

# Recommended Books

Introduction to Data Science

Nina Zumel

John Mount



# Lesson Goals

- Present an overview of a number of helpful and relevant books

(You certainly do not need to buy all, or even any of them, but it is good to know they are available)



R



# *Hands-On Programming with R*

**Garrett Grolemund**

A basic introduction to programming and R,  
using a project-oriented approach.

O'REILLY\*

Copyrighted Material



## Hands-On Programming with R

WRITE YOUR OWN FUNCTIONS AND SIMULATIONS

Garrett Grolemund  
Foreword by Hadley Wickham

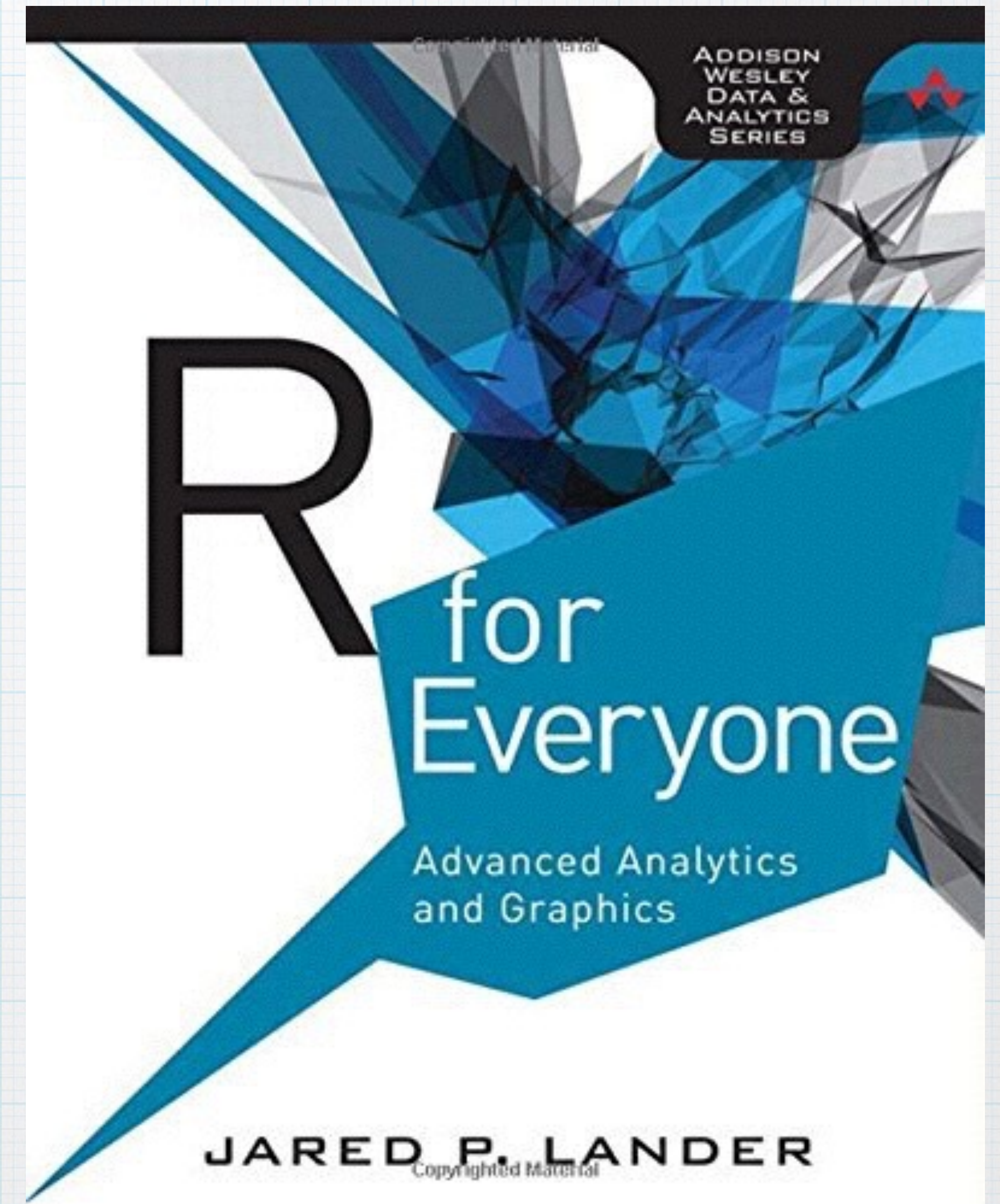
Copyrighted Material



# *R for Everyone* Jared P. Lander

Well-liked, very popular R book.

Covers programming basics and basic R structures. Good survey of basic statistics and modeling in R.



