

Making Your Educational Data Visual

Tavinder K. Ark, PhD

Jorge A. Rodriguez, MD (@translatedmed)

Brent Thoma, MD, MA (@Brent_Thoma)

The Challenge

Graduate medical educators collect a vast amount of educational data about applicants, learners, faculty, programs, and systems. Using this data effectively can be difficult. Typically, we view our data in a table or Excel spreadsheet, draw a conclusion, and then take action. Yet optimizing the use of our data requires iterative, ongoing interaction with the stakeholders and data analysis skills. This iterative process helps us decide which features of the data are most relevant to represent visually and which visual presentation structure is best to communicate the data story.

What Is Known

Data visualization involves translating information into a visual context such as a graph, chart, or map, to render the data easier to understand and to gain insights. Effective data visualization can quickly communicate large amounts of information and complex relationships, engage viewers, and facilitate opportunities to share insights and conclusions.¹⁻³ In medical education, uses of data visualization range from assessment of resident competence in dashboards to communicating statistical findings from a research study.

Often, data visualization is viewed as the last step in data analysis, used to present key findings rather than integrated as part of the data analysis process. Actively using data visualizations throughout the analysis process can prompt recognition of relationships between data points, additional questions, and new analyses. Data can be explored in many ways, such as descriptively (eg, plotting all points, mean, or median) and in more complex ways (eg, exploring the relationships between variables and reanalyzing data by subgroup), with the goal of identifying insights that meet the needs and objectives of the audience. Graphs, such as bar charts for categorical data and boxplots for continuous data, can inform preliminary visualizations.^{1,4,5} Once finalized, visual mockups are tested with the target audience(s) to ensure that the data addresses their needs and complements the compelling narrative within the written report. Interactive data representation, such as longitudinal competency dashboards, can meaningfully inform stakeholder objectives over time as well as the data analyst's understanding of the data.^{1,3}

DOI: <http://dx.doi.org/10.4300/JGME-D-21-00944.1>

RIP OUT ACTION ITEMS

1. Explore your educational data and identify priorities for visualization.
2. Use subgroups, interactivity, color choice, and annotation to communicate a data narrative.
3. Regularly revisit the purpose and effectiveness of data representation through stakeholder engagement.

How You Can Start TODAY

1. **Identify your stakeholders.** Stakeholder perspectives are a vital consideration when visualizing data. Identify those individuals who are responsible or accountable for decisions emerging from the data and meet with them throughout the visualization process. Consider who needs to be consulted versus who just needs to be informed. In a graduate medical education (GME) context these individuals may include trainees, core faculty members, program directors, Clinical Competency Committee (CCC) members, vice chairs for education, and an institution's GME Council.
2. **Engage stakeholders.** Engage stakeholders to understand what data is available, why the data was collected, and how they envision using it in the short and long term. During this process, seek answers to 2 questions: "Is the goal to understand a problem or to make decisions?" and "Is the data collected for a one-time decision or as part of an ongoing process that requires routine data visualization updates?"
3. **Play with the data using visualizations.** Identify points and variables that you intend to represent visually, with the ultimate goals of deepening understanding, exploring relationships, and illuminating gaps. Visualize each variable independently, then in combination with other variables, to explore similarities and differences between relevant subgroups (eg, dividing a GME program's evaluation data by training year or gender might reflect relevant subgroups). This process informs the choice of data visualization. For example, when choosing a graph, you may find that either a boxplot or violin configuration best fits the data.⁵ If selecting design features, such as color, interactivity, or annotation, decide what best illuminates the narrative derived from the data.^{6,7} Through this exploration, data limitations may emerge and further sharpen understanding.

4. **Test your visuals.** Create a series of visual mockups with different representations of the same data using relevant subgroups.^{4,5} Share these mockups with your stakeholders to see if their interpretations match your intentions. How stakeholders interpret a data visualization represents a pivotal point in deciding which graphical representation best communicates the information. This iterative process minimizes data misinterpretation and inaccurate data use. For example, before presenting milestone graphs at a CCC meeting, share multiple representations with select faculty to gauge clarity.
5. **Develop a data narrative.** Work with stakeholders to collaboratively develop the data story. Meet frequently with stakeholders to determine how they interpret a given graph. These insights shape the final data visualizations. Clarifications and revisions are to be expected.



The Development of Data Visualization With Ongoing Stakeholder Engagement

concepts visually increases impact and stakeholder engagement.¹

What You Can Do LONG TERM

1. **Combine data sources.** Merging different datasets into a visualization can provide deeper insights, give broader context, or stimulate new questions. For example, combining trends on different diversity metrics into one visualization can help a program track progress in increasing diversity over time.
2. **Incorporate advanced visualizations and interactivity.** Interactivity allows stakeholders to explore. This can be as simple as allowing users to select from different graphs or other displays of the same data, or to mouse over to see more detailed data, such as means or percentages. You can also allow users to select a graph type and which variables they prefer to visualize.^{1,2,5} See the Opportunity Atlas for an interactive visual example.⁸
3. **Maintain and govern your data.** Data visualizations may need maintenance and refreshing to support your stakeholders' needs longitudinally. As part of this iterative process, establish the frequency of visualization updates up front. To maintain privacy and security, establish who has data access and ensure it is stored in a secure location; create and maintain the data workflow from access to visualization.
4. **Visualization does not always involve data.** Effective visuals can communicate text, such as educational processes or overall goals for a program. Examples include word clouds, timelines, and a program's recruitment and selection processes. Conveying

References and Resources for Further Reading

1. Kirk A. *Data Visualisation: A Handbook for Data Driven Design*. 2nd ed. Los Angeles, CA: SAGE Publications LTD; 2019.
2. Tufte ER. *Envisioning Information*. 6th ed. Cheshire, CT: Graphic Press; 1990.
3. Knaflic CN. *Storytelling With Data: A Data Visualization Guide for Business Professionals*. 1st ed. Hoboken, NJ: Wiley; 2015.
4. Cleveland WS, McGill R. Graphical perception: theory, experimentation, and application to the development of graphical methods. *J Am Statistical Assoc*. 1984;79(387):531–554. doi:10.2307/2288400
5. Holtz Y, Healy C. From Data to Viz. <https://www.data-to-viz.com/>. Accessed September 14, 2021.
6. Brewer CA. Color use guidelines for data representation. *Proceedings of the Section on Statistical Graphics, American Statistical Association*. <http://www.personal.psu.edu/cab38/ColorSch/ASApaper.html>. Accessed September 14, 2021.
7. Brewer CA, Harrower M. Colorbrewer 2.0: color advice for cartography. <https://colorbrewer2.org/>. Accessed September 14, 2021.
8. The Opportunity Atlas. <https://www.opportunityatlas.org/>. Accessed September 14, 2021.



Tavinder K. Ark, PhD, is Assistant Professor and Director, Data Science Lab at Robert D. and Patricia E. Kern Institute for the Transformation of Medical Education, Medical College of Wisconsin; **Jorge A. Rodriguez, MD**, is Instructor in Medicine, Harvard Medical School and Brigham and Women's Hospital; and **Brent Thoma, MD, MA**, is Associate Professor of Emergency Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, and Clinician Educator, Royal College of Physicians and Surgeons, Ottawa, Ontario, Canada.

Corresponding author: Tavinder K. Ark, PhD, Medical College of Wisconsin, tark@mcw.edu