Industrial and Organizational Psychology: Perspectives on Science and Practice
2021, Vol. 14, No. 4, 510-513

Not the version of record. Final article available at:
https://doi.org/10.1017/iop.2021.115

Spreading the word: Equipping I-O students to use descriptive statistics for effective data visualization

Afra S. Ahmad* and Steven Zhou*
George Mason University
* equal authorship listed by alphabetical order

The focal article by Murphy (2021) succinctly articulates the impressive advancements and improvements made in statistical methods over the past few decades, with the application of these complex tools appearing in many top-tier I-O psychology journals. No doubt, these methods are incredibly powerful and important to answer the complicated questions posed by researchers. However, Murphy (2021) makes the argument that, in favor of these more complicated methods, researchers have neglected the importance of the "simpler" descriptive statistics (i.e., "Table 1") and their impact on interpretation and application of findings to applied practice. We concur, and in the following pages, we focus on the role that statistical graduate training has in highlighting the importance of descriptive statistics for applied data science practitioners and data visualization.

Drawing from our experience designing and instructing an online applied Master's program in I-O psychology, we first highlight the types of jobs that I-O practitioners are engaging in today and the statistical skills needed for such jobs (spoiler: mostly descriptive statistics). Then, we focus on the rapidly growing science of data visualization and the statistical skills necessary for effective data visualization (another spoiler: mostly descriptive statistics). In doing so, we hope to reinforce Murphy's (2021) suggestion that scholars, especially I-O graduate students, should not neglect the importance of Table 1 even as the more advanced methodologies are taught and mastered.

Where we are: The current state of applied I-O statistical training

More than 100 Industrial-Organizational Psychology graduate programs in the United States offer either a Master's, Doctorate, or both. I-O graduates are employed across four employment sectors (academia, consulting, industry, and government) with a majority going applied (Zelin et al., 2015). In recent years, there has also been growth in applied Master's programs, sometimes called Masters of Professional Studies (MPS). Employers continue to recognize the benefits of leveraging and incorporating I-O knowledge and competencies to a variety of jobs. Consequently, graduate programs are preparing students for a wide range of jobs including traditional I-O positions (e.g., consulting, applied research) and those in the HR and management.

As graduate I-O programs produce students who are increasingly entering the applied workforce, it is important to evaluate the specific competencies that employers desire. In developing the curriculum plan for our own MPS I-O program, we reviewed KSAOs from numerous job advertisements that were recruiting Masters-level I-O practitioners. The most common KSAOs included "apply expertise in people research, quantitative analysis, data science and data visualization to provide insights on talent management and leadership development initiatives," and "track record in interpreting data creatively and delivering impactful insights (e.g., going beyond the 'what' of a research inquiry – into the 'so what' and 'now what'… and 'what haven't we thought of yet?'" In alignment with these job ads, SIOP's Professional Practice Committee conducted a "Career Study" and found that *oral communication* is a top competency for I-O psychologists across employment sectors (Zelin et al., 2015). This evidently growing interest in the *communication* of data and results leads to the question, does our statistical graduate training adequately prepare students for the applied workforce?

The Guidelines for Education & Training (SIOP, 2016) recommends covering both descriptive and inferential statistical methods, spanning both parametric and nonparametric statistical methods and covering both quantitative and qualitative research methodology.

Correspondence concerning this paper should be addressed to Steven Zhou, Department of Psychology, George Mason University, Fairfax, VA 22030. Email: szhou9@gmu.edu

However, a cursory review of several syllabi from various I-O graduate programs reveal that there is much more emphasis on advanced statistical courses focused on, for example, structural equation modeling, multilevel analysis, and longitudinal analysis. This anecdotally reflects Murphy's (2021) findings that the focus of statistical analyses in I-O, both in research and in education, tends to be towards these "advanced" methods. In contrast, relatively little time is spent on understanding, interpreting, and using descriptive statistics, reflected both in research (Murphy, 2021) and in education (our review of syllabi).

