

統計學一上： PYTHON介紹、 作業以及期中/期末考試環境說明

Hsin-Min Lu

盧信銘

台大資管系

Agenda

- Python介紹
- WinPython安裝
- 作業以及期中/期末考試環境說明

Python短史



- 1991年由Guido van Rossum創造
- 注重程式可讀性
- Python 2.0於2000年釋出，
2020年正式退役。
- Python 3.0於2008年釋出，語法與2.0
不完全相容。
- Python 3.8為目前最新版。
- Python直譯器有數種實作，最常見的是
由C語言寫成的CPython。
- Python為當前從事資料處理與分析、機
器學習、深度學習的主流語言。

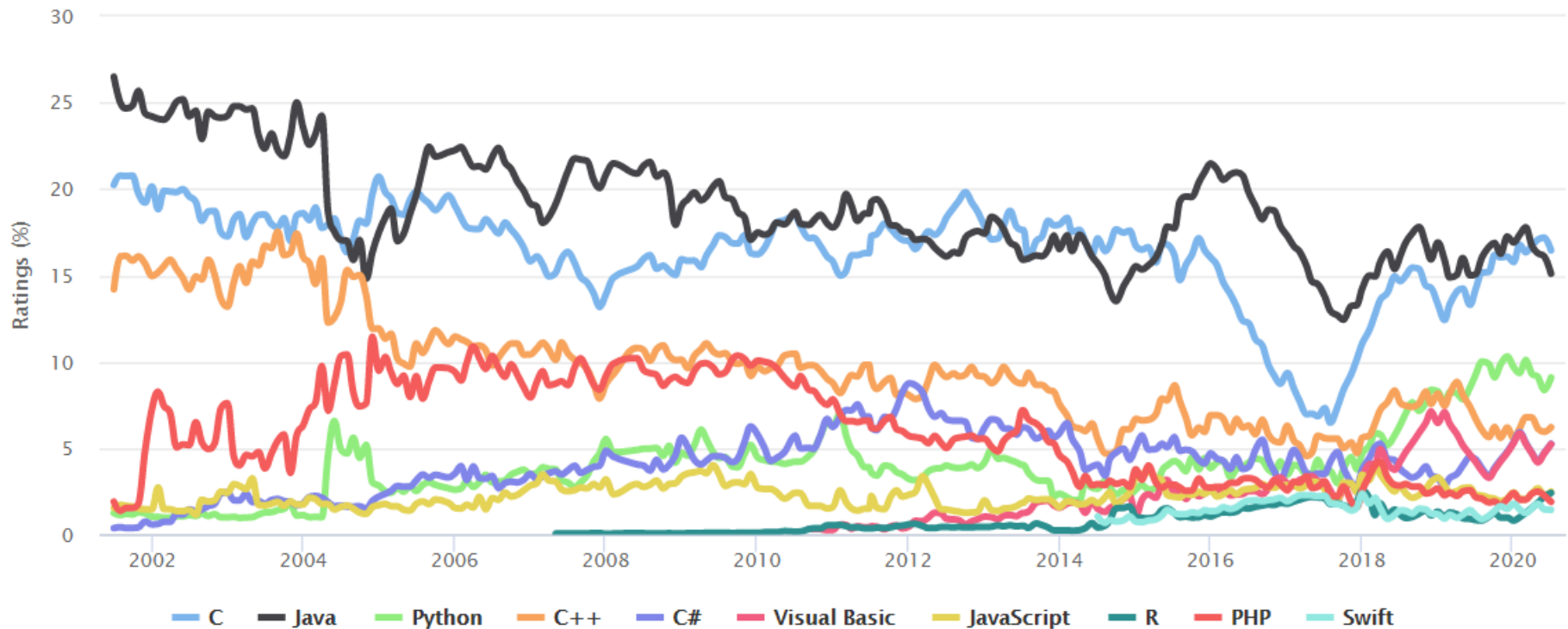


Python是前三大程式語言

- 前三大分別是C, Java, Python。

TIOBE Programming Community Index

Source: www.tiobe.com



都是程式語言，但Python很不一樣

- C語言是“高效能”但較難開發與維護的語言
- C語言典型的功用是開發作業系統(Operation Systems)，為製作Windows 10與Linux的主要程式語言。
- Java跟C語言一樣是“通用型”、“高效能”的程式語言。
- Java語法嚴謹、但寫起來攏長，為人(我)詬病。
- Java典型的功用是開發Andriod平台APP。
- Python是“直譯式”的程式語言，執行效率不及C與Java。但卻在近年備受喜愛。
- 常是“當你需要一個程式語言”時的選擇。

