

Sören Göckel

Lügen mit Zahlen und Karten

Inhaltsverzeichnis

Statistiken lesen lernen.....	2
Lügen mit Diagrammen.....	2
Verkürzte Y-Achse.....	3
Günstig gewählte Zeitreihe.....	5
Skalierung in zu vielen Dimensionen.....	8
Lügen mit Landkarten.....	9
Falsch gewählte Kartenprojektion.....	9
Absolutwerte als Einfärbung.....	10
Gewichtung nach Fläche statt Wert.....	11
Lügen mit Zahlen.....	12
Verzerrte Stichproben und Self-Selection.....	12
Korrelation ist nicht Kausalität.....	12
Anteil ist nicht Absolutwert.....	12
Fünf Rückfragen an Statistiken.....	13
Weiterführende Literatur.....	13

Statistiken lesen lernen

„Statistik ist für mich das Informationsmittel der Mündigen. Wer mit ihr umgehen kann, kann weniger leicht manipuliert werden. Der Satz ‚Mit Statistik kann man alles beweisen‘ gilt nur für die Bequemen, die keine Lust haben, genau hinzusehen.“

- Elisabeth Noelle-Neumann, Gründerin des Allensbach-Instituts für Demoskopie

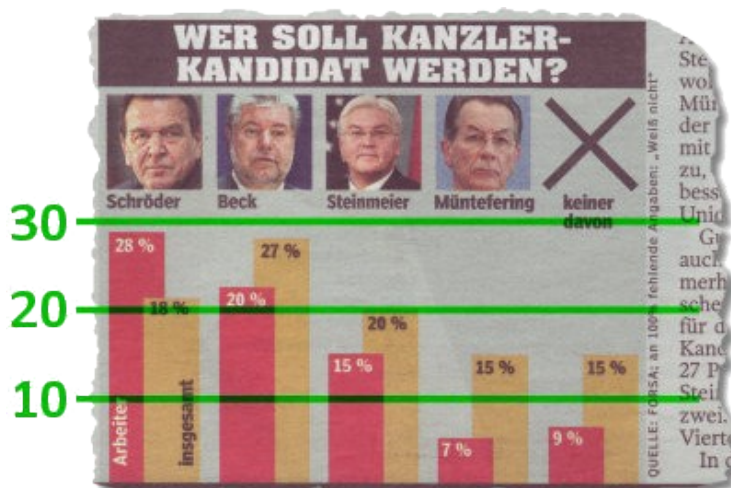
Sätze wie „Glaube keiner Statistik, die du nicht selbst gefälscht hast“ suggerieren, dass sich Statistiken beliebig fälschen ließen. Und tatsächlich lassen häufig sowohl die reinen Zahlen als auch deren Darstellung großen Spielraum für Manipulationen. Hier sollen einige der häufigsten vorgestellt werden, in der Hoffnung, den Blick für derartige Fälschungen zu schärfen.

Lügen mit Diagrammen

Die einfachste Möglichkeit, das Unterbewusstsein des Betrachters zu beeinflussen, liegt in einer verzerrten Darstellung der - an sich korrekten - Zahlenwerte. Nicht immer ist die Fälschung so offensichtlich wie bei diesen beiden Schaubildern:



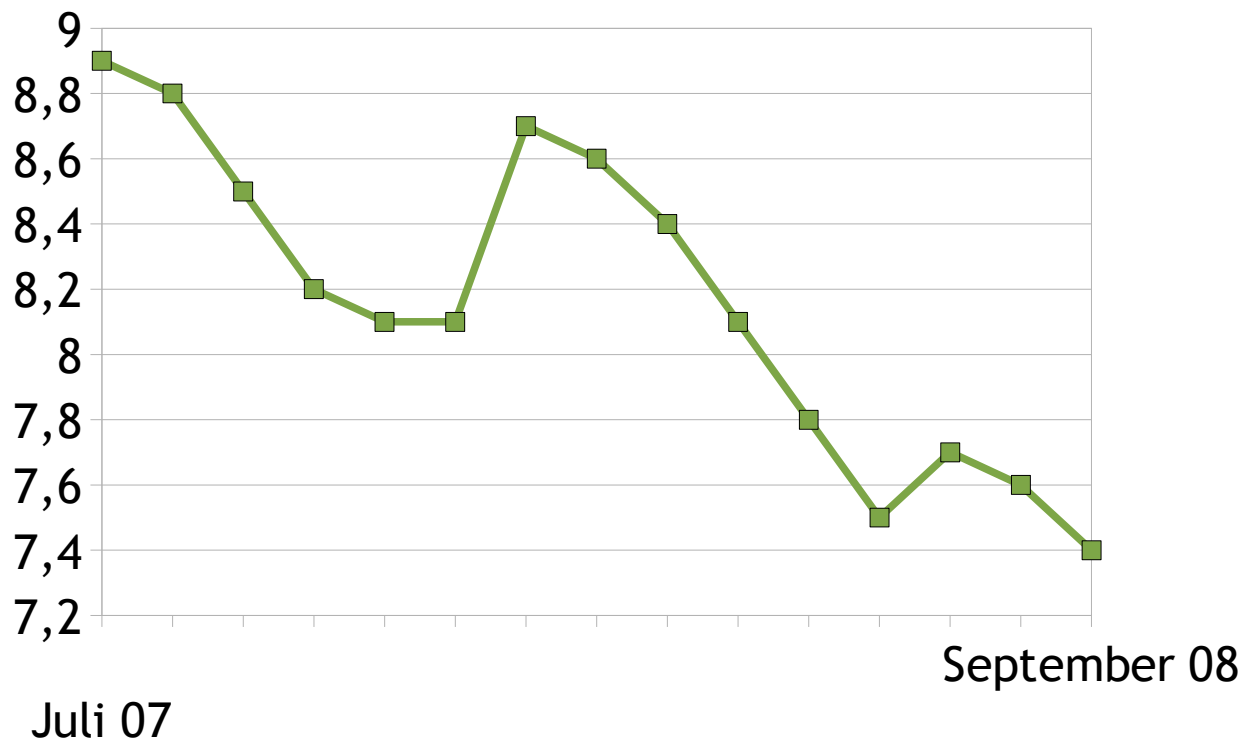
Quelle: bild.de via bildblog.de



Quelle: BILD via bildblog.de

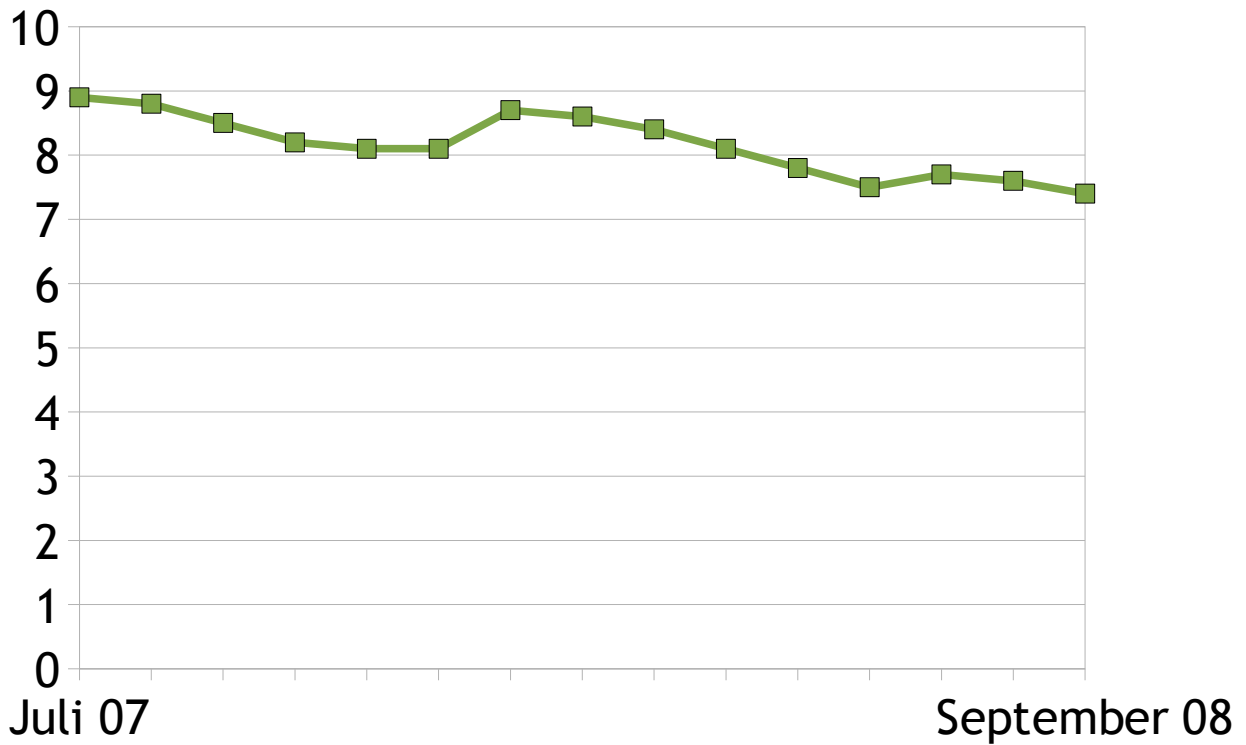
Verkürzte Y-Achse

Sehr häufig zu sehen ist eine Verkürzung der Y-Achse:



Entwicklung der Arbeitslosenquote in Deutschland, Quelle: statista.de

Durch die Verkürzung der Y-Achse wirkt die Senkung der Arbeitslosenquote wesentlich eindrucksvoller, als wenn die Y-Achse tatsächlich bei 0 beginnt:

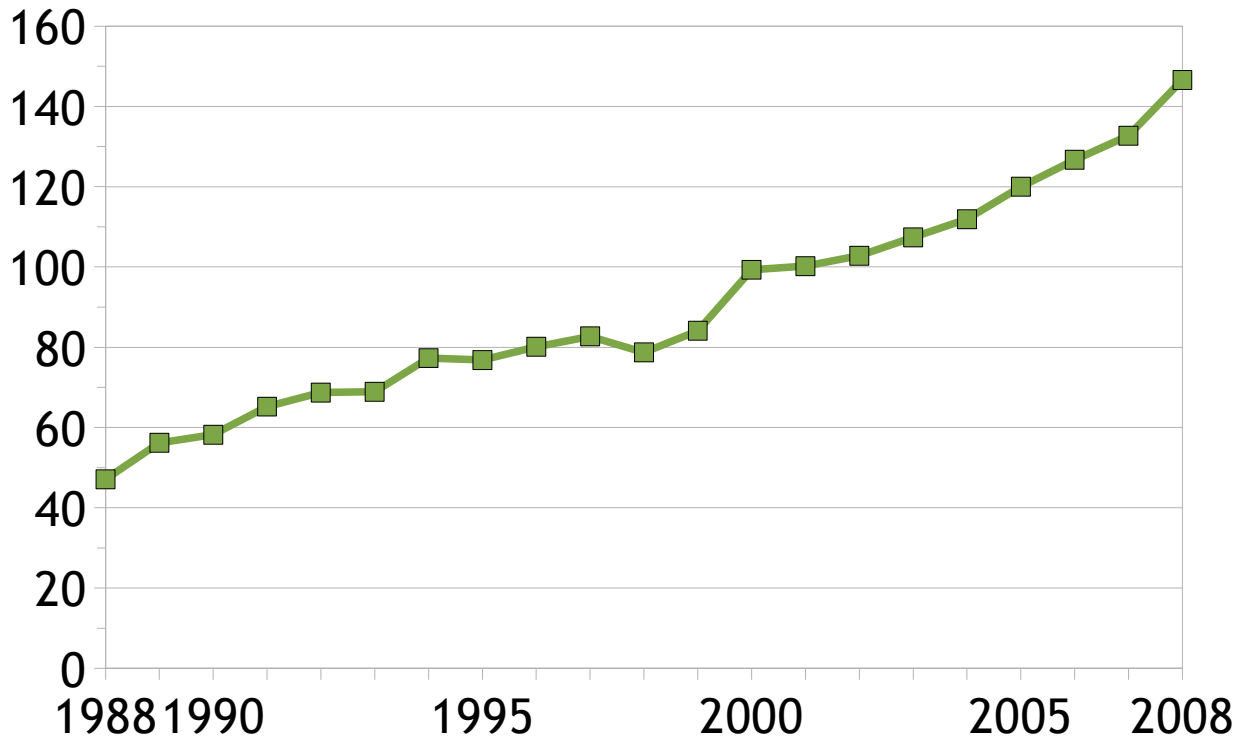


Die gleiche Entwicklung mit 0-basierter Y-Achse

Diese Darstellungsweise ist häufig anzutreffen, wenn Verbesserung oder Verschlechterungen eines Zustands bedeutender herausgestellt werden sollen, als sie tatsächlich sind.

