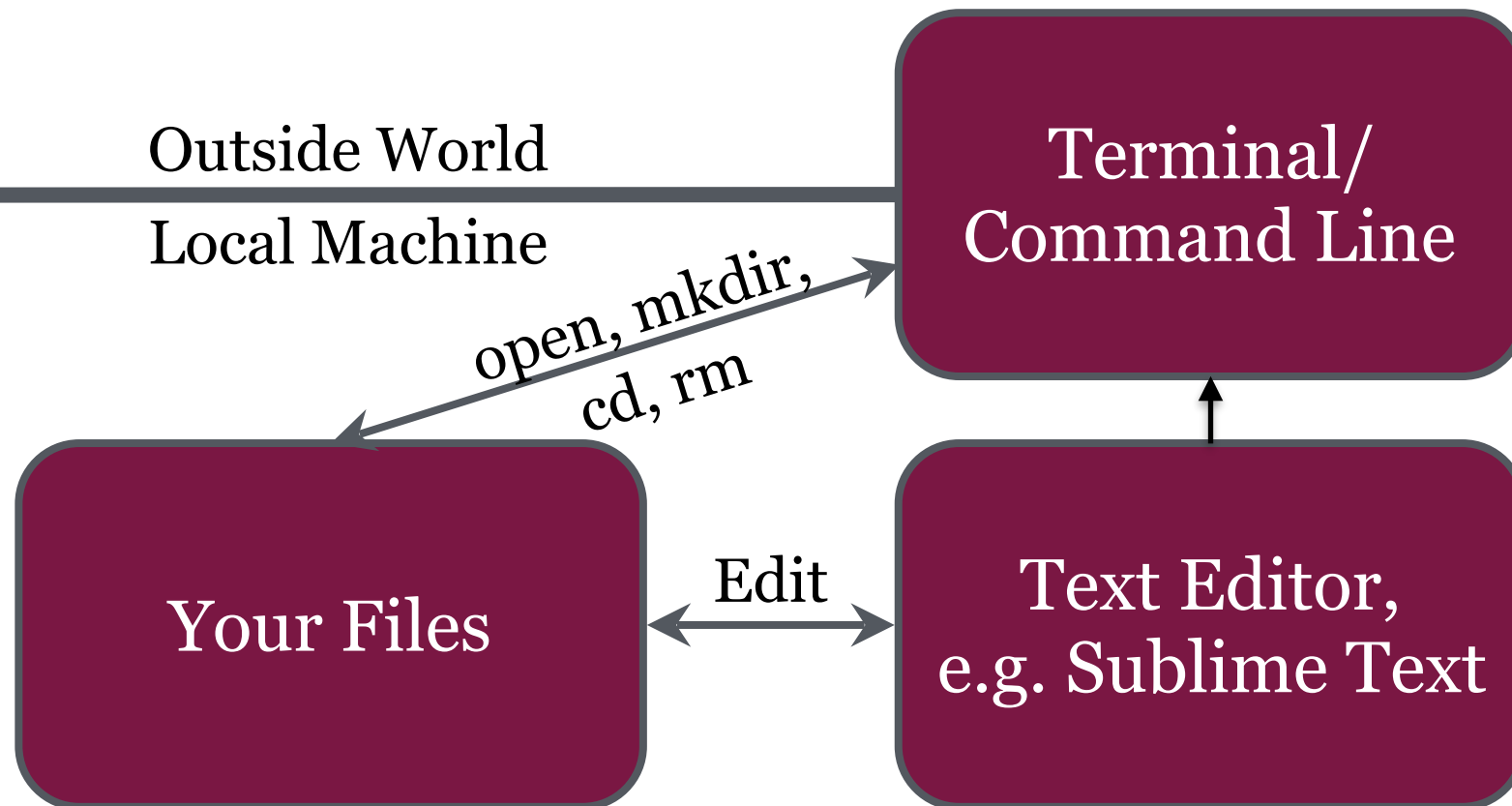   ‣find

    ‣sudo

    ‣cp

    ‣mv

COMMAND LINE

Action	Microsoft Windows (cmd.exe)	Microsoft Windows (Windows PowerShell)	Unix/Linux/Mac OS X (bash)	Unix/Linux/Mac OS X (POSIX shell)
Scroll through History of typed commands	↑/↓	↑/↓	↑/↓ or Ctrl+P / Ctrl+N	ESC K / ESC J
Signal end-of-file	Ctrl+Z		Ctrl+D	
Abort current command/typing	Ctrl+C	Ctrl+C	Ctrl+C	Ctrl+C
Erase word to the left	Ctrl+← Backspace		Ctrl+W or Esc+← Backspace	
Erase word to the right	Ctrl+Delete		Alt+D	
Erase line to the left	Ctrl+Home	Ctrl+Home	Ctrl+U	
Erase line to the right	Ctrl+End	Ctrl+End	Ctrl+K	
Yank/paste previously erased string			Ctrl+Y	
Move one word to the left (backward)	Ctrl+←	Alt+←	Alt+B	
Move one word to the right (forward)	Ctrl+→	Alt+→	Alt+F	
Move to beginning of line	Home	Home	Ctrl+A or Home	
Move to end of line	End	End	Ctrl+E or End	
