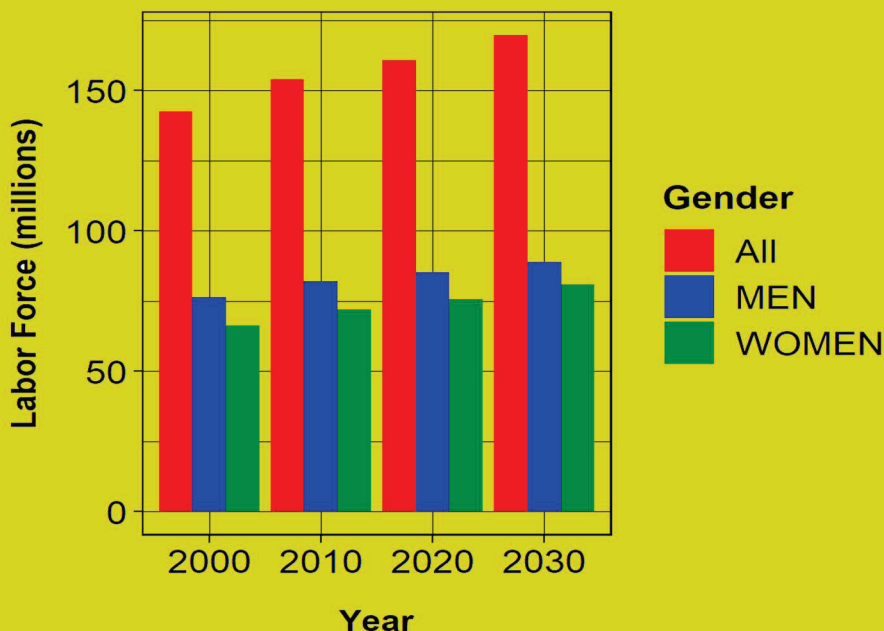


Using R For Excel Analysts

Tools for Increasing Productivity



Kilem L. Gwet, Ph.D.

USING R FOR EXCEL ANALYSTS

Tools for Increasing Productivity

First Edition

Other Books by the author:

- ▶ Beginner's Guide to Principal Components: Applications with Microsoft Excel. ISBN: 0970806256 / 978-0970806253.
- ▶ Inter-Rater Reliability using SAS: A Practical Guide for Analyzing Categorical & Quantitative Ratings. ISBN: 1792374887 / 978-1792374883.
- ▶ HANDBOOK OF INTER-RATER RELIABILITY, (5th Edition)
The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters. VOLUME 1: Analysis of Categorical Ratings. ISBN: 1792354630 / 978-1792354632
- ▶ HANDBOOK OF INTER-RATER RELIABILITY, (5th Edition)
The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters. VOLUME 2: Analysis of Quantitative Ratings. ISBN: 1792354649 / 978-1792354649

<https://agreestat.com/books/>

USING R FOR EXCEL ANALYSTS

Tools for Increasing Productivity

Kilem Li Gwet, Ph.D.

AgreeStat Analytics
Maryland, USA

Copyright © 2022 by Kilem Li Gwet, Ph.D. All rights reserved.

Published by AgreeStat Analytics. Printed and bound in the United States of America.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by an information storage and retrieval system – except by a reviewer who may quote brief passages in a review to be printed in a magazine or a newspaper – without permission in writing from the publisher. For information, please contact AgreeStat Analytics at the following address:

AgreeStat Analytics

20416 Davencroft Ct,

Montgomery Village, MD 20886-4379 USA

e-mail: contact@agreestat.com

This publication is designed to provide accurate and authoritative information in regard of the subject matter covered. However, it is sold with the understanding that the publisher assumes no responsibility for errors, inaccuracies or omissions. The publisher is not engaged in rendering any professional services. A competent professional person should be sought for expert assistance.

Publisher's Cataloguing in Publication Data:

Gwet, Kilem Li

Using R for Excel Analysts

Tools for Increasing Productivity / By Kilem Li Gwet

p. cm.

Includes bibliographical references and index.

