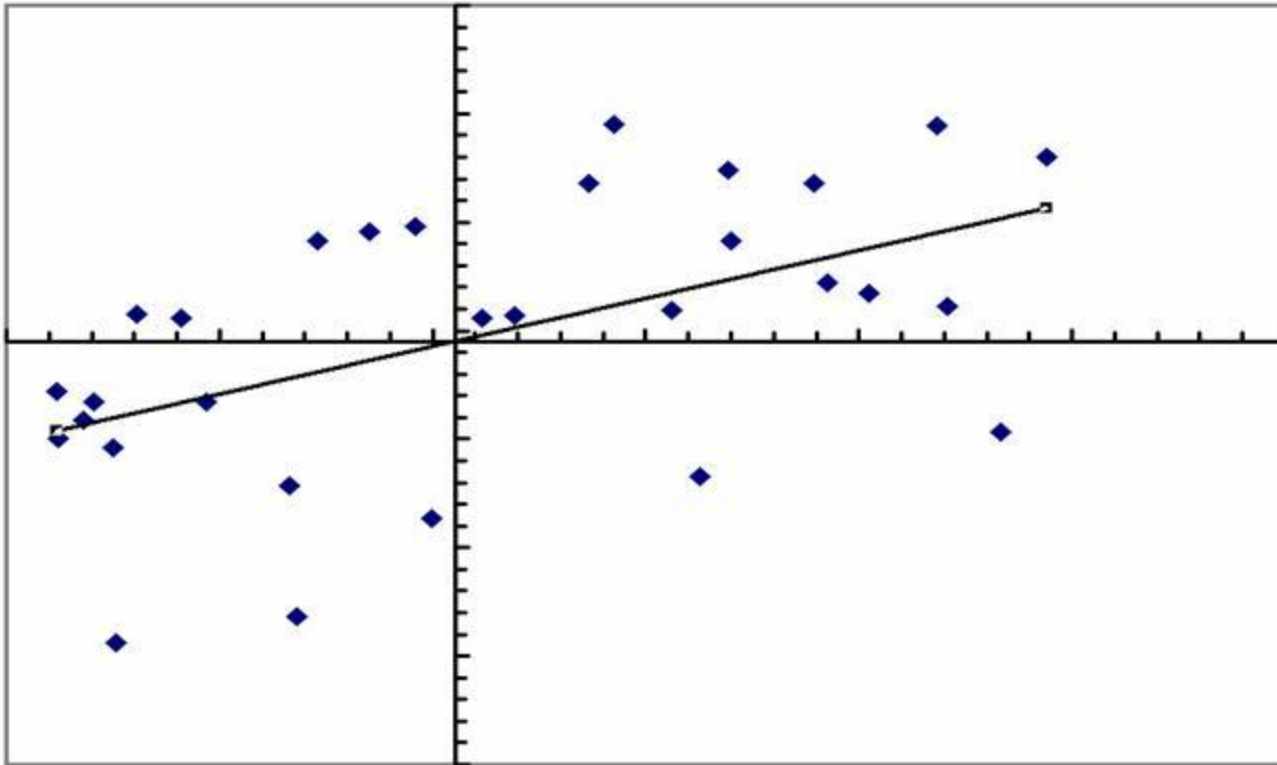


# Scatter Diagram

Paul Klein CSSBB, CMQ/OE, CQA  
Kleincox@yahoo.com

# Scatter Diagram



# Scatter Diagram Uses

- Scatter Diagram is used when you have 2 numerical factors that you desire to see if they are correlated with line or simple curve. The closer to a single line the higher the correlation.
- Data is collected in pairs and this x-y analysis does not take into account the aspect of time nor other parameters.
- Correlation is established before control charts are created for parameters not able to be measured directly.

# Collect data

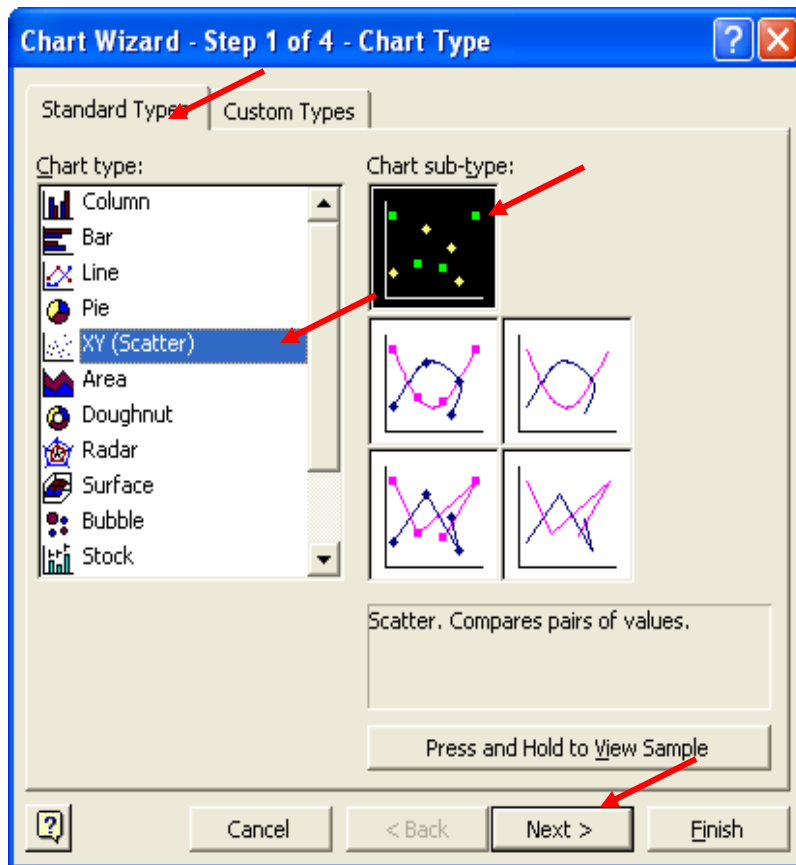
- No other process parameters should change during the test.
- Document the other process parameter setting so that the test can be replicated or modified under known conditions.

