"As the 2016 presidential campaign lurches forward, students need data-comprehension skills more than ever to support democratic decision-making."

Data Literacy Strategies to Bolster Student Election Understanding

KRISTIN FONTICHIARO

Have you ever worried that with the post–Common Core influx of larger quantities of informational text, your students know to extract numbers from reading but cannot explain what those numbers mean or whether their size indicates significance? Have numbers and statistics in a news story ever seemed really big or small, and you aren't sure why?

Many Americans (including educators and their students) struggle with data literacy—the ability to "read" and "write" effectively with data, specifically quantitative data. As the 2016 presidential campaign lurches forward, students need data-comprehension skills more than ever to support democratic decision-making. Here are three high-leverage data literacy strategies that can help.

RECOGNIZE THAT AVERAGE ISN'T ALWAYS AVERAGE

It is common to hear about the "average American." Yet there are three ways to calculate the average, and each can be utilized at different times to be more or less representative of the data: mean, median, and mode. The difference between them is distinct:

- *Mean* is the most common way to calculate the average. Add up all of the numbers, then divide by the number of numbers you had.
- *Median* lines up all the numbers in ascending order. Find the middle number. That is the median. If there are two middle numbers, average them together.
- *Mode* is the most frequent number in the series. If no number repeats, you cannot calculate a mode.

Let's use some presidential data to see how these work in practice. In 2015, *Forbes* released its calculations of the net worth of each presidential candidate. Here is a list of the presidential candidates (as of March 14, 2016) and their *Forbes*-calculated net worth. (Note: Trump has disputed *Forbes*'s algorithms for several years [O'Brien, 2005] and, in his filing with the Federal Election Commission, pegged his net worth at $10 billion [Mullany, 2015]).

- Donald Trump, $4.5 billion
- Hillary Clinton, $45 million
- John Kasich, $10 million
- Ted Cruz, $3.5 million
- Bernie Sanders, $700,000
- Marco Rubio, $100,000

The *mean* rounds up to $760 million. Yet five of the six candidates have a net worth less than that, so this approach to averaging is misleading.

Because we have an even number of data points, we take the average of the two middle numbers, Kasich's $10M and Cruz's $3.5M, to find the *median*: $6.75 million. This is a much more accurate number.

We cannot calculate *mode* because no numbers repeat. However, had we expanded the list to include Rand Paul and Rick Santorum, both of whom have a *Forbes*-estimated net worth of $2 million (Fontevecchia, 2015), then the mode would be $2 million. (Coincidentally, these add-on candidates would shift the median from $6.75 million to $2.75 million and the mean down to $459 million. Adding just a few data points to a set can make a big difference!)

So which is the "best" average to use? It depends on the point to be made: A political commentator might want to use the mean to make the case that all candidates are out-of-touch "fat cats." This is *technically* true even as it is misleading. Rubio (now withdrawn) could have argued that with a net worth $6 million below the median, he is most representative of the middle class—it's true both in terms of averages and in candidate rankings. How might using different versions of averaging impact the data for a variety of objective and nefarious purposes?

Bottom line: When students see "average," ask, "How was it calculated?"

COMPARE NUMBERS WITH STATISTICAL BENCHMARKS

Understanding numbers in context is difficult for novice researchers who do not have prior knowledge against which to weigh information. In *Stat-Spotting: A Field Guide to Identifying Dubious Data* (2013), Joel Best suggests that when you first confront a number, you compare it to what he calls *statistical benchmarks*: baseline data about population, geography, economics, income, or other aspects. Numbers in the millions can seem huge, but they might not be if only we could see those numbers compared to other data. Statistical benchmarks can help us narrow the comprehension gap and serve as "rules of thumb" (Abilock, 2012) to guide our students to higher comprehension (see Figure 1).

Let's see how statistical benchmarks might improve evaluation of some hypothetical campaign trail statements. Compare these claims to statements against the statistical benchmarks in Figure 1. Would you vote for this candidate?

• "When elected, I will add $400 million to support full-day kindergarten." Is this a significant investment? Well, if we know that there are about 4 million children born annually, then a $400 million budget divided by 4 million kids equals about $100 per child. Is that likely to cause a significant impact on kindergarten education? Probably not.

