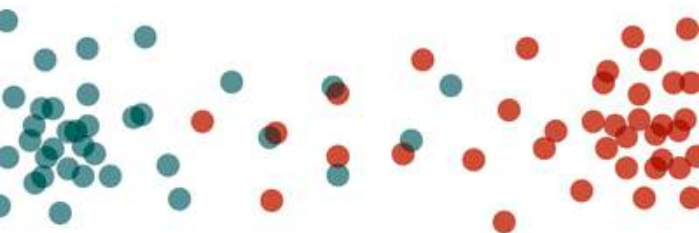


Die KUNST der STATISTIK



Was uns Daten wirklich sagen und wie
wir dies im Alltag nutzen können

David Spiegelhalter
Die Kunst der Statistik

David Spiegelhalter

Die Kunst der Statistik

Was uns Daten wirklich sagen und
wie wir dies im Alltag nutzen
können

Übersetzung aus dem Englischen
von Nikolas Bertheau

REDLINE | VERLAG

Bibliografische Information der Deutschen Nationalbibliothek

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

Für Fragen und Anregungen:
info@redline-verlag.de

1. Auflage 2020

© 2020

by Redline Verlag, ein Imprint der Münchner Verlagsgruppe GmbH,
Nymphenburger Straße 86
D-80636 München
Tel.: 089 651285-0
Fax: 089 652096

© der Originalausgabe by David Spiegelhalter 2019

Die englische Originalausgabe erschien 2019 bei Penguin Books LTD, einem Imprint von Penguin Random House UK unter dem Titel *The Art of Statistics*.

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlages reproduziert oder unter Verwendung elektronischer Systeme gespeichert, verarbeitet, vervielfältigt oder verbreitet werden.

Übersetzung: Nikolas Bertheau
Redaktion: Matthias Michel, Wiesbaden
Umschlaggestaltung: Laura Osswald, München
Satz: abavo GmbH, Buchloe
Druck: GGP Media GmbH, Pößneck
Printed in Germany

ISBN Print 978-3-86881-775-1

ISBN E-Book (PDF) 978-3-86414-96267-178-5

ISBN E-Book (EPUB, Mobi) 978-3-86414-96267-179-2

Weitere Informationen zum Verlag finden Sie unter

www.redline-verlag.de

Beachten Sie auch unsere weiteren Verlage unter www.m-vg.de

*Den Statistikern dieser Welt mit ihren lebenswerten
Eigenschaften der Pedanterie, Großzügigkeit und
Integrität – stets bestrebt, aus den verfügbaren Daten
das denkbar Beste zu machen.*

INHALTSVERZEICHNIS

EINLEITUNG

Warum wir Statistik brauchen.	11
Die Welt in Daten verwandeln	16
Statistik als Unterrichtsfach	21
Dieses Buch.	24

1. KAPITEL

Die Dinge ins richtige Verhältnis setzen – Kategoriale Daten und Prozentzahlen.	27
--	----

2. KAPITEL

Zahlen zusammenfassen und kommunizieren. Große Zahlenmengen.	45
---	----

3. KAPITEL

Warum schauen wir uns die Daten überhaupt an? Populationen und Messverfahren	75
---	----

4. KAPITEL

Was führt zu was?	95
-----------------------------	----

5. KAPITEL

Beziehungen mittels Regression modellieren	117
--	-----

6. KAPITEL

Algorithmen, Analytik und Vorhersage	133
--	-----

7. KAPITEL

Wie sicher kennen wir die Wirklichkeit? Schätzer und Intervalle	171
---	-----

8. KAPITEL

Wahrscheinlichkeit – die Sprache der Unsicherheit und der Streuung. .	185
---	-----

9. KAPITEL

Wahrscheinlichkeit <i>meets</i> Statistik.	205
--	-----

10. KAPITEL	
Fragen, Antworten und Entdeckungen	225
11. KAPITEL	
Auf Bayessche Art aus Erfahrungen lernen	267
12. KAPITEL	
Was schiefgehen kann.	295
13. KAPITEL	
Wie können wir Statistik besser machen	315
14. KAPITEL	
Fazit.	329
DANK	331
ÜBER DEN AUTOR	333
GLOSSAR.	335
ABBILDUNGSVERZEICHNIS.	357
TABELLENVERZEICHNIS.	361
ANMERKUNGEN	363
INDEX	377

EINLEITUNG

Die Zahlen sprechen nicht für sich. Wir sprechen für sie. Wir verleihen ihnen einen Sinn.

– Nate Silver, *Die Berechnung der Zukunft*¹

WARUM WIR STATISTIK BRAUCHEN

Kein verurteilter britischer Mörder hatte so viele Menschen auf dem Gewissen wie Harold Shipman, obwohl er nicht das typische Profil eines Serienkillers aufwies. Der freundliche Hausarzt, der in einem Vorort von Manchester praktizierte, injizierte zwischen 1975 und 1998 mindestens 215 seiner zumeist betagten Patienten starke Überdosen an Opiaten. Zuletzt beging er den Fehler, das Testament einer Patientin dahin gehend zu ändern, dass sie ihm einen Teil ihres Vermögens vermachte. Ihre Tochter, eine Rechtsanwältin, schöpfte Verdacht, und die gerichtlich angeordnete Untersuchung seines Computers ergab, dass er Patientenakten nachträglich gefälscht hatte, um seine Patienten kränker erscheinen zu lassen, als sie in Wirklichkeit gewesen waren. Shipman war bekannt für seine Technikbegeisterung, jedoch reichte sein technischer Sachverstand nicht aus, um sich dessen bewusst zu sein, dass jede Änderung, die er vornahm, mit einem Zeitstempel versehen wurde (nebenbei ein gutes Beispiel für Daten, die einen versteckten Sinn enthalten können).

