Program Evaluation Toolkit
Module 8, Chapter 2: Visualizing Your Data

Regional Educational Laboratory Central

From the National Center for Education Evaluation at IES

Speaker 1:

Welcome to the second chapter of module 8. This chapter covers best practices in data visualization, which involves using graphical representations to make data easier to understand.

Data visualization is a critical part of disseminating the findings of your evaluation. Graphs, diagrams, icons, and other visual representations can help make the data more understandable. Data visualization also aids in the interpretation of the data, allowing your audience to see trends and outliers.

The *Forum Guide to Data Visualization: A Resource for Education Agencies*, available on the resource page of the website, is an excellent overview of the principles of data visualization. This chapter will focus on a few of the guide's key considerations, principles, and recommendations that are most critical to effectively disseminate the findings of your evaluation.

In chapter 1 of this module, you considered the audience, message, approach, timing, and responsible party to develop a dissemination plan. Keep the same concepts in mind when you design a data visualization.

First, who is your audience? What are their needs?

Second, what message do you want to communicate? What findings do you want to emphasize?

Next, what approach or type of data visualization is appropriate for communicating the message?

Finally, what timing is best for your overall dissemination plan?

The following slides cover each of these considerations in more detail.

To consider the audience, ask the question "Who needs or wants to learn more about your data and findings?" Once you have identified the intended audience, think about their expectations, capabilities, and needs so that you can design a data visualization that effectively communicates your message.

First, what does your audience care about? How much information would they like to have? What do they expect to see in the data visualization? It is not always necessary to share every piece of information that you gather in your evaluation. For example, in the AMMP! evaluation, it may not be important to caregivers to know the numbers of students attending the program. But this information may be vital to the funder. Thinking about what the audience cares about and what data they want to see will help guide your decisions about what to include in a data visualization.

Finally, what are the language or cultural needs of your target audience? Do you need to provide translations of the data visualization? Are there icons, images, or colors you should use or avoid?

Failing to consider your audience's needs and tailor the data visualization to those needs can cause the audience to ignore your work or to critically misunderstand the message you are trying to communicate.

After considering your audience, think about the message of the data visualization. What do you want your audience to take away from the visualization? What data do you want to share? What conclusion should the audience draw from the data?

Keep in mind that the message must be rooted in the evaluation findings. It is unethical to create a message that is not supported by the data. However, you are not obligated to share all the findings if they are not of interest to your audience.

The approach or type of data visualization you choose matters. Not every type of data visualization is best for every type of data, so choose the most appropriate visualization for sharing your data and communicating your message. This chapter covers the most common types of data visualizations, including line graphs, bar charts, scatter plots, pie charts, and tables.

A line graph is the best choice for displaying change in a value over time. Multiple lines can be used to display and compare different values. A line graph allows interpretation of small differences between values. This example graph shows student attendance data for AMMP!. The graph displays data for six students, each line showing how many AMMP! sessions a student attended each month. It is easy to see differences in the number of sessions each student attended in this example, but what if the graph included 25, 50, or 100 students? A line graph can quickly become cluttered with too many data points, so think carefully about data you are displaying before choosing a line graph.

A bar chart, also called a column chart, is used to compare quantities from different categories of data. A stacked bar chart can be used to display the values of subcategories within larger categories. Stacked bar charts are commonly used for reporting survey results. If the purpose of a bar chart is to highlight differences, the chart should be scaled so that the bars significantly differ in length or height. An audience will find it difficult to accurately distinguish small differences in height. As a general guideline, include an appropriately scaled y-axis that starts at 0 and that will not exaggerate or distort the differences in bars. Additional guidelines for scaling axes are covered later in this chapter.

This example of a bar chart displays the average number of grade 8 students attending AMMP! during each trimester during the 2020/21 school year. Because the data are in separate categories—trimesters—a line graph is not appropriate. A bar chart is more appropriate for displaying data in different categories.

A scatter plot demonstrates a relationship between the values for two variables. Dots represent the values obtained for the two variables—one plotted on the x-axis and one plotted on the y-axis. Once all values are plotted, a trend line can be added to more clearly show the relationship

between two variables or changes over time in the data. A scatter plot can also show how closely grouped data points are and may highlight outliers in the data.

In this example, homework completion rates are plotted in relation to AMMP! students' scores on the high school math placement test. A trend line shows the relationship between the two variables. The plot suggests that, as homework completion rates go up, math placement test scores increase.

A pie chart displays the parts of a whole. It is most commonly used for percentages. Because it can be difficult to interpret or compare segments accurately, a pie chart can be misinterpreted when segments are similarly sized or when some segments are small. A pie chart should always have as few categories as possible and be clearly labeled.

This example pie chart displays the total number of AMMP! participants by race or ethnicity.

Use a table when it is necessary to display the exact values for each variable, such as in an evaluation report. However, a table can be less user friendly when it presents too much information, making it difficult to share a specific message or highlight important data. A table that displays a lot of information might be best placed in an appendix rather than in the main body of a report.

