

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

29. Juli 2020

Warum nicht nur dem US-Präsidenten während der Pandemie mehr Datenkompetenz gut täte.

Für jeden hat sich während der Coronapandemie etwas geändert. Ich zum Beispiel habe meinen Umgang mit Medien verändert – vor allem, was die Coronaberichterstattung angeht. Warum? Vielleicht gibt die Auswahl folgender Schlagzeilen die Antwort preis:

- > »Lateinamerika wird zum Corona-Epizentrum« #1
- > »USA global am schlimmsten von Corona betroffen« #2
- > »Die Pandemie trifft den Balkan mit voller Wucht« #3

All diese Überschriften erschienen im Abstand von 3 Tagen zueinander. Balkan, Brasilien, USA – dort soll Corona derzeit jeweils »am schlimmsten« wüten.

Welches Land hat Corona wirklich »am schlimmsten« getroffen? Diese Frage ist offenbar nicht so trivial, wie es zunächst scheint. Ein datenkritischer Blick auf 3 Schlagzeilen während des Lockdowns soll dir dabei helfen, dir selbst ein Bild machen zu können.

Der Trump in uns – Warum du Zahlen stets relativ sehen solltest

Glaubt man vielen Berichten der letzten Wochen, schneidet niemand auf der Welt bei der Reaktion auf die Pandemie derzeit schlechter ab als das Land von »The Donald«. Das ist natürlich ein verlockendes Narrativ. »Die USA ist am schlimmsten von dem Coronavirus Sars-CoV-2 betroffen«, schreibt etwa der Münchner Merkur online, nur eines von vielen ähnlich lautenden Beispielen. #4

Aber was sagen die Daten wirklich?

Richtig ist, dass es in *absoluten* Zahlen in keinem anderen Staat mehr bestätigte Tote und bestätigte Infizierte in Zusammenhang mit Covid-19 gibt als in den USA. Fast 150.000 Menschenleben.

Richtig ist aber beispielsweise auch, dass in den USA mehr Weiße als Schwarze von Polizisten erschossen werden – in absoluten Zahlen. Eine geschmacklose Relativierung, für die der US-Präsident vor Kurzem (zu Recht!) kritisiert wurde. #5

Schließlich verschleiert diese *absolute* Darstellung, dass es als Schwarzer in den USA (13% Bevölkerungsanteil, aber 23% Anteil der durch die Polizei Erschossenen) *relativ* betrachtet mehr als doppelt so wahrscheinlich ist, durch die Kugel eines Polizisten zu sterben wie für weiße US-Amerikaner (60% Bevölkerungsanteil, 45% der Schusswaffenopfer). Natürlich kommt es auf die relative Zahl an, will man strukturellen Rassismus erforschen.

Bei Rassismus weiß das (fast) jeder.

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

Wer also über Coronatote oder Infizierte in einer bestimmten Region berichtet, sollte mit der gleichen Begründung Zahlen *relativ* darstellen. Wie steht es in dem Fall um die strauchelnde Supermacht?

- > **Relative Durchseuchung:** Anteilig belegen die USA, was die absolute Anzahl der insgesamt bislang Infizierten angeht, einen weniger schlagzeilentauglichen Rang 9 aller Länder weltweit (Stand 27. Juli 2020, etwa 1,3% bestätigte Durchseuchung der Gesellschaft). Der unrühmliche erste Rang gehört derzeit mit weitem Abstand dem Emirat Katar im Nordosten der arabischen Halbinsel (etwa 3,8%). Die Zahlen sind jedoch mit Vorsicht zu betrachten, denn sie hängen unter anderem von der Anzahl durchgeführter Tests ab.
- > **Relatives Sterberisiko:** Beim relativen Sterberisiko belegen die USA (Stand 27. Juli 2020) derzeit den elften Platz. Und die 10 Plätze davor? Die Plätze 1 bis 7 werden ausschließlich von europäischen Staaten belegt. Chile und Peru belegen Rang 8 und 9, danach folgt vor den USA Frankreich. Für Bewohner Belgiens etwa war es seit Ausbruch doppelt so wahrscheinlich, an Corona zu sterben, wie für Bewohner der USA. Im Kleinstaat San Marino in Norditalien war das Sterberisiko gar 3-mal so hoch wie in Übersee.

