Chapter 2: Building Your Analytics Toolbox: A Primer on Using R & Python For Security Analysis

If you add a little to a little and do this often, soon the little will become great.

Hesiod

Before we jump right into the various use cases in the book it’s important to ensure you at least have a basic familiarity with the two most prominent languages featured in practically all of the scenarios: Python (<http://www.python.org/>) and R (http://www.r-project.org/). While there are an abundance of tools available for data analysis, we feel these two provide nearly all the features necessary to help you go from data to discovery with the least amount impedance.

A sub-theme throughout the book and the process at the heart of security data science is: **idea**, **exploration**, **trial** (and *error*) and **iteration**. It is ineffective at best to attempt to shoehorn this process into a traditional edit/compile/run workflow found in most traditional languages and development environments. The acts of performing data analysis tasks and creating informative visualizations are highly interactive endeavors. Despite all of their positive features, standalone Python and R do not truly enable rich, dynamic interaction with code and data; but, when coupled with IPython (<http://ipython.org/>) and RStudio (http://www.rstudio.com/), respectively, they are both transformed into powerful exploration tools, enabling rapid development and testing of everything from gnarly data transformations to generating sophisticated visualizations.

This chapter will provide pointers to installation resources for the tools, introduce core features of each language and development environment and explain the structure of the examples you will find in the remaining chapters of the book.