



EvalFest How-Tos for Graphical Exploratory Data Analysis in Excel

Graphical Exploratory Data Analysis in Excel

These instructions build on videos in our Data Management series. In the Database Organization and Password Protection and Data Cleaning and Recoding videos, we discuss how to prepare your data for analysis. And the Pivot Tables and Basic Analysis video demonstrates how to compute basic frequencies for each answer category associated with a variable (or item).

Now you might be thinking, why bother? Looking at the data in various graphical forms is another way of ensuring that the data your data were collected and coded as expected. If you produce a graph of the data that is very unexpected, you will want to go back to the raw data to try to determine what is wrong.

This step is really about looking at the data in a graphical format to ensure that nothing is amiss. If you want more information about formatting graphs or creating other graphs, we suggest Stephanie Evergreen's blog, 'How to Build Data Visualizations in Excel' (<http://stephanieevergreen.com/how-to/>) and Ann K Emery's online tutorials (<http://annkemery.com/tag/tutorials/>) as good resources.

Exploring the data graphically

For now, we're still getting to know our data, and so let's get started.

1. Single Bar graph
 - a. Let's look at a single bar graph for the variable Gender.
 - i. Bar graphs are typically used to display categorical data.
 - b. First, you will create a table of counts, or frequency table, from which to create the bar chart. We will use the COUNTIF function to create the table.
 - c. You can go to an empty area of the spreadsheet, or even a different worksheet if you'd like. Type in or copy your category names.
 - d. In the cell next to the first category name, begin to type =COUNTIF. Choose COUNTIF from the list.
 - e. First, highlight the range of cells you want Excel to search and, subsequently, count. These are in the worksheet that contains the data.
 - f. Type a comma and then click on the cell that contains the first category name.
 - g. Press Return (or Enter).
 - h. The cell should populate with the count of the category, in this case, Males.

- i. You can drag the small green square to the cells below to produce counts for the other category, in this case Female.
 - j. To create the bar graph, highlight the category names and corresponding counts.
 - k. Go to the Insert tab.
 - l. Scroll over to the middle of the menu options and click on the arrow next to the bar graph.
 - m. Select the first option, the basic 2D bar graph.
 - n. Just as with numerical summaries, graphical displays allow you to see the data to ensure that nothing seems off. For example, if we knew that this event was attended by almost all women, this graph wouldn't make sense. If your graph appears very unlike you would expect, you would need to take a closer look at the data before you run any statistical tests.
 - o. Because we are simply exploring and getting to know our data, we don't worry about formatting our display at this stage. We recommend reformatting a display such as this before including it in a presentation or report.
2. Side-by-side bar graph
- a. Just as with a single variable bar graph, you will have to create a table before creating a side-by-side bar graph. For this graph, you will need to create a pivot table, as explained in the Pivot Tables and Basic Analysis video.
 - b. Let's look at OverallCode again, but this time, by Gender. You will need to first create the pivot table.
 - c. Once you have the pivot table created, click on any cell in the table.
 - d. Go to the Insert tab.
 - e. Scroll over to the middle of the menu options and click on the arrow next to the bar graph.
 - f. Select the first option, the basic 2D bar graph.
 - g. Again, look for patterns in the graph that you may not have expected. Does everything look okay? Should you proceed? These are questions you want to ask at this step.
 - h. As you can see, Very Good (4) and Excellent (5) are the most frequent ratings, regardless of Gender. If we think back to the mean, median, and standard deviation we computed for OverallCode in the Numeric Exploratory Data Analysis video, these values align with what we see in this graph. The mean of 4.388 is less than the median of 5, the standard deviation is relatively small (.7586), and the data are left skewed, showing most of our data at the upper end of the rating scale.



UNC
MOREHEAD PLANETARIUM
AND SCIENCE CENTER



karen peterman
CONSULTING

