WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach

# 1 Was sind eigentlich Daten?

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



Wenn Sie fragen, was Daten sind, dann antworten die meisten Menschen mit einer vagen Beschreibung, die an ein Tabellenblatt oder einen Zahlenwust erinnert. Die Technikaffinen erwähnen vielleicht Datenbanken oder Data-Warehouses. Dies ist jedoch nur das Format, in dem die Daten vorliegen und wie sie gespeichert werden. Es sagt nichts darüber aus, was Daten sind oder was jeder einzelne Datensatz repräsentiert. Dies ist eine Falle, in die man leicht gerät. Wenn man nämlich nach Daten fragt, bekommt man in der Regel eine Computerdatei. Sich eine Ausgabedatei aber als etwas anderes als eine Datei vorzustellen, ist schwierig. Wenn Sie jedoch über die Dimension der Datei hinausdenken, dann bekommt das Ganze einen Sinn.

## Was Daten repräsentieren

Daten sind mehr als Zahlen und, um sie zu visualisieren, muss man wissen, was sie darstellen. Daten repräsentieren die Realität. Sie sind eine Momentaufnahme der Welt, ebenso wie ein Fotograf einen kurzen Augenblick einfängt.

Sehen Sie sich Abbildung 1.1 an. Wenn Ihnen dieses Foto in die Hände fallen würde, ohne weitere Informationen dazu und ohne dass ich Ihnen etwas dazu erzählte, dann könnten Sie sich keinen rechten Reim darauf machen. Es ist einfach nur ein Hochzeitsfoto. Für mich war es jedoch ein glücklicher Moment an einem der besten Tage meines Lebens. Das links ist meine Frau, wunderschön herausgeputzt, und rechts, das bin ich – mal etwas anders gekleidet als

**Abbildung 1.1** Ein Foto, ein Datenpunkt

nur in Jeans und T-Shirt. Der Pfarrer, der uns traut, ist der Onkel meiner Frau, der dieser Zeremonie eine persönliche Note verlieh. Der Mann im Hintergrund ist ein Freund der Familie, der es sich nicht hat nehmen lassen, so viel wie möglich zu fotografieren, obwohl wir einen Fotografen engagiert hatten. Die Blumen und der Torbogen stammen von einem Blumengeschäft, das eine Stunde von dem Ort des Geschehens entfernt liegt. Die Hochzeit fand im Frühsommer in Los Angeles statt.

Das sind sehr viele Informationen nur zu einem Bild und ebenso funktioniert dies mit Daten. (Für manche – so auch für mich – sind auch Bilder Daten.) Zu jedem einzelnen Datenpunkt kann ein Wer, Was, Wann, Wo und Warum gehören, daher kann es leicht sein, dass eine Ziffer in mehr als nur einem Behältnis landet. Dennoch ist das Extrahieren von Informationen aus einem Datenpunkt nicht so einfach wie das Betrachten eines Fotos. Sie können raten, was auf dem Foto passiert. Wenn Sie jedoch Mutmaßungen über Daten anstellen, etwa, wie genau sie sind oder in welchem Verhältnis sie zu ihrer Umgebung stehen, dann können Sie am Ende einen falschen Eindruck von dem gewinnen, was die Daten eigentlich darstellen. Sie müssen alles im Umfeld betrachten, den Kontext finden und herausfinden, wie Ihr Datensatz als Ganzes aussieht. Wenn Sie das Gesamtbild sehen, dann ist es viel einfacher, über einzelne Punkte besser zu urteilen.

Stellen Sie sich vor, ich hätte Ihnen nicht diese Dinge über mein Hochzeitsfoto erzählt. Wie hätten Sie dann mehr herausfinden können? Was wäre gewesen, wenn Sie Fotos sehen könnten, die davor und danach aufgenommen wurden?

Jetzt sehen Sie mehr als nur eine Momentaufnahme. Sie sehen mehrere Momente, die zusammengenommen den Teil der Hochzeit darstellen, den Einzug der Braut, das Eheversprechen und die Teezeremonie mit den Eltern und meiner Großmutter, die bei chinesischen Hochzeiten üblich ist. Wie beim ersten Foto hat auch jedes dieser Fotos seine eigene Geschichte, etwa mein Schwiegervater, der zu Tränen gerührt war, als er seine Tochter übergab, oder wie glücklich ich mich fühlte, als ich zusammen mit meiner Braut zum Altar schritt. Viele der Fotos fingen Augenblicke ein, die ich während der Hochzeit von meinem Standpunkt aus nicht gesehen habe. Daher fühle ich mich fast wie ein außenstehender Betrachter und so wird es Ihnen vermutlich auch gehen. Doch je mehr ich Ihnen über diesen Tag erzähle, umso mehr Licht scheint auf den einzelnen Punkt.

Dennoch sind dies nur Momentaufnahmen und Sie wissen nicht, was zwischen den einzelnen Fotos passiert ist. (Auch wenn Sie es erraten könnten.) Um die ganze Geschichte zu kennen, hätten Sie entweder dort sein müssen oder ein Video ansehen müssen. Doch selbst dann hätten Sie die Zeremonie nur aus bestimmten Blickwinkeln gesehen, weil es häufig nicht möglich ist, jedes einzelne Detail aufzuzeichnen. Während der Zeremonie herrschte beispielsweise fünf Minuten lang etwas Durcheinander, weil wir versuchten, eine Kerze anzuzünden, die der Wind ständig wieder ausblies. Schließlich hatten wir keine Streichhölzer mehr und der

**Abbildung 1.2** Fotogalerie

Hochzeitsplaner versuchte hektisch, welche zu besorgen. Zum Glück war unter unseren Gästen ein Raucher, der sein Feuerzeug hervorholte. Diese Fotogalerie zeigt das jedoch nicht, weil sie, wie erwähnt, nur eine Abstraktion der Realität ist.

Dies ist der Zeitpunkt für die Stichprobenerhebung. Häufig ist es nicht möglich, alles zu zählen oder aufzunehmen, weil es zu viel kostet oder die Arbeitskraft dafür fehlt (oder beides). Daher nehmen Sie die Einzelteile und suchen dann nach Mustern oder Verbindungen, um eine gut begründete Vermutung darüber anzustellen, was die Daten darstellen. Die Daten sind eine Vereinfachung – eine Abstraktion – der Realität. Wenn Sie also Daten visualisieren, dann visualisieren Sie eine Abstraktion der Welt oder zumindest einen winzigen Teil davon. Datenvisualisierung ist eine Abstraktion von Daten. Letztendlich erhalten Sie also eine Abstraktion einer Abstraktion, was eine interessante Herausforderung darstellt.

Dies soll jedoch nicht heißen, dass Datenvisualisierung Ihre Sichtweise blockiert – ganz im Gegenteil. Datenvisualisierung kann helfen, den Fokus von den einzelnen Datenpunkten zu lösen und diese aus einem anderen Blickwinkel zu erforschen – also den Wald doch vor lauter Bäumen zu sehen. Bleiben wir also bei dem Beispiel mit diesem Hochzeitsfoto. Abbildung 1.3 verwendet den gesamten Datensatz der Hochzeit, von dem Abbildung 1.1 und Abbildung 1.2 nur Teilmengen sind. Jedes Rechteck ist ein Foto aus unserem Hochzeitsalbum. Sie sind zeitlich geordnet und ihre Farben richten sich nach dem häufigsten Farbton in dem jeweiligen Foto.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



Bei einem Zeitreihenlayout sehen Sie die Höhepunkte der Hochzeit, wenn unser Fotograf mehr Fotos aufgenommen hat, sowie die Flauten, in denen nur wenige Fotos gemacht wurden. Die Spitzen in dem Diagramm sind natürlich, wenn es etwas zu fotografieren gibt, beispielsweise als ich zum ersten Mal meine Frau im Hochzeitskleid sah oder als die Zeremonie begann. Nach der Zeremonie wurden die üblichen Gruppenfotos mit Freunden und Familienangehörigen aufgenommen, daher gab es dann eine weitere Spitze. Später gab es etwas zu essen und die Aktivität ließ nach, insbesondere als die Fotografen kurz vor 16 Uhr eine Pause einlegten. Alles kam dann wieder in Gang mit dem typischen Hochzeitswalzer und der Tag ging etwa gegen 19 Uhr zu Ende. Meine Frau und ich ritten dann dem Sonnenuntergang entgegen.

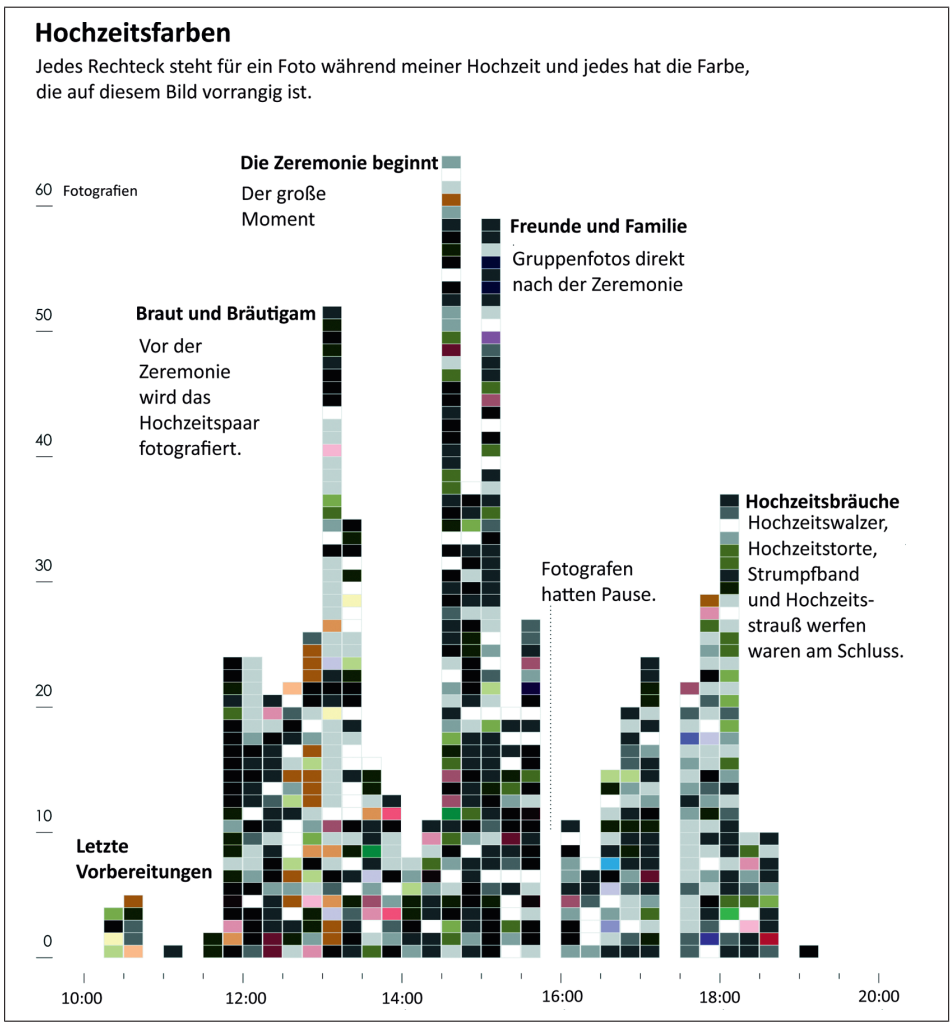



**Abbildung 1.3** Farben bei der Hochzeit

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



Im Rasterlayout ist dieses Muster unter Umständen aufgrund der linearen Darstellung nicht zu erkennen. Alles scheint mit gleichen Abständen zu passieren, obwohl eigentlich die meisten Bilder während der spannenden Momente geschossen wurden. Sie bekommen außerdem auf einen Blick ein Gefühl für die Farben der Hochzeit: Schwarz für die Anzüge, Weiß für das Brautkleid, Rot für die Blumen und die Brautjungfern und Grün für die Bäume, die die Freilufthochzeit und den Empfang umrahmten. Bekommen Sie die Details, die Sie wollten, aus den eigentlichen Fotos? Nein. Aber manchmal ist diese Ebene zu Beginn nicht notwendig. Manchmal müssen Sie erst die allgemeinen Muster erkennen, bevor Sie ins Detail gehen können. Manchmal wissen Sie nicht, dass ein einzelner Datenpunkt wichtig ist, bis Sie etwas anderes sehen und erkennen, wie dieser in die Grundgesamtheit passt.

Doch damit muss es noch nicht genug sein. Betrachten Sie eine andere Ebene und konzentrieren Sie sich nur auf die Anzahl der gemachten Fotos – ohne Farben oder Einzelbilder (Abbildung 1.4). Dieses Layout haben Sie bestimmt schon gesehen. Es ist das Säulendiagramm, das dieselben Höhen und Tiefen zeigt wie Abbildung 1.3, aber es sieht anders aus und vermittelt eine andere Botschaft. Das einfache Säulendiagramm unterstreicht die Anzahl der geschossenen Fotos im 15-Minuten-Takt, während Abbildung 1.3 auch etwas von der Stimmung des Fotoalbums vermittelt.

Das Wesentliche ist jedoch, dass alle vier Ansichten dieselben Daten zeigen: Sie stellen alle meinen Hochzeitstag dar. Jede Grafik stellt den Tag anders dar und konzentriert sich auf verschiedene Facetten der Hochzeit. Die Interpretation der Daten verändert sich mit der visuellen Form, die diese einnehmen. Traditionelle Daten untersuchen und erforschen Sie in der Regel mit einem Säulen- oder Balkendiagramm. Dies muss jedoch nicht bedeuten, dass Sie das Gefühl für den einzelnen Datenpunkt – das Einzelbild – verlieren müssen. Manchmal bedeutet dies, sinnvolle Anmerkungen hinzuzufügen, die es den Lesern ermöglichen, die Daten besser zu interpretieren. Manchmal ist die Botschaft in den Zahlen auch eindeutig und lässt sich aus der Datenvisualisierung deutlich erkennen.

Die Verbindung zwischen Daten und was diese darstellen, ist der Schlüssel zu einer Visualisierung, die etwas bedeutet. Es ist der Schlüssel zu einer ausgefeilten Datenanalyse. Es ist der Schlüssel zum tieferen Verständnis Ihrer Daten. Computer erledigen den Großteil der Arbeit und machen aus Zahlen Formen und Farben. Sie müssen jedoch die Verbindung zwischen Daten und Realität herstellen, damit Sie oder diejenigen, für die Sie die Grafiken erstellen, einen Wert daraus ziehen können.

