



Chapter 2 Python for Data Science

CSS 341 Introduction to
Data Science

Chukiat Worasuchep

Important Notice

การเรียนการสอนหัวข้อนี้ ผ่านทางสื่อออนไลน์ (Online meeting)
และมีการบันทึกภาพและเสียงเพื่อประโยชน์ทางการศึกษาต่อไปในอนาคต.
หากท่านไม่ยินยอมให้มีการเผยแพร่การบันทึกดังกล่าว ขอให้แจ้งให้ผู้สอนทราบภายใน 36 ชั่วโมง.

Learning objectives

- เข้าใจการใช้งาน Python เพื่องาน data science
- หัดทดลองใช้ Jupyter notebook เพื่องาน data science
- ทบทวน data structures เช่น tuple, list, dictionary สำหรับงาน data science
- แนะนำ numpy และ pandas เบื้องต้น

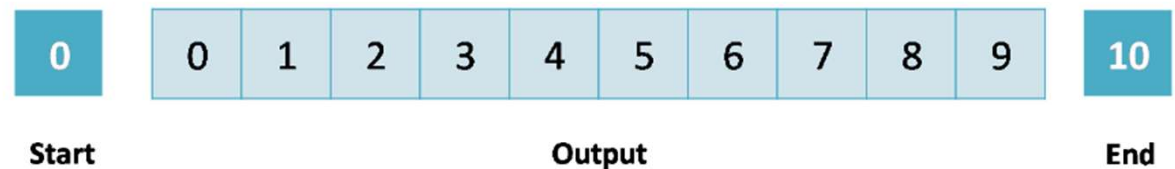
Outline

- ☐ Review basic Python
- ☐ Tools esp. Jupyter notebook
- ☐ Introduce numpy and pandas libraries

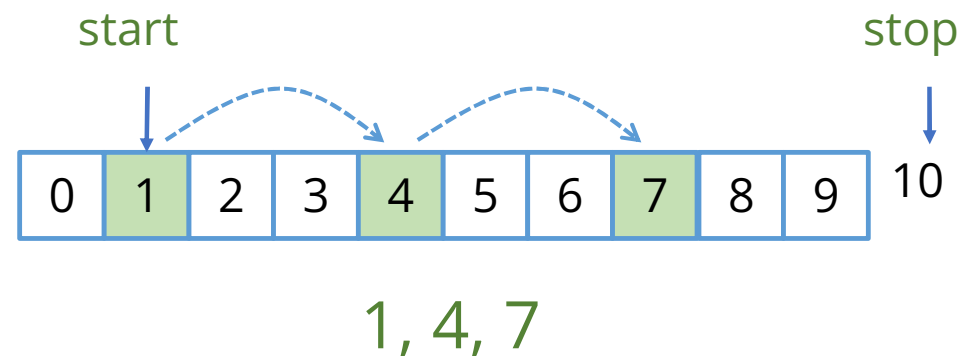
range()

```
range(start, stop, step_size)
```

