

3.6 Fazit

Data Mining kombiniert verschiedene Techniken, die jeweils eine ungerichtete, hypothesenfreie Datenauswertung unterstützen. Typische Auswertungsziele liegen in einer Aufteilung von Datenräumen durch Identifikation entsprechender Segmente, der automatischen Zuordnung neuer Datensätze zu vordefinierten Gruppen durch Klassifizierung oder dem Auffinden von Zusammenhängen durch Assoziation. Data Mining kann im Rahmen eines interaktiven Prozesses zur Entdeckung bisher unbekannter Mustern in Daten führen, die dann von jeweiligen Experten des Anwendungsgebietes zu bewerten sind. Hier wird am ehesten deutlich, welche Rolle Data Mining bei der Auffindung neuer Zusammenhänge spielen kann. Durch die entsprechenden Algorithmen können Datenmengen, die aufgrund von Größe oder Komplexität basierend auf Hypothesen nicht mehr geeignet ausgewertet werden können, schnell nach Zusammenhängen und Mustern durchsucht werden. Die Bewertung der Nützlichkeit der so in den Daten gefundenen Strukturen allerdings erfordert umfassende Fachkenntnisse über das jeweilige Anwendungsgebiet, um interessante und potentiell wertvolle Muster von unbrauchbaren zu unterscheiden und verwertbare Erkenntnisse ableiten zu können; Data Mining kann hier also allenfalls unterstützend wirken und potentiell interessante Hinweise geben, nicht jedoch Expertenwissen und menschliche Erfahrung ersetzen. Nicht zuletzt sind Konzepte wie Data Mining und Data Warehousing aufgrund von Tendenzen zu ihrer Anwendung auf personenbezogene Daten mit Zielen, die von der Absatzoptimierung bis zur automatischen Klassifizierung potentieller Straftäter reichen, aus datenschutzrechtlichen Gründen auch kritisch zu bewerten.

4 Computergestützte Datenvisualisierung

„Scientists need an alternative to numbers. A technical reality today and a cognitive imperative tomorrow is the use of images.“¹¹⁰

„The purpose of visualization is insight, not pictures.“¹¹¹

Kapitel 4 behandelt das Gebiet der Visualisierung, das sich mit der Nutzung der menschlichen visuellen Fähigkeiten bei der Auswertung von Daten durch computergestützte graphische Repräsentationen befasst. Kap. 4.1 dient zur Einführung in dieses Gebiet und zur Abgrenzung der Untergebiete Scientific Visualization und Informationsvisualisierung; in Kap. 4.2 werden mit Expressivität, Effektivität und Angemessenheit zentrale Anforderungen an eine Datenvisualisierung vorgestellt. Kap. 4.3 skizziert den Visualisierungsprozess und Unterscheidungsmerkmale von Visualisierungstechniken; auf Herausforderungen, die durch eine Visualisierung von Daten mit vielen Dimensionen und / oder hohem Wertebereich aufgeworfen werden, wird in Kap. 4.4 eingegangen. In Kap. 4.5 werden Basistechniken zur Datenvisualisierung vorgestellt, eine Darstellung von Konzepten zur Visualisierung multivariater¹¹² Daten (Kap. 4.6) sowie zur Visualisierung von Zeit- und Raumbezug (Kap. 4.7) schließt sich an. In Kap. 4.8 werden ausgewählte Beispiele aus dem Untergebiet der Informationsvisualisierung vorgestellt; ein abschließendes Fazit wird in Kap. 4.9 gezogen.

4.1 Einführung

4.1.1 Visualisierung

Die visuelle Wahrnehmung nimmt im menschlichen Sinnenapparat eine besondere Rolle ein. So enthält das menschliche Auge über 100 Millionen Photorezeptoren (Stäbchen und Zapfen) [Poggio 1986]; die pro Sekunde in beide Augen eingehende Datenmenge wird auf ca. 50 Milliarden Bit beziffert (im Vergleich zu ca. 1 Million Bit pro Sekunde über beide Ohren) [Kurzweil 1993, 255ff.]. Hiermit korrespondiert die besondere Rolle, die Bilder in der Kulturgeschichte des Menschen schon sehr früh¹¹³ eingenommen haben. Bilder erfüllen für den Menschen verschiedene Aufgaben; Schuck-Wersig unterscheidet hier zwischen den kulturellen Bildfunktionen Magie, Orientierung, Identifikation/Projektion, Wissensrepräsentation und Sinnlichkeit [Schuck-Wersig 1993, 67ff.]. Mit den Fortschritten der Computertechnologie und ihrem sich bislang stetig verbessernden Preis-/ Leistungsverhältnis ist seit wenigen Jahrzehnten eine zunehmend effiziente Erzeugung computergenerierter graphischer Repräsentationen aus vorhandenen Daten möglich. Der computergestützte Einsatz von aus Daten generierten graphischen Repräsentationen mit dem Ziel, einen besseren Zugang zu den in ihnen enthaltenen Informationen zu erhalten, wird als *Visualisierung* (*Visualization*)¹¹⁴ bezeichnet:

- ▶ „*Visualization: The use of computer-supported, interactive, visual representations of data to amplify cognition.*“ [Card et al. 1999, 6]

¹¹⁰ McCormick, B.H.; DeFanti, T.A.: *Visualization in Scientific Computing* [McCormick, DeFanti 1987, 7].

¹¹¹ Card, S.K.; Mackinlay, J.D.; Shneiderman, B.: *Readings in Information Visualization: Using Vision to think* [Card et al. 1999, 6].

¹¹² Zur Begrifflichkeit vgl. Kap. 4.4.1

¹¹³ So sind frühe Höhlenzeichnungen bereits im Jungpaläolithikum (um 30.000 v.Chr.) nachweisbar; Tätowierungen wurden an ägyptischen Mumien aus der Zeit um 2.000 v.Chr. festgestellt, wurden wahrscheinlich aber bereits vom Cro-Magnon-Mensch vor 40.000 Jahren ausgeführt. Körperbemalungen fanden sich möglicherweise bereits vor 400.000 Jahren beim Homo erectus (vgl. [Schuck-Wersig 1993, 49ff.]).

¹¹⁴ Benachbarte Gebiete sind bspw. Computer Graphik [Foley et al. 1990] sowie Computer Vision [Fischler, Firschein 1987].

Datenvisualisierungen können nach Card et al. Erkenntnisprozesse auf vielfältige Weise unterstützen. Sie erlauben eine direkte Nutzung der Fähigkeiten des visuellen Systems sowie eine Entlastung der Gedächtnisanforderungen durch eine externe, visuelle Ablage von Informationen, ebenso die Reduzierung des Suchaufwandes nach spezifischen Daten bspw. durch Gruppierung oder die Herstellung visueller Bezüge, ein verbessertes Erkennen von Mustern in den dargestellten Daten oder das Beobachten von Veränderungen [Card et al. 1999, 16f.].

Geeignete graphische Darstellungen von Daten offenbaren dem Betrachter die in diesen enthaltenen Strukturen oft weitaus intuitiver als Zahlenkolonnen. Dies soll zunächst an einem einfachen Beispiel verdeutlicht werden. Die Tabelle in Abb. 4.1a enthält insgesamt 15 Zahlenwerte für den Werteverlauf von drei Variablen A, B und C über die Jahre 1980 bis 1984. Bereits für diese sehr kleine Datenmenge sind die in ihr enthaltenen Informationen anhand der tabellarischen Darstellung nicht leicht zu entdecken. Weder ist hier sofort zu erkennen, in welchem Jahr jeweils die größten oder kleinsten Werte für die einzelnen Variablen anfallen, noch ob Zusammenhänge zwischen der Werteentwicklung der einzelnen Variablen bestehen. Eine entsprechende graphische Darstellung dieser Daten offenbart diese Zusammenhänge hingegen gleichsam *auf einen Blick*. Das aus den Daten generierte Liniendiagramm (Abb. 4.1b) erlaubt nicht nur, schnell die jeweiligen Minima und Maxima der einzelnen Variablen zu identifizieren, sondern veranschaulicht zudem die Korrelationen im Werteverlauf. Es ist deutlich zu erkennen, dass die Werte von Variable A jeweils steigen oder sinken, wenn dies auch für die Werte von Variable B der Fall ist, während die Werteentwicklung von Variable C einer entgegengesetzten Tendenz folgt.

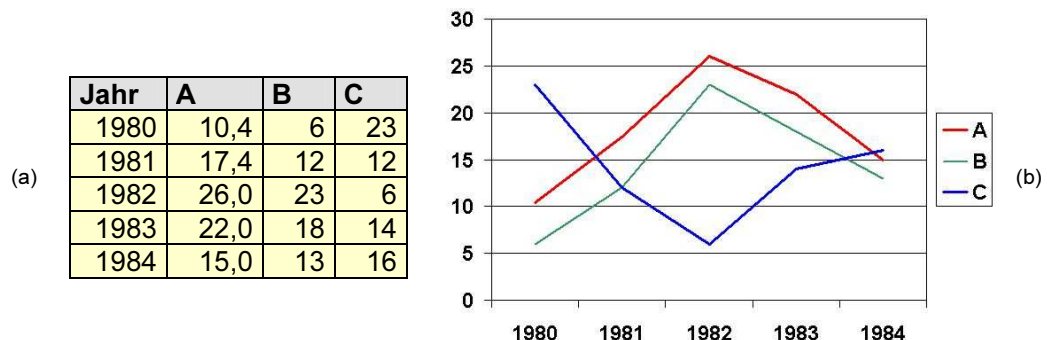


Abb. 4.1 - Vorteile einer Datenvisualisierung: (a) Ausgangsdaten, (b) graphische Darstellung.

Die bildliche Darstellung von Daten ist keine Erfindung des Computerzeitalters. Eine *computergestützte* Visualisierung von Daten bietet jedoch gegenüber statischen Darstellungen beispielsweise auf Papier entscheidende Vorteile. Sie ermöglicht die Entkopplung von *Werkzeugentwicklung* und *Werkzeugeinsatz* und erlaubt so die Erstellung von visuellen Artefakten zur Zeit der Benutzung und für Daten, die dem Erzeuger des Visualisierungswerkzeuges nicht bekannt sein müssen. Zudem erlaubt sie eine direkte Veränderung der graphischen Darstellung durch Interaktion des Benutzers. Card et al. konstatieren, dass die hieraus resultierenden Möglichkeiten, mehr Repräsentationen in einer gegebenen Zeit zu erforschen sowie den Aufwand zur Erstellung der Repräsentationen auf Maschinen zu verlagern und sich auf die Auswirkungen von Modifikationen zu konzentrieren, den Prozess des Verstehens von Daten fundamental verändern können [Card et al. 1999, 231]. Im Kontext der Visualisierung von Daten wird gerne auf Richard Hamming's Satz

- ▶ „The purpose of computation is insight, not numbers.”¹¹⁵

¹¹⁵ Zitiert nach [Card et al. 1999, 6], als Quelle wird dort [Hamming 1973] angegeben.

verwiesen (vgl. bspw. den für die Entwicklung der computergestützten Visualisierung einflussreichen Report *Visualization in Scientific Computing* von [McCormick, DeFanti 1987, 3] und bezugnehmend hierauf [Hennings 1991, 175, Fußnote 191]). Card et al. wandeln die Aussage von Hamming ab und betonen:

- ▶ „*The purpose of visualization is insight, not pictures.*“ [Card et al. 1999, 6]

Die Autoren nennen Entdeckung (Discovery), Entscheidungsfindung (Decision Making) sowie Erklärung (Explanation) als wesentliche Ziele einer solcherart erreichbaren Einsicht. Eine visuelle Repräsentation von Daten kann dabei verschiedenen Zwecken dienen; nach [Schumann, Müller 2000, 62f.] lassen sich die folgende Problemklassen identifizieren, die zugleich als Bearbeitungsziele mit einer gewissen Allgemeingültigkeit angesehen werden können:

- | | |
|-----------------|---|
| Identifikation | ▶ Welchen Wert haben Daten in einem bestimmten Gebiet? |
| Lokalisierung | ▶ Wo liegen Daten eines bestimmten Wertes? |
| Korrelation | ▶ Gibt es Zusammenhänge zwischen zwei oder mehr Variablen oder zwischen Datenwerten und bestimmten Gebieten des Beobachtungsraumes bzw. bestimmten Zeitpunkten? |
| Vergleich | ▶ Wie unterscheiden sich die Datenwerte in einem bestimmten Gebiet oder zu unterschiedlichen Zeitpunkten? |
| Verteilung | ▶ Wo liegen Extremwerte und Ausreißer? Lassen sich Muster in den Datenwerten bzw. Trends erkennen? |
| Häufigkeit | ▶ Welche Datenwerte treten besonders häufig auf? |
| Gruppierung | ▶ Welche Datenwerte lassen sich anhand gemeinsamer Eigenschaften zusammenfassen, das heißt, welche Cluster treten auf? |
| Kategorisierung | ▶ Welche Datenwerte müssen auf Grund unterschiedlicher Eigenschaften separiert werden, das heißt, welche Klassifizierungen können vorgenommen werden? |

Die Visualisierung von Daten erfolgt durch geeignete Abbildung auf sog. *visuelle Variablen* [Schumann, Müller 2000, 125ff.], zu denen *Position, Größe, Orientierung, Form, Struktur, Farbe* sowie *Textur* gerechnet werden. Die Wirkungen jeder dieser Variablen ist spezifisch und kann anhand der folgen drei Formen unterschieden werden:

- | | |
|-----------------------|---|
| Selektive Wirkung | ▶ Der Betrachter kann unterschiedliche Datenwerte spontan in Gruppen aufteilen und unterscheiden. |
| Ordinale Wirkung | ▶ Der Betrachter kann unterschiedliche Datenwerte spontan in eine Ordnung bringen. |
| Proportionale Wirkung | ▶ Der Betrachter kann unterschiedliche Datenwerte spontan in eine Ordnung bringen und zugleich die Größe der zugrundeliegenden Datenwerte direkt assoziieren. |

Dabei bilden Position, Größe und Orientierung den Autoren zufolge die effektivsten visuellen Variablen; Positions- und Größenvergleiche können vom Menschen besonders genau durchgeführt werden. Farbe und Textur werden als *retinale Variablen* bezeichnet, die von geometrischen Elementen unabhängig wahrgenommen und daher zusätzlich zu diesen eingesetzt werden können. Sie zeichnen sich dadurch aus, dass sie, anders als die anderen visuellen Variablen, ohne kognitive Belastung wahrgenommen werden und daher gut in Kombination mit diesen eingesetzt werden können. Der Einsatz von Farbe birgt allerdings

zugleich stets ein Risiko von Fehlinterpretationen. So werden einige Farben schneller wahrgenommen als andere¹¹⁶; eine entsprechende Auswahl hat daher direkte Auswirkungen auf die Rezeption der dargestellten Informationen¹¹⁷. Ferner beeinflussen die Farben der Umgebung die Farbwahrnehmung eines Objektes¹¹⁸; Größe und Farbe eines Objektes können ebenfalls die Wahrnehmung beeinflussen. So hat die Größe eines Objektes Einfluss auf die Reinheit, mit der dessen Farbe wahrgenommen wird; die Farbe eines Objektes kann wiederum Einfluss auf seine wahrgenommene Größe haben¹¹⁹. Die Wahrnehmung von Farbe kann zudem durch berufsspezifische¹²⁰ oder kulturelle¹²¹ Konventionen bestimmt und nicht zuletzt durch verschiedene Formen von Farbfehlsichtigkeit¹²² des Betrachters beeinträchtigt sein.

