media/rep/

Repositorium für die Medienwissenschaft

Karl H. Müller; Armin Reautschnig

Die Visuelle Datenanalyse (VDA) in der vergleichenden sozialwissenschaftlichen Forschung

2011

https://doi.org/10.25969/mediarep/19842

Veröffentlichungsversion / published version

Sammelbandbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:
Müller, Karl H.; Reautschnig, Armin: Die Visuelle Datenanalyse (VDA) in der vergleichenden sozialwissenschaftlichen Forschung. In: Theo Hug, Andreas Kriwak (Hg.): *Visuelle Kompetenz Beiträge des interfakultären Forums Innsbruck Media Studies.* Innsbruck: Innsbruck University Press 2011, S. 236–250. DOI: https://doi.org/10.25969/mediarep/19842.

Nutzungsbedingungen:
Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under a Deposit License (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual, and limited right for using this document. This document is solely intended for your personal, non-commercial use. All copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute, or otherwise use the document in public.
By using this particular document, you accept the conditions of use stated above.

DFG Deutsche Forschungsgemeinschaft

Philipps Universität Marburg

Die Visuelle Datenanalyse (VDA) in der vergleichenden sozialwissenschaftlichen Forschung

Karl H. Müller & Armin Reautschnig

Zusammenfassung

Dieser Artikel entwirft eine knappe Skizze für einen neuartigen, aber sehr konsequenzenreichen Methodenbereich innerhalb der Sozialwissenschaften, der als Visuelle Datenanalyse (VDA) bezeichnet wird. Die Visuelle Datenanalyse wurde in den letzten Jahren mit dem Fokus auf große vergleichende Survey- und Paneldatensätze aufgebaut, ihr Anwendungsbereich erstreckt sich aber weit über den Survey- oder Panelbereich hinaus. Der Artikel stellt die Grundzüge dieser Visuellen Datenanalyse vor und vermittelt zwei konkrete Beispiele, welche auch den potentiellen Nutzen wie die komparativen Vorteile dieses visuellen Verfahrens verdeutlichen sollen. Am Schluss werden noch die weiteren Entwicklungspotentiale der Visuellen Datenanalyse zusammengefasst und der neue Zusammenhang von quantitativen, qualitativen und visuellen Methoden erläutert, der als ein triadischer und generativer, d.h. sich wechselseitig selbst erzeugender Komplex betrachtet werden sollte.

In den letzten Jahren ist es zu einem starken Anstieg von visuellen Methoden in den Sozialwissenschaften gekommen, in deren Zentrum Bilder unterschiedlicher Genres – Filme, Fotografien, künstlerische Werke – stehen (Müller 2003). Relativ selten sind hingegen visuelle Ansätze geblieben, gesellschaftliche Daten, Bilder und sozialwissenschaftliche Analysen zu verbinden. Einer der frühen und nicht weiter entwickelten Versuche stellt hier insbesondere Otto Neurath und die von ihm geschaffene „Wiener Methode der Bildstatistik“ und später ISOTYPE (International System of Typographic Education) dar (Müller 1991a,b), allerdings wurde der Neurathsche Ansatz mit seinem Tod im Jahr 1945 ebenfalls nicht mehr weiter entwickelt. Gegenwärtig beschränkt man sich in den meisten Fällen darauf, im Rahmen der normalen sozialwissenschaftlichen Produktion in Zeitschriftenartikeln oder in sozialwissenschaftlichen Büchern datenbasierte Grafiken zu integrieren, die aber über nur wenig kognitiven Mehrwert gegenüber Tabellen oder Beschreibungen verfügen und im Wesentlichen repräsentative Funktionen besitzen. Der folgende Artikel möchte einen knappen Überblick zur Visuellen Datenanalyse (VDA) als einem eigenständigen Cluster von sozialwissenschaftlichen Methoden vermitteln, die nicht *ex post* zur Illustration von Forschungsergebnissen eingesetzt wird, sondern die eigenständig Forschungsergebnisse produziert. Den Anwendungsbereich, in dem diese Visuelle Datenanalyse vorgestellt und erläutert wird, bildet die vergleichende Forschung, die speziell im letzten Jahrzehnt im europäischen Kontext einen dramatischen Aufschwung genommen hat. (Zur Übersicht vgl. Harkness, Van de Vijver & Mohler 2003)

1. Virtuelle Labors als Orte der Visuellen Datenanalyse

Dieser Artikel wird seinen Anfang aber zunächst damit nehmen, einen wissenschaftssoziologisch neuerdings zentralen Begriff einzuführen, nämlich den des Labors, der in einer seiner gängigen Umschreibungen im Wesentlichen das konkrete Setting wissenschaftlicher Arbeits- und Produktionsbedingungen umfasst.

"The laboratory allowed ... to consider the technical activities of science within the wider context of equipment and symbolic practices within which they are embedded ... In other words, the study of laboratories has brought to the fore the full spectrum of activities involved in the production of knowledge." (Knorr 1995, S. 3)

Laboratorien sind gemäß dieser Umschreibung auf allen möglichen Ebenen des szientifischen Tätigkeitsspektrums vorhanden und instanziierbar, von den Teilchenbeschleunigern der Hochenergiephysik bis hin zu den Interviewstudios der Meinungsforschung; von den bench-mark-Laboratorien der Molekularbiologie bis hin zu den Rechenanlagen für komplexe dynamische Systeme – und den Gemächern des philosophischen Diskurses (Knorr-Cetina 1992).

