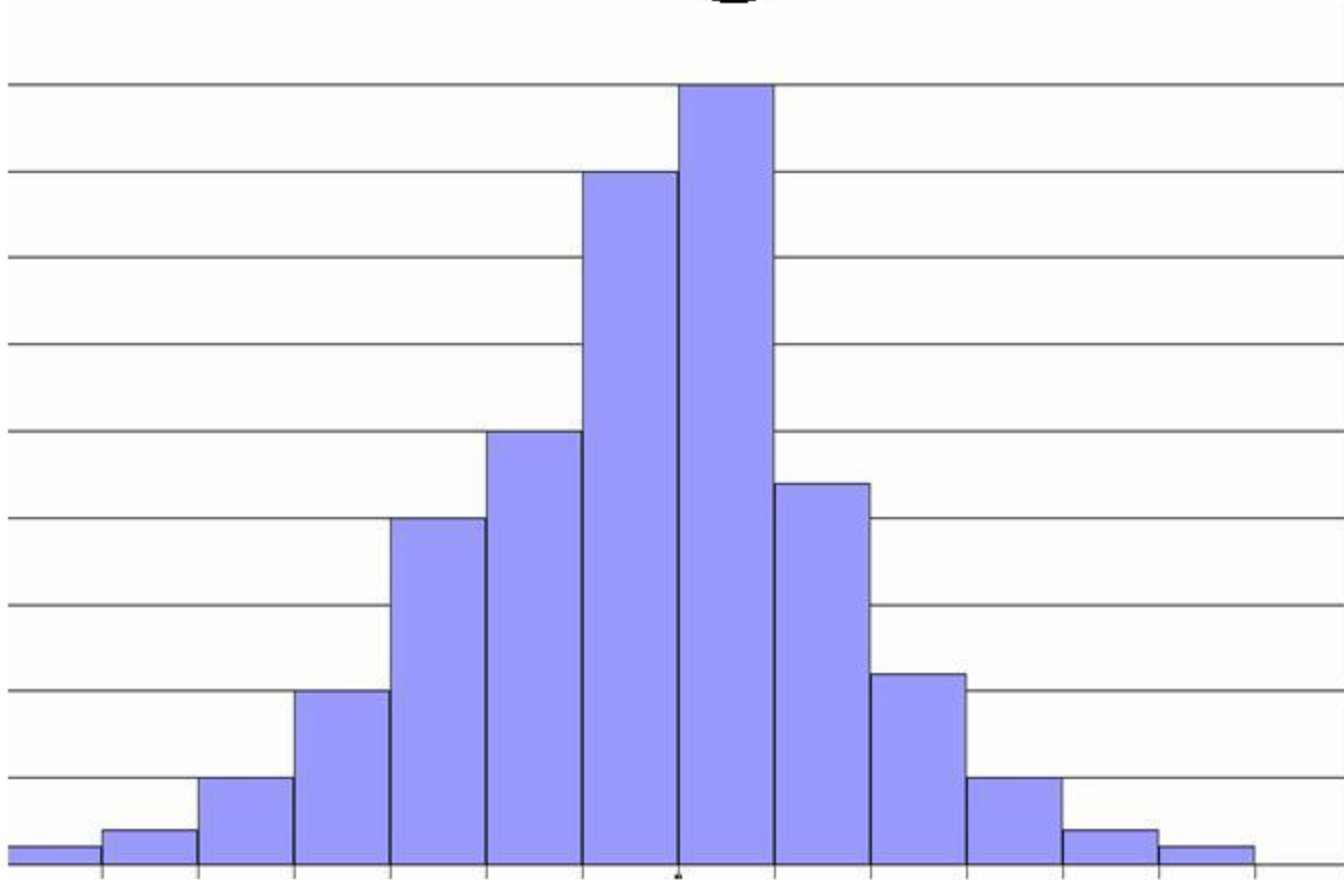


Histogram

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

Histogram



Histogram Uses

- Scatter Diagram is used when you have 1 numerical factor that you desire to see what shape/type of population that the data represents.
- Data is collected individually and this x only analysis does not take into account the aspect of time nor other parameters.
- Shapes of the distribution curve help understand the process characteristics. Examples are normal (gaussian), weibull, hyperbolic, bimodal, skewed, abnormal, poisson

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.
- Collect at least 50 data points but the more that are collected the better defined the shape will be. 300+ data points are not abnormal in this type of analysis especially in the age of automation.
- Be aware of autocorrelation errors. These are often caused by taking data so quick that the process is more dependent on where the last data point was than being an independent point.

Transfer data to Excel Spreadsheet

data	column #
627	9
118	2
680	10
135	2
53	1
283	4
543	8
44	1
93	2
418	6
379	6
258	4
902	13



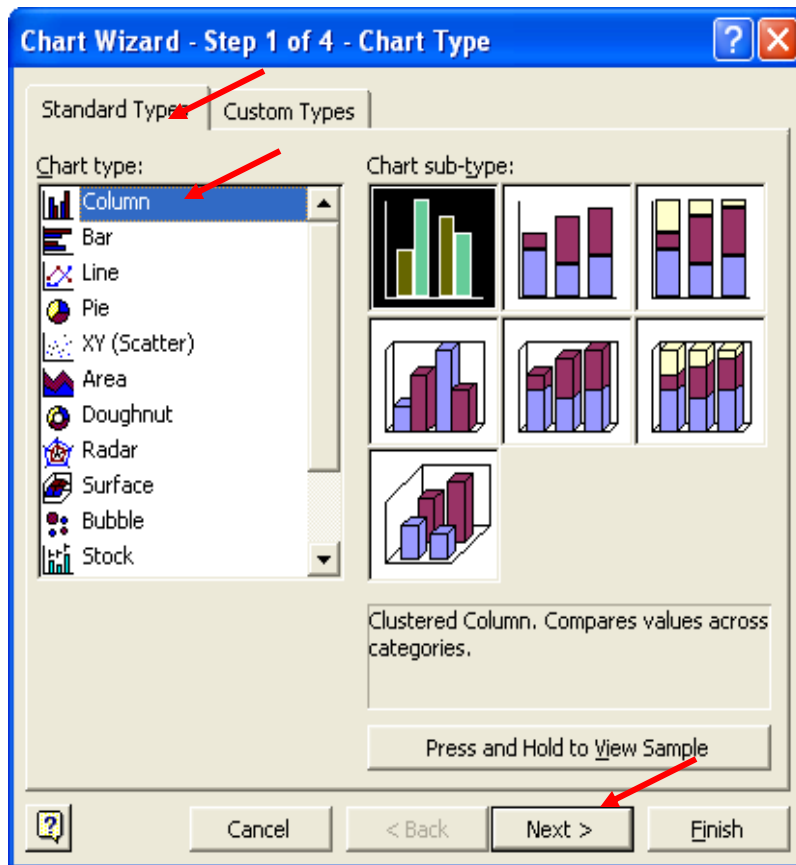
796	12
889	13
456	7
161	3
438	7

- Data will be individual X's
- We are evaluating how a parameter varies during the collection timeframe. For example temperature variation.
- No need to sort or prioritize data.
- The more data points the better. Plan on at least 50 data points preferably several hundred.

Histogram

- Now some basic data evaluations.
 - Find the **minimum** and **maximum values** and the **total number of data points**.
 - For distributions of 50 the guideline is to use 7 columns.
 - For distributions of 100 the guideline is to use 10 columns.
 - For distributions of 300+ the guideline is to use 14 columns.
 - Take the **maximum value** minus the **minimum value** and divided this by the **number of columns**. This will be the range for each column.
 - Tally the number of points in each group. The next several pages will show how to automatically do this if you have not added the histogram add-in on excel.