Diese Verbindung ist manchmal schwer zu erkennen, wenn Sie auf einem großen Maßstab die Daten nach Tausenden von Fremden durchsuchen. Sie ist jedoch viel offensichtlicher, wenn Sie in den Daten nach einem Einzelnen suchen. Sie können zu dieser Person fast eine Beziehung aufbauen, selbst wenn Sie sie noch nie getroffen haben. Der in Portland lebende Entwickler Aaron Parecki hat mit seinem Telefon 2,5 Millionen GPS-Punkte in 3,5 Jahren zwischen

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



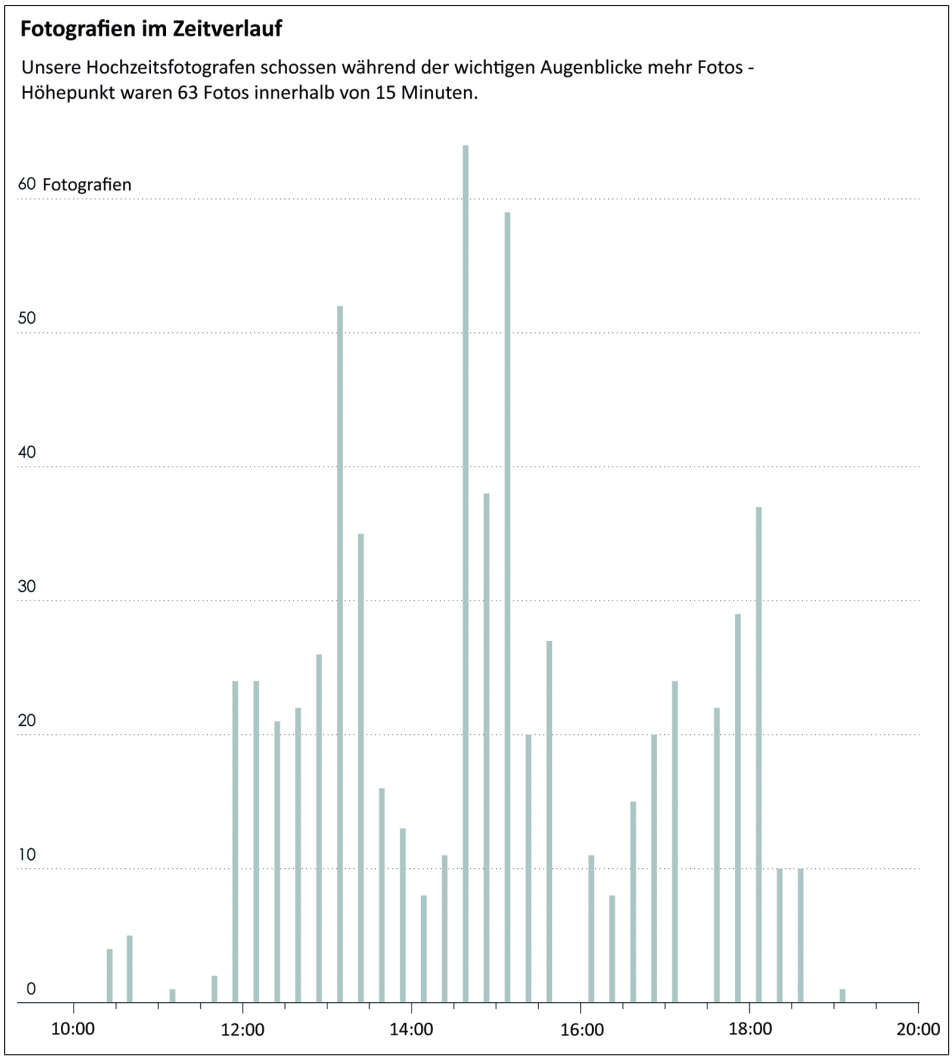



**Abbildung 1.4** Fotos im Zeitverlauf

2008 und 2012 gesammelt, also etwa einen Punkt alle zwei bis sechs Sekunden. Abbildung 1.5 zeigt eine Karte dieser Punkte, wobei die Jahre jeweils andere Farben haben.

Wie erwartet, zeigt die Karte ein Netz aus Straßen und Bereiche, die Parecki häufig besuchte und die farblich stärker hervorgehoben sind als andere. Er ist ein paar Mal umgezogen und Sie erkennen, dass sich sein Reiseverhalten mit den Jahren verändert hat. Zwischen 2008 und 2010 (blau) sind die Fahrten verteilter, während Parecki 2012 (gelb) in einigen enger zusammenliegenden Gebieten geblieben ist. Ohne weiteren Kontext ist es schwierig, mehr dazu zu

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



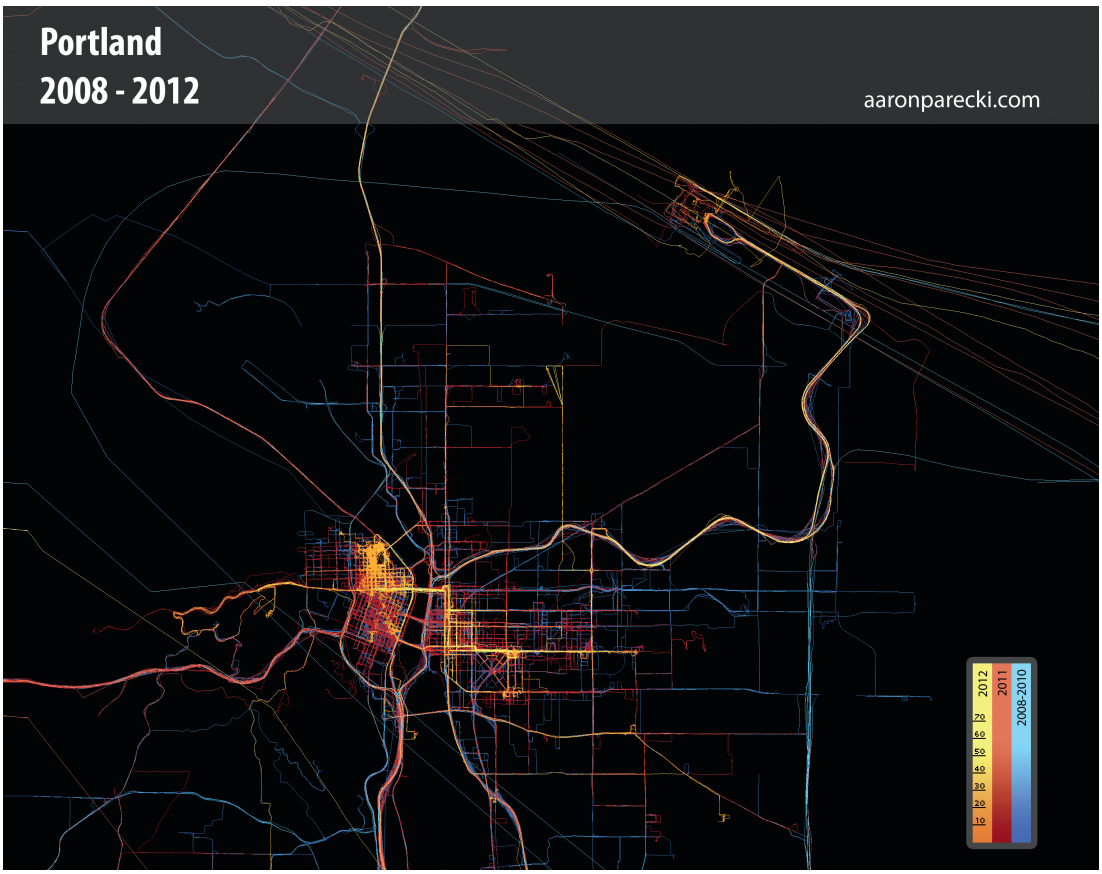



**Abbildung 1.5** GPS-Spuren, gesammelt von Aaron Parecki, http://aaronparecki.com

sagen, weil Sie nichts weiter als den Standort sehen, aber für Parecki sind die Daten etwas Persönliches (so wie das einzelne Hochzeitsfoto für mich). Sie sind sein Fußabdruck, den er in mehr als drei Jahren in einer Stadt hinterlassen hat. Weil er Zugriff auf Log-Protokolle hatte, die mit Zeitangaben verknüpft waren, konnte er auch bessere Entscheidungen basierend auf diesen Daten erstellen, beispielsweise wann er von der Arbeit nach Hause fahren sollte.

Was wäre jedoch, wenn mehr Informationen als nur Zeit- und Standortdaten damit verknüpft wären? Was wäre, wenn Sie neben diesen Daten auch noch notieren würden, was sich während oder nach einer bestimmten Zeit ereignet hat? Dies hat der Künstler Tim Clark zwischen 2010 und 2011 für sein Projekt *Atlas of the Habitual* getan. Ebenso wie Parecki zeichnete Clark seinen Standort an 200 Tagen mit einem GPS-Gerät auf, das in Bennington im US-Staat Vermont etwa 2000 Meilen abdeckte. Clark betrachtete danach seine Standortdaten und benannte bestimmte Reisen und Menschen, mit denen er Zeit verbrachte, und schlüsselte alles nach bestimmten Zeiten im Jahr auf.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



Der Atlas (Abbildung 1.6), mit per Mausklick zu aktivierenden Kategorien und Zeiten, zeigt Clarks Spuren an 200 Tagen und liest sich wie ein persönliches Tagebuch. Wählt man »Erledigungen«, dann steht dort: »Alltägliches erledigen wie zum Lebensmittelladen gehen oder sogar 30 Meilen bis zum einzigen Fahrradgeschäft im Süden Vermonts fahren, der sonntags geöffnet hat.« Die Spuren bleiben in der Umgebung der Stadt, bis auf zwei lange Ausnahmen, die aus dem Rahmen fallen.

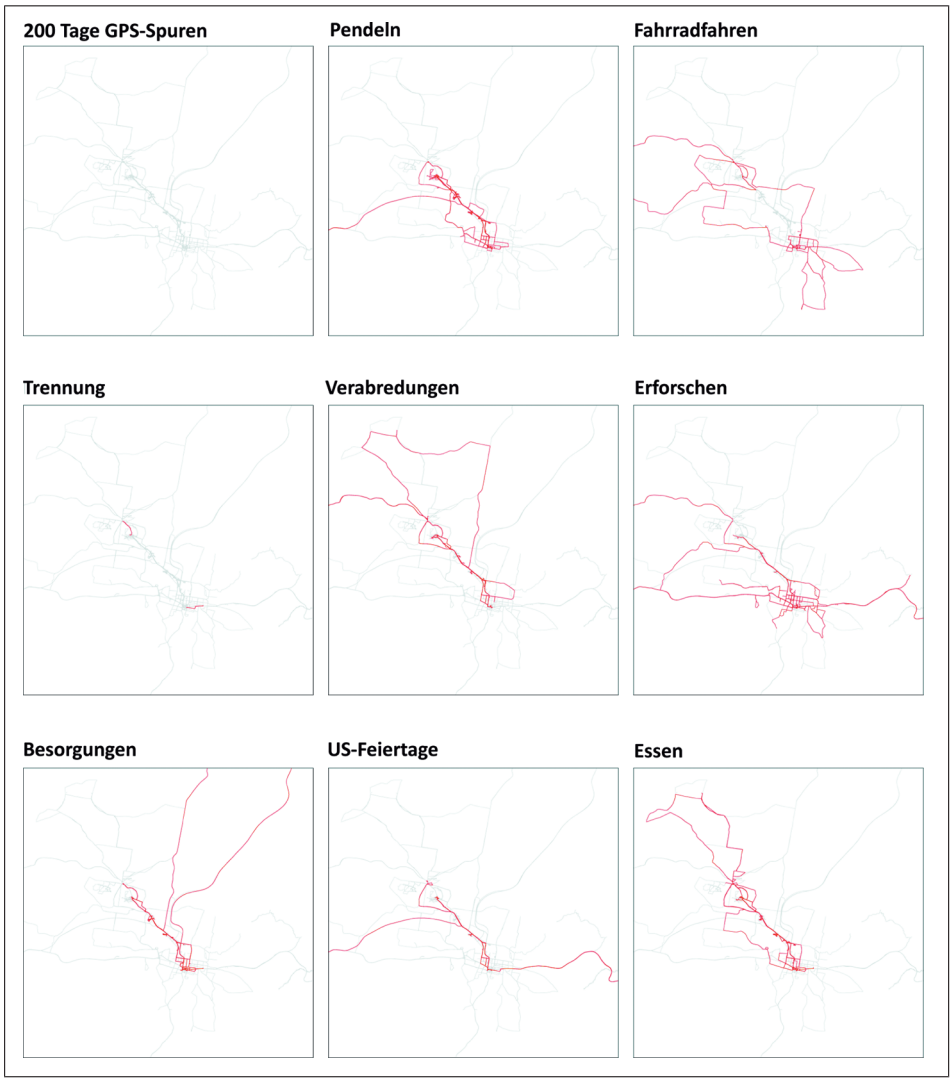
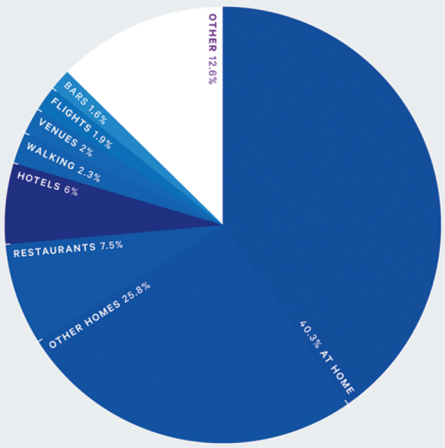



**Abbildung 1.6** Ausgewählte Karten aus »Atlas of the Habitual« von Tim Clark, www.tlclark.com/atlasofthehabitual

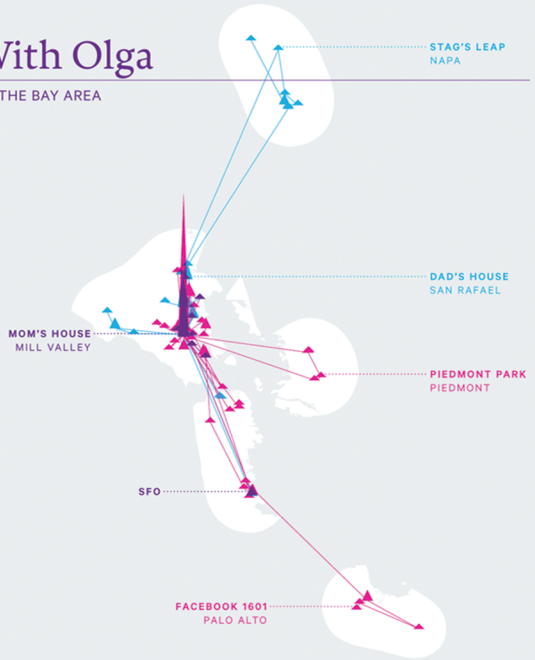
# With Olga

EVERYWHERE




# With Olga

IN THE BAY AREA




## DAYS TOGETHER

**191¼**

315 different encounters

## MOST TIME SPENT TOGETHER

- **MANHATTAN** — 83¾ DAYS
- **BROOKLYN** — 51½ DAYS
- **MILL VALLEY** — 9 DAYS
- **ANCHORAGE** — 7½ DAYS
- **SYDNEY** — 4¼ DAYS

## MOST VISITED PLACE TOGETHER

**Old Apartment**

194 visits

## DIFFERENT CITIES VISITED TOGETHER

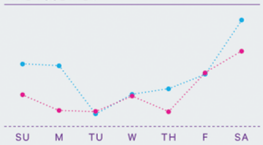
**56**

In 3 countries, 9 states and Washington D.C.

