

打造自己的 Jupyter Notebook

在資料科學(Data science)的領域中, R 與 Python 是很常用的 2 種語言, 一般編寫 R 語言, Rstudio 是公認的最佳 IDE 之一, 而 Python 可用的 IDE 也很多, 其中 Jupyter notebook 的方便性, 也深受大家的喜愛。

另外因為機器學習的套件, 如: Keras, TensorFlow...等, 支援 Python, 加上 Python 語言易學的特性, 使得 Python 受歡迎的層度, 大大的提升。



圖片來源: [IEEE Spectrum](#)

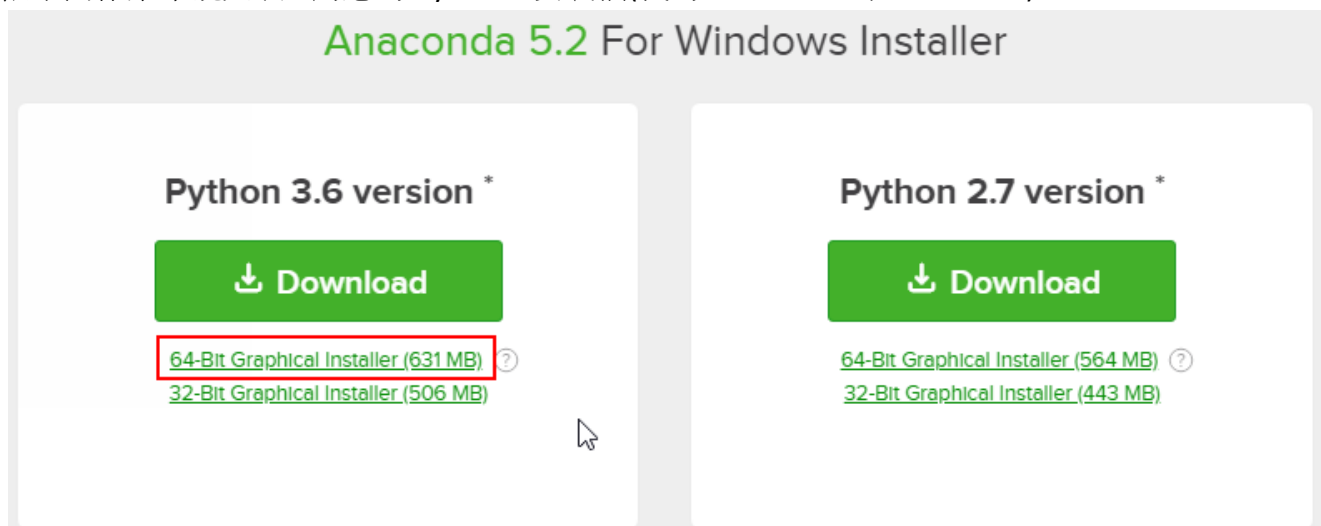
不過近期 Keras, TensorFlow 也都有支援 R, 接下來我們要做 2 件事:

1. 將 R 語言整合至 Jupyter notebook。
2. 利用 R interface to Keras, 重寫手寫字跡辨視。

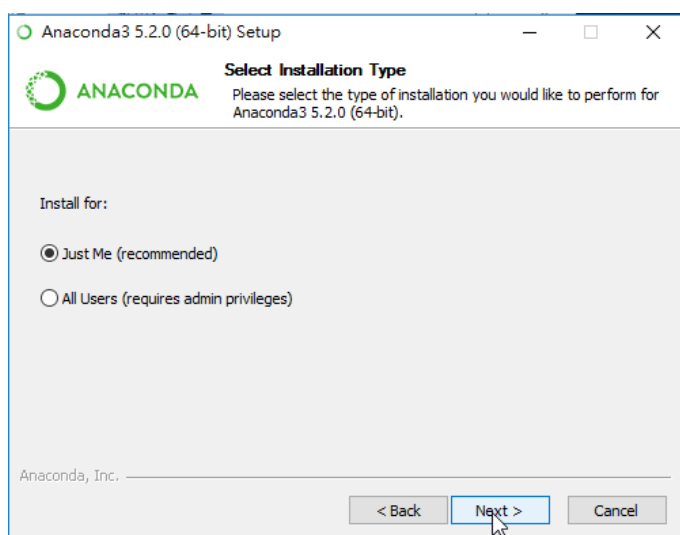
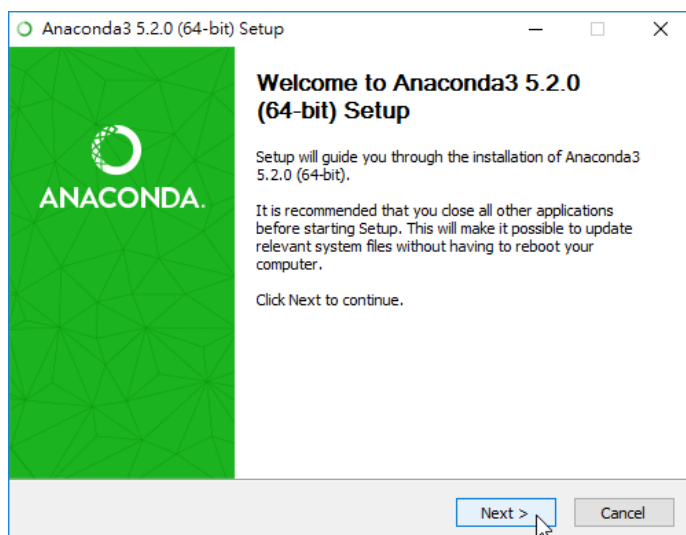
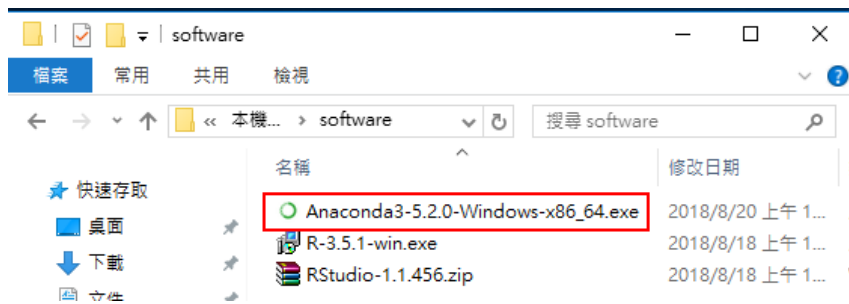
安裝 Anaconda

