

Framework for Comparative Research on Relational Information Displays

Sung Park¹ and Richard Catrambone²

School of Psychology

&

Graphics, Visualization, and Usability Center (GVU)

Georgia Institute of Technology

ABSTRACT

We identify critical issues in comparative research on relational information displays (RIDs). The key argument is that when conducting an analysis of the cognitive process of people viewing different displays, their perceptual processes must be held constant so that they do not affect the results. We propose that in order to help researchers more easily compare display types (e.g., graphs) for how effectively they convey information, two factors must be considered. First, each element (e.g., each bar in a bar graph) in graphs that are being compared has to be equally discriminable. Second, the number of elements in the graphs being compared has to be the same; the maximum number of elements is limited by from the graph that uses a presentation format (e.g., density) that has the fewest number of discriminable levels. We present a psychophysics experiment that identified differential discrimination thresholds for density levels.

CR Categories: I. 3. 6 [Computer Graphics]: Methodology and Techniques-Ergonomics

Keywords: relational information display, human factors, psychophysics

1 INTRODUCTION

People often have trouble finding the right information when looking at graphs. Sometimes this leads to frustration, causing people to discontinue their search or resulting in them acquiring false information due to the poor display design. A clear understanding of which display is best suited for various types of information enhances the accuracy and efficiency of the user's task.

A variety of graphic displays have been used to represent different types of information, such as line graphs, bar charts, pie charts, scatter plots, matrices, tables, networks, maps, and many others. These displays can be categorized under a common name, relational information displays (RIDs), which are displays that represent relations among types of information (e.g., a bar graph including a person's name on the x-axis and her finish in a race on the y-axis) [1].

To calibrate the effectiveness among different RIDs, well-designed comparative research is required. In this study, critical issues in designing comparative research on RIDs will be discussed by analyzing prior RID research. This will lead to an argument for why additional grounding research in psychophysics is critical. Finally, a psychophysics experiment was conducted that provides initial foundational data.

2 THEORY

Many studies were conducted on RIDs (for a review, see [2]). Some studies are considered as foundational work because they provide basic understanding on how to leverage our capabilities and limitations of visual perception so that we can design effective graphs or displays. Of this research, we focus on studies on displays associated with different types of data (data conveying different scale types such as quantitative, ordinal, and nominal).

2.1 Ranking Quantitative Tasks

Cleveland and McGill [3] developed principles for statistical graphics. They developed the ranking of quantitative tasks based on psychophysics and their own empirical study (See quantitative ranking in Figure 1).

In psychophysics, Steven's Law portrays the relationship between an actual physical magnitude, x , and its perceived magnitude, $p(x)$:

$$p(x) = c \times x^\beta$$

The β weight differs from the particular physical magnitude being perceived. Experiments indicate that the average β for length judgments range from 0.9 to 1.1; for area, from 0.6 to 0.9; for volume, from 0.5 to 0.8. This difference indicates that people in general are more accurate at length judgments than area and volume judgments.

Cleveland and McGill also conducted an empirical study where participants were asked to judge the percent difference of each display. The result supported the quantitative ranking in Figure 1.

¹ email: gtg116s@mail.gatech.edu

² email: rc7@prism.gatech.edu

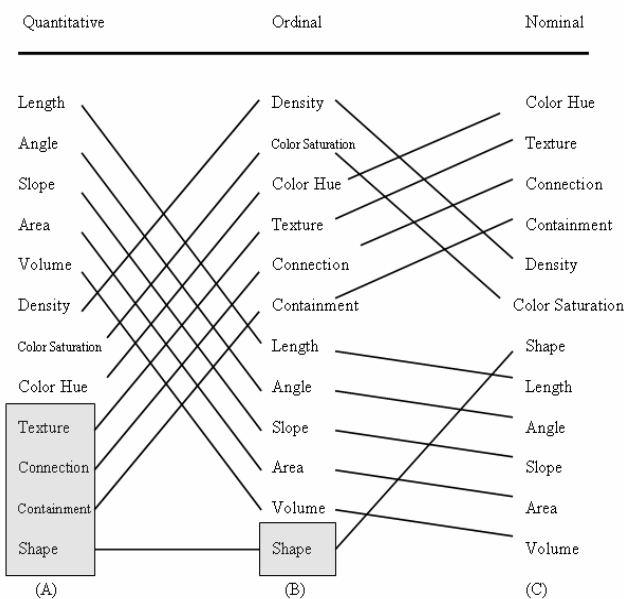


Figure 1. Ranking of different approaches for displaying information. The higher an approach is in the column, the better its accuracy. The approaches shown in the gray boxes are not relevant to these types of data. The ranking of quantitative data (A) is analyzed by Cleveland and McGill [3] and the rankings of ordinal and nominal data (B and C) are investigated by Mackinlay [4].

2.2 Ranking Non-Quantitative (Ordinal and Nominal) Tasks

Mackinlay [4] conducted a psychological analysis of displays presenting ordinal and nominal information to complement Cleveland and McGill's study (See ordinal and nominal rankings in Figure 1).

He suggested that density conveys ordinal information better than length primarily because length judgment suffers from the distance effect while density does not. For example, in Figure 2, the comparison of size (ordinal information) in lines (length) is easily perceived in (A) but is more difficult to do so in (B). On the other hand, density judgment does not suffer from the distance effect.

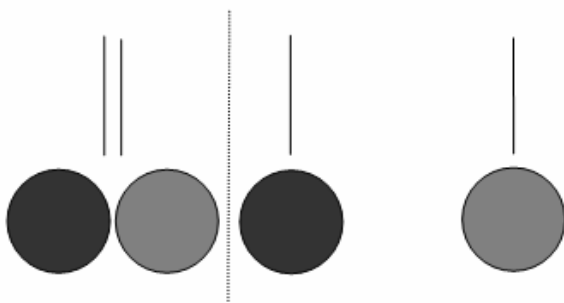


Figure 2. An example of the distance effect

Due to this distance effect, ordinal size¹ perception has a severe stepping limitation meaning that the maximum number of values that can be encoded without the distance effect is limited. Finally, