

Data Visualization

Michael Hopkins

Rochester School Department, New Hampshire

Presentation Outline

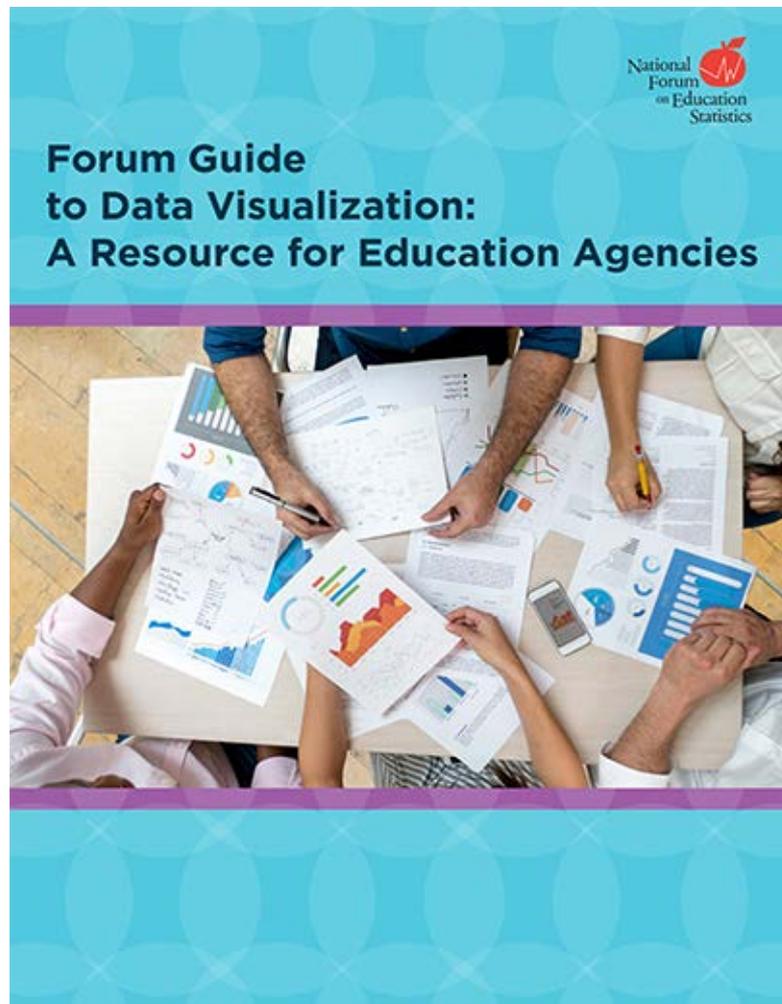
- Forum Resources on Data Visualization
- Data Visualization in Education Organizations
- Data Visualization to Advance Data Analysis
- Data Visualization to Improve Communications
- Implementing the Data Visualization Process

Data Visualization is Important

Effective data visualization can

- improve communications;
- make data more user friendly; and
- increase stakeholder access to data.

Forum Guide to Data Visualization: A Resource for Education Agencies



Document Purpose & Audience

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.

The document will be useful for

- program and data staff;
- data analysts and researchers;
- administrators and policymakers; and
- related roles analyzing or presenting data for public consumption.

Table of Contents

Chapter 1. Data Visualization in Education Organizations

- defines the concept of data visualization;
- describes how data visualization blends both science and art; and
- explains how data visualization can improve education data use.

Chapter 2. Data Visualization to Advance Data Analysis

- illustrates how data visualization can be a sound method for analysts to identify trends, patterns, and cues in data.

Chapter 3. Data Visualization to Improve Communications

- presents four key principles and seven practical recommendations that will improve efforts to visualize data for stakeholders who need to understand and use education data to make decisions.

Chapter 4. Implementing the Data Visualization Process

- illustrates a six-step process for visualizing data for both analytical and communications purposes.

10 and 2

- 10 minutes of presentation and at least 2 minutes of reflection.
- Mental processing time is needed.
- Present information for 10 minutes.
- Allow students to process information individually or collaboratively for 2 minutes.
- Retention increases with the 2 minutes reflection and sharing with others.

Forum Data Visualization Online Course

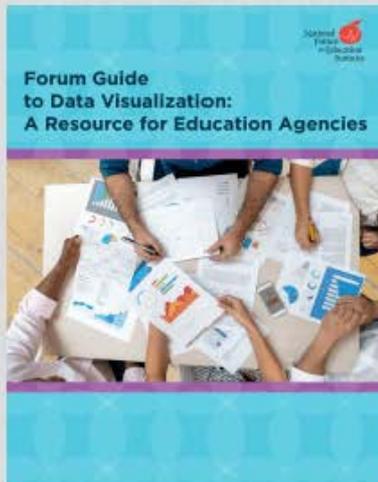
Module 1: Introduction to Data Visualization

- introduces 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 common education stakeholders;
- introduces the key principles and characteristics of effective data visualization; and
- explains how effective data visualization techniques can contribute to a culture of data quality throughout an education agency.

>> COURSE: Data Visualization

LESSON MENU

Module 1. Introduction to Data Visualization



This free course is part of a suite of resources on data visualization, which includes the *Forum Guide to Data Visualization: A Resource for Education Agencies*. The audience for this course includes staff in local, state, and federal education agencies whose responsibilities include any aspect of analyzing data or sharing meaning through data with education stakeholders. The course may also be of interest to others concerned about the utility of elementary and secondary education data.

This module is divided into four lessons:

- Lesson 1: What is Data Visualization?
- Lesson 2: Key Principles for Effective Data Visualization
- Lesson 3: Six-Step Process for Data Visualization
- Lesson 4: Data Visualization and Education Data: A Simple Example

After completing this course, you will be able to identify the principles for the application of data visualization techniques to the presentation of education data.

Ch. 1: Data Visualization in Education Organizations



What is Data Visualization?

Data visualization is the transformation of data into information through visual presentation and analysis.

- It may culminate in a figure or image, but should not be viewed simply as a graphical product.
- 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.

Ch. 1: Data Visualization in Education Organizations

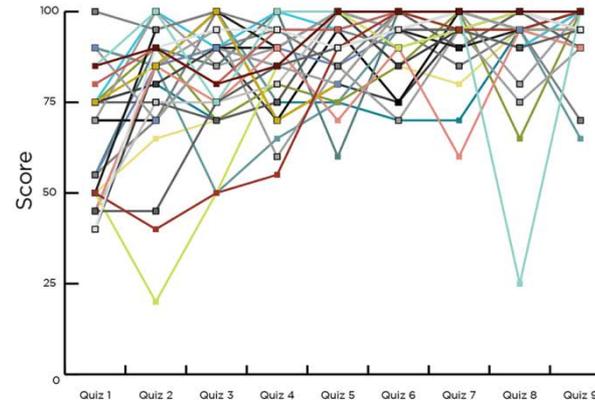
How many words is that picture worth?

Quiz Performance, Class 3B

	A	B	C	D	E	F	G	H	I	J
1	Student	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5	Quiz 6	Quiz 7	Quiz 8	Quiz 9
2	1	85	90	80	85	100	100	100	100	100
3	2	40	90	95	95	90	95	100	100	95
4	3	80	90	80	95	95	100	100	100	95
5	4	85	100	75	100	100	100	100	25	100
6	5	90	70	90	85	80	100	100	95	95
7	6	90	40	50	55	90	100	95	95	100
8	7	75	85	100	70	80	100	95	95	95
9	8	45	45	75	85	90	95	95	90	95
10	9	50	20	50	85	100	90	95	100	100
11	10	45	85	75	90	70	90	60	95	90
12	11	100	95	100	95	95	90	95	100	100
13	12	75	80	90	85	100	85	95	100	95
14	13	70	100	85	90	100	95	100	80	100
15	14	50	95	100	90	100	95	90	95	100
16	15	45	75	75	80	100	95	100	100	95
17	16	75	75	70	75	100	100	90	100	90
18	17	70	70	90	90	100	100	100	95	95
19	18	75	100	90	100	95	100	95	90	100
20	19	55	70	90	95	85	100	95	80	100
21	20	55	85	90	90	85	95	95	95	100
22	21	75	85	95	60	85	70	95	75	90
23	22	55	90	85	100	60	100	100	100	95
24	23	75	90	70	80	75	85	100	65	100
25	24	90	85	50	65	75	95	100	95	100
26	25	90	85	100	70	95	75	100	95	100
27	26	45	85	100	75	100	95	95	95	85
28	27	70	70	100	75	100	95	85	95	70
29	28	40	85	90	70	80	75	95	95	95
30	29	50	65	70	80	100	85	80	95	100
31	30	75	80	70	75	75	70	70	95	90

