C.A.L.F
STEWARDSHIP

CATTLE. ANTIMICROBIAL. LAND. FINANCIAL.

Microsoft Excel for the Veterinary Practitioner: Creating Charts and Graphs

W. Isaac Jumper, DVM

Microsoft Excel for the Veterinary Practitioner:
Creating Charts and Graphs
W. Isaac Jumper, DVM

Introduction:

Graphs and charts are useful tools to the veterinary practitioner for displaying data, identifying patterns, and communicating analysis results. Microsoft Excel has many graphing capabilities, two of which are bar graphs and scatter plots. Bar graphs are useful for displaying categorical data (i.e. nominal or ordinal data) whereas scatter plots are useful for displaying continuous data. The objective of this article is to describe the process of creating bar graphs and scatter plots of animal health data in Microsoft Excel. systems.

Creating a bar graph:

It is important to match the type of graph created to the type of data that has been collected. Bar graphs represent data in categories; the x-axis (i.e. horizontal axis) typically displays each category, and the y-axis (i.e. vertical axis) typically displays counts or percent. Histograms are a special type of bar graph that can represent continuous data organized into categories (e.g. calf weaning weights that are continuous from 450 lbs. to 650 lbs. might be organized into 50 lb. increments). Scatter plots are used to graph quantitative data, where the x and y-axes each may represent a discrete or continuous data measurement (e.g. body weight vs. age in days).

Prior to inserting a graph or chart, it is recommended to click/select a blank cell outside of your data in the spreadsheet. If a cell is selected that contains data when the graph is added, Excel will populate the graph with data that is often wrong. To insert a graph into a spreadsheet, click on the "Insert" tab at the top of the spreadsheet. There is a "Charts" group in the middle of the "Insert" tab ribbon, and various types of charts can be found in this group. Hovering the cursor over the bar graph image will provide a description of the bar graph option seen (Figure 1). Clicking on the bar graph icon will then produce a drop-down menu with the available options for a bar or column chart. One of the more commonly used options is the 2-D Clustered Column graph. After clicking on the 2-D Clustered Column option, a graph will appear in the spreadsheet. If a blank cell was selected prior to adding the chart, the cha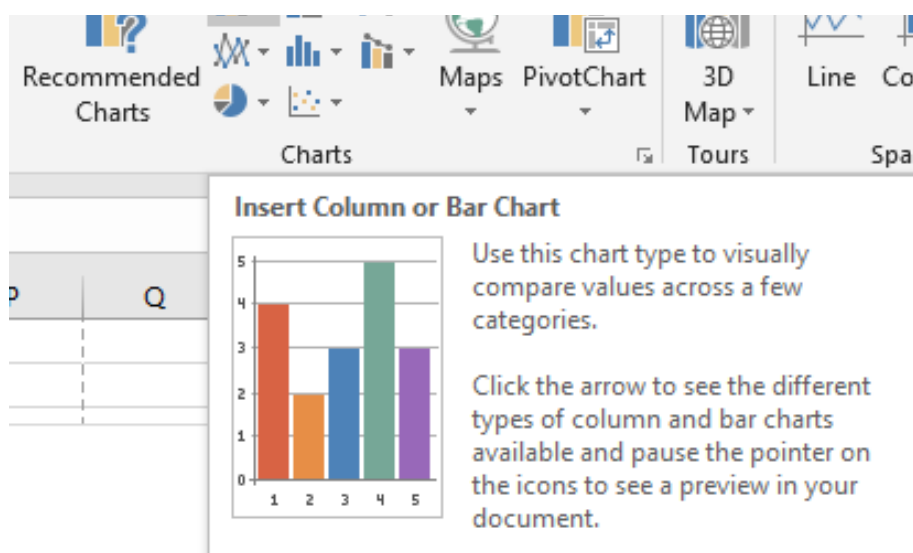rt will be blank. Once a graph has been added to the spreadsheet, the user must select the data that will be displayed in the graph. To select the data that will be used in the graph, either right-click on the graph itself, and select the "Select Data" option, or click on the graph and under the "Design" tab at the top of the spreadsheet, click on the "Select Data" button. Either method will produce the same result; a menu will open for the user to select data. Note that when a graph is added to a spreadsheet, and when the graph is selected, two new tab options, "Design" and "Format", appear

Figure 1: With the "Insert" tab selected, the various chart options can be seen displayed within the "Charts" group.

at the top of the spreadsheet. These tabs include options to change the design or the format of the graph or chart, and when the graph or chart is not selected, these tab options are not present.

