

---

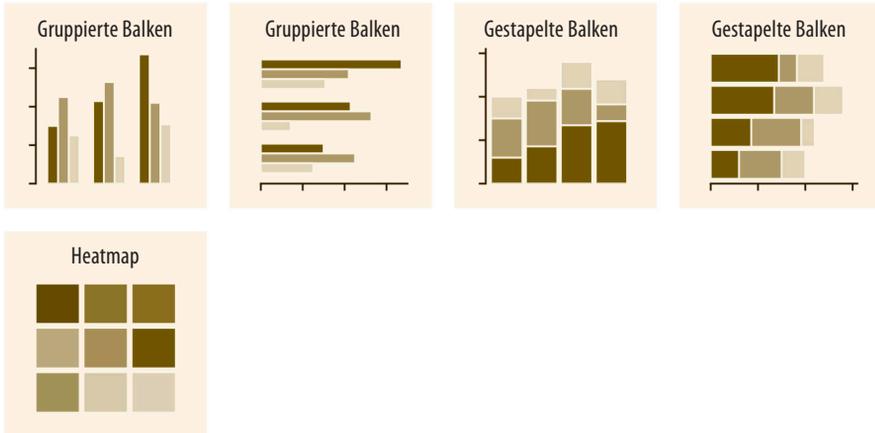
# Ausgewählte Visualisierungen

Dieses Kapitel bietet anhand von typischen Beispielen einen schnellen Überblick über die verschiedenen Diagramme und Darstellungen, die häufig zur Visualisierung diverser Datentypen verwendet werden. Es dient sowohl zum Nachschlagen, für den Fall, dass Sie nach einer bestimmten Visualisierung suchen, deren Namen Sie möglicherweise nicht kennen, wie auch als Inspirationsquelle, falls Sie Alternativen zu den routinemäßig erstellten Diagrammen brauchen.

## Quantitative Werte

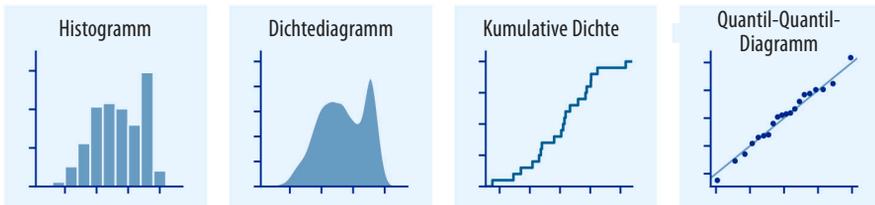


Der gebräuchlichste Ansatz zur Visualisierung von Mengen (d. h. von Zahlenwerten, die für bestimmte Kategorien angezeigt werden) ist die Verwendung von vertikal oder horizontal angeordneten Balken (Kapitel 6). Anstatt Balken zu verwenden, können wir jedoch auch Punkte an der Stelle platzieren, an der die entsprechenden Balken enden würden (Kapitel 6).



Wenn es zwei oder mehr Kategorien gibt, für die Mengen angezeigt werden sollen, können wir die Balken gruppieren oder stapeln (Kapitel 6). Wir können die Kategorien auch auf die  $x$ - und  $y$ -Achse abbilden und die Mengen über eine Heatmap nach Farben anzeigen (Kapitel 6).