Fünfzehn seiner Patienten, die nicht feuerbestattet worden waren, wurden exhumiert. In ihren Körpern fand man tödliche Mengen an Diamorphin, der medizinischen Variante von Heroin. Als Shipman daraufhin im Jahr 1999 wegen fünfzehnfachen Mordes angeklagt wurde, entschied er sich dafür, auf eine Verteidigung zu verzichten. Während des gesamten Prozesses äußerte er sich nicht einmal. Er wurde für schuldig befunden und zu lebenslanger Haft verurteilt. Eine

Untersuchung sollte klären, welcher weiteren Verbrechen er sich möglicherweise schuldig gemacht hatte und ob man ihm früher hätte auf die Spur kommen können. Ich gehörte damals zu den Statistikern, die als Gutachter vor den Untersuchungsausschuss geladen wurden, welcher am Ende zu dem Ergebnis kam, dass Shipman mit Sicherheit 215 seiner Patienten und möglicherweise noch weitere 45 ermordet hatte.²

In diesem Buch soll es darum gehen, wie wir mithilfe der **Statistik*** jene Art von Fragen beantworten können, die sich uns stellen, sobald wir versuchen, die Welt besser zu verstehen. Manche dieser Fragen werden wir in grauen Kästen hervorheben. Um Shipmans Verhalten besser zu verstehen, bietet sich als erste Frage beispielsweise diese an:

Was waren das für Menschen, die Shipman tötete,
und wann starben sie?

Die öffentliche Untersuchung lieferte Angaben zum Alter, Geschlecht und Sterbedatum der einzelnen Opfer. Abbildung 0.1 ist eine ziemlich anspruchsvolle Visualisierung dieser Daten in Form einer Punktwolke entlang der Achsen für das Sterbealter und das Sterbedatum der Opfer. Der unterschiedliche Schwärzegrad zeigt an, ob es sich um einen Mann oder eine Frau handelt. Die Achsen wurden um Balkendiagramme ergänzt, die die Häufigkeiten der einzelnen Jahres- und Altersangaben (letztere in 5-Jahres-Blöcken) zusammenfassen.

Für manche Schlussfolgerungen reicht ein kurzer Blick auf die Abbildung. Es gibt mehr schwarze als graue Punkte, was bedeutet, dass die Mehrzahl von Shipmans Opfern Frauen waren. Die Balken rechts der Punktwolke zeigen, dass die meisten Opfer zwischen 65 und 90 Jahre alt waren. Aus der Verteilung der Punkte wird dann aber ersichtlich, dass die Opfer zwar anfangs ausschließlich älter waren, dass sich mit den Jahren aber auch jüngere Fälle einschlichen. Die Balken oberhalb der Punktwolke weisen eine deutliche Lücke rund um das Jahr 1992 auf, für das

* **Fett** gedruckte Begriffe enthält das Glossar am Ende des Buches, das sowohl einige grundlegende Definitionen als auch eine Reihe weiterer Fachtermini erklärt.

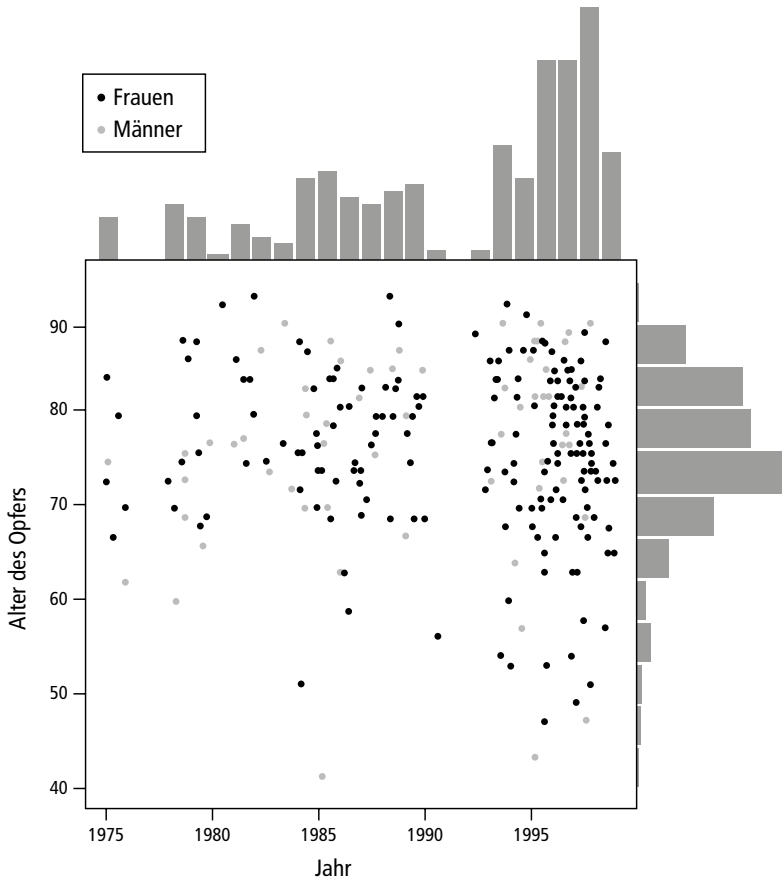


Abbildung 0.1

Eine Punktwolke, die Sterbealter und Todesjahr der 215 bestätigten Opfer Harold Shipmans darlegt. Die Balkendiagramme entlang der Achsen zeigen die entsprechenden Häufigkeitsdichten.

kein Mord belegt ist. Die Erklärung ist, dass Shipman bis dahin in einer Gemeinschaftspraxis mit anderen Ärzten tätig gewesen war, dann aber – möglicherweise, weil er argwöhnte, man könne ihn unter Verdacht haben – eine eigene Praxis für Allgemeinmedizin eröffnete. Anschließend verstärkten sich seine Aktivitäten, wie das obere Balkendiagramm zeigt.

Diese Analyse der von der Untersuchungskommission identifizierten Opfer gibt Anlass zu weiteren Fragen über die Art und Weise, wie er seine Morde durch-

führte. Statistisch aufschlussreich sind beispielsweise die auf den Totenscheinen verzeichneten Tageszeiten, zu denen seine vermutlichen Opfer gestorben sind. Abbildung 0.2 ist ein Kurvendiagramm, das die Verteilung der Tageszeiten vergleicht, zu denen Shipmans Patienten starben und zu denen eine Stichprobe von Patienten anderer Hausärzte aus derselben Gegend starben. Hier lässt sich auch ohne subtile Analyse ein Muster erkennen – eine so gewonnene Erkenntnis wird gelegentlich auch als »interokular« bezeichnet, weil sie den Betrachter »zwischen die Augen« trifft. Shipmans Patienten starben in ihrer überwiegenden Mehrheit am frühen Nachmittag.

