Paper 228

Visual Communication Art & Science: The Design Guide and Gallery for Clear, Convincing Graphs, Tables, Maps, & Text

LeRoy Bessler
Visual Communication Power, Bessler Consulting & Research, bessler@execpc.com

Abstract and Introduction

Learn and see how to design information delivery that informs and delivers.

Technology is no substitute for design. It offers an ever-growing, bewildering array of software options and features. Templates and "wizards" eliminate design decisions, but don't necessarily make good ones. Web sites and color printers ease visual communication, but don't guarantee communication effectiveness. Special Effects impress, but don't inform or influence, missing the real objectives of visual communication.

Simple design (optimal options) and subtractive design (default removal) offer the best use of preparation and viewing time. Additive design can be both justifiable and desirable, but only if it provides leveraged return on communication investment. Use of color, despite how intuitive or straightforward it may seem, requires care.

Become your own "wizard". Learn design guidelines that are based on observation, experience, and research. The principles and concepts are largely software-independent, and are relevant for all experience levels. See poster illustrations—some innovative, others enhancements of the familiar—of how to make technology the servant of visual communication. These "Solutions That Work" were developed with SAS* and SAS/GRAPH*.

The illustrations for this paper, many of which are in color, are only in the poster. SUGI Proceedings has page count limits, and is printed in black-and-white. Upon request, the author will mail illustration copies.

Design for Communication

Put it before them

- briefly so they will read it,
- clearly so they will appreciate it,
- picturesquely so they will remember it,
- and, above all, accurately so they will be guided by its light.

Joseph Pulitzer

Simplicity is like an oasis in the desert.

Jan White

Design to inform and influence, not to impress.
A powerful image is "a visual sound bite".

LeRB

Acknowledgements

I thank Chris Potter, who, while a Graphics Section Chair back in 1991, encouraged me to share my ideas on this subject. This now perennial paper continues a tradition started by Chris during his series of SUGI graphics papers (1986-1990), in which he was the first to emphasize good SAS/GRAPH design (1989).

I thank Steven J. Subichin, who introduced me to his pioneering work on multi-line annotation and blanking for maps, and Gary F. Plazyk, who suggested I investigate cluster analysis as a tool to solve The N Color Map Problem.

Related Resources at SUGI 24

For an introduction to SAS/GRAPH software at SUGI, see the Beginning Tutorial "SAS/GRAPH Procedures: An Overview", by Phil Mason.

See also at SUGI the Advanced Tutorial "Show Them What's Important: Solutions for a Finite Workday in an Era of Information Overload", by LeRB.

Defaults Versus Customization

Software and hardware are power tools, but defaults can produce lots of sub-optimal results quickly.

Customize to:

- Focus on the message
- Focus on the data
- Suppress inessential graphic elements

Software defaults reflect the graph-paper-grid-based laboratory report tradition.

Overriding defaults requires more work, but avoiding decoration requires less work.

Special Effects Are For Movies

Communication and productivity are the real goals. Good design, interesting data can stand on their own. Omit drop-shadows, shaded background, clip art, etc.

Just Say "No" to the Designer Drug 3D

3D pie charts **always distort** the relative size of shares of the whole, the portrayal of which is exactly the unique visual purpose of a pie chart. Why would anyone design to defeat that objective?

3D bar charts introduce needless complexity to a simple image, and are usually more difficult to interpret than 2D.

For maps, use the straightforward two-dimensional CHOROPLETH map. The 3D alternatives—SURFACE, PRISM, and BLOCK maps—are picturesque, but impractical. SURFACE maps are too vague for serious communication. PRISM and BLOCK maps suffer from the response for some "high" states hiding that for "low" states.

Exception: Use PROC G3D for a 3-variable plot. It is ironic that, despite the abundance of unnecessary and hard-to-interpret 3D graphs, there are exceedingly few presentation graphs built to show the relationship between three variables, the only valid justification for use of 3D.

