

Climate Change's Big Impact on Big Data

The problem of climate change is not the amount of data to manage, the technology trying to tame that data, or even the people who use the technology. The problem has to do with science, or more acutely, certain populations' unwillingness to believe in science, that poses the greatest challenge to big data. According to the most recent poll from the Pew Research Center, only 40% of Americans believe that global warming is primarily caused by human activities that pump excessive amounts of CO2 into the atmosphere. If you belong to this minority, keep reading. And if you do not belong to this minority, definitely keep reading. Those who understand the scientific community have a sense of how elusive consensus among scientists can be. By its very nature, science is based on debunking science with better science. Thus, to have a consensus among the overwhelming majority of the world's scientists is extremely significant. To be sure, there is still ample room for debate. That debate, however, is more focused on the magnitude of climate change's impact than its origins.

For the purpose of this article, I will give corporate America the benefit of the doubt by assuming that it is more scientifically inclined than the population at large (allotting for variations based on industry, geography, and other factors). But this paper is not trying to accurately quantify the scientific inclination of businesses. Instead, it is about the alarming proportion of science naysayers (both active and passive) who undermine data-driven, actionable insights. Based on my own experiences with a variety of businesses, I estimate that about half the people in corporate America fall into this category.

If correct, then every one in two businesspeople in America disregards data regardless of how reasonable or compelling its conclusions. After all, these people trust their gut more than any scientific consensus. Sadly, reason and empirically-derived conclusions are often dismissed even before any analysis is considered. If such results are eventually given consideration, it feels as if any subsequent

validation or further analysis is asking too much. These trends reflect a different enemy of science: the irrepressible urge for short-term results. The data-driven professional understands that the initial pass at the data may have been incomplete. Hypotheses must be revised. Assumptions must be revisited. New data sources must be considered. Qualitative opinions must be factored in. In other words, science is about continuous improvement in the search for answers. When an answer is found, science immediately attempts to debunk it with whatever methods make sense. If the debunking effort fails, the answer will remain as the best answer, at least for the moment. The data scientist will not dismiss an answer just because it is a non-quantifiable, emotional response. Science will accept an answer as the best answer only if alternative means to disprove it have been extensively explored. In this sense, science is a non-judgmental “show-me-what-you-got” discipline.

Indeed, other non-quantifiable attributes should be considered in lieu of or in addition to the data and methodologies used to harness it. In fact, any conclusion should be treated with skepticism, be it data-driven or not. Such is the way of science. However, we must discourage knee-jerk rejections of the empirical approach merely upon the basis that it doesn't feel right. A blanket-policy of skepticism towards testing, profiling, modeling, and other forms of data analysis is not helpful.

Data can actually be used to validate hunches and predict events using empiric methodologies. In the end, the data-driven approach employs the tried-and-true scientific method to attempt to augment, if not supersede, what we think is true. Inevitably, some results will seem counterintuitive to conventional corporate wisdom. This is the fun part for data scientists (that is, unless political motivations lurk behind the scenes). It seems as if the resistance to big data in corporate America is what holds its practitioners

back from being fully integrated into mainstream corporate functions, a sad state of affairs given that big data has a lot to offer.

If my 50-50 ratio of science supporters to science naysayers is even remotely accurate, that's a big problem. We can prove whether a particular idea worked out using a well-designed statistical test. Any arguments against taking the actions indicated by the results should be based on the flaws of the data itself or the methods used to analyze it, not based on an individual's intuition. Succinctly, challenge science with better science. In some cases, there may be a mismatch between the data and what is being measured. The science naysayer may argue that some human behavior is just not conducive to measurement. Even so, there should still be dialogue about how to improve the data as well as analytical practices, not a blatant rejection of their merit. Engaging in that dialogue depends on valuing a data-driven, scientific approach. So long as there are people who continue to resist the broad scientific consensus about climate change, you can be sure that data scientists will remain undervalued members of corporate America.