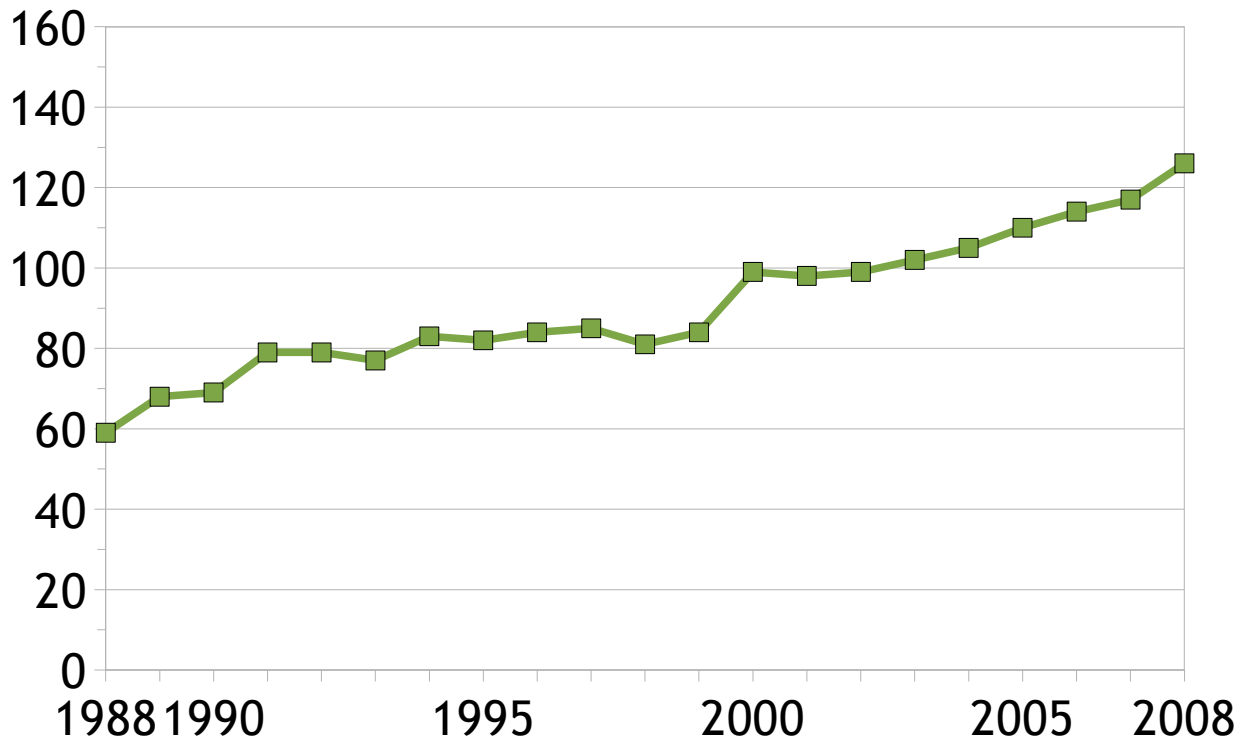
Günstig gewählte Zeitreihe

Folgendes Schaubild leidet gleich an zwei Probleme (wobei beide mehr in den zugrunde liegenden Zahlen als der optischen Darstellung zu suchen sind):



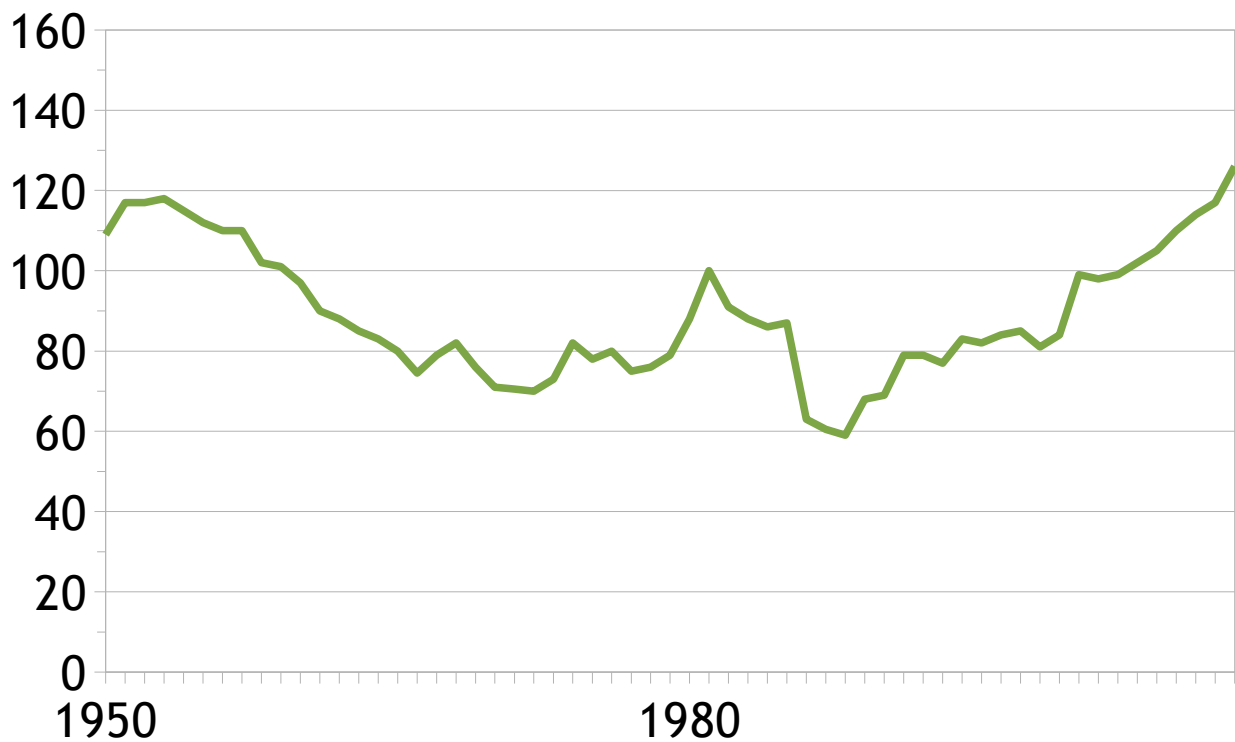
Benzinpreis für Normalbenzin in Eurocent; Quelle: Wikipedia

