

1 Mit Daten Geschichten erzählen

Denken Sie einen Moment an die weitverbreiteten Datenvisualisierungen – diejenigen, über die Sie in Veranstaltungen oder Vorlesungen hören, über die Sie in Blogs lesen, und an diejenigen, die Ihnen spontan eingefallen sind, als Sie diesen Satz gelesen haben. Was haben all diese Visualisierungen gemeinsam? Alle erzählen eine interessante Geschichte. Vielleicht bestand ihr Zweck darin, Sie von etwas zu überzeugen. Oder es ging darum, Sie zum Handeln zu bewegen, Sie mit neuen Informationen zu versorgen oder Ihre vorgefasste Einstellung zu einem bestimmten Aspekt unserer Wirklichkeit infrage zu stellen. Egal, was es ist, gute Datenvisualisierungen – ob groß oder klein, ob im Rahmen eines Kunstwerks oder für eine Präsentationsfolie – helfen Ihnen, die Geschichte hinter den Daten zu erkennen.

Mehr als nur Zahlen

Machen wir uns nichts vor. Daten können einfach nur langweilig sein, wenn Sie nicht wissen, wonach Sie suchen, oder wenn Sie von vornherein nicht wissen, dass die Daten interessante Informationen enthalten. Daten sind dann nichts weiter als eine Mischung aus Zahlen und Wörtern, die außer den reinen Daten keine Bedeutung zu haben scheinen. Das Tolle an Statistik und Datenvisualisierung ist, dass sie Ihnen helfen können, hinter die Daten zu blicken. Denken Sie daran: Daten repräsentieren immer das echte Leben. Es handelt sich bei Daten nicht nur um reine Zahlen. Daten haben eine Bedeutung, zeigen die Realität und können sogar einen Funken Schönheit beinhalten. Wie auch im realen Leben kann die Geschichte manchmal sehr einfach und geradlinig sein, in anderen Fällen ist sie komplex und verzweigt. Manche Geschichten passen besser in ein Sachbuch, andere eher in einen Roman. Es liegt an Ihnen, dem Statistiker, dem Programmierer, Designer oder Datenwissenschaftler, welche Geschichte Sie erzählen.

Dies war eines der ersten Dinge, die ich während des Statistikstudiums gelernt habe. Ich muss zugeben, dass Statistik für mich vor meinem Studienbeginn die reine Analyse von Daten war. Daten waren für mich das Ergebnis eines rein mechanischen Prozesses. In den meisten Fällen trifft dies auch zu. Im Hauptstudium habe ich Elektrotechnik studiert; daher ist es nicht verwunderlich, dass ich Daten aus dieser Perspektive betrachtete.

Verstehen Sie mich nicht falsch. Diese Perspektive ist an sich nicht verkehrt, jedoch habe ich im Laufe der Jahre gelernt, dass Daten, auch wenn sie objektiv sind, immer auch eine menschliche Dimension besitzen.

Nehmen Sie noch mal das Beispiel zu den Arbeitslosenzahlen aus der Einleitung dieses Buches. Es ist einfach, die durchschnittliche Arbeitslosenquote eines bestimmten Bundesstaates zu ermitteln. Sie haben jedoch auch gesehen, dass es innerhalb eines Staates große Abweichungen und Unterschiede geben kann. Ja, sogar zwischen benachbarten Gemeinden oder Städten lassen sich Unterschiede erkennen. Möglicherweise kennen Sie jemanden, der in den letzten Jahren seinen Job verloren hat. Der- oder diejenige ist dann nicht nur einfach eine weitere Nummer in der Statistik, nicht wahr? Die Zahlen repräsentieren Menschen, und daher sollten Sie mit den Daten auch auf entsprechende Weise umgehen. Sie brauchen nicht die Geschichte jeder einzelnen Person zu erzählen. Dennoch besteht ein gravierender Unterschied zwischen der Aussage, dass die Arbeitslosenquote um fünf Prozentpunkte gestiegen ist, und der Tatsache, dass hierdurch

Hunderttausende ihren Arbeitsplatz verloren haben. Die erste Teilaussage liest sich wie eine Zahl ohne Inhalt, wohingegen man sich mit der zweiten Teilaussage eher identifizieren kann – sie ist besser zuordenbar.

Journalismus

Ein Praktikum in der Grafikabteilung der *New York Times* hat mir einige Dinge klar werden lassen. Das Praktikum dauerte nur drei Monate in dem Sommer nach dem zweiten Jahr auf der Graduiertenschule, jedoch hat es bleibenden Eindruck auf die Art und Weise hinterlassen, wie ich mit Daten umgehe. Ich habe nicht nur gelernt, wie ich Infografiken für eine Meldung erstelle. Ich habe gelernt, wie ich aus Daten Nachrichten machen kann. Damit einhergehend habe ich dann viel über Design, Organisation, Faktenprüfung, Detektivarbeit und über Recherche gelernt.

Es gab Tage, an denen meine Aufgabe lediglich darin bestand, drei Zahlen in einem Datensatz zu verifizieren, da die Grafikabteilung der *New York Times* vor dem Erstellen einer Grafik immer sicherstellt, dass die Daten, auf denen sie basiert, auch korrekt sind. Erst wenn wir wussten, dass die Daten verlässlich sind, haben wir uns um die Präsentation der Daten gekümmert. Es ist eben diese Detailgenauigkeit, die die Grafiken in *der New York Times* so gut macht.

Schauen Sie sich eine x-beliebige Grafik der *New York Times* an. Die Daten werden klar, prägnant und sehr angenehm präsentiert. Was bedeutet das genau? Wenn Sie sich eine Infografik ansehen, können Sie die dargestellten Daten verstehen. Wichtige Punkte oder Bereiche werden hervorgehoben; Symbole und Farben werden sorgfältig in einer Legende oder in anderen Beschriftungen erklärt. Außerdem macht es die *Times* ihren Lesern einfach, die Geschichte in den Daten zu erkennen. Es ist nicht nur einfach ein Graph, ein Diagramm, sondern eine Grafik.

Die Grafik in Abbildung 1.1 ist ein typisches Beispiel für das, was Sie auch in der *New York Times* finden. Sie illustriert die zunehmende Wahrscheinlichkeit, dass Sie, abhängig vom Alter, innerhalb eines Jahres sterben werden.

Diese Grafik basiert auf einem einfachen Liniendiagramm. Jedoch helfen zahlreiche Designelemente, die Geschichte besser zu erzählen. Die Be-

www



Einige der besten Infografiken aus *The New York Times* finden Sie unter <http://datafl.ws/nytimes0>.

Hinweis



Schauen Sie sich die englischsprachige Videodokumentation *Journalism in the Age of Data* von Geoff McGhee an, in der beschrieben wird, wie Journalisten Daten verwenden, um über aktuelle Ereignisse zu berichten. In der Dokumentation finden Sie auch Interviews mit einigen der Besten in der Branche.

schriftungen und Hinweise liefern den Kontext und führen Ihnen vor Augen, warum die Daten interessant sind. Die Linienstärken und Farben lenken das Auge auf die wichtigen Stellen.

