Die Explorative Datenanalyse als Lern- und Erkenntniswerkzeug

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der

Philosophisch-
Sozialwissenschaftlichen
Fakultät der

Universität Augsburg

vorgelegt von
Ulrich Fahrner aus
Augsburg
2009

Erstgutachter: Professor Dr. Gabi Reinmann

Zweitgutachter: Professor Antony Unwin, Ph. D.

Tag der mündlichen Prüfung: 4. Februar 2009

Vorwort

Seit 2002 zieht sich die Explorative Datenanalyse wie ein roter Faden durch meine Arbeit. Die Faszination der Explorativen Datenanalyse entsteht dadurch, dass ein erfahrener Datenanalyst mit wenigen Handgriffen aus einem unscheinbaren Datensatz überraschende Ergebnisse herausarbeitet. Dieses mühelose Jonglieren mit unübersichtlichen Datensätzen sah ich zum ersten Mal 2001 in einer Veranstaltung von Antony Unwin. 2002 begleitete ich zusammen mit meinem Studienkollegen Robert Schmied mit der Videokamera eine Tagung des Lehrstuhls für Rechnerorientierte Statistik und Datenanalyse, bei der weltweit führende Wissenschaftler, die auf diesem Gebiet arbeiten, vortrugen. Die dabei entstandenen Filme zeigen auf faszinierende Weise, wie Experten mit Hilfe von Interaktiver Statistischer Graphik Probleme lösen. Damals wurde mir bewusst, dass sich die Explorative Datenanalyse kaum mit Mitteln eines linearen Textes, ähnlich einer Betriebsanleitung, vermitteln lässt. Aus diesen Überlegungen heraus beschäftigte sich meine Diplomarbeit im Bereich der Mathematik mit den Möglichkeiten einer multimedialen Einführung in die Explorative Datenanalyse. Daraus resultierte die Erkenntnis, dass die Prinzipien und Gesetze, nach denen eine Datenanalyse abläuft, viel komplizierter sind, als sie an der Oberfläche scheinen. Die vorliegende Arbeit soll einen Beitrag dazu leisten, diese Phänomene genauer zu untersuchen. Auf diesem Weg begleiteten mich vor allem Gabi Reinmann und Antony Unwin, die mir bei vielen Fragen konstruktive und hilfreiche Diskussionspartner waren. Die Arbeit in kurzer Zeit abzuschließen, war nur mit der konstanten Unterstützung meiner Kolleginnen und Kollegen im Medienlabor der Universität Augsburg möglich. Meine Familie, die mir den Rücken frei hielt, um diese Aufgabe zu meistern, war für mich von unschätzbarem Wert. Dafür sei allen herzlich gedankt!

Ulrich Fahrner

Augsburg, Januar 2009

Inhaltsverzeichnis

1. Einleitung

Das erste Kapitel dieser Arbeit gliedert sich in folgende Teile:

1. *Motivation – Empirie - eine Säule der Wissenschaft*
2. *Problemanalyse - Statistik und die Klarheit der Zahlen*
3. *Mathematikkompetenz - eine Schlüsselqualifikation*
4. *Mathematik, die Kunst des Lernens*
5. *Die Explorative Datenanalyse als Problemlösewerkzeug*
6. *Die Explorative Datenanalyse als Lernwerkzeug*
7. *Zusammenfassung der Fragestellung*

1.1 Motivation: Empirie - Säule der Wissenschaft

„Der Mangel an mathematischer Bildung gibt sich durch nichts so auffallend zu erkennen, wie durch maßlose Schärfe im Zahlenrechnen."

Carl Friedrich Gauß

„Alles Wissen stammt aus der Erfahrung" (W. Ostwalds,1902, S. 52) dieser Satz von Immanuel Kant steht gewissermaßen als Präambel am Anfang meiner Arbeit, in der ich untersuche, inwieweit sich das sehr erfolgreiche Erkenntniswerkzeug der *Explorativen Datenanalyse* dazu eignet, als Lernwerkzeug für Studierende eingesetzt zu werden. Der Begriff „Werkzeug" ist in diesem Zusammenhang als Instrument zu verstehen, das Nutzern ermöglicht, komplexe Datensätze zu betrachten, zu manipulieren und Aspekte empirischer Untersuchungen sichtbar zu machen, um daraus neue Erkenntnisse zu gewinnen. Seit den Arbeiten von John W. Tukey (J.W. Tukey, 1962, S. 1-67) gehört das *Suchen* (Exploration) neben dem *Beschreiben* (Deskription) und dem *Schließen* (Induktion) zu den Grundaufgaben der Statistik. Die sich daraus entwickelte Strategie wird als *„Explorative Datenanalyse"* bezeichnet. Die Erfolge dieser Methode bei der Bearbeitung realer Datensätze und die steigenden Nutzerzahlen im Bereich der Anwender waren für mich Anlass, Chancen und Grenzen der *Explorativen Datenanalyse* als Lernwerkzeug zu untersuchen. Seit nunmehr dreihundert Jahren hält der Siegeszug der empirischen Forschungsmethoden an. Ausgehend von den Naturwissenschaften über die Ingenieurwissenschaften bis hin zu den Sozialwissenschaften sind empirische Methoden aus der Wissenschaft nicht mehr wegzudenken. Die große Aufgabe, die aus der Anwendung von empirischen Forschungsmethoden erwächst, ist die sinnvolle Auswertung der erhobenen Daten für neuen Erkenntnisgewinn. Viele der entwickelten Methoden liegen an der Nahtstelle zwischen den Disziplinen Mathematik und Physik oder zwischen Mathematik und den Ingenieurwissenschaften. Als Beispiel hierfür sei das Buch von Carl Friedrich Gauß „Theoria combinationis observationum erroribus minimis obnoxiae: erroribus minimis obnoxiae“ (1823) genannt, in dem, ausgehend von astronomischen Problemen, statistische Werkzeuge entwickelt und angewendet werden.

Der Begriff *„Statistik“* stammt aus dem Lateinischen und leitet sich von *statisticum*, den Staat betreffend, ab. Es ging in den Anfängen primär zunächst darum, Daten über den Staat zu sammeln, darzustellen und für die Abwägung politischer Entscheidungen bereit zu stellen. Damit wurden Entscheidungen unabhängiger von subjektiven Einschätzungen. Als einer der Ersten befasste sich Blaise Pascal (1623 - 1662) systematisch mit Wahrscheinlichkeiten (H.Wußing, 2008, S. 280ff) im Glücksspiel und begründete mit diesen Forschungsleistungen Wahrscheinlichkeitsrechnung als eigenständige Disziplin innerhalb der Statisik. Ab Anfang des 18. Jahrhunderts wurde mit der Entwicklung der *Normalverteilung* die entsprechenden Grundlagen von Carl Friedrich Gauß und Pierre-Simon Laplace („Théorie Analytique des Probabiliés“ (1812)) geschaffen. Ausgehend vom Konstrukt des Zufallsexperimentes, z. B. eines Würfels, eines Münzwürfels wurde auf mathematischem Gebiet versucht, den Zufall zu bändigen und eine geschlossene mathematische Darstellung zu finden. Trotz aller Bemühungen gelang es bis Anfang des 20. Jahrhunderts nicht dieses Problem zu lösen. In der berühmten Liste von 23 Problemen, die David Hilbert auf dem Mathematikerkongress im Jahre 1900 in Paris vortrug, war unter dem Problem Nummer sechs die Wahrscheinlichkeitstheorie im Rahmen der Mechanik verzeichnet (H. Wußing, 2008, S. 446ff). Bis dahin war es nicht möglich, die Wahrscheinlichkeitstheorie klar axiomatisch darzustellen, wie es z.B. bei den Teildisziplinen lineare Algebra und anderen, damals schon weit entwickelten Bereichen der Mathematik möglich war. Eine axiomatische Begründung der Wahrscheinlichkeitstheorie wurde erst von Andrei Nikolajewitsch Kolmogorow Anfang 1930 entwickelt und in seinem Buch „Grundbegriffe der Wahrscheinlichkeitsrechnung" 1933 zusammengefasst (H.Wußing, 2008, S. 441ff). Erst mit seinen Arbeiten gelang es, auf mathematischem Gebiet den Zufall, Verteilungen, Wahrscheinlichkeiten axiomatisch zu begründen, darzustellen und gesichert mathematisch abzuleiten. Die Wahrscheinlichkeitstheorie als Grundlage für die schließende Statistik ist eine der wenigen Teilgebiete der Mathematik, die sich erst sehr spät entwickelt haben und uns trotzdem täglich begegnen. Dies liegt meiner Meinung nach daran, dass der Mensch einen Hang zur Gewissheit hat (G.Gigerenzer, 2000, S. 237ff). Wahrscheinlichkeit oder Wahrscheinlichkeitsaussagen bereiten den Menschen Unbehagen. Ein Verständnis dafür entwickelt sich erst relativ spät, anders, als z.B. bei geometrischen oder physikalischen Zusammenhängen.

Zur Verdeutlichung dieser These soll hier ein einfaches Beispiel genannt werden:
Kindern zwischen zwei und drei Jahren ist es möglich, eine schiefe Ebene und den Verlauf einer herunter rollenden Kugel oder eines herunter fahrenden Wagens in Abhängigkeit vom Winkel zu verstehen und experimentelle Vorhersagen zu machen: Je steiler die schiefe Ebene steht, um so schneller rollt die Kugel oder um so schneller fährt der Wagen. Den Kindern ist dieser geometrische Zusammenhang bewusst, selbst wenn sie den dahinter liegenden physikalischen Zusammenhang nicht verbalisieren oder darstellen können. Bei Wahrscheinlichkeitsaussagen kommt ein derart intuitives Verständnis erst sehr spät oder in manchen Fällen überhaupt nicht zum Tragen. Ein typisches Beispiel hierfür ist die große Anzahl von Teilnehmern am Lottospiel. Trotz der schlechten Gewinnquote von 1 : 140 000 000 probieren jede Woche sehr viele Menschen ihr Glück, denn sie glauben intuitiv Gewinnen und Verlieren habe irgendetwas mit „fifty-fifty" zu tun. Das liegt unter anderem daran, dass sie keinen intuitiven Zugang zur Wahrscheinlichkeit haben und nicht abschätzen können, wie unwahrscheinlich ein Gewinn sein wird. Der Begriff der Wahrscheinlichkeit ist für den Großteil der Menschen viel sperriger als geometrische oder einfache arithmetische Begriffe. Das liegt in erster Linie daran, dass der Mensch von seiner individuellen Erfahrung ausgeht und seine Geschichte, sein Leben sowie seine Erfahrungen nur retrospektiv betrachtet. In dieser „rückblickenden" Betrachtungsweise spielt Zufall nur sehr selten bewusst eine Rolle, denn alles was in einer Biografie passiert, hat sich tatsächlich ereignet und die Realisationen der einzelnen Ereignisse, die passiert sind, treffen immer mit der Wahrschein-

lichkeit “1” ein. Die Frage, ob verschiedene Ereignisse in der Biografie auch anders entstehen hätten können oder ob man an der einen oder anderen Stelle vielleicht „Glück” hatte, stellt sich im Nachhinein kaum. Deshalb ist es nicht verwunderlich, dass viele statistische Fragestellungen von Studierenden als schwierig und unübersichtlich empfunden werden. Studenten und angehende Wissenschaftler wenden statistische Verfahren oft nur mechanisch an (G.Gigerenzer, 2000, S. 59ff). Diese Beobachtung ist besonders offensichtlich, wenn im jeweiligen Studienfach die mathematische Grundbildung keinen großen Raum einnimmt. Dieses mechanische Anwenden führt dazu, dass statistische Aussagen oft falsch interpretiert werden und wissenschaftliche empirische Studien nicht richtig gelesen werden können. Neue Methoden und statistische Verfahren werden häufig aus Unsicherheit und wegen der Komplexität des gestellten Problems kaum angewendet und deshalb führen auch explorative mathematische Methoden oft ein Schattendasein.

Die Explorative Datenanalyse als Lernwerkzeug steht im Mittelpunkt der vorliegenden Arbeit. Dabei geht es in erster Linie darum zu untersuchen, ob die Explorative Datenanalyse ein geeignetes Werkzeug für die statistische Grundbildung ist und hilft, empirische Forschungsmethoden sowohl bei den Studierenden als auch bei jungen Wissenschaftlern besser zu verankern.

1.2 Problemanalyse - Statistik und die Klarheit der Zahlen

„Wir benutzen die Statistik wie ein Betrunkener einen Laternenpfahl:
Vor allem zur Stütze unseres Standpunktes und
weniger zum Beleuchten eines Sachverhalts.”

Andrew Lang

Bücher, die zeigen, wie man mit Statistik „lügen” kann oder wie leicht Statistiken manipulierbar sind, haben Konjunktur. Ein Beispiel dafür ist der Titel „Das Einmaleins der Skepsis” von Gerd Gigerenzer (2002a). Bei der Interpretation von Statistiken und stati-

stischen Aussagen steht bei Vielen die Angst im Hintergrund, dass Statistiken nicht verlässlich sind, und die Aussagen vom Statisiker manipuliert werden und dass es in der „Natur" von Statistiken läge, dass diese von Menschen manipuliert werden. Das ehrliche Herangehen an statistische Sachverhalte, das sachgerechte Abbilden eines Problems in einem Datensatz und das gewissenhafte Vorgehen bei der Modellbildung scheint selbst von Fachleuten immer wieder angezweifelt zu werden. Gerd Gigerenzer (2002b) hat zu diesem Thema zahlreiche Studien zusammengetragen und selbst mehrere Untersuchungen dazu durchgeführt. Dabei geht es um das Statistikverständnis bei Ärzten, Juristen und Psychologen. In den Untersuchungen soll überprüft werden, inwieweit Ärzte, Juristen und Psychologen Studien, in denen Statisitiken, Daten und Einschätzungen vorkommen, zu bewerten und daraus richtige Handlungen abzuleiten. Als Illustration soll an dieser Stelle die Untersuchung zur Berufsgruppe der Ärzte näher erläutert werden. (G. Gigerenzer, 2002a, S. 59-72). In mehreren Studien in Deutschland und in den USA kommt es bei folgendem Versuchsaufbau immer zum gleichen Ergebnis: Die Probanden sind kompetente Fachärzte, die in ihrer täglichen Arbeitsroutine mit dem nötigen Fachwissen agieren. Ihnen werden mehrere Studien (vgl. G. Gigerenzer, 2002a, S. 65ff) vorgelegt. Sie werden dann dazu aufgefordert, mit einem konkreten Untersuchungsergebnis, das sich mit der vorgelegten Studie begründet, eine Handlungsempfehlung für Patienten zu geben. Eine der Testaufgaben lautet wie folgt:
In einem Brustkrebs-Screening wird eine Patientin außerhalb einer Risikogruppe als positiv getestet. Nun soll der Arzt dieser Patientin aus der Altersgruppe zwischen 30 und 40 eine Empfehlung geben, da sie positiv bei der Mammographie getestet worden ist. Dem Arzt stehen folgende Angaben zur Verfügung:

1. Bei einem Mammographie-Screening 10% der Patientinnen positiv getestet werden.
2. In der Altersgruppe von 30 bis 40 ist Brustkrebs eher selten, dass heißt, eine von 100 Patientinnen ist wirklich an Brustkrebs erkrankt.

Nun kommt die Patientin zum Arzt und möchte wissen, mit welcher Wahrscheinlichkeit sie wirklich krank ist. Die Rechnung ist mit Hilfe der bedingten Wahrscheinlichkeit ganz einfach zu lösen. Von 100 Patientinnen werden zehn Frauen positiv getestet. Wir wissen aber, dass nur eine von 100 Frauen tatsächlich an Brustkrebs erkrankt ist. So hat also die positiv getestete Frau mit einer Wahrscheinlichkeit von 0,1 also 10 % wirklich Brustkrebs. 80 bis 90 Prozent der Ärzte geben den Frauen eine falsche Einschätzung wieder und behaupten, dass sie mit 80 bis 90 Prozent wirklich an Brustkrebs leiden (G. Gigerenzer, 2002b, S. 65).

Dieses Ergebnis und Versuche bei Juristen und Psychologen mit ähnlichen Ergebnissen zeigen ganz klar, dass es Nichtmathematikern kaum möglich ist, Statistiken richtig zu lesen und dass sie keinen richtigen Umgang mit Zahlen und Risiken haben (G. Gigerenzer, 2002b, S.129ff). Trotzdem entscheiden sie jeden Tag mit Hilfe solcher Studien, was zu tun ist. Diese Fehlerquote ist deshalb so erschreckend hoch, weil man sich bei der Ausbildung der genannten Berufsgruppen alle Mühe gegeben hat, ihnen beizubringen, wie statistische Aussagen in Studien ihres Fachbereichs zu bewerten und einzuschätzen sind. Für viele Wissenschaftler wird nie deutlich, wo der Unterschied zwischen Wahrscheinlichkeitsverteilung und relativer Häufigkeit ist. Es wird nie klar, wo der Zusammenhang liegt zwischen den beschreibenden Instrumentarien, der beschreibenden deskriptiven Statistik auf der einen Seite und auf der anderen Seite der Wahrscheinlichkeitstheorie, die auf Axiomen aufgebaut ist. Einen Grund sehe ich darin, dass genau diese Schnittstelle zwischen der Mathematik einerseits und der realen physischen Welt „draußen" andererseits eines der schwierigsten philosophischen Probleme ist. Genau an dieser Schnittstelle treffen unterschiedliche Forschungslogiken aufeinander und jedem, der Statistik und Wahrscheinlichkeitstheorie richtig anwenden will, jedem der sozusagen die Vorzüge der Mathematik und die Erkenntnis aus der Mathematik für ein empirisches Problem nützen mag, muss klar sein, wo hier die Probleme liegen. Nur wer im Prinzip weiß, wie Mathematik funktioniert, wie Mathematik sich von empirischen Wissenschaften abgrenzt, welche Rolle die Mathematik im Erkenntnisprozess hat und welche eigenen Methoden die Mathematik bereit stellt, um Erkenntnis zu generieren, kann einordnen, wie der Zusammenhang zwischen Wahrscheinlichkeitstheorie und empirischer Forschung ist.

1.3 Mathematikkompetenz - eine Schlüssenqualifikation

„Die Mathematik kann nichts von ihrer Würde einbüßen,
wenn sie als bloßes Objekt der Spekulation,
als unanwendbar zur Auflösung praktischer Aufgaben betrachtet wird.”

Alexander von Humboldt

Mathematik gehört zu den unbeliebtesten Fächern in in unseren Schulen. Bitkom-Präsident Prof. August Wilhelm Scheer (Computerwoche, 2008) weist darauf hin, dass fehlende Mathematikkenntnisse ein wesentlicher Grund für die hohen Abbrecherquoten in technischen Studiengängen sind. Im Studiengang Mathematik beenden rund 60 % der Studienanfänger ihr Studium vorzeitig oder wechseln in eine andere Disziplin. Diese Zahlen sind alamierend und zeigen, dass sowohl die Ausbildung in den naturwissenschaftlichen Studiengängen als auch bei den didaktischen Ansätzen dringender Handlungsbedarf besteht, um dem Negativimage mit seinen drastischen Folgen für die Gesellschaft der Mathematik um zukehren. Dabei stellt sich die Frage, wie sich Mathematikkompetenz definieren läßt. Die OECD definiert „Mathematikkompetenz” (J. Baumert, 2003, S. 118f) als die Fähigkeit,

> *„die Rolle zu erkennen und zu verstehen, die die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und sich auf eine Weise mit der Mathematik zu befassen, die den Anforderungen des gegenwärtigen und künftigen Lebens einer Person als konstruktiven, engagierten und reflektierten Bürger entspricht.”*

Wie ist es um die Mathematikkompetenz in Deutschland bestellt?

Die Pisa- und die TIMSS-Studien sehen bei den Schülern vor allem Defizite im Bereich des konzeptionellen Verständnisses und im Verständnisses naturwissenschaftlicher Arbeitsweisen bei den Schülern (J. Baumert, 2003, S. 55f). Im Bereich der mathematischen Grundbildung ist der Anteil von Schüler, die das Fähigkeitsniveau der Beherrschung einfacher Routinen nicht überschreitet, also einfache Rechenaufgaben lösen kann, mit 70 % sehr groß. 30 % jeder Altersgruppe bleiben auf dem Niveau rein rechnerischen Denkens und kommen nie in den Bereich, der mathematischen Grundbildung, die ausreicht, um komplexere mathematische oder naturwissenschaftliche Zusammenhänge zu durchdringen (vgl. M. Prenzel, 2003). Selbst im Bereich der Medizin wird dieses Defizit ersichtlich. Medizinischen Journals, in denen Forschungsergebnisse von Forschergruppen veröffentlicht werden, konnte in einer Untersuchung aus dem Jahr 2004 von dem Journal of the American Medical Association nachgewiesen werden, dass 15 % aller Studien statistische Resultate falsch anwenden, falsch bewerten oder auswerten (R. Matthews, 2006, S. 26). Es konnte nachgewiesen werden, dass in diesen Studien ein signifikanter Zusammenhang zwischen den Studienergebnissen und der Art, wie die Studien finanziert werden, besteht. Des Weiteren konnte nachgewiesen werden, dass eine Vielzahl der Ergebnisse zwar statistisch signifikant sind, sich aber als bedeutungslose Zufälle im Gesamtbild erweisen. Diese Beispiele zeigen, dass das Problem viel tiefer liegt als lediglich in mangelhaften schulischen Mathematikleistungen. Je außergewöhnlicher eine Behauptung ist, desto aufwändiger muss die Beweisführung sein und das kann natürlich nicht nur mit standardmäßigem Anwenden mechanisch auswendig gelernter Statistikkenntnisse geschehen.

Mögliche Ansatzpunkte zur Steigerung der Statistikkompetenz bei Studierenden und Wissenschaftlern

Die Statistik ist ein mächtiges Werkzeug für die empirische Forschung die solide mathematische Grundkenntnisse und mathematisches Verständnis voraussetzt. Fehlen diese Grundlagen, werden statistische Verfahren mechanisch und oberflächlich angewendet werden, ohne sie in der Tiefe zu durchdringen und ohne wirklich die mathematische Forschungslogik zu verstehen (G. Gigerenzer, 2002b, S. 3-24), auf der das Gebilde der Wahrscheinlichkeitstheorie und vieler statistischer Verfahren fußt. Eine Konsequenz daraus ist, dass statistische Aussagen oft falsch interpretiert werden und die Übersetzung von der Sprache der Mathematik in den eigentlichen Forschungszusammenhang oft nicht gelingt. Auf Grund der Komplexität moderner statistischer Verfahren werden diese kaum angewendet und mit Verfahren ersetzt, die dem eigentlichen statistischen Problem nicht angemessen sind. Da sie aber leicht anzuwenden oder in einer bestimmten Forschungstradition üblich sind, findet man sie in vielen Bereichen der empirischen Wissenschaften (G. Gigerenzer, 2002b, S. 27-43). Die explorativen Datenanalysemethoden führen dagegen oft ein Schattendasein, da die Komplexität des Vorgehens ein großes mathematisches und datenanalytisches Grundverständnis erfordert. Eine Konsequenz daraus ist, dass viele Informationen, die in den Daten liegen, für immer verborgen bleiben. In dieser Arbeit wird zum einen untersucht, warum die Explorative Datenanalyse im Bereich der empirischen Ausbildung von Wissenschaftlern gelehrt werden soll und zum anderen, inwieweit die Explorative Datenanalyse ein Lernwerkzeug innerhalb der Statistikausbildung sein kann. Neben dieser fachspezifischen Fragestellung zur Statistikausbildung zieht sich als roter Faden die Frage durch die Arbeit, wie aus Erfahrung Wissen entsteht. Interessant sind dabei die Prozesse die sich abspielen, wenn aus Individualerfahrungen Modelle, Konzepte und „verlässliche" Erklärungsmodelle der Welt entstehen. Genauer muss der Frage nachgegangen werden, welches mathematische und statistische Grundwissen und welche Basiskompetenzen ein Forscher haben muss, um sowohl die Explorative Datenanalyse als auch schließende statistische Verfahren richtig anzuwenden.

1.4 Mathematik ist die Kunst des Lernens

„Erzähle mir und ich vergesse.
Zeige mir und ich erinnere.
Lass es mich tun und ich verstehe."

Konfuzius, chinesischer Philosoph, 551 - 479 v. Chr.

Die vorliegende Arbeit berührt die Schnittstellen von Mathematik, Statistik, Lernpsychologie, Medienpädagogik und Informatik. Diese interdisziplinäre Herangehensweise macht es notwendig, Begriffe zu vereinheitlichen und manchmal anders anzuwenden, als es in dem jeweiligen Ursprungsfach üblich ist. Ein nicht zu unterschätzendes Problem stellt unter anderen die ideologische Überfrachtung einzelner Konzepte und Begriffe in der Psychologie und Pädagogik dar. Verschiedene, an sich nützliche und in ihrem Ursprungskontext sehr treffende Begriffe wie „System“, „Gestaltprinzipien“, „Lernen“, „Wissen“, „Erfahrung“, „quantitative und qualitative Methoden“, „Messen“, „Beobachten“ usw. sind durch eine nicht begründbare Extrapolation in viele Forschungsbereiche und eine ideologische Überhöhung nur noch schwierig vorurteilsfrei verwendbar. Der Wunsch, mit *einer* „Theorie” die ganze Welt erklären zu wollen, hat meiner Meinung nach in einigen Bereichen der Sozialwissenschaften großen Schaden angerichtet. Das wird daran erkennbar, dass viele Einzeltheorien bestehen, die nicht miteinander verknüpft sind. Deshalb beanspruchen die in dieser Arbeit verwendeten und vielleicht in der allgemeinen Diskussion auch überfrachteten Begriffe nur für den kleinen und hier im Einzelnen belegten Rahmen ihre Gültigkeit. Ich möchte in dieser Arbeit mit den Begriffen und Definitionen möglichst nah an den Ideen und Konzepten der jeweiligen Disziplinen bleiben. Um dies zu verdeutlichen, sei hier Beispiel gegeben: Der Begriff des „dynamisches Systems” ist modern und gegenwärtig und dazu verdammt, von der Biologie bis hin zur Unternehmensführung alles erklären zu müssen, wobei die Maxime gilt: „Alles ist irgendwie systemisch”. Ich dagegen möchte diesen Begriff nur in seiner mathematischen Bedeutung verwenden (siehe Abschnitt 3.7), da nur diese für meine Arbeit Relevanz besitzt.

1.5 Die Explorative Datenanalyse als Problemlösewerkzeug

„Überall geht ein früheres Ahnen
dem späteren Wissen voraus.”

Alexander von Humboldt

Die Explorative Datenanalyse wurde in den 1970er Jahren von John W. Tukey (M.Theus, 1996 S.10 ff) eingeführt und begründet. Im Laufe der letzten Jahrzehnte wurden zahlreiche Werkzeuge entwickelt, z.B. Boxplots, Mosaicplots oder Parallelkoordinatenplots, die allesamt die Aufgabe haben, dem Forscher, dem Mathematiker, dem

Datenanalysten dabei zu helfen, Hypothesen zu generieren, die anschließend mit anderen Mitteln, z.B. der schließenden Statistik, bzw. mit Rückgriff auf andere Theorien, bewiesen und ausgebaut werden. Das Hauptwerkzeug der Explorativen Datenanalyse ist die Interaktive Statistische Graphik (ISG), mit der es möglich ist, auf schnelle und interaktive Weise, Datensätze zu manipulieren und grafisch darzustellen. In den letzten Jahrzehnten von 1970 bis Heute hat sich gezeigt, dass die Explorative Datenanalyse dadurch in der Lage ist, dem Experten, dem Mathematiker, Wege zu interessanten neuen Modellen aufzuzeigen (M.Theus, S. Urbanek, 2009 oder A. Unwin, M.Theus, H. Hofmann 2006). Deshalb liegt die Vermutung nahe, dass es auch für den Lernenden nützlich sein könnte, sich schon in der schulischen und universitären Ausbildung mit Explorativer Datenanalyse zu beschäftigen. Gerade im Umgang mit realen Datensätzen und einem Kontext zeigt die Explorative Datenanalyse ihre Stärken und Vorzüge. Mit ihr wird es möglich, Ansatzpunkte für komplexe Probleme zu finden, die jenseits von Schulbuchbeispielen oder konstruierten Exempeln für statistische Verfahren liegen. Da gerade der Modellbildungsprozess und die erste Auswertung der Daten beim Lernenden und beim Forschenden die meisten Probleme hervorbringen, scheint es ein lohnender Ansatzpunkt zu sein, die interaktive statistische Graphik an realen Datensätzen im Lernbereich und Lehrbereich einzusetzen.

Die Explorative Datenanalyse findet ein immer größer werdendes Anwendungsfeld in allen Bereichen der Wissenschaft. Die Auswertung empirischer Daten ist eine Schlüsselqualifikation, die Mathematiker, Naturwissenschaftler, Ingenieure, Sozial- und Wirtschaftswissenschaftler benötigen, aber häufig erst während ihres Studiums oder den ersten Forschungsprojekten mühevoll erlernen müssen.

Die interaktive statistische Software hat sich zum zentralen Hilfsmittel für Analyseverfahren entwickelt. In diesem Zusammenhang bedeutet Interaktivität nicht nur menügesteuerte Benutzerführung, sondern auch eine unmittelbare Abhängigkeit (dynamische Verknüpfung) zwischen einzelnen Anwendungsfenstern, d.h. eine Veränderung in einem Fenster hat eine direkte Auswirkung auf die anderen Fenster. Das Verfahren der Explorativen Datenanalyse liegt mit seinem wichtigsten Werkzeug, der interaktiven statistischen Graphik, an der Nahtstelle zwischen realen Fragestellungen und der formalen und abstrakten Sprache der Mathematik. Mit Hilfe der Explorativen Datenanalyse wird es bei vielen realen Problemen erst möglich, tief greifende mathematische Erkenntnisse in der Praxis anzuwenden und deren Aussagen für reale Fragestellungen zu nutzen. Dies ist deshalb so wichtig, weil sich sowohl der Lernende als auch der Forschende bei jeder empirischen Untersuchung mit zwei Arten der Modellbildung konfrontiert sieht. Einerseits wird die reale Welt über ein empirisches Untersuchungswerkzeug in einen Datensatz überführt, was bereits zu einer Auswahl und zu einer Modellbildung führt, da nie alle Daten, die einen bestimmten Sachverhalt charakterisieren, erhoben werden können. Andererseits wird bei einem vorhandenen Datensatz eine zweite Modellbildung angewendet, die sich die Hilfsmittel, welche die Statistik bereit stellt, zu Nutze macht. Der zweite Bereich der Modellbildung ist theoretisch fundiert und kann mit Hilfe der Wahrscheinlichkeitstheorie, der schießenden Statistik und mit den Ansätzen der Explorativen Datenanalyse gut und einfach gelöst werden. Der erste Bereich der Modellbildung, in dem erhobene Daten von der realen Welt in einen Datensatz überführt wurden, stellt das wesentlich größere Problem dar. Theoriegestützt wird hier ein Modell der Wirklichkeit in einem Datensatz abgebildet. Dabei wird in den meisten Fällen nicht überprüft, ob der entstandene Datensatz wirklich den Grundbedingungen und Grundannahmen der theoriegestützten Modellbildung entspricht. Eine derartige Überprüfung ist jedoch mit Hilfe der Ansätze der Explorativen Datenanalyse möglich.

Funktionen der interaktiven statistischen Graphik im Erkenntnisprozess

Die Forschungsfrage, die geklärt werden soll lautet:

„Welche Aufgaben übernimmt die interaktive statistische Graphik im Erkenntnisprozess von realen und multivariaten Datensätzen?”

Schon bei der Formulierung der Frage kristallisieren sich zwei Erkenntnisse heraus, denen Theorien zugrunde liegen heraus. Auf der einen Seite stehen die Funktionen der interaktiven statistischen Graphik, die als Teil der Statistik ihre Erkenntnisse axiomatisch begründet. Auf der anderen Seite gibt es die praktischen Funktionen im Bereich des empirischen Erkenntnisprozesses. Hinzu kommen des Weiteren die psychologischen Aspekte, die durch die starke Interaktion zwischen den Graphiken, den dahinter liegenden mathematischen Konzepten und dem Datenanalysten, der die Interaktivität des Systems nutzt, entstehen. In diesem Spannungsfeld wird die interaktive statistische

Graphik als Erkenntniswerkzeug untersucht und ihre Funktion in der Explorativen Datenanalyse herausgearbeitet. Die Interaktivität der interaktiven statistischen Graphik beschränkt sich auf ein sehr technisches Niveau. Durch verschiedene, im Rechnersystem abbildbare, Funktionen lassen sich die Interaktionen in und zwischen den Graphiken realisieren, so dass solche Verknüpfungen immer nur einzelne Aspekte einer realen Fragestellung zeigen können. Die Explorative Datenanalyse eines realen Datensatzes kann aber wesentlich mehr: Der Zugewinn an Erkenntnis hängt ganz entscheidend von der Kreativität und der Erfahrung des Datenanalysten ab. Zudem gibt es nützliche Heuristiken, die für viele Datenanalysen viel versprechend sind und erste Anhaltspunkte für das weitere Vorgehen bieten. Wie und unter welchen Bedingungen dabei die Explorative Datenanalyse den Erkenntnisprozess unterstützt, umreißt der zweite Bereich der vorliegenden Arbeit mit folgenden konkreten Fragen:

1. Wie kann der Erkenntnisprozess während einer Datenanalyse beschrieben werden?
2. Welche Aspekte spielen bei diesem Vorgang eine Rolle?
3. Gibt es Vorgehensheuristiken, die sich im Ablauf von Datenanalysen identifizieren lassen?
4. Gibt es Parallelen zum Lernprozess in Mathematik und Statistik?

1.6 Die Explorative Datenanalyse als Lernwerkzeug

„Es ist nicht das Wissen, sondern das Lernen,
nicht das Besitzen, sondern das Erwerben,
nicht das Dasein, sondern das Hinkommen,
was den größten Genuß gewährt."

Carl Friedrich Gauß

Wie bereits ausgeführt, ist die Explorative Datenanalyse ein erfolgreiches Erkenntniswerkzeug beim Lösen von konkreten und realen Fragestellungen in der statistischen

Anwendung. Da in der Mathematik und damit auch in der Statistik der Lernprozess sehr eng mit dem eigenen individuellen Erkenntnisprozess verbunden ist, stellt sich die Frage, ob die Explorative Datenanalyse auch als Lernwerkzeug in der universitären Ausbildung eingesetzt werden kann. In den Untersuchungen PISA 2003 - Mathematikschwerpunkt und TIMSS war eine wesentliche Erkenntnis der Wissenschaftler, dass es bei vielen Schülern an konkreten Problemlösekompetenzen mangelt (J. Baumert, 2003, S. 55f). Konkret heißt dies, dass zwar die mathematischen Konzepte klar sind, aber nicht eindeutig auf ein neues, unbekanntes Problemfeld transferiert werden können.

Welche Rolle kann die interaktive statistische Graphik im Lernprozess spielen?

Dabei sind wiederum mehrere Themenfelder zu unterscheiden:

1. Die Interaktivität solcher Systeme,
2. Im Lernprozess die Möglichkeiten, kooperativ miteinander an Problemen zu arbeiten und Ideen zu entwickeln
3. Die große Individualität der einzelnen Lernwege innerhalb einer Explorativen Datenanalyse und die damit verbundene Individualisierung des Erkenntnisweges sowie des Lernweges.

Welche Funktion können reale Beispiele im Lernprozess übernehmen?

Bei der Explorativen Datenanalyse ist ein entscheidender Vorteil, dass schwierige und unübersichtliche reale Fragestellungen sinnvoll bearbeitet werden können. Die Problemlösekompetenz scheint bei Schülern und Studierenden am geringsten ausgebildet zu sein und bedarf offensichtlich dringend der Verbesserung, um sich im Lernprozess schneller und effizienter Wissen anzueignen und zu verstehen. Davon abgeleitet ist die Frage, welche Funktion in reale Kontexte eingebettete Beispiele im Lernprozess spielen.

1.7 Zusammenfassung der Fragestellung

Die Explorative Datenanalyse hat sich in den vergangenen 40 Jahren als nützliches Erkenntniswerkzeug im Bereich großer multivariater Datensätze erwiesen. Die Explorative Datenanalyse wurde in den 1970er Jahren von John W. Tukey eingeführt und begründet. Im Laufe der letzten Jahrzehnte wurden zahlreiche Werkzeuge entwickelt, die allesamt unterstützend eingesetzt werden können. Die Explorative Datenanalyse hat als Hauptwerkzeug die interaktive statistische Graphik, mit der es möglich ist, auf schnelle und interaktive Weise Datensätze zu manipulieren und grafisch darzustellen. In den letzten Jahrzehnten hat sich gezeigt, dass die Explorative Datenanalyse dadurch in der Lage ist, dem Nutzer Wege zu interessanten neuen Modellen aufzuzeigen. Bedingt durch die Erfolge im Bereich der Forschung drängt sich mehr und mehr die Frage auf, ob nicht auch in der Lehre und beim Lernen die Explorative Datenanalyse und die damit eng verknüpften Erkenntnisse der interaktive statistische Graphik nutzbringend eingesetzt werden könnte. Dabei gibt es zwei Bereiche, die unterschieden und getrennt voneinander betrachtet werden müssen. Es handelt sich um die Explorative Datenanalyse als Erkenntniswerkzeug in der Forschung und um die Chance, die Explorative Datenanalyse als Lernwerkzeug einzusetzen.

2. Strategie zur Problemlösung in Mathematik und Datenanalyse

Das zweite Kapitel dieser Arbeit gliedert sich in folgende Teile:

1. *Von der Natur der Mathematik*
2. *Heuristiken des mathematischen Problemlösens*
3. *Die Graphik in der Mathematik*
4. *Die Idee der Explorativen Datenanalyse*
5. *Klassen statistischer Probleme*
6. *Werkzeuge der Explorativen Datenanalyse*
7. *Vorgehensmodell der Explorativen Datenanalyse*
8. *Ein Beispiel für das Arbeiten mit der Explorativen Datenanalyse*
9. *Strategie der Problemlösung in der Datenanalyse*

2.1 Von der Natur der Mathematik

„Es ist die Logik, mit der man beweist;
es ist die Intuition mit der man erfindet."
Henri Poincaré

Zu Beginn dieser Untersuchung sollen ein paar grundsätzliche Überlegungen zur Entwicklung der Mathematik, der Statistik und deren philosophische Basis anstellt werden, denn der Begriff Mathematik stammt von dem altgriechischen Adjektiv μαθηματική [τέχνη] und mathēmatikē [téchnē] ab, was soviel bedeutet wie „[die Kunst des] Lernen[s], zum Lernen gehörig" (W. Gemoll, 2009). Schon der Begriff „Mathematik" zeigt die enge Verknüpfung zwischen Mathematik und Lernen, die in der folgenden Arbeit näher beleuchtet wird. Das wichtigste Hilfsmittel, um neue mathematische Erkenntnisse zu erlangen, ist das „Problemlösen", das seit jeher das Hauptlernwerkzeug in der Lehre der Mathematik ist. Die Mathematik nimmt unter den Wissenschaften eine Sonderrolle ein, da sie weder eine Naturwissenschaft noch eine Geisteswissenschaft im eigentlichen Sinn ist. Das grundlegende Werkzeug der Mathematik, nämlich das Schließen und Beweisen, das zu „immer währenden Wahrheiten" führt, macht die Mathematik, von einem epistemologischen Standpunkt aus, zu einem komplizierten Forschungsfeld. Bei vielen Schülern und Studierenden führt dieser Sonderstatus zu Schwierigkeiten. So ist es häufig nicht klar, wie die Mathematik und die in der Schule und im Studium erworbenen Fähigkeiten einzuordnen sind (Rayn, 1984). Die Spannweite der philosophischen Ansichten über Mathematik reichen vom Realismus bzw. Platonismus, also der Ansicht, dass die mathematischen Gegenstände eine von Menschen unabhängige Existenz besitzen, über den Formalismus, der die Mathematik nur als Spiel mit Zeichen und Regeln begreifen, bis hin zum Konstruktivismus, der davon ausgehen, dass Mathematik nur in den Köpfen der Mathematiker stattfindet. Für diese Arbeit soll folgender pragmatischer Standpunkt gelten, der nach Meinung des Verfassers, für die meisten Anwender und Mathematiker einen guten Mittelweg darstellt:

> Die mathematischen Objekte, Sätze und Kalküle sind von den Mathematikern oder, je nach persönlicher Einstellung (ontologischer oder konstruktivistischer Standpunkt), erfundene oder entdeckte Strukturen, die in vielen Fällen nützliche und interessante Aussagen über die uns umgebende Welt außerhalb der Mathematik geben können. Die Entdeckung oder die Erfindung dieser Strukturen findet als aktiver Konstruktionsprozess in den Köpfen derjenigen statt, die sich mit Mathematik beschäftigen. Das Offenbarwerden mathematischer Strukturen bedarf also des aktiven Konstruktionsprozesses in jedem Lernenden und in dessen kognitiver Struktur.

Imre Lakatos (I. Lakatos, 1979, S. 14) schreibt dazu 1979 im Vorwort seines Buches „Beweise und Widerlegungen":

> *„In der Geistesgeschichte geschieht es häufig, das beim Erscheinen einer neuen mächtigen Methode das Studium jener Probleme rasch voranschreitet und in den Brennpunkt des Interesses rückt, die mit der neuen Methode behandelt werden können, während der Rest Gefahr läuft, nicht beachtet oder sogar vergessen zu werden, und sein Studium der Verachtung anheim fällt."*

In diese Lage scheint die Philosophie der Mathematik in unserem Jahrhundert als Ergebnis der dynamischen Entwicklung der Mathematik geraten zu sein. Der Hauptgegenstand der Metamathematik ist die Abstraktion von der Mathematik, in der mathematische Theorien durch formale Systeme ersetzt werden, Beweise als korrekt gebildete Zeichenreihen verstanden werden und Definitionen als Kürzel, die „theoretisch entbehrlich", jedoch „typografisch üblich" sind. Gleichzeitig gibt es Probleme, die aus dem Bereich der metamathematischen Probleme herausfallen. Dazu gehören sämtliche Probleme, die sich auf die inhaltliche Mathematik und ihren Fortschritt und sämtliche Probleme, die sich auf die Situationslogik des mathematischen Problemlösens beziehen. Damit formuliert Lakatos auch ein Hauptproblem für Lernende mathematischer Inhalte, dass das wichtigste Lernwerkzeug der Mathematik in der formalen Darstellung des mathematischen Wissensgebäudes seit Mitte des 19. Jahrhunderts nur noch eine untergeordnete Rolle spielt. Dies ist aber allein in erkenntnistheoretischer Hinsicht der Fall, denn bei Betrachtung des einzelnen, Mathematik treibenden Individuums ist natürlich im Erkenntnisprozess die Verdeutlichung von Sachverhalten an Beispielen nach wie vor von entscheidender Bedeutung. Felix Klein schreibt 1898 (F. Klein 1889, S. 3):

> *„Indem ich für das Recht der Anschauung im Gebiete meiner Wissenschaft kämpfe, will ich die Bedeutung der logischen Entwicklung keineswegs hintansetzen. Nur da findet die Mathematik nach der Auffassung, die ich vertrete, ihre volle Geltung, wo beide Seiten nebeneinander zur Entfaltung kommen."*

Klein vertritt also die Auffassung, dass nur durch die mathematische Anschauung sich mathematische Kreativität entfalten kann und dadurch neue Ideen für das formale Gebäude der Mathematik entstehen.

Abbildung 1: Graphik zur Veranschaulichung eines topologischen Begriffs aus Felix Klein Protokolle, 1872-1912 (Band I, Seite 113, Seminar: "Über die Gruppe der Modulargleichung fuer Transformation pter Ordnung und specielle über die Transformation 25ter de elliptischer functionone." Seminar of Sunday, February 14, 1880, Quelle: Clay Mathematics Institute)

2.2 Heuristiken des mathematischen Problemlösens

„Wissen und Erkennen sind die Freude und die Berechtigung der Menschheit."

Alexander von Humboldt

Der Weg zur Lösung eines mathematischen Problems – z.B. eines mathematischen Satzes – ist in der Lösung selbst oft unsichtbar. Beim Lesen eines Beweises ist die Präzision, mit der Argumente und Schlüsse zusammengefügt sind, zu sehen und alle Teile fügen sich zu einer Einheit zusammen. Für den Leser ist oft nicht klar, wie man zu einer solchen Lösung kommt. Der fertige Beweis hat das Gesicht einer fehlerfrei zusammengefügten Zeichenkette. Selbige, einmal gefunden erklärt einerseits, formal alles, ohne in mathematischem Sinne den Hauch eines Zweifels zu lassen. Andererseits verrät sie manchmal kaum etwas über den Entstehungsprozess. Im Laufe der Beschäftigung mit Mathematik eignet sich jeder Mathematiker individuelle Heuristiken an, die mehr oder weniger nützlich sind, um neue mathematische Probleme zu lösen. Es entsteht ein Arsenal an Vorgehensweisen, die helfen, Lösungen zu finden. Zu Beginn der Beschäftigung mit Mathematik sind diese Vorgehensweisen wie ein Spiel mit den formalen Möglichkeiten der mathematischen Darstellung. Beispielsweise lassen sich viele Beweise in der Algebra „einfach so herunterrechnen”, das heißt, die formal saubere Formulierung der Fragestellung und die klare Definition des Ziels legen den Lösungsweg nahe. Die formale Formulierung wirkt also als „Denkbeschleunigung”, die es dem mathematischen Problemlöser ermöglicht, seine Gedanken durch das Formulieren zu fokussieren. Mit steigender Komplexität der Fragestellungen muss das Vorgehen immer differenzierter werden. Zwar ist es noch Ziel, am Ende eine abstrakte und lückenlose, formale mathematische Formulierung zu finden, aber die Argumente dafür reihen sich allein aus der Anwendung des Formalismus nicht mehr aneinander. An dieser Stelle kommt das (mathematische) Problemlösen nicht ohne Kreativität aus.

Mathematiker wie Gottfried Wilhelm Leibniz, Leonhard Euler, René Descartes und Bernard Bolzano, um nur einige zu nennen, haben sich damit beschäftigt, wie diese individuellen Heuristiken zum Problemlösen in der Mathematik beitragen. Im 20. Jahrhundert befasste sich der Mathematiker George Pólya mit Problemlösestrategien in der Mathematik mit Hilfe heuristischer Ansätze. Sein Buch „Schule des Denkens – Vom Lösen mathematischer Probleme” aus dem Jahre 1944 ist für viele Forscher, die sich mit diesen Fragestellungen beschäftigen, ein viel zitierter Ausgangspunkt (J. Mason et. al., 2008). Die Kernaussagen nach Pólya (G. Pólya, 1944) lassen sich so zusammenfassen: Problemlöseheuristiken in der Mathematik legt sich jeder Mathematiker individuell zurecht, die Hauptelemente der Lösungsfindung sind:

- Analogien,
- Beispiele,
- Zerlegung in Teilprobleme,
- Reduzierung der Fragestellung auf „zeichenbare” Darstellungen.

Obwohl das eigentliche Vorgehen sehr individuell ist, lässt sich doch ein Grundmuster erkennen, das für viele, die mathematisches Problemlösen erlernen wollen, nützlich ist und in die Problemlöseheuristiken Vieler einzuordnen sind (J. Mason et. al., 2008).

Abbildung 2: Grundstruktur des Problemlöseprozesses in der Mathematik nach George Pólya

Struktur von Problemlöseheuristiken in der Mathematik

Die Struktur, wie man das Lösen von Problemen in der Mathematik angehen kann, lässt sich grob in fünf Schritte (Beginn-Aktion-Reflexion-Aktion-Einsicht) einteilen, wobei in zwei Bereichen Kreisprozesse beim konkreten vorgehen auftreten, die beliebig oft durchlaufen werden können, um eine Lösung zu finden. Da diese Schritte auf den Problemlöser bzw. den Lernenden bezogen werden sollen, beschreibt Pólya (G. Pólya, 1944) jeden der Schritte als Zustand des Individuums, das sich mit Mathematik beschäftigt. Bei dieser Betrachtung gehen wir von einer bereits gestellten Fragestellung aus, die etwa durch einen Lehrer oder von der wissenschaftlichen Community vorgegeben ist.

Grundstruktur

Zustand: Beginn – Man muss die Frage verstehen

Prozess: Vermuten, d.h. es werden Vermutungen in alle Richtungen angestellt und Assoziationen gebildet, die Ansatzmöglickeiten für mögliche Pläne liefern.

Phase: Planung, d.h. es wird geplant, aber noch nicht ausprobiert; man lotet aus, ob die Voraussetzungen für angedachte Ansätze vorhanden sind.

Handlung: Spezialisieren, d.h. man versucht mit einfachen Spezialfällen der Problemstellung näher zu kommen und diese für Lösungsansätze zu nutzen.

Vorgehen: Das Vorgehen strukturiert sich in folgenden iterativen Kreisprozess Probieren –Versuche – Analogie – überprüfen im Kontext – Probieren ...

Fragen: Was ist unbekannt? Was ist gegeben? Wie lautet die Bedingung? Ist die Bedingung ausreichend, um die Unbekannte zu bestimmen? Ist die Aufgabenstellung unterbestimmt? Passen alle gegebenen Bedingungen zusammen? Kann ich ein Beispiel finden?

Aktivitäten: Zeichne eine Figur! Führe eine passende Bezeichnung ein! Lässt sich ein passendes Beispiel ausrechnen? Kann man die Bedingungen vereinheitlichen oder auftrennen? Kann man Teile der Bedingungen hinschreiben? Usw.

Zustand: Aktion – Ausdenken eines Plans

Prozess: Vermuten, d.h. es werden Vermutungen in aussichtsreiche Richtungen angestellt und Assoziationen gebildet, die Ansatzmöglickeiten für mögliche Pläne liefern (dieser Prozess ist ein Kreisprozess, der öfter durchlaufen werden muss, um eine Idee für einen Plan zu bekommen).

Phase: Planung, d.h. es wird geplant aber noch nicht ausprobiert, man lotet aus, ob die Voraussetzungen für angedachte Ansätze vorhanden sind.

Handlung: Spezialisieren, d.h. man versucht mit einfachen Spezialfällen der Problemstellung näher zu kommen und diese für Lösungsansätze zu nutzen.

Vorgehen: Das Vorgehen strukturiert sich in folgenden iterativen Kreisprozess Probieren – Versuche – Analogie – überprüfen im Kontext – Probieren ...

Fragen: Habe ich eine solche Aufgabe früher schon einmal gesehen? Kenne ich verwandte Aufgaben? Welche Sätze aus dem zu bearbeitenden Bereich könnten hilfreich sein? Habe ich diese verstanden? Kann man das Ergebnis von ähnlichen Aufgaben ver-

wenden? Kann man die Bedingungen elementarer hinschreiben? Kann man die Aufgabe anders ausdrücken? Gibt es eine allgemeinere Fragestellung, die das Problem löst? Ist es nützlicher, abstrakter oder konkreter vorzugehen? Benutze ich in meinem Plan alle gegebenen Bedingungen?
Aktivitäten: Aufschreiben aller relevanten Sätze aus dem Themengebiet, Ausrechnen einiger konkreter Beispiele, Zeichnen einer Planfigur, Abspalten möglicher Teilprobleme, Aufstellen eines Plans.

Zustand: Reflexion – Reflexion über den Plan
Prozess: Übergang vom Vermuten zum Beweisen, d.h. eine Richtung ist eingeschlagen und wird zielstrebig verfolgt; mathematische Präzision wird immer wichtiger.
Phase: Durchführung, d.h. der Plan wird kritisch gewürdigt und auf mögliche Probleme untersucht; Begriffe werden exakt definiert und von ihren Eigenschaften her eingegrenzt.
Handlung: Übergang vom Spezialisieren zum Verallgemeinern, d.h. die Charakteristika der Spezialfälle werden untersucht; die einfachen Spezialfälle des letzten Zustands werde mehr und mehr verallgemeinert.
Vorgehen: Das Vorgehen strukturiert sich in folgenden iterativen Kreisprozess Probieren – Versuche – Analogie – überprüfen im Kontext und Kontext verändern – Probieren....
Fragen: Wo liegen bei meinem Plan die schwierigen Stellen? Was bedeutet die Ausführung des Plans für mein „mathematisches Weltbild”? Was ist der Kernpunkt an den schwierigen Stellen meines Plans? Habe ich so ähnliche Pläne schon einmal umgesetzt? Gibt es Stellen, an denen ich Rechenaufwand vermeiden kann? Sind meine Abkürzungen und Begriffe geeignet?
Aktivitäten: Reflektieren über den Plan; Nachdenken, ob man analoge Pläne schon einmal ausgeführt hat.

Zustand: Aktion – Ausführung des Plans
Prozess: Beweisen, d.h. eine Richtung ist eingeschlagen und wird zielstrebig verfolgt; mathematische Präzision ist unabdingbar und oberste Maxime des Handelns (dieser Prozess ist ein Kreisprozess, da es beim Ausführen des Planes zu unvorhergesehenen Schwierigkeiten kommen kann, die einen neuen Plan erfordern).