Python

- 既然效率沒有C跟Java來的好，為何受到喜愛？
- 主要原因：
 - 語法乾淨、容易了解
 - 開發與除錯簡易
 - 社群發達、容易找到資源
 - 跨平台
 - 資料處理、分析、與數值運算的功能完整，資源豐富

資料處理與數據分析軟體比較

	軟體工具	資料清理 與整併	視覺化	資料探勘與 機器學習	深度學習
套裝軟體	Excel*	Yes (手動、 小量)	Yes	No	No
	Tableau	Partial	Yes	No	No
	Power BI	Partial	Yes	Partial	Partial
	SAS	Yes	Yes	Yes	Partial
	RapidMiner	Yes	Yes	Yes	Partial
程式語言	Python	Yes (Pandas)	Yes (Matplotlib)	Yes (sklearn與 其他)	Yes (Tensorflow, Pytorch)
	R	Yes	Yes (ggplot)	Yes (多種套件)	Yes (MXNetR與其他)

* Excel 可以使用 Visual Basic for Application (VBA) 自動化，但執行速度極慢，不建議在大量數據下使用。

我愛Excel，還需要Python嗎？

- 我愛Excel，我也愛Python
- Excel可以處理80%的日常資料處理問題。
- 現在Excel的圖表也算漂亮。
- Excel只要拖拉點選，加上靈活應用公式，可以達到很多功能。
- 但只使用Excel有一些缺點：
 - 重複性低
 - 資料表合併(Merge)困難
 - 一些統計工作較難執行
 - 無法處理大量資料

Python跟R的比較

- 套裝軟體大部分有不同形式的費用，可以依自己的需求斟酌採買。
- 程式語言具有較大彈性，能適應各種狀況，在資料分析任務較複雜時能發揮。

	Python	R
定位	通用程式語言	以結構性資料起家
執行速度	中等	中等
套件豐富度	極高	極高
是否跨平台	是	是
結構資料處理	強	強
非結構資料處理	強	中等
中文處理	正常	有問題
視覺化套件	豐富	豐富
資料探勘與機器學習	豐富	豐富
深度學習	豐富	較弱
大量資料處理	較有彈性	較沒彈性

Python的選擇

- Python主要有第二版與第三版。第二版已經退役。
- 強烈建議使用Python第三版 (Python3)。
- Python的環境設置隨個人喜好不同，會有差異，大致上可以區分為幾個類型：
- **最簡設置**：Python直譯器與文字編輯器 (如Notepad++或VS Code)
- **整合環境**：使用整合的圖形介面，如Pycharm
- **報告生成環境 (原生)**：以生成報告為主要目的的開發環境，如Jupyter Notebook或Jupyter Lab
- **報告生成環境 (Anaconda)**：使用第三方公司所建構好的設置，如Anaconda。
- 另有安裝簡易的Winpy (<http://winpython.github.io/>)，檔案拷貝完成即可執行。

Python環境比較

	優點	缺點
最簡設置	環境單純、觀念清楚	不適合發展大型程式
整合環境	功能完整，適合中大型程式發展	功能較複雜，初學者較易混淆
報告生成環境 (原生)	適合紀錄工作流程與重複執行	不適合模組化開發
報告生成環境 (Anaconda)	安裝簡易	可能會受提供廠商約束
報告生成環境 (winpy)	安裝最簡易	未來更新不確定性

- 本課程將會使用報告生成環境(Jupyter Lab)
- 本課程作業與考試使用Winpy。

Python 環境設定

- Winpy: <http://winpython.github.io/>
- 建議使用Python 3.8或更新的版本
- 如果為64bit Windows PC，建議下載64bit版本。