After clicking on "Select Data", the menu displayed in Figure 2a will appear. Clicking on "Add" will bring up a smaller window (Figure 2b) where the "Series name" and "Series values" can be specified. Click in the "Series values" box, and remove the default "={1}" from the box. If the default "={1}" is not removed from the "series values" box, the graph will not display the intended data properly. A "Series name" can be added, but is not necessary to complete the graph. In the example seen in Figure 2b, the "Series name" would be the title for the data that is being graphed, or "ww%". After adding the appropriate "Series name" and "Series values",
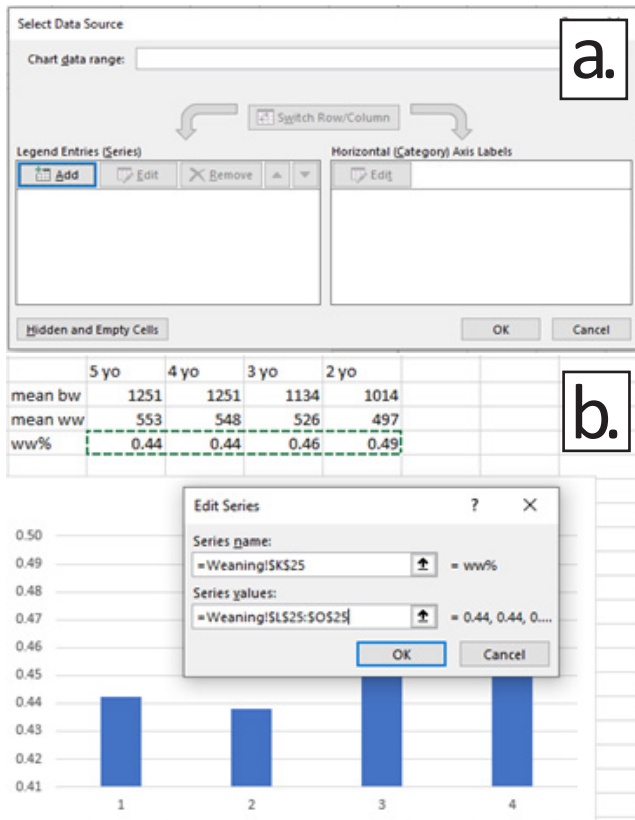
clicking the "OK" button will return the user to the menu seen in Figure 2a. Now, the data you wish to graph should appear as "Series1", or whatever the user entered as the "Series name", under the "Legend Entries (Series)" box. The horizontal axis labels for the data in the graph will appear by default as integers ordered numerically according the number of categories (Figure 2b). In Figure 2b, you can see that the horizontal axis is simply numbered 1 through 4. These horizontal axis labels are changed by clicking the "Edit" button under the "Horizontal (Category) Axis Labels" seen in Figure 2a. After clicking the "Edit" button, the user can select the labels that match the data in the graph. In the example in Figure 2b, these labels are the cells containing "5 yo" (i.e. 5-year-old), "4 yo" (i.e. 4-year-old), "3 yo" (i.e. 3-year-old) and "2 yo" (i.e. 2-year-old). The axis labels in a bar graph can only be modified by using the "Edit" button under the "Select Data" feature of the graph.

In the example in Figure 2b, the example data being used is data that could be collected during weaning and pregnancy examinations on a cow-calf operation. In this case, dam body weight, and weaning weight of calves have been collected, and the average dam body weight and calf weaning weight have been calculated using the =AVERAGE() formula for each age cow age category. The "ww%" was then calculated to determine on average, what percentage of their body weight each group of cows weaned. This is done by creating a formula that divides "mean ww" for each age category by "mean bw". These weaning weight percentages can then be graphed for each cow age category.