Phase: Durchführung, d.h. die geplanten Beweisschritte werden Zug um Zug in die formale Sprache der Masthematik umgesetzt und überprüft.
Handlung: Verallgemeinern, d.h. die Spezialfälle vom Beginn spielen nun keine Rolle mehr, außer, dass sie nützliches Anschauungsmaterial für die Ausführung des Plans sind.
Vorgehen: Plan ausführen, d.h. Beweisschritte umsetzen und den „Beweis" formal fehlerfrei „hinschreiben".
Fragen: Kann ich an jeder Stelle begründen, dass mein Plan richtig ist? Finde ich für jeden meiner Schritte eine kurze Begründung? Könnte ich jede dieser Begründungen beweisen?
Aktivitäten: Plan ausführen und jeden einzelnen Schritt des Plans kontrollieren.

Zustand: Einsicht – Ausführung des Plans
Prozess: Beweisen, d.h. eine Richtung ist eingeschlagen und wird zielstrebig verfolgt; mathematische Präzision ist unabdingbar und oberste Maxime des Handelns.
Phase: Rückblick, d.h. der „hingeschriebene" Beweis wird Schritt für Schritt überprüft und kritisch gewürdigt.
Handlung: Verallgemeinern, d.h. die Spezialfälle vom Beginn spielen nun keine Rolle mehr, außer, dass sie nutzliches Anschauungsmaterial für die Ausführung des Plans sind.
Vorgehen: Lösung prüfen, d.h. die gedungene lösung wird mit Hilfe einer „Probe", einer Folgerung, einer Einordnung und dem reduzieren auf Spezialfälle auf Plausibilität überprüft.
Fragen: Kann ich das Resultat, das ich erhalten habe, kontrollieren? Kann ich den Beweis kontrollieren? Kann ich Folgerungen ziehen, die mir zeigen, dass mein Beweis richtig ist? Kann ich den bewiesenen Satz in anderen Zusammenhängen anwenden? Kann ich das Ergebnis noch auf eine andere Weise ableiten? Ist mein verwendeter Plan auch für andere Aufgaben und Zusammenhänge nützlich?
Aktivitäten: Satz Anwenden, Überlegen, wie das Resultat in mein „mathematisches Weltbild" passt? „Die Probe machen".

Die oben beschriebene Problemlöseheuristik stammt in ihrer Urform von George Pólya (1949). Sie wurde von John Mason, Leone Burton und Kaye Stacey fortgeführt, die mit ihrem Buch „Mathematisch denken – Mathematik ist keine Hexerei” (J. Mason, 2008) eine sehr erfolgreiche Einführung für Mathematikstudenten in diesem Bereich verfasst haben.

Das Suchen vieler Lehrender nach einem Konzept, mit dem sie ihren Studierenden das Lösen mathematischer Probleme nahe bringen können, zeigt, dass der mathematische Formalismus allein wenig zum kreativen Suchen und Finden von Lösungen beiträgt. Es zeigt vielmehr das grundsätzlich philosophische Problem der Mathematik, das Imre Lakatos in seinem Buch „Beweis und Wiederlegung - die Logik mathematischer Entdeckungen” beschreibt (I. Lakatos, 1976). Er zeigt an mehreren, für die Mathematik wichtigen Beispielen auf, dass der formale Beweis eines Satzes nur begrenzt zum Verstehen des Problems beiträgt. Eines der Grundprobleme beim Erlernen von mathematischen Konzepten ist, dass der Mensch zunächst immer von seiner individuellen Erfahrung, von einem erlebten Einzelfall oder von einer konkreten eigenen Handlung ausgeht. Das grundlegende Prinzip der Mathematik ist es, Sachverhalte ein für alle Mal zu beweisen und in der Folge deduktiv vom Allgemeinen auf den Einzelfall zu schließen. Diese Vorgehensweise der Mathematik ist für die Wissenschaftliche von substanzieller Bedeutung und macht den Kern mathematischen Denkens aus. Aber „der“ Mathematiker als Individuum geht in den Beweisen, in den Schlüssen und in der Anwendung von anschaulichen Einzelphänomenen aus, die z.B. durch Induktion verallgemeinert werden. D.h., der kreative Prozess des Suchens und Findens nach Lösungen von mathematischen Fragestellungen verläuft innerhalb einer ganz anderen Logik als derjenigen der Mathematik, wenn sie als Wissenschaft des formalen Schließens interpretiert wird. Die Kunst besteht darin, zwischen beiden Polen ein Gleichgewicht herzustellen. Eine Möglichkeit ist das Zeichnen von Graphiken und Planfiguren, die einen Sachverhalt auf eine konkretere Weise als die formale Beschreibung darstellen. Ein Weiteres Konzept ist, sich einen Sachverhalt durch konkrete und reale Beispiele exemplarisch zu verdeutlichen. Um diese beiden Hilfestellungen bei der Lösung mathematischer Probleme soll es im nächsten Abschnitt gehen.

2.3 Die Graphik in der Mathematik

„Je planmässiger die Menschen vorgehen,
desto wirksamer vermag sie der Zufall zu treffen.“
Friedrich Dürrenmatt

Das Fach Mathematik und speziell der Themenschwerpunkt „Stochastik“ gehören in den allgemein bildenden Schulen zu den unbeliebtesten Fächern schlechthin. Für viele Menschen ist der Umgang mit Wahrscheinlichkeiten ein Buch mit sieben Siegeln. Es scheint, dass das Denken in Wahrscheinlichkeiten den als Kind gemachten Erfahrungen widerspricht. Beispielsweise halten viele Menschen beim Lotto spielen eine ungeordnete Zahlenkombination wie 12, 34, 42, 3, 28 und 44 für wesentlich wahrscheinlicher als die Zahlen 1, 2, 3, 4, 5, 6. Dies kommt daher, dass der Mensch intuitiv die ungeordnete Zahlenfolge in eine Kategorie einordnet und die geordnete in eine andere. Der intuitive Eindruck, dass eine ungeordnete Folge wahrscheinlicher ist, kommt daher, dass viel mehr ungeordnete, auf einander folgende Zahlenfolgen existieren, als es geordnete Folgen gibt. In Wirklichkeit sind beide Kombinationen gleich wahrscheinlich. Selbst einfache Prozentrechnungen stellen viele Menschen vor ungeahnte Probleme, wie folgende Beispiele aus der Presse belegen.

- Aus der Agenturmeldung: „Sechs Prozent aller Münchener Theaterkarten sind Freikarten” wird in einem Artikel der Süddeutschen Zeitung zu, „Jede sechste Karte ist eine Freikarte.” (P. Dittrich, 2008)
- Oder aus dem Nachrichtenmagazin Focus: „Statt sieben registrierten die Forscher 14 Todesfälle, einen Anstieg um 50%.”(P. Dittrich, 2008)
- Oder aus einer Presseerklärung der SPD „ ...die erfolgreiche Haushaltskonsolidierungspolitik von Hans Eichel muss fortgesetzt werden, da heute schon fast jeder zweite Euro (d.h. jede vierte Mark) für den Schuldendienst verbraucht wird.” (P. Dittrich, 2008)

Einfache Fragestellungen aus der Geometrie sind für viele Menschen wesentlich klarer zu begreifen als Fragen, die in irgendeiner Weise mit Wahrscheinlichkeiten zu tun haben. Dies liegt darin begründet, dass ein Mensch in seiner frühkindlichen Entwicklung sehr wohl schon geometrische Zusammenhänge wahrnimmt und diese erkennen kann, obwohl er weder die Begriffe noch den mathematischen Hintergrund kennt. Wie z.B. ein Kind, das mit Bauklötzen spielt und die Begriffe „rund" und „rollen" nicht kennt, aber ganz klar die Zusammenhänge an einer schiefen Ebene erfasst. Daher erscheint es nicht überraschend, dass es in der Menschheits- und Wissenschaftsgeschichte unzählige Beispiele für die geometrische und zeichnerische Darstellung der Häufigkeitsauswertungen von Naturbeobachtungen gibt. Diese frühen Informationsgraphiken stellen für viele Menschen einsichtig und schlüssig komplexe empirische Sachverhalte dar. Die Himmelsscheibe von Nebra mit einem Alter von 3500 Jahren ist hierfür ein Beispiel[1]. Die Himmelsscheibe zeigt die Beobachtungsergebnisse von vielen Jahrzehnten (H. Wußing, 2008 S. 15), wenn nicht gar Jahrhunderten in einer Form, die nicht nur den Experten (Herstellern der Scheibe) zugänglich war.

Abbildung 3: Himmelsscheibe von Nebra - eine frühe Informationsgraphik ; gefunden bei Halle ca. 3500 Jahre alt (H. Wußing 2008, S. 16)

[1] eine von vielen Spekulationen zur Funktion der Scheibe (H. Wußing, 2008 S.16)

Die Darstellung von Daten mit Hilfe von Graphiken und das Zurückführen von statistischen Fragestellungen auf ein geometrisches Problem ist ein naheliegendes Mittel der Problemlösung (Theus 1996, S.1-10).

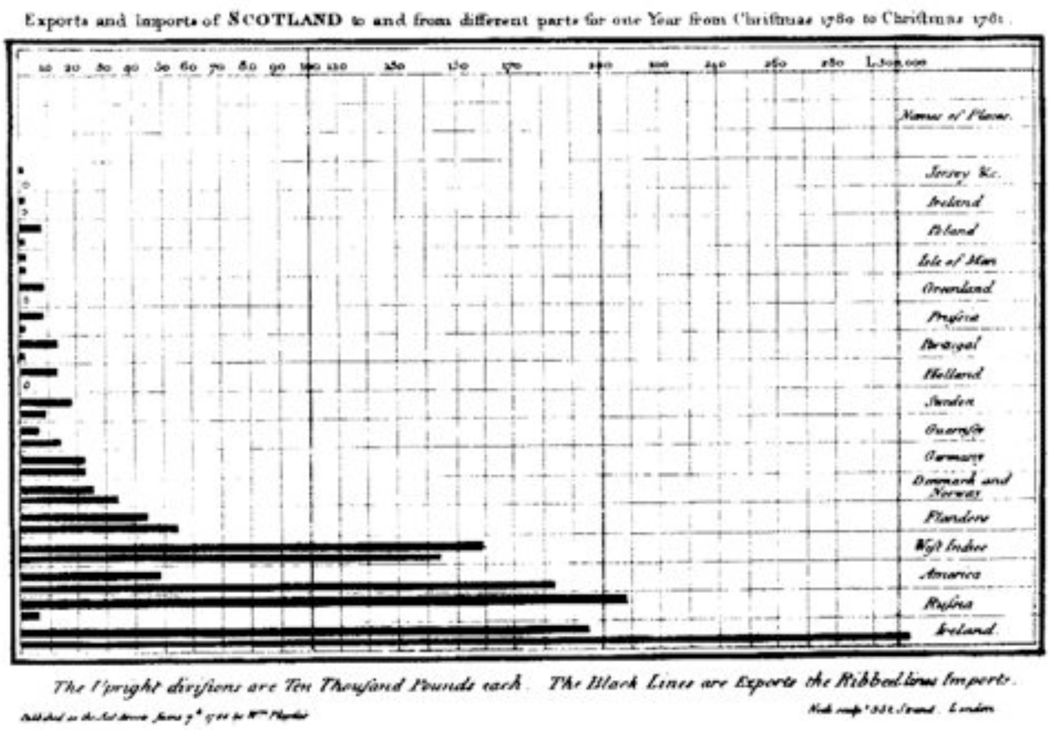


Abbildung 4: Ein frühes Säulendiagramm (Barchard): William Playfair zeichnete 1780 dieses Säulendiagramm

Die moderne Statistik entwickelte sich im Spannungsfeld zwischen graphischer Darstellung von Daten in immer neuen Formen von Diagrammen und der mathematischen Präzision der Maßtheorie (Theus 1996, S. 3ff). Diese beiden Bereiche stellen zwar methodisch Gegenpole zwischen Anschauung und Abstraktion dar, bilden aber inhaltlich bei der Bearbeitung konkreter Probleme eine Einheit. Die Darstellung von Daten in Graphiken hat zwei Aufgaben: Erstens die Aufarbeitung und Darstellung von Forschungsergebnissen auf eine Weise, die viele Menschen auch ohne mathematische Expertenkenntnis nachvollziehen können und zweitens die Aufarbeitung von Daten zur Ideengenerierung für den mathematischen Experten. Letzteres war vor der Entwicklung der interaktiven statistischen Graphik äußerst mühevoll oder durch den enormen Aufwand der Darstellung von Millionen Fällen mit Hunderten von Merkmalen aus praktischen Gründen völlig unmöglich. Dennoch zeigen einzelne, hervorragende statistische Graphiken auf Papier das Potenzial, das die graphische Darstellung für die Bearbeitung realer statistischer Fragestellungen hat. Im 18. Jahrhundert erlebte die statistische Graphik eine sprunghafte Weiterentwicklung. William Playfair und Charles Joseph Minard erstellten Graphiken von hervorragender Qualität, die neben ihrer Bedeutung für die Statistik, richtige Kunstwerke sind (Theus, 1996, S. 3).

Abbildung 5: Eine Karte mit statistischen Einträgen: Dr. John Snow analysierte eine Choleraepidemie 1854 mit Hilfe einer Karte in der die Todesfälle eingetragen waren.

In der Zeit um 1854 wurden, Plots wie Histogramme, Balkendiagramme (Barchart), Karten und Zeitreihendarstellungen systematisch angewendet und weiter entwickelt. Auch statistische Laien wie der Arzt John Snow lösten mit grafischen Darstellungen aktuelle Probleme der damaligen Zeit. Die Karte von John Snow zeigt z.B., dass die Krankheit vor allem von einem verschmutzten Brunnen in einem bestimmten Bereich der Stadt ausging. Im Laufe der Entwicklung der Statistik während der letzten 400 Jahre gab es ein periodisches Auf und Ab der Methoden (Theus, 1996, S. 3). Statistische Problemlösung mit Hilfe von Graphiken wechselte sich mit analytischen Methoden ab. Bedingt durch den ungeheuren Aufwand, Graphiken für eine große Fallzahl abzubilden, wurden sie meist nur dort verwendet, wo es keine andere Form der Darstellung mehr gab. Eine größere Verbreitungsdichte hatten statistische Graphiken in Form von Infographiken, mit denen die Ergebnisse empirischer Forschung einem breiten Publikum plausibel gemacht werden sollten. Erst seit Mitte der 1970er Jahre werden interaktive statistische Graphiken zur Problemlösung eingesetzt. Die Konjunktur graphischer Methoden hat vor allem praktische Gründe: Durch leistungsfähige Computer, die in der Lage sind, schnell und effizient Graphiken mit vielen Datenpunkten zu erzeugen, ist es dem Datenanalysten möglich, zeitnah und flexibel Graphiken für die Problemlösung einzusetzen. Frühere Konzepte, die sich im 18. Jahrhundert nur aus praktischen Gründen nicht zur Problemlösung einsetzen ließen, wurden in Software umgesetzt und erfolgreich verwendet.

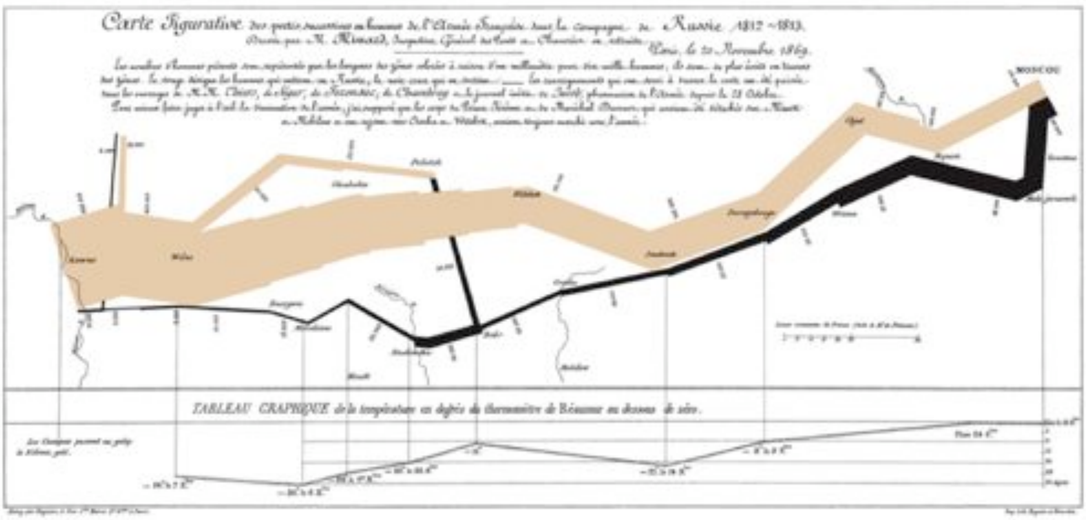


Abbildung 6: Eine Infographik zu Napoleons Russlandfeldzug 1812- 1813: Charles Joseph Minardes erstellte 1869 eine Infografik, die die Verluste Napoleons während des Russlandfeldzuges zeigt.

Zusammenfassung: Strategie des Problemlösens in der Mathematik

Die Mathematik ist auf das Engste mit „Lernen" verknüpft, was sich in der Wortetymologie nachverfolgen lässt. Das wichtigste Instrument in der Mathematik ist die Problemlösung und das immer weitere Ableiten unumstoßbarer Sätze. Daraus lässt sich die Sonderstellung der Mathematik in der Erkenntnistheorie ableiten. Der Formalismus und die Abstraktion, die das mathematische Schließen kennzeichnen, haben den Vorteil, dass sie das Denken und das Problemlösen beschleunigen können, und den Nachteil, dass sie manchmal mögliche Lösungsansätze verschleiern. Jeder Mathematiker entwickelt individuelle Problemlöseheuristiken, die ihm helfen, mit dem mathematischen Formalismus Probleme zu lösen. Diese Heuristiken lassen sich in einer Grundstruktur zusammenfassen. Diese ist von zwei dynamischen Kreisprozessen gekennzeichnet, die den Mathematiker über mehrere Iterationsschritte seiner Lösung näher bringen. Ein erfolgreiches Werkzeug, das sich die Mathematiker zurecht gelegt haben, ist das Zurückführen von abstrakten Zusammenhängen auf geometrische Probleme. In der Statistik gibt es dazu eine analoge Vorgehensweise, in der man unübersichtliche Datensätze in statistische Graphiken übersetzt. Diese Graphiken haben eine lange Tradition, eignen sich aber aufgrund des Aufwands ihrer Herstellung erst seit der Erfindung des Computers zur Problemlösung im großen Umfang.

2.5 Die Idee der Explorative Datenanalyse

„Die gefährlichste Weltanschauung
ist die Weltanschauung derjenigen,
die die Welt nicht angeschaut haben."

Alexander von Humboldt

Die Explorative Datenanalyse ist eine adaptive Vorgehensweise. Ihr Ziel ist, die Information in Datensätzen zugänglich und sichtbar zu machen, indem man statistische Modelle und Graphiken einsetzt, um Dateneigenschaften zu erkennen. Obwohl statistische Verfahren einen wichtigen Beitrag in der Explorative Datenanalyse leisten, spielt die interaktive Graphik eine zentrale Rolle, weil sie in der Lage ist, unerwartete Einsichten zu gewähren und auf neue Ideen hinzuweisen. Die ist vor allem bei multivariaten Datensätzen von entscheidender Bedeutung, da sonst viele Zusammenhänge, die in den Daten liegen, unentdeckt bleiben. Dadurch ergänzen sich Graphiken und Modelle (im statistischen Sinne) gegenseitig. Ein Resultat, das einer Graphik entstammt, soll womöglich mit einem Modell überprüft werden und umgekehrt. Interaktive graphische

Methoden sind ein besonders wirksamer Ansatz, da sie mehrere Ansichten gleichzeitig anbieten; daher kann man auf gewonnene Einsichten flexibel durch die Anpassung der Darstellung reagieren.

Die interaktive Graphik liegt an der Schnittstelle zwischen der Abbildung eines realen Problems in einem Datensatz und gleichzeitig an der Schnittstelle zwischen Datensatz und Modell. Um die grundsätzliche Idee der Explorativen Datenanalyse zu verdeutlichen, sollen an dieser Stelle zwei Analogien angeführt werden. Dabei werden die Analogien aus der Perspektive des Faches sowie aus der Perspektive des einzelnen Forschers betrachtet.

Analogie des Vorgehens in der Evolutionsforschung und in der Datenanalyse

Das Vorgehen bei der Exploration kann man am ehesten mit dem Vorgehen in der Biologie, genauer mit dem Erforschen evolutionärer Zusammenhänge, vergleichen. Zuerst sammelt man Daten, versucht diese zu klassifizieren und ihre Besonderheiten herauszufinden. So wird jeder Datensatz zu einem Puzzleteil, dessen Anknüpfungspunkte bekannt und klassifizierbar sind. Die große Aufgabe besteht darin, die einzelnen Puzzleteile in einem konstruktiven Prozess zu einer These zusammenzufügen. Das Verknüpfen vieler einzelner Thesen führt schließlich zu einem Gesamtbild der Evolution, das als Ganzes empirisch überprüfbar wird.

Analogie zur Verdeutlichung des Vorgehens bei der explorativen Datenanalyse

Das Vorgehen bei der Explorativen Datenanalyse ist vergleichbar mit dem Vorgehen eines Naturforschers aus Europa, der zum ersten Mal in seinem Leben in den brasilianischen Regenwald kommt und mit seinen Forschungen beginnt.
Zunächst ist er überwältigt von den vielen neuen Eindrücken und ihm fehlt jeder Überblick über das Gesehene. Nach und nach betrachtet er einzelne Pflanzen und Tiere und vergleicht Merkmale mit der Flora und Fauna seiner Heimat. Er erkennt Übereinstimmungen und Unterschiede und beginnt nach einiger Zeit, Phänomene zu klassifizieren.

Nach einigen Tagen und Wochen ist er mit der neuen Umgebung vertraut und hat sich ein neues gedankliches Gerüst konstruiert, in das er die neuen Erkenntnisse einordnet und miteinander verknüpft.
Im nächsten Schritt wird er beginnen, gezielt Experimente zu machen, Wirkzusammenhänge innerhalb des Urwaldes zu beobachten und die von ihm konstruierten Klassen für Tiere und Pflanzen in Beziehung zu setzen. So bilden sich Erklärungsmuster heraus, denen die eigenen stichprobenartigen Beobachtungen über die Zusammenhänge im Urwald zugrunde liegen. Diese Vermutungen müssen sich mit standardisierten Beobachtungen überprüfen lassen. Passt eine Hypothese nicht zu den neu gemachten Beobachtungen, wird diese verändert und um die neu gewonnenen Erkenntnisse erweitert.
Der Naturforscher wird ständig seine Hypothesen und seine Erkenntnisse mit dem vorhandenen Wissen über die Pflanzen und Tiere seiner Heimat vergleichen. Findet er Übereinstimmungen, wird er davon ausgehen, dass es sich um allgemein gültige Gesetze handelt. Findet er Widersprüche, so werden diese für ihn als neue Forschungsgegenstände interessant.

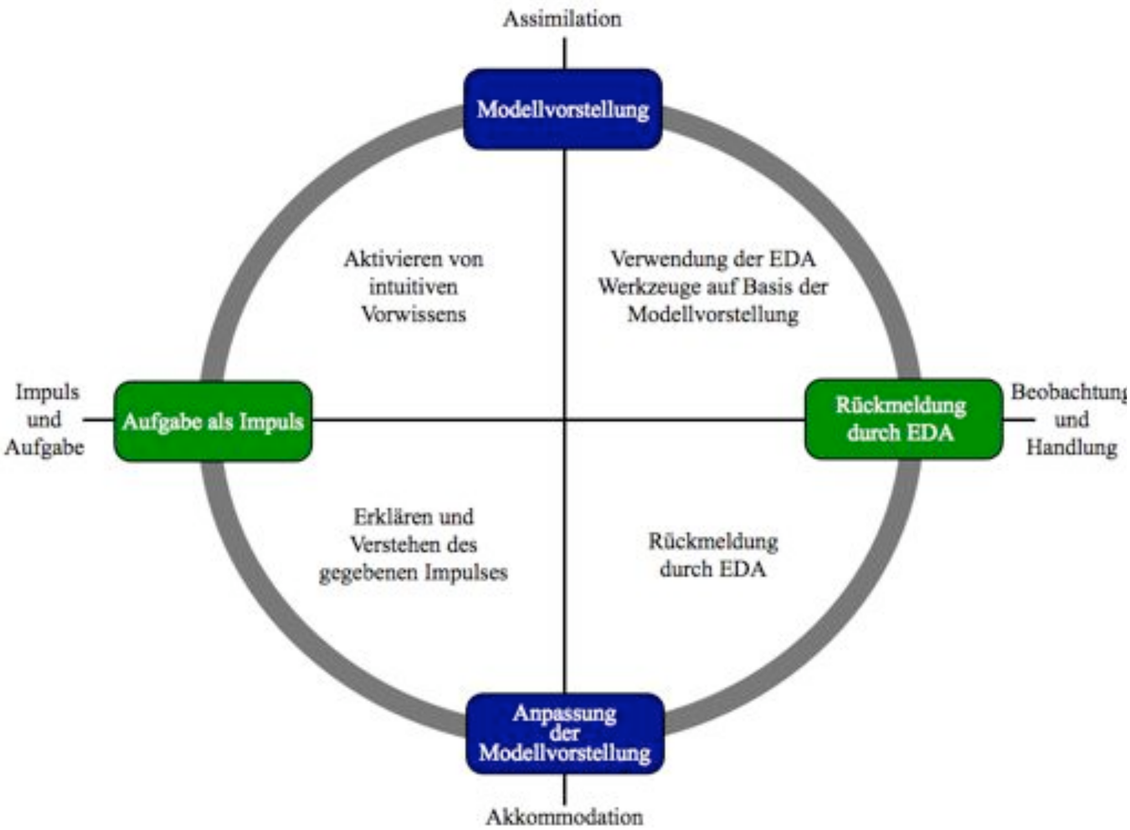


Abbildung 7: Verbesserung von statistischen Konzepten durch Datenerfahrungen

Die Explorative Datenanalyse wirkt in zwei Richtungen. Die eine Richtung führt vom Forscher zum Untersuchungsgegenstand, denn die Explorative Datenanalyse eröffnet dem Forscher in dieser Blickrichtung ein Fenster, aus dem er den Forschungsgegenstand in neuen Blickwinkeln betrachten kann. Die andere Ausrichtung zeigt vom Untersuchungsgegenstand auf den Forscher. Wenn er sich die unmittelbaren Eindrücke durch dieses Fenster nicht erklären kann, muss er seine Konzepte anpassen. Der Vorzug der Explorativen Datenanalyse liegt in diesem Fall darin, dass die Eindrücke, die auf den Forscher wirken, sehr direkt sind und ohne weitere Abstraktionsebene auf ihn wirken. Das gelingt der Explorativen Datenanalyse dadurch, das ihr Hauptwerkzeug, die interaktive statistische Graphik, stark an die Wahrnehmungsfähigkeiten des Menschen angepasst ist. Durch interaktive Manipulationsmöglichkeiten kann jeder Forscher seinen individuellen Weg für die Annäherung an den Forschungsgegenstandes finden.

2.6 Klassen statistischer Probleme

„Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen,
sind sie nicht sicher, und insofern sie sicher sind,
beziehen sie sich nicht auf die Wirklichkeit."
Albert Einstein

Seit den Arbeiten von John W. Tukey (1962, S. 1-67) gehört das Suchen (Exploration) neben dem Beschreiben (Deskription) und dem Schließen (Induktion) zu den Grundaufgaben der Statistik. Die daraus abgeleitete Strategie der Statistik wird als Explorative Datenanalyse bezeichnet. Es liegt in der Natur eines derart komplexen und interaktiven Vorganges wie der Suche nach einem statistischen Modell für einen realen Datensatz, dass er sich nicht schematisch abhandeln lässt. Das Leitbild, das Tukey für die Explorative Datenanalyse geprägt hat, ist das eines Detektivs, welcher ausgehend von einem Problem in den Daten, interessante Strukturen und Besonderheiten aufdeckt, Hinweisen nachgeht und Hypothesen entwickelt (Tukey 1962). Die deskriptive Statistik begnügt sich damit, die Verteilung eines Merkmals zu beschreiben. Die Explorative Datenanalyse jedoch geht einen Schritt weiter und versucht, Fragen zu klären wie: „Was ist an einer Verteilung eines Merkmals ungewöhnlich oder bemerkenswert?” oder „Wo bieten sich Ansätze für ein statistisches Modell?”.

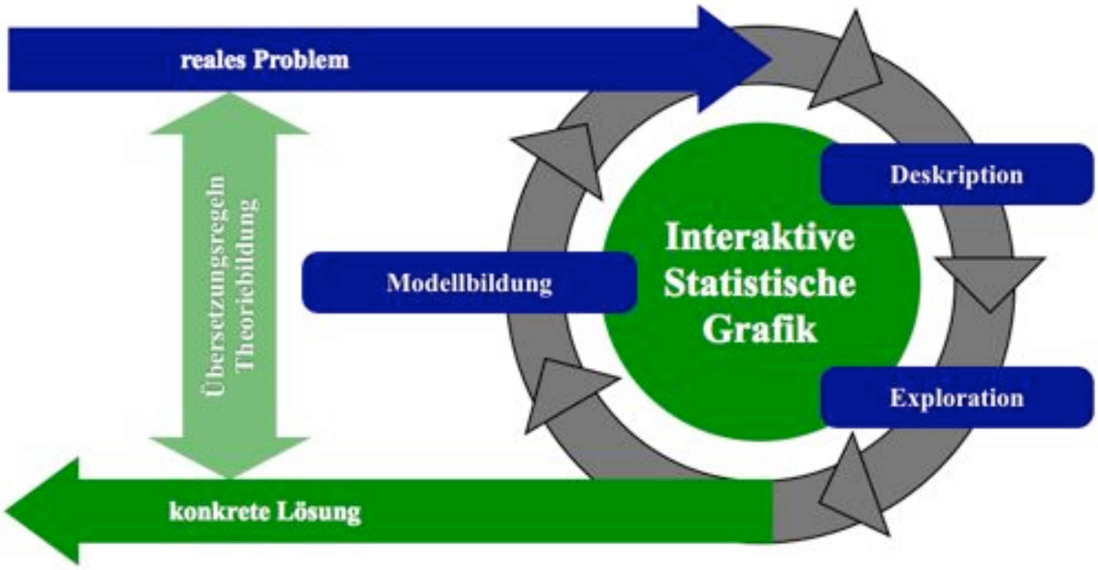


Abbildung 8: Problembearbeitung mit Hilfe der explorativen Datenanalyse

Erst durch das Beantworten dieser beiden Fragen ist es möglich, die gesamte Vielfalt und den Reichtum der Mathematik im Hinblick auf die induktive Statistik auf reale und komplexe Datensätze anzuwenden. Um eine reale Situation zu analysieren und Ansatzpunkte für komplexe Modelle zu finden, die über Erkenntnisse von Laborexperimenten hinausgehen, ist es notwendig, umfangreiche Datenmengen zu sammeln, um die entscheidenden Daten für ein bestimmtes Phänomen zu generieren. Werden die mit großem Aufwand gesammelten Daten nur mit vorgefertigten Modellvorstellungen aus der Theorie ausgewertet und nicht detailliert untersucht, so bleiben in vielen Fällen fundierte Informationen, die hinter Daten stecken, unentdeckt. Die Exploration schließt sozusagen die Lücke zwischen Deskription und Induktion. Für den Betrachter, der sich noch nicht mit der Exploration beschäftigt hat, stellt sich die Frage, wie groß diese Lücke ist und bei welcher Art von Fragestellungen die Vorzüge der Exploration zum Tragen kommen und wie das Ganze konkret funktionieren soll. Dabei hilft es, die statistischen Fragestellungen in drei Problemklassen einzuteilen.

Klasse 1: Auswertung von Laborexperimenten

Die Definition für ein Laborexperiment lautet: „Ein Laborexperiment ist dadurch charakterisiert, dass unter kontrollierten Bedingungen in einem abgeschlossenen künstlichen Raum wenige Variablen variiert werden, um eindeutige kausale Wirkungen zu untersuchen und zu quantifizieren.“ (U. Fahrner & A.Unwin, 2008, S. 156f)
Bei der Auswertung von Laborexperimenten führen sachlogische Überlegungen zu einem qualitativen Zusammenhang und zu einem konkreten mathematischen Modell, dessen Parameter und seine Allgemeingültigkeit bestimmt werden müssen. Die Aufgabe der Statistik ist es, das Modell zu quantifizieren, d.h. die Größe der Parameter zu ermitteln und dafür Fehlerschranken anzugeben. Außerdem müssen durch die Versuchsplanung Aussagen zur Allgemeingültigkeit (Signifikanz) des Modells gemacht werden, sofern diese nicht schon durch sachlogische Überlegungen ermittelt wurden. Beispiele hierfür sind ein physikalisches Experiment der klassischen Mechanik, ein psychologisches Experiment in der Art der Versuche von Pawlow, Bestimmungen der Festigkeit im Bereich der Ingenieurwissenschaften usw. All diese Beispiele haben gemeinsam, dass die Randbedingungen und Fragestellungen klar durch das Problem vorgegeben sind.

Tabelle 1: Charakterisierung von Laborexperimenten (U. Fahrner & A.Unwin, 2008 S. 156f)

	Größenordnung	**Beispiel: Freier Fall**
Anzahl der Fälle	viele	100
Anzahl der Variablen	wenige	2
Modell	aus der Theorie bekannt	$g = 2h/t^2$

Klasse 2: Auswertung von Feldexperimenten

Die Definition des Begriffes „Feldexperiment“ lautet: „Ein Feldexperiment ist dadurch charakterisiert, dass unter teilweise kontrollierten Bedingungen in einem abgeschlossenen natürlichen Raum (Feld: Althochdeutsch für weites Land und für abgezäunte Weide) Variablen beobachtet werden, um ihre Wirkungen zu untersuchen und zu quantifizieren.“ (U. Fahrner & A.Unwin, 2008, S. 156f).

Feldexperimente, die mit klassischen Methoden ausgewertet werden, sind durch relativ übersichtliche Datensätze gekennzeichnet. Das bedeutet, dass sie bestenfalls aus sehr vielen Fällen, aber nur aus sehr wenigen Variablen mit wenigen Kategorien bestehen, sodass die Fragestellungen schon relativ konkret und klar sind. Durch Vorüberlegungen aus der Theorie und/oder mit Pilotexperimenten kann sichergestellt werden, dass die betrachteten Variablen das Feldexperiment vollständig beschrieben haben und die nicht benutzten Informationen für das Ergebnis irrelevant sind. Zu Beginn der Auswertung sind sowohl die Modellbildung und die dabei gemachten Fehler als auch die Ansätze für die Interpretation bekannt. Fragen der Art von „Welche Informationen liegen außerhalb der konkreten Modellvorstellung in einem multivariaten Datensatz vor?" können nur schwer beantwortet werden. Zusammengefasst kann man konstatieren, dass zwar Hypothesen bestätigt, aber kaum neue Hypothesen generiert werden können. Feldexperimente zeichnen sich dadurch aus, dass in Wissensgebieten, in denen es schon Modellvorstellungen gibt, die Modellparameter relativ präzise bestimmt werden können und das resultierende Ergebnis Allgemeingültigkeit besitzt.

Tabelle 2: Charakterisierung von Feldexperimenten (U.Fahrner & A.Unwin, 2008, S. 156f)

	Größenordnung	**Beispiel: Mendelsche Vererbungsgesetze**
Anzahl der Fälle	viele	100000
Anzahl der Variablen	viele	4
Modell	aus der Theorie bekannt	Kombinatorisches Modell: Aa Ab Ba Bb

Klasse 3: Untersuchungen in der realen Welt

Die Definition von „Untersuchungen in der realen Welt" lautet: „Eine Untersuchung in der realen Welt ist dadurch charakterisiert, dass unter beobachtbaren, aber nicht zu kontrollierenden Bedingungen Variablen beobachtet werden, um ihre Wirkungen zu untersuchen und zu quantifizieren." (U. Fahrner & A.Unwin, 2008, S. 156f).

Der zu untersuchende Datensatz ist unübersichtlich. Vermutungen über qualitative Zusammenhänge sind schwierig, weil nur wenige, eher diffuse Randbedingungen vorgegeben sind; die Fragestellungen sind kaum konkret; man befindet sich oft auf wissenschaftlichem Neuland, und man hat kaum oder gar keine Ideen und Anhaltspunkte für Hypothesen. Fragestellungen der Klasse 3 lassen sich mit den Werkzeugen der Explorativen Datenanalyse statistisch bearbeiten, indem man sie mit Hilfe der Exploration in Probleme der Klasse 2 zerlegt und diese schließlich in Probleme der Klasse 1 überführt. Konkrete Szenarien, die untersucht werden sollen, sind häufig der Klasse 3 zuzuordnen. Die „Kunst“ bei der Lösung dieser Probleme besteht darin, das Gesamtsystem in Subsysteme zu unterteilen, die sich nicht gegenseitig beeinflussen und deren Wirkmechanismen sich durch Superposition mit dem Wirkmechanismus des Gesamtsystems verknüpfen lassen. Das Unterteilen in Subsysteme muss sowohl datenanalytisch als auch sachlogisch begründet und untermauert werden können. Dadurch wird das willkürliche „Suchen” nach Daten ausgeschlossen.

Tabelle 3: Charakterisierung von Untersuchungen in der realen Welt (U. Fahrner & A.Unwin, 2008, S. 156f)

	Größenordnung	**Beispiel: Eine Unterrichtsstunde**
Anzahl der Fälle	wenige	Klassenstärke
Anzahl der Variablen	viele	unbekannt
Modell	unbekannt	unbekannt

2.7 Werkzeuge der Explorativen Datenanalyse

„Ein Mann, der recht zu wirken denkt,
muss auf das beste Werkzeug halten.”

Johann Wolfgang von Goethe

Das wichtigste Hilfsmittel (M.Theus, 1996, S. 24ff) der interaktiven Graphik ist der Computer. Mit leistungsfähigen Rechnern ist es möglich, sehr große Datensätze z.B. (Millionen von Fällen, tausenden von Variablen) aus verschiedenen Perspektiven zu betrachten. Mit „Perspektiven" sind dabei verschiedene interaktive Plots (d.h. Graphiken und Diagramme mit interaktiven und über die Daten vernetzten Funktionalitäten) gemeint, die alle miteinander „vernetzt" sind. In den letzten Jahren wurde eine große Zahl von Plots in interaktiver Software umgesetzt und es wurden neue Arten von Plots entwickelt, welche zu den wichtigsten Werkzeugen der interaktiven Graphik gehören. Dieses Phänomen geht von einfachen Streudiagrammen, Histogrammen, Barcharts (= Histogramme für kategorielle Variablen) über Boxplots (= eine graphische Darstellungsweise für Mittelwert, Streuung und Spannweite) bis hin zu komplexen Darstellungen von Daten in Mosaikplots oder Trees (= Baumdarstellungen, die Hinweise auf die Relevanz von Variablen geben). Allen gemeinsam ist das „Highlighting": Selektierte Daten werden in allen Plots gleichzeitig markiert. Eine Veränderung in einem Diagramm wirkt sich so unmittelbar in allen anderen erstellten Diagrammen aus. An dieser Stelle sollen die Schlüsselbegriffe der interaktiven statistischen Graphik noch einmal zusammengefasst werden, um den hohen Interaktionsgrad darzustellen und Ansatzpunkte für den Einsatz „neuer Medien" aufzuzeigen. Die Schlüsselbegriffe der interaktiven statistischen Graphik sind die sechs Grundelemente „Highlighting“, „Selektion“, „Abfragen“, „Warnungen“, „Linking“ und „Umordnungen“.

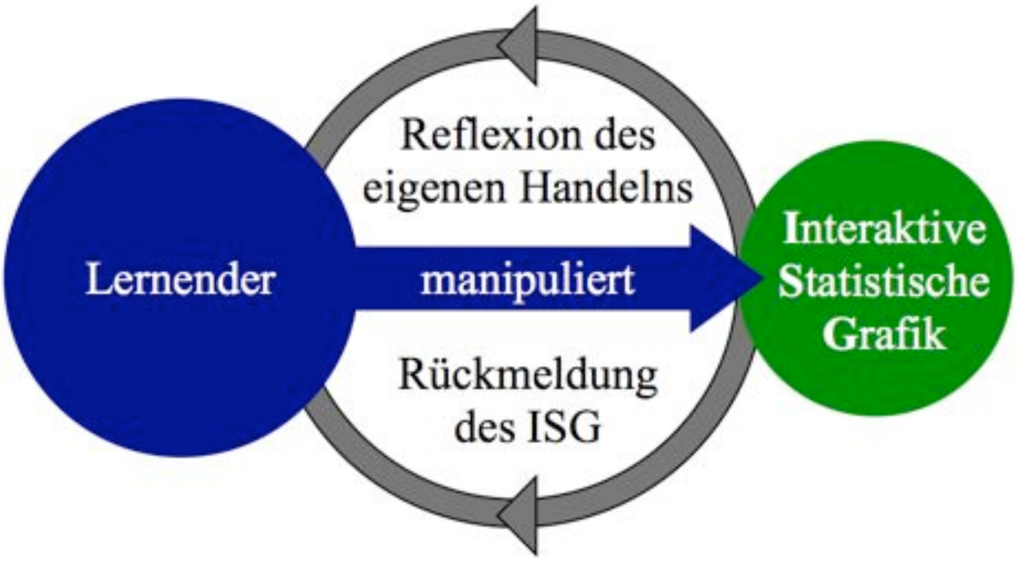


Abbildung 9: Interaktion zwischen Lernendem und interaktiver statistischer Graphik

1. Highlighting

„Highlighting“ bezeichnet eine hervorgehobene Darstellung einer Untermenge der Daten in allen Graphiken. Diese Hervorhebung kann auf verschiedene Art und Weise geschehen. Ein Beispiel sind Dotplots, in denen die Plotsymbole vergrößert sind oder anderenoder Plotsymbole verwendet werden oder durch eine Farbänderung der Symbole gekennzeichnet sind. Da die Kennzeichnug der gehighlighteten Untermenge in allen dargestellten Graphiken möglich sein soll und die Darstellung für den Betrachter möglichst ohne große Erklärungen konsistent sein muss, ist in vielen Programmpaketen das Highlighting durch eine Farbänderung umgesetzt (M. Theus, 1996, S. 42ff)

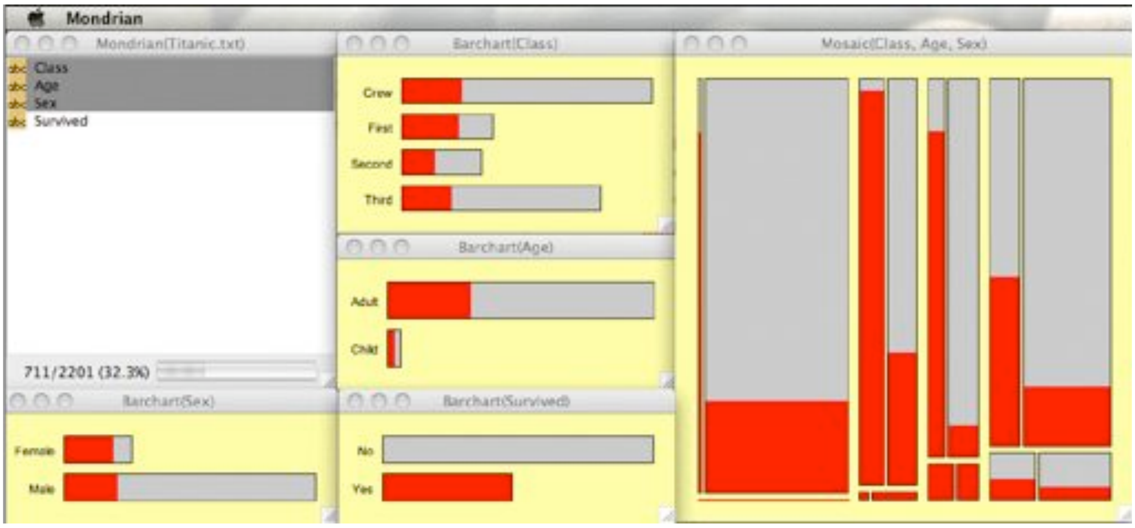


Abbildung 10: Ein Beispiel für Highlighting und Linking – alle Diagramme sind miteinander verbunden: D.h. in allen werden diejenigen Fälle mit der Merkmalsausprägung „Survived” rot markiert

2. Selektion

Selektion bedeutet die Auswahl einer Untermenge von Daten in einer Graphik mit Hilfe verschiedener Werkzeuge, die dann gehighlightet werden. In vielen Programmpaketen stehen folgende Selektionswerkzeuge zur Verfügung:

Lasso: Mit dem Lasso lassen sich beliebige Formen erstellen und damit verschiedene Untermengen einfangen. Start- und Endpunkt werden immer verbunden.

Drag-Box: Die Drag-Box kann Rechtecke in Graphiken selektieren.

Brush: Mit dem Werkzeug Brush (Bürste) kann man rechteckige Bereiche dynamisch highlighten.

Pointer: Der Pointer selektiert einzelne Punkte.

Slicer: Mit dem Slicer (Messer) können achsenparallele Intervalle selektiert werden.

Selektions-Modi: Mit denunterschiedlichen Selektions-Modi ist es möglich die Auswahl verschiedener oder gleicher Werkzeuge zu kombinieren. In den meisten Programmen stehen die Modi „einfach” (es wird genau der gerade ausgewählte Bereich selektiert), „Schnitt” (es wird die Schnittmenge aus mehreren ausgewählten Bereichen selektiert), „Vereinigung” (es wird die Vereinigungsmenge aus mehreren ausgewählten Bereichen selektiert) und „nicht” (es wird die Komplementmenge ausgewählt) (M. Theus ,1996, S. 42ff).

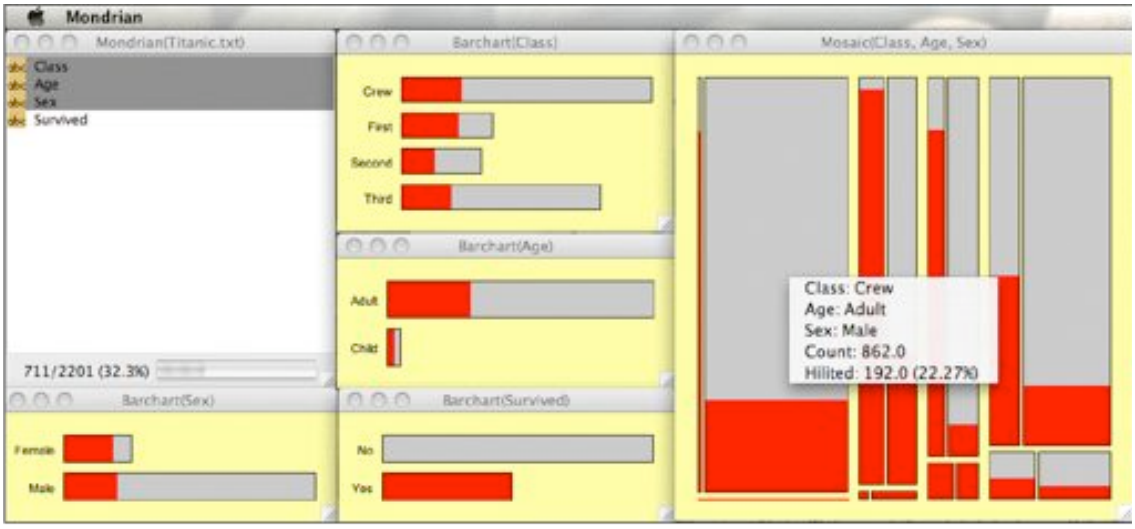


Abbildung 11: Ein Beispiel für Abfragen in statistischen Graphiken – in den verschiedenen Diagrammen sind zu jedem Element die passende Basis oder Metadaten abfragbar.

3. Abfragen

Durch Abfragen kann der Benutzer weiter diferenzierte Informationen zu den Objekten erhalten. Diese werden auf Wunsch des Nutzers mit einfachen Mausklicks zur Verfügung gestellt, ähnlich wie dem Kontextmenüs in allen gängigen Rechnerbetriebssystemen (M. Theus, 1996, S .42ff).

4. Warnungen

Warnungen sind nicht offensichtliche Informationen, die vom System geliefert werden, da der Nutzer (beispielsweise aufgrund der begrenzten Auflösung des Bildschirms) Datenpunkte oder kleine Untermengen des Datensatzes vom System nicht mehr dargestellt bekommt oder nicht mehr wahrnehmen kann. Ähnliche Probleme können bei einer interaktiven dynamischen Skalierung oder Parameterisierung in einer Graphik auftreten (M. Theus, 1996, S. 42ff).

5. Linking

Linking bezeichnet den Sachverhalt, dass alle Graphiken, Tabellen und Modelle miteinander verbunden sind und eine Änderung in einer Darstellung alle anderen damit verbundenen Darstellungen des Datensatzes ändert (M. Theus, 1996, S. 42ff).

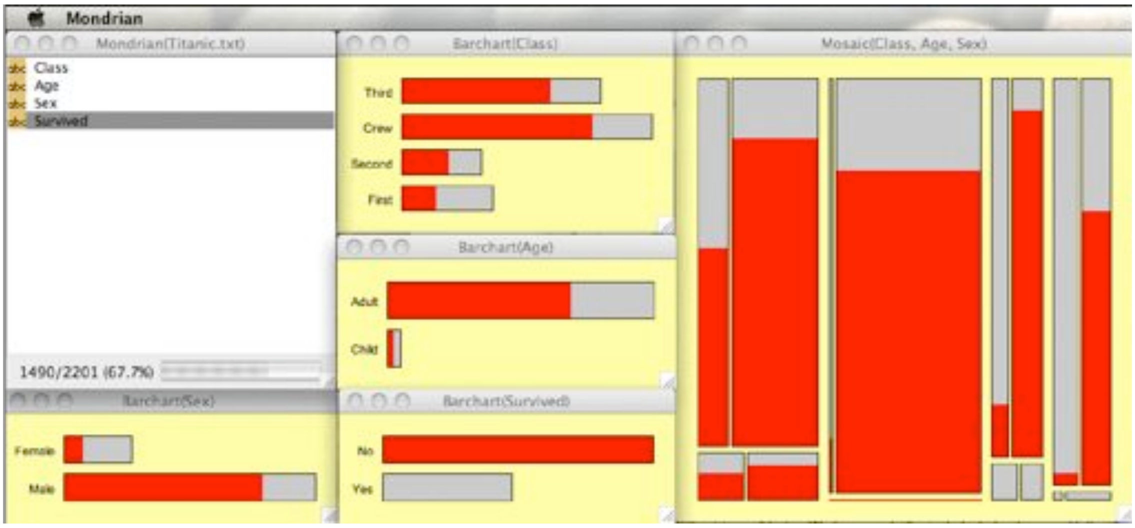


Abbildung 12: Umordnen von Elementen in statistischen Graphiken – hilft die graphische Darstellung der Fragestellung anzupassen.

6. Umordnen von Elementen

Das bedeutet beispielsweise das Vertauschen der x- und y-Achsen im Scatterplot, das Umsortieren der Achsen im Parallel Coordinate Plot, das Ändern der Reihenfolge der Kategorien in Barcharts bzw. der Reihenfolge von Variablen in Mosaikplots. Durch Gruppieren einzelner Kategorien können auch neue Einblicke in die Datenstruktur gewonnen werden.

Zusätzlich zu den einzelnen Plots ist ein weiteres wichtiges Prinzip, dass in den Darstellungen der Daten logisch gezoomt werden kann. Das bedeutet im einfachsten Fall (M. Theus ,1996, S.42ff):

1. Der Nutzer kann jeden einzelnen Datenpunkt abfragen.

2. Alle zu diesem Datenpunkt verfügbaren Informationen werden vom Rechnersystem dargestellt.
3. Es werden, abhängig vom betrachteten Datenausschnitt, mehr oder weniger viele Informationen angezeigt.
4. Mit solchen interaktiven Graphiken ist es möglich, Makro- und Mikrophänomene in den Daten bis hin zu jedem einzelnen Fall so darzustellen, dass Strukturen erkennbar werden, ohne dass man dabei den Gesamtüberblick verliert.

2.8 Vorgehensmodell der Explorativen Datenanalyse

Das Hauptwerkzeug der Explorativen Datenanalyse ist die interaktive statistische Graphik. Im vorigen Abschnitt wurden die grundlegenden Werkzeuge (Highlighting, Selektion, Linking usw.) der interaktiven statistischen Graphik kurz skizziert. Alle daraus entwickelten interaktiven Plots könnte man natürlich auch manuell erstellen, was aber dem Vorgehen der Exploration mit vielen iterativen Arbeitsschritten widerspricht. Bei der Exploration kommt es darauf an, die Daten aus möglichst vielen Blickwinkeln zu betrachten. Dies ist mit einem vernünftigen Zeit- und Arbeitsaufwand nur möglich, wenn man einen geeigneten Rechner und eine intuitive graphische Oberfläche eines statistischen Softwarepaketes verwendet. Da der Rechner das wichtigste Instrument für die Explorative Datenanalyse ist, liegt es nahe, das Erlernen der einzelnen Analyseschritte möglichst nah am „Werkzeug” durchzuführen.
Es liegt in der Natur der Exploration, dass es verschiedene Wege und Methoden gibt, ein und dasselbe Problem zu lösen. Das heißt, verschiedene Datenanalysten bilden im Laufe des Erlernens der Explorativen Datenanalyse individuelle Vorlieben einer Herangehensweise an ein statistisches Problem heraus. Anders als die Wahrscheinlichkeitstheorie und die schließende Statistik haben die deskriptive Statistik und die Explorative Datenanalyse keine fundierte Theorie, mit der sie endgültig definiert werden. Dennoch ist das Vorgehen bei der Datenanalyse nicht beliebig und es können sogar unterschiedliche Fragestellungen, die analog zur schließenden Statistik sind, beantwortet werden.