Die Daten können uns nicht sagen, *warum* so viele der Patienten zu dieser Zeit starben, aber eine weitere Untersuchung ergab, dass Shipman seine Hausbesuche nach dem Mittagessen vornahm, wenn er mit seinen älteren Patienten zumeist allein war. Er bot ihnen dann eine Spritze an, die ihnen, wie er ihnen versicherte, Erleichterung verschaffen würde, die aber in Wahrheit eine tödliche Menge Diamorphin enthielt. Jedes Mal, wenn ein Patient friedlich in seiner Gegenwart gestorben war, änderte er anschließend seine Patientenakte, um es so aussehen zu lassen, als handelte es sich um einen natürlichen Tod, der zu erwarten gewesen war. Die Leiterin der öffentlichen Untersuchung, Dame Janet Smith, sagte später: »Es ist einfach nur entsetzlich und übersteigt jede Vorstellungskraft, wie er da Tag für Tag den wunderbar fürsorglichen Arzt mimte und doch in der Tasche seine tödliche Waffe mit sich führte ..., um sie schließlich herauszuholen, als wäre nichts dabei.«

Shipman ging dabei ein gewisses Risiko ein, denn eine einzige Obduktion hätte genügt, um ihn zu entlarven. Angesichts des Alters seiner Patienten und der scheinbar natürlichen Todesursachen wurde eine solche jedoch niemals durchgeführt. Seine Beweggründe für diese Morde konnten niemals geklärt werden. Weder sagte er während des Prozesses gegen ihn aus, noch sprach er jemals mit einem Familienangehörigen oder jemand anderem über seine Untaten. Im Gefängnis nahm er sich schließlich das Leben – passenderweise zur richtigen Zeit, damit seine Frau seine Pension beziehen konnte.

Wir können uns diese Form des Forschens als »forensische« Statistik vorstellen, und in diesem Fall handelten wir Statistiker ja tatsächlich im gerichtlichen Auftrag. Wir haben es hier mit keiner Mathematik und keiner Theorie zu tun, sondern ausschließlich mit der Suche nach Mustern, die möglicherweise noch mehr interes-

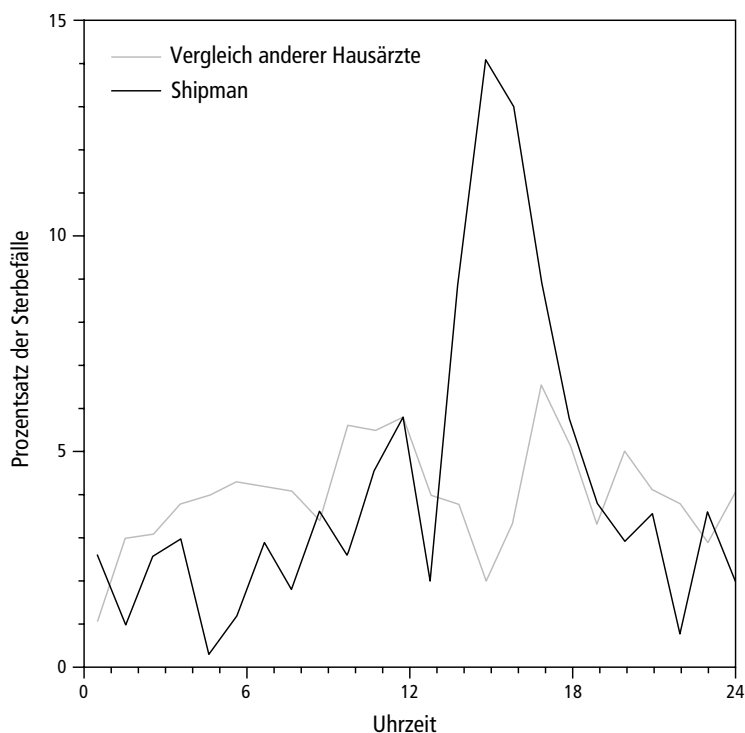


Abbildung 0.2

Die Tageszeiten, zu denen Harold Shipmans Patienten starben, verglichen mit den Tageszeiten, zu denen die Patienten anderer Hausärzte aus derselben Gegend starben. Es bedarf keiner großartigen statistischen Analyse, um hier ein Muster zu erkennen.

sante Fragen aufwerfen. Während die Details von Shipmans Untaten anhand der jeweiligen fallspezifischen Indizien ermittelt wurden, lieferte diese Form der Datenanalyse zugleich generelle Einblicke in die Art und Weise, wie er seine Verbrechen beging.

Später im Buch – im 10. Kapitel – werden wir sehen, ob eine statistische Analyse nach allen Regeln der Kunst hätte helfen können, Shipman früher auf die Schliche zu kommen.* Aber auch so zeigt die Geschichte des mörderischen Haus-

* Für die Ungeduldigen: fast sicher ja!

arztes sehr schön, wie Daten helfen können, die Welt besser zu verstehen und bessere Urteile zu fällen. Und genau davon handelt die Wissenschaft der Statistik.

DIE WELT IN DATEN VERWANDELN

Um Harold Shipmans Verbrechen statistisch erfassen zu können, mussten wir uns von der langen Liste individueller Tragödien lösen, für die er verantwortlich war. Wir mussten aus den unverwechselbaren Facetten des Lebens und Sterbens dieser Menschen bestimmte Fakten und Zahlen extrahieren, die wir anschließend zählen und in Graphen visualisieren konnten. Das mag zunächst kalt und unmenschlich erscheinen, aber wenn wir mit Mitteln der Statistik Licht in die Welt bringen wollen, müssen wir das täglich Erlebte in Daten verwandeln, und das ist nur möglich, indem wir Geschehnisse kategorisieren und etikettieren, Messungen dokumentieren, Ergebnisse analysieren und Schlussfolgerungen kommunizieren.

Allein schon der erste Schritt des Kategorisierens und Etikettierens erweist sich mitunter als äußerst schwierig. Betrachten wir die folgende elementare Frage, die jeden interessieren sollte, dem unsere Umwelt am Herzen liegt:



Wie viele Bäume gibt es auf unserem Planeten?



Bevor wir uns darüber Gedanken machen, wie wir eine Antwort auf diese Frage finden können, müssen wir allerdings eine noch grundlegendere Frage beantworten: Was ist ein »Baum«? Vielleicht entgegnen Sie nun, ein Baum lasse sich doch schon vom bloßen Anschauen erkennen. Aber was Sie für einen Baum halten, mag in den Augen anderer ein Busch oder eine Staude sein. Um also aus Erlebtem und Erfahrenem Daten zu gewinnen, müssen wir zuerst einmal mit stringenten Definitionen arbeiten.