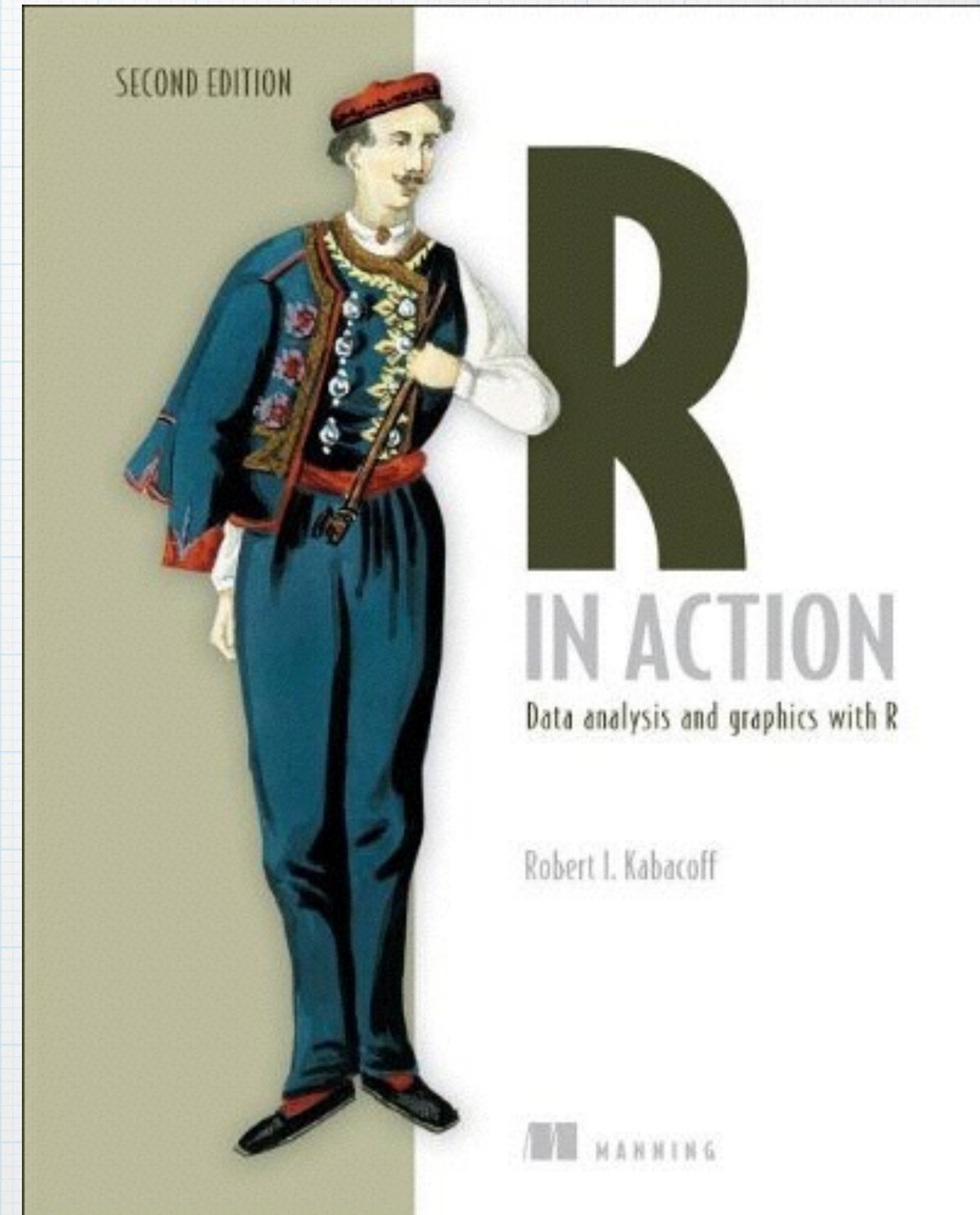


# *R in Action, Second Edition*

## Robert I. Kabacoff

Our go-to R reference.

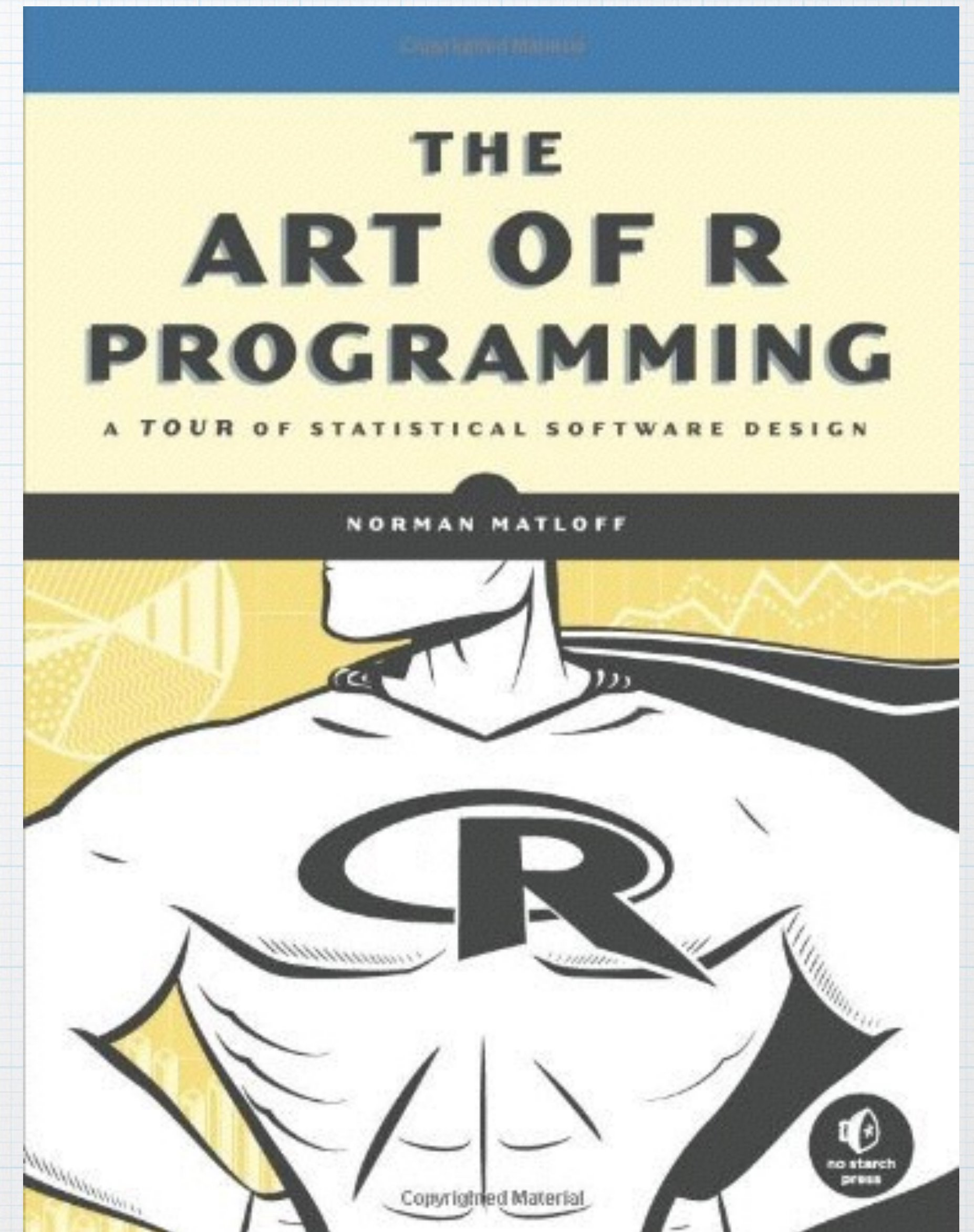
Covers R programming structures, data management and statistical functions. Less machine learning than Lander, but comprehensive coverage of classical statistical analysis as done in R.





# *The Art of R Programming* Norman Matloff

Discusses R from a Computer Science (programming) rather than Statistical perspective. Includes a chapter on R's debugging environment — very useful!



