MAX-PLANCK-INSTITUT FÜR GESELLSCHAFTSFORSCHUNG
MAX PLANCK INSTITUTE FOR THE STUDY OF SOCIETIES

A HOPEFULLY GENTLE INTRODUCTION TO PYTHON

Max Planck Institute for the Study of Societies

20 February 2024

Joshua Cova cj@mpifg.de

OBJECTIVES FOR TODAY

- Learn how to operate Python on VS Code
- Become acquainted with the Python programming environment (scripts, terminal and packages)
- Learn the fundamentals of Python programming
  - Operations
  - Data structures: Lists, dictionaries
  - Read in tabular data (e.g. .csv)
  - Basic functionalities with the Pandas library

SOME REASONS TO LEARN PYTHON AS SOCIAL SCIENTISTS

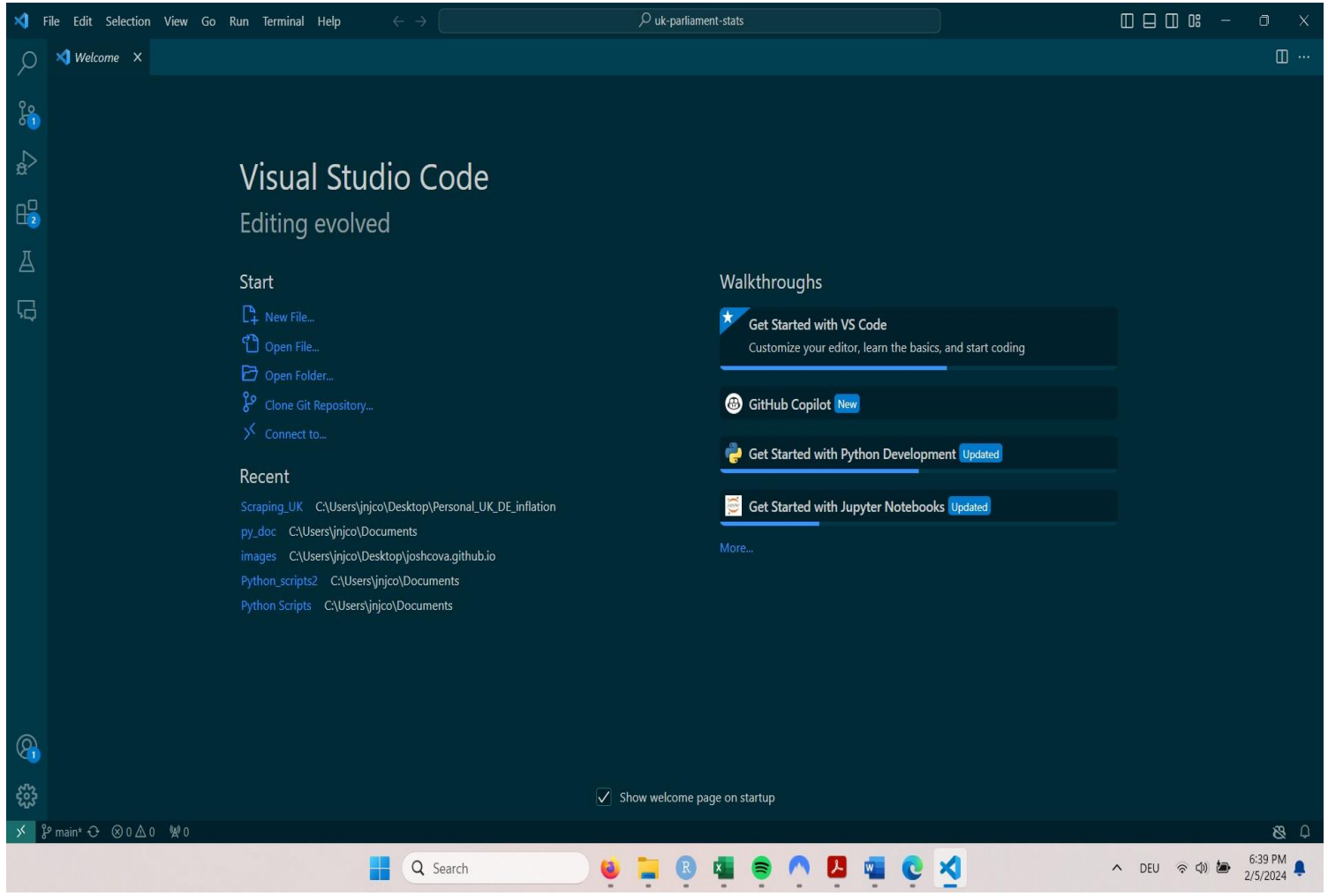
- Some use cases…
  - Data collection: web scraping
  - Data cleaning and preprocessing
  - Statistical analysis (e.g. regressions)
  - Text analysis, natural language processing (NLP)
  - Data visualization
- …And some motivation
  - It has become a lot easier to learn Python and coding since the arrival of ChatGPT and other LLMs
  - Tons of resources and a good community
  - Great entry point to other programming languages