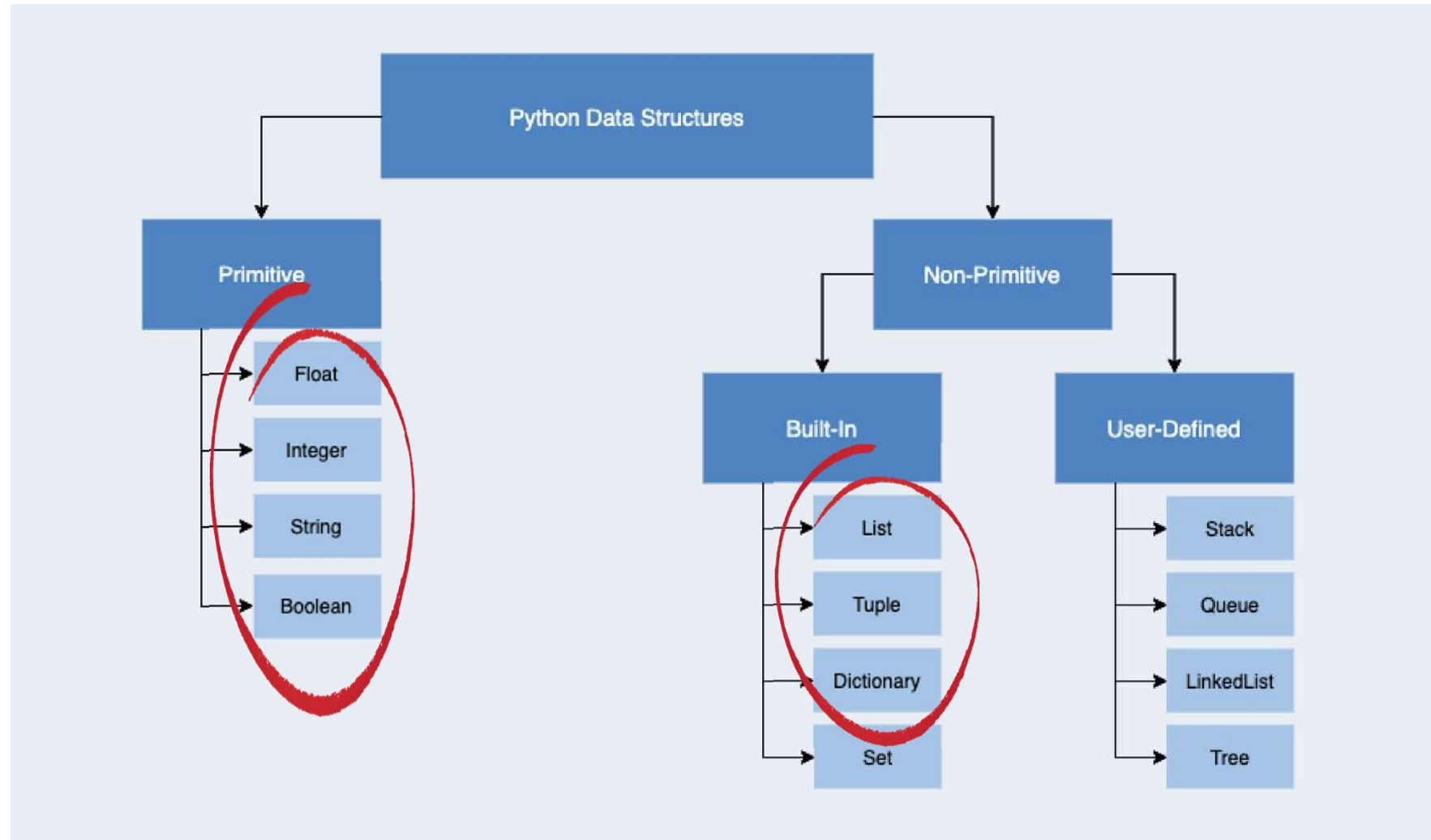
Python range(10)



