R WIthout Statistics

R Without Statistics

Visualize, Communicate, and Automate with R

by David Keyes

No Starch Press

R WITHOUT STATISTICS. Copyright © 2024 by David Keyes.

All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner and the publisher.

Printed in the United States of America

First printing

28 27 26 25 24     1 2 3 4 5

ISBN-13: 978-1-7185-0332-8 (print)  
ISBN-13: 978-1-7185-0333-5 (ebook)

COMP: insert nsp logo to the left of the address block; ® is 10pts

Published by No Starch Press®, Inc.  
245 8th Street, San Francisco, CA 94103  
phone: +1.415.863.9900  
www.nostarch.com; info@nostarch.com

Publisher: William Pollock  
Managing Editor: Jill Franklin  
Production Manager: Sabrina Plomitallo-González  
Production Editor: Sydney Cromwell  
Developmental Editor: Frances Saux  
Cover Illustrator: ???  
Interior Design: Octopod Studios  
Technical Reviewer: Rita Giordano  
Copyeditor: Rachel Monaghan  
Proofreader: ???  
Indexer: ???

Library of Congress Cataloging-in-Publication Data

<CIP DATA GOES HERE.>

For customer service inquiries, please contact info@nostarch.com. For information on distribution, bulk sales, corporate sales, or translations: sales@nostarch.com. For permission to translate this work: rights@nostarch.com. To report counterfeit copies or piracy: counterfeit@nostarch.com.

No Starch Press and the No Starch Press logo are registered trademarks of No Starch Press, Inc. Other product and company names mentioned herein may be the trademarks of their respective owners. Rather than use a trademark symbol with every occurrence of a trademarked name, we are using the names only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

[X]

COMP: dedication.

For my wife, Rachel, and my children, Leila and Elias.

About the Author

David Keyes is the founder of R for the Rest of Us (https://rfortherestofus.com), where he develops courses, conducts corporate trainings, and works with organizations to harness the power of R. While leading a team of consultants, he has overseen the creation of many R-based reports. As a self-taught R user with a qualitative background, he helps people who don’t think of themselves as R users learn to use this powerful tool.

About the Technical Reviewer

Rita Giordano is an independent data visualization consultant and LinkedIn instructor based in the UK. She is a physicist with a PhD in statistics applied to structural biology and has extensive research and data science experience.

Brief Contents

COMP: Run in 1st pass

Contents in Detail

COMP: Run in 1st pass

ACKNOWLEDGMENTS

This book is a testament to the many members of the R community who share their knowledge freely and encourage others generously. I call myself self-taught, but really what I am is community-taught. Throughout this book, you will read about several R users from whom I have learned so much; still, many others go unmentioned. To everyone who has worked to develop R, share your knowledge about R, and make R a welcoming place, thank you.

I’d also like to thank the team at R for the Rest of Us. Working directly with talented R users has taught me so much about what is possible with R.

Finally, I’d like to thank people who have provided feedback as I’ve written this book. Technical reviewer Rita Giordano has helped me make sure everything works and suggested great ideas for improvement. My editor, Frances Saux, has provided fantastic input along the way. To Bill Pollock and the entire team at No Starch: thank you for taking a flyer on me and my strange idea to write a book about nonstatistical uses of a tool created for statistics.

Introduction

In early 2020, as the world struggled to contain the spread of COVID-19, one country succeeded where others did not: New Zealand. There are many reasons New Zealand was able to tackle COVID-19. One was the R programming language (yes, really).

This humble tool for data analysis helped New Zealand fight COVID-19 by enabling a Ministry of Health team to generate daily reports on cases throughout New Zealand. Based on the information in these reports, officials were able to develop policies that kept the country largely free of COVID-19. The team was small, however, so producing the reports every day with a tool like Excel wouldn’t have been feasible. As team leader Chris Knox told me, “Trying to do what we did in a point-and-click environment is not possible.”

Instead, a few staff members wrote R code that they could run every day to produce updated reports. These reports did not involve any complicated statistics; they were literally counts of COVID-19 cases. Their value came from everything else that R can do: data analysis and visualization, report creation, and workflow automation.

This book explores the many ways that people use R to communicate and automate tasks. You’ll learn how to do the following:

* Make professional-quality data visualizations, maps, and tables.
* Replace a clunky multi-tool workflow to create reports with R Markdown.
* Use parameterized reporting to generate multiple reports at once.
* Produce slideshow presentations and websites using R Markdown.
* Automate the process of importing online data from Google Sheets and the US Census Bureau.
* Create your own functions to automate tasks you do repeatedly.
* Bundle your functions into a package that you can share with others.

Best of all, you’ll do all of this without performing any statistical analysis more complex than calculating averages.

Isn’t R Just for Statistical Analysis?

