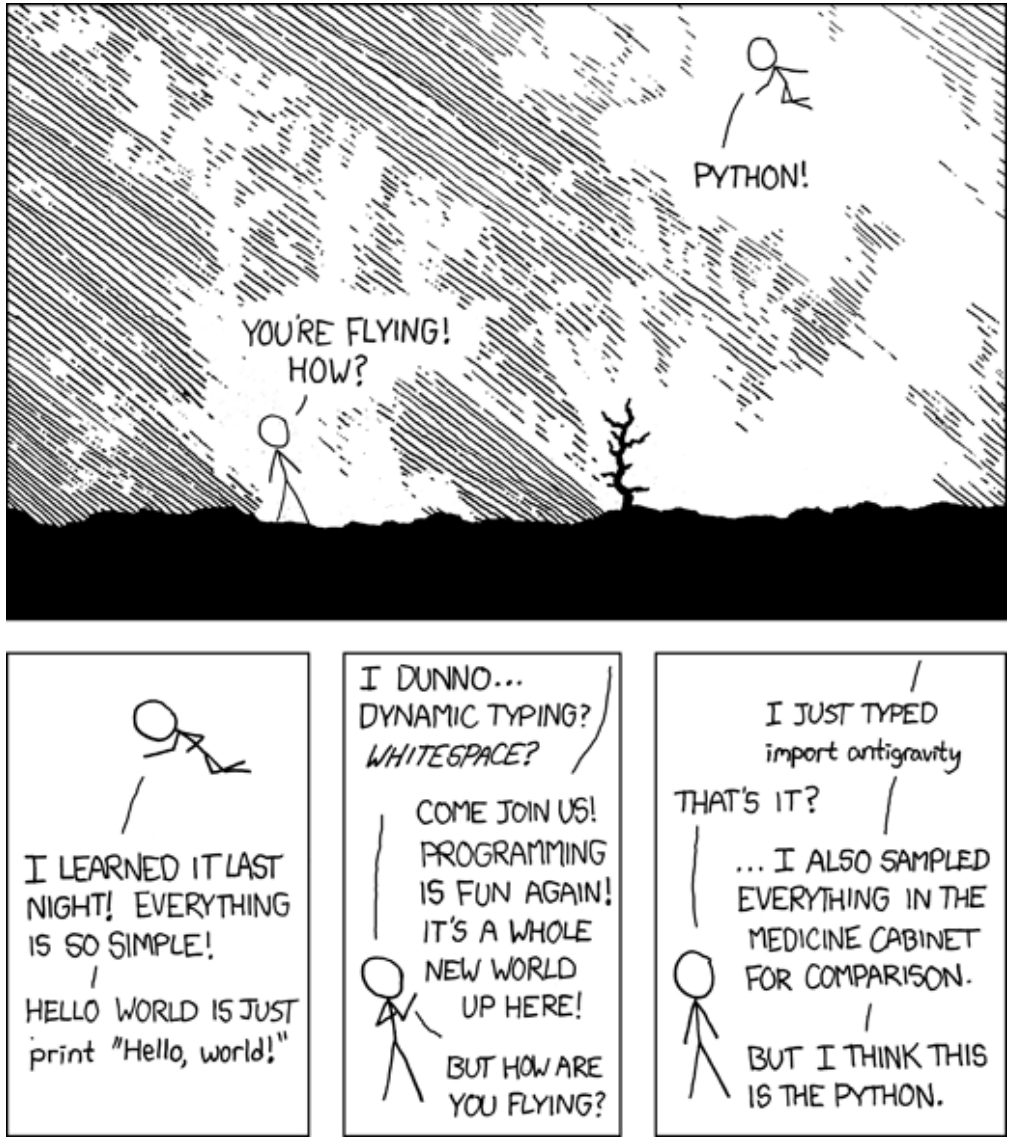
Introduction to Python

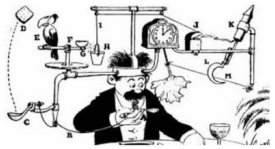
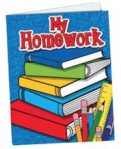
Tom Paskhalis

RECSM Summer School 2021, Introduction, Part 1, Day 1



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

AS SEEN BY USERS OF …

	STATA	R	SAS	Python	SPSS
STATA					
R					
SAS					
Python					
SPSS					

About me

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

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?

