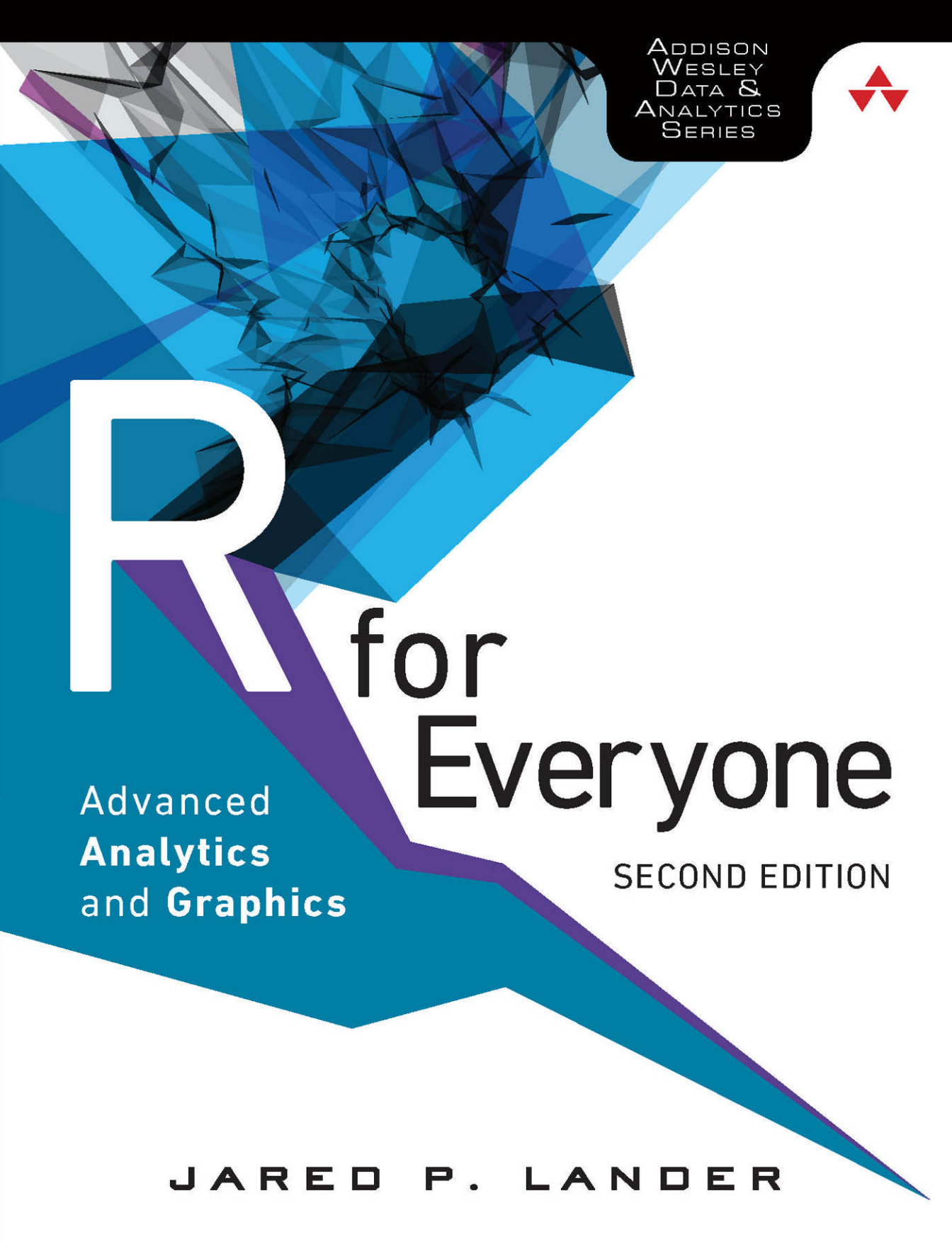




ADDISON
WESLEY
DATA &
ANALYTICS
SERIES



R for Everyone

Advanced
Analytics
and **Graphics**

SECOND EDITION

J A R E D P . L A N D E R

R for Everyone

Second Edition

R for Everyone: Advanced Analytics and Graphics

Table of Contents

Cover

Title Page

Copyright Page

Contents

Foreword

Preface

Acknowledgments

About the Author

1 Getting R

1.1 Downloading R

1.2 R Version

1.3 32-bit vs. 64-bit

1.4 Installing

1.5 Microsoft R Open

1.6 Conclusion

2 The R Environment

2.1 Command Line Interface

2.2 RStudio

2.3 Microsoft Visual Studio

2.4 Conclusion

3 R Packages

Table of Contents

3.1 Installing Packages

3.2 Loading Packages

3.3 Building a Package

3.4 Conclusion

4 Basics of R

4.1 Basic Math

4.2 Variables

4.3 Data Types

4.4 Vectors

4.5 Calling Functions

4.6 Function Documentation

4.7 Missing Data

4.8 Pipes

4.9 Conclusion

5 Advanced Data Structures

5.1 data.frames

5.2 Lists

5.3 Matrices

5.4 Arrays

5.5 Conclusion

6 Reading Data into R

6.1 Reading CSVs

6.2 Excel Data

6.3 Reading from Databases

6.4 Data from Other Statistical Tools

6.5 R Binary Files

Table of Contents

6.6 Data Included with R

6.7 Extract Data from Web Sites

6.8 Reading JSON Data

6.9 Conclusion

7 Statistical Graphics

7.1 Base Graphics

7.2 ggplot2

7.3 Conclusion

8 Writing R functions

8.1 Hello, World!

8.2 Function Arguments

8.3 Return Values

8.4 do.call

8.5 Conclusion

9 Control Statements

9.1 if and else

9.2 switch

9.3 ifelse

9.4 Compound Tests

9.5 Conclusion

10 Loops, the Un-R Way to Iterate

10.1 for Loops

10.2 while Loops

10.3 Controlling Loops

10.4 Conclusion

11 Group Manipulation

Table of Contents

11.1 Apply Family

11.2 aggregate

11.3 plyr

11.4 data.table

11.5 Conclusion

12 Faster Group Manipulation with dplyr

12.1 Pipes

12.2 tbl

12.3 select

12.4 filter

12.5 slice

12.6 mutate

12.7 summarize

12.8 group_by

12.9 arrange

12.10 do

12.11 dplyr with Databases

12.12 Conclusion

13 Iterating with purrr

13.1 map

13.2 map with Specified Types

13.3 Iterating over a data.frame

13.4 map with Multiple Inputs

13.5 Conclusion

14 Data Reshaping

14.1 cbind and rbind

Table of Contents

14.2 Joins

14.3 reshape2

14.4 Conclusion

15 Reshaping Data in the Tidyverse

15.1 Binding Rows and Columns

15.2 Joins with dplyr

15.3 Converting Data Formats

15.4 Conclusion

16 Manipulating Strings

16.1 paste

16.2 sprintf

16.3 Extracting Text

16.4 Regular Expressions

16.5 Conclusion

17 Probability Distributions

17.1 Normal Distribution

