No.4 Data Delusion

JvM
GREEN PAPERS

THOUGHTS ON A WORLD IN THE MIDST OF RADICAL CHANGE

NO. 4 DATA DELUSION.   **APRIL, 2020**

HOW YOU HAVE BEEN MISUNDERSTANDING THE DATA BEHIND COVID-19 AND WHY DATA LITERACY IS THE MAKE-IT-OR-BREAK-IT SKILL OF OUR TIMES.

Our world has never been in a crisis of such magnitude with so much information so widely accessible, including by-the-minute updates on the number of COVID-19 infected and deceased. And it has never been the case that so much of this information was simply unreliable. We find ourselves amidst not just a pandemic, but an "infodemic" as well.

The rise of COVID-19 has seen an increase in overall news consumption.[1] This has created a frenzy for "live figures" among members of the public, with each one of us rigorously checking the latest dashboards, graphs and visualisations filled with updates on the COVID-19 numbers in our town, municipality, country and the world.[2] It has also given a massive push to data visualisation, with many newspapers and institutions creating "appealing-to-eye" and "easy-to-understand" interactive maps that show the harrowing constantly-growing numbers.

Watching the number of infected double every 3 days in New York, bringing it from 967 cases to over 238,831 cases within 32 days[3] has sent shivers down our spines, because we believe we understand the numbers that we see. At the end of the day, each one of us has primal trust in the objectivity of numbers, their comparability and their solidity.[4] But this trust is deceiving, because these numbers portray only part of the whole truth.

The problem with hard numbers is that we have never been taught to look beyond them and into the context around them. But it is looking at the context that reveals two fundamental problems with data that we ignore on a daily basis: measurement method and perspective.

MEASUREMENT METHOD.

Understanding how a certain statistic has been collected is essential to being able to understand

what it actually means. Let's take a look at the number of people infected by COVID-19. As media and governmental institutions report on the rising or decreasing number of infected across countries, they throw around hard-to-comprehend absolute values, omitting the most important number. That is the number of people being tested.

Interpreting the rate of infections without knowing the rate of testing is completely misleading: the only thing it actually reflects is how many infections have been recorded and shows very little insight on how spread the infection is in the country.[5]

Germany, where the disparity between the number of infected and deceased is so unprecedented, has been turning heads with many wondering what is going there. The answer is that Germany is catching a lot more infections through the superior number of tests being done. On the 12th of April 2020, Robert-Koch-Institut confirmed that they were testing at a 20.94 tests-per-thousand-people rate in comparison to UK testing at a 5.54-per-thousand-people rate.[6] Virologists suggest that that number amounts to a discovery rate of about 15.6 % in Germany, meaning that the real number of cases is somewhere over 930,000 and with an average discovery rate of just 6 % across the world,[7] the real number of cases globally could be far above 40 million.[8]

Another gloomier example is the number of corona-related deaths. There has been no internationally-uniform definition on what constitutes a corona-related death. The two approaches to measurement are: "COVID-19 positive deaths" and "death through COVID-19". While different definitions are applied across the world, the measurement sometimes differs within a single country. In Germany, Robert-Koch-Institute counts all deaths that are "COVID-19 positive", whereas the municipality of Hamburg registers only those deaths that have been identified as "deaths through COVID-19" post-autopsy.[9]

This measurement divergence leads to a disparate mortality rate between Germany and Hamburg (Germany = 3.19 %, Hamburg = 2.09 %).[10] Projecting into the future, these rates would predict a contrasting number of victims, with a difference of more than half a million deaths. A difference that can be critical when deciding how long the restrictions should be in place for.

SETTING THE DATA INTO PERSPECTIVE.

Finding the right frame of reference is key to interpreting the data at hand with certainty. Let's go back in time to February 2020, as COVID-19 was just working its way to Europe. Back then many critics were chastising scientists for raising panic, they were referencing the yearly number of global deaths from respiratory diseases – namely 650.000 people[11] — suggesting that COVID-19 was not even worth talking about.

