



Thomas Rahlf

# Datenvisualisierung mit R

111 Beispiele

*2. Auflage*

**EXTRAS ONLINE**



Springer Spektrum

# Datenvisualisierung mit R

Thomas Rahlf

# Datenvisualisierung mit R

111 Beispiele

2., erweiterte Auflage

 Springer Spektrum

Thomas Rahlf  
Rheinische Friedrich-Wilhelms-Universität  
Bonn  
Bonn, Deutschland

Ergänzendes Material finden Sie auf  
<http://www.springer.com/de/book/9783662548196>.  
oder auf  
<http://extras.springer.com>.

ISBN 978-3-662-54819-6 ISBN 978-3-662-54820-2 (eBook)  
<https://doi.org/10.1007/978-3-662-54820-2>

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

Springer Spektrum

Ursprünglich erschienen bei Open Source Press, München, 2014

© Springer-Verlag GmbH Deutschland 2018

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung: Iris Ruhmann

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Spektrum ist Teil von Springer Nature

Die eingetragene Gesellschaft ist Springer-Verlag GmbH Deutschland

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

# Vorwort zur zweiten Auflage

Die Rückmeldungen zu der vergriffenen, 2014 im open source press Verlag publizierten ersten Auflage sowie zu der zwischenzeitlich erschienenen englischen Ausgabe waren so erfreulich, dass ich dem Angebot des Springer Verlages, das Buch in aktualisierter Neuauflage wieder verfügbar zu machen, gerne nachgekommen bin. Das Konzept der Erläuterung vollständiger Beispiele sowie die Beschränkung auf Base Graphics wurden beibehalten. Gegenüber der ersten Auflage wurden 11 Beispiele neu aufgenommen, so dass der Untertitel nun „111 Beispiele“ lautet. Der Haupttitel wurde der englischen Ausgabe angepasst. Zwei wesentliche Ergänzungen haben sich ergeben. Zum einen wurde das Kapitel zu kategorialen Daten um einen Abschnitt zur Visualisierung von Netzwerkbeziehungen erweitert. Neben Beispielen für klassische Netzwerkdiagramme, einer angepassten Heatmap und einem multiplen Balkendiagramm findet man hier auch ein Chord-Diagramm und einen Riverplot. Auf den ersten Blick ungewöhnlich erscheinend, haben sie mittlerweile aber auch Eingang in Publikationen von angesehenen Wissenschaftszeitungen wie *Science*, *Nature* oder *Cell* gefunden.

Das Kapitel zu Karten wurde um ein Beispiel zur Verwendung von georeferenzierten Rasterformaten sowie eines zu Cartograms erweitert. Weiterhin wurden drei Beispiele für die Einbindung von mit R erstellten Daten in interaktive JavaScript-Abbildungen ergänzt. R stellt mittlerweile mehrere Konzepte bzw. Pakete bereit, mit denen mehr oder weniger direkt JavaScript-Visualisierungen erzeugt werden können. Solche Pakete bilden letztendlich eine Art Container in R, bei dem eine jeweils spezifische und eigens entwickelte Syntax die so geschriebenen Skripte in die notwendige Notation der JavaScript-Bibliothek übersetzt. Man ist dadurch auf den Sprachumfang des R-Paketes sowie die Qualität und Flexibilität der Übersetzungsroutinen angewiesen. Ich bin nicht sicher, ob das der richtige Weg ist. Hier wurde ein anderer gewählt. In Kapitel 12 werden in drei Beispielen die Daten mit R so aufbereitet, dass sie in vorhandenen, nur leicht angepassten JavaScript-Code eingebunden werden. Dafür werden Highcharts und Mapael verwendet, zwei JavaScript-Bibliotheken, mit denen bei geringem Änderungsaufwand „out of the box“ sehr ästhetische Abbildungen erzeugt werden können.

Vorrangiges Ziel des Buches ist weiterhin die Erläuterung der Erstellung von Präsentationsgrafiken. Für die explorative Visualisierung im Rahmen der Datenanalyse sei auf das nunmehr erschienene Buch von Antony Unwin „Graphical Data Analysis with R“ (CRC Press, 2015) verwiesen.

Mein erster Dank geht an Agnes Herrmann und Iris Ruhmann vom Springer-Verlag, die die vorliegende Auflage ermöglicht haben. Für hilfreiche Anregungen, Hinweise und Gedankenaustausche zu dieser Auflage danke ich darüber hinaus Alberto Cairo, Martin S. Fischer, Sebastian Jeworutzki, Nikola Sander, Antony Unwin, January Weiner und wiederum Stefan Fichtel.

January Weiner hat freundlicherweise eine Anregung in seinem Paket riverplot aufgenommen, die Beispiel 6.4.4 zugute kommt.

Bonn, Juni 2017

Thomas Rahlf

# Vorwort zur ersten Auflage

Als ich vor fast zwanzig Jahren eine Reihe von Büchern zur statistischen Grafik und grafisch gestützten Datenanalyse rezensiert habe, war alles noch ganz anders: Formate waren proprietär, Betriebssysteme und ihre Zeichensätze inkompatibel, Grafik- und Statistiksoftware teuer. Seit der Jahrtausendwende änderte sich die Lage grundlegend: Das Internet war den Kinderschuhen entwachsen, Open-Source-Projekte gewannen unter diesem neuen Begriff immer mehr Anhänger und eine Handvoll Enthusiasten stellte Version 1.0 der freien Statistik-Programmiersprache R zur Verfügung. Viele Entwickler ließen sich für eine Mitarbeit an diesem Projekt begeistern. 2013 hat R die Versionsnummer 3 erreicht, neben der Basis-Software gibt es aktuell über 4000 frei verfügbare Erweiterungspakete. Firmen und Organisationen wie Google, Facebook oder die CIA verwenden R zur Datenanalyse. Als besondere Stärke werden immer wieder die Grafikfähigkeiten hervorgehoben. Praktisch alle für die Datenvisualisierung relevanten Technologien werden zeitnah in R integriert. Man kann mit zahlreichen Funktionen jede nur erdenkliche Abbildung detailgenau konstruieren, Karten erstellen und vieles mehr.

Man muss nur wissen, wie – und dazu möchte dieses Buch einen Beitrag leisten.

*Was dieses Buch sein möchte – und was nicht* Das vorliegende Buch ist *keine* Einführung, die systematisch die Grafikwerkzeuge von R erläutert. Es möchte vielmehr anhand von 100 vollständigen Skript-Beispielen in die Grundlagen der Gestaltung von Präsentationsgrafiken einführen und zeigen, wie Balken- und Säulendiagramme, Bevölkerungspyramiden, Lorenzkurven, Boxplots, Streudiagramme, Zeitreihendarstellungen, Radialpolygone, Gantt-Diagramme, Heatmaps, Bumpcharts, Mosaik- und Ballonplots sowie eine Reihe verschiedener thematischer Kartentypen mit dem *Base Graphics System* von R erstellt werden.<sup>1</sup> Jedes Beispiel verwendet reale Daten und erläutert die Abbildung und deren Programmierung Schritt für Schritt. Die Auswahl orientiert sich an meinem persönlichen Erfahrungsschatz – sicher wird der ein oder andere die eine oder andere Abbil-

---

<sup>1</sup> Für die anderen in R verfügbaren Grafikansätze wie `grid` und, darauf aufbauend, `lattice` und `ggplot2` sei auf bereits vorhandene Einführungen verwiesen.

derung vermissen, anderes als zu ausführlich empfinden. Dennoch sollte ein großer Anwendungsbereich abgedeckt sein.

*Das Buch richtet sich an R-Kenner:* Für Sie sind insbesondere die Beispiele nützlich, besonders der Code. Teil I können Sie vermutlich überspringen.

Leserinnen und Leser, die von R schon gehört, es auch schon einmal ausprobiert und keine Angst vor dem Programmieren haben; Sie profitieren von beiden Teilen.

Anfänger: Ihnen helfen vor allem die fertigen und hier abgebildeten Grafiken. Sie sehen, was mit R möglich ist. Oder anders gesagt: Sie sehen, dass es R überhaupt gibt und dass sich damit Grafiken erzeugen lassen, die Sie schon lange einmal erzeugen wollten – Sie wussten bloß nicht, wie. Der Code ist Ihnen zwar zu kompliziert, aber Sie können evtl. andere damit beauftragen, Grafiken für Sie in R zu programmieren.

*Windows, Mac und Linux* Alle Skripte und Bearbeitungsschritte führen unter Windows, Mac OS X und Linux zu identischen Ergebnissen. Die Beispiele wurden mit Mac OS X erstellt und anschließend unter Ubuntu 12.04 sowie einer Evaluierungskopie von Windows 8.1 getestet.