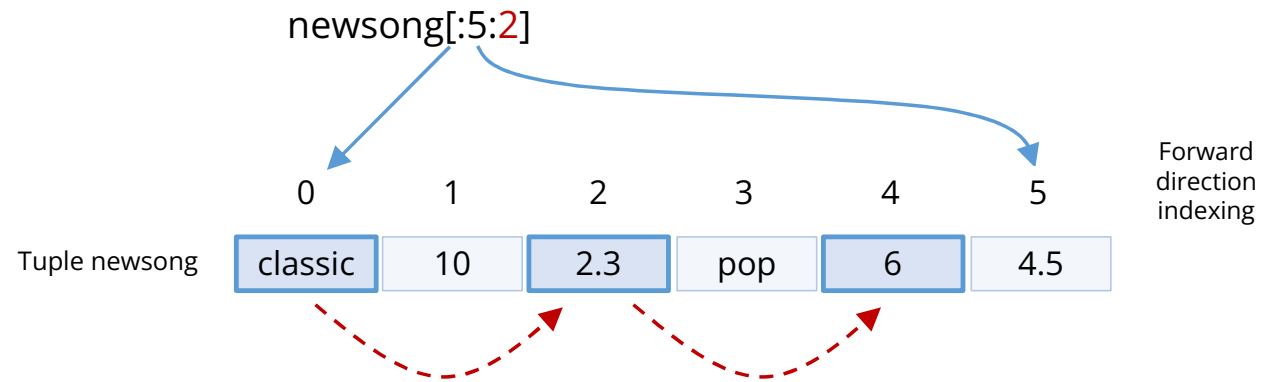
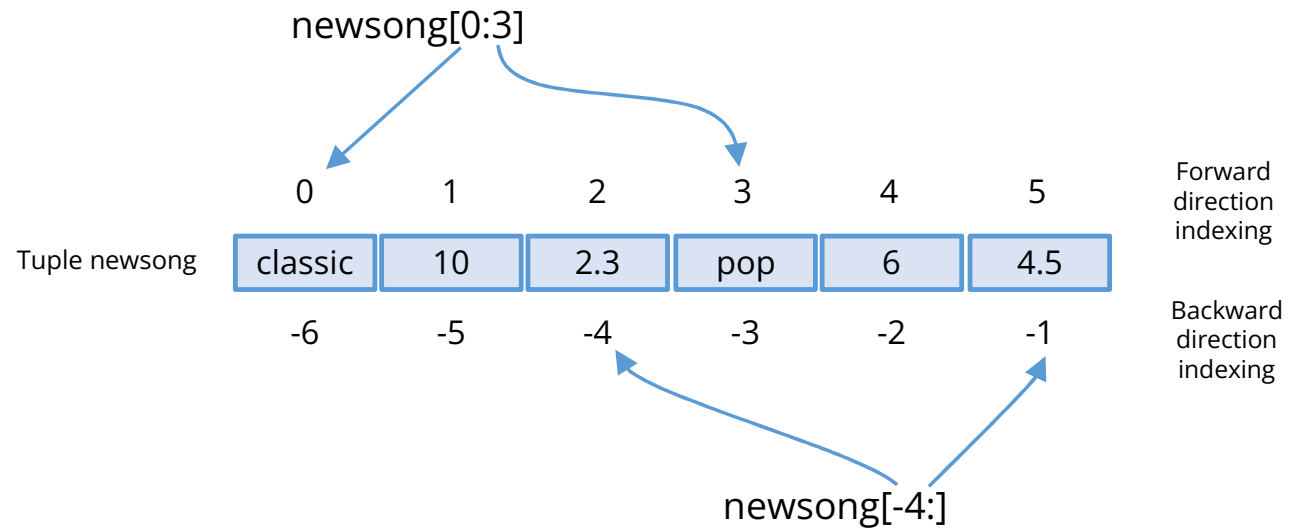
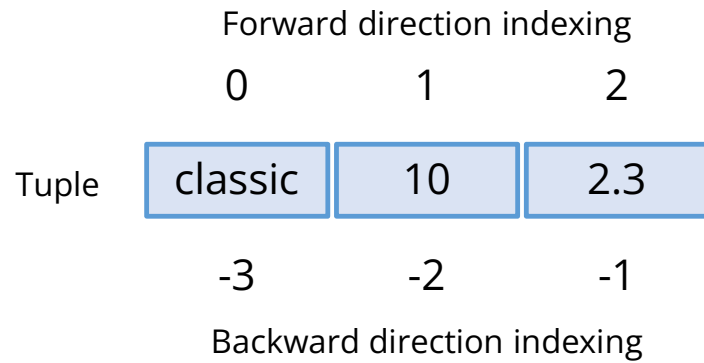
range(1, 10, 3)



Python data structures



Tuple



Chukiat Worasucheep

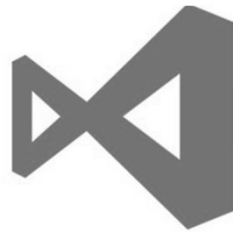
Outline

- ☐ Review basic Python
- ☐ Tools esp. Jupyter notebook
- ☐ Introduce numpy and pandas libraries

