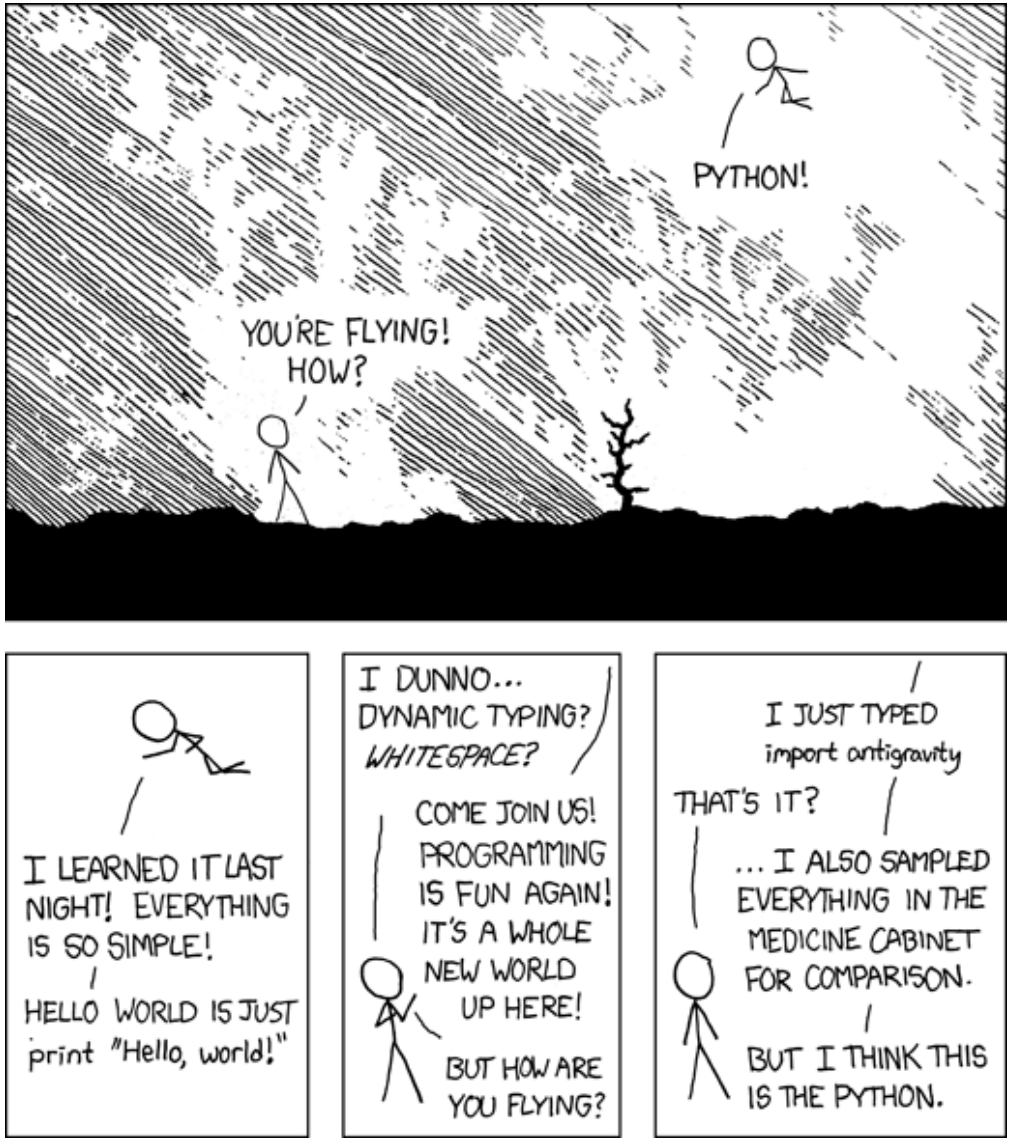
Day 1, Part 1: Introduction

Introduction to Python

Tom Paskhalis

RECSM Summer School 2023



Source: https://xkcd.com/353/

AS SEEN BY USERS OF …

About me

- Assistant Professor in Political Science and Data Science, Trinity College Dublin
    - Before: Postdoctoral Fellow, New York University
    - PhD in Social Research Methods, London School of Economics and Political Science
- My research:
    - Political communication, social media, interest groups
    - Text analysis, machine learning, record linkage, data visualization
- Contact
    - tom@paskhal.is
    - tom.paskhal.is
    - @tpaskhalis

About you

- Name?
- Affiliation?
- Research interests?
- Previous Experience with Python?
- Why are you interested in this course?

