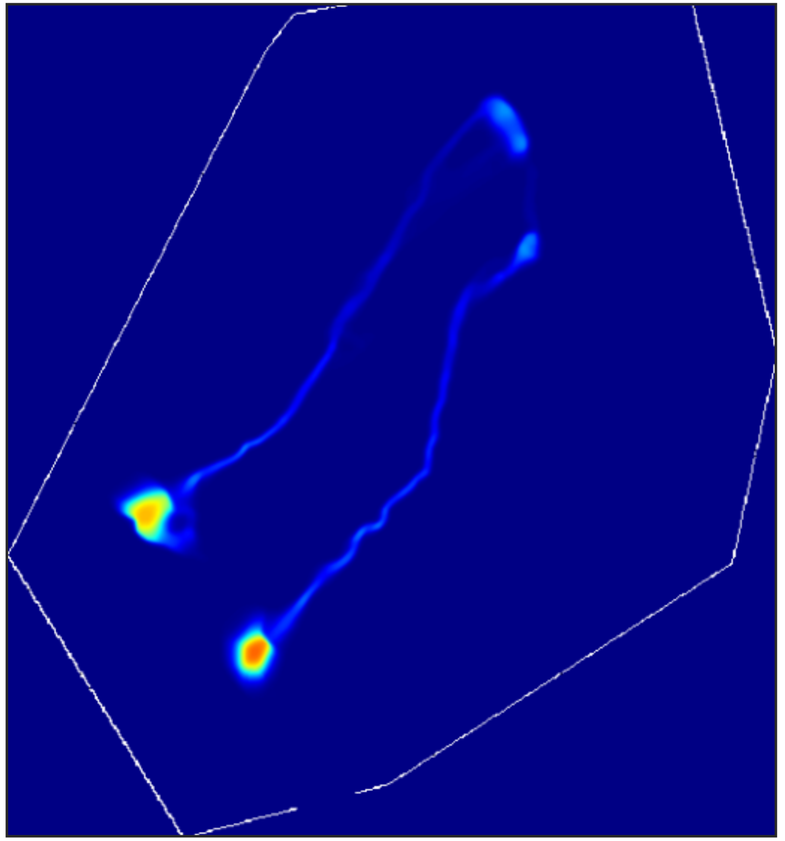
Introduction to Data Science and Programming, Fall 2019

