



python

For Data Science

Structure

- Part 1 - Introduction, recap & warm-up exercises
- Part 2 - NumPy (Numerical Python)
- Part 3 - Basic plotting
- Part 4 - pandas (Python Data Analysis Library)

Ask!

The art and science of asking questions is the source of all knowledge.
- Thomas Berger

- Do not hesitate to ask!
- During exercises (you can also ask others).



Image by [mohamed Hassan from Pixabay](#)

Now let me ask something..

- Why do you want to learn Python/programming?
- What would you use Python for?

Failure

- Coding is all about trial and error.
- Don't be afraid of it.
- Error messages aren't scary, they are useful.



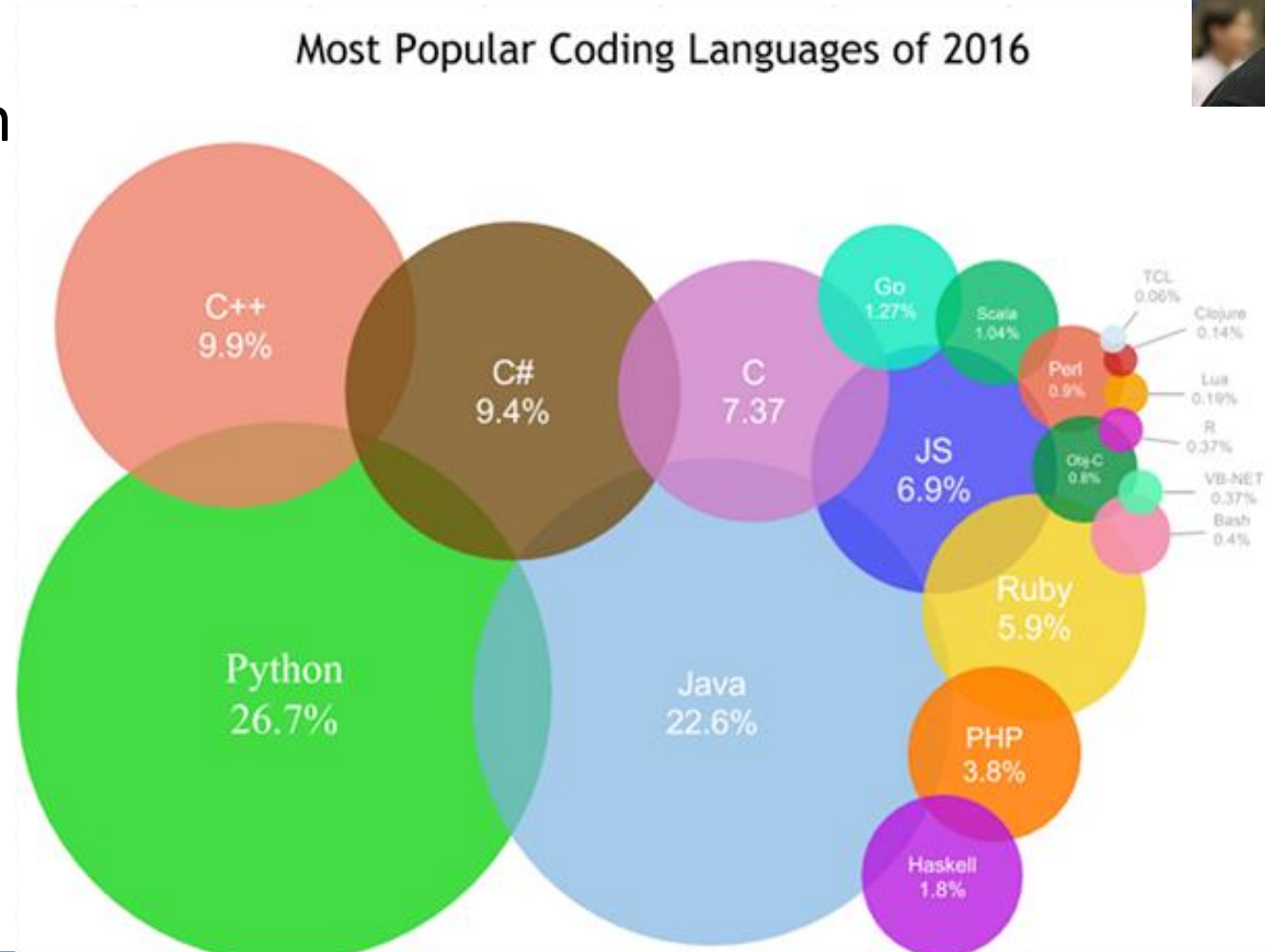
[Python natalensis by A. Smith](#) on Wikimedia Commons

History

- Started by Guido Van Rossum as a hobby
- Now widely spread
- Open Source! Free!
- Versatile



