Installation and Introduction to Jupyter & RStudio

CSE 4/587 Data Intensive Computing
Prepared by Jacob Condello (UTA) Spring 2017, updated 2019

1 Anaconda/Jupyter Installation

1.1 What is Anaconda?

Anaconda is a freemium 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]

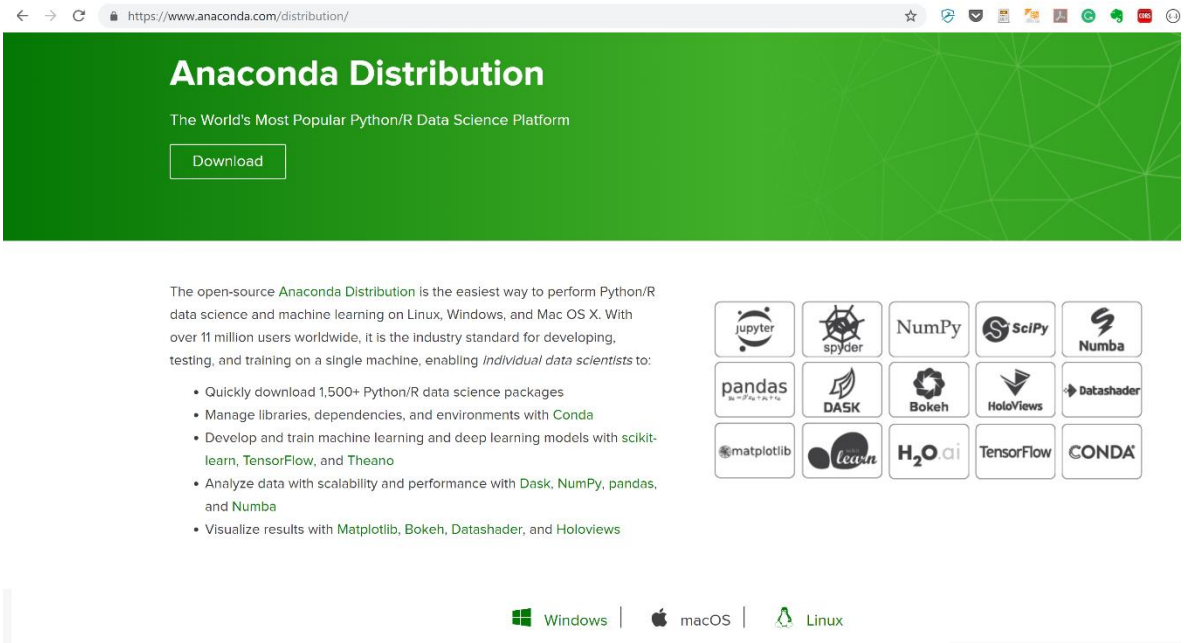
1.2 What is Jupyter?

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.anaconda.com/distribution/
More information on installing can be found at this link: https://docs.anaconda.com/anaconda/install/



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

Link for installing on Windows: https://docs.anaconda.com/anaconda/install/windows/

1.3.2 Mac OS X

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

Link for installing on Mac: https://docs.anaconda.com/anaconda/install/mac-os/

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

Link for installing on Linux: https://docs.anaconda.com/anaconda/install/linux/

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
```

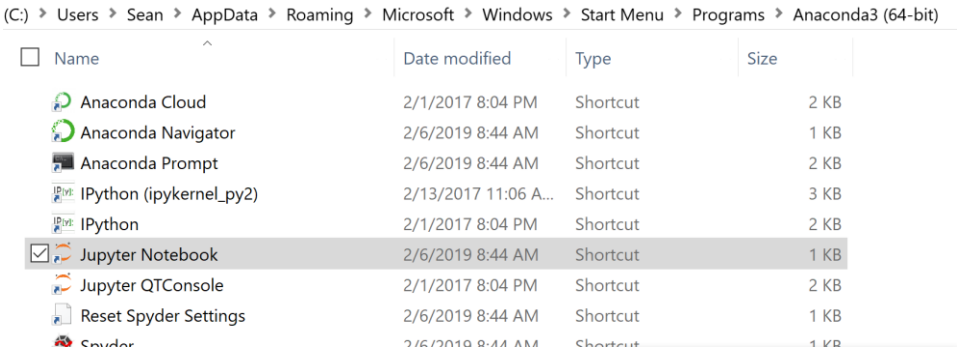
Link for information on the command: https://docs.anaconda.com/anaconda/packages/r-language-pkg-docs/

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
```

An alternative is to directly go to the directory and select Jupyter notebook application.