1. Data Analysis
2. Statistical Computing
3. R Software
4. Microsoft Excel

ISBN 978-1-7923-7489-0

Contents

Preface	1
About the Author	5
1 Introduction	7
1.1 <i>Short Overview of R and Spreasheets</i>	8
1.2 <i>Spreadsheets: Advantages and Disadvantages</i>	9
1.2.1 <i>Advantages of Excel Spreadsheets</i>	10
1.2.2 <i>Problems with Spreadsheets</i>	13
1.3 <i>Using R to Automate Spreadsheets</i>	19
1.3.1 <i>Interaction Between R and Excel</i>	21
1.3.2 <i>Creating and Sharing Analytic Reports Programmatically</i>	24
1.4 <i>Concluding Remarks</i>	27
2 Setting Up the R Environment	29
2.1 <i>Introduction</i>	30
2.2 <i>RStudio Cloud</i>	30
2.3 <i>RStudio Desktop</i>	35
2.4 <i>Using RStudio</i>	39
2.4.1 <i>R Commands in RStudio</i>	39
2.4.2 <i>RStudio Workflow</i>	41
2.5 <i>Basic Use of R in RStudio</i>	44
2.5.1 <i>R Objects</i>	45
2.5.2 <i>Built-In Functions</i>	62

CONTENTS

CONTENTS

2.6 *Concluding Remarks* 62

3 R Datasets 65

3.1 *Introduction* 66

3.2 *Manipulating Datasets in R* 68

3.2.1 *R Built-In Datasets* 69

3.2.2 *Creating New Datasets in R* 72

3.2.3 *Saving R Objects to a Disk in R Data Formats* 75

3.2.4 *Exporting R Datasets* 86

3.3 *Projects and Working Directories* 88

3.3.1 *The Problem* 89

3.3.2 *Using Projects in R* 92

3.3.3 *Organizing Project Directories* 96

3.4 *Concluding Remarks* 98

4 Exploratory Data Analysis with Tidyverse 101

4.1 *Introduction* 103

4.2 *Basic Analysis of R Datasets* 104

4.2.1 *Exploring R Datasets* 107

4.2.2 *Slicing Datasets* 108

4.2.3 *Using the Pipe Operator* 119

4.2.4 *Producing Summary Statistics* 123

4.2.5 *Miscellaneous Useful Functions* 138

4.3 *Combining Datasets* 146

4.3.1 *Using Base R* 148

4.3.2 *Using the Tidyverse Package* 152

4.4 *Concluding Remarks* 160

5 Data Visualization 163

5.1 *Introduction* 165

5.2 *Graphics with ggplot2* 165

5.2.1 *Changing the Plot Theme and Adding Titles and Subtitles* . 170

CONTENTS

CONTENTS

5.2.2	<i>Different Types of Plots</i>	172
5.3	<i>Scatterplot and Lineplots</i>	173
5.3.1	<i>Simple Scatterplots</i>	174
5.3.2	<i>Lineplots</i>	183
5.3.3	<i>Connected Scatterplots</i>	190
5.4	<i>Bar Charts</i>	193
5.4.1	<i>Vertical barchart showing counts (Figure 5.14)</i>	195
5.4.2	<i>Vertical barchart showing percentages and labels</i>	197
5.4.3	<i>Horizontal barchart showing percentages and labels</i>	198
5.4.4	<i>Vertical & Horizontal Barcharts, on the same Figure</i>	200
5.4.5	<i>Vertical and Horizontal Stacked Barcharts with Labels in Figures 5.18 & 5.19</i>	204
5.4.6	<i>Vertical and Horizontal Dodged Bar Charts with Labels in Figures 5.20 & 5.21</i>	208
5.5	<i>Pie Charts</i>	211
5.5.1	<i>Two-dimensional Piecharts</i>	211
5.5.2	<i>Three-dimensional Piecharts</i>	213
5.6	<i>Draw an equation as a continuous curve</i>	218
5.7	<i>Concluding Remarks</i>	221
6	Programmatic Reporting	223
6.1	<i>Introduction</i>	225
6.2	<i>Installing R Markdown</i>	226
6.3	<i>Creating Word, PDF and HTML Reports</i>	227
6.3.1	<i>Create a New R Markdown Document</i>	228
6.3.2	<i>Processing R Markdown Documents to Create Reports</i>	231
6.3.3	<i>Adding R Code to R Markdown Documents</i>	234
6.3.4	<i>Review of the R Markdown Template</i>	236
6.3.5	<i>R Markdown Document Example: laborforce.Rmd</i>	238
6.4	<i>Spreadsheet Automation</i>	243
6.4.1	<i>Data Analysis Report in Excel</i>	244
6.4.2	<i>Using Microsoft OneDrive</i>	246

