# Dealing with Data using Python

Aeint Thet Ngon

October 9, 2015

# Statistical and Data Mining Packages

- numpy
- pandas
- matplotlib
- scikitlearn

# Numpy

- Open source module
- A large library of high level mathematical functions
- Efficiently implemented multi-dimensional arrays
- Designed for scientific computation

# Pandas

- Data set merging and joining
- Tools for reading and writing data between in-memory data structures and different file formats
- Importing data sets
- Time series-functionality

# Matplotlib

- Built on Numpy
- Excel tools: utilities for exchanging data with Microsoft Excel
- Plotting made easier: Scatter graph, 2D/3D graphs, heat maps, geographical maps

# SciKitLearn

- Built on NumPy, SciPy, and matplotlib
- Tools for Data mining and data analysis: Regression, Clustering, Model Selection