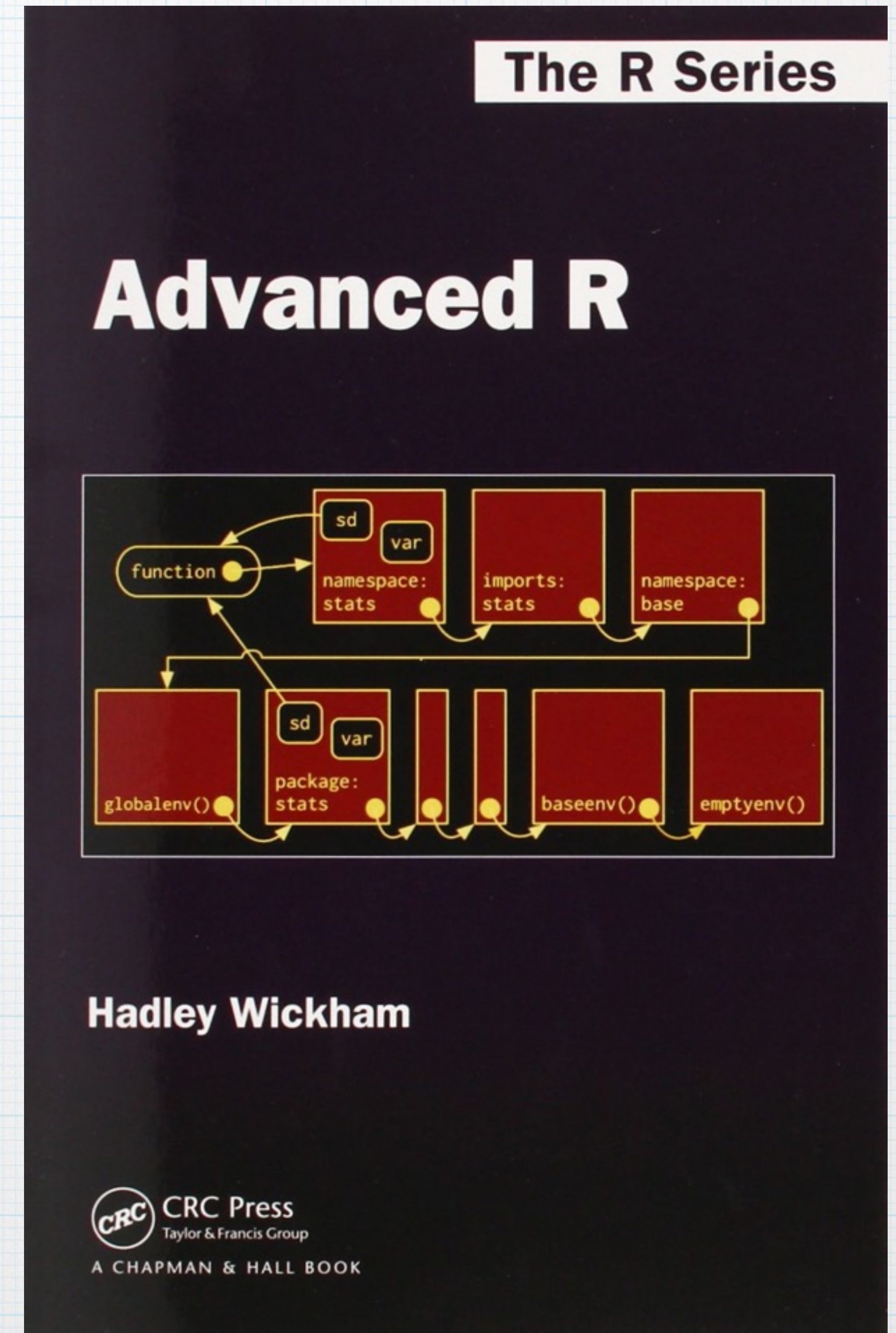


# *Advanced R*

## Hadley Wickham

Good deep dive into the internals of R as a programming language. A must if you are interested in developing R packages.

Also available online: <http://adv-r.had.co.nz>



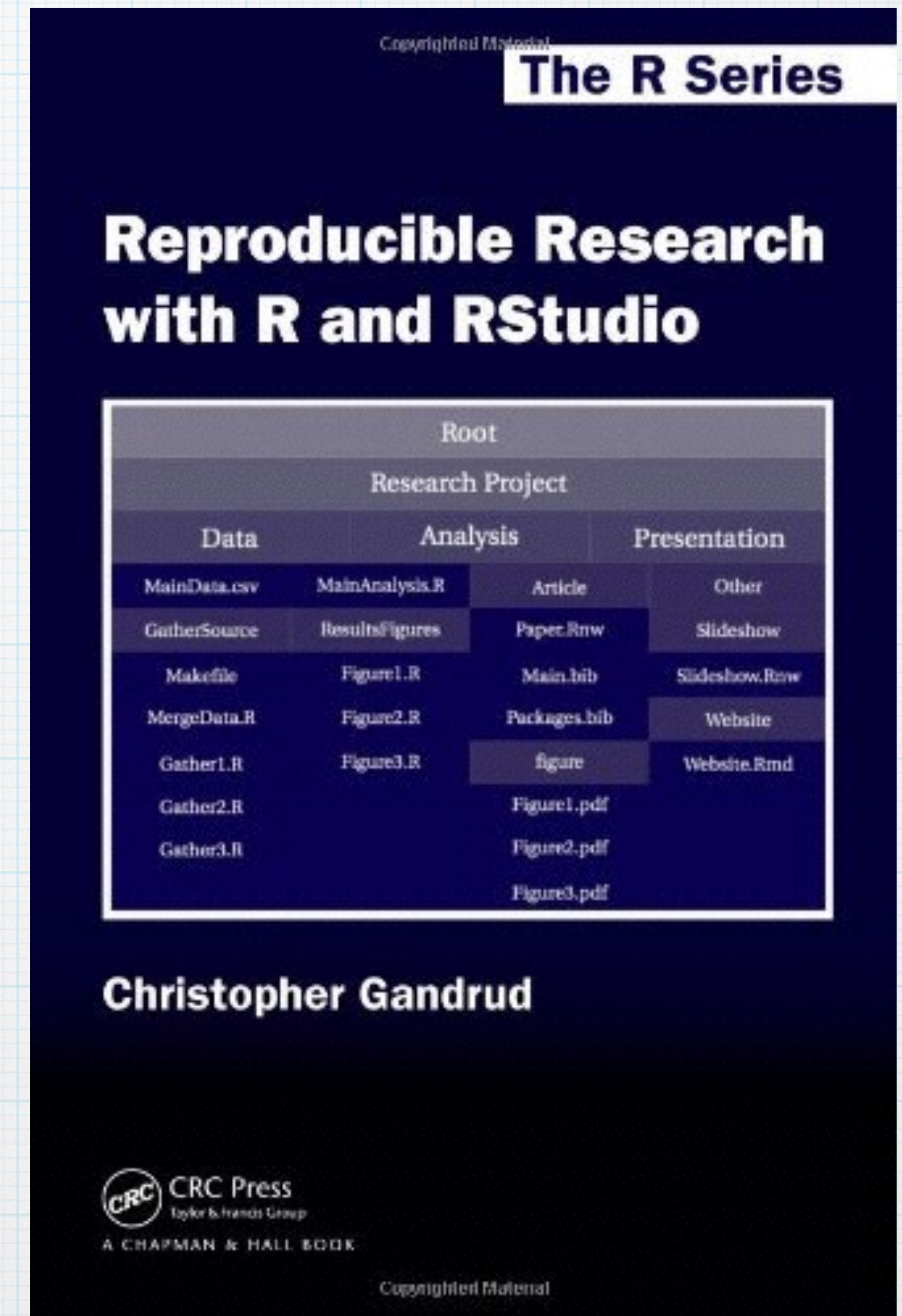


# Reproducible Research with R and RStudio

## Christopher Gandrud

Covers R markup and *knitr* within the RStudio environment.