Tools for developing Python programs



PyCharm
Community



VS Code
with Python Extension



Jupyter
Notebook



(C) Chukiat Worasuchee

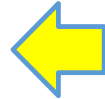
Install program jupyter

d:\>pip install jupyter 

```
Command Prompt for Notebooks
D:\notebooks>pip install jupyter
Defaulting to user installation because normal site-packages is not writeable
Collecting jupyter
  Using cached jupyter-1.0.0-py2.py3-none-any.whl (2.7 kB)
Requirement already satisfied: qtconsole in c:\users\chuki\appdata\roaming\python\python38\site-packages (from jupyter) (5.1.1)
Requirement already satisfied: notebook in c:\users\chuki\appdata\roaming\python\python38\site-packages (from jupyter) (6.4.10)
Requirement already satisfied: ipykernel in c:\users\chuki\appdata\roaming\python\python38\site-packages (from jupyter) (6.2.0)
Requirement already satisfied: nbconvert in c:\users\chuki\appdata\roaming\python\python38\site-packages (from jupyter) (6.4.4)
Requirement already satisfied: jupyter-console in c:\users\chuki\appdata\roaming\python\python38\site-packages (from jupyter) (6.4.0)
Requirement already satisfied: ipywidgets in c:\users\chuki\appdata\roaming\python\python38\site-packages (from jupyter) (7.6.3)
Requirement already satisfied: jupyter-client<8.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipykernel->jupyter) (7.0.1)
Requirement already satisfied: matplotlib-inline<0.2.0,>=0.1.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipykernel->jupyter) (0.1.2)
Requirement already satisfied: tornado<7.0,>=4.2 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipykernel->jupyter) (6.1)
Requirement already satisfied: traitlets<6.0,>=4.1.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipykernel->jupyter) (5.0.5)
Requirement already satisfied: debugpy<2.0,>=1.0.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipykernel->jupyter) (1.4.1)
Requirement already satisfied: ipython<8.0,>=7.23.1 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipykernel->jupyter) (7.26.0)
Requirement already satisfied: widgetsnbextension~=3.5.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipywidgets->jupyter) (3.5.1)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipywidgets->jupyter) (1.0.0)
Requirement already satisfied: nbformat>=4.2.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from ipywidgets->jupyter) (5.1.3)
Requirement already satisfied: pyparsing>=2.0.2 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from packaging->bleach->nbconvert->jupyter) (2.4.7)
Installing collected packages: jupyter
Successfully installed jupyter-1.0.0
WARNING: You are using pip version 22.0.3; however, version 22.2.2 is available.
You should consider upgrading via the 'c:\program files\python38\python.exe -m pip install --upgrade pip' command.
D:\notebooks>
```

Install program Jupyter lab

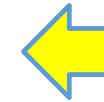
d:\>pip install jupyterlab



```
Administrator: Command Prompt
C:\Windows\system32>pip install jupyterlab
Requirement already satisfied: jupyterlab in c:\users\chuki\appdata\roaming\python\python38\
Requirement already satisfied: packaging in c:\users\chuki\appdata\roaming\python\python38\s
Requirement already satisfied: jupyterlab-server~=2.3 in c:\users\chuki\appdata\roaming\pytho
) (2.7.1)
Requirement already satisfied: tornado>=6.1.0 in c:\users\chuki\appdata\roaming\python\python
Requirement already satisfied: jinja2>=2.1 in c:\users\chuki\appdata\roaming\python\python38
Requirement already satisfied: nbclassic~=0.2 in c:\users\chuki\appdata\roaming\python\python
)
Requirement already satisfied: jupyter-core in c:\users\chuki\appdata\roaming\python\python38
Requirement already satisfied: jupyter-server~=1.4 in c:\users\chuki\appdata\roaming\python\p
1.10.2)
Requirement already satisfied: ipython in c:\users\chuki\appdata\roaming\python\python38\site
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\chuki\appdata\roaming\python\pytho
terlab) (2.0.1)
Requirement already satisfied: ipython-genutils in c:\users\chuki\appdata\roaming\python\pytl
1.4->jupyterlab) (0.2.0)
```


Install commonly used packages for data science works

d:\>pip install numpy pandas matplotlib sklearn scipy



```
Administrator: Command Prompt
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>pip install numpy pandas matplotlib sklearn scipy
Requirement already satisfied: numpy in c:\users\chuki\appdata\roaming\python\python38\site-packages (1.19.5)
Requirement already satisfied: pandas in c:\users\chuki\appdata\roaming\python\python38\site-packages (1.3.1)
Requirement already satisfied: matplotlib in c:\users\chuki\appdata\roaming\python\python38\site-packages (3.4.2)
Requirement already satisfied: sklearn in c:\users\chuki\appdata\roaming\python\python38\site-packages (0.0)
Requirement already satisfied: scipy in c:\users\chuki\appdata\roaming\python\python38\site-packages (1.5.4)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from pandas) (2022.1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: pillow>=6.2.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from matplotlib) (9.0.1)
Requirement already satisfied: cyclizer>=0.10 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from matplotlib) (0.10.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: scikit-learn in c:\users\chuki\appdata\roaming\python\python38\site-packages (from sklearn) (0.24.2)
Requirement already satisfied: six in c:\program files\python38\lib\site-packages (from cyclizer>=0.10->matplotlib) (1.16.0)
Requirement already satisfied: joblib>=0.11 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from sklearn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\chuki\appdata\roaming\python\python38\site-packages (from sklearn) (3.1.0)
WARNING: You are using pip version 22.0.3; however, version 22.0.4 is available.
You should consider upgrading via the 'c:\program files\python38\python.exe -m pip install --upgrade pip' command.

C:\Windows\system32>
```

