Why Shouldn't We Let Data Speak for Themselves?*

A Brief Discisson about the Neutrality of Data

Rae Zhang

19 March 2023

*Code and data are available at: https://github.com/JunweiZhang130/Why-we-should-not-let-the-data-speak-for-themselves-.git

1 Tutorial Topic

To what extent do you think we should let the data speak for themselves? Please write at least two pages.

2 Discussion

When there is a clear research question and transparent methods for collecting and analyzing data, data can be a powerful tool for telling stories and communicating information. In his article "Artificial Intelligence - The Revolution Hasn't Happened Yet,"(Jordan 2019) Jordan emphasizes the significance of interdisciplinary approaches to data analysis in the AI industry. While computational and statistical disciplines have traditionally been at the forefront of data analysis, social sciences, cognitive sciences, and humanities have much to offer in this field.

Jordan's argument underscores the limitations of relying solely on data-driven narratives, as it risks oversimplifying or misrepresenting complex issues. Data can be a powerful tool for communicating information, but critical thinking and social analysis are necessary to fully understand the data and its implications for society. In essence, data analysis must go beyond numbers to appreciate the human aspect of the analysis.

A prime example of the dangers of relying solely on data to inform decision-making can be found in the education sector. Aparicio's paper, "Measuring efficiency in education: The influence of imprecision and variability in data on DEA estimates,"(Aparicio, Cordero, and Ortiz 2019) highlights how different schools collect data that can cause a different interpretation in the final data results. While summarizing data from the original observations (students), this approach neglects the existing dispersion of data, which may become a serious problem if variability across schools is high. External factors may also impact the final data results, which may result in an inaccurate picture of the schools' performance. Failure to take these factors into account may lead to an unfair labelling of certain schools as "failing," which can hinder opportunities to provide resources and support to those communities.

One of the limitations of data analysis is that it only reveals correlations between variables and does not necessarily capture the underlying reasons for these correlations. To understand the deeper meaning behind data, it is necessary to incorporate social scientific and humanistic perspectives into data analysis. By doing so, researchers can move beyond simple correlations to gain a more nuanced understanding of the data and its implications.

Furthermore, data-driven narratives can sometimes overlook the "human part" of the analysis. It is essential to engage with people who are impacted by the data and to consider their perspectives in any analysis. The inclusion of social scientific and humanistic perspectives can help researchers to better understand how data affects individuals and communities.

Data about depression can provide important information about patterns and trends, such as demographic groups that are more likely to be affected or changes in prevalence rates over time. However, this data cannot fully capture the lived experiences of individuals who suffer from depression or the cultural and societal factors that contribute to its prevalence.

In conclusion, Jordan's argument underscores the importance of adopting a holistic approach to data analysis that incorporates social scientific and humanistic perspectives. While data can be a valuable tool for generating insights and ideas, it should not be the only source of inspiration or guidance. Data-driven narratives can be misleading, and it is necessary to engage in critical thinking and social analysis to fully understand the complexities of the data and its implications for society. By adopting a holistic approach, researchers can gain a more nuanced understanding of the data, its limitations, and its possibilities.

References

Aparicio, Juan, Jose M. Cordero, and Lidia Ortiz. 2019. "Measuring Efficiency in Education: The Influence of Imprecision and Variability in Data on DEA Estimates." *Socio-Economic Planning Sciences* 68: 100698. https://doi.org/https://doi.org/10.1016/j.seps.2019.03.004.

Jordan, Michael I. 2019. "Artificial Intelligence—The Revolution Hasn't Happened Yet." *Harvard Data Science Review* 1 (1).