

5.4 Datenvisualisierung

Abstract: Die visuelle Projektion von heterogenen (z. B. Forschungs-)Daten auf einer 2-dimensionalen Fläche, wie etwa einem Bildschirm, wird als Datenvisualisierung bezeichnet. Datenvisualisierung ist ein Oberbegriff für verschiedene Arten der visuellen Projektion. In diesem Kapitel wird zunächst der Begriff definiert und abgegrenzt. Der Fokus des Kapitels liegt auf Informationsvisualisierung und Visual Analytics. In diesem Kontext wird der Prozess der visuellen Transformation vorgestellt. Es soll als Grundlage für eine wissenschaftlich valide Generierung von Visualisierungen dienen, die auch visuelle Aufgaben umfassen. Anwendungsszenarien stellen den Mehrwert der hier vorgestellten Konzepte in der Praxis vor. Der wissenschaftliche Beitrag liegt in einer formalen Definition des visuellen Mappings.

Einleitung

Datenvisualisierung umfasst jegliche Art der visuellen Repräsentation von digitalen Daten in computerbasierten Systemen. Auch wenn sogenannte Infografiken nicht immer eine computerbasierte Projektion aufweisen, können diese ebenfalls der Datenvisualisierung zugeordnet werden. Somit ist Datenvisualisierung ein Oberbegriff für die visuelle Projektion von Daten auf einer 2-dimensionalen Ebene, auch wenn die Darstellung einen 3-dimensionalen Raum simuliert. Für die Klassifikation von Datenvisualisierung sind zum einen die zugrundeliegenden Daten und die Art der Projektion von Bedeutung. Eine Projektion eines Realweltobjekts, wie etwa des menschlichen Herzens mittels Computertomographie, ist eine virtuelle Darstellung, die oft als „Scientific Visualization“ bezeichnet wird.¹ Das Ziel hierbei ist durch computerbasierte grafische Simulationen neue oder unbekannte Erkenntnisse über einen bestimmten meist wissenschaftlichen Sachverhalt zu erlangen.² Werden dagegen abstrakte Daten visualisiert, etwa Text, Bilder, Ontologien etc., spricht man im Allgemeinen von „Informationsvisualisierung“.³ Der bedeutende Unterschied liegt nicht nur in den Daten, die keine reine Realweltobjekte mehr sind, sondern auch in der Art der visuellen Projektion. Sie visualisiert abstrakte Daten unter besonderer Berücksichtigung der menschlichen Kognition und Wahrnehmung und ermöglicht per Definition menschliche Interaktionen mit der visuellen Repräsentation, um die

1 Vgl. West 1999, 15.

2 Vgl. Earnshaw und Wiseman 1992, 1–7.

3 Vgl. Card, Mackinlay und Shneiderman 1999, 7.

Kognition zu stärken, neue Erkenntnisse aus den Daten zu erhalten und diverse analytische Aufgaben durchzuführen. Die Visualisierung abstrakter Daten kann weiter klassifiziert werden, etwa unter Berücksichtigung des Grades der menschlichen bzw. der maschinellen Verarbeitung der visuellen Transformation bzw. der grafischen Projektion. Eine solche Klassifikation wurde von Kohlhammer et al. aufgestellt.⁴ Diese unterscheiden zwischen Information Design (Infografik), Informationsvisualisierung, Semantik Visualisierung, Visual Analytics und den meist nicht-visuellen „Knowledge Discovery in Databases“.

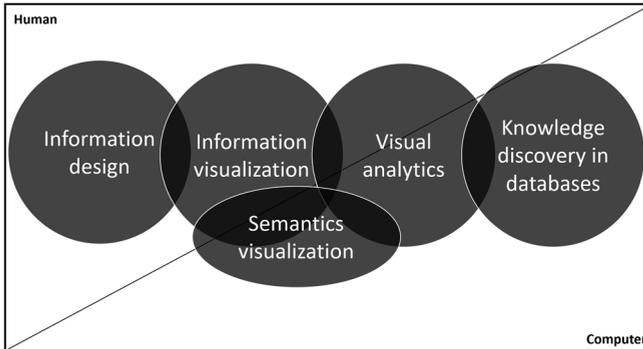


Abb. 1: Visualisierungsklassifikation nach Kohlhammer et al.⁵

Dieses Kapitel geht auf Informationsvisualisierung und somit auch auf Visual Analytics ein, die allein aufgrund der enormen Menge an Daten eine essentielle Rolle spielen. Forschungsdaten werden meist genau mit diesen Methoden der Datenvisualisierung verarbeitet und dargestellt. Zunächst wird eine kurze Abgrenzung und Definition der Informationsvisualisierung vorgestellt mit dem Ziel Informationsvisualisierung und Visual Analytics besser differenzieren zu können. Anschließend werden die Kernaspekte, wie etwa Daten und Datenklassifikation, visuelle Repräsentation und visuelle Aufgaben eingeführt. Diese sollen dazu befähigen, ein grundlegendes Verständnis zur Erzeugung von Informationsvisualisierungen zu erlangen. Im weiteren Verlauf werden Anwendungsbeispiele für verschiedene Szenarien gegeben, um den heterogenen Einsatz der Visualisierungen nachvollziehen zu können.

⁴ Vgl. Kohlhammer et al. 2012, 85–87.

⁵ Vgl. Kohlhammer et al. 2012, 85.

1 Informationsvisualisierung

Informationsvisualisierung ist die wohl am meisten verbreitete Form der Datenvisualisierung. Eine klare Abgrenzung wurde bereits 1999 von Card et al. eingeführt, die Informationsvisualisierung als computer-basierte, interaktive visuelle Repräsentation von abstrakten Daten zur Stärkung der Kognition definierten⁶. Ein wesentlicher Unterschied zur Datenvisualisierung besteht hier in der Einführung des Terms *abstrakt*, der eine offensichtliche räumliche Projektion der Daten auf einer 2-dimensionalen Fläche ausschließt. Ohne eine solche räumliche Zuordnung, besteht die Herausforderung darin, die Daten in eine ausdrucksstarke visuelle Repräsentation zu überführen. Die Stärkung der Kognition, definierten sie als Erwerb und Nutzung des menschlichen Wissens. Diese soll zu Erkenntnissen (insights) führen, etwa durch Erkundung, Analyse, Entscheidungsfindung oder Erläuterung.⁷ Um diesen Prozess zu beschreiben, führten sie das Referenzmodell der Informationsvisualisierung ein (siehe Abb. 2).

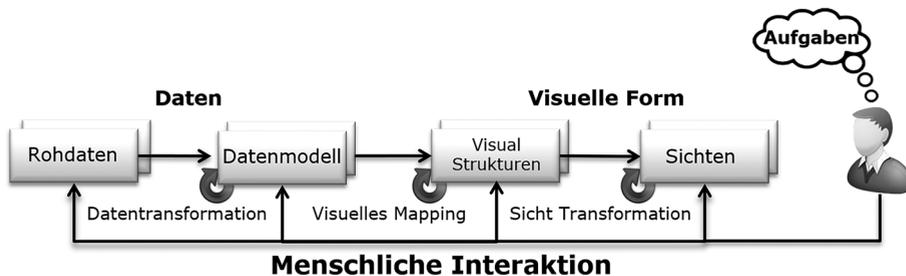


Abb. 2: Referenzmodell der Informationsvisualisierung nach Card et al.⁸

Das Referenzmodell sieht eine Transformation von Rohdaten bis hin zu interaktiven Visualisierungen vor, die hier als Sichten (Views) dargestellt sind. Das Modell sieht drei Transformationsschritte vor. Die Datentransformation überführt die Daten in eine für Visualisierung adäquate Form. Hier finden meist lernende und statistische Verfahren Einsatz. So können etwa aus unstrukturiertem Text mittels Latent Dirichlet Allocation⁹ Topics extrahiert werden.¹⁰ Die Datenqualität ist hierbei entscheidend. Auch die Extraktion von Variablen, die eine aspektorientierte Visualisierung erlauben, etwa temporale oder geographische Visualisierungen etc., spielen eine

⁶ Vgl. Card, Mackinlay und Shneiderman 1999, 7.

⁷ Vgl. Nazemi 2016, 4; Card und Mackinlay 1997, 6.

⁸ Vgl. Card, Mackinlay und Shneiderman 1999, 17.

⁹ Vgl. Blei, Ng und Jordan 2003, 996–1002.

¹⁰ Vgl. Nazemi et al. 2015b, 6.

wesentliche Rolle. Die Datenstruktur ist in der Stufe des visuellen Mappings ausschlaggebend für die Erzeugung der visuellen Struktur. Sie stellt die zugrundeliegenden Daten idealerweise unter Berücksichtigung der Variablen und einer für den Menschen gut wahrnehmbaren visuellen Repräsentation dar.¹¹ Die letzte Transformationsstufe ermöglicht die Interaktion der Benutzer mit der graphischen Repräsentation und auch mit jedem Zwischenschritt des Modells. Dabei sind die Aufgaben der Benutzer von besonderer Bedeutung. So sollen je nach Bedarf auch etwa die Daten verändert (reduziert/erweitert) oder alternative visuelle Strukturen ausgewählt werden können.

Visual Analytics verbindet Informationsvisualisierung mit automatischen Analysetechniken, um ein effektives Verständnis, Schlussfolgerung und Entscheidungsfindung zu ermöglichen.¹² Demnach erlaubt es im direkten Vergleich zu Informationsvisualisierung auch die Interaktion mit den Transformationsschritten und wird definiert als die Wissenschaft der analytischen Schlussfolgerung durch interaktive Visualisierungen.¹³ Keim et al. haben dazu ein Referenzmodell¹⁴ erstellt, das über die Jahre verschiedene Revisionen und Erweiterungen je nach Anwendungsfall erhielt.¹⁵ Eine klare Unterscheidung zur Informationsvisualisierung kann demnach durch (1) die Ziele, die sich auf analytische Aufgaben fokussieren, (2) eine direkte Kopplung der Methoden der automatischen Analyse und Datenmodellierung und (3) die Visualisierung sehr großer Datenmengen entstehen.