Good introduction to the tool ecosystem around R, and gives good advice on setting up an R-based project data science project.

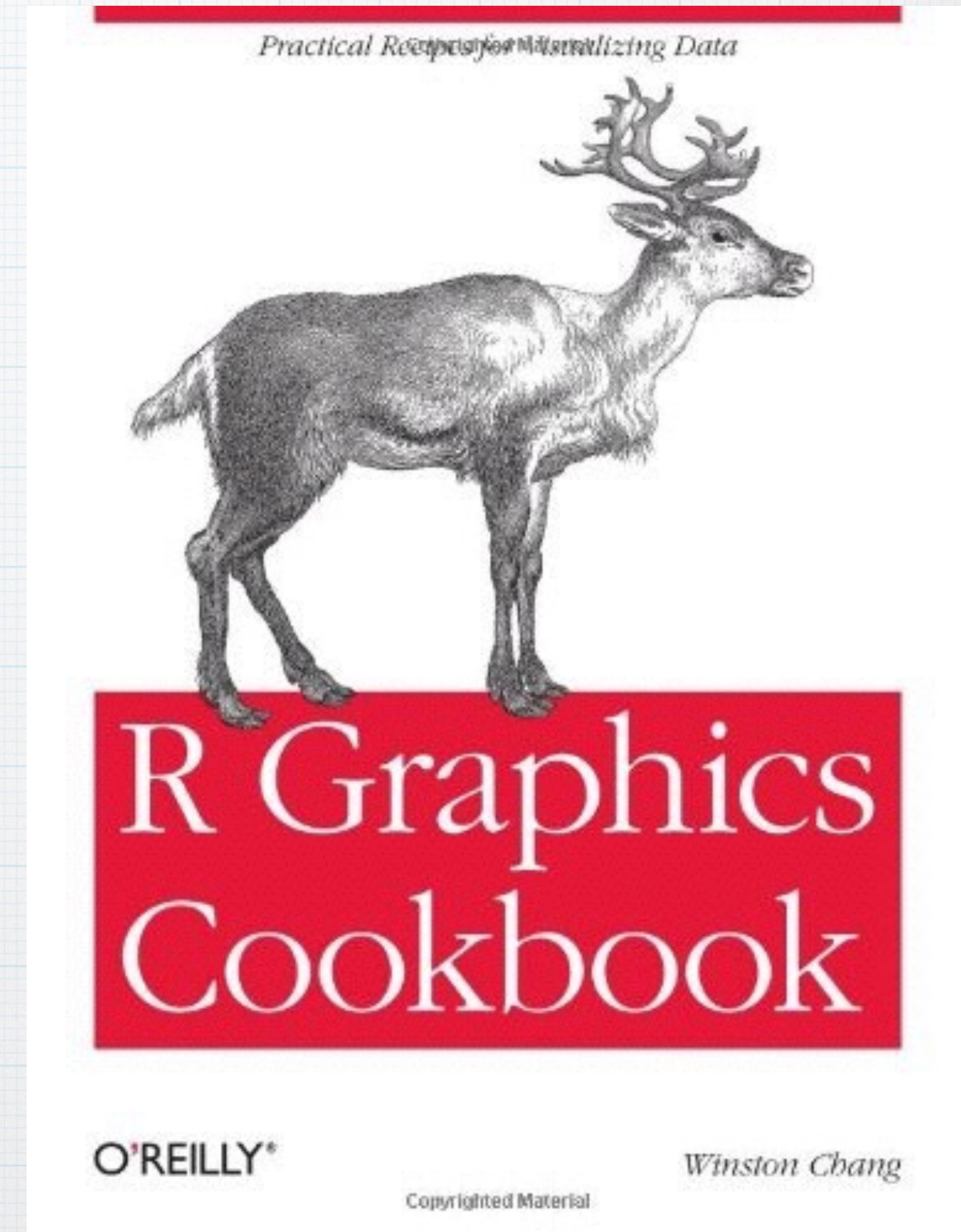




# *R Graphics Cookbook*

## Winston Chang

Useful collection of recipes for creating graphics in R, primarily in ggplot2.