Figure 2: a.) The "Select Data Source" menu allows the user to manage the data series that will be displayed in the graph. b.) When adding a data series through the "Select Data Source" menu, the default "={1}" must be removed from the "Series values" box, then the data to be displayed by the bar graph can be selected.
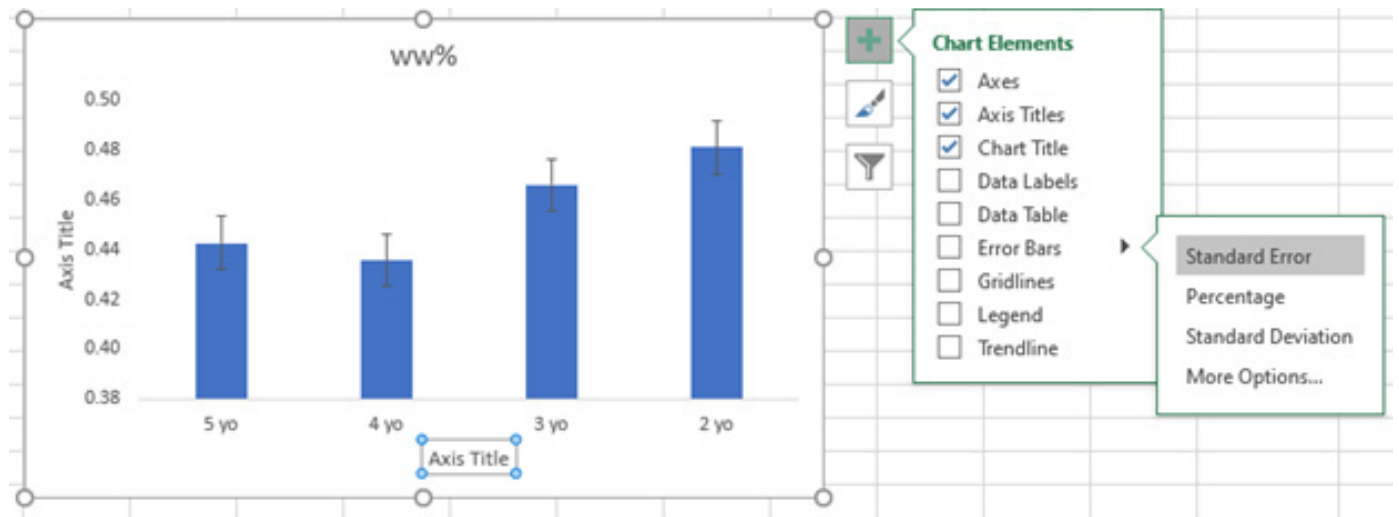
Figure 3: Chart Element options are displayed by clicking the green plus sign to the right of the chart. More options for each of the chart elements can be seen by right-clicking on any element (e.g. right clicking on the Axis Title box) and selecting the "Format" option. Axis Titles can be edited by clicking on the title text box and typing the desired title. Right-clicking anywhere on the chart will produce additional chart modification options. The "Design" and "Format" tabs at the top of the screen also offer the user options to modify appearances of the chart.

Creating a scatterplot:

Scatterplots are used to graph pairs of numerical data. A scatterplot can be useful in determining whether two variables are related to one another, or for identifying patterns between two sets of continuous data. One example is the relationship between weaning weight and age in days. To create a scatterplot, click on the "Insert" tab at the top of the spreadsheet, and go to the "Charts" group. Select the "Insert Scatter (X,Y) or Bubble Chart" option. Several options will appear in the drop-down menu shown in Figure 4; this article will use the simple "Scatter" plot. After clicking on the "Scatter" plot options shown in Figure 4, a graph will appear in the spreadsheet. Similar to when adding the bar graph, if data were selected prior to inserting the scatterplot, the newly inserted scatterplot will contain data. It is recommended that a blank scatterplot be

inserted. To do this, the user should make sure an empty cell anywhere in the spreadsheet is selected prior to inserting the scatterplot. This will ensure the graph is blank, and allow the user to begin building the scatterplot without having to first remove incorrect data.