Creating a chart



- Highlight data in columns A, and B
- Select chart icon. Choose “Standard Types” and select “Column”.

Excel formula steps

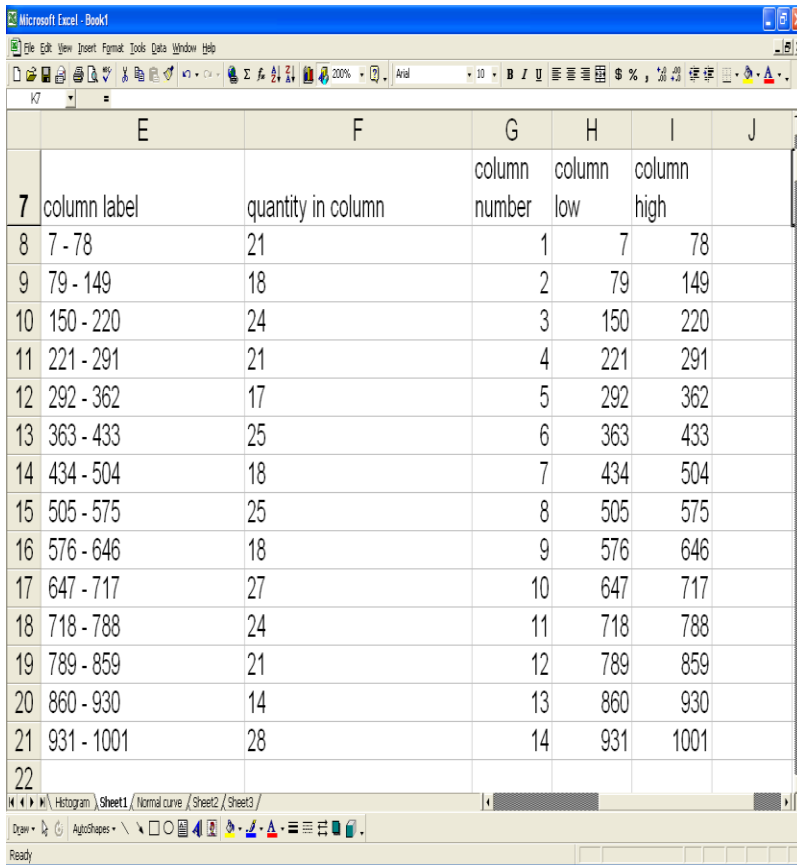
The collected data has been placed in cells B4 through B304
Below you can see the formulas placed into G2 through G6

	D	E	F	G
1				
2		maximum value:	=MAX(\$B\$4:\$B\$304)	995
3		minumum value:	=MIN(\$B\$4:\$B\$304)	7
4		number of data points:	=counta(\$B\$4:\$B\$304)	301
5		number of bars	=VLOOKUP(F4,columns,2)	14
6		Range per bar		71

Columns	
50	7
100	10
300	14

Excel formula steps

On the left you can see the cells e7 through I21 as values and on the next page there are examples of formulas used in each column.



	E	F	G	H	I	J
7	column label	quantity in column	column number	column low	column high	
8	7 - 78	21	1	7	78	
9	79 - 149	18	2	79	149	
10	150 - 220	24	3	150	220	
11	221 - 291	21	4	221	291	
12	292 - 362	17	5	292	362	
13	363 - 433	25	6	363	433	
14	434 - 504	18	7	434	504	
15	505 - 575	25	8	505	575	
16	576 - 646	18	9	576	646	
17	647 - 717	27	10	647	717	
18	718 - 788	24	11	718	788	
19	789 - 859	21	12	789	859	
20	860 - 930	14	13	860	930	
21	931 - 1001	28	14	931	1001	
22						

Excel formula steps

Bear with me, I know the formulas are daunting, but I can help. If you need help write or call me.

column label	quantity in column	column number	column low	column high
= " &H8& " - "&I8	=COUNTIF(\$C\$4:\$C\$304,G8)	1	=I8-\$G\$6	=\$G\$3+G8*\$G\$6
= " &H9& " - "&I9	=COUNTIF(\$C\$4:\$C\$304,G9)	2	=I8+1	=\$G\$3+G9*\$G\$6
= " &H10& " - "&I10	=COUNTIF(\$C\$4:\$C\$304,G10)	3	=I9+1	=\$G\$3+G10*\$G\$6
= " &H11& " - "&I11	=COUNTIF(\$C\$4:\$C\$304,G11)	4	=I10+1	=\$G\$3+G11*\$G\$6
= " &H12& " - "&I12	=COUNTIF(\$C\$4:\$C\$304,G12)	5	=I11+1	=\$G\$3+G12*\$G\$6
= " &H13& " - "&I13	=COUNTIF(\$C\$4:\$C\$304,G13)	6	=I12+1	=\$G\$3+G13*\$G\$6
= " &H14& " - "&I14	=COUNTIF(\$C\$4:\$C\$304,G14)	7	=I13+1	=\$G\$3+G14*\$G\$6
= " &H15& " - "&I15	=COUNTIF(\$C\$4:\$C\$304,G15)	8	=I14+1	=\$G\$3+G15*\$G\$6
= " &H16& " - "&I16	=COUNTIF(\$C\$4:\$C\$304,G16)	9	=I15+1	=\$G\$3+G16*\$G\$6
= " &H17& " - "&I17	=COUNTIF(\$C\$4:\$C\$304,G17)	10	=I16+1	=\$G\$3+G17*\$G\$6
= " &H18& " - "&I18	=COUNTIF(\$C\$4:\$C\$304,G18)	11	=I17+1	=\$G\$3+G18*\$G\$6
= " &H19& " - "&I19	=COUNTIF(\$C\$4:\$C\$304,G19)	12	=I18+1	=\$G\$3+G19*\$G\$6
= " &H20& " - "&I20	=COUNTIF(\$C\$4:\$C\$304,G20)	13	=I19+1	=\$G\$3+G20*\$G\$6
= " &H21& " - "&I21	=COUNTIF(\$C\$4:\$C\$304,G21)	14	=I20+1	=\$G\$3+G21*\$G\$6

Excel formula steps

Here is the formula for calculating the column number.

Basically it is the (max-min)/range per column plus 1.

This formula needs to be filled in for column C.

In my example the columns are assumed to be integers. This can be adjusted easily if the columns are not integers.

If you need help write or call me.