## FAVORITE BEVERAGES WITH OLGA

- **FILTER COFFEE** — 111 SERVINGS
- **RED WINE** — 78 SERVINGS
- **DALE'S PALE ALE** — 35 SERVINGS
- **CHAMPAGNE** — 30 SERVINGS
- **LATTE** — 26 SERVINGS

## TIME TOGETHER




## BRIEFEST MONTH TOGETHER

**June 2011**

40¾ hours

## MOST CONSECUTIVE HOURS TOGETHER

**247**

Australia trip — February 2010

## TIME SPENT WITH OLGA AND...

- **SARAH** — 6¼ DAYS
- **MOM** — 6¼ DAYS
- **BRIAN** — 5¾ DAYS
- **OLGA'S MOM** — 5 DAYS
- **RYAN** — 4½ DAYS

## WEDDINGS ATTENDED TOGETHER

**Seven**

Aaron & Jessica, Charlie & Bret, Glenn & Mariana, Lewis & Ange, Randy & Allison, Rob & Elise and Toby & Harriet

## DAYS TOGETHER IN THE BAY AREA

**13½**

Approximately 7% of total time together

## BAY AREA PLACES VISITED TOGETHER

**77**

18 stores, 13 restaurants, 10 homes, 6 outdoor places, 3 coffee shops, 3 grocery stores, 2 airport terminals, 2 bars, 2 gas stations, 2 hospitals, 2 hotels, 2 liquor stores, 2 parking garages, 2 parking lots, a cinema, a deli, a drug store, a laundromat, a library, a museum, a park and work

## FAVORITE BAY AREA BOTTLESHOP

**Vintage Wine & Spirits**

Visited twice

## FAVORITE BAY AREA BEER WITH OLGA

**Lagunitas IPA**

5 servings

## BAY AREA MUSEUMS VISITED TOGETHER

**The Exploratorium**

With Marina — July 9, 2011

## MOST PLAYED ARTIST TOGETHER

**The Beach Boys**

25 songs listened to from *Christmas with the Beach Boys*

## TIME TOGETHER IN THE BAY AREA


## MOST FREQUENTED CITY TOGETHER

**Mill Valley**

68% of time in the Bay Area

## MOST VISITED BAY AREA PLACES

- **MOM'S HOUSE** — 35 VISITS
- **MARIN GENERAL HOSPITAL** — 6 VISITS
- **CHEVRON MILL VALLEY** — 5 VISITS
- **SFO INTERNATIONAL TERMINAL** — 4 VISITS
- **DAD'S HOUSE** — 3 VISITS

## CRISES INVOLVING A TICK

**One**

Spotted by Olga, removed by Mom

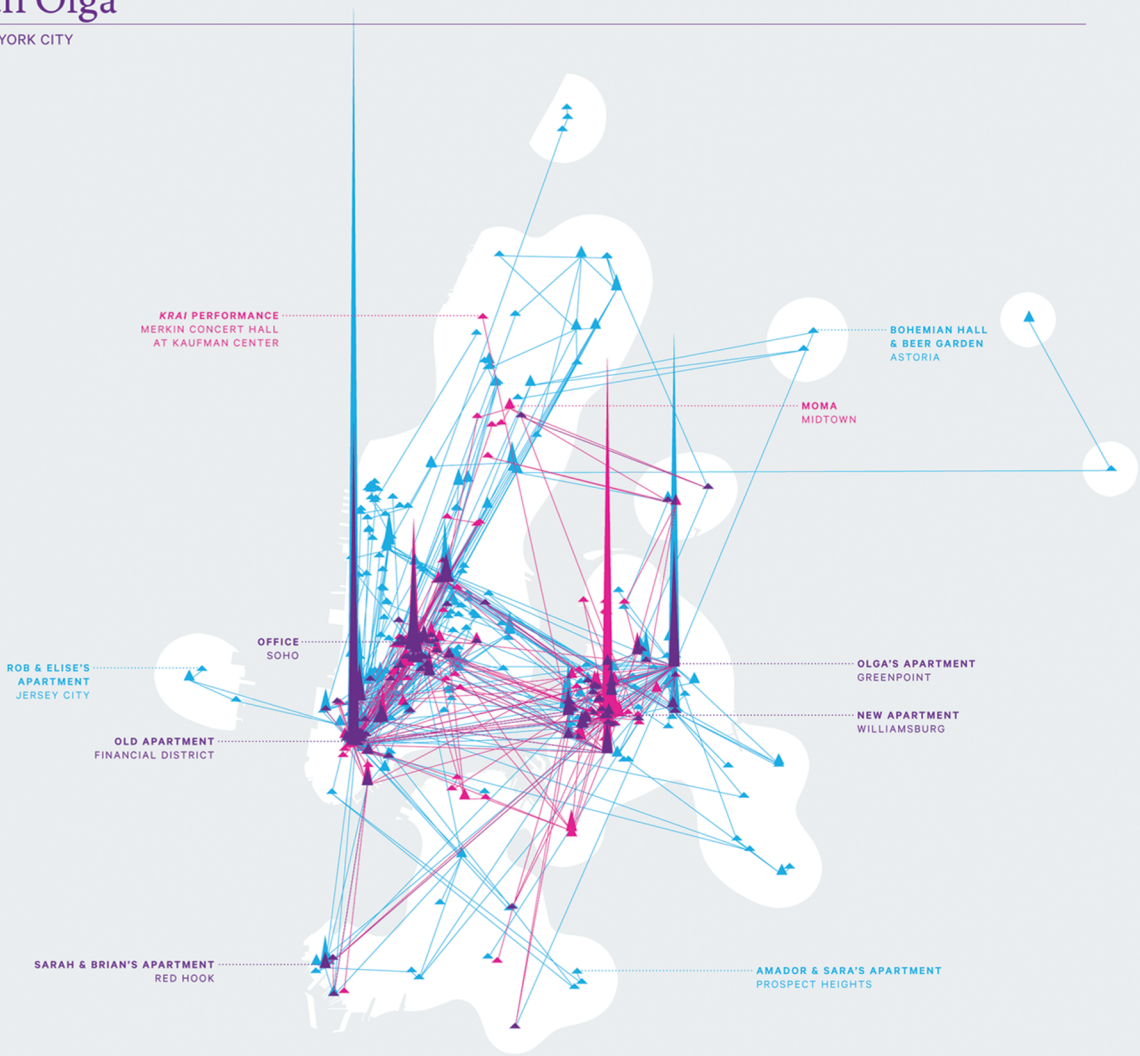
## MOST VISITED RESTAURANTS TOGETHER

**Le Garage, Picante and Sushi Ran**

Each visited twice

# With Olga

IN NEW YORK CITY




## DAYS TOGETHER IN NEW YORK CITY

136¾

Approximately 72% of total time together

## MOST VISITED NYC PLACES

- **OLD APARTMENT** — 194 VISITS
- **OLGA'S APARTMENT** — 84 VISITS
- **NEW APARTMENT** — 67 VISITS
- **THE OFFICE** — 35 VISITS
- **TAKAHACHI TRIBECA** — 21 VISITS

## TIME TOGETHER IN NEW YORK CITY


## TIME IN NEW YORK SPENT WITH OLGA

31%

5% of time together spent in transit

## MOST VISITED NYC RESTAURANTS

- **TAKAHACHI TRIBECA** — 21 VISITS
- **LES HALLES ON JOHN STREET** — 9 VISITS
- **DINER / ENID'S** — 7 VISITS
- **MILLER'S TAVERN / FIVE LEAVES** — 6 VISITS
- **RABBIT HOLE** — 5 VISITS

## FAVORITE NYC COCKTAIL WITH OLGA

Bloody Mary

22 servings

## NYC PERFORMANCES WITH OLGA

Twenty-Eight

Bell (11), Bear in Heaven (3), Baths + How to Dress Well + Zola Jesus, Blonde Redhead + Pantha du Prince, Dexter Lake Club Band, Jason Nazary, Knights on Earth, Olga Bell *Krai*, Little Women, Nathan Fake + Four Tet, Now Ensemble + Matmos, Owen Pallett, Panda Bear, Pierre-Laurent Aimard, Sleigh Bells and *The Nose*

## SIGNIFICANT NYC MISHAPS

Five

Abandoned keyboard stand, muddled dinner invitation date, missed ferry, shattered martini glass and smashed iPhone

Es gibt einen Eintrag mit der Überschrift »Trennung«, zu dem Clark schreibt: »Meine langjährige Freundin und ich haben uns getrennt – unmittelbar, bevor ich wegzog. Dies waren die Momente, in denen ich wirklich Schwierigkeiten hatte, mich damit abzufinden, dass ich wegziehen musste.« Zwei kleine Wege, einer innerhalb der Stadtgrenzen und einer außerhalb, und die Daten fühlen sich unglaublich persönlich an.

Dies ist vielleicht der Reiz an der »Quantified Self«-Bewegung, deren Ziel es ist, Technologie einzusetzen, um Daten über seine eigenen Aktivitäten und Verhaltensmuster zu sammeln. Manche Menschen verfolgen ihr Gewicht, andere, was sie essen und trinken oder wann sie zu Bett gehen. Ihr Ziel ist dabei, gesünder und länger zu leben. Andere verfolgen eine größere Vielfalt an Kennzahlen, weil sie mehr von sich erfahren wollen, als sie im Spiegel sehen. Die Sammlung persönlicher Daten wird am Ende des Tages so etwas wie ein Tagebuch der Selbstreflexion.

Nicholas Felton ist eine der bekannteren Persönlichkeiten auf diesem Gebiet, da er Jahresberichte über sich selbst veröffentlicht, die sowohl seine gestalterischen Fähigkeiten als auch seine disziplinierte Sammlung persönlicher Daten hervorhebt. Er verfolgt nicht nur seinen Standort, sondern auch, mit wem er seine Zeit verbringt, in welchen Restaurants er isst, welche Filme er anschaut, welche Bücher er liest, und viele andere Dinge mehr, die er jedes Jahr offenlegt. Abbildung 1.7 zeigt eine Seite aus Feltons Jahresbericht für 2010/2011.

2005 verfasste Felton seinen ersten Jahresbericht, dem seitdem jährlich weitere folgen. Jeder Bericht ist für sich genommen schön anzuschauen und stillt eine merkwürdige Sehnsucht, am Leben eines Fremden teilzuhaben. Was ich jedoch am interessantesten finde, ist die Entwicklung seiner Berichte zu einer persönlichen Darstellung sowie die immer größer werdende Datenfülle. Wenn man seinen ersten Bericht ansieht (Abbildung 1.8), dann hat man das Gefühl, dass es sich eher um eine Gestaltungsübung handelt, in die Spuren von Feltons Persönlichkeit eingebettet sind. Im Grunde genommen geht es jedoch hauptsächlich um Zahlen. Mit jedem Jahr fühlen sich die Daten jedoch weniger wie ein Bericht, sondern eher wie ein Tagebuch an.

Dies ist am offensichtlichsten in seinem *Jahresbericht 2010*. Feltons Vater war im Alter von 81 Jahren gestorben. Anstatt sein eigenes Jahr zusammenzufassen, entwarf Felton einen Jahresbericht (Abbildung 1.9), der das Leben seines Vaters basierend auf Kalendern, Dias, Postkarten und anderen persönlichen Dingen katalogisierte. Obwohl die Person im Fokus hier wiederum ein Fremder sein kann, ist es leicht, die Zahlen mit Gefühlen zu verbinden.

Wenn Sie Arbeiten wie diese sehen, dann ist es einfach, den Wert persönlicher Daten für den Einzelnen zu verstehen, und vielleicht, aber nur vielleicht, ist es gar nicht so verrückt, Datenschnipsel über sich selbst zu sammeln. Die Daten sind für Sie vielleicht nicht direkt nützlich, aber vielleicht sind sie es in zehn Jahren. So wie es nützlich ist, wenn einem zufällig wieder ein altes Tagebuch aus der Jugendzeit in die Hände fällt. Die Erinnerung hat ihren Wert. Auf vielerlei

**Abbildung 1.7** (vorhergehende Doppelseite) Eine Seite aus dem Jahresbericht 2010/2011 von Nicholas Felton, http://feltron.com

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach




**Abbildung 1.8** Ausgewählte Seiten aus dem Jahresbericht 2005 von Nicholas Felton, http://feltron.com

Arten protokollieren Sie ja bereits Bruchstücke Ihres Lebens, wenn Sie soziale Netze wie Twitter, Facebook oder Foursquare nutzen. Eine Statusaktualisierung oder ein Tweet sind wie eine winzige Momentaufnahme von dem, was Sie gerade in einem bestimmten Augenblick tun. Ein geteiltes Foto mit einem Zeiteindruck kann in einigen Jahrzehnten vielleicht viel bedeuten, und beim Check-in werden ihre digitalen Bits auf Dauer in die reale Welt gestellt.

Sie haben ja bereits gelesen, wie wertvoll diese Daten für den Einzelnen sein können. Was aber, wenn Sie aggregierte Daten vieler Einzelpersonen betrachten? Das statistische Bundesamt der USA erhebt alle zehn Jahre die offiziellen Bevölkerungszahlen im gesamten Land. Die Daten sind eine wertvolle Quelle

**Abbildung 1.9** (folgende Doppelseite) Ausgewählte Seiten aus dem Jahresbericht 2010 von Nicholas Felton, http://feltron.com

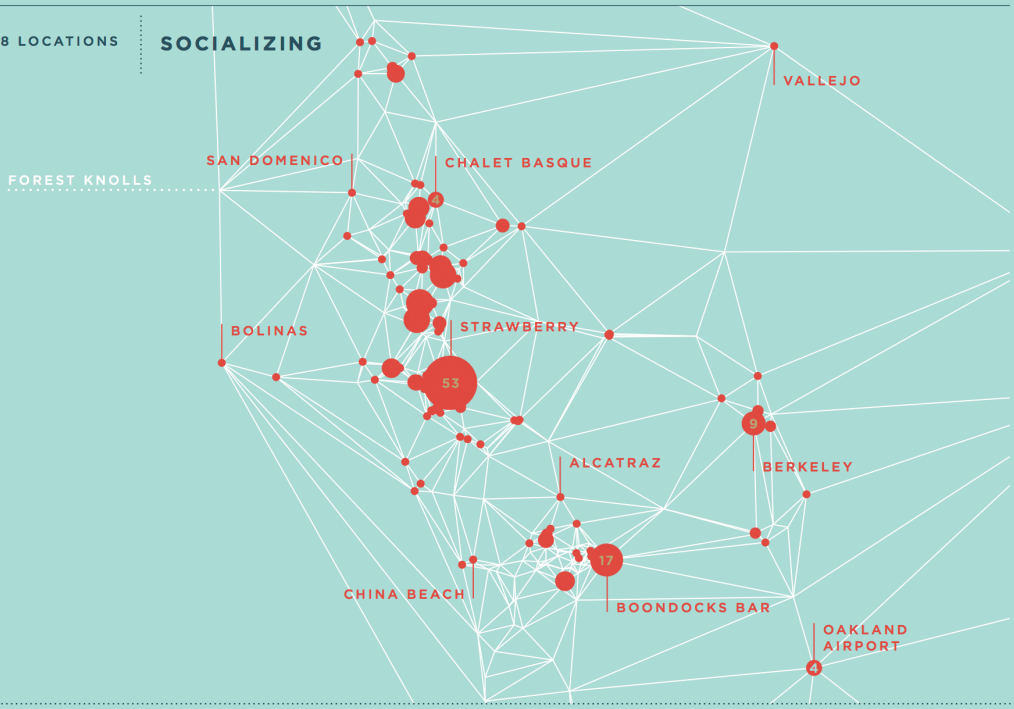
WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach

2000–2010
SAN RAFAEL

# THE 21ST CENTURY

138 LOCATIONS

## SOCIALIZING




**JAN 5, 2001**
71 YEARS,
6 MONTHS
AND 1 DAY

PERSON SEEN THE MOST
**MARINA**
117 TIMES

BLACK PANTHERS MET
**ONE**
BOBBY SEALE

WALKS RECORDED
**THIRTY-FIVE**
AND 1 HIKE

**2009–2010**
GOLDEN GATE TOLL BOOTH PREFERENCE
**LANE 6**
14 VISITS

ENTERTAINMENT

- **123** MOVIES
- **45** MUSIC
- **38** LECTURES
- **29** DANCES
- **29** POKER
- **22** PLAYS
- **19** TELEVISION
- **8** SLIDE SHOWS
- **5** ACROBATS
- **3** MIMES

MOST WATCHED TV SHOW
**THE OSCARS**
8 TIMES

LAST DAY
**SEP 12, 2010**
81 YEARS, 2 MONTHS
AND 8 DAYS OLD

WEATHER
SEP 12, 2010
3:20 PM
**49.8° F AND OVERCAST**
LARKSPUR, CALIFORNIA

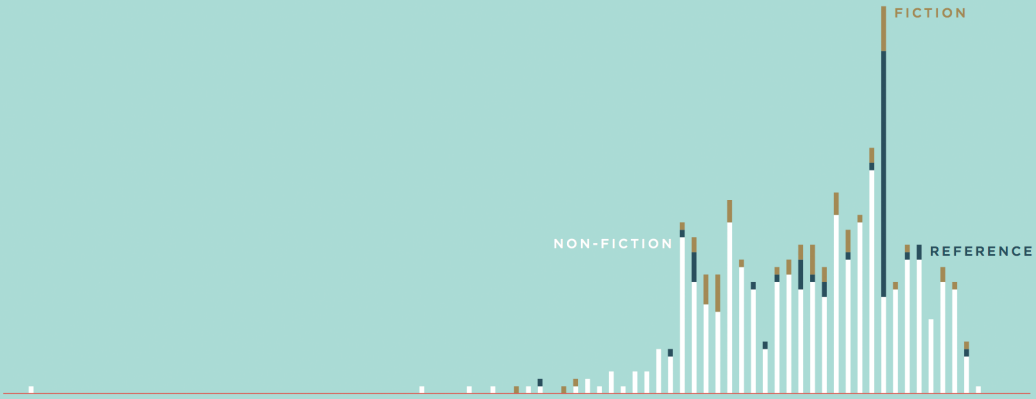
WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach

1848–2009

# BOOKS

536 BOOKS

## DATE PUBLISHED




BOOKS

**561**

SPANNING 161 YEARS

MOST REPRESENTED AUTHOR

**MARTIN GILBERT**
5 BOOKS

TYPES OF BOOKS

- **88** TRAVEL
- **77** HISTORY
- **42** MACHINES
- **37** GEOGRAPHY
- **32** ENCYCLOPEDIA
- **26** RELIGION
- **23** HEALTH
- **23** NOVEL
- **22** SCIENCE
- **17** HOW-TO

BIOGRAPHIES

**8**

FROM ARMSTRONG TO STALIN

MEDIAN PUBLISHING DATE

**1983**
11 BOOKS

REGION WITH MOST TRAVEL BOOKS

**RUSSIA**
6 BOOKS

TRAVEL BOOKS FOR UNVISITED PLACES

**SIX**

AUSTRALIA, ICELAND, GREENLAND, IRAN, PAKISTAN AND VENEZUELA

COOKBOOKS

**FIVE**

WAR-RELATED BOOKS

**51**

35 BOOKS ABOUT WORLD WAR 2

ELEVATOR BOOKS

**TWELVE**
1941–1991

HOW-TO TOPICS

**FOURTEEN**

BICYCLES, CLEANING, CROSS COUNTRY SKIING, DOING EVERYTHING RIGHT, HANDICRAFT, HOME REPAIR, PEST CONTROL, PHOTOGRAPHY, PREVENTING AND SURVIVING FIRES, SAILING, SURVIVAL AND TAI CHI

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



bei der Zuweisung staatlicher Mittel. Und von Zensus zu Zensus helfen die Fluktuationen in der Bevölkerung aufzuzeigen, wohin die Menschen innerhalb des Landes ziehen, wie sich die Zusammensetzung der Nachbarschaft verändert und wie Gegenden wachsen oder schrumpfen. Die Daten zeichnen also ein Bild der Menschen, die in den USA leben. Die Daten, die die Regierung erhebt und pflegt, können nur bestimmte Dinge über die Einzelpersonen aussagen. Wer die Menschen tatsächlich sind, ist schwer zu fassen.

Was sind ihre Vorlieben und Abneigungen? Was für eine Mentalität haben sie? Gibt es bedeutende Unterschiede zwischen benachbarten Städten und Gemeinden? Der Medienkünstler Roger Luke DuBois hat in seinem Werk *A More Perfect Union* eine andere Art von Zensus gewählt – und zwar 19 Millionen Kontaktprofile. Wenn Sie sich bei einer Partnerbörse anmelden, dann beschreiben Sie sich zunächst selbst: wer Sie sind, woher Sie kommen und was Ihre Interessen sind. Nachdem Sie diese Informationen etwas widerwillig angegeben haben und sich vielleicht entschlossen haben, die eine oder andere Information nicht preiszugeben, beschreiben Sie, wie Ihr Wunschpartner sein soll. Nach den Worten von DuBois erzählen Sie im letzteren Fall die Wahrheit, während Sie im ersteren lügen. Wenn man also die Kontaktprofile dieser Menschen zusammenfasst, dann erhält man eine Kombination daraus, wie Menschen sich selbst sehen und wie sie gesehen werden wollen.

In *A More Perfect Union* kategorisiert DuBois Kontaktprofile, digitale Zusammenfassungen von Hoffnungen und Träumen, nach Postleitzahlen und suchte dann nach dem Wort, das für jedes Gebiet am charakteristischsten war. Mithilfe einer Kopie einer Straßenkarte ersetzte DuBois jeden Städtenamen durch dieses besondere Wort dieser Stadt und zeichnete damit ein anderes Bild der USA – ein besser erkennbares und privateres.

Abbildung 1.10 zeigt Südkalifornien, das Reich der Tonfilme. Hier tauchen Wörter wie *acting*, *writer* und *entertainment* (Schauspielerei, Schriftsteller und Unterhaltung) auf. In Washington D.C. hingegen (Abbildung 1.11) sind es Wörter wie *bureaucrat*, *partisan* und *democratic* (Bürokrat, Parteigänger und demokratisch). Die meisten sind Berufe, aber in manchen Gebieten beschreiben die Wörter persönliche Eigenschaften, wichtige Dinge oder wesentliche Ereignisse.

In Louisiana (Abbildung 1.12) stechen Wörter wie *Cajun* und *curvy* (kurvig) ins Auge, ebenso wie *crawfish* (Languste), *bourbon* und *gumbo* (Okra). Allerdings war in New Orleans das charakteristischste Wort *flood* (Flut), eine Folge der Auswirkungen des Hurrikans Katrina 2005.

Menschen werden anhand gewöhnlicher demografischer Daten definiert, etwa nach Rasse, Alter oder Geschlecht, aber sie identifizieren sich auch mit dem, was sie gerne in ihrer Freizeit tun, was ihnen passiert ist oder mit wem sie gerne ausgehen. Das Großartige an *A More Perfect Union* ist, dass man dies auf einem landesweiten Maßstab in den Daten sehen kann.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



Dieses Gefühl – wo Datenpunkte Erinnerungen und Berichte Porträts und Tagebücher sind – gibt es auch in Feltons Berichten, in Clarks Atlas und in Pareckis GPS-Spuren. Statistiker und Entwickler nennen dies Analysen. Für Künstler und Designer ist es Storytelling – das Erzählen einer Geschichte. Für das Extrahieren der Informationen aus den Daten – um zu verstehen, was in den Zahlen steckt – sind Analyse und Storytelling jedoch ein und dasselbe.

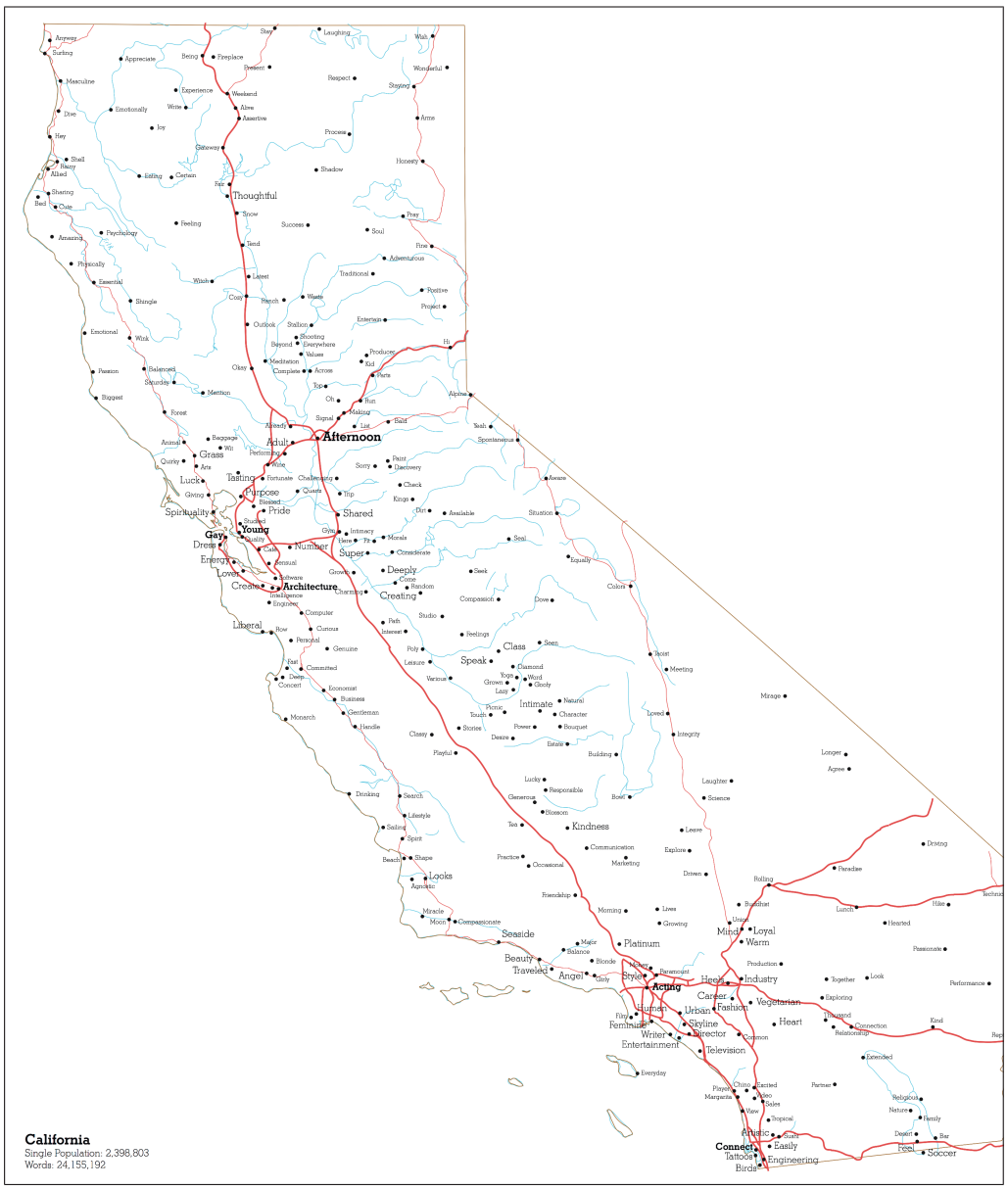



**Abbildung 1.10** Karte von Kalifornien aus »A More Perfect Union« (2011) von R. Luke DuBois, mit freundlicher Genehmigung des Künstlers und der bitforms gallery, New York City, http://perfect.lukedubois.com

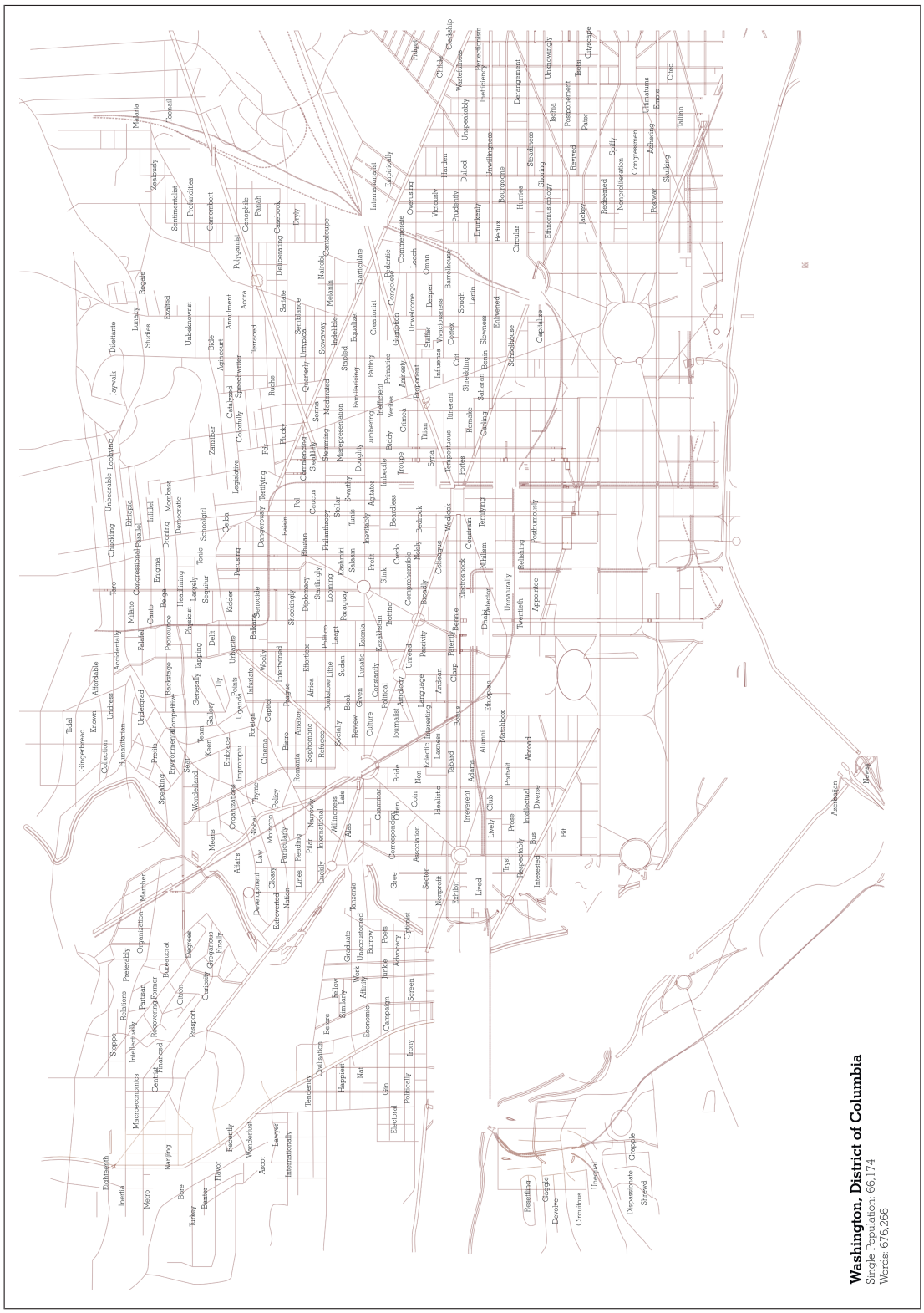



**Abbildung 1.11** Karte von Washington, D.C. aus »A More Perfect Union« (2011)

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



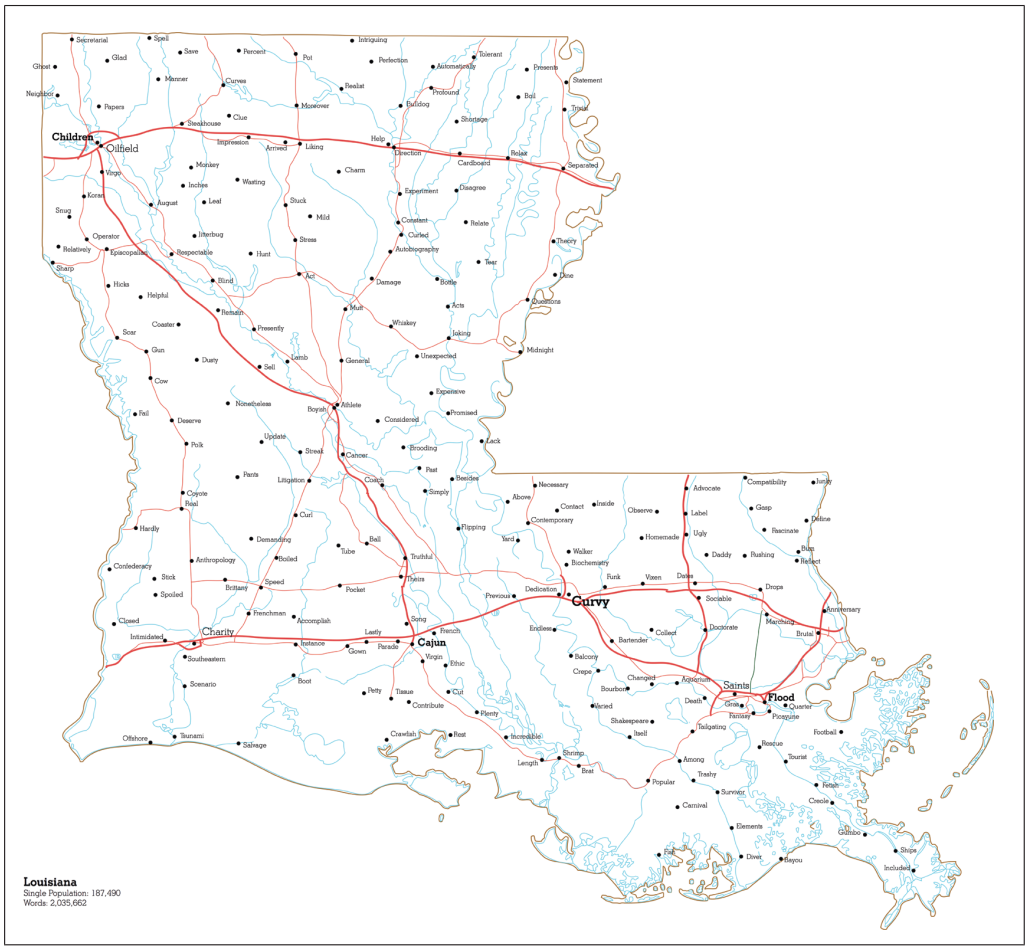



**Abbildung 1.12** Karte von Louisiana aus »A More Perfect Union« (2011)

Ebenso wie das, was es darstellt, können Daten aufgrund von Variabilität und Unsicherheitsfaktoren komplex sein. Wenn Sie jedoch alles im richtigen Kontext betrachten, dann wird es allmählich plausibel.

# Variabilität

In einer Kleinstadt in Deutschland begibt sich der Physiker und Amateurfotograf Kristian Cvecek nachts mit seiner Kamera in den Wald. Mit langen Belichtungszeiten fängt er auf seinen Bildern die Bewegungen der Glühwürmchen ein, die zwischen den Bäumen tanzen. Das Insekt (Abbildung 1.13) ist winzig und tagsüber kaum wahrnehmbar. Doch in der Dunkelheit ist es nur schwer zu übersehen.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



**Abbildung 1.13** Ein Glühwürmchen in der Nacht von Kristian Cvecek, http://quit007.deviantart.com/gallery

Obwohl für einen Beobachter jeder Moment im Flug wie ein beliebiger Flug im Raum erscheint, ergibt sich aus Cveceks Fotos ein Muster (Abbildung 1.14). Es scheint, als würden sich die Glühwürmchen mit einem vorbestimmten Ziel den Pfad entlangbewegen und um die Bäume kreisen.

Es gibt jedoch auch Zufälligkeiten. Sie können raten, wohin ein Glühwürmchen aufgrund seiner Flugbahn als Nächstes hinfliegt, aber wie sicher sind Sie dabei? Ein Glühwürmchen kann jeden Moment nach links, rechts, oben oder unten abbiegen. Diese Variabilität, die jeden Flug zu etwas Einzigartigem macht, führt dazu, dass das Beobachten von Glühwürmchen Spaß macht und das Bild so schön ist. Der Pfad ist das Entscheidende an der Sache. Der Endpunkt, der Startpunkt und die durchschnittliche Position sind nur halb so wichtig.

Bei Daten finden Sie Muster, Trends und Zyklen, aber es ist nicht immer (eigentlich sogar selten) ein gerader Pfad von Punkt A zu Punkt B. Gesamtwerte, Mittelwerte oder sonstige aggregierte Messergebnisse können interessant sein, aber sie sind nur Teil der Geschichte. Die Fluktuationen in den Daten könnten hingegen das Interessanteste und Wichtigste sein.

Zwischen 2001 und 2010 gab es laut der National Highway Traffic Safety Administration (NHTSA) in den USA 363.839 tödliche Autounfälle. Zweifellos wiegt dieser Gesamtwert von mehr als einem Drittel einer Million schwer, da er für die verlorenen Leben von noch mehr als diesen steht. Wenn Sie sich ausschließlich auf diese eine Zahl konzentrieren (Abbildung 1.15), dann werden Sie nachdenklich oder denken vielleicht sogar über Ihr eigenes Leben nach.

Gibt es jedoch irgendetwas, das Sie aus diesen Daten lernen können – außer dass Sie vorsichtig fahren sollten? Die NHTSA stellt Daten bis hin zu Einzelunfällen bereit, die angeben, wann und wo jeder Unfall passiert ist. Somit können Sie genauer hinsehen.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



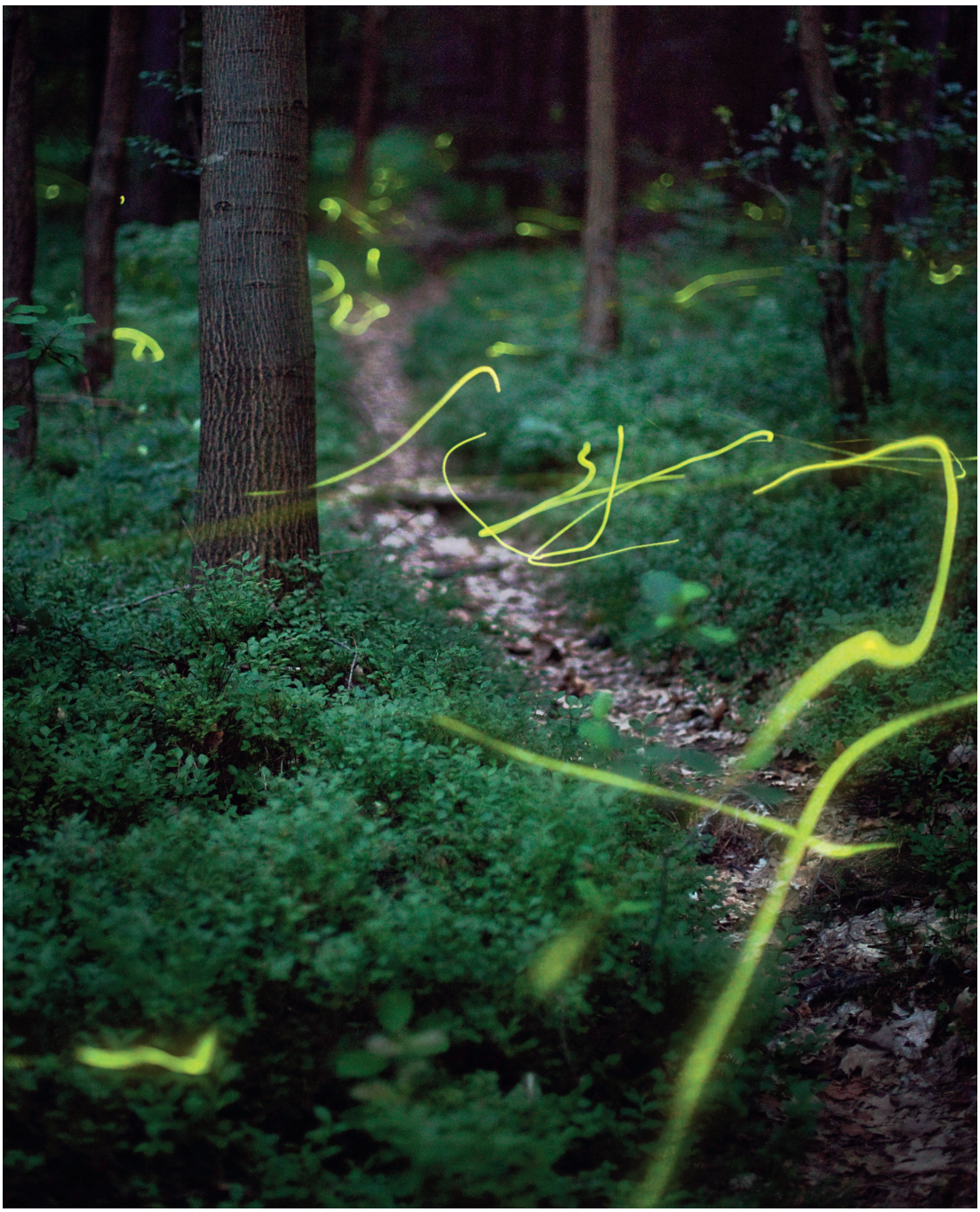


**Abbildung 1.14** Weg eines Glühwürmchens von Kristian Cvecek, http://quit007.deviantart.com/gallery

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach




**Abbildung 1.15** Ein Gesamtwert

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



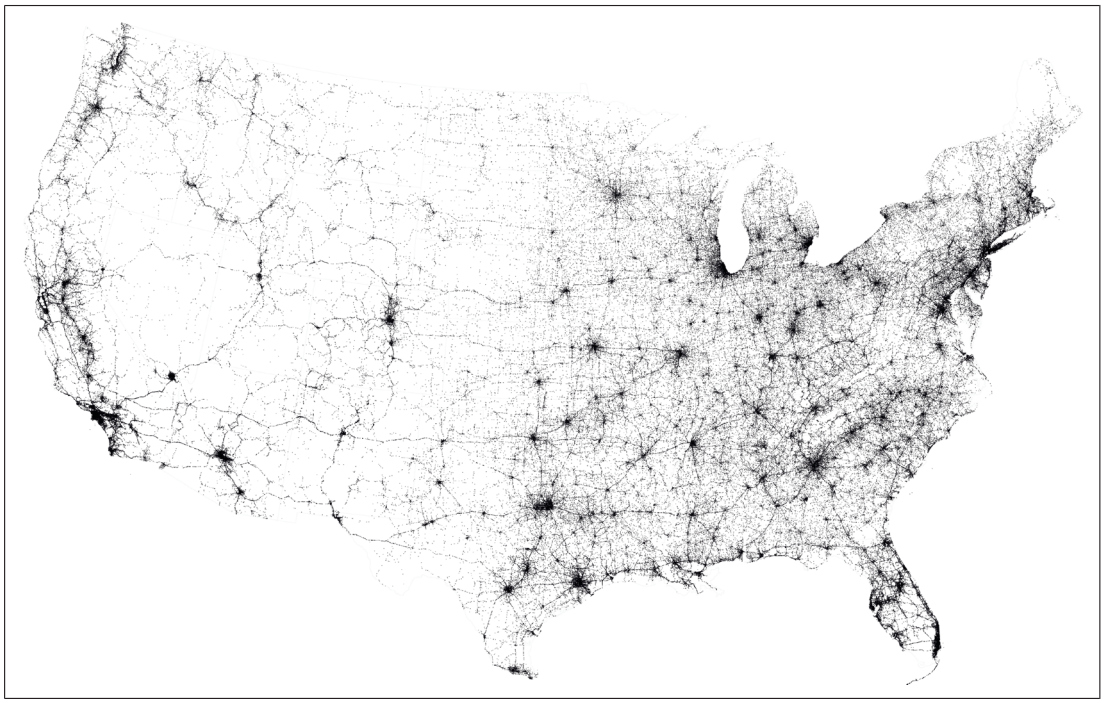


**Abbildung 1.16** Alles auf einen Blick

Abbildung 1.16 zeigt jeden tödlichen Unfall im Gebiet der USA zwischen 2001 und 2010. Jeder Punkt steht für einen Unfall. Erwartungsgemäß gibt es eine höhere Konzentration von Unfällen in großen Städten und auf stark frequentierten Highways. Es gibt weniger Unfälle, wo es weniger Menschen und Straßen gibt. Obwohl dies nicht auf die leichte Schulter genommen werden sollte, sagt die Karte mehr über das Straßennetz des Landes aus als über die Unfälle.

Ein Blick auf die Unfälle im Zeitverlauf verlagert den Fokus auf die Ereignisse an sich. Abbildung 1.17 zeigt beispielsweise die Anzahl der Unfälle pro Jahr, was eine ganz andere Geschichte erzählt als die Gesamtanzahl in Abbildung 1.15. Noch immer ereignen sich jährlich Zehntausende von Unfällen, aber es gab einen signifikanten Rückgang zwischen 2006 und 2010 und auch die Zahl der tödlichen Unfälle pro 100 Millionen Fahrzeuge (nicht dargestellt) ging zurück.

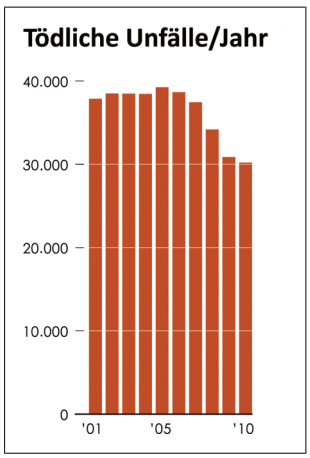



**Abbildung 1.17** Tödliche Unfälle pro Jahr

Saisonale Zyklen werden bei einer monatlichen Granularität sichtbar (Abbildung 1.18). Die Ereignisse steigen steil an während der Sommermonate,

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



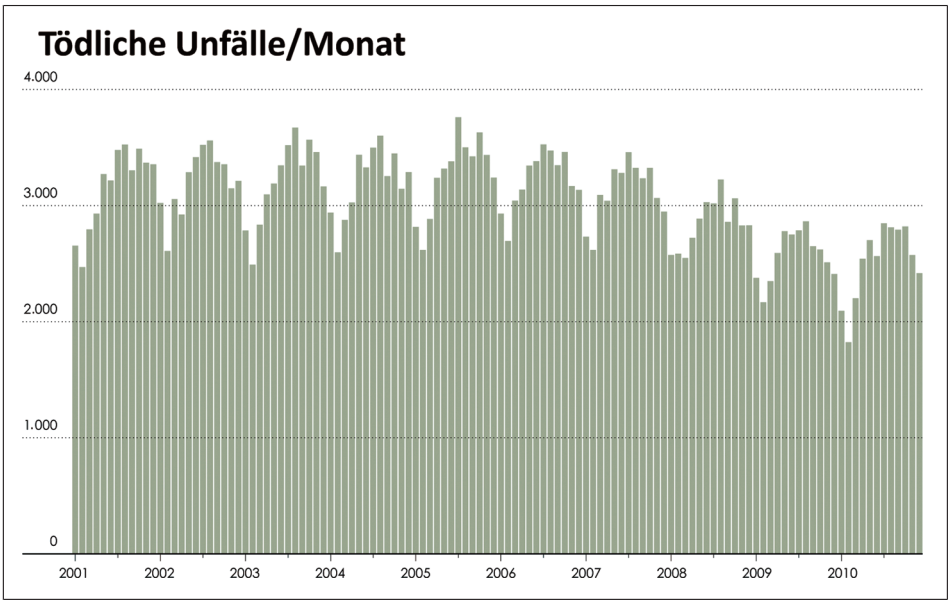



**Abbildung 1.18** Tödliche Unfälle pro Monat

wenn alle in den Urlaub fahren und mehr Zeit im Freien verbringen. Im Winter hingegen fahren weniger Menschen, also gibt es weniger Unfälle. Das ist jedes Jahr gleich. Dennoch sieht man gleichzeitig den jährlichen Gesamtrückgang zwischen 2006 und 2010.

Es gibt jedoch eine Variabilität, wenn Sie bestimmte Monate über die Jahre hinweg vergleichen. So ereigneten sich beispielsweise 2001 die meisten Unfälle im August mit einem kleinen, relativen Rückgang im darauffolgenden Monat. Das passierte ebenso zwischen 2002 bis 2004. Zwischen 2005 und 2007 waren die meisten Unfälle jedoch im Juli. Zwischen 2008 und 2010 war es dann wieder der August.

Auf der anderen Seite ereigneten sich im Februar, dem Monat mit den wenigsten Tagen, jedes Jahr die wenigsten Unfälle – mit Ausnahme von 2008. Es gibt also saisonale Schwankungen und Schwankungen innerhalb der Saison.

Wenn Sie die täglichen Unfälle ansehen (Abbildung 1.19), dann sehen Sie eine noch größere Variabilität, die aber nicht nur Rauschen ist. Es gibt noch immer ein Muster mit Höhen und Tiefen. Obwohl die saisonalen Muster schwieriger auszumachen sind, erkennen Sie einen wöchentlichen Zyklus mit mehr Unfällen während der Wochenenden als zur Wochenmitte. Der Spitzentag in jeder Woche schwankt zwischen Freitag, Samstag und Sonntag.

