Designing multivariate displays: a human factors approach.

Name: Paul Gokin

Course: HF 730, Spring 2005

Instructor: William Gribbons, Ph.D.

Date: March 29, 2005.

Colin Ware commends data-based visual displays on exploiting the brain's natural pattern-finding ability in "presenting information so that structures, groups, and trends can be discovered among hundreds of data values" (Ware, 2004, p. 225). Not all displays are created equal, however. William Cleveland cautions that "no matter how intelligent the choice of information, no matter how ingenious the encoding of the information… a graph is a failure if the visual decoding fails" (Cleveland, 1985, p. 229). Similarly, Jock Mackinlay makes the following distinction between expressiveness and effectiveness of a graphical presentation: while *expressiveness* depends on accuracy and completeness of the display, *effectiveness* depends on how well an already expressive presentation matches the viewer's perceptual and cognitive abilities (Mackinlay, 1986, pp. 119-124). In light of this, the purpose of this paper is to explore ways in which strengths and weaknesses of human perceptual and cognitive abilities guide the design of data displays.

It's all about comparisons. "The purpose of display is comparison (recognition of phenomena), not numbers," writes John Tukey, "much of what we want to know about the world is naturally expressed as *phenomena,* … which can be described in *non* numerical words" (Tukey, 1990, pp. 329-330). Therefore, the kinds of questions that visual displays answer best are questions like: "Is the value small, medium, or large?" or "Is the difference, or change, up, down, or neutral?" or "Is the difference, or change, small, medium, or large?" or "Do the successive changes grow, shrink, or stay roughly constant?" or "Does the vertical scatter change as we move from left to right?" (Tukey, 1990, p. 331). Answers to all of these questions involve comparisons. In fact, it is by comparing positions, distances, lengths, slopes, densities, colors that one can see structure, patterns, and trends in the data.

How we "decode" visualizations.

William Cleveland divided graphical decoding tasks into graphical-perception tasks and graphical-cognition tasks. Graphical-perception tasks include judging relative position of marks, slopes of connecting lines, areas, densities, colors, etc. (Cleveland, 1986, pp. 230-231, 235). Graphical-cognition tasks, on the other hand, involve using graph's scales and labels to read off data values and perform rapid mental calculations and quantitative reasoning (Cleveland, 1986, pp. 232). Not all graphical-perception tasks can be performed with the same degree of accuracy, however. In fact, Cleveland, through a combination of theoretical and experimental approaches, was able to order the basic graphical-perceptual tasks from most accurate to least accurate:

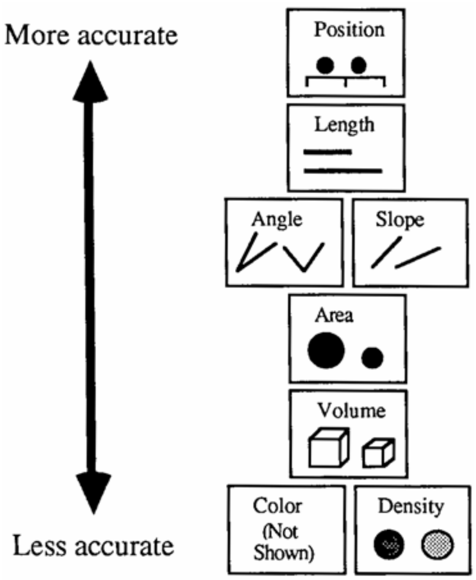


Figure 1. Cleveland's ranking of elementary quantitative perceptual tasks (from Mackinlay, 1986, p. 125).