Class 11: Scientific programming with numpy

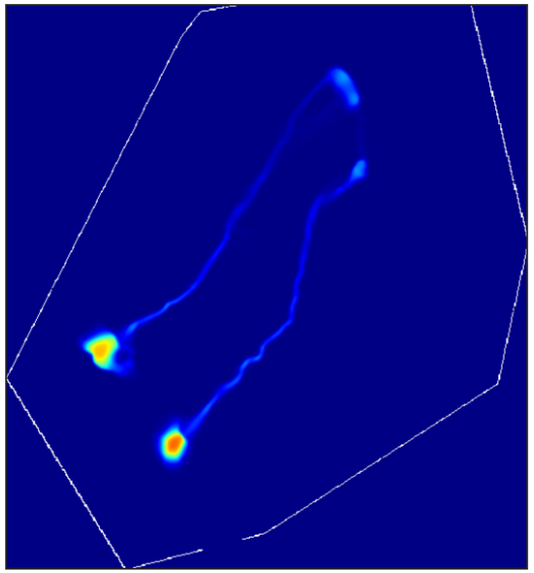
Instructor: Michael Szell
Oct 2, 2019



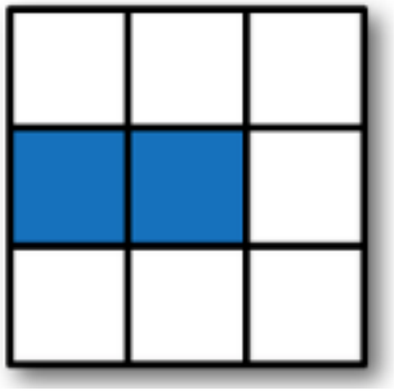
IT UNIVERSITY OF COPENHAGEN

Today you will learn why, when, and how to use numpy

Scientific computing



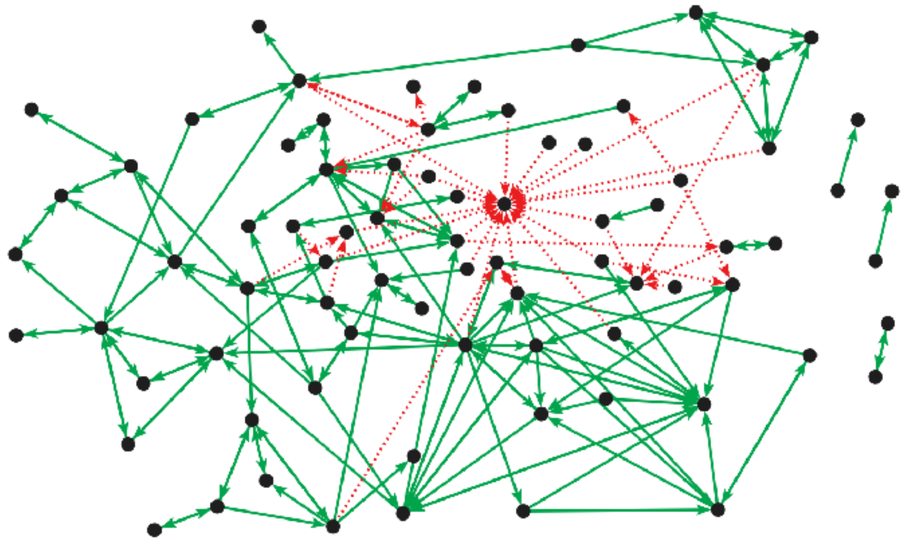
Array manipulation

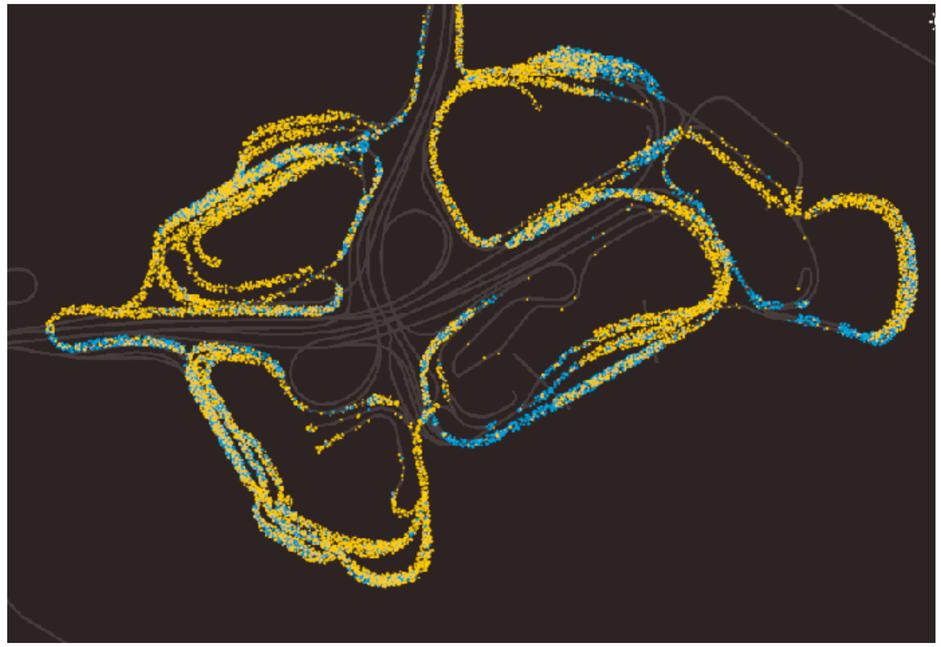


Data processing

Introduction: Michael Szell

Data Scientist researching:





Social networks

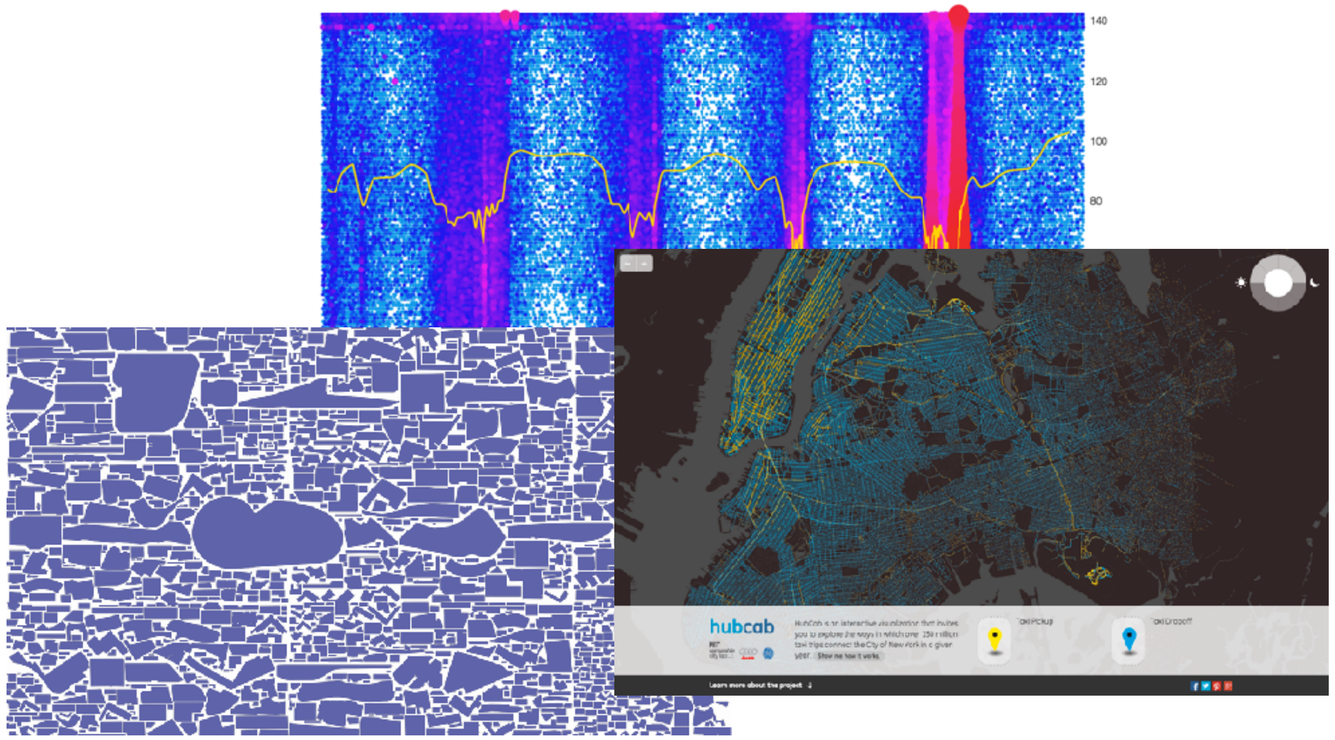
MATLAB, Python

Urban mobility

Python

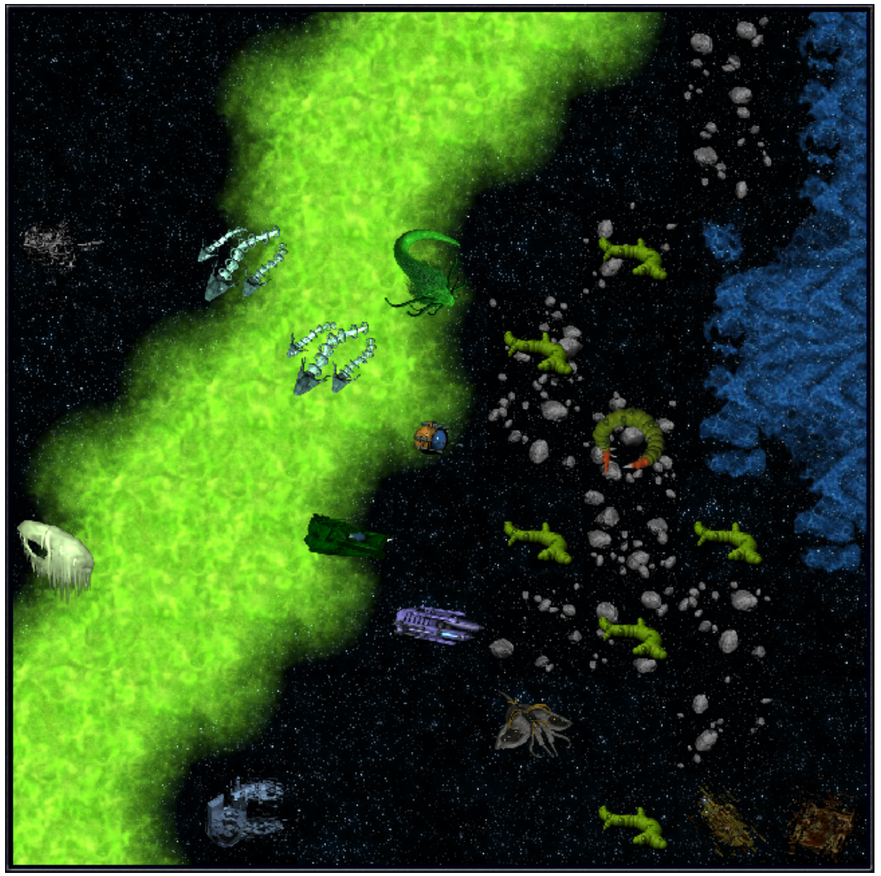
Introduction: Michael Szell

Data Scientist creating:



Data Visualizations

Python, Javascript, MongoDB



MMOG

PHP, C/C++, MySQL

Introduction: Michael Szell

Data Scientist having background in / working with:

Physicists

Architects, Urban planners

Industrial designers

Computer scientists

Game developers

Mathematicians

Economists

Medical doctors

Contact me at

First: Student colleagues, TAs, LearnIT Forum

4E04, Mondays 13:00-14:00 (or after appointment)
misz@itu.dk

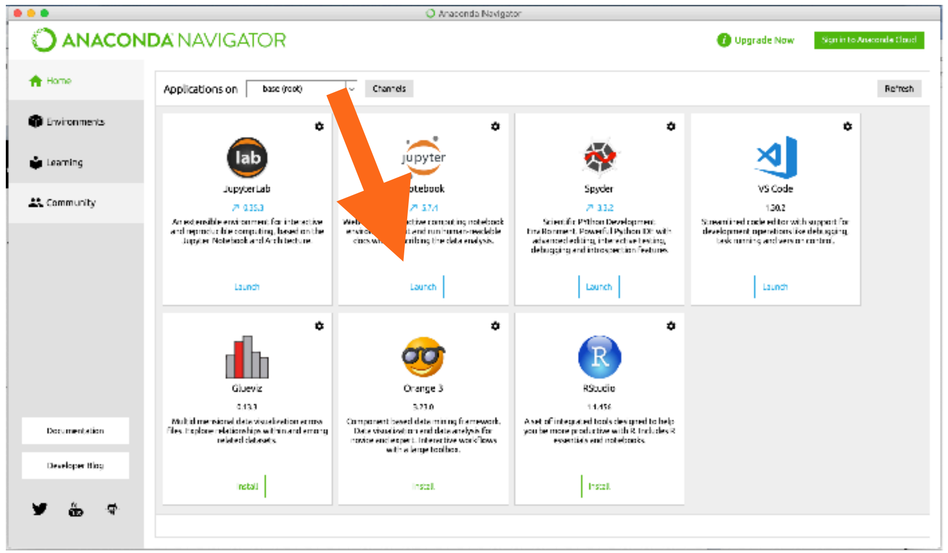
Questions?

