**Got It!**

**1. Scatter plots**

The last chapter focused on visualizing one variable. Now, we'll move on to two variables, beginning with scatter plots.

**2. When should you use a scatter plot?**

Scatter plots should be used when you have two continuous variables, and you want to know about their relationship. For example, if one variable increases, does the other one increase too, or does it decrease?

**3. Los Angeles County home prices**

Here's a dataset on home prices in four cities in Los Angeles County in 2012. The dataset includes the number of bedrooms, the sale price in millions of dollars, and the area in square feet.

**4. Prices vs. area**

Here's a scatter plot with the price on the y-axis and the area on the x-axis. To verbally describe this plot, you'd say it's a scatter plot of "price versus area". It's OK, but all the points are clustered in the bottom left, making it hard to read. Let's use a logarithmic scale for each axis. On the logarithmic plot, notice that moving right one grid line doubles the area, or moving up one grid line multiples the price by a factor of ten. Now the points are more evenly spread throughout the plot.

**5. Correlation**

One important concept when interpreting scatter plots is the idea of correlation. Roughly speaking, correlation is a measure of how well you can draw a straight line through the points. If that straight line goes upwards as you move to the right, it's called a positive correlation. If the line goes down as you go to the right, it's called negative correlation. Here are five theoretical datasets. The red line in each panel shows what perfect negative correlation would look like. The green lines show perfect positive correlation. In the left-most panel, you can see an example of strong negative correlation. That means that as the x values increase, the y values decrease. In the right-most panel, you can see strong positive correlation, meaning that as x increases, so does y. The middle panels show intermediate states. In the third panel, showing no correlation, the values of y are completely unrelated to the values of x.

**6. Sometimes correlation isn't helpful**

Here's the Datasaurus Dozen again. Recall that each dataset had the same correlation, despite looking very different. Correlation makes the most sense if there is a straight line relationship between the x and y values. If you have a more complicated shape, you'll need to be more creative in how you describe the relationship. For example, "x and y have a slight negative correlation" is not as good a description as "the plot looks like a dinosaur".

**7. Adding trend lines**

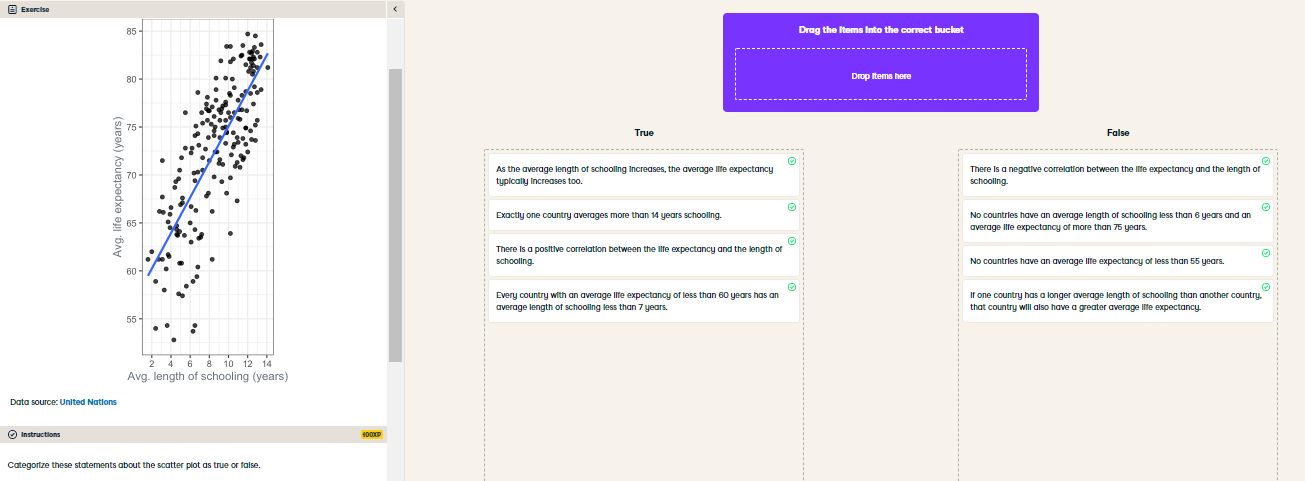
Adding a straight line to a scatter plot is a great way to see you if you really do have a linear relationship between the x and y variables. Here, with the logarithmic scales, the trend line has a close fit to the points, suggesting that as the logarithm of the area increases, you get a linear increase in the logarithm of the price.

**8. Adding smooth trend lines**

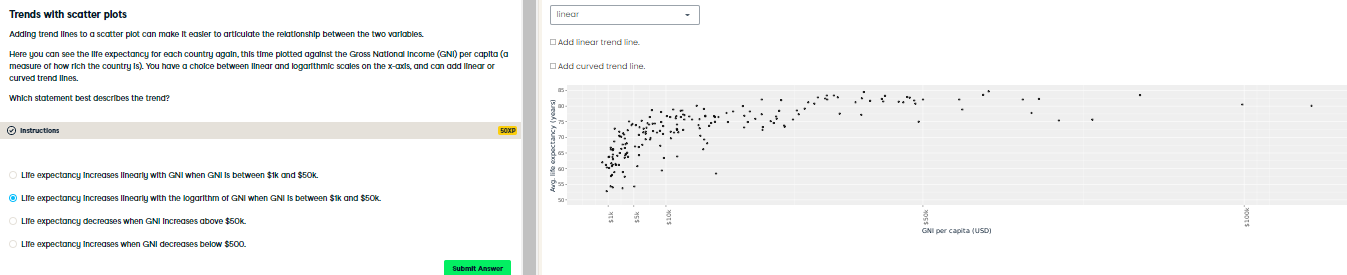
Sometimes a straight line might be a terrible fit. Here, in the price versus area plot using a linear scale, the line completely misses the more expensive homes. When a straight trend line is a poor fit, one alternative is to use a curve. Having a curve like this can help you find a way to describe the relationship. Here, by seeing the trend line curve upwards, you can say "as area increases, the price increases faster than linearly".

**9. Let's practice!**

Let's see some examples.



Skillful scatter plot interpreting! Scatter plots are excellent for exploring the relationship between two continuous varaibles.



Terrific trend line usage! By comparing the linear and LOESS curved trend lines you can easily see when the data is increasing faster or slower than linearly, which can help you describe the relationship between the variables.