



香港浸會大學
HONG KONG BAPTIST UNIVERSITY

Tools Installation

JOUR7280/COMM7780

Big Data Analytics for Media and Communication

Instructor: Dr. Xiaoyi Fu

Getting started: Tools

- Talking to your computer: Command line interface (CLI)
- Text Editor: vscode, sublime, notepad++, or others
- Platform for publishing and socializing: Git and GitHub, and Markdown language
- The tool: Python 3.x (Anaconda 3) and Jupyter Notebook

Command Line Interface (CLI) basics

- Open CLI:
 - MacOS: spotlight search – “terminal”
 - Windows: start menu search - “cmd”
- An interesting tutorial ([macOS](#), [Windows](#))

Command Line Interface (CLI) basics

- Basic commands of CLI

- MacOS

- - pwd
- - date
- - mkdir
- - echo
- - ls
- - cd
- - cp
- - rm
- - mv

- Make 10 folders at once: `mkdir -p test{1..10}/{1,2,3}`

- Windows

- - echo %cd%
- - date
- - mkdir
- - echo
- - dir
- - cd
- - copy
- - rmdir
- - move

Git

- Git is a “version control system”: records changes to a file or set of files over time and users can recall specific versions later.
 - Useful for collaboration project
- Download link ([mac](#), [windows](#))
- Setting up your Git account
 - `git config --global user.email "your@email.com"`
 - `git config --global user.name "your name"`

GitHub

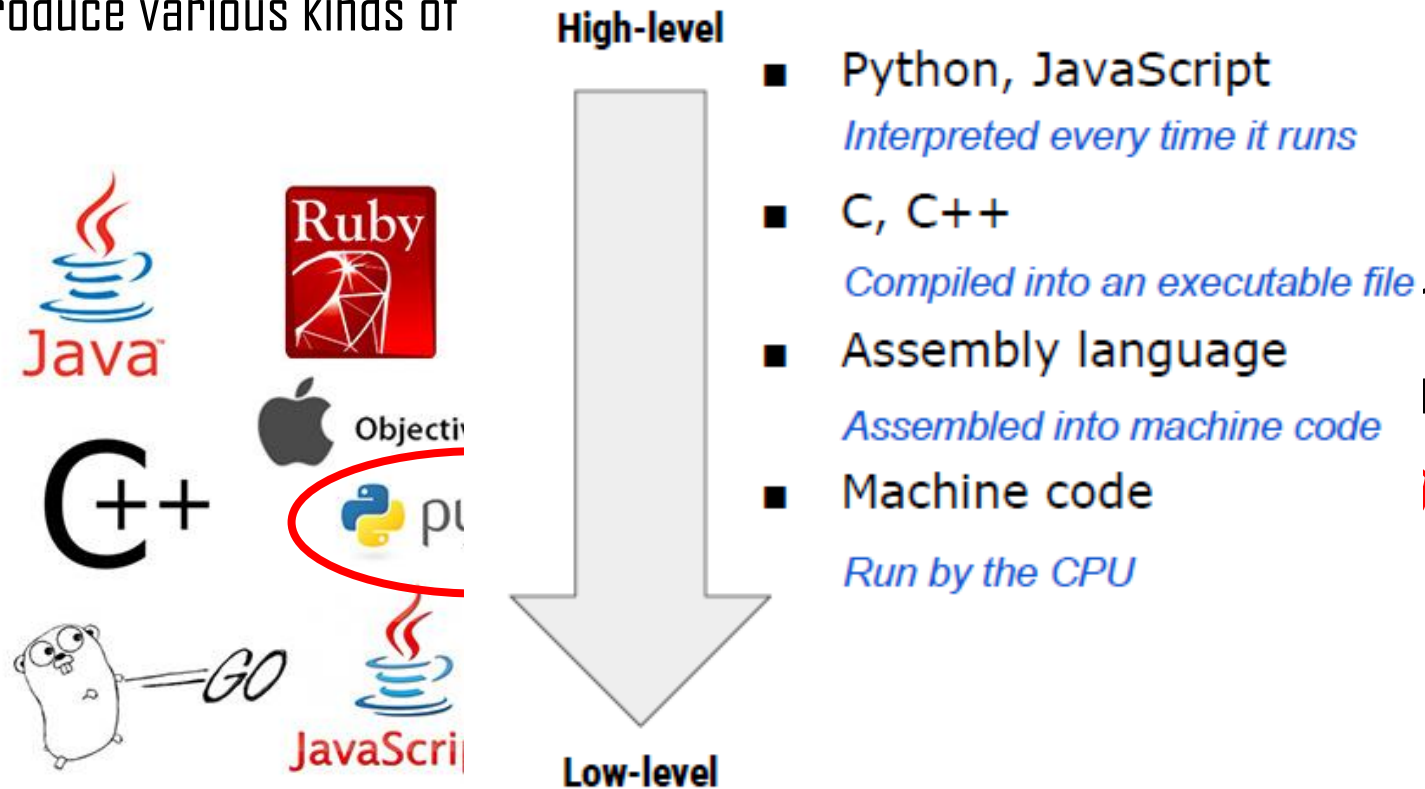
- GitHub is a web-based hosting service for software development projects that use Git version control system." ([What is GitHub](#))
- Push and pull
- Public/Private repositories
- The social aspects of GitHub (share, fork, star)
 - A profile that shows your portfolio

