

Datenvisualisierung und Data Mining

Daniel A. Keim

Universität Konstanz und AT&T Shannon Research Labs

180 Park Ave, Florham Park, NJ, 07932, USA

Telefon (+1) 973 360 8482, Fax (+1) 973 360 8077

keim@research.att.com

1 Einleitung

Die rasante technologische Entwicklung der letzten zwei Jahrzehnte ermöglicht heute die persistente Speicherung riesiger Datenmengen durch den Computer. Forscher an der Universität Berkeley haben berechnet, dass jedes Jahr ca. 1 Exabyte (= 1 Million Terabyte) Daten generiert werden - ein großer Teil davon in digitaler Form. Das bedeutet aber, dass in den nächsten 3 Jahren mehr Daten generiert werden als in der gesamten menschlichen Entwicklung zuvor. Die Daten werden oft automatisch mit Hilfe von Sensoren und Überwachungssystemen aufgezeichnet. So werden beispielsweise alltägliche Vorgänge des menschlichen Lebens, wie das Bezahlen mit Kreditkarte oder die Benutzung des Telefons, durch Computer aufgezeichnet. Dabei werden gewöhnlich alle verfügbaren Parameter abgespeichert, wodurch hochdimensionale Datensätze entstehen. Die Daten werden gesammelt, da sie wertvolle Informationen enthalten, die einen Wettbewerbs-Vorteil bieten können. Das Finden der wertvollen Informationen in den großen Datenmengen ist aber keine leichte Aufgabe. Heutige Datenbankmanagementsysteme können nur kleine Teilmengen dieser riesigen Datenmengen darstellen. Werden die Daten zum Beispiel in textueller Form ausgegeben, können höchstens ein paar hundert Zeilen auf dem Bildschirm dargestellt werden. Bei Millionen von Datensätzen ist dies aber nur ein Tropfen auf den heißen Stein.

Vorteile der visuellen Datenexploration

Für ein effektives Data Mining ist es wichtig, den Menschen in den Datenexplorationsprozess mit einzubinden, um die Fähigkeiten des Menschen – Flexibilität, Kreativität und das Allgemeinverständnis – mit den enormen Speicherkapazitäten und Rechenleistungen moderner Computersysteme zu kombinieren. Die Grundidee der visuellen Datenexploration ist die geeignete Darstellung der Daten in visueller Form, die es dem Menschen erlauben, einen Einblick in die Struktur der Daten zu bekommen, Schlussfolgerungen aus den Daten zu ziehen sowie direkt mit den Daten zu interagieren. Visuelle Data Mining Verfahren haben in den letzten Jahren einen hohen Stellenwert innerhalb des Forschungsbereichs Data Mining erhalten. Ihr Einsatz ist immer dann sinnvoll, wenn wenig über die Daten bekannt ist und die Explorationsziele nicht genau spezifiziert sind. Dadurch dass der Mensch direkt am Explorationsprozess beteiligt ist, können die Explorationsziele bei Bedarf verändert und angepasst werden.

Die visuelle Datenexploration kann als ein Prozess zur Generierung von Hypothesen aufgefasst werden. Die Visualisierung der Daten ermöglicht dem Menschen ein tieferes Verständnis für die Daten,

wodurch er neue Hypothesen über die Daten aufstellen kann. Die Hypothesen können dann wiederum mit Hilfe visueller Datenexplorationsverfahren untersucht und verifiziert werden. Die Verifikation kann jedoch auch mit Hilfe von Techniken aus dem Bereich der Statistik und der künstlichen Intelligenz durchgeführt werden. Die Hauptvorteile der Einbindung des Menschen in den Prozess der Datenexploration im Vergleich zu vollautomatischen Verfahren aus der Statistik bzw. Künstlichen Intelligenz sind:

- der visuelle Datenexplorationsprozess kann stark inhomogene und verrauschte Daten verarbeiten
- der Benutzer benötigt keine Kenntnisse von komplexen mathematischen oder statistischen Algorithmen und Parametern, und deshalb kann die Datenexploration auch durch Nicht-Spezialisten durchgeführt werden

Zusammenfassend kann man feststellen, dass die visuelle Datenexploration in vielen Fällen eine einfachere Exploration der Daten erlaubt und oft auch bessere Ergebnisse erzielt, insbesondere wenn die herkömmlichen automatischen Algorithmen nur unzureichende Ergebnisse liefern. Die visuelle Datenexploration bietet darüber hinaus ein besseres Verständnis des Datenexplorationsprozesses sowie der erzielten Ergebnisse. Visuelle Datenexplorations-Techniken werden deshalb in vielen Anwendungsbereichen eingesetzt und, in Verbindung mit automatischen Algorithmen, sind sie ein unentbehrliches Verfahren zur Exploration wichtiger Informationen aus großen Datenbanken.

Das Paradigma der visuellen Datenexploration

Nach [Shn96] kann die visuelle Datenexploration in 3 Schritte untergliedert werden:

- overview
- zoom and filter
- details-on-demand

Diese Gliederung wird auch als Information Seeking Mantra [Shn96] bezeichnet. Der Benutzer braucht beim visuellen Datenexplorationsprozess zuerst einen Überblick über die Daten (*overview*). In dieser ersten visuellen Darstellung kann der Benutzer interessante Muster in den Daten erkennen, die dann im folgenden mit Hilfe von Zoom- und Selektionstechniken (*zoom and filter*) genauer untersucht werden. Für eine genaue Analyse der Muster benötigt der Benutzer eine Möglichkeit, auf Details der Daten zuzugreifen (*detail-on-demand*). In allen 3 Schritten der visuellen Datenexploration können Visualisierungstechniken eingesetzt werden: Visualisierungstechniken können einfach einen Überblick über die Daten erzeugen und erlauben es einem Benutzer, interessante Teilmengen innerhalb der Visualisierung schnell zu erkennen. Während des Fokussierens auf interessante Teilmengen ist es wichtig, einen Überblick über die Daten beizubehalten, was zum Beispiel durch eine interaktive Verzerrung der visuellen Überblicksdarstellung bezüglich der Foki erfolgen kann. Für die weitere Exploration interessanter Teilmengen benötigt der Datenanalyst eine Möglichkeit, die Daten genauer zu betrachten, um Details zu verstehen. Es ist in diesem Zusammenhang wichtig, dass Visualisierungstechniken nicht nur grundlegende Verfahren für alle 3 Schritte bereitstellen, sondern auch die Schwierigkeiten bei den Übergängen zwischen diesen Schritten überbrücken helfen.

2 Klassifizierung Visueller Data Mining Techniken

Visuelles Data Mining bzw. Informations-Visualisierung konzentriert sich auf Daten, die keine 2D oder 3D Semantik besitzen und damit keine Standardabbildung auf die zweidimensionale Darstellung des

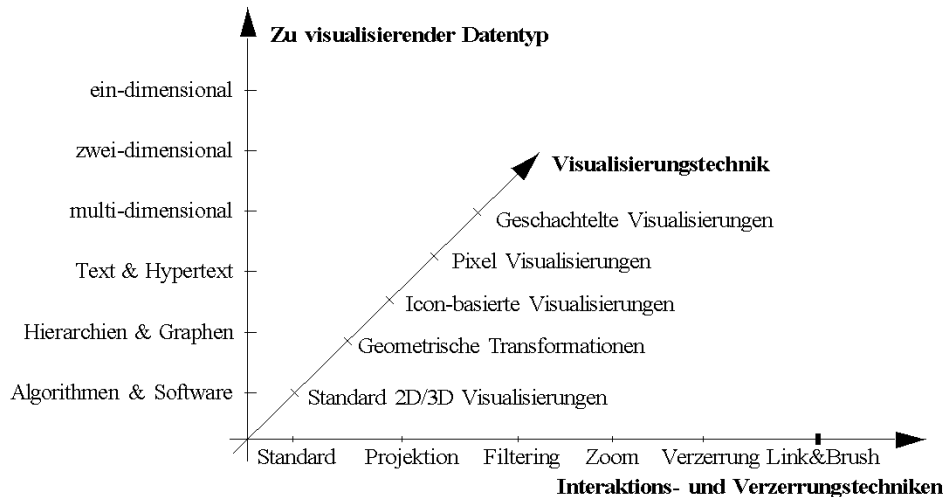


Abbildung 1: Klassifikation visueller Data Mining Techniken (vgl. [Kei01])

Bildschirms besitzen. Für solche Datenmengen gibt es mittlerweile eine Reihe guter Visualisierungstechniken wie zum Beispiel x-y-Plots, Liniendiagramme und Histogramme. Diese Techniken können für die visuelle Datenexploration hilfreich sein, jedoch sind sie im allgemeinen beschränkt auf relativ kleine und niedrigdimensionale Datenmengen. In den vergangenen Jahren wurden eine Vielzahl neuartiger Techniken für hochdimensionale Datenmengen ohne interne 2D oder 3D Semantik entwickelt. Ein Überblick über diese Verfahren wird in den aktuellen Büchern wie zum Beispiel [CMS99] [War00] [Spe00] [SM00] gegeben. Die Techniken können anhand folgender 3 Kriterien klassifiziert werden [Kei01] (siehe Abbildung 1) : (1) der zu visualisierende Datentyp, (2) die verwendete Visualisierungstechnik und (3) die verwendeten Techniken für Interaktion und Verzerrung.

