Scientific Tools for Python

Bartosz Teleńczuk

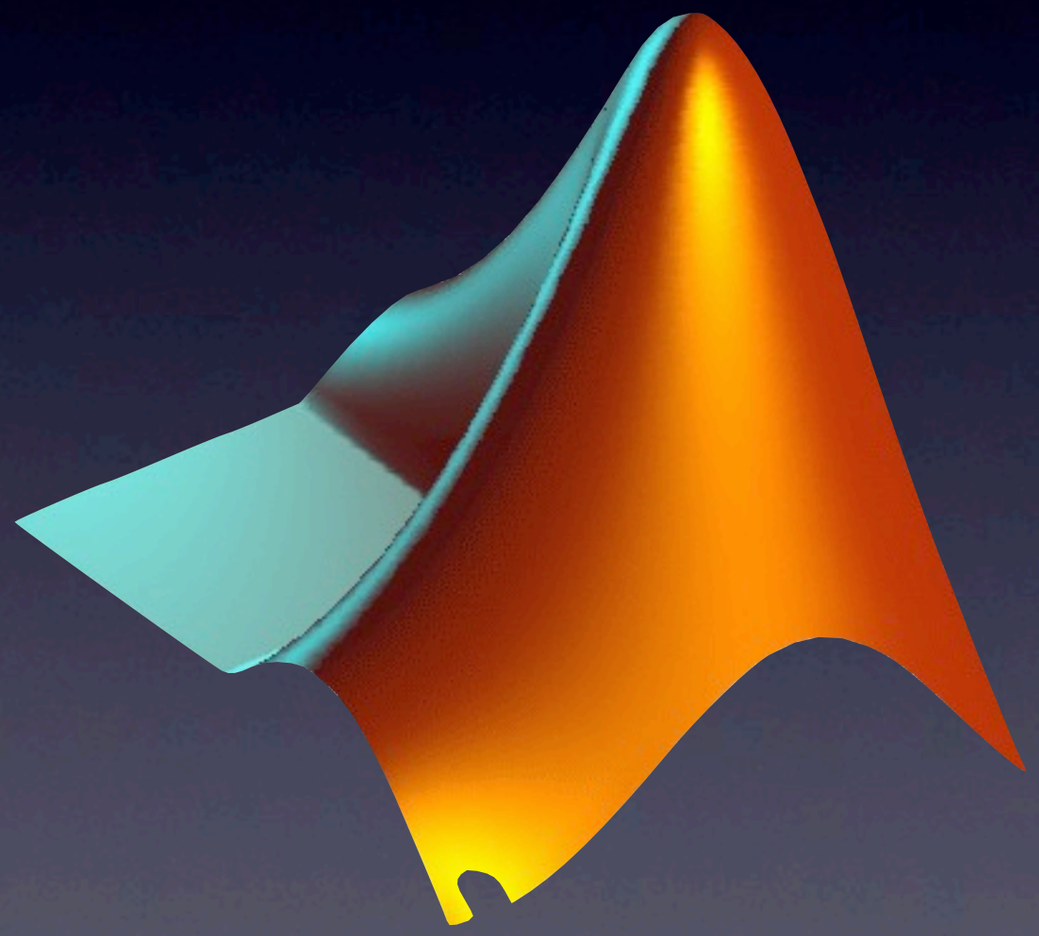
Advanced Scientific Programming in Python

Warsaw 2010

Python was rapidly adopted

- web development
- database programming
- prototyping
- scripting language
- game programming
- GUIs

