CHAPTER I

RATIONALE

Since the term “scientific visualization” was coined by a panel of the Association for Computing Machinery (ACM) in 1987 (McCormick, DeFanit, & Brown, 1987), both hardware and software developers have invested resources to enhance computer graphing capabilities. These developments can be seen as a logical extension of the work of John Tukey (1977, 1980, 1986a, 1986b, 1986c, 1988), who argued for the value of high dimensional graphics in social sciences.

While many scientific researchers believe that dynamic visualization is a promising data analysis tool for a variety of applications (Alexander & Winarsky, 1989; Scientific visualization comes down to earth, 1992), psychological studies concerning advanced graphical representation have been unable to support the effectiveness of these displays (Marchak & Whitney, 1990; Marchak & Marchak, 1991; Wickens, Merwin, & Lin, 1994). To address this discrepancy, this paper reviews (a) the cognitive and data analytical issues relevant to data visualization, (b) suggests an explanation for these different opinions, and (c) proposes and reports the results of an empirical test of this explanation. The central thesis of this work is that visualization is only effective when there is an appropriate alignment of data type, task type (research goal) and type of graphical representation.

Previous Research

Many guidelines on graphical format (e.g. Spear, 1969; Bertin, 1983; Tufte, 1983; Tufte, 1990) are based on authors’ intuitions instead of empirical findings (Spence, 1990). In addition, they are typically concerned with graphical presentation as communicating a result of the analysis rather than data visualization as a process of discovery. In the domains of cognitive psychology and ergonomics, researchers have examined particular perceptual or cognitive tasks such as pattern perception, semantic analysis and graph schemata (e.g. Cleveland, 1985; Kosslyn, 1989; Pinker, 1990). However, Lohse (1991) argued that the focus of most of these studies are too narrow to make generalizations to broader sets of phenomena. Other researchers emphasizing continuous information display in applications such as aviation, production process control, and failure detection also suffer limited generalization (e.g. Buttigieg & Sanderson, 1991; Sanderson, Flack, Buttigieg, & Casey, 1989; Sanderson, Haskell, & Flach, 1992).

Among such researchers, C. Wickens and his colleagues have established one of the most theoretically grounded and programmatic research endeavors (cf. Wickens, 1986; Carswell & Wickens 1987; Barnett & Wickens, 1988, Wickens, Todd, 1990; Goettl & Wickens & Kramer, 1991; Wickens, 1992; Wickens, Merwin & Lin, 1994). Building on his theory of the proximity compatibility principle, Carswell and Wickens (1987) predicted that integral displays are suited for integrative tasks while separable displays facilitate focus tasks. An integral display combines several dimensions in a single object whereas separable displays show data of different dimensions in different panels. Integral and separable displays are relative concepts. For instance, when there are bivariate data, a scatterplot is considered integral while two histograms are viewed as separable. If there are three variables in the data set, a 3D plot is treated as integral while three scatterplots are regarded as separable. In Wickens' view, the merits of integral and separable graphs depend on task nature. Tasks that require attention on a portion of the data, such as reporting a single datum value of a variable, are called focus tasks. On the other hand, tasks that aim to integrate information across several variables, such as locating an observation which has the highest values in several variables, are named integrative tasks. Based upon Wickens' theory, it is predicted that the more integrative the task is, the more an integral graph can help, and vice versa.

However, in several empirical studies this prediction was not fulfilled. Wickens interpreted these results as a failure of high-dimensional graphical devices in general. In this paper I argue that Wickens' negative results are not due to the general ineffectiveness of visualization procedures, but rather to the misalignment of data type, research goal (task type) and graphical type employed in his research. The supposition of this paper is that the quality of graphical display is dependent on the proper alignment of these three aspects of data visualization.

This paper consists of four sections. The first part briefly presents the alignment framework of data type, research goal, and graphical format. The second part introduces Wickens' theory and evaluates three of his studies in the light of the alignment framework. Next, an empirical study based on this framework is proposed. Finally, the results are reported and the implications and recommendations from expected results are discussed. A more detailed account of the alignment framework is presented in a review of the literature in Appendix A.