R/Stata/SPSS is great, why learn Python?

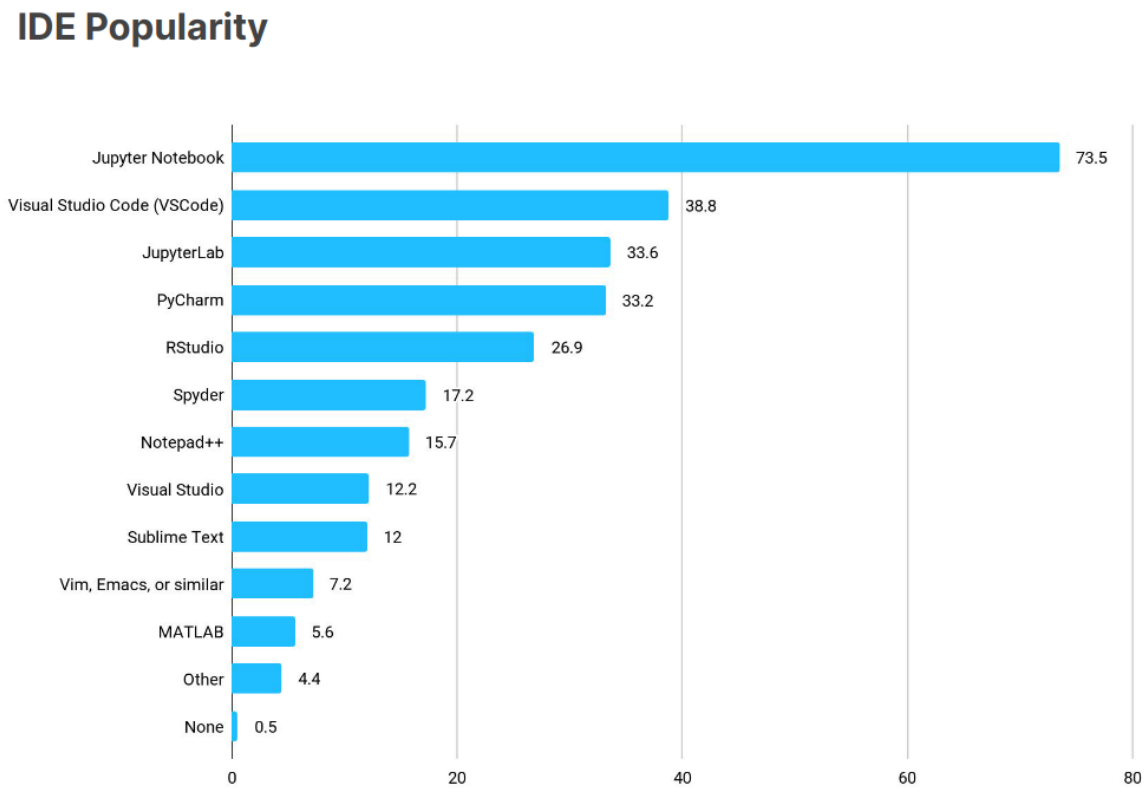
- Python is free and open source
- Python is a truly versatile programming language
- Python offers a great library ecosystem (>300K)
- Python is widely used in the industry
- Python is well-known outside academia/data science

Popularity of programming languages

Jun 2022	Jun 2021	Change	Programming Language	Ratings	Change
1	**2**	▲	**Python**	**12.20%**	**+0.35%**
2	1	▼	C	11.91%	-0.64%
3	3		Java	10.47%	-1.07%
4	4		C++	9.63%	+2.26%
5	5		C#	6.12%	+1.79%
6	6		Visual Basic	5.42%	+1.40%
7	7		JavaScript	2.09%	-0.24%
8	10	▲	SQL	1.94%	+0.06%
9	9		Assembly language	1.85%	-0.21%
10	16	⇈	Swift	1.55%	+0.44%

Source: https://www.tiobe.com/tiobe-index/

Popularity of data analysis software



Source: https://www.kaggle.com/kaggle-survey-2021

Python and Development Enviroments

- There is a number of integrated development environments (*IDE*s) available for Python (IDLE, Spyder, PyCharm)
- As well code editors with Python-specific extensions (Visual Studio Code, Atom, Sublime Text, Vim)
- Try different ones and choose what works best for you!