4.1.2 Die Visualisierung wissenschaftlicher Daten – Scientific Visualization

Von besonderer Bedeutung ist die Visualisierung von Daten im wissenschaftlichen Kontext. Sie kann hier wesentlich sowohl zum Entdecken und Verstehen von in Daten enthaltenen Phänomenen, zur Kommunikation wissenschaftlicher Ergebnisse sowie zur wissenschaftlichen Ausbildung beitragen. Ferner kann die Erzeugung geeigneter graphischer Repräsentationen insbesondere auch bei der Auswertung anfallender, sehr großer Datenmengen unterstützen [McCormick, DeFanti 1987] [DeFanti et al. 1989]. Die Visualisierung von Daten im wissenschaftlichen Kontext (vgl. die Beispiele in Abb. 4.2a und 4.2b) wird als *Scientific Visualization* bezeichnet:

- ▶ „*Scientific visualization: Use of interactive visual representations of scientific data, typically physically based, to amplify cognition.*“¹²³ [Card et al. 1999, 7, Tab. 1.1]
- ▶ „*Die wissenschaftlich-technische Visualisierung, auch als Scientific Visualization bezeichnet, hat die Aufgabe, geeignete visuelle Repräsentationen einer gegebenen Datenmenge zu erzeugen, um damit eine effektive Auswertung zu ermöglichen. Es soll die Analyse, das Verständnis und die Kommunikation von Modellen, Konzepten und Daten in der Wissenschaft und im Ingenieurbereich erleichtert werden.*“¹²⁴ [Schumann, Müller 2000, 5]

Visualisierung als organisiertes Gebiet der Computerwissenschaft ist eng mit dem Feld der Scientific Visualization verknüpft. Als Initialzündung dieses Bereichs kann der Workshop on Visualization in Scientific Computing der amerikanischen National Science Foundation (NSF)¹²⁵ im Jahr 1987 betrachtet werden [Card et al. 1999, 6, 37f.]. Der entsprechende

¹¹⁶ Eine Anordnung von Farben nach der Reaktionszeit auf diese (von kurz nach lang) ist: Gelb → Weiß → Rot → Grün → Blau (vgl. [Schumann, Müller 2000, 88]).

¹¹⁷ Details, die farblich besonders hervorgehoben werden, können eine unangemessene Überbewertung zur Folge haben [Meyer 1999, 169].

¹¹⁸ Ein Beispiel hierfür ist der Helmholtz-Kohlrausch- oder Farbglut-Effekt – ein farbiges Objekt vor einem grauen Hintergrund scheint im Vergleich zum Grauwert zu glühen (vgl. [Schumann, Müller 2000, 87]).

¹¹⁹ So wirkt ein kleines rotes Objekt vor schwarzem Hintergrund intensiver als größere Objekte gleicher Farbe; durch den sog. Irradiationseffekt wirkt hingegen ein weißes Rechteck neben einem schwarzen Rechteck bei gleicher Größe länger (vgl. [Schumann, Müller 2000, 85]).

¹²⁰ Bspw. wird in den Ingenieurwissenschaften oder der Wirtschaft die Farbe Rot oft mit Gefahr, Blau mit neutraler und Grün mit positiver Bedeutung verbunden, in der Medizin hingegen Grün und Blau mit Infizierung und Tod assoziiert und Rot als Farbe des Lebens positiv wahrgenommen [Schumann, Müller 2000, 115].

¹²¹ Die Farbe Weiß wird im abendländischen Kulturkreis mit positiven Werten wie Freude und Reinheit assoziiert, Schwarz hingegen mit Trauer; eine umgekehrte Bedeutungszuordnung findet sich in Indien [Schumann, Müller 2000, 115].

¹²² Zu Visualisierung und Farbfehlsichtigkeit vgl. [Schumann, Müller 2000, 97ff.].

¹²³ Hervorhebungen im Original.

¹²⁴ Hervorhebung im Original.

¹²⁵ <http://www.nsf.gov/>

NSF-Report von McCormick und DeFanti, veröffentlicht in der November-Ausgabe von *IEEE Computer Graphics* 1987, bewertete Scientific Visualization als bedeutendes wissenschaftliches Werkzeug mit einer Vielzahl von Anwendungsgebieten von der Modellierung von Molekülen über Medizin, Geowissenschaften, Mathematik bis zur Astrophysik [McCormick, DeFanti 1987]. DeFanti und Brown, die Koautoren des Reports, äußerten sich einige Jahre darauf überrascht über den durch diesen ausgelösten Entwicklungsschub (vgl. [DeFanti, Brown 1994]). Scientific Visualization hat sich seither zu einem komplexen Gebiet mit vielfältigen Ansätzen und spezifischen Herausforderungen entwickelt (für einen Überblick vgl. bspw. [Rosenblum et al. 1994]).

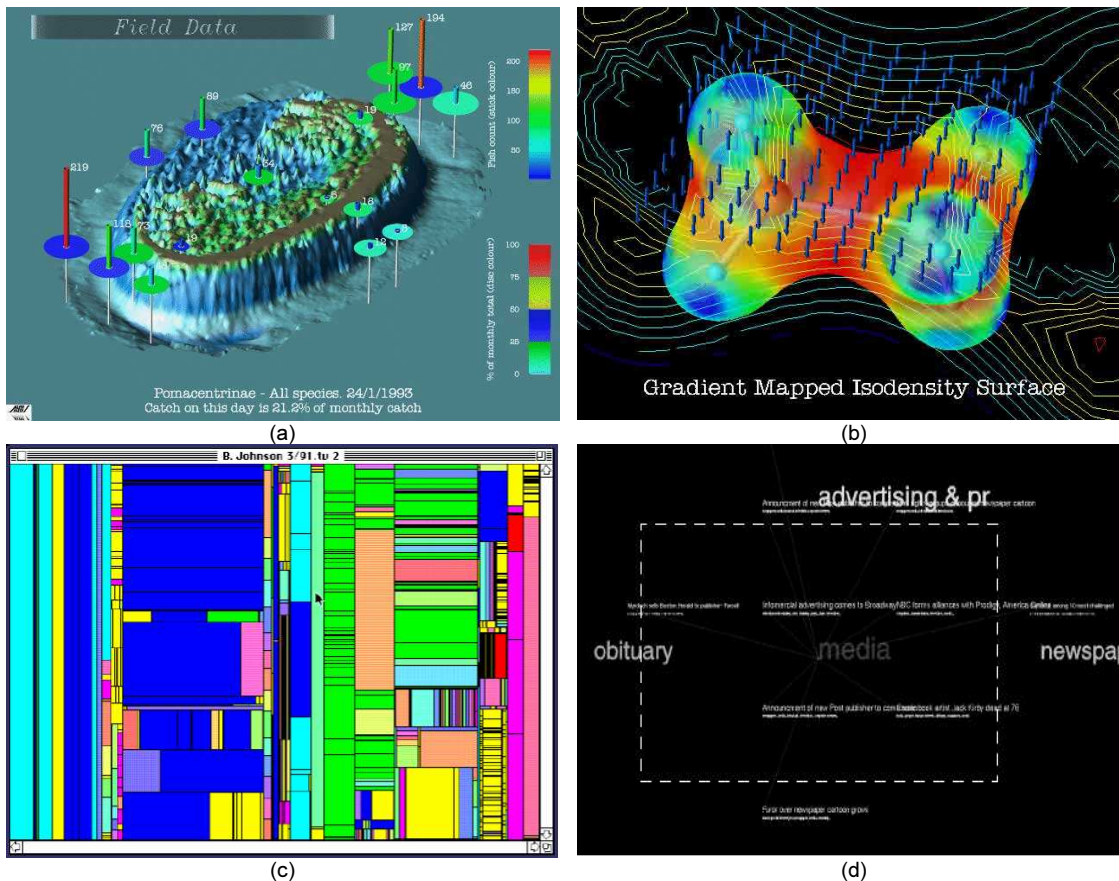


Abb. 4.2 - Beispiele für Scientific Visualization (oben) und für Informationsvisualisierung (unten):
 (a) Ozeanographie; (b) Chemie¹²⁶; (c) Darstellung eines Dateisystems durch eine Tree-Map¹²⁷;
 (d) Galaxy of News zur Navigation über Dokumentensammlungen¹²⁸.

Scientific Visualization wird oft in Verbindung gebracht mit der Analyse sehr großer Datenmengen [Willim 1989, 631] [Däßler 1999], die zudem zumeist einen *physikalischen Bezug* aufweisen [Card et al. 1999, 6f., 38] [Spence 2001, 4]. Obwohl dies für viele Anwendungsbeispiele - etwa aus dem Bereichen Strömungsmechanik, Medizin oder Astrophysik - zutrifft, ist zu beachten, dass Visualisierung in wissenschaftlichen Prozessen auch für die Auswertung kleinerer und kleiner Datenmengen ein wichtiges Instrument darstellt, ebenso wie auch *nichtphysikalische* Daten (etwa typische Variablen im Bereich der Psychologie) von wissenschaftlicher Relevanz sein können.

Nach [Schumann, Müller 2000, 5f.] lassen sich drei zentrale Bereiche für den Einsatz der

¹²⁶ Beide Abbildungen wurden der Homepage für das Visualisierungswerkzeug OpenDX entnommen, auf der sich eine Vielzahl weiterer Beispiele befindet (<http://www.opendx.org/highlights.php>).

¹²⁷ Abbildung übernommen aus [Shneiderman 1992].

¹²⁸ Abbildung übernommen aus [Rennison 1994, Fig.1d].

Visualisierung wissenschaftlicher Daten unterscheiden: Als *explorative Analyse* wird die Suche nach zunächst unbekanntem Strukturen und Eigenschaften in Daten bezeichnet; ihr Ziel ist es, zu Hypothesen über die in den Daten enthaltenen Eigenschaften zu gelangen. Die *konfirmative Analyse* dient hingegen zur zielgerichteten Untersuchung von Daten mit dem Ziel, bereits vorhandene Hypothesen zu verifizieren. Das dritte Einsatzgebiet der Visualisierung besteht in *Präsentation und Kommunikation* mit dem Ziel einer anschaulichen Vermittlung gefundener Ergebnisse an Dritte.

4.1.3 Die Visualisierung abstrakter Daten – Informationsvisualisierung

In den 90er Jahren formierte sich eine neue Forschungsrichtung innerhalb des Gebietes der Visualisierung, die als *Informationsvisualisierung (Information Visualization)* bezeichnet wird. Die Abgrenzung dieses Gebietes ist dabei „[...] *noch nicht befriedigend gelöst und wird keinesfalls einheitlich vorgenommen*“ [Schumann, Müller 2000, 341]. Entsprechend finden sich in der Literatur unterschiedliche Definitionen (vgl. bspw. die Zusammenstellung in [Däßler 1999]), die auch aus einer gewissen Überlappung mit dem Gebiet der Scientific Visualization [Spence 2001, 4] resultieren mögen. An dieser Stelle soll die Definition von Card et al. wiedergegeben werden:

- ▶ „*Information Visualization: The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.*“ [Card et al. 1999, 7]

Das Gebiet der Informationsvisualisierung befasst sich mit der Entwicklung von Techniken und Werkzeugen zur Visualisierung abstrakter Konzepte oder Zusammenhänge, die nicht unbedingt einen direkten physikalischen Bezug besitzen [Spence 2001, 4] [Card et al. 1999, 7]. Im Anschluss an die Behandlung grundlegender Konzepte der „klassischen“ Visualisierung von Daten werden in Kap. 4.8 einige Beispiele aus diesem Gebiet aufgeführt (vgl. auch Abb. 4.2c und 4.2d).

4.2 Anforderungen an eine Datenvisualisierung

Nicht jede visuelle Repräsentation ist gleichermaßen für die Analyse oder Präsentation der zugrundeliegenden Daten geeignet. Um die Erzeugung von Bildern zu vermeiden, die zu falschen Schlussfolgerungen bezüglich der dargestellten Daten führen, müssen die jeweils geeigneten Techniken vielmehr für jedes Anwendungsgebiet sorgfältig ausgewählt werden. Als zentrale Kriterien einer Visualisierung wurden von [Mackinlay 1986] die Forderungen nach Expressivität und Effektivität formuliert; [Schumann, Müller 2000] fordern zusätzlich zu diesen die Angemessenheit einer Visualisierung.