• "Two hundred U.S. representatives stand with me on this issue." Although 200 sounds like a large number, it is less than half of the 435 U.S. representatives.

• "Last year, two million Americans died of cancer." Can this be right, if only 2.5 million people total died in the U.S. last year? It *could* be that 80% of deaths come from cancer, but we don't know for sure. Without knowing for sure, we had better do some more research, because something seems suspicious here. Bottom Line: Posting statistical benchmarks in libraries, classrooms,

Fig. 1. Statistical Benchmarks

Here are some sample benchmarks that can be useful for comparisons:

• U.S. population: Just over 323 million (U.S. Census, 2016)

• World population: Just over 1.6 billion (U.S. Census, 2016)

• Number of births each year: Just under 4 million (NCHS, 2014a). This means, on average, there are about 4 million first graders, 4 million kids who will be eligible for a driver's license, and 4 million young adults who can vote for the first time in any given year.

• Number of deaths each year: Just over 2.5 million (NCHS, 2014b)

• Total acres of land in the United States: 2.3 billion (Lubowski, Vesterby, Bucholtz, Baea, & Roberts, 2006)

• Total number of United States representatives: 435 (U.S. House, n.d.)

• Total number of United States senators today: 100 (U.S. Senate, n.d.)

Need additional benchmark data? Try these sites:

• U.S. Census QuickFacts: http://www.census.gov/quickfacts/table/PST045215/00

• CDC's National Center for Health Statistics: http://www.cdc.gov/nchs/nchs_for_you/general_public.htm

• USA.gov portal for statistics about the United States: https://www.usa.gov/statistics

Samples should be of adequate size. You need *enough* units in the sample that you can't just write off any findings as coincidence.

and computer labs can help students quickly compare data they encounter and evaluate its effectiveness.

STUDY SAMPLING

A *sample* is the pool of people studied, polled, interviewed, or counted to come up with data results. Sampling is important because it is rare that any researcher, news outlet, or political pollster can talk to every single person. But sampling also comes with challenges, and we need our students to be alert to them.

Sample Size. Samples should be of adequate size. You need *enough* units in the sample that you can't just write off any findings as coincidence. There is no hard and fast rule for this, more of a rule of thumb. If you want to be able to say that a certain percentage of Americans trusts the Electoral College (and how many Americans are there? Check your statistical benchmarks!), you can't just ask the people who live on your street. Five or six households are insufficient to represent over 300 million people.

Randomly Selected. Five households are a poor sample for another reason: they are not *randomly* selected. Choosing every fifth house address from a city's phone book will get you a better random sample—*for your city.* This was a problem during presidential election years when pollsters were still relying on home phone numbers for polling at a time when many young voters had transitioned to cell phone–only access (Jackson, 2008). By only contacting those who had landlines, pollsters were deliberately excluding younger voters—not random!

Representative. There is another important reason you cannot merely poll your street: it is not *representative* of America as a whole. Many Americans live on different kinds of streets from yours: more or less rural or urban, more or less balanced between owners and renters. What about the homeless? Those from different cultural backgrounds? What about seniors who have moved out of single-family homes and into senior apartment complexes a few streets away?

Representative sampling isn't just being *inclusive*, however; sometimes, it means knowing who *not* to include. One of the big surprises in Iowa caucus years is that polling predictions can be wildly different from voting patterns. What distinguishes the polls of the *Des Moines Register* and its pollster Ann Selzer? Unlike many 2016 polls that called random citizens (which is faster and easier to do), Selzer sampled only registered voters. By eliminating those who would not vote, she could better predict the behaviors of who *would* vote (Cohn, 2016; Malone, 2016). Sometimes, less is more.

Defining the Sample. One of the challenges of running any kind of poll, survey, or data-collection project over time is definitions. Census data, for example, can be difficult to compare over time because the definitions of different racial and ethnic groups have greatly changed over the years. Consider the February 2016 Republican debate. With five candidates onstage (Marco Rubio, Donald J. Trump, John Kasich, Ted Cruz, and Ben Carson), Rubio said, "On this stage tonight there are two descendants of Cuban origin, and an African American. We are the party of diversity."

Rubio was defining diversity along social/racial lines, and 60% of candidates being non-Caucasian was a significant accomplishment for the Republicans. Others might look at those five candidates and say, "No women? Not diverse!" Different definitions of diversity make it hard to compare information. Which would be more diverse, an all-male Republican slate with mixed socioethnic backgrounds or an all-white Democratic slate that was half female?

