

O'REILLY®

Daten- visualisierung Grundlagen & Praxis

Wie Sie aussagekräftige
Diagramme und Grafiken
gestalten



Claus O. Wilke
Übersetzung von Bilgehan Gür

Inhalt

Cover

Titel

Impressum

Inhalt

Vorwort

1 Einleitung

Hässliche, schlechte und falsche Abbildungen

TEIL I: Von den Daten zur Visualisierung

2 Datenvisualisierung: die Darstellung von Daten gestalten
(Aesthetics)

Aesthetics (Gestaltungselemente) und Datenarten

Skalen bilden Datenwerte auf Aesthetics ab

3 Koordinatensysteme und Achsen

Kartesische Koordinaten

Nichtlineare Achsen

Koordinatensysteme mit gekrümmten Achsen

4 Farbskalen

Farbe als Unterscheidungsmerkmal

Farbe zur Darstellung von Datenwerten

Farbe als Werkzeug zur Hervorhebung

5 Ausgewählte Visualisierungen

Quantitative Werte

Verteilungen

Proportionen

x-y-Relationen

Geodaten

(Mathematische) Unsicherheit

6 Visualisierung quantitativer Werte

Balkendiagramme

Gruppierte und gestapelte Balken

Punktdiagramme und Heatmaps

7 Visualisierung von Verteilungen: Histogramme und Dichtediagramme

Visualisierung einer einzelnen Verteilungsgröße

Gleichzeitige Visualisierung mehrerer Verteilungsgrößen

8 Visualisierung von Verteilungen: Empirisch kumulative Häufigkeitsverteilung und Q-Q-Plots

Empirisch kumulative Häufigkeitsverteilung

Stark verzerrte Verteilungen

Quantil-Quantil-Diagramme

9 Gleichzeitige Visualisierung mehrerer Verteilungen

Visualisierung von Verteilungen entlang der vertikalen Achse

Visualisierung von Verteilungen entlang der horizontalen Achse

10 Visualisierung von Proportionen

Ein Fall für Kreisdiagramme

Ein Fall für nebeneinander angeordnete Balken (engl. Side-by-side bars)

Ein Fall für gestapelte Balken und gestapelte Dichten

Proportionen separat als Teile der Summe visualisieren

11 Visualisierung verschachtelter Proportionen

Fehlerhaft verschachtelte Proportionen

Mosaikdiagramme und Treemaps

Verschachtelte Kreisdiagramme

Parallel Sets

12 Visualisierung von Korrelationen zwischen zwei oder mehr quantitativen Variablen

Streudiagramme (engl. Scatterplots)

Korrelogramme

Dimensionsreduktion

Gepaarte Daten

13 Visualisierung von Zeitreihen und anderen Funktionen einer unabhängigen Variablen

Einzelne Zeitreihen

Mehrere Zeitreihen und Dosiswirkungskurven

Zeitreihe von zwei oder mehr Antwortvariablen

14 Visualisierung von Trends

Glättung

Trends mit einer definierten funktionalen Form anzeigen

Trendbereinigung und Zerlegung von Zeitreihen

15 Visualisierung von Geodaten

Projektionen

Schichten

Choroplethenkartierung

Kartogramme

16 Visualisierung von Ungenauigkeiten

Wahrscheinlichkeiten als Häufigkeiten darstellen

Visualisierung der Unsicherheit von Punktschätzungen
Visualisierung der Unsicherheit von Kurvenanpassungen
Hypothetische Ergebniskurven

TEIL II: Prinzipien des Graphen-Designs

17 Das Prinzip proportionaler Farbflächen

Visualisierungen entlang linearer Achsen
Visualisierungen entlang logarithmischer Achsen
Direkte Flächendarstellungen

18 Umgang mit überlappenden Punkten

Teiltransparenz und Jittering
2D-Histogramme
Konturlinien

19 Typische Fallstricke beim Gebrauch von Farbe

Anzeige zu vieler oder irrelevanter Informationen
Verwendung nichtmonotoner Farbskalen zur Darstellung von Datenwerten
Ungeeignete Abbildungen für Leser mit Farbsehschwäche

20 Redundante Codierung

Gestaltung von Legenden mit redundanter Codierung
Gestaltung von Abbildungen ohne Legenden

21 Multipanel-Diagramme

Small Multiples
Zusammengesetzte (modulare) Abbildungen

22 Titel, Beschriftungen und Tabellen

Titel und Beschriftungen
Achsen- und Legendenbeschriftung

Tabellen

23 Bringen Sie Daten und Kontext in Einklang

Den Kontext in angemessenem Umfang bereitstellen

Hintergrundgitter

Gepaarte Daten

Zusammenfassung

24 Verwenden Sie größere Achsenbeschriftungen

25 Vermeiden Sie Linienzeichnungen

26 Vermeiden Sie 3D

Vermeiden Sie unnötige 3D-Darstellungen

Vermeiden Sie 3D-Positionsskalen

Angemessene Verwendung von 3D-Visualisierungen

TEIL III: Verschiedene Themen

27 Die am häufigsten verwendeten Bilddateiformate verstehen

Bitmap- und Vektorgrafiken

Verlustfreie und verlustbehaftete Datenkomprimierung von Bitmapgrafiken

Von einem Bildformat in ein anderes konvertieren

28 Auswahl der richtigen Visualisierungssoftware

Reproduzierbarkeit und Wiederholbarkeit

Datenexploration versus Datenpräsentation

Trennung von Inhalt und Design

29 Eine Geschichte erzählen und Erkenntnisse auf den Punkt bringen

Was ist eine Geschichte?

Erstellen Sie eine Abbildung für die Generäle

Komplexe Abbildungen aufbauen

Gestalten Sie Ihre Abbildungen einprägsam

Seien Sie konsistent, aber wiederholen Sie sich nicht

A Kommentierte Bibliografie

Nachdenken über Daten und Visualisierung

Programmierbücher

Statistikbücher

Ältere Bücher

Bücher zu verwandten Themen

B Technische Hinweise

C Referenzen

Index

Über den Autor

Kolophon

Koordinatensysteme und Achsen

Für jede Art von Datenvisualisierung müssen Positionsskalen definiert werden, die bestimmen, wo sich in einer Grafik unterschiedliche Datenwerte befinden. Wir können Daten nicht visualisieren, ohne verschiedene Datenpunkte an verschiedenen Orten zu platzieren, auch wenn wir sie nur entlang einer Linie nebeneinander anordnen. Für normale 2D-Visualisierungen sind zwei Zahlen erforderlich, um einen Punkt eindeutig zu spezifizieren, und daher benötigen wir zwei Positionsskalen. Diese beiden Maßstäbe sind normalerweise, aber nicht unbedingt, die x - und die y -Achse des Diagramms. Wir müssen auch die relative geometrische Anordnung dieser Skalen angeben. Üblicherweise verläuft die x -Achse horizontal und die y -Achse vertikal, aber wir könnten auch andere Anordnungen wählen.

Beispielsweise könnte die y -Achse in einem spitzen Winkel zur x -Achse verlaufen, oder eine Achse könnte in einem Kreis und die andere in radialer Richtung verlaufen. Die Kombination eines Satzes von Positionsskalen und ihrer relativen geometrischen Anordnung wird als *Koordinatensystem* bezeichnet.

Kartesische Koordinaten

Das am häufigsten verwendete Koordinatensystem für die Datenvisualisierung ist das *kartesische 2D-Koordinatensystem*, bei dem jeder Ort durch einen x - und einen y -Wert eindeutig angegeben wird. Die x - und die y -Achse verlaufen orthogonal zueinander, und

die Datenwerte werden in einem gleichmäßigen Abstand entlang beider Achsen platziert (Abbildung 3-1). Die beiden Achsen sind durchgehende Positionsskalen, und sie können sowohl positive als auch negative reelle Zahlen darstellen. Um das Koordinatensystem vollständig zu definieren, müssen wir den Zahlenbereich angeben, den jede Achse abdeckt. In Abbildung 3-1 verläuft die x -Achse von $-2,2$ bis $3,2$ und die y -Achse von $-2,2$ bis $2,2$. Alle Datenwerte zwischen diesen Achsengrenzen werden an der entsprechenden Stelle in der Abbildung platziert. Alle Datenwerte außerhalb der Achsengrenzen werden verworfen.

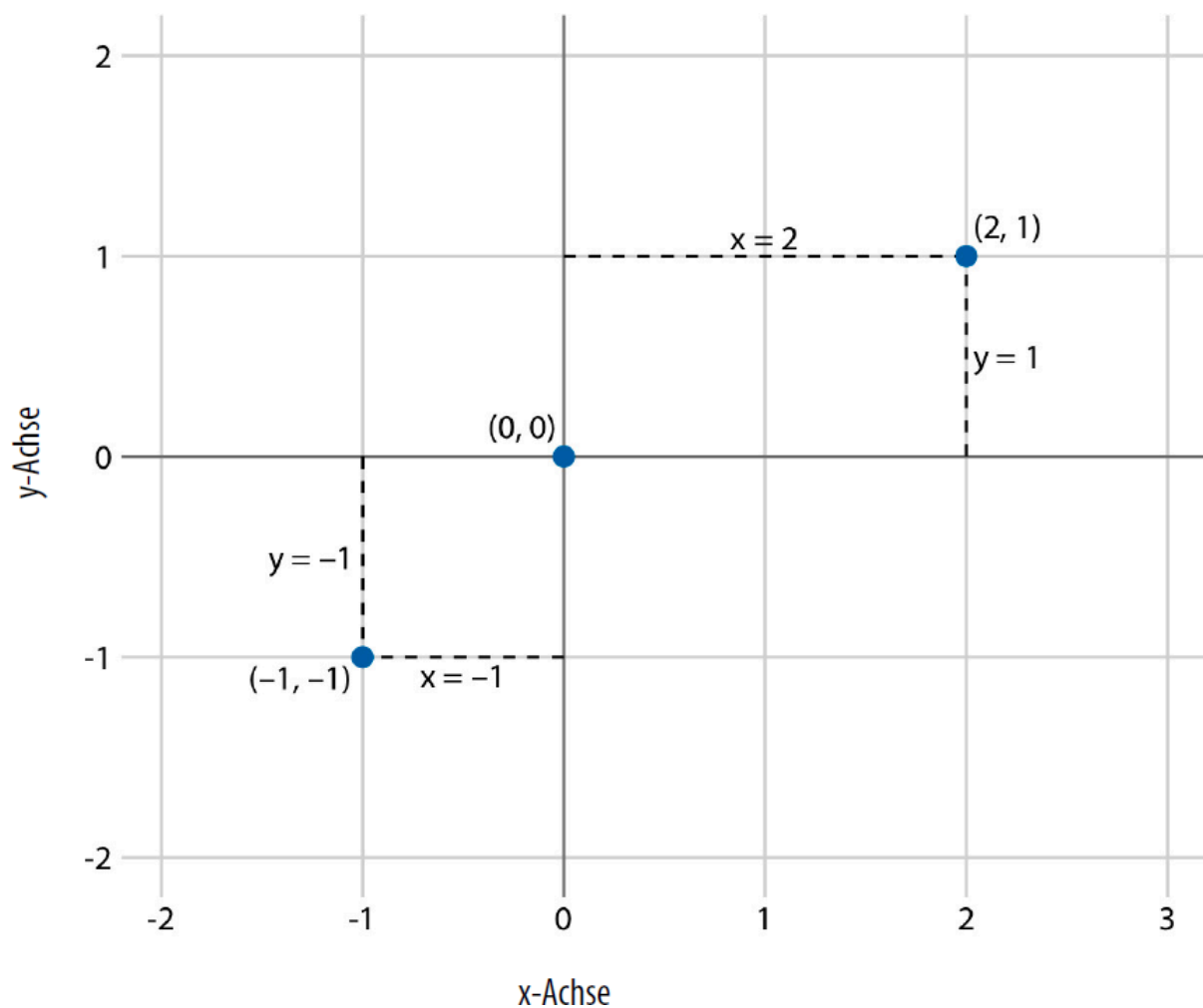


Abbildung 3-1: Kartesisches Standardkoordinatensystem. Die horizontale Achse wird herkömmlicherweise als x und die vertikale Achse als y bezeichnet. Die beiden Achsen bilden ein Gitter mit gleichem Abstand. Hierbei werden sowohl die x - als auch die y -Gitterlinie durch Einheiten von eins getrennt. Der Punkt $(2, 1)$ befindet sich zwei x -Einheiten rechts und eine y -Einheit über dem Ursprung $(0, 0)$. Der Punkt $(-1, -1)$ befindet sich eine x -Einheit links und eine y -Einheit unterhalb des Ursprungs.

