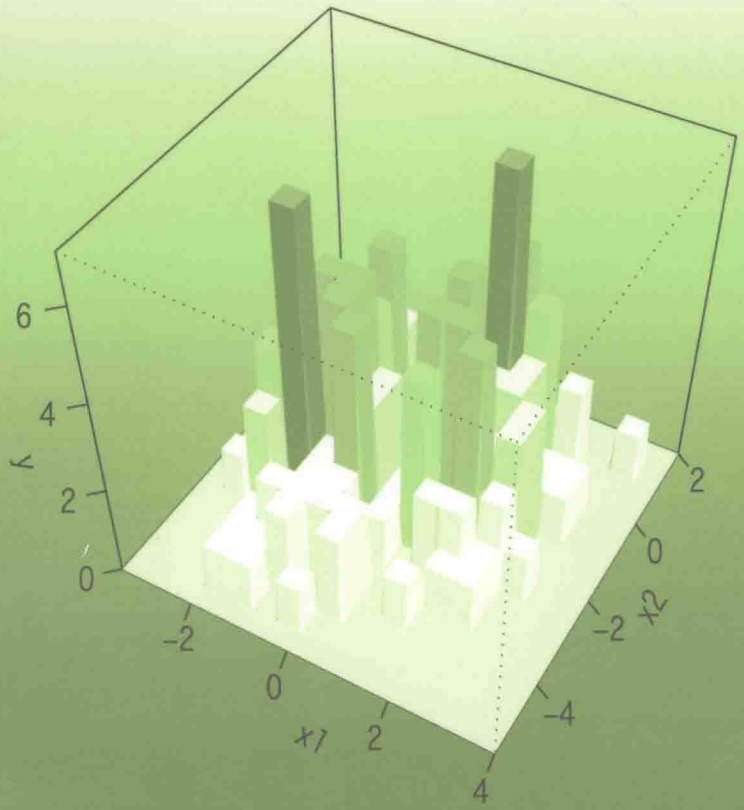
Guidebook to R Graphics Using Microsoft® Windows



Kunio Takezawa

WILEY

Guidebook to R Graphics Using Microsoft® Windows

Kunio Takezawa

National Agricultural Research Center,
National Agriculture and Food Research Organization
Graduate School of Life and Environmental Sciences
University of Tsukuba, Ibaraki
Tsukuba, Japan

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Takezawa, Kunio, 1959–
Guidebook to R graphics using Microsoft Windows / Kunio Takezawa.
p. cm.
Includes index.
ISBN 978-1-118-02639-7 (pbk.)
1. Computer graphics. 2. R (Computer program language) 3. Microsoft Windows (Computer file)
I. Title.
T385.T346 2012
006.6'633—dc23 2011049806

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

PREFACE

> Carol Marcus: Let me show you something that will make you feel young as when the world was new. (Star Trek: The Wrath of Khan (1982) http://www.imdb.com/title/tt0084726/quotes)

Construction of appropriate graphs plays an important role in data analysis. Pertinent graphs often reveal the conditional implications of data clearly even if the summarization of data by deriving a small number of values can show limited aspects of the characteristics of data. In addition, persuasive graphs are an indispensable tool for the presentation of scientific papers, the description of commodities, patent applications, project proposals, lectures, training courses, business meetings, negotiations, legal actions, etc. The graphical presentation of features of data and structures of concepts can strengthen arguments.

Therefore, it is well known that producing high-quality graphs is a requisite in diverse fields. Demand for richly expressive graphs has grown. Moreover, people's need for better graphs at a lower cost has increased markedly because the extraordinary development of computer technology has enabled the drawing of complicated graphs in a short period of time. Therefore, drawing graphs using a PC is no longer a specialist skill that requires professional expertise. It should be a fairly commonplace technique, comparable to the creation of simple documents using a word processor.

The free software "R" is now widely used for statistical calculation and graphics. R is equipped with various functions for constructing graphs. Moreover, installing software packages enables a wider range of graphs to be produced. The importance of R as a tool for graphics has increased. However, those unfamiliar with the use of R for statistical calculations tend to avoid the software because its wide range of functions may make it appear daunting.

