

# Big Data, Open Data und Zivile Statistik

MANFRED BOROVCNIK, KLAGENFURT

Mit staatlichen Initiativen, Daten von allgemeiner Relevanz öffentlich zu machen (Open Data) und der breiten Verfügbarkeit von Daten, die en passant entstehen und häufig auch zugänglich sind, haben wir einen Zugang zu Daten wie nie zuvor. Die öffentliche Diskussion greift immer öfter auch auf diese Daten zu und verwendet sie als evidenz-basierte Argumente, um Entscheidungen in die eine oder andere Richtung zu beeinflussen. Statistik für demokratische Entscheidungen oder demokratische Teilhabe zu verwenden, ist ein Anliegen, das dem ehemaligen Bildungsziel Politische Bildung, das auch mit dem Mathematikunterricht verknüpft wurde, entspricht. Es hat sich – europaweit – ein Projekt „ProCivicStat“ damit beschäftigt, Zivile Statistik als eigenständige Disziplin innerhalb der Statistik aufzuwerten und für den Unterricht im Sinne politischer Bildung zugänglich zu machen. In diesem Beitrag geht es darum, die Eigenart von Ziviler Statistik auszuloten und insbesondere auch die Ziele des Projekts ProCivicStat auf ihre Brauchbarkeit hin zu untersuchen. Dabei wird sich erweisen, dass Anwendungen der Statistik im Bereich Ziviler Statistik grundsätzlich verschieden sind von deren Verwendung in Wirtschaft oder Technologie. Die Interessensvertreter, welche – direkt oder virtuell – beteiligt sind, mögen keineswegs dieselben Interessen teilen, sie können sich in ihren Wertvorstellungen grundsätzlich unterscheiden und sie sind in aller Regel von den Konsequenzen der Entscheidungen völlig anders betroffen. Die vorgebrachten Thesen und der grundsätzliche Charakter der Statistik können daher in einem fundamentalen Konflikt mit der Idee von Statistik für die Stärkung ziviler Bürgerbeteiligung (Civic Empowerment) stehen.

## 1. Einleitung

Einige Thesen mögen die Komplexität von Ziviler Statistik unterstreichen: • Der Charakter von Daten unterscheidet sich von Fakten. • Nicht alle Kriterien für Entscheidungen sind verhandelbar. • Entscheidungen sind eng verbunden mit Werturteilen. • Häufig reduzieren falsch gestellte Alternativen die Rationalität von Entscheidungen. Darüber hinaus erfordert ein gesellschaftlicher Gebrauch von Statistik, um Bürger in ihrer Verantwortung zu stärken und um das Niveau der politischen Diskussion zu verbessern, weit mehr Kompetenzen als weithin in der Bevölkerung gegenwärtig vorhanden sind. Es mag auch weise sein, einen Blick in die Geschichte zu werfen: Statistiken wurden ursprünglich erhoben, um jenen, die an der Macht waren, zu informieren, damit sie besser ihre Macht ausüben konnten, und eben nicht für jene, die regiert wurden. Der grundsätzliche Charakter der Statistik kann daher in einem fundamentalen Konflikt mit der Idee von Statistik für Civic Empowerment stehen, deren Versprechen – zivile Verantwortung und Teilhabe an demokratischen Entscheidungsprozessen – damit von vornherein kaum einlösbar sind. Wir gehen auf die Kritikpunkte im Folgenden ein.

## 2. Big Data und Open Data

Big Data spielen in der Statistik eine enorme Rolle als neuartige Quelle von Information, die jedoch nicht den traditionellen Kriterien entspricht. Open Data ist einfach ein Schlagwort, um zu charakterisieren, dass die Daten öffentlich zugänglich sind; da spielen Staaten und insbesondere Statistische Ämter im Sinne von Transparenz mit. Beide Arten von Daten kommen auch ganz wesentlich im Rahmen von Ziviler Statistik vor, die sich damit beschäftigt, wie demokratische Entscheidungen, ja wie überhaupt Meinungsbildung in Demokratien zustande kommen.

### 2.1 Big Data

Bei Big Data handelt es sich um Massendaten, die oft aus automatisiertem Monitoring entstehen, die teils en passant – ohne eigentliches Schema – erhoben werden. Dazu zählen auch Daten aus Social Media. Es kommt dabei zu einer Mischung aller traditionellen Datentypen; ungeheure Mengen von

Daten werden erzeugt und verändert. Nicht nur die fehlende Struktur der Daten macht es schwierig, die Daten entsprechend zu verwerten. Allein die große Masse an Daten erfordert neue und größere Computersysteme und insbesondere auch neue Algorithmen (auch wenn die Methoden dieselben sind) zu deren Analyse. Insbesondere ist auch eine Kalibrierung der Daten schwierig. Dazu zwei Beispiele: Bei gemeinsamen Wanderungen haben wir die App *Health* auf i-Phones genutzt und auf 20 km bis zu 2 km Differenz entdeckt. Daten von Sonden, die einen Schadstoffgehalt messen, sind nicht ganz so einfach auszulesen, wie man das gerne hätte: sowieso ist problematisch, wo die Sonden aufgestellt werden. Es erweist sich als sehr schwierig, die Messwerte zu kalibrieren, das heißt, eine Messkurve zu eichen, die es erlaubt, die Daten der Sonde in reale Daten über den eigentlichen Schadstoffgehalt umzurechnen. Man braucht präzise Zusatzmessungen mit geeichten Messinstrumenten; diese werden selten gemacht. Stattdessen werden oft über Plausibilitätsüberlegungen Eichkurven angenähert.

Ein Hauptproblem von Big Data ist Bias, während das Problem von traditionellen statistischen Daten ihre Variabilität an sich ist (Wild 2017).

## **2.2 Big Data – Erfolge sind oft klein, summieren sich aber**

Für Big Data gilt: • Algorithmen müssen sehr geschickt sein, um Strukturen zu finden. • Viele PCs werden zusammengelegt, um Strukturen zu suchen. • Der Erfolg ist eine – oft gar nicht so große – Verbesserung der Situation. Da aber auch ein kleiner Erfolg viel bringt, wenn man ihn flächendeckend einführt, hat das Thema solche Brisanz. Etwa bei Kundentracking, wenn etwas angeboten wird, was eigentlich nicht wirklich passt, so kann der Kunde das dennoch übernehmen, ganz einfach weil er sich „persönlich angesprochen fühlt“. Für verschiedene Einschätzungen erfordert es nur wenig Aufwand, um zu guten Ergebnissen zu kommen – so etwa bekommt man die politische Einstellung relativ schnell durch ein paar Schlüsselfragen und Daten dazu heraus.

Das Problem bei Big Data ist ein kleiner Nutzen für die Individuen, dem ein unvergleichbar hoher Nutzen bei den „Großen“ gegenüber steht. Noch sind wenige „Übergriffe“ aus der Störung der Privacy bekannt, jedoch, wer weiß?

## **2.3 Big Data als Basis für Diagnosen**

Es bestehen große Hoffnungen, mit Hilfe von Big Data Bilddiagnose in der Medizin zu revolutionieren. Mit Hilfe von an riesigen Bildmengen trainierten Algorithmen hat man die automatische Diagnostizierung von Röntgenbildern auch tatsächlich schon sehr weit vorangebracht. Ein geschulter Röntgenologe braucht länger zur Diagnose etwa eines CTs und hat dabei höhere Übersehensraten. Allerdings haben die Algorithmen eine viel höhere Falsch-positiv-Rate. Das führt zu vielen unnötigen weiteren Untersuchungen und auch zu einer nicht unbeträchtlichen Zahl von Behandlungen; jedenfalls, wenn man die Systeme autonom einsetzen würde.

Die Frage für die Zukunft, wenn Algorithmen flächendeckend die Diagnose unterstützen werden, wird lauten: Was passiert im Konfliktfall zwischen Algorithmus und Arzt? Wird der Arzt sich trauen, gegen den Algorithmus zu entscheiden? Und: Wie wird ein Arzt zur Verantwortung gezogen, wenn der Algorithmus falsch gelegen ist und er dem gefolgt ist?

## **2.4 Open Data**

Open Data sind nicht nur Daten, die man verwenden kann, ohne Urheberrechte zu verletzen; sie sind einfach öffentlich zugängliche Daten. Und wir haben heute – nicht nur von statistischen Ämtern – viel mehr Zugriff auf Daten als früher. Das Problem bleibt, dass man viele Daten wiederum nicht bekommt, weil sie nicht vorhanden sind. Damit entsteht eine Verzerrung der Welt. Wenn man die Wirksamkeit von Screening bei verschiedensten Arten von Krebs untersuchen möchte, so hat man in

der Regel keine Daten über unnötige Eingriffe und deren Folgen im Nachklang von falsch-positiven Diagnosen. Man hat natürlich auch keine Daten über die tatsächliche Lebensverlängerung, wenn eine Krebsdiagnose durch Screening gestellt wurde. Es gibt ja in-situ-Formen, die sich sehr spät oder nur sehr langsam entwickelt hätten – da ist frühzeitige Erkennung sogar ein Nachteil (Gigerenzer 2013).