首先下載 [Anaconda](#), 因為它提供完整的解決方案, 預先安裝好 Python 資料科學應用模組, 包含 NumPy、pandas、matplotlib 與 sklearn 等。

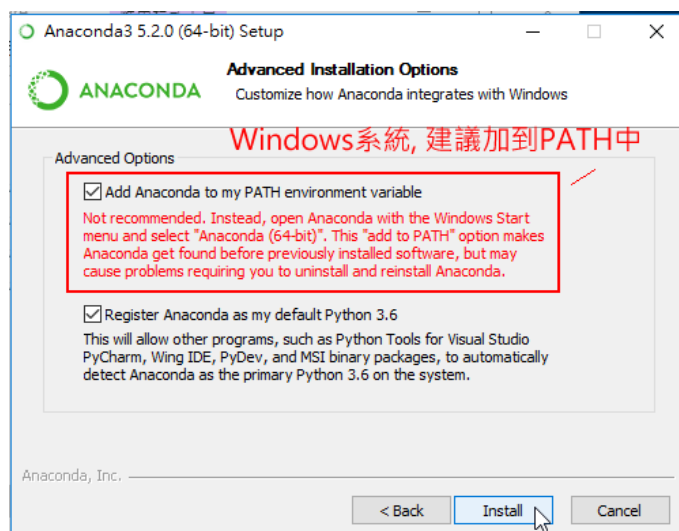
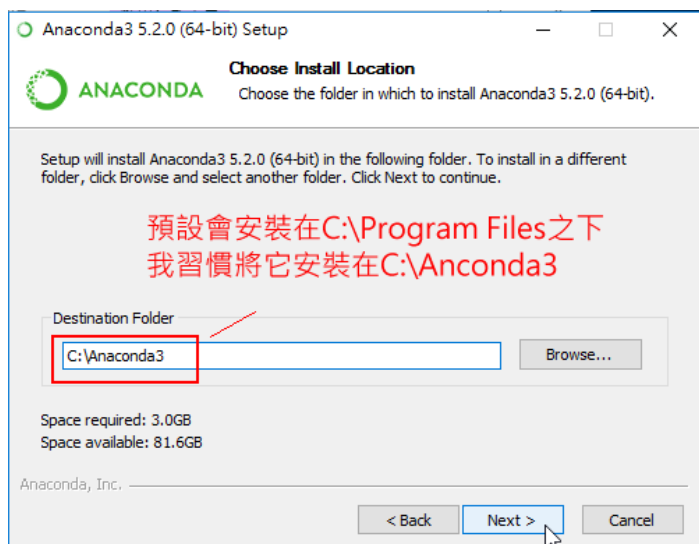
請依不同作業系統, 點選對應的 Python3 安裝檔(我的 Notebook 為 Win10 x64)。



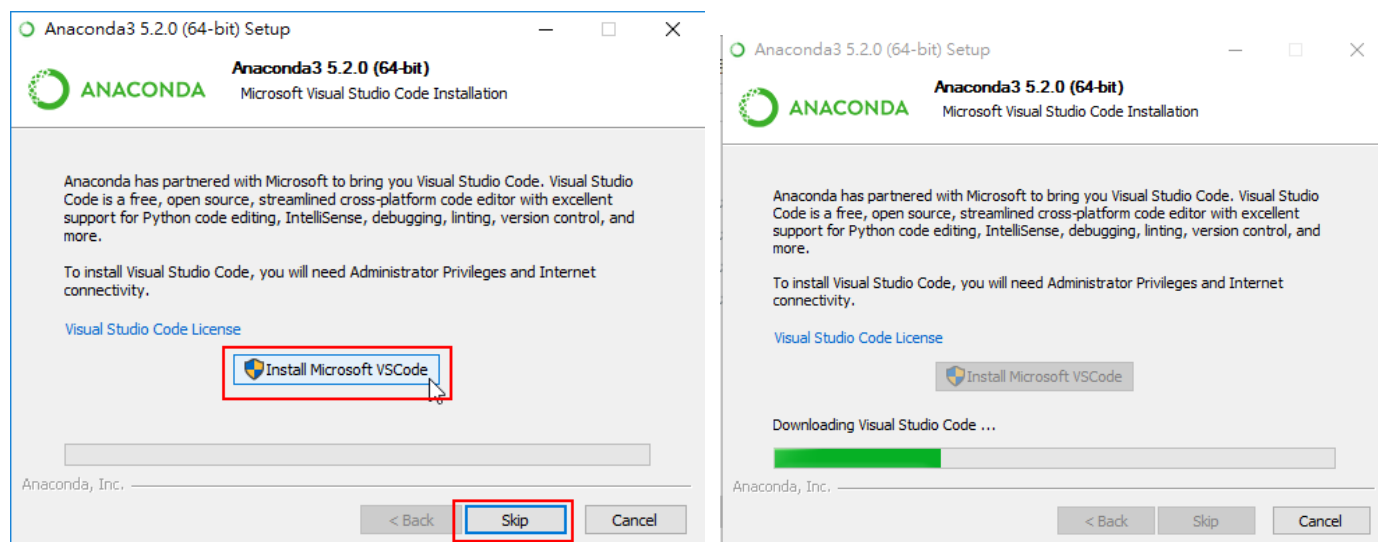
下載完成後的安裝，雙擊 Anaconda3-5.2.0-Windows-x86_64.exe



