CHAPTER IV

Discussion

The results largely conform to the predictions derived from the alignment framework and substantiate the effectiveness of 3D graphics for data visualization with appropriate data and tasks. These findings counter the conclusions of Wickens et al. (1994) that 3D graphs do not have an advantage over 2D graphs for scientific visualization. In this study 3D graphics outperformed their 2D counterparts for both outlier detection and relationship examination across all sample sizes. While I hypothesized that the superiority of 3D displays would be realized in medium and large data sets only, it turns out that the superiority of 3D graphics was also present in small data sets.

Outlier Detection

When the mean scores of two tests were used as the outcome variable, the result did not match the prediction that identical performances between 2D and 3D graphics would be found in the small data set. Even in this case, 3D mesh was the best among the three graphical formats (see Figure 16a). One possible explanation is that when there were only a few data points, both 2D and 3D spin plots fail to suggest a cluster of the majority of data. As a result, the examinees could not find a reference to tell which observation is an outlier. In contrast, the outlier “popped up” when a mesh surface was used as a reference. Nonetheless, this conclusion is tentative because inconsistent results occur when two tests were examined separately. According to the logistic regression with exact test, in Test 1 graphical formats did not significantly differ in the small data size ($p = .1334$).

In both medium and large data sets, the results conform to the prediction that 3D graphics are better than 2D ones for spotting outliers. In both cases 3D mesh was as effective as 3D spin. Based on my review of the literature, it was expected that mesh surfaces are used primarily for examining relationships rather than detecting outliers. Nonetheless, it seems that users found the mesh surface helpful as a depth cue, which is an important feature for determining distances among observations.

It was found that in 2D plots, performance of outlier detection decreased as the sample size increased (see Figure 16b). This outcome may be understandable. When there were just a few observations on the three 2D scattergrams, it was easier to determine which observations departed from the rest of data. When the sample size increased and the three 2D plots were jammed with dots, without rotation it was difficult to tell the actual distance between the suspected outlier and the majority in all dimensions.

Scores of Test 1 and 2 were computed separately for confidence intervals (see Figure 17). In the small sample size of Test 1, performance among the three types of graph were indistinguishable. In the medium sample size performance of 3D spin was even better than that of 3D mesh plots. However, in Test 2 3D mesh plots led to superior performance than 2D plots in both small and medium sample sizes. Also, in the small data size of Test 2 the difference between 3D spin and 3D mesh plot approached significance. Further, in the medium data size of Test 2 the relationship between 3D spin and 3D mesh plots was opposite to that in Test 1 i.e. 3D mesh plot outperformed 3D spin plot. This is consistent with the idea that students improved their skill for interpreting 3D mesh plots after some practice.

<u>Relationship Examination</u>

Like the result in the outlier identification, in the small data set 3D mesh plots surprisingly outperformed 3D spin and 2D plots in a significant manner for the task of relationship examination (see Figure 18a). It is possible that in the small data set, data points alone failed to suggest a relationship because too few observations scattered apart from each other. However, a mesh surface could smooth those data points and show the patterns among variables no matter how sparse the data points were.

Nonetheless, expectations were fulfilled in both medium and large data sets. According to correlated t-tests, 3D mesh plot was the best for the task of relationship examination, 3D spin, the second best, and 2D plots, the worst. In the large data set, the mean score of 2D plot was slightly higher than that of 3D spin plot. The possible explanation of the equal performance between 2D and 3D spin plots in the large data set is that when there were adequate observations on the graph, the pattern became obvious even in 2D scattergrams.

It was found that virtually in all sizes of data sets, performance improved as the sample size increased (see Figure 18b). The exception is that in 2D plots scores of small and medium sample sizes were exactly equal. One explanation is that the large data set provided enough observations to form a more obvious function while the small and medium data sets failed to suggest a pattern.

It is interesting to notice that in both outlier detection and relationship examination, sample size did not make a significant difference in 3D mesh (see Figure 16b and Figure 18b). However, the large data set was still the best, the medium, the second, and the small, the worst. It is concluded that regardless of sample size, a mesh surface was helpful in examining relationships among variables and also detecting multiple outliers.

The learning effect that occurred in outlier detection was not found in relationship examination. To a large extent the patterns in

Test 1 were equivalent to those of Test 2 (see Figure 19). It suggests that the use of high-dimensional graphs were more intuitive for the research goal of relationship examination than for outlier identification. This phenonmenon may also be caused by random errors from randomization of scenarios and data sets. Subjects might encounter difficult exercises in the second test with no improvement of scores.

