

Forum Guide to Data Visualization: A Resource for Education Agencies



Forum Guide to Data Visualization: A Resource for Education Agencies





National Cooperative Education Statistics System

The National Center for Education Statistics (NCES) established the National Cooperative Education Statistics System (Cooperative System) to assist in producing and maintaining comparable and uniform information and data on early childhood, elementary, and secondary education. These data are intended to be useful for policymaking at the federal, state, and local levels.

The National Forum on Education Statistics (Forum) is an entity of the Cooperative System and, among its other activities, proposes principles of good practice to assist state and local education agencies in meeting this purpose. The Cooperative System and the Forum are supported in these endeavors by resources from NCES.

Publications of the Forum do not undergo the same formal review required for products of NCES. The information and opinions published here are those of the Forum and do not necessarily represent the policy or views of the National Center for Education Statistics or the U.S. Department of Education.

October 2016

This publication and other publications of the National Forum on Education Statistics may be found at the websites listed below.

The NCES Home Page address is <http://nces.ed.gov>

The NCES Publications and Products address is <http://nces.ed.gov/pubsearch>

The Forum Home Page address is <http://nces.ed.gov/forum>

This publication was prepared in part under Contract No. ED-CFO-10-A-0126/0002 with Quality Information Partners, Inc. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. government.

Suggested Citation

National Forum on Education Statistics. (2016). *Forum Guide to Data Visualization: A Resource for Education Agencies*. (NFES 2017-016). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Technical Contact

Ghedam Bairu

(202) 502-7304

ghedam.bairu@ed.gov



Working Group Members

This online publication was developed through the National Cooperative Education Statistics System and funded by the National Center for Education Statistics of the U.S. Department of Education. The Data Visualization Working Group of the National Forum on Education Statistics is responsible for the content.

Chair

Mike Hopkins, Rochester School Department (NH)

Members

Clare Barrett, New Jersey Department of Education

Heather Boughton, Ohio Department of Education

Wendy Geller, Vermont Agency of Education

Chandra Haislet, Maryland State Department of Education

Laurel Krsek, San Ramon Valley Unified School District (CA)

Zenaida Napa Natividad, Guam Department of Education

John Q. Porter, Mississippi Department of Education

Grady Wilburn, United States Department of Education

Susan Williams, Virginia Department of Education

Consultant

Tom Szuba, Quality Information Partners

Project Officer

Ghedam Bairu, National Center for Education Statistics

Acknowledgements

The National Forum on Education Statistics would like to thank everyone who reviewed or otherwise contributed to the development of the *Forum Guide to Data Visualization: A Resource for Education Agencies*. We would especially like to acknowledge the contributions of Ebony Walton and Lauren Musu-Gillette of the National Center for Education Statistics who shared comments and suggestions that improved this document.



Foreword

Document Purpose

The purpose of this document is to recommend data visualization practices that will help education agencies communicate data meaning in visual formats that are accessible, accurate, and actionable for a wide range of education stakeholders. Although this resource is designed for staff in education agencies, many of the visualization principles apply to other fields as well. Our focus is on tailoring visualization recommendations to meet common needs of the education community, as determined by the collective experience of our working group members.

This resource strives to

- introduce the concept of data visualization and the ways in which it can improve how education data are viewed, analyzed, communicated, and understood by a range of education stakeholders;
- describe key data visualization principles and practices that can be applied to education data; and
- explain how the data visualization process can be implemented to support effective data analysis and communication throughout an education agency.

It should be noted that this document focuses on the needs of the education community and does not reflect the full spectrum of data visualization strategies that may be used in other industries.

Intended Audience

The *Forum Guide to Data Visualization: A Resource for Education Agencies* will be of interest to anyone concerned about the utility of elementary and secondary education data. More specifically, this document is intended for staff in local, state, and federal education agencies whose responsibilities include any aspect of analyzing data or sharing data meaning with education stakeholders. This audience includes program and data staff, researchers, administrators, policymakers, and related roles associated with analyzing or presenting data for public consumption.

Development of Forum Products

Members of the Forum establish working groups to develop best practice guides in data-related areas that may be of interest to federal, state, and local education agencies. They are assisted in this work by NCES, but the content of the guides comes from the collective experience of working group members who review all products iteratively throughout the development process. After a working group completes the content and reviews a document a final time, publications are subject to examination by members of the Forum standing committee that sponsors the project. Finally, Forum members (approximately 120 people) review and formally vote to approve all documents prior to publication. NCES provides final review and approval prior to online publication. The information and opinions published here are the product of the National Forum on Education Statistics and do not necessarily represent the policies or views of the U.S. Department of Education or the National Center for Education Statistics. Readers may modify, customize, or reproduce any or all parts of this document.



About This Document

The guide is presented in the following chapters and appendices:

Chapter 1: Data Visualization in Education Organizations defines data visualization for the purposes of this document, describes how data visualization blends components of both science and art, and explains how data visualization can improve data use in the field of education.

Chapter 2: Data Visualization to Advance Data Analysis describes how data visualization can be a productive and sound technique for analysts looking to identify trends, patterns, and cues in data.

Chapter 3: Data Visualization to Improve Communications introduces four key principles and seven practical recommendations that, if adhered to, will improve the effectiveness of any effort to visualize data for audiences who need to understand and use education data to make decisions.

Chapter 4: Implementing the Data Visualization Process presents a six-step process for visualizing data for both analytical and communications purposes. Using graduation rates as an example, the document's key principles and recommended practices are implemented and illustrated.

Appendix A: Data Visualization Handouts presents handout-ready summaries of the key points of this document.

Appendix B: Citations and Additional Resources lists the citations of resources referenced in the text as well as related publications, including web materials available from the National Forum on Education Statistics, the National Center for Education Statistics (NCES), and other organizations.

The Historical Origins of Data Visualization

Engineer and economist William Playfair published the first bar chart in 1786, thus ushering in the era of data visualization. By 1801, his publication *Statistical Breviary* was credited with the display of the first pie chart. Over the hundreds of years that followed, the graphical presentation of data was largely limited to the domain of economists, statisticians, engineers, and related professionals who analyzed data and interpreted their meaning. In the past several decades, however, the nearly universal application of computing power and the tremendous volume of data it has generated has led people in other industries, including education, to consider the same problem that faced William Playfair: Is there a better way than numerical tables to analyze and communicate the meaning of large amounts of data?

Do your data inspire this response from viewers?

An education agency can reduce the likelihood of seeing this type of reaction from its data users by applying the sound data visualization practices described throughout this resource.



Contents

National Cooperative Education Statistics System	ii
Working Group Members	iii
Foreword	iv
Document Purpose	iv
Intended Audience	iv
Development of Forum Products	iv
About This Document	v
Chapter 1: Data Visualization in Education Organizations	1
Introduction	1
What is Data Visualization?	1
The Science and Art of Perception	4
Data Visualization Can Improve Data Use in Education	6
Data Visualization in Your Organization	8
Summary	9
Chapter 2: Data Visualization to Advance Data Analysis	10
Forget the Art and Focus on the Science	12
An Example of Data Visualization for Analysis	14
A Word of Caution About Causation	16
Summary	18
Chapter 3. Data Visualization to Improve Communications	19
What Belongs in a Data Visualization?	19
Key Principles for Effective Data Visualization	20
Key Principle 1: Show the data	20
Key Principle 2: Reduce the clutter	22
Key Principle 3: Integrate text and images	26
Key Principle 4: Portray data meaning accurately and ethically	27
Recommended Practices for Data Visualization	29
Recommendation 1: Capitalize on consistency	29
Recommendation 2: Data that should not be compared should not be presented side by side	29
Recommendation 3: Don't limit your design choices to default graphing programs	30
Recommendation 4: Focus on the take-home message for the target audience	30
Recommendation 5: Minimize jargon, acronyms, and technical terms	31
Recommendation 6: Choose a font that is easy to read and will reproduce well	31
Recommendation 7: Recognize the importance of color and the benefits of Section 508 compliance	31
Summary	32



Chapter 4: Implementing the Data Visualization Process	33
Step 1. Question: Someone Needs Information.....	34
Step 2. Research: Data Exploration and Analysis.....	34
Step 3. Findings: Data Meaning/ Answer.....	35
Step 4. Customization: Audience-Specific Messaging.....	36
Step 5. Visualization: Present Data Meaning Clearly and Accurately.....	36
Step 6. User Feedback: Review and Refine Efforts.....	37
Applying the Data Visualization Process to a Real-World Example.....	37
Step 1. Question: Someone Needs Information.....	39
Step 2. Research: Data Exploration and Analysis.....	39
Step 3. Findings: Data Meaning/ Answer.....	42
Step 4. Customization: Audience-Specific Messaging.....	45
Step 4. Customization: Audience-Specific Messaging.....	45
Step 5. Visualization: Present Data Meaning Clearly and Accurately.....	47
Step 6. User Feedback: Review and Refine Efforts.....	49
Summary.....	50
Appendix A. Data Visualization Handouts	51
Appendix B. Citations and Additional Resources	53

Chapter 1:

Data Visualization in Education Organizations

Introduction

Every day, 2.5 quintillion (2,500,000,000,000,000,000) bytes of data are uploaded to the Internet, meaning that 90 percent of the data in the world was generated in the last two years (IBM 2016). How are people expected to access, analyze, and interpret the meaning of such vast stores of data? In some industries, large and complex datasets (“big data”) are mined by supercomputers. In other fields in which there is an increasing need to analyze or communicate the meaning of data through visual markers, patterns, and trends, there is data visualization. Once the realm of statisticians and data specialists, the graphical display of information has become the cornerstone of basic analysis and communications in an age in which datasets have become so large and complex that they cannot be understood or expressed by routine analytical techniques (Few 2014).

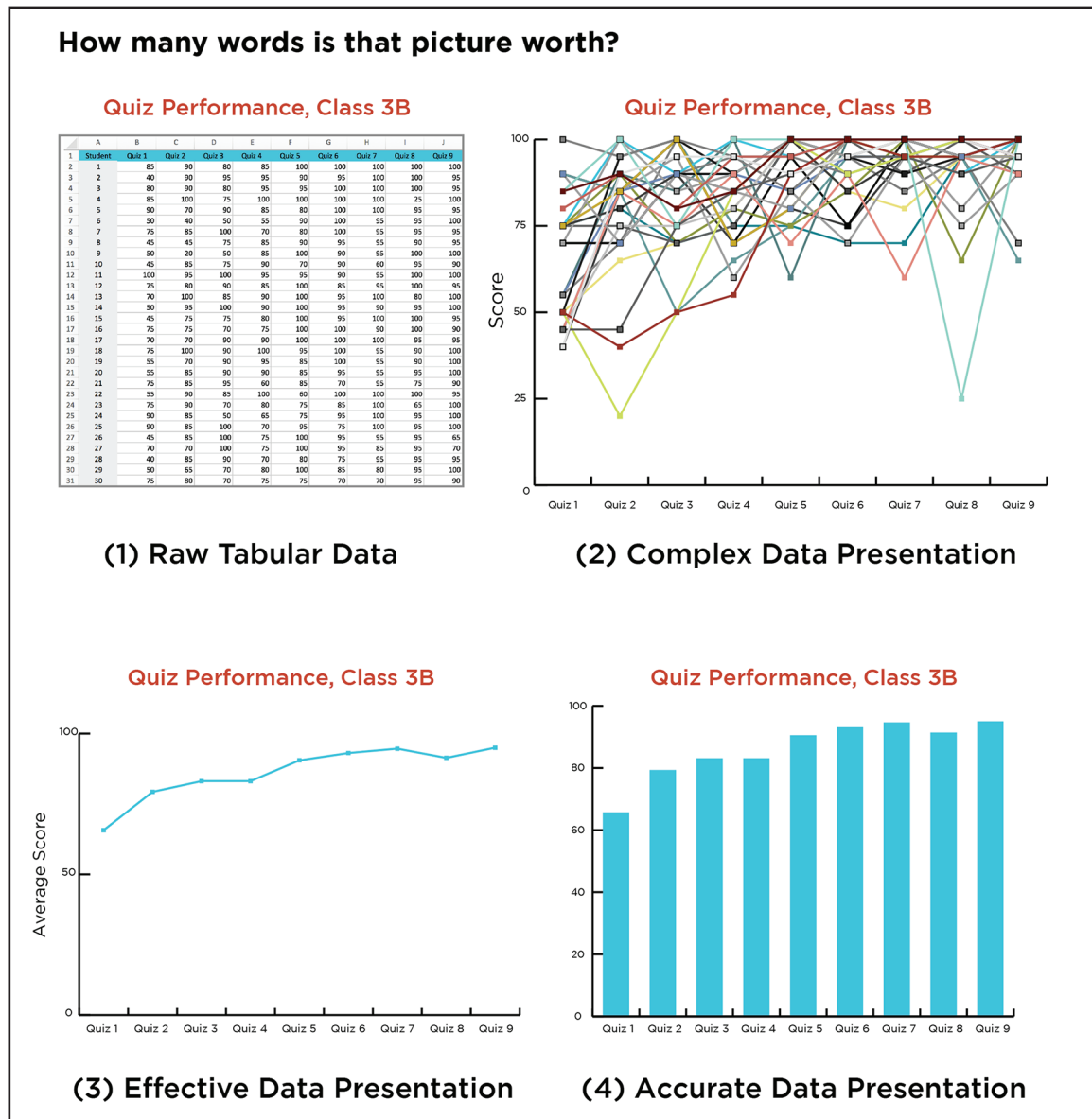
As expectations for visualized data continue to evolve, many professions have found that both expert and non-expert audiences benefit from seeing data visualized. The education community is no exception. Given the detailed data that are collected about the inputs, processes, and outcomes of the education enterprise, it is not surprising that discerning the meaning of data is a challenge for education stakeholders, including practitioners, policymakers, researchers, parents, and the general public. Although websites and textbooks about how to visualize data are readily available (e.g., Cleveland 1993; Tufte 2001; Few 2009, 2012; Evergreen 2014), they are often written for specialists in information architecture or graphic design. In contrast, this document has been customized to meet the specific needs of the education data and research communities—professionals who are engaged in interpreting data and communicating their meaning to a wide range of education stakeholders.

What is Data Visualization?

Data visualization is the transformation of data into information through visual presentation and analysis. Data visualization may culminate in a figure or image, but it should not be viewed simply as a graphical product—rather, it is the process of using a wide range of communications methods, presentation technologies, and media formats to visually reveal the meaning of data to viewers (see figure 1.1).

Data visualization is the process of graphically presenting data to reveal its patterns, trends, and meaning.

Figure 1.1. Some pictures are worth a thousand words. Others need a thousand words to interpret what they mean. Above all else, the goal of data visualization is to accurately reveal and convey data meaning that might otherwise go unnoticed or be misinterpreted in datasets and data tables.

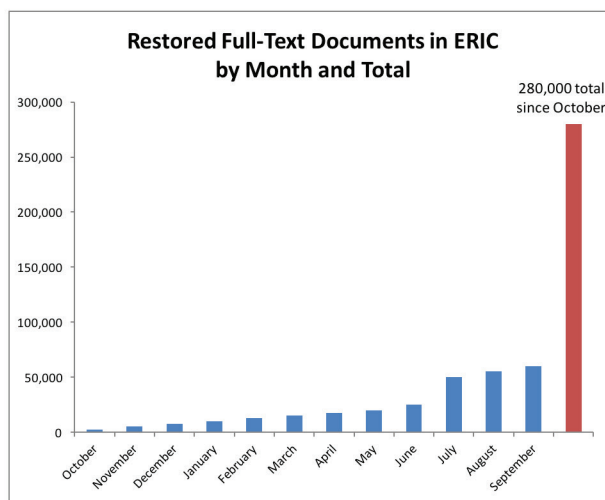


Analysis: Raw tabular data (image 1) is both detailed and comprehensive but, for most viewers, understanding what it means is nearly impossible at a glance and quite difficult even after prolonged review. A complex data presentation (image 2) may be easier to comprehend than raw tabular data, but still needs to be studied in order to be understood even though the data are presented visually. A more effective data visualization accurately portrays data in a manner that can be clearly understood by intended viewers with minimal effort or expertise, such as recognizing a trend in a class's quiz performance data (image 3). If, however, a viewer wishes to focus on differences in average class performance on successive quizzes, an alternative, more customized, presentation may make more sense (image 4).

The terms “data visualization” and “infographic” (short for “information graphic”) are often used interchangeably and incorrectly. While similar in concept, data visualizations are designed specifically to convey the meaning of datasets, whereas infographics are intended to help spread information about facts and opinions (see figure 1.2). Data visualizations differ from infographics in several important ways (Illinsky and Steele, 2011):

- *Quantity of Data:* Data visualizations tend to present larger amounts of data than infographics, which usually focus on only a few pieces of data.
- *Reusability / Regeneration:* Data visualizations can usually be repurposed for other datasets, whereas infographics tend to be tailored to convey the meaning of a specific data value or values.
- *Degree of Aesthetic Treatment:* The primary purpose of data visualization is to clarify the meaning of data (with an emphasis on data accuracy), whereas an infographic often employs more aesthetic design to display a data-driven point or argument in a more compelling manner.

Figure 1.2. Data visualizations and infographics are similar in concept, but differ in intent, construction, and outcome.



(1) Data Visualization



(2) Infographic

Analysis: Data visualizations tend to focus on presenting and clarifying the meaning of data (example 1), whereas infographics use data to make a point or support an argument (example 2).

The Science and Art of Perception

The human brain is capable of spontaneously processing particular types of visual stimuli—such as certain colors, shapes, size, and contrast—without consciously focusing attention on doing so (Treisman 1985, 1986; Wolfe and Robertson 2012). In other words, some images can catch a person’s eye before he or she even realizes it (Evergreen 2014). This phenomenon, referred to as pre-attentive visual processing, allows the human brain to simultaneously perceive and interpret the basic meaning of some visual elements in as few as 200 milliseconds, well before the conscious mind is aware of what has happened (Healey et al. 1996). Effective data visualization incorporates the science of visual cues in a way that promotes efficient and accurate understanding and communication (see figure 1.3).

Figure 1.3. Even simple visual cues can significantly streamline the brain’s visual processing. How long does it take you to process the answers to questions 1, 2, and 3?

The human eye can perceive some visual presentations more easily than others.

(1) How many occurrences of the number 3 are in this sequence? How long did it take you to determine the answer?

158453874516315484946155604564351225843689751155468910538925832451

(2) How long did it take for this sequence?

158453874516315484946155604564351225843689751155468910538925832451

(3) How long did it take for this sequence?

15845**3**874516**3**15484946155604564**3**5122584**3**6897511554689105**3**89258**3**2451

Analysis: There are six instances of the number 3 in each sequence. The number sequences are identical except for their visual presentation—the only differences are the application of color (blue or black), the contrast (the use of bold text), and the font size (the larger presentation in the third sequence). After Schwabish, J. (2014). An Economist’s Guide to Visualizing Data. *Journal of Economic Perspectives*: 28(1): 209-234.

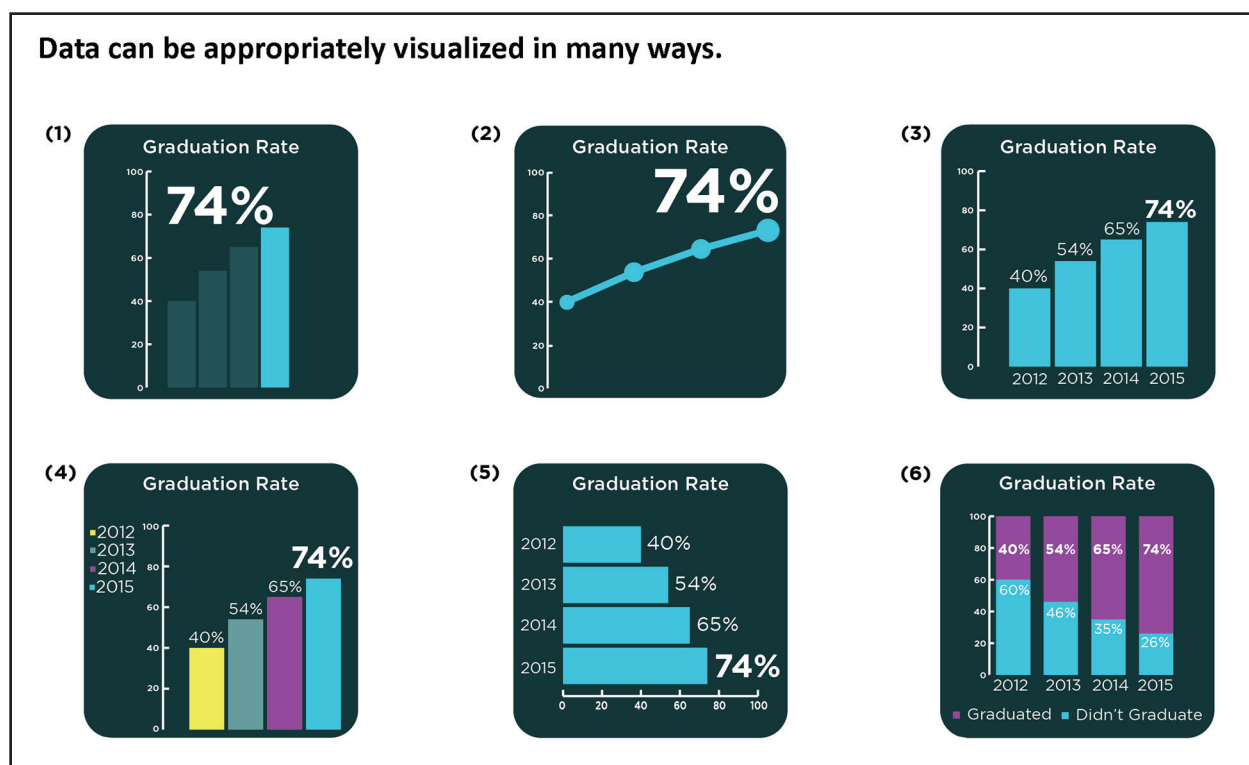
But data visualization is not just a science—aesthetics also play an important role in effectively presenting data in a graphical format. Some presentations are simply more visually appealing than others, and some artistic designs are more effective at attracting attention, improving insight, and conveying meaning to a viewer (Munari 1966; Kosara 2007). Such presentations may connect on an emotional or intellectual level with a viewer because of the type of information they contain, because of their color scheme, because they resemble previously recognized patterns or conventions, or merely—and this is difficult to define—because of subjective qualities that are difficult to quantify but nonetheless “look right.”