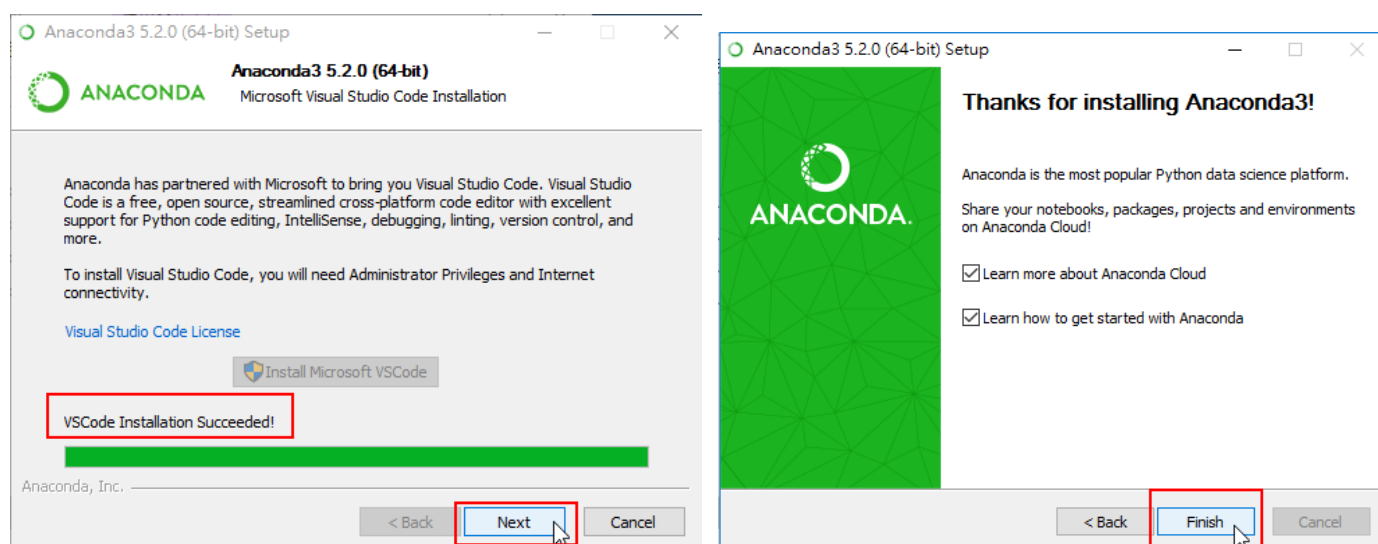
程式預設會安裝在 C:\Program Files 之下，也可以裝到其他目錄(例如: C:\Anaconda3), Windows 系統建議將路徑加到 PATH 中，方便在命令提示字元模式中方便操作，預設是沒有勾選。



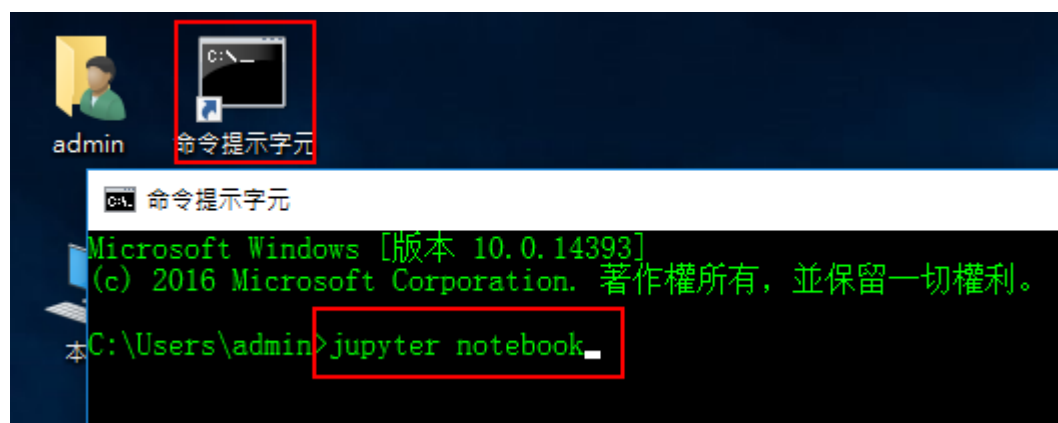
是否要安裝 Visual Studio Code, 可自行選擇, 若要安裝請點[Install Microsoft VSCode], 若要跳過請點 [Skip], 這裏點安裝 VSCode