Python and Jupyter Notebook

- Jupyter Notebook is language-agnostic web-based interactive computational environment
- Is available with backends (*kernels*) for different programming languages (**Ju**lia, **Py**thon, **R** = **Jupy**te**r**)
- Can be used both locally and remotely
- Good for ad-hoc data analysis and visualization

Jupyter Notebook

- Notebooks allow writing, executing and viewing the output of Python code within the same environment
- All notebook files have `.ipynb` extension for **i**nteractive **py**thon **n**ote**b**ook
- The main unit of notebook is *cell*, a text input field (Python, Markdown, HTML)
- Output of a cell can include text, table or figure

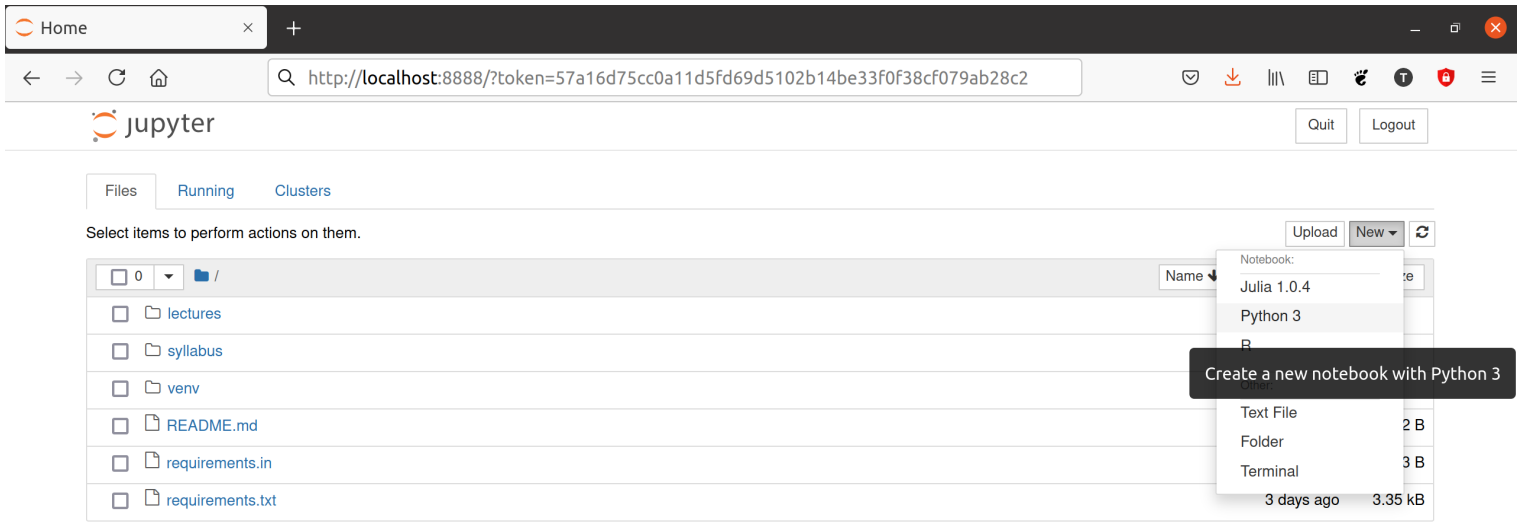
Jupyter Notebook online

- For this workshop I recommend using one of the online platforms for working with Jupyter Notebooks:
    - Google Colab, a cloud platform for hosting Jupyter Notebooks. You need to have a Google account, but it does not require any local installations.
    - Kaggle Code, a platform for sharing and exploring data-science-focussed Jupyter Notebooks. Although technically owned by Google, you can register just for Kaggle website.

