Create a Histogram using Functions in Excel 2013

Milo Schield

Member: International Statistical Institute

US Rep: International Statistical Literacy Project

Director, W. M. Keck Statistical Literacy Project

Slides and Demo output at: www.StatLit.org/pdf/Excel2013-Create-Histogram-Using-Functions-slides.pdf
Excel2013-Create-Histogram-Using-Functions-demo.pdf

Histogram versus Bar/Column Chart

A bar (or column) chart involves bars that are separated because the data is categorical (male/female) or discrete numeric (# of kids in family).

A histogram involves bars (horizontal or vertical) that can touch because the data is continuous numeric (heights or weights).

Ordinal data (small, medium, large) can be either type.

The Goal and Approach

Goal: Summarize data using functions: COUNTIF and FREQUENCY (#8 and 9). Create histogram (#17).

Excel 2013 has two ways to summarize continuous data:
1. Using functions: COUNTIF or FREQUENCY
2. Using a command: **Histogram** in Data Analysis

Functions have a big advantage over commands.
- Functions update automatically when data changes.
- Commands require a manual update.

This presentation demonstrates both of the functions.

Create Excel Histogram of Q7 from this data: B1:I241

Data for Q1-Q4 (B-E) is Binary: 0=No, 1=Yes.
Data for Q5-Q6 (F-G) is Ordinal (discrete): 1-5.
Data for Q7-Q8 (H-I) is quantitative continuous

	A	B	C	D	E	F	G	H	I
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2	1	0	1	0	0	3	5	67	5
3	2	0	1	0	1	4	1	62	4
4	3	0	1	0	1	3	4	60	5
5	4	0	1	1	0	4	5	60	4
6	5	0	0	1	0	3	1	71	3

Data is at: www.StatLit.org/xls/Excel2013-Histogram-Functions-Data.xls

Steps in Creating a Histogram of Q7 (Col G) using functions

Summarize data into bins:
A. Determine the number of bins and the bin width
B. Setup bin ranges, bin maximums and bin counts
C1 Insert COUNTIF function to generate bin counts
C2 Insert FREQUENCY function to generate counts

Create a histogram chart using this summary data:
1a. Select bin range, maximum and data on spreadsheet
1b. From Insert ribbon, insert recommended chart
2. Delete extraneous series (Max series)
3. Create histogram: eliminate gaps between bars.

A: Determine the # of bins and the width of bins

	J	K	L	M	N	O	P	Q
5								
6	Step A:	Determine # of bins						
7		# of data values		240	=COUNTA(G2:G241)			
8		Est. # of bins		7.9	=LOG(M7,2)		# data values = 2^(#bins)	
9		Select # bins		8	Manual entry			
10								
11		Maximum		100	=MAX(G2:G241)			
12		Minimum		34	=MIN(G2:G241)			
13		Range		66	=M11-M12		Range = Max - Min	
14		Bin Width		8.25	=M13/M9		Bin width = Range/#bins	
15		Select width		8	Manual entry			
16								
17	J	K	L	M	N	O	P	Q

B: Generate K25 and K26. Copy K26 down to K33.

	J	K	L	M	N	O	P	Q
17								
18	Step B:	Set bin maximums ascending				*[Instructor already did this. See below.]*		
19		K25	=M$12+M$15-1			K41	=M$12+M$15-1	
20		K26	=K25+M$15			K42	=K41+M$15	
21		**Pull down K26 to K33**				**Pull down K42 to K49.**		

	J	K	L	M	N	O	P	Q
22								
23	Step C1:	Calculate bin count using COUNTIF						
24	Range	BinMax	Count					
25	33-41	41		=Countif(G$2:G$241,"<="&K25)				
26	41-49	49		=Countif(G$2:G$241,"<="&K26)-Sum(L$25:L25)				
27	49-57			* Pull-down L26 to L33				
28	57-65							
29	65-73							
30	73-81							
31	81-89							
32	89-97							
33	97-105							

Inserting the range before the bin maximums allows the chart to use the function-generated counts as the source. This means the chart will auto-update whenever the underlying data changes.

C1: Use COUNTIF function Enter L25 & L26. Pull L26 down.

	J	K	L	M	N	O	P	Q
22								
23	Step C1:	Calculate bin count using COUNTIF						
24	Range	BinMax	Count	Cell	**Formulas under Count for each bin****			
25	33-41	41	5	L25	=COUNTIF(G$2:G$241,"<="&K25)			
26	41-49	49	10	L26	=COUNTIF(G$2:G$241,"<="&K26)-SUM(L$25:L25)			
27	49-57	57	48	L27	=COUNTIF(G$2:G$241,"<="&K27)-SUM(L$25:L26)			
28	57-65	65	62	L28	=COUNTIF(G$2:G$241,"<="&K28)-SUM(L$25:L27)			
29	65-73	73	56	L29	=COUNTIF(G$2:G$241,"<="&K29)-SUM(L$25:L28)			
30	73-81	81	37	L30	=COUNTIF(G$2:G$241,"<="&K30)-SUM(L$25:L29)			
31	81-89	89	14	L31	=COUNTIF(G$2:G$241,"<="&K31)-SUM(L$25:L30)			
32	89-97	97	6	L32	=COUNTIF(G$2:G$241,"<="&K32)-SUM(L$25:L31)			
33	97-105	105	2	L33	=COUNTIF(G$2:G$241,"<="&K33)-SUM(L$25:L32)			
34			240	L34	=SUM(L25:L33)			

No $ sign

"&" links text strings

This use of the Max is the same as used by the Excel Histogram command in the Data-Analysis Toolpak.

