How to choose the right encodings

Let's now talk about certain essential principles in visualization use for effective communication. If you remember during this lecture, we talked about the encoding slayer. Right? The way that we represent data through the variation of certain properties of objects: length, height, position, area, angle and so on and so forth.

Now the question is how do you choose? How do we choose whether we should use length, height, position, area, or color shade to encode our data? Well, fortunately, there are certain tricks, certain principles, certain guidelines, that you can use to guide yourself through that decision. The first thing to think about is that, again, that idea that I presented earlier, which is a visualization is certainly something that should look beautiful, nice, attractive, engaging. But the main purpose of a visualization is to make data clear and to enable people to see trends and patterns in the data. Therefore, a visualization is a tool. Right?

Therefore, whenever we design a visualization, we need to ask ourselves, what is the purpose of this visualization? Or what is it that I want to show? What is it that I want people to see in the data?

And based on the answer to that question, you may choose one encoding or a completely different encoding. Let me show you an example. This is a chart that was designed by the European Court of Auditors in Europe. This is one of the branches of the European Union. It's a chart that is showing what percentage of migrants that arrived to Greece in 2016, where they came from. So what percentage came from Syria? What percentage came from Afghanistan? What percentage came from Iraq? And so on and so forth. And as you can see, they are showing the data using this variation of a pie chart that we could call a donut chart because it has a hole in the middle.

Now, the question is, is this a good visualization or a bad visualization? Well, the answer to that question is: it depends. It all depends on what you want to emphasize. It all depends on what you want to show. It all depends on what you want people to do with your data. This is the way that I explain it.

If the purpose of this visualization is to let people see that nearly half of the migrants who arrived to Greece in 2016 came from Syria, and then the other half came from other countries. But you don't care that much about which countries. You could care more about this sort of half-versus-half message then this chart is fine. I mean, the only changes that I would make to this chart would be to emphasize that message even more, half and half, by using one color for Syria and then another color, a single color, for all of the other countries in this mix. Half versus half.

But what if the purpose of this chart is not to show half versus half? What about if the purpose of this chart is to let readers compare these countries to each other without having to read every single number? So you want to estimate sort of the size of Syria, the percentage of people who came from Afghanistan, from Iraq, and from so on and so forth.

Well, this chart is not that effective for that. If you don't read the numbers, if you tried to compare the segments of the chart to each other, you will need to squint a little bit and then use your fingers, to sort of like, you will need to rotate your fingers like that. "Oh this is

bigger. This is slightly smaller. This is slightly smaller." It's hard work. It's hard work to do. Right. So if the purpose of this visualization was comparison, perhaps the pie chart that is using angle, an area, to encode the data is not the right solution. Perhaps you will need to encode the data, perhaps using length or height, like in a bar graph, so people can compare these countries to each other.

Third alternative. What about if the purpose of this visualization is not to enable comparison, it is not to show parts of a whole, and show half versus half. What about if the purpose of the visualization were to show where all these countries are, and whether there is some sort of relationship between the percentage of people coming from these countries to Greece and the distance between these countries and Greece?

In that particular case, perhaps the way to represent this data, the right way to represent this data, will be a map, showing where all these countries are. So this is basically shows you these three alternatives.

The pie chart, as I said before, I think that is fine if the message is to show half versus half. It's only that we change the colors. We emphasize Syria, and then we deemphasize all the other countries by coloring them with the same color, with that gray color. That's fine if half versus half. If e want to compare and rank these countries, though, the bar chart is a little bit better. We have sort of empirical evidence that shows that when the purpose of a visualization is comparison, accurate comparison, the bar graph is actually better than the pie chart. Because we are able to compare lengths or heights much better than we are able to compare angles or areas, which are again the encodings that are used in a pie chart. And then if the purpose of the visualization is to show where those countries are, then the map is the right solution.

So purpose is always the key word when we think about the design of any data visualization. What is my visualization for? What it is that I want people to see or to do with the visualization that I'm presenting to them.