Alignment Framework

It is generally agreed that no graphical format can be appropriately applied to all situations. Therefore, there is a need to have a framework describing how various graphical formats are related to various situations. Until recently, no comprehensive framework was able to explain how crucial aspects of visualization contribute to effective scientific practice. I have recently proposed that effective data visualization is an outcome of the proper combination of graphical technique, research goal, and data type. In each of these three aspects, there are several sub-categories that must be considered as portrayed in Figure 1. Each of these aspects and sub-categories are discussed briefly below.

Taxonomy of Visualization Techniques

There are hundreds of types of graphs, charts and diagrams for statistical visualization. To make sense out of this large pool of graphical representations one can consider them in terms of dimensionality and noise-smooth (Yu & Behrens, 1995). As shown in Figure 2, visualization techniques can be considered to fall somewhere between two poles of the noise-smooth continuum, as well as the level of dimension. Noise level and dimensionality together dictate the complexity of the graph.

Dimensionality. The number of dimensions of data is a common criterion for classifying graphical formats (e.g. Watson & Driver, 1983; Yorchak, Alison, & Dodd, 1984; Lee & MacLachlan, 1986; Barfield & Robless, 1989; Spence, 1990). According to these researchers graphical formats correspond to the number of variables, and therefore, can be categorized as one-dimensional, two-dimensional, and multi-dimensional.

The level of data dimension conditions the choice of graphing technique. For instance, a single histogram is often an effective summary of one-dimensional data (Scott, 1992). With bivariate data a scatterplot is a logical candidate for visualization because histograms cannot depict bivariate functions. Multivariate data sets usually require even more sophisticated visualization techniques.

Dimension reduction is considered by some researchers an effective means to enhance graphical representation of multivariate data sets (e.g. Gabriel, 1981; Cook & Weisberg, 1994; Dawkins, 1995). Dimension reduction literally reduces the number of dimensions. For example, the Gabriel biplot, which is a visual equivalent to principal components analysis, collapses variables into a few factors to optimize the variance explained. Detailed discussion of dimension reduction is presented in Appendix A.

The focus of this chapter is on dimension integration because it is a coherent theme across studies by Wickens and his colleagues. Dimension integration retains all variables but portrays them by a single object for the facilitation of understanding related information (Carswell & Wickens, 1987; Barnett and Wickens, 1988). This strategy, named as integral display or object display, is widely used in multi-dimensional visualization techniques such as Chernoff faces, star plots, radar plots, 3D spin plots, and spider plots.

Noise-smooth continuum. Apart from dimensionality, noise level is another useful criterion to classify graphical displays. Noise-smooth can be thought of as how much summary versus raw data is portrayed. When accuracy of prediction is required, data analysts have to display every single datum point (e.g. Palya, 1991). In many situations, however, depicting all data is not desirable. As Kosslyn (1989) said, “too much information is as bad as too little information.” (p.189) Too much information may hinder the researcher from unveiling the underlying data pattern. Therefore, the researcher often faces the tension between retaining and sacrificing detail. In mathematical statistics much work has examined the noise level of graphical representation (Nadaraya, 1965; Bowman & Foster, 1993; Scott, 1992; Hardle, 1991; Silverman, 1986), which is closely associated with bandwidth and structure imposition.

The bandwidth is the basic scaling level in which observations are grouped in a graphical object. The more levels a graph has, the noisier it is. For instance, Figure 3 shows how bandwidth determines the complexity of a histogram. In the top panel, the histogram with a narrow interval (small bandwidth) shows many ups and downs across the bars. In order to obtain a different view of the underlying structure, one can use wider bandwith (fewer bars) to obtain a smoother histogram as shown in the two lower panels. Of course the danger here is that too much detail may be smoothed over.

In addition to bandwidth choice, structure imposition is another way of suppressing noise. A typical example is the superimposition of a regression line on top of data points in a scatterplot.

It is a common misunderstanding that higher-dimensional graphs are always more complicated than lower-dimensional graphs. This taxonomy helps to clarify that in each dimension noise level also contributes to complexity. For example, a 3D plot with a smoothed mesh surface may appear to be less complicated than three separate histograms with many intervals. This is because a 3D mesh plot portrays a function, as shown in the top panel of Figure 4, whereas histograms depict the raw data, as shown in the bottom panel of Figure 4.

