

The Use and Potential Misuse of Data in Higher Education

A Compilation of Examples

Karen L. Webber, Ph.D.

RAPID INCREASES IN TECHNOLOGY have led to a heightened demand for and use of data collection and data visualization software. There are a number of data analyses or visualisations that, when presented without the proper context or that do not follow principles of good graphic design, have resulted in inaccurate conclusions. Principles of cognition and good graphic design are available a number of scholars, including Amos Tversky, Daniel Kahneman, Edward Tufte, Stephen Kosslyn, and Alberto Cairo. In addition to good graphic design, context is critical for correct interpretation, as there are many questions around the legitimacy, intentionality, and even the ideology of data use within higher education (Calderon 2015). Inaccurate conclusions can lead to incorrect discussions for possible program development and/or policy changes. You may wish to read more detail from several important scholars; see references at the end of this document. Below are some examples found in books or on the internet that may help you see the possibility of misrepresentation or misunderstanding. I do not verify the accuracy of these images or charts. Many were retrieved from public sources, and I include a source or reference for each.

Mindful of tenets put forth by the above scholars, I believe the potential misuse or misunderstanding of data in higher education can occur in six categories: 1) incorrect statistical treatment of data; 2) not using consistent data definitions; 3) incorrect use of color, scale, and size/proportion; 4) incorrect charts/graphs; 5) not representing the proper context; and 6) multiple challenges, most often with infographics. Although I am not an expert on data visualizations, the examples below seek to remind others of the principles of good graphic design that can help ensure clear visuals that will (hopefully) lead to less or no misinterpretation.

Examples of Poor Visuals That Do Not Follow Good Design Principles

1. Statistical Treatment of Data

1a. Misinterpretation based on different calculations used:

**Cumulative Amount Borrowed for Graduate Education
U.S. Institutions, 2012**

Graduate Degree Recipients	2011-2012		
Graduate Degree	% Borrowing	Mean Debt (All)	Mean Debt (Borrowers)
Total	63.85	34,412.80	53,894.33
Degree Type			
Master's degree (all)	62.95	25,075.66	39,837.16
Professional degree (all)	84.76	109,385.40	129,053.20
Doctoral degree (all)	49.72	36,587.93	73,582.36
Gender			
Male	62.97	34,778.65	55,234.19
Female	64.45	34,167.62	53,017.00

Source: NPSAS 2012, from Webber & Burns, 2018

Figures for Mean Debt ALL versus BORROWERS can be significantly different. For example, the mean debt for students completing a Master's degree in 2011-12 based on ALL students in that group is \$25,076. However, only 62.95% of the students in that group borrowed for education, and when one examines the mean debt by BORROWERS ONLY, the value increases to \$39,837. It is important to know how a mean or average score were derived.

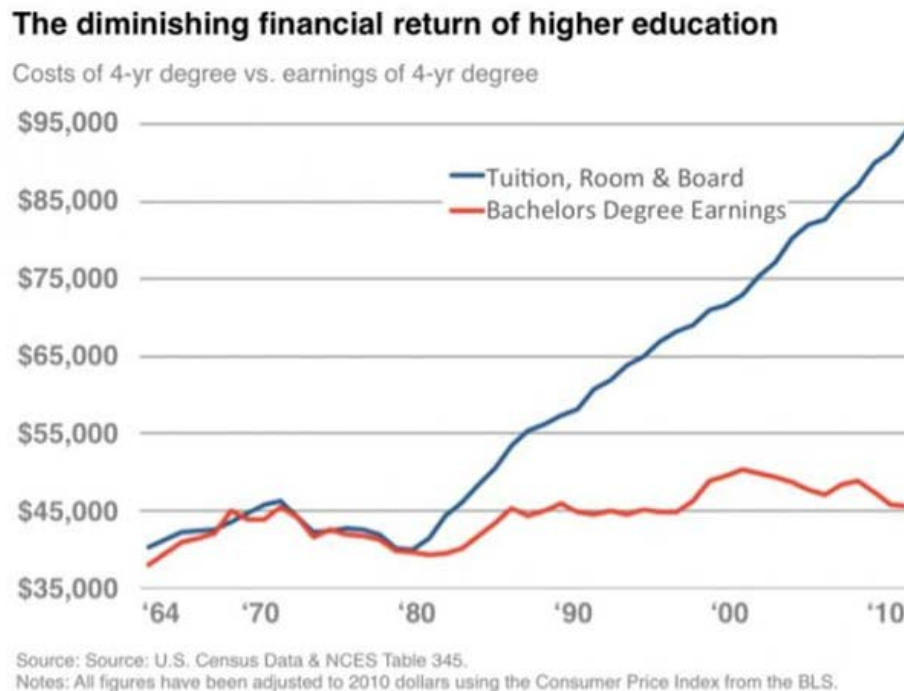
1b. Infographic with Pie Chart That Totals Greater than 100%



Infographic retrieved from Chibani (2018)

Pie charts are designed to show proportions that total 100, so on first review, the pie chart below may be confusing to some readers. It appears that the three items shown in the pie chart are separate questions, but if so, it would be better to show them separately rather than in one pie chart.

1c. Cost Vs. Benefit of a College Education – The Diminishing Financial Return of a College Education

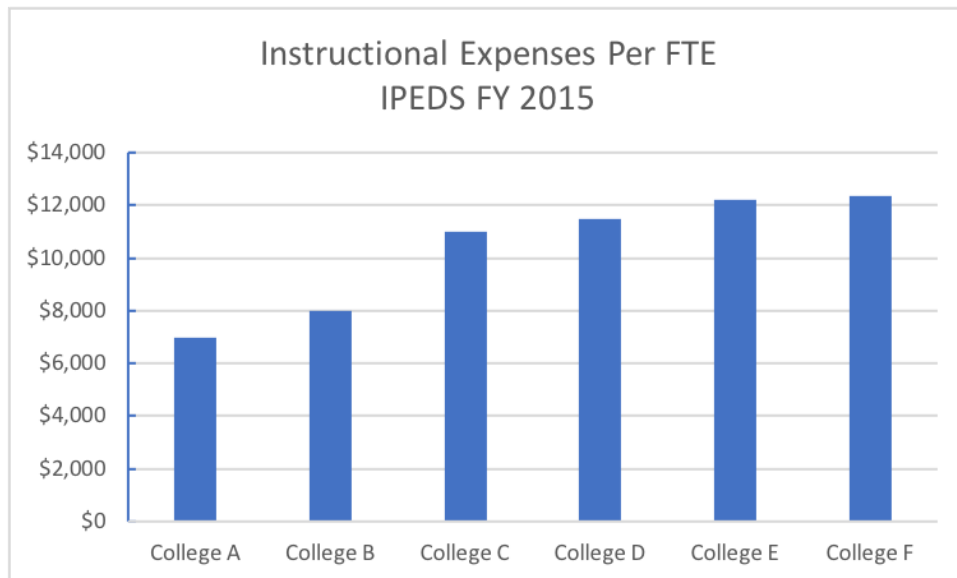


<https://io9.gizmodo.com/11-most-useless-and-misleading-infographics-on-the-inte-1688239674>