Consistency: Define Your Style

- For titles, footnotes, notes, font choices & sizes, symbols, line types, etc. (i.e., everything)
- **Consistency breeds/speeds comprehension**: viewer need not "recalibrate" perceptual and interpretive apparatus from page to page, or report to report
- **Spared over-choice, preparer is more productive**
- Implement standards with custom SAS macros

Text Is Essential: Handle It With Care

- **If the letters or numbers aren't readable, change the design or abandon the chart**
- Usually use black, it's the most readable color (that's why books are black on white)
- Emphasize with *italics*, **bold**, underline
- Be careful using ALL CAPS, or a different font
- Use mixed upper and lower case (it's the written communication standard, because it's easier to read)
- Keep it brief: **Focus attention with sparse text**
- If not doing science, suppress decimals (footnote any imperfect sum of rounded values)
- **Make the title your headline: the main message of your graph**

Limit Font Styles to Two, Sizes to Three

- Fancy font: maybe title, footnotes, legend
- Fancy software fonts increase processing time, print file size, & print time
- Use Default Characters (see explanation below) if you like the printed result
- Titles (usually all one size), maybe H > 1
- Footnotes small *only if to be downplayed*
- Body text usually H = 1, smaller if dense

Default Characters require F=NONE (or FTEXT=NONE for parts where the F parameter is unavailable), or not specifying F= (FTEXT=) at all. They require override of the fancy font that is the default for TITLE1.

What about Color?

In principle, color should be added to a graphic to communicate. In practice, it is usually added to decorate, or simply because it's available. (We confuse the possible with the necessary.)

Web publishing has caused an explosion of color communication. An intranet exhibits your color use enterprise-wide. The Internet exhibits your color use worldwide. Color printing has been available more than fifteen years for computer graphic applications. In recent years manufacturers have been producing moderately priced full-color laser printers that use toner (not wax or wax transfer) and plain paper. Also now available are comparatively cheap ink jet printers that provide very good quality output, but require special paper for best results. Advances in technology will continue to increase quality, speed, and function with an increasing performance/price ratio.

Industry analysts say untrained color users are unlikely to gain productivity or competitive advantage. To do so requires good design.

Always Remember These Facts about Color

Prof. Jay Neitz of the Eye Institute of the Medical College of Wisconsin: 8 to 10 percent of American males have some form of color blindness. Due to genetic differences, only about one-half percent of American females.

The commonest form of color blindness cannot distinguish red from green.

Reported Benefits of Color

- increased readership
- increased reading speed & comprehension
- faster learning
- reduced error rates
- improved recognition
- improved recall
- people prompted to act

Jan White on Color Communication

- If everybody screams, all you get is noise.
- The less color used, the more effective it is.
- Color consistency provides recognition.
- Use color to sort and/or link information.
- Make large areas pale, small areas bright.
- Colored text or lines must be thicker than black.
- Don't waste color on titles; for emphasis, use large or bold print instead.
- On a graph, use color to emphasize an important **visual** message.

Michael Turton on Color Communication

- Color works better with space around it.
- Color is more effective in bigger masses, since it is less dense than black.
- Color prioritizes information, whether meant to or not.
- Color is more effective if it has some function; it confuses if used arbitrarily.
- Color can guide the viewer.

Aaron Marcus on Color Communication

Use blue for large areas, not text or lines. Blue-sensitive color receptors are the least numerous in the retina's central focusing area.

Use red or green in the center of the visual field. The edges of the retina are not very sensitive to these colors.

Use thin, light text or lines on dark backgrounds for viewing in low light situations; use dark text or lines on light backgrounds for viewing in high light situations.

In an image, use the highest contrast for text. For the best contrast, use black and white.

LeRB on Color Communication

- Color does not improve bad design.
- Design color applications to communicate, not to decorate.

