Day 7

Jupyter, Pandas, (if enough time: Numpy, SciPy)

Today’s Agenda

- Jupyter – God of Thunder and Data Analysis
  - Summarizing analysis and Code in a Notebook
- Pandas – Not **bear**ly tables
  - Fast and efficient table handling
- NumPy – All things highly efficient
- SciPy – Scientific Python, need I say more?
  - All your Stats and Maths needs

What are Jupyter Notebooks?

Combined document and interactive Python console

Embed Python code with Text, Images, ...

Combine code from different languages

Easy to combine an analysis with explanations

What is Pandas?

Pandas is a port of R data frames

An easy & efficient way to handle tables

Very important package for data analysis

Noteworthy about Pandas

Performing operations by whole columns:

Highly efficient, simple to write

Looping over DataFrame rows or columns:
Inefficient, avoid if possible

Use built-in ways of doing things!

What is NumPy?

NumPy = **Num**erical **Py**thon

Highly optimized handling of linear algebra

Operations on Matrices, Vectors

When are we using NumPy?

Often runs unknown in background

Pandas DataFrames use Numpy

Doing lots of calculations?

Vectorize them with NumPy!

What is SciPy?

Collection of scientific functions

Tightly integrated with NumPy

Everything from image processing
to statistics, integration, optimization, ...

When will I use it?

Most common: Statistics (scipy.stats)

Stochastic functions

Descriptive statistics & hypothesis testing

And all your other needs