Though science and art can reveal much about which types of graphics are more or less likely to achieve their communications purposes, there is not a single best approach to data visualization. In fact, so many factors are involved in visualizing data—the visualization’s purpose, its intended audience, the types of data it seeks to illuminate, the media

used to portray it, each viewer's personal response to its artistic and intellectual elements, and so on—that simple, hard-and-fast rules for data visualization do not, and cannot, exist. Yet there are general principles, best practices, and recommendations that can and should inform the design of a data visualization (see chapter 2 and chapter 3). Ideally, the data visualization process is most effective when the science of perception and objective data standards are integrated with subjective artistic and communications choices to meet the specific information needs of an intended audience (see figure 1.4)

Data visualization is a lot like writing in that there are few hard-and-fast rules for success, but there are several key principles and practical recommendations that can help one to produce more effective communications tools (in visualized or written formats).

Figure 1.4. Although there are various ways to visualize data, subjective interpretation plays a large role in visualization choices, and becomes especially critical when determining what information can be shared without obscuring or otherwise deviating from the primary meaning of the data.



Analysis: Each of these examples has strengths and weaknesses, depending on the intended audience and the message to be communicated. Good data visualization strives to present the right amount of information and aesthetic style for a specific message and audience. The same visualization might be appropriate for one message or audience, but not for another.

Data Visualization Can Improve Data Use in Education

A host of different types of stakeholders in the education community routinely use data to make decisions:

- Teachers look at student performance data to identify knowledge gaps and customize instruction.
- Administrators view enrollment and coursetaking records to create class schedules.
- School board members assess fiscal data to ensure equitable resources across school campuses.
- Researchers scrutinize outcome data to evaluate the effectiveness of curricula and instruction.
- Parents examine school- and district-level graduation rates to determine where to purchase a home.
- Community members consider expenditure and revenue data to decide how to vote on tax increases.

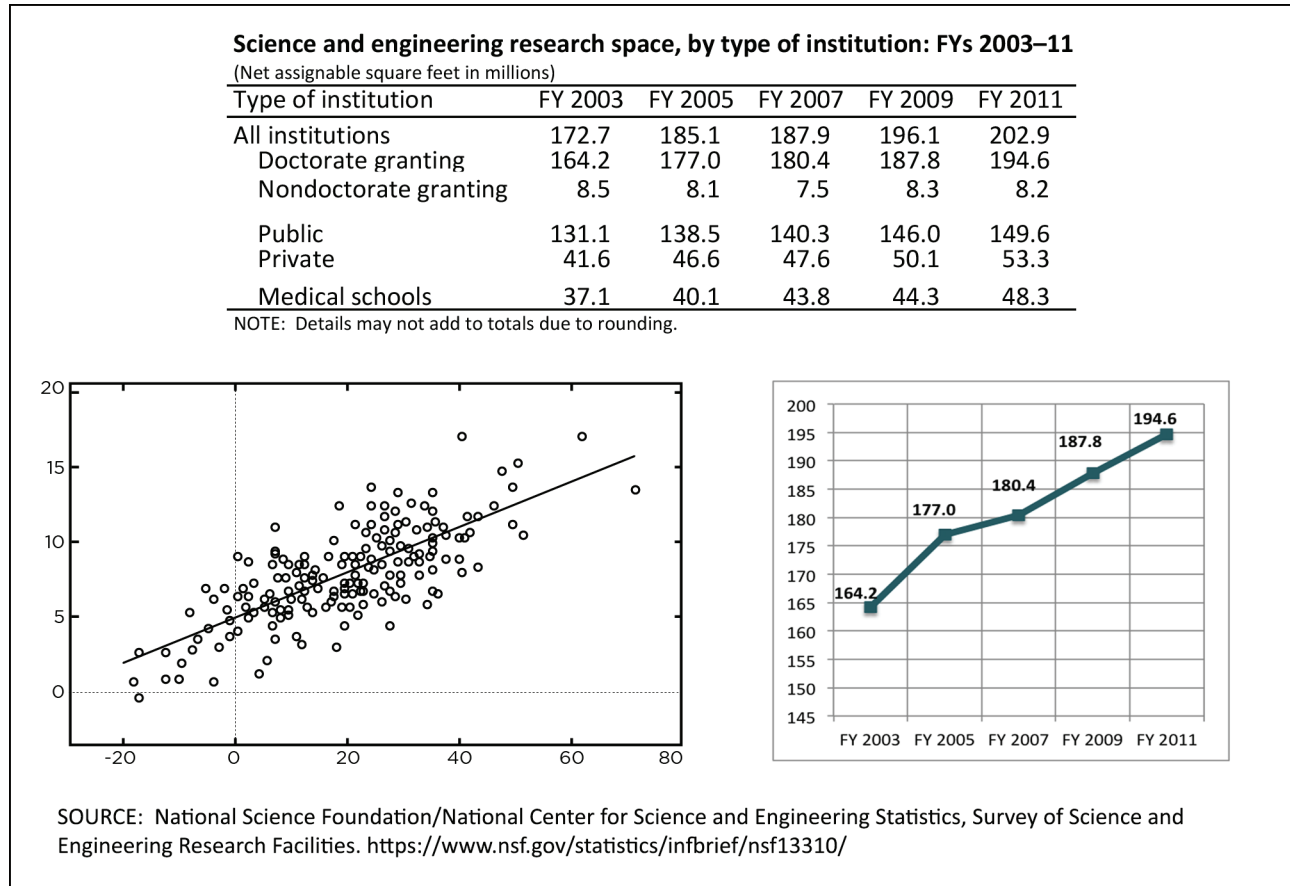
In each of these examples, it is critical that the data are accurate, reliable, and timely (collectively referred to as “high quality”) because of the high stakes consequences of those decisions on various stakeholders in the education system (National Forum on Education Statistics, 2012 and 2015). But even when high-quality data are available, they need to be presented in a way that meets each audience’s unique information needs. After all, teachers, board members, researchers, and parents each bring different information needs and expertise to their use of data.

Many of these stakeholders will find data to be more understandable when they are presented in a visually accessible manner. As such, any and all data generated by your agency for decisionmaking purposes may be a candidate for visualization; however, while data visualization is usually a useful step in identifying and communicating data meaning in education agencies, it is not always necessary. In some cases, it could be appropriate to share data in “pure” form as raw data. In other instances, data might only need to be minimally treated, as occurs when it is presented as a statistical value. But for most stakeholders in education settings, data will be easier to understand, interpret, and use when they have been visualized (see figure 1.5).


More examples of how visualized data can be used in education agencies, including the pros and cons of various visualization choices, are included in chapter 2, chapter 3, and chapter 4. Common education data topics such as the following are addressed as examples, case studies, and hypothetical scenarios:

- test scores
- student attendance
- classes missed for extracurricular activities
- student enrollment
- dropout rates
- child poverty rates (by state)
- graduation rates

Figure 1.5. For many stakeholders, the meaning of visualized data is often more clear than untreated data tables, but presentation choices should reflect the information needs of the intended audience.



Analysis: Although data in tables are “pure” in the sense that all of the values are visible, unanalyzed rows and columns of numbers leave a reader to his or her own methods for distilling meaning (top image), which often requires some degree of analytical expertise and can lead to misinterpretation. A graph of data that have undergone statistical treatment (bottom left) can be precisely what is needed by research audiences, but other viewers may have more success understanding a more intuitive visual presentation of a dataset (bottom right).



It is a sound communications practice to customize messages to meet the needs of as many audiences as possible. In 1998, Congress amended the Rehabilitation Act of 1973 to require federal agencies, as well as organizations receiving federal funds, to make their electronic and information technology accessible to people with disabilities (Section 508 of 29 U.S.C. § 794 (d)). With respect to data visualization, graphics that cannot be interpreted by assistive technologies such as screen readers and magnifiers may not be accessible to audiences who are blind, color blind, or otherwise disabled. As such, compliance with federal Section 508 Accessibility Guidelines is not just encouraged—it and comparable state and local regulations are often required by law.¹

Data Visualization in Your Organization

Once an education agency has gone to the effort of collecting data, failing to use it squanders an opportunity to put valuable information into action. Implementing data visualization throughout an agency will substantially improve data use—and is therefore likely to be an emerging leadership priority in many education organizations.

If data visualization is a priority for senior leadership, it will be recognized as a priority by management, program, research, data, and communications staff, as well.


The data visualization process works best when it is envisioned, implemented, and managed at an organizational level, rather than within a single department or by individual staff members. Regardless of the size of your agency, it is important to view data visualization as an initiative that will be carried out by a team from across all major data and reporting departments in the agency. It is not a one-time or one-person exercise.

Visualization activities should be aligned with the organization's broader data governance framework, and team members should include staff involved in policymaking, research, data, communications, and program content (such as people with expertise in curriculum, instruction, and program areas). As such, the organization should recognize the different, yet critical, roles that each team member is assigned in the data visualization process. For example, some people in an organization may be charged with discerning the meaning of the data, while others work on validating the accuracy of this analysis and interpretation. Meanwhile, some staff translate findings into more understandable formats, while others confirm that these presentations are appropriate for intended audiences. Given the range of roles, specialized skills, and organizational authority needed to implement such a process, senior leadership will want to determine how data visualization responsibilities are assigned. They will also want to ensure that staff engaged in data visualization activities adhere to applicable data governance and communications policies, either through collegial encouragement or more formal requirements.²

Organization-wide success will be more likely when a staff training program is implemented. To ensure an efficient and effective program, it is important that the training be tailored to each role and responsibility in the visualization production and dissemination process. After all, the staff member who uses data visualization to improve analysis does not need the same skills and training as the person who designs visualizations for reports to parents and community members.

¹ For more information about the application of Section 508 Accessibility Guidelines, see: National Forum on Education Statistics. (2011). *Forum Guide to Ensuring Equal Access to Education Websites: An Introduction to Electronic Information Accessibility Standards* (NFES 2011–807). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Available at http://nces.ed.gov/forum/pub_2011807.asp.

² The National Forum on Education Statistics has produced best practice guides for the education community on a range of topics relating to data governance, data management, data privacy, data quality, and data use. This document extends and customizes recommendations from several of these Forum resources to include data visualization (data communications) as a critical component of the education community's efforts to improve the quality, comparability, and utility of education data. Visit <http://nces.ed.gov/forum/publications.asp> to access these and other free Forum resources.



As data visualization becomes a more integrated aspect of an organizational culture, leaders can strengthen the organization's data visualization practices in numerous ways. They might, for example, offer ongoing opportunities for staff development, encourage staff collaboration for the purpose of streamlining processes, or routinely share examples of more—and less—effective visualization products.

With respect to measuring the effectiveness of data visualization efforts, the guiding metric cannot be simply the percentage of data that have been visualized. Visualization is a highly customized endeavor, and your staff might determine that some types of data can be understood by their audiences without being visualized.

A more effective way to measure the success of a data visualization initiative is to ask your internal and external audiences whether visualization efforts are improving data analysis, communications, and understanding. Some organizations may choose to accomplish this through formal data user councils that are convened on a regular basis to provide feedback. Other agencies might solicit feedback via online surveys or through appraisal forms that accompany print products. In any case, the most important thing to remember is that the most important feedback will come from your audiences—and just as there is not a single “right” way to visualize data, there is not a single “effective” formula for gathering input from target audiences. The ideal approach is frequently whatever is most practical for your organization and its stakeholders.

Data visualization for public consumption should aim to be “no training required” for viewers—in other words, audiences should not need to possess specific skills to understand visualized data.

Summary

Education agencies share data with students, staff, parents, community members, policymakers, and researchers because the information is judged to be of value. Accordingly, providing education stakeholders with clear and accurate information about education organizations, processes, and performance is a fair, necessary, empowering, and healthy component of our education system.

The key to effective data visualization is to customize best practice visualization techniques in a manner that is most likely to meet the specific information needs of each intended audience.

The ability to create customized, audience-specific data visualizations can become a vital component of a broader organization-wide data analysis and communications strategy. Data visualization focuses on presenting information in a way that is not only accurate, reliable, timely, and appropriately comprehensive, but also understandable and actionable for each of your intended audiences. When appropriately applied, the data visualization approaches described in this document will improve a data consumer's ability to understand and analyze data, extract information, and use that information to make data-driven decisions.

Effective data visualization is

- valuable as an analytical and communications tool because of the insights it can provide through visually apparent cues, patterns, and trends;
- customized to meet the information needs of specific intended audiences; and
- designed to reduce the likelihood of viewers misunderstanding or misinterpreting data.

Effective data visualization is *not*

- emphasizing presentation over message in a way that distorts or distracts from meaning; or
- more complex or creative than it needs to be to accurately convey data meaning.

Chapter 2:

Data Visualization to Advance Data Analysis

This chapter focuses on the needs of a data analyst or group of analysts who are trying to determine what a particular set of data, or multiple datasets, might mean, but are not intending to share their analytical products with other audiences.

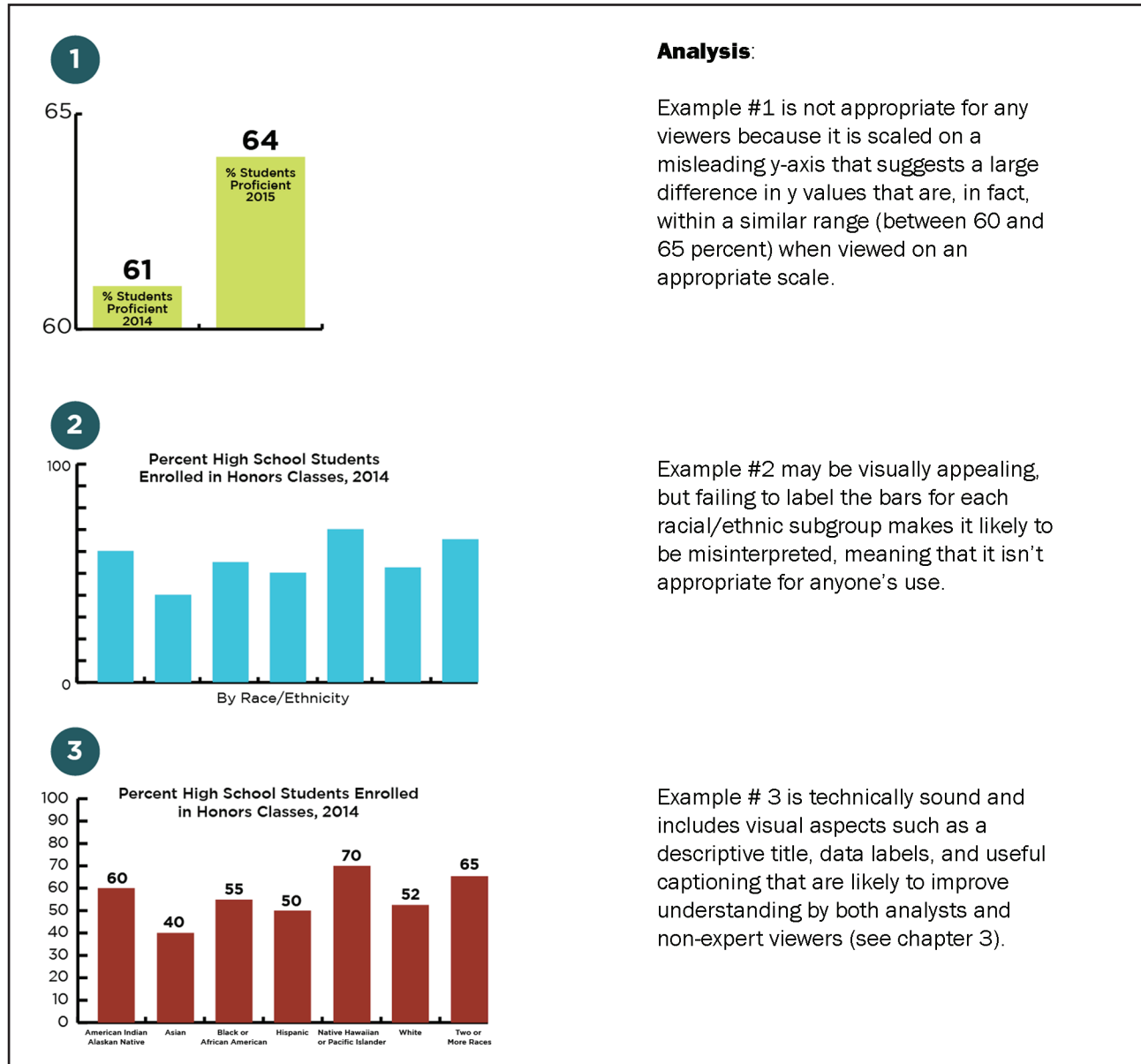
Although Chapter 1 describes data visualization as a blend of science and art, the intent of data visualization is insight, not pictures.³

In contrast to some staff members who are tasked with communicating the meaning of data to other users, data analysts do not always need to make their visualizations attractive to the eye. A data analyst is a technical expert, studying the numbers and striving to make sense of them—thus, accuracy and simplicity are often more appropriate goals for analytical purposes than aesthetic appeal. By far the most important visualization priority for data analysts is preserving the integrity of data without introducing features that distort or distract from their meaning (see chapter 3 for appropriate approaches to communicating data in an accurate, but more visually appealing manner for general audiences).

While data visualizations for analysis may not need to be works of art, they cannot be haphazard or disorderly (see figure 2.1). It is important that graphics accurately represent the data. For instance, using an inappropriate scale for the x-axis or y-axis can obscure data meaning. Similarly, labeling axes, lines, or bars incorrectly, or not labeling them at all, can cause confusion and misunderstanding—and failing to account for how data were collected or defined can lead to erroneous conclusions, even if every point is plotted accurately, every label is precise, and the scale is appropriate.

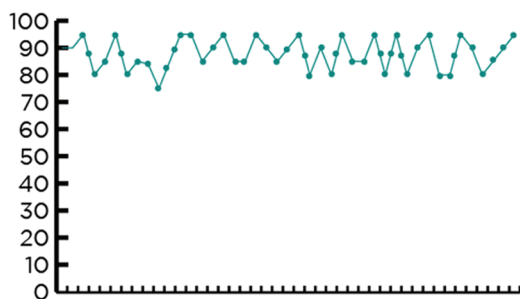
³The quote “The intent of data visualization is insight, not pictures” is widely attributed to Ben Shneiderman in Card, S.K., Mackinlay, J.D., and Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Academic Press: San Diego, CA.

Figure 2.1. Data visualizations must accurately portray the data but, depending on the intended use, visualizations for analytical purposes do not necessarily need to be aesthetically appealing. The following data visualizations might “look” better or worse to varying degrees, but are not equally suitable for all types of audiences.



4

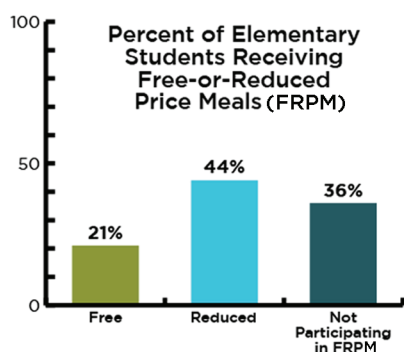
Average Daily Attendance Percentage



Example #4 appears to be accurate, but its lack of visual features, such as axis labels, makes it inappropriate for non-expert viewers, although it may be of use to analysts and other technical experts for coarse analysis.

5

Percent of Elementary Students Receiving Free-or-Reduced Price Meals (FRPM)



Example #5 may include the visually helpful features of the previous examples, but the data values are imprecise—as recognizable through partial percentage values that do not add up to 100 percent. Without explanation (such as rounding errors), it is inaccurate and, therefore, may detract from the credibility of an agency's analysis or conclusions.

