

# Van Gogh Your Data: Data Visualization Methods with SAS® Business Intelligence

Natalie Parsons, Scott McQuiggan, SAS Institute Inc., Cary, NC

Presented by Sanjay Matange, SAS Institute Inc., Cary, NC

## ABSTRACT

From the halls of academia to the walls of museums, data visualization is an art form with ever increasing prominence. But how do we take data visualization concepts from these expert designs and create real-world business solutions? Using methods from distinguished data visualization experts and SAS BI, this paper presents contemporary techniques to visualize your data and simplify over-complicated reports with ample examples for both novice and expert report writers. We explore the utility of simplicity in BI reports, design guidelines that capitalize on understandings of visual perception, and optimizations for report consumption on various display media. While these techniques are applicable across industries, we will focus our examples on illustrations using data fundamental to the education industry

## INTRODUCTION

With guidance from data visualization experts, we are going to cover three principles that will help report designers surface data in a way that not only grabs the users' attention, but also clearly communicates the data's message. By first looking at intent, we will follow the advice of Edward Tufte as he advises on the importance of content in a graphic. We will identify questions which help report creators clearly identify the intent of their consumers, and therefore choose content which speaks to their consumer's needs. After identifying intent, we will look to Colin Ware and Steven Few for guidance on how users perceive visuals. Using this information we will determine a list of requirements to guide us in choosing the most efficient format, color, and size for visuals. Finally, after looking at intent and perception we will again look to Few and Tufte for counsel on simplicity. Looking to the utility of simplicity and Tufte's data-ink ratio to create report visuals which allow consumers to interpret the data's message without distraction and waste.

## Intent

When creating visualizations, it is important to identify the intent of the users. Edward Tufte tells us "The overwhelming fact of data graphics is that they stand or fall on their content." When attempting to identify user intent, ask yourself questions such as "who is going to use these reports?", "what does this visual need to present?", and "is this information meaningful and relevant to the user's domain?" Asking these types of questions help you determine the intent of the end user and the visualization needed to satisfy that intent.

Using higher education as an example, let us consider a user responsible for student enrollment. A perfectly designed report visual showing employee time and attendance information would not have the same relevance as the report below showing admitted students versus enrolled students. The report shows a cumulative trend for applications, admittance, and enrollment for 2009. The report answers the question "is this information meaningful and relevant to the user's domain?"

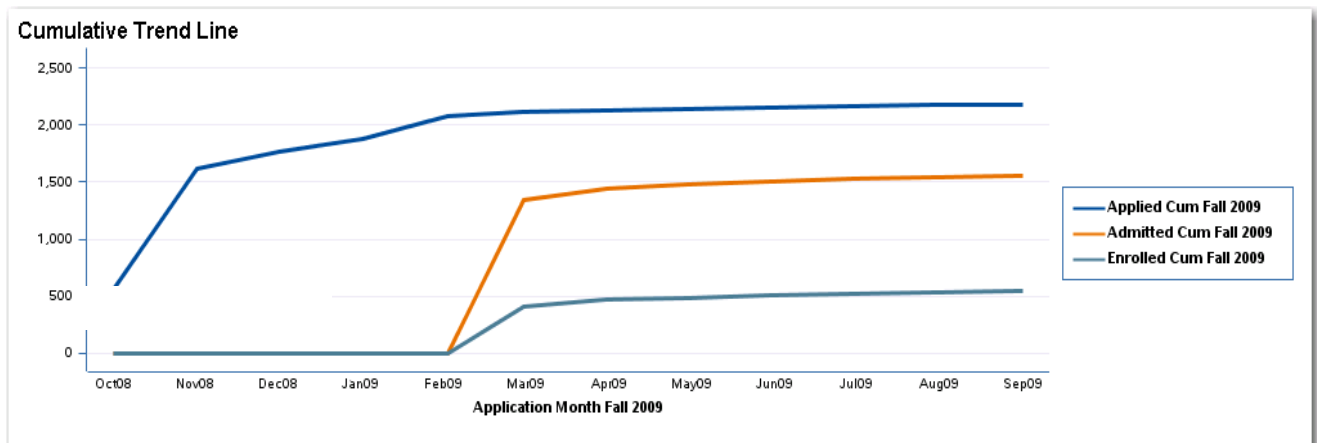


Figure 1. SAS BI Line Graph

## Perception

On the importance of visual perception Colin Ware says: “The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers.”

