How You're Being Lied to with Statistics

In 1965, Darrell Huff testified before the U.S. Senate to explain how statistics could be used to mislead. He gave clever examples like how in parts of the Netherlands, the number of children born in a house was positively correlated with the number of nearby stork nests.

Did that mean storks were delivering babies? Of course not.

As Huff explained, larger houses were more likely to have both more children and more chimney pots which were perfect for nesting storks. It wasn't causation. It was a coincidence disguised as statistics.

Huff knew the tricks because he literally wrote the book: ***How to Lie with Statistics,*** the best-selling statistics book of all time.

But he didn't tell the Senate everything.

He didn't mention that the tobacco industry had paid him to be there with a mission to discredit studies linking smoking to cancer. And while he said he was working on a follow-up book, he didn't tell them the title: *How to Lie with Smoking Statistics*.

That book was never published. But the story shows how easy it is to be fooled by data.

The Bottom Line

Next time you're looking at data, stop people from tricking you. Ask yourself: How would this look if I zoomed out and saw more context? What might have been left out? And who collected the data in the first place?

3 Tests for Misleading Data

So how do you know if someone is misleading you with stats? Here are three quick tests.

First, think of data as a picture. Why? Because all data is from the past—it's a snapshot in time. And just like photos, people tend to show you only the best ones.

1. The Zoom Test

Ask: How different would this number look if I zoomed in or out?

Would the trend hold if you looked at a different time frame, zoomed out to include more people, or broke it down by geography or demographics?

2. The Crop Test

Ask: What's missing from the frame?

What's been left out that would change how you see the picture?

3. The Source Test

Ask: Who took the picture?

If the same person or organization collecting the data is also presenting it, check their incentives. Like a selfie, statistics can be flattering when the subject is also behind the camera.

Sources: United States. Congress. Senate. Committee on Commerce. Cigarette Labeling And Advertising: Hearings, Eighty-ninth Congress, First Session. Washington: U. S. Govt. Print. Off., 1965. Reinhart, Alex. "Huff and Puff: A Statistical Counterattack to Public-Health Fears about Tobacco." *Significance*, vol. 11, no. 3, 2014, pp. 14–17. Wiley, doi:10.1111/j.1740-9713.2014.00765.x.

3 Tests for Misleading Data in Action

Let's look at an example and apply the three tests.

A recent Shapiro Administration press release claimed:

"Currently, proposed changes to Medicaid could strip health coverage from more than 300,000 Pennsylvanians..."

1. The Zoom Test

- The administration used the highest possible estimate from their own projections.

2. The Crop Test

- The 300,000 figure includes healthy adults who could lose coverage only if they fail to meet a new work requirement: 20 hours a week of work, training, or volunteering. The estimate assumes most people won't meet that standard, even with years to prepare before the policy takes effect in 2027.
- The number also includes people who would be removed because they are not eligible for Medicaid. If someone doesn't qualify for the program, then they shouldn't be on it. How is that a cut?

3. The Source Test

- The figure comes from the Shapiro Administration and is based on Congressional Budget Office estimates. But it's framed in a way that supports the administration's argument and leaves out important context.

With any data you are given, it's important to be skeptical. **The final test is to do a gut check.**

Do I want this data to be true?

We're most easily fooled when the numbers confirm what we already hope is right. In 1965, Southern senators wanted to believe that smoking didn't cause cancer. Tobacco was a major industry in their states. Their friends and neighbors worked in it. Believing the data meant accepting hard truths they didn't want to face.

Next time you're looking at data, ask yourself:

- How would this look if I zoomed out and saw more context?
- What might have been left out?
- And who collected the data in the first place?