Zusammenfassend ist festzustellen, dass drei wesentliche Aspekte das Design und die Entwicklung von Informationsvisualisierungen besonders beeinflussen: (1) Daten sind von besonderer Bedeutung, denn diese sind die Grundlage zur Wahl der visuellen Struktur und der visuellen Variablen, etwa Farbe oder Größe, (2) die Visualisierung selbst, die im Idealfall eine einfache Interpretation der Daten ermöglicht, und (3) die zu lösenden Aufgaben, die die Wahl der Visualisierung stark beeinflussen.

2 Daten und Datenklassifikation

Der Ausgangspunkt jeglicher Visualisierungen sind die zugrundeliegenden Daten.¹⁶ Dabei kann zwischen Datentyp, Daten-Dimensionalität und Art der Daten unterschieden werden (siehe Abb. 3).

¹¹ Vgl. Nazemi 2016, 111–114.

¹² Vgl. Keim 2010, 7.

¹³ Vgl. Thomas und Cook 2005, 4.

¹⁴ Vgl. Keim 2010, 10.

¹⁵ Vgl. Stoffel et al. 2014, 1604; Nazemi 2018, 187; El-Assady et al. 2020, 1002.

¹⁶ Vgl. Keim 2010, 3; Keim 2001, 1–5; Card und Mackinlay 1997, 1; Card, Mackinlay und Shneiderman 1999, 3.

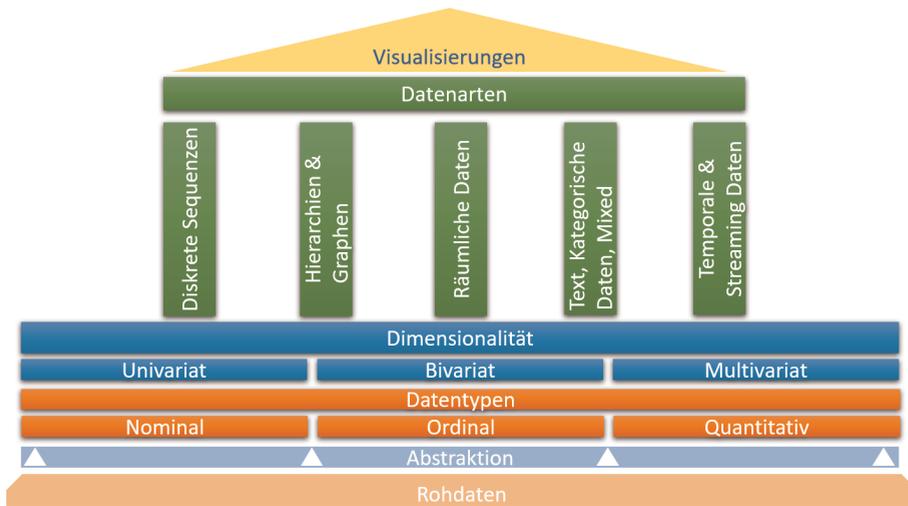


Abb. 3: Schematische Darstellung der Zusammenhänge zwischen verschiedenen Datentypen, der Dimensionalität und Datenarten in Beziehung zu Rohdaten

Datentypen beziehen sich darauf, ob in den Daten eine natürliche Ordnung der extrahierten oder bereits vorhandenen Variablen vorliegt, ob sie mit lexikalischen, semantischen etc. Transformationen generiert werden können oder ob eine solche Ordnung nicht vorhanden und möglich ist. In der Informationsvisualisierung werden folgende Datentypen unterschieden:¹⁷

- Nominal: Daten ohne Ordnung (Operatoren: = und \neq)
- Ordinal: Daten besitzen eine natürliche Ordnung (Operatoren: < und >)
- Quantitativ: Numerische Werte mit natürlicher Ordnung und der Möglichkeit der arithmetischen Rechenoperationen (Operatoren: +, -, *, /)

Datentypen, die keine natürliche Ordnung besitzen, sind *nominal*.¹⁸ Sie dienen zur Unterscheidung von Entitäten z. B. in Form von Kategorien oder Namen. Dabei soll beachtet werden, dass Hilfsvariablen und statistische Verfahren dazu dienen können eine Ordnung zu schaffen, etwa eine lexikalische Ordnung für Namen oder die physikalische Ordnung der Farben. Datentypen, die eine natürliche Ordnung besitzen, sind *ordinal*,¹⁹ etwa eine kategorisierte Tagestemperatur in heiß, warm und kalt. Dabei können ordinale Datentypen, binär (0 und 1), diskret oder kontinuierlich

¹⁷ Vgl. Card und Mackinlay 1997, 1; Card, Mackinlay und Shneiderman 1999, 12.

¹⁸ Vgl. Stevens 1946, 678.

¹⁹ Vgl. Ward, Grinstein und Keim 2010, 46.

sein²⁰ und können mit den Operatoren < und > in genau einem Vorgänger und einem Nachfolger unterteilt werden. Numerische Werte, die eine arithmetische Operation zulassen werden generell als *quantitativ* bezeichnet und weisen ebenfalls eine durch die Zahlen gegebene natürliche Ordnung auf. Diese Arten von Daten können anhand ihrer Wertigkeit geordnet werden (z. B. natürliche Zahlen). Die Einteilung (nominal, ordinal, quantitativ) ist nicht unumstritten,²¹ kann jedoch durch vielfache Nutzung und starke Verbreitung in der Informationsvisualisierung als wissenschaftlich akzeptiert angesehen werden.

Jeder Datensatz wiederum kann mittels der Dimensionalität kategorisiert werden. So können Datensätze in 1-dimensional, 2-dimensional oder multi-dimensional unterschieden werden.²² Dabei beziehen sich die Dimensionen auf die Anzahl der Variablen in den Daten.

Tab. 1: Kategorisierung der Daten mittels unterschiedlicher Dimensionen²³

Daten Dimensionalität	
1-dimensional (univariat)	Lineare Datentypen
2-dimensional (bivariat)	Planare Datentypen
Multi-dimensional (multivariat)	Daten mit mehr als drei Dimensionen, auch multivariate Daten genannt

2.1 Eindimensionale Daten

Eindimensionale Daten können in diskreten Sequenzen, Text, kategorischen Werten oder in temporalen und Streaming Daten enthalten sein und verfügen über eine Variable. Siehe Abb. 4 für eine eindimensionale Ereigniskette mit natürlicher Ordnung (gegeben durch die Jahreszahl). Nominale Listen ohne Ordnung sind ebenfalls Teil der Daten. Diskrete Strukturen, z. B. Abfolgen – kodierte Gene innerhalb des Genoms – können ebenfalls eindimensional vorliegen. Die Visualisierung geschieht über die relative Position des Gens. Reiner Text ohne Vorverarbeitung ist ebenfalls als nominal eindimensional einzustufen.

²⁰ Vgl. Ward, Grinstein und Keim 2010, 46.

²¹ Vgl. Velleman und Wilkinson 1993, 5–7.

²² Vgl. Shneiderman 1996, 337–338; Ward und Keim 2010, 382; Keim, Panse und Sips 2005, 3–4.

²³ Vgl. Keim, Panse und Sips 2005, 3–4.

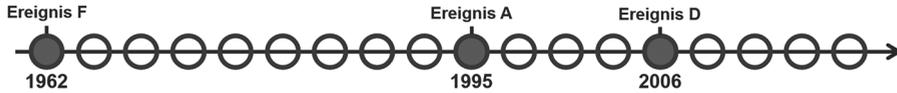


Abb. 4: Eindimensionale Ereigniskette

2.2 Zweidimensionale Daten

Zweidimensionale Daten haben genau zwei Variablen, die miteinander verknüpft sind. Dabei wird häufig eine abhängige Variable zu einer unabhängigen Variablen in Korrelation gesetzt. Die Werte können z. B. mittels zwei Spalten innerhalb einer Tabelle repräsentiert werden. Die Variablen selbst können wiederum ordinal, nominal oder quantitativ sein. Typische Beispiele sind temporale Daten, Streaming Daten oder räumliche planare Daten (siehe Abb. 3). Zweidimensionale Daten lassen sich vereinfacht in X-Y-Plots darstellen. Die Visualisierung 2-dimensionaler Daten gestaltet sich recht einfach, etwa mit Balkendiagrammen oder Line-Charts (siehe Abb. 5).

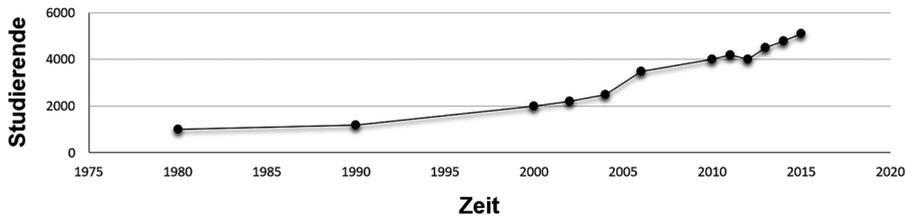


Abb. 5: Zeitreihen als Beispiel zweidimensionaler Datensätze

2.3 Multidimensionale Daten

Multidimensionale Datensätze oder auch multivariate Datensätze enthalten drei und mehr Variablen. Beispiele finden sich als temporale Daten, Streaming Daten, Hierarchien und Graphen oder als räumliche Daten. Beispielsweise existieren innerhalb einer Fertigungsstraße hunderte von Sensoren, die synchronisiert Daten erfassen, etwa Temperaturwerte oder Bewegungen. Das Abfrageergebnis kann hunderte Spalten umfassen. Die Variablen der Spalten können wieder ordinal, nominal oder quantitative sein. Auch wenn nur drei Dimensionen vorliegen, sollten 3-dimensionale Projektionen vermieden werden, diese führen oft zu Interpretationsschwierigkeiten. Statt der Einführung eines X-Y-Z-Plots kann der Einsatz visueller Variablen in X-Y-Plots die Wahrnehmung stark vereinfachen und die Interpretation der Daten

erleichtern (siehe Abb. 6). Eine Erhöhung der Anzahl der visuellen Variablen geht mit einer Komplexitätssteigerung innerhalb der Visualisierung einher. Multivariate Daten mit mehr als drei Dimensionen lassen sich oft sehr gut mit Matrizenvisualisierungen darstellen. Dabei stehen zur besseren Übersicht jeweils zwei Variablen in Korrelation (siehe Abb. 6). Es existieren verschiedene Ansätze der Interaktion mit multidimensionalen Visualisierungen,²⁴ die die Lösung analytischer Aufgaben ermöglichen.