In short, this taxonomy of visualization technique guides the researcher to contemplate questions such as: (a) how many dimensions do I want to display? Should I reduce or integrate several dimensions? (b) how much data should be depicted? (c) How do I balance noise versus summary? Decisions about dimension reduction, dimension integration, and noise suppression are often driven by the research goal and data type, which are discussed below.

Research Goals and Visualization

In addition to the preceding taxonomy of noise-smooth (raw data versus summary), I propose a classification of task type/research goal in terms of how much data need to be depicted. Generally speaking, there are six major categories of research goals as described in Appendix A. They are spotting outliers, discriminating clusters, checking distributional and other assumptions, examining relationships, comparing group differences and observing a time-based process. Only spotting outliers and examining relationships will be address here because they are examined in the proposed experiment. As the six tasks fall somewhere along the noise-smooth continuum, spotting outliers and examining relationships are located near the two opposite poles of the continuum i.e. spotting outliers requires focusing on all raw data whereas examining relationships focuses on a smooth structure underlying the data.

Spotting outliers. It is often helpful for a researcher to begin visualization with the goal of detecting outliers, because outliers generally need to be accounted for if parametric statistics are to be used. Moreover, this procedure should be implemented before discriminating clusters, checking assumptions and doing remedial data transformation for a number of reasons. First, some clustering procedures such as Ward's method perform poorly in the presence of outliers (Everitt, 1993). Also, the need for data transformation can be unduly influenced by one or two outliers (Atkinson, 1985). With the presence of outliers, some transformation methods may be ineffective to normalize data (Rasmussen, 1989). After removing outliers the distributional assumptions may no longer be broken. During exploratory stages of data analysis it is preferred to include all observations and thus a noisier graph is desirable.

Examining relationships. Once a researcher examines individual data points as done when detecting outliers, it is appropriate for subsequent research objectives to be more global such as determining relationships. In order to see a global picture of the relationships, the visualizer may suppress some noise by imposing a function on the data. However, this does not imply that the visualizer can ignore the raw data. Instead the analyst should plot both the raw data and the function to learn about their relationship. Hence, the visualization techniques applied to examining relationships can be both noisy and smooth.

In contrast to the above taxonomy, Wickens et al. (1990, 1994) classified task type in terms of the degree of information integration. However, his categorization is of little practical use because pure focus and medium integration tasks such as reporting data values are seldom used in data visualization, and his high integration tasks emphasize on individual observations. This will be more fully discussed below in the section regarding Wickens' research.

Data Type and Visualization

Besides research goals, data type also affects the choice of visualization techniques. There are at least four relevant aspects of data type, namely, origin, format, complexity, and distribution. Data format is concerned with measurement scale while data distribution is related to normality and non-normality. Scale and normality are not related to Wickens' studies, and therefore they are omitted from this section. Readers who are interested in data format should refer to Appendix A.

Data origin. Data origin is concerned with discipline-specific interpretation of the meaningfulness of data. Keller and Keller (1993) argued that the visualizer should put aside the concept of data origin and simply regard numbers as the computational materials for plotting graphs. However, data from different origins contain different inherent structures that should be taken into account in the choice of visualization techniques (cf. Behrens & Smith, in press). For instance, most of Wickens et al.'s studies are related to aviation. With this kind of continuous data, the economy and clarity of display is the primary criterion for a successful visualization. It is therefore not surprising that Casey and Wickens (1986) found Chernoff faces were inferior to bar graphs for data of continuous process. Because of the differences in type of data and goals across disciplines, it is doubtful whether the visualization techniques used in aviation can always be well-applied to data visualization in other subjects.

Data complexity. Data complexity is governed by the level of dimension, the number of observations, and the structure of the data. The need for dimension integration and noise filtering in graphics arise from the complexity of data sets. If all data sets are one-dimensional, contain as little as six to eight observations, and provide a nicely bell-shaped curve or a linear function, there will be little struggle between noisy and smooth displays. However, when the data set is multi-dimensional, contains hundreds or even thousands of observations, has non-normal distributions and non-linear relationships among variables, advanced visualization techniques should be used to deal with all three sources of data complexity.