4.2.1 Expressivität

Das *Expressivitätskriterium* einer Visualisierung [Schumann, Müller 2000, 9ff.] [Mackinlay 1986] ist die Forderung, nur die in den Daten enthaltenen Informationen darzustellen¹²⁹. Um Verfälschungen bei der Interpretation der Daten auszuschließen, bildet das Expressivitätskriterium das zentrale Kriterium für jede Visualisierung; es kann dabei für eine gegebene Datenmenge oft durch unterschiedliche Darstellungsformen erreicht werden. Ein klassisches Beispiel hierzu findet sich in [Mackinlay 1986]: Die Darstellung in Abb. 4.3a verstößt gegen das Expressivitätskriterium, da die gewählte Darstellungsform als Balken-

¹²⁹ Bei [Mackinlay 1986] findet sich ein Ansatz zur Formalisierung von Expressivität (Expressiveness) und Effektivität (Effectiveness) graphischer Präsentationen. Ausgehend von der Beobachtung, dass jede Kommunikation auf gewissen Konventionen zwischen den Teilnehmern beruht, setzt Mackinlay graphische Präsentationen mit Sätzen einer graphischen Sprache gleich. Er definiert: „*A set of facts is expressible in a language if it contains a sentence that*

(1) *encodes all the facts in the set,*
 (2) *encodes only the facts in the set.*“

diagramm eine fälschliche Interpretation der Balkenlängen als bedeutungstragend impliziert (Schweden erscheint als „mehr“ oder „besser“ als bspw. Japan); hier ist nach Mackinlay einer Darstellung der Daten als Scatterplot (vgl. Kap. 4.5.5) wie in Abb. 4.3b der Vorzug zu geben.

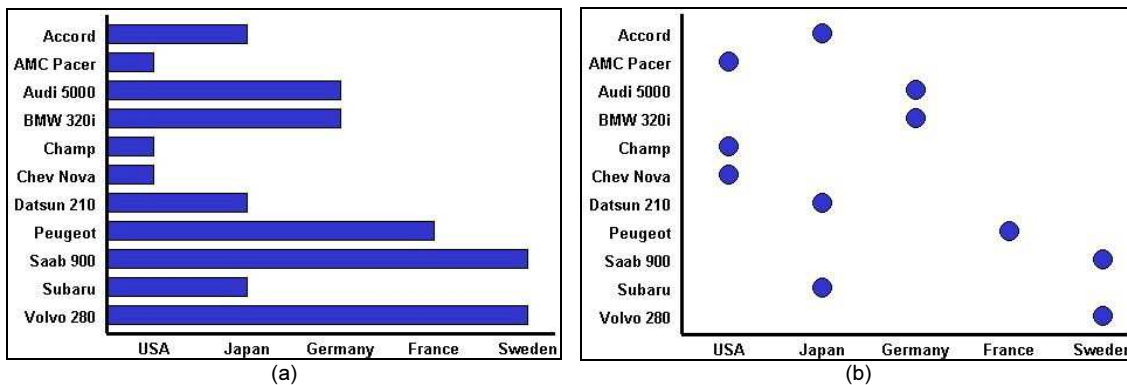


Abb. 4.3 - Expressivitätskriterium¹³⁰: (a) Verletzung; (b) korrekte Darstellung.

4.2.2 Effektivität

Das *Effektivitätskriterium* [Schumann, Müller 2000, 11f.] [Mackinlay 1986] fordert die Auswahl der jeweils am besten geeigneten visuellen Darstellungsform für eine gegebene Datenmenge. Für eine effektive Visualisierung sind Ziel und Anwendungskontext der Visualisierung ebenso von Bedeutung wie die individuellen Fähigkeiten des Betrachters oder die Eigenschaften des verfügbaren Ausgabemediums. Die Darstellung in Abb. 4.4 ist nicht effektiv, da für eine korrekte Interpretation der Symbole die Legende erforderlich ist; eine effektivere Darstellung wäre hier nach Schumacher und Müller eine Verwendung unterschiedlich großer Kreissignaturen.

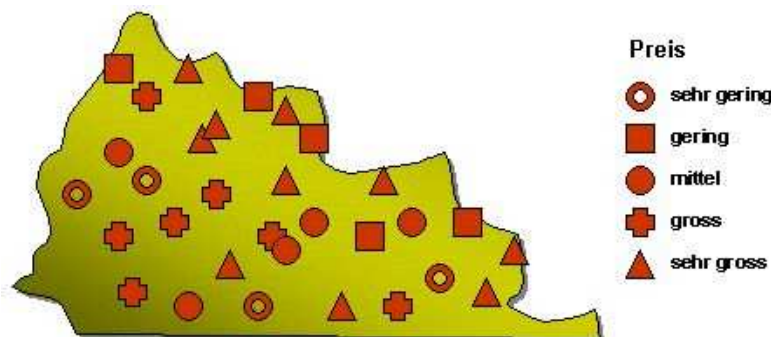


Abb. 4.4 - Beispiel für eine Verletzung des Effektivitätskriteriums¹³¹.

4.2.3 Angemessenheit

Das *Angemessenheitskriterium* einer Visualisierung [Schumann, Müller 2000, 12] schließlich fordert, dass der Aufwand zur Erstellung einer Visualisierung, bspw. die erforderliche Rechenleistung, der Zeitaufwand oder die kognitive Belastung des Anwenders, in einem angemessenen Verhältnis zum durch diese erreichbaren Nutzen stehen soll. [Meyer 1999, 162ff.] verweist darauf, dass bei einem Einsatz von Visualisierung für die Informationsvermittlung im Management das Verhältnis von Kosten und Nutzen nicht zu objektivieren sei und ein Einsatz „[...] daher der (subjektiven) Abwägung durch den Entscheider überlassen bleiben“¹³² müsse.

¹³⁰ Abbildungen erstellt in Anlehnung an [Mackinlay 1986, Fig. 11 und 12].

¹³¹ Abbildung erstellt in Anlehnung an [Schumann, Müller 2000, 11, Abb. 1.2].

¹³² Zitat aus [Meyer 1999, 166].

4.3 Visualisierungsprozess und Unterscheidungsmerkmale von Visualisierungstechniken

4.3.1 Der Visualisierungsprozess

Der Prozess der Visualisierung von Daten [Schumann, Müller 2000, 15ff.] kann vereinfacht als eine Abfolge von Teilschritten beschrieben werden, die die Ausgangsdaten sukzessive transformieren, bis schließlich eine visuelle Präsentation der Daten vorliegt. Die drei wesentlichen Schritte werden als Filtering (Datenaufbereitung), Mapping (Erzeugung eines Geometriemodells) und Rendering (Bildgenerierung) bezeichnet; sie können in einer sog. *Visualisierungspipeline* angeordnet werden (vgl. Abb. 4.5):



Abb. 4.5 - Die Stufen der Visualisierungspipeline¹³³.

Durch die einzelnen Schritte werden die Ausgangsdaten, auch als Rohdaten bezeichnet, zunächst in aufbereitete Daten, dann in Geometriedaten und schließlich in Bilddaten transformiert (vgl. Abb. 4.6). In einem ersten Schritt, der als *Filtering* oder *Datenaufbereitung* bezeichnet wird, werden die Rohdaten zu aufbereiteten Daten transformiert und für die Visualisierung vorbereitet. Dabei werden die Daten erforderlichenfalls ergänzt oder durch eine Auswahl aus der Gesamtmenge reduziert; ferner werden in diesem Schritt eventuelle Fehler in den Daten korrigiert. Die durch das Filtering entstandenen aufbereiteten Daten werden in einem zweiten Schritt, der als *Mapping* bezeichnet wird, auf geometrische Primitive einschließlich zugehöriger Attribute wie Farbe abgebildet; das Mapping besitzt daher entscheidende Bedeutung für die visuelle Repräsentation der Daten. Die anschließende Umwandlung der durch das Mapping entstandenen Geometriedaten in Bilddaten wird als *Rendering* bezeichnet; in diesem Schritt werden bspw. Blickpunkt und Blickrichtung oder Projektionen festgelegt.

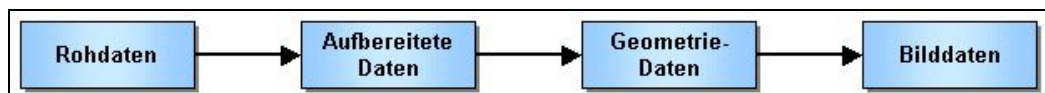


Abb. 4.6 - Datenfluss in der Visualisierungspipeline¹³⁴.

Nach [Schumann, Müller 2000, 175ff.] können Visualisierungstechniken anhand von drei Kriterien unterschieden werden, die sich auf die Dimensionalität des Darstellungsraumes, die Zeitabhängigkeit der Darstellung sowie ihre Vollständigkeit beziehen. Diese drei Kriterien werden nachfolgend kurz dargestellt.

4.3.2 Zwei- oder dreidimensionale Darstellungen

Trotz aller Fortschritte in der Leistungsfähigkeit von Hardware sind heutige Computerbildschirme immer noch zweidimensional¹³⁵. Dieses flache Ausgabemedium kann entweder für die Erzeugung zweidimensionaler, also ebenfalls flacher, oder dreidimensionaler Darstellungen genutzt werden, wobei im letzteren Fall der Eindruck räumlicher Tiefe durch geeignete perspektivische Darstellungen erreicht wird. Entsprechend können Visualisierungstechniken danach unterschieden werden, ob sie zweidimensionale (2D) oder dreidimensionale (3D) Darstellungstechniken verwenden. Bei 2D-Darstellungstechniken wird ein Bild aus zweidimensionalen graphische Elementen zusammengesetzt. 3D-Darstellungstechni-

¹³³ Abbildung erstellt in Anlehnung an [Schumann, Müller 2000, 15, Abb. 2.1].

¹³⁴ Abbildung erstellt in Anlehnung an [Schumann, Müller 2000, 17, Abb. 2.2].

¹³⁵ Eine Ausnahme bilden Ausgabegeräte wie Head-mounted Displays aus dem Bereich der sog. Virtual Reality (vgl. bspw. [Bormann 1994] [Kalawsky 1993]), auf den hier nicht eingegangen werden soll.

ken erschließen hingegen eine dritte räumliche Dimension durch den Einsatz dreidimensionaler graphischer Elemente, die auf die Ausgabeebene projiziert werden. Dieser Ansatz stellt erhöhte Anforderungen an Hard- und Software; zudem können perspektivische Verzerrungen und insbesondere Verdeckungen von Teilen des Bildes auftreten, so dass Möglichkeiten zur interaktiven Veränderung der Betrachtungsperspektive erforderlich werden. 2D-Darstellungstechniken können nur zwei räumliche Dimensionen für die Visualisierung nutzen, sind dafür jedoch nicht mit Projektionen - und nur in geringem Maße mit Verdeckungen¹³⁶ - konfrontiert und erlauben meist eine korrekte Ablesung der dargestellten Datenwerte. Aufgrund ihrer jeweiligen Vor- und Nachteile besitzen beide Techniken ihre Berechtigung; die Entscheidung für einen Einsatz wird daher meist problembezogen getroffen [Schumann, Müller 2000, 175f.].

4.3.3 Statische oder dynamische Darstellungen

Ebenfalls problembezogen muss entschieden werden, ob statische oder dynamische Darstellungstechniken verwendet werden. Statische Darstellungstechniken sind insofern statisch, als sie nur über Nutzerinteraktionen verändert werden; sie erlauben damit eine Betrachtung des jeweils aktuellen Bildes ohne zeitliche Begrenzung und eignen sich für quantitative Aussagen über die visualisierten Daten. Dynamische Darstellungstechniken verändern hingegen die Darstellung automatisch über die Zeit und erlauben so die Beobachtung von Veränderungen in den Datenwerten [Schumann, Müller 2000, 176f.].

4.3.4 Vollständige oder unvollständige Darstellungen

Die vollständige Visualisierung einer gegebenen Datenmenge führt bei großen Datenmengen schnell zu überladenen, wenig aussagekräftigen Bildern. Eine Verbesserung kann durch die Zusammenfassung einzelner Ausprägungen zu Klassen erreicht werden; genügt dies nicht, kann auf unvollständige Darstellungen zurückgegriffen werden, die nur eine echte Teilmenge¹³⁷ der gesamten Daten visualisieren. Der hieraus resultierende Informationsverlust muss ausgeglichen werden, dies kann durch automatische Veränderung des visualisierten Ausschnittes über die Zeit erfolgen oder durch Interaktionstechniken, die einen nutzerdefinierten Abruf von weiteren Details erlauben [Schumann, Müller 2000, 177f.].

4.4 Dimensionalität und Visualisierung

4.4.1 Begriffsklärungen

Für die Zwecke einer Visualisierung von Daten ist eine genauere Betrachtung und Aufteilung ihrer Dimensionalität sinnvoll. So unternehmen [Schumann, Müller 2000, 27ff.] eine getrennte Betrachtung der Dimensionen des Raumes, in dem die Daten erhoben wurden (sog. unabhängige Variablen), sowie der Dimensionen, die die erhobenen Daten charakterisieren (sog. abhängige Variablen). Die Autoren unterscheiden dabei zunächst zwischen Beobachtungsraum und Merkmalen:

Beobachtungsraum ▶ Als *Beobachtungsraum* wird der Raum bezeichnet, in dem Daten erhoben werden. Der Beobachtungsraum ist stets ein metrischer Raum; hingegen kann davon abstrahiert werden, ob es sich um einen physikalischen oder einen abstrakten Raum handelt sowie auf welche Weise die Datenerhebung erfolgt, also ob es sich um beobachtete, gemessene, berechnete oder künstlich entworfene

¹³⁶ Verdeckungen treten bei 2D-Darstellungstechniken nur dann auf, wenn sich einzelne Elemente überlagern, etwa wenn die Zahl der zugleich darzustellenden Objekte sehr groß wird.

