

O'REILLY®

Daten- visualisierung Grundlagen & Praxis

Wie Sie aussagekräftige
Diagramme und Grafiken
gestalten



Claus O. Wilke
Übersetzung von Bilgehan Gür

Papier
plus⁺
PDF.

Zu diesem Buch – sowie zu vielen weiteren O'Reilly-Büchern – können Sie auch das entsprechende E-Book im PDF-Format herunterladen. Werden Sie dazu einfach Mitglied bei oreilly.plus⁺:

www.oreilly.plus

Datenvisualisierung – Grundlagen und Praxis

*Wie Sie aussagekräftige
Diagramme und Grafiken gestalten*

Claus O. Wilke

*Deutsche Übersetzung von
Bilgehan Gür*

O'REILLY®

Claus O. Wilke

Lektorat: Ariane Hesse

Übersetzung: Bilgehan Gür

Fachliche Unterstützung: Claus O. Wilke, Marcel Fiebelkorn (Analytics Endeavor), Joachim Zuckarelli, Kristian Rother

Korrektur: Friederike Daenecke, Zülpich

Satz: III-satz, www.drei-satz.de

Herstellung: Stefanie Weidner

Umschlaggestaltung: Michael Oréal, www.oreal.de

Druck und Bindung: mediaprint solutions GmbH, 33100 Paderborn

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.

ISBN:

Print 978-3-96009-121-9

PDF 978-3-96010-381-3

ePub 978-3-96010-382-0

mobi 978-3-96010-383-7

Dieses Buch erscheint in Kooperation mit O'Reilly Media, Inc. unter dem Imprint »O'REILLY«.

O'REILLY ist ein Markenzeichen und eine eingetragene Marke von O'Reilly Media, Inc. und wird mit Einwilligung des Eigentümers verwendet.

1. Auflage

Translation Copyright für die deutschsprachige Ausgabe © 2020 dpunkt.verlag GmbH

Wiebinger Weg 17

69123 Heidelberg

Authorized German translation of the English edition of *Fundamentals of Data Visualization*

ISBN 978-1-492-03108-6 © 2019 Claus O. Wilke. This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

Hinweis:

Dieses Buch wurde auf PEFC-zertifiziertem Papier aus nachhaltiger Waldwirtschaft gedruckt. Der Umwelt zuliebe verzichten wir zusätzlich auf die Einschweißfolie.



Schreiben Sie uns:

Falls Sie Anregungen, Wünsche und Kommentare haben, lassen Sie es uns wissen: komentar@oreilly.de.

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen.

Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Verlag noch Übersetzer können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

5 4 3 2 1 0

Inhalt

Vorwort	XI
1 Einleitung	1
Hässliche, schlechte und falsche Abbildungen	2
<hr/>	
TEIL I: Von den Daten zur Visualisierung	5
2 Datenvisualisierung: die Darstellung von Daten gestalten (Aesthetics)	7
Aesthetics (Gestaltungselemente) und Datenarten	7
Skalen bilden Datenwerte auf Aesthetics ab	10
3 Koordinatensysteme und Achsen	13
Kartesische Koordinaten	13
Nichtlineare Achsen	16
Koordinatensysteme mit gekrümmten Achsen	21
4 Farbskalen	25
Farbe als Unterscheidungsmerkmal	25
Farbe zur Darstellung von Datenwerten	27
Farbe als Werkzeug zur Hervorhebung	30
5 Ausgewählte Visualisierungen	33
Quantitative Werte	33
Verteilungen	34
Proportionen	35
x-y-Relationen	36
Geodaten	38
(Mathematische) Unsicherheit	38

6	Visualisierung quantitativer Werte	41
	Balkendiagramme	41
	Gruppierte und gestapelte Balken	45
	Punktogramme und Heatmaps	49
7	Visualisierung von Verteilungen: Histogramme und Dichtediagramme	55
	Visualisierung einer einzelnen Verteilungsgröße	55
	Gleichzeitige Visualisierung mehrerer Verteilungsgrößen	60
8	Visualisierung von Verteilungen: Empirisch kumulative Häufigkeitsverteilung und Q-Q-Plots	65
	Empirisch kumulative Häufigkeitsverteilung	65
	Stark verzerrte Verteilungen	68
	Quantil-Quantil-Diagramme	71
9	Gleichzeitige Visualisierung mehrerer Verteilungen	75
	Visualisierung von Verteilungen entlang der vertikalen Achse	75
	Visualisierung von Verteilungen entlang der horizontalen Achse	82
10	Visualisierung von Proportionen	85
	Ein Fall für Kreisdiagramme	85
	Ein Fall für nebeneinander angeordnete Balken (engl. Side-by-side bars)	88
	Ein Fall für gestapelte Balken und gestapelte Dichten	90
	Proportionen separat als Teile der Summe visualisieren	92
11	Visualisierung verschachtelter Proportionen	95
	Fehlerhaft verschachtelte Proportionen	95
	Mosaikdiagramme und Treemaps	97
	Verschachtelte Kreisdiagramme	100
	Parallel Sets	102
12	Visualisierung von Korrelationen zwischen zwei oder mehr quantitativen Variablen	105
	Streudiagramme (engl. Scatterplots)	105
	Korrelogramme	109
	Dimensionsreduktion	112
	Gepaarte Daten	115
13	Visualisierung von Zeitreihen und anderen Funktionen einer unabhängigen Variablen	119
	Einzelne Zeitreihen	119
	Mehrere Zeitreihen und Dosiswirkungskurven	122
	Zeitreihe von zwei oder mehr Antwortvariablen	124

14 Visualisierung von Trends	131
Glättung	131
Trends mit einer definierten funktionalen Form anzeigen	137
Trendbereinigung und Zerlegung von Zeitreihen	140
15 Visualisierung von Geodaten	147
Projektionen	147
Schichten	154
Choroplethenkartierung	156
Kartogramme	160
16 Visualisierung von Ungenauigkeiten	163
Wahrscheinlichkeiten als Häufigkeiten darstellen	163
Visualisierung der Unsicherheit von Punktschätzungen	168
Visualisierung der Unsicherheit von Kurvenanpassungen	180
Hypothetische Ergebniskurven	183
<hr/>	
TEIL II: Prinzipien des Graphen-Designs	187
17 Das Prinzip proportionaler Farbflächen	189
Visualisierungen entlang linearer Achsen	190
Visualisierungen entlang logarithmischer Achsen	194
Direkte Flächendarstellungen	197
18 Umgang mit überlappenden Punkten	199
Teiltransparenz und Jittering	199
2D-Histogramme	202
Konturlinien	204
19 Typische Fallstricke beim Gebrauch von Farbe	209
Anzeige zu vieler oder irrelevanter Informationen	209
Verwendung nichtmonotoner Farbskalen zur Darstellung von Datenwerten	213
Ungeeignete Abbildungen für Leser mit Farbsehschwäche	214
20 Redundante Codierung	219
Gestaltung von Legenden mit redundanter Codierung	219
Gestaltung von Abbildungen ohne Legenden	225
21 Multipanel-Diagramme	229
Small Multiples	229
Zusammengesetzte (modulare) Abbildungen	234