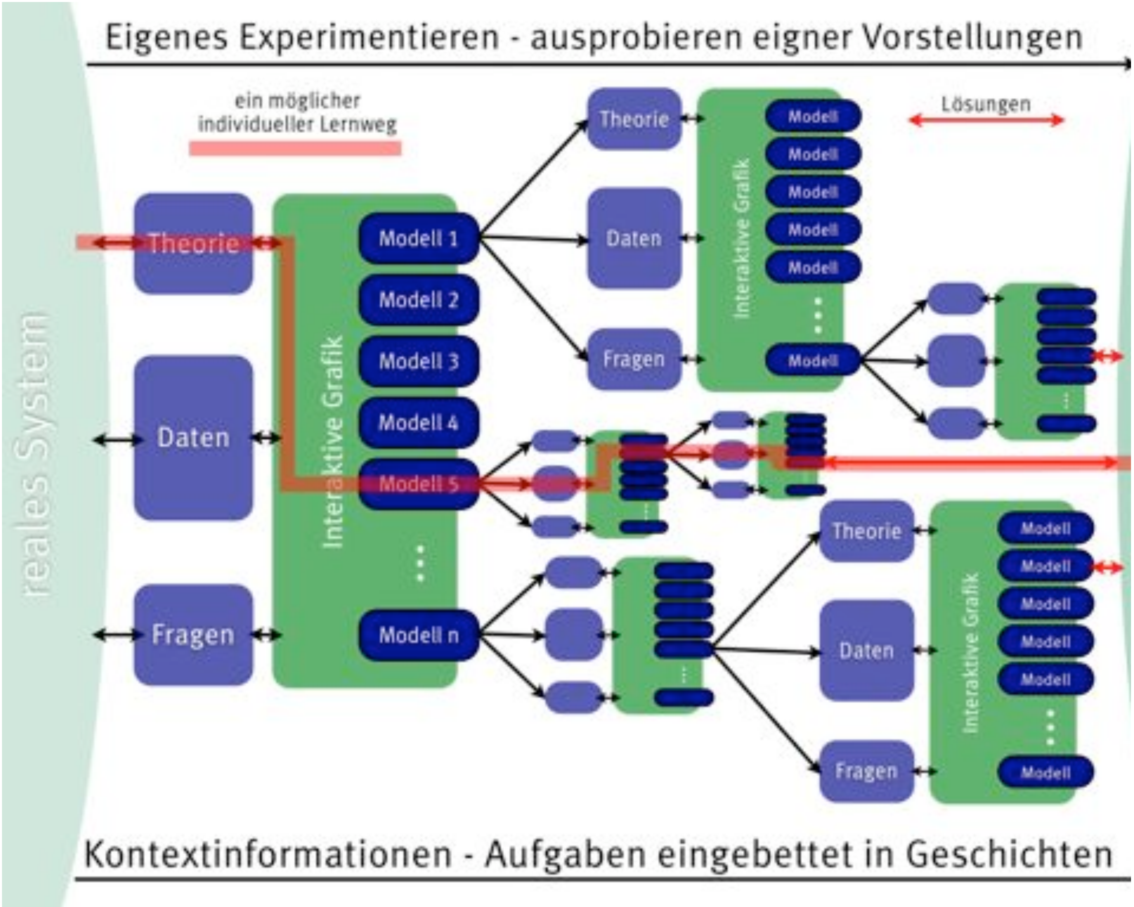


Abbildung 13: Lernwege und Lösungswege in der explorativen Datenanalyse - Die obige Darstellung versinnbildlicht die unterschiedlichen Lern- und Lösungswege bei einer Explorativen Datenanalyse. Der Weg von der Fragestellung bis zur Lösung ist nicht nur problemabhängig, auch Vorerfahrung und Vorlieben des Lernenden spielen eine wichtige Rolle (U. Fahrner & A.Unwin, 2008, S. 163f)

Ein weiterer wichtiger Faktor bei der Exploration ist neben der Aussagekraft, die interaktive Graphiken für geschulte Statistiker besitzen, die Möglichkeit, anhand der aufgearbeiteten Graphiken mit den „Inhabern der Datensätze" die Ergebnisse der Exploration zu diskutieren. Dadurch wird es möglich, zusätzliche Metadaten aus dem jeweiligen Fachbereich, in dem die Datensätze erhoben wurden in die Analyse mit einzubeziehen. Die so gewonnenen sachlogischen Überlegungen können neue wertvolle Aspekte für das weitere Vorgehen der Exploration aufzeigen und helfen bei der Interpretation der Ergebnisse. In einer ersten Näherung besteht eine Datenanalyse aus drei Schritten (U. Fahrner & A.Unwin, 2008 S. 166f):

1. Übersicht verschaffen

Im ersten Schritt der Datenanalyse schafft man sich eine Übersicht über die einzelnen Variablen. Im Fall einer diskreten Variablen verwendet man hierzu einen Barchart. Bei einem Barchart ist die Höhe der Balken proportional zur Anzahl der Fälle der jeweiligen Kategorie. Die Breite der Balken ist für jede Kategorie gleich. Durch das Ordnen der Barcharts nach verschiedenen Kriterien ist es leicht, einen ersten Überblick zu gewinnen. Bei der Untersuchung einer stetigen Variablen ist ein Histogramm hilfreich. Dabei werden aus den stetigen Daten Klassen gebildet und deren Häufigkeit senkrecht zum Klassenintervall angetragen. Dabei ist zu beachten, dass die Struktur eines Histogramms stark von der willkürlichen Wahl der Klassenbreite und des Ursprungs der ersten Klasse abhängt. Deshalb sind beide Parameter (Klassenbreite und Ursprung) in interaktiven statistischen Softwarepaketen direkt veränderbar. Dadurch wird es möglich, verschiedene Struktureigenschaften der Variablen zu untersuchen:

- Treten bestimmte Werte gehäuft auf?
- Gibt es bestimmte Intervalle, die keine Werte enthalten?
- Wie stark sind die Werte diskretisiert?
- Gibt es irgendwelche Symmetrien oder folgt das Histogramm zumindest optisch irgendeiner Verteilungsfunktion?

Die Antworten auf diese Fragen geben Hinweise auf das weitere Vorgehen.

2. Konkretisierung der ersten Eindrücke

Die ersten Eindrücke werden nun konkretisiert und durch weitere Plots verfeinert, die den Daten unter Umständen angemessener sind. Zur Auswahl stehen eine Vielzahl von verschiedenen Plots, von denen hier zwei genannt werden sollen.

1. Wenn die Variablen stark diskretisiert sind, sollte man sich die einzelnen Datenpunkte in diesem Schritt in einem Dotplot darstellen lassen. So können vielleicht einzelne Gruppen unterschieden oder zusammengefasst werden.
2. Gewinnt man durch das Histogramm den Eindruck, dass die einzelnen Datenpunkte sehr dicht liegen, wenige Lücken aufweisen und sich symmetrisch

verteilen könnten, könnte man als Werkzeug z.B. einen Boxplot wählen, der einem einen ersten Eindruck der Lage- und Streuparameter gibt.

Nachdem die ersten Eindrücke durch verschiedene Plots konkretisiert wurden, geht es im nächsten Schritt darum, die Abhängigkeitsstrukturen zwischen einzelnen Variablen zu untersuchen.

3. Erkennen von Abhängigkeiten

Um Abhängigkeiten zu untersuchen, stehen gelinkte Barcharts, Spinplots und Mosaikplots als naheliegendste Vertreter bei diskreten Daten und Scatterplots sowie gelinkte Histogramme bei stetigen Variablen zur Verfügung. Auch beliebige Kombinationen zwischen den Plots sind möglich. Wenn man z.B. die Abhängigkeit einer stetigen Variablen von einer diskreten untersuchen möchte, bietet es sich an, die diskrete Variable in einem Barchart darzustellen und die stetige in einem Histogramm. Nun kann man durch Selektieren eine Variable im Barchart highlighten, dazu zeigen die gehighlighteten Flächen im Histogramm die Abhängigkeit von der diskreten Variable. Beim Vergleich von zwei diskreten Variablen würde sich zum Beispiel die Kombination von einem Barchart für die erste Variable und einem Spinplot für die zweite Variable eignen. Das Vorgehen ist ähnlich dem stetigen Fall. Hier wird eine Kategorie im Barchart gehighlightet und dazu die gehighlighteten Anteile im Spinplot der Variable zwei verglichen.

Aus diesem Vorgehen heraus entwickelte sich die Idee, Mosaikplots zu erstellen, denn Mosaikplots sind die mehrdimensionale Verallgemeinerung von Spinplots. Der Mosaikplot wird durch das abwechselnde vertikale und horizontale Unterteilen des Plots konstruiert, entsprechend den Anteilen der hinzugefügten Variablen. Die Entwicklung des parallelen Koordinatenplots ermöglicht es auch, in einem Plot Abhängigkeiten von diskreten und stetigen Variablen abzulesen. An dieser Art Daten zu plotten wird deutlich, dass hier nur mit einer schnellen interaktiven Darstellung auf einem Rechner sinnvoll gearbeitet werden kann und dass es von entscheidender Bedeutung ist, dass die Software intuitiv und direkt bedient werden kann (M.Theus, 2008, S. 69). Zusammenhänge werden bei dieser Art von Plot nur dann deutlich, wenn man die Daten schnell umordnen, sortieren, skalieren und highlighten kann.

Durch die Kombination einzelner Plots erhält man schrittweise Erklärungsmuster, wenn sachlogische Überlegungen aus dem Wissensgebiet einbezogen werden, aus dem die Daten stammen (M.Theus, 2008, S. 49). Diese Erklärungsmuster verdichten sich Schritt für Schritt, bis schließlich erste Ansatzpunkte für Hypothesen sichtbar werden.

Diese Hypothesen können nun mit der schließenden Statistik bestätigt werden. Dabei stehen die ganze Vielfalt und der mathematische Reichtum der Statistik zur Verfügung. Bei der Exploration geht man aber noch weiter: Man betrachtet die Residuen, die sich aus den statistischen Modellen ergeben, und untersucht diese auf ihre Strukturen. Lassen sich Strukturen finden, so ergeben sich Ansatzpunkte für neue Hypothesen und verbesserte statistische Modelle. Lassen sich keine Strukturen finden, bedeutet dies, dass alle Strukturen, die im Datensatz liegen, erkannt wurden.

Je nach Qualität und Aussagekraft des Datensatzes kann es dazu kommen, dass der oben beschriebene Erkenntnisprozess schon vorher abbricht, d.h. dass der Informationsgehalt des Datensatzes nicht ausreicht, um schließende statistische Methoden anzuwenden. Bei vielen Fragestellungen, in denen es zunächst darum geht, Wirkungszusammenhänge in einer Stichprobe (z.B. bei psychologischen, aber auch bei physikalischen Experimenten) herauszufinden, ist es durchaus ausreichend, aus den Wirkungszusammenhängen eine Erklärungshypothese zu generieren und diese mit theoretischen Überlegungen zu untermauern. Dieses Vorgehen entspricht dem klassischen Vorgehen beim physikalischen Experiment. Dabei wird aus den in der Laborsituation erzeugten Daten (Stichprobe) eine These aufgestellt, diese in das vorhandene theoretische Gerüst physikalischer Erkenntnisse eingebunden, um dann in einem größeren Zusammenhang bestätigt zu werden.

Das Modell der EDA, die Tukey (1962) vorlegte, ist durch die Entwicklung moderner Rechner und graphischer Software durchführbar geworden. Für diese Arbeitsweise sind schnelle, flexible Werkzeuge unerlässlich. Neuskalierungen, Einzoomen, Sortierungen, Formatierungen, Abfragen und – am allerwichtigsten – Linking müssen unmittelbar geschehen (Unwin, 1999). Standardgraphiken wie Histogramme, Säulendiagramme, Boxplots und Streudiagramme sind durch das Hinzufügen interaktiver Fähigkeiten und durch ihre Implementation in graphische Software zu leistungsfähigen Werkzeugen geworden. Andere Graphiken, wie Mosaikplots für multivariate kategorielle Daten und Parallelkoordinaten-Plots für stetige Daten, sind durch Interaktivität erst in der Praxis anwendbar geworden (M. Theus, 2008, S. 11). Es muss möglich sein, viele Fenster schnell zu zeichnen, sie übersichtlich zu verwalten und sofort aktualisieren zu können. Wenn tiefer gehende Analysen, weiter führende Untersuchungen oder Erkenntnis bringende Darstellungsänderungen notwendig sind, soll das mit direkten graphischen Zugriffen zügig durchgeführt werden. In der klassischen Statistik bieten Kommandozeilen-Oberflächen (d.h. der Nutzer muss jeden einzelnen Befehl wissen und eintippen, um ihn ausführen zu lassen. Es gibt keinerlei grafisch unterstützende Oberfläche) einen sinnvollen, strukturierten und mächtigen Zugang. Dieser Ansatz ist von den Konzepten der Explorativen Datenanalyse zu weit entfernt. Deshalb werden graphische Benutzeroberflächen benötigt, welche ein intuitives Arbeiten ermöglichen.

Graphische Benutzeroberflächen haben Nachteile, wenn es darum geht, die gemachten Analyseschritte zu dokumentieren bzw. eine große Anzahl von Einstellungsmöglichkeiten übersichtlich darzustellen. Diese Nachteile werden in dem Bereich aufgewogen, in dem es darum geht, intuitive neue Informationen zu entdecken und Ansatzpunkte für Hypothesen zu finden. Der Mensch denkt nicht in Kommandozeilen, sondern in Bildern. Erst wenn Ansatzpunkte für Hypothesen gefunden sind, wird es notwendig, diese mit induktiven statistischen Verfahren zu belegen und die dafür geeigneten Kommandozeilenwerkzeuge (z.B. der Statistiksoftware R) zu benutzen.

Rechnerkapazitäten verbessern sich ständig, trotzdem werden sie nie allen Ansprüchen gerecht. Zur Zeit können Datensätze mit einer Million Fälle mit interaktiven graphischen Methoden praktisch analysiert werden (Unwin et al., 2006). Für größere Datensätze werden andere Strategien gebraucht, z.B. eine Aufteilung in Untergruppen und die Verwendung von Stichproben. In dieser Hinsicht unterscheidet sich EDA von der klassischen Statistik kaum. Es ist schwer, die interaktive Vorgehensweisen in gedruckter Form wiederzugeben. Die Anzahl der betrachteten Graphiken, die ausprobierten Variationen, die Abfragen und die vielen verschiedenen Linkings tragen zum besseren Verständnis der Daten und zur Aufdeckung neuer Information bei.

2.9 Ein Beispiel für das Arbeiten mit der Explorativen Datenanalyse

„Grau, teurer Freund, ist alle Theorie,
und grün des Lebens goldner Baum."

Johann Wolfgang von Goethe

Nachfolgendes Beispiel haben Antony Unwin und der Verfasser für einen Artikel mit dem Titel „Design-Based-Research – Explorative Datenanalyse als Analysemethode in realen Lehr-Lern-Systemen" in dem Buch „Der Nutzen wird vertagt ..." ausgearbeitet. (Reinmann & Kahlert 2008, S. 159-162)

Kennzeichnend für interaktive Analysen mit EDA ist die Schnelligkeit und Flexibilität. Zur Erläuterung wird ein Datensatz betrachtet, der aus dem vertraulichen Bericht über die Mortalität der Soldaten im Krim-Krieg von Florence Nightingale an die britische Regierung aus dem Jahre 1858 stammt (www.florence-nightingale-avenging-angel.co.uk). Dieser Datensatz außerhalb des Lehr-Lern-Kontextes wurde gewählt, um zu zeigen, wie man Daten aus einer realen Situation ohne festgelegten Versuchsplan auswerten kann. Außerdem wird durch die multivariate Ansicht dieses Datensatzes deutlich, wie durch eine solche Betrachtungsweise des Datensatzes Fragen zu dessen Qualität untersucht werden können. Dieser Datensatz und die weiteren Erhebungen von Florence Nightingale hatten Einfluss auf die damalige Politik und auf die gesellschaftliche Entwicklungen und zeigt somit auch die Kraft, die empirischen Untersuchungen innewohnt.

Ein Auszug dieses Datensatzes ist in Abbildung 14 wiedergegeben. Insgesamt sind die Daten von 63 Corps aufgeführt, welche über sieben Monate vom Oktober 1854 bis April 1855 auf der Krim stationiert gewesen sind.

Division and Corps		Average Strength	Total: Admitted into Hospital	Died in the Crimea, &c.	Sent to Scutari, &c.	Died at Scutari, &c.	Invalided to England, &c.	Remarks
2nd Division	30th Foot	522	934	108	308	93	99	
	55th "	695	1,462	61	265	96	100	
	62nd "	430	949	96	135	42	24	
	95th "	417	1,250	199	345	155s	114	
	41st "	684	1,323	104	320	94	81	
	47th "	637	1,223	91	280	71	102	
	49th "	655	1,071	66	274	90	89	
3rd Division	1st "	771	1,048	229	354	118	63	
	14th "	423	878	8	42	2	6	a)
	38th "	689	1,728	149	319	118	73	

Abbildung 14: Auszug aus Tabelle B aus dem Bericht von Florence Nightingale über die Mortalität im Krim-Krieg[2]

Abbildung 15 zeigt ein Histogramm der Stärken der Corps und ein Säulendiagramm für deren Typ. Es ist charakteristisch für die interaktive Graphik, dass mehr als eine Graphik betrachtet und dass die Graphiken gelinkt (d.h. die Graphiken sind miteinander verbunden) werden.

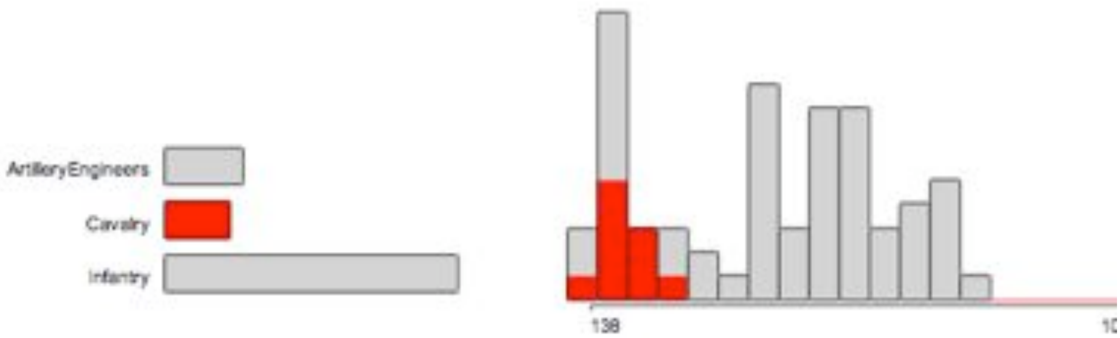


Abbildung 15: Durchschnittliche Stärke des Corps und Corpstyp[3]

Die Kavallerie-Corps sind selektiert worden und es wird dadurch deutlich, dass diese Kavallerie-Corps alle kleiner waren. Man hätte auch in die andere Richtung linken kön-

2 Die Spalten sind: Corps, durchschnittliche Stärke der Corps, in das Krankenhaus eingelieferte Soldaten, auf der Krim gestorbene Soldaten, nach Scutari verschickte Soldaten, in Scutari gestorbene Soldaten, nach England als Invaliden heimgekehrte Soldaten

3 Die Kavalleriecorps sind selektiert worden und rot gekennzeichnet. Die rote Linie rechts ist eine Warnung, dass nicht alle Datensätze hier abgebildet wurden.

nen, d.h. z.B. die höheren Werte im Histogramm selektieren, um zu sehen, von welchem Corps-Typ sie stammen.

Abbildung 16: Corps-Sterbeziffern auf der Krim (Histogramm) und ein Auszug aus dem Säulendiagramm der Corps-Stärken.[4]

Die Sterbeziffern der Soldaten innerhalb der Corps in der Krim selbst sind in der Abbildung 16 dargestellt. Der Extremwert wurde selektiert und im begleitenden Säulendiagramm als das 63. Foot Corp erkannt. Es wird nur ein Ausschnitt des Säulendiagramms gezeigt; das liegt daran, dass interaktiv eingezoomt (d.h. die Darstellung zeigt detailliertere Informationen) wurde, um sich auf den relevanten Teil konzentrieren zu können. Falls mehrere Corps selektiert worden wären, hätte man anschließend interaktiv sortieren können, um die interessanten Fälle zusammenzustellen. Das Hauptziel von Florence Nightingale war, der Regierung die hohen Sterbeziffern im Krankenhaus in Scutari vor Augen zu führen.

Abbildung 17 stellt die Sterbeziffern im Scutari Krankenhaus und auf der Krim dar. Man erkennt, dass in allen Corps (außer dem schon erwähnten 63. Foot Corp) die Sterbeziffern in Scutari höher waren. Die Achsen sind gleich skaliert worden, um direkt vergleichen zu können. Zwei Corps hatten besonders hohe Sterbeziffern in Scutari. Die Abfrage zeigt, dass das Grenadier Guards Corp am härtesten getroffen wurde. Nach der Bewegung des Cursors auf das nächste Corps entdeckt man, dass dies das 93rd Foot Corp war. Die Möglichkeit, jedes statistische Objekt in einer Graphik direkt abfragen zu können, ermöglicht eine sehr schnelle Zuordnung zwischen Diagramm und Daten. In

---

[4] Die rote Linie rechts ist eine Warnung, dass nicht alle Datensätze hier abgebildet wurden.

diesem Fall wurde eine erweiterte Abfrage verwendet, um Information über Variablen sichtbar zu machen, die nicht in der Graphik selbst erscheinen. Man hätte sich auch Werte für mehrere Variablen anzeigen lassen können.

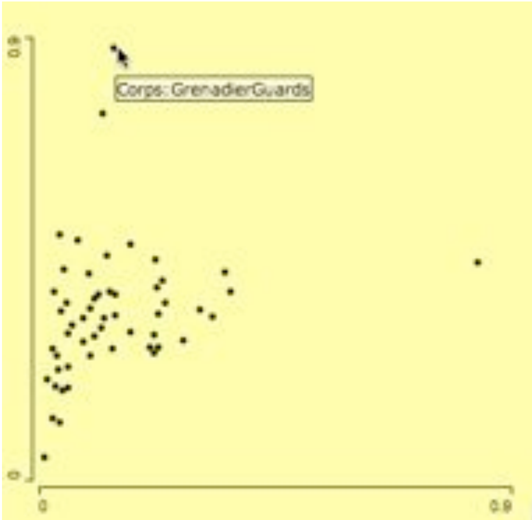


Abbildung 17: Sterbeziffern der Corps-Soldaten in Scutari und auf der Krim[5]

Es ist leicht, die bis hierhin berichteten Analysen ohne Fragen zu akzeptieren. Aber es kommen Zweifel an der Datenqualität auf. Wie hätte man z. B. die „durchschnittliche" Stärke eines Corps genau messen können? Sind die Statistiken unter den Umständen des Kriegs wirklich genau gesammelt worden? Wie sind die Todesfälle klassifiziert worden, die sich auf der Schifffahrt zwischen der Krimhalbinsel und Scutari ereignet haben (auf einer berüchtigten Fahrt sind 47 von 130 Patienten gestorben)? Um die Resultate richtig interpretieren zu können, ist die Hilfe entsprechender Experten mit dem nötigen qualitativen Hintergrundwissen notwendig. Auch hier kann die interaktive Graphik hilfreich sein, um Kommunikation zwischen Datenanalysten und Historiker zu unterstützen. Dass dies tatsächlich funktioniert, hat sich in den letzten Jahren in unserer Arbeit immer bestätigt (A. Unwin, 2006)

[5] Das Corps mit der höchsten Sterbeziffer in Scutari wurde abgefragt.

2.10 Strategie der Problemlösung in der Datenanalyse

Das Verfahren der Explorativen Datenanalyse liegt mit seinem wichtigsten Werkzeug, der interaktiven Graphik, an der Nahtstelle zwischen der realen Fragestellung und der formalen Sprache der Mathematik und deren abstrakten Modelle. Mit Hilfe der Explorativen Datenanalyse wird es bei vielen realen Problemen erst möglich, tief greifende mathematische Erkenntnisse in der Praxis anzuwenden und deren Aussagen für reale Fragestellungen zu nutzen. Der Kernleitsatz der Explorativen Datenanalyse sagt aus, dass man die Daten aus verschiedenen Perspektiven „multivariat" ansieht und darin Strukturen erkennt (M. Theus, 2003, S. 9ff). Der Gesamtvorgang, der als Explorative Datenanalyse bezeichnet wird, teilt sich dabei in folgende wesentliche Schritte auf:

1. Übersicht über die Daten gewinnen

Durch die Verwendung verschiedener interaktiver Graphiken und das Darstellen der Daten in verschiedenen Plots gewinnt man eine erste Übersicht über die Daten. Die Vernetzung der Plots untereinander und die Einbeziehung der Metadaten des Datensatzes zusammen mit Expertenwissen aus dem jeweiligen Wissensgebiet führen zu qualitativen Informationen über den Datensatz. Durch diese Kombination ist es oft möglich, ein Gesamtproblem in mehrere Teilprobleme zu zerlegen. Durch das explizite Verwenden von Expertenwissen können mitunter in diesem Schritt Modelle gefunden werden, die sich relativ leicht sachlogisch begründen und interpretieren lassen.

2. Erstellung von Arbeitsmodellen

In diesem Schritt werden die erkannten Strukturen aus dem vorhergehenden Schritt dazu verwendet, statistische Modelle auszuwählen und unter Einbeziehung der Daten eine konkrete Modellrealisation zu erstellen. Außerdem lassen sich hier oft schon grobe Abschätzungen des Gültigkeitsbereichs und Schranken für Fehlertoleranzen der Modelle angeben. In diesem Schritt sollte es in den meisten Fällen möglich sein, die Komplexität des Gesamtsystems und der einzelnen Subsysteme abzuschätzen bzw. einzelne bearbeitbare Subsysteme vom Gesamtsystem abzuspalten. Man geht je nach Komplexität des Problems parallel vor, aggregiert mehrere Modellvorstellungen und kombiniert diese mit sachlogischen Überlegungen aus dem jeweiligen Wissensgebiet.

3. Übersicht über das Residuum der Arbeitsmodelle gewinnen

In diesem Schritt werden die Residuen der Arbeitsmodelle der Subsysteme mit den Werkzeugen der interaktiven statistischen Graphik betrachtet und Hinweise für die Verbesserung der Modelle gesucht. Die Modelle der Subsysteme werden dann zu einem Gesamtarbeitsmodell aggregiert und Interpretationen für das Gesamtmodell gesucht.

4. Induktive Statistik anwenden

Die gefundenen konkreten Modellrealisationen werden nun mit Hilfe aller Möglichkeiten und Erkenntnisse der induktiven Statistik getestet. Erkenntnisse aus Residuen und aus dem Vergleich verschiedener Arbeitsmodelle führen zu einer verbesserten Modellvorstellung. Diese fließt wieder in Arbeitsmodelle und die Modelle der Subsysteme ein.

5. Schritte wiederholen

Die oben genannten Schritte können sich viele Male in unterschiedlichster Reihenfolge wiederholen, bis schließlich eine interpretierbare Lösung gefunden ist, die zum Verstehen des realen Problems beiträgt. Dieses parallele und vielschichtige Vorgehen ist notwendig, um die multivariaten Strukturen in Datensätzen aufzuspüren. Der Begriff der Explorativen Datenanalyse beschreibt also eine Strategie zur Problemlösung in einem realen Sachverhalt.

3. Wahrnehmung, Lernen und Gedächtnis

Das dritte Kapitel dieser Arbeit gliedert sich in folgende Teile:

3.1 Das Likelihood-Prinzip der Wahrnehmung

„Genie ist in Wahrheit kaum mehr als die Fähigkeit,
auf ungewöhnliche Weise wahrzunehmen."
William James

Wie die Beispiele des vorigen Abschnitts zeigen, können geschickte graphische Darstellungen erstaunliche Beiträge zur Problemlösung leisten. Es gibt aber auch graphische Darstellungen, welche die visuelle Wahrnehmung des Menschen in die Irre führen und nicht zur Aufklärung von Sachverhalten führen, sondern im Gegenteil diese verschleiern. Es drängen sich Fragen nach den Fähigkeiten und Grenzen unseres visuellen Wahrnehmungssystems auf, denn nur eine genaue Kenntnis der Rahmenbedingungen der visuellen Wahrnehmung bietet die Chance, Graphiken zur Problemlösung nutzbringend einzusetzen. Durch viele Experimente sind Rahmenbedingungen der visuellen Wahrnehmung und Informationsverarbeitung geklärt und es gibt Hinweise darauf, wie die visuelle Informationsverarbeitung grundsätzlich funktionieren könnte. Bis heute gibt es aber keine geschlossene Theorie, die es ermöglicht, alle einzeln experimentell beobachteten Phänomene einzuordnen.

Abbildung 18 Das Liklelihood-Prinzip der Wahrnehmung: Unbewusste Schlüsse führen zur Lösung mit der größten subjektiven Wahrscheinlichkeit (Goldstein, E. B. (2008): , S. 107 - 130) Das Bild vom Dalmatiner gilt in der Psychologie als ein Beispiel für die ganzheitliche Erfassung der Gestalt in der Wahrnehmung; es werden nicht zuerst die einzelnen Gliedmaßen erkannt, und dann zum Hund zusammengesetzt, sondern das Bild emergiert aus den Punkten..

Neuere Arbeiten gehen davon aus, dass es einen engen Zusammenhang zwischen der Objektwahrnehmung und der Steuerung von Handlungen gibt (Goldstein, 2008, S. 130).

Man geht davon aus, dass über die motorische Interaktion mit Objekten, die Fähigkeiten der Objektwahrnehmung stark beeinflusst wird (Helbig, 2006, S. 221-228). Den bisherigen Wissensstand kann man als „Likelihood-Prinzip der Wahrnehmung" (Goldstein, 2008, S. 127) zusammenfassen: „Unbewusste Schlüsse führen zur Lösung mit der größten subjektiven Wahrscheinlichkeit." Die Urform des Likelihood-Prinzips der Wahrnehmung wurde von Hermann Helmholtz Anfang des 20. Jahrhunderts aufgestellt, um das Problem der Mehrdeutigkeit visueller Stimuli zu umgehen. Die Grundidee von Helmholtz war, den Wahrnehmungsprozess als Problemlöseprozess aufzufassen (Goldstein, 2008, S. 127). Die Fragestellung für das Individuum besteht nun darin herauszufinden, welches Objekt ein bestimmtes Reizmuster verursacht hat. Da das Individuum in seiner Umwelt neben den visuellen Reizen die Möglichkeit hat, mit den es umgebenden Objekten zu interagieren, entwickelt sich im Laufe des Lebens ein Wissen über die „naheliegendste Antwort" für das Problem. Unter dieses Prinzip lassen sich auch die „Gestaltgesetze" oder besser „Gestaltprinzipien" subsumieren. Diese sind mit vielen Experimenten herausgearbeitet worden und geben einen guten Anhaltspunkt, wie die visuelle Wahrnehmung arbeitet und wo deren Grenzen und Möglichkeiten sind. Dabei ist nicht zu vergessen, dass diese Gestaltprinzipien auch nur den Charakter von Heuristiken haben und für sich nicht beanspruchen können, eine „Erklärung" für das visuellen Objektwahrnehmungssystem des Menschen zu sein. Diese „Gestaltprinzipen" entstehen auch durch einen individuellen Konstruktionsprozess, d.h., dass sich die Wahrnehmungsprinzipen an neue Gegebenheiten anpassen können. Durch die Verknüpfung der visuellen Wahrnehmung und die Möglichkeit motorischer Interaktion entstehen biologischen Grenzen, stabile, für jeden Menschen individuelle Wahrnehmungsheuristiken, die unbewusst ablaufen und dafür sorgen, dass man einen schnellen und groben Überblick über eine komplexe visuelle Situation bekommt. Folgende Gestaltprinzipien geben einen groben Rahmen für die visuelle Wahrnehmung (Goldstein, 2008, S. 107-111):

Prinzip der Einfachheit

Beschreibung:

Visuelle Strukturen werden meist so verknüpft, dass das resultierende Muster so einfach wie möglich aus den Erfahrungen des Individuums erklärt werden kann. Darin liegt das Hauptproblem, denn das Erkennen von neuen Mustern hängt stark von gemachten Erfahrungen ab und kann den Bereich möglicher Erklärungen von Anfang an einschränken, weil die Wirklichkeit nur noch verzerrt wahrgenommen wird.

Der Nutzen für die interaktive statistische Graphik:

Erfahrungen von schon gelösten Problemen helfen beim Lösen unbekannter Fragestellungen.

Mögliche Fehlschlüsse sind dabei:

1. Bei 3d-Darstellungen auf einem 2d Monitor kann es durch Überlagerungen zu Fehleinschätzungen kommen.
2. Verinnerlichte Strukturen, die nichts mit dem Problemkontext zu tun haben, können als Lösungen herangezogen werden und führen somit zu Fehleinschätzungen.

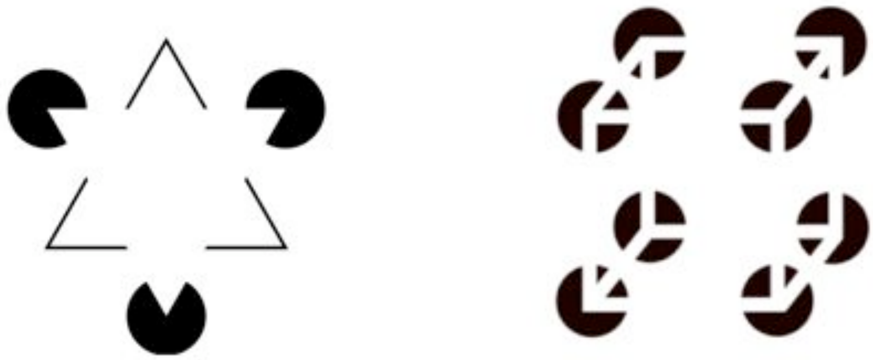


Abbildung 19: Das Liklelihood-Prinzip der Wahrnehmung: Unbewusste Schlüsse führen zur Lösung mit der größten subjektiven Wahrscheinlichkeit (Goldstein, E. B. (2008): , S. 107 - 130). Ein Beispiel für das Prinzip der Einfachheit; Links: z.B. ein weißes Dreieck liegt über den schwarzen Kreisen und der Dreieckskontur; Rechts: z.B. ein weißer Würfel liegt wird durch acht Löcher in der Bildebene betrachtet

Prinzip der Ähnlichkeit

Beschreibung:

Objekte, die gemeinsame Merkmalsausprägungen besitzen, werden zu Gruppen verbunden und, wenn möglich, einer Struktur zugeordnet. Bei diesem Prinzip entsteht ein Problem, wenn visuelle Strukturen mehrere ähnliche Merkmale haben. D.h., eine Gruppe könnte mit dem Merkmal „Farbe” gebildet werden, eine mit dem Merkmal „Größe”, eine mit dem Merkmal „Form” oder „Orientierung”. Folglich wird diese Gruppenbildung entweder dem Problem nicht unbedingt angemessen sein oder aber für verschiedene Individuen eindeutig sein.

Der Nutzen für die interaktive statistische Graphik: Cluster mit gemeinsamen Merkmalsausprägungen gebildet leicht können werden, wenn die Merkmale als Muster dargestellt werden.

Mögliche Fehlschlüsse:

1. Gruppenzuordnungen müssen nicht eindeutig sein
2. Gruppen, die für das Individuum nicht auffällige Merkmale tragen, werden übersehen.

Prinzip der Fortsetzung

Beschreibung:

Objekte oder Punkte, die sich stetig zu einer Linie verbinden lassen, werden als Einheit wahrgenommen. Linien bzw. Grafen, die im IR^3 differenzierbar sind, werden als wahrscheinlichere Struktur erkannt, als nur stetige oder unstetige.

Der Nutzen für die interaktive statistische Graphik:

Visuell lassen sich leicht Ausgleichsgeraden oder Ansätze für Glättungen finden und Tendenzen erkennen.

Mögliche Fehlschlüsse:

1. Bei 3d-Darstellungen auf einem 2d-Monitor können Linien durch Überlagerungen als zusammengehörig gesehen werden, die in der Tiefendimension gegeneinander versetzt sind.
2. Auch unstetige Strukturen können in einem speziellen Problem eine Lösung sein.

Prinzip der Nähe

Beschreibung:

Objekte, die einen kleineren Abstand zueinander haben, werden als zusammengehörig betrachtet. Wenn mehrere Merkmale vorliegen, die für eine Gruppierung nach dem Prinzip der Ähnlichkeit sprechen, ist es eine Frage der individuellen Präferenz, wie die Objekte gruppiert werden.

Nutzen für die interaktive statistische Graphik:

Strukturen in Scatterplots können leicht entdeckt werden. Bei Mosaikplots können die Abstände zwischen den Farbflächen als Strukturierungshilfe genutzt werden.

Mögliche Fehlschlüsse:

1. Bei Fragestellungen, die mit Regressionen zu tun hat, ist es für den Betrachter kaum möglich, die Summe der Abstände von zwei Punktwolken zu vergleichen.
2. Das enge Zusammenliegen von Objekten kann auch nur von der Skalierung des Plots herrühren und nicht von einem inhaltlichen Zusammenhang.

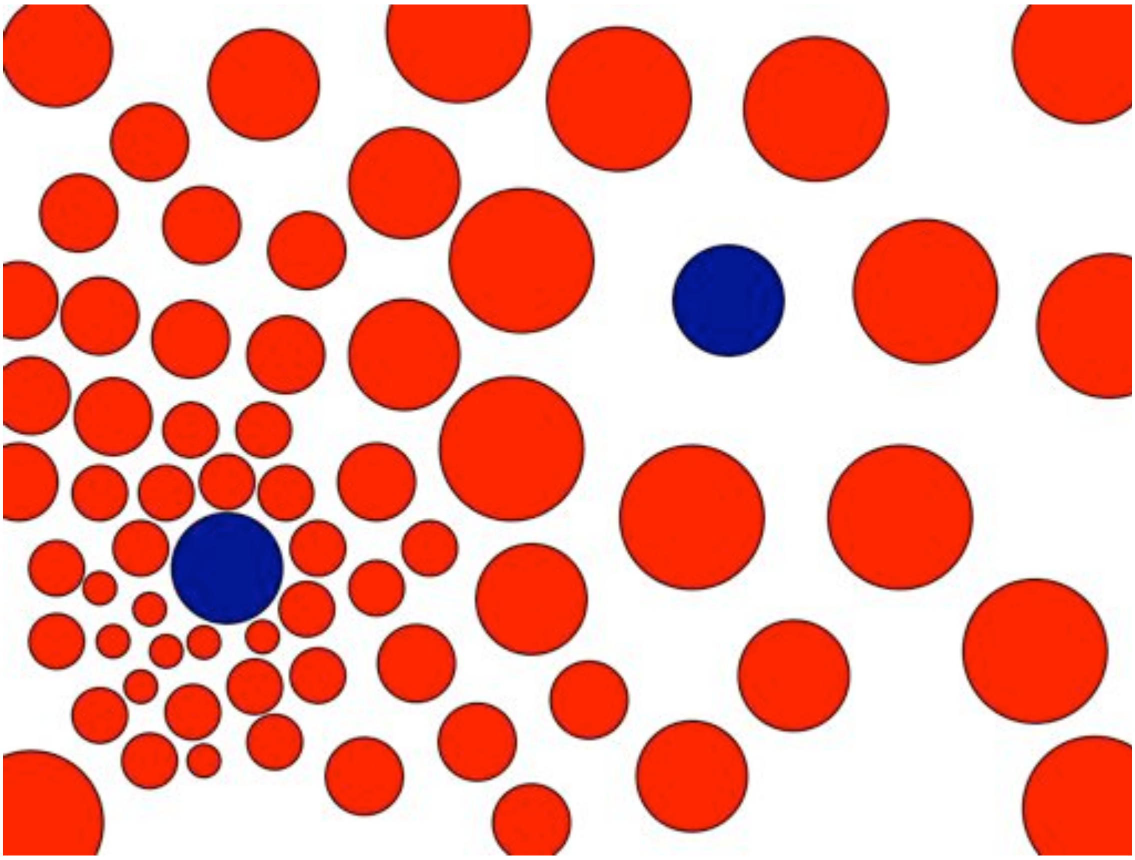


Abbildung 21: Ein Beispiel für das Prinzip der Nähe (Goldstein, E. B. (2008): , S. 205 - 213)

Prinzip des gemeinsamen Schicksals

Beschreibung:

Objekte, die sich entlang einer gemeinsamen Graphenschar bewegen, erscheinen zusammengehörig. Wenn man z.B. eine Autobahn von oben betrachtet, werden die Fahrspuren als eine Einheit wahrgenommen. Ein anderes Beispiel dafür sind zwei sich durchdringende Fischschwärme, bei denen der Beobachter aufgrund der Bewegung die Individuen dem jeweiligen Schwarm zuordnet.

Nutzen für die interaktive statistische Graphik:

Werden in der interaktiven statistischen Graphik zwei Plots mittels Linking verbunden und Teile davon periodisch gehiglightet, so ergibt sich eine Pseudobewegung, z.B. in einem Scatterplot, durch welche Gruppen unterschieden werden können. Dabei hilft das Prinzip des gemeinsamen Schicksals.

Möglicher Fehlschluss:

Die gemeinsame Bewegung bedeutet noch keinen kausalen Zusammenhang

Prinzip der Bedeutung

Beschreibung:

Objekte werden zu Gruppen zusammengefasst, wenn die Gruppe vertraut erscheint oder ihnen eine konstruierte Bedeutung zugeordnet werden kann, die nicht notwendiger Weise kausal aus dem Kontext begründet werden kann.

Nutzen für die interaktive statistische Graphik:

Geometrische Strukturen wie Geraden, Rechtecke, Kreise können als innere Struktur in Plots erkannt werden.

Abbildung 20: Ein Beispiel für das Prinzip der Bedeutung (Goldstein, E. B. (2008): , S. 205 - 213)

Mögliche Fehlschlüsse:

1. Die interpretierte Bedeutung ist in starkem Maße von der eigenen Erfahrungswelt abhängig, die nicht zwingend nützlich für das zu lösende Problem ist.
2. Die durch die Zuordnung von geometrischen Figuren, die in einem Plot zu sehen sind, entstehenden Vorurteile über die Zusammenhänge bei einer konkreten Fragestellung, sind nur sehr schwer zu relativieren.

Abbildung 22: Beispiel für das Prinzip der Einfachheit, bei dem diese Heuristik nicht funktioniert. Tiefenwahrnehmung und Skalierung: Die Person in dieser Montage ist immer gleich groß, die unterschiedliche Größeneinschätzung kommt von der Erfahrung des Sehens in einem wirklichen 3d-Raum (Goldstein, E. B. (2008): , S. 205 - 213)

Prinzip der Gemeinsamen Region

Beschreibung:

Objekte, die durch zusätzliche visuelle Strukturen abgegrenzt sind, werden zu einer Gruppe verbunden.

Nutzen für die interaktive statistische Graphik:

Grundstrukturen von Plots (z.B. Koordinatenachsen von parallel Koordinatenplots) werden als ordnendes Element erkannt und akzeptiert.

Möglicher Fehlschluss:

Feste Ordnungsstrukturen brauchen dem Problem nicht angemessen zu sein, so dass Zugänge zur Problemlösung verschleiert werden.

Prinzip der Verbundenheit der Elemente

Beschreibung:

Objekte, die durch zusätzliche visuelle Strukturen verbunden sind, werden als Einheit wahrgenommen.

Nutzen für die interaktive statistische Graphik:

Farben und Markierungen können Zusatzinformationen transportieren.

Möglicher Fehlschluss:

Bei eine Überfrachtung können die Informationen nicht mehr richtig eingeordnet werden.

Prinzip der Synchronität

Beschreibung:

Visuelle Ereignisse, die zur selben Zeit auftreten, werden als zusammengehörig betrachtet. Visuelle Ereignisse, die zeitlich versetzt auftreten, werden als nicht zusammengehörig betrachtet.

Nutzen für die interaktive statistische Graphik:

Werden in der interaktiven statistischen Graphik zwei Plots mittels Linking verbunden und Teile davon periodisch gehiglightet, so ergibt sich eine Pseudobewegung, z.B. in einem Scatterplot, der Gruppen unterscheidet. Dabei hilft das Prinzip des gemeinsamen Schicksals.

Möglicher Fehlschluss:

Die gemeinsame Bewegung bedeutet noch keinen kausalen Zusammenhang. Die Gestaltprinzipien sind zugleich Chance und Last für die Explorative Datenanalyse, denn sie vermischen Alltagsheuristiken mit Erfahrungen aus Datenanalysen.

Bewertung und Einordung der Gestaltprinzipien für die Datenanalyse

Vorteile:

Die Wahrnehmungsheuristiken werden durch das Bearbeiten von Fragestellungen und das Lösen von Problemen ständig verbessert sodass die unbewussten Einschätzungen über Zusammenhänge liegen immer näher an den wirklichen Lösungen. So entwickelt sich durch das Bearbeiten vieler Datenanalyseaufgaben ein intuitives Verständnis für Fragestellung der Datenanalyse. Dadurch, dass die in der interaktiven statistischen Graphik entwickelten Plots der menschlichen Wahrnehmung angepasst sind, ist es möglich, viele Informationen gleichzeitig zu transportieren und zu verarbeiten.

Nachteile:

Die Wahrnehmungsheuristiken sind stark von der Alltagserfahrung des Individuums geprägt und entsprechen nicht zwingender Weise den mathematischen Bedingungen der verwendeten Plots. Die Reduktion von 3d-Darstellungsweisen auf 2d-Monitore sind ohne Interaktionsmöglichkeiten grundsätzlich zum Scheitern verurteilt, weil sie unserer Alltagserfahrung widersprechen. Damit verbunden ist das Problem, dass es viel Übung und großer Erfahrung bedarf, um sich bei der Arbeit mit der interaktiven statistischen Graphik von den Gestaltprinzipien zu lösen, wenn es die Problemstellung erfordert.

Die genannten Nachteile können durch Interaktion mit der statistischen Graphik ausgeglichen werden. Dazu sind in verschiedenen Softwarepaketen folgende Prinzipien verankert.

1. Der Nutzer kann alle graphischen Elemente in einem Diagramm abfragen und über diese Option die abstrakten Daten anzeigen lassen.
2. Die Plots können umskaliert und in der Darstellungsweise angepasst werden, so dass gestellte Fragen direkt abgelesen werden können, ohne dass der Nutzer eine weitere manuelle Rechnung vornehmen muss.
3. Das System zeigt Warnungen an, wenn es im Datensatz zu Ungereimtheiten kommt. D.h. Daten für einen bestimmten Plot fehlen, also sind die Daten für eine bestimmte Darstellungsweise nicht geeignet usw.

4. Die interaktive statistische Graphik ist so gestaltet, dass sie je nach Fragestellung eine geeignete Auswahl an Informationen, d.h. Daten, Fälle, Merkmale anzeigen kann. Der Nutzer kann, je nach seinen Bedürfnissen, logisch zoomen.

Der Übergang von einer konkreten grafischen Darstellung hin zu einener abstrakteren Beschreibung durch Kennzahlen geht immer mit einem Verlust an Informationen einher, wie die beiden folgenden Abbildungen zeigen. Beim Betrachten dieser Figuren wird deutlich, dass je nach Definition von „Größe" bzw. „Mächtigkeit" und dem Verständnis des mentalen Mess- oder Beobachtungskonzepts sich die Antworten unterscheiden (A. Zimpel, 2008).

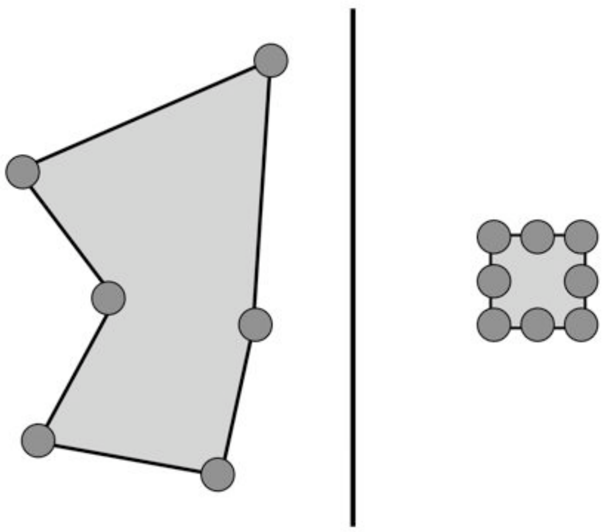


Abbildung 23: Welches visuelle Objekt ist größer ?

Die Punkte in Abbildung 23 und 24 können entweder als Eckmarkierungen einer Fläche interpretiert werden oder als abzählbare Punktmenge. Je nach Standpunkt ändert sich die Anwort auf die Frage: „Was ist größer?". Ein spezielles mentales Mess- oder Beobachtungskonzepts ist die Fähigkeit des Zählens. Heinz von Förster (1967) konnte zeigen, dass das neuronale Netz der menschlichen Retina (Netzhaut des Auges) in der Lage ist, Objekte zu unterscheiden und zu „zählen". Am Biological Computer Laboratory der University of Illinois konstruierten Heinz von Förster und Mitarbeiter eine künstliches neuronales Netz nach Vorbild der menschlichen Retina, das zählen konnte (H. Förster, 2003, S. 169).

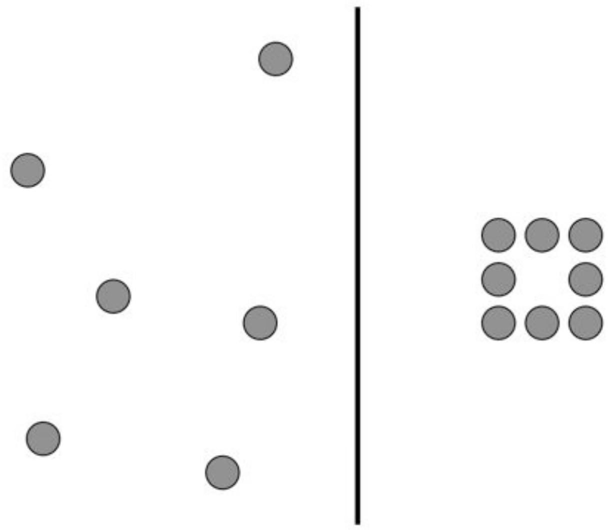


Abbildung 24: Wo befinden sich mehr Punkte ?

Diese Tatsache verdeutlicht, dass das mentale Konzept des Zählens ein tief durch evolutionäre Anpassung verankertes mentales Modell ist (E. Kandel, 1996, S. 366ff). Deshalb werden die Fähigkeit des Zählens und die damit verbundenen Informationsverarbeitungsprozesse im nächsten Abschnitt ausführlich beschrieben.

3.2 Die Fähigkeit des Zählens und die mathematische Abstraktion

„Eigentlich sollte man nämlich sagen:
<von etwas abstrahieren>, nicht <etwas abstrahieren>."

Immanuel Kant

Die Fähigkeit zu Zählen ist nicht nur dem Menschen vorbehalten. Der Verhaltensforscher Otto Köhler (1889-1974) konnte schon früh zeigen, dass verschiedene Vogelarten in der Lage sind, bis fünf oder sechs zu zählen (Zimpel, 2008, S. 29). Studien mit Katzen und Primaten konnten diese Ergebnisse bestätigen. Diese Erkenntnisse werfen die Frage auf, wodurch sich der zählende Mensch auszeichnet. Die Antwort ist die Entwicklung eines Systems aus Zahlzeichen, die den zählenden Tieren verschlossen bleiben. Diese Entwicklung begann vor mehr als 200000 Jahren in Afrika, noch bevor sich der Homo sapiens sapiens auf der Erde verbreitete. Aus dieser Zeit sind Funde eingekerbter Elefantenkochen erste Belege für die „mathematische Begabung" des Homo erectus (Zimpel 2008, S. 29). Zu den ersten Belegen für das empirische Forschen des Homo sapiens sapiens zählt sicher der Ishango-Kochen aus Zentralafrika, der auf ein Alter zwischen 9000 und 20000 Jahre geschätzt wird. Auf ihm sind strukturierte Ritzungen und Kerbungen zu sehen, die ihn je nach Deutung, als Rechenhilfsmittel oder Beobachtungsinstrument für die Mondphasen ausweisen (Wußing, 2008, S. 10). Beide dieser möglichen Interpretationen zeigen deutlich, das mathematisch empirische Beschreibungsformen der Umwelt schon in der Frühzeit der Menschheitsgeschichte eine wichtige Rolle spielten und die Menschen versuchten, für das zahlenmäßige Erfassen ihrer Umwelt geeignete Beschreibungsformen zu finden.