We will use Jupyter notebooks with Python 3.7

In the terminal: `jupyter notebook`

or `jupyter lab`

or With anaconda navigator:

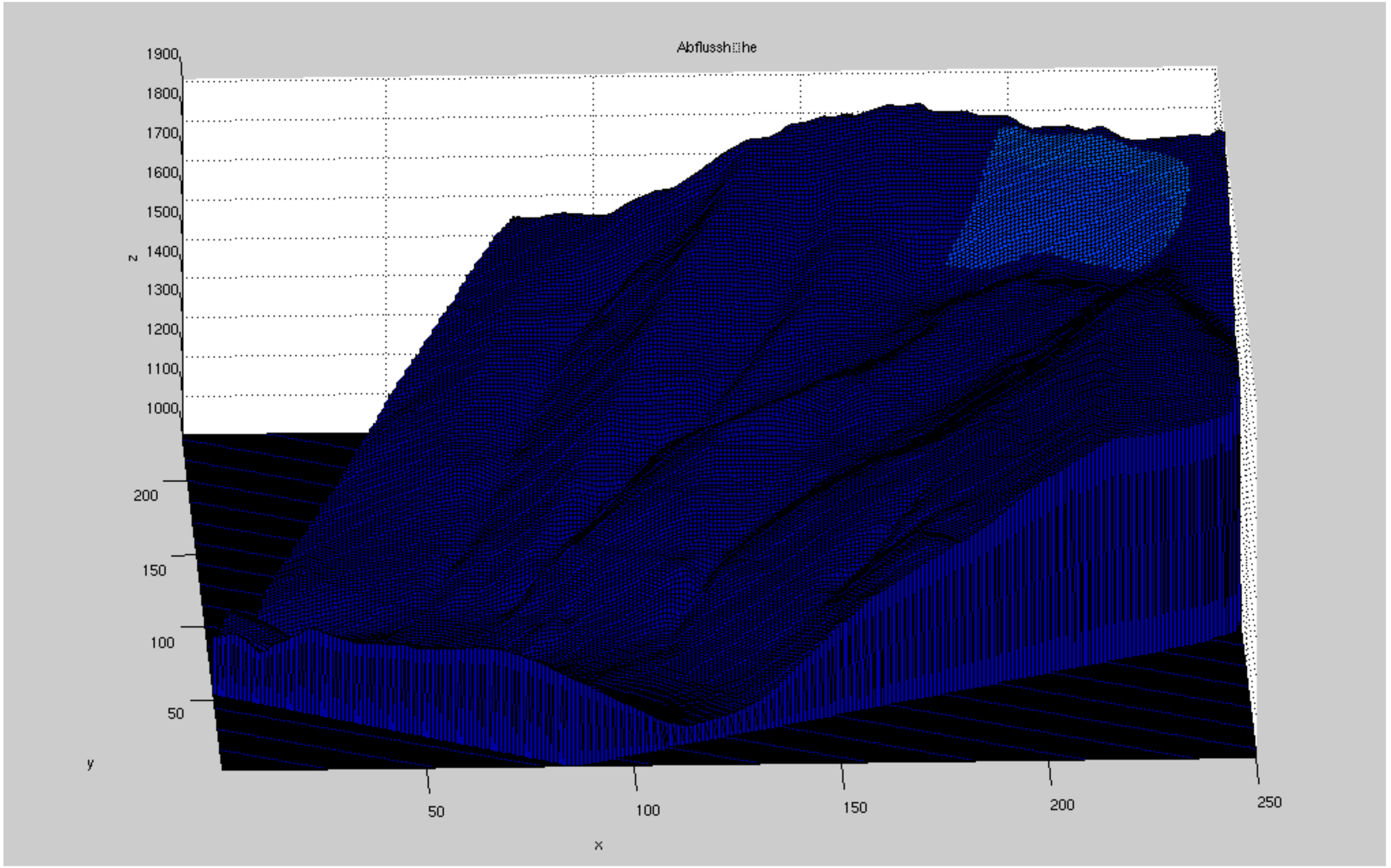


We will use Jupyter notebooks with Python 3.7

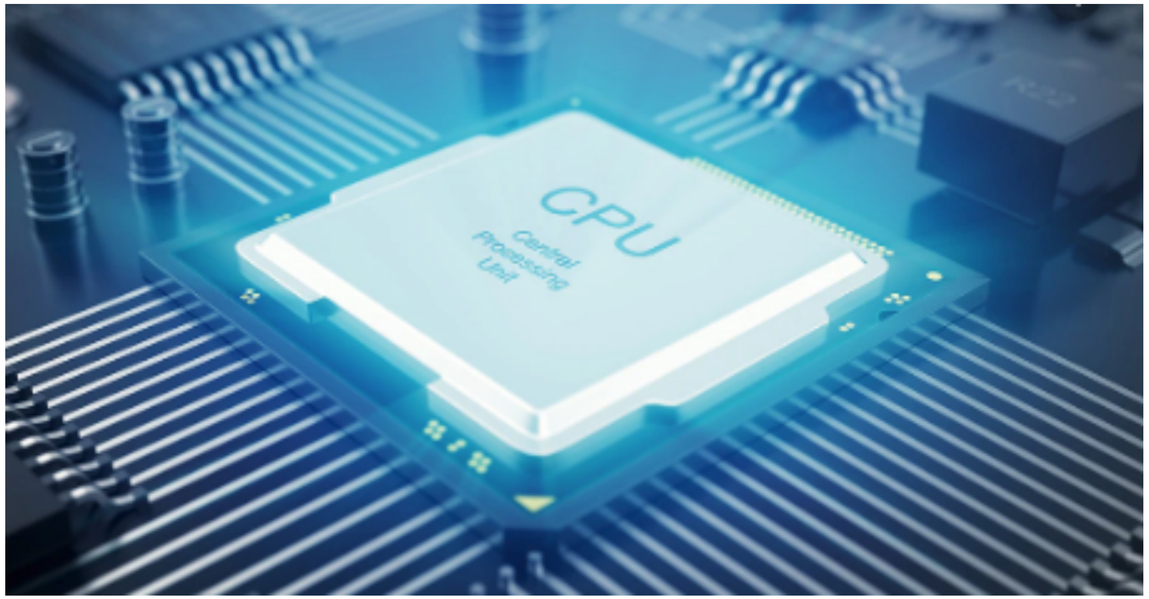
You are responsible for setting up a working Jupyter environment

Before each class, download and unzip the classmaterial
from learnit, and start the Jupyter notebook

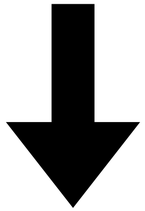
When we do repeated calculations on matrices, standard Python is inefficient



Vectorization is the ability to run operations on vectors instead of single numbers




```
c = []
for i in range(n):
    c.append(a[i]*b[i])
```



```
c = a*b
```

numpy is optimized for vectorized operations

NumPy = Numerical Python

- `ndarray` with vectorized operations
- Vectorized mathematical functions
- Data I/O
- Linear algebra, random numbers
- Using pre-compiled C