## Danksagung

Für Hinweise, Kommentare, Rückmeldungen, Daten, Diskussionsgelegenheiten oder Hilfe danke ich Gregor Aisch, Insa Bechert, Evelyn Brislinger, Giuseppe Casalicchio, Arnulf Christl, Katja Diederichs, Günter Faes, Mira Hassan, Mark Heckmann, Daniel Hienert, Bruno Hopp, Duncan Temple Lang, Uwe Ligges, Lorenz Matzat, Meinhard Moschner, Stefan Müller, Paul Murrell, David Phillips, Martijn Tennekes, Patrick R. Schmid, Thomas Schraitle, Valentin Schröder, Torsten Steiner, Michael Terwey, Katrin Weller, Bernd Weiss, Nils Windisch, Benjamin Zapilko und Lisa Zhang.

Ganz besonders hat das Manuskript vom Austausch mit einem Infografiker und einem Datenjournalisten profitiert. Stefan Fichtel hat alle Abbildungen angesehen und kritisch kommentiert. Für einzelne Abbildungen hat er eigene Vorschläge gestaltet. Das war mir eine unschätzbare Hilfe. Nicht in allen Fällen waren wir einer Meinung, und an der ein oder anderen Stelle habe ich mich über seinen Rat hinweggesetzt. Verbliebene Fehler und Unzulänglichkeiten gehen daher zu meinen Lasten.

Björn Schwentker hat sich die Mühe gemacht und große Teile des Manuskripts gründlich gegengelesen. Ihm verdanke ich wertvolle Hinweise, die den Text an einigen Stellen mit Sicherheit klarer und lesbarer gemacht haben.

Schließlich gilt mein Dank Markus Wirtz, dass er das Experiment gewagt hat, aus all dem am Ende doch ein Buch zu drucken.



## **Im Internet**

Die Abbildungen sind für unterschiedlichste Endausgaben konzipiert. Insbesondere bei den Karten und Radialsäulendiagrammen ist manches aufgrund des Buchformates grenzwertig klein. Gerade für solche Fälle sei auf die Website des Buches verwiesen, auf der alle Abbildungen in hoher Auflösung bzw. als Vektorgrafik im PDF-Format bereitgestellt werden: <http://www.datenvisualisierung-r.de>

# Inhaltsverzeichnis

|          |  |   |
|----------|--|---|
| <b>1</b> | <b>Daten für alle</b> . . . . .                                  | 1 |
| 1.1      | Datenvisualisierung zwischen Wissenschaft und Journalismus . . . | 1 |
| 1.2      | Warum R? . . . . .   | 3 |
| 1.3      | Das Konzept des Datendesigns . . . . .                           | 4 |

## Teil I Grundlagen und Technik

|          |  |    |
|----------|--|----|
| <b>2</b> | <b>Aufbau und technische Voraussetzungen</b> . . . . .   | 9  |
| 2.1      | Begriffe und Elemente . . . . .  | 9  |
| 2.2      | Gestaltungsraster . . . . .  | 9  |
| 2.3      | Perzeption . . . . .   | 12 |
| 2.4      | Schriften . . . . .  | 16 |
| 2.4.1    | Fonts . . . . .  | 18 |
| 2.4.2    | Freie Schriften . . . . .  | 19 |
| 2.5      | Symbole . . . . .  | 20 |
| 2.5.1    | Symbolfonts . . . . .  | 22 |
| 2.5.2    | Symbole im SVG-Format . . . . .  | 24 |
| 2.6      | Farbe . . . . .  | 24 |
| 2.6.1    | Farbmodelle . . . . .  | 24 |
| 2.6.2    | Farbe in statistischen Abbildungen . . . . .   | 26 |
| <b>3</b> | <b>Umsetzung in R</b> . . . . .  | 29 |
| 3.1      | Installation . . . . .   | 29 |
| 3.2      | Grundkonzepte in R . . . . .   | 30 |
| 3.2.1    | Datenstrukturen . . . . .  | 31 |
| 3.2.2    | Import von Daten . . . . .   | 35 |
| 3.3      | Grafikkonzepte in R . . . . .  | 43 |
| 3.3.1    | Paper-Pencil-Prinzip des Base Graphics System:<br>High-Level- und Low-Level-Funktionen . . . . . | 48 |
| 3.3.2    | Einstellung von Grafikparametern . . . . .   | 51 |

|          |  |     |
|----------|--|-----|
| 3.3.3    | Randeinstellungen von Abbildungen und Grafiken . . . . .                         | 58  |
| 3.3.4    | Mehrfachgrafiken: Panels mit <code>mfrow</code> und <code>mfcop</code> . . . . . | 59  |
| 3.3.5    | Komplexere Anordnungen mit <code>layout</code> . . . . .                         | 60  |
| 3.3.6    | Schrifteinbindung . . . . .  | 63  |
| 3.3.7    | Ausgabe mit <code>cairo_pdf</code> . . . . .                                     | 65  |
| 3.3.8    | Unicode in Abbildungen . . . . .   | 66  |
| 3.3.9    | Farbeinstellungen . . . . .  | 70  |
| 3.4      | R-Pakete und -Funktionen in diesem Buch . . . . .                                | 72  |
| 3.4.1    | Pakete . . . . .   | 73  |
| 3.4.2    | Funktionen . . . . .   | 77  |
| 3.4.3    | Schematische Vorgehensweise . . . . .  | 85  |
| <b>4</b> | <b>Über R hinaus</b> . . . . .   | 87  |
| 4.1      | Ergänzungen mit LaTeX . . . . .  | 87  |
| 4.2      | Manuelle Nachbearbeitung und Symbolfonterstellung in Inkscape . . . . .          | 92  |
| 4.2.1    | Nachbearbeitung . . . . .  | 92  |
| 4.2.2    | Symbolfonterstellung . . . . .   | 94  |
| <b>5</b> | <b>Zu den Beispielen</b> . . . . .   | 99  |
| 5.1      | Versuch einer Systematik . . . . .   | 99  |
| 5.2      | Die Skripte zum Laufen bringen . . . . .   | 101 |

## Teil II Beispiele

|          |   |     |
|----------|---|-----|
| <b>6</b> | <b>Kategoriale Daten</b> . . . . .  | 105 |
| 6.1      | Balken- und Säulendiagramme . . . . .   | 105 |
| 6.1.1    | Balkendiagramm einfach . . . . .  | 106 |
| 6.1.2    | Balkendiagramm für Mehrfachantworten –<br>die ersten beiden Antwortkategorien . . . . . | 111 |
| 6.1.3    | Balkendiagramm für Mehrfachantworten –<br>alle Antwortkategorien . . . . .              | 116 |
| 6.1.4    | Balkendiagramm für Mehrfachantworten –<br>alle Antwortkategorien, Variante . . . . .    | 119 |
| 6.1.5    | Balkendiagramm für Mehrfachantworten –<br>alle Antwortkategorien (Panel) . . . . .      | 121 |
| 6.1.6    | Balkendiagramm für Mehrfachantworten –<br>Symbole für Individuen . . . . .              | 124 |
| 6.1.7    | Balkendiagramm für Mehrfachantworten –<br>alle Antwortkategorien, gruppiert . . . . .   | 127 |
| 6.1.8    | Säulendiagramm mit zweizeiliger Beschriftung . . . . .                                  | 133 |
| 6.1.9    | Säulendiagramm mit 45-Grad-Beschriftung . . . . .                                       | 135 |
| 6.1.10   | Profildiagramm für Mehrfachantworten –<br>Mittelwerte der Antworten . . . . .           | 136 |
| 6.1.11   | Dotchart für drei Variablen . . . . .   | 138 |