In light of this situation, this book aims to demonstrate that producing graphs using R is an easy-to-master technique. Hence, this book often does not describe R commands and their range of arguments exhaustively, but rather exemplifies typical methods for constructing graphs and their results. This makes the book a guide that aims to foster a feeling of confidence that most graphs can be produced using R. As is generally the case with learning word processing software, learners should obtain an overall picture of the graphical abilities of R in a short time by gaining familiarity with its main features rather than by acquiring details of each function; then they can realize a greater diversity of graphics by referring to references and articles on the Internet. This way of learning is effective provided learners are not aspiring to careers as specialists in the software.

To achieve the aims of this book, procedures for excuting R are presented as R programs that contain a series of R commands. That is, this book, unlike other books in this field, does not proceed with procedures by inputting R commands sequentially. Batch processing is focused on rather than sequential processing (real-time processing). Sequential processing has the advantage of being able to construct graphics while seeing the graphs in progress. However, batch processing is clearly superior to sequential processing because inappropriate manipulation and keystroke errors often occur during the work. In batch processing, R programs (a series of R commands) are recorded regularly, making it easy to rerun a procedure after modifying the program. In addition, learners can review R programs that they have learned, and use or develop them after accumulating techniques in the form of R programs.

Fortunately, if learners transform their R programs into text files, the programs can be easily found when needed by full-text searching. For example, we assume that a folder called `D:\GraphicsR` contains `.RData` (a work image file; this file stores R programs and data). When R programs are accumulated in `.RData`, it will be useful for learners to search for R programs that they need among the programs in `.RData`. For this purpose, learners should activate `.RData` and run the R program below, for instance.

```
function()
{
  ob1 <- objects(pos=1)
  nd <- length(ob1)
  for(ii in 1:nd) {
    print(paste("D:\\GraphicsR\\", ob1[ii], ".txt", sep = ""))
    dump(ob1[ii], paste("D:\\GraphicsR\\", ob1[ii], ".txt",
```

```
      sep = ""))
  }
}
```

This procedure transforms all objects (R programs and data files) in `.RData` into text files and outputs these text files in `D:\GraphicsR`. Extensions of these text files will be `.txt`. If a full-text search of these text files is carried out with R commands as key words, the R programs that use these R commands will be found. Additionally, if comments such as "draw a circle" or "change a font" are written in R programs, a full-text search with the key words of "circle" or "font" will find R programs that contain the above functions.

Moreover, this book explains the method of producing GUI (interactive) programs. Although many people need to construct graphs or carry out statistical calculations, some would rather not go to the trouble of inputting R commands sequentially based on a process of trial and error while considering the function of each R command, and they would have little ability to cope with R programs consisting of several tens of lines. To allow such people to benefit from the useful functions of R, interactive R programs should be prepared to create an environment in which R can be used free from the consciousness of its presence. This is a goal of this sort of program. In addition, if interactive R programs are used in lectures on statistical analysis or as exercise materials for students taking such a course, the contents of statistical analysis can be conceived as a smooth flow of concepts and statistical analysis will attract a high level of interest. The author hopes that a great number of readers will appreciate this intention to construct R programs along this line to increase the familiarity with R.

For these purposes, the R programs listed in this book are elementary ones that allows readers to understand each function of R. New R programs can be developed by adapting these R programs according to readers' needs and interests. By transforming them into text files by the method described above, these text files will guide readers in producing graphs using R. This book provides a platform for readers' originality and ingenuity in this manner; it is not an introductory book that must be read in page order. Such constructive usage of this book is expected to expand the possibilities for creating graphics using R programs, regardless of whether this book is read alone, used as a reference, or used as a textbook in a course on statistical analysis.

The FTP site for this book can be found at the following URL:
`ftp://ftp.wiley.com/public/sci_tech_med/guidebook_r-graphics`.

The R programs listed in this book assume that the size of the graphics window on the display of your PC is roughly 16 cm (roughly 6.3 inch) $\times$ 16 cm (please refer to the figure below). If a larger graphics window than this is needed, or the graphics window cannot be as large as this because of the limited size of your display, the R programs should be modified.