It is important to keep in mind that Cleveland's task accuracy continuum applies only to judgments made along a quantitative scale. What about judgments along nominal and ordinal scales? Jock Mackinlay can be credited for building on Cleveland's work and creating the following rankings for three of the major scales: [1]

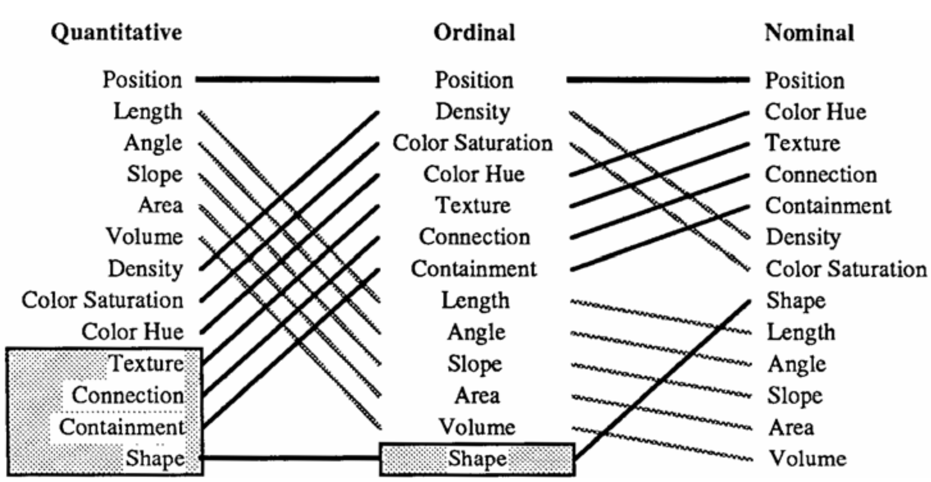


Figure 2. Cleveland's and Mackinlay's combined ranking of elementary perceptual tasks (from Mackinlay, 1986, p. 125).

---

[1] A *nominal* scale (also called a categorical scale) arranges objects (i.e. measurements) into arbitrary "equivalence classes," so that objects falling into the same class are perceived qualitatively equivalent. The only permissible statistic here is determining equality—whether an object is equal to the one it's being compared to. An *ordinal* scale orders the objects. It makes it possible to compare objects (greater or less, but not by how much) and also calculate the median and percentiles. Objects arranged on an *interval* scale make it possible to determine the distance between them, allowing for calculating the mean, standard deviation, an so on. Finally, a *ratio* scale is created by adding a point of origin to the interval scale. (From Halstead-Nussloch, 1995, based on Stevens' taxonomy). Interval and ratio scales will be collectively referred to as quantitative throughout this paper.

In particular, Mackinlay noticed that the three components of color—density (lightness), hue, and saturation—while poor at encoding quantitative data, are particularly good for encoding values on an ordinal scale. In addition, hue and texture have been found to work well for encoding nominal data. Judging position, however, remains the easiest perceptual task for all three data types.

The reason behind this ranking has to do with the way the human brain—the visual cortex in particular—is "wired." Let's take a closer look at some of the perceptual and cognitive principles that Stephen Kosslyn uses as the basis for discussion of graph design in "Elements of Graph Design."

The mind is not a camera. Salience, detectability, and discriminability are some of the most basic principles here, based on the fact that neurons are difference detectors, responding best to a *change* in stimulation (Kosslyn, 1994, p. 24). Therefore, the most salient features of the graphical display—larger marks, thicker lines, taller bars, etc.—will be noticed first and continue to command our attention as we look at a graphic. At the other end of the stimulus discriminability spectrum, differences that are too subtle might not be registered at all, given the amount of random "noise" that neural activity generates, as well as the fact that "neurons that detect edges inhibit each other," preventing a cell that would typically detect a stimulus from responding (Kosslyn, 1994, pp. 269). What all this means for graph design, then, is that visualizations that rely on difference rather than absolute value judgments for interpretation will be more successful in exploiting this ability of the human mind. Since differences in position along a common scale and length are among the easiest to detect accurately, it only makes sense that we try to use them first to encode quantitative data. This can be done effectively with scatter plots and bar charts. A word of caution about using differences in length: large percentage differences are easier to detect than small percentage differences even when the differences are the same in absolute terms (Cleveland, 1985, p. 241).

