

I'm John Johnson, CEO of Edgeworth Economics, and co-author of “Everydata: The Misinformation Hidden in the Little Data you Consume Every Day.” Let's talk data (and how it's misrepresented and misinterpreted)! AMA!

johnjohnsoneverydata ¹ and r/Science AMAs¹

¹Affiliation not available

April 17, 2023

Abstract

Hey Reddit! I am John Johnson, founder and CEO of the economic consulting firm Edgeworth Economics, which is known for its work in antitrust, labor, and intellectual property consulting. Edgeworth models all kinds of big data, from football player injuries to chocolate prices. With Edgeworth, I work as an expert witness, requiring that I explain both simple and complex data concepts to lawyers and juries that knew little about how data could be used to misrepresent a subject. My work explaining data inspired me to work with Mike Gluck to co-write a book: “Everydata: The Misinformation Hidden in the Little Data you Consume Every Day.” Everydata is about how all kinds of data is misrepresented and misinterpreted. Recently I wrote an op-ed for The Hill about the flaws in a particular political poll. In my “spare time,” I am chairman of the board at Appleseed, a nonprofit dedicated to social justice. In my ACTUALLY spare time, I follow professional wrestling and baseball. PROOF I'll be back around 2:30 PM ET to answer all your questions about data (visualizations), how data is misused, econometrics, Everydata, Rampart, or whatever else your heart desires! Edit: I have a meeting to get to, but I'll stop by tomorrow to answer any more questions that I get, or have missed so far Edit 2: I think I officially have to call it at this point. If you have any more questions, you can still post them here or PM me and I'll try to get around to them at some point. Thanks so much to everybody who participated! Also thanks to u/rhiever who set this whole thing up. Appreciate your mods, they're really great!

[REDDIT](#)

I'm John Johnson, CEO of Edgeworth Economics, and co-author of "Everydata: The Misinformation Hidden in the Little Data you Consume Every Day." Let's talk data (and how it's misrepresented and misinterpreted)! AMA!

JOHNJOHNSONEVERYDATA [R/SCIENCE](#)

Hey Reddit! I am [John Johnson](#), founder and CEO of the economic consulting firm [Edgeworth Economics](#), which is known for its work in antitrust, labor, and intellectual property consulting. Edgeworth models all kinds of big data, from football player injuries to chocolate prices. With Edgeworth, I work as an expert witness, requiring that I explain both simple and complex data concepts to lawyers and juries that knew little about how data could be used to misrepresent a subject.

My work explaining data inspired me to work with Mike Gluck to co-write a book: "[Everydata: The Misinformation Hidden in the Little Data you Consume Every Day](#)." Everydata is about how all kinds of data is misrepresented and misinterpreted.

Recently I wrote an [op-ed for The Hill](#) about the flaws in a particular political poll.

In my "spare time," I am chairman of the board at [Appleseed](#), a nonprofit dedicated to social justice. In my ACTUALLY spare time, I follow professional wrestling and baseball.

[PROOF](#)

I'll be back around 2:30 PM ET to answer all your questions about data (visualizations), how data is misused, econometrics, Everydata, [Rampart](#), or whatever else your heart desires!

Edit: I have a meeting to get to, but I'll stop by tomorrow to answer any more questions that I get, or have missed so far

Edit 2: I think I officially have to call it at this point. If you have any more questions, you can still post them here or PM me and I'll try to get around to them at some point. Thanks so much to everybody who participated! Also thanks to [u/rhiever](#) who set this whole thing up. Appreciate your mods, they're really great!