Forget the Art and Focus on the Science

Why do mathematicians—people who spend their lives manipulating numbers, looking for numerical patterns, and thinking about how values can be expressed—like to graph equations? Why do they plot points, draw vector arrows, and diagram multidimensional objects? After all, a graph is not a series of ciphers, but a picture: a visual representation of something real or imagined—a rendering on paper, whiteboard, or screen that favors lines, curves, and dimensionality over numerals and symbols.

The answer lies in the fact that a mathematician's graph is not art. While such a graph may contain many artistic elements, and while those artistic features may be vital to the task of communicating mathematical concepts to a more general audience (see chapter 3), the mathematician's overarching purpose for graphing numbers is to illuminate the patterns, trends, and cues inherent in those numeric values. That is, the mathematician's aim is to discover things of significance that were previously not visible or were difficult to see. At heart, such a search for meaning is science, not art.

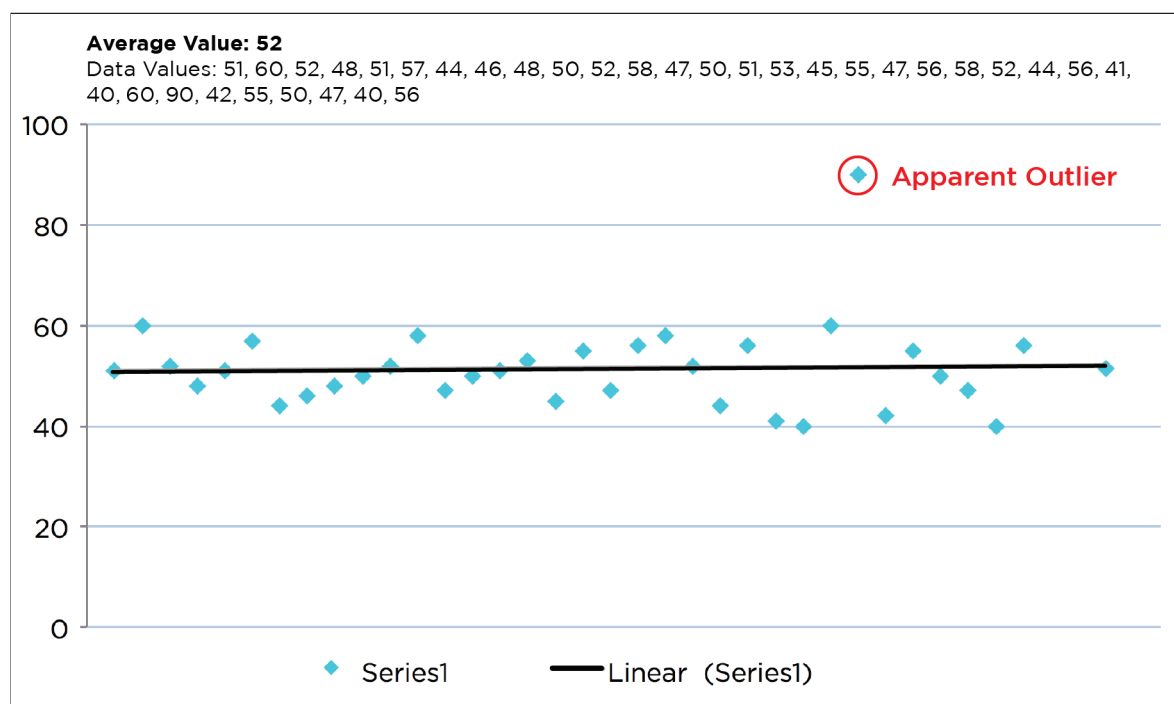
The word “illustrate” has two common meanings: (1) to draw a picture and (2) to show meaning or clarify an idea. That's because drawing a picture of a concept often improves understanding of that concept. Even if you're a mathematician,

statistician, researcher, scientist, analyst, or other kind of data expert, graphing numbers is likely to generate fresh insight into what numbers mean. While not every set of numbers needs to be graphed to be understood, graphing is an important tool for transforming numerical and statistical values into more understandable and meaningful concepts, even for mathematical, data, and other analytical experts. In this way, data visualization is essential even for those whose job it is to study and interpret data.

Imagine a large set of numbers that have broadly similar values, with the exception of one extreme value—an outlier. Scanning or averaging the values in the dataset is unlikely to expose the one point that stands out, but graphing those points would offer a good chance of revealing the outlier (see figure 2.2).

Over the past several decades, the increasing amounts of data being collected in the education field have led to larger datasets that need to be analyzed and interpreted. Spreadsheets and other data processing tools can readily perform mathematical and statistical functions such as distribution and regression analysis. However, when a dataset is large, a graphical rendering of the data often reveals potentially meaningful features that are difficult to identify with routine statistical summaries.

Figure 2.2. Even small datasets can be easier to interpret when viewed in graphical form.



Analysis: A visual review of the data reveals the outlying data point (90) much more quickly than an assessment of the average value (52) or a quick review of the original data points. This example includes only 34 integer values, but many datasets are much larger and more complex, which further complicates the identification and analysis of meaningful features in the dataset.

Of course, small and simple datasets are usually easier to interpret in tabular form than large and complicated ones. Interestingly, though, even when the dataset is small and relatively simple, data visualization can uncover important features that might otherwise go unnoticed (Card et al. 1999). This is especially true when parsing meaning between and across multiple datasets. For example, an analyst might gain new insights from viewing a small dataset overlaid on another small dataset in a single visualization. An analyst might then overlay a third graph to discover even more pertinent information (discussed later in this chapter.)

An Example of Data Visualization for Analysis

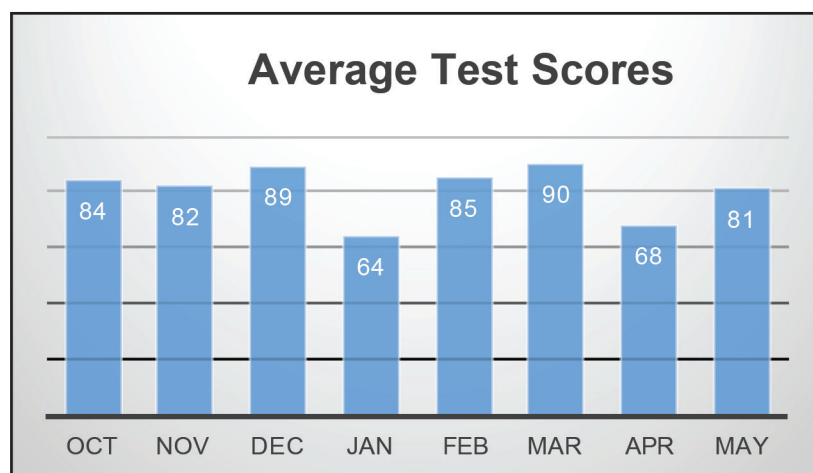
One common method for analyzing data is to overlay two or more visualized datasets to create a new, and possibly more illuminating, data visualization. Seeing different sets of data in a visual format, juxtaposed together on one graph, can spark revelations about the patterns and trends in the data (that is, possible ways to interpret what is observed in the data). Analysts can describe the patterns they see, and then use this information to inform subsequent analysis and understanding—with such observations potentially yielding new hypotheses and lines of inquiry.

Consider an example of how such an application of data visualization might enable a data analyst to identify meaning within and across multiple datasets.

Suppose that Hypothetical Middle School tests its students at the end of each month. The average score was relatively high in most months, but in January and April, the scores were significantly lower (see figure 2.3). Why might the scores have dropped in those two months?

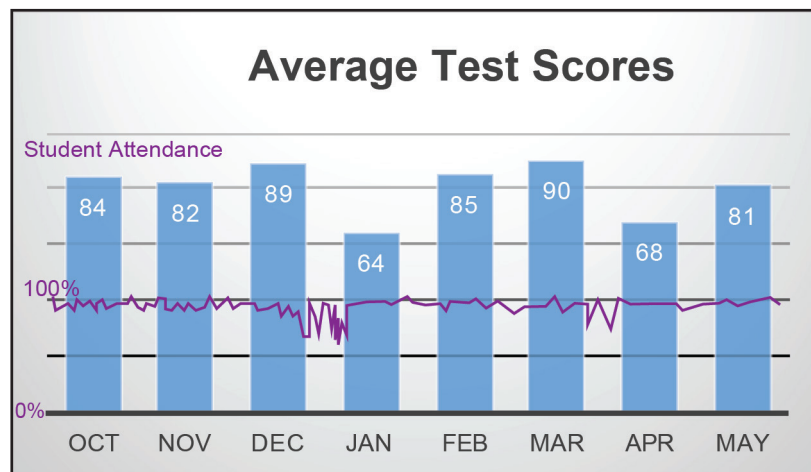
When properly applied, data visualization as a tool for observing patterns in data is a valid application of the scientific method for advancing knowledge by observing a phenomenon (such as a pattern or trend in data), identifying research questions, generating hypotheses, testing hypotheses, and then producing new insights and understanding that fuel future observation, hypothesis, and research.

Figure 2.3. To illustrate an example of how data visualization can be used for data analysis, consider student test scores in which two months (January and April) appear to be lower than other monthly averages.



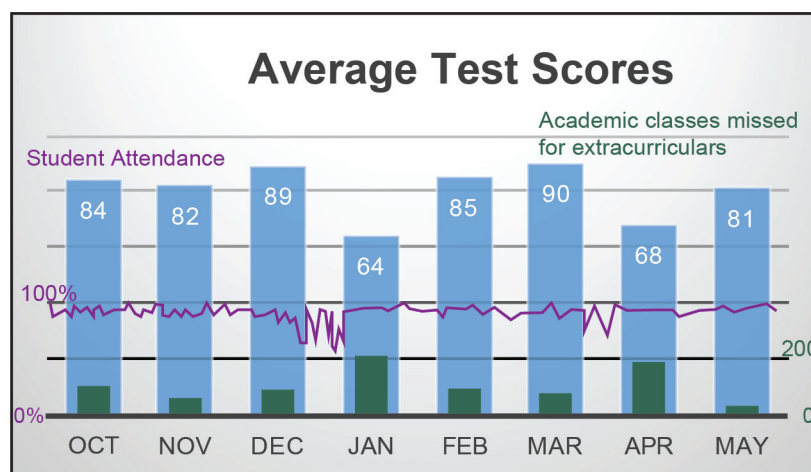
A data analyst looks into the matter by incorporating a range of available data. The first dataset the analyst incorporates into the visualization is daily attendance over the entire school year (see figure 2.4). The analyst notices that student attendance was lower than usual in the month prior to each of the two test dates that had lower scores. The analyst does not assume that the lower test scores were caused by the increased number of absences, but notes that the idea is interesting and may have merit as a plausibly related factor. It is concluded that it might be a good hypothesis to test.


Figure 2.4. An overlay of daily attendance data shows the appearance of lower attendance immediately preceding the months with lower average test scores.



The analyst understands that many factors can affect test scores, so other data are assessed as well. This time, the analyst creates a bar graph that indicates the number of academic classes students missed because of excused absences due to extracurricular activities throughout the year (see figure 2.5). The bar graph reveals that in the months of January and April, there were almost twice as many excused absences for extracurricular activities as there were during the other months. Without jumping to conclusions about causation, the analyst speculates whether excused absences due to extracurricular activities might be another plausible cause for the lower test scores during those months.

Figure 2.5. The number of classes missed because of extracurricular activities peaks during months with lower average test scores.





The analyst understands that correlation is not causation but believes that preliminary data analysis suggests three rational, related, observation-based hypotheses (that is, based on patterns seen in actual data) that may need to be studied more formally: (1) decreases in student attendance in certain months resulted in lower average test scores in the following month; (2) increases in excused absences for extracurricular activities resulted in lower average monthly test scores; and (3) decreases in student attendance and increases in excused absences for extracurricular activities combined to result in lower average monthly test scores.

The analyst documents these correlations and preliminary hypotheses in a report to the principal of Hypothetical Middle School—careful to present this information in terms that the principal and administrators will find to be clear and useful. As such, the report offers a few cautious recommendations that reflect identifiable trends in the data that may be relevant. The analyst explains that the research team has been looking into the possibility of whether increased numbers of absences (as seen in daily attendance and extracurricular activities) might have had a negative effect on the school's average test scores during specific months, and advises the principal to take these factors into account when reflecting on the test score data. The analyst also recommends that the principal keep apprised about extracurricular scheduling and take that into account when planning for testing. Finally, the principal is reminded that although these factors appear to be correlated, they are not necessarily connected causally—and can't be considered causes for the lower scores without conducting additional research (see “A Word of Caution about Causation” below).

A Word of Caution About Causation

“Correlation is not causation.” As every statistician knows, just because there's a pattern in the data does not mean that that pattern has significance or that fluctuation in one variable causes a change in another variable. When two datasets (such as the hypothetical sets A and B) are correlated—that is, when they tend to fluctuate or vary in similar patterns—any of the following possibilities may be true:

- A causes B.
- B causes A.
- A and B cause each other (in a cycle).
- C causes both A and B.
- A causes C which causes B.
- B causes C which causes A.
- Another form of causation is at work.
- There is no causation at all (it's a coincidence).

Not obeying the statistician's mantra that “correlation is not causation” can lead to mistaken interpretations of data meaning (see figure 2.6). Having acknowledged this, the appearance of patterns, trends, and cues in visualized data, including correlations, can suggest a need to further investigate relationships that might otherwise go unnoticed in datasets.

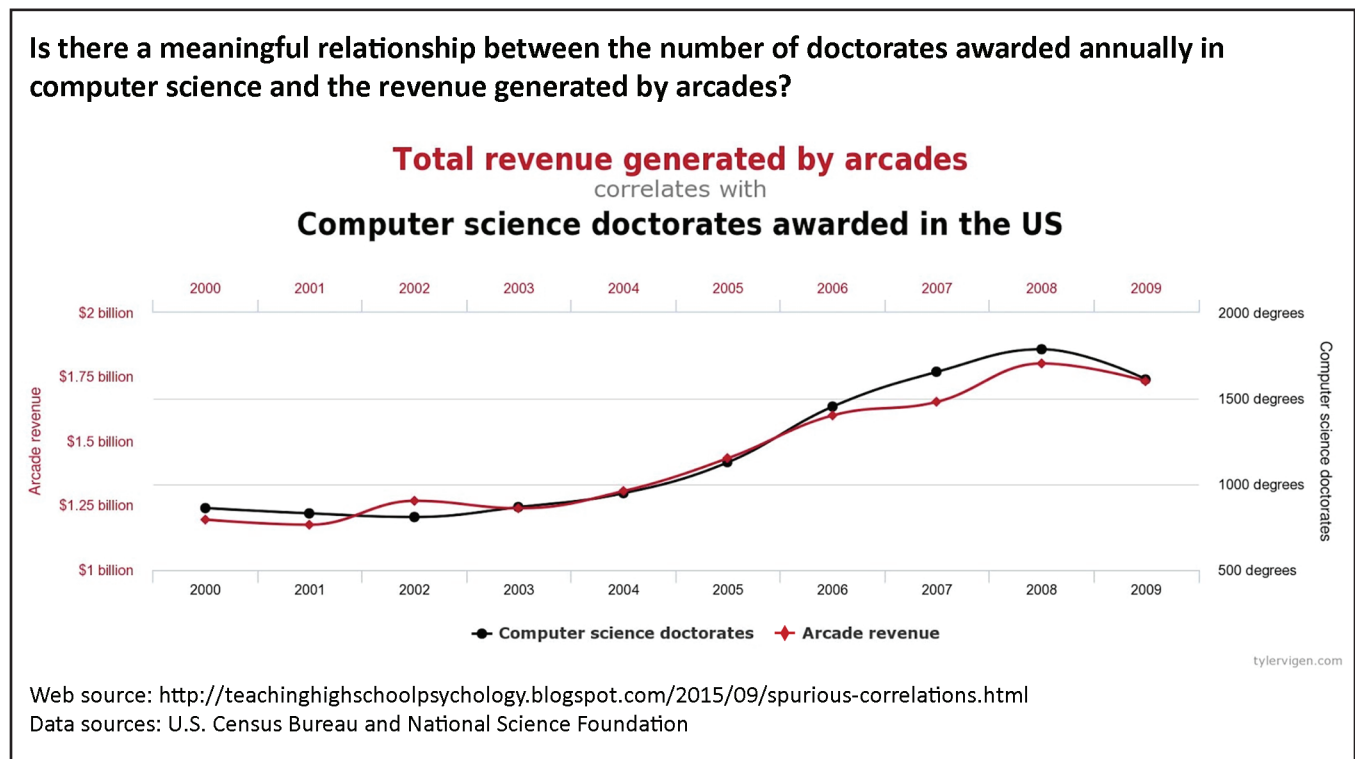
Sound data analysis does not jump to conclusions about causation simply because of the appearance of visual patterns, trends, and cues. Rather, a wise data analyst notices correlations in the data (often with the help of data visualization techniques), describes those correlations, and makes cautious hypotheses about potential relationships that are intended to serve as starting points for further inquiry.

Mistakenly assuming causation is not the only pitfall that a data analyst might encounter. Remember that a data visualization is merely an alternate way of looking at numbers. Any inaccuracies inherent in the data will, naturally, be

reproduced when the numbers are visualized in a graphic format. While a lengthy discussion of data integrity is beyond the scope of this document, the following tips should be kept in mind when interpreting visualized data:

- What is the quality of the data? For example, do the data measure what they purport to measure? Can the data be reproduced consistently over time? Are they timely enough to be relevant? Was there any bias in the data source(s) or bias in how the data were collected (large or small samples), treated (statistical manipulation), or analyzed (objectively or to make a point)?
- Are the data relevant to the question being studied? Has this relevance been proven or assumed?
- What are the data's limitations? Is it applicable to your setting, population, and circumstances?⁴

Figure 2.6. The statisticians are right: correlation is not causation.



Analysis: The data in the visualization above originate from trusted sources such as the U.S. Census Bureau and the National Science Foundation—and it is evident from the visual appearance of the data points that both sets of lines vary in similar patterns. Nonetheless, the adage “correlation is not causation” reminds analysts that similar changes over time do not justify claims that either of the variables caused fluctuations in the other. In fact, many observers would suggest that these highly correlated values do not have any meaningful relationship with each other based simply on the correlation of the data points.

⁴ For more information about data quality, see the National Forum on Education Statistics. (2005). *Forum Guide to Building a Culture of Quality Data: A School and District Resource* (NFES 2005–801). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Available online at <http://nces.ed.gov/pubs2005/2005801.pdf> and the National Forum on Education Statistics. (2005). *Forum Guide to Education Indicators* (NFES 2005–802). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Available online at <http://nces.ed.gov/pubs2005/2005802.pdf>.



Summary

Data analysts use visualization as a research tool. While there are many methods of statistical analysis that do not require visualization, viewing data in graphical form often discloses patterns, trends, and cues that might otherwise remain unnoticed. While visualization is an especially helpful tool for identifying meaning in large datasets, even small datasets can reveal unexpected features when graphed. This is especially true when visualization is used to overlay (and compare) multiple datasets.

Scientific fidelity is essential when creating data visualizations for any purpose. Accuracy and simplicity are more important than aesthetic appeal for visualizations intended for analytical purposes. Finally, analysts should take time to think about the quality of the data and, at all times, remember that correlation is not causation.

Chapter 3:

Data Visualization to Improve Communications

In addition to being a valuable tool for enhancing data analysis, data visualization is essential for presenting information in a manner that communicates data meaning to a range of audiences—especially non-expert viewers. Computing technologies, in combination with data systems yielding great stores of valuable information, are available to visualize data in ways that were not previously possible with traditional paper-and-ink methods. While visualization technologies range from relatively unsophisticated online graphing programs to state-of-the-art visualization applications, technology is only a tool for presenting data in an appropriate format.⁵ Effectively communicating meaning reflects sound principles and proven practices for meeting the information needs of intended audiences, which is the focus of this chapter.

What Belongs in a Data Visualization?

The ability to create customized, audience-specific data visualizations is an important aspect of a broad, organization-wide analytical and communications strategy. Data visualization emphasizes presenting information in a way that is not only accurate, reliable, timely, and appropriately comprehensive, but also exceptionally understandable, for your intended audience.

There are many methodologies for visualizing data. However, in many ways, data visualization for communications purposes boils down to the following four principles that serve as the foundation for helping viewers more readily understand information:⁶

1. **Show the data.**
2. **Reduce the clutter.**
3. **Integrate text and images.**
4. **Portray data meaning accurately and ethically.**

These key principles of data visualization will help viewers who are not data experts more accurately understand the meaning of the data being displayed.

These key principles are presented as overarching points of emphasis that should be adhered to under all circumstances when developing visualizations. The seven recommended practices described later in this chapter are generally helpful but do not apply to all data visualizations in the same way as a key principle.

⁵ NCES's Create-A-Graph tool (<http://nces.ed.gov/nceskids/createagraph/>) is an example of a popular tool for helping students and other members of the education community present data in graphical form. While useful for many purposes, including instructional uses, it does not apply the full spectrum of best practice principles for effective data visualization.

⁶ The first three principles originate in Schwabish, Jonathan A. (2014). An Economist's Guide to Visualizing Data. *Journal of Economic Perspectives*, 28(1), 209–234. <http://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.1.209>.

