Free Training Transcript: Trend Lines

Welcome to this video on Trend Lines. You can download the Exercise workbook to follow along in your own copy of Tableau.

Adding Trend Lines

Trends can provide important insight into an analysis, for example, here answering the question "by how much does our power output go up with wind speed"? We can see the relationship here, when wind speed increases, the power output increases. But how exactly? It's easy to add trend lines to the view. Click on the Analytics Pane, and drag Trend Line to the type of model you desire. Removing trend lines is just as easy – simply drag it off the view. But we want our trend line, so let's undo that.

Trend Line Options

By default, trend lines are per pane and per color. Back on the Data tab, if we bring another dimension into the view, such as Location, we'll get a trend line per pane, or per scatterplot. Similarly, if we were to move Location to Color, our trend line breaks out into three. If we want to see our Locations on Color but only have one overall trend, we can modify the trend line.

Editing the trend line takes a simple right click, select Trend Lines, and Edit Trend Lines. Unchecking "allow a trend line per color" will bring us back to a single overall trend line. We can also uncheck Show Confidence Bands to simplify the view.

There are several other things we can do with this dialog box. First, the model type. These are the same options as when we first bring out Trend Line from the Analytics pane. These options tell Tableau to build a linear regression model based on this type of transformation on one or both variables. Linear refers to the coefficients, not the relationship of the variables. For information on the model types and transformations, the Online Help article on Trend Line Model Types provides some excellent information. The Confidence Bands shows the 95% confidence interval for the model. And we also have the option to force the y-intercept through 0.

Trend Line Significance

It's important to assess if a trend line provides valuable information. Hovering over the trend line brings up a tooltip with the trend line equation, the p-value, and R-squared.

In statistics, p-value is the number assigned to the concept of significance. If the p-value is less than a cutoff value, usually 0.05, the results are interpreted as significant. A

large p-value (on a scale of 0–1) can indicate that the apparent trend in the data is due to chance, not the factors in the model.In this example, the trend line has a very small p-value, which is good.

However, to correctly assess if a model is a good fit, we need to know more than the p-value. We also have the R-squared, which effectively it tells us how our model fits the data. R-squared values can range from 0-1 and higher values are considered better. We see in our example that the R-squared is very high, at 0.956. This means our model fits the data very well – an R-squared of 1 indicates a perfect fit, but beware that if your R-squared is incredibly high, for example 0.999, your model may be misleading. A common indicator of an artificially high R-squared value is low degrees of freedom, or just having too many observations.

Trend Line Residuals

To decide if a trend line accurately represents the data, it's not enough to have a small p-value or a large R-squared. Not all of our points will fall on the predicted trend line. The distance from a given point to its predicted value is the error, or residual.

In a correct model, those residuals should follow a random normal distribution around the zero line when plotted against the explanatory variable. If this residual plot isn't normally distributed, it indicates there are trends in how the data fail to line up with the predicted values, which means the model isn't the best.

To get the residuals values for a view with a trend line, go to Worksheet > Export > Data. This will prompt us to save a file (the only format option is Microsoft Access), we'll name it Trend Line Residuals, and hit save. We'll choose "connect after export". This data source has the original data from our scatter plot as well as the predicted values (from the trend line) and the residuals. The residual plot is constructed as the explanatory variable on the horizontal axis (Wind Speed goes to Columns) and the residuals on the vertical axis (on Rows). And we'll bring Windmill to Level of Detail. Remember, a good model has a normally distributed scatter around zero. Clearly, our current model isn't very good at predicting our Power based on our Wind Speed values. Although the p-value and R-squared for the trend line are good, the residual plot is pretty awful.

Conclusion

That's an overview of trend lines in Tableau and how to interpret the relevant statistics around them. Thank you for watching and we invite you to continue with the Free Training videos to learn more.