Guido Van Rossum
by [Doc Searls on Flickr](#) CC-BY-SA



Python today

- Developed a large and active scientific computing and data analysis community
- Now one of the most important languages for
 - Data science
 - Machine learning
 - General software development
- Packages: NumPy, pandas, matplotlib, SciPy, scikit-learn, statsmodels

2 Modes

1. IPython

Python can be run interactively

Used extensively in research

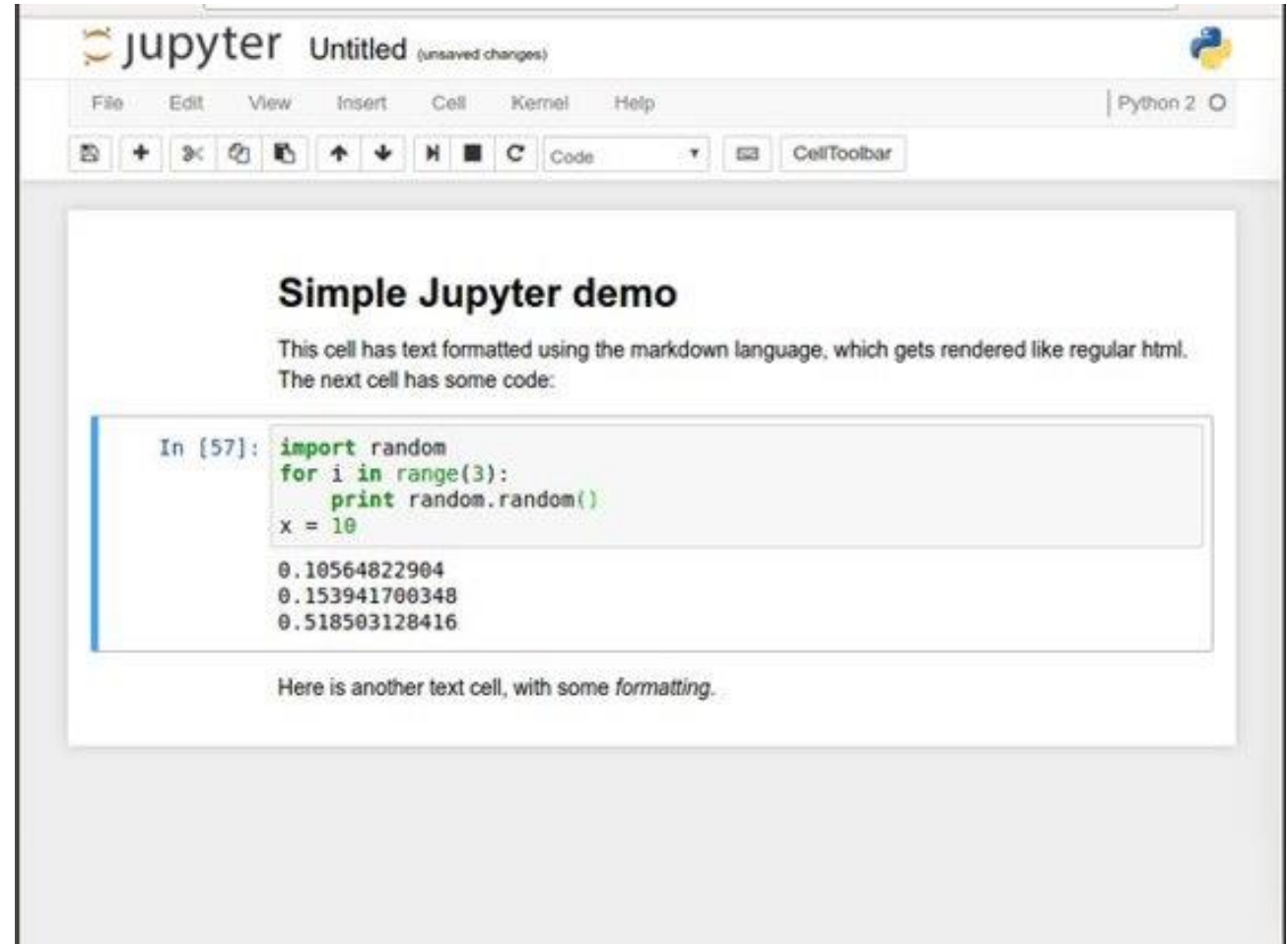
2. Python scripts

What if we want to run more than a few lines of code?

Then we must write text files in .py

Noteable (Jupyter notebooks)

- Easy to use environment
- Web-based
- Combines both text and code into one
- Come with a great number of useful packages



1. Start Noteable

The screenshot shows the 'Python for Data Science' course interface. The left sidebar contains a menu with 'Python for Data Science' and 'Course Content'. The main content area is titled 'Course Content' and displays a 'Welcome' message. Below the welcome message, the 'Noteable' section is highlighted with an orange box. The 'Noteable' section includes a sub-header 'Noteable' and a description: 'Web-based Python development environment. Please select Status'. Below this, the 'Setup Instructions' section is visible, listing steps for opening Noteable and setting up the environment. A modal window titled 'Connect to Noteable' is overlaid on the right side of the screen. This modal contains a form with three fields: 'Send your user ID?' (checked), 'Send your name?' (unchecked), and 'Send your email address?' (unchecked). The 'Submit' button is highlighted with an orange box.

