**Correlation in Tableau**

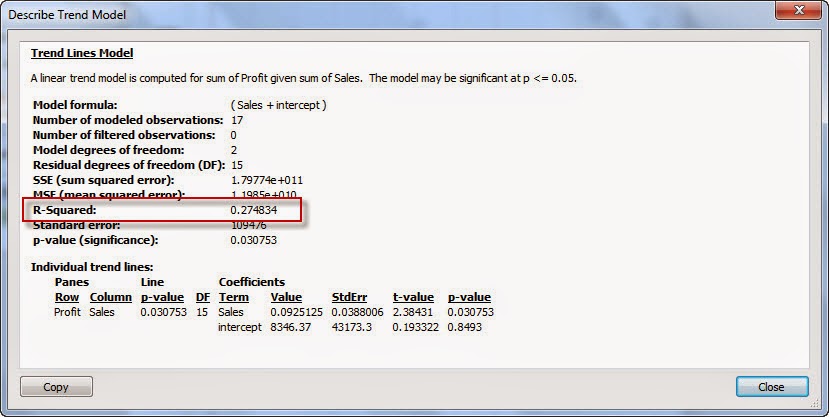
**What is a correlation coefficient?**

A correlation coefficient is a value that quantifies the relationship of two or more variables. In linear correlation the coefficient quantifies the strength and direction of the correlation between the variables.

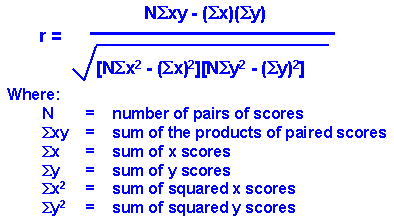
One type of correlation coefficient is the Pearson product-moment correlation coefficient, also known as r, which measures linear correlation and provides a value between -1 and +1.

* 1 is total positive correlation
* 0 is no correlation
* −1 is total negative correlation

I will start this post out by saying I am by no means a statistics expert.  I took a lot of math courses while getting my physics degree, but statistics was never my forte.  However, I know enough to be able to translate mathematical formulas into Tableau's calculated field syntax.  
  
Before getting into the nitty gritty of the math, I just want to point out that Tableau will automatically compute the r-squared value if you add a trend line to any scatterplot (right-click the trend line and choose 'Describe Trend Model'):

[](http://3.bp.blogspot.com/-g702TiWYPqM/VCwzw7tTNSI/AAAAAAAAb3g/MdAT-5P7t_A/s1600/10-1-2014%2B9-44-29%2BAM%2Br%2Bsqu.jpg)

While you can copy/paste this into an annotation or some other calculation, it wont dynamically update when your data updates.  Therefore if we wanted to compute the correlation so we can use it in other places in Tableau we must compute it in Tableau's calculated fields.  
  
First let's start with the formula for correlation ('r'):

[](http://1.bp.blogspot.com/--yEe3epBEIk/VCw1IBCtd7I/AAAAAAAAb3o/IGLdxh1NoPA/s1600/corrform1.gif)

Now let's map those to Tableau calculated field functions:

N = size()

Sigma XY = window\_sum(sum(X)\*sum(Y))