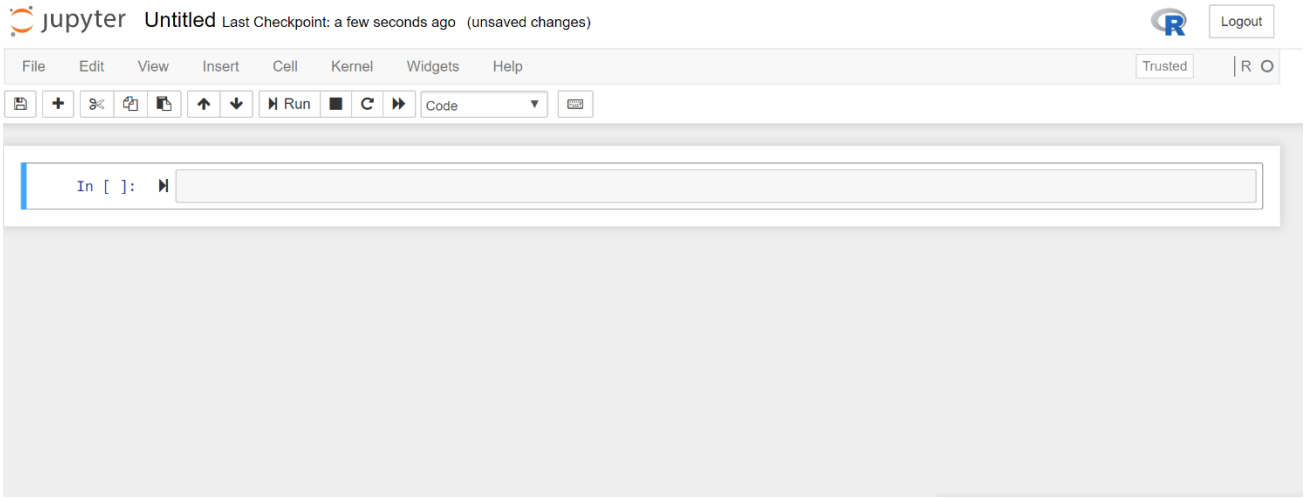


Select New in the upper right, then select R under Notebooks

1.6 Using Jupyter

1.6.1 Intro

Each new notebook starts 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:

jupyter Untitled Last Checkpoint: 6 minutes ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted R