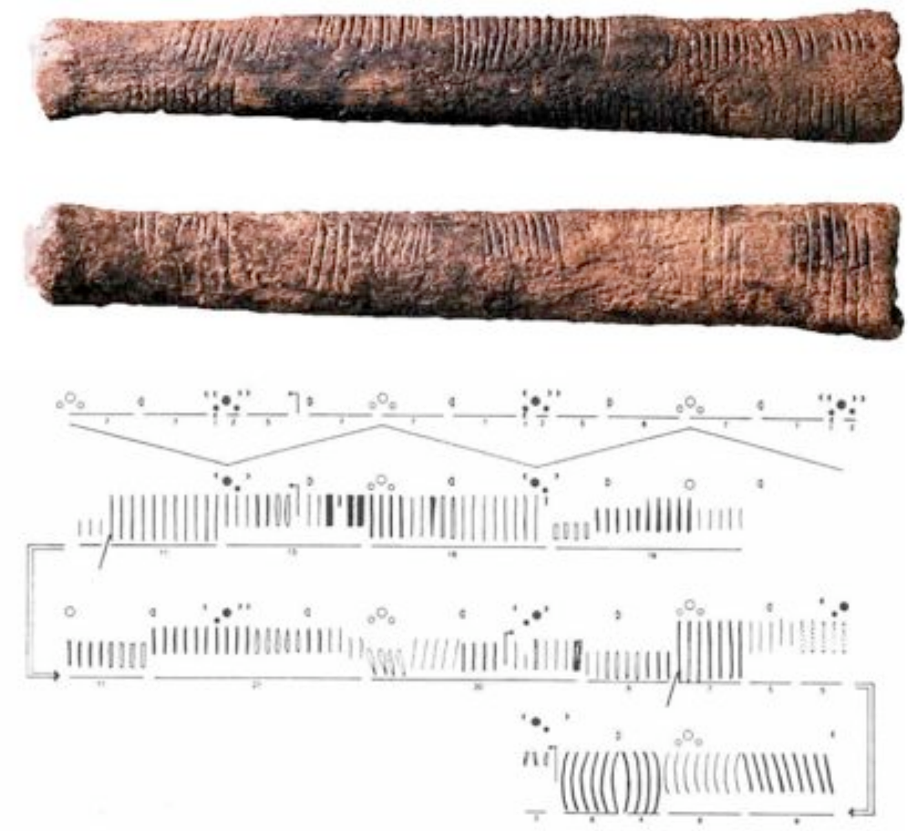


Abbildung 25: Beispiel einer frühen empirischen Beobachtung: Ishango-Knochen aus zentral Afrika ca. 9000 - 20000 Jahre alt (H. Wußing 2008, S. 10)

Die anfänglich benutzten Kerbungen auf Knochen oder Strichlisten auf anderen Materialien haben alle das Problem, dass sehr bald (ab einer Anzahl zwischen fünf und acht) die Strichliste aus dem Aufmerksamkeitsfenster der zählenden Person tritt und die Zahl nicht mehr auf Anhieb erkannt werden kann. Durch die Umordnung der Stichliste und der Anordnung in Blöcken z.B. bei Dreierblöcken ist es wieder möglich, die Strichlisten ganz ins Aufmerksamkeitsfenster zu holen und die entsprechende Anzahl auf Anhieb zu erkennen, z.B. Anordnung der Augen auf einem Würfel (Zimpel, 2008, S. 35). Aus diesen Umordnungen entwickelten sich neue Zeichen für Zahlenblöcke und schließlich Symbole, d.h. Superzeichen für Zahlen.

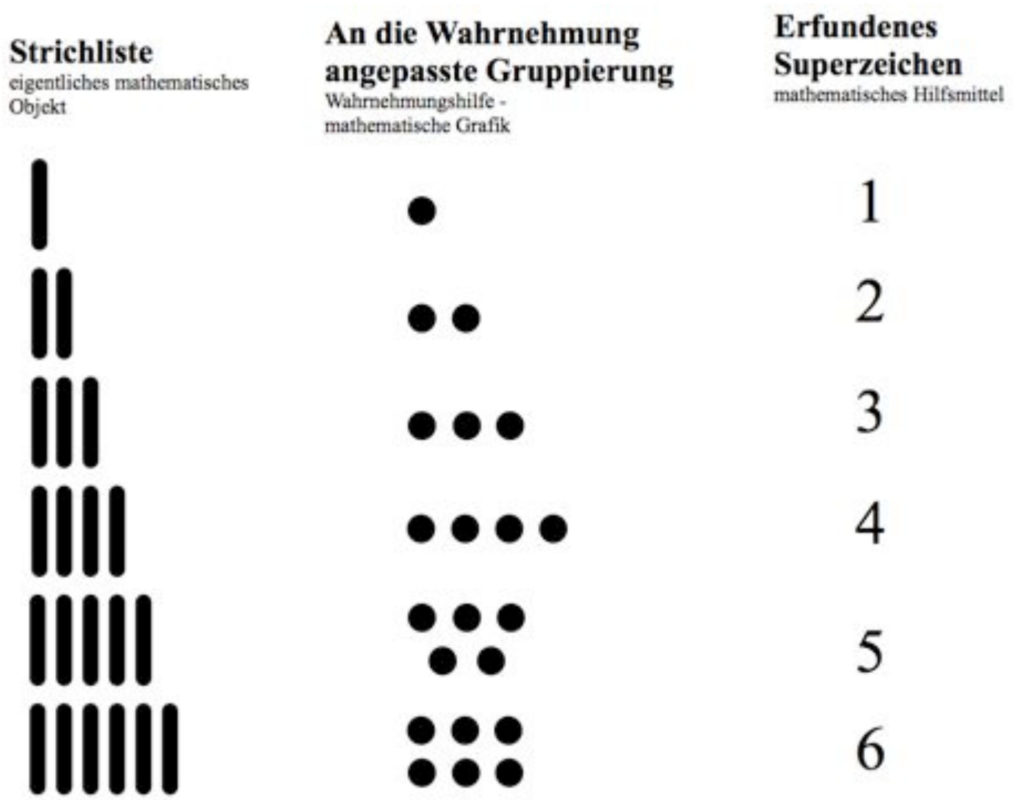


Abbildung 26: Beispiel zum Wahrnehmungsfenster des Menschen

Ohne diese Superzeichen und ihre Benennung mit Zahlwörtern ist der menschliche Zahlensinn überraschend dürftig. Dies wird bei Völkern und gesellschaftlichen Gruppen augenscheinlich, die in ihrer Sprache keine differenzierten Zahlwörter entwickelt haben (Zimpel, 2008, S. 33). Das konnte der Verhaltensforscher Peter Gordon (2004, S. 496-499). mit seinen Arbeiten über das Jäger- und Sammlervolk der Pirahas zeigen Die Pirahas kennen in ihrer Sprache nur Zahlwörter für „eins", „zwei" und „viele". Wenn sie nun die Aufgabe erhielten, Mengen mit mehr als sechs Gegenständen zu vergleichen, verloren sie den Überblick. Dies zeigt, dass die Einführung der Zahlwörter und der Superzeichen, und die damit verbundene Abstraktion ein entscheidender Schritt für die Entwicklung mathematischer Fähigkeiten ist.

Die Entwicklung unseres „arabischen” Zahlensystems hat seinen Anfang in strukturierten Strichlisten. Aus Strichlisten und Zahlwörtern entwickelten sich in Indien die Brahmizahlen, die um das Jahr 522 durch den indischen Mathematiker Bhaskara I durch die Null als Zeichen ergänzt wurden (Wußing, 2008, S. 98). Über den arabischen Kulturraum kamen diese Ziffernsymbole und die Idee des Dezimalsystems nach Europa. Unsere heute gültigen Ziffernsymbole haben sich aus einer westarabischen Form des 11. Jahrhunderts entwickelt und wurden schnell populär, da das Rechnen mit einem Stellenwertsystem das Denken beschleunigt. D.h., dieses System erhöht die Geschwindigkeit, mit der Rechenoperationen durch Individuen ausgeführt werden können.

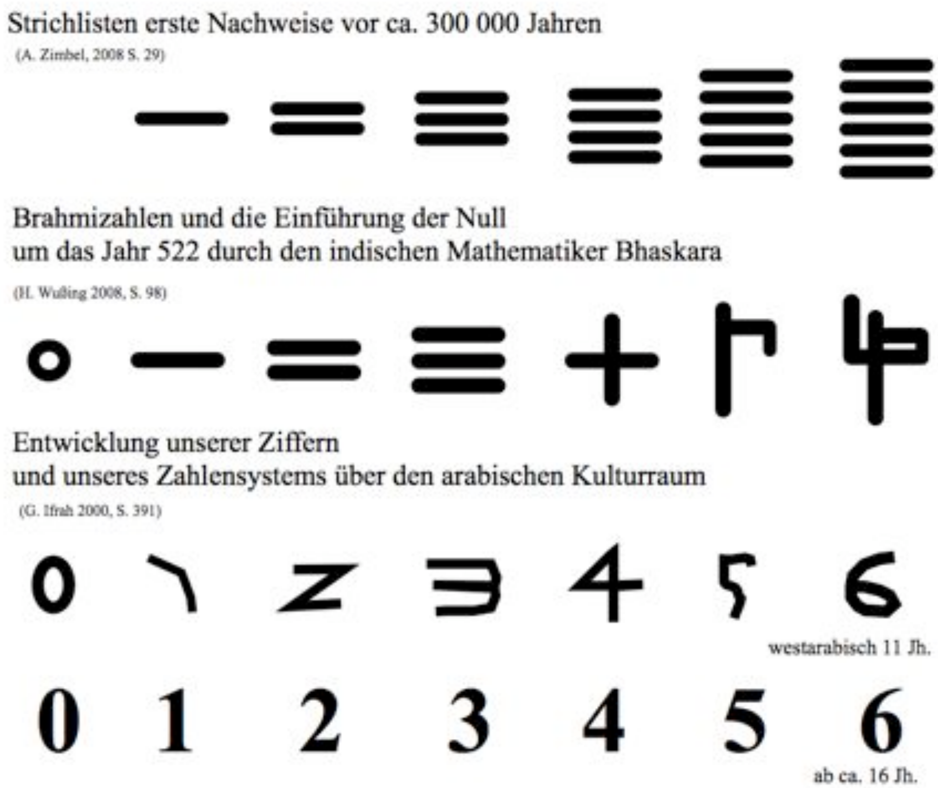


Abbildung 27: Die Entwicklung der Zahlen

Abstraktion als „Denkbeschleuniger”

Anhand von Abbildung 27 kann der Leser selbst sehr leicht nachvollziehen, dass die Differenzbildung in einem Zahlensystem, das einer Strichliste entspricht, sehr viel mühsamer ist als mit einem Stellenwertsystem, das in seiner Struktur der menschlichen Wahrnehmung angepasster ist. Ein Zeichensystem, das als „Denkbeschleuniger“ in der Mathematik wirken soll, muss einerseits an die menschliche Wahrnehmung angepasst und andererseits zweckmäßig für den mathematischen Gegenstand sein. Ein kleines Beispiel dafür soll diese Aussage illistrieren: Das römische oder das altgriechische Zahlensystem sind zwar der menschlichen Wahrnehmung angepasst, da sie aus den Zahlwörtern des Lateinischen bzw. des Griechischen entstanden sind. Diese Zahlensysteme sind für die Entwicklung einer abstrakten Mathematik ungeeignet, da sie die Eigenschaften der natürlichen Zahlen als unendliche kommutative Gruppe eben nicht in natürlicher Weise reflektieren. Deshalb ist das Rechnen mit diesen Zahlensystemen viel umständlicher als das Rechnen mit dem arabischen Ziffernsystem.

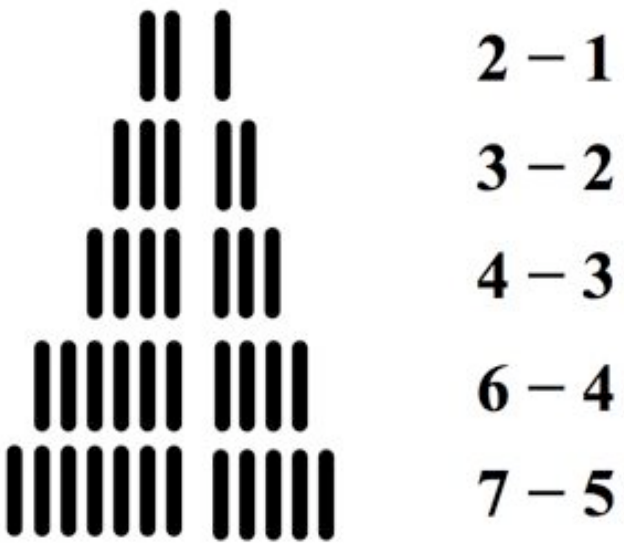


Abbildung 28: Arabische Zahlen als Denkbeschleuniger

Mathematische Superzeichen (angefangen vom Zahlensystem bis hin zu komplizierten Tensorausdrücken) müssen also den Spagat zwischen menschlicher Wahrnehmung und mathematischer Angemessenheit bewältigen, um vorteilhaft bei der mathematischen Problemlösung wirken zu können. Das Aufmerksamkeitsfenster des Menschen ist sehr begrenzt: Es liegt zwischen drei und fünf Einheiten. Wir können drei bis fünf Striche auf ein Mal erkennen und ihnen eine Zahl zuordnen. (Zimpel, 2008, S.165). Wie verhält es sich dann aber mit der Sprache oder diesem Text, der offensichtlich aus sehr vielen Zeichen besteht und dennoch gelesen werden kann?

Wahrnehmung von Schrift

Die Tatsache, dass Schrift gut, schnell und flüssig gelesen werden kann, liegt an zwei Faktoren. Der Prozess ist wird durch viel Üben automatisiert und die Phänomene Schrift und Sprache haben sich stark an das Aufmerksamkeitsfenster des Menschen angepasst bzw. sich kompatibel entwickelt. Im Hintergrund steht, dass das menschliche Gehirn die Schrift und die Sprache „erfunden" hat und dass durch einen evolutionären Aushandlungsprozess zwischen den sich verständigenden Individuen ein „für alle Gehirne geeigneter" Schrift- und Sprachgebrauch sich entwickelt haben – inklusive regionaler Besonderheiten in den unterschiedlichen Sprachfamilien. Das Gehirn ist zunächst nicht für das Lesen gebaut (Spitzer, 2003, S. 243). Durch seine hohe Plastizität ist es aber möglich, dass sich die Strukturen im Gehirn so anpassen, dass das Lesen hoch automatisiert vor sich geht. Es ist nicht möglich, Schrift nicht zu lesen. Dieser automatisierte Prozess des Lesens verläuft sehr rasch und ist nicht zu unterdrücken.

Abbildung 29: Dominanz automatisierter Prozesse: Das Stroop-Experiment

Dies kann durch das „Stroop-Experiment" (Stroop, 1935, S. 643-662) gezeigt werden. In diesem Experiment sollen verschieden farbige geometrische Objekte mit ihrer jeweiligen Farbe benannt werden. Sind die dabei verwendeten Objekte irgendwelche Flächen oder unspezifische Zeichen, so fällt es relativ leicht, die richtige Farbe zuzuordnen. Handelt es sich dabei um Farbwörter, die nicht mit der Farbe übereinstimmen, ist es für Versuchspersonen schwer, die Farben richtig zu benennen. Dies liegt daran, dass ein geübter Leser die Form und Struktur des Wortes wahrnimmt, ohne das Wort Buchstabe für Buchstaben zu lesen. Das Denken läuft der Wahrnehmen sozusagen voraus. Aber wie kommt es zu diesem Phänomen?

Wahrscheinlichkeitsschätzungen beim Lesen und Informationsgehalt von Zeichenketten

Um diese Frage zu lösen, ist es zunächst hilfreich, den Informationsgehalt einer Zeichenkette abzuschätzen. Diese Abschätzung liefert wichtige Informationen darüber, wie sich das Lesen „mathematischer Texte” (hier meint „mathematischer Text“ eine Kombination aus Schrift und mathematischen Superzeichen) und das Lesen niedergeschriebener Sprache unterscheiden. Die nachfolgende detaillierte Modellrechnung gibt eine Größenabschätzung für die Komplexität der Verarbeitungsprozesse in Retina und Gehirn an. Die ausführliche Herleitung zeigt, mit wie wenig Voraussetzungen die Verarbeitung von Sprache und mathematischen Superzeichen unterschieden werden kann. Dabei liegen die aufgezeigten Unterschiede zwischen der Verarbeitung von Sprachzeichen und mathematischen Superzeichen auf einer grundsätzlich logischen Ebene und nicht auf der Individualebene des Lernenden. Da die Unterscheidung von Sprachzeichen und mathematischen Superzeichen nicht naheliegend ist und nur eine vollständige Darstellung das Problem zeigt, folgt im nächsten Abschnitt eine Herleitung aus elementaren Grundannahmen.

Claude Elwood Shannons (C. Shannons, 1949, S. 31-64) zeigte zusammen mit Warren Weaver in seinem Buch „The Mathematical Theory of Communication“, dass sich das Maß für Information durch eine Entropiefunktion $H(\{P_i\})$ beschreiben lässt. Diese Funktion gibt nur ein untere bzw. obere Abschätzung für eine im Detail viel kompliziertere Fragestellung. Mit der Entropiefunktion $H(\{P_i\})$ lässt sich also das Funktionieren von Sprache nicht erklären, sondern es können nur einige Effekte quantifiziert werden.

Shannon-Entropie

Der Ansatzpunkt für die Shannon-Entropie ist eine Menge Z (z.B. ein Alphabet), wobei jedem Element z_i (z.B. der i-te Buchstabe des Alphabets) der Menge $Z=\{z_1,z_2,z_3,...,z_n\}$ eine Auftretenswahrscheinlichkeit P_i zugeordnet werden. Shannon geht von einer gedächtnislosen Zeichenquelle aus, die unablässig und zufällig Zeichen sendet. Unter diesem Fall ist die Shannon-Entropie (Wachter, 1998, S. 442-444 und Shannon, 1949, S. 116-119) durch folgende Bedingungen festgelegt:

(1) $H(\{P_i\})$ ist eine kontinuierliche, differenzierbare und eindeutige Funktion der Wahrscheinlichkeiten $\{P_i\}$.

(2) Bei N gleichwahrscheinlichen Ereignissen $P_i=1/N$, ist die Entropie H eine monoton wachsende Funktion von N. Diesen Spezialfall kann man mit I(N) bezeichnen. Also

$$I(N) = H\left(\left\{P_i = \frac{1}{N}\right\}\right)$$

(3) Die Unsicherheit in einem Satz von Wahrscheinlichkeiten bleibt unverändert, wenn man diesen in Wörter einteilt:

$$w_1 = \sum_{i=1}^{n_1} P_i \,, \quad w_2 = \sum_{i=n_1+1}^{n_2} P_i \,, \ldots , \quad w_k = \sum_{i=n_{k-1}+1}^{n_k=N} P_i$$

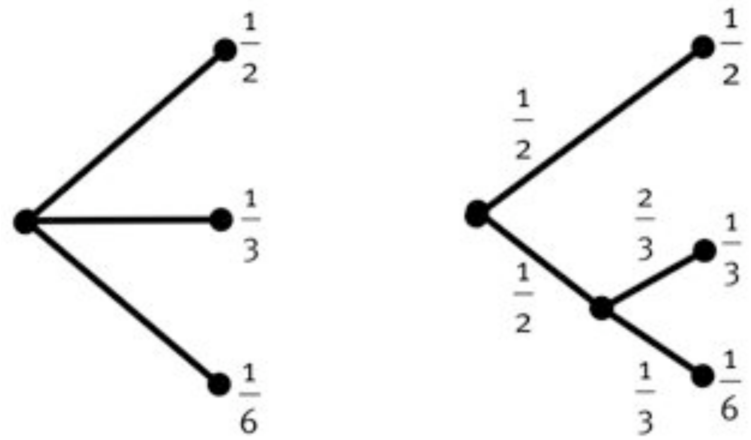
So gilt:

$$H(\{P_i\}) = H(\{w_i\}) + \sum_{j=1}^{k} w_i H\left(\left\{\frac{P_i}{w_j}\right\}\right)$$

Hier ein Beispiel zur Verdeutlichung des Sachverhalts. Seien $p_1=1/2$, $p_2=1/3$, $p_3=1/6$ so folgt nach obiger Definition:

$$H\left(\frac{1}{2},\frac{1}{3},\frac{1}{6}\right) = H\left(\frac{1}{2},\frac{1}{2}\right) + \frac{1}{2}H\left(\frac{2}{3},\frac{1}{1}\right)$$

Diesen Sachverhalt kann man sich durch einen Wahrscheinlichkeitsbaum verdeutlichen:



Um H({Pi}) konkret angeben zu können, betrachten wir zunächst den Fall gleicher Wahrscheinlichkeiten

$$P_i = \frac{1}{nm}, \ldots, P_{nm} = \frac{1}{nm}, n, m \in \mathbb{N}$$

die in m gleichgroße Wörter unterteilt werden:

$$w_i = \sum_{i=1}^{n} P_i = \frac{1}{m}, \ldots, w_m = \sum_{i=(m-1)n+1}^{nm} P_i = \frac{1}{m}$$

mit (2) können wir schreiben:

$$I(nm) = I(m) + \sum_{j=1}^{m} \frac{1}{m} I(n) = I(m) + I(n)$$

aus der Bedingung (1) folgt, dass man H({Pi}) nach ableiten können. So folgt, wenn man p=nm einsetzt:

$$m \frac{d}{dp} I(p) = \frac{d}{dn} I(n)$$

mal n auf beiden Seiten liefert:

$$p\frac{d}{dp}I(p)=n\frac{d}{dn}I(n)=const.$$

ist konstant, da man p ändern kann ohne n zu variieren. Also folgt:

$$I(n)=k\ln n$$

wobei k eine Konstante ist.

Wenn man jetzt den Fall gleicher Wahrscheinlichkeiten mit unterschiedlich großen Wörtern betrachtet, kann man schreiben:

$$P_i=\frac{1}{n},\; w_j=\frac{\alpha_j}{n},\; \sum_j\alpha_j=n,\; n,\alpha_j\in\mathbb{N}$$

mit (3) folgt:

$$I(n)=H\left(\left\{w_j\right\}\right)+\sum_j\left(\frac{\alpha_j}{n}\right)I\left(\alpha_j\right)$$

$$\Rightarrow H\left(\left\{w_j\right\}\right)=\sum_j\left(\frac{\alpha_j}{n}\right)\left[I(n)-I\left(\alpha_j\right)\right]=-k\sum_j\left(\frac{\alpha_j}{n}\right)\left[\ln\alpha_j-\mathrm{l}\right.$$

$$=-k\sum_j\left(\frac{\alpha_j}{n}\right)\ln\left(\frac{\alpha_j}{n}\right)=-k\sum_j w_j\ln w_j$$

mit $w_i \rightarrow P_i$ und $k=1/\ln2$ (man möchte den Informationsgehalt in bit) kann man die Shannon-Entropie sinnvoll definieren:

$$H=-\frac{1}{\ln2}\sum_i P_i\ln P_i,\; k>0,\; \sum_i P_i=1$$

mit

$$\log_b r=\frac{\log_a r}{\log_a b}$$

folgt:

$$H = -\sum_i P_i \frac{\ln P_i}{\ln 2} = -\sum_i P_i \log_2 P_i$$

Für ein Alphabet mit n gleichwahrscheinlichen Zeichen folgt also:

$$H(n) = -\sum_i P_i \log_2 \frac{1}{n} = -\log_2 \frac{1}{n} \sum_i P_i, \ mit \sum_i P_i = 1$$

$$H(n) = -\log_2 \frac{1}{n} \ mit \ \ln \frac{1}{x} = -\ln x \ gilt: H(n) = \log_2 n$$

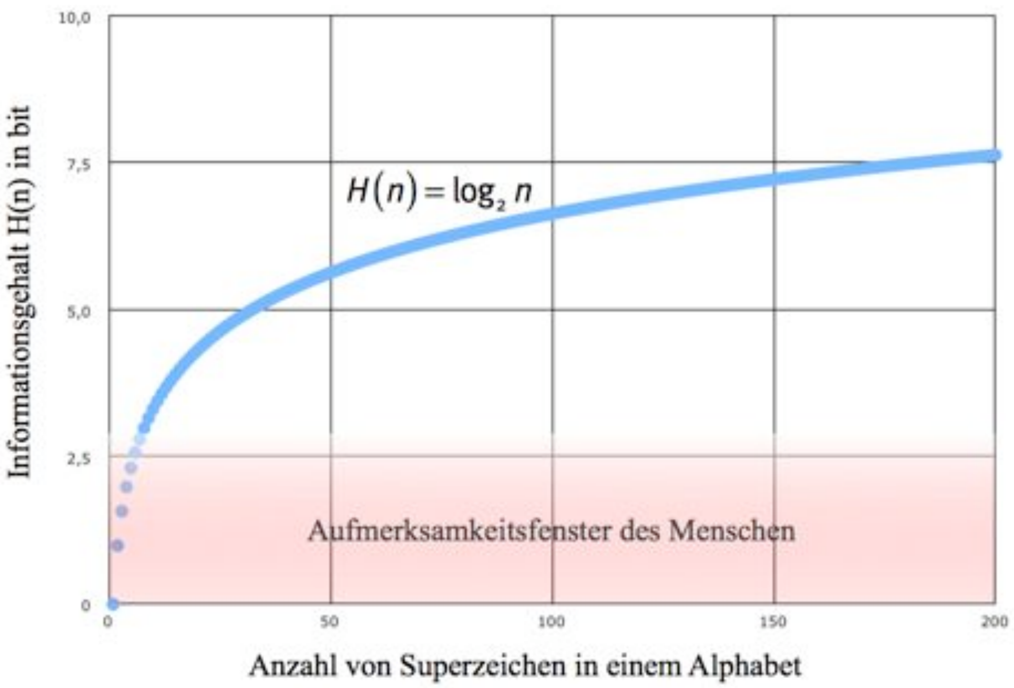


Abbildung 30: Die Shannon-Entropie in Abhängigkeit von der Anzahl der Superzeichen

Das Aufmerksamkeitsfenster des Menschen liegt im Duchschnitt bei drei bis vier Einheiten (z.B. Formen, Objekte, Zeichen). D.h., dass nach der Shannon-Entropie Probleme in der Größenordnung von 1,6 bis 2 Bit parallel verarbeitet werden können. Unser Alphabet mit 26 Zeichen liegt bei der Zeichenschätzung bei 4,7 bit mit der Annahme, das alle Zeichen gleich wahrscheinlich wären. Dies ist aber nicht der Fall, wie folgendes Diagramm zeigt:

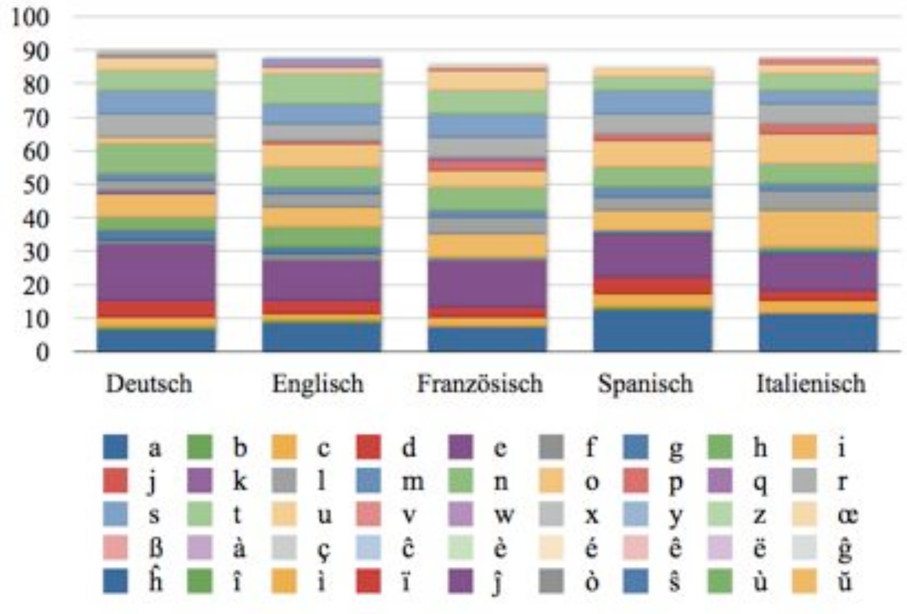


Abbildung 31: Die charakteristische Häufigkeitsverteilung von Buchstaben in verschiedenen Sprachen

Die Verteilung der Buchstaben ist für jede Sprache charakteristisch. Für ein einzelnes Zeichen mit der Auftrittswahrscheinlichkeit P_i ergibt sich aus der Shannon-Entropie:

$$H\left(P_i\right) = -\log_2 P_i$$

Je seltener das jeweilige Zeichen ist, um so mehr Information steckt in ihm. Folgendes Diagramm zeigt am Beispiel der Buchstaben E und Q aus einem deutschen Text den jeweiligen Informationsgehalt. Beim Lesen von Texten nützt der Mensch diesen Effekt und schätzt (nicht nur in der Abhängigkeit der Buchstabenhäufigkeit, sondern auch abhängig von Silben, Buchstabenfolgen und Wortlängen) ab, welche Zeichen im noch nicht gelesenen Text vorkommen könnten. Die Sprache und die Schrift haben sich im Laufe der evolutionären Entwicklung und kulturellen Anpassung so entwickelt, dass sie genau in das Aufmerksamkeitsfenster des Menschen passen. Dies konnte nicht nur an europäischen Sprachen gezeigt werden, sondern an fast jeder Sprache der Welt (Zimpel, 2008 S.??). Bei der Entwicklung der Schrift ist ein ganz ähnlicher Anpassungsprozess zu beobachten: Sowohl die auf Lauten basierenden Schriftbilder von Sprachen als auch symbolisierte Schriftsysteme passen zum Aufmerksamkeitsfenster des Menschen. Dieser Sachverhalt zwischen Häufigkeit und Auffälligkeit lässt sich mit der Shannon-Entropie begründen. Setzt man den Beitrag, den ein Zeichen zur Gesamtunsicherheit eines Zeichensystems leistet, mit seiner Auffälligkeit gleich, lässt sich dieses Optimierungsproblem einfach lösen. Dabei zeigt sich, das der höchste Wert dieses Maximierungsproblems bei drei liegt, also innerhalb des Aufmerksamkeitsfensters des Menschen.

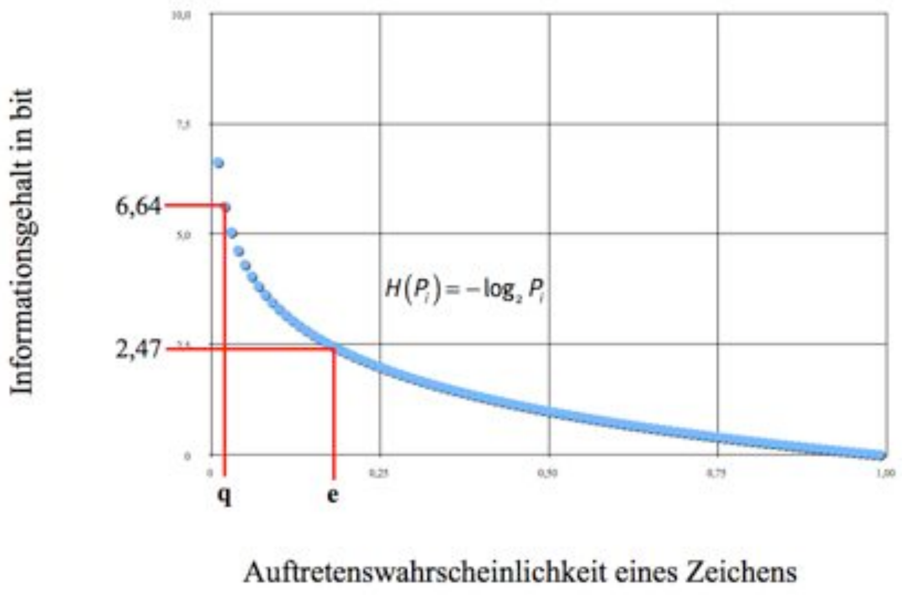


Abbildung 32: Informationsgehalt in Abhängigkeit der Auftretungswahrscheinlichkeit eines Zeichens

Die Entropie H_n für Wörter w der Länge n ist gegeben durch:

$$H_n = -\frac{1}{n} \sum_{w \in Z^n} p_w \cdot \log_2 p_w$$

Ein Text wird nach den beschriebenen mathematischen Regeln wie folgt aufgebaut:

1. Man generiert gemäß der Häufigkeit, beispielsweise der deutschen Sprache, eine zufällige Zeichenkette aus dem Alphabet.

2. Diese Zeichenfolge teilt man in Wörter der Länge n ein. n wird zufällig nach den Worthäufigkeiten der deutschen Sprache erzeugt.
3. Man geht nun die Zeichenkette durch und sorgt mittels Zufallsgenerator dafür, dass die Buchstabenkombinationen (z.B. un, qu, en, ie) der Verteilung in der deutschen Sprache entsprechen. Daraus entsteht ein Text, der vermeintlich ein deutsches Sprach- und Klangmuster besitzt. (Brian, 1988, S. 103).

Damit lässt sich jetzt der „Stroop-Effekt” erklären. Wenn wir lesen, überholt unser Denken die Wahrnehmung. Dadurch ist es möglich, Texte schnell und flüssig zu lesen. Dies gelingt auch mit Texten, bei denen Buchstaben fehlen oder die orthografisch nicht richtig sind. Die Fehlererkennung des Menschen ist in dieser Hinsicht „perfekt”, und um Größenordnungen besser als jedes heute verfügbare Computersystem.

In welchem der folgenden „Texte“ fällt der Stern am meisten auf?

1. Lorem ipsum dolor sit amet, ligula suspendisse nulla pretium, rhoncus tempor placerat fermentum, enim integer ad vestibulum volupitat. Nislirhoncus turpis est, vel elit, congue wisi enim nunc ultricies sit, magna ad vestibulum eros in vel, volupitat nec el pellentesque leo, temporibus scelerisque nec.
2. *************************** ****************************Litora************* lacinia*************** ********* ******* mauris nulla ****** **** ******** ****************** ****************Mauris at suspendisse******* ************* ************************ Vitae vehicula **pien***** ********et******a*t*au*t**** ******platea***** *****suspen*********disse id.
3. Montes et m*etus adipiscing placerat consectetuer nunc. Non libero nam et al dolor. Nascetur quis ut, tristique libero odio sit tempus, a*c ut in et felis convallis. Pe*llentesque dignissim ame*t commodo, nec tur*pis dignissim torquent, laor*eet el orci unde aptent ten*etur, dol*or sit. Sed sed mau*ris duis. Quis e*nim ut, cursus dol*or id arcu explicabo ligula, quis*que natus mauris sed nul*la in, ac sed ve*hicula.

4. Pell*entesq*ue e*u al*iqu*et v*el i*n vi*tae ultr*icies. V*itae vehi*c*ula proin arcu *at lo*bortis. Ultr*ici*es mo*les*t*ie lib*e*ro d*ign*issi*m id m*a*uri*s, m*u*s n*ec te*m*p**us l*o***rem, l**acinia v*estib***m nec elem***entum, s**apien et a*t a*u**t pl*at*ea su*spen*dis**se id. El**it a i*d, at po***suer***e vel penatibu**s el or**ci saepe o**rci, ad***ipisc***ing, v***ivamu**s in. V**itae ve*h**icu**la pr***oi*n ar**cu a*t lo***bo*r*tis.

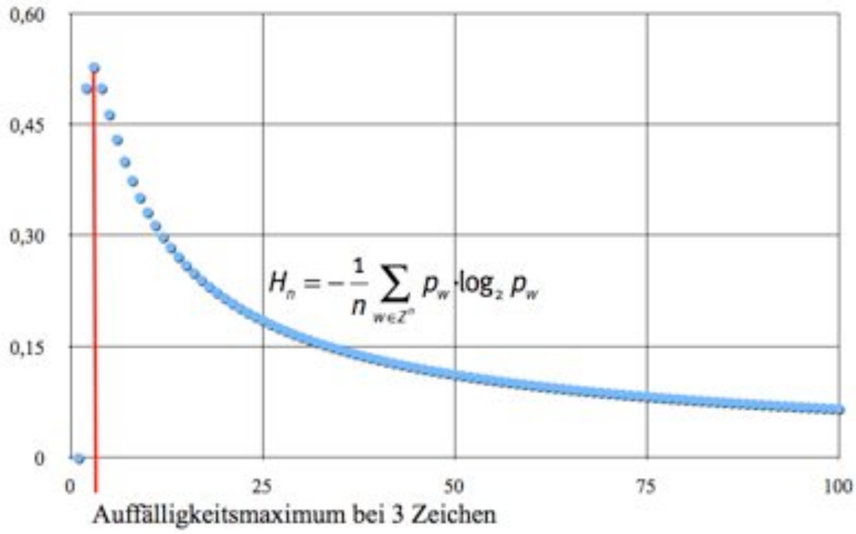


Abbildung 33: Auffälligkeit von Zeichen in Zeichenketten

Mathematische Zeichenketten und die menschliche Wahrnehmung

Fasst man Zahlen und mathematische Symbole zu einem Zeichensystem zusammen, wie es üblicherweise in der Mathematik verwendet wird, so ist dieses nicht optimal an die menschliche Wahrnehmung angepasst. Folgende Probleme treten in den Vordergrund:

1. In diesem Zeichensystem gibt es mehr als 26 Zeichen.
2. Die Häufigkeitsverteilung mathematischer Zeichen sagt nichts oder nur sehr wenig aus über den mathematischen Satz, den sie beschreiben.

3. In einem mathematischen Beweis kann aus der Zeichenkette kaum geschlossen werden, welcher nächste Beweisschritt nötig ist und welche Beweisidee zugrunde liegt.
4. Dies zeigt, dass mit mathematischen Texten anders umzugehen ist als mit Texten menschlicher Sprache. Der mathematische Formalismus und die in mathematische Zeichen übersetzten Beweise sind der menschlichen Wahrnehmung nicht so stark angepasst wie Sprache und Schrift. Mathematische Zeichen, Zahlen und Kalküle müssen dem mathematischen Gegenstand angepasst sein, um als „Denkbeschleuniger" wirken zu können. Bei einer schnellen und effektiven Handhabung mathematischer Zeichensysteme darf die Anpassung an die menschliche Wahrnehmung nicht zu kurz kommen. Aus diesen Überlegungen folgt, dass die Mathematik nur sehr begrenzt so erlernt werden kann wie eine Sprache. Zwar hat der mathematische Formalismus an einigen Stellen Eigenschaften einer Fremdsprache, wesentliche Merkmale einer Sprache fehlen aber. Allein dieses Argument spricht gegen einen Mathematikunterricht, der die sprachliche Rezeption in den Mittelpunkt stellt. So wie beim Erlernen einer Sprache ein wichtiges Element des Lernens das eigene Sprechen ist, so ist beim Lernen mathematischer Zusammenhänge sie eigene Auseinandersetzung mit den mathematischen Objekten unabdingbar (V. Ulm, 2008, S. 11f) Das Öffnen individueller Lernwege, die es dem Einzelnen ermöglichen, durch direktes Manipulieren und durch „in die Hand nehmen" mathematische Objekte und deren Eigenschaften zu entdecken, löst dieses Problem (V. Ulm, 2008, S. 25f). An den Universitäten werden im Curriculum der mathematischen Studiengänge meist Vorlesung und Übung kombiniert. Studierende der Mathematik wissen, dass ohne aktive Teilnahme an den Übungen der Besuch einer Vorlesung nicht sehr Erfolg versprechend sein wird.

Um die kognitiven Vorgänge besser modellieren zu können, werden im nächsten Abschnitt Grundlagen von Lernen und Gedächtnis kurz beschrieben.

3.3 Ergebnisse der Überlegungen zu Wahrnehmung und Abstraktion

Die vorangegangenen beiden Abschnitte stellen zwei wichtige Erkenntnisse für die Analyse des Lernprozesses der Explorativen Datenanalyse zur Verfügung:

Die visuelle Wahrnehmung des Menschen kann in ganz natürlicher Weise mit Wahrscheinlichkeiten umgehen und sehr leistungsfähig zweidimensionale Muster und Strukturen erkennen. Mit den beschriebenen Gestaltprinzipien ist es möglich, Plots dahingehend zu untersuchen, ob sie sich als Problemlösewerkzeug eignen.

Die Abstraktion von „Etwas" hin zu mathematischen Objekten, z.B. Zahlen, kann das Denken beschleunigen und ebenfalls der Problemlösung dienen. Die Abstraktion vom eigentlichen Objekt birgt auch die Gefahr, dass ein abstraktes Zeichensystem außerhalb des Wahrnehmungsfensters des Menschen liegt und es viel Übung und eigener Aktivität bedarf, um solch ein Zeichensystem zu verstehen.

3.4 Funktionelle Asymmetrie des Gehirns

„Das Tier taugt zu allem, was es soll, vollkommen –
der Mensch zu nichts recht, als was er lernt, liebt und übt."

Johann Heinrich Pestalozzi

Das bewusste Denken des Menschen, auch das kognitive Lernen, findet in der Großhirnrinde (Cortex) statt (W.Edelmann, 2000, S. 5ff) . Neben dem bewussten Lernen sind für den Lernerfolg die Motivation und weitere soziale Komponenten wesentliche Faktoren. Die Zentren für Motivation liegen in tieferen Hirnarealen. Die Großhirnrinde teilt sich in zwei Hemisphären auf, auf welchen die Mehrzahl der sensorischen und motorischen Funktionen symmetrisch angeordnet sind. Einige für das Lernen ganz wesentliche Funktionen sind dagegen asymmetrisch nur einer Gehirnhälfte zugeordnet. Bei Rechtshändern befinden sich in der linken Großhirnhälfte die Sprachzentren, denen auch die Bereiche für analytisches Denken einschließlich das Rechnen zugeordnet sind. In der rechten Hirnhälfte befinden sich die Zentren für räumliches Denken und Mustererkennung. Bei den zwei Prozent Linkshändern liegt das Sprachzentrum in der rechten Hirnhälfte und die Bereiche für räumliches Denken in der linken Hirnhälfte.

Das Wissen über die Funktionsverteilung im Großhirn wurde mit Hilfe von „Split-Brain-Patienten“ untersucht (J. Pinel, 2007, S. 530ff). Bei diesen Menschen sind die Gehirnhälften durch Unfall, Krankheit oder von Geburt an getrennt, so dass den einzelnen Gehirnhälften die genauen Funktionen zugeordnet werden können.

Insgesamt lassen sich die Leistungen der linken Hemisphäre zusammenfassen:

- Herstellung der Verbindung zum Bewusstsein
- Zuständigkeit für sprachliches und begriffliches Denken
- Zuständigkeit für arithmetische Berechnungen
- Zuständigkeit für abstraktes und analytische Zusammenhänge

Die Leistungen der rechten Hemisphäre lassen sich zusammenfassen in:

- musikalische Fähigkeiten
- Bild- und Mustererkennung
- geometrisches und räumliches Denken
- einheitliche und konkrete Zusammenhänge

In der älteren Literatur wird davon gesprochen, dass die linke Hirnhälfte die dominante und die rechte die subdominante (untergeordnete) Hemisphäre sei. Es stimmt, dass die linke Hemisphäre die herausragenden Leistungen des Schreibens, Rechnens und abstrakten Denkens übernimmt (W.Edelmann, 2000, S. 5ff). Die Einteilung in „dominant“ und „untergeordnet“ ignoriert aber die Tatsache, dass auch die rechte Hemisphäre einen hoch entwickelten Teil des menschlichen Gehirns darstellt. Bei dem Versuch, ein schlichtes geometrisches Muster aus einfachen farbigen Blöcken nachzubilden, versagt die linke Hemisphäre, während die rechte diese Aufgabe rasch und exakt erfüllt. Das gleiche gilt für die Mustererkennung über den Tastsinn. Weiterhin kann nachgewiesen werden, dass die rechte Hemisphäre bei der Erkennung bildlicher Darstellungen besonders leistungsstark ist. Für manche Aufgabenarten ist sie sogar der eindeutig überlegene Teil (J. Pinel, 2007, S. 543ff). Immer wenn es um den Umgang mit Flächen, Körpern und die umfassende Orientierung auf anschaulicher Grundlage geht, kommen ganz wesentliche Teile der kognitiven Leistungen von der rechten Gehirnhälfte.

Bei normalen Gehirnen ist es unmöglich, nur eine Hälfte zu entwickeln und die andere weitgehend zu vernachlässigen. In einem Statistik-Seminar wird die rechte Hemisphäre ebenso gefördert wie die linke. Umgekehrt wird die linke Hemisphäre in einem Musik- oder Malkurs nicht weniger beansprucht als die rechte (J. Pinel, 2007, S. 560ff). Trotz dieser prinzipiell ganzheitlichen Funktionsweise des menschlichen Gehirns gibt es gewisse Anhaltspunkte dafür, dass sich Menschen im Ausmaß der Aktivierung der beiden Hemisphären unterscheiden. Manche Menschen neigen dazu, die Dinge eher linkshemisphärisch-analytisch oder eher rechtshemisphärisch-intuitiv zu erkennen. Die Lehrpläne an allgemeinbildenden Schulen sind besonders auf das analytisch-begriffliche Denken der linken Hemisphäre ausgerichtet. Neuere Untersuchungen gehen aber davon aus, dass eine stärkere Einbeziehung der rechten Hemisphäre mit ihrem intuitiven und bildhaften Denken und ihrem Bezug zu emotionalen Prozessen zu einer wesentlichen Steigerung des Lernpotentials führen könnte (M.Arnold, 2002, S. 107ff). Diese Steigerung des Lernpotentials versucht man in der Pädagogik durch multiple Präsentation nutzbar zu machen.
Die Intuition ist in vielen Lebensbereichen eine wichtige Erkenntnisquelle. Gekennzeichnet ist Intuition durch die Unmittelbarkeit der Erkenntnis und eine eingebungsartige Erfassung des Wesentlichen (W. Edelmann, 2000, S. 142ff). So ist die Aufgabe, in einem größeren Wissensgebiet Zusammenhänge systematisch darzustellen, ohne eine zunächst intuitive Gesamtschau unmöglich. Intuitives Denken beruht gewöhnlich auf einer Vertrautheit mit dem fraglichem Wissensbereich und mit dessen Struktur. Dies ermöglicht, Stufen auszulassen und Abkürzungen zu gehen. Der Lernende tut dies auf intuitive Art und Weise, die er später analytisch überprüft und in seine Schlussfolgerung einbaut. Intuition ist ein wesentliches Merkmal des produktiven Denkens und Problemlösens.

3.5 Gedächtnis und Übung

„Im Menschen ist nicht alles Gedächtnis,
sondern Erinnerung.“

Thomas von Aquin

Das Gedächtnis ist kein passiver Speicher. Man unterscheidet das Kurz- und Langzeitgedächtnis. Das Kurzzeitgedächtnis speichert zunächst alle Informationen, die von außen über die sensorischen Zentren der Großhirnrinde im Gehirn ankommen. Die Informationen werden nur im Erregungszustand eines Netzwerkes aus Nervenzellen gespeichert. Im Gegensatz dazu werden die Informationen in der physiologischen Struktur des Langzeitgedächtnisses in Form von neuen Verbindungen (Synapsen) zwischen Nervenzellen gespeichert (E. Kandel, 1996, S. 358f). Gelernte Gedächtnisinhalte sind an vielen verschiedenen Stellen des Gehirns gespeichert (W. Edelmann, 2000, S. 133ff). So befinden sich sprachliche Informationen in einem anderen Bereich als visuelle. Das bedeutet z.B., dass das Wissen über einen Gegenstand, etwa eine Blume, im Gehirn verteilt abgelegt wurde. Bei Bedarf werden diese vielen Einzelinformationen wie Form, Bezeichnung und Geruch wieder zusammen geführt. Konkret bedeutet dies, wenn eine Information in Form eines wahrnehmbaren Reizes auf eine Sinneszelle trifft, wird dieser als elektrischer Impuls an eine Nervenzelle (Neuron) weiterleitet. Dabei wird ein bestimmter Energiewert überschritten und die Nervenzelle gibt den Reiz über einen faserartigen Fortsatz (Axon) an ein oder mehrere Neuronen weiter, die ihn ihrerseits weiterleiten können. Die Information hinterlässt so also charakteristische Spuren. Durch häufiges Nachziehen dieser Spuren (Üben, Wiederholen) verstärken sich diese Verbindungen (Synapsen) zwischen den Zellen, es entsteht also ein bleibendes Muster (E. Kandel, 1996, S. 358f). Das Gehirn passt sich plastisch den neuen Anforderungen an. Auf der beschriebenen physiologischen Struktur „baut” die logische Struktur auf. Dabei werden Informationen im Gehirn in einer hierarchischen Struktur (Baum) abgelegt. Beim Lernvorgang ist es nun wichtig, neue Begriffe, Erklärungen und Modelle in der vorhandenen hierarchischen Struktur sinnvoll zu verankern. Durch ein einfaches Selbstexperiment kann jeder Rezipient ausprobieren, was damit gemeint ist: Wenn es z.B.

darum geht, möglichst viele Lebewesen aufzuzählen, kann man anfangen aufzuzählen und wird dabei feststellen, dass man relativ bald am Ende ist bzw. verfällt man unwillkürlich in ein anderes Muster, indem man ausgehend von Oberbegriffen beginnt immer feinere Differenzierungen zu finden. Als Beispiel dienen: Säugetiere – Huftiere – Paarhufer – Rinder – Schafe – Schweine usw. (W. Edelmann, 2000, S. 133ff).
Aus dieser Erkenntnis leitet sich ein einfacher Lerngrundsatz ab: Neues muss sinnvoll in vorhandenem Wissen verankert werden. Jean Piaget (1976) nennt diesen Vorgang Assimilation. Im Folgenden werden die Begriffe zu Lehren und Lernen in Anlehnung an die Arbeiten von Jean Piaget verwendet. Als Bezugspunkt dient mir dabei das Buch: Piaget, Jean (1976): Die Äquilibration der kognitiven Strukturen. 1. Aufl. Stuttgart: Klett (Konzepte der Humanwissenschaften).

3.6 Das Konzept der Assimilation

„Indem sich das Denken den Dingen anpasst,
strukturiert es sich selbst, und indem es sich selbst strukturiert,
strukturiert es auch die Dinge.“

Jean Piaget

Um zu erklären, was Assimilation bedeutet, ist es zunächst wichtig, einige Begriffe zu erläutern. Begonnen werden soll mit dem Begriffspaar „mechanisch – sinnvoll”, bei dem es um die begriffliche Verankerung der Informationen geht (W. Edelmann, 2000, S. 139ff).

mechanisch: Die Informationen werden wörtlich gelernt und nicht mit vorhandenem Wissen verknüpft.

sinnvoll: Die Informationen werden inhaltlich gelernt und mit dem vorhandenem Wissen verknüpft.

Beim Begriffspaar „rezeptiv – entdeckend” unterscheidet man, ob der Lernende aufgearbeitetes Wissen übernimmt oder selbst entdeckt .

rezeptiv: Die Informationen werden dem Lernenden in einer geschlossenen aufgearbeiteten Darstellung angeboten.

entdeckend: Durch geschickte Aufgabenstellung und den Einbau von Hinweisen entdeckt der Lernende die Sachverhalte und Informationen selbst.

Ein Lehrstoff kann potenziell sinnvoll sein und trotzdem mechanisch gelernt werden. Lernen beschreibt Ausubel (Ausubel nach W. Edelmann, 2000, S. 133ff) mit folgender Definition: „Es ist das Wesentliche eines sinnvollen Lernprozesses, dass symbolisch ausgedrückte Vorstellungen zufallsfrei und inhaltlich (nicht wörtlich) bezogen werden auf das, was der Lernende bereits weiß – nämlich auf einige bestehende relevante Aspekte seiner Wissensstruktur." „Sinnvolles Lernen" ist demnach durch folgende beiden Eigenschaften gekennzeichnet:

1. Es muss inhaltlich, d.h. nicht wörtlich gelernt werden.
2. Der neue Lernstoff muss zufallsfrei auf das bisherige Wissen bezogen werden.

Beispiele zur Assimilation

- Mechanisches rezeptives Lernen: Das Einprägen einer uns neuen Telefonnummer.
- Sinnvolles rezeptives Lernen: Der Lehrer behandelt das Thema „Kompass", die Schüler ordnen die neuen Erfahrungen in die zuvor gelernten Regeln über Magnetismus ein.
- Mechanisch entdeckendes Lernen ist: Ein Kind entdeckt zufällig, dass eine Kerzenflamme erlischt, wenn man ein Glas darüber stülpt. Es kann diesen Vorgang nicht begreifen, erfreut sich aber am erzielten Effekt.
- Sinnvolles entdeckendes Lernen: Schüler erhalten Landkarten, statistische Daten und andere Informationen und entdecken z.B. die Regel: „Jede größere Stadt bildet den kulturellen, verwaltungsmäßigen und wirtschaftlichen Mittelpunkt einer Region."

