National Institute for Learning Outcomes Assessment
Making Learning Outcomes Usable & Transparent

Assessment in Practice

Achieving Excellence in Data Visualization

Dr. Martha J. Larkin & Dr. George R. Larkin
Walden University

INTRODUCTION

Assessment, like most other areas of the education field, has experienced a dramatic increase in the amount of data available for analysis. The data deluge and new demands from the public for educational accountability have required educational assessment professionals to examine new ways to present assessment data and analysis. Technological advancements over the past few decades have revolutionized the way we gather, analyze, and present data. Access to data, analytic tools, and graphic software are available to almost anyone with a computer and an internet connection. Unfortunately, access does not always equal excellence. The new wealth of data and analysis represents both promises and problems for the educational assessment community. Assessment professionals are faced with the conundrum of having more data to communicate; however, they must present it in a manner that informs rather than overwhelms the stakeholders. Visualization offers one possible solution for presenting data in an understandable format and new technological tools have simplified the task of producing visual representations of data.

Sue and Griffin (2016) noted that “data visualization is a general term that describes the use of visual representations to help people analyze and interpret data” (p. 5). SAS.com defined it as “the presentation of data in a pictorial or graphical format.” As more and more data are collected and analyzed, decision makers at all levels welcome data visualization that enables them to see analytical results presented visually, find relevance among numerous variables, communicate concepts and hypotheses to others, and even predict the future.

Data visualization has become an important part of business intelligence and more recently has become an area of interest in academia. Increased amounts of data need to get to people in an appealing and useable form (Few, 2007). Also, 75% of what we know comes to us through our visual senses (Hurt, 2011).The primary uses of visual representations are to: (1) find and examine meaningful patterns to gain understanding, and (2) communicate that information to your audience in a manner that meets their needs (Azzam, Evergreen, Germuth, & Kistler, 2013; Evergreen, 2017; Few, 2004, 2006; Kirk, 2016). Achieving excellence in data visualization requires planning, knowledge of established principles, and the appropriate use of available tools.

PLANNING FOR EXCELLENCE IN DATA VISUALIZATION

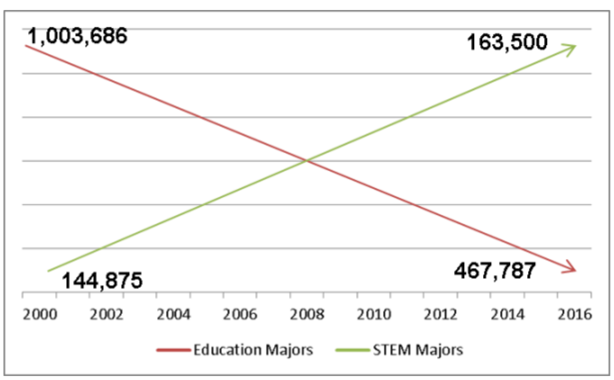
Achieving excellence in data visualization requires planning that addresses the decision makers' purpose for using the data and the display needs of the data set. Fry (2008) outlined six areas to consider when planning for data visualization.

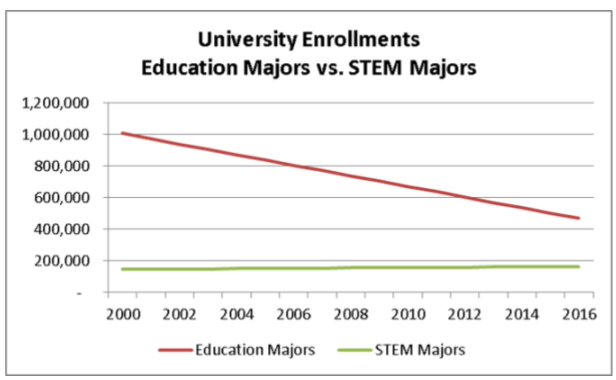
- The first area is dealing daily with *too much information* that results in information overload.
- Second, we are improving at *data collection* but collection alone is not enough. Data must be visualized for maximum benefit.
- The third refers to the lack of *sophisticated thinking about data* which is crucial for establishing meaning.
- The fourth indicates that *data never stay the same* because most real-world data are dynamic.
- Fifth, *what is the question* refers to technology enabling increased capabilities of creating and storing data to the extent that the data may easily be dis-associated from the purpose for collecting it.
- Sixth is *a combination of many disciplines*. Data are complex and need insights from diverse fields to acquire meaningful solutions.