	B	C
3	data	column #
4	627	9
5	118	=INT((B5-\$G\$3)/\$G\$6)+1
6	680	10

Chart Creation steps

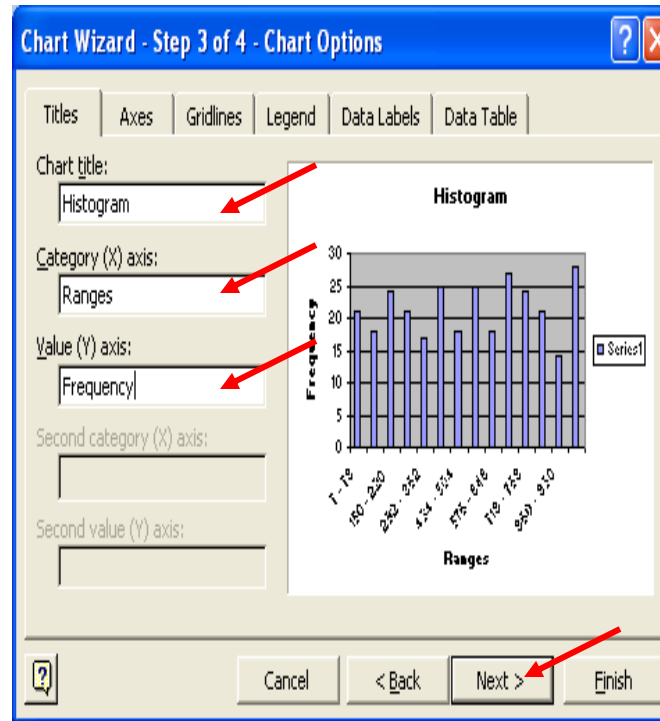
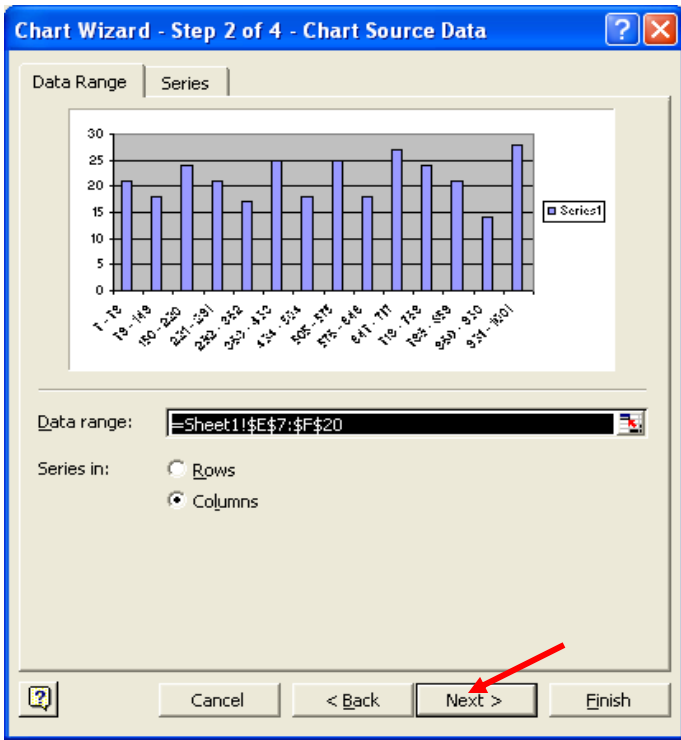


Chart Creation steps

- Using Legend tab. De-select “Show Legend”.