22	Titel, Beschriftungen und Tabellen	241
	Titel und Beschriftungen.	241
	Achsen- und Legendenbeschriftung	244
	Tabellen	247
23	Bringen Sie Daten und Kontext in Einklang	251
	Den Kontext in angemessenem Umfang bereitstellen	251
	Hintergrundgitter	256
	Gepaarte Daten.	261
	Zusammenfassung	263
24	Verwenden Sie größere Achsenbeschriftungen	265
25	Vermeiden Sie Linienzeichnungen	269
26	Vermeiden Sie 3D	277
	Vermeiden Sie unnötige 3D-Darstellungen.	277
	Vermeiden Sie 3D-Positionsskalen	279
	Angemessene Verwendung von 3D-Visualisierungen.	285
<hr/>		
TEIL III: Verschiedene Themen		287
27	Die am häufigsten verwendeten Bilddateiformate verstehen	289
	Bitmap- und Vektorgrafiken	289
	Verlustfreie und verlustbehaftete Datenkomprimierung von Bitmapgrafiken	291
	Von einem Bildformat in ein anderes konvertieren	293
28	Auswahl der richtigen Visualisierungssoftware	295
	Reproduzierbarkeit und Wiederholbarkeit.	296
	Datenexploration versus Datenpräsentation	298
	Trennung von Inhalt und Design	300
29	Eine Geschichte erzählen und Erkenntnisse auf den Punkt bringen	303
	Was ist eine Geschichte?	304
	Erstellen Sie eine Abbildung für die Generäle.	307
	Komplexe Abbildungen aufbauen	311
	Gestalten Sie Ihre Abbildungen einprägsam.	313
	Seien Sie konsistent, aber wiederholen Sie sich nicht	315

A	Kommentierte Bibliografie	321
	Nachdenken über Daten und Visualisierung	321
	Programmierbücher	322
	Statistikbücher	322
	Ältere Bücher	323
	Bücher zu verwandten Themen	324
B	Technische Hinweise	325
C	Referenzen	327
	Index	331

Vorwort

Wenn Sie Wissenschaftler, Analyst, Berater oder sonst jemand sind, der technische Dokumente oder Berichte erstellen muss, ist eine der wichtigsten Fähigkeiten, die Sie benötigen, die Fähigkeit, Daten überzeugend zu visualisieren, im Allgemeinen in Form von Abbildungen. Abbildungen tragen normalerweise das Gewicht Ihrer Argumente. Sie müssen klar, attraktiv und überzeugend sein. Der Unterschied zwischen guten und schlechten Abbildungen kann den Unterschied ausmachen zwischen einem einflussreichen oder einem unklaren Artikel, einem Zuschuss bzw. Vertrag, den Sie erhalten oder nicht, oder zwischen einem gut oder schlecht verlaufenen Vorstellungsgespräch. Und doch gibt es überraschend wenige Ressourcen, die Ihnen vermitteln, wie Sie überzeugende Datenvisualisierungen erstellen. Nur wenige Hochschulen bieten Kurse zu diesem Thema an, und es gibt auch nicht so viele Bücher zu diesem Thema. (Einige existieren natürlich.) Tutorials einer Visualisierungssoftware konzentrieren sich in der Regel darauf, wie bestimmte visuelle Effekte erzielt werden, anstatt zu erklären, warum bestimmte Optionen bevorzugt werden und andere nicht. In Ihrer täglichen Arbeit müssen Sie einfach wissen, wie man gute Abbildungen erstellt; und wenn Sie Glück haben, haben Sie einen geduligen Mentor, der Ihnen ein paar Tricks beibringt, wenn Sie Ihre ersten wissenschaftlichen Arbeiten schreiben.

Erfahrene Textredakteure sprechen vom »Ohr«, der Fähigkeit, beim Lesen innerlich zu »hören«, ob ein Text gut geschrieben ist. Ich denke, wenn es um Abbildungen und andere Visualisierungen geht, brauchen wir in ähnlicher Weise ein »Auge«, also die Fähigkeit, eine Abbildung zu betrachten und festzustellen, ob sie ausgewogen, klar und überzeugend ist. Und ebenso wie beim Beurteilen von Text kann die Fähigkeit, zu sehen, ob eine Abbildung funktioniert oder nicht, erlernt werden. Ein »Auge« zu haben bedeutet in erster Linie, dass Sie sich einer größeren Sammlung einfacher Regeln und Prinzipien für eine gute Visualisierung bewusst sind und auf kleine Details achten, die anderen möglicherweise nicht auffallen.

Meiner Erfahrung nach entwickeln Sie – genau wie beim Beurteilen von Text – noch kein Auge, wenn Sie am Wochenende mal ein Buch lesen. Es ist ein lebenslanger Prozess, und Konzepte, die für Sie heute zu komplex oder zu subtil sind,

können Ihnen in fünf Jahren viel sinnvoller erscheinen. Ich kann aus eigener Erfahrung sagen, dass sich mein Verständnis bezüglich der Gestaltung und Aufbereitung von Abbildungen weiterentwickelt. Ich versuche routinemäßig, mich mit neuen Ansätzen zu beschäftigen, und ich achte auf die visuellen und gestalterischen Entscheidungen, die andere in ihren Abbildungen treffen. Ich bin auch offen dafür, meine Meinung zu ändern. Ich mag heute eine bestimmte Abbildung für großartig halten, aber nächsten Monat könnte ich einen Grund dafür finden, sie zu kritisieren. Bitte verstehen Sie in diesem Sinne keine meiner Äußerungen als absolute Äußerung. Denken Sie kritisch über meine Argumentation für bestimmte Strategien nach, und entscheiden Sie selbst, ob Sie diese übernehmen wollen oder nicht.

Obwohl der Inhalt in diesem Buch in einem logischen Ablauf präsentiert wird, können die meisten Kapitel auch für sich studiert werden, und es ist nicht erforderlich, das Buch vollständig durchzulesen. Sie können jederzeit einen Abschnitt auswählen, an dem Sie gerade interessiert sind, oder einen Abschnitt, der sich mit einer bestimmten Designauswahl befasst, über die Sie nachdenken. Tatsächlich denke ich, dass Sie das Beste aus diesem Buch herausholen, wenn Sie nicht alles auf einmal lesen, sondern es über längere Zeiträume stückweise studieren. Versuchen Sie, einige Konzepte umzusetzen, und schlagen Sie es später erneut auf, um etwas über andere Konzepte zu lernen oder um Abschnitte über Konzepte, die Sie vor einiger Zeit kennengelernt haben, erneut zu lesen. Möglicherweise werden Sie feststellen, dass dasselbe Kapitel Ihnen andere Informationen bietet, wenn Sie es nach einigen Monaten erneut lesen.

