Print

Tree Maps and Mosaic Plots

In previous videos, you learned about packed bar charts.

You can use packed bar charts when you have many levels of a categorical variable.

In the S&P 500 example, there are 500 companies, grouped into 12 sectors.

In this packed bar chart, the bars are sized by market capitalization in billions of US dollars.

In an earlier video we asked, "How can we see Company, Sector, and Market Cap all in one graph?"

Company and Sector are categorical, and Market Cap is continuous.

One option is to create a mosaic plot.

Recall that a mosaic plot is a special type of stacked bar chart for exploring the relationship between two categorical variables. But mosaic plots aren't the best choice when you have a lot of levels of your variables.

Here's a mosaic plot for the S&P data, where the bars are sized by Market Cap. With 500 companies, it's difficult to learn much from this graph.

An alternative for graphing two categorical variables with many levels is a tree map. In this tree map, the size of the colored boxes represents the Market Cap for the particular company.

When you color the tree map by Sector, you can see which companies fall into each sector.

You can see that four of the six largest companies are in the Information Technology sector. These are pink (or fuchsia). But this tree map is a little hard to read, and it can be difficult to distinguish between the different colors.

Each company belongs to, or is nested within, one sector.

A nested tree map enables you to easily see the largest sectors in terms of Market Cap.

You can also see the number of companies per sector and the biggest companies within each sector.

Information Technology, Financials, and Health Care are the largest sectors contributing to the overall market capitalization.

Real Estate, Materials, and Telecommunications are the smallest.

Let's look at an industrial example. Remember that the data set Scrapped Parts.jmp includes information on 811 batches with one or more scrapped parts per batch. The data table includes the following information: the number of pieces scrapped per batch, the product line, the product family, and the total value of the scrapped parts per batch.

This is a mosaic plot for the number of pieces scrapped by Product Line and Product Family.

A majority of the scrapped parts were from the Small product family.

Of these, most were from product lines A2 and A3.

Notice how this picture changes when the mosaic plot is based on total value instead of pieces! From a value perspective, large parts from product line B2 is clearly the biggest issue.

If you want to see Product Line, Product Family, the number of pieces scrapped, and the total value of the scrapped parts in the same graph, you can use a tree map.

Earlier, you saw a nested tree map.

In this example, the product lines aren't unique to the different product families, so you won't use a nested tree map. But a tree map is equally effective in displaying these multidimensional data.

In this tree map, the boxes are sized by the total value, and they are colored by the number of pieces.

In this one graph, you can see that B2 accounts for the highest value (it is the largest box), and that for the Small product family, A2 and C1 had the most scrapped parts (these are the darkest in color). You explore this scenario more in a practice exercise.

Earlier, you learned about bar charts, packed bar charts, and Pareto plots for graphing one categorical variable at a time.

In this video, you learned how to use tree maps to plot two categorical variables at one time. You also learned that tree maps are an alternative to mosaic plots when you have many levels of your categorical variables.

You see how to create tree maps in a JMP demonstration video. Note that there are many other options for graphing categorical data.

For a more comprehensive list, see the Read About It for this module.

Close