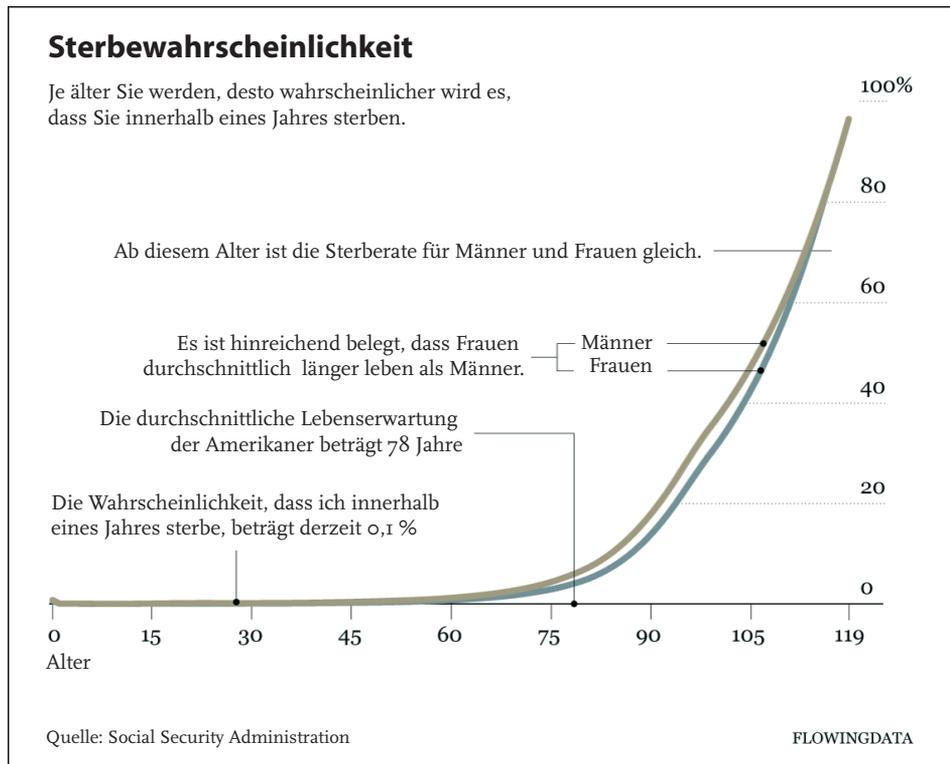


Abbildung 1.1: Die Sterbewahrscheinlichkeit hängt vom Alter ab

Beim Erstellen von Diagrammen und Grafiken geht es nicht nur um die Visualisierung von Statistiken, sondern auch darum, zu erklären, was die Visualisierung zeigt.

Kunst

The New York Times ist objektiv. Sie stellt die Daten dar, liefert Ihnen die Fakten und in beidem macht sie einen guten Job. Auf der anderen Seite des Spektrums geht es bei der Visualisierung weniger um Analyse und nüchterne Fakten, sondern darum, Ihre Gefühle anzusprechen. Jonathan Harris und Sep Kamvar haben dies in einem fast wörtlichen Sinn mit *We Feel Fine* gemacht (siehe Abbildung 1.2).



Abbildung 1.2: »We Feel Fine« von Jonathan Harris und Sep Kamvar

Diese interaktive Seite extrahiert Sätze und Ausdrücke aus öffentlichen, persönlichen Blogs und visualisiert diese als herumfliegende Blasen. Jede Blase entspricht einem Gefühl und ist in einer bestimmten Farbe codiert. Auf den ersten Blick sieht es so aus, als ob Menschen im Weltraum herumfliegen. Warten Sie ein wenig und Sie werden sehen, wie die Blasen Cluster bilden. Verwenden Sie die Optionen oben auf der Seite, um die Anzeige zu filtern. Klicken Sie eine einzelne Blase an, um den zugehörigen Texteintrag zu lesen; diese Einträge sind zugleich poetisch und aufschlussreich.

Es gibt viele andere Beispiele wie Golan Levins Website *The Dumpster*, die Blogeinträge untersucht, in denen die Trennung vom Lebenspartner erwähnt wird. *Sumedicina* von Kim Asendorf und Jana Lange erzählt die fiktive Geschichte eines Mannes, der versucht, einer korrupten Organisation zu entfliehen; hierzu werden keine Texte, sondern Graphen und Diagramme verwendet. Andreas Nicolas Fischer hat verschiedene Skulpturen erschaffen, die auf Daten basieren; eine zeigt beispielsweise das Abflauen der Konjunktur in den USA.

www 
Interagieren Sie mit den »Emotionsblasen« und untersuchen Sie sie im Live- und Online-Kunstprojekt von Jonathan Harris und Sep Kamvar auf <http://wefeelfine.org>.

www 
Auf [FlowingData](http://flowingdata.com) finden Sie weitere Beispiele zum Thema Daten und Kunst: <http://datafl.ws/art>.

Der entscheidende Punkt ist, dass es sich bei Daten und deren Visualisierung nicht immer um kalte, harte Fakten handeln muss. Manchmal sind Sie nicht auf der Suche nach analytischen Erkenntnissen. Stattdessen können Sie eine Geschichte auch manchmal aus einer emotionalen Perspektive erzählen und so die Betrachter anregen, sich in den Daten wiederzuerkennen. Es ist so ähnlich wie bei Filmen. Nicht alle Filme müssen Dokumentationen sein und nicht alle Visualisierungen müssen in Form von klassischen Diagrammen daherkommen.

Unterhaltung

Neben dem Journalismus und der Kunst hat die Datenvisualisierung auch im Bereich Unterhaltung Einzug gehalten. Wenn Sie sich Daten in der abstrakten Bedeutung des Wortes vorstellen, losgelöst von Tabellenkalkulationsprogrammen und CSV-Dateien, und auch Statusaktualisierungen und Fotos einschließen, lässt sich das schnell erkennen.

So hat beispielsweise Facebook die Statusaktualisierungen seiner User verwendet, um den glücklichsten Tag des Jahres zu ermitteln. Die Online-Datingsite OkCupid hat Online-Informationen benutzt, um die Anzahl der Lügen und Unwahrheiten einzuschätzen, mit denen Leute versuchen, ihr digitales Selbst besser aussehen zu lassen als das reale (siehe Abbildung 1.3). Diese Analysen hatten nicht das Ziel, das Geschäftsmodell zu verbessern, den Umsatz zu steigern oder Unzulänglichkeiten in einem System zu finden. Dennoch haben sie sich im Web wie ein Lauffeuer verbreitet, und zwar wegen ihres Unterhaltungswertes. Die Daten haben uns und der Gesellschaft, in der wir leben, ein wenig den Spiegel vorgehalten.

Laut Facebook ist übrigens Thanksgiving der glücklichste Tag des Jahres, und OkCupid fand zum Beispiel heraus, dass Leute bei ihrer Körpergröße um circa fünf Zentimeter schummeln.

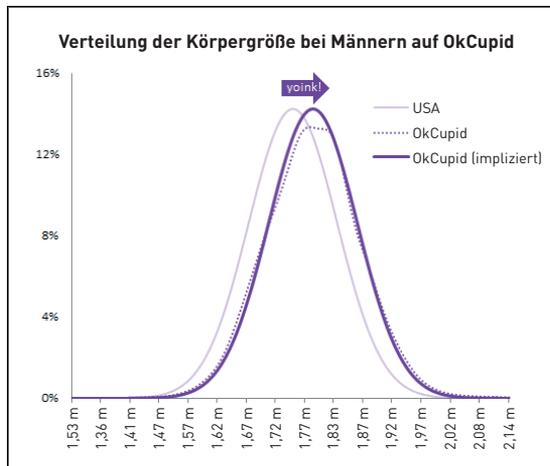


Abbildung 1.3: Prozentuale Verteilung der Körpergröße von Männern auf OkCupid

www



Schauen Sie sich einmal den OkTrend-Blog an, um mehr Erkenntnisse über das Online-Dating zu erhalten, beispielsweise was Weiße wirklich mögen und wie man sich davor bewahren kann, versehentlich hässlich auszusehen:
<http://blog.okcupid.com>.

Fesselnd

Selbstverständlich geht es beim Erzählen von Geschichten nicht nur darum, Leute zu informieren oder zu unterhalten. Manchmal haben sie den Zweck, auf die Dringlichkeit bestimmter Themen hinzuweisen oder Leute auf Trab zu bringen. Wer kann die Szene in der Dokumentation *Eine unbequeme Wahrheit* (original: An Inconvenient Truth) vergessen, in der Al Gore auf der Scherenarbeitsbühne steht, um den zunehmenden Ausstoß von Kohlenstoffdioxid in die Atmosphäre zu verdeutlichen.

Meiner Meinung nach gibt es niemanden, der dies besser gemacht hat als Hans Rosling, Professor für International Health und Direktor der Gapminder Foundation. Rosling verwendet ein Tool mit dem Namen Trendalyzer, das Sie in Abbildung 1.4 sehen, um die Veränderungen bei der Entwicklung der Armut nach Ländern unterteilt aufzuzeigen. Er lässt diese Animation während eines Vortrags ablaufen, der einen zunächst einmal mit den ganzen Daten vertraut macht, und am Ende erhält er fast immer Standing Ovations. Es ist ein wunderbarer Vortrag; falls Sie ihn noch nicht gesehen haben, empfehle ich Ihnen unbedingt, ihn sich einmal anzuschauen (http://www.ted.com/talks/hans_rosling_reveals_new_insights_on_poverty.html).