### 3. Eigenheiten der Statistik und ihrer Anwendungen

Wir wollen zuerst einige Eigenheiten von Statistik im Lichte ihrer Anwendungen betrachten. Dies soll im späteren Verlauf als Rahmen für die Bewertung von Ideen und Methoden einer Zivilen Statistik dienen. Wir gehen auf den Modellcharakter ein, die Voraussetzungen, die an Daten gestellt werden müssen, damit sie eine Verbindlichkeit im Sinne von Relevanz und Repräsentativität haben; wir behandeln auch das Problem von falschen Alternativen und von Entscheidungskriterien, welche Entscheidungen wesentlich beeinflussen und auch verzerren können; abschließend äußern wir uns zu Datenexploration und Datenreduktion, beides Ansätze, die unterschiedlicher nicht sein könnten, um aus Daten lesbare und verwertbare Information zu erhalten.

#### 3.1 Modell oder Szenarien

Mathematische Modellbildung im Sinne des naturwissenschaftlichen Paradigmas lässt es als vernünftig erscheinen, das angestrebte Modell, das die Wirklichkeit und das Problem darin abbilden soll, möglichst gut an die Realität anzupassen. Die Kreislaufmodelle beginnend mit Schupp (1982) oder Blum und Kirsch (1989) verbessern wiederkehrend die Passung von Modell und Wirklichkeit. Schon Borovcnik (1986) hat sich dahingehend geäußert, dass es oft von Vorteil ist, eine reale Situation aus dem Blickwinkel mehrerer Modelle zu betrachten und in der Differenz der Modelle und der darin erbrachten „Lösungen“ das gestellte Ausgangsproblem besser zu verstehen. Es ist daher oft unangebracht die Überlegungen und Entscheidungen auf ein einziges Modell zu basieren. Dies trifft für gesellschaftspolitische Entscheidungen umso mehr zu als es für technologische oder wirtschaftliche Anwendungen der Fall ist, weil – anders als in den letztgenannten Gebieten – gemeinsame Ziele kaum vorauszusetzen sind. Borovcnik (2009) spricht in diesem Zusammenhang auch von Szenarien, mit denen man das Problem besser durchleuchten kann, um zu einer „Lösung“ zu gelangen.

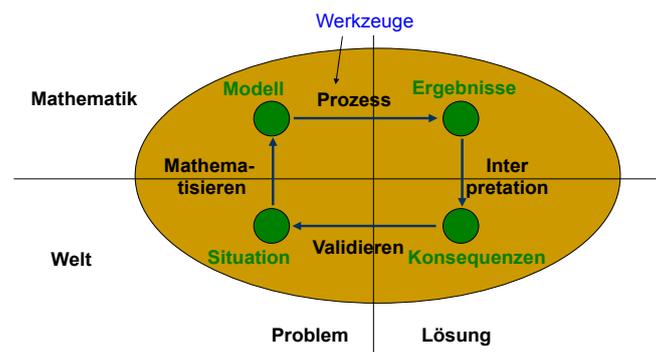


Abb. 1: Kreislaufschema der mathematischen Modellierung nach Schupp (1985) oder Blum und Kirsch (1989)

#### 3.2 Der Charakter von Daten

Borovcnik (2014) wirft die Frage auf, wie artifiziell Daten eigentlich sein müssen, damit sie geeignet sind, anstehende Fragen zu beantworten. Daten müssen in ihrer Erhebung sehr vielen Annahmen und Anforderungen gerecht werden, damit sie brauchbar sind. Daten als vorgegebene Entitäten, Daten als Fakten, die man nicht mehr hinterfragen kann, das sind Sichtweisen, die ignorieren, dass man Daten nicht einfach ohne konzeptionellen Hintergrund (wie das bei Big Data oft erfolgt) erheben kann. Daten passieren nicht einfach. Einige knapp gehaltene Beispiele mögen das illustrieren.

Daten sind nicht immer direkt messbar; oft entsprechen sie Konstrukten oder Indikatoren. Intelligenz wird z. B. als Konstrukt davon abhängig, welches Testinstrument man verwendet. Arbeitslosigkeit wird davon abhängig, wie man sie definiert – und es gibt unzählige Definitionen, die am Wesentlichen vorbeigehen. Das Risiko von Krankheiten, etwa von Krebs oder Leukämie wird allein schon deswegen nicht direkt einem Schadensereignis wie dem Super-Gau in Fukushima zuschreibbar, weil das alles seltene Erkrankungen sind, weil Messungen der Belastungen nicht einmal generell flächendeckend gemacht worden sind, weil die historische Belastung einzelner Personen nicht aufgezeichnet wurde, und weil zwischen Belastung und Erkrankung eine längere Zeitspanne liegt. Da reden dann schon die einen das Risiko klein während andere das Risiko überbewerten (Martignon 2016).

HCB ist ein Gift, das praktisch nicht löslich ist; es kann in kleinsten Dosen irreparable Schäden im menschlichen Organismus hervorrufen. Bei der Verbrennung von Abfällen wurde in einem Tal in Kärnten 2014 eine hohe HCB-Dosis an die Umwelt abgegeben. Offizielle Messungen ergeben jedoch nun, vier Jahre nach dem Vorfall, Entwarnung (Umweltbundesamt 2018). Wenn man weiß, dass HCB praktisch nicht in den Boden eindringt, so ist die Dicke der Bodenproben entscheidend, welche Belastung man misst. Je dicker die Schichte, umso kleiner die gemessenen Belastungen. Die Bevölkerung arbeitet mit lokalen Statistikern zusammen, um einen Belastungskatalog auch über Urinproben von Kindern zu erstellen, um das individuelle Risiko zu erfassen. Was ist nun Fakt an den Daten?

Viele Daten, die „informed decisions“ (informierte Entscheidungen) zugrunde liegen, sind verzerrt oder fehlen ganz. Wenn man über Länder hinweg vergleicht, so findet man unterschiedliche Definitionen oder Messverfahren, sodass eine Vergleichbarkeit lediglich vorgetäuscht wird aber kaum vorhanden ist. Man könnte auch sagen, dass die Arbeit in angewandter Statistik zu 80% aus der Datenproduktion besteht inklusive dem Design der Studien und nur zu 20% aus der Analyse der Daten.

### **3.3 Falsche Alternativen**

Oft wird der Entscheidungsspielraum vor der Analyse eingeengt. Was unter der Prämisse, das Problem traktabel zu machen, durchgezogen wird, hat oft manipulativen Charakter. Das heißt, die Alternativen werden so eingeengt und zugespitzt, dass nur mehr das als sinnvoll erscheint, was man „durchsetzen“ will. So etwa kamen in einer angestrebten Reform des Spitalsdienstes an Wochenenden in England die folgenden Behauptungen auf (Ridgway, Arnold, Moy & Ridgway 2016, S. 2):

„Wir haben an Wochenenden [...] ungefähr 200 vermeidbare Todesfälle [...] verschiedene Behauptungen: Einlieferungen an Wochenenden haben eine höhere Sterblichkeit, [...] die Todesfälle sind vermeidbar, [...] es gibt eine Ursache: zu wenig Personal an Wochenenden.“

Die eingeschränkte Alternative war, gleich viele oder mehr Ärzte am Wochenende in Krankenhäusern. Allerdings zeigte eine tiefere Analyse, dass einerseits die Art der Erfassung der Sterblichkeit das Ergebnis verzerrt und dass es ferner gar nicht an der Präsenz von Ärzten sondern am Labor und an der Verfügbarkeit von CT-Untersuchungen liegt, wenn man Verbesserungen erreichen möchte.

Eine andere falsche Alternative jüngerer Zeit scheint die Frage nach dem friedlichen Nutzen von Atomenergie oder Ausstieg aus der Atomenergie zu sein. Zu viele offene, unbeantwortete Fragen sind ausgeklammert worden. Die Frage ist doch eher, unter welchen Bedingungen könnte atomare Energie genützt werden. Während Deutschland sich abmüht, die Infrastruktur für alternative Energien aufzubauen (etwa ein riesiges Stromnetzwerk, das die flüchtige Energie von den Windrädern in die Berge bringt und durch Hochpumpen von Wasser in entsprechenden Behältern speichert), war eine entsprechende Frage in Frankreich nicht zu stellen, so abhängig ist das Land von Atomenergie. Was Sicherheitsfragen anbelangt, so ist bei Zwischenfällen die Distanz von deutschen Regionen zu französischen Kernkraftwerken zu klein, um die Frage innerhalb eines Landes zu lösen. Allerdings, so sorglos wie man im Betrieb von Kernkraftwerken in Japan umgegangen ist, kann es wohl nicht gehen. Auch eine falsche Alternative: sorgloser, gefährlicher Umgang gegen vollkommener Ausstieg.