Run Jupyter notebook

```
Command Prompt for Notebooks - jupyter notebook

D:\notebooks>jupyter notebook
[I 10:54:14.235 NotebookApp] Serving notebooks from local directory: D:\notebooks
[I 10:54:14.235 NotebookApp] Jupyter Notebook 6.4.10 is running at:
[I 10:54:14.235 NotebookApp] http://localhost:8888/?token=2f70ede29eebeff0751ddc34d94725fd1891b2e577a040d8
[I 10:54:14.235 NotebookApp] or http://127.0.0.1:8888/?token=2f70ede29eebeff0751ddc34d94725fd1891b2e577a040d8
[I 10:54:14.235 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 10:54:14.261 NotebookApp]

To access the notebook, open this file in a browser:
  file:///C:/Users/chuki/AppData/Roaming/jupyter/runtime/nbserver-13012-open.html
Or copy and paste one of these URLs:
  http://localhost:8888/?token=2f70ede29eebeff0751ddc34d94725fd1891b2e577a040d8
  or http://127.0.0.1:8888/?token=2f70ede29eebeff0751ddc34d94725fd1891b2e577a040d8
```


Jupyter notebook home

```

Select Command Prompt for Notebooks

D:\notebooks>dir
Volume in drive D is Data
Volume Serial Number is A234-7DEE

Directory of D:\notebooks

08/12/2022  04:11 PM    <DIR>          .
08/12/2022  04:11 PM    <DIR>          ..
09/16/2021  07:35 AM    <DIR>          .ipynb_checkpoints
08/22/2021  07:30 AM    <DIR>          c111-ExploringCS
11/11/2021  11:26 AM    <DIR>          c341-DS
09/23/2021  02:00 PM    <DIR>          covid
11/30/2021  04:11 PM    <DIR>          dl-with-python-chollet
10/02/2021  07:11 AM    11 jup.bat
11/30/2021  01:38 PM    <DIR>          masters
08/12/2022  04:13 PM    <DIR>          matlab2python
08/07/2021  07:36 AM    <DIR>          port-rew
02/25/2022  04:31 PM    <DIR>          pycaret
09/13/2021  02:15 PM    <DIR>          sharpe
07/06/2022  11:22 AM    <DIR>          tests
09/08/2021  10:07 PM    <DIR>          th-covid
               1 File(s)            11 bytes
              14 Dir(s) 166,980,165,632 bytes free

D:\notebooks>
```