This chart from a *Business Insider* article examines the diminishing returns of a college education, but it leaves out a very important fact: that the prospects of non-college graduates are even worse than those of graduates. This chart omits the US Dept of Education numbers on the fact that non-college-graduates can expect a significantly smaller amount of lifetime earnings than graduates.

2. Misinterpretation Due to Inconsistent Data Definitions

- 2a. IPEDS Definition for Salaries and Wages — Definition says: “amounts paid as compensation for services...for department research and public service...includes compensation for academic instruction, occupational and vocational instruction, community education, adult education, and remedial instruction.”



Source: fictitious data mirroring IPEDS

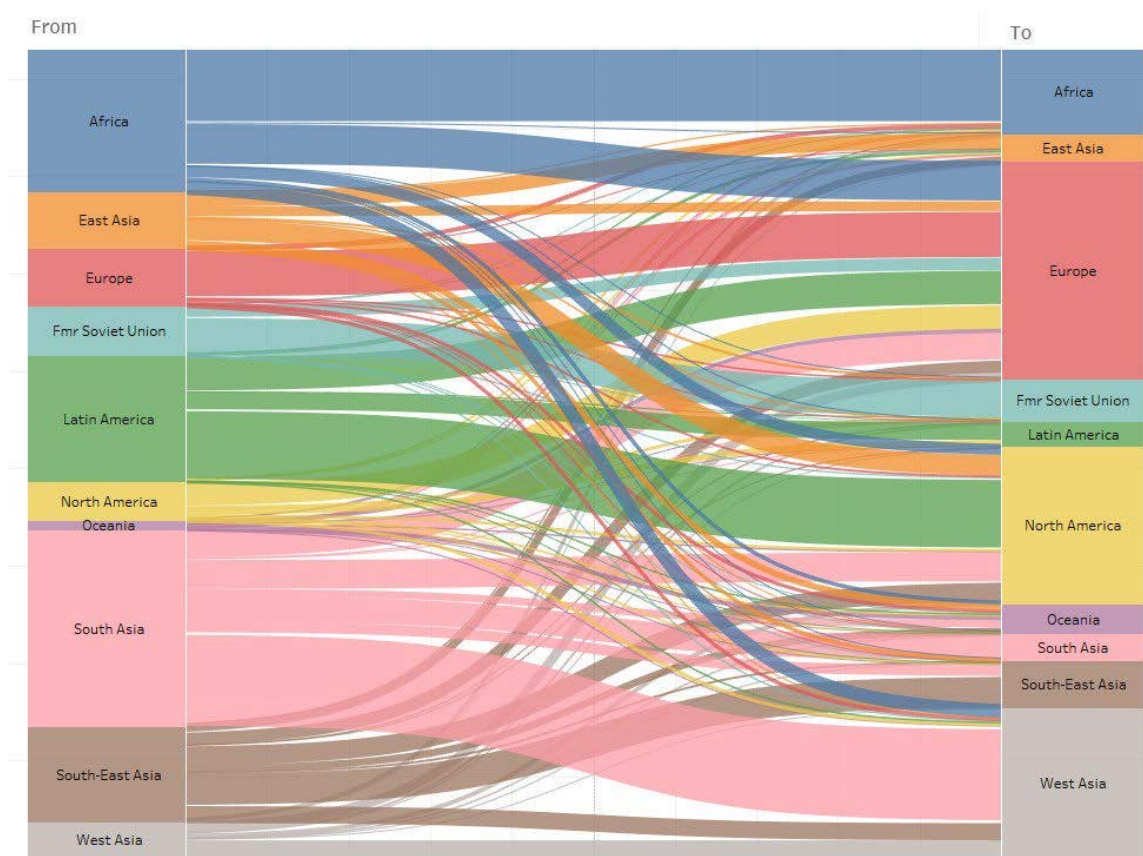
However, some of these institutions include 100% of faculty member salary in instruction while others apportion based on a budgeted percentage. Therefore, the comparison is flawed.