Rubio's savvy debate jab is also, sadly, an example of poor sample size. By using only five people to represent a single party's racial composition, he overlooked a 2012 Gallup poll showing that 89% of those who identified as Republicans were also non-Hispanic white (Newport, 2012).

Bottom line: Look at who is being studied and ask, "Is this representative?"

CONCLUSION

Having a few data strategies in your toolkit will help you better support students as they move through informational text, close reading strategies, and research materials. To learn more, download Compound Interest's "Rough Guide to Spotting Bad Science" at http://www.compoundchem.com/wp-content/uploads/2014/04/Spotting-Bad-Science.pdf. This one-

page poster covers twelve common data red flags you can use for mini-lessons, post in libraries and labs, and share with colleagues. Or choose one concept as a building focus each month, and you'll go a long way toward developing your school's data literacy skills.

Additionally, enroll in the free online 4T Data Literacy conference July 14–15, 2016. Learn more and subscribe to our data literacy blog at http://dataliteracy.si.umich.edu.

REFERENCES

Abilock, D. (2012) True or not? *Educational Leadership, 69*(6), 70–74.

Best, J. (2013). *Stat-spotting: A field guide to identifying dubious data*. Berkeley, California: University of California Press.

Cohn, N. (2016, January 31). Why this was the Iowa poll everyone was waiting for. *New York Times*. Retrieved from http://www.nytimes.com/2016/01/31/upshot/why-this-is-the-iowa-poll-that-everyones-waiting-for.html

Fontevecchia, A. (2015, September 25). *Forbes*' 2016 presidential candidate wealth list. *Forbes*. Retrieved from http://www.forbes.com/sites/afontevecchia/2015/09/29/forbes-2016-presidential-candidate-net-worth-list/print/

Jackson, B. (2008, February 21). Cell phones and political polls. *FactCheck*. Retrieved from http://www.factcheck.org/2008/02/cell-phones-and-political-polls/

Lubowski, R. N., Vesterby, M., Bucholtz, S., Baea, A., & Roberts, M. (2006). Major uses of land in the United States, 2002. Economic Research Service of the United States Department of Agriculture. Retrieved from http://www.ers.usda.gov/publications/eib-economic-information-bulletin/eib14.aspx

Malone, C. (2016, January 27). Ann Selzer is the best pollster in politics. *Five Thirty Eight*. Retrieved from http://fivethirtyeight.com/features/selzer/

Mullany, G. (2015, July 15). Donald Trump claims his wealth exceeds "TEN BILLION DOLLARS." Retrieved from http://www.nytimes.com/politics/first-draft/2015/07/15/donald-trump-says-hes-worth-more-than-10-billion/

National Center for Health Statistics, Center for Disease Control (NCHS). (2014a). Births and natality. Retrieved from http://www.cdc.gov/nchs/fastats/births.htm

National Center for Health Statistics, Center for Disease Control (NCHS). (2014b). Deaths and mortality. Retrieved from http://www.cdc.gov/nchs/fastats/deaths.htm

Newport, F. (2012). Democrats racially diverse; Republicans mostly white. *Gallup*. Retrieved from http://www.gallup.com/poll/160373/democrats-racially-diverse-republicans-mostly-white.aspx

O'Brien, T. L. (2005, October 23). What's he really worth? *New York Times*. Retrieved from http://www.nytimes.com/2005/10/23/business/yourmoney/whats-he-really-worth.html

United States Census. (2016). U.S. and world population clock. http://www.census.gov/popclock/

United States House of Representatives. (n.d.) The House explained. Retrieved from http://www.house.gov/content/learn/

United States Senate (n.d.) Home page. Retrieved from http://www.house.gov/content/learn/

Kristin Fontichiaro, clinical assistant professor at the University of Michigan School of Information, runs the Michigan Makers (http://michiganmakers.si.umich.edu) and Making in Michigan Libraries project (http://makinglibraries.si.umich.edu), made possible in part by the Institute of Museum and Library Services RE-05-15-0021-15. Additionally, she is series editor for the acclaimed Makers as Innovators series for Cherry Lake Publishing. Email: font@umich.edu