# Transfer data to Excel Spreadsheet

Independent	Dependent
48.34	8.9
82.64	8.25
545.79	4.61
271.91	7.36
188.27	2.59
873.75	3.38
398.72	0.97
446.42	5.1
99.62	6.19
808.8	5.22
976.63	7.41
384.45	9.29
770.52	5.5
476.5	9.91
163.59	9.21
122.93	5.21
292.43	1.93
265.61	8.17
72.95	6.8
341.21	2.46
679.78	3.46
883.11	3
102.81	9.19
624.83	7.13
47.4	7.3
650.06	5.4
570.3	1.86
933.36	7.84
677.24	4.58
757.78	4.65

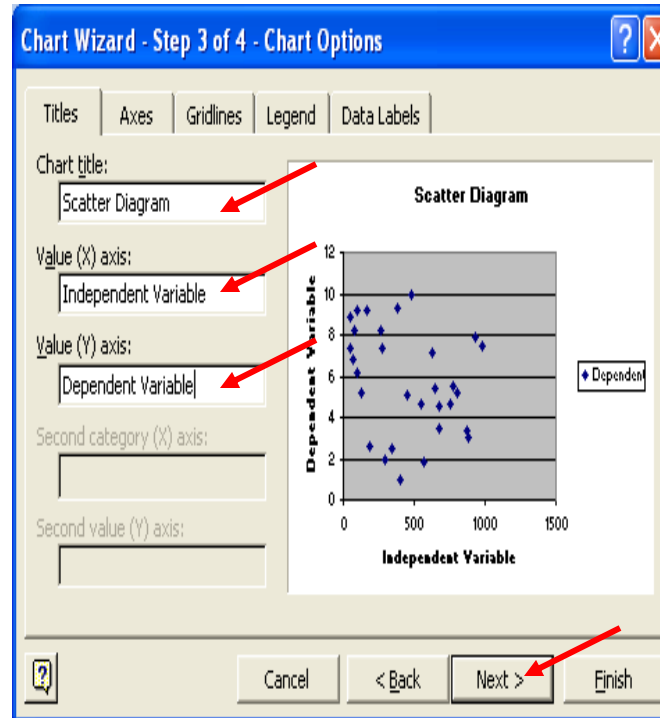
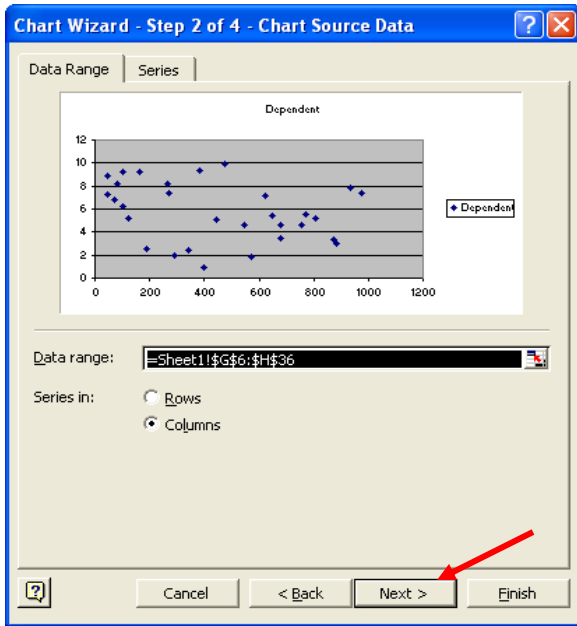
- Data will be in pairs of X (independent variable) and Y dependent variable)
- We are evaluating if I change the process setting “X” what results did we obtain for “Y”.
- No need to sort or prioritize data.
- The more data points the better. Plan on at least 30 data points if possible.

# Creating a chart



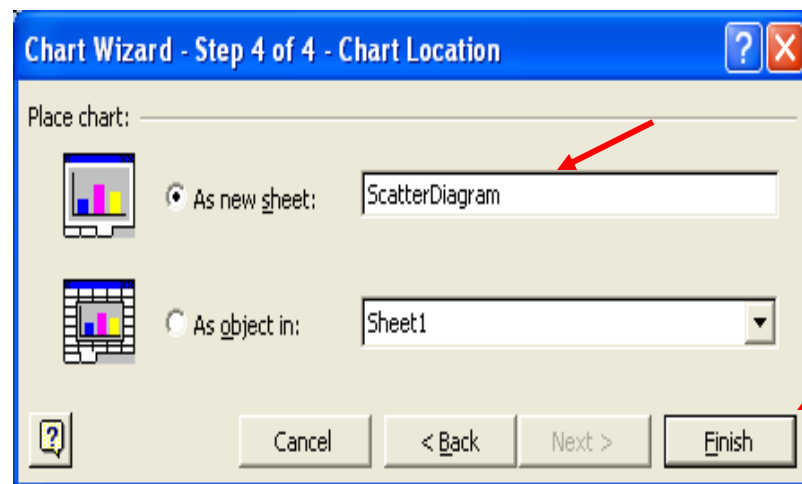
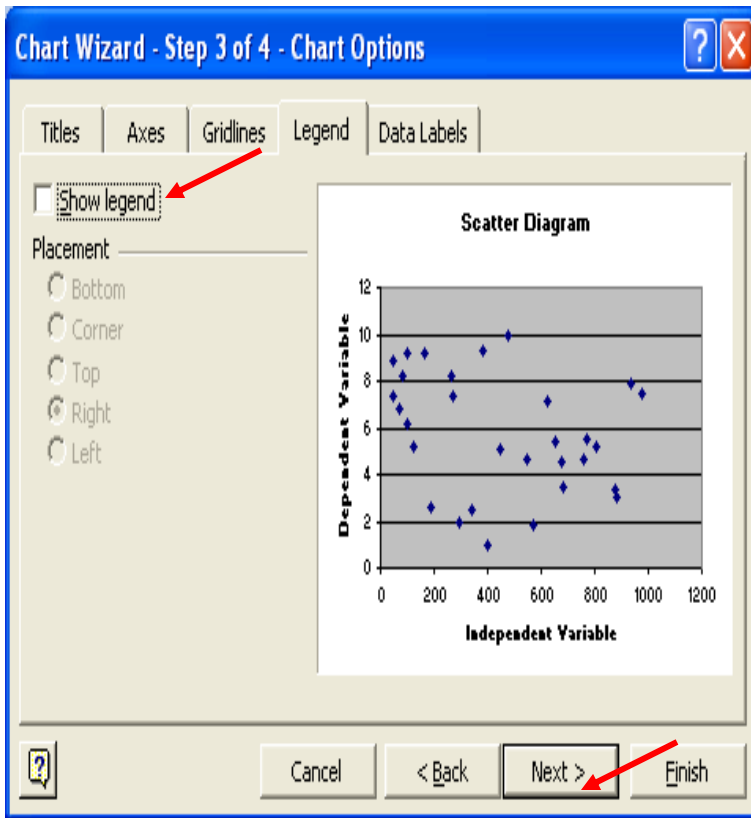
- Highlight data in columns A, and B
- Select chart icon. Choose “Standard Types” and select “XY (Scatter)”.