[READ REVIEWS](#)

[WRITE A REVIEW](#)

CORRESPONDENCE:

DATE RECEIVED:

June 23, 2016

DOI:

10.15200/winn.146661.14278

ARCHIVED:

June 22, 2016

CITATION:

johnjohnsoneverydata ,
r/Science , I'm John Johnson,
CEO of Edgeworth Economics,
and co-author of "Everydata:
The Misinformation Hidden in
the Little Data you Consume
Every Day." Let's talk data (and

Hi John,

I read the op-ed on the political poll with interest. When I came across your layman's description of the confidence interval, it didn't seem quite right to me

Margin of error is one common way to measure statistical uncertainty from polling. It's a way of answering the question, "How sure are you?" In this case, the pollsters were only "sure" within 4 points - anywhere from 47 percent to 55 percent of people might think the comments are racist. In other words, it's quite possible a minority - not a majority - held this view.

Moreover, it typically means that if the study were repeated 100 times, 95 times you would get findings within the margin of error

I read this to mean: given a margin of error from a single survey, we could expect to be able to repeat the survey 100 times and find the repeated surveys' estimates would fall within the original survey's margin of error 95% of the time. That's not how I understand confidence intervals, so I decided to do a simulation experiment in R to be sure.

how it's misrepresented and misinterpreted)! AMA!, *The Winnower* 3:e146661.14278, 2016, DOI: [10.15200/winn.146661.14278](https://doi.org/10.15200/winn.146661.14278)

© et al. This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



First, we set up a "population," in this case 1000 yes/no answers with something near 50% of each.

```
set.seed(45)
pop <- rbinom(1000, 1, .5)
popMean <- mean(pop)
popMean
```

```
## [1] 0.487
```

Now we'll perform 100 surveys, each randomly sampling 100 subjects from the population and determining the mean estimate and it's 95% binomial confidence interval.

```
# do an experiment of calculating mean and 95% confidence interval from a random
# sample of size 100 and repeat that experiment 100 times
cis <- t(sapply(rep(100, 100), function(n) {
  s <- sample(pop, n)
  ci <- binom.test(sum(s), n, conf.level = 0.95)
  r <- c(ci$estimate, ci$conf.int)
  names(r) <- c("estimate", "lower", "upper")
  r
}))
```

Here we test what I read of your definition: how often a given confidence interval will contain the mean estimate of the other 99 surveys.

```
mean(sapply(1:nrow(cis), function(x) {
  cis[-x,"estimate"] >= cis[x, "lower"] &
  cis[-x, "estimate"] <= cis[x, "upper"] })))
```

```
## [1] 0.8408081
```

That's not very close to the 95% we should get. Next, we'll try my understanding of the confidence interval: how often do the confidence intervals contain the actual population mean?

```
mean(cis[, "lower"] <= popMean & cis[, "upper"] >= popMean)
```

```
## [1] 0.95
```

That's more like it, but my definition above isn't exactly in layman's terms. I think I would describe it like this:

The survey's margin of error is an expression of uncertainty. The 4 point margin on the given survey means that it's likely that we'd get an answer within 4 points (between 47% and 55%) if we were able to get an answer from every single voter in the country. The key word, however, is "likely." For every 100 surveys done, we expect 5 of them to get it wrong and give you a range that doesn't line up with reality. We call that a 95% confidence interval because 95 out of 100 surveys will get it right, but we can't actually know if any single survey is one of the 95 right ones or one of the 5 wrong ones.

[datatitian](#)

Wow. That is pretty impressive. Let me look at what you did more closely since you spent so much time on it, and I will get back to you.

In light of the fact that no major election in U.S. history has been decided by a single vote, it often seems pointless to show up at the ballot box, at least for an individual voter. As each individual vote has a tangible cost (time, gas, convenience, etc.), how should a statistically literate citizen view voting?

[2nd bike concussion](#)

This is interesting question. If every voter believed their vote did not count, eventually either no one would vote, or it would converge to someone. Local elections can be heavily influenced, but even a larger election--politicians are jockeying to get your vote, so you do have some influence.