Python for Data Science Course Content

Welcome

Welcome to the Python for Data Science LEARN course, run by the Digital Skills and Training team in Information Services to complement their classroom course. This is an introductory course in using Python for Data Science applications. In the recent years Python has become extremely popular within the data science communities mainly due to its ease of use, open source nature and the fact that it is completely free. This course aims to introduce newcomers to the most popular packages used today - numpy, pandas and matplotlib. Note that it assumes you have some basic knowledge of Python.

This is an extra-curricular course. You won't receive any credits for completing it.

Please note that this course is an *introduction* at best and does not cover all the topics you might need for a job or further study.

This is entirely self-contained and self-paced. You can do it in your own time. It serves the purpose of an introduction into the world of data science.

This course uses the Noteable service as a teaching environment. You will have to upload the course material to Noteable. Please follow the instructions below.

Noteable

Web-based Python development environment. Please select Status

Setup Instructions

To be ready for the course, first you will need to set up your environment. Follow the steps below:

1. Open Noteable by the link provided on this page.
2. You will now be taken to your Jupyter Homepage.
3. Click the button +GitRepo near the top right.
4. Type in `https://git.ecdf.ed.ac.uk/digital-skills/python-data-science` and click OK.

Now return to your Jupyter Homepage you will see a new folder has appeared called `python-data-science`, inside are all files needed for the course!

The file `python-data-jupyter-readme.ipynb` explains how Noteable and Jupyter notebooks work.

Connect to Noteable

* Indicates unsaved changes.

PERSONAL DATA

Send your user ID? ☒

Send your name? ☐
Check the box to send your name to the external tool.

Send your email address? ☐

Click Submit to proceed.

Cancel Submit



Edina

Start Personal Container

Standard Notebook ▼

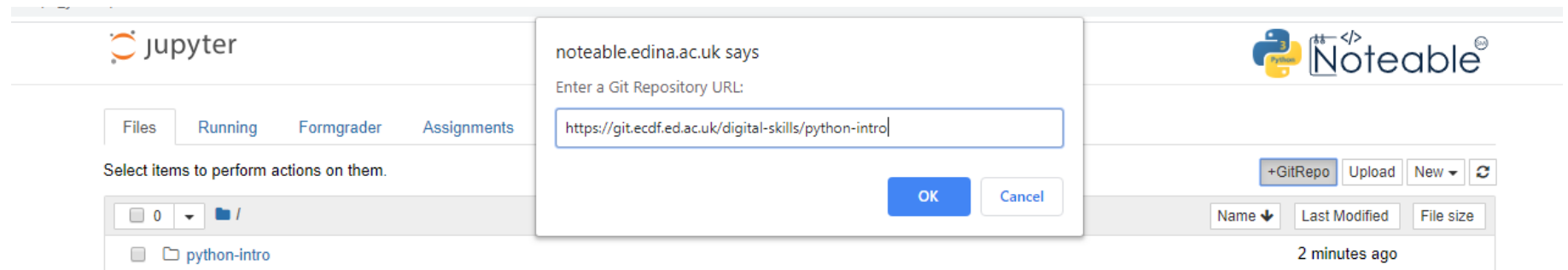
Start

You have a running container:

Reconnect

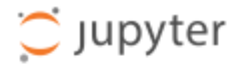
Shut Down

2. Clone GitRepo(recommended)



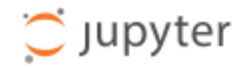
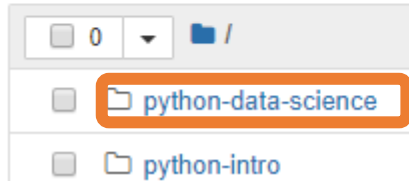
Pull down a Git repository

