Perceptual Principles for Effective Visualizations

Penny Rheingans
Martin Marietta Services Group
US EPA Visualization Center
RTP NC, 27709
rheingans@vislab.epa.gov

Chris Landreth
MCNC
RTP NC, 27709
landreth@ncsc.org

1 Introduction

Visualization is the science of representing data visually in order to enhance communication or understanding. In this way, the complex and powerful capabilities of the human visual system can be harnessed to aid in the comprehension of potentially huge quantities of information. Since the information contained in visualizations must pass through the perceptual system, careful attention to the system's characteristics can greatly improve the effectiveness of visualizations.

This paper presents a number of perceptual principles for the construction of effective visualizations. Principles are demonstrated in example images. In the example images, value is represented by color, height, position, or opacity. Other representation parameters are certainly possible, but are not covered here. Additional cues which can enhance comprehension are also spotlighted.

2 Single-Variable Visualizations

The simplest class of visualizations show the value of a single scalar variable over a domain. The purpose of such images is usually to show the pattern of data values or the locations of data features (such as extreme values) in relation to geographic landmarks (such as state boundaries or anatomical features). In addition to such qualitative information, quantitative information about values at particular locations may be of interest.

2.1 Exploit familiar scenarios

Visualizations are most effective when their layout, lighting and choreography contain visual elements that are associated with common perceptual experience. Previous exposure toosimilar conditions lays the gourndwork for a quicker and deeper comprehension of a visualization's geometric features.

Lighting. Through repeated experience, humans have become accustomed to illumination sources which originate from above the objects in a scene. For example, we usually view scenes illuminated by the sun or overhead light fixtures. Ramachandran [88] found that when subjects viewed images of objects that were ambiguously either concave or convex, the subjects always resolved the ambiguity based on the perceived direction of the light source from the objects.

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

Figure 1 illustrates the point. Two plates are displayed with circular objects visible within their bounds. The illusion is that the circular objects on the left plate appear as mounds, while the objects on the right appear as indentations. In fact, all the circular objects are mounds; the lighting on the right plate is from below. In general, lighting of complex 3D objects from below may cause similar confusion in the perception of those objects and should therefore be avoided.

Shadows. Displaying a 3D object in an environment where its shadow projects onto nearby flat surfaces is a powerful method of conveying three-dimensional structure and placement. Conversely, when shadows are absent, the object's structure or placement within the environment may be significantly misinterpreted. Figure 2 shows a comparison between a 3D isosurface rendered without (Fig. 2a) and with (Fig. 2b) a projected soft shadow onto an adjacent flat surface. Viewers who observed the scene without shadows were generally unable to conclude where the isosurfaces were located with respect to the map surface, but they could readily locate the same isosurfaces, when accompanied by shadows, as residing directly on top of the map surface.

Other cues. Some other shape and position cues of the physical world that can be effectively used in visualizations include: hidden line and surface removal, perspective, intensity depth-cuing, and stereo display. Hidden line and surface removal gives cues to the relative distances to objects along the same line of sight because nearer objects obscure farther ones. Perspective, intensity depth-cuing, and stereopsis all give depth cues.

2.2 Emphasize the interesting

Designers of visualizations should take care that the most striking features of the image are also the most important. Representations which draw the viewer's eye to unimportant features may cause more interesting features to be overlooked. Features likely to catch the eye are those that are brightly colored, moving or changing, defined by sharp boundaries, or highly saturated.

The common spectrum color scale maps the middle values to yellow, a particularly striking color. In applications where the location of middle values is of particular interest, this is appropriate. Such applications are not very common, however. More often, the high or low values are of greatest interest, and middle values are of least interest.

Double-ended Color Scales. Data sets with both positive and negative values can have a zero point representing no change, average, or expected value. In such data, deviation from zero (and the pattern of such deviation) is what is interesting. Figure 3 shows such a data set. Positive values show deposition of sediment from a column of water. Negative values indicate erosion and resuspension of sediment back into the water. Figure 3a uses a standard spectrum color scale to display the data. The distribution of erosion and deposition is not immediately obvious. A conscious distinction must be drawn between orange and red (indicating deposition) and the other hues (indicating erosion). Figure 3b shows the same data mapped with a double-ended

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

color scale. In this image, the areas characterized by erosion and deposition are clearly and immediately defined. Areas with no change in sediment are an unobtrusive grey.

