Lecture 23

Produced by Dr. Worldwide

Welcome to the 305

Descriptive Statistics

- A statistic is a quantity computed from a sample AKA a function of the data

- Descriptive statistics refers to the analysis of data that helps to describe, visualize, or summarize features in a sample

- Purposes for descriptive statistics
  - Understand the data without seeing the entire dataset
  - Quantify the center of the data
  - Quantify the spread of the data
  - Capture patterns in the data

- A frequency table is an array that gives that lists all possible values with counts of how often they occur in the sample (shows the sample distribution

- For a numeric variable it is helpful to organize possible values in bins

Ex: Starting Salaries

- Suppose we have starting salaries of 50 recently graduated students with B.S. degrees in data analytics, recorded the in thousands of dollars per year

- Download Salaries.xlsx from link Sheet 1 on course website

- Focus on sheet named Frequency

- Raw data contained in cells A4:A53

- Frequency table contained in cells D4:E17

- To create, use the following Excel function

  =FREQUENCY(data array, bins array)

- For more help, see Link 1 on course website

	Salaries
3	**Salaries**
4	84.8181
5	90.4642
6	82.7153
7	83.319
8	89.9589
9	99.1958
10	85.382
11	92.2283
12	88.6465
13	96.7041
14	83.4656

Bins	Frequency
80	2
82	3
84	6
86	6
88	2
90	7
92	8
94	3
96	5
98	2
100	4
102	0
104	0
	2

Ex: Starting Salaries

- Bins need to be specified beforehand

- The first bin counts the number of values smaller than the bin value and all other bins count the number of values larger than the previous bin but smaller than the current one

- To use FREQUENCY, you need to select the array where the results will be displayed and press
  - CTRL+SHIFT+ENTER   (Windows)
  - CTRL+U or ⌘ +RETURN   (Mac)

- Current frequency table breaks up the range of values into bins of size 2

- Go to sheet named Practice to create frequency table with bins of size 3

Ex: Starting Salaries

- Bins need to be created
  - Select the three cells C4:C6
  - Hover cursor over the bottom right corner until you see "+"
  - Click and drag cursor down to C12

Bins	Frequency
80	
83	
86	
	104
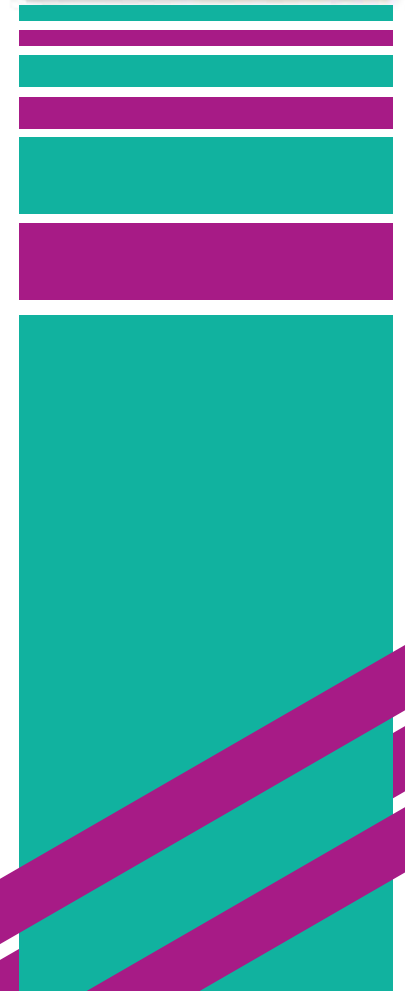
Bins	Frequency
80	
83	
86	
89	
92	
95	
98	
101	
104	

- Excel detects and continues patterns

Ex: Starting Salaries

- Frequencies need to be added to the table
  - Select D4 (or D4:D13) and input the formula

    =FREQUENCY(A4:A53, C4:C12)

  - You may need to use shortcuts to create vector of frequencies

- Q: Why is the frequency table better than the original data?

