

Harvard Data Science Review • Issue 2.2, Spring 2020

Can the Coronavirus Prompt a Global Outbreak of “Distributional Thinking” in Organizations?

Nathan E. Sanders¹

¹Applied Analytics Division, WarnerMedia, Boston, Massachusetts, United States of America

The MIT Press

Published on: Sep 29, 2020

DOI: <https://doi.org/10.1162/99608f92.a577296b>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

Media reporting and public discourse in the spring of 2020 has been dominated by the discussion of statistics relating to the COVID-19 outbreak and how to interpret them. Reasoning about these numbers has inspired fear as well as hope in communities worldwide. This environment provides a lens, and a rare scale, for data scientists to investigate how complex statistical topics are communicated to, understood by, and acted upon by diverse audiences. In particular, this crisis has put a premium on ‘distributional thinking,’ a mindset for reasoning about variation that is front and center in the response to the coronavirus as well as broadly relevant to organizations. This kind of thinking is already widespread among data scientists, but the challenge we face is to instill it across our organizations to equip them to tackle complex problems whose response should be informed by data and evidence. Fortunately, ours is not the first domain to encounter this challenge. I suggest learning from the example of modern social justice movements, who have evolved strategies to generate widespread appreciation of issues with distributional considerations, like the disparate impacts of environmental pollution and inequities in policing. I point to movement-building techniques like participatory research and shared leadership for lessons on how to grow the capacity for distributional thinking within companies, NGOs, agencies, and other organizations.

Keywords: communication, industry, coronavirus, equity

1. Introduction

The outbreak of the coronavirus SARS-CoV-2 responsible for the disease designated COVID-19 (Chinazzi et al., 2020; C. Wang et al., 2020) has had tragic impacts already around the world, with further devastation threatened in the months to come (Ferguson et al., 2020). But so too has it created opportunities for galvanization of communities and societywide dialog that we can look to for positive signs about our collective path forward.

To the data scientist, one noteworthy outcome from this pandemic is that it has prompted many millions of people to think in terms of distributions. Indeed, this way of thinking has been instrumental to the debate about and design of our governmental and civic response to this crisis. For data scientists, this is a rare opportunity to attract attention to and learn from how complex statistical ideas are communicated to, interpreted by, and responded to across diverse sectors.

The idea that thinking statistically in terms of distributions is important to decision makers and actors of all types is, of course, hardly new. Fields including statistics, risk management, decision science, statistical physics, and more are largely organized around probabilistic reasoning of this kind (see, e.g., Tversky & Kahneman, 1974). To quote Snee (1999), “If there was no variation, there would be no need for statistics and statisticians.” On this subject, the business literature has for nearly a decade featured admonitions to avoid the

“flaw of averages,” Sam Savage’s maxim for overturning the ‘show me *the* number’ corporate culture. Savage’s message is that a bottom line based on average assumptions is not necessarily the average bottom line (Savage, 2002, 2012). The literature on ‘probabilistic’ and ‘statistical thinking’ formalizes cognitive models of how “general knowledge and beliefs, along with descriptions of situations, lead to mental models that are used to assess probabilities” (Johnson-Laird, 1994 p.206; see also Wild & Pfannkuch, 1999).

However, education researchers have repeatedly documented the difficulty learners face in internalizing thinking about distributions (Ben-Zvi et al., 2004). In the public square, whether or not they are the norm, examples abound of media reporting that embraces the simplicity of singular answers to complex questions (Pew Research Center, 2009; Secko et al., 2013). From my own experience in industry (Sanders, 2019b), I can attest that, far too often, businesses resort to this kind of thinking too. In environments where decisions need to be made quickly and available data is limited, data scientists are prone to resort to it as well.

I suggest the term ‘distributional thinking’ as a useful construct for statistical thinking in terms of distributions in the organizational context. I define the term in Section 2. In Section 3, I examine case studies of how the coronavirus outbreak has prompted widespread use of distributional thinking. These case studies illustrate for data scientists the challenges in communicating effectively about similar nonsingular problems that arise in organizations, which I explore in Section 4.

We need not restrict our attention to the present crisis to find excellent role models for propagating distributional thinking. We are beneficiaries of decades of progress in this area by social movements. In tackling topics from income inequality to environmental justice, these movements have successfully appealed to policymakers and citizens to consider the variation in impact of actions taken across segments of our population. I conclude by searching for productive parallels in effective statistical communication between the response to the coronavirus, the practice of data science in organizations, and these ongoing social movements in Section 5 that lead to recommendations in Section 6.

2. Defining Distributional Thinking

In terms that will be familiar to data scientists in organizations across varied domains, distributional thinking can be defined as the frame of mind for considering the outcome of a process as not just a singular state of being, but rather a pattern of alternatives and their likelihoods. This means holding in mind not only the probability of a particular outcome to occur, but also the fact that the process will have a pattern of impact on different agents or elements of the system depending on a variety of baseline characteristics and possible interventions. Furthermore, a distributional thinker must appreciate that this variation may have both random and systematic components. Distributional thinking has an opposite that we can call ‘singular’ or ‘deterministic thinking.’

2.1. Examples of Distributional Thinking

The best examples of distributional thinking are ones that suggest a demonstrably better answer than a singular thinking approach. Consider local weather forecasts that try to predict the likelihood that it will rain tomorrow in a given city. The appropriate behavior for a resident of that city based on the forecast should depend on several factors and how they are distributed.

Suppose the forecasted likelihood of rain is low over the course of the day, but highly peaked during the morning commute. A singular thinker might risk leaving their umbrella at home for the day and get wet, but a distributional thinker will recognize that their chance of exposure is high and be prepared. Likewise, if the overall likelihood of rain is low, but is spatially distributed in concentration over the resident’s path to work.

The distributional thinker should also consider their particular context and the available interventions. A resident carrying a waterproof case to work risks less than one transporting something more delicate. And the impact of being exposed to rain might be mitigated if they can choose a reliable train over one that is more likely to keep them waiting. These are all straightforward considerations, but, even for so universal a target of reasoning as the weather, there are challenges in communicating information that enables all audiences to think distributionally in an accurate way (Gigerenzer et al., 2005).

As an example with higher stakes for organizations (and individuals), consider long-term financial investment. While seeking to maximize yield, the optimal investment strategy requires minimization of risk at some level. The singular strategy (picking the single asset with the highest predicted rate of return) is vulnerable to the isolated performance of that asset. The distributional thinking approach does not necessarily dictate any particular investment strategy, but it does require consideration of diversification. The distributional thinker would evaluate the risk profile of each individual asset and the exposure propagated from it to their portfolio. That means not investing singularly in the asset with the highest predicted yield, attractive though it may be.