The concept of double-ended color scales extends naturally to bivariate color scales, i.e., mappings from two scalar values to a color. One such scale might map the value of one variable to brightnesses of a hue, such as green. The value of the other variable would be mapped to brightnesses of the complementary hue, in this case purple. The contributions of the two variables are summed additively to give the display color. The resulting color scale contains three clearly discernible classes of colors. Greys represent places where the values of the two variables are comparable because equal chromatic contributions of the two hues cancel each other, producing grey. Dark greys are formed when both values are low, while light greys result when both values are high. Places where one variable is significantly larger are colored green, while places where the other predominates are colored purple. Such a color scale is useful in situations where the two variables are expected to be correlated. Places where this relationship does not hold are immediately apparent.

Missing Values. Sometimes things that are not shown can be distracting. Figure 4a shows a height-mapped, pseudo-colored representation of ozone concentrations over the southern hemisphere. Places where no data value was available are made invisible, or *iblanked*. While iblanking seems to be the most accurate way to represent missing values, the sharp boundaries of the holes and the very different values peeking through them draw a viewer's eyes to the holes and away from the values that are actually present. The effect is even stronger when the visualization is animated to show changes in the distribution over time. In fact, when the visualizer showed this representation to the atmospheric scientists studying the data, they promptly asked that the holes be "filled in." Figure 4b shows the same data set with missing values estimated by an adaptive filter. No valid values were changed.

Interpolation is not without dangers, however. A visualization showing interpolated values can misrepresent the smoothness of data values, the density of data values, and the likelihood that displayed values are accurate. Ideally, both the interpolated and iblanked visualizations would be available to the researcher. Additionally, a visualization that mapped value to color and/or height and mapped certainty about estimated values (related inversely to distance from real data values) to opacity would be valuable.

2.3 Say it again (Use redundant mappings)

Visualizations that represent data values using multiple display parameters have the potential to portray the data more effectively than visualizations that map each data variable to a single display parameter. There are a number of compelling reasons why this should be true:

1. Different display parameters convey different types of information most efficiently. For example, brightness conveys shape more effectively than hue, but hue provides more accurately distinguishable display levels than does brightness.

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

2. Multiple display parameters can overcome visual deficiencies. If one display parameter is ambiguous or ineffective because of the visual deficiency, another may compensate. For example, a person with color deficiencies would likely find it easier to unambiguously judge the value represented by a color using a redundant hue and brightness color scale than one using a standard spectrum scale which varies only in hue.

3. Multiple display parameters reinforce each other. In this way, areas with differing values have greater visual difference from one another.

Color and Height. For example, Figure 5b shows a representation in which concentrations of hydrogen peroxide (H_2O_2) have been mapped to both height and color. This redundant representation conveys the shape of the statistical surface more clearly than one mapping values to color (Figure 5a), because humans are more experienced judging the pattern of a surface from its shape than from its color. Additionally, the redundant representation conveys the location of global maxima (or near maxima) more clearly than a representation using just height-mapping (Figure 5c). By using a redundant representation, different perceptual channels can simultaneously process features of the data distribution.

Redundant Color Scales. Redundant representations need not employ parameters as distinct as height and color. Values can be represented redundantly using only color. For example, data values can be mapped to both hue and lightness. Figure 6b shows such a visualization. Compared to an image using only hue (Figure 6a), this representation shows the location and swirling structure of high areas more clearly. This redundant representation also has the advantage that it can be unambiguously interpreted by someone with a dichromatic color deficiency. For example, a protanope (commonly called red-green color blind) viewing Figure 6a would find it difficult to distinguish the values near 100 (green) from the values near 170 (red). The same person viewing Figure 6b would be able to distinguish the two values based on lightness; the higher values would appear brighter. Because a legend is included, the value encoded by each color is unambiguous (to the limits of discrimination).

The utility of redundant color scales has been emprically confirmed. Ware [88] conducted three experiments comparing a linear grey scale, a perceptual grey scale, a saturation scale, a spectrum scale, and a red-to-green scale for univariate data representation. In the first experiment, subjects were asked to judge the metric value of a colored patch surrounded by a contrasting area. The spectrum scale produced significantly more accurate metric value readings. In the second experiment, subjects were asked to judge the effectiveness of the color scales in revealing information about the surface properties of simulated surfaces. In general, the grey scales were judged to be more effective. In the third experiment, the five original color scales were compared to a redundant experimental scale that cycled through the hues while it increased monotonically in lightness (a rainbow scale similar to that in Figure 6b). When the experimental scale was used for a metric query task, subject accuracy was similar to that of the spectrum scale (which had no monotonic lightness variation) and significantly better than the others. This suggests that a color scale that varies in both luminance and hue can be used to accurately represent both metric and surface properties.

