R WIthout Statistics

R Without Statistics

Book Subtitle

by David Keyes

No Starch Press

BOOK TITLE. Copyright © 2024 by Author Author.

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

27 26 25 24 23     1 2 3 4 5

ISBN-13: 978-1-xxxx-xxxx-x (print)  
ISBN-13: 978-1-xxxx-xxxx-x (ebook)

PROD: hyphenation of ISBN should be 978-1-59327-xxx-x or 978-1-7185-xxxx-x depending on which publisher code is used

Publisher: William Pollock  
Managing Editor: Jill Franklin  
Production Manager: Sabrina Plomitallo-González  
Production Editor: ???  
Developmental Editor: ???  
Cover Illustrator: ???  
Interior Design: Octopod Studios  
Technical Reviewer: ???  
Copyeditor: ???  
Compositor: ???  
Proofreader: ???  
Indexer: ???

PROD: Compositor and Interior Design should be checked and potentially changed if it is not a yellow and black series book

The following images are reproduced with permission: Figure x-x from example.com; Figure x-x courtesy of Organization; Figure x-x photo by Photographer Name

PROD: delete image credits or change as needed

For information on distribution, bulk sales, corporate sales, or translations, please contact No Starch Press, Inc. directly at info@nostarch.com or:

No Starch Press, Inc.  
245 8th Street, San Francisco, CA 94103  
phone: 1.415.863.9900  
www.nostarch.com

Library of Congress Cataloging-in-Publication Data

<CIP DATA GOES HERE.>

PROD: hide or remove summary from cip block before copying here

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.

PROD: only include this lego paragraph in lego books (and leave out parts, for example mindstorms, if not relevant)

LEGO®, MINDSTORMS®, the brick configuration, and the minifigure are trademarks of the LEGO Group, which does not sponsor, authorize, or endorse this book.

The information in this book is distributed on an “As Is” basis, without warranty. While every precaution has been taken in the preparation of this work, neither the author(s) nor No Starch Press, Inc. shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in it.

PROD: only include one of these fiction paragraphs if it is a work of fiction (and choose which one based on the use of real figures).

All characters in this publication are fictitious or are used fictitiously.

All characters in this publication are fictitious, and any resemblance to real persons, living or dead, is purely coincidental.

PROD: Add printer initial in brackets

[X]

Start 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, where he develops courses, conducts corporate trainings, and works with organizations to harness the power of R. 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

The TR is also cool.

Brief Contents

TK after correction pass.

Contents in Detail

TK after correction pass.

A frontmatter could have a foreword and/or a Preface.

**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, so many others go unmentioned. To everyone who has worked to develop R, share your knowledge about R, and make R a welcoming place for everyone, 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 provided great ideas for improvement. My editor, Frances Saux, has provided fantastic feedback 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 non-statistical 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 of these was the R programming language (yes, really).

How did a humble tool for data analysis help New Zealand fight COVID-19? It allowed a team at the Ministry of Health to generate daily reports on cases throughout New Zealand. These reports enabled officials to develop policies that kept the country largely COVID-19 free. The team was small, however, so producing these 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 re-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 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 activities like the following:

* Make professional-quality data visualizations, maps, and tables.
* Replace a clunky multi-tool workflow for creating 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 a Tool for Statistical Analysis?

Many people think of R as simply a tool for hardcore statistical analysis. But, over a quarter of a century since its creation, R can do much more than manipulate numerical values. After all, every R user must illuminate their findings and communicate their results somehow, whether via data visualizations, reports, websites, or presentations. Also, the more you use R, the more you’ll find yourself wanting to automate tasks you used to 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 R is good at these things. 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 replace other tools, R can allow you to do things, like generate reports and tables, that you’re already probably doing, and it can do it 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 are either a current R user keen to explore uses of R 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 even if you’ve never written a line of R code. But if you have written many lines of R code, the book should help you learn plenty of new techniques to up your R 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 we cover. We’ll dive into their code, breaking the programs down to help you understand how they works, and suggest ways of going beyond the example. The book has three parts:

Part 1:Visualizations

In the first part, you’ll learn about ways to use R to visualize data.

Chapter 1: An R Programming Crash Course

Introduces you to the R Studio 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, it introduces you to the ggplot2 package for data visualization and addresses important principles that can help you to make high-quality graphics.

Chapter 3: Making Your Own ggplot Theme

Describes how journalists at the BBC made a custom theme for the ggplot2 data visualization package. We’ll walk through the package they created, and in the process, you’ll learn how to make your own theme.

Chapter 4: Creating Maps

Walks through 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: Crafting High-Quality Tables

Shows you how to use the gt package to make high-quality tables in R. We draw from a conversation with R table connoisseur Tom Mock to learn how to apply design principles that ensure your tables communicate effectively.

Part 2: Reports, Presentations, and Websites

The second part of the book focuses on using R Markdown to communicate efficiently. You’ll learn how to incorporate visualizations like the ones discussed in Part 1 into complete reports, slideshow presentations, and static websites generated entirely using R code.

Chapter 6: Writing Reports in R Markdown

Introduces R Markdown, a tool that allows you to generate a professional report in R. This chapter will cover the structure of an R Markdown document, show 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: Using Parameters to Automate Reports

Covers one of the advantages of using R Markdown: the fact that you can produce multiple reports at the same time using a technique called parameterized reporting. We explore 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: Making Slideshow Presentations with xaringan

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: Building Websites with distill

Shows you how to create your own website with R Markdown and the distill package. We explore how distill works by considering a website about COVID-19 rates in Westchester County, New York. In the process, we cover how to create pages on your site, add interactivity through R packages, and deploy your website using several options.

Chapter 10: Reproducible Reporting with Quarto

Explains how to use Quarto, the next-generation version of R Markdown. You'll learn to use Quarto to do all of the things you did previously in R Markdown (reports, parameterized reporting, slideshow presentations, and websites).

Part 3: Automation and Collaboration

The last part of the book focuses on ways you can use R to automate your work and share it with others.

Chapter 10: 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 United States Census Bureau data. You’ll learn how the packages work and how to use them to automate the process of accessing data.

Chapter 11: Creating Your Own R Packages

Shows you how to create your own functions and packages to share your code with others. One of the major benefits of R is that you can create your own functions to automate common tasks, then bundle them into a package other users can import. This chapter covers a few example functions. Then you’ll learn how to create your own package by learning from R developers who built packages to improve the work of researchers at the Moffitt Cancer Center.

By the end of the book, you should be able to use R for a wide range of non-statistical tasks. You’ll know 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 yourself can develop. Let’s dive in.