Was den Begriff der virtuellen Laboratorien auszeichnet, betrifft dabei einige besondere Ausstattungsmerkmale, welche für diese Art des Labors konstitutiv werden. Im Wesentlichen lassen sich dazu zwei Bereiche identifizieren:

"A virtual laboratory can be divided into two components: the application programs, data files and textual descriptions that comprise the experiments; and the system support that provides the framework on which these domain-dependent experiments are built." (Prusinkiewicz & Lindenmayer 1990, S. 194)

Zu den basalen Elementen von virtuellen Laboratorien zählen, je nach intendierten Anwendungsbereichen, einerseits eine reichhaltige Programm- und Hypertextumgebung und andererseits hinreichend sophistizierte grafische PC-Workstations samt den entsprechenden grafikfähigen Peripherien. Für virtuelle Laboratorien steht auf der Programmebene derzeit bereits eine Fülle an Verfahren offen, die von biomorphen, ontogenetischen Modellen, evolvierenden Computer-Prozessen oder Netzwerken bis hin zu Robotics-Verfahren, viralen Replikationsformen oder zellularen Automaten reichen und die auf jeweils unterschiedliche Problemstellungen unterschiedlich zufriedenstellende Heuristiken und Methoden bereithalten.

Von virtuellen sozialwissenschaftlichen Laboratorien soll demgemäß dann die Rede sein, wenn die zu analysierenden Prozesse in gesellschaftlichen Milieus angesiedelt sind, seien diese nun auf den Mikroniveaus – auf individuellen Meinungen wie Einstellungen, auf Kommunikationsprozessen, auf Kleingruppeninteraktionen, etc. – oder auch auf den Makrodomänen beheimatet, nämlich auf den gerne als ausdifferenzierte Subsysteme moderner Gesellschaften titulierten Feldern wie Wirtschaft, Politik, Wissenschaft, Kultur und dergleichen mehr. Solche virtuelle sozialwissenschaftliche Laboratorien sind in den letzten Jahrzehnten zunehmend mit der entsprechenden Hard- wie Software ausgestattet worden und verfügen ein über beträchtliches Analyseinstrumentarium im Bereich der vergleichenden sozialwissenschaft-

lichen Forschung, aber in der Regel über keine visuellen Analyseprogramme. (Toš & Müller 2009).

2. Visuelle Datenanalyse als eine eigenständige Methode im virtuellen sozialwissenschaftlichen Labor

Bislang stehen visuelle Komponenten im Kontext virtueller sozialwissenschaftlicher Laboratorien in Form von Grafikprogrammen zur Verfügung, die aber zumeist repräsentative Funktionen besitzen und keine analytischen Aufgaben übernehmen. Üblicherweise werden nach Abschluss der analytischen Tätigkeiten ausgesuchte Ergebnisse eigens in visueller Form repräsentiert.

Was die Visuelle Datenanalyse primär auszeichnet, ist ihr Operationsgebiet als analytische Methode, deren Grundlagen durch die Abbildung 1 dargestellt werden können (vgl. dazu auch Card, Mackinlay & Shneiderman 1999 oder Chen, Härdle & Unwin 2007). Aus dieser Abbildung wird ersichtlich, dass im Kern der Visuellen Datenanalyse ein visuelles Programm – im konkreten Fall WISDOMIZE 2.0 – steht, das sozialwissenschaftlich relevante Daten – im konkreten Fall Survey- oder Paneldaten – in Muster transformiert, die dann seitens einer kompetenten sozialwissenschaftlichen ForscherIn analysiert werden. Der wichtigste Prozess der Visuellen Datenanalyse besteht in der visuellen Interaktion, die einen rekursiv-geschlossenen Prozess bildet und der in der Regel zu stabilen Eigenwerten, nämlich zu speziellen Mustererkennungen führt.

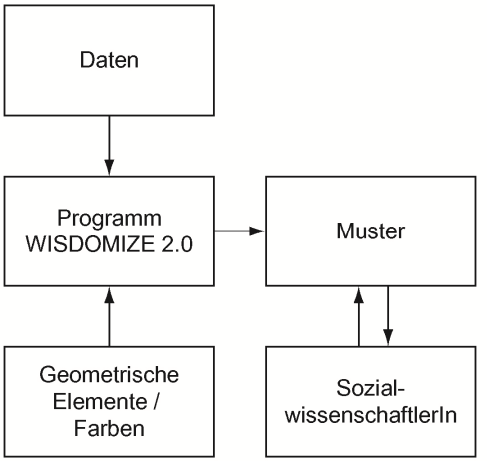

Abbildung 1: Grundlagen der Visuellen Datenanalyse

Die Visuelle Datenanalyse kann auf zwei Wegen eingesetzt werden, nämlich einerseits als generatives Verfahren zur Datenerzeugung und andererseits als transformatorische Methode zur Datenanalyse.

Generative Methoden dienen zur Datenerzeugung und in diesem Segment besitzt die Visuelle Datenanalyse ein bislang nur wenig ausgeschöpftes Potential als Ersatz zu den konventionellen Verfahren der textbasierten Fragebogenkonstruktion in der Survey- oder Panelforschung. Die Abbildung 2 zeigt eine typische Musterkollektion, die im Rahmen der Survey- oder Panelforschung für komplexere Fragestellungen wie die bisherige Lebensgeschichte oder den beruflichen Karriereweg eingesetzt werden kann. RespondentInnen wählen üblicherweise spontan, vollständig und hoch reliabel eines der Muster aus der Abbildung 2, das ihrem bisherigen Lebens- oder Karriereweg am besten entspricht.

