

How We Visualize Data and How to Apply Those Findings in SAS® Visual Analytics.

Ryan Kumpfmiller, Zencos Consulting

ABSTRACT

With data discovery tools becoming more useful and elaborate each year, the capabilities of displaying data and designing reports have never been better. We have gotten to the point where we can now create interfaces that end users view and interact with. To get the most out of these capabilities, just like with data, we need to know what is going on behind the scenes.

Now that we are building interfaces with data discovery tools such as SAS® Visual Analytics, it's time to understand the way that we view data and incorporate that research into how we build reports.

INTRODUCTION

With every new business, social network, device, etc. that comes out, additional sources of data begin to feed the already colossal volume of data that the world has created. The technology industry has made great advances with building tools to store and structure this data. However, what about the end result of all of this work? After storing, processing, and querying through vast amounts of data, the goal is to be able to have consumable information that can help us make better decisions moving forward.

Being able to interpret this information in an effective manner is an important part of the process. Data can be transformed in any way needed, but it will not matter if the end consumer cannot make a conclusion. This is where data visualization becomes paramount to finishing the process from source to user. Users can look at a table and be able to view a few values, but with data visualization users have the ability to consume large amounts of data at once.

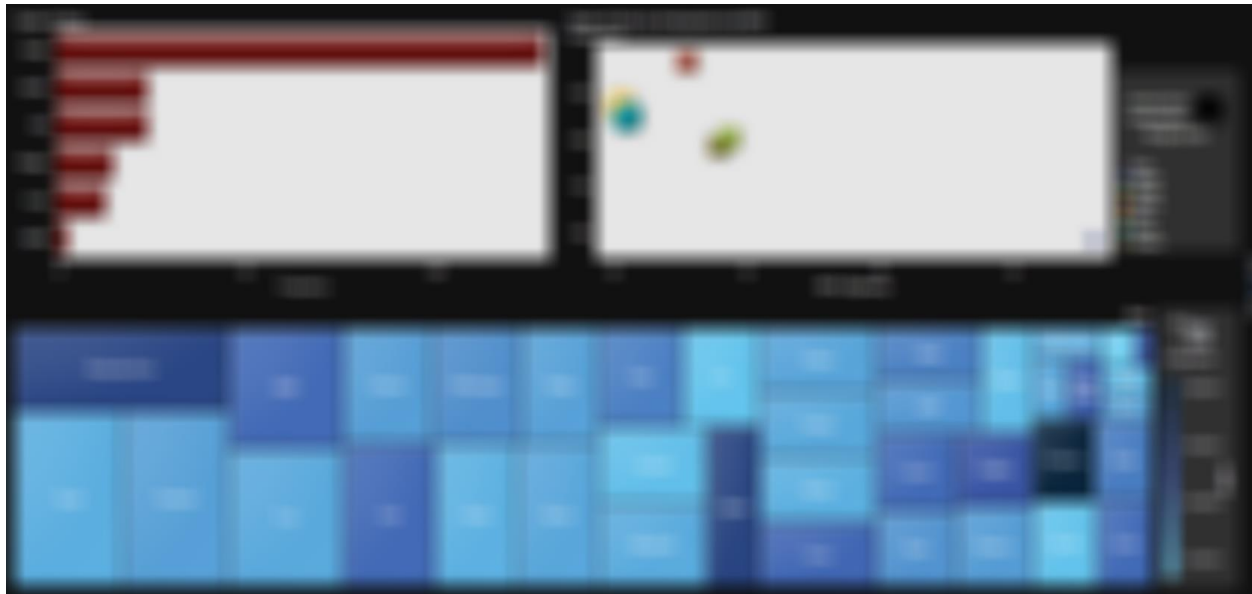
With that ability, this gives us the capability to easily identify problems and get key takeaways from the data. It is imperative, however, that the right data visualization techniques are used, or else those findings might not be as straightforward. With tools like SAS® Visual Analytics it is possible to get carried away or use the wrong method for displaying data. This is something to avoid. As Edward Tufte says "Above all else show the data".

HOW WE VISUALIZE DATA

Before jumping into SAS® Visual Analytics and showing some better visualization techniques to use, it is helpful to understand the process of how our minds perceive data. Once this is comprehended, it is easier to apply the concept to all areas of the tool instead of just knowing that a bar graph is better to use in most cases than a pie chart.