Three factors which have a strong impact on user perception are format, color, and size.

### FORMAT

The format of your visual should communicate five things:

- how the values relate
- accurate quantities
- easy comparison of quantities
- the ranked order of quantities
- the use of information.

(Steven Few, Data Visualization for Human Perception)

To explain this concept, Steven Few uses an example of a pie chart versus a bar chart. Using the SAS BI examples in the following figures, which one communicates this list the best?

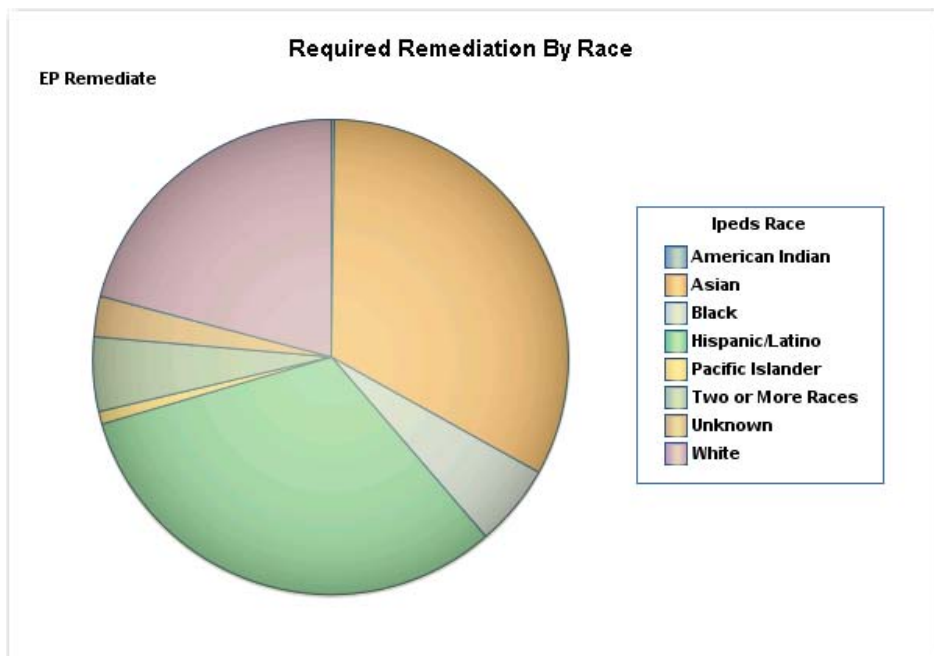


Figure 2. SAS BI Pie Chart

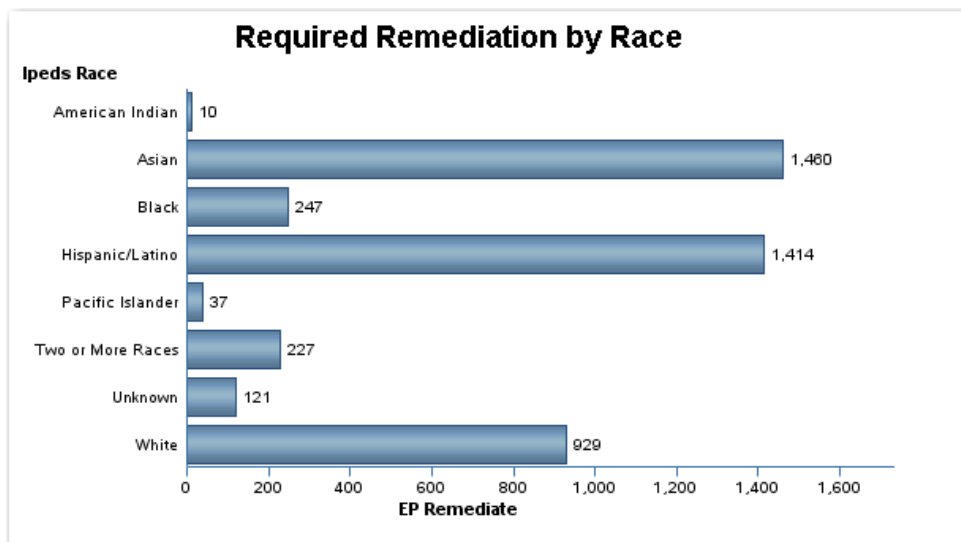


Figure 3. SAS BI Bar Chart

In the bar chart, we can see that all five criteria are satisfied, whereas in the pie chart they are not. In the pie chart, we cannot see accurate quantities or the rank order. At first glance, it appears that we can easily compare quantities, but on further contemplation, what are we looking at? The area of the pies compared to each other or the angle? In the bar chart, the bars are on a uniform axis which makes them easily comparable. The human perception of visual format helps the user understand what the data is trying to say. Using these five criteria by Few helps you create more efficient visuals.

**COLOR**

Color should have a purpose such as highlighting key information, providing distinction between categories of information, or to show scale. It should not be over used or it will lose its effect. Robert Spence warns against using color “with the hysterical abandon of a child discovering its first paint box.”

| Student Group  |                | First Time Freshman |              |                  |                          |                  |                  |                          |               |               |
|----------------|----------------|---------------------|--------------|------------------|--------------------------|------------------|------------------|--------------------------|---------------|---------------|
|                |                | Applied 2009        | Applied 2010 | Applied Variance | Applied Percent Variance | Applied Cum 2009 | Applied Cum 2010 | Applied Cum Pct Variance | Admitted 2009 | Admitted 2010 |
| App Month 2009 | App Month 2010 |                     |              |                  |                          |                  |                  |                          |               |               |
| OCT08          | OCT09          | 565                 | 735          | 170              | 30.1%                    | 572              | 739              | 29.2%                    | 440           | 534           |
| NOV08          | NOV09          | 1,053               | 1,395        | 342              | 32.5%                    | 1,625            | 2,134            | 31.3%                    | 747           | 945           |