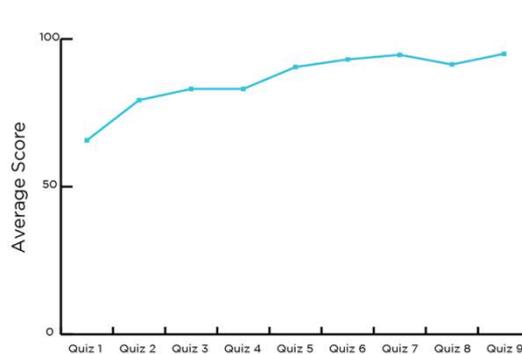
(1) Raw Tabular Data

Quiz Performance, Class 3B



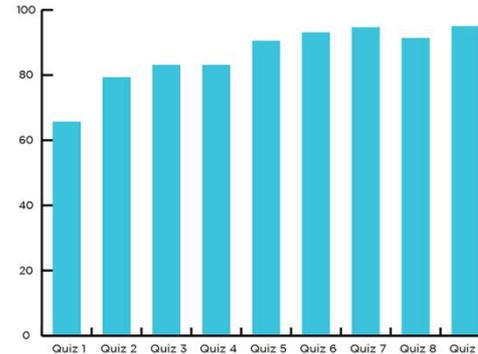
(2) Complex Data Presentation

Quiz Performance, Class 3B



(3) Effective Data Presentation

Quiz Performance, Class 3B



(4) Accurate Data Presentation

Ch. 1: Data Visualization in Education Organizations

Effective data visualization *is*

- valuable as an analytical and communications tool because of the insights it provides through visually apparent cues, patterns, and trends.
- customized to meet the information needs of specific intended audiences.
- intentionally 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.
- more complex or creative than it needs to be to accurately convey data meaning.

Ch. 2: Data Visualization to Advance Data Analysis

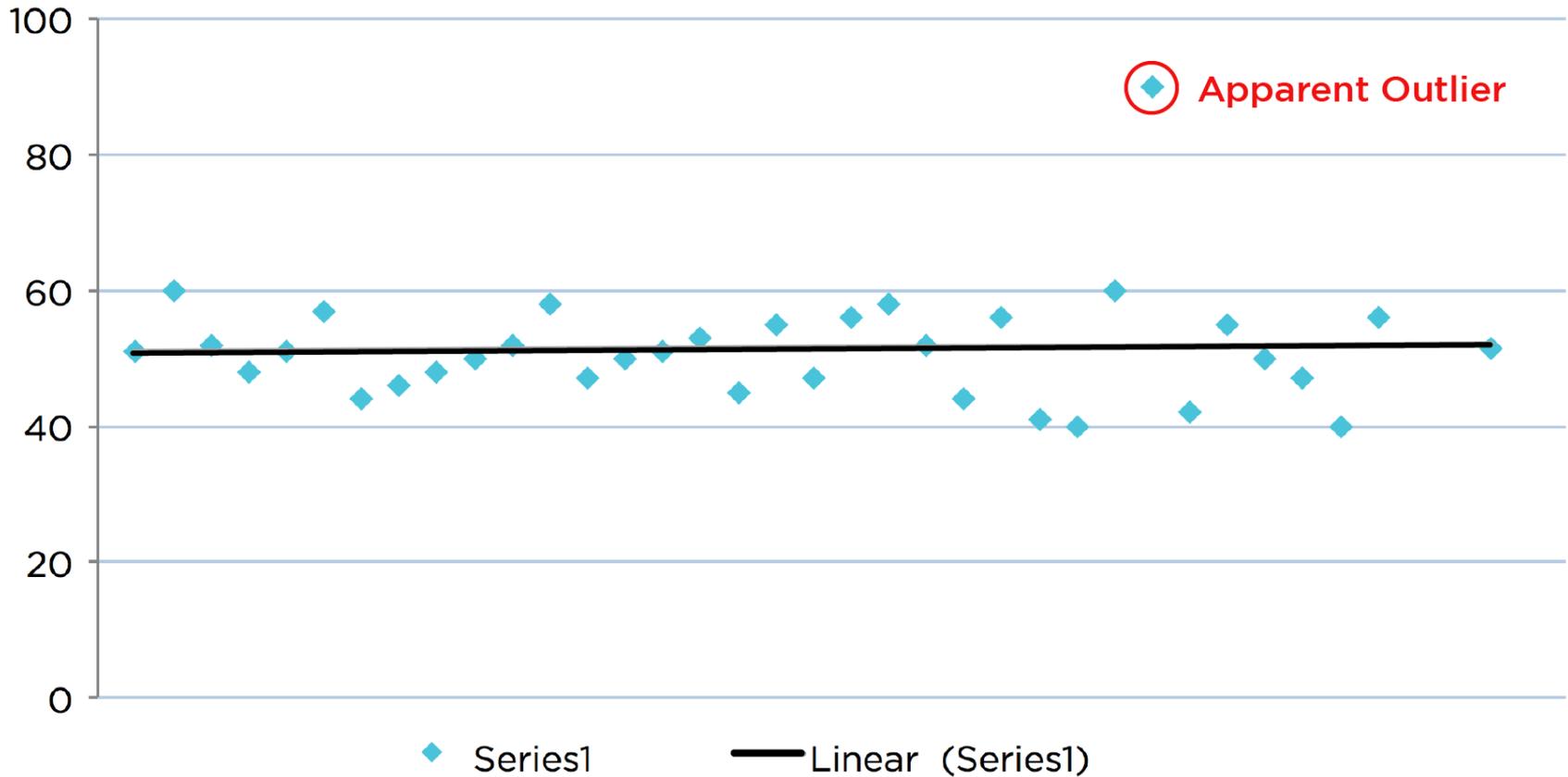
Find the outlier in this (very small) dataset ...

- Average Value = 52.
- Data Values =
51, 60, 52, 48, 51, 57, 44, 46, 48, 50, 52,
58, 47, 50, 51, 53, 45, 55, 47, 56, 58, 52,
44, 56, 41, 40, 60, 90, 42, 55, 50, 47, 56

Ch. 2: Data Visualization to Advance Data Analysis

Average Value: 52

Data Values: 51, 60, 52, 48, 51, 57, 44, 46, 48, 50, 52, 58, 47, 50, 51, 53, 45, 55, 47, 56, 58, 52, 44, 56, 41, 40, 60, 90, 42, 55, 50, 47, 40, 56