Once the scatterplot is inserted, the process of selecting data to populate the chart is similar to that used to select the data included in a bar chart. Right-clicking on the empty scatterplot will produce a menu from which the "Select Data" option is available. A menu similar to that in Figure 2a will appear, and the user can add data series by selecting the "Add" button. For a scatterplot, both an x and a y data series must be added, compared to only an x data series when creating a bar chart.
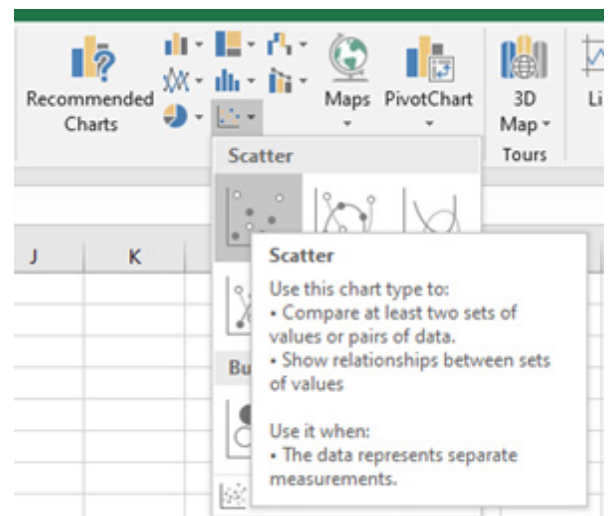

Figure 4: Menu options available when adding a Scatterplot to the spreadsheet.headings. data.

The example scatterplot shown in Figure 5 is a plot of calf age in days (x-axis) by calf weaning weight (y-axis). In this graph, each individual calf in the dataset is represented by an individual blue dot on the graph. Scatterplots are useful for observing the distribution of the data. For instance, in Figure 5, a large cluster of calves can be seen around the 200-day mark, with many of these calves being around 500 lbs. In general, the user can see that as calf age increases, weight increases as well. The dotted line shown at a diagonal through the data points is the trend line. This line, along with the equation and R2 values shown on the graph, can be added by right clicking on any of the blue dots in the graph and selecting the "Add Trendline" option from the right-click menu. After clicking on "Add Trendline", a menu will appear to the right of the screen similar to that used to modify elements of the chart. Several different options are available for formulating the equation of the line, but by default, the "Linear" option will be selected. The line shown in Figure 5 is a linear trendline, and the equation displayed on the graph is obtained by clicking the "Display Equation on chart" and "Display R-squared value on chart" options at the bottom of the Format Trendline menu. The equation of the line for a linear trendline follows the format below:

D	E	F	G
gStat	Gest(Mon)	Dry	Date_checked
	5	0	8/19/2020
	5.5	0	8/19/2020
		0	7/30/2020
	5.5	0	7/30/2020
	5.5	0	7/30/2020
	6	0	7/30/2020
	5	0	7/30/2020
	5	0	7/30/2020
	5	0	7/30/2020
	5.5	0	7/30/2020

$$y=m(x)+b$$

Figure 5: Scatterplot of weaning weight by calf age. A trendline, equation for the line, and R2 value have been added to the graph.

Where y is the y-axis value (i.e. weaning weight in Figure 5), m is the slope of the line (i.e. 2.73 in Figure 5), x is the x-axis value (i.e. calf age in Figure 5), and b is the y-intercept of the line (i.e. 6.15 in Figure 5). The slope of the line is particularly useful, as the slope is defined as the relative change in units of the y-axis variable for every 1 unit change in the x-axis variable. In other words, in Figure 5, the slope describes a 2.73 lbs. increase in the weight of the average calf in the group for every 1 day change in average age (i.e. the average daily gain of the group of calves is 2.73 lbs./day).

The R2 value displayed on the graph in Figure 5 also helps describe the data displayed in the scatterplot. The R2 value is the coefficient of determination, and describes the amount of variation in one variable that can be directly described by another variable. In Figure 5, the R2 value describes the amount of variation in the y-axis variable (i.e. weaning weight) that can be described by the x-axis variable (i.e. calf age). The coefficient of determination is measured on a scale of 0 to 1 and is often described as a percentage. In Figure 5, the R2 of 0.34 means that about 34% of the variation in weaning weight can be described by calf age.

The coefficient of determination is especially useful when making comparisons between similar datasets. The data displayed in Figure 6 is from the same dataset as that in Figure 5, only Figure 6 is limited to weaning data on calves born to 2-year-old dams, or first-calf heifers (blue dots) and 4-year-old dams (orange dots). This segregation of the data can be achieved in several different ways. In this dataset, the age of the dam of each calf was a separate column within a spreadsheet, and two different data series were entered using the "Select Data" feature of the scatterplot. The first series was only weaning weights and ages for calves from 2-year-old dams, and the second series was weaning weights and ages for calves from 4-year-old dams. In Figure 5, all

calves, without respect to the age of the dam, were included in only one data series.

Note that each data series in Figure 6 has a trendline, equation of the trendline, and R2 value. These are unique to each data series, and can be used to compare the two groups of calves. For example, it can be seen from Figure 6 that calves born to 2-year-old dams experienced an average daily gain (ADG) of 2.87 lbs./day, while calves born to 4-year-old dams experienced an ADG of 2.28 lbs./day. Also, roughly 45% of the variation in weaning weight among those calves born to 2-year-old dams can be explained by age, whereas only 18% of the variation in weaning weight can be explained by age in calves born to 4-year-old dams. These differences would not have been apparent had the data not been partitioned out by age of the dam. These data would suggest there were other factors influencing the weaning weight of those calves born to 4-year-old dams, making the effect of age less impactful on overall weaning weight. Other insights include that calves born to 2-year-old dams outgained those born to 4-year-old dams. These insights provide information to the veterinary practitioner about potential problem areas in health and reproduction that may warrant further exploration, while also providing good tools for communicating findings to the producer.

Creating a histogram:

Histograms are special cases where bar graphs are used to approximately represent the frequency distribution of numerical data. By convention, histograms are typically in the shape of a bar graph with no (or minimal) space between the bars. Histograms typically have no space between the bars, because each bar represents some category or continuum of data, commonly called a "bin", (i.e. weaning weights from 450 to 549 may be represented by one bar, and the width of the bar represents the range of the data contained by it). Newer versions of Microsoft Excel now have a built-in graphing function to create histograms located in the "Charts" group under the "Insert" tab. Clicking on the "Insert Statistics Chart" will give the user the option of inserting a histogram, pareto, or box-and-whisker plot. If inserting a histogram using this method, the

Figure 6: Scatterplot of weaning weight data limited to only calves from 2-year-old dams (i.e. heifers; blue dots) and 4-year-old dams (orange dots). The upper left equation and R2 value describes the data on calves from 2-year-old dams, while the lower right equation and R2 value describes data on calves from 4-year-old dams. Notice the difference between the slope and R2 values for each line.