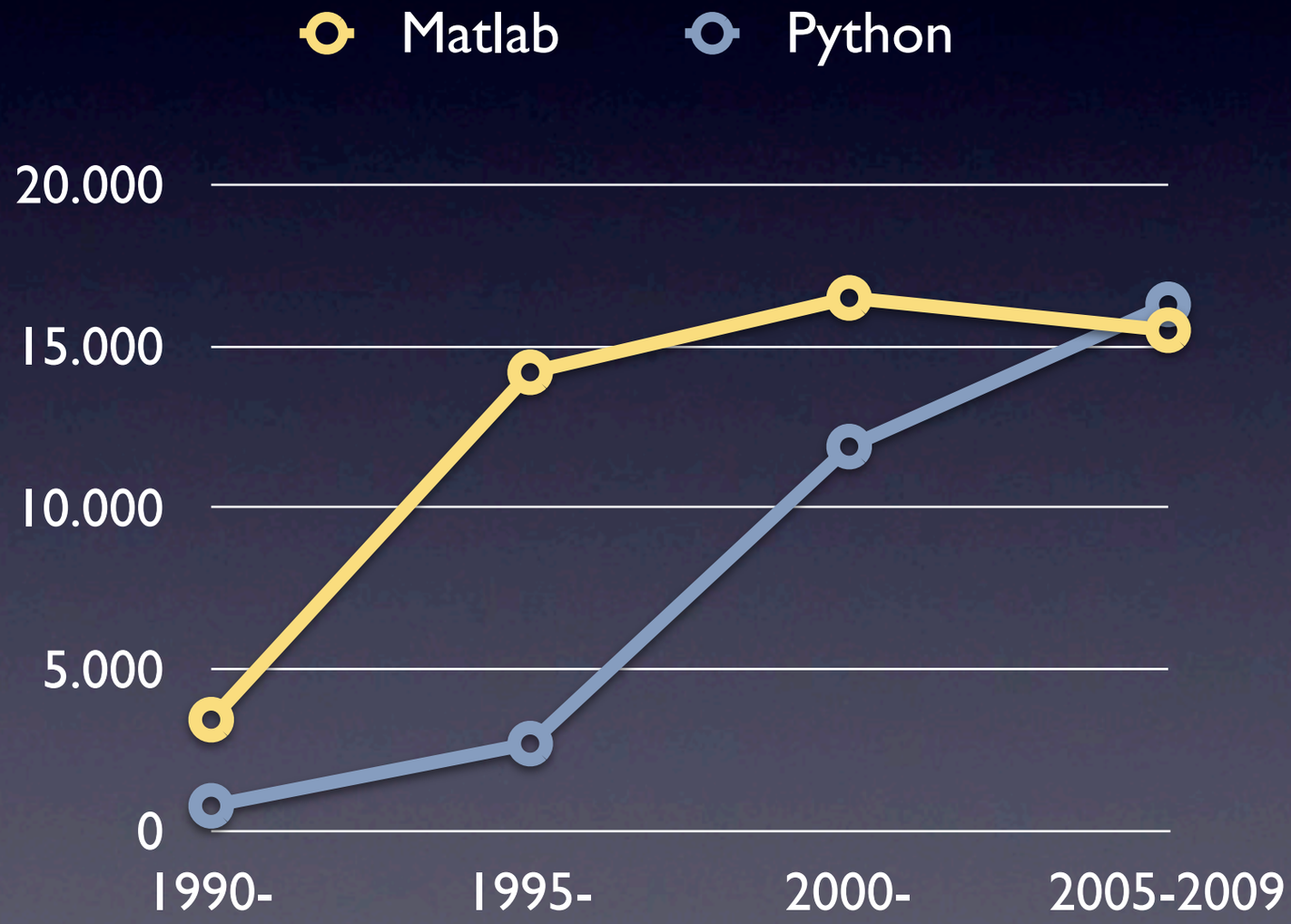
What about science?



...but WHY?!?!