17.2 Binomial Distribution

17.3 Poisson Distribution

17.4 Other Distributions

17.5 Conclusion

18 Basic Statistics

18.1 Summary Statistics

18.2 Correlation and Covariance

18.3 T-Tests

18.4 ANOVA

18.5 Conclusion

Table of Contents

19 Linear Models

19.1 Simple Linear Regression

19.2 Multiple Regression

19.3 Conclusion

20 Generalized Linear Models

20.1 Logistic Regression

20.2 Poisson Regression

20.3 Other Generalized Linear Models

20.4 Survival Analysis

20.5 Conclusion

21 Model Diagnostics

21.1 Residuals

21.2 Comparing Models

21.3 Cross-Validation

21.4 Bootstrap

21.5 Stepwise Variable Selection

21.6 Conclusion

22 Regularization and Shrinkage

22.1 Elastic Net

22.2 Bayesian Shrinkage

22.3 Conclusion

23 Nonlinear Models

23.1 Nonlinear Least Squares

23.2 Splines

23.3 Generalized Additive Models

23.4 Decision Trees

Table of Contents

23.5 Boosted Trees

23.6 Random Forests

23.7 Conclusion

24 Time Series and Autocorrelation

24.1 Autoregressive Moving Average

24.2 VAR

24.3 GARCH

24.4 Conclusion

25 Clustering

25.1 K-means

25.2 PAM

25.3 Hierarchical Clustering

25.4 Conclusion

26 Model Fitting with Caret

26.1 Caret Basics

26.2 Caret Options

26.3 Tuning a Boosted Tree

26.4 Conclusion

27 Reproducibility and Reports with knitr

27.1 Installing a LaTeX Program

27.2 LaTeX Primer

27.3 Using knitr with LaTeX

27.4 Conclusion

28 Rich Documents with RMarkdown

28.1 Document Compilation

28.2 Document Header

Table of Contents

28.3 Markdown Primer

28.4 Markdown Code Chunks

28.5 htmlwidgets

28.6 RMarkdown Slideshows

28.7 Conclusion

29 Interactive Dashboards with Shiny

29.1 Shiny in RMarkdown

29.2 Reactive Expressions in Shiny

29.3 Server and UI

29.4 Conclusion

30 Building R Packages

30.1 Folder Structure

30.2 Package Files

30.3 Package Documentation

30.4 Tests

30.5 Checking, Building and Installing

30.6 Submitting to CRAN

30.7 C++ Code

30.8 Conclusion

A: Real-Life Resources

A.1 Meetups

A.2 Stack Overflow

A.3 Twitter

A.4 Conferences

A.5 Web Sites

A.6 Documents

Table of Contents

A.7 Books

A.8 Conclusion

B: Glossary

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Table of Contents

X

List of Figures

List of Tables

General Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

Table of Contents

W

X

Y

Z

Index of Functions

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

Table of Contents

V

W

X

Y

Index of Packages

A

B

C

D

F

G

H

J

K

L

M

N

P

Q

R

S

T

U

V

W

X

Table of Contents

Index of People

Data Index