推薦書目

- Think Python 2e (作者Allen B. Downey)
- 英文入門書，全文下載：
<https://greenteapress.com/wp/think-python-2e/>
-
- Python程式設計：從入門到進階應用 (黃建庭著)
- 入門書，<https://www.books.com.tw/products/0010818402>
-
- Python技術者們練功 (作者Naomi Ceder，張耀鴻譯)
- 進階書，<https://www.books.com.tw/products/0010834816>
- 善用Python官網文件與Google

WINPYTHON安裝

WinPython Installation Process

- Goto WinPython Website: <http://winpython.github.io/>
- Download WinPython 3.8.5:
<https://github.com/winpython/winpython/releases/download/3.0.20200808/Winpython64-3.8.5.0.exe>
- Execute [Winpython64-3.8.5.0.exe](#)

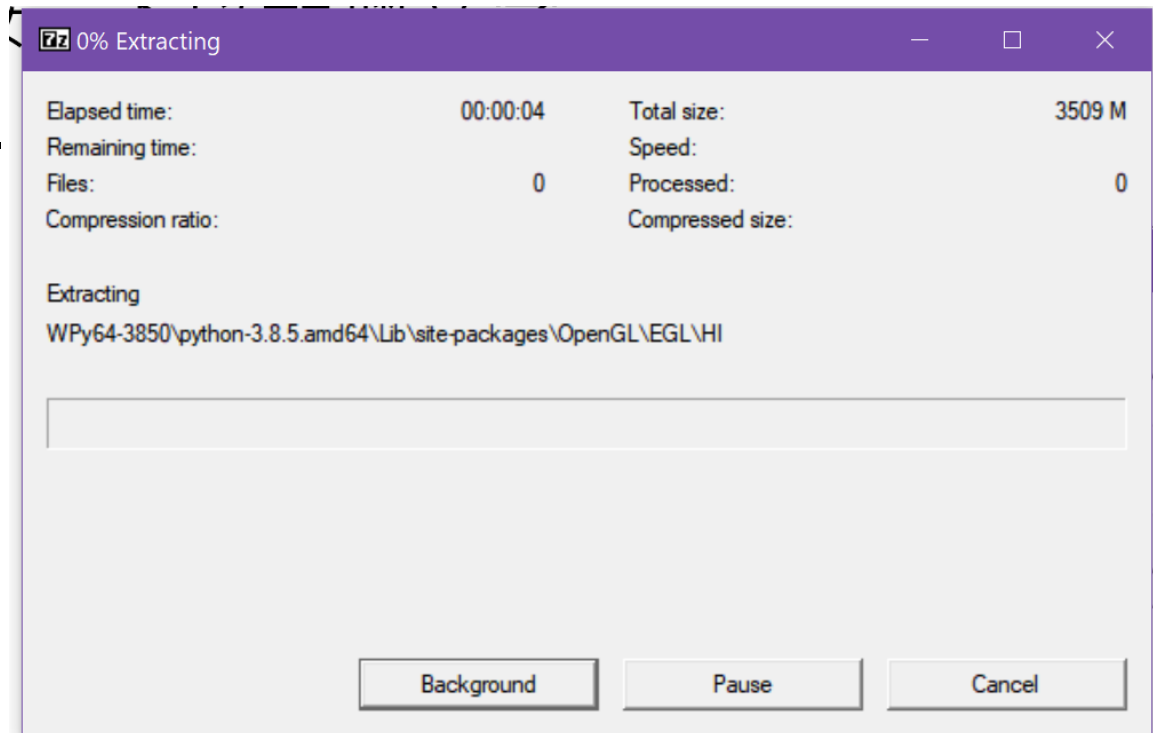
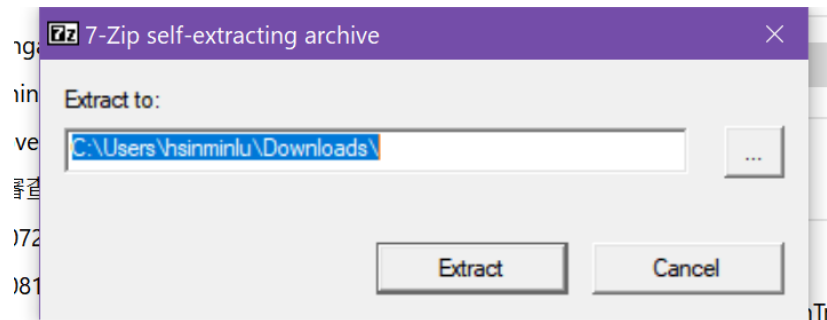
Just In Case