- Select data labels and click “show value”.

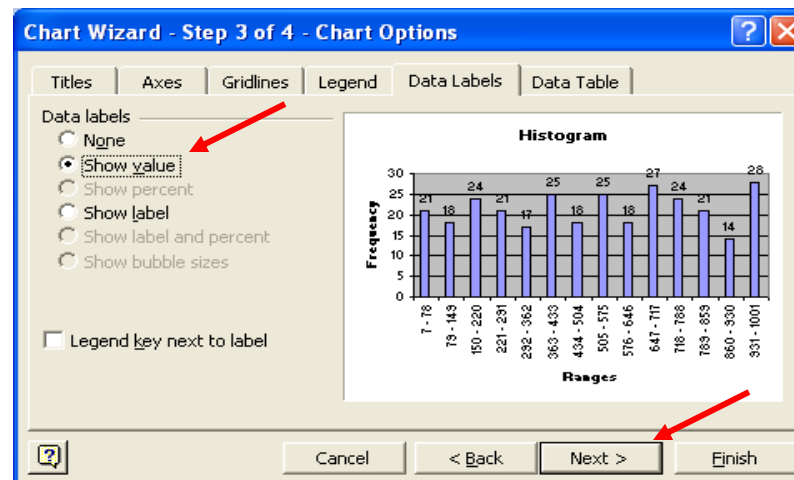
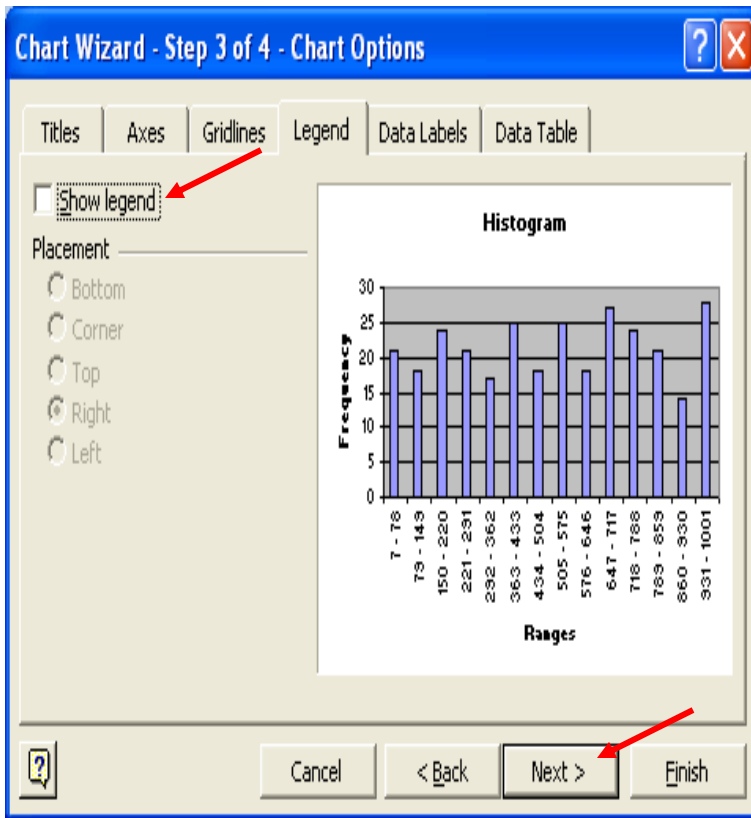
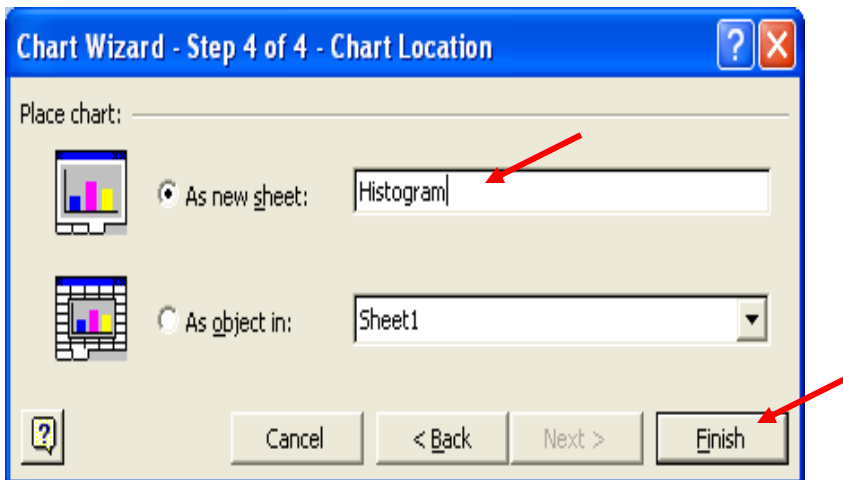
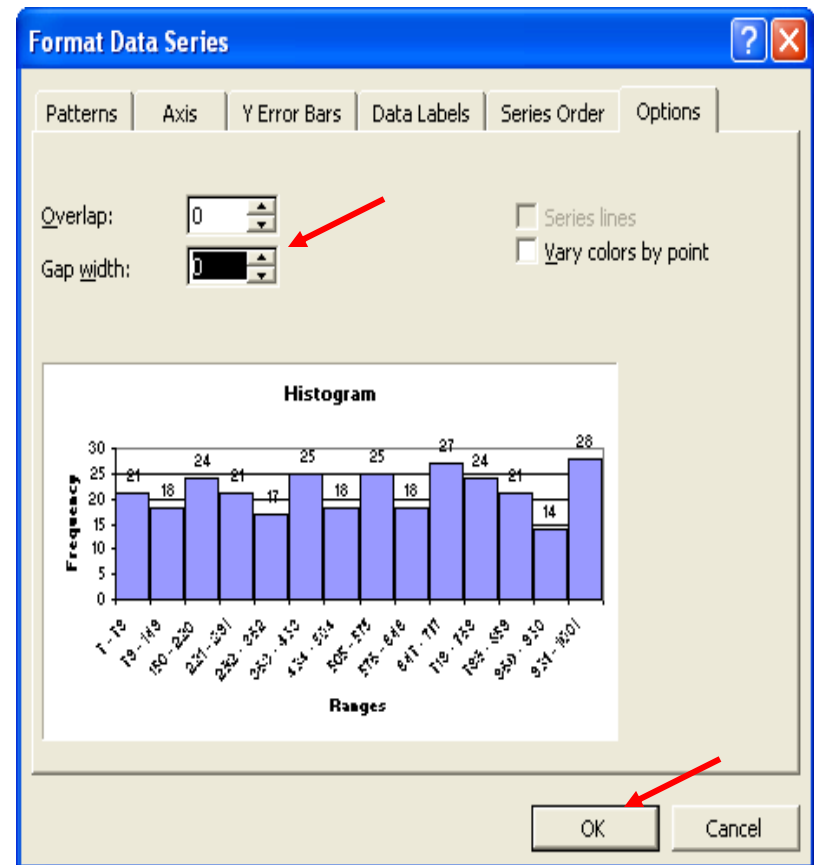
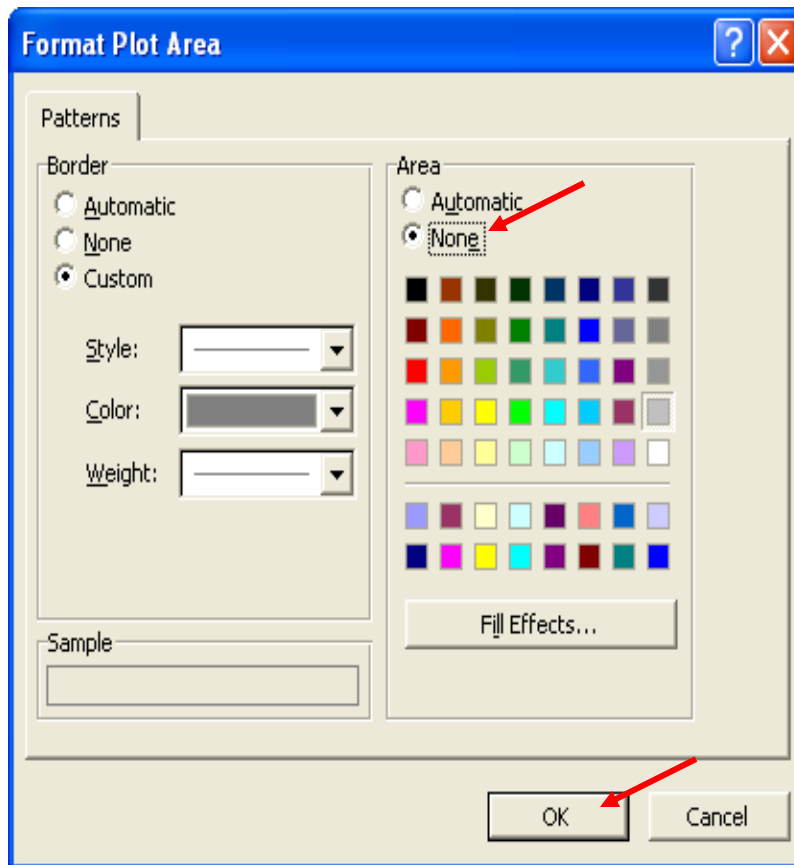