Understanding our perception is significant to learning how the eyes and brain interpret information that is viewed on a screen. In Colin Ware's book, *Information Visualization*, he explains the course of perceptual processing in a three step procedure. It is in this process that our eyes and brain work together through three steps to bring together a thought or conclusion to what we see. Now let's walk through each step.

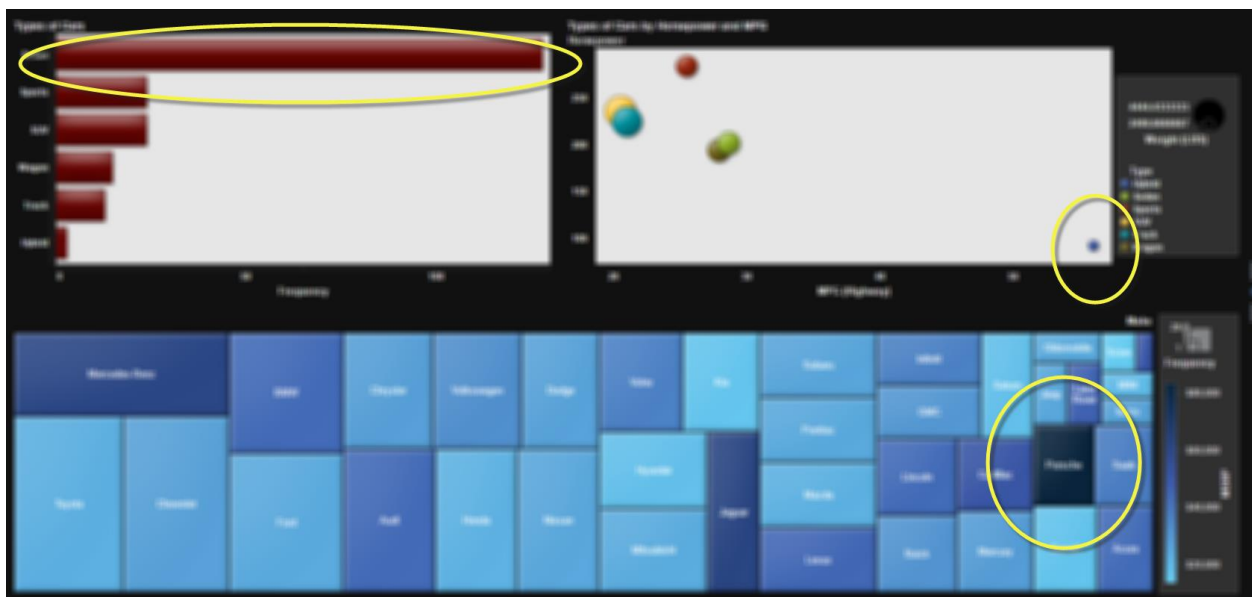
In the first step, unconsciously our eyes and brain quickly scan through the screen.



Display 1. Step 1 of the Perceptual Process

Here the neurons in the mind pick up the pre-attentive attributes of the screen such as form, color, position, and motion. So when the above screen is first glanced at, the brain will instantly pick up those detailed level aspects of the screen. It can tell that there are three sections to this screen, all of which are colored in a different manner, and each section is a separate type of graph.

After the initial scan, the brain will then take a high level look over the screen shown below in Display 2.



Display 2. Step 2 of the Perceptual Process

Since the first step broke out the structure, this step picks up those sections and looks at them individually. The brain's perception is selective and working memory is limited, so this is where any patterns of the data that are out of place will be acknowledged. For example, the first bar in the bar graph, blue bubble in the right chart, and dark blue square in the tree map will all be instantly recognized as items to take a closer look at.

The third and final step is where long term memory comes in to add context to what is happening on the screen. This is shown in Display 3 below.



Display 3. Step 3 of the Perceptual Process

The working memory of the mind is holding the few distinct points from step 2 and now looks to add background to what is going on. What are the data elements that are out of place? What measures are being used to make them stick out? Is there a logical reason to this? These are the questions looking to be answered that long term memory will help build a conclusion around.

