



KNOW YOUR EVERYDATA™

How To Avoid Being Misled
By the “Little Data”
You Consume Every Day

John H. Johnson, PhD & Mike Gluck

Status reports. Emails. Weather reports. From the moment your alarm jolts you awake, you're bombarded with data.

Here's an eye-opening fact for you: the amount of data you likely consume in a day—[34 gigabytes](#)—would fill dozens of pickup trucks if you printed it all out.

So what's the problem with all this data? This: the majority of numbers you read in newspapers, hear on TV, and see at work are either wrong or misleading—or both.

How do we know? Because we've spent the past three years researching hundreds of ways in which data is misreported and misinterpreted. From [headlines that rarely tell the full story](#) to [less-than-forthcoming graphs](#) proffered by Fortune 100 companies to [a false Twitter report](#) that wiped out “\$136 billion in value” from the stock market, it's almost impossible to avoid an ever-growing flood of misleading information in today's data-driven, media-saturated world.

The Little Data Dilemma

So what do you do? You probably can't change the amount of misinformation that fills your inbox and clogs your smartphone. But you can change the way you interpret data—and learn how to use data to make better decisions for your business. The good news is you don't have to be a trained statistician. Just knowing a few key concepts—and, perhaps more importantly, being willing to dig a bit deeper—will put you far ahead of your peers.

Following are two of our favorite case studies. Each real-world example illustrates how a little insight can go a long way in changing how you understand data.

“ You probably can't change the amount of misinformation that fills your inbox and clogs your smartphone. But you can change the way you interpret data—and learn how to use data to make better decisions for your business.

Show Me The Money

Let's start with a research study focusing on how much people earn in the workplace, and how this data may—or may not—be related to their athletic prowess in high school.

An article in *The Atlantic* asked the question, "[Why Do Former High School Athletes Make More Money?](#)" According to the piece, people who were athletes in high school end up with "higher status careers" compared to those who weren't, with wages "between 5 and 15 percent higher than those of the poor trombonists and Yearbook Club presidents." Nice. As if the kids who played trombone in high school didn't have it bad enough already.

This article struck John's wife as exactly the type of question that raised a flag for sound consumers of data, and at her urging we investigated the study and the underlying numbers. The basic premise of the article is based on [studies](#) by researchers at Cornell and Southern Illinois University who looked at two unique data sets on biodata. In psychology, biodata is self-reported data about one's self (i.e., "biographical data"). There are two studies—in the first one, a sample of 66 adults participated in a survey about leadership traits and past experience in extracurriculars. This part of the survey was to capture subjective views about participation in athletics, and asked about how a person's extra-curricular activities related to qualities such as self-confidence, leadership, and self-respect.

From this small sample of 66 people, the authors conclude that “people tend to expect former student-athletes to demonstrate greater leadership ability as well as organizationally beneficial personality traits; however, former student athletes are not expected to be altruistic with respect to others.” We won’t spend a lot of time on the first study, other than to say there is always controversy about these types of attitudinal surveys across fields (economists tend to be very skeptical, but those in industrial organizational psychology utilize them quite often). From a statistical standpoint, however, it does appear that the key result may be affected by combining all sports activities into one bucket, and all non-sports activities into a second bucket.

That said, let’s focus on the second study, which conducted a statistical analysis of the 2000 University of Illinois Veteran’s Survey. According to the study, this sample contained information on 931 World War II veterans who were ages 71 to 93 at the time of completion of the survey in 2000. The key conclusion from the authors here is that participation in sports as a kid has a positive effect that “persist[s] for more than 55 years.” The authors describe “a positive relationship between participation in competitive youth sports and several measures of long-term personal success and prosociality.”

Here are a few observations:

First, there is no part of either study one or study two that looks at or measures the actual effect of high school sports on wages. That data does not exist. Rather, the study looks at how self-reported data of participating in a college sport 55 years ago correlates with leadership metrics, and with trade jobs versus upper-management jobs. This is a fairly inexact methodology by which to measure whether one gets a better job.

Second, the premise of the study is that sports participation more than 55 years ago correlates with future job outcomes. But there is only a somewhat limited set of explanatory variables; in fact, the only other explanations that the study can control for are age and size of hometown. This immediately raises the question of omitted variables—what if high school sports participation by a subset of male veterans ages 71 to 93 was correlated with any number of other factors, such as their benefits under the GI Bill, their education level, their participation in World War II, their ability, or the income level of their parents?

