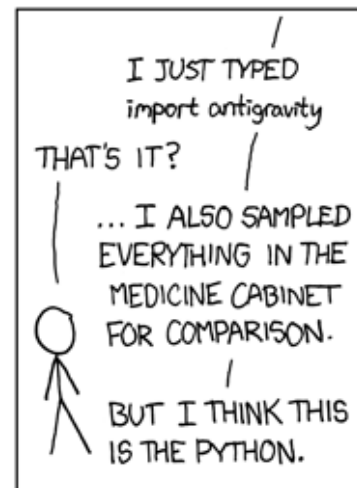
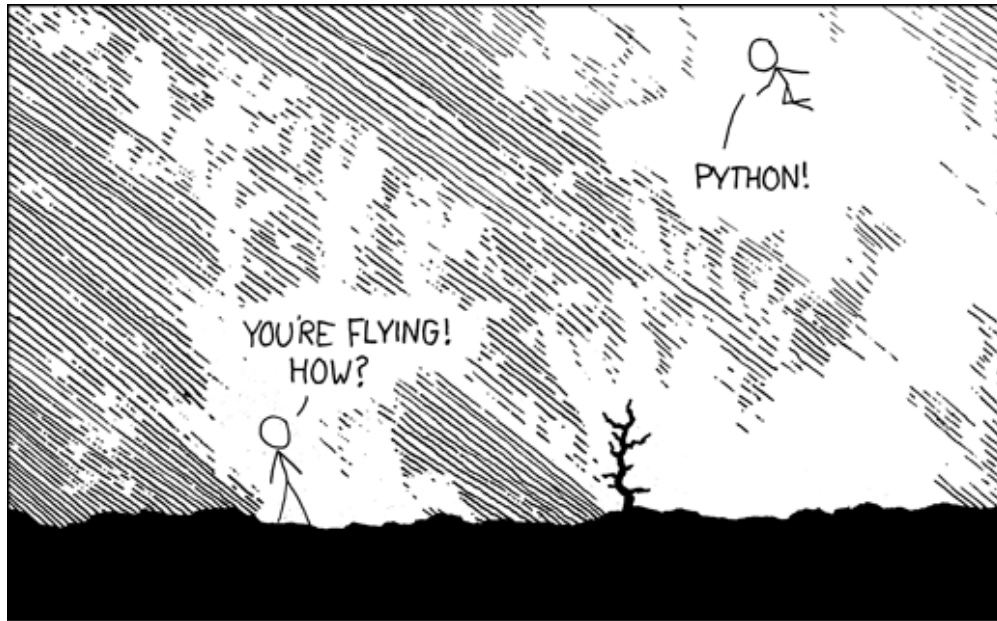


# Introduction to Python

Tom Paskhalis

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





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

# AS SEEN BY USERS OF ...

STATA

R

sas

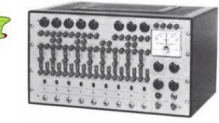
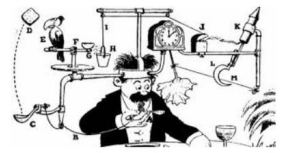
python

SPSS

STATA



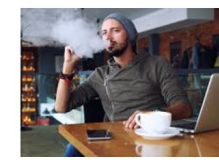
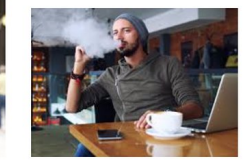
R