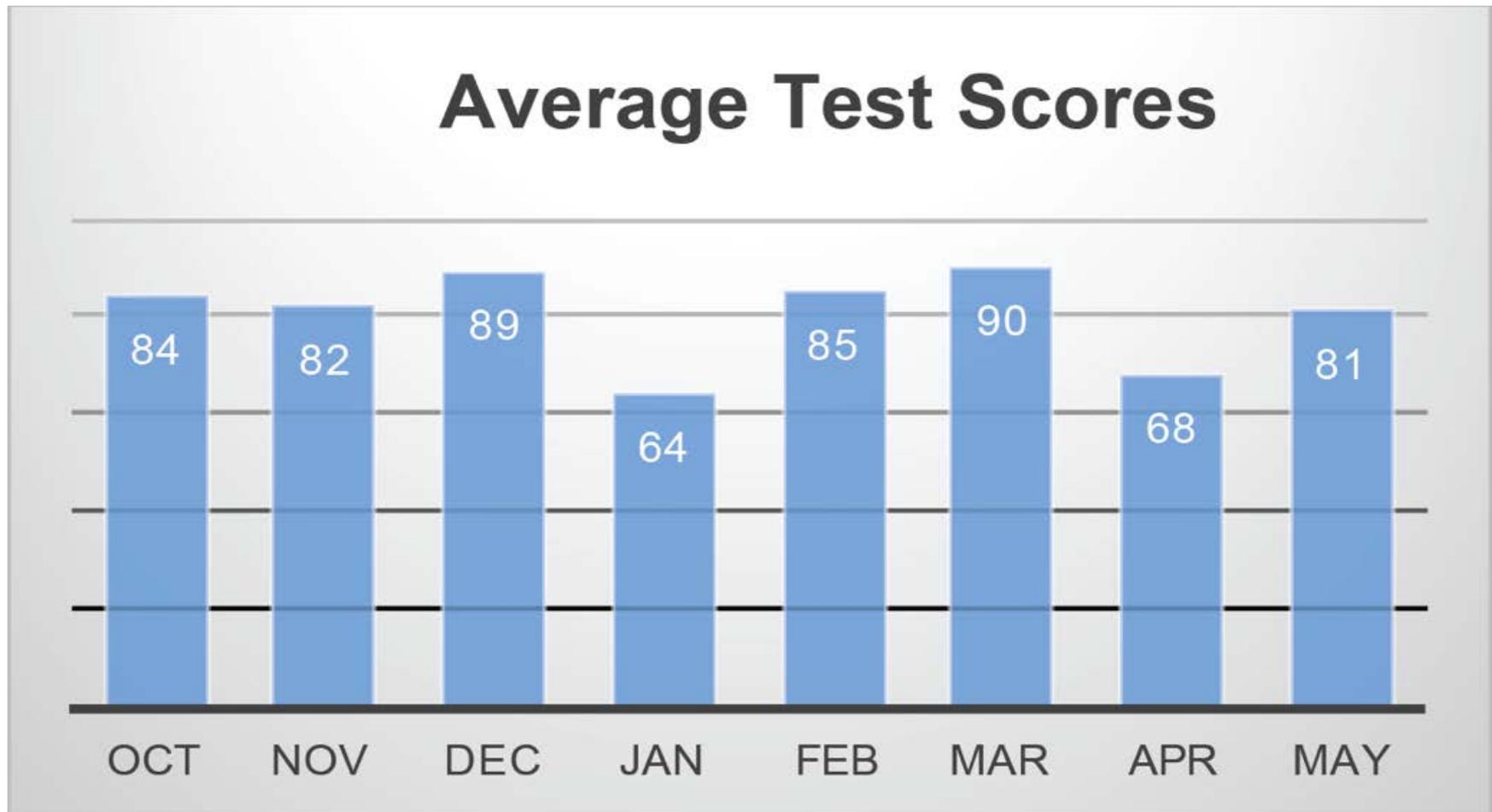


Ch. 2: Data Visualization to Advance Data Analysis

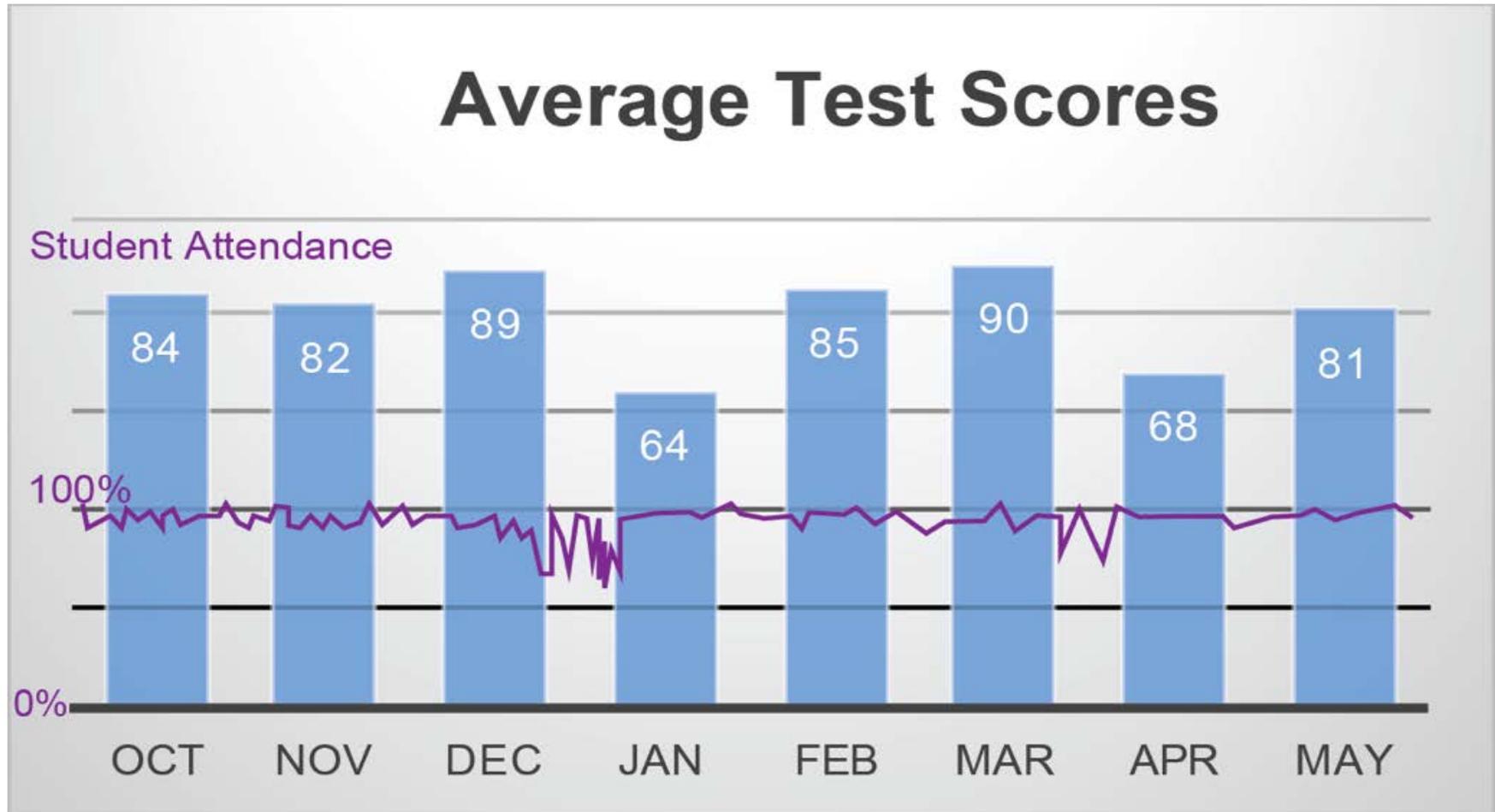
Data visualization to advance data analysis...

- focuses on the needs of data analysts who are trying to determine what a particular set of data, or multiple datasets, might mean.
- is not intended to share their analytical methods as communications tools.

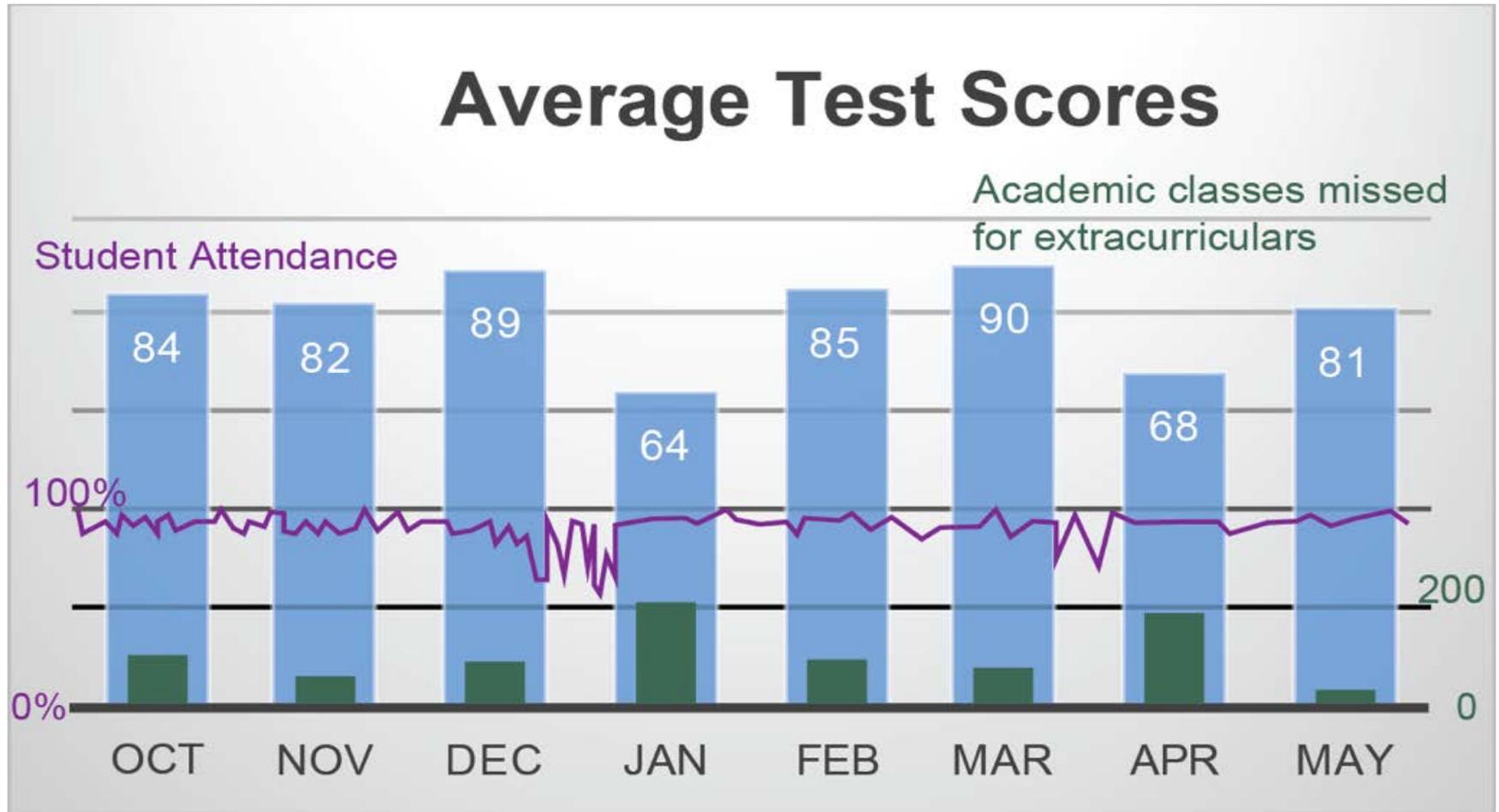
Ch. 2: Data Visualization to Advance Data Analysis