Chart Creation steps

- Select “As new sheet” and finish creating a chart.



Formatting continued

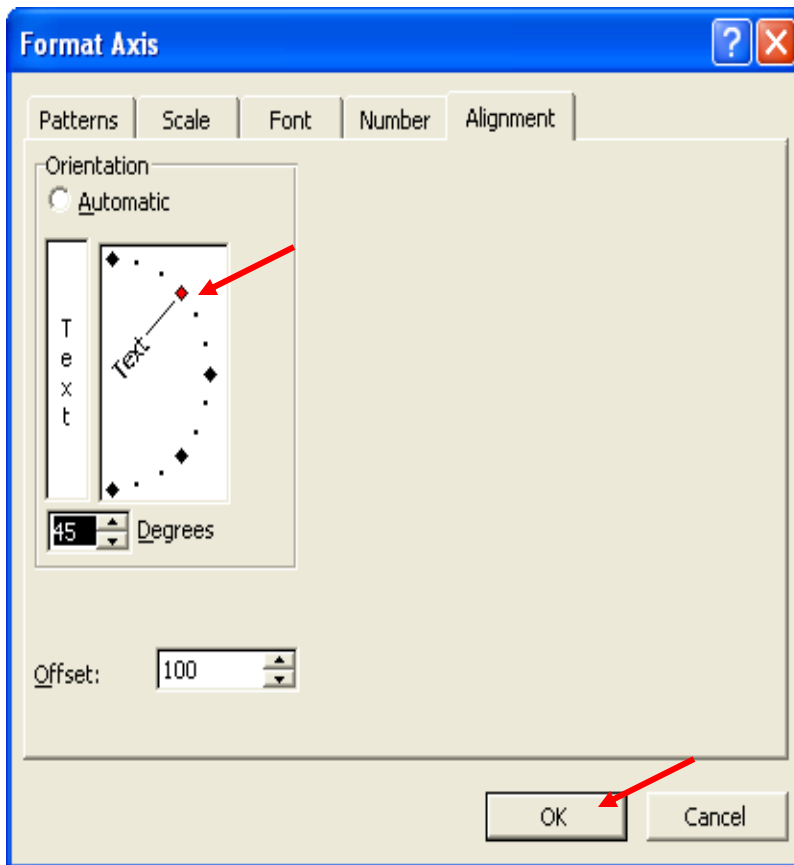


Format adjustments

- 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

Formatting continued

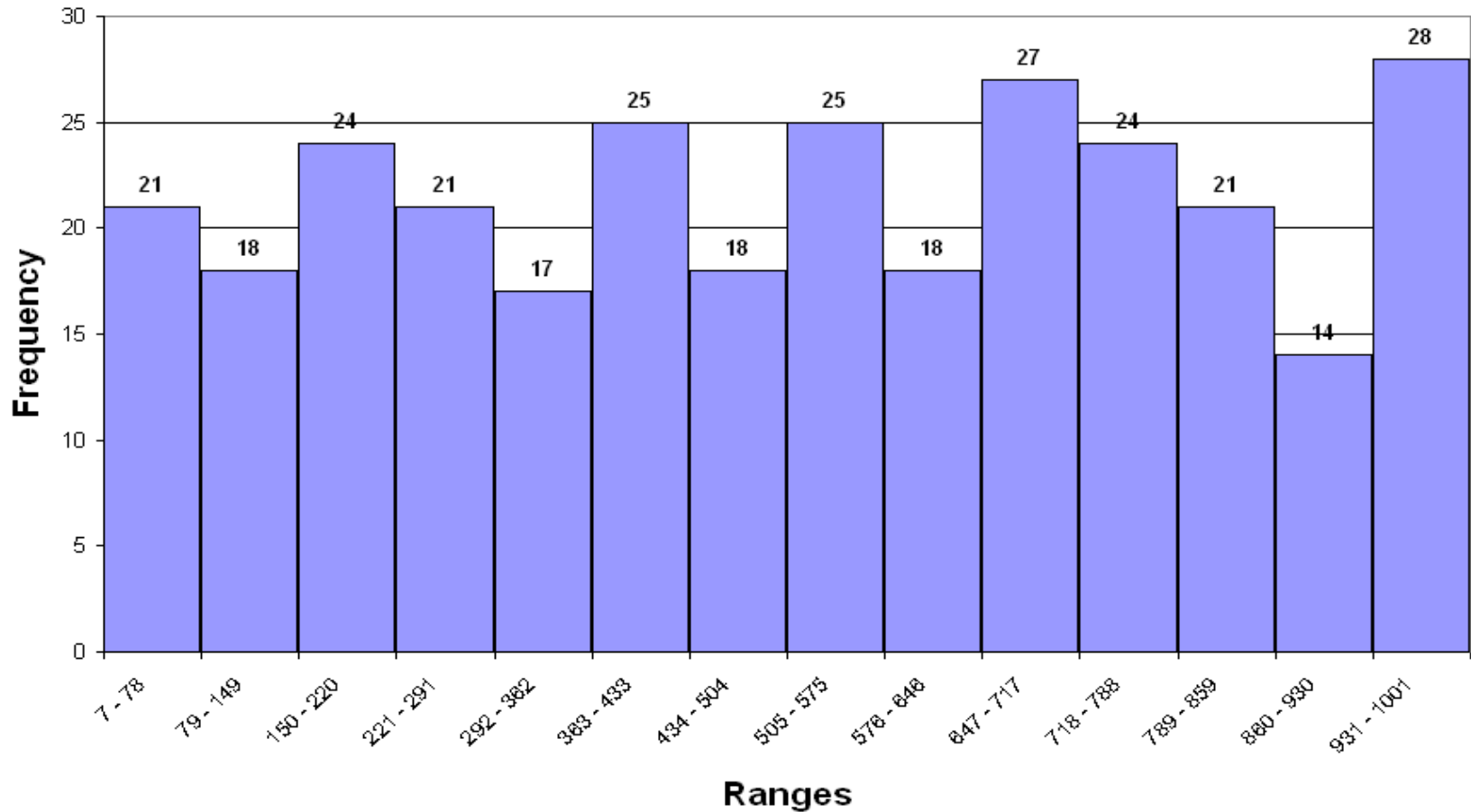
X-axis



Histogram example: uniform distribution

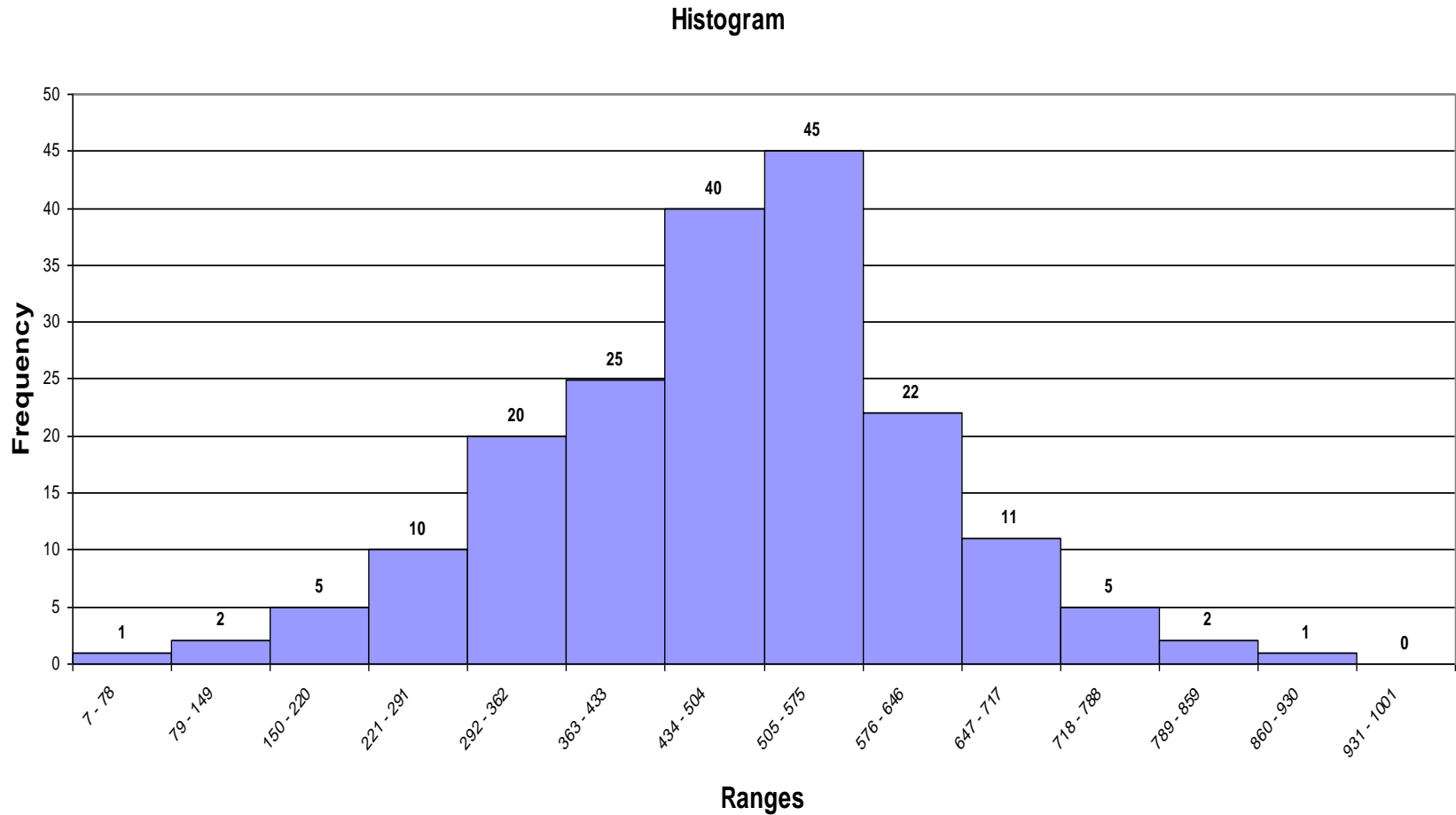
[lottery chance of pulling one specific number]

Histogram



Histogram example: Normal (Gaussian)