Hi John, big fan of your book, Everydata. I'm a huge baseball fan and have always been interested in the transition of Major League scouting from the "Old School" to the "New School" of sabermetrics. I'm not sure how familiar you are with the game, but if you are: If you were part of the conversation back in the late 1800s or early 1900s that led to the creation of the Batting Average, which was then used as the ultimate arbiter of talent until the last few years, what would you say?

[yeoman29](#)

I am a huge baseball fan actually! If I were around back at the turn of the century, I would caution that although batting averages can contain valuable information, it is important to be aware that averages can lie. You might misread a player's talent by ignoring power hitting, or overemphasizing outlier performances.

Hi John,

Thanks for taking the time to do an AMA!

- As an expert witness, what are some of the most-used tools in your toolbox?
- As a data scientist, What are some emerging data analytics tools that you think folks should know about?
- Does your style/approach change depending upon whether you're dealing with lawyers or dealing with juries? How so?
- What's your take on microsimulation?

[Imaotsetung](#)

Being a good expert witness requires the ability to synthesize information and explain concepts carefully. Although I rely on my statistical training daily, my ability to teach is critical to the being an expert witness.

Advances in cloud computing are fascinating to me. The speed and size of our data sets expand on a daily basis.

Yes and no. My job is to give my objective opinion, so that doesn't change. But, attorneys think about issues from a certain perspective which is different than a general audience found in a jury.

Microsimulations have a valuable place in our statistical literacy and advancing our knowledge, but of course, as a true empirical economist, I love real data.

Thanks for doing the AMA! What are your thoughts on different macroeconomic indicators that are used in news and politics. Are there any that we should completely abandon ? I've noticed often media fails to include confidence intervals on economic projections and act surprised when projection don't match reality.

Also, any thoughts on Bayesian approach to data reporting / interpretation?

[Warlord Achilles](#)

With any set of macro data, I think the important point is to view numbers in their totality. There is no magic "indicator" that tells us everything we want to know about the state of the economy.

I am glad you mentioned confidence intervals-- labor numbers, for example, are reported down to the nearest 1000 employees, but the confidence intervals can be in the 100,000. People don't pay attention to that enough.

Hey John,

This is actually the first I've heard of you, but your book "Everydata" sounds quite interesting. I was wondering if you had any thoughts as to what the most accurate polling method is at this point. I was listening to the 538 Podcast and their discussion of the difficulties of phone polls vs internet polls, and I was curious what thoughts you had towards that subject. Is there a better method we should be using, or are we stuck with just variations on these two?

[ChillBro69](#)

Polling is a big area of interest to me, and I have been looking at some of the recent polls. First, I don't think you can generalize that phone or internet are necessarily always better. Phone polls have the well known bias that the samples tend to be people who own phones--skewing older. Internet polls have the advantage of a broader sample perhaps, but are skewed towards those who choose to participate. As an aside, fascinating issue in Europe right now with polls on Brexit where the internet polls and phone polls consistently give completely different results.

We are always looking for new ways to gather information and survey. The key is sometimes conducting good polls requires money and time.

How does the "misrepresented" data that you talk about come up in economic consulting?

[finfan96](#)

Economic consulting as a broad field involves a wide range of both litigation work and other business advisory work. Since so much of the work is empirical in nature, the ability to think about numbers carefully and understand what they mean (and what they might not mean) is a part of our daily work.

Hi John,

Completely off-topic but how do you feel about your first name being almost the same as your last name? Would you rather your parents gave you a different name?

[vulpa](#)

Actually, that's pretty funny. I am actually the IV, which means I am the fourth generation of John Johnsons!

What's your thought on how data is systematically manipulated for political ends?

Just on reddit alone, to cite some provocative examples:

- We have folks who can cite the percentage of people in federal prisons for drug charges, but who don't know that the vast majority of people in prison are held in state prisons, and not for drug charges.
- We have folks that say that there's no example of unemployment going up after you raise the federal minimum wage. As long as you don't count 2007, 2008, and 2009, that is. (Clearly those don't count.)
- We have folks that think that the top 1% of taxpayers pay lower income tax rates than the average man on the street, even though they don't, because they just know someone is paying 15% capital gains tax rates. That were raised almost four years ago.