C2a: If K41:K49 entered, skip this step else pull K42 down

	J	K	L	M	N	O	P	Q
35								
36	Step C2:	Calculate count for each bin using FREQUENCY						
37		a) Select L41:L49.				**[Array function will not fill all rows unless this is done]**		
38		b) In the formula bar, enter =Frequency(G$2:G$241, K41:K49) [Don't press Enter]						
39		c) Press and hold CTRL-SHIFT and then press ENTER (CSE command)						
40	Range	BinMax	Count	**Formulas under Count for each bin****				
41	33-41	41						
42	41-49	49						
43	49-57							
44	57-65							
45	65-73							
46	73-81							
47	81-89							
48	89-97							
49	97-105							

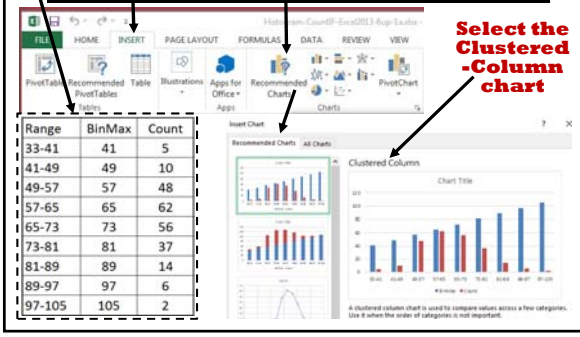
C2b: Use FREQUENCY function to calculate Counts per Bin

	J	K	L	M	N	O	P	Q
35								
36	Step C2:	Calculate count for each bin using FREQUENCY						
37		a) Select L41:L49.				**[Array function will not fill all rows unless this is done]**		
38		b) In the formula bar, enter =Frequency(G$2:G$241, K41:K49) **[Don't press Enter]**						
39		c) **Press and hold CTRL-SHIFT and then press ENTER: three-fingered CSE command.**						
40	Range	BinMax	Count	Formulas under Count for each bin**				
41	33-41	41						
42	41-49	49						
43	49-57	57						
44	57-65	65						
45	65-73	73						
46	73-81	81						
47	81-89	89						
48	89-97	97						
49	97-105	105						

Step a) Select area: Most common mistake
Step b) Do not press Enter!!
Step c) This is the trickiest operation in Excel.

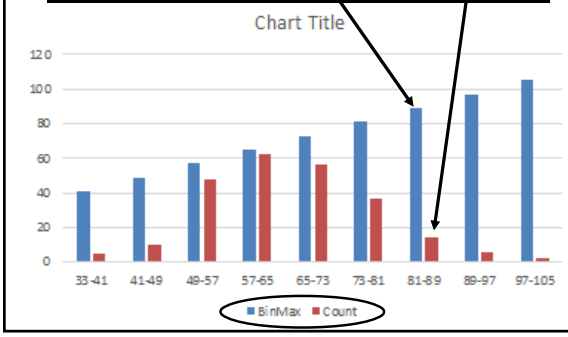
The FREQUENCY function is an array function: very tricky!

1: Manually select bin data; Insert "Recommended Chart"

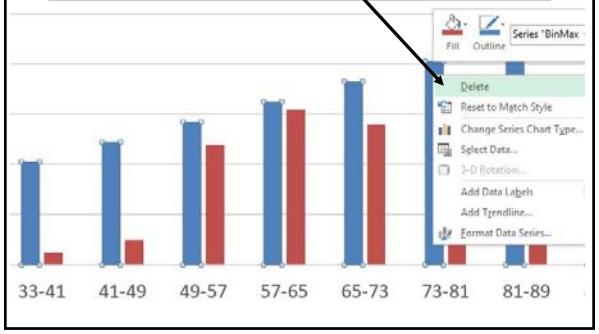


Range	BinMax	Count
33-41	41	5
41-49	49	10
49-57	57	48
57-65	65	62
65-73	73	56
73-81	81	37
81-89	89	14
89-97	97	6
97-105	105	2

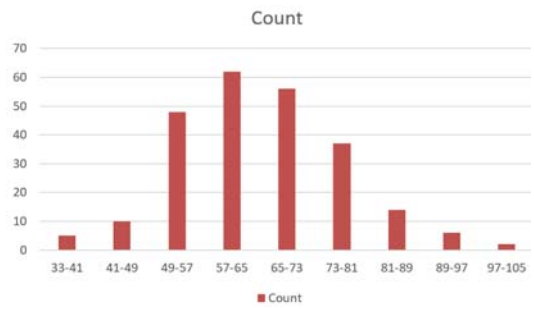
1b: This clustered-column Chart has two Series: Max and Count



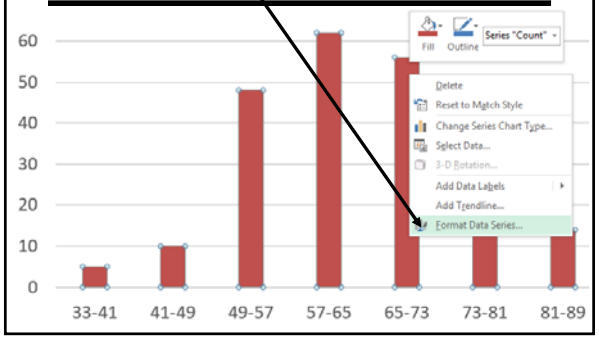
2a: Right-mouse Max-series bars; Select "Delete"



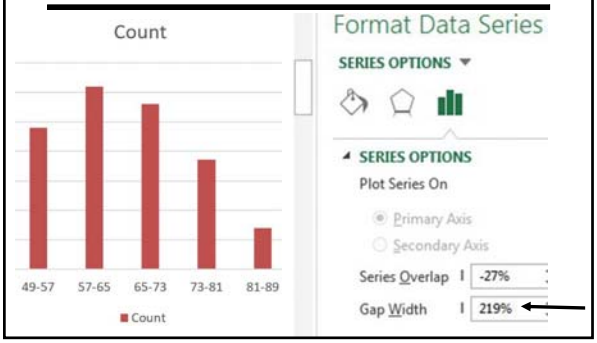
2b: Deleting the "Max" series yields a Count-series bar chart