Ware cites research studies indicating that "the brain contains large arrays of neurons that filter for orientation and size information at each point in the visual field" (Hubel & Wiesel cited in Ware & Knight, 1995, p. 4). Each array can be thought of as an input channel. This means that in order to be easily distinguished, stimuli must activate different channels. Studies suggest that the angle difference of 30 or more degrees will ensure activation by a different channel, and therefore, easy discrimination. (Ware & Knight, 1995, p. 5). Similarly, easily discriminated spatial frequency ratio is between 10:1 and 4:1; and the total number of channels is estimated to be between four and ten (Ware & Knight, 1995, pp. 5-6). One implication of these findings are for design of textures, which Mackinlay identified as one the easiest ways of encoding nominal data: we must ensure that texture orientation and density differ enough to be easily discriminated and that the total number of textures used in a graphic does not exceed the number of input channels used to decode them.

Differences in input channel tuning also enable reading graphics at different levels of granularity. For example, in a line plot, finely-tuned receptors "read" local, high-frequency fluctuations in the shape of the line, while coarsely-tuned receptors attend to the general shape of the entire graph (Kosslyn, 1994, p. 5). This also explains Tufte's observation that adding detail does not reduce our ability to interpret the display, provided that the detail is "properly arranged" (Tufte, 1990, p. 37).

Experimental evidence shows that area and volume judgments suffer from perceptual distortion, with both being consistently underestimated (Kosslyn, 1994, p. 268; Cleveland, 1985, p. 243). This makes area and volume poor candidates for encoding quantitative, as well as ordinal and nominal data.

Hue, saturation, and density, while at the bottom of the list for encoding quantitative data—it is very difficult to judge color dimensions in absolute terms—can all be used to encode ordinal data. However, Ware cautions that in the case of hue, only short sections of the spectrum are perceptually ordered—those that increase or decrease continuously on both red-green and yellow-blue channels (Ware, 2004, p. 128). On the other hand, hue can be used to encode nominal data clearly. Unfortunately, studies show that only about 8 unique hues can be identified reliably (Post and Greene cited in Ware, 2004, p. 113), which constrains the use of hues as labels for nominal data. While providing a reference scale may increase this number (Morse, 1979, p. 97), it is still a good idea to "use colors that are well-separated in the spectrum" (Kosslyn, 1994, p. 162).

Another problem with color is that the three components of color—hue, lightness, and saturation—are not highly perceptually separable (Ware, 2004, p. 137). That is, we can't easily attend to one aspect of color and ignore the others. This makes color coding multiple variables in a single glyph independently of one another using different components of color much less effective. Ware suggests using highly separable dimensions like texture or height difference in conjunction with color to encode multiple variables (Ware, 2004, p. 137). Combinations of highly separable dimensions can be effectively used in layering graphical elements. For example, color and texture are used on geographical maps to increase information density without creating clutter and overwhelming the viewer.

One area where color, especially hue, can be used effectively is to relate and group *visually* the elements that are related *logically*: objects that share the same hue are perceived to be related. The Gestalt principle of similarity is at work here. However, this also means that objects that are not related in meaning, should differ in visual treatment.[2]

Prior knowledge: the mind judges a book by its cover. The set of principles that fall under this heading share the notion that "our visual and memory systems tend to make a direct connection between the properties of a pattern and the properties of the entities symbolized by that pattern" (Kosslyn, 1994, p. 8). For example, the principle of informative changes (or meaningful differences) states that the brain "registers differences and tries to interpret them" (Kosslyn, 1994, p. 269). The implication for graph design is straightforward: to highlight differences, make them visible, and vice versa. Also, accurately graphing the data—using scales properly, avoiding gratuitous decoration of data, etc.—will ensure that the detected visual differences accurately represent the differences in the data.