Reverse search of history	F8	F8	Ctrl+R	
Stop execution of the current job			Ctrl+Z	
Insert the next character typed verbatim			Ctrl+V	
Autocomplete command/file name	Tab ⇆ (enabled by default in Windows XP and later)	Tab ⇆	Tab ⇆ (usually once)	Esc (usually twice)
Paste contents of clipboard at cursor	Alt+Space then E then P		⇧ Shift+Ins	
Scroll window up	⇧ Shift+PageUp (may not work in some versions of Windows XP)		⇧ Shift+PageUp	
Scroll window down	⇧ Shift+PageDown (may not work in some versions of Windows XP)		⇧ Shift+PageDown	
Clear screen			Ctrl+L	

DATA SCIENCE TOOLS

DATA SCIENCE TOOLS

ACTIVITY: KNOWLEDGE CHECK

ANSWER THE FOLLOWING QUESTIONS

EXERCISE

1. What command line command is used to force an action?
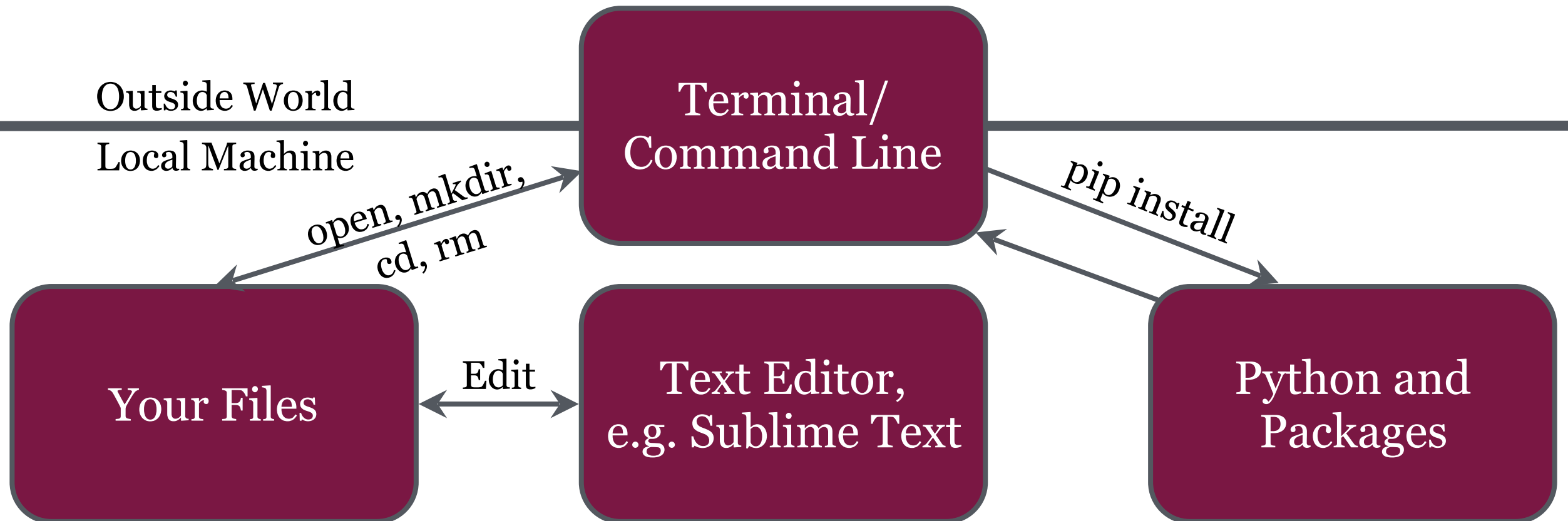
DELIVERABLE

Answers to the above questions

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*.