Der **zu visualisierende Datentyp** kann wie folgt untergliedert werden [Shn96]

- **Ein-dimensionale Daten**, wie zum Beispiel zeit-abhängige Daten (vgl. ThemeRiver Visualisierung [HHNW02] zeitabhängiger Nachrichten in Abbildung 2)
- **Zwei-dimensionale Daten**, wie zum Beispiel geographische Karten (vgl. Gridfit Visualisierung [KH98] von Telefondaten in Abbildung 3)
- **Multi-dimensionale Daten**, wie zum Beispiel tabellarische Daten aus relationalen Datenbanken (vgl. Parallele Koordinaten Visualisierung [ID90] in Abbildung 4)
- **Text und Hypertext**, wie zum Beispiel Nachrichten oder Web-Dokumente (vgl. ThemeView Visualisierung [WTP⁺95] [Wis99] von Text Dokumenten in Abbildung 5)
- **Hierarchien und Graphen**, wie zum Beispiel Telefon- oder Internetverbindungen (vgl. Skitter Visualisierung [HBk⁺01] in Abbildung 6)
- **Algorithmen und Software**, wie zum Beispiel Debugging-Operationen (vgl. Tarantula Software Visualisierung [EHJS01] in Abbildung 7).

Die **Visualisierungstechniken** können wie folgt untergliedert werden

- **Standard 2D/3D Visualisierungen**, wie zum Beispiel Balkendiagramme oder X-Y-Diagramme (vgl. Abbildung 3)
- **Geometrische Transformationen**, wie zum Beispiel künstliche Landschaften [Wis99] (vgl. Abbildung 5) und Parallele Koordinaten [ID90] (vgl. Abbildung 4)
- **Icon-basierte Visualisierungen**, wie zum Beispiel die "Strichmännchen"-Visualisierung [PG88] (vgl. Abbildung 8)
- **Pixel-Visualisierungen**, wie die Recursive Pattern oder Circle Segments Techniken [Kei00] (vgl. Abbildung 10 und Abbildung 11)
- **Geschachtelte Visualisierungen**, wie zum Beispiel Treemaps [Shn92] [JS91] (vgl. Abbildung 13) oder Dimensional Stacking [War94] (vgl. Abbildung 12).

Die dritte Kategorie der Klassifikation sind die **Interaktions- und Verzerrungstechniken**. Sie erlauben es dem Benutzer, direkt mit den Visualisierungen zu interagieren. Interaktions- und Verzerrungstechniken können wie folgt untergliedert werden

- **Interaktive Projektion** wie sie zum Beispiel im GrandTour System [Asi85] verwendet wird.
- **Interaktive Selektion** wie sie zum Beispiel im Polaris System [STH02] verwendet wird.
- **Interaktives Zooming** wie es zum Beispiel im Spotfire System [Shn99] verwendet wird.
- **Interaktive Verzerrung** wie sie zum Beispiel im Hyperbolic Tree [LRP95] [MB95] verwendet wird.
- **Interaktives Linking and Brushing** wie es zum Beispiel im XGobi System [SCB92] [BSC96] verwendet wird.

Man beachte, dass die drei Dimensionen unserer Klassifikation als orthogonal betrachtet werden können. Orthogonal bedeutet in diesem Zusammenhang, dass für eine zu visualisierende Datenmenge eine beliebige Visualisierungstechnik in Verbindung mit einer beliebigen Interaktions- und Verzerrungstechnik verwendet werden kann. Man beachte ferner, dass ein System mehrere unterschiedliche Datentypen und eine Kombination mehrerer Visualisierungs- und Interaktionstechniken unterstützen kann.

3 Zu visualisierende Datentypen

Die im Bereich Informations-Visualisierung vorkommenden Daten besitzen in der Regel eine große Anzahl von Datensätzen. Jeder Datensatz entspricht dabei einer Beobachtung, wie zum Beispiel einer Messung bei einem physikalischen Experiment oder einer Transaktion in einem E-Commerce System, und besitzt eine feste Anzahl an Attributen¹. Die Anzahl der Attribute kann dabei stark variieren – von einigen wenigen Attributen bis hin zu Tausenden von Attributen. Da bestimmte Attribute wie zum Beispiel die Zeit oder die geographischen Koordinaten eine besondere Bedeutung haben, werden diese in der Regel auch besonders behandelt. Im Folgenden sollen die verschiedenen Kategorien genauer vorgestellt werden.

¹Im Bereich der Informations-Visualisierung spricht man in der Regel von Dimensionen.

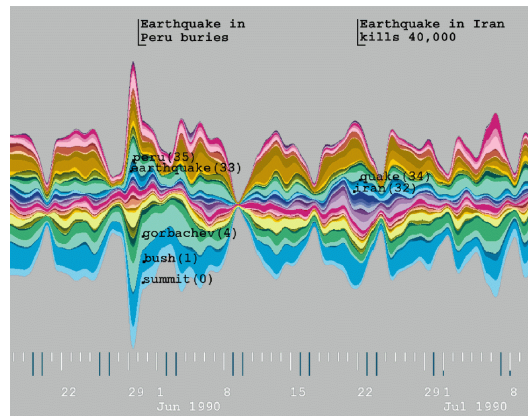


Abbildung 2: Die ThemeRiver Visualisierungstechnik [HHNW02] stellt die thematischen Veränderungen in großen Mengen von Textdokumenten über die Zeit dar. Der Fluss fließt von links nach rechts. Die sich veränderte Breite der einzelnen Segmente des Flusses visualisiert die thematischen Veränderungen. In unserem Beispiel wird ein Archiv von Associate Press Nachrichtenmeldungen von Juni-Juli 1990 visualisiert. Die wichtigsten Ereignisse dieses Zeitraums, wie das Erdbeben in Peru und das Gipfeltreffen zwischen Bush und Gorbatschow, können dabei gut identifiziert werden. ©IEEE

Ein-dimensionale Daten

Eindimensionale Daten besitzen in der Regel ein kontinuierliches Attribut, das eine vollständige Ordnung auf den Daten definiert. Ein typisches Beispiel für eindimensionale Daten sind zeitabhängige Daten. Jedem Zeitpunkt können dabei mehrere Datenwerte zugeordnet sein. Beispiele für eindimensionale Daten sind Aktienkurs-Verläufe (vgl. Recursive Pattern Visualisierung [KKA95] in Abbildung 10) oder Zeitreihen von Zeitungsmeldungen (vgl. ThemeRiver Visualisierung [HHNW02] in Abbildung 2).

Zwei-dimensionale Daten

Zweidimensionale Daten besitzen zwei spezielle Dimensionen, die jeden Punkt eindeutig charakterisieren. Ein klassisches Beispiel für zweidimensionale Daten sind geographische Koordinaten. Für die Darstellung von zweidimensionalen Daten eignen sich zum Beispiel Standard X-Y-Plots, wobei einfache geographische Karten als ein spezieller Typ von X-Y-Plots betrachtet werden können. Auf den ersten Blick scheint die Visualisierung von zeitabhängigen oder geographischen Daten einfach zu sein. In vielen Fällen kommt es aufgrund der großen Datenmengen jedoch zu einem hohen Überlappungsgrad, wodurch die Visualisierungen schwierig zu verstehen ist. Ein Beispiel für eine Visualisierungstechnik, die große Mengen zweidimensionaler Daten überlappungsfrei visualisiert, ist die Gridfit Visualisierungstechnik [KH98] (siehe Abbildung 3).

