PYTHON, PANDAS AND JUPYTER TUTORIAL

DSA 2019

Claire Babirye and Neema Mduma

June 3rd, 2019

Outline

Python

Python Packages for Data Science

Practical session

Introduction

A very popular general-purpose programming language.

- Open source general-purpose language
- Dynamically semantics (rather than statically typed like Java or C/C++)
- Interpreted (rather than compiled like Java or C/C++)

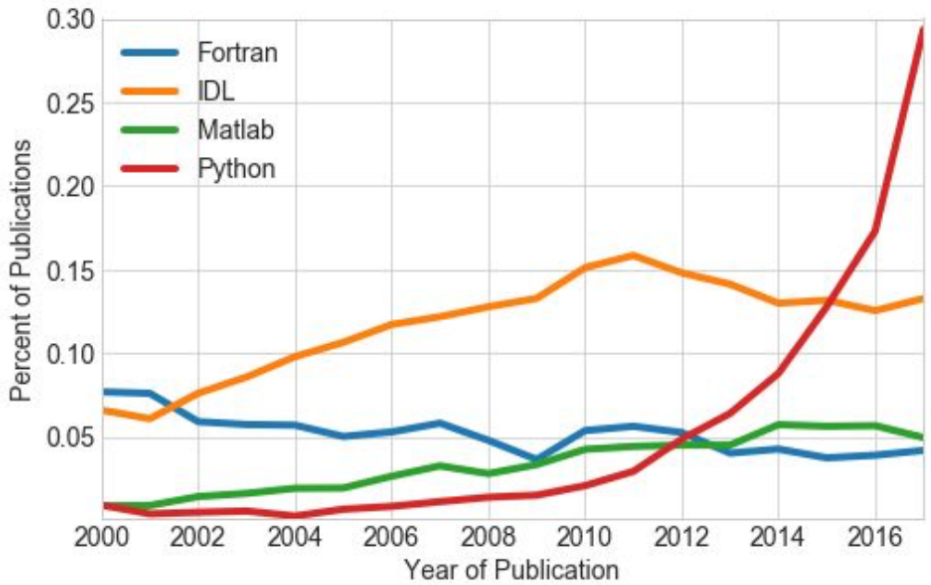
Python Applications

- Web development (Django)
- Web scraping (Beautiful Soup)
- Scripting language.
- Scientific programming and numeric computing.
- Automation and embedded system.
- Desktop GUIs and 3D modeling.

Importance of Python

- Python is a teaching language
- Created to bridge the gap between the shell and C

Mentions of Software in Astronomy Publications:



Compiled from NASA ADS (code).

Thanks to Juan Nunez-Iglesias, Thomas P. Robitaille, and Chris Beaumont.

