Print

Packed Bar Charts and Data Filtering

As you learned in the previous videos, sorted bar charts and Pareto plots are useful when you want to explore categorical variables with many levels.

Pareto plots were developed to make it easier to focus on the biggest issues, or signals, through the noise of all of the smaller issues.

A more modern and flexible alternative to Pareto plots is packed bar charts. In this packed bar chart for the engine defect data, the three largest categories really stand out. The smaller categories, which aren't of primary interest, are gray.

Look at the gray shaded bars. You can see that there are many smaller defect categories, although they are not labeled.

Now, look at the lengths of the bars for the three largest issues compared to the combined length of the bars for all of the other issues.

Clearly, if you can address these three defect categories, you would eliminate a majority of the engine defects that occur.

Notice that, when you change the structure and add labels, the packed bar chart conveys essentially the same information as a Pareto plot. But a packed bar chart is more compact and efficient than a Pareto plot, and there's less unused white space in the graph.

In this example, we have only 21 defect categories.

Consider an example involving hundreds or even thousands of categories. For instance, how would you graphically display sales information for hundreds of products or customers in one graph?

Here's an example involving the top 500 companies in the S&P 500 (or, the S&P). The data are from March,

2017. The S&P is a report of the market value (or market capitalization) for the top 500 companies in the US stock market.

You can easily see the 12 largest companies in the S&P. You also get a rough idea of the overall market capitalization that comes from these companies, relative to the entire S&P.

For this packed bar chart, we have one categorical variable, Company.

Each company is grouped into one of twelve market sectors. How can we see which companies in the S&P fall into the different sectors?

One technique is to use the packed bar chart with a data filter.

You can use a data filter to stratify a graph or an analysis by the values of other variables in the data set.

For example, here you can see a packed bar chart for only the companies in the Consumer Discretionary sector. There are 82 companies in this sector, and the largest by far is Amazon.

There are only five Telecommunications companies in the S&P. Of these, AT&T and Verizon are by far the biggest.

Notice that, for these two sectors, the scale of the X axis, Market Cap in billions, is different.

The market cap for Amazon is approximately $400 billion US dollars, whereas the market cap for AT&T is approximately $250 billion.

How can we see Company, Sector, and Market Cap, all in one graph?

We can use a type of graph called a tree map. You learn about tree maps, and revisit mosaic plots, in an upcoming video.

In the next JMP demonstration videos, you learn how to create packed bar charts. You also learn how to use the local data filter to update a graph or analysis based on values of other variables.

Close