Bins	Frequency
80	2
83	4
86	11
89	7
92	10
95	6
98	4
101	4
104	0
	2

Descriptive Statistics

- A histogram is a graph based off a frequency table for a numeric variable

- The histogram is the most common graph used to understand the variability in numeric data

- The histogram is built off rectangles whose heights represent the number of observations within each bin

- Histogram gives us an estimate of the distribution in the population

- Three things we learn from a histogram
  - Shape of the distribution (left skewed or right skewed)
  - Range of values
  - Outlier identification
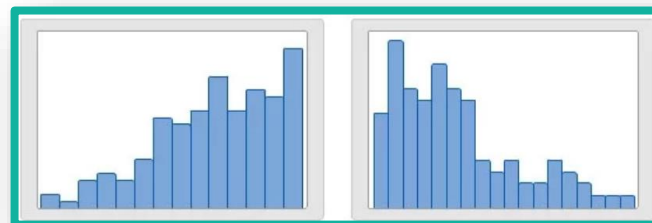
Ex: Starting Salaries

- Various ways to build a histogram in Excel

- Created from frequency table
  - Copy and paste frequency values (not formulas)

Bins	Frequency	Bins	Frequency
80	2	<= 80	2
83	4	80-83	4
86	11	83-86	11
89	7	86-89	7
92	10	89-92	10
95	6	92-95	6
98	4	95-98	4
101	4	98-101	4
104	0	101-104	0
	2	>=104	2

  - Highlight the two columns of bin names and frequencies

MIAMI

Ex: Starting Salaries

- Created from frequency table
  - Select appropriate chart (clustered column) through Insert menu
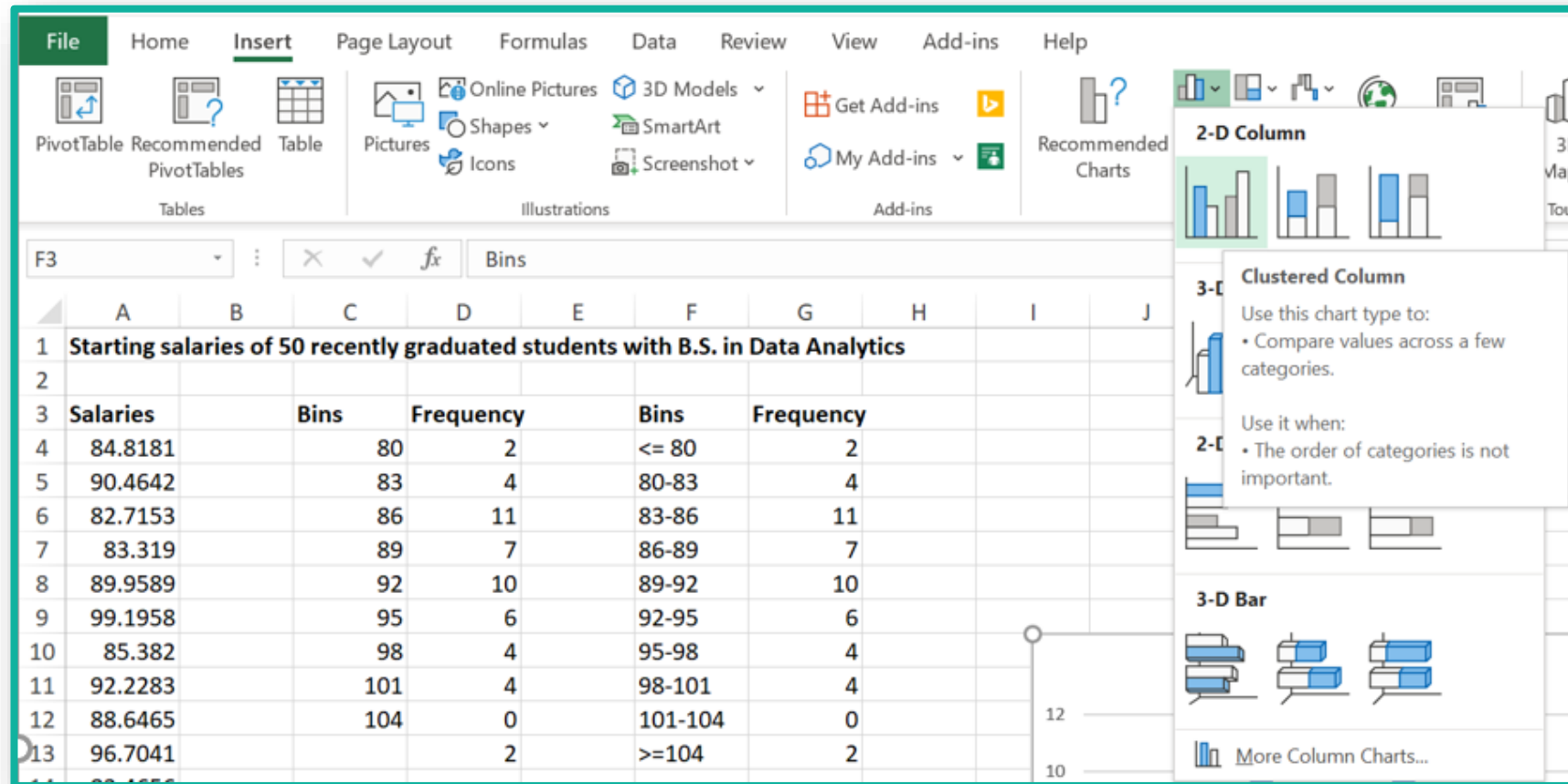
Ex: Starting Salaries