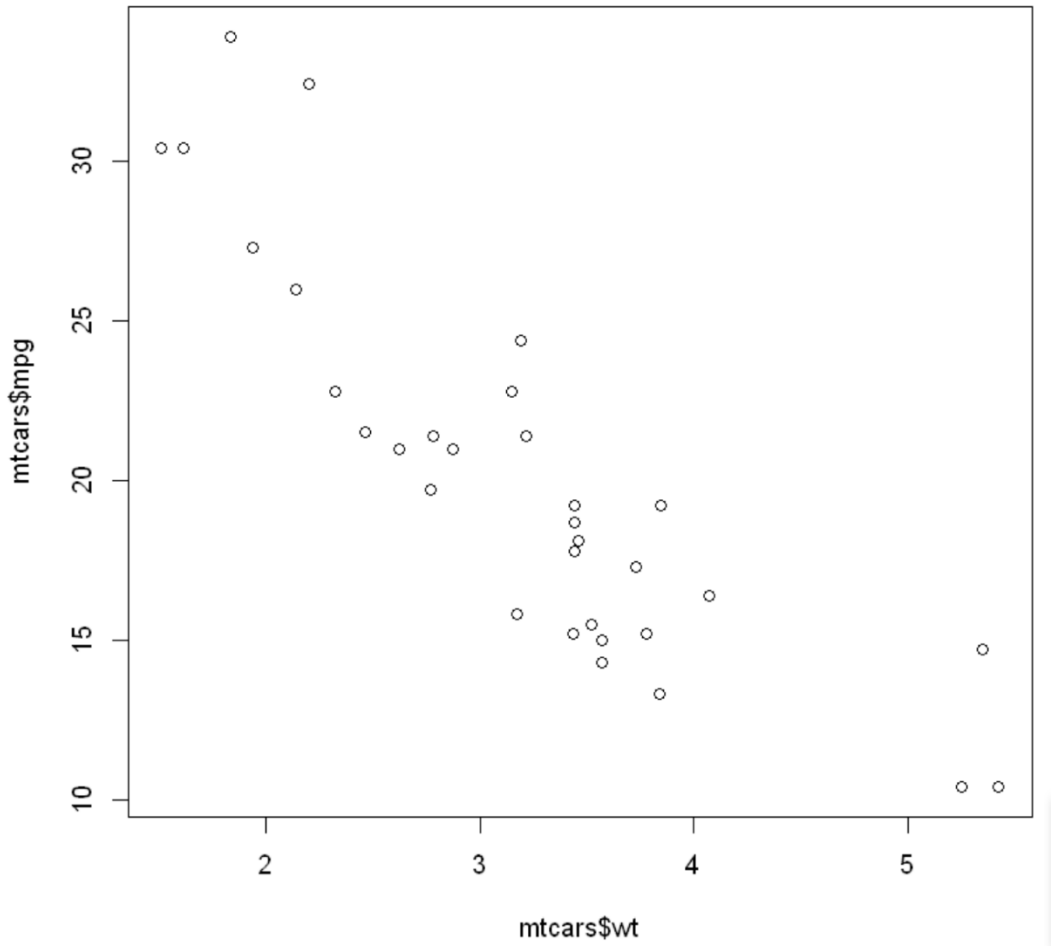
Run Code

In [2]:
```
library(datasets)  #to include a few provided datasets e.g.  mtcars
```

In [3]:
```
mtcars  #should print the dataframe with rows as  cars and columns as specs
```

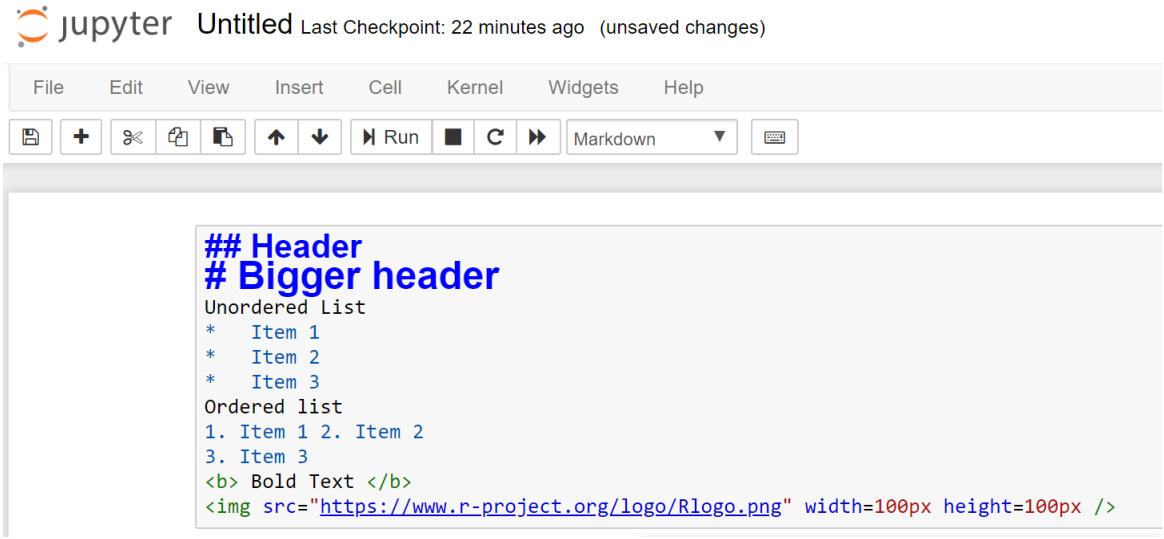
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4

In [4]:
```
plot(x=mtcars$wt, y=mtcars$mpg) #plot miles per gallon vs weight for all
```

