COMP41680

Preliminary Material

Derek Greene

UCD School of Computer Science

Spring 2017

Overview

- Module Details
- Why Python?
- Installing Python 3 via Anaconda
- Running Python Code
- Using IPython Notebooks
  - Getting Started
  - Code Cells
  - Markdown Cells

COMP41680 Module Outline

- **Weeks 1-3:** Crash course in Python 3
  - Working with IPython Notebooks
  - Language fundamentals
  - Data structures
  - Input/output
- **Weeks 4-12:** Practical Data Science in Python
  - Introduction to data science
  - Accessing and preparing data
  - Numerical computing and statistics in Python
  - Machine learning in Python
  - Data visualisation

COMP41680 Practical Details

- Lectures Mondays 12-1pm B002.
- Labs Wednesdays 11am-12pm B002. **Laptop required.**
- Moodle page currently open for registration via self-enrolment. Password is "**py2017ds**": https://csmoodle.ucd.ie/moodle/course/view.php?id=474
- Module marked based on 2 continuous assessment assignments. No end of semester exam.

50%	Assignment 1: Data Collection & Preparation
50%	Assignment 2: Data Exploration & Machine Learning

! All assignment deadlines are **hard deadlines.**

! Plagiarism will be treated seriously. Any evidence of plagiarism in an assignment will result in a **0 mark**.

Why Python?

- Open source, freely available online
- Clean, concise, unambiguous syntax
  - Often referred to as "executable pseudo-code"
- Supports Rapid Application Development
- Supports a variety of programming paradigms
  - Simple scripts
  - Modular & Object-oriented programming
  - Interactive notebooks
- Strong library support
  - Comprehensive built-in library provides many functions
  - Many third-party packages available, particularly for numerical computing, data analysis, and visualisation.

Installing Python

- In the module we will use **Python 3.5**
- Python 3.x is recommended for new code and fixes many of the issues and inconsistencies from Python 2.
- Be aware: Python 3.x code is not fully backwards compatible with Python 2.
- Install Python via the **Anaconda** distribution which provides a version of Python tailored for data analytics, with easy installation of many third party packages.

  https://www.continuum.io/downloads

- Download and run Anaconda the graphical or terminal installer for Python 3.5 (not 2.7!) for your operating system - Windows, OSX or Linux.

Running Python Code

Several different ways to run Python code…

1. Type `python` at the terminal to start the basic Python interactive shell.

```
> python
Python 3.5.2 |Anaconda 4.2.0 (x86_64)| (default, Jul  2 2016, 17:52:12)
[GCC 4.2.1 Compatible Apple LLVM 4.2 (clang-425.0.28)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> 10 * 20
200
>>>
```

2. Type `ipython` at the terminal to start the enhanced IPython interactive shell.

```
> ipython
Python 3.5.2 |Anaconda 4.2.0 (x86_64)| (default, Jul  2 2016, 17:52:12)
Type "copyright", "credits" or "license" for more information.

IPython 5.1.0 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]: 10 * 20
Out[1]: 200

In [2]:
```

3. Run full script files line by line from the terminal using:

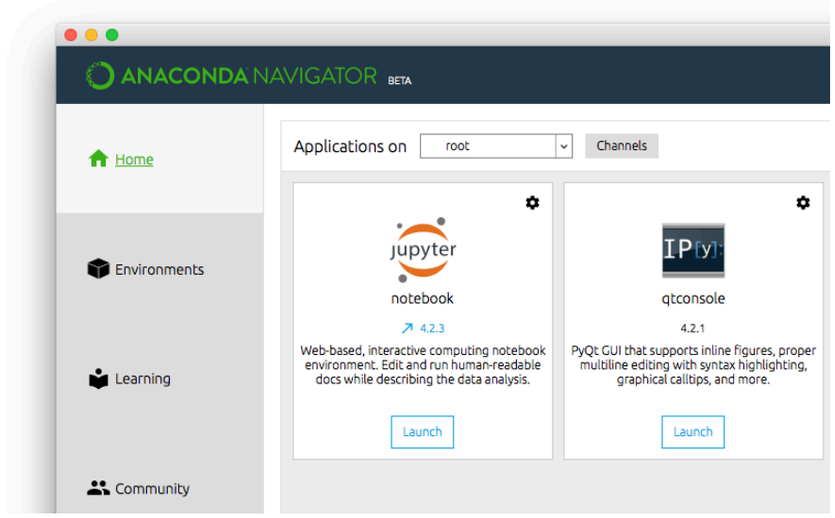
```
python <script_file.py>
```

```
> python hello.py
Hello World
>
```