Wie eine kurze Recherche ergibt, gilt eine Pflanze offiziell, das heißt in behördlicher Definition als »Baum«, wenn sie einen hölzernen Stamm hat, der auf Brusthöhe (etwa 1,30 m) einen hinreichend großen Durchmesser aufweist, den sogenannten Bruthöhendurchmesser oder kurz BHD. Die meisten Behörden verwenden einen BHD von 10 Zentimetern, der U.S. Forest Service allerdings verlangt von einer Pflanze einen BHD von mindestens 12,7 Zentimetern, bevor er sie zum Baum erklärt.

Jetzt können wir aber nicht über den ganzen Planeten wandern und sämtliche Pflanzen mit hölzernem Stamm einzeln vermessen, um auf diese Weise die Anzahl derer zu ermitteln, die diesem Kriterium genügen. Die Wissenschaftler, die sich mit dieser Frage beschäftigten, wählten folglich einen pragmatischeren Ansatz. Sie betrachteten zunächst eine Reihe von Gebieten mit jeweils einem einheitlichen Landschaftstyp, Ökozonen genannt, zählten die Bäume, die sie hier fanden, und teilten sie durch die Zahl der Quadratkilometer, die jedes dieser Gebiete umfasste, um die Bäume pro Quadratkilometer zu ermitteln. Anschließend schätzten sie anhand von Satellitenbildern für jede dieser Ökozonen ihre Gesamtfläche auf dem Planeten. Mithilfe einiger komplexer statistischer Modelle kamen sie schließlich auf eine geschätzte Gesamtzahl von 3,04 Billionen (oder 3.040.000.000.000) Bäumen auf der Erde. Das klingt nach viel; jedoch vermuteten die Wissenschaftler, dass es einmal doppelt so viele gewesen waren.*³

Wenn sich die amtlichen Stellen schon nicht einig sind, was unter einem Baum zu verstehen ist, sollte es uns nicht überraschen, wenn weniger klare Begriffe noch schwierigere Fragen aufwerfen. Um ein extremes Beispiel zu nennen: Die offizielle Definition von »Arbeitslosigkeit« veränderte sich in Großbritannien in den Jahren 1979 bis 1996 mindestens 31-mal.⁴ Die Definition des Bruttoinlandsprodukts (BIP) unterliegt ständigen Revisionen. So werden beispielsweise in Großbritannien seit 2014 auch der Drogenhandel und die Prostitution im BIP erfasst; die Schätzverfahren greifen dabei auf einige ungewöhnliche Datenquellen wie bei-

* Diese Zahl wird mit einer Fehlermarge von 0,1 Billionen angegeben, was bedeutet, dass sich die Wissenschaftler hinreichend sicher waren, dass der wahre Wert im Bereich von 2,94 und 3,14 Billionen liegt. (Ich gestehe, dass mir dies angesichts der vielen im Modell gemachten Annahmen als etwas zu exakt erscheint.) Sie schätzten zudem, dass jedes Jahr 15 Milliarden Bäume gefällt werden und der Planet seit Bestehen der menschlichen Zivilisation 46 Prozent seines Baumbestandes verloren hat.

spielsweise Punternet – eine Website zur Bewertung von Prostitutionsdienstleistungen, die auch Preise für die unterschiedlichen Aktivitäten angibt – zurück.⁵

Selbst unsere persönlichsten Gefühle lassen sich kodifizieren und einer statistischen Analyse unterziehen. Von Oktober 2016 bis September 2017 wurden 150 000 britische Bürger im Rahmen einer Erhebung gefragt: »Wie glücklich und zufrieden fühlten Sie sich gestern auf Ganze gesehen?«⁶ Auf einer Skala von 0 bis 10 lag die Durchschnittsantwort bei 7,5 und damit leicht über dem Wert von 7,3 aus dem Jahr 2012, was möglicherweise auf die wirtschaftliche Erholung seit dem Börsencrash von 2008 zurückzuführen ist. Den niedrigsten Wert lieferten die 50- bis 54-Jährigen, den höchsten Wert die 70- bis 74-Jährigen, was in Großbritannien einem typischen Muster entspricht.*

Glück und Zufriedenheit zu messen, fällt schwer; ob jemand lebt oder tot ist, sollte sich hingegen problemlos ermitteln lassen. Wie die Beispiele in diesem Buch zeigen werden, sind Überlebens- und Sterberaten ein häufiges Thema statistischer Betrachtungen. In den Vereinigten Staaten jedoch kann jeder Bundesstaat seine eigene juristische Definition von Tod haben, und obwohl im Jahr 1981 der Uniform Declaration of Death Act (das »Gesetz zur Vereinheitlichung der Todeserklärung«) erlassen wurde, bleiben kleinere Unterschiede bestehen. Jemand, der in Alabama bereits für tot erklärt wurde, könnte zumindest theoretisch nach der Überquerung der Grenze zu Florida seine juristische Qualifizierung als Toter verlieren, müssen dort doch zwei qualifizierte Ärzte den Eintritt des Todes bestätigen.⁷

Diese Beispiele zeigen, dass Statistiken immer auch Entscheidungen zugrunde liegen, die bis zu einem gewissen Grade willkürlich sind, und dass es eine Illusion ist zu glauben, man könnte die Komplexität der individuellen Erfahrung zweifelsfrei kodieren und in eine Tabelle oder eine andere Form von Software eingeben. Und so schwer es fällt, Eigenschaften von uns Menschen und der Welt um uns herum zu definieren, zu zählen und zu messen, sind das dann immer noch erst »Daten« und damit lediglich der Ausgangspunkt unserer Bemühungen, die Welt besser zu verstehen.

Daten als Quelle des Erkenntnisgewinns unterliegen zwei entscheidenden Einschränkungen. Erstens erfassen sie die Dinge, die uns in Wahrheit interessieren, fast immer nur unvollständig: Das emotionale Wohlbefinden der Nation fangen

* Was mir, sofern ich »Durchschnitt« bin, etwas gibt, auf das ich mich freuen kann.

wir schwerlich damit ein, dass wir die Menschen fragen, wie glücklich und zufrieden sie in der zurückliegenden Woche auf einer Skala von 0 bis 10 gewesen sind. Zweitens wird alles, was wir zu messen beschließen, örtlichen, personellen und zeitlichen Schwankungen unterliegen und das Problem ist dann, aus dieser scheinbar willkürlichen **Streuung** bedeutsame Schlussfolgerungen zu ziehen.