# Chart Creation steps

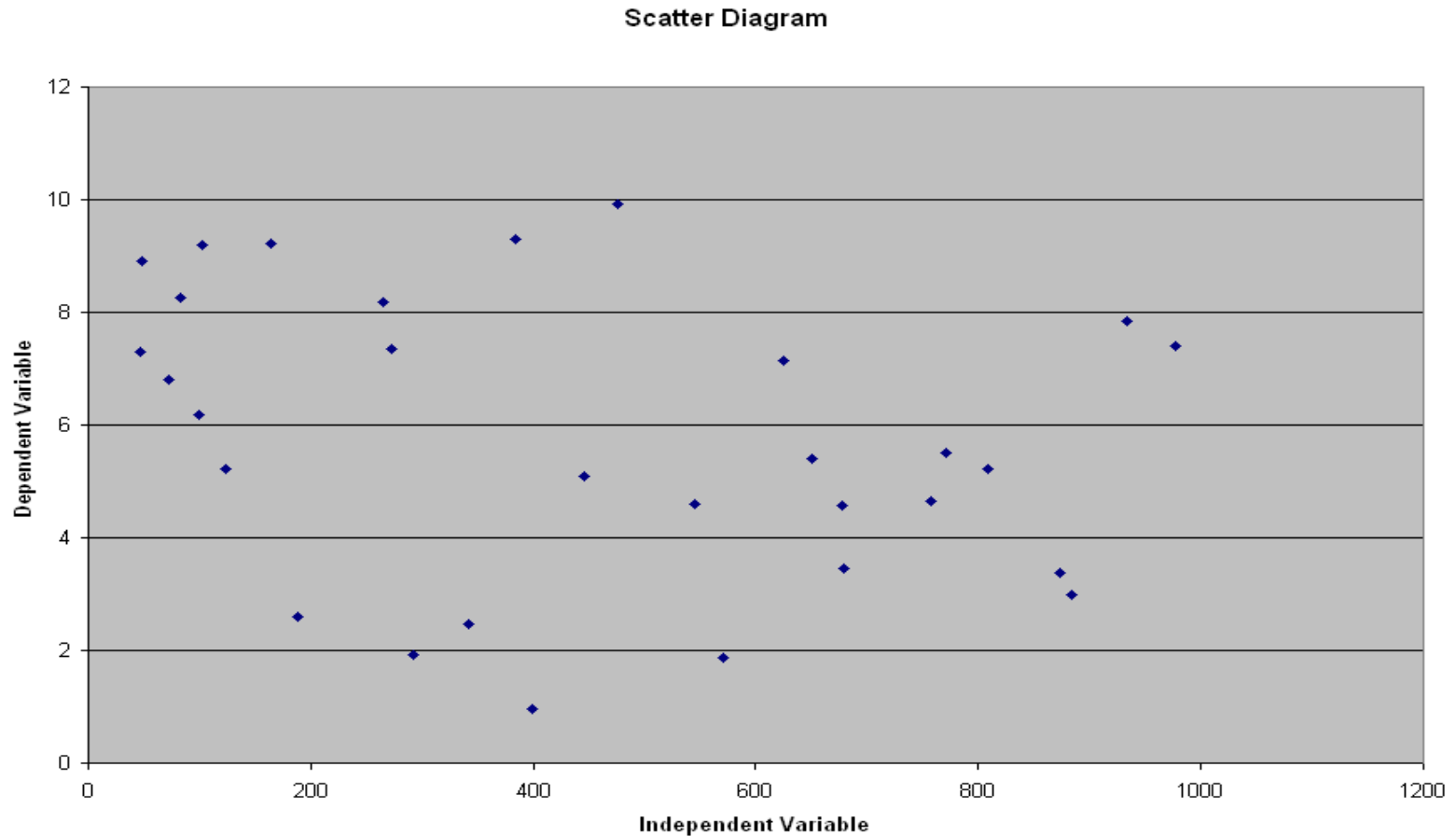


# Chart Creation steps

- Using Legend tab. De-select “Show Legend”.
- Select “As new sheet” and finish creating a chart.