This example table displays the number of AMMP! sessions six students attended each month. A table can be useful when it displays a smaller set of values, often aggregated to a higher level so that the audience can interpret trends and patterns between groups (for example, mean test scores and standard deviations for AMMP! versus non-AMMP! students), over time (for example, mean test scores and standard deviations across several years for AMMP! versus non-AMMP! students), or across content areas (such as reading and math scores).

Regardless of the type of data visualization, there are four principles related to enhancing a display and making it easier to read and interpret:

- Show the data.
- Reduce the clutter.
- Integrate text and visualizations.
- Portray the meaning of the data accurately and ethically.

You can use the *Data Visualization Checklist*, available on the resources page of the website, to check your data visualizations against these principles. Download the checklist so that you can reference it as the principles are covered in more detail on the following slides. These principles and the checklist are adapted from the *Forum Guide to Data Visualization*, referenced earlier in this chapter.

A data visualization should show the data collected and provide enough information for the audience to fully understand both the data and the message. To better show the data, follow these best practices.

First, include data labels. Labels with the exact values of the data are important so that the audience doesn't have to guess based on the axes. However, in some data visualizations, such as

scatter plots that show multiple data points clustered together, the labels would appear cluttered and be of little use to the audience.

Integrate a legend or key into the chart. The audience must be able to determine what the different parts of the visualization mean. The legend can be integrated into the visualization by labeling each part.

Also include the data source and metadata in a table or figure note if appropriate. Telling the audience where the data come from helps them make informed judgments about the conclusions and provides them with a way to find more information. Metadata are data about the data—where did the data come from and how were the data calculated. Including metadata in the visualization or in the accompanying text allows the audience to better understand the data visualization and make informed judgments.

Finally, use appropriately scaled axes. The axis scales can easily hide or exaggerate differences between values, obscuring the meaning of the data. Use an appropriate scale that allows the audience to easily see differences between the data without exaggeration. Intentionally manipulating the axes not only leads to misinterpretation but is unethical.

This data visualization shows grade 8 student attendance in AMMP!. The data are labeled so that the audience knows the exact number of students who attended each trimester of the school year. In addition, the legend is integrated into the chart so that the audience knows what the three bars represent. The y-axis includes an appropriate scale that clarifies the amount of growth for the audience. If the scale went up to 100, the growth would be more difficult to discern.

The source of the data in this bar chart is the AMMP! evaluation. Including a source may not be necessary in this case, but it is shown here to reinforce the idea that having a source is helpful. Also, the word “average” in the title of the bar chart provides metadata. In other words, the word tells the audience how these data were calculated.

Including every data element, label, and grid line tends to clutter up a visualization and make it harder to read, especially for a wide audience. To reduce the clutter, consider the following strategies.

Show only essential data. You don’t need to display every piece of data you have gathered. You have likely gathered data that do not contribute to the message you want to communicate, so don’t display those data. Be strategic about the data you display so that they focus on the message.

Remove unnecessary chart elements. Experts in data visualization refer to the data–ink ratio. Simply put, ink that does not display data is wasted. The data–ink ratio should be high: more data, less ink. Extra markings make a visualization cluttered, hard to read, and less appealing to the audience. For example, grid lines may look attractive, but they do not show data and can make visualizations harder to read. Make your visualization as clean and simple as possible by removing most grid lines, using only necessary tick marks on the axes, removing redundant labels, and eliminating the border. Avoid using decorations such as clip art. Also, 3D elements and other effects can distort data and lead to misinterpretation.

Finally, use the accompanying text. Some details can be moved to the surrounding paragraphs and table and figure notes to help declutter a visualization. However, be careful to move only noncritical information. An audience often looks at a visualization without reading the text, so all critical information should remain in the visualization.

Here are two examples of data visualizations related to the AMMP! evaluation. On the left is a clustered bar chart with a poor data–ink ratio. The example is a bit excessive, but it demonstrates how too many chart elements can make a data visualization difficult to read.

On the right is a simplified version of the data visualization. The first rule of simplification is to show only necessary data. Because the message is about grade 8 student attendance in AMMP!, grade 6 and 7 have been eliminated to focus on only the data that match the message. Next, most of the grid lines and axis labels have been eliminated, and the legends and labels are integrated. The result is a clean and easy-to-read data visualization that conveys only the intended message.

When a data visualization is integrated into a report or another dissemination approach with text, the visualization and any accompanying text should be closely integrated so that they support each other and help to reinforce audience understanding. The visualization should also stand on its own so that the audience has all the necessary information to understand it without looking at the accompanying text. Consider the following strategies for integrating the text and visualizations.

First, be strategic about the text used in the visualization. All text must convey critical information and, at the same time, be limited. Use text sparingly but for strategic purposes, such as integrated legends or limited callout boxes to highlight important pieces of information. Callout boxes are data labels that fall outside of the bars in a graph. All text should be horizontal; however, the y-axis label can be vertical. Consider turning a data visualization, especially a bar chart with a large number of bars, sideways to make it more readable.

Use descriptive titles. A title should summarize the key takeaway from a visualization and convey the meaning of the data. It is a useful way to highlight the message you want the audience to take away. Make sure to write the title in plain language so that it is accessible to all audience members.