Compatibility is another "prior knowledge" principle, and occupies a central place in graph design. In simplest terms, compatibility requires that presentation match viewer's expectations (Morse, 1979, p. 95). For example, an increase in quantity usually occurs in the upward (rather than downward) direction, passage of time occurs from left to right, etc. Graphs should not violate these basic expectations. In addition, designing graph elements to match their real-world position, color, etc. can also help comprehension. For example, when an independent variable to be graphed is geographical in nature, it often helps to arrange the glyphs to mimic their position on a map (or to use a statistical map).

Going beyond the basic conceptual and cultural compatibility, there is task compatibility—a principle requiring that the graphical display should be tailored to the user's task. Alan Morse cites a study by Schutz showing that superimposed graphs facilitated comparisons, while separating the plots made reading data values easier (Morse, 1979, p. 96). More specifically, Morse argues that only the information that the viewer wants to know about should be displayed (Morse, 1979, p. 95). In support of this recommendation, Morse refers to a study by Posner who observed a decline in performance with the increase in the amount of information that needed to be reduced out (Morse, 1979, p. 96).

Working memory: the spirit is willing, but the mind is weak. Eliminating unneeded complexity also reduces cognitive load by relieving the viewer from holding irrelevant details in working memory. Some of the ways that the complexity inherent in the data can be reduced is are sampling, filtering, averaging, aggregating (Morse, 1979, p. 96). Tufte cautions, however, that the choice of intervals and end points used for aggregating data can significantly distort it (Tufte, 1997, p. 35-37).

---

[2] Perceptual grouping, while essential to graph design, is outside of the this paper's scope.

Reducing clutter—artificial complexity—by maximizing "data-ink" and reducing "chartjunk" is another good way of achieving clarity, as Tufte demonstrated by redesigning Tukey's box plots (Tufte, 1983, p. 125). The amount of allowable "chartjunk" depends on the viewer's task, however. For example, Kosslyn recommends reducing the number of tick marks, but not when specific values are important (Kosslyn, 1994, p. 88).

When multiple plots are graphed together, visual complexity can be reduced by plotting them in such a way that (a) they can be easily grouped at the perceptual level—through proximity, similar shape, etc.—and arranged so as to produce (b) the simplest and/or (c) most meaningful pattern (Kosslyn, 1994, pp. 196, 74). Doing this extends the viewer's ability to devote more cognitive capacity to analyzing the data, rather than perceiving it and holding it in memory. Sometimes, however, simplicity can only be achieved by splitting up a graph into two or more sections (as Minard did by placing the temperature plot in the Napoleon invasion graphic onto a separate, parallel "pane") or into separate plots (Tukey, 1990, p. 332).

Legends are yet another assault on WM: to understand the graph, the reader must first memorize the meaning of the symbols used in the graph (or constantly refer to the legend until memorization takes place). A better approach is to use direct labels whenever possible (Tufte, 1997, pp. 98-99). If, however, using direct labels results in too much clutter, then, a compromise approach may be appropriate: making the symbol resemble the thing it represents. This way, the symbol will not be an arbitrary mark, but a cue for recalling the variable it represents, reducing the need to refer to the legend, resulting in less WM load.

Choosing the best graph type.

So what is the ideal display type? Based on what we learned about the accuracy of elementary graphical perception tasks, the simple answer is: the display that will rely on the most accurate perceptual tasks for decoding the data type: position/length for quantitative (Cleveland, 1985, p. 254), position/density/saturation for ordinal, and position/hue/texture for nominal data (Mackinlay, 1986, p. 125). For example, both a pie chart and a divided bar can expressively convey the relationship of part to the whole. However, the divided bar is more effective in enabling the viewer to judge the sizes of the parts more accurately, because it requires length, rather than area judgments (Kosslyn, 1994, p. 28).

Sometimes, being able to encode data to be decoded by the most accurate decoding task requires transformations of data, or what Tukey would call precomputation. For example, if the task is to see changes in rates over time, it is better to calculate the *rate* and display it using a type of graph that relies on position judgments (such as the line graph), rather than plotting the *running total* using a line graph, asking the viewer to make judgments about very fine changes in the line's slope. Naturally, precomputation also helps with graphical-cognitive tasks by relieving the viewer from performing mental calculations. The key is to know what tasks the viewer will want to accomplish with the graph and precompute to make those tasks easier.