process is similar to that of adding a bar chart or scatterplot. The "Select Data" function is then used to add a data series, and the histogram will automatically set the size of the bins displayed on the x-axis. Bin size can be modified by right-clicking on the axis, and selecting "Format axis".
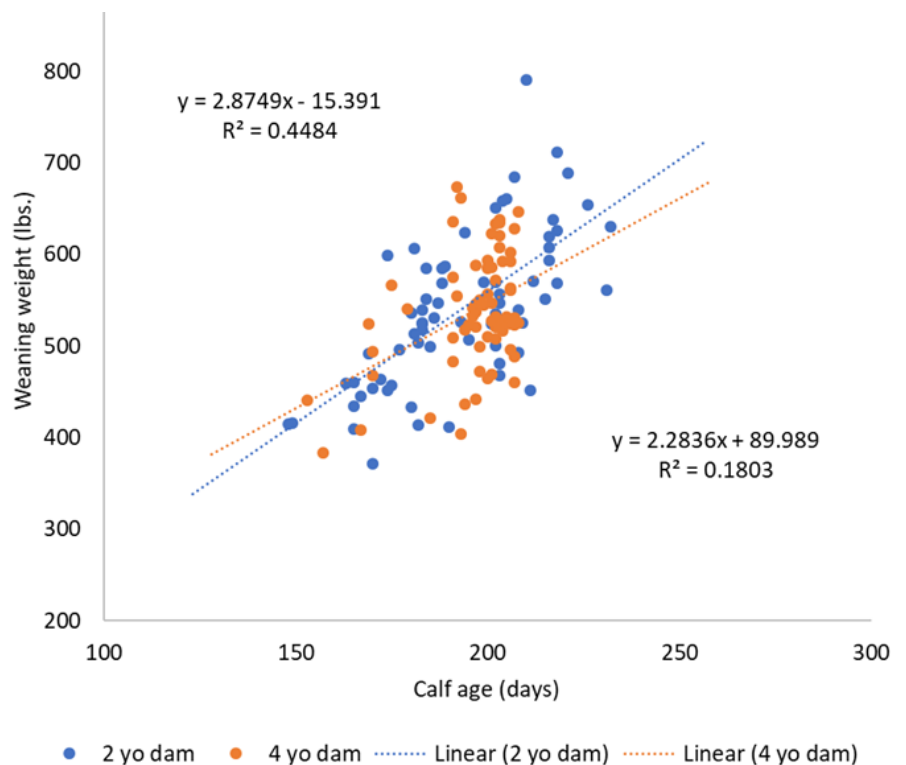
Alternatively, the user can create a histogram by using the =FREQUENCY formula. In this method, the user must first decide the minimum bin and increment values of interest (e.g. first bin starts at 400 lbs., each bin represents an increment of 50 lbs.). Once the bins have been established, the =FREQUENCY formula is used to determine how many observations within the dataset fit into each bin. Figure 7 shows the steps in developing a frequency distribution of weaning weights. In Figure 7a, the bins of interest are established. Notice that the formula shown references the "50" in cell T96. This cell is locked using dollar signs ($) so that when the formula is copied down, the reference to the cell containing "50" is not moved. This method of building the bins allows the user to change the bin increment, without needing to redo the bins (i.e. the 50 could be changed to 25 instead, and the bins would automatically update). In Figure 7b, a second formula is used to complete the range that will serve as the labels for the x-axis of the histogram. Figure 7c displays the use of the =FREQUENCY formula. The =FREQUENCY formula requires two components: the data array (i.e. weaning weights in this
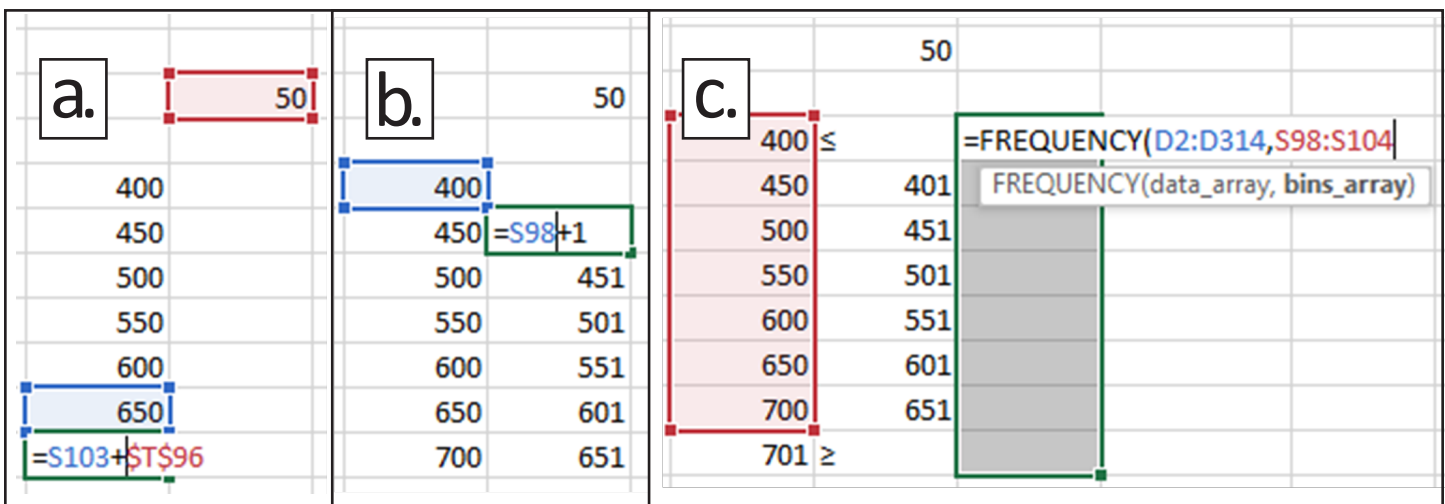