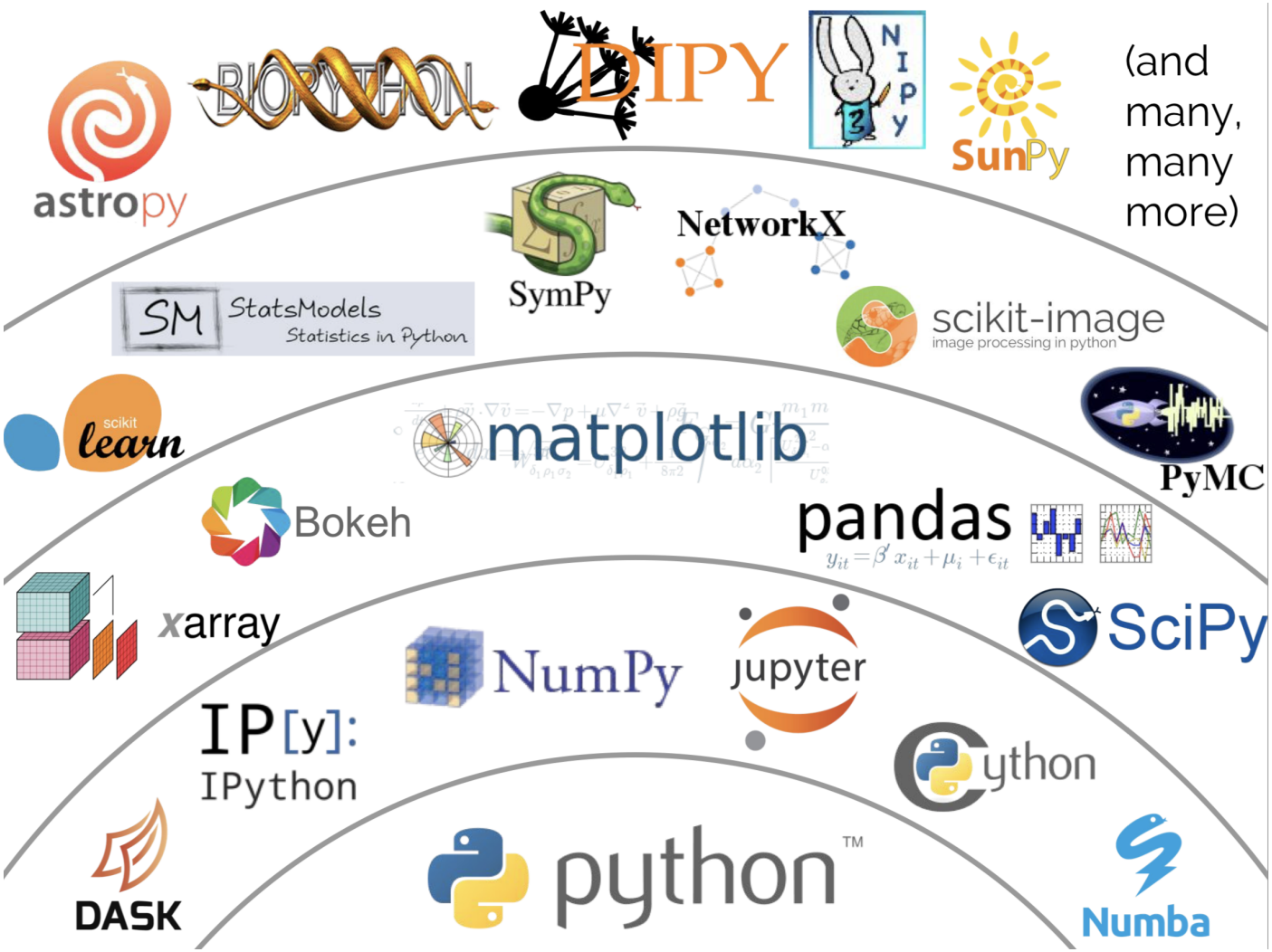
Outline

Python

Python Packages for Data Science

Practical session

Python Scientific Stack



Jupyter

Open-source web application for interactive and exploratory computing.

- Allows to create and share documents that contain live code, equations, visualizations and explanatory text.
- It is a platform for Data Science at scale.
- Covers all the life-cycle of scientific ideas to publications.
- Demo

Numpy and Sci-py

Numpy: The fundamental Python package for scientific computing.

- Provide high-performance vector, matrix and higher-dimensional data structures.

Sci-py: Collections of high level mathematical operations

- Linear algebra
- Optimization
- Integration etc

Matplotlib

Matplotlib is an excellent 2D and 3D graphics library for generating scientific figures.

- It provides both a very quick way to visualize data from Python and publication-quality figures in many formats.

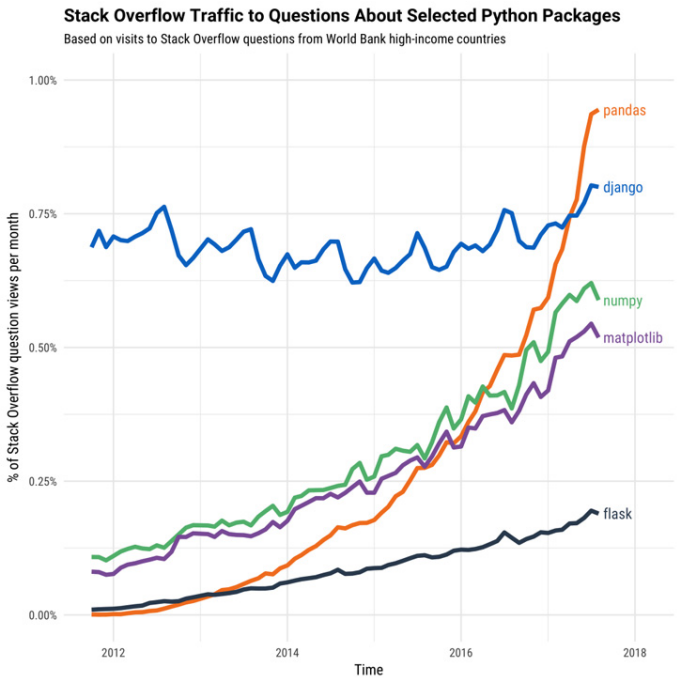
Other data visualization packages: Seaborn and Bokeh.

Pandas

A python package providing fast, flexible, and expressive data structures for data analysis.

- A fundamental high-level building block for doing practical, real world data analysis in Python.
- Designed to work with relational or labeled data or both.

Pandas



Outline

Python

Python Packages for Data Science

Practical session

Practical session