The alignment framework anticipated that for discovering patterns such as relationship examination, 3D spin plots should be more useful than 2D plots as the data set increases the number of observations. However, this prediction was not fulfilled. As shown in Figure 19, the equal performance of 2D and 3D spin plots in the large sample size of both tests suggests that adequate data points could unveil the function regardless of graph type.

It is important to note that the underlying functions of data play an important role in determining the usefulness of high-dimensional graphs. If the function is linear, 2D scattergrams may be as effective as 3D graphs in spotting outliers and determining relationships. The complication of functions leads to visual obstacles in 2D graphs, and therefore 3D graphs compensate this weakness. In other words, the conclusion of this study should not be over-generalized to data sets which do not have non-linear functions.

Nevertheless, most of the findings in this study confirm the notion that the usefulness of visualization technique is tied to the task nature and the data type, which is predicted by the alignment framework. Although the effectiveness of 3D mesh in the small data set was not foreseen, this result further supports the alignment framework rather than weakening it. At first it was argued that the ineffectiveness of 3D graphics in Wickens et al.’s studies (1994) is due to the use of a few observations on the graphs. However, this study counteracted Wickens' conclusion because even with a small data set 3D graphics are still superior to 2D ones.

Limitations

This study has several limitations. First, because subjects for this study must meet certain requirements in order to participate, the sample size was inevitably small. Data visualization is not a common topic in most introductory statistics courses. This made recruitment of qualified subjects more difficult. Also, all subjects for this study were majoring in social sciences. The homogeneous backgrounds of subjects hinder the conclusion from generalizing to physical science majors.

Moreover, although subject responses were stable across the two tests, inconsistency was found in the responses to Test item 11. After a careful examination, I found no obvious defect in the question design. This phenomenon may be due to chance fluctuations or a real design problem. Further data collection should be made to refine the instrument, if necessary.

Last, the setting of this study is artificial--the data were fictitious; the location of outliers were arbitrary; there was a time limit to complete the task. In authentic data visualization, usually the data sets are much larger and it requires a long time to dig out insights. Given a much large data set and a longer time, the results may be very different.

Recommendations

In light of present findings, more research should be conducted to clarify the following issues:

1. For the task of outlier detection, performance with 2D graphs declined when the data size increased. When scores of Test 1 and 2 were analyzed separately, sample size effect for outlier detection became

inconclusive. More levels of repeated measures should be made to check the consistency of data size effect.

2. For examining relationship, 2D scattergrams and 3D spin plot had equal performance in the case of a large data set. This phenomenon is outside the expectation of the alignment framework. The efficacy of 2D and 3D spin plots for relationship examination should be further investigated with the large data sets.

3. In this study 3D graphs were found to be effective when two-way interactions were used to generate data. Future research may employ data sets with linear, spline, polynomial, curvilinear and other functions to further evaluate the worth of various graph types under different circumstances.

To compensate the limitations of this study mentioned earlier, future research may adopt the following strategies:

1. In future investigations, more subjects should be recruited to improve the statistical power of the tests. In addition, subjects may be recruited from a wide variety of disciplines. Comparing performance in data visualization between social science and natural science majors may bring more insights to research of statistical education.

2. In future research authentic data rather than data generated by functions may be used. Also, the number of scenarios may be reduced in order to let users spend more time for a few tasks. Results of this study can be verified by realistic simulations of data analysis.

As indicated in Figure 1, the alignment framework is composed of three major aspects and in each aspect there are many sub-categories. This study only tackled a corner of all possible combinations: With regard to graphical format, the noise level was addressed by using noisier scatterplots and smoother mesh surfaces. However, in dimensionality only 2D and 3D, but not 1D graphics, were included in this study. Pertaining to the six research goals, only spotting outliers and examining relationships were selected for evaluation. With regard to data type, the data origin was assumed constant in this study and the data format was continuous. Although the number of observation varied in this experiment, another factor of data complexity, the number of variable, was ignored. Also, the data sets used in this study had normal distributions. Obviously, there is a lot of room in the alignment framework for further investigation. Future examinations of the alignment framework may address the following issues:

1. In this study only two tasks, outlier identification and relationship examination, were chosen. Future work could compare 3D spin plots and other cluster detection techniques introduced in the literature review.

2. In this experiment the level of data complexity was altered by varying the number of observations. All graphics employed in this investigation were somehow confined to three variables only. However, 3D mesh surface can be extended to four dimensions by applying animation. Animated mesh surfaces are discussed in the literature review (Appendix A). The efficacy of animated mesh surface and other high-dimensional visualization techniques should be examined as well.