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

Other redundant color scales are the heated-object [Pizer and Zimmerman 83] and optimal color scales [Levkowitz 88]. The heated-object scale goes from black through red, orange, and yellow to white, with brightness increasing monotonically. The heated-object scale has more distinguishable display values and more contrast between different levels than a grey scale [Pizer and Zimmerman 83]. The heated-object scale has a stronger perceived natural ordering than the rainbow scale because of the monotonic increase in brightness and because the color order is based on experience. Levkowitz's optimal color scales increased monotonically in both brightness and RGB components while being linearized with respect to just noticeable differences (JNDs). Levkowitz experimentally compared the optimal color scale to a linearized grey scale and a linearized heated-object scale. He found the grey scale to result in significantly more accurate identifications of simulated lesions in medical images. While this result does not exactly validate the use of redundant color scales, the expect results of the experimental task (shape perception) was that the grey scale would excel.

Explicit Redundancy. In contrast to mapping a given data field to a single object having redundant attributes (e.g. color and height), the same data field may be mapped to two or more redundant objects, each having a single distinct attribute. Figure 7a shows a visualization in which the surface of the planet Jupiter has been mapped to two images, one with true color and one with a hue/value pseudo-colored field. With multiple representations such as this, accuracy in metric readings (pseudo-colored image) is maximized while maintaining optimal clarity of the object's form (true color image). In Figure 7b, a single molecule is redundantly displayed four times: as a single solid object, and as three orthogonal projections of the solid object. The object's form is best revealed by the solid representation, while quantitative information of the object's geometry in Cartesian space is revealed by the orthogonal projections. In both examples, redundancy is used to eliminate the need to compromise mutually exclusive visual elements of a single representation by allowing simultaneous display of all elements.

More Redundant Techniques. Other examples of redundant representations include combining size and color, opacity and color, and color and texture. A vector data set can be represented using both color and vector length to redundantly encode vector magnitude. Color and opacity can be used redundantly to represent value, mapping more important values to higher opacity so that interesting places are more apparent.

2.4 Minimize illusions

The human visual system is not immune to confusion. Specifically, it is susceptible to a number of illusions. These include:

1. The perceived size of an object may be influenced by its color.
2. The perceived hue of a color may be influenced by its saturation.
3. The perceived saturation of a color may be influenced by its hue.
4. The perceived depth of an object may be influenced by its color.
5. The perceived color of an object may be influenced by the color of surrounding objects.

Whenever possible, the conditions which give rise to these illusions should be avoided.

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

Color-size Effects. Some visual experiments have suggested that the color of an object can influence the perceived size of that object. Tedford, Berguist, and Flynn [77] surveyed studies of the effect of color on perceived size, noting that researchers differed in their conclusions about whether an effect existed as well as the relative ordering of color-size effects. They concluded that the disagreement could be attributed to lack of consistency of other stimulus characteristics such as saturation and brightness. They conducted their own experiments under precisely controlled conditions and found a significant color-size effect. Specifically, rectangles of the same size, saturation, and brightness appeared to have different sizes when colored red-purple, yellow-red, purple-blue, or green (in order of decreasing apparent size). At high saturations, this effect was statistically significant for all color pairs except yellow-red and purple-blue. At low saturations, only the difference between yellow-red and green rectangles was significant. In trials where hue was held constant and saturation varied, rectangles with higher saturations were consistently judged to be smaller than less saturated rectangles. Generalizing from the surveyed studies, they observed that warm colors (red, orange, yellow) appear larger than cool colors (blue, green).

Cleveland and McGill [83] investigated the implications of the color-size illusion for statistical maps. Subjects were shown a map of Nevada in which counties where colored red or green with the total area of red and green nearly equal. Subjects were asked to judge which color, if any, represented the larger land area. Each subject was shown ten maps. On the average, subjects judged that the red areas were larger more often than they judged the areas the same or the green areas larger. When the experiment was repeated using low-saturation tones of red and green (formed by adding yellow), no such bias was observed. Their results suggest that the color of a region influences the perceived size of the region, and the effect is strongest for very saturated colors.

2.5 Control level of detail

The amount of detail in a visualization should be appropriate to the form and content of informations displayed. Simply put : include enough detail, but not too much. Detail in visualization can take many forms, including contour lines, surface detail, additional variables, and color scales with high frequency components.

Segmented vs. Holistic representation. Whether to represent a 2D or 3D field as discretized steps or as a continuous gradation depends on the necessity of displaying the segmented or holistic structure of the data. Contours or discretized color fields (2D datasets) and isosurfaces (3D datasets) present segmented structures; they utilize our perceptive ability to evaluate quantity at specific locations in 2D or 3D space. Continuous color fields (2D datasets) and voxel fields (3D datasets) present holistic structures; they utilize our perceptive ability to evaluate detailed form globally.