Das erste Problem ist, dass der Einfluss der allgemeinen Teuerung auf den Benzinpreis vernachlässigt wird; dieser Einfluss kann durch Anpassung der Preise auf einen einheitlichen Geldwert herausgerechnet werden:



Benzinpreis für Normalbenzin in Eurocent (Wert von 2000); Quelle: Eigenberechnung auf Basis von statista.de, Wikipedia

Das zweite Problem ist die Wahl des betrachteten Zeitbereichs: Durch die Wahl eines Bereichs von 1988-2008 erscheint der aktuelle Benzinpreis als „höchster Preis aller Zeiten“. Wird die Zeitreihe dagegen in die Vergangenheit erweitert, so zeigt sich, dass das Bild eines stetig ansteigenden Preises falsch ist:

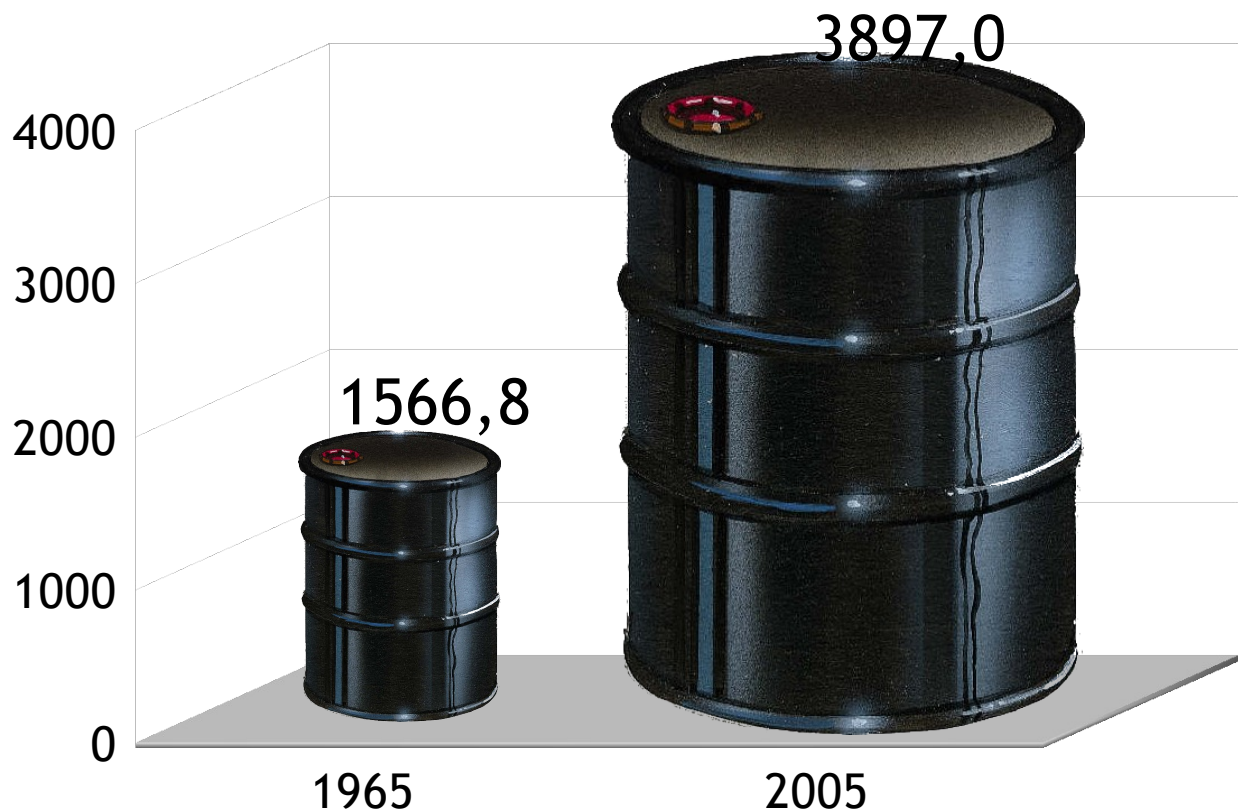


Die gleiche Zeitreihe seit 1950; Datenquellen: *statista.de*, *Wikipedia*

Verkürzte (oder „günstig gewählte“) Zeitreihen werden häufig eingesetzt, wenn z.B. die Performance von Wertpapieren in einem positiven Licht dargestellt werden soll.