Use arrows or callout boxes to emphasize critical information and provide additional information. Be careful, though. Use arrows or callout boxes only for the most important information. You still need to keep the simplification rule in mind.

Also, use color to emphasize the data you want to show. For example, you might use color for only the data you want to emphasize and use shades of gray for other data. If you need more than one color, be intentional about which colors you use because the audience may ascribe meaning to color. For example, red is generally understood to represent “stop” or “bad.” Using red to emphasize improvement in the data may confuse the audience and some individuals may have difficulty seeing that color. Also be mindful of color blindness and avoid red-green or blue-yellow combinations. Additionally, consider how the visualization may be viewed. If the document may be printed, there is no guarantee that a person will print it in color. Having good

color contrast will make the data visualizations easier to read not only when viewed on a screen but also when printed in black and white.

Finally, use heavier line weights or larger font sizes for emphasis without using color. Having a hierarchy of font sizes or line weights can help highlight the most important information. However, limit the number of line weights or font sizes to no more than three so that the data visualization is easy to read and understand.

This is the same data visualization of average AMMP! attendance for grade 8 students as you saw on previous slides, but there are a few changes to better integrate the text and the data visualization so that it is a complete source of information. First a descriptive title highlights the critical takeaway. The important information, such as the metadata, is still included, but now the audience knows the exact, intended message of the visualization. The legend is already integrated into the graph. Callout boxes for the labels are used to draw more attention to them without cluttering the visualization.

In addition, color is used to highlight the third trimester average. Yellow is used on the spring bar to prompt the audience to focus on grade 8 attendance in the third trimester, a point of emphasis. In addition, the yellow strongly contrasts with the blue and gray in the other bars, so the emphasis still works in black and white.

You have a responsibility to portray the meaning of data accurately and ethically in a data visualization. Misrepresenting data by introducing bias or communicating a message that is inconsistent with your findings is unethical and unadvisable. Let's look at some common concerns related to accuracy and ethics.

The first concern involves hiding negative data. It is okay to show the data that support the message you want to communicate. However, it is not okay to deliberately leave out data points that do not support your message. Cherry-picking data, or selecting only those data points that support a predetermined message, is a violation of research ethics.

Another concern is manipulating how data are visually displayed to change the meaning. Deliberate manipulation of a data visualization, such as changing an axis to exaggerate the difference between data points, is unethical.

The next concern involves using language that suggests a conclusion not supported by the data. Remember that conclusions drawn from the data must be informed by the evaluation design. It is inaccurate and unethical to make a causal claim in a descriptive figure or table title when the evaluation design does not support it.

There are also a set of general practices for creating data visualizations. These practices may not apply to every visualization you create, but they are good to keep in mind. These practices are adapted from the *Forum Guide to Data Visualization*.

First, use consistent data visualizations over time. Displaying new data in a familiar visualization allows your audience to process and understand the data quickly and easily. Be consistent in the scales on the axes, the use of color, formatting, and other aspects of the visualization.

Do not display data side by side if the data should not be compared. An audience will automatically compare data that are placed together or side by side, even if they are warned not to do so. You are responsible for ensuring that data are not compared if a comparison is not appropriate.

Also, think beyond the default data visualization. Many applications and data programs have built-in data visualizations. Although these visualizations may be a good start, the default settings are not optimized to your data and the message you want to communicate. In addition, the default settings frequently introduce a lot of unnecessary clutter into a visualization. Take the time to go beyond the default and customize the data visualization to your needs.

Focus on the message for the intended audience. One of the consistent lessons of this module has been to focus on the message and audience. It is critical that you design a data visualization with your audience in mind and communicate a focused message that will be useful to that audience. The data visualization should emphasize that message.

At the same time, use plain language. Jargon, acronyms, and technical terms tend to interfere with your message rather than enhance it. Plain language, as discussed in chapter 1 of this module, makes your data visualization accessible to a wide audience.

In addition, carefully choose fonts that display and reproduce well. Never use fonts solely for stylistic reasons because they tend to distract from the message. Different fonts should be used only to create emphasis or highlight certain information. In general, use the fewest number of fonts possible—never more than three.

Finally, use color wisely. Color should be used to highlight or emphasize information. However, relying solely on color can be problematic because some colors may not be accessible to people with color blindness. In addition, a data visualization may be printed in black and white. A better practice is to use contrast rather than color to differentiate data.

As discussed in chapter 1 of this module, it is important to ensure that your data visualizations can be understood by a wide audience. Some audience members may require narrative descriptions, called *alternative text*, of any data visualizations in your dissemination materials. Creating alternative text that describes the data being displayed will help those audience members fully engage with your materials. The *Key Considerations for Accessibility* handout, available on the resources page of the website, can help you assess whether your materials are accessible to your audience.

This concludes module 8 as well as the module series on program evaluation. For more tools from the Institute of Education Sciences, visit ies.ed.gov. We hope that you have enjoyed the module series.

This handout was prepared under Contract ED-IES-17-C-0005 by Regional Educational Laboratory Central, administered by Marzano Research. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.