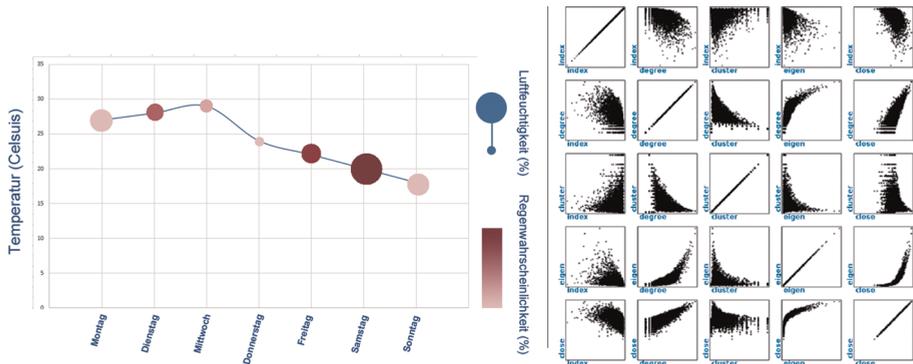


Abb. 6: Visualisierung multivariater Daten, links: Visualisierung von vier Dimensionen durch Nutzung von visuellen Variablen. Rechts: Visualisierung von fünf Dimensionen mit einem Scatterplot-Matrix nach Viau et al.²⁵

Die Klassifikation der Daten in Dimensionen wird oft gleichgestellt mit der Datenartklassifikation.²⁶ So wird oft im Kontext der Datendimensionen auch in Graphen, Hierarchien, Text oder Netzwerke unterschieden.²⁷ Diese Differenzierung ist für das visuelle Mapping von Bedeutung. Liegt etwa eine klare Hierarchie in den Daten vor, ist evtl. eine andere Visualisierung besser geeignet.

3 Visuelles Mapping

Visuelles Mapping beschreibt die Projektion der zugrundeliegenden Daten auf möglichst effektive visuelle Repräsentationen. Es bezieht sich auf die Fragestellung „wie“ Daten visualisiert werden sollten. Dabei spielen zwei Faktoren der Daten eine

²⁴ Vgl. Inselberg und Dimsdale 1990, 361–378; Gahegan 1998, 43–56; May und Kohlhammer 2008, 911–918; May, Davey und Kohlhammer 2010, 985–994; Viau et al. 2010, 1100–1108.

²⁵ Vgl. Viau et al. 2010, 1100.

²⁶ Vgl. Keim, Panse und Sips 2005, 2–3.

²⁷ Vgl. Keim, Panse und Sips 2005, 2–3; Nazemi 2016, 142–143.

Schlüsselrolle: (1) die Art der Daten und somit die Dimensionen bzw. etwaige Strukturen in den Daten und (2) die Datentypen, die vereinfacht in nominal, ordinal und quantitativ differenziert werden können. Das visuelle Mapping ist somit ein zweistufiger Prozess, das auf die Arbeiten von Bertin zurückgeht²⁸ und erweitert wurde.²⁹ Dabei wird zunächst die Positionierung der graphischen Objekte auf einer 2-dimensionalen Ebene je nach Datendimension oder der vorliegenden Struktur vorgenommen (Imposition). Diese Positionierung wird häufig als „Placement“ oder „Layouting“ bezeichnet.³⁰ Das Layouting stellt somit das „Skelett“ einer Visualisierung dar. Dabei können bei multivariaten Daten oder komplexen Strukturen Layout- oder Placement-Algorithmen durchaus kombiniert werden. Als Beispiel soll hier die in Abb. 6 dargestellte Scatterplot-Matrix dienen. Diese ist eine Kombination aus einer Matrix-Visualisierung und dem Scatterplot. Diese Layouts können sowohl nebeneinander liegen (juxtaposing Layouts), wie in diesem Fall, aber auch übereinander (superimposing Layouts).³¹

Datentypen sind besonders für die Wahl der retinalen bzw. visuellen Variablen von großer Bedeutung. Frühe Arbeiten hierzu gehen ebenfalls auf Bertin zurück,³² die für die unterschiedlichen Datentypen die entsprechenden Variablen bezüglich ihrer Eignung ordnen. Dabei ist stets die Position, die durch das Layout definiert wird für jeden Datentypen am besten geeignet (siehe Abb. 7).

	Quantitativ	Ordinal	Nominal
besser geeignet ↑	Position	Position	Position
	Länge	Farbdichte	Farbe
	Winkel	Sättigung	Farbdichte
	Neigung	Farbe	Sättigung
	Fläche	Länge	Form
	Farbdichte	Winkel	Länge
	Sättigung	Neigung	Winkel
	Farbe	Fläche	Neigung
weniger geeignet ↓	Form	Form	Fläche

Abb. 7: Retinale Variablen sortiert nach ihrer Eignung für verschiedene Datentypen (adaptiert nach Mackinlay³³)

²⁸ Vgl. Bertin 1983, 6.

²⁹ Vgl. Mackinlay 1986, 125; Nazemi 2016, 113–114.

³⁰ Vgl. Nazemi 2016, 220–223.

³¹ Vgl. Nazemi 2016, 55, 222.

³² Vgl. Bertin 1983, 42–44.

³³ Vgl. Mackinlay 1986, 125.

Da die Position für alle Datentypen am besten geeignet ist und Bertin³⁴ eine Differenzierung zwischen Positionierung (Imposition) und den retinalen Variablen (Implantation) propagiert hat, wird die Position hier nicht weiter als Bestandteil der retinalen Variablen betrachtet. Hilfreicher ist eine Differenzierung zwischen Layout und Präsentation. Das Layout stellt somit die Positionierung der graphischen Objekte auf dem Bildschirm, deren Relation zum Bildschirm und zueinander dar, wobei Präsentation die retinalen Variablen ohne Position enthält.

Visualisierung in Form des visuellen Mappings kann somit als Kreuzprodukt aus Layout und Präsentation beschrieben werden. Es ermöglicht sowohl das Übereinanderlegen (superimposing) von visuellen Layouts als auch eine Darstellung nebeneinander (juxtaposing). Das Layout wird anhand der Dimensionalität und der Datenstruktur ermittelt. So können beispielsweise bivariate Daten mit Hilfe von Scatter-, Line-, Area-, Bar-Charts etc. visualisiert werden. Multivariate Daten können z. B. mit Line Plots, Stacked Bars, Sliced River, Stacked River, Theme River etc. dargestellt werden, vgl. Abb. 8 für eine Auswahl an möglichen Layouts und Präsentationen.

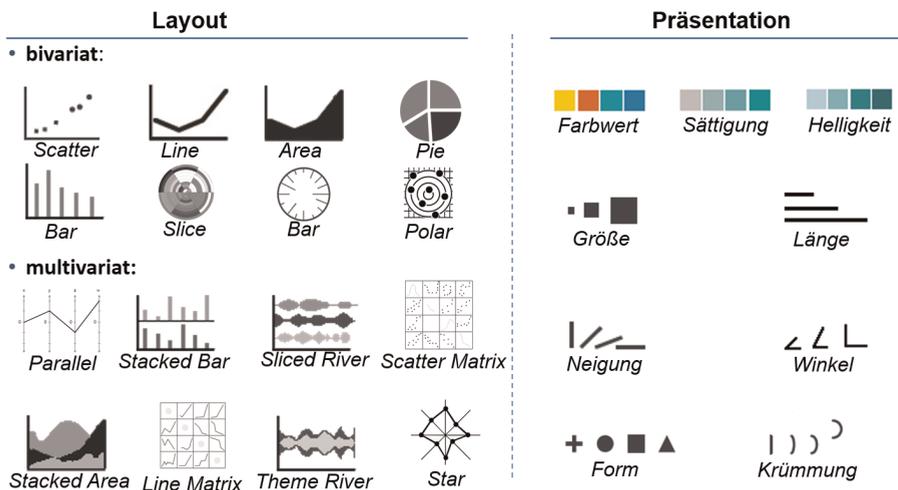


Abb. 8: Unterteilung des visuellen Layouts als Tupel aus Layout und Präsentation zur gezielten und effektiven Gestaltung von Visualisierungen

³⁴ Vgl. Bertin 1983, 50–52, 79, 189.

Die Präsentation setzt sich zusammen aus den retinalen Variablen wie etwa *Farbwert, Sättigung, Helligkeit, Größe, Länge, Neigung, Winkel, Form* und der *Krümmung*. Eine Visualisierung (V) kann als Kreuzprodukt aus Layout (L) und Präsentation (P) wie folgt definiert werden:

$$V = L \times P$$

Somit wird z. B. bei vierdimensionalen Daten, eine vierdimensionale Projektion in den zweidimensionalen Raum möglich. In Abb. 6 (links) liegt die Zeit auf der X-Achse und der Temperaturwert auf der Y-Achse, zusätzlich wird die Luftfeuchtigkeit per Größe und die Regenwahrscheinlichkeit per Sättigung dargestellt. Mackinlay präzierte die einzelnen Präsentationsmöglichkeiten und ordnete diese nach Genauigkeit bzw. Aussagekraft (siehe Abb. 7).³⁵

Dies wird nützlich, um möglichst aussagekräftige Visualisierungen für verschiedene Datentypen zu erstellen. Eine Visualisierung mit der bestmöglichen Aussagekraft für verschiedene Daten der Datentypen D ist somit definiert als:

$$V = L \times P = \{(l_i, f(d_i)) | i \in N \wedge l \in L \wedge d \in D\}$$

Dabei liefert die Funktion $f: D \rightarrow P$ eine bestmögliche Präsentation P für einen bestimmten Datentyp D . Gegeben sei als Beispiel ein X-Y-Plot mit einem Datensatz bestehend aus quantitativen, ordinalen und nominalen Daten ([0...99, klein...groß, ProduktA...ProduktC]). Die quantitativen und nominalen Daten werden der X, sowie der Y-Achse zugeordnet, somit werden die Datenwerten mittels der Position abgebildet, so bleibt für die ordinalen Daten nur noch die Darstellung über die Dichte und für die nominalen Daten die Darstellung über verschiedene Farbwerte, um die Aussagekraft der Visualisierung dieses Beispiels zu maximieren.