Use numpy whenever you run into limitations with lists and dicts

Speed

>1D-data

Masking / slicing

Use numpy whenever you run into limitations with lists and dicts

Speed

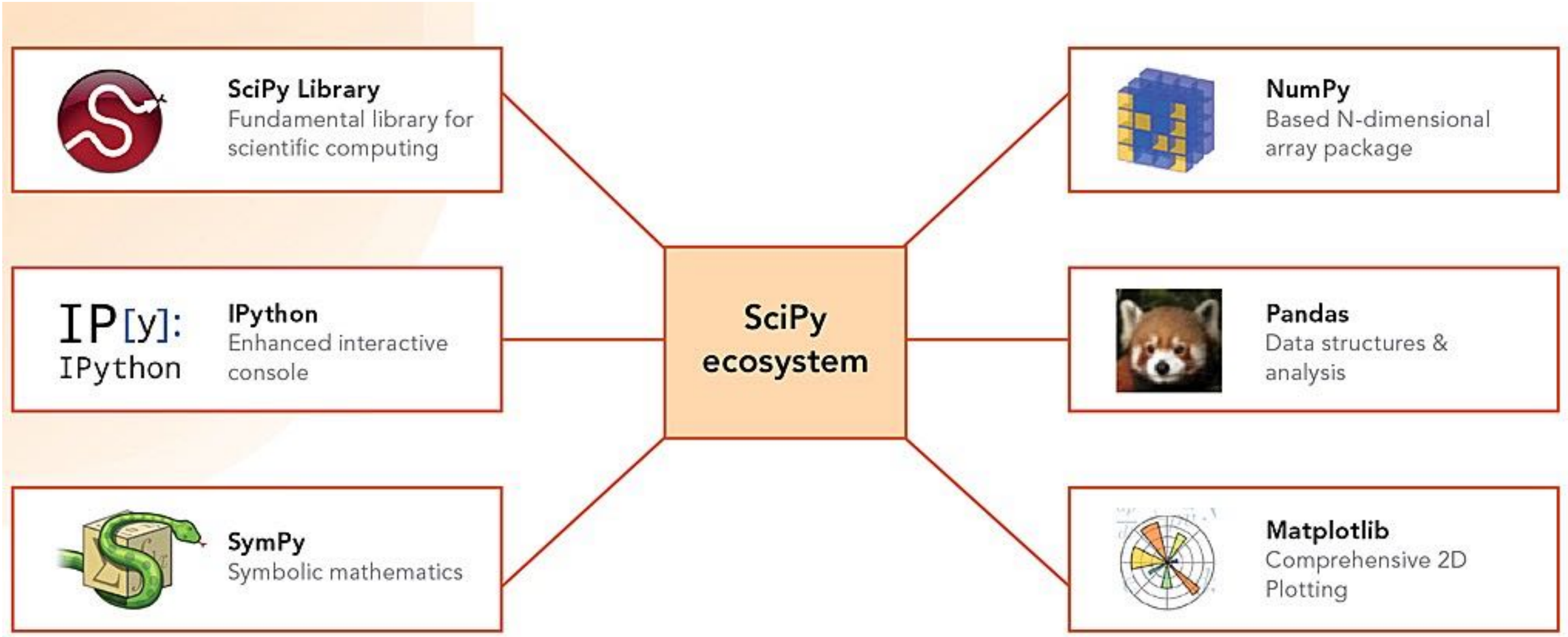
>1D-data

Masking / slicing

If you run into limitations with numpy, try Pandas (not covered here):

