**Grouped scatter plot**

In this video, I’m going to show you how to make a grouped scatterplot, and just have with a regular scatterplot, both you x-axis and y-axis variables must be interval or ratio and that means they should be defined as scale in SPSS. Now if you want to know more about why that is and why it doesn’t work if it is ordinal, have a look at my scatterplot video. Now what makes a grouped scatterplot different is that I’m splitting it by categorical variable, so here I have gender, and I’m looking at the relationship between age and competency in males and females separately but on the same graph, so I’ve got blue bubbles for males and green bubbles for females. I’m also going to show you how to insert a trend line for both of your sub groups. Just as in the other video, I’m going to use age and competency, and this is for employees so there’s a total competency score out of 100 when they started their job, and we want to look at is there a relationship between how old someone is and how competent they are when they start their job and then see if that relationship is different from males and females. So if we go to Graphs and Chart Builder. We want a ScatterDot, and we’re going to choose the second option. So double click or drag it into your gallery and age is my independent variable because it will not change based on my competency, so if I become more or less competent I’m not going to become older or younger, and my competency when I start my job is dependent so it’s going into the y axis, because it may depend upon my age. Now in order to get those different coloured bubbles for males and females, I’m going to take gender and drag it to my set colour. Go ahead and click ‘okay’. So now I can see the relationship between age and competency for males, which are the blue bubbles and females, which are the green bubbles, and we can see this looks like a similar trend because all the blue and green bubbles are mixed up together. It does look like there’s a few females here that are outside the regular trend, but otherwise, they’re both increasing. So as males and females get older, they tend to be more competent at their job, and again that’s probably because they have more experience and education.

To add our trend lines, double click to open the Chart Editor and up here at the top, you’ve got a icon. It says ‘Add Fit line at Sub Group’ and that’s we want, because I want a separate line for males and females, and I want linear because my trend here is clearly a straight line. It’s not curved, it not quadratic or cubic or anything like that so I’m going to leave it as linear and go ahead and close that if I want. Now anytime you add a Fit line, you should only do a regression or you have done a regression analysis already, so with r-squared values what we do is multiply them by 100 in order to get a percent, so if we take males for example, that would be 8.5%, and that tells us that would be males, 8.5% of the variation in competency scores can be explained by age. And that’s not very high, that’s very low, only 8.5% because it means there’s roughly 91 to 92% of the variation in competency not explained by age for males, and it’s even smaller for females. Be very careful about putting in trend lines. You should really only do it if you’re doing regression, and also, consider if there are any outliers such as these three points here, that might be throwing off the slope of your line. So it looks like our trend should be a bit more steeper but because of these dots down here, it’s made my line a bit more flat so I would consider moving these three individuals while I create my graph and trend line.

END.