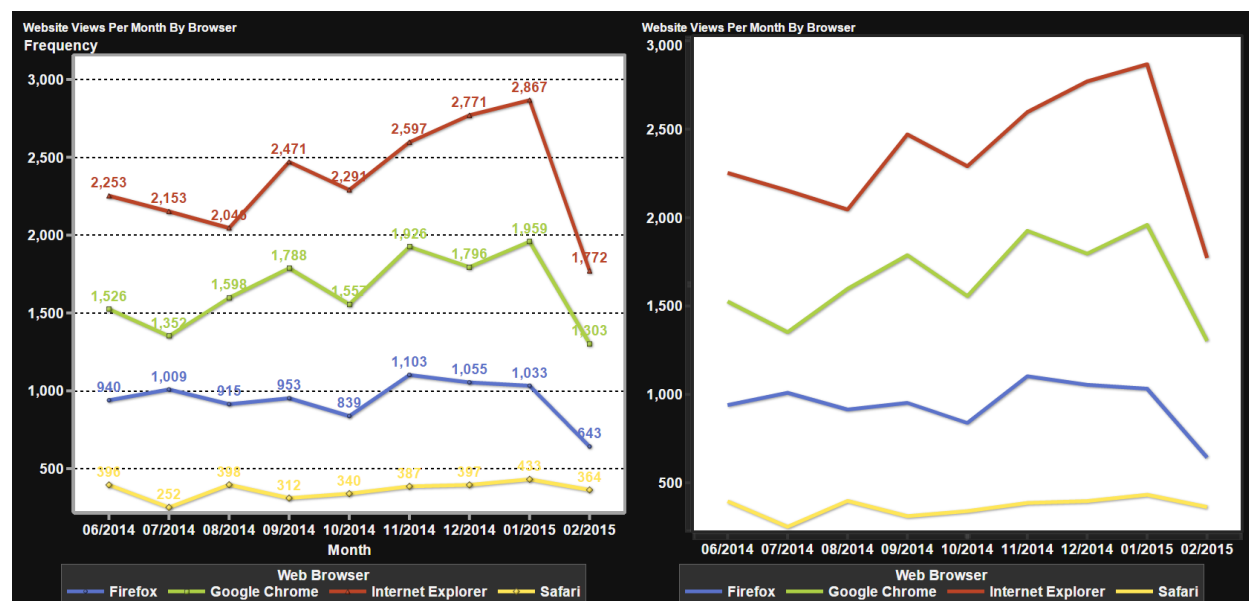
Memory plays a vital role with data visualization. As mentioned in step two, the working memory of the brain is limited to only holding a few items at once. When looking at a table of data, this means that the brain can only hold a few data elements. Even if they are calculated items in a dataset, that is still only skimming a fraction of what is available. By creating reports, those few items turn into conclusions that come from consuming large amounts of data graphically displayed. Therefore the goal of proper data visualization should be to make those conclusions obvious so that they can be reached as soon as possible.

EFFECTIVE METHODS FOR INTERPRETING DATA

Now that the process to how the mind visualizes what is on a screen has been covered, it is time to look into methods to improve that process. In order to be the most efficient in bringing out conclusions, step 1 in the perception process should be instant. Also, steps 2 and 3 should be apparent and supportive to what is being displayed. Time will be wasted if end users have to figure out what they are looking for or cannot tell what they are looking at.

CHART JUNK

Before looking into ways to improve reports, it would be good to first look into what shouldn't be done to graphs and charts that could hinder visualization. In Edward Tufte's book, *The Visual Display of Quantitative Information*, he references what he calls "chart junk" on multiple occasions. What he means by chart junk is anything added to the space of the graph that adds nothing to the data. When looking at a graph we want to see the data and nothing else. This includes grid lines, borders, axis marks, legends and labels that can already be understood in the graph, and any other items that do not add any context to the data being seen. These features are only going to slow down the first steps of the perceptual process.



