

Global Big Data Conference

BIG DATA 3 DAYS ON MAY 09TH 10TH 11TH 2014 **BOOTCAMP**

Irving Convention Center, 500 W Las Colinas Blvd,
Irving, TX 75039

Python for Data Analytics and Big Data

Benjamin Zaitlen
Continuum Analytics

Global Big Data Conference

All Materials in Github

<https://github.com/quasiben/BDD14>

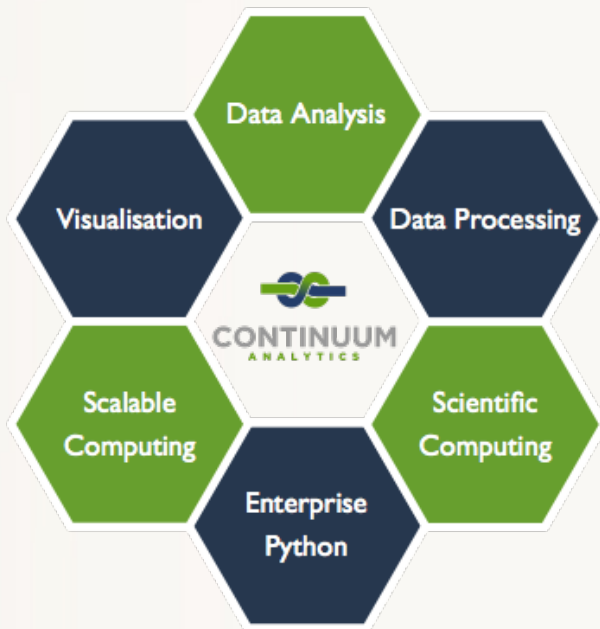
Global Big Data Conference

Agenda

- Environment Setup
- Python/IPython Overview
- NumPy
- Pandas
- MapReduce in Python
- Other useful libraries

Global Big Data Conference

About Continuum Analytics



Technologies

- Array computing
- Distributed computing & HPC
- GPUs, vectorized hardware
- Interactive visualization

Products:

- Anaconda Python
- Wakari Analytics Platform
- NumbaPro - GPU compiler
- IOPro - Fast I/O Library