Finally, the authors of the veterans study also note their results are potentially picking up correlations but no causation. There are a lot of two-way analyses in this study—leadership correlated with athletics, self-confidence correlated with self-respect, etc.

Here is one possible explanation: What if kids who have time for sports are better off financially (i.e., they don't have to work after school)? Being better off financially when you are young could explain success later in life. The point is, this is a complicated question, and it is highly unlikely that these particular metrics are capturing a true causal relationship.

We don't mind biometric data—and the study is certainly very interesting. But this is yet another example where the data captured in headlines can't quite convey the nuanced interpretation of the underlying study.

What's the takeaway here? First of all, it's important to read past the headline. (Most Americans don't.) But even once you get into the details of the article—and the underlying study—two key issues should be readily apparent: self-reported data is subject to manipulation, and correlation is not proof of causation. If you're relying on misleading data as the foundation for making decisions about HR (or any other aspect of your business, for that matter), you may find yourself hiring the wrong people, focusing on the wrong metrics, and limiting your ability to achieve your strategic goals.

“The data captured in headlines can't quite convey the nuanced interpretation of the underlying study ... it's important to read past the headline.”

Sold on Data

Next, let's turn our attention to an industry-leading company whose data is consistently misinterpreted by the general public, despite the fact that the company clearly explains how to understand the numbers.

The company? Zillow.

If you're trying to buy or sell a home, you've probably heard of Zillow. With a database of more than 110 million homes in the U.S., Zillow lets you see how much nearly any home in the country is worth.¹

Well, sort of.

Here's how it works: when you look up a home, you'll see a "Zestimate," which is Zillow's beautifully branded name for a home's estimated market value.

1. We say "nearly" any home because Zillow doesn't appear to track every home in the U.S., based on data we compared from the American Housing Survey from the U.S. Census Bureau. (The United States Census Bureau, American Housing Survey (AHS), last revised May 14, 2015.) That said, the differences we found could be due to time (the data was collected approximately two years apart), or due to definitions in terms of what constitutes a "housing unit" (for the bureau) versus a "home" (for Zillow). The data we looked at includes: [Table C-01-AH](#), American Housing Survey, accessed August 6, 2015. This number is itself based on a sample of 70 million housing units. We also looked at "[Zestimate](#)," accessed August 6, 2015, Zestimate_Accuracy_2015_03_31.

To come up with the Zestimate, Zillow uses a proprietary formula that relies on both public and user-reported data. According to its website, the company looks at all sorts of data, including:

- Square footage
- Location
- Number of bedrooms and bathrooms
- Property tax data
- Past sale prices
- Comparable sales of other homes in the same area

In fact, you can probably look up your home right now on Zillow and see how much the company thinks it's worth—even if you bought it years ago.

If there's not enough data for a Zestimate, sometimes Zillow simply doesn't give one. For example, when we looked at Zillow's data, none of the homes in Vermont had a Zestimate.² Zero. Zip. Zilch. If you live in New Jersey, however, you're in luck—Zillow has a Zestimate for 99.4 percent of the homes in the Garden State.

2. Zillow offers quite a bit of information about its Zestimates—including a downloadable Microsoft Excel spreadsheet that shows their accuracy (we found the link [here](#)) and [a website dedicated to data](#).

Even though Zillow points out that a Zestimate is not an appraisal, many people still rely on it when buying or selling a house. “If a house for sale has a Zestimate of \$350,000, a buyer might challenge the sellers’ list price of \$425,000,” noted [a Los Angeles Times article](#). Similarly, sellers might ask an agent why they should list their house for far below the Zestimate.³

“Not a week goes by that we don’t encounter a consumer who is fixated on a particular value for a home because that’s what Zillow says it is,” said an officer of a real estate agency in [a piece for the Washington Post](#), where he also called Zillow’s predicted values “wildly inaccurate and inconsistent.” (In [a rebuttal piece in the Post](#), Zillow’s chief economist notes that it’s nearly impossible for anyone to predict a home’s sale price with 100 percent certainty, while conceding that real estate agents outperformed Zillow in a D.C.-area study. “Is a well-informed human better at pricing an individual home than a computer?” he asked. “The answer is yes, of course. But it’s closer than you might think.”⁴)

3. Keep in mind that real estate agents may have less to lose than you do when it comes to reducing the price of your house. For example, if your house sells for \$300,000, your agent might get \$5,250 (a 7 percent listing fee = 3.5 percent for the buyer’s agent, and 3.5 percent for the seller’s agent; of that 3.5 percent, half may go to the agent, and half to the agency, so each agent ends up with 1.75 percent). If you lower the price of your house and it sells for \$250,000, your agent only loses \$875 (1.75 percent of \$50,000) but you’ve lost a lot more.