PRINCIPLES FOR DATA VISUALIZATION

Much of this section is based on the work of Edward Rolf Tufte, an American statistician and professor emeritus of political science, statistics, and computer science at Yale University. He is noted for his writings on information design and as a pioneer in the field of data visualization. In order to achieve excellence in data visualization, Tufte formulated six principles of graphical integrity (Tufte, 2001).

Principle 1: The representation of numbers, as physically measured on the surface of the graph itself, should be directly proportional to the numerical quantities represented.



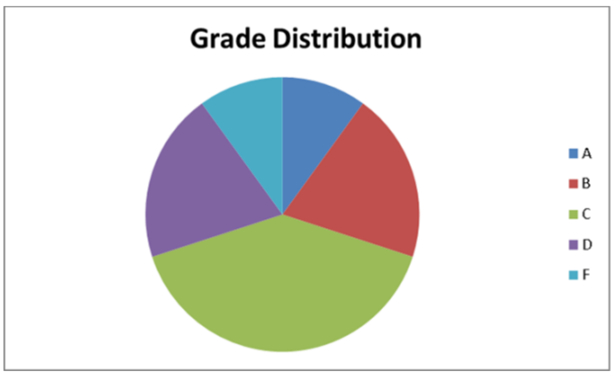


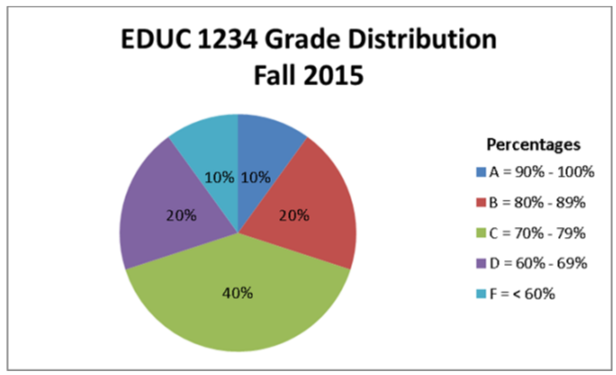
The two graphs above present hypothetical information regarding enrollments for Education and STEM majors. The graph on the left is an example of a chart that violates Tufte's Principle 1. The reader's initial reaction to the graph on the left might be that in 2016 the Universities enrolled more STEM majors than Education majors. The reader's distorted perception of reality occurs because the physical measurement of the numbers on the graph is not in direct proportion to the numerical quantities being measured. Between 2000 and 2016 the number of

STEM majors increased by 11.4% while the number of Education majors decreased by 53.4%; however, there are still over two and one-half times more Education majors than STEM majors.

The graph on the right is a more accurate representation of the same hypothetical data. The physical measure of the numbers on the graph to the right is in direct proportion to the numerical quantities being represented. The graph on the right accurately displays that from 2000 to 2016 there was a sharp decline in the number of Education majors and a modest increase in the number of STEM majors. The graph on the right also shows that the number of STEM majors still is far fewer than the number of Education majors.

Principle 2: Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity.