When to Use Color on a Graph

- No response levels/categories: black & white
- Few levels or categories: gray shades maybe
- Many levels or categories: color necessary

Benefits of Supposedly Boring Black-and-White

- In a color-cluttered, color-chaotic environment, well-designed black-and-white can be distinctive, "impactful", and memorable
- Faster, cheaper, more reliable equipment
- Easier to use (simpler equipment, no agonizing over color strategy)
- More copyable (more, cheaper, faster copiers): **Good Graphs Get Copied**

Color Differs on Different Media

With hardcopy versus transparency or slide or PC projectors, the same color looks different due to the difference between reflected and transmitted color.

With monitor versus hardcopy or transparency, the same color may look different because a monitor uses Red, Green, and Blue (RGB) additive primaries to mix colors, whereas printing uses Cyan, Magenta, Yellow, and Black (CMYK) subtractive primaries.

Furthermore, with different technology from different vendors, even for devices in the same class the same color may look different.

Given such confusion, it's no surprise that many vendors of software and hardware for color have begun to adopt standard ways and tools to specify and calibrate colors across all the facilities one might use, so that a color will (should) look the same everywhere.

Use Reliable SAS/GRAPH Color Names

For good control of color design, use H*hhhllss* color names, if available for your device.

- *hhh*= Hue
- *ll*= Lightness
- *ss*= Saturation
- *hhh*, *ll*, *ss* are hexadecimal values
- *hhh*, *ll*, *ss* ranges are 000-168, 00-FF, 00-FF
- *hhh* = 000 - 168 defines a "wheel of hues", 0 - 360 degrees (this makes more sense if you see the illustration in the SAS/GRAPH manual)
- ll = 00 (0%) is always black
- ll = FF (100%) is always white
- ss = FF (100%) is full saturation
- ss = 00 (0%) is always gray
- llss = 80FF is what I call the "true color"

Beware of Confusing SAS/GRAPH Color Names

Consider the pair

C = VIOLET (violet)
C = LIV (light violet)

"Under the covers", the HLS colors assigned are (spaces inserted in color name for clarity):

C = H 01E B0 67 for VIOLET
C = H 022 75 4E for LIV

Since 75 < B0, it's no surprise that "LIGHT" VIOLET is darker than VIOLET. And why does the hue vary?

Use HLS color names: they are reliable.

SAS/GRAPH Names for Grays

For GRAY, "LIGHT", "MEDIUM", and "DARK" do work as expected, but, even with WHITE and BLACK, may not provide enough colors. If so, use color names of the form GRAY*ll*, where *ll* is hexadecimal with range 00-FF. FF (hex for decimal 255) is 0% gray, i.e., WHITE. 00 (hex for decimal 0) is 100% gray, i.e., BLACK. 80 (hex for decimal 128) is 50% gray. Here are other correspondences for your convenience:
D5 = 17%, CC = 20%, C0 = 25%, AA = 33%,
99 = 40%, 66 = 60%, 55 = 67%, 40 = 75%,
33 = 80%, 2B = 83%.

How to Choose and Use SAS/GRAPH Grays

Gray shades too close together are difficult or impossible to distinguish. Decide how many (N) grays are needed for the chart, divide 256 by N - 1, and use the quotient (in hexadecimal) as increment from 00 to FF for *ll* in GRAY*ll* assignments. Subsets of the values provided in the section above can produce equally spaced grays for sets of 3, 4, 5, 6, or 7 PATTERN statements.

A photographer told me that the human eye cannot reliably distinguish more than five shades of gray (or of any other color). In my personal experience with printing and perception, a five-gray-shade map has no distinguishability problem, a six-gray-shade map is marginal, and a seven-gray-shade map can be hard to interpret. (I am counting black and white as shades of gray, which they are.)

With color, you can support more area fills. But then you need a strategy to pick colors, unless you use one hue and vary the lightness, but then you still have a distinguishability problem for sufficiently large N.

Beware Those Supposedly Safe Shades of Gray

- Text on gray-shade background often hard to read
- Gray shades sometimes do not photocopy well
- Black area fill on a map hides shared boundaries

Area-Fill Colors for Signed Response Ranges

If the response value for a bar chart, pie chart, or choropleth map is signed numeric, my suggestion for area-fill colors to suit the ranges of response is: Blue for high range; Light Blue for other positive; Red for low range; Light Red for other negative; and White for zero.