- If you see this,
- Click “More info”
- Then click
“Run Anyway”



Select you Install Location

- And wait.....
- It takes 3 – 5 minutes to finish copying files.
- It takes longer is the target location is a cheap USB drive.
- (USB drive usually have a slower write speed)



安裝完成後，啟動Jupyter-Lab

- 開啟WPy-64-3850目錄
- 雙擊“Jupyter Lab.exe”
(如果你沒有開啟顯示延伸檔名，
會只有看到“Jupyter Lab”)

› Users › hsinminlu › Downloads › WPy64-3850

Name	Date
n	7/22
notebooks	8/15
python-3.8.5.amd64	8/15
scripts	8/15
settings	8/15
t	8/15
IDLE (Python GUI).exe	8/15
IDLEX.exe	8/15
IPython Qt Console.exe	8/15
Jupyter Lab.exe	8/15
Jupyter Notebook.exe	8/15
license.txt	3/17
Pyzo.exe	8/15
Qt Designer.exe	8/15
Qt Linguist.exe	8/15
Spyder reset.exe	8/15
Spyder.exe	8/15
VS Code.exe	8/15
WinPython Command Prompt.exe	8/15
WinPython Control Panel.exe	8/15
WinPython Interpreter.exe	8/15
WinPython Powershell Prompt.exe	8/15