Skalierung in zu vielen Dimensionen

Häufig werden für die Darstellung von Wertsteigerungen skalierte Objekte an Stelle von Linien- oder Balkengraphiken verwendet:



Weltweite Erdölförderung in Mio. t, Datenquelle: Wikipedia

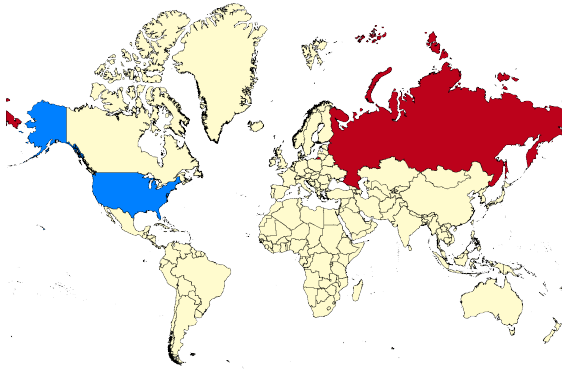
Hierbei ist zu beachten, dass das Objekt nicht nur in der Höhe, sondern auch in der Breite (und bei einigermaßen realistischer 3D-Darstellung für das Gehirn auch noch in der Tiefe) skaliert, d.h. bei einer Verdoppelung der Höhe wird das Objekt nicht nur doppelt, sondern viermal oder gar achtmal so „groß“. Orientiert sich die Skalierung nur an der Höhe des Objekts, so erscheinen Steigerungen oder Verringerungen des Wertes wesentlich eindrucksvoller, als sie dies in Wirklichkeit sind.

Lügen mit Landkarten

Richtig eingesetzt, sind thematische Landkarten eine hervorragende Möglichkeit, Daten mit Raumbezug intuitiv verständlich aufzubereiten. Falsch eingesetzt, verbergen sie Zusammenhänge oder zeigen falsche Zusammenhänge auf.

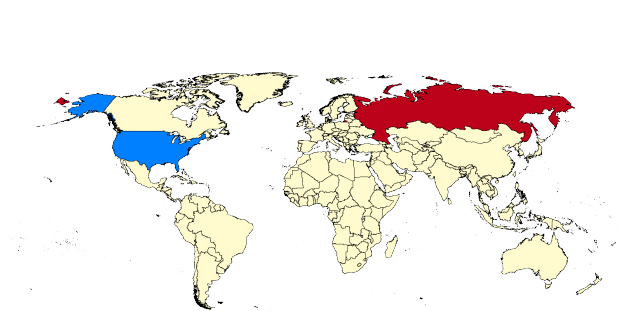
Falsch gewählte Kartenprojektion

Welche der beiden folgenden Weltkarten stellt die Erde richtig dar?



Mercatorprojektion

Quelle: GfK GeoMarketing GmbH



Eckert-III-Projektion

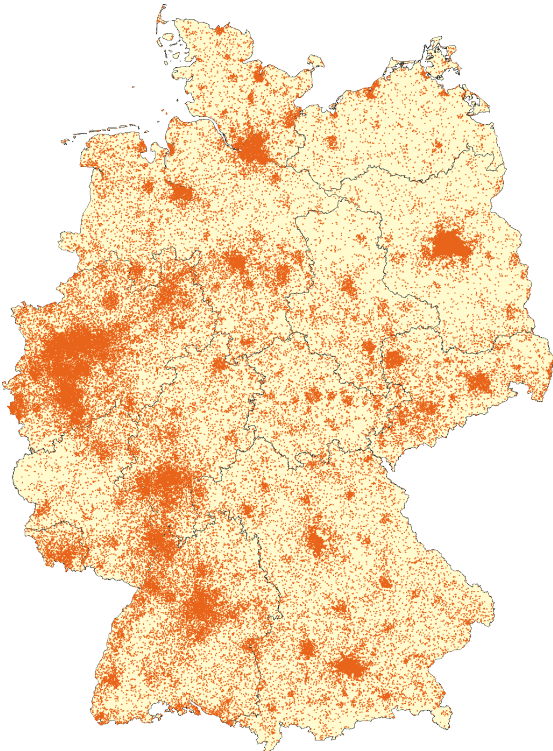
Quelle: GfK GeoMarketing GmbH

Die Antwort lautet: Keine oder beide, je nachdem, was man als „richtig“ bezeichnet. Die Erde auf einer flachen Landkarte verzerrungsfrei darzustellen, ist nicht möglich, da die Weltkugel nicht verzerrungsfrei in die Ebene abgewickelt werden kann. Daher ist es immer notwendig, eine Projektion durchzuführen, die zwangsläufig zu Verzerrungen bei Formen, Entfernungen oder Größenverhältnissen führt. So bleiben in der links gezeigten Mercatorprojektion zwar die Formen von Objekten erhalten, dies wird jedoch mit einer starken Vergrößerung der Objekte in Polnähe erkauft. Deshalb fand man diese Projektion häufig in amerikanischen Landkarten im Kalten Krieg, da durch die Verzerrung die (nördlicher als die USA gelegene) Sowjetunion größer erschien und somit das Bild einer Bedrohung besser vermittelt werden konnte.

Die rechts dargestellte Eckert-III-Projektion zeigt alle Länder im richtigen Flächenverhältnis, weist dafür jedoch starke Formverzerrungen im Randbereich auf.