This color strategy reflects an American habit of using red as a negative indicator, avoids the red-green color blindness problem, and provides a color combination with reasonable contrast.

How to Choose Colors

Apart from the suggestion just above, nothing else specific is offered here. This topic deserves a treatise in itself. If you care about color significance across national boundaries, it's a quagmire. If you don't trust your intuition, ask the author for references.

Remove Axis Clutter

- Turn off axis lines: they tell nothing
- Turn off tick marks
- Label (invisible) tick marks sparingly
- If not turning off axis labels, supply your own
- Use Sparse Annotation whenever possible (more about this later)

Control Axis-Range Effects on Your Message

Usually start the vertical axis at zero, *not* at the default.

Zero de-accentuates fluctuations. *Prevent needless anxiety, questions.* Concern should be triggered by a measurement that fails or crosses management's pre-defined goal or threshold, not by insignificant dips or bumps.

Zero de-accentuates change. *Prevent needless elation or alarm.* Growth or decline should be judged by the size of the absolute or percent change, and by the practical effect of that change, not by the *visual* steepness of the slope of a trend, which is always controlled by an arbitrary choice of axis range.

On an analytical (rather than presentation) graph, one might actually want to be able to visually determine the Y value of the data points from the vertical axis. In that case, spreading the actual used data range across the full vertical space of the page/screen can be appropriate. However, the only absolutely reliable way to determine values is from a companion table, not from a visual impression instrument (i.e., a graph).

For percents, use the range 0 to 100. Bar length then is a "visual percent", and 100 is a natural choice as it is the absolute maximum value.

For a trend chart issued monthly, use a fixed number of months: January to December (for the same year, or across years), or Report Month N Years Ago to Current Report Month.

Usually Omit Area Fill

- Beneath line(s) *always*
- In pie slices, unless for a presentation, or for my New, Improved Pie Chart
- On simple bar charts, but maybe light gray, especially if the bars are close together

Make It "Easy on the Eyes" With Area Fills

- To carry information, use solid colors or grays
- Use parallel lines or cross-hatching *only in desperation*, and *never* use them on maps

Use of parallel lines or cross-hatching usually, if not always, yields an ugly, or at least visually disturbing, image. For maps, not only is the image already complex and rich with contour (i.e., line) detail, but also parallel lines or cross-hatching can even confuse boundaries with area-fill elements.

For some mapping, use of area fills to encode levels of response is functionally inappropriate. For the use of dot maps or bubble maps, see G. F. Plazyk, "Using the Annotate Facility with Maps: A Tutorial", in *Proceedings of MWSUG '91*, MidWest SAS Users Group (Fox Point, Wis.), 1991.

See also the discussion earlier of color, and the suggestion for one particular area-fill color palette.

Best Choice: Graph versus Table

A chart can both depict relative size, and supply detail. Presentations or reports that deliver both image (impact) and numbers (precision) are memorable, quickly and easily comprehended, and both influencing and reliable for decisions.

Perhaps as a surprise to nobody, research has indeed confirmed the intuition that graphs lead to quicker decisions, and tables lead to more reliable decisions.

The poster demonstrates several effective ways to supply detail for a graph, but sometimes a companion table is the best solution.

So, **the best answer to the question of graph versus table is often "graph *and* table".** And you can put them on the same page.

The traditional SAS-plus-SAS/GRAPH way to make graph-and-table composites is to feed a PROC PRINT or PROC TABULATE output to PROC GPRINT for loading into a graphic catalog, and then to PROC GREPLAY the GPRINT and a cataloged graph through a TEMPLATE.

Other possible and effective composite solutions (e.g., for a multi-line trend chart) entail putting significant values (either averages over the trend period, or ending values for the trend period) in a set of footnotes or legend entries. In this way, the footnotes or legend combine descriptive text with data to create a simple, but useful, table of significant information.