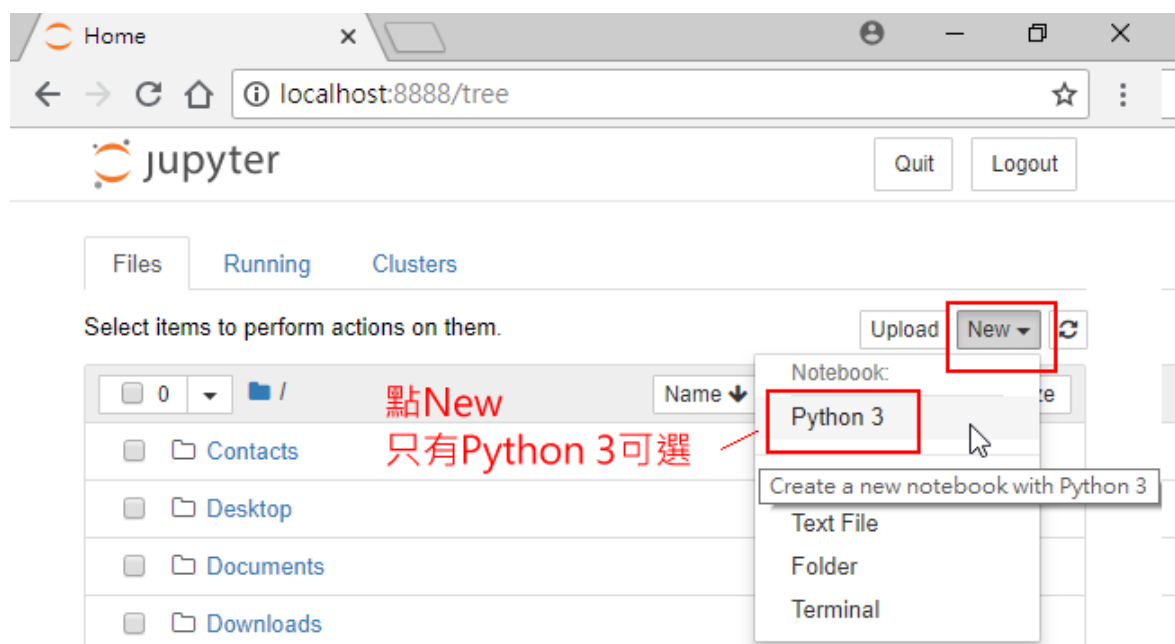
安裝完 VSCode 後點 Next, 繼續安裝 Anaconda3, 最後按 Finish 結束安裝。



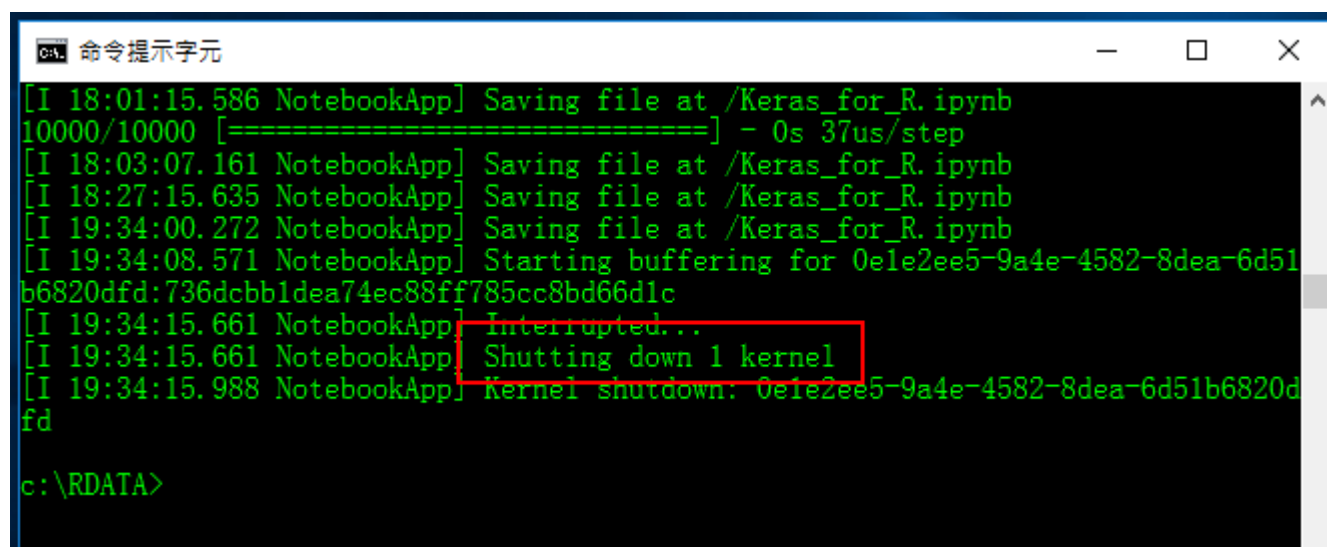
雙擊[命令提示字元], 在命令列模式輸入 jupyter notebook



啟動 Jupyter notebook 後, 點 New, 目前只有 Python 3 可以選擇。

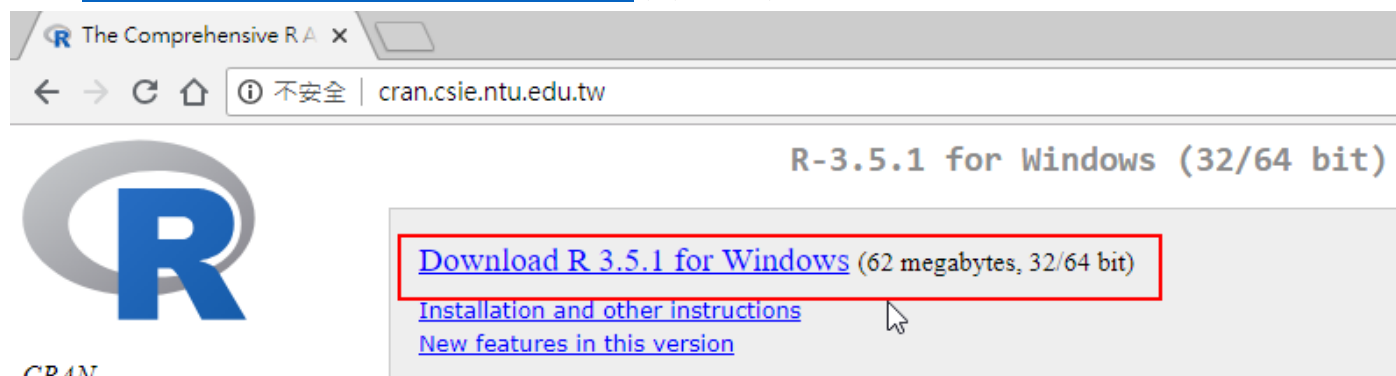


關閉 Browser 後回到命令提示字元模式, 按 Ctl + C 確認終止服務。



安裝 R 語言及 RStudio

開啟 [CRAN \(Comprehensive R Archive Network\)](https://cran.r-project.org/) 下載 R 語言