¹³⁷ Dies kann bspw. durch die Auswahl von Datensätzen oder von Attributen erfolgen.

Daten handelt.

- Merkmale ▶ Die verschiedenen Größen, die im Beobachtungsraum erhoben werden, werden als *Merkmale* bezeichnet.

Eine genauere Betrachtung von Beobachtungsraum und Merkmalen führt zu den Begriffen der Beobachtungspunkte sowie der unabhängigen und abhängigen Variablen:

- Beobachtungspunkte ▶ Diejenigen Punkte des Beobachtungsraumes, für die Daten vorhanden sind, werden als *Beobachtungspunkte* bezeichnet; ihr Verbund¹³⁸ und ihr jeweiliger Wirkungskreis¹³⁹ sind bei der Visualisierung zu berücksichtigen.

- Unabhängige Variablen ▶ Die verschiedenen Dimensionen, die den Beobachtungsraum aufspannen, werden als *unabhängige Variablen* bezeichnet.

- Abhängige Variablen ▶ Im Gegensatz dazu werden die erhobenen Merkmale, da ihre Werte von der jeweiligen Position im Beobachtungsraum abhängt, als *abhängige Variablen* bezeichnet.

Aufbauend auf diesen Begriffen kann nun die Dimensionalität von abhängigen und unabhängigen Variablen begrifflich genauer gefasst werden. [Schumann, Müller 2000, 171] verwenden den Begriff *multi* zur Abgrenzung zwischen den Ausprägungen *1* bzw. *2 und mehr (multi)*¹⁴⁰. Darauf aufbauend wird der Begriff der Multiparameterdaten eingeführt:

- Multiparameterdaten ▶ *Multiparameterdaten* liegen vor, wenn mindestens zwei abhängige Variablen vom Typ Skalar gegeben sind. An die Eigenschaften des Beobachtungsraumes werden keine gesonderten Anforderungen gestellt, so dass dieser auch eindimensional sein kann.

Die Autoren verweisen in diesem Zusammenhang darauf, dass in der Literatur auch die Begriffe *mehrdimensional* und *multivariat*¹⁴¹ und zudem in nicht einheitlicher Weise verwendet werden. Sie unterscheiden wie folgt zwischen mehrdimensionalen Daten und multivariaten Daten [Schumann, Müller 2000, 171f.]:

- Mehrdimensionale Daten ▶ Der Begriff *mehrdimensional* ist unabhängigen Variablen vorbehalten, bezieht sich also auf den Beobachtungsraum. Entsprechend beschreiben mehrdimensionale Daten keine Merkmale, sondern lediglich einen Beobachtungsraum.

- Multivariate Daten ▶ Im Gegensatz dazu ist der Begriff *multivariat* für abhängige Variablen zu verwenden und bezieht sich also auf die Merkmale. Danach enthalten Daten, die als *multivariat* bezeichnet werden, lediglich eine Beschreibung der Merkmale, nicht jedoch eine Beschreibung des Beobachtungsraumes.

¹³⁸ Beobachtungspunkte können in bestimmten Anordnungen (bspw. regelmäßige oder unregelmäßige Gitter) oder beliebiger Verteilung (gitterfrei) vorliegen (zu verschiedenen Gittertypen siehe [Schumann, Müller 2000, 34f.]).

¹³⁹ Der Wirkungskreis eines Beobachtungspunktes wird danach unterschieden, ob die an ihm erhobenen Daten nur für den Beobachtungspunkt Gültigkeit besitzen (punktuelle Wirkungskreis), für eine gewisse Umgebung (lokaler Wirkungskreis) oder für den gesamten Beobachtungsraum (globaler Wirkungskreis) [Schumann, Müller 2000, 29f.].

¹⁴⁰ [Spence 2001, 34] schlägt hingegen eine Trennung zwischen den Ausprägungen *one*, *two* und *many* vor.

¹⁴¹ So finden sich bspw. bei [Spence 2001, 34ff.] die Begriffe *univariate data*, *bivariate data*, *trivariate data* und *hypervariate data*.

Je nach Ansatzpunkt ist allerdings eine getrennte Behandlung von unabhängigen und abhängigen Variablen nicht immer möglich. So dient etwa die Anwendung VisDB von Keim und Kriegel (vgl. Kap. 4.6.4 sowie Abb. 4.11) zur Visualisierung von Daten aus Datenbanktabellen. Dabei korrespondiert die Zahl der ausgewählten Attribute mit der Zahl der Dimensionen; die Autoren verweisen darauf, dass bei einer solchen Sichtweise eine Aufteilung in unabhängige und abhängige Dimensionen oft unklar ist [Keim, Kriegel 1994].

4.4.2 Herausforderung Dimensionalität und Wertebereich

Besondere Herausforderungen für die Visualisierung entstehen, wenn Beobachtungsraum und / oder Merkmale durch eine hohe Dimensionalität gekennzeichnet sind oder die möglichen Ausprägungen von Merkmalen einen großen Wertebereich umfassen. Die Darstellung eines Beobachtungsraumes, der durch mehr als zwei unabhängige Variablen aufgespannt wird, erfordert eine Projektion und ist daher mit dem Problem potentieller Mehrdeutigkeiten und Verdeckungen konfrontiert. Ferner ist eine geeignete Darstellung besonders dann schwierig zu erreichen, wenn die Zahl der Werte, die pro Beobachtungspunkt dargestellt werden müssen, sehr groß wird. Einen weiteren Faktor bildet die Anzahl der Ausprägungen, die ein Merkmal annehmen kann; ein großer Wertebereich eines Merkmals kann die visuelle Unterscheidung individueller Ausprägungen zusätzlich erschweren [Schumann, Müller 2000, 29ff.].

Nachfolgend werden beispielhaft einige Techniken der Visualisierung vorgestellt. Ausgehend von Basistechniken (Kap. 4.5) werden Ansätze zur Visualisierung multivariater Daten (Kap. 4.6) sowie Aspekte der Darstellung des Raum- und Zeitbezuges von Multiparameterdaten (Kap. 4.7) behandelt¹⁴². Kap. 4.8 stellt ausgewählte Beispiele aus dem Gebiet der Informationsvisualisierung vor.

4.5 Basistechniken

Für die Visualisierung von Daten existiert eine Vielzahl von Basistechniken, die nicht nur im wissenschaftlichen Bereich eingesetzt werden. Die vorgestellten Diagrammformen¹⁴³ basieren primär auf den visuellen Variablen Position, Größe und Orientierung.

4.5.1 Balken- und Säulendiagramme

Balkendiagramme (Abb. 4.7a) visualisieren Werte in Form horizontaler Balken; *Säulendiagramme* (Abb. 4.7b) entstehen hingegen durch eine Darstellung der Werte als senkrechte Säulen. Balkendiagramme eignen sich insbesondere für Rangfolgenvergleiche, während Säulendiagramme für die Darstellung der Entwicklung von Werten über die Zeit oder Häufigkeitsvergleiche eingesetzt werden können (vgl. [Meyer 1999, 42f.]). Nach [Meyer 1999, 169] sind Balkendiagramme gegenüber Säulendiagrammen „*vielfach vorzuziehen, da Individuen zur Überbewertung der Säulenlänge neigen*“. Ferner finden sich zwei Varianten von Säulendiagrammen mit dreidimensionalen Säulen, die die dritte Dimension unterschiedlich nutzen. *2½D-Säulendiagramme* (Abb. 4.7c) verwenden ein zweidimensionales Koordinatensystem und bieten außer einem optischen Effekt keinen zusätzlichen Effektivitätsgewinn. *3D-Säulendiagramme* (Abb. 4.7d) können hingegen zur gleichzeitigen Darstellung mehrerer Variablen genutzt werden; allerdings besteht hier die Gefahr der Verdeckung von

¹⁴² Die Aufteilung in Basistechniken, Techniken für multivariate Daten und die Behandlung von Raum- und Zeitbezügen orientiert sich an der Systematik in [Schumann, Müller 2000]; von der dort ebenfalls vorgenommenen Darstellung spezieller Visualisierungstechniken für Volumendaten sowie für Strömungsdaten wird in dieser Arbeit allerdings abgesehen (für eine ausführliche Beschreibung dieser Gebiete siehe [Schumann, Müller 2000, 251ff.]).

¹⁴³ Für Beispiele, die auf Struktur und Form, Farbe sowie Texturen basieren, vgl. [Schumann, Müller 2000, 141ff.].

einzelnen Variablenwerten (vgl. [Schumann, Müller 2000, 134f.] [Meyer 1999, 43f.]).



Abb. 4.7 - Verschiedene Diagrammformen:

- (a) Balkendiagramm; (b) Säulendiagramm; (c) 2 1/2-D-Säulendiagramm;
 (d) 3D-Säulendiagramm; (e) Histogramm; (f) kumulatives Histogramm;
 (g) Kreisdiagramm; (h) Tortendiagramm; (i) Liniendiagramm;
 (j) Kurvendiagramm; (k) Scatterplot; (l) Verbunddiagramm.

4.5.2 Histogramme

Histogramme (vgl. [Schumann, Müller 2000, 135ff.]) bilden eine spezielle Form der Säulendiagramme, die zur Darstellung von Häufigkeitsverteilungen eingesetzt werden (Abb. 4.7e). Quantitative Daten werden hierzu in Klassen eingeteilt und die Häufigkeiten in den Klassen dargestellt. Eine oft verwendete Variante sind *kumulative Histogramme* (Abb. 4.7f), bei denen jede Säule die Summe der Datenwerte der zugehörigen Klasse und aller vorangehenden (links von dieser Säule dargestellten) Klassen präsentiert.

4.5.3 Kreisdiagramme

Kreisdiagramme (Abb. 4.7g) können zur Visualisierung des Verhältnisses einzelner Werte zu einer Gesamtheit eingesetzt werden; wird ein Kreisdiagramm perspektivisch als 2 1/2-D-Diagramm dargestellt, wird es als *Kuchen-* oder *Tortendiagramm* (Abb. 4.7h) bezeichnet (vgl. [Meyer 1999, 47ff.] [Willim 1989, 61]). Für einen sinnvollen Einsatz von Kreisdiagrammen ist eine bestimmte Mindestgröße der Darstellung erforderlich; zudem sollten nicht

mehr als fünf oder sechs Variablen zugleich dargestellt werden [Schumann, Müller 2000, 138].

4.5.4 Linien- und Kurvendiagramme

Liniendiagramme (Abb. 4.7i) stellen Datenwerte in einem zweidimensionalen Koordinatensystem dar, wobei jeweils benachbarte Punkte mit Liniensegmenten verbunden werden; *Kurvendiagramme* (Abb. 4.7j) interpolieren zwischen den Punkten statt der Verbindungslinien Kurvensegmente. Beide Formen können auch mehrere Variablen zugleich darstellen und eignen sich, um Trends, lokale Strukturen sowie die allgemeine Verteilung von Werten zu erkennen [Schumann, Müller 2000, 133f.] [Willim 1989, 55ff.].

4.5.5 Scatterplots

Scatterplots oder *Punktediagramme* (Abb. 4.7k) visualisieren die Werte zweier Variablen als graphische Symbole in einem zweidimensionalen Koordinatensystem. Sie eignen sich besonders, um Korrelationen oder Cluster in den dargestellten Werten abschätzen zu können. Scatterplots können auf die Darstellung von mehr als zwei Variablen erweitert werden, wenn mehrere Variablen die selben Wertebereiche besitzen; zur Unterscheidung der Variablen kann dann auf unterschiedliche Farben und/oder Symbole zurückgegriffen werden [Schumann, Müller 2000, 131f.] [Meyer 1999, 45f.].

4.5.6 Verbunddiagramme

Verbunddiagramme (Abb. 4.7l) entstehen durch die Kombination verschiedener Diagrammtypen [Meyer 1999, 52f.]. Ein typisches Beispiel ist die gleichzeitige Visualisierung von Niederschlags- und Temperaturwerten in einem Diagramm, wobei Niederschläge als Säulen- und Temperaturen als Liniendiagramm dargestellt werden.

4.6 Visualisierung multivariater Daten

„However, an image has only three dimensions. And this barrier is impassable“ [Bertin 1977/1981]. Diese Feststellung des französischen Kartographen Bertin, die noch aus der Zeit vor der computergestützten Visualisierung stammt, beschreibt treffend die Herausforderungen, die auftreten, wenn Merkmale mit mehreren Dimensionen in einem Bild zu codieren sind. Nachfolgend werden einige ausgewählte Basiskonzepte vorgestellt, die diese Problematik adressieren¹⁴⁴. Zu beachten ist, dass diese Konzepte jeweils spezifische Vor- und Nachteile¹⁴⁵ aufweisen, so dass ihr Einsatz anwendungsbezogen entschieden werden muss [Schumann, Müller 2000, 247].

4.6.1 Panel-Matrizen

Visualisierungsformen, die bivariate Darstellungen in Form einer Matrix anordnen, werden als *Panel-Matrizen* bezeichnet (vgl. [Schumann, Müller 2000, 179]). Stellvertretend¹⁴⁶ für dieses Konzept soll hier die Scatterplot-Matrix vorgestellt werden. Scatterplots (vgl. Kap. 4.5.5) dienen zur Visualisierung der Beziehungen von jeweils zwei Variablen zueinander; das Konzept der *Scatterplot-Matrix* (vgl. [Schumann, Müller 2000, 179ff.] [Spence 2001, 42f.]) erweitert diese Darstellungsform auf mehrere Variablen. Für jede mögliche Zweierkombination von Variablen wird ein Scatterplot erzeugt; die einzelnen Diagramme werden in Form einer Matrix angeordnet (vgl. Abb. 4.8). Damit können Korrelationen zwischen je-

¹⁴⁴ Vgl. bspw. auch die Arbeiten von [Mihalisin et al. 1991] sowie von [Spoerri 1993].

