OakenDoor

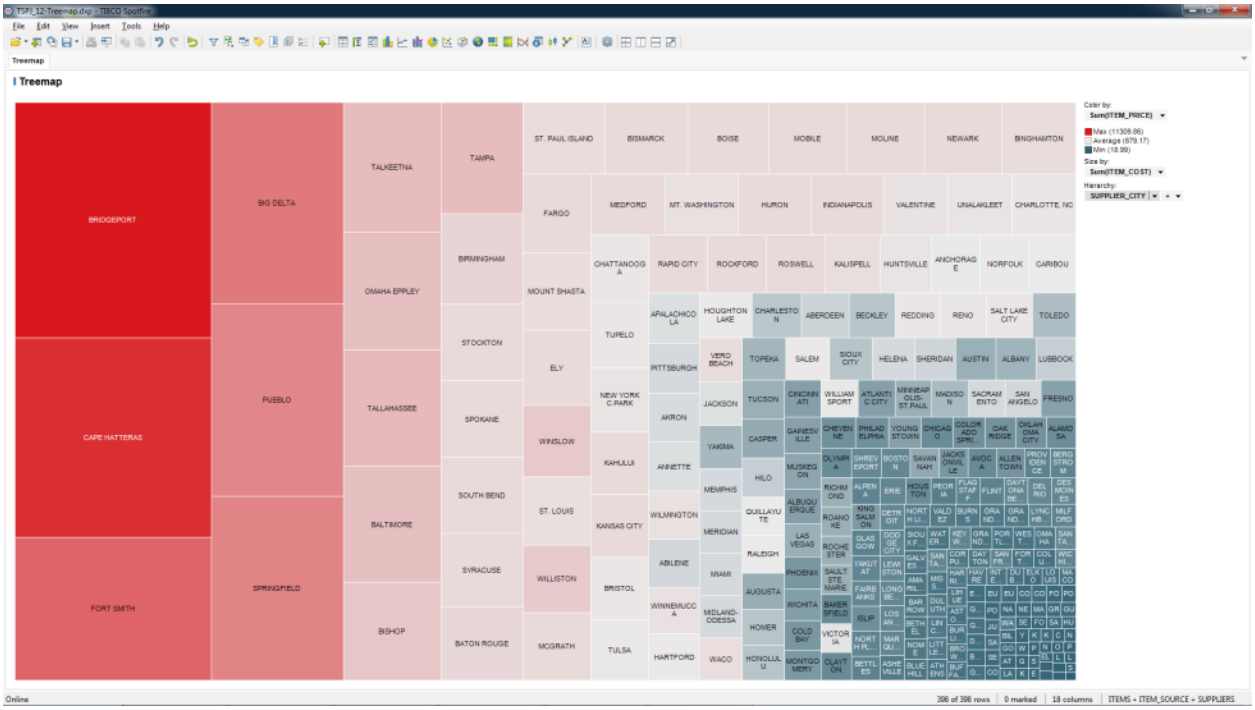
Chart Analysis

Shaun McNamara

2022

Table of Contents

Tree Map

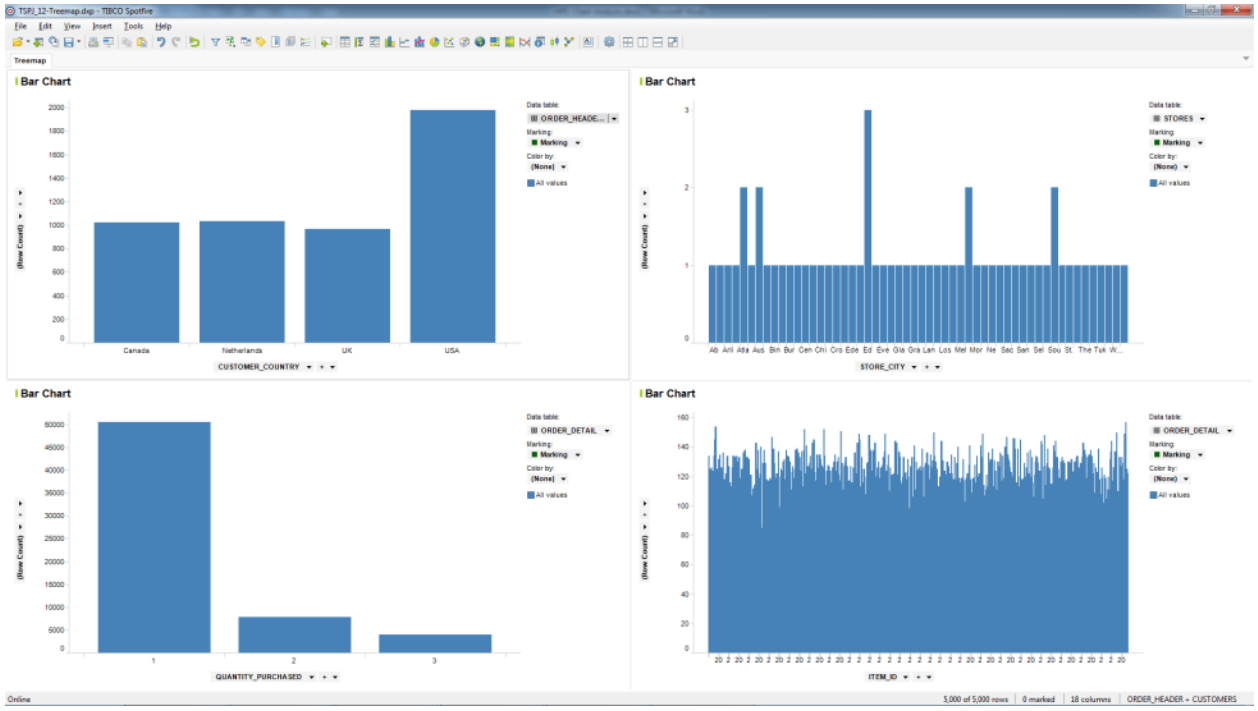


Tree maps are a tool for hierarchical data analysis invented by Ben Schneiderman. Basically tree maps segment a rectangular area into smaller rectangular areas and encode three quantitative variables: one using the height, one using the width of each smaller segmented rectangles, and of course, color.

Separable visual dimensions allow us to view some pairs of visual attributes independently, but others only as a whole. We can easily see the area and the color of objects independently, quickly finding all of the large rectangles, all of the blue rectangles, or all of the large blue rectangles in a display that consists of red and blue rectangles of varying sizes. Other pairs of attributes are called “integral” visual dimensions. In this case, we tend to perceive the dimensions holistically, not independently. The lengths and widths of rectangles are perceived holistically as their areas. If we try to find all of the tall rectangles in a display, our eyes will be drawn to the tall rectangles with the largest area, even though there might be other rectangles of similar heights that we have trouble noticing because they have small widths. In order to search for differences in only width or only height, we’re required to work harder and spend more time than we would if focusing on the differences of two separable visual dimensions. This problem might not seem significant in a tree map with just a few rectangles, but as the amount of data increases, it becomes more and more of a problem. That being said, some tree maps successfully encode more than 1 Million items across two variables.

After a user views the tree map as a whole, they'll probably want to make individual comparisons between specific boxes. However, accurate comparisons of area are not something that humans do well. Comparative information cannot be accurately perceived without spatial separation and further orientation. Basically, it's difficult to accurately make comparisons of the width or height of boxes that are not arranged next to one another along a common baseline.