- Created from frequency table
  - Observe the default plot which needs to be fixed
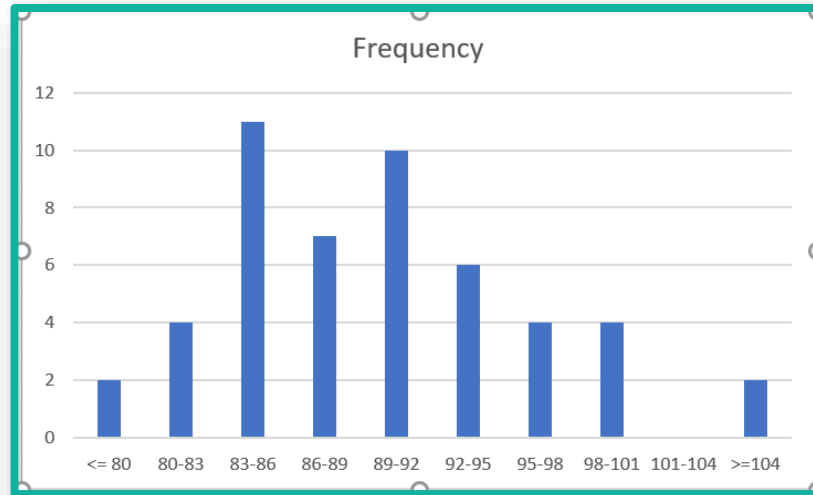
  - Double click on plot to get sidebar menu for easy editing of elements