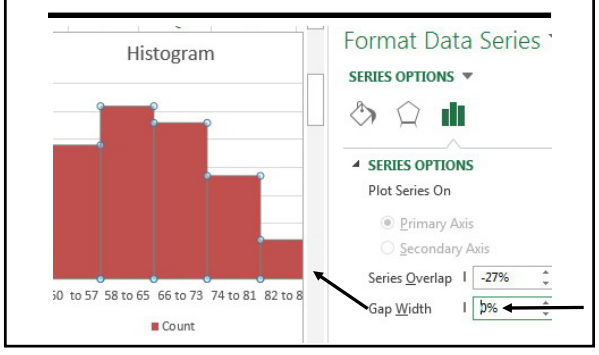
3a: Right-Mouse a "Count" bar; Select "Format Data Series"



3b: The default Gap Width is Double the Bar Width

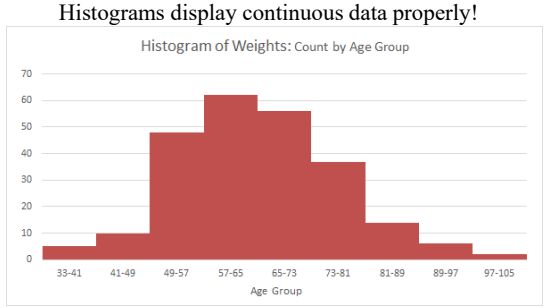


3c: Change Gap Width to zero. Result is a Histogram.



Conclusion

Histograms display continuous data properly!



Create a Histogram using Functions in Excel 2013

Milo Schield

Member: International Statistical Institute

US Rep: International Statistical Literacy Project

Director, W. M. Keck Statistical Literacy Project

Slides and Demo output at: www.StatLit.org/pdf/
Excel2013-Create-Histogram-Using-Functions-slides.pdf
Excel2013-Create-Histogram-Using-Functions-demo.pdf

Histogram versus Bar/Column Chart

A bar (or column) chart involves bars that are separated because the data is categorical (male/female) or discrete numeric (# of kids in family).

A histogram involves bars (horizontal or vertical) that can touch because the data is continuous numeric (heights or weights).

Ordinal data (small, medium, large) can be either type.

The Goal and Approach

Goal: Summarize data using functions: COUNTIF and FREQUENCY (#8 and 9). Create histogram (#17).

Excel 2013 has two ways to summarize continuous data:

1. Using functions: COUNTIF or FREQUENCY
2. Using a command: **Histogram** in Data Analysis

Functions have a big advantage over commands.

- Functions update automatically when data changes.
- Commands require a manual update.

This presentation demonstrates both of the functions.

Create Excel Histogram of Q7 from this data: B1:I241

Data for Q1-Q4 (B-E) is Binary: 0=No, 1=Yes.
Data for Q5-Q6 (F-G) is Ordinal (discrete): 1-5.
Data for Q7-Q8 (H-I) is quantitative continuous

	A	B	C	D	E	F	G	H	I
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2	1	0	1	0	0	3	5	67	5
3	2	0	1	0	1	4	1	62	4
4	3	0	1	0	1	3	4	60	5
5	4	0	1	1	0	4	5	60	4
6	5	0	0	1	0	3	1	71	3

Data is at: www.StatLit.org/xls/Excel2013-Histogram-Functions-Data.xls

Steps in Creating a Histogram of Q7 (Col G) using functions

Summarize data into bins:

A. Determine the number of bins and the bin width
B. Setup bin ranges, bin maximums and bin counts
C1 Insert COUNTIF function to generate bin counts
C2 Insert FREQUENCY function to generate counts

Create a histogram chart using this summary data:

1a. Select bin range, maximum and data on spreadsheet
1b. From Insert ribbon, insert recommended chart
2. Delete extraneous series (Max series)
3. Create histogram: eliminate gaps between bars.

A: Determine the # of bins and the width of bins

	J	K	L	M	N	O	P	Q
5								
6	Step A:	Determine # of bins						
7		# of data values		240	=COUNTA(G2:G241)			
8		Est. # of bins		7.9	=LOG(M7,2)		# data values = 2^(#bins)	
9		Select # bins		8	Manual entry			
10								
11		Maximum		100	=MAX(G2:G241)			
12		Minimum		34	=MIN(G2:G241)			
13		Range		66	=M11-M12		Range = Max - Min	
14		Bin Width		8.25	=M13/M9		Bin width = Range/#bins	
15		Select width		8	Manual entry			
16								
17	J	K	L	M	N	O	P	Q

B: Generate K25 and K26. Copy K26 down to K33.

	J	K	L	M	N	O	P	Q
17								
18	Step B:	Set bin maximums ascending				***[Instructor already did this. See below.]***		
19		K25	=M$12+M$15-1			K41	=M$12+M$15-1	
20		K26	=K25+M$15			K42	=K41+M$15	
21		**Pull down K26 to K33**				**Pull down K42 to K49.**		
22	J	K	L	M	N	O	P	Q
23	Step C1:	Calculate bin count using COUNTIF						

24	Range	BinMax	Count
25	33-41	41	
26	41-49	49	
27	49-57		
28	57-65		
29	65-73		
30	73-81		
31	81-89		
32	89-97		
33	97-105		

=Countif(G$2:G$241,"<="&K25)

=Countif(G$2:G$241,"<="&K26)-Sum(L$25:L25)

* Pull-down L26 to L33

Inserting the range before the bin maximums allows the chart to use the function-generated counts as the source. This means the chart will auto-update whenever the underlying data changes.

C1: Use COUNTIF function
Enter L25 & L26. Pull L26 down.