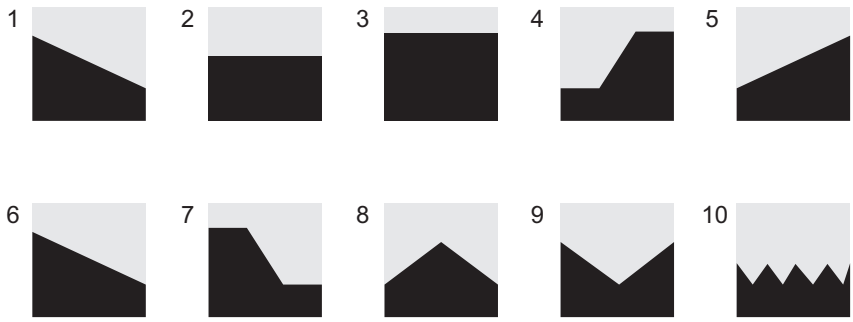

Abbildung 2: Visuelle Muster für den bisherigen Lebensverlauf

Daneben können RespondentInnen gebeten werden, selbst einfache visuelle Muster in ein-, zwei- oder seltener dreidimensionalen Konfigurationen zu erzeugen, beispielsweise die Relevanz von Lebensbereichen auf einer Linie mit den Endpunkten sehr hoher Relevanz einerseits und völliger Irrelevanz. In diesem Sinne können im Prinzip Survey- oder Panelfragebögen konstruiert werden, die sich im großen Ausmaß solcher visueller Muster und Mustererzeugungen durch RespondentInnen bedienen.

Neben der visuellen Datenanalyse als generativer Methode steht ihr Aufgabenbereich als transformatorische Methode, worin Datensets mit Hilfe der Visuellen Datenanalyse in Forschungsergebnisse transformiert werden und die im Zentrum der weiteren Ausführungen stehen wird.

3. Die Visuelle Datenanalyse als Transformationsmethode

Die Visuelle Datenanalyse als Transformationsmethode wird im Weiteren hinsichtlich einer Datenklasse skizziert, die sich in den letzten Jahrzehnten in einer Reihe von sozialwissenschaftlichen Disziplinen in den Vordergrund geschoben hat, nämlich von Survey- und Paneldaten. Die Tabelle 1, die auf Kalkulationen von Willem E. Saris und Irmtraud N. Gallhofer (2007, S. 2f.) basiert, gibt einen Überblick dazu, wie stark diese Daten mittlerweile die Produktion von sozialwissenschaftlichen Zeitschriftenartikeln dominieren.

	Ökonomie (39.4%)	Soziologie (59.6%)	Politik wissenschaften (28.9%)	Sozial-psychologie (48.7%)	Öffentliche Meinung (95.0%)
1949/50	5.7%	24.1%	2.6%	22.0%	43.0%
1964/65	32.9%	54.8%	19.4%	14.6%	55.7%
1979/80	28.7%	55.8%	35.4%	21.0%	90.6%
1994/95	42.3%	69.7%	41.9%	49.9%	90.3%

Tabelle 1: Prozentwerte von sozialwissenschaftlichen Zeitschriftenartikeln mit Survey- oder Paneldaten 1949/50–1994/95

Genauer geht es bei der transformatorischen Visuellen Datenanalyse darum, reichhaltige Muster zu erzeugen, mit denen kompetente BeobachterInnen aus unterschiedlichen sozialwissenschaftlichen Disziplinen interagieren können, um in diesem Prozess wesentliche Eigenschaften von Datensätzen zu erkennen.

Daten → Visuelle Transformation → Muster ↔ SozialwissenschaftlerIn

Einen gewichtigen theoretischen Hintergrund für die Visuelle Datenanalyse bilden demgemäß auch die Kognitionswissenschaften – und hier im Besonderen jene Bereiche, die sich mit Mustererzeugung, Mustererkennung und Wahrnehmung befassen (Foerster 1985, Hofstadter 1982 & 1985, Kosslyn 1980, Osherson, Kosslyn & Hollerbach 1990, Ware 2004), aber auch die Kunsttheorie (Arnheim 1969, Goodman 1973 & 1984) oder das Informationsdesign (Pickover 1990, Thalmann 1990, Tufte 1983 & 1991, Young, Valero-Mora & Friendly 2006).

Von den zugrundeliegenden Datensätzen kann die Visuelle Datenanalyse mit nahezu beliebig großen Datensätzen operieren, die von wenigen Variablen hin zu sehr großen Survey- oder Paneldatensätzen reichen. Was nun diese interaktiven Muster betrifft, so lassen sie sich in insgesamt fünf Hauptgruppen untergliedern, welche für jeweils unterschiedliche Frage- und Problemstellungen geeignet sind:

- Extremwerte – visuelle Muster zur schnellen Entdeckung von Extremwerten und Outliers in einem gegebenen Datensatz
- Abweichungen vom Mittelwert – die rasche Mustererkennung unter- wie überdurchschnittlicher Regionen einer Datenmenge
- Visuelle Kohärenzanalyse – die schnelle Erkennung von Zusammenhängen speziell zwischen mehreren Datengruppen
- Visuelles Clustering – die rasche Spezifizierung von Gruppen oder Typen in einer Datenmenge
- Visuelle Distanzen – große Abstände oder Nähen zwischen einzelnen Datengruppen

Für jede dieser fünf Hauptgruppen steht jeweils eine größere Anzahl von unterschiedlichen Mustern zur Verfügung, die vom Programm WISDOMIZE 2.0 zur Verfügung gestellt werden und die im Prozess der visuellen Datenanalyse forscherseitig ausgesucht werden können.

Jede dieser Hauptgruppen von Mustern kann zudem nach zwei grundsätzlichen Aspekten differenziert werden, nämlich nach statischen und nach dynamischen Analysen. Im statischen Fall steht ein Datensatz zur Verfügung, der zu einem spezifischen Zeitpunkt generiert wurde und über nur einen einzigen Beobachtungszeitpunkt verfügt. Im dynamischen Fall sind hingegen Datensätze vorhanden, die im Minimalfall Beobachtungen für zwei verschiedene Zeitpunkte enthalten, aber im Prinzip sich auch über eine größere Anzahl von solchen Messzeitpunkten erstrecken können.

