

Nr. 4 Data
Delusion

JvM

GREEN PAPERS

THOUGHTS ON A WORLD IN THE MIDST OF RADICAL CHANGE



NR. 4 DATA DELUSION.

APRIL 2020

WIE DIE DATEN HINTER COVID-19 MISSVERSTANDEN WERDEN UND WARUM DATENKOMPETENZ DIE MAKE-IT-OR-BREAK-IT-FÄHIGKEIT UNSERER ZEIT IST.

Nie zuvor hat unsere Welt eine Krise erlebt, die ein derartiges Ausmaß hat und begleitet wird von einer nie da gewesenen Menge an Informationen, inklusive minütlicher Updates zur Zahl der COVID-19-Infizierten und daran Verstorbenen. Und nie zuvor waren so viele unzuverlässige Informationen im Umlauf. Wir haben es aktuell nicht nur mit einer Pandemie zu tun, sondern auch mit einer „Infodemie“.

Der Ausbruch des Coronavirus brachte nicht nur einen Anstieg des Nachrichtenkonsums¹ mit sich, sondern führte auch zu einer allgemeinen Besessenheit von Echtzeitdaten. Innerhalb kürzester Zeit haben wir uns angewöhnt, ständig Dashboards, Graphen und Infografiken zu checken, die Aufschluss über den aktuellen Stand der COVID-19-Fälle in der eigenen Heimatstadt und Region, im eigenen Land und in der Welt geben.² Viele Zeitungen und Institutionen gaben der Datenvisualisierung einen enormen Schub durch ansprechende, verständliche und interaktive Karten, die die erschreckende Ausbreitung sichtbar machten.

Die Verdopplung der Infektionen in New York alle drei Tage, mit einem Anstieg von 967 auf 238.831 Fälle in knapp 32 Tagen³, ließ uns im April einen kalten Schauer über den Rücken laufen. Denn wir glauben, dass wir die Zahlen, die wir sehen, verstehen zu können. Am Ende des Tages haben wir alle ein Urvertrauen in die Objektivität von Zahlen, ihre Vergleichbarkeit und Stichhaltigkeit.⁴ Doch dieses Vertrauen täuscht, denn die Zahlen stellen nur einen Teil der Wahrheit dar.

Das Problem mit den Zahlen ist die Tatsache, dass uns nie beigebracht wurde, sie objektiv zu bewerten, statt ihnen per se zu vertrauen. Doch gerade die objektive Bewertung von Daten wirft zwei grundlegende Probleme auf, die wir täglich ignorieren: Messmethode und Kontext.

MESSMETHODEN VERSTEHEN.

Um zu begreifen, was eine bestimmte Statistik tatsächlich bedeutet, ist es wichtig, zu verstehen, wie sie erhoben wurde. Sehen wir uns einmal die Zahl der COVID-19-Infizierten an. Während Medien und Regierungsinstitutionen über steigende oder fallende Zahlen von Infizierten in verschiedenen Ländern berichten, werfen sie mit absoluten Werten um sich, lassen dabei jedoch oft den wichtigsten Wert aus den Augen: die Anzahl der getesteten Personen.

Den Anteil der Infizierten zu interpretieren, ohne dabei die Testrate zu kennen, ist extrem irreführend: Tatsächlich sagt diese Zahl nur etwas über die Anzahl der erfassten Infektionen aus und sehr wenig über die Verbreitung des Virus im Land.⁵

Im Fall von Deutschland führte das merkwürdige Missverhältnis zwischen der Zahl der Infizierten und derjenigen der Verstorbenen zu großer Verwunderung. Die Erklärung beruht auf der Tatsache, dass Deutschland durch die enorme Zahl an Tests auch deutlich mehr Infektionen erfasst. Am 12. April 2020 hat das Robert-Koch-Institut bestätigt, dass die Testrate in Deutschland bei 20,94 Tests pro tausend Einwohner liegt, verglichen mit einem Wert von 5,54 in Großbritannien.⁶ Virologen gehen davon aus, dass diese Zahl einer Entdeckungsrate von 15,6 % in Deutschland und einer weltweiten Entdeckungsrate von 6 % entspricht.⁷ Das würde bedeuten, dass die tatsächliche Zahl der Fälle in Deutschland zu diesem Zeitpunkt bei ungefähr 930.000 lag, während es weltweit schon deutlich über 40 Millionen Fälle gegeben haben könnte.⁸