An effective data visualization is designed to

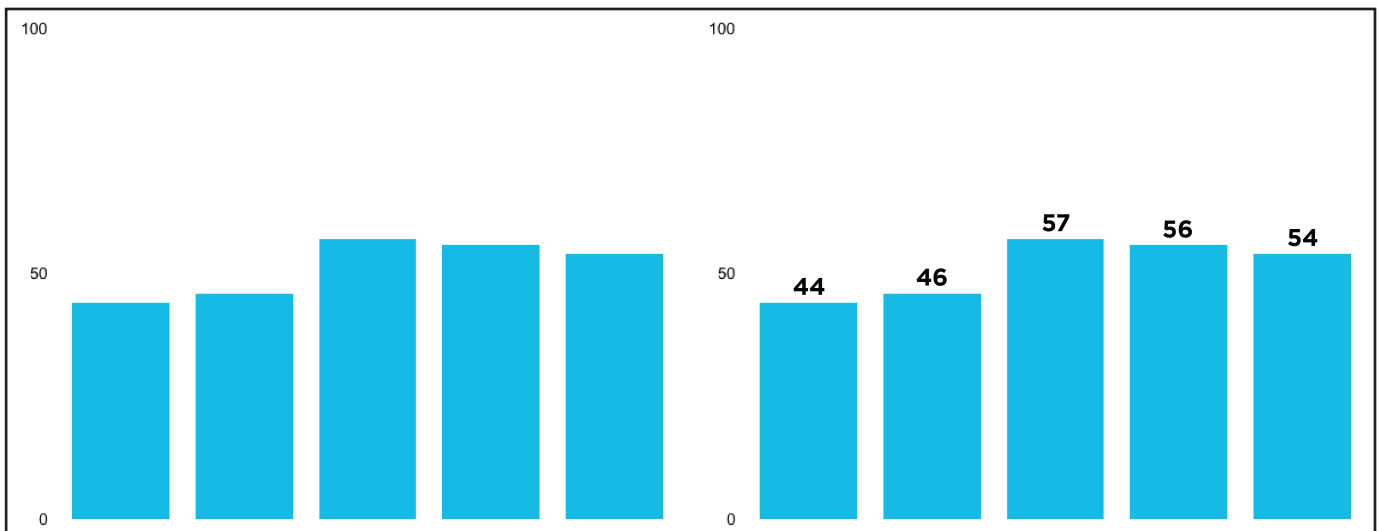
- display data;
- focus on the primary message by avoiding purely cosmetic “bells and whistles”;
- present images and textual descriptions in a complementary manner; and
- represent data meaning in an accurate and ethical manner at all times.

Key Principles for Effective Data Visualization

Key Principle 1: Show the data.

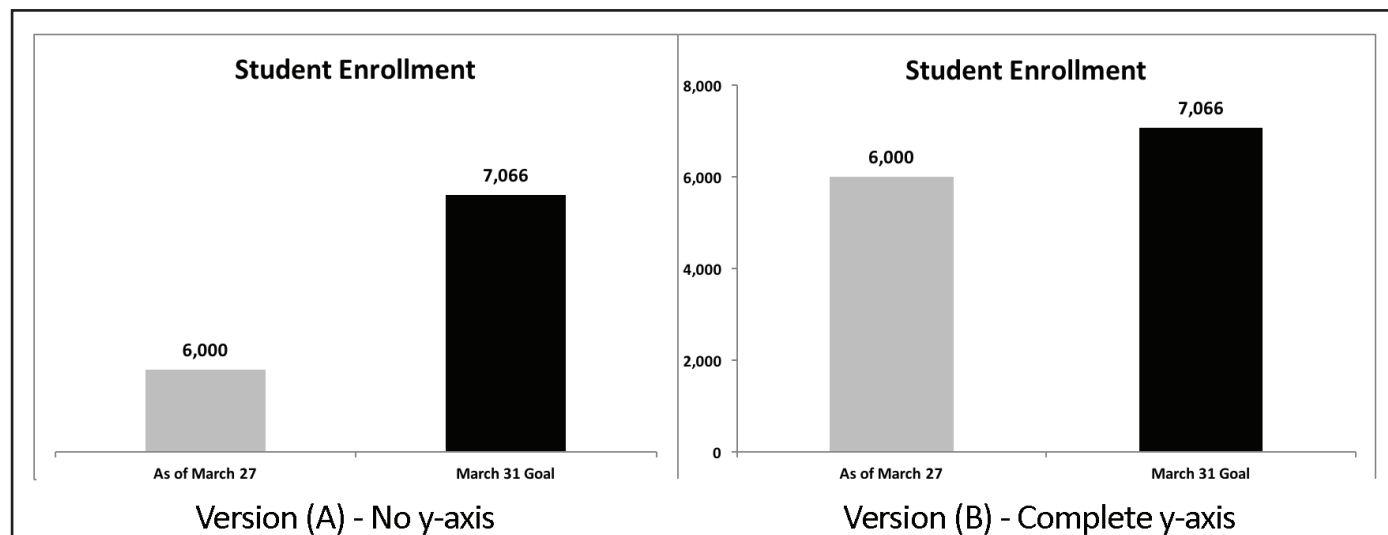
Visualizations such as bar graphs and line plots are frequently seen in the media, but oftentimes viewers must guess what the exact values might be because the salient points are not labeled. While audiences can sometimes make fairly good estimates of data values by looking at the scale on an appropriate axis, many experts believe that the data values that underlie a visualization are important enough to be labeled because showing the data values increases understanding and comprehension among readers (see figure 3.1).

Figure 3.1. Including actual data values is a key principle of data visualization. See below for an example of how otherwise identical bar charts are perceived differently simply by the addition of data values.



A corollary to the key principle of “show the data” is the need to include related information that is necessary to fully understand the data. A legend (or key), the data source, and appropriately scaled axes help a viewer more accurately put data values into perspective. For example, figure 3.2a shows a data visualization that includes data labels but scales the y-axis in a manner that distorts the relationship between the two bars being compared, as is more apparent in figure 3.2b.

Figure 3.2. Although data values are presented in each visualization, the missing y-axis scale in figure A tends to exaggerate or otherwise misrepresent the relationship between the two bars (data values), which differs from the impression given in figure B with a complete y-axis scale.



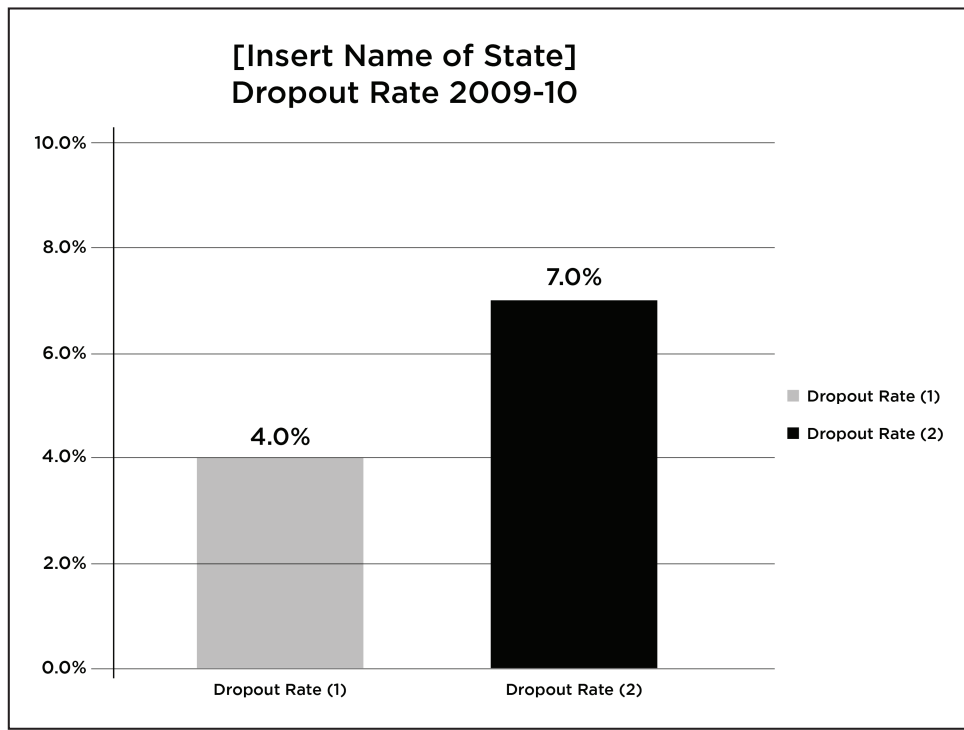
Metadata, or “data about data,” can also help to provide the appropriate context in which to interpret data and information.⁷ Metadata that might be relevant in a data visualization include the data source, data definition, formulas used in calculations (for transparency or replication), collection dates, and other information that could help a viewer more fully understand the meaning of the data. Metadata may be included in the visualization itself or in accompanying text.

In figure 3.3, two types of dropout rates are presented for the 2009-10 school year. Both rates are correct—Dropout Rate (1) is 4.0 percent and Dropout Rate (2) is 7.0 percent—but represent substantially different values. Including metadata that describe how the two rates are defined (and, presumably, why they could both be accurate and yet different) is a critical component of the visualization being understandable to a viewer who would otherwise be unlikely to know the difference between an annual and a cohort dropout rate.

Including metadata in a visualization or accompanying text is a critical component of making information understandable. In figure 3.3, for example, it is integral to a reader knowing that there are two different formulas for calculating a dropout rate.

⁷The National Information Standards Organization (NISO), an association accredited by the American National Standards Institute (ANSI), defines metadata as structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage information. <http://www.niso.org/>. For more information about metadata, see National Forum on Education Statistics. (2009). *Forum Guide to Metadata: The Meaning Behind Education Data* (NFES 2009–805). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Figure 3.3. Why might metadata be helpful?

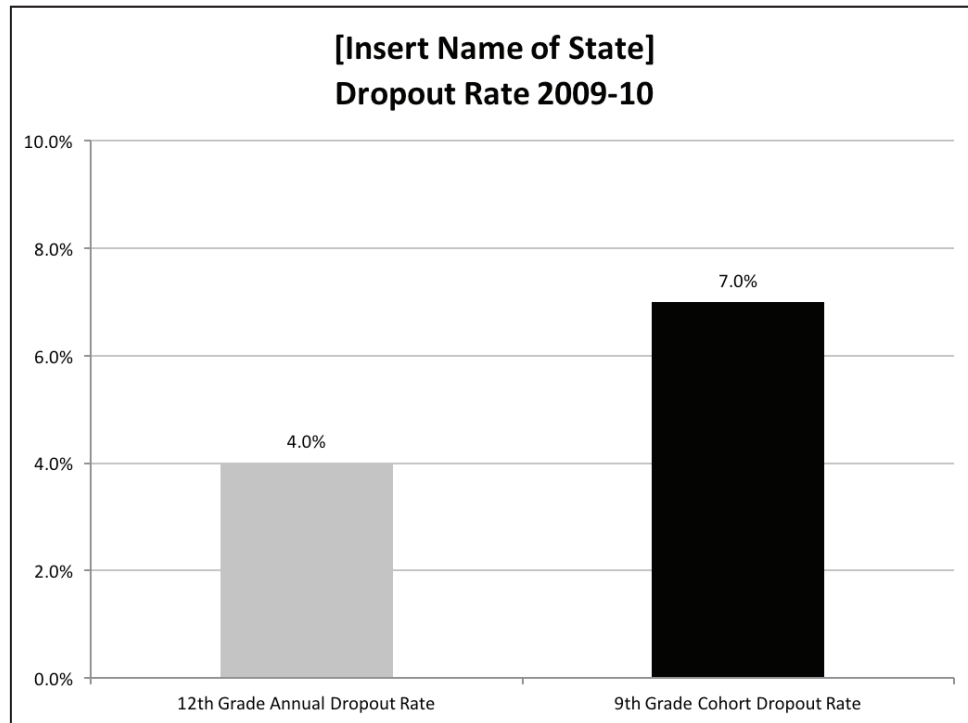


Analysis: The two dropout rates shown are for the same school district, population, and school year. The difference in values reflects variation in definitions and formulas. Dropout Rate (1) is a 12th Grade Annual Dropout Rate, defined as the percentage of students who were enrolled in 12th grade at some time but who did not graduate from high school or complete a state- or district-approved educational program and did not transfer to another public school district, private school, or state- or district-approved educational program (including correctional or health facility programs); have a temporary absence due to suspension or school-excused illness; or die. Dropout Rate (2) applies the same definition to a cohort of students entering 9th grade but dropping out prior to graduation of the cohort (usually 4 or 5 years later). Thus, both bars represent “the dropout rate” in the same school district, population, and school year, but they count different students over different periods of time. See figure 3.4 for enhancements that would further clarify this image and improve the understanding of the data.

Key Principle 2: Reduce the clutter.

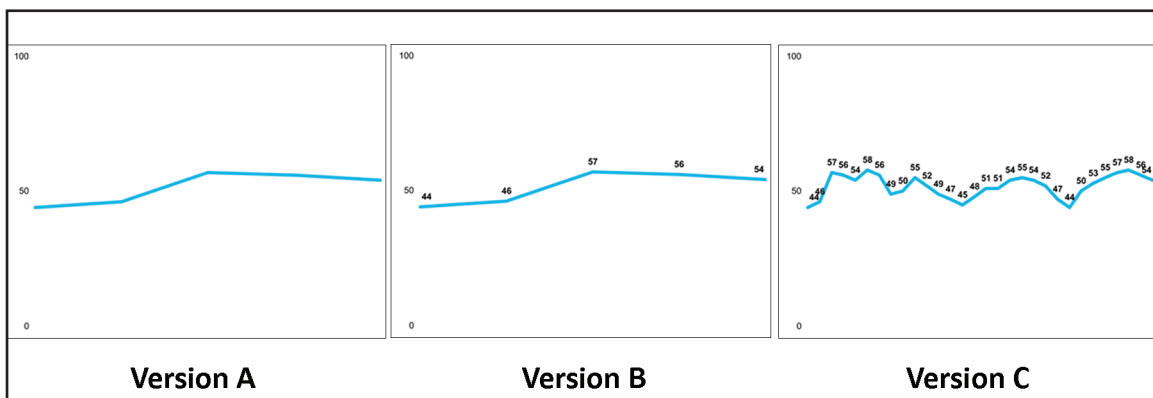
There is a limit to how much information can be displayed without overwhelming a visualization—and, subsequently, a viewer. Rather than being helpful, extraneous information can distract from the primary meaning of the image. Reducing clutter usually requires subjective judgment by the person or team designing the data visualization about whether to show certain components and, if so, how to do so. After all, data values, definitions, sources, and other forms of metadata are often necessary to fully understand the meaning of the data (Key Principle 1), but they must be incorporated judiciously so as to not distract from the primary meaning that the image is intended to convey (Key Principle 2) (see figures 3.4 and 3.5).

Figure 3.4. Contrast this image with the metadata presented in figure 3.3. Simply labeling the columns as 12th Grade Annual Dropout Rate and 9th Grade Cohort Dropout Rate makes it clear to a viewer that two different pieces of information are being presented. Because the data source is visible, a viewer who wishes to learn more about the two rates knows where to find more information.



Source: U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), "State Dropout and Completion Data File", 2009-10 v.1a.

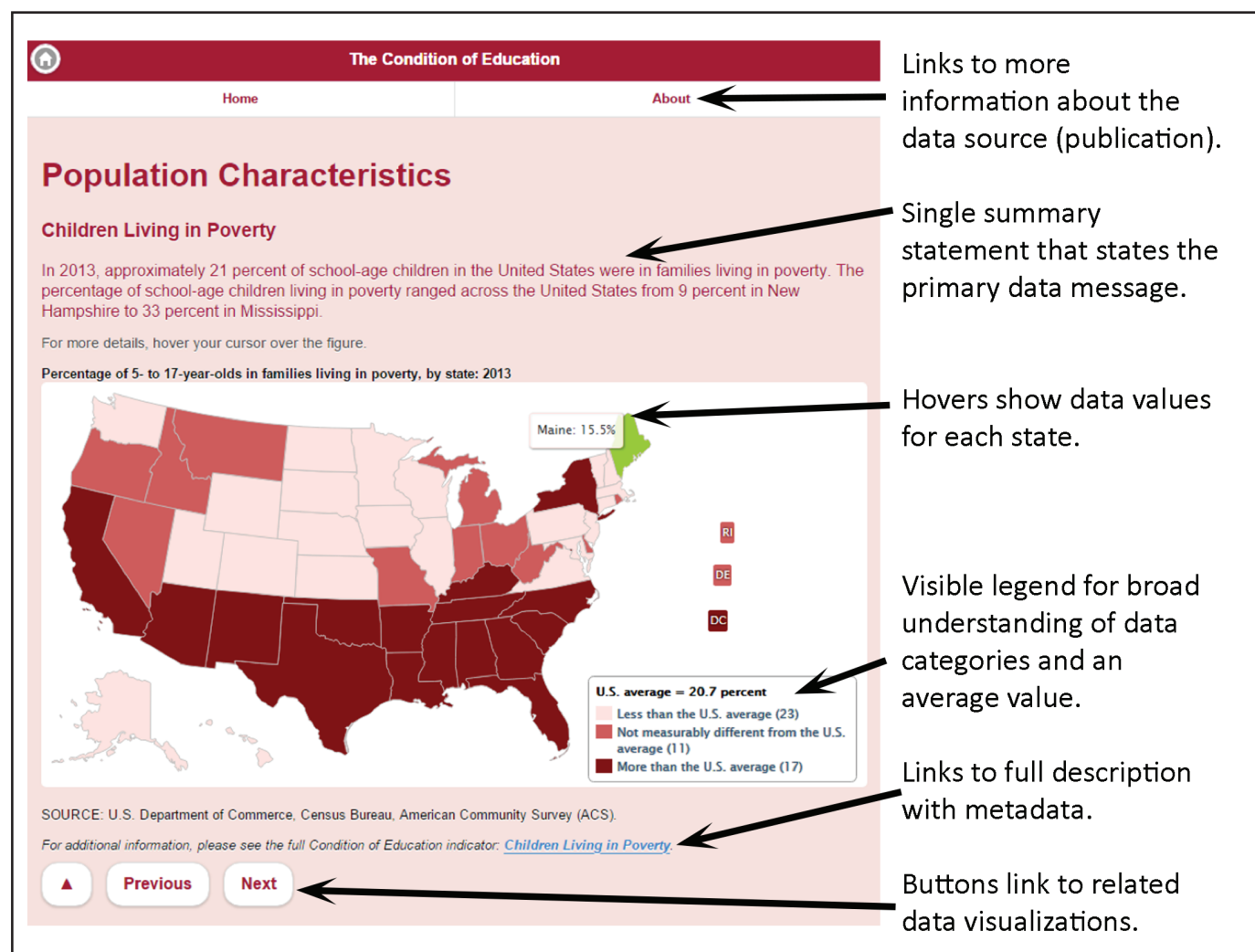
Figure 3.5. Which figure is most effective? There is a balance between presenting useful information in a comprehensive manner and overwhelming a viewer with more stimuli than can be processed. Designers must strike a balance based on the message to be conveyed and the needs of the audience.



Analysis: A line graph is a useful format for presenting time series data (Version A). Note, however, that line graphs convey more meaning when data values are shown (Version B), although too many data points become clutter, are likely to overwhelm a viewer, and should be avoided (Version C).

The media through which a visualization is shared can play an important role in balancing the tension between showing the data (Key Principle 1) and reducing the clutter (Key Principle 2). For example, visualizations presented in digital formats, such as on a website, often display metadata through hovers and links to additional information that do not distract from the primary data message (see figure 3.6). Through the use of such tools, audiences requiring details about statistical methods and other contextual information can easily access it, but viewers who would not find the information to be useful do not need to see or link to it.

Figure 3.6. The digital display of data visualizations, such as on a website, permits the use of hovers, buttons, and links to present critically valuable metadata without distracting from the primary meaning the image is intended to convey.



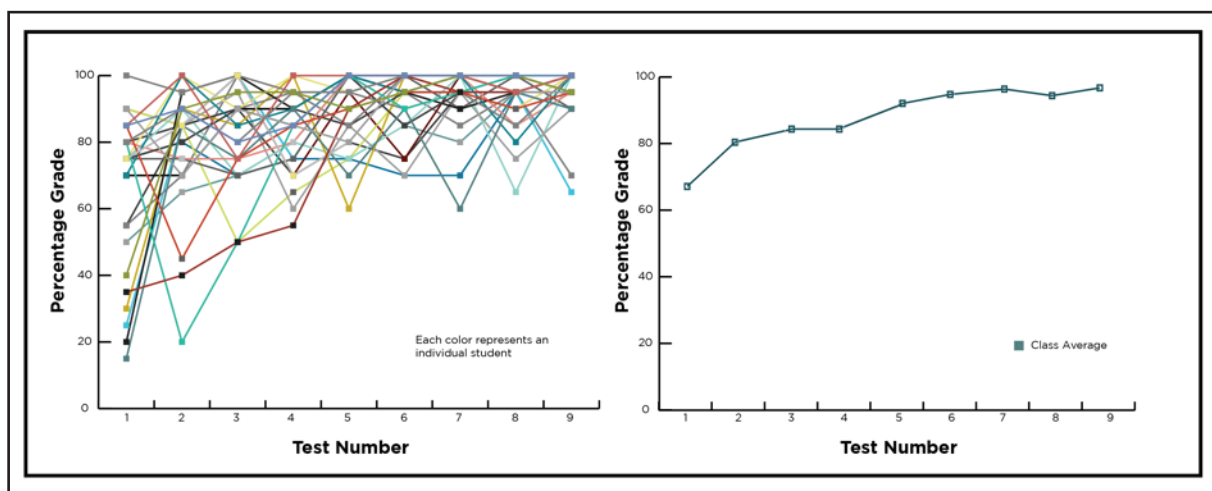
Source: Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., Wilkinson-Flicker, S., Barmer, A., and Dunlop Velez, E. (2015). The Condition of Education 2015 (NCES 2015-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC.

