

Python for Data Analysis, Preface and Ch 1

The Preface and Chapter 1 of *Python for Data Analysis* make the case that Python has the most important feature of a successful programming language—an active community. Python has become the language of choice among scientists and engineers, and has an ever-growing library of tools for data analysis. Since the Preface doesn't contain much content, I'll discuss six important points from the first chapter. 1) Python can work as glue code for a large volume of legacy C programs for performing heavy mathematical operations. Because these programs have been debugged over years, and run quickly because C is fast, there is no point to porting them. Python provides a way to capture and combine the results of those programs to do other analysis. 2) As powerful as Python is for data analysis, it is not a good language for multithreaded applications, such as web applications, because the Global Interpreter Lock(GIL) enforces the rule that only a singly byte code can execute at a time. 3) The Numpy library provides an *ndarray* object for fast array processing. R is in fact a subset of this vectorization, combined with Pandas' data frame semantics. 4) Matplotlib provides powerful plotting capabilities. 5) IPython is a Mathematica-style notebook for quickly writing scripts and displaying the output for presentation. 6) Scipy is an extension of Numpy's math functionality, with tools for integration, optimization, and signal processing.