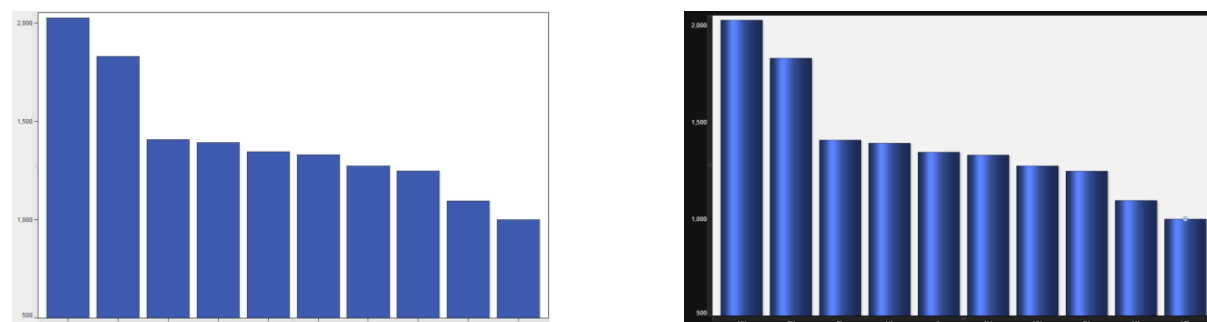
Display 4. Example of Chart Junk

Above in Display 4, this shows a side by side example of chart junk being used (left) and taken away (right). Notice how the data labels and grid lines make you focus more on details that aren't relevant to what the graph is trying to show. The exact numbers each month don't mean as much as noticing the changes for each category every month. Also, the right chart has removed axis lines and labels which allows more space for the actual chart itself. The labels for each axis do not need to be on the report if they can be easily understood from the title.

In addition to chart junk, Tufte also talks about variance over detail. With this he means that the details around the data are much less important compared to being able to recognize the differences in the data. Essentially, this comes down to knowing what is wanted out of a certain graph. If the focus is only a select few values of a certain element, then only show those ones instead of everything and build the chart around being able to display an outlier if it comes along. After all that is the end goal, to show the major differences in the important data to then be able to do further investigation.

SHADING AND TEXTURE

Once all of the un-necessary features are removed from the charts and only what is needed is being shown, then changes can be made to the data elements to help speed up the perceptual process. One visual aspect that is important is surfaces, but surface is a 3D component that we see in our real world environment. When dealing with a 2D space, the only way to create surface on objects is to change the shading and texture of the objects. When applying these qualities, the objects on the screen get a popped out look that makes them more visually appealing to the eye.

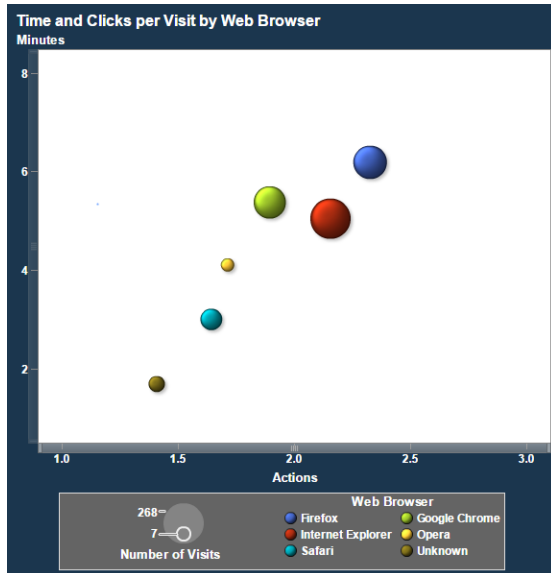


Display 5. Example of Added Styles

Above in Display 5, a style was added to the right bar graph that added some texture to the bars which in turn makes it easier to view. This may not seem like that big of a difference now, but when viewing a large report with multiple dashboards and graphs the change can make it easier on the eye in noticing positioning, shape, and layout. Thinking back to the second step of the perceptual process, these changes will make the data stand out even more to notice the discrepancies. Also, when it comes to shading and lighting of objects, the lighter the object is compared to its surroundings is always going to make it easier to view. This makes it better to use dark screens for your reports with

lighter objects.

MULTIFUNCTIONAL ELEMENTS



Another effective way in being able to visualize data from Tufte, is the use of multifunctioning elements. When using pie charts and bar graphs, only one measure is being interpreted at a time. There's nothing wrong with that, but you can save space by utilizing charts that use data points to handle multiple measures. Then you are clearing space for another dashboard or graph and packing more data into a smaller area. Bubble plots and heat maps are examples of objects that can handle additional data fields by using the size of the categories that they are representing.