Comparative Bar Charts

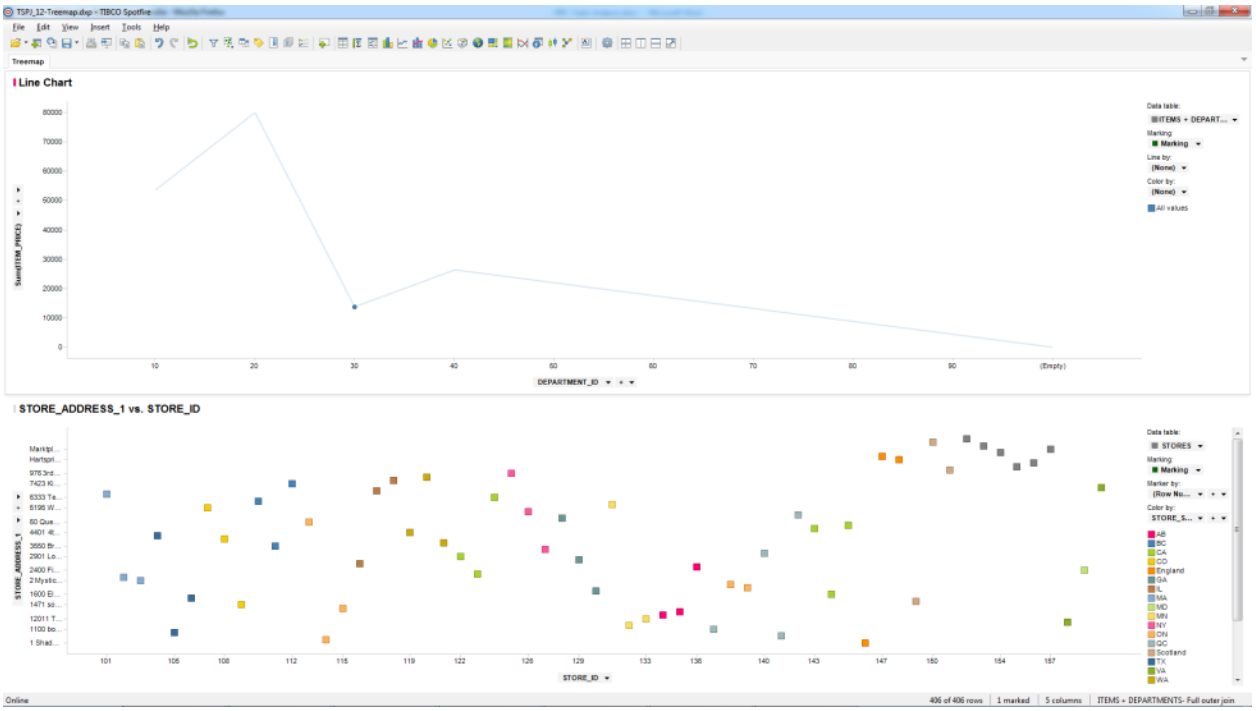


In this case, comparative bar charts provide a simple redesign and rely on humans' perceptual strengths instead of perceptual weaknesses.