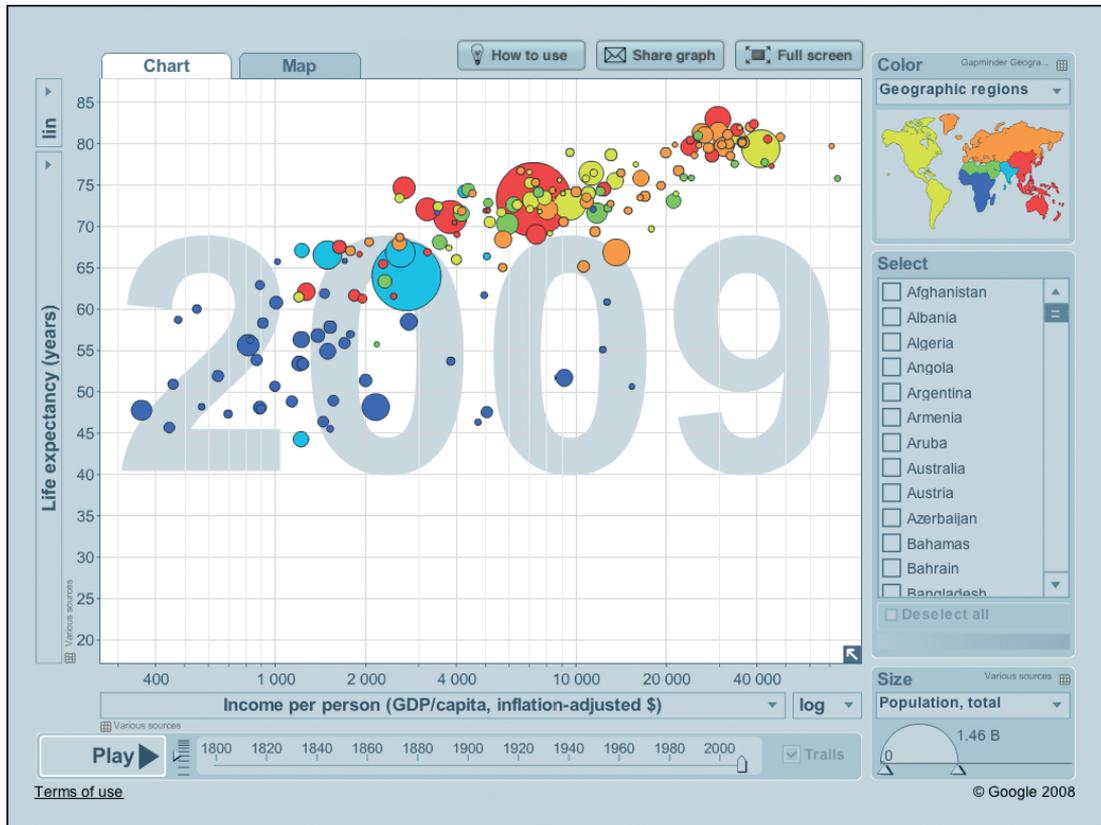


Abbildung 1.4: Trendalyzer von der Gapminder Foundation

Die Visualisierung an sich ist ziemlich einfach: Es handelt sich um ein Bewegungsdiagramm. Die einzelnen Blasen repräsentieren die Länder und die Bewegung basiert auf dem durchschnittlichen Jahreseinkommen. Warum ist der Vortrag dann so populär? Der Grund: Hans Rosling spricht mit Überzeugungskraft und Begeisterung. Er erzählt eine Geschichte. Wie oft haben Sie schon eine Präsentation mit Grafiken und Diagrammen über sich ergehen lassen müssen, bei denen fast jeder eingeschlafen ist. Hans Rosling hingegen erfasst die Bedeutung der Daten und nutzt dies zu seinem Vorteil. Das Schwertschlucken am Ende seines Vortrags verdeutlicht die Kernaussage unmissverständlich: Das scheinbar Unmögliche kann möglich werden. Nachdem ich Roslings Vortrag gesehen hatte, wollte ich unbedingt die Daten haben, die er verwendete. Das war eine Geschichte, die ich unbedingt weiter untersuchen wollte.

Zu einem späteren Zeitpunkt sah ich einen Gapminder-Vortrag zum gleichen Thema mit den gleichen Visualisierungen, jedoch durch einen anderen Redner vorgelesen. Dieser Vortrag war bei Weitem nicht so beeindruckend wie der von Hans Rosling. Um ehrlich zu sein: Der Vortrag war gähnend langweilig. Es gab fast keine Emotionen. Ich habe beim Sprecher keine Begeisterung in Bezug auf die Daten und keine Überzeugungskraft gespürt. Es sind also nicht nur die Daten allein, die etwas interessant machen. Ob sich jemand an Ihre Ausführungen erinnert, liegt vor allem daran, wie Sie die Daten präsentieren und darstellen.

Zusammenfassend lässt sich feststellen, dass es wichtig ist, sich immer die folgenden Punkte vor Augen zu führen: Gehen Sie an die Visualisierung heran, als ob Sie eine Geschichte erzählen wollten. Welche Art von Geschichte wollen Sie erzählen? Ist es ein sachlicher Bericht oder eine Novelle? Wollen Sie andere davon überzeugen, dass es wichtig ist, in die Gänge zu kommen?

Denken Sie an Charakterentwicklung. Hinter jedem Datenpunkt befindet sich eine Geschichte, so wie jeder Charakter in einem Buch eine Vergangenheit, Gegenwart und Zukunft besitzt. Es gibt Interaktionen und Beziehungen zwischen diesen Datenpunkten. Es liegt an Ihnen, diese zu finden. Allerdings müssen auch erfahrene Geschichtenerzähler zuerst lernen, Sätze zu konstruieren, bevor sie einen Roman schreiben können.

Wonach Sie suchen sollten

Okay – nach Geschichten. Das ist klar. Aber welche Art von Geschichte möchten Sie mit Daten erzählen? Nun, die Details hängen immer von den jeweiligen Daten ab. Allgemein ausgedrückt sollten Sie unabhängig davon, wofür Sie die Infografik verwenden, immer nach zwei Dingen Ausschau halten: nach Mustern und nach Beziehungen.

Muster

Dinge ändern sich im Laufe der Zeit. Sie werden älter, Sie bekommen graue Haare und Ihr Sehvermögen lässt nach (siehe Abbildung 1.5). Logos ändern sich. Neue Firmen und Geschäftsmodelle entstehen. Firmen verschwinden. Manchmal treten diese Änderungen plötzlich und ohne Vorwarnung ein. Manchmal verlaufen die Veränderungen so langsam, dass sie kaum bemerkt werden.

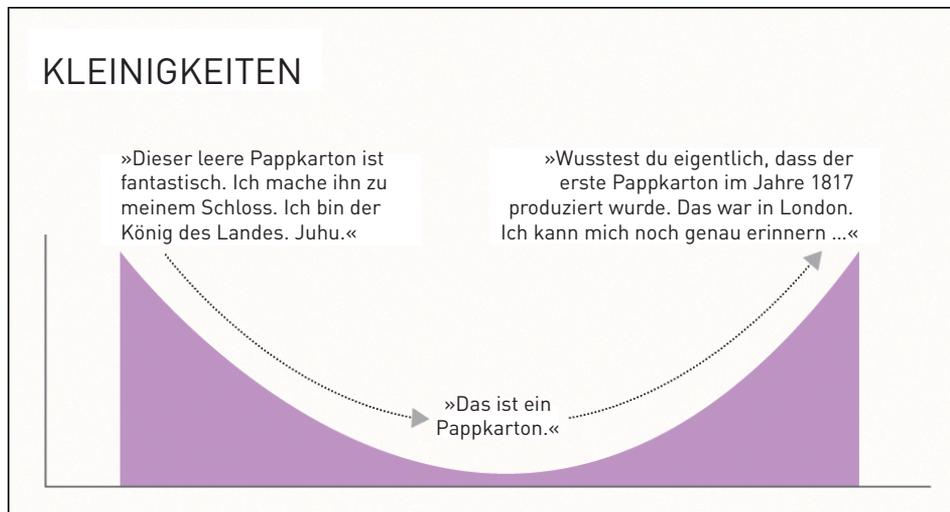


Abbildung 1.5: Ein witziger Blick auf das Altern

Unabhängig davon, was genau Sie sich anschauen, kann sowohl die Veränderung interessant sein als auch der Veränderungsprozess selbst. Auf diese Weise können Sie Muster über einen längeren Zeitraum untersuchen. Angenommen, Sie schauen sich die Aktienkurse in einer bestimmten Periode an. Die Kurse fallen und steigen, das ist offensichtlich. Wie sieht es aber mit den Änderungen pro Tag aus, wie mit denen pro Woche oder mit denen pro Monat? Gibt es Zeiträume, in denen die Börsenkurse mehr stiegen als üblich? Wenn ja, warum ist das so? Gab es bestimmte Ereignisse, die diese Änderung hervorgerufen haben?