  - Select the x-axis labels for modification and try to figure out how to rotate the text to -21 degrees
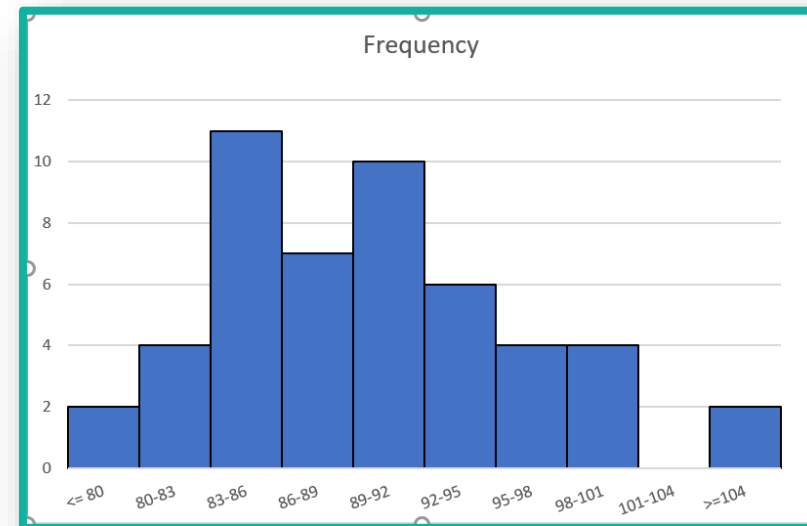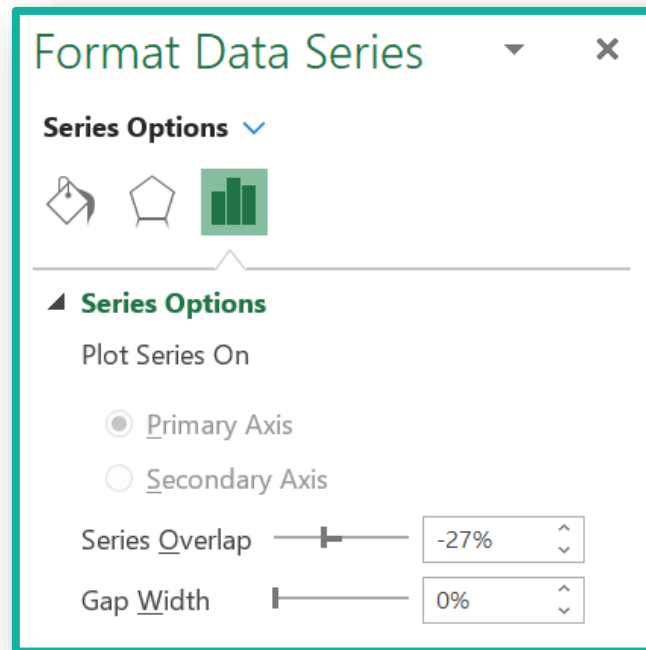
Ex: Starting Salaries

- Created from frequency table
  - Select text options, then textbox, then make custom angle -21 degrees
  - Select the rectangles and give boxes a 1pt black border
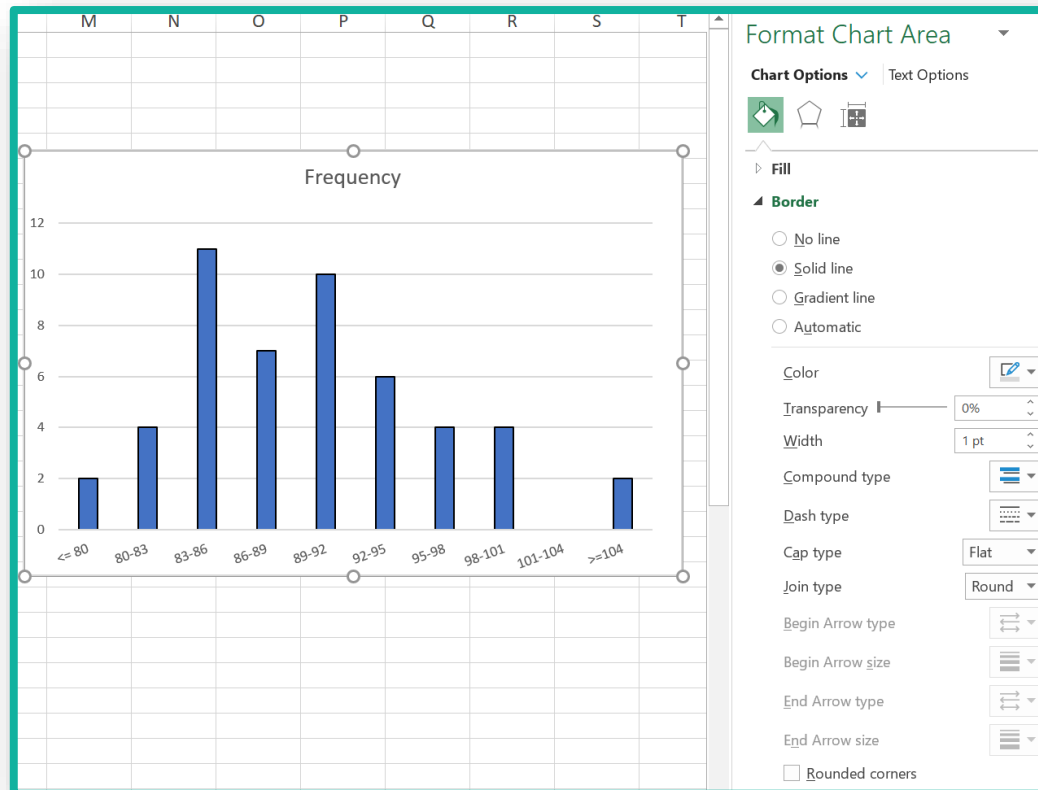
Ex: Starting Salaries