Absolutwerte als Einfärbung

Häufig sieht man Karten, bei denen z.B. die Einwohnerzahl eines Gebiets in Form einer bestimmten Farbe dargestellt wird. Da die absolute Einwohnerzahl jedoch von der Größe ebenso wie von der Einwohnerdichte abhängig ist, kann die Färbung eines Gebietes je nach Zuschnitt und Größe der Einzelgebiete variieren. Besser ist hier die Wahl einer Punktdichte-Darstellung, die (im Beispiel der Einwohnerzahlen) die Dichteinformation quasi von allein erzeugt:

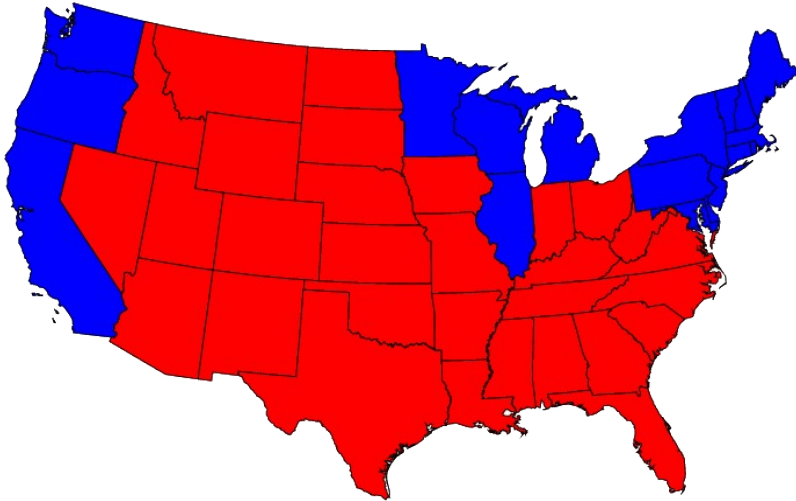


*Ein Punkt steht für 500 Einwohner.
Quelle: GfK GeoMarketing GmbH*

Möglich wäre hier ebenfalls, die Bevölkerungsdichte statt der Absolutwerte zu verwenden und in Farben umzusetzen.

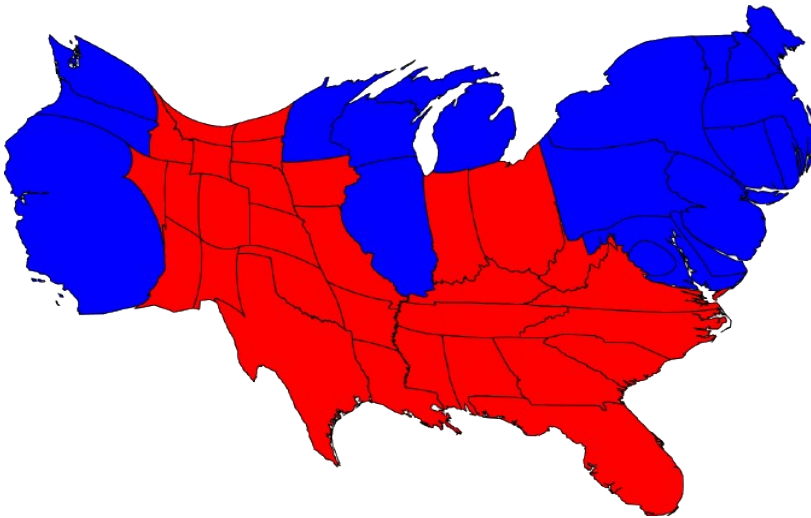
Gewichtung nach Fläche statt Wert

Die folgende Karte zeigt die Ergebnisse der einzelnen Bundesstaaten bei der US-Präsidentschaftswahl 2004. Rote Staaten gingen an George W. Bush, blaue an John Kerry:



Quelle: Mark Newman, University of Michigan

Offensichtlich hat Bush zu Recht gewonnen - die rote Fläche ist wesentlich größer als die blaue. Damit setzen wir jedoch die Fläche eines Staates proportional zu seiner Bedeutung; dabei ist die Zahl der Wahlmänner, die der Staat stellt, das entscheidende Kriterium. Da gerade die vielen großen Staaten des mittleren Westens sehr dünn besiedelt sind (und damit nur wenige Wahlmänner entsenden), zeichnet die Verzerrung der Karte, so dass die Fläche proportional zur Zahl der Wähler ist, ein ganz anderes Bild:



Quelle: Mark Newman, University of Michigan

Ein solches Kartogramm, auch wenn es keine realistische Darstellung der Geographie mehr ist, transportiert die Information in diesem Fall wesentlich besser als die reine Karte.

Lügen mit Zahlen

Endgültig zu spät ist es natürlich, wenn schon die der Darstellung zu Grunde liegenden Zahlen falsch sind. Auch hierfür gibt es zahlreiche Möglichkeiten; im Folgenden sollen nur drei dargestellt werden.