4 Visuelle Aufgaben

Eine Interaktion durch die Benutzenden mit einer Visualisierung dient der Beantwortung einer bestimmten Frage oder Hypothese. Einer Interaktion liegt dabei eine Fragestellung zugrunde, die als zu lösende Aufgabe verstanden werden kann. Diese Fragestellung wird auch als visuelle Aufgaben bezeichnet und bezieht sich auf die Frage, „warum“ soll etwas auf diese Art visualisiert werden. Erstmals formalisierte Bertin diese Fragestellung.³⁶ Bertin konstatierte, eine Frage ist aufgeteilt in den Fragentypen (Question Type) und das Abstraktionslevel (Level Of Reading). Die In-

³⁵ Vgl. Mackinlay 1986, 125.

³⁶ Vgl. Bertin 1983, 141, 178.

formation, die gefunden werden muss, bestimmt der Fragentyp.³⁷ Zusätzlich existieren drei Abstraktionslevel zu jedem Typ:

- Elementar (Elementary)
- Intermediär (Intermediate)
- Übergreifend (Overall)

Elementare Fragen beziehen sich auf einen Zeitpunkt oder einen Datenwert. Bei intermediären Fragen umfasst das Ergebnis eine Gruppe von Daten, jedoch bezieht sich die Frage nicht auf den kompletten Datensatz. Wohingegen übergreifenden Fragen den kompletten Datensatz betreffen. Dazu zwei Beispiele visueller Aufgaben:³⁸

- Wie hoch ist der Preis einer Aktie X an einem bestimmten Datum?
- Zu welchem Datum wurde ein bestimmter Preis für eine Aktie X erreicht?

Frage 1 ist eine elementare Frage, da die Frage auf ein Datum bzw. einen Datenwert abzielt. Als übergreifend kann Frage 2 bezeichnet werden, da die Frage den kompletten Datensatz umfasst. Andrienko und Andrienko unterscheiden hier nochmals in grundlegende Aufgaben (Elementary Tasks) und zusammenfassende Aufgaben (Synoptic Tasks³⁹). Aigner et al. fasst die Erkenntnisse von Andrienko et al. in einer Übersicht zusammen (vgl. Abb. 9).⁴⁰ Grundlegende Aufgaben beziehen sich hier auf einen Punkt oder Datenwert. Nachschlagen (Lookup) bezeichnet hier das Suchen nach einem bestimmten Datenwert. In der Beziehungssuche (Relation Seeking) werden Aufgaben zusammengefasst, die eine Relation zwischen den Entitäten offenlegen sollen, während vergleichende Aufgaben (Comparison) die Charakteristiken einzelner Datenpunkte oder Gruppen gegenüberstellt. Sowohl das Nachschlagen von Werten wie auch das Vergleichen kann sowohl direkt oder invers passieren. Dazu gibt Aigner et al. folgendes Beispiel für das Nachschlagen:⁴¹

- Direkt: What was the price of Google stocks on January 14?
- Invers: On which day(s) was the lowest stock price for Amazon in 2010?

³⁷ Vgl. Andrienko und Andrienko 2006, 50.

³⁸ Vgl. Bertin 1983, 141, 178; Andrienko und Andrienko 2006, 50.

³⁹ Vgl. Andrienko und Andrienko 2006, 136.

⁴⁰ Vgl. Aigner et al. 2011, 72–75.

⁴¹ Vgl. Aigner et al. 2011, 72–75.

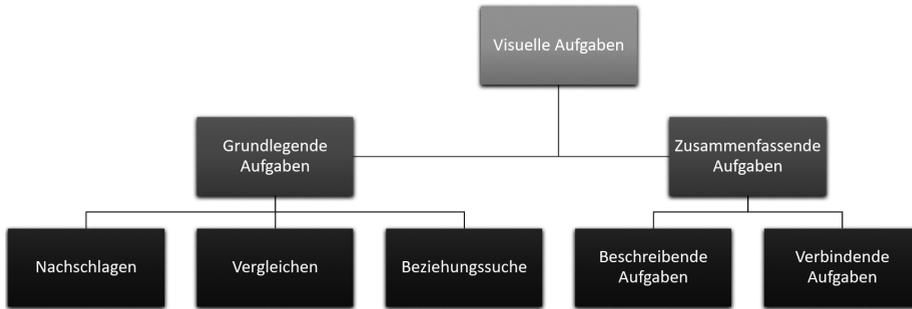


Abb. 9: Visualization Task Categorization, adaptiert von Aigner et al.⁴²

Zusammenfassende Aufgaben (Synoptic Tasks) bestehen aus beschreibende Aufgaben (Descriptive Tasks) und verbindende Aufgaben (Connective Tasks). Beschreibende Aufgaben geben Charakteristika oder Referenzen von Gruppen wieder, wohingegen verbindende Aufgaben zwei oder mehr Datensätze in Beziehung zueinander setzen. Aigner et al. machen hier folgende Beispiele:⁴³

- Beschreibend: What was the trend of Oracle stocks during January?
- Verbindend: Is the behavior of Nokia stocks influencing the behavior of Motorola stocks?

Munzner geht noch einen Schritt weiter und formalisiert die visuelle Aufgabe.⁴⁴ Eine visuelle Aufgabe besteht aus einer Aktion und einem Ziel.⁴⁵ Munzner führt dazu die gruppierenden Aktionen: Analysieren (Analyze), Suchen (Search) und Anfragen (Query) ein. Dazu werden zusätzlich die gruppierenden Ziele: alle Daten (All Data), Attribute (Attributes), Netzwerkdaten (Network Data) und räumliche Daten (Spatial Data) definiert. Die Analyse unterscheidet in vereinnnehmende Aktionen, wie das Entdecken und Präsentieren, sowie in produzierende Aktionen, wie das Annotieren oder die Aufnahme von Daten. In der Suche wird differenziert nach Nachschlagen, Durchsuchen, Lokalisieren und Explorieren. Als Anfrage ist definiert: Identifizieren, Vergleichen und Zusammenfassen.⁴⁶ Als mögliche Ziele auf der Gesamtheit der Daten wird das Erkennen von Trends, Ausreißern oder Features genannt. Innerhalb von Attributen können Verteilungen, Extremwerte, Abhängigkeiten, Korrelationen oder Ähnlichkeiten gefunden werden. In Netzwerkdaten können Topologien oder Pfade analysiert werden und in räumlichen Daten kann die Darstellung von Umrissen sinnvoll sein. Das durch Munzner erzeugte Framework⁴⁷ bietet

⁴² Vgl. Aigner et al. 2011, 74.

⁴³ Vgl. Aigner et al. 2011, 75.

⁴⁴ Vgl. Munzner 2014, 43–61.

⁴⁵ Vgl. Munzner 2014, 43.

⁴⁶ Vgl. Munzner 2014, 54.

⁴⁷ Vgl. Munzner 2014, 43.

die Möglichkeit der Erweiterung. So können weitere visuelle Aufgaben durch Hinzufügen von Aktionen und Zielen erzeugt werden und eine visuelle Aufgabe (T) kann demnach als Tupel von Aktion (A) und Ziel (Z) definiert werden.

$$T = (A, Z)$$

Anhand der vereinfachten Darstellung in Abb. 10 können Aufgaben identifiziert werden, die der Benutzer mit einer Visualisierung lösen möchte.

👉 Aktion

- Analyse:



Entdecken



Präsentieren



Vergleichen



Erkennen



Kategorisieren



Ableiten
- Suche:

	Ziel bekannt	Ziel unbekannt
Position Bekannt	Lookup 	Browse 
Position unbekannt	Locate 	Explore 

🎯 Ziel



Trends



Anomalien



Eigenschaften



Verteilung



Extrema



Abhängigkeit



Korrelationen



Ähnlichkeiten

Abb. 10: Definition einer visuellen Aufgabe (vereinfacht), mit Erlaubnis adaptiert von Munzner⁴⁸

5 Anwendungsszenarien

Informationsvisualisierung und Visual Analytics spielen in fast allen Bereichen eine essentielle Rolle. Dabei werden nicht nur sehr unterschiedliche Daten herangezogen, sondern auch sehr heterogene Aufgaben gelöst. Dieser Abschnitt verdeutlicht einige der möglichen Anwendungsszenarien zur Verdeutlichung des Mehrwerts solcher interaktiven Visualisierungen anhand einiger ausgewählter Anwendungsdomänen. Ein Großteil der Visualisierungen ist mit JavaScript umgesetzt, wobei hier Scalable Vector Graphics (SVG) eingesetzt wurden.

⁴⁸ Vgl. Munzner 2014, 43.

5.1 Visualisierung wissenschaftlicher Literaturdaten

Durch die massive Open-Access-Bewegung stellen heute bibliothekarische Daten eine wichtige Ressource für verschiedene Anwendungsbereiche dar. Insbesondere mit Methoden des maschinellen Lernens in Kombination mit interaktiven Visualisierungen, also Visual Analytics, lässt sich nicht nur nach bestimmten Publikationen, Personen oder Themen suchen, diese Daten erlauben zudem komplexe analytische Aufgaben. Als Beispiel soll die Erkennung von Trends und Technologien dienen. Basierend auf das bereits vorgestellte Referenzmodell der Informationsvisualisierung⁴⁹ wurde zunächst ein Transformationsmodell erstellt, das beginnend mit der Extraktion von Literaturdaten aus dem Web (Web-Mining) bis hin zu interaktiven Visualisierungen den gesamten Transformationsprozess abdeckt.⁵⁰ Die Analysemöglichkeiten wurden mit lernenden Modellen erweitert und ein neuer Algorithmus zur Trenderkennung entwickelt.⁵¹ Dabei wurde auch der Prozess erweitert und an den speziellen Fall der Literaturdaten angepasst (vgl. Abb. 11).