Best Choice: Lines versus Bars versus Pies

Line charts (plots) show trends or relationships. A side-by-side bar chart works better than a multi-line chart if there would be too much crossing. When annotating, a simple bar chart is better than a jagged single-line chart, to avoid obscuring the values. Moreover, you can get values at the top of vertical bars as a simple option, without annotation.

Bar charts show changes and compare magnitudes. SAS/GRAPH pie charts lose slice-related text if slices are too many or too small. A Custom Horizontal Bar Chart or Annotated Ranked Horizontal Bar Chart solves that problem. (See the poster.)

Make Plots Easy to Interpret and Look At

- Use Sparse Annotation/Tickmarking
  Sparse Annotation makes the graph talk
- Details are best provided in a table
- Use V=NONE for plots, if possible
- For point detection: V=DOT (a BIG dot), V=CIRCLE, or V=- (circle around dot)
- Avoid grid lines; if not, use fine line L=33
- Plot lines should be thicker than reference lines
- Used dashed lines for extrapolation or forecasts
- Color lines need to be thicker than black lines to be equally visible

What is Sparse Annotation/Tickmarking?

With Sparse Annotation/Tickmarking, you suppress almost all the tick mark values on both axes.

It is most likely to be applicable for a trend chart, i.e., a graph of a response variable on the vertical axis versus a temporal value (date or time) on the horizontal axis.

When trying to draw an inference from, or influence a decision with, a trend chart, you are interested in only: (a) portraying the trend; and (b) identifying the response level at critical points, not at every single temporal instance. (The complete list of precision data belongs in a table, not a graph.)

The critical points along a trend are:

1. starting value
2. ending value
3. intermediate maximum value, if any
4. intermediate minimum value, if any
5. points of inflection, if any
6. points along the trend where something of special significance occurs

Also, you can use vertical reference lines, or other labeled eyecatchers, to indicate temporal instances where a significant environmental change occurred, with expected or obvious impact on response. You might even highlight an event to emphasize that it had *no* effect on response.

Best Ways to Use Vertical Bar Charts

Always put the values at the tops of the bars. Omit the vertical axis, tick marks, and tick mark labels. Make things easy and unmistakable for the viewer.

When you have two or more response variables, **use side-by-side bars, not stacked bars.** With stacked bars, you can easily compare size only for the bottom set of bars. And with stacked bars any vertical axis and tick marks are useless for determining values except for the bottom set of bars.

Use Sequence of Bars or Slices to Communicate

The default order for SAS/GRAPH bar and pie charts is alphabetic order of bar and slice name (MIDPOINT value). To enable quick assessment of significance, order the bars or slices by decreasing size left-to-right (or increasing size, if smaller is better than bigger).

"De-alphabetize" the DESCENDING Pie Chart

PATTERNs are assigned by SAS/GRAPH to pie chart slices in alphabetic order of slice name. If, however, you want DESCENDING slices, and the colors to be arranged, e.g., from light to dark, then, to get pattern colors ordered by slice size, you must first determine the "size order for slice names".

It is a fact of visual perception that light areas appear bigger than dark areas. So, if bigness of response is what you want to emphasize, light fill makes it apparently even bigger. If smallness of response is what you want to emphasize, dark fill makes it apparently even smaller.

Control Pie Chart Labels

- Specify NOHEADING and OUTSIDE
- *Do not* match label color to slice (yellow text on white paper is illegible; black text is always most readable)
- SAS/GRAPH appends .0 to integer VALUEs: Suppress it with a FORMAT statement

Coping With Pie Chart PERCENT Feature

% only displayed at tenths or hundredths:

- if input VALUE to tenths, % to tenths
- if input VALUE to hundredths, thousandths, or "finer", % to hundredths

Get % at tenths for any VALUE finer than tenths by reformatting input to GCHART:
TOGCHART = ROUND(VALUE,0.1);

If the input VALUE is integer, or integer with zero(s) to right of decimal point, SAS/GRAPH insists on displaying % to hundredths, and there are *no circumventions*.