3. Starting a notebook



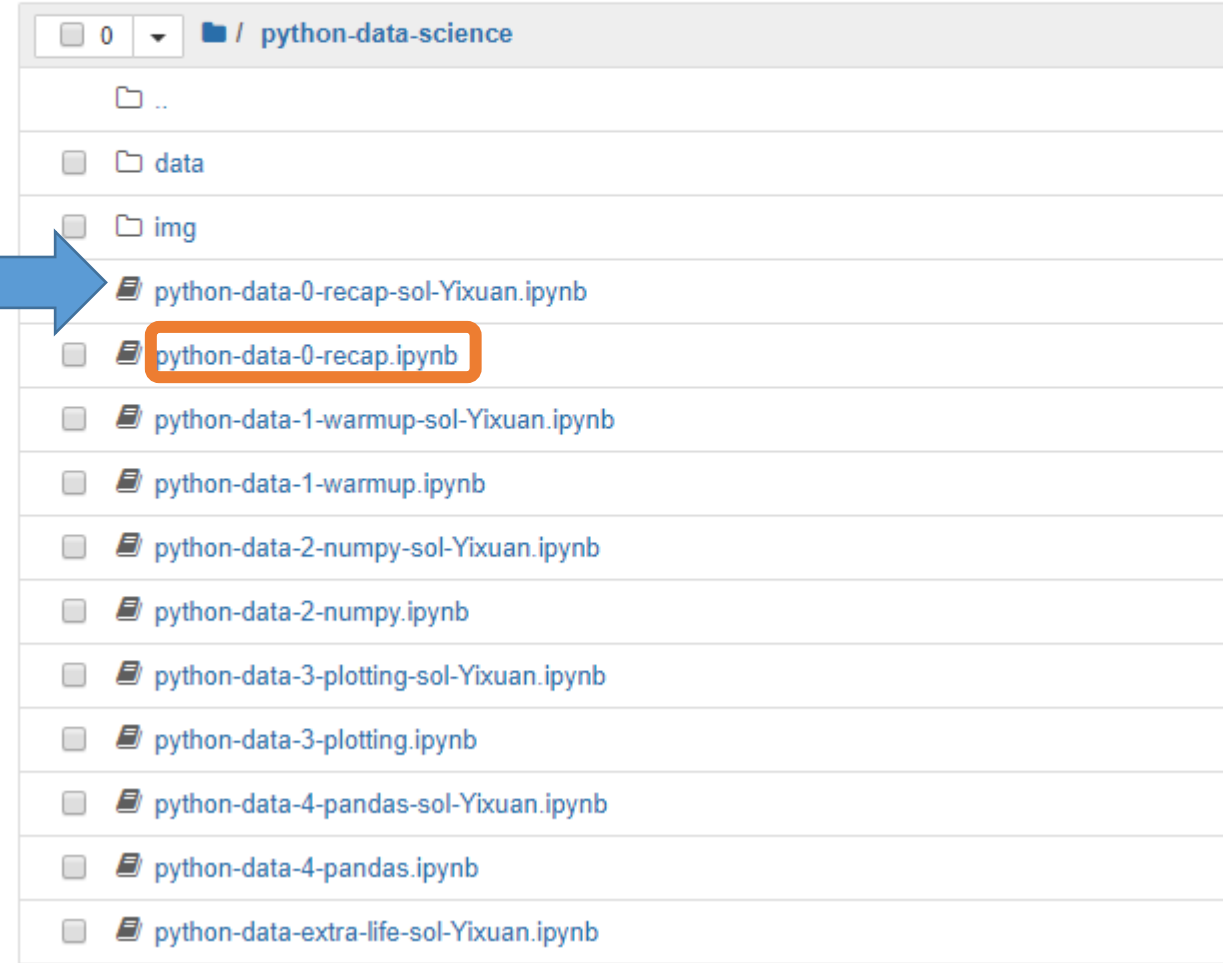
Files Running Assignments

Select items to perform actions on them.