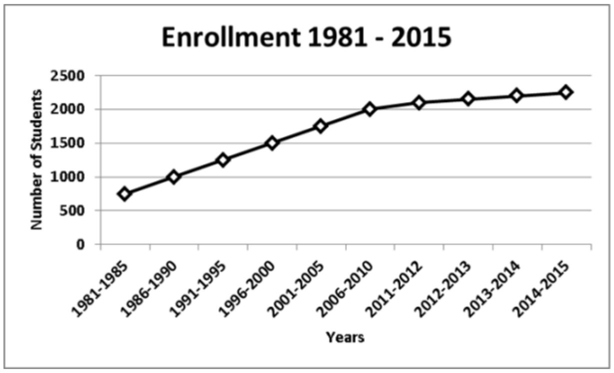


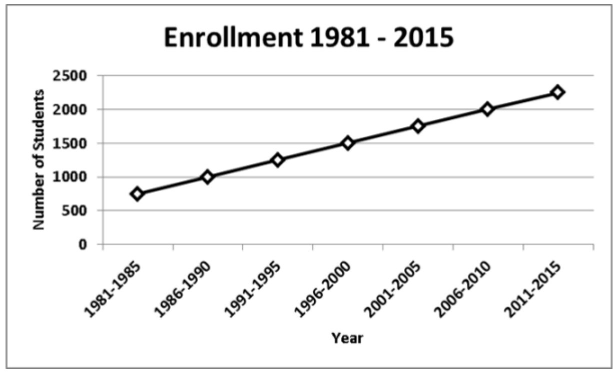


The pie chart on the left is a poor representation of Principle 2 because minimal labeling results in potential graphical distortion and ambiguity of the hypothetical grade distribution data. All the reader can determine is that the chart shows a grade distribution ranging from A though F with each colored section of the pie proportionally representing a corresponding letter grade. Unfortunately, the reader is left with many unanswered questions. What is the course for the grade distribution and what term is represented? Numerically what constitutes each letter grade? Perhaps an analogy is shopping for over the counter medications in a store. A consumer would not think of buying an unlabeled product because he or she would not know the contents of the unlabeled container and the intended use for the product. Likewise, unlabeled data can be equally useless and potentially harmful.

The pie chart on the right is a better representation of Principle 2 because it shows clear, detailed, and thorough labeling that provides the reader with more information about the data that can lead to accurate understanding of the data. Not only can the reader determine the course (EDUC 1234) and term (Fall 2015) for the data, but also what percentages constitutes each letter grade (e.g., A= 90-100%, B= 80-89%, C=70-79%, etc.) . In addition the reader can glean from the labels on each colored pie section what percentage of the total pie is represented by each letter grade (e.g., 10% A's, 20% B's, 40% C's, etc.). When appropriate, a clear, detailed, and thoroughly labeled chart might be used for data comparison inside and outside the institution which could not be done accurately with a minimally or poorly labeled chart.

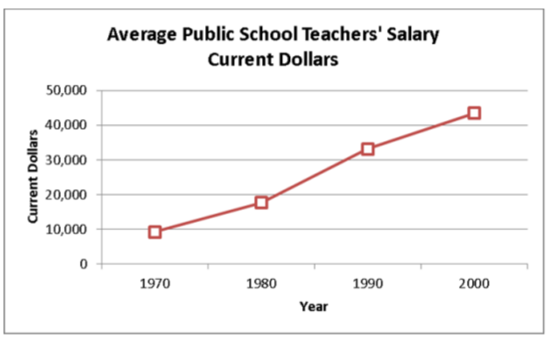
Principle 3: Show data variation, not design variation.

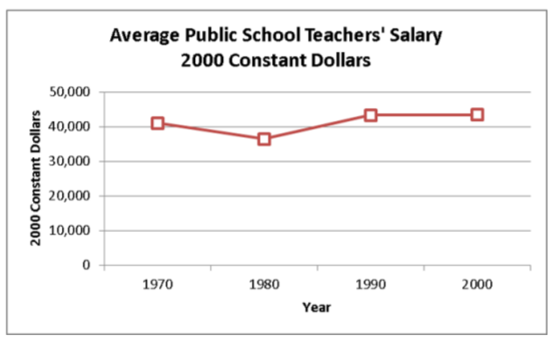




At first glance, the two line graphs showing hypothetical enrollment data from 1981-2015 appear very similar. The vertical axis representing the number of students is identical, but the difference between the two line graphs occurs in the horizontal axis. Both line graphs begin with multi-year intervals and continue with equally spaced intervals through 2006-2010. The line graph on the left begins labeling in yearly increments starting with 2011-2012 and continues the yearly increments for the remainder of the horizontal axis. The line graph on the right continues consistent increments for the remainder of the horizontal axis and does not change to yearly increments in midstream as did the graph on the left. Obviously, the graph on the left is a poor representation of Principle 3 because it includes design variation that can distort the data. The graph on the right is a better example of adherence to Principle 3 because it uses consistent time increments on the horizontal axis for the entire graph. As a result the reader is able to focus on the accurate data variation on the line graph on the right because there is no design variation.

Principle 4: In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.