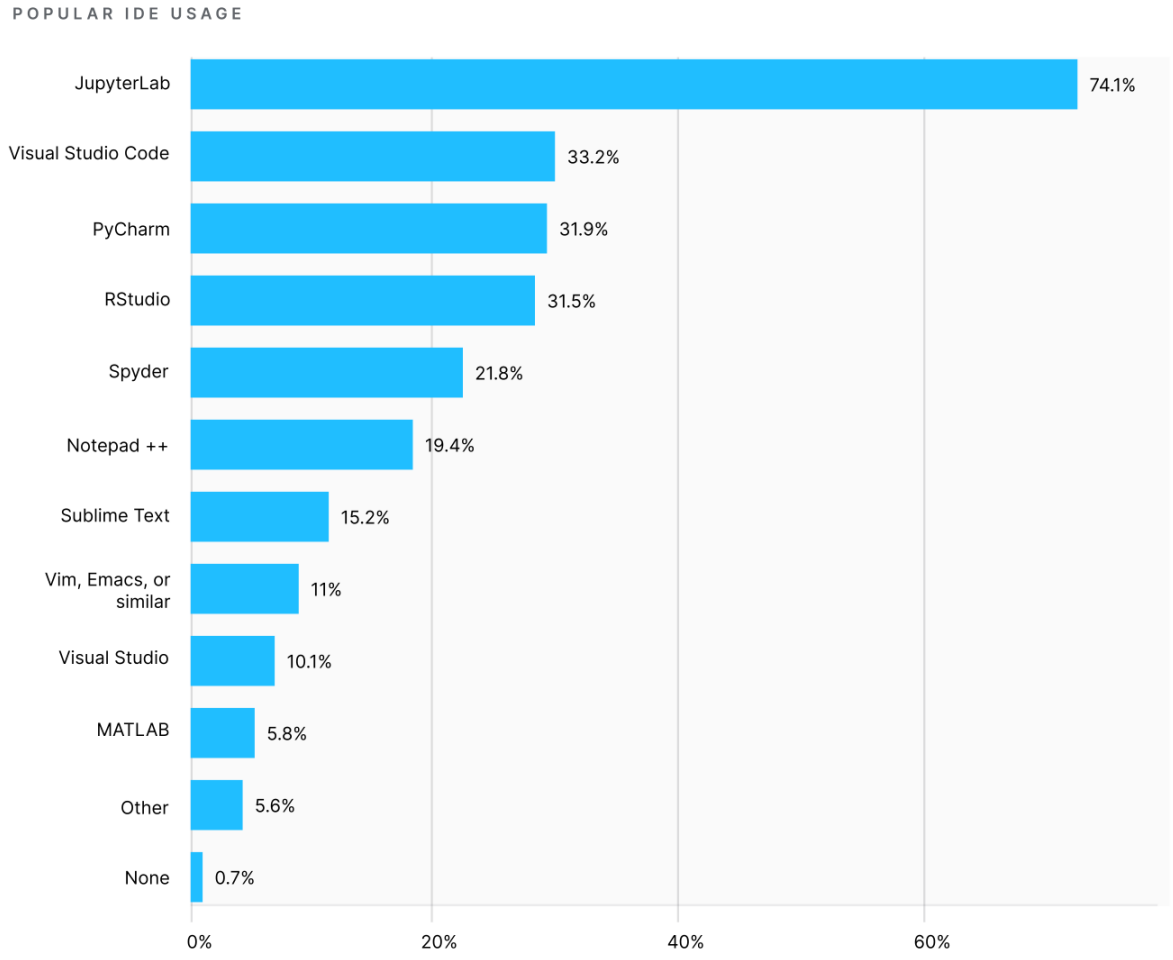
- Python is free and open source
- Python is a truly versatile programming language
- Python offers a great library ecosystem (>300K)
- Python is widely used in the industry
- Python is well-known outside academia/data science

Popularity of programming languages

Jun 2021	Jun 2020	Change	Programming Language	Ratings	Change
1	1		C	12.54%	-4.65%
2	3	↑	Python	11.84%	+3.48%
3	2	↓	Java	11.54%	-4.56%
4	4		C++	7.36%	+1.41%
5	5		C#	4.33%	-0.40%
6	6		Visual Basic	4.01%	-0.68%
7	7		JavaScript	2.33%	+0.06%
8	8		PHP	2.21%	-0.05%
9	14	⇈	Assembly language	2.05%	+1.09%
10	10		SQL	1.88%	+0.15%

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

Popularity of data analysis software



Source: https://www.kaggle.com/kaggle-survey-2020

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 (Vim, Atom, Sublime Text, Visual Studio Code)
- 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

