Visualizing Data

In business, you may be called upon to give presentations in which you present the results of specific analyses or present data to stakeholders. Often times, you are given the data in raw form. It is important to know how to organize this data so that you can then choose the appropriate visual form for presentation. Why should you be concerned with visualizing data? Well, the topics you will learn about in this course will make you accustomed to the terminology of statistics. However, when you are presenting data in a business setting, not all persons in the room will have the same level of knowledge. But most will be able to look at a chart, a graph, or some type of visually and have an “Ah Ha!” moment when they can SEE the big picture.

In this section, we will look at creating the following in Excel:

1. Frequency Distribution and Histogram
2. Time Series Graph
3. Pareto Chart

**Frequency Distribution and Histogram**

**Histogram**

* is a bar chart for grouped numerical data in which you use vertical bars to represent frequencies or percentages in each group.

**Frequency Distribution**

* summarizes numerical values by tallying them into a set of numerically ordered classes.
* To create a useful frequency distribution, you must think about how many classes are appropriate for your data and also determine a suitable width for each class interval.

In general, you should have at least 5 classes and no more than 15. To determine the class interval width, we’ll use the following formula:

**Interval Width = (higher value – lowest value) / number of classes**

The data below shows the sales (in thousands) of 39 U.S. National Parks in Column A. We have decided to use 6 classes, and the class boundaries are provided as well.

* Construct a frequency distribution and histogram for the data using 6 classes.
1. Insert the following data into column A in Excel

|  |
| --- |
| Sales |
| 5 |
| 13 |
| 27 |
| 28 |
| 36 |
| 41 |
| 46 |
| 52 |
| 52 |
| 61 |
| 64 |
| 66 |
| 70 |
| 75 |
| 77 |
| 77 |
| 94 |
| 106 |
| 132 |
| 143 |
| 169 |
| 183 |
| 196 |
| 217 |
| 220 |
| 233 |
| 233 |
| 236 |
| 265 |
| 308 |
| 338 |
| 402 |
| 462 |
| 505 |
| 520 |
| 539 |
| 650 |
| 760 |
| 775 |

1. Enter the class upper boundaries for the first seven classes in cells B2:B8. The numbers are 133.50, 262.50, 391.50, 520.50, 649.50.
2. Click on the DATA tab. In the analysis group, click on Data Analysis.
3. Select Histogram and click OK.
4. Select the input range (Acreage data).
5. Select the bin range (B2:B8); Excel will add one class.
6. Select the output range (any unused cell on the worksheet)
7. Select the chart output (look in the lower left corner).
8. Click OK.
9. Edit the chart title and axis labels. To eliminate the spaces between the bars, right click on one of the histogram bars and select “Format data series…”, then move the gap width slider all the way to the left.
* Your final result should look like this:



**Time Series Graph**

Times-series graph

* plots the values of a numerical variable on the Y axis and plots the time period associated with each numerical value on the X axis.
* A time-series plot can help explore trends in data that occur over time.

Using the time series sheet, we are provided with sales for a 5 year period. Construct a time series graph for the data.

1. Insert the following data into Column A

Year
2001
2002
2003
2004
2005

1. Insert the following data into Column B

Sales
150,000
200,000

300,000
100,000

50,000

1. Highlight the data in both columns
2. Click Insert, the Scatter, then choose Scatter with Straight Lines and Markers
* Your output should look like this



As a business person, we would interpret this as having strong sales as we approached 2003 but a visible dip in sales during 2004 and 2005. As a manager, your response is “WHY and HOW do we fix it?”

**Bar Chart**

* compares different categories by using individual bars to represent the tallies for each category.

**Pie Chart**

* uses parts of a circle to represent the tallies of each category.

**Pareto Chart**

* In a Pareto Chart, the tallies for each category are plotted as vertical bars in descending order, according to their frequencies.
* A Pareto Chart can reveal situations in which the Pareto principle occurs.

**Pareto Principle**

* exists when the majority of items in a set of data occur in a small number of categories and the few remaining are spread out over a large number of categories.
* These two groups are often referred to as the “vital few” and the “trivial many”. With a Pareto Chart, you can separate these so you can focus on the important categories.

|  |
| --- |
| **Sample Data Table:Types of Errors Discovered During Surgical Setup** |
| **Error Type** | **Frequency** | **Percent** | **Cumulative %** | **Cut off** |
| Wrong Supplier | 67 | 36.8% | 36.8% | 0.8 |
| Excess Count | 45 | 24.7% | 61.5% | 0.8 |
| Too Few Count | 35 | 19.2% | 80.8% | 0.8 |
| Wrong Size | 10 | 5.5% | 86.3% | 0.8 |
| Wrong Sterile Instrument Set | 9 | 4.9% | 91.2% | 0.8 |
| Missing Item | 8 | 4.4% | 95.6% | 0.8 |
| Damaged Item | 6 | 3.3% | 98.9% | 0.8 |
| Other | 2 | 1.1% | 100.0% | 0.8 |
| Total | 182 | 100% |  |  |





In this example presenting the data using a Pareto Chart, displays how three (3) of the error type account for 80% of all errors: **Wrong Supplier, Excess Count, and Too Few Count.**

**Getting the Data Ready**

**Step 1**: Prepare the Data.

In order to make the Pareto Chart in excel, first you must have the data ready. Once we have the values for each cause, in the example above Surgical Setup Errors we need to calculate the data for the percent column and the cumulative percentage column.