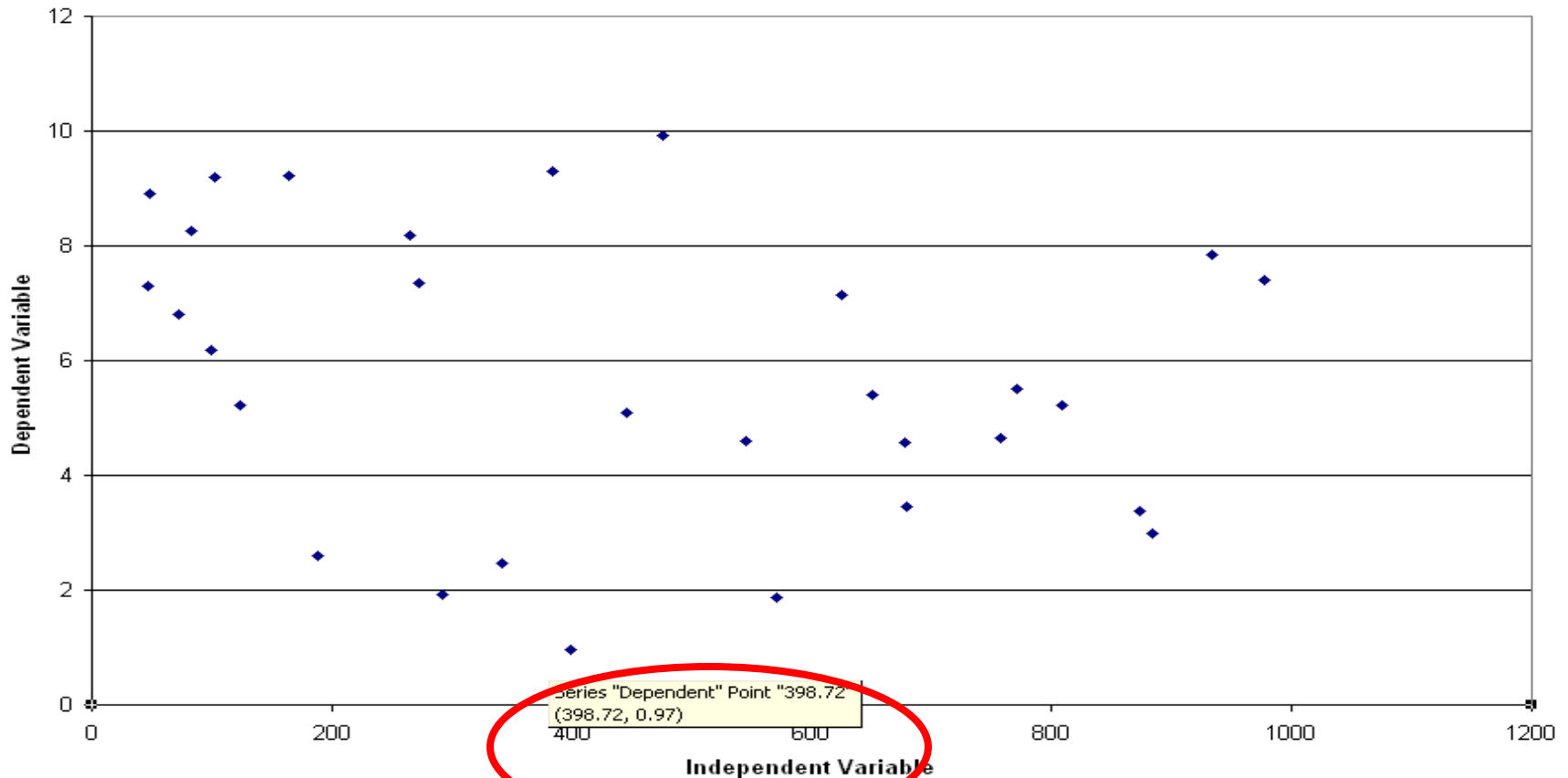


# Chart before formatting

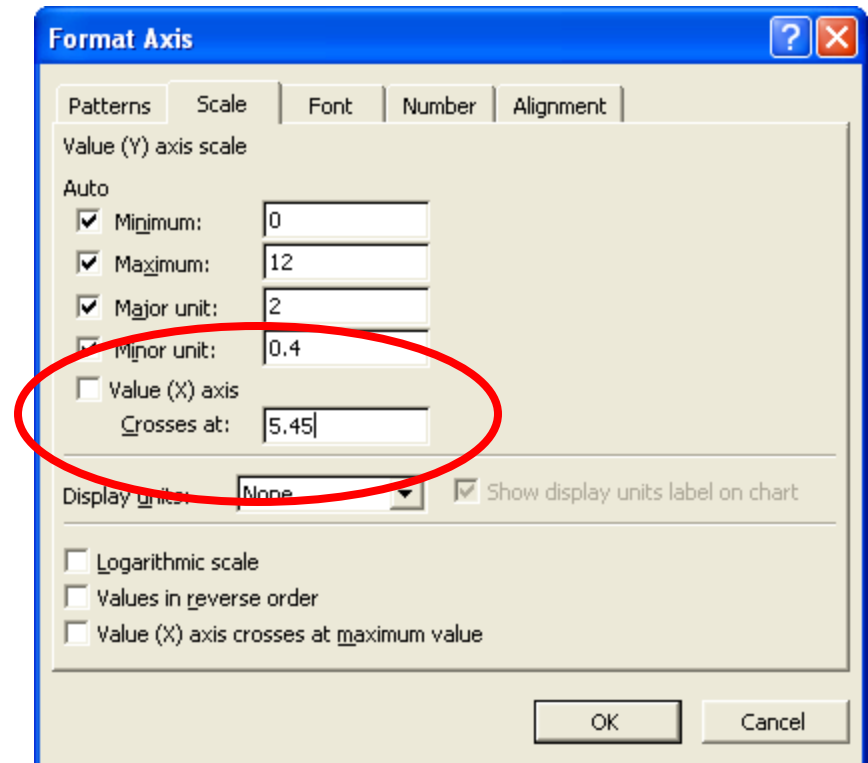
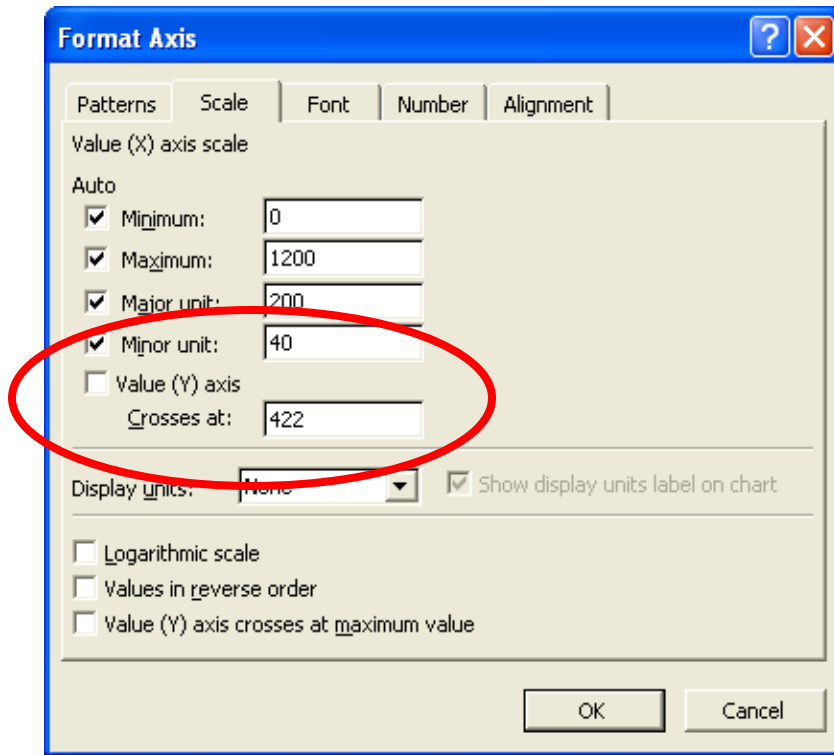