Markdown language

- Markdown is a lightweight markup language with plain text formatting syntax
- `readme.md`
- Markdown language quick guide [[Link](#)]
- Work with markdown language
 - vscode Preview, jupyter notebook
 - GIT MD Syntax [[Link](#)]
 - GIT Deeper MD Syntax [[Link](#)]
 - GIT MD Emojis! [[Link](#)]

Programming Language

- A programming language is a formal language, which comprises a set of instructions used to produce various kinds of



n programming language refers to *high-level* *ies*, such as C, C++, Java, Matlab...

Introduction to Python

- **Python** is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation.

Increasingly popular!

Aug 2018	Aug 2017	Change	Programming Language	Ratings	Change
1	1		Java	16.881%	+3.92%
2	2		C	14.966%	+8.49%
3	3		C++	7.471%	+1.92%
4	5	⬆	Python	6.992%	+3.30%
5	6	⬆	Visual Basic .NET	4.762%	+2.19%
6	4	⬇	C#	3.541%	-0.65%
7	7		PHP	2.925%	+0.63%
8	8		JavaScript	2.411%	+0.31%
9	-	⬆	SQL	2.316%	+2.32%
10	14	⬆	Assembly language	1.409%	-0.40%

Jan 2020	Jan 2019	Change	Programming Language	Ratings	Change
1	1		Java	16.896%	-0.01%
2	2		C	15.773%	+2.44%
3	3		Python	9.704%	+1.41%
4	4		C++	5.574%	-2.58%
5	7	⬆	C#	5.349%	+2.07%
6	5	⬇	Visual Basic .NET	5.287%	-1.17%
7	6	⬇	JavaScript	2.451%	-0.85%
8	8		PHP	2.405%	-0.28%
9	15	⬆	Swift	1.795%	+0.61%
10	9	⬇	SQL	1.504%	-0.77%

TIOBE Rankings

Why popular & Features

- Code readability, shorter codes, ease of writing
 - fewer lines of code in comparison to languages such as C++ or Java.
- "Simplicity is the best"
 - Closer to English language; Easy to learn
 - More emphasis on the solution to the problem rather than the syntax
- Interpreted language
 - Directly run the program from the source code.
 - No separate compilation and execution steps like C and C++.
- Rich Library Support
 - The Python Standard Library is very vast.

```
public class HelloWorld
{
    public static void main (String[] args)
    {
        System.out.println("Hello, world!");
    }
}
```

Java Code

```
print("Hello, world!") # Python version 3
```

Python Code

What Python can do



SciPy.org

Scientific Computing Tools for Python

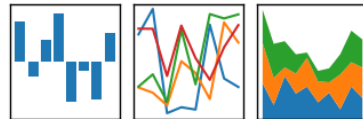
SciPy refers to several related but distinct entities:

- The *SciPy ecosystem*, a collection of open source software for scientific computing in Python.
- The *community* of people who use and develop this stack.
- Several *conferences* dedicated to scientific computing in Python - SciPy, EuroSciPy and SciPy.in.
- The *SciPy library*, one component of the SciPy stack, providing many numerical routines.