Öffentliche Entscheidungen durch Nudges zu beeinflussen kann einer Gruppe durchaus vernünftige Entscheidungen nahebringen, die sie ansonsten nicht getroffen hätte (Thaler & Sunstein 2009). Allerdings kann Nudging auch rasch als manipulative Bevormundung interpretiert werden, wodurch der „Nudger“ längerfristig seine Glaubwürdigkeit verliert. Hier läuft Statistik auch Gefahr, Nudges durch „rationale“ Argumente zu unterfüttern und damit gesellschaftliche Entscheidungen zu manipulieren. Die einseitige Verengung des Spielraums der Entscheidungen mag das Problem leichter erfassen lassen, sie verzerrt allerdings sowohl das Problem als auch die Entscheidung.

### **3.4 Entscheidungskriterien**

Wie sehr sich Entscheidungskriterien auf die letztlich getroffene Entscheidung auswirken, kann man in Borovcnik (2015) nachlesen. Verschiedene Stakeholder, die in einer Entscheidungssituation aufeinander treffen, haben danach auch zu Recht verschiedene Maßstäbe; das ist insbesondere im Rahmen gesellschaftspolitischer Entscheidungen (wozu auch der Gesundheitssektor gehört) schwer abzugleichen. Es gibt nicht die für alle beste Entscheidung. Ohnehin ist es von der Sache her schwer, vernünftige Kriterien zu finden. Das zeigen die folgenden Beispiele.

Ridgway et al. (2016) berichten von der Bewertung einer allfälligen Reduktion von Ärzten in der Ambulanz an Sonntagen. Soll man zur Bewertung der Konsequenzen 1-Tages-, 2-Tages- oder 30-Tages-Überlebensraten heranziehen? Diese Kriterien sind nicht beliebig, sondern sollten mit dem Risiko, das aus geringerer ärztlicher Unterstützung in Spitälern resultiert, verbunden sein. Gigerenzer (2013) berichtet über die Schwierigkeiten in der Interpretation von 5-Jahres-Überlebensraten bei Krebs in Screening-Programmen. Wenn Krebs früher entdeckt wird (auch nicht-aggressive Arten), so wird bei Behandlung die 5-Jahres-Überlebensrate größer sein als bei Personen, die nicht am Screening teilnehmen und daher erst später im Krankheitsverlauf die Erstdiagnose erhalten.

Interessant ist auch, dass Kriterien und Wahrscheinlichkeiten von der Rolle der Beteiligten abhängen (Radfahrer-Autofahrer-Phänomen). Auch haben die Stakeholder andere Konsequenzen zu tragen: Ärzte haben eine Haftpflicht, Patienten haben die direkten Folgen zu tragen. Interessant ist auch, dass manche Leute denken, die Kriterien für Entscheidungen seien „politisch verhandelbar“.

### **3.5 Datenexploration versus komplexere Datenreduktion**

Datenkomprimierung beruht auf sehr komplexen statistischen Verfahren (Regression, Faktorenanalyse etc.), welche zur sachlichen Interpretation der vorhandenen Daten nötig sind, aber üblicherweise einfach übergangen werden. Datenexploration dagegen bedient sich sehr einfacher Techniken, die ohne jegliche Voraussetzungen zu prüfen, eingesetzt werden. Wir können mit elementaren Techniken – Prozentsätzen und einfachen Diagrammen – sehr weit kommen. Man muss jedoch festhalten, dass schon Prozentsätze schlecht verstanden werden, Funktionen und Skalen in Graphen werden noch häufiger missverstanden. Außerdem werden die verwendeten Visualisierungen immer komplexer und überdies noch dynamisiert (Engel 2019).

So etwa wird im Gapminder die Entwicklung der Menschheit durch eine Animation „erklärt“: Eine Steigerung des Wohlstands (gemessen durch das Brutto-Inlandsprodukt pro Kopf) bringt über 300 Jahre hinweg eine drastische Steigerung der Lebenserwartung (Rosling 2009). Der Film ist so beeindruckend, dass man danach gar nicht wagt, die verwendeten Begriffe zu hinterfragen. Was kann ein Brutto-Inlandsprodukt über mehrere hundert Jahre hinweg eigentlich bedeuten? Was kann eine Lebenserwartung eigentlich bedeuten, wenn diese für China innerhalb einer Dekade um mehrere Jahre auf und ab pendelt? Wie wird die Entwicklung denn verständlich, wenn man das Inlandsprodukt auf einer logarithmischen Skala darstellt, auf der die Hälfte des dargestellten Bereiches (in Originalwerten) auf einem Achtel der Bildbreite erscheint: die Distanz von 400 zu 40000 erscheint wie 400 zu 4000; die gezeigte ansteigende Gerade demnach eine langgezogene stark abflachende Kurve ist.

Auch verwischt die Animation, dass der starke Anstieg der Lebenserwartung um 1900 eigentlich zufolge der verbesserten Hygiene, der Ausmerzungen der Säuglings- und Kindersterblichkeit und der besseren Ernährung eintrat (siehe auch Blastland & Spiegelhalter 2013). Der jüngere, kleinere Anstieg nach 1990 ist der verbesserten Altersmedizin zu verdanken, wobei der Zuwachs an Lebenserwartung hier oft auch zulasten eingeschränkter Lebensqualität erfolgt. Das bedeutet, dass eine naive „je älter die Menschen werden, umso besser“-Deutung nicht aufrecht zu erhalten ist.

Es war schon die Rede davon, dass die Datenexploration einer Datenkomprimierung gegenübersteht. Verfahren der Datenkomprimierung sind noch schwieriger zu interpretieren, wären aber oft zur sachlichen Interpretation der vorhandenen Daten zielgerichteter. In Anbetracht der Schwierigkeiten mit dem Verständnis von Prozenten muss man sich die didaktische Frage stellen, wie man Meta-Wissen zu diesen Verfahren vermitteln kann, damit eine verständige Würdigung und Wertschätzung der Methoden und der damit erzielten Ergebnisse ermöglicht wird. Und eine Kritikfähigkeit bei den „Konsumenten“ gesellschaftlicher Entscheidungen erhalten bleibt.

## **4. Zivile Statistik**

Hierbei geht es um einen jüngeren Ansatz, Statistik als Basis der demokratischen Meinungsbildung zu gebrauchen. Also im Sinne politischer Bildung. Dazu hat es ein EU-weites und EU-finanziertes Projekt Pro-Civic-Stat gegeben, das sich grundsätzlich mit der Legitimierung von Ziviler Statistik beschäftigt hat, das auch entsprechende Kompetenzen beschrieben hat, die für die Anwendung von Statistik in gesellschaftspolitischen Fragen erforderlich sind. Im Rahmen des Projekts wurden auch umfangreiche Materialien entwickelt.

### **4.1 Der Ansatz von Pro-Civic-Stat zum Aufbau einer Zivilen Statistik**

Engel (2019, S. 2) beschreibt die Grundidee von Ziviler Statistik so

“Demokratie lebt von Argumenten, die auf nachweisbaren Fakten basieren. Fehlinformationen, Fake News, Informationsmangel und Ignoranz sind Bedrohungen für unsere Lebensweise. In einer zunehmend komplexeren Welt ist das Engagement von Zivilbürgern eine kritische Ressource bei öffentlichen Entscheidungen [...] Das Verständnis solcher Themen ist für das bürgerschaftliche Engagement in modernen Gesellschaften von großer Bedeutung, basiert aber oft auf komplexen multivariaten Daten, [...] Kompetenzen im Bereich Zivilstatistik sind zur informierten Partizipation in demokratischen Gesellschaften nötig.“

#### **Pro-Civic Stat und Politische Bildung**

Das hatten wir in den 1980ern unter Projektorientiertem Unterricht und Politischer Bildung ange-mahnt. Neu ist die Möglichkeit zur tiefergehenden eigenen Datenanalyse anhand weiterer Daten mit viel mächtigeren statistischen Werkzeugen als dazumal. Mit der Explorativen Datenanalyse sind damals auch Ziele einer politischen Bildung und Emanzipation verknüpft worden. Damals hatten wir noch das Ziel, einfache mathematische Darstellungen zu verwenden, damit die Begrifflichkeiten und Visualisierungen in der Kommunikation (mit dem Problem in der Lösungsphase, zwischen Experten und Öffentlichkeit) direkt verstehbar sind und die sich aus dem Modell ergebende Sichtweise auch mitteilbar, sprich verständlich wird (Borovcnik & Ossimitz 1987). Damals wollten wir auch noch die Mathematik im Hintergrund vermitteln (vielleicht nicht jedem, aber doch einigen). Das ist im Ansatz von ProCivicStat anders: man will versuchen, mittels dynamischer Visualisierung von multivariaten Datensätzen wesentliche Erkenntnisse über dahinterstehende gesellschaftliche Phänomene aufzubauen. Auch benutzt werden komplexe Techniken zur Datenkomprimierung wie Faktorenanalyse und Regression Trees, ohne dass man diese auch nur ansatzweise erklären oder verstehen könnte.