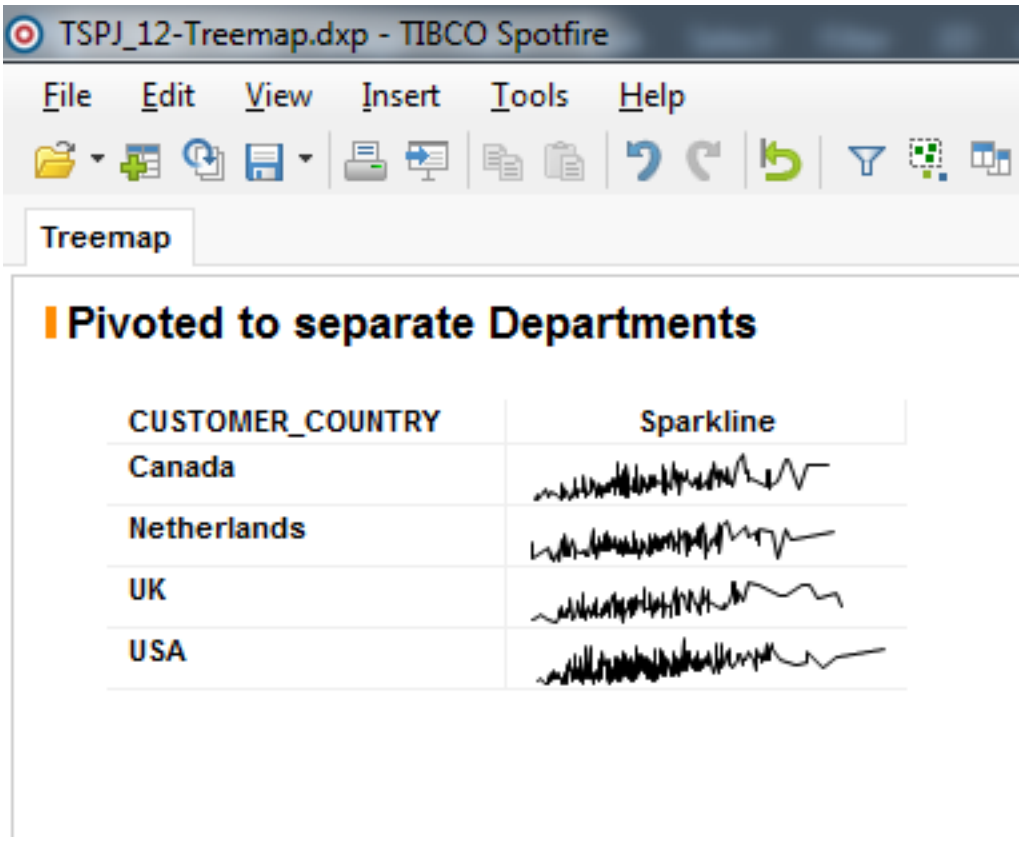
Features in categorization jump out immediately – customer country compared to store city. The difference would not have been readily apparent in the original. For our purposes it would be easier to make comparisons between sites in a country for a study.

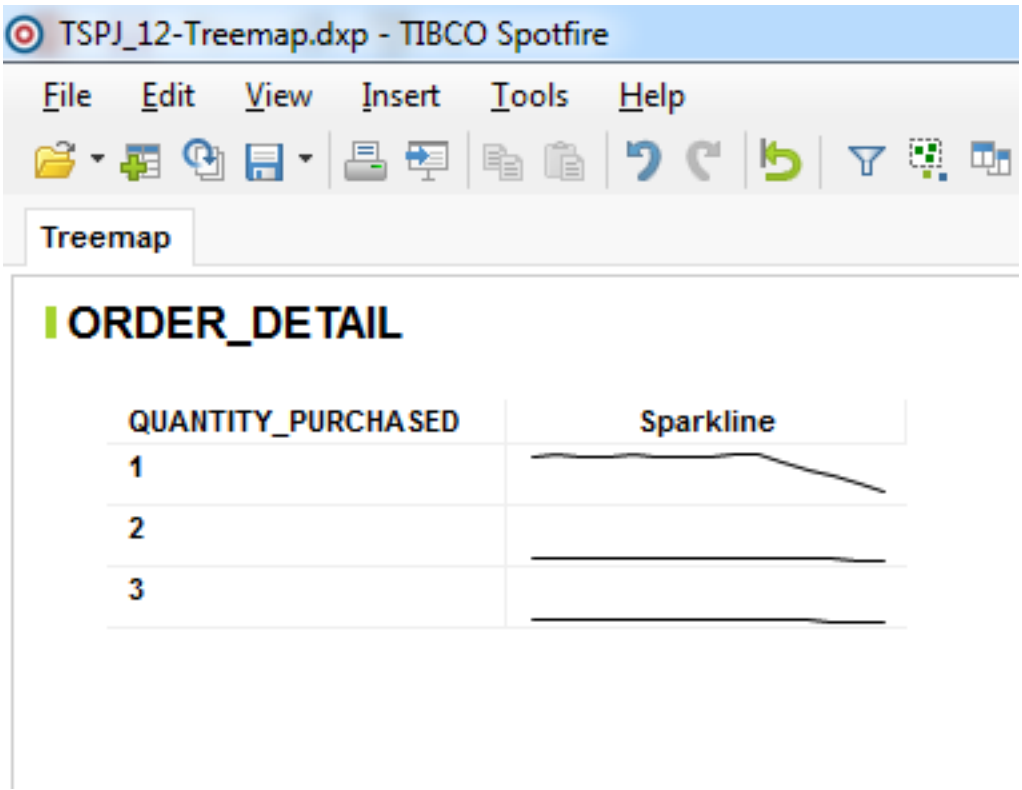
Line Charts, area charts, scatter plots, and sparklines