Sie können die Granularität auch auf Unfälle pro Stunde erhöhen. Abbildung 1.20 zeigt die genauen Daten. Jede Zeile stellt ein Jahr dar. Somit zeigt jede Zelle in dem Raster eine Stundenzeitreihe des jeweiligen Monats.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



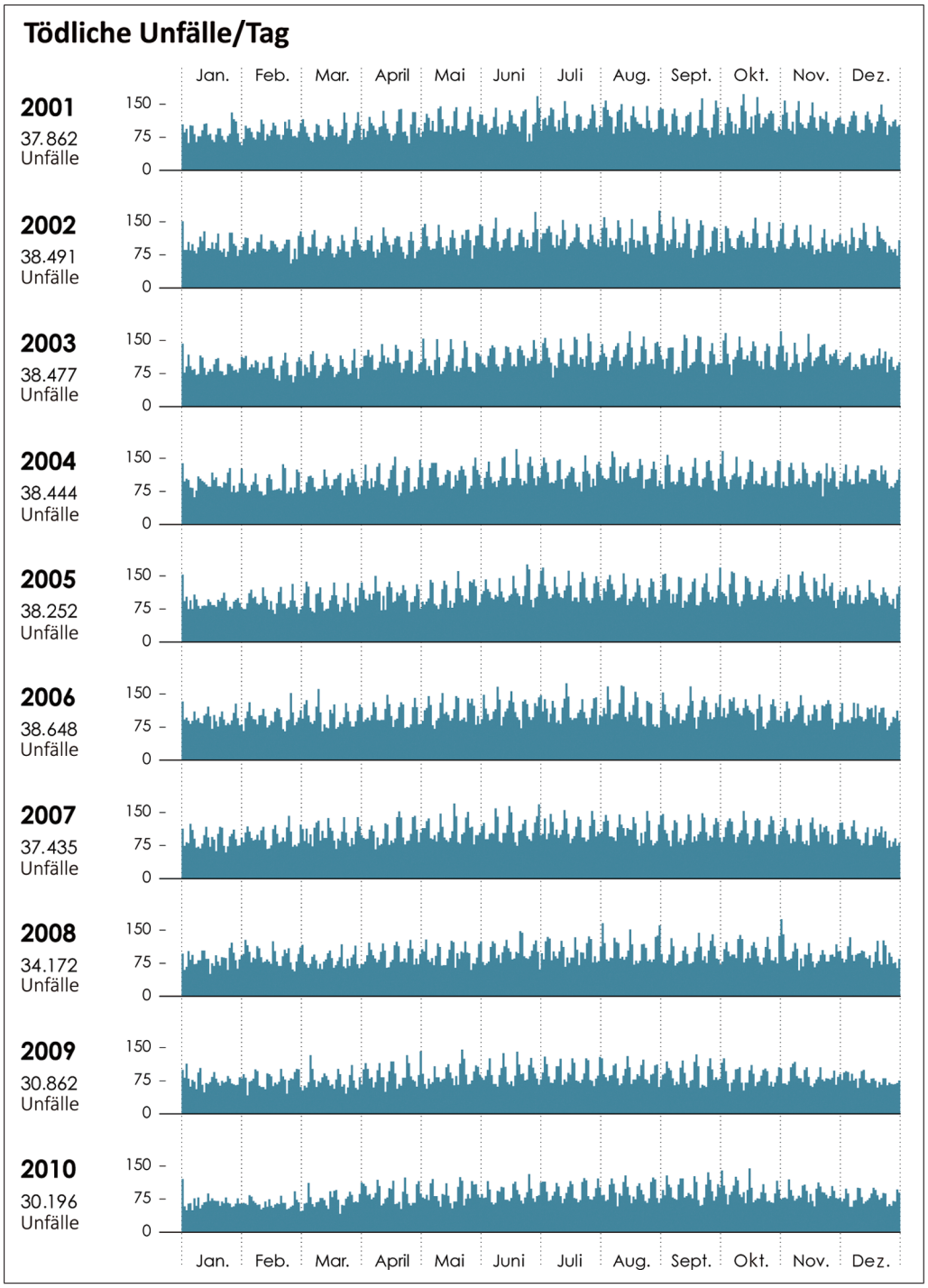



**Abbildung 1.19** Tödliche Unfälle pro Tag

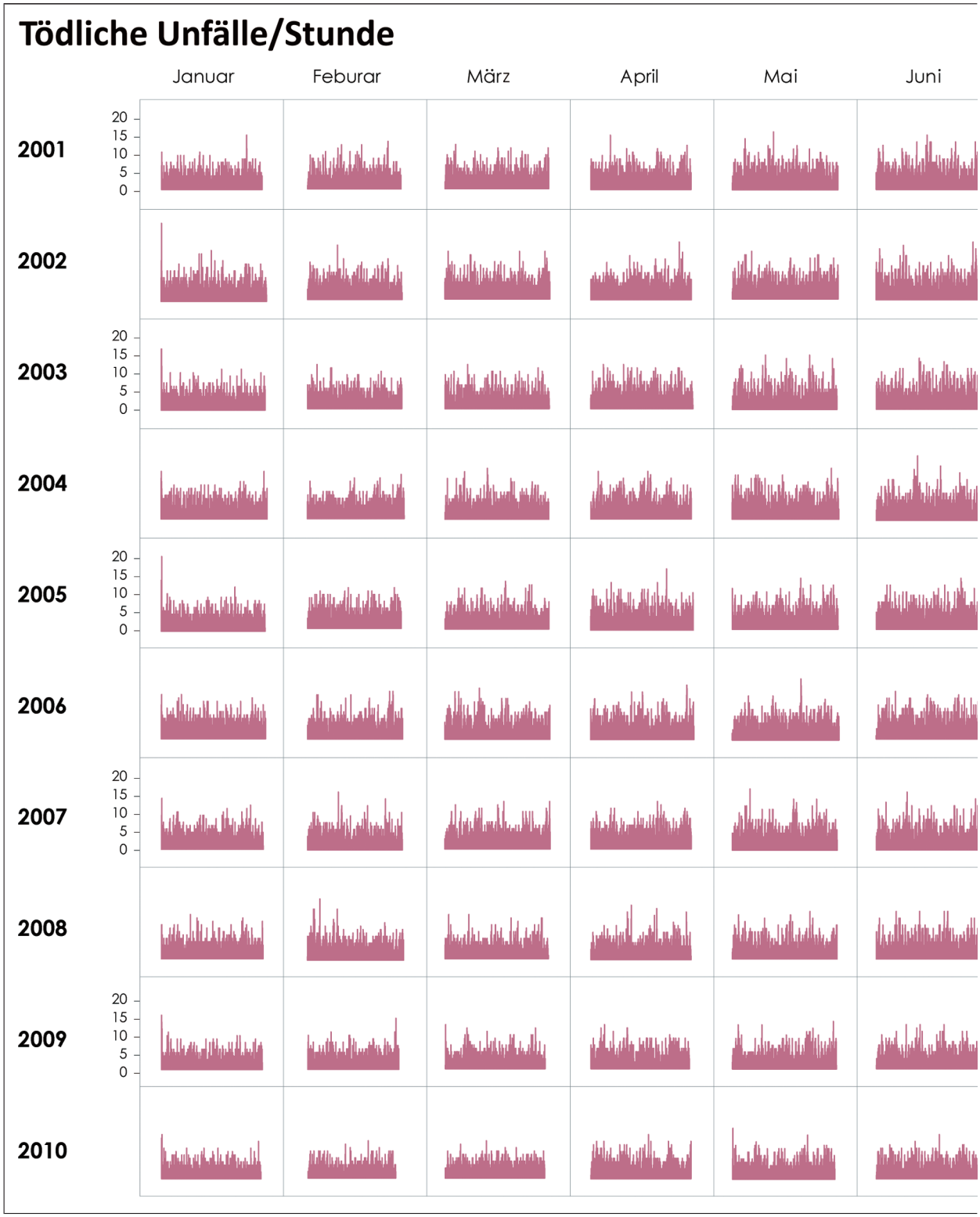



**Abbildung 1.20** Tödliche Unfälle pro Stunde

WILEY
Yau, Visualisierung
c01

Kühn & Weyh Satz&Medien

12. 3. 2014

M. Kaltenbach

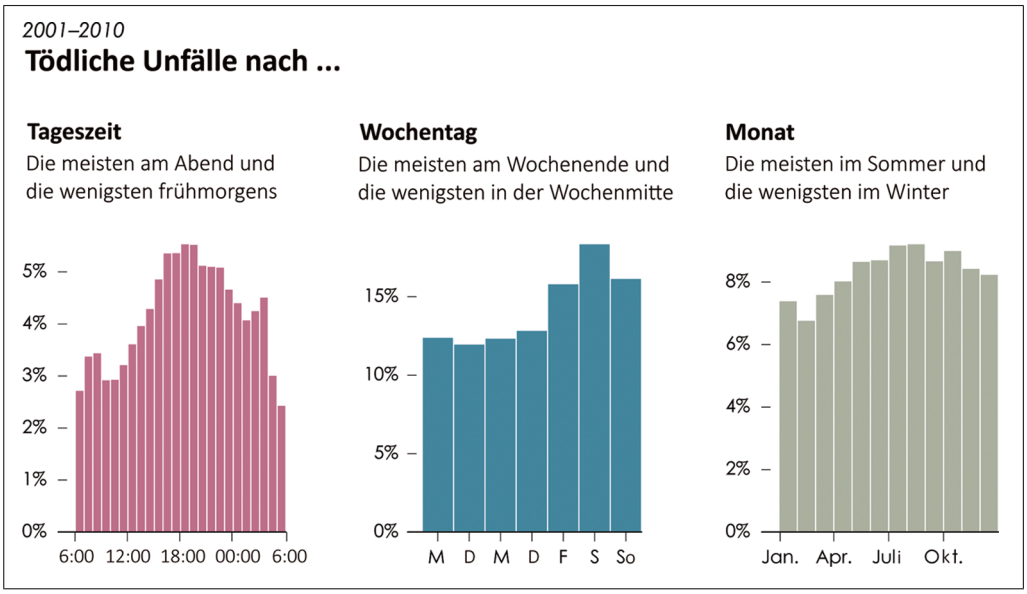



**Abbildung 1.21** Verteilung der Unfälle im Zeitverlauf

Mit Ausnahme einer Spitze an Neujahr um Mitternacht lässt sich auf dieser Ebene aufgrund der Variabilität nur schwer ein Muster ausmachen. Das monatliche Diagramm ist ebenfalls schwer zu interpretieren, wenn man nicht weiß, wonach man suchen soll. Es gibt jedoch klare Muster, wenn die Daten aggregiert werden wie in Abbildung 1.21. Anstatt Werte jede Stunde, jeden Tag oder jeden Monat anzuzeigen, können Sie bestimmte Zeitsegmente aggregieren, um die Verteilungen zu erforschen.

Was schwer zu unterscheiden war oder zuvor wie Rauschen aussah, ist nun leicht zu erkennen. Es gibt einen kleinen Anstieg während des morgendlichen Berufsverkehrs, aber die meisten tödlichen Unfälle passieren am Abend nach der Arbeit. Wie Abbildung 1.19 zeigt, gibt es mehr Unfälle während des Wochenendes, aber in der Zusammenfassung wird dies deutlicher. Sie können auch saisonale Muster erkennen, allerdings deutlicher mit einer größeren Anzahl an Unfällen im Sommer als im Winter.

Es hat also seinen Wert, sich die Daten jenseits von Mittelwert, Median oder Gesamtwert anzusehen, weil Ihnen diese Messwerte nur einen kleinen Teil der Geschichte erzählen. Meistens verdecken jedoch aggregierte Zahlen oder Werte, die einfach nur angeben, wo die Mitte einer Verteilung ist, die interessanten Details, auf die Sie sich eigentlich konzentrieren sollten, und zwar sowohl zur Entscheidungsfindung als auch zum Erzählen von Geschichten.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



Ein abweichender Wert, der aus der Menge heraussticht, könnte etwas sein, das Sie bereinigen müssen oder dem Sie sich stärker widmen sollten. Vielleicht sind die Veränderungen im Zeitverlauf ein Signal, dass etwas Gutes (oder Schlechtes) im System passiert. Zyklen oder regelmäßige Vorkommnisse könnten beim Blick in die Zukunft helfen. Manchmal hilft es jedoch nicht, zu viel Variabilität zu sehen. Dann drehen Sie die Granularität auf Verallgemeinerungen und Verteilungen zurück. Sie verlieren diese Informationen – die Filetstücke –, wenn Sie aus zu weiter Entfernung auf die Daten blicken.

Betrachten Sie es mal aus dieser Warte: Wenn Sie auf Ihr Leben zurückblicken, würden Sie sich dann lieber daran erinnern, wie Ihre Tage so durchschnittlich verliefen, oder sind es eher die Höhen und Tiefen, die am wichtigsten sind? Ich wette, es ist eine Kombination aus beidem.

## Unsicherheit

Viele Daten sind Schätzungen und keine absoluten Endzahlen. Ein Analyst betrachtet den Beweis (beispielsweise eine Stichprobe) und stellt dann eine auf Tatsachen beruhende Vermutung über die Grundgesamtheit an. Diese wohlbegründete Vermutung birgt auch eine gewisse Unsicherheit. Sie tun dies ständig in Ihrem Alltag. Sie mutmaßen etwas, weil Sie es wissen, gelesen oder von irgendjemandem gehört haben. Und Sie können mit einer gewissen (eher ungefähren) Sicherheit sagen, dass Sie recht haben. Sind Sie absolut sicher oder haben Sie eigentlich keinen Schimmer? Genauso ist das mit den Daten.

**Hinweis**

Es ist verlockend, Daten als absolute Wahrheit zu betrachten, weil wir mit Zahlen Fakten assoziieren. Meistens sind Daten jedoch einfach eine wohlbegründete Vermutung. Ihr Ziel ist die Verwendung von Daten, denen keine große Unsicherheit anhaftet.

Nach meinem Bachelorstudium (Ingenieurwesen mit Nebenfach Statistik) hatte ich neun Monate Zeit bis zum Beginn meines Masterstudiums. Ich nahm ein paar Jobs an, bei denen ich etwas mehr als Mindestlohn verdiente. Sie waren aber todlangweilig, daher beschäftigte sich mein Gehirn zwangsläufig mit spannenderen Dingen.

Eines Tages dachte ich mir: »Ich kenne mich ein wenig mit Statistik und Wahrscheinlichkeitsrechnung aus, und ein Deck Karten habe ich auch. Ich werde ein versierter Blackjack-Spieler werden, so wie die Jungs vom MIT. Schluss mit diesem langweiligen Job. Ich werde reich werden.« Daraufhin begann meine vierwöchige Blackjack-Besessenheit. (Um Sie nicht auf die Folter zu spannen: Ich bin nicht reich geworden und es ist keineswegs so spannend, wie es im Kino aussieht.)

Falls Sie dieses Spiel nicht kennen, dann hier eine Kurzeinführung: Es gibt einen Geber und einen Spieler. Der Geber gibt jedem zwei Karten (eine seiner Karten liegt verdeckt vor ihm). Die Summe aller Karten muss nun möglichst nahe an 21 Punkte herankommen, ohne diese zu übersteigen. Sie können weitere Karten (*hit*) oder keine mehr verlangen (*stay*). Manchmal

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



kann man seine beiden Karten auch teilen (*split*) und mit »geteilter Hand« weiterspielen. Sie können auch Ihren Wetteinsatz verdoppeln. Je mehr Sie setzen, umso mehr können Sie gewinnen. Wenn Sie jedoch mit Ihren Karten den Wert 21 überschreiten (*bust*), dann haben Sie automatisch verloren. Wenn nicht, werden weitere Karten ausgegeben oder auch nicht, und wer am nächsten an den 21 Punkten dran ist, hat gewonnen.

Das Spiel ist so ausgelegt, dass der Geber im Vorteil ist, aber wenn Sie im richtigen Moment Karten nehmen oder passen, dann verringert sich dieser Vorteil. Diese Regeln basieren auf Durchschnittswerten, aber wie Ihnen jeder sagen kann, der schon einmal Blackjack gespielt hat, gibt es bei jedem Blatt eine Unsicherheit. Sie können trotzdem verlieren, auch wenn Sie den richtigen Zug machen. Stellen Sie sich vor, dass Sie eine 5 und eine 6 bekommen, was zusammen 11 macht, und der Geber zeigt eine 6. Der richtige Zug wäre, zu verdoppeln, weil es für Sie unmöglich ist, mit einer weiteren Karte 21 zu überschreiten. Die Chance, 21 Punkte zu erreichen, ist jedoch nicht schlecht. Es kann auch durchaus sein, dass der Geber mit der gezeigten 6 die 21 überschreitet.

Sie verdoppeln also und bekommen eine 3, was zusammen 14 ergibt. Ach, herrje. Nicht gut. Ihre einzige Hoffnung ist jetzt, dass der Geber übers Ziel hinausschießt. Er dreht also seine verdeckte Karte um. Es ist eine 10, sodass er insgesamt 16 hat. Den Regeln nach muss er eine Karte ziehen. Es ist eine 5. Der Geber hat also insgesamt 21. Verloren! Hätten Sie nicht verdoppelt, dann hätten Sie nur halb so viel Geld verloren, als wenn Sie regulär gespielt hätten. Aber wenn das Gewinnen so leicht wäre, dann würde kein Kasino dieses Spiel anbieten.

Es besteht auf beiden Seiten Unsicherheit, weil Sie gegen die Verteilung spielen, oder anders ausgedrückt, Sie kennen nur die ungefähre Wahrscheinlichkeit beim Ziehen der Karten. Sie wissen unter Umständen, welche Karten im Spiel sind, aber Sie können nur vermuten, welche Karte als Nächstes kommt.

**Hinweis**

Wenn Sie Karten zählen oder verfolgen, welche Karten noch im Stapel sind, dann verändert sich die Wahrscheinlichkeit, wenn Sie Ihren Einsatz aufgrund Ihres Vorteils verändern, aber die Unsicherheit bleibt.