Obwohl fast alle Abbildungen in diesem Buch mit *R* und *ggplot2* erstellt wurden, betrachte ich dies nicht als ein *R*-Buch. Ich stelle Ihnen allgemeine Prinzipien für das Erstellen von Abbildungen vor. Welche Software verwendet wurde, um die Abbildungen in diesem Buch anzufertigen, ist nebensächlich. Sie können jede beliebige Plot-Software verwenden, um die Arten von Abbildungen zu generieren, die ich hier zeige. Mit *ggplot2* und ähnlichen Paketen sind jedoch viele der von mir verwendeten Techniken viel einfacher umzusetzen als mit anderen Bibliotheken. Dennoch sollten Sie wissen, dass dies kein *R*-Buch ist und ich daher an keiner Stelle in diesem Buch Code oder Programmier Techniken erläutere. Ich möchte, dass Sie sich auf die Konzepte und Abbildungen konzentrieren, nicht auf den Code. Wenn Sie neugierig sind, wie eine der Abbildungen entstanden ist, können Sie den Quellcode des Buches in seinem GitHub-Repository nachlesen.

Gedanken zu Grafiksoftware und Pipelines zur Vorbereitung und Generierung von Abbildungen

Ich habe über zwei Jahrzehnte Erfahrung in der Vorbereitung von Abbildungen für wissenschaftliche Veröffentlichungen und habe Tausende von Abbildungen erstellt. Wenn es in diesen zwei Jahrzehnten eine Konstante gegeben hat, dann war es die Änderung der Pipelines für die Vorbereitung und das Generieren von Abbildungen

und Grafiken. Alle paar Jahre wird eine neue Grafik-Bibliothek entwickelt oder ein neues Paradigma entsteht, und große Gruppen von Wissenschaftlern wechseln zu dem brandneuen Toolkit. Ich habe Abbildungen erstellt mit Gnuplot, Xfig, Mathematica, Matlab, Matplotlib in Python, Basis R, ggplot2 in R und möglicherweise mit weiteren Tools, an die ich mich gerade nicht erinnere. Mein derzeit bevorzugter Ansatz ist ggplot2 in R, aber ich erwarte nicht, dass ich es bis zum Rentenalter weiter benutze.

Dieser ständige Wechsel der Softwareplattformen ist einer der Hauptgründe, warum dieses Buch kein Programmierbuch ist und warum ich alle Codebeispiele ausgelassen habe. Ich möchte, dass dieses Buch für Sie nützlich ist, unabhängig von der verwendeten Software, und ich möchte, dass es auch dann noch wertvoll ist, wenn ggplot2 durch die nächste neue Entwicklung ersetzt wird. Mir ist klar, dass diese Wahl manche Benutzer von ggplot2 frustrieren kann, die gern wissen möchten, wie ich eine Abbildung erstellt habe. Jeder, der sich für meine Codierungstechniken interessiert, kann dann den Quellcode des Buches nachlesen. Er ist verfügbar. Ferner werde ich in Zukunft möglicherweise eine ergänzende Veröffentlichung erstellen, die sich nur mit dem Code befasst.

Eine Sache, die ich im Laufe der Jahre gelernt habe, ist, dass Automatisierung Ihr Freund ist. Ich meine, Abbildungen sollten als Teil der Datenanalyse-Pipeline (die auch automatisiert werden sollte) automatisch generiert werden. Die Abbildungen sollten aus der Pipeline kommen und man sollte sie an den Drucker senden können, ohne dass eine manuelle Nachbearbeitung erforderlich ist. Ich sehe viele Nachwuchskräfte, die grobe Entwürfe ihrer Abbildungen erstellen und diese dann in Illustrator importieren, um sie dort aufzupolieren. Es gibt mehrere Gründe, warum dies keine gute Idee ist: Sobald Sie eine Abbildung manuell bearbeiten, kann die endgültige Version nicht mehr reproduziert werden. Ein Dritter kann nicht genau die gleiche Abbildung generieren wie Sie. Wenn Sie lediglich die Schriftart der Achsenbeschriftungen geändert haben, spielt dies möglicherweise keine Rolle. Die Grenzen sind jedoch verschwommen und es kann leicht passieren, in einen Bereich zu geraten, wo die Abgrenzungen weniger klar sind. Angenommen, Sie möchten kryptische Labels manuell durch besser lesbare Labels ersetzen. Möglicherweise kann ein Dritter nicht überprüfen, ob das Ersetzen der Labels angemessen war.

Wenn Sie zudem Ihrer Pipeline für die Abbildungen viel manuelle Nachbearbeitung hinzufügen, werden Sie zögern, Änderungen vorzunehmen oder Ihre Arbeit zu wiederholen. Auf diese Weise könnten Sie vernünftige Änderungsanforderungen von Mitarbeitern oder Kollegen ignorieren oder versucht sein, eine alte Version wiederzuverwenden, obwohl Sie tatsächlich alle Daten neu generiert haben. Dritens könnten Sie selbst vergessen, was Sie genau getan haben, um eine bestimmte Abbildung zu erstellen; oder Sie können möglicherweise nicht eine neue Abbildung mit neuen Daten generieren, die visuell genau mit Ihrer früheren Abbildung übereinstimmt. Dies sind keine erfundenen Beispiele. Ich habe gesehen, wie sie alle bei echten Menschen und echten Publikationen vorgekommen sind.

Aus all diesen Gründen sind interaktive Grafikprogramme eine schlechte Idee. Diese Programme zwingen Sie, Ihre Abbildungen manuell vorzubereiten. Tatsächlich ist es wahrscheinlich immer noch besser, einen Entwurf automatisch generieren zu lassen und ihn anschließend in Illustrator aufzupeppen, als die gesamte Abbildung in einer interaktiven Plot-Software von Hand zu erstellen. Bitte beachten Sie, dass Excel auch ein interaktives Plot-Software ist, allerdings nicht für die Erstellung von Abbildungen (oder für die Datenanalyse) empfohlen wird.

Eine wichtige Komponente in einem Buch zur Datenvisualisierung ist die Machbarkeit der vorgeschlagenen Darstellungen. Es ist schön, eine elegante neue Art der Visualisierung zu erfinden, aber wenn niemand mit dieser Visualisierungsform relativ leicht Abbildungen erzeugen kann, ist es nicht sinnvoll, sie zu verwenden. Als Tuftte zum Beispiel zum ersten Mal Sparklines vorschlug, hatte niemand eine einfache Möglichkeit, sie herzustellen. Wir brauchen zwar Visionäre, die die Welt vorantreiben, indem sie die Grenzen des Möglichen ausreizen, aber ich möchte, dass dieses Buch praktisch und direkt anwendbar für Data Scientists ist, die geradedabei sind, Abbildungen für ihre Veröffentlichungen vorzubereiten. Daher können die Visualisierungen, die ich in den folgenden Kapiteln vorschlage, mit wenigen Zeilen R-Code über ggplot2 und leicht erhältliche Erweiterungspakete erstellt werden. Tatsächlich wurde fast jede Abbildung in diesem Buch – mit Ausnahme einiger weniger in den Kapiteln 26, 27 und 28 – genau wie gezeigt automatisch generiert.