開啟 [RStudio Download](#), 下載 RStudio 安裝包, 這裏我選擇免安裝的 ZIP 檔, 下載後直接解壓縮即可使用。



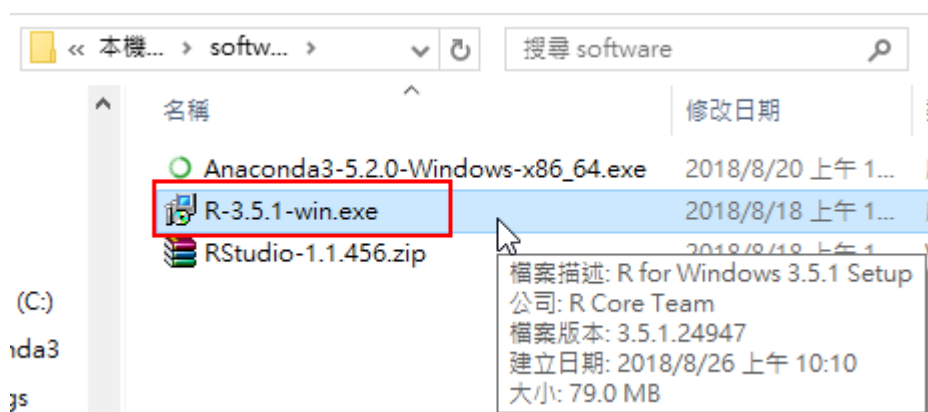
Installers for Supported Platforms

Installers	Size	Date
RStudio 1.1.456 - Windows Vista/7/8/10	85.8 MB	2018-07-19
RStudio 1.1.456 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-07-19

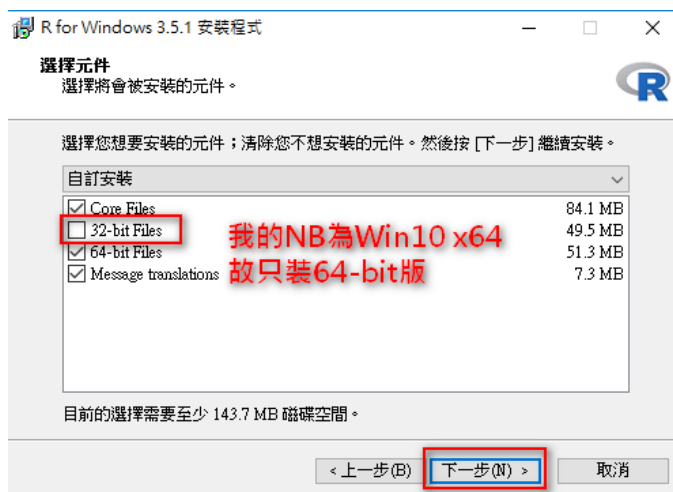
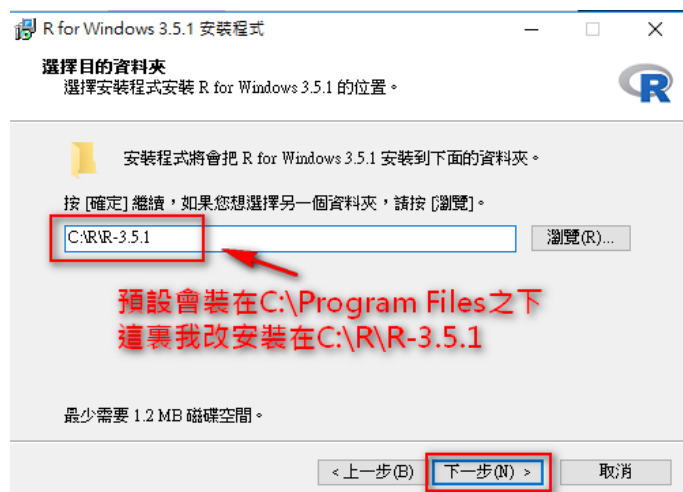
Zip/Tarballs

Zip/tar archives	Size	Date
RStudio 1.1.456 - Windows Vista/7/8/10	122.9 MB	2018-07-19
RStudio 1.1.456 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-07-19