Datenwerte sind jedoch normalerweise nicht nur Zahlen. Sie beinhalten Einheiten. Wenn wir zum Beispiel die Temperatur messen, können die Werte in Grad Celsius oder Fahrenheit gemessen werden. In ähnlicher Weise können wir, wenn wir die Entfernung messen, die Werte in Kilometern oder Meilen angeben, oder, wenn wir die Dauer messen, die Werte in Minuten, Stunden oder Tagen messen.

In einem kartesischen Koordinatensystem entspricht der Abstand zwischen Gitterlinien entlang einer Achse diskreten Schritten in diesen Dateneinheiten. In einer Temperaturskala haben wir beispielsweise alle 10 Grad Fahrenheit eine Gitterlinie, und in einer Entfernungsskala gibt es möglicherweise alle 5 Kilometer eine Gitterlinie.

Ein kartesisches Koordinatensystem kann zwei Achsen haben, die zwei verschiedene Einheiten darstellen. Diese Situation tritt recht häufig auf, wenn wir zwei verschiedene Variablentypen auf x und y abbilden. In Abbildung 2-3 hatten wir beispielsweise die Temperatur gegen die Tage des Jahres aufgetragen. Die y -Achse in Abbildung 2-3 gibt die Temperatur in Grad Fahrenheit mit Gitterlinien in 20° -Abständen an, und die x -Achse zeigt Monate, mit Gitterlinien am Ersten jedes dritten Monats. Wenn die beiden Achsen in unterschiedlichen Einheiten gemessen werden, können wir eine Achse relativ zu der anderen strecken oder komprimieren und eine gültige Visualisierung der Daten beibehalten (Abbildung 3-2). Welche Version vorzuziehen ist, hängt unter Umständen von der Botschaft ab, die wir übermitteln wollen.

Ein hoher und schmaler Verlauf betont die Veränderung entlang der y -Achse, und ein kurzer und breiter Verlauf bewirkt das Gegenteil. Idealerweise sollten wir ein Seitenverhältnis wählen, das sicherstellt, dass alle wichtigen Positionsunterschiede erkennbar sind.

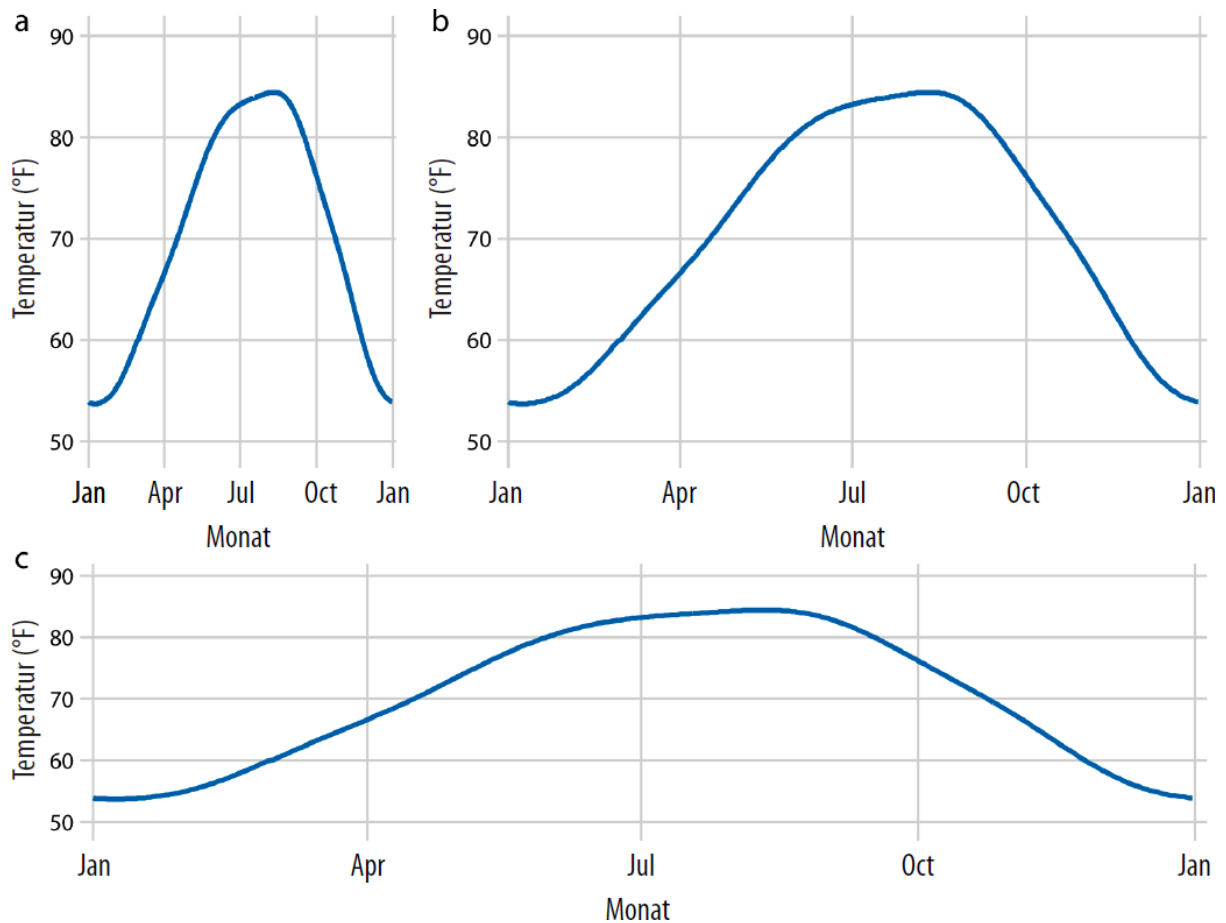


Abbildung 3-2: Tägliche Temperaturnormalen für Houston, TX. Die Temperatur wird auf der y-Achse und der Tag des Jahres auf der x-Achse abgebildet. Die Abbildungen (a), (b) und (c) zeigen den gleichen Verlauf in verschiedenen Seitenverhältnissen. Alle drei Grafiken sind gültige Darstellungen der Temperaturdaten. (Datenquelle: NOAA)

Wenn hingegen die x- und die y-Achse in den gleichen Einheiten dargestellt werden, sollten die Gitterabstände für die beiden Achsen gleich sein, sodass der gleiche Abstand entlang der x- oder y-Achse der gleichen Anzahl von Dateneinheiten entspricht. Als Beispiel können wir die Temperatur in Houston, TX, im Vergleich zur Temperatur in San Diego, CA, für jeden Tag des Jahres aufzeichnen (Abbildung 3-3a). Da auf beiden Achsen die gleiche Menge aufgetragen wird, müssen wir sicherstellen, dass die Gitterlinien perfekte Quadrate bilden, wie dies in Abbildung 3-3a der Fall ist.

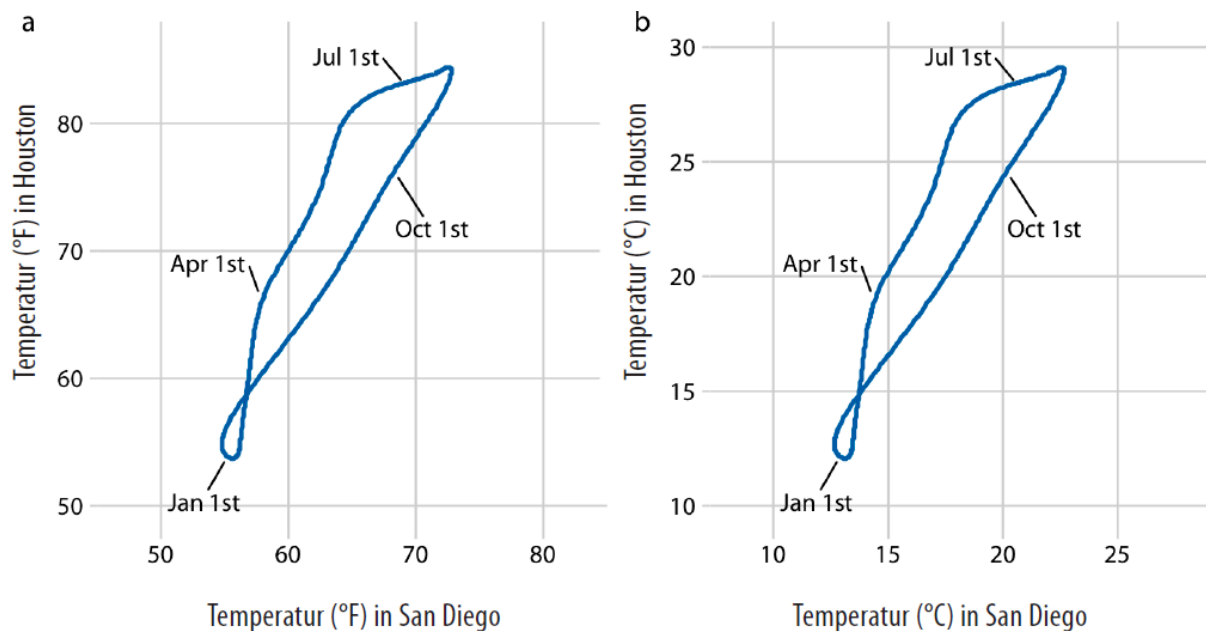


Abbildung 3-3: Tägliche Temperaturnormalen für Houston, TX, aufgetragen gegen die jeweiligen Temperaturnormalen von San Diego, CA. Die ersten Tage der Monate Januar, April, Juli und Oktober werden hervorgehoben, um einen zeitlichen Bezug herzustellen. (a) Die Temperaturen sind in Grad Fahrenheit angegeben. (b) Die Temperaturen sind in Grad Celsius angegeben. (Datenquelle: NOAA)

Sie fragen sich vielleicht, was passiert, wenn Sie die Einheiten Ihrer Daten ändern. Die Einheiten sind schließlich willkürlich und Ihre Präferenzen können sich von denen anderer unterscheiden. Eine Änderung der Einheiten ist eine lineare Transformation, bei der eine Zahl zu allen Datenwerten addiert oder von ihnen subtrahiert wird und/oder bei der alle Datenwerte mit einer anderen Zahl multipliziert werden. Glücklicherweise sind kartesische Koordinatensysteme unter solchen linearen Transformationen invariant. Daher können Sie die Einheiten Ihrer Daten ändern, ohne dass sich die resultierende Abbildung ändert, solange Sie die Achsen entsprechend ändern. Vergleichen Sie als Beispiel die Abbildungen 3-3a und 3-3b. Beide zeigen die gleichen Daten, aber in Teil (a) sind die Temperatureinheiten in Grad Fahrenheit und in Teil (b) in Grad Celsius angegeben. Obwohl sich die Rasterlinien an verschiedenen Stellen befinden und die Zahlen entlang der Achsen unterschiedlich sind, sehen beide Darstellungen genau gleich aus.

