The Scatter Template

So in this video, I'm going to be talking kind of more in-depth about the scatter template, and this is a template which is possibly the most wide ranging in its uses of any Flourish template. We always kind of joke when we're developing a new template like, oh, maybe we can just add it to the scatter chart, scatter plot, rather, because it can just do so many things. And I wanted to start here in the template chooser, because I wanted to show you visually how many different things scatter can do. So these are all the starting points that we have with scatter, and these are all sort of like different types of charts, really generally actually different types of charts that you can make with the scatter template. I think the only one that's not on here that you can do is beeswarm plots.

So I'm going to kind of go through them and explain the basic structure of the scatter template and then kind of show you how each of the other plots are made. But just know that if you want to make one of these, you don't need to 100 percent make it, just throwing your data into this basic plot, and then scatter plot, and then trying to sort of do the settings yourself in the column settings. There are these starting points which should be good guides.

So I'm going to start here in a basic scatter plot which is showing, I think, GDP per capita and life expectancy from a bunch of different countries. So each of these dots is the country. And I wanted to show you this because I want to show you the differences between a basic scatter plot and a bubble chart. So here's the basic scatter, also good to know is that in the scatter template, the settings panel is kind of enormous and a bit daunting. But don't be too scared by that. A lot of these things, if you're not making that kind of chart, you don't really need to touch them, so yeah.

So here we are in the basic sort of X Y scatter plot with GDP per capita on one axis and life expectancy on the other. And they're doing a little bit of coloring by continent, or region, or whatever. But if we go to the data we can see, if you just ignore everything else, you can just see that the X values are here bound to GDP and the Y values are the life expectancy. And then we're naming each of them and giving them a region for the color. But that's really just a basic, even if I just got rid of these two settings, you can see that these are just a normal X Y scatter plot.

So, I'm going to throw those back on there though, so we've already talked about color and stuff like that, so I don't need to go over that again, but this is a basic scatter plot. If I want to make this into a bubble chart, I need to add a third sort of value dimension, and that would be the value that the bubbles would be scaled by. So in this case, that is the size column setting right here, and I'm going to make that column E. So we're going to be sizing by this population column that we have yet to sort of touch so far. And there you go. So now we can see that the bubbles are being sized by population and it should come up in the pop-up population now.

A couple of setting things to know about the dots, you can kind of change the color palette like normal here, but you can also change in the style. So I like to do low opacity for the dots because, especially if you're using a color palette that's a bit darker, so you can see where the overlaps are. I think it looks quite nice. You can also add a bit of an outline to the dots if you want. I like to keep that quite minimal just to give them a bit of dimension to them. But yes, that's just sort of a basic bubble chart. There are starting points for basic

scatter and bubble charts, so you should be totally fine making those. The other thing I want to point out before I move on, on this template is that we're currently on the x axis using a logarithmic scale, log scale, just to point out that you can do stuff like that in this chart.

OK, so the next one is a strip plot. And this just sort of shows, it's a good way to sort of display data by category and one other variable. It's something that database people, a chart type that they really like to use. So we have it in this template as well.

The binding that you're going to be using for that is going to be this series connect with line binding, but before I talk about that, we should talk about what the X and Y values are. So to make this possible, you see that the X value is this sort of numeric value, and then the Y value is going to be a categorical value. So it's going to be a category of like Europe and Central Asia. And that's what makes these sort of lines here along the Y axis. And then the X value sort of places the dots along there. And you'll see that these are actual like countries and stuff. So that is what the C column is doing there with name.

And then to make these connect. So if I get rid of this and watch this here in the corner, if I get rid of this, it doesn't really do very much, but what it's doing is kind of like making these into a category and it's connecting them with a line. So this makes more sense if you look at it on the dot plot, actually. But just know that the strip plot is basically you have one categorical variable and then one numeric variable, and you can do the same with changing the opacity and everything with these. These definitely having lower opacity is really quite useful, I think, because there's lots of interesting overlap between the dots often.

So now I'm moving on to the dot plot. Here's another example. This is like a very similar chart. So you have this along the x axis, you have this sort of numeric value and then the Y axis is the category. But the y axis, make sure you understand in this one, is also the name. So instead of having one sort of category and then placing named dots along it, like in the strip plot, in this we have the category and then we're placing two other data points along it for that same category. So in this case, we have the country, and then we have the year and the income in that year, and we're connecting them with the dot, if that makes sense. So 2014 and 2004.

And make sure you have this series. In this case, we're not actually connecting by the value like we did in the other one. We're connecting on the country. So if you want this to be by country, make sure that that is the column that you have in this series connect with line. So if I get rid of that, it removes the lines. It just has the dots placed along, sort of like we had in the strip plot. If we put it back, it sort of draws the line between them. And there's all sorts of styling options. There's all sorts of styling options for the line and stuff like that.

The second to last one that I'm going to show you is the distribution option, and that is the box, violin, and beeswarm plots. So basically all you need to do for these is to enter X and Y values. Your X value is most likely going to be a sort of category again, as you can see. And then the Y values are going to be like the value of, from each of those categories, some sort of value. And these are great to sort of show the distribution of values across different categories. So unlike in the strip and dot plots, you're usually going to have your categorical variables along the X and then your values, your numeric values, are going to be the Y.

And if you go on the settings panel to this box, violin, and beeswarms, these are all just sort of different, more sort of statistical options, visualization options that science and statistics people like to use to show distribution. So we have beeswarm, which is where you kind of beeswarm out the values. And then you also have, so you can see they're actually on top of each other here, and when you beeswarm, it spreads them out to make them look kind of like a beeswarm.

And then you can also have, we can also have the box plot to sort of show the median distribution, and then you can also have a violin plot, which is kind of similar to a beeswarm, if you show them, it kind of outlines the beeswarm basically. So those are kind of advanced options and they're not something that a lot of people use, but if you're sort of into science and statistics, they're definitely useful and they're there for you.

And then finally, I just want to show this because I'm not really going to explain it, it's kind of complicated. But we also have this Hans Rosling, which was a famous statistician, and this chart just sort of animates through different years and shows different points of development of different countries in the world. And this is an option. You do a lot of it here in this time slider. And there's this extra time binding column setting here that you have to use. I'm not going to really show it to you now, because if you're into this, you'll be able to figure it out.

But that's kind of a nice place to end. This is probably going to be the longest video because this is kind of our most versatile template, but once you know what it can do, it's really quite powerful. And so thanks for listening.