Figure 8 shows two displays of a pair of coexisting scalar datasets representing pollutant concentrations in 3D space. The isosurfaces in Figure 8a reveal the form of the SO_2 pollutant field at 4 and 8 ppb, and the SO_4 field at 4 ppb only. Obtaining precise information from the isosurfaces as to where the pollutants attain or exceed a

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

given threshold (e.g. 4 ppb) is possible, but precisely determining the overall distribution of pollutants is not. By comparison, the voxel fields in Figure 8b reveal substantially detailed information of the pollutant structure in 3D space but do not clearly reveal any given quantity at a given location in space.

Recent perception research by Livingstone and Hubel [88] indicates that visual information is processed in at least three separate cognitive pathways in the human brain. One pathway, "blob-thin-stripe-V4,"processes perceived spatial distribution of colors. A second pathway, "parvo-interblob-pale-stripe-V4," evaluates high-resolution shape information. The third pathway, "magno-4B-thick-stripe," processes movement and stereoscopic depth. These three pathways are subsequently integrated so that one sees a unified environment in 3D space. This three-fold separation of visual information suggests that discrete representations of data (e.g. contours or isosurfaces) are likely processed primarily by a separate cognitive pathway than continuous data representations (e.g.. color or voxel fields): the shapes of contours and isosurfaces are processed by the shape-resolving "parvo" pathway, while the overall forms of continuous fields are processed by the "blob" pathway. Thus the choice of creating a discretized or continuous representation is tied to utilizing one of two distinct cognitive processes.

This distinction between two perceptual processes appears to be particularly striking when the datasets are animated as time sequences. When isosurface or contour representations are used and the datasets are animated, the viewer's attention seems to focus on the moving *edges* of the shapes rather than on the overall dynamics of the field. When continuous representations are utilized, the viewers attention seems to be more globally directed, making the perception of many different events simultaneously more possible than with discrete representations.

Color Scale Detail. Adding detail can sometimes make the fine structure of a value distribution more apparent. One way to do this is by using color scales with high frequency components. One type of high frequency color scale cycles rapidly and repeatedly through a sequence of colors. Such a scale makes small value differences more apparent because the colors representing them differ more than those from a lower frequency color scale. A repeating high frequency color scale, however, makes it impossible to look up the value represented by a color because a single color can represent several non-contiguous values. Adding non-repeating high frequency components such as contour lines or other sharp transitions can provide discrimination of small differences and metric lookup capabilities.

Figure 9a shows a visualization of suspended sediment concentrations using a spectrum color scale. It conveys a smoothly changing distribution where values drop off monotonically with distance from the two points of high value. Figure 9b represents the same data using a striped color scale. This color scale is composed of nine narrow bands of highly saturated color. Between each pair of bands is a section containing less saturated stripes of the two colors. In the image, the saturated bands give rough contours, the striped sections give an indication of the overall value distribution, and the stripes themselves show the fine structure of the value distribution.

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

Surface Detail. Adding surface detail to a height-mapped or iso-value surface can also facilitate the perception of fine surface structure. Figure 10 shows two height-mapped, pseudo-colored representations of ozone concentrations over the Mid-Atlantic states. The white lines below the ozone surface show state boundaries. In Figure 10a the transparent ozone surface allows the map to be seen, but small features of the surface are difficult to perceive. Figure 10b uses a texture map that modulates opacity to give the impression of a tangible surface while allowing the map to show through. Compare the lower right and upper left of the two images to see the differences in discernible detail of surface shape.

3 Multivariate Visualizations

Multivariate visualizations show two or more variables over a single spatial domain. The complexity of many scientific domains requires multiple variables to model phenomena or processes. For example, the complexity of environmental processes require simultaneous visualization of multiple quantities in order to represent the interactions among the various components. Multivariate visualizations in the environmental sciences can show joint distributions of two chemical species to compare patterns of pollution; airborne species in conjunction with weather conditions to see how weather affects pollutant transport; variables from two model domains (such as air and water) together in order to explore their interrelationships; species concentrations in conjunction with topography to see how terrain characteristics affect transport; or modeled data together with field samples of the same quantity for the purposes of model validation.

The basic challenges in the visualization of multivariate data are to:

1. Clearly show the spatial relationship of different variables.
2. Keep the different variable representations from interfering with one another. Specifically, the single-variable distributions should still be visible.
3. Facilitate the understanding of joint distributions.
4. Show as many variables as can be effectively displayed.