|          |   |            |
|----------|---|------------|
| 6.1.12   | Säulendiagramm mit Anteilen   | 142        |
| 6.2      | Kreis- und Radialdiagramme  | 145        |
| 6.2.1    | Einfaches Kreisdiagramm   | 146        |
| 6.2.2    | Kreisdiagramme, Beschriftung innen (Panel)                              | 148        |
| 6.2.3    | Sitzverteilung (Panel)  | 150        |
| 6.2.4    | Spie Chart  | 153        |
| 6.2.5    | Radialpolygone (Panel)  | 156        |
| 6.2.6    | Radialpolygone (Panel) – andere Spaltenanordnung                        | 158        |
| 6.2.7    | Radialpolygone übereinander   | 159        |
| 6.3      | Grafiktabellen  | 160        |
| 6.3.1    | Vereinfachtes Gantt-Diagramm  | 161        |
| 6.3.2    | Vereinfachtes Gantt-Diagramm – Farben nach Personen                     | 165        |
| 6.3.3    | Bumpchart   | 167        |
| 6.3.4    | Heatmap   | 170        |
| 6.3.5    | Mosaikplot (Panel)  | 173        |
| 6.3.6    | Ballonplot  | 175        |
| 6.3.7    | Treemap   | 177        |
| 6.3.8    | Treemaps für zwei Ebenen (Panel)  | 180        |
| 6.4      | Netzwerkbeziehungen   | 184        |
| 6.4.1    | Ungerichtetes Netzwerk  | 185        |
| 6.4.2    | Chord Diagram   | 190        |
| 6.4.3    | Gerichtetes Netzwerk  | 195        |
| 6.4.4    | Riverplot   | 198        |
| 6.4.5    | Heatmap für Beziehungen   | 201        |
| 6.4.6    | Multiple Barchart   | 204        |
| <b>7</b> | <b>Verteilungen</b>   | <b>207</b> |
| 7.1      | Histogramme und Boxplots  | 207        |
| 7.1.1    | Histogramme übereinander  | 207        |
| 7.1.2    | Säulendiagramme mit Colorbrewer gefärbt (Panel)                         | 209        |
| 7.1.3    | Histogramme (Panel)   | 213        |
| 7.1.4    | Boxplots für Gruppen – absteigend sortiert                              | 215        |
| 7.1.5    | Boxplots für Gruppen – absteigend sortiert, Vergleich zweier Erhebungen | 219        |
| 7.2      | (Bevölkerungs-)Pyramiden  | 223        |
| 7.2.1    | Pyramide mit mehreren Farben  | 225        |
| 7.2.2    | Pyramiden – Betonung der äußeren Bereiche (Panel)                       | 228        |
| 7.2.3    | Pyramiden – Betonung der inneren Bereiche (Panel)                       | 231        |
| 7.2.4    | Pyramiden mit eingezeichneter Linie (Panel)                             | 234        |
| 7.2.5    | Pyramide mit Zusammenfassungen  | 235        |
| 7.2.6    | Balkendiagramme als Pyramiden (Panel)                                   | 238        |
| 7.3      | Ungleichheit  | 241        |
| 7.3.1    | Einfache Lorenzkurve  | 242        |
| 7.3.2    | Lorenzkurven übereinander   | 244        |
| 7.3.3    | Lorenzkurven (Panel)  | 246        |

|          |  |     |
|----------|--|-----|
| 7.3.4    | Vergleich von Einkommensanteilen mit Balkendiagramm (Quintile) . . . . .       | 249 |
| 7.3.5    | Vergleich von Einkommensanteilen mit Balkendiagramm (Dezile) . . . . .         | 251 |
| 7.3.6    | Vergleich von Einkommensanteilen mit Panel-Balkendiagramm (Quintile) . . . . . | 253 |
| <b>8</b> | <b>Zeitreihen</b> . . . . .  | 257 |
| 8.1      | Kurze Zeitreihen . . . . .   | 257 |
| 8.1.1    | Säulendiagramm für Entwicklungen . . . . .                                     | 257 |
| 8.1.2    | Säulendiagramm mit Anteilen für Wachstumsentwicklungen . . . . .               | 260 |
| 8.1.3    | Quartalswerte als Säulen . . . . .   | 263 |
| 8.1.4    | Quartalswerte als Linien mit Werte-Beschriftungen . . . . .                    | 265 |
| 8.1.5    | Kurze Zeitreihen übereinander . . . . .  | 267 |
| 8.2      | Flächen unter und zwischen Zeitreihen . . . . .                                | 269 |
| 8.2.1    | Flächen zwischen zwei Zeitreihen . . . . .                                     | 269 |
| 8.2.2    | Fläche als Korridor mit Zeitreihen (Panel) . . . . .                           | 271 |
| 8.2.3    | Prognoseintervalle (Panel) . . . . .   | 274 |
| 8.2.4    | Prognoseintervalle Index (Panel) . . . . .                                     | 278 |
| 8.2.5    | Zeitreihen mit gestapelten Flächen . . . . .                                   | 280 |
| 8.2.6    | Flächen unterhalb einer Zeitreihe . . . . .                                    | 283 |
| 8.2.7    | Zeitreihen mit Trend (Panel) . . . . .   | 286 |
| 8.3      | Darstellung von Tages-, Wochen- und Monatswerten . . . . .                     | 289 |
| 8.3.1    | Tageswerte mit Beschriftungen . . . . .  | 289 |
| 8.3.2    | Tageswerte mit Beschriftungen und Wochensymbolen (Panel) . . . . .             | 291 |
| 8.3.3    | Tageswerte mit Monatsbeschriftung . . . . .                                    | 296 |
| 8.3.4    | Zeitreihen aus Wochenwerten (Panel) . . . . .                                  | 298 |
| 8.3.5    | Monatswerte (Panel) . . . . .  | 301 |
| 8.3.6    | Monatswerte mit Monatsbeschriftung . . . . .                                   | 303 |
| 8.3.7    | Monatswerte mit Monatsbeschriftung (Layout) . . . . .                          | 306 |
| 8.4      | Sonderfälle und Spezielles . . . . .   | 309 |
| 8.4.1    | Zeitreihen als Streudiagramm (Panel) . . . . .                                 | 309 |
| 8.4.2    | Zeitreihen mit fehlenden Werten . . . . .                                      | 311 |
| 8.4.3    | Saisonspannweiten (Panel) . . . . .  | 315 |
| 8.4.4    | Saisonspannweiten übereinander . . . . .                                       | 317 |
| 8.4.5    | Saisonfigur (Seasonal Subseries Plot) mit Datentabelle . . . . .               | 319 |
| 8.4.6    | Zeitliche Spannweiten . . . . .  | 323 |
| <b>9</b> | <b>Streudiagramme</b> . . . . .  | 325 |
| 9.1      | Varianten . . . . .  | 327 |
| 9.1.1    | Streudiagramm Variante 1: Vier Quadranten farblich unterschieden . . . . .     | 327 |
| 9.1.2    | Streudiagramm Variante 2: Ausreißer farblich hervorgehoben . . . . .           | 330 |

|           |  |            |
|-----------|--|------------|
| 9.1.3     | Streudiagramm Variante 3: Bereiche farblich hervorgehoben                    | 333        |
| 9.1.4     | Streudiagramm Variante 4: Eingezeichnete Ellipse                             | 335        |
| 9.1.5     | Streudiagramm Variante 5: Verbundene Punkte                                  | 338        |
| 9.2       | Sonderfälle und Spezielles   | 340        |
| 9.2.1     | Streudiagramm mit wenigen Punkten  | 340        |
| 9.2.2     | Streudiagramm mit selbst definierten Symbolen                                | 343        |
| 9.2.3     | Karte von Deutschland als Streudiagramm                                      | 346        |
| <b>10</b> | <b>Karten</b>  | <b>349</b> |
| 10.1      | Einführende Beispiele  | 349        |
| 10.1.1    | Karten von Deutschland: Ortsnetzbereiche<br>und Postleitzahlengebiete        | 349        |
| 10.1.2    | Gefilterte Postleitzahlenkarte   | 352        |
| 10.1.3    | Europakarte Nuts 2006 (Ausschnitt)   | 354        |
| 10.2      | Punkte, Diagramme und Symbole in Karten                                      | 356        |
| 10.2.1    | Karte von Deutschland mit ausgewählten Orten und Umriss<br>(Panel)           | 356        |
| 10.2.2    | Karte von Deutschland mit ausgewählten Orten<br>(Kreisdiagramme) und Umriss  | 359        |
| 10.2.3    | Karte von Deutschland mit ausgewählten Orten (Säulen)<br>und Umriss          | 362        |
| 10.2.4    | Karte von Deutschland als dreidimensionales<br>Streudiagramm                 | 365        |
| 10.2.5    | Karte von Nordrhein-Westfalen mit ausgewählten Orten<br>(Symbole) und Umriss | 369        |
| 10.2.6    | Karte von Tunesien mit selbst definierten Symbolen                           | 371        |
| 10.3      | Choroplethenkarten   | 374        |
| 10.3.1    | Choroplethenkarte von Deutschland auf Kreisebene                             | 375        |
| 10.3.2    | Choroplethenkarte von Deutschland auf Kreisebene (Panel)                     | 377        |
| 10.3.3    | Choroplethenkarte von Europa auf Länderebene                                 | 383        |
| 10.3.4    | Choroplethenkarte von Europa auf Länderebene (Panel)                         | 386        |
| 10.3.5    | Weltchoroplethenkarte: Regionen  | 389        |
| 10.4      | Sonderfälle und Spezielles   | 391        |
| 10.4.1    | Weltkarte mit Orthodromen  | 392        |
| 10.4.2    | Stadtkarten mit OpenStreetMap-Daten (Panel)                                  | 394        |
| 10.4.3    | Georeferenzierte Karte im Rasterformat                                       | 400        |
| 10.4.4    | Cartogram (Panel)  | 407        |
| <b>11</b> | <b>Illustratives</b>   | <b>411</b> |
| 11.1      | Tabelle mit Symbolen der Schrift „Symbol Signs“                              | 411        |
| 11.2      | Radialsäulendiagramme mit Beschriftung (Panel)                               | 413        |
| 11.3      | Radialsäulendiagramme ohne Beschriftung (Panel)                              | 420        |
| 11.4      | Radialsäulendiagramm (Poster)  | 423        |
| 11.5      | Nacht-Karte von Deutschland als Streudiagramm                                | 428        |
| 11.6      | Streudiagramm Gapminder  | 430        |