Scientific Computing

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



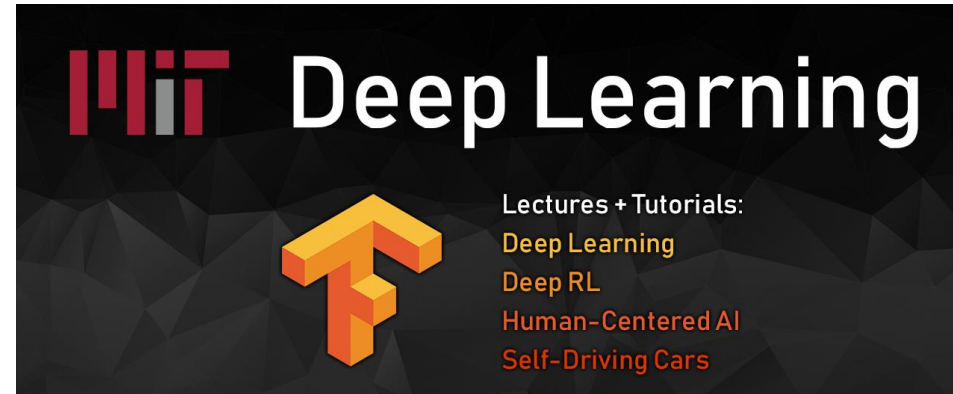
[home](#) // [about](#) // [get pandas](#) // [documentation](#) // [community](#) //

Python Data Analysis Library

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](#) programming language.

pandas is a [NumFOCUS](#) sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to [donate](#) to the project.