| DEC08          | DEC09          | 146                 | 154          | 8                | 5.5%                     | 1,771            | 2,288            | 29.2%                    | 94            | 77            |
| JAN09          | JAN10          | 109                 | 125          | 16               | 14.7%                    | 1,880            | 2,413            | 28.4%                    | 69            | 58            |
| FEB09          | FEB10          | 200                 | 129          | -71              | (35.5%)                  | 2,080            | 2,542            | 22.2%                    | 133           | 83            |
| MAR09          | MAR10          | 37                  | 30           | -7               | (18.9%)                  | 2,117            | 2,572            | 21.5%                    | 21            | 17            |
| APR09          | APR10          | 15                  | 16           | 1                | 6.7%                     | 2,132            | 2,588            | 21.4%                    | 11            | 11            |
| MAY09          | MAY10          | 13                  | 23           | 10               | 76.9%                    | 2,145            | 2,611            | 21.7%                    | 8             | 11            |
| JUN09          | JUN10          | 9                   | 19           | 10               | 111%                     | 2,154            | 2,630            | 22.1%                    | 7             | 13            |
| JUL09          | JUL10          | 9                   | 17           | 8                | 88.9%                    | 2,163            | 2,647            | 22.4%                    | 4             | 8             |
| AUG09          | AUG10          | 14                  | 7            | -7               | (50.0%)                  | 2,177            | 2,654            | 21.9%                    | 7             | 5             |
| SEP09          | .              | 6                   | .            | .                | .                        | 2,183            | .                | .                        | 4             | .             |

Figure 4. SAS BI Table

In this table, there is an abundance of information in this case, highlighting key information is important. Using red and green to indicate progress toward a goal for the application variance year over year, the user can quickly identify progress.

The tile chart in the next figure uses a grey scale to show a comparison of admits of students for 2010 compared to the application which are shown by the size of the tile. In these examples, we can see how color can assist in the

visual perception of information.