- 11.7 Karte von Napoleons Rußlandfeldzug von 1812/13 von Charles Joseph Minard, 1869 . . . . . 435
- 12 Interaktive Visualisierung mit JavaScript: Highcharts und Mapael . 439**
  - 12.1 Streudiagramm in Highcharts . . . . . 441
  - 12.2 Zeitreihe in Highcharts . . . . . 449
  - 12.3 Choroplethenkarten mit Mapael . . . . . 456
    - 12.3.1 Installation der Javascript-Bibliotheken von Mapael . . . . . 458
- Anhang A. Verwendete Daten . . . . . 477**
- Literatur . . . . . 483**
- Sachverzeichnis . . . . . 487**

# Kapitel 1

## Daten für alle

### 1.1 Datenvisualisierung zwischen Wissenschaft und Journalismus

Art und Umfang von Daten, unsere Einstellung zu ihnen sowie ihre Verfügbarkeit haben sich in den vergangenen Jahren grundlegend gewandelt. Noch nie gab es so viele Daten wie heute. Noch nie waren sie so leicht verfügbar. Und noch nie waren die Möglichkeiten der Analyse, Aufbereitung und Präsentation größer.

Manche Wissenschaftler, wie etwa der Mathematiker Stephen Wolfram, glauben, dass man den Prozess der Datenanalyse weitgehend automatisieren kann, und sprechen in diesem Zusammenhang sogar von einer Demokratisierung der Wissenschaft. Andere, wie Googles Chefökonom Hal Varian, meinen hingegen, dass dafür mehrere Fähigkeiten erlernt werden müssen und diese zukünftig zentrale Schlüsselqualifikationen darstellen: „The ability to take data – to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it’s going to be a hugely important skill in the next decades (...)“.<sup>1</sup>

In den letzten Jahren ist eine Fülle von Websites, Büchern und anderen Publikationen entstanden, die sich der Visualisierung von Daten widmen. Dabei steht deren erzählende, nicht die explorative Visualisierung im Vordergrund. Eines der bekanntesten Beispiele ist die Mission von Hans Rosling, dem Autor und Erfinder von GAPMINDER, Statistiken zu weltweiten gesellschaftlichen Entwicklungen einem breiten Publikum eingängig zu veranschaulichen. Hans Rosling wurde 2012 vom *Time Magazine* zu den „100 Most Influential People in the World“ gezählt. Nahezu in Vergessenheit geratene Sozialwissenschaftler, die sich mit der didaktischen Visualisierung von statistischen, gesellschaftlichen Zusammenhängen befasst haben, allen voran Otto Neurath, werden wiederentdeckt.<sup>2</sup>

---

<sup>1</sup> [http://www.mckinsey.com/insights/innovation/hal\\_varian\\_on\\_how\\_the\\_web\\_challenges\\_managers](http://www.mckinsey.com/insights/innovation/hal_varian_on_how_the_web_challenges_managers).

<sup>2</sup> Eve, Matthew/Burke, Christopher (Hrsg.)/Otto Neurath (2010): *From Hieroglyphics to Isotype: A Visual Autobiography*. London: Hyphen Press.



Dabei ist es nicht so, dass das Rad neu erfunden wurde. In der Wissenschaft haben Datenvisualisierungen seit jeher und kontinuierlich eine wichtige Rolle gespielt. Bildgebende Verfahren gehören zum festen Bestandteil vieler Analysen in der Medizin, praktisch alle Naturwissenschaften nutzen bildliche Darstellungen von Daten zur visuellen Kommunikation von Ergebnissen. Die Zeitschrift *Nature* bietet Interessenten im Internet als Kaufanreiz für ihre Artikel neben einem Abstract kleine Voransichten der enthaltenden Abbildungen („Figures at a glance“).

Im Rahmen der statistischen Methodik haben eine Reihe von Wissenschaftlern schon vor vielen Jahren Grundlagenforschung zur statistischen Grafik betrieben: Bahnbrechend war neben den Arbeiten von William S. Cleveland das Buch von Edward Tufte, *The Visual Display of Quantitative Information*. Das Buch erschien 1983 und erlebte bereits in der ersten Auflage sechzehn Nachdrucke. Zusammen mit zwei in der Folge erschienenen Werken, *Envisioning Information* und *Visual Explanations*, hat Edward Tufte damit den Maßstab für das Thema auf eine sehr genuine Weise definiert.

Auch in der Wirtschaft gibt es eine lange Tradition der Präsentation von Daten. Seit vielen Jahren werden in Unternehmen für interne Zwecke nicht nur Daten gesammelt und ausgewertet, sondern auch in Abbildungen umgesetzt. Nach außen werden in besonders aufbereiteten Publikationen Präsentationsgrafiken in Geschäftsberichten möglichst eindrucksvoll zur Schau gestellt.<sup>3</sup>

Schließlich bemüht sich die amtliche Statistik seit vielen Jahren erfolgreich, ihre Ergebnisse nicht nur in tabellarischer Form bereitzustellen, sondern auch grafisch aufzubereiten. Hier kann man sowohl national als auch international eine nahezu von Jahr zu Jahr fortschreitende Tendenz zur stärkeren Visualisierung des offiziellen Datenmaterials feststellen.

Die Flut von Daten, die auf uns einströmt und uns ihre Auswertung aufdrängt, hat einen Nebeneffekt: Mit ihrer neuen, potentiellen Verfügbarkeit und Offenheit geht ein Umdenken in Bezug auf die Nutzungsrechte und Einsichtsmöglichkeiten einher. Zunehmend wird die Offenheit nicht nur von amtlichen, sondern auch von Unternehmensdaten gefordert. Umwelt- und Wetteraufzeichnungen, Verbrauchsdaten oder solche aus den Bereichen Gesundheit oder Bildung, Abstimmungen in Landtagen, Gesetzestexte, Daten zur Verkehrslage oder Fahrpläne sollen frei und offen zugänglich sein. Gegenüber den USA, Großbritannien oder auch der Schweiz hat Deutschland hier noch Nachholbedarf.<sup>4</sup>

Big Data und Open Data erfordern neue Methoden und neue Herangehensweisen. Eine innovative Variante, die sich die Bezeichnung *Data Science* zu eigen gemacht hat, versteht darunter eine Kombination aus Programmierfähigkeiten, mathematisch-statistischen Kenntnissen und substanzwissenschaftlicher Expertise. Drew Conway hat diese Kombination in Form eines Venn-Diagramms dargestellt, das uns auch sehr anschaulich die Schnittmengen veranschaulicht.

---

<sup>3</sup> In ihrer schönsten Form zusammengestellt bei Rädeker, Jochen/Dietz, Kirsten (2011): Reporting, Unternehmenskommunikation als Imageträger – ausgesuchte Finanz- und Nachhaltigkeitsberichte weltweit. Mainz: Hermann Schmidt Verlag.

<sup>4</sup> <https://index.okfn.org>.

Diese *Data Science* ist in aller Regel hochmathematisch und elaboriert. Aber auch der journalistische Bereich zeigt ein stark gewachsenes Interesse an Daten. Die allen voran von der *New York Times* und dem *Guardian*, in Deutschland von der *ZEIT* und anderen Medien angebotenen Recherchen und Visualisierungen sind unter dem Begriff *Datenjournalismus* im Aufwind.

So genannte Infografiken, häufig auch animiert und interaktiv, verbreiten sich geradezu explosionsartig im Internet. Seriöse und Maßstäbe setzende Angebote basieren dabei auf der Arbeit umfangreicher Experten-Teams und werden selbst Gegenstand der Forschung.

Daneben erfreuen sich individuelle Angebote von „Information Designern“ wie Catherine Mulbrandon, Stephen Few, Robert Kosara, Ben Fry oder Nathan Yau großer Beliebtheit, die eigene Datenvisualisierungssoftware entwickeln, Consulting-Firmen gründen, weltweit Workshops anbieten oder Blogs mit zigtausenden registrierten Nutzern aufbauen.<sup>5</sup>

Aus Sicht der eher „traditionellen“ statistischen Grafik schießt das eine oder andere dabei über das Ziel hinaus: So manches wird nicht nur als zu bunt, zu verspielt oder zu überladen empfunden, sondern auch als verwirrend oder gar verfälschend. Hier ist in jüngster Zeit eine Diskussion entstanden, von der am Ende sicher beide Seiten profitieren werden.<sup>6</sup>

## 1.2 Warum R?

In diesem Buch werden sämtliche Daten mit der freien Statistik-Software R visualisiert. Unter Wissenschaftlern ist die Programmiersprache inzwischen weit verbreitet und sehr beliebt. Doch jenseits der Forschung ist ihr Potenzial, maßgeschneiderte Grafiken zu produzieren, wenig bekannt. Das ist kein Wunder, denn Grafiker oder Journalisten tun sich mit dem Programmieren bekanntlich schwer. Es wäre sicherlich auch falsch zu behaupten, man lerne R so schnell, dass man in wenigen Minuten die erste ansprechende Grafik erstellt.