Summary of the Alignment Framework

The proper combination of graphical technique, research goal and data type is crucial to a successful visualization. For example, if the task is to spot multiple outliers or discriminate clusters from a huge data set, a 3D plot that integrates three dimensions and depicts noisy data can be helpful. On the other hand, a smoothed graph such as a density curve or a histogram with large bandwidth runs the risk of hiding outliers. In light of the above taxonomies, we can predict under which circumstance particular techniques can succeed or fail. In the following section I will discuss the work of Wickens and his colleagues, and point out the mismatch of visualization technique, goal and data type in their research.

Wickens’ Theory and Studies

Proximity Compatibility Principle. Wickens and his colleagues conducted a series of studies concerning the appropriateness of graphical display, task type, and the integration of dimensions. They termed their theory the “proximity compatibility principle.” In this view, an optimal display should be both physically and perceptually proximate and compatible. Proximity is defined in terms of sharing of features between displayed attributes such as closeness in space, identity in color or similarity of semantic meaning (Carswell & Wickens, 1987).

Although Wickens’ primary interest is with displays used in aviation tasks, Wickens and Todd (1990) compared displays used in aviation to those used in scientific visualization in general. Also, most of their experiments employed scenarios outside of aviation such as financial analysis (Wickens, Merwin & Lin, 1994), economics (Wickens, Todd, 1990), and human performance evaluation (Goettl & Wickens & Kramer, 1991).

Separable Display, Integral display and Configural display. Carswell and Wickens (1987) classified graphical displays into two categories based on proximity, namely, separable and integral displays. In separable displays, different dimensions of data are presented in separate graphical panels, and thus there is a low degree of proximity among different dimensions. In contrast, integral displays combine several dimensions into a single graphical object, and consequently entails a higher degree of proximity. In their terminology, an integral display is also referred to as an object display. For example, when there are two variables, X and Y, one may show their values by two bar charts as separable displays, or compress the information into a single object--a scatterplot with X and Y axes, as portrayed in Figure 5. When there are three variables--X, Y and Z, one may use three separate pairwise scatterplots (X,Y; X,Z; Y,Z), or employ a 3D plot integrating all three dimensions, as shown in Figure 6.

If different physical dimensions in a graph correspond to a single cognitive code, a graph is considered integral. A successful collection of dimensions would bring out an emergent feature that facilitates data analysis. However, occasionally these dimensions may be too artificial to be blended mentally. In such cases a graph that maintains separate perceptual codes in different dimensions is said to be configural (Carswell & Wickens, 1990).

Serial Processing and Parallel Processing. According to Wickens, in an integral display, the dimensions are perceptually compatible i.e. cognitive effort devoted to one dimension would lead to the increase of cognitive resources invested into the other. However, if the graph is separable, cognitive energy spent in one dimension is at the expense of the others. Using a computer analogy, the cognitive processing of separable graphs is thought of as serial processing while the integrated perceptual processing is viewed as parallel processing (Carswell & Wickens, 1987). Although parallel processing seems to be a good thing for human performance, this ability can be detrimental if parallel processing is mandatory rather than optional (Wickens, 1992).

Focus Task and Integrative task. According to the proximity compatibility principle, the merits of separable and integral displays, as well as serial processing and parallel processing, are dependent on the task nature--focus tasks versus integrative tasks. If the visualization task requires a focus on independent dimensions such as locating individual data values, a separable graph that is likely to facilitate serial processing is desirable. If the job demands a judgment based upon integrated information from various dimensions, then an integral graphical object, which can enhance parallel processing, should be superior to a separable one.

Wickens and his colleagues conducted a series of experiments to verify these predictions. Throughout these experiments, the proximity compatibility principle was only partially supported. Some studies revealed a consistency between the theory and the empirical data (Carswell & Wickens, 1987, Barnett & Wickens, 1988), but some yielded opposing results (Wickens & Todd, 1990, Barnett & Wickens, 1988; Goettl & Wickens, 1991, Wickens, Merwin & Lin, 1994). Because the studies with negative results have been interpreted as negating claims for the validity of multivariate visualization techniques, three of these studies are now introduced in detail.