▸Let's `pip install` a package with the command line. We'll install Beautiful Soup, a HTML/XML parsing package.

```
pip install beautifulsoup4
```
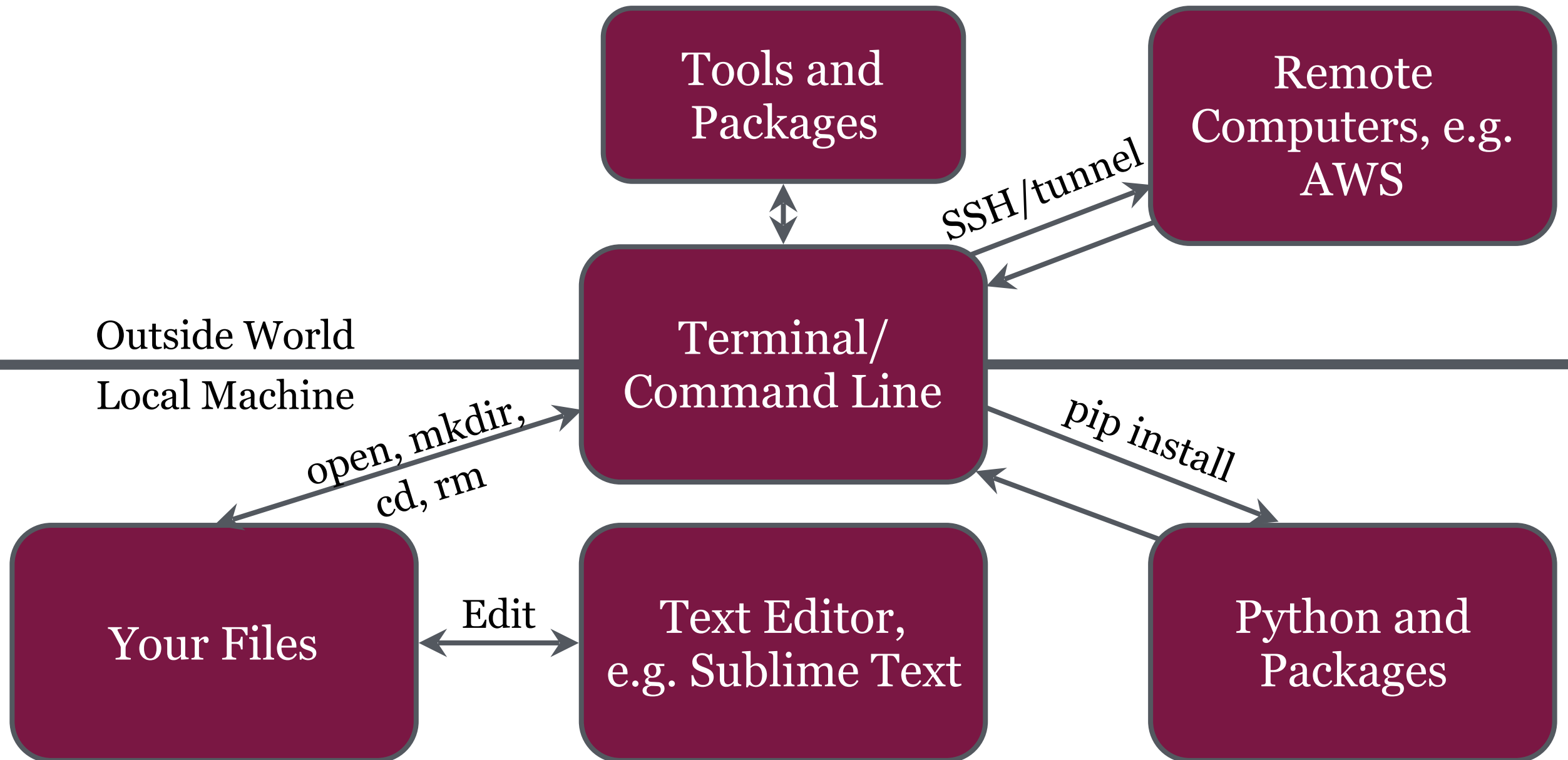
DATA SCIENCE TOOLS

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.

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 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.