In diesem Buch verwendete Konventionen

Die folgenden typografischen Konventionen werden in diesem Buch verwendet:

Kursiv

Kennzeichnet neue Begriffe, URLs, E-Mail-Adressen, Dateinamen und Datei-erweiterungen.

Festbreitenschrift

Verweist auf Programmelemente wie Variablen- oder Funktionsnamen, Anweisungen und Schlüsselwörter.



Dieses Element kennzeichnet einen Tipp oder Vorschlag.



Dieses Element kennzeichnet einen allgemeinen Hinweis.



Dieses Element kennzeichnet eine Warnung oder weist auf Fallstricke hin.

Verwenden von Codebeispielen

Zusätzliches Material kann unter <https://github.com/clauswilke/dataviz> heruntergeladen werden.

Dieses Buch soll Ihnen dabei helfen, Ihre Arbeit zu erledigen. Wenn in diesem Buch Beispielcode angeboten wird, können Sie ihn im Allgemeinen in Ihren Programmen und in der Dokumentation verwenden. Sie müssen uns nicht um Erlaubnis bitten, es sei denn, Sie reproduzieren einen wesentlichen Teil des Codes. Schreiben Sie beispielsweise ein Programm, das mehrere Codestücke aus diesem Buch verwendet, ist hierfür keine Erlaubnis erforderlich. Der Verkauf oder Vertrieb von Beispielen aus O'Reilly-Büchern erfordert jedoch eine Genehmigung. Die Beantwortung einer Frage durch Zitieren dieses Buches und Zitieren des Beispielcodes erfordert keine Erlaubnis. Das Einfügen einer erheblichen Menge an Beispielcode aus diesem Buch in die Dokumentation Ihres Produkts erfordert hingegen wieder eine Genehmigung.

Wir wissen es zu schätzen, wenn Sie auf dieses Buch verweisen, verlangen es jedoch nicht ausdrücklich. Eine Nennung der Quelle enthält normalerweise den Titel, den Autor, den Verlag und die ISBN. Zum Beispiel: »Fundamentals of Data Visualization, Claus O. Wilke (O'Reilly). Copyright 2019 Claus O. Wilke, 978-1-492-03108-6«.

Möglicherweise werden Sie feststellen, dass weitere Formen der Nutzung als faire Verwendung einzustufen sind (z.B. die Wiederverwendung einiger Abbildungen aus dem Buch). Wenn Sie der Meinung sind, dass Ihre Verwendung von Codebeispielen oder anderen Inhalten nicht der fairen Verwendung oder der oben erteilten Erlaubnis entspricht, wenden Sie sich bitte an permissions@oreilly.com.

Danksagung

Dieses Projekt wäre ohne die fantastische Arbeit des RStudio-Teams, das R-Universum zu einer erstklassigen Publishing-Plattform zu machen, nicht möglich gewesen. Insbesondere muss ich mich bei Hadley Wickham für die Erstellung von ggplot2 bedanken, der Software, mit der alle Abbildungen in diesem Buch erstellt wurden. Ich möchte auch Yihui Xie für die Erstellung von R Markdown und für das Schreiben der knitr- und bookdown-Pakete danken. Ich glaube nicht, dass ich dieses Projekt ohne diese Tools gestartet hätte. Das Schreiben von R-Markdown-Dateien macht Spaß und es ist einfach, Material zu sammeln und an Schwung zu gewinnen. Besonderer Dank geht an Achim Zeileis und Reto Stauffer für colorspace, Thomas Lin Pedersen für ggforce und gganimate, Kamil Slowikowski für ggrepel, Edzer Pebesma für sf und Claire McWhite für ihre Arbeiten zu colorspace und colorblindr, um Farbsehschwächen in zusammengesetzten R-Abbildungen zu simulieren.

Mehrere Personen haben hilfreiche Rückmeldungen zu Entwürfen dieses Buches gegeben. Vor allem haben Mike Loukides, mein Lektor bei O'Reilly, und Steve Haroz jedes Kapitel gelesen und kommentiert. Ich erhielt auch hilfreiche Kommentare von Carl Bergstrom, Jessica Hullman, Matthew Kay, Tristan Mahr, Edzer Pebesma, Jon Schwabish und Hadley Wickham. Len Kiefers Blog und Kieran Healy's Buch und Blogpostings haben zahlreiche Inspirationen für die Erstellung von Abbildungen und Datensätzen geliefert. Eine Reihe von Personen wies auf kleinere Probleme oder Tippfehler hin, darunter Thiago Arrais, Malcolm Barrett, Jessica Burnett, Jon Calder, Antônio Pedro Camargo, Daren Card, Kim Cressman, Akos Hajdu, Thomas Jochmann, Andrew Kinsman, Will Koehrsen, Alex Lalejini, John Leadley, Katrin Leinweber, Mikel Madina, Claire McWhite, S'busiso Mkhondwane, Jose Nazario, Steve Putman, Maëlle Salmon, Christian Schudoma, James Scott-Brown, Enrico Spinielli, Wouter van der Bijl und Ron Yurko.

Des Weiteren möchte ich auch allen anderen Mitwirkenden an der tidyverse- und der R-Community im Allgemeinen danken. Es gibt wirklich ein R-Paket für jede Visualisierungsherausforderung, der man begegnen kann. Alle diese Pakete wurden von einer weitreichenden Community von Tausenden von Datenwissenschaft-

lern und Statistikern entwickelt, und viele von ihnen haben in irgendeiner Form zur Erstellung dieses Buches beigetragen.

Abschließend möchte ich meiner Frau Stefania dafür danken, dass sie viele Abende und Wochenenden geduldig ausgeharrt hat, während ich Stunden vor dem Computer verbracht habe, um ggplot2-Code zu schreiben, winzige Details bestimmter Abbildungen zu überdenken und Kapiteldetails auszuarbeiten.

Einleitung

Datenvisualisierung ist teils Kunst und teils Wissenschaft. Die Herausforderung besteht darin, die Kunst richtig zu machen, ohne die Wissenschaft falsch zu machen, und umgekehrt. Eine Datenvisualisierung muss in erster Linie die Daten präzise darstellen. Sie darf nicht irreführen oder verfälschen. Wenn eine Zahl doppelt so groß ist wie eine andere, beide Zahlen in der Abbildung aber ungefähr gleich aussehen, dann ist die Darstellung falsch. Gleichzeitig soll eine Datenvisualisierung ästhetisch ansprechend sein. Optisch gelungene Präsentationen unterstreichen die Aussagekraft der Datenvisualisierung. Wenn eine Abbildung störende Farben, unausgewogene visuelle Elemente oder andere ablenkende Merkmale enthält, fällt es dem Betrachter schwerer, das Diagramm richtig zu interpretieren.

Meiner Erfahrung nach wissen Wissenschaftler häufig (wenn auch nicht immer!), wie man Daten visualisiert, ohne dass die Darstellung grob irreführend ist. Möglicherweise haben sie jedoch keinen ausgeprägten Sinn für visuelle Ästhetik und treffen versehentlich Entscheidungen, die ihre gewünschte Botschaft beeinträchtigen. Designer hingegen können optisch ansprechende Darstellungen erstellen, gehen aber mit Daten zu leichtfertig um. Mein Ziel ist es, beiden Gruppen nützliche Informationen zur Verfügung zu stellen.

