Installation and Introduction to Jupyter & RStudio

CSE 4/587 Data Intensive Computing Spring 2017 Prepared by Jacob Condello

1 Anaconda/Jupyter Installation

1.1 What is Anaconda?

Anaconda is a free mium open source distribution of the Python and R programming languages for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment.^[1]

It also is the recommended installation method for Jupyter.^[2]

1.2 What is Jupyter?

The Jupyter Notebook is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text.^[3]

I have tested and verified installing Anaconda and launching Jupyter on all three platforms using the Python 3.5 version.

1.3 Install Anaconda

The Anaconda installer can be found at this link: https://www.continuum.io/downloads

O Download Anaconda Now! C: ×		θ
$\epsilon ightarrow {f C}$ $\begin{tabular}{lllllllllllllllllllllllllllllllllll$		* 🖬 🚱 :
	Anaconda changelog	
	Download for Windows Download for OSX	Download for Linux
	Anaconda 4.2.0	Python 3.5 version
	For OSX	GRAPHICAL INSTALLER (407M)
	Anaconda is BSD licensed which gives you permission to use Anaconda commercially and for redistribution.	COMMAND-LINE INSTALLER (349M)
	Graphical Installer	64-Bit
	 Download the graphical installer Double-click the downloaded .pkg file and follow the instructions 	Python 2.7 version
7.4	Command Line Installer	GRAPHICAL INSTALLER (403M)
	Download the command-line installer Optional: Verify data integrity with MD5 or SHA-256 Mo	COMMAND-LINE INSTALLER (346M)
	instructions: Python 3.5 version	64-Bit
	bash Anaconda3-4.2.0-MacOSX-x86_64.sh	
	Python 2.7 version	
	bash Anaconda2-4.2.0-MacOSX-x86_64.sh	
	NOTE: Include the "bash" command even if you are not using the bash s	hell.

1.3.1 Windows

Download GUI installer for Python 3.5 and follow on-screen instructions choosing to add the install location to your PATH

1.3.2 Mac OS X

Download GUI installer for Python 3.5 and follow on-screen instructions

1.3.3 Linux

Download the python 3.5 version and run the following:

```
bash Anaconda3-4.2.0-Linux-x86_64.sh
```

Follow on screen prompts, hitting enter when necessary Select yes for prepending install location to PATH in .bashrc

1.4 Installing the R kernel

Jupyter comes installed with a couple python kernels. The easiest way to get R in Jupyter is through conda, which is the package manager used by Anaconda. The package we will be installing is called r-essentials which includes 80 of the most used R packages for data science.^[4] To Install the r-essentials package into the current environment, first restart the terminal then run:

conda install -c r r-essentials

1.5 Running Jupyter

The following command starts a local server and opens a browser window listing all files in the current working directory, so run it in a directory where you wish to save your notebooks:

jupyter notebook

Select New in the upper right, then select R under Notebooks

O Mome x C Home x	Θ
$\leftarrow \rightarrow \mathbb{C}$ (\bigcirc localhost.8888/tree	≪☆ 🖬 🚺 :
💆 jupyter	
Files Running Clusters Conda	
Select items to perform actions on them.	Upload New - 2
	Text File
Notebook list empty.	Folder
	Terminal
	Notebooks
	Python [conda root]
	Python [default]

1.6 Using Jupyter

1.6.1 Intro

Each new notebook starts of with one cell. Cells can be in one of two states: command mode or edit mode. To go into edit mode press the enter key or click on a cell. Notice it turned green. To return to command mode press esc. Notice it turned blue.

While a cell is selected, it can be run by selected "Cell ->Run Cell and Select Below". To avoid doing this every time the shortcut can be used. To look at the shortcuts select "Help ->Keyboard Shortcut". Notice after running a cell the number in the square brackets is populated. This is to indicate the order of execution as cells do not need to be run top to bottom. When there is an asterisk in the square brackets that means the cell is still running.

To rename a notebook you can click the current title, usually "Untitled" by default.

1.6.2 R Example

Try generateing a plot by typing the commands below, one per cell:

library(datasets)		#to inc	lude a	few	provid	ed d	latase	ts e.	.g.	mtcar	s			
mtcars		#should	print	the	datafr	ame	with	rows	as	cars	and	columns	as	specs
<pre>plot(x=mtcars\$wt,</pre>	y=mtcars\$mpg)	#plot m	iles p	er g	allon v	vs w	eight	for	all	cars				

1.6.3 Markdown

Markdown cells are useful for providing information to the reader in between code. To change the cell to a markdown cell select "Cell ->Cell Type ->Markdown". Markdown is a language that is a superset of HTML.^[5] Try typing in the following to see some of the possibilities:

```
## Header
# Bigger header
Unordered List
* Item 1
* Item 2
* Item 2
* Item 3
Ordered list
1. Item 1
2. Item 1
2. Item 2
3. Item 3
```


• • • / C Home	x C Untitled x	Θ
$\leftrightarrow \rightarrow \mathbf{C}$ (localhost:8	888/notebooks/Untitled.ipynb	@ # 🗊 🗘 :
	JUPYTET Untitled Last Checkpoint: 2 hours ago (unsaved changes)	R
	File Edit View Insert Cell Kernel Widgets Help	RO
	B + 3< 2 B ↑ ↓ H B C Code + B CellToolbar Δ 0 D	
	mtcars\$wt	
	Header	
	Bigger header	
	Unordered List	
	Item 1	
	• Item 3	
	Ordered list	
	1. Item 1	
	2. Item 2 3. Item 3	
	Bold Text	
	R	
	In []:	