Andererseits: Der Einstieg ist leichter als in viele andere Programmiersprachen, weil R speziell für Daten und Statistik gemacht ist – und damit auch für deren Visualisierung. Es bietet einige Vorteile, die auch für Redakteure oder Datendesigner Gold wert sein können und die eine Software wie Excel nicht bietet:

- Alle Grafiken lassen sich im Vektorformat speichern (z.B. PDF, EPS oder SVG) und mit gängigen Vektorgrafikprogrammen wie Adobe Illustrator oder dem freien Inkscape sofort weiterverarbeiten, so dass jedes Grafikelement einzeln anpassbar ist.

---

<sup>5</sup> <http://visualizingeconomics.com>; <http://www.perceptualedge.com>; <http://kosara.net>; <http://benfry.com>; <http://flowingdata.com>.

<sup>6</sup> Gelman, Andrew/Unwin, Antony, *Infovis and Statistical Graphics: Different Goals, Different Looks*, in: *Journal of Computational and Graphical Statistics* 22/1 (2013), S. 2–28. Diskussionsbeiträge von Robert Kosara, Paul Murrell, Hadley Wickham S. 29–44, Antwort S. 45–49.

- Jedes Element der Grafik lässt sich durch R fast beliebig in Farbe oder Form verändern. Es lassen sich nach belieben Text, Symbole, Pfeile oder ganze Zeichnungen hinzufügen oder verschiedene Diagramme kombinieren.
- Die Grundformen der wichtigsten Diagrammtypen, wie Säulen-, Linien- oder Kreisdiagramme, lassen sich für einen ersten Eindruck oft schnell durch einen einzigen Befehl erzeugen.
- R beherrscht auch Karten und lässt so beliebige Geo-Visualisierungen zu. Das Kartenmaterial dafür kann zum Beispiel im geläufigen Format von Shape-Dateien eingeladen werden.
- Da Grafiken in R komplett programmiert sind, lässt sich jeder Schritt nachvollziehen, jeder Fehler finden und Änderungen sind leicht möglich. Dies ermöglicht auch eine Qualitätskontrolle durch Dritte und eine Offenlegung des Grafik-Sourcecodes im Sinne maximaler Open-Data-Transparenz.
- R ist kostenlos.
- R ist offen.
- R ist durch viele Programm-Module (Packages) erweiterbar, um besondere Grafiktypen darzustellen oder fortgeschrittene Datenanalysen vorzuschalten. Eine wachsende internationale Community stellt im Internet immer mehr Erweiterungen zur Verfügung.
- R-Grafiken können auch als Grundlage für interaktive Online-Grafiken dienen, indem beispielsweise den als SVG gespeicherten Diagrammelementen mit einem JavaScript-Paket wie D3.js interaktives Leben eingehaucht wird. Alternativ gibt es inzwischen ein komplettes JavaScript-Paket namens *Shiny*<sup>7</sup>, mit dem sich interaktive Datenanwendungen im Netz direkt in R schreiben lassen.

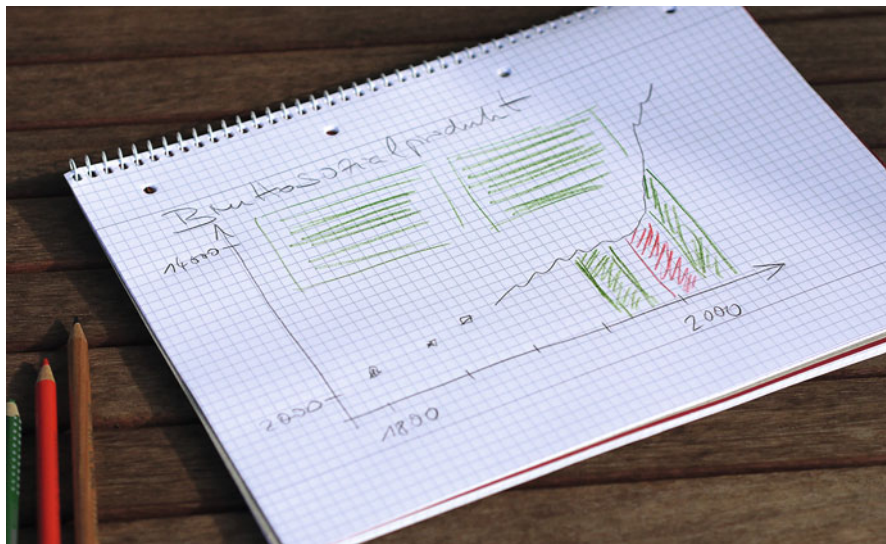
### 1.3 Das Konzept des Datendesigns

Das Buch verfolgt einen 100-Prozent-Ansatz: Alle Beispiele zeigen die vollständige Gestaltung einer konkreten Abbildung. Es wird immer vom Ergebnis ausgegangen: Die Ausgangsfragen waren jeweils: Wie muss eine bestimmte Grafik aussehen oder wie können vorhandene Daten am ehesten visualisiert werden? Dabei wurde unabhängig von einer konkreten Software stets mit einer Skizze begonnen. Erst der nächste Schritt bestand dann darin, sich nach den dafür benötigten Werkzeugen (Paketen und Funktionen) umzusehen und diese anzuwenden.

Die verwendeten Daten stammen ganz überwiegend aus der Sozialwissenschaft und der amtlichen Statistik, einige aus der Betriebswirtschaft, der Makroökonomie, der Politik, der Medizin, der Meteorologie oder den sozialen Medien. Mein Bestreben war, für alle ausgewählten Darstellungsformen geeignete Daten zu finden. Das ist sicher mal mehr, mal weniger gelungen. Die Daten wurden aber nicht „vorfriert“, sondern in der Form verwendet, in der sie zur Verfügung standen. Dadurch ist zwar der Skriptumfang manchmal etwas größer als unter Laborbedingungen mit

---

<sup>7</sup> <http://www.rstudio.com/shiny/>.



**Abb. 1.1** Skizze einer Abbildung

jeweils für die Aufgabe schon optimal aufbereiteten Daten. Andererseits ist das lebensnäher und kann Ihnen bei dem ein oder anderen Ihrer Daten-Fallstricke nützlich sein.

Alle Abbildungen sind als PDF-Datei konzipiert, so dass sie möglichst verlustfrei und flexibel weiterzuverwenden sind.

Im Durchschnitt waren für die Erstellung des Ergebnisses 40 Zeilen Code nötig. Von der ersten Idee bis zur fertigen Umsetzung verging pro Abbildung in der Regel ein Tag, manchmal eine Woche. Wenn Sie mit Ihren Daten etwas kommunizieren möchten, lohnt es sich meiner Ansicht nach, diese Zeit zu investieren.

**Teil I**

**Grundlagen und Technik**

## Kapitel 2

# Aufbau und technische Voraussetzungen

Bevor wir uns der konkreten Umsetzung in R zuwenden, wollen wir zunächst einige Erläuterungen zum Aufbau von Abbildungen voranstellen. Nach zwei Beispielen für die unterschiedliche Perzeption von Grafiken folgt eine Definition der Elemente von Abbildungen anhand schematischer Übersichten, die wir in Anlehnung an das Grafikdesign als „Gestaltungsraster“ bezeichnen. Anschließend folgen Erläuterungen zu wichtigen „Hilfselementen“ von Abbildungen, den verwendeten Schriften und Symbolen sowie der Farbe.

### 2.1 Begriffe und Elemente

Eine Abbildung kann eines oder mehrere *Diagramme* oder *Grafiken* enthalten. Die beiden letzten Begriffe werden hier also synonym verwendet.

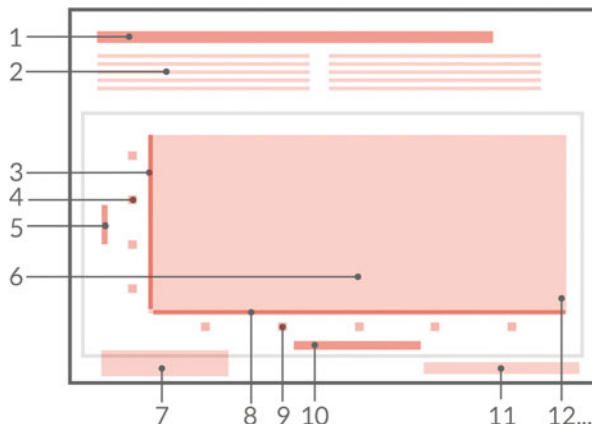
Ein Diagramm besteht aus einem *Datenbereich* (in R: *Plot Region*) und optional Achsen, Achsenbeschriftungen, Achsenbezeichnungen, Punktbezeichnungen, Legenden, Über- und Unterschriften.