3. Use of Color, Scale, Size, Proportion

3a. Too Many Colors

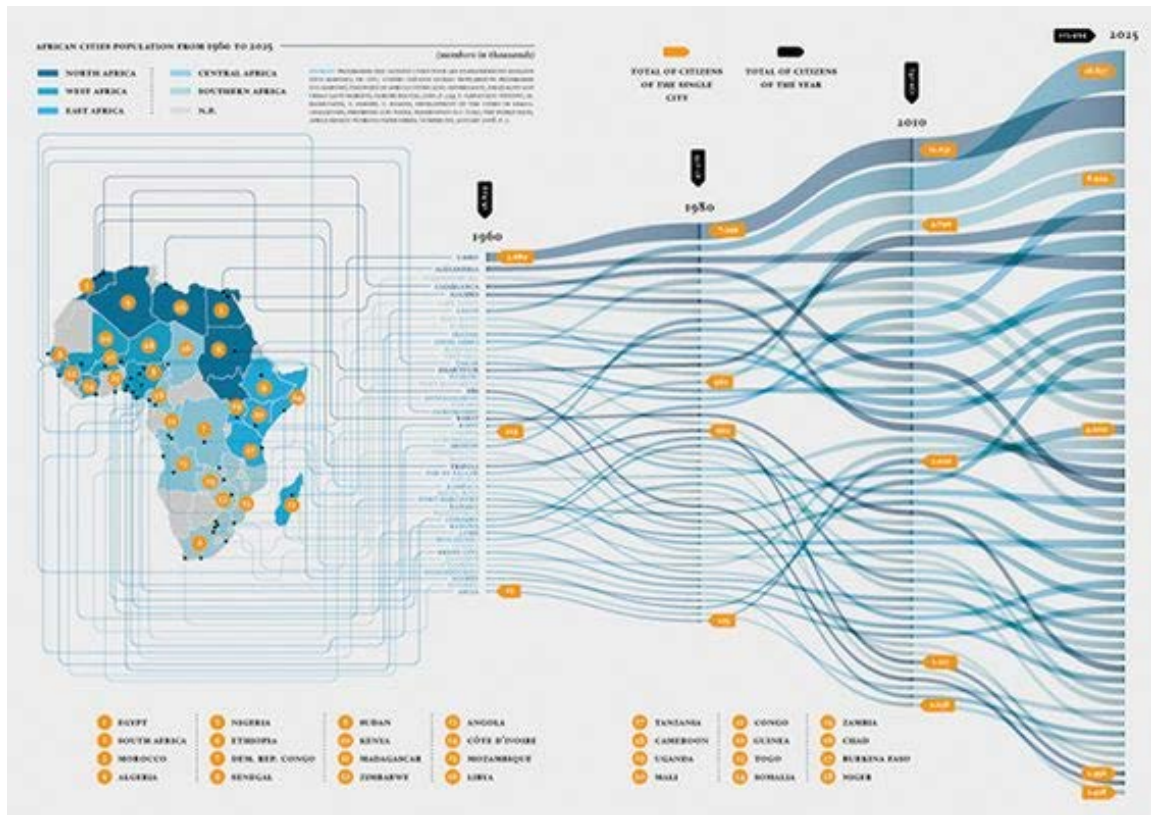
The Sankey below is a frequently-used type of diagram to visually show movement across two or more points. While the intent in this diagram is to examine flow across countries, there are so many data points and similar color chords, that it is difficult to follow some of the flow lines. Perhaps a Sankey with a small number of initial countries would offer better ability to see the flow to from one country to another.

Global Migration Patterns 2005 - 2009



Source: Quantum Analytics AG on Twitter: Global migration patterns 2005 - 2009

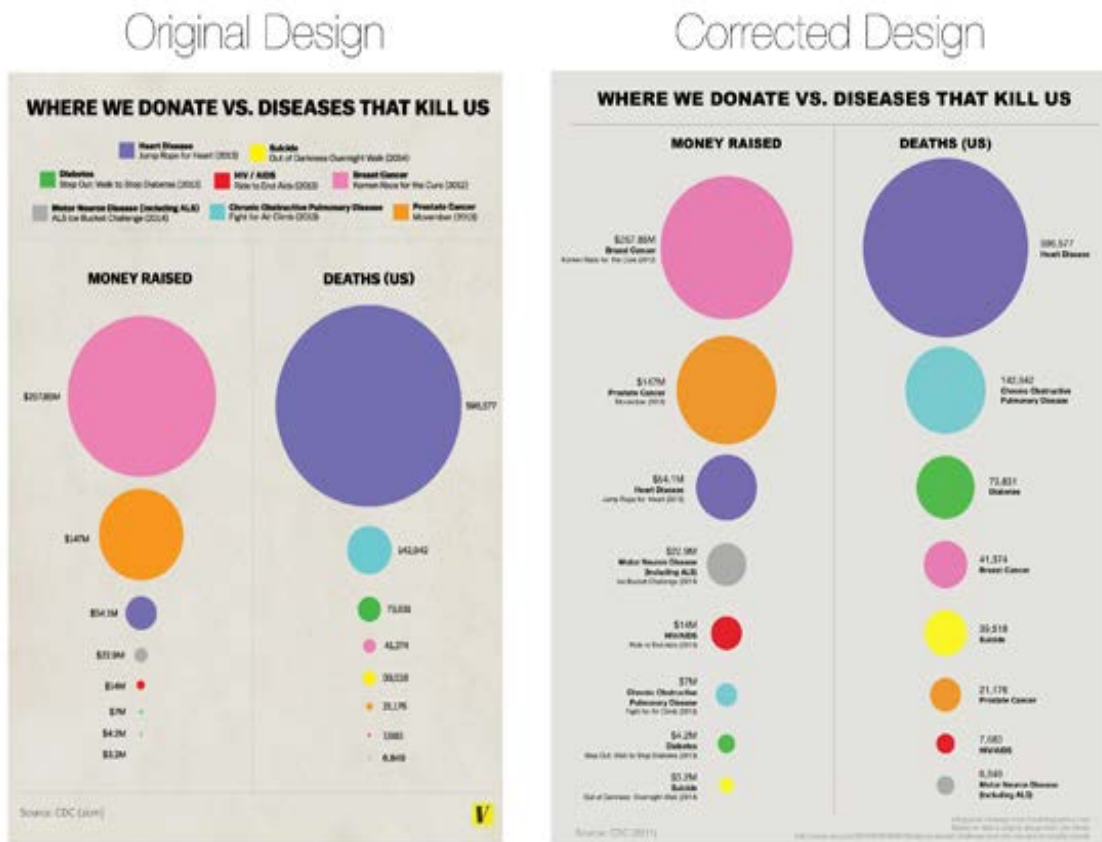
Too Little Color



Source: <https://www.behance.net/gallery/20553607/AFRICA-Big-Change-Big-Chance-Triennale-di-Milano>

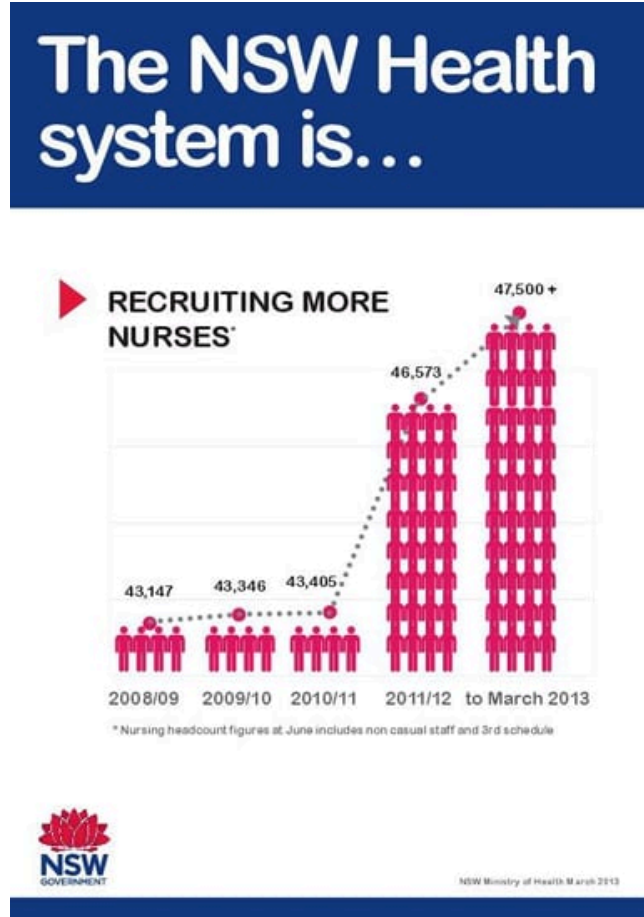
3b. Comparisons Using Incorrect Size Proportions

The scale of size must accurately represent the data. Below, the size of the bubbles do not accurately reflect the relationship between the amount of money donated to the treatment of a disease and the numbers of deaths caused by it.