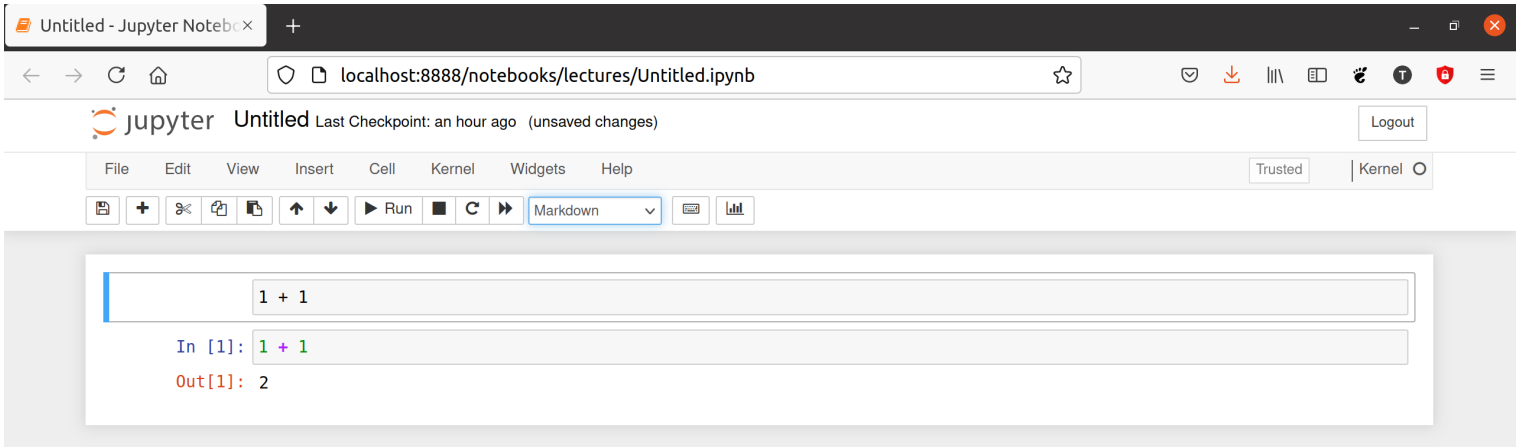
Jupyter Notebook installation

- If you would prefer to install Jupyter Notebook on your local machine, there are two main ways to do this:
    - pip
    - conda
- Unless you have prior experience with Python, I recommend installing Anaconda distribution, which contains all the packages required for this course.

Jupyter Notebook demonstration



Jupyter Notebook demonstration



Course outline

Date	Time (CEST)	Topic
26 June	09:00-10:45	Introduction to Python objects and data types
	10:45-11:15	Break
	11:15-13:00	Pandas, data input/output
27 June	09:00-10:45	Exploratory data analysis, data visualization
	10:45-11:15	Break
	11:15-13:00	Regression analysis, communicating results

Materials

- All materials for this workshop can be found:
    - In this GitHub repository: github.com/tpaskhalis/RECSM_Introduction_Python
    - Alternative shortlink: bit.ly/RECSM_Python
- For your convenience you might want to choose to clone this repository to your local macihine.
- All slides and exercises were created using Python and Jupyter Notebooks.

Additional materials

- There are many great online resources and published books on programming in Python.
- Some of them also provide a good coverage of using Python for data analysis.
- Here are some pointers to start from.

Books

- Guttag, John. 2021 *Introduction to Computation and Programming Using Python: With Application to Computational Modeling and Understanding Data*. 3rd ed. Cambridge, MA: The MIT Press

- McKinney, Wes. 2022. *Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter*. 3rd ed. Sebastopol, CA: O'Reilly Media

- Sweigart, Al. 2019. *Automate the Boring Stuff with Python*. 2nd ed. San Francisco, CA: No Starch Press

Online

- The Hitchhiker's Guide to Python
- Python For You and Me
- Python Wikibook
- Python 3 Documentation (intermediate and advanced)

Next

- Basic Python types
- Operations
- Object manipulations