## Verteilungen

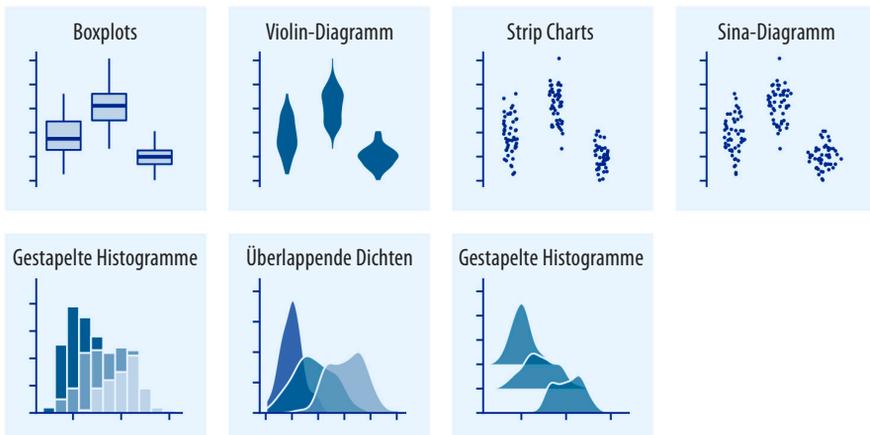


Histogramme und Dichtediagramme (Kapitel 7) bieten die intuitivsten Darstellungen einer Verteilung, jedoch erfordern beide eine bewusste, explizite Festlegung von Parameterwerten und können irreführend sein.

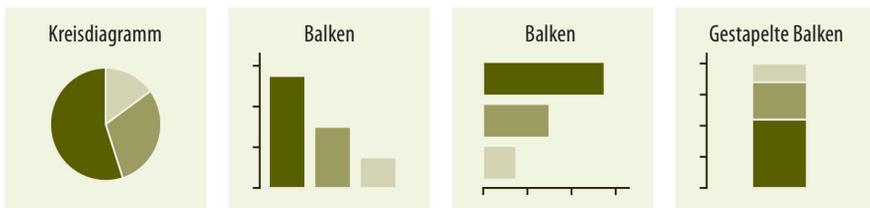
Kumulative Dichten und Quantil-Quantil-Diagramme (Q-Q-Diagramme) (Kapitel 8) geben die Daten immer originalgetreu wieder, sind jedoch möglicherweise schwieriger zu interpretieren.

Box-Plots (Kastengrafiken), Violin-Plots, Streifen- und Sina-Diagramme sind nützlich, wenn wir viele Verteilungen gleichzeitig visualisieren möchten und/oder hauptsächlich an Gesamtverschiebungen zwischen den Verteilungen interessiert sind (siehe »Visualisierung von Verteilungen entlang der vertikalen Achse« auf Seite 75). Gestapelte Histogramme und überlagernde Dichten ermöglichen einen detaillierteren Vergleich einer kleineren Anzahl von Verteilungen. Gestapelte Histo-

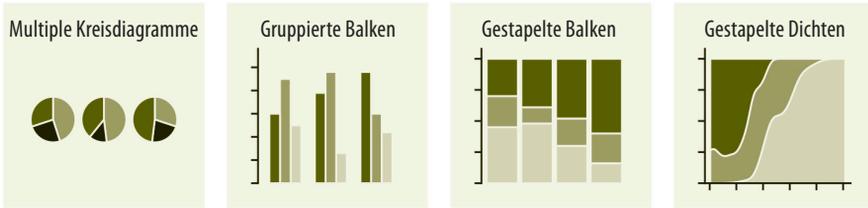
gramme sind jedoch möglicherweise schwer zu interpretieren und sollten am besten vermieden werden (siehe »Gleichzeitige Visualisierung mehrerer Verteilungsgrößen« auf Seite 60). Ridgeline-Diagramme können eine nützliche Alternative zu Violin-Plots sein. Dies ist häufig hilfreich, wenn eine sehr große Anzahl von Verteilungen oder Änderungen der Verteilungen im Zeitverlauf angezeigt werden sollen (siehe »Visualisierung von Verteilungen entlang der horizontalen Achse« auf Seite 82).