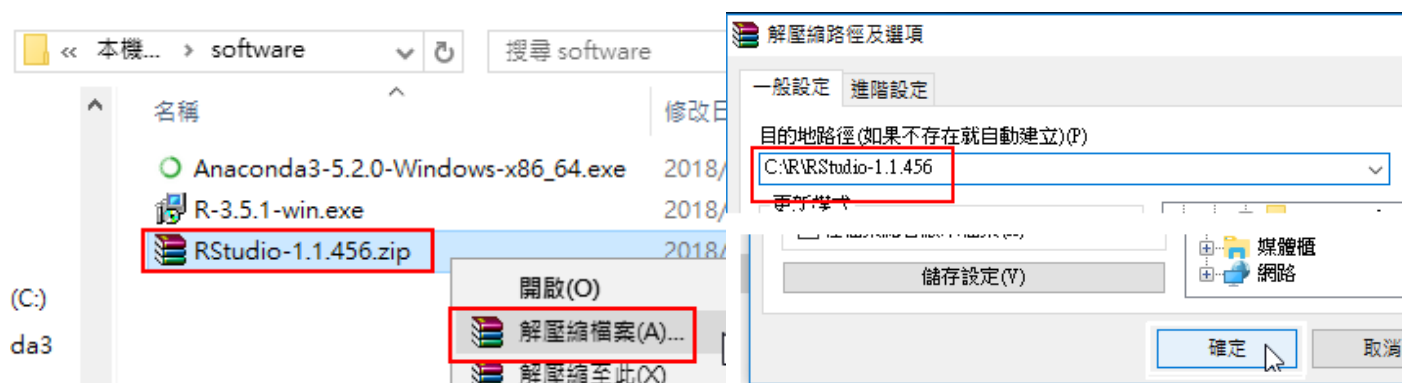
雙擊 R-3.5.1-win.exe, 開始安裝 R 語言



這裏我選擇安裝在 C:\R\R-3.5.1 之下, 僅安裝 64-bit Files

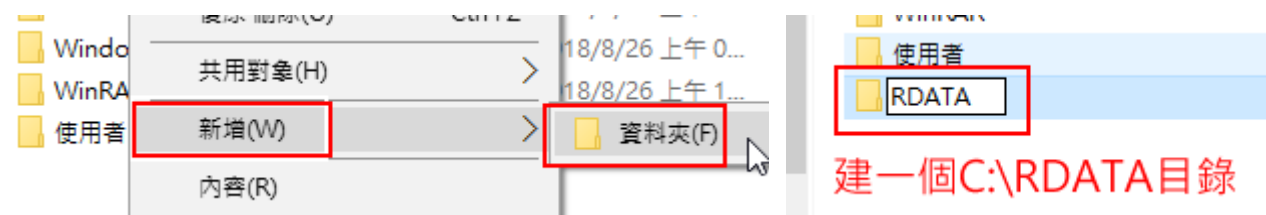


解壓縮 RStudio 到 C:\R\RStudio-1.1.456 目錄下

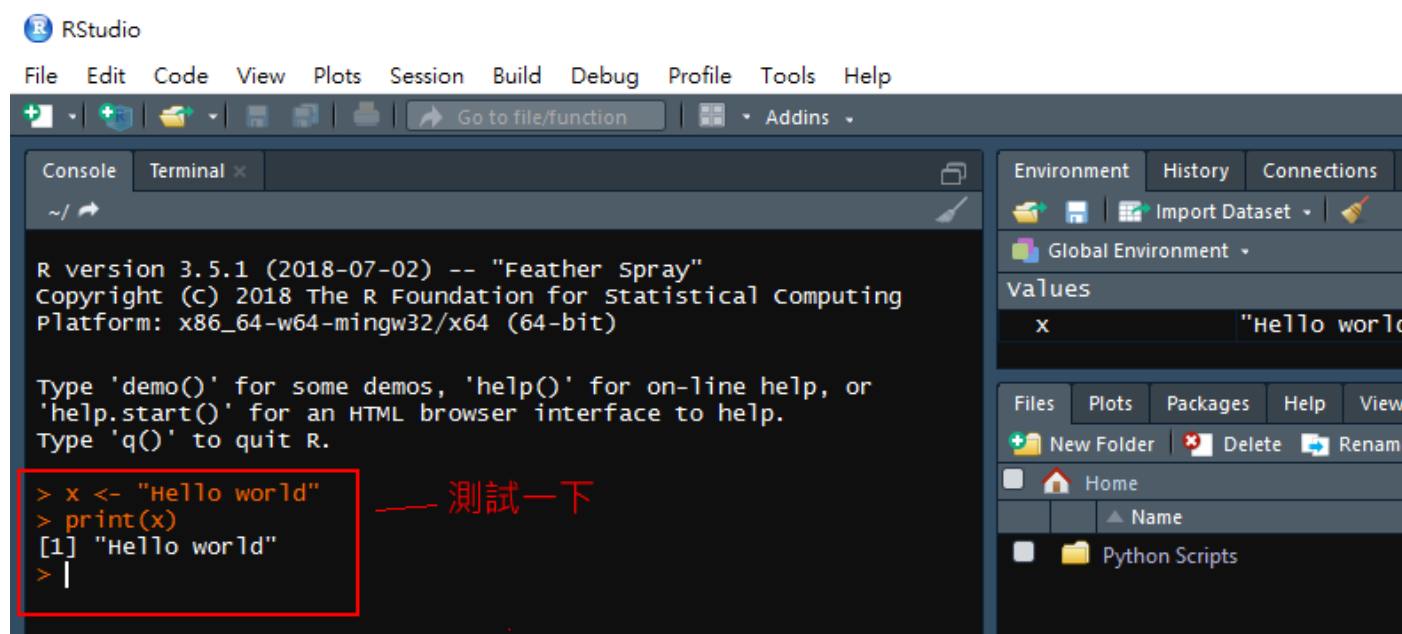
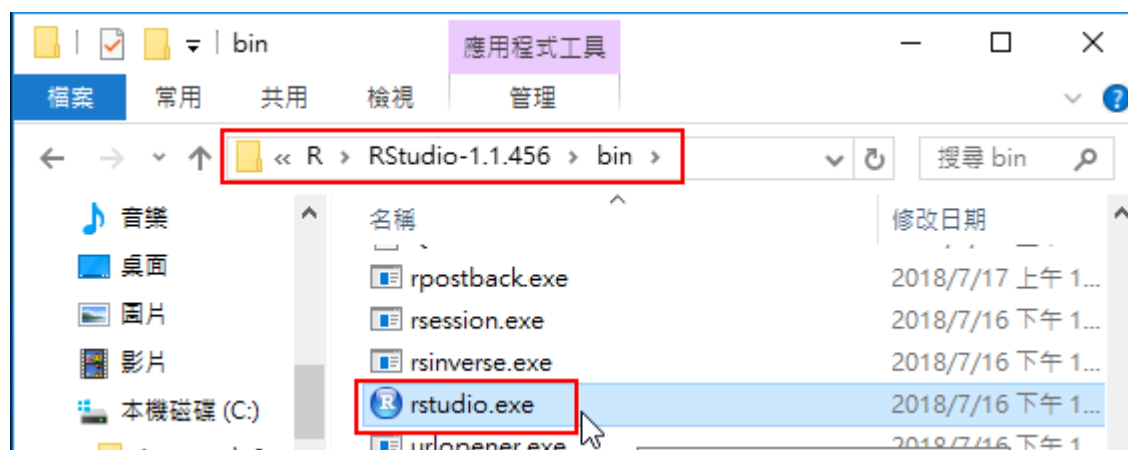