## **Pro-Civic Stat und Visualisierung**

Heute werden immer komplexere Darstellungen verwendet; auch visuell, ja, dynamisch müssen sie auch noch sein. Natürlich ist dies dem Druck der Anwendungen geschuldet! Da werden Darstellungen auch verwendet, um das Gegenüber auszutricksen. Aus den Schriften zum Projekt geht indirekt hervor (Engel 2019), dass man unterstellt, dass allein aus der Visualisierung – ohne Modellbildung sowie ohne Begriffe – das Phänomen verstanden werden kann. Da treffen gleich mehrere Fäden aufeinander:

- Die Daten sind nicht Fakten, werden aber angesichts der visuellen Darstellung als real aufgefasst.
- Die dynamische Veränderung verleiht der Visualisierung Authentizität.
- Der Einzelne ist so damit befasst, die – oft dynamische – Visualisierung zu verstehen, sodass für Kritik an der Vorgangsweise kaum Platz bleibt.

Neben Dispositionen kommen auch Emotionen ins Spiel; da wird man sich gegen die Mehrheit, die das schon „verstanden“ hat, kaum kritisch äußern können. Schließlich bewegen sich die bunten Kugeln so schön.

## **4.2 Charakterisierung von Ziviler Statistik von Gal**

Iddo Gal ist einer der führenden Protagonisten von ProCivicStat. Gal (2018) charakterisiert den Gegenstand und die Methoden von Ziviler Statistik als einen Zugang zu Daten und Entscheidungen, die von gesellschaftspolitischer Relevanz sind. Die verwendeten Daten sind multivariat und oft auch aggregiert und entstammen vielfältigsten Quellen, auch für ein und dasselbe Problem muss man Daten aus verschiedenster Provenienz zusammenführen. Dazu kommt noch, dass die Daten oft mit Texten unterschiedlichster Qualität eingekleidet sind. Der Text enthält Folgerungen aus Analysen, die oft auch visuell orientiert sind. Meist geht es auch um die Untersuchung oder Bewertung von Zusammenhängen, die sich auch noch über die Zeit hinweg verändern. Im Einzelnen kann man Gegenstand und Methoden von PCS wie folgt beschreiben. Dabei gehen wir gleich auf die damit verbundenen Schwierigkeiten ein.

### **Zivile Statistik ist multivariat**

Die simple Analyse von Tabellen und Graphen reicht selten. Es braucht geeignete Techniken, um den Aufgaben gerecht zu werden; dazu zählen auch Korrelation, Regression, Cluster- und Faktorenanalyse. Immer wird zu einer realistischeren Modellbildung auf komplexere Methoden zurückgegriffen. Es ist möglich, aber schwierig, brauchbares Meta-Wissen darüber zu vermitteln. Speziell bei Korrelation treten die Fragen nach Richtung und Kausalität auf – und die gehören zu den heikelsten – sind aber *nicht* in der Statistik zu erörtern sondern Fragen, die innerhalb der Substanzwissenschaft aus dem Kontext der Daten zu behandeln sind.

### **Zivile Statistik verwendet ausgiebig aggregierte Daten**

Da bei Ziviler Statistik oft aggregierte Daten im Spiel sind, muss man Artefakte über aggregierte Daten kennen. Gruppieren kann viel höhere (niedrigere) Korrelation zwischen Merkmalen vortäuschen. Indikatoren bilden das interessierende Konstrukt-Merkmal selten brauchbar ab (etwa die Arbeitslosenrate).

### **Zivile Statistik untersucht dynamische Zusammenhänge**

Das trifft auf Veränderungen in der Zeit. Um den Charakter der Daten direkt zu zeigen, werden daher oft dynamische – und das gleich multivariate – Visualisierungen gezeigt. Diese Visualisierungen bringen neben der Anschaulichkeit Bedeutungsveränderungen mit sich: Die Visualisierung macht die Daten pseudo-konkret trotz ihres artifiziellen und komplexen Charakters. Die dynamische Veränderung schließt einen kognitiven Ansatz aus, weil sie ganz stark die emotionalen Sinne anspricht. Es läuft alles wie in einem Film ab, wo sich bunte Kugeln bewegen (Rosling 2009).

### **Zivile Statistik verwendet multiple Quellen**

Die meisten dieser Daten sind entweder Indikatoren, aggregierte Daten oder wahllose Big Data aus dem Internet. Das erhebt oft Zweifel über ihre Definition (was erfassen sie denn eigentlich?) sowie ihre Relevanz und ihre Repräsentativität.

### **Zivile Statistik verwendet unterschiedlichste Texte**

Da werden oft extrem lange Texte verwendet. Meist werden in diesen Texten implizit Wertevorstellungen der Autoren miteingeschleust. Die Sprache ist schon immer als Kriterium für Komplexität in Zusammenhang mit der Statistik identifiziert worden. Die Überlappung mit Wertevorstellungen, die implizit mittransportiert werden, erhöht die Komplexität enorm.

### **Zivile Statistik verwendet unterschiedlichste neue Visualisierungen**

Die Visualisierungen werden besonders im Zusammenhang mit der Darstellung multivariater Daten sehr komplex. Aber sogar mit einfachen Graphen sind die Defizite im richtigen Lesen (oder gar Erstellen) bekannt. Wie können wir da erwarten, dass die um vieles komplexeren Graphen richtig verstanden werden. Was noch schwerer wiegt, die dynamischen Visualisierungen sprechen nicht emotional statt rational an.

## **5. Kategorien, die das Konstrukt Zivile Statistik beschreiben**

Innerhalb des Projekts ProCivicSat gibt es ein starkes Interesse, Zivile Statistik als eigenständige Disziplin zu rechtfertigen. Daher auch der Versuch, Zivile Statistik als Konstrukt von eigenen Kompetenzen zu beschreiben, die es nahelegen, Unterricht und Lernsequenzen zu entwickeln, welche gerade diese Kompetenzen herausbilden lassen. Es gibt mehrere Quellen, in denen diese Kompetenzen beschrieben werden, die innerhalb einer Zivilen Statistik wesentlich sind, und die auch begründen, warum diese als erstrebenswerte Ziele für einen Unterricht in Ziviler Statistik gelten können: Ridgway, Nicholson & Gal (2017), Gal, Ridgway & Nicholson (2018), Engel (2019). Diese Arbeiten erschließen ein Konstrukt Zivile Statistik durch insgesamt 11 Komponenten, welche wiederum in Teile zerfallen. Die Komponenten und unsere Ansichten dazu werden erläutert. Wir folgen dabei den Ausführungen in Borovcnik (2018). Die Kategorien werden wie folgt in drei Bereiche zusammengefasst:

#### *Engagement und Aktion*

- # 1 Bereitschaft zum gesellschaftlichen Engagement (Meaning for society and policy)
- # 2 Kritische Bewertung und Reflexion
- # 3 Dispositionen

#### *Wissen*

- # 4 Statistik und Risiko
- # 5 Modelle und Darstellungen (Models, patterns, and representations)
- # 6 Methodik und Forschungsprozess

#### *Ermöglichungsprozesse*

- # 7 Erweiterungen im Bereich Amtlicher Statistik
- # 8 Kontextbezogenes gesellschaftliches Wissen
- # 9 Informations- und Kommunikationstechnologie und Informationsrecherche (ICT)
- #10 Quantitativer Kern
- #11 Lese- und Textverständnis und Kommunikation (Literacy)

## 5.1 Engagement und Aktion

### #1 Bereitschaft zum gesellschaftlichen Engagement (Meaning for society)

Statistik diene ursprünglich den Herrschenden als Instrument. Wenn falsche Alternativen die Modellierung verengen, wird das Ergebnis beeinflusst. Die Subjektivität in der Modellierung macht bei Anwendungen auf gesellschaftlicher Ebene Schwierigkeiten. Experten sind nicht neutral sondern haben Interessen. Die Vereinfachung der Situation, damit breitere Kreise das verstehen, führt zu Verzerrungen und der Charakter von Lösungen als theoretischer Input in die Diskussion kann insbesondere durch Visualisierung in Fakten gedreht werden. Entscheidungen verlangen, die Evidenz verschiedener Optionen, das Risiko (Wahrscheinlichkeit), subjektiven Nutzen und Expertenbewertungen zu evaluieren.