Nichtlineare Achsen

In einem kartesischen Koordinatensystem sind die Gitterlinien entlang einer Achse sowohl in Dateneinheiten als auch in der resultierenden Visualisierung gleichmäßig verteilt. Die Positionsskalen in diesen Koordinatensystemen bezeichnen wir als *linear*. Während lineare Skalen im Allgemeinen eine genaue Visualisierung der Daten liefern, gibt es Szenarien, in denen nichtlineare Maßstäbe bevorzugt werden. In einer nichtlinearen Skala entspricht ein gleichmäßiger Abstand in Dateneinheiten einem ungleichmäßigen Abstand in der Visualisierung oder umgekehrt.

Die am häufigsten verwendete nichtlineare Skala ist die *logarithmische Skala*. Logarithmische Skalen sind linear im multiplikativen Raum, sodass eine Einheit auf der Skala einer Multiplikation mit einem festen Wert entspricht. Um eine logarithmische Skala zu erstellen, müssen wir die Datenwerte logarithmieren und dabei die Zahlen, die entlang der Achsengitterlinien angezeigt werden, potenzieren. Dieser Vorgang wird in Abbildung 3-4 gezeigt, in der die Zahlen 1, 3.16, 10, 31.6 und 100 auf linearen und logarithmischen Skalen dargestellt sind. Die Zahlen 3.16 und 31.6 mögen seltsam erscheinen, aber sie wurden ausgewählt, weil sie auf einer logarithmischen Skala genau zwischen 1 und 10 und zwischen 10 und 100 liegen. Wir können dies daran beobachten, dass $10^{0.5} = \sqrt{10} \approx 3.16$ und äquivalent $3.16 \times 3.16 \approx 10$ ist. In ähnlicher Weise ist $10^{1.5} = 10 \times 10^{0.5} \approx 31.6$.

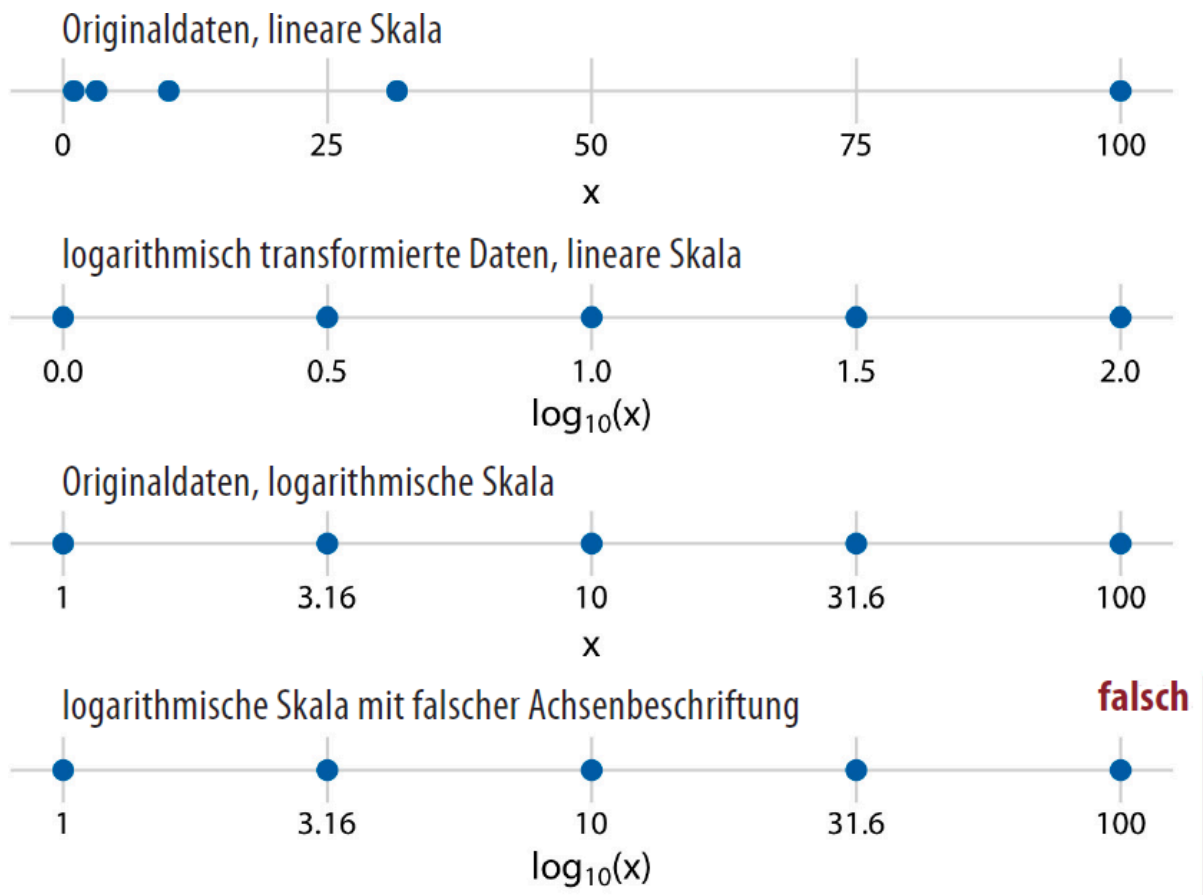


Abbildung 3-4: Beziehung zwischen linearen und logarithmischen Skalen. Die Punkte entsprechen den Datenwerten 1, 3.16, 10, 31.6 und 100, die auf einer logarithmischen Skala Zahlen mit gleichem Abstand darstellen. Wir können diese Datenpunkte auf einer linearen Skala anzeigen, wir können sie logarithmisch transformieren und sie dann auf einer linearen Skala zeigen oder wir können sie auf einer logarithmischen Skala zeigen. Wichtig ist, dass für eine korrekte Beschriftung einer logarithmischen Achse der Name der angezeigten Variablen verwendet wird und nicht der Logarithmus dieser Variablen.

Mathematisch gibt es keinen Unterschied zwischen dem Auftragen der logarithmisch transformierten Daten auf einer linearen Skala oder der Darstellung der Originaldaten auf einer logarithmischen Skala (Abbildung 3-4). Der einzige Unterschied besteht in der Beschriftung der einzelnen Achsenabschnitte und der gesamten Achse. In den meisten Fällen ist eine logarithmische Skala vorzuziehen, da die entsprechende Beschriftung dem Leser die Interpretation der Zahlen anhand der Achsenmarkierungen erleichtert. Es besteht auch weniger Verwechslungsgefahr hinsichtlich der Basis des Logarithmus. Wenn wir mit logarithmisch transformierten Daten arbeiten, könnten wir uns schnell fragen, ob

die Daten jetzt mit dem natürlichen Logarithmus oder dem Logarithmus zur Basis 10 transformiert wurden. Und es ist nicht ungewöhnlich, dass Beschriftungen mehrdeutig sind – z.B. $\log(x)$, was gar keine Basis angibt. Ich empfehle Ihnen, immer die Basis zu überprüfen, wenn Sie mit logarithmischen Daten arbeiten. Geben Sie beim Erstellen einer Abbildung mit logarithmischen Daten immer die Basis an, wenn Sie die Achse beschriften!

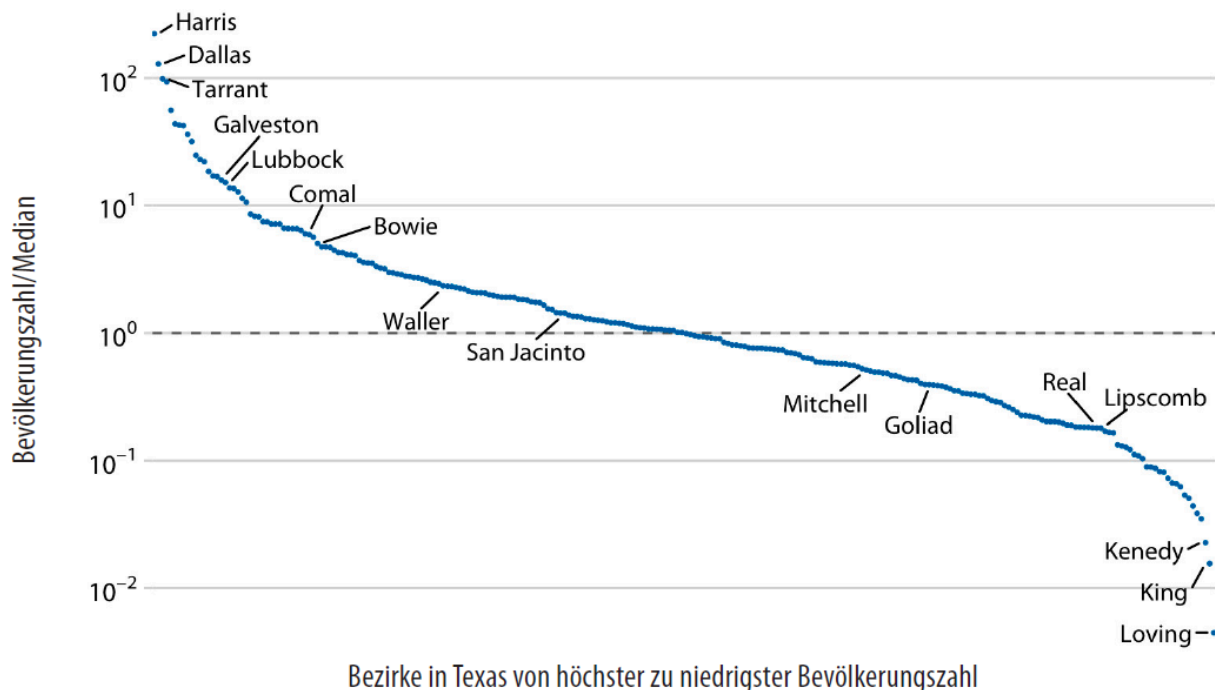


Abbildung 3-5: Bevölkerungszahl der Bezirke in Texas im Verhältnis zu ihrem Median. Ausgewählte Bezirke werden durch den Namen hervorgehoben. Die gestrichelte Linie gibt ein Verhältnis von 1 an, welches einem Bezirk mit einer Bevölkerungszahl des Medians entspricht. Die bevölkerungsreichsten Bezirke haben ungefähr 100-mal mehr, die bevölkerungsärmsten Bezirke ungefähr 100-mal weniger Einwohner als der Median-Bezirk. (Datenquelle: US Decennial Census 2010)

Da die Multiplikation auf einer logarithmischen Skala wie eine Addition auf einer linearen Skala aussieht, sind logarithmische Skalen die natürliche Wahl für alle Daten, die durch Multiplikation oder Division erstellt wurden. Insbesondere sollten das Verhältnis und die Proportionen (*engl. ratios*) mehrerer Größen im Allgemeinen auf einer logarithmischen Skala angezeigt werden. Als Beispiel habe ich die Einwohnerzahl in jedem Bezirk in Texas genommen und durch die mittlere Einwohnerzahl (der Median) in allen Bezirken in

Texas geteilt. Das resultierende Verhältnis ist eine Zahl, die größer oder kleiner als 1 sein kann. Ein Verhältnis von genau 1 impliziert, dass der entsprechende Bezirk genau die mittlere Einwohnerzahl hat. Bei der Darstellung dieser Verhältnisse auf einer logarithmischen Skala können wir sehen, dass die Bevölkerungszahlen in texanischen Gemeinden symmetrisch um den Median verteilt sind, und dass die bevölkerungsreichsten Gemeinden über 100-mal mehr Einwohner haben als der Median, während die bevölkerungsärmsten Gemeinden mehr als 100-mal weniger Einwohner haben (Abbildung 3-5).