Multi-dimensionale Daten

Viele Daten besitzen mehr als drei Attribute und können von daher nicht mittels einfacher 2D oder 3D Darstellungen visualisiert werden. Typische Beispiele für mehrdimensionale Daten sind Tabellen in relationalen Datenbanken, die oft mehrere hundert oder sogar tausend Attribute besitzen. In den meisten Fällen existiert keine einfache Abbildung dieser Attribute in die zweidimensionale Ebene, so

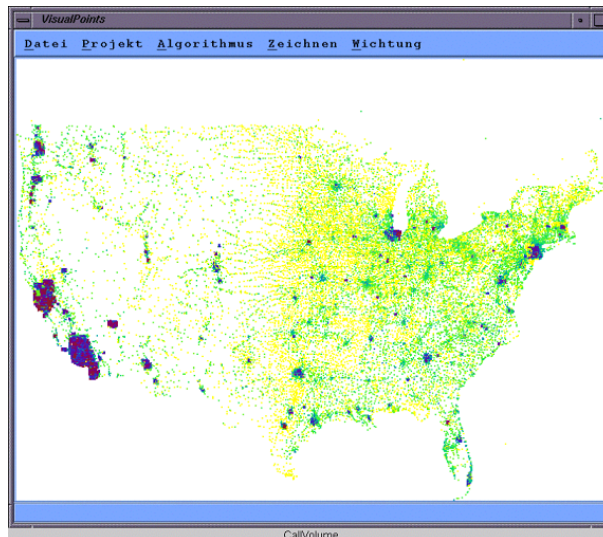


Abbildung 3: Die Gridfit Visualisierungstechnik [KH98] ist eine spezielle Technik für die Visualisierung Geographie-bezogener Daten. Die Gridfit Technik sorgt dafür, dass Überlagerungen von Pixeln in Gebieten mit einer hohen Dichte vermieden werden, die geographische Position der Pixel aber weitgehend erhalten bleibt. Im Bild wird eine Visualisierung von Telefon-Anruf-Volumen-Daten in den USA gezeigt. ©IEEE

dass neuartige Techniken für ihre Visualisierung benötigt werden. Die Parallele Koordinaten Technik [ID90] ist ein Beispiel für eine Visualisierungstechnik, die auf einfache Art und Weise die Visualisierung von multidimensionalen Daten erlaubt (siehe Abbildung 4).

Text und Hypertext

Im Zeitalter des World Wide Web sind Text und Hypertext zwei wichtige Datentypen. Text und Hypertext Datentypen unterscheiden sich von den bisher vorgestellten Datentypen, da sie nur schwierig durch einfache Datentypen mit fest definierter Größe beschrieben werden können. Viele der bekannten Standard-Visualisierungstechniken können deshalb nicht für ihre Visualisierung verwendet werden. Damit aber Text und Hypertext dennoch visualisiert werden können, werden sie in den meisten Fällen in sogenannte Beschreibungsvektoren (Feature-Vektoren) transformiert. Ein Beispiel für eine einfache Transformation ist das Zählen aller nicht-trivialen Wörter im Text. Diese einfache Transformation wird in der Praxis oft mit einer Hauptkomponentenanalyse oder Multidimensionalen Skalieren kombiniert, um die Daten dann mit Hilfe von Standard 2D/3D-Visualisierungen darzustellen (vgl. ThemeView Visualisierung [WTP⁺95] in Abbildung 5).

Hierarchien und Graphen

Eine Vielzahl von Datensätze lässt sich nicht durch die bisherigen Datentypen beschreiben, da die Datensätze komplexe Beziehungen untereinander besitzen. Diese Beziehungen können mit Hilfe von Graphen modelliert werden. Ein Graph besteht aus einer Menge von Objekten, den Knoten, und Verbindungen zwischen diesen Objekten, den Kanten. Eine Hierarchie von Objekten kann dabei als eine spezielle Art von Graph aufgefasst werden, bei der die Verbindungen nur in Top-Down Richtung verlaufen.

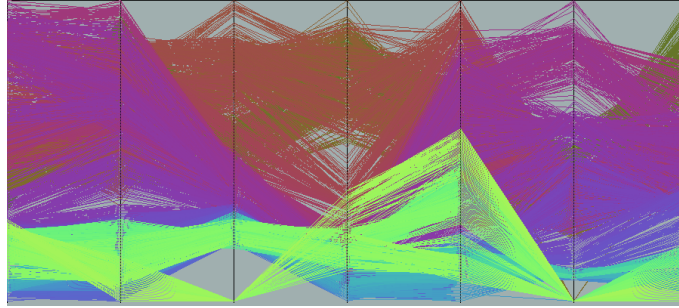


Abbildung 4: Die Parallelen Koordinaten [ID90] bilden den k -dimensionalen Raum auf die zwei Bildschirmdimensionen ab, indem jeder Datensatz als polygonale Linie in einem System paralleler Achsen dargestellt wird. Jede der k äquidistanten parallelen Achsen entspricht einer Dimension und ist linear skaliert vom Minimum bis zum Maximum der auftretenden Datenwerte der Dimension. Jeder Datensatz wird durch eine polygonale Linie repräsentiert, welche die vertikalen Achsen (Dimensionen) an der Stelle schneidet, die dem jeweiligen Datenwert der dazugehörigen Dimension entspricht.

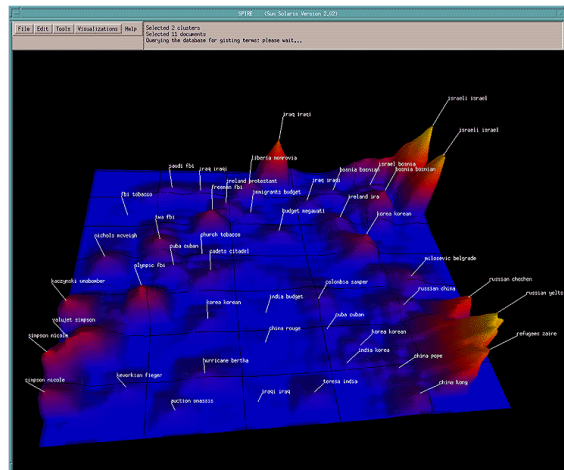


Abbildung 5: Die *ThemeView*TM Visualisierungstechnik [WTP⁺95] stellt eine große Menge von Textdokumenten als eine *Landschaft* dar, wobei Berge Themengebieten entsprechen, die in der Menge von Text-Dokumenten häufig auftreten. Die ThemeView Visualisierung generiert Feature-Vektoren aus den Text-Dokumenten, die dann geometrisch transformiert werden, um zwei Dimensionen zu ermitteln, die für die Verteilung der Datensätze in der Darstellung geeignet sind. ©PNNL

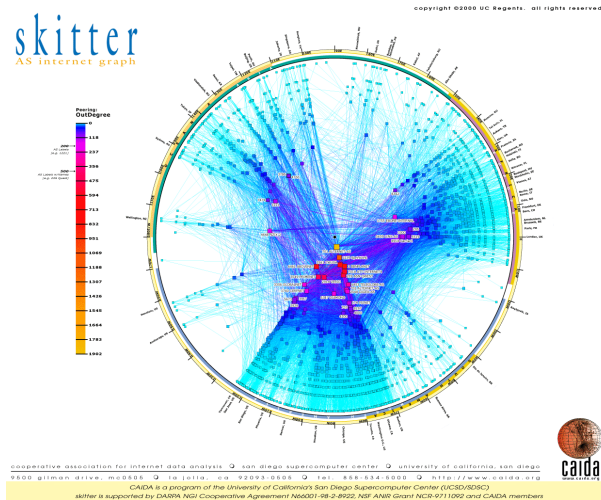


Abbildung 6: Die im Skitter Projekt [HBkc⁺01] verwendete Visualisierungstechnik zeigt ein globalen Internet-Graphen. Die Knoten sind in Polarkoordinaten angeordnet, wobei die wichtigen Knoten mit einer hohen Anzahl von Verbindungen weiter innen liegen als die Knoten mit einer geringen Anzahl von Verbindungen. ©CAIDA

fen. Beispiele für Hierarchien und Graphen sind die Verbindungen in Telefon- oder Computernetzwerken, das Kaufverhalten von Kunden in E-Commerce Angeboten, das Filesystem auf Festplatten und die Hyperlinks im World Wide Web. Für Graphen und Hierarchien existieren aufgrund ihrer Bedeutung eine Reihe von speziellen Visualisierungstechniken (vgl. Skitter Internet Graph [HBkc⁺01] in Abbildung 6). Ein Überblick über graph-basierte Visualisierungstechniken wird in [Che99] gegeben, Informationen über Web Visualisierungstechniken sind in [Dod01] zu finden, und in [BETT99] werden die algorithmischen Aspekte des Zeichnens von Graphen ausführlich behandelt.