Ch. 2: Data Visualization to Advance Data Analysis



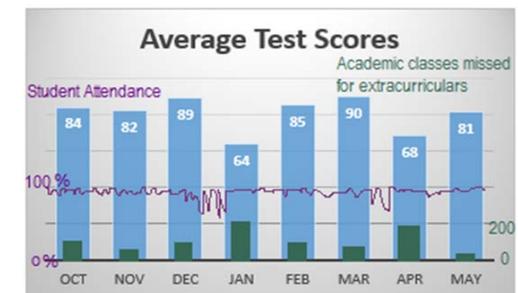
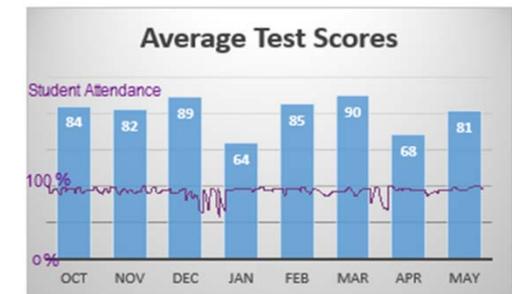
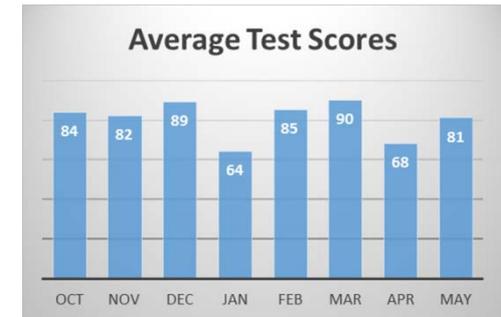
Ch. 2: Data Visualization to Advance Data Analysis



Ch. 2: Data Visualization to Advance Data Analysis

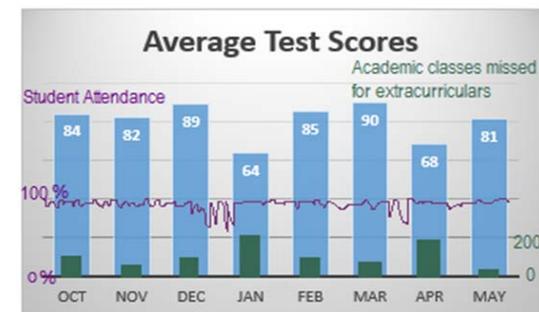
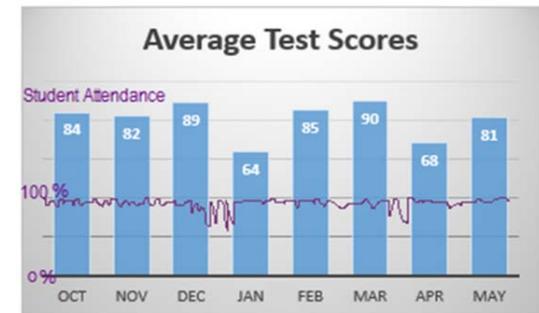
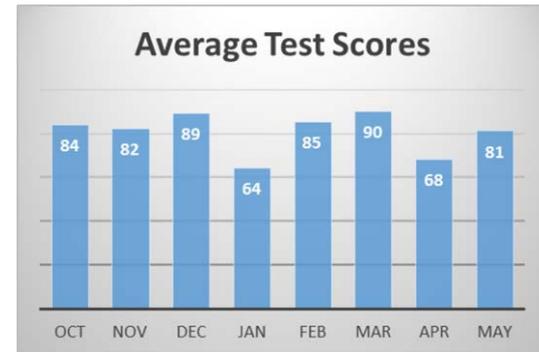
Preliminary data analysis suggests three rational, observation-based hypotheses that may need more formal study:

- (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.
- (3) Decreases in student attendance and increases in excused absences for extracurricular activities combined to result in lower average monthly test scores.



Ch. 2: Data Visualization to Advance Data Analysis

- **Take Away 1:** Attendance and excused absence patterns *might* have had a negative effect on the school's average test scores during specific months.
- **Take Away 2:** It is acceptable and *even encouraged* to take these factors into account preliminarily when planning test dates and reflecting on the test score data.
- **Take Away 3:** Although these factors appear to be correlated, they are not necessarily connected causally—and cannot be considered causes for lower scores without additional research.



Ch. 3: Data Visualization to Improve Communications

- Data visualization is essential for presenting information in a manner that communicates data meaning to a range of audiences—especially non-expert viewers.
- *4 key principles* ensure that data meaning is communicated effectively:
 - 1) *Show the data.*
 - 2) *Reduce the clutter.*
 - 3) *Integrate text and images.*
 - 4) *Portray data meaning accurately and ethically.*

Ch. 3: Data Visualization to Improve Communications

Key Principle 1: *Show the Data*

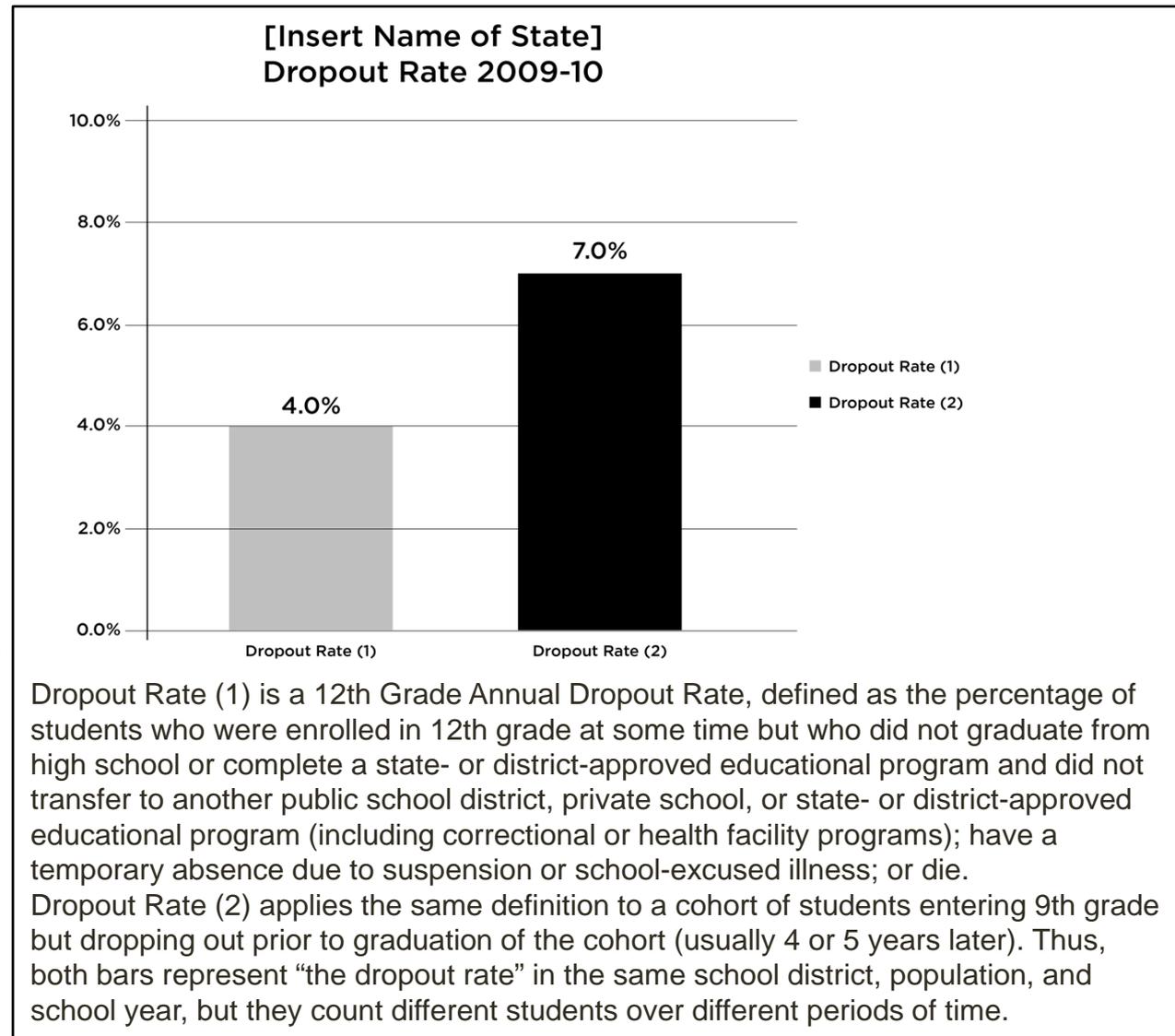
- Data values that underlie a visualization are important enough to be labeled because showing the data values improves reader comprehension.
- A corollary to the key principle of “show the data” is the need to include related information, including metadata, that is necessary to fully understand the data.