Data analysis



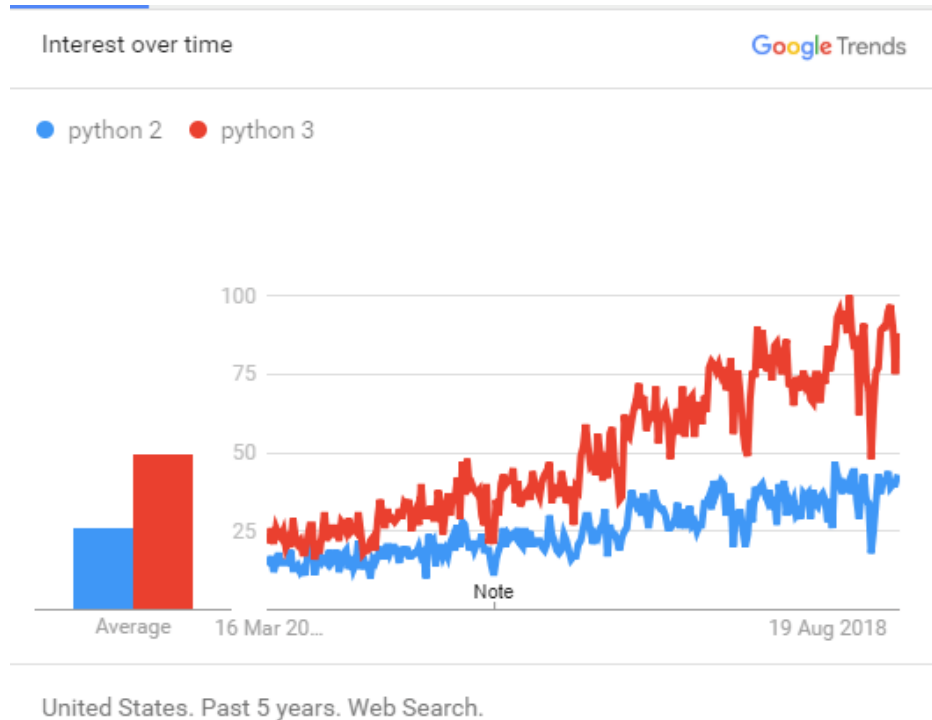
Machine Learning

matplotlib

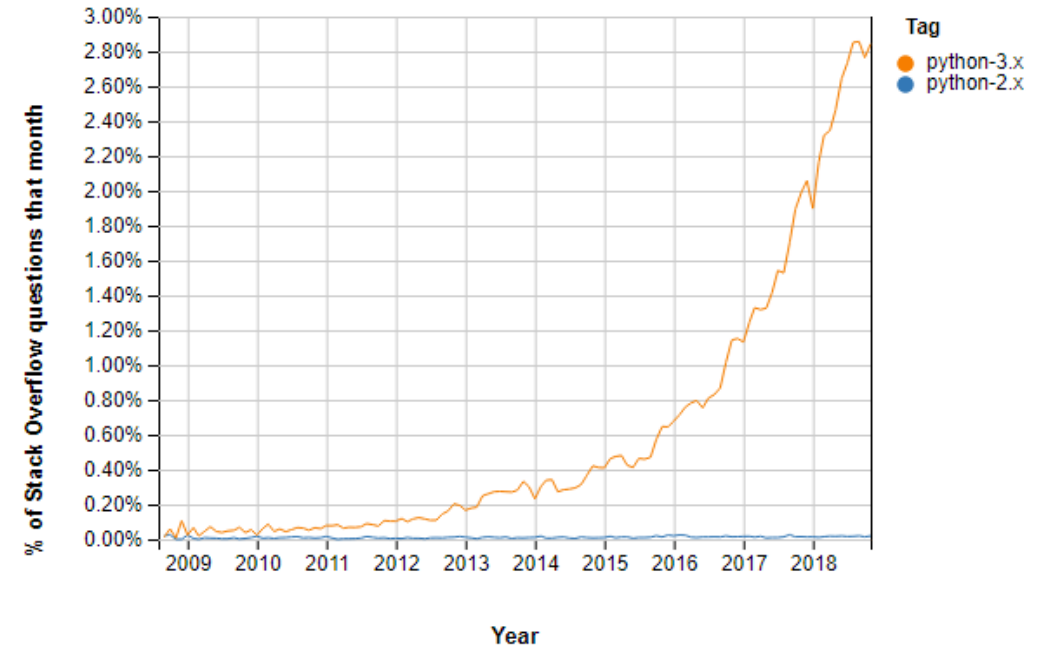
Version 3.0.2

Data Visualization

Python 2 vs 3