等待環境初始化

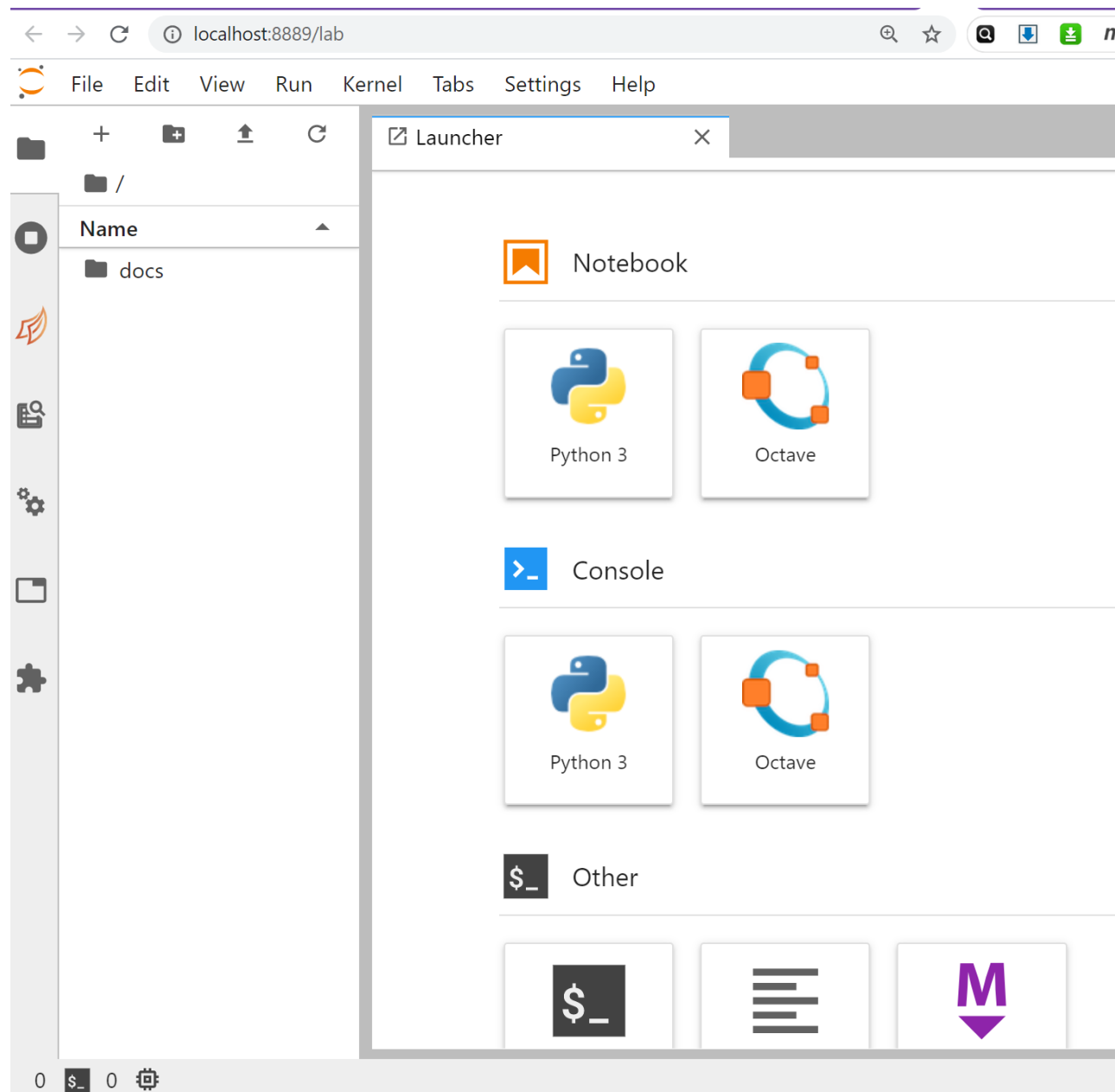
- 第一次使用Jupyter-Lab需要等待環境初始化
- Python會將Library Compile成bytecode。
- 這在USB Drive上會有點久 (e.g. 10分鐘)。

```
[I 09:30:35.076 LabApp] Writing notebook server cookie secret to C:\Users\hsinminlu\Downloads\WPy64-3850\settings\runtim
e\notebook_cookie_secret
[I 09:30:36.302 LabApp] The port 8888 is already in use, trying another port.
[I 09:30:36.728 LabApp] Loading IPython parallel extension
[I 09:30:37.296 LabApp] JupyterLab extension loaded from C:\Users\hsinminlu\Downloads\WPy64-3850\python-3.8.5.amd64\lib\
site-packages\jupyterlab
[I 09:30:37.296 LabApp] JupyterLab application directory is C:\Users\hsinminlu\Downloads\WPy64-3850\python-3.8.5.amd64\s
hare\jupyter\lab
[I 09:30:37.380 LabApp] Serving notebooks from local directory: C:\Users\hsinminlu\Downloads\WPy64-3850\notebooks
[I 09:30:37.380 LabApp] Jupyter Notebook 6.1.3 is running at:
[I 09:30:37.381 LabApp] http://localhost:8889/?token=9c86ea346b51d265eab175c638f12fe15d9bea775c94c78f
[I 09:30:37.381 LabApp] or http://127.0.0.1:8889/?token=9c86ea346b51d265eab175c638f12fe15d9bea775c94c78f
[I 09:30:37.381 LabApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 09:30:37.480 LabApp]

To access the notebook, open this file in a browser:
    file:///C:/Users/hsinminlu/Downloads/WPy64-3850/settings/runtime/nbserver-12608-open.html
Or copy and paste one of these URLs:
    http://localhost:8889/?token=9c86ea346b51d265eab175c638f12fe15d9bea775c94c78f
    or http://127.0.0.1:8889/?token=9c86ea346b51d265eab175c638f12fe15d9bea775c94c78f
Generating new fontManager, this may take some time...
[I 09:30:42.293 LabApp] Build is up to date
```