WHY USE VISUAL STUDIO CODE AS A CODE EDITOR?

- Free
- Open source
- Relatively light
- Supports many languages (e.g. R, HTML)
- Numerous extensions
- Maintained by Microsoft

LET'S GET STARTED: HOPEFULLY YOU SEE SOMETHING LIKE THIS…

- But first let us check whether Python is properly installed:
  - Window users: Open the command prompt and then type: python --version
  - Mac users: Open the terminal and then type: python --version
- In case you have run into difficulties installing/setting up VS Code and/or Python, you might want to use Google Colab as a workaround



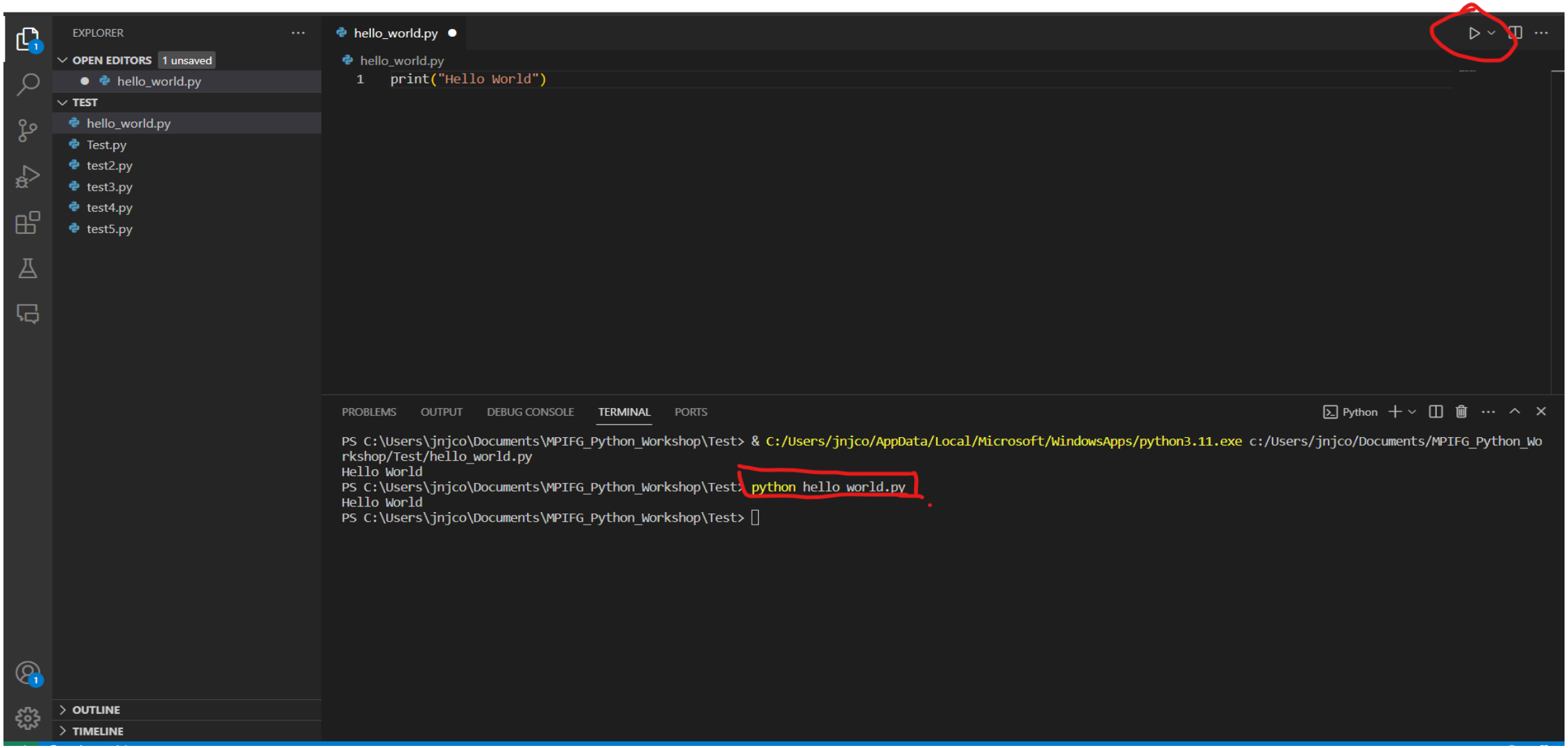
VS CODE: SETTING THINGS UP

- Select the Extension Icon:
  - Install the Python extension, which will make it easier to code in Python and debug code
  - Install Jupyter, a special Python file type, which makes it easier to visualize output (Notebook/markdown documents)
- If you do not like the dark color of the editor, feel free to change it:
  - Settings → Themes → Color Theme
- Create a folder for this session and then open it in VS Code
  - File → Open Folder
- Create a new file
  - File → Create new file

File extensions

Python scripts → **.py**
Jupyter notebook → **.ipynb**

OUR FIRST SCRIPT



OPERATIONS

- You can use Python as a calculator in the terminal: **+**; **-**; ***; **/**….
- Note the role of **=** as an assignment operator and note the difference between **=** and **==**
- But a better use then using the terminal as a calculator would be to assign numeric values to variables and then run a program
- Remember to comment you code with **#**, it is better for future you and for your colleagues !

DATA TYPES

- The way in which Python stores your variables is important → We know when a number is a number, Python does not necessarily know it if you do not tell it.
- You can always check the data type with the *type* function → *type*(*variable*)