Count the number of points in the spreadsheet and find the median (equal number of points above and below) value. You can also place the cursor over a data point to find the x and y values of it.)

Scatter Diagram



Right click the axis and enter the values found in the cell noted below



Shut off grid lines and retain major axis lines  
(This needs to be done on X and Y axis)  
[can also right click gridlines and hit “clear”]

The image shows the 'Chart Options' dialog box in Microsoft Excel, specifically the 'Gridlines' tab. The dialog is divided into several sections: 'Titles', 'Axes', 'Gridlines', 'Legend', and 'Data Labels'. The 'Gridlines' tab is active, and the 'Major gridlines' checkboxes for both the 'Value (X) axis' and 'Value (Y) axis' are unchecked. Red arrows point to these checkboxes. The 'Minor gridlines' checkboxes are also unchecked. A preview window on the right shows a scatter plot titled 'Scatter Diagram' with a regression line. The regression equation is  $y = 1.5434x + 446.46$  and the coefficient of determination is  $R^2 = 0.1845$ . The x-axis is labeled 'Independent Variable' and ranges from 0 to 1200. The y-axis is labeled 'Dependent Variable' and ranges from -1500 to 3500. The 'OK' and 'Cancel' buttons are visible at the bottom right of the dialog.

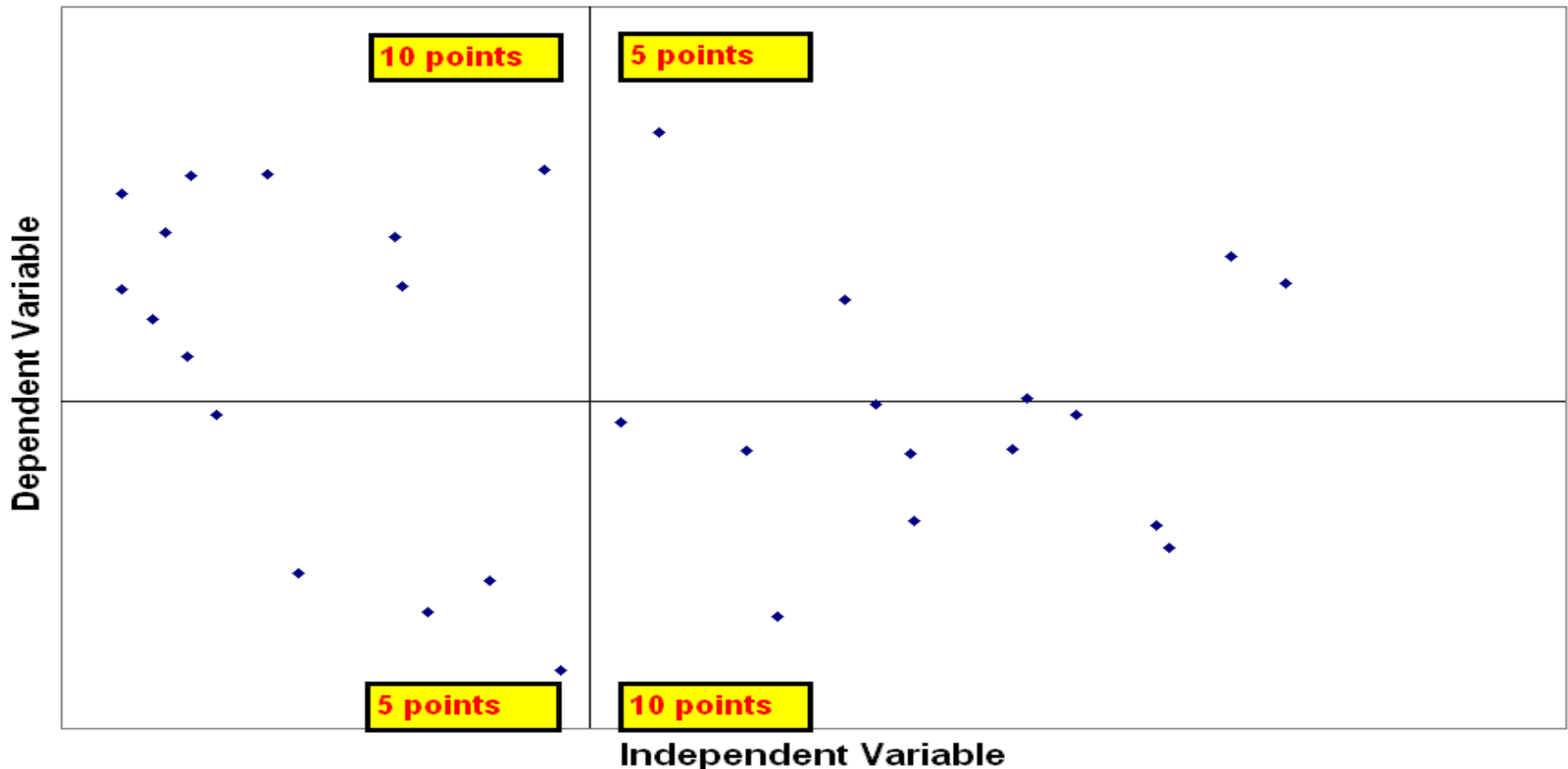
# Format adjustments

- Right click the background of the chart and select “format plot area” and click “None”.
- Click on titles and change them to be at least 16 font.
- Click on the X axis then right click and click alignment as shown.
- Change font on data labels and categories to be bold

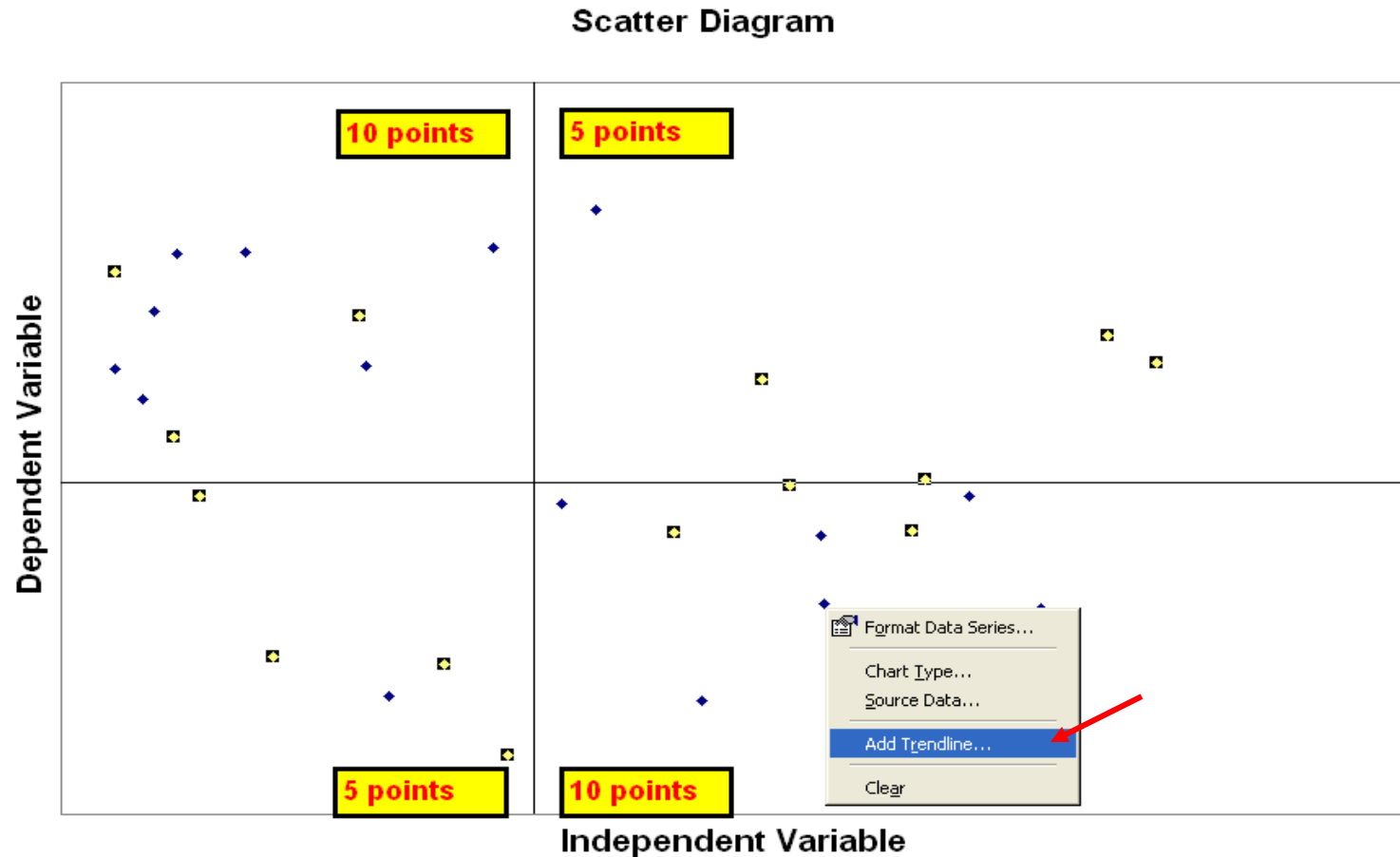
From the chart add the number of points in each sector. The ratio of the total points in each diagonal given a very rough estimation of how the variables are related.

Diagonal with maximum number of points divided by diagonal with minimum number of points or in this cast  $(10+10)/(5+5) = 2$  Values less than 1.5 have little or no correlation while values above 3 have correlation and values above 5 have significant correlation.