¹⁴⁵ Für einen Vergleich der Basiskonzepte im Hinblick auf die jeweils darstellbare Datenmenge sowie die jeweilige Eignung für Aufgaben der Identifikation und das Erkennen von Zusammenhängen vgl. [Schumann, Müller 2000, 211ff.].

¹⁴⁶ Zu den weiteren Varianten der Panel-Matrizen zählen die Prosection Matrix [Spence 2001, 168 ff.] sowie die Technik der Hyperslices (vgl. [Schumann, Müller 2000, 183ff.]).

weils zwei Variablen gut wahrgenommen werden; allerdings ist wegen des erforderlichen Platzbedarfes für die einzelnen Diagramme schnell eine Obergrenze der zugleich visualisierbaren Variablen erreicht.

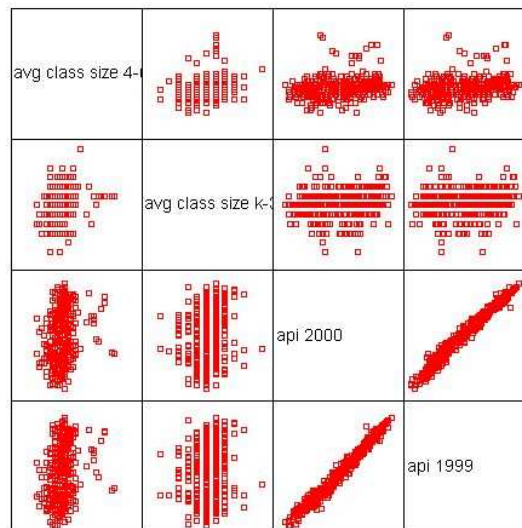


Abb. 4.8 - Scatterplot-Matrix¹⁴⁷.

4.6.2 Streckenzüge

Streckenzüge (vgl. [Schumann, Müller 2000, 185ff.]) konstruieren eine eigene, entsprechend skalierte Achse für jede Variable des Merkmalsraumes, tragen die Ausprägungen eines Merkmals als Punkte auf die jeweiligen Achsen auf und verbinden diese mit Linien. Je nach Ausrichtung der Achsen lassen sich hierbei die Techniken der Sternförmigen Koordinaten und der Parallelen Koordinaten unterscheiden.

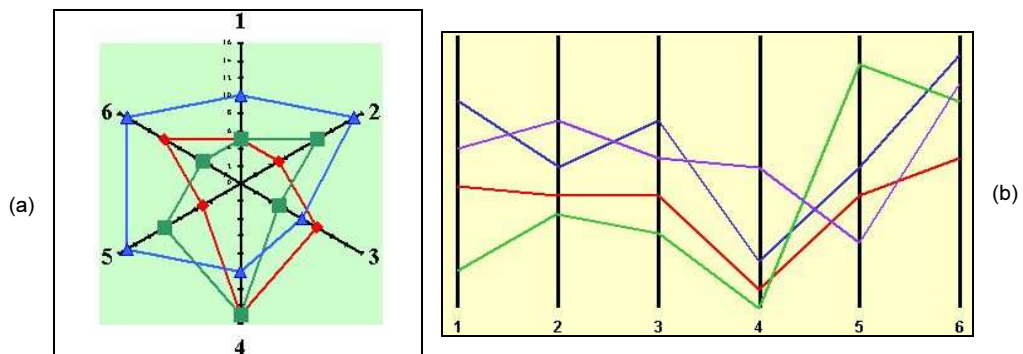


Abb. 4.9 - Das Prinzip von Sternförmigen Koordinaten (a) und Parallelen Koordinaten (b).

▪ Sternförmige Koordinaten

Sternförmige Koordinaten (Netzdiagramme, Starplots) (vgl. [Schumann, Müller 2000, 185f.] [Meyer 1999, 55] [Spence 2001, 50]) erlauben die Visualisierung einiger Merkmale mit mehreren Dimensionen in einer Darstellung, indem sie die Achsen, die die verschiedenen Dimensionen aufspannen, sternförmig anordnen (vgl. Abb. 4.9a). Die Werte jedes Merkmals werden dabei auf die entsprechenden Achsen aufgetragen und über Linienzüge miteinander verbunden, so dass für jedes Merkmal eine bestimmte Struktur entsteht.

▪ Parallele Koordinaten

Parallele Koordinaten [Inselberg 1997] verwenden ebenfalls für jede Dimension eine ei-

¹⁴⁷ Abbildung übernommen von <http://www.ats.ucla.edu/stat/spss/webbooks/reg/chapter1/spssreg1.htm>.

gene Koordinatenachse, allerdings werden hier die Achsen parallel zueinander angeordnet (vgl. Abb. 4.9b). Da bei dieser Technik Datensätze mit identischen Werten auf den selben Streckenzug abgebildet und so nicht voneinander unterschieden werden können, adressieren Erweiterungen bspw. eine Integration von Histogrammen in die Darstellung (Parahistogramme, vgl. [Schumann, Müller 2000, 186ff.]).

4.6.3 Ikonen

Eine weitere Variante der Visualisierung multivariater Daten stellen *Ikonen* dar (vgl. [Schumann, Müller 2000, 192ff.]). Ikonen, auch als *Glyphen* bezeichnet, sind graphische Primitive, die Werte bspw. über Längen, Formen, Winkel oder Farben abbilden und die exakt positioniert werden können. Abb. 4.10 zeigt Beispiele¹⁴⁸ für unterschiedliche Arten von Ikonen.

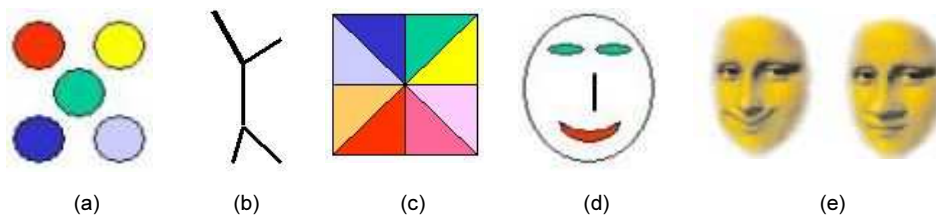


Abb. 4.10 - Verschiedene Ikonen¹⁴⁹: (a) Kreispalette; (b) Stick-Figure-Ikone; (c) Farbkone; (d) Chernoff Face; (e) verzerrte Mona-Lisa-Gesichter.

Eine interessante Variante der Ikonen stellen die Gesichtskonon dar. Bereits aus den 70er Jahren des 20. Jahrhunderts stammt das Konzept der *Chernoff Faces* (vgl. Abb. 4.10d), bei denen die menschlichen Fähigkeiten zur Unterscheidung von Gesichtsmerkmalen genutzt werden. Dazu werden die Ausprägungen verschiedener Attribute über entsprechend dargestellte Elemente wie Kopfform, Mund oder Nase codiert und zu einem Gesicht zusammengesetzt. Chernoff definierte 18 unterschiedliche Gesichtselemente, die entsprechend der darzustellenden Datenwerte jeweils innerhalb vorgegebener Grenzen verzerrt werden, so dass multivariate Daten mit bis zu 18 Dimensionen dargestellt werden können. Er berichtet von Anwendungen dieses Konzeptes anhand zweier Beispiele, der Darstellung fossiler Daten mit 6 Dimensionen sowie geologischer Daten (Mineralgehalte anhand von 12 Variablen) [Chernoff 1973]. Eine Variante dieses Ansatzes besteht in der Verwendung vertrauter Gesichter wie dem der Mona Lisa, die entsprechend verzerrt werden [Alexa, Müller 1998] (vgl. Abb. 4.10e). Die Autoren, die auch eine Visualisierung über Hände (Handikonon) vorstellen, ordnen die Mona Lisa-Gesichter auf einer Karte der USA an, um so für verschiedene Städte zugleich Eigenschaften wie Kriminalitätsrate, Wirtschaftslage und Gesundheitsversorgung zu visualisieren (vgl. Abb. 4.12a und 4.12b).

4.6.4 Pixelbasierte Techniken

Der in gewisser Weise radikalste Ansatz zur Ausnutzung des Darstellungsmediums Bildschirm ist die Abbildung jedes Datenwertes auf jeweils genau *ein* Pixel. Die Anwendung VisDB [Keim, Kriegel 1994] wurde für die visuelle Auswertung von großen Datenmengen in Datenbanken entwickelt; dabei werden Pixel so gefärbt und arrangiert, dass sie die Relevanz von Datensätzen auf eine konkrete Anfrage hin darstellen. Gelbe Pixel indizieren Datensätze, die die Bedingungen der Anfrage zu 100 Prozent erfüllen und werden in der Mitte der Visualisierung dargestellt; Datensätze, die die Anfrage weniger gut erfüllen, werden in anderen Farben spiralförmig um dieses Zentrum angeordnet. Über Schieberegler können

¹⁴⁸ Weitere Ansätze aus diesem Kontext sind die dreidimensionalen sog. Data Jacks sowie das Shape Coding (vgl. [Schumann, Müller 2000, 194ff.]).

¹⁴⁹ Abbildungen (a) bis (d) erstellt in Anlehnung an [Schumann, Müller 2000, 194, Abb. 6.11], Abbildung (e): Detail aus [Alexa, Müller 1998, Fig. 1]).

Anfrageparameter interaktiv verändert werden, um die Auswirkungen anhand der sich verändernden Pixelmuster zu bewerten (vgl. Abb. 4.11).

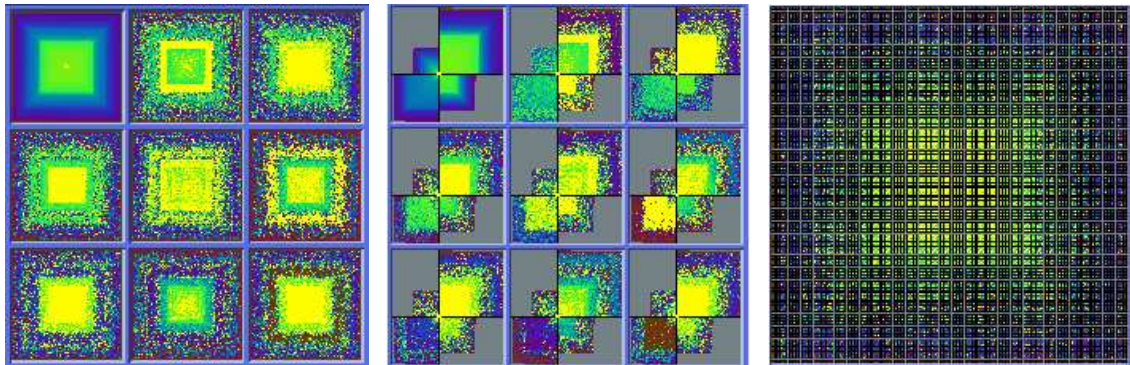


Abb. 4.11 - Drei mit VisDB erzeugte Darstellungen aus 7000 Datensätzen mit acht Dimensionen¹⁵⁰.

4.6.5 Hierarchische Techniken

Einen weiteren Ansatz zur Visualisierung multivariater Daten bilden die sog. *hierarchischen* Techniken, die Datenwerte jeweils auf unterschiedliche Weise hierarchisch anordnen (vgl. [Schumann, Müller 2000, 205ff.]). Sie können danach unterschieden werden, ob sie den Präsentationsraum oder den Merkmalsraum hierarchisieren. Zur ersten Gruppe zählen die Autoren bspw. die ineinandergeschachtelten heterogenen Koordinatensysteme von Worlds within Worlds [Feiner, Beshers 1990], in die zweite Gruppe werden ConeTrees [Robertson et al. 1993] zur Visualisierung von Baumstrukturen (vgl. Kap. 4.8.1 sowie Abb. 4.13a) eingeordnet.

Panel-Matrizen, Streckenzüge, Ikonen, pixelbasierte sowie hierarchische Techniken werden von [Schumann, Müller 2000, 178ff.] als Basiskonzepte für die Visualisierung multivariater Daten eingestuft. Erweiterungen dieser Konzepte können bspw. im Einsatz mehrerer, interaktiv miteinander verknüpfter Darstellungen (Linking & Brushing)¹⁵¹, aber auch in einer Abbildung von Datenwerten auf akustische Parameter (Sonifikation) bestehen (vgl. [Schumann, Müller 2000, 215ff.]).

4.7 Visualisierung von Zeit- und Raumbezug

Um zu aussagekräftigen Darstellungen von Multiparameterdaten zu gelangen, ist eine gleichzeitige Visualisierung von abhängigen und unabhängigen Variablen sinnvoll. Dabei sind für eine geeignete Visualisierung von Zeit- sowie Raumbezug multivariater Daten jeweils spezielle Techniken erforderlich, die hier kurz charakterisiert werden sollen.

4.7.1 Visualisierung des Zeitbezuges

Eine *Zeitabhängigkeit* [Schumann, Müller 2000, 234ff.] in zu visualisierenden Daten ist gegeben, wenn eine Dimension des Beobachtungsraumes (also eine unabhängige Variable) mit der physikalischen Größe *Zeit* assoziiert ist. Die Visualisierung des Zeitbezuges ist dabei nicht an die physikalische Größe *Zeit* gebunden. Die Autoren unterscheiden daher zwischen dem Zeitbezug in den Daten und dem Zeitbezug in der Darstellung. Der *Zeitbezug in den Daten* kann dabei folgendermaßen unterteilt werden:

- Statisch
- ▶ Ein *statischer Zeitbezug* ist gegeben, wenn die Zeitachse nur einen festen Zeitpunkt oder ein Intervall, in dem die vorliegenden

¹⁵⁰ Abbildungen übernommen aus [Keim, Kriegel 1994, Fig. 7].