Home Page - Select or create a notebook

localhost:8888/tree

jupyter

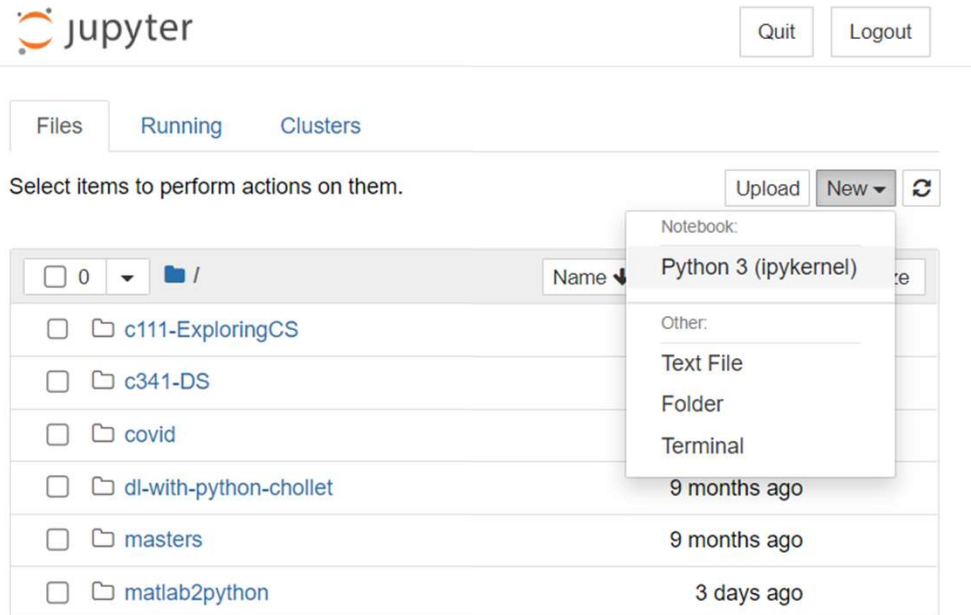
Quit Logout

Files Running Clusters

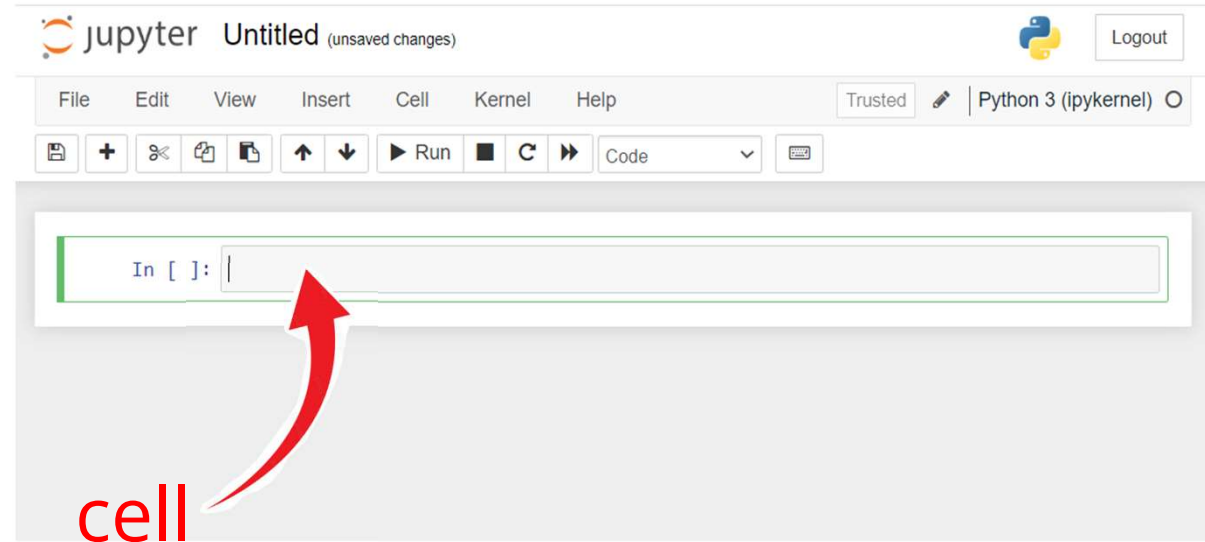
Select items to perform actions on them. Upload New

<input type="checkbox"/>	0	▼	📁 /	Name ▼	Last Modified	File size
<input type="checkbox"/>			📁	c111-ExploringCS	a year ago	
<input type="checkbox"/>			📁	c341-DS	9 months ago	
<input type="checkbox"/>			📁	covid	a year ago	
<input type="checkbox"/>			📁	dl-with-python-chollet	9 months ago	
<input type="checkbox"/>			📁	masters	9 months ago	
<input type="checkbox"/>			📁	matlab2python	3 days ago	
<input type="checkbox"/>			📁	port-rew	a year ago	
<input type="checkbox"/>			📁	pycaret	6 months ago	
<input type="checkbox"/>			📁	sharpe	a year ago	
<input type="checkbox"/>			📁	tests	a month ago	
<input type="checkbox"/>			📁	th-covid	a year ago	
<input type="checkbox"/>			📄	jup.bat	10 months ago	11 B

Create a new Python *notebook* file



The image shows the Jupyter web interface's 'Files' tab. At the top, there's a 'jupyter' logo and 'Quit' and 'Logout' buttons. Below are tabs for 'Files', 'Running', and 'Clusters'. A message says 'Select items to perform actions on them.' There are 'Upload', 'New', and a refresh icon. A file list shows folders like 'c111-ExploringCS', 'c341-DS', 'covid', 'dl-with-python-chollet', 'masters', and 'matlab2python' with checkboxes and timestamps. A 'New' dropdown menu is open, showing 'Notebook: Python 3 (ipykernel)' and 'Other: Text File, Folder, Terminal'.