Trending over time is best perceived using these types of charts. One advantage in line charts over bar charts is that line charts do not need to start at zero. For instance, if one bar represents a value of 10 and another 9, when the bars start at zero we can see that this is a relatively minor difference. However, if the bars start at 8, then the one that represents 10 would appear to be twice as high as the one that represents 9. Line charts and scatter plots are fine for this, but with bar charts the length of the bar encodes its value.

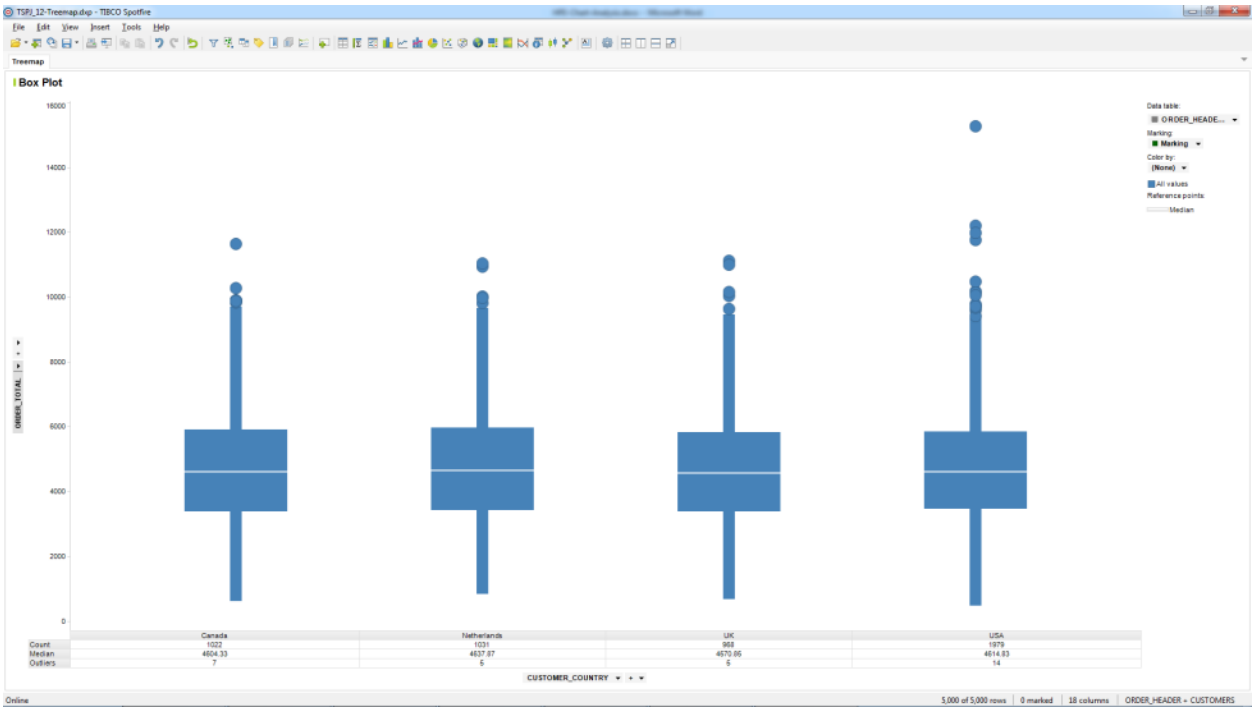


Sparklines are useful for quick doses of trending over time that then lead to greater analysis of other trending charts during a data deep dive. For example, Risk Score “Sparklines” present the general shape of the variation (over time) in a simple, highly condensed way. It is a simple design that sparks a user to action. They are highly effective in this context. While a typical chart is designed to show as much data as possible, and is set off from the flow of text, sparklines are intended to be succinct, memorable, and located where they are discussed. Often times textual content is more accurately remembered because the small graphic/visual cue is also embedded in the mind.