Importantly, this latter skill is a critical KSAO necessary to prepare the growing number of students entering applied practice. We surveyed about 100 Masters and PhD I-O graduates working in applied practice, asking them what software and analytical techniques they used. The most popular software was Tableau for data visualization (81 used it regularly in their jobs), followed by Excel (79) and then R (48). The most popular analyses used were correlation (62 answered "5" on a five-point scale from "not at all" to "a lot"), data visualization (55), regression (49), and t-tests (41). Remarkably, the majority of practitioners only used advanced techniques (e.g., SEM, longitudinal, multivariate) "not at all" or "a little" in their daily jobs. Clearly, the industry is calling for data visualization as a critical skill, potentially even more important than advanced analysis methods. In the following section, we dive deeper into what we argue is missing in I-O graduate training: using descriptive statistics for data visualization.

Where we need to go: The importance of data visualization

The "academic-practitioner" divide is well-known, with dozens of articles, commentaries, and scholarly discussions in recent years bemoaning the lack of translation of the "science of I-O" to actual workplace settings and calling for movements to change this (e.g., Anderson et al., 2001; Aguinis et al., 2017; Latham, 2019). The most recent articulation of SIOP's strategic goal acknowledges this, emphasizing the importance of "translat[ing] scientific knowledge to promote individual and organizational health and effectiveness" (SIOP, 2021). Put simply, the complex and advanced statistical models used in most top I-O journals are valuable, but unless the results are effectively summarized and communicated succinctly to an applied (and largely non-statistical) audience, such research is likely to have little impact. In order for SIOP to grow as an organization and our field to build influence in the workplace, effective and concise communication of data-driven results must be a preeminent skill for all I-O psychologists. In most situations with an applied audience, communication of results does not take the form of blocks of dense text or tables, but rather of high-quality visual graphs and charts, i.e., data visualization.

Data visualization is built almost entirely from descriptive statistics. Granted, this is an over-simplification, as there are certainly ways to incorporate inferential statistical models into visuals. But even then, these are often not much more than regression lines or error bars to depict standard errors (see Healy, 2018). It would be tough to visualize a latent state-trait model with multiple indicators and autoregression across four measurement occasions in a simple, succinct, and "beautiful" way. Instead, the best visuals, in terms of effectively communicating and persuading the audience, focus on not much more than means (e.g., bar charts), variances (e.g., scatterplots), and groupings (e.g., social networks). Just as Murphy (2021) demonstrated in the focal article, descriptive statistics can and often should be used to communicate the results of more advanced statistical methods. In a modern world driven by highly engaging, interactive, and easy-to-understand charts and graphs, this becomes even more important.

Data visualization also exemplifies the rule of "less is more" (Berinato, 2016; Knaflic, 2015). Such a sentiment often goes in direct contrast to the approach taken in most advanced statistical methods. Take for example the classic problem of endogeneity in most cross-sectional and panel studies. The solution? Researchers must add the appropriate control variables, and they must use a variety of statistical methods to correct for and control endogeneity bias (e.g., Zaefarian et al., 2017; Antonakis et al., 2021). To be clear, this is not to say that endogeneity is not important; it is, and these recommended corrections are highly valuable. However, we bring this up as an example of how data visualization tries to keep things simple, but the advanced methods found in research studies require the use of more complicated techniques to get the right answer. This tension is not often recognized or valued, which is why we argue for more attention to be paid to when and why we should try to keep things simple.

To clarify, we are not saying that data visualization is easy; it is incredibly difficult to do well. For example, Sawicki (2020) articulates a 10-point "grammar of data visualization" that should generally be followed when creating data visuals. Moreover, bad visualizations can be found everywhere. At best, they are not effective in communicating what they are trying to communicate. At worst, they are misleading and can perpetuate misinformation among the public; the recent trend of misleading data visuals related to COVID-19 is a prime example (Leybzon, 2020). In some ways, poorly conducted data visualization (and by extension, poorly used descriptive statistics) can have even more widespread negative effects, due to public accessibility and audience size, than a poorly conducted SEM. As such, if we as I-O psychologists hope to pursue SIOP's goals of bridging the academic-practitioner gap, much more focus needs to be put on the importance of descriptive statistics and data visualization.