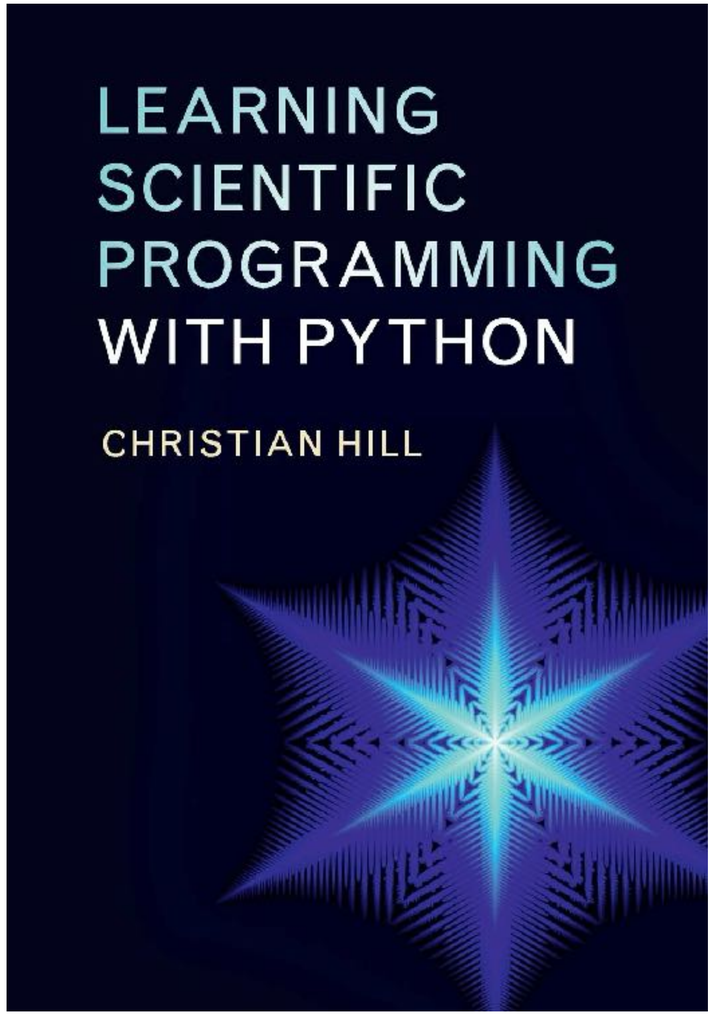
Merging or reshaping data sets

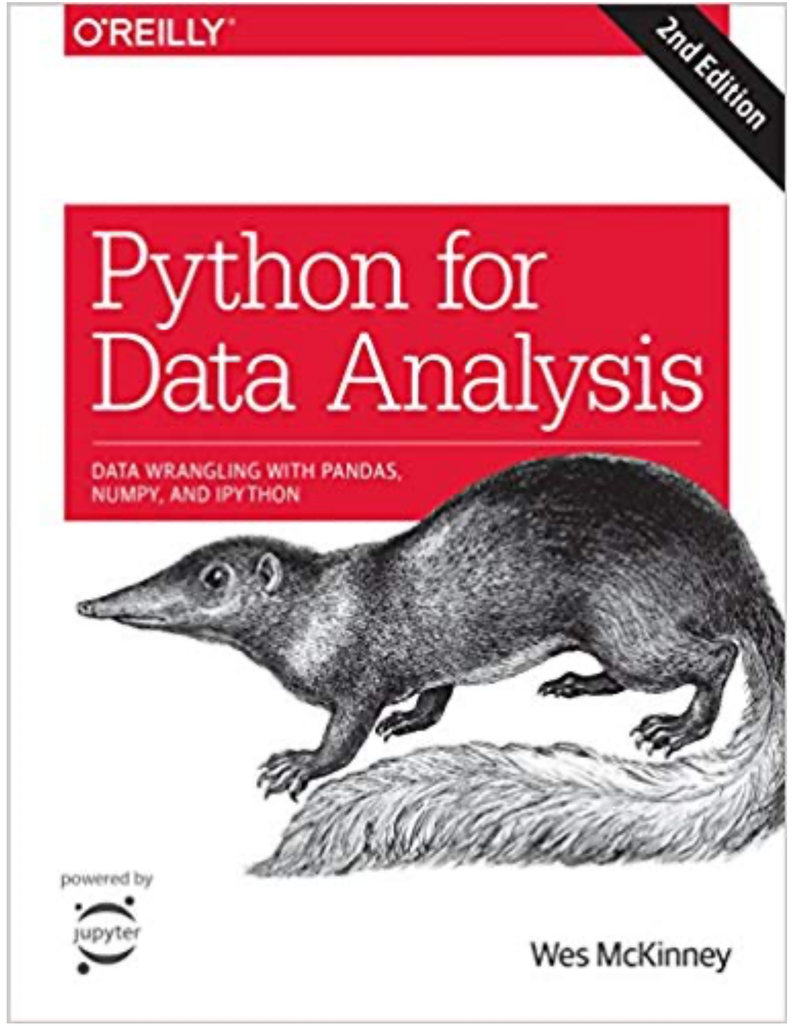
More I/O options (Excel, SQL,..)



https://blog.k2datascience.com/ds-fundamentals-scipy-stack-dcfc48576f3

Sources and further materials for today's class





https://www.oreilly.com/library/view/python-for-data/9781449323592/ch04.html