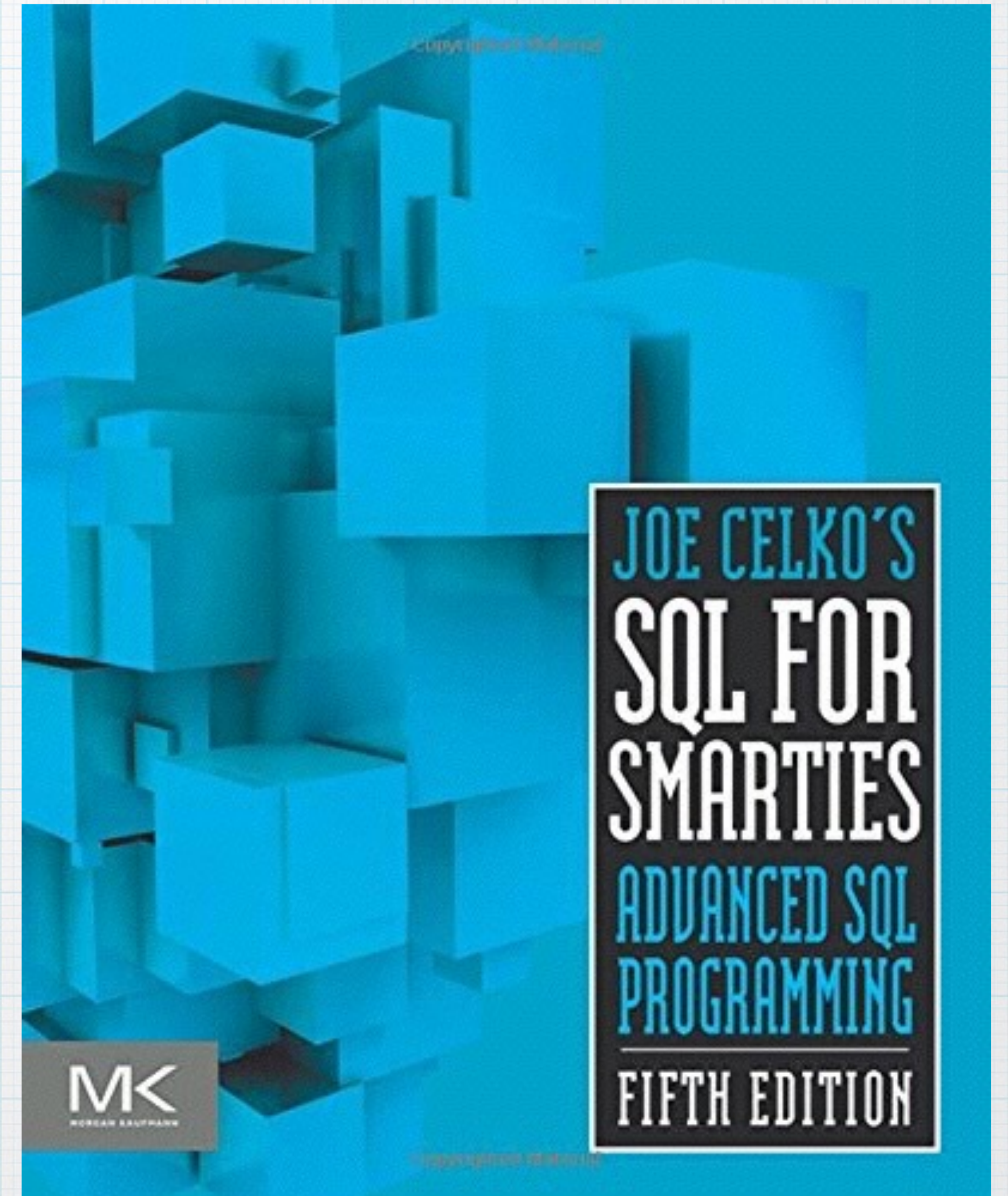
SQL



# *SQL for Smarties, Fourth Edition*

**Joe Celko**

Our go-to SQL reference.





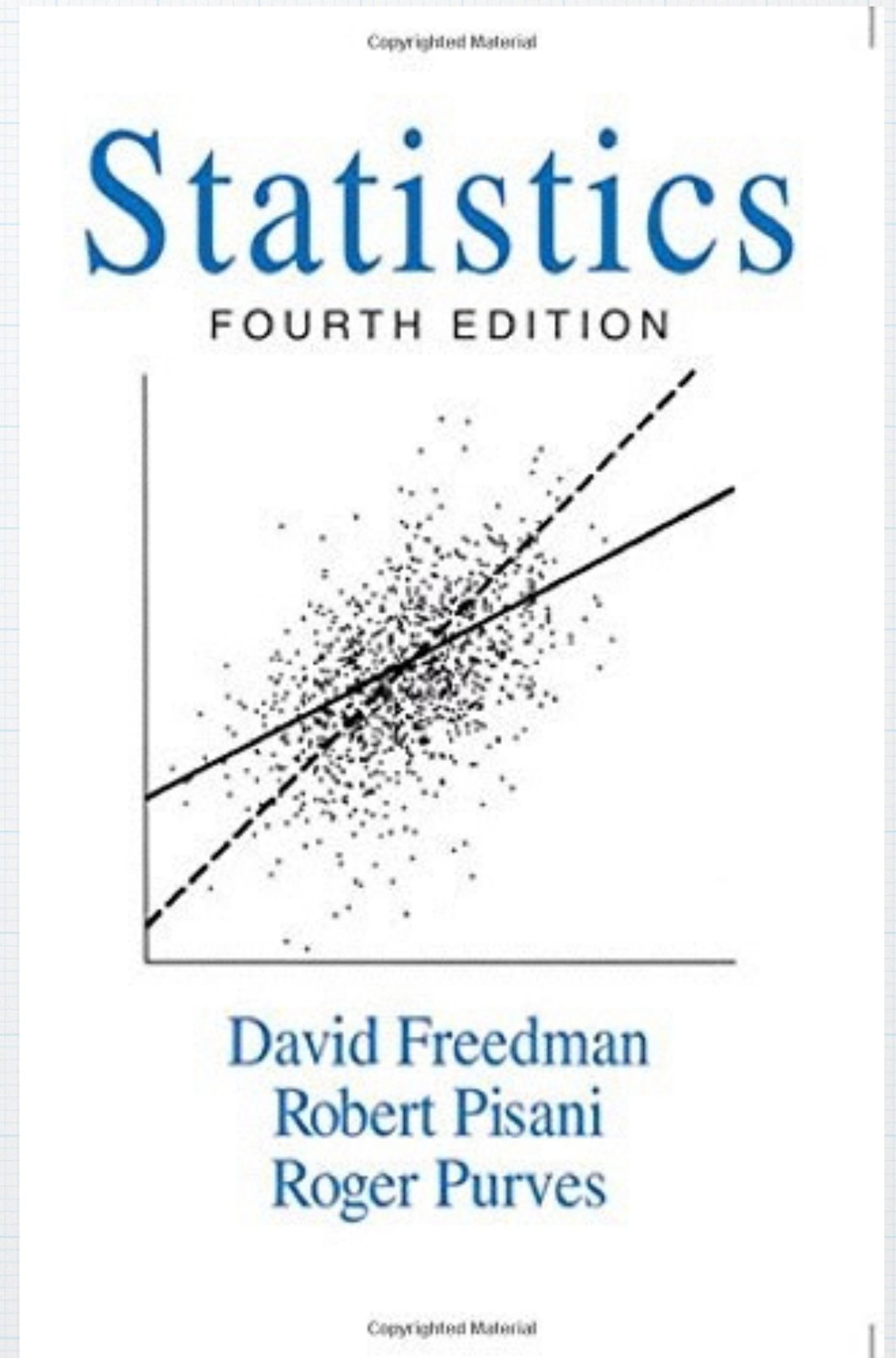
# Statistics



*Statistics, Fourth Edition*  
**David Freedman, Robert Pisani,  
Roger Purves**

Good introduction to Statistics and Probability.

Intuitive approach that relies on illustrations and examples, rather than formulas.