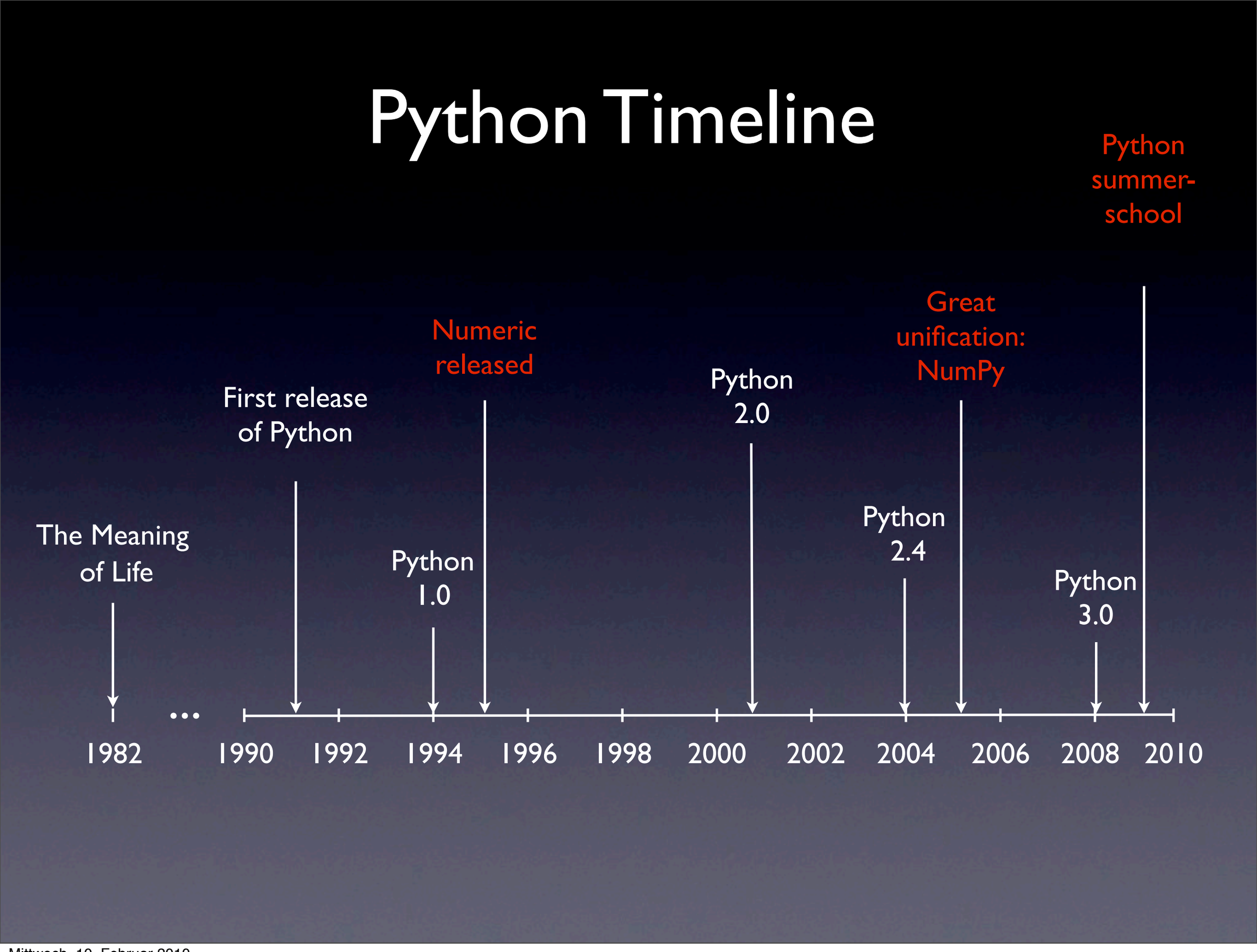
- complex data types
- numerical algorithms (linear algebra, etc...)
- plotting
- user-contributed functions
- good documentation
- full IDE

Number of publications



Source: Google Scholar

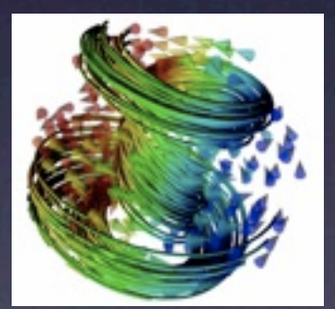
Python Timeline



Big BOOM!!!!!

Numerical Optimization

Symbolic mathematics



3D Scientific Visualization

Data Mining

What is NumPy?

"What really makes Python excel as a language for scientists and engineers is the NumPy extension."

Travis Oliphant

- efficient implementation of an array object
- I/O function
- basic statistics, linear algebra, ...