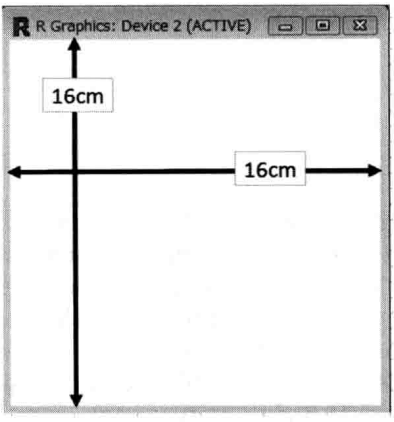


Figure 0.1 Graphics window

Warning: When the R programs listed in this book are executed or modified by trainees in a lecture or a workshop, the copyright holder has made it a condition that all trainees have purchased this book.

KUNIO TAKEZAWA

Tsukuba, Japan
March, 2012

ACKNOWLEDGMENTS

The original edition of this book was published as "Graphics and GUI Operations Using R" ("R niyoru gazouhyougen to GUI sousa"), written in Japanese. The publisher is CUTT System Development Laboratory Inc. (Hyakunin-cho 4-9-7 8th floor, Shinjuku, Tokyo 169-0073, Japan. President: Mr. Katsutoshi Ishizuka, Editor: Mr. Tomohiro Takei). CUTT System Development Laboratory Inc. kindly agreed to our request to publish an English edition of this book. This English edition was not just a faithful translation, but underwent processes of additions and alterations on the advice of the editors at Wiley.

To all these wonderful people I owe a deep sense of gratitude, especially now that this project has been completed.

K.T.

CONTENTS

CHAPTER 1

BASIC GRAPHICS

1.1 INTRODUCTION

This chapter first describes the procedures to start up R, produce R programs, and run them. This is followed by explanations of R programs used to construct simple graphs. New techniques are introduced by adding new methods to already-known materials. Hence, learning is followed by consulting, testing, and modifying the listed R programs sequentially. Alternatively, learners can find graphs that roughly suit their purpose and experiment with them. Previous articles may be referred to if unknown commands or functions are used in the programs. In addition, the last part of this chapter introduces techniques to share displayed graphs with other application software packages and to save graphs as digital files.

1.2 DOWNLOADING AND INSTALLATION OF R

The procedure below installs R on a PC loaded with a Windows OS.

Guidebook to R Graphics Using Microsoft Windows, First Edition. By Kunio Takezawa

1. Access "the R Project for Statistical Computing" web page (`http://www.r-project.org/`).
2. Click "CRAN" under "Download, Packages" listed in the menu on the left side.
3. Choose one of the mirror sites on the "CRAN Mirrors" page. Most of the mirror sites are identical. For example, choose University of Tsukuba (`http://cran.md.tsukuba.ac.jp/`).
4. Click "Windows" under "Download and Install R" on "The Comprehensive R Archive Network" web page.
5. Click "base" on the same line as "base Binaries for base distribution (managed by Duncan Murdoch)" on the "R for Windows" page.
6. Click "Download R 2.11.0 for Windows (32 megabytes)" (the version may be different) on the "R-2.11.0 for Windows" page.
7. R-2.11.0-win32.exe (or a different version) is now available for installation. Construct a folder in the hard disk in your PC and download the file to the folder.
8. Double-click R-2.11.0-win32.exe to start the installation process of R.
9. Agree to the "GNU General public license" and specify the location of installation.
10. Select "Full installation" on the "Select components" page.
11. After a few further selections, installation starts.

Sets of programs called "packages" have been prepared for R. Packages can be added to the version of R installed on your PC. If a PC is connected to the Internet, select "Packages" in a menu after R is booted. Then, choose "Set CRAN mirror...". Many mirror sites appear. Then, choose "Japan (Tsukuba)", for example, and click "OK". Return to "Packages" in the menu. Then, select "Load package" to display the names of many packages. Choose the names of packages that you need and click "OK". The selected packages are installed in this way.

If packages are installed in this manner, the names of packages displayed are those that fit the version of R installed on a PC. However, some packages that fit an older version of R can be used in a newer version of R. For example, the "gtools" package that will be dealt with later fits version 2.10 of R but does not fit version 2.11 of R. Hence, when the "gtools" package is used, it is safe to employ version 2.10 of R. However, the use of the "gtools" package with version 2.11 of R can be attempted. For this purpose, the method described above cannot be used. A method for a PC that is not connected to the

Internet is useful for installing packages for older versions of R. This method is as follows:

1. In the fifth step of the installation procedure of R described above, click "contrib" on the same line of "contrib Binaries of contributed packages (managed by Uwe Ligges)".
2. The "Index of /bin/windows/contrib" page appears. Select a version of R that fits the packages that you will use (for example, 2.10).
3. The names of the files of packages compressed in zip format are displayed. Click the names of the files of packages needed for downloading them.
4. R is booted. Select "Packages" in the menu. Then, select "Install package(s) from local zip files...".
5. Select the files that were downloaded beforehand to install packages.