Given that too many features on a graph can be distracting and reduce effectiveness, designers must choose which information to present. Often, it is fairly easy to identify potentially distracting elements because they reflect efforts to accomplish too much with a single visualization (see figure 3.7). To reduce clutter, thoughtful designers should ask questions such as the following:

When trying to reduce clutter in a data visualization, look for distracting elements and other instances in which it appears that the image is trying to do too much.

- What is the primary take-home message that the image is intended to convey?
- How much data are too much to show to your audience in one visualization, without distracting from that primary take-home message?
- Does each feature included in the visualization improve or diminish the likelihood of a viewer understanding the take-home message?
- What features and information must be in the visualization?
- What can be taken out?
- Have one or more representatives of the target audience confirmed that the visualization conveys what it is intended to say?

Figure 3.7. Designers should ask themselves how much information becomes too much and begins to distract from the take-home message that a visualization is intended to convey to an audience.



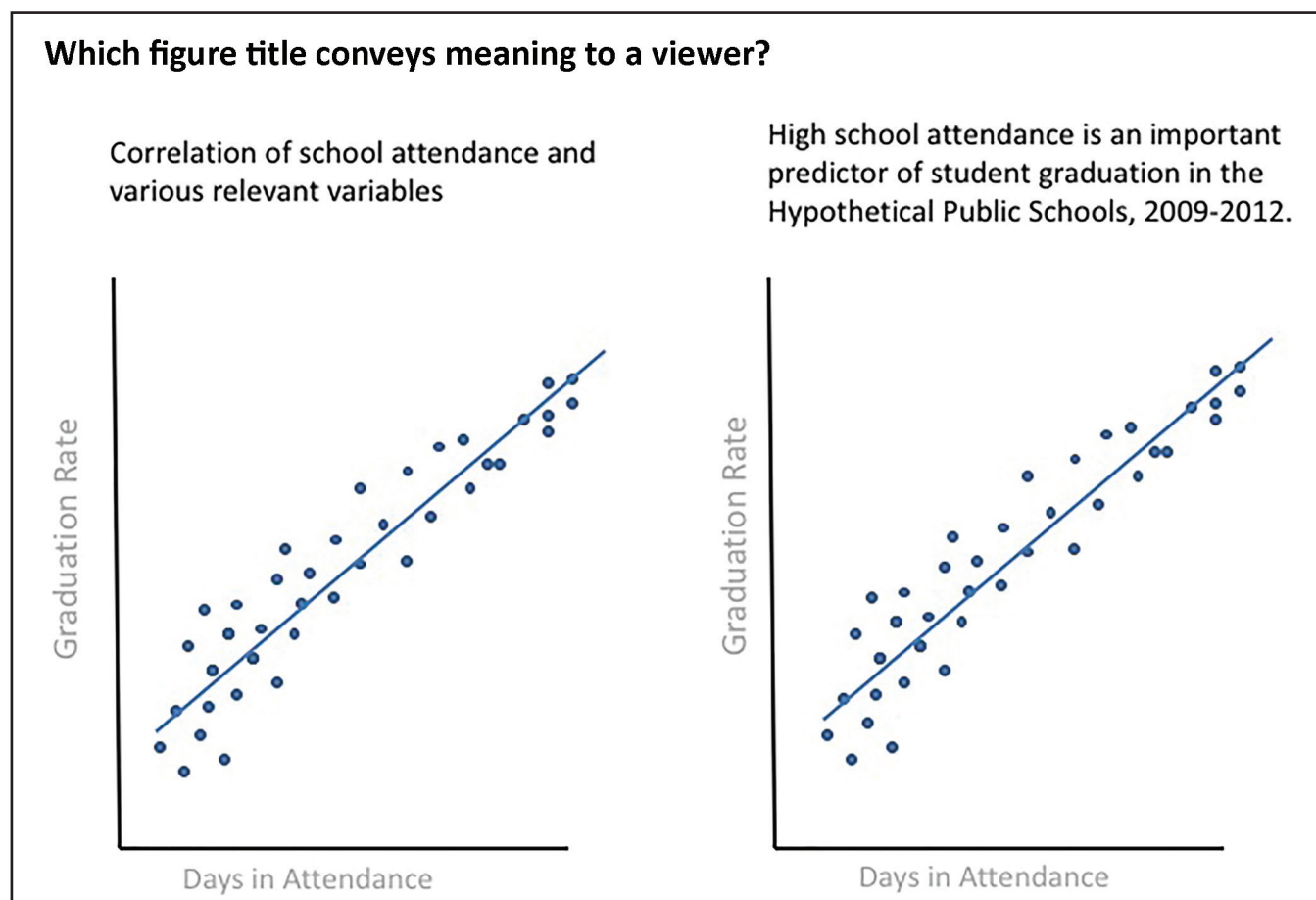
Analysis: The meaning of the data can become unclear when a data visualization tries to accomplish too much—for example, using a different color for each individual student in a dataset. While the graph may accurately reflect the data, reducing the clutter by presenting the class average, rather than numerous individual grades, would likely improve understanding for most audiences.

Key Principle 3: Integrate text and images.

If a data visualization is intended to convey data meaning to a viewer, it cannot suffer from a sideshow effect in which the image does not appear to be connected to the text in a relevant manner. Images and related text, whether in an attached report, on a web page, or integrated into the visualization, should clearly connect and reinforce each other to enhance understanding.

While a visualization should complement related text, it should also be able to stand on its own as a complete piece of information. Including legends and captions, which are critical for defining and explaining an image, improves clarity. The wise use of descriptive figure titles, variable (data) names, captions, and callout boxes contributes to the effective communication of the take-home message for a viewer (see figure 3.8).

Figure 3.8. Every aspect of imagery and text, including figure titles and captions, should point viewers toward a better understanding of the primary take-home message of the visualization.



Analysis: The figure title to the left does not convey data meaning. In contrast, the figure title to the right states the take-home message in plain language so that a viewer would understand the meaning of the data even if the values could not be seen.



Key Principle 4: Portray data meaning accurately and ethically.

The correct presentation of data is a fundamental aspect of data visualization, but even data that are technically accurate can be presented unethically. Common methods for intentionally misleading viewers and introducing a bias include

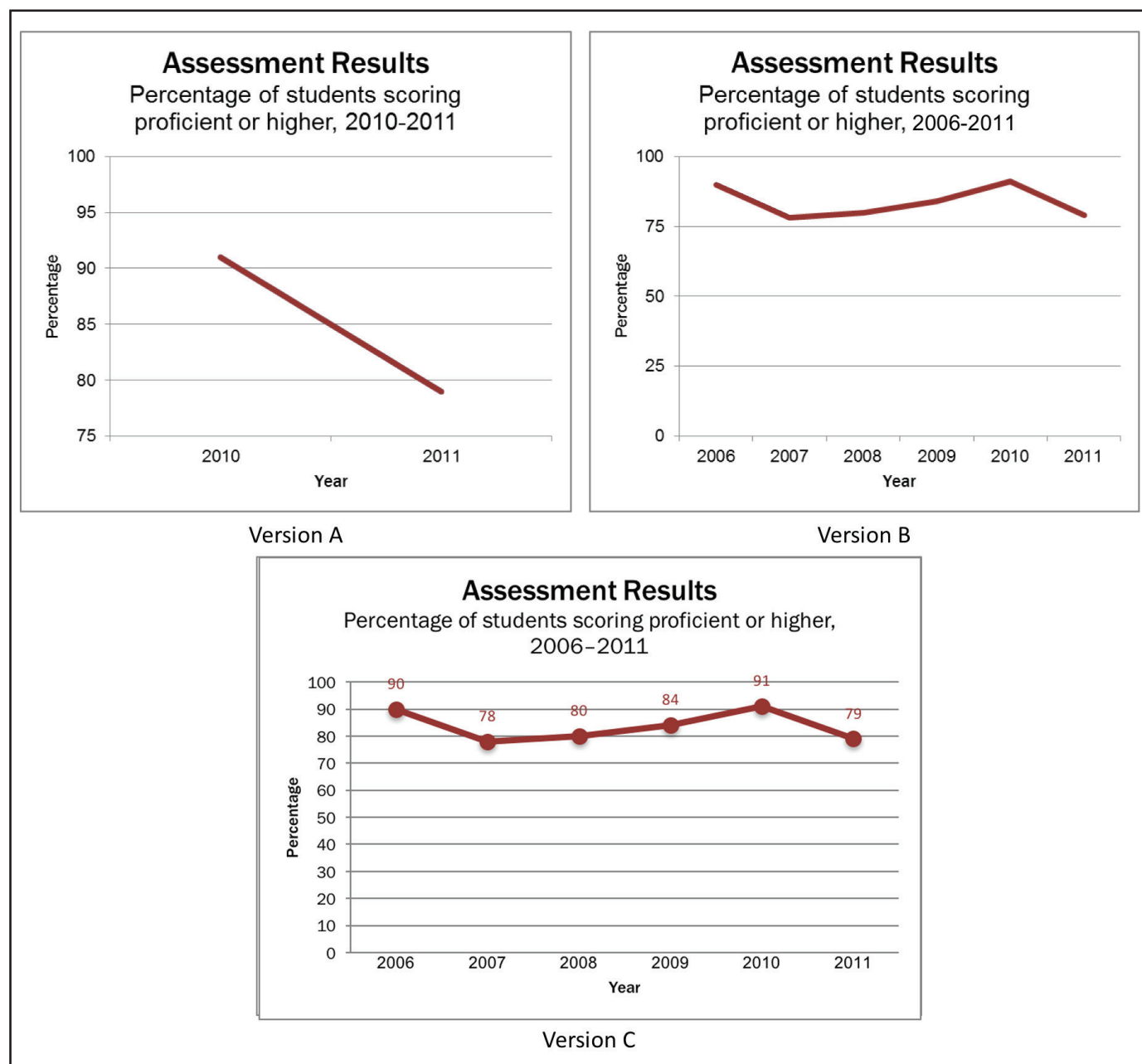
- limiting which data are seen (e.g., overemphasizing specific subsets of data or patterns in the data by only showing parts of one or both axes—see also cherry-picking data below);
- manipulating how the data are presented visually (e.g., suggesting that certain types of data are continuous over time rather than discrete across time to suggest relationships that are not valid); and
- using language that suggests a conclusion that is not substantiated by the data (e.g., referring to a trend or pattern that does not fully describe a variable) (figure 3.9).

Cherry-picking, or otherwise selecting only data points that support a particular point of view (sometimes referred to as a bias), is another form of unethical presentation, even if what is said is technically accurate (but perhaps not fully so). For example, suppose that a local official declared that a school district no longer had a student behavior and discipline problem because the data showed that only 1.5 percent of students had received out-of-school suspension in the last year. Such a statement might be a cause for celebration, but would the cheers be as loud if the audience knew that in-school suspensions had doubled during that same period because of an administrative decision to substitute in-school suspensions for behavior that, in the past, would have warranted out-of-school suspensions? Similarly, what if certain demographic subgroups had out-of-school suspension rates of 8 and 10 percent, but those populations were so low in number that the school-wide average was only 1.5 percent? Is it still fair to suggest that the district no longer had a discipline problem?

Thus, the accuracy of a data visualization is determined in its totality—with the overall message, as well as the specific data values, accurately reflecting data meaning without introducing bias or prejudicial slant. Effective data visualizations are intentionally and proactively designed to minimize the likelihood of any such foreseeable misinterpretation or misuse.

Ethics play a vital role in ensuring that data visualization does not inadvertently or intentionally introduce bias or otherwise misrepresent data meaning.

Figure 3.9. Data can be presented accurately in the technical sense of the word, but still be misleading.



Analysis: All three visualizations are technically accurate with respect to the data they present. However, version A suggests a significant downward trend through the manipulation of the x axis (showing only two years of data) and the y-axis (shortening the scale from 0-100 to 75-100 percent, which accentuates the slope of the trend line). In contrast, version B shows the full range of the x- and y-axes for the same data—showing a much less dramatic trend. Moreover, the presentation of separate testing dates in versions A and B as a line suggests continuous change over time that is not a fully accurate representation of discrete data points. Version C is more likely to convey a complete picture, which shows some variation in discrete data (represented by symbols and data values), but no signal of a downward trend over successive years of testing.

Recommended Practices for Data Visualization

In the developing field of data visualization, knowledge is being uncovered and expertise is advancing on an ongoing basis. New approaches, techniques, and models are being shared about the comprehensive strategies and detailed nuances of imparting the meaning of data in a visual way. Within this continuously improving discipline, at least seven generally agreed upon recommended practices are recognized to meet many communications needs in education agencies.

Recommendation 1: Capitalize on consistency.

Recommendation 2: Data that should not be compared should not be presented side by side.

Recommendation 3: Don't limit your design choices to default graphing programs.

Recommendation 4: Focus on the take-home message for the target audience.

Recommendation 5: Minimize jargon, acronyms, and technical terms.

Recommendation 6: Choose a font that is easy to read and will reproduce well.

Recommendation 7: Recognize the importance of color and the benefits of Section 508 compliance.

Recommendation 1: Capitalize on consistency.

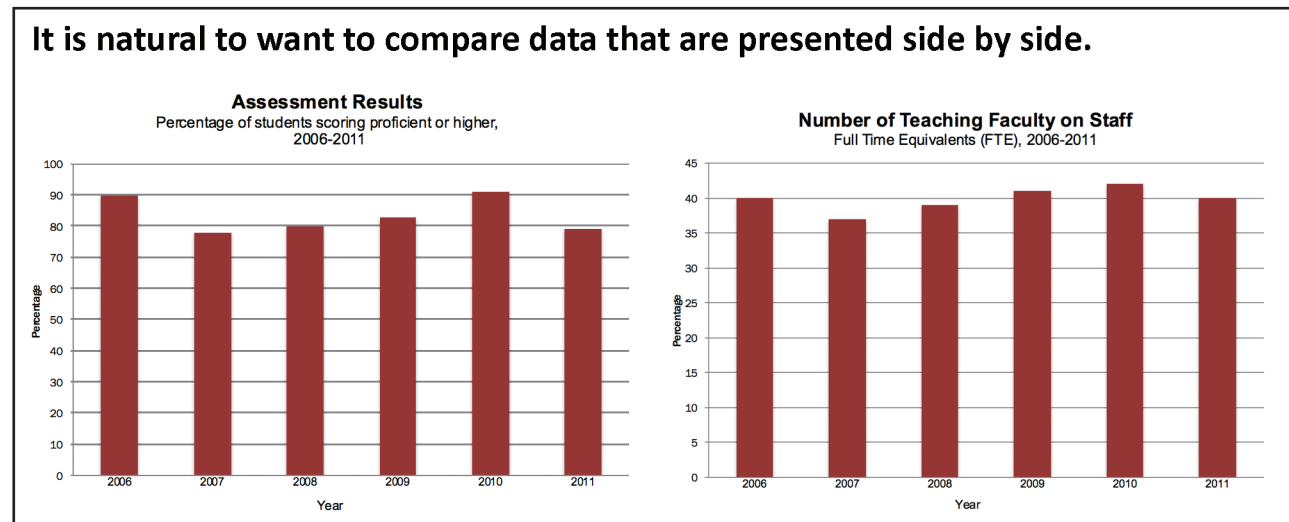
People tend to process information more quickly when it is presented in a familiar manner, so approaches to data visualization that are consistent over time help to establish and reinforce audience expectations for how data are viewed. For example, if there is a standard way for your organization to display visualized data, such as when trends over time are presented as line graphs with time on the x-axis and the variable of interest on the y-axis, readers will become accustomed to this format and may eventually become adept at interpreting the meaning of data in these types of visualizations. Following such widely recognized standards and conventions means that once a viewer understands how to read one data visualization, he or she will be better prepared to understand other data presented in a similar format. Consistent presentation is especially important for education agencies that expect to report the same types of data (for example, performance, attendance, and financial data) to similar audiences over time. Similarly, figures that are intended to be compared should be presented with consistent scaling, formatting, and related presentation choices so that visual differences are readily observable and substantive rather than aesthetic in nature. Doing so streamlines the production process and permits easier comparisons within and across organizations and jurisdictions.

The recommendations shared in this chapter are important, but are only effective when the four key principles described above are adhered to as standard practice.

Recommendation 2: Data that should not be compared should not be presented side by side.

It is not advisable to place two (or more) datasets in the same graph, table, figure, or other context, unless the intent of the data visualization is for the viewer to note the similarities and differences between the datasets. Displaying data in a single image encourages comparison, which is appropriate if that is the goal of the visualization. If, however, datasets are not intended to be compared, side-by-side presentation may promote misuse or misapplication (see figure 3.10). For example, if an agency's end-of-year assessment changes, displaying performance data from the old assessment next to the new assessment on the same x-axis will likely result in some readers viewing it as if it were trend analysis, regardless of warnings, caveats, and cautions against doing so in the text or footnotes of the image.

Figure 3.10. Presenting data side by side encourages a reader to associate and even compare data regardless of whether there is an actual relationship between the datasets.



Analysis: Although common sense suggests that the number of FTE teaching staff in a school may influence the academic performance of students, the relationship between the two sets of data is complex and is, in fact, affected by a host of other variables in the school setting, including, for example, staff and student characteristics, grade level, courses taught and attended, and subject matter expertise. Presenting the two datasets side by side may encourage a reader to associate student performance and FTE counts without regard for those other factors that have been shown to have a significant influence on how to interpret these data.

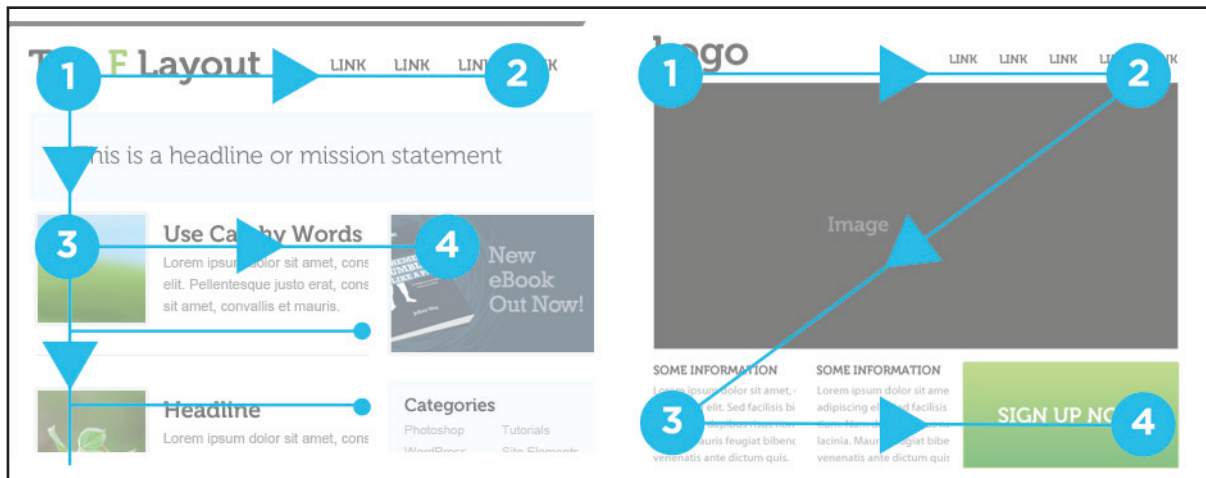
Recommendation 3: Don't limit your design choices to default graphing programs.

Many common data and statistical applications have a default mode for graphing that, to more and lesser degrees, is designed to be generally appropriate for a range of data across multiple purposes. In other words, while an application may generate a visual representation of the data, the product is not likely to be a thoroughly considered, carefully constructed, customized visualization that meets the specific information needs of a particular target audience. Moreover, the default mode of many applications introduces unnecessary clutter, such as a different symbol and color for every dataset, that is not appropriate for many types of education stakeholders. While these software products may have tools that can be used to present data or customize visual design, doing so requires that the user advance beyond the default production mode.

Recommendation 4: Focus on the take-home message for the target audience.

Keep the most important information at the forefront of the visualization in order to focus on the primary message. Emphasize important text through the wise application of font selection (see below) and surround that text with white space so that it stands out—noting that many designers employ F and Z patterns to help them locate the most important parts of the message in positions of prominence (see figure 3.11).

Figure 3.11. Many designers rely on the natural patterns of how a person reads a page, scanning in predictable “F” and “Z” patterns to glean information quickly and efficiently.