NumPy Examples

interactive session

Again M...b®

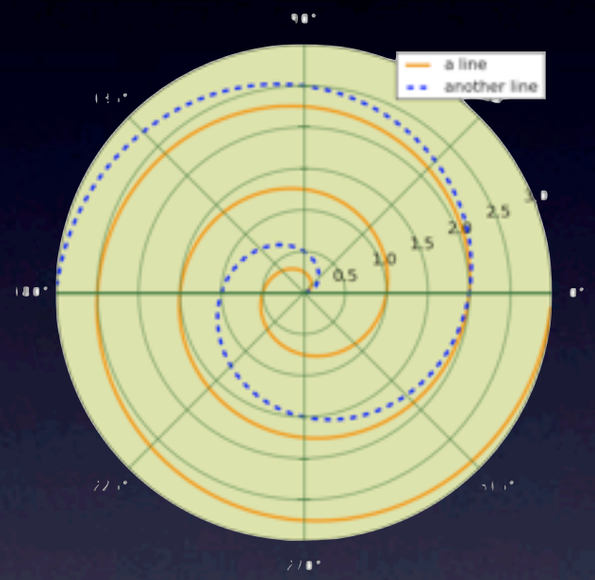
- ~~complex data types~~ → NumPy
- numerical algorithms
- plotting
- user-contributed functions
- good documentation
- full IDE

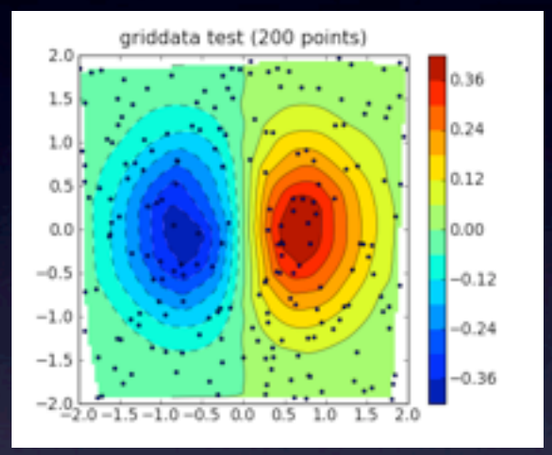
scipy.stats	statistical functions
scipy.integrate	integration routines
scipy.optimize	optimization tools
scipy.signal	signal processing tools
scipy.sparse	sparse matrices
scipy.cluster	clustering algorithms

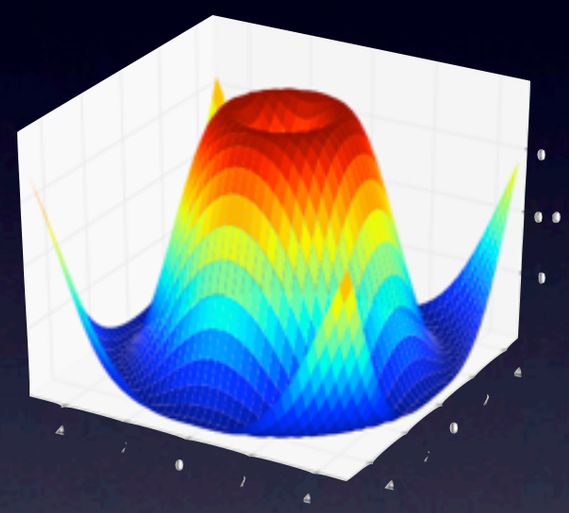
Again M...b®

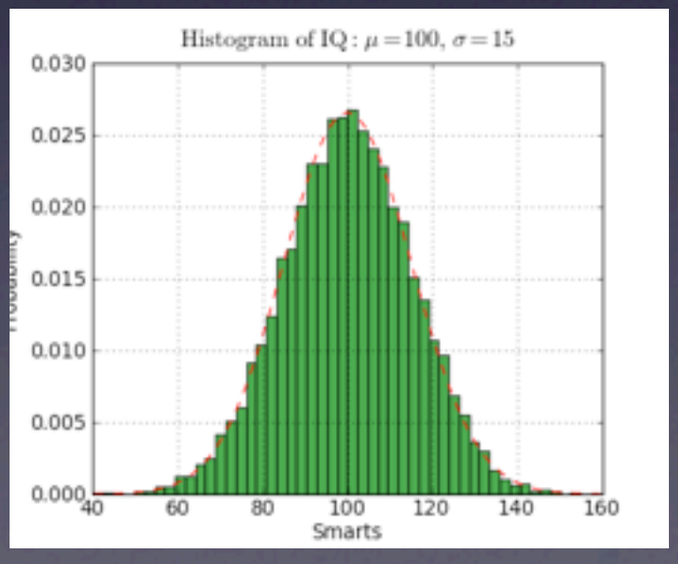
- ~~complex data types~~ → NumPy
- ~~numerical algorithms~~ → SciPy, ...
- plotting
- user-contributed functions
- good documentation
- full IDE