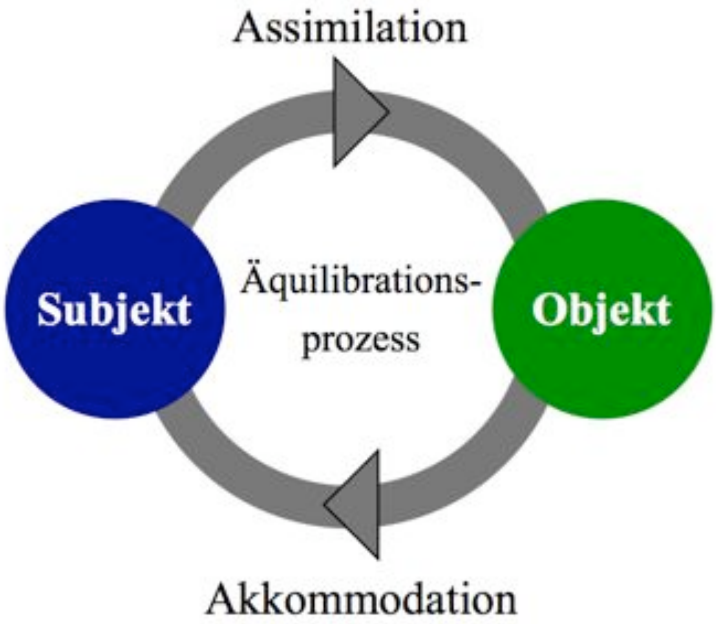


Abbildung 34: Der Äquilibrationsprozess, J. Piaget (1976) in „Die Äquilibration der kognitiven Strukturen“

Das wesentliche Ziel sinnvollen Lernens besteht also im Aufbau einer klar gegliederten kognitiven Struktur, bei der die einzelnen Wissenselemente nicht unverbunden nebeneinander stehen, sondern in einem System von aufeinander bezogenen Bedeutungen organisiert sind. Dabei sollen möglichst viele Bedeutungen so ökonomisch wie möglich gespeichert werden (W. Edelmann, 2000, S. 133ff). Neu zu lernende Ideen werden mit mehr als einer vorhandenen Bedeutung in Zusammenhang gebracht werden. So entsteht ein dichtes Netz von Verbindungen. Diese kognitive Struktur spiegelt das Resultat aller Lern-, Behaltens- und Vergessensprozesse wider. Nur so kann das Ziel erreicht werden, eine klar organisierte und stabile Wissensmenge im Gedächtnis zu verankern.

Herausforderungen beim Erlernen mathematischer Zusammenhänge

Mehr als in vielen anderen Bereichen der Wissensvermittlung ist es beim Erlernen mathematischer Zusammenhänge von elementarer Bedeutung, eine klar organisierte, hierarchische Wissensstruktur im Gedächtnis zu verankern. Um dies zu erreichen, stehen zwei (grundsätzliche) Lernmethoden zur Verfügung:

1. Das induktive Lernen, bei dem man, von einem speziellen Beispiel ausgehend, eine Vorstellung für den allgemeinen Zusammenhang bekommt.

2. Das deduktive Lernen, bei dem man vom allgemeinen auf ein einzelnes Beispiel schließt.

Im schulischen Mathematikunterricht steht deduktives Lernen im Vordergrund. Die Schüler bekommen von ihrem Lehrer Beispiele geboten, anhand derer ein Zusammenhang erklärt wird. Dieser Zusammenhang wird aber in den allerseltensten Fällen wirklich bewiesen oder allgemein begründet. Deshalb bleibt vielen Schülern die Allgemeingültigkeit vieler Zusammenhänge verborgen und es dominiert bei ihnen die Meinung, dass es möglich wäre, Mathematik mechanisch-rezeptiv zu lernen. Aber gerade bei mathematischen Zusammenhängen ist es nötig, dass nach einer ersten Problemlösephase in der an Beispielen gearbeitet wird, um eine Idee für einen Zusammenhang zu bekommen, der Versuch folgt diesen Zusammenhang schlüssig zu beweisen und bis ins Detail zu begründen. Danach ist es dringend erforderlich, durch weitere Beispiele und Folgerungen den Vorteil dieses Satzes plastisch zu manifestieren (V.Ulm, 2008, S. 25ff).

Multiple Repräsentation

Schon Johann Amos Comenius (1592 – 1670) wusste, dass es für effektives Lernen wichtig sei, vom reinen Wortunterricht abzukommen und Lerninhalte multipel zu repräsentieren. In seiner „Didactia magna” von 1657 schreibt der tschechische Pädagoge:

> „Wenn man nun eine Sprache lernt, die Muttersprache nicht ausgenommen, so müssen die Dinge, die mit den Wörtern bezeichnet werden sollen, gezeigt werden. Umgekehrt sollen die Schüler was sie sehen, hören, fühlen oder schmecken durch Worte ausdrücken lernen, so dass Sprache und Verständnis parallel sich entwickeln und ausgefeilt werden.”

Das didaktische Prinzip der Anschauung und ein Lernen mit allen Sinnen war auch ein Hauptansatzpunkt von Johann Heinrich Pestalozzi (1746 – 1827). Neuere Forschungen zeigen, dass eine parallele Form der Informationsaufnahme, Verarbeitung und Speicherung sich vorteilhaft auf den Lernprozess auswirkt. Bilder und verbale Prozesse stellen alternative Codierungssysteme für einen identischen Begriff oder eine Begriffskette (Regel) dar (W. Edelmann, 2000, S. 153ff).

Die Verknüpfung von bildlicher Vorstellung (rechte Hemisphäre) und verbaler Ausdrucksform (linke Hemisphäre) wird als multiple Repräsentation bezeichnet. Um die Vorteile der multiplen Repräsentation für die Lehre zu verdeutlichen, sei an dieser Stelle das Erlernen einer Fremdsprache genannt: Das rein analytische Erlernen von Regeln und Vokabeln führt in den seltensten Fällen zum exakten Beherrschen einer Fremdsprache, wohin gegen die Kombination mit der täglichen Anwendung in der Sprache, d.h. das Sprechen der Sprache in Alltagssituationen, zu schnellem Verständnis und sicherem Beherrschen der Sprache führt (M.Arnold, 2002, S.109).

Je konkreter oder dinghafter eine Information ist, desto eher findet eine bildhafte Codierung (Vorstellung, nicht-sprachliches Denken) statt. Bildhafte Prozesse können zur Lösung abstrakter Probleme genauso beitragen wie zur Lösung konkreter Probleme. Die rechte Gehirnhälfte verarbeitet die bildlichen Informationen simultan (gleichzeitig). Dies bedeutet, dass eingehende Informationen in einem Schritt verarbeitet werden und nur das Ergebnis dieser Verarbeitung über die linke Gehirnhälfte ins Bewusstsein tritt. Das gilt nicht nur für bildliche Informationen, sondern für alles, was unter dem Begriff „Intuition“ verstanden wird (M.Arnold, 2002, S. 133ff). Die Ergebnisse dieser Intuition werden im Nachhinein von der linken Gehirnhälfte analytisch begründet. Als typisches Beispiel hierfür sei folgendes Exempel genannt: Menschen werden spontan in „sympathisch“ und „unsympathisch“ eingeteilt. Wenn man einen Menschen zum ersten Mal trifft, stellt man Sympathie bzw. Antipathie innerhalb einer sehr kurzen Zeitspanne fest. Nachträglich versucht man dieses (Vor-)Urteil analytisch zu begründen. Dies führt zum Hauptproblem der simultanen Verarbeitung – das Ergebnis kann, muss aber nicht zwangsläufig richtig sein. Aus eigener Erfahrung kennt der Leser sicher Fälle, bei denen sich auf den ersten Blick unsympathische Menschen nach näherer Betrachtung als sympathisch herausstellen. Dieses Phänomen, auf die Mathematik übertragen, hat Henri Poincaré (H. Poincaré, 1921, S. 4) mit folgendem Satz gemeint: „Es ist die Logik, mit der man beweist; es ist die Intuition mit der man erfindet.”

Multiple Repräsentation bei der Explorativen Datenanalyse

Das zentrale Werkzeug für die Explorative Datenanalyse ist die interaktive statistische Graphik. Aus psychologischer Sicht geht es bei der interaktiven statistischen Graphik, neben dem mathematischen Analysieren des Sachverhalts, auch darum, geometrische

Muster zu erkennen, Bilder simultan zu verarbeiten und ein „Gefühl” für die Daten zu entwickeln. Die Konsequenz aus dieser Erkenntnis ist, dass in der Lehre der Explorativen Datenanalyse beide Gehirnhälften gefördert werden müssen und das Ziel sein muss, die Potenziale der Beiden sinnvoll zu verknüpfen (M.Arnold, 2002, S. 109). Wie bereits beschrieben, liegt es in der Natur der Sache, dass die rechte Gehirnhälfte nicht über den Umweg einer Verbalisierung zu trainieren ist, sondern nur über das konkrete „Denken in Bildern”.

Aus dieser Feststellung erwachsen folgende Forderungen für die Lehre der Explorativen Datenanalyse:

1. Der Lernende benötigt zunächst ein fundiertes Wissen in den Bereichen Wahrscheinlichkeitstheorie und deskriptive Statistik.
2. Der Lernende muss an konkreten Beispielen und an echten Datensätzen üben.
3. Der Lernende muss die interaktive Graphik selbst anwenden.
4. Der Lernende muss die Ergebnisse seiner Exploration immer einer kritischen Analyse durch induktive statistische Verfahren unterziehen.

Um dem Lernen des Explorativen Datenanalyse näher zu kommen soll im nächsten Abschnitt näher untersucht werden, von welcher „Natur” ein Lehr-Lernsystem ist, in dem das eigen- und selbständige Lernen im Mittelpunkt steht. Der Begriff des Systems meint in diesem Zusammenhang den mathematischen Begriff des Systems und nicht pädagogische oder psychologische Systembegriffe.

3.7 Lehr-Lernsysteme als dynamische Systeme

„Alles Urdenken geschieht in Bildern:
darum ist die Phantasie ein so notwendiges Werkzeug desselben,
und werden phantasielose Köpfe nie etwas Großes leisten, es sei denn in der Mathematik."

Arthur Schopenhauer

Lernen ist wohl der komplexeste Vorgang, den das menschliche Gehirn hervorbringt, denn es umfasst das Speichern von Informationen, das Kombinieren von Daten zu völlig neuen Zusammenhängen, das Reflektieren von Gelerntem und schließlich das Einordnen des Gelernten in das eigene Sein bis hin zum Erforschen des biologischen Systems „Gehirn", welche das eigene Bewusstsein hervorgebracht hat. Die Lehr-Lernforschung steht nun vor dem Problem, diese komplexen Vorgänge zu analysieren, Modelle zu planen und aus jenen optimale Rahmenbedingungen und Methoden zu entwickeln, die dem Lernenden die Chance geben, möglichst effektiv und mit Freude und Lust zu lernen. Diese Überlegungen machen es fast unmöglich, das Lernen quantitativ zu erforschen und mit wenigen Variablen zu beschreiben. Der einzig sinnvolle Weg ist, aus der Erfahrung neue Methoden des Lehrens zu entwickeln. Dieses Vorgehen ist im besten Sinne des Wortes „empirisch" (griechisch: „auf Erfahrung beruhend"). Forschungsmethoden und mathematische Verfahren, die der Komplexität des Lernvorgangs gerecht werden, müssen allein auf die Forschungsfrage und den Beobachtungsgegenstand konzentriert sein, auch wenn es sich um komplexe Vorgänge handelt. Implizite Festlegungen durch die Wahl von Forschungsverfahren müssen dem Forscher bewusst und sachlogisch begründbar sein. Dazu gehört etwa die Frage, ob ein Problem diskret (z. B. eine Notenskala) oder stetig (z. B. die Zeitskala in einer Unterrichtsstunde) ist. Der Forscher muss sich darüber im Klaren sein, in welchem mathematischen Raum sein Modell „lebt". Es ist zu unterscheiden, ob es sich um einen chaotischen Vorgang, einen diskontinuierlichen Vorgang oder um einen einfachen linearen Zusammenhang handelt, der über die Grenzen der Untersuchung hinaus extrapoliert werden kann. Die gewählten (beobachteten) Größen, welche den Lernvorgang beschreiben sollen, müssen diesen Vorgang möglichst exakt und vollständig beschreiben. Eine „Stückelung” in Teilsysteme, deren Verhalten man nicht genau kennt und die das Gesamtsystem nicht disjunkt zerlegen, führt dazu, dass Aussagen über die Teilsysteme keine Aussagekraft für das Gesamtsystem haben und wenig zur Theoriebildung beitragen (U.Fahrner & A.Unwin, 2008, S. 181ff).

Abbildung eines Lehr-Lernsystems in einen Datensatz

Der Dreh- und Angelpunkt einer jeden empirischen Untersuchung ist die Abbildung eines realen Lehr-Lern-Phänomens mittels Datenerhebung in der Sprache der Mathematik. Die Realität wird mit der Abbildung in einen Datensatz in mathematische Zusammenhänge übersetzt; diese werden mit Hilfe von mathematischen Methoden ausgewertet, neu kombiniert und zu neuen Aussagen synthetisiert; diese müssen in die reale Welt „rückübersetzt“, d.h. interpretiert werden. Erst durch diese Interpretation wird es möglich, Aussagen einer realen Lehr-Lern-Situation zu untermauern (U.Fahrner & A.Unwin, 2008, S. 181ff). Eine zutreffende Interpretation von mathematischen Aussagen ist nur möglich, wenn ganz zu Beginn der Forschungsarbeiten die Abbildung in den Datensatz und die dabei gemachten impliziten Fehler klar sind und für diese maximale Grenzen angegeben werden können. Schränkt man sich hier am Anfang zu stark ein, indem man z.B. eine zu kleine Klasse mathematischer Beschreibungsformen wählt, werden unter Umständen die entscheidenden Größen von der Beschreibung ausgeschlossen und führen so zu einem aussagelosen Modell.

Lernen ist – wie der wissenschaftliche Erkenntnisprozess – kein kontinuierlicher Vorgang, sondern von Sprüngen gekennzeichnet z.B. zwischen diskreten „Aha”-Erlebnissen, die nicht durch stetige Variablen dargestellt werden können. Daher ist der Lern- und Erkenntnisprozess dem Grunde nach ein chaotischer, diskontinuierlicher Prozess. Dies entspricht der Erfahrung vieler Lehrender in Schulen und an den Universitäten. Der Lernprozess ist mit einfachen, linearen, stetigen mathematischen Modellen nur in wenigen Randbereichen ansatzweise beschreibbar. Mathematische Verfahren, die solche Vorgänge beschreiben können, wurden ab den 1960 Jahren entwickelt. Alle Daten müssen so im Modell verwendet werden, wie sie erhoben wurden, d.h. diskrete Variablen bleiben diskret, stetige Variablen bleiben stetig, beispielsweise müssen Fragen in Fragebögen mit Auswahlantworten immer diskret ausgewertet werden. Die in vielen empirischen Untersuchungen gemachten Vereinfachungen dienen oft nur der Rechenvereinfachung und sind nicht aus dem Problem heraus begründet. Sie entziehen der Untersuchung viel von ihrer wissenschaftlichen Aussagekraft. Neben den rein mathematischen Werkzeugen zur Beschreibung des Lehr-Lernprozesses ist die wissenschaftstheoretische Herangehensweise im Umbruch begriffen. Durch Studien wie PISA werden Lernszenarien in verschiedenen Ländern für Politik und Gesellschaft vergleichbar. Bildungspolitiker begnügen sich immer weniger damit, von der Forschung zu erfahren, welche Lernszenarien nicht funktionieren; sie erwarten Antworten auf drängende Fragen einer Wissensgesellschaft geben. Viele Forscher im Lehr-Lernbereich nehmen diese Herausforderung an und orientieren sich bei der Konstruktion von konkreten Lernszenarien an der Herangehensweise der Ingenieurwissenschaften. Ein konkretes Lehr-Lernszenario kann nur durch eine sehr große Anzahl von Dimensionen (im mathematischen Sinn) eindeutig charakterisiert werden. Diese Anzahl von Dimensionen übersteigt in einer realen Situation jede Anzahl von Variablen, die erhoben und in einem konventionellen statistischen Modell bearbeitet werden können. Dies liegt zum Teil daran, dass man nicht alle Variablen kennt, dass von Anfang an nicht klar ist, welche Variablen die gestellte Forschungsfrage beantworten können, oder dass keine geeigneten Werkzeuge zur Messung dieser Variablen zur Verfügung stehen. Erschwerend kommt hinzu, dass durch den Messvorgang selbst und durch den Einsatz bestimmter Messwerkzeuge die Observablen (also die beobachteten Variablen) verfälscht werden können und der Messvorgang für die Observablen nicht immer Teil der Theoriebildung ist (vgl. Hawthorne-Effekt; Mayo, 1933). Vor ähnlichen Problemen stand die Physik zu Beginn des 20. Jahrhunderts: Damals musste sich die Physik von dem unbeteiligten Beobachter, der keinen Einfluss auf das Messergebnis hatte, und von deterministischen Erklärungsmodellen im atomaren Bereich trennen. Erst durch das Integrieren des Messvorgangs in die Theorie und die Wahl einer wesentlich größeren Klasse an statistisch mathematischen Modellen war es möglich, die Quantenmechanik zu entwickeln und bis dahin nicht erklärbare Phänomene zu beschreiben sowie für diese konkrete Voraussagen zu treffen. Startpunkt für das Entwickeln neuer Modelle und Beschreibungsformen für den Lernvorgang ist eine genaue Analyse von komplexen Datensätzen mit einer großen Zahl von Variablen, welche die Komplexität des Lernvorgangs beschreiben.

Grundlagen dynamischer Systeme

Ein dynamisches System ist ein sehr allgemeines mathematisches Konstrukt (B. Aulbach, 2004, S. 169ff). Dies geht von einem System aus, was heißt, dass sich zumindest prinzipiell eine Grenze zwischen einem Innen und einem Außen ziehen lässt. Des Weiteren geht es um die dynamische Veränderung des Systems, was heißt, dass es sich während einer Zeit t von einem Zustand in einen anderen Zustand verändert. Wenn man diese Beschreibung mathematisch etwas präzisiert, besteht ein dynamisches System aus drei Teilen. Der erste Teil zeigt einen Zustandsraum X, der das System beschreiben kann. Dieser kann von sehr allgemeiner Natur sein, zum Beispiel der uns umgebende Raum, ein unendlichdimensionaler Vektorraum, ein Funktionenraum oder eine andere topologische Struktur. Der zweite Teil zeigt eine Zeitmenge T. Diese kann, je nach Struktur des zu behandelnden Problems, entweder kontinuierlich (d.h. $T = \mathbb{R}$) oder diskret (d.h. $T = \mathbb{Z}$) sein. Drittens zeigt sich eine Familie von einparametrigen Abbildungen $\varphi_t = X \to X, t \in T$ mit folgenden beiden Eigenschaften:

(1) $\varphi_0 = id_X$

(2) $\varphi_t \circ \varphi_s = \varphi_{t+s} \quad \textit{für alle } t, s \notin T$

Je nach Art der Zeitmenge $\mathbb{R}$ oder $\mathbb{Z}$ benennt man die Systeme als kontinuierliche oder diskrete dynamische Systeme.

Das mathematische Konstrukt der dynamischen Systeme eignet sich dazu, chaotische und unübersichtliche Naturphänomene oder Sozialphänomene in bestimmten Teilbereichen zu beschreiben. Im Bereich der Physik lassen sich damit Eigenschaften von Strömungen oder chaotische Bewegungen beschreiben. In der Biologie wendet man das Verfahren bei Evolutionsphänomenen und genetischen Sachverhalten an. Für die Kognitionswissenschaft ist das Konzept der dynamischen Systemen im Umgang mit neuronalen Netzen nützlich. Die Vielzahl der Möglichkeiten der Anwendung dynamischer Systeme zeigt, neben der Universalität des Konzepts, aber auch, dass die daraus entwickelten Modelle entweder sehr elementare Zusammenhänge aufzeigen oder noch viel präziser ausdifferenziert werden müssen, um konkrete Aussagen im jeweiligen Forschungsbereich treffen zu können.

3.8 Äquilibration als dynamisches System

„Erfahrung ist eine teure Schule,
aber Narren wollen anderswo nicht lernen.“

Benjamin Franklin

Um mit Hilfe des Konzeptes “dynamisches System” den Äquilibrationsprozess zu bearbeiten, ist es hilfreich, sich mit einem weiteren mathematischen Konzept, dem der Maschine, auseinander zu setzen.
Unter einer trivialen Maschine versteht man einen Operator, der einen Eingangszustand X mit einem Ausgangszustand Y in eindeutiger Weise verbindet (vgl. H. von Foerster, 2008). Charakteristisch für eine triviale Maschine ist, dass die Zuordnung zwischen Input und Output über die Zeit konstant ist. Bei einer trivialen Maschine ist es also immer möglich vorauszusagen, was die Maschine auf einen bestimmten Input antwortet. Die meisten uns umgebenden Maschinen, z.B. ein Auto, ein Computer oder eine Kaffeemaschine, sind triviale Maschinen: Auf den Input, „ich drehe das Lenkrad nach links“ fährt das Auto nach links. Es kommt nie vor, dass eine triviale Maschine einen unvorhergesehenen Output liefert. Selbst das Ergebnis komplexer Computerprogramme ist immer voraussehbar und eindeutig definiert. Die Maschine führt sozusagen kein Eigenleben und passt sich nicht unvorhergesehen, neuen Gegebenheiten an. Um die Eigenschaften einer trivialen Maschine genauer zu untersuchen, betrachten wir folgendes Beispiel: Gegeben sei eine triviale Maschine mit den Eingangszuständen A, B, C, D. Diese sind die Ursachen für die Ausgangszustände 1, 2, 3, 4. Wenn also am Eingang unserer trivialen Maschine A anliegt, gibt sie 1 zurück, wenn B anliegt, gibt sie 2 zurück, wenn C anliegt, gibt sie 3 zurück, und wenn D anliegt, gibt sie schließlich 4 zurück. Das heißt, geben wir an den Eingang eine Folge BADCAD, so erhält man immer 214314. An unserer trivialen Maschine kann man die Haupteigenschaften einer trivialen Maschine ablesen: Sie ist synthetisch determiniert, analytisch determinierbar, vergangenheitsunabhängig und dadurch voraussagbar.

Komplexe biologische Systeme, z.B. Tiere und Menschen, die eine kognitive Struktur ihr Eigen nennen, können nicht als triviale Maschine beschrieben werden. Das liegt daran, dass diese nicht-trivialen Maschinen ihre Zuordnungen und Verknüpfungen über die

Zeit ändern. Nicht-triviale Maschinen ändern ihre inneren Zustände. Ein Beispiel: Wenn ein Lehrer einen Schüler fragt: „Wie viel ist zwei mal drei”, kann der Schüler verschiedene Antworten geben. Er kann z.B. die vom Lehrer erwartete Antwort, „sechs” geben. Und sich dann wie eine triviale Maschine verhalten. Oder er kann antworten: „Zwei mal drei ist das gleiche wie drei mal zwei”, und kann diese Feststellung auf höchst unterschiedliche Weise begründen. Er kann aber auch Antworten geben, die völlig außerhalb des vorhersagbaren Bereichs des Lehrers liegen, z.B.: „In fünf Minuten ist Pause”! Mathematisch ist eine nicht-triviale Maschine dadurch modellierbar, dass man neben der Zuordnung zwischen Input und Output noch eine zeitliche Zustandsänderung der Maschine in Betracht zieht.
Das bedeutet, wenn wir unsere triviale Maschine von oben, die immer A auf 1, B auf 2, C auf 3 und D auf 4 zuordnet, um eine Variable Z erweitern, die einen inneren Zustand darstellen und sich in Abhängigkeit der Eingaben ändern soll, entsteht aus der trivialen eine nicht-triviale Maschine. Nehmen wir der Einfachheit halber an, es gäbe nur zwei innere Zustände, einen Zustand I und einen Zustand II, die in nachfolgender Tabelle zugeordnet werden:

Tabelle 4: Zuordnungstabelle für die Zustände einer einfachen nicht trivialen Maschine

Maschine im Zustand I			**Maschine im Zustand II**		
x	y	z	x	y	z
A	1	I	A	4	I
B	2	II	B	3	I
C	3	I	C	2	II
D	4	II	D	1	II

Für einen Betrachter von außen, der unsere, zugegebener Maßen einfache, nicht-triviale Maschine untersuchen will, sehr schwierig, die Maschine und ihre inneren Zustände zu beschreiben. Stellt man sich einen Beobachter vor, der die nicht-triviale Maschine von außen betrachtet: Ist die Maschine im inneren Zustand I, den der Beobachter weder sehen noch messen kann, wird er auf einen Input B einen Output 2 bekommen. Wird er die Messung wiederholen, dann hat sich die Maschine von einem Zustand I in einen Zustand II verändert und der Beobachter erhält von seinem gleichen Input B nun den Output 3. Wenn man obige Tabellen betrachtet, hat man den Eindruck, dass dieses Problem sehr übersichtlich und nicht besonders schwer zu lösen ist. Dieser intuitive Eindruck trügt, denn schätzt man die Anzahl der möglichen nicht-trivialen Maschinen, die durch Experimente bestimmt werden sollen, in Abhängigkeit der Eingangs- und Ausgangssymbole ab, so erhält man bei nur zwei Eingangssymbolen, schon mögliche 2^{16} = 65536, bei vier Symbolen 2^{8192}=10^{2466}, bei acht Symbolen sind es schon $2^{3 \cdot 2^{30}} = 10^{969685486}$ mögliche Maschinen die nur durch eine gleiche Anzahl von Versuchen unterschieden werden können (H. von Foerster, 2008, S. 65). Das bedeutet, wenn ein Forscher die konstruierte, nicht-triviale Maschine findet und von der besprochenen Versuchsmaschine nichts weiß, dann muss der Forscher bei vier Eingangssignalen, wie gerade berechnet, 10^{2466} Experimente durchführen, um diese einfache, nicht-triviale Maschine zu bestimmen. Um diese Zahl einordnen zu können, muss man präsent haben, dass das Universum ca. $3 \cdot 10^{20}$ Sekunden alt ist. Wenn also ein Forscher für einen Versuch (Inputzeichen abschicken, Outputzeichen ablesen) eine Sekunde benötigt, dann ist unsere nicht-triviale Maschine, die nur aus vier Eingangssignalen und vier Ausgangssignalen besteht, nicht während der ganzen Existenz unseres Universums bestimmbar, denn $3 \cdot 10^{20}$ ist kleiner als 10^{2466}. Diese Beobachtung hat Arthur Gill (1962) in seinem Unbestimmbarkeitsprinzip zusammengefasst und bewiesen. Damit ist auf recht einfache Weise gezeigt, dass ein bloßes Betrachten von Eingangs- und Ausgangszuständen für die Untersuchung von kognitiven Systemen ungeeignet und sogar völlig aussichtslos ist. Aus diesem Beispiel kann man die Grundeigenschaften von nicht-trivialen Maschinen ableiten. Nicht-triviale Maschinen sind synthetisch determiniert, analytisch unbestimmbar, vergangenheitsabhängig und unvoraussagbar.

Dieser Exkurs in das Konzept der nicht trivialen Maschinen zeigt uns zwei Dinge:

1. Die Untersuchung von kognitiven Systemen mit Hilfe einer einfachen In- und Outputbetrachtung kann nur in Grenzbereichen erfolgreich sein, wenn man davon ausgeht, dass das Konzept der nicht-trivialen Maschine eine Untergrenze für die Komplexität und zeitliche Veränderungen eines kognitiven Systems darstellt.

2. Die Verständigung, zwischen zwei Menschen oder der Umgang eines Menschen mit seiner Umgebung muss wesentlich „intelligenter" organisiert sein als eine nicht-triviale oder eine triviale Maschine.

Das einfachste und noch Erfolg versprechende Konzept ist das der Äquilibration im Sinne von J. Piaget (H. von Foerster, 2008, S. 65). Im Gegensatz zur konstruierten nicht-trivialen Maschine mit vier Ein-Ausgangssignalen und zwei inneren Zuständen, die in zwei Tabellen festgelegt sind, ist der Prozess der Äquilibration, der zur Anpassung eines kognitiven Systems führt, durch seine Eigenschaften als Regelkreis gekennzeichnet. Startet man mit einer beliebigen vorgegebenen kognitiven Struktur (z.B. durch biologische oder genetische Grenzen), passt sich diese über den Prozess der Assimilation und Akkommodation an neue Umgebungsgegebenheiten an (H. Maturana, 1987, S. 196).

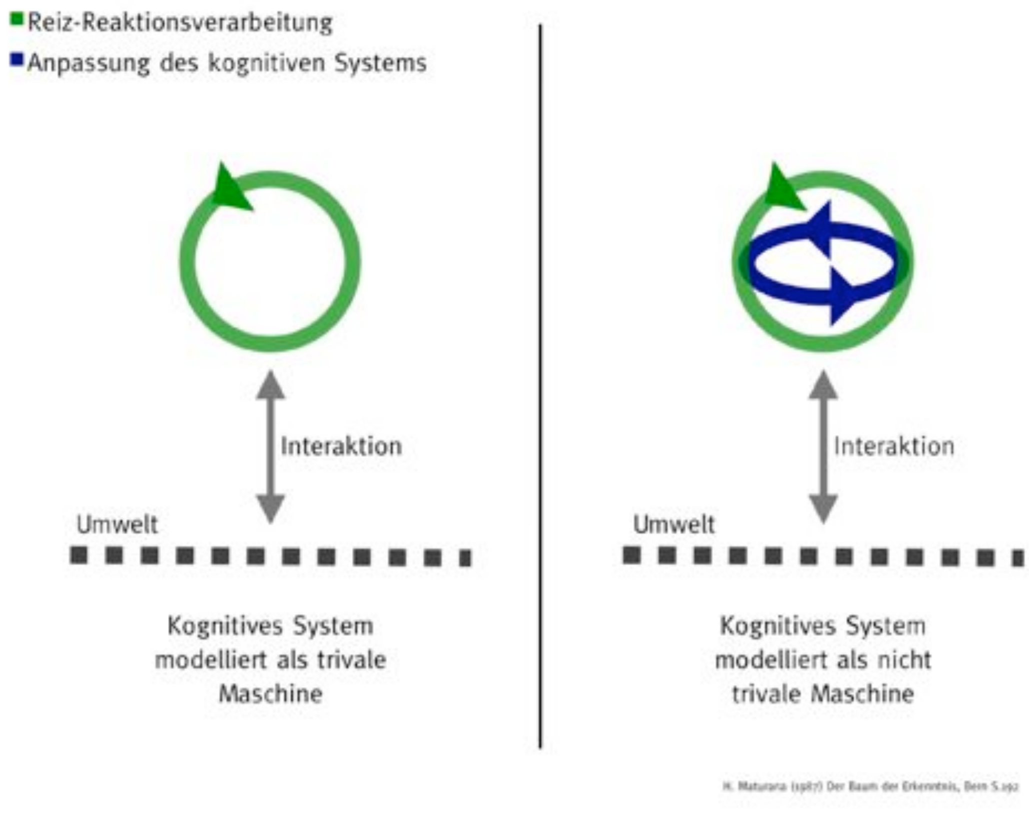


Abbildung 35: Organisation eines kognitiven Systems (H. Maturana (1987) Der Baum der Erkenntnis, Bern S.192)

Bei dieser Betrachtung gibt sowohl der von Piaget beschriebene Äquilibrations-Prozess als auch die trivialen Maschine im mathematischen Sinn eine Untergrenze für die Komplexität des Lernprozesses an. An dieser Stelle soll der Äquilibrations-Prozess an der einfachen Wechselwirkung zwischen einem Subjekt und einem Objekt erläutert werden. Piaget (1967) geht von einem Gleichgewicht zwischen kognitiver Struktur und Umwelt und von einem geschlossenen Regelkreis zwischen einem Subjekt und seiner Umwelt aus. Ein Individuum, das seine Umwelt und die darin enthaltenem Objekte untersucht, hat zwei Möglichkeiten mit Neuem umzugehen:

1. Die Assimilation, was bedeutet, dass das Individuum die neuen Umwelteindrücke an seine schon vorhandenen kognitiven Strukturen anpasst und sie damit vollständig erklären kann.
2. Die Anpassung der Umwelteindrücke an die kognitive Struktur gelingt nicht, was zur Folge hat, dass das Individuum versucht, das dadurch ausgelöste Ungleichgewicht auszugleichen. Dies geschieht durch Akkommodation, indem die kognitiven Strukturen an die neuen Umwelteindrücke angepasst werden.

Aus diesen beiden Handlungsoptionen lässt sich nun folgender Regelkreis konstruieren: Aus einem neuen Sinneseindruck folgt der Wunsch nach Assimilation. Gelingt dies nicht, wird das kognitive System mittels Akkommodation angepasst und der neue Eindruck mit der angepassten kognitiven Struktur in Einklang gebracht. Bei der Vielzahl möglicher Eindrücke, die ein Individuum zu verarbeiten hat, geht dies nicht in einem Schritt, sondern das kognitive System passt sich über mehrere Iterationsschritte den veränderten Konditionen an (siehe J. Piaget, 1976).

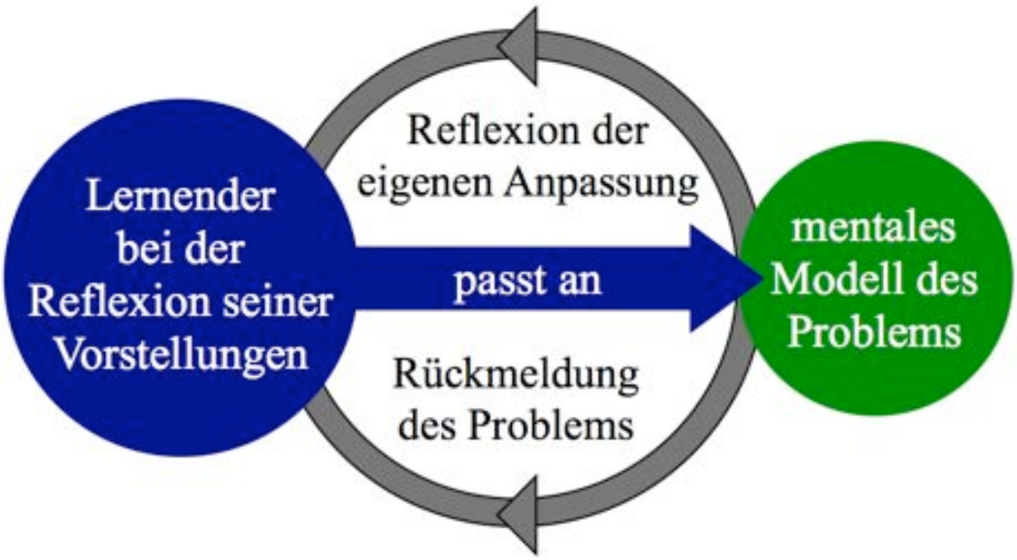


Abbildung 36: Akkommodation bei einer explorativen Datenanalyse

Der Prozess der Anpassung lässt sich mit Hilfe eines Operators als dynamisches System mathematisch beschreiben. Abbildung 37 zeigt die Interaktion zwischen einem Subjekt und einem Objekt. Obs.S bezeichnet die Observable im Bezug auf die Aktion des Subjekts. Unter Obs.O versteht man die auf das Objekt bezogene Observable. Für die elementare mathematische Betrachtung kann man die Sequenz Obs.O → Obs.S zu Obs. zusammenfassen. Koord.S bezeichnet die Koordinationen der Aktionen oder Operationen des Subjekts und analog dazu bezeichnet Koord.O die folgenden Koordinationen zwischen Objekten, die physikalischer kausaler Natur sind und den Objekten zugeschrieben werden können. Die Sequenz aus Koord.S → Koord.O fassen wir zu dem Operator Koord. zusammen, der den Bedingungen einer Abbildung eines dynamischen Systems genügt (H. von Foerster, 2003, S. 261ff).

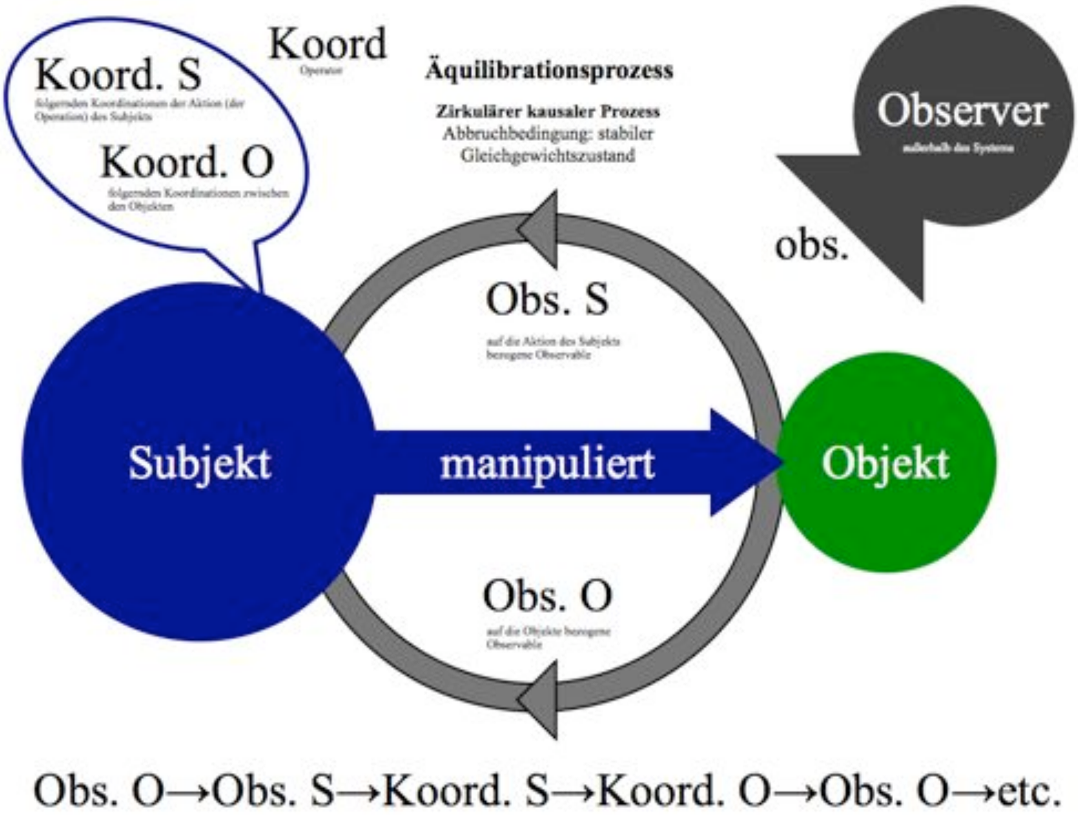
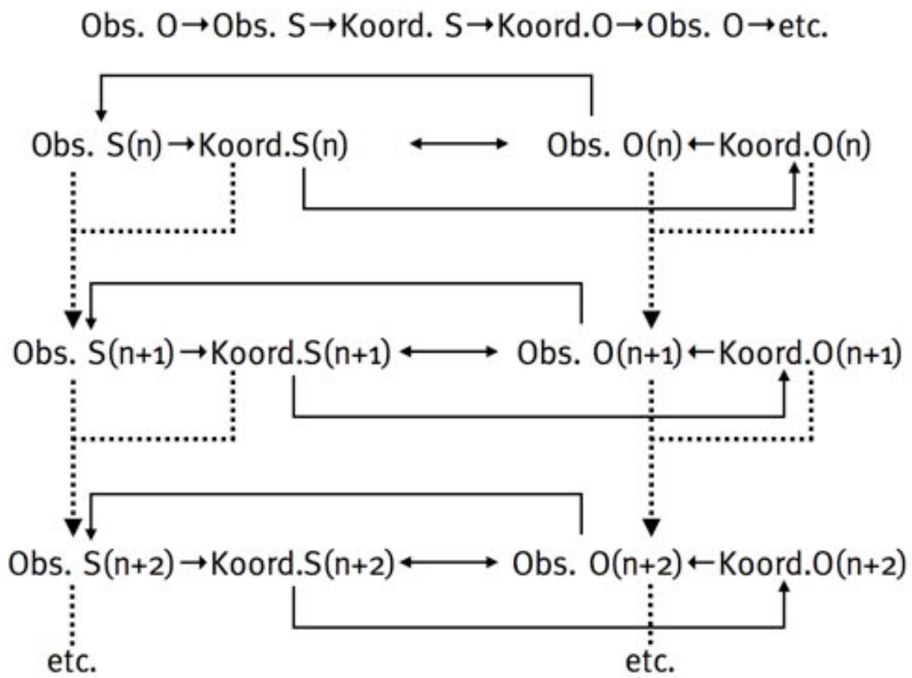
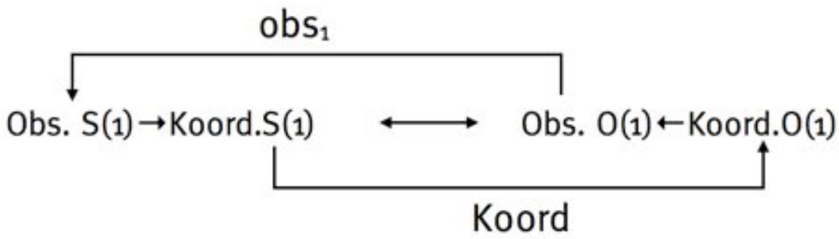


Abbildung 37: Der Äquilibrationsprozess (J. Piaget (1976)Die Äquilibration der kognitiven Strukturen, Stuttgart; H. Foerster (2003), Objects: Tokens von (Eigen-)behavior in Understanding Understanding: Essays on Cybernetics and Cognition)



Beobachtet man den Regelkreis über mehrere Iterationsschritte, so folgt daraus, dass die Sequenz Obs.S(n) → Koord.S(n) den Ausgangspunkt für Obs.S(n+1) und Koord.O(n) → Obs.O(n) der Eingangswert für Obs.O(n+1) ist. Diese Iterationsgleichung lässt sich so lange weiterführen, bis sich schließlich ein stabiles Verhalten einstellt und die innere kognitive Struktur des Individuums an die äußere Problemstellung angepasst ist (J.Piaget, 1976, S. 48ff).



$$\Rightarrow \quad obs_2 = Koord(obs_1) = Koord(Koord(obs_0))$$

$$\Rightarrow \quad obs_n = \underbrace{Koord(Koord(Koord(....(obs_0))}_{n\ mal}$$

$$\Rightarrow \quad obs_n = Koord^{(n)}(obs_0))$$

$$\Rightarrow \quad obs_\infty = \lim_{n\to\infty} Koord^{(n)}(obs_0))$$

Die Beobachtung, dass sich ein stabiles Verhalten einstellt, ist der Ausgangspunkt für die mathematische Betrachtung des Äquilibrations-Prozesses. Obwohl unbekannt ist, ob der Operator Koord. eine Abbildung ist, die den Bedingungen für dynamische Systeme gehorcht, d.h.

(1) $\text{Koord.}_o = \text{id}_{obs.}$ und

(2) $\text{Koord.}_m \circ \text{Koord.}_n = \text{Koord.}_{m+n}$

können einige qualitative Aussagen über das System getroffen werden:
Das sich einstellende stabile Verhalten und das damit verbundene Wissen oder die Kompetenz werden durch zusätzliche Iterationsschritte nicht weiter verändert. Daraus folgt, dass das System einen Fixpunkt oder, wie Heinz von Foerster (2003, S.263 f) schreibt einen Zustand des „Eigenbehaviors", erreicht hat. Dieses Eigenbehavior ist eine elementare Eigenschaft der kognitiven Struktur des betrachteten Individuums. Setzt man weiterhin voraus, dass sich das Verhalten des Individuums eines endlichen Zustandsraums beschreiben lässt, spannen alle Eigenbehaviors die kognitive Struktur auf und bilden in natürlicher Weise ein Koordinatensystem.

Äquilibration als Lernmodell der Explorativen Datenanalyse

Fragestellungen in der Explorativen Datenanalyse sind „offen" gestellte Fragen, d.h. es gibt für ein- und dieselbe Frage oft mehre Lösungen und Lösungswege. Um zu verdeutlichen, was in diesem Zusammenhang „offen" bedeutet, hier ein Beispiel aus der Schulalgebra *(V. Ulm, 2008, S. 41)*:

Eine geschlossen gestellte Frage ist zum Beispiel:

- Löse die Gleichung: 7x-11=24

Offen gestellte analoge Fragestellungen sind:

- Stelle einige Gleichungen mit x=5 auf.
- Stelle Exponenxialgleichungen mit der Lösung x=5 auf
- Stelle quadratische Gleichungen mit den Lösungen 1 und 5 auf. Beschreibe alle hierzu möglichen quadratische Gleichungen.
- Erfinde zur Gleichung 7x-11=24 eine Textaufgabe
- usw.

Geschlossen gestellte Fragestellungen sind dadurch charakterisiert, das sie stark „didaktisiert" sind und im negativen Sinn „Schulbuch-Charakter" besitzen. In ihnen kommt nur ein Thema vor, und es gibt eine einfach abprüfbare Lösung. Zum Lösen solcher Aufgaben wird meistens bloße Rechenfertigkeit benötigt, mathematische Kompetenzen treten in den Hintergrund. Sie sind gut geeignet, Routinen einzuschleifen und zu festigen.

Offen gestellte Fragestellungen sind meist „reale" Fragestellungen. Sie müssen kreativ bearbeitet werden und verlangen ein grosses Maß an mathematischer Kompetenz. Sie schaffen Motivation beim Lernenden und erfordern intensive Auseinandersetzung mit den Inhalten der jeweiligen Wissensgebiete. Bei der Explorativen Datenanalyse wird vom Datenanalysten erwartet, das er flexibel auf Fragestellungen reagieren kann und reale Datensätze, die praktisch nie idealen Modellvorstellungen der mathematischen Statistik entsprechen, einordnen kann. Wie in Kapitel 2 beschrieben, ist ein Hauptcharakteristikum der interaktiven statistischen Graphik (ISG) der hohe Grad an Interaktionsmöglicheiten zwischen den Daten und den Lernenden. So werden die Daten über Plots, die an das visuelle System des Menschen angepasst sind und vom Nutzer auf seine Bedürfnisse angepasst werden können, dargestellt und der Nutzer kann die Daten manipulieren und mit ihnen „spielen". D.h., der Nutzer der interaktiven statistischen Graphik kann mit abstrakten Datensätzen so umgehen wie mit ihm umgebenden Objekten. Sammelt ein Lernender mit Hilfe der interaktiven statistischen Graphik Wissen über reale Datensätze, so besteht die Hoffnung, dass sich bei ihm kognitive Strukturen manifestieren, die ihm zu einem größeren Verständnis von Statistik und Datenanalyse verhelfen.

Abbildung 38: Offen gestellte Fragestellungen fördern eigene Lernwege

Wie soll nun dieses Modell im Detail funktionieren?

Bearbeitet ein Lernender mit Hilfe der interaktiven statistischen Graphik einen Datensatz, so bekommt er ständig Rückmeldungen von der Statistiksoftware. Diese visuellen Eindrücke sind an seine Wahrnehmung angepasst und durch die Interaktionsmöglichkeit kann der Lernende die Darstellung in bestimmten Grenzen an seine Bedürfnisse anpassen. Der Lernende wird zunächst als Problemlöser versuchen, die Rückmeldungen der Software in Einklang mit seinen Vorstellungen und Konzepten zu bringen. Dies wird nicht immer gelingen und er erfährt relativ direkt Rückmeldungen vom System, die sich mit seiner Vorstellungswelt nicht in Einklang bringen lassen. Da es sich bei dem zu bearbeitenden Datensatz um eine Teilinformation eines realen Problems handelt, kann der Lernende versuchen, diese Widersprüche entweder innerhalb oder außerhalb der Statistik zu lösen.

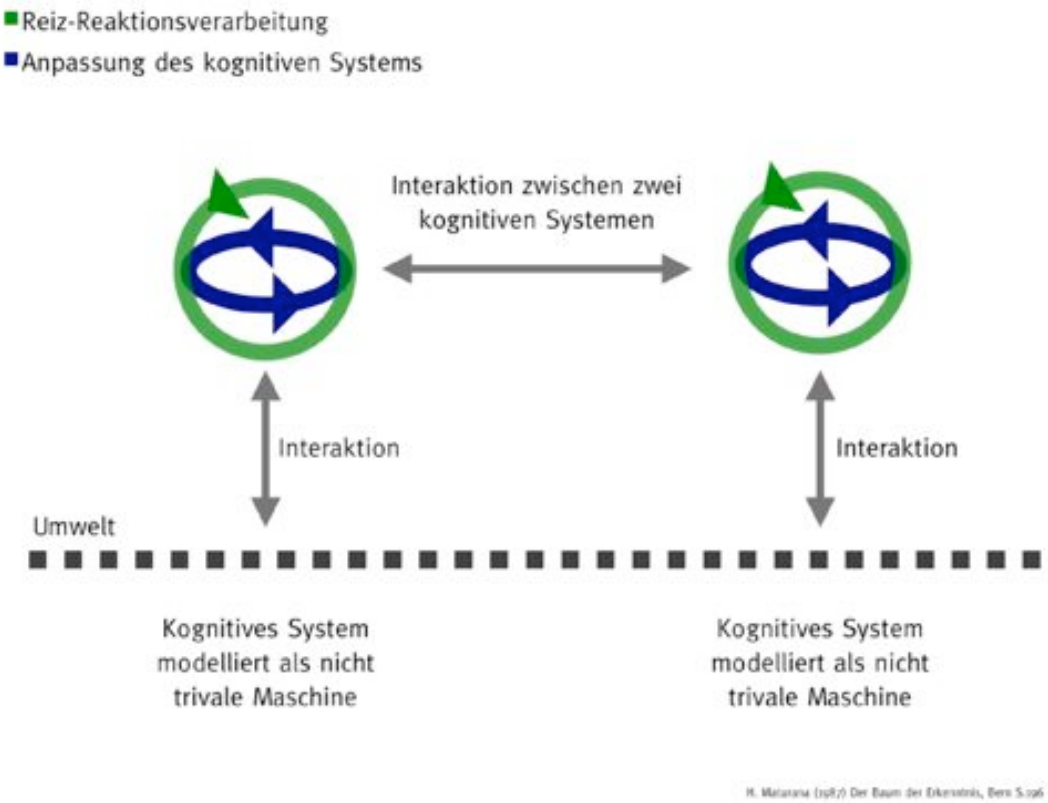


Abbildung 39: Kommunikation zwischen zwei Lernenden bei der Explorativen Datenanalyse

Innerhalb der Statistik kann er, neben seinem statistischen Fachwissen und Kompetenzen, aus der schließenden Statistik auf andere Darstellungsformen der interaktiven statistischen Graphik zurückgreifen.
Außerhalb der Statistik kann er versuchen, durch Plausibilitätsüberlegungen seinen inneren Konflikt aufzulösen. Wie gut das gelingt, hängt von seinem Vorwissen und sehr stark von dem zu bearbeitenden Beispiel ab.
Beschäftigt sich der Lernende nicht nur mit einer Aufgabe, sondern mit einer, nach Schwierigkeit gestaffelten Aufgabenfolge, sollte sich mit der Zeit ein stabiles Verhalten des Problemlösens einstellen und damit statistische Problemlösekompetenz entstehen. Dieser geschlossene Vorgang zwischen Lernenden und zu bearbeitender Aufgabe darf aber nicht darüber hinwegtäuschen, dass innerhalb dieses Systems nicht alle Lernschwierigkeiten gelöst werden können. Neben diesen Vorteilen für einen einzelnen Lerner ergibt sich noch für das Lernen in der Gruppe der Vorzug einer erleichterten Kommunikation über Lösungsansätze und Lösungswege. Durch die zusätzliche Möglichkeit, die interaktive statistische Graphik als Kommunikationswerkzeug einzusetzen,

eröffnet sich die Chance, mehrere Lernende relativ leicht und konstruktiv miteinander zu verbinden. Diese Lerngemeinschaft hat neben der Kommunikation über die Sprache, gezeichnete Diagramme und mathematische Kalküle noch die Möglichkeit kooperativ an einem Datenanalyseproblem mit der interaktiven statistischen Graphik zu arbeiten.

3.10 Ergebnisse der theoretischen Überlegungen

„ Wahrlich es ist nicht das Wissen, sondern das Lernen,
nicht das Besitzen sondern das Erwerben,
nicht das Da-Seyn, sondern das Hinkommen,
was den grössten Genuss gewährt."