Verzerrte Stichproben und Self-Selection

„Laut einer Umfrage des ADAC ärgern sich 70% aller Autofahrer über ihr Navigationsgerät.“

Kann man dieser Zahl Glauben schenken? Eher nicht, dafür stecken einfach zu viele Unbekannte dahinter: Wie war die Frage formuliert? („Ärgern Sie sich regelmäßig über Ihr Navi?“ vs. „Haben Sie sich schon einmal über Ihr Navi geärgert?“) Und vor allem: Wer hat an der Umfrage überhaupt teilgenommen? Falls es sich um eine Leserumfrage in der ADAC-Mitgliederzeitschrift oder auf der ADAC-Homepage gehandelt hat, melden sich Autofahrer, die sich schon einmal geärgert haben, sicher häufiger als solche, die rundum zufrieden sind.

Seriöse Studien bemühen sich stets, die Teilnehmer möglichst zufällig und gleichverteilt („repräsentativ“) aus der Gesamtmenge (Grundgesamtheit) auszuwählen. Falls das Auswahlverfahren dafür nicht geeignet ist (Auswahl aus Telefonbüchern findet z.B. nur Teilnehmer mit Festnetzanschluss und ohne Geheimnummer) oder sich die Teilnehmer selbst für die Studie melden können, ist eine repräsentative Aussage normalerweise nicht möglich.

Korrelation ist nicht Kausalität

„Kleinere Menschen haben längere Haare.“

Auf den ersten Blick ist diese Aussage völlig absurd, aber statistisch betrachtet stimmt sie. Der Grund dafür ist, dass Frauen (im Durchschnitt) kleiner sind als Männer, aber die Haare länger tragen. Die Aussage ist also korrekt, das Problem ist die Formulierung, die suggeriert, die Haarlänge habe ihre *Ursache* in der kleineren Körpergröße - und das ist natürlich falsch.

Zwei statistische Variablen, die miteinander zusammenhängen (sich also mehr oder weniger parallel oder aber gegenläufig entwickeln), werden als *korreliert* bezeichnet. Für die Korrelation kann es jedoch mindestens vier Gründe geben:

- A ist die Ursache für B (z.B. steigen bei sinkenden Temperaturen im Winter die Heizkosten).
- B ist die Ursache für A.
- Beides lässt sich auf eine gemeinsame Ursache zurückführen (wie im Beispiel mit der Haarlänge).
- Die Korrelation ist zufällig.

Anteil ist nicht Absolutwert

Durch den Bau einer modernen Kläranlage steigt der Anteil der Nutzfische in den betroffenen Gewässern von 20% auf 80%. Ist dies ein Zeichen für die Effektivität der Kläranlage? Es kann es sein, muss aber nicht. Denn ebenso gut, wie die Anzahl der Nutzfische gestiegen sein kann, kann auch die Anzahl der Nicht-Nutzfische gefallen sein. Wenn Absolutwerte in Statistiken durch Anteile ersetzt werden, ist dies immer ein Alarmzeichen.

Fünf Rückfragen an Statistiken

Darrell Huff definiert in seinem Standardwerk „Wie lügt man mit Statistik“ fünf Fragen, die man sich bei jeder Statistik stellen sollte, um Manipulationen zu erkennen:

- **Sagt wer?** Wer hat die Statistik in Auftrag gegeben? Hat er ein Interesse an einem bestimmten Ergebnis? An welchem?
- **Woher weiß er das?** Wie kamen die Zahlen zustande? Ist die Auswahl der Stichprobe okay? Ist die Studie / Statistik handwerklich in Ordnung?
- **Was fehlt?** Fehlen in der Statistik bestimmte Werte, die man erwarten würde? Sind unpassende Teilergebnisse unter den Tisch gefallen?
- **Hat jemand stillschweigend das Thema geändert?** Beweist die Statistik tatsächlich das, was sie vorgibt? Oder zielen die angegebenen Zusammenhänge knapp am Thema vorbei? Versucht jemand, Anteile als Absolutwerte zu verkaufen?
- **Ist das Ergebnis sinnvoll?** Ergibt die Aussage einen Sinn? Falls nein, liegt das an falschen Erwartungen oder einer fehlerhaften Studie?

Weiterführende Literatur

- Huff, Darrell: „How to lie with statistics“; ISBN 0393310728 - Der Klassiker zum Thema „Lügen mit Statistik“.
- Huff, Darrell: „Wie lügt man mit Statistik“; ISBN 3725403074 - Die deutsche Version des Buches.
- Monmonier, Mark: „How to lie with maps“; ISBN 0226534219 - Über Manipulationsmöglichkeiten bei Landkarten.
- Monmonier, Mark: „Eins zu einer Million. Die Tricks und Lügen der Kartographen“; ISBN 3764353910 - Deutsche Ausgabe des Buches.
- Beck-Bornholdt, Hans-Peter; Dubben, Hans-Hermann: „Der Hund, der Eier legt. Erkennen von Fehlinformationen durch Querdenken“; ISBN 3499611546 - Über statistische Tricks, mit einem Schwerpunkt auf medizinischen Studien.