Seit Jahrhunderten stellt sich die Statistik dieser doppelten Herausforderung und spielt eine führende Rolle im Versuch, die Welt mit wissenschaftlichen Mitteln zu begreifen. Sie liefert die Grundlage für eine – notwendigerweise stets unvollständige – Interpretation der Daten mit dem Ziel, aus dem Hintergrundrauschen, das sich daraus erklärt, dass kein Mensch wie der andere ist, wichtige Beziehungen und Gesetzmäßigkeiten herauszudestillieren. Aber die Welt bleibt nicht stehen; es entstehen neue Fragen, neue Datenquellen werden verfügbar und somit muss sich auch die Statistik weiterentwickeln.

Menschen haben schon immer gezählt und gemessen. Als eigene Disziplin jedoch nahm die moderne Statistik ihren Anfang in den 1650er-Jahren, als Blaise Pascal und Pierre de Fermat zum ersten Mal begriffen, was **Wahrscheinlichkeit** wirklich bedeutet, wie wir im 8. Kapitel sehen werden. Auf der Grundlage dieses mathematisch soliden Ansatzes für den Umgang mit der Mannigfaltigkeit und Unbeständigkeit der Welt entwickelte sich die Wahrscheinlichkeitslehre von da an in raschem Tempo weiter. Aus den Daten zum Sterbealter der Menschen ließen sich nun Pensionen und Renten berechnen. Die Astronomie machte einen enormen Satz vorwärts, nachdem Wissenschaftler erkannten, wie sie mithilfe der Wahrscheinlichkeitstheorie dem Problem von Messungenauigkeiten begegnen konnten. Im viktorianischen Zeitalter begann man mit Begeisterung, Daten über den menschlichen Körper (und tausend andere Dinge) zu sammeln und starke Bezüge zwischen statistischer Analyse und Genetik, Biologie und Medizin herzustellen. Im 20. Jahrhundert wurde die Statistik mathematischer. Zum Leidwesen vieler Studenten und praktischer Anwender wurde sie irgendwann gleichbedeutend mit der mechanischen Verwendung einer ganzen Reihe von häufig nach irgendwelchen exzentrischen und streitlustigen Statistikern benannten statistischen Instrumenten, die wir im Lauf dieses Buches kennen lernen werden.

Diese verbreitete Vorstellung von Statistik als einem »Köcher voller Tools« stößt mittlerweile an ihre Grenzen. Erstens befinden wir uns im Zeitalter der *Data Science* oder **Datenwissenschaft**, in der umfangreiche und komplexe Datenmen-

gen aus Routinequellen wie Verkehrsüberwachungskameras, Social-Media-Posts und Internetkäufen zusammengetragen und als Basis für technologische Innovationen wie optimierte Verkehrsführung, zielgerichtete Werbung oder Kaufempfehlungssysteme genutzt werden. **Algorithmen** auf der Basis von **Big Data** werden wir im 6. Kapitel behandeln. Statistik erscheint zunehmend als ein Unterbereich der Datenwissenschaft, benötigt doch der moderne Datenwissenschaftler darüber hinaus Fähigkeiten in Datenmanagement, Programmierung und Algorithmenentwicklung sowie ein gründliches Verständnis des untersuchten Gegenstandsbereichs.

Dem traditionellen Verständnis von Statistik setzt zudem die starke Zunahme wissenschaftlicher – insbesondere biomedizinischer und soziologischer – Forschungstätigkeit zu sowie der Erwartungsdruck, in hochkarätigen Fachzeitschriften zu publizieren. Die Folge sind Zweifel an der Zuverlässigkeit von Teilen der wissenschaftlichen Literatur und Behauptungen, dass sich viele »Entdeckungen« nicht reproduzieren lassen – siehe beispielsweise den anhaltenden Streit darüber, ob eine selbstbewusste Körperhaltung, eine *Power Pose*, hormonelle und andere Veränderungen bewirken kann.⁸ Für die sogenannte Reproduzierbarkeits- oder Replikationskrise wird nicht zuletzt die unsachgerechte Anwendung statistischer Standardmethoden verantwortlich gemacht.

Mit der zunehmenden Verfügbarkeit gewaltiger Datenmengen und anwenderfreundlicher Analysesoftware könnte man meinen, es bestünde weniger Trainingsbedarf in statistischen Methoden. Das wäre jedoch eine extrem naive Annahme. Größere Datenmengen und die Zunahme der Zahl und Komplexität wissenschaftlicher Studien erübrigen nicht statistische Fähigkeiten, sondern ganz im Gegenteil erschweren sie es, die richtigen Schlussfolgerungen zu ziehen. Mehr Daten bedeuten, dass wir noch sorgfältiger schauen müssen, was ihre Aussagekraft wirklich wert ist.

Eine intensive Analyse von Datensätzen, die aus Routinedaten gewonnen wurden, kann beispielsweise die Möglichkeit falscher »Entdeckungen« erhöhen, weil entweder bereits die Ausgangsdaten eine systematische Verzerrung aufweisen oder weil von einer Vielzahl von Analysen nur diejenigen Ergebnisse berichtet werden, die »interessant« erscheinen – eine Vorgehensweise, die auch als *p*-Hacking bezeichnet wird. Um veröffentlichte wissenschaftliche Arbeiten und erst recht die Medienberichte, denen wir täglich begegnen, richtig einordnen zu können, schadet es nicht, sich der Gefahren des selektiven Berichtens, der Notwendigkeit einer Repli-

zierbarkeit wissenschaftlicher Behauptungen durch unabhängige Kollegen und der Gefahr einer Überinterpretation einzelner aus dem Kontext gerissener Studien bewusst zu sein.

Alle diese Einsichten lassen sich unter dem Stichwort der **Datenkompetenz** zusammenfassen: der Fähigkeit, nicht nur Probleme des echten Lebens einer statistischen Analyse zu unterziehen, sondern auch die Schlussfolgerungen, die andere aus statistischen Überlegungen ziehen, zu verstehen und richtig einzuordnen. Die Datenkompetenz der breiten Bevölkerung können wir jedoch nur verbessern, indem wir Statistik anders unterrichten.

STATISTIK ALS UNTERRICHTSFACH

Generationen von Schülern litten und leiden unter einer trockenen Vermittlung von Statistik im Mathematikunterricht, die sich auf die Vermittlung einer Reihe von Techniken zur Anwendung in bestimmten Situationen beschränkt. Vor lauter mathematischer Theorie kommen praktische Fragen, warum beispielsweise eine Formel verwendet wird oder welche Probleme auftauchen können, wenn wir Fragen mittels Daten zu beantworten versuchen, häufig zu kurz.