In addition to the viewer's perceptual abilities, the graph's purpose will also influence the choice of the graph type (and design). Tukey makes a distinction between two different functions that a display can serve. One is prospecting, which requires presenting the data to allow the viewer to explore the data for the presence or absence of phenomena (Tukey, 1990, p. 331). Another is transfer—designing the visualization to communicate phenomena (Tukey, 1990, p. 331). Strategies for designing prospecting (or exploratory) displays include maximizing data ink to let the data stand out, plotting multiple data sets together to explore correlations, and avoiding excessive data reduction (i.e. aggregation) to prevent data distortions. On the other hand, transfer displays require a certain amount of orchestrated visual impact to direct the viewer's attention to those aspects of the graph that are the most important in communicating the message (Tukey, 1990, p. 332). For these types of displays, Tukey challenges Tufte's obsession with maximizing data ink to the point that all impact of the graphic is all but gone: "[w]e do need to reduce busyness, but we must retain impact" (Tukey, 1990, p. 333).

The following table summarizes some of the "enabling qualities" of selected graph types (based on Kosslyn, 1994, pp. 24-56; Tukey's [transfer, prospecting] and Tufte's [chartjunk, data-ink] terms are also used):

Chart type	Supported task(s)	Advantages and disadvantages	Purpose and data
Pie	See approximate relative amounts	**Advantages:** high visual impact. **Disadvantages:** accurate comparisons are difficult because they're based on area (and angle) judgments.	*Purpose*: transfer *Data*: quantitative
Exploded pie	Emphasize a relatively small, single part.	**Advantage:** impact: the area stands out due to a broken contour; **Limitation:** can't be used effectively for emphasizing larger portions.	*Purpose*: transfer *Data*: quantitative
Divided bar	Estimate relative amounts with a degree of accuracy.	**Advantages:** judgments are based on length and can be made relatively accurately; the bar's area shading adds visual impact. **Disadvantage:** the visual impact of the bar's area may make length judgments less accurate.	*Purpose*: transfer *Data*: quantitative (bar length); nominal/ordinal (bar position).
Layer graph	See cumulative totals	**Advantage over divided bar charts:** can use quantitative data for an independent variable.	*Purpose*: mostly transfer *Data*: quantitative (bar length and position);