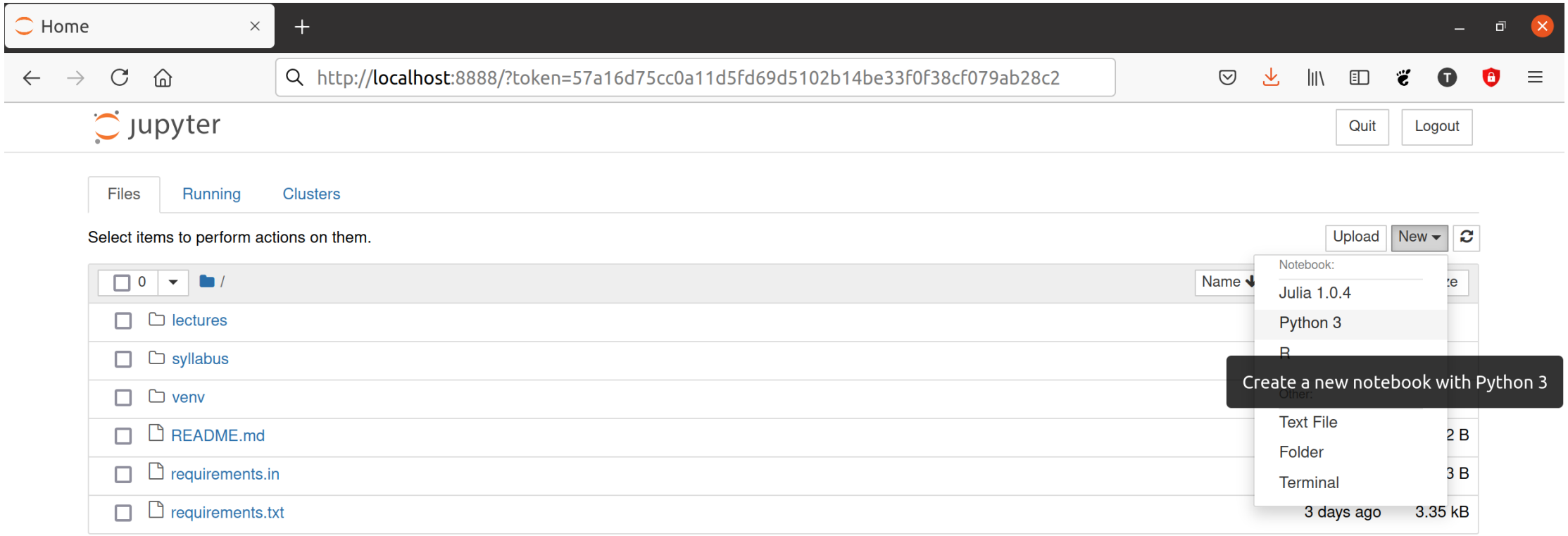
Jupyter Notebook

- 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

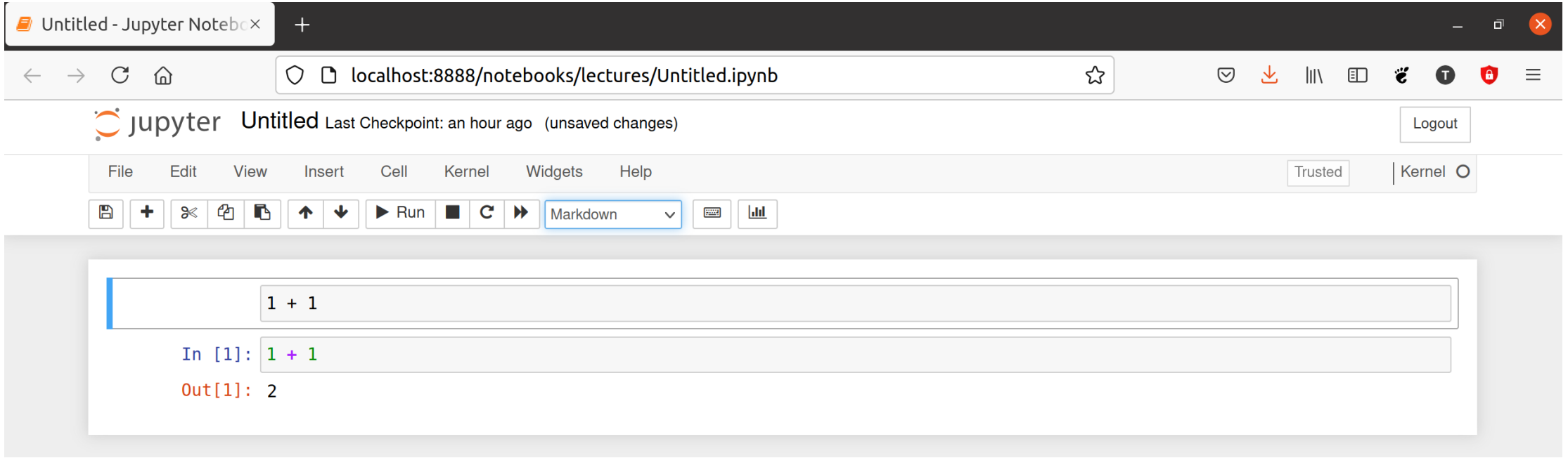
Jupyter Notebook Installation

- There are two main ways to install Jupyter Notebook: pip and conda. Unless you have prior experience with Python, I recommend installing Anaconda distribution, which contains all the packages required for this course.
- Alternatively, you may choose to use Google Colab, a cloud platform for hosting Jupyter Notebooks. Its interface is slightly different and you need to have a Google account, but it does not require any local installations.

Jupyter Notebook Demonstration



Jupyter Notebook Demonstration



Course Outline

Date	Time (CET)	Topic
28 June	14:00-15:50	Introduction to Python objects and data types
	15:50-16:10	Break
	16:10-18:00	Pandas, data input/output
29 June	14:00-15:50	Exploratory data analysis, data visualization
	15:50-16:10	Break
	16:10-18:00	Regression analysis, communicating results

Materials

- All materials for this workshop can be found in this GitHub repository: github.com/tpaskhalis/RECSM_Introduction_Python
- For your convenience you might want to choose to clone this repoistory to your local macihine.
- It is worth noting that 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:

Next

- Basic Python types
- Operations
- Object manipulations