¹⁵¹ Brushing bezeichnet dabei Möglichkeiten, bspw. in einer Scatterplot-Matrix in einem Diagramm Werte anzuwählen und diese dadurch auch in den anderen Diagrammen hervorzuheben (vgl. dazu auch [Spence 2001, 42f.]).

Daten gültig sind, enthält.

- Quasistatisch ▶ Ein *quasistatischer Zeitbezug* liegt vor, wenn die Zeitachse mehrere diskrete Zeitpunkte oder -intervalle enthält, etwa vier Messungen pro Jahr.
- Dynamisch ▶ Bei einem *dynamischen Zeitbezug* ist die Zeitachse hingegen kontinuierlich skaliert.

Der *Zeitbezug in der Darstellung* kann in statische oder dynamische Repräsentationen unterschieden werden, die jeweils nach Ziel der Visualisierung auszuwählen sind. Bei statischen Repräsentationen verändert sich die graphische Darstellung nicht automatisch über die Zeit; bei dynamischen Repräsentationen verändert sie sich hingegen über die Zeit ohne Zutun des Betrachters (vgl. Kap. 4.3.3). Dynamische Darstellungen können dabei durch

- ▶ wechselnde Attributbelegungen (Flashlighting),
 - ▶ Veränderungen der Ansicht bspw. durch Kamerafahrten,
 - ▶ Objektbewegungen und Deformationen,
 - ▶ sequentielles Hintereinanderschalten *gleichartiger* graphischer Darstellungen sowie
 - ▶ sequentielles Hintereinanderschalten *verschiedenartiger* graphischer Darstellungen
- erreicht werden. Die Autoren verweisen allerdings darauf, dass gängige Visualisierungssysteme die Behandlung des Parameters Zeit nicht hinreichend unterstützen und diese im allgemeinen vollständig dem Anwender überlassen.

4.7.2 Visualisierung des Raumbezuges

Raumbezogene Daten liegen vor, wenn „*der Beobachtungsraum Ortskoordinaten enthält, die ein 2- oder 3-dimensionales räumliches Bezugssystem definieren*“ [Schumann, Müller 2000, 220]. Für die Erstellung raumbezogener Visualisierungen können Geoinformationssysteme oder Visualisierungswerkzeuge verwendet werden, wobei beide Ansätze jeweils spezifische Vor- und Nachteile bieten. So stellen Geoinformationssysteme (GIS) effiziente Werkzeuge für die Visualisierung von Daten im geographischen Kontext dar, erlauben eine Integration unterschiedlicher Arten von Daten und stellen in der Regel Datenbankanbindungen sowie Analysefunktionen bereit. Zur Visualisierung werden Karten verwendet, wobei verschiedene Schichten (sog. Layer) flexibel übereinander gelegt werden können¹⁵². Ihr Nachteil besteht in einer zumeist gegebenen Beschränkung auf einen zweidimensionalen Raumbezug sowie in fehlenden modernen Methoden zur Visualisierung von Multiparameterdaten. Visualisierungswerkzeuge stellen diese hingegen zur Verfügung und sind zudem gut für die Darstellung dreidimensionaler Raumbezüge geeignet; auf der anderen Seite unterstützen sie Karten nur unzureichend, zudem fehlen hier meist Datenbankanbindungen und geeignete Analysefunktionen [Schumann, Müller 2000, 221f.].

Bei einer raumbezogenen Visualisierung ist zwischen einer direkten und einer indirekten Darstellung des Raumbezugs zu unterscheiden. Bei einer *direkten Darstellung* des Raumbezugs werden Merkmale und räumliches Bezugssystem in einer visuellen Repräsentation vereint; der Vorteil dieses Ansatzes liegt darin, dass auf diese Weise die Abhängigkeit der dargestellten Datenwerte von ihrem räumlichen Kontext intuitiv wahrnehmen lässt. Der räumliche Bezug wird über entsprechende Positionierungen, bspw. durch die Platzierung von Ikonen auf einer Karte (vgl. Abb. 4.12), hergestellt. Allerdings können bei einer direkten Darstellung des Raumbezugs aufgrund sonst entstehender Überlappungen nicht allzu viele Werte dargestellt werden; für die Darstellung multivariater Daten können zudem nur Ikonen oder hinreichend kleine Varianten sternförmiger Koordinaten (Sternikonen) eingesetzt werden. Weitere Varianten einer direkten Darstellung des Raumbezugs entstehen durch Ein-

¹⁵² Typische Beispiele sind Möglichkeiten zum Ein- und Ausblenden von Strassen- oder Flussnetzen.

färben von Flächen einer Karte oder die Verwendung von Kartogrammen [Schumann, Müller 2000, 223ff.].

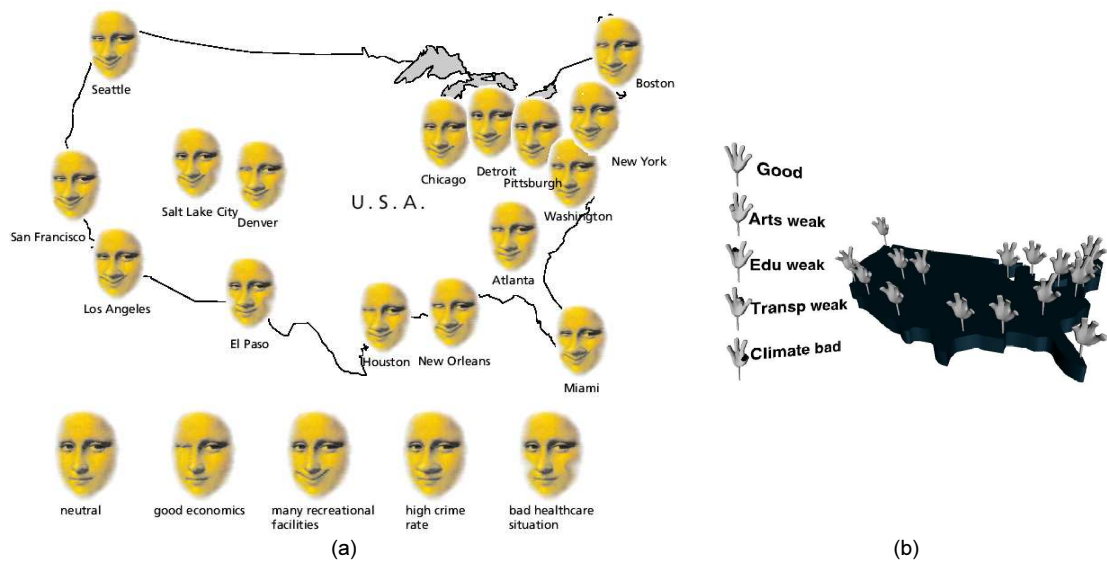


Abb. 4.12 - Einsatz von Ikonen bei der direkten Visualisierung des Raumbezuges multivariater Daten¹⁵³: (a) verzerrte Mona-Lisa-Gesichter; (b) Hand-Ikonen.

Wenn die Zahl der darzustellenden Datenwerte zu groß wird, kann auf eine *indirekte Darstellung* des Raumbezugs zurückgegriffen werden. Hierbei werden Merkmale und räumliches Bezugssystem in unterschiedlichen visuellen Repräsentationen dargestellt; eine Zuordnung erfolgt dann bspw. über Farbassoziationen, Verbindungslinien oder Annotationen. Vorteile dieses Ansatzes liegen in der größeren darstellbaren Datenmenge sowie in der Möglichkeit, alle Basiskonzepte für die Visualisierung multivariater Daten einsetzen zu können; der Nachteil liegt im vergleichsweise höheren kognitiven Aufwand zur Interpretation solcher Darstellungen [Schumann, Müller 2000, 229ff.].

Die Visualisierung von Daten in ihrem Raumbezug ist nach [Schumann, Müller 2000, 233f.] oft „*unbedingt erforderlich*“, um bspw. die Einordnung der Daten zu erleichtern oder räumliche Korrelationen zu veranschaulichen. Aufgrund der Komplexität der darzustellenden Elemente kommt hier ferner geeigneten Interaktionstechniken besondere Bedeutung zu.

4.8 Beispiele aus der Informationsvisualisierung

Nachfolgend werden beispielhaft einige Ansätze aus dem Bereich der Informationsvisualisierung vorgestellt. Für einen ausführlichen Überblick sei auf [Card et al. 1999], die auch eine Sammlung von Reprints zu wichtigen Arbeiten aus diesem Gebiet bereitstellen, sowie auf [Spence 2001] verwiesen; für eine Einführung eignen sich bspw. [Shneiderman 1998, 522ff.] oder [Däßler 1999]. Beinahe schon paradoxerweise wird eine übersichtliche Darstellung der verschiedenen Ansätze der Informationsvisualisierung dadurch erschwert, dass eine mögliche Systematisierung bspw. sowohl anhand der Dimensionalität der zugrunde liegenden Daten, den zur Visualisierung eingesetzten Techniken, der Art der Interaktionen, der Anwendungsbereiche etc. erfolgen kann, wobei häufig Mehrfachzuordnungen einzelner Arbeiten möglich ist (vgl. etwa die Zuordnung einzelner Artikel zu Themengebieten in [Card et al. 1999]). Die hier ausgewählten Beispiele streben daher lediglich einen Eindruck von der Vielfalt der von der Informationsvisualisierung adressierten Aspekte ab; weitere, hier nicht ausgeführte Ansätze befassen sich bspw. mit der visuellen Darstel-

¹⁵³ Abbildungen übernommen aus [Alexa, Müller 1998, Fig.1 u. Fig.2].

lung von sog. Information Workspaces¹⁵⁴, von allgemeinen Netzwerken¹⁵⁵ oder von Internet bzw. World Wide Web¹⁵⁶. Ferner sei angemerkt, dass Arbeiten wie die ConeTrees (vgl. Kap. 4.8.1) wie auch die bereits vorgestellten Ansätze der parallelen Koordinaten von Inselberg (vgl. Kap. 4.6.2) und der pixelbasierten Visualisierung von Keim und Kriegel (vgl. Kap. 4.6.4) sowohl im Kontext der (wissenschaftlichen) Visualisierung (vgl. [Schumann, Müller 2000]) wie der Informationsvisualisierung (vgl. [Card et al. 1999] [Spence 2001]) behandelt werden und damit auch Beispiele für die unscharfe Abgrenzung beider Gebiete gegeneinander darstellen.

4.8.1 Baumstrukturen

Baumstrukturen treten in vielfältigen Anwendungsgebieten, bspw. zur Darstellung von hierarchischen Dateisystemen, auf; entsprechende graphisch-interaktive Schnittstellen wie der Windows Explorer zählen heute zum Standard bei Desktop-Oberflächen. Graphische Darstellungen von Bäumen sind dabei insbesondere mit zwei Herausforderungen konfrontiert - sie wachsen deutlich schneller in die Breite als in die Tiefe, und sie enthalten viel leeren Platz, der notwendig ist, um die Knoten des Baumes visuell zu organisieren [Card et al. 1999, 149ff.]. Um hier zu effizienteren Darstellungen zu gelangen, wurden unterschiedliche Ansätze vorgeschlagen (vgl. auch den Hyperbolic Browser in Kap. 4.8.3). ConeTrees (Kegelbäume) [Robertson et al. 1993] (vgl. Abb. 4.13a) stellen Teilbäume in Form dreidimensionaler Kegel dar und erhöhen so beträchtlich die Zahl der Knoten, die zugleich dargestellt werden können. Sie sind nach [Card et al. 1999, 150] besonders effektiv für die Visualisierung der Gesamtstruktur eines Baumes, wobei in der Darstellung allerdings jeweils Teile der Knoten verdeckt sind.

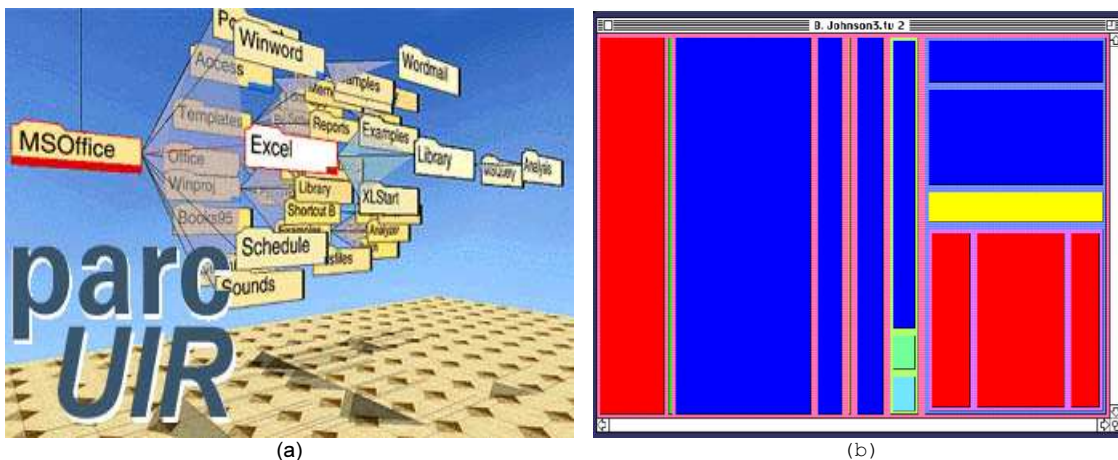


Abb. 4.13 - Visualisierung von Baumstrukturen: (a) ConeTree¹⁵⁷; (b) Tree-Map¹⁵⁸.

Ein anderer Ansatz zur Visualisierung von Baumstrukturen besteht in der Generierung sog. Tree-Maps, die den jeweils verfügbaren Darstellungsraum vollständig ausnutzen. Die Tree-Map [Johnson, Shneiderman 1991] [Shneiderman 1992], ursprünglich motiviert durch Fehlen geeigneter Tools zur Visualisierung der Strukturen großer Festplatten-Laufwerke, teilt die Darstellungsfläche entsprechend der Knoten der obersten Baumebene in Rechtecke

¹⁵⁴ Vgl. für einen Überblick [Card et al. 1999, 513ff.] sowie die Arbeiten von [Risch et al. 1997] [Card et al. 1996] [Bederson, Hollan 1994] [Robertson et al. 1993].