In the Display 6 to the left you can see that the size of the bubble gives us an additional measure of number of visits, but we can also use the X and Y axis for a total of three measures all fit into one chart. Depending on the data and how it is related, using these types of graphs can be very effective in getting a lot of quality data on the screen to be used for analysis. Then when it comes to the third step in the perceptual process, even more data is being consumed at once to come to a conclusion.

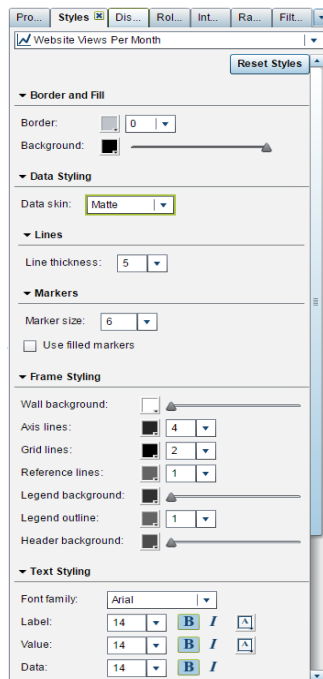
Display 6. Example of Multifunctional Elements

Additional points to consider to building out more effective charts:

- Make reports read from left to right and not up to down. Similarly to reading a book, the eyes are much quicker going horizontal.
- Use a color scheme that relates to the information being shown. For example, use red/yellow/green to signify bad to good and not the other way around.
- Bar charts are often better to use compared to pie graphs even when displaying data in percent form.

HOW TO APPLY TO SAS® VISUAL ANALYTICS

After looking into how the mind perceives information and the ways to enhance that process, those findings can now be leveraged into the SAS® reporting and visualization tool which is currently Visual Analytics.

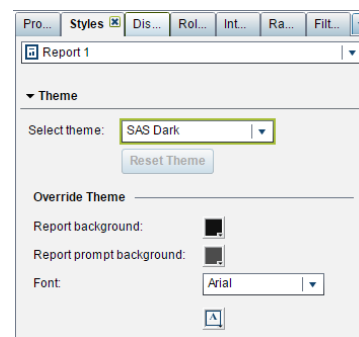


Display 7. Object Styles Tab

Within the report designer in SAS® Visual Analytics, there is tabs that give the ability to change the properties and styles of the objects. This is where the customization can be done to make objects more resourceful. When looking back to Figure 4 with two line charts, all of the chart junk can be removed within these two tabs.

Grid lines, markers, and axis labels can all be turned on and off from the Properties Tab. Axis lines can be removed from within the Styles tab (Shown on Display 7) by changing the color to the background of the report. Notice that there is also an option for a Data skin. This is how the texture and shading can be changed on the elements within a graph. Back in Figure 5 when this was changed to Matte it made the bar appear more 3D. There is 4-5 different options depending on the object at hand and each option has a different variety of shading and texture to it. The options stay mostly the same for all of the objects so that developers can keep a consistent feel across the report.

The report background can also be changed at the Styles tab when the first drop down box is selected to the highest level of the report. SAS has three different themes (Light, Dark, High Contrast) or that can be overridden and customized to any color as well. This is where a darker background can be applied to the report using the SAS Dark theme or a custom color. This was touched upon earlier as a way to have the audience focus in on the graphs more by using a darker background and lighter colors for the graphs.