Analysis: Placing the most important information in these prominent positions increases the likelihood that it will be noticed given that most people read or skim content in an orderly progression from points 1-4 in both an “F” pattern (left) for text and a “Z” pattern (right) for webpage content. Adapted from <http://webdesign.tutsplus.com/articles/understanding-the-f-layout-in-web-design-webdesign-687>.

Recommendation 5: Minimize jargon, acronyms, and technical terms.

The language used in a visualization should explain concepts, clarify data meaning, and enhance understanding. Thus, for most audiences, text should not include jargon, lingo, technical terms, acronyms, or other words that are not commonly understood by non-expert audiences. For example, it is not realistic to expect many target audiences to know how to interpret terms such as “disaggregated,” “performance index,” “FTE,” “CRT,” “chi-square,” “p-value,” or “coefficient.” Efforts to simplify language can be as simple as spelling out words rather than using acronyms and providing easy access to the definitions of terms.

Recommendation 6: Choose a font that is easy to read and will reproduce well.

A good rule of thumb is that any font that looks “fancy” or that is selected for “style” rather than simplicity should be examined closely for readability. Multiple fonts can sometimes make sense as a tool for engaging a viewer or emphasizing a point, but too many different fonts tends to distract from a message. Some experts suggest three as a reasonable limit (Evergreen 2014). Font styles, such as boldface and italics, should be avoided except in headings or to highlight specific words or phrases. Similarly, underlined text should not be used unless it is intended to signal to a viewer that they should “click here” to link to other material.

Recommendation 7: Recognize the importance of color and the benefits of Section 508 compliance.

An important aspect of Section 508 compliance is the use of color. While colors can be engaging and support messaging within a visualization, they can also present challenges. In all likelihood, a substantial portion of your audience prints reports only in black and white. Making images distinguishable on the basis of contrast rather than color ensures that your graphics are distinguishable in gray scale and in black and white. Moreover, some readers have physical disabilities that limit the recognition of colors. Being insensitive to, for example, red/green color blindness limits the accessibility of your visualizations to large segments of your readers. When your organization complies with Section 508 guidance, its

electronic products will be accessible to viewers regardless of disability—and be easier to copy or otherwise exchange across media; reflect a consistent look and feel; transition more readily to new platforms (e.g., handhelds); enhance usability for all stakeholders; and reflect proactive data governance within the organization.⁸

Section 508 refers to a federal law (29 U.S.C. § 794 (d)) requiring federal agencies, as well as organizations receiving federal funds, to make their electronic and information technology accessible to people with disabilities. Comparable practices are required by state and local policies, regulations, and laws in many jurisdictions.

Summary

Four key principles and seven recommended practices for effective data visualization are presented in this chapter. Note that the nature of customizing visualizations to meet the specific information needs of particular audience types means that subjective decisionmaking is still an important part of the development process.

Four Key Principles for Effective Data Visualization

Key Principle 1: Show the data.

Key Principle 2: Reduce the clutter.

Key Principle 3: Integrate text and images.

Key Principle 4: Portray data meaning accurately and ethically.

Seven Recommended Practices for Data Visualization

Recommendation 1: Capitalize on consistency.

Recommendation 2: Data that should not be compared should not be presented side by side.

Recommendation 3: Don't limit your design choices to default graphing programs.

Recommendation 4: Focus on the take-home message for the target audience.

Recommendation 5: Minimize jargon, acronyms, and technical terms.

Recommendation 6: Choose a font that is easy to read and will reproduce well.

Recommendation 7: Recognize the importance of color and the benefits of Section 508 compliance.

Effective data visualizations are

Appropriate (for the intended audience)

Accurate (in the presentation of data meaning)

Actionable (because the information is useful)

⁸ For more information about the application of Section 508 Accessibility Guidelines in the education community, see: National Forum on Education Statistics. (2011). *Forum Guide to Ensuring Equal Access to Education Websites: An Introduction to Electronic Information Accessibility Standards* (NFES 2011–807). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Available at http://nces.ed.gov/forum/pub_2011807.asp.

Chapter 4: Implementing the Data Visualization Process

Improving how your organization implements the data visualization process requires both expertise and effort. The effective presentation of data is well worth this commitment on the part of agency staff when you consider the many ways in which stakeholders will use these data to make critical decisions about the future of the education enterprise. Accurate and effective presentation choices will help to improve decisions about students' educational programs; assessments of school quality and satisfaction; and legislative action that will direct school funding, instructional options, administrative policies, and school management practices.

This chapter is organized around figure 4.1, which illustrates the data visualization process, from recognizing a question or information need through reviewing and refining the data visualization process. Interim steps incorporate efforts to find the right data to address the issue, analyze those data to determine their meaning, customize the take-home message to meet audience needs, and present a visualization that is accurate and unambiguous (so that it does not mistakenly encourage misunderstanding or misapplication of the data).


Data visualization should not be viewed as a product; instead, it should be approached as a process through which data are transformed into meaningful information—in this context, a message that is visually understandable for specific data audiences.

Figure 4.1. The six steps to effective data visualization described in this chapter.

Six-Step Process for Data Visualization

- Step 1. Question: Someone Needs Information
- Step 2. Research: Data Exploration and Analysis
- Step 3. Findings: Data Meaning/Answer
- Step 4. Customization: Audience-Specific Messaging
- Step 5. Visualization: Present Data Meaning Clearly and Accurately
- Step 6. User Feedback: Review and Refine Efforts

Analysis: Note that skipping steps and other shortcuts introduce risk to the production process and increase the likelihood that a visualization will not meet the information needs of the intended audience. When this occurs, viewers are more likely to misunderstand or misapply the data in their decisionmaking.



The final product of the data visualization process is the transformation of data into meaningful information that educates and informs a specific audience. In other words, the data that otherwise might confuse or be misunderstood by a viewer will become more likely to accurately convey data meaning and, in turn, influence practice, inform policy, increase learning, and improve research. The following section presents each step as a distinct, but related, part of the data visualization process.

Step 1. Question: Someone Needs Information

In local, state, and federal education agencies, “information needs” encompass a wide range of public reporting activities. In some cases, reporting is mandated or required, such as annual surveys or accountability collections to a state department of education or the U.S. Department of Education. Other reporting activities may be aimed at the general public, such as a school district’s annual report to its taxpaying community.

What is the question?

Is this a one-time information need or a routine data request that will likely be repeated?

Questions can come from a variety of stakeholders with an interest in education. This includes public policymakers, school policymakers, school administrators, parents, students, community members, advocacy organizations, and researchers. Note that even when these audiences request information about the same issue, their different perspectives sometimes create the need for similar, but not identical, data. For example, if the community is generally concerned about students dropping out of high school, a question from a parent might center on aggregate dropout rates in their school, while an inquiry from the principal might focus on data in student-level warning systems. Similarly, the school superintendent could request data about resources available to students at risk of dropping out and a school board member might look into data about the success rates of dropout prevention programs across the state. Each of these stakeholders is concerned about the same issue, but their information needs vary based on their unique perspectives.

Step 2. Research: Data Exploration and Analysis

In order to respond to a question or information need, you must determine whether the data needed to address the issue are available and, if so, how to access them. While it may be tempting to find any data that might tangentially meet the information need, experienced data staff understand that the quality of the data will determine the accuracy of analysis and, therefore, the findings (message), decisions, and actions that will be made based on the information.

What data and analysis are needed?

Is high-quality data available for exploration and analysis?

Education data collected and maintained by schools, districts, states, and the U.S. Department of Education do not represent the universe of data available for answering questions. Workforce data, geographic data, demographic data, and economic data are just a few examples of the many types of information from other industries that may be collected reliably and adapted responsibly to address important issues in education. When working with another agency to acquire data needed to answer the question or satisfy the information need, both agencies will wish to establish binding agreements that permit the legal exchange of the data. Such agreements often specify narrow conditions regarding the use or publication of the data, especially if there is any reason to think that personally identifiable information (PII) or sensitive data are involved.

If the data required to answer the question are not available, you may choose to develop a new data collection method. If doing so or otherwise locating relevant, high-quality data is deemed to be unmanageable, you may need to revise your plan to answer the question or acknowledge that appropriate data are not available.

Once the data are in hand, analysis generally focuses on identifying patterns in the data that are relevant to the research question. This analysis frequently is iterative in nature, with each step in the process requiring reconsideration and modification as the understanding of the data unfolds—sometimes leading to new or modified searches for additional relevant data. During data exploration and analysis, the analyst will need to have knowledge of basic research and analytical methods, such as which statistical approaches or analytical assumptions work best with a particular type of data. As analysis advances, it is also necessary to thoroughly understand the broader context in which the data have been collected and reported. For example, do the data reflect different groups of students over time, the same cohort of students over time, or different time periods completely? The agency’s data stewards, research staff, and program experts should be consulted as necessary to confirm the validity of all data analysis and interpretation.

For more information about data quality, see:

- *Forum Guide to Building a Culture of Quality Data: A School & District Resource*
http://nces.ed.gov/forum/pub_2005801.asp
- *Forum Curriculum for Improving Education Data: A Resource for Local Education Agencies*
http://nces.ed.gov/forum/pub_2007808.asp

Step 3. Findings: Data Meaning/Answer

Once the question or information need has been defined and you have determined that relevant high-quality data are available to answer that question or satisfy that information need, determine what the data tell you about the issue—that is, the most relevant values or patterns in the data that answer your question. Data analysts and other researchers apply a wide range of analytical practices to generate sound, reliable interpretations of data values, trends, features and, ultimately, meaning. Sometimes relatively “simple” methods are sufficient (e.g., plotting purely descriptive data over time) whereas in other instances more complex analysis is necessary (e.g., advanced statistical methods). When high-quality data are subjected to sound analytical practices, the product is a valid conclusion that reflects the meaning of the data as accurately as possible—sometimes referred to as the “take-home message” that needs to be conveyed to an audience.

What is the take-home message from the data?

That is, what is the core message in the data that answers the question or addresses the information need?

At an organizational level, you must determine who has responsibility for determining data meaning (i.e., analyzing the data accurately, consistently, and without bias). Data stewards are often the most qualified staff members for analyzing data and determining data meaning, but other staff may also need to be involved depending on the nature of the research question and the manner in which decisions are made in your organization. For example, some questions are politically sensitive and senior leadership may have strong opinions regarding who is authorized to determine the agency’s message. Although the data value may not change, decisionmakers may decide that the messenger should be the state department of education or the school board president rather than a building principal (or vice versa). Note that such decisions should not bias interpretation—rather, they simply acknowledge that many factors contribute to how data can best be shared with stakeholders.

Regardless of the message and where it originates, data staff should engage in quality reviews in all visualization activities. Because data staff are most familiar with the data, they are likely to be the best qualified to evaluate data accuracy and integrity throughout the visualization process. These data experts are also likely to realize when additional explanation and complementary data are needed to more completely present data meaning.

Step 4. Customization: Audience-Specific Messaging

Given that you have now identified the take-home message from the data, it is time to consider who needs to hear it: your target audience. While identifying your target audience's specific information needs depends heavily on the question and message, having a clear sense of the audience's specific needs, expectations, and capabilities is critical for each visualization. For example, a message intended for the parents of non-English-speaking students are likely to benefit from translation into languages they are able to speak and read.

Who is your audience?

To whom is the message being conveyed? What is the most appropriate way to communicate with this audience?

Researching the needs of common types of audiences is a good practice. Frequently targeted audiences for education agencies might include parents, students, community members, instructional staff, administrators, and policymakers. What do you know about them? If you are routinely messaging information to them, isn't it worth the effort to learn more about their particular needs? It might become advantageous to conduct surveys or focus groups to gather information about common audiences in order to better customize your messages.

For each of your target audiences, determine the following types of profiles:

- How would you characterize their ability to understand data meaning?
- What, if any, technical expertise can you assume about them?
- How much accompanying explanation is appropriate for the audience?
- Would some audience members benefit from data visualizations presented in another language?
- What other information might the audience need in order to understand the message?
- Are there particular media that are more or less likely to be accessible by that audience (e.g., posting information to a website may not help an audience that doesn't have Internet service in their home)?

No Training Required

Data visualizations for general audiences should be designed so that they do not require training for viewers to understand the take-home message.

Step 5. Visualization: Present Data Meaning Clearly and Accurately

Once data, message, and audience have been considered, it is time to determine how that message can be presented in a way that is appropriate, accurate, and actionable for the audience who will see it. As described in chapter 3, the fundamental principles of visually presenting data messages are: (1) show the data; (2) reduce the clutter; (3) integrate text and images; and (4) portray data meaning accurately and ethically. The implementation of these principles is adjusted depending on the nature of the message and audience. Is the visualization intended to inform parents as they make choices about program participation? To inform policymaking by local or state decisionmakers? To improve research? Each of these different uses affects how the data might be communicated. Sometimes these differences are dramatic and other times they are nuanced—but in most cases they should be addressed in the visualization. For example, an image designed for school administrators may not need details about the location of each school building (as might be appropriate for community members) given that this is where principals go to work each day.

How will you present your message?

That is, what is the most effective way to visualize the data for your audience?

It is critical to note that, regardless of message, it may not be appropriate to share some types of data with all audiences. For example, the visualization of assessment results intended for teacher use (such as in a data dashboard) is likely to include details about individual student performance. This information is never appropriate for other audience types (such as community members) who do not have a verifiable “need to know” as defined by federal and state privacy laws.

The Family Educational Rights and Privacy Act (FERPA) is a federal law that requires parental consent for the release of individual student data (with some exceptions). For more information about FERPA, visit <http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>.

Finally, it is a sound communications practice to accommodate the physical limitations of disabled viewers. Otherwise, the meaning of your data may not reach all members of your community. For example, graphics printed in red and green may not be visible to those with color blindness. As such, compliance with federal Section 508 Accessibility Guidelines is not just encouraged—it and comparable state and local regulations are required by law.⁹

Step 6. User Feedback: Review and Refine Efforts

Once you have customized your visualization for the target audience, there is only one proven way of confirming that you were successful: ask your audience for feedback! Establish relationships with representatives of common target audiences and ask them for formative feedback (that is, feedback you can use to improve a visualization while it is still under development) as well as summative feedback (that is, feedback on the final visualization that you can use to improve future efforts). If a significant number of parents agree that a visualization is clear, you can be confident in its use. If, however, they don’t understand the data, meaning, or message, you can be certain that others will also find the visualization to be ineffective. In cases in which it is not appropriate to survey representative audiences, colleagues who understand the needs of potential viewers can serve as proxy reviewers to assess clarity, readability, and applicability.

How can you ensure that your visualization is effective?

Ask your users for feedback and iterate, iterate, iterate based on that feedback.

Applying the Data Visualization Process to a Real-World Example

The remainder of this chapter focuses on the process of applying data visualization practices to answer a relatively common question about the education system at local, state, and national levels. Developing a data visualization to answer such a question can be accomplished through the six-step process described in this chapter and the key principles and recommended practices presented in chapter 3 and shown below.

Six-Step Process for Data Visualization

- Step 1. Question: Someone Needs Information
- Step 2. Research: Data Exploration and Analysis
- Step 3. Findings: Data Meaning/ Answer
- Step 4. Customization: Audience-Specific Messaging
- Step 5. Visualization: Present Data Meaning Clearly and Accurately
- Step 6. User Feedback: Review and Refine Efforts

⁹ For more information about the application of Section 508 Accessibility Guidelines, see: National Forum on Education Statistics. (2011). *Forum Guide to Ensuring Equal Access to Education Websites: An Introduction to Electronic Information Accessibility Standards* (NFES 2011–807). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Available at http://nces.ed.gov/forum/pub_2011807.asp.



Four Key Principles for Effective Data Visualization

Key Principle 1: Show the data.

Key Principle 2: Reduce the clutter.

Key Principle 3: Integrate text and images.

Key Principle 4: Portray data meaning accurately and ethically.

Seven Recommended Practices for Data Visualization

Recommendation 1: Capitalize on consistency.

Recommendation 2: Data that should not be compared should not be presented side by side.

Recommendation 3: Don't limit your design choices to default graphing programs.

Recommendation 4: Focus on the take-home message for the target audience.

Recommendation 5: Minimize jargon, acronyms, and technical terms.

Recommendation 6: Choose a font that is easy to read and will reproduce well.

Recommendation 7: Recognize the importance of color and the benefits of Section 508 compliance.

The remainder of this chapter focuses on the process of applying data visualization practices to answer a relatively common question about the education system.

Step 1. Question: Someone Needs Information

Many education stakeholders request information about high school graduation rates in an effort to evaluate whether schools are effective in one of their core missions: graduating students. Questions about high school graduation can range from inquiries about a specific school to broader assessments of how efforts to graduate students across the state compare to peer states. For example, a question from a state policymaker, administrator, or education advocacy organization might be:

How does our state's high school graduation rate compare to other states' high school graduation rates?

Step 2. Research: Data Exploration and Analysis

Your state knows its high school graduation rates because it collects these data from every school district on an annual basis or calculates the rates from its own statewide longitudinal data system that has been populated by school district data.¹⁰ But the question in this example requires the comparison of your state's graduation rates to those of other states in the nation. Fortunately, your state education agency submits graduation data to the National Center for Education Statistics (NCES)¹¹ each year and knows that other states are expected to do the same.

A visit to the NCES website at <http://nces.ed.gov> reveals the availability of the U.S. Department of Education's *EDFacts* Consolidated State Performance Report, which includes public high school 4-year adjusted cohort graduation rate (ACGR) for the United States, the 50 states and the District of Columbia: School years 2010-11 to 2012-13 (see table 4.1).¹²

Although there is a range of valuable data visualization products available to education staff, all of the figures presented in this document were constructed in a basic spreadsheet application (MS Excel 2010) to illustrate the importance of designer decisionmaking (over technical tools) in the data visualization process.

¹⁰ Visit <http://nces.ed.gov/programs/SLDS/> for more information about the development and use of statewide longitudinal data systems in state education agencies across the nation.

¹¹ The National Center for Education Statistics (NCES) is part of the U.S. Department of Education and the Institute of Education Sciences, and is the primary federal entity for collecting and analyzing data related to education in the U.S. and other nations. Visit <http://nces.ed.gov/> for more information about NCES.

¹² *EDFacts* is a U. S. Department of Education initiative to put performance data at the center of policy, management and budget decisions for all K-12 educational programs. Visit <http://www2.ed.gov/about/inputs/ed/edfacts/index.html> for more information about *EDFacts* and its data collection activities. The Consolidated State Performance Report is available at <http://www2.ed.gov/admins/lead/account/consolidated/index.html>. The school 4-year adjusted cohort graduation rate (ACGR) data for 2010-11 to 2012-13 are available at http://nces.ed.gov/ipeds/data/ACGR_2010-11_to_2012-13.asp. All data used in this example were downloaded in January 2016.

Table 4.1. Public high school 4-year adjusted cohort graduation rate (ACGR) for the United States, the 50 states and the District of Columbia: School years 2010-11 to 2012-13.

State	Adjusted Cohort Graduation Rate		
	2010-11	2011-12	2012-13
United States ¹	79	80	81
Alabama	72	75	80
Alaska	68	70	72
Arizona	78	76	75
Arkansas	81	84	85
California ²	76	79	80
Colorado	74	75	77
Connecticut	83	85	86
Delaware	78	80	80
District of Columbia	59	59	62
Florida	71	75	76
Georgia	67	70	72
Hawaii ²	80	81	82
Idaho ³	—	—	—
Illinois	84	82	83
Indiana	86	86	87
Iowa	88	89	90
Kansas	83	85	86
Kentucky ³	—	—	86
Louisiana	71	72	74
Maine	84	85	86
Maryland	83	84	85
Massachusetts	83	85	85
Michigan	74	76	77
Minnesota	77	78	80
Mississippi	75	75	76
Missouri ²	81	84	86
Montana	82	84	84
Nebraska	86	88	88
Nevada	62	63	71
New Hampshire	86	86	87
New Jersey	83	86	88
New Mexico	63	70	70
New York	77	77	77
North Carolina	78	80	83
North Dakota	86	87	88
Ohio	80	81	82

— Not available.

¹ The United States 4-year ACGR was estimated using both the reported 4-year ACGR data from reporting states and the District of Columbia and using imputed data for Idaho, Kentucky, and Oklahoma for school years 2010-11 and 2011-12, and imputed data for Idaho for school year 2012-13.

² School year 2011-12 data for California, Hawaii, and Missouri were revised subsequent to the publication of these data in NCES 2014-391. The estimated United States ACGR includes these revisions.

³ The Department of Education's Office of Elementary and Secondary Education approved a timeline extension for these states to begin reporting 4-year ACGR data, resulting in the 4-year ACGR not being available in one or more of the school years shown.