Conclusion

We conclude by offering a few key action-oriented takeaways for readers to consider. First, educators should make space for descriptive statistics and data visualization in their graduate statistics courses. As a start, we should review syllabi and incorporate or expand these topics to provide adequate training in these areas. Second, students must be open to learning *both* descriptive statistics and advanced statistics, with the goal of understanding when and why each should be used. The ability to conduct advanced and complex statistics is incredibly important to answer complicated research questions, but unless students are able to subsequently communicate their results and their data to a non-statistical audience, there will be limited ability for the research to make an impact on actual business practice. Moreover, our brief survey of I-O graduates suggests that data visualization skills may be even more valuable for getting a job than advanced analytical methods. Finally, as a field, we must try and inform the public on how to interpret data visualizations. Unfortunately, misleading data visualizations (e.g., base rate fallacy, omitted variables, scaling) are rampant in popular culture. As scientists, we should be aiming to improve public understanding and awareness of how to use data visualization and when it can go wrong. One way this can be accomplished is to continue to broaden inclusivity in our field and cater to students from all types of educational and career backgrounds (see Zhou & Ahmad, 2020) so that they can leverage I-O knowledge in their respective workplaces. In short, Murphy (2021) is correct to point out the underappreciated nature of descriptive statistics and their importance to scientific research. Hopefully, we have built on his argument to illustrate one direct and powerful impact of good descriptive statistics: the communication of data in the form of data visualizations.

References

Anderson, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner-researcher divide in Industrial, Work and Organizational (IWO) psychology: Where are we now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, *74*(4), 391-411. https://doi.org/10.1348/096317901167451

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, *24*(2), 443-483. https://doi.org/10.1177/1094428119877457

Aguinis, H., Ramani, R. S., Campbell, P. K., Bernal-Turnes, P., Drewry, J. M., & Edgerton, B. T. (2017). Most frequently cited sources, articles, and authors in industrial-organizational psychology textbooks: Implications for the science-practice divide, scholarly impact, and the future of the field. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *10*(4), 507-557. https://doi.org/10.1017/iop.2017.69

Berinato, S. (2016). *Good charts: The HBR guide to making smarter, more persuasive data visualizations*. Harvard Business Review Press.

Healy, K. (2018). *Data visualization: A practical introduction*. Princeton University Press.

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

Latham, G. P. (2019). Perspectives of a practitioner-scientist on organizational psychology/organizational behavior. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*, 1-16. https://doi.org/10.1146/annurev-orgpsych-012218-015323

Leybzon, D. D. (2020, May 29). Bad data visualization in the time of COVID-19. *Medium*. Retrieved from https://medium.com/nightingale/bad-data-visualization-in-the-time-of-covid-19-5a9f8198ce3e

Murphy, K. R. (2021). In praise of Table 1: The importance of making better use of descriptive statistics. *Industrial and Organizational Psychology: Perspectives on Science and Practice*. Advance online publication. Retrieved from https://www.siop.org/Research-Publications/IOP-Journal

Sawicki, J. (2020, July 28). Why is this chart bad? The ultimate guide to data visualization evaluation using GoDVE (Grammar of Data Visualization Evaluation). *Towards Data Science*. Retrieved from https://towardsdatascience.com/why-is-this-chart-bad-5f16da298afa

SIOP. (2016). Guidelines for education and training in industrial-organizational psychology. *Society for Industrial and Organizational Psychology*. Retrieved from https://www.siop.org/Events-Education/Graduate-Training-Program/Guidelines-for-Education-and-Training

SIOP. (2021). SIOP vision, mission, values, and goals. *Society for Industrial and Organizational Psychology*. Retrieved from https://www.siop.org/About-SIOP/Mission

Zaefarian, G., Kadile, V., Henneberg, S. C., & Leischnig, A. (2017). Endogeneity bias in marketing research: Problem, causes and remedies. *Industrial Marketing Management*, *65*, 39-46. https://doi.org/10.1016/j.indmarman.2017.05.006

Zelin, A., Lider, M., & Doverspike, D. (2015). *SIOP Career Study Executive Report*. Center for Organizational Research, The University of Akron. Retrieved from http://www.siop.org/tip/april15/techreport.pdf

Zhou, S., & Ahmad, A. (2020). Who's your audience? Expanding I-O teaching to non-I-O students. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 13*(4), 585-589. https://doi.org/10.1017/iop.2020.98