在 C 磁碟新增一個 RDATA 目錄

我個人習慣將 R 相關程式碼都放到 RDATA 目錄下



在 C:\R\RStudio-1.1.456\bin, 雙擊 rstudio.exe, 執行 RStudio



R & RStudio 安裝正常, 關閉 RStudio。

透過 R 語言套件 IRkernel(Interactive R Kernel)
將 R 語言新增至 Jupyter Notebook 的運算核心
雙擊命令提示字元,
切換到 C:\R\R-3.5.1\bin 目錄下,
執行 R.exe

```
命令提示字元
Microsoft Windows [版本 10.0.14393]
(c) 2016 Microsoft Corporation. 著作權者保留。
C:\Users\admin>cd \
C:\>cd r
C:\R>cd r*
C:\R\R-3.5.1>cd bin
C:\R\R-3.5.1\bin>
```

```
C:\R\R-3.5.1\bin>dir
磁碟區 C 中的磁碟沒有標籤。
磁碟區序號: F830-619B

C:\R\R-3.5.1\bin 的目錄
2018/08/26 上午 10:47 <DIR> .
2018/08/26 上午 10:47 <DIR> ..
2018/07/02 上午 10:04 11,377 config.sh
2018/07/02 上午 10:07 88,576 R.exe
2018/07/02 上午 10:07 88,576 Rscript.exe
2018/08/26 上午 10:47 <DIR> x64
3 個檔案 188,529 位元組
3 個目錄 82,925,142,016 位元組可用

C:\R\R-3.5.1\bin>R.exe
```

進入 R 命令模式, 輸入 `install.packages('devtools')`

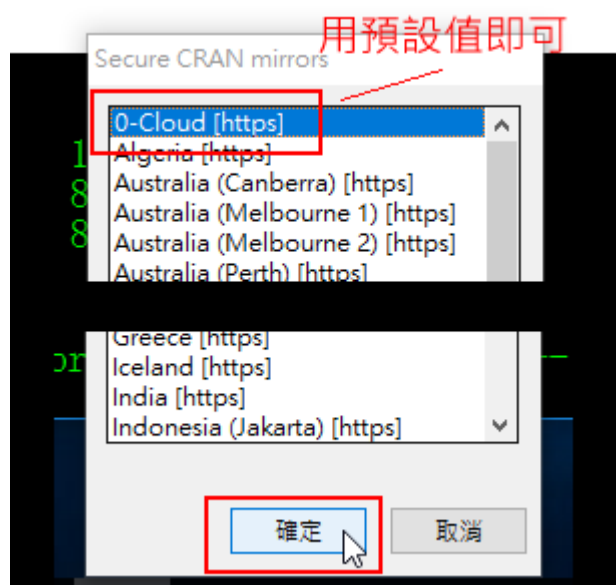
```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R OK0nAOC
bYUziHNC
'license()' 'licence()' oC

R OX@pA\hHXF~mC
'contributors()' pB
'citation()' |iDzpbX~Ta R R MC

'demo()' @d{A 'help()' uWUA
'help.start()' zL HTML sUC
'q()' } RC

> install.packages('devtools')
```



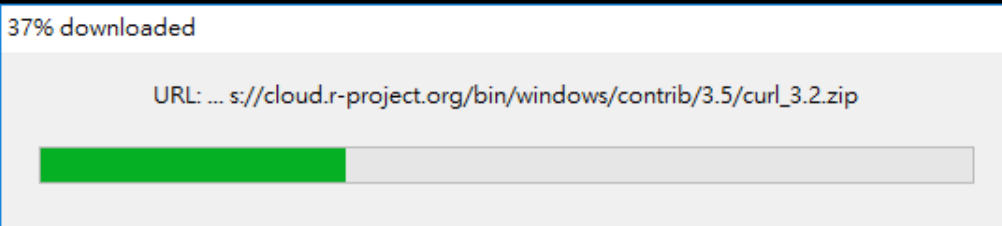
系統會由 mirrors site 安裝相關套件。

```
'demo()' @d{A 'hel
'help.start()' zL
'q()' } RC

> install.packages(
--- Please select a
also installing the dependencies 'mime', 'curl', 'openssl', 'R6', 'httr',
r', 'digest', 'rstudioapi', 'jsonlite', 'git2r', 'withr'

URL 'https://cloud.r-project.org/bin/windows/contrib/3.5/mime_0.5.zip'
Content type 'application/zip' length 46959 bytes (45 KB)
downloaded 45 KB

URL 'https://cloud.r-project.org/bin/windows/contrib/3.5/curl_3.2.zip'
Content type 'application/zip' length 2986282 bytes (2.8 MB)
```



由 GitHub 下載 IRkernel 套件, 輸入 **devtools::install_github('IRkernel/IRkernel')**

```
package 'jsonlite' successfully unpacked and MD5 sums checked
package 'git2r' successfully unpacked and MD5 sums checked
package 'withr' successfully unpacked and MD5 sums checked
package 'devtools' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\TEMP\RtmpEn4rPT\downloaded_packages

> devtools::install_github('IRkernel/IRkernel')
```

安裝完成, 如下圖所示

```
main                                     html
** building package indices
** testing if installed package can be loaded
* DONE (IRkernel)
Warning message:
GitHub repo contains submodules, may not function as expected!
```

