

Using R Statistics

Byron A. Ellis – March 23, 2019



R does not limit users to preprogrammed procedures as in commercial statistical software.

R is a free open source software. It is downloadable from the [Comprehensive R Archive Network](#) (CRAN). R is a programming language and a software environment for data analysis and statistics (Cotton, 2013). Users can download R from mirror sites around the world. Cotton noted that R refers to two things, the programming language and the software that runs programs written in R.

Ross Ihaka and Robert Gentleman created R in the early 1990s from S language developed at Bell Laboratories in the 1970s (Cotton, 2013; Kleiber & Zeileis, 2008). Base R extends itself by using packages stored in libraries (Kleiber and Zeileis, 2008). Faraway (2002) indicated that the greatest disadvantage in R is that it is not easy to learn.

Below is the initial R screen.

The screenshot shows the RGui (64-bit) application window. The title bar reads 'RGui (64-bit)'. The menu bar includes 'File', 'Edit', 'View', 'Misc', 'Packages', 'Windows', and 'Help'. Below the menu bar is a toolbar with icons for file operations (open, save, print, etc.). The main window is titled 'R Console' and contains the following text:

```
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> |
```

Deepanshu Bhalla from [Listen Data](#), in 2016, listed companies using R; top tier companies on his list included Facebook, Google, Twitter, Microsoft, Uber, and so on. IT companies included Accenture, Amadeus IT Group, Capgemini, Cognizant, and more. Consulting companies included A.T. Kearney, Absolute Data, AC Nielsen, Bain and Company, IBM, and others. Financial institutions included American Express, Bank of America, Citibank, and so on. [Make Me Analyst](#) also has a more recent list of companies using R and indicated that the most popular language for data science analysis is R, followed by Python, SQL, and Excel.

[Revolutions](#), in 2017, noted that “CRAN, the global repository of open-source packages that extend the capabilities of R, reached a milestone today. There are now more than 10,000 R packages available for download.” It also confirmed that R is the most popular language for data scientists.

The basic unit in R is a vector, and functions in R operates on vectors (Kleiber & Zeileis, 2008). Users construct vectors by using the function `c ()`, `c` stands for combine or concatenate. So, we can form a vector `z` in the following fashion:

```
> z <- c(1, 2, 3, 4, 5)
```

The `<-` is known as an assignment operator, users can use an equal sign instead, but programmers prefer the assignment operator. Function `length ()` finds the length of the vector:

```
> length(z), it returns
```

```
[1] 5
```

Everything in R is an object (Kleiber & Zeileis, 2008). Objects are variables, vector, matrices, functions, and so on. The function `object ()` displays the names of the objects stored in the user’s workspace. All objects created in a session can be permanently stored for later retrieval. Vectors’ value must be of the same *mode*. Objects in R have a *class* that can be discovered by the function `class`.

Users use the command `library ()` to load packages, and packages in Base R can be seen by typing `library ()`; they can create histograms, bar plots, pie charts, boxplots, perform regression analyses, and so on. The code for a histogram is:

```
> s <- hist(variable, freq = FALSE)
> lines(density(variable))
```

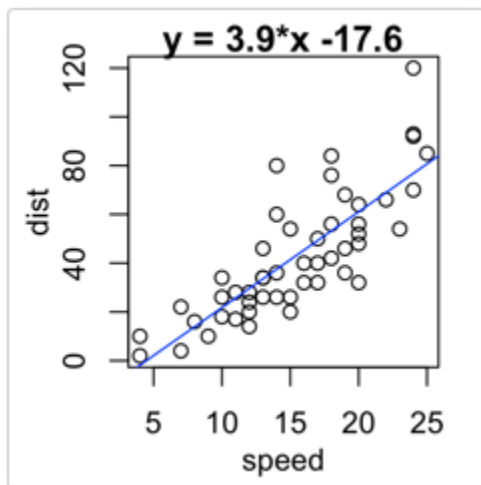
Users can also write their functions, as well as take advantage of useful packages for quality control charting. For instance, the `qcc` package provides plotting for Shewhart quality control and Cusum and EWMA charts, draws operating characteristic curves, performs process capability analyses, draws Pareto charts, and cause-and-effect diagrams (Scrucca, 2004). R also has Six Sigma packages; McDonough (2011) noted that the Six Sigma project involves a significant amount of statistical analysis and that Black Belts usually use Minitab or JMP to perform these analyses. He said that commercial software is limited in capability and quite expensive. Thus, the importance of due diligence when acquiring statistical packages is essential.

Many researchers use R for regression analyses, including econometricians. For instance, given a model

$$y = X\beta + \epsilon$$

where y is the dependent variable, and the X matrix contains the independent variable(s). R uses the function `lm ()` for finding the regression coefficient(s), residuals, and so on. User would use

`> s_lm = lm (y ~ x)` to find the coefficients, standard errors, t values, probabilities, R-Square and F values of the regression. `s_lm` stores these values, and the function `summary ()` or `summary (s_lm)` makes them available. Users use the functions `plot ()` and `abline ()` to plot the graph and draw the regression line, `plot (y ~ x)`, and `abline (s_lm)`, respectively.



[From STHDA](#)

Organizations that use and pay for multiple licenses of preprogrammed commercial software should consider the cost savings and flexibility of the free R open-source software versus its learning slope.

References

- Cotton, R. (2013). Learning R. O'Reilly: Sebastopol, CA.
- Faraway, J. J. (2002). Practical Regression and Anova using R.
- Kleiber, C. & Zeileis, A. (2008). Applied Econometrics with R. Springer: New York, NY.
- McDonough, J. (2011). Six Sigma with R – A Tutorial. *The Fedora Project*.
- Scrucca, L. (2004). qcc: An R Package for Quality Control Charting and Statistical Process Control. *R News*, Vol. 4/1, p. 11-20.

Copyright of TJP is the property of The Jethro Project, and its contents may not be copied or emailed to multiple sites or posted to a list server without the copyright holder's express written permission. Users, however, may print, download, or email articles for individual use.