Carl Friedrich Gauß

Die Modellierung des Lernprozesses als dynamisches System gibt eine Untergrenze für die Komplexität des Lernprozesses an. Die aus dem Lernprozess entstehende kognitive Struktur muss weder diesem Komplexitätsgrad entsprechen, noch muss der kognitive Prozess dem beschriebenen dynamischen System gleichen. Der angegebene Fixpunkt, oder wie Foerster sagt, die „Eigenbehavior“, gibt nur eine untere Grenze für die Komplexität der kognitiven Struktur des Lernenden an. Dies ist aus elementar logischen Überlegungen offensichtlich, denn in die Konstruktion des dynamischen Systems sind keine spezifische Eigenschaften des Menschen oder eines anderen Lebewesens eingeflossen. Die einzige Verbindung zum Lernenden ist, dass sich ein stabiles Verhalten für ein komplexes Problem herausbildet. Durch die sehr grobe Beschreibung des Lernenden als einer nicht-trivialen Maschine kann kein besseres Ergebnis erwartet werden. Dennoch zeigt schon diese grobe Abschätzung, dass bei der Gestaltung der Lernumgebung folgende Randbedingungen beachtet werden müssen:

1. Lernen erfolgt auf individuellen Lernwegen, denn die Iterationsschritte bei einem chaotischen System sind sensitiv in den Anfangswerten, so dass jeder Lernende, ausgehend von seinem Vorwissen, einen eigenen Lernweg beschreitet.

2. Nach u.U. sehr vielen Iterationsschritten stellt sich ein „Eigenbehavior" ein, das charakteristisch für den Lernenden ist. Im Lernprozess des Individuums bilden sich diese individuellen Eigenbehaviors heraus, die auch als das für dieses Individuum angepasste „Koordinatensystem" bezeichnet wird. Eine deterministische vollständige Beschreibung des Lernprozesses aus elementar sachlogischen Argumenten ist damit ausgeschlossen.

3. Für die Kommunikation zwischen zwei Individuen ist es nicht notwendig, dass sie das gleiche „Koordinatensystem" besitzen. Vielmehr ist es erforderlich, die Individuen mit gleichen Problemen zu konfrontieren, damit sie zumindest in Teilen „Eigenbehaviors" ausbilden, die dann eine gemeinsame Schnittmenge abdecken.

4. Um möglichst wenige Iterationsschritte zum Aufbau einer kognitiven Struktur vollziehen zu müssen, ist es von Vorteil, wenn der Lernende Probleme, mit denen er sich beschäftigt, selbst wählen kann. Die Anzahl der zu vollziehenden Iterationsschritte bis zum stabilen Zustand eines dynamischen Systems, hängt stark von der gewählten Anfangskonfiguration ab. Das bedeutet im Extremfall, dass ein Lernender bei einer geschickt gestellten Aufgabe ein stabiles Verhalten zeigt, und ein anderer Proband mit dieser Aufgabe sehr lange braucht, um ans Ziel zu kommen.

4. Empirische Untersuchung

Das vierte Kapitel dieser Arbeit gliedert sich in folgende Teile:

1. Design der empirischen Untersuchung
2. Methodisches Vorgehen
3. Filmanalyse und Videobeobachtung
4. Beispiel einer Filmanalyse anhand des „PRIM 9“- Films
5. Untersuchungergebnisse aus den Filmanalysen
6. Studierendenexperiment – Erhebungsinstrumente
7. Ergebnisse des Studentenexperiment
8. Expertenbefragung – Erhebungsinstrumente
9. Beispiel einer tabellarischen Ergebniszusammenfassung
10. Ergebnisse der Expertenbefragung

4.1. Design der empirischen Untersuchung

„Das Wissen ist Kind der Erfahrung."
Leonardo da Vinci

Die nachfolgend dargestellte empirischen Untersuchung beschränkt sich auf die statistische Kompetenzentwicklung von Studierenden an Universitäten. Es gibt zwar Ansätze, die Explorative Datenanalyse auch für die Statitikausbildung bei Schülern zu nutzen (S. Hußmann, 2003, S. 194), aber diese Bemühungen stehen noch völlig am Anfang. Da selbst an Universitäten die Explorative Datenanalyse nicht flächendeckend verbreitet ist und es folglich auch im Universitätskontext schwierig ist, geeignete Daten zu gewinnen, beschränkt sich nachfolgende Untersuchung auf Lehrende und Studierende der Universität Augsburg sowie einige Experten aus dem Umfeld der Universität Augsburg. Welchen Einfluss die Explorative Datenanalyse auf das intuitive Verständnis statistischer Sachverhalte hat, ist relativ schwierig zu ermitteln, denn die Entwicklung mathematischer Kompetenz ist stets ein Wechselspiel zwischen praktischer Anwendung, intuitiven Verständnis und dem formalen, korrekten Wissen über die mathematischen Hintergründe. Das Eine kommt ohne das Andere nicht aus. Eine Lehrstrategie, die auf zu viel Praxis setzt, ist genauso zum Scheitern verurteilt wie eine Strategie, die nur das formale, theoretische Gebäude sieht. Aus diesem Grund ist es schwierig, die Kompetenzentwicklung bei Studierenden einer konkreten Ursache im Lehrbetrieb zuzuordnen. Es gelingt nur eine relativ grobe und ungefähre Zuordnung. Die Beobachtungen, die dieser Arbeit zu Grunde liegen lauten wie folgt:

1. Eine Untersuchung einer Gruppe Studierender, die entweder eine Vorlesung oder ein Seminar bei Prof. Unwin besuchen, in denen die Explorativen Datenanalyse eingesetzt wird. Hier wurde folgendes untersucht:

    (a) die Lernbiographie, d.h. welche schulische Vorbildung vorhanden ist und welche Interessen und Neigungen vorliegen.

    (b) die Beobachtung der Problemlösekompetenzen in Form eines Tests

    (c) ein Experiment, bei dem mit Hilfe der Explorativen Datenanalyse ein Problem bearbeit werden soll. Bei dieser Aufgabe werden die Studierenden videografiert und die Ergebnisse in Form von Dokumenten analysiert.

    (d) Jeder videografierte Studierende wurde zusätzlich anhand eines Leitfadeninterviews befragt.

2. Eine Expertenbefragung in Form von Leitfadeninterviews mit acht Hochschullehrern aus dem Umfeld der Universität Augsburg, wobei vier die Explorative Datenanalyse in ihren Veranstaltungen nicht einsetzen und vier die Explorative Datenanalyse ganz regulär in ihren Lehrveranstaltungen verwenden.

Das beschriebene Forschungsdesign ist ein Kompromiss in mehrerlei Hinsicht:

1. Die Gruppe der Studierenden, die untersucht wurde, ist klein. Es konnten Daten von insgesamt 13 Studierenden erhoben werden. Die Probanden haben alle Lehramt Gymnasium mit dem Schwerpunkt Mathematik, aber mit unterschiedlichen Nebenfächern, studiert. Hauptsächlich waren die Fächerkombinationen Mathematik/Sport, sowie Mathematik/Physik vertreten. Trotz dieser kleinen Gruppe erfüllt die Auswahl der Studierenden die Kriterien einer zufälligen Stichprobe, da fast alle Studierenden befragt wurden, die in diesem Jahrgang (4. Semester) den Kurs „Statistik I für Lehrämter" hören sollten. Da sich die Universität Augsburg weder in ihrer Struktur noch in ihrer Qualität stark von anderen bayerischen Universitäten unterscheidet, kann man davon ausgehen, dass sich die Situation Bayernweit nicht wesentlich unterscheidet.
2. Aus Gründen der Validität wäre es wünschenswert, eine größere Studierendengruppe zu untersuchen. Dies hätte aber vorausgesetzt, dass Explorative Datenanalyse regulärer Teil der Statistikausbildung an bayerischen Universitäten ist. Da dies nicht der Fall ist, muss man sich folglich mit der vorliegenden Stichprobe begnügen und versuchen, damit zumindest eine grobe Abschätzung der Chancen und Möglichkeiten der Explorativen Datenanalyse zu bestimmen.
3. Neben den üblichen Befragungen und einer Leistungsermittlung in Form eines Prüfungsbogens wurden die Testpersonen zusätzlich in einem Experiment beobachtet, um direkte Daten von der Interaktion zwischen Testperson, Interaktiver Statistischer Graphik und realer Fragestellung zu bekommen.
4. Um einen grundsätzlichen Anhaltspunkt für das Verhältnis zwischen praktischem Bezug und abstrakter Darstellung zu bekommen, wurde eine Expertenbefragung an die Untersuchung der Kohorte der Studierenden angeschlossen. Die Befragung war so konstruiert, dass die eine Hälfte der Experten Pioniere der Explorativen Datenanalyse sind und die andere Hälfte renommierte Wissenschaftler, die die Explorative Datenanalyse zwar kennen,

aber selbst in der Lehre nicht anwenden und ihre Forschungsschwerpunkte auf andere Bereiche der Statistik legen.

Welche Ergebnisse sind mit dem Forschungsdesign zu erzielen?
Um präzise Aussagen darüber zu treffen, wie groß der Anteil der Explorativen Datenanalyse anhand realer Beispiele an der Statistikkompetenz der Studierenden ist, müsste man ein 2x2 Design mit einer großen Zahl von Studierenden implementieren. Das Problem ist jedoch, dass die Kurse, die die Studenten besuchen, entweder vollkommen ohne Explorative Datenanalyse und reale Datensätze auskommen müssten, oder der Kurs mit Explorativer Datenanalyse und geeigneten realen Beispielen ideal gestaltet sein müsste. Dies geschieht unter der Annahme, dass es genug Studierende, genug Professoren und genügend Mittel für eine solche Untersuchung gäbe. Daraus ergibt sich ein Widerspruch, denn die Wirkung und die Rahmenbedingungen sollen durch die durchgeführte Untersuchung erst bestimmt werden. Die vorliegende Untersuchung beschränkt sich deshalb auf das Ziel, Lehrenden einen groben Anhaltspunkt für die Integration der Explorativen Datenanalyse in ihre Veranstaltungen zu geben.

4.2 Methodisches Vorgehen

„Zur Erforschung der Wahrheit bedarf
es notwendig der Methode."
René Descartes

Am Beginn der Untersuchung steht die Analyse der Explorativen Datenanalyse und ihrem Hauptwerkzeug, der Interaktiven Statistischen Graphik als Problemlösewerkzeug. Für diese Analyse konnte der Verfasser auf Vorarbeiten des Lehrstuhls für Rechnerorientierte Statistik und Datenanalyse zurückgreifen. Vor allem die Arbeit von Martin Theus (1996) mit dem Titel „Theorie und Anwendung interaktiver statistischer Graphik" und die eigene Diplomarbeit mit dem Titel „Multimediale Einführung in die Explorative Datenanalyse"(U. Fahrner, 2005) halfen in der Anfangsphase den Forschungsgegenstandes zu präzisieren.

Abbildung 40: Beispiel für ein interaktives Video

Die Explorative Datenanalyse ist seit Anfang der 1970er Jahre ein sehr erfolgreiches Erkenntniswerkzeug im Bereich der Naturwissenschaften. Frühe Anwendungen für das erste interaktive System, in dem die Interaktive Statistische Graphik implementiert wurde, waren Beschleunigerdaten eines großen Linearbeschleunigers in Stanford. Die Arbeitsgruppe um John Tukey entwickelte das PRIM-9-System, das extra nur für die Auswertung der Beschleunigerdaten entwickelt wurde. Ausgehend von diesem ersten Projekt entwickelte sich im Laufe der Jahrzehnte eine wissenschaftliche Community, die sich schwerpunktmäßig mit der Explorativen Datenanalyse beschäftigte und viele ihrer zukunftsweisenden Entwicklungen in Form von Filmen festhielt. Die Beschreibung der interaktiven Suche eines Modells innerhalb eines Datensatzes mit Hilfe der Interaktiven Statistischen Graphik lässt sich nur schwer in einem linearen Text darstellen. Die Analyse der Filme der „Video Lending Library" der „American Statistics Association" und von Filmen, die im Oktober 2002 im Workshop „Data Visualization for large data sets and Data Mining" (http://rosuda.org/Workshop2002/WS.html) entstanden waren, ergaben erste Hinweise auf das iterative Vorgehen bei einer Explorativen

Datenanalyse (vgl. Kapitel 2). Im Anschluss an diese Analyse wurden, in Kooperation mit Antony Unwin, vier Studierende zu ihrem Graphikverständnis befragt. In diesem Pretest ging es um Fragen, inwieweit Graphiken eindeutig verstanden werden und ob eine Einbettung von Graphiken und statistischen Fragestellungen in reale Aufgaben Vorteile für den Lernenden bietet. Das Ergebnis dieses Pretests lässt sich kurz zusammenfassen:

Werden statistische Fragestellungen mit Hilfe realer Datensätze in einen Kontext eingebettet, so fällt es den Studierenden leichter, diese Aufgaben zu bearbeiten. Ist die Aufgabe offen gestellt, so dass es mehrere Wege zu einer oder zu mehreren Lösungen gibt, ist der Lerneffekt bei den Studierenden groß. Die stark didaktisierten Aufgaben mit einem klar vorgegebenen Lösungsweg waren für die Studierenden nicht anregend. Ausgehend von diesem Pretest wurde deutlicher, welchen Kriterien eine Untersuchung der Explorativen Datenanalyse als Lernwerkzeug genügen muss. Es wurde beobachtet, dass sich mit der Explorativen Datenanalyse individuelle Lernwege öffnen lassen, die den Studierenden anregen, sich sowohl mit den theoretischen Werkzeugen als auch mit den praktischen Anwendungen näher zu beschäftigen.

Abbildung 41: Beispiel für das Analysefenster

Aus diesem Grund musste ein Design gewählt werden, bei dem der individuelle Lernweg beobachtet werden kann. Als ideales Werkzeug für eine solche Fragestellung wurde ein Experiment vorgesehen, das folgendermaßen aufgebaut ist: Der Testperson wird ein offen gestelltes Problem vorgelegt. Dieses Problem soll die Testperson mit einer Kombination aus Interaktiver Statistischer Graphik, dem realen Datensatz, dem dazugehörigen realen Kontext und seiner internalisierten statistischen Kompetenz lösen. Der Prozess des „Aufgabenlösens” wird mit einer Videokamera beobachtet. Die Manipulation, die die Testperson am Rechner ausführt, wird synchron dokumentiert. Im Anschluss an diesen Test wird ein Interview geführt, in dem die Testperson den Problemlösevorgang aus ihrer subjektiven Sicht beschreibt. Durch den Vergleich dieser drei Datenquellen ist es möglich, den Problemlösevorgang in groben Zügen zu rekonstruieren und daraus Schlüsse zu ziehen. Um festzustellen, von welchem Kompetenz- und Motivationsniveau jede einzelne Testperson ausgeht, wurden mittels eines Fragebogens die Lernbiographie und die Grundeinstellung zum Fach Mathematik erfasst. Als weitere Vergleichsmöglichkeit stand ein Kompetenztest, ähnlich dem schon durchgeführten Pretest, zur Verfügung. Auf Grund der Versuchsergebnisse aus der Untersuchung der Studierendengruppe und der Filmanalyse entstand ein Leitfaden für die sich anschließende Expertenbefragung. Durch die Kombination von Experimenten mit dem Wissen von Experten, die über jahrzehntelange Lehrerfahrung verfügen, ist es trotz des kleinen Stichprobenumfanges gelungen, die Rahmenbedingungen für den Einsatz der Explorativen Datenanalyse als Lernwerkzeug vorgeben.

4.3. Filmanalyse und Videobeobachtung

„Bevor man beobachtet,
muß man sich Regeln für seine Beobachtungen machen."

Jean-Jacques Rousseau

Insgesamt wurden 37 Dokumentationen von Anfang der 1970er Jahre bis ca. 1990 analysiert. In den Filmen sieht der Betrachter den Forscher beim Interagieren mit der Software und mit einem Datensatz. In vielen dieser Analysebeispiele tritt der spielerische Aspekt der Datenanalyse hervor. Spiel ist in dem Sinne zu verstehen, etwa der spielerische Umgang mit einem Baukasten, mit dem Kinder versuchen immer wieder neue Objekte zu bauen und zu konstruieren. Wie mit einem aus Baukasten aus Plots, Datensätzen und Metadaten versuchen die Datenanalysten eine möglichst schlüssige Antwort auf eine gestellte Frage zu geben oder sich selbst neue Fragen und interesessante Zielsetzungen zu überlegen. Das Ergebnis dieser Analyse war, dass die Wege wie man mittels der Explorativen Datenanalyse ans Ziel kommt, höchst unterschiedlich sind und das kreative Suchen nach Lösungswegen eines der Hauptcharakteristika darstellt.

Beispiel einer Dokumentenanalyse anhand des PRIM-9-Films

Historie und Umfeld:

Die Entwicklung der Explorativen Datenanalyse ist eng mit John W. Tukey verknüpft. Anfang der 1960er Jahre begann er die statistische Datenanalyse neu zu definieren. Meilensteine bei der Entwicklung der rechnerorientierten explorativen Datenanalyse waren der Artikel „The Future of Data Analysis” und die Implementierung interaktiver statistischer Graphik im PRIM-9-Projekt. PRIM-9 steht für Picturing (Auswahl einer zweidimensionalen Projektion), Rotation, Isolation (Einschränkung auf den Unterraum der Variablen), Masking (Gruppenbildung und Ausreihererkennung) und stand damit Pate für grundlegende methodische Ansätze interaktiver Graphik. Zu diesem Projekt entstand ein 16mm Film, der die damals revolutionären Ansätze dokumentierte und die interaktiven Möglichkeiten des Systems zeigt.

Technik:

Die Bild und Tonqualität des Films zum PRIM-9-Projekt entspricht der Qualität eines Dokumentarfilms der 1960er und 1970er Jahre. Der Film wurde auf 16mm Filmmaterial aufgenommen. Die Ausleuchtung der Szenen ist gleichmäßig und sorgt an jeder Stelle des Films für ein schatten- und reflexionsfreies Bild. Kameraführung, Schnitt und Überblendungen sind professionell und in einem zweiten Abschnitt erläutert John W. Tukey die Projektvorgaben und Ziele, die einem klaren Drehbuch folgen. Die sehr gute Tonqualität wird zum Teil durch Ansteckmikrophone, Richtmikrophone und eine Nachvertonung erreicht.

Inhalt:

Der inhaltliche Aufbau des Film ist an Lernzielen orientiert. Der Inhalt steht im Vordergrund und an ihm richten sich die eingesetzten formalen Elemente aus. Mit filmtechnischen Effekten wird deshalb äußerst sparsam umgegangen. Der Film ist inhaltlich klar in folgende Abschnitte gegliedert:

Uni – Projekttitel – Personen:

Zu Beginn des Films werden das Labor, der Projekttitel und die Mitarbeiter kurz vorgestellt. In der ersten Szene sieht der Zuschauer Luftaufnahmen des Linearbeschleunigers (Stanford Linear Acceleration Center) und erhält so einen Eindruck, aus welchem Kontext die Daten stammen. Der Projekttitel wird erklärt: PRIM 9 steht für Picturing, Rotation, Isolation, Masking in neun Dimensionen. Danach werden die Mitarbeiter des Projekts mit ihren zugehörigen Aufgabenfeldern vorgestellt.

Begriffserklärung und grundlegende Definitionen:

In einem zweiten Abschnitt erläutert John W. Tukey die Projektvorgaben und Ziele und gibt einen kleinen Ausblick auf die weiter reichenden Entwicklungsmöglichkeiten Interaktiver Statistischer Graphik.

Datenanalyse mit Hilfe des Rechnersystems PRIM-9:

Anhand realer Daten zeigt John W. Tukey das prinzipielle Vorgehen bei der Explorativen Datenanalyse. Er verknüpft die Darstellung interaktiver statistischer Graphik mit der konkreten Implementierung im PRIM-9-System. Dadurch werden in kurzer Zeit die Vorteile interaktiver Werkzeuge klar. Nebenbei werden die Bedienungselemente des Rechners erklärt, so dass der Zuschauer rasch versteht, dass Interaktivität in der statistischen Graphik mehr ist als eine technische Spielerei, sondern vielmehr neue innovative Wege in der Datenanalyse eröffnet.

Umsetzung der Ergebnisse in ein physikalisches Modell:

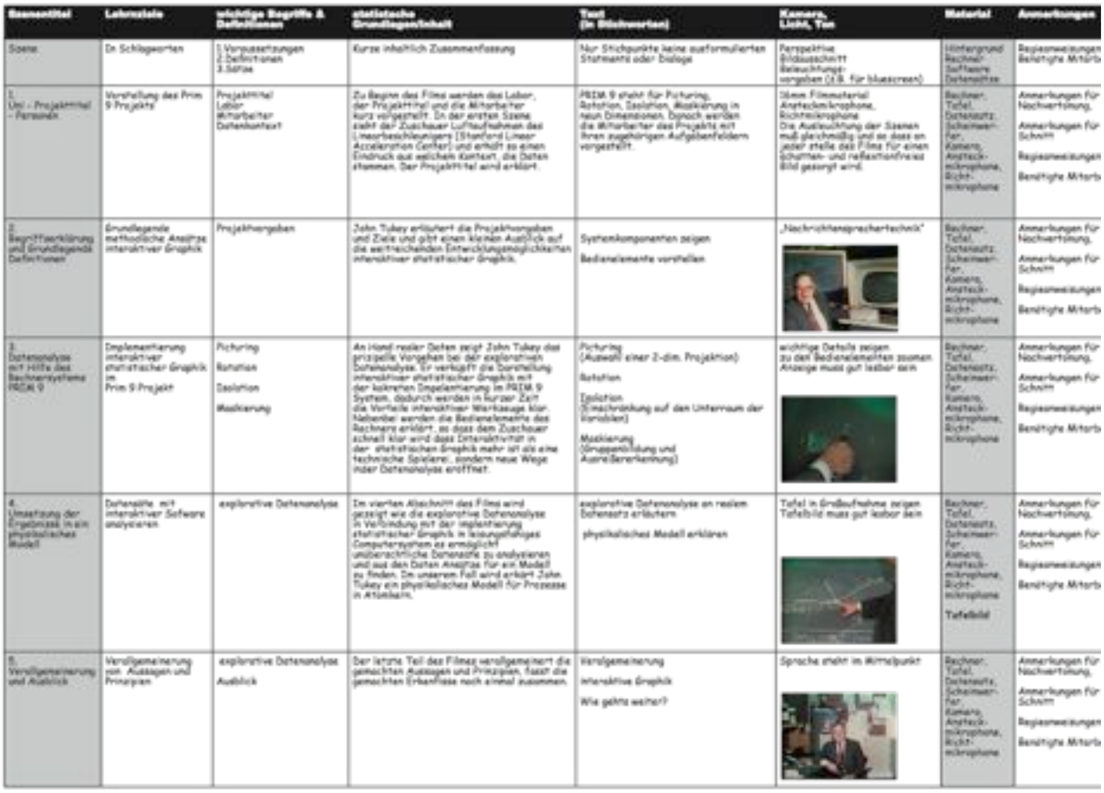
Im vierten Abschnitt des Films wird gezeigt, wie die Explorative Datenanalyse in Verbindung mit der Implementierung statistischer Graphik und einem leistungsfähigen Computersystem ermöglichen, unübersichtliche Datensätze zu analysieren und aus den Daten Ansätze für ein Modell zu finden. In unserem Fall erklärt John W. Tukey ein physikalisches Modell für Prozesse im Atomkern.

Verallgemeinerung und Ausblick:

Der letzte Teil des Filmes verallgemeinert die Aussagen und Prinzipien und fasst die Erkenntnisse noch einmal zusammen. Im Film setzt John W. Tukey das Leitbild um, das er für die Explorative Datenanalyse geprägt hat. Der Forscher arbeitet wie ein Detektiv, der, ausgehend von einem Problem, in den Daten interessante Strukturen und Besonderheiten aufdeckt, Hinweisen nachgeht und Hypothesen entwickelt.

Formale Elemente und eingesetzte Medien:

Im Film wird mit vier prinzipiell verschiedenen Szenensettings gearbeitet. Abwechselnd werden Personen, Personen mit Rechner (John W. Tukey erklärt dabei die interaktiven Bedienelemente von PRIM 9.), der Bildschirm und Personen an der Tafel gezeigt. Immer dann, wenn Personen in den Szenen vorkommen, wird versucht den Zuschauer durch direkten Blickkontakt einzubeziehen. Die Situation gleicht der eines Nachrichtensprechers, der ebenso im direkten Blickkontakt mit den Zuschauern steht. Durch diese „Nachrichtensprechertechnik" wirkt der Film sehr lebendig und zieht das Interesse des Zuschauers auf sich. Das verstärkt sich durch die wechselnden Szenensettings, die dazu führen, dass trotz der Filmlänge von 25 Minuten der Betrachter bei der Sache bleibt und der Film nicht langweilig oder langatmig wirkt. Die eingesetzten Medien werden in der jeweiligen Szene für einen klar umrissenen inhaltlichen Abschnitt verwendet, je nachdem, ob die „Lernziele" besser visuell, audiovisuell oder textlich vermittelt werden können. Zusätzlich zur Analyse der historischen Filmdokumente bot sich im Oktober 2002 im Workshop „Data Visualization for large data sets and Data Mining"(http://rosuda.org/Workshop2002/WS.html), der vom Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse der Univeristät Augsburg veranstaltet wurde, die Möglichkeit, weltweit wichtige Vertreter der Explorativen Datenanalyse bei ihrer Arbeit zu beobachten. Bei diesem Workshop entstanden 17 Filme, in denen Datenanalyseexperten, mit der von ihnen entwickelten Software, Datensätze analysieren und die Vorzüge der Explorativen Datenanalyse aus ihrer Sicht darstellen. Diese Filme ergänzten die Eindrücke, die sich schon aus der Analyse des historischen Filmmaterials ergeben haben.



Szenentitel	Lehrziele	wichtige Begriffe & Definitionen	statistische Grundlagen/Inhalt	Text (in Stichworten)	Kamera, Licht, Ton	Material	Anmerkungen
Szene	In Schlagworten	1.Voraussetzungen 2.Definitionen 3.Sätze	Kurze inhaltlich Zusammenfassung	Nur Stichpunkte keine ausformulierten Statments oder Dialoge	Perspektive Bildausschnitt Beleuchtungs-vorgaben (z.B. für bluescreen)	Hintergrund Rechner Software Datensätze	Regieanweisungen Benötigte Mitarb
1. Uni - Projekttitel - Personen	Vorstellung des Prim 9 Projekts	Projekttitel Labor Mitarbeiter Datenkontext	Zu Beginn des Films werden das Labor, der Projekttitel und die Mitarbeiter kurz vorgestellt. In der ersten Szene sieht der Zuschauer Luftaufnahmen des Linearbeschleunigers (Stanford Linear Acceleration Center) und erhält so einen Eindruck aus welchem Kontext, die Daten stammen. Der Projekttitel wird erklärt.	PRIM 9 steht für Picturing, Rotation, Isolation, Maskierung in neun Dimensionen. Danach werden die Mitarbeiter des Projekts mit ihren zugehörigen Aufgabenfeldern vorgestellt.	16mm Filmmaterial Ansteckmikrophone, Richtmikrophone Die Ausleuchtung der Szenen muß gleichmäßig und so dass an jeder stelle des Films für einen Schatten- und reflexionfreies Bild gesorgt wird.	Rechner, Tafel, Datensatz, Scheinwer-fer, Kamera, Ansteck-mikrophone, Richt-mikrophone	Anmerkungen für Nachvertonung, Anmerkungen für Schnitt Regieanweisungen Benötigte Mitarb
2. Begriffserklärung und Grundlegende Definitionen	Grundlegende methodische Ansätze interaktiver Graphik	Projektvorgaben	John Tukey erläutert die Projektvorgaben und Ziele und gibt einen kleinen Ausblick auf die weitreichenden Entwicklungsmöglichkeiten interaktiver statistischer Graphik.	Systemkomponenten zeigen Bedienelemente vorstellen	„Nachrichtensprechertechnik"	Rechner, Tafel, Datensatz, Scheinwer-fer, Kamera, Ansteck-mikrophone, Richt-mikrophone	Anmerkungen für Nachvertonung, Anmerkungen für Schnitt Regieanweisungen Benötigte Mitarb
3. Datenanalyse mit Hilfe des Rechnersystems PRIM 9	Implementierung interaktiver statistischer Graphik im Prim 9 Projekt	Picturing Rotation Isolation Maskierung	An Hand realer Daten zeigt John Tukey das prizipelle Vorgehen bei der explorativen Datenanalyse. Er verknüpft die Darstellung interaktiver statistischer Graphik mit der kokreten Implementierung im PRIM 9 System, dadurch werden in kurzer Zeit die Vorteile interaktiver Werkzeuge klar. Nebenbei werden die Bedienelemente des Rechners erklärt, so dass dem Zuschauer schnell klar wird dass Interaktivität in der statistischen Graphik mehr ist als eine technische Spielerei, sondern neue Wege inder Datenanalyse eröffnet.	Picturing (Auswahl einer 2-dim. Projektion) Rotation Isolation (Einschränkung auf den Unterraum der Variablen) Maskierung (Gruppenbildung und Ausreißererkennung)	wichtige Details zeigen zu den Bedienelemetten zoomen Anzeige muss gut lesbar sein	Rechner, Tafel, Datensatz, Scheinwer-fer, Kamera, Ansteck-mikrophone, Richt-mikrophone	Anmerkungen für Nachvertonung, Anmerkungen für Schnitt Regieanweisungen Benötigte Mitarb
4. Umsetzung der Ergebnisse in ein physikalisches Modell	Datensätze mit interaktiver Sofware analysieren	explorative Datenanalyse	Im vierten Abschnitt des Films wird gezeigt wie die explorative Datenanalyse in Verbindung mit der implentierung statistischer Graphik in leistungsfähiges Computersystem es ermöglicht unübersichtliche Datensätze zu analysieren und aus den Daten Ansätze für ein Modell zu finden. Im unserem Fall wird erklärt John Tukey ein physikalisches Modell für Prozesse in Atomkern.	explorative Datenanalyse an realem Datensatz erläutern physikalisches Modell erklären	Tafel in Großaufnahme zeigen Tafelbild muss gut lesbar sein	Rechner, Tafel, Datensatz, Scheinwer-fer, Kamera, Ansteck-mikrophone, Richt-mikrophone Tafelbild	Anmerkungen für Nachvertonung, Anmerkungen für Schnitt Regieanweisungen Benötigte Mitarb
5. Verallgemeinerung und Ausblick	Verallgemeinerung von Aussagen und Prinzipien	explorative Datenanalyse Ausblick	Der letzte Teil des Filmes verallgemeinert die gemachten Aussagen und Prinzipien, fasst die gemachten Erkenntisse noch einmal zusammen.	Verallgemeinerung interaktive Graphik Wie gehts weiter?	Sprache steht im Mittelpunkt	Rechner, Tafel, Datensatz, Scheinwer-fer, Kamera, Ansteck-mikrophone, Richt-mikrophone	Anmerkungen für Nachvertonung, Anmerkungen für Schnitt Regieanweisungen Benötigte Mitarb

Abbildung 42: Beispiel des verwendeten Analyseschemas

4.3 Untersuchungsergebnisse aus den Filmanalysen

„Zu den Dingen, welche einen Denker in Verzweiflung bringen können,
gehört die Erkenntnis, dass das Unlogische für den Menschen nötig ist,
und dass aus dem Unlogischen vieles Gutes entsteht."

Friedrich Nietzsche

Anhand der Dokumentenanalyse der Filme ist zu erkennen, dass die Explorative Datenanalyse durch folgende Aktionen gekennzeichnet ist:

1. Probieren

Zu Beginn einer jeden Analyse starten die Experten damit, ihrer Intuition folgend, differente Plots zu testen und sich eine Übersicht zu verschaffen. Dieser Vorgang kann mit dem Aufwärmen eines Sportlers verglichen werden: Es werden einfache Standardplots und Basisheuristiken angewandt, um sich einen ersten Eindruck von dem gestellten Problem zu verschaffen. Das Vorgehen ist meist noch nicht zielgerichtet oder orientiert sich an einer zu beantwortenden Frage.

2. Spielen

Nach der ersten „Aufwärmphase” ändert sich der Charakter der Datenanalyse. Es ähnelt nun mehr einem Spiel, bei dem man versucht durch die zielgerichtete Manipulation der vorhandenen Bausteine (Plots, Datensatz, Metadaten) näher kennen zu lernen und einen Plan für die Lösung der Aufgabe zu gewinnen. In dieser Phase entstehen neue Fragen, denen der Datenanalyst nachgeht, wenn er sie für interessanter als die Ausgangsfrage erachtet.

3. Reflektieren

Der Datenanalyst hat sich durch die vorhergehende Phase Ansatzpunkte und einen Plan zur Lösung der gestellten Frage überlegt und überprüft nun diese Überlegungen kritisch. Dabei fließen seine Erfahrung, sein Wissen über den Datensatz sowie seine Statistikkompetenz ein. Reichen ihm in dieser Phase die Ergebnisse noch nicht aus oder fühlt er sich mit dem Problem noch nicht genügend vertraut, beginnt er wieder mit Phase 2. Der Experte versucht durch einen spielerischen Umgang mit den einzelnen Elementen der Interaktiven Statistischen Graphik mehr über das Problem herauszufinden.

4. Verwerfen

In dieser Phase verwirft der Experte alle Ansätze und Überlegungen, die nicht geradlinig zur Lösung des zu untersuchenden Problems führen. Alle Nebenüberlegungen, Irrwege und Fragen, die *en passant* beantwortet wurden, werden aus dem Lösungsweg entfernt.

5. Vergleichen

Die entstandene Lösung wird auf ihre Plausibilität hin überprüft, indem die Ergebnisse mit ähnlich gelagerten Fragestellungen verglichen oder die Ergebnisse mit anderen statistischen Methoden (der schließenden Statistik) getestet werden.

6. Bewerten

Am Ende der Analyse werden sowohl das Ergebnis als auch der Lösungsweg kritisch gewürdigt. Es wird versucht, diesen in einen größeren Kontext zu stellen. Es taucht immer wieder die Frage auf, wie der Datensatz, die Metadaten und die verwendete Software aussehen müssten, um das gezielte Ergebnis zu verbessern.

Aus den beobachteten Aktionen der Experten lässt sich eine Taxonomie für die Explorative Datenanalyse aufstellen (in Anlehnung auf Bloom, 1964).

Abbildung 43: Taxonomie für die Explorative Datenanalyse (in Anlehnung an Bloom, 1964)

Die erste Phase ist vom „Ausprobieren" gekennzeichnet. Dafür ist es notwendig, die Möglichkeiten und Grenzen der einzelnen Plots zu kennen und zu wissen, welche Möglichkeiten die Explorative Datenanalyse hat. Deswegen kann man als Basis für die erarbeitete Taxonomie den Begriff „Wissen" verwenden. In der ersten und zweiten Phase wird mit den Werkzeugen der Interaktiven Graphik probiert und gespielt. Das bedeutet nichts anderes, als Wissen aus der ersten Stufe der entwickelten Taxonomie anzuwenden. Daraus folgt, dass die zweite Stufe der Taxonomie das Anwenden ist. In der dritten Phase einer Datenanalyse ist die dominante Aktion das „Reflektieren", die der Analyse

der angewendeten Methoden und Werkzeuge entspricht. Deshalb steht auf der dritte Stufe der Taxonomie das „Analysieren”. Die Phasen vier und fünf sind durch die Aktionen „Verwerfen” und „Vergleichen” charakterisiert. Bei genauer Betrachtung geht es in diesen beiden Phasen durch Wegstreichen und Vergleichen mit anderen statistischen Methoden um das Herausarbeiten einer klar strukturierten Lösung. Dies kann man auch als Synthese eines Lösungsweges begreifen. Deshalb bezeichnen wir die nächste Stufe in der Taxonomie als „Synthetisieren”. Die letzte Aktion in einer Explorativen Datenanalyse ist das „Bewerten”. Dieses Bewerten setzt auf die Einordnung der erzielten Ergebnisse in einen größeren Zusammenhang und setzt daher voraus, dass der Bewertende einen genauen Überblick und „Einsicht” in das bearbeitete Themenfeld hat. Daher ist die oberste Stufe der vorgeschlagenen Taxonomie das Bewerten. Diese Taxonomie ist so zu verstehen, dass ausgehend von der Stufe „Wissen” über die Stufen „anwenden”, „analysieren”, „synthetisieren” und „bewerten” der Grad der „Kompetenz” steigt.
Diese Taxonomie bildet die Grundlage für den Fragebogen zur Kompetenzbewertung der untersuchten studentischen Stichprobe. Insgesamt gibt es in diesem Testfragebogen sechs verschiedene Aufgabentypen. Diese unterscheiden sich durch steigende Anforderungen im Sinne der angegebenen Taxonomie.

4.4 Studentenexperiment – Erhebungsinstrumente

„In Wahrheit heißt etwas wollen,
ein Experiment machen, um zu erfahren, was wir können;
darüber kann uns allein der Erfolg oder Misserfolg belehren."

Friedrich Nietzsche

Die durchgeführte Untersuchung an einer Stichprobe von dreizehn Studierenden besteht aus folgenden Teilen:

1. Erhebung der Lernbiographie mittels Fragebogen und qualitativem Interview.

2. Ermittlung einer Kompetenzstufe nach der entwickelten Taxonomie.
3. Experiment, bei dem mittels Videographie die Interaktion zwischen einem Lernenden, einem Datensatz und einer interaktiven statistischen Software beobachtet und untersucht wird.
4. Erhebung der subjektiven Einschätzung des eigenen Lernens und Problemlösens mit Hilfe von Graphiken, einem Fragebogen und einem qualitativen Interview.

Die Erhebungsinstrumente wurden auf Grundlage von theoretischen Überlegungen, einem Pretest und einer vom Verfasser technisch betreuten Diplomarbeit mit dem Titel „Usability von Statistik-Software (USS) am Beispiel von Mondrian" konstruiert, die Arnold Brutler am Lehrstuhl für Rechnerorientierte Datenanalyse und Statistik geschrieben hat. In dieser Arbeit werden die technischen Grundlagen für die Beobachtung der Interaktion zwischen einem Datenanalyseprogramm und einem Datenanalysten entwickelt. Arnold Brutler und der Verfasser haben auf Grundlage von Quicktime eine Playerkonfiguration umgesetzt, mit der es möglich ist, Annotationen, Computeraktivität und die interagierende Person synchron darzustellen. Weiterhin ist es damit möglich, solche Beobachtungsdokumente über die Zeit mit Kapiteleinschüben zu strukturieren.

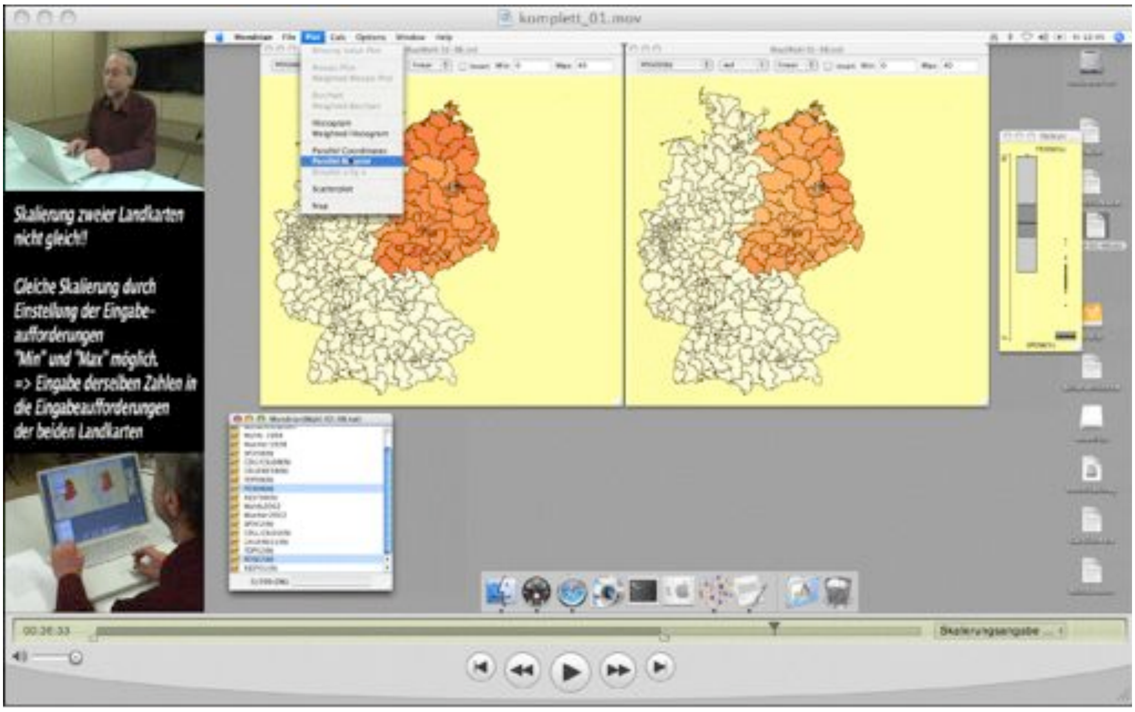


Abbildung 44: Beispiel für ein Analysefenster mit Navigation und Annotation

Auf Grund der aufgezeichneten Beobachtungen werden die jeweiligen Lernwege rekonstruiert. Zusätzlich zu den vom Verfasser entwickelten Erhebungsinstrumenten wurden Teile von Standardtests zur Lerntypeneinteilung verwendet.
Diese erwiesen sich jedoch als wenig nützlich, da sie in einer kleinen Stichprobe (wie in der untersuchten Studierendengruppe) keine eindeutigen Ergebnisse liefern oder grundsätzlich nicht funktionieren.

Fragebogen zu den Lernstrategien

Der Fragebogen zur Lernstrategie der Studierenden besteht aus drei Teilen. Im ersten Teil werden Basisdaten erhoben: schulischer Werdegang, Lieblingsfächer, das Verhältnis zur Mathematik und das Verhältnis zu graphischen Darstellungen. Im zweiten Teil sollen die Probanden einschätzen oder aus ihrer Erfahrung Auskunft geben, wie sie schwierige mathematische Probleme lösen und dazu Graphiken nutzen. Der dritte Teil besteht aus einem Lerntypenfragebogen, der untersuchen möchte, inwieweit der Proband visuelles, auditives, handelndes oder rezeptives Lernen präferiert.
Dieser Fragebogen wurde von Werner Stangl vom Institut für Pädagogik und Psychologie von der Johannes Keppler Universität Linz übernommen. Der wichtigste Teil dieses Fragebogens ist der Abschnitt zwei, in dem es um das Problemlösen geht. Die Operationalisierung der beiden zentralen Fragen basiert auf dem Konzept von George Pólya (vgl. Kapitel 3.2) und auf den empirischen Befunden des Pretests, der zusammen mit Antony Unwin durchgeführt wird.

Kompetenzfragebogen

Der Kompetenzfragebogen wurde Aufgrund der vorangegangenen Videoanalyse und der darauf aufbauenden Taxionomie so konstruiert, dass jedem Aufgabenteil eine Kompetenzstufe zugeordnet werden kann. Dies wird durch sechs verschiedene Aufgabentypen möglich, die mit jeweils mit einer Kompetenzstufe korrespondieren.

Aufgabentypen

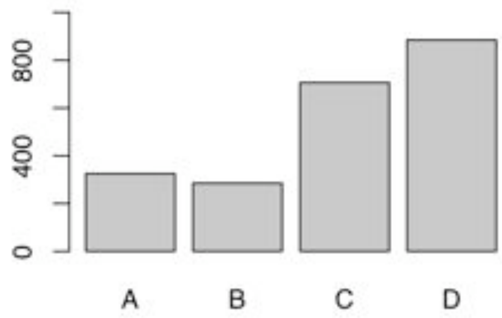
Aufgabentyp 1: Einfache Aufgabe zum elementaren Verständnis einer Graphik

Bei diesem Aufgabentyp geht es darum zu überprüfen, inwieweit der Lernende die grundsätzliche Struktur eines Diagramms verstanden hat.

Der Aufgabentyp ist dadurch charakterisiert, dass nur eine sehr einfache Frage gestellt wird, welche die Kerneigenschaft des Diagramms enthält. Es wird nach der eingeführten Taxonomie nur Wissen abgefragt.

Beispiel:

Aufgabe: Was können Sie aus folgendem Diagramm schließen?



Anzahl in den Gruppen A, B, C, D

- Es gibt mehr in Gruppe D als in jeder anderen Gruppe.
- Es gibt mehr in Gruppe C als in jeder anderen Gruppe.
- Es gibt mehr in Gruppe D als in allen anderen Gruppen zusammen.
- Es gibt weniger in Gruppe D als in jeder anderen Gruppe.
- Keine dieser Antworten.

Erforderliche Kompetenzen

Tabelle 5: Kompetenzschema Aufgabentyp 1

Aufgabe 1	Bildverständnis	Kontextverständnis	Statistikverständnis
Wissen	Barcharts	-	Nominalskala
Anwenden	Barcharts	-	Nominalskala
Analysieren	Barcharts	-	Nominalskala
Synthetisieren	-	-	-
Bewerten	-	-	-

Besonderheiten

Im Pretest hat sich gezeigt, dass selbst einfache Variationen der graphischen Darstellung (z.B. eine andere Art die Koordinatenachsen anzutragen) zu Irritationen bei den Testpersonen führt.

Bewertung

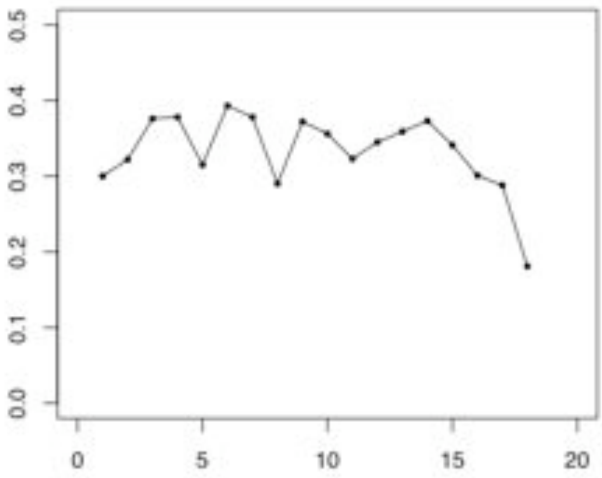
Leider ist es mit diesem Aufgabentyp nur schwer möglich, die Basiskompetenz Wissen global abzufragen. Das liegt daran, dass ein erfahrener Datenanalyst einen Plot wie ein Werkzeug benutzt und seine Eigenschaften – aus der Erfahrung heraus – besser abschätzen kann.

Aufgabentyp 2: Einfache Ausgabe zum fundierten Verständnis einer Graphik

Dieser Aufgabentyp ist dadurch charakterisiert, dass man alle Eigenschaften des Plots kennen muss und aus der eigenen Anwendung heraus abschätzen kann, wo die Defizite des Plots liegen und welche Stellen besonderer Aufmerksamkeit bedürfen.

Beispiel:

Aufgabe : Was können Sie aus folgendem Diagramm schließen?



Zeitreihe von jährlichen Werten

- Die Werte weisen steil nach unten.
- Die Reihe ist über die Jahre meistens auf gleichem Niveau geblieben.
- Der letzte Wert ist ein Ausreißer.

- Die Variabilität ist zu hoch, um Schlüsse zu ziehen.
- Keine dieser Antworten

Erforderliche Kompetenzen

Tabelle 6: Kompetenzschema Aufgabentyp 2

Aufgabe 4	**Bildverständnis**	**Kontext-verständnis**	**Statistik-verständnis**
Wissen	Histogramm	-	Intervallskala, Ausreißer, Verteilung, Symmetrie
Anwenden	Histogramm	-	Intervallskala, Ausreißer
Analysieren	Histogramm	-	Intervallskala
Synthetisieren	-	-	-
Bewerten	-	-	-

Besonderheiten

Für den Aufgabensteller ist es bei dieser Art von Fragen schwer, das richtige Maß zwischen Basiswissen und Anwendungskompetenz zu finden.

Bewertung

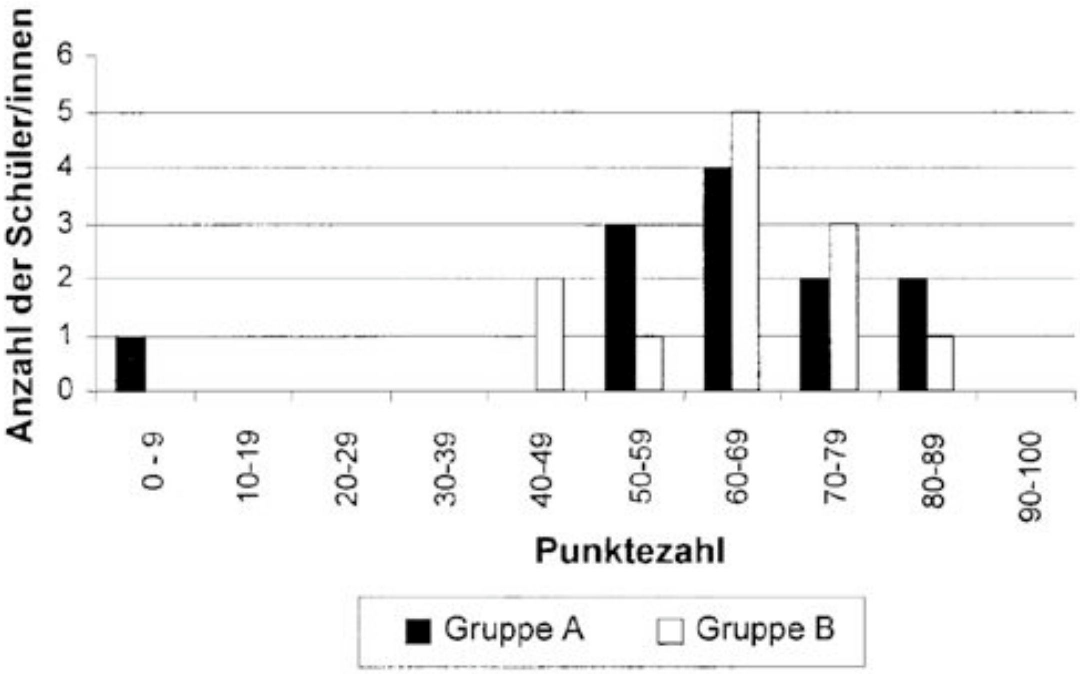
Durch das diffizile Gleichgewicht zwischen Wissens- und Anwendungskompetenz ist es schwierig, ohne ausgiebigen Pretest valide Aufgaben zu stellen.

Aufgabentyp 3: Aufgabe zum fundierten Verständnis einer Graphik und Verwendung theoretischer Argumente

Aufgaben des Typs 3 beinhalten die erste Kompetenzstufe, in der der Plot in eine größere Aufgabenstellung eingebettet ist. Neben dem Plot stehen dem Probanden noch Metainformationen in Form einer Geschichte zur Verfügung. Bei diesem Aufgabentyp soll überprüft werden, ob ein Diagramm in Verbindung mit einem stark didaktisierten Text zum besseren Verständnis beiträgt.

Beispiel:

Aufgabe : Das nachfolgende Diagramm zeigt die Ergebnisse eines Physiktests für zwei Gruppen, die als Gruppe A und Gruppe B bezeichnet werden. Die durchschnittliche Punktezahl von Gruppe A ist 62,0 und der Durchschnitt für Gruppe B ist 64,5. Schüler/innen haben den Test bestanden, wenn ihre Punktezahl bei 50 oder darüber liegt.



Der Lehrer betrachtet das Diagramm und behauptet, dass Gruppe B beim Test besser abgeschnitten hat als Gruppe A. Die Schüler/innen der Gruppe A sind mit ihrem Lehrer nicht einer Meinung. Sie versuchen den Lehrer zu überzeugen, dass Gruppe B nicht unbedingt besser abgeschnitten hat.

Geben Sie ein Argument an, das die Schüler/innen aus Gruppe A benutzen können, indem Sie das Diagramm verwenden.

Erforderliche Kompetenzen

Tabelle 7: Kompetenzschema Aufgabentyp 3

Aufgabe 6	Bildverständnis	Kontextverständnis	Statistikverständnis
Wissen	Histogramm	einfacher, konkreter und sachlich richtiger Inhalt	Intervallskala Ausreißer Verteilung Symmetrie
Anwenden	Histogramm	einfacher, konkreter und sachlich richtiger Inhalt	Intervallskala Ausreißer Verteilung Symmetrie
Analysieren	Histogramm	einfacher, konkreter und sachlich richtiger Inhalt	Intervallskala Ausreißer Verteilung Symmetrie
Synthetisieren	-	einfacher, konkreter und sachlich richtiger Inhalt	Intervallskala, Ausreißer, Verteilung, Symmetrie
Bewerten	-	-	-

Besonderheiten

Dieser Aufgabentyp zeichnet sich dadurch aus, dass sowohl das Diagramm als auch der Text passend zum abgefragten Inhalt erfunden sind. D.h. der zu testende Inhalt, das Diagramm und der Text passen so zusammen, dass es nur eine klar umrissene eindeutige Lösung gibt. Dieser Aufgabentyp entspricht den klassischen Schulbuchbeispielen.