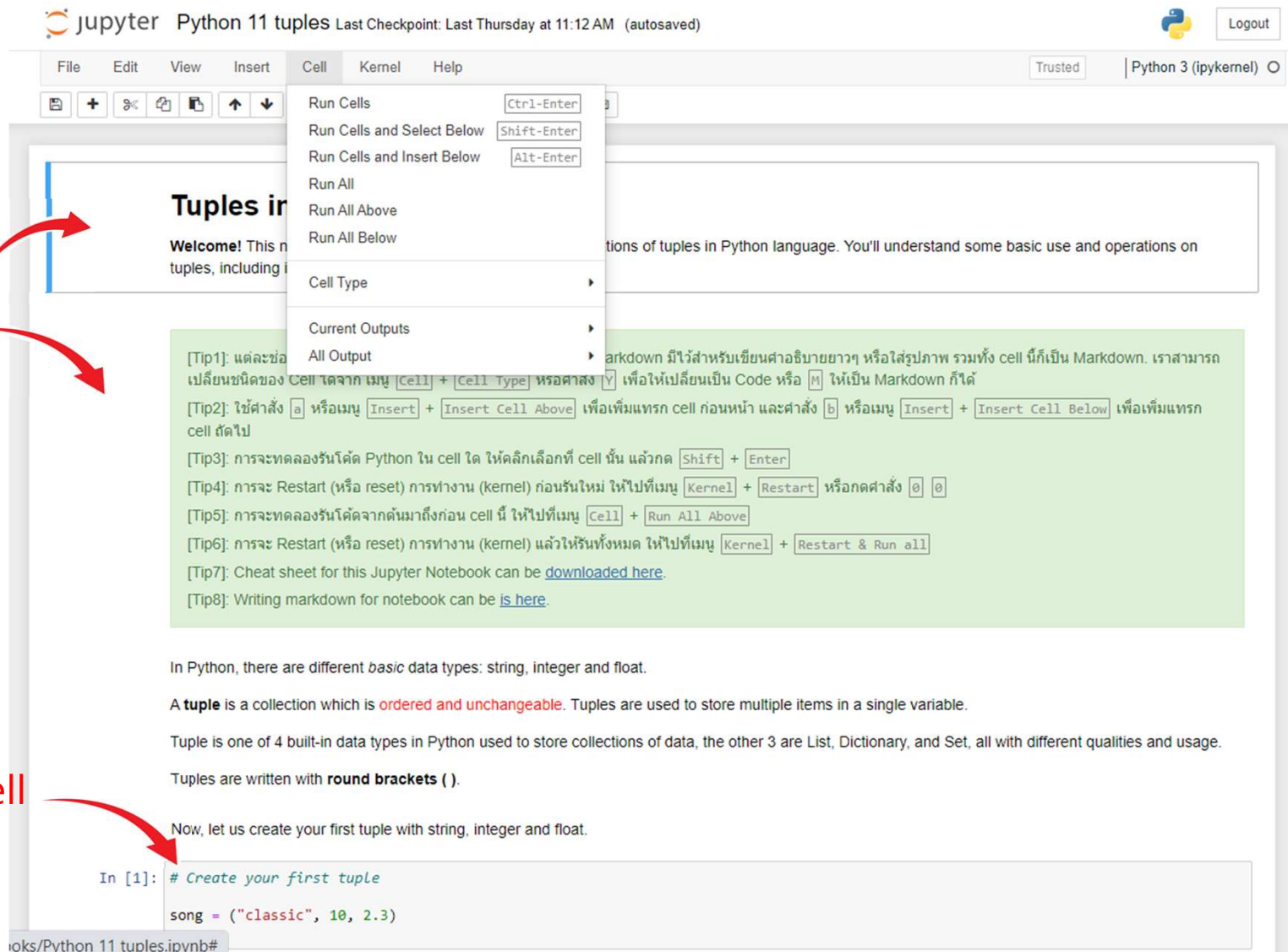


The image shows the Jupyter Notebook editor interface. At the top, it says 'jupyter Untitled (unsaved changes)' with a 'Logout' button. Below are tabs for 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', and 'Help'. There's a 'Trusted' status and 'Python 3 (ipykernel)' kernel info. A toolbar contains icons for saving, adding, zooming, copying, pasting, undo, redo, and running. A code cell is selected, showing 'In []: |'. A red arrow points from the word 'cell' at the bottom left to the code cell.

Code cells and markdown cells

Markdown cell

Code cell



Jupyter Python 11 tuples Last Checkpoint: Last Thursday at 11:12 AM (autosaved)

File Edit View Insert Cell Kernel Help Trusted Python 3 (ipykernel)

Run Cells (Ctrl-Enter)
Run Cells and Select Below (Shift-Enter)
Run Cells and Insert Below (Alt-Enter)
Run All
Run All Above
Run All Below
Cell Type
Current Outputs
All Output

Tuples in Python

Welcome! This notebook introduces the concept of tuples in Python language. You'll understand some basic use and operations on tuples, including...

