Introduction

Janis Keuper

Introduction to ML

ML Summer School -Janis Keuper

INSTITUTE FOR MACHINE LEARNING AND ANALYTICS
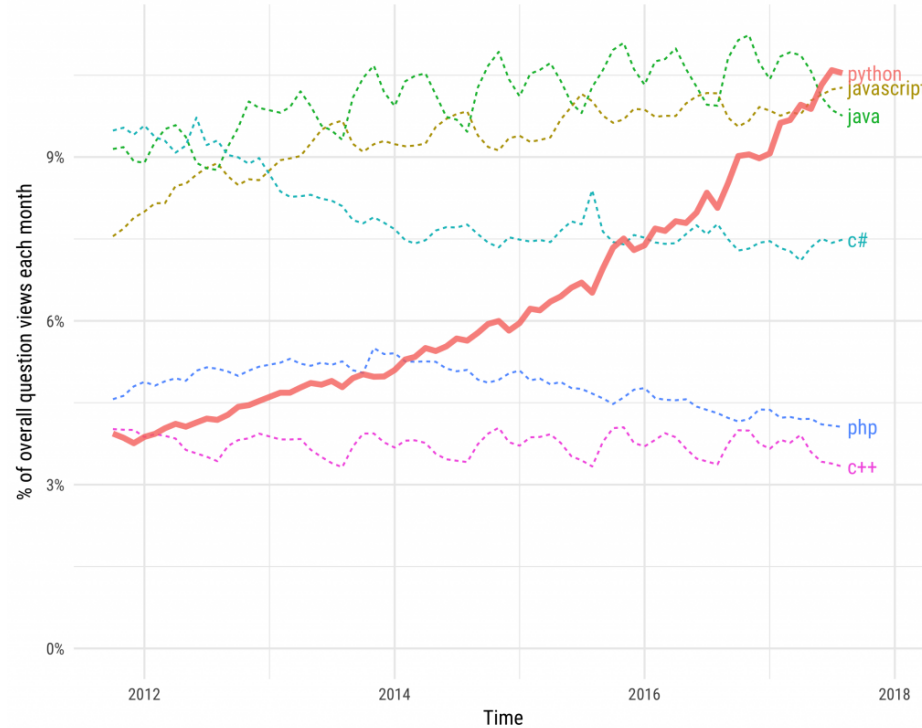
Introduction

Python for Data Science

- We will use Python (more exactly tools with python interface) for this course

- Python is the most dominant language used in Data Science, Dats Analytics and Machine Learning to day
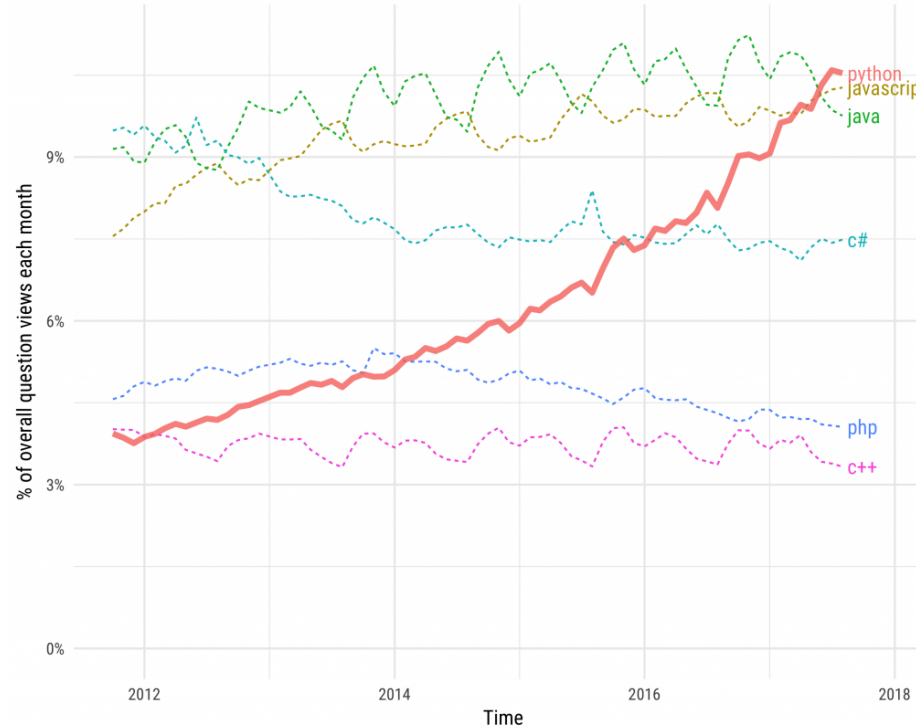
- Alternatives: R and Julia

Python for Data Science

Growth of major programming languages
Based on Stack Overflow question views in World Bank high-income countries

[2]

Python for Data Science

- Triggered by ML + Data Science

- Python pushes convergence of technologies:
  - HPC
  - Cloud
  - Big-Data

Growth of major programming languages

Based on Stack Overflow question views in World Bank high-income countries
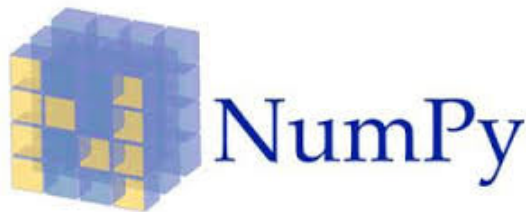

[2]

Python for Data Science

- Very low entry barrier (easy to learn)

- Quite universal in programming approaches

- Easy to interface existing high peformance libs

- Huge Community

- **Libraries !!!**

- **Python as „glue code" for rapid development and markup**

Python for Data Science

https://www.anaconda.com/distribution/

- Easy installation on Linux/Mac/Win
  - See read.me in Moodle
  - User space install (no admin needed)
- Conda Python package management
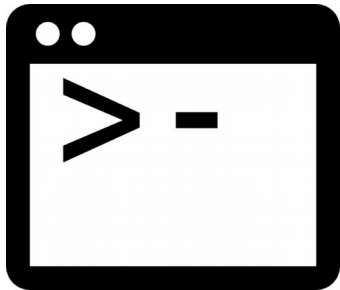 - Simple install/update/sharing of software
- All open source!

Python for Data Science

NOTE:

There are two versions of Python:

Python2 and Python3

Support for Python2 will end 31.12.19

→ get started with Python3!

Python for Data Science

Interactive Python Introduction

→ Start lab session now!