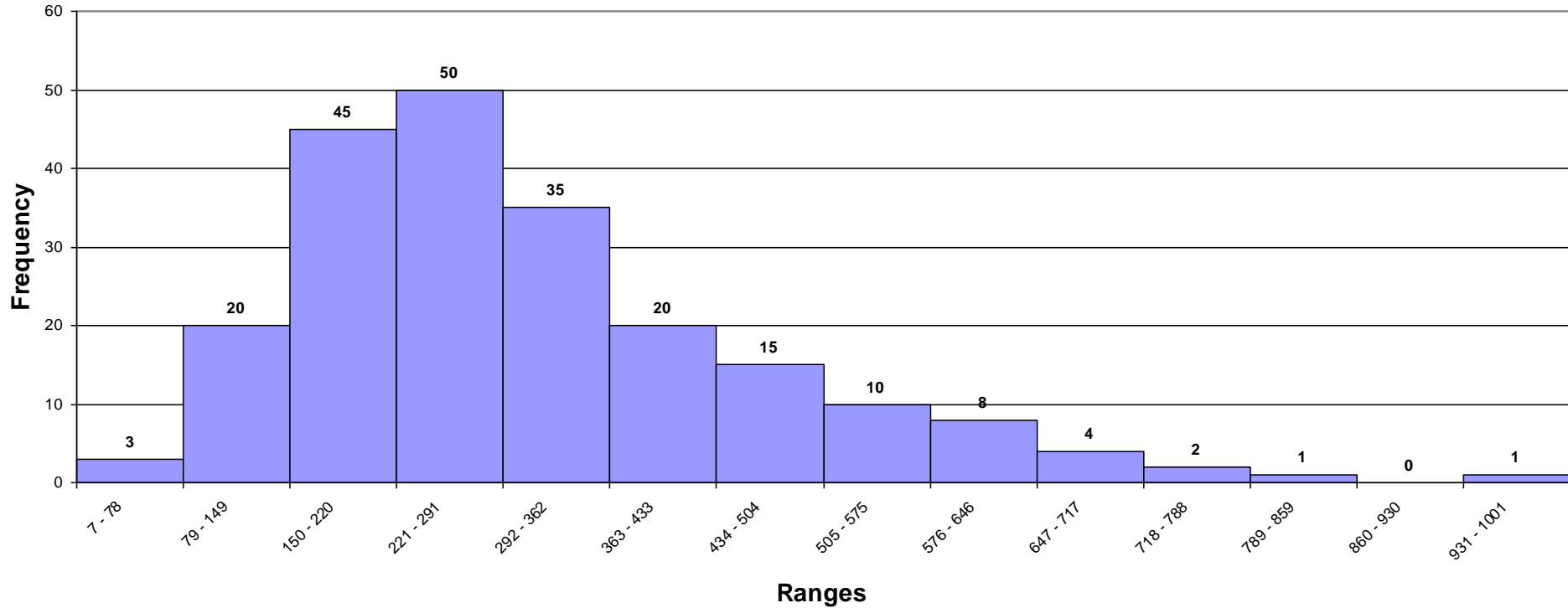
[density of pressed parts when process in in control]



Histogram example: Weibull

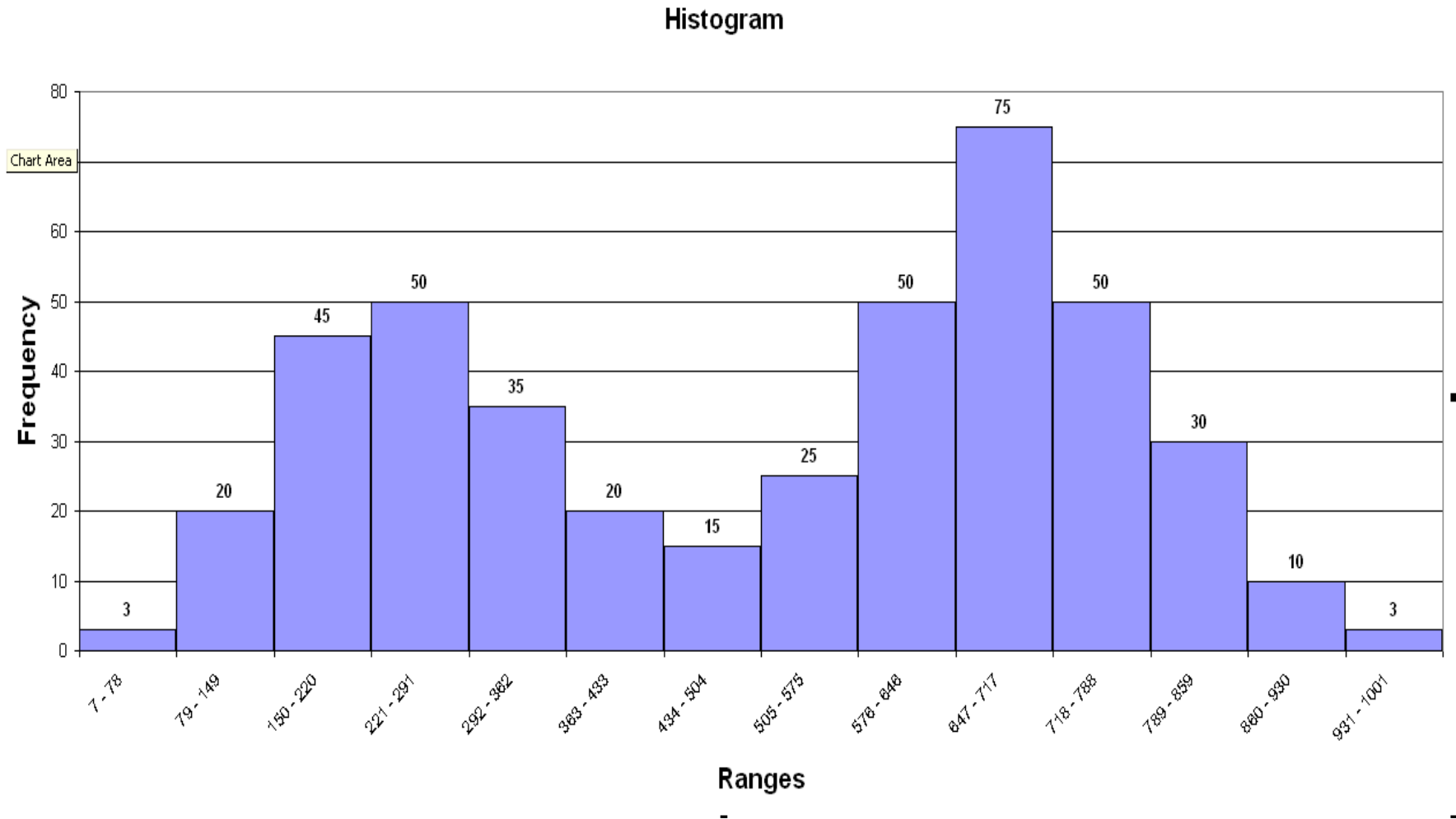
[lot arrival times in kiln area]