2.2. Connections to Other Paradigms

Distributional thinking invokes several important statistical concepts including:

1. Probability, that is, reasoning according to a probability distribution and eschewing point estimators for distributional ones.
2. Dependency, that is, recognizing and attempting to quantify the effect of one component of a system on another.
3. Robustness, that is, anticipating that a quantity may not be distributed in a straightforward or normal way because some effects may have highly skewed impacts.
4. Nonlinearity, that is, identifying when the impact of an effect may extrapolate to new cases in complex ways.
5. Communication of uncertainty, that is, presenting the plausible variation of an inference or prediction using visualizations or other means to emphasize its significance.

However, to be broadly relevant to all of the roles that contribute to decision making in organizations, distributional thinking should not extend to all the expansive tenets of statistical thinking. Statistical thinking at its most general involves transnumeration between multiple representations of data, interrogation via mathematical modeling, and many more elements that go beyond our purposes (Wild & Pfannkuch, 1999). The distinction is this: a data scientist needs to engage in technical processes like modeling to contribute their part to the organization, but to take advantage of the outputs of this modeling it is sufficient for their stakeholders to think distributionally. Distributional thinking most closely corresponds to the “consideration of variation” within the Wild and Pfannkuch (1999) framework.

Though it has its own “holes” (Gelman & Yao, 2020), the Bayesian method provides a natural computational approach to distributional thinking. Bayesian models can marginalize over a robust range of possible values of model parameters to generate the posterior distribution of a quantity of interest in a way that addresses all the statistical concepts enumerated above. Of course, the Bayesian framework does not have a monopoly on probabilistic reasoning, and distributional estimators can also be generated from frequentist methods (Xie & Singh, 2013). Moreover, probabilistic modeling is increasingly an active area of research within deep learning (Y. Gal & Ghahramani, 2016; Tran et al., 2018; Wilson & Izmailov, 2020).

Lastly, there is a robust literature and practice around ‘systems thinking.’ This is a broad framework that, while notoriously difficult to define (Arnold & Wade, 2015), parallels statistical thinking in many ways. A systems thinking approach involves seeing ‘wholes rather than parts’ and recognizing interconnectedness.

3. Coronavirus Case Studies

In the days after the COVID-19 coronavirus outbreak was declared a global pandemic on March 11, 2020 (Sohrabi et al., 2020), three topics have dominated much of the media and public conversation:

1. How the individual risk from infection varies with age (see, e.g., Chen et al., 2020; Lu et al., 2020; Riou et al., 2020; Verity et al., 2020; W. Wang et al., 2020; Wu & McGoogan, 2020).
2. How we can all contribute to ‘flattening the curve’ of the virus’s spread (see, e.g., Ferguson et al., 2020; Tang et al., 2020).
3. Whether or not nations are equipped to perform the diagnostic testing needed for a sufficient response (Binnicker, 2020; World Health Organization, 2020).

Engaging with each of these topics requires a significant level of distributional thinking.¹

3.1. Distributional Thinking and COVID-19: Age-Dependence

Take the question of age first. Probably all contemporaneous readers of this piece will have already asked themselves, what danger does an infection pose to me, or my parent, or my child? An elderly person who only considers the overall mean fatality rate would underestimate their personal risk from infection (W. Wang et al.,

2020), and a young person who only considers their specific fatality risk would undervalue the community impact of their potential to spread the virus (Tang et al., 2020). Clearly, this is a case where an emphasis on a single value misses the forest for the trees.

Adding complexity to this macabre calculation is that much uncertainty remains in the measurement of both the baseline and the age-dependence of the case fatality rate (Battegay et al., 2020; Kobayashi et al., 2020; Lin et al., 2020), and the true relationship depends on yet other factors such as personal medical history and access to care (Chen et al., 2020; Wu & McGoogan, 2020). Accounting for this uncertainty compounds the age-distributed nature of this risk and requires multilevel distributional thinking.

3.2. Distributional Thinking and COVID-19: Flattening the Curve

As a second example, consider the question that countless public service messages, editorials, and media reports have sought to address in the past few weeks (e.g., Anderson et al., 2020; Burkert & Loeb, 2020; Fisher & Heymann, 2020, to cite just a few), why should I care about ‘flattening the curve?’

As so many of us have come to learn, this epidemiological term of art refers to the distribution of diagnosed infection cases over time. The point of flattening the curve is not merely delaying when the outbreak will peak (the mode of the time distribution), but rather decreasing the concentration (the amplitude of the peak) so that hospitals are never overwhelmed with cases exceeding their capacity at any one time (Mikolajczyk et al., 2009). Understanding the advice of public health officials around the world to practice “social distancing” in order to flatten the curve (Wilder-Smith & Freedman, 2020) requires distributional thinking, as well.

3.3. Distributional Thinking and COVID-19: Testing

Lastly, distributional thinking of a yet higher order is needed to appreciate the role that testing can play in coronavirus response. Ideally, a mass testing regime would allow groups to interact with each other without fear that they are encountering an infectious carrier of the virus. It could at best even allow individuals known to have recovered from the disease to interact ‘normally,’ freed from social distancing measures. This ideal scenario requires that those previously infected and recovered can no longer transmit the virus to others, that those previously infected are immune from reinfection, and that we can achieve full testing coverage and record and share testing results between all those who may come in contact. At present, it must be noted, it remains uncertain whether and after how long previous carriers are no longer infectious (Lan et al., 2020; Woelfel et al., 2020).

However, there are further, distributional complications. As is familiar to immunologists and statisticians alike, a positive or negative test result does not fully collapse the probability function of infection because every test will have some rate of false-negatives and false-positives (e.g., Andreotti et al., 2003; Hoffrage et al., 2000). That means we must think of the risk from contact with any individual as uncertain, even when their test result is known. Furthermore, diagnostic failure rates may vary with both physical characteristics (e.g., age and prior

state of health) and environmental/situational ones (e.g., how carefully the test was done, diet and hygiene, and how much time has transpired since the individual’s infection; Klarkowski et al., 2014). The risk of a false-negative propagates from the individual to the population: if the unknown carrier is allowed to socially interact, they may infect others and raise the number of future cases that can be expected in the general population (Quade et al., 1980; W. Zhang et al., 2020). For all these reasons, the appropriate action to take upon testing outcomes will have complicated dependencies arising from risks that can be mitigated, but not eliminated.