Using this method, even if R version 2.11 is installed on a PC, packages for R version 2.10 may be used.

1.3 START-UP OF R, AND CONSTRUCTION AND EXECUTION OF R PROGRAMS

We assume that the folder used for storing data files and the results of calculations is D:\GraphicsR. Save all files created with R in this folder.

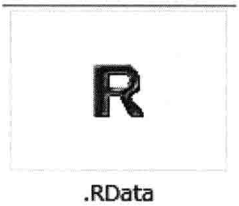


Figure 1.1 Work image file.

The file for storing R programs is called the "work image file" (Fig. 1.1). Confirm the presence of this file in D:\GraphicsR. If the work image file is not located in D:\GraphicsR, search for a file named ".RData" (which will be somewhere in your hard disk if R has been installed correctly) and copy the file by saving it in D:\GraphicsR. If the drive where .RData is originally located is the "D" drive, hold the Ctrl key and copy the file. If the Ctrl key is not held, the file is not copied but moved. That is, the original .RData is deleted. When multiple .RData files are placed in the same PC, each .RData

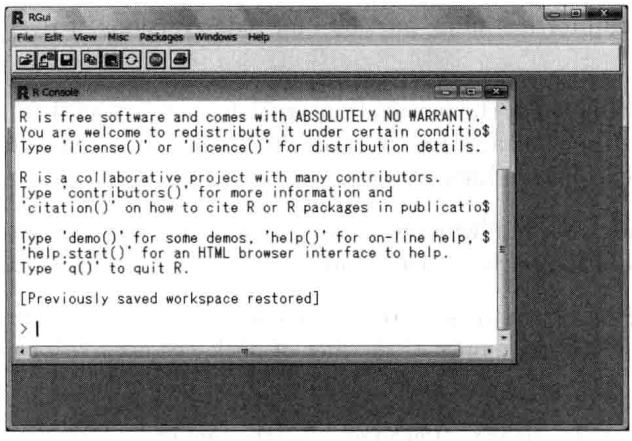


Figure 1.2 Window immediately after "R" is booted.

file stores its own R programs. Even if the version of R is upgraded, the same .RData file can be used.

R is booted by double-clicking .RData in D:\GraphicsR. R can also be booted by double-clicking the shortcut button assigned to .RData on the desktop or other places. Upon booting R, the window shown in Fig. 1.2 appears.

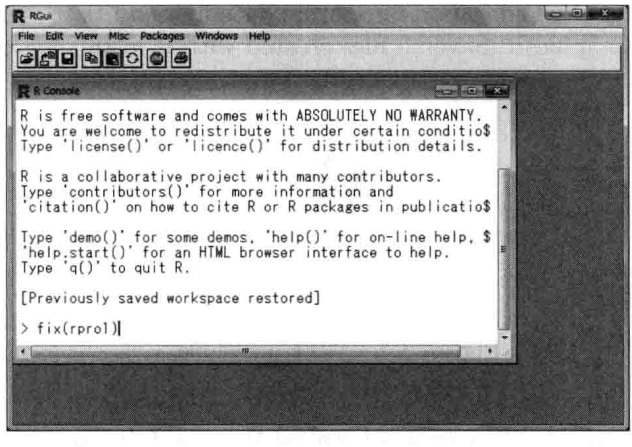


Figure 1.3 Construction of R program.

The inner window in Fig. 1.2 is called the console window. In this window, R commands and R programs are executed. For example, when an R program named `rpro1()` is produced, type `fix(rpro1)` and click the return key in the console window (Fig. 1.3). Then, the display in Fig. 1.4 appears. The new

window is an editor. When R is installed in the standard manner, "notepad" is used as an editor. However, users can set up another editor for this purpose. An R program is a series of R commands listed between "`function () {`" and "`}`". Arguments can be specified in "`()`" of "`function () {`". For example, `function (aa) {` is an R program that uses `aa` as an argument. By setting numerical values or text as `aa`, an R program with this argument is run.