Read more at <http://blog.visme.co/bad-infographics/#Zgib2ZPyLEvq3FAe.99>

3d. Inconsistent Use of Scale



Sources:

<https://io9.gizmodo.com/11-most-useless-and-misleading-infographics-on-the-inte-1688239674>

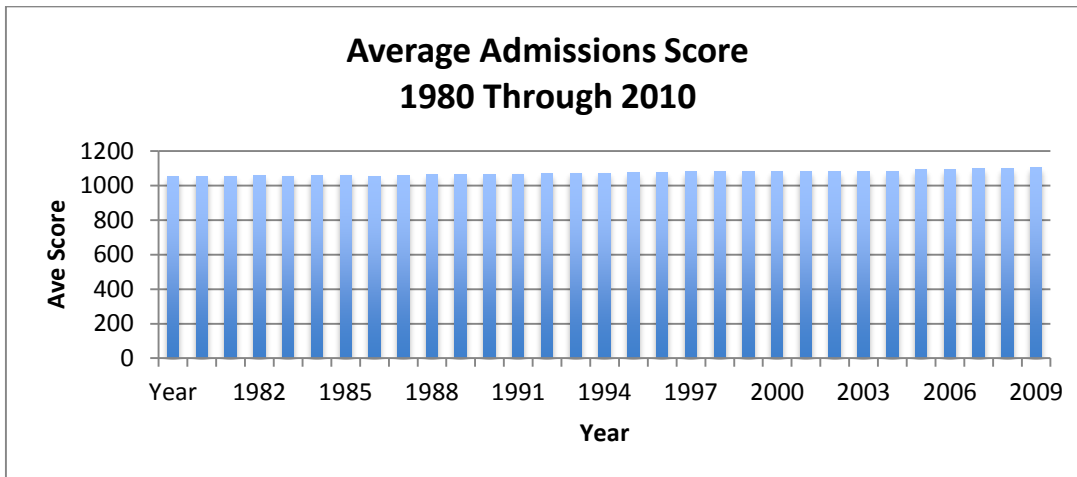
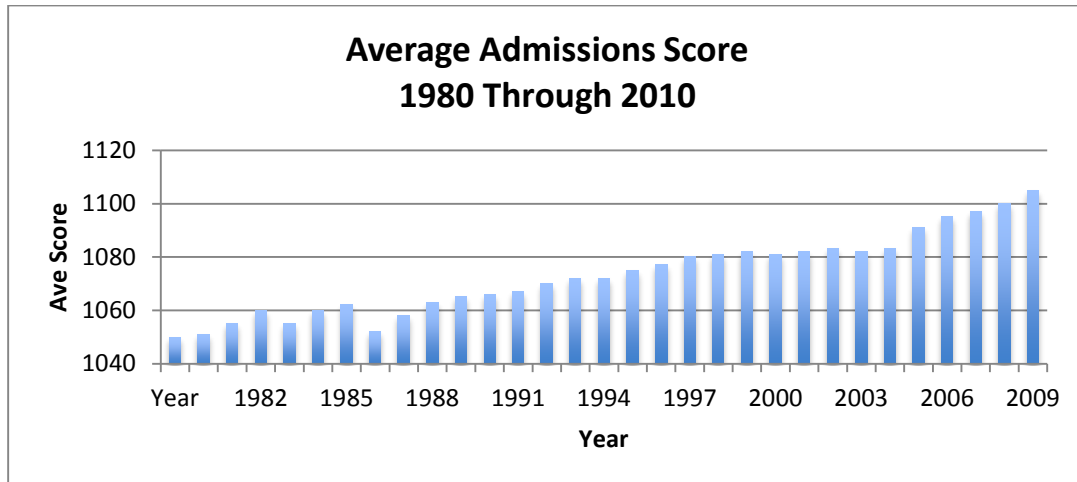
also:

<https://www.theguardian.com/news/datablog/gallery/2013/aug/01/16-useless-infographics#img-3>