Histogram



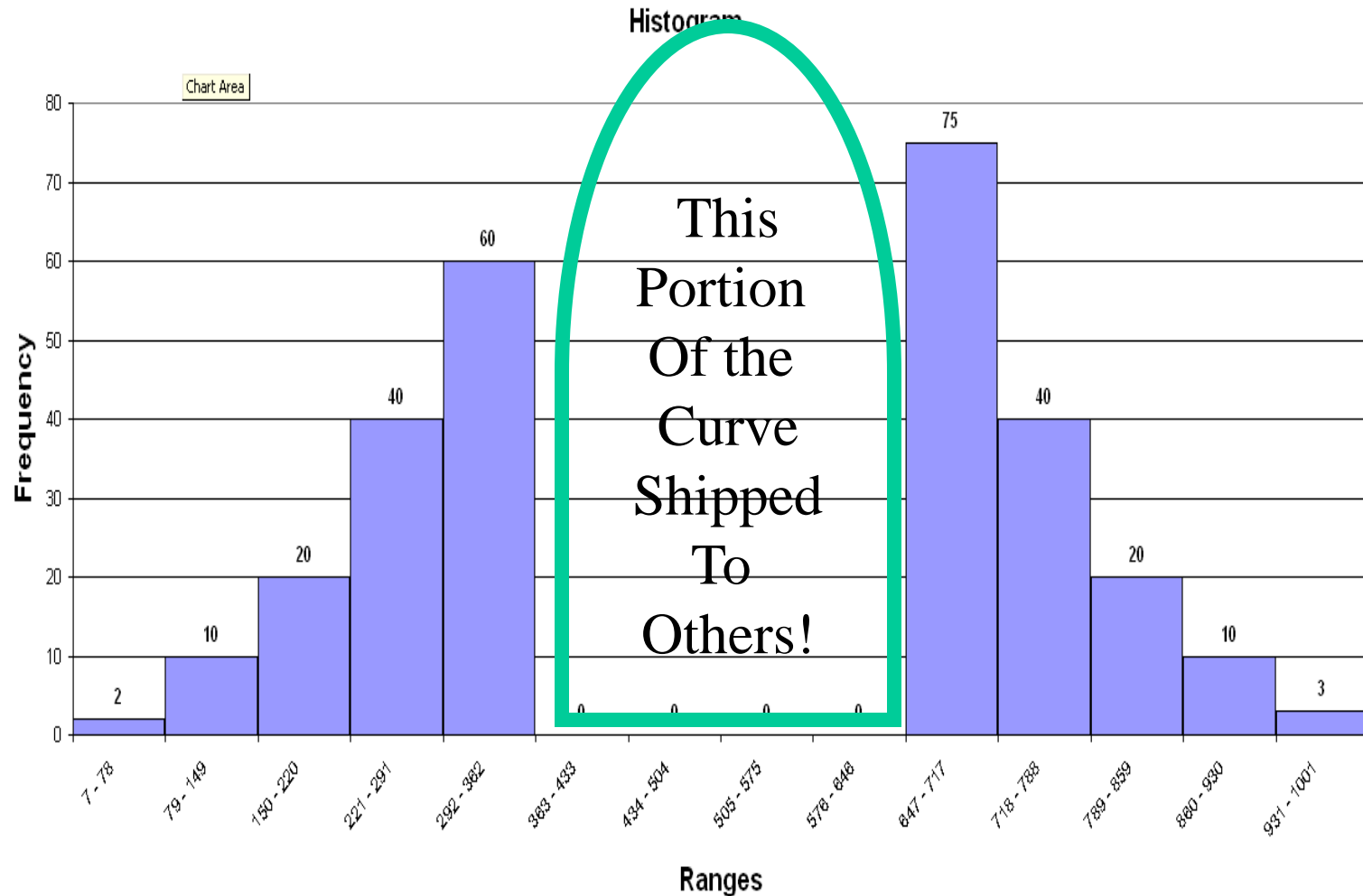
Histogram example: Bimodal

[alignment position when drive has loose bearing shifting left and right]



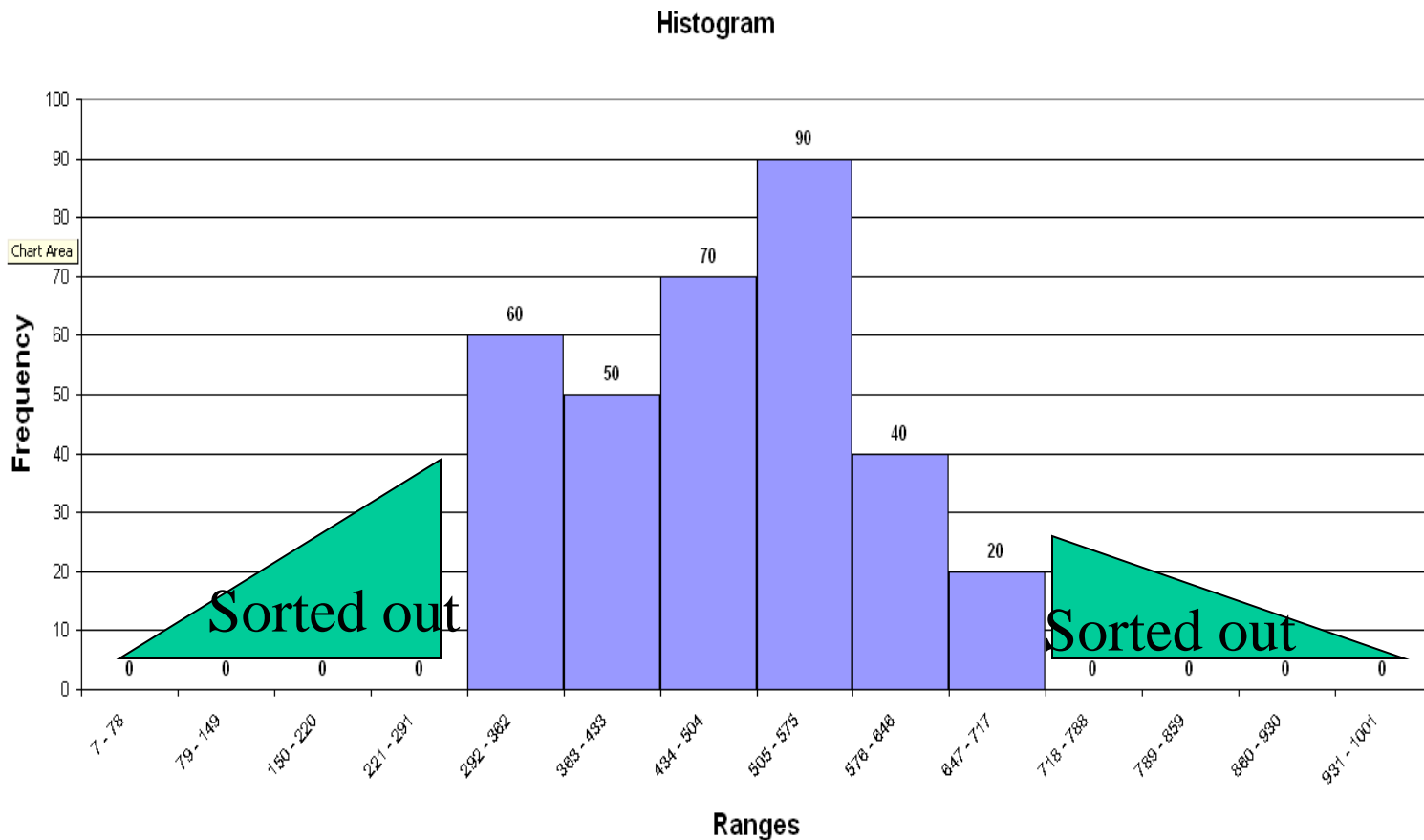
Histogram example: pre-sorted

(shipped the good stuff to the preferred customer)



