

Thomas Rahlf

Datendesign mit R

100 Visualisierungsbeispiele

1. Auflage

Open Source Press

Alle in diesem Buch enthaltenen Programme, Darstellungen und Informationen wurden nach bestem Wissen erstellt. Dennoch sind Fehler nicht ganz auszuschließen. Aus diesem Grunde sind die in dem vorliegenden Buch enthaltenen Informationen mit keiner Verpflichtung oder Garantie irgendeiner Art verbunden. Autor(en), Herausgeber, Übersetzer und Verlag übernehmen infolgedessen keine Verantwortung und werden keine daraus folgende Haftung übernehmen, die auf irgendeine Art aus der Benutzung dieser Informationen – oder Teilen davon – entsteht, auch nicht für die Verletzung von Patentrechten, die daraus resultieren können. Ebenso wenig übernehmen Autor(en) und Verlag die Gewähr dafür, dass die beschriebenen Verfahren usw. frei von Schutzrechten Dritter sind.

Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. werden ohne Gewährleistung der freien Verwendbarkeit benutzt und können auch ohne besondere Kennzeichnung eingetragene Marken oder Warenzeichen sein und als solche den gesetzlichen Bestimmungen unterliegen.

Dieses Werk ist urheberrechtlich geschützt. Alle Rechte, auch die der Übersetzung, des Nachdrucks und der Vervielfältigung des Buches – oder Teilen daraus – vorbehalten. Kein Teil des Werkes darf ohne schriftliche Genehmigung des Verlags in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder einem anderen Verfahren), auch nicht für Zwecke der Unterrichtsgestaltung, reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Copyright © 2014 Open Source Press, München
Gesamtlektorat: Dr. Markus Wirtz
Satz: Open Source Press & Thomas Schraitle
Umschlaggestaltung: Olga Saborov, Open Source Press
Gesamtherstellung: Kösel, Krugzell

ISBN: 9783955390945 (gedruckte Ausgabe)

<http://www.opensourcepress.de>

Inhaltsverzeichnis

Vorwort	11
1 Daten für alle	15
1.1 Datenvisualisierung zwischen Wissenschaft und Journalismus.	15
1.2 Warum R?	18
1.3 Das Konzept des Datendesigns.	20
I Grundlagen und Technik	23
2 Aufbau und technische Voraussetzungen	25
2.1 Begriffe und Elemente.	25
2.2 Gestaltungsraster.	26
2.3 Perzeption.	29
2.4 Schriften.	33
2.4.1 Fonts.	35
2.4.2 Freie Schriften.	37
2.5 Symbole.	38
2.5.1 Symbolfonts.	41
2.5.2 Symbole im SVG-Format.	41
2.6 Farbe.	43
2.6.1 Farbmodelle.	43
2.6.2 Farbe in statistischen Abbildungen.	45

3 Umsetzung in R	49
3.1 Installation	50
3.2 Grundkonzepte in R	50
3.2.1 Datenstrukturen	51
3.2.2 Import von Daten	55
3.3 Grafikkonzepte in R	62
3.3.1 Paper-Pencil-Prinzip des Base Graphics System: High-Level- und Low-Level-Funktionen	67
3.3.2 Einstellung von Grafikparametern	70
3.3.3 Randeinstellungen von Abbildungen und Grafiken	77
3.3.4 Mehrfachgrafiken: Panels mit mfrow und mfcop	79
3.3.5 Komplexere Anordnungen mit layout	80
3.3.6 Schrifteinbindung	82
3.3.7 Ausgabe mit cairo_pdf	83
3.3.8 Unicode in Abbildungen	85
3.3.9 Farbeinstellungen	88
3.4 R-Pakete und -Funktionen in diesem Buch	90
3.4.1 Pakete	91
3.4.2 Funktionen	95
3.4.3 Schematische Vorgehensweise	105
4 Über R hinaus	107
4.1 Ergänzungen mit LaTeX	107
4.2 Manuelle Nachbearbeitung und Symbolfonterstellung in Inkscape	113
4.2.1 Nachbearbeitung	113
4.2.2 Symbolfonterstellung	114
5 Zu den Beispielen	119
5.1 Versuch einer Systematik	119
5.2 Die Skripte zum Laufen bringen	121