Statistik kann Bürger informieren, damit sie politische Entscheidungen besser mitgestalten können. Jedoch, eine Situation zu evaluieren, ist viel schwieriger als zwei Optionen miteinander zu vergleichen. Oft wird aber zudem die Situation auf eine falsche Alternative (zwei Optionen) reduziert, anstatt dies selbst als Modellierungsschritt darzustellen. Wahrscheinlichkeit ist schwer zu verstehen und interpretieren. Bei Nutzen ist das wohl kaum besser. Beides sind Begriffe, die ganz stark subjektiven Vorstellungen und Bewertungen unterliegen. Wie soll man diese Subjektivität auf einem gesellschaftlichen Niveau miteinander versöhnen? Sind Experten wirklich frei von Eigeninteresse? Komplexere Situationen müssen vereinfacht werden, damit sie die Leute weithin verstehen können. Die Situation kann dadurch stark verzerrt werden und ihren theoretischen Charakter verlieren, speziell durch die – dynamische – Visualisierung bzw. Materialisierung.

### #2 Kritische Bewertung und Reflexion

Experten und deren Ergebnisse kritisch in Frage zu stellen, sozusagen als Grundhaltung, reicht nicht. Es braucht spezifisches Wissen aus Kontext und Methodologie, um richtige Fragen zu stellen. Die Bewertung umschließt auch Werte; Statistik kann daher Entscheidungen nicht „automatisieren“.

„Folgen sie einer Agenda?“ „Sind die Daten glaubwürdig?“ Kritisches Denken mag auf der Metaebene durchaus durch solche Fragen angeregt werden. Doch, um Experten herauszufordern, um mehr und validere Information zu erhalten, braucht es mehr als solch rituelles Fragen. Die Experten wären bald auf solche Standardfragen vorbereitet. Es bedarf Wissen, aus dem Kontext, der Disziplin und über statistische Begriffe und die Methodologie (Repräsentativität, Stichproben, Konzeptualisierung der Variablen, potentieller Einfluss weiterer Faktoren, Kausalität vs. Korrelation). Viele dieser entscheidenden Faktoren sind nicht durch die Daten abgedeckt, sodass sie gar nicht untersucht werden können. Statistik kann Werte nicht ersetzen; mit anderen Wertevorstellungen führen dieselben Statistiken zu anderen Schlüssen und Entscheidungen. Schon in einfachsten Fällen ist kritische Bewertung eine komplexe Aufgabe. Die folgenden Beispiele beziehen sich auf Gal (2018).

„Noch nie dagewesen: arktischer Winter der wärmste – Rekord.“ Man muss diese Aussage einbetten – nicht das sehen, was andere wollen. Die Alpen, z.B., erleben jüngst eine Wärmeperiode, aber war da es im Mittelalter weit wärmer. Zeugnis davon geben die Dörfer und die Almwirtschaft, die sich zwischen 1300 und 1700 m Seehöhe abspielte.

„Wir sehen eine hohe Wahrscheinlichkeit für Erdbeben von 70%.“ Dies war eine Vorhersage. Es ist schwierig zu verstehen, dass die Vorhersage korrekt gewesen sein kann gemessen an bestimmten (!) Modellen und dass es dennoch im entsprechenden Zeitraum ein Beben gegeben hat.

„Deutschland und Österreich sind zwei der reichsten Länder weltweit.“ Was bedeutet es für eine Person, in einem reichen Land zu leben. Diese Person mag sogar von der Infrastruktur wenig haben, abgesehen von der Tatsache, dass die Person nach üblichen Maßstäben arm sein kann.

### #3 Dispositionen

Dispositionen beeinflussen das Verständnis. Allein die Frage, warum man an einer berichteten Zahl interessiert sein sollte, kann Hintergründe aufdecken. Allerdings werden Dispositionen (u.a. des Wohlfühlens) auch genutzt, um Botschaften zu transportieren: die tanzenden bunten Kreise zeigen doch die „Entwicklung“ (Rosling 2009). Wozu braucht man da noch die verwendeten Konzepte hinterfragen?

„The Israelian railways announce to have an 84.21% accuracy of trains.“ Gal (2018) knüpft an dieses Beispiel folgende kritische Fragen: Warum wird die Zahl mitgeteilt? Brauchen die Eisenbahnen mehr Geld? Weiters: Was bedeutet denn Zuggenauigkeit? Wann ist ein Zug verspätet? Oder, was bedeutet der Prozentsatz? Bezieht er sich auf den Anteil aller Züge, die pünktlich waren, egal in welcher Region, oder bezieht er sich auf verschiedene Tageszeiten (Stoßzeiten inkludiert)? In diesem Sinne scheinen Dispositionen ein Schlüssel für herausfordernde Fragen zu sein.

Die Gapminder-Animation (Rosling 2009) über die Beziehung zwischen Lebenserwartung und Einkommen zeigt den negativen Einfluss von Dispositionen. Sie erzeugt ein synthetisches Verständnis des Phänomens. Sie macht die Konzepte Einkommen und Lebenserwartung lebendig und entkleidet sie ihres theoretischen Charakters. Allerdings wird eine logarithmische Skala (!) für das Einkommen benutzt. Wo liegt dann die Hälfte des Einkommensbereichs – wohl doch nicht in der Mitte der Graphik? Was kann Einkommen über 500 Jahre quer über Kulturen und Wirtschaftssysteme denn eigentlich bedeuten? Was kann Lebenserwartung bedeuten, wenn sie innerhalb von 5 Jahren mehrfach um ca. 10 Jahre schwankt?

## 5.2 Wissen

### #4 Statistik und Risiko

Werden Risiken analysiert oder kommuniziert, so spielt die Rolle der Beteiligten einen entscheidenden Faktor. Wahrscheinlichkeit bedeutet für sie genuin etwas anderes (nicht nur, dass sie subjektiv andere Werte hätten). Es gibt auch einen Unterschied, ob man eine Entscheidung einmal oder sehr oft trifft (Borovcnik 2015). Wir haben es häufig mit sehr kleinen Risiken zu tun, für die man keinerlei Erfahrung haben kann. Risiken sind oft nur als Hazard vermutbar, d.h., man kann die negativen Folgen nicht direkt auf eine Gefährdung zurückführen. Das macht individuelle Entscheidungen und Verantwortung wichtig, was aber gerade in gesellschaftlichen Kontexten schwierig ist.

Gal (2018) nennt unter Statistik und Risiko: Stichprobenziehen mag die Glaubwürdigkeit beeinträchtigen; er setzt mit Begriffspaaren fort: Beschreiben–Vergleichen, Korrelation–Kausalität, Beschreibung–Prognose. Aus gesellschaftlicher und individueller Perspektive ergeben sich folgende Caveats:

- Die Rolle, die jemand in einer Entscheidung spielt, beeinflusst nicht nur dessen Sicht sondern auch dessen Entscheidung.
- Wahrscheinlichkeit bedeutet nicht dasselbe für die verschiedenen Stakeholder – das geht über die üblichen Bedeutungsunterschiede hinaus. Für den Einzelnen machen relative Häufigkeiten des Systems keinen Sinn!
- Es gibt einen großen Unterschied in der Rationalität zwischen Einzelentscheidungen und wiederholten Entscheidungen.

Solche Eigenschaften spielen eine entscheidende Rolle, wie Risiken aufgegriffen und interpretiert werden (etwa beim Screening nach Krankheiten).

Einen Sonderfall stellen kleine Wahrscheinlichkeiten dar: Wir haben keinerlei Daten darüber und werden niemals ausreichende Daten dazu haben. Wir müssen uns daher auf modell-basierte Zahlen (keine Schätzungen!) für das Risiko verlassen. Historisch gab es viele Vorschläge, kleine Wahrscheinlichkeiten einfach zu ignorieren. *Kleine Wahrscheinlichkeiten stehen aber heute im Blickwinkel von Risikobetrachtungen.*

Korrelation und Kausalität: Es lohnt sich, einen Blick in die historische Entwicklung von Regression und Korrelation zu machen und zu erkennen, wie diese Methoden dazu entwickelt wurden, um einen empirischen Beweis kausaler Beziehungen zu liefern (MacKenzie 1981). Und der „Beweis“ wurde dadurch vervollständigt, dass der Reversionskoeffizient (das was wir in diesem Spezialfall heute als Korrelation identifizieren) für ähnliche Probleme immer rund um die 0.50 lag. *Korrelation ist nur ein Werkzeug, um die Spreu vom Weizen zu trennen, um jene Probleme zu isolieren, wo es sich lohnen mag, über allfällige kausale Zusammenhänge durch interne Betrachtungen nachzuforschen.*

#### **#5 Modelle und Darstellungen (Models, patterns, and representations)**

Alle Modelle sind falsch, einige aber sind nützlich (Box nach Box und Draper 1987). Das Zitat ist aber *keine* Einladung, die Prüfung der Annahmen eines Modells zu vernachlässigen! Es ist besonders wichtig, dass sich „nützlich“ auf den Zweck bezieht, und der Zweck wiederum basiert auf Wertevorstellungen.