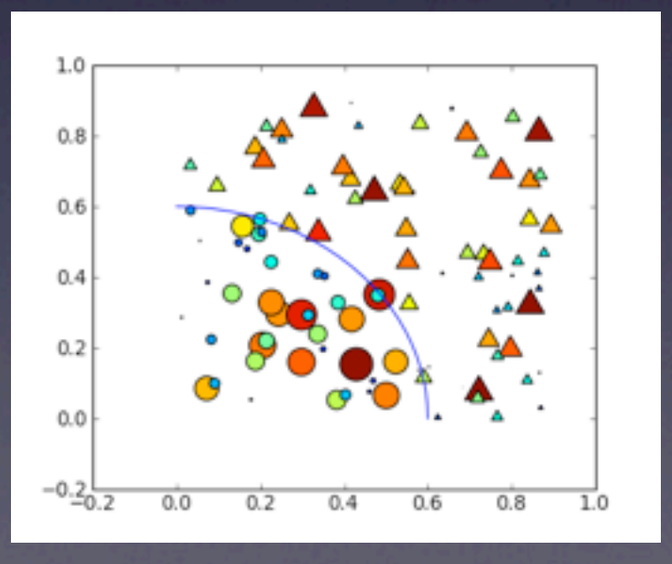
matplotlib

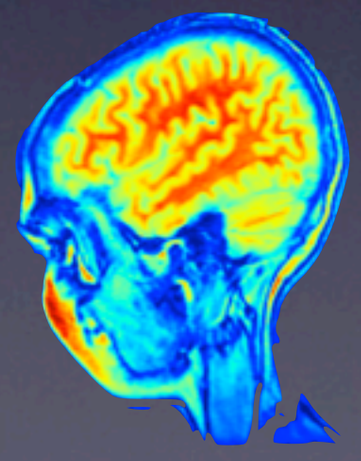

a line
another line
0.5
1.0
1.5
2.0
2.5


griddata test (200 points)
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0
0.36
0.24
0.12
0.00
−0.12
−0.24
−0.36