Glücklicherweise beginnt sich das zu ändern. Die Anforderungen der Datenwissenschaft und der Datenkompetenz rufen nach einem stärker problemorientierten Ansatz, bei dem die Anwendung bestimmter statistischer Instrumente lediglich als eine Komponente unter vielen in einem umfassenderen Untersuchungsprozess gesehen wird. Ein möglicher Problemlösungsansatz ist die **PPDAC**-Struktur, an der wir uns in diesem Buch orientieren werden.⁹ Abbildung 0.3 basiert auf einem Beispiel aus Neuseeland, einem Vorreiter in Sachen Statistikunterricht in den Schulen.

Die erste Station des Zyklus ist die Spezifizierung des Problems. Die statistische Untersuchung beginnt stets mit einer Fragestellung – siehe unsere Fragen nach einem Muster in den Morden Harold Shipmans und nach der Anzahl der Bäume

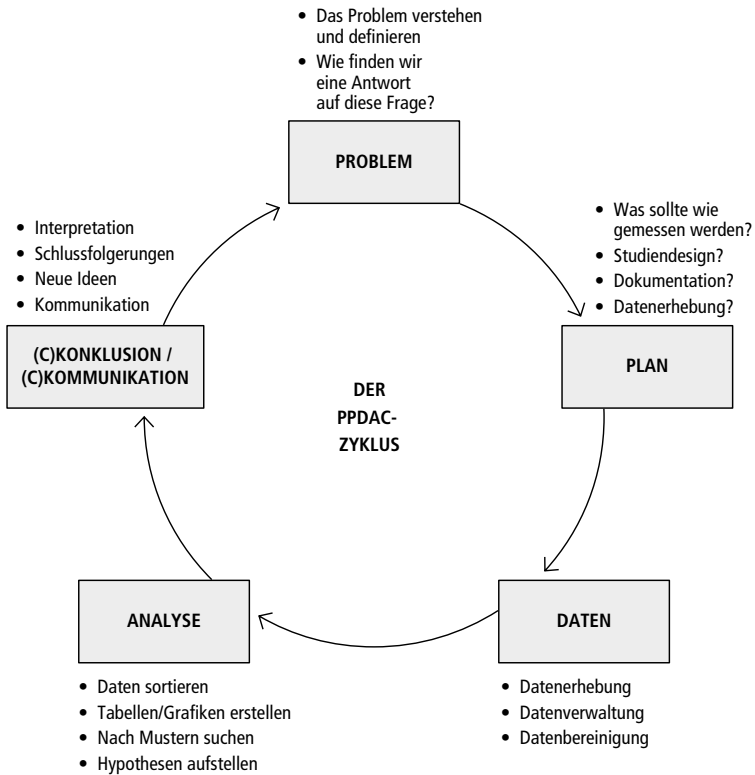


Abbildung 0.3

Der PPDAC-Problemlösungszyklus mit den Stationen Problem, Plan, Daten, Analyse und (C)Konklusion (*conclusion*; Schlussfolgerung) bzw. (C)Kommunikation (*communication*). Nach jedem Durchlauf beginnt ein neuer Zyklus.

auf der Erde. Später im Buch wollen wir uns mit Fragen beschäftigen wie der nach dem erwartbaren Nutzen verschiedener Therapien unmittelbar im Anschluss an eine Brustkrebsoperation oder der nach dem Grund, warum ältere Männer große Ohren haben.

Mitunter könnte man versucht sein, auf eine sorgfältige Planung zu verzichten. Die Shipman-Frage verlangt von uns lediglich, so viele Daten wie möglich über seine Opfer zusammenzutragen. Aber die Leute, die die Bäume zählten, legten großen Wert auf präzise Definitionen und klare Messverfahren, weil nur eine gut

gestaltete Studie verlässliche Schlussfolgerungen zulässt. In der Eile, die nötigen Daten zusammenzutragen und mit der Analyse beginnen zu können, kommt jedoch häufig die sorgfältige Studienplanung – das »Studiendesign« – zu kurz.

Den zum Erfassen »guter« Daten erforderlichen organisatorischen und programmiertechnischen Fähigkeiten kommt in der Datenwissenschaft eine zunehmend wichtige Rolle zu – unter anderen deshalb, weil aus Routinequellen bezogene Daten häufig erst noch aufwendig »bereinigt« werden müssen, bevor sie sich analysieren lassen. Datenerhebungssysteme verändern sich im Lauf der Zeit, es schleichen sich offensichtliche Fehler ein und so weiter: Die Formulierung »gefundene Daten« bringt hübsch zum Ausdruck, in welchem ungeordneten Zustand sie sich mitunter befinden – wie etwas, das man von der Straße aufliest.

Die eigentliche »Analyse« stellt traditionell den Schwerpunkt des Statistikunterrichts dar, und auch wir werden in diesem Buch eine Reihe von Analysetechniken vorstellen. Gelegentlich jedoch reicht bereits eine geeignete Visualisierung wie in Abbildung 0.1. Und zu guter Letzt ist der Schlüssel zu guter Statistik, die richtigen Schlüsse zu ziehen, die der begrenzten Aussagekraft der Datengrundlage Rechnung tragen, und sie klar zu kommunizieren, wie wir dies in den grafischen Veranschaulichungen der Shipman-Daten gesehen haben. Jede Schlussfolgerung wirft wieder neue Fragen auf, und so beginnt der Zyklus von Neuem – wie in dem Moment, als wir uns für die Uhrzeiten zu interessieren begannen, zu denen Shipmans Patienten gestorben sind.

Auch wenn der PPDAC-Zyklus in der Praxis nicht immer genau so befolgt wird, wie wir ihn in Abbildung 0.3 dargestellt finden, unterstreicht er dennoch, dass formale statistische Analysetechniken lediglich einen Teil der Arbeit eines Statistikers oder Datenwissenschaftlers ausmachen. Statistik ist weit mehr als ein Zweig der Mathematik voller esoterischer Formeln, mit denen sich Generation um Generation von Schülern und Studenten herumquält.

DIESES BUCH

Als ich in den 1970er-Jahren in Großbritannien studierte, gab es genau drei Fernsehsender, Computer hatten die Größe von zwei Kleiderschränken und das Wikipedia-Ähnlichste, was wir hatten, war das Handgerät aus Douglas Adams' (bemerkenswert vorausschauendem) Roman *Per Anhalter durch die Galaxis*. Zum Zwecke der Selbstoptimierung griffen wir deshalb zu Pelican-Büchern, deren kultige blaue Rücken jedes Studentenbücherregal schmückten.