Weiters wird für die Visuelle Datenanalyse wichtig, dass sie sich auf eine eigene Testtheorie stützen kann, die ausschließlich mit visuellen Mustern operiert. Für jede der Hauptgruppen können nämlich *ex ante* die Bedingungen formuliert werden, unter denen ein Muster als erkennbar oder nicht erkennbar zu qualifizieren ist. Wird nun ein spezielles Muster in einer der Hauptgruppen generiert, so erhält die Forscherseite eine Reihe von Hinweisen, ob die Voraussetzungen für ein bestimmtes Muster gegeben sind – oder nicht. Einer visuellen Testtheorie fehlen zwar die Vorzüge ihres statistischen Pendants in der Angabe von Signifikanzniveaus, etc., doch kann eine solche visuelle Testtheorie so aufgebaut werden, dass sie weitgehend frei von zufälligen oder geschmäcklerischen Musterspezifizierungen bleibt und wertvolle begleitende Hinweise für den Prozess der interaktiven visuellen Datenanalysen offeriert.

4. Die Visuelle Datenanalyse im Kontext der vergleichenden sozialwissenschaftlichen Forschung

Wahrscheinlich den größten kognitiven Mehrwert bringt die Visuelle Datenanalyse im Bereich der vergleichenden Forschung, die in der Regel über Datensätze aus einer größeren Menge von regionalen Einheiten – Staaten, aber auch Regionen oder Städte – verfügt. Die vergleichende Forschung operiert gegenwärtig nahezu ausschließlich mit statistischen Methoden unterschiedlicher Komplexitätsgrade, einschließlich der Konstruktion von einfachen linearen und seltener:

nicht-linearen Modellen. Und genau in diesem Kontext kann die Visuelle Datenanalyse höchst gewinnbringend eingesetzt werden. Im Weiteren sollen zwei typische Beispiele vorgeführt werden, in denen die Vorteile der Visuellen Datenanalyse klar demonstriert werden können.

4.1 Multidimensionale visuelle Kohärenzanalysen

Das erste Beispiel führt in das Feld der visuellen Kohärenzanalysen, die für eine Reihe unterschiedlicher Problemstellungen eingesetzt werden können. Von den Daten her werden mehrere Bereiche von vergleichenden Surveys oder Panels benötigt, wobei sich jeder dieser Bereiche oder Feldern aus mehreren Variablen oder Dimensionen zusammensetzen sollte. Solche Bereiche können sich einerseits aus der Struktur von vorhandenen Surveys oder Panels ergeben, lassen sich im Prinzip aber frei forscherseitig definieren und kompilieren.

Die zwei grundlegenden Forschungsfragen für multidimensionale Kohärenzanalysen lauten, ob sich zwischen zwei oder mehreren Feldern deutliche Kohärenzen – oder Inkohärenzen – erkennen lassen. Je nach erkenntnisleitender Forschungsfrage wird eine visuelle Testtheorie so aufgebaut, dass zunächst forscherseitig die Auswahl Kohärenz beziehungsweise Inkohärenz getroffen wird. Die visuelle Nullhypothese lautet nun, dass – in Abhängigkeit von der gewählten Option – ein konträres Muster vorhanden ist, dessen Eigenschaften durch das Programm selbst – im konkreten Fall WISDOMIZE 2.0 – festgesetzt werden.

Das konkrete Beispiel besitzt als Datengrundlage die verschiedenen Wellen des Europäischen Sozialen Survey (ESS), der vor allem Einstellungen und Lebensbedingungen auf eine qualitativ sehr hochstehende Art abfragt und dessen Daten gegenwärtig als europäische best practice-Beispiele im Bereich der vergleichenden Forschung firmieren. Der ESS lässt sich im Prinzip in eine sehr große Zahl von Feldern segmentieren, die für entsprechende Kohärenzanalysen herangezogen werden könnten. Im vorliegenden Fall wurden wegen einer schon lange laufenden Debatte über objektive und subjektive Survey-Dimensionen die zwei Felder von tendenziell objektiven Arbeitsbedingungen und stärker subjektiven Arbeitsbewertungen ausgewählt, wobei die theoretische Erwartung zunächst dahin geht, dass beide Felder sich durch eine Kohärenzbeziehung auszeichnen.

Für jedes dieser Felder wurden nun die entsprechenden Dimensionen ausgesucht, die zu einem Muster in Gestalt von Abbildung 3 führen. Für die konkrete Darstellung wurden für die Kohärenzanalyse die Abweichungen von einem benutzerdefinierten Referenzwert gewählt. Und als Referenzpunkt wurde der Skalenmittelwert von 5 selektiert, da alle zwölf Dimensionen der beiden Felder eine Skala von 0 bis 10 verwendeten. Demgemäß sieht man in der Abbildung 3 in jedem europäischen Land auf der linken Seite den Bereich der eher objektiven Arbeitsbedingungen (Flexibilität, Möglichkeit der Selbstbestimmung von Arbeitsabläufen, Einfluss auf Arbeitsumgebung, Einfluss auf Entscheidungsabläufe, Möglichkeit auf einen ähnlichen Beruf, Möglichkeit des Selbständigwerdens) und auf der rechten Seite das Feld der stärker subjektiven Arbeitsevaluationen (Arbeitszufriedenheit, Zufriedenheit mit dem Verhältnis von Arbeit und Leben, Beruf als interessante Tätigkeit, Beruf und Stress, Möglichkeit der Arbeitslosigkeit, Bewertung der Entlohnung). Sogar auf den ersten Blick wird in der Abbildung 3 erkennbar,

dass sich das linke Feld quer durch Europa stark differenziert: In den skandinavischen Ländern ist das linke Feld deutlich über dem Skalenmittelwert gelegen, wogegen in den mittel- und osteuropäischen, aber auch in südwesteuropäischen Ländern die entsprechenden Ausprägungen deutlich unter dem Skalenmittelwert liegen. Konträr verhält es sich hingegen beim rechten Feld, das sich quer über Europa nahezu in den gleichen Verteilungen präsentiert.

