

EconoFact Chats: The Art and Science of Data Visualization

Ben Casselman, The New York Times

Published on 10th October 2021

Michael Klein:

I'm Michael Klein, Executive Editor of EconoFact, a non-partisan, web-based publication of the Fletcher School at Tufts University. At EconoFact, we bring key facts and incisive analysis to the national debate on economic and social policies, publishing work from leading economists across the country. You can learn more about us and see our work at www.econofact.org.

Michael Klein:

There's a saying attributed to Mark Twain -- "there are lies, damned lies, and statistics." While it's true that some statistics mislead rather than inform, it's through statistics that we understand much of the world around us, especially in the area of economics. Statistics enable us to make sense of vast amounts of data, but in doing so, they necessarily mask certain information as well. And the graphical representation of data in statistics is a particularly powerful way to convey information.

Michael Klein:

In recent years, newspapers have upped their game in the presentation of data and statistics. One of the people working in this area is Ben Casselman of the New York Times, my guest today. Ben writes about economics and other business topics with a particular focus on stories involving data. He previously served as chief economics writer for the data journalism website, FiveThirtyEight, and before that, as a reporter for the Wall Street Journal.

Michael Klein:

Ben, welcome to EconoFact chats.

Michael Klein:

Ben, many people think that economics is all about numbers, and to a certain extent they're correct. Much of economics involves empirical analysis. Interpreting and presenting data is an art as well as a science. What do you consider when presenting results from data to readers when many of them have little or no background in statistics?

Ben Casselman:

Yeah, so this is something that I think I've thought a lot about in my career and that certainly the Times has thought a lot about. I think that the central question that we have to approach things as journalists is what information are we trying to convey to readers, and then what do they need to know in order to understand that information?

Ben Casselman:

In a sense it's not that different from what a researcher might be doing, an economist might be doing in a paper, trying to understand the world and then present that to their audience. Except that in economics, there's this whole language that exists around communicating data. We talk about statistical significance and magnitudes and all these things, and in journalism, we can't use that language because it's not going to be understood very well by our readers.

Ben Casselman:

And so we have to come up with ways to communicate clearly, information that's clear and accurate in a language that people can understand. And I think crucially we have to convey uncertainty, because uncertainty is central to any sort of statistical understanding, but we can't use the language of statistical significance.

Ben Casselman:

And so we have to try to help people understand qualitatively. We have to help people understand the difference between something that we know and something that we might have some evidence for, but the jury's still out. And what we might just have an intuition about and a theory about. And my job is to make sure that readers understand where on that spectrum we are on any given topic.

Michael Klein:

So that's what we try to do with EconoFact as well, because our readership are people who are not necessarily well steeped in statistics, but our contributors are people, economists who are often distilling work from their own research.

Michael Klein:

Some statistics that are presented weekly, like the new unemployment claims, or monthly like the unemployment rate and the inflation rate, garner a lot of attention, but they're often issues with these numbers. And this goes to your point about having people understand.

Michael Klein:

For example, last year we had the striking statistic that GDP declined by about 31% in the second quarter at an annual rate, and rose by about 33% in the third quarter, again at an annual rate, which actually meant that at the end of the third quarter, GDP was still lower than at the beginning of the second quarter. How would you work to make these regular statistics, especially this example of GDP, understandable to the general public?

Ben Casselman:

So I think this is a great example, because it's one where the statistic as it's conventionally delivered could really mislead people. So let me make a couple of points here.

Ben Casselman:

One is that 31% decline in the second quarter of last year. I think it's easy to hear that as a lay reader and think, well, economic output shrank by a third, shrank by nearly a third. That's not right, because we measure conventionally, we report GDP at an annual rate. So that means that if the declines continued for a full year then we would lose a third of output. And of course we knew in the spring of last year, or at least we certainly hoped, that we weren't going to continue to see that dropping knife at that rate. And so there was no realistic world in which we were going to see economic output declined by a third.

Ben Casselman:

And so we at the Times made a decision to report that at a non-annualized rate, which is about a 9% decline, which to be clear is still an enormous decline, and to highlight that figure rather than the annualized number.