4. Use web-based interactive notebooks…

Jupyter & IPython Notebooks

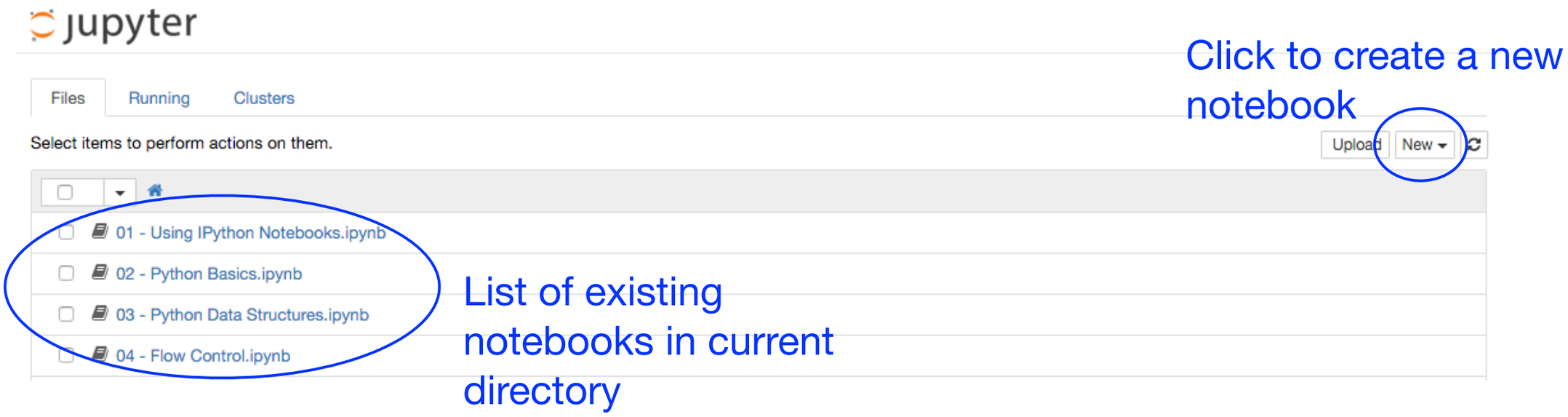
- The Jupyter project is a web application for interactive data science and scientific computing.
- IPython Notebooks provide an engine for running Python code under Jupyter.
- We will use IPython Notebooks for many of the labs and assignments in COMP41680.
- To start the Notebook server, either:
  1. In the terminal, type `jupyter notebook`
  2. Click the Anaconda Navigator icon, then choose **jupyter-notebook** from the list of apps.



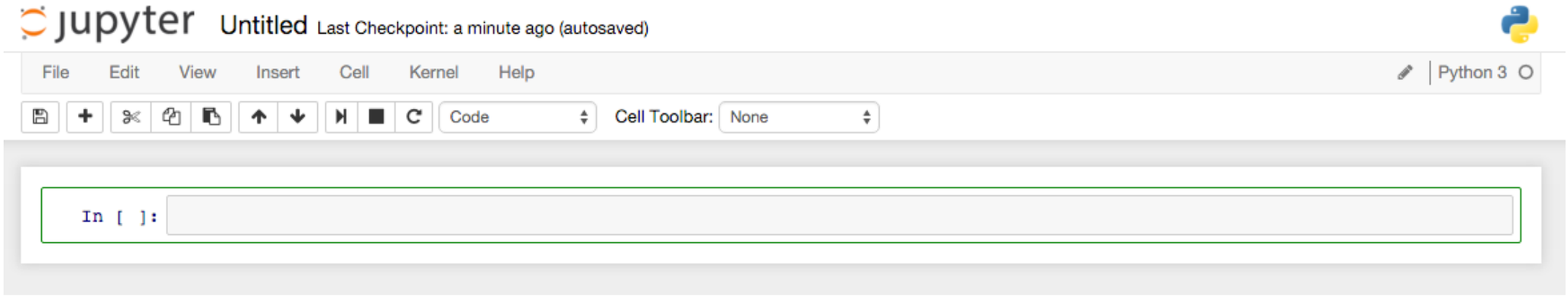
- This should load the IPython Notebook dashboard in your browser. Later you can also manually go to http://localhost:8888