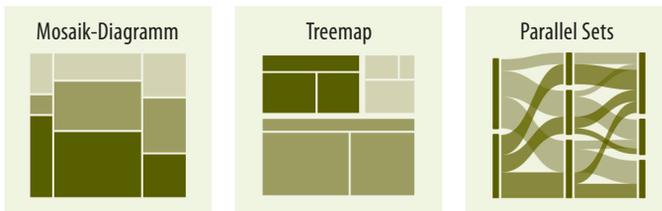
## Proportionen



Proportionen können als Kreisdiagramme, nebeneinanderliegende Balken oder gestapelte Balken dargestellt werden (Kapitel 10). Genauso wie bei Mengen können wir, wenn wir Proportionen mit Balken visualisieren, die Balken entweder vertikal oder horizontal anordnen. Kreisdiagramme unterstreichen, dass die einzelnen Teile ein Ganzes ergeben und heben einfache Anteile hervor. Jedoch ist es leichter, die einzelnen Fragmente in nebeneinander angeordneten Balken zu vergleichen. Gestapelte Balken können schwieriger zu lesen sein, wenn nur eine Größe in ihre Anteile zerlegt wird; sie können jedoch beim Vergleich der jeweiligen Proportionen mehrerer Größen hilfreich sein.

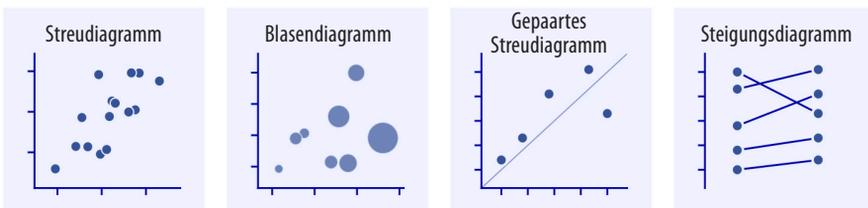


Wenn mehrere Proportionen oder Veränderungen der Proportionen über geänderte Bedingungen hinweg dargestellt werden sollen, sind Kreisdiagramme eher ineffizient. Häufig sind sie ungünstig oder ungeeignet, um Relationen deutlich erkennbar zu machen. Gruppierte Balken funktionieren gut, solange die Anzahl der verglichenen Bedingungen moderat ist, und gestapelte Balken können für eine große Anzahl von Bedingungen funktionieren. Gestapelte Dichten (Kapitel 10) sind geeignet, wenn sich die Anteile entlang einer kontinuierlichen Variablen ändern.

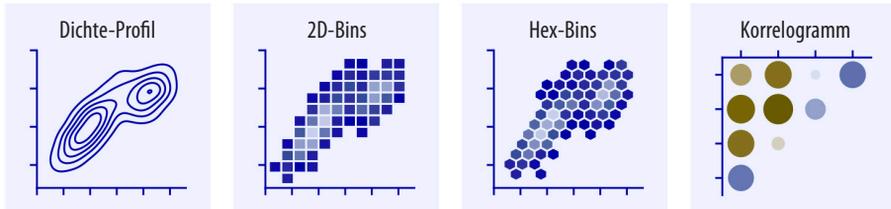


Wenn Proportionen anhand mehrerer Gruppierungsvariablen angegeben werden, sind Mosaikdiagramme, Treemaps oder Parallel Sets nützliche Visualisierungsansätze (Kapitel 11). In Mosaikdiagrammen wird davon ausgegangen, dass jede Ebene einer Gruppierungsvariablen mit jeder Ebene einer anderen Gruppierungsvariablen kombiniert werden kann, wohingegen Treemaps eine solche Annahme nicht treffen. Treemaps funktionieren auch dann gut, wenn sich die Unterteilungen einer Gruppe von den Unterteilungen einer anderen Gruppe unterscheiden. Parallel Sets funktionieren besser als jedes Mosaikdiagramm oder Treemaps, wenn mehr als zwei Gruppierungsvariablen vorhanden sind.