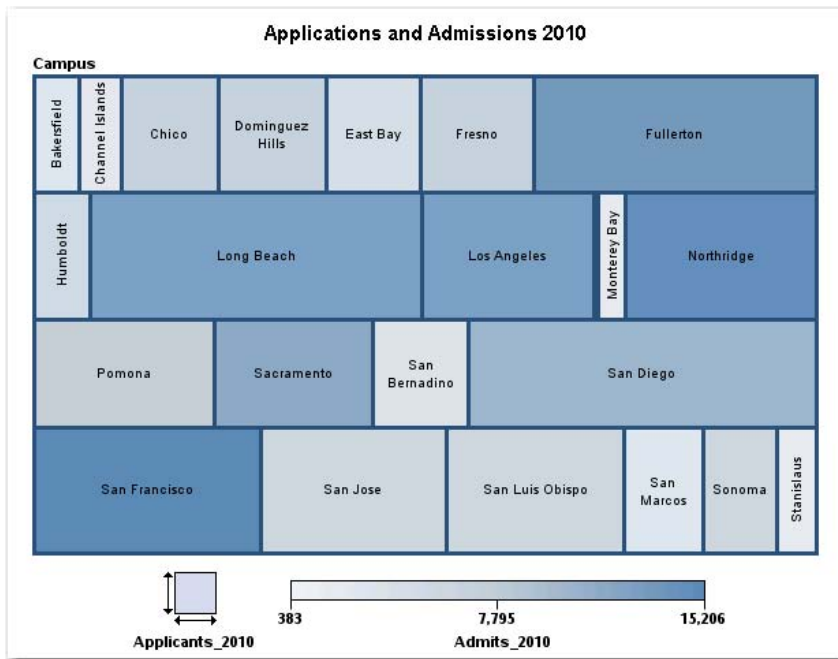


Figure 5. SAS BI Tile Chart

### Size

Size can be used to assist users in quickly determining exceptions in data and assign limited quantitative values to visuals. In the previous example, we see where size and color are working together to convey applications versus admission to each campus. The user can see that the largest tile (most applicants) is not necessarily the darkest (most admits). This chart quickly identifies the exceptions in the data. It does not, however, give you the exact ratio or numerical difference. The same can be seen in a line graph. The width of a line hints at quantitative information but does not give a specific numerical answer. In some cases, this might be the level of information the user needs. In those cases, visuals that use size can be valuable.

When choosing format, color, and size in the design of your visualization, it is important to keep in mind that they should never make a visual hard to read.

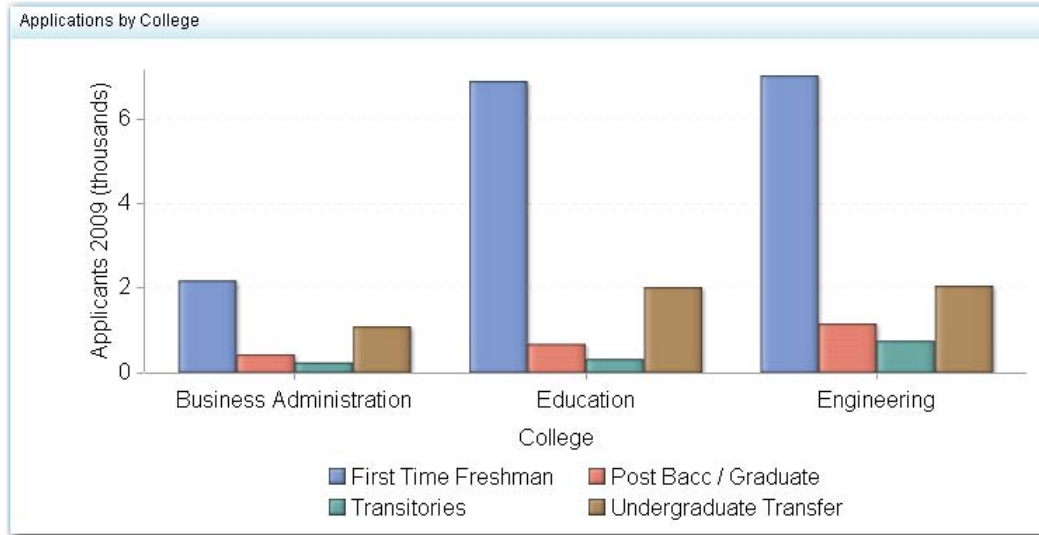
### SIMPLICITY

Steven Few: A display should say what needs to be said and no more, say it directly, clearly, and without distraction.