Notebook Dashboard

- The IPython dashboard provides a mini filesystem interface for creating and accessing notebooks.
- Note: The dashboard shows notebooks in the directory where you launched the notebook server.

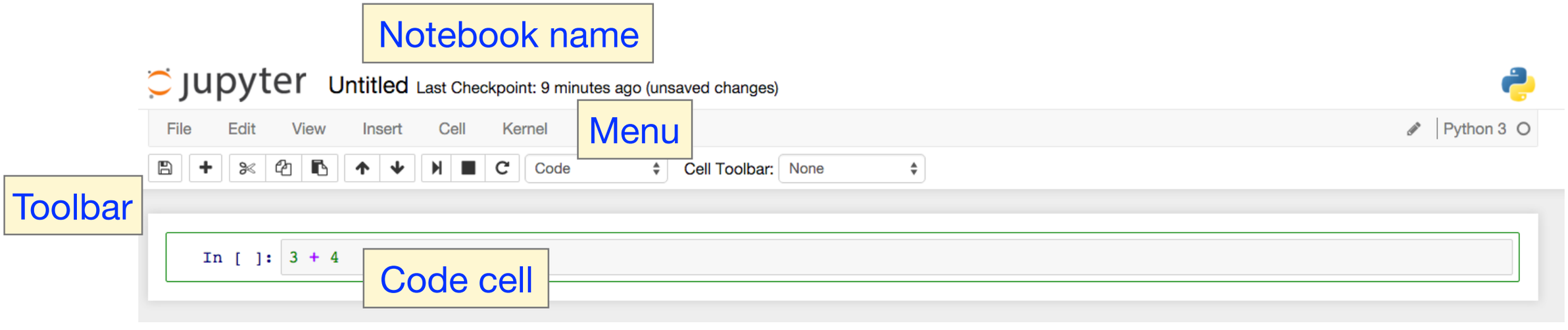


- To start writing code, create **New → Python 3 Notebook**

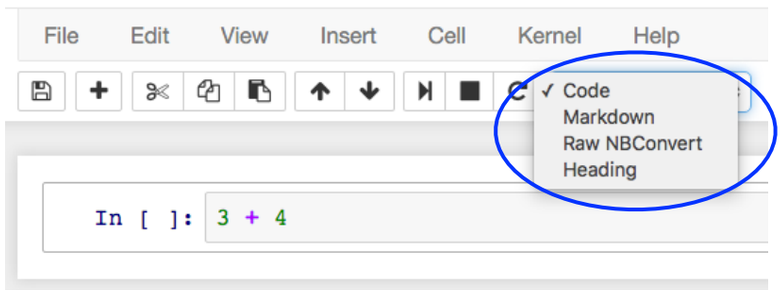


Notebook Interface

- When you create a new notebook, you will be presented with the notebook name, a menu bar, a toolbar and an empty code cell.



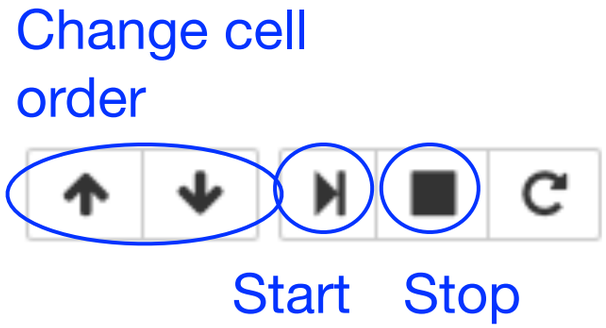
- IPython notebooks have two fundamental types of cells:
  1. **Markdown cells:** Contain text content for explaining a notebook.
  2. **Code cells:** Allow you to type and run Python code.