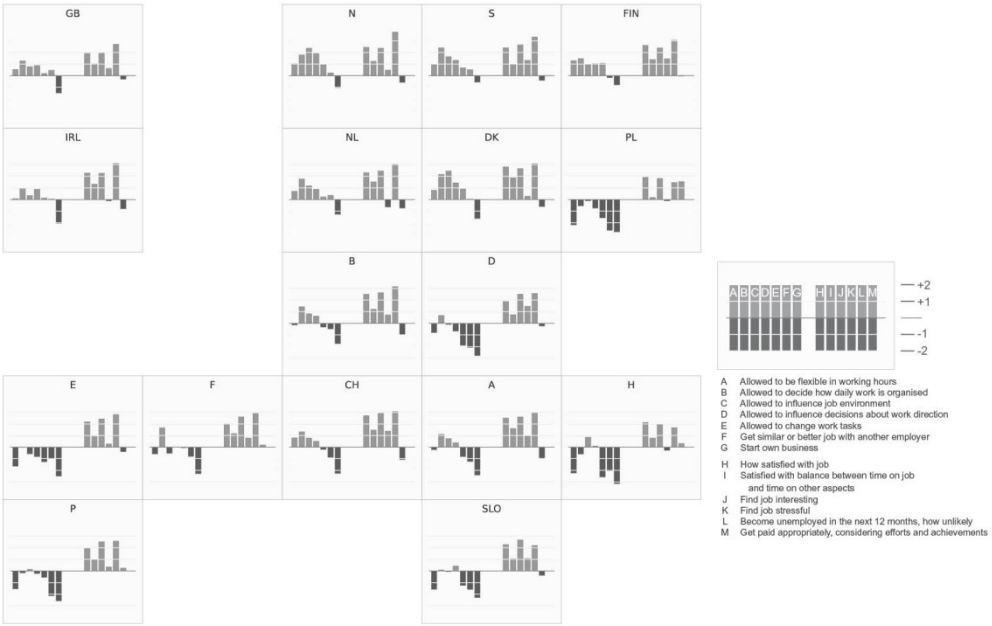

Abbildung 3: Eine visuelle Kohärenzanalyse zwischen subjektiven und objektiven Arbeitsbedingungen in Europa (Quelle: Europäischer Sozialer Survey)

In einer solchen Konstellation – ein Feld mit starken Binnendifferenzierungen, ein zweites Feld mit einem sehr hohen Ausmaß an Homogenität – führt die visuelle Kohärenzanalyse zum Ergebnis, dass keinerlei Kohärenz, auch keine schwache, zwischen den eher subjektiven und den stärker subjektiven Arbeitsbedingungen besteht, ein Ergebnis, das durch die mitlaufende visuelle Testtheorie auch abgestützt wird.

Ein solches Ergebnis kann dann als Ausgangspunkt für weitergehende visuelle oder statistische Explorationen genommen werden, aber auch für theoretische Überlegungen, wie sinnvoll die Inklusion eines Feldes ist, wenn sich die Werte nahezu in gleicher Weise quer über Europa reproduzieren, obwohl die jeweiligen Arbeitsumgebungen – das linke Feld in der Abbildung 2

– durchaus unterschiedlich gestaltet sind und auch sehr unterschiedlich wahrgenommen werden.

4.2 Multidimensionales Visuelles Clustern

Das zweite Beispiel eines multidimensionalen Clusterns setzt in der Regel komplexere Datenmengen voraus. Die Datengrundlagen bestehen normalerweise aus m Feldern mit n Dimensionen für o unterschiedliche regionale Einheiten. In ihrer visuell am leichtesten erkennbaren Form können Felder und Dimensionen quadratisch angeordnet sein, es sind aber auch Rechteckformen sowie – allerdings um den Preis ungleich schwierigerer Mustererkennungen – unregelmäßige Vielecke spezifizierbar. Im vorliegenden Fall wurde – wieder auf dem Hintergrund der Daten für den Europäischen Sozialen Survey – eine 4 x 6 Darstellung ausgesucht – vier Felder mit jeweils sechs verschiedenen Dimensionen. Diese Felder betreffen die Bereiche:

- Arbeitsbedingungen (Flexibilität, Möglichkeit der Selbstbestimmung von Arbeitsabläufen, Einfluss auf Arbeitsumgebung, Einfluss auf Entscheidungsabläufe, Möglichkeit auf einen ähnlichen Beruf, Möglichkeit des Selbständigwerdens)
- Arbeitsbewertungen (Arbeitszufriedenheit, Zufriedenheit mit dem Verhältnis von Arbeit und Leben, Beruf als interessante Tätigkeit, Beruf und Stress, Möglichkeit der Arbeitslosigkeit, Bewertung der Entlohnung)
- Vertrauen in Institutionen (Vertrauen in das nationale Parlament, in das Rechtssystem, in die Polizei, in PolitikerInnen, in politische Parteien, in das Europäische Parlament)
- Subjektiver Gesamtzustand (Vertrauen in andere, Hilfe durch andere, Lebenszufriedenheit, Glück, Soziale Teilnahme, Gesundheit)

In diesem Fall wird visuell nach Clustern von Ländern gesucht. Die visuelle Testtheorie wird in diesem Fall so aufgebaut, dass forscherseitig zunächst eine bestimmte Anzahl von Clustern festgelegt wird und dass benutzerseitig ein Vorschlag für eine Gruppenbildung mit der selektierten Gruppenzahl vorgenommen wird. Der Test in diesem Fall besteht darin festzustellen wie gut das forscherseitig vorgenommene Clustering ausfällt – und ob bessere Alternativen für die gewählte Anzahl von Clustern zur Verfügung stehen.