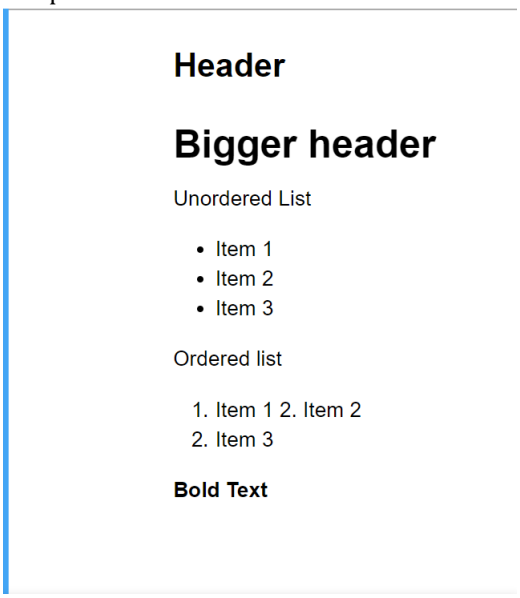


1.6.3 Markdown

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:

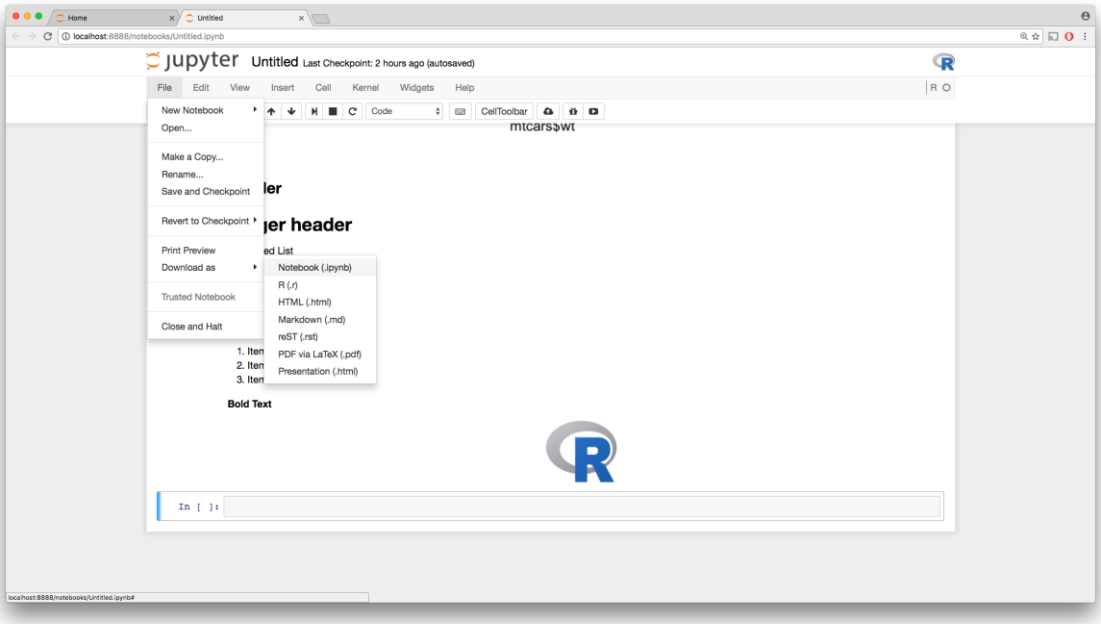


The output for the above code is as follows:



1.6.4 Exporting

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



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 di culty. 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`