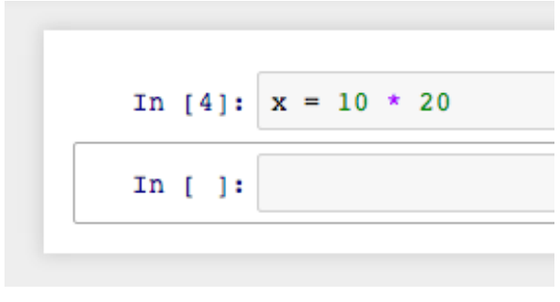
Every new cell starts off being a code cell. But this can be changed by using the drop-down on the toolbar

Code Cells

- In a code cell, you can enter one or more lines of Python code. Run the code by hitting Shift-Enter or by pressing the **Play** button in the toolbar.



- You can modify and re-run code cells multiple times in any order.
- When a code cell is executed, the code it contains is sent to the **kernel** associated with the notebook - i.e. the Python instance running in the background.
- The results returned from this computation are displayed as the cell's output. Note that some code will not have an output.



- Restarting the kernel associated with a notebook clears all previous history (e.g. variable values).

Markdown Cells

- It can be helpful to provide explanatory text in notebooks.
- **Markdown** is a lightweight type of markup language with plain text formatting syntax which can be rendered as HTML.
- IPython supports a set of common Markdown commands. HTML tags and LaTeX formulae can also be included.
- When a Markdown cell is executed, the Markdown code is converted into the corresponding formatted rich text.

```
This is normal text.
```

This is normal text.

```
*This is italics*.
```

This is italics.

```
And **this is bold**.
```

And **this is bold**.

```
# Heading 1
## Heading 2
### Heading 3
```

Heading 1
Heading 2
Heading 3

```
Example <font color='red'>HTML use</font>
```

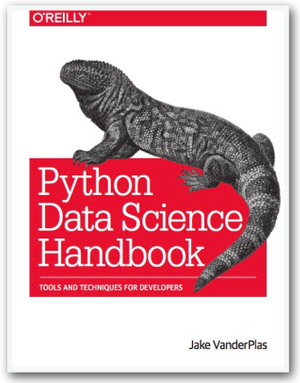
Example HTML use

```
Formula: $x=\frac{y}{z}$
```

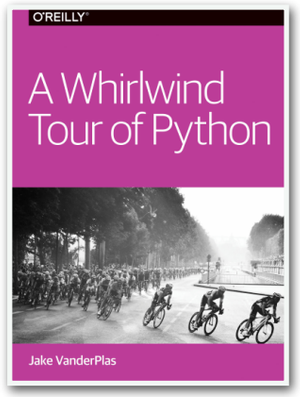
Formula: $x = \frac{y}{z}$

Book Resources

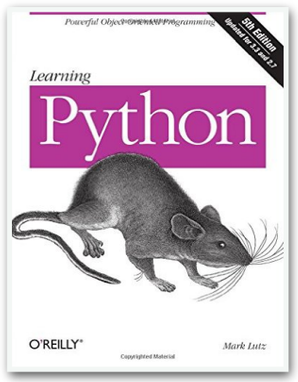
No single textbook for this module. A range of good Python books are available. Make sure the book covers Python 3.x.



Python Data Science Handbook
Jake VanderPlas
http://shop.oreilly.com/product/0636920034919.do



A Whirlwind Tour of Python
Jake VanderPlas
http://www.oreilly.com/programming/free/a-whirlwind-tour-of-python.cs



Learning Python 5th Edition
Mark Lutz
http://shop.oreilly.com/product/0636920028154.do

Online Resources

- **Python**
  - Official Python 3 documentation
    https://docs.python.org/3/
  - SciPy lectures notes
    http://www.scipy-lectures.org
- **IPython Notebooks**
  - Official documentation
    http://ipython.readthedocs.org/en/stable/overview.html
- **Markdown**
  - Github guide to Markdown
    https://help.github.com/articles/markdown-basics
  - Original Markdown syntax specification
    http://daringfireball.net/projects/markdown/syntax/