Global Big Data Conference

About Continuum Analytics

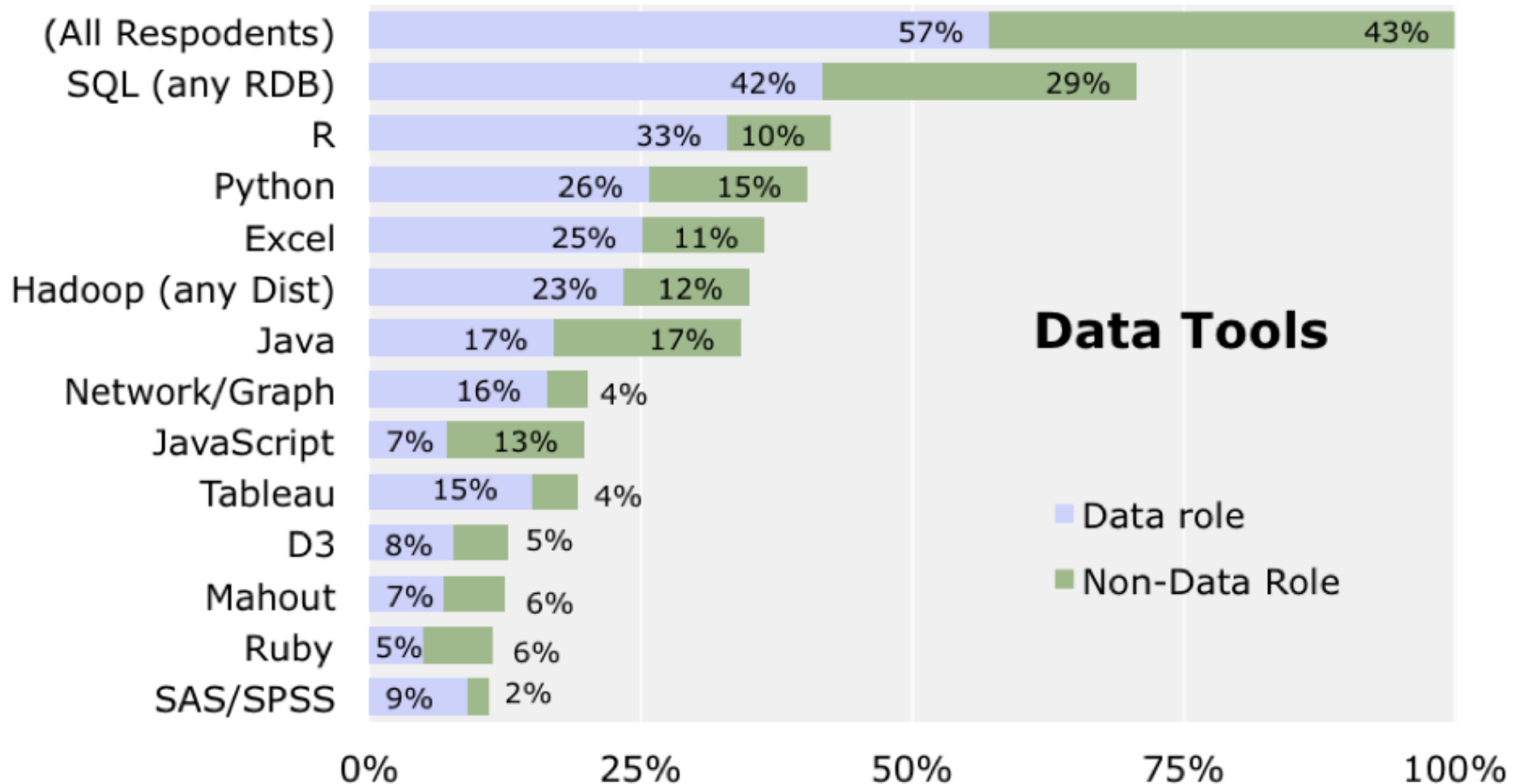
Domains

- Finance
- Defense & Government
- Geophysics
- Advertising/Marketing Analytics
- Scientific computing
- and more!

<http://continuum.io>



Global Big Data Conference



O'Reilly Strata "Data Science Salary Survey 2013"

Python: A glue language

- Scripting
- Web Apps
- RADs
- Scientific Computing
- Statistical Computing
- Big Data Computing

Global Big Data Conference

Agenda

- Python/IPython Overview
- NumPy
- Pandas
- MapReduce in Python
- Visualization along the way
- Other useful libraries

Intro to Python/IPython

Basic Language Concepts
Distributions and IDEs

NumPy

- Arrays: shape creation, core concepts
- Slicing, views, and multi-dimensional arrays
- Reshaping
- Array methods
- Data model

Pandas

- History and concepts
 - Series
 - DataFrame
- Aggregations & Plotting

MapReduce

- Overview
- Python Extensions
- Canonical Example
- Data Philosophy

Machine Learning

- Scikit-Learn
- Pylearn2
- Shogun
- WiseRF

Mapping and GIS

- Basemap: <http://matplotlib.org/basemap/>
- Shapely: <https://github.com/sgillies/shapely>
- Fiona - read/write data <https://github.com/sgillies/fiona>
- Python for ArcGIS: <http://resources.arcgis.com/en/communities/python/>
- GeoDjango - Build GIS web apps: <http://geodjango.org/>

Global Big Data Conference

More Resource for Learning

- Scipy Wiki — Topical Software: <http://scipy.org/topical-software.html>
- SciPy Lectures: <http://scipy-lectures.github.io/>
- PyData Videos: <http://vimeo.com/pydata>
- IPython Notebooks: <http://nbviewer.ipython.org/>

Follow me of Twitter: @quasiben

Email me: ben.zaitlen@continuum.io

Global Big Data Conference

Thank you to the Organizers

And big thanks to you!