- Created from frequency table
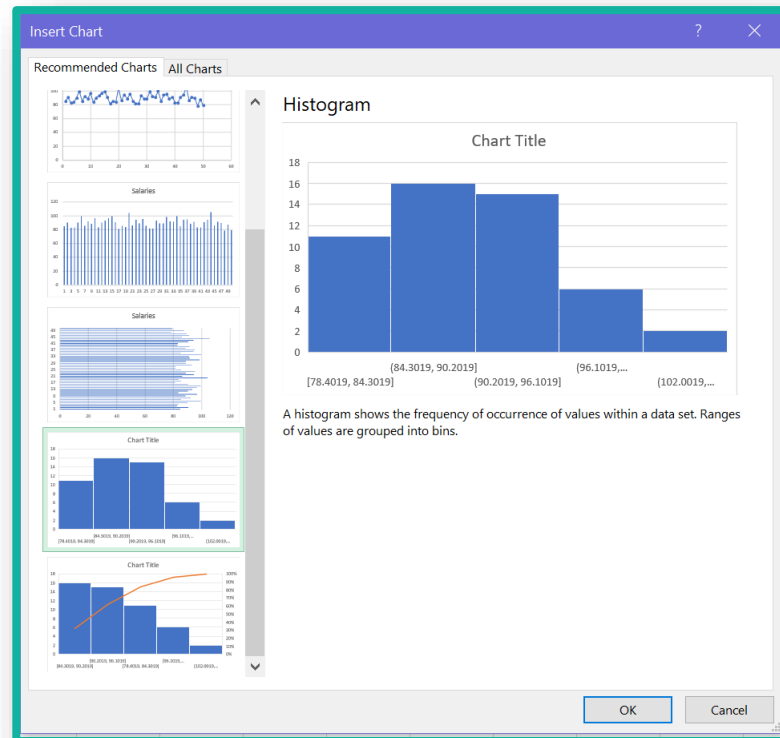 - When rectangles are selected, select Series Options, and make Gap Width=0