But none of these critics understood the relation between this absolute number and the infection and fatality rate, which actually are decisive factors for how each disease can affect our society. Ignoring that frame of reference let many underestimate the danger of COVID-19.[12]

Now, in April we still do not know the exact mortality rate of COVID-19, but studies show that the value lies somewhere between 0.5 % and 13 %, and is probably closer to 1 %.[13] That 1 % makes us breathe out, thinking: "Well, 1 % is not even that much — I definitely wouldn't be in that 1 %!" But putting this 1 % into a global context can show us the true dimensions.

By looking at the bigger picture, one can quickly calculate what the global number of victims of COVID-19 by the end of the pandemic will be. With 1 % mortality rate and 60 % infection rate,

the number of victims will be 45 million people. A number that equals to more than half of all deaths in World War II.[14] A number that doesn't make one feel so invincible anymore. A number that should be enough to show why each one of us should #staythef***athome.

These aforementioned examples show us one simple truth — data can be misleading without context. Misunderstanding the measurement method can provide false information about the state of the crisis. Failing to set data within a broader perspective can make us believe that something is less relevant than it actually is. And this truth applies to every datapoint around us: from the matters of life and death up to the conventional consumption data.

Such as consumption of toilet paper. Newspapers, social media feeds and the world around us have been pointing fingers at people hoarding toilet paper, suggesting how selfish and unreasonable they are being. And whilst there is definitely a proportion of society who are guilty, we have to look for the context of these sales figures to see what is really happening here.

The toilet paper industry is split into two, largely separate, markets — B2B and B2C. The toilet paper made for the commercial market is a fundamentally different kind of paper to that what you buy in the supermarket: it is produced in huge rolls, the paper itself is thinner and more utilitarian and it is not packaged in multiple-roll packages. The B2B and B2C paper each own about 50 % of the whole toilet paper market.

Self-isolation has led to a shift in consumption: as a high amount of people are staying at home, they are not using restrooms at work, restaurants, bars or clubs, meaning that the demand for B2C toilet paper has increased by a staggering 40 %.[15] This means that whilst the demand for consumer toilet paper has increased, the supply of commercial toilet paper cannot satisfy it, causing a deficit. So whilst many of us have been blaming that guy next door for buying all the toilet paper, we might have missed the bigger picture all along.[16]

Understanding the context has never been harder than now: the current situation is completely out-of-the-ordinary. It creates trends that are so short-lived that we can only make wild guesses in an attempt to understand the context around them.

Nevertheless, there are 3 questions that everyone can ask in order to get one step closer to championing data literacy:

1. HOW WAS THIS DATA COLLECTED? Understanding how data has been collected and how a variable is measured gives us understanding of what that data means. Don't be afraid to be critical – if the collection and measurement methods seem off, it is quite possible that they are.

2. WHAT IS THE BACKGROUND CONTEXT ON THE MATTER YOU ARE LOOKING AT? Think about what other figures can be important in relation to the value, question the baseline or the comparison value used, search outside-the-box for reasons that might explain why the values are different than expected (like the B2C and B2B toilet paper market).

3. HOW RELIABLE IS THE SOURCE THAT YOU ARE USING? Check to see if the figure has been referenced in other reliable sources and if it has, scrutinize if the context has been altered or omitted. Remember, it could be even a human copy-and-paste error rather than malicious behaviour that can change the meaning of figures.

However exceptional our situation is now, one thing is clear: we live in a world, where data is not only omnipresent, but more fluid, dynamic and complex than ever before. In this world, it becomes essential for everyone to be able to identify the context of each data point we interpret. Or at least demand that context from the news providers. The current crisis will pass,

but we are doomed to live in a constant "infodemic", surrounded by information manipulations, fake news and panic, if we do not get better at understanding the context of data.

Sasha Kulek
Data Strategist, Jung von Matt AG

Sources:

[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: 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 (1996): Trust in Numbers: The Pursuit of Objectivity in Science and Public Life, [online] https://books.google.nl/books?id=oK0QpgVfIN0C.

[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 , [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: 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: 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].