Introduction to Python

Michele Tizzoni

Department or Sociology and Social Research

PhD Programme in Sociology and Social Research

Tedious bio

Hello everyone! I am **Michele Tizzoni**, Assistant Professor in computational social science at the University of Trento.

- I am affiliated with the C2S2 (`https://c2s2.unitn.it/`)
- At UNITN since September 2022
- Before that, I was Senior Research Scientist at the ISI Foundation in Turin (`www.isi.it`)

Research Interests

My research deals with the computational study of human behaviour and in particular how human behaviour explains the **dynamics of infectious diseases**.

Some keywords that are relevant to my work:

- Computational and digital epidemiology
- Human mobility from digital traces (e.g. mobile phone data)
- Complex networks
- Wearable sensors

Overall (Tentative) Outline

- **Lesson 1:** Intro to Python and Basics
- **Lesson 2:** Data Cleaning and Processing; Exploratory Data Analysis; Visualization
- **Lesson 3:** Elements of Data Science

Course Goals

The course seeks to:

- Help you familiarize with the most relevant concepts that are necessary for using Python from a quant/data scientist perspective
- Offer you an incremental guide to learning
- Foster your (critical) curiosity about the language

By the end of the course you should be able to:

- Distinguish Data Types
- Upload data
- Perform data cleaning, processing
- Visualize results/patterns
- Deploy a model (and, most importantly, know where to look to deploy one)

Python: An Historical Sketch

Python is a widely-used, interpreted, object-oriented programming language designed to be general-purpose.

First release: 1991 (v0.9.0). Currently: v3.10.0

Created by Guido Van Rossum, which identified the language's goals as:

- *Easy and intuitive, but powerful*
- *Open Source*
- *Understandable*
- *Suitable for all kinds of everyday tasks*

Python Today/1

Today Python is one of the most utilized programming languages (according to different rankings, stably among the 5 most common ones).