1.6.4 Exporting

To export select "File ->Download as ->…"

••• C Home x C Untitled	×	θ
← → C ③ localhost:8888/notebooks/Untitled.ipynb		@☆ 🖬 🚺 :
🔁 jupyter	Untitled Last Checkpoint: 2 hours ago (autosaved)	R
File Edit View	r Insert Cell Kernel Widgets Help	RO
New Notebook	↑ ↓ H ■ C Code CellToolbar	
Open	mtcarsąwt	
Make a Copy Rename Save and Checkpoin	ler	
Revert to Checkpoin	' jer header	
Print Preview	ed List	
Download as	Notebook (.ipynb)	
Trusted Notebook	R (.r) HTML (.html)	
Close and Halt	Markdown (.md) reST (.rst)	
1. 2. 3.	terr PDF via LaTeX (pdf) terr Presentation (.html) terr	
Bold	Text	
In ():	R	
loca host 8888/notebooks/Untitled /pynb#		

Note: It appears pandoc and Latex need to be installed along with several packages to export to pdf. I was able to get it to work on all platforms with varying degree of difficulty. These links should provide a good start should you need to export to pdf:

https://www.latex-project.org/get/

http://pandoc.org/installing.html

2 RStudio Installation

2.1 What is RStudio?

RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics.^[6]

2.2 Install RStudio

Installing RStudio is straight forward. Just download and run the installer at this link: https://www.rstudio.com/products/rstudio/download/

O Ownload RStudio - RStudio ×			e
\leftarrow \rightarrow \mathfrak{C} $\begin{bmatrix}$ a Secure $\end{bmatrix}$ https://www.rstudio.com/products/rst-	udio/download/		☆ 🗟 😭 :
	R Studio rstudio	conf Products Resources Pricing About Us Blogs Q	
	Installers for Supported Platforms		
	Hatalies Ritudio 1.0.136- Window Vista/7/0/10 Ritudio 1.0.136- Wat CSX 126- (6-6-30) Ritudio 1.0.136- Viotna 126-4/ (6-6-30) Ritudio 1.0.136- Roton 19/ Ritudio 7 / hoper/SDZ 11.1 (6-6-30) Ritudio 1.0.136- Roton 19/ Ritudio 7 / hoper/SDZ 11.1 (6-6-30)	Size Date MOS R1.9140 2016-12-21 392.81267975597.33177.646364611.40993ba R1.2140 2016-12-21 322.6126712517.84636461 R0.5140 2012-12-21 322.612643423 R0.5140 2012-12-21 322.61267124 R0.5140 2012-12-21 32.61267145 R0.7140 2016-12-21 AntiParticipart	
	Zip/Tarballs		
	Zipytar archives Risolatio 1.0.136 - Windows Vistar/18/10 Risolatio 1.0.136 - Wuntu 1.2.044/Debian 8+ (E2-bit) Risolatio 1.0.136 - Wondows 1944-Refshort 74 percentization Risolatio 1.0.136 - Redox 194/Refshort 74-percentizati 1.1.4+ (E2-bit) Risolatio 1.0.136 - Redox 194/Refshort 74-percentizati 1.1.4+ (E4-bit)	Size Date M65 1175480 2016-124 4513598356120ebab127e7efbb6969bbe 662.00 2016-122 fcs379593bd612596bb127e7efbb6967b403bbc27b4135bc12145 052.00 2016-122 0cs092593bd612596bb125bc125bc125bc125bc125bc125bc125bc125	
	Source Code		
	A tarball containing source code for RStudio v1.0.136 ca	i be downloaded from here	
	250 Northern Ave, Boston, MA 02210 844-448-1212 info@studio.com	DMCA Missed #studioconf? Here are some tips from IDE engineer Support @Revin_usityers (ides from all	
	R Studio	ECCN talks forthicoming, #rstats ECCN twicks 1 week ago	
		🛩 in f እ 🖂 ር እ	
https://download1.rstudio.org/RStudio-1.0.136.exe			^

2.3 Using RStudio

RStudio includes four windows: Script (top left), Console (bottom left), Plot/Help (bottom right), and Workspace (top right). If you only see 3 you may need to select "File ->New File ->R Script". To try a similar example as before you can type the following in a script. To run it select "Source":

library(datasets)#to include a few provided datasets e.g. mtcarsmtcars#should print the dataframe with rows as cars and columns as specsplot(x=mtcars\$wt, y=mtcars\$mpg)#plot miles per gallon vs weight for all carsx <- rnorm(100)</td>#x contains 100 random numbers following a standard normal distribution



Notice 1) the output generated from the script in the console window 2) the plot generated in the plot/help window 3) the value of x in the workspace window

3 Sources

- [1] https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)
- [2] http://jupyter.org/install.html
- [3] http://jupyter.org/
- $[4] \ https://conda.io/docs/r-with-conda.html$

 $[5] \ http://jupyter-notebook.readthedocs.io/en/latest/examples/Notebook/Working\%20With\%20Markdown\%20Cells. html$

 $[6] \ {\tt https://en.wikipedia.org/wiki/RStudio}$