Google Trends Python 2 vs. Python 3



Stack Overflow Questions Python 2 vs. Python 3

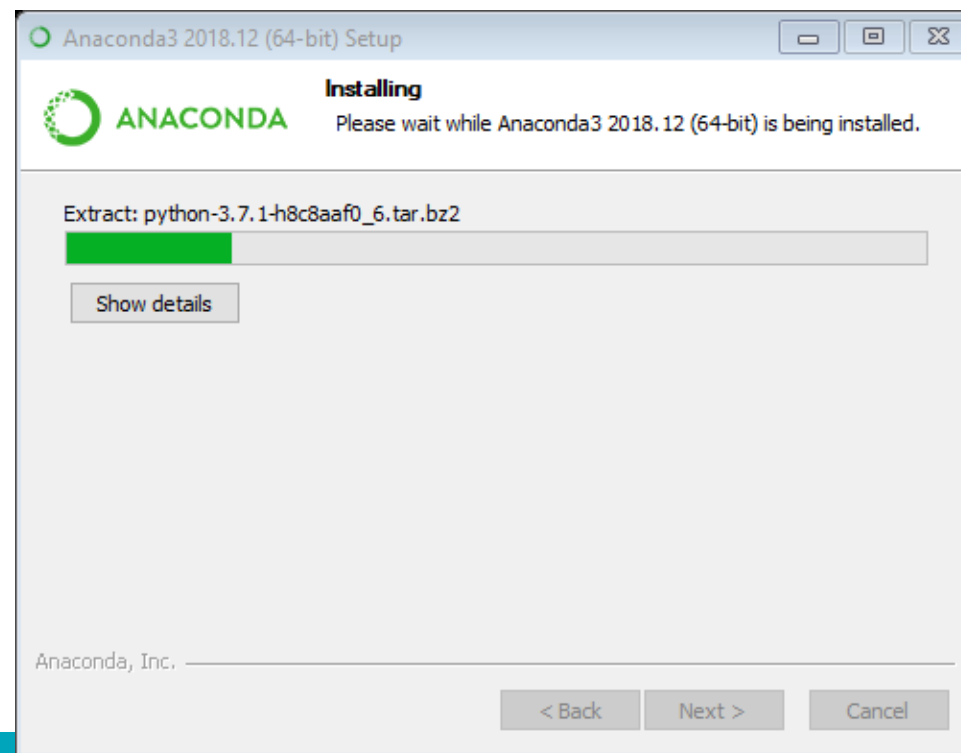
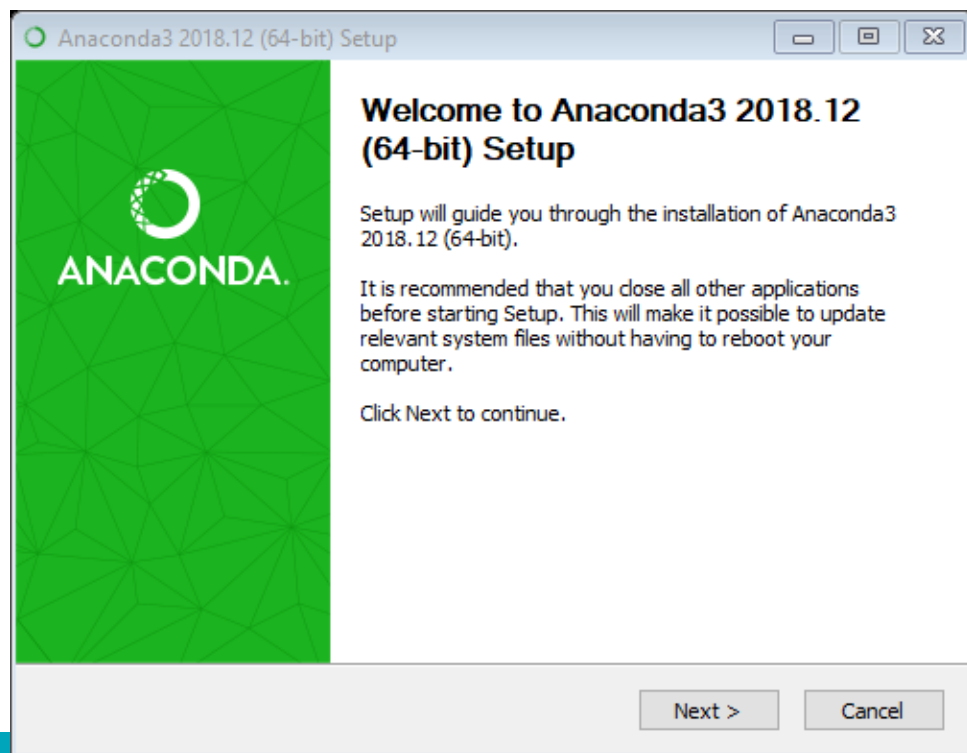
Installing Python and Jupyter Notebook