<u>Comparing bar graph, 3D plot, triangle display, and 3D surface</u>. In a simulation of economic analysis, Wickens and Todd (1990) employed four graphical formats and five types of information integration. The hypothetical data were the development value, lender value, and resource value of four nations. Separable displays were bar graphs and 3D plots whereas object graphs were triangle displays and 3D surface plots, as shown in Figure 7 and 8, respectively. Readers should be familiar with all preceding graphical formats except triangle displays. A triangle display, as shown in Figure 9, shows the value of three variables of one subject. The three points are connected to form a triangle and thus the object is named a triangle display.

Tasks assigned to participants in this study were categorized into five types according to the degree of information integration:

1. Pure focus task, e.g. What is the blue country's lender value?

2. Integration within a dimension, e.g. Is blue's development value larger than red's development value?

3. Integration within an observation, e.g. Do the three values of the red country's variables exceed a given range?

4. Integration across variables and observations by summing, e.g. Which country has the highest sum of all three dimensions?

5. Integration across variables and observations by examining relationships, e.g. As development value increases do lender values also increase?

The results only partially conformed to the proximity compatibility principle, which predicted a progressively greater advantage for more integrative displays. The performance which suffered the most with the bar graphs was the pure focus one. Also, it was found that 3D displays provided no additional advantages and so the authors suggested that 3D graphs should be used with caution.

Comparing bar graph, scatterplot and triangle display. In an investigation ostensibly simulating exploration of scientific data, Goettl, Wickens, and Kramer (1991) conducted three related experiments. In the first experiment two bar graphs and one scatterplot were employed to show two variables. In each panel there were only two observations. Participants were asked to predict the value of the third observation given that the function was linear. This task was considered integrative in nature since integration across observations and dimensions were required. Results indicated that scatterplots, which were integral displays, yielded better performance than separable bar graphs.

In the second experiment, the conditions were the same as before except that the task was focused i.e. participants were asked to report a portion of the data. Again, the proximity compatibility principle was confirmed because separable displays led to better performance than integral displays in this focus task.

In the last experiment bar graphs and triangles were compared. The values of three variables were shown as bars in the bar chart whereas the same three values were represented as three connected points in the triangle display. An integrative task was assigned to subjects. Given a formula and the values of three variables, participants were asked to estimate the value of criterion variable. However, the results contradicted the prediction derived from the proximity compatibility principle, because performance with triangle displays was not superior to that of bar graphs for this task.

Comparing stereopsis, motion, and mesh. In a recent study also aimed at simulating scientific visualization, Wickens, Merwin and Lin (1994) found that certain integral displays were not more supportive than separable ones in integrative analysis. This study consisted of two experiments. The first one was a comparison between 2D (separable) and 3D (integral) displays with six data points on each graph whereas the second one examined the effectiveness of stereo, mesh and rotation with eight data points on each graph. While the first experiment was similar to Wickens et al.'s previous studies, the second one carried new aspects of greater consequence. In the second experiment, three types of display were employed--stereopsis, 3D plot with the feature of rotation, 3D plot with a mesh surface. A still 3D plot was used as the control condition. The task types were also classified in terms of the degree of information integration:

1. Pure focus task, e.g. What is the earnings value of the blue company?

2. Integration across dimensions in one observation, e.g. Is the green company's debt value greater than its earnings value?

3. Integration across observations in one dimension, e.g. How much greater is blue's price than red's price?

4. Integration across dimensions and observations, e.g. Which company has the highest total value of all three variables?

It was found that the long term retention of abstract knowledge of the data failed to benefit from the 3D display exposure. Also, the rotation of a 3D graph and the presence of a mesh surface did not support performance in integrative processing. Wickens et al. stated that "it is an article of faith in many scientific visualization products that scientists should be able to explore their data interactively" (p.47) while in their empirical study animated motion did not provide any benefit for understanding data.

<u>Criticism to Wickens' Studies</u>

In reaction to these results, Sanderson and her colleagues (Sanderson et al., 1989; Buttigieg & Sanderson, 1991; Sanderson, Haskell, & Flach, 1992) argued that separable graphs such as bar charts could be superior to an integral display in an integration task if separable displays could constitute a strong emergent feature. This argument notwithstanding, I argue that the major problem of Wickens' studies is a mis-match of graph, task, and data.

<u>Pattern-seeking versus value-reporting.</u> Wickens classified research goal/task type in terms of degree of information integration. In practice, graphical presentations are rarely useful in pure focus tasks and medium integration tasks such as reporting single values. Several researchers pointed out that graphs should be used to convey an

overall pattern while tables are better for looking up data points (Ware and Beatty, 1986, 1988; Kosslyn, 1994).