4. The study found that initial Zestimates “were within 5 percent of the ultimate sale price 46 percent of the time,” while real estate agents’ initial list prices were within 5 percent “76 percent of the time.”

So how should you interpret all of this information?

As you start to dig into the data, you'll see that Zillow provides something it calls a median error rate for each geographic area. The median error rate is a percentage, and it tells you that half of the Zestimates in that area will be closer to the final sale price than the error rate, and half will be further.

For example, when we looked at the data for Los Angeles, the median error rate was 7 percent. This means half of the homes sold for a price within 7 percent of their Zestimate (keep in mind it could be up to 7 percent lower or higher—so on a \$800,000 house that's \$744,000 to \$856,000). Of course, the median error rate also means that half the homes sold for an amount outside this range—either less than 7 percent of the Zestimate, or more than 7 percent. Note that we looked at data from a specific day, and that the values may have changed between then and whenever you're reading this.

In some cases Zillow also gives what it calls a “Value Range,” which shows the estimated high and low values for a home. This range, as it explains on its website, is in fact a 70 percent confidence interval, meaning that the home value should be within that range in 70 percent of an infinite number of samples.

In other words, a Zestimate should be considered in the context of its median error rate. To its credit, Zillow displays the median error rate and explains what it is. It even simplifies everything with a four-star rating system that tells you how accurate the Zestimate is (as you might expect, the stars seem to correspond to the median error rate).⁵

The issue is that many people simply focus on the Zestimate without putting it in context, just as many people looking at a poll don't consider the margin of error—they just look at who is “winning.” But by ignoring the additional data—whether it's a median error rate, margin of error, or confidence interval—you're missing the big picture. Let's say you live in Washington, D.C., which had the highest Zestimate accuracy when we looked at Zillow's top metro areas. (Including a four-star rating and the lowest median error rate for top metro areas, which was 5.0 percent when we reviewed it on September 1, 2015.) If you're trying to sell a home with a Zestimate of \$500,000, even if your house is within the median error rate of 5.0 percent (and remember, only half of the homes are), that means it could be \$25,000 (5.0 percent of \$500,000) more or less than the estimate. That's a \$50,000 range.

“By ignoring the additional data ... you're missing the big picture.”

5. Zillow says stars are “tied to” the median error rate. In the data we reviewed, a median error rate of 5.3 percent to 7.4 percent got four stars, 7.6 percent to 8.9 percent got three stars, 9.1 percent to 11.8 percent got two stars; and areas with one star didn't have a median error rate given.

Zestimates are also interesting because some people use them as a forecast, trying to determine what a house will sell for in the future. But Zillow [clearly notes on its website](#) that the Zestimate is based on what the home is worth “today.” (Although we should note that Zillow does offer a [Zestimate forecast](#) for some properties, and it has a [Zillow Home Value Index](#) that offers predictions about the housing market.) Given that Zestimates are typically updated three times a week, it’s quite possible that home sales in your area (or other factors) could change a home’s Zestimate between the time you review it and the time your home sells. There are two lessons here—first of all, figure out if the data you’re seeing is truly a prediction or not before you use it as such. And secondly, make sure that prediction error is accounted for, whether it’s already added in, or you’re adjusting for it.

Finally, consider what the Zestimate is based on. It’s a proprietary formula, which means you don’t know all the factors that go into it. It relies at least in part on self-reported data, which we know (in some cases) may not be 100 percent reliable. And the amount of data available may change from region to region, or even from week to week within the same region. When you don’t have data, you need to work with the sample that you have. If there are more transactions in a certain area, that gives Zillow more data, and the Zestimates presumably become more accurate. (Or, as we think they should say—Zaccurate.)