Im Gegensatz dazu verwischt eine lineare Skala für dieselben Daten die Unterschiede zwischen einem Bezirk mit einer dem Median entsprechenden Bevölkerungszahl und einem Bezirk mit einer im Vergleich dazu viel geringeren Bevölkerungszahl (Abbildung 3-6).



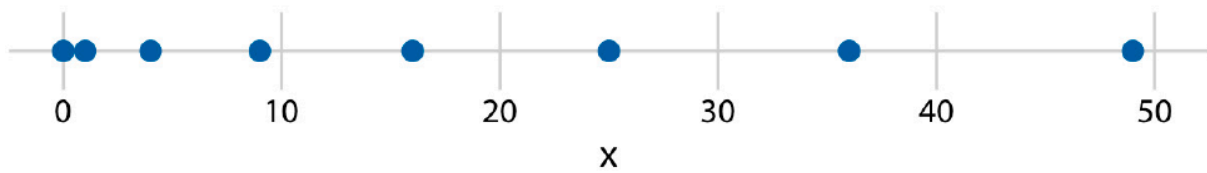
Abbildung 3-6: Bevölkerungsgrößen der Gemeinden von Texas im Verhältnis zu ihrem Median. Durch den Gebrauch einer linearen Skala zur Darstellung der relativen Anteile haben wir Werte > 1 überbetont und Verhältnisse < 1 undeutlich gemacht. In der Regel sollten solche »Ratios« nicht auf einer linearen Skala angezeigt werden. (Datenquelle: US Decennial Census 2010)

Auf einer logarithmischen Skala ist der Wert 1 der natürliche Mittelpunkt, ähnlich dem Wert 0 auf einer linearen Skala. Wir können

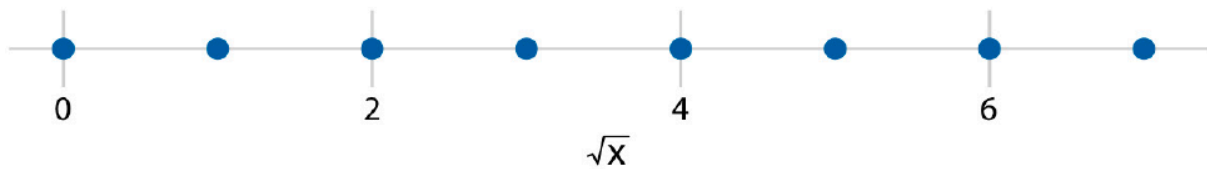
uns Werte größer als 1 als Multiplikationen und Werte kleiner als 1 als Divisionen vorstellen. Zum Beispiel können wir $10 = 1 \times 10$ und $0.1 = 1/10$ schreiben. Der Wert 0 hingegen kann niemals auf einer logarithmischen Skala erscheinen. Es liegt unendlich weit von 1 entfernt. Eine Möglichkeit, um sich das klarzumachen, ist, sich zu vergegenwärtigen, dass $\log(0) = -\infty$ ist. Alternativ können Sie auch annehmen, dass für den Übergang von 1 zu 0 eine unendliche Anzahl von Divisionen durch einen endlichen Wert (z.B. $1/10/10/10/10/10/10/10/10/10\dots = 0$) oder eine Division durch »Unendlich« erforderlich ist (d.h. $1/\infty = 0$).

Logarithmische Skalen werden häufig verwendet, wenn der Datensatz Werte mit sehr unterschiedlichen Größenordnungen enthält. Für die in den Abbildungen 3-5 und 3-6 gezeigten Bezirke in Texas hatte der bevölkerungsreichste Bezirk (Harris) 4.092.459 Einwohner bei der US-Volkszählung 2010, während der bevölkerungsärmste Bezirk (Loving) 82 Einwohner hatte. Eine logarithmische Skala wäre also auch dann angemessen gewesen, wenn wir die Bevölkerungszahl nicht durch den Median geteilt hätten, um diese in Größenverhältnisse umzuwandeln. Aber was würden wir tun, wenn es einen Bezirk mit 0 Einwohnern gäbe? Dieser Bezirk könnte nicht auf der logarithmischen Skala angezeigt werden, da sie bei minus unendlich liegen würde. In dieser Situation wird manchmal empfohlen, eine sogenannte *Quadratwurzel-Skala* zu verwenden, bei der anstelle einer logarithmischen Transformation eine Quadratwurzeltransformation verwendet wird (Abbildung 3-7). Diese komprimiert genau wie eine logarithmische Skala größere Zahlen in einen kleineren Bereich, erlaubt aber im Gegensatz zu einer logarithmischen Skala das Vorhandensein des Wertes 0.

Originaldaten, lineare Skala



Quadratwurzel-transformierte Daten, lineare Skala



Originaldaten, Quadratwurzel-Skala

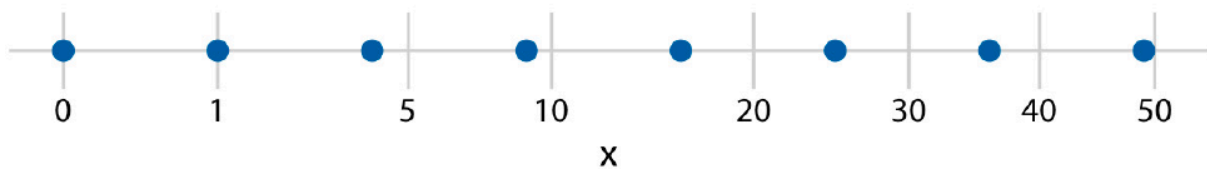


Abbildung 3-7: Beziehung zwischen linearen und Quadratwurzel-Skalen. Die Punkte entsprechen den Datenwerten 0, 1, 4, 9, 16, 25, 36 und 49, die auf einer Quadratwurzel-Skala Werte mit gleichem Abstand sind, da sie die Quadrate der ganzen Zahlen von 0 bis 7 sind. Wir können diese Datenpunkte auf einer linearen Skala anzeigen, wir können sie mit der Quadratwurzel transformieren und sie dann auf einer linearen Skala darstellen, oder wir können sie auf einer Quadratwurzel-Skala auftragen.

Ich sehe in Bezug auf Quadratwurzel-Skalen zwei Probleme:

Erstens: Während auf einer linearen Skala ein Einheitsschritt einer Addition oder Subtraktion eines konstanten Werts entspricht und er auf einer logarithmischen Skala einer Multiplikation mit einem konstanten Wert oder einer Division durch einen konstanten Wert entspricht, gibt es für eine Quadratwurzel-Skala keine solche Regel. Die Bedeutung eines Einheitsschritts auf einer Quadratwurzel-Skala hängt vom Skalenwert ab, bei dem wir beginnen.

Zweitens ist unklar, wie für gleichmäßige Intervalle Teilstriche (engl. ticks) am besten auf einer Quadratwurzel-Skala platziert werden können. Um gleichmäßig verteilte Abschnitte zu erhalten, müssten wir die Teilstriche bei den Werten der Quadratzahlen platzieren, aber Teilstriche an beispielsweise den Positionen 0, 4, 25, 49 und 81 (jede zweite Quadratzahl) wären nicht intuitiv. Alternativ könnten wir sie in linearen Intervallen platzieren (10, 20, 30 usw.), aber dies würde

entweder zu wenige Teilstriche nahe dem unteren Ende der Skala oder zu viele nahe dem oberen Ende ergeben. In Abbildung 3-7 habe ich die Achsenmarkierungen an den Positionen 0, 1, 5, 10, 20, 30, 40 und 50 auf der Quadratwurzel-Skala platziert. Diese Werte sind willkürlich, decken jedoch den Datenbereich angemessen ab.

Trotz dieser Probleme mit Quadratwurzel-Skalen sind sie gültige Positionsskalen, und ich schließe die Möglichkeit, dass sie geeignete Anwendungen haben, nicht aus. Genauso wie eine logarithmische Skala die natürliche Skala für Proportionen ist, könnte man argumentieren, dass die Quadratwurzel die natürliche Skala für Daten ist, die in Quadratzahlen angegeben werden. Ein Szenario, in dem Daten von Natur aus Quadratzahlen sind, finden wir bei der Darstellung geografischer Regionen. Wenn wir die Größe geografischer Regionen auf einer Quadratwurzel-Skala darstellen, heben wir die räumliche Ausdehnung der Regionen von Ost nach West oder von Nord nach Süd hervor. Diese Ausdehnung könnte z.B. relevant sein, wenn wir uns fragen, wie lange es dauern könnte, durch eine bestimmte Region zu fahren. Abbildung 3-8 zeigt die Gebiete der Bundesstaaten im Nordosten der USA auf einer linearen und auf einer Quadratwurzel-Skala. Obwohl die Flächen dieser Bundesstaaten sehr unterschiedlich sind (Abbildung 3-8a), wird die relative Zeit, die benötigt wird, um durch die jeweiligen Gebiete zu fahren, durch die Abbildung auf der Quadratwurzel-Skala (Abbildung 3-8b) genauer dargestellt als durch die Abbildung auf der linearen Skala (Abbildung 3-8a).

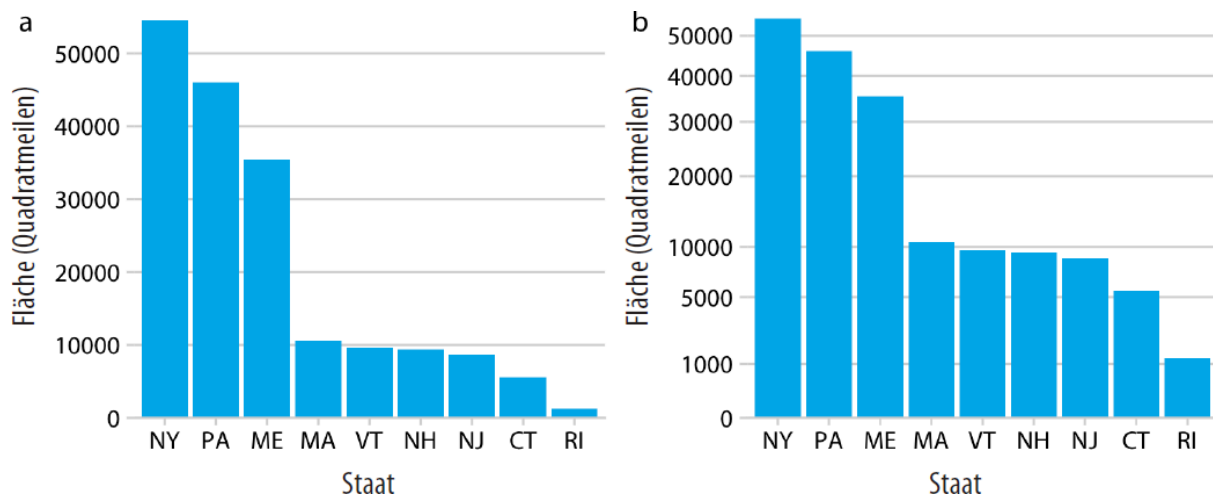


Abbildung 3-8: Gebiete in nordöstlichen US-Bundesstaaten. (a) Flächengröße auf einer linearen Skala. (b) Flächengröße auf einer Quadratwurzel-Skala. (Datenquelle: Google)

Koordinatensysteme mit gekrümmten Achsen

Alle Koordinatensysteme, auf die wir bisher gestoßen sind, haben zwei gerade Achsen verwendet, die im rechten Winkel zueinander stehen, auch wenn die Achsenwerte das Ergebnis einer (nichtlinearen) Abbildung waren. Es gibt jedoch auch andere Koordinatensysteme, bei denen die Achsen selbst gekrümmt sind. Insbesondere in einem *Polarkoordinatensystem* geben wir Positionen über einen Winkel und einen radialen Abstand zum Ursprung an. Daher ist die Winkelachse kreisförmig (Abbildung 3-9).