Algorithmen und Software

Eine weitere interessante Klasse von Datentypen sind Algorithmen und Software. Die Erstellung großer Softwareprojekte ist nach wie vor eine große Herausforderung. Das Ziel der Visualisierung ist die Vereinfachung der Softwareentwicklung durch ein besseres Verständnis der Algorithmen und des Quellcodes. Das kann zum Beispiel durch eine geeignete Visualisierung der Struktur der Software (z.B. des Funktions-Aufrufgraphen) geschehen oder durch eine Visualisierung aller von Fehlern betroffenen Teile des Programms erreicht werden (vgl. Tarantula Software Visualisierungssystem [EHJS01] in Abbildung 7). Es gibt eine Vielzahl von Systemen und Tools, welche die Visualisierung von Algorithmen und Software unterstützen. Einen Überblick gibt [Tri01].

4 Visualisierungstechniken

Für die Visualisierung von Daten existieren eine Vielzahl von Visualisierungstechniken. Neben den weitverbreiteten Standard 2D/3D Techniken, wie zum Beispiel x-y (bzw. x-y-z) Diagrammen, Balkendiagrammen, Liniendiagrammen usw., stehen heute eine Reihe weiterentwickelter Visualisierungstechniken zur Verfügung. Im Folgenden sollen die verschiedenen Visualisierungstechniken genauer vorgestellt werden.

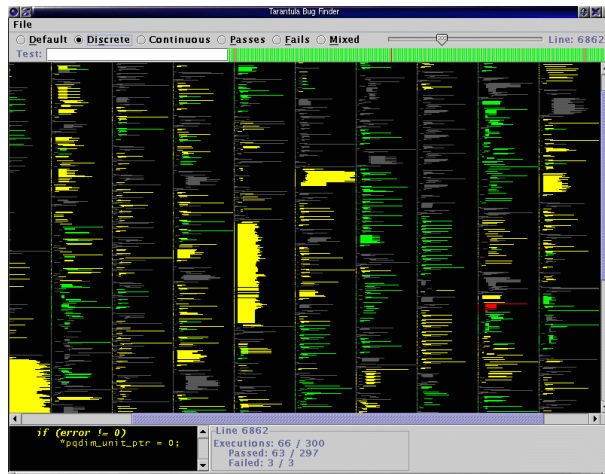


Abbildung 7: Tarantula [EHJS01] ist ein Software-Visualisierungssystem, das die Ergebnisse einer Folge von Tests visuell darstellt. Tarantula zeigt dabei, welche Teile des Quellcodes durch fehlerfreie bzw. fehlerhafte Tests durchlaufen wurden. Dadurch hilft Tarantula Softwareentwicklern, schnell Fehler in umfangreichen Software-Projekten zu finden, was ohne diese Unterstützung erfahrungsgemäß schwierig und zeitintensiv ist. ©GVU Center, Georgia Tech University

Geometrische Transformationen

Geometrische Transformationen versuchen interessante Projektionen der multidimensionalen Datenmenge zu finden, um sie dann visuell darzustellen. Die Klasse der Visualisierungstechniken, die auf geometrischen Transformationen basieren, umfassen Verfahren der explorativen Statistik wie zum Beispiel Scatterplot Matrizen [And72] [Cle93] und Techniken, die unter dem Oberbegriff *Projection Pursuit* [Hub85] zusammengefasst werden. Weitere Visualisierungstechniken, die auf geometrischen Transformationen basieren, sind die Prosection Views [FB94] [STDS95], Hyperslice [vWvL93], und die bereits erwähnten Parallelen Koordinaten [ID90] (siehe Abbildung 4).

Icon-basierte Visualisierungen

Die Idee der Icon-basierten Visualisierungstechniken ist die Abbildung der Attributwerte auf die Eigenschaften eines sogenannten *Icons*. Eine visuelle Darstellung wird dabei durch die Abbildung der Attribute eines Datensatzes auf die Eigenschaften eines Icons erzeugt. Im Fall der Strichmännchen Icons [PG88] (siehe Abbildung 8), zum Beispiel, werden die Dimensionen eines Datensatzes auf die zwei Darstellungsdimensionen sowie die Winkel und Länge der Arme und Beine des Strichmännchens abgebildet. Liegen die Datensätze bezüglich der beiden Darstellungsdimensionen dicht zusammen, entstehen in der resultierenden Visualisierung charakteristische Textur-Muster, welche die Eigenschaften und Charakteristiken der Daten beschreiben und durch den Menschen leicht identifiziert werden können. Die verwendeten Icons können frei definiert werden: Es können Gesichter [Che73], Nadel-Icons [AK02], Stern-Icons [War94], Strichmännchen Icons [PG88], Farb-Icons [Lev91], [KK94] oder TileBars [Hea95] sein.

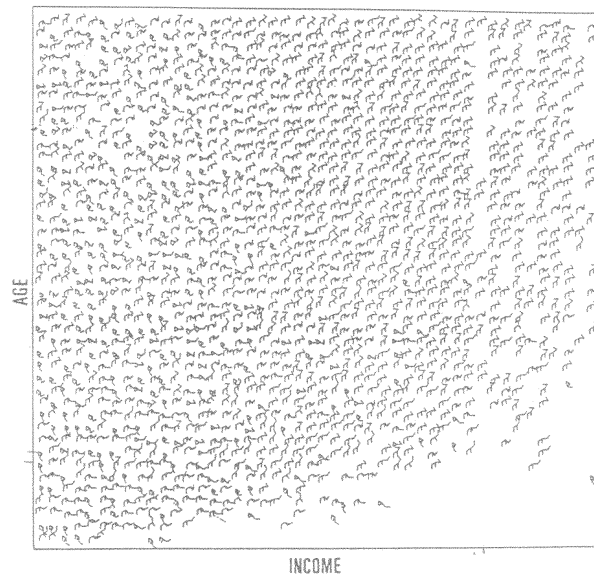


Abbildung 8: Die Strichmännchen Visualisierung [PG88] zeigt Volkszählungsdaten der USA, wobei die Strichmännchen auf der x-Achse nach dem Einkommen und auf der y-Achse nach dem Alter angeordnet sind. Für das Attribut wurden zwei verschiedene Strichmännchen verwendet (das weibliche Strichmännchen besteht aus einem Dreieck mit Schwanz). Die übrigen Attribute wie Abstammung, Schulbildung, etc. sind auf die Längen und Winkel der Arme und Beine abgebildet. Es ist interessant, dass für höhere Einkommen eine relativ homogene Struktur zu sehen ist, wohingegen die Struktur für niedrigere Einkommen sehr diffus ist. ©IVPR, University of Massachusetts at Lowell

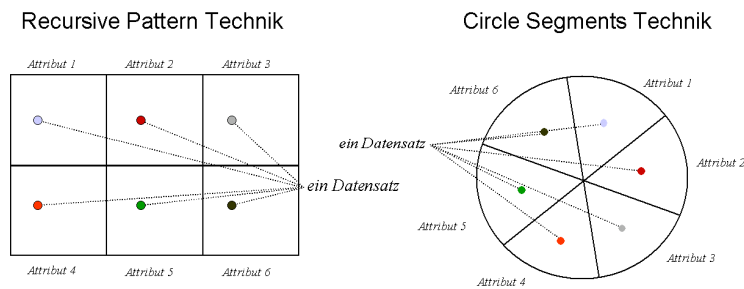


Abbildung 9: In Pixel-Visualisierungen [Kei00] werden die Pixel gemäß der Dimensionen gruppiert, bei der Recursive Pattern Technik [KKA95] in rechteckigen Teilbereichen und bei der Circle Segments Technik [AKK96] in Kreissegmenten. Die Pixel, die zu einem Datensatz gehören, sind dementsprechend über die Teilbereiche verstreut und stehen nur über die relative Position innerhalb der Teilbereiche in Beziehung.

Pixel-Visualisierungen

Die Idee der Pixel-Visualisierungen [Kei00] ist die Abbildung jedes Datenwerts auf ein farbiges Pixel. Dabei werden die Pixel gemäß der Dimensionen gruppiert (siehe Abbildung 9), bei der Recursive Pattern Technik [KKA95] in rechteckigen Teilbereichen und bei der Circle Segments Technik [AKK96] in Kreissegmenten. Die Pixel, die zu einem Datensatz gehören, sind dementsprechend über die Teilbereiche verstreut und stehen nur über die relative Position innerhalb der Teilbereiche in Beziehung. Über die visuelle Beziehung zwischen den Teilbereichen ist es möglich, lokale Beziehungen zwischen den Attributen, Korrelationen und Ausnahmen zu finden. Wird jeder Datenwert auf genau ein farbiges Pixel abgebildet, können mit Pixel-Visualisierungen auf heutigen Bildschirmen bis zu 1.000.000 Datenwerte dargestellt werden. Beispiele für pixelorientierte Visualisierungen sind die Spiral Technik [KK94], die Recursive Pattern Technik [KKA95] (siehe Abbildung 10), die Circle Segments Technik [AKK96] (siehe Abbildung 11) sowie die Pixel Bar Chart Technik [KHDH02].