Bewertung

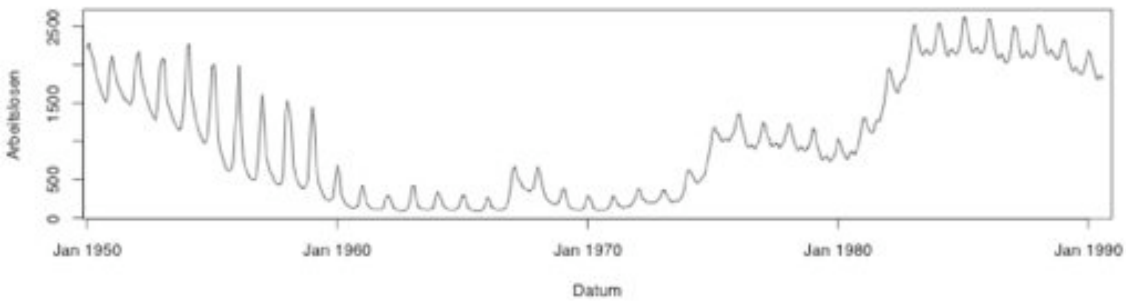
Durch die Ausrichtung dieses Aufgabentyps auf einen klar definierten Inhalt können mit Aufgaben dieser Art die wichtigen Querverbindungen zwischen einzelnen Themenfeldern nicht überprüft werden.

Aufgabentyp 4: Aufgabe zum fundierten Verständnis einer Graphik, eingebettet in einen Kontextzusammenhang

Diese Aufgaben stammen aus einem realen Zusammenhang, d.h. die Daten, Diagramme und Metainformationen stammen aus einem tatsächlichen Erhebungs-Kontext. Der Lernende kann die Informationen in der Aufgabe mit Wissen aus anderen Themenbereichen verknüpfen und dadurch die Lösung der Aufgabe auf einem anderen Themengebiet als der Mathematik auf Plausibilität überprüfen.

Beispiel

Aufgabe: Nachfolgendes Diagramm zeigt die monatlichen Arbeitslosenzahlen für Deutschland vom Januar 1950 bis Juni 1990. Was kann man daraus schließen?



Monatliche Arbeitslosenzahlen in Tausend in Deutschland von 1950 bis 1990.

- Die Saisonalität war in den1950er Jahren sehr hoch.
- Die Arbeitslosenzahlen waren in den 1960er Jahren sehr niedrig.
- Die Wirkungen der Ölpreiskrisen in den 1970er Jahren sind klar zu erkennen.
- Die Saisonalität ist immer gleich geblieben.
- Der langfristige Trend in den Arbeitslosenzahlen zeigt immer nach oben.

Erforderliche Kompetenzen

Tabelle 8: Kompetenzschema Aufgabentyp 4

Aufgabe 16	**Bildverständnis**	**Kontextverständnis**	**Statistikverständnis**
Wissen	Zeitreihendarstellung	einfacher, konkreter und sachlich richtiger Inhalt	Intervallskala, Ausreißer, Trend, Variabilität

Aufgabe 16	**Bildverständnis**	**Kontextverständnis**	**Statistikverständnis**
Anwenden	Zeitreihendarstellung	einfacher, konkreter und sachlich richtiger Inhalt	Intervallskala, Ausreißer, Trend, Variabilität
Analysieren	Zeitreihendarstellung	einfacher, konkreter und sachlich richtiger Inhalt	Intervallskala, Ausreißer, Trend, Variabilität
Synthetisieren	-	einfacher, konkreter und sachlich richtiger Inhalt	Intervallskala, Ausreißer, Trend, Variabilität
Bewerten	-	-	-

Besonderheiten

Da der Grad der Allgemeinbildung bei den unterschiedlichen Probanden individuell verschieden ist, kann man nicht von gleichen Voraussetzungen ausgehen. Es ist möglich, dass ein Proband die Aufgabe auf Grund seines Wissens in einem anderen Themenbereich wesentlich schneller und besser löst und nicht auf Grund seiner Datenanalysekompetenzen.

Bewertung

Es ist schwer, für eine heterogene und unbekannte Gruppe Aufgaben dieses Typs zu erstellen, da unter Umständen das unterschiedliche Vorwissen sehr stark variiert. Dieser Aufgabentyp hat den Vorteil, dass er individuelle Lösungswege ermöglicht und den Lernenden zum eigenen Denken anregt.

Aufgabentyp 5: Aufgabe zum fundierten Kennntis elementarer statistischer Zusammenhänge

Aufgaben folgender Art kommen völlig ohne graphische Darstellung aus. Mit ihnen soll ermittelt werden, inwieweit die Lernenden abstrakte Datenanalysekompetenz besitzen. Die Aufgaben werden so gestellt, dass eine selbst gestellte Planfigur oder Graphik das Lösen der Aufgabe stark vereinfacht. Es liegt am Probanden, die Aufgabe in geeigneter Weise umzustrukturieren.

Beispiel

Aufgabe 12: Ab einem Alter von 40 Jahren empfehlen viele Ärzte ihren Patientinnen an einem Mammographie-Screening (Röntgen-Reihenuntersuchung zur Brustkrebsfrüherkennung) teilzunehmen. Diese Empfehlung beruht auf der Aussage, dass das Mammographie-Screening das Risiko an Brustkrebs zu sterben in dieser Altersgruppe um 25% verringert. Eine Patientin wird nun bei einer Mammographie positiv getestet. Beurteilen Sie den Sachverhalt aufgrund von Tabelle A und Aussage B und beantworten Sie die Frage: Mit welcher Wahrscheinlichkeit wird eine als positiv getestete Frau tatsächlich an Brustkrebs sterben? Beurteilen Sie die Ausgangsthese, dass das Sterberisiko um 25% sinkt?

Tabelle A

Behandlung	Todesfälle (bei jeweils 1000 Frauen)
Kein Mammographie-Screening	4
Mammographie-Screening	3

Aussage B

Bei einem Mammographie-Screening werden 10% der untersuchten Frauen positiv getestet.

Erforderliche Kompetenzen

Tabelle 9: Kompetenzschema Aufgabentyp 5

Aufgabe 11	Bildverständnis	Kontextverständnis	Statistikverständnis
Wissen	-	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Mittelwert, natürliche Häufigkeiten, bedingte Wahrscheinlichkeit
Anwenden	-	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Mittelwert, natürliche Häufigkeiten, bedingte Wahrscheinlichkeit

Aufgabe 11	Bildverständnis	Kontextverständnis	Statistikverständnis
Analysieren	-	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Mittelwert, natürliche Häufigkeiten, bedingte Wahrscheinlichkeit
Synthetisieren	-	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Mittelwert, natürliche Häufigkeiten, bedingte Wahrscheinlichkeit
Bewerten	-	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Mittelwert, natürliche Häufigkeiten, bedingte Wahrscheinlichkeit

Besonderheiten

Bei dieser Art von Aufgaben kann der Proband durch das Anwenden abstrakten statistischen Wissens die Aufgabe rechnerisch lösen.

Bewertung

Bei diesen Aufgaben kann der Proband frei entscheiden, ob er eine rechnerische oder graphische Lösung bevorzugt. Bei der Auswertung ist interessant, inwieweit graphische und analytische Methoden kombiniert werden.

Aufgabentyp 6: Aufgabe zur fundierten Kenntnis graphischer und statistischer Zusammenhänge

Bei folgendem Aufgabentyp soll der Lernende sein Wissen über graphische Darstellungen so anwenden, dass er mögliche Schlüsse und Bewertungen von Graphiken vorgegeben bekommt und diese bewertet. Es muss also eingeordnet werden, ob Schlüsse und Folgerungen plausibel sind oder nicht.

Beispiel

Aufgabe 20: Unter dem Stichwort „Methusalems machen Kasse” war einmal im Handelsblatt zu lesen: „Ein langes Studium zahlt sich in barer Münze aus. Zu diesem überraschenden Ergebnis kommt eine Studie über Einstiegsgehälter von Berufsanfängern, für die die, „Deutsche Gesellschaft für Personalführung 44 Firmen, befragt hat.”

Kann die Aussage richtig sein?

Die dem Artikel zugrunde liegende Untersuchung zeigt eine starke positive Korrelation zwischen Einstiegsgehalt und Semesterzahl.

Wie könnte der zugehörige Scatterplot aussehen?

Erforderliche Kompetenzen

Tabelle 10: Kompetenzschema Aufgabentyp 6

Aufgabe 20	Bildverständnis	Kontextverständnis	Statistikverständnis
Wissen	Scatterplot	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Intervallskala, Verteilung, Regression, Korrelation
Anwenden	Scatterplot	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Intervallskala, Verteilung, Regression, Korrelation
Analysieren	Scatterplot	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Intervallskala, Verteilung, Regression, Korrelation
Synthetisieren	Scatterplot	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Intervallskala, Verteilung, Regression, Korrelation
Bewerten	Scatterplot	einfacher, konkreter und sachlich richtiger unvollständiger Inhalt	Intervallskala, Verteilung, Regression, Korrelation

Besonderheiten

Dieser Aufgabentyp erfordert die größte Datenanalysekompetenz und die größte Kompetenz im Umgang mit Graphiken, denn im Mittelpunkt steht die oberste Kompetenzstufe der Taxonomie, das Bewerten.

Bewertung

Es ist schwierig, solche Aufgaben in realen Kontexten zu finden.

Reflexionfragebogen

Nach der Kompetenzerhebung mussten die Probanden einen Fragebogen bearbeiten, bei dem die Reflexion über die eben bearbeiteten Aufgaben im Mittelpunkt stand. Dieser Fragebogen besteht aus zwei Teilen. Im ersten Teil geht es darum, dass die Lernenden aus ihrer subjektiven Perspektive einschätzen, welcher Aufgabentyp ihnen leicht und welcher schwer fiel. Im zweiten Teil geht es um die Emotionen beim Bearbeiten des Kompetenzfragebogens und die Lernenden sollen zusätzlich einordnen, welche Art von Aufgaben sie mit positiven bzw. mit negativen Emotionen verbinden. Die Antworten werden bei dem an das Interaktionsexperiment anschließenden Interview genauer hinterfragt.

Interaktionsexperiment

Nach den theoretischen Betrachtungen aus Kapitel 3 sind die Observablen Obs.O, d.h. Reaktion des manipulierten Objekts, und Obs.S, d.h. Reaktion des Subjekts, im Lernprozess von Bedeutung und müssen daher möglichst exakt erfasst werden. Praktisch geschieht dies im Versuch mit den Videoaufzeichnungen der Interaktionen des Probanden mit der statistischen Software. Am Anfang des Experiments wurden die Probanden aufgefordert, ihre subjektiven Eindrücke während des Experiments zu verbalisieren. Dies gelingt den Versuchspersonen unterschiedlich gut und hat eine Wechselwirkung auf das Experiment. Die von der Versuchsperson zu untersuchende Problemstellung musste folgenden Bedingungen genügen:

1. Die gestellte Aufgabe sollte einen realen Hintergrund haben.
2. Das verwendete Material sollte nicht „didaktisiert“ vorliegen, sondern die Testpersonen hatten die Aufgabe, sich mit realen Datensätzen und professioneller statistischer Software auseinanderzusetzen.
3. Das Problem sollte eine qualitative Fragestellung beinhalten, d.h. die Lösung der Aufgabe sollte nicht nur durch eine Zahl sondern eine verbal interpretierte Aussage in einem realen Kontext ausgedrückt werden.

4. Die zu bearbeitende Fragestellung sollte so bekannt sein, dass jeder Testperson auf Grund ihrer Allgemeinbildung der Kontext rund um die Fragestellung bekannt ist.

Die obigen Rahmenbedingungen erfüllt der Titanic-Datensatz, der auf die Frage hin untersucht werden soll, ob bei der Schiffsevakuierung Frauen und Kinder bevorzugt wurden. Alle Testpersonen aus der Stichprobe waren mit dem Sachverhalt „Untergang der Titanic” vertraut. Der zur Verfügung gestellte Datensatz enthielt das Merkmal „Klasse” mit den Ausprägungen, „1. Klasse”, „2. Klasse”, „3. Klasse”, „Mannschaft”, das Merkmal „Geschlecht” mit den Ausprägungen „männlich” und „weiblich”, das Merkmal „Alter” mit den Ausprägungen „Kind” und „Erwachsener” und das Merkmal „überlebt” mit den Ausprägungen, „ja” und „nein”. Für die Bearbeitung stand die Software „Mondrian“ zur Verfügung, die alle Probanden bereits aus der Vorlesung kennen. Die Aufgabenstellung lautet:

„Auf ihrer Jungfernfahrt kollidierte die Titanic in der Nacht vom 14. auf den 15. April 1912 gegen 23:40 Uhr mit einem Eisberg und versank zwei Stunden und 40 Minuten nach dem Zusammenstoß im Nordatlantik. Bei Umgebungstemperaturen unter 0 °C starben zwischen 1490 und 1517 von über 2200 an Bord befindlichen Personen. Angesichts der hohen Opferzahl zählt der Untergang der Titanic zu den großen Katastrophen der Seefahrt


THE NEW YORK HERALD.

THE TITANIC SINKS WITH 1,800 ON BOARD; ONLY 675, MOSTLY WOMEN AND CHILDREN, SAVED

Sie haben einen Datensatz zur Verfügung der die Variablen Alter, Geschlecht, Klasse und Überleben enthält. Ausgehend von diesen Daten sollen sie zeigen oder widerlegen, ob Frauen und Kinder bevorzugt Plätze in den Rettungsbooten bekommen haben, wie in obigem Zeitungsartikel behauptet wird.

Erforderliche Kompetenzen

Tabelle 11: Kompetenzschema des Interaktionsexperiments

Stufe	Bildverständnis	Kontextverständnis	Statistikverständnis
Wissen	Barcharts, Spine-plot, Mosicplot, Abfragen, High-lighting, Linking	konkreten und sachlich rich-tigen aber unvollständigen Inhalt verstehen, Text lesen und Wichtiges von Unwich-tigem trennen, Fragestellung verstehen, hinterfragen und bewerten, Metainformation aus der eigenen Erfahrung einfließen lassen	Skalen, Nominalskala, ab-solute und relative Häufig-keiten, bedingte Wahrscheinlichkeit, stocha-stische Unabhängigkeit, Unabhängigkeitstest, stati-stische Modelle
Anwenden	Barcharts, Spine-plot, Mosicplot, Abfragen, High-lighting, Linking	konkreten und sachlich rich-tigen aber unvollständigen Inhalt verstehen, Text lesen und Wichtiges von Unwich-tigem trennen, Fragestellung verstehen, hinterfragen und bewerten, Metainformation aus der eigenen Erfahrung einfließen lassen	Skalen, Nominalskala, ab-solute und relative Häufig-keiten, stochastische Unabhängigkeit, Unabhän-gigkeitstest,
Analysieren	Barcharts, Spine-plot, Mosicplot, Abfragen, High-lighting, Linking	konkreten und sachlich rich-tigen aber unvollständigen Inhalt verstehen, Text lesen und Wichtiges von Unwich-tigem trennen, Fragestellung verstehen, hinterfragen und bewerten, Metainformation aus der eigenen Erfahrung einfließen lassen	Skalen, Nominalskala, ab-solute und relative Häufig-keiten, stochastische Unabhängigkeit, Unabhän-gigkeitstest,
Synthetisieren	Barcharts, Spine-plot, Mosicplot, Abfragen, High-lighting, Linking	konkreten und sachlich rich-tigen aber unvollständigen Inhalt verstehen, Text lesen und Wichtiges von Unwich-tigem trennen, Fragestellung verstehen, hinterfragen und bewerten, Metainformation aus der eigenen Erfahrung einfließen lassen	Skalen, Nominalskala, ab-solute und relative Häufig-keiten

Stufe	Bildverständnis	Kontextverständnis	Statistikverständnis
Bewerten	Barcharts, Spineplot, Mosicplot, Abfragen, Highlighting, Linking	konkreten und sachlich richtigen aber unvollständigen Inhalt verstehen, Text lesen und Wichtiges von Unwichtigem trennen, Fragestellung verstehen, hinterfragen und bewerten, Metainformation aus der eigenen Erfahrung einfließen lassen	Skalen, Nominalskala, absolute und relative Häufigkeiten

Besonderheiten

Aufgaben dieser Art leben von der Wechselwirkung zwischen realen Problemen außerhalb der Mathematik und dem Potential der Explorativen Datenanalyse. Viele Lernende, die zum ersten Mal mit der Explorativen Datenanalyse in Kontakt kommen, sind von dieser Wechselwirkung verblüfft. Es beginnt ein unscheinbarer Datensatz zu sprechen, der mit Standardmethoden nur sehr sperrig zu bearbeiten ist. Gerade an solchen Beispielen wird die Kraft und das Potential der Explorativen Datenanalyse besonders deutlich.

Im Anschluss an das Interaktionsexperiment wurde mit jedem Probanden ein qualitatives Leitfadeninterview geführt. In diesem wurde erfragt, welcher Arbeitsschritte und Aktionen sich der Proband bewusst ist und ob er die ausgeführten Arbeitsschritte begründen und logisch einordnen kann. Zusätzlich wurden die Studierenden nach ihren grundsätzlichen Einstellungen zu Graphiken und mathematischem Problemlösen befragt.

Bewertung

Reale Beispiele zu finden, die obigen Randbedingungen genügen, ist schwierig und abhängig von der zu untersuchenden Stichprobe. Dadurch ist es nicht möglich, standardisierte Tests mit einer großen Zahl von Items zu erstellen.

4.5 Ergebnisse des Studentenexperiment

„Keine Wirkung in der Natur ist ohne Vernunftgrund.
Erkenne den Vernunftgrund,
und du bedarfst nicht des Experiments."
Leonardo da Vinci

Wie im Abschnitt 4.1 über das Forschungsdesign beschrieben, besteht der verwendete Fragebogen zu den Lernstrategien aus drei Teilen. Im ersten Teil geht es darum, die Teilnehmer zu charakterisieren. Der zweite Teil beinhaltet die konkreten Problemlösestrategien beim Lösen mathematischer Probleme. Im dritten Teil beantworten die Probanden einen Fragebogen zur Lerntypeneinteilung. Aufgrund der geringen Anzahl von Probanden wird darauf verzichtet, die vorliegenden Daten mit statistischen Standardmethoden auszuwerten. Die Ergebnisse werden durch eine Kombination von Graphiken und beschreibendem Text ausgewertet.

Fragebogen Teil 1

Die untersuchte Gruppe ist hinsichtlich ihrer schulischen Vorbildung sehr homogen. Alle Probanden haben das Gymnasium besucht und gleich nach der Schul- bzw. Wehrdienstzeit mit dem Studium begonnen. Keiner der Teilnehmer weist Besonderheiten in seinem Bildungsweg auf. Die untersuchten Teilnehmer teilen sich in zwei Gruppen ein: Die eine Gruppe studiert Mathematik/Physik, die andere Mathematik/Sport für das Lehramt an Gymnasien. Um mehr über ihre Interessen zu erfahren, sollten die Probanden zwei ihrer Lieblingsfächer aus der Schulzeit nennen und zwei Gründe anführen, weshalb ein Fach als Lieblingsfach kategorisiert wird. Bei den meisten Studierenden entsprechen die angegebenen Lieblingsfächer der gewählten Fächerkombination im Studium. In Abbildung 45 zeigt der linke Mosaikplot, dass die meisten Befragten Sport und Mathematik begeistert sind und die zweitgrößte Gruppe ihren Schwerpunkt in Physik und Mathematik sehe. Der rechte Mosaikplot weist auf, dass die meisten das Fach Mathematik aus intrinsischen Interesse schätzen und für eine zweite Gruppe sind die Persönlichkeit des Lehrers und die Zensuren von entscheidender Bedeutung.

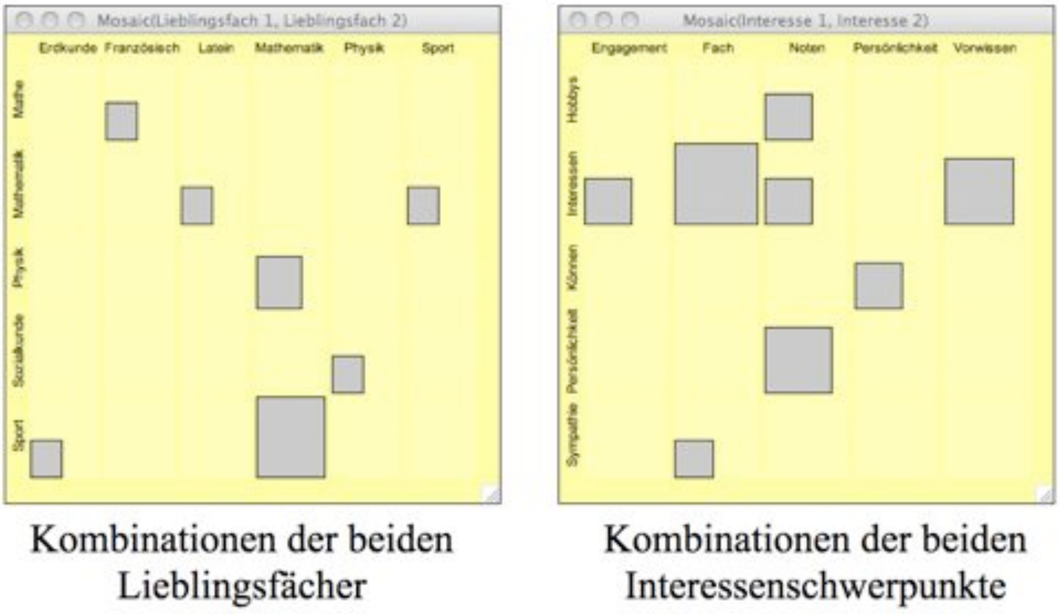


Abbildung 45: Lieblingsfächer und Interessen der Probanden

Für die Auswahl eines Fachs als Lieblingsfach sind verschiedene Faktoren verantwortlich. Neben dem Fach sind insbesondere die Persönlichkeit des Lehrers und die Noten in wichtig. Die Einstellung zu graphischen Darstellungen und zur Mathematik zeigen, dass den meisten Probanden Mathematik schon immer leicht fiel und sie Spaß am strukturierten Denken hatten und haben. Alle Teilnehmer sehen in Graphiken ein nützliches Werkzeug, um sich entweder einen schnellen Überblick zu verschaffen oder sie als Werkzeug bei der Problemlösung zu verwenden. Keiner der Studierenden findet statistische Graphiken überflüssig. Für alle stellen sie eine Hilfe beim Verstehen und Problemlösen dar. Wo der einzelne seinen Schwerpunkt legt und an welchen Stellen Diagramme besonders nützlich für ihn sind, ist sehr individuell.

Fragebogen Teil 2

In zweiten Teil der Untersuchung ging es darum, ob den Probanden die Problemlöseheuristiken, die sie anwenden, bewusst sind und ob sie diese strukturiert einsetzen.

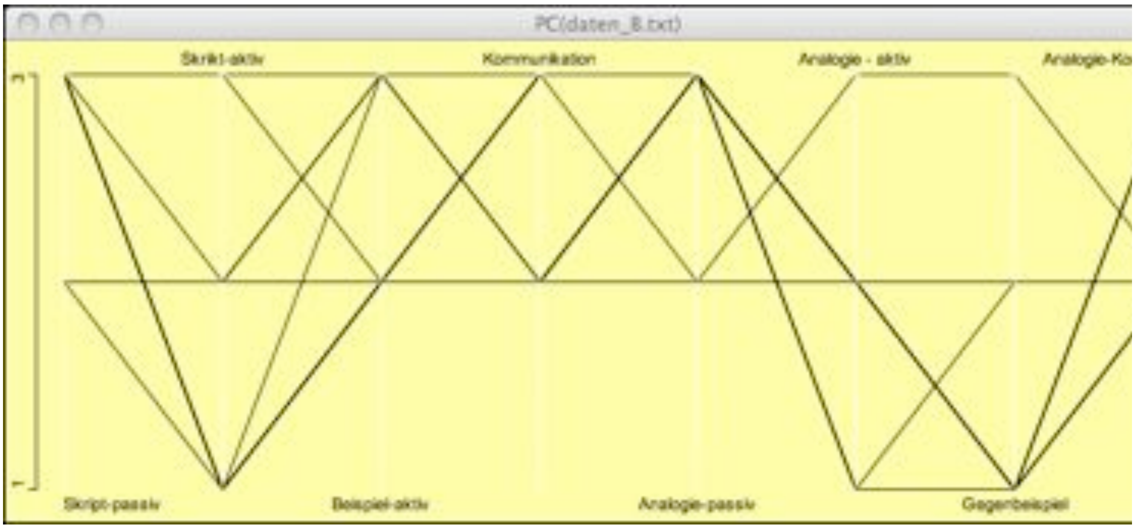


Abbildung 46: Parallelkoordinatenplot zur Lernstrategie

In der untersuchten Stichprobe lassen sich die Lernenden grob in zwei Gruppen einteilen. Die eine Gruppe ist noch sehr im schulischen Alltag verhaftet und gewöhnt, Schulbuchbeispiele mit Musterlösungen zu rechnen. Der andere Teil der Gruppe verwendet verschiedene Problemlöseheuristiken, denen sie sich nur zum Teil bewusst sind. Dieser Eindruck wird auch von dem nachfolgenden Kompetenztest bestätigt und spiegelt sich auch in den qualitativen Interviews wider.

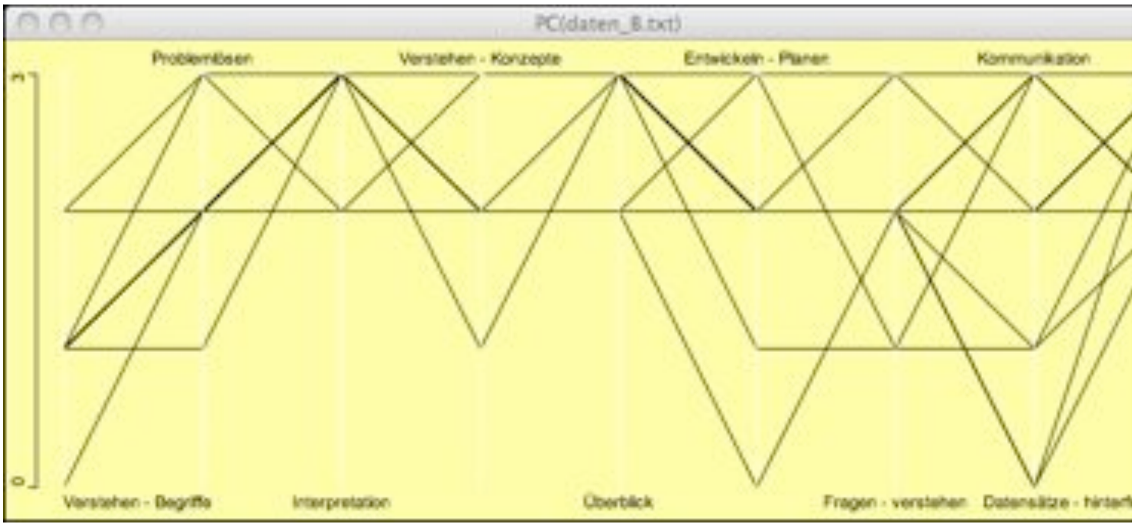


Abbildung 46: Parallelkoordinatenplot zur Lernstrategie - Dieser Parallelkoordinatenplot zeigt auf den parallelen y-Achsen die verschiedenen subjektiven Bewertungen von verschieden Lernstrategien. Die schwarzen Linien verbinden jeden einzelnen Fall.

Fragebogen Teil 3

Der verwendete Lerntypentest geht auf von Frederic Vester (1975) zurück, der vier verschiede Lerntypen beschrieben hat:

1. einen auditiven Typ, der hauptsächlich durch Zuhören lernt.
2. den optisch/visuellen Typ, der hauptsächlich durch Betrachten und Beobachten lernt.
3. den haptischen Typ, der vor allem durch motorische Aktionen lernt, also ein Lerntyp, der alles anfassen, befühlen und bewegen muss, um Dinge zu verstehen.
4. einen rezeptiven/intellektuellen Typ, der vor allem durch Lesen und Nachdenken Dinge begreift.

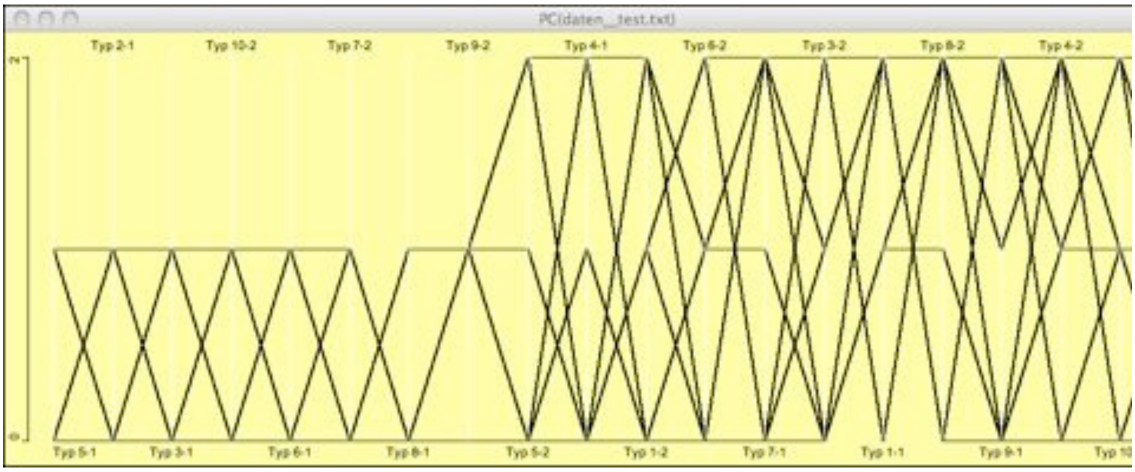


Abbildung 48: Parallelkoordinatenplot zur Lerntypeneinteilung - Dieser Parallelkoordinatenplot zeigt auf den parallelen y-Achsen die verschiedenen Lerntypenfragen. Die schwarzen Linien verbinden jeden einzelnen Fall.

Seit langem ist bekannt, dass diese Lerntypeneinteilung von Vester oder auch differenziertere Typeneinteilungen aus forschungslogischen Gründen schwierig sind. Es gibt Zweifel darüber, ob die Einteilung in Lerntypen überhaupt nützlich und möglich ist, da der Lernprozess ein sehr individueller Vorgang ist. Der Test wurde von Werner Stangl vom Institut für Pädagogik und Psychologie der Johannes Kepler Universität Linz übernommen. http://www.stangl-taller.at/ARBEITSBLAETTER/ . Die erhobenen Daten wurden mit Hilfe der Explorativen Datenanalyse ausgewertet. Es konnte keine eindeutige Typeneinordnung getroffen werden. Bei der vorhergehenden Stichprobe zeigte sich vielmehr, dass die „Typeneinteilung" tendenziell eher von den Fragen abhängt als von den Testpersonen. In der vorliegenden Arbeit wird deshalb auf eine weitergehende Analyse verzichtet.

Auswertung – Kompetenzen

Der Fragebogen enthält sechs Aufgabentypen mit jeweils zugeordnetem Kompetenzniveau nach der in dieser Arbeit entwickelten Taxonomie. Die Basisaufgaben auf der Kompetenzstufe „Wissen" vom Typ 1 können von allen Teilnehmern der Untersuchung problemlos gelöst werden. Demnach verfügen alle Teilnehmer über ein solides Basiswissen über die Grundeigenschaften von Graphiken. Der Aufgabentyp 2 ist gekennzeichnet durch Kenntnis der Basiseigenschaften der Graphiken und deren Schwachpunkte in verschiedenen Anwendungssituationen. Bei diesem Aufgabentyp werden den Teilnehmern keine Kontextinformationen zu den betrachteten Daten zur Verfügung gestellt. Dieser Aufgabentyp wird nicht von allen Teilnehmern richtig beantwortet. Einige Teilnehmer haben die Grenzen und Schwierigkeiten der Plots nicht richtig erkannt, sondern haben die „naive" Auswahlantwort gewählt. Die Aufgaben des Typs 3 stellten für die meisten Teilnehmer keinerlei Schwierigkeit dar. Man kann vermuten, dass sie durch ihre schulische Ausbildung an diesen Typ von Aufgaben gewöhnt sind und diese besser bearbeiten als die Aufgaben des Typs 2, obwohl ihnen dafür mehr Standardwissen abverlangt wird. Aufgaben vom Typ 4, die neben dem Wissen und dem Anwenden von Datenanalysekompetenz auch noch die Analyse des Lösungsweges erfordern und in einen realen Kontext eingebettet sind, fielen den Teilnehmern relativ leicht. Zwei Drittel der Teilnehmer können bei dieser anspruchsvollen Art der Aufgaben eigene kreative und richtige Lösungswege vorweisen. Aufgaben des Typs 5, die ohne Graphiken auskommen und fundiertes datenanalytisches Wissen voraussetzen, stellen zwei Drittel der Teilnehmer vor große Probleme. Nur die Probanden, die sich über Planfiguren eine vernünftige Übersicht verschafften, haben gute Lösungen erzielt. Für die Mehrzahl der Probanden sind die Lösungen dieser Aufgaben durch ein relativ unstrukturiertes Ausprobieren gekennzeichnet. Beim schwierigsten Aufgabentyp der Stufe 6 kann nur noch ein Teilnehmer eine kreative und richtige Lösung finden. Die Bewertung eines Lösungsweges und eine Interpretation stellten die meisten vor unlösbare Probleme.

Auswertung – Reflexion

Mit dem Reflexionsbogen sollen die Teilnehmer mit zwei Fragen kurz einschätzen, was ihnen leicht bzw. schwer fällt. Zwei weitere Fragen sollen die Emotionen klären, die damit verbunden sind. Den meisten fällt es besonders schwer, Antworten, die sie aus Statistik und Mathematik erhalten, richtig innerhalb des Ursprungskontextes der Aufgabe zu interpretieren. Den meisten fällt es leicht, Graphiken zu lesen und Sätze aus der Statistik anzuwenden. Auf der emotionalen Ebene wird das Lesen der Graphiken und das Ausknobeln des richtigen Lösungsweges mit positiven Emotionen, die Interpretation und der Umgang mit fehlenden Angaben mit negativen Emotionen belegt.

Auswertung – Interaktionsexperiment

Als wichtigster Teil der gesamten Untersuchung stellt sich das Interaktionsexperiment heraus. Das liegt daran, dass bei diesem Experiment der Problemlöseprozess des Lernenden direkt beobachtet werden kann und damit der zentrale Vorgang während einer Explorativen Datenanalyse. Insgesamt konnten neun Probanden beobachtet werden.

Versuchsanordnung

Die Probanden können an einem selbst gewählten Termin die Datenanalyseaufgabe bearbeiten. Nach einer kurzen Begrüßung und dem Vorstellen des Versuchsaufbaus nimmt die Testperson ihren Platz ein und beginnt mit dem Lesen der Aufgabe. Als Werkzeuge stehen ein Aufgabenblatt mit der Rahmengeschichte und der Fragestellung zur Verfügung. Der Datensatz und die statistische Software „Mondrian" sind auf einen Apple Macintosh Rechner installiert. Bei einigen Teilnehmern gibt es Probleme, sich mit den Bedienungsbesonderheiten des Apple Macintosh Rechners zurecht zu finden. Wenn dies der Fall ist, werden kurze Instruktionen gegeben. Im Anschluss bearbeiteten die Probanden die Aufgabe und werden dabei von einer Kamera, einem Aufzeichnungsgerät für die Bildschirmaktionen beobachtet. Ein Mikrofon diente den Tonaufnahmen. Alle Signale werden synchron aufgezeichnet und für die Auswertung bereitgestellt. Es besteht keinerlei zeitliche Begrenzung. Jeder Teilnehmer kann nach eigenem Ermessen ausprobieren und bestimmt selbst das Ende des Versuchs. Nach Abschluss der Aufgabenbearbeitung wird mit dem Probanden ein Interview geführt, das die Eindrücke während des Experiments und seine Problemlösestrategie zum Inhalt hat.

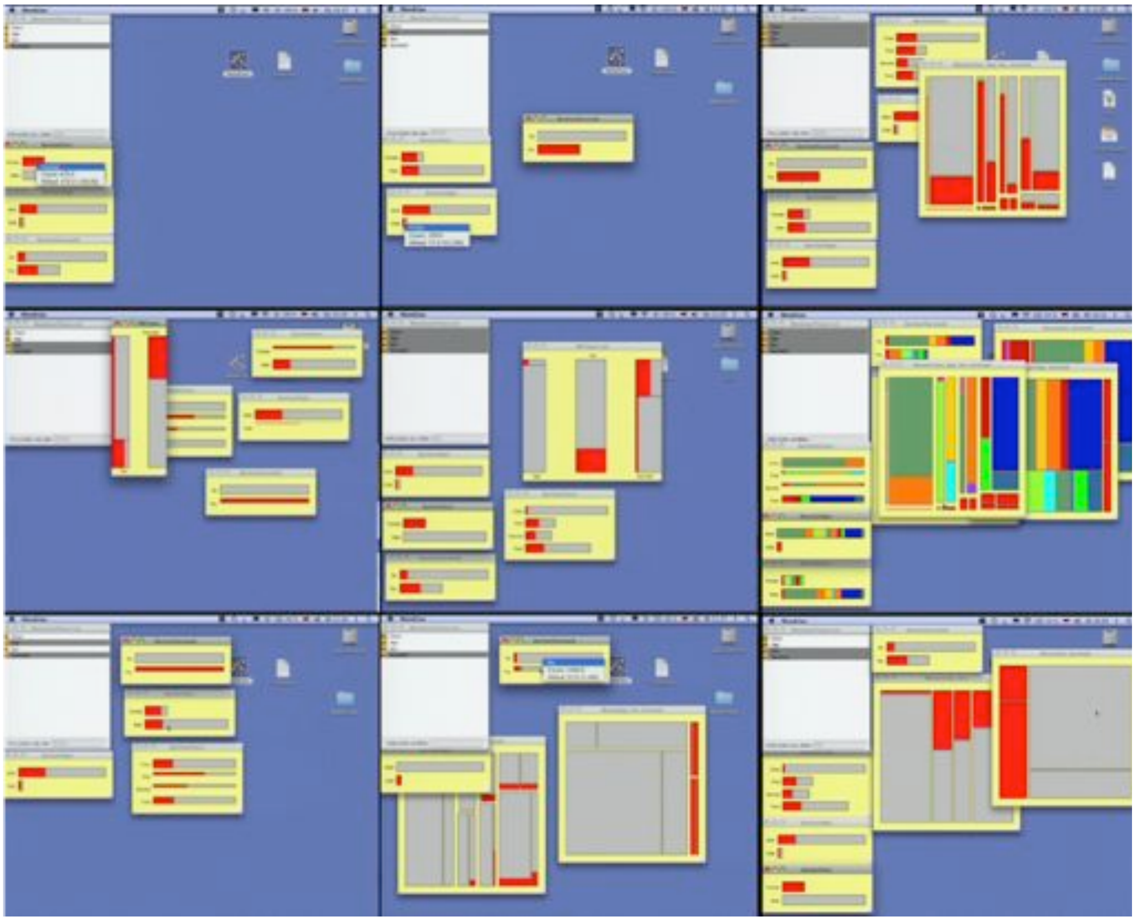


Abbildung 49: Die verschiedenen Lernwege der Probanden

Ergebnisse

Alle neun untersuchten Teilnehmer lösten die Aufgabe auf individuellen Wegen. Kein Lösungsweg gleicht dem anderen. Jeder Proband findet einen für sich geeigneten Lernweg, alle führen zum richtigen Ergebnis. Die benötigte Zeit variiert zwischen fünf und fünfundzwanzig Minuten, wobei kein Zusammenhang zwischen der Basiskompetenz der Teilnehmer und der Bearbeitungszeit festgestellt werden kann. Diejenigen, die sich länger mit dem Datensatz beschäftigen, verfolgen noch „Nebenaspekte", die nicht unbedingt für die Beantwortung der Frage notwendig gewesen sind. Nach dem Experiment haben alle Teilnehmer emotional eine positive Grundstimmung, da es ihnen leicht fällt, die Aufgabe zu lösen und viele beim praktischen Handeln die Zeit aus den Augen verlieren. Trotz dieser stark individualisierten Lösungswege läßt sich ein Grundschema bei allen Probanden beobachten.

Typisches Vorgehen

Erster Schritt: Verstehen der Fragestellung

Abbildung 50: Verstehen der Fragestellung

Folgende Fragen klären die Probanden für sich: Welche Software habe ich zur Verfügung? Wie sieht der Datensatz aus? Wie lautet die Fragestellung? Welche Metainformationen habe ich zur Verfügung? Welche Art von Antwort wird erwartet? Zwei Drittel der Probanden klären diese Fragen durch praktisches Handeln. Sie öffnen z.B. den Datensatz mit dem Texteditor und machen sich mit der Statistik Software vertraut, indem sie diese öffnen und den Datensatz hochladen. Das andere Drittel der Probanden ging eher lesend – rezeptiv vor, d.h. sie müssen nach dem Laden der statistischen Software zunächst in der Hilfe nachlesen, um sich mit dem Funktionsumfang der Software vertraut zu machen. Manche erfragen beim Versuchsleiter die Funktionalität. Keiner aus dieser Gruppe öffnet den Datensatz mit dem Texteditor, um nachzusehen, wie die native Struktur des Datensatzes beschaffen ist.

Zweiter Schritt: Überblick verschaffen

Im zweiten Schritt beginnen alle Teilnehmer sich einen Überblick über den Datensatz zu verschaffen, indem sie einfache Plots Merkmal für Merkmal anwenden. In unserem Fall werden in der Mehrzahl Barchats als primäre Möglichkeit für eine erste Übersicht verwendet. Viele der Teilnehmer wandeln nach kurzen Interaktionstests (d.h. sie markierten Elemente in einem Plot und beobachteten die Auswirkungen auf andere Plots im

Arbeitsbereich) die Barchats in die verwandten Spinplots. So können sie absolute und relative Häufigkeiten direkt nebeneinander vergleichen.

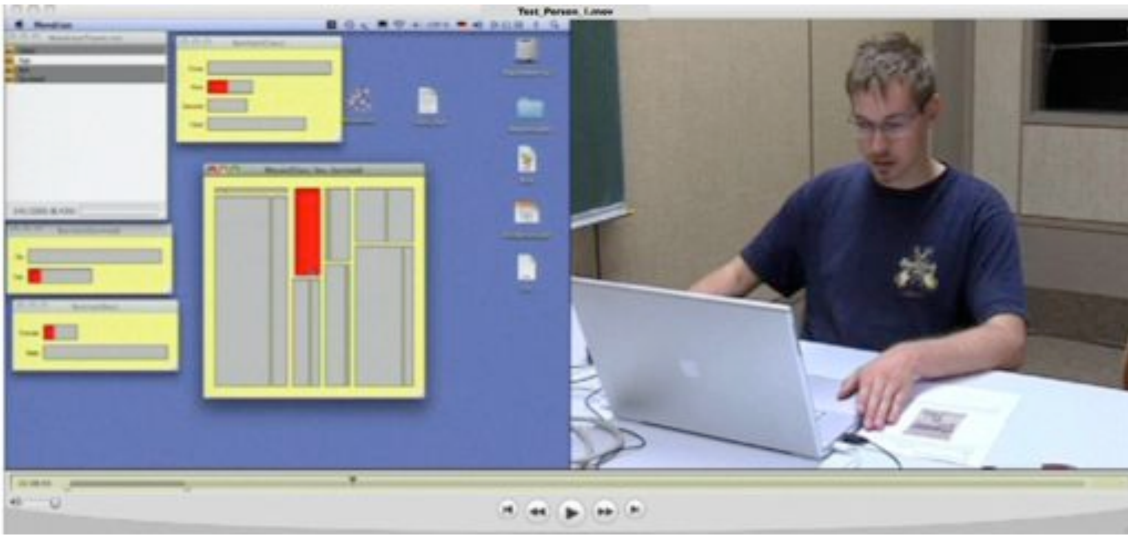


Abbildung 51: Überblick verschaffen

In diesem Schritt gibt es die Möglichkeit, die einzelnen Elemente der Plots abzufragen. Dieses Werkzeug wurde viel genutzt. Gerade bei denjenigen Studenten ist diese Funktion beliebt, die nach eigener Aussage noch wenig Erfahrung mit der Interaktiven Statistischen Graphik besitzen.

Dritter Schritt: Plan schmieden

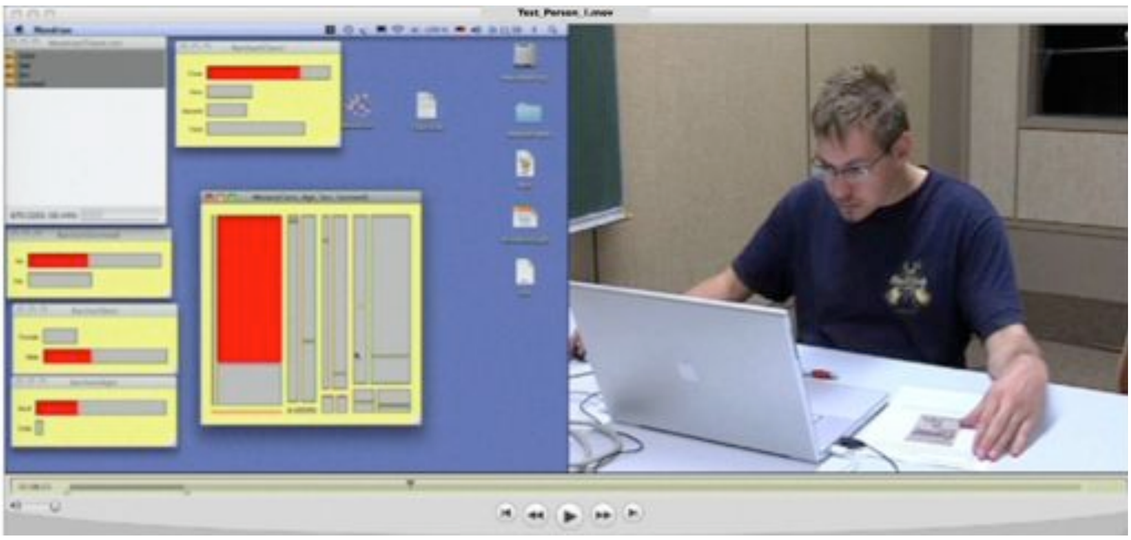


Abbildung 52: Plan schmieden

Der dritte Schritt ist durch eine Phase der Reflexion gekennzeichnet. Die Probanden überlegen, wie sie die verschiedenen Plots kombinieren können. Diese Überlegungen äußeren manche der Teilnehmer auch verbal, so dass sie aufgezeichnet und dokumentiert werden. Einige Teilnehmer verwenden für ihre Überlegungen in dieser Phase das bereitliegende Aufgabenblatt, um sich Werte aus Abfragen zu notieren. Das Ende dieser Phase kann genau beobachtet werden, nämlich dann, wenn das Verhalten der Probanden von einem eher ruhigen Nachdenken in eine zügige Aktion wechselt. Dieser Umschwung im Handeln markiert den Übergang zum nächsten Schritt.

Vierter Schritt: Plan ausführen

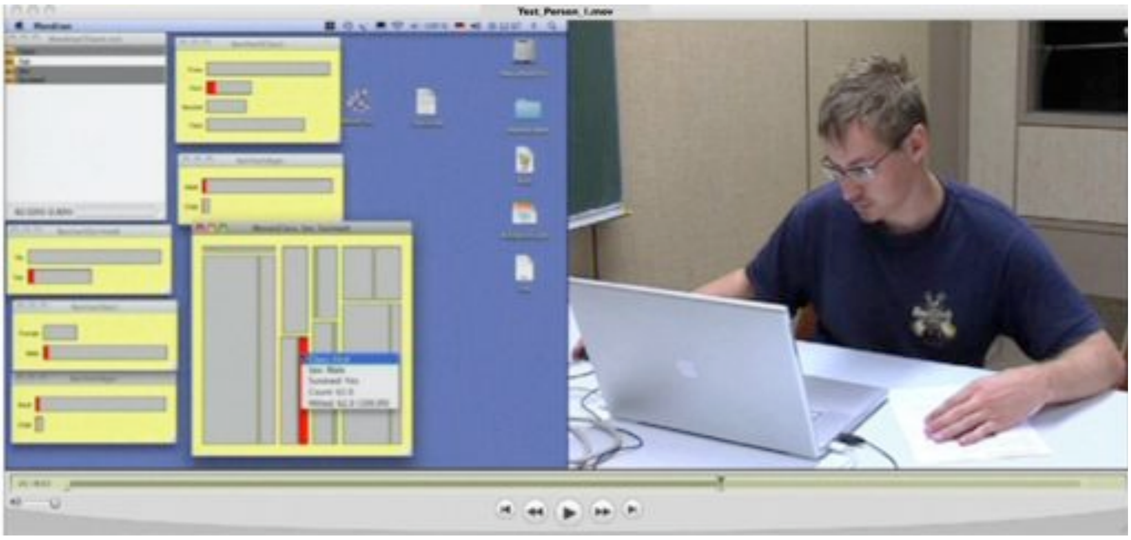


Abbildung 53: Plan ausführen

Der vierte Schritt beginnt mit dem zügigen Umsetzen des Planes. Sehr zielgerichtet werden Plots gezeichnet und miteinander verlinkt. Bei etwa der vier der Teilnehmer kann beobachtet werden, dass zu Beginn dieses Schrittes alle Fenster der bis dahin gezeichneten Plots geschlossen und die neuen Fenster schnell und zielgerichtet erstellt werden. Fast alle Teilnehmer ordnen in diesem Schritt die Plots neu an, d.h. sie verschieben die einzelnen Fenster auf dem Desktop so, dass sie ihren Überlegungen entspricht und optisch einen geordneten Eindruck machte.

Fünfter Schritt: Lösungen ablesen und interpretieren

Abbildung 54: Lösung ablesen

Im fünften Schritt werden die Ergebnisse aus der Datenanalyse mittels Abfragen abgelesen und als Notizen auf dem Aufgabenblatt vermerkt. Viele der Probanden schreiben dann einen kleinen Text, der ihre Überlegungen zusammenfasst, und die Aufgabe beantwortet. Einige der Teilnehmer formulieren ihre Antwort viel ausführlicher, als es die Aufgabenstellung verlangt hätte. Bei etwa der fünf der Teilnehmer gewinnt der Verfasser den Eindruck, dass sie alle Erkenntnisse, die sie gewonnen haben, nun auch mitteilen möchten. Etwa ein drei der Teilnehmer beenden an dieser Stelle die Aufgabenbearbeitung und geben ihr Lösungsblatt ab. Alle anderen der Teilnehmer schließen noch eine kurze Reflexionsphase an.

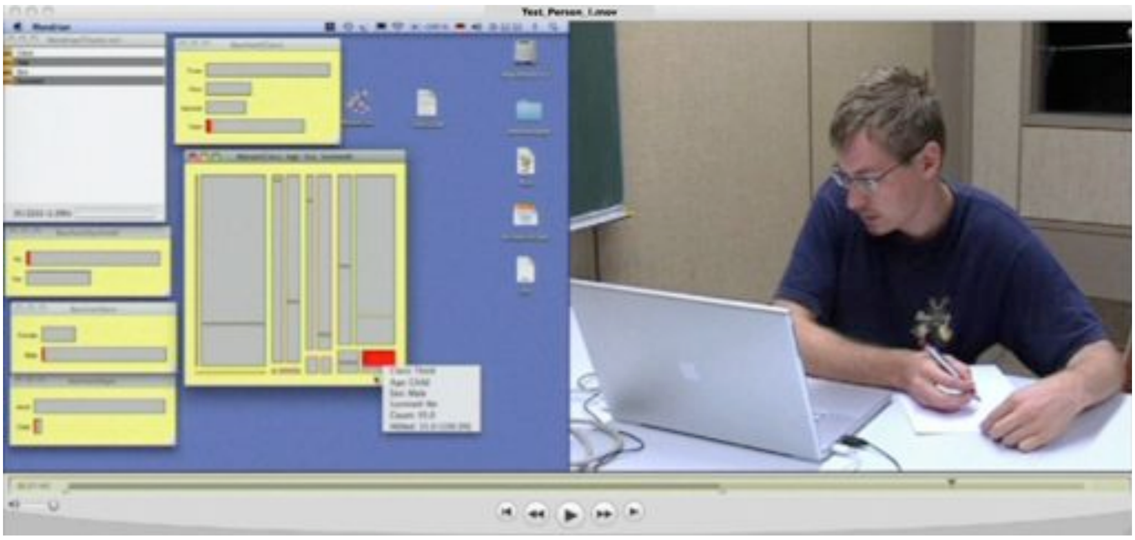


Abbildung 55: Lösung interpretieren

Sechster Schritt: Eigene Lösung kritisch bewerten

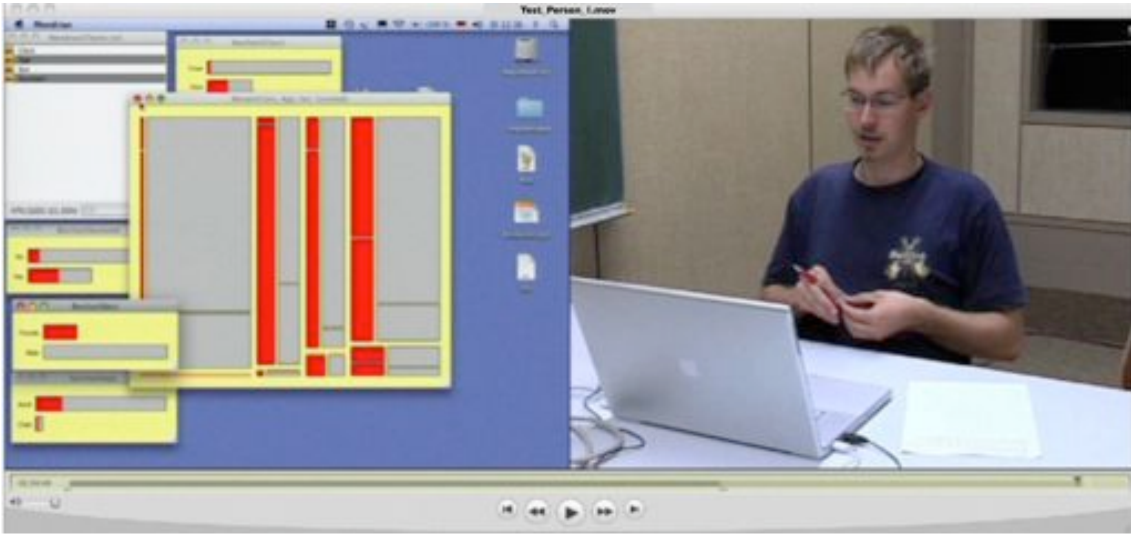


Abbildung 56: Eigene Lösung kritisch bewerten

Im sechsten und letzten Schritt reflektieren die Teilnehmer ihre Lösung, indem sie z.B. versuchen, in irgendeiner Weise „die Probe” zu machen oder sie lesen sich ihre Antwort nochmals durch und verbessern sie.