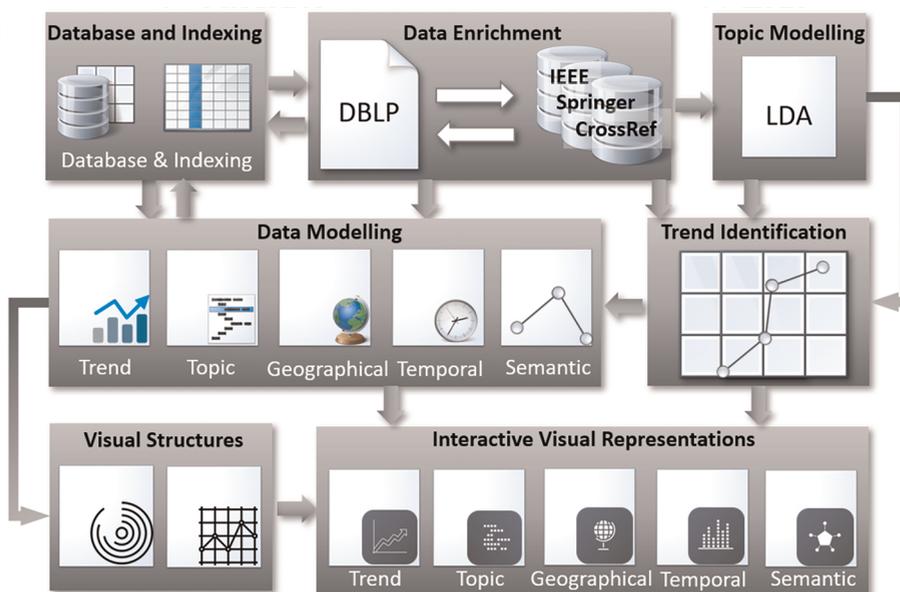


Abb. 11: Transformationsmodell für wissenschaftliche Literatur nach Nazemi und Burkhardt⁵²

⁴⁹ Vgl. Card, Mackinlay und Shneiderman 1999, 17.

⁵⁰ Vgl. Nazemi et al. 2015b, 3.

⁵¹ Vgl. Nazemi und Burkhardt 2019b, 194–200.

⁵² Vgl. Nazemi und Burkhardt 2019b, 194.

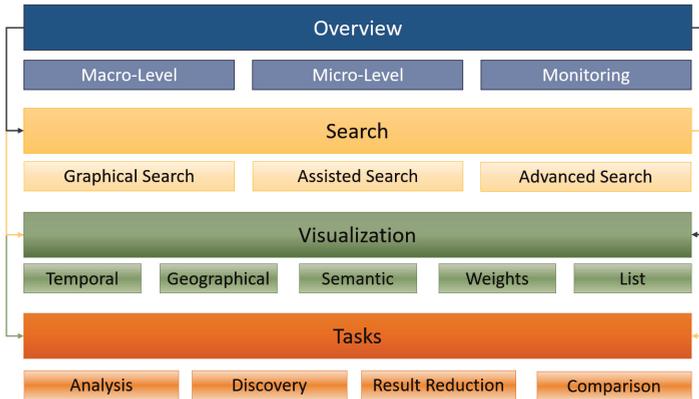


Abb. 12: Analyse-, Interaktions- und Visualisierungsmöglichkeiten zur visuellen Trendanalyse wissenschaftlicher Literatur⁵³

In Abb. 11 erkennt man, dass die Datenmodellierung aspektorientiert durchgeführt wurde. Dazu wurden insgesamt fünf Datenmodelle erstellt. Basierend darauf wurde ein Modell der Suche und Exploration eingeführt, das die unterschiedlichen Such-, Visualisierungs- und Interaktionsmöglichkeiten abbildet.⁵⁴ Basierend auf den Erkenntnissen aus dem Technologie- und Innovationsmanagement wurden hier Überblicke über den gesamten Datenbestand, über die Resultate einer Suche und weitere Explorationsmöglichkeiten eingeführt, die auch das Lösen analytischer Aufgaben erlauben (siehe Abb. 12).

Die visuelle Projektion erlaubt das Lösen sehr unterschiedlicher Aufgaben, die über das Recherchieren hinausgehen. Dazu wurden nicht nur sehr unterschiedliche Visualisierungen realisiert, die unterschiedliche Aspekte der Daten aufzeigen, wie etwa temporale oder semantische Visualisierungen. Es wurden diverse Interaktionskonzepte eingeführt, die den Benutzenden erlauben sogenannte „emerging Trends“ zu entdecken. Abb. 13 zeigt einige Visualisierungen und auch die Interaktionstechniken.

⁵³ Vgl. Nazemi und Burkhardt 2019a, 286.

⁵⁴ Vgl. Nazemi und Burkhardt 2019a, 285–288.

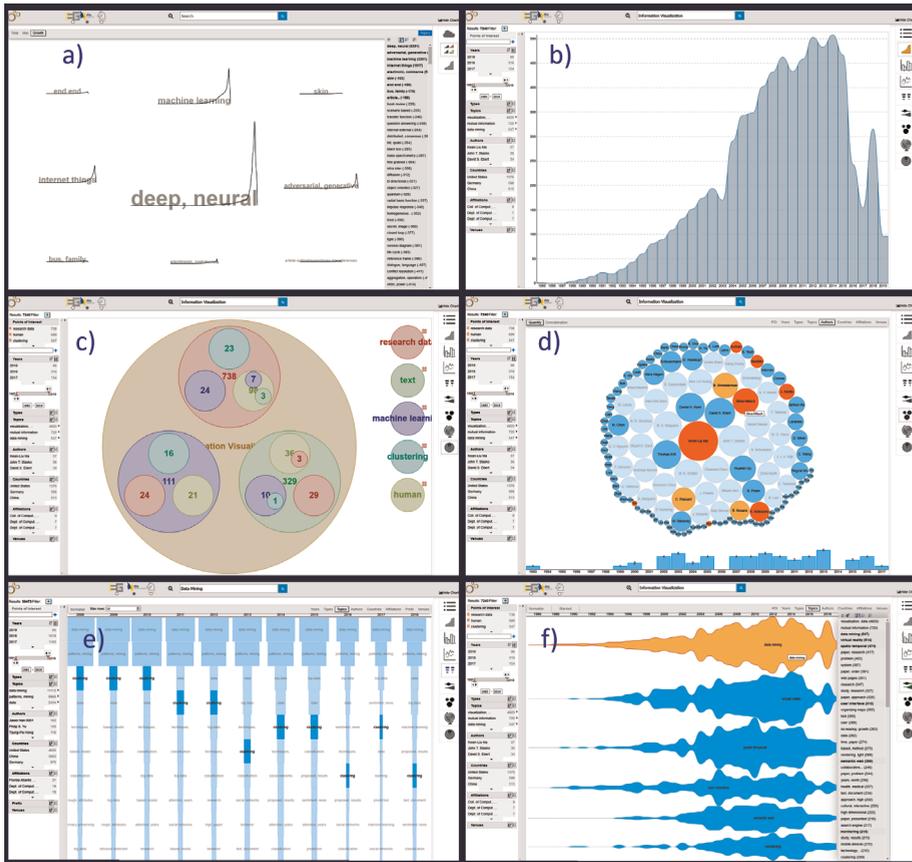


Abb. 13: a) Macro-Level Übersicht der „emerging Trends“ in der gesamten Datenbank; b) temporale Übersicht (Micro-Level) nach einer Suche; c) graphical search: graphische Suche nach eigenen Termen; d) temporale Übersicht der Subthemen zu einer Technologie; e) stacked-graph zur temporalen Übersicht extrahierter Themen; f) semantische Visualisierung etwa zur Koautoren-Relation etc.⁵⁵

5.2 Visualisierung von Government Data

Im Rahmen von Open Data Initiativen haben Behörden weltweit begonnen, Teile ihre Daten öffentlich bereitzustellen. Speziell im EU-Open-Data-Portal,⁵⁶ in dem auch sehr umfassende Daten von EuroStat⁵⁷ und EUR-Lex⁵⁸ einbezogen sind, lassen

⁵⁵ Vgl. Nazemi und Burkhardt 2019a, 287–293; Nazemi und Burkhardt 2019b, 196–198.

⁵⁶ S. <https://data.europa.eu/euodp/>. Letztes Abrufdatum der Internet-Dokumente ist der 15.11.2020.

⁵⁷ S. <https://ec.europa.eu/eurostat/>.

sich sehr große Mengen an öffentlichen Daten für diverse Anwendungsmöglichkeiten abfragen und in Systeme einbinden. Zudem ist davon auszugehen, dass die veröffentlichte Datenmenge aufgrund diverser Transparenzregelungen in allen Bereichen staatlichen Handelns weiter zunehmen.⁵⁹ Dabei gibt es eine Reihe von Anwendungsmöglichkeiten, darunter solche, die die Transparenz und Partizipation und somit die Qualität staatlicher Entscheidungen in beträchtlichem Maße unterstützen.⁶⁰ Grundsätzlich lassen sich diese in die drei aufeinander aufbauenden Stufen gliedern: (1) *e-Enabling*, (2) *e-Engaging* und (3) *e-Empowering*,⁶¹ wobei die unterste Stufe sich im Wesentlichen auf die reine Bereitstellung von Daten beschränkt, um eine grundsätzliche Teilhabe von Bürgern und Organisationen zu ermöglichen. Wenn basierend auf den bereitgestellten Daten eine Meinungsbildung stattfinden soll, etwa im Rahmen von öffentlichen Debatten über eine politische Agenda, so erfordert dies das *e-Engaging*. Dieses kann gestärkt werden, in dem man Bürgerinnen und Bürger sowie Organisationen zusätzlich motiviert, sich an der politischen Agenda und den Debatten zu beteiligen – diese Form bezeichnet man als *e-Empowering*. Insbesondere *e-Engaging* und *e-Empowering* liefert in der Regel zusätzliche Daten in Form von Stellungnahmen und Meinungen, die ihrerseits Veröffentlichung finden und im Rahmen der Visual Analytics genutzt werden können.