In diesem Buch versuche ich, die wichtigsten Prinzipien, Methoden und Konzepte abzudecken, die zur Visualisierung von Daten für Veröffentlichungen, Berichte oder Präsentationen erforderlich sind. Da die Datenvisualisierung ein weites Feld ist und in seiner weitesten Definition auch Themen wie schematische technische Zeichnungen, 3D-Animationen und Benutzeroberflächen umfassen kann, musste ich den Umfang notwendigerweise einschränken. Ich befasse mich speziell mit statischen Visualisierungen, sei es in gedruckter Form, online oder als Folien. Das Buch behandelt keine interaktiven Bilder oder Filme, außer in einem kurzen Abschnitt in Kapitel 16. Deshalb werde ich in diesem Buch die Wörter »Visualisierung« und »Abbildung« abwechselnd verwenden.

Auch enthält das Buch keine Anleitungen, wie Sie Abbildungen mit vorhandener Visualisierungssoftware oder Programmierbibliotheken erstellen. Die kommen-

tierte Bibliografie am Ende des Buchs enthält Verweise auf geeignete Texte zu diesen Themen.

Das Buch ist in drei Teile gegliedert. Der erste Teil, »Von den Daten zur Visualisierung«, beschreibt verschiedene Arten von Plots und Diagrammen, z.B. Balkendiagramme, Streudiagramme und Kreisdiagramme. Der Schwerpunkt liegt hier auf dem wissenschaftlichen Teil der Visualisierung. In diesem Teil diskutiere ich einen Kernsatz von Darstellungen, auf die Sie wahrscheinlich in Veröffentlichungen stoßen werden und/oder die Sie in Ihrer eigenen Arbeit benötigen. Ich verzichte bewusst darauf, mit einem enzyklopädischen Ansatz jede Art von Visualisierung vorzustellen. Bei der Organisation dieses Teils habe ich versucht, Visualisierungen nach der Art der von ihnen übermittelten Botschaft zu gruppieren und nicht nach der Art der darzustellenden Daten. In statistischen Texten wird die Datenanalyse und -visualisierung häufig nach den vorliegenden Datentypen erörtert, wobei das Material nach Anzahl und Typ der Variablen (eine kontinuierliche Variable, eine diskrete Variable, zwei kontinuierliche Variablen, eine kontinuierliche und eine diskrete Variable usw.) organisiert wird. Ich glaube, dass nur Statistiker diese Form der Organisation hilfreich finden.

Die meisten anderen Menschen gehen von den Aussagen aus, die sie vermitteln wollen – wie groß etwas ist, aus welchen Teilen es sich zusammensetzt, wie es sich auf etwas anderes bezieht und so weiter.

Im zweiten Teil, »Prinzipien des Graphen-Designs«, werden verschiedene Entwurfsprobleme behandelt, die beim Zusammenstellen von Datenvisualisierungen auftreten. Der primäre, aber nicht ausschließliche Schwerpunkt liegt auf dem gestalterischen Aspekt der Datenvisualisierung. Sobald wir den geeigneten Diagrammtyp für unseren Datensatz ausgewählt haben, müssen wir Entscheidungen über die optischen Elemente wie Farben, Symbole und Schriftgrößen treffen. Diese Auswahlmöglichkeiten können sich sowohl auf die Klarheit als auch auf die Eleganz einer Visualisierung auswirken. Die Kapitel in diesem zweiten Teil befassen sich mit den häufigsten Problemen, die in der Praxis wiederholt auftreten.

Der dritte Teil, »Verschiedene Themen«, behandelt einige verbleibende Aspekte, die nicht in die ersten beiden Teile passten. Hier werden Dateiformate erläutert, die häufig zum Speichern von Bildern und Plots verwendet werden. Darüber hinaus werden Überlegungen zur Auswahl der Visualisierungssoftware und zum Platzieren einzelner Abbildungen im Kontext eines größeren Dokuments angestellt.

Hässliche, schlechte und falsche Abbildungen

In diesem Buch zeige ich häufig verschiedene Versionen derselben Abbildungen. Einige dienen als Beispiele für die Erstellung einer guten Visualisierung, andere als abschreckende Beispiele für das, was man vermeiden sollte.

Um Ihnen eine einfache optische Richtlinie zu geben, welche Beispiele vorzugsweise verwendet und welche vermieden werden sollten, bezeichne ich problematische Abbildungen als »hässlich«, »schlecht« oder »falsch« (Abbildung 1-1):

Hässlich

Eine hässliche Abbildung weist gestalterische Probleme auf, ist aber ansonsten klar und informativ.

Schlecht

Eine schlechte Abbildung ist problematisch im Hinblick darauf, wie sie wahrgenommen wird. Beispielsweise ist sie unklar, verwirrend, übermäßig kompliziert oder irreführend.

Falsch

Bei einer falschen Abbildung stimmt etwas mit der Mathematik nicht. Sie ist damit objektiv und inhaltlich falsch.

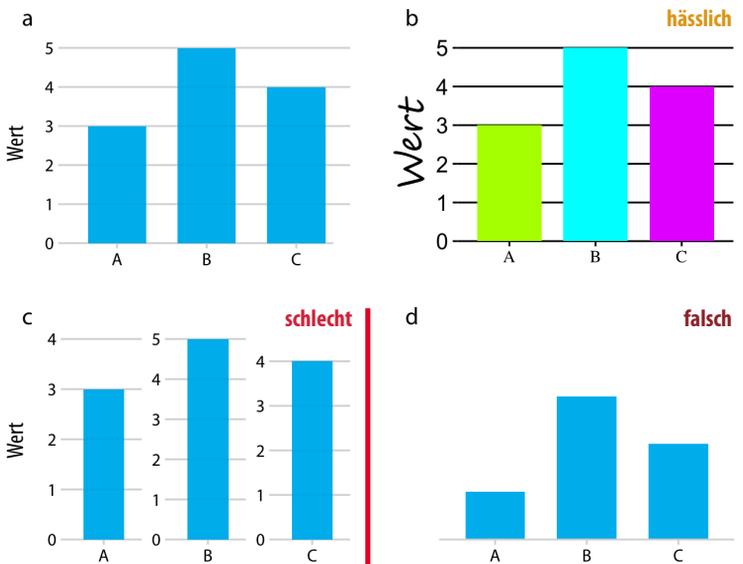


Abbildung 1-1: Beispiele für hässliche, schlechte und falsche Abbildungen. (a) Ein Balkendiagramm mit drei Werten (A = 3, B = 5 und C = 4). Dies ist eine vernünftige Visualisierung ohne größere Mängel. (b) Eine hässliche Version von Diagramm (a): Während die Abbildung technisch korrekt ist, ist sie nicht ästhetisch ansprechend. Die Farben sind zu hell und nicht brauchbar. Das Hintergrundraster ist zu stark ausgeprägt. Der Text wird mit drei verschiedenen Schriftarten in drei verschiedenen Größen angezeigt. (c) Eine schlechte Version von Diagramm (a): Jeder Balken wird mit einer eigenen Y-Achsenkala angezeigt. Da die Skalen nicht ausgerichtet sind, wirkt die Abbildung irreführend. Man kann leicht den Eindruck gewinnen, dass die drei Werte näher beieinanderliegen, als es tatsächlich der Fall ist. (d) Eine falsche Version von Diagramm (a): Ohne eine explizite y-Achsenkala können die durch die Balken dargestellten Zahlen nicht ermittelt werden. Die Balken scheinen die Längen 1, 3 und 2 zu haben, obwohl die angezeigten Werte 3, 5 und 4 sein sollen.