Ch. 3: Data Visualization to Improve Communications

Key Principle 1 *Show the Data*

Why might metadata be helpful?

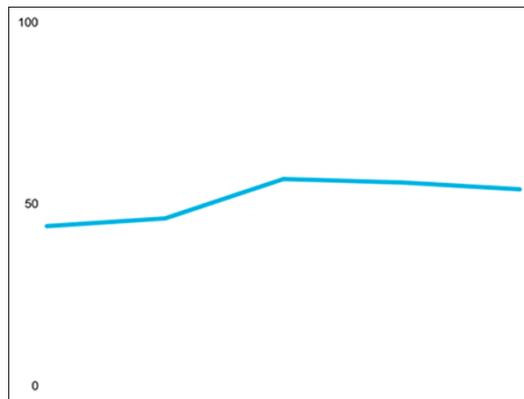
The two dropout rates are for the same school district, population, and year.



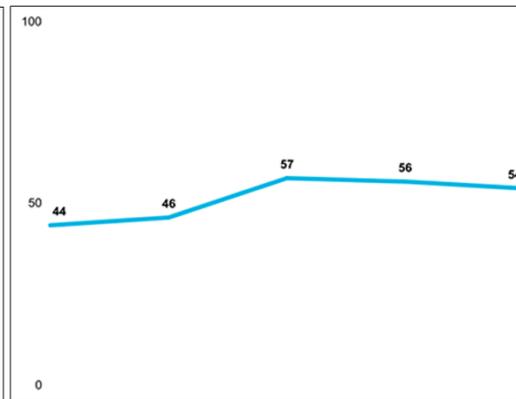
Ch. 3: Data Visualization to Improve Communications

Key Principle 2: *Reduce the Clutter*

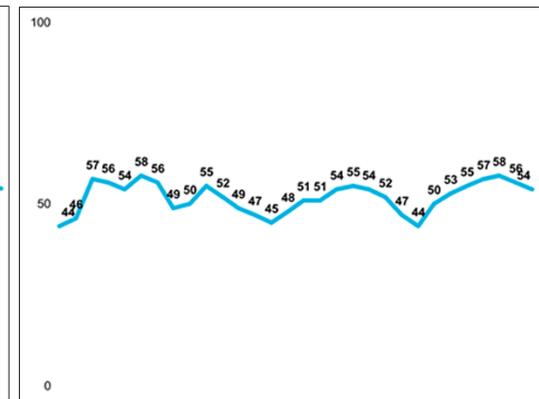
- Unnecessary bells and whistles, purely cosmetic features, and other extraneous information distract from the primary meaning of the data.
- Which figure below is most effective?
Strike a balance between showing the data and overwhelming the audience.



Version A



Version B



Version C

Ch. 3: Data Visualization to Improve Communications

Key Principle 3: *Integrate Images and Text*

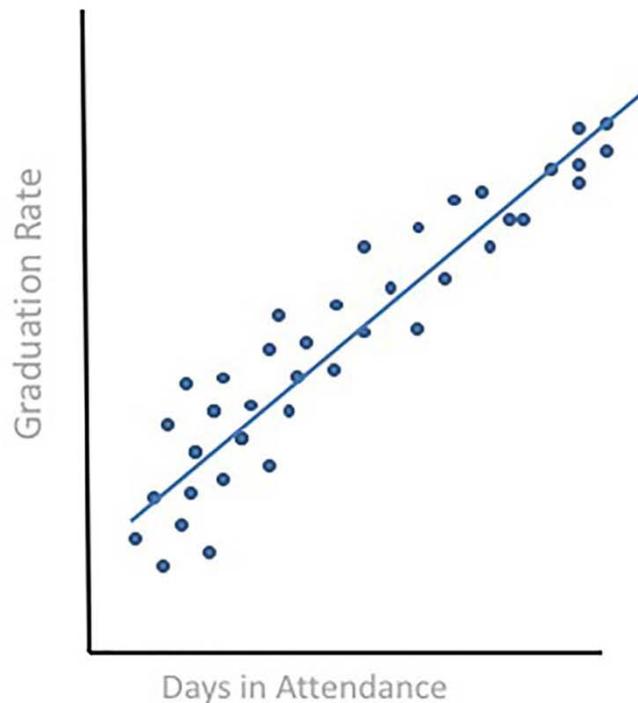
- Visualizations should stand on their own as a complete piece of information and be clearly connected to the text.
- Every aspect of imagery and text should point viewers toward a better understanding of the take-home message.

Ch. 3: Data Visualization to Improve Communications

Key Principle 3

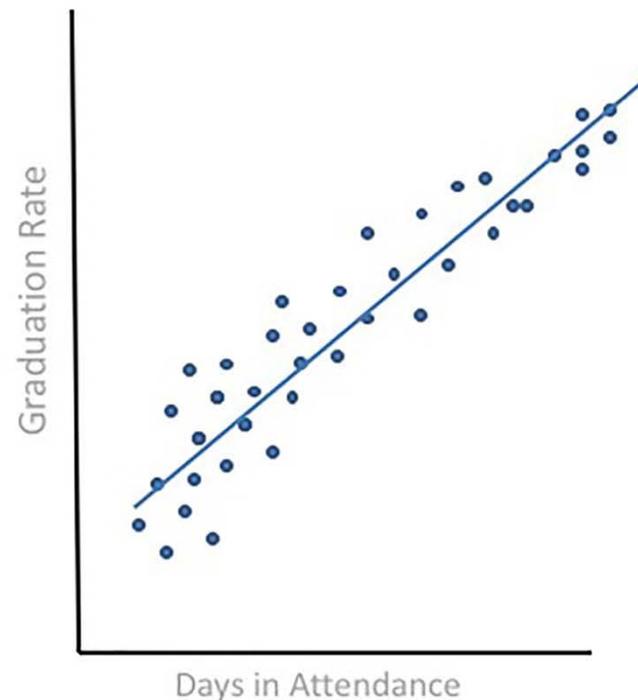
Integrate Images and Text

Correlation of school attendance and various relevant variables



Which figure title conveys meaning to a viewer?

High school attendance is an important predictor of student graduation in the Hypothetical Public Schools, 2009-2012.