Type	Python operationalization	Example
Text	*str*	abc = 'Hello World'
Numeric	*int, float*	num1 = 1
Logical	*bool*	print(1==1) TRUE

IF STATEMENTS AND VARIABLE ASSIGNMENT

- If statement are a common application in programming, let's try it out!
- Reserved words: if, elif, else
- Indent code in Python → either use the Tab key or press the Spacebar 4 times.

LET US SWITCH TO JUYPTER NOTEBOOKS AND TAKE A CLOSER LOOK AT DATA STRUCTURES

- Today, we will focus on lists, dictionaries and data frames
  - Lists: Mutable and can contain different data types.
    - Create a list with square brackets → e.g. *names = ['Adam', 'Bob', 'Claire']*
    - Index a value with [ ]
    - Methods: *.append*
  - Dictionaries: {key : value} format, think of an old-fashioned phone book.
    - *sentiments = {"positive": 1, "neutral": 0, "negative": -1}*

PYTHON LIBRARIES

Collection of domain-specific and reusable Python functions and methods (think of a toolbox)

Some useful libraries for social scientists:

- Pandas, Numpy: data analysis, wrangling & manipulating
- Matplotlib/Seaborn: data visualization
- NLTK, scikit-learn, SpaCy: QTA/NLP
- Beautiful Soup, Selenium: Web scraping

To install libraries/packages open a terminal window and type → *pip install 'package name'*

Alternatively if you have the Anaconda environment → *conda install 'package name'*

ADDITIONAL RESOURCES

There a great number of free resources to learn Python:

- Python tutorial → https://wiki.python.org/moin/BeginnersGuide/Programmers
- YouTube videos, especially the data science ones
- Python QTA/NLP applications for computational social sciences
  - Wouter van Atteveldt, Damian Trilling & Carlos Arcila (2022) *Computational Analysis of Communication*

MAX-PLANCK-INSTITUT FÜR GESELLSCHAFTSFORSCHUNG
MAX PLANCK INSTITUTE FOR THE STUDY OF SOCIETIES

Max Planck Institute for the Study of Societies

Paulstr. 3 | 50676 Köln | www.mpifg.de | info@mpifg.de