Wie Sie sehen, führt eine einzige Frage, die als Ausgangspunkt dient, fast automatisch zu weiteren Fragen. Dies trifft nicht nur für Zeitreihendaten zu, sondern für alle Arten von Daten. Versuchen Sie, sich Ihren Daten eher aus einer forschenden und fragenden Sicht zu nähern, da Sie so in den meisten Fällen mehr und interessantere Antworten finden.

Sie können Ihre Zeitreihendaten auf unterschiedliche Arten aufteilen. In manchen Fällen ist es sinnvoll, stündliche oder tägliche Werte zu verwenden. In anderen Fällen ist es hingegen besser, die Daten auf Monats- oder Jahresbasis zu betrachten. Beim ersten Format erhalten Sie möglicherweise ein Diagramm, das mehr Rauschen enthält, wohingegen der Einsatz größerer Zeiträume zu einer aggregierten Sicht der Daten führt.

Diejenigen von Ihnen, die Websites betreiben, auf denen Daten über die Besuche und Seitenaufrufe gesammelt werden, finden sich hier schnell wieder. Wenn Sie sich den Datenverkehr auf Ihrer Website mit den täglichen Daten ansehen, wird das Diagramm etwas zackelig. Es gibt viel mehr Schwankungen (siehe Abbildung 1.6).

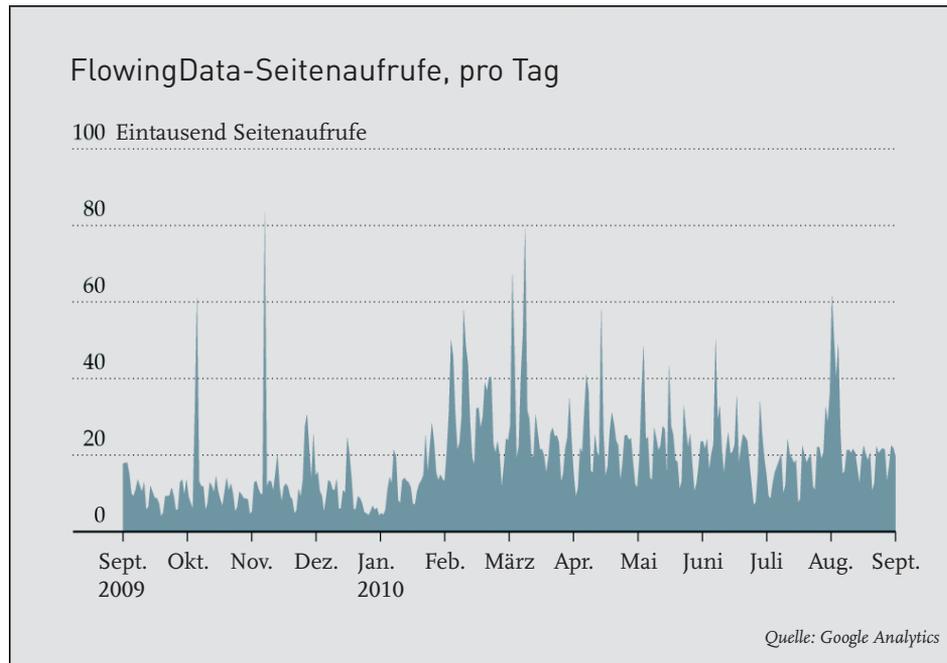


Abbildung 1.6: Unique Visitors auf FlowingData auf Tagesbasis

Wenn Sie sich die gleichen Daten auf einer Monatsbasis ansehen (was Abbildung 1.7 zeigt), befinden sich in dem Diagramm, das den gleichen Zeitraum abdeckt, weniger Datenpunkte, wodurch das Diagramm gleichmäßiger aussieht.

Ich will damit nicht sagen, dass das eine Diagramm per se besser ist als das andere. In Wahrheit ergänzen sie einander. Wie Sie die Daten aufteilen, hängt davon ab, wie viel Details Sie darstellen wollen (oder eben nicht).

Muster über Zeiträume sind selbstverständlich nicht die einzigen, nach denen Sie suchen sollten. Sie können Muster auch in Aggregationen entdecken, mit denen Gruppen, Personen oder Dinge verglichen werden. Was essen oder trinken Sie regelmäßig jede Woche? Welche Themen behandelt der US-amerikanische Präsident in seiner Rede an die Nation? Welche US-Bundesstaaten wählen üblicher-

weise die Republikaner? In diesen Fällen ist es sinnvoll, nach Mustern in geographischen Regionen zu suchen. Auch wenn sich die Fragen und Datentypen unterscheiden, ist der grundsätzliche Ansatz immer sehr ähnlich, wie Sie in den folgenden Kapiteln noch sehen werden.

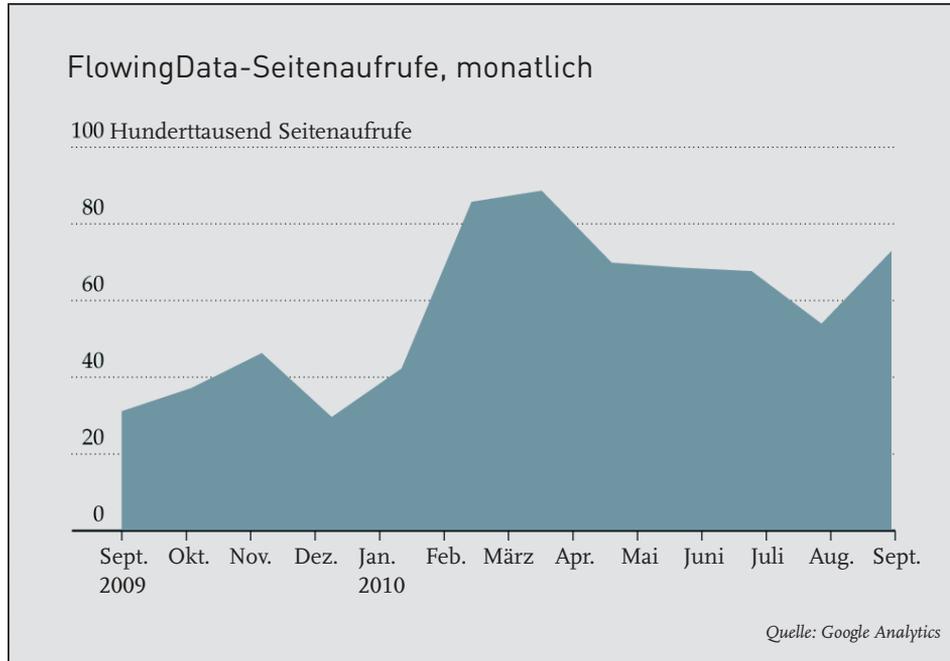


Abbildung 1.7: Unique Visitors auf FlowingData pro Monat

Beziehungen

Haben Sie schon mal eine Infografik gesehen, die aus zahlreichen Diagrammen bestand und so aussah, als ob die Diagramme willkürlich platziert wurden? Ich meine damit die Grafiken, die so wirken, als ob ihnen etwas Spezielles fehlt, als ob der Grafiker sich nur wenig mit den Daten befasst und dann eine Grafik zusammengeschnitten hat, um eine Deadline einzuhalten. In vielen Fällen ist das Spezielle die Beziehungen in den Daten.