Ein weiteres besorgniserregendes Beispiel ist die Zahl Corona-bedingter Todesfälle. International gibt es keine Klarheit darüber, was einen Corona-bedingten Todesfall ausmacht. Die beiden Ansätze zur Erfassung sind: „Tod COVID-19-Erkrankter“ und „Tod durch COVID-19“. Während weltweit unterschiedliche Definitionen verwendet werden, gibt es in manchen Fällen sogar unterschiedliche Messmethoden innerhalb eines Landes. In Deutschland definiert das Robert-Koch-Institut einen Corona-bedingten Todesfall als „Tod COVID-19-Erkrankter“, während Hamburg nur diejenigen Todesfälle als Corona-bedingt registriert, bei denen nach einer Autopsie der „Tod durch COVID-19“ bestätigt wird.⁹

Diese unterschiedlichen Methoden führen zu einer deutlichen Abweichung der Messwerte für Deutschland von denen für Hamburg. In die Zukunft projiziert, sagen die Raten voraus, dass die Fallzahlen um etwa mehr als eine halbe Million auseinandergehen werden (Deutschlands Sterberate von 3,19 % und Hamburgs Sterberate von 2,09 %).¹⁰ Ein Unterschied, der maßgeblich sein kann, wenn über die Verlängerung der Beschränkungen entschieden wird.

DATEN IN DEN RICHTIGEN KONTEXT SETZEN.

Den richtigen Bezugsrahmen zu finden ist der Schlüssel dazu, Daten richtig zu interpretieren. Blicken wir einmal zurück in den Februar 2020, als COVID-19 sich gerade seinen Weg nach Europa bahnte. Damals unterstellten kritische Stimmen den Wissenschaftlern, nur Panik verbreiten zu wollen. Sie bezogen sich auf die Zahl der jährlichen Todesfälle durch Lungenkrankheiten – die bei 650.000¹¹ liegt – und behaupteten, es würde sich kaum lohnen, ernsthaft über COVID-19 zu diskutieren.

Dabei verstand keiner der Kritikerinnen und Kritiker die Beziehung zwischen der absoluten Zahl und der Infektions- und Sterberate, die die eigentlich entscheidenden Faktoren sind, wenn es darum geht, die Auswirkungen einer Krankheit auf die Gesellschaft zu beurteilen. Die Ignoranz des Bezugsrahmens ließ viele die Auswirkungen von COVID-19 unterschätzen.¹²

Jetzt, im April 2020, kennen wir immer noch nicht die genaue Sterberate von COVID-19, doch Studien zeigen, dass der Wert irgendwo zwischen 0,5 % und 13 % liegt und sich vermutlich näher an 1 % befindet.¹³ Dieses „1 %“ lässt uns aufatmen und denken: „Ach, 1 % ist gar nicht so viel – ich werde sicher nicht dazugehören!“ Doch setzen wir diese 1 % in den globalen Kontext, können wir die wahre Dimension der Pandemie erkennen.

Betrachtet man den größeren Zusammenhang, kann man leicht berechnen, wie hoch die Zahl der globalen Todesfälle zum Ende dieser Pandemie sein wird: Bei einer Sterberate von 1 % und einer Infektionsrate von 60 % wird die Summe der Todesfälle bei rund 45 Millionen liegen. Eine Zahl, die der Hälfte der Opfer im Zweiten Weltkrieg entspricht.¹⁴ Eine Zahl, angesichts derer wir uns schnell so gar nicht mehr unbesiegtbar fühlen. Eine Zahl, die uns allen unbestreitbar klarmachen sollte: „#StayTheF***AtHome!“

Die oben genannten Beispiele erschließen eine sehr einfache Wahrheit: Daten können ohne Kontext irreführend sein. Wenn wir die Messmethoden falsch verstehen, erhalten wir falsche Informationen über den aktuellen Status der Krise. Betrachten wir Daten nicht im richtigen Kontext, stufen wir wichtige Dinge als unwichtig ein. Diese Tatsache gilt für jeden Datenpunkt, dem wir begegnen: von Daten über Leben und Tod bis hin zu konventionellen Konsumdaten.