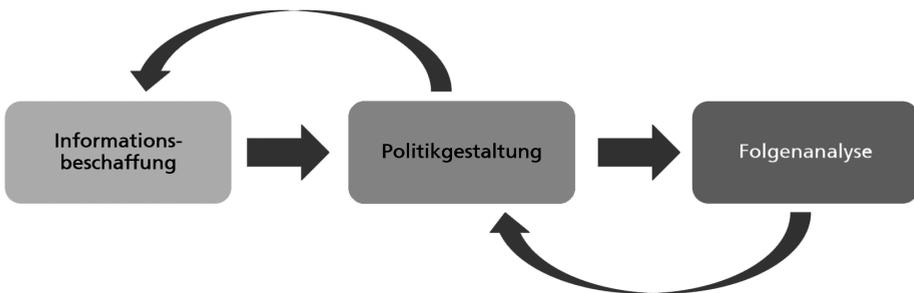


Abb. 14: Ein vereinfachter Prozess zur Modellierung von politischen Maßnahmen. Alle drei Stufen beinhalten heterogene Datenquellen, um die Analyse verschiedener Standpunkte, Meinungen und Möglichkeiten zu ermöglichen.⁶²

58 S. <https://eur-lex.europa.eu>.

59 Ein anschauliches Beispiel bietet etwa der Bereich der europäischen Chemikalienregulierung. Das dort seit dem Jahr 2016 eingeführte „lernende System“ sieht einen öffentlichen Zugang zu Stoffdaten, Stellungnahmen und Begründungen an diversen Stellen des Entscheidungsprozesses vor. Siehe mit weiteren Nachweisen Below 2018.

60 Siehe Für die Hintergründe, Gestaltungskriterien und Möglichkeiten der Transparenz und Partizipation am Beispiel der Chemikalienregulierung Below 2018.

61 Vgl. Macintosh 2004, 2–3; Burkhardt et al. 2014, 4–8.

62 Vgl. Kohlhammer et al. 2012, 85.

Wie bereits erwähnt, ist eine Grundvoraussetzung der Bürgerbeteiligung die Möglichkeit des Zugriffs auf politische Daten. Speziell in Bezug auf Visualisierungssysteme für politische Daten, lassen sich dabei drei elementare Stufen zum Einsatz von visuellen Systemen definieren:⁶³ (1) Die Informationsbeschaffung, etwa im Zuge der Recherche zu bestehenden Problemen oder zu allgemeinen gesellschaftlichen Herausforderungen, (2) die Politikgestaltung, eine Phase, die konkrete politische Maßnahmen definiert, und (3) die Folgenanalyse, die die Auswirkung von definierten politischen Maßnahmen ermittelt und prüft, ob eine Maßnahme das beabsichtigte Problem adressiert.⁶⁴ Abhängig von den aktuellen Analysephasen können unterschiedliche visuell-analytische Methoden sinnvoll sein.

Bei der Visualisierung von Government Daten lassen sich darüber hinaus formelle Daten von erhobenen bzw. maschinell extrahierten Daten unterscheiden. Formelle Daten sind dabei konkrete Beschreibungen etwa von Gesetzen, Maßnahmen oder gesellschaftspolitischen Zielen. Viele dieser formellen Daten werden beispielsweise in Amtsblättern oder Bundesanzeigern veröffentlicht und sind Ergebnisse politischer Entscheidungen. Für solche Daten ist eine Visualisierung oftmals schwierig (wie z. B. Abb. 15), da sie vergleichsweise sehr spezifisch und fachbezogen sind. Daher gibt es nur wenige spezialisierte Visualisierungsmöglichkeiten.⁶⁵ Für erhobene Daten gibt es in der Regel deutlich mehr Visualisierungen, da sie auf gängigen Formen basieren. Den Großteil stellen statistische Daten dar. Zahlen zur Bevölkerungsentwicklung, Migration, Bruttoinlandsprodukt oder diverse Wirtschaftsindikatoren lassen sich temporal visualisieren. Aber auch Abhängigkeiten zwischen verschiedenen Daten zu Entitäten wie zu Behörden oder konkreten Ansprechpartnern lassen sich teilweise abfragen und visualisieren.

63 Vgl. Kohlhammer et al. 2012, 84–85.

64 Ein Beispiel für Folgenanalysen ist die in Deutschland über § 44 Absatz 1 der Gemeinsamen Geschäftsordnung der Bundesministerien (GGO) rechtlich vorgeschriebene Gesetzesfolgenabschätzung (GFA). Sie befasst sich im klassischen Sinne mit der Aufgabe, wahrscheinliche Folgen und Nebeneffekte von Regelungsvorhaben (prospektive GFA), entstehenden (begleitende GFA) oder bestehenden Gesetzen (retrospektive GFA) zu ermitteln und diese zu beurteilen. Siehe zur Praxis der Gesetzesfolgenabschätzung Böhret und Konzendorf, 2001, S. 1 oder das Impact Assessment der Europäischen Union unter https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/impact-assessments_en. Ein konkretes Anwendungsbeispiel findet sich in Below 2018.

65 Vgl. Burkhardt und Nazemi 2018, 157–160.

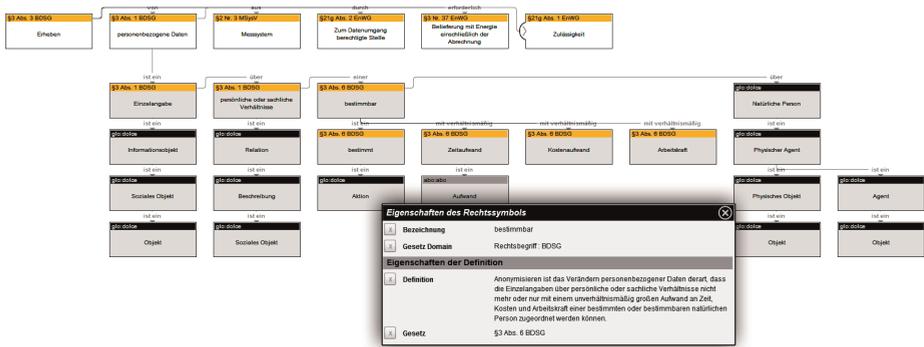


Abb. 15: Visualisierung von formalen Government Daten, wie z. B. Gesetze, sind oftmals sehr speziell und erfordern individuelle Ansätze⁶⁶

5.3 Visualisierung von verknüpftem Wissen

Ein spezielles Feld nimmt die Verknüpfung von Wissen ein.⁶⁷ Da durch die Verknüpfung über verschiedene Datenbanken hinweg, die Daten stetig anwachsen, ist deren Verwendung sehr weitreichend. Sie eignen sich für eine Vielzahl an Auswertungen und haben ebenfalls eine Vielzahl an potentiellen Einsatzszenarien. Am bekanntesten ist dabei das Semantic Web, wie es etwa durch DBpedia⁶⁸ umgesetzt ist. Dabei handelt es sich bei DBpedia im Wesentlichen um eine semantisch-angereicherte Variante der Wikipedia,⁶⁹ jedoch wird sie von diversen weiteren Datenbanken ergänzt. Zur Identifikation von Ressourcen über die Datenbanken hinweg werden sogenannte URI (Abkürzung für: Uniform Resource Identifier) verwendet. Über diese URI können Drittquellen Ressourcen um weitere Information ergänzen oder sie in Relation zu anderen Ressourcen stellen. Damit entsteht ein großes Informationsnetz mit heterogenen Daten. Aus diesem Wissensnetz lassen sich verschiedene Visualisierungen entwickeln, von analytischen Systemen bis hin zu explorativen Systemen (in Abb. 16 dargestellt).

⁶⁶ Vgl. Burkhardt und Nazemi 2018, 159–160.

⁶⁷ Vgl. Nazemi et al. 2015a, 76–77.

⁶⁸ S. <https://www.dbpedia.org>.

⁶⁹ S. <https://www.wikipedia.org>.

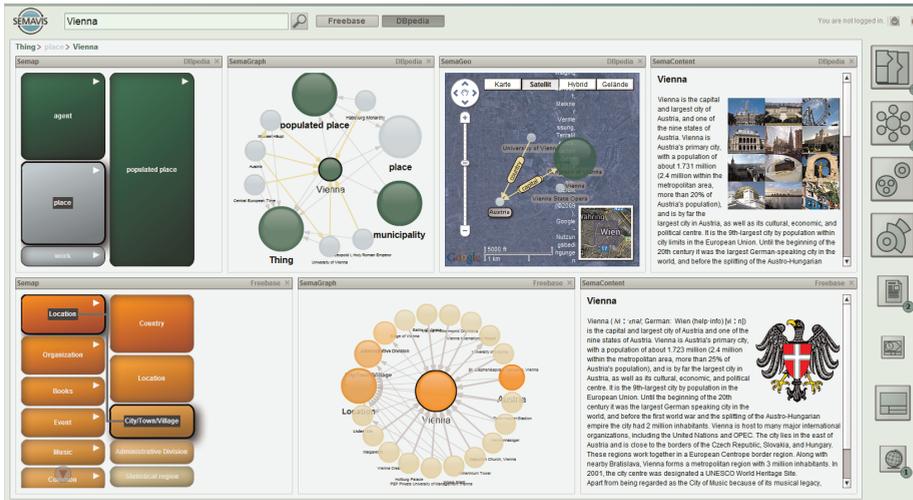


Abb. 16: Screenshot von einer Semantic-Web Visualisierungslösung, die sich aus verschiedenen Linked-Open Daten Quellen die Ergebnisse visualisiert⁷⁰

5.4 Visualisierung von Produktions- und Manufakturdaten

Informationsvisualisierung und Visual Analytics in Smart Manufacturing bekommt immer mehr Aufmerksamkeit durch die Wissenschaft. Eine hohe Anzahl an Sensoren und die Verknüpfung untereinander führt zu immer größeren Datenmengen. Im Fehlerfall präzise und schnell reagieren zu können, kann helfen die Ausfallzeit zu vermindern und Kosten einzusparen.⁷¹ Ein Fehlerfall geht meist im Vorfeld mit Anomalien innerhalb des Datenstroms einher. Diese Anomalien zu erkennen, bevor der Fehlerfall eintritt, bildet dabei den Schlüssel zum Erfolg.