In der Statistik ist hiermit Korrelation und Kausalität gemeint. Mehrere Variablen können auf verschiedene Weise miteinander in Beziehung stehen. Kapitel 6 »Beziehungen visualisieren« behandelt diese Konzepte und zeigt, wie sie visualisiert werden können.

Auf einer abstrakteren Ebene, auf der Sie sich nicht um Formeln und Hypothesentests kümmern, können Sie Ihre Grafiken so gestalten, dass Werte verglichen und gegenübergestellt und Verteilungen sichtbar werden. Ein einfaches Beispiel sehen Sie in Abbildung 1.8, das den Ausschnitt *Technologie* aus dem *World Progress Report* zeigt.

Das Beispiel enthält verschiedene Histogramme, die die Anzahl der Internet-User, Internet-Abonnenten und Breitbandanschlüsse je 100 Einwohner zeigen. Beachten Sie, dass der Bereich für Internet-User (0 bis 95 je 100 Einwohner) viel größer ist als bei den beiden anderen Datensätzen.

www 

Der *World Progress Report* ist ein grafisch aufbereiteter Bericht basierend auf Daten von UNdata, der den Fortschritt auf der Welt vergleicht. Den vollständigen Report finden Sie unter <http://dataf1.ws/12i>.

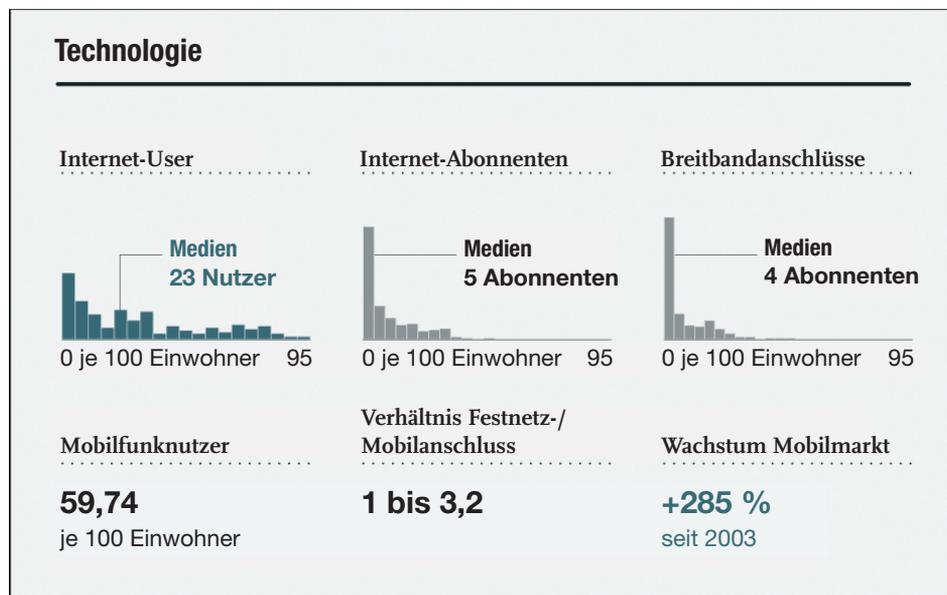


Abbildung 1.8: Weltweite Technologienutzung

Der schnelle und einfache Weg hätte darin bestanden, Ihre Software entscheiden zu lassen, welche Wertebereiche für die einzelnen Histogramme benutzt werden sollen. Jedoch wurden die einzelnen Histogramme im Beispiel so gestaltet, dass alle den gleichen Wertebereich verwenden, auch wenn es keine Länder gab, die weder 95 Internet-Abonnenten noch Breitbandanschlüsse je 100 Einwohner hatten. Diese Art der Darstellung macht es viel einfacher, die Verteilungen zwischen den Gruppen miteinander zu vergleichen.

Falls Sie also mit mehreren und unterschiedlichen Datensätzen arbeiten, stellen Sie sie sich als mehrere Gruppen vor und nicht als getrennte Einheiten, die nicht miteinander interagieren. Dies führt meistens zu interessanteren Ergebnissen.

Fragwürdige Daten

Wenn Sie nach den Geschichten in Ihren Daten suchen, sollten Sie das, was Sie sehen, immer infrage stellen. Vergessen Sie nicht: Nur weil es sich um Zahlen handelt, müssen diese nicht unbedingt wahr sein.

Ehrlicherweise muss ich zugeben, dass die Überprüfung nicht gerade zu meinen Lieblingsaufgaben beim Erstellen einer Grafik zählt. Wenn jemand, eine Gruppe oder ein Auftraggeber Sie mit Daten versorgt, liegt es doch bei ihm, dass die Daten stimmen. Dennoch zeichnet die Überprüfung der Daten die Arbeit eines guten Grafikdesigners aus. Schließlich verwenden zuverlässige Bauarbeiter auch keinen minderwertigen Zement für das Fundament eines Hauses; daher sollten auch Sie Ihre Infografiken nicht auf minderwertigen Daten aufbauen.

Die Überprüfung und Verifizierung der Daten ist eine der wichtigsten, wenn nicht sogar die wichtigste Teilaufgabe beim Entwerfen von Infografiken.

Wonach Sie immer suchen sollten, sind Dinge, die keinen Sinn ergeben. Möglicherweise ist bei der Eingabe der Daten ein Fehler aufgetreten und jemand hat eine weitere Null angefügt oder eine vergessen. Oder es gab Probleme mit der Internetanbindung während des Daten-Scraping, wodurch an einigen zufälligen Stellen die Daten verfälscht wurden. Was immer es auch sein mag, in Fällen, in denen Ihnen die Daten oder bestimmte Teile merkwürdig vorkommen, sollten Sie sich bei der Datenquelle rückversichern.

Die Person, von der Sie die Daten erhalten haben, hat üblicherweise eine Ahnung davon, was zu erwarten ist. Wenn Sie selbst die Daten zusammengetragen haben, fragen Sie sich selbst, ob die Daten so sinnvoll sind, wie sie sind. Angenommen, nur ein Staat hat 90 Prozent von was auch immer und die Werte aller anderen Staaten liegen im Bereich zwischen 10 und 20 Prozent. Woran liegt das?

In vielen Fällen ist die Ursache einer Abweichung ein einfacher Tippfehler. In anderen Fällen handelt es sich um einen interessanten Punkt in Ihrem Datensatz, den Sie als Aufhänger für Ihre Geschichte verwenden können. Achten Sie darauf, dass Sie wissen, um welche Art der Abweichung es sich handelt.

Design

Wenn all Ihre Daten in Ordnung sind, können Sie mit der Visualisierung loslegen. Egal, ob Sie eine Grafik für einen Bericht erstellen, eine Infografik für das Web oder ein Datenkunstwerk, Sie sollten dabei ein paar Regeln folgen. Bei allen Regeln gibt es genügend Spielraum. Daher sollten Sie die folgenden Absätze wie ein Gerüst verwenden, und nicht wie ein in Stein gemeißeltes Gesetz. Wenn Sie gerade erst mit der Erstellung von Datengrafiken anfangen, sind die folgenden Informationen ein guter Ausgangspunkt.

Erklären Sie Ihre Codierung

Das Design jeder Grafik erfolgt immer in ähnlichen Schritten. Sie erhalten die Daten; Sie codieren die Daten als Kreise, Balken und/oder mittels Farben; abschließend wird die Grafik von anderen gelesen. An dieser Stelle müssen die Leser oder Betrachter Ihre Codierung wieder decodieren. Was stellen die Kreise, Balken und Farben dar?

William Cleveland und Robert McGill haben ausführliche Informationen zum Thema Codierung verfasst. Manche Codierungen funktionieren besser als andere. Wenn Ihre Leser jedoch nicht wissen, was sich hinter der Codierung verbirgt, spielt es keine Rolle, für welche Art der Codierung Sie sich entscheiden. Wenn sie es nicht decodieren können, war die Zeit, die Sie mit dem Erstellen der Grafik zugebracht haben, sinnlos vergeudet.

Das Fehlen des Zusammenhangs finden Sie ab und zu bei Grafiken, die im Bereich zwischen Datenkunst und Infografik angesiedelt sind. Bei reiner Datenkunst ist dieses Phänomen fast immer anzutreffen. Eine Beschriftung oder eine Legende kann die Atmosphäre und Ausstrahlung des Kunstwerks ziemlich vermurksen. Aber Sie können stattdessen zumindest einige Informationen in einen kurzen, erklärenden Text aufnehmen. Dies hilft anderen dabei, Ihren Aufwand zu würdigen.

