

Bad Bar Graphs

11 DEC 2024 · 1:49 PM ET · BY [DEREK LOWE](#) · 2 MIN READ · [COMMENTS](#)



I hadn't seen [this preprint](#) until [this Nature piece](#) called my attention to it, but it's certainly worth a look. The authors have gone back over the recent biology literature (3387 articles from 2023 in 15 high-profile journals) and they find that 88% of these use some sort of bar graph to present data. And I can well believe it! But unfortunately, about a third of those have presentation mistakes in them.

The two most common mistakes are (1) not starting the Y axis at zero and (2) log-scale problems. In general, the first of those will tend to make differences between two data sets look larger than they really are (because you end up "zooming in" on the top of the actual bars), and the second will tend to make them look smaller (since you need much more of a change to show up on a log-scale plot). The authors found that the log-scale mistakes cause more distortion than the axis-zeroing ones, and I can believe that, too, since presenting data in logarithmic space can get out of hand pretty fast.

For what it's worth, the journals with the most visualization problems were *Nature Biomedical Engineering*, *Cell*, *Science Translational*

Medicine, and *ACS Nano*. That last journal seemed to make mostly axis-zeroing mistakes (*Nature Plants* and *Nature Neuroscience* had a similar pattern), while the other leaders were more evenly divided, except for *Nature Cancer*, which stood out as the only one of the 15 journals that had a solid preference for log-scale mistakes. I don't know if we can draw many conclusions from all that, since this is only one time period's worth of data (i.e., perhaps other years might have other winners - well, losers).

The authors provide a useful flow chart of advice in their Figure 2C about using bar graphs (and when to avoid them). There are situations where they're the appropriate way to show your data, of course, but not when you're reaching for things like a non-zeroed y axis. Their Extended Data Figure 2 presents the same example data set presented in a number of different formats, most of which are distortions and should be avoided. It's not that bar graphs are inherently misleading or that log axes are, either: but either one of them can be misleading when misapplied, and it's very easy to misapply them even if you're not actually trying to hide anything. Sticking with these principles would improve the literature, but I'm world-weary enough to think that if the authors revisit the same journals in (say) three years they will find the same problems occurring at similar rates. I hope I'm wrong!

ABOUT THE AUTHOR



[Derek Lowe](#)