Histogram example: pre-sorted

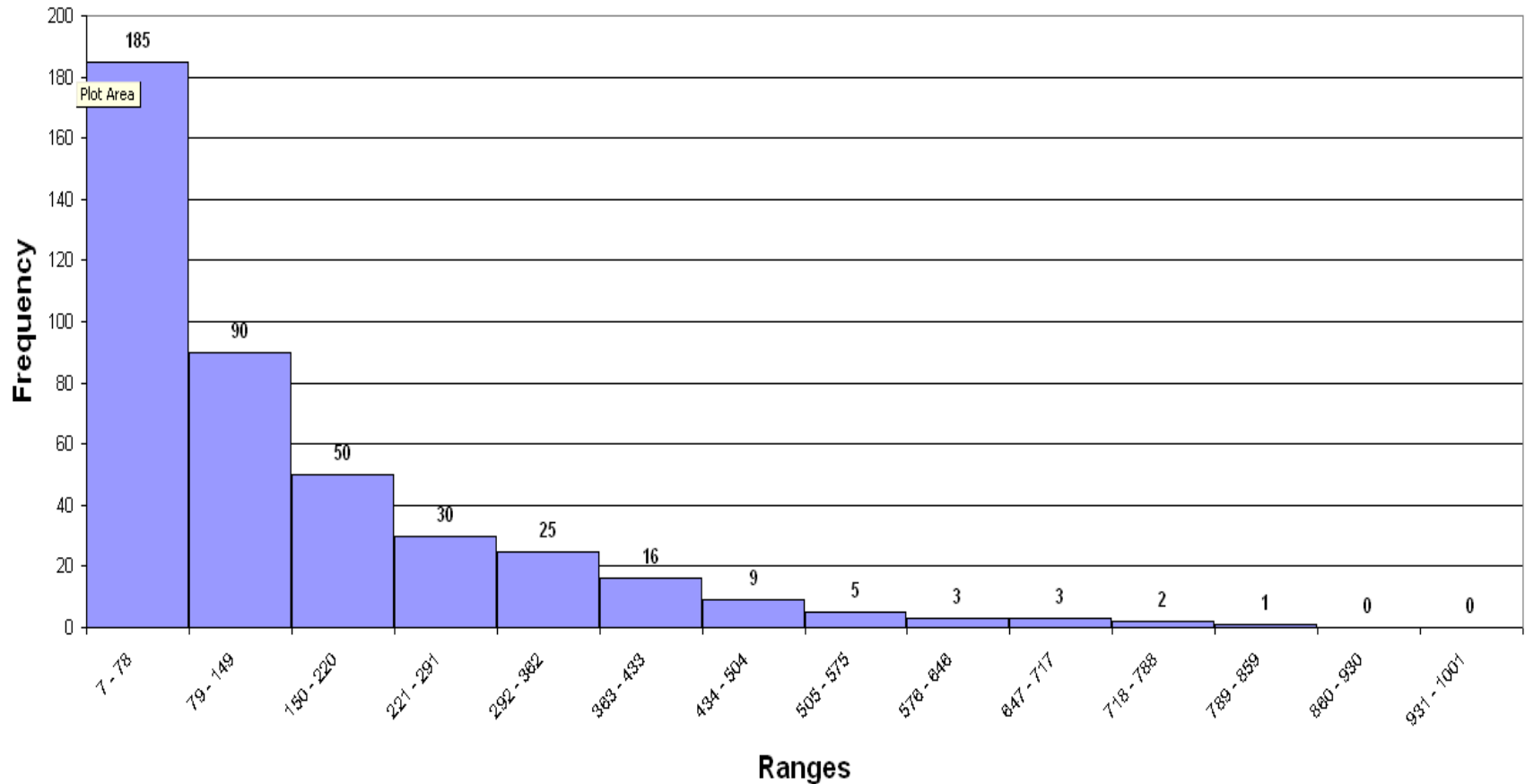
(you're the preferred customer and vendor can't control process)



Histogram example: Skew

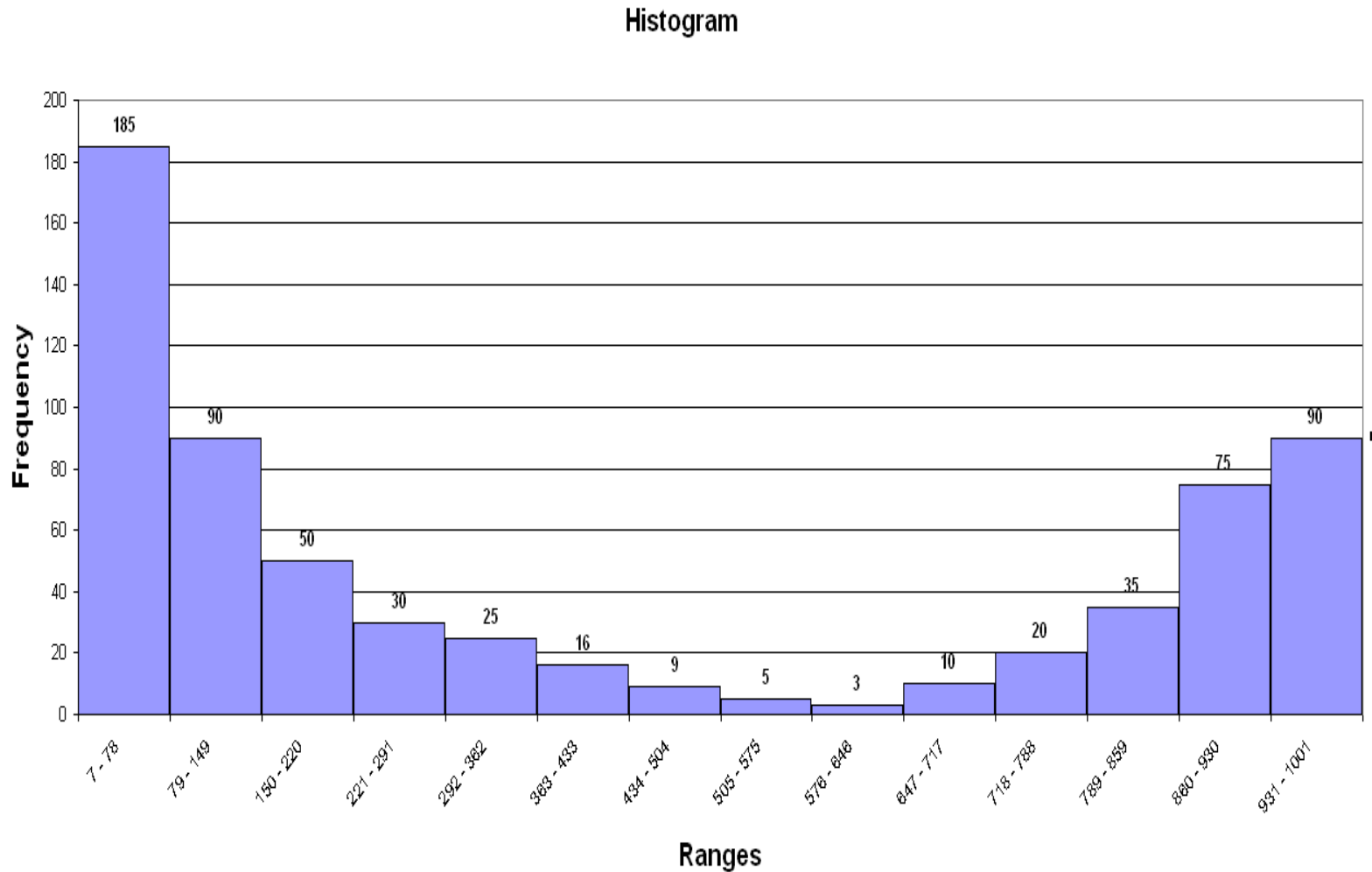
[one sided limit. Bouncing off of a hard stop collar]

Histogram



Histogram example: Bathtub

(reliability curve)



Selected Histogram types

- Uniform: equally distributed chance: Dice
- Normal: Traditional process under control
- Weibull: Queueing theory arrival times
- Pre-sorted: Pre-selected non-natural populations.
- Skew: data with a bias. Response times. Loaded dice.
- Bathtub: Reliability curve for describing infant mortality through wear-out.
- Poisson: distributions only allow 1 or 0 events so distributions will approach appropriate distribution at large sample size
- Exponential: arrival times

Percentages under the curve
or use of standard deviation apply
ONLY
to the specific curve type the data exhibits.

Normal curve

$$\pm 1\sigma = 68.27\%$$

$$\pm 2\sigma = 95.45\%$$

$$\pm 3\sigma = 99.73\%$$

$$\pm 4\sigma = 99.99733\%$$

$$\pm 5\sigma = 99.999943\%$$

Please refer to a statistics reference for percentages under the curve for other distributions.

Thank You

Paul Klein

kleincox@yahoo.com