Polarkoordinaten können für periodische Daten nützlich sein, sodass Datenwerte an einem Ende der Skala logisch mit Datenwerten am anderen Ende verknüpft werden können. Betrachten Sie beispielsweise die Tage in einem Jahr. Der 31. Dezember ist der letzte Tag des Jahres, aber er ist auch ein Tag vor dem ersten Tag des Jahres. Wenn wir zeigen wollen, wie sich ein Wert im Laufe des Jahres ändert, kann es angebracht sein, Polarkoordinaten zu verwenden, wobei die Winkelkoordinate jeden Tag angibt. Wenden wir nun dieses Konzept auf die Temperaturnormalen in Abbildung 2-3 an. Da es sich bei den Temperaturnormalen um Durchschnittstemperaturen handelt, die nicht an ein bestimmtes Jahr gebunden sind, kann der 31. Dezember als 366 Tage später als

der 1. Januar (Temperaturnormalen umfassen den 29. Februar) und auch als 1 Tag früher angesehen werden. Indem wir die Temperaturnormalen in ein Polarkoordinatensystem zeichnen, betonen wir ihre zyklische Eigenschaft (Abbildung 3-10). Im Vergleich zu Abbildung 2-3 zeigt die Polarversion, wie ähnlich die Temperaturen in Death Valley, Houston und San Diego vom späten Herbst bis zum frühen Frühling sind. Im kartesischen Koordinatensystem ist diese Aussage nicht so leicht erkennbar, weil die Temperaturwerte Ende Dezember und Anfang Januar in entgegengesetzten Teilen der Abbildung dargestellt sind und daher keine visuelle Einheit bilden.

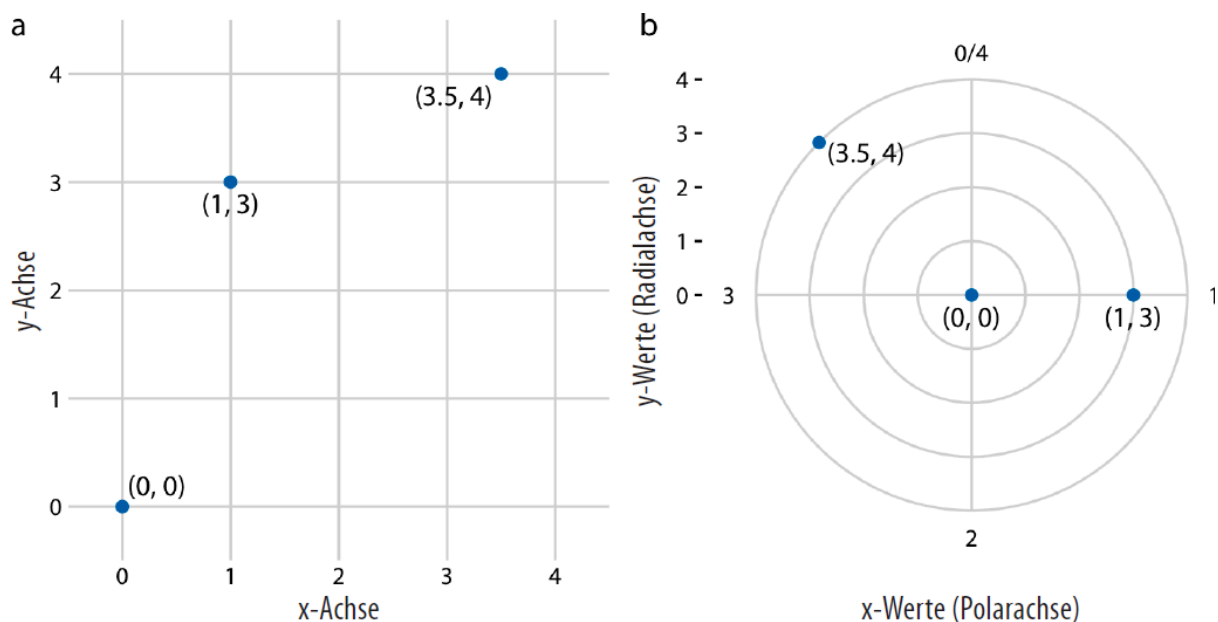


Abbildung 3-9: Beziehung zwischen kartesischen und Polarkoordinaten. (a) Drei Datenpunkte in einem kartesischen Koordinatensystem. (b) Dieselben drei Datenpunkte in einem Polar-Koordinatensystem. Wir haben die x-Koordinaten von Abbildung (a) als Winkelkoordinaten und die y-Koordinaten von Abbildung (a) als Radialkoordinaten verwendet. Die Kreisachse verläuft in diesem Beispiel von 0 bis 4, und daher sind $x = 0$ und $x = 4$ die gleichen Stellen in diesem Koordinatensystem.

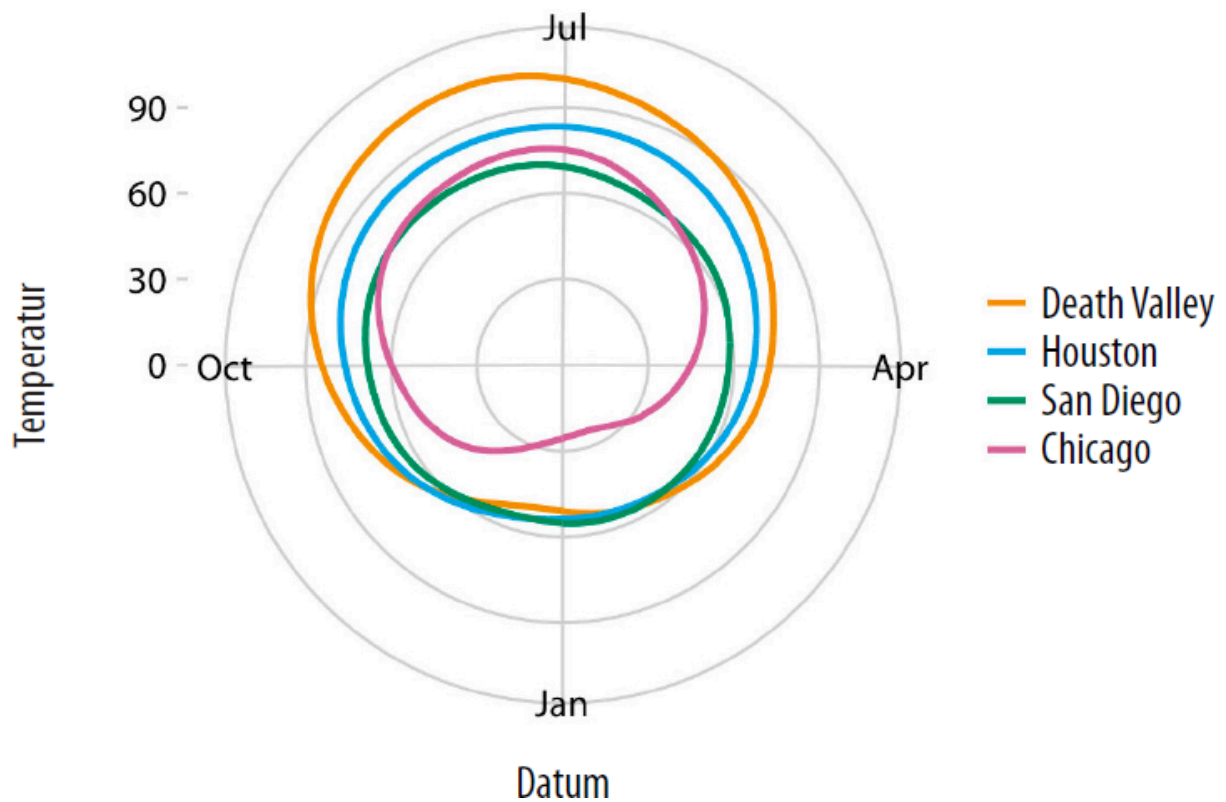
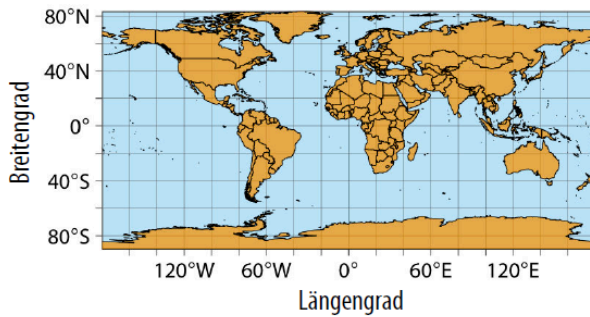
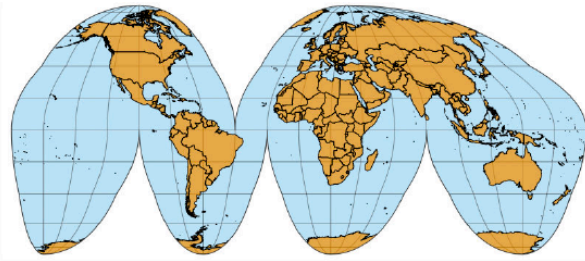


Abbildung 3-10: Tägliche Temperaturnormalen für vier ausgewählte Orte in den USA, angezeigt in Polarkoordinaten. Der radiale Abstand vom Mittelpunkt gibt die Tagestemperatur in Fahrenheit an, und die Tage des Jahres werden beginnend mit dem 1. Januar auf der 06:00-Position gegen den Uhrzeigersinn angeordnet. (Datenquelle: NOAA)

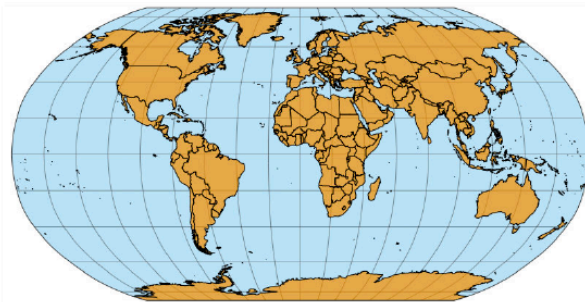
kartesisches Längen- und Breitengradsystem



Goode-Homolosine



Robinson



Winkel-Tripel

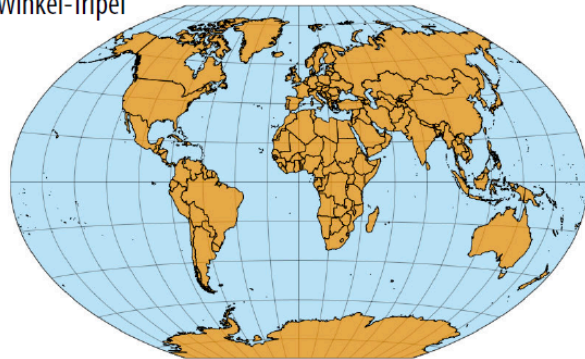


Abbildung 3-11: Die Weltkarte, dargestellt in vier verschiedenen Projektionen. Das kartesische Längen- und Breitengradsystem bildet die Längen- und Breitengrade jedes Ortes auf ein reguläres kartesisches Koordinatensystem ab. Diese Abbildung führt sowohl bei Flächen als auch bei Winkeln zu erheblichen Verzerrungen relativ zu ihren wahren Werten auf dem 3D-Globus. Die Goode-Homolosine-Projektion repräsentiert perfekt die wahren Oberflächenbereiche, wobei einige Landmassen, insbesondere Grönland und die Antarktis, in separate Teile geteilt werden müssen. Sowohl die Robinson-Projektion als auch die Winkel-Tripel-Projektion stellen ein Gleichgewicht zwischen Winkel- und Flächenverzerrungen her und werden üblicherweise für Karten des gesamten Globus verwendet.