Solange die Modellierung innerhalb einer Firma etwa erfolgt, gibt es ein gemeinsames Ziel. Wenn in der Gesellschaft verschiedenste Stakeholder über den Zweck eines Modells Übereinkunft finden sollen, dann kann das eine harte Auseinandersetzung zur Folge haben. Dazu kommt noch, wie man modellieren kann, wenn man die Konstituenten des Modells nicht versteht. Dasselbe gilt für die Ergebnisse. Allerdings gilt auch, wie wir schon angemerkt, dass Elementarisierung den Charakter sowohl der verwendeten Modelle als auch der Lösung völlig verändern kann. Die verschiedenen Rollen, die Beteiligung, das Engagement, Wissen und die Folgen, die man von Entscheidungen basierend auf einem „falschen“ Modell „erleidet“, all das ist so unterschiedlich, dass eine einfache Lösung für gesellschaftliche Probleme durch Statistik wohl außer Reichweite steht.

Der PPDAC-Zyklus (Problem – Plan – Data – Analysis – Conclusions) ist von eingeschränkter Hilfe. Üblicherweise wird er Wild & Pfannkuch (1999) zugeschrieben, aber der Modellbildungskreislauf ist eine alte Idee aus der deutschen Modellierungsdiskussion der 1980er Jahre (Schupp 1982, Borovcnik 1986, Blum & Kirsch 1989; verfeinert in Leiß & Blum 2006). Der Ansatz kann durch mehrere Durchläufe verfeinert werden, aber das stellt die Konzeptualisierung der Basisbegriffe keineswegs in Frage. Der Ansatz lässt die nötigen Annahmen, um zu dem Modell zu gelangen, einfach weg (lokale Verfeinerung vs. globale Diskussion der Passung eines Modells). Wie soll man denn modellieren, wenn der Modelleur die Bestandteile und Voraussetzungen des Modells gar nicht versteht?

Die Notwendigkeit und Nützlichkeit von guten Darstellungen ist unbestritten. Jedoch muss man sich ausreichend damit beschäftigen, wie Darstellungen den Charakter der Situation und der Ergebnisse verändern können. Muster werden oft nur durch spezielle Darstellungen entdeckt und mögen auch einfach ein Artefakt darstellen. Die verwendete Darstellung (insbesondere dynamisch) kann die Glaubwürdigkeit eines Musters ungebührlich erhöhen.

#### **#6 Methodik und Forschungsprozess (Methods and inquiry process)**

Dazu zählen nach Gal (2018) Forschungsmethoden, Stichprobenmethoden, Qualitative Forschung und Ethik. Die üblichen Methoden werden für PCS für die spezifischen Anforderungen an die Datenproduktion adaptiert. Daten in PCS verletzen jedoch oft diese Voraussetzungen oder leiden an einem Definitionsproblem für die Merkmale. Weder sind die üblichen Methoden für Daten über die ganze Population geeignet (kein Signifikanztest etwa), noch sind sie geeignet für nicht-zufällige Daten, die üblicherweise verwendet werden; schon gar nicht passen sie für aggregierte Daten.

## 5.3 Ermöglichungsprozesse

### #7 Amtliche Statistik

Indexzahlen über lange Zeiträume zu vergleichen, ist oft wenig sinnvoll (siehe auch das Gapminder-Beispiel, Rosling 2009). Korrelation wird durch Aggregation von Daten verändert und täuscht so stärkere oder schwächere Zusammenhänge vor. Arbeitslosigkeit mag der falsche Indikator für die Beschäftigungslage einer Wirtschaft sein; über geleistete Arbeitsstunden gibt es aber wenig zuverlässige Statistiken. Aggregierte Daten auf nationalem Niveau zeigen ein stabileres Muster von Zusammenhängen als auf regionalem Niveau (Aggregation eliminiert das Rauschen aus den Daten). Indikatoren sind auf Individuen nicht übertragbar!

Beispiel: Indikatoren leisten nicht das, wofür sie entwickelt wurden. Arbeitslosigkeit: Über die letzten 10 Jahre haben sich die bezahlten Stunden nicht geändert. Aber da die Anzahl der Angestellten zugenommen hat, haben die Stunden pro Person abgenommen. Gar nicht zu sprechen vom Einkommen für Personen, die als beschäftigt gezählt werden. Daher ist Arbeitslosigkeit der falsche Indikator für Beschäftigung in einem Wirtschaftsraum. Die Zahl der geleisteten Arbeitsstunden in Beschäftigungsverhältnissen ist da schon besser, noch interessanter ist, diese Zahl aufzugliedern nach dem Verdienst, der damit verbunden ist. Das entspricht viel besser der wirtschaftlichen Aktivität.

Beispiel: Unfälle werden durch Alkohol, durch Motorräder „verursacht“. Keine anderen Ursachen werden gelistet. Es kann daher gar keine anderen Ursachen geben. Wenn eine Alkoholisierung registriert wird, *ist* sie die Unfallursache. Auch dann, wenn es klar ist, dass der Unfall nicht vermeidbar gewesen wäre, wenn der Lenker nüchtern gewesen wäre.

Wie Statistiken aufgezeichnet werden, ist vorgegeben und kann kaum erweitert oder verändert werden. Fehlende Transparenz der Aufzeichnung bildet häufig Kritik in gesellschaftlich weniger konsensualen Bereichen. Fragen, die über die verfügbaren Daten hinausgehen, entziehen sich der Analyse. Gibt es keine Daten zu einem Phänomen, „gibt“ es kein Phänomen

### #8 Kontextbezogenes gesellschaftliches Wissen

Um etwa den Gender Wage Gap zu analysieren, reicht es nicht, die simplen Daten aus Einkommensstatistiken zu untersuchen. Soll man nicht ohnehin die Unterschiede in der Einkommensverteilung analysieren anstatt nur die Differenz der Mittelwerte zu betrachten? Faktoren, die üblicherweise untersucht wurden: Beruf, Alter, Teilzeit, Geschlecht. Faktoren, die nicht untersucht wurden (teilweise fehlen die Daten dazu): Risiko (Unfälle), Flexibilität, Arbeitsbedingungen, Überstunden „on demand“ etc. sowie persönliche Merkmale, die das Einkommen beeinflussen: Intelligenz, Anpassbarkeit, etc.

Die festgestellten Unterschiede könnten zufolge Diskriminierung oder zufolge persönlicher Entscheidungen sein. Am Ende der Analysen stellen sich Fragen wie „Was empfiehlt man jungen Frauen? Was empfiehlt man der Gesellschaft?“ Solche Empfehlungen implizieren immer Wertevorstellungen.

### #9 Informations- und Kommunikationstechnologie und Informationsrecherche (ICT)

Manches Mal bekommt man den Eindruck, es geht den Verfechtern von PCS einfach nur um „push the right button“ in der bereitgestellten Software. Wer jedoch passt auf, prüft oder würde erkennen, wenn der falsche Knopf gedrückt worden ist. Die Ergebnisse sind leicht abzurufen, sie erhalten den Charakter von Fakten. Und: Die Glaubwürdigkeit der Ergebnisse steigt bei den Nutzern sogar, weil sie mit Hilfe von Technologie erhalten wurden.

Jedoch gilt immer noch: diese Ergebnisse hängen von Modellen und Annahmen ab und entsprechen einer eingeschränkten Sicht der Welt und des Problems. Und die Ergebnisse bedürfen einer sachgerechten Interpretation!

## #10 Quantitativer Kern

Die Frage nach dem wichtigsten statistischen Begriff beantworten viele mit Prozent. Allerdings reduzieren Prozente den Sachverhalt auf einen Wert. Viel wichtiger ist die Verteilung. Was kann die Angabe von 84.21% Zugpünktlichkeit bedeuten? Welche Zugtypen, welche Tageszeiten, welche Region, welche Variabilität? Genauso wenig ist die Angabe einer 0.8% Inzidenz von Brustkrebs für Frauen zu verstehen. Variabilität ist der Kern der Statistik – Mittelwerte und Prozente ebnen das aus, treffen aber offenbar auf ein archetypisches Verlangen nach Sicherheit (siehe Batanero & Borovcnik 2016).