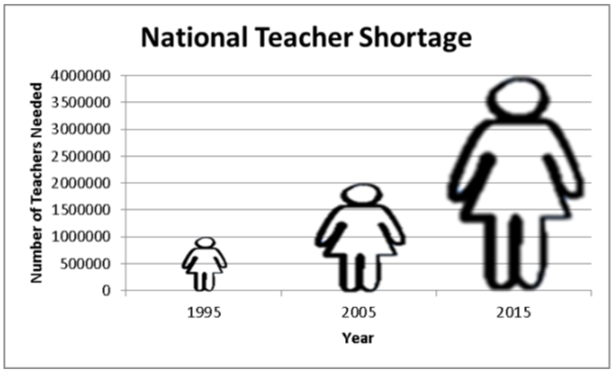


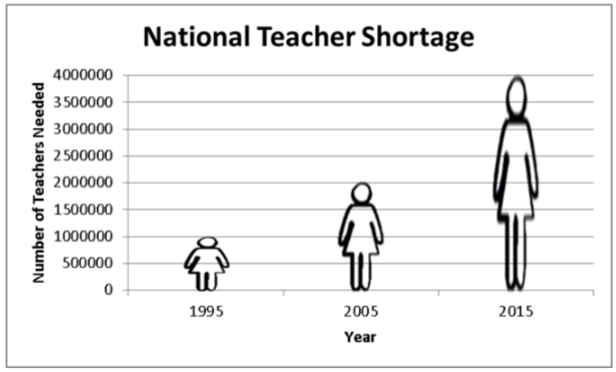


The value of money changes over time. A dollar today does not have the same purchasing power that it did when we were younger. The graph on the left shows the increase in the average public school teachers' salaries from 1970 to 2000 in current dollars. The graph on the right displays the same information in 2000 constant dollars. The increases shown on the right graph appear much less dramatic that the increases shown on the left graph. Displaying monetary values over time in constant dollars or some other standardized measure gives the reader a more accurate representation of the increases and decreases that occurred. The data for these graphs were obtained from the National Center for

Education Statistics website:
https://nces.ed.gov/programs/digest/d09/tables/dt09_078.asphttps://nces.ed.gov/programs/digest/d09/tables/dt09_07 8.asp. The CPI inflation calculator at U.S. Department of Labor Bureau of Labor Statistics www.bls.gov/data/inflation_calculator.htm was used to determine the constant value of money.

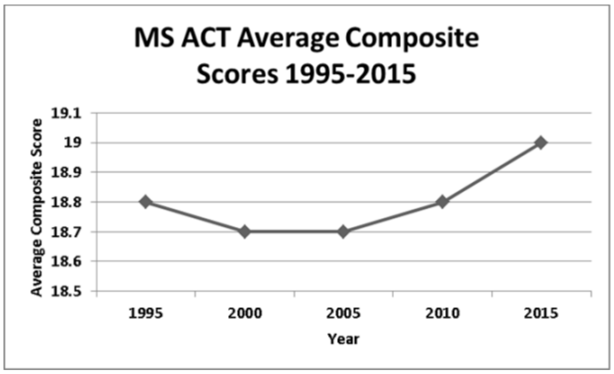
Principle 5: The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.

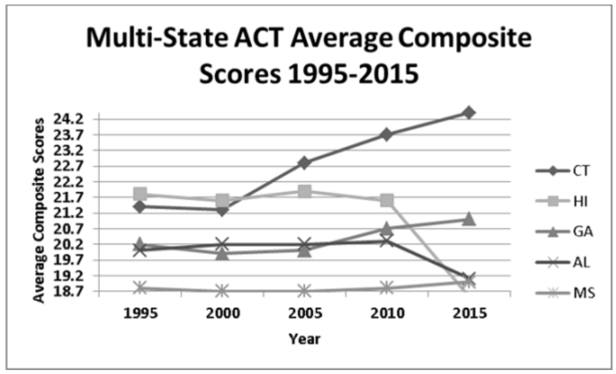




The two bar graphs above display the same hypothetical data regarding an increase in the teacher shortage that our nation faced between 1995 and 2015. Teacher shortage is a single variable and should be depicted in a single dimension. The graph on the left is misleading because it shows the variable with increases in both the vertical and horizontal dimensions of the teacher icon. The graph on the right is a better illustration of Principle 5 because it shows an increase only in the vertical dimension. Graphic integrity should not be sacrificed for artistic sophistication.

Principle 6: Graphics must not quote data out of context.