- Select the title Frequency and change to Starting Salaries

Ex: Starting Salaries

- Created from raw data
  - Highlight raw data and variable name in A3:A53
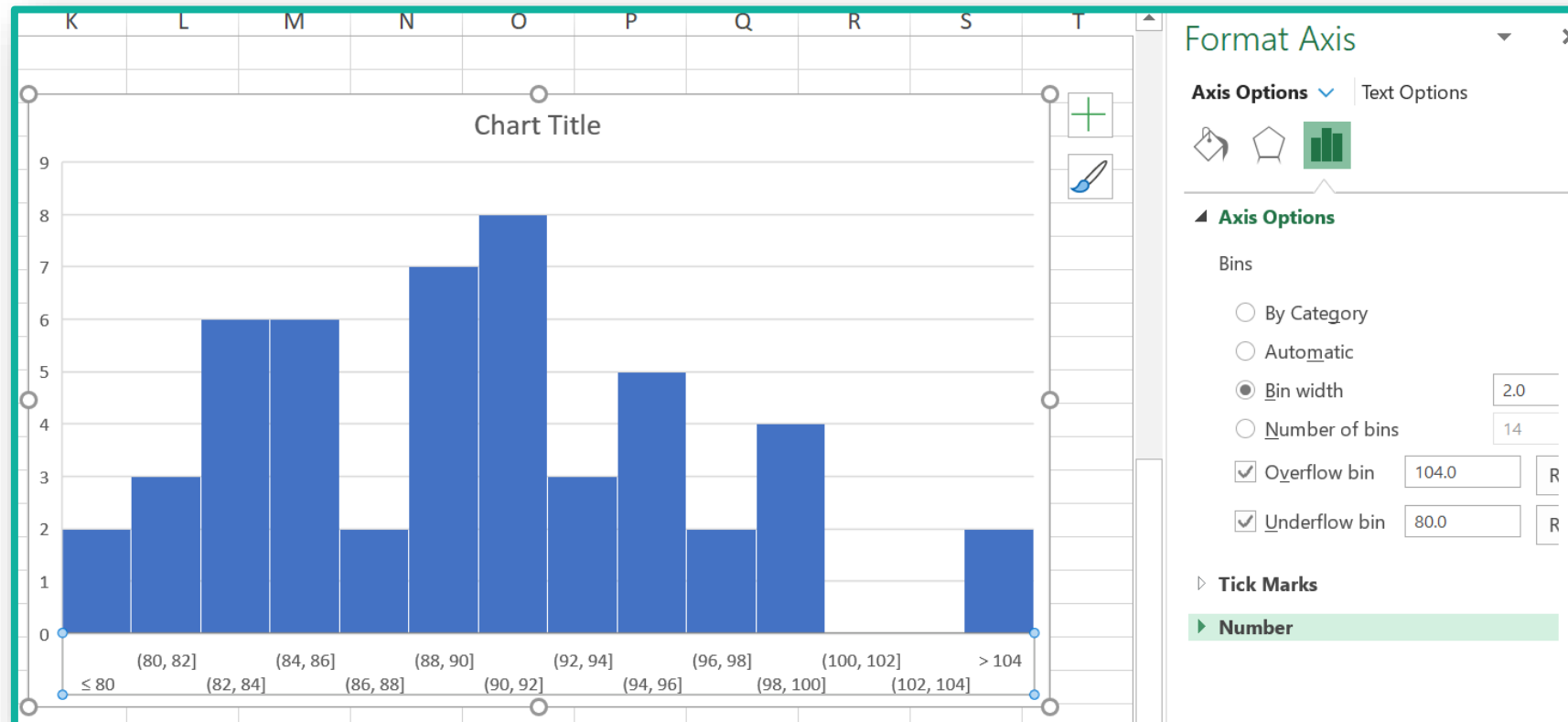 - Select Recommended Charts in Insert menu
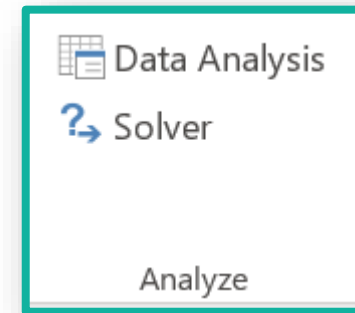 - Select Histogram and select OK

Ex: Starting Salaries

- Created from raw data
  - Select x-axis and modify Axis Options
  - We can build a histogram similar to previous histogram