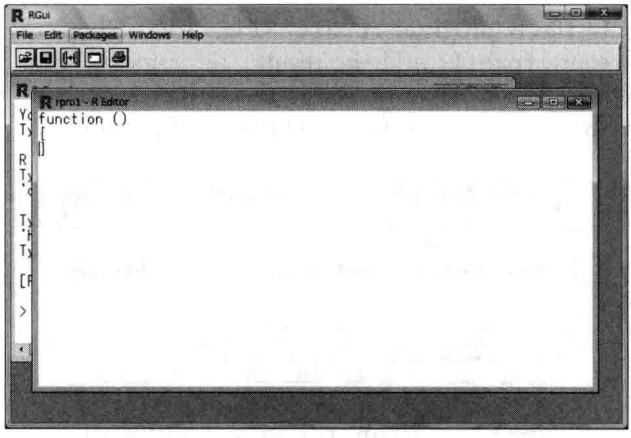


Figure 1.4 Editor.

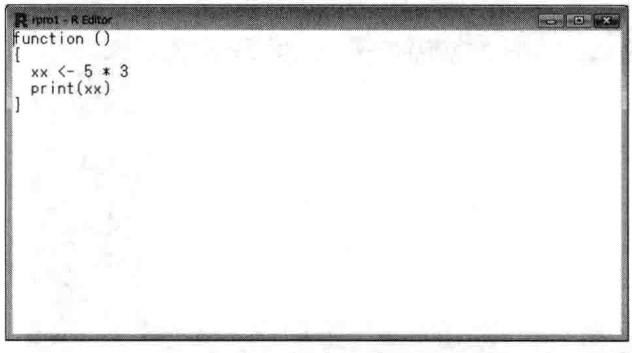


Figure 1.5 R program.

Let us produce an R program for multiplying 5 and 3 and displaying the result in the console window. Fig. 1.5 shows an example of an R program for this purpose. The editor is closed to execute this R program. For this, "×" (Fig. 1.6) located of the upper right of the editor is clicked. Then, a dialogue box asking for a selection is displayed (Fig. 1.7). "Yes" or the return key is clicked. Only the console window is then displayed (Fig. 1.8). `rpro1()` is

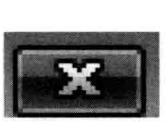


Figure 1.6 Symbol clicked to close.

typed to execute the R program `rpro1()` (Fig. 1.9). Then, the return key is clicked to execute `rpro1()` and the result "15" appears (Fig. 1.10). This series of procedures summarizes the basic use of R: the start-up of "R", the construction of R programs, and the execution of R programs.

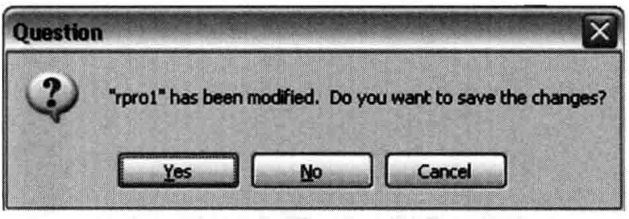


Figure 1.7 Dialogue box asking for selection.

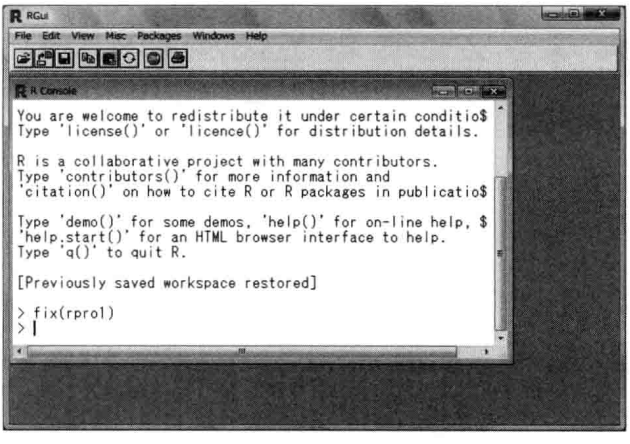


Figure 1.8 Return to the console window.

When graphics windows are displayed in the console window (Fig. 1.11), the execution of `graphics.off()` in the console window (Fig. 1.12) clears all graphics windows, and the console window remains (Fig. 1.13). If R is shut down and rebooted, only the console window appears. Even if only the