¹⁵⁵ Vgl. für einen Überblick [Card et al. 1999, 187ff.] [Spence 2001, 137ff.] sowie die Arbeiten von [Becker et al. 1995] [Eick, Wills 1993] [Fairchild et al. 1988].

¹⁵⁶ Vgl. für einen Überblick [Card et al. 1999, 465ff.] sowie die Arbeiten von [Bray 1996] [Andrews 1995] [Hendley et al. 1995].

¹⁵⁷ Abbildung übernommen von der Homepage des Forschungsgruppe User Interface Research (UIR) des Palo Alto Research Center (Xerox-PARC); <http://www2.parc.com/istl/projects/uir/>.

¹⁵⁸ Abbildung übernommen aus [Shneiderman 1992].

mit einer Größe relativ zu deren Inhalt auf. Dann wird jede so entstandene Fläche für die nächste Knotenebene rekursiv aufgeteilt, wobei bei jedem Schritt eine Drehung der Schnitt-richtung um 90 Grad erfolgt (vgl. Abb. 4.13b sowie Abb. 4.2c). Auf diesem Ansatz basierende Arbeiten sind bspw. SeeSys [Baker, Eick 1995] für die Darstellung von Softwarestatistiken und TennisViewer [Jin, Banks 1997], der zur Visualisierung des Verlaufs von Tennisspielen deren hierarchische Struktur (Punkt, Spiel, Satz, Match) anhand von Tree-Maps darstellt.

4.8.2 Gleichzeitige Darstellung von Orientierungs- und Detailinformationen

Wenn große Mengen von Informationen zugleich und gleichrangig dargestellt werden, können die resultierenden Bilder schnell unübersichtlich werden. Zwei Ansätze, um hier zu besser zugänglichen Visualisierungen zu gelangen, sind die Techniken Überblick und Detail sowie Fokus und Kontext (vgl. Kap. 4.8.3). Mit der Technik von *Überblick und Detail* (*Overview and Detail*) wird angestrebt, einen Überblick über die gesamten Daten mit Möglichkeiten einer interaktiven Auswahl einer Untermenge der Daten zu verbinden, für die dann Details angezeigt werden. Dabei können zwei Varianten unterschieden werden: Beim Time Multiplexing wird zu jedem Zeitpunkt entweder ein Überblick *oder* Details angezeigt; beim Space Multiplexing werden hingegen zur gleichen Zeit Überblick *und* Details dargestellt [Card et al. 1999, 285]. Beispielhaft sollen zwei Anwendungen für unterschiedliche Einsatzgebiete kurz vorgestellt werden. LifeLines [Plaisant et al. 1996] wurde für die Visualisierung biographischer Daten entworfen; die Autoren beschreiben Anwendungen für Informationen über jugendliche Straftäter des Maryland Department of Juvenile Justice (vgl. Abb. 4.14a) sowie für Krankenakten.

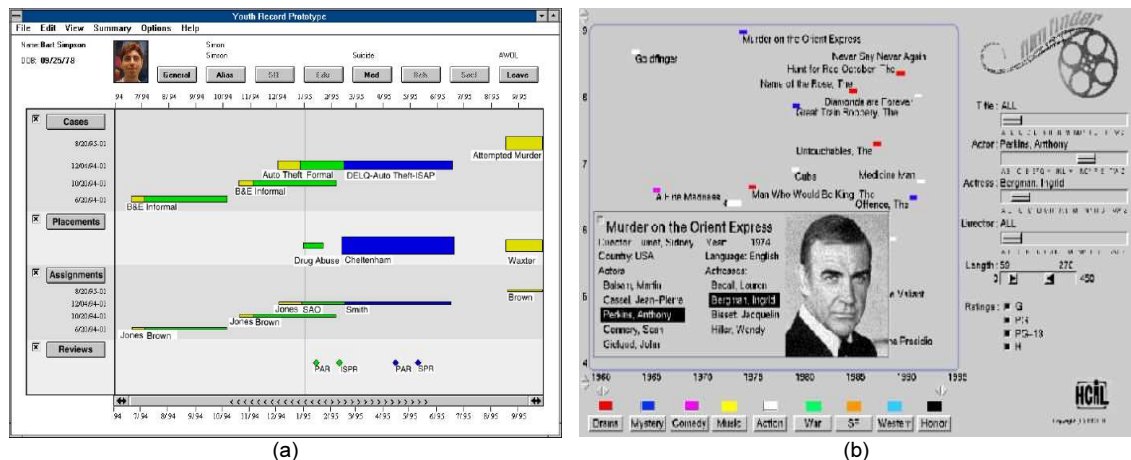


Abb. 4.14 - Überblick und Detail: (a) LifeLines¹⁵⁹, (b) FilmFinder¹⁶⁰.

Der FilmFinder [Ahlberg, Shneiderman 1994] wurde hingegen für die interaktive Erschließung von Informationen über Filme konzipiert (vgl. Abb. 4.14b). Er verbindet die Technik von Überblick und Detail mit dem von Ben Shneiderman entwickelten Konzept der *Dynamic Queries* [Shneiderman 1994]. Hierbei bestimmt der Anwender ausschließlich durch Manipulation von graphischen Bedienelementen wie Schieberegler die aus einer Datenbank auszuwählenden Daten, die direkt in eine visuelle Darstellung umgesetzt werden. Jede Veränderung eines Eingabeparameters löst erneut eine entsprechende Datenbankabfrage aus und verändert so direkt die visuelle Darstellung des Ergebnisses; Voraussetzung hierfür sind extrem schnelle Zugriffsmöglichkeiten auf die Datenbank.

¹⁵⁹ Abbildung übernommen aus [Plaisant et al. 1996, Fig. 2].

¹⁶⁰ Abbildung übernommen aus [Ahlberg, Shneiderman 1994, Color plate 3].

4.8.3 Fokus und Kontext

Sog. *Fokus und Kontext*-Techniken (*Focus & Context*) verfolgen einen anderen Ansatz, um es dem Betrachter zu ermöglichen, alle Daten im Auge zu behalten und zugleich detaillierte Analysen auf ihnen durchzuführen. Dabei wird in Anlehnung an die Funktionsweise des menschlichen visuellen Systems das interaktive Fokussieren auf einen Teil der Darstellung unter Beibehaltung des Kontextes in einer dynamischen Darstellung kombiniert [Card et al. 1999, 307ff.]. Hierzu werden Verzerrungen¹⁶¹ von Teilen des Bildes eingesetzt; frühe Arbeiten, die diesen Ansatz verfolgen, sind Fisheye-View [Furnas 1981] sowie Bifocal Lens [Spence, Apperley 1982]. Umsetzungen von Fokus und Kontext-Techniken finden sich in den ConeTrees (vgl. Kap. 4.8.1), dem ebenfalls bei [Robertson et al. 1993] beschriebenen Perspective Wall oder der Document Lens [Robertson, Mackinlay 1993] (vgl. Kap. 4.8.4). Hier sollen beispielhaft kurz zwei weitere Umsetzungen dieses Ansatzes vorgestellt werden.

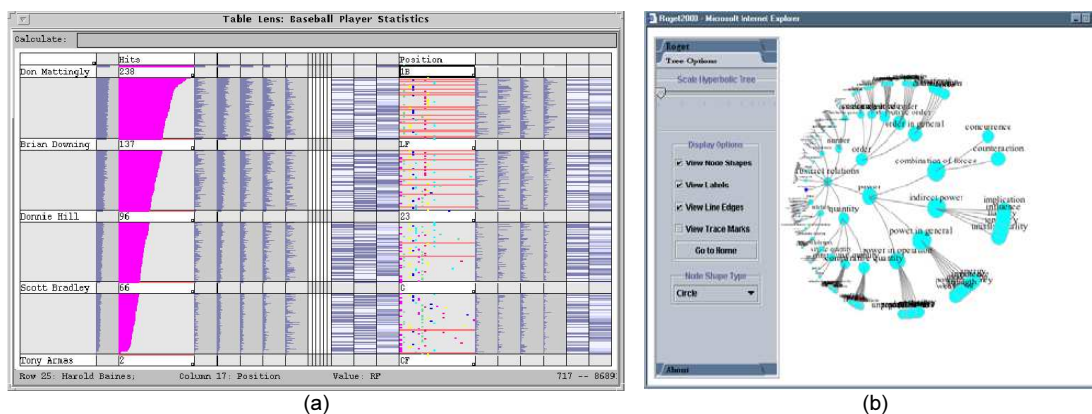


Abb. 4.15 - Beispiele für Fokus und Kontext: (a) Table Lens¹⁶², (b) Hyperbolic Browser¹⁶³.

Die Table Lens [Rao, Card 1994] [Pirolli, Rao 1996] wurde für die für die Analyse sehr großer Datentabellen entwickelt; die Autoren demonstrieren dies am Beispiel einer Anwendung auf Statistiken über Baseball-Spieler, bei der 323 Spieler mit 23 Variablen zugleich dargestellt wurden (vgl. Abb. 4.15a). Die einzelnen Zeilen der Tabelle werden dabei extrem in der Höhe gestaucht und die Werte der Zellen graphisch über Balken repräsentiert; Fokus-Regionen bieten detaillierte Informationen, indem sie auswählbare Zeilen in normaler Größe darstellen. Nach [Card et al. 1999, 341] erlaubt die Table Lens eine effektive Visualisierung von Daten mit vielen Dimensionen und ermöglicht auf dem Bildschirm eine Darstellung von bis zu 100mal umfangreicheren Spreadsheets. Der Hyperbolic Browser von [Lamping, Rao 1996] dient zur Darstellung großer Baumstrukturen und kombiniert dazu nicht-euklidische und euklidischer Geometrie. Durch interaktives Strecken und Stauchen können so bei gleichzeitiger Darstellung von Details im Fokus Bäume mit rund 1000 Knoten - mit rund 50 Knoten im Fokus - visualisiert werden. Abb. 4.15b zeigt eine Anwendung dieses Ansatzes aus dem Jahr 2000 zur Visualisierung eines Thesaurus [Baumgartner, Waugh 2000].

4.8.4 Visualisierung von Dokumenten und Dokumentensammlungen

Ein weiteres von der Informationsvisualisierung adressiertes Anwendungsgebiet ist die Visualisierung von Dokumenten oder Dokumentensammlungen. An dieser Stelle sollen stellvertretend einige Ansätze kurz dargestellt werden. Ein bekanntes Beispiel ist die Do-

¹⁶¹ Einen Überblick über verschiedene visuelle Transferfunktionen geben [Leung, Apperley 1994]; [Carpendale et al. 1997] analysieren die Erweiterung visueller Transferfunktionen auf 3D.

¹⁶² Abbildung übernommen aus [Rao, Card 1994, color plate 2].

¹⁶³ Abbildung übernommen aus [Baumgartner, Waugh 2000, Fig. 4c].

document Lens [Robertson, Mackinlay 1993], die eine Fokus und Kontext-Technik einsetzt. Hier kann eine virtuelle Lupe interaktiv über die Darstellung bewegt werden, um jeweils einen Ausschnitt eines Dokuments detailliert zu betrachten, während umgebende Seiten verzerrt dargestellt werden (vgl. Abb. 4.16a). SeeSoft [Eick et al. 1992] erlaubt es hingegen, bis zu 50.000 Zeilen Software-Code simultan zu analysieren. Einzelne Code-Files werden dabei nebeneinander als senkrechte, dünne Rechtecke dargestellt, jede Code-Zeile als dünne wagerechte Linie (vgl. Abb. 4.16b). Die Darstellung kann durch Auswahl verschiedene statistischer Auswertungen interaktiv verändert werden. Die einzelnen Linien werden jeweils entsprechend eingefärbt; so können bspw. Alter der Programmzeilen, jeweiliger Programmierer oder Anzahl der Aufrufe in einem kürzlich durchgeführten Test betrachtet werden. Nach [Card et al. 1999, 451] besitzt diese direkte Abbildung von Text in graphische Form den Vorteil einer intuitiven Visualisierung, limitiert allerdings die Informationsmenge, die dargestellt werden kann.

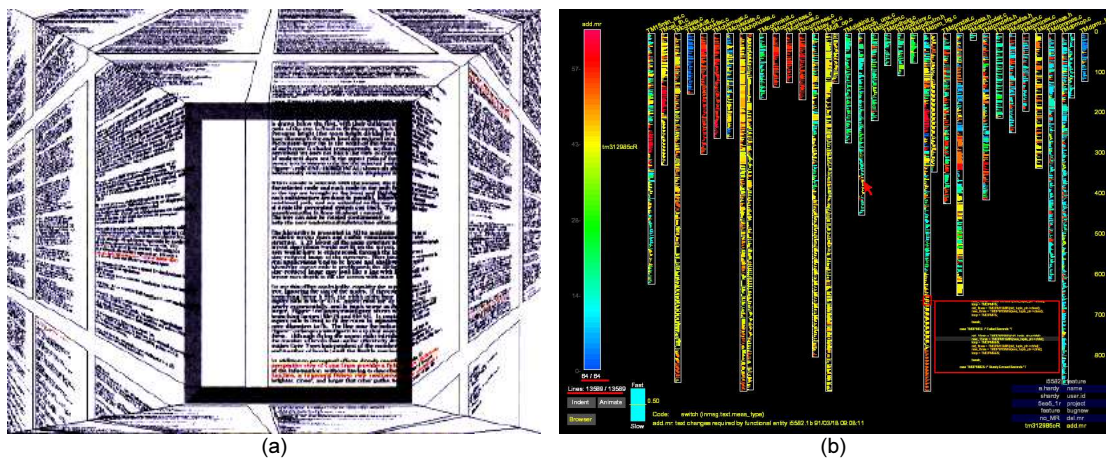


Abb. 4.16 - Dokumentenvisualisierung: (a) Document Lens¹⁶⁴ (b) SeeSoft¹⁶⁵.