Geschachtelte Visualisierungen

Die geschachtelten Visualisierungen (Stacked Display Techniques) partitionieren die Daten gemäß eines oder mehrerer Attribute und visualisieren die Daten dann in hierarchischer Form, wobei die Wertebereiche der Attribute ineinander geschachtelt werden. Die Attribute, die für den Aufbau der Hierarchie verwendet werden, müssen sorgfältig ausgewählt werden, da sie einen großen Einfluss auf die Qualität des Ergebnisses haben. Beispiele für geschachtelte Visualisierungstechniken sind Dimensional Stacking [LWW90] (siehe Abbildung 12), World-within-Worlds [FB90] und Treemaps [JS91] [Shn92] (siehe Abbildung 13).

5 Interaktions- und Verzerrungstechniken

Für eine effektive Datenexploration sind Interaktions- und Verzerrungstechniken unverzichtbar. Durch Verwendung von Interaktionstechniken kann der Datenanalyst die Visualisierungen gezielt bezüglich der Explorationsziele verändern. Interaktionstechniken erlauben zudem eine Kombination verschiedener Visualisierungstechniken. Verzerrungstechniken helfen bei der Fokussierung auf Details, ohne dabei

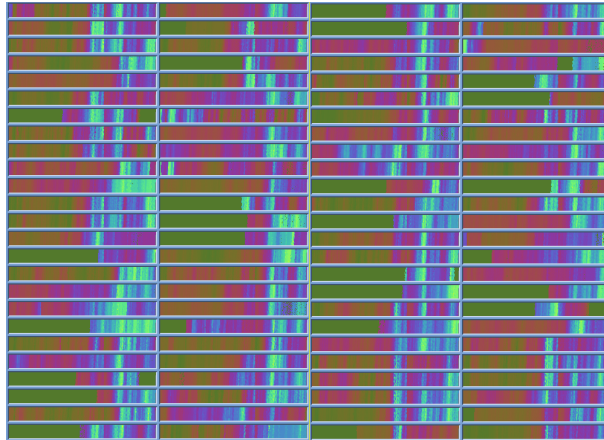


Abbildung 10: Die *Recursive Pattern* Technik [KKA95] basiert auf einer rekursiven Verallgemeinerung einer zeilen- und spalten-orientierten Anordnung der Pixel. Auf einer Rekursionsstufe werden die Pixel von links nach rechts angeordnet, dann eine Zeile tiefer von rechts nach links usw. angeordnet. Die Anordnung der Pixel erfolgt nach dem gleichen Prinzip auf allen Rekursionsstufen mit dem Unterschied, dass die Basiselemente, die auf der nächst höheren Rekursionsstufe angeordnet werden, aus Pixelarrays bestehen, die sich aus den tieferen Rekursionsstufen ergeben. Die Visualisierung zeigt die tagesgenauen Kurse der 100 Aktien des FAZ-Index über einen Zeitraum von 20 Jahren (Jan. 74 - Apr. 95). Die Abbildung der Kurswerte auf die Farbskala wurde so gewählt, dass helle Farben hohen Kursen und dunkle Farben niedrigen Kursen entsprechen. ©IEEE

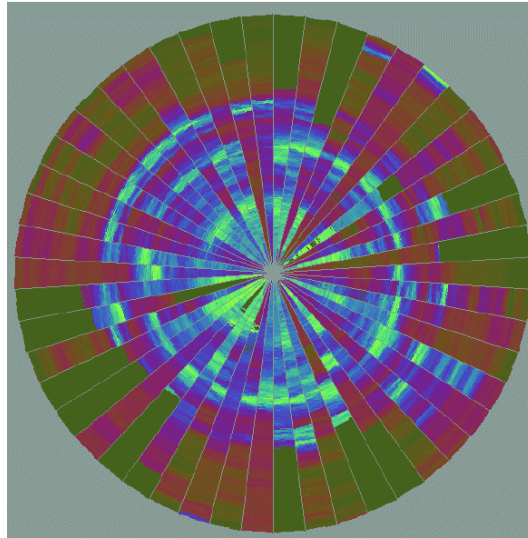


Abbildung 11: Die *CircleSegments* Technik [AKK96] ordnet die Pixel, die zu einem Attribut gehören, in den Segmenten eines Kreises an. Die Anordnung der Pixel innerhalb eines Segmentes beginnt im Zentrum des Kreises und endet am Kreisrand und erfolgt nach einem links-rechts Schema orthogonal zur Segment-Halbierenden. Die Visualisierung zeigt die tagesgenauen Kurse von 50 Aktien des FAZ-Index über einen Zeitraum von 20 Jahren. Deutlich zu erkennen sind ähnliche Aktienkursverläufe, die Hochpreisphasen (helle kreisförmige Ringe), sowie Aktien, die sich gegen den Trend verhalten. ©IEEE

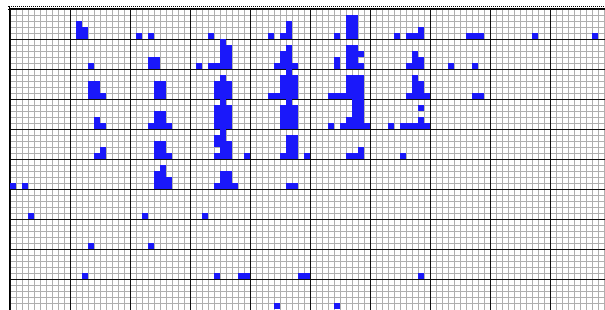


Abbildung 12: Die Idee der Dimensional Stacking Technik [LWW90] ist die Einbettung eines Koordinatensystems in ein anderes Koordinatensystem. Die visuelle Darstellung wird durch die Aufteilung des äußersten Koordinatensystems in rechteckige Zellen erzeugt. Innerhalb dieser rechteckigen Zellen spannen zwei weitere Attribute ein inneres Koordinatensystem auf. Das innere Koordinatensystem wird dann wiederum in rechteckige Zellen zerlegt usw. In unserem Beispiel werden verschiedene Attribute von Ölförderdaten wie Längengrad, Breitengrad, Tiefe und Qualität visualisiert. ©IEEE

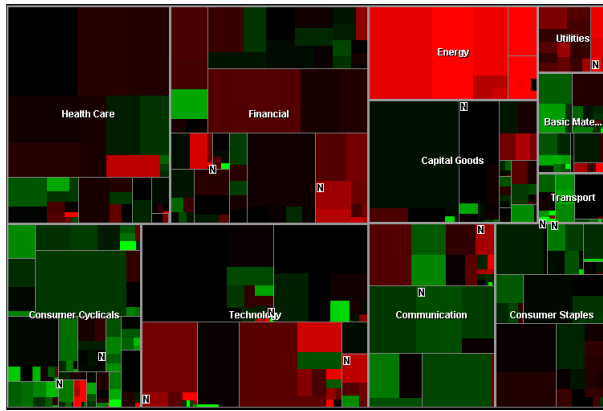


Abbildung 13: Die Treemap Visualisierungstechnik [JS91] [Shn92] unterteilt den Bildschirm abwechselnd in x- und y-Richtung gemäß der auf den Daten definierten Hierarchie. Die Größe und Farbe der entstehenden Gebiete entspricht weiteren Attributen. Die Visualisierung zeigt Finanzdaten, wobei die Anordnung nach Marktsegmenten erfolgt und die Größe der Marktkapitalisierung entspricht [Sma01]. Die Farben zeigen an, wie sich die Aktienwerte in den Segmenten verändert haben – rot entspricht sinkenden Kursen und grün entspricht steigenden Kursen. ©SmartMoney.com

den Überblick über die Daten zu verlieren. Die Idee von Verzerrungstechniken (distortion techniques) ist das Hervorheben von Ausschnitten der Visualisierung mit vielen Details unter Beibehaltung der Darstellung aller übrigen Teile der Visualisierung mit geringerem Detaillierungsgrad. In diesem Zusammenhang unterscheidet man zwischen dynamischen und interaktiven Techniken. Während bei dynamischen Techniken Veränderungen der visuellen Darstellung automatisch vorgenommen werden, erfolgen bei interaktiven Verfahren Veränderungen an der visuellen Darstellung durch Benutzerinteraktion. Im Folgenden sollen die verschiedenen Interaktions- und Verzerrungstechniken genauer vorgestellt werden.