Eine zweite Form, bei der wir auf gekrümmte Achsen stoßen, betrifft Geodaten, d.h. Karten. Standorte auf dem Globus werden durch ihren Längen- und Breitengrad angegeben. Da die Erde jedoch eine Kugel ist, ist es irreführend und nicht empfehlenswert, Breiten- und Längengrade als kartesische Achsen zu zeichnen (Abbildung 3-11). Stattdessen verwenden wir verschiedene Arten von nichtlinearen Projektionen, die versuchen, Artefakte zu minimieren, und auf unterschiedliche Weise ein Gleichgewicht herzustellen, und zwar zwischen Flächen oder Winkeln, die in Bezug auf die tatsächlichen Linien auf dem Globus ihre Form und Größe beibehalten (Abbildung 3-11).

2D, Projektion von 3D-Objekten auf 277, 280

2D-Bins 37

2D-Fehlerbalken 38

2D-Histogramm 202, 204

3D-Balken, Probleme mit 278

3D-Diagramm 277, 285

- 3D-Positionsskalen vermeiden 279, 285

- angemessene Verwendung von 285

- Probleme mit 3D-Balken 278, 284

- unnötiges 3D vermeiden 277, 279

- Verzerrung von Objekten bei der Projektion auf 2D 277

A

Abbildungen, manuelle Schritte bei der Vorbereitung von 299

absolute Zahlen vs. prozentuale Anteile 91

Achsen 16, 24

- Achsenbereich bei Small Multiples 231

- Koordinatensysteme mit gekrümmten Achsen 21, 24

- nichtlinear 16, 21

- Titel 244, 246

 - Beispiel einer Abbildung mit 244

 - übertreiben bei 246

 - weglassen von 245

- Visualisierung mehrerer Verteilungen auf der horizontalen Achse 82, 83

- Visualisierung mehrerer Verteilungen auf der vertikalen Achse 75, 81
 - weglassen in Abbildungen 246
- Achsenbeschriftung große, Verwendung von 265, 268
- Achsenbeschriftung, *siehe auch* Achsen, Titel 244
- Aesthetics 7, 298
 - kontinuierliche oder diskrete Daten darstellen 8
 - üblich bei Datenvisualisierungen 7
- Aesthetics, Merkmale der
 - Codierung einzelner Variablen bei mehreren Merkmalen der Aesthetics, Legenden und 228
- Akzentfarbskala 30
- Albers-Projektion 152
- Alterspyramide 63
- antikorrelierende Variablen 109
- Antwortvariable 75
- A-Posteriori-Ansatz 177
- A-Posteriori-Verteilungen 176
 - Visualisierung von 179
- approximierte Linien 39
- A-Priori-Ansatz 177 Äquator 147
- ausfüllen 269
- Automatisierung
 - programmgesteuert generierte Abbildungen 297

B

Balken

- auf linearen Achsen 190
- in Visualisierungen entlang linearer Achsen 192
- Visualisierung quantitativer Werte 33
- Visualisierung von Proportionen 35

Balkendiagramm 41, 42

- 3D, Probleme mit 282
- auf logarithmischen Achsen 194
 - Visualisierung von Mengen 194
 - Visualisierung von Verhältnissen 196
- Einschränkungen gestapelter Balkendiagramme 99
- fehlerhafte Ansätze zur Visualisierung verschachtelter Proportionen 96
- Hintergrundgitter in 260
- Linienzeichnungen in 269
- mit Fehlerbalken 175
- ohne explizite Achsen 246
- Visualisierung quantitativer Mengen 45
- Visualisierung quantitativer Werte 41
 - gestapelte Balken 47
 - gruppierte Balken 45
 - horizontale Balken 43
 - Probleme mit langen Beschriftungen bei vertikalen Balken 43
 - Reihenfolge von Balken 43
 - vertikale Balken 42
- Visualisierung von Proportionen 86
 - ein Fall für gestapelte Balken 90
 - ein Fall für nebeneinander angeordnete Balken 88, 90
- vs. Kreisdiagramm 197
- Bandbreite 57
- bayessche Statistik 170
 - Evaluierung von Unsicherheiten 176
 - prinzipieller Vorteil 177
- Beschriftung 243
 - Abbildungstitel als erstes Element 243
 - Achsen 244, 265, 268
 - einzelner Module 234

- Gebrauch von Direktbeschriftung statt Legenden 225
- Gebrauch von direkter Beschriftung anstatt Farbe bei mehr als 8 Kategorien 209
 - in die Kernabbildung integrieren 242
 - Platzierung in Abbildungen und Tabellen 249
 - vertikale Balken in Balkendiagrammen 43
- Bildbearbeitungs- oder Illustrationsprogramm 299
- Bilddateiformate 289, 294
 - Bitmap- und Vektorgrafik 289, 291
 - Konvertierung zwischen Formaten 293
 - verlustfreie und verlustbehaftete Komprimierung von Bitmapgrafiken 291, 292
- Bildunterschriften 241
- Binning (Kategorisierung) von Daten
 - Bins (Kategorien) in Histogrammen 55
 - in Sechsecke 204
- Bitmapgrafik 289
 - JPEG und PNG-Formate 289
 - speichern als Vektorgrafik 293
 - verlustfreie und verlustbehaftete Komprimierung von 291, 292
- Blasendiagramm 37
 - Visualisierung von Korrelationen zwischen quantitativen Variablen 106
- Blau-Gelb-Farbsehschwäche 214, 220
- Box-Plot 34
 - Tönung 274
 - Visualisierung mehrerer Verteilungen 77, 78
- C**
- Choroplethen 27
 - Choroplethenkarte 147, 157, 160
 - Gebrauch in Karten 38
- Color Vision Deficiency (CVD) 215

ColorBrewer-Farbskala 25, 27

CVD, *siehe auch* Color Vision Deficiency 215

D

Darstellungsmerkmale, Anwendung auf Karten mit Geodaten 156

Daten

in Abbildungen, *siehe auch* Daten und Kontext in Einklang bringen 251

Kategorien von 8

kategorisierbare 9

Gebrauch von Farben 25, 209

kontinuierliche 8

zu viele Daten in Abbildungen 308

Daten und Kontext in Einklang bringen 251, 263

Abbildungen mit gepaarten Daten 261, 263

Hintergrundgitter 256, 261

richtige Menge an Kontext bereitstellen 251, 256

Abbildung mit zu viel Farbflächen von Nicht-Daten 252, 253

Datenexploration vs. Datenpräsentation 298, 300

Daten-Farbflächen-Verhältnis 251

Datenquellen 242

Datentransformation (bei 3D-Visualisierungen) 280

Datentypen, Ausprägungen von Merkmalen 8

Datenvisualisierungsprogramme, Glättungsfunktionen 136

Datenvisualisierungssoftware auswählen 295, 302

Datenexploration vs. Datenpräsentation 298, 300

Reproduzierbarkeit und Wiederholbarkeit 296

Reproduzierbarkeit und Wiederverwendbarkeit 297

Trennung von Inhalt und Design 300, 302

Datenwerte

farbige Darstellung 38

Nutzung von Farbe zur Darstellung von 27, 29

- Datum und Uhrzeit 9
- Design und Inhalt, Trennung von 302
- Deterministic Construal Errors 172
- Deuteranomalie/Deuteranopie 214, 220
- Dichtediagramm
 - direkt beschriftete Kurven 226
 - Gebrauch als Ersatz für Legenden in Streudiagrammen 226
 - Linienzeichnungen in 271
 - Visualisierung einer einzelnen Verteilung 57
 - Dichtediagramm oder Histogramm auswählen 59
 - Kerndichte und Bandbreite 58
 - Visualisierung von Verteilungen 34, 62, 64
 - abgebildete Dichten 93
 - gestapelte Dichten 91
- Dichten, kumulative 34
- Dimensionsreduktion 112, 115
 - Gebrauch mit höherdimensionalen Zeitreihen in verbundenen Streudiagrammen 127
- direkte Beschriftung
 - als Alternative zu Legenden 225
 - als Alternative zum Gebrauch von Farbe 209
- direkte Flächendarstellung 197, 198
 - Kreisdiagramm 197
- diskrete Daten 8
- diskrete Elemente oder Gruppen farbig unterscheiden 25, 27
- diskrete Ergebnisvisualisierung 165
- divergente Farbskala 28
 - Beispiel 29
- divergierende Farbskala 215
- Dosiswirkungskurven, Visualisierung mit mehreren Zeitreihen 122, 124

E

Einfachheit, Nachteile von 313

einprägsame Abbildung, generieren 313, 315

empirisch kumulative Häufigkeitsverteilung (ECDF) 65, 68

- normalisierter Rang 67

- stark verzerrte Verteilung 68, 71

- steigende und fallende ECDFs 66

Environmental Systems Research Institute (ESRI) 149

Erkundung von Daten *siehe auch* Datenexploration vs. Datenpräsentation 298

European Petroleum Survey Group (EPSG) 149

Eye-Diagramm 179

Eye-Plot 39

F

Facettieren 229

falsche Abbildungen 3

Farbe 7

- als Unterscheidungsmerkmal 25, 27

- als Werkzeug zur Hervorhebung 30, 31

- Differenzierung einer kategorisierbaren Variable nach Balkenfarbe 47

- farbige Abbildung von Datenwerten in Heatmaps 51

- farbige Visualisierung von Geodaten 38

- grundlegende Nutzung in der Datenvisualisierung 25

- in fotografischen Bildern 292

- Intensität 211

- Linienzeichnungen und 269

- redundante Codierung 219

- typische Fallstricke beim Gebrauch von 209, 217

 - farbige Kennzeichnung von zu vielen oder irrelevanten Informationen 209, 212

 - unzureichendes Design bei Farbsehschwäche 214, 217

- Verwendung nichtmonotoner Farbskalen 213, 214
- Verwendung von grauem Hintergrund in Abbildungen 257
 - zur Darstellung von Datenwerten 27, 29
- Farbflächen, Prinzip der proportionalen 189, 198, 251
 - Definition von Farbfläche 189
 - direkte Flächendarstellung 197, 198
 - Visualisierung entlang linearer Achsen 190, 194
 - Visualisierung entlang logarithmischer Achsen 194, 196
- Farbsehschwäche 215, 217, 219, 228
 - Color Vision Deficiency (CVD) 214
 - Deuteranomalie/Deuteranopie 214, 220
 - Protanomalie/Protanopie 214, 220
 - Tritanomalie/Tritanopie 214, 220
- Farbskala 25, 31
 - Akzent- 30
 - divergent 28
 - qualitativ 25
 - sequenziell 27
- Färbung
 - in der Visualisierung von Zeitreihen entlang linearer Achsen 191
 - in Visualisierungen entlang linearer Achsen 192
- Fehlerbalken 38, 76, 163
 - abgestufte 171
 - in Ridgeline-Diagrammen 179
 - Schwächen von 170
 - Variationen von 174
 - Vorteile gegenüber komplexeren Unsicherheitsanzeigen 175
- Fehlerbalken, abgestufte 38, 171
 - Vor- und Nachteile 174
- flächengleiche Projektion 149