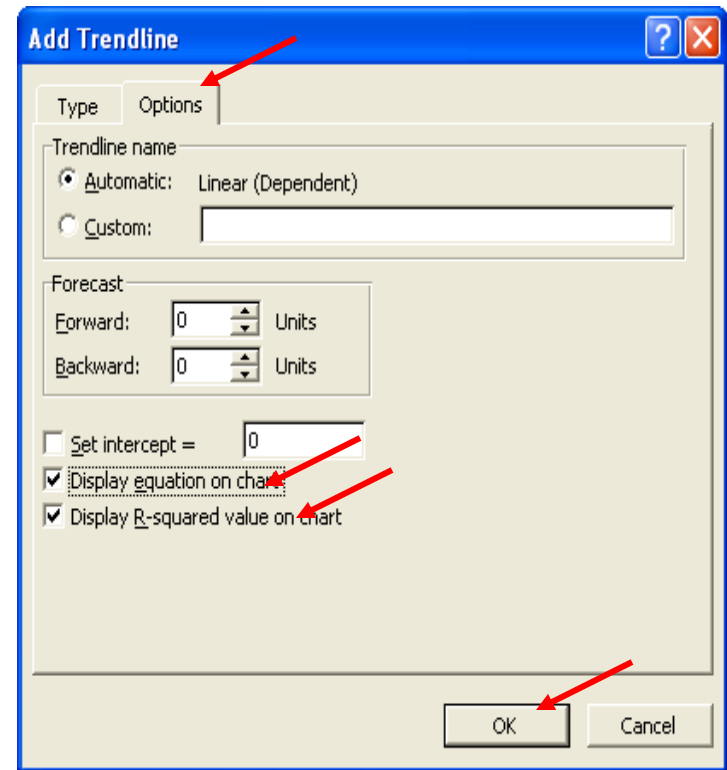
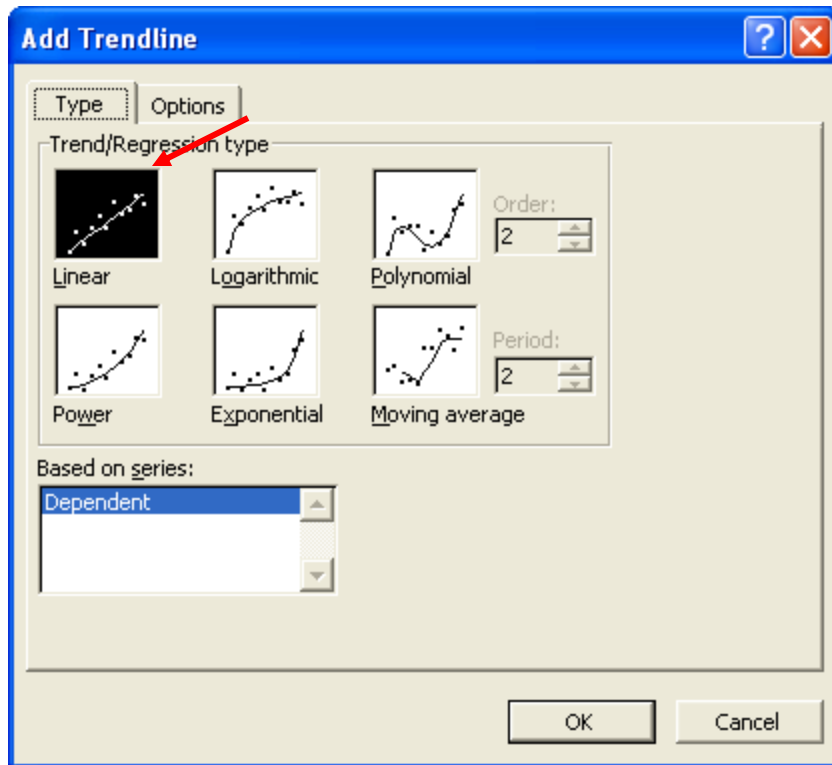
**Scatter Diagram**



If desired you can add a computer generated trendline by right clicking a datapoint and select “add trendline”.

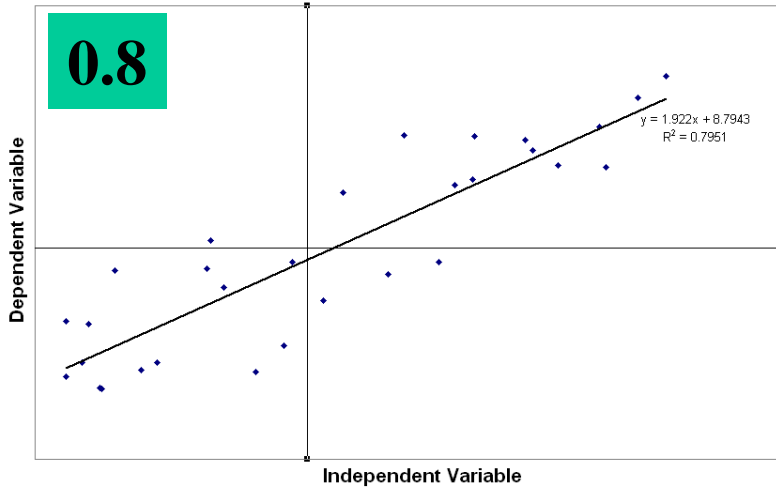


Select the trendline based on your knowledge of the process and how the data looks. You can see differing correlation values (R value) as well as defining the estimated formula defining the correlation. R value of 1 is perfectly correlated, and correlations less than 0.3 are very weak and other factors should be studied. Note that polynomial has options of higher levels but typically level 2 and 3 are the most common for polynomials.

