

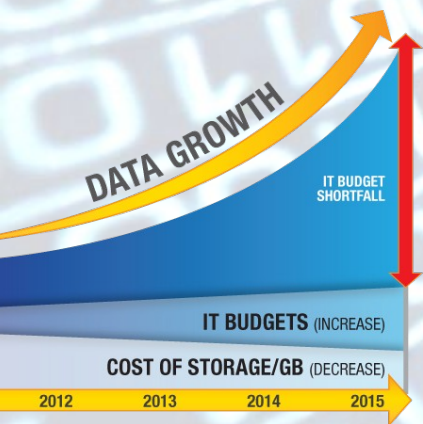
Introduction to Python Tools

André Panisson - ISI Foundation

andre.panisson@isi.it

Complex Networks Thematic School, Les Houches, April 7-18, 2014

- There is a common belief that
“the data speak by themselves”
(Peter Gould)

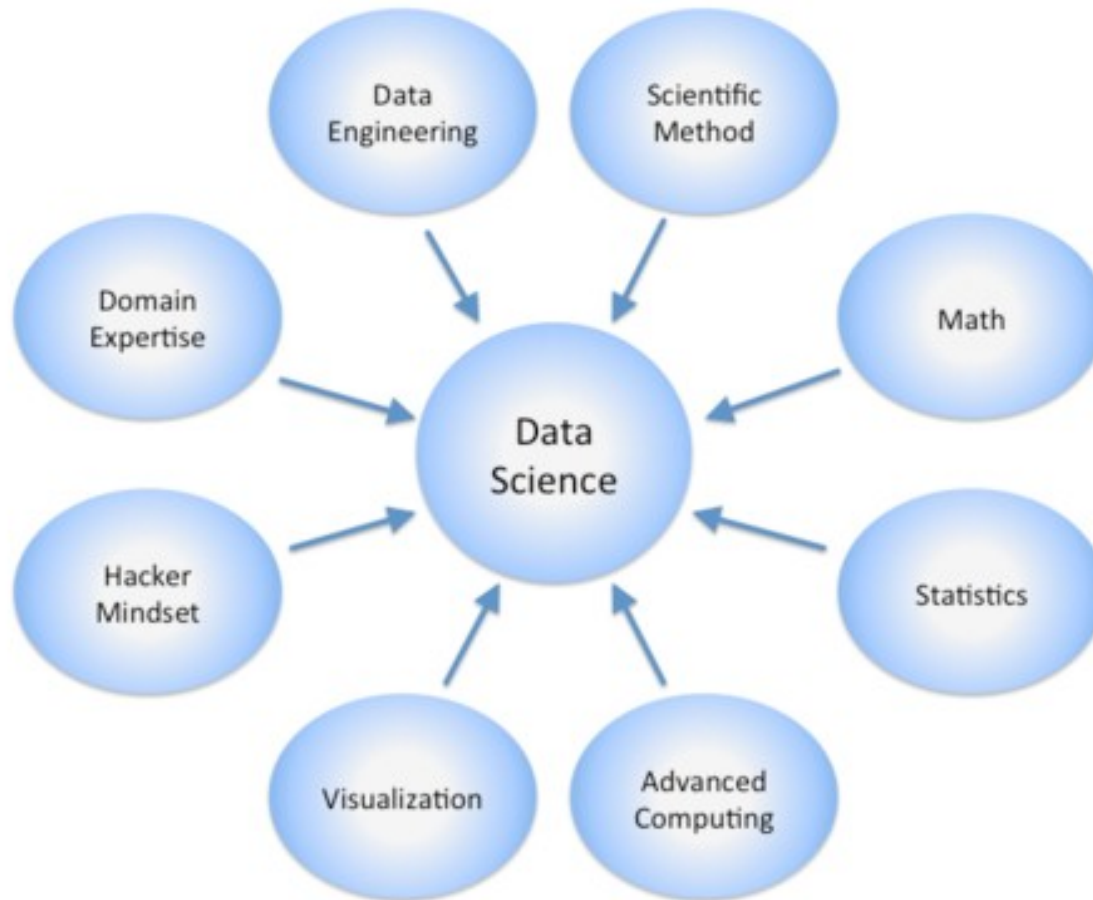


The Data Deluge

“The ability to generate vast quantities of data is outpacing the development of infrastructures and tools adequate to support data-driven 21st century science and technology.”

(US President’s Council of Advisors on Science and Technology, 2007)

The World of Data Science



Tools

Name	Advantages	Disadvantages	Open source?
R	Library support visualization	Steep learning curve	Yes
Matlab	Elegant matrix support visualization	Expensive; incomplete statistics support	No
SciPy/NumPy/ Matplotlib/Pandas	Python	Under heavy development	Yes
Excel	Easy; visual; flexible	Large datasets	No
SAS	Large datasets	Expensive; outdated programming language	No
Stata	Easy statistical analysis		No
SPSS	Like Stata but more expensive and worse		

Pandas

Scikit-Learn

NetworkX

Matplotlib

Scipy

Numpy

Python

<http://continuum.io/downloads>

github.com/panisson/LesHouches

<http://leshouches.panisson.com/>
Password: **13sh0uches**