It is doubtful whether focus tasks and medium integration tasks should be implemented with visualization at all. For instance, in the third study reported above (Wickens, Merwin & Lin ,1994) the focus question is: “What is the earning value of the blue company?” This question can be easily answered by simply looking up the values on a spreadsheet. The medium integration questions is “How much greater is blue’s price than red’s price?” By looking at the graph only, one can hardly tell what the exact difference between the blue’s and red’s price is. It is more accurate to subtract the two values from a spreadsheet. Even in the supposed high integration task, the focus is still on individual values rather than the relationships among variables: “Which company has the highest total value of all three variables?” Indeed it is more efficient to sum the values of all three variables, sort the data by summed values, and then list all the cases.

Although a 3D spin plot can be used for closely examining particular observations such as detecting outliers, the major concern of spotting outliers is the relative distance between the extreme cases and the majority of the observations. In other words, the analyst cares more about the overall data structure than individual cases. However, in Wickens et al.’s experiment, the focus on individual observations led subjects away from the global picture. Accordingly, the graphical format of the 3D spin plot and the task type of value-reporting are misaligned and poor performance should be expected.

Similar concerns arise when Wickens employed the 3D mesh. A 3D mesh is usually a smoothed surface. However, in Wickens et al.’s graph only individual data points are connected. It is qualified to be a surface plot or a perspective plot, rather than a mesh plot. A surface plot is not appropriate for aiding data value reporting, because in a large data set the surface would appear to be rough and tracing the exact co-ordinate is difficult. Even in a small data sets as Wickens et al. portrayed, exact co-ordinates are difficult to perceive. Again, Wickens et al.'s experiment tasks represent a misalignment of research goal and graphing technique.

Tasks given in Wickens et al.'s experiments, ranging from pure focus to so-called high integration, are not compatible with the goal of pattern seeking in data visualization. Given all the preceding misalignments, it is not surprising that Wickens et al. found no advantage of 3D spin and 3D mesh, because they are the wrong tools for the assignments that concentrate on individual values.

<u>Small number of data point.</u> It is important to note that in most of Wickens' studies the data sets in the experiments were as small as two to eight. As mentioned in the taxonomy of graphical format, the choice of visualization techniques are often driven by the goal of handling massive and messy data. Triangle displays in the first and second studies have little practical value because in most situations either the number of variables or observations would exceed three.

For example, in the third experiment discussed above, Wickens reported no advantage for use of a 3D plot. This is not unexpected under the circumstances described by Wickens, because 3D plots were developed primarily to solve the problem of overplotting and perspective limitation in multiple dimensions. In this case of small sample size, no overplotting and viewpoint obstruction occured and the advantages of the 3D plot could not be gained. In this way Wickens' tasks represent a misalignment of the plot made for high volume data with a very small data set.

The size of the data set is also an issue determining the appropriate use of a mesh surface. When there are many observations in a 3D plot, the perception of trend, which is based upon our perception of depth, is not easily formed. In this case connecting neighboring points to construct a mesh surface is helpful because it can provides depth cues. Moreover, if there are many peaks and holes in the plot, a smoothed mesh surface derived from a function is also desirable. In the third study reported above Wickens et al. (1994) found no advantage to this plot, but again, they had not properly aligned the plot with the type of data set size appropriate for the plot. An advantage of the type Wickens might have expected is only likely to occur when the size of the data set is large--an attribute that did not hold when only eight data points were used.

Summary of Criticism

As mentioned before, a successful visualization results from the proper alignment of graphical representation, task type, and data type. In the first study if Wickens had used forty instead of four data points, and the integrative tasks were concerned with examining relationships among variables rather than reporting the values of a particular nation, I would expect that the 3D mesh surface would yield superior performance to 2D displays. In the triangle display used in the second study, only one value of each dimension can be presented at one time. Due to this limitation, the appropriate task can only be estimating the value of criterion variable one by one, rather than estimating an overall pattern of the outcome variable. If Wickens had used a 3D mesh surface plot, which could depict more than one data values at once, and had altered the task to be the estimation of an overall pattern, the results would be consistent with the proximity compatibility principle. Again, in the third study, if Wickens had used forty data points and the integrative tasks were concerned with overall patterns rather than one particular observation, 3D rotation and mesh surface would be found superior to other displays.