Adjusted Cohort Graduation Rate

State	2010-11	2011-12	2012-13
Oklahoma ³	—	—	85
Oregon	68	68	69
Pennsylvania	83	84	86
Rhode Island	77	77	80
South Carolina	74	75	78
South Dakota	83	83	83
Tennessee	86	87	86
Texas	86	88	88
Utah	76	80	83
Vermont	87	88	87
Virginia	82	83	84
Washington	76	77	76
West Virginia	78	79	81
Wisconsin	87	88	88
Wyoming	80	79	77

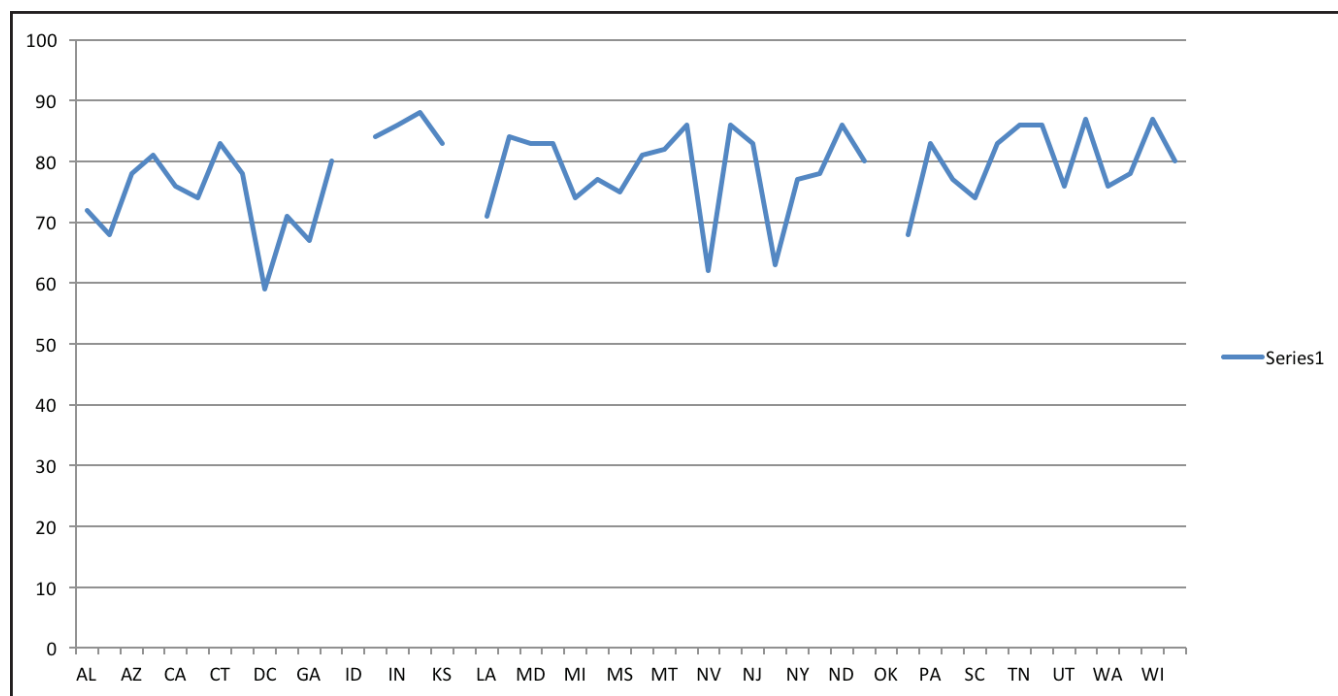
NOTE: The 4-year ACGR is the number of students who graduate in 4 years with a regular high school diploma divided by the number of students who form the adjusted cohort for the graduating class. From the beginning of 9th grade (or the earliest high school grade), students who are entering that grade for the first time form a cohort that is adjusted by adding any students who subsequently transfer into the cohort and subtracting any students who subsequently transfer out, emigrate to another country, or die.

SOURCE: EDFacts/Consolidated State Performance Report, school years 2010-11, 2011-12, and 2012-13, <http://www2.ed.gov/admins/lead/account/consolidated/index.html>. This table was prepared January 2015.

Step 3. Findings: Data Meaning/Answer

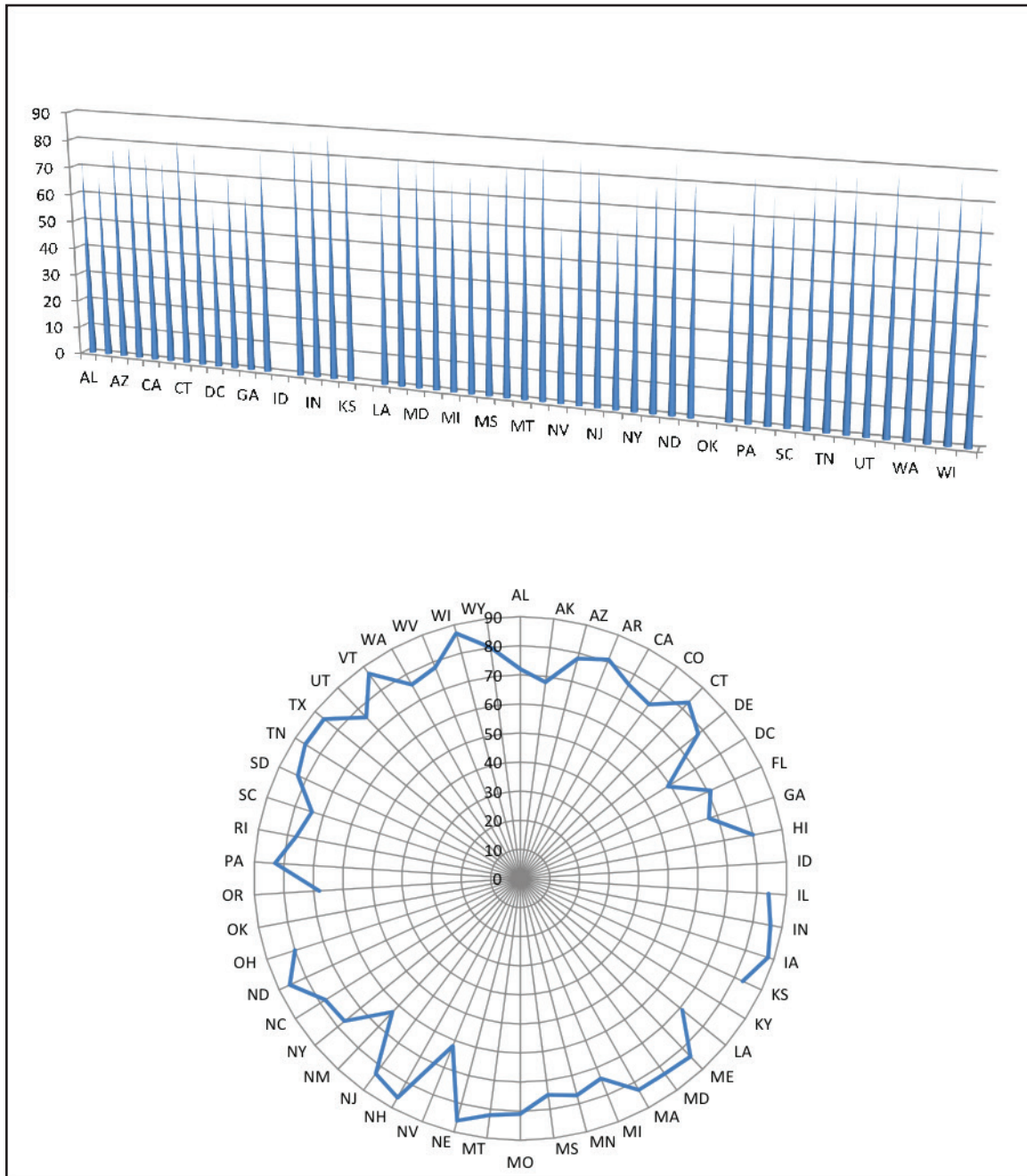
While the data presented in tabular form in table 4.1 are appropriate for some types of audiences, especially members of the research community who may be looking for a repository of all original data values, even seasoned analysts are likely to find it difficult to identify patterns, trends, or cues in such a table. Thus, an astute analyst will likely wish to visualize these data (see chapter 2) in order to more accurately compare the graduation rate of any single state with the rest of the states in the table (and nation) (see figures 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8 for perspectives in the advancement of possible visualization choices).

Figure 4.2. The default setting in a common spreadsheet tool produces a graph with many features that are likely to lead to misunderstanding or misinterpretation of the data.



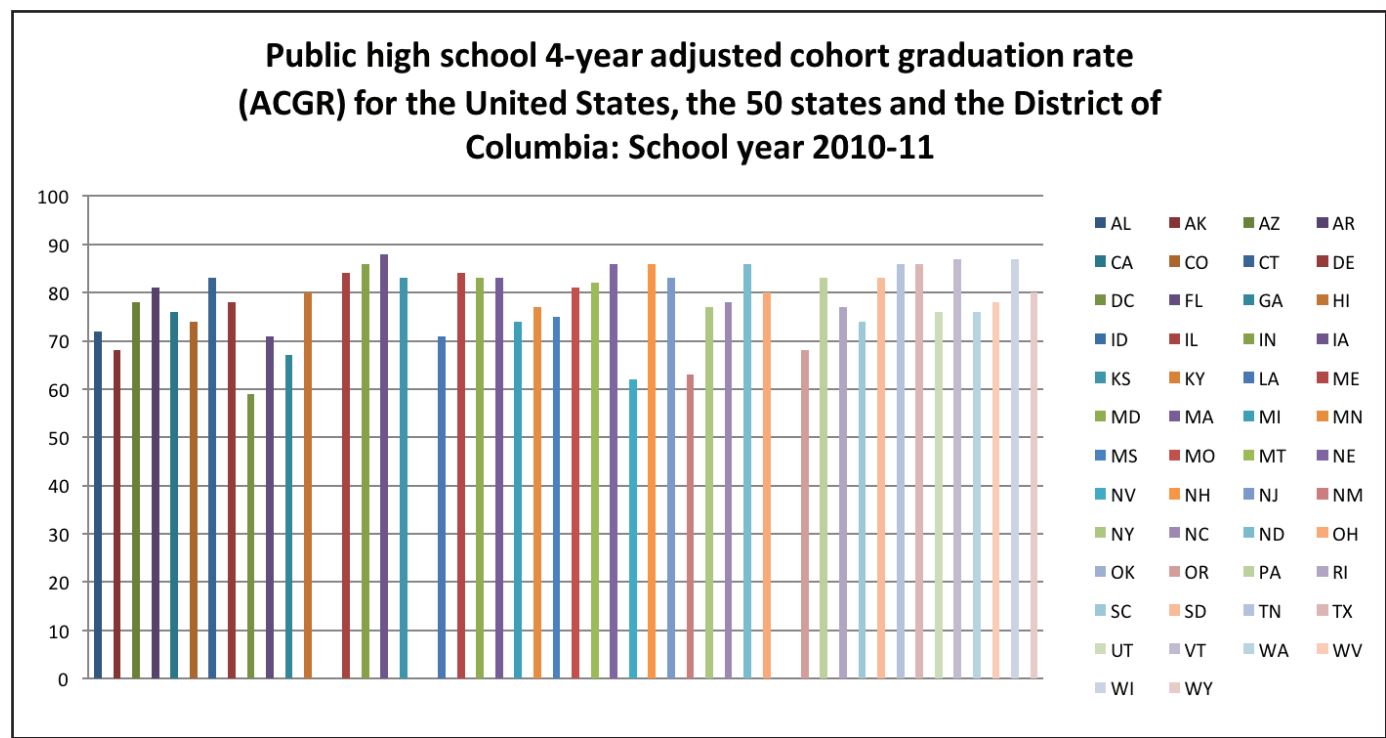
Analysis: The default graph from a spreadsheet does not include a title or label to identify what is being viewed (for example, what year's data are being presented?). Most viewers would probably recognize that the x-axis represents state postal codes, although it appears that the axis only displays every other abbreviation, most likely as a default given the width of the figure. The units on the y-axis are not labeled. From the perspective of a skilled data analyst, however, the more egregious mistake is that the default creates a line graph, suggesting that each state's separate rate is connected to other rates as continuous variables rather than the discrete data values that they actually represent. Gaps in the line (data) are not explained.

Figure 4.3. Other default chart types produce figures that fail to improve understanding to varying degrees.



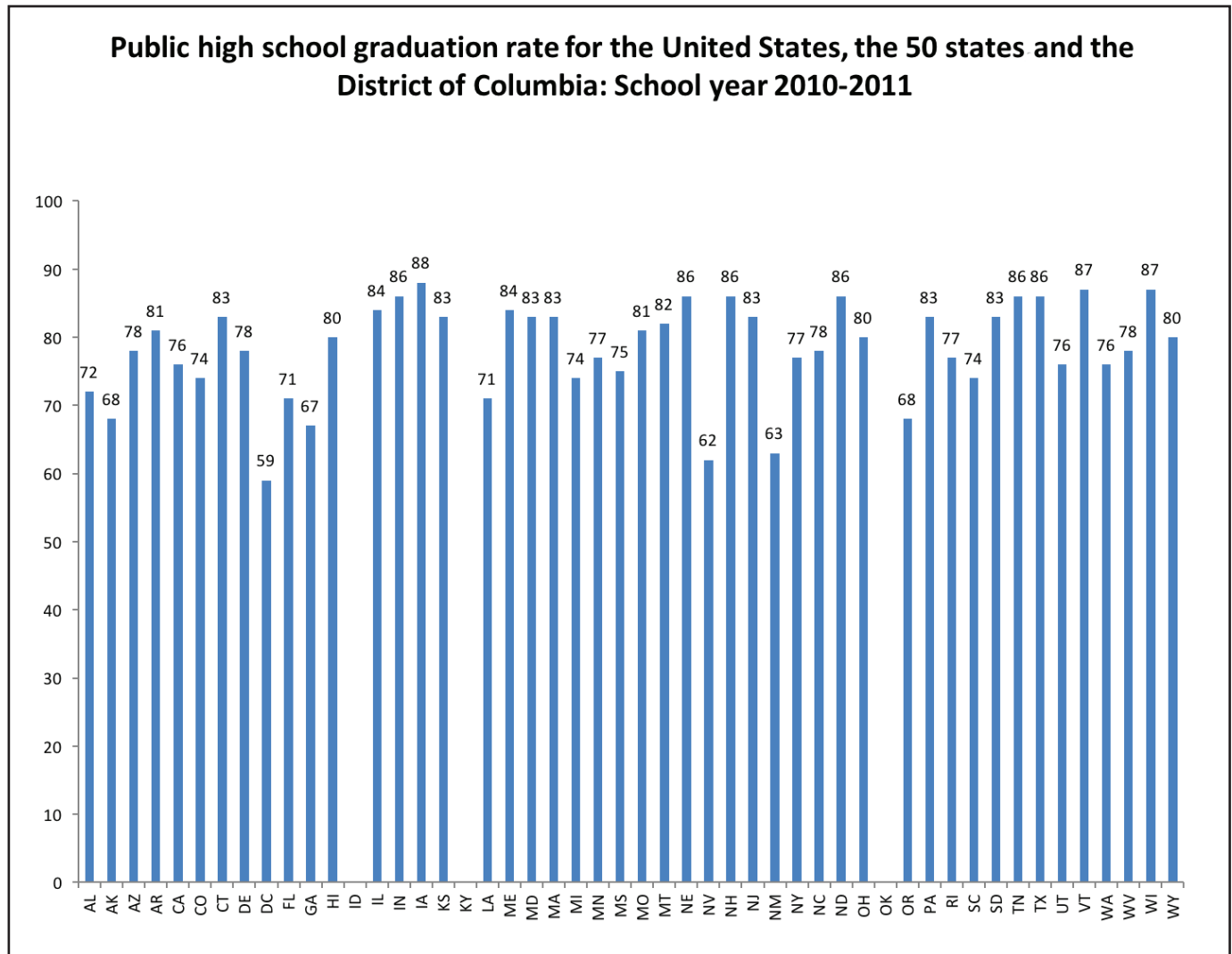
Analysis: How does a 3-D cone format improve understanding in the top image? At which data value does the pinnacle of each cone end? The thin lines make discerning actual data values practically impossible. Moreover, why would 2-dimensional data (state name and graduation rate) require a 3-D presentation? In the bottom image, the radar chart is another attempt to put style over substance. Such presentations are not only unfamiliar to many audiences—they also violate Key Principle 2 (reduce the clutter) by introducing visual elements simply for stylistic purposes.

Figure 4.4. A bar chart (for discrete data) and figure title begin to improve understanding and analysis.



Analysis: A bar chart correctly shows that each state’s ACGR is a discrete value (a major and necessary technical correction). The multicolored key, while logical in theory, is impractical in reality. It is not reasonable to think that a viewer can match the nuanced variations in 50 colors in the key to the data bars they represent. It is equally impractical to think that a printed version of the image will allow viewers to discriminate between slight variations in color.

Figure 4.5. More effective visualization choices, including the application of key principles covered in chapter 3, improve the visual appeal and analysis of the data.

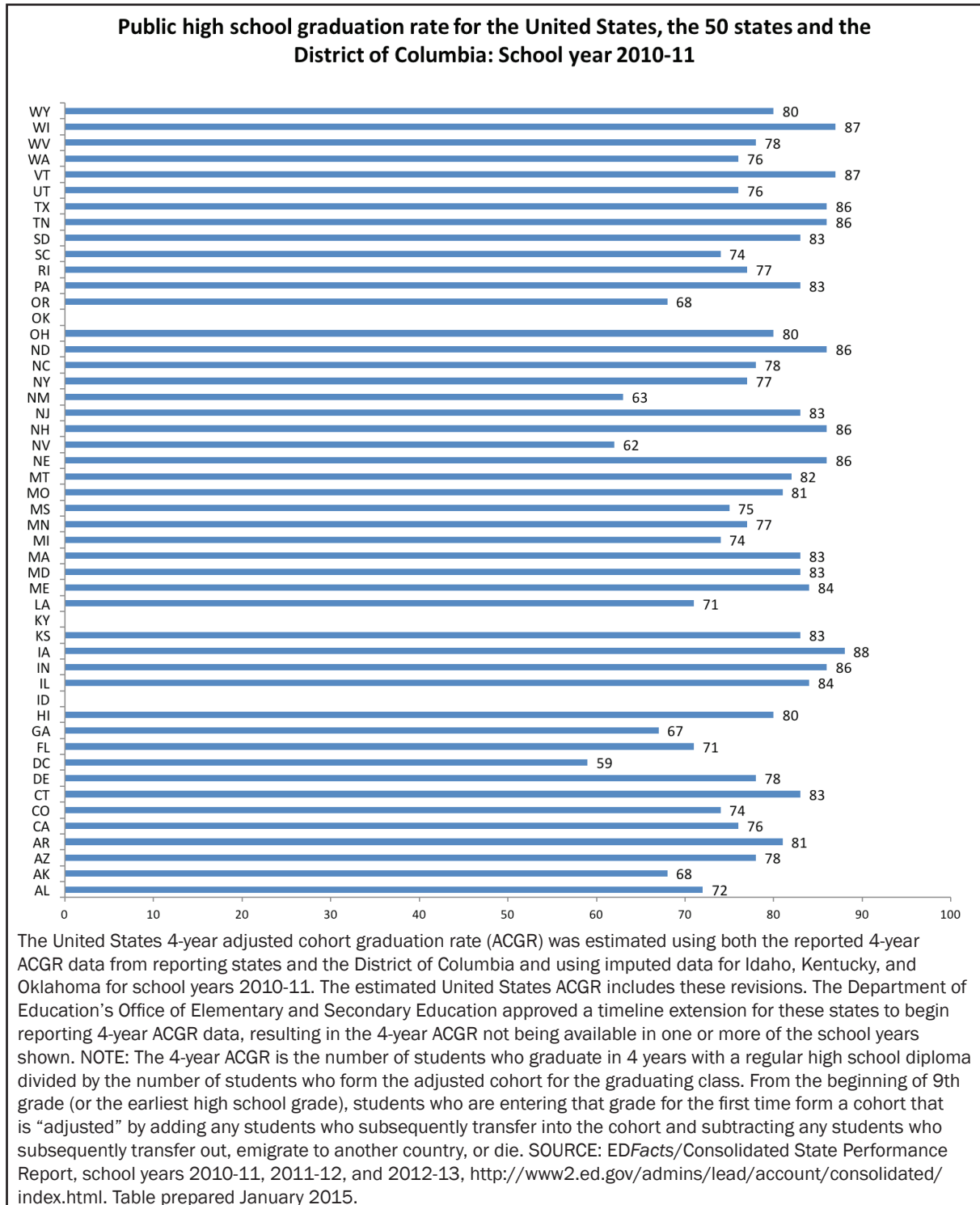


Analysis: This bar chart is a next step toward presenting the data in a useful format for many audiences. For example, data values for each state replace horizontal y-axis grid lines (Key Principle 1: Show the Data). A single color for the bars is visually less distracting for many viewers and the use of state abbreviations on the x-axis permits the removal of the key (Key Principle 2: Reduce Clutter).

Step 4. Customization: Audience-Specific Messaging

While figure 4.5 may meet the needs of a data analyst, the question in step 1 originated from a more general audience (policymakers, administrators, or education advocates), which means that visualization designers cannot assume any data or statistical expertise on the part of the viewer. Because such an audience warrants a “no training required” approach to visualization, planners should apply as many key principles from chapter 3 as make sense for this data message and audience. Thus, figure 4.6 integrates text with the figure (Key Principle 3) to ensure that data are accurately portrayed by including the data source and a definition of the 4-year adjusted cohort graduation rate (Key Principle 4).