In an acute sense, our individual safety and collective health during the present pandemic depends on our own, our neighbors’, and our leaders’ ability to internalize and act upon these complex topics. Surely there is a positive role for statisticians and data scientists to play in informing political leaders and the public about these considerations. I have no doubt there will be rigorous studies of the effectiveness of public messaging on these topics in the months and years ahead. For now, we are left to look at examples from prior incidents and other domains for guidance, and to seek multiple and innovative approaches to building a shared understanding of these issues across the population.

4. Distributional Thinking in Organizations

The same type of distributional thinking challenges facing the public confronting the coronavirus arise constantly in industry and other organizational domains, particularly when potential consequences are highly unequal across different outcomes or when highly skewed outcomes are possible. Simple, common examples can illustrate each effect.

4.1. Distributional Thinking Is Critical to Business

Revenue projection is a fundamental financial task for any organization, be it a large corporation, small business, or government (e.g., Duran, 2008; Reddick, 2004). Any such projection should be thought of in distributional terms because of the unequal risk across the spectrum of outcomes. Often, the potential consequences of a low-end outcome, for example, missing payroll, may far outweigh the potential benefits of a high-end one. An informed executive, thinking distributionally, will ask a data scientist to forecast not only what the firm’s revenue will likely be in the next quarter, but also what the probability is that revenue might fail to meet certain strategically important thresholds, and how that risk is distributed across products or other components of the business. We should certainly hope that our political leaders are considering the risks, and distribution of risks, associated with coronavirus response policies with a similar level of consideration today.

Predictions from recommendation engines like those used in online content personalization or streaming media services (Ricci et al., 2015; S. Zhang et al., 2019) can also benefit from probabilistic modeling to address the so-called long tail problem (see, e.g., Agarwal et al., 2019). A recommendation with a small chance of extreme liking (e.g., one belonging to the long tail of a right-skewed distribution) may be much more useful to the consumer (and to the firm) than recommendations with a high chance of mediocre liking (e.g., one falling near the mode of the same skewed distribution). This is the difference between the user finding their new favorite

band or TV show, which they will then invest dozens of hours on using the service, versus scrolling by it to the next recommendation. An effective implementation of a recommendation system should consider the full potential preference distribution of its users, and there are a variety of probabilistic models available for this purpose (Abdollahpouri et al., 2019; Santos et al., 2010; Valcarce et al., 2016).

Note that two apparently-distinct aspects of distributional thinking are evident in these examples, and both are critically important. These are 1) the dependence of the outcome on predictor variables and 2) the distribution of probability given any one set of dependencies (the stochasticity). As an example of the former, one can make point estimates conditioned on various factors by adding independent variables to a regression routine. This allows modelers to move beyond asserting a mean fatality rate for COVID-19, for example, to estimating how risk depends on age. An example of the latter would be a model capable of predicting a full probability distribution of health outcome rates, which could itself be a function of age.

4.2. Barriers to Distributional Thinking in Organizations

Unfortunately, data scientists in industry and other fields routinely neglect these components of distributional thinking for understandable reasons.

Hullman (2020) highlighted that the presentation of uncertainty in data analysis in media and beyond is “an exception rather than the rule.” Boukhelifa et al. (2017) documented practices for coping with uncertainty across 12 different applied domains, revealing systematic differences in approach across fields and (alongside many positive examples) a tendency among some to ignore uncertainty. Both Boukhelifa et al. (2017) and Borghouts et al. (2019) found that analysts generally focus on “minimizing” uncertainty (for example, by removing outliers) rather than “exploiting” uncertainty (in the positive sense of viewing uncertainty as a source of information). Oxbury (2018) provided a pedagogical tour of the implications of variability in modern data science workflows, undercutting the ‘religious observance’ granted to certain point estimation and goodness-of-fit measures. He showed examples where significant “variation occurs for a single problem, in a single population, using a single modelling method, with fixed parameters” and emphasized that the outcomes of modeling processes conventionally approached as deterministic (like gradient descent and multilayer perceptrons) can vary with factors like data ordering.

There are certain instances where ignoring elements of distributional thinking will seem natural. Potentially important conditional dependencies may be overlooked when data describing them are not available. It may be expedient to rely on maximum likelihood point estimators for machine learning models when probabilistic implementations would take longer or have not been developed. But even when probabilistic information is available, data scientists in organizations often communicate only mean estimates. After their detailed qualitative study of current practice, Boukhelifa et al. (2017) called for “data workers” to more fully integrate uncertainty in the reasoning process going forward. In their playfully titled paper “Ignorance Is Bliss,” Pappenberger and Beven (2006) argued that it is untenable to neglect the analysis of

variability and presented a code of practice for scientists that centers on communication of uncertainty to users of data.

Even when there are justifiable limiters on the full proliferation of distributional modeling, they need not be preventative of distributional thinking. Data scientists can estimate and communicate the likely impact of neglecting these distributional effects even when sufficient data or ideal modeling techniques are out of reach (see, e.g., Mousavi & Gigerenzer, 2014). In cases where conclusions are unaltered with or without probabilistic modeling, these behaviors will still reinforce an instinct for distributional thinking that will be valuable to the organization.

It must be acknowledged that simplicity is also a virtue in statistical communication and that an emphasis on nuance is not beneficial if it impairs fundamental understanding. But this impairment, if recognized, need not be considered permanent. Applying social representation theory to computational communication through data visualization, Foucault Welles and Meirelles (2015) explain that collective understanding is achieved through a “process involv[ing] two central steps: objectification, where a relatively complex concept becomes simplified to the point where common understanding is possible, and anchoring, where the simplified object is interpreted through a lens of preexisting understandings about related objects.” It is, naturally, possible for unfamiliar and difficult concepts to become more familiar, at a cost of time and effort that will often be well justified.

5. Lessons on Building Capacity for Distributional Thinking

If corporate decision makers struggle to internalize distributional thinking, they are certainly not alone. Some failures during the chaotic onset of the coronavirus crisis notwithstanding, the examples from Section 3 demonstrate the remarkable capacity for a wide range of consumers of information to become familiar with distributional topics when it becomes imperative to do so. In general, data scientists can benefit from much prior work on how to create distributional thinkers.