Neuere Visual Analytics Ansätze (vgl. Abb. 17), wie bei Xu et al. vorgestellt,⁷² ermöglichen eine Echtzeitüberwachung von gesamten Fertigungsstrecken mit Hilfe einer durchdachten Oberfläche. Mit KI-Methoden können schon heute große Datenmengen analysiert werden.⁷³ Dabei kann ein Visual Analytics-System aktiv dabei unterstützen die richtigen Parameter für eine KI-Anwendung zu setzen und die Ergebnisse zu evaluieren.⁷⁴ Diese zwei Publikationen stehen stellvertretend für zwei Forschungsrichtungen innerhalb des Smart Manufacturing Forschungsfeldes. Zum

⁷⁰ Vgl. Nazemi 2016, 284.

⁷¹ Vgl. Kaupp et al. 2017, 83–97; Beez et al. 2018, 163–180.

⁷² Vgl. Xu et al. 2017, 291–300.

⁷³ Vgl. Kaupp et al. 2019, 55–65.

⁷⁴ Vgl. Xu et al. 2018, 109–119.

einen, von der Maschine ausgehend, den kompletten Fertigungsprozess zu visualisieren,⁷⁵ zum anderen die verwendeten Methoden transparenter zu gestalten.⁷⁶

Beide setzen damit auf Anomaly Detection oder auch Novelty Detection. Wobei im Kontext des Smart Manufacturing große Mengen multivariater Daten meist unsupervised anomaly detection Algorithmen genutzt werden. Dabei können auch Verfahren des maschinellen Lernens eingesetzt werden. Ein Datensatz innerhalb des maschinellen Lernens besteht dabei immer aus Eingangswerten (Ausgangssituation) und der passenden Beschreibung (Label). Bei großen Datenmengen (in der Regel mehrere Terabytes) jeden Datensatz händisch zu annotieren, ist dabei sehr zeit- und arbeitsintensiv, also schlussendlich kostenintensiv und fehleranfällig. Weshalb hier auf unüberwachte (unsupervised) Techniken zurückgegriffen wird. Hier fehlt das Label und die Anomalie wird mittels mathematischer Operationen auf Basis der Ausgangssituation bestimmt. Ke Xu et al. erzeugen automatische Verbünde (Ensembles) von vordefinierten Anomalie-Detektions-Algorithmen auf Basis der verschiedenen Datentypen und visualisieren deren Effektivität. So können für bestimmte Daten die besten Ensembles ausgewählt werden, um innerhalb der großen Menge multivariater Daten effektiv Anomalien zu entdecken. Mit dieser Methodik wird der Weg von den Daten zur Anomalie transparent gestaltet.

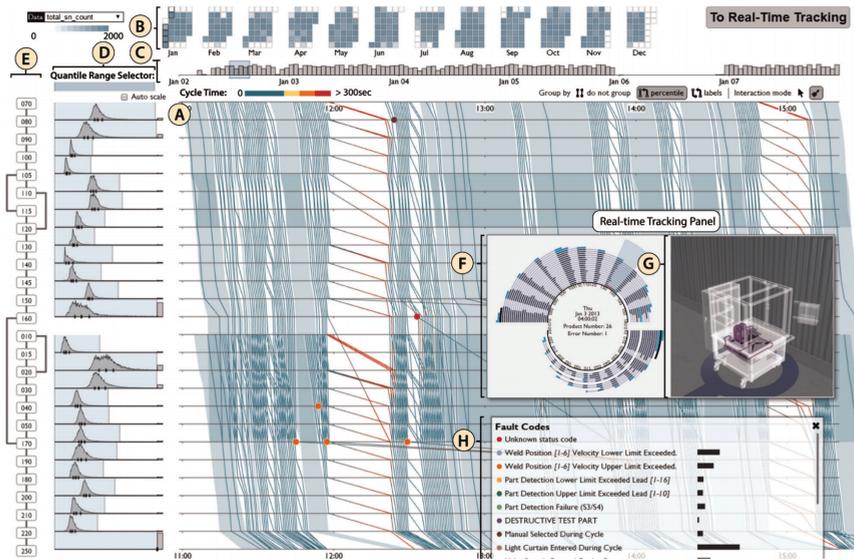


Abb. 17: ViDX Visual Diagnostics of Assembly Line Performance in Smart Factories⁷⁷

75 Vgl. Xu et al. 2017, 291–300.

76 Vgl. Xu et al. 2018, 109–119.

77 Vgl. Xu et al. 2017, 291.

Xu et al. zeigen mit ViDX (Visual Diagnostics of Ansembly Line Performance for Smart Factories) ein Dashboard, das den Fertigungsprozess visualisiert. Unterteilt ist die Oberfläche in fünf Abschnitte; die Station-Map, Histogramme, einen erweiterten Marey-Graphen, darüberliegend eine Timeline mit Kalender (vgl. Abb. 17). Mittels Station Map werden (links außen in Abb. 17) der Ablauf und die Verbindungen der einzelnen Stationen visualisiert. Das Histogramm zeigt die Auslastung der Station. Der erweiterte Marey-Graph gibt durch die parallele Betrachtung aller Stationen gleichzeitig einen Überblick über Ausfälle (Lücken) oder Verzögerungen (zusammenlaufender Graph mit verzögerter Weiterverarbeitung). Anhand der Timeline kann der Ausschuss der Fertigungsstraße verglichen und Anomalien können so schneller gefunden werden. Der Kalender aggregiert dabei die Timeline auf Monatsbasis. Ke Xu et al.⁷⁶ überwachen in der Historie, also über die Zeit, die physische Anordnung und den Ausstoß einer ganze Fabrik. Heutige Fabriken emittieren meist noch komplexere Datensätze, welche noch speziellere Visualisierungen in Zukunft nötig machen. Die Datenvisualisierung bildet damit eine wesentliche Grundlage innerhalb des Smart Manufacturing und sorgt dafür sehr komplexe Sachverhalte sinnbringend darzustellen.

Fazit

Die steigende Menge an Daten stellt eine große Herausforderung zur Bearbeitung, Analyse und Gewinnung neuer Erkenntnisse dar. Methoden der Datenvisualisierung, insbesondere jene der Informationsvisualisierung und Visual Analytics bieten hier Möglichkeiten, aus enorm großen Datenmengen Erkenntnisse zu gewinnen, Wissen zu generieren und Muster in den Daten zu erkennen. Dieser Trend hat dazu geführt, dass Visual Analytics und Informationsvisualisierung längst zu etablierten Methoden der Datenanalyse gehören. Dies beschränkt sich nicht nur auf Forschung, sondern erweitert den Nutzerkreis auf Akteure der Wirtschaft. So gibt es diverse Werkzeuge von fast allen Softwareanbietern im Bereich der Business Intelligence, die teilweise für wissenschaftliche Zwecke frei genutzt werden können. Beispielhaft sollen hier die bekanntesten Systeme genannt werden. Microsoft bietet mit PowerBI⁷⁸ eine Lösung an, die vergleichsweise einfach strukturierte Daten mit verschiedenen visuellen Layouts darstellen kann. Dabei können visuelle Variablen auch personalisiert werden. Die Nutzung der Desktop-Variante ist kostenlos.⁷⁹ Tableau⁸⁰ ist ebenfalls eine Lösung zur Visualisierung strukturierter Daten. Diese hat ähnliche Funktionen wie PowerBI und ist zu Lehr- und Forschungszwecken als Desktop-Vari-

⁷⁸ S. <https://powerbi.microsoft.com/de-de/>.

⁷⁹ S. https://powerbi.microsoft.com/de-de/desktop/?WT.mc_id=Blog_Desktop_Update.

⁸⁰ S. <https://www.tableau.com/de-de>.

ante kostenlos. Microsoft Azure⁸¹ steht an den meisten Hochschulen kostenlos zur Verfügung. Der Trend des kombinierten Einsatzes von maschinellen Lernverfahren bzw. Methoden der künstlichen Intelligenz und Visualisierungen ist in der Forschung zum Zeitpunkt des Erscheinens dieses Buches von enormer Bedeutung und wird in Zukunft zu vielen weiteren Entwicklungen im Bereich Visual Analytics führen, die auch zur Laufzeit die Auswahl der Methode und Parametrisierungen erlauben wird. Um die oben genannten Werkzeuge adäquat nutzen zu können, wurden in diesem Kapitel kanonische und wissenschaftlich etablierte Aspekte der Visualisierung dargestellt.

Literaturverzeichnis

Letztes Abrufdatum der Internet-Dokumente ist der 15.11.2020.