What's interesting about Zillow from a business perspective is that the company is—by many measures—exceedingly transparent in how it shares and explains its data. Information about median error rates and other statistical matters is fairly easy to uncover on their website. Yet, home buyers and sellers alike routinely latch on to a Zestimate as gospel, despite clear disclaimers that it is anything but.

When you share data—whether it's with your customers, your employees, or your leadership team—it's important to recognize not only your audience's ability to understand the data, but also their appetite for doing so. In this case, surely some Zillow users agree with the Zestimate they find, and refuse to look for any signs that it could be less than 100% accurate. Other users may see Zillow's explanations regarding self-reported data and median error rates, but may not fully realize how these factors may impact the results. Perhaps some users are intimidated by the mere thought of statistics. Regardless of the reason, each of these scenarios ends with the same net result: despite Zillow's efforts to educate users about what a Zestimate is (and isn't), people may still reach inaccurate conclusions regarding what the data truly means.

“It's important to recognize not only your audience's ability to understand the data, but also their appetite for doing so.”

Be a Sound Consumer—and Creator—of Data

Now that you've seen just a few of the ways in which data can be misunderstood, we leave you with two recommendations:

First of all, challenge yourself to go deeper when you're consuming data. Read the footnotes. Ask more questions. Understand where the data is coming from, and who is presenting it. Make sure you're asking the right questions to get the answers you need.

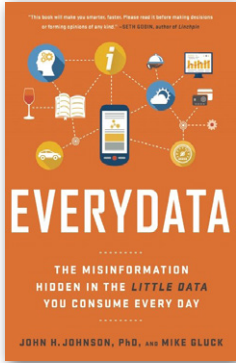
Secondly, recognize that different people have wildly different abilities to create, consume, and interpret data—and understand how this impacts your business. Many people simply will not take the time, possess the skill, or have the skepticism and curiosity needed to fully comprehend anything beyond the most basic statistics. For better or worse, this means that the power to change the way your audience consumes information is in your hands. Whether you're creating a quarterly report for your supervisor, writing an article for your company newsletter, or developing spec sheets for your clients, you can often change the story dramatically simply by paying close attention to the ways in which you select and present the data.

Start Now

The amount of information in our lives (and the lives of your customers) will only continue to increase for the foreseeable future. With 24/7 news cycles and the world at our fingertips, there has never been a more critical time to understand how to interpret and distribute this data.

By becoming a more discerning consumer of data, and recognizing the impact that data has on our everyday lives, you can help ensure continued success for your company—and your career. 📊

Info



BUY THE BOOK | Get more details or buy a copy of [Everydata](#).

ABOUT THE AUTHORS | **John H. Johnson, PhD** is President and CEO of Edgeworth Economics, and a professional economist, expert witness, author, and speaker known internationally for his ability to explain highly sophisticated concepts in a simple, straightforward manner. Through his leadership, Edgeworth Economics has become one of the world's premier economic consulting firms.

Mike Gluck is an award-winning writer and marketer who has led the creation of multi-million dollar advertising campaigns, and worked on behalf of Time Warner Cable, Fisher-Price, the Pittsburgh Penguins, and Riverside Health System, among others. As President of Gluckworks—a copywriting and marketing firm in Buffalo, NY—Gluck's expertise is making complex topics easy to understand.

→ **SEND THIS** | [Pass along a copy](#) of this manifesto to others.

→ **SUBSCRIBE** | [Sign up for e-news](#) to learn when our latest manifestos are available.

This document was created on June 8, 2016 and is based on excerpts from Everydata, published April 12, 2016. The copyright of this work belongs to the author, who is solely responsible for the content. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License. To view a copy of this license, visit [Creative Commons](#) or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA. Cover image from [Adobe Stock](#). You are given the unlimited right to print this manifesto and to distribute it electronically (via email, your website, or any other means). You can print out pages and put them in your favorite coffee shop's windows or your doctor's waiting room. You can transcribe the author's words onto the sidewalk, or you can hand out copies to everyone you meet. You may not alter this manifesto in any way, though, and you may not charge for it.

About ChangeThis

ChangeThis is a vehicle, not a publisher. We make it easy for big ideas to spread. While the authors we work with are responsible for their own work, they don't necessarily agree with everything available in ChangeThis format. But you knew that already.

800ceoread

ChangeThis is powered by the love and tender care of 800-CEO-READ. Visit us at 800ceoread.com, and keep up with the latest developments in business books on our review site, [In the Books](#).