CONTENTS

CONTENTS

6.4.3	<i>Connecting R to OneDrive with Microsoft365R</i>	251
6.4.4	<i>Data Analysis Report in Google Sheets</i>	263
6.4.5	<i>Connecting R and Excel</i>	265
6.4.6	<i>Connecting R and Google Sheets</i>	267
6.5	<i>Concluding Remarks</i>	269
7	Connecting R to Excel with the Openxlsx Package	271
7.1	<i>Introduction</i>	272
7.2	<i>Reading an Excel Workbook into an R Data Frame</i>	272
7.3	<i>Writing an R Object to Excel</i>	278
7.3.1	<i>Writing an R Object to an Excel Worksheet</i>	279
7.3.2	<i>Changing the Color of Worksheet Tabs</i>	284
7.3.3	<i>Alternative Methods for Writing a Data Frame to Excel</i>	288
7.4	<i>Formatting Individual Cells</i>	295
7.5	<i>Inserting R Graphs Into an Excel Worksheet</i>	305
7.6	<i>Concluding Remarks</i>	308
8	Connecting R to Excel with the xlsx Package	311
8.1	<i>Introduction</i>	312
8.2	<i>Analysis of Excel Data with R</i>	313
8.3	<i>Manipulating Individual Cells</i>	321
8.3.1	<i>Writing to and Reading from Cells</i>	322
8.3.2	<i>Formatting Individual Cells</i>	331
8.3.3	<i>More Worksheet Formatting Options</i>	336
8.4	<i>Options for Creating Cell Styles</i>	343
8.5	<i>Adding a Plot to an Excel Worksheet</i>	349
8.6	<i>Concluding Remarks</i>	351
9	Connecting R to GoogleSheets	353
9.1	<i>The Googlesheets4 Package</i>	354
9.2	<i>Read a Sheet into an R data frame</i>	354
9.3	<i>Read a Specific Sheet into a Data Frame</i>	357

CONTENTS

CONTENTS

9.4 *Writing an R Data Frame to a Google Spreadsheet* 360

9.5 *Writing an R data frame to a specific range of cells* 363

9.6 *Manipulating Google Spreadsheets* 367

9.7 *Concluding Remarks* 370

10 Using GitHub to Manage Versions 373

10.1 *Introduction* 375

10.2 *What is GitHub?* 376

10.3 *Using the GitHub Desktop* 377

10.3.1 *Creating a GitHub Account* 378

10.3.2 *Installing GitHub Desktop* 382

10.3.3 *Basic Use of GitHub Desktop* 387

10.4 *How to Use GitHub for Version Control?* 395

10.4.1 *Committing, Reviewing and Pushing Changes* 397

10.4.2 *Branching in GitHub* 403

10.4.3 *Merging and Pulling Branches* 406

10.5 *Cloning and Forking* 415

10.5.1 *Cloning / Forking Your Own Repositories* 416

10.5.2 *Cloning repositories created by others* 419

10.6 *Alternatives to GitHub* 422

10.7 *Concluding Remarks* 425

A Operators in R 427

A.1 *Selected Built-In Functions in R* 428

A.2 *Relational Operators* 434

B Additional Data and Scripts 437

B.1 *US Quarterly Gross Domestic Product (GDP)* 438

B.2 *A Model R Markdown Document* 440

Bibliography 442

List of R Functions 445

CONTENTS

CONTENTS

Author Index	449
Subject Index	451

Preface

In this book, you will learn how to use R to automate tedious and repetitive tasks in **Excel** and **Google Sheets**. In addition to using R to leverage Excel's capability in data management and analysis, you will discover all the power of R in data visualization and programmatic reporting. You will see how R can automate the entire workflow from raw data to the final report in PDF, HTML or Word formats. This will allow you to update your analysis report with new data in one click. The use of **OneDrive** and **GoogleDrive** as a way of sharing your reports with others is extensively discussed. As a bonus, I added a chapter on the important issue of version control with **GitHub**. Here is where you will learn how to manage the different versions of your R scripts and share them with others if needed, in a collaborative work environment. These are some of the essential skills you must develop to be an effective analyst in the new world of data science.