Weil ich Statistik studierte, befanden sich in meiner Pelican-Sammlung *Facts from Figures* (dt.: *Einführung in die Statistik*) von M. J. Moroney (1951) und *How to Lie with Statistics* (dt.: *Wie lügt man mit Statistik*) von Darrell Huff (1954). Diese ehrwürdigen Ausgaben verkauften sich zu Hunderttausenden, was sowohl für das vorhandene Interesse an der Statistik als auch für den Mangel an Alternativen sprach. Während sich diese Klassiker während der fast sieben Jahrzehnte seit ihrem Erscheinen erstaunlich gut behaupteten, erfordert das heutige Zeitalter einen anderen, an den oben beschriebenen Prinzipien ausgerichteten Unterrichtsansatz.

Dieses Buch nimmt deshalb das Lösen von lebenspraktischen Problemen als Ausgangspunkt für die Vorstellung statistischer Konzepte. Einige dieser Konzepte mögen intuitiv einleuchten, während andere subtilerer Art sind und etwas mehr geistige Anstrengung erfordern, wenngleich mathematische Fähigkeiten im engeren Sinne nicht vonnöten sein werden. Verglichen mit klassischen Texten richtet dieses Buch den Fokus mehr auf konzeptionelle Fragestellungen als auf technische Fertigkeiten und verwendet lediglich ein paar unschuldige Gleichungen unterstützt von einem Glossar. Software spielt in jeder Art von Datenwissenschaft und Statistik eine unverzichtbare Rolle, ist aber kein Schwerpunkt dieses Buches. Anleitungen für frei verfügbare Software wie R oder Python lassen sich unschwer finden.

Die in Kästen hervorgehobenen Fragen lassen sich allesamt mehr oder weniger gut mittels statistischer Analyse beantworten, auch wenn sich ihre Bedeutungsrahmen stark unterscheiden. Bei manchen handelt es sich um wichtige wissenschaftliche Hypothesen, beispielsweise, ob das Higgs-Boson existiert oder ob es in

der Tat hinreichend schlüssige Beweise für außersinnliche Wahrnehmungen (ASW) gibt. Bei anderen geht es um Gesundheitsthemen: ob stark frequentierte Krankenhäuser bessere Überlebenschancen bieten oder ob ein Screening auf Eierstockkrebs zu weniger Erkrankungen führt. Manchmal sind wir lediglich an quantitativen Schätzungen interessiert: der Erhöhung des Krebsrisikos durch den Verzehr von Speck-Sandwiches, der Zahl der Sexualpartner der Menschen in Großbritannien während ihres gesamten Lebens oder dem Nutzen der täglichen Einnahme eines Statins.

Und manche Fragen sind einfach nur spannend: die Identifizierung des *Titanic*-Überlebenden, dem das Glück am holdesten war, die Frage, ob man Harold Shipman früher hätte auf die Schliche kommen können, oder die Bestimmung der Wahrscheinlichkeit, dass ein bei einer Ausgrabung auf einem Parkplatz in Leicester gefundenes Skelett wirklich als die Gebeine von König Richard III. identifiziert werden kann.

Dieses Buch richtet sich sowohl an Schüler und Studenten der Statistik, die sich für eine nichttechnische Einführung in die Grundthemen interessieren, als auch allgemein an Leser, die mehr Einblick in die Hintergründe der statistischen Zahlen und Fakten erhalten wollen, denen sie in Beruf und Alltag begegnen. Mein Hauptaugenmerk liegt auf dem gekonnten und sorgfältigen Umgang mit Statistiken: Zahlen mögen als kalte und harte Fakten daherkommen, aber die Versuche, Bäume, Glück und Zufriedenheit oder Todesfälle zu messen, haben bereits gezeigt, dass der Umgang mit ihnen viel Fingerspitzengefühl erfordert.

Statistiken können Erkenntnisgewinn und eine klare Sicht auf Problemstellungen ermöglichen, aber wir alle wissen auch (oder ahnen es zumindest), dass sie sich missbrauchen lassen, um beispielsweise bestimmte Sichtweisen gezielt zu pushen oder einfach nur Aufmerksamkeit zu heischen. Die Fähigkeit, statistische Behauptungen auf ihre Stichhaltigkeit hin zu überprüfen, ist, so scheint mir, in der heutigen Welt eine Schlüsselkompetenz. So hoffe ich, dass dieses Buch seinen Beitrag dazu leisten wird, dass die Menschen den Mut und die Kraft finden, die Zahlen, die täglich auf sie einprasseln, mit der nötigen Skepsis zu behandeln und gebührend zu hinterfragen.

ZUSAMMENFASSUNG

- Erfahrenes und Erlebtes lässt sich nicht in eindeutiger Weise in Daten übersetzen und jedes Datenabbild der Wirklichkeit muss immer unvollständig bleiben.
- Die Statistik als Wissenschaft blickt auf eine lange und erfolgreiche Geschichte zurück, ändert jedoch mit der Zunahme der verfügbaren Daten ihren Charakter.
- Jeder Datenwissenschaftler benötigt umfangreiche Kenntnisse in statistischen Methoden.
- Im Zentrum der statistischen Ausbildung stehen heute nicht mehr nur mathematische Methoden, sondern der vollständige Problemlösungsprozess.
- Der PPDAC-Zyklus liefert dazu das passende Rahmenkonzept: Seine Stationen lauten Problem – Plan – Daten – Analyse und (C)Konklusion (conclusion) bzw. (C)Kommunikation (communication).
- Datenkompetenz ist in der Welt von heute unverzichtbar.

1. KAPITEL

DIE DINGE INS RICHTIGE VERHÄLTNIS SETZEN – KATEGORIALE DATEN UND PROZENTZAHLEN

Was geschah mit den Kindern, die zwischen 1984 und 1995
in Bristol am Herzen operiert wurden?