Gute Abbildungen kennzeichne ich nicht ausdrücklich. Jede Abbildung, die nicht als fehlerhaft gekennzeichnet ist, kann also zumindest als akzeptabel angesehen werden: als eine Abbildung, die informativ ist, ansprechend aussieht und so gedruckt werden kann, wie sie ist. Beachten Sie, dass es bei guten Abbildungen immer noch Qualitätsunterschiede gibt und einige besser sind als andere.

Ich gebe im Allgemeinen meine Gründe für bestimmte Bewertungen an, aber einige sind Geschmackssache. Generell ist die »hässliche« Bewertung subjektiver als die »schlechte« oder »falsche« Bewertung. Darüber hinaus ist die Grenze zwischen »hässlich« und »schlecht« fließend. Manchmal können schlechte Designentscheidungen die menschliche Wahrnehmung bis zu einem Punkt beeinträchtigen, an dem eine »schlechte« Bewertung angemessener ist als eine »hässliche« Bewertung. Auf alle Fälle ermutige ich Sie, Ihren eigenen Blickwinkel zu entwickeln und meine Entscheidungen kritisch zu bewerten.

Von den Daten zur Visualisierung

Datenvisualisierung: die Darstellung von Daten gestalten (Aesthetics)

Wenn wir Daten visualisieren, nehmen wir Datenwerte und wandeln sie systematisch und logisch in visuelle Elemente um, aus denen die endgültige Abbildung besteht. Obwohl es viele verschiedene Arten von Datenvisualisierungen gibt und auf den ersten Blick ein Streudiagramm, ein Kreisdiagramm und eine Heatmap nicht viel gemeinsam zu haben scheinen, können all diese Visualisierungen mit einer gemeinsamen Sprache beschrieben werden, indem Datenwerte in Tintenkleckse auf Papier oder farbige Pixel auf einem Bildschirm umgewandelt werden. Die wichtigste Erkenntnis ist die folgende: Alle Datenvisualisierungen ordnen Datenwerte quantifizierbaren Merkmalen zu, die in der resultierenden Grafik dargestellt werden.

Diese Merkmale bezeichnen wir als *Aesthetics*, deren Bedeutung mit dem deutschen Begriff »Ästhetik« höchst unzureichend wiedergegeben wäre. Im Weiteren verwenden wir daher den englischen Fachbegriff, der letztlich jene Gestaltungsmerkmale meint, mit der die Datenwerte in der Grafik visualisiert werden.

Aesthetics (Gestaltungselemente) und Datenarten

Die Gestaltung beschreibt jeden Aspekt eines bestimmten grafischen Elements. Einige Beispiele finden Sie in Abbildung 2-1. Eine kritische Komponente jedes grafischen Elements ist natürlicherweise seine *Position*, die beschreibt, wo sich das Element befindet. In üblichen 2D-Grafiken beschreiben wir Positionen durch einen *x*- und einen *y*-Wert, aber andere Koordinatensysteme und ein- oder dreidimensionale Visualisierungen sind ebenfalls möglich. Als Nächstes haben alle grafischen Elemente eine *Form*, eine *Größe* und eine *Farbe*. Auch wenn wir eine Schwarzweiß-Zeichnung vorbereiten, müssen grafische Elemente eine Farbe haben, die sichtbar ist: zum Beispiel schwarz, wenn der Hintergrund weiß ist oder weiß, wenn der Hintergrund schwarz ist. Wenn wir Linien zur Visualisierung von Daten verwenden, können diese Linien unterschiedliche Breiten oder Strich-Punkt-Muster aufweisen. Über die in Abbildung 2-1 gezeigten Beispiele hinaus gibt es viele andere gestalterische Aspekte, die bei einer Datenvisualisierung auftreten können. Wenn wir bei-

spielsweise Text anzeigen möchten, müssen wir möglicherweise Angaben zu Schriftart und Schriftgröße machen, und wenn sich grafische Objekte überlappen, müssen wir möglicherweise angeben, ob sie teilweise transparent sind.

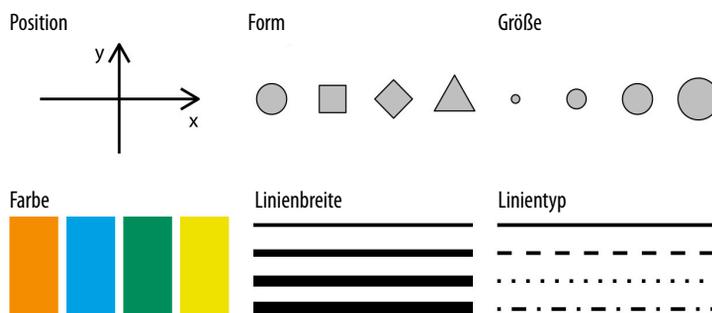


Abbildung 2-1: Häufig verwendete Gestaltungselemente in der Datenvisualisierung: Position, Form, Größe, Farbe, Linienbreite, Linientyp. Einige dieser Elemente können sowohl kontinuierliche als auch diskrete Daten (Position, Größe, Linienbreite, Farbe) darstellen, während andere normalerweise nur diskrete Daten (Form, Linientyp) darstellen können.

Alle Gestaltungselemente fallen in eine von zwei Gruppen: diejenigen, die kontinuierliche Daten darstellen können, und diejenigen, die dies nicht können. *Kontinuierliche Datenwerte* sind Werte, für die beliebig feine Zwischenprodukte existieren. Zum Beispiel ist die Zeitdauer ein kontinuierlicher Wert. Zwischen zwei beliebigen Zeitpunkten, beispielsweise 50 Sekunden und 51 Sekunden, gibt es beliebig viele Zwischenstufen, wie 50,5 Sekunden, 50,51 Sekunden, 50,50001 Sekunden und so weiter. Im Gegensatz dazu ist die Anzahl der Personen in einem Raum ein *diskreter Wert*. Ein Raum bietet Platz für 5 oder 6 Personen, nicht jedoch für 5,5 Personen. In den Beispielen von Abbildung 2-1 können Position, Größe, Farbe und Linienbreite kontinuierliche Daten darstellen, aber Form und Linientyp können normalerweise nur diskrete Daten darstellen.