Despite its limitations, Microsoft Excel remains one of the most widely-used data analysis tool among professionals in the business world. It is intuitive and the learning curve is not steep. Moreover, you typically get what you see. You can add formulas in a worksheet, and obtain new results in realtime when you change your input data. Therefore, Excel cannot be ignored when it comes to distributing your work to a large audience. If you want to share data tables or small datasets, then using the spreadsheet (Excel or Google Sheets) remains one of the best options at your disposal.

However, the use of Excel as your main tool for analysis will make it more

difficult for you to tackle several challenges frequently encountered in data analysis projects. Here are 2 examples:

- *Data Management*

Most basic data analysis tasks in Excel start with a static rectangular data table in a worksheet. Excel formulas such as `=SUM(E3:H3)` can be added to the worksheet to do some calculations. However, your input data must sometimes come from many different worksheets or workbooks. This requires that you sort and merge different data sources before you can build a complete dataset, which is ready for analysis. If manually done, such a task can become tedious and error-prone.

To resolve this problem, Excel offers 2 solutions. You can use Excel Visual Basic for Applications (or VBA), which is the programming language that comes with Excel. Alternatively, you can use Excel Power Query, which is an Excel Add-on that can help you resolve this problem through a point-and-click menu driven approach. While VBA requires some programming skills, Power Query also allows you to write queries in addition to allowing you to do the work manually.

R is a non-Excel option for performing the same tasks and much more. R offers more diverse and more powerful data management tools than Excel. With R being an open source software, independent developers across the world add new tools to R almost every day. Moreover, you can receive free help from an extensive community of R users.

- *Special Analyses*

When you have a standard analysis to conduct involving basic arithmetic and commonly-use statistical techniques, you will likely find an Excel formula you can use. However, if you want to implement a special technique, then only the use of VBA will solve your problem. I am an experienced VBA programmer and was able to develop some fancy applications with it. VBA learning curve is steep and developing a VBA solution can be

time-consuming due to the lack of a rich library of functions and the inability to use solutions developed by others in a convenient manner.

Developing an R solution for implementing a special analysis will generally be much faster than developing a similar VBA solution. Almost all statistical techniques you can think of are already implemented in R either through base R or through one of the many packages developed by third parties.

I strongly recommend investing time to learn R and use it as your language of choice for automating Excel and Google Sheets. Your code will be independent of Excel and Google Sheets, can easily be perused by colleagues, will be easy to maintain, and above all, will be highly portable. The skills you will develop, will be transferable outside of the Microsoft and Google ecosystems of software products.

R is free and RStudio is a user-friendly platform for writing R code. With the development of RStudio Cloud, the Internet is all you need to run R. That is, downloading and installing the desktop versions of R and RStudio is no longer necessary to use R and automate Excel, Google Sheets and more. There is so much you can do with R that only your ambition and imagination can limit you. I hope this book will trigger your interest in R as a supplement to Excel, and stimulate your willingness to become a better data analyst.

Kilem Li Gwet, Ph.D.

About the Author

I am a mathematical statistician and a quantitative management consultant. I obtained my PhD in Mathematics and Statistics in 1997 from the Carleton University's School of Mathematics and Statistics, Ottawa (Canada). My initial specialization was the design and analysis of statistical surveys. However, after working for various organizations for over 20 years, my interests have considerably shifted in the past few years. My focus now is on the fields of statistical computing (with R, SAS and Excel), and inter-rater reliability analysis.

In the field of inter-rater reliability, I wrote many books and published several papers in peer-reviewed journals. I am also the developer of the widely-used Excel and R-based Cloud versions of the **AgreeStat360** software dedicated to the analysis of inter-rater reliability data. The link <https://agreestat.com/> gives you a glimpse into my work in the field of inter-rater reliability analysis.