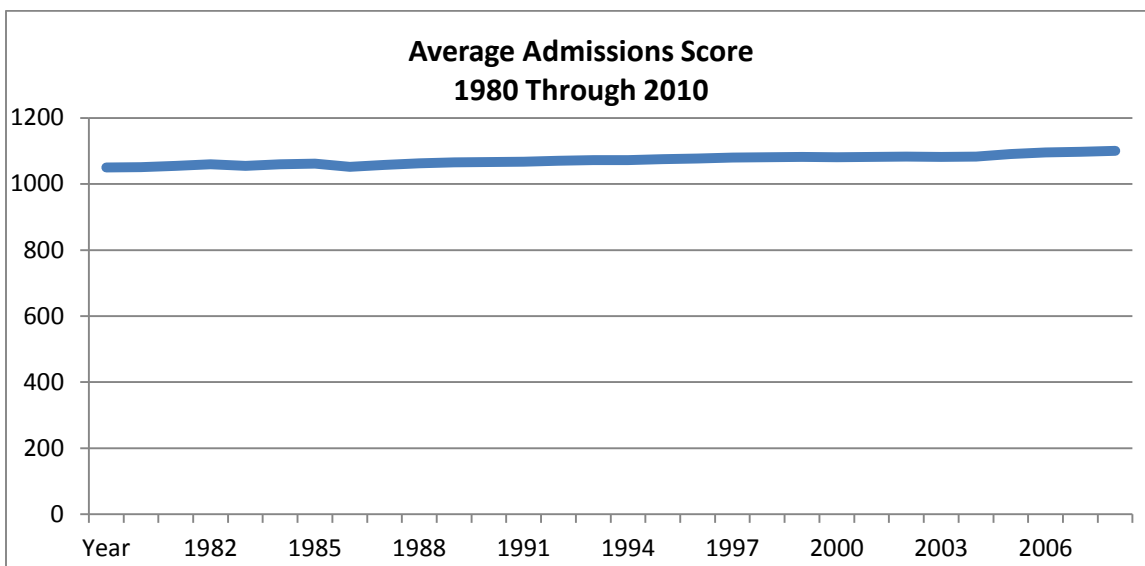
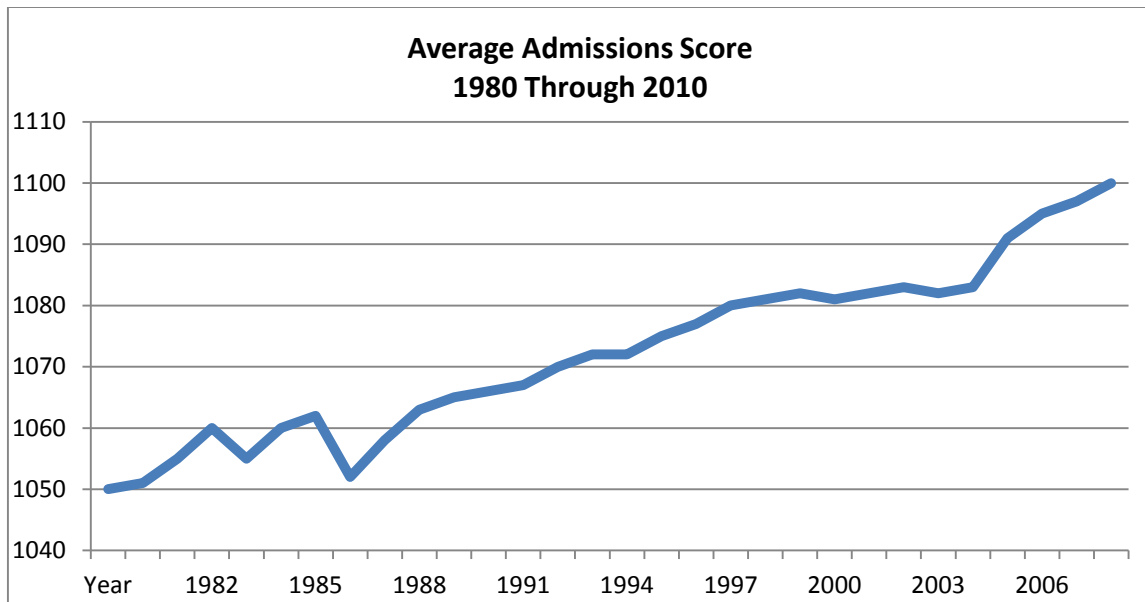
In this graphic above, four stick people stand in for 43,000 nurses in 2008/09 – 2010/11 — but then 28 stick people stand in for an additional 3,000 nurses in 2011/12. This makes a 7 percent increase look like a 7,000 percent increase.

3d. Misinterpretation Based on Y-Axis Scale

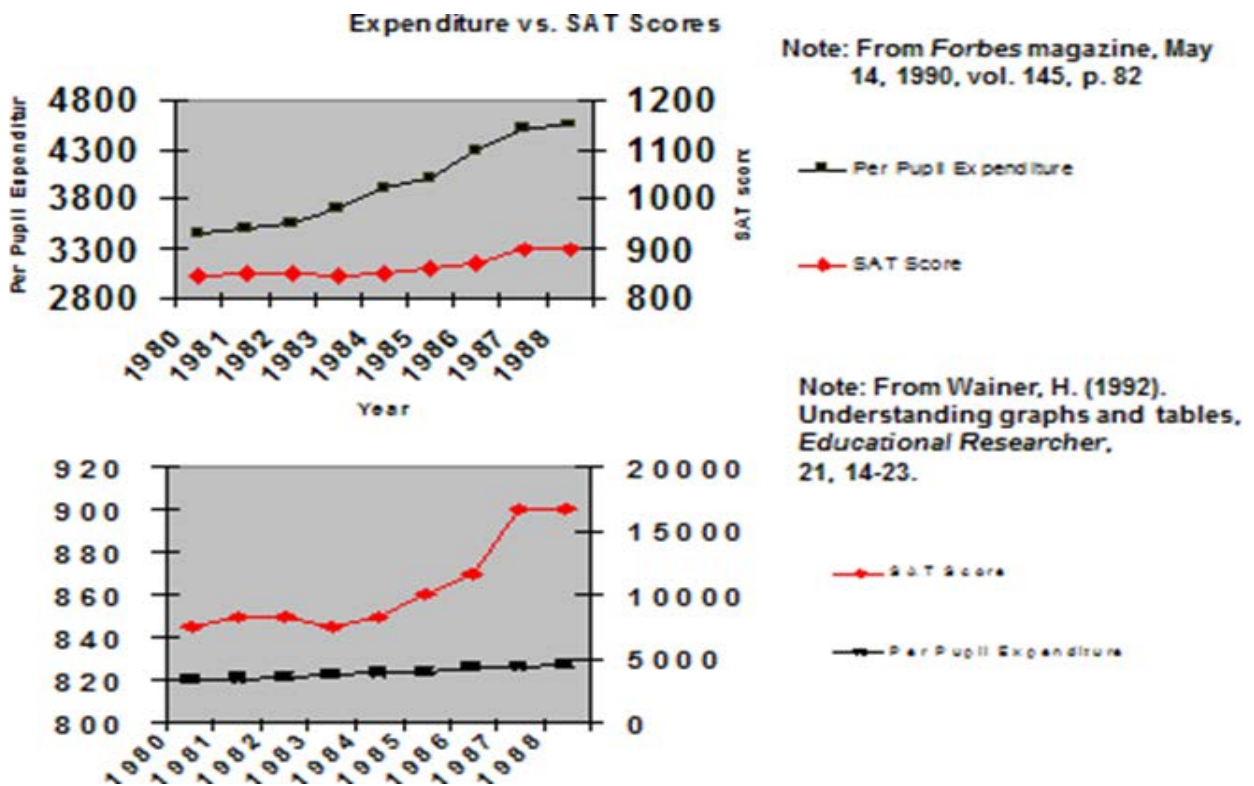
Compare the two charts below. How would you interpret the trend over time based on what is shown? A number of scholars recommend that all or most charts start the scale at zero. In this case, you can see how a change in the minimum Y axis greatly changes the perceived difference over time.