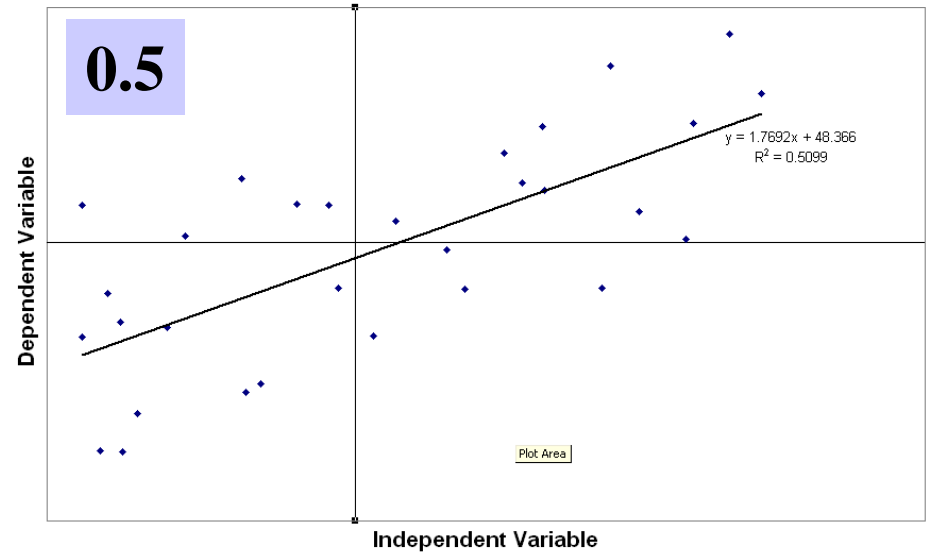


# R values examples

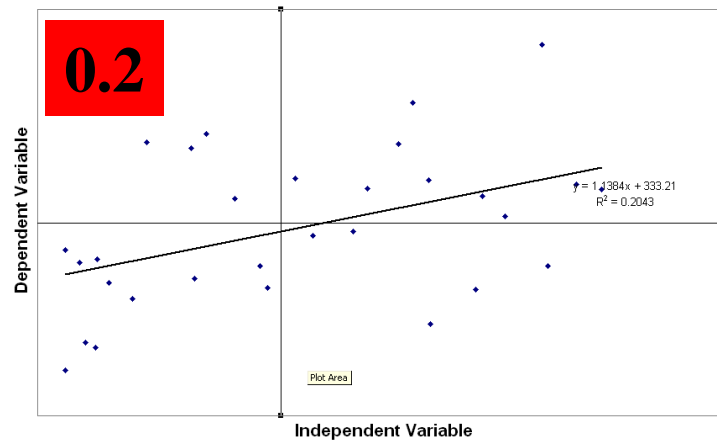
Scatter Diagram



Scatter Diagram



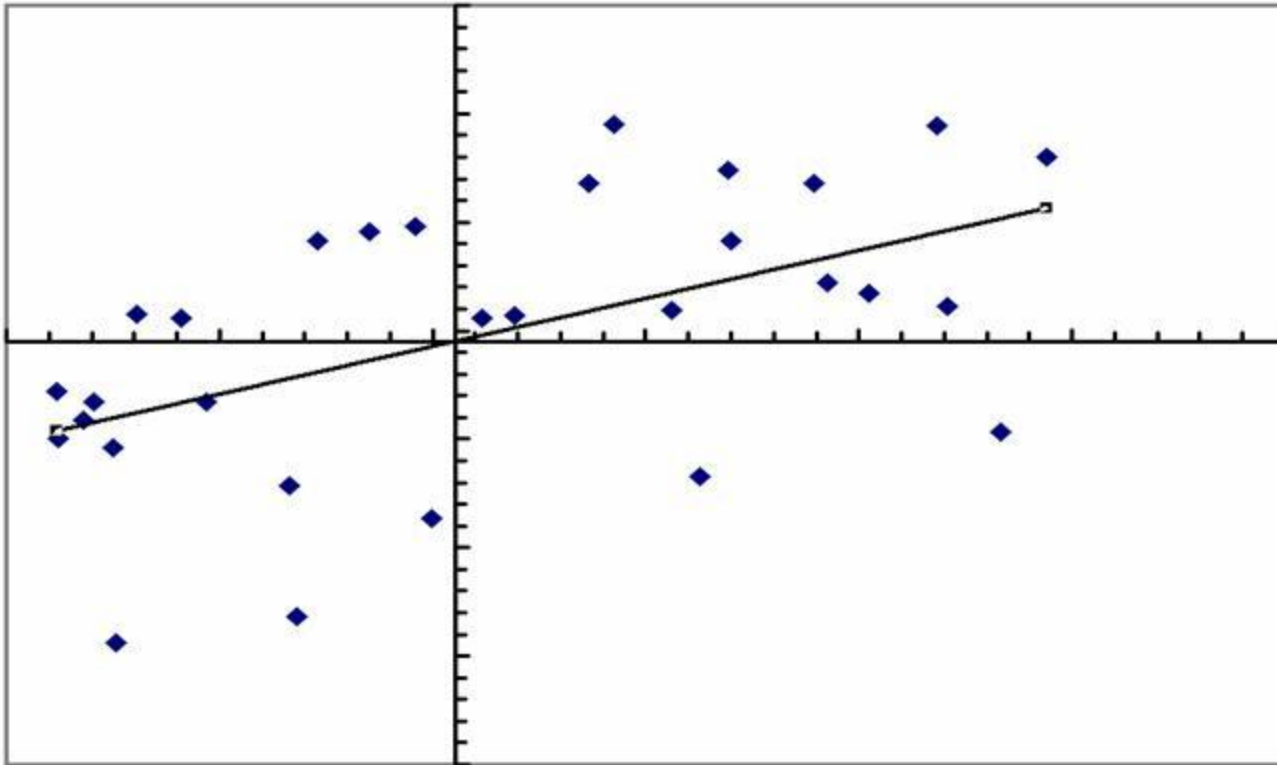
Scatter Diagram



Note that having no correlation is not bad data. Sometimes it allows you to change a parameter and know that it can optimize some other aspect without impacting the dependent factor.

An example of this occurred at when traditional adjustment of viscosity was through pH changing the zeta potential of a fluid. The pH adjustment also degraded the binder that was being used. So to control the spray drying conditions we had to degrade the binder. A new chemical allowed us to de-link the pH from the viscosity adjustment. This allowed manufacturing to dial in each parameter independently.

# Scatter Diagram



# Thank You

Paul Klein

[kleincox@yahoo.com](mailto:kleincox@yahoo.com)