II Beispiele	123
6 Kategoriale Daten	125
6.1 Balken- und Säulendiagramme.	125
6.1.1 Balkendiagramm einfach.	126
6.1.2 Balkendiagramm für Mehrfachantworten – die ersten beiden Antwortkategorien.	131
6.1.3 Balkendiagramm für Mehrfachantworten – alle Antwortkategorien.	135
6.1.4 Balkendiagramm für Mehrfachantworten – alle Antwortkategorien, Variante.	137
6.1.5 Balkendiagramm für Mehrfachantworten – alle Antwortkategorien (Panel).	140
6.1.6 Balkendiagramm für Mehrfachantworten – Symbole für Individuen.	142
6.1.7 Balkendiagramm für Mehrfachantworten – alle Antwortkategorien, gruppiert.	145
6.1.8 Säulendiagramm mit zweizeiliger Beschriftung.	150
6.1.9 Säulendiagramm mit 45-Grad-Beschriftung.	152
6.1.10 Profildiagramm für Mehrfachantworten – Mittelwerte der Antworten.	153
6.1.11 Dotchart für drei Variablen.	156
6.1.12 Säulendiagramm mit Anteilen.	159
6.2 Kreis- und Radialdiagramme.	161
6.2.1 Einfaches Kreisdiagramm.	162
6.2.2 Kreisdiagramme, Beschriftung innen (Panel).	163
6.2.3 Sitzverteilung (Panel).	165
6.2.4 Spie Chart.	168
6.2.5 Radialpolygone (Panel).	171
6.2.6 Radialpolygone (Panel) – andere Spaltenanordnung.	173
6.2.7 Radialpolygone übereinander.	174
6.3 Grafiktabelle.	175

6.3.1	Vereinfachtes Gantt-Diagramm.	177
6.3.2	Vereinfachtes Gantt-Diagramm – Farben nach Personen.	181
6.3.3	Bumpchart.	182
6.3.4	Heatmap.	185
6.3.5	Mosaikplot (Panel).	188
6.3.6	Ballonplot.	190
6.3.7	Treemap.	192
6.3.8	Treemaps für zwei Ebenen (Panel).	194
7	Verteilungen	199
7.1	Histogramme und Boxplots.	199
7.1.1	Histogramme übereinander.	199
7.1.2	Säulendiagramme mit Colorbrewer gefärbt (Panel).	201
7.1.3	Histogramme (Panel).	204
7.1.4	Boxplots für Gruppen – absteigend sortiert.	207
7.1.5	Boxplots für Gruppen – absteigend sortiert, Vergleich zweier Erhebungen.	211
7.2	(Bevölkerungs-)Pyramiden.	215
7.2.1	Pyramide mit mehreren Farben.	216
7.2.2	Pyramiden – Betonung der äußeren Bereiche (Panel).	218
7.2.3	Pyramiden – Betonung der inneren Bereiche (Panel).	221
7.2.4	Pyramiden mit eingezeichneter Linie (Panel).	224
7.2.5	Pyramide mit Zusammenfassungen.	225
7.2.6	Balkendiagramme als Pyramiden (Panel).	228
7.3	Ungleichheit.	231
7.3.1	Einfache Lorenzkurve.	231
7.3.2	Lorenzkurven übereinander.	233
7.3.3	Lorenzkurven (Panel).	236

7.3.4	Vergleich von Einkommensanteilen mit Balkendiagramm (Quintile)	238
7.3.5	Vergleich von Einkommensanteilen mit Balkendiagramm (Dezile)	240
7.3.6	Vergleich von Einkommensanteilen mit Panel-Balkendiagramm (Quintile)	242
8	Zeitreihen	245
8.1	Kurze Zeitreihen.	245
8.1.1	Säulendiagramm für Entwicklungen.	245
8.1.2	Säulendiagramm mit Anteilen für Wachstumsentwicklungen.	248
8.1.3	Quartalswerte als Säulen.	251
8.1.4	Quartalswerte als Linien mit Werte-Beschriftungen.	253
8.1.5	Kurze Zeitreihen übereinander.	255
8.2	Flächen unter und zwischen Zeitreihen.	257
8.2.1	Flächen zwischen zwei Zeitreihen.	257
8.2.2	Fläche als Korridor mit Zeitreihen (Panel).	259
8.2.3	Prognoseintervalle (Panel).	262
8.2.4	Prognoseintervalle Index (Panel).	265
8.2.5	Zeitreihen mit gestapelten Flächen.	267
8.2.6	Flächen unterhalb einer Zeitreihe.	269
8.2.7	Zeitreihen mit Trend (Panel).	272
8.3	Darstellung von Tages-, Wochen- und Monatswerten.	275
8.3.1	Tageswerte mit Beschriftungen.	275
8.3.2	Tageswerte mit Beschriftungen und Wochensymbolen (Panel).	277
8.3.3	Tageswerte mit Monatsbeschriftung.	282
8.3.4	Zeitreihen aus Wochenwerten (Panel).	283
8.3.5	Monatswerte (Panel).	286
8.3.6	Monatswerte mit Monatsbeschriftung.	288
8.3.7	Monatswerte mit Monatsbeschriftung (Layout).	291