Try My New Improved Pie Chart

Release 6.10 of SAS/GRAPH brought a legend for pie slice names, but the legend in my New, Improved Pie Chart provides more function. (See the poster.)

Use the Powerful "Pac-Man Pie Chart"

The idea of a two-part pie chart may seem trivial, if not silly. But if the share important to your message is either tiny or huge, the image is very "impactful" and, therefore, memorable. (See the poster.)

Images stick, long after numbers are forgotten.

Images, added to text, have been found to improve, e.g., effectiveness of fundraising and memory of the request. (Such images were thematic symbols, *not* photos staged or picked for emotional response.)
In a case where "Other" is the huge slice, you can easily satisfy curiosity (if any) about its content with a table displayed below the chart.

It is essential to not blunt the visual message by splitting a big wedge into a lot of little ones, which may be as small as or smaller than the wedge whose smallness you wish to emphasize.

Tips for Tables

For details and examples related to tables, please see, also by LeRB, the SUGI 24 Advanced Tutorial "Show Them What's Important: Solutions for a Finite Workday in an Era of Information Overload".

In a prior paper—"Effective and Efficient Information Delivery for Executive Management", in *Proceedings of the Seventeenth Annual SAS Users Group International Conference*, SAS Institute Inc. (Cary, N.C.), 1992—I emphasized that, typically, a small subset of the observations accounts for a large majority, or almost all, of the total response. A Top 10 or Top NN Report (i.e., some one-page-or-less report) usually suffices, often accounting for 80% to 99% of the total response. This suggests the first two tips.

Subset the Data. "Let part stand for the whole." If you subset the data, be sure to include a subtitle which reassures the viewer/reader: "Includes PP% of all the data", or words to that effect. Only rarely will your audience demand the insignificant.

Rank the Data. Show them what's important. If you subset the data, you must make sure that whatever is better (high or low), all the highest or lowest responses are listed. If you show all of the data, still show them the more important data first. To learn about and get the code for the interesting device of my Nested Ranking Report, and a subtitled Top NN Report, be sure to consult the aforementioned Advanced Tutorial.

Assure Alignment. If you are using a font that is normally variable width (for example, when passing a PROC PRINT or PROC TABULATE output through PROC GPRINT), use its UNIFORM version. All the SAS/GRAPH fonts are available with a "U" suffix. This will assure that your decimal points and digit positions line up from row to row.

Usually Suppress Any Grid. PROC TABULATE by default produces a spreadsheet-like grid of cell separators. Unless horizontal lines are needed to guide the eye from row descriptor to column data items, these separators only distract (and marginally increase creation and display/print times). In PROC PRINT, you can double-space. In PROC TABULATE, you can either replace all or some of the grid-drawing symbols with blanks, or remove just horizontal lines with the NOSEPS option.

Why and When Maps?

Most SAS user sites have large amounts of data that include geographic unit area designators (in the USA, state code is probably commonest). Though this data can be reported in various tabular (or graphic) formats, a geographic effect (e.g., that of proximity) is not easily revealed without an InfoGeographic. Also, for presentation, a visual image is more interesting, stimulating, and memorable than a mere listing.

Good map design enables more effective, more efficient exploration and presentation of all this geo-keyed information in your data warehouse.

"InfoGeographics" is my name for thematic mapping, statistical mapping, Business Geographics, or the functions of a Geographic Information System (GIS).

Key Design Issues for InfoGeographics

3D and area fill for maps have already been discussed above, and are not taken up again here.

The remaining issues are more technical and intensely code-oriented in their resolution. For the details of much of my work in this area, see "Map Smart: Design and Build Effective InfoGeographics Using PROC GMAP and Software Intelligence" in *Proceedings of the Twenty-Second Annual SAS Users Group International Conference*, SAS Institute Inc. (Cary, N.C.), 1997. The poster at SUGI 24 exhibits maps that demonstrate those solutions. Discussions of the design issues of response-range assignment, annotation, and blanking/readability, and of how they are best addressed, are given below.