Anaconda Installers		
Windows 	MacOS 	Linux 
Python 3.8	Python 3.8	Python 3.8
64-Bit Graphical Installer (466 MB)	64-Bit Graphical Installer (462 MB)	64-Bit (x86) Installer (550 MB)
32-Bit Graphical Installer (397 MB)	64-Bit Command Line Installer (454 MB)	64-Bit (Power8 and Power9) Installer (290 MB)

- Python 3.8
 - Anaconda [\[Link\]](#)
 - A free and open-source distribution of the Python and R programming languages
- Jupyter Notebook
 - Included in the Anaconda installation

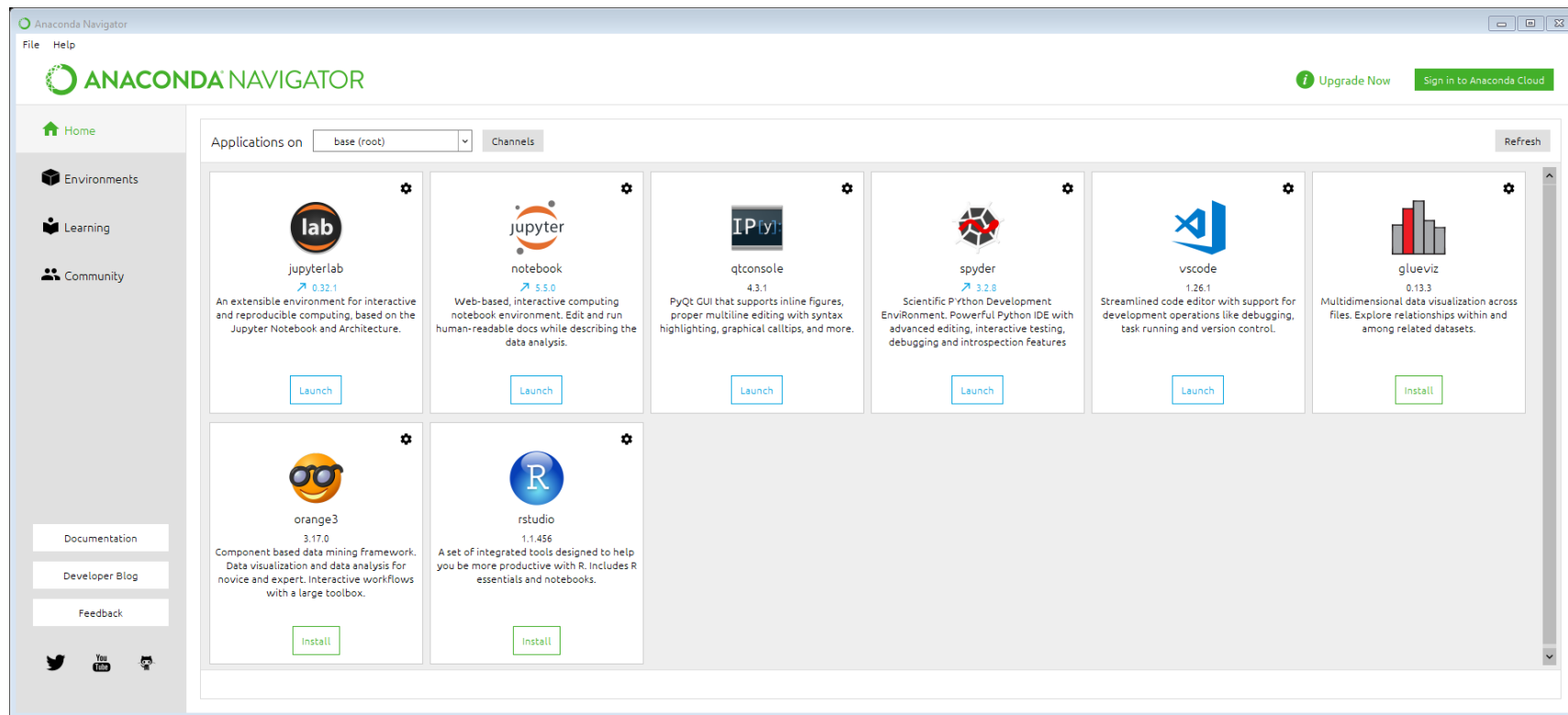
Install Anaconda

- Run the Anaconda installer.
 - PS: turn off anti-virus software, especially 360 Safeguard (360安全卫士) before the installation
 - PS: if you have installed Python individually before, remove it first



Anaconda

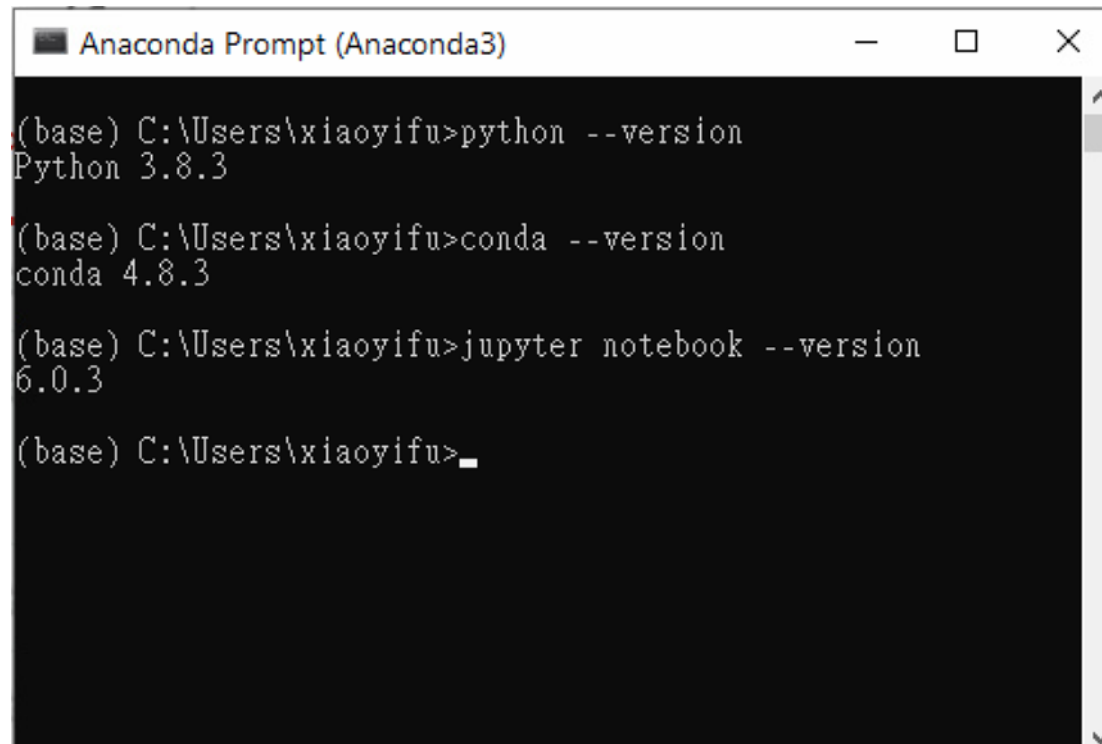
- Open the *Anaconda Navigator* after installation. You can install and launch different environment for later development.