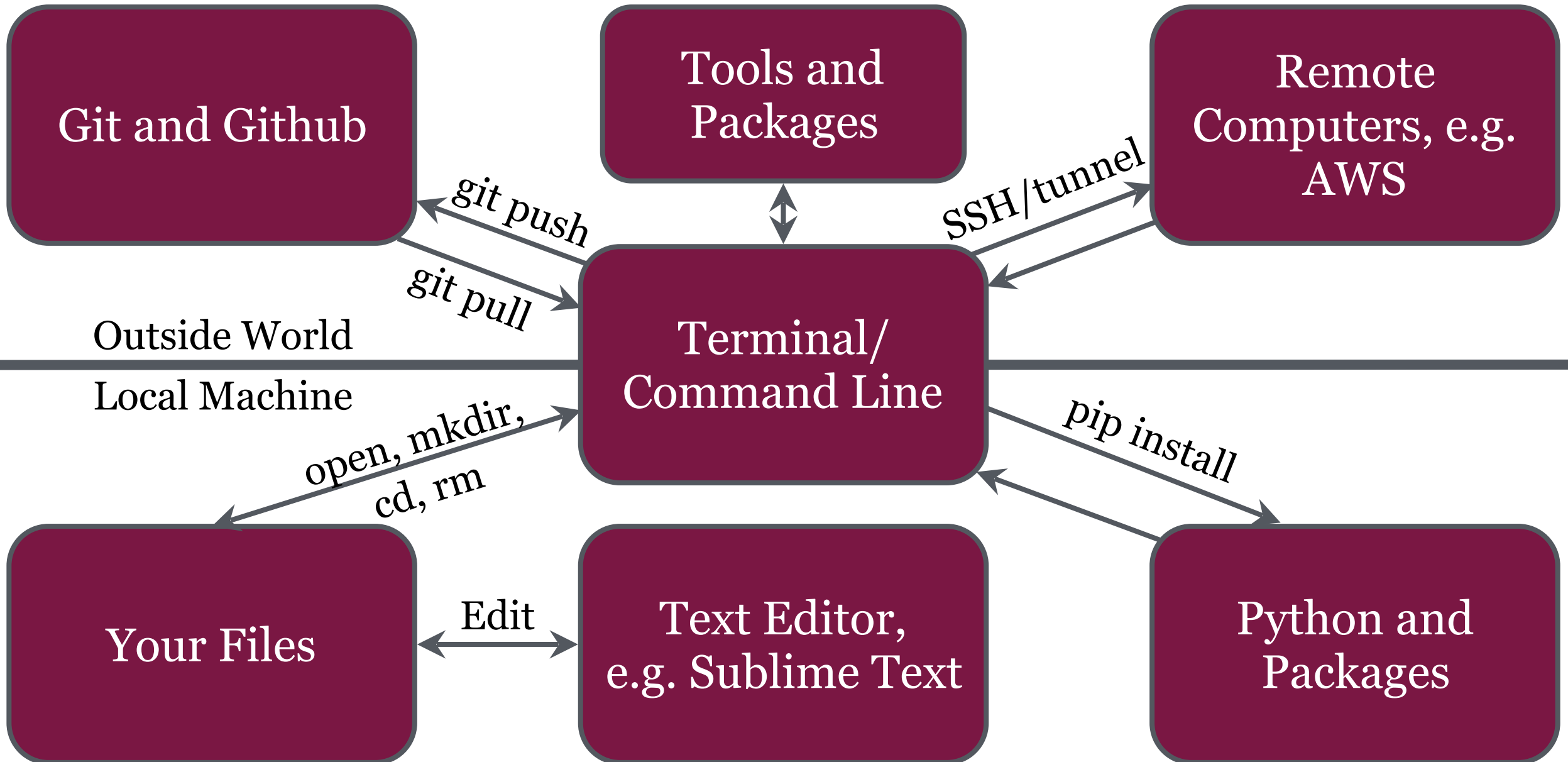▸ Let's see a demonstration of this.

DATA SCIENCE TOOLS

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

EXERCISE

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

DELIVERABLE

Answers to the above questions

GIT AND COMMAND LINE

ACTIVITY: GIT AND COMMAND LINE

EXERCISE

DIRECTIONS (20 minutes)

1. Let's review the exercises from Codecademy Python.
2. Let's review the exercises from the GA command line tutorial.
3. Are there any questions?

DELIVERABLE

Questions

TOPIC REVIEW

REVIEW

‣ What are some common data science tools?

‣ Why are these tools useful?

‣ Any other questions?

BEFORE NEXT CLASS

DUE DATE

‣ Project:

CREDITS

CITATIONS

▸ Title, Author: link
▸ Title, Author: link
▸ Title, Author: link

Q & A

EXIT TICKET

DON'T FORGET TO FILL OUT YOUR EXIT TICKET

THANKS!

NAME

‣ Optional Information:

‣ Email?

‣ Website?

‣ Twitter?