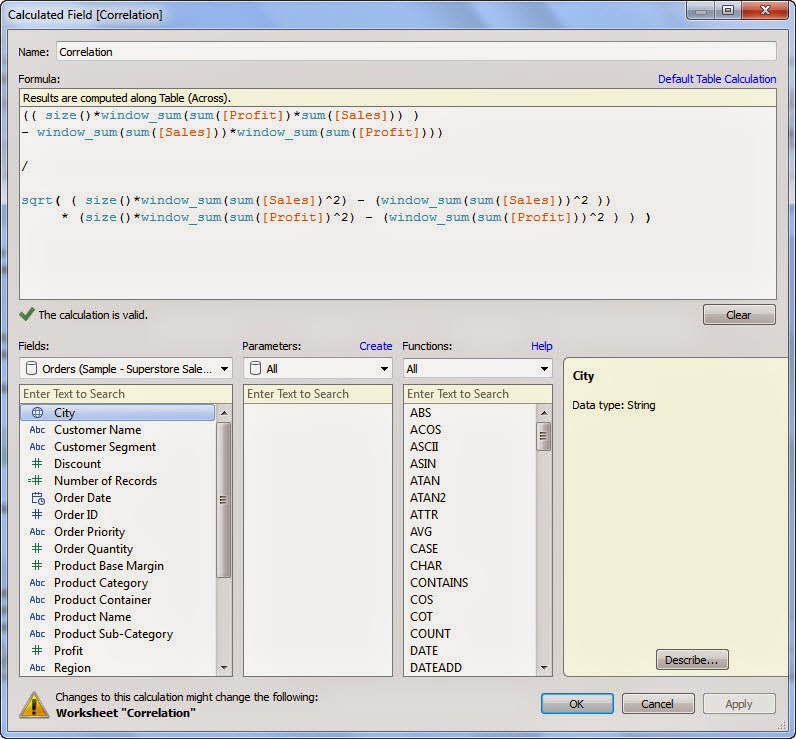
Sigma X = window\_sum(sum(X))

Sigma Y = window\_sum(sum(Y))

Sigma X^2 = window\_sum(sum(X)^2)

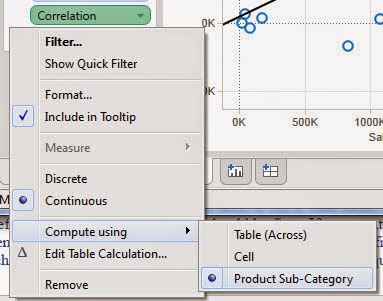
Sigma Y^2 = window\_sum(sum(Y)^2)

Putting it all together results in a pretty gnarly calc, but it's do-able:

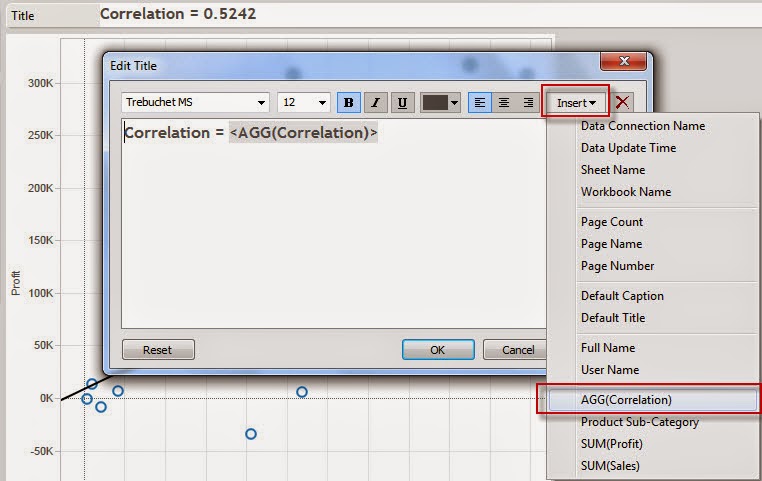
[](http://4.bp.blogspot.com/-bhBUxmTn_Rg/VCw1vn0Z71I/AAAAAAAAb3w/OoFlD8aCSRg/s1600/10-1-2014%2B10-10-58%2BAM.jpg)

Once you've written your calculation using whatever your X and Y fields represent then drag your new field onto the Detail shelf to include it in the view.

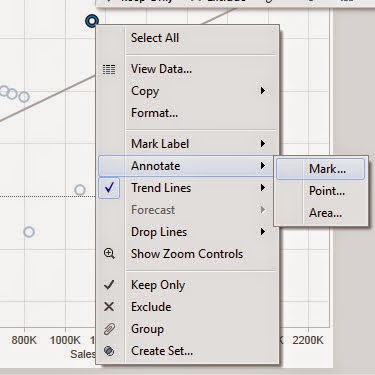
Note:  since this uses table calculation functions, we must tell Tableau what to compute along. After adding the field to your view, right-click it and choose Compute Using->[dimension that makes each point unique in your scatter].  You could probably also leave it on the default 'Table (Across)' but in case that doesn't work, select the specific dimension that is making all the points in your scatter:

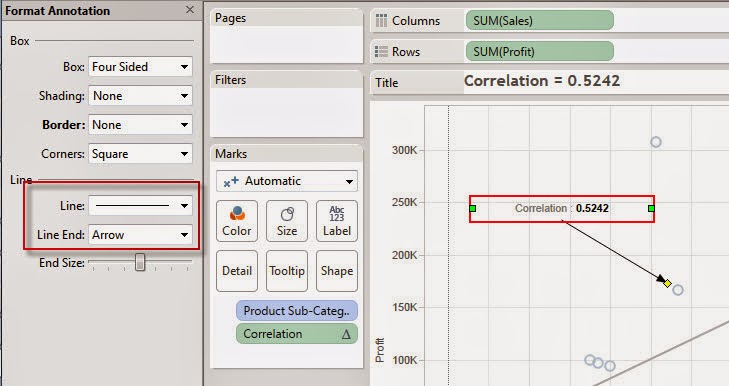
[](http://1.bp.blogspot.com/-isJlwAmP_BI/VCw3Vk14OwI/AAAAAAAAb34/jsgBuO1S-uA/s1600/10-1-2014%2B10-15-38%2BAM.jpg)

Now, where to display it?  While it wont let you insert this field into Area or Point annotations, you can insert it into the Title or into a Mark annotation.  If you use the Title, just edit the title of the view and use the Insert button in the upper right to stick the Correlation field in there:

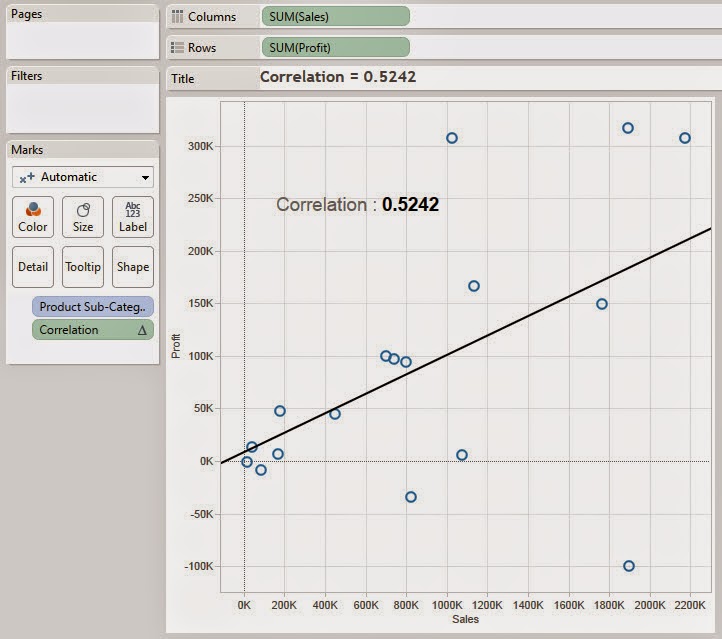
[](http://3.bp.blogspot.com/-97Mxm_4TZOw/VCw6XmTIKEI/AAAAAAAAb4g/2UgDrVokVCk/s1600/10-1-2014%2B10-30-35%2BAM.jpg)

If you use the mark annotation, just pick any of the marks in your scatter and choose 'Annotate->Mark'.  You can then delete all the other fields in there and just show the correlation.

[](http://3.bp.blogspot.com/-WstVtA9kBIM/VCw4se-wuiI/AAAAAAAAb4E/ptBUjw_72OM/s1600/10-1-2014%2B10-21-41%2BAM.jpg)  
  
  
  
  
  
  
  
  
  
  
  
  
Lastly, format it to remove the 'line' and 'line end' so it appears as its floating (like an Area annotation):

[](http://2.bp.blogspot.com/-twOb1HYXB7Q/VCw4sXUSXFI/AAAAAAAAb4I/JrwgBfe4SYQ/s1600/10-1-2014%2B10-22-36%2BAM.jpg)

Final product:

[](http://3.bp.blogspot.com/-QKtyI424nI0/VCw5V96YWMI/AAAAAAAAb4U/qAVDSXqrSV8/s1600/10-1-2014%2B10-26-19%2BAM.jpg)