Unsicherheit trifft natürlich auch auf Dinge außerhalb der Karten zu und hat vielerlei Gesichter. Das Wetter ist hier ein gutes Beispiel. Wie oft haben Sie sich vor dem Kofferpacken die Wettervorhersage für den nächsten Tag oder die nächste Woche angesehen, nur um letztendlich zu erkennen, dass das Wetter nicht wie erwartet war?

Oder die Anzeige im Auto, auf der man ablesen kann, wie lange man mit der momentanen Tankfüllung noch fahren kann? Ich war mit meiner Frau beim Einkaufen und laut Anzeige hätte ich noch schätzungsweise 16 Meilen fahren können. Nach Hause waren es aber 18 Meilen. Was tun? Anstatt nun an der nächsten Tankstelle anzuhalten, fuhr ich zu der, die am nächsten von meinem Zuhause entfernt ist. Die Anzeige zeigte zwei Meilen lang, dass wir nur noch null Meilen fahren könnten. Aber es hat

gereicht! (Zum Glück! Denn irgendjemand hat konsequent darauf bestanden, dass ich das Auto würde schieben müssen.)

Wiegen Sie sich mehrmals. Unter Umständen zeigt die Waage ständig etwas anderes an. Das ist ganz normal, obwohl ein paar Sekunden atmen weder zu Gewichtsverlust noch Gewichtszunahme führt. Die Akkuanzeige im Laptop kann in Stundenschritten springen, auch wenn nur ein paar Minuten vergangen sind. Aus dem Lautsprecher wird verkündet, dass die nächste U-Bahn in zehn Minuten ankommt, doch sie kommt in elf. Eine Lieferung wird für Montag angekündigt, doch sie kommt erst am Mittwoch.

Wenn Sie Daten haben, die eine Serie aus Mittelwerten und Medianen sind, oder eine Sammlung aus Schätzwerten, die auf einer Stichprobe basieren, dann sollten Sie sich stets Gedanken über deren Unsicherheit machen.

Dies ist besonders wichtig, wenn gewichtige Entscheidungen, die Millionen betreffen, auf Schätzungen basieren, beispielsweise bei nationalen oder globalen Bevölkerungszahlen. Die Erstellung und Finanzierung von Programmen basiert häufig auf diesen Zahlen, daher kann bereits ein sehr geringer Fehler einen großen Unterschied machen.

**Hinweis**

Zahlen scheinen konkret und absolut zu sein, aber Schätzungen bergen Unsicherheiten. Daten sind eine Abstraktion dessen, was sie repräsentieren, und der Grad an Genauigkeit variiert.

Das statistische Bundesamt der USA veröffentlicht Daten über das Land zu Themen wie Migration, Armut und Wohnraum, die sich auf Stichproben aus der Grundgesamtheit beziehen. (Dies unterscheidet sich vom 10-Jahres-Zensus, bei dem jeder Einwohner der USA gezählt werden soll.) Jeder Schätzwert hat eine Fehlerspanne, daher liegt der tatsächliche Zählwert oder Prozentsatz wahrscheinlich innerhalb eines gewissen Bereichs. Abbildung 1.22 zeigt beispielsweise die Schätzwerte zu den Haushalten. Die Fehlerspanne bei den Gesamthaushalten beträgt fast eine viertel Million.

| | Schätzungen | Fehlerspanne |
|---|---|---|
| **Haushalte - gesamt** | 114.235.996 | ± 248.114 |
| **Familien - gesamt** | 76.254.318 | ± 230.785 |
| **Familiengröße - Durchschnitt** | 3,17 | ± 0,01 |
| **Familienhaushalte mit Ehepaaren** | 56.655.412 | ± 293.638 |
| **Verheiratet, 15 Jahre und länger** | 50,2% | ± 0,2 |
| **Geschieden, 15 Jahre und länger** | 10,5% | ± 0,1 |

Quelle: 2010 American Community Survey

**Abbildung 1.22** Haushaltsschätzungen 2010

Stellen Sie sich ein Behältnis mit Kaugummikugeln vor, in das Sie nicht hineinsehen können. Sie wollen aber wissen, wie viele von jeder Farbe darin sind. (Warum sollten Sie sich für die Verteilung von Kaugummikugeln interessieren? Keine Ahnung. Setzen Sie Ihre Fantasie ein. Sie sind ein Kaugummiexperte, der für eine Kaugummifabrik arbeitet. Sie wetten mit Ihrem eingebildeten Statistikkollegen, dass in Ihrem Prozess in jedem Behältnis die Kugeln gleichmäßig verteilt sind. Es geht also um Stolz und Bargeld.)

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



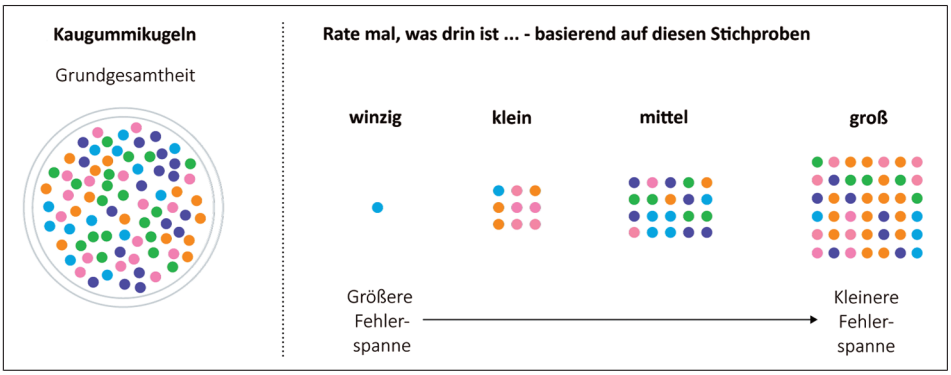



**Abbildung 1.23** Kaugummikugeln und Fehlerspanne

Wenn Sie alle Kaugummikugeln auf den Tisch ausleerten und zählten, dann müssten Sie nicht raten, denn dann würden Sie die Gesamtzahl erhalten. Sie können aber auch nur eine Handvoll herausnehmen, dann müssten Sie raten, wie viele Kugeln in dem Behältnis sind – ausgehend von der Menge in Ihrer Hand. Je größer die Handvoll, umso leichter fällt das Raten, denn man hat zweifellos eine bessere Abbildung des Gesamtinhalts. Sie könnten natürlich genauso gut nur eine Kugel herausnehmen. Es wäre dann allerdings schwieriger herauszufinden, was sonst noch in dem Behältnis ist.

Bei nur einer Kaugummikugel wäre die Fehlerspanne hoch. Bei einer großen Hand voller Kugeln wäre die Fehlerspanne niedriger. Und wenn Sie alle Kugeln zählten, dann wäre die Fehlerspanne gleich null.

Wenn Sie das auf Millionen von Kaugummikugeln in Tausenden unterschiedlich großen Behältern mit verschiedenen Verteilungen und großen und kleinen Handvoll anwenden, wird die Schätzung immer komplexer. Ersetzen Sie dann Kaugummis durch Menschen, die Behälter durch Städte, Gemeinden und Länder sowie die Handvoll durch beliebig verteilte Umfragen – schon erhält ein Mittelwert mit einer Fehlerspanne mehr Gewicht.

Laut Gallup waren 48 Prozent der Amerikaner mit Obamas Job zwischen dem 11. und 13. Juni 2012 nicht einverstanden. Es gab jedoch eine Fehlerspanne von drei Prozent, was den Unterschied zwischen einem mangelnden Einverständnis von mehr als der Hälfte und weniger als der Hälfte der Bevölkerung bedeutet. Während des Wahlkampfes schätzen Meinungsumfragen, welcher Kandidat vorne liegt. Wenn die Fehlerspanne groß ist, können die Ergebnisse mehr als eine Person vorne sehen, was irgendwie den Zweck von Meinungsumfragen zunichtemacht.

Schätzungen werden schwierig, wenn Sie Personen, Orte oder Dinge einstufen, vor allem wenn Sie Messungen kombinieren (und statische Modelle mit zahlreichen Variablen erstellen). Nehmen Sie beispielsweise die Evaluation von Bildungseinrichtungen, die ständig untersucht wird. Städte, Schulen und Lehrer werden häufig miteinander verglichen, aber wie wird eine

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



gute Erziehung definiert oder was macht eine ganze Stadt schlau? Ist es der Anteil an Abiturienten? Oder der Anteil an Studierenden? Ist es die Anzahl an Universitäten, Bibliotheken oder Museen pro Kopf? Wenn es all das ist, wiegt dann eine Zählung mehr als die andere? Oder wiegen alle gleich schwer? Die Antworten verändern sich, je nachdem, wen Sie fragen. Und ebenso ist es mit den Bewertungen.

2011 gab das Bildungsministerium von New York City einen sogenannten Lehrerdatenbericht heraus, der die Unterrichtsqualität zu messen versuchte. Die Berichte wurden ursprünglich nur an Schulen und Lehrer weitergegeben, wurden aber Anfang 2012 öffentlich zugänglich gemacht. Die Schätzungen haben mehrere Faktoren berücksichtigt. Einer der wesentlichen war jedoch die Veränderung der Testperzentilen von der siebten zur achten Klasse.

**Hinweis**

Meine Heimatstadt galt aufgrund einer Veröffentlichung, die hier ungenannt bleiben soll, als die »dümmste« Stadt der USA. Die Einstufungen waren Schätzungen, die auf Werten mit zweifelhafter Unsicherheit beruhten.

So kam es, dass die Lehrerin Carolyn Abbott als schlechteste Mathelehrerin der Stadt bekannt wurde – mit einer Bewertung auf der nullten Perzentile. Ihre Siebtklässler erzielten jedoch Werte auf der 98. Perzentile. Wie konnte das sein?

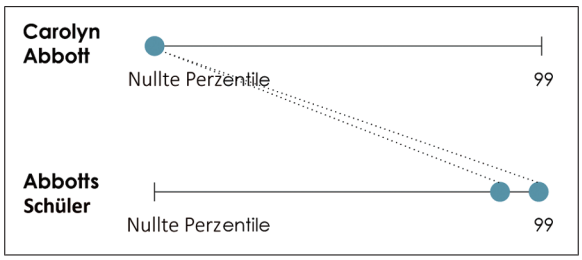



**Abbildung 1.24** Carolyn Abbotts Bewertung im Vergleich zu der ihrer Schüler

Die Schüler sollten in der 8. Klasse auf der 97. Perzentilen landen, doch stattdessen landeten sie auf der 98. Perzentilen, was gemäß dem Statistikmodell kein Fortschritt war. Die meisten stimmten überein, dass Schüler mit einem schlechten Lehrer diese Werte sicherlich nicht erhalten würden. Das Problem ist, dass es bei den Lehrerbewertungen Unsicherheiten und Variabilität gibt. Eine Bewertung stellt eine Verteilung von Lehrern dar, die aufgrund von Schätzungen, denen eine gewisse Unsicherheit anhaftet, in eine Rangfolge gebracht werden. Dennoch werden die Bewertungen als absolut betrachtet. Ein allgemeines Publikum würde dieses Konzept nicht verstehen, daher liegt es an Ihnen, es deutlich zu kommunizieren.

Wenn Sie nicht bedenken, was Ihre Daten wirklich darstellen, dann ist es ein Leichtes, diese versehentlich falsch zu interpretieren. Sie müssen immer Unsicherheitsfaktoren und Variabilität einbeziehen. An dieser Stelle kommt der Kontext ins Spiel.

## Kontext

Wenn Sie sich den nächtlichen Himmel ansehen, dann sehen die Sterne aus wie Punkte auf einer flachen Oberfläche. Das Fehlen visueller Tiefe macht die Übertragung des Himmels auf Papier relativ einfach, daher kann man sich auch die Sternbilder leichter vorstellen. Man ver-

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



bindet einfach die Punkte. Obwohl man Sterne so wahrnimmt, als seien sie alle gleichermaßen weit weg, sind sie eigentlich verschiedene Lichtjahre weit entfernt.

Wenn Sie weiter als die Sterne fliegen könnten, wie würden dann die Sternbilder aussehen? Darüber hat sich Santiago Ortiz Gedanken gemacht, als er die Sterne aus einer anderen Perspektive (Abbildung 1.25) dargestellt hat.

Die anfängliche Sicht zeigt die Sterne in einem globalen Layout, so wie wir sie sehen. Sie schauen durch die Sterne hindurch auf die Erde, allerdings so, als hätten sie alle den gleichen Abstand zu unserem Planeten. In der Vergrößerung sehen Sie die Sternbilder, so wie Sie sie von der Erde aus sehen, wenn Sie eingewickelt im Schlafsack in den Bergen liegen und in den sternenklaren Himmel blicken.

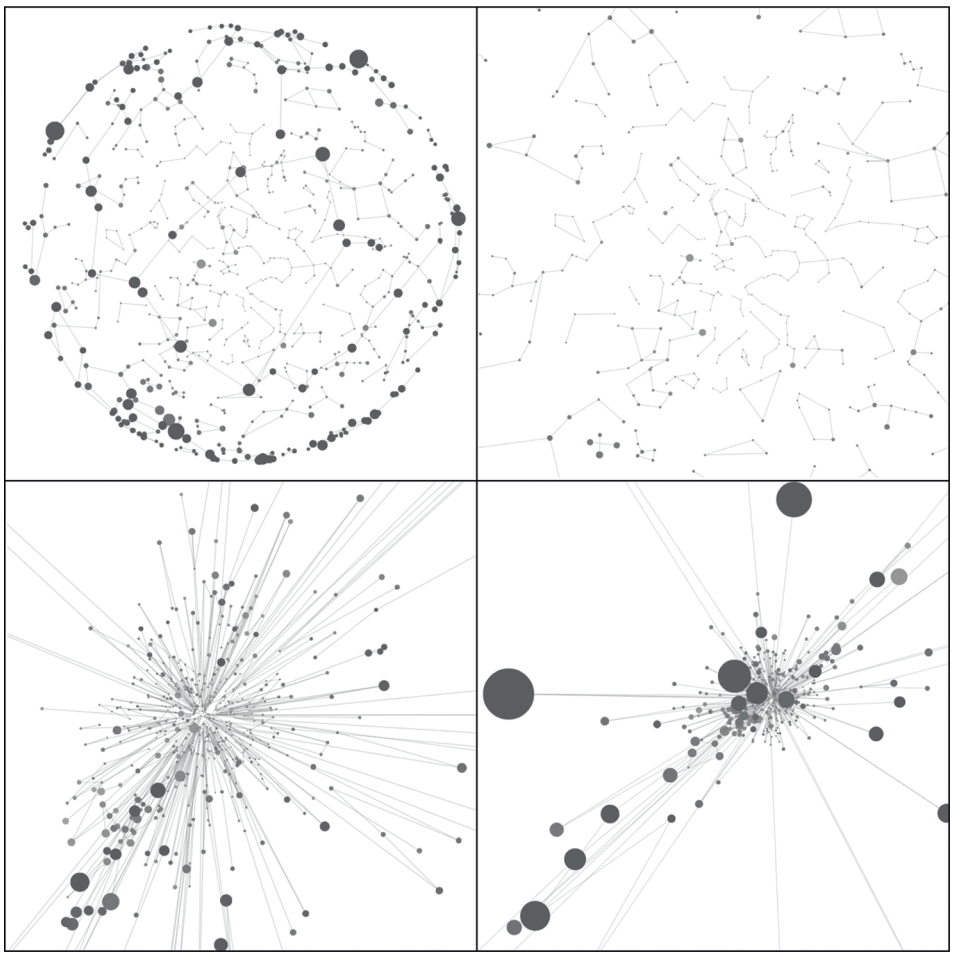


**Abbildung 1.25** »View of the Sky« von Santiago Ortiz, http://moebio.com/exomap/viewsofthesky/2/

Es macht Spaß, die wahrgenommene Sicht zu betrachten, aber wenn Sie die Perspektive ändern, um die tatsächliche Entfernung zu zeigen, wird es interessant. Der Sternenwechsel und die deutlichen Sternbilder sind praktisch nicht mehr erkennbar. Die Daten sehen aus diesem neuen Blickwinkel anders aus.

