

How visualizations can lie

Let's talk now about how charts may mislead. This is a very common problem. And this is the topic that I covered in my most recent book, *How Charts Lie*, which is the first one that I have written for the general public to be better prepared to read charts, and graphs, and maps more effectively. It's a book sort of like presented this way, a book that teaches people how to become better readers of charts. And I believe that this is an essential skill, a fundamental skill if you want to work in data journalism. Many of you may end up not designing charts and graphs and maps for a living, but all of us will read charts, and graphs, and maps on a regular basis. So it is very important, I think, to start abandoning certain myths that surround data visualization.

For example, how many times have you heard the saying a picture is worth a thousand words? This is not true, as I hope that I will prove to you in just one minute. Or visualization is intuitive, we have sort of this bias to think that, you know, graphs and maps and charts can be understood in the blink of an eye just by taking a quick look at them, and that is wrong. Charts need to be read not just seen in order to be understood correctly. Or the last one is very common in the world of business analytics. The data should speak for itself, just show me the data. Well, that is wrong, data never speaks for itself. Data needs to be made to talk in order to present, in order to be able to extract meaning from that data. So we should abandon, abandon these myths. Or as I usually say in my classes and also in the book, perhaps we could append, append a second portion or a second element to all these sentences saying, well, yeah, a picture may be worth a thousand words, but only if you know how to read it. If you don't know how to read it or if you don't pay attention to the chart, the chart will not be worth a thousand words. The chart needs to be explained. You need to explain it to yourself. What is the chart saying? What is the chart conveying? So let me show you a couple of examples of what I mean.

Take a look at this chart. This chart is as straightforward as it gets. It's a chart similar to one that we just, we just have seen. It's a chart that shows the homicide rate in the United States from 1960 up to 2015. So what the chart is showing is a story that many people in the United States already know, which is that violent crime rates and in particular particular homicide rates increased very sharply, very drastically during the 70s and the 80s, so the line goes up. Then the line went down. Homicide rates declined over the 90s for different reasons. There are many reasons for these, 90s. They stayed more or less the same during the 2000, and then in the past two or three years homicide rates have begun increasing again. This is a challenge and it's a challenge that we need to deal with. What can we do about this? What can we do to avoid having so many homicides in the United States? The problem, though, is that this chart, if you don't know how to read it, can be really, really misleading.

Let me tell you how some people read this chart. You take a look at this chart. You notice that it shows you the homicide rate, homicide rate in the United States. And you read this chart as saying the United States is becoming a much more dangerous country. Is this true? No, it is not, because this is basically showing you the average over the entirety of the country. But that doesn't really mean that every single place in the United States has a similar homicide rate. It may happen that certain places in United States have very low homicide rates, but other places have very high homicide rates, and when you calculate the national rate, basically these places that have very high homicide rates are skewing the national rate. This is what is actually happening in the in the country at the moment.

If you talk to statisticians who deal with crime statistics on a regular basis, probably they will tell you that most places, most cities, most towns, most neighborhoods in the United States are pretty safe. Homicide rates are quite low in most places in the United States. If we could plug them on the chart, right? We could imagine each one of these places is a little dot on the chart, right? As I'm seeing on the screen right now, most places in United States are very, very, very low on that homicide rate scale. What is the challenge, though, that there are certain cities and in particular a certain area, certain neighborhoods in certain cities and towns in the United States that in the past few years have become so dangerous and so violent that if we tried to plug them on the chart, they will go through the roof. It will be impossible to place those neighborhoods on that scale because the homicide rates in those places are so, so high. Now, these places are skewing the national rate, right? They are making the national rate increase. These are the places where homicide rates are increasing the most as an effect of that, they are also making the national rate increase. But most places in the United States are still relatively safe.

Now, is the original chart a lie? No, it is not a lie, but it is a lie. It will lie to us if we don't know how to read it. If we don't know how to put the data in context. And here comes a second part of this thought. If you are a designer, if you are a visualization designer or if you are a data journalist, it is your responsibility not to show just the original chart, the homicide rate at a national level. It is your responsibility not to simplify the information, but to clarify the information. How do you clarify? By showing the national rate, but then talking about all these outliers and extreme values that may be distorting the national rate. You need to talk about both. You need to talk about the national rate and also the exceptions to that national rate. By adding more information so people can figure out what's going on in this particular story. Again, I'm going to emphasize this again. Never think that data visualization or data journalism at large is about simplifying information. It is about clarifying information. And when we want to clarify, very often, we need to increase the amount of data that we show to the public. The amount of annotations, the amount of explainers, the amount of clarifications that we put in our stories rather than reducing the amount of information that we present to people.