Auch diese relativen Zahlen haben noch einen Haken: Sie zeigen nur das *kumulierte* Risiko, also über den gesamten Zeitraum der Pandemie zusammengezählt. Für eine Einschätzung des aktuellen Risikos müssen ausschließlich die aktuellen Daten betrachtet werden. Einen gewissen Aufschluss gibt der Durchschnitt der letzten 7 Tage. ^{#6} Bezogen auf die letzte Woche führen die Liste der **bestätigten Neuinfektionen** der Oman und Bahrain an (260 bzw. 230 Fälle pro Million Einwohner). Die USA folgen hinter Panama und Südafrika auf dem immer noch katastrophalen fünften Platz (etwa 200 Infizierte pro Million Einwohner). Allerdings haben in diesem Zeitraum weltweit nur 3 andere Länder mehr Tests durchgeführt als die USA (2,4 Tests pro 1.000 Einwohner). ^{*1} Dadurch rücken sie in dieser Liste weiter nach oben.

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

Die Entwicklung der Todeszahlen hinkt dem »Infektionsgeschehen«, also der täglichen Ansteckungsrate, etwa 3 Wochen hinterher. ^{*2} Hier belegen die USA derzeit im 7-Tages-Mittel mit 2,8 Toten pro Million Einwohner Rang 13 hinter überwiegend süd- und mittelamerikanischen Ländern. Am lebensgefährlichsten ist Corona derzeit in Peru und Panama (täglich jeweils über 6 Tote pro Million Einwohner) ^{*3} und Kirgisistan (5,8).

Die Anzahl der durch (oder mit) Corona *unmittelbar* Gestorbenen gibt indes keine genaue Auskunft darüber, inwieweit die Sterblichkeit einer Gesellschaft *mittelbar* beeinflusst wird. Zum Beispiel, weil es weniger Verkehrstote oder mehr Suizide gibt. Aufschluss darüber gibt die Kennzahl der **Übersterblichkeit**, die in der aktuellen Debatte um das Coronavirus nur selten vorkommt. ^{*7} In dieser Kennzahl enthalten sind auch Virusopfer, die nicht mit, aber wegen Corona starben – etwa aufgrund allgemein schlechterer gesundheitlicher Versorgungslage oder weil Schwerkranke coronabedingt Angst haben, ins Krankenhaus zu gehen.

Die Übersterblichkeit ist die aussagekräftigste der 3 Kennzahlen, aber leider auch die, deren Berechnung am spätesten erfolgt. Sie vergleicht *alle* Todesfälle aus dem Pandemiezeitraum mit einem Mittelwert der Todesfälle aus vergangenen Jahren, ggf. bereinigt, um etwa Veränderungen der Altersstruktur einer Gesellschaft oder dem Umstand gestiegener Lebenserwartung Rechnung zu tragen. Auf diese Weise wird beispielsweise rückwirkend das Ausmaß der jährlichen Grippewelle berechnet. Die Übersterblichkeit gibt am besten Auskunft darüber, ob eine Gesellschaft aus gesundheitlicher Sicht *insgesamt* gut auf die Pandemie reagiert hat.

Die folgende Grafik setzt bei ausgewählten Ländern die Übersterblichkeit ins Verhältnis zu den gemeldeten Coronatoten. Liegt der Wert bei nahezu 100%, wie beispielsweise in Frankreich (97%), so entsprechen die gemeldeten Coronatoten (Frankreich: ca. 30.200) in etwa der gleichzeitigen (Gesamt-)Übersterblichkeit – also zusätzliche Tote im Vergleich zum Durchschnitt der letzten Jahre (Frankreich: ca. 31.100). Ein

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

niedrigerer Wert wie in Spanien bedeutet, dass die offizielle Anzahl der an Covid-19 Verstorbenen nur 60% der Übersterblichkeit darstellt, dass also statt *nur* 28.500 Coronatoten im Pandemiezeitraum etwa 47.500 mehr Menschen gestorben sind als normalerweise.

Schlau aus den Daten wird man also erst dann, wenn man sich vertieft damit beschäftigt. *4 Dann erkennt man auch leichter sogenanntes »Daten-Cherry picking«, also das bewusste Ausblenden von Darstellungen, die dem gewünschten Ergebnis widersprechen.