Betrachten wir einmal den Konsum von Toilettenpapier. In Zeitungen, Social-Media-Feeds sowie im gesamten öffentlichen Gespräch empört man sich über Menschen, die Toilettenpapier hams-tern, und verurteilt ihr selbstsüchtiges, unverantwortliches Handeln. Während es definitiv einen Teil der Gesellschaft gibt, der sich dieser Anklage schuldig macht, sollten wir versuchen, die Verkaufszahlen in den richtigen Kontext zu stellen, um zu verstehen, was wirklich passiert ist.

Die Toilettenpapierindustrie ist in zwei völlig voneinander unabhängige Märkte geteilt – B2B und B2C. Toilettenpapier, das für den gewerblichen Markt produziert wird, unterscheidet sich fundamental vom Toilettenpapier im Supermarkt: Es wird in viel größeren Rollen produziert, das Papier selbst ist dünner, grober und wird nicht in kleinen Gebinden verkauft. B2C- und B2B-Toilettenpapier machen jeweils rund 50 % des Gesamtmarktes aus.

Die Selbstisolation führte nun zu einer Umverteilung des Konsums: Ein großer Teil der Menschen, die zu Hause bleiben, nutzt eben keine Toiletten mehr am Arbeitsplatz, in Restaurants, Bars, Clubs oder an Flughäfen. Das steigerte die Nachfrage nach B2C-Toilettenpapier um 40 %.¹⁵ Während der Bedarf im Privatgebrauch steigt, ist das Angebot für gewerblichen Gebrauch aufgrund der genannten Unterschiede nicht in der Lage, diesen Privatverbrauch zu decken, was zu einem Defizit führt. Wenn wir uns also darüber auslassen, dass unsere Nachbarn alle Vorräte im Supermarkt für sich beanspruchen, haben wir dabei womöglich den Gesamtzusammenhang aus den Augen verloren.¹⁶

Den Kontext zu verstehen war nie so schwer wie heute: Die aktuelle Situation ist in jeder Hinsicht außergewöhnlich. Sie bringt Trends hervor, die von solch kurzer Dauer sind, dass schon der Versuch schwerfällt, sie objektiv zu bewerten.

Es gibt jedoch drei Fragen, die wir uns alle stellen können, um der persönlichen Datenkompetenz einen Schritt näher zu kommen:

1. WIE WURDEN DIE DATEN GESAMMELT? Erst wenn wir uns damit auseinandersetzen, wie Daten erhoben wurden und wie einzelne Werte definiert sind, können wir verstehen, was diese Daten wirklich aussagen. Dabei dürfen wir uns nicht scheuen, kritisch zu sein – wenn die Messmethoden nicht logisch wirken, ist es durchaus möglich, dass sie es tatsächlich nicht sind.

2. IN WELCHEM KONTEXT WIRD DAS THEMA BETRACHTET? Um den Kontext richtig zu verstehen, hilft es, sich zu fragen, welche anderen Aspekte für die Fragestellung wichtig sein könnten. Nicht zuletzt sind Auffälligkeiten in den Daten oft auch auf andere Erklärungen zurückzuführen (wie bei unserem B2B-/B2C-Beispiel). Auch ein Blick auf die Größe der Datenbasis und was genau zum Vergleich herangezogen wird, kann helfen.
3. WIE VALIDE IST DIE INFORMATIONSQUELLE? Eine kurze Recherche zeigt schnell, ob auch andere Medien die Daten verwenden und der ursprüngliche Kontext womöglich ein anderer ist. Nicht zu vergessen ist, dass neben absichtlich falschen Darstellungen auch simple Copy-Paste-Fehler das Problem sein können.

Wie außergewöhnlich die aktuelle Situation auch sein mag, eins steht fest: Wir leben in einer Welt, in der Daten nicht nur omnipräsent, sondern auch beweglicher, dynamischer und komplexer sind als jemals zuvor. In dieser Welt wird die richtige Einordnung von Daten zur essenziellen Fähigkeit für uns alle. Mindestens aber die Fähigkeit, den richtigen Kontext von Nachrichtenmedien zu fordern. Die Pandemie wird ein Ende haben, doch wir sind dazu verdammt, in einer andauernden „Infodemie“ zu leben – umgeben von manipulierten Informationen, Fake-News und Panik-, wenn wir nicht bald lernen, Daten objektiv zu bewerten.