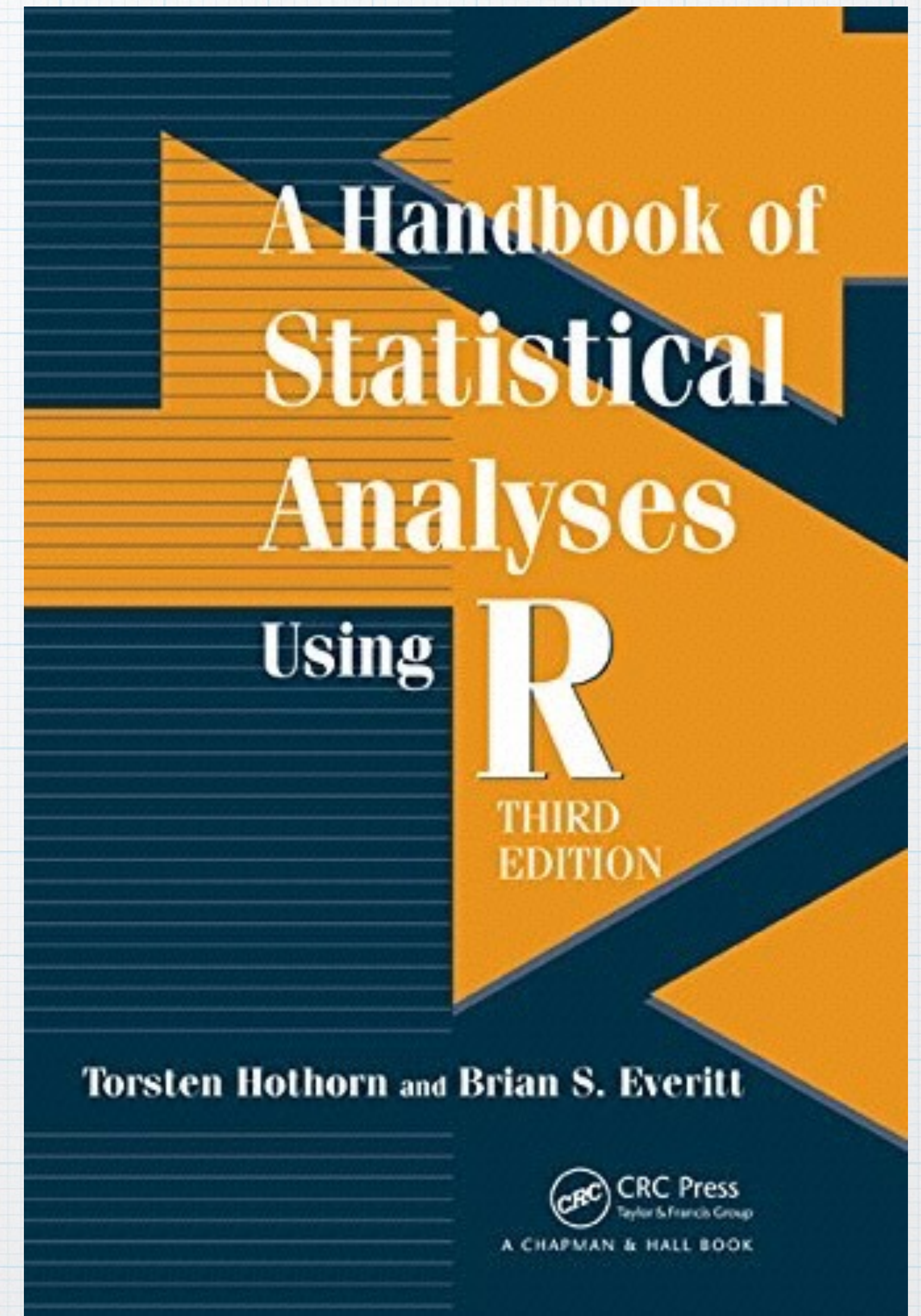


# *A Handbook of Statistical Analyses using R*

**Torsten Hothorn,  
Brian S. Everitt**

Intermediate level (assumes familiarity with R  
and basic statistics).

Covers several techniques in statistical data  
analysis, including linear and logistic regression,  
cluster analysis, and multidimensional scaling.





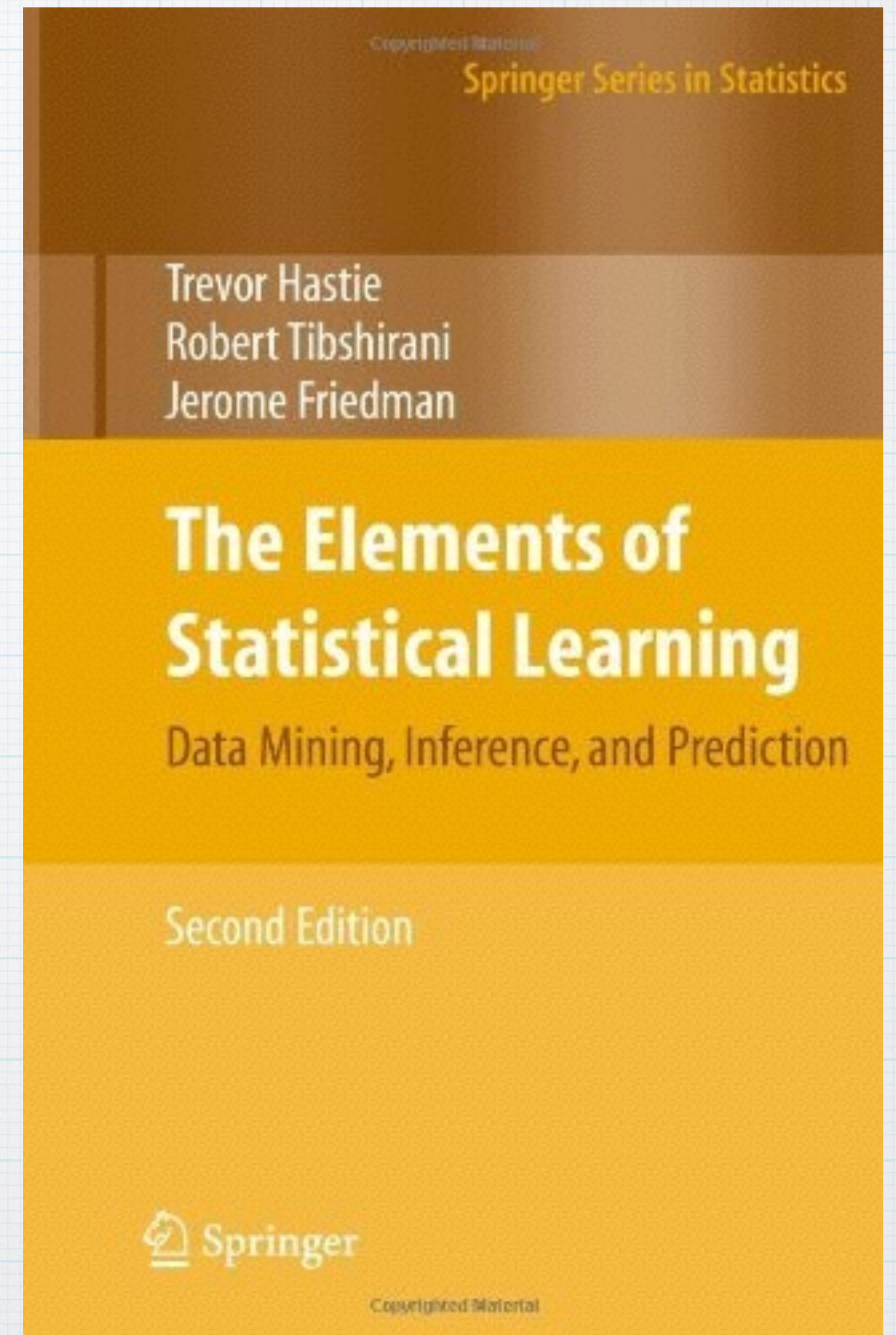
# Theory



*The Elements of Statistical  
Learning, 2nd edition*  
**Trevor Hastie, Robert Tibshirani,  
Jerome Friedman**

Our go-to machine learning reference.