	J	K	L	M	N	O	P	Q
22								
23	Step C1:	Calculate bin count using COUNTIF						
24	Range	BinMax	Count	Cell	Formulas under Count for each bin**			
25	33-41	41	5	L25	=COUNTIF(G$2:G$241,"<="&K25)			
26	41-49	49	10	L26	=COUNTIF(G$2:G$241,"<="&K26)-SUM(L$25:L25)			
27	49-57	57	48	L27	=COUNTIF(G$2:G$241,"<="&K27)-SUM(L$25:L26)			
28	57-65	65	62	L28	=COUNTIF(G$2:G$241,"<="&K28)-SUM(L$25:L27)			
29	65-73	73	56	L29	=COUNTIF(G$2:G$241,"<="&K29)-SUM(L$25:L28)			
30	73-81	81	37	L30	=COUNTIF(G$2:G$241,"<="&K30)-SUM(L$25:L29)			
31	81-89	89	14	L31	=COUNTIF(G$2:G$241,"<="&K31)-SUM(L$25:L30)			
32	89-97	97	6	L32	=COUNTIF(G$2:G$241,"<="&K32)-SUM(L$25:L31)			
33	97-105	105	2	L33	=COUNTIF(G$2:G$241,"<="&K33)-SUM(L$25:L32)			
34			240	L34	=SUM(L25:L33)			

No $ sign

"&" links text strings

This use of the Max is the same as used by the Excel Histogram command in the Data-Analysis Toolpak.

C2a: If K41:K49 entered, skip this step else pull K42 down

	J	K	L	M	N	O	P	Q
35								
36	Step C2:	Calculate count for each bin using FREQUENCY						
37		a)	Select L41:L49.		**[Array function will not fill all rows unless this is done]**			
38		b)	In the formula bar, enter =Frequency(G$2:G$241, K41:K49) [Don't press Enter]					
39		c)	Press and hold CTRL-SHIFT and then press ENTER (CSE command)					
40	Range	BinMax	Count	**Formulas under Count for each bin****				
41	33-41	41						
42	41-49	49						
43	49-57							
44	57-65							
45	65-73							
46	73-81							
47	81-89							
48	89-97							
49	97-105							

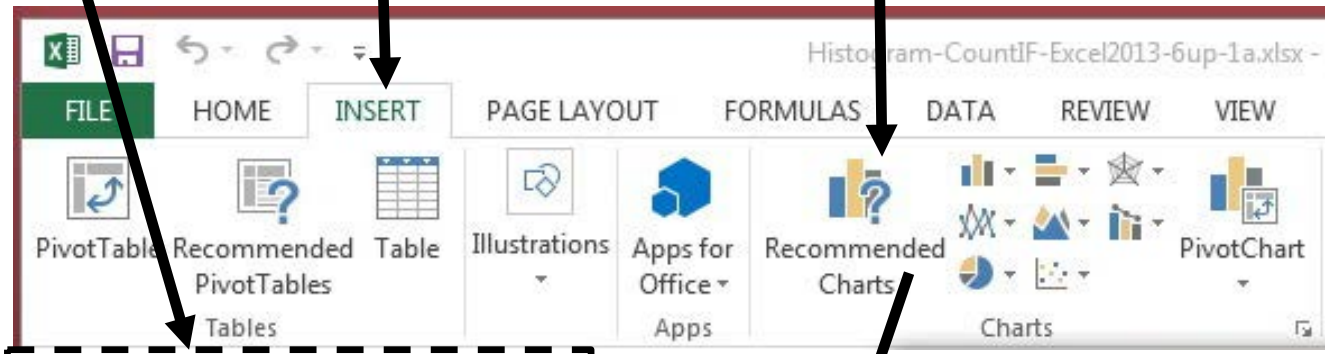
C2b: Use FREQUENCY function to calculate Counts per Bin

	J	K	L	M	N	O	P	Q
35								
36	Step C2:	Calculate count for each bin using FREQUENCY						
37		a)	Select L41:L49.			[Array function will not fill all rows unless this is done]		
38		b)	In the formula bar, enter =Frequency(G$2:G$241, K41:K49) [Don't press Enter]					
39		c)	Press and hold CTRL-SHIFT and then press ENTER: three-fingered CSE command.					
40	Range	BinMax	Count	Formulas under Count for each bin**				
41	33-41	41						
42	41-49	49						
43	49-57	57						
44	57-65	65						
45	65-73	73						
46	73-81	81						
47	81-89	89						
48	89-97	97						
49	97-105	105						

Step a) Select area: Most common mistake
Step b) Do not press Enter!!
Step c) This is the trickiest operation in Excel.

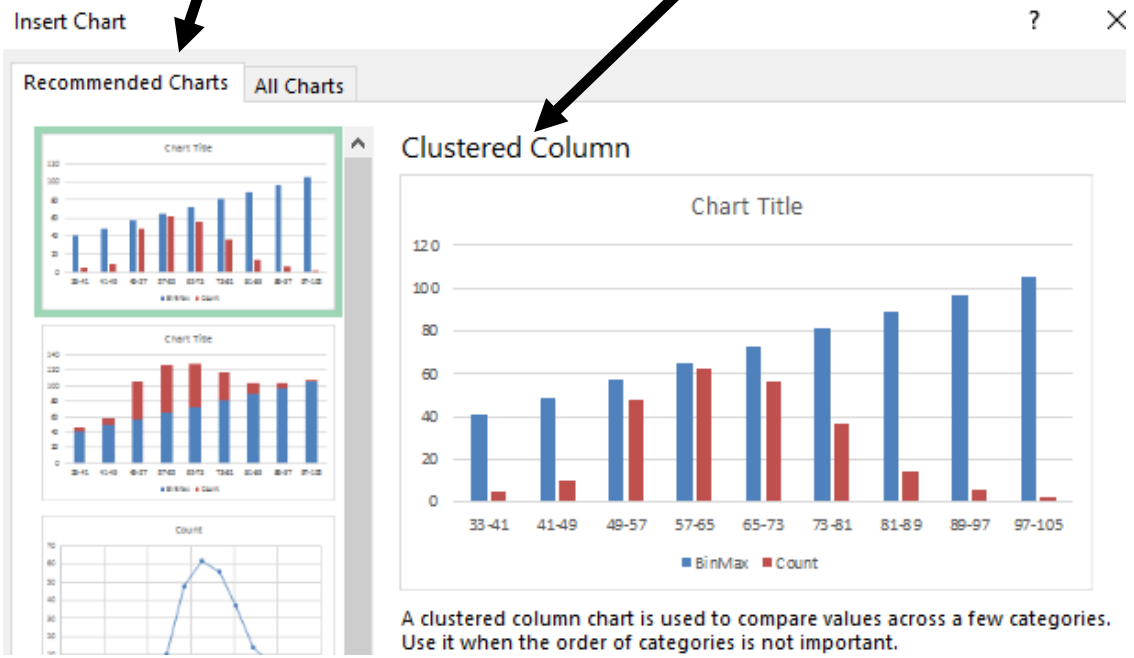
The FREQUENCY function is an array function: very tricky!