Ben Casselman:

So that's one thing. The next thing is the one that you highlighted Michael, which is that we saw this 31% decline and then we saw 33% increase. And I think it's easy to hear that and to think that that means, well, we dropped by a third and we're back by a third, and so now we're back to zero. Which is again not true. If you start at a 100 and you lose 50%, you're down to 50. If you go up by 50%, you get to 75, right. You're still a gap. And so we made a really big point at the Times in emphasizing both in our language and in our graphics where we were relative to where we had begun.

Michael Klein:

And you were right, of course that the 9% decline in the quarter didn't continue. So it was really good that you had the foresight to do that and explain things more clearly. Ben another problem with standard data is what is called base effects. When there's an unusual year, like 2020, the annual statistics for 2021 will be compared to that year. For example, inflation looked really high when we compare 2021 prices to those in 2020, but much less so when we consider what happened, say in 2019. Do you think going beyond standard government announcements would be an important way to put statistics in context?

Ben Casselman:

We often, I think in furthering that goal, that I laid out at the beginning of sort of helping to highlight the information that readers need and to give them the context they need. That means often presenting numbers either differently than the way they're conventionally reported or providing some extra context so that people can understand them. And sometimes it just means making sure that we are taking a step back and helping people understand the way these numbers are reported. So the base effects example with inflation is a really good one. We knew this was coming right. Everybody knew that we'd seen this unusual drop in prices in the spring of 2020 and so that then a year later it was going to make it look like there had been this big gain in prices. And so we highlighted that ahead of time, right?

Ben Casselman:

We wrote a story that ran before the numbers came out to sort of say, you're going to see these big, scary inflation numbers, but don't get too excited. Here's what you need to understand. And then when we reported the numbers, we made a point of calling out the base effects. We actually had a really great graphic I think that illustrated the part of this that was base effect and the part of it that was real and that presented those numbers in a way that people understood that the central points that they actually needed to take away. In this case, it did require getting into the weeds a little bit in a way we don't normally, because it was actually crucial to understanding those numbers.

Michael Klein:

So there's a saying "unusual times, call for unusual measures." They don't usually mean economic measures, but in this case, I guess that's exactly on point. I know you're involved, Ben, in presenting other data in innovative ways as well, beyond those statistics that are just regularly released. For example, I know you were involved in creating a very impressive series of graphics that demonstrated for different types of families, whether the 2017 Tax Act would increase or decrease their tax liabilities, and by how much. Could you briefly describe those graphs and what it took to create them, and what they showed?

Ben Casselman:

For context, this was around the tax bill that was being debated in 2017. And there was a ton of debate at that time about how the bill was going to affect the middle class, specifically. And most of that ended up looking like either just some numbers on the share of Americans. X percent of Americans or X percent of middle-class Americans will get a tax cut or a tax increase under the bill. Or else there were a lot of pieces

that sort of cherry-picked typical examples, they'd say for a family of four earning \$75,000, here's what it would mean. The thing is, is that taxes are really complicated, and people's incomes are also really complicated. And so what that means is that any change in tax law can have really different effects, even for people who look relatively similar in many ways.

Ben Casselman:

And we really wanted to help people understand that and highlight that. And so what we did was we created this graphic that took 25,000 American households and simulated what the tax bill would mean for them, and then showed them as dots where you could see how much people made and whether they would see a gain or a loss under the tax bill. And what it meant was that you could really clearly see two things. You could see that most people were getting a tax cut, that there were more dots on the tax cut side than on the tax increase side. And that was important for people to understand. But then it was also clear that you could see that there was one household earning about \$80,000 that would get a \$5,000 tax cut. And another household earning about \$80,000 would get a \$3,000 tax increase. And so it helps people see the breadth of experiences and why, then we could sort of go through and take some time to understand sort of why this was all so complicated.

Michael Klein:

Yeah. That was a very impressive graphic. And it did show how it was complicated. And also, you know, as you mentioned, the heterogeneity of things, so things are not always as simple. So that graph was really good, but they're also graphs that aren't so good. And we've been talking about successes in presenting data, but there are data and statistics that are particularly challenging to present, or even which are presented poorly. Edward Tufte is well-known for a series of books on the presentation of quantitative information. And along with the good examples, he also presents examples where the graphics misrepresent the data. For example, by having the data disproportionately presented in the figures, presenting figures that lack context or presenting inconsistent data. Do you have any pet peeves about the way certain data are presented? And are there any areas in which you think it's particularly difficult to make data and its associated statistics hard to understand?