TSPJ_12-Treemap.dxp - TIBCO Spotfire
File Edit View Insert Tools Help
Treemap
Pivoted to separate Departments
CUSTOMER_COUNTRY
Sparkline
Canada
Netherlands
UK
USA


TSPJ_12-Treemap.dxp - TIBCO Spotfire
File Edit View Insert Tools Help
Treemap
ORDER_DETAIL
QUANTITY_PURCHASED
Sparkline
1
2
3

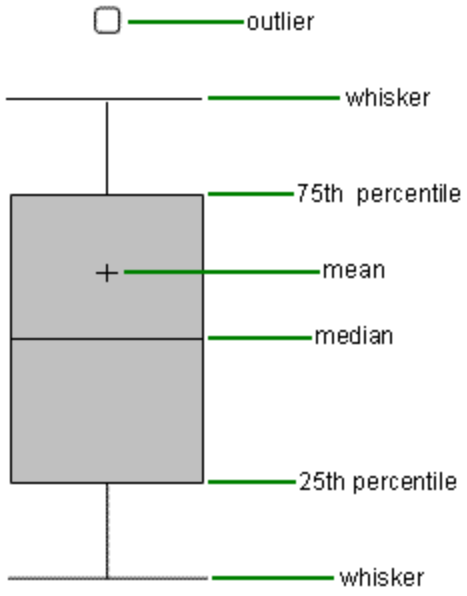
Box Plots



Box plots are usually associated with descriptive analytics, and as such they are highly detailed, have sections and points that are easily recognizable, and are non-parametric – meaning they don't assume the data has a characteristic structure. For example, box plots could be used to show how similar workloads are completed by various CRAs or a single CRA. They can also be used for rating/ranking sites within a study. Parametric chart structure would be better for what we know of risk score vs. workload score or risk over time.

While box plots can have a learning curve, the key to understanding box plots lies in understanding their quartiles – the three points that divide the data set into four equal groups of a ranked set of data values. A simple representation, as shown next, can quickly raise a user's understanding and open up the rich information that box plots have to offer. Another useful element of a box plot is the perception of outliers.