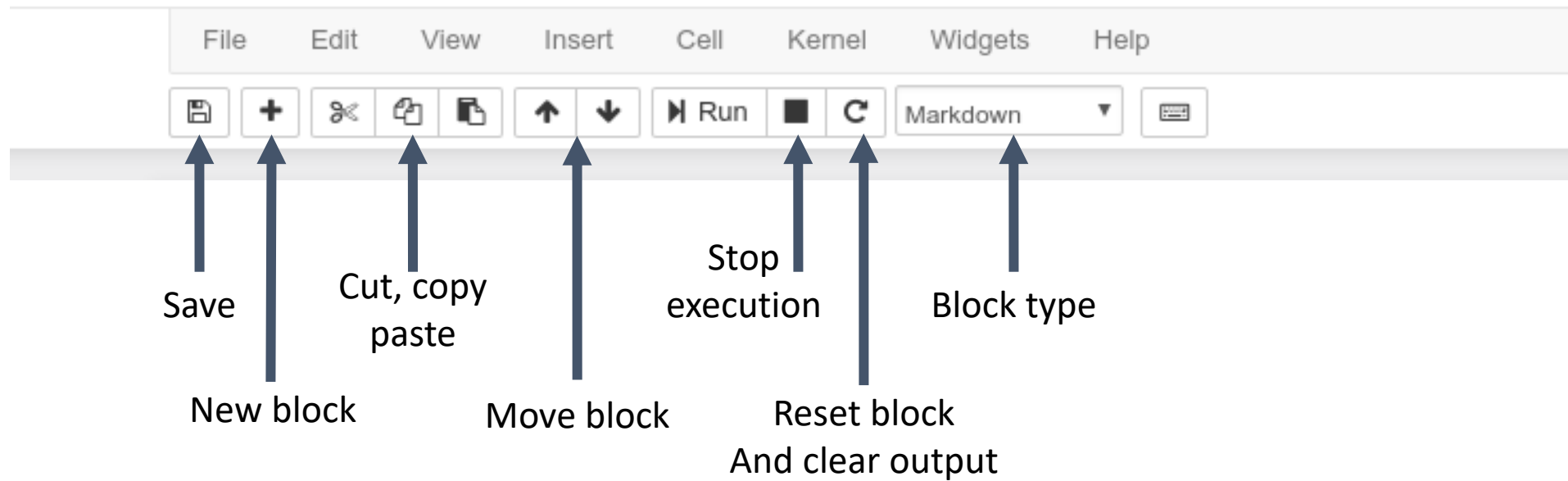


Files Running Assignments

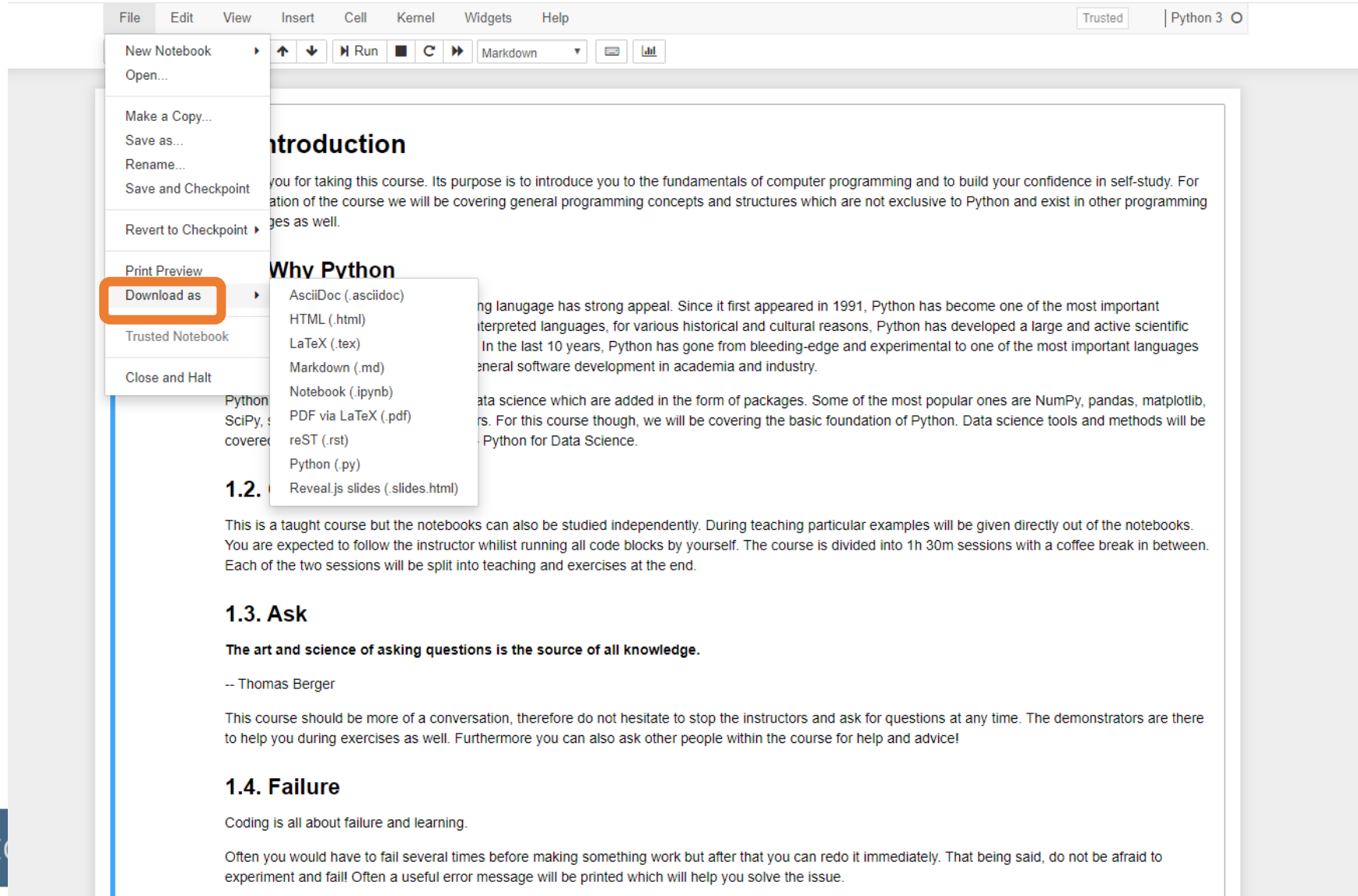
Select items to perform actions on them.