Die Berichterstattung über die USA ist dafür ein gutes Beispiel: Nur in (die tatsächliche Durchseuchung verzerrenden) *absoluten* Fallzahlen liegt das viertgrößte Land der Welt tatsächlich an der traurigen Spitze. Betrachtet man die eigentlich relevanten Kennzahlen, reicht es *nur* für einen Platz im vordersten Feld. Das hält von Tagesschau über Deutschlandfunk bis hin zu Spiegel, Zeit und so ziemlich jedem anderen Medium aber niemanden davon ab, weiter anderslautend zu titeln.

Warum ist diese Relativierung *gegen* Trump ein probates Mittel, während ihm der gleiche statistische Kniff (zu Recht!) zum Vorwurf gemacht wird?

Verunsichert? Gut so!

Wo also ist Corona »am schlimmsten«? Der Vergleich der *relativen* Übersterblichkeit beantwortet diese Frage zu einem gewissen Teil. Vieles verschweigt die Kennzahl aber: Sie sagt nichts aus über die Höhe eines Wirtschaftseinbruchs, eventuelle Auswirkungen auf die psychische Gesundheit (und deren Folgen), über Arbeitslosigkeit oder sonstige sozioökonomische Konsequenzen von Corona. Auch positive Effekte der Pandemie werden spätestens aus der Retrospektive ein Teil der Betrachtung sein; viele davon können wir heute unmöglich »messen«: Welchen Effekt hat diese Phase auf unsere Arbeit, eine Beschleunigung der Digitalisierung, den Umweltschutz, auf die Vereinbarkeit von Familie und Beruf oder auf zwischenmenschliche Beziehungen? #8 Welchen Einfluss hat all das auf unsere Lebenserwartung? #9

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

Es ist eine Herausforderung, bei so viel Komplexität nicht dem eigenen Lieblingsnarrativ zu erliegen. Das wissen Datenanalysten wie Max Roser nur zu genau. Er und sein Team betreiben die Website, von der ich mich im Augenblick zu Corona am besten informiert fühle: Our World in Data.

Von ihr stammen auch die Graphen, die du bisher schon im Text gesehen hast. Das Besondere an Our World in Data ist nicht nur die schiere Menge der verfügbaren aktuellen Daten, die Interaktivität oder die Tatsache, dass Roser sie kostenlos ins Internet stellt, sondern auch die Transparenz der Website. Quellen werden offengelegt und Zweifel an der Datenqualität transparent aufgezeigt (allerdings leider nur auf Englisch).

Vom sprichwörtlichen »Bauchgefühl« hält Roser nicht viel: »Die psychologische Forschung hat in den vergangenen 20, 30 Jahren gezeigt, dass wir keine guten intuitiven Statistiker sind«, ^{#10} erklärt er in einem Interview mit der Wochenzeitung Der Freitag. Einen nüchternen Blick, der viele Aspekte mitdenkt und auch Widersprüchlichkeiten und Unschärfen sichtbar macht und möglichst frei von eigenen politischen Überzeugungen ist – den hätten nur wenige Menschen und schon gar nicht auf so emotional aufgeladene Themen wie eine akute Pandemie, führt der Datenforscher aus.

Eine weitere Schwierigkeit beim Umgang mit Daten ist Roser wichtig: die Datenqualität. »Man muss hinterfragen, woher die Daten kommen, welche Auswahl getroffen wurde und was miteinander in Verbindung gesetzt wird«, rät er.

Doch trotz ihrer Ungenauigkeit sind Coronadaten genauer und transparenter als viele andere Informationen, auf deren Grundlage Entscheidungen sonst getroffen werden (zum Beispiel Befragungen und Konjunkturprognosen wie der »ifo-Index«). Selten waren die Datensätze zugänglicher, die Darstellungsmöglichkeiten vielfältiger. Aber absolut sicher sind auch sie nicht.

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

Warum Datenqualität wichtig ist – und wie du sie hinterfragst

Je transparenter die Datenquelle, desto besser die »Datenqualität«. Jede Kennzahl ist immer mit Hinblick auf die den Daten zugrunde liegenden Unsicherheiten zu beurteilen. Auch die Daten zu Corona haben unterschiedliche Qualität und bergen gewisse Unsicherheiten:

- > **Die Infektionszahlen** sind offizielle Zahlen der jeweiligen Regierungen – und einige davon waren schon in der Vergangenheit für das sorgfältige Frisieren so mancher Daten bekannt. #11 Nordkorea beispielsweise meldete vor wenigen Tagen erst den ersten Infizierten. #12 Hinzu kommt, dass Datenquellen zwar einigermaßen standardisiert sind, Länder sie aber dennoch nach etwas unterschiedlichen Kriterien erfassen dürfen. #13 Vor allem hängt dieser Wert von der Anzahl durchgeführter Tests ab. Die tatsächliche Fallzahl liegt stets um einen Faktor x höher, der gewaltige regionale Unterschiede aufweist. Der größte Vorteil dieser Daten: Innerhalb ihres regionalen Bezugsrahmens lassen sich an ihnen Trends ablesen – wo sind in den nächsten Wochen mehr Krankenhausbetten nötig? Werden die Todeszahlen sinken?
- > **Die Todeszahlen** sind insgesamt verlässlicher, weil Tote mit vorherigen Symptomen aktuell meist auf Corona getestet werden. Doch auch hier drohen *Manipulation* (die aus statistischer Sicht überaus unglaubwürdige Erfolgsstory aus China – angeblich 4 Tote von Anfang Mai bis Mitte Juli – ist wahrscheinlich staatlich manipuliert) und *Inkonsistenz* (teilweise werden nur in Krankenhäusern Verstorbene gezählt oder Daten geschätzt). Viele Länder haben außerdem keine ausreichende Infrastruktur, um verlässliche Daten zu erfassen. Und doch: Diese Daten sind das Beste, was wir haben – und sie sind tagesaktuell verfügbar.

Eigentlich müsste also gerade die Zeit für Datenjournalismus, kritisches Hinterfragen und Differenzierung sein. Eigentlich. Stattdessen

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

bombardieren Medien ihre Konsumenten immer noch zu oft mit absoluten Zahlen ohne Kontext: »In Deutschland ist die Zahl der mit Corona Infizierten auf 204.964 gestiegen.« Eine solche Information fördert ohne Einordnung keinerlei Verständnis und erst recht kein Interesse für einen vertieften Blick auf die Daten.

Das Einmaleins der Datenkompetenz

Wie könnte es besser gehen? Wenn du versuchst, Fakten aus Daten abzuleiten, verwende einfach folgende Checkliste:

- > Formuliere eine möglichst klare Frage (zum Beispiel: »Wie hoch ist das aktuelle Sterberisiko?«).
- > Lege eine Perspektive fest (weltweit im Zeitverlauf oder der Vergleich aktueller Daten auf Kreisebene).
- > Finde eine aussagekräftige Kennzahl (relative Übersterblichkeit, hilfsweise bestätigte Tote).
- > Am wichtigsten: Ordne dein Ergebnis ein (Datenqualität und »normale« Sterblichkeit im selben Zeitraum).

Damit lässt sich auch besser verstehen, wie gefährlich die Pandemie aktuell noch ist. Momentan sterben coronabedingt weltweit wieder über 6.000 Menschen am Tag – etwa alle 14 Sekunden ein Leben, zumindest nach offiziellen Angaben.

Doch auch dies ist eine absolute Zahl. An einem »normalen« Tag sterben weltweit durchschnittlich über 160.000 Menschen. Derzeit etwa 6.000 Coronatote bedeuten einen Anstieg der *durchschnittlichen* Todesgefahr um knapp 4%, für die Risikogruppen liegt er um ein Vielfaches höher. In der zweiten Aprilhälfte lag der 7-Tage-Mittelwert schon einmal eine Weile bei etwa 7.500 täglichen Toten, also etwa um 1/4 höher.

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

Haben wir also das Schlimmste hinter uns und sollten optimistisch sein?

Global gesehen muss diese Frage derzeit klar mit Nein beantwortet werden. Nachdem die Todeszahlen bis Ende Mai rückläufig waren (auf bis zu unter 4.000 Verstorbenen pro Tag), steigt diese Zahl seitdem anfangs langsamer, jüngst schneller wieder an. Der starke Anstieg der weltweiten Infektionszahlen garantiert, dass dieser Trend in den nächsten Wochen anhält.

In einer global vernetzten Welt ist jedes Land und fast jeder Erdbewohner in irgendeiner Weise durch eine Pandemie dieses Ausmaßes betroffen. Natürlich interessiert für das eigene Infektionsrisiko zunächst nur die Situation vor der eigenen Haustür. Gleichzeitig wird diese Krise fast alle betreffen, solange das Virus weltweit wütet und es keinen Impfstoff gibt.