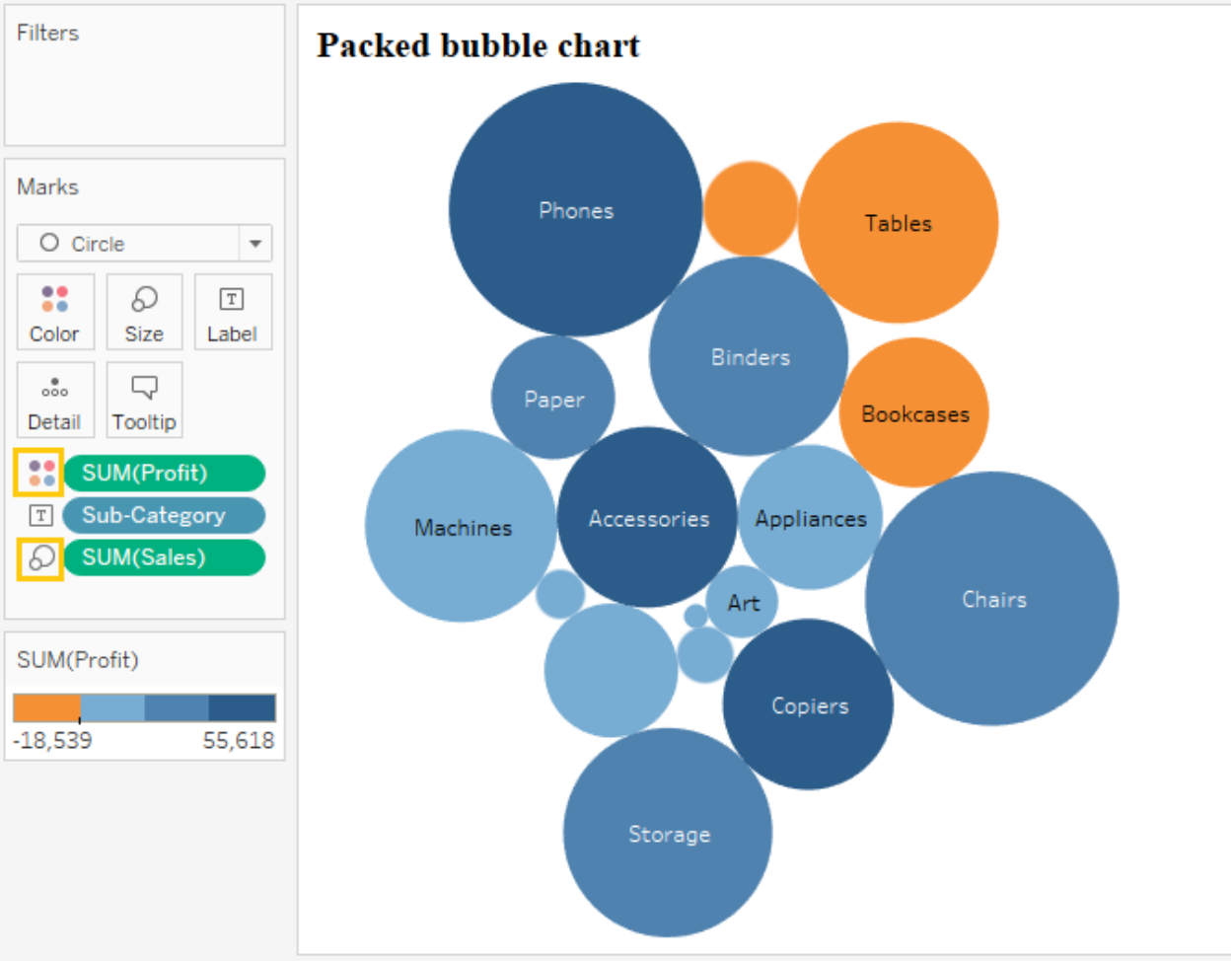
SAEs at various sites over the course of a study could have great value in understanding and action when represented in box plots.



Charts that were not fully analyzed include heat maps and parallel coordinate plots. However, preliminary review of heat maps shows that they can be good for some trending, such as what Web pages or parts of a page are frequently scanned by users (they're best for biometric feedback). Much like tree maps, users that have a spectrum of color blindness lose in both instances as the chart will not have one of its quantitative aspects – accurate, perceptible color representation. I personally feel that when it comes to data, heat maps are best when coupled with another form such as a geographical map or a surface plot. I have not seen evidence that Spotfire can produce a heat + surface map or incorporate a heat map into its map charts.

As for parallel coordinate plots, it's my initial understanding that area charts are better. Parallel coordinate plots can become dense and meaningless quickly – hiding meaningful data, and acquiring a line with the mouse can be a challenge. Similarly, 3D scatter plots, although cool looking, can tend to hide valuable data.

Bubble Plot/Packed Bubble Chart



Bubble charts have many benefits. Comparing easily spotted categories by size is one benefit. They are also a simple data visualization technique to provide insight into potentially dense information in a visually attractive format.

The benefits of a Bubble chart include:

- Explaining a complex data set easily.
- Analyzing data sets with multiple inputs more easily.
- Helping to visualize patterns and uncover trends using data analysis.
- Helping to correlate data as part of a data series.
- Being dynamic enough to analyze scenarios beyond time-series or whole-to-whole comparisons using line and bar charts.
- Effectively displaying a relatively large amount of information.

- Depicting the relationship between three or even four variables that can change over time.

Bubble charts are not ideal if it's important to depict exact values. Additionally, if data is too complex, or a lot of bubbles are adjacent to each other, the following can happen:

- A bubble chart can become difficult to understand.
- The overlapping of bubbles can make it difficult to distinguish them.
- Users may find it hard to depict zero or negative values.
- It can be hard to ascertain exact values using circle sizes.