Albers-Projektion 152

Form 7

frequentistische Statistik 170

G

Gaußsche Glockenkurve 57

Gaußscher Kernel 57

Generalized Additive Models (GAMs) 136

Geodaten 147, 162

 Darstellung auf Karten 38

 Visualisierung anhand von Choroplethenkarten 157, 160

 Visualisierung anhand von Kartogrammen 160, 162

 Visualisierung mit Kartenprojektionen 147, 153

geografische Regionen

 Quadratwurzel-Skala für 21

 Unterscheidung von Datenwerten 27

gepaarte Daten 115, 117

 Visualisierung mit einem Steigungsdiagramm 115

 Visualisierung mit einem Streudiagramm 115

Geschichten

 Funktion beim Argumentieren und Erinnern 303

 verstehen, was eine Geschichte ist 304, 307

Geschichten und Abbildungen 303, 305, 319

 Abbildungen für die Generäle 307, 310

 einprägsame Abbildung generieren 313, 315

 konsistent bleiben, ohne sich zu wiederholen 315, 319

 Pfad zu komplexeren Abbildungen 311, 313

 verstehen, was eine Geschichte ist 304, 307

gestapelte Balken 47

 Visualisierung mehrerer Proportionen 35

gestapelte Balkendiagramme

- Einschränkungen 99
- Visualisierung von Proportionen 86, 90
- gestapelte Dichten
 - Visualisierung von Proportionen 91
 - Visualisierung, wenn sich Anteile entlang einer kontinuierlichen Variablen ändern 36
- gestapelte Histogramme 34
 - Visualisierung mehrerer Verteilungen 60
- ggplot2
 - facet grid-Funktion 229
 - Hintergrundgitter 256
 - Hue-Skala 25
 - Trennung von Inhalt und Design durch Themes 301
- Gitter, Hintergrund *siehe auch* Hintergrundgitter 256
- Glättung 131, 136
 - anhand von Datenvisualisierungsprogrammen 136
 - Gebrauch von Spline-Modellen 135
 - Generalized Additive Models (GAMs) 136
 - Locally Estimated Scatterplot Smoothing (LOESS) 134
 - Tankkapazitäten, dargestellt mit einem expliziten Analysemodell 137
 - unvorhersehbares Verhalten von Allzweck-Glättungsformen 137
 - Vorsicht bei der Interpretation der Ergebnisse von geglätteten Funktionen 136
- Goode-Homolosine-Projektion 150
- graduelles Konfidenzintervall 39
- Größe 7
 - der farbigen grafischen Elemente 217
- Grundgesamtheit (Population) 168
- gruppierte Balken 45
- Gruppierungsvariable 75

H

- Half-Eye-Diagramm 179
- Half-Eye-Plot 39
- handgezeichnete oder manuell nachbearbeitete Abbildungen 299
- Handlungsbogen 304
- Hässliche, schlechte und falsche Abbildungen 3
- Häufigkeiten, Framing-Wahrscheinlichkeiten als 163, 168
- Hauptkomponenten 112
- Hauptkomponentenanalyse (PCA) 112
 - Visualisierung höherdimensionaler Zeitreihen im Hauptkomponentenraum 127
- Heatmap 34
 - Kartogramm 161
 - Visualisierung quantitativer Mengen 51, 52
- Hex-Bins 37
- Hintergrund
 - Box vom Hintergrund trennen 274
 - Empfehlungen zur Verwendung von 263
 - grauen Hintergrund und Gitterlinien entfernen 257
 - Verwendung von grauem Hintergrund und weißen Gitterlinien 256
- Hintergrundgitter 253, 256, 261
 - diagonale Linie anstelle eines Gitters in Streudiagrammen verbundener Daten 261
 - Empfehlungen zur Verwendung von 263
 - entlang beider Achsen, bei Streudiagrammen 261
 - für Abbildungen zur Änderung der y-Werte 257
 - Gitterlinien senkrecht zur gewünschten Schlüsselvariablen 260
 - grauen Hintergrund und Gitterlinien entfernen 257
 - horizontale Referenzlinie 257
 - in Balkendiagrammen 260
 - in ggplot2-Software 256

in Streudiagrammen verbundener Daten, Verwendung einer diagonalen Linie mit 263

Verwendung von grauem Hintergrund und weißen Gitterlinien 256

Histogramm 34, 166

2D 202, 204

gestapeltes 34

in Ridgeline-Diagrammen 83

Linienzeichnungen in 269

Visualisierung einer einzelnen Verteilung 56, 57

Dichtediagramm oder Histogramm auswählen 59

Visualisierung mehrerer Verteilungen mit gestapelten Histogrammen 60, 62

HOPs, *siehe auch* hypothetische Ergebnisplots 183

hypothetisches Ergebnisdiagramm (hypothetical outcome plot, HOP) 183, 185

I

Inhalt und Design, Trennung von 300, 302

Isotyp-Diagramm 315

J

Jittering (Zittern) 80

Gebrauch bei überlappenden Punkten 200

JPEG-Datei 292

speichern als/von andere(n) Formate(n) 294

K

Karten 38, 147

aus mehreren Schichten mit unterschiedlichen Arten von Informationen 154

Choropleth 157, 160

Kartogramme 162

Projektionen 147, 153

Kartesische Koordinaten 13, 16

Kartogramm 38, 147, 160, 162

Heatmap 161

- mit deformierten Formen 160
- Kategorisierung (Binning) von Daten
 - Kategorien (Bins) in Histogrammen 55, 57
 - Breite auswählen 56
- Kerndichteschätzung 57, 166
 - Tücken 59
- Knoten (von Splines) 135
- Komprimierung, verlustfrei und verlustbehaftet, von Bitmapgrafiken 291, 292
 - verlustbehaftete Komprimierung in JPEGs 294
- Konfidenzband 180
 - abgestuftes 181, 182
- Konfidenzintervall 39, 163, 177
- Konfidenzniveaus 181
- Konfidenzstreifen, als Alternative zu Fehlerbalken 174
- Konsistenz, erreichen, ohne sich zu wiederholen 315, 319
- Kontext in Abbildungen, *siehe auch* Daten und Kontext in Einklang bringen 251
- Konturlinien 37, 204, 208
 - in eigenem Diagrammfenster zeichnen 207
 - mehrere Sätze in verschiedenen Farben zeichnen 207
- Koordinaten 7
- Koordinatensysteme und -achsen 13, 16, 24
 - kartesisch 13
 - Koordinatensysteme mit gekrümmten Achsen 21, 24
 - nichtlineare Achsen 16, 21
- Korrelationen, Visualisierung von 105, 117
 - Dimensionsreduktion 112, 115
 - gepaarte Daten 115, 117
 - Korrelogramm 109, 111
 - Streudiagramm 105, 109
- Korrelationskoeffizient

- Berechnung von 109
- Darstellung als Korrelogramm 37
- korrelierende oder antikorrelierende Variablen 109
- Korrelogramm 37
 - Einschränkungen 110
 - Visualisierung von Korrelationen zwischen quantitativen Variablen 109, 111
- Kredibilitätsintervall 176
- Kreisdiagramm
 - Balkendiagramm vs. 197
 - direkte Flächendarstellung 197
 - fehlerhafte Ansätze zur Visualisierung verschachtelter Proportionen 96
 - ohne explizite Achsen 246
 - unnötige Verwendung von 3D in 277
 - verschachtelte Kreisdiagramme, Visualisierung verschachtelter Proportionen 100, 102
 - Visualisierung von Proportionen 35, 85, 88
 - Balkendiagramme anstelle von Kreisdiagrammen nutzen 87
 - Hervorhebung einfacher Anteile 87
- Kurvenanpassungen, Visualisierung von Ungenauigkeiten bei 180, 183

L

- Landkarten 27
- Legende
 - bei modularen Abbildungen 237
 - Beschriftung zu klein 265
 - Gestaltung mit redundanter Codierung 219, 225
 - Gestaltung von Abbildungen ohne 225, 228
 - Titel 244, 246
 - Beispiel einer Abbildung mit 244
 - übertreiben bei 246
 - weglassen von 245

- lineare Achse, Visualisierung entlang einer 190, 194
 - Balken 190
 - Zeitreihe 191
- Lineare Skala 16
 - Proportionen auf 19
 - Vergleich zur logarithmischen Skala 20
- Linear-log-Abbildung 140
- Linienbreite 7
- Liniendiagramm 37
 - anderer Gebrauch als zur Visualisierung einer Zeitreihe 123
 - Visualisierung einer Zeitreihe 120
 - Visualisierung einer Zeitreihe mit zwei oder mehr Antwortvariablen anhand zweier Abbildungen 125
 - Visualisierung mehrerer Zeitreihen 122
 - Visualisierung von Dosiswirkungskurven 124
- Linienzeichnungen vermeiden 269, 274
- LOESS (Locally Estimated Scatterplot Smoothing) 134
 - saisonale Zeitreihenzerlegung nach 144
- log (logarithmische) Skala 138
- logarithmische Achse, Visualisierung entlang einer 194, 196
 - Balkendiagramme auf logarithmischen Achsen 194
- logarithmische Normalverteilung 70, 73
- logarithmische Skala 17
 - Datensatzwerte mit sehr unterschiedlichen Größenordnungen 19
- log-lineare Abbildung 140
- Log-log-Abbildung 140

M

- Mengen, Visualisierung von
 - auf logarithmischen Achsen, Probleme bei 194
- menschliche Wahrnehmung

- beurteilt Entfernungen besser als Flächen 197
- Farbsehschwäche 214
- verlustbehaftete Komprimierung und 292
- visuell komplexere und einzigartigere Abbildungen sind einprägsamer 313
- Mercator-Projektion 149
- Merkmale 8, 9, 11
 - Ausprägungen eines 8, 11
 - geordnete 9
- Mittelwert 168
- modulare Abbildung 234, 238
 - Beschriftung einzelner Module 234
 - wie einzelne Module zusammenpassen 236
- Mosaikdiagramm 36
 - Einschränkungen 99
 - ohne explizite Achsen 246
 - Visualisierung verschachtelter Proportionen 97
 - vs. Treemap 98
- Multipanel-Diagramm 229, 238
 - Abbildungsmodule mit Rahmen darstellen 253
 - modulare Abbildung 234, 238
 - Small Multiples 229, 234
 - Verwendung von Small Multiples anstelle von 3D-Diagrammen 284
- Multipanel-Diagramme, Konturlinien in eigenem Diagrammfenster zeichnen 207

N

- Nachbearbeitung von ersten Entwurfsabbildungen 299
- nichtlineare Beziehung von Daten, funktionale Form für 138
- nicht-monotone Farbskala 213
- Nullhypothese 177
- Numerische Daten 9

