

Foundations 2

Christopher G. Wilson, Ph.D.

Loma Linda University
Dept. of Pediatrics and Center for Perinatal Biology

April 8, 2015

Outline

1. Review
2. IPython
3. NumPy (www.numpy.org)
4. SciPy (www.scipy.org)
5. Pandas
6. Data visualization tools in IPython

Review from last week

- ▶ CodeAcademy for Python
- ▶ Using the text editor *edit*
- ▶ Version control with *git*

Quiz Time!

Take your quiz at:

<http://goo.gl/forms/2wC0r9jgZP>

Why *IPython*

- ▶ *IPython* is an interactive data exploration and visualization shell that supports the inclusion of code, inline text, mathematical expressions, 2D/3D plotting, multimedia, and dynamic widgets.
- ▶ *IPython* is a suite of tools designed to cover scientific workflow from interactive data transformation and analysis to publication quality output.
- ▶ The *IPython* notebook uses a web browser as its display front-end and provides a rich interactive environment similar that seen in Mathematica.

Why *IPython* continued...

- ▶ *IPython* notebooks makes it possible to save analysis procedures and output—providing reproducible, curatable data analysis, and an easy way to share algorithms/methods.
- ▶ *IPython* supports parallel coding and distributed data analysis to take advantage of cloud/high-performance clusters.

NumPy provides a foundation for scientific computing

- ▶ *NumPy* is great for matrix math
- ▶ Provides numeric primitives like random number generation and Fourier transforms
- ▶ The *NumPy* tutorial is available here:
http://wiki.scipy.org/Tentative_NumPy_Tutorial

SciPy provides an ecosystem of useful analysis packages

- ▶ *SciPy* is built on top of *NumPy*
- ▶ Tools considered part of *SciPy* include: Matplotlib, IPython, Sympy, and pandas
- ▶ The tutorial is available at:
<http://docs.scipy.org/doc/scipy/reference/tutorial/index.html>

Pandas for data analysis

- ▶ Pandas is the *Python Data Analysis Library*
- ▶ The project website is: <http://pandas.pydata.org/>
- ▶ Pandas provides easy-to-use data structures and analysis tools that are key to handling much of the data we are interested in
- ▶ The 10 minute tutorial for Pandas is here:
<http://pandas.pydata.org/pandas-docs/stable/10min.html>

Plotting/visualization tools

- ▶ There are a few visualization modules for Python
- ▶ For 2D graphics the easiest to use is matplotlib (matplotlib.org)
- ▶ The beginner's guide is available at:
<http://matplotlib.org/users/beginner.html>
and includes a tutorial.
- ▶ For 3D plotting, *Mayavi* is a great tool
- ▶ Documentation can be found at:
<http://docs.enthought.com/mayavi/mayavi/application.html#using-the-mayavi-application>

For next week. . .

1. Work through the *IPython* tutorial
2. Try the listed *SciPy/NumPy* exercises
3. Work through the *matplotlib* tutorial
4. A helpful book:
<http://www.greenteapress.com/thinkpython/thinkpython.html>