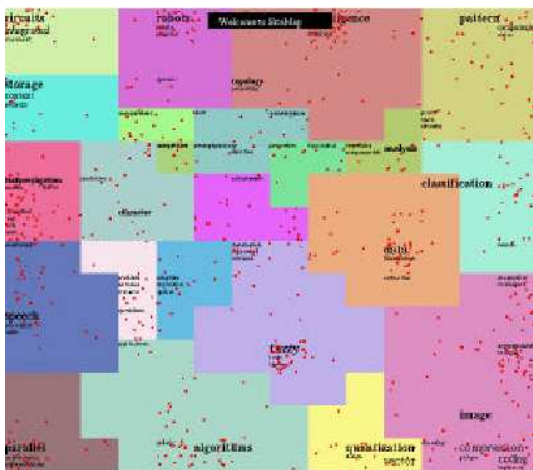
Ansätze zur Visualisierung von Dokumentensammlungen streben eine Darstellung an, die thematisch ähnliche Dokumente in einer Darstellung nahe beieinander gruppiert und so das Auffinden relevanter Dokumente erleichtern soll. Der klassische Ansatz zur automatischen Bewertung von Ähnlichkeiten zwischen Dokumenten basiert auf Vektorraum-Analyse (vgl. Kap. 1.2.3) und stammt aus dem Kontext des Information Retrieval [Salton et al. 1995]; allerdings betont Spence, dass eine Visualisierung von Dokumenten nicht mit Information Retrieval gleichzusetzen sei: Hier werde eine Unterstützung von Nutzern angestrebt, die das Ziel einer Suche nicht genau spezifizieren können, in dem ihnen zunächst ein grober Überblick über das vorhandene Dokumentenmaterial gegeben wird, das nachfolgend sukzessive exploriert werden kann [Spence 2001, 178f.].

Für die Umsetzung in visuelle Darstellungen wurden unterschiedliche Ansätze vorgeschlagen. So verwendet [Lin 1992] ein neuronales Netz (Kohonen-Karte), um eine zweidimensionale *Dokumentenkarte* zu generieren. Eine gegebene Sammlung von Dokumenten wird dabei so auf einer Karte angeordnet, dass miteinander in Beziehung stehende Dokumente nahe bei einander dargestellt werden (vgl. Abb. 4.17a). Eine andere Darstellungsform sind *Galaxies*. Hierbei handelt es sich um zweidimensionale Scatterplots, in denen zuvor als ähnlich eingestufte Dokumente oder Dokumenten-Cluster als sog. Docupoints nahe beieinander dargestellt werden und wie Sterne an einem Nachthimmel erscheinen [Wise et al. 1995]. Ein dritter Ansatz besteht in der Generierung von *Themespaces* (vgl. Abb. 4.17b), die als abstrakte dreidimensionale Landschaften aus von Bergen und Tälern sowohl die

¹⁶⁴ Abbildung übernommen aus [Däßler 1999, Bild 2].

¹⁶⁵ Abbildung übernommen aus [Eick et al. 1992].

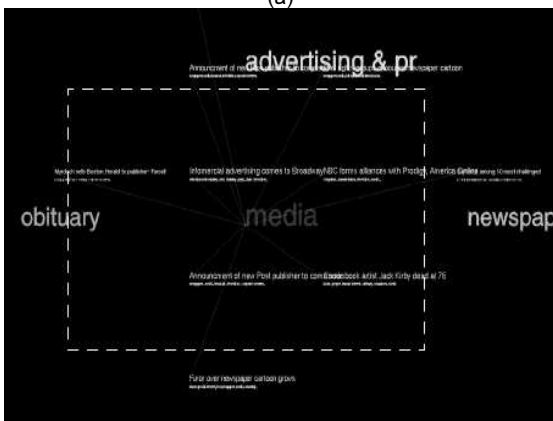
primären Themen einer beliebig großen Dokumentensammlung sowie die relative Häufigkeit der diesen zugeordneten Dokumente darstellen [Wise et al. 1995].



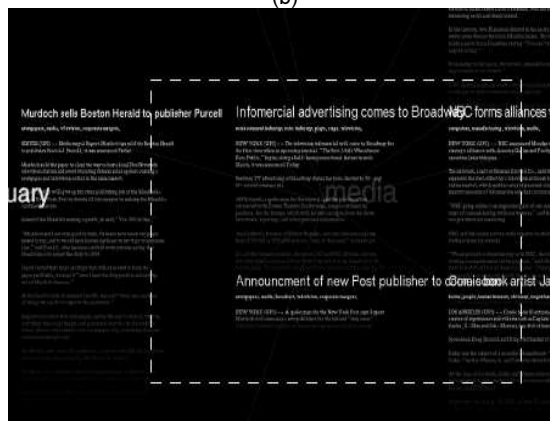
(a)



(b)



(c)



(d)

Abb. 4.17 - Visualisierung von Dokumentensammlungen: (a) Dokumentenkarte¹⁶⁶; (b) Themescape¹⁶⁷; (c) und (d) Galaxy of News¹⁶⁸.

Einen anderer Ansatz liegt Galaxy of News [Rennison 1994] zugrunde. Hier werden aus eingehenden Nachrichtenartikeln zunächst Schlüsselbegriffe herausgefiltert und darauf basierend Beziehungen zwischen den Artikeln hergestellt. Die einzelnen Artikel werden bei der Visualisierung nicht einem einzelnen Ort im Raum zugeordnet: Der Betrachter kann sich sukzessive durch eine Hierarchie aus Themen, Unterthemen bis hin zu Artikeln bewegen. Navigiert der Betrachter zu Unterthemen, bewegen sich die zugehörigen Artikel in den Blickpunkt des Betrachters; zunächst werden Schlagzeilen angezeigt, bei weiterem Hereinzoomen beginnen die Textkörper zu erscheinen, bis schließlich der gesamte Artikel sichtbar ist (vgl. Abb. 4.17c und 4.17d sowie Abb. 4.2d).

Die Visualisierung des *Inhalts* von Dokumenten wirft allerdings erhebliche Probleme auf. [Däßler 1999] betont, dass das im Kontext von Visualisierung häufig zitierte Argument, ein Bild sage mehr als 1000 Worte, gerade auf die Visualisierung von Textdokumenten nicht zutrefte. Er verweist auf die Schwierigkeiten, Text anhand von graphischen Symbolen darzustellen, sowie den erheblichen Lernaufwand, der erforderlich wäre, um eine solche Zeichensprache zu beherrschen.

¹⁶⁶ Abbildung übernommen aus [Däßler 1999, Bild 13].

¹⁶⁷ Abbildung übernommen aus [Däßler 1999, Bild 12].

¹⁶⁸ Abbildungen übernommen aus [Rennison 1994, Fig.1d und 1e].

4.9 Fazit

Der Einsatz interaktiver computergenerierter visueller Repräsentationen kann die effiziente und intuitive Erschließung von Daten in hohem Maße unterstützen und als vielfältig anwendbares Konzept eingestuft werden. Visualisierung kann Menschen bei einem Zugang zu den in Daten enthaltenen Informationen auf eine Weise unterstützen, die etwa von rein textuellen Darstellungen nicht geleistet werden kann. Sie kann im wissenschaftlichen Bereich sowohl zur Auffindung von bisher unbekanntem Zusammenhängen in Daten, zur Überprüfung von Hypothesen sowie zur Kommunikation von Ergebnissen an Dritte verwendet werden. Dabei ist im Kontext der Kommunikation wissenschaftlicher Ergebnisse durch die heute verfügbaren technischen Möglichkeiten bereits eine gewisse Inflation entsprechender bildlicher Darstellungen zu beobachten. In der Januarausgabe 2003 von *Nature* findet sich ein Kommentar von J.M. Ottino, der zu sorgfältigem Umgang mit dem Einsatz von Illustrationen zur Veranschaulichung wissenschaftlicher Ergebnisse mahnt [Ottino 2003] (vgl. auch [Marsiske 2003]). Ottino konstatiert, dass deren Anzahl in Fachzeitschriften wie *Nature* oder *Science* in den letzten zwanzig Jahren so stark zugenommen hat, dass es heute schwer sei, einen entsprechenden Artikel *ohne* Bilder zu finden. Er führt Beispiele attraktiv anmutender wissenschaftlicher Illustrationen an - darunter ein 2002 preisgekröntes computergeneriertes Bild einer „Nanolaus“ (vgl. Abb. 4.18a) sowie einen Nanoschaltkreis auf einem Titelbild von *Science* (vgl. Abb. 4.18b) -, die Ottino zufolge einen jeweils unkorrekten Eindruck beim Betrachter erwecken.



Abb. 4.18 - Kritisierte künstlerische Freiheit in der wissenschaftlichen Illustration: (a) Preisgekröntes Bild einer Nanolaus; (b) Nanoschaltkreis. Nach Ottino kann die Nanolaus, so wie dargestellt, nicht funktionieren; im Nanoschaltkreis wiederum sind Kohlenstoffatome dargestellt, die - in der Realität viel größeren - Goldatome hingegen nicht¹⁶⁹.

Bei der Generierung geeigneter graphischer Repräsentationen aus Daten sind vielfältige Faktoren zu beachten, um Verfälschungen der in diesen enthaltenen Informationen - und damit die Gefahr von Fehlinterpretationen - auszuschließen. Aufgrund der inhärenten Dreidimensionalität der menschlichen visuellen Wahrnehmung bestehen zudem besondere Herausforderungen bei einer geeigneten graphischen Darstellung von Daten mit vielen Beschreibungsdimensionen, die sich ihrerseits aus den Dimensionen von Merkmalen und Beobachtungsraum zusammensetzen. In diesem Zusammenhang sei auf einige weitere Faktoren verwiesen, die Herausforderungen für den Einsatz von Visualisierungswerkzeugen darstellen. So darf nicht übersehen werden, dass Visualisierungssysteme, die mächtige Werkzeuge zur Bildgenerierung aus Daten bereitstellen (vgl. bspw. [Thompson et al. 2001]), vertiefte Kenntnisse voraussetzen, um das in ihnen enthaltene Potential vollständig

¹⁶⁹ Abbildungen übernommen aus [Ottino 2003, Fig. 3 und Fig. 5].

ausnutzen zu können. Entsprechend ist es oft erforderlich, basierend auf diesen Systemen weitere, von Anwendern ohne Spezialkenntnisse intuitiv bedienbare Anwendungen zu entwickeln. Ebenfalls sollte nicht verkannt werden, dass die Interpretation komplexer visueller Metaphern (vgl. etwa die Beispiele zur wissenschaftlichen Visualisierung in Abb. 4.2) ebenfalls Hintergrundkenntnisse und Erfahrung des jeweiligen Anwenders voraussetzt.

Eine interessante Erweiterung zur „klassischen“ Datenvisualisierung stellt das in den 90er Jahren des 20. Jahrhunderts entstandene Gebiet der Informationsvisualisierung dar. Die bisher dort hervorgebrachten Entwicklungen machen deutlich, dass der Fundus an neuen interessanten Metaphern für die Visualisierung erweitert und bereichert und zudem auf weitere Gebiete - etwa die Darstellung der Struktur von Daten - ausgeweitet werden kann. Der bisher erreichte Stand verdeutlicht allerdings auch gegenwärtige Grenzen und Herausforderungen visueller Datenrepräsentationen. So bestehen Probleme neben der Darstellung des Inhalts von Dokumenten bspw. bei einer geeigneten Visualisierung von Netzwerkstrukturen, die bei einer hinreichend großen Anzahl von Verbindungen zwischen einzelnen Knoten schnell unübersichtlich werden [Card et al. 1999, 187ff., 639] [Spence 2001, 137ff.] sowie generell bei der Visualisierung großer Mengen von Information [Däßler 1999]. Däßler stellt in diesem Zusammenhang kritisch fest, dass die Informationsvisualisierung bisher primär Prototypen entwickelt und noch keine breite Anwendung gefunden hat. [Card et al. 1999, 639] äußern sich in ihrem Schlusswort hingegen optimistisch und prognostizieren: „*Information Visualization will enter the mainstream*“.

Visualisierung soll Menschen bei der Interpretation von Daten unterstützen und ist damit an der hierfür jeweils möglichen Leistungsfähigkeit zu messen. Festzuhalten bleibt daher, dass der geeignete Einsatz ihrer Konzepte eine wertvolle Unterstützung beim Zugang zu vielen Arten von Daten darstellen kann, andere Formen der Darstellung jedoch nicht obsolet werden lässt. In vielen Fällen wird daher eine an den Bedürfnissen der jeweiligen Anwender ausgerichtete *Kombination* geeigneter graphischer *und* textueller Darstellungen von Datenwerten vorzuziehen sein, die eine sinnvolle Ergänzung der spezifischen Stärken beider Darstellungsformen erlaubt. Dass es letztlich stets auf den Anwender sowie den Anwendungskontext ankommt, ob und welche Form einer Visualisierung für bestimmte Daten geeignet ist, soll abschließend an einem Beispiel verdeutlicht werden, das [Shneiderman 1998, 522] entnommen wurde. So kann eine interaktive visuelle Darstellung von 1000 Städten der USA auf einer entsprechenden Karte den schnellen und intuitiven Abruf von Touristeninformationen durch Zeigen auf eine Stadt erlauben – *vorausgesetzt*, der Anwender ist mit den geographischen Gegebenheiten der USA vertraut. Anwender, die zwar den Namen einer Stadt, nicht jedoch ihre Position kennen, werden diese hingegen schneller in einer alphabetischen Liste identifizieren können.