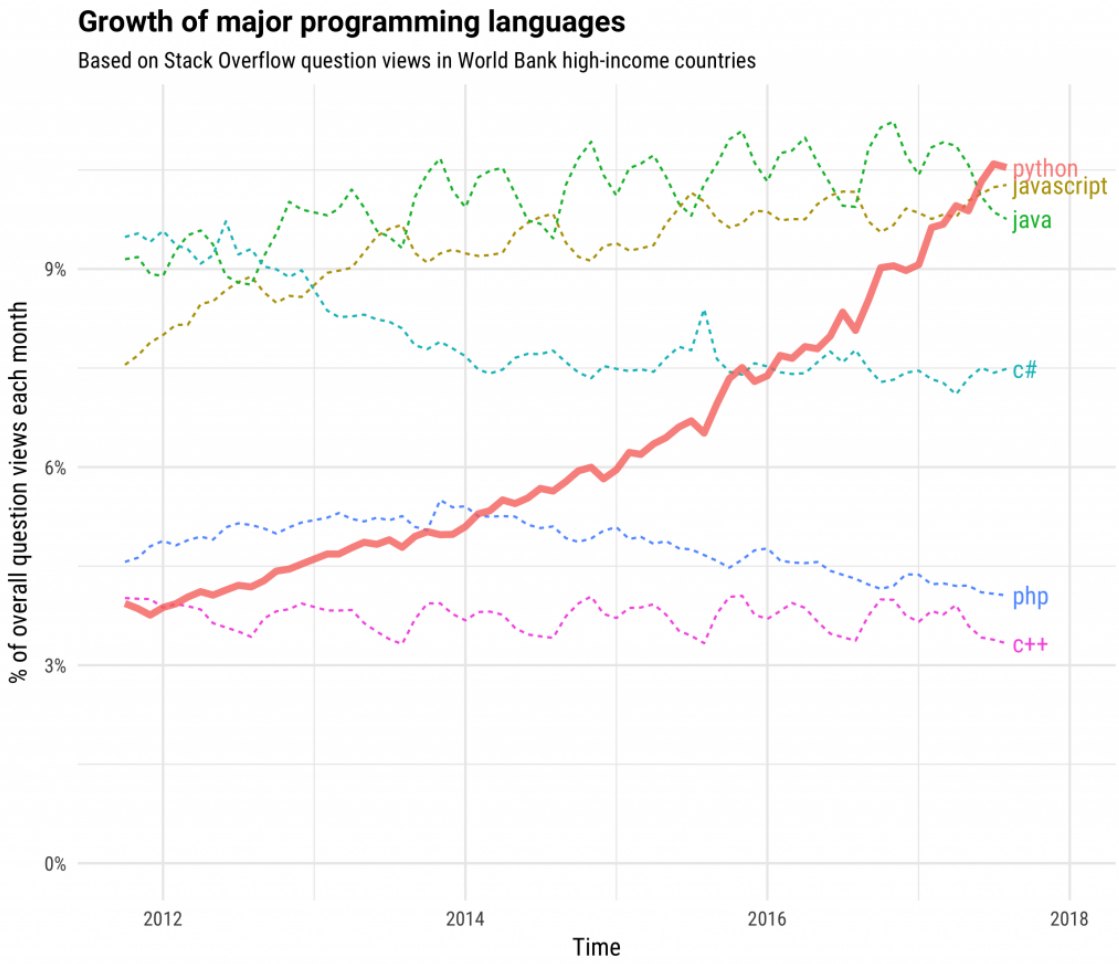
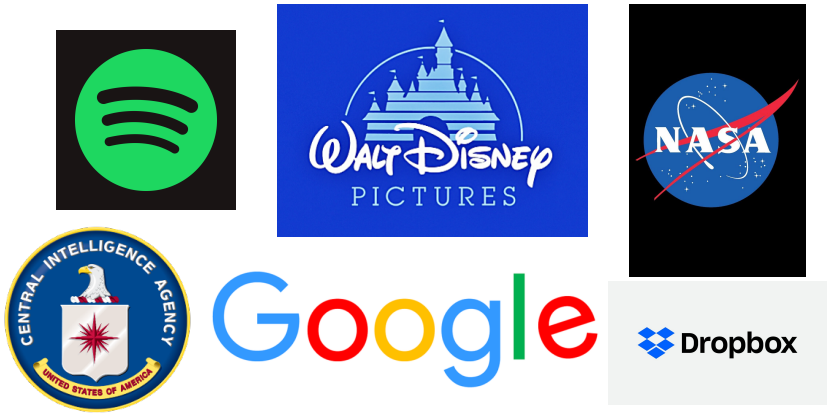


Figure: Python Growth. Source: Stack Overflow

Python Today/2

Who uses Python?