Covers a wide variety of statistical machine learning algorithms, including their derivation, strengths and weaknesses.





# Practice

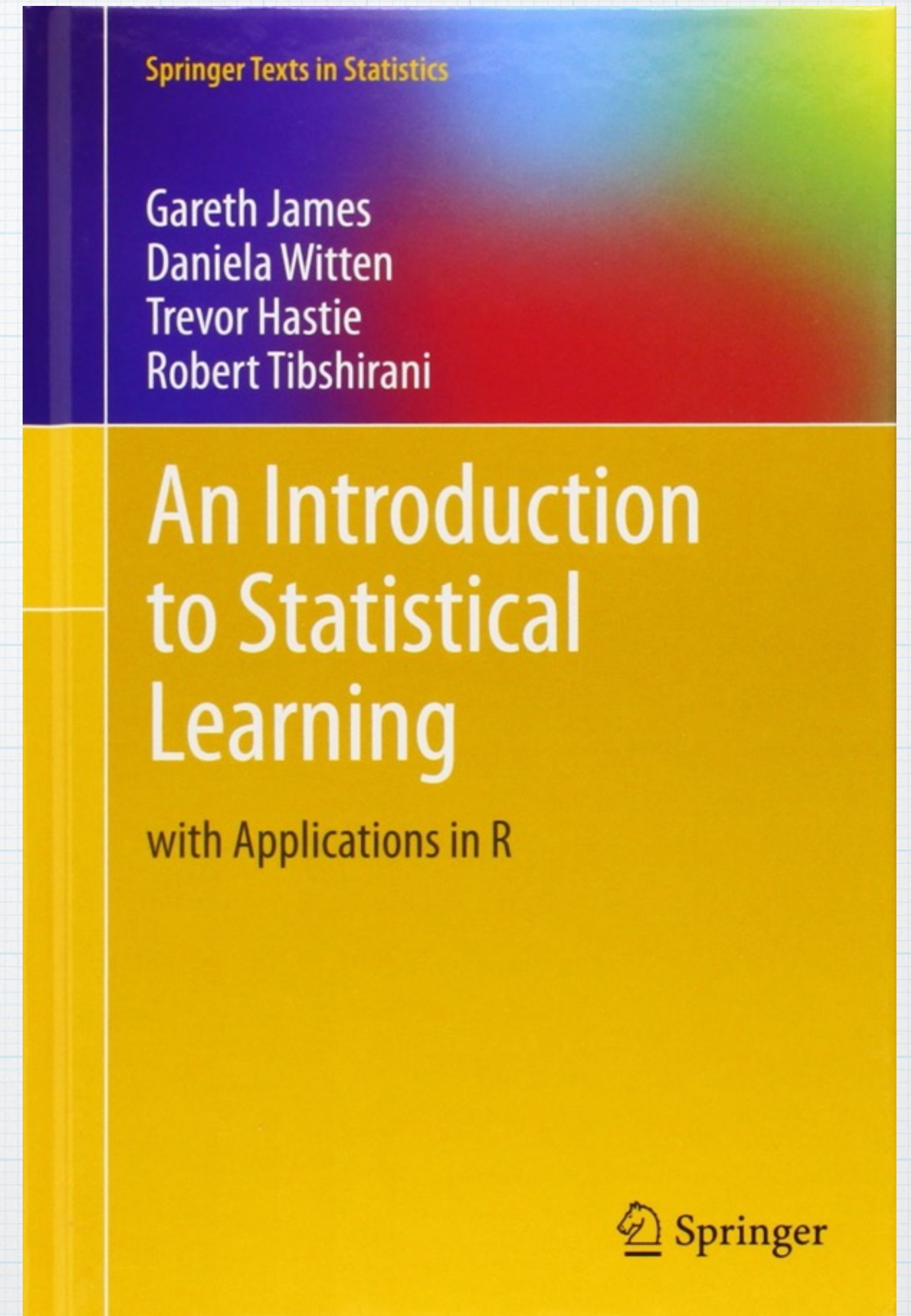


# *An Introduction to Statistical Learning*

**Gareth James, Daniela Witten,  
Trevor Hastie, Robert Tibshirani**

Companion text to *The Elements of Statistical Learning*.

More application oriented, with example code  
in R.