5.1. Conceptualizing Variation Is Central to Statistics Education

The difficulty in thinking distributionally has been well documented through quantitative and qualitative studies of students (and teachers) by education researchers (Ben-Zvi et al., 2004) and conceptualizing variation has long played a central role in statistics education (I. Gal & Garfield, 1997). Lee and Meletiou (2003), for example, documented foundational challenges in undergraduate statistics students’ understanding of histograms, and their ability to move beyond a “deterministic mindset” to recognize the stochasticity involved in generating data sets. The statistics education literature suggests a variety of techniques for understanding and improving individuals’ reasoning about variation. Peters (2011), Pfannkuch and Wild (2004), Reading and Reid (2006), and Wild and Pfannkuch (1999), among others, have presented detailed cognitive frameworks for educators seeking to help students build a robust understanding of variation. A theme within this literature is the need to progress learners along a cognitive pathway from singular to distributional thinking, to ‘nurture’

their conception of variation (Reading & Reid, 2006), by presenting notions of variation in a variety of contexts over time.

Generalizing beyond statistics, Brown et al. (2014) emphasize the importance of practice, and “self-testing” on complex tasks, to achieving learning and long-term memory. This dovetails with appeals to persistent engagement that arise from another domain, social movements.

5.2. Social Movements as Distributional Thinker Generators

Many social movements have a powerful focus on equity, meaning fairness in the provision of benefits that takes into account individual circumstances (Espinoza, 2007). The Black Lives Matter movement centers on racial equity in policing practices (Taylor, 2016). The Occupy movement has focused on equity in the distribution of income and wealth (Hammond, 2015). The Environmental Justice movement urges policymakers to consider not only the bulk impact of pollution, but also its specific impact on the most vulnerable and overburdened populations, especially when they are in the minority (Brulle & Pellow, 2006).

Consideration of equity inherently requires distributional thinking because it entails acknowledging the differential impact of policies and the potential for asymmetric outcomes depending on individual characteristics. Successfully making change on issues of equity requires the transformation of diverse stakeholder groups into distributional thinkers, each prepared to recognize and evaluate the disparate impacts of their and societies’ actions and policies. In a sense, the purpose of social movements is to spread distributional thinking and focus it on issues of key societal concern.

5.3. Successful Techniques Evolved by Social Movements

In a broad overview of lessons from her experience in three different social movements (1. immigrants’ rights, 2. Occupy, and 3. LGBTQ and Two-Spirit movements), Costanza-Chock (2018) identified maximizing participation as a recurring theme to overcoming cross-cutting challenges encountered universally by movements. One set of techniques adopted widely by movements to facilitate the shift to distributional thinking are ‘participatory research’ strategies. These approaches seek to engage a broad and representative swath of the affected community in research about issues of concern to the movement, for example, pollution in the context of Environmental Justice (Lockie, 2018; Minkler et al., 2008; Schlosberg, 2003; Shepard, 2002).

The goal of this participatory engagement is to increase the representativeness and responsiveness of research practice to the actual needs of the community and, simultaneously, to increase the capacity of members of the community to decide about, advocate, and mobilize for their own interests. In a meta-analysis of 25 community-based participatory research studies, Spears Johnson et al. (2016) highlighted the need for true engagement to realize actual positive social impacts and to avoid ‘top-down’ approaches that reinforce existing hierarchies.

Environmental Justice research examples illustrate how quantitative methods and social movements can intersect to enable communities to advocate for their specific interests (e.g., Davis & Burgoon, 2015; Mah, 2017). Binet et al. (2019) describe how they recruited and engaged “resident researchers” throughout the instrument design, data collection, and data analysis of their Healthy Neighborhoods Study in greater Boston. Part of my own work in the same region has been to democratize access and facilitate the ability to analyze complex environmental regulatory data sets so that local community organizations and smaller NGOs can probe and understand the local and distributional impacts of policies and pollution in New England (Sanders, 2019a).

In their work promoting equity in higher education, Kezar and Holcombe (2017) have promoted the model of “shared leadership.” In this context, the distributional thinking challenge is to be mindful of how traditional practices of academic administration confer unequal resources, burdens, social capital, and prestige to students, faculty, and other stakeholders as a function of traits like gender and race. The shared leadership approach eschews the leader/follower binary by creating organizational structures supporting members at all levels of authority to take on aspects of leadership and to contribute their perspectives.

Lessons from social justice movements are already carrying over to data science through the growing field of data ethics (Olteanu et al., 2019). As a result, a two-way bridge of knowledge is emerging between movement advocacy and data science, particularly as relates to the appropriate use of data and avoiding social bias in modeling practices (Diakopoulos et al., 2017; Leonelli, 2016). This may lead to a broader proposition of data science for political action and social impact (Green, 2018). If data science is to serve this role, we should look to social movements to inform the engagement model for our work as data scientists in addition to our data handling and modeling practices.

6. Recommendations for Organizations

Practitioners in all of the domains discussed above encounter problems where distributional thinking—careful analysis of uncertainty, contemplation of disparities of impact, and consideration of how outcomes depend on multiple factors—is critical. Participatory methodologies emerging from social movements can be a valuable tool for organizations to build the capacity for distributional thinking and, in doing so, better integrate data science in their decision making and other practices.

6.1. Recognize That Movement Engagement Principles Fit With Organizations

In his recent guide for companies looking to make the most of data science teams, Berinato (2019) urged businesses to move beyond “unicorn”-centric frameworks. In this incumbent model, one individual is expected to have all “talents” necessary to the success of a data science project. Instead, he recommends using a collaborative team approach to tackle data-centric projects, even when it requires individuals to abandon functional boundary lines and learned strictures. Addressing six distinct talents ranging from data analysis to

storytelling, Berinato wrote: “Overcoming culture clashes begins with understanding others’ experiences...this exposure is meant to create empathy among team members with differing talents. Empathy in turn creates trust, a necessary basis for effective teamwork.”

With the perspective of an organization as a platform for establishing empathy among different stakeholders in order to contribute varied skills to address a shared project, the relation between the applied setting for data science and the social movement becomes clear. Berinato provides some useful case studies of this collaborative approach from the corporate sphere. Augmenting these case studies with the examples from mission-driven organizations and social movements discussed in Section 5 suggests the following recommendations.

6.2. Use a Participatory Engagement Model

The key lesson to be learned from social movements for data scientists operating within organizations is this: imparting a mindset for distributional thinking requires intensive engagement.

Evidence-based strategic decision making and technical product development is enhanced by distributional thinking among all stakeholders. The evolved practice of social movements suggests that bidirectional communication and true collaboration is required to generating a widespread capacity to think distributionally. One-way instruction from or to the data science team does not achieve this. The more participatory and collaborative the data measurement and modeling process is, the more representative it can be of concerns and needs from across the organization and the more relevant the resulting product or information will be to those who need it.