#### Summarization

Using the visual to count, sum or average data instead of listing all values helps reduce large sets of data. When detailed information is not required, it is more effective to summarize data because smaller amounts are easier for users to process.

When a data set is particularly large, look for possible group breaks. For example, group by student group instead of looking at all students as a whole. Looking at data in an isolated group can sometimes show a pattern that is missed in the data set as a whole. In the following example, the bar chart shows application by college and student group. Because we separated the data by college and student group, we can see that both the college of education and engineering have more than twice as many freshman application than the business college. This is something that would be lost if the students were not broken into groups.



**Figure 6. SAS BI Bar Chart**

Keep Few's simplicity axiom above in the back of your mind, and give users all the data they need but not more. You want the visual to effectively communicate the data's message while allowing the user to reach that message conveniently and quickly.

**Data-ink ratio:**

Edward Tufte gives us the data-ink ratio for simplifying visuals.

"Data-ink ratio= data-ink/total ink used to print the graphic= 1.0 [minus the] proportion of a graphic that can be erased without loss of data-information..."

This means that the ink used to print the visual should have as much data value as possible. Tufte advises that new ink should nearly always mean new information is being presented. This again allows the user to reach the data message without distraction and waste.

Two big offenders in this area are gridlines and background color. Upon first glance, a graphic with gridlines and background color can seem more pleasing to the eye. However, over time, these become distractions from the message that the data is attempting to convey. At times, space for visuals is limited and each pixel of real estate is important. Eliminating unnecessary visual information allows the most important to stand out.

**CONCLUSION**

By combining the principles of intent, perception, and simplicity, report creators can create dynamic visualizations to convey the data's message. Identifying intent will keep the visual relevant to the audience's needs. Considering user perception of format, color and size will allow users to accurately read reports. Simplifying report content, data and graphical ink, will show data without confusion and waste. Looking at these principles and their examples in SAS BI, report creators can start designing reports to surface the data message for the past, present and future.

**REFERENCES**

Tufte, Edward. *The Visual Display of Quantitative Information*. Cheshire, CT:Graphics Press LLC,1986.

Spence, Robert. *Information Visualization: Design for interaction*. Essex, England: Pearson Education Limited, 2001.

Ward, Matthew., Georges Grinstein, and Daniel Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications*. Natick, MA: A. K. Peters, Ltd, 2010.

Few, Stephen. *Information Dashboard Design: The Effective Visual Communication of Data*. Sebastopol, CA: O'Reilly Media, Inc., 2006.

Few, Stephen. *Show Me the Numbers; Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press, 2004.

Ahuja, Jaspreet S., and Jane Webster. *Perceived disorientation: an examination of a new measure to assess Web design effectiveness*. *Interacting with Computers* 14 (2001): 15-2. Web

Gwizdka, Jacek., and Ian Spence. *Implicit Measures of Lostness and Success in Web Navigation*. *Interacting with Computers*. 19 (2007) 357-369. Web

### **Recommended Reading**

- SAS Institute Inc. 2011. *Base SAS® 9.3 Procedures Guide*. Cary, NC: SAS Institute Inc.
- McDaniel, Stephen, and Chris Hemedinger. *SAS For Dummies*. 2<sup>nd</sup> ed. John Wiley Sons Inc.

### **CONTACT INFORMATION**

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

Name: Natalie Parsons  
Enterprise: SAS  
Address: 100 SAS Campus Drive  
City, State ZIP: Cary, NC 27513  
Work Phone: (919) 531-2058  
E-mail: [Natalie.parsons@sas.com](mailto:Natalie.parsons@sas.com)

Name: Scott McQuiggan, Ph.D  
Enterprise: SAS  
Address: 100 SAS Campus Drive  
City, State ZIP: Cary, NC 27513  
Work Phone: (919) 531-1119  
E-mail: [Scott.McQuiggan@sas.com](mailto:Scott.McQuiggan@sas.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.