Und damit kann zur Abbildung 3 übergeschwenkt werden, welche ein 4 x 6 Muster für die einzelnen ESS-Länder reproduziert, das sich im Prinzip visuell in zwei, drei oder mehrere Cluster separieren lässt. Dieses Muster wurde so konstruiert, dass für jede einzelne Dimension eines Feldes die Länderverteilung in drei annähernd gleich große Gruppen separiert wird und die Farbgebung diese Gruppenseparierung wiedergibt: weiß für das obere, grau für das mittlere und schwarz für das untere Länderdrittel in einer spezifischen Dimension. Diese Separierung wird für jede der ausgewählten Dimension eines Feldes sowie über alle Felder durchgeführt, so dass jedes Land als ein Muster mit insgesamt 4 x 6 weißen, grauen und schwarzen Flächen erscheint.

An Hand der Abbildung 3 lässt sich dann beispielsweise ein visuelles Clustern mit drei Gruppen durchführen, das benutzerseitig zwar in unterschiedliche Sets separiert werden kann, im vorliegenden Fall aber zu den folgenden Clustern führt:

- Cluster 1: Norwegen, Schweden, Finnland, Dänemark, Niederlande, Schweiz
- Cluster 2: UK, Irland, Frankreich, Belgien, Slowenien, Österreich
- Cluster 3: Portugal, Spanien, Polen, Ungarn, Deutschland

Wie immer das Ergebnis des visuellen Clusterns ausfällt, man erhält forscherseitig einerseits eine Rückmeldung, ob das gewählte Clustern konsistent durchgeführt wurde und ob sich im Prinzip homogenere Cluster spezifizieren ließen. In diesem Fall läuft die visuelle Testtheorie über eine Gesamtbewertung pro Land, worin die Verteilung der hellen, grauen und dunklen in Zahlenwerte (-1, 0, +1) transformiert wird und daraus ein Gesamtwert errechnet wird, der seinerseits zu einer Länderverteilung führt, die dann – in Abhängigkeit von der forscherseitig vorgenommenen Anzahl von Clustern – zu einer spezifischen Clusterbildung führt, die mit den forscherseitig vorgenommenen Clustern numerisch verglichen werden kann.

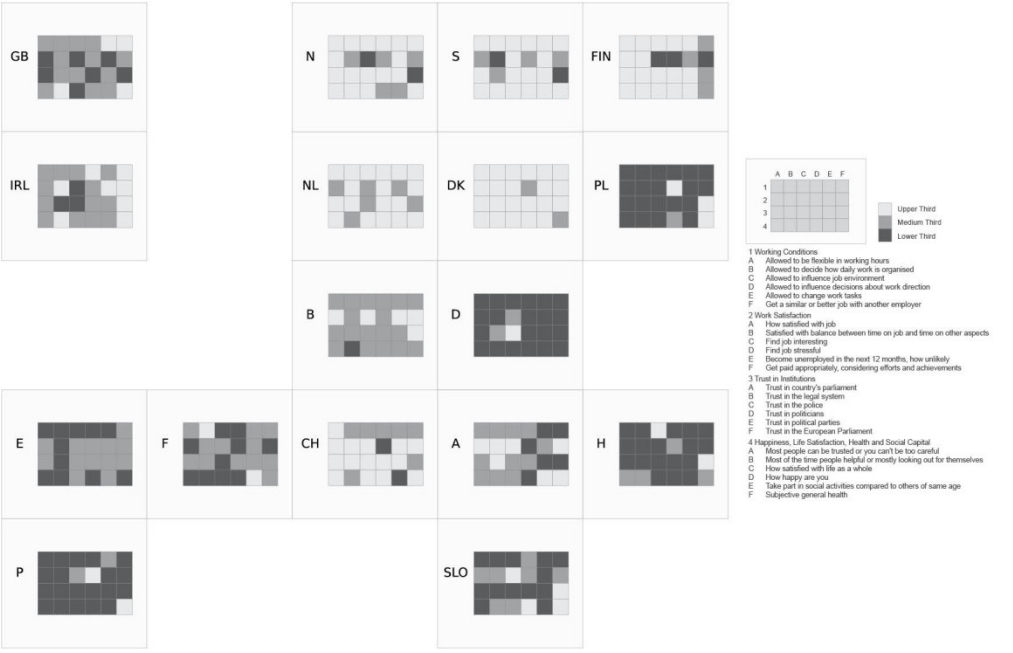

Abbildung 4: Visuelles Clustern mit vier Feldern und jeweils sechs Dimensionen (Quelle: Europäischer Sozialer Survey)

Damit wären zwei Beispiele für die Visuelle Datenanalyse im Kontext der vergleichenden Forschung etwas näher vorgestellt.

5. Komparative Vorteile der Visuellen Datenanalyse

Mit den visuellen Kohärenzanalysen und dem visuellen Clustern wurden konkrete Anwendungsfälle der Visuellen Datenanalyse gezeigt, die gleichzeitig auch den Nutzen dieser Methode vor Augen führen. Zusammenfassend lassen sich die potentiellen Vorteile der Visuellen Datenanalyse im Kontext der vergleichenden Forschung in folgende Punkte zusammenfassen.

Der wahrscheinlich gewichtigste Vorteil liegt darin, dass die Visuelle Datenanalyse eine sehr schnelle Einsicht in grundlegende komplexe Dateneigenschaften vermittelt. Speziell das Beispiel mit dem visuellen Clustern verdeutlicht, dass an sich komplexe multidimensionale Daten sehr schnell in entsprechende Muster transformiert werden können, die dann sofort visuell analysiert werden können. Im konkreten Beispiel wurden 4x6 Dimensionen verwendet, es könnten aber – je nach Datensätzen und Forschungsinteressen – auch 7x7 oder 12 x 12 Felder und Dimensionen sein, mit denen eine Visuelle Datenanalyse vergleichsweise schnell operieren kann.