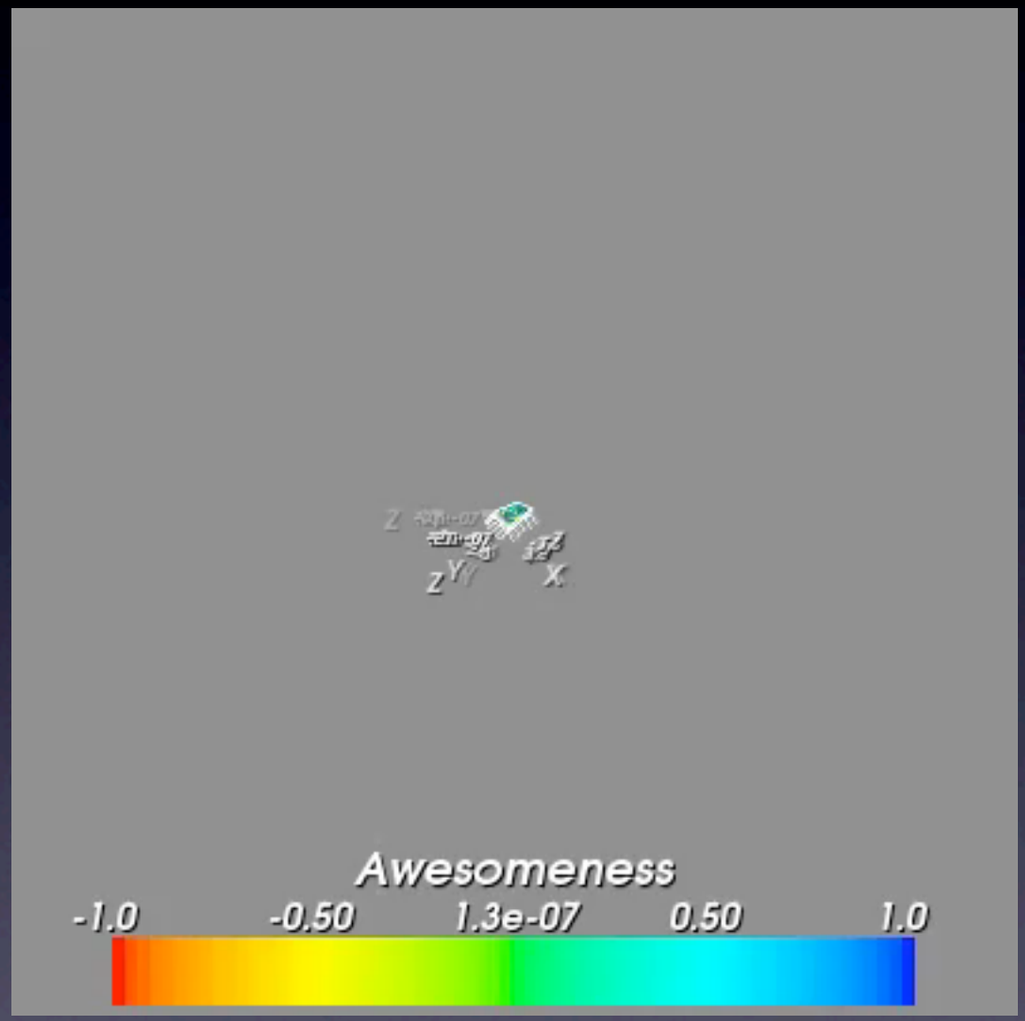

Histogram of IQ: $\mu = 100$, $\sigma = 15$
Probability
0.030
0.025
0.020
0.015
0.010
0.005
0.000
40
60
80
100
120
140
160
Smarts


1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.2
0.0
0.2
0.4
0.6
0.8
1.0



Matplotlib examples

MayaVI



Contributed by Eillif Miller

Again M...b®

- ~~complex data types~~ → NumPy
- ~~numerical algorithms~~ → SciPy, ...
- ~~plotting~~ → matplotlib, ...
- user-contributed functions
- good documentation
- full IDE

SciKits

Docs

Again M...b®

- ~~complex data types~~ → NumPy
- ~~numerical algorithms~~ → SciPy, ...
- ~~plotting~~ → matplotlib, ...
- ~~user-contributed functions~~
- ~~good documentation~~
- full IDE

Serialization

Scientific data

- large data sets
- inhomogenous
- meta-data
- searchable
- platform independent
- collaborative

State of the art

- closed formats
- set of unrelated and badly-labelled files
- metadata stored separately (if at all)
- data repositories closed to outsiders

Python to rescue?

- pickles
- numpy (record) arrays
- relational databases
- HDF5

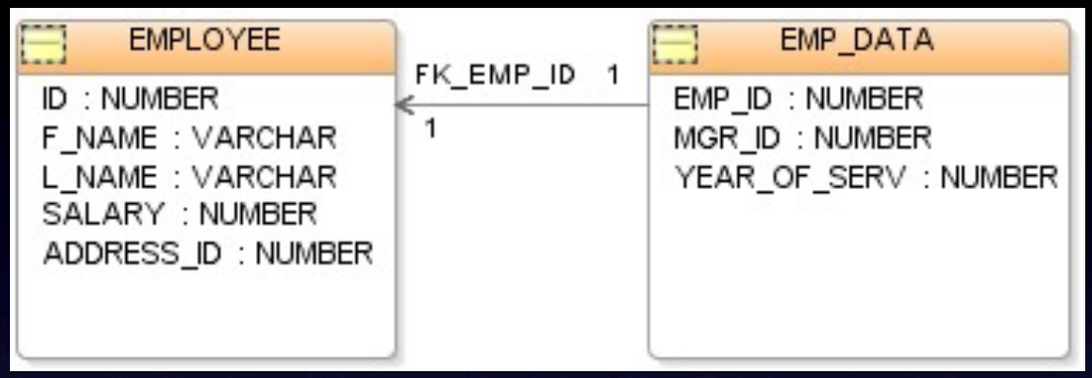
pickle

- standard library
- works with (almost) all Python objects
- Python-specific
- insecure

NumPy Record Arrays

- spreadsheet-like data structures
- picklable (or stored in binary format)
- Python-specific

Relational Databases



- define tables with data description and relations between them
- implement fast search, grouping and sorting
- implemented by many database systems

- hierarchical dataset.
- built on top of HDF5 (API for C, C++, F90, Java)
- very efficient on large datasets
- object-oriented design