1: Manually select bin data; Insert "Recommended Chart"

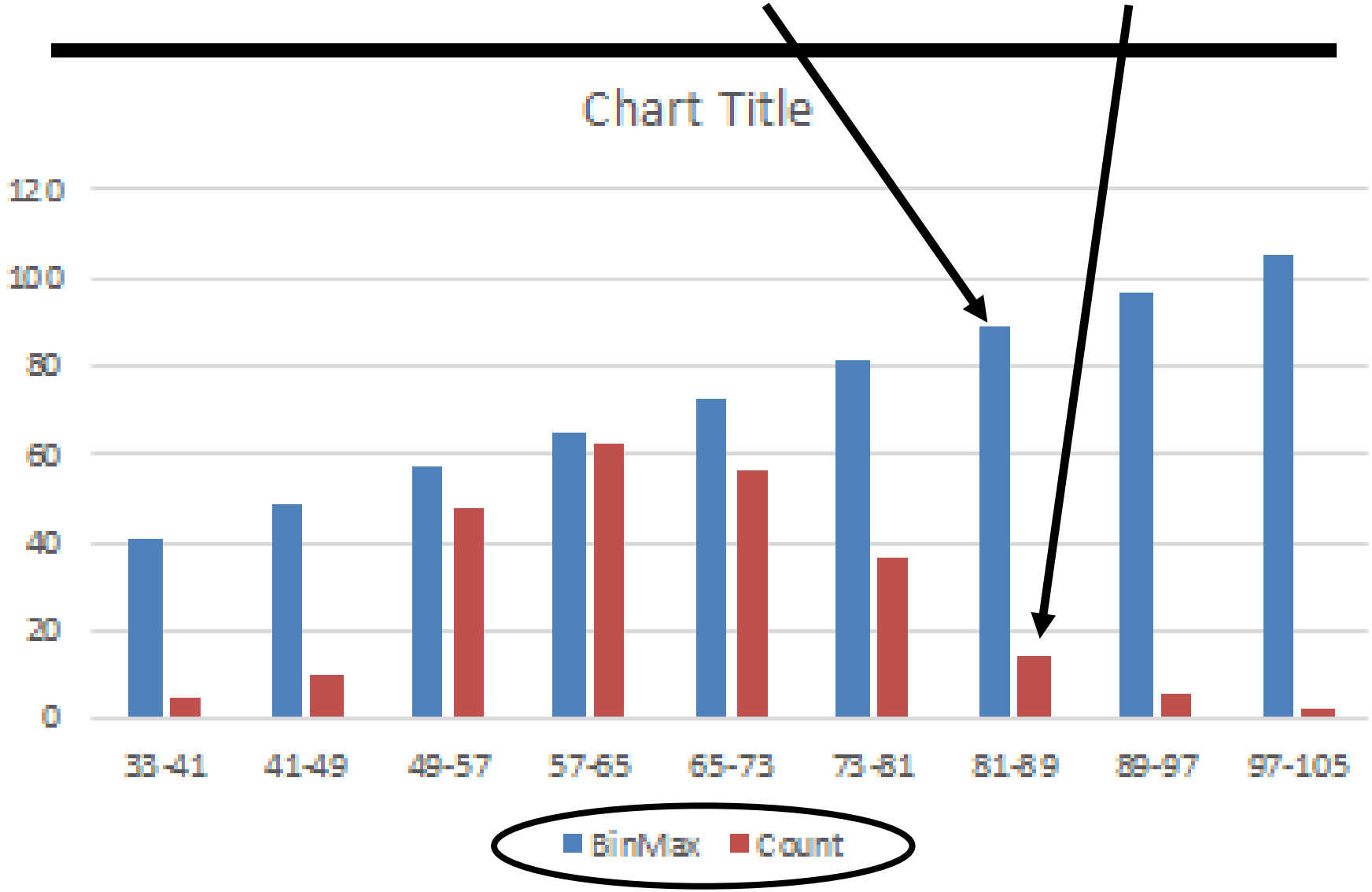


Select the Clustered -Column chart

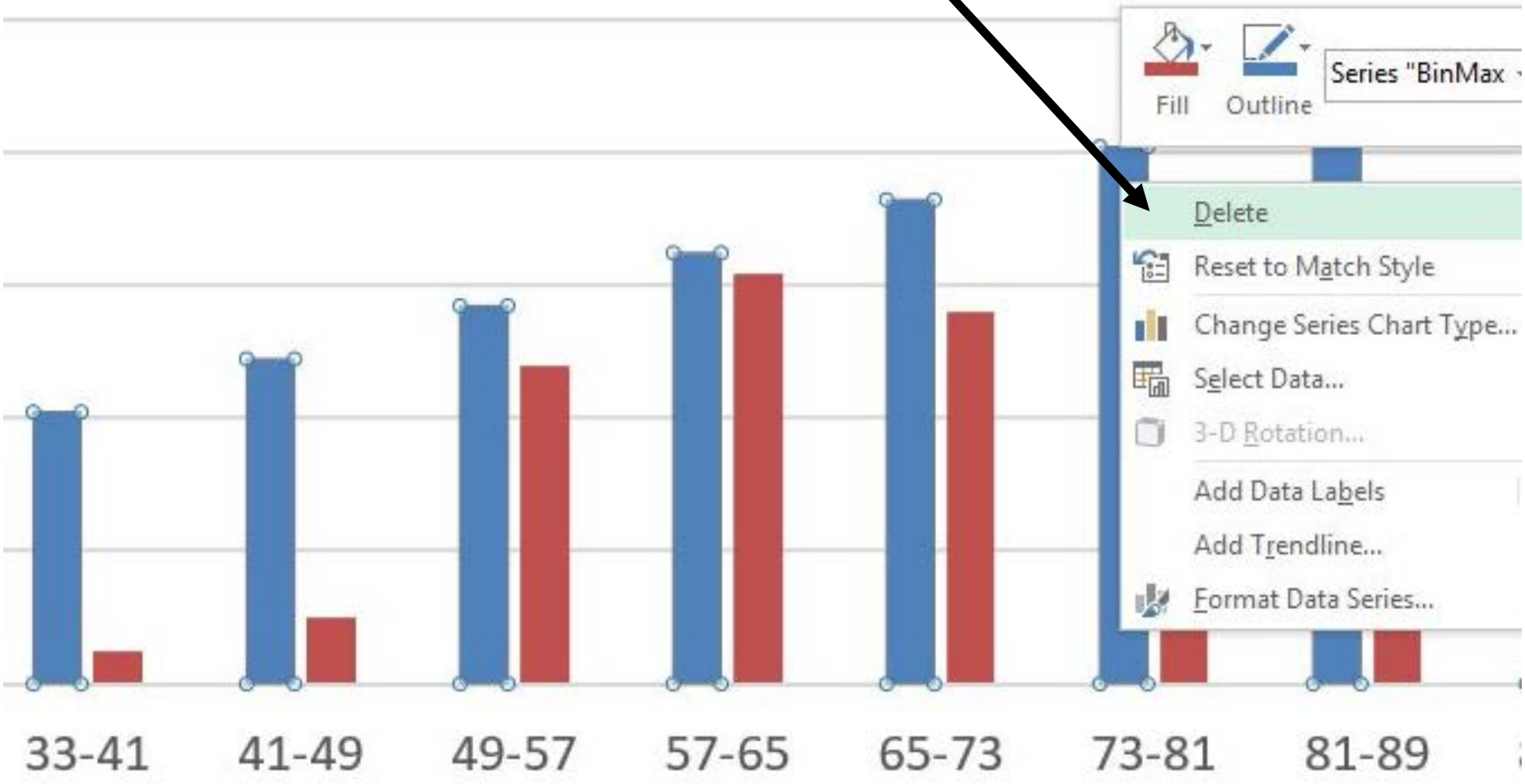
Range	BinMax	Count
33-41	41	5