Dynamische Projektion

Die Idee der dynamischen Projektionstechniken ist die dynamische Änderung der Projektionen einer multidimensionalen Datenmenge. Ein klassisches Beispiel ist das GrandTour System [Asi85]. Im GrandTour System wird versucht alle "interessanten" zweidimensionalen Projektionen einer multidimensionalen Datenmenge in einer Serie von Scatterplots darzustellen. Man beachte, dass die Zahl der möglichen Projektionen exponentiell in der Anzahl der Dimensionen wächst. Für eine große Anzahl von Dimensionen ist es deshalb im Allgemeinen nicht möglich, alle Projektionen darzustellen. Die Serie der Scatterplots kann zufällig, manuell oder in Abhängigkeit von den Daten erzeugt werden. Beispiele für Systeme, die dynamische Projektionstechniken unterstützen, sind XGobi [SCB92] [BSC96], XLispStat [Tie91] und ExplorN [CWL96].

Interaktive Filterung

Im Prozess der visuellen Exploration großer Datenmengen ist eine interaktive Partitionierung der Daten in Segmente und ein Fokus auf interessante Teilmengen wichtig. Die Teilmengen können entweder durch eine direkte Selektion der gewünschten Teilmenge (*browsing*) oder durch die Spezifikation von Eigenschaften der gewünschten Teilmenge (*querying*) spezifiziert werden. Eine direkte Auswahl der

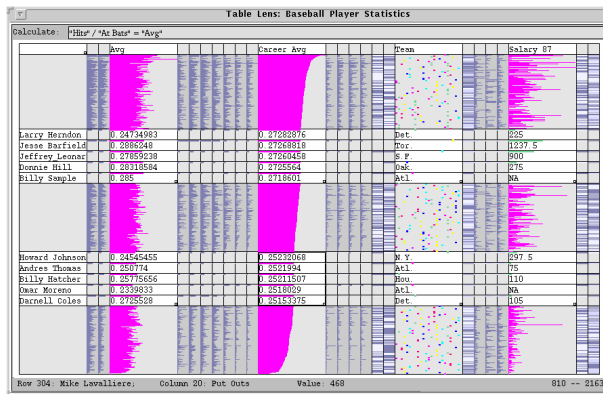


Abbildung 14: Die *Table Lens* Technik [RC94] stellt die Zeilen und Spalten einer Tabelle als komprimierte Balken dar. Das Zooming erlaubt es dann, Zeilen und Spalten mit höherem Detaillierungsgrad zu betrachten. ©ACM

gewünschten Teilmenge ist bei großen Datenmengen schwierig, und ein Problem bei der Spezifikation der gewünschten Teilmenge ist, dass sie oft nicht das gewünschte Ergebnis liefert. Es wurden deshalb eine Vielzahl von Interaktionstechniken für eine bessere Unterstützung der interaktiven Filterung entwickelt. Beispiele sind Magic Lenses [BSP⁺93] [FS95], InfoCrystal [Spo93], Dynamic Queries [AS94] [Eic94] [GR94] und Polaris [STH02].

Interaktives Zooming

Sobald aber große Datenmengen verarbeitet werden, ist es wichtig, die Daten in komprimierter Form darzustellen, damit ein Überblick über die Daten dargestellt wird. Zooming ist eine weitverbreitete Technik, um die Daten – ausgehend von der Überblicksdarstellung – genauer zu betrachten. Zooming erlaubt eine variable visuelle Darstellung der Daten in unterschiedlichen Auflösungen. Zooming bedeutet aber nicht nur, dass die Datenobjekte einfach größer dargestellt werden, sondern auch, dass automatisch die Repräsentation der Daten geändert wird: Je höher der gewählte Zooming-Faktor, desto mehr Details werden auch dargestellt. So können die Datenobjekte zum Beispiel bei einem niedrigen Zooming-Faktor durch einzelne Pixel, bei einem mittleren Zooming-Faktor als Icons und bei einem hohen Zooming-Faktor durch beschriftete Objekte dargestellt werden. Der TableLens-Ansatz [RC94] ist ein interessantes Beispiel für die Verwendung der Zoom-Idee auf große tabellarische Datensätze (siehe Abbildung 14). Weitere Beispiele für Techniken und Systeme, die interaktives Zooming unterstützen, sind PAD++ [PF93] [Bed94], IVEE/Spotfire [BH94] und DataSpace [ADLP95]. Eine vergleichende Darstellung von Fisheye und Zooming Techniken findet sich in [SDZ⁺93].

Interaktive Verzerrung

Die Interaktiven Verzerrungstechniken (distortion techniques) unterstützen den Datenexplorationsprozess durch die Beibehaltung eines Überblicks über die Daten während ein Teil der Daten genauer betrachtet wird. Die Idee der Interaktiven Verzerrungstechniken ist die visuelle Darstellung von Ausschnitten der Daten mit größerem Detaillierungsgrad im Kontext der Überblicksvisualisierung. Bekannte interaktive Verzerrungstechniken sind hyperbolische und sphärische Verzerrungen. Diese Techniken finden bei der Visualisierung von Hierarchien und großen Graphen Verwendung. Sie sind auch teilweise mit

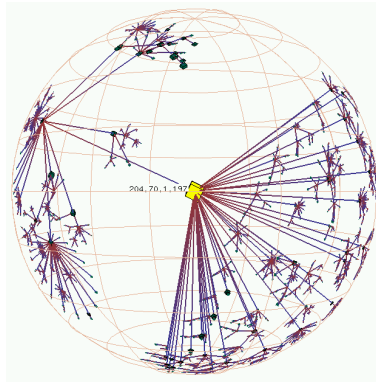


Abbildung 15: Hyperbolische Visualisierung von Internet-Graphen (Bild 1) ©CAIDA

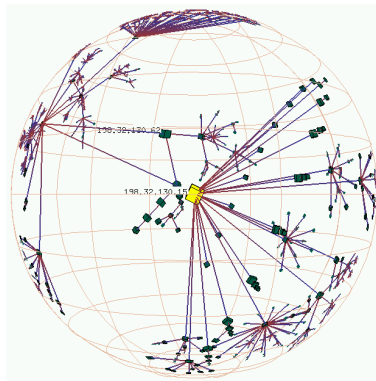


Abbildung 16: Hyperbolische Visualisierung von Internet-Graphen (View 2) ©CAIDA

anderen Visualisierungstechniken integriert. Ein Überblick über die interaktive Verzerrungstechniken findet sich in [LA94] und [CCF97]. Beispiele für Verzerrungstechniken sind Bifocal Displays [SA82], Perspective Wall [MRC91], Graphical Fisheye Views [Fur86] [SB94], Hyperbolische Visualization [LRP95] [MB95] und Hyperbox [AC91].

Interaktives Linking and Brushing

Im Abschnitt über Visualisierungstechniken 4 haben wir gesehen, dass es eine Reihe von Techniken für die Visualisierung hochdimensionaler Daten gibt, die alle ihre Stärken und Schwächen haben. Die Idee von Linking and Brushing (Verknüpfung und Einfärbung) ist die Zusammenführung verschiedener Visualisierungstechniken, um die Nachteile der einzelnen Visualisierungstechniken auszugleichen. So können zum Beispiel zwei verschiedene Scatterplots einer hochdimensionalen Datenmenge durch die Einfärbung und Verknüpfung von Punkten in allen Projektionen kombiniert werden. Nach dem gleichen Prinzip können fast alle in diesem Artikel beschriebenen Visualisierungstechniken durch Linking and Brushing miteinander kombiniert werden. Das Ergebnis ist eine hervorgehobene Darstellung der markierten Daten in allen Visualisierungen, wodurch Abhängigkeiten und Korrelationen in den Daten erkennbar werden. Interaktive Veränderungen in einer Visualisierung werden sofort in den anderen Vi-

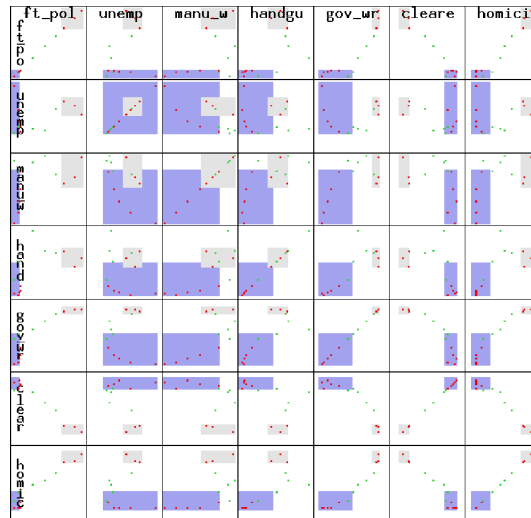


Abbildung 17: Linking & Brushing mit dem XMDV-Tool ©Matthew Ward, Worcester Polytechnic Institute

sualisierungen sichtbar. Es ist interessant, dass man in der durch Linking und Brushing verknüpften Visualisierung mehr erkennen kann, als in der Summe der Einzelvisualisierungen. Typische Beispiele für Visualisierungen, die mit Hilfe von interaktivem Linking and Brushing verknüpft werden, sind Scatterplots, Balkendiagramme, Parallele Koordinaten, und Pixel-Visualisierungen. Die meisten interaktiven visuellen Datenexplorations-Systeme unterstützen interaktives Linking and Brushing. Beispiele für solche Systeme sind Polaris [STH02], Scalable Framework [MKS02], SPlus [BCW88], XGobi [SCB92] [BCS96], Xmdv [War94] und DataDesk [Vel92] [WUT95].