[Tip1]: แต่ละช่องเปลี่ยนชนิดของ Cell ได้จากเมนู [Cell] + [Cell Type] หรือคำสั่ง `Y` เพื่อให้เปลี่ยนเป็น Code หรือ `M` ให้เป็น Markdown ก็ได้

[Tip2]: ใช้คำสั่ง `a` หรือเมนู [Insert] + [Insert Cell Above] เพื่อเพิ่มแทรก cell ก่อนหน้า และคำสั่ง `b` หรือเมนู [Insert] + [Insert Cell Below] เพื่อเพิ่มแทรก cell ถัดไป

[Tip3]: การจะทดลองรันโค้ด Python ใน cell ใด ให้คลิกเลือกที่ cell นั้น แล้วกด `Shift` + `Enter`

[Tip4]: การจะ Restart (หรือ reset) การทำงาน (kernel) ก่อนรันใหม่ ให้ไปที่เมนู [Kernel] + [Restart] หรือกดคำสั่ง `0 0`

[Tip5]: การจะทดลองรันโค้ดจากต้นมาถึงก่อน cell นี้ ให้ไปที่เมนู [Cell] + [Run All Above]

[Tip6]: การจะ Restart (หรือ reset) การทำงาน (kernel) แล้วให้รันทั้งหมด ให้ไปที่เมนู [Kernel] + [Restart & Run all]

[Tip7]: Cheat sheet for this Jupyter Notebook can be [downloaded here](#).

[Tip8]: Writing markdown for notebook can be [is here](#).

In Python, there are different *basic* data types: string, integer and float.

A **tuple** is a collection which is **ordered and unchangeable**. Tuples are used to store multiple items in a single variable.

Tuple is one of 4 built-in data types in Python used to store collections of data, the other 3 are List, Dictionary, and Set, all with different qualities and usage.












Tuples are written with **round brackets ()**.

Now, let us create your first tuple with string, integer and float.

```
In [1]: # Create your first tuple
song = ("classic", 10, 2.3)
```

books/Python 11 tuples.ipynb#

Notebook files for learning

Name	Date modified	Type	Size
 .ipynb_checkpoints	11/8/2565 10:08	File folder	
 Python 04 strings.ipynb	28/6/2565 9:23	IPYNB File	33 KB
 Python 11 tuples.ipynb	11/8/2565 11:12	IPYNB File	83 KB
 Python 12 lists.ipynb	28/6/2565 9:25	IPYNB File	203 KB
 Python 13 dictionaries.ipynb	28/6/2565 9:26	IPYNB File	98 KB
 Python 21 conditions.ipynb	28/6/2565 10:14	IPYNB File	1,346 KB
 Python 22 loops.ipynb	28/6/2565 10:15	IPYNB File	167 KB
 Python 23 functions.ipynb	28/6/2565 10:16	IPYNB File	122 KB
 Python 24 classes.ipynb	28/6/2565 10:20	IPYNB File	25 KB
 Python 31 numpy1D.ipynb	28/6/2565 10:22	IPYNB File	74 KB
 Python 32 numpy2D.ipynb	28/6/2565 10:26	IPYNB File	26 KB

Outline

- ☐ Review basic Python
- ☐ Tools esp. Jupyter notebook
- ☐ Introduce numpy and pandas libraries

numpy

- A Python open-source library, created in 2005 by Travis Oliphant, used for working with arrays.
- NumPy stands for Numerical Python.
- Widely used in domain of linear algebra, matrices, Fourier transform, image processing.
- Mainly used for n-dimensional arrays.
- Implemented in C, and thus is up to 50x faster than traditional Python lists.
- Great place for further study:
 - https://www.w3schools.com/python/numpy/numpy_intro.asp
 - <http://mscbio2025.csb.pitt.edu/notes/numpy.slides.html#/>
 - <https://cloudxlab.com/blog/numpy-pandas-introduction/>
 - <https://www.codecademy.com/article/introduction-to-numpy-and-pandas>



pandas

- pandas is a fast, powerful, flexible and easy to use open-source data analysis and manipulation tool.
- It has functions for analyzing, cleaning, exploring, and manipulating data.
- *pandas* stands for “*Python Data Analysis*” and was created by Wes McKinney in 2008.
- The source code for Pandas: <https://github.com/pandas-dev/pandas>
- Great place for further study: <https://www.w3schools.com/python/pandas/default.asp>
- <https://pandas.pydata.org/>



What we've learnt today

- Review basic Python
- Tools esp. Jupyter notebook
- Introduce numpy and pandas libraries