41-49	49	10
49-57	57	48
57-65	65	62
65-73	73	56
73-81	81	37
81-89	89	14
89-97	97	6
97-105	105	2



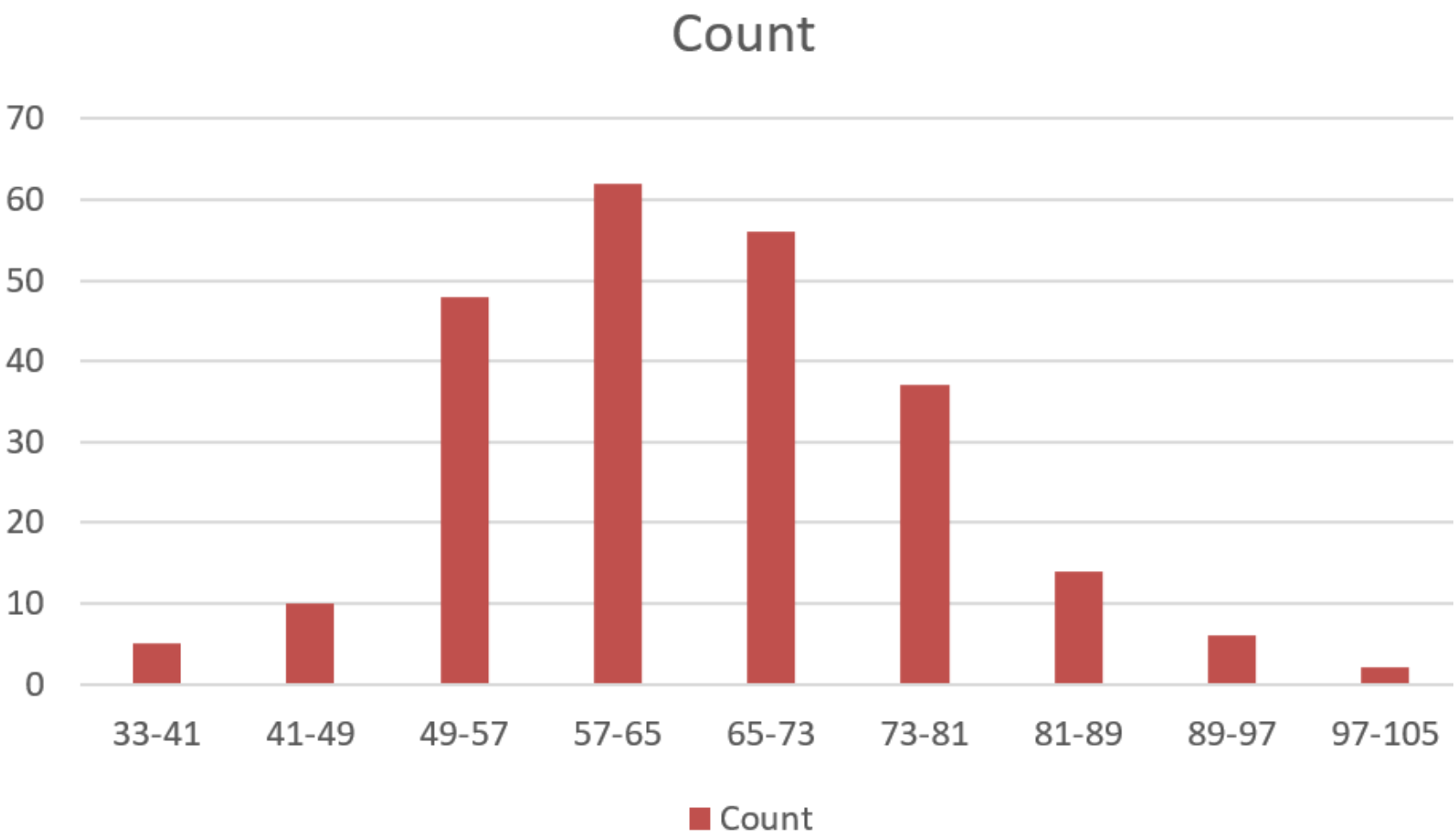
1b: This clustered-column Chart has two Series: Max and Count