6 Zusammenfassung und Ausblick

Die Exploration großer Datenmengen ist ein sehr wichtiges, aber schwieriges Problem. Informations-Visualisierungs-Techniken können helfen, dieses Problem zu lösen. Die Verwendung von Informations-Visualisierungs-Techniken in zahlreichen Systemen hat gezeigt, dass sie zum Aufspüren interessanter Informationen (wie zum Beispiel Korrelationen, Cluster, funktionale Abhängigkeiten und Ausnahmen) in großen Datenmengen beitragen können. Die Zahl der Anwendungen, die Informations-Visualisierungs-Techniken für eine verbesserte Datenexploration verwenden, steigt stetig an. Beispiele für Bereiche, in denen die visuelle Datenexploration heute schon erfolgreich eingesetzt wird, sind Betrugserkennung, Marketing, und Data Mining zum Beispiel in bio-molekularen Datenbanken. Die Aufgabe zukünftiger Forschung ist die Integration von Informations-Visualisierungs-Techniken mit traditionellen Techniken aus den Bereichen Statistik, maschinelles Lernen und Operations-Research. Erste Ansätze in diese Richtung sind in [PPJ99] [MMHJ99] [HKW99] [HAA00] [MMH00] zu finden und auch kommerzielle Systeme [DBM01] [Gra01] [SAS01] [SPS01] beginnen sich in diese Richtung zu orientieren. Der Vorteil einer solchen Integration ist eine Steigerung der Qualität und Effizienz des Datenexplorationsprozesses. Zusätzlich müssen die visuellen Data Mining Techniken mit großen Datenbank und Data Warehouse Systemen verbunden werden. Das ultimative Ziel ist ein integriertes, leicht bedienbares und verständliches Datenexplorationssystem, das eine schnelle Exploration sehr großer Datenmengen ermöglicht.

Literatur

- [AC91] B. Alpern and L. Carter. Hyperbox. In *Proc. Visualization '91, San Diego, CA*, pages 133–139, 1991.
- [ADLP95] V. Anupam, S. Dar, T. Leibfried, and E. Petajan. Dataspace: 3D visualization of large databases. In *Proc. Int. Symp. on Information Visualization, Atlanta, GA*, pages 82–88, 1995.
- [AK02] J. Abello and J. Korn. Mgv: A system for visualizing massive multi-digraphs. *Transactions on Visualization and Computer Graphics*, 2002.
- [AKK96] M. Ankerst, D. A. Keim, and H.-P. Kriegel. Circle segments: A technique for visually exploring large multidimensional data sets. In *Proc. Visualization 96, Hot Topic Session, San Francisco, CA*, 1996.
- [And72] D. F. Andrews. Plots of high-dimensional data. *Biometrics*, 29:125–136, 1972.
- [AS94] C. Ahlberg and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA*, pages 313–317, 1994.
- [Asi85] D. Asimov. The grand tour: A tool for viewing multidimensional data. *SIAM Journal of Science & Stat. Comp.*, 6:128–143, 1985.
- [BCS96] R. A. Becker, W. S. Cleveland, and M.-J. Shyu. The visual design and control of trellis display. *Journal of Computational and Graphical Statistics*, 5(2):123–155, 1996.
- [BCW88] R. Becker, J. M. Chambers, and A. R. Wilks. The new s language, wadsworth & brooks/cole advanced books and software. *Pacific Grove, CA*, 1988.
- [Bed94] B. Bederson. Pad++: Advances in multiscale interfaces. In *Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA*, page 315, 1994.
- [BETT99] G. D. Battista, P. Eades, R. Tamassia, and I. G. Tollis. *Graph Drawing*. Prentice Hall, 1999.
- [BH94] B. B. Bederson and J. D. Hollan. Pad++: A zooming graphical interface for exploring alternate interface physics. In *Proc. UIST*, pages 17–26, 1994.
- [BSC96] A. Buja, D. F. Swayne, and D. Cook. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics*, 5(1):78–99, 1996.
- [BSP+93] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. DeRose. Toolglass and magic lenses: The see-through interface. In *Proc. SIGGRAPH '93, Anaheim, CA*, pages 73–80, 1993.
- [CCF97] M. S. T. Carpendale, D. J. Cowperthwaite, and F. D. Fracchia. Ieee computer graphics and applications, special issue on information visualization. *IEEE Journal Press*, 17(4):42–51, July 1997.
- [Che73] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal Amer. Statistical Association*, 68:361–368, 1973.

- [Che99] C. Chen. *Information Visualisation and Virtual Environments*. Springer-Verlag, London, 1999.
- [Cle93] W. S. Cleveland. *Visualizing Data*. AT&T Bell Laboratories, Murray Hill, NJ, Hobart Press, Summit NJ, 1993.
- [CMS99] S. Card, J. Mackinlay, and B. Shneiderman. *Readings in Information Visualization*. Morgan Kaufmann, 1999.
- [CWL96] D. B. Carr, E. J. Wegman, and Q. Luo. Explorn: Design considerations past and present. In *Technical Report, No. 129, Center for Computational Statistics, George Mason University*, 1996.
- [DBM01] DBMiner. Dbminer software. <http://www.dbminer.com>, 2001.
- [Dod01] M. Dodge. Web visualization. http://www.geog.ucl.ac.uk/casa/martin/geography_of_cyberspace.html, Oct 2001.
- [EHJS01] J. Eagan, M. J. Harrold, J. A. Jones, and J. Stasko. Visually encoding program test information to find faults in software. In *Technical Report, Georgia Institute of Technology, GIT-GVU-01-09*, 2001.
- [Eic94] S. G. Eick. Data visualization sliders. In *Proc. ACM UIST*, pages 119–120, 1994.
- [FB90] S. Feiner and C. Beshers. Visualizing n-dimensional virtual worlds with n-vision. *Computer Graphics*, 24(2):37–38, 1990.
- [FB94] G. W. Furnas and A. Buja. Prosections views: Dimensional inference through sections and projections. *Journal of Computational and Graphical Statistics*, 3(4):323–353, 1994.
- [FS95] K. Fishkin and M. C. Stone. Enhanced dynamic queries via movable filters. In *Proc. Human Factors in Computing Systems CHI '95 Conf., Denver, CO*, pages 415–420, 1995.
- [Fur86] G. Furnas. Generalized fisheye views. In *Proc. Human Factors in Computing Systems CHI 86 Conf., Boston, MA*, pages 18–23, 1986.
- [GR94] J. Goldstein and S. F. Roth. Using aggregation and dynamic queries for exploring large data sets. In *Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA*, pages 23–29, 1994.
- [Gra01] Silicon Graphics. Mineset software. <http://www.sgi.com/software/mineset>, 2001.
- [HAA00] Hofmann H., Siebes A., and Wilhelm A. Visualizing association rules with interactive mosaic plots. *ACM SIGKDD Int. Conf. On Knowledge Discovery & Data Mining (KDD 2000)*, Boston, MA, 2000.
- [HBkc⁺01] B. Huffaker, A. Broido, k. claffy, M. Fomenkov, S. McCreary, D. Moore, and O. Jakubiec. Visualizing internet topology at a macroscopic scale. In http://www.caida.org/analysis/topology/as_core_network, 2001.
- [Hea95] M. Hearst. Tilebars: Visualization of term distribution information in full text information access. In *Proc. of ACM Human Factors in Computing Systems Conf. (CHI'95)*, pages 59–66, 1995.