Both line graphs represent hypothetical ACT average composite scores for 1995-2015. The difference is that the graph on the left shows scores for only one state, MS, whereas the graph on the right shows the scores for a total of five states. If the reader views the graph reporting data for only one state, it is difficult to determine what the data mean. The scores for that state appear to be somewhat stable during the time frame represented with a slight increase beginning somewhere between 2005 and 2010, so the conclusion might be that the state of MS is holding its own for the ACT average composite scores. By viewing the graph on the right the reader can determine how any given state is doing in comparison to the others. For example, MS has the lowest scores overall of the five states represented for 1995-2015, yet the lines for MS, AL, GA, and HI all remain relatively stable from 1995-2010 with more movement between 2010 and 2015. MS and GAs scores were trending upward between 2010 and 2015 whereas AL and HI scores were trending downward during that time. CT also had upward trending but it began in 2000 which was sooner than the upward or downward trending of the other four states. The line graph on the right illustrates how much more information can be gleaned by viewing data in context and how the perspective could change substantially from viewing isolated data to viewing it in context.

TOOLS FOR DATA VISUALIZATION

Several tools are available to assist those interested in displaying data in a visual format. The charts function of Microsoft Excel is one of the most commonly used data visualization tools. The charts function has a wide variety of options for data visualization and has the capacity to produce excellent results. As with most tools, the quality of the visual produced by Excel is more a function of the skill level of the chart maker than the technical capabilities of the software. Unfortunately, the Microsoft website offers very limited resources regarding the principles of good data visualization. Other data analysis tools such as SAS, STAT, and Tableau have data visualization tools built into the software and their websites offer some excellent resources regarding best practices in data visualization.

Environmental Systems Research Institute (ESRI) is a worldwide leader in the field of Geographic Information Systems (GIS). Software produced by ESRI allows users to store, analyze, and display data that contain geographic components. The use of GIS in the educational assessment field should increase as geographically referenced data and GIS software become more available.

William Jacoby, a political science professor at Michigan State University, maintains a website that contains several useful data visualization resources. Finally, Coursera and several other online venues offer instruction on data visualization. The website URLs for Coursera and all the other resources mentioned here are listed below.

- Excel https://www.microsoft.com/en-us
- SAS http://www.sas.com/en_us/insights/big-data/data-visualization.html
- STATA http://www.stata.com/
- Tableau http://www.tableau.com/
- GIS ESRI http://www.esri.com/
- William Jacoby http://polisci.msu.edu/jacoby/
- Coursera https://www.coursera.org/course/datavisualization

CONCLUSION

Data visualization is one means of increasing the impact of assessment in the field of education. It enables educators to communicate findings in an understandable format and new technological tools have simplified the task of producing visual representations of data. A key to achieving excellence in data visualization is planning that addresses the decision makers' purpose for using the data and the display needs of the data set. Also, established principles provide a guide for maintaining graphical integrity in the data visualization process. Finally, examples illustrate how findings can be communicated effectively. The planning considerations, established principles, and examples outlined in this article provide a starting point for those seeking to improve their data visualization skills.

REFERENCES

Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. In T. Azzam & S. Evergreen (Eds.), Data visualization, part 1*. New Directions for Evaluation, 139,* 7-32.

Evergreen, S. D. H. (2017). *Effective data visualization: The right chart for the right data*. Thousand Oaks, CA: Sage Publications, Inc.

Few, S. (2004, September 4). *Common mistakes in data presentation*. Retrieved from www.perceptualedge.com/articles/ie/data_presentation.pdf

Few, S, (2006, April 11). *The surest path to visual discovery*. Retrieved from www.perceptualedge.com/articles/b-eye/path_to_visual_discovery.pdf

Few, S. (2007, January 10). *Data visualization: Past, present, and future. Cognos Innovation Center for Performance Management*. Retrieved from www.perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf

Fry, B. (2008). *Visualizing data*. Sebastopol, CA: O'Reilly Media, Inc.

Hurt, J. (2011, July 26). *We are all visual learners*. Retrieved from velvetchainsaw.com/2011/07/26/all-visual-learners/

Kirk, A. (2016). *Data visualization: A handbook for data driven design*. Thousand Oaks, CA: Sage Publications, Inc.

National Center for Education Statistics. (n.d.). Digest of Education Statistics. Retrieved from https://nces.ed.gov/programs/digest/d09/tables/dt09_078.asp.

SAS (n.d.). *Data visualization: What it is and why it matters*. SAS Institute, Inc. Retrieved from www.sas.com/en_us/insights/big-data/data-visualization.html

Sue, V. M., & Griffin, M. T. (2016). *Data visualization & presentation with Microsoft Office*. Thousand Oaks, CA: Sage Publications, Inc.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

U.S. Department of Labor Bureau of Labor Statistics (n.d.). *CPI Inflation calculator*. Retrieved from www.bls.gov/data/inflation_calculator.htm