Gelegentlich findet man das Problem des fehlenden Kontexts auch bei reinen Datengrafiken. Dies kann Ihre Leser frustrieren, und das ist das Letzte, was Sie wollen. Manchmal vergessen Sie es vielleicht, nachdem Sie sich ja lange mit den Daten auseinandergesetzt haben und daher genau wissen, was sie bedeuten. Ihre Leser jedoch kennen den Zusammenhang nicht, den Sie aus der Analyse entwickelt haben.

Hinweis



Mehr Informationen darüber, wie Formen und Farben codiert werden, finden Sie im Artikel *Graphical Perception and Graphical Methods for Analyzing Scientific Data* von Cleveland/McGill. Einen Link zum betreffenden PDF finden Sie unter anderem hier: <http://flowingdata.com/2010/03/20/graphical-perception-learn-the-fundamentals-first/>.

Wie können Sie also sicherstellen, dass die Leser Ihre Codierungen decodieren können? Erklären Sie deren Bedeutung mithilfe von Beschriftungen und Legenden. Was Sie am besten verwenden, hängt von der Situation ab. Nehmen Sie als Beispiel die Weltkarte in Abbildung 1.9, die die Verbreitung von Firefox nach Ländern zeigt.

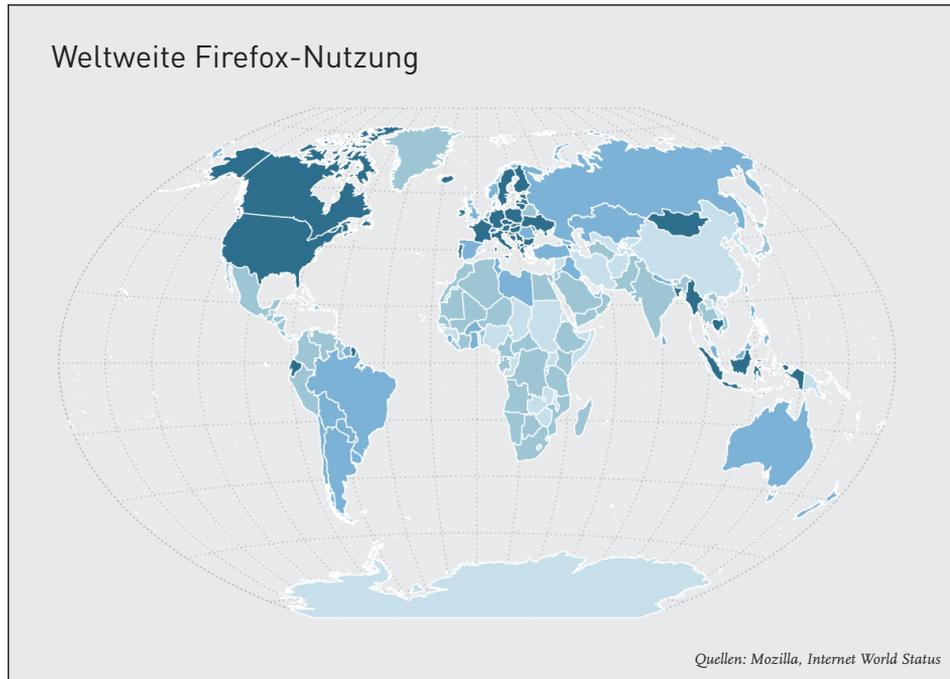


Abbildung 1.9: Die Nutzung von Firefox nach Ländern

Die verschiedenen Blauschattierungen für verschiedene Länder sind zu erkennen, aber was bedeuten sie? Entspricht ein dunkles Blau weiter oder weniger verbreitet? Wenn Dunkelblau bedeutet, dass Firefox weiter verbreitet ist, was genau bedeutet dann weiter verbreitet? In ihrem derzeitigen Zustand ist die Karte ziemlich nutzlos. Falls Sie jedoch eine Legende ergänzen, wie sie Abbildung 1.10 zeigt, werden die Dinge deutlich. Die Farblegende hat sogar eine doppelte Funktion, da sie gleichzeitig ein Histogramm ist, das die Häufigkeitsverteilung je 1.000 Internet-User zeigt.

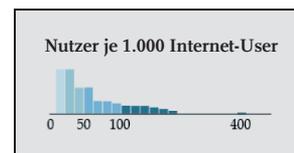


Abbildung 1.10: Legende für die Karte der Firefox-Nutzung

Alternativ können Sie Formen und Objekte in der Grafik auch direkt beschriften, falls Ihnen ausreichend Platz zur Verfügung steht und Sie nicht zu viele Kategorien verwenden. Ein Beispiel zeigt Abbildung 1.11. Dieses Diagramm zeigt die Anzahl der Oscarnominierungen, bevor der Schauspieler den Oscar als bester Schauspieler erhält.

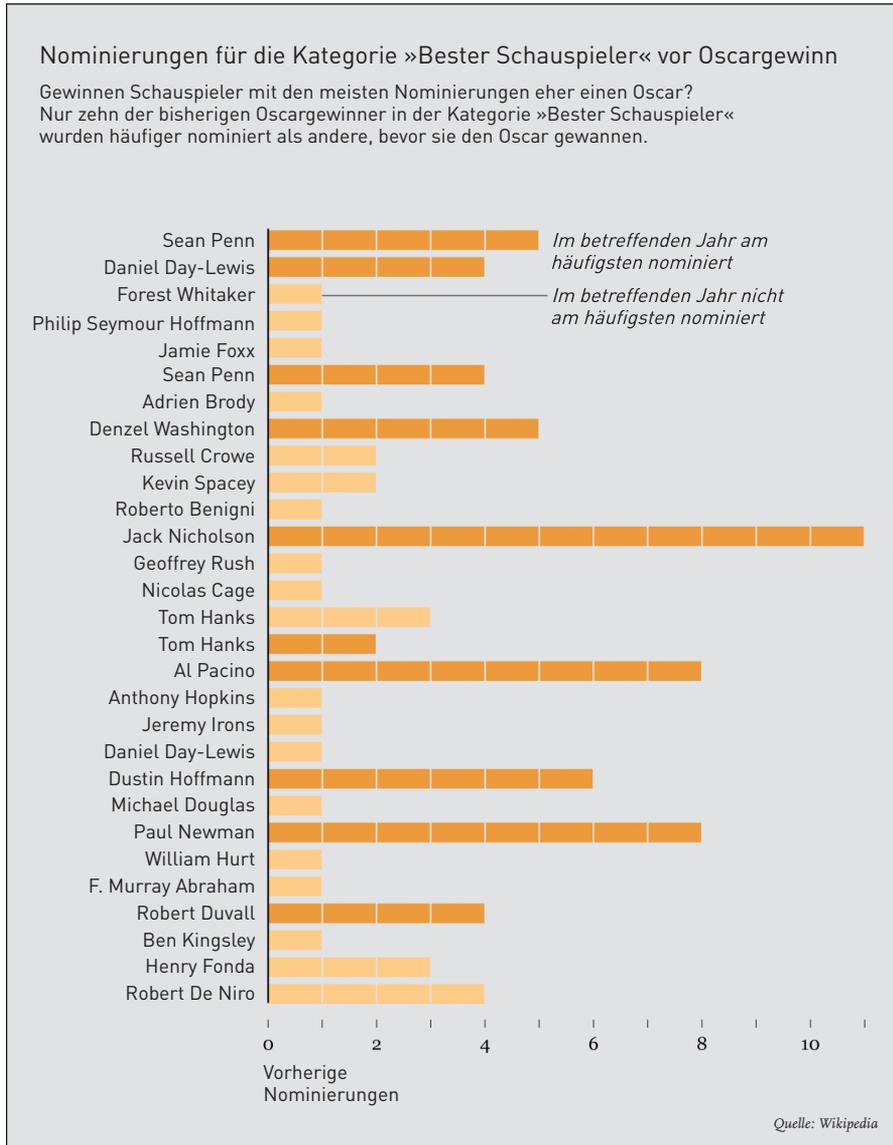


Abbildung 1.11: Beispiel mit direkt beschrifteten Objekten

Im Internet kursierte die Theorie, dass die Schauspieler, die verglichen mit ihren Kollegen in einem bestimmten Jahr am häufigsten nominiert wurden, dann üblicherweise den Oscar gewannen. An der Beschriftung in der Grafik erkennen Sie, dass für Schauspieler mit den meisten Nominierungen ein dunkles Orange verwendet wird, für die anderen ein helles Orange.

