

## Sources of Open Source in Machine Learning



There are approximately 13,000 add-on packages for the programming language R. While discussing with friends and colleagues about a few useful and popular sources of open source I realized it's not easy to refer to a list that is short in description and handy for a quick reference. I also realized that the terms 'module', 'package' & 'library' which are often loosely used amongst machine learning practitioners are defined differently by different programming languages. To restrict our definition to Python, the most popular ML language, a module "is a .py file that defines one or more function/classes which you intend to reuse in different codes of program". In Python package refers to "a directory of Python module(s)". While a library is used "loosely to describe a collection of the core modules".

I'll attempt to bring forward a few Machine Learning 'sources' (avoiding 'module', 'package' & 'library') which are open source. I've attempted to refer to a credible download source from CRAN (Comprehensive R Archive Network) along with a brief description and an interesting tip! The table below is restricted to R sources only with the aim to add more sources. I've a plan to create a similar list for Python sources too.

Download Source	Short Description	Important Tips!
<a href="#">dplyr</a>	Useful for data manipulations for R users	Don't miss <a href="#">data-wrangling</a> cheat sheet
<a href="#">dplython</a>	The python version of dplyr	Functionality of Dplyr on top of Python's

		pandas
<a href="#">data.table</a>	Provides an enhanced version of data.frames	Useful <a href="#">Github</a> link
<a href="#">Hmisc</a>	For high-level graphics, advanced table making etc.	Useful <a href="#">documentation list</a>
<a href="#">reshape2</a>	To transform data between wide and long formats	Comparison with similar package <a href="#">tidyr</a>
<a href="#">stringi</a>	Useful for text sorting, comparing, extracting words	Provides platform-independent functions
<a href="#">stringr</a>	Contains functions to manipulate strings	Useful <a href="#">CheatSheets</a>
<a href="#">stringdist</a>	For string distance calculation & string matching	
<a href="#">ggplot2</a>	Popular data visualization package in R	Don't miss the <a href="#">cheatsheet</a>
<a href="#">rvest</a>	To scrape data from html, XML web pages	Scrape information about <a href="#">The Lego Movie</a>
<a href="#">tidyr</a>	To arrange column (variable) & observation (row)	Makes messy data usable by dplyr
<a href="#">rmarkdown</a> <a href="#">knitr</a>	File format for making dynamic documents with R R Integration with LaTeX, LyX, HTML, AsciiDoc etc.	Supports dozens of static / dynamic output format Dynamic report generation with R
<a href="#">shiny</a>	To build interactive graphs and web apps	Don't miss the existing <a href="#">gallery</a>
<a href="#">reticulate</a>	Provides interoperability between Python and R	Calling Python from R in variety of ways
<a href="#">r2d3</a>	Translate R objects into D3 friendly data structures	Useful <a href="#">pre-built examples</a>
<a href="#">DT</a>	R data objects can be displayed as tables on HTML	Provides filtering, pagination, sorting