Figure 4.6. Wiser visualization choices improve the likelihood of identifying meaning in the data.



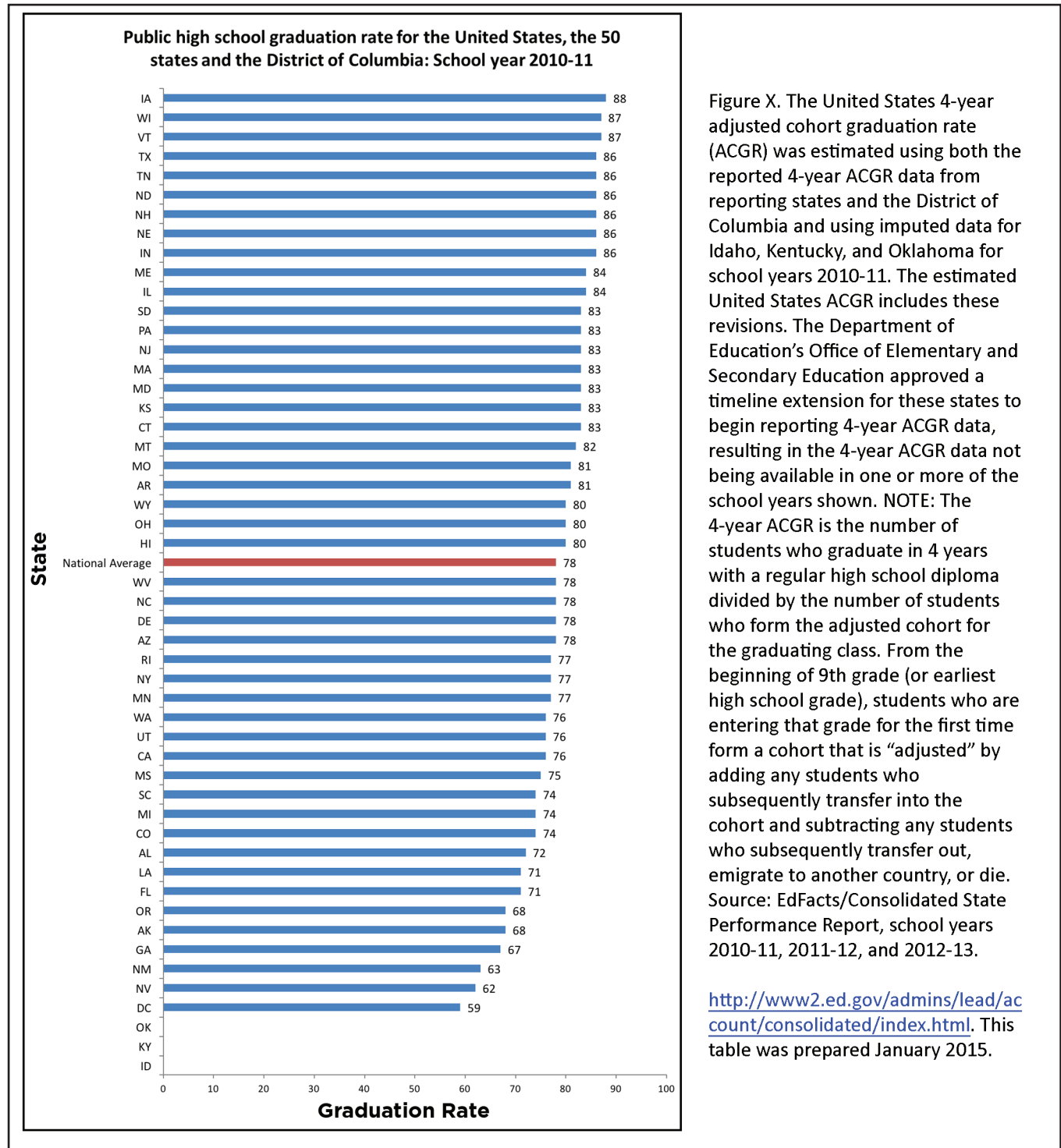
Analysis: A horizontal presentation of the bars is likely to enhance the comparability of state graduation rate values for many viewers. The inclusion of source notes allows the graphic to stand alone as a piece of information (Key Principle 3: Integrate text and figures).



Step 5. Visualization: Present Data Meaning Clearly and Accurately

Although the data are presented accurately (including data sources and definitions), the application of several recommended practices from chapter 2 will further clarify meaning. For example, figure 4.7 illustrates that the visual power of reordering the states from highest to lowest data values (Recommendation 3: don't limit your design choices), inserting a national average value (Recommendation 4: focus on the take-home message for the target audience), and highlighting that national average in another color to simplify comparisons (Recommendation 7: recognize the importance of color) will all contribute to better understanding of the take-home message. Depending on the media in which the visualization is released, including numerical and text values for all information in the visualization will be important for ensuring that the image complies with Section 508 accessibility expectations (Recommendation 7: recognize the benefits of Section 508 compliance).

Figure 4.7. Final visualization choices improve the presentation of the data, especially for non-expert viewers.

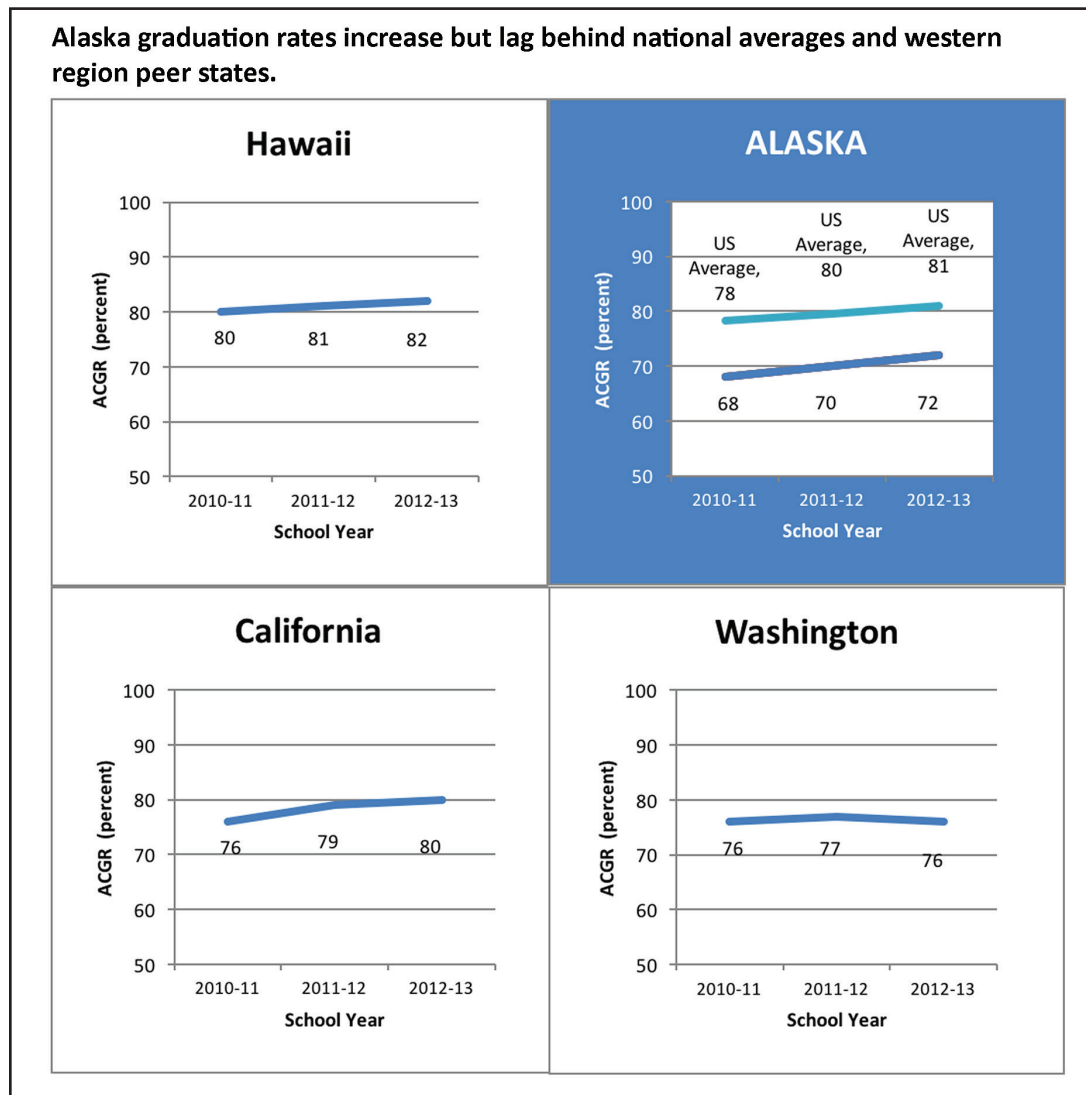


Analysis: Rank ordering the states (each bar) and adding a national average value in a contrasting color facilitates easy comparison and completes the visualization. Note that even non-expert viewers can quickly ascertain where their state graduation rate ranks relative to other states in the nation and against the national average. The inclusion of source data confirms the stand-alone nature of the visualization.

Step 6. User Feedback: Review and Refine Efforts

Because getting stakeholders the information they need to make sound decisions is a priority of the organization's senior leadership, the communications team convenes a focus group of representative data users to solicit feedback on the effectiveness of their visualization effort. They learn that figure 4.7 is generally understandable by a wide range of people, but isn't as specific as it needs to be for the public to compare their state to both the nation and other states in their region. Options for emphasizing specific data within a broader body of information include the use of color, borders, bolded items, highlighted text, and font size to more prominently display subsets of data or other critical messages in the data. Staff evaluated these options and decided to refine the visualization. The product of this refinement, and the finalized draft of the visualization, is presented in figure 4.8.

Figure 4.8. Feedback from representative members of an intended audience can help staff to refine the design of a visualization to more precisely meet the audience's information needs.



Analysis: A narrower presentation of data, focusing on a single state (Alaska) and its regional peers, shows a three-year trend in graduation rates compared to national averages and peer states. A highly descriptive title explicitly explains the take-home message for a viewer.

Summary

Data visualization is the process through which data are transformed into visually meaningful information for intended audiences. This chapter illustrated a six-step process for visualizing data that reflects sound research methods for analyzing and communicating high-quality data:

The ultimate purpose of collecting and sharing data is to enable stakeholders to use the information to improve the education system. Data are meant to be used to make decisions.

Six-Step Process for Visualizing Data

- Step 1. Question: Someone Needs Information
- Step 2. Research: Data Exploration and Analysis
- Step 3. Findings: Data Meaning/ Answer
- Step 4. Customization: Audience-Specific Messaging
- Step 5. Visualization: Present Data Meaning Clearly and Accurately
- Step 6. User Feedback: Review and Refine Efforts

Data visualization supports efficiency in the education system. Once an education organization has gone to the effort of collecting data, failing to use the information to inform instructional, administrative, and policy-related decisionmaking devalues a precious information resource. Taking action with data—the right data at the right time in the right format and in the right context—can be a powerful tool for anyone needing to make choices about how our educational system serves students and communities (National Forum on Education Statistics 2012). Data visualization is a critical component of the data analysis and communications process for many education stakeholders.

Appendix A: Data Visualization Handouts

The Data Visualization Process in Your Education Agency

Given the detailed data that are collected about the inputs, processes, and outcomes of the education enterprise, it is not surprising that discerning the meaning of data is a challenge for education stakeholders, including practitioners, policymakers, researchers, parents, and the general public.

The ability to create customized, audience-specific data visualizations can become a fundamental and powerful aspect of a broader organization-wide analytical and communications strategy. Data visualization focuses on presenting information in a way that is not only accurate and appropriately comprehensive, but also understandable and actionable for each of your intended audiences.

When applied effectively, the sound data visualization approaches below will improve a viewer's ability to understand, analyze, and retain information and, subsequently, use that knowledge to make decisions.

Four Key Principles for Effective Data Visualization

- Key Principle 1: Show the data.
- Key Principle 2: Reduce the clutter.
- Key Principle 3: Integrate text and images.
- Key Principle 4: Portray data meaning accurately and ethically.

Education organizations share data with stakeholders because the information is judged to be of value. Providing clear and accurate information about education settings, processes, and performance is a fair, necessary, empowering, and healthy component of our education system.

Seven Recommended Practices for Data Visualization

- Recommendation 1: Capitalize on consistency.
- Recommendation 2: Data that should not be compared should not be presented side by side.
- Recommendation 3: Don't limit your design choices to default graphing programs.
- Recommendation 4: Focus on the take-home message for the target audience.
- Recommendation 5: Minimize jargon, acronyms, and technical terms.
- Recommendation 6: Choose a font that is easy to read and will reproduce well.
- Recommendation 7: Recognize the importance of color and the benefits of Section 508 compliance.

Six-Step Process for Data Visualization

- Step 1. Question: Someone Needs Information
- Step 2. Research: Data Exploration and Analysis
- Step 3. Findings: Data Meaning/Answer
- Step 4. Customization: Audience-Specific Messaging
- Step 5. Visualization: Present Data Meaning Clearly and Accurately
- Step 6. User Feedback: Review and Refine Efforts

For more information about these data visualization process, principles, and recommended practices, download the free *Forum Guide to Data Visualization: A Resource for Education Agencies* at <http://nces.ed.gov/forum/publications.asp>.

Data Visualization

- D**isplay (or show) the data (Key Principle 1)
- A**void or reduce clutter (Key Principle 2)
- T**ext and images must be integrated (Key Principle 3)
- A**ccurately and ethically portray data meaning (Key Principle 4)

- V**erify quality of data
- I**nvest in more than default programs
- S**ide-by-side presentations only when comparison is intended
- U**se a format that is easy to read
- A**void multiple fonts
- L**ess is more
- I**mages distinguishable by contrast not color
- Z** patterns and **F** patterns are effective
- A**ccessibility necessitates Section 508 compliance
- T**ake-home message is the priority
- I**nsight is preferable to hindsight
- O**bserve size and position hierarchy to indicate importance
- N**ot all data need to be visualized

Created by Zenaida Napa Natividad, Guam Department of Education

For more information about the data visualization process, principles, and recommended practices, download the free *Forum Guide to Data Visualization: A Resource for Education Agencies* at <http://nces.ed.gov/forum/publications.asp>.

Appendix B: Citations and Additional Resources

Citations

- Beniger, J.R. and Robyn, D.L. (1978). Quantitative graphics in statistics: A brief history. *The American Statistician*. 32: pp. 1–11.
- Card, S.K., Mackinlay, J.D., and Schneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Academic Press: London, UK.
- Cleveland, W. (1993). *Visualizing Data*. Hobart Press, Lafayette, IN.
- Dynarski, M., & Kisker, E. (2014). Going public: Writing about research in everyday language. (REL 2014-051). U.S. Department of Education: Washington, DC. Retrieved July 2016 from <http://files.eric.ed.gov/fulltext/ED545224.pdf>.
- Ernst, J. V., & Clark, A. C. (2009). Technology-based content through virtual and physical modeling: A national research study. *Journal of Technology Education*, 20(2), 23-36. Retrieved July 2016 from <http://eric.ed.gov/?id=EJ898827>.
- Evergreen, S.D.H (2014). *Presenting Data Effectively: Communicating Your Findings for Maximum Impact*. Sage Publications. Thousand Oaks, CA.
- Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press. Oakland, CA.
- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Second Edition. Analytics Press. Oakland, CA.
- Few, S. (2014): Data Visualization for Human Perception. In Soegaard, Mads and Dam, Rikke Friis (eds.). *The Encyclopedia of Human-Computer Interaction*, 2nd Ed. Aarhus, Denmark: The Interaction Design Foundation. Retrieved March 2015 from https://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html.
- Healey, C. G., Booth, K. S., and Enns, J. T. (1996). High-Speed Visual Estimation Using Preattentive Processing. *ACM Transactions on Human Computer Interaction* 3(2), pages 107-135, 1996. Retrieved July 2016 from <http://www.csc.ncsu.edu/faculty/healey/download/tochi.96.pdf>.
- IBM (2016): *What is Big Data? Bringing Big Data to the Enterprise*. Retrieved May 2016 from <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>.
- Illinsky, N. and Steele, J. (2011). *Designing Data Visualizations*. O'Reilly Media: Sebastopol, CA.
- Kosara, R. (2007). Visualization Criticism - The Missing Link Between Information Visualization and Art, in *Information Visualization*, 2007. IV '07. 11th International Conference, vol., no., pp.631-636, 4-6.

- Masud, L., Valsecchi, F., Ciuccarelli, P., Ricci, D., & Caviglia, G. (2010). *From Data to Knowledge – Visualizations as Transformation Processes within the Data-Information-Knowledge Continuum*. 14th International Conference Information Visualization (IV), 445 – 499. Retrieved July 2016 from http://www.researchgate.net/profile/Paolo_Ciuccarelli/publication/221360612_From_Data_to_Knowledge_-_Visualizations_as_Transformation_Processes_within_the_Data-Information-Knowledge_Continuum/links/0fcfd5006b40145ed6000000.pdf.
- Munari, B. (1966). *Design as Art*. Penguin Group: London.
- National Forum on Education Statistics. (2005). *Forum Guide to Education Indicators* (NFES 2005–802). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved July 2016 from http://nces.ed.gov/forum/pub_2006807.asp.
- National Forum on Education Statistics. (2011). *Forum Guide to Ensuring Equal Access to Education Websites: An Introduction to Electronic Information Accessibility Standards* (NFES 2011–807). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved July 2016 from http://nces.ed.gov/forum/pub_2011807.asp.
- National Forum on Education Statistics. (2009). *Forum Guide to Metadata: The Meaning Behind Education Data* (NFES 2009–805). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved July 2016 from http://nces.ed.gov/forum/pub_2009805.asp.
- National Forum on Education Statistics (2012). *Forum Guide to Taking Action with Education Data*. (NFES 2013-801). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved July 2016 from http://nces.ed.gov/forum/pub_2013801.asp.
- National Forum on Education Statistics (2015). *Forum Guide to Building a Culture of Data Quality*. (NFES 2005-801). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved July 2016 from http://nces.ed.gov/forum/pub_2005801.asp.
- Schwabish, Jonathan A. (2014). An Economist's Guide to Visualizing Data. *Journal of Economic Perspectives*, 28(1), 209–234. Retrieved July 2016 from <http://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.1.209>.
- Treisman, A. (1985). Preattentive Processing in Vision. *Computer Vision, Graphics, and Image Processing*, 31(2):156-177.
- Treisman, A. (1986). Features and Objects in Visual Processing. *Scientific American*, 255(5):114-125.
- Tufte, E.R. (2001). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT.
- Wolfe, J. and Robertson, L (Eds). 2012, *From Perception to Consciousness: Searching with Anne Triesman*. Oxford University Press: Oxford, England.
- Wong, D.M. (2010). *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures*. W.W. Norton & Company, Inc. New York.



Additional Resources

Data.Gov. (2015). *Education*. <http://www.data.gov/education/>

Fast, E. F., Blank, R. K., Potts, A., & Williams, A. (2002). *A Guide to Effective Accountability Reporting*. Council of Chief State School Officers: Washington, DC. <http://programs.ccsso.org/content/pdfs/GEAR.pdf>

Google Public Data Explorer (2011). https://www.youtube.com/watch?v=AM6w_tUlln4

Institute of Education Sciences (2014). *Graphic design for researchers*. Washington, DC: Institute of Education Sciences. <http://eric.ed.gov/?id=ED545255>

Krauss, J. (2012). Infographics: More than words can say. *Learning & Leading with Technology*, 39(5), 10-14. <http://eric.ed.gov/?id=EJ982831>

National Center for Education Statistics (2012). *The Nation's Report Card: What every parent should know about NAEP*. (NCES 2012-469). Institute of Education Sciences. U.S. Department of Education. <http://eric.ed.gov/?id=ED532973>

National Center for Education Statistics (2014). *Are the nation's twelfth-graders making progress in mathematics and reading?* (NCES 2014-087). Institute of Education Sciences. U.S. Department of Education. <http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014087.pdf>

Plain Language Action and Information Network (2011). Federal Plain Language Guidelines. <http://www.plainlanguage.gov/howto/guidelines/FederalPLGuidelines/FederalPLGuidelines.pdf>

Statewide Longitudinal Data Systems Grant Program (2015). *SLDS spotlight: Data Use through Visualizations and Narratives*. Statewide Longitudinal Data System Program. National Center for Education Statistics. Institute of Education Sciences. U.S. Department of Education. http://nces.ed.gov/programs/slds/pdf/spotlight_Data_use_visualization.pdf

U.S. Department of Education (2015). *Federal Student Aid: Financial aid toolkit*. <http://www.financialaidtoolkit.ed.gov/tk/outreach/social-media/infographics.jsp>

What Works Clearinghouse (2012). *What Works Clearinghouse Reporting Guide for Study Authors*. Institute of Education Sciences. U.S. Department of Education. <http://ies.ed.gov/ncee/wwc/Document/235>