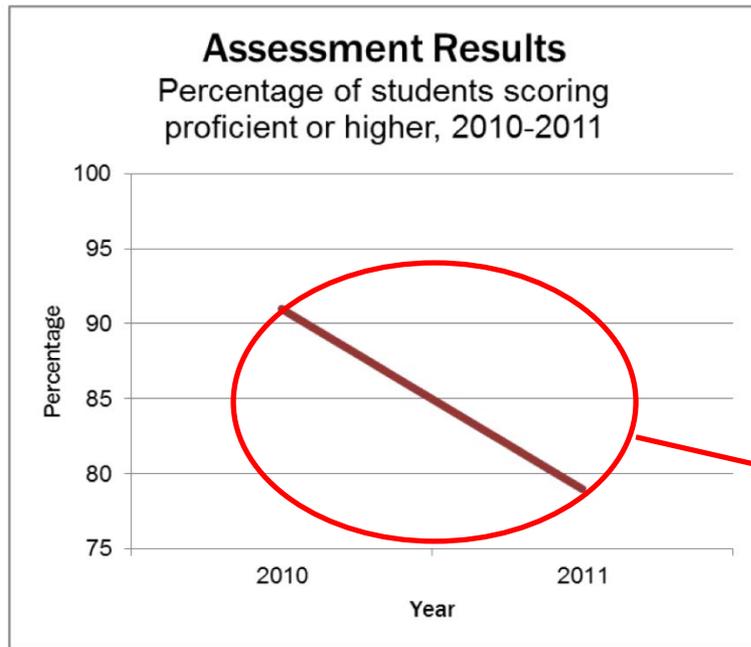


Ch. 3: Data Visualization to Improve Communications

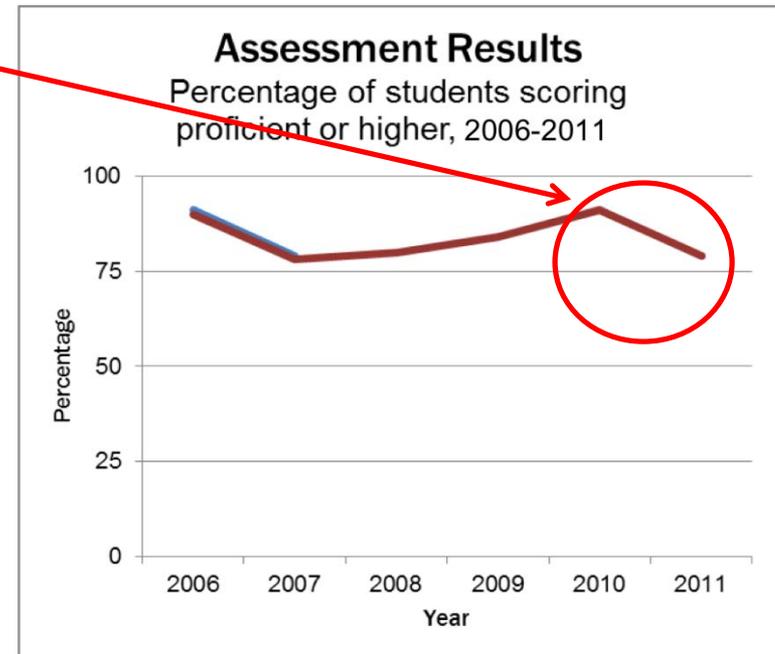
Key Principle 4: ***Portray the Data Accurately and Ethically***

- Even accurately presented data can be presented unethically, such as
 - limiting which data are seen (overemphasizing subsets or “patterns” in data by only showing parts of an axis);
 - suggesting a conclusion that does not reflect the data (referring to a “trend” that does not fully describe the data); and
 - manipulating how the data are presented visually (suggesting that certain types of data are continuous over time rather than discrete across time to suggest relationships that are not valid).

Ch. 3: Data Visualization to Improve Communications



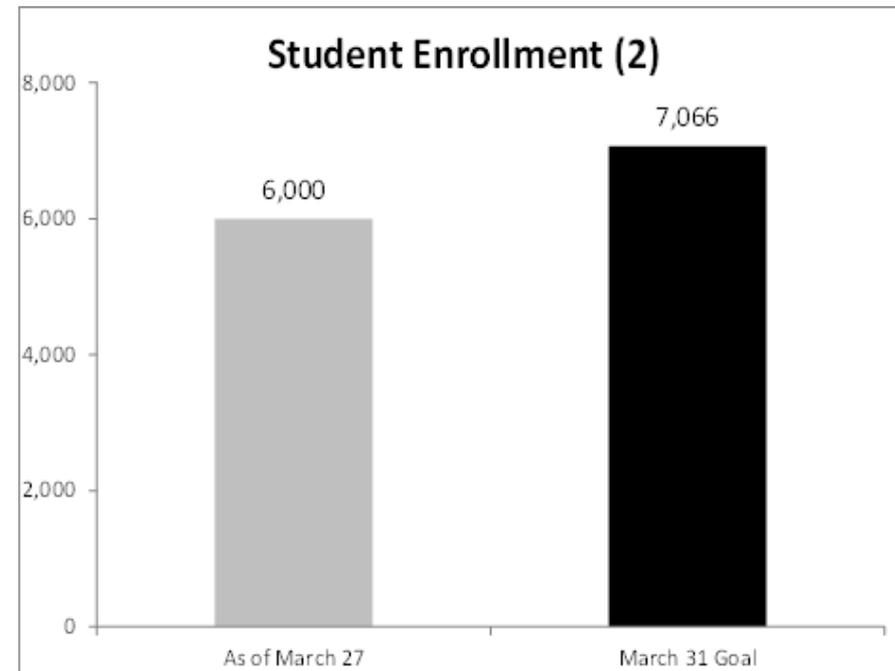
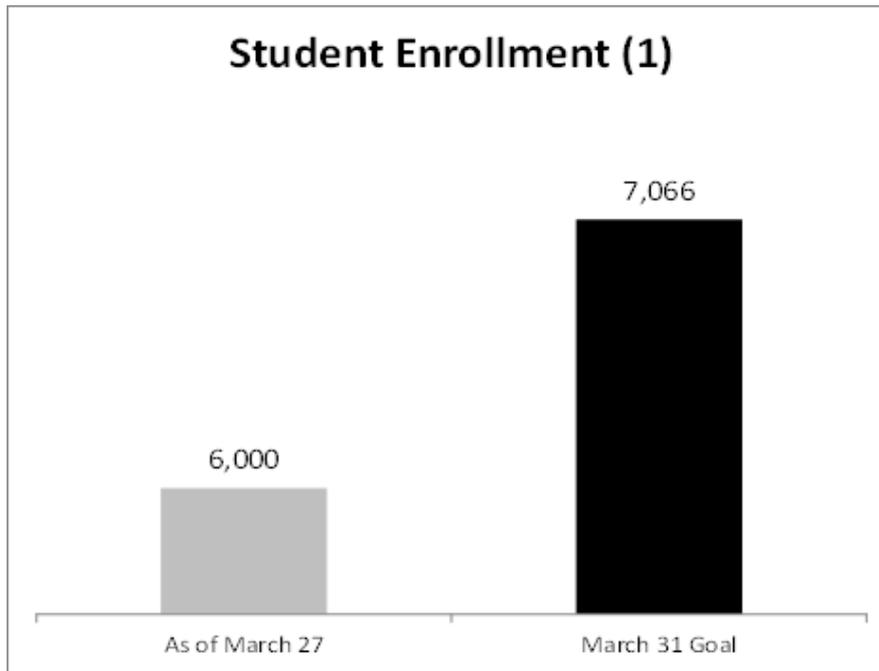
Key Principle 4
Portray the Data
Accurately and Ethically



Ch. 3: Data Visualization to Improve Communications

Key Principle 4

Portray the Data Accurately and Ethically



Ch. 3: Data Visualization to Improve Communications

Seven Recommended Practices for Data Visualization

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

Ch. 4: Implementing the Data Visualization Process

Data Visualization: A Six-Step Process

1. Question: Someone Needs Information
 - What is the question? Is this a one-time information need or a routine request that will likely be repeated?
2. Research: Data Exploration and Analysis
 - What data and analysis are needed? Are high-quality data available for relevant exploration and analysis?
3. Findings: Data Meaning/Answer
 - What is the “take-home message” from the data? That is, what is the core message in the data that you wish to convey?

Ch. 4: Implementing the Data Visualization Process

Data Visualization: A Six-Step Process

4. Customization: Audience Specific Messaging
 - Who is your audience? To whom is the message being conveyed? What is the most appropriate way to communicate with this audience?
5. Visualization: Present Data Meaning Clearly and Accurately
 - How will you present your message? That is, what is the most effective way to visualize the data for your audience?
6. User Feedback: Review and Refine Efforts
 - How can you ensure that your visualization is effective? Ask your users for feedback and iterate, iterate, and iterate based on that feedback.

Ch. 4: Implementing the Data Visualization Process

Step 1. Someone Needs Information

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

Step 2. Data Exploration and Analysis

A visit to the NCES website reveals the availability of the *EDFacts* Consolidated State Performance Report that includes public high school 4-year adjusted cohort graduation rate (ACGR) for the US, the 50 states and DC: School years 2010-11 to 2012-13 (see figure).