Establish a Rationale for Response Ranges

(Though I have had no occasion to do so to date, it should be understood that **Software-Intelligent solutions developed for range assignment in a map are equally applicable to a bar chart or pie chart for a continuous numeric response variable**.)

Whenever you are dealing with a continuous numeric response variable (as opposed to a discrete variable, where responses are "YES" or "NO", "female" or "male", etc.), whether you are using a map or bar chart or pie chart to graph the data, you get into the quandary of how to structure the response ranges.

To make a deliberate choice of ranges, in principle, requires knowledge of the data distribution. Before creating the map, one can first do a PROC SORT and PROC PRINT, and inspect the data. However, that is inconvenient, time-consuming, and laborious, and can result in an arbitrary decision anyhow.

What dramatized the possible severity of the range assignment problem was trying to map by state the SUGI 14 attendance counts. The SAS/GRAPH default range assignments lead to a ridiculous outcome. There were six response range midpoints, but only three of the ranges had any responses. This was rooted in the details of data distribution: one state at 736, another state at 447, and the remaining 48 in the range 1-178. But the defaults divided 0 to 780 into six equal ranges. It was a useless statistical map.

My reaction to this was to do what I called, with great pretentiousness and with tongue in cheek, solving "The Four Color Map Problem", a longstanding problem of mathematics. I built a macro to do Automated Rationale-based Response Range Assignment (ARbRRA). You simply point it at your response dataset, and tell it, among other things, at which percentiles you want the three breaks between the four response ranges.

Percentile-based ranges create a talking point for the map. Software defaults or arbitrary breakpoints cannot provide concept-based defendability.

My favorite percentile is the 50th, i.e., the median. What I call **"The Power of the Median"** is its representative centrality. The influence of outliers suffered by the regrettably popular average is absent. Regardless of the specific percentiles chosen, it is natural to break up the total range based on percentiles. One can use, e.g., the 20th percentile, median (the 50th percentile), and 80th percentile. The resulting four ranges may be called, e.g., Very Low, Below Median, Above Median, and Very High.

Other rationales can be built-in instead. E.g., one might prefer to use the mean and a multiple of the standard deviation to develop ranges. In that connection, consider how you might use Tchebychev's Theorem. Certainly, in the case of an intrinsically bounded response variable (e.g., percents, air temperatures, human body temperatures) unrepresentative outliers are unlikely or impossible, so that the mean then is a good measure of central tendency.

The **legend entries** are labeled with either user-specified text, or with the "trimmed" numeric ranges. **"Trimmed" ranges use actual data values, emphasizing the inter-range separation.** Traditional ranges (and default ranges) are always contiguous.

Even without the extremely anomalous result that inspired me to construct the ARbRRA solution, it is always better to make a deliberate choice of ranges, based on a rationale.

One thing lead to another, and over the years I have built other ARbRRA macros, to deal with other situations, among them being—

1. **The Five Color Map Problem.** The solution divides signed responses into five ranges: N highest, remaining positive, Zero, N lowest, remaining negative. The Software-Intelligent legend handling is very user-friendly, but too sophisticated to permit a brief explanation here (because it needs to handle 30 different situations of content and format).

2. **The N Color Map Problem.** The solution uses cluster analysis to determine the optimal bounds of N ranges, where N is user-specifiable. Except for the choice of N, it is impossible to consider this response-range assignment technique as in any way arbitrary.

3. **The Top NN Map.** This solution highlights only the NN areas (NN is a number of your choice) with the highest response and/or only those above a certain threshold, whichever is fewer. It can be extended to a simplification/variation of Case 1, by highlighting the Top NN and Bottom NN.

Make Maps Information-Rich and Readable

The general-purpose InfoGeograghics exhibited in the poster and code-documented elsewhere communicate as much information as possible with a single image, by multi-line annotation in the geographic unit areas. In effect, they build a table into the map. The table is a geographic grid, rather than a rectangular grid.