Eine Abbildung kann mehrere Diagramme enthalten. In diesem Fall können in jedem Einzel-Diagramm Über- und Unterschriften, Achsen, Legende etc. vorhanden sein; darüber hinaus gibt es Über- und Unterschriften, die sich auf die gesamte Abbildung beziehen. Enthält eine Abbildung mehrere Diagramme, sprechen wir im folgenden in Ermangelung eines sinnvollen oder üblichen deutschen Begriffs von *Panel*.

### 2.2 Gestaltungsraster

Eine Abbildung besteht grundsätzlich aus einem Titel (1), einem Untertitel (2), einer Y-Achse (3) inklusive Beschriftung (4) und Namen (5), dem eigentlichen Datenbereich (6), einer Legende (7), einer X-Achse (8) inklusive Beschriftung (9) und

**Abb. 2.1** Elemente einer Abbildung

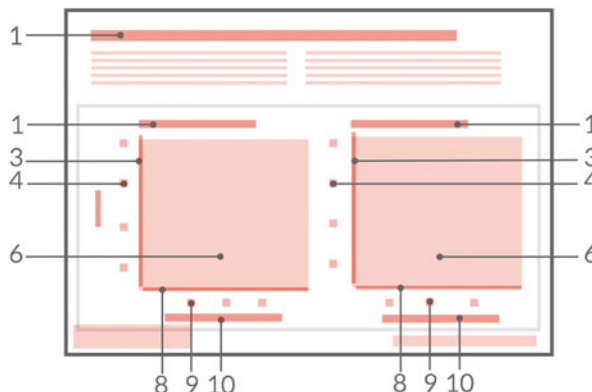


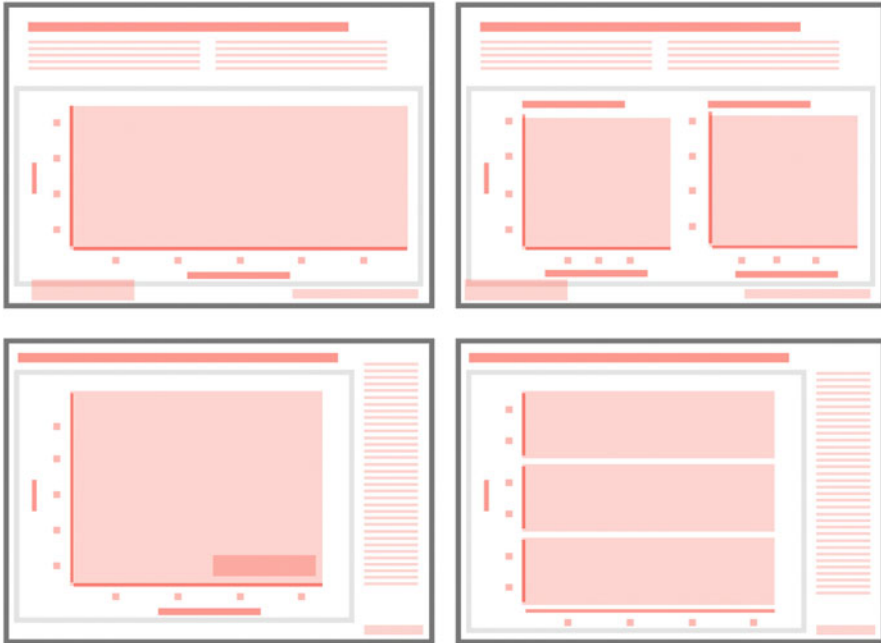
Namen (10), schließlich einer Quellenangabe (11). Darüber hinaus können Abbildungen weitere Elemente wie Annotationen, Linien oder Symbole enthalten.

Die erste Überlegung sollte das Seitenverhältnis der Abbildung betreffen. Wenn z.B. in einem Streudiagramm beide Größen Prozentangaben sind und der Wertebereich jeweils von 0 bis 100 dargestellt werden soll, dann ist es naheliegend, dass die Achsen gleich lang sind, der Datenbereich also quadratisch gezeichnet wird. In anderen Fällen ist eine Entscheidung nicht ohne weiteres möglich. In R haben Sie die Möglichkeit, bei der Erstellung einer Grafik diese Größen exakt anzugeben (Abschn. 3.3.3, Abschn. 3.3.7).

Braucht man eine Legende? Wann? Wohin? Am besten ist es, wenn man auf eine Legende verzichten kann. Das ist in aller Regel bei Zeitreihendiagrammen möglich, denn hier können die Bezeichnungen direkt an die Daten geschrieben werden: Diese sind ja per Linien verbunden und somit eindeutig. Das ist bei Punktdiagrammen nicht der Fall. Hier müssen die Bedeutungen der Farben bei Streudiagrammen mit einer Legende erläutert werden. Nahezu beliebige Einstellungsmöglichkeiten für die Form und Platzierung einer Legend bietet in R die Funktion `legend()`.

**Abb. 2.2** Elemente einer Abbildung mit zwei Diagrammen





**Abb. 2.3** Beispielhafte Anordnung einzelner Elemente

Wenn wir mehrere Grafiken in eine Abbildung aufnehmen, sprechen wir von einem *Panel*. In diesem Fall können bestimmte Elemente wiederholt auftreten (Abb. 2.2).

Die Anordnung der einzelnen Elemente kann variieren, ebenso die Anzahl der Grafiken, die in einer Abbildung enthalten sind (Abb. 2.3).

In dem vorliegenden Buch werden wir Beispiele für Abbildungen zeigen, die über 40 Grafiken enthalten. In R gibt es für die Definition solcher Panels verschiedene Möglichkeiten (Abschn. 3.3.4, Abschn. 3.3.5).

Man kann sicher keine universell gültigen Vorschriften für die Erstellung eines Gestaltungsrasters definieren. Die folgenden Hinweise sollten aber bedacht werden:

1. Es macht einen Unterschied, ob Grafiken *frei stehen* oder in einen Fließtext eingebunden sind. In letzterem Fall ist die Überschrift anders, die Schriftgrößen der einzelnen Elemente müssen angepasst werden, eine erläuternde Unterüberschrift sowie erläuternde Beschriftungen und Pfeile entfallen oder werden sparsamer verwendet.
2. In aller Regel gibt es nicht nur *eine* angemessene Darstellung der Daten, sondern *mehrere*. Ob man etwa Balken stapelt oder in einem Panel mehrere Balkendiagramme darstellt, muss im *konkreten* Einzelfall anhand der *konkreten* Daten entschieden werden.

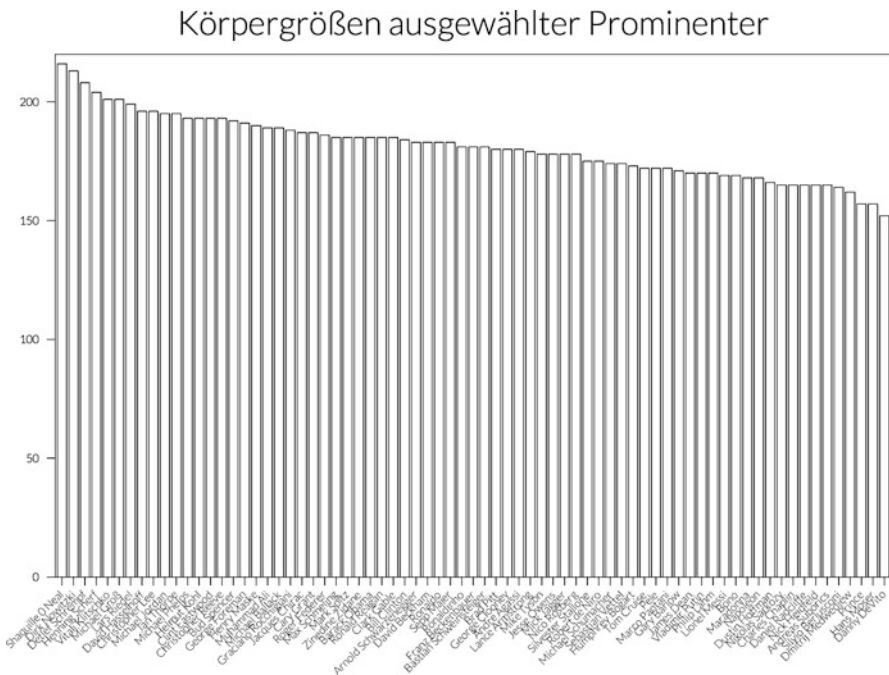


3. Quellenangabe und Titel einer Abbildung innerhalb eines Aufsatzes, Buches oder einer Website können entfallen, wenn diese Angaben dort erfolgen, wo die Abbildung eingebunden wird.