Ergebnisse der Befragung

Die Fragen sind sehr offen gestellt. Im Laufe des Interviews wird immer konkreter nachgefragt. Die erste Frage lautet: „Wie setzen Sie Graphiken und Diagramme beim Problemlösen ein? Oder kommen Sie ganz ohne graphische Veranschaulichung aus?” Hier eine typische Antwort: *„Also zum Problemlösen, ich beginne erst einmal mit Graphiken, um mir einen Überblick zu schaffen, also wenn ich ein Problem gestellt kriege, dann zeichne ich mir eine Graphik auf, um einen Überblick zu haben, um dann eventuell auch Lösungsskizzen daran zu machen, anhand von diesen dann auch praktisch vorzugehen. Das war jetzt aber eher alles einfache Analysis praktisch, und sonst wenn ich mir jetzt Graphiken anschaue, die ich z.B. gerade gemacht habe für eine Statistik, dann ist das von mir selber schon die Problemlösestrategie bei mir und nicht nur eine Hilfestellung.”*

Alle neun befragten Studierenden finden Graphiken und Diagramme nützlich für den Problemlöseprozess. Zwei von ihnen geben an, dass der Aufwand zum Zeichnen von

Graphiken und Diagrammen so groß sei, dass sie von ihnen nur selten eingesetzt würden.

In der zweiten Teil ging es darum, ob Graphiken beim Lernen eingesetzt werden. Für einen Teil der Studenten war dies völlig selbstverständlich und sie erläuterten die verschiedenen Einsatzmöglichkeiten von Diagrammen, Mindmaps und Planfiguren. Eine typische Antwort in diesem Bereich lautete: *„Ja das ist bei mir so, das visuelle Lernen, man prägt sich eben ein Bild ein und daran erkennt man auch viele Wege praktisch, die man gedanklich gegangen ist, um etwas zu lernen”.* Sieben der Befragten setzen Mindmaps und Planfiguren, die mit Bleistift und Papier gezeichnet werden, für die Problemlösung ein. Das manuelle Zeichnen von Diagrammen und statistischen Graphiken ist nach den Einschätzungen der Meisten händisch zu mühsam. Nur die Studierenden, die als zweites Fach Physik studieren, kennen das *„Von-Hand-Zeichnen”* umfangreicher Graphiken und Diagramme. Eine typische Aussage dazu: *„...im Anfängerpraktikum müssen wir ja auch immer Graphiken machen und die Versuchsergebnisse irgendwie veranschaulichen... Das bringt mir schon was, vor allem macht mir das auch Spaß, zuerst die Graphik zu machen und dann sieht man dann gleich mal, ob es die richtige Steigung hat oder mein Ergebnis ganz daneben liegt.”*

Auch der Einsatz von Graphiken und Diagrammen als Wissensanker wird beschrieben: *„Ja, es hilft schon. Insbesondere in der Physik. Ich finde, man kann sich Graphiken leichter merken als irgendwelche Formeln. Und wenn man sich die Graphik dazu merken kann, dann sieht man eben auch, geht es gegen den Grenzwert oder so, ohne dass man die Formel explizit diskutieren oder im Kopf haben muss.”*

Auf die Frage, wie sie die eben bearbeitete Datenanalyseaufgabe wahrgenommen haben, schildern die meisten Probanden positive Eindrücke. Keine der Testpersonen findet den Umgang mit den Werkzeugen der Interaktiven Statistischen Graphik unangenehm oder ist mit ihnen überfordert. Hier eine typische Aussage: *„Also, ich fand es toll. Ja, man hat eben ohne jetzt viele Zahlen auf den Bildschirm zu haben, hat man finde ich, sehr schnell einen Überblick, wie schaut es aus. Klickt man die Kinder an und sieht dann gleich oben, so viele haben überlebt und so viele sind gestorben. Und ich finde, es macht es deutlich einfacher, wie wenn man da mit Tabellen oder extra für jeden Zusammenhang einzeln die Tabellen erstellen müsste.”*

Zum Thema Interaktivität meinen die meisten Probanden, das es schneller geht und dass das mühevolle Zeichnen von Graphiken per Hand und deren Vernetzung es erst ermöglicht, Graphiken sinnvoll als Problemlösewerkzeug einzusetzen. Eine Aussage dazu lautet beispielhaft: *„Ich finde es geht mit der Auswertung schnell, man kommt schneller zum Ziel”*.

Zusammenfassung

Die Ergebnisse des gesamten Studierendenexperiments lassen sich in folgende Hauptpunkte zusammenfassen:

1. Die Graphiken im Lernprozess werden mindestens auf drei verschiedene Arten verwendet und eingesetzt:
    - als Planfigur, d.h. man vergegenwärtigt sich mit einer Skizze einen mathematischen Sachverhalt.
    - als eine Art Mindmap, die dazu dient Wissen zu ordnen und zu strukturieren.
    - als Diagramm, das den Charakter eines Werkzeugs bzw. einer mathematischen Methode hat.
2. Alle drei Arten der Graphikanwendung haben gemeinsam, dass einem abstrakten Sachverhalt ein anschauliches geometrisches Objekt zugeordnet wird. Die Graphik hat in diesen drei Fällen den Status einer Analogie; d.h. man versucht etwas Unanschauliches auf ein „geometrisches Problem” zu reduzieren.
3. Kein Lernender der untersuchten Gruppe kommt ohne Anschauung aus. Eine Graphik, ein Bild oder ein Beispiel dienen immer als Anker für abstraktere Konzepte.
4. Das Verhältnis von „abstrakt” zu „konkret” muss bei einer Problemstellung, die das Lernen erleichtern soll, im richtigen Verhältnis stehen. Wie der Vergleich der Aufgaben zweiter und dritter Art gezeigt hat, fällt es Lernenden schwerer mit abstrakten Definitionen und Graphiken umzugehen, wenn die Einbettung in einen grosseren Kontext

fehlt. Aufgaben mit einem realen Hintergrund, z.B. das Interaktionsexperiment, haben großen Einfluss auf das Lernen.

5. Das Vorwissen spielt bei der Lösung von Aufgaben, bei denen den Lernenden keine direkte Rückmeldung gegeben wird, eine entscheidende Rolle. Davon abhängig ist, inwieweit schwierige Aufgaben fünfter und sechster Art gelöst werden können. Bekommt der Lernende durch ein System, z.B. die interaktive statistische Graphik, rasch Rückmeldungen von seinen Manipulationen, so können auch die leistungsschwächeren Studierenden schwierige Aufgaben ohne Probleme lösen.
6. Durch den Erfolg beim Lösen einer komplizierten Aufgabe, die schnelle und konkrete Rückmeldung der Interaktiven Statistischen Graphik sowie die Möglichkeit, schnell und mit wenig Arbeitsaufwand eigene Überlegungen prüfen zu können, steigt die Motivation beim Aufgabenlösen.

4.6 Expertenbefragung – Erhebungsinstrumente

„Ein Fachmann ist ein Mann,
der einige der gröbsten Fehler kennt,
die man in dem betreffenden Fach machen kann
und der sie deshalb zu vermeiden versteht."
Werner Heisenberg

Für die Bearbeitung eines Forschungsgegenstandes, über den es noch sehr wenige Untersuchungen und kaum gesicherte Theorien gibt, ist die Expertenbefragung als methodisches Instrument der Erkenntnisgewinnung. An der Universität Augsburg gibt es zwei Arbeitsgruppen, die sich mit statistischer Forschung und Lehre beschäftigen. Eine dieser Arbeitsgruppen setzt die Explorative Datenanalyse für die Lehre ein, die andere Arbeitsgruppe nicht. Beide Arbeitsgruppen halten periodisch abwechselnd den

Vorlesungszyklus „Statistik für Mathematiker" und den Zyklus „Statistik für Lehramtsstudenten".

Für diese Arbeit werden je vier Vertreter der beiden Arbeitsgruppen befragt und deren Einschätzungen miteinander verglichen. Die Experteninterviews werden in Form von Leitfadeninterviews geführt. Der Leitfaden hierfür wird aus den Experimenten mit den Studierenden entwickelt. Themengebiete, die der Experte von sich aus nicht ansprach, werden mit Nachfragen abgearbeitet. Die Interviews wird aufgezeichnet und transkriptiert. Mittels eines Tabellenkalkulationsprogrammes wird das Transkript in Sätze unterteilt und diese durchnummeriert. In drei Stufen findet eine Inhaltsanalyse statt. In der ersten Stufe werden alle Sätze entfernt, in denen der Interviewte von den Fragestellungen abschweift, sich wiederholt oder die der sozialen Interaktion zwischen den Interviewpartnern dient.

In der zweiten Stufe werden die Kernaussagen herausgefiltert, welche die Forschungsfragen betreffen. In der dritten Stufe werden diese Aussagen verallgemeinert, so dass sie zusammengefasst die Aussagen des Experten zur Forschungsfrage sind. Durch die sehr offene Art der Befragung und den Expertenstatus der Interviewpartner sind Antworten, die nur sozialer Erwünschtheit entsprechen, sehr unwahrscheinlich.

Folgende Experten werden befragt:

Name	**Termin/Ort**	**Bemerkung**
Prof. Dr. Adalbert FX Wilhelm Professor of Statistics Commerzbank Chair of Information Management School of Humanities and Social Sciences Jacobs University, Bremen	Datum: 01.07.2008 Beginn: 15:24 Ort: Augsburg	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema
Prof. Antony Unwin, Ph.D. Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse Institut für Mathematik, Universität Augsburg	Datum: 23.07.2008 Beginn: 11:00 Ort: Augsburg	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema
Prof. Dr. Friedrich Pukelsheim Lehrstuhl für Stochastik und ihre Anwendungen Institut für Mathematik, Universität Augsburg	Datum: 11.07.2008 Beginn: 11:00 Ort: Augsburg	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema

Name	Termin/Ort	Bemerkung
Prof. Dr. Lothar Heinrich Lehrstuhl für Stochastik und ihre Anwendungen Institut für Mathematik, Universität Augsburg	Datum: 06.08.2008 Beginn: 11:00 Ort: Augsburg	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema
Dr. Thomas Klein Lehrstuhl für Stochastik und ihre Anwendungen Institut für Mathematik, Universität Augsburg	Datum: 08.07.08 Beginn: 14:00 Ort: Augsburg	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema
Dr. Robert Schmied Professur für Mathematik und Operations Research Fakultät für Elektrotechnik und Informationstechnik Universität der Bundeswehr München	Datum: 24.07.2008 Beginn: 10:00 Ort: Augsburg	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema
Senior Technical Staff Member Dr. Martin Theus O2, München, Statistics Research, AT&T Labs	Datum: 26.07.08 Beginn: 15:00 Ort: München	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema
Dr. techn. ALI ÜNLÜ Dipl.-Math. Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse Institut für Mathematik Universität Augsburg	Datum: 11.07.08 Beginn: 10.30 Ort: Augsburg	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema

Tabelle 12: Befragte Experten

Beispiel einer tabellarischen Ergebniszusammenfassung

Experteninterview		
Zur Person:	Name:	XXXXXXXXXXXX
	Funktion:	XXXXXXXXXXXX
	Status:	XXXXXXXXXXXXXXX
	Institution:	XXXXXXXXXXXX
	Bemerkungen:	XXXXXXXXXXXX
Zum Interview:	Datum:	01.07.2008
	Uhrzeit:	15:24
	Ort:	Augsburg
	Länge:	48:46 min

	Bemerkungen:	keine besonderen Vorkommnisse, keine Gespräche davor oder danach zum Thema
Zusammenfassung		
Lehrtätigkeit:	Fragen:	Welche Statistikkurse haben Sie unterrichtet? Für welche Zielgruppe haben Sie Statistik unterrichtet?
	Stichworte:	Einführung in die deskriptive Statistik, induktive Statistik, explorative Statistik, Wahrscheinlichkeitstheorie, Einführung in empirische Forschungsmethode usw. für Mathematiker für Sozialwissenschaftler für Geisteswissenschaftler für Wirtschaftswissenschaftler für Naturwissenschaftler für Lehrer
	Aussagen:	Ein kompletter Zyklus der empirischen Forschungsmethoden und Statistik in den Sozialwissenschaften. Die Mehrheit der Kursteilnehmer ist der Mathematik abgeneigt.
Vorwissen:	Fragen:	Welches Vorwissen hatten die Teilnehmer?
	Stichworte:	Vorwissen der Kursteilnehmer und deren kulturelle Prägung: Eindeutigkeit von grafischen Darstellungen / Kulturen/ Frauen/Männer, Teilnehmerwissen und Schulvorwissen: vernetztes Mathematikwissen - fragmentiertes Wissen - Vorteile von Studenten, Intuitives Verständnis von Statistischen und Wahrscheinlichkeitstheoretischen Konzepten, teilweise schwieriges Vorwissen aus der Schule, Verständnis von Hintergrundkonzepten .... mathematische Sprache, Irrglaube der Gewissheit
	Aussagen:	Welches Vorwissen hatten die Teilnehmer? Intuitives Verständnis von Statistischen und Wahrscheinlichkeitstheoretischen Konzepten Verständnis von Hintergrundkonzepten und der mathematische Sprache Viele Studierende kommen mit großen Vorurteilen über Statistik und Datenanalyse an die Universität. Es gibt zwei ganz große Probleme: Zum ersten eine große Heterogenität an Vorwissen, das die Studierenden aus der Schule mitbringen. Zum zweiten ist man in der Lehre noch zu sehr versucht davon auszugehen, dass man praktisch bei Null anfangen kann. Muss man erst falsche Konzepte wieder abbauen und neues Verständnis schaffen? Schwierig ist es Alternativen zu dem Schulwissen zu verankern.
Lehrorganisation:	Fragen:	Wie waren die von Ihnen gehaltenen Kurse organisiert?
	Stichworte:	Verhältnis Übung - Vorlesung - Denkstrukturen der Lernenden: Welche Funktion haben in ihren Kursen reale Datensätze?, zur Hand nehmen und lösen von Datenanalyseaufgaben um Konzepte zu festigen, aktives Wissen - passives Wissen, z.B. Sprache, Motivation - praktische Anwendung, Lust auf Problemlösung, Problemlöse- heuristiken

	Aussagen:	Wie waren die von Ihnen gehaltenen Kurse organisiert? Verhältnis zwischen Übung und Vorlesung Die Zweiteilung zwischen Übungen und Vorlesung hat sich bewährt. Die Studierenden müssen individuell Aufgaben bearbeiten und sich so mit den Sachverhalten auseinandersetzen. Übung und Vorlesung sind zwei parallele Stränge, die möglichst miteinander verzahnt sind. EDA, reale Beispiele und Theorie müssen sich ergänzen. Das Verhältnis von aktives und passiven Wissen. Die Studierenden haben durchaus ein großes aktives Wissen in der Anwendung und setzen das gut um. Sie sind aber vergleichsweise unsicher beim passiven Wissen wenn es wirklich um die theoretischen Konzepte und diese ganzen Entwicklungen geht. Für die Studierenden in den Sozialwissenschaften ist der Zugang über die mathematisch theoretische Seite mehr oder weniger komplett unpopulär und auch völlig uninteressant. Wahrscheinlich ist er auch einfach zu abstrakt und zu ungewohnt.
Problemlöse Konzepte:	Fragen:	Wie benutzen Sie interaktive statistische Graphiken bei der Datenanalyse? Worin sehen Sie den Hauptvorteil von Interaktivität von statistischer Graphiken im Problemlöseprozess? Worin sehen Sie den Hauptvorteil von Interaktivität von statistische Graphiken im Problemlöseprozess?
	Stichworte:	Graphik als problemlöse Werkzeuge, Mathematik - kreativ suchendes problemlöse Konzept, Problemlöseung als Haltung in der Mathematik, Lust auf Problemlösung ... Problemlöseheuristiken, Mindemaps - Planfiguren, Sie haben bei meinem Vorgehen keine Funktion, Sie sind für mich ein Werkzeug, um eine Übersicht über die Daten gewinnen, Sie sind für mich ein Werkzeug, der Erstellung von Erstellung von Arbeitsmodellen, Sie sind für mich ein Werkzeug, um Ansatzpunkte für induktive Statistik zu finden.
	Aussagen:	Problemlöseheuristiken: Die Problemlösung als Haltung in der Mathematik Für die Studierenden in den Sozialwissenschaften ist die Haltung der Mathematik zum Probleme lösen eine fremde Welt. Der axiomatische Aufbau ist an Ihr Vorwissen und ihre Grundhaltung nicht anschlussfähig. Der Zugang zu den Problemstellungen muss um erfolg zu haben letztendlich über Beispiele und reale Anwendungen erfolgen. Die Problemlösung als Haltung in der Mathematik Ist das eigentlich für die Studenten nachvollziehbar, dass Mathematik eher so was kreativ suchendes ist oder ist ihnen so ein Zugang fremd, weil sie nur die schönen Schulbuchbeweise oder Buchbeweise kennen. Die meisten Studierenden wissen nichts oder ganz wenig auch über die axiomatische Strukturierung der Mathematik. Das Problem, dass ich manchmal eben auch an der Datenanalyse oder in der Statistik sehe, man bleibt an den Details hängen und kann sich natürlich an den Details auch wunderbar bis zum Umfallen aufarbeiten, aber der Überblick und die Konzepte werden nicht vermittelt.
Kommunikation:	Fragen:	Benutzen Sie interaktive statistische Graphiken als Kommunikationswerkzeug?

	Stichworte:	
	Aussagen:	Benutzen Sie interaktive statistische Graphiken als Kommunikationswerkzeug? Ich mach die Erfahrung, es gibt ein paar Graphiken, ein paar wenige Graphiken die häufig benutzt werden, ich meine ein Balkendiagramm, ein Histogramm, ein Streudiagramm. Viele Studierende sind über das Interpretationsvolumen von so einem komplexe Plots überrascht. Die grundsätzliche Problematik, die ich wirklich sehe ist, dass gerade bei der Statistik, und das ist jetzt unerheblich, ob ich das über Graphiken mache oder ob das über Zahlen geht, letztendlich Studierenden Schwarz-Weiß sehen. D.h., die sind eigentlich zunächst mal überkritisch, wenn sie dann selber Analysen rechnen, vergessen sie, aber dies Hyperkritik, sondern da sind sie dann der Meinungen, oh, das Ergebnis das ich jetzt gefunden hab muss jetzt aber absolut wahr sein. Der Erkenntnisprozess wird nur von wenigen Studenten als philosophisches Problem erkannt. Mindmaps sind beim Lernen im allgemeinen vielleicht eine Hilfe.
Lernkonzepte:	Fragen:	Spielen oder spielten in ihren eigenen Lernprozess Graphiken und Diagramme eine Rolle? Welche Funktion haben in ihren Kursen statistische Graphiken? Welche Funktion haben in ihren Kursen reale Datensätze? Welche Funktion hat in ihren Kursen die interaktive statistische Graphik? Spielen oder spielten in ihren eigenen Lernprozess Graphiken und Diagramme eine Rolle? Worin sehen Sie den Hauptvorteil von interaktiver statistische Graphiken und realer Datensätze im Lernprozess Ihrer Studenten?
	Stichworte:	Sie haben keine Funktion, sie dienen als Anschauungsmaterial, sie dienen als Anker für mathematisches Zusammenhänge, sie dienen zur Illustration von Beispielen, sie dienen zur Motivation von Konzepten, Sie dienen zur Festigung schon gelernter Konzepte, sie dienen zusammen mit realen Datensätzen zur Motivation von Fragestellungen, sie dienen als Problemlösewerkzeug.
	Aussagen:	Welche Funktion haben in ihren Kursen reale Datensätze? In den meisten Fällen ist der Großteil des W-theoretischen-Aufbaus ist für die Statistik meines Erachtens für den Anwender nicht notwendig. Man braucht ein intuitives Verständnis was Wahrscheinlichkeit ist, die Studierenden mit.
Bemerkungen:		

Tabelle 13: Beispiel einer qualitativen Auswertung

Ergebnisse der Expertenbefragung

Wie oben bereits beschrieben werden vier Experten ausgewählt, die die Explorative Datenanalyse aktiv in ihren Vorlesungen und Seminaren benutzen und vier Experten, die zwar die Explorative Datenanalyse kennen, diese jedoch nicht in der Lehre einsetzen. Trotz dieses, für den Untersuchungsgegenstand grundlegenden Unterschieds ergeben sich keine fundamental verschiedene Ansichten darüber, wie der mathematische Lehrbetrieb optimal organisiert werden soll. Eine der wichtigsten Kernaussagen, die alle Experten teilen ist, dass eine Vorlesung, die mathematischen und statistischen Inhalt vermitteln soll, auf jeden Fall mit einer Übung begleitet werden muss. Aus Sicht der befragten Experten ist es für das Erlangen mathematischer Kompetenz unabdingbar, mit den Objekten und Kalkülen der Mathematik direkt umzugehen. Diese Meinung vertreten auch diejenigen Experten, die neben den gängigen Statistikzyklen an deutschen Universitäten auch Statistikvorlesungen für Lehramtsstudenten, Sozialwissenschaftler und Ingenieure geben. Die übereinstimmende Aussage aller Experten ist, dass ohne das *„zur-Hand-nehmen"* von mathematischen Objekten, Sätzen und Kalkülen keine Statistikkompetenz erworben werden kann. Eine typische Aussage für den oben genannten Sachverhalt lautet: *„Das Defizit oder der Schwachpunkt, den ich sehe ist, dass der Vortrag in der Vorlesung eben doch diese Vermittlung oder deduktive Vermittlung des mathematischen Wissens in den Vordergrund stellt."* Die Experten sehen für die Ausbildung von Mathematikern zwei Bereiche, die im Gleichgewicht sein müssen: Das abstrakte, formal geschlossene und abgeleitete Gebäude der Mathematik und die mathematische Anwendung, die sich nicht nur auf „plumpe Textaufgaben" und reale Beispiele bezieht, sondern auch auf Anwendungen innerhalb der Mathematik. Die beiden Bereiche stehen nicht unabhängig nebeneinander, sondern beeinflussen sich gegenseitig sowohl auf der wissenschaftlichen Ebene der Wissenschaft als auch beim Erlernen der Mathematik. Ein Experte beschreibt den Sachverhalt so: *„Also, ich würde es eher als zwei parallele Stränge sehen, die immer wieder möglichst häufig miteinander verzahnt werden. Also, die Vorlesung zielt im Wesentlichen darauf ab, die theoretischen Konzepte zu verstehen..."*. Neben dieser Verzahnung von Vorlesung und Übung sehen die Experten eine Schwierigkeit darin, zwischen Abstraktion und Anschaulichkeit die richtige Balance zu finden. Die Abstraktion ist das Hauptwerkzeug des mathematischen Schließens. Die Ideen und Ansätze für Beweise, Problemlösungen und neue Fragen kommen in erster Linie von einzelnen Beobachtungen unabhängig davon, welchen Abstraktionsgrad diese Beobachtung erreicht. Zu Graphiken und graphischen Darstellungen gibt es unterschiedliche Meinungen darüber, wo und wie diese am besten eingesetzt werden können. Einig sind sich alle Experten darüber, dass eine Zeichnung, eine Planfigur und ein Diagramm im Gedächtnis wie ein Anker wirken, an dem man theoretische Konzepte und praktische Anwendungen festmachen kann. Die Nützlichkeit von Planfiguren innerhalb der Mathematik ist bei allen Experten unumstritten. Alle sehen darin ein geeignetes Mittel, um mathematischen Problemlösungen näher zu kommen. Die größten Kontroversen löst die Frage nach Mindmaps aus. Für einen Teil sind diese ein geeignetes Mittel, um Stoff gut zu strukturieren, während der Rest in ihnen undefinierte Zeichnungen und Zeitverschwendung sieht. Der Wert der Interaktiven Statistischen Graphik als Problemlösewerkzeug in der Forschung wird allgemein anerkannt und respektiert. Über ihren Einsatz in der Lehre ist die Meinung geteilt. Zwei Experten postulieren, das Problemstellungen, bei denen die Interaktive Statistische Graphik ein nützliches Werkzeug ist, kein Teil der Mathematikerausbildung sein sollte. Zwei der Befragten gehen davon aus, dass die Interaktive Statistische Graphik ein großes Potenzial für die Lehre hätte, wenn die Übungen entsprechend gestaltet und betreut wären. Die Hälfte der Experten ist der Meinung, dass die Interaktive Statistische Graphik eine wichtige Ergänzung zur „Standardlehre“ in der Statistik ist. Ähnlich geteilt ist die Meinung über die Verwendung von realen Beispielen in der Statistikausbildung. Für sechs der acht Experten sind reale Beispiele, die in einen Kontext eingebettet sind, eine Bereicherung für die Lehre in der Statistik.

5. Ergebnisse und Ausblick

„Die Kenntnis der Ursachen bewirkt
die Erkenntnis der Ergebnisse."
Cicero

Die Forschungsfrage, wie der Erkenntnisprozess während einer Datenanalyse beschrieben werden kann, lässt sich auf zwei Ebenen beantworten:

1. Durch die Beobachtung der Datenanalysten während des Erkenntnisprozesses wird ein Schema herausgearbeitet, mit dem die Explorative Datenanalyse mit den stattfindenden Aktionen beschrieben werden kann. Auf einer zweiten, tieferliegenden Ebene lässt sich mit Hilfe des Äquilibrationsansatzes von Piaget die Explorative Datenanalyse als Interaktion zwischen einem Subjekt und einem Objekt darstellen. Durch diesen Kreisprozess der Assimilation und Akkommodation kann das Problemlösen während der Explorativen Datenanalyse als Lernprozess beschrieben werden. Mit Hilfe des Eigenbehavior-Ansatzes von Heinz von Foerster kann gezeigt werden, dass sich über mehrere Iterationsschritte eine individuelle kognitive Struktur beim Datenanalysten ausbildet. Stabiles Verhalten wird erkennbar. Aus dieser Überlegung kann allerdings nicht geschlossen werden, dass sich wirklich eine kognitive Struktur herausbildet, wie sie Piaget mit seinem Lernprozess beschreibt. Diese Überlegungen können leider nur eine untere Grenze für die Komplexität des kognitiven Systems angeben. Es ist also durchaus möglich, dass der Datenanalyseprozess in Wirklichkeit „intelligenter" organisiert ist.

Die zweite Forschungsfrage, welche Aspekte im Datenanalyseprozess für den Aufbau eines stabilen Problemlöseverhaltens eine Rolle spielen, kann in dieser Arbeit nur sehr oberflächlich beantwortet werden. Ein wichtiger Aspekt scheint die Direktheit der Interaktionsmöglichkeiten zwischen Individuum (Subjekt) und mathematischem Objekt (Datensatz, Software) zu sein. Zu dieser Fragestellung kann die Wahrnehmungspsychologie mit ihren Gestaltprinzipien nur einen begrenzten Beitrag liefern. Im Abschnitt 5.1 wird gezeigt, dass die Gestaltprinzipien keine Gestaltgesetze sind, sondern stark von individuellen Bedingungen abhängen können. Dies ist um so mehr der Fall, je weniger die visuellen Eindrücke mit der Erfahrung des alltäglichen Lebens zu tun haben. Die schon bekannten Vorgehensheuristiken für den Ablauf einer Explorativen Datenanalyse konnten mit Hilfe von Videoanalysen sowohl bei Experten als auch bei Studierenden bestätigt werden. Die Frage, ob der Lernprozess in Mathematik und Statistik Parallelen aufweist, kann bestätigt werden. Alle theoretischen Herleitungen der Arbeit werden durch die empirischen Befunde gestützt. Auf Grund der kleinen Stichprobengröße und der Kompromisse beim Forschungsdesign des empirischen Teils bleiben offene Fragen, die durch detailliertere Untersuchungen mit viel mehr Probanden geklärt werden müssen.

Abbildung 57: Regelkreis in einer Universitätssituation; Biggs, J.B. (1994b). Student learning research and theory: Where do we currently stand? In G. Gibbs (Ed)., Improving student learning: Theory and practice, (pp. 1-19). Oxford: The Oxford Centre for Staff Development.

Welche Folgerungen kann man aus dieser Arbeit für die Gestaltung eines Lernszenarios ziehen, das die Explorative Datenanalyse als ein Lernwerkzeug einschließt ? Der Fokus muss darauf liegen, einen Rahmen für die Lernaktivität zu schaffen, der für jeden Lernenden individuell angemessen ist. Die Modellierung des Äquilibrationsprozesses als dynamisches System hat gezeigt, dass dieser individuell und von den Anfangswerten sensitiv abhängig sein kann. Daraus erwächst die Forderung, dass der Lernende in seiner Lernumgebung die Möglichkeit hat, individuelle Lernwege zu gehen. Dies setzt im speziellen Fall voraus, dass die Problemstellungen, mit denen die Lernenden konfrontiert werden, offen gestellt werden. Dies hat sich im Interaktionsexperiment bestätigt. Die Abbildung 57 visualisiert grob, welche Regelkreise in einer konkreten Universitätssituation zu beachten sind. Mathematisch gesehen, steckt in einem komplexen System ein anderes komplexes System. Diese verschachtelten Systeme sind schwer zu regeln und im Konkreten schwierig zu untersuchen. Für ein solches System Rahmenbedingungen anzugeben, die es stabilisieren und die Anzahl der Iterationsschritte verringern, müsste Aufgabe einer für eine weiteren Untersuchung sein.

Datenanalysekompetenz ist eine der Schlüsselkompetenzen für die Sozial- und Naturwissenschaften und Grundlage jeden ingenieurwissenschaftlichen Handelns. Deshalb ist es unabdingbar, die Studierenden dieser Fächer so gut wie möglich auf die Herausforderungen während ihrer beruflichen Tätigkeit vorzubereiten. Schulbuchbeispiele helfen hier nur bedingt weiter. Reale Fragestellungen und reale Problemlösekonzepte müssen gelernt und beherrscht werden. Die Datenanalyse und die statistische Basiskompetenz werden in vielen Bereichen nur als formaler Ballast verstanden. Aber gerade die Datenanalyse erfordert Kreativität, präzises theoretisches Wissen und praktische Handlungskompetenz. Wie bereits festgestellt, gehören Mathematik und Statistik zu den unbeliebtesten Fächern an allgemeinbildenden Schulen. Durch neue anwendungsbezogene Ausbildungskonzepte, bei denen die Explorative Datenanalyse elementarer Bestandteil ist, könnte man schon im Studium den Studierenden die Bedeutung der Datenanalyse vor Augen führen und sie für die Statistik begeistern. Da die Mathematik grundsätzlich und die Statistik im Besonderen auf Grund ihrer erkenntnistheoretischen Sonderstellung schwer zu vermittelnde Lehrgegenstände sind, bedarf es besonderer Anstrengung in der Vermittlung. Im Grundschul- und Sekundarstufenbereich hat die didaktische Forschung durch das Konzept der offenen Lernwege schon positive Veränderungen bewirkt. Im Bereich der universitären Mathematikerausbildung wird durch die Kombination von Vorlesung und Übung ein wichtiger Beitrag zu einer handlungsorientierten Mathematikausbildung geleistet. In den Studiengängen, in denen die Mathematik nur Werkzeug ist, besteht meiner Meinung nach noch erheblicher Verbesserungsbedarf. Gerade in den Sozialwissenschaften ist es eine schwierige Gratwanderung zwischen theoretischen Konzepten und statistischer Anwendung, bei der häufig eine fundierte Datenanalysekompetenz auf der Strecke bleibt. In diesen Bereichen könnte die Explorative Datenanalyse mit ihrem Hauptwerkzeug, der Interaktiven Statistischen Graphik, nützliche Impulse geben.

Abbildungsverzeichnis

Tabellenverzeichnis

Literaturverzeichnis

Arnold, M. (2002). *Aspekte einer modernen Neurodidaktik: Emotionen und Kognitionen im Lernprozess. Schriften der Philosophischen Fakultäten der Universität Augsburg.* München: Vögel.

Atteslander, P., & Cromm, J. (2003). *Methoden der empirischen Sozialforschung* (10., neu bearb. und erw. Aufl., 104. - 111. Tsd.). *De-Gruyter-Studienbuch.* Berlin: de Gruyter.

Aulbach, B. (2004). *Komplexe Dynamik,* Universität Augsburg, Augsburg.

Backhaus, K., Erichson, B., Plinke, W., Weiber, R., & Backhaus-Erichson-Plinke-Weiber (2006). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (11., überarb. Aufl.). *Springer-Lehrbuch.* Berlin: Springer.

Baer, M. (2006). *Didaktik auf psychologischer Grundlage: Von Hans Aeblis kognitionspsychologischer Didaktik zur modernen Lehr- und Lernforschung* (1. Aufl.). Bern: h.e.p.-Verl.

Barzel, B. (2008). *Mathematik-Methodik: Handbuch für die Sekundarstufe I und II* (2. Aufl.). Berlin: Cornelsen Scriptor.

Baumert, J. (2003). *Zusammenfassung zentraler Befunde: PISA 2000 - ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland.* Berlin: Max-Planck-Inst. für Bildungsforschung.

Beschreibenede Statistik und explorative Datenanalyse.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals* (1st ed.). New York: Longmans Green.

Bortz & Döring (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler,* from http://dx.doi.org/10.1007/978-3-540-33306-7.

Bühner, M. (2008). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktualisierte und erw. Aufl., [Nachdr.]). *psMethoden/Diagnostik.* München: Pearson Studium.

Cicero, M. T., & Bayer, K. (1993). *Topica: Die Kunst, richtig zu argumentieren ; lateinisch und deutsch.* Darmstadt: Wiss. Buchges.

computerwoche (2008). *Bitkom droht/fordert: Mehr Mathe und Informatik in Schule und Uni: Der IT-Verband Bitkom hat zum "Jahr der Mathematik" gefordert, den Anteil technischer Fächer in Schulen und Hochschulen zu steigern. Bei der Vermittlung der Lerninhalte scheint es aber noch Probleme zu geben,* from IDG Business Media GmbH: http://www.computerwoche.de/job_karriere/1853482/index.html.

Das Reich des Zufalls: Wissen zwischen Wahrscheinlichkeiten Häufigkeiten und Unschärfen (1999). Heidelberg: Spektrum Akad. Verl.

Das Reich des Zufalls: Wissen zwischen Wahrscheinlichkeiten Häufigkeiten und Unschärfen (1999). Heidelberg: Spektrum Akad. Verl.

Edelmann, W. (2000). *Lernpsychologie* (6., vollst. überarb. Aufl.). *Lehrbuch.* Weinheim: Beltz PVU.

Ehmann, M., & Miller, C. (2006). *Geonext für Einsteiger: Dynamische Mathematiksoftware für den Unterricht ; [mit CD-ROM]* (1. Aufl.). *Mathematik - Verstehen* *durch Handeln.* Seelze: Friedrich.

Einstein, A. (2005). *Mein Weltbild:* Ullstein Taschenbuch Verlag.

Fahrmeir, L., & Brachinger, W. (1996). *Multivariate statistische Verfahren* (2., überarb. Aufl.). Berlin: de Gruyter.

Faller, A. (1984). *Der Körper des Menschen: Einführung in Bau und Funktion* (10., überarb. Aufl.). *Flexible Taschenbücher.* Stuttgart: Thieme [u.a.].

Fathom 2: Eine Einführung (2006). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Flick, U. (2007). *Qualitative Sozialforschung: Eine Einführung* (Orig.-Ausg., vollst. überarb. und erw. Neuausg.). *Rororo Rowohlts Enzyklopädie: Vol. 55694.* Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl.

Gassmann, O., & Enkel, E. (2006). Open Innovation: Die Öffnung des Innovationsprozesses erhöht das Innovationspotenzial. *Zeitschrift Führung + Organisation, 75*(3), 132–138.

Gauss, C. F., Bolyai, W., & Schmidt, F. (1972). *Briefwechsel zwischen Carl Friedrich Gauss und Wolfgang Bolyai* (Repr. d. Ausg. Leipzig 1899.). New York: Johnson.

Gauss, C. F., Bolyai, W., & Schmidt, F. (1899). *Briefwechsel zwischen Carl Friedrich Gauss und Wolfgang Bolyai.* Leipzig: Teubner.

Gemoll: Griechisch-Deutsches Schul- und Handwörterbuch (Neubearbeitung) (2006): Oldenbourg, R.

Gigerenzer, G. (2002). *Adaptive thinking: Rationality in the real world. Evolution and cognition.* Oxford: Oxford Univ. Press.

Gigerenzer, G. (2002). *Das Einmaleins der Skepsis: Über den richtigen Umgang mit Zahlen und Risiken* (2. Aufl.). Berlin: Berlin-Verl.

Gigerenzer, G. (2007). *Bauchentscheidungen: Die Intelligenz des Unbewussten und die Macht der Intuition* (1. Aufl.). München: Bertelsmann.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1998). *Das Reich des Zufalls: Wissen zwischen Wahrscheinlichkeiten, Häufigkeiten und Unschärfen:* Spektrum Akademischer Verlag GmbH.

Goethe, J. W. von (2007). *Werke:* Insel Verlag.

Goldstein, E. B. (2008). *Wahrnehmungspsychologie: Der Grundkurs* (7. Aufl.). Berlin, Heidelberg: Spektrum Akad. Verl.

Hans-Joachim Mittag, D. S. *Beschreibenede Statistik und explorative Datenanalyse.*

Harenberg, B. (1988). *Das grosse Personenlexikon zur Weltgeschichte in Farbe.* Dortmund: Chronik-Verl.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: Data mining, inference, and prediction. Springer series in statistics.* New York, NY: Springer.

Haustein, H. D. (2004). *Quellen der Meßkunst: Zu Maß und Zahl, Geld und Gewicht:* de Gruyter.

Heisenberg, W. (2006). *Der Teil und das Ganze: Gespräche im Umkreis der Atomphysik:* Piper.

Huber, O. (2005). *Das psychologische Experiment: Eine Einführung:* Huber, Hans.

Humboldt, A. von (2006). *Kosmos: Entwurf einer physischen Weltbeschreibung:* Eichborn.

Hußmann, S. (2003). *Mathematik entdecken und erforschen: Theorie und Praxis des Selbstlernens in der Sekundarstufe II* (1. Aufl.). Berlin: Cornelsen.

Kadunz, G. (2003). *Visualisierung: Die Verwendung von Bildern beim Lernen von Mathematik* (1. Aufl.). *Klagenfurter Beiträge zur Didaktik der Mathematik: Vol. 3.* München: Profil Verl.

Kandel, E. R. (1996). *Neurowissenschaften: Eine Einführung. Spektrum-Lehrbuch.* Heidelberg, Berlin: Spektrum Akad. Verl.

Krauthausen, G., & Scherer, P. (2001). *Einführung in die Mathematikdidaktik. Mathematik Primar- und Sekundarstufe.* Heidelberg: Spektrum Akad. Verl.

Kütting, H., & Sauer, M. (2007). *Elementare Stochastik: Mathematische Grundlagen und didaktische Konzepte:* Spektrum Akademischer Verlag in Elsevier.

Lakatos, I., & Worrall, J. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery.* Cambridge: University Press.

Lakatos, I., & Worrall, J. (1979). *Beweise und Widerlegungen: Die Logik mathematischer Entdeckungen. Wissenschaftstheorie, Wissenschaft und Philosophie: Vol. 14.* Braunschweig: Vieweg.

Lamnek, S. (2006). *Qualitative Sozialforschung: Lehrbuch* (4., vollst. überarb. Aufl., [Nachdr.]). Weinheim: Beltz PVU.

Leuders, T. (2003). *Mathematik-Didaktik: Praxishandbuch für die Sekundarstufe I und II.* Berlin: Cornelsen-Scriptor.

Mason, J., Burton, L., Stacey, K., & Riederle, M. (2008). *Mathematisch denken: Mathematik ist keine Hexerei* (5., verb. Aufl.). München: Oldenbourg.

Mathematik und plausibles Schliessen (1962). *Wissenschaft und Kultur.* Basel: Birkhäuser.

Mathematik und plausibles Schliessen (1962). *Wissenschaft und Kultur.* Basel: Birkhäuser.

Mathematik und plausibles Schliessen (1963). *Wissenschaft und Kultur.* Basel: Birkhäuser.

Maturana, H. R., & Varela, F. J. (a 1996). *Der Baum der Erkenntnis: Die biologischen Wurzeln des menschlichen Erkennens* (Genehmigte Taschenbuchausg, [Nachdr.]). München: Goldmann.

Paul, D. (2005). *PISA, Bach, Pythagoras: Ein vergnügliches Kabarett um Bildung, Musik und Mathematik:* Vieweg, F.

Pestalozzi, H., & Seyffarth, L. W. (18691895). *Pestalozzi's sämmtliche Werke.* Brandenburg a.H.: Müller [u.a.].

Pinel, J. P. (2007). *Biopsychologie:* Addison Wesley in Pearson Education Deutschland.

Pólya, G. (1949). *Schule des Denkens: Vom Lösen mathematischer Probleme. Sammlung Dalp.* Bern: Francke.

Pólya, G. ((1949)). *Schule des Denkens. (Sammlung Dalp: 36).* Bern Francke.

Reinmann, G. (2007). *Der Nutzen wird vertagt ...: Bildungswissenschaften im Spannungsfeld zwischen wissenschaftlicher Profilbildung und praktischem Mehrwert.* Lengerich u.a: Pabst Science Publ.

Sachs & Hedderich (2006). *Angewandte Statistik: Methodensammlung mit R,* from http://dx.doi.org/10.1007/978-3-540-32161-3.

Schoenfeld, A. H. (1994). *Mathematical thinking and problem solving. Studies in mathematical thinking and learning.* Hillsdale, NJLawrence Erlbaum.

Schopenhauer, A. (2005). *Die Welt als Wille und Vorstellung* (3., verb. und beträchtlich verm. Aufl.). Paderborn: Voltmedia.

Schrage, M. D. (1999). *Serious Play: How the Worlds Best Companies Simulate to Innovate:* Harvard Business School Publishing.

Steland (2007). *Basiswissen Statistik: Kompaktkurs für Anwender aus Wirtschaft, Informatik und Technik,* from http://dx.doi.org/10.1007/978-3-540-74206-7.

Theus, M. (1996). *Theorie und Anwendung interaktiver statistischer Graphik. Augsburger mathematisch-naturwissenschaftliche Schriften: Vol. 14.* Augsburg: Wißner.

Theus, M., & Urbanek, S. (2009). *Interactive graphics for data analysis: Principles and examples.* Boca Raton, Fla.: CRC Press.

Ulm, V. (2008). *Mathematikunterricht für individuelle Lernwege öffnen: Sekundarstufe* (3. Aufl.). *SINUS-Transfer.* Seelze-Velber: Kallmeyer.

Unwin, A., Theus, M., & Hofmann, H. (2006). *Graphics of Large Datasets: Visualizing a Million:* Springer.

Vester, F. (2007). *Denken, Lernen, Vergessen: Was geht in unserem Kopf vor, wie lernt das Gehirn, und wann lässt es uns im Stich?* (32. Aufl., aktualisierte Neuausg.). München: Dt. Taschenbuch-Verl.

Wallrabenstein, W., & Seifert, M. (2001). *Offene Schule - offener Unterricht: Ratgeber für Eltern und Lehrer* (9. Aufl., aktualisierte Aufl. 1994.). Reinbek bei Hamburg: Rowohlt.

Wussing, H. (2008). *6000 Jahre Mathematik: Eine kulturgeschichtliche Zeitreise - 1. Von den Anfängen bis Leibniz und Newton:* Springer Berlin.

Wußing (2009). *6000 Jahre Mathematik: Eine kulturgeschichtliche Zeitreise - 2. Von Euler bis zur Gegenwart.*

Zech, F. (2002). *Grundkurs Mathematikdidaktik: Theoretische und praktische Anleitungen für das Lehren und Lernen von Mathematik* (10., unveränd. Aufl.). Weinheim, Basel: Beltz.

Zimpel, A. F. (2008). *Der zählende Mensch: Was Emotionen mit Mathematik zu tun haben.* Göttingen: Vandenhoeck & Ruprecht.

Historische Dokumente

Carl Friedrich Gauss, Societas Regia Scientiarum Gottingensis, Theoria combinationis observationum erroribus minimis obnoxiae: erroribus minimis obnoxiae, Veröffentlicht von H. Dieterich, 1823, Original von Universität Gent, Digitalisiert am 25. Okt. 2007,90 Seiten

Felix Klein, Über Aufgabe und Methode des mathematischen Unterrichts an den Universitäten, 1898, Leipzig, Teubner

Felix Klein, Protokolle Band 1, 1880

Filme der ASA Statistical Graphics Video Lending Library

(http://cm.bell-labs.com/cm/ms/departments/sia/video-library/index.html)

Multidimensional Scaling
J.B. Kruskal
AT&T Bell Laboratories (1962)

Image of a Thunderstorm
Anne Freeny and John Gabbe
AT&T Bell Laboratories (1966)

Real-time Rotation
Jih-Jie Chang
AT&T Bell Laboratories (1970)

Prim-9
J.W. Tukey, J.H. Friedman and M.A. Fisherkeller
Stanford Linear Accelerator (1973)

Ozone in the Northeast
Richard A. Becker, William S. Cleveland, Beat Kleiner & Jack L. Warner
AT&T Bell Laboratories (1978)

Dynamic Displays of Data
Richard A. Becker and Robert McGill
AT&T Bell Laboratories (1985)

Brushing a Scatter Plot Matrix
Richard A. Becker and Robert McGill
AT&T Bell Laboratories (1985).

Data Analysis Networks in DINDE
R.W. Oldford and S.C. Peters
University of Waterloo (1986).

Brushing and Rotation on an Iris
Richard A. Becker, William S. Cleveland and Gerald Weil
AT&T Bell Laboratories (1987)

Dataviewer: A Program for Looking at Data in Several Dimensions
Andreas Buja and Paul Tukey
Bell Communication Research (1987)

Plot3d: Data Animation on a Sun Workstation
Paul Tukey and Vonn Marsch
Bell Communication Research (1987)

Automatic Tracing of Ice Floes on Satellite Images
Jeff Banfield
University of Washington (1987)

Antelope: Data Analysis with Object Oriented Programming and Constraints
John McDonald
University of Washington (1987)

Use of the Grand Tour in Remote Sensing
John A. McDonald and Steve Willis
University of Washington (1987)

Odds Plots: Finding Associations between Views of a Data Set
Werner Stuetzle
University of Washington (1987)

Higher Hierarchical Views of Statistical Objects
C. Hurley and R.W. Oldford
University of Waterloo (1988)

Graphical Programming
G.D. DesVignes and R.W. Oldford
University of Waterloo (1988)

Treetools
Marilyn Becker, Linda A. Clark, and Daryl Pregibon
AT&T Bell Laboratories (1989)

Visualizing Multivariate Structure with VISUALS/Pxpl
Forrest W. Young & Penny Rheingans
University of North Carolina at Chapel Hill (1990)

Thin Plate Splines and the Analysis of Biological Shapes
Fred L. Bookstein and William Jaynes
University of Michigan Medical Center (1990)

Focusing & Linking as Paradigms for the Visualization of High-Dimension Data
Andreas Buja, Bellcore,
and John McDonald, John Michalak, Werner Stuetzle and Steve Willis
University of Washington (1991)

XGobi: Dynamic Graphics for Data Analysis
Deborah F. Swayne, Dianne Cook, and Andreas Buja
Bellcore (1991)

A Simple Dynamic Graphical Diagnostic Method for Almost Any Model
George S. Easton
University of Chicago (1991)

Visualizing Panel Data
Martin Koschat and Deborah F. Swayne
Bellcore (1992)

Smoother's Workbench
Lise Manchester
Dalhousie University (1992)

Tokyo Data Map
Planned by the Tokyo Government
Produced by Dentsu Inc. and Dentsu Prox Inc. (1992)

SeeNet: See a Network
Richard A Becker, Allan Wilks, and Steven Eick
AT&T Bell Labs (1992)

Edge Information at Landmarks in Medical Images
Fred L. Bookstein and William D. K. Green
University of Michigan Medical Center (1993)

Grand Tour and Projection Pursuit
Dianne Cook, Andreas Buja, Javier Cabrera, and Deborah F. Swayne
Bellcore (1993)

Manufacturing Process Data
William F. Eddy and Audris Mockus
Carnegie-Mellon University (1993)

Exploring Time Series Using Interactive Graphics
Andrew McDougall and Dianne Cook
Rutgers University (1994)

Incidence of Disease Mumps
William F. Eddy and Audris Mockus
Carnegie-Mellon University (1994)

Computer Graphics in Statistics: The Last 30 Years in Brief
Dianne Cook
Iowa State University (1995)

Display of U.S. Air Traffic
William F. Eddy and Shingo Oue
Carnegie-Mellon University (1994)

Spatial CDF Estimation & Visualization with Applications to Forest Health Monitoring
James J Majure, Dianne Cook, Noel Cressie, Mark Kaiser, Soumendra Lahiri, Juergen Symanzik
Iowa State University (1995)

Dynamic Graphics in a GIS: Analyzing and Exploring Multivariate Data
Juergen Symanzik, James J Majure, Dianne Cook
Iowa State University (1995)

An Interactive Environment for the Graphical Analysis of Spatial Data
James J. Majure, Dianne Cook, Juergen Symanzik, and Inna Megretskaia
Iowa State University (1995)

Missing Data in Interactive High-Dimensional Visualization
Deborah F Swayne, Bellcore; Andreas Buja, AT&T Bell Labs (1996)

Lebenslauf

Angaben zur Person

Nachname(n) / Vorname(n)	Fahrner Ulrich
Geburtsdatum	28.02.1969

Berufserfahrung

Zeitraum	2004 →
Beruf oder Funktion	wissenschaftlicher Mitarbeiter am Medienlabor
Wichtigste Tätigkeiten und Zuständigkeiten	Leitung des Studiobetriebs im Medienlabor Leitung des ITS-Teilprojekts "Präsentieren in Forschung und Lehre" Aufbau der Lernplattform Digicampus Konzeption und Durchführung der Seminare: - Einführung in die Videoarbeit - Technische Aspekte der Videoarbeit - Unireport Konzeption und Durchführung der Vorlesung - Grundkurs Statistik für Pädagogen
Zeitraum	1998 - 2004
Beruf oder Funktion	studentische Hilfskraft an der Universität Augsburg
Wichtigste Tätigkeiten und Zuständigkeiten	Betreuung und Aufbau des Macintosh-Pools des Instituts für Mathematik Betreuung und Aufbau des Macintosh-Pools des Sprachenzentrums
Zeitraum	1998 - 2004
Beruf oder Funktion	Übungsgruppenleiter an der Fachhochschule Augsburg
Wichtigste Tätigkeiten und Zuständigkeiten	für die Fächer: - Physik für Elektroingenieure - Physik für Informatiker - Mathematik für Maschinenbauingenieure

Zeitraum	1992 - 1998
Beruf oder Funktion	Berufstätigkeit als Elektroinstallateur
Wichtigste Tätigkeiten und Zuständigkeiten	Automatisierung Netzwerktechnik insbesondre Glasfaserübertragung Medientechnik und Sendetechnik
Zeitraum	1991 - 1992
Beruf oder Funktion	Zivildienst
Schul- und Berufsbildung	
Zeitraum	1974 - 1984
Bezeichnung der erworbenen Qualifikation	Qualifizierender Hauptschulabschluss
Zeitraum	1984 - 1986
Bezeichnung der erworbenen Qualifikation	Ausbildung zum Elektrogerätemechaniker
Zeitraum	1986 - 1988
Bezeichnung der erworbenen Qualifikation	Ausbildung zum Elektrogeräteelektroniker
Zeitraum	1988 - 1989
Bezeichnung der erworbenen Qualifikation	Fachschulreife
Zeitraum	1989 - 1991
Bezeichnung der erworbenen Qualifikation	Fachgebundene Hochschulreife
Zeitraum	1998 - 2005
Bezeichnung der erworbenen Qualifikation	Diplom Mathematiker
Hauptfächer/berufliche Fähigkeiten	Nebenfach Physik Diplomarbeit zum Thema: "Multimediale Einführung in die explorative Datenanalyse"