1. To calculate the percentage column values divide each error type by the total of all the error types.
2. To calculate the Cumulative Percentage column in cell D9, type +C9, then move down and in cell D10 type +D9+C10, in D11 type +D10+C11, and repeat this process through cell D16.



**Step 2**: Add Cutoff.

* Add a column in your data table titled “Cutoff” and insert the value .8 for all error categories.
* Your table should look like this now:



**Making the Chart**

**Step 1:** Select the frequency data by highlighting cells B8 through B16 above and click on the “Insert” tab above and then in the “Charts” menu click on “Column” and under 2-D Columns select “Clustered Column.” (Hover over column boxes to view type). Click on the legend and then right click and select “Delete” to remove the legend.



**Step 2:** Add labels to the horizontal axis.

* Click on the chart and click on “Chart Tools” and “Select Data”. On the Horizontal “Axis Labels” side of the pop-up box click on “Edit” and highlight cells A9 through A16. Click “OK” and “OK” again.



**Step 3:** Resize the labels so that they are easy to read.

* Right click on the horizontal axis and select “Font” and change the size to a size that allows the labels to be viewed as shown here. In this example the size is 6. Click “OK.”



**Step 4:** Add the Percentages.

* Click on the chart and click on “Chart Tools” above and click on “Select Data”. In the “Select Data Source” pop up box in the left hand side under ‘Legend Entries” click on “Add” and in the popup box under “Series Values” highlight cells D9 through D16 and click on “OK” and then press “OK” in the “Select Data Source” pop up box. Click “OK” and “OK” again.



**Step 5:** Format percentages to display on the secondary axis.

* Right click on the red bars you just created and select “Format Data Series.” With “Series Options” selected on the left, in the “Plot Series On” box (on the right) click on “Secondary Axis.” Click “Close.”



**Step 6:** Format line.

* Right click on the red bars and select “Change Series Chart Type” and with Line Type selected on the left, click on the “Line with Markers” box on the left. (Hover over box to view line type). Click “OK.”



**Step 7:** Make sure the secondary axis is formatted correctly to 100%.

* Right click on the secondary axis and select “Format Axis” and under “Axis Options” under Maximum change the fixed number to 1 and under “Number” change decimal places to 0 and click on “Close”. (If you have trouble with this, make sure that the box “Linked to Source” is NOT checked off.



**Step 8:** Add a title.

* Add a title to the graph by double clicking on the graph and under “Chart Tools” under the “Layout” tab select “Chart Title” and click on “Above Chart”. Edit the text to say “Surgical Setup Errors by Type” and change the font size by right clicking on the title, select font and select font size. In this example the font size is 10.



**Step 9:** Add the cut off line by double clicking on the chart and under “Chart Tools” click on “Select Data” and Click on “Add” under the “Legend Entries” side of the pop up box and under “Series Values” highlight cells E9 through E16 and then click on “OK” and in the “Select Data Source” box click on “OK” and “OK” again.



**Step 10:** Format the cutoff line.

* Right click on the line you just created and select “Format Data Series” and under “Marker Options” select “None”. Under “Line Color” select “Solid Line” and change the color to “Black”. Lastly, under “Line Style” change the width to 2pt. and change the “Dash Type” to “Round Dot”.
* Your graph should now look like this.



**Step 11:** Highlight 80% of errors.

* Change the color of the bars to grey for the categories Wrong Size to Other which will help to highlight the 3 error types that account for 80% of the errors. Double click on the graph and then double click on the bar for “Wrong Size” and click on the bar one more time to highlight only that bar now right click and select “Format Data Point” and under “Fill” select ‘Solid Fill” and change the color to grey. Follow these steps to change the bar color for the remaining bars.
* Now your chart should look like this.



**Step 12:** Final touches. We like to keep things as simple as possible, so we want to simplify the line by eliminating the boxes on it.

* Right click on the line and select “Format Data Series.” Select “Marker Style and either make the markers smaller or select “No Marker” like we did. Use the “Chart Layout” and “Axis Titles” to the primary and secondary axis.
* And remember, just like the chart title, if you can “click it” you can edit it. We simply added labels and then clicked inside the labels to edit the text.
* You can also click on the labels and drag them to the position you want. We place the labels on top of the axis so that they are easy to read.
* If your chart shifts, simply click anywhere in the chart until “Plot Area” comes up and then you can drag and resize your chart to the proportions that look best for displaying the data.
* We also lightened the grid lines by simply clicking on them and then right clicking, select “Format Gridlines” and changed the color to a very soft gray.