2a: Right-mouse Max-series bars; Select "Delete"

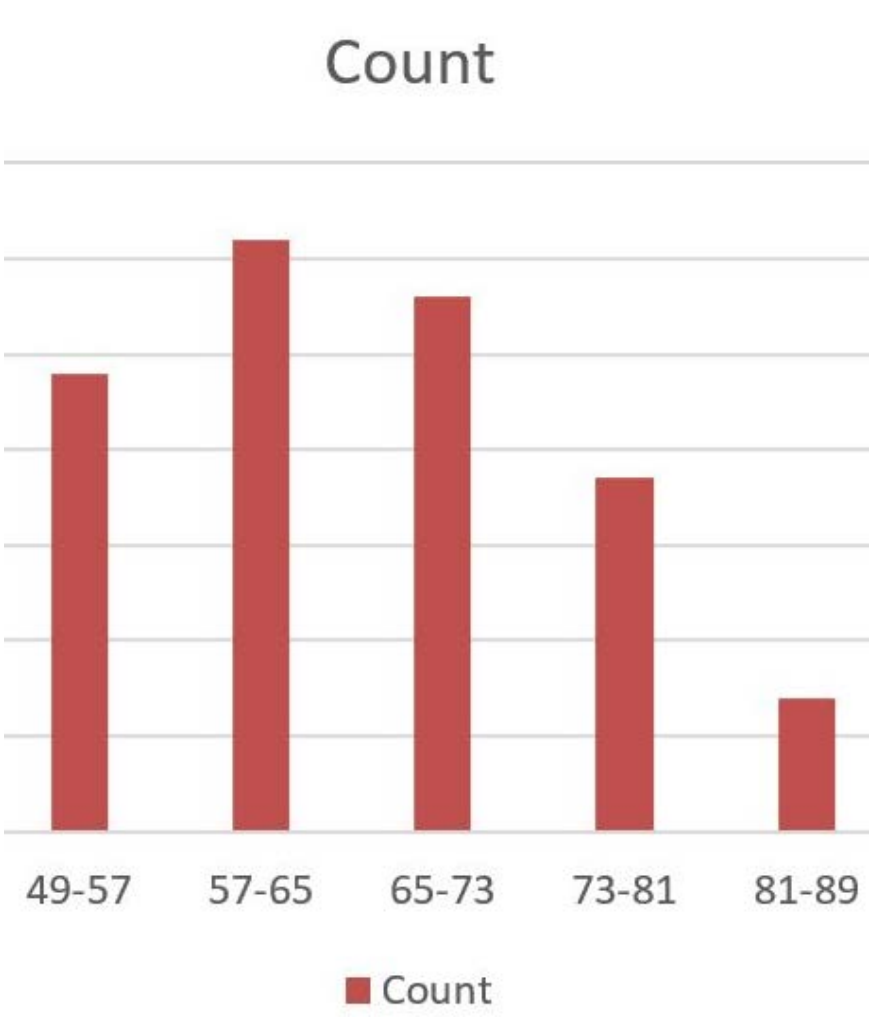


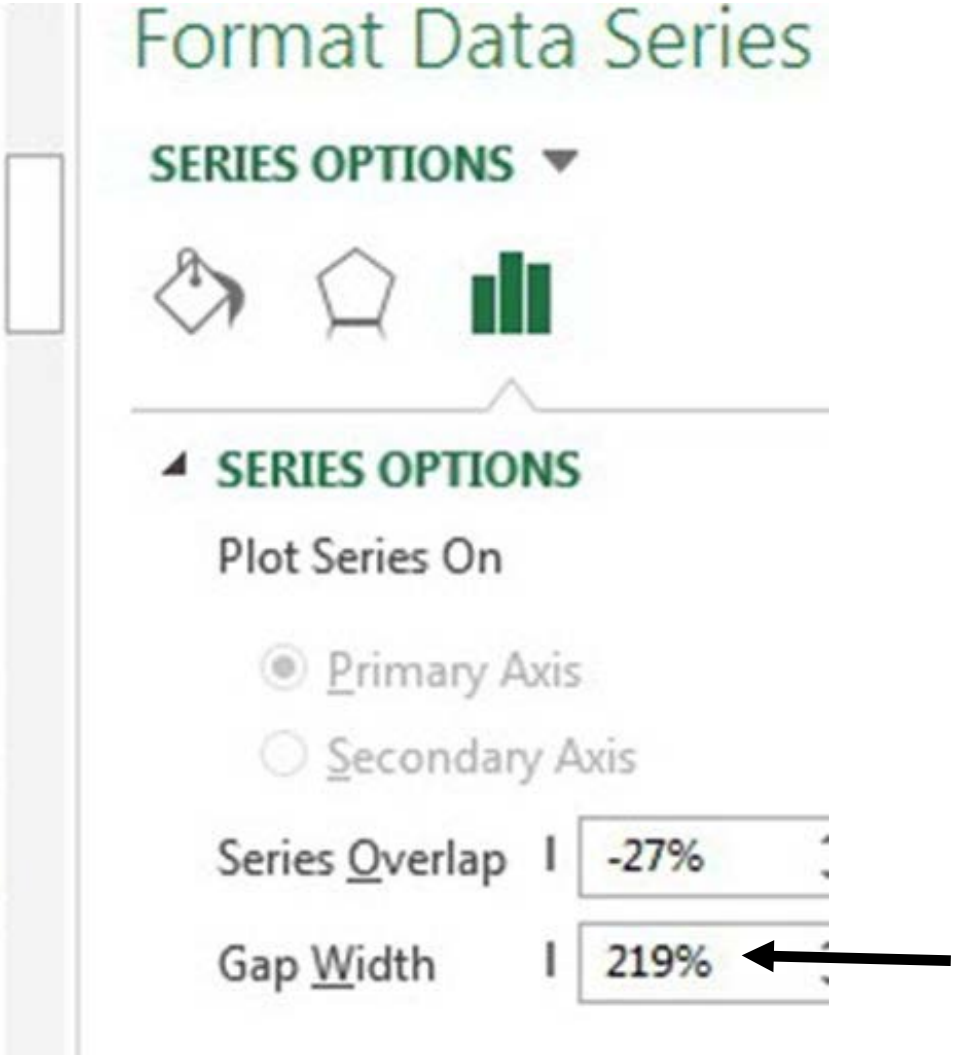
2b: Deleting the "Max" series yields a Count-series bar chart



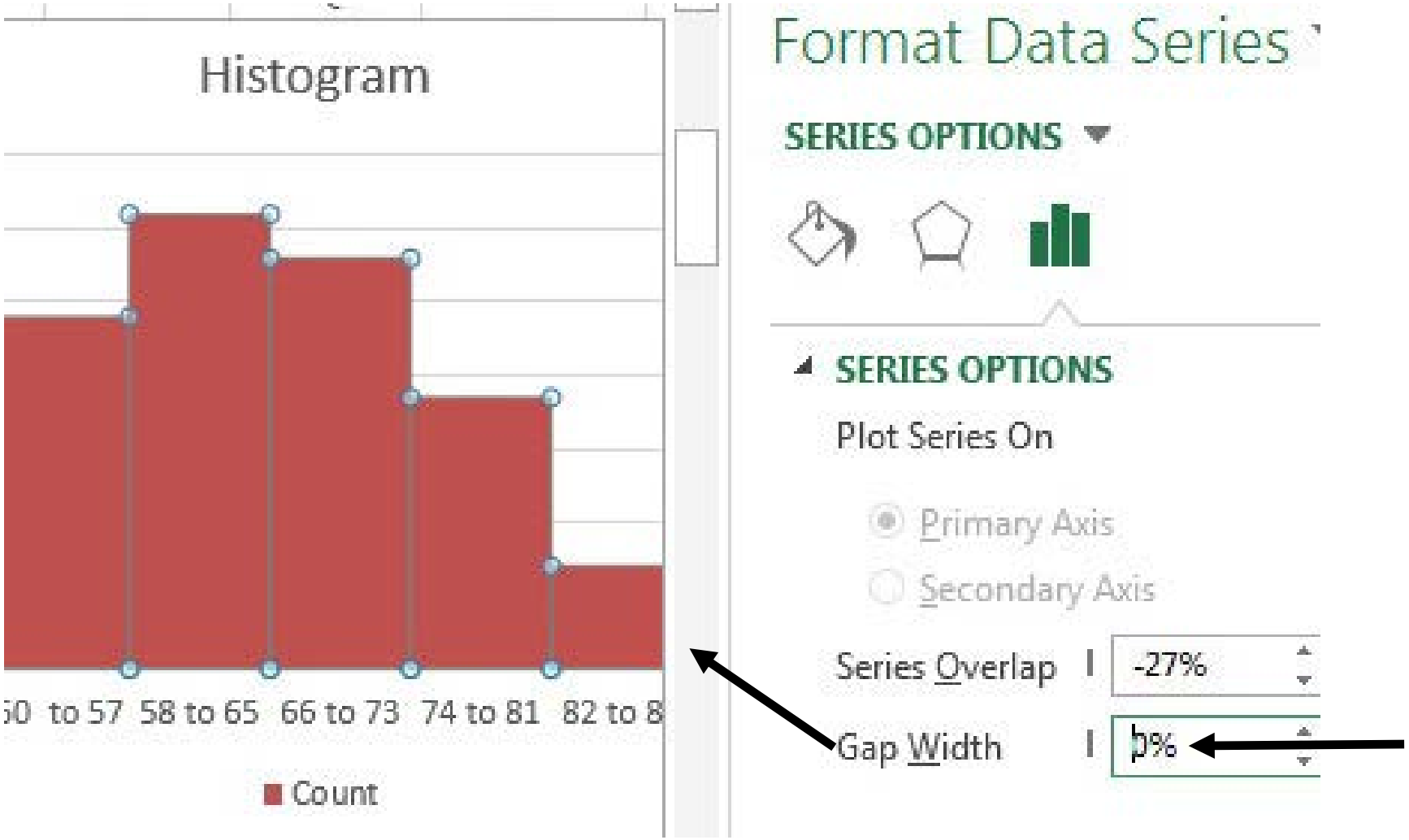
3a: Right-Mouse a "Count" bar; Select "Format Data Series"

3b: The default Gap Width is Double the Bar Width





3c: Change Gap Width to zero. Result is a Histogram.



Conclusion

Histograms display continuous data properly!