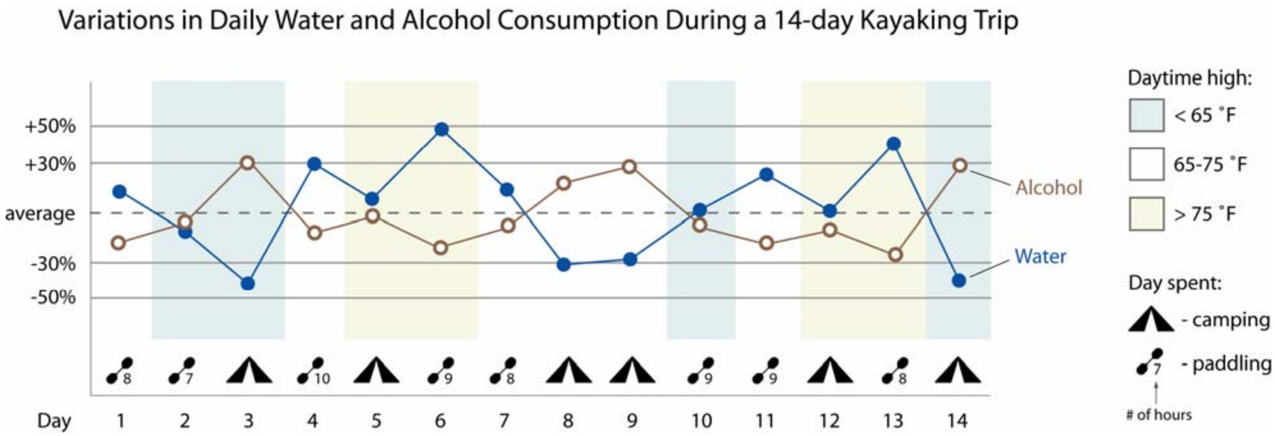
Chart type	Supported task(s)	Advantages and disadvantages	Purpose and data
Bar	See relative point values on a quantitative scale for entities on nominal or ordinal scales.	**Advantages:** high visual impact due to the bar's area; precise judgments can be made based on the position (of the bar ends) along a common scale. **Disadvantage:** too much non-data ink due to bar width and the rectangular shape that results.	*Purpose*: mostly transfer *Data*: quantitative (bar length); nominal/ordinal (bar position).
Side by side graph	Make pairwise comparisons	**Advantage over bar charts:** each pair of bars is naturally perceived as one perceptual unit (chunking), simplifying the graph (less WM load).	*Purpose*: transfer & prospecting *Data*: quantitative (bar length); nominal/ordinal (bar position).
Line	See variations along a continuous scale	**Advantages:** shows trends effectively by design (hence the use of trend lines on scatter plots); allows for precise judgments—based on position along a common scale.	*Purpose*: transfer & prospecting *Data*: quantitative
Scatter	Get an overall impression of the relation between two variables	**Advantages:** allows for highest data-ink ratio by design; allows for precise judgments—based on position along a common scale.	*Purpose*: mostly prospecting *Data*: quantitative

Multivariate considerations.

The perceptually and cognitively-driven guidelines we've already discussed apply to multivariate displays as well (i.e. the symbols and lines used to represent each data series must be easy to differentiate and axes easy to match with the corresponding data series). We've also already looked at a few applications of those specifically to multivariate displays (i.e. placing visually similar plots together to facilitate chunking). However, a few unaddressed issues remain.

The main criterion for deciding to plot two or more data series on a single plot should be that the data series are related and must be compared. After all, why plot unrelated sets of data? What's more, the graph should be designed so that these comparisons can be made easily and accurately. That is, we must enable "apples-to-apples" comparisons. One way to achieve this is to use an identical scale for all data series. For data that don't share the same scale, this may be possible through algebraic transformation of data (see the display created for this paper).

Since a multivariate display has the potential to show many sets of data simultaneously in the same physical area, data clutter can render the display useless. One way to cope with this is to use multiple panels or plots, arranged parallel to each other, in a matrix, or other meaningful way that makes direct comparisons possible while making it easy to differentiate between data sets. The thing these implementations have in common is that the overall framework and labels of each plot doesn't change—only the data changes. This way the viewer can focus on "the changes in the data, not changes in data frames" (Tufte, 1990, p. 67). Of course, plotting all of these small multiples "within the eyespan" puts less strain on WM, letting the viewer "make comparisons at a glance" for "uninterrupted visual reasoning" (Tufte, 1990, p. 67).



We now turn to the discussion of the multivariate plot, specifically dreamed up[3] and designed to illustrate many of the concepts discussed in this paper.

The main purpose: to allow the viewer explore the relationship between water and alcohol consumption throughout the trip, as well as see how consumption fluctuates with time, main daily activity (camping or paddling), the number of paddling hours (on paddling days), and daytime high temperature.

Hypotheses: higher temperature and paddling increases water consumption, lower temperature and camping increases alcohol consumption, and vice versa. In addition, more hours paddling increase water consumption. All hypotheses are substantiated by the data.

