



Datenvisualisierung – Bedeutsam für die Wissenschaft und die Wirtschaft

Aktuelle Schätzungen gehen davon aus, dass heute pro Sekunde über 13.000 Terabyte an Daten generiert werden¹ – Tendenz steigend. Diese riesigen Datenmengen zu verwalten und zu verstehen wird deshalb zunehmend zur Herausforderung. In vielen Bereichen unseres Alltags ist Visualisierung nicht mehr wegzudenken: Navigations- und Wetterkarten, Visualisierung in der Medizin oder den Ingenieursdisziplinen. Für die Ökonomen gewinnt Visualisierung in den Bereichen Finanzdaten und Marktanalyse seit Jahren an Bedeutung.

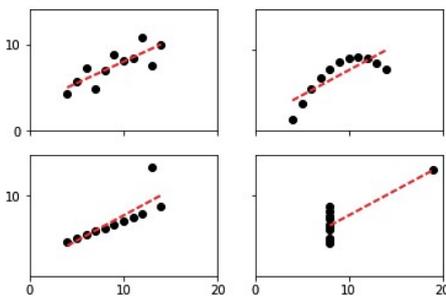
von Valentin Bruder

Warum Visualisierung wichtig ist

In der Medizin werden Messdaten beispielsweise von Computer- oder Magnetresonanztomographie für Diagnose und Therapie verbildlicht. Ein weiteres Anwendungsgebiet sind Ingenieursdisziplinen, in denen zum Beispiel Strömungen visualisiert werden, um den Kraftstoffverbrauch von Flugzeugen oder Autos zu optimieren. Visualisierung kann einen entscheidenden Teil dazu beitragen, trotz großen Datenbestands unterschiedliche Muster in den Daten zu erkennen und daraus entscheidende Schlussfolgerungen zu ziehen.

Ein klassisches Beispiel dafür, dass beschreibende Statistik alleine nicht immer für eine Datenanalyse ausreicht, ist *Anscombes Quartett*. Die Abbildung unten zeigt vier verschiedene Datensätze, deren Durchschnittswert, Varianz, Korrelation und einfache lineare Regressionsgerade beinahe gleich sind.

Anscombe's Quartet

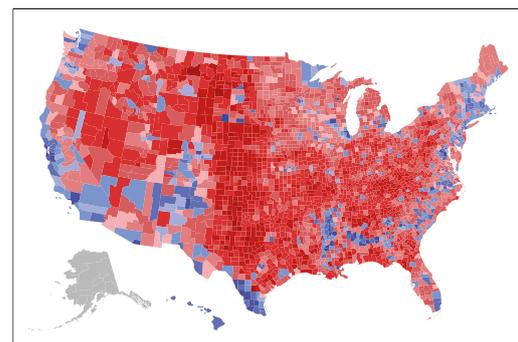
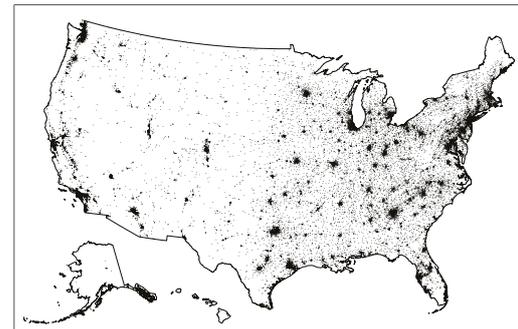


Anscombe, F. J. (1973). „Graphs in Statistical Analysis“. *American Statistician*. 27(1): 17–21.

Wie man in der Visualisierung erkennen kann, bilden die Datenpunkte jedoch sehr unterschiedliche Muster. Das Beispiel verdeutlicht warum eine Datenvisualisierung oft einen Mehrwert zu deskriptiver Statistik bietet.

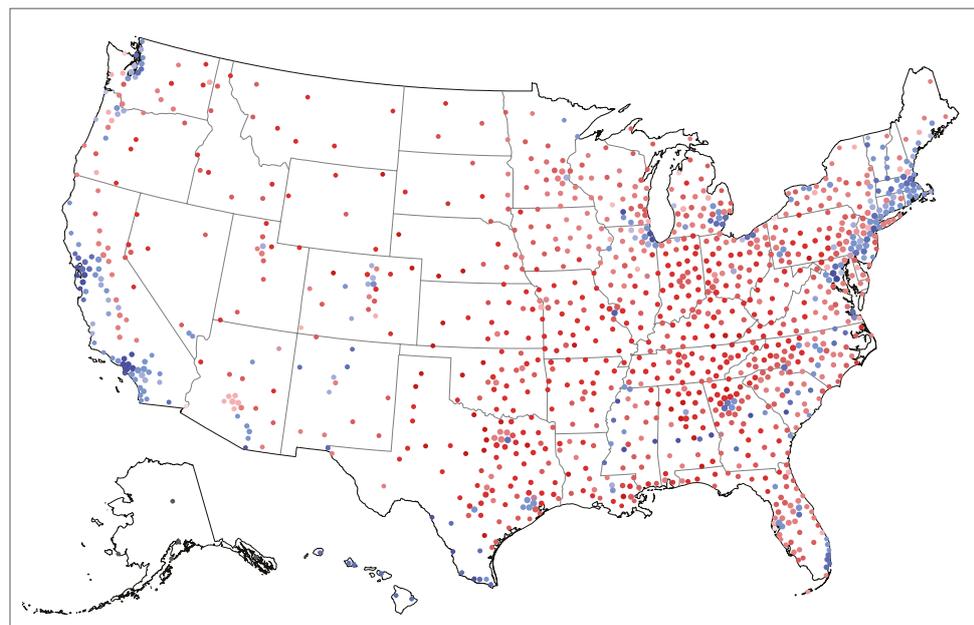
Beispiel aus der Forschung: Die USA-Präsidentenwahl visualisiert

