Python for Data Science

Giulio Rossetti, EUI, Firenze 1-2 February 2018

How many people here know a programming language?

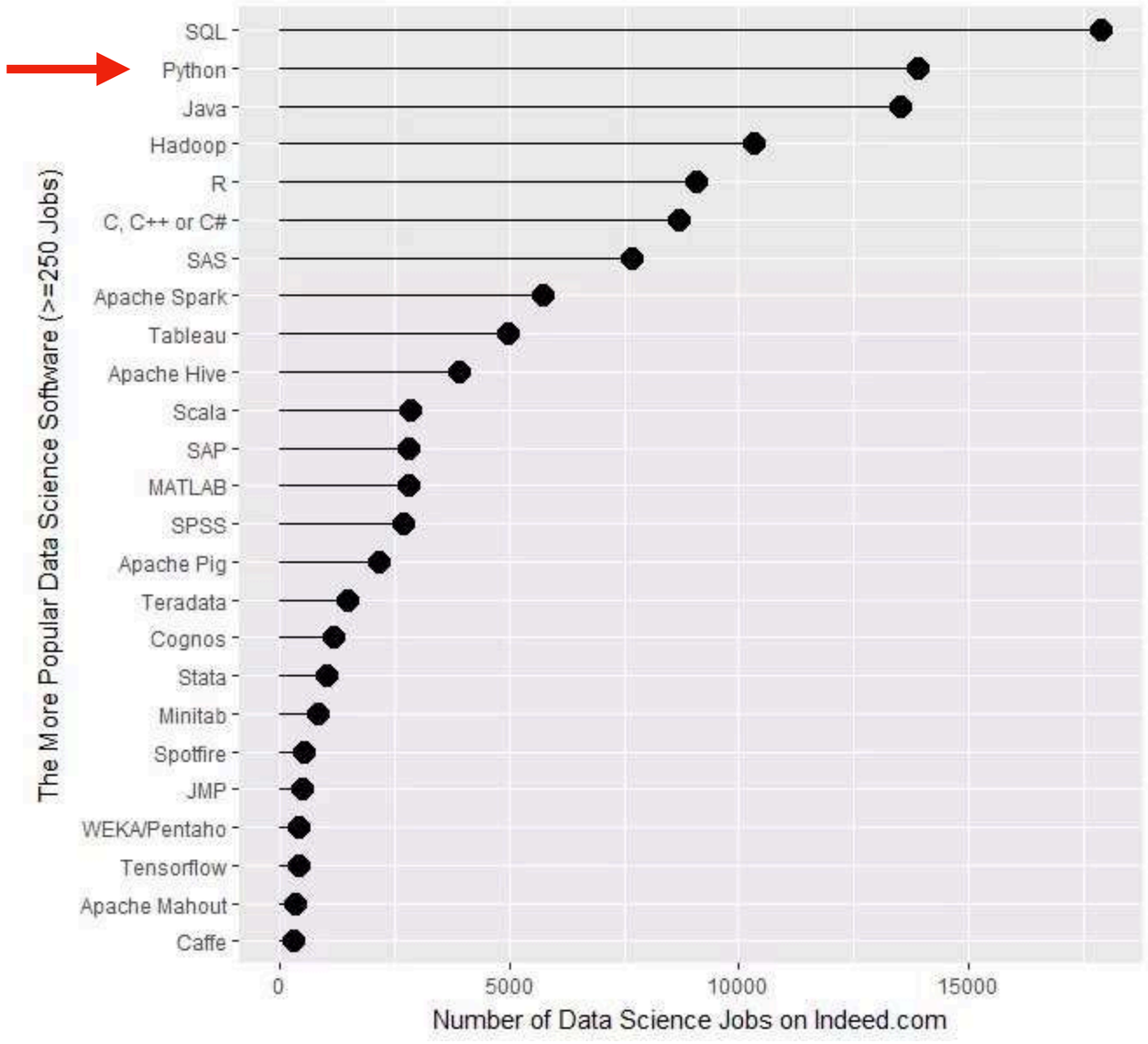
How many people here know a programming language?