Die obigen Beispiele zeigen, dass Ihnen zahlreiche Optionen zur Verfügung stehen. Sie sind einfach einsetzbar. Gleichzeitig machen diese kleinen Details Ihre Grafik sehr viel besser lesbar.

Beschriften Sie die Achsen

Genauso wie Sie Ihre Codierungen erläutern sollten, sollten Sie auch immer die Achsen eines Diagramms beschriften. Fehlt die Beschriftung oder Erläuterung, sind die Achsen reine Dekoration. Geben Sie bei der Achsenbeschriftung an, auf welche Skala sich die Datenpunkte beziehen. Ist sie logarithmisch, inkrementell, exponentiell oder bezieht sie sich auf je 100 Toiletten? Ich persönlich gehe, falls die Beschriftung fehlt, immer von Letzterem aus.

Wenden wir uns einem kleinen Wettbewerb zu, den ich vor ein paar Jahren auf FlowingData durchgeführt habe und der diesen Punkt verdeutlichen soll. Ich hatte die Grafik, die Sie in Abbildung 1.12 sehen, gepostet und die Leser aufgefordert, eine möglichst unterhaltsame Achsenbeschriftung zu ersinnen.

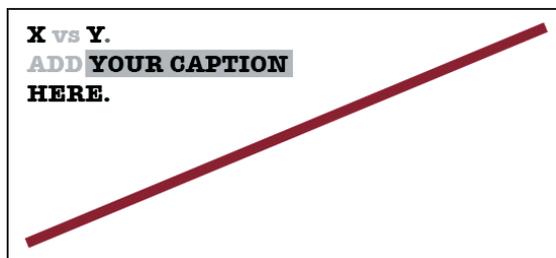


Abbildung 1.12: Ergänzen Sie Ihre Beschriftung.

Es gab mehr als 60 verschiedene Beschriftungen für ein und dasselbe Diagramm. Abbildung 1.13 zeigt eine kleine Auswahl.

An den Beispielen können Sie gut erkennen, dass – obwohl alle das gleiche Diagramm verwenden – eine Änderung der Achsenbeschriftungen eine komplett andere Geschichte erzählt. Zugegeben, dies war nur ein Spiel. Stellen Sie sich nun

vor, dass Ihr Diagramm ernst genommen werden soll. Ohne Beschriftungen ist Ihr Diagramm bedeutungslos.

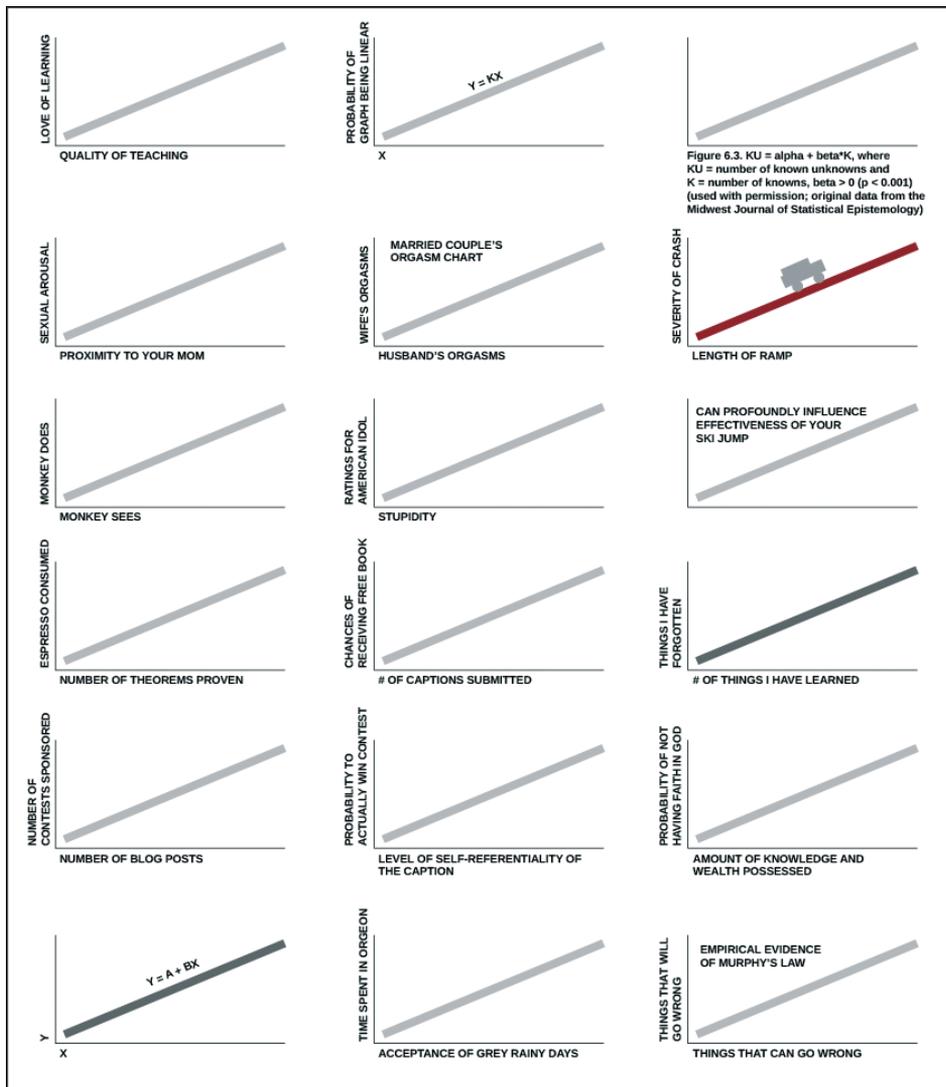


Abbildung 1.13: Einige der Einsendungen zu einem Beschriftungswettbewerb auf FlowingData

Halten Sie die Geometrie unter Kontrolle

Beim Entwerfen eines Diagramms verwenden Sie geometrische Formen. Ein Säulendiagramm benutzt Rechtecke, wobei die Höhe der einzelnen Rechtecke Werte repräsentiert. In einem Punktdiagramm wird der Wert durch die Position des Punkts symbolisiert, ähnlich wie bei einem Standardzeitreihendiagramm. Ein Kreisdiagramm verwendet Winkel, um den Wert anzugeben, und die Summe der einzelnen Werte entspricht hier immer 100 Prozent (siehe Abbildung 1.14). Dies ist alles recht einfach; bleiben Sie aber dennoch auf der Hut, da es genauso einfach ist, die Dinge zu vermurksen. Wenn Sie nicht sorgfältig genug sind, passieren Ihnen Fehler. Und wenn Sie Fehler machen, haben andere, besonders im Web, nur wenig Hemmungen, Sie deutlich darauf hinzuweisen.