Check your existing versions

```
In [2]: import sys  
print("Python version: ", sys.version)
```

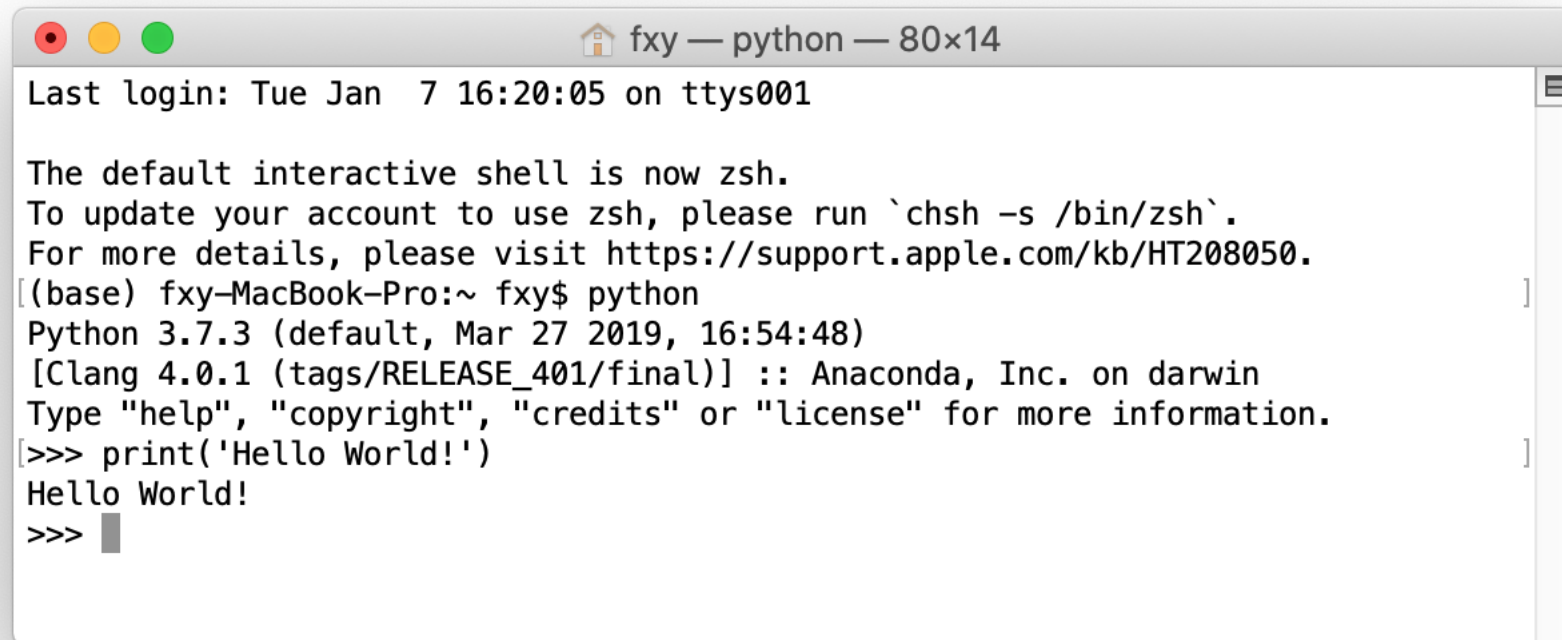
```
Python version: 3.8.5 (default, Aug 5 2020, 03:39:04)  
[Clang 10.0.0 ]
```



```
Anaconda Prompt (Anaconda3)  
  
(base) C:\Users\xiaoyifu>python --version  
Python 3.8.3  
  
(base) C:\Users\xiaoyifu>conda --version  
conda 4.8.3  
  
(base) C:\Users\xiaoyifu>jupyter notebook --version  
6.0.3  
  
(base) C:\Users\xiaoyifu>_
```