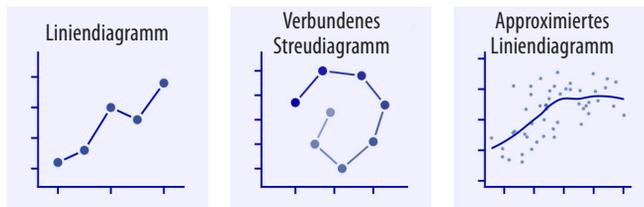
## x-y-Relationen



Streudiagramme (Kapitel 12) stellen den visuellen Prototyp dar, wenn eine quantitative Variable relativ zu einer anderen dargestellt werden soll. Wenn wir drei quantitative Variablen haben, können wir eine auf die Punktgröße abbilden und eine Variante des Streudiagramms erstellen, die als Blasendiagramm bezeichnet wird. Für gepaarte Daten, bei denen die Variablen entlang der  $x$ - und  $y$ -Achse gemessen werden, ist es prinzipiell hilfreich, eine Linie hinzuzufügen, die  $x = y$  angibt (siehe »Gepaarte Daten« auf Seite 115). Gepaarte Daten können auch als Steigungsdiagramme (engl. *Slopegraphs*) gepaarter Punkte angezeigt werden, die durch gerade Linien verbunden sind.

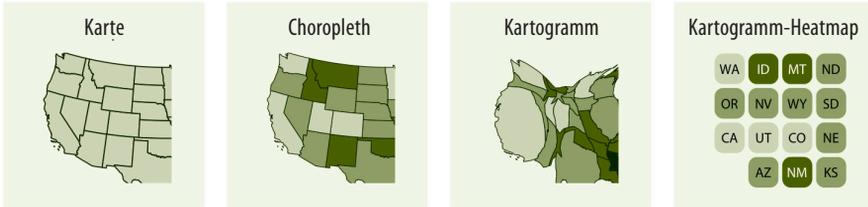


Bei einer großen Anzahl von Punkten kann es vorkommen, dass gewöhnliche Streudiagramme aufgrund der Überlagerung von Datenpunkten nicht mehr aussagekräftig sind. In diesem Fall können Konturlinien, 2D-Bins oder Hex-Bins eine Alternative darstellen (Kapitel 18). Wenn wir andererseits mehr als zwei Größen visualisieren möchten, können wir Korrelationskoeffizienten in Form eines Korrelogramms anstelle der zugrunde liegenden Rohdaten darstellen (siehe »Korrelogramme« auf Seite 109).



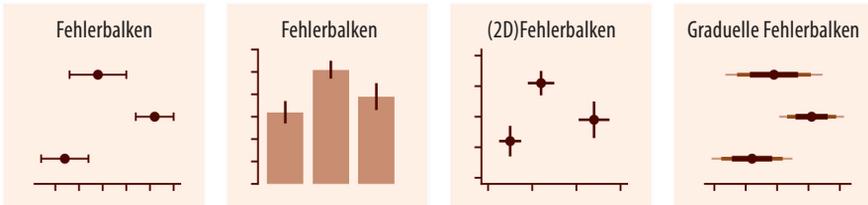
Wenn die  $x$ -Achse die Zeit oder eine kontinuierlich ansteigende Größe darstellt, wie z. B. eine Behandlungsdosis, zeichnen wir üblicherweise Liniendiagramme (Kapitel 13). Wenn wir eine zeitliche Folge von zwei Größen haben, können wir ein sogenanntes verbundenes Streudiagramm zeichnen, indem wir zuerst die Werte beider Größen in ein Streudiagramm zeichnen und dann die Punkte verbinden, die benachbarten Zeitpunkten entsprechen (siehe »Zeitreihe von zwei oder mehr Antwortvariablen« auf Seite 124). Und wir können durchgehende Linien verwenden, um Trends in einem größeren Datensatz darzustellen (Kapitel 14).