Joshua L. war 16 Monate alt und litt unter einer Transposition der großen Arterien: einem schweren angeborenen Herzfehler, bei dem Hauptschlagader und Lungenschlagader mit jeweils der falschen Herzkammer verbunden sind. Joshua benötigte eine Operation zur »Vertauschung« der Arterien. Am 12. Januar 1995 verabschiedeten sich seine Eltern um kurz nach 7 Uhr von ihm und schauten zu, wie er in den Operationssaal der Bristol Royal Infirmary geschoben wurde. Was Joshuas Eltern nicht wussten, war, dass seit den frühen 1990er-Jahren Geschichten über die geringen Überlebenschancen bei solchen Operationen speziell in Bristol in Umlauf waren. Niemand hatte ihnen erzählt, dass Krankenschwestern und -pfleger sogar gekündigt hatten, nur um nicht länger Eltern berichten zu müssen, dass ihr Kind gestorben war. Sie wussten auch nicht von der spätabendlichen Teambesprechung am Vortag, auf der erwogen worden war, Joshuas Operation abzusagen.¹

Joshua verstarb auf dem Operationstisch. Im Jahr darauf startete das General Medical Council (die britische Ärztekammer) aufgrund der Beschwerden von Joshuas und anderen Eltern, die ihre Kinder verloren hatten, eine Untersuchung. Im Jahr 1998 wurden zwei Chirurgen und der frühere Krankenhausleiter des schwe-

ren ärztlichen Fehlverhaltens für schuldig befunden. Die Öffentlichkeit blieb jedoch alarmiert, sodass schließlich eine amtliche Untersuchung eingeleitet wurde, in deren Rahmen ein Team von Statistikern die traurige Aufgabe erhielt, die Überlebensraten in Bristol zwischen 1984 und 1995 mit denen anderer Orte in Großbritannien zu vergleichen. Ich leitete dieses Team.

Wir ermittelten als Erstes, wie viele Kinder am Herzen operiert worden waren und wie viele dabei ihr Leben gelassen hatten. Das klingt nicht besonders kompliziert, aber wie die Einleitung gezeigt hat, kann bereits das simple Zählen von Ereignissen seine Tücken haben. Wer ist »Kind«? Was zählt als »Operation am Herzen«? Wann lässt sich ein Todesfall eindeutig als Folge einer Operation bestimmen? Und selbst nachdem wir alle diese Dinge definiert hatten, blieb die Frage: Lassen sich die Zahlen so ohne Weiteres ermitteln?

Wir definierten »Kind« als »noch nicht 16-jährig« und beschränkten uns auf Operationen am offenen Herzen, bei denen das Herz gestoppt und seine Funktion von einem kardiopulmonalen Bypass übernommen worden war. Einige der jungen Patienten hatten während eines Krankenhausaufenthalts mehrere Operationen gehabt, die wir jedoch wie eine einzige behandelten. Todesfälle rechneten wir den Operationen zu, wenn sie binnen 30 Tagen nach der letzten Operation eintraten, unabhängig davon, wo der Tod eintrat und ob ein Zusammenhang mit der Operation offensichtlich war oder nicht. Wir wussten, dass der Tod ein unvollkommenes Maß für die Qualität des Operationsresultats war, weil damit nicht erfasst wurde, ob die Kinder infolge des Eingriffs Hirn- oder andere Schäden davongetragen hatten, aber zu den langfristigen Auswirkungen fehlten uns die Daten.

Unsere Hauptdatenquelle war die Datenbank »Hospital Episode Statistics« (HES, die britische Krankenhausstatistik), deren Einträge sich wiederum aus Verwaltungsdaten speisten, die von Hilfskräften eingegeben wurden. Während die HES bei Ärzten in schlechtem Ruf stand, hatte sie als Quelle den Vorteil, dass sie sich mit dem nationalen Sterberegister verknüpfen ließ. Daneben gab es ein paralleles System von Daten, welche unmittelbar an die vom chirurgischen Berufsverband eingerichtete Cardiac Surgical Registry (CSR) übermittelt wurden.

Obgleich diese beiden Datenquellen theoretisch dieselben Vorgänge dokumentierten, wiesen sie beträchtliche Abweichungen auf: Für die Jahre 1991–1995 sprach die HES von 62 Todesfällen bei 505 offenen Operationen (14 Prozent), während in der CSR 71 Todesfälle bei 563 Operationen (13 Prozent) verzeichnet

waren. Nicht weniger als fünf weitere lokale Datenquellen waren verfügbar, von der Dokumentation der Anästhesisten bis zu den persönlichen Logbüchern der Chirurgen. In Bristol mangelte es nicht an Daten, aber keine dieser Datenquellen konnte für sich beanspruchen, die »Wahrheit« gepachtet zu haben, und niemand hatte sich die Mühe gemacht, die Operationsergebnisse zu analysieren und daraus irgendwelche Konsequenzen zu ziehen.

Wir errechneten, dass, hätten die Patienten in Bristol eine Überlebenswahrscheinlichkeit gleich der durchschnittlichen Überlebenswahrscheinlichkeit aller übrigen Orte in Großbritannien gehabt, für den betrachteten Zeitraum 32 statt der in der HES verzeichneten 62 Todesfälle zu erwarten gewesen wären, weshalb wir damals in unserem Bericht von »30 zusätzlichen Todesfällen zwischen 1991 und 1995« sprachen.* Die exakte Zahl variierte je nach zugrunde gelegten Datenquellen und es mag schon für sich genommen ungeheuerlich erscheinen, dass es uns nicht möglich war, so elementare Fakten wie die Zahl der Operationen und ihre Ausgänge eindeutig zu ermitteln, was beim Stand der modernen Dokumentationstechnik eigentlich hätte möglich sein müssen.

Diese Befunde fanden ein breites Presseecho. *Eine* Konsequenz besagter Untersuchung waren weitreichende Veränderungen in der Art und Weise, wie die klinische Leistung dokumentiert wurde: Es war der ärztlichen Zunft nicht länger überlassen, sich selbst zu reglementieren. Mechanismen zur amtlichen Dokumentation klinischer Überlebensdaten wurden etabliert – wenngleich, wie wir als Nächstes sehen werden, allein schon die Art und Weise, wie Daten visualisiert und präsentiert werden, Einfluss darauf hat, wie sie vom Publikum aufgenommen werden.

Zahlen und Zahlenverhältnisse kommunizieren

Daten, die nichts weiter festhalten, als dass bestimmte Ereignisse stattgefunden oder nicht stattgefunden haben, werden auch als **Binärdaten** bezeichnet, weil sie nur zwei Werte annehmen können; wir sprechen hier üblicherweise von »wahr«

* Heute bedauere ich die Formulierung »zusätzliche Todesfälle« (*excess deaths*), die die Zeitungen später als »vermeidbare Todesfälle« interpretierten. Allein dem Zufall ist es zuzuschreiben, dass rund die Hälfte der Krankenhäuser Todesfallzahlen oberhalb des Durchschnitts aufweisen, und nur die wenigsten davon dürften vermeidbar gewesen sein.