8.4	Sonderfälle und Spezielles.	294
8.4.1	Zeitreihen als Streudiagramm (Panel).	294
8.4.2	Zeitreihen mit fehlenden Werten.	296
8.4.3	Saisonspannweiten (Panel).	299
8.4.4	Saisonspannweiten übereinander.	301
8.4.5	Saisonfigur (Seasonal Subseries Plot) mit Datentabelle.	303
8.4.6	Zeitliche Spannweiten.	306
9	Streudiagramme	309
9.1	Varianten.	311
9.1.1	Streudiagramm Variante 1: Vier Quadranten farblich unterschieden.	311
9.1.2	Streudiagramm Variante 2: Ausreißer farblich hervorgehoben.	314
9.1.3	Streudiagramm Variante 3: Bereiche farblich hervorgehoben.	317
9.1.4	Streudiagramm Variante 4: Eingezeichnete Ellipse.	319
9.1.5	Streudiagramm Variante 5: Verbundene Punkte.	321
9.2	Sonderfälle und Spezielles.	324
9.2.1	Streudiagramm mit wenigen Punkten.	324
9.2.2	Streudiagramm mit selbst definierten Symbolen.	326
9.2.3	Karte von Deutschland als Streudiagramm.	329
10	Karten	333
10.1	Einführende Beispiele.	333
10.1.1	Karten von Deutschland: Ortsnetzbereiche und Postleitzahlengebiete.	333
10.1.2	Gefilterte Postleitzahlenkarte.	335
10.1.3	Europakarte Nuts 2006 (Ausschnitt).	338
10.2	Punkte, Diagramme und Symbole in Karten.	339
10.2.1	Karte von Deutschland mit ausgewählten Orten und Umriss (Panel).	339

10.2.2	Karte von Deutschland mit ausgewählten Orten (Kreisdiagramme) und Umriss.	342
10.2.3	Karte von Deutschland mit ausgewählten Orten (Säulen) und Umriss.	345
10.2.4	Karte von Deutschland als dreidimensionales Streudiagramm.	348
10.2.5	Karte von Nordrhein-Westfalen mit ausgewählten Orten (Symbole) und Umriss.	351
10.2.6	Karte von Tunesien mit selbst definierten Symbolen.	354
10.3	Choroplethenkarten.	357
10.3.1	Choroplethenkarte von Deutschland auf Kreisebene.	357
10.3.2	Choroplethenkarte von Deutschland auf Kreisebene (Panel).	360
10.3.3	Choroplethenkarte von Europa auf Länderebene.	364
10.3.4	Choroplethenkarte von Europa auf Länderebene (Panel).	367
10.3.5	Weltchoroplethenkarte: Regionen.	370
10.4	Sonderfälle und Spezielles.	372
10.4.1	Weltkarte mit Orthodromen.	373
10.4.2	Stadtkarten mit OpenStreetMap-Daten (Panel).	375
11	Illustratives	381
11.1	Tabelle mit Symbolen der Schrift „Symbol Signs“.	381
11.2	Radialsäulendiagramme mit Beschriftung (Panel).	383
11.3	Radialsäulendiagramme ohne Beschriftung (Panel).	389
11.4	Radialsäulendiagramm (Poster).	392
11.5	Nacht-Karte von Deutschland als Streudiagramm.	396
11.6	Streudiagramm Gapminder.	398
11.7	Karte von Napoleons Rußlandfeldzug von 1812/13 von Charles Joseph Minard, 1869.	403

III Anhang	407
A Verwendete Daten	409
A.1 ZA2391: Politbarometer 1977-2011 (Partielle Kumulation). . . .	409
A.2 ZA4753: European Values Study 2008: Germany (EVS 2008).	410
A.3 ZA4804: European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008).	410
A.4 ZA4972: Eurobarometer 71.2 (May-Jun 2009).	410
A.5 BetterLifeIndex_Data_2011V6.xls.	411
A.6 weltenergiemix.xlsx.	411
A.7 personen.xlsx.	412
A.8 v_frauen_maenner.	412
A.9 gadm.org.	412
A.10 NUTS-Karten.	413
B Literatur	415
B.1 Bücher und Artikel.	415
B.2 Websites zum Einstieg.	419
Index	421