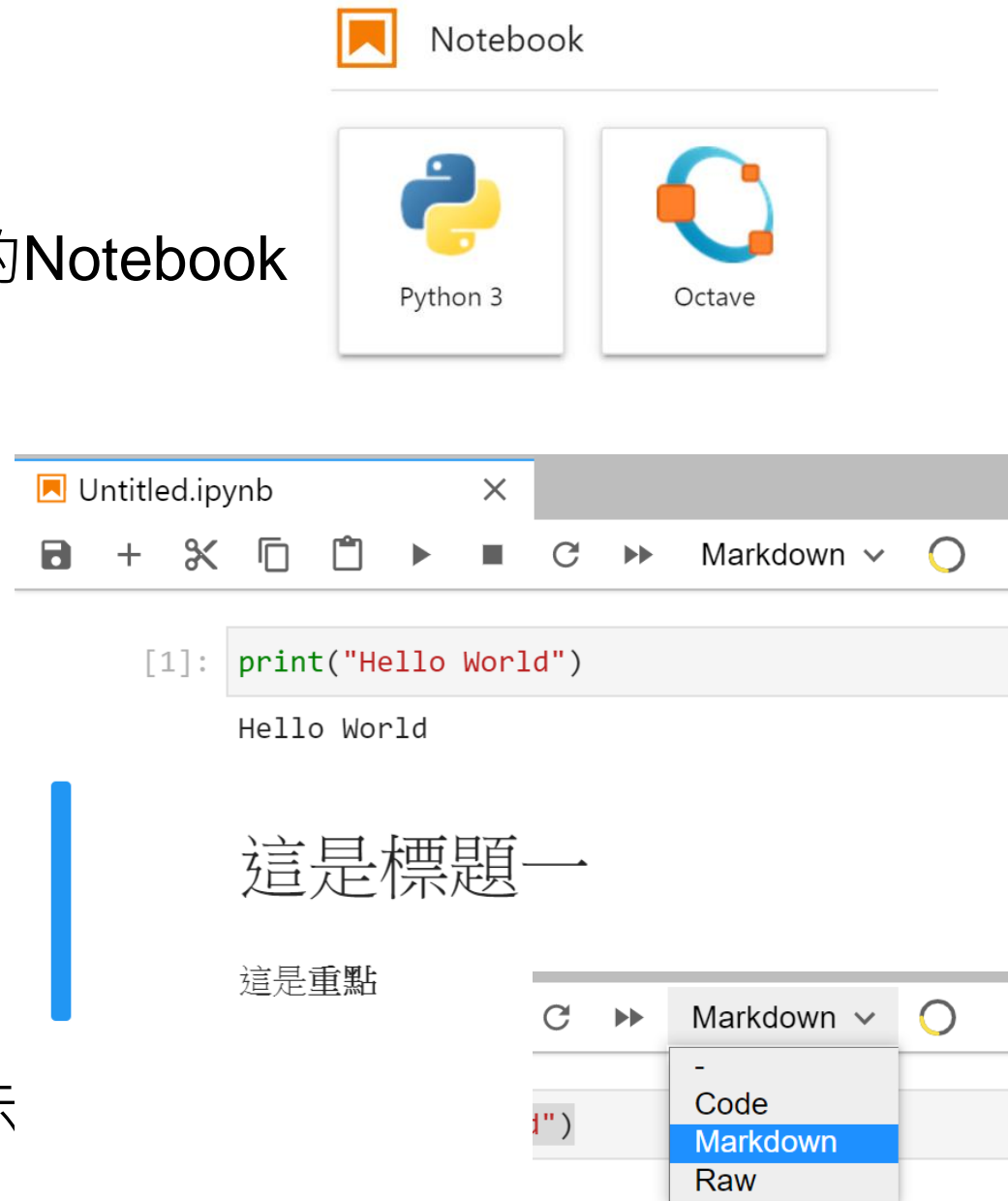
Jupyter-Lab環境出現在瀏覽器中

- 建議使用 Chrome 瀏覽器。
(需設為預設瀏覽器)
- Firefox 也可以。
- 請不要用 IE
(母湯喔)
- 也許可以考慮新的 Edge...