Als Nächstes werden wir die Datentypen betrachten, die wir in unserer Visualisierung darstellen möchten. Sie können sich Daten als Zahlen vorstellen, aber numerische Werte sind nur zwei von verschiedenen Datentypen, auf die wir möglicherweise stoßen. Neben kontinuierlichen und diskreten numerischen Werten können Daten in Form von diskreten Kategorien, in Form von Datums- oder Uhrzeitangaben und als Text vorliegen (Tabelle 2-1). Wenn Daten numerisch sind, nennen wir sie auch *quantitativ*, und wenn sie kategorisierbar sind, nennen wir sie *qualitativ*. Variablen, die qualitative Daten enthalten, sind kategoriale *Merkmale*, und die verschiedenen Kategorien werden als *Ausprägungen* bezeichnet. Die Ausprägungen eines Merkmals sind meistens ohne Reihenfolge (wie im Beispiel von *Hund*, *Katze*, *Fisch* in Tabelle 2-1), aber es können auch geordnete Merkmale sein, falls es eine geordnete Reihenfolge zwischen den Ausprägungen des Merkmals gibt (wie im Beispiel von *gut*, *angemessen*, *schlecht* in Tabelle 2-1).

Tabelle 2-1: Arten von Variablen, die in typischen Szenarien zur Datenvisualisierung vorkommen

Variablentyp	Beispiel	Skala	Beschreibung
Quantitativ/ numerisch kontinuierlich	1,3; 5,7; 83; $1,5 \times 10^{-2}$	Kontinuierlich	Beliebige numerische Werte. Diese können ganze, rationale oder reelle Zahlen sein.
Quantitativ/ numerisch diskret	1, 2, 3, 4	Diskret	Zahlen in diskreten Einheiten. Diese sind meistens, aber nicht notwendigerweise ganze Zahlen. Zum Beispiel können die Werte 0,5 sowie 1,0 und 1,5 auch als diskrete Werte angesehen werden, wenn im gegebenen Datensatz keine dazwischen liegenden Werte existieren können.
Qualitativ/ kategorisch ungeordnet	Hund, Katze, Fisch	Diskret	Diskrete und eindeutige Kategorien ohne feste Reihenfolge. Diese Variablen werden auch als <i>Merkmale</i> bezeichnet.
Qualitativ/ kategorisch geordnet	gut, angemessen, schlecht	Diskret	Diskrete und eindeutige Kategorien mit fester Reihenfolge. Zum Beispiel liegt »angemessen« immer zwischen »gut« und »schlecht«. Diese Variablen werden auch als <i>geordnete Merkmale</i> bezeichnet.
Datum oder Zeit	5. Jan 2018, 08:03h	Kontinuierlich oder diskret	Spezifische Tage und/oder Zeiten. Hierzu zählen auch allgemeine Datumsangaben ohne Jahr wie »4. Juli« oder »25. Dezember«.
Text	Franz jagt im komplett verwahrlosten Taxi quer durch Bayern.	Keine oder diskret	Freiformtext. Kann bei Bedarf als kategorisierbar behandelt werden.

In Tabelle 2-2 finden Sie ein konkretes Beispiel für diese verschiedenen Datentypen. In ihr werden die ersten Zeilen eines Datensatzes mit den täglichen Temperaturnormalen (durchschnittliche Tagestemperaturen über ein 30-Jahres-Fenster) für vier US-amerikanische Standorte angezeigt. Diese Tabelle enthält fünf Variablen: *Monat*, *Tag*, *Ort*, *Stations-ID* und *Temperatur* (in Grad Fahrenheit). *Monat* ist ein geordneter Faktor, *Tag* ist ein diskreter numerischer Wert, *Position* ist ein ungeordneter Faktor, *Stations-ID* ist ebenfalls ein ungeordneter Faktor und *Temperatur* ist ein kontinuierlicher numerischer Wert.

Tabelle 2-2: Die ersten 8 Zeilen eines Datensatzes mit täglichen Temperaturnormalen für vier Wetterstationen.

Monat	Tag	Ort	Stations-ID	Temperatur (°F)
Jan	1	Chicago	USW00014819	25.6
Jan	1	San Diego	USW00093107	55.2
Jan	1	Houston	USW00012918	53.9
Jan	1	Death Valley	USC00042319	51.0
Jan	2	Chicago	USW00014819	25.5

Tabelle 2-2: Die ersten 8 Zeilen eines Datensatzes mit täglichen Temperaturnormalen für vier Wetterstationen. (Fortsetzung)

Monat	Tag	Ort	Stations-ID	Temperatur (°F)
Jan	2	San Diego	USW00093107	55.3
Jan	2	Houston	USW00012918	53.8
Jan	2	Death Valley	USC00042319	51.2

Skalen bilden Datenwerte auf Aesthetics ab

Um Datenwerte auf Aesthetics abzubilden, müssen wir angeben, welche Datenwerte welchen spezifischen Werten auf dieser Skala entsprechen. Wenn unsere Abbildung beispielsweise eine x -Achse hat, müssen wir angeben, welche Datenwerte auf bestimmte Positionen entlang dieser Achse fallen. In ähnlicher Form müssen wir möglicherweise angeben, welche Datenwerte durch bestimmte Formen oder Farben dargestellt werden sollen. Diese Zuordnung von Datenwerten und Gestaltungselementen erfolgt über *Skalen*. Eine Skala definiert eine eindeutige Zuordnung von Daten und Gestaltungselement (Abbildung 2-2).

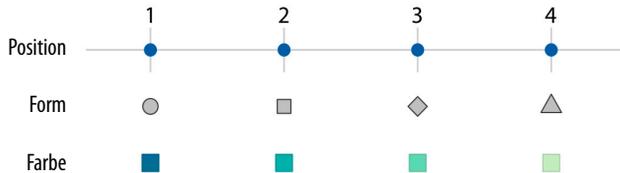


Abbildung 2-2: Skalen verknüpfen Datenwerte mit Gestaltungselementen. Hier wurden die Nummern 1 bis 4 auf eine Positionsskala, eine Formskala und eine Farbskala abgebildet. Für jede Skala entspricht jede Zahl einer eindeutigen Position, Form oder Farbe und umgekehrt.

Wichtig ist, dass eine Skala eins zu eins angelegt ist, sodass es für jeden spezifischen Datenwert genau ein Gestaltungselement gibt und umgekehrt. Wenn eine Skala nicht eins zu eins interpretierbar ist, wird die Datenvisualisierung mehrdeutig.

Lassen Sie uns das Besprochene in die Praxis umsetzen: Wir können den in Tabelle 2-2 gezeigten Datensatz nehmen, die Kartentemperatur auf die y -Achse, den Tag des Jahres auf die x -Achse und den Standort auf die Farbe übertragen und somit die Aesthetics mithilfe von durchgezogenen Linien visualisieren. Das Ergebnis ist ein Standardliniendiagramm, das die Temperaturnormalen und deren Änderung im Laufe des Jahres an den vier Standorten zeigt (Abbildung 2-3).

