Why Python for Data Analysis?

School of Information Studies
Syracuse University

Data Analysis

Nuts and Bolts of Data Analysis

- Manipulating
- Processing
- Cleaning
- Crunching

Types of Data

Multidimensional arrays

Tabular or spreadsheet data

Multiple tables related by key columns

Evenly or unevenly spaced time series

Why Python?

Easy to use

Most popular dynamic languages

Distinguished by active scientific computing community

Significant growth since 2000

Excellent choice for data-centric applications

Libraries Used

NumPy—used for scientific computing

Pandas—makes working with structured data very fast

Matplotlib—helps produce 2D data visualizations

Ipython—ties everything together

SciPy—collection of packages

- Scipy.integrate—numerical integration routines
- Scipy.linalg—linear algebra routines
- Scipy.optimize —function optimizers (minimizers)
- Scipy.stats—probability distributions

So Why Python?

General-purpose programming language

- Beautifully designed
- Intuitive
- Exceedingly powerful

Python Is…

Real

A general-purpose programming language

Used by Google and Dropbox in core applications