### 2.3 Perzeption

Der wichtigste Aspekt bei der Gestaltung von Abbildungen ist die richtige Wahrnehmung der Daten. Diese kann durch eine unglückliche Darstellungsform stark beeinträchtigt werden. Zwei Beispiele:

Im ersten Beispiel werden die Körpergrößen ausgewählter prominenter Personen dargestellt (Abb. 2.4).<sup>1</sup>



**Abb. 2.4** Körpergrößen ausgewählter Prominenter

Die Skalierung der Y-Achse beginnt, wie häufig gefordert, bei Null. Insgesamt wird dadurch der Eindruck erweckt, als liegen die Körpergrößen dieser Personen recht nahe beieinander. Der Effekt wird verstärkt durch die – hier noch nicht einmal ausgefüllten – Säulen, deren Gesamtvolumen einen Großteil der gesamten Fläche der Abbildung einnimmt.

<sup>1</sup> Quelle: [celebrityheights.com](http://celebrityheights.com).

# Körpergrößen ausgewählter Prominenter

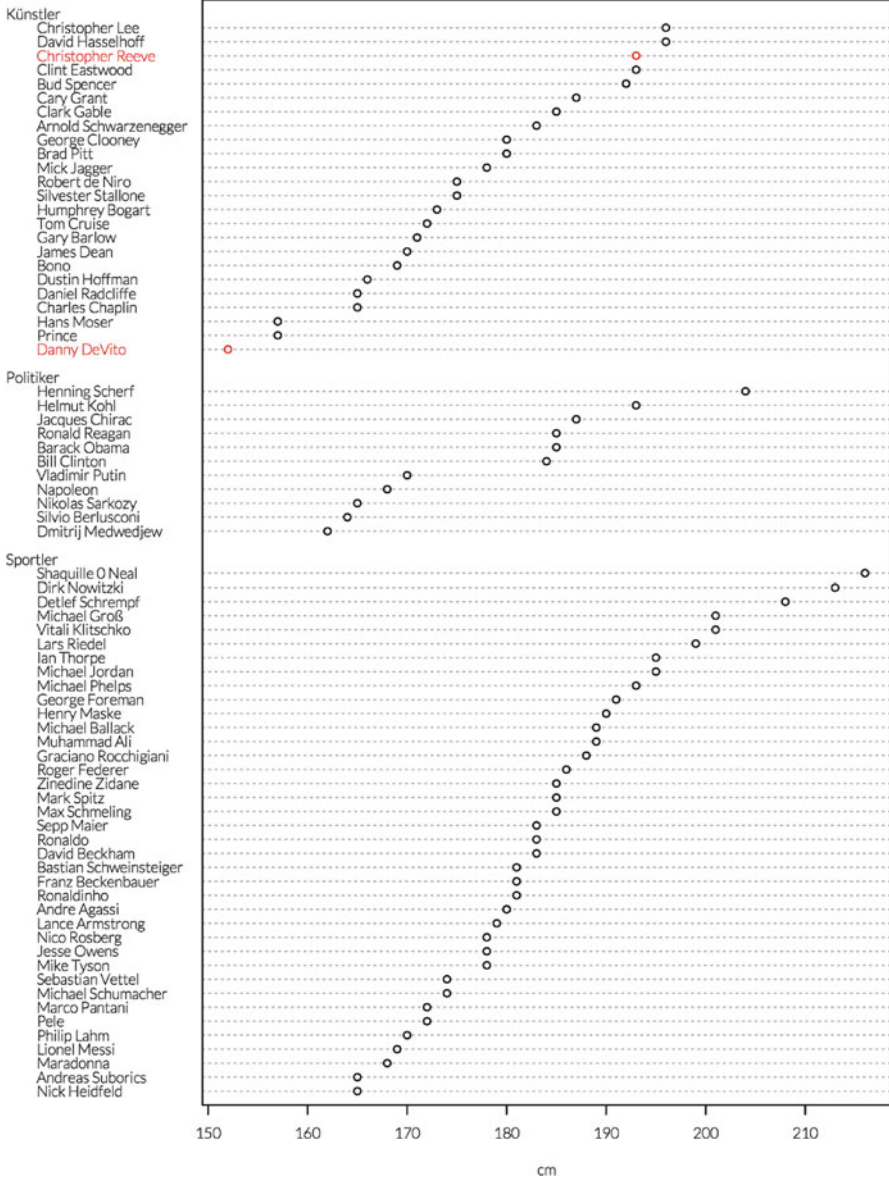


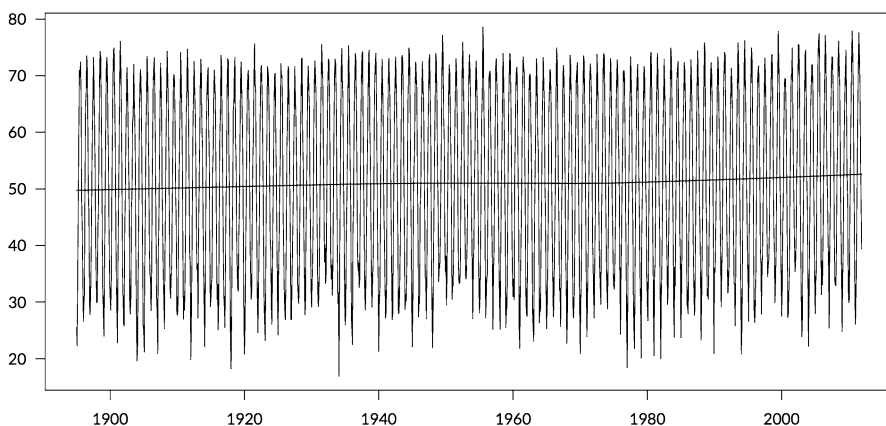
Abb. 2.5 Körpergrößen ausgewählter Prominenter als Dotchart

Das widerspricht unserer Alltagserfahrung, die doch beachtliche Unterschiede zwischen Körpergrößen wahrnimmt. Im Internet findet man ein Bild, auf dem Danny de Vito neben Christopher Reeve abgebildet ist. Vermutlich werden die meisten Betrachter dieses Bildes der Auffassung sein, dass das Säulendiagramm die Unterschiede in Körpergrößen nicht angemessen abbildet. Sinnvoller für diese Daten ist der folgende, vor allem von William S. Cleveland mehrfach empfohlene *Dotchart* (Abb. 2.5).

Vier Unterschiede zu dem Säulendiagramm verbessern die Wahrnehmung erheblich:

1. Anstelle von Säulen wird die Information der Körpergröße durch Punkte abgebildet.
2. Eine Gruppierung nach „Prominententyp“ liefert eine zusätzliche Informationsebene und sorgt durch die Gruppierung insgesamt für eine größere Übersichtlichkeit.
3. Die Skalierung beginnt nicht bei Null, sondern bei den Daten.
4. Durch die horizontale Anordnung sind die Namen der Personen besser lesbar.

Ein zweites Beispiel betrifft Zeitreihen. William S. Cleveland hat mit dem Begriff „Banking“ eine Vorgehensweise beschrieben, die für Liniendiagramme eine geeignete Darstellungsform sicherstellen soll. Die Grundidee ist, dass man das Charakteristische der Daten am besten wahrnimmt, wenn die Datenlinien im Mittel möglichst nahe bei einem 45-Grad-Winkel liegen. Wir illustrieren das an einem Beispiel, das die Monatstemperaturen in New Jersey von 1895 bis 2011 zeigt (Abb. 2.6).