Extension of Wickens' Studies

This study tested these suppositions by exposing subjects to aligned and mis-aligned graph-task-data situations. This study is an extension of Wickens et al.'s third study (1994) with three modifications. First, the interest of this study is on 3D spin and 3D mesh plots rather than stereopsis displays. This is because stereopsis displays are rare and the availability of 3D plots is increasing greatly. Second, the data sets to be used in the experiment will be composed of eight, fifty and one hundred cases, respectively, in contrast to only six to eight cases in Wickens et al.'s study. This is important because a large data set is more realistic in data visualization work and effective plot use depends on appropriate alignment with data set size. Third, as in Wickens' studies the task types (research goals) will also be classified as obtaining an overall impression of data or dealing with individual cases, the latter task type will be spotting outliers instead of recalling values. As mentioned above, focused and medium integration tasks such as reporting values should be accomplished with the use of tables. Therefore in this study spotting outliers, which is a highly integrative task, will be used instead. Outlier detection is considered highly integrative for two reasons. First, in a multivariate case, outliers are defined in terms of their coordinates in more than one dimensions, and therefore spotting outliers demands a high level of information integration. In addition, for detecting outliers an analyst is interested in the relationship between the extreme cases and the majority of the data. Thus, the task is global in nature.

Purpose and Hypotheses

In light of the alignment framework presented above, the following hypotheses were tested:

1. It is expected that in small sample sizes, 2D scatterplots would have equal performance with 3D spin and mesh plots in both outlier detection and relationship examination while performance is measured by an objective test.

2. It is anticipated that superior performance would be found for 3D spin and mesh plots for outlier detection across medium and large data sets.

3. It is hypothesized that 3D mesh plots are superior to 3D spin and 2D plots for examining relationships across medium and large data sets.

CHAPTER II

METHOD

Design

This study used a 3 data-size X 3 plot-type X 2 research-goal completely within-subjects factorial design as illustrated in Table 1. The factor of data size has levels of small (eight cases), medium (50 cases), and large (100 cases). The factor of graphical format includes 2D, 3D spin and 3D mesh plots with spotting outliers and examining relationships as the levels of the research goal/task.

The alignment framework suggests differential effectiveness across graphs at each level of research goal/task and sample size. In the small sample size conditions, performance is expected to be equivalent across graphs. In the medium and large sample sizes performance is expected to be inferior for 2D plots versus the other plots. Therefore, the hypotheses and analyses of interest are based on specific interactions of graph type and sample size rather than main effects.

Subjects

The sample is twenty-three graduate students with prior coursework in quantitative research methods, including low-dimensional data visualization, at Arizona State University. Participation was voluntarily. The mean age of the subjects was 33.2 with a standard deviation of 8.29. Six of the subjects were males while fourteen were females. The average age of the male group was 34, with a standard deviation of 9.46. In the female group, the mean was 32.86 and the standard deviation was 8.09. The age distribution by sex is presented in Figure 10.

Materials

Data sets. Nine different fictitious data sets were randomly constructed for the experiment. All were tri-variate data sets with three each consisting of one hundred, fifty, and eight cases from each of the following functions:

$$Y = 3.05372 + (-.05266\ X1) + (.21302\ X2) + (-.12967\ X1\ X2) + e \quad \text{(Function A)}$$

$$Y = 3.15458 + (.05929\ X1) + (-.01096\ X2) + (-.14112\ X1\ X2) + e \quad \text{(Function B)}$$

$$Y = 3.36628 + (.10321\ X1) + (.14369\ X2) + (-.29778\ X1\ X2) + e \quad \text{(Function C)}$$

These functions are depicted in Figure 11. In the above equations *e* stands for random errors. The initial random numbers were generated by the function *normal-rand* in XLISP-STAT with a mean of 0 and a standard deviation of 1. However, these values were too small and adding them to the model did not make the data points scatter. Therefore, a modification of the random error function was adopted as follows:

(floor (* 15 (normal-rand *n*))))

In this experiment participants were asked to spot outliers and examine relationships. Thus, the data sets were designed to match the task nature. As indicated in the preceding functions, two-way interaction effects were present in the data sets in order to make the conditional relationships among the three variables fairly complicated. In other words, the relationship between two variables varied across levels of the third variable.