- Aggarwal, Charu C., Hg. 2017. *Outlier Analysis*. 2nd ed. 2017. Cham: Springer International Publishing.
- Aigner, Wolfgang, Silvia Miksch, Heidrun Schumann und Christian Tominski. 2011. *Visualization of Time-Oriented Data*. London: Springer.
- Andrienko, Natalia und Gennady Andrienko. 2006. *Exploratory Analysis of Spatial and Temporal Data*. Berlin, Heidelberg: Springer-Verlag.
- Beez, Ulrich, Lukas Kaupp, Tilman Deuschel, Bernhard G. Humm, Fabienne Schumann, Jürgen Bock und Jens Hülsmann. 2018. „Context-Aware Documentation in the Smart Factory.“ In *Semantic applications: Methodology, technology, corporate use*. Bd. 23, hg. v. Thomas Hoppe, Bernhard Humm und Anatol Reibold, 163–80. Berlin: Springer.
- Bertin, Jacques. 1983. *Semiology of Graphics: Diagrams, Networks, Maps*. Übersetzt von J. B. Berg. Madison, Wis. Univ. of Wisconsin Press.
- Below, Nicola. 2018. *Partizipation und Transparenz der europäischen Chemikalienregulierung - Juristische Analyse der Inklusionsleistung der europäischen Chemikalienregulierungsbehörden*, FORUM Wirtschaftsrecht 24, Kassel: kassel university press.
- Blei, David M., Andrew Y. Ng und Michael I. Jordan. 2003. „Latent Dirichlet Allocation.“ *The Journal of Machine Learning Research* 3: 993–1022.
- Böhret, Carl; Konzendorf, Götz. 2001. *Handbuch Gesetzesfolgenabschätzung (GFA), Gesetze, Verordnungen, Verwaltungsvorschriften*, Baden-Baden: Nomos Verlag.
- Burkhardt, Dirk und Kawa Nazemi. 2018. „Visualizing Law – A Norm-Graph Visualization Approach based on Semantic Legal Data.“ In *The 4th International Conference of the Virtual and Augmented Reality in Education: I3M*.
- Burkhardt, Dirk, Kawa Nazemi, Jan R. Zilke, Jörn Kohlhammer und Arjan Kuijper. 2014. „Fundamental Aspects for E-Government.“ In *Handbook of research on advanced ICT integration for governance and policy modeling*, hg. v. Dirk Burkhardt, Giorgio Prister, Kawa Nazemi, Peter Sonntagbauer und Susanne Sonntagbauer, 1–18. (Advances in Electronic Government, Digital Divide, and Regional Development.) Hershey, Pa.: IGI Global.

⁸¹ S. <https://azure.microsoft.com/de-de/>.

- Card, S. K. und J. Mackinlay. 1997. „The Structure of the Information Visualization Design Space.“ In *Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*, 92–99. Los Alamitos: IEEE Computer Society Press.
- Card, Stuart K., Jock D. Mackinlay und Ben Shneiderman, Hg. 1999. *Readings in Information Visualization: Using Vision to Think*. [Nachdr.]. (The Morgan Kaufmann series in interactive technologies.) San Francisco, Ca.: Morgan Kaufmann.
- Earnshaw, Rae A. und Norman Wiseman. 1992. „What Scientific Visualization Can Do!“ In *An Introductory Guide to Scientific Visualization*, hg. v. Rae A. Earnshaw und Norman Wiseman, 5–19. Berlin, Heidelberg: Springer.
- El-Assady, Mennatallah, Rebecca Kehlbeck, Christopher Collins, Daniel Keim und Oliver Deussen. 2020. „Semantic Concept Spaces: Guided Topic Model Refinement Using Word-Embedding Projections.“ *IEEE transactions on visualization and computer graphics* 26 (1): 1001–11. doi:10.1109/TVCG.2019.2934654.
- Gahegan, Mark. 1998. „Scatterplots and scenes: visualisation techniques for exploratory spatial analysis.“ *Computers, Environment and Urban Systems* 22 (1): 43–56. doi:10.1016/S0198-9715(98)00018-0.
- Inselberg, A. und B. Dimsdale. 1990. „Parallel coordinates: a tool for visualizing multi-dimensional geometry.“ In *Proceedings of the 1st Conference on Visualization '90*, 361–78. VIS '90. Los Alamitos, CA. <http://dl.acm.org/citation.cfm?id=949531.949588>.
- Kaupp, Lukas, Ulrich Beez, Jens Hülsmann und Bernhard G. Humm. 2019. „Outlier Detection in Temporal Spatial Log Data Using Autoencoder for Industry 4.0.“ In *Engineering Applications of Neural Networks*. Bd. 1000, hg. v. John Macintyre, Lazaros Iliadis, Ilias Maglogiannis und Christina Jayne, 55–65. Cham: Springer International Publishing.
- Kaupp, Lukas, Ulrich Beez, Bernhard G. Humm und Jens Hülsmann. 2017. „From Raw Data to Smart Documentation: Introducing a Semantic Fusion Process for Cyber-Physical Systems.“ In *CERC2017 Collaborative European Research Conference: Proceedings*, 83–97. Karlsruhe.
- Keim, Daniel, Hg. 2010. *Mastering the Information Age: Solving Problems with Visual Analytics*. Goslar: Eurographics Association.
- Keim, Daniel, Christian Panse und Mike Sips. 2005. „Information Visualization: Scope, Techniques and Opportunities for Geovisualization.“ In *Exploring Geovisualization*. Dykes, J. et al. (Hrsg.). – Oxford, Elsevier, 2005. – S. 1–17.
- Kohlhammer, Jörn, Kawa Nazemi, Tobias Ruppert und Dirk Burkhardt. 2012. „Toward Visualization in Policy Modeling“ *IEEE computer graphics and applications* 32 (5): 84–89. doi:10.1109/MCG.2012.107.
- Macintosh, A. 2004. „Characterizing E-Participation in Policy-Making.“ In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences: Abstracts and CD-ROM of full papers: 5–8 January, 2004, Big Island, Hawaii*, hg. v. Ralph H. Sprague. Los Alamitos, Ca.: IEEE Computer Society Press.
- Mackinlay, Jock. 1986. „Automating the design of graphical presentations of relational information.“ *ACM Trans. Graph.* 5 (2): 110–41. doi:10.1145/22949.22950.
- May, T. und J. Kohlhammer. 2008. „Towards closing the analysis gap: Visual generation of decision supporting schemes from raw data.“ *Computer Graphics Forum* 27 (3): 911–18. doi:10.1111/j.1467-8659.2008.01224.x.
- May, Thorsten, James Davey und J. Kohlhammer. 2010. „Combining Statistical Independence Testing, Visual Attribute Selection and Automated Analysis to Find Relevant Attributes for Classification.“ In *IEEE Symposium on Visual Analytics Science and Technology (VAST), 2010: 25–26 Oct. 2010, Salt Lake City, Utah, USA; proceedings /sponsored by IEEE Computer Society Visualization and Graphics Technical Committee*. Ed. by Alan MacEachren, 239–40. Piscataway, NJ: IEEE Press.

- Munzner, Tamara. 2014. *Visualization Analysis and Design*. AK Peters Visualization Series. Hoboken: Taylor and Francis.
- Nazemi, Kawa. 2016. *Adaptive Semantics Visualization*. Studies in computational intelligence Volume 646. ORT: Springer.
- Nazemi, Kawa. 2018. „Intelligent Visual Analytics – a Human-Adaptive Approach for Complex and Analytical Tasks.“ In *Advances in Intelligent Systems*, hg. v. Ahram Karwowski und Tareq Ahram, 180–190. Cham: Springer International Publishing.
- Nazemi, Kawa und Dirk Burkhardt. 2019a. „A Visual Analytics Approach for Analyzing Technological Trends in Technology and Innovation Management.“ In *Advances in Visual Computing*. Bd. 11845, hg. v. George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Daniela Ushizima, Sek Chai, Shinjiro Sueda et al., 283–294. (Lecture Notes in Computer Science.) Cham: Springer International Publishing.
- Nazemi, Kawa und Dirk Burkhardt. 2019b. „Visual Analytics for Analyzing Technological Trends from Text.“ In *Information Visualisation: Biomedical visualization and geometric modelling & imaging*, hg. v. E. Banissi, 191–200. Los Alamitos, CA: IEEE Computer Society, Conference Publishing Services.
- Nazemi, Kawa, Dirk Burkhardt, Egils Ginters und Jorn Kohlhammer. 2015a. „Semantics Visualization – Definition, Approaches and Challenges.“ *Procedia Computer Science* 75: 75–83. doi:10.1016/j.procs.2015.12.216.
- Nazemi, Kawa, Reimond Retz, Dirk Burkhardt, Arjan Kuijper, J. Kohlhammer und Dieter W. Fellner. 2015b. *Visual trend analysis with digital libraries: Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business – i-KNOW '15, New York, USA*, ACM Press, 2015.
- Shneiderman, B. 1996. „The eyes have it: a task by data type taxonomy for information visualizations.“ In *Proceedings 1996 IEEE Symposium on Visual Languages*, 336–43.
- Stevens, S. S. 1946. „On the Theory of Scales of Measurement.“ *Science* 103 (2684): 677–680. doi:10.1126/science.103.2684.677.
- Stoffel, Sacha, Andreas Stoffel, Florian Kwon, Bum Chul, Geoffrey Ellis und Daniel Keim. 2014. „Knowledge Generation Model for Visual Analytics.“ *IEEE transactions on visualization and computer graphics* 20 (12): 1604–1613. doi:10.1109/TVCG.2014.2346481.
- Thomas, James J. und Kristin A. Cook. 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. United States US Department of Homeland Security: National Visualization and Analytics Center, IEEE Press.
- Velleman, Paul F. und Leland Wilkinson. 1993. „Nominal, Ordinal, Interval, and Ratio Typologies are Misleading.“ *The American Statistician* 47 (1): 65–72. doi:10.1080/00031305.1993.10475938.
- Viau, Christophe, Michael McGuffin, Yves Chiricota und Igor Jurisica. 2010. „The FlowVizMenu and Parallel Scatterplot Matrix: Hybrid Multidimensional Visualizations for Network Exploration.“ *IEEE transactions on visualization and computer graphics* 16 (6): 1100–1108. doi:10.1109/TVCG.2010.205.
- Ward, Matthew O., Georges Grinstein und Daniel Keim. 2010. *Interactive Data Visualization: Foundations, Techniques, and Applications*. Hoboken: 360 Degree Business.
- West, Thomas G. 1999. „Images and reversals: James Clerk Maxwell, working in wet clay“ *SIGGRAPH Comput. Graph.* 33 (1): 15. doi:10.1145/563666.563671.
- Xu, Ke, Meng Xia, Xing Mu, Yun Wang und Nan Cao. 2018. „EnsembleLens: Ensemble-Based Visual Exploration of Anomaly Detection Algorithms with Multidimensional Data.“ *IEEE transactions on visualization and computer graphics*. doi:10.1109/TVCG.2018.2864825.
- Xu, Panpan, Honghui Mei, Liu Ren und Wei Chen. 2017. „ViDX: Visual Diagnostics of Assembly Line Performance in Smart Factories.“ *IEEE transactions on visualization and computer graphics* 23 (1): 291–300. doi:10.1109/TVCG.2016.2598664.