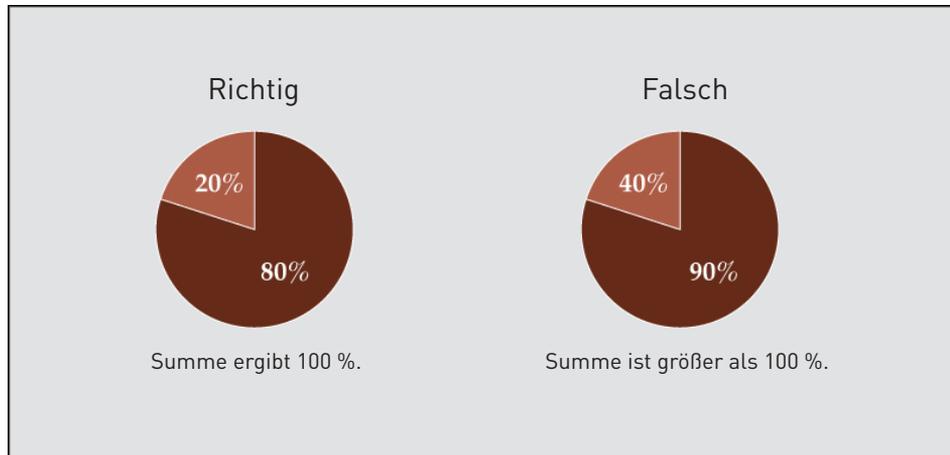


Abbildung 1.14: Der richtige und der falsche Weg, ein Kreisdiagramm zu erstellen

Ein weiterer häufig vorkommender Fehler besteht darin, bei der Darstellung von Werten zweidimensionale Formen zu verwenden, aber die Größe der Formen so einzustellen, als ob es sich um eindimensionale Formen handelt. Die Rechtecke eines Säulendiagramms haben immer zwei Dimensionen, jedoch wird immer nur die Länge verwendet, um den Wert anzugeben. Die Breite des Rechtecks hat keine Bedeutung. Wenn Sie jedoch ein Blasendiagramm erstellen, verwenden Sie die Fläche, um Werte darzustellen. Anfänger verwenden stattdessen häufig den Radius oder den Durchmesser, was zu einer falschen Skalierung führt.

Abbildung 1.15 zeigt zwei Kreise, bei denen der Wert durch die Kreisfläche dargestellt wird. Dies ist der korrekte Weg.

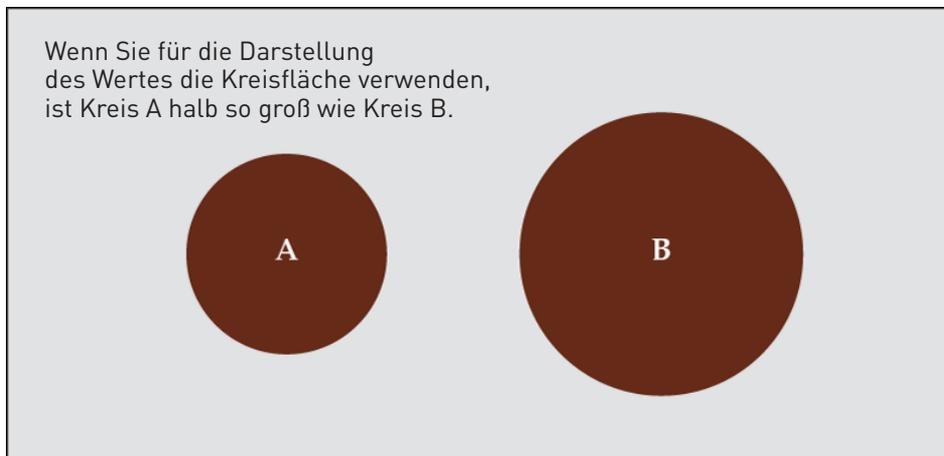


Abbildung 1.15: Der richtige Weg, um die Größe von Blasen einzustellen

Abbildung 1.16 hingegen zeigt zwei Kreise, bei denen der Durchmesser zur Darstellung des Wertes verwendet wurde. Der zweite Kreis hat den doppelten Durchmesser des ersten, jedoch ist seine Fläche viermal so groß.

Das Gleiche gilt für Rechtecke, die in einer Treemap verwendet werden. Dort verwenden Sie die Fläche der Rechtecke, um den Wert darzustellen und nicht ihre Länge oder Breite.

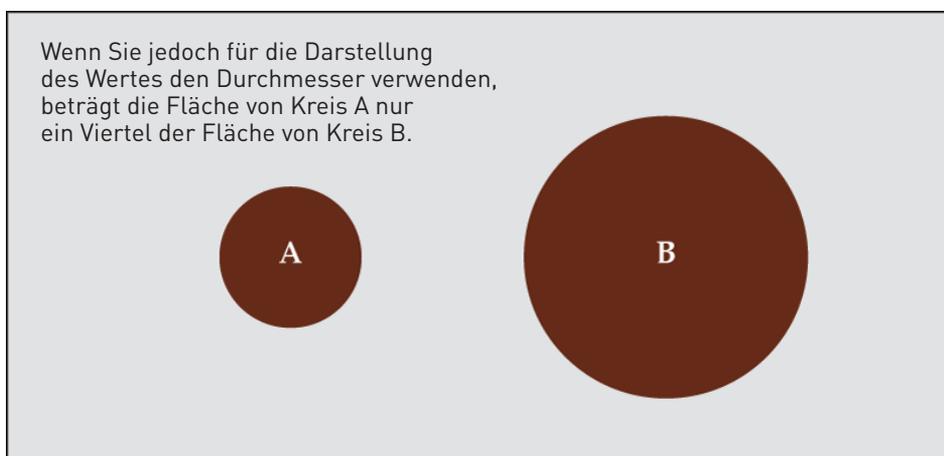


Abbildung 1.16: Der falsche Weg, um einen Wert in der Größe von Blasen einzustellen

Geben Sie die Datenquellen an

Eigentlich sollte dies selbstverständlich sein, dennoch vergessen es viele Leute. Woher stammen die Daten? Wenn Sie sich Grafiken in Zeitungen ansehen, wird die Datenquelle dort irgendwo angegeben, meist in kleiner Schriftgröße im unteren Bereich der Grafik. Sie sollten das Gleiche tun. Andernfalls wissen Ihre Leser nicht, wie zuverlässig Ihre Grafik ist.

Die Leser wissen ansonsten nicht, ob die Daten nicht einfach nur erfunden wurden. Selbstverständlich würden Sie so etwas nie tun, aber nicht alle wissen das. Die Angabe der Datenquelle macht Ihre Grafik nicht nur glaubwürdiger, sondern ermöglicht es anderen auch, die Daten zu prüfen und zu analysieren.

Das Benennen der Datenquelle stellt die Zahlen gleichzeitig in einen Kontext. Es ist offenkundig, dass eine Befragung, die auf einem Volksfest oder Jahrmarkt durchgeführt wurde, anders interpretiert wird, als eine, bei der die Daten im Rahmen einer Volkszählung erhoben wurden.

Berücksichtigen Sie Ihre Zielgruppe

Berücksichtigen Sie immer Ihre Zielgruppe und den Zweck Ihrer Grafik. So sollte ein Diagramm, das für eine Folienpräsentation erstellt wird, eher einfach gehalten sein. Zwar können Sie eine Menge Details in das Diagramm einfügen, jedoch werden nur die Leute in den ersten Reihen in der Lage sein, sie zu erkennen. Falls Sie hingegen ein Poster entwerfen, das studiert und genau betrachtet werden soll, können Sie viel mehr Details unterbringen.

Arbeiten Sie an einem Geschäftsbericht? Dann sollten Sie nicht versuchen, das schönste Datenkunstobjekt zu erstellen, das die Welt jemals gesehen hat. Erstellen Sie stattdessen eine aussagekräftige Datengrafik, die direkt zur Sache kommt. Verwenden Sie die Grafik im Rahmen der Datenanalyse? Dann ist die Grafik nur für Ihre Augen bestimmt und es ist vermutlich nicht erforderlich, viel Zeit für die Ästhetik und die Beschriftung der Daten aufzuwenden. Soll Ihre Grafik einem breiten Publikum zugänglich gemacht werden? Werden Sie in diesem Fall nicht zu kompliziert und erläutern Sie anspruchsvolle Konzepte.

Zusammenfassung

Beginnen Sie mit einer Frage, untersuchen Sie Ihre Daten mit kritischem Blick, machen Sie sich den Zweck Ihrer Grafik deutlich und für wen sie bestimmt ist. Diese Vorgehensweise hilft, eine deutliche Grafik zu erstellen, die es wert ist, dass die Leute sich damit befassen – und dies gilt unabhängig von der Art der Grafik.

Wie Sie dies bewerkstelligen, wird in den weiteren Kapiteln beschrieben. Sie werden den Umgang mit Daten und deren Visualisierung erlernen. Sie werden von A bis Z lernen, wie Sie Grafiken erstellen. Anschließend können Sie das Erlernte auf Ihre eigenen Daten anwenden. Machen Sie sich klar, welche Geschichte Sie erzählen wollen, und erstellen Sie die Grafik so, dass sie zu der Geschichte passt.