Data scales and data transformations. All four of Stevens' data scales are represented: nominal (daily activity), ordinal (daytime high ranges), interval (day, # of hours padding), and ratio (fluctuation in fluid consumption). The two data series (water and alcohol consumption variation) are plotted together on a single graph using an identical scale on both axes, making it easy to not only see how one fluctuates in relationship with the other, but also how much. In this particular case, the viewer can see that in percentage terms, alcohol consumption does not fluctuate as much about its average as water consumption does: the deviation from average of alcohol consumption never exceeds 30%, while that of water does. Expressing the variation in absolute terms (and corresponding units like gallons and ounces) would have made this kind of comparison impossible at a glance (by turning the comparison into a graphical-cognitive task).

Chart type: line. Line chart works well for showing changes along a continuous scale, such as time. It also allows relatively precise judgments of quantities. The addition of the dashed average line helps see how much above or below daily average consumption was on each day. The addition of the 30% and 50%

[3] The data used in the plot are fictional, specifically created to highlight the relationships being explored.

lines is also not arbitrary: the 30% line forms a boundary which alcohol consumption variation never exceeds; the 50% line serves the same function for water consumption.

Symbol design. Simple round shapes are used. The symbols for alcohol consumption have a white middle, making them distinct from water symbols. This allows the viewer to better focus selectively on each series if needed. Real-world color of water—blue—is used in the color of its symbol, making it easy to memorize. Since there are only two data sets, the viewer only has to memorize one to interpret the display without referring to the series labels. Color-matched, direct labels are used to label each series. Labels are placed next to the legend. This puts all data necessary to interpret the coding used in one place. The principle of pictorial realism is used in the design of the daily activity symbols—the symbols for each activity are iconic representations of a paddle and a camping tent—reducing WM load.

In addition to encoding main data series, color is also used to encode temperature ranges. Since the goal here is to let the viewer see how the temperature fluctuated about the average of 65-75 degrees, two hues at the opposing sides of the blue-yellow channel are used. Middle is signified by lack of pigment (white). This scheme is scalable to several more pairs of temperature ranges, if needed (by varying saturation and lightness). Real-world metaphor used: yellow, a "warm" color, is used to encode a warmer temperature range, while "blue," a cool color encodes the cooler temperature range.

Variation: what if transfer, rather than prospecting, was the graph's primary purpose? Fewer variables would be plotted on a single plot, to focus the viewer's attention on the phenomenon being communicated. Further precomputation, including properly isolating the effects of daily activity (paddling vs. camping), would also be required. For example, if we wanted to show how water consumption varied with temperature, we could have plotted, on a bar graph, water consumption per hour (bar height) vs. temperature (bar position), ordering the bars from low to high temperature along the x axis.

Bibliography.

Cleveland, W. (1985). *The Elements of Graphing Data.* Monterey, CA: Wadsworth.

Halstead-Nussloch, R. (1995). CS610 Research Methods – Measurement. Retrieved on March 20, 2005 from http://tapestry.spsu.edu/drrich/classes/cs6023/cs610mea.html

Kosslyn, S. (1994). *Elements of Graph Design.* New York, NY: W.H. Freeman and Company.

Mackinlay, J. (April, 1986). Automating the Design of Graphical Presentations of Relational Information. *ACM Transactions on Graphics, 5(2),* pp. 110-141.

Morse, A. (August, 1979). Some principles for the effective display of data. *ACM SIGGRAPH Computer Graphics, Proceedings of the 6th annual conference on Computer graphics and interactive techniques, 13(2)*, pp. 94-101.

Tukey, J. (August, 1990). Data-Based Graphics: Visual Display in the Decades to Come. *Statistical Science, Vol. 5, No. 3,* pp. 327-339.

Tufte, E. (1985). *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

Tufte, E. (1990). *Envisioning Information.* Cheshire, CT: Graphics Press.

Tufte, E. (1997). *Visual Explanations.* Cheshire, CT: Graphics Press.

Ware, C. & Knight, W. (January, 1995). Using Visual Texture for Information Display. *ACM Transactions on Graphics, 14(1),* 3-20.

Ware, C. (2004). *Information Visualization.* San Francisco: Morgan Kaufmann.