- [HHNW02] S. Havre, B. Hertzler, L. Nowell, and P. Whitney. Themeriver: Visualizing thematic changes in large document collections. *Transactions on Visualization and Computer Graphics*, 2002.
- [HKW99] A. Hinneburg, D. A. Keim, and M. Wawryniuk. HD-Eye: Visual mining of high-dimensional data. *Computer Graphics & Applications Journal*, 19(5):22–31, Sep/Oct 1999.
- [Hub85] P. J. Huber. The annals of statistics. *Projection Pursuit*, 13(2):435–474, 1985.
- [ID90] A. Inselberg and B. Dimsdale. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proc. Visualization 90, San Francisco, CA*, pages 361–370, 1990.
- [JS91] B. Johnson and B. Shneiderman. Treemaps: A space-filling approach to the visualization of hierarchical information. In *Proc. Visualization '91 Conf*, pages 284–291, 1991.
- [Kei00] D. Keim. Designing pixel-oriented visualization techniques: Theory and applications. *Transactions on Visualization and Computer Graphics*, 6(1):59–78, Jan–Mar 2000.
- [Kei01] D. Keim. Visual exploration of large databases. *Communications of the ACM*, 44(8):38–44, 2001.
- [KH98] D. Keim and A. Herrmann. The gridfit approach: An efficient and effective approach to visualizing large amounts of spatial data. In *Proc. Visualization 98, Research Triangle Park, NC*, pages 181–189, 1998.
- [KHDH02] D. A. Keim, M. C. Hao, U. Dayal, and M. Hsu. Pixel bar charts: A visualization technique for very large multi-attribute data sets. *Information Visualization Journal*, 1(1):1–14, Jan. 2002.
- [KK94] D. A. Keim and H.-P. Kriegel. Visdb: Database exploration using multidimensional visualization. *Computer Graphics & Applications*, 6:40–49, Sept. 1994.
- [KKA95] D. A. Keim, H.-P. Kriegel, and M. Ankerst. Recursive pattern: A technique for visualizing very large amounts of data. In *Proc. Visualization 95, Atlanta, GA*, pages 279–286, 1995.
- [LA94] Y. Leung and M. Apperley. A review and taxonomy of distortion-oriented presentation techniques. In *Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA*, pages 126–160, 1994.
- [Lev91] H. Levkowitz. Color icons: Merging color and texture perception for integrated visualization of multiple parameters. In *Proc. Visualization 91, San Diego, CA*, pages 22–25, 1991.
- [LRP95] J. Lamping, Rao R., and P. Pirolli. A focus + context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proc. Human Factors in Computing Systems CHI 95 Conf.*, pages 401–408, 1995.
- [LWW90] J. LeBlanc, M. O. Ward, and N. Wittels. Exploring n-dimensional databases. In *Proc. Visualization '90, San Francisco, CA*, pages 230–239, 1990.
- [MB95] T. Munzner and P. Burchard. Visualizing the structure of the world wide web in 3D hyperbolic space. In *Proc. VRML '95 Symp, San Diego, CA*, pages 33–38, 1995.
- [MKS02] N. Lopez M. Kreuzeler and H. Schumann. A scalable framework for information visualization. *Transactions on Visualization and Computer Graphics*, 2002.

- [MMH00] Ankerst M., Ester M., and Kriegel H.P. Towards an effective cooperation of the computer and the user for classification. *ACM SIGKDD Int. Conf. On Knowledge Discovery & Data Mining (KDD 2000)*, pages 179–188, 2000.
- [MMHJ99] Ankerst M., Breunig M., Kriegel H.P., and Sander J. Optics: Ordering points to identify the clustering structure. *SIGMOD'99, Int. Conf on Management of Data, Philadelphia, PA*, pages 49–60, 1999.
- [MRC91] J. D. Mackinlay, G. G. Robertson, and S. K. Card. The perspective wall: Detail and context smoothly integrated. In *Proc. Human Factors in Computing Systems CHI '91 Conf., New Orleans, LA*, pages 173–179, 1991.
- [PF93] K. Perlin and D. Fox. Pad: An alternative approach to the computer interface. In *Proc. SIGGRAPH, Anaheim, CA*, pages 57–64, 1993.
- [PG88] R. M. Pickett and G. G. Grinstein. Iconographic displays for visualizing multidimensional data. In *Proc. IEEE Conf. on Systems, Man and Cybernetics, IEEE Press, Piscataway, NJ*, pages 514–519, 1988.
- [PPJ99] Wong P., Whitney P., and Thomas J. Visualizing association rules for text mining. *Proc. Symposium on Information Visualization*, 1999.
- [RC94] R. Rao and S. K. Card. The table lens: Merging graphical and symbolic representation in an interactive focus+context visualization for tabular information. In *Proc. Human Factors in Computing Systems CHI 94 Conf., Boston, MA*, pages 318–322, 1994.
- [SA82] R. Spence and M. Apperley. Data base navigation: An office environment for the professional. *Behaviour and Information Technology*, 1(1):43–54, 1982.
- [SAS01] SAS. Sas software. <http://www.sas.com>, 2001.
- [SB94] M. Sarkar and M. Brown. Graphical fisheye views. *Communications of the ACM*, 37(12):73–84, 1994.
- [SCB92] D. F. Swayne, D. Cook, and A. Buja. *User's Manual for XGobi: A Dynamic Graphics Program for Data Analysis*. Bellcore Technical Memorandum, 1992.
- [SDZ⁺93] Schaffer, Doug, Zuo, Zhengping, Bartram, Lyn, Dill, John, Dubs, Shelli, Greenberg, Saul, and Roseman. Comparing fisheye and full-zoom techniques for navigation of hierarchically clustered networks. In *Proc. Graphics Interface (GI '93), Toronto, Ontario, 1993, in: Canadian Information Processing Soc., Toronto, Ontario, Graphics Press, Cheshire, CT*, pages 87–96, 1993.
- [Shn92] B. Shneiderman. Tree visualization with treemaps: A 2D space-filling approach. *ACM Transactions on Graphics*, 11(1):92–99, 1992.
- [Shn96] B. Shneiderman. The eye have it: A task by data type taxonomy for information visualizations. In *Visual Languages*, 1996.
- [Shn99] B. Shneiderman. Dynamic queries, starfield displays, and the path to spotfire. In <http://www.cs.umd.edu/hcil/spotfire>, 1999.

- [SM00] H. Schumann and W. Müller. *Visualisierung: Grundlagen und allgemeine Methoden*. Springer, 2000.
- [Sma01] SmartMoney. Marketcap treemap visualization. <http://www.smartmoney.com/marketmap/>, 2001.
- [Spe00] B. Spence. *Information Visualization*. Pearson Education Higher Education publishers, UK, 2000.
- [Spo93] A. Spoerri. Infocrystal: A visual tool for information retrieval. In *Proc. Visualization '93, San Jose, CA*, pages 150–157, 1993.
- [SPS01] SPSS. Spss software. <http://www.spss.com>, 2001.
- [STDS95] R. Spence, L. Tweedie, H. Dawkes, and H. Su. Visualization for functional design. In *Proc. Int. Symp. on Information Visualization (InfoVis '95)*, pages 4–10, 1995.
- [STH02] C. Stolte, D. Tang, and P. Hanrahan. Polaris: A system for query, analysis and visualization of multi-dimensional relational databases. *Transactions on Visualization and Computer Graphics*, 2002.
- [Tie91] L. Tierney. Lispstat: An object-orientated environment for statistical computing and dynamic graphics. In *Wiley, New York, NY*, 1991.
- [Tri01] J. Trilk. Software visualization. <http://wwwbroy.informatik.tu-muenchen.de/~trilk/sv.html>, Oct 2001.
- [Vel92] P. F Velleman. *Data Desk 4.2: Data Description*. Data Desk, Ithaca, NY, 1992, 1992.
- [vWvL93] J. J. van Wijk and R.. D. van Liere. Hyperslice. In *Proc. Visualization '93, San Jose, CA*, pages 119–125, 1993.
- [War94] M. O. Ward. Xmdvtool: Integrating multiple methods for visualizing multivariate data. In *Proc. Visualization 94, Washington, DC*, pages 326–336, 1994.
- [War00] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufman, 2000.
- [Wis99] J. A. Wise. The ecological approach to text visualization. *Journal of the American Society for Information Science*, 50(13):1224–1233, 1999.
- [WTP+95] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, Schur A., and V. Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proc. Symp. on Information Visualization, Atlanta, GA*, pages 51–58, 1995.
- [WUT95] A. Wilhelm, A.R. Unwin, and M. Theus. Software for interactive statistical graphics - a review. In *Proc. Int. Softstat 95 Conf., Heidelberg, Germany*, 1995.