Ein weiterer gewichtiger Vorteil der Visuellen Datenanalyse liegt darin, dass sie auf einem höchst komplexen und differenzierten System aufbaut, nämlich der menschlichen Wahrnehmungsfähigkeit mit ihren praktisch unendlich feinen Distinktionen. Dieser Punkt ist auch deswegen so interessant, weil die Visuelle Datenanalyse durchaus immer wieder auch nicht-intendierte Resultate zu generieren vermag, weil man beispielsweise forscherseitig in einem Muster für visuelles Clustern auch andere relevante Eigenschaften in einem Datensatz entdeckt, die gar nicht im intendierten Anwendungsbereich der visuellen Clusteranalyse liegen.

Der dritte nicht unbeträchtliche Vorteil liegt darin, dass die Visuelle Datenanalyse auch mit sehr großen sozialwissenschaftlichen Datensätzen umgehen kann. Beispielsweise lassen sich auch mehrere ESS-Wellen heranziehen und mit denselben Haupttypen der Visuellen Datenanalyse untersuchen, zumal das Programm WISDOMIZE spezielle Darstellungen für mehrere Zeitpunkte bereithält.

Ein weiterer und nicht unwichtiger Vorteil liegt darin, dass sich viertens Visuelle Datenanalysen direkt mit dem eingelebten Repertoire an quantitativen Methoden verknüpfen lassen. Beispielsweise können dem visuellen Clustern die statistischen Clusteranalysen als anschlussfähiger Bereich gegenübergestellt werden und tiefergehende Analysen auf der Basis der Ergebnisse des visuellen und des statistischen Clusterns unternommen werden.

Derselbe Punkt kann fünftens – ebenfalls unter der Rubrik der komparativen Vorteile – für die Seite der qualitativen Daten vorgebracht werden, worin sich ein ähnlich anschlussfähiges Potential verbirgt, das allerdings im Kontext der gegenwärtigen Programmarchitekturen noch nicht genutzt werden kann. Aber im Prinzip ist es vorstellbar, einen Kontext von

Qualitative Daten → Visuelle Transformation → Muster ↔ SozialwissenschaftlerIn

aufzuspannen und qualitative Daten in visuelle Muster zu transformieren, die dann forscherseitig interaktiv analysiert werden können.

Gerade die letzten beiden Punkte können noch generalisiert und zur Abbildung 5 verdichtet werden. Quantitative und qualitative Methoden lassen sich über visuelle Methoden in einen geschlossen triadischen Verbund transformieren, deren Elemente sich wechselseitig erzeugen oder generieren. Anders ausgedrückt können erst über visuelle Methoden die Potentiale bisheriger sozialwissenschaftlicher Verfahren voll genutzt werden, da erst – so die Abbildung 5 – der Weg von Triangulierungen ziel- und ergebnisgerichtet beschritten werden kann: Visuelle Analysen können zu quantitativen Untersuchungen führen, die zu weiteren qualitativen Studien anregen. Quantitative Forschung kann in visuellen Explorationen überleiten, die in qualitative Forschung münden. Oder qualitative Studien werden zum Anlass für visuelle Designs genommen und quantitativ weiter umgesetzt.

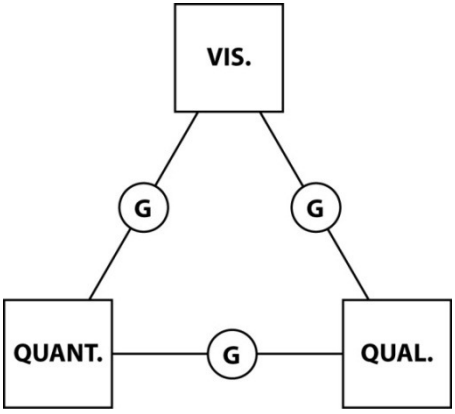

Abbildung 5: Der triadische Verbund von quantitativen, qualitativen und visuellen Methoden und seinen generativen Relationen (G)

6. Ausblicke

Die Visuelle Datenanalyse im Kontext von virtuellen sozialwissenschaftlichen Laboratorien könnte, wird und muss um eine Reihe weiterer Komponenten erweitert werden, die bislang noch gar nicht thematisiert worden sind und die zumindest ausblickhaft angerissen werden sollen.

Die erste Erweiterung könnte als interne bezeichnet werden, betrifft sie doch vornehmlich Aspekte im Umkreis des bisherigen Programms WISDOMIZE 2.0. Diese Ausbauten werden sich in den folgenden Feldern vollziehen:

- eine Vergrößerung der Datenbasis jenseits des bisherigen Fokus auf den Europäischen Sozialen Survey um weitere europäische und vergleichende Datensätze, aber auch nationale Surveys mit geeigneten regionalen Partitionierungen
- der Ausbau der Haupttypen von visuellen Mustern um neue wie beispielsweise multidimensionale *Faces*, visuelle Trends, etc.
- die Erweiterung der visuellen Testtheorie
- die Rekombination von bestehenden visuellen Mustern zu komplexeren Mustern.

Die zweite Erweiterung lässt sich als prozessorientiert apostrophieren und handelt von bewegten Mustern und der Visualisierung von gesellschaftlichen Prozessen, die aber immer streng im Kontext von sozialwissenschaftlichen Methoden und Analysen ablaufen und in denen es um die visuelle Identifizierung von Prozessverläufen geht.

Ein weites Feld für die Visuelle Datenanalyse bieten darüber hinaus die verschiedenen Verfahren der explorativen Datenanalyse, in denen es summarisch ausgedrückt um die schnelle Erkennung basaler Muster in sehr großen Datensätzen geht. Solche Ansätze werden für die Sozialwissenschaften besonders dann relevant, wenn verstärkt prozessgenerierte elektronische Massendaten zur Verfügung stehen, welche die bisherigen Survey- oder Paneldaten zumindest teilweise ablösen. Eine besonders interessante und für die Sozialwissenschaften schon derzeit relevante Methode stellt im explorativen Feld jene der Parallelen Koordinaten dar, mit denen die Variablen von Datensätzen als jeweils vertikale Dimensionen parallel arrangiert werden (Inselberg 2009) und mit denen auch geometrische Algorithmen erzeugt werden können.