Die folgenden Abbildungen zeigen ein aktuelles Beispiel aus der Forschung. Zu sehen ist eine Visualisierung der Ergebnisdaten der Präsidentschaftswahl in den USA 2016 (unten). Diese kombiniert dabei die Bevölkerungsdichte in der Form einer Punktstich Visualisierung (oben rechts) mit dem Wahlunterschied für jeden Wahlkreis in der Form eines Flächenkartogramms (Mitte rechts). Die kombinierte Visualisierung kann dabei beide Aspekte kompakt darstellen und in Relation setzen.



Herausforderungen in der Visualisierungsforschung

Die Visualisierung ist ein aktives Forschungsfeld welches sich durch seine fachliche Breite, sowie Interdisziplinarität auszeichnet. Neben der beinahe komplet-



Kombinierte Visualisierung (Bild: Görtler, J., Spicker, M., Schulz, C., Weiskopf, D., & Deussen, O. (2019). *Stippling of 2D Scalar Fields*. *IEEE transactions on visualization and computer graphics*, 25(6), 2193-2204.

¹ <https://www.domo.com/learn/data-never-sleeps-6#/>



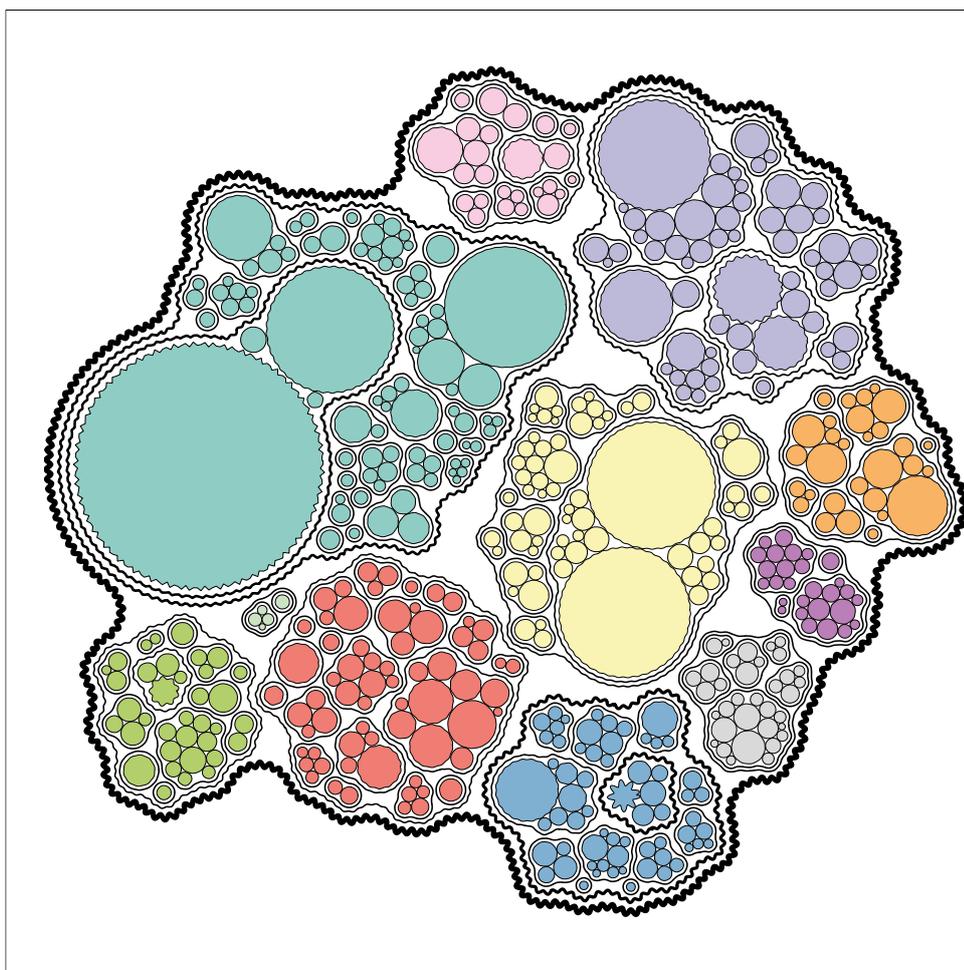
ten Bandbreite der Informatikforschung, spielen auch mathematische Themen, wie numerische Algorithmen, sowie Psychophysik und Kognitionspsychologie eine entscheidende Rolle.