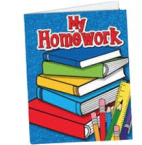
sas



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

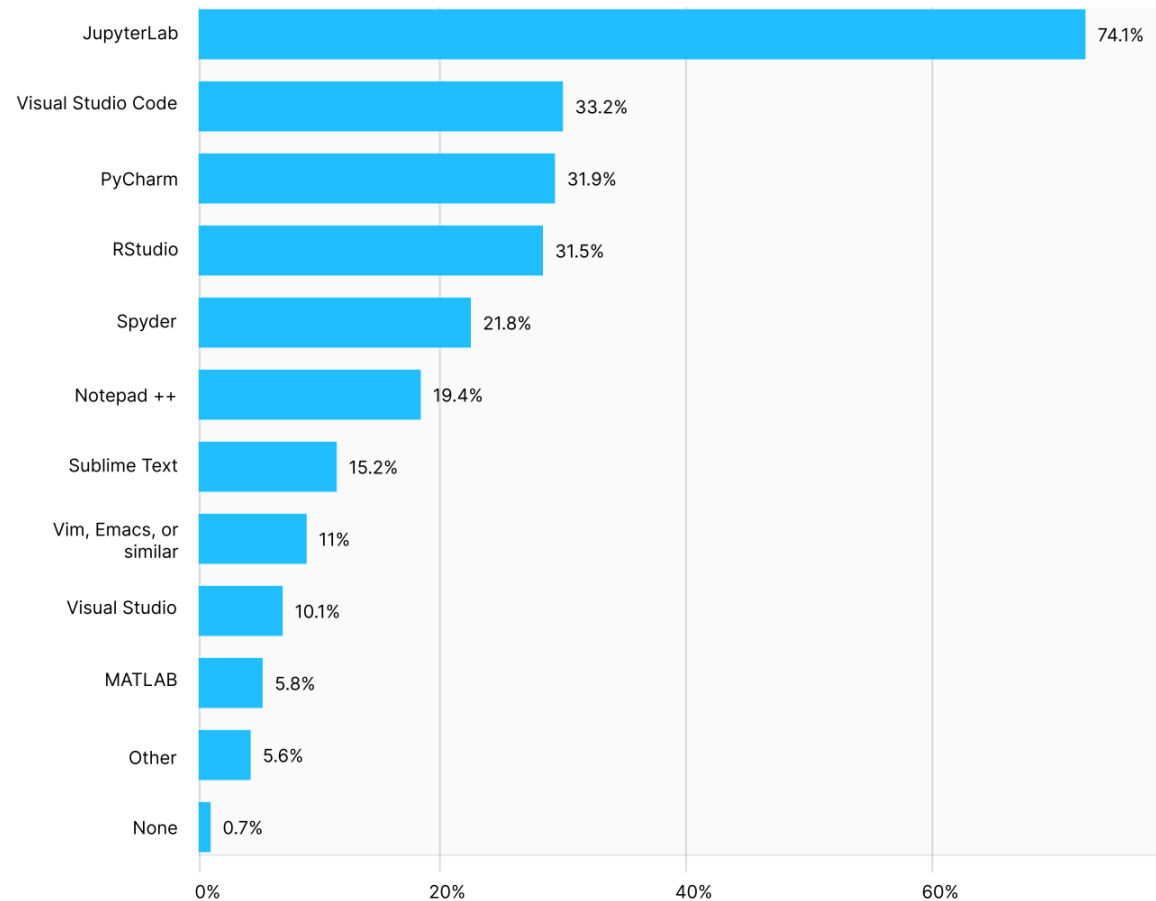
Jun 2021	Jun 2020	Change	Programming Language	Ratings	Change
1	1		 C	12.54%	-4.65%
2	3	▲	 Python	11.84%	+3.48%
3	2	▼	 Java	11.54%	-4.56%
4	4		 C++	7.36%	+1.41%
5	5		 C#	4.33%	-0.40%
6	6		 Visual Basic	4.01%	-0.68%
7	7		 JavaScript	2.33%	+0.06%
8	8		 PHP	2.21%	-0.05%
9	14	▲	 Assembly language	2.05%	+1.09%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

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



# Popularity of data analysis software

POPULAR IDE USAGE



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

# Python and Development Enviroments

- There is a number of integrated development environments (*IDEs*) 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 (**Julia**, **Python**, **R** = **Jupyter**)
- 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 interactive **python notebook**
- 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

The screenshot displays the Jupyter Notebook web interface in a browser window. The address bar shows the URL `http://localhost:8888/?token=57a16d75cc0a11d5fd69d5102b14be33f0f38cf079ab28c2`. The Jupyter logo and 'Quit'/'Logout' buttons are at the top. Below the navigation tabs (Files, Running, Clusters), the 'Files' tab is active, showing a file browser. The browser displays a directory structure with folders 'lectures', 'syllabus', and 'venv', and files 'README.md', 'requirements.in', and 'requirements.txt'. A 'New' dropdown menu is open, showing options: 'Notebook:', 'Julia 1.0.4', 'Python 3', 'R', 'Text File', 'Folder', and 'Terminal'. A dark overlay text reads 'Create a new notebook with Python 3'.

Home × +

← → ↻ 🏠 🔍 `http://localhost:8888/?token=57a16d75cc0a11d5fd69d5102b14be33f0f38cf079ab28c2` 📁 📄 🖨️ 🌐 🛡️ ☰

jupyter Quit Logout

Files Running Clusters

Select items to perform actions on them.

Upload New ↻

0 ▾ /

	Name	Size
<input type="checkbox"/>	lectures	
<input type="checkbox"/>	syllabus	
<input type="checkbox"/>	venv	
<input type="checkbox"/>	README.md	2 B
<input type="checkbox"/>	requirements.in	3 B
<input type="checkbox"/>	requirements.txt	3.35 kB

Notebook:  
Julia 1.0.4  
Python 3  
R  
Text File  
Folder  
Terminal

Create a new notebook with Python 3

# Jupyter Notebook Demonstration

The screenshot displays a web browser window with a Jupyter Notebook interface. The browser's address bar shows the URL `localhost:8888/notebooks/lectures/Untitled.ipynb`. The notebook's title bar indicates the file is "Untitled" and shows the last checkpoint was "an hour ago" with "unsaved changes". A "Logout" button is visible in the top right corner of the notebook interface.

The main menu bar includes "File", "Edit", "View", "Insert", "Cell", "Kernel", "Widgets", and "Help". Below the menu is a toolbar with icons for saving, creating new cells, deleting cells, copying, pasting, undo, redo, running the cell, and a dropdown menu currently set to "Markdown".

The notebook content area shows a single code cell. The input code is `1 + 1`. Below the code cell, the execution output is displayed as `In [1]: 1 + 1` followed by `Out[1]: 2`.



# 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](https://github.com/tpaskhalis/RECSM_Introduction_Python)
- For your convenience you might want to choose to clone this repository to your local machine.
- 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