Figure 7: Steps in developing a frequency distribution of weaning weights. a.) Establishing bins in 50lbs. increments; basing the bins on the formula copied will allow automated updates if bin size is changed. b.) the second of values formed completes the ranges that will serve as labels for the x-axis of the histogram. c.) the =FREQUENCY formula requires the array of data (i.e. weaning weights in this example), and the bins (highlighted in red). Note that while completing the =FREQUENCY formula, the same number of cells plus one additional cell below the last bin must be selected. When the formula is completed, the user must press Control + Shift + Enter to produce results.

example) and the bins array (i.e. bins created in Figure 7a). Note that when entering the =FREQUENCY formula, the same cells parallel to the bins are selected, plus one additional cell below the last bin (i.e. 8 total cells are selected in Figure 7c, including the cell where the =FREQUENCY formula is being entered, whereas only 7 bins are included; the cell containing "701" below the last bin is included for labeling purposes).

When the =FREQUENCY formula is completed, the user must press the Control + Shift + Enter keys in succession. Pressing the Enter key only will not complete the formula. Once completed, the =FREQUENCY will produce the data series highlighted in red in Figure 8. Graphing this data series in a bar graph will result in the graph shown in Figure 8. Editing the "Horizontal (Category) Axis Labels" in the "Select Data" menu of the bar graph will allow the user to edit the x-axis labels to reflect the bins created previously.