Stuff like that.

And of course there are misstatements the other way, but I figure the audience here is informed on those more frequently.

[yes_its_him](#)

In a political year, the proliferation of "bad statistics" and "biased numbers" is something everyone needs to be aware of. When I speak about statistical literacy (including in my book) I talk about the fact that heightened awareness of (1) where numbers come from (2) the source (3) how they are potentially cherry-picked is vital to not getting misled by numbers.

What do you think of Richard Thaler and the associated ideas behind behavioral economics? It seems as though your skillsets fall in line with their ideologies in regards to economics; but our current US political cycle seems to ignore this way of thinking. What are your thoughts?

[darkgrey](#)

Behavior economics is a powerful set of tools amongst economists, but like all things, it has to be applied thoughtfully. The notion that our theoretical models can more closely approximate real human behavior and decision-making is a very good thing, on net.

I will have to keep an eye out for your book. It appeals to my sense of curiosity regarding data and its use in the 21st century. Keep up the great work!

[Chairsniffa](#)

Thanks. I got word yesterday it is available in Australia now.

What's the most misrepresented statistic used today?

[tombrady4prez](#)

Where do I begin???

Let me couch it this way. When I see in the newspaper the following phrases, I pause:

"New study says..." "4 out of 5" or "9 out of 10" "Trust me..."

Hey! What is one of the hardest things about your job, and what would you recommend to a teenager to get started in becoming a data scientist? Thanks for taking the time out of your day to look at my question.

[imanapple1](#)

Good data work requires meticulous attention to every detail. From shaping and cleaning a data set to framing the question to conducting the analysis. As someone who wants to become a data scientist, start with math and programming courses. Learning how to think analytically is a critical skill.

I feel like I see headlines about studies showing how amazing red wine is for you yet I also see some that say the inverse. So how should I interpret headlines about statistical studies and any tips of how to tell which ones I should believe?

[unbrokenwindow](#)

You will see these studies almost every day--I found 2000 studies on coffee claiming it both cured and prevented cancer. So, how do you know what to interpret--first, look at the source. Is it from a reputable journal or reputable University? Is it funded by a specific interest group? Also, be wary of the "shocking new" headline that seems to overturn a generation of research.

Whenever I read the term "social justice" these days, my mind immediately conjures up images of millennials acting self-righteous about complicated economic issues that they've learned about entirely through social media.

As a professional researcher who chairs the board of a social justice nonprofit, how do you view the modern state of the "social justice" conversation? And furthermore, how do you believe that social and economic justice can be constructively included in pragmatic policy dialogues?

[Indifferent2Apathy](#)

There is a wide range of social justice organizations that address a tremendous number of societal problems. The particular organization I am involved in focuses on systemic change and data based research and solutions. What I have learned as a Chair of a non-profit board is that there are a wide range of practitioners who can bring their skills to bear on these issues.

Hi Mr. Johnson. My dad, Bill D., works a few rooms down from you. I told him I would post here. Can you tell me how awesome it is working with such a cool guy?

[PhrygianHalfCadence](#)

Yes, it is.

Hi John, thanks for sharing your time!

What has your career path been like? At what point did you decide to found Edgeworth, and what factors led to that? What advice do you have for people considering a similar path?

[viscount16](#)

I started an academic back in the late 1990s. I was always more interested in real world problems. I

started working at a large consulting firm for several years, and had a very positive experience. But, i always wanted to build my own firm with its own unique culture. So, in 2009, I started Edgeworth with 6 employees. Today, we have over 80 employees.

If you are an entrepreneur at heart, I would suggest that you make sure you have a good plan in place, and that you develop a strong business plan. I have also been very well served by my training as a professional economist and statistician.