The shared leadership notion of Kezar and Holcombe (2017), developed in the higher education equity context, has notable similarity to the team structure recommended by Berinato (2019) for industrial data science projects. Both call for different perspectives or talents to be elevated in service of achieving goals of the organization despite their location at different levels of the social hierarchy or functions in the business. Through this kind of collaboration, data scientists and their counterparts can establish a common dialog about data collection issues, model design choices, and other statistical topics that will leave the whole organization better equipped to practice distributional thinking.

6.3. Work to Find a Common Language

Part of the work of establishing a participatory and collaborative relationship between data science and counterpart teams is to align on a shared language for the topics under study and the metrics for assessing them (Malone, 2020). In the climate and Environmental Justice domain, communicating how to reason about the distributed risk and uncertainty associated with climate change is a fundamental, long-studied challenge (Palmer, 2000; Stern, 2014). Findings indicate that even common words such as ‘likely’ can be interpreted

differently across different groups (Budescu et al., 2009), suggesting significant linguistic challenges to a shared distributional mindset.

Informed by the social debate around response to climate change, Budescu et al. (2009), Morton et al. (2011), Spiegelhalter (2017), and many others have established both verbal and numerical recommendations for communicating more effectively under these conditions. These recommendations are, in general, not based on brilliant insights of individual scientists or discoveries in the lab. They stem from study of how stakeholders actually understand and respond to technical and nontechnical terms when used in different contexts and presented with different frames. It may not typically be practical for data science teams to do carefully controlled trials of response to verbal or visual communication tools in their organizations, but highly informative qualitative data can be compiled regularly through a participatory engagement process.

6.4. Plan for the Participatory Approach to Take Longer

For the participatory approach to succeed, industrial data scientists must successfully integrate the considerations raised by business stakeholders and also communicate their thought process around technical decisions at each step in the process. This engagement will add time and complexity to projects that may not be practical in every case. But the lesson suggested by the Environmental Justice literature compiled by Spears Johnson et al. (2016) and social science research on generating collective understanding of complex data (Foucault Welles & Meirelles, 2015) is that participatory processes and repeated engagement will yield long-term organizational benefits.

6.5. The Path Forward

This kind of participatory dialog will be important to coronavirus response, as well. Communications failure is unfortunately all too common in disaster response (Donahue & Tuohy, 2006). In the infectious disease context, specifically, the exchange of information between public health officials and stakeholders, including hospitals, clinicians, and ultimately individual patients, is a longstanding concern (Akhlaq et al., 2016). The challenge is exacerbated for more vulnerable populations that may experience disproportionate impact from a pandemic (Nick et al., 2009). Communicating statistical information effectively, generating an “informed and rational internal dialog” in others, may be especially important during the COVID-19 pandemic to help channel fear and apprehension toward productive mitigations (Meng, 2020).

In public health, too, participatory design processes have proven relevant (Revere et al., 2014). The widespread uptake of interactive epidemiological simulation and data visualization tools (e.g., Cicalò & Valentino, 2019; JHU CSSE, 2020; Stevens, 2020) to grasp and anticipate the spread of COVID-19 demonstrates that enabling others to use the technical tools of our trade can facilitate broader shared understanding.

At the intersection of social and technical disciplines lies opportunity to advance data science and equity alike through the widespread use of new quantitative methods. With some luck and much effort, the exchange of

ideas between practitioners in these fields may even be part of the solution to the pandemic that faces us today.

Acknowledgments

I thank the anonymous reviewers for their very helpful comments as well as the following individuals for insightful feedback on this work: Jonathan Foster, Rose Hendricks, Matt Marolda, Xiao-Li Meng, Shannon Morey, and Arjun Sanghvi. I thank my colleagues at the American Institute of Physics (AIP) TEAM-UP initiative, particularly Arlene Modeste Knowles and Ed Bertschinger, for pointing out important connections between social equity movements and the research enterprise. I thank my colleagues affiliated with the Communicating Science Conference series (ComSciCon) for 8 years of collaboration that has been fundamental to forming my personal perspective on communication described here.

Disclosure Statement

Nathan Sanders has no financial or non-financial disclosures to share for this article.

References

- Abdollahpouri, H., Burke, R., & Mobasher, B. (2019). Managing popularity bias in recommender systems with personalized re-ranking. *arXiv*. <https://doi.org/10.48550/arXiv.1901.07555>
- Agarwal, P., Sreepada, R. S., & Patra, B. K. (2019). A hybrid framework for improving diversity and long tail items in recommendations. In B. Deka, P. Maji, S. Mitra, D. Bhattacharyya, P. Bora, & S. Pal (Eds.), *Lecture Notes in Computer Science: Vol 11942. PReMI 2019: Pattern Recognition and Machine Intelligence* (pp. 285–293) Springer. http://doi.org/10.1007/978-3-030-34872-4_32
- Akhlaq, A., Sheikh, A., & Pagliari, C. (2016). Defining health information exchange: Scoping review of published definitions. *BMJ Health & Care Informatics*, 23(4), 684–764. <https://doi.org/10.14236/jhi.v23i4.838>
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, 395(10228), 931–934. [https://doi.org/10.1016/S0140-6736\(20\)30567-5](https://doi.org/10.1016/S0140-6736(20)30567-5)
- Andreotti, P. E., Ludwig, G. V., Peruski, A. H., Tuite, J. J., Morse, S. S., & Peruski Jr, L. F. (2003). Immunoassay of infectious agents. *Biotechniques*, 35(4), 850–859. <https://doi.org/10.2144/03354ss02>
- Arnold, R. D., & Wade, J. P. (2015). A definition of systems thinking: A systems approach. *Procedia Computer Science*, 44, 669–678. <https://doi.org/10.1016/j.procs.2015.03.050>
- Battegay, M., Kuehl, R., Tschudin-Sutter, S., Hirsch, H. H., Widmer, A. F., & Neher, R. A. (2020). 2019-novel coronavirus (2019-nCoV): Estimating the case fatality rate—A word of caution. *Swiss Medical Weekly*, 150,

Article w20203. <https://doi.org/10.4414/smw.2020.20203>

Ben-Zvi, D., & Garfield, J. B. (Eds.). (2004). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3–16). Dordrecht, The Netherlands: Kluwer academic publishers.