Ben Casselman:

So Tufte's work has been really influential on me, and I think on newspaper data visualization, media data visualization people broadly. A lot of the things that he highlights, I think we think a lot about. You see examples of graphics that for example really zoom way in to make small changes look really big in a way that I think can be misleading, sometimes intentionally misleading, other times by accident. The place where I've thought a lot about it, I think is any time that we're showing relationships. I think there's we have a couple of challenges. One is that infographics, data visualizations are really powerful. People really remember them and they take a lot from them. But people are often not very good at reading them.

Ben Casselman:

And so there's a risk that if we don't do a good job of making something clear on a graphic, that the one thing people will remember from a story turns out to be inaccurate. So I talked about sort of relationship between a couple of variables, right? If we're trying to think about the relationship between taxes and growth, for example, we're trying to think about the relationship between any sort of two variables. If we throw a line onto a chart, people are going to look at that and often take it very literally, even if we're really not very sure about that relationship. They're going to assume causality where we really may not have much evidence to support it. And so we have to think really carefully about whether to present that relationship. And then if we do, how do we make really clear how certain we are in it, and whether we believe there's a causal relationship.

Michael Klein:

So that takes me to my next point. As you know, economists try to analyze the effects of one thing while holding other things constant. And the Latin phrase *ceteris paribus* meaning all else equal, is probably familiar to most of our listeners who have taken at least one economics course. If you don't hold all else equal, like the tax and income example you were saying, then you can get misleading statistics. Another example, when you're trying to analyze the effects of discrimination on wages, you get too big a number if you don't control for factors, other than say, race or gender. How do you account for this when you present statistics to an audience that might not have learned the two words *ceteris paribus*? And also, how do you inform your readers about this important caveat when looking at statistics?

Ben Casselman:

So I think there are a couple of things here. One of them is we have to think really hard as journalists about what it is that we're actually trying to communicate to our readers, which may or may not be the same thing that a researcher is trying to study in their paper. So a researcher, to take your example about discrimination, might be trying to zoom in very carefully to understand if two people are applying for the same job, is there a difference in the likelihood of them getting that job, or is there a difference in the wage that they might be paid for that job. And they're trying to control for everything so they can really study that question. That might be the question we're asking this journalist, but we might also be asking a question about the broader equality equity in society, where it may well be that we don't want to control for all those things. We may want to understand, you know, the relative incomes of black and white households, for example, not controlling for everything.

Ben Casselman:

And so I think one thing we have to stop and think about is what is it that we're actually trying to measure. But then to the extent that we are trying to understand things, *ceteris paribus*, trying to understand things with controls, we have to do a lot of work on our end, under the hood to make sure that we understand the subject fully so that then we can present the much boiled down version to readers. I may read a whole lot of economics papers. I may do some of my own data analysis. I may do data analysis that I can share with an outside economist to say, does this make sense. That may all get boiled down into a couple of sentences in an article, but there's a couple of sentences that I'm really confident in. And so that when I present a simple relationship to people, that I know that under the hood, there's a lot more going on there.

Michael Klein:

So it's like the saying somebody once wrote to somebody else "I'm writing you this letter if I had more time, I'd write you a shorter letter" because it's hard to present things very concisely.

Ben Casselman:

No, we have a responsibility to do that work and to take that time, to be able to say it in a short way.

Michael Klein:

Yeah, that's what we do at EconoFact as well, where we have to boil things down for an audience that won't have the background necessarily that the original research was written for people who do have the background. So Mark Twain's quote that I mentioned in the introduction can be interpreted as scoundrels using statistics to make false claims. But another interpretation is that some statistics are just flawed because it's difficult to collect data. What statistics do you consider either the most reliable or the most unreliable and ones that raise big questions about the accuracy of what they're supposed to measure?