將 R Kernel 安裝到 Jupyter Notebook, 輸入 **IRkernel::installspec()**

```
* DONE (IRkernel)
Warning message:
GitHub repo contains submodules, may not function as expected!

> IRkernel::installspec()
[InstallKernelSpec] Installed kernelspec ir in C:\Users\admin\AppData\Roaming\jupyter\kernel
s\ir
```


輸入 `q()`，不儲存工作空間，完成新增 R 語言運算核心設定。

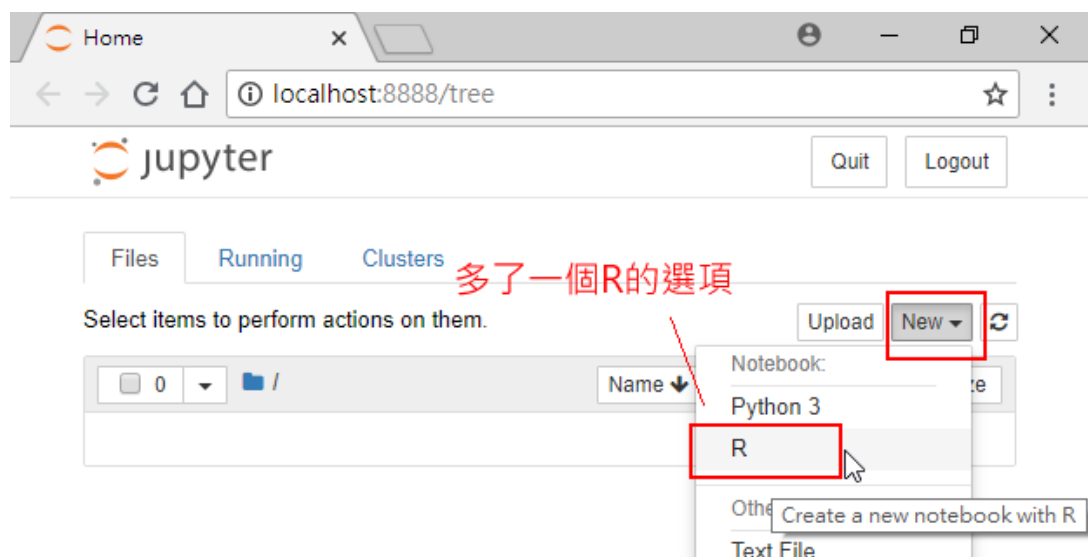
```
> IRkernel::installspec()
[InstallKernelSpec] Installed kernelspec ir in C:\Users\ac...s\ir
>
> q()
儲存工作空間圖案? [y/n/c]: n
```

切換到 `C:\RDATA` 目錄下，啟動 Jupyter Notebook

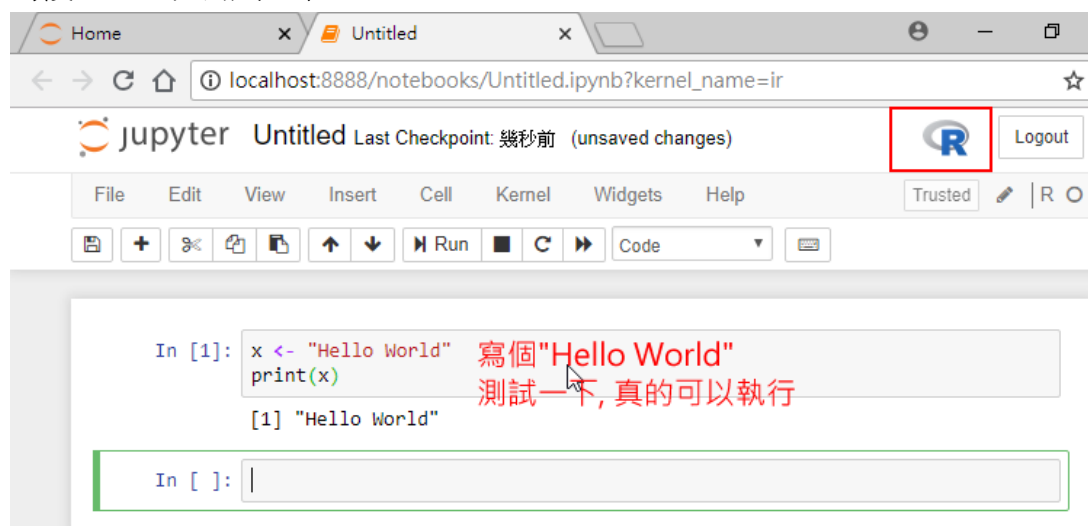
```
> q()
儲存工作空間圖案? [y/n/c]: n
C:\R\R-3.5.1\bin>cd c:\rdata
c:\RDATA>jupyter notebook
```

切換到C:\RDATA目錄
啟動jupyter notebook

點 New，清單中多了一個 R 語言，如下圖所示。



寫段 R Code，測試一下



Iris 為 R 內建資料集

Iris資料集為R內建，現在來測試一下

```
In [2]: summary(iris)
```

```
Sepal.Length    Sepal.Width    Petal.Length    Petal.Width
Min.   :4.300    Min.   :2.000    Min.   :1.000    Min.   :0.100
1st Qu.:5.100    1st Qu.:2.800    1st Qu.:1.600    1st Qu.:0.300
Median :5.800    Median :3.000    Median :4.350    Median :1.300
Mean   :5.843    Mean   :3.057    Mean   :3.758    Mean   :1.199
3rd Qu.:6.400    3rd Qu.:3.300    3rd Qu.:5.100    3rd Qu.:1.800
Max.   :7.900    Max.   :4.400    Max.   :6.900    Max.   :2.500

Species
setosa   :50
versicolor:50
virginica :50
```

```
In [3]: plot(iris)
```