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

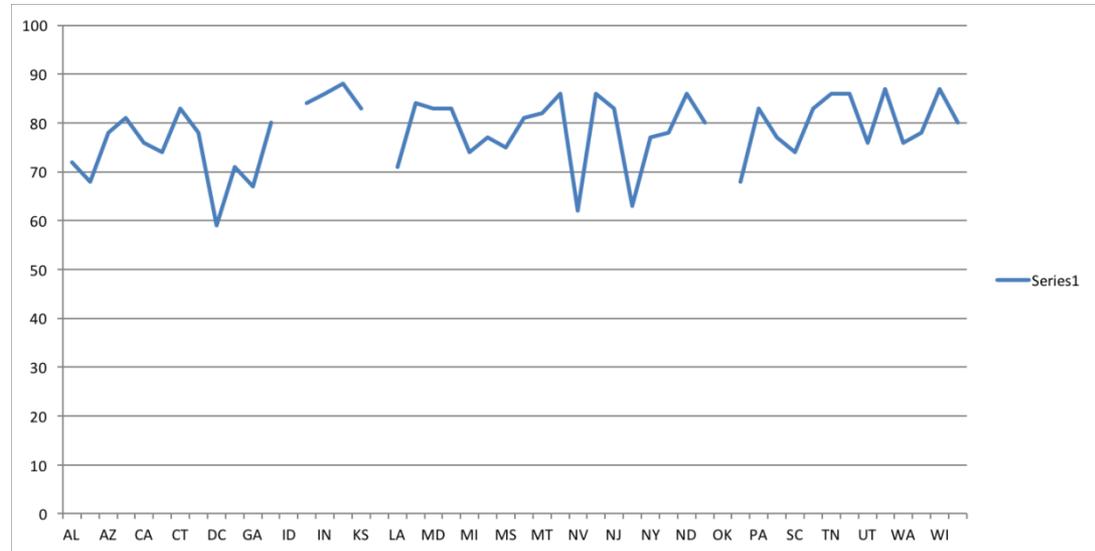
State	Adjusted Cohort Graduation Rate		
	2010-11	2011-12	2012-13
United States ¹	79	80	81
Alabama	72	75	80
Alaska	68	70	72
Arizona	78	76	75
Arkansas	81	84	85
California	76	79	80
Colorado	74	75	77
Connecticut	83	85	86
Delaware	78	80	80

Ch. 4: Implementing the Data Visualization Process

Step 3. Data Meaning/Answer

While the data presented in tabular form in table 4.1 are appropriate for some types of audiences, even seasoned analysts are likely to find it difficult to identify any patterns, trends, or cues in such a table.

The default setting in a spreadsheet tool produces a graph with many features that are likely to lead to misunderstanding or misinterpretation of the data (see figure).

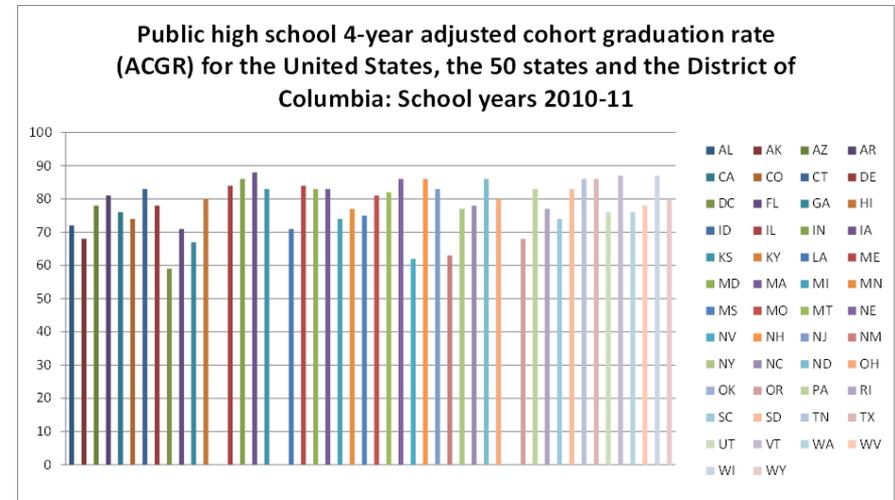
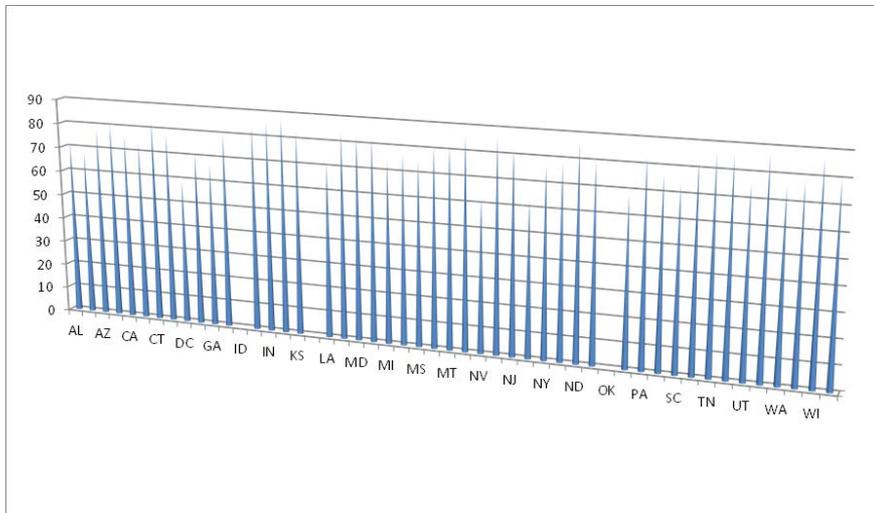
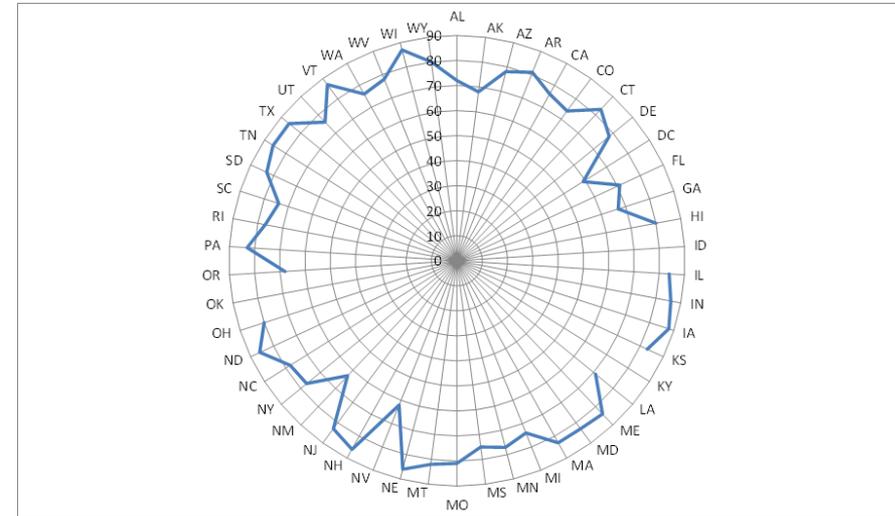


Ch. 4: Implementing the Data Visualization Process

Step 3. Data Meaning/Answer

Other default graphs are visually interesting but are more likely to obscure rather than clarify meaning.

- Why 3D for 2D data? What data values do the tips of the cones indicate?
- How does a radar chart clarify ACGR?
- Which bars map to which colors in the legend?