Je besser möglichst viele von uns dabei die Daten verstehen, desto besser können wir als Gemeinde, als Nation und als Weltgemeinschaft reagieren. Und desto weniger Unsinn wird verbreitet.

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

Zusätzliche Informationen

- *1 Nur Luxemburg (12,5 Tests pro 1.000 Einwohner binnen der vergangenen 7 Tage), Bahrain (5,8) und die Vereinigten Arabischen Emirate (4,6) haben relativ gesehen noch mehr getestet.
- *2 Die meisten Todesfälle treten erst einige Wochen nach Krankheitsausbruch auf. Hinzu kommt eine Meldeverzögerung von einigen Tagen, die außerdem je nach Wochentag Schwankungen unterliegt (am Wochenende wird zum Beispiel weniger gemeldet).
- *3 Peru hat am 24. Juli etwa 3.500 frühere Tote nachgemeldet. Das verzerrt den aktuellen Mittelwert auf ein Ergebnis von 22 Toten pro Million Einwohner und Tag. Aus der hier genannten Zahl sind diese Toten wieder herausgerechnet.
- *4 Auf Deutschland bezogen bietet der Deutschlandfunk [eine gut gemachte Übersicht](#), die wöchentlich donnerstags aktualisiert wird.

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

Quellen und weiterführende Links

#1 N-tv über die Coronalage in Lateinamerika (2020)

 <https://www.n-tv.de/panorama/Lateinamerika-wird-zum-Corona-Epizentrum-article21908182.html>

#2 Der Stern über die Coronapandemie in den USA (2020)

 <https://www.stern.de/gesundheit/corona--usa-am-schlimmsten-betroffen---viele-neue-infektionen-im-sueden-9313676.html>

#3 Die SZ über die Coronakrise im Balkan (2020)

 <https://www.sueddeutsche.de/politik/balkan-coronavirus-1.4953841>

#4 Der Merkur gibt seine Einschätzung zu Corona weltweit (2020)

 <https://www.merkur.de/welt/coronavirus-news-infizierte-tote-zahlen-statistik-vergleich-laender-welt-europa-amerika-schweden-china-zr-13813046.html>

#5 Der Spiegel kritisiert Trumps Analyse zur Gewalt gegen Schwarze (2020)

 <https://www.spiegel.de/politik/ausland/donald-trump-us-praesident-relativiert-polizeigewalt-gegen-schwarze-a-55cc5ef9-cef7-4af4-af79-50067bb4d28d>

#6 Daten zu Genesenen und noch Infizierten sind zum Beispiel auf der Seite

 der Johns Hopkins Universität zu finden (englisch, 2020)
<https://coronavirus.jhu.edu/map.html>

#7 Our World in Data befasst sich in einem ausführlichen Artikel mit der

- Übersterblichkeit (englisch)

<https://ourworldindata.org/covid-excess-mortality>

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY

- #8** »Lagerkoller« daheim? Wie du dir kostenlos helfen lassen kannst,
 schreiben unsere Gastautorin Christine Müller und Dirk Walbrühl
<https://perspective-daily.de/article/1195/>
- #9** Welche Zutaten für ein langes, gesundes Leben nötig sind, verraten uns am
 besten die 100-Jährigen, schreibt Maren Urner
<https://perspective-daily.de/article/355/>
- #10** Das Interview mit Max Roser in Der Freitag (2016)
 <https://www.freitag.de/autoren/christine-kaeppler/99-prozent-koennen-irren>
- #11** Der Spiegel berichtet über Zweifel an den Zahlen aus China (2020)
 <https://www.spiegel.de/politik/ausland/corona-faelle-in-china-peking-zweifelhafte-zahlen-a-295ef443-5df3-406b-acff-360746eb27be>
- #12** Nordkoreas offizielle Coronazahlen in der Meldung bei der Tagesschau
 (2020)
<https://www.tagesschau.de/ausland/corona-nordkorea-101.html>
- #13** Our World in Data verwendet die Daten des European Center for Disease
 Prevention and Control (ECDC) und klärt über Unsicherheiten auf
(englisch, 2020)
<https://bit.ly/2P5IFH7>

Frederik v. Paepcke

Hast du wirklich mehr Ahnung von Daten als Donald Trump?

perspective-daily.de/article/1350/MHLexOoY