

The elements of a visualization

Let's talk about the different elements within a data visualization. When you see a chart, a graph, a data map in news media or in social media, what are the elements within that graphic? Right? So I usually explain that any data visualization is usually made of several layers of content, so we can begin talking about these layers from the bottom up. So the most fundamental, the most basic, the foundational layer of a visualization is usually what I call the scaffolding layer.

The scaffolding layer is the layer that contains all those elements in a visualization that support the content of that visualization. Then on top of that, you have the encodings layer, and we will get to what that means in just one minute. And then on top of the encoding layers, you have the annotation layer that will contain all the textual elements that you will put in your chart to put the data in context. And then on top of that layer, you may also have what I like to call the "me", quotation marks, the "me" layer on top of that graphic. So, let's let's take each one of these layers one by one and explain what they mean.

First of all, we have the scaffolding layer. What is this scaffolding layer? Take a look at this chart. This chart was designed by The Wall Street Journal. The topic of the chart is completely beside the point, it's a chart about unemployment rates in different sectors of the economy. Don't focus on the content of the chart itself, focus on those elements that support the content. And by that, I mean, for example, the color legend on the upper right corner, the little labels on the axes and the axes themselves, the little tick marks on the axis of the chart, and so on and so forth. That's the scaffolding layer. That is basically what determines how the content is going to be presented, and it puts everything sort of in context, right? It's like you have that below your content.

And then on top of that, on top of those elements, the scaffolding elements, then you put on top of that, you put the content itself. The content itself is what we call the encodings layer. The idea of encoding is fundamental in data visualization. As I have already explained in this lecture, data visualization basically consists of mapping numbers onto certain features or certain properties of objects. So this is a little bit of an abstract idea, but it's very easy to understand. Let's suppose that you begin with a bunch of numbers, a bunch of figures, a data set, and then you choose an object to represent those numbers, a series of objects to represent those numbers. Let's say a series of rectangles. You have a series of rectangles of the same size and then you have the numbers over here. And then when you map those numbers onto those rectangles, there will be one rectangle per observation in your data set. The height of those bars will change according to the numbers that you have on your data set. The result of doing that is a bar graph, right? This size, not the size, the height, the height of those rectangles will vary in proportion to the data that you had on your data set. In that case, we will say that the encoding that we are using is height.

So the encoding, the encoding in data visualization is not the object that you choose to represent your data. It is the property of those objects that you vary according to the data. So in data visualization we could use many different kinds of methods of encoding. This is like, by the way, if you want to learn a little bit more about encoding, I discuss them at length in my book, *The Truthful Art*. There are many different kinds of encoding, one of them is height, right? Like in bar graphs, height or length, length is also used in bar graphs when the bars are horizontal.

But there are other many encodings, for example, position. You could use as objects to represent your data, you can use little dots and then you can vary the relative position of those dots next to each other in proportion to the data that you have in your data set. In that case, you will say that the encoding is not the dots, the encoding is the position, encoding is position, right? You could use area, for example, when you create, say, maps in which you put bubbles on top to represent your data, unemployment rate in the United States and you have little bubbles that vary in size according to those numbers. The encoding in that case is not the bubble. The encoding in that case is the area, area is the encoding, right?

Encoding is, sorry, area is also the method of encoding used in a kind of graphic called the tree map. The tree might be sort of that rectangular shape that you have there on the right hand side of the image, right? You show a total, right? That's the big rectangle, and then the subdivisions of that rectangle shows you the, you know, the relative the relative size of the components of that total in that area. In that case, the area is also the encoding of that chart. You can also use angle, right? Like in pie charts the encoding is angle and as a consequence of that, the area of the segments of that pie chart are also, is also proportional to the data that you're trying to, that you're trying to represent. Line weight, for example, is another way to represent your data and then color, color, both color, hue, different colors or color shade, the intensity of those colors can be used to represent your data. But there are many other kinds of encoding, these are just the most common ones in data visualization.