Same (fictitious) data in line chart:



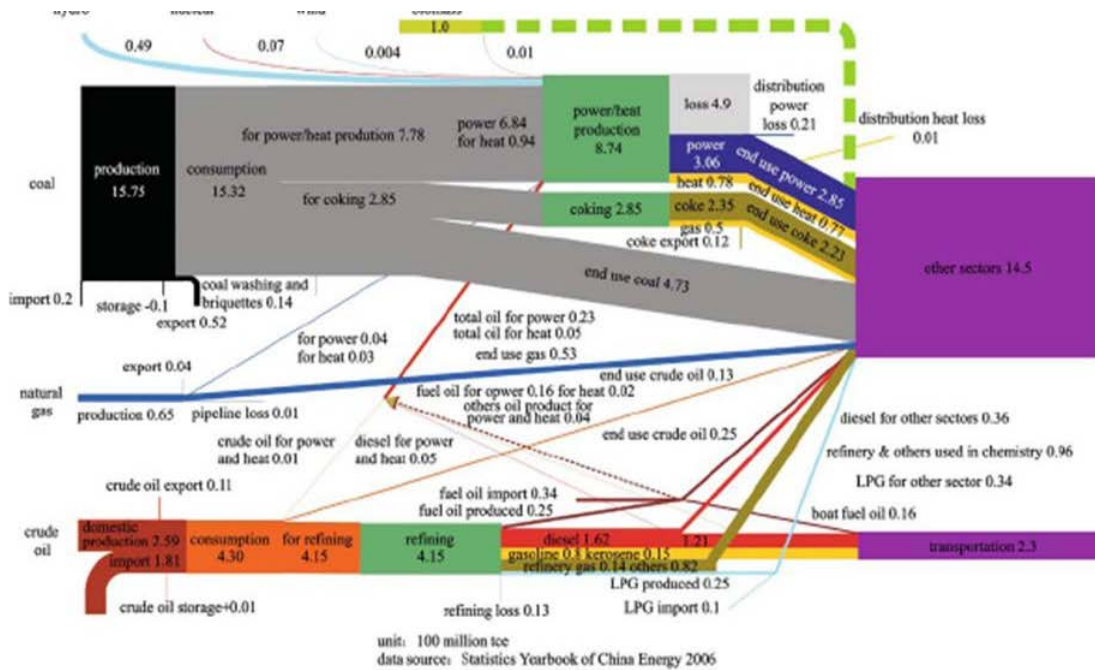
3d. Misinterpretation Due to Different Scaling on Double Y-Axis Chart



Including two Y-axes is challenging to interpret, and then changing the scale for the Y axes clearly shows that the reader can interpret the same data differently. The interpretation has implications for possible policy changes, and you can see that the interpretation could lead to incorrect outcomes.

3e. Multiple Issues, Generally Confusing Visual

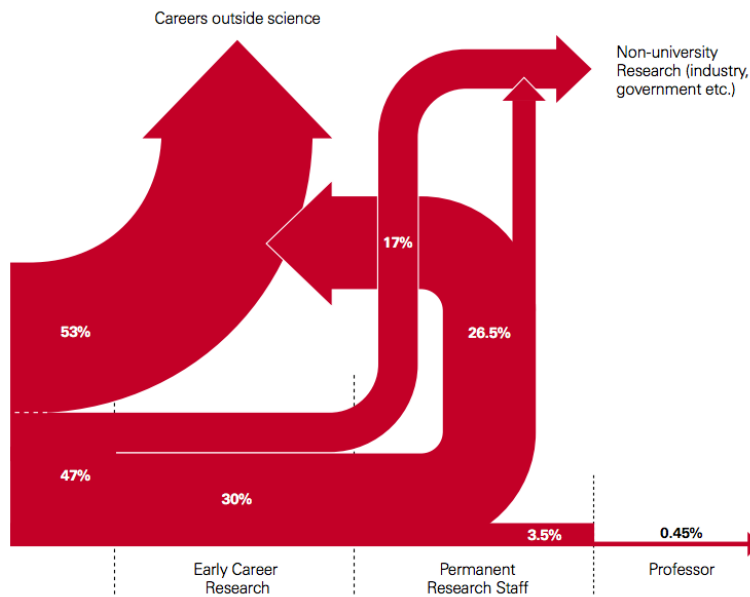
In addition to requiring substantial mental effort to understand, the examples of a Sankey diagrams below ignore several principles of good practice. In the first diagram, there are multiple flow points and many colors. The diagram may require too much mental work to grasp the full information attempting to be conveyed.



Source: http://www.wrsc.org/attach_image/china-energy-flow-chart-2005

3e. Confusing Flow Lines/Patterns

Figure 1.6 Careers in and outside science

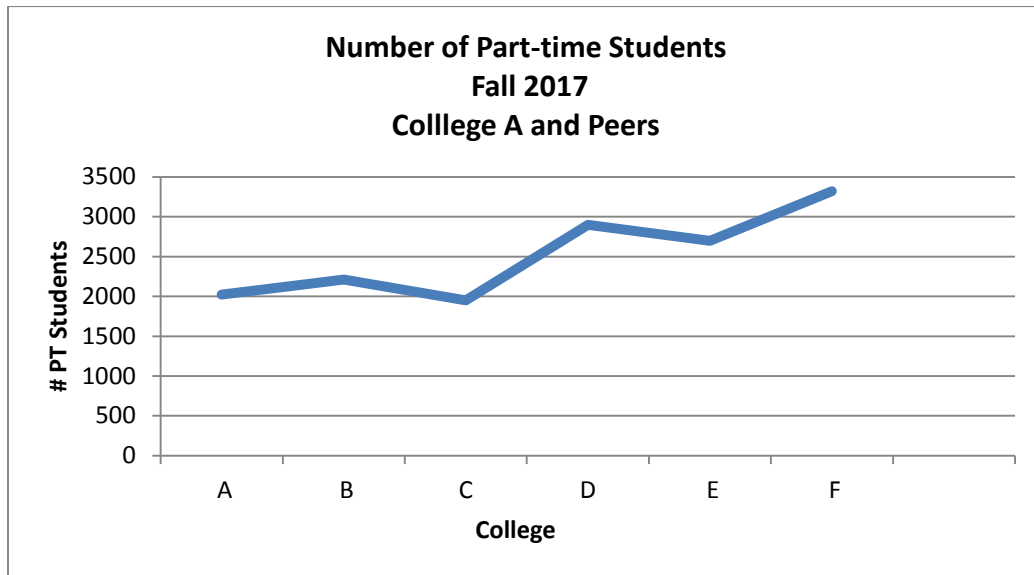


Source: Royal Society (2010)