...and many more companies, organizations and institutions

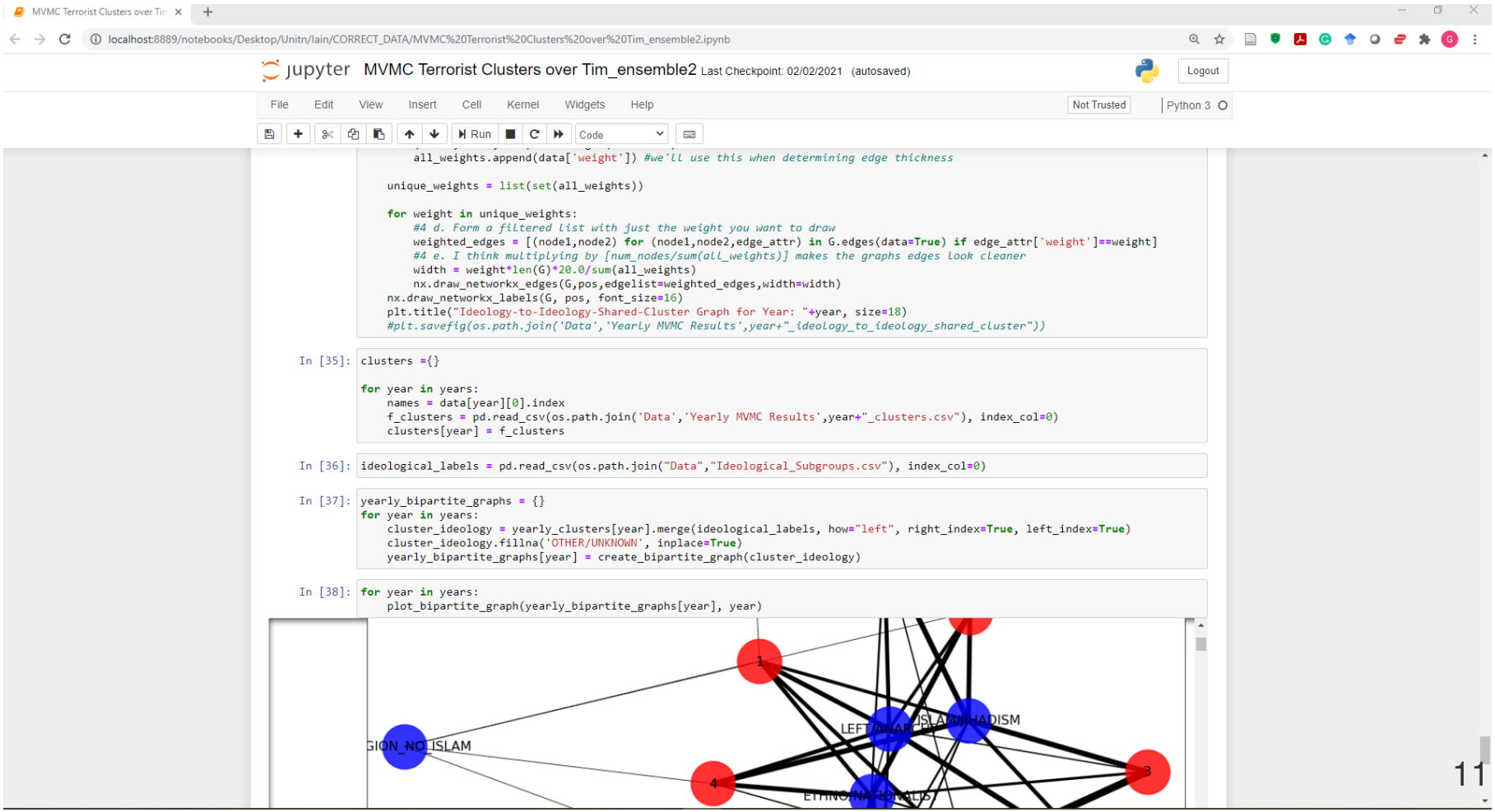
Why Choosing Python?

Scholars (especially those in the social sciences) often face the Hamletic doubt: *should I learn R or should I learn Python?* TL;DR answer: **is it really necessary to choose one?**

Python	**R**
General-purpose language / broad approach to data science	Specifically designed for statistical analysis
Often new advances in ML are first developed in Python	Often new advances in stats are first developed in R
Supports all kinds of data formats	Designed mostly for Excel, csv, text files
Heavily relies on libraries/packages	Data analysis functionalities are mostly built-in
Big community of developers with many different expertise	Community more focalized on statistical methods/theory

Presenting Jupyter Notebook/1

JN is a web-based application for interactive computing that allows code development, documentation and execution.



Presenting Jupyter Notebook/2

Let's open it.

- Open your Anaconda Prompt
- type `jupyter notebook`
- Click *new* on the top right of the screen

Pros and Cons of Jupyter Notebooks

Pros:

- Write, run, analyze everything in the same place
- Great illustrative power (good for teaching/presentations)
- Documentation + code together

Cons:

- Hard to test on long/asynchronous tasks
- Cells can run out of order
- No inspection of variables

Sean J. Taylor @seanjtaylor · 11h

Every notebook I work on long enough becomes a horror show.

12 10 187