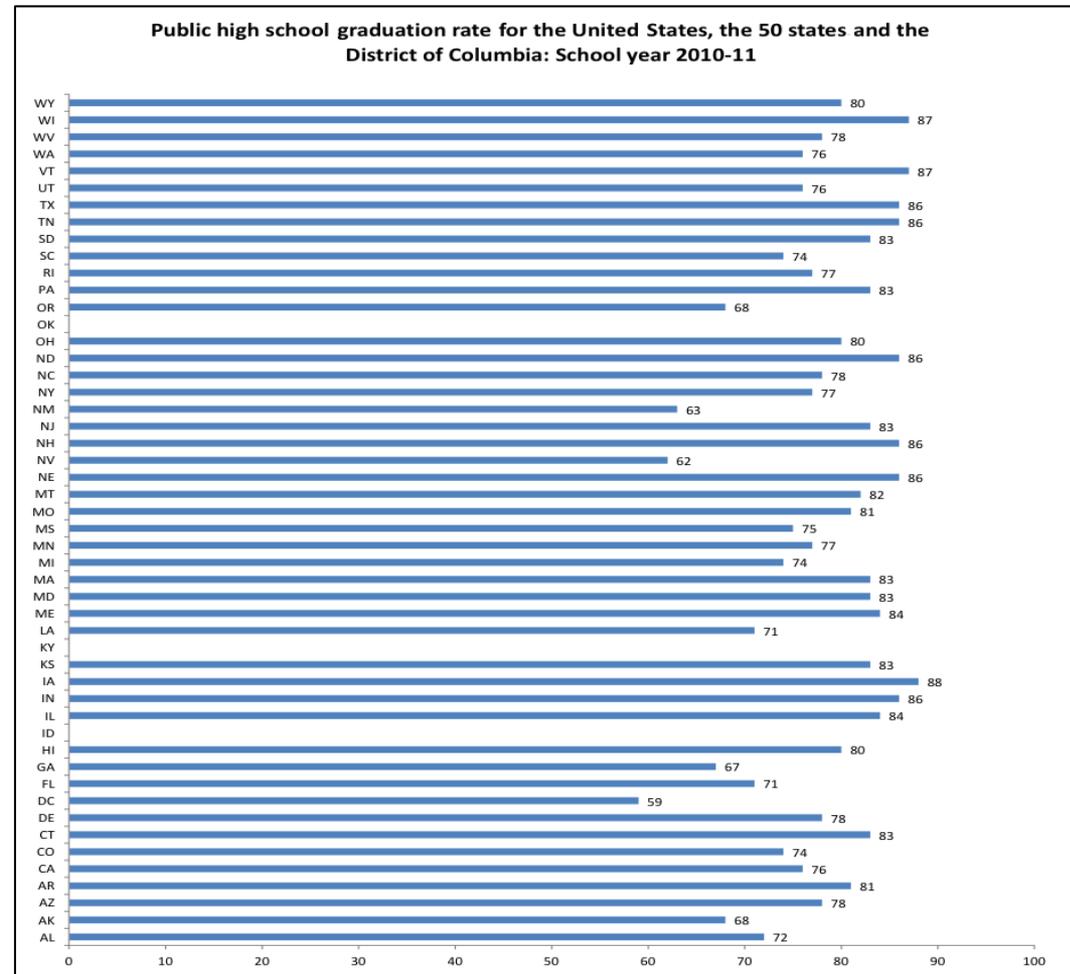


Ch. 4: Implementing the Data Visualization Process

Step 4. Customization: Audience-Specific Messaging

The question in step 1 originated from the general public. Such a broad audience indicates that designers can't assume any data, statistical, policy, or education-related expertise on the part of the viewer.

This audience warrants a “no training required” approach to visualization.



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>.
Table prepared January 2015.

Ch. 4: Implementing the Data Visualization Process

Step 5. Present Data Meaning Clearly and Accurately

The visual power of

- reordering the states from highest to lowest data values (recommendation 3)
- inserting a national average value (recommendation 4), and
- highlighting that national average in another color to simplify comparisons (recommendation 7)

will all contribute to better understanding of the take home message.

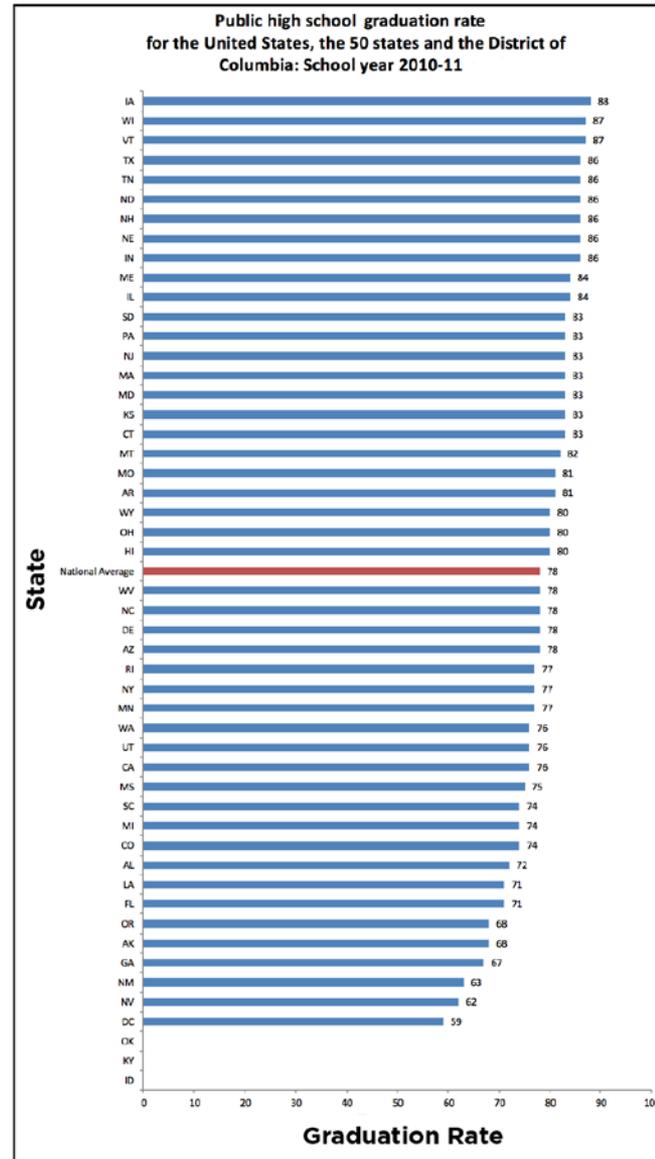


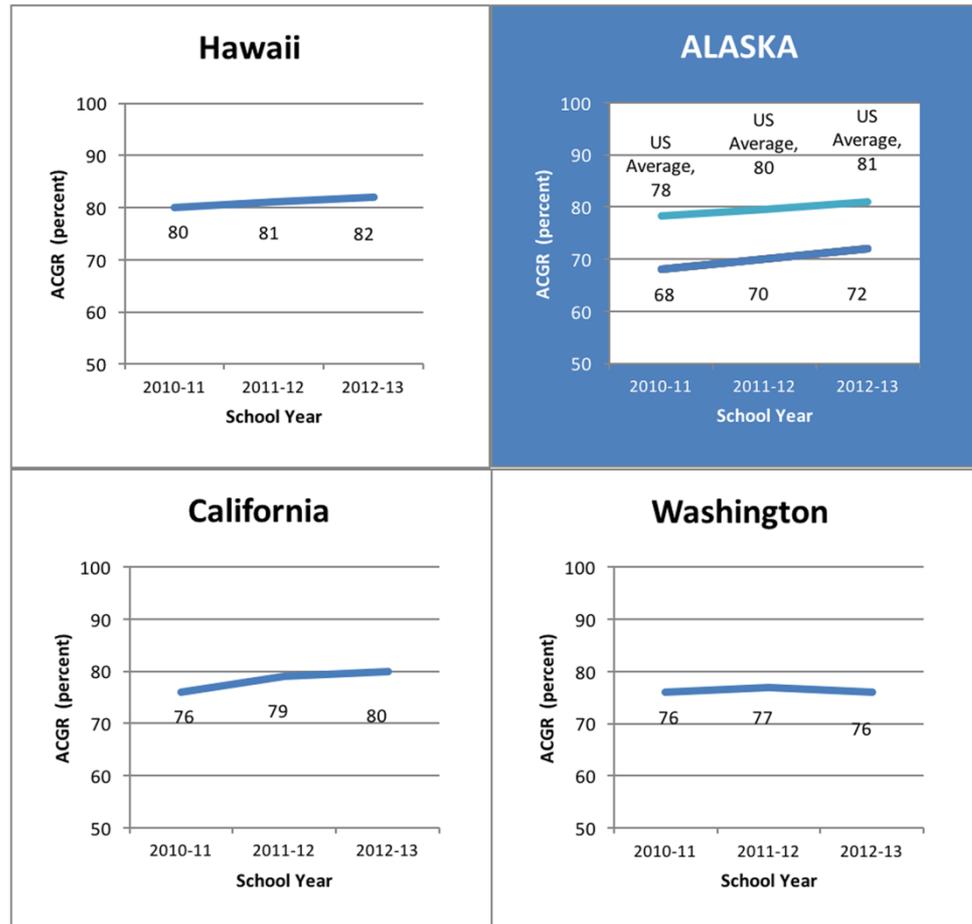
Figure X. The United States 4-year adjusted cohort graduation rate (ACGR) was estimated using both the reported 4-year ACGR data from reporting states and the District of Columbia and using imputed data for Idaho, Kentucky, and Oklahoma for school years 2010-11. The estimated United States ACGR includes these revisions. 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 data not being available in one or more of the school years shown. 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 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/ac/count/consolidated/index.html>. This table was prepared January 2015.

Ch. 4: Implementing the Data Visualization Process

Step 6. Review and Refine Efforts

Alaska graduation rates increase but lag behind national averages and western region peer states.



Summary

- 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 decisions about how our educational system serves students and communities.
- Data visualization is a critical component of the data analysis and communications process for many education stakeholders.

Questions?



Thank you

Download the *Forum Guide to Data Visualization: A Resource for Education Agencies* and access other free Forum resources at:

<http://nces.ed.gov/forum/publications.asp>