Beispiel: 0.8% Inzidenz von Brustkrebs für Frauen. Für die institutionelle Perspektive mag das passen, um damit die Rationalität eines Screening-Programmes zu untermauern. Und es dann mit ziemlicher „Aggressivität“ zu bewerben. Für die einzelnen Frauen im System mag es viel wichtiger sein, zu wissen, welcher Untergruppe sie angehören. *Variabilität* ist im Kern von Statistik – Mittelwerte nivellieren diese Variabilität, obwohl sie das angesprochene archetypische Bedürfnis nach Sicherheit abdecken.

## #11 Statistical Literacy

Wir stimmen vielen Zielen einer Zivilen Statistik zu, wie sie im Projekt ProCivicStat formuliert werden. Allerdings möchten wir sie nicht als spezifische Gründe pro Zivile Statistik nutzen. Wir hätten gerne, dass ein üblicher Statistikkunterricht solche Ziele realisiert und einige haben dort auch ihren Platz:

- Reale und authentische Probleme.
- Modellieren der Situation und der Fragestellung.
- Über die Datenproduktion nachdenken, damit sie valide Daten liefert, die geeignet sind, die anstehenden Fragen teilweise zu beantworten.
- Die Technologie passend einsetzen, auch im Sinne des Durchrechnens mehrerer Modelle und des Vergleichens der Ergebnisse.
- Interpretieren der Ergebnisse im Verhältnis zu den Annahmen und prüfen, ob diese valide sind.
- Diskutieren weiterer Probleme, die bei der Erforschung der Ausgangsfrage aufgeworfen wurden.

## 6. Ziele von Ziviler Statistik und Charakter von Statistik sind nicht vereinbar

### 6.1 Die Fähigkeiten und Kompetenzen

Die Ausbildung bestimmter Kompetenzen, die von der Zivilen Statistik als wichtig eingestuft und forciert werden, sind auch Ziele für die übliche Lehre der Statistik. Und sie können im üblichen Statistikkunterricht erworben werden. Obwohl wir dem zustimmen, dass diese Kompetenzen in vielen Fällen nicht wirklich gut vermittelt werden. Civic Stat macht, wie jetzt vorgeschlagen wird, umfangreichen Gebrauch von dynamischen, multivariaten Visualisierungen. Als ob der Betrachter direkt Zeuge einer Entwicklung werden könnte, als ob er ein Video sehen würde, das die Realität zeigt. Es sind aber die Auswirkungen eines Modells, die hier gezeigt werden, und nicht die Realität. Die Annahmen des Modells werden in der Visualisierung versteckt, die den Charakter der Widerspiegelung der Realität erhält (ein Video erzählt mehr als 1000 Bilder, ein Bild erzählt mehr als 1000 Worte). Wir bezweifeln, dass die Betrachter die verwendeten Visualisierungen verstehen können. Wir vergleichen das mit dem weitverbreiteten Mangel an grundlegenden Fähigkeiten wie dem Verstehen von Prozentsätzen oder einfachen Diagrammen.

*Wir bezweifeln auch, dass die Lehrer die erforderlichen Kompetenzen haben werden, um Zivile Statistik zu unterrichten.*

## 6.2 Wahrheit: Statistik ist kein neutraler Ansatz für die Welt

Wir haben historisch eine Entwicklung darin gesehen, wer die Wahrheit festlegt und mitteilt: In älteren Zeiten sagte der Priester, was die Wahrheit ist, später der Kaiser (mit oder ohne Kirche), dann der Richter im Rechtssystem, schließlich der Arzt. Und nun sollte der Statistiker diese Rolle übernehmen. Das ist weder die Aufgabe der Statistik, noch wollen wir die Rolle des Statistikers zur Wahrheitsfindung missbrauchen lassen. Statistiker können versuchen, die Menschen dazu zu bringen, sich geeignete Daten zu beschaffen (die meisten Daten sind nicht geeignet für das, wozu sie verwendet werden) und sie richtig zu analysieren. Aber dann muss der Statistiker den „Kunden“ mit der Frage allein lassen, was er tun soll und wie er die Ergebnisse im Kontext interpretieren soll.

*Die Interpretation ist an Kontext und Werte gebunden.* Der Statistiker macht diese Werte nicht überflüssig. Wir sehen Ansprüche für Civic Stat in dem Sinne, dass die Daten sagen können, wie die Welt ist. *Die Statistiker werden die übertriebenen Versprechungen nicht einlösen können und – als Gegenreaktion – wird in der Zukunft niemand mehr der Statistik trauen.* Ein Fall, den wir am meisten bedauern würden.

Die Interpretation statistischer Ergebnisse erfordert *Werte*. Die Statistik ist *beschränkt* durch die (Qualität der) Daten, die Modelle und die inhärenten Annahmen. Alle diese Einheiten bedeuten Einschränkungen bei der Analyse und für die Interpretation. Die Experten, die darüber entscheiden, wie die Daten gewonnen werden, welche Modelle verwendet werden, welche Aspekte der Ergebnisse berichtet werden, haben ein eigenes Interesse. Man wird immer einen Experten finden können, der seine Ansicht teilt und sie durch auf Daten basierende Schlussfolgerungen unterstützt. Das führt zu einer Abwertung von Experten, was der Gesellschaft insgesamt gar nicht zugutekommt und daher nicht im Interesse eines demokratisch verfassten Staates liegen kann.

*Es ist wichtig zu beachten, dass Experten ihre eigenen Interessen verfolgen* (entweder in der Sache, oder im Sinne ihrer Karriere, oder um damit Geld zu verdienen).

## 6.3 Handlungen und Entscheidungen sind abhängig von Werten

Es ist nicht die Aufgabe des Statistikers, Werte zu ersetzen oder Werte vorzugeben.

Beispiel: Das geschlechtsspezifische Lohngefälle (Gender Wage Gap, GWG; FullFact 2017), wie es aus Daten berechnet wurde, kann auf Diskriminierung oder auf persönliche Entscheidungen zurückzuführen sein. Der Statistiker kann keine der beiden Positionen teilen und weder die eine noch die andere unterstützen. Dies ist keine Frage der Statistik. Außerdem, wenn wir über den GWG Bescheid wissen, welchen Rat sollten wir einer jungen Frau geben? Die Werte sind die Entscheidung der Nutzer von Ziviler Statistik. Wir vermissen diesen Punkt in Zivilstatistik. Außerdem sehen wir, dass das Material (siehe ProCivicStat, o. D.) den Eindruck erweckt, dass man aus den Daten alles herauslesen kann.

*Die Interpretation der statistischen Ergebnisse hängt eben nicht nur vom Kontext, sondern auch von den Werten ab.*

*Für Zivile Statistik:* In der Gesellschaft treffen unterschiedliche Stakeholder aufeinander. Die Werte werden untereinander nicht geteilt; in einer Demokratie werden die Werte meist verhandelt und nicht vorgegeben. Das bedeutet auch, dass verschiedene Stakeholder die gleichen Ergebnisse unterschiedlich interpretieren können. Dies trifft auch für die Screening-Systeme im Gesundheitssystem zu: der Patient, der Arzt, das medizinische System, das Gesundheitsministerium usw., alle Beteiligten haben unterschiedliche Interpretationen über die Risiken und den Nutzen von Vorsorgeuntersuchungen für Krankheiten. Und sie interpretieren die gleichen Ergebnisse unterschiedlich. Die Beteiligten haben unterschiedliche Interessen und Rollen und sind von den Entscheidungen ganz unterschiedlich betroffen (persönliches Leid, Haftung, Behandlung lege artis etc., abgesehen davon, dass es auch um Geld geht).

*In der Angewandten Statistik ist dies anders:* Der Statistiker arbeitet für einen Kunden, der eine optimierte oder bessere Entscheidung auf der Basis empirischer Evidenz haben möchte. Die Analyse wird eindeutig für *einen* Stakeholder durchgeführt. Es besteht daher kein Interessenkonflikt.