Abbildung 2-3 ist eine Standardvisualisierung für eine Temperaturkurve und wahrscheinlich die Visualisierung, die die meisten Datenwissenschaftler zuerst intuitiv auswählen würden. Es liegt jedoch an uns, welche Variablen auf welchen Skalen abgebildet werden. Zum Beispiel könnten wir, anstatt die Temperatur auf die y -Achse und den Standort farbige abzubilden, das Ganze auch umgekehrt gestalten.

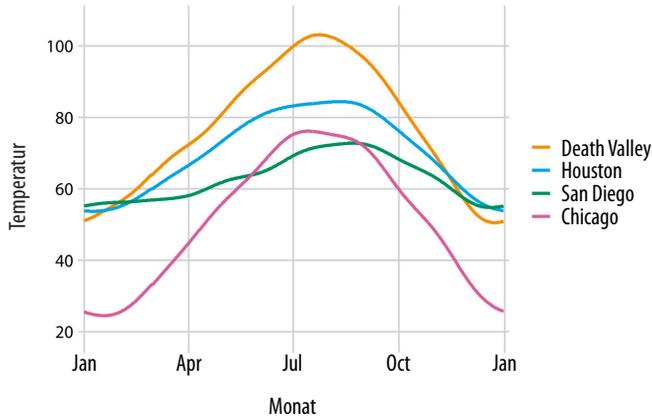


Abbildung 2-3: Tägliche Temperaturnormalen für vier ausgewählte Standorte in den USA. Die Temperatur wird auf die y-Achse, der Tag des Jahres auf die x-Achse und der Standort auf die Linienfarbe abgebildet. (Datenquelle: NOAA)

Da nun die entscheidende Schlüsselvariable (Temperatur) als Farbe angezeigt wird, müssen ausreichend große Farbflächen für die Farben angezeigt werden, um nützliche Informationen zu vermitteln [Stone, Albers Szafir und Setlur 2014]. Aus diesem Grund habe ich für diese Visualisierung in Abbildung 2-4 Quadrate anstelle von Linien ausgewählt: ein Quadrat für jeden Monat und Ort, und die Quadrate habe ich dann mit der durchschnittlichen Temperaturnormalen für jeden Monat gefärbt.

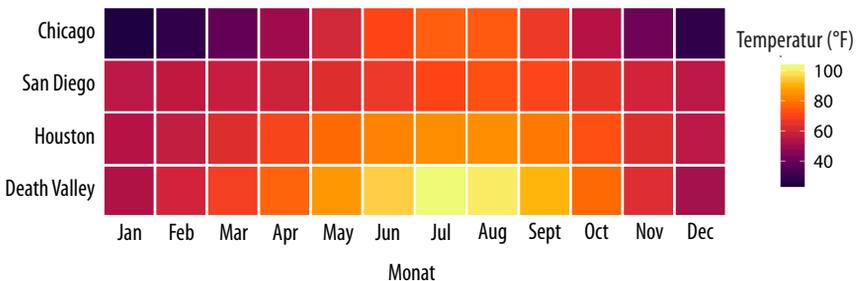


Abbildung 2-4: Monatliche normale Durchschnittstemperaturen für vier Standorte in den USA. (Datenquelle: NOAA)

Ich möchte betonen, dass in Abbildung 2-4 zwei Positionsskalen (Monat entlang der x-Achse und Standort entlang der y-Achse) verwendet werden, doch keine von ihnen ist eine kontinuierliche Skala. Der Monat hat zwölf Ausprägungen und ist ein geordnetes Merkmal (d.h., seine Ausprägungen lassen sich in eine natürliche Reihenfolge bringen); der Standort ist ein ungeordnetes Merkmal mit 4 Ebenen. Daher sind beide Positionsskalen diskret. Bei diskreten Positionsskalen platzieren wir die verschiedenen Ebenen des Merkmals im Allgemeinen in gleichem Abstand entlang der Achse. Wenn das Merkmal geordnet ist (wie hier für den Monat), müs-

sen die Ebenen in der entsprechenden Reihenfolge platziert werden. Wenn das Merkmal ungeordnet ist (wie hier für den Standort), dann ist die Reihenfolge beliebig und wir können jede gewünschte Reihenfolge wählen. Ich habe die Orte vom insgesamt kältesten (Chicago) bis zum insgesamt heißesten Ort (Death Valley) gelistet, um eine angenehme Farbfolge zu generieren. Ich hätte jedoch auch eine andere Reihenfolge wählen können – die Abbildung wäre dennoch gleichermaßen gültig gewesen.

In den beiden Abbildungen 2-3 und 2-4 wurden insgesamt drei Skalen (zwei Positionsskalen und eine Farbskala) verwendet. Dies ist eine typische Anzahl von Skalen für eine grundlegende Visualisierung, aber wir können auch mehr als drei Skalen gleichzeitig verwenden. In Abbildung 2-5 werden fünf Skalen verwendet (zwei Positionsskalen und jeweils eine Farb-, Größen- und Formsкала), und jede Skala repräsentiert eine andere Variable aus dem Datensatz.

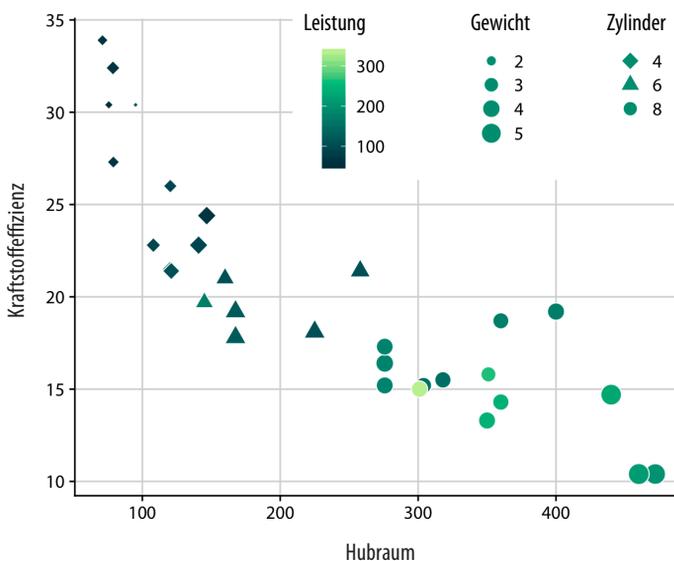


Abbildung 2-5: Kraftstoffeffizienz versus Hubraum für 32 Autos (Modelle 1973–74). Diese Abbildung verwendet fünf getrennte Skalen, um Daten darzustellen: (i) die x-Achse (Verschiebung), (ii) die y-Achse (Kraftstoffeffizienz), (iii) die Farbe der Datenpunkte (Leistung), (iv) die Größe der Datenpunkte (Gewicht) und (v) die Form der Datenpunkte (Anzahl der Zylinder). Vier der fünf angezeigten Variablen (Hubraum, Kraftstoffverbrauch, Leistung und Gewicht) sind numerisch kontinuierlich. Die verbleibende Variable (Anzahl der Zylinder) kann entweder als numerisch diskret oder qualitativ geordnet betrachtet werden. (Datenquelle: Motor Trend, 1974)