測試你的新環境

- 按下“Python 3”，開啟新的Notebook
- 輸入 `print("Hello World")`
- 按 **Ctrl-Enter**
- 確認程式正常執行。
- 按 “+”新增Cell
- 轉換為Markdown Cell (上方選單)
- 輸入
這是標題一
這是**重點**
- 按**Ctrl-Enter**。
- 確認Markdown有正確顯示



Jupyter-Lab檔案位置

- 在Jupyter-Lab環境中，預設的跟目錄是安裝目錄下的“notebooks”目錄。
- 如: C:\Users\hsinminlu\Downloads\WPy64-3850\notebooks
- 建議把所有的Jupyter-lab檔案放在這個目錄下。
- 如果你需要把檔案放在別的地方，可以使用mklink /d指令建立目錄連結。如此可以直接在在notebooks目錄下存取不在這裡的目錄與檔案。
- mklink /d類似Linux下的symbolic link
- mklink /d不是建立捷徑。捷徑只是一個普通檔案...
- Cf. <https://magicjackting.pixnet.net/blog/post/223136040>
<https://docs.microsoft.com/zh-tw/windows-server/administration/windows-commands/mklink>

給Mac使用者.... (註: 我主要使用 Windows 10、Ubuntu 18 LTS、與iPhone)

- 自行處理Jupyter-lab安裝，可選用原生環境或Anaconda。
- 仍需準備一個裝有WinPython的USB Drive，以備期中考與期末考使用。
- 我不懂Mac的好。但以下幾點供你參考。
- Linux (e.g. Ubuntu, Redhat, Arch Linux)的硬體支援不及Windows 10，圖形介面仍然沒有Mac好用。但做為後端系統或數值計算平台很強大。
- Windows 10的Windows Subsystem for Linux Version 2 (WSL2)可以讓你在Windows環境下執行你喜歡的Linux環境，並且與原生Windows系統有良好整合。
- 如果你買得起Mac，也許可以考慮高階的Windows筆電，如Surface Pro、Dell XPS 13之類的。(cf. <https://www.cnet.com/news/best-laptop-for-2020/>)

如果你對Python不熟...

- NTUCOOL上有Python的介紹:
- Python 101 (影片與ipynb檔案)
- Numpy Introduction (影片與ipynb檔案)
- 另外我們還會用到pandas, scipy.stats, statsmodels等。

作業以及期中/期末考試環境 說明

作業繳交

- 作業使用Jupyter Lab完成。
- 上傳由IPYNB轉換的HTML檔案至COOL作業區。
- COOL上傳是唯一的繳交方式。你不能印出來以紙本繳交，或帶隨身碟來要求助教拷貝檔案。

作業 工作環境設定

- (強烈)建議使用WinPython 64bit 3.8
- Official Site: <https://winpython.github.io>
- Link:
<https://github.com/winpython/winpython/releases/download/2.1.20190928/Winpython64-3.7.4.1.exe>
- 這是一個Portable的版本，安裝只是將檔案解壓縮至指定資料集中，完成後執行“Jupyter Lab.exe”即可開始使用。
- 如果解壓縮至隨身碟，則可以隨插即用。
- Mac與Linux使用者: 請自行安裝Python 3.8與Jupyter Lab。

考試

- 考試(有很高機會)會在管院電腦教室
- 考試時需自行攜帶裝有WinPython環境(且已設定好)的隨身碟應考。
- 你不應該使用電腦室電腦中安裝的Python與Jupyter-lab。

作業答題注意事項

- 繳交檔案最前面請註明你的姓名學號。
- 所有作業請使用Jupyter Lab回答，並將HTML檔案繳交至COOL作業區。
- 除了概念討論，簡答題等沒有牽涉到計算的題目外，所有的計算需要以Python程式碼在IPYNB檔案中進行，並有適當的說明。如果沒有達到這個要求，作業成績將會斟酌扣分(打對折或更多)。
- 舉例而言，如果要求單尾檢定 $\alpha = 0.05$ 的z值，應該以`scipy.stats.norm.ppf(1-0.05)`求得，而不是去查表然後抄下1.645這個神奇數字。如果題目牽涉到資料的分析(比如說計算樣本均數)，這些分析都必須要在IPYNB檔案中完成，而不是在其他地方計算完成後將答案抄回。
- 另外提醒答題應該使用“三明治”答題法，也就是在計算之前說你要做什麼 (e.g., Your Hypotheses and what methods to use)，然後使用Python算出你要的結果，最後說明結果的意義 (e.g., Hypotheses Test Results)。

QUESTIONS?