4. Toolbar



5. Download files



The screenshot shows a Jupyter Notebook interface. The 'File' menu is open, and the 'Download as' option is highlighted with an orange rectangle. A sub-menu is visible, listing various file formats for download: AsciiDoc (.asciidoc), HTML (.html), LaTeX (.tex), Markdown (.md), Notebook (.ipynb), PDF via LaTeX (.pdf), reST (.rst), Python (.py), and Reveal.js slides (.slides.html). The background content of the notebook includes a title 'Introduction', a paragraph about the course purpose, a section 'Why Python', and several numbered sections (1.2, 1.3, 1.4) discussing the course structure and the importance of asking questions and handling failure.

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

New Notebook
Open...

Make a Copy...
Save as...
Rename...
Save and Checkpoint
Revert to Checkpoint

Print Preview
Download as
Trusted Notebook
Close and Halt

AsciiDoc (.asciidoc)
HTML (.html)
LaTeX (.tex)
Markdown (.md)
Notebook (.ipynb)
PDF via LaTeX (.pdf)
reST (.rst)
Python (.py)
Reveal.js slides (.slides.html)

Introduction

you for taking this course. Its purpose is to introduce you to the fundamentals of computer programming and to build your confidence in self-study. For the duration of the course we will be covering general programming concepts and structures which are not exclusive to Python and exist in other programming languages as well.

Why Python

Python is a programming language that has strong appeal. Since it first appeared in 1991, Python has become one of the most important interpreted languages, for various historical and cultural reasons, Python has developed a large and active scientific community. In the last 10 years, Python has gone from bleeding-edge and experimental to one of the most important languages for general software development in academia and industry.

Data science tools and methods are added in the form of packages. Some of the most popular ones are NumPy, pandas, matplotlib, SciPy, etc. For this course though, we will be covering the basic foundation of Python. Data science tools and methods will be covered in a later course.

1.2. Course Structure

This is a taught course but the notebooks can also be studied independently. During teaching particular examples will be given directly out of the notebooks. You are expected to follow the instructor whilst running all code blocks by yourself. The course is divided into 1h 30m sessions with a coffee break in between. Each of the two sessions will be split into teaching and exercises at the end.

1.3. Ask

The art and science of asking questions is the source of all knowledge.

-- Thomas Berger

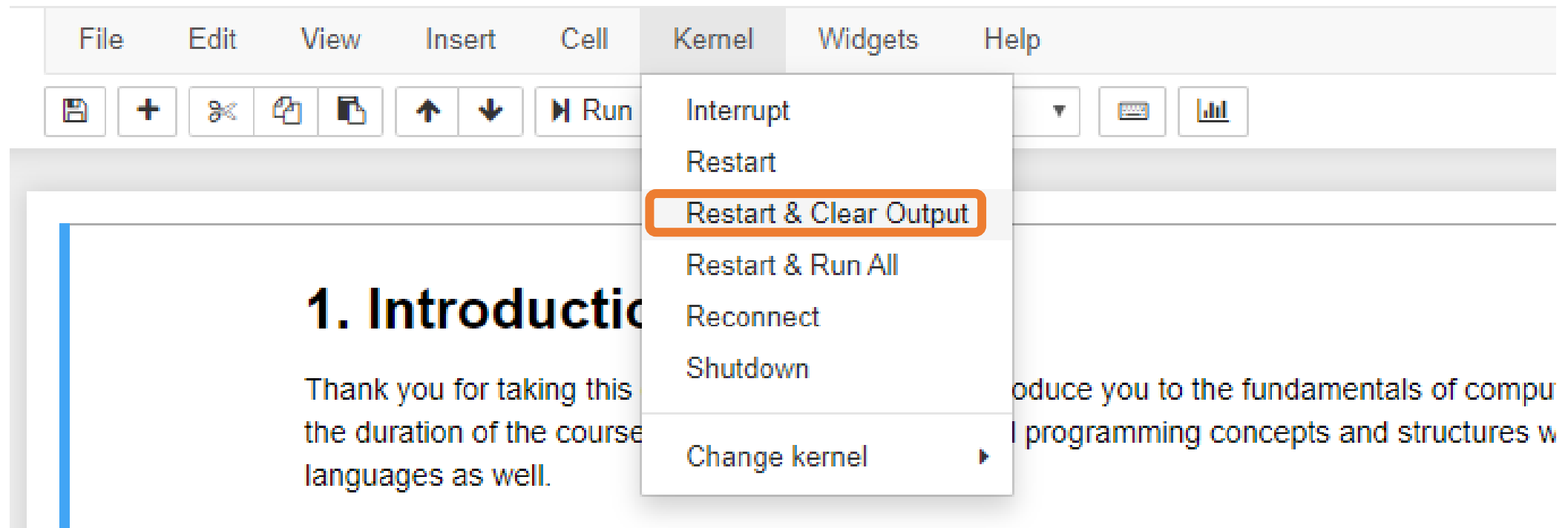
This course should be more of a conversation, therefore do not hesitate to stop the instructors and ask for questions at any time. The demonstrators are there to help you during exercises as well. Furthermore you can also ask other people within the course for help and advice!