To counter a response bias toward finding outliers, in two of the nine data sets pseudo-outliers were inserted. Pseudo-outliers are defined as observations whose values in two or more variables are about one and a half standard deviations above or below the means. In the rest of data sets, one outlier and one pseudo-outlier were added. For this case, an outlier is defined as an observation whose values in two or more dimensions are at least two standard deviations above or below the means. Because some extreme cases are outliers in two dimensions while some are outliers in three dimensions, without rotating the 3D plot it would be difficult to detect their existence.

Displays. Each data set was displayed using three types of graphs on Apple Macintosh computers. These graph type were:

1. A set of 2D scatterplots portraying three variables in a pairwise manner as seen in Figure 12.

2. A 3D plot portraying a cloud of data points with a spin option as shown in Figure 13. Four control functions were available in the spin plot for the viewer to find out the relative positions of data points. A “rock” button, which was located at far left corner of the bottom control bar, enabled the user to make the graph look as though the data were being shaken. When clicking a “pitch” button, the user could make the plot lean towards or turn away from the viewer. The “roll” feature allowed the user to move the plot clockwise or anti-clockwise. Last, the “yaw” feature served the function of moving the plot in a circle to either the left or right direction.

3. A 3D plot with a mesh surface conforming to the underlying function is shown in Figure 14. Because a mesh surface is perspective-dependent, the spin functions such as “rock,” “pitch,” “roll,” and “yaw” were also provided.

Procedure

The experiment was administered in several computer labs at the Arizona State University. Students were currently enrolled in graduate statistics courses and recruited by appeal from a graduate student. No course credit was given though they were informed they would be helping a student complete a dissertation and would gain additional experience with multivariate visualization. One to five subjects participated in the experiment at a time. Upon arrival, participants were asked to read an informed consent form and sign it if they agreed to participate. All participants accepted the terms and conditions. Next, the researcher explained the instructions to the participants and demonstrated how to use the software. Participants could proceed to the simulation when they indicated that they understood the task nature and the software features. The researcher encouraged participants to ask for help as needed. During the data collection the researcher observed whether individuals had difficulty in performing the task or using the software.

In this experiment I will refer to a combination of graph(s) and question as a scenario. Subjects were exposed to all eighteen scenarios as described in the design section. In order to avoid carry over effects, the data used in each scenario were randomly drawn from data sets created by the three functions described earlier. The order of scenarios was likewise randomized for each individual.

For each scenario, subjects viewed a graph or several graphs, and a dialog box with a multiple-choice question. They were told to answer the question according to the information shown on the graph(s), and were permitted to manipulate the graphics as appropriate to the graph type. After an answer was selected, another set of graph(s) and problem were presented. The process ended when all eighteen conditions were exhausted. The subjects were allowed to explore the data and answer the online question for about thirty minutes. Afterwards, they repeated the same procedure with a different sequence of the scenarios and different data sets drawn randomly. The entire process took approximately one hour.

Instrument

An objective test, which was embedded within the computer program, was given to the subjects as they were exposed to each condition. The test consisted of eighteen questions--one for each scenarios. The formats of Question 1 to 9, which was concerned with outlier detection were the same but the outliers varied. For the task of spotting outliers, the graph showed one observation each colored red and green. Participants chose one of the following answers: (a) the red observation is an outlier, (b) the green observation is an outlier, or (c) there is no outlier.

For the task of examining relationship, participants were given values of two variables and asked for a third. Again, the subjects had three choices: high, medium, and low. Questions of examining relationships (Question 10 to 18) are as follows:

10. What is the value of Z when both X and Y are high?
11. What is the value of Z when X is high and Y is low?
12. What is the value of Z when X is high and Y is medium?
13. What is the value of Z when X is low and Y is high?
14. What is the value of Z when both X and Y are low?
15. What is the value of Z when X is low and Y is medium?
16. What is the value of Z when X is medium and Y is high?
17. What is the value of Z when X is medium and Y is low?
18. What is the value of Z when both X and Y are medium?