Berinato, S. (2019). Data science and the art of persuasion. *Harvard Business Review*, (January–February 2019), 126–137. <https://hbr.org/2019/01/data-science-and-the-art-of-persuasion>

Binet, A., Gavin, V., Carroll, L., & Arcaya, M. (2019). Designing and facilitating collaborative research design and data analysis workshops: Lessons learned in the healthy neighborhoods study. *International Journal of Environmental Research and Public Health*, 16(3), 324. <https://doi.org/10.3390/ijerph16030324>

Binnicker, M. J. (2020). Emergence of a novel coronavirus disease (COVID-19) and the importance of diagnostic testing: Why partnership between clinical laboratories, public health agencies, and industry is essential to control the outbreak. *Clinical Chemistry*, 66(5), 664–666. <https://doi.org/10.1093/clinchem/hvaa071>

Borghouts, J., Gordon, A. D., Sarkar, A., O’Hara, K. P., & Toronto, N. (2019). Somewhere around that number: An interview study of how spreadsheet users manage uncertainty. *arXiv*. <https://doi.org/10.48550/arXiv.1905.13072>

Boukhelifa, N., Perrin, M.-E., Huron, S., & Eagan, J. (2017). How data workers cope with uncertainty: A task characterisation study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3645–3656). <https://doi.org/10.1145/3025453.3025738>

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick*. Harvard University Press.

Brulle, R. J., & Pellow, D. N. (2006). Environmental justice: Human health and environmental inequalities. *Annual Review of Public Health*, 27(1), 103–124. <https://doi.org/10.1146/annurev.publhealth.27.021405.102124>

Budescu, D. V., Broomell, S., & Por, H.-H. (2009). Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological Science*, 20(3), 299–308. <https://doi.org/10.1111/j.1467-9280.2009.02284.x>

Burkert, A., & Loeb, A. (2020, March 17). Flattening the COVID-19 curves. *Scientific American*. Retrieved from <https://blogs.scientificamerican.com/observations/flattening-the-covid-19-curves/>

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet*, 395(10223), 507–513. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Rossi, L., Sun, K., Viboud, C., Xiong X., Yu, H., Halloran, M. E., Longini Jr., I. M., & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), 395–400. <https://doi.org/10.1126/science.aba9757>

Cicalò, E., & Valentino, M. (2019). Mapping and visualisation on of health data. the contribution on of the graphic sciences to medical research from New York yellow fever to China coronavirus. *DISEGNARECON*, 12(23), 12–1. <https://doi.org/10.20365/disegnarecon.23.2019.12>

Costanza-Chock, S. (2018). Key lessons from participatory communications research with the immigrant rights, occupy, and LGBTQ and two-spirit movements In G. Meikle (Ed.), *The Routledge Companion to Media and Activism*. Routledge (ch. 7). Taylor & Francis.

Davis, J. A., & Burgoon, L. D. (2015). Can data science inform environmental justice and community risk screening for Type 2 diabetes? *PLOS One*, 10(4), Article e0121855. <https://doi.org/10.1371/journal.pone.0121855>

Diakopoulos, N., Friedler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., Jagadish, H.V., Unsworth, K., Venkatasubramanian, S., Wilson, C., Yu, C., & Zevenbergen, B. (2017). *Principles for accountable algorithms and a social impact statement for algorithms*. FAT/ML. <https://www.fatml.org/resources/principles-for-accountable-algorithms>

Donahue, A., & Tuohy, R. (2006). Lessons we don’t learn: A study of the lessons of disasters, why we repeat them, and how we can learn them. *Homeland Security Affairs*, 2(2), Article 4. <https://www.hsaj.org/articles/167>

Duran, R. E. (2008). Probabilistic sales forecasting for small and medium-size business operations. In B. Prasad (Ed.), *Studies in Fuzziness and Soft Computing: Vol 230. Soft computing applications in business* (pp. 129–146). Springer. https://doi.org/10.1007/978-3-540-79005-1_8

Espinoza, O. (2007). Solving the equity–equality conceptual dilemma: A new model for analysis of the educational process. *Educational Research*, 49(4), 343–363. <https://doi.org/10.1080/00131880701717198>

Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L., Van Elsland, S., Thompson, H., ... Ghani, A. (2020). *Impact of non-pharmaceutical interventions (NPIS) to reduce COVID-19 mortality and healthcare demand (Tech. Rep. No. 9)*. MRC Centre, Imperial College London. Retrieved from <https://spiral.imperial.ac.uk/handle/10044/1/77482>

Fisher, D., & Heymann, D. (2020). Q&A: The novel coronavirus outbreak causing COVID-19. *BMC medicine*, 18(1), Article 57. <https://doi.org/10.1186/s12916-020-01533-w>

- Foucault Welles, B., & Meirelles, I. (2015). Visualizing computational social science: The multiple lives of a complex image. *Science Communication*, 37(1), 34–58. <https://doi.org/10.1177/1075547014556540>
- Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In Gal, I., & Garfield, J. B. (Eds.), *The assessment challenge in statistics education* (pp. 1–13). IOS Press.
- Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In M.F. Balcan & K.Q. Weinberger (Eds.), *Proceedings of The 33rd International Conference on Machine Learning: Vol 48* (pp. 1050–1059).
- Gelman, A., & Yao, Y. (2020). Holes in Bayesian statistics. *arXiv*. <https://doi.org/10.48550/arXiv.2002.06467>
- Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). “A 30% chance of rain tomorrow”: How does the public understand probabilistic weather forecasts? *Risk Analysis: An International Journal*, 25(3), 623–629. <https://doi.org/10.1111/j.1539-6924.2005.00608.x>
- Green, B. (2018). Data science as political action: grounding data science in a politics of justice. *arXiv*. <https://doi.org/10.48550/arXiv.1811.03435>
- Hammond, J. L. (2015). The anarchism of occupy wall street. *Science and Society*, 79(2), 288–313. <https://doi.org/10.1521/SISO.2015.79.2.288>
- Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, 290(5500), 2261–2262. <https://doi.org/10.1126/science.290.5500.2261>
- Hullman, J. (2020, Jan). Why authors don’t visualize uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 130–139. <https://doi.org/10.1109/TVCG.2019.2934287>
- JHU CSSE. (2020, 03). *COVID-19 global case tracker*. John Hopkins Center for Systems Science and Engineering. John Hopkins Center for Systems Science and Engineering. <https://coronavirus.jhu.edu/map.html>
- Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition*, 50(1-3), 189–209. [https://doi.org/10.1016/0010-0277\(94\)90028-0](https://doi.org/10.1016/0010-0277(94)90028-0)
- Kezar, A. J., & Holcombe, E. M. (2017). *Shared leadership in higher education*. Washington, DC: American Council on Education.
- Klarkowski, D., O’Brien, D. P., Shanks, L., & Singh, K. P. (2014). Causes of false-positive HIV rapid diagnostic test results. *Expert Review of Anti-infective Therapy*, 12(1), 49–62. <https://doi.org/10.1586/14787210.2014.866516>