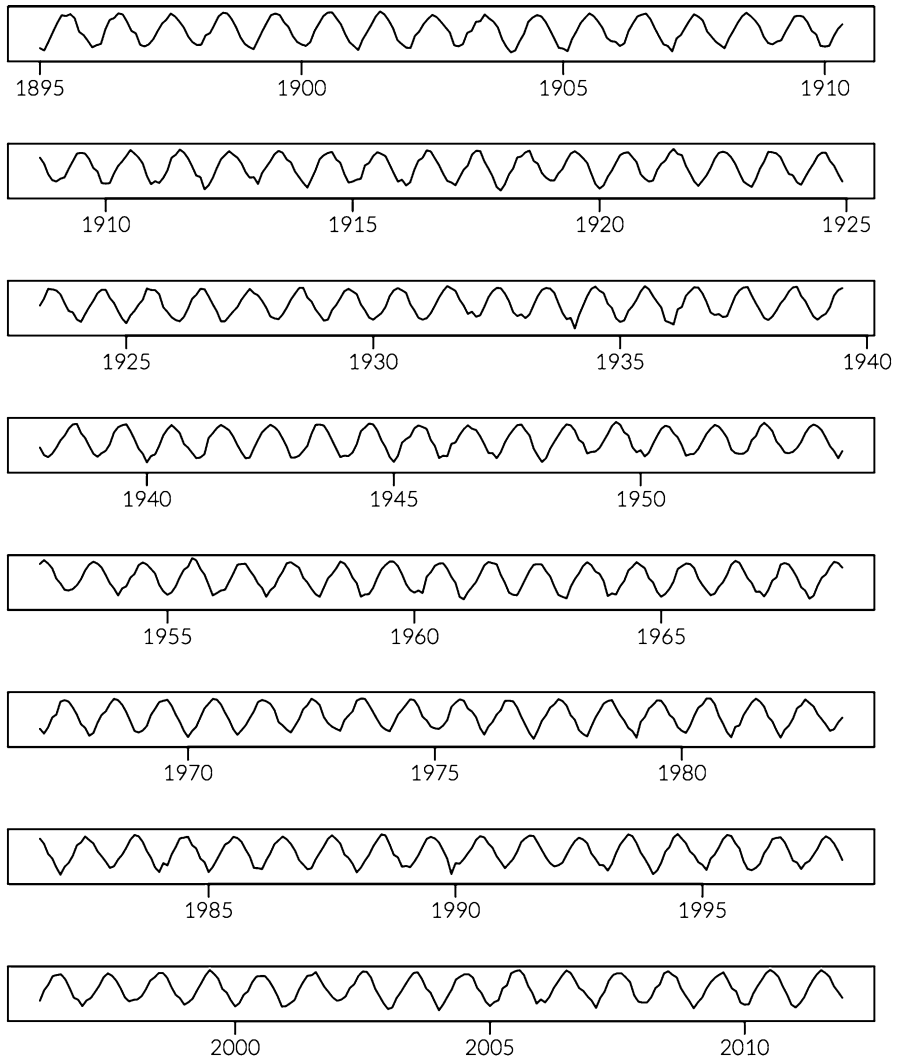


**Abb. 2.6** Monatstemperaturen in New Jersey von 1895 bis 2011 mit Trend

Bei diesem bewusst extrem gewählten Beispiel sind die Linien so gestaucht, dass die genaue Verlaufsform *der eigentlichen Daten* praktisch nicht erkennbar ist. Sehr wohl kann man dagegen den Trendverlauf erkennen, der hier sehr leicht, aber dennoch eindeutig nach oben zeigt.

Einen ganz anderen Eindruck erhält man dagegen, wenn man die Abbildung „auseinanderzieht“ und daraus einen so genannten Cut-and-Stack-Plot erstellt (Abb. 2.7).

Hier ist die zyklische Verlaufsform der Monatstemperaturen sehr gut erkennbar. Andererseits kann man dieser Abbildung keinerlei Trendverlauf mehr entnehmen. Es kommt also auch auf die gewünschte Aussage an, welche Abbildungsform vorzuziehen ist.



**Abb. 2.7** Monatstemperaturen in New Jersey von 1895 bis 2011 als Cut-and-Stack-Plot

## 2.4 Schriften

Schriften bilden einen nicht unerheblichen Teil in Abbildungen. Leider werden sie in aller Regel sehr stiefmütterlich behandelt. Man kann aber mit der Verwendung der richtigen Schrift einen erheblichen Gewinn an Klarheit erzielen. Eine interessante Untersuchung verdanken wir Sven Neumann vom Fachbereich Gestaltung an der HTW Berlin. Er hat sich mit der Leserlichkeit von Schrift im öffentlichen Raum befasst und bei der Befragung von über 100 Personen festgestellt, dass der Abstand, ab dem eine Schrift lesbar ist, von Schrift zu Schrift deutlich variiert. Das ist nicht nur für Verkehrsschilder relevant, auch Abbildungen profitieren von lesbaren Schriften.<sup>2</sup>

Viele Anwender beschränken sich in der Auswahl ihrer Schriften für Texte und erst recht für Abbildungen auf die Vorgaben ihrer Software oder ihres Betriebssystems. Das hat nicht nur pragmatische, sondern auch finanzielle Gründe: Wenn Sie eine hochwertige Schrift wie die Frutiger jeweils in einer regulären, kursiven und fetten Variante in drei verschiedenen Stärken kaufen, müssen Sie schon mehrere hundert Euro ausgeben – bei unklarer Rechtslage, was Sie damit eigentlich machen dürfen.<sup>3</sup> Glücklicherweise gibt es eine ganze Reihe freier und qualitativ hochwertiger Alternativen, deren Einsatz auch bei der Erstellung von Abbildungen sinnvoll ist. Bevor wir uns diese näher ansehen, wollen wir zunächst einen Überblick über die wichtigsten Eigenschaften von Schriften geben.

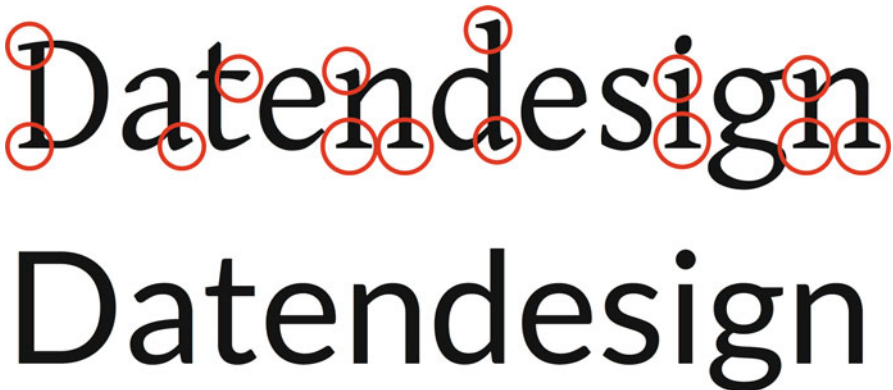


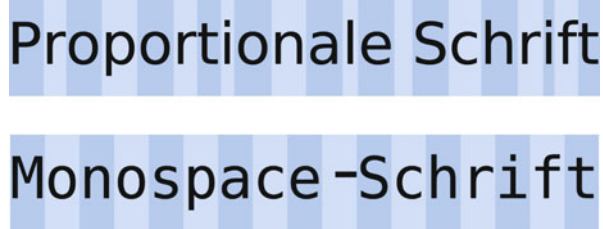
Abb. 2.8 Serifen- und serifenlose Schrift

Aktuell werden in Deutschland Schriften nach der DIN 16518 in elf Gruppen eingeteilt. Für den Hausgebrauch reicht jedoch eine gröbere Klassifikation. Grundsätzlich unterscheidet man *proportionale* und *nichtproportionale* Schriften. Insbesondere erstere werden noch einmal in *Serifen-* und *serifenlose* Schriften unterteilt.

<sup>2</sup> <http://kd.htw-berlin.de/abschlussarbeiten/leserlichkeit-von-schriften/>.

<sup>3</sup> <http://www.typografie.info/2/content.php/152-Mythos-Schriftlizenzen>.

**Abb. 2.9** Proportionale und nichtproportionale Schrift



Serifen sind auf den ersten Blick so etwas wie Verzierungen eines Buchstabens: kleine, feine Linien, die quer zu den größeren Linien eines Buchstabens stehen.

Solche Schriften werden in aller Regel für lange Texte verwendet, da lange Texte in Serifenschriften erwiesenermaßen angenehmer zu lesen sind. Serifenlose Schriften verwendet man dagegen für Überschriften oder kurze Texte. Eine proportionale Schrift ist dadurch gekennzeichnet, dass die einzelnen Buchstaben unterschiedlichen Platz in der Breite beanspruchen. Ein kleines „l“ oder „i“ braucht weniger Platz als ein „m“. Bei den nichtproportionalen Schriften ist dagegen jeder Buchstabe gleich breit.<sup>4</sup>

Die vermutlich bekannteste nichtproportionale (Serifen-)Schrift ist Courier, die man als Schreibmaschinenschrift kennt – die heute wohl bekannteste serifenlose Proportionalchrift die Helvetica von Max Miedinger und Eduard Hoffmann. Bereits ein Vierteljahrhundert zuvor wurde die bis heute bekannteste proportionale Serifenschrift, die Times von Stanley Morison und Victor Lardent gestaltet.

|                    | mit Serifen | ohne Serifen    |
|--------------------|-------------|-----------------|
| proportional       | Times       | Helvetica       |
| nicht proportional | Courier     | Liberation Mono |

**Abb. 2.10** Schriftenbeispiele

<sup>4</sup> Quelle: Wikipedia. Urheber: Algos.