3.1 Show multiple surfaces

In some multivariate visualizations, the contributions of different variables naturally occupy different regions of space. Some examples are isosurfaces of different variables, isosurfaces at different value levels, situations where different variables are modelled or sampled at different locations, or height-surfaces of different variables. In such cases, the challenge is to display each variable in a way that does not obscure the others rather than to find a way to combine display parameters in the same space.

Most commonly, surfaces are made transparent so that other objects can be seen through them. A disadvantage of this technique is that discerning the shape of a transparent surface is more difficult than an opaque one. One reason for this is that obscuration cues are lost. Figure 11a shows the solvent-accessible surface of a molecule along with a ball and stick representation of the atoms and bonds. The solvent-accessible surface was made transparent to allow the ball and stick object to be seen. Figure 11b shows the same molecule with the surface textured with an opacity-

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

modulated texture. Using this technique, the ball and stick object can be seen through the holes in the texture while the opaque sections provide surface shape cues. This technique also works for nested surfaces over an interior object (such as solvent-accessible surfaces using different radii), but the interior object is difficult to see clearly unless the representation is rotated interactively or in an animation.

3.2 Use orthogonal display parameters

Displaying two or more sets of data in a single visualization if often useful or necessary. Such an example was given in Figures 8a and 8b, which showed two pollutants coincident in 3D space. In such visualizations, a single consistent representation method (e.g. isosurfaces in Figure 8a) is most suitable for conveying the interaction between the multiple datasets. Often, however, the datasets will differ spatially or qualitatively, or will have little or no numerical correlation. It is necessary in these cases to represent the respective datasets with visual distinctness, or orthogonality to minimize visual confusion within the scene. In Figure 12, a visualization is constructed to convey three distinct numerical simulations simultaneously: pollutant concentration in 3D space, rainfall intensity on the 2D land projection, and pollutant deposition (also on the 2D land projection). The 3D pollutant data is mapped to an isosurface that propagates over the 2D land mass. Because the isosurface has distinct geometrical edges, representations for the 2D data which do not cause confusion with these geometric features are chosen. For the 2D deposition data, therefore, a continuous 2D color field is employed rather than contours because the gradual forms of the color field provide the needed visual distinctness to the well-defined form of the isosurface. Likewise, a segmented height field, composed of clear rectangular shafts, is chosen for the 2D rainfall data because its appearance provides significant orthogonality with respect to the other two representations.

In the perception model put forth by Livingstone and Hubel [88], the use of orthogonality is consistent with optimal utilization of the separate pathways of visual processing. According to this model, the distinct and sharp-edged shape of the isosurface in Figure 11 is processed largely by use of the shape-resolving parvo system. The form of the 2D pseudo-colored surface, with its gradual spatial changes in hue and value, is apprehended largely through use of the color-processing blob system, while the height/depth field of colorless shafts representing 2D rainfall is largely processed through the depth-processing magno system.

4 Conclusions

Attention to perceptual principles is essential to the construction of effective visualizations. These principles mandate the use of familiar paradigms, de-emphasis of uninteresting features, redundant mappings, appropriate level of detail, and orthogonal display parameters.

Acknowledgments

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.

We would like to thank Dudley Bromley, Mark Bolstad, and Todd Plessel of Martin Marietta Technical Group at the EPA Scientific Visualization Center.

References

Cleveland, William S., and Robert McGill (1983), A Color-Caused Optical Illusion on a Statistical Graph, *The American Statistician*, vol. 37, no. 2, pp. 101-105.

Levkowitz, Haim (1988), Color in Computer Graphic Representation of Two-dimensional Parameter Distributions, PhD dissertation, University of Pennsylvania.

Livingstone, Margaret, and David Hubel (1988), Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception, *Science*, vol. 249, pp. 740-749.

Pizer, Stephen M., and John B. Zimmerman (1983), Color Display in Ultrasonography, *Ultrasound in Medicine and Biology*, vol. 9, no. 4, pp. 331-345.

Ramachandran, V. S. (1988), Perception of Shape from Shading. *Nature,* vol. 331, no. 6152, pp. 133-166.

Tedford, W. H. Jr, S. L. Berquist, and W. E. Flynn (1977), The Size-Color Illusion, *The Journal of General Psychology*, vol. 97, pp. 145-149.

Ware, Colin (1988), Color Sequences for Univariate Maps: Theory, Experiments, and Principles, *IEEE Computer Graphics and Applications*, Sept. , pp. 41-49

Reprinted with permission from *Perceptual Issues in Visualization*, Georges Grinstein and Haim Levkowits, eds., Springer-Verlag, 1995, pp. 59-74.