Ben Casselman:

It's definitely true that there are statistics that are outright flawed. Data that's collected badly. But I think the thing we actually have to guard against the most is misinterpreting statistics, or having statistics that don't measure the thing that we're actually interested in. So this could be the case with a government statistic, we actually have a lot of confidence in the data collection and the work that's been done. But that just doesn't do a very good job of measuring what we're interested in. The unemployment rate is an example that I often give. The government definition of unemployment is fairly narrow. It only counts people who are actively looking for work. Which means that in this moment that we're living in right now, if people can't work because their kids are home from school, doing remote school, they may not be looking for work. They're not going to be counted as part of the unemployment rate, but they are going to all the same, be a sign of the challenges that we're dealing with.

Michael Klein:

Yeah. We have an EconoFact memo by myself and Kit Baum that looks at these two different measures of unemployment rate, the conventional one, and one that's called U-6 and we make that point.

Ben Casselman:

I think that's one really important example here. Another one is we've used during the pandemic, a lot of private data that has, I think been really useful to understand the state of the economy. People have looked at OpenTable reservations to understand the impact on dining. They've looked at TSA check-ins to understand airlines. They've looked at data from providers of time clocks to see how many people are working. But there's a risk with those that they may not really be representative of everything that's happening in the economy. If the time clock provider mostly works with a certain type of company or in a certain region, then they're going to miss what's happening in the larger economy. That's not because the data is flawed. It's just because it doesn't tell us what we might be most interested in. And so we have to make sure we understand those shortcomings in order to report them and interpret them accurately.

Michael Klein:

So Ben, you raised the issue of COVID and you've written a lot about the effects of COVID on the economy and the efforts to mitigate the economic downturn resulting from the pandemic. What do you find most surprising after analyzing the data, something that the data illustrated that you didn't expect?

Ben Casselman:

I think that the clearest sort of most important takeaway in some ways from the data is just how successful the economic policy response was to the pandemic. Not that it was perfect by any stretch of the imagination. And we could talk all day about ways in which it fell short. But I think that you could sometimes get the impression if you just followed the sort of anecdotal reporting that poverty was rising, hunger was rising, we saw lines at food banks, we saw all of this, but when we've gotten the data, we've actually been able to see that poverty, at least by one measure, fell to a record low last year in large part because of the supports that we provided. Hunger did not rise over the course of the year. Incomes actually rose. And so in a lot of ways, the economic policy response was tremendously effective at mitigating the damage of the pandemic. And that's something that we really only know because we had the data to illustrate it.

Michael Klein:

So the data showed that the experience with COVID wasn't at all typical. And so I guess that means there was a really important time to let the data tell us what's going on rather than to rely on past relationships. But we've seen that with issues related directly to the pandemic, there's a real skepticism among some

people about the data and a feeling among some people that the government is not playing it straight. Do you think that this kind of view will start to affect people's acceptance of other statistics as well - not just those related to COVID? Like the growth rate of the economy, the rate of inflation or joblessness?

Ben Casselman:

Yeah. I mean, I think we've already seen it. I think we saw it last year during the pandemic, there was a lot of skepticism of the employment data. For example, I had a long back and forth with, among other people, Howard Dean, the former Vermont governor and presidential candidate.

Michael Klein:

Did he scream at you?

Ben Casselman:

He Twitter screamed at me.

Michael Klein:

He's known for screaming. Of course

Ben Casselman:

He's known for screaming. Yeah. I mean there were some known challenges with data collection during the pandemic, but that got sort of spun into this conspiracy theory about how the numbers were being cooked in a way that was not true and I think it was not helpful to helping to understand things. We've seen this with inflation for a very long time, skepticism of the numbers. And I think we as journalists and people who communicate this to the public, have a responsibility to help people understand the real flaws in the data, but also the ways in which it is to knock down sort of baseless conspiracy theories.

Michael Klein:

Well, Ben your work at the Times and ours at EconoFact are related. We're both trying to raise the level of public debate by getting facts and good analysis in the public sphere. So thanks very much for speaking with me today and also thanks for your efforts along these lines.

Ben Casselman:

Thanks so much for having me. I enjoyed it.

Michael Klein:

This has been EconoFact Chats to learn more about EconoFact and to see the work on our site, you can log into www.econofact.org. You can subscribe on our site to our newsletter that will let you know when we publish new memos and new podcast episodes. Please feel free to share this podcast and our memos with friends, colleagues and on social media. EconoFact is a publication of the Fletcher School at Tufts University. Thanks for listening.