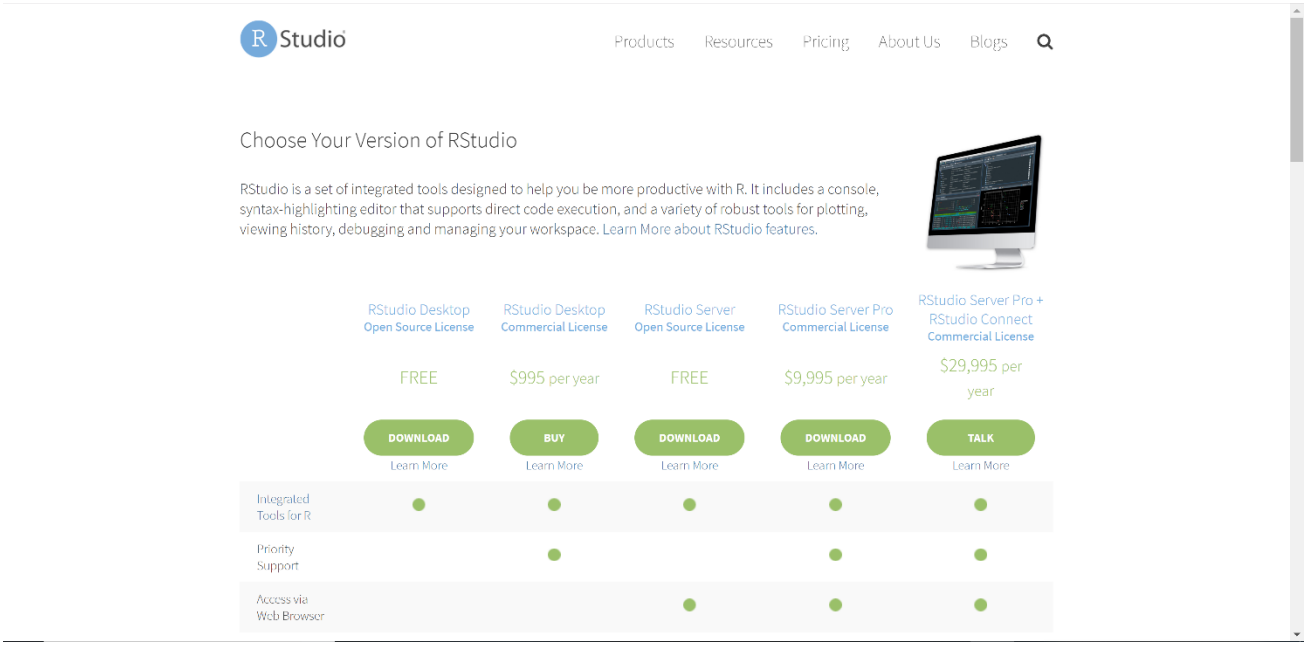
2 RStudio Installation

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

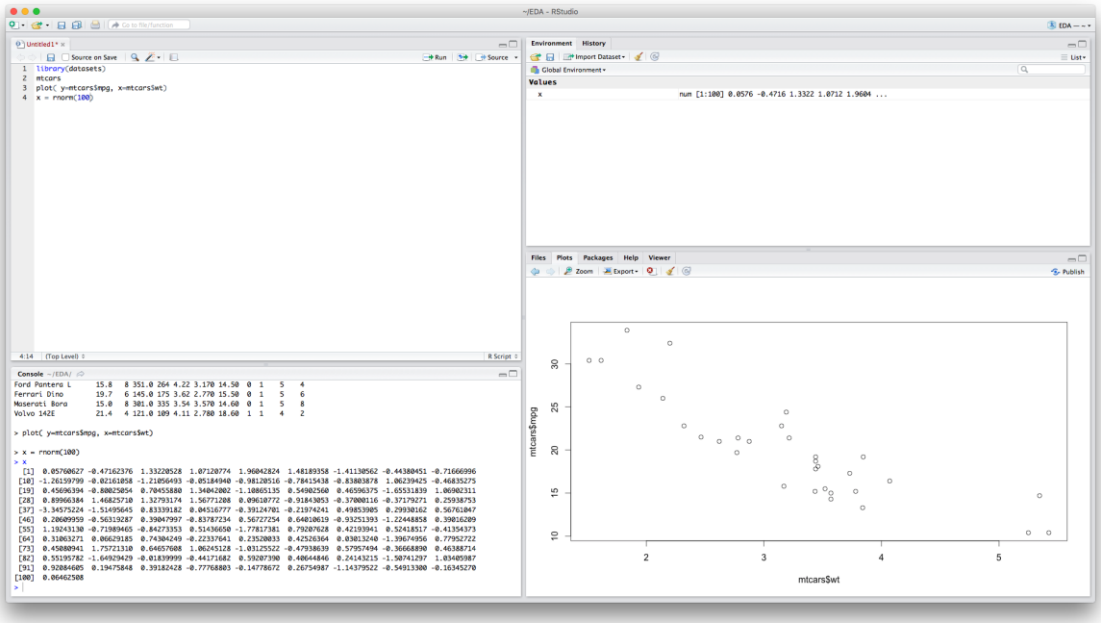
Installing RStudio is straight forward. Just download and run the installer at this link: https://www.rstudio.com/products/rstudio/download/



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. mtcars
mtcars               #should print the dataframe with rows as cars and columns as specs
plot(x=mtcars$wt, y=mtcars$mpg) #plot miles per gallon vs weight for all cars
x <- rnorm(100)      #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

Note: In certain cases you might get a error saying: Error in plot.new() : figure margins too large
What this error is telling you is that you need to expand your plot 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] https://en.wikipedia.org/wiki/RStudio