

Use R!

Advisors:

Robert Gentleman • Kurt Hornik • Giovanni Parmigiani

For other titles published in this series, go to
<http://www.springer.com/series/6991>

Hadley Wickham

ggplot2

Elegant Graphics for Data Analysis



Springer

Hadley Wickham
Rice University
Department of Statistics
Houston, TX
77005-1827
USA
hadley@rice.edu

Series Editors

Robert Gentleman
Program in Computational Biology
Division of Public Health Sciences
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue, N. M2-B876
Seattle, Washington 98109
USA

Kurt Hornik
Department of Statistik and Mathematik
Wirtschaftsuniversität Wien Augasse 2-6
A-1090 Wien
Austria

Giovanni Parmigiani
The Sidney Kimmel Comprehensive Cancer
Center at Johns Hopkins University
550 North Broadway
Baltimore, MD 21205-2011
USA

ISBN 978-0-387-98140-6 e-ISBN 978-0-387-98141-3
DOI 10.1007/978-0-387-98141-3
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009928510

© Springer Science+Business Media, LLC 2009

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

1	Introduction	1
1.1	Welcome to ggplot2	1
1.2	Other resources	2
1.3	What is the grammar of graphics?	3
1.4	How does ggplot2 fit in with other R graphics?	4
1.5	About this book	5
1.6	Installation	6
1.7	Acknowledgements	6
2	Getting started with qplot	9
2.1	Introduction	9
2.2	Datasets	10
2.3	Basic use	11
2.4	Colour, size, shape and other aesthetic attributes	12
2.5	Plot geoms	13
2.5.1	Adding a smoother to a plot	14
2.5.2	Boxplots and jittered points	16
2.5.3	Histogram and density plots	18
2.5.4	Bar charts	20
2.5.5	Time series with line and path plots	20
2.6	Faceting	22
2.7	Other options	23
2.8	Differences from plot	26
3	Mastering the grammar	27
3.1	Introduction	27
3.2	Fuel economy data	28
3.3	Building a scatterplot	29
3.4	A more complex plot	34
3.5	Components of the layered grammar	35
3.5.1	Layers	37
3.5.2	Scales	37
3.5.3	Coordinate system	38
3.5.4	Faceting	38
3.6	Data structures	39

4	Build a plot layer by layer	41
4.1	Introduction	41
4.2	Creating a plot	42
4.3	Layers	42
4.4	Data	45
4.5	Aesthetic mappings	46
4.5.1	Plots and layers	46
4.5.2	Setting vs. mapping	47
4.5.3	Grouping	49
4.5.4	Matching aesthetics to graphic objects	52
4.6	Geoms	55
4.7	Stat	55
4.8	Position adjustments	59
4.9	Pulling it all together	59
4.9.1	Combining geoms and stats	60
4.9.2	Displaying precomputed statistics	62
4.9.3	Varying aesthetics and data	62
5	Toolbox	65
5.1	Introduction	65
5.2	Overall layering strategy	66
5.3	Basic plot types	66
5.4	Displaying distributions	68
5.5	Dealing with overplotting	72
5.6	Surface plots	77
5.7	Drawing maps	77
5.8	Revealing uncertainty	80
5.9	Statistical summaries	81
5.9.1	Individual summary functions	84
5.9.2	Single summary function	84
5.10	Annotating a plot	85
5.11	Weighted data	88
6	Scales, axes and legends	91
6.1	Introduction	91
6.2	How scales work	92
6.3	Usage	93
6.4	Scale details	96
6.4.1	Common arguments	96
6.4.2	Position scales	98
6.4.3	Colour	102
6.4.4	The manual discrete scale	107
6.4.5	The identity scale	109

6.5	Legends and axes	110
6.6	More resources	112
7	Positioning	115
7.1	Introduction	115
7.2	Faceting	115
7.2.1	Facet grid	116
7.2.2	Facet wrap	120
7.2.3	Controlling scales	121
7.2.4	Missing faceting variables	123
7.2.5	Grouping vs. faceting	123
7.2.6	Dodging vs. faceting	125
7.2.7	Continuous variables	129
7.3	Coordinate systems	129
7.3.1	Transformation	130
7.3.2	Statistics	133
7.3.3	Cartesian coordinate systems	133
7.3.4	Non-Cartesian coordinate systems	136
8	Polishing your plots for publication	139
8.1	Themes	139
8.1.1	Built-in themes	140
8.1.2	Theme elements and element functions	142
8.2	Customising scales and geoms	147
8.2.1	Scales	147
8.2.2	Geoms and stats	148
8.3	Saving your output	148
8.4	Multiple plots on the same page	151
8.4.1	Subplots	152
8.4.2	Rectangular grids	153
9	Manipulating data	157
9.1	An introduction to <code>plyr</code>	157
9.1.1	Fitting multiple models	161
9.2	Converting data from wide to long	164
9.2.1	Multiple time series	164
9.2.2	Parallel coordinates plot	167
9.3	<code>ggplot()</code> methods	169
9.3.1	Linear models	170
9.3.2	Writing your own	173
10	Reducing duplication	177
10.1	Introduction	177
10.2	Iteration	177
10.3	Plot templates	178
10.4	Plot functions	181

Appendices

A	Translating between different syntaxes	185
A.1	Introduction	185
A.2	Translating between qplot and ggplot	185
A.2.1	Aesthetics	185
A.2.2	Layers	186
A.2.3	Scales and axes	186
A.2.4	Plot options	187
A.3	Base graphics	187
A.3.1	High-level plotting commands	187
A.3.2	Low-level drawing	189
A.3.3	Legends, axes and grid lines	190
A.3.4	Colour palettes	190
A.3.5	Graphical parameters	190
A.4	Lattice graphics	190
A.5	GPL	192
B	Aesthetic specifications	195
B.1	Colour	195
B.2	Line type	195
B.3	Shape	196
B.4	Size	196
B.5	Justification	196
C	Manipulating plot rendering with grid	199
C.1	Introduction	199
C.2	Plot viewports	199
C.3	Plot grobs	201
C.4	Saving your work	201
	References	203
	Index	207
	R code index	211