Your first Python program

- Switch to English keyboard before coding!
- Two ways of printing "Hello World!"
 - via CLI

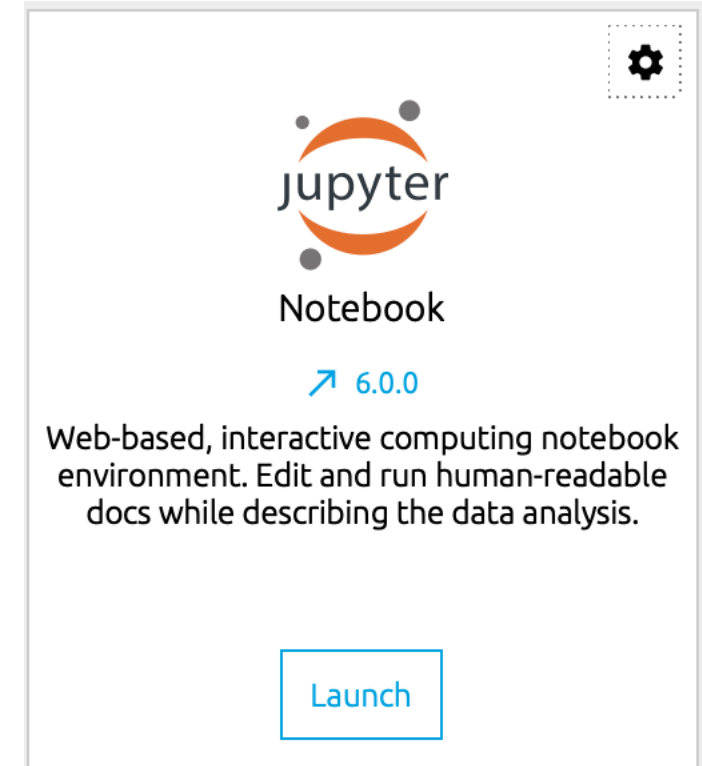


```
fxy — python — 80x14
Last login: Tue Jan  7 16:20:05 on ttys001

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
(base) fxy-MacBook-Pro:~ fxy$ python
Python 3.7.3 (default, Mar 27 2019, 16:54:48)
[Clang 4.0.1 (tags/RELEASE_401/final)] :: Anaconda, Inc. on darwin
Type "help", "copyright", "credits" or "license" for more information.
[>>> print('Hello World!')
Hello World!
>>> █
```

Jupyter Notebook

- Jupyter Notebook
 - Introduced in 2011. Inspired by scientific programs like Mathematica or Sage, the Notebook offers a modern and powerful web interface to Python.
 - Contain both code and rich text elements, such as figures, links, equations, ...
 - The ideal place to bring together an analysis description, and its results
 - We will use Jupyter Notebook in this course



Your first Python program

- Two ways of printing "Hello World!"
 - via Jupyter Notebook

The screenshot displays the Jupyter Notebook web interface. At the top, the Jupyter logo is on the left, and 'Quit' and 'Logout' buttons are on the right. Below this is a tab bar with 'Files', 'Running', and 'Clusters'. The 'Files' tab is active, showing a file browser for the '/ anaconda3' directory. A dropdown menu is open, showing options: 'Notebook: Python 3' and 'Other: Text File, Folder, Terminal'. The file browser lists directories: '..', 'Anaconda-Navigator.app', 'bin', and 'conda-meta'. Below the file browser, the Jupyter logo is followed by 'Untitled' and 'Last Checkpoint: 2 minutes ago (unsaved changes)'. On the right, there is a Python logo and a 'Logout' button. The main interface has a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. Below the menu bar is a toolbar with icons for saving, creating new, opening, saving as, undo, redo, run, and a code editor dropdown. The code editor shows a single cell with the code `In [1]: print('Hello World')` and the output `Hello World`. Below the code cell is an empty input field for the next cell, labeled `In []:`.

Exercise

- Install Anaconda & Jupyter notebook
- Try "Hello World!" program

References

- Anaconda installation FAQ [[Link](#)]
- Add Anaconda3 to path (mac) [[link](#)]
- Git tutorial [[link](#)]
- Change Jupyter notebook working directory [[link](#)]
- Learning website
 - <https://www.learnpython.org/>
 - <https://stackoverflow.com/>

This content is copyright protected and shall not be shared, uploaded or distributed.

Thank You