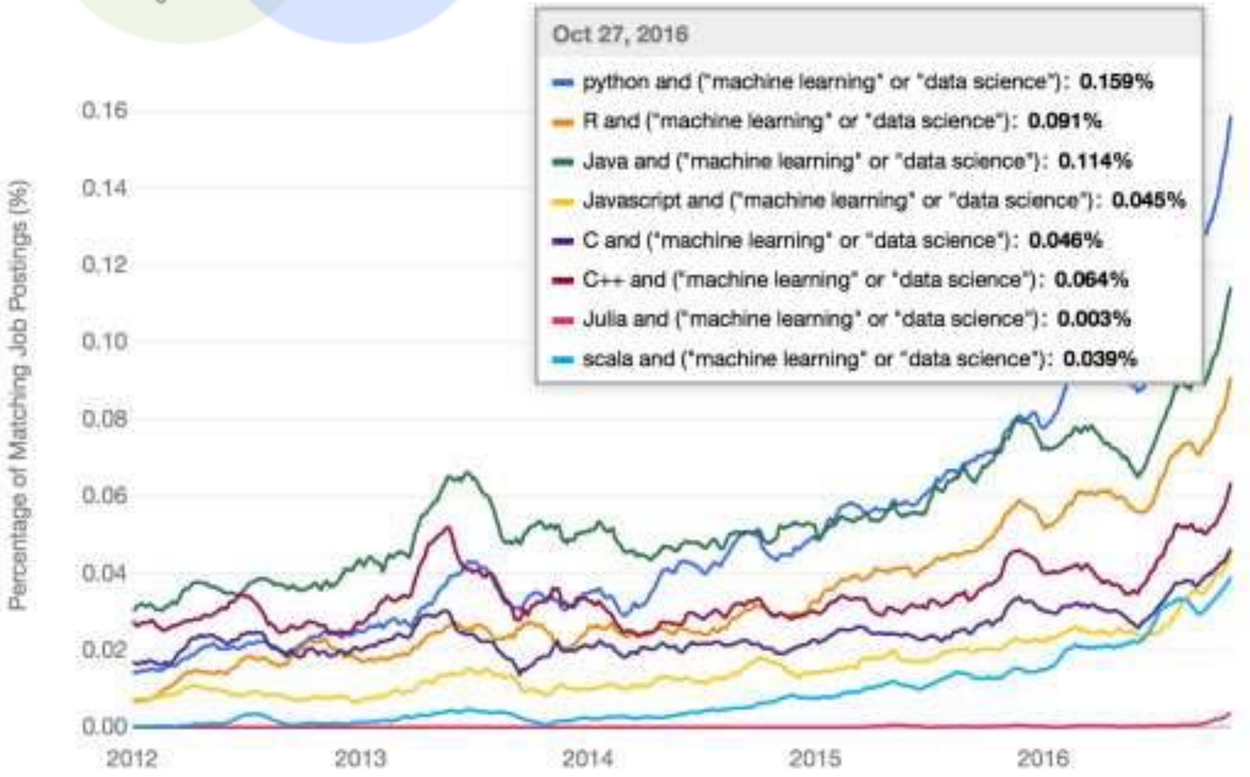
How many people here have used Python?

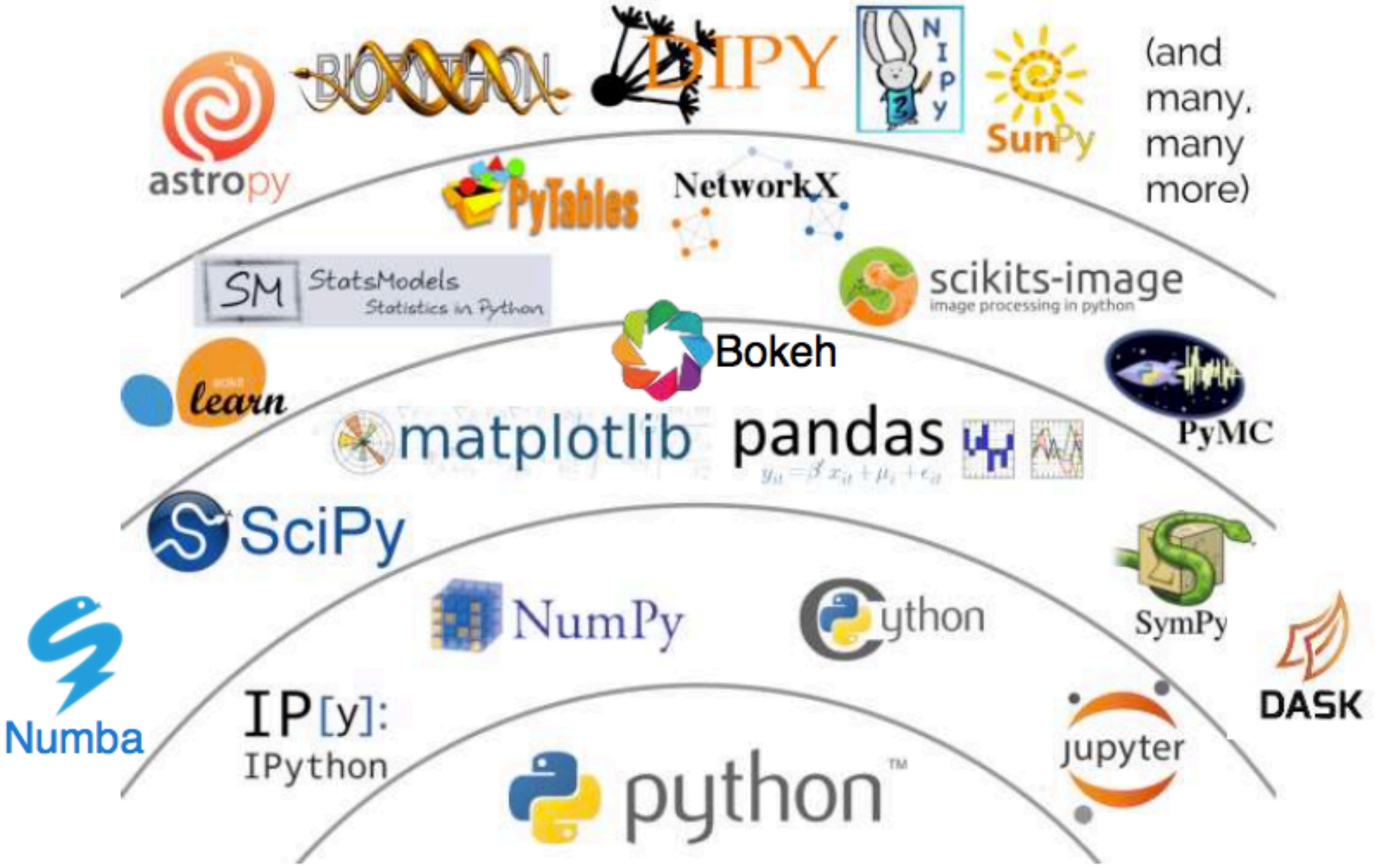
Python is a general-purpose language,
it can be used to build just about anything!