Sasha Kulek
Data Strategist Jung von Matt AG

Quellen:

- ¹ Coronavirus news is dominating readers' attention (2020): in: Vox, [online] <https://www.vox.com/recode/2020/3/17/21182770/news-consumption-coronavirus-traffic-views> [11.04.2020].
- ² Why isn't the government publishing more data about coronavirus deaths? (2020): in: The Guardian, [online] <https://www.theguardian.com/commentisfree/2020/apr/02/government-publish-data-coronavirus-deaths> [13.04.2020].
- ³ COVID-19 United States Cases by County (o. J.): in: Johns Hopkins Coronavirus Resource Center, [online] <https://coronavirus.jhu.edu/us-map> [20.04.2020].
- ⁴ Porter, Theodore M. (1996): Trust in Numbers: The Pursuit of Objectivity in Science and Public Life, [online] <https://books.google.nl/books?id=oK0QpgVfINOC>.
- ⁵ Oft völlig überbewertet: Vorsicht bei Corona-Statistiken (2020): in: Heise online, [online] <https://www.heise.de/newsticker/meldung/Oft-voellig-ueberbewertet-Vorsicht-bei-Corona-Statistiken-4701773.html> [14.04.2020].
- ⁶ To understand the global pandemic, we need global testing — the Our World in Data COVID-19 Testing dataset (o. J.): in: Our World in Data, [online] <https://ourworldindata.org/covid-testing#germany> [20.04.2020].
- ⁷ Hohe Dunkelziffer: Zahl der Infizierten in Deutschland möglicherweise [...] (2020b): in: Deutsches Ärzteblatt, [online] <https://www.aerzteblatt.de/nachrichten/111854/Hohe-Dunkelziffer-Zahl-der-Infizierten-in-Deutschland-moeglicherweise-schon-bei-460-000> [15.04.2020].
- ⁸ COVID-19 United States Cases by County (o. J.): in: Johns Hopkins Coronavirus Resource Center, [online] <https://coronavirus.jhu.edu/us-map> [20.04.2020].
- ⁹ Gestorben „mit“ oder „an“ Covid-19?: Warum in Deutschland so wenige Corona-Tote obduziert werden (2020): in: Der Tagesspiegel, [online] <https://www.tagesspiegel.de/wissen/gestorben-mit-oder-an-covid-19-warum-in-deutschland-so-wenige-corona-tote-obduziert-werden-/25726918.html> [12.04.2020].
- ¹⁰ Coronavirus-Karte: Deutschlandweite Fallzahlen in Echtzeit (2020): in: Der Tagesspiegel, [online] <https://interaktiv.tagesspiegel.de/lab/karte-sars-cov-2-in-deutschland-landkreise/> [20.04.2020].
- ¹¹ Up to 650 000 people die of respiratory diseases linked to seasonal flu each year (2017): in: World Health Organization: WHO, [online] <https://www.who.int/news-room/detail/14-12-2017-up-to-650-000-people-die-of-respiratory-diseases-linked-to-seasonal-flu-each-year> [12.04.2020].
- ¹² How Bad Will the Coronavirus Outbreak Get? Here Are 6 Key Factors (2020): in: New York Times, [online] <https://www.nytimes.com/interactive/2020/world/asia/china-coronavirus-contain.html> [12.04.2020].
- ¹³ Lower death rate estimates for coronavirus, especially for non-elderly, provide glimmer of hope (2020): in: STAT, [online] <https://www.statnews.com/2020/03/16/lower-coronavirus-death-rate-estimates/> [13.04.2020].
- ¹⁴ Research Starters: Worldwide Deaths in World War II (o. J.): in: The National WWII Museum | New Orleans, [online] <https://www.national-ww2museum.org/students-teachers/student-resources/research-starters/research-starters-worldwide-deaths-world-war> [13.04.2020].
- ¹⁵ How a global pandemic led to a toilet paper shortage — and when it gets better (2020): in: New York Post, [online] <https://nypost.com/2020/04/09/how-a-global-pandemic-lead-to-a-toilet-paper-shortage/> [12.04.2020].
- ¹⁶ Oremus, Will (2020): What Everyone's Getting Wrong About the Toilet Paper Shortage, in: Medium, [online] <https://marker.medium.com/what-everyones-getting-wrong-about-the-toilet-paper-shortage-c812e1358fe0> [10.04.2020].