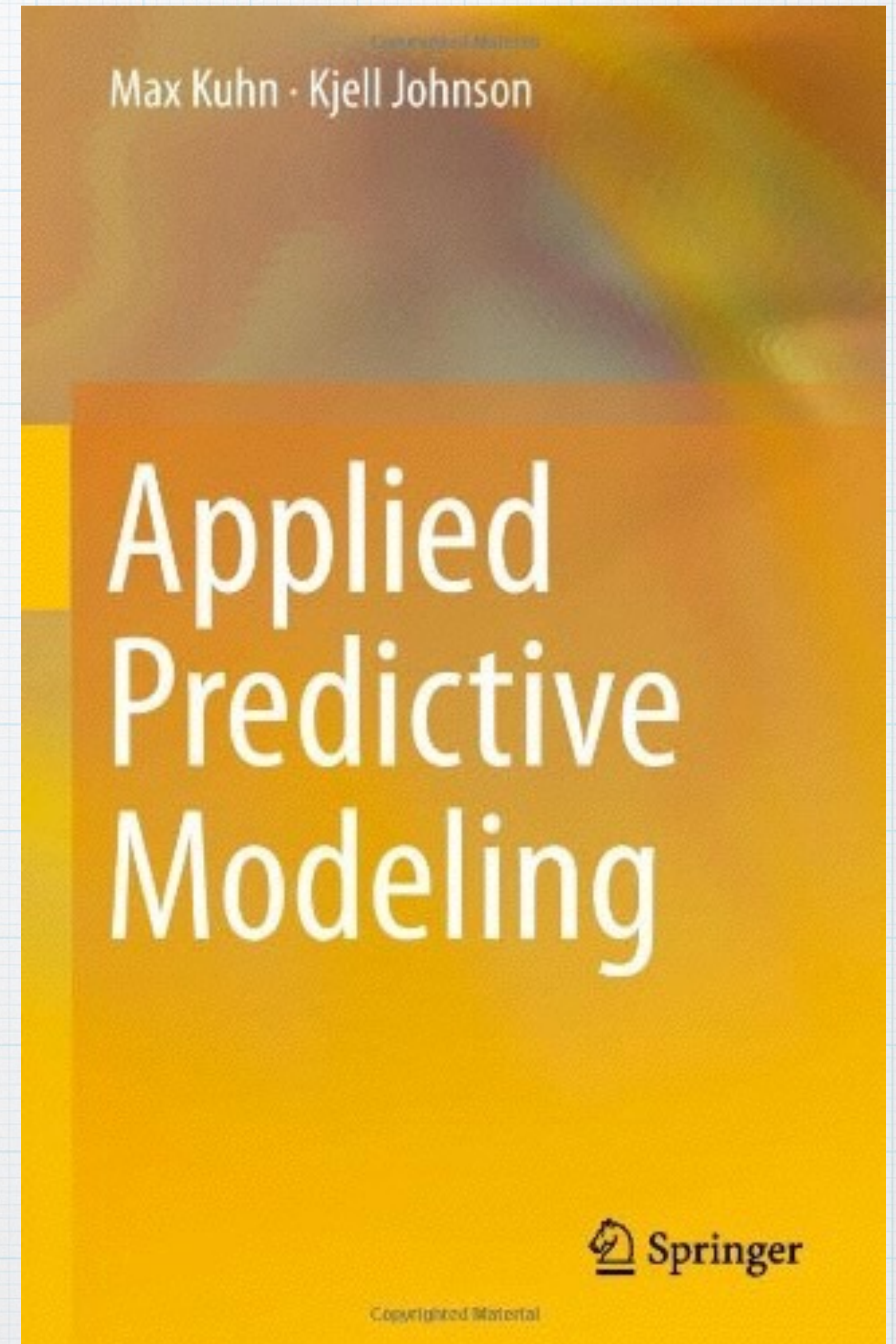


# *Applied Predictive Modeling*

**Max Kuhn, Kjell Johnson**

Practice-oriented approach to predictive modeling.

Good combination of theory and practice. Also covers data treatment, model evaluation, and feature selection. Example code in R.





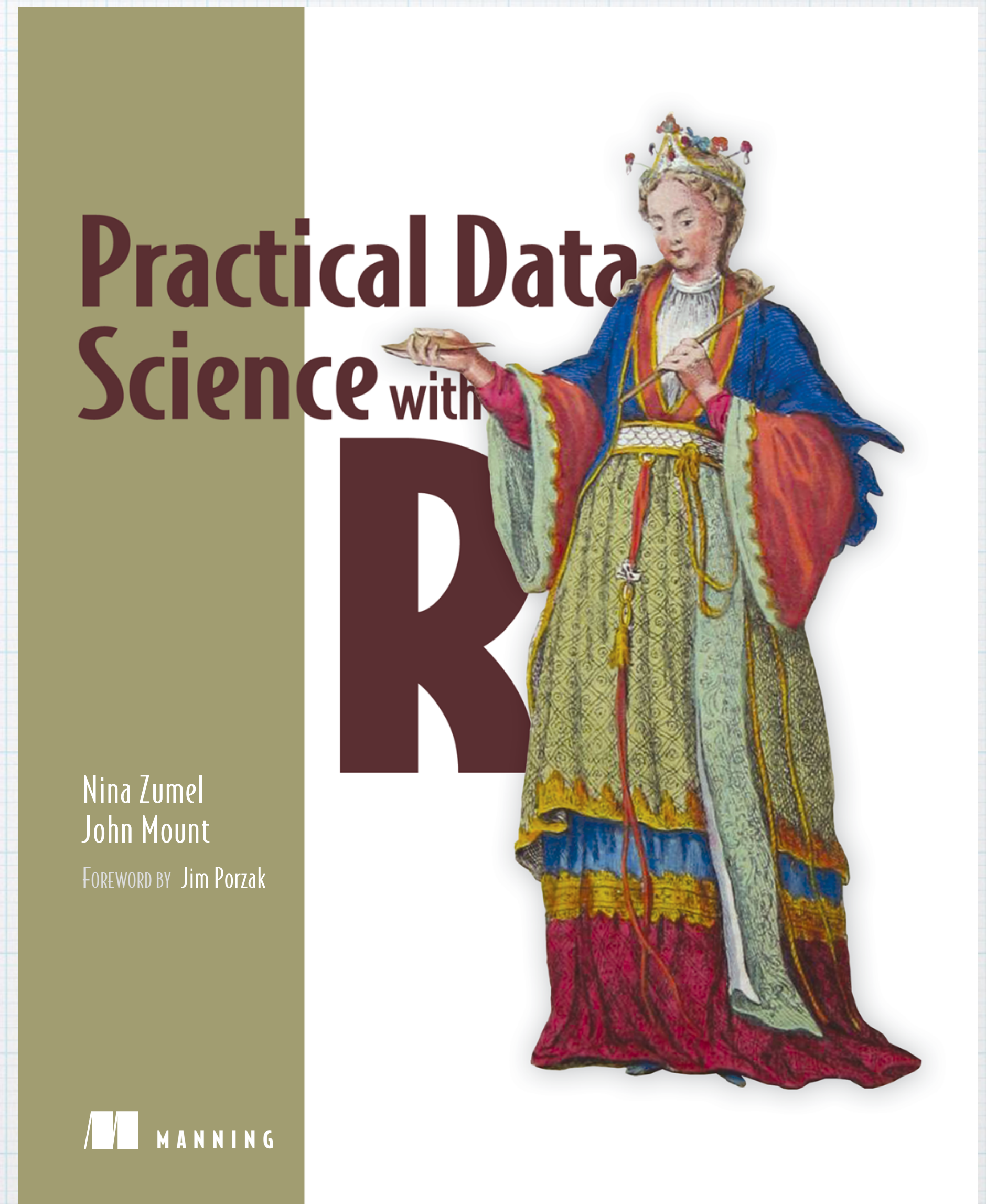
# *Practical Data Science with R*

## Nina Zumel, John Mount

Our book!

A practitioner-oriented view of data science.

Covers all aspects of a data science project: project management, data treatment, modeling and model evaluation, deployment and reporting. Extensive example code in R.





# And more

- More good books (our list is not comprehensive)
- Some web resources:
  - <http://www.r-bloggers.com>
  - <http://www.statmethods.net> (Quick-R)
  - <http://cran.r-project.org>



# What you should take away

- There are a *lot* of available resources
- Research what you actually need and you may need to get one or two extra books.