Many people think of R as simply a tool for hardcore statistical analysis, but it can do much more than manipulate numerical values. After all, every R user must illuminate their findings and communicate their results somehow, whether that’s via data visualizations, reports, websites, or presentations. Also, the more you use R, the more you’ll find yourself wanting to automate tasks you currently do manually.

As a qualitatively trained anthropologist without a quantitative background, I used to feel ashamed about using R for my visualization and communication tasks. But the fact is, R is good at these jobs. The ggplot2 package is the tool of choice for many top information designers. Users around the world have taken advantage of R’s ability to automate reporting to make their work more efficient. Rather than simply replacing other tools, R can perform tasks that you’re probably already doing, like generating reports and tables, better than your existing workflow.

Who This Book Is For

No matter your background, using R can transform your work. This book is for you if you’re either a current R user keen to explore its uses for visualization and communication or a non-R user wondering if R is right for you. I’ve written R Without Statistics so that it should make sense whether or not you’ve ever written a line of R code. But even if you’ve written entire R programs, the book should help you learn plenty of new techniques to up your game.

R is a great tool for anyone who works with data. Maybe you’re a researcher looking for a new way to share your results. Perhaps you’re a journalist looking to analyze public data more efficiently. Or maybe you’re a data analyst tired of working in expensive, proprietary tools. If you have to work with data, you will get value from R.

About This Book

Each chapter focuses on one use of the R language and includes examples of real R projects that employ the techniques covered. I’ll dive into the project code, breaking the programs down to help you understand how they work, and suggest ways of going beyond the example. The book has three parts:

In Part I, you’ll learn how to use R to visualize data.

Chapter 1: An R Programming Crash Course

Introduces the RStudio programming environment and the foundational R syntax you’ll need to understand the rest of the book.

Chapter 2: Principles of Data Visualization

Breaks down a visualization created for Scientific American on drought conditions in the United States. In doing so, this chapter introduces the ggplot2 package for data visualization and addresses important principles that can help you make high-quality graphics.

Chapter 3: Custom Data Visualization Themes

Describes how journalists at the BBC made a custom theme for the ggplot2 data visualization package. As the chapter walks you through the package they created, you’ll learn how to make your own theme.

Chapter 4: Maps and Geospatial Data

Explores the process of making maps in R using simple features data. You’ll learn how to write map-making code, find geospatial data, choose appropriate projections, and apply data visualization principles to make your map appealing.

Chapter 5: Effective Tables

Shows you how to use the gt package to make high-quality tables in R. With guidance from R table connoisseur Tom Mock, you’ll learn the design principles to present your table data effectively.

Part II focuses on using R Markdown to communicate efficiently. You’ll learn how to incorporate visualizations like the ones discussed in Part I into reports, slideshow presentations, and static websites generated entirely using R code.

Chapter 6: R Markdown Reports

Introduces R Markdown, a tool that allows you to generate a professional report in R. This chapter covers the structure of an R Markdown document, shows you how to use inline code to automatically update your report’s text when data values change, and discusses the tool’s many export options.

Chapter 7: Parameterized Reporting

Covers one of the advantages of using R Markdown: the ability to produce multiple reports at the same time using a technique called parameterized reporting. You’ll see how staff members at the Urban Institute used R to generate fiscal briefs for all 50 US states. In the process, you’ll learn how parameterized reporting works and how you can use it.

Chapter 8: Slideshow Presentations

Explains how to use R Markdown to make slides with the xaringan package. You’ll learn how to make your own presentations, adjust your content to fit on a slide, and add effects to your slideshow.

Chapter 9: Websites

Shows you how to create your own website with R Markdown and the distill package. By examining a website about COVID-19 rates in Westchester County, New York, you’ll see how to create pages on your site, add interactivity through R packages, and deploy your website in multiple ways.

Chapter 10: Quarto

Explains how to use Quarto, the next-generation version of R Markdown. You’ll learn how to use Quarto for all of the projects you previously used R Markdown for (reports, parameterized reporting, slideshow presentations, and websites).

Part III focuses on ways you can use R to automate your work and share it with others.

Chapter 11: Automatically Accessing Online Data

Explores two R packages that let you automatically import data from the internet: googlesheets4 for working with Google Sheets and tidycensus for working with US Census Bureau data. You’ll learn how the packages work and how to use them to automate the process of accessing data.

Chapter 12: Creating Functions and Packages

Shows you how to create your own functions and packages and share them with others, which is one of R’s major benefits. Bundling your custom functions into a package for other users to import can streamline their work, as with the R developers you’ll read about who built packages for researchers working at the Moffitt Cancer Center.

By the end of this book, you should be able to use R for a wide range of nonstatistical tasks. You’ll know how to effectively visualize data and communicate your findings using maps and tables. You’ll be able to integrate your results into reports using R Markdown, as well as efficiently generate slideshow presentations and websites. And you’ll understand how to automate many tedious tasks using packages others have built or ones you develop yourself. Let’s dive in!