Why **multi-line, multi-function annotation**? Well, there are lots of things you might want to associate with a geographic unit area: name or some sort of alphabetic or numeric ID; response value; rank; percent of whole. Name/ID and value are obvious. I have previously emphasized the importance of ranking/ordering in bar charts, pie charts, and tables. Percent of whole has obvious use—it's the reason for the popularity of pie charts, and, e.g., a 50-slice pie chart for the USA just won't work.

When annotating areas filled with gray shades, or with dark or intense colors, custom-developed **"blanking" provides an inset box of white space, to assure readability.**

Multi-line annotation with blanking was first reported by S. J. Subichin in "Enhanced Useability for Annotation on SAS/GRAPH Maps", in *WISAS Proceedings*, Volume 5, June Issue, WISAS Inc. (South Milwaukee, Wis.), 1993.

I built on Subichin's pioneering work in two ways.

I provided further adjustments to the USA state-center coordinates in the vendor-supplied SAS/GRAPH data set MAPS.USCENTER. The adjustments permit more annotation to be inlaid without crossing the state boundary. The annotation box is as equidistant as possible from all near-points of the boundary.

I developed **"boundary-respecting blanking"** by trimming the otherwise rectangular boundaries of the annotation box for the difficult states of West Virginia, Florida, and Tennessee. Strictly rectangular boxes exceed those state boundaries if the annotated white space is kept at sufficient size.

Annotation Without Blanking

Blanking is not needed for area fills that use light colors (e.g., light pink, light yellow, light blue, etc.). Nowadays, one presumably can find devices to render sufficiently light colors (which was not the case in the early years of computer graphic hardcopy).

Continued Need for Blanking

Even if someday SAS/GRAPH blanking is provided, the vendor implementation may not offer the flexibility achievable with the custom solution developed by Subichin and enhanced by LeRB.

Many publications (e.g., SUGI Proceedings) do not accept color illustrations. Also, a very interesting, informative map is one that someone may want to copy. As has been previously discussed, color copiers are still not as widespread, cheap, and fast as black-and-white. Thus, annotated gray-shade maps usually are most practical.

For readability, whether of originals or of copies, blanking is always preferred, if not a necessity, when annotating in black over gray-shade area fill.

The N Color Map Problem and PROC CLUSTER

If there is nothing inherent in the responses (e.g., they are not signed numbers), nor in your or the audience's preference as to how to show or see the data, then you have no guide for choice of response ranges. Your only recourse is to inspect the data to find groups of responses. Intuitively, "natural" groups or classes are distinguishable by sufficient separation between their ranges. Choice of the *number* N of ranges is still arbitrary, but at least the boundaries of well-separated ranges can be justified.

An automated rationale-based tool that reveals those natural ranges, for any given N, is PROC CLUSTER. It offers eleven statistical methods for cluster analysis. The one I use is METHOD=CENTROID. It copes with outliers well. I can't say which is the "best" clustering method. For me, reasonability of results is the measure of adequacy of any method used. However, I built a macro, which among other things, permits user specification of METHOD=. If you have an expert opinion as to the best choice of clustering method, please let me know what it is.

Constraints on Good Design

If there is someone whom you cannot persuade as to the rightness of what you regard as good design, and whose wishes cannot be ignored, always remember:
The customer is always right,
even when the customer is wrong.

Trademarks

Author and Related Work in Progress

LeRoy Bessler, Ph.D.
Visual Communication Power
Bessler Consulting & Research
PO Box 96
Milwaukee, WI 53201-0096, USA
bessler@execpc.com
414-351-6748

Born to See, Meant to Look
- Faust

Dr. LeRoy Bessler is a SAS consultant, and frequent speaker, with interests in visual communication, graphic design, information visualization, color, InfoGeographics, and Software-Intelligent Application Development.

An award winner for papers on graphic design and visual communication, Dr. Bessler is writing a book to be published by SAS Institute, titled "Chart Smart: Design Guide and Solution Toolkit for SAS Graphs, Tables, and Maps That Inform and Influence".