Und darüber hinaus können schließlich auch noch regelbasierte Modelle visuell umgesetzt werden, womit sich vielfältige Möglichkeiten für sozialwissenschaftliche Experimente ergeben. Solche Erweiterungen, die in das Feld von genuinen Bildersprachen führen, produzieren visuelle Gestalten von rekursiven dynamischen Prozessen. Nicht selten bietet hier die Visualisierung – und nur die Visualisierung – Lösungskapazitäten bei Modell- und Simulationsproblemen, in denen herkömmliche alpha-numerische Darstellungen notwendigerweise versagen müssen.

Damit wäre eine Erstübersicht zu den Potentialen einer Visuellen Datenanalyse abgeschlossen, welche – so eine der Hauptbotschaften dieses Artikels – ein autonomes Methodenspektrum innerhalb der Sozialwissenschaften begründet. Mehr noch, in den Welten der virtuellen sozialwissenschaftlichen Laboratorien entstehen derzeit Musterbildungen und Bilder-Sprachen, welche die bisherigen Arbeitsformen innerhalb der Sozialwissenschaften nicht nur grundlegend erweitern und ihre Explorations- und Simulationsmöglichkeiten potenzieren werden, sondern welche auch viele der früheren Versuche der Integration von Bildern, Zahlen und Sprachen auf eine Weise umsetzen, in der in vergangenen Jahrzehnten oder Jahrhunderten nicht einmal geträumt werden konnte.

Literatur

Arnheim, Rudolf (1969): *Visual Thinking*. Berkeley: University of California Press.

Card, Stuart K.; Mackinlay, Jock D. & Shneiderman, Ben (1999): *Information Viusualization. Using Vision to Think*. San Francisco: Morgan Kauffman Publishers.

Chen, Chun-houh; Härdle, Wolfgang & Unwin, Antony (2007) (eds.): *Handbook of Data Visualization*. Berlin: Springer.

Foerster, Heinz von (1985): *Sicht und Einsicht. Versuche zu einer operativen Erkenntnistheorie*. Braunschweig: Friedrich Vieweg & Sohn.

Goodman, Nelson (1973): *Sprachen der Kunst. Ein Ansatz zu einer Symboltheorie*. Frankfurt am Main: Suhrkamp.

Goodman, Nelson (1984): *Of Mind and Other Matters*. Cambridge, Mass.: Harvard University Press.

Harkness, Janet A.; Van de Vijver, Fons J.R. & Mohler, Peter P. (2003) (eds.): *Cross-Cultural Survey Methods.* Hoboken: Wiley.

Hofstadter, Douglas R. (1982): *Gödel, Escher, Bach: an Eternal Golden Braid*. Harmondsworth: Penguin Books.

Hofstadter, Douglas R. (1985): *Metamagical Themas: Questing for the Essence of Mind and Pattern*. New York: Basic Books.

Inselberg, Alfred (2009): *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications.* Berlin: Springer.

Knorr-Cetina, Karin (1992): The Couch, the Cathedral, and the Laboratory: On the Relationship between Experiment and Laboratory in Science. In: Pickering, Andrew (ed.): *Science as Practice and Culture*. Chicago: Chicago University Press, pp. 113–138.

Knorr-Cetina, Karin (1995): Laboratory Studies. The Cultural Approach to the Study of Science. In: Jasanoff, Sheila; Markle, Gerald E.; Petersen, James. C. & Pinch, Trevor J. (eds.): *Handbook of Science, Technology and Society*. Los Angeles: Sage Publications, pp. 1–15.

Kosslyn, Stephen M. (1980): *Image and Mind*. Cambridge, Mass.: Harvard University Press.

Müller, Marion G. (2003): *Grundlagen der visuellen Kommunikation*. Konstanz: UVK Verlagsgesellschaft.

Müller, Karl H. (1991): *Symbole - Statistik- Computer - Design. Otto Neuraths Bildpädagogik im Computerzeitalter*. Wien: Hölder-Pichler-Tempsky.

Müller, Karl H. (1991b): Neurath's Theory of Pictorial-Statistical Representation In: Uebel, Thomas (ed.): *Rediscovering the Forgotten Vienna Circle. Austrian Studies on Otto Neurath and the Vienna Circle*. Dordrecht: Kluwer Academic Publishers, pp. 223–251.

Osherson, Daniel N.; Kosslyn, Stephen M. & Hollerbach, John M. (1990)(eds.): *Visual Cognition and Action. An Invitation to Cognitive Science*, Vol. 2. Cambridge: The MIT Press.

Pickover, Clifford A. (1990): *Computers, Pattern, Chaos and Beauty. Graphics from an Unseen World*. New York: St.Martin's Press.

Prusinkiewicz, Przemyslaw & Lindenmayer, Aristid (1990): *The Algorithmic Beauty of Plants*. New York: Springer.

Saris, Willem & Irmtraud Gallhofer (2007): *Design, Evaluation, and Analysis of Questionnaires for Survey Research.* Hoboken: Wiley Interscience.

Thalmann, Daniel (1990) (ed.): *Scientific Visualization and Graphics Simulation*. New York: John Wiley.

Toš, Niko & Müller, Karl H. (2009) (eds.): *Three Roads to Comparative Research: Analytical, Visual, Morphological*. Wien: edition echoraum.

Tufte, Edward R. (1983): *The Visual Display of Quantitative Information*. Cheshire, Conn.: Graphics Press

Tufte, Edward R. (1991): *Envisioning Information*. Cheshire, Conn.: Graphics Press.

Ware, Colin (2004): *Information Visualization. Perception for Design.* 2nd edition. Amsterdam: Elsevier.

Young, Forrest W.; Valero-Mora, Pedro M. & Friendly, Michael (2006): *Visual Statistics. Seeing Data with Dynamic Interactive Graphics*. Hoboken: John Wiley.