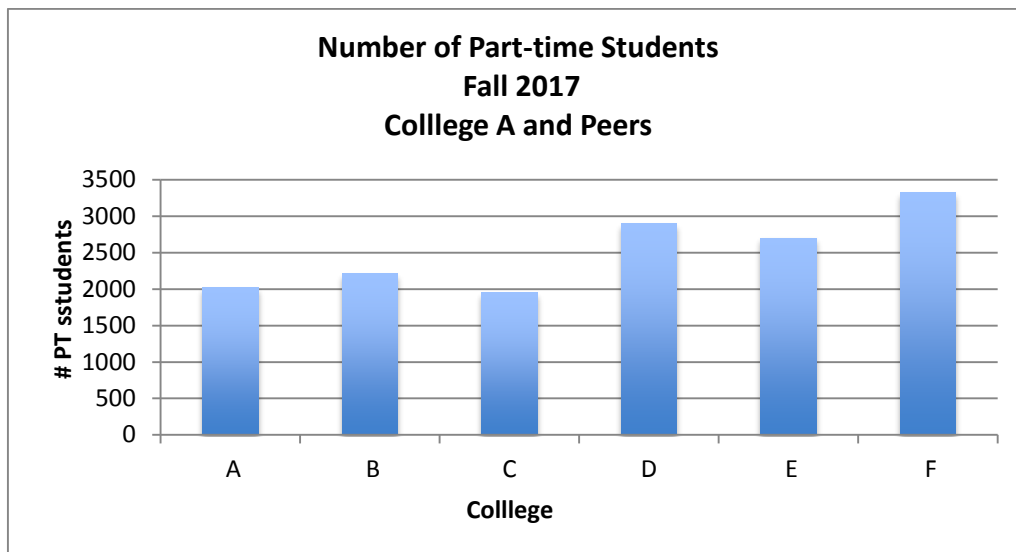
In the diagram above, several items must be considered and require the reader to make assumptions. First, the thickness of the arrows is not proportional to its percentage weight. Second, the endpoints are not clear (there are three endpoints: careers outside science, non-university research, and permanent academic research staff of which professors are a sub-population). Third, some lines overlap, possibly suggesting individuals who fall in these categories (early career researchers who move into careers outside science or non-university research) are taking a step-back or not moving forward in their careers (which could or could not be the case). And fourth, the division of some larger groups into two or more smaller groups is somewhat confusing.

4. Choosing the Wrong Type of Chart

4a. Use of a Line Chart Instead of Bars



Line charts depict a value over time. The line chart above incorrectly attempts to show a comparison of one time point of data for multiple sites. A bar chart (shown below) or table are more correct options.

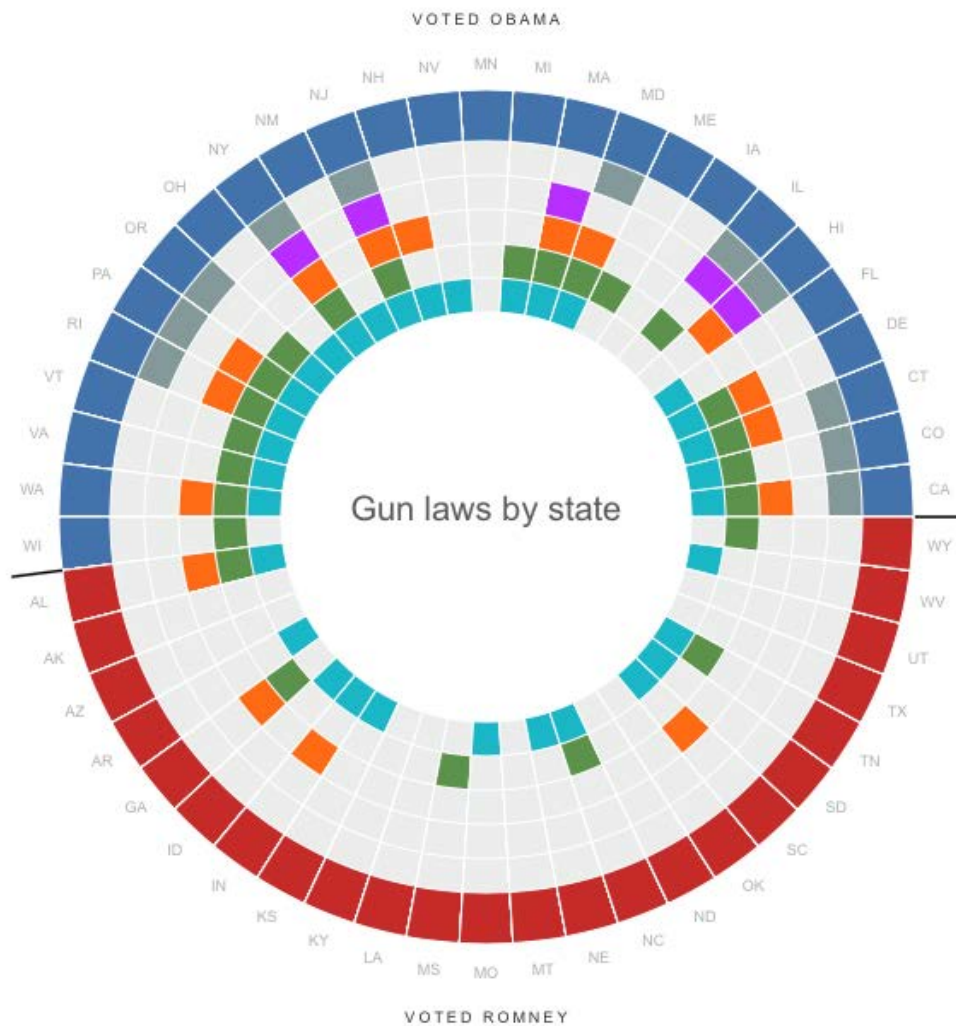


To choose the right type of chart, ask yourself if you want to:

Compare values:	Bar chart
	Line chart
Show the individual parts that make up a whole:	Pie chart
	Stacked bar
	Stacked column
Understand how the data is distributed:	Scatter plot
	Line chart
	Bar chart
Analyze Trends:	Bar chart
	Line chart
Comprehend the relationship between data sets:	Scatter plot
	Line chart
	Bubble chart

Read more at: <http://blog.visme.co/bad-infographics/#Zgib2ZPyLEvq3FAe.99>

4b. Use of Circular Graph Instead of Table



Original graphic from: <https://www.theglobeandmail.com/news/world/gun-control-in-america-a-state-by-state-breakdown/article6465107/>

Source: *Globe and Mail*

This circular graph is complex and very hard to understand. The data presented in a table would be easier to interpret.