Nach wie vor gibt es viele Herausforderungen und Forschungsfragen, die sich nicht zuletzt aus der Breite der Disziplin ergeben. Neben algorithmischen Herausforderungen wie dem Entwurf von Datenstrukturen, die in der Lage sind, die großen Datenmengen zu speichern und effizient zu verarbeiten, müssen Techniken zunehmend auch auf spezielle Hardwaresysteme (z.B. Grafikkarten oder Hochleistungsrechner) und deren Anforderungen angepasst und optimiert werden. Neue grafische Metaphern zu finden, die hochdimensionale Daten anschaulich und verständlich präsentieren, ist ein weiteres Hauptziel aktueller Forschung. Die Entwicklung neuer Methoden passiert meist im Kontext, beziehungsweise an der Schnittstelle zu verschiedenen Anwendungsdisziplinen. Dabei unterscheiden sich deren Anforderungen miteinander deutlich. Zunehmend rückt auch

Was ist Was???

Mit **Datenvisualisierung** werden Techniken bezeichnet, die benutzt werden, um Erkenntnisse aus Daten mit Hilfe visueller Darstellung zu erlangen. Damit ist sie integraler Bestandteil der Datenwissenschaften. Grundlegende Visualisierungstechniken wie Koordinatensysteme oder Balken- und Kuchendiagramme wurden bereits im 17. und 18. Jahrhundert entwickelt. Das Hauptziel der Visualisierung ist, große Datenmengen, wie sie heute in vielen Bereichen anfallen (z.B. Sensordaten, Simulationsdaten und Nutzerdaten), besser zu verstehen und komplexe Zusammenhänge innerhalb der Daten zu erkennen. Dabei werden die Daten in Grafiken überführt und dem Data Scientist in einer abstrakten Form präsentiert.

Die Datenvisualisierung hat sich in den letzten Jahrzehnten als eigener Forschungszweig etabliert. Relativ neue Einsatzbereiche für Datenvisualisierung sind u.a. Digital Humanities, in denen Visualisierung in Kombination mit Methoden der Computerlinguistik zur Analyse großer Textbestände verwendet wird, sowie der Einsatz von Visualisierung zum Verständnis von maschinellem Lernen.



Die Visualisierung der Daten von Standard&Poor's zeigt 500 Aktienwerte unterteilt in Sektoren und Unternehmen. (Bild: Görtler, J., Schulz, C., Weiskopf, D., & Deussen, O. (2017). Bubble treemaps for uncertainty visualization. *IEEE transactions on visualization and computer graphics*, 24(1), 719-728)

die Effektivität der Visualisierungen in den Vordergrund, das heißt, eine Visualisierungstechnik sollte den Anwendenden einen Mehrwert bei der Datenanalyse bieten. Hierfür werden unter anderem Aspekte aus Kognitions- und Perzeptionsforschung berücksichtigt und Methoden in Benutzerstudien evaluiert.

Aktiendaten nach Wirtschaftssektor

Die Visualisierung rechts oben zeigt den S&P 500 Index (Aktiendaten einer Woche) in Form einer Bubble Treemap. Dabei werden die Aktien durch Kreise repräsentiert, wobei deren Fläche proportional zum durchschnittlichen Schlusskurs ist. Zusätzlich wird die Standardabweichung anhand der Umrisse visualisiert. Die Gruppierung ergibt sich aus den Wirtschaftssektoren. Die Visualisierung hilft beispielsweise dabei, einen Sektor mit geringer Unsicherheit zu finden (grün), sowie einen mit hoher Unsicherheit (blau), inklusive des Unternehmens, welches dafür hauptverantwortlich ist.

Datenvisualisierung wird bereits in vielen Bereichen eingesetzt und beeinflusst täglich unseren Blick auf die Welt. Mit immer größeren Datenmengen wird es zunehmend wichtiger diese effektiv und effizient zu analysieren, um im wissenschaftlichen als auch Business-Kontext Daten besser zu verstehen und neue Erkenntnisse zu gewinnen.

Valentin Bruder ist wissenschaftlicher Mitarbeiter am Visualisierungsinstitut der Universität Stuttgart (VISUS). Seine Forschung konzentriert sich auf die Bewertung, Modellierung und Vorhersage der Performanz von Visual-Computing-Systemen. Zur Zeit arbeitet er im SFB Transregio 161 am Projekt A02 zu Quantifying Visual Computing Systems.
Web: <https://vbruder.github.io>