Documentation, tutorials available at:
www.python.org

Data Science is growing with Python leading the way






astropy
BIOPYTHON
IPY
NIPY
SunPy
(and many, many more)
PyTables
NetworkX
SM StatsModels Statistics in Python
scikits-image image processing in python
Bokeh
learn
matplotlib
pandas
PyMC
SciPy
NumPy
ython
SymPy
DASK
Numba
IP[y]: IPython
jupyter
python™


astropy
BIOPYTHON
IPY
NIPY
SunPy
(and many, many more)
PyTables
NetworkX
SM StatsModels Statistics in Python
scikits-image
image processing in python
Bokeh
learn
matplotlib
pandas
PyMC
SciPy
NumPy
Cython
SymPy
DASK
Numba
IP[y]: IPython
jupyter
python™


astropy
BIOPYTHON
IPY
NIPY
SunPy
(and many, many more)
PyTables
NetworkX
SM StatsModels Statistics in Python
scikits-image
image processing in python
Bokeh
learn
matplotlib
pandas
PyMC
SciPy
NumPy
ython
SymPy
DASK
Numba
IP[y]: IPython
python
jupyter


astropy
BIOPYTHON
IPY
NIPY
SunPy
(and many, many more)
PyTables
NetworkX
SM StatsModels Statistics in Python
scikits-image
image processing in python
Bokeh
learn
matplotlib
pandas
PyMC
SciPy
NumPy
ython
SymPy
DASK
Numba
IP[y]: IPython
python
jupyter


astropy
BIOPYTHON
IPY
NIPY
SunPy
(and many, many more)
PyTables
NetworkX
SM StatsModels Statistics in Python
scikits-image
image processing in python
learn
Bokeh
matplotlib
pandas
PyMC
SciPy
NumPy
ython
SymPy
DASK
Numba
IP[y]: IPython
jupyter
python™