Once the frequency distribution data series has been added to the graph, the graph elements can be modified by selecting the graph, right-clicking on it, and selecting "Format Chart Area" from the right-click menu. Histograms are characterized by a lack of space between bars in the graph. This lack of space graphically displays the range of values encompassed by each bin category.
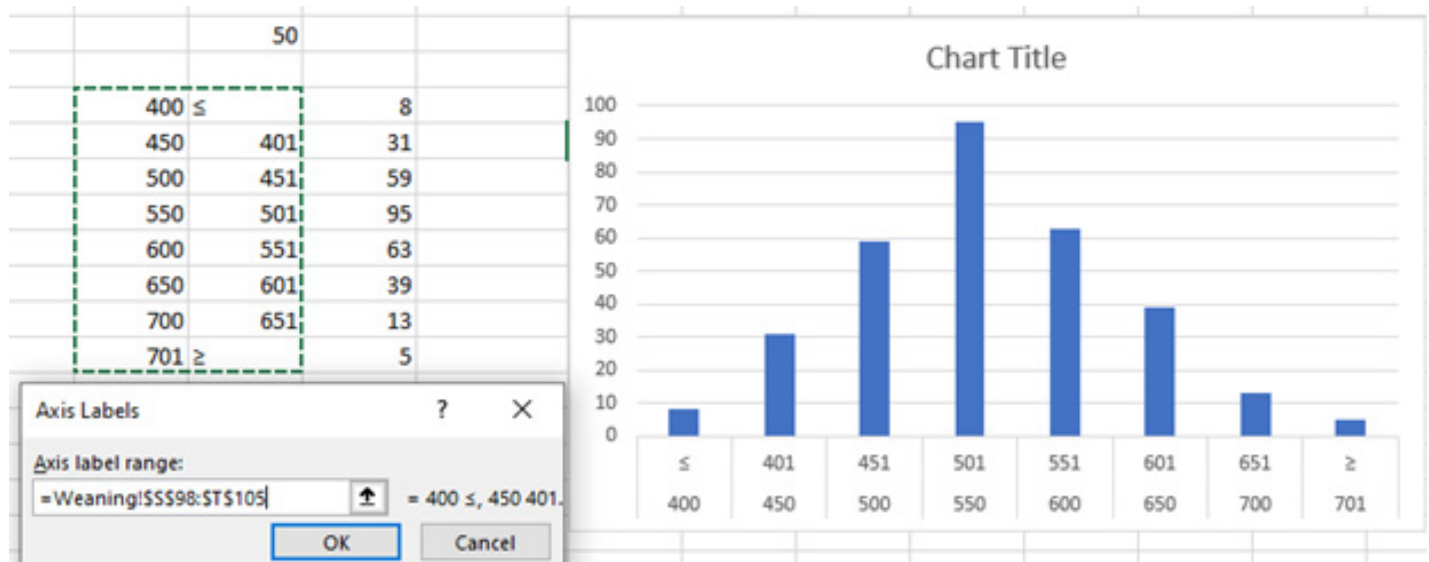
Figure 8: When selecting the "Horizontal (Category) Axis Labels", the bins used in the =FREQUENCY formula as well as the ranges established in the adjacent column must be highlighted. This ensures the x-axis of the graph shown above is labelled properly.

Figure 9 shows a competed histogram with the space removed from between the bars. Space was removed by right-clicking on any of the bars in the graph, , selecting the "Format Data Series" option, and decreasing the Gap Width to 0 in the "Series Option" menu. selecting the "Format Data Series" option, and decreasing the Gap Width to 0 in the "Series Option" menu.

Conclusion:

Graphs are useful for describing the distribution of data, identifying patterns or trends, and communicating findings. Bar graphs, scatterplots, and histograms are three types of graphs that are particularly useful to veterinary practitioners. Determining the type of graph needed to properly display the data of interest can assist the veterinary practitioner in their exploration of an animal health or production problem, and offer a compelling method of communicating the data.
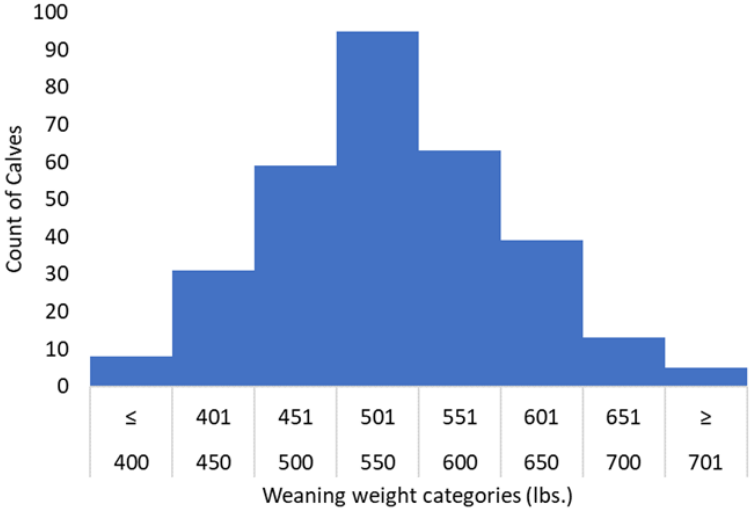

Figure 9: Completed histogram displaying an approximate normal distribution of calf weaning weights. Because each weight category (i.e. bin) represents calves that range from the high to low ends of the bin, the width of each bar in the graph extends the width of each bin.