Kobayashi, T., Jung, S.-m., Linton, N. M., Kinoshita, R., Hayashi, K., Miyama, T., Anzai, A., Yang, Y., Yuan, B., Akhmetzhanov, A. R., Suzuki, A., & Nishiura, H. (2020). Communicating the risk of death from novel coronavirus disease (COVID-19). *Journal of Clinical Medicine*, 9(2), 580. <http://doi.org/10.3390/jcm9020580>

Lan, L., Xu, D., Ye, G., Xia, C., Wang, S., Li, Y., & Xu, H. (2020). Positive RT-PCR test results in patients recovered from COVID-19. *JAMA*, 323(15), 1502–1503. <https://doi.org/10.1001/jama.2020.2783>

Lee, C., & Meletiou, M. (2003). Some difficulties of learning histograms in introductory statistics. In *2003 Proceedings of the American Statistical Association, Statistics Education Section* (pp. 2326–2333). Alexandria, VA: American Statistical Association

Leonelli, S. (2016). Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), Article 20160122. <https://doi.org/10.1098/rsta.2016.0122>

Lin, C.-Y., et al. (2020). Social reaction toward the 2019 novel coronavirus (COVID-19). *Social Health and Behavior*, 3(1), 1–2. http://doi.org/10.4103/SHB.SHB_11_20

Lockie, S. (2018). Privilege and responsibility in environmental justice research. *Environmental Sociology*, 4(2), 175–180. <https://doi.org/10.1080/23251042.2018.1460936>

Lu, X., Zhang, L., Du, H., Zhang, J., Li, Y. Y., Qu, J., Zhang, W., Wang, Y., Bao, S., Li, Y., Wu, C., Liu, H., Liu, D., Shao, J., Peng, X., Yang, Y., Liu, Z., Xiang, Y., Zhang, F., Silva, R.M., ... Wong, G.W.K. (2020). SARS-CoV-2 infection in children. *New England Journal of Medicine*, 382(17), 1663–1665. <https://doi.org/10.1056/NEJMc2005073>

Mah, A. (2017). Environmental justice in the age of big data: Challenging toxic blind spots of voice, speed, and expertise. *Environmental Sociology*, 3(2), 122–133. <https://doi.org/10.1080/23251042.2016.1220849>

Malone, K. (2020, 1 31). When translation problems arise between data scientists and business stakeholders, revisit your metrics. *Harvard Data Science Review*, 2(1). <https://doi.org/10.1162/99608f92.c2fc310d>

Meng, X.-L. (2020, March 31). XL-Files: COVID coping and the law of most people. *IMS Bulletin*, 49(3), 6. Retrieved from <https://imstat.org/2020/03/31/xl-files-covid-coping-and-the-law-of-most-people/>

Mikolajczyk, R., Krumkamp, R., Bornemann, R., Ahmad, A., Schwehm, M., & Duerr, H.-P. (2009). Influenza —Insights from mathematical modelling. *Deutsches Ärzteblatt International*, 106(47), 777–782. <https://doi.org/10.3238/arztebl.2009.0777>

Minkler, M., Vásquez, V. B., Tajik, M., & Petersen, D. (2008). Promoting environmental justice through community-based participatory research: The role of community and partnership capacity. *Health Education &*

Behavior, 35(1), 119–137. <https://doi.org/10.1177/1090198106287692>

Morton, T. A., Rabinovich, A., Marshall, D., & Bretschneider, P. (2011). The future that may (or may not) come: How framing changes responses to uncertainty in climate change communications. *Global Environmental Change*, 21(1), 103–109. <https://doi.org/10.1016/j.gloenvcha.2010.09.013>

Mousavi, S., & Gigerenzer, G. (2014). Risk, uncertainty, and heuristics. *Journal of Business Research*, 67(8), 1671–1678. <https://doi.org/10.1016/j.jbusres.2014.02.013>

Nick, G. A., Savoia, E., Elqura, L., Crowther, M. S., Cohen, B., Leary, M., Wright, T., Auerbach, J., & Koh, H. K. (2009). Emergency preparedness for vulnerable populations: people with special health-care needs. *Public Health Reports*, 124(2), 338–343. <https://doi.org/10.1177/003335490912400225>

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, 2, 13. <https://doi.org/10.3389/fdata.2019.00013>

Oxbury, W. (2018). Does data science need statistics? In N. Adams and E. Cohen (Eds.), *Statistical Data Science* (pp. 1–19). https://doi.org/10.1142/9781786345400_0001

Palmer, T. N. (2000). Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics*, 63(2), 71–116. <https://doi.org/10.1088/0034-4885/63/2/201>

Pappenberger, F., & Beven, K. J. (2006). Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Water Resources Research*, 42(5), Article W05302. <http://doi.org/10.1029/2005WR004820>

Peters, S. A. (2011). Robust understanding of statistical variation. *Statistics Education Research Journal*, 10(1), 52–88. <https://psycnet.apa.org/doi/10.52041/serj.v10i1.367>

Pew Research Center. (2009, July 9). Public praises science; scientists fault public, media. *Pew Research Center for the People & the Press*. <https://www.pewresearch.org/politics/2009/07/09/public-praises-science-scientists-fault-public-media/>

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17–46). Springer. https://doi.org/10.1007/1-4020-2278-6_2

Quade, D., Lachenbruch, P. A., Whaley, F. S., McClish, D. K., & Haley, R. W. (1980, 05). Effects of misclassifications on statistical inferences in epidemiology. *American Journal of Epidemiology*, 111(5), 503–515. <https://doi.org/10.1093/oxfordjournals.aje.a112930>

Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal*, 5(2), 46–68. <https://iase->

[web.org/documents/SERJ/SERJ5\(2\)_Reading_Reid.pdf](http://web.org/documents/SERJ/SERJ5(2)_Reading_Reid.pdf)

Reddick, C. G. (2004). Assessing local government revenue forecasting techniques. *International Journal of Public Administration*, 27(8–9), 597–613. <https://doi.org/10.1081/PAD-120030257>

Revere, D., Dixon, B. E., Hills, R., Williams, J. L., & Grannis, S. J. (2014). Leveraging health information exchange to improve population health reporting processes: Lessons in using a collaborative-participatory design process. *EGEMS*, 2(3), 1082. <https://doi.org/10.13063/2327-9214.1082>