1.4. Failure

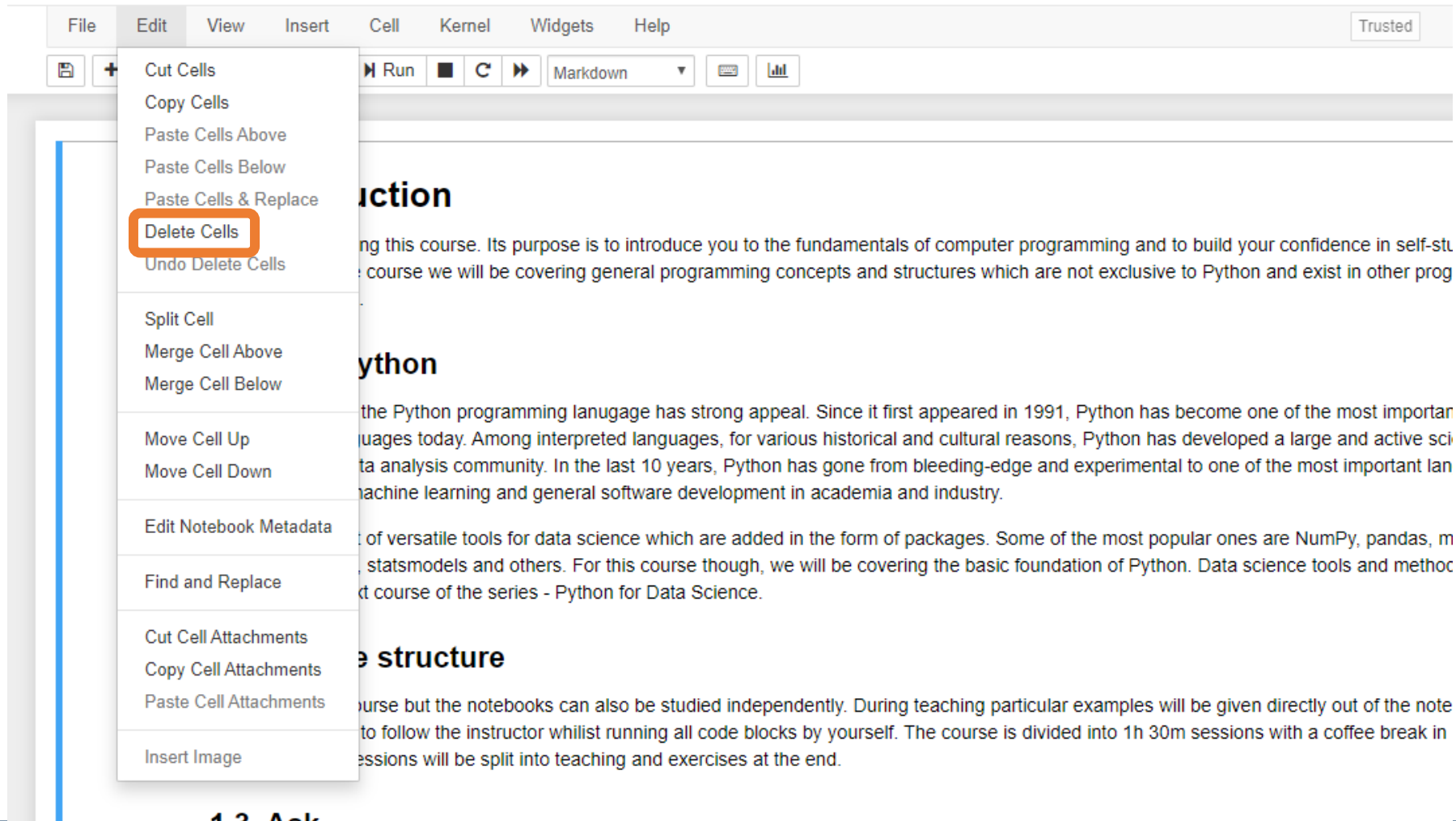
Coding is all about failure and learning.

Often you would have to fail several times before making something work but after that you can redo it immediately. That being said, do not be afraid to experiment and fail! Often a useful error message will be printed which will help you solve the issue.