O

Okabe-Ito-Skala 25, 216

Overplotting 199

P

Parallel Sets 36

Parallel-Sets-Diagramm

 Visualisierung verschachtelter Proportionen 102, 103

Parameter 168

partielle Transparenz, Gebrauch bei überlappenden Punkten 200

PDF-Datei

 Renderzeit bei großen Dateien 291

 speichern als/von andere(n) Formate(n) 294

Phasenporträt 125

PNG-Datei 292

 gegenüber JPEG zu bevorzugen 292

 speichern als/von andere(n) Formate(n) 294

Polarkoordinaten 22

Pole 147

Population (Grundgesamtheit) 168

Position 7

Positionsskala 11

 3D, angemessene Verwendung von 285

 3D, vermeiden von 279, 285

Potenzgesetzverteilung 68

Präsentation von Daten *siehe auch* Datenexploration vs. Datenpräsentation 299

Prinzip proportionaler Farbflächen 189

programmgesteuert generierte Abbildungen *siehe auch* Automatisierung 297

Projektionen (Karten) 147, 153

 Albers-Projektion, flächengleich 152

 Goode-Homolosine 150

 konform und flächengleich 149

- Mercator-Projektion und Varianten 149
- Verwaltung von Registern durch Normungsgremien 149
- Proportionen, Visualisierung von 85, 94
 - ein Fall für gestapelte Balken und gestapelte Dichten 90, 91
 - ein Fall für Kreisdiagramme 88
 - ein Fall für nebeneinander angeordnete Balken 88, 90
 - separate Visualisierung als Teile der Summe 92
 - verschachtelte Proportionen 95, 103
 - fehlerhafte Ansätze 95, 97
 - Visualisierung mit Mosaikdiagrammen und Treemaps 97, 100
 - Visualisierung mit Parallel-Sets-Diagrammen 102, 103
 - Visualisierung mit verschachtelten Kreisdiagrammen 100, 102
 - Visualisierung von 35
- Protanomalie/Protanopie 214, 220
- Proteindaten-Visualisierung 285
- prozentuale Anteile vs. absolute Zahlen 91
- Punktendiagramm
 - auf logarithmischen Achsen 195
 - Punkte im Vergleich zum Text zu klein 267
- Quantil 39
- Visualisierung einer Zeitreihe 119
 - Verbindung einzelner Punkte mit Linien 120
- Visualisierung quantitativer Mengen 49, 51
- Visualisierung quantitativer Werte 33
- Punkte
 - gefüllte vs. nicht gefüllte 273
 - Umgang mit überlappenden Punkten 199
- Punktschätzung 168, 180
- Q**
- Q-Q-Diagramme 34

Q-Q-Plot

Visualisierung von Verteilungen 71

Quadratwurzel-Skala 20

geeignete Anwendungen 21

Probleme 20

qualitative Daten 8

qualitative Farbskala 25, 215

Beispiel zum Gebrauch von 25

Okabe-Ito-Skala 216

Quantil-Punktdiagramm 39, 167

Quantil-Quantil-Diagramm *siehe auch* Q-Q-Plot

quantitative Daten 8

Visualisierung von Korrelationen zwischen quantitativen Variablen 105

quantitative Werte, Visualisierung von 33, 41, 52

Balkendiagramm 41, 45

gruppierte und gestapelte Balken 45, 48

Punktdiagramme und Heatmaps 49, 52

R

Rastergrafik *siehe auch* Bitmapgrafik 289

redundante Codierung 219, 228

Gestaltung von Legenden mit 219, 225

Regenbogenskala 213

Relationen (x-y), Visualisierung von 37

Reproduzierbarkeit 296

angewandt auf Datenvisualisierung 296

Schwierigkeit bei interaktiver Visualisierungssoftware 297

Ridgeline-Diagramm 35, 39

Anzeigen von bayesschen A-Posteriori-Verteilungen 179

Visualisierung mehrerer Verteilungen 82, 83

Rot-Grün-Farbsehschwäche 214, 220

rotierte Beschriftungen 43

S

saisonale Zeitreihenzerlegung nach LOESS 144

Sättigung von Farben 211, 220

Schätzung, Visualisierung von 38, 168

schlechte Abbildungen 3

Schraffur 269

Sehvermögen, Farbsehschwäche 214, 217, 220

Simulation von Legenden für 224

sequenzielle Farbskala 27, 213

für Menschen mit Farbsehschwäche 215

signifikanter Unterschied 172

Sina-Diagramm 34, 80

Skala 10

Farbe 25, 31

lineare 16

logarithmische 17

Positionsskala 11

Quadratwurzel 20

Skalierung bei Small Multiples 231

skalierungsfreie Verteilung 68

Small Multiples 229, 234

anstelle von 3D-Visualisierungen 284

Mehrfachdarstellungen 311

mit zu wenig Kontext 253

Software, zur Datenvisualisierung, automatische Generierung von Legenden 223

spezifische Elemente in Daten farbig hervorheben 30, 31

Spline 135

Standardabweichung 168

vs. Standardfehler 170

Standardfehler 169

stark verzerrte Verteilung 68, 71

Steigungsdiagramm 115

Stichprobe 168

Stichprobenerhebung 168

Stichprobengröße 169

Storytelling

 Action–Background–Development–Climax–Ending 304

 Eröffnung – Herausforderung – Aktion – Lösung 304

Streifendiagramm 34

Streudiagramm

 3D, Probleme mit 282

 3D, und 3D-Positionsskala 279

 Angabe von linearen Trendlinien an den Datenpunkten 138

 Direktbeschriftung statt Legenden 226

 gepaarter Daten, diagonale Linie anstelle eines Gitters 261

 Gitterlinien entlang beider Achsen 261

 Linienzeichnungen in 272

 Locally Estimated Scatterplot Smoothing (LOESS) 134

 mit Fehlerbalken 175

 verbundenes 37, 125

 Visualisierung einer großen Anzahl von Punkten 37

 Visualisierung verbundener Daten 115

 Visualisierung von Korrelationen zwischen quantitativen Variablen 105, 109

 vs. Punktdiagramm einer Zeitreihe 120

Strip-Chart

 Visualisierung mehrerer Verteilungen 80

T

Tabellen 247, 249

 Beispiel 247

- Beschriftung relativ zum Anzeigeelement 249
- Formatierung in Abbildungen 247
 - visuelle Unordnung durch horizontale und vertikale Linien 249
- Teildichte, Visualisierung von Proportionen 93
- Text 9
 - Ersetzen von Schriftarten 290
 - lesbar, ausbalanciert im Vergleich zum Rest der Abbildung 267
 - wahrgenommene Dunkelheit 257
 - zu klein im Vergleich zum Rest der Abbildung 266
- Themes 301
- TIFF-Datei 292
- Titel
 - Abbildungstitel und Beschriftungen 241, 243
 - Achsen und Legenden 244, 246
- Tönung, bei Box-Plots 274
- Topografie, in 3D 285
- Transparenz
 - Gebrauch von partieller Transparenz bei überlappenden Punkten 200
 - unzureichende bei hoher Anzahl an Punkten 202
- transversale Mercator-Projektion 150
- Treemap 36
 - Einschränkungen 99
 - menschliche Wahrnehmung, Probleme der 198
 - ohne explizite Achsen 246
 - Visualisierung verschachtelter Proportionen 98
 - vs. verschachtelte Kreisdiagramme 102
 - vs. Mosaikdiagramm 98
- Trellis-Plot 229
- Trendbereinigung und Zerlegung von Zeitreihen 140, 145
- Trends, Visualisierung von 37, 131, 145

- Glättung einer Zeitreihe 131, 136
- in x-y-Relationen 37
- Trendbereinigung und Zerlegung von Zeitreihen 140, 145
- Trends mit einer definierten funktionalen Form 137, 140

Tritanomalie/Tritanopie 214, 220

U

überlagernde Dichten 34

überlappende Punkte, Umgang mit 199, 208

- Gebrauch von 2D-Histogrammen 202, 204

- Gebrauch von Konturlinien 204, 208

- Gebrauch von partieller Transparenz und Jittering 199, 202

Ungenauigkeiten, Visualisierung von 163, 185

- bei Kurvenanpassungen 180, 183

- Framing-Wahrscheinlichkeiten als Häufigkeit 163, 168

- für Punktschätzungen 168, 180

- hypothetisches Ergebnisdiagramm (hypothetical outcome plot, HOP) 183, 185

Unsicherheit (mathematische), Visualisierung von 38

USA

- Karte mit Albers-Projektion 152

V

Variablen, näherungsweise lineare Beziehungen zwischen 137

Vektorgrafik 289, 290, 291

- enorme Dateigrößen bei großen/komplexen Abbildungen 291

- PDF, EPS und SVG Bilddateiformate 289

- speichern als Bitmap 293

- Unterschiede in der Darstellung bei zwei verschiedenen Programmen 290

verbundene Daten 37

- Daten und Kontext in Einklang bringen 261, 263

verbundenes Streudiagramm 37

- Visualisierung einer Zeitreihe mit zwei oder mehr Antwortvariablen 125

Visualisierung höherdimensionaler Zeitreihen mit
Hauptkomponentenanalyse (PCA) 127

Verhältnisse

auf einer logarithmischen Skala *siehe auch* Proportionen

Visualisierung auf logarithmischen Achsen 194

mit Balken 196

verlustbehaftete Komprimierung

Grundidee 292

in JPEG-Bildern 292, 294

verlustfreie Komprimierung 291

verschachtelte Proportionen 95, 103

fehlerhafte Ansätze zur Visualisierung 95, 97

Visualisierung mit Mosaikdiagrammen und Treemaps 97, 100

Visualisierung mit Parallel-Sets-Diagrammen 102, 103

Visualisierung mit verschachtelten Kreisdiagrammen 100

Verteilungen 34

empirisch kumulative Häufigkeitsverteilung und Q-Q-Plots 65, 73

stark verzerrte Verteilungen 68, 71

Histogramme und Dichtediagramme 55, 64

simultane Visualisierung mehrerer Dichten 60, 64

Visualisierung einer einzelnen Verteilung 55

mehrere Verteilungen gleichzeitig 75, 83

auf der horizontalen Achse 82, 83

auf der vertikalen Achse 75, 81

Violin-Plot 34, 39

mit Fehlerbalken 179

Visualisierung mehrerer Verteilungen 78, 80

Visualisierungen

ausgewählte 33

häufig genutzte 33

Proportionen 35

quantitative Werte 33

Verteilungen 34

Visualisierungssoftware *siehe auch* Datenvisualisierungssoftware

Visualisierungstypen

Relationen 37

Unsicherheit (mathematische) 38

visuelle Sprache, einheitlich und konsistent 236

W

Wahrscheinlichkeit, Framing-Wahrscheinlichkeiten als Häufigkeit 163, 168

Wahrscheinlichkeitsverteilung 166

Web-Mercator-Projektion 150

Wiederholbarkeit 296

angewandt auf Datenvisualisierung 296

Schwierigkeit bei interaktiver Visualisierungssoftware 297

Wiederholungen vermeiden

aber gleichzeitig konsistent bleiben 315

aber konsistent bleiben 319

X

x-y-Relationen, Visualisierung von 37

Z

Zeitreihen 119, 129

mit zwei oder mehreren Antwortvariablen, Visualisierung von 124, 129

Trends, Visualisierung von 131

Visualisierung einer einzelnen Zeitreihe 119, 122

Visualisierung entlang linearer Achsen 191

Visualisierung mehrerer Zeitreihen und Dosiswirkungskurven 122, 124

Visualisierung von x-y-Relationen 37

Zeitreihen, Zerlegung von 142

Zittern (Jittering)

Gebrauch bei überlappenden Punkten 200