## 7. Schluss: Statistik – wozu sie dienen kann

Das Projekt ProCiviStat will mit der „Zivilstatistik“ eine Facette statistischer Bildung beschreiben, die weder darauf angelegt ist, den Stochastikunterricht zu ersetzen, noch klar auf den schulischen Bereich beschränkt ist. Man will auch die Statistikausbildung in mathematik-ferneren Studiengängen miteinbeziehen. Es gibt sogar Proponenten des Projekts, welche die Zivilstatistik im Mathematikunterricht als schlecht untergebracht ansehen und daher die Statistikausbildung in Fächer wie Geographie oder direkt Politische Bildung verlagern wollen. Der durch Zivilstatistik beschriebene Anspruch zieht – im Sinne der Proponenten des Projekts – nicht nur eine Bringschuld auf Seiten der Rezipienten, sondern auch auf Seite der Produzenten von Zivilstatistik nach sich. Mit den vorliegenden Erläuterungen und der vorgebrachten Kritik soll einerseits gezeigt werden, wie wenig angemessen die Vorstellungen des Projekts mit Blick auf eine Realisierung entsprechender Elemente im schulischen Stochastikunterricht sind, andererseits soll auch gezeigt werden, dass der Anspruch des Projekts überzogen ist, dass Statistik in gesellschaftspolitischen Fragen – insbesondere im Dissens – eine neutrale Rolle spielen kann, um endgültige Antworten (Wahrheiten?) zu finden.

Statistik kann helfen, gute Fragen zu stellen; Antworten allerdings sind an einen Kontext und an Wertvorstellungen gebunden. Je stärker die Statistik aus spezifischen Anwendungen (Wirtschaft, Technik etc.) heraus in gesellschaftspolitische Bereiche eindringt, umso weniger trifft der enge Kontext zu, umso diffuser werden Wertvorstellungen. Wenn über Werte kein Konsens besteht – und der besteht gerade in einer Demokratie nicht sondern wird in Absprachen ausgehandelt – werden Antworten von der Statistik nicht zu erwarten sein. Es gibt zu hohe Erwartungen, die auch durch Experten geweckt werden, etwa um ihre eigene Rolle aufzuwerten.

- Statistik kann insbesondere den demokratischen Prozess des Werteabgleichs keineswegs ersetzen.
- Statistik ist nicht in der Lage, die Stelle des Vertreters der Wahrheit einzunehmen, wie dies historisch etwa in folgender Reihenfolge geschehen ist: Religion – Medizin – Jurisprudenz – Medien – Statistik.

Statistik soll nicht dazu instrumentalisiert werden, die gesellschaftliche Wahrheit zu verwalten. Die vielen Fact-Checks weltweit (etwa FullFact 2017) zeigen allerdings in eine andere Richtung. Das bedeutet nicht, dass man Zivile Statistik im Unterricht aussparen soll. Viele Ziele einer Zivilen Statistik sind auch für den allgemeinen Statistikunterricht relevant. Man muss jedoch wissen, wie schwierig die Einschätzung vorgeblich objektiver Ergebnisse ist und wie sehr die Interpretation dieser von Wertvorstellungen abhängt. Abschließend soll zum Ausdruck gebracht werden, dass man Statistik so vermitteln soll, dass man lernt, wie man Modelle nützlich macht im Sinne von G.E.P. Box (Box & Draper 1987, S. 424): „Essentially, all models are wrong, but some of them are useful.“

Der Autor dankt Herrn Andreas Vohns für die kritische Durchsicht des Beitrags, welche die Lesbarkeit der Ideen erheblich verbessert hat.

## Literatur

- Batanero, C. & Borovcnik (2016): *Statistics and probability in high school*. Rotterdam: Sense Publishers.
- Blastland, M. & Spiegelhalter, D. J. (2013): *The norm chronicles*. London: Profile Books.
- Blum, W. & Kirsch, A. (1989). The problem of the graphic artist. In: W. Blum, J. S. Berry, R. Biehler, I. D. Huntley, G. Kaiser-Messmer & L. Profke (Hrsg.), *Applications and modelling in learning and teaching mathematics*. Chichester: Horwood, S. 129-135.
- Borovcnik, M. (1986): Zum Teilungsproblem. *Journal für Mathematik-Didaktik*, 7(1), 45–70.

- Borovcnik, M. (2009): Aufgaben in der Stochastik – Chancen jenseits von Motivation. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG)* 42, 20–42.
- Borovcnik, M. (2014): Vom Nutzen artifizieller Daten. In: U. Sproesser, S. Wessolowski & C. Wörn (Hrsg.): *Daten, Zufall und der Rest der Welt*. Berlin: Springer, S. 27–43.
- Borovcnik, M. (2015): Risk and decision making: The „logic“ of probability. *The Mathematics Enthusiast* 12(1,2&3), 113-139.
- Borovcnik, M. (2018): Die Rolle der Statistik in ziviler Verantwortung. In: *Beiträge zum Mathematikunterricht 2018*. WTM-Verlag, S. 333–336.
- Borovcnik, M. & Ossimitz, G. (1987): *Materialien zur beschreibenden Statistik und explorativen Datenanalyse*. Wien: Hölder-Pichler-Tempsky.
- Box, G. E. P. & Draper, N. R. (1987): *Empirical model-building and response surfaces*. New York: Wiley.
- Engel, J. (2019): Statistical Literacy und Gesellschaft: Was ist Zivilstatistik? *Stochastik in der Schule* 39(1), 2–12.
- FullFact (2017, 16.11.): *Do women earn less than men in the UK?* UK’s Independent Factchecking Charity. Online: [fullfact.org/economy/UK\\_gender\\_pay\\_gap/](http://fullfact.org/economy/UK_gender_pay_gap/) (Zugriff: 20.12.2019).
- Gal, I., Ridgway, J. & Nicholson, J. (2018, März): A Framework for understanding data about society. *Vortrag Minisymposium 29: Statistical Literacy and Civic Engagement: Teaching and Learning with Data about Society*. GDMV Paderborn.
- Gigerenzer, G. (2013): *Risiko. Wie man die richtigen Entscheidungen trifft*. München: Random House.
- Leiß, D. & Blum, W. (2006): Beschreibung zentraler mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller (Hrsg.): *Bildungsstandards Mathematik: konkret: Sekundarstufe I: Aufgabenbeispiele, Unterrichts Anregungen, Fortbildungsideen*. Berlin: Cornelsen Verlag, S. 33–50.
- MacKenzie, D. A. (1981): *Statistics in Britain – 1865-1930 – The social construction of scientific knowledge*. Edinburgh: Edinburgh University Press.
- Martignon, L. (2016): Empowering citizens against typical misuse of data concerning risks. In: J. Engel (Hrsg.): *Promoting understanding of statistics about society*. IASE Roundtable Berlin.
- ProCivicStat (o.D.): Is there racial bias in European football? *ProCivicStat Materials*. Online: [community.dur.ac.uk/procivic.stat/wp-content/uploads/2015/12/Racial\\_Bias\\_in\\_Football\\_.pdf](http://community.dur.ac.uk/procivic.stat/wp-content/uploads/2015/12/Racial_Bias_in_Football_.pdf) (Zugriff: 20.12.2019).
- Ridgway, J. Arnold, P., Moy, W. & Ridgway, R. (2016): Deriving heuristics from political speeches for understanding statistics about society. In: J. Engel (Hrsg.): *Promoting understanding of statistics about society*. IASE Roundtable Berlin.
- Ridgway, J., Nicholson, J. & Gal, I. (2017): Conceptual Framework for Civic Statistics. Übersetzung „Aufgabenanalyse: Facetten von Statistical Literacy“. Online: [community.dur.ac.uk/procivic.stat/index.php/resources/conceptual-framework/](http://community.dur.ac.uk/procivic.stat/index.php/resources/conceptual-framework/) (Zugriff: 20.12.2019).
- Rosling, H. (2009): 200 years that changed the world. *Gapminder*. Online: [www.gapminder.org/videos/200-years-that-changed-the-world/](http://www.gapminder.org/videos/200-years-that-changed-the-world/) (Zugriff: 20.12.2019).
- Schupp, H. (1982): Zum Verhältnis statistischer und wahrscheinlichkeitstheoretischer Komponenten im Stochastik-Unterricht der Sekundarstufe I. *Journal für Mathematik-Didaktik*, 3(3-4), 207–226.
- Thaler, R. H. & Sunstein, C. R. (2009). *Nudge. Wie man kluge Entscheidungen anstößt*. Berlin: Econ.
- Umweltbundesamt (2018).Görtschitztal. Untersuchungen an Mensch und Umwelt. Report 0652.Wien: Umweltbundesamt. [www.umweltbundesamt.at/news\\_180328](http://www.umweltbundesamt.at/news_180328) (Zugriff: 20.12.2019).
- Wild, C. (2017): Statistical literacy as the earth moves. *Statistics Education Research Journal* 16(1), 31–37.
- Wild, C. J. & Pfannkuch, M. (1999): Statistical thinking in empirical enquiry (with discussion). *International Statistical Review* 67(3), 223–265.

## Verfasser

Manfred Borovcnik

Alpen-Adria-Universität Klagenfurt, Institut für Statistik, Universitätsstraße 65, 9020 Klagenfurt

[manfred.borovcnik@uni-klu.ac.at](mailto:manfred.borovcnik@uni-klu.ac.at)