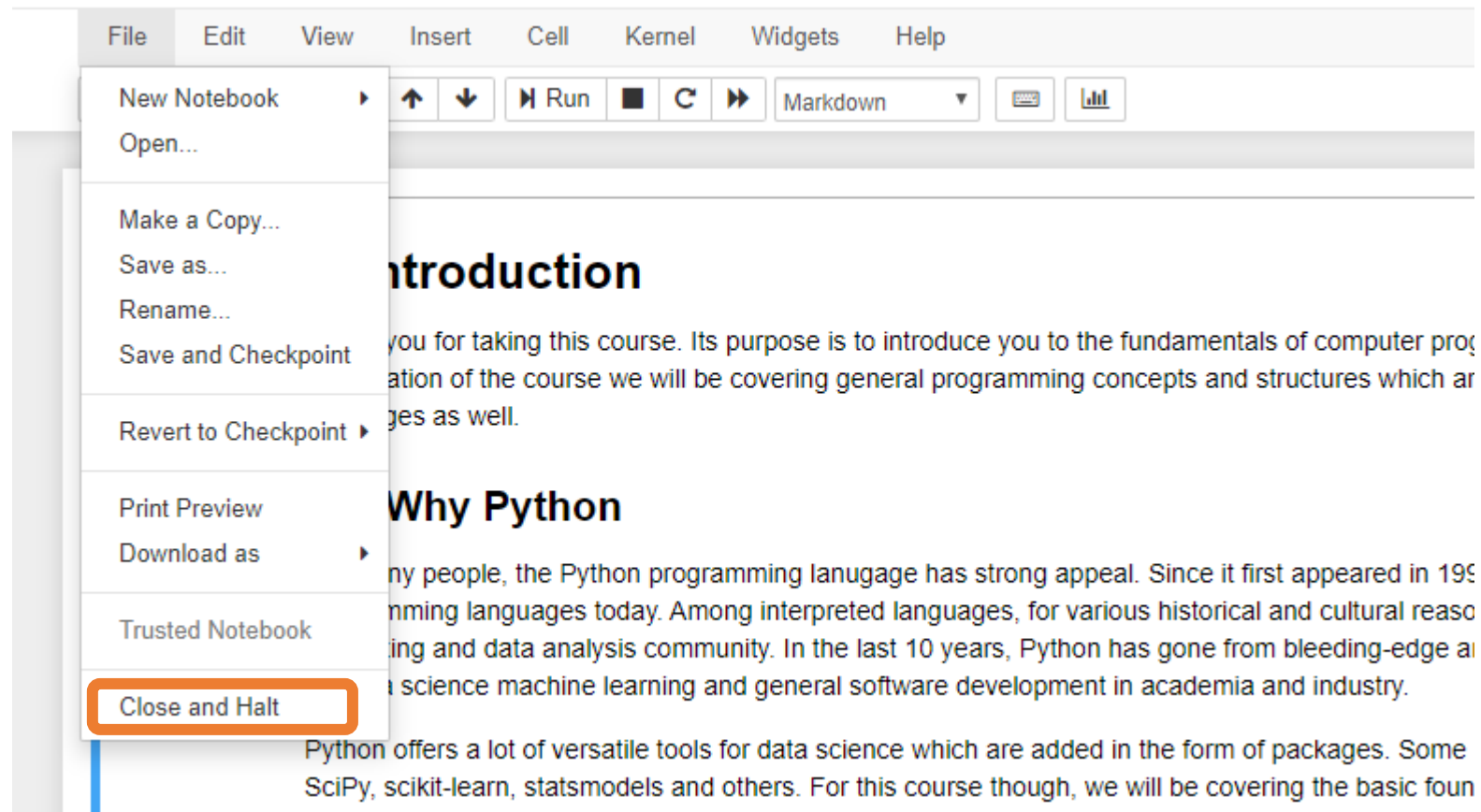
6. Kernel/Restart & Clear output



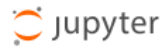
7. Edit/Delete Cell



8. File/ Close & Halt



9. Create a folder



Files Running

Select items to perform actions on them.

0

The notebook list is empty.

Upload

New

Notebook:

Python 3

R

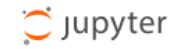
Other:

Text File

Folder

Terminal

10. Rename



Files Running

Rename

Move



Upload

New



1



Name

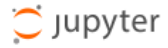
Last Modified



Untitled Folder

seconds ago

11. Upload files



Files [Running](#)

Select items to perform actions on them.

0

/ Introduction to Python

Name

Last Modified

..	seconds ago
----	-------------

The notebook list is empty.

Upload

New



Files [Running](#)

Select items to perform actions on them.

0

/ Introduction to Python

Name

Last Modified

The notebook list is empty.

python-intro-0.ipynb	Upload	Cancel
python-intro-1.ipynb	Upload	Cancel
python-intro-2.ipynb	Upload	Cancel
..		seconds ago
python-intro-exercises.ipynb	Upload	Cancel

Running blocks

- By pressing the Run button
- Shift + Enter – runs block
- Alt + Enter – creates a new block

Other operations

- File/Save and Checkpoint
- File/Revert to Checkpoint
- Tab completion
- Introspection

Let us start

To follow along, you need to open your own notebook. But please try to keep up with my presentation, as you still have time for exercises during the teaching.

Thank you for attending!

- You can continue to access the 'Python for Data Science' Learn course and notebooks after the course.
- There are solution notebooks for all the topics we have covered today, plus notebooks covering additional topics (text analysis, signal processing, regular expressions, network analysis, machine learning & Conway's game of life).
- Visit the Digital Skills Framework to find more Python learning resources: <https://www.digitalskills.ed.ac.uk/>