# Geodaten

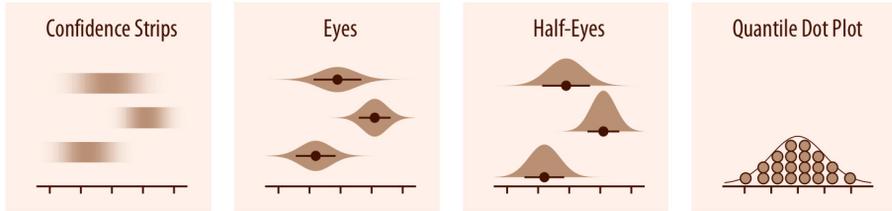


Geodaten werden hauptsächlich in Form einer Karte angezeigt (Kapitel 15). Eine Karte nimmt Koordinaten des Globus auf und projiziert sie auf eine flache Oberfläche, sodass Formen und Abstände auf dem Globus entsprechend in der 2D-Darstellung visualisiert werden. Zusätzlich können wir Datenwerte in verschiedenen Regionen anzeigen, indem wir gemäß der Daten diese Regionen in der Karte färben. Eine solche Karte wird als Choroplethenkarte (oder auch Flächenkartogramm) bezeichnet (siehe »Choroplethenkartierung« auf Seite 156). In einigen Fällen kann es hilfreich sein, die verschiedenen Regionen gemäß einer anderen Größe (z.B. der Bevölkerungszahl) zu verzerren oder jede Region in quadratischer Form zu vereinfachen. Solche Visualisierungen nennt man Kartogramme (Siehe »Kartogramme« auf Seite 160).

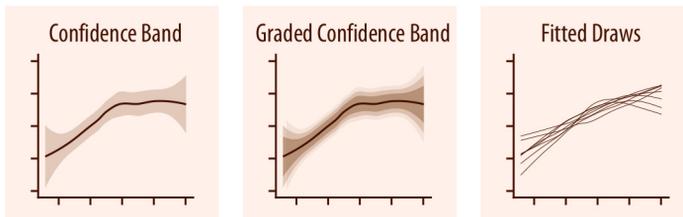
## (Mathematische) Unsicherheit



Fehlerbalken sollen den Bereich der wahrscheinlichen Werte für eine Schätzung oder Messung anzeigen. Sie erstrecken sich horizontal und/oder vertikal von einem Bezugspunkt, der die Schätzung oder Messung darstellt (Kapitel 16). Referenzpunkte können auf verschiedene Arten angezeigt werden, z.B. durch Punkte oder Balken. Abgestufte Fehlerbalken zeigen mehrere Bereiche gleichzeitig an, wobei jeder Bereich einem unterschiedlichen Konfidenzintervall entspricht. Es handelt sich im Grunde um mehrere Fehlerbalken mit unterschiedlichen Strichstärken, die übereinander aufgetragen sind.



Um eine detailliertere Visualisierung zu erzielen, als dies mit Fehlerbalken oder abgestuften Fehlerbalken möglich wäre, können wir die tatsächlichen Konfidenzintervalle oder A-posteriori-Verteilungen visualisieren (Kapitel 16). Konfidenzintervalle vermitteln ein visuelles Gefühl der Ungenauigkeit, sind jedoch schwer präzise abzulesen. Sogenannte *Eye*- und *Half-Eye*-Darstellungen kombinieren Fehlerbalken mit Visualisierungsansätzen für Verteilungen (Violinen bzw. Ridgelines) und zeigen somit sowohl genaue Bereiche für einige Konfidenzintervalle als auch die Verteilung der Gesamtunsicherheit an. Ein sogenanntes Quantil-Punktplot kann als alternative Visualisierung einer Unsicherheitsverteilung dienen (siehe »Wahrscheinlichkeiten als Häufigkeiten darstellen« auf Seite 163). Da diese Darstellungsform die Verteilung in diskreten Einheiten zeigt, ist das Quantil-Punktplot nicht so genau, kann aber leichter abgelesen werden als die kontinuierliche Verteilung, die durch ein Violin- oder Ridgeline-Diagramm gezeigt wird.



Bei kontinuierlichen Liniendiagrammen fungiert ein Konfidenzintervall als Äquivalent eines Fehlerbalkens (siehe »Visualisierung der Unsicherheit von Kurvenanpassungen« auf Seite 180). Es zeigt eine Reihe von Werten an, die die Linie bei einem bestimmten Konfidenzniveau durchlaufen könnte. Wie bei Fehlerbalken können wir graduelle Konfidenzintervalle zeichnen, die mehrere Konfidenzniveaus gleichzeitig anzeigen. Wir können auch anstelle oder zusätzlich zu den Konfidenzbändern einzelne approximierende Linien anzeigen.