Rich Gruss

Week 6: Python for Data Analysis, Ch. 3 and 4

Ch. 3

1. IPython is a Python development environment that provides an interactive shell, a web-ready notebook, and a fast parallel computing engine. It promotes an execute-explore model of development suited for Python projects.

2. Important IPython features include introspection with ?, and magic commands (%timeit, %run, %logstart, %debug).

3. Two principles of "Zen of Python": flat is better than nested. Keep code as modular and decoupled as possible to make testing and debugging easier. Overcome a fear of longer files.

Ch. 4

Numpy's central data structure is the ndarray, which provides vectorized arithmetic operations. The element-wise array functions are called Universal Functions (ufuncs) and include unary ufuncs (sqrt(), exp()), and binary ufuncs (maximum()). These vectorized operations are much faster than pure Python equivalents.

Ndarrays are created with np.array(list) or np.matrix(list). The type (accessible through the attribute 'dtype') is inferred from the data, but if you'd like to coerce a type, you can use this syntax: np.array(list, dtype=np.float64).

An extremely import point about slices: slices create a list of references to the original data, not a new list. So if you update any element in your slice, those same elements will be modified in your original array.