Das kann Kontext erzielen. Er kann Ihre Perspektive eines Datensatzes vollständig ändern. Er kann bei der Entscheidung helfen, was die Zahlen darstellen und wie sie zu interpretieren sind. Nachdem Sie wissen, worum es bei den Daten geht, hilft Ihnen Ihr Verstand, die faszinierenden Teile zu finden, die zu einer wertvollen Datenvisualisierung führen.

Ohne Kontext sind Daten nutzlos und jede Visualisierung, die Sie damit erstellen, wird ebenfalls nutzlos sein. Wenn man Daten verwendet, ohne außer den Werten selbst etwas über sie zu wissen, dann ist das so, als würde man aus zweiter Hand eine gekürzte Version eines Zitats hören und dieses dann in einem Aufsatz als wesentlichen Diskussionspunkt zitieren. Das mag in Ordnung sein, aber Sie riskieren, dass Sie später herausfinden, dass der Sprecher das Gegenteil von dem meinte, was Sie dachten. Sie müssen die Metadaten oder die Daten über die Daten kennen – also Wer, Was, Wann, Wo und Wie –, bevor Sie wissen können, worum es bei den Zahlen geht.

**Wer:** Ein Zitat in einer großen Zeitung ist von größerer Bedeutung als eines, das in einem Prominentenklatschblatt erscheint, das dafür bekannt ist, die Wahrheit großzügig auszulegen. Entsprechend implizieren Daten aus einer vertrauenswürdigen Quelle in der Regel eine stärkere Genauigkeit als eine zufällige Online-Umfrage.

Die Gallup Organization, die seit den 1930ern die öffentliche Meinung misst, ist vertrauenswürdiger als beispielsweise jemand (möglicherweise ich), der mit einer kleinen, einmaligen kurzzeitigen Twitter-Stichprobe spätnachts experimentiert. Während Gallup darauf bedacht ist, Stichproben zu erstellen, die für eine Region repräsentativ sind, sind es bei den Twitter-Stichprobennehmern Unbekannte.

Doch es ist nicht nur wichtig, wer die Daten sammelt, sondern auch, um wen es sich bei den Daten handelt. Wie in dem Kaugummibeispiel ist es häufig finanziell nicht durchführbar, Daten über jeden und alles in einer Grundgesamtheit zu sammeln. Die meisten Menschen haben keine Zeit, Tausend, geschweige denn eine Million, Kaugummikugeln zu zählen und zu kategorisieren, daher nehmen sie Stichproben. Entscheidend ist, gleichmäßig über die Grundgesamtheit verteilt Stichproben so zu sammeln, dass sie für das Ganze repräsentativ sind. Haben das die Datensammler getan?

**Wie:** Häufig wird auf eine Methodik verzichtet, weil sie zu komplex und für eine technisch versierte Zielgruppe gedacht ist, dennoch ist es sinnvoll, das Wesentliche darüber zu wissen, wie bestimmte Daten gesammelt wurden.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



Wenn Sie derjenige sind, der die Daten sammelt, dann ist alles in Ordnung, aber wenn Sie sich online einen Datensatz schnappen, vielleicht von jemandem, den Sie nicht kennen, wie wollen Sie dann wissen, ob dieser etwas taugt? Vertrauen Sie dem sofort oder forschen Sie nach? Sie müssen die genauen statistischen Modelle hinter jedem Datensatz nicht kennen, aber suchen Sie nach kleinen Stichproben, großer Fehlerspanne und unpassenden Annahmen über die Themen, etwa Indizes oder Rangfolgen, die uneinheitliche oder zusammenhangslose Informationen einbeziehen.

Manchmal werden Indizes erzeugt, um die Lebensqualität in Ländern zu messen und eine Kennzahl wie Alphabetisierung wird als ein Faktor verwendet. Ein Land verfügt vielleicht nicht über aktuelle Informationen zur Alphabetisierung, daher verwendet der Datenerheber einfach einen Schätzwert aus einer vorherigen Dekade. Das wird zu Problemen führen, weil dann der Index nur unter der Annahme funktioniert, dass die Alphabetisierungs-Quote vor einem Jahrzehnt mit der gegenwärtigen vergleichbar ist, was wahrscheinlich (eigentlich sicherlich) nicht der Fall ist.

**Was:** Sie wollen letztendlich auch wissen, worum es bei Ihren Daten geht, aber vorher sollten Sie wissen, was die Zahlen umgibt. Sprechen Sie mit Fachleuten, lesen Sie Fachaufsätze und jegliche Begleitdokumentation.

In den Einführungskursen zur Statistik lernt man in der Regel etwas über die Analysemethoden, beispielsweise über statistische Tests, Regression und Modellierung, aber in einem Vakuum, denn eigentlich sollen die Mathematik und die Konzepte vermittelt werden. Wenn es dann jedoch um die echten Daten geht, verlagert sich das Ziel auf das Sammeln von Informationen. Es verlagert sich von »Was steckt in den Zahlen?« zu »Was repräsentieren diese Daten in der Realität? Sind sie sinnvoll? Wie wirkt sich dies auf andere Daten aus?« Ein großer Fehler besteht darin, jeden Datensatz gleich zu behandeln und dieselben abgedroschenen Methoden und Werkzeuge zu verwenden. Machen Sie es anders!

**Wann:** Die meisten Daten sind irgendwie zeitlich verknüpft, weil es sich vielleicht um eine Zeitreihe oder eine Momentaufnahme eines bestimmten Zeitraums handelt. In beiden Fällen müssen Sie wissen, wann die Daten erfasst wurden. Eine Schätzung, die vor Jahrzehnten erhoben wurde, kann nicht mit einer aus der heutigen Zeit gleichgesetzt werden. Das scheint zwar offensichtlich, aber es passiert ganz häufig, dass alte Daten genommen und als neue weitergegeben werden, weil keine anderen verfügbar sind. Dinge ändern sich, Menschen ändern sich und Orte ändern sich, also ändern sich zwangsläufig auch Daten.

**Wo:** Dinge können sich in Städten und Staaten ebenso wie im Zeitverlauf ändern. Am besten vermeidet man beispielsweise globale Verallgemeinerungen, wenn die Daten nur aus einigen wenigen Ländern kommen. Dies gilt ebenso für digitale Orte. Daten von Websites wie Twitter oder Facebook verstecken das Verhalten ihrer Nutzer und sind daher nicht unbedingt auf die reale Welt übertragbar.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach




**Abbildung 1.26** »A History of the World in 100 Seconds« (Eine Geschichte der Welt in 100 Sekunden) von Gareth Lloyd, http://datafl.ws/24a

Obwohl die Lücke zwischen digital und real immer kleiner wird, ist der Zwischenraum noch immer offensichtlich. Eine animierte Karte, die die »Geschichte der Welt« darstellt und auf dem georeferenzierten Wikipedia basiert, zeigte in einem geografischen Raum für jeden Eintrag hervorspringende Punkte. Abbildung 1.26 zeigt das Ende des Videos.

Das Ergebnis ist beeindruckend und es gibt sicherlich eine Korrelation zur realen Zeitleiste. Es ist jedoch klar, dass die Karte, da Wikipedia-Inhalte in englischsprachigen Ländern bekannter sind, in jenen Gegenden mehr Punkte zeigt.

**Warum:** Zu guter Letzt müssen Sie wissen, warum die Daten gesammelt wurden – meistens nämlich als Plausibilitätsprüfung des Bias. Manchmal werden Daten gesammelt oder sogar fabriziert, weil es die Tagesordnung verlangt. In diesen Fällen ist Vorsicht geboten. Regierung und Wahlen sind unter Umständen das Erste, was einem dazu einfällt. Doch die sogenannten Informationsgrafiken im Internet, die mit Schlüsselwörtern versehen sind und von den Websites veröffentlicht werden, um Zugriffe zu bekommen und im Google-Ranking zu steigen, gehören mittlerweile auch zu den Übeltätern. (In meinen Anfängen als Blogger für FlowingData bin ich ein paar Mal darauf hereingefallen, aber ich habe meine Lektion gelernt.)

Bringen Sie so viel Sie können über Ihre Daten in Erfahrung – Ihre Analyse und Visualisierung wird besser werden. Danach können Sie das, was Sie wissen, an die Leser weitergeben.

Doch nur weil Sie Daten haben, bedeutet das nicht, dass Sie eine Grafik erstellen und diese der Welt mitteilen sollten. Kontext kann Ihnen helfen, eine Dimension – eine Informationsebene – zu Ihren Datengrafiken hinzuzufügen, aber manchmal ist es besser, sich zurückzuhalten, weil es das Richtigere ist.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



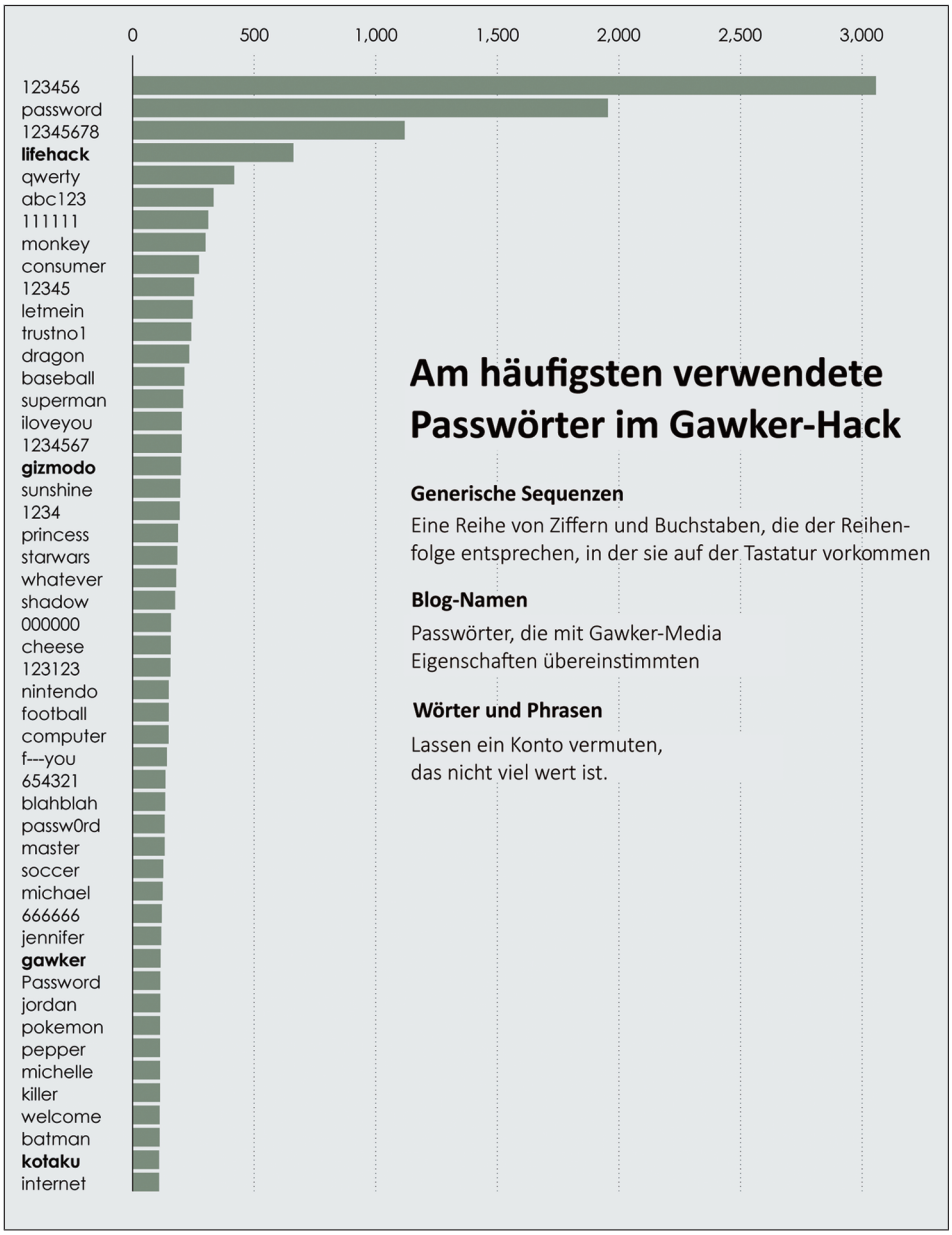



**Abbildung 1.27** Am häufigsten verwendete Passwörter im Gawker-Hack

2010 wurde Gawker Media, die große Blogs wie Lifehacker und Gizmodo betreiben, gehackt und 1,3 Millionen Nutzernamen und Passwörter drangen an die Öffentlichkeit. Sie konnten über BitTorrent heruntergeladen werden. Die Passwörter waren verschlüsselt, aber die Hacker

knackten 188.000, was mehr als 91.000 eindeutige Passwörter an die Öffentlichkeit brachte. Was würden Sie mit dieser Art von Daten tun?

Gemein wäre es, die Nutzernamen mit einfachen (also schwachen) Passwörtern hervorzuheben. Oder könnte man sogar eine Anwendung erstellen, die anhand eines Nutzernamens das Passwort errät? Man könnte vielleicht auch einfach die häufigen Passwörter hervorheben (Abbildung 1.27). Das bietet etwas Einblick in die Daten, ohne es zu leicht zu machen, sich in ein fremdes Konto einzuloggen. Es könnte auch als eine Warnung für andere dienen, damit sie ihre Passwörter in etwas weniger offensichtliche ändern. Sie wissen schon, irgendetwas mit mindestens zwei Symbolen, einer Ziffer und einer Mischung aus Klein- und Großbuchstaben. Die Regeln für Passwörter heutzutage sind lächerlich. Doch ich schweife ab.

Mit Daten wie dem Gawker-Satz ist eine tiefe Analyse unter Umständen interessant, aber sie könnte auch mehr Schaden als Nutzen anrichten. In diesem Fall ist Datenschutz wichtiger, daher ist es besser, die Dinge, die sie zeigen und betrachten, zu begrenzen.

Ob Sie Daten nutzen sollten, ist jedoch nicht immer klar umrissen. Manchmal ist die Trennung zwischen falsch und richtig nicht ganz deutlich, daher müssen Sie die Entscheidung treffen. Am 22. Oktober 2010 veröffentlichte Wikileaks, eine Online-Organisation, die Privatdokumente und Medien aus anonymen Quellen veröffentlicht, 391.832 geheime Dokumente der US-Armee, die heute als Iraq War Logs (Tagebuch des Irakkriegs) bekannt sind. Die Dokumente berichteten von 66.081 toten Zivilisten unter insgesamt 109.000 offiziell registrierten Toten in der Zeit von 2004 bis 2009.

Das Leck brachte Vorkommnisse von Missbrauch und falscher Berichterstattung ans Licht, beispielsweise dass zivile Opfer als »Feind im Kampf gefallen« klassifiziert wurden. Andererseits kann es ungerechtfertigt erscheinen, Ergebnisse über geheime Daten zu veröffentlichen, die man sich auf wenig feine Art beschafft hat.

Vielleicht sollte es eine goldene Regel für Daten geben: Behandele anderer Menschen Daten, wie du auch möchtest, dass man deine Daten behandelt.

Letztendlich geht es wieder darum, was die Daten darstellen. Daten sind eine Abstraktion der Realität und diese kann kompliziert sein, aber wenn Sie ausreichend Kontext gesammelt haben, dann können Sie sich zumindest ernsthaft bemühen, ihnen einen Sinn zu geben.

WILEY
Yau, Visualisierung
c01
Kühn & Weyh Satz&Medien
12. 3. 2014
M. Kaltenbach



## Zusammenfassung

Datenvisualisierung wird häufig als eine Übung in Grafikdesign oder als ein reines Problem der Computerwissenschaft betrachtet, aber die beste Arbeit ist stets in den Daten verankert. Um Daten zu visualisieren, müssen Sie wissen, was sie sind, was sie in der Realität darstellen und in welchem Kontext Sie sie interpretieren sollten.

Daten gibt es in verschiedenen Formen und Größen, mit unterschiedlichen Granularitäten. Unsicherheiten sind mit ihnen verbunden, was bedeutet, dass Gesamtsummen, Durchschnittswerte und Mediane nur ein kleiner Teil dessen sind, was ein Datenpunkt ist. Daten verbiegen sich. Daten drehen sich. Daten verändern sich. Daten können privat und sogar poetisch sein. Folglich gibt es auch Datenvisualisierungen in vielen Formen.