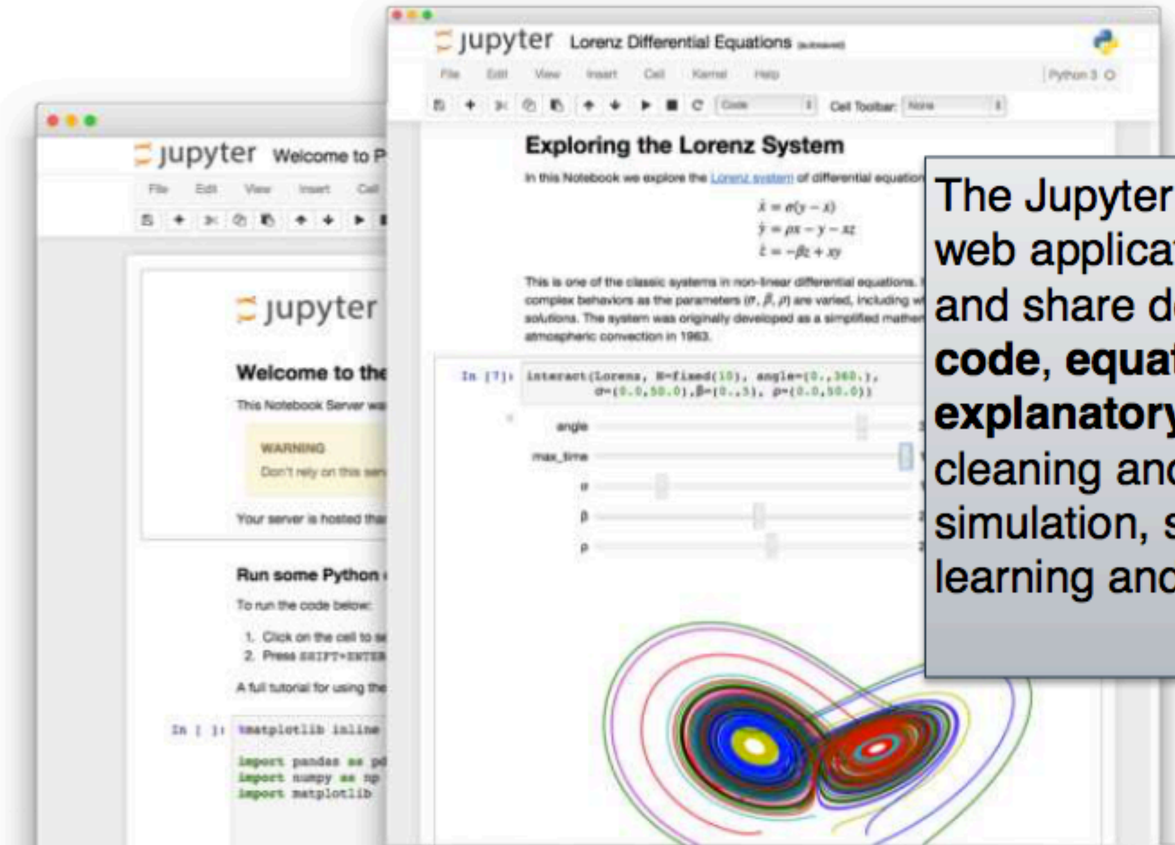
Course organisation and materials

This course will be completely hands-on!

To do so we will use Jupyter notebooks
(available within Anaconda)

Notebooks allow to:

- *Interactively* execute Python code
- Visualise results *step-by-step*



The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live **code**, **equations**, **visualizations** and **explanatory text**. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.

Course Outline

The course is organised in 3 sessions

Course Outline

The course is organised in 3 sessions

Session 1: Python Basics

Learn the basic concepts behind modern programming languages. We'll talk about:

- Variables & data structures
- Flow control
- I/O and exceptions

Course Outline

The course is organised in 3 sessions

Session 1: Python Basics

Learn the basic concepts behind modern programming languages. We'll talk about:

- Variables & data structures
- Flow control
- I/O and exceptions

Session 2: Data manipulation in Pandas

Learn how to load, access and transform datasets. We'll talk about:

- DataFrames
- Slicing, Indexing, Grouping…
- …

Course Outline

The course is organised in 3 sessions

Session 1: Python Basics

Learn the basic concepts behind modern programming languages. We'll talk about:

- Variables & data structures
- Flow control
- I/O and exceptions

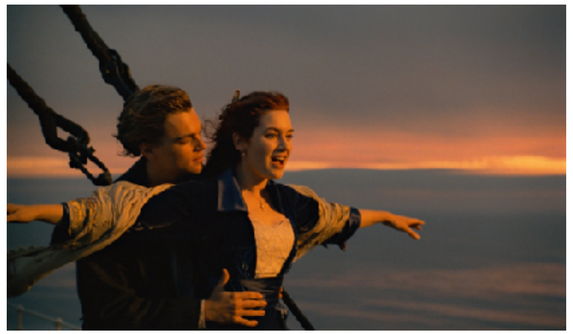
Session 2: Data manipulation in Pandas

Learn how to load, access and transform datasets. We'll talk about:

- DataFrames
- Slicing, Indexing, Grouping…
- …

Session 3: Case Study

Who is more likely to survive to the infamous Titanic shipwreck?



Let’s begin!

Install **Anaconda**
(Python 3 version!)
https://www.anaconda.com/download/

Anaconda is a python distribution for data scientists.

It provides out-of-the box support for:

- Data modeling
- Analysis
- Visualisation
- …