LAWS ON FILE

If no colour appears, there is no such law on file

- 2012 election results
- Background check law
- Permit required to purchase
- Licence required to sell
- Records kept on file
- Firearms banned from workplace

5. Correct Labelling and Correct Context

5a. Labels are Switched

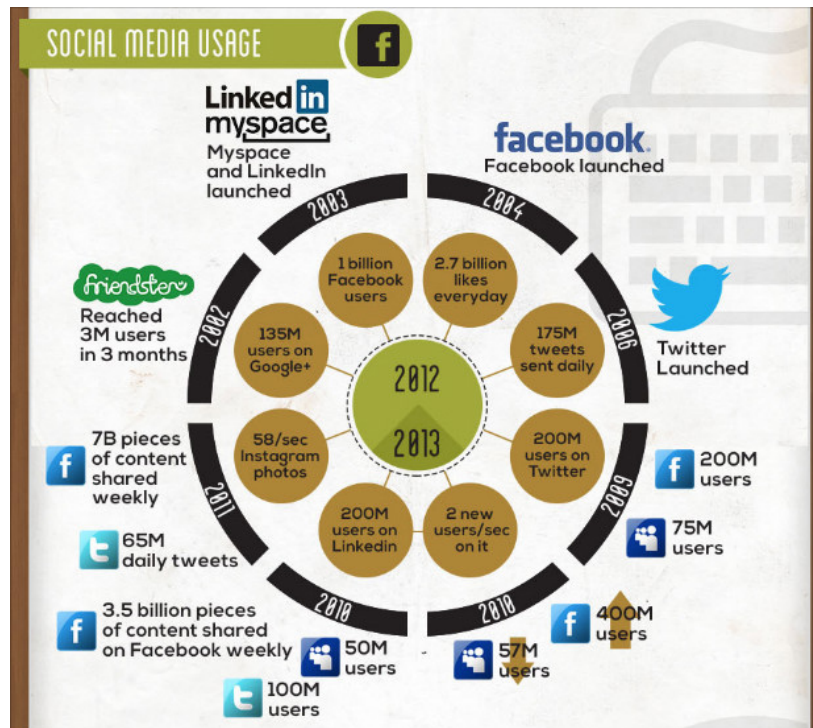
This chart was published in multiple sources a few years ago, including *Upworthy* and *The Atlantic*. But as *The Atlantic* later pointed out, that the content is factually incorrect—the numbers for “prison” and “Princeton” are reversed, and Ivy League tuition hasn't been that low in a long time.



Source: <https://io9.gizmodo.com/11-most-useless-and-misleading-infographics-on-the-inte-1688239674>

6. Multiple Challenges with InfoGraphics

Infographics are challenging to do well, and often cited for their bad design. Here is an example of an infographic that requires the reader to do too much mental work, thus increasing the likelihood for misunderstanding (from Chibani, 2018).



Source: Chibani (2018)

Top Six Reasons Why Infographics are Terrible

http://www.slate.com/blogs/future_tense/2013/11/08/infographic_top_6_reasons_infographics_are_terrible_infographic.html

Important Readings (Not an Exhaustive List)

- Cairo, A. (2013). *The functional art: An introduction to information graphics and visualizations*. Berkeley, CA: New Riders.
- Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. Berkeley, CA: New Riders.
- Calderon, A. (2015). In light of globalization, massification, and marketization: Some considerations on the uses of data in higher education. In K. Webber and A. Calderon (Eds.), *Institutional research and planning in higher education: Global contexts and themes* (pp.288-306). New York: Routledge Press.
- Chibani, N. (2018). 11 Common Infographics Mistakes. Retrieved at: <http://blog.visme.co/bad-infographics/>
- Drake, B., Pytlarz, I., & Patel, M. (2018) Let me paint you a picture: Utilizing visualizations to make data more accessible. In K. Webber (Ed.) *Building capacity in institutional research and decision support in higher education*. Knowledge Studies in Higher Education, 4. Cham: Springer. 81-93.
- Few, S. (2012). *Show me the numbers: Designing tables and graphics to enlighten* (2nd edition). Oakland, CA: Analytics Press.
- French, K. (2018a). How to Fix the 15 Most Common Infographic Design Mistakes. Retrieved from: <https://www.columnfivemedia.com/how-to-fix-the-15-common-infographic-design-mistakes>
- French, K. (2018b). 25 Tips to Instantly Improve Your Data Visualization Design. Retrieved from: <https://www.columnfivemedia.com/25-tips-to-upgrade-your-data-visualization-design>
- Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3, 446-474.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Strauss & Giroux.
- Kosslyn, S. (1994). *Elements of graphic design*. New York: W.H. Freeman & Co.
- Lester, J., Klein, C., Rangwala, H. & Johri, A. (2017). Learning analytics in higher education. *ASHE higher education report*. 43 (5), 9-135.
- Mathies, C. (2018). Uses and misuses of data. In K. Webber (Ed.) *Building capacity in institutional research and decision support in higher education*. Knowledge Studies in Higher Education, 4. Cham: Springer. 95-111.
- Tufte, E. (2001). *The visual display of quantitative data*. Cheshire, CT: Graphics Press.
- Plus E.Tufte other books: *Envisioning Information; Visual Explanations; Beautiful Evidence*. All from Graphics Press. See: <http://www.edwardtufte.com>
- Tversky, A. (2001). Spatial schemas in depictions. In M. Gattis (Ed.) *Spatial schemas and abstract thought*. (pp. 79-112). Cambridge, MA: MIT Press.
- Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.
- Tversky, A. & Kahneman, D. (1986). Rational choice and framing of decisions. *The Journal of Business* 59(4). Part 2: The behavioral foundations of economic theory (Oct. 1986), 5251-5278.
- Webber, K.L. & Burns, R.A. (2018). Graduate student debt 2000 to 2012: Trends and implications. Submitted for consideration to *Peer-Reviewed Journal*, April 2018.