Ex: Starting Salaries

- Created using the Analysis ToolPak add-in
  - Look at Link 2 for help with loading the Analysis ToolPak
  - This may be automatically loaded on your computer
  - Above Excel Solver you will find Data Analysis
  - Select Histogram and fill out fields

| Data Analysis |
| Solver |
| Analyze |

Histogram   ? ✕

Input

Input Range: $A$3:$A$53   ▲

Bin Range: $C$3:$C$12   ▲

☑ Labels

OK
Cancel
Help

Output options

○ Output Range: ▲

◉ New Worksheet Ply: Plot Example

○ New Workbook

☐ Pareto (sorted histogram)

☐ Cumulative Percentage

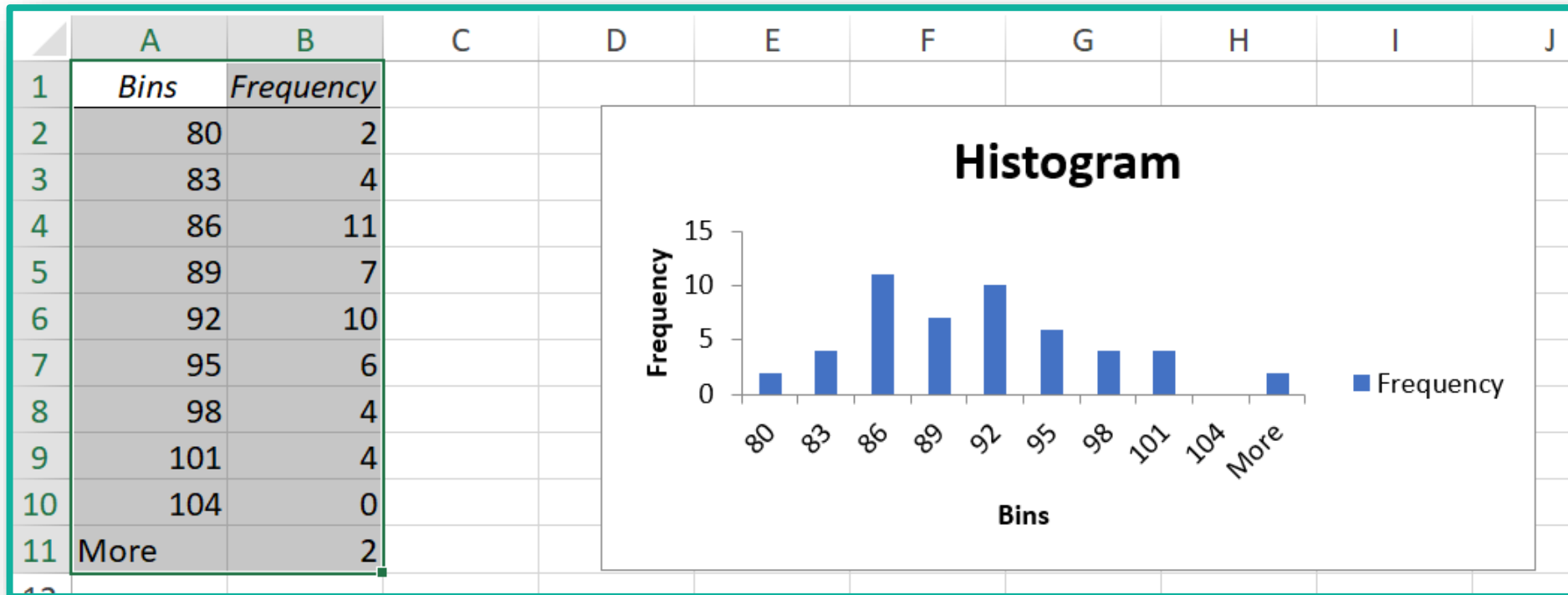☑ Chart Output

Ex: Starting Salaries

- Created using the Analysis ToolPak add-in
  - Result found in new sheet named Plot Example

- Requires more modification

The End

Dale