Display 8. Report Styles Tab

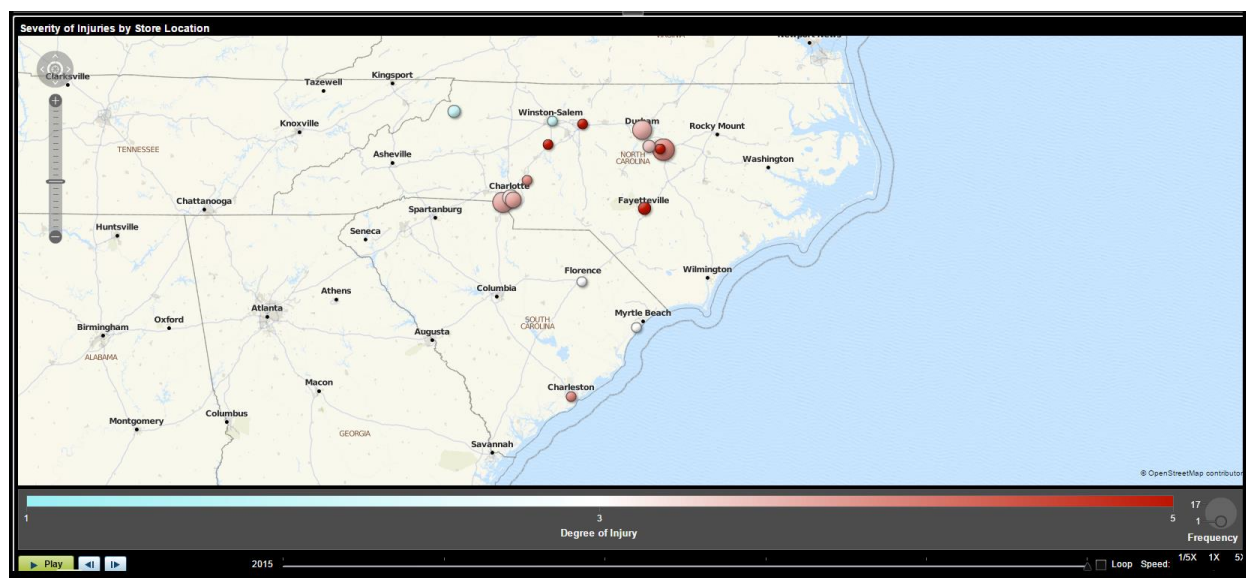
Another great aspect of the SAS® Visual Analytics tool is the ability to provide multiple means of analysis within tables. It was noted earlier that tables are bad for visualization because they limit the amount of information that the mind can handle at once. However, tables in this tool have additional capabilities that allows the user to pack more information in.

Regions	Incidents	Incidents Per Store	Incident Trends	% Claims Per Incident	Customer Claims Count	Employee Claims Count	Percent Employee Claims	Payout Sum
Central	387	25.8			111	86		\$5,186,993
Eastern	74	12.3			13	5		\$146,668
Western	64	12.8			10	5		\$53,980

Display 9. List Table Example

By utilizing the sparklines, calculated fields, and the display rules abilities of the table, two gauges and a line chart have been included into this small table shown in Display 9. This allows for a lot more data to be captured into such a small area and the gauges also help by clearly indicating what is going on with the data.

In the previous section, multifunctional elements were mentioned as a way to add more data into graphs. SAS® Visual Analytics also has the added capability to do animation over time in bubble charts and maps.



Display 10. Bubble Map Example.

In Display 10, there are buttons at the bottom where you can control the animation. When the play button is clicked, the bubbles for each of the data points will then grow and change color in respect to that time period's data. This can be done for days, months, years, etc. However, one thing to be careful of is missing data. If there is missing data for certain time periods then the bubbles will start to flash on and off. Anything that flashes repeatedly has an adverse effect on the eyes and brain.

CONCLUSION

Data visualization plays the essential final part in getting data into the hands of those who need it. If not done correctly, then the end users might end up missing the discrepancies in the data that they are after. By understanding the perceptual process, it becomes easier to imagine how our unconscious and conscious mind picks up what is on the interface. From there, developers can start creating meaningful charts by knowing how to keep chart junk out, using the more eye catching styles, and packing as much relevant data into the reports as possible. SAS® Visual Analytics is a great tool for this type of reporting since it gives the developer all the creative options needed in order to produce effective reports.

REFERENCES

- Ware, Colin. (2004). Information visualization perception for design (3rd ed.). San Francisco, CA: Morgan Kaufman.
- Tufte, Edward. (1983). The visual display of quantitative information (2nd ed.). Cheshire, Conn. (Box 430, Cheshire 06410): Graphics Press.
- Few, Stephen. (2009). Now you see it: Simple visualization techniques for quantitative analysis.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Ryan Kumpfmiller
 Enterprise: Zencos Consulting
 Address: 2005 Cambridge Beltway Drive
 City, State ZIP: Charlotte, NC
 Work Phone: 412-952-9240
 E-mail: rkumpfmiller@zencos.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.