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In Ricci F., Rokach L., Shapira B. (Eds.), *Recommender systems handbook* (pp. 1–35). Springer. http://doi.org/10.1007/978-0-387-85820-3_1

Riou, J., Hauser, A., Counotte, M. J., & Althaus, C. L. (2020). Adjusted age-specific case fatality ratio during the COVID-19 epidemic in Hubei, China, January and February 2020. *medRxiv*. <https://doi.org/10.1101/2020.03.04.20031104>

Sanders, N. E. (2019a). AMEND: Open source and data-driven oversight of water quality in New England. *Media and Communication*, 7(3), 91–103. <http://doi.org/10.17645/mac.v7i3.2136>

Sanders, N. E. (2019b). A balanced perspective on prediction and inference for data science in industry. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.5e126552>

Santos, R. L., Macdonald, C., & Ounis, I. (2010). Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 881–890). <https://doi.org/10.1145/1772690.1772780>

Savage, S. L. (2002). The flaw of averages. *Harvard Business Review*, 80(11), 20–21. <https://hbr.org/2002/11/the-flaw-of-averages>

Savage, S. L. (2012). *The flaw of averages: Why we underestimate risk in the face of uncertainty*. John Wiley & Sons.

Schlosberg, D. (2003). The justice of environmental justice: Reconciling equity, recognition, and participation in a political movement. In A. Light and A. De-Shalit (Eds.), *Moral and Political Reasoning in Environmental Practice* (pp. 125–156). MIT Press.

Secko, D. M., Amend, E., & Friday, T. (2013). Four models of science journalism: A synthesis and practical assessment. *Journalism Practice*, 7(1), 62–80. <https://doi.org/10.1080/17512786.2012.691351>

Shepard, P. (2002). Advancing environmental justice through community-based participatory research. *Environmental Health Perspectives*, 110(Suppl. 2), 139.

- Snee, R. D. (1999). Discussion: Development and use of statistical thinking: A new era. *International Statistical Review/Revue Internationale de Statistique*, 67(3), 255–258 <https://doi.org/10.2307/1403703>
- Sohrabi, C., Alsafi, Z., O’Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, R. (2020). World health organization declares global emergency: A review of the 2019 novel coronavirus (Covid-19). *International Journal of Surgery*, 76, 71–76. <https://doi.org/10.1016/j.ijsu.2020.02.034>
- Spears Johnson, C. R., Kraemer Diaz, A. E., & Arcury, T. A. (2016). Participation levels in 25 community-based participatory research projects. *Health Education Research*, 31(5), 577–586. <https://doi.org/10.1093/her/cyw033>
- Spiegelhalter, D. (2017). Risk and uncertainty communication. *Annual Review of Statistics and Its Application*, 4(1), 31–60. <https://doi.org/10.1146/annurev-statistics-010814-020148>
- Stern, N. (2014). Ethics, equity and the economics of climate change paper 1: Science and philosophy. *Economics & Philosophy*, 30(3), 397–444. <http://doi.org/10.1017/S0266267114000297>
- Stevens, H. (2020, March 14). Why outbreaks like coronavirus spread exponentially, and how to “flatten the curve.” *The Washington Post*. <https://www.washingtonpost.com/graphics/2020/world/corona-simulator/>
- Tang, B., Bragazzi, N. L., Li, Q., Tang, S., Xiao, Y., & Wu, J. (2020). An updated estimation of the risk of transmission of the novel coronavirus (2019-ncov). *Infectious Disease Modelling*, 5, 248–255. <https://doi.org/10.1016/j.idm.2020.02.001>
- Taylor, K.-Y. (2016). *From #blacklivesmatter to black liberation*. Haymarket Books.
- Tran, D., Dusenberry, M. W., van der Wilk, M., & Hafner, D. (2018). Bayesian layers: A module for neural network uncertainty. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alche-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (pp. 14633–14645). Curran Associates, Inc.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Valcarce, D., Parapar, J., & Barreiro, Á. (2016). Item-based relevance modelling of recommendations for getting rid of long tail products. *Knowledge-Based Systems*, 103, 41–51. <https://doi.org/10.1016/j.knosys.2016.03.021>
- Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P., Fu, H., Dighe, A., Griffin, J., Cori, A., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba, Z. M., Fitzjohn, R., Gaythorpe, K. A. M., . . . Ferguson, N. (2020). Estimates of the severity of covid-19 disease. *The Lancet*, 20(6), 669–677. [https://doi.org/10.1016/s1473-3099\(20\)30243-7](https://doi.org/10.1016/s1473-3099(20)30243-7)

- Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223), 470–473. [https://doi.org/10.1016/s0140-6736\(20\)30185-9](https://doi.org/10.1016/s0140-6736(20)30185-9)
- Wang, W., Tang, J., & Wei, F. (2020). Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *Journal of Medical Virology*, 92(4), 441–447. <https://doi.org/10.1002/jmv.25689>
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–248. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>
- Wilder-Smith, A., & Freedman, D. O. (2020, 02). Isolation, quarantine, social distancing and community containment: Pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *Journal of Travel Medicine*, 27(2), Article taaa020. <https://doi.org/10.1093/jtm/taaa020>
- Wilson, A. G., & Izmailov, P. (2020). Bayesian deep learning and a probabilistic perspective of generalization. *arXiv*. <https://doi.org/10.48550/arXiv.2002.08791>
- Woelfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Mueller, M. A., . . . Wendtner, C. (2020). Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster. *medRxiv*. <https://doi.org/10.1101/2020.03.05.20030502>
- World Health Organization. (2020). *Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: Interim guidance*, 2 March 2020. <https://apps.who.int/iris/handle/10665/331329>
- Wu, Z., & McGoogan, J. M. (2020, 02). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) Outbreak in China: Summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*, 323(13), 1239–1242. <https://doi.org/10.1001/jama.2020.2648>
- Xie, M.-g., & Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review*, 81(1), 3–39. <https://doi.org/10.1111/insr.12000>
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, 52(1), 1–38. <https://doi.org/10.1145/3285029>
- Zhang, W., Du, R.-H., Li, B., Zheng, X.-S., Yang, X.-L., Hu, B., Wang, Y., Xiao, G., Yan, B., Shi, Z., & Zhou, P. (2020). Molecular and serological investigation of 2019-ncov infected patients: Implication of multiple shedding routes. *Emerging Microbes & Infections*, 9(1), 386–389. <https://doi.org/10.1080/22221751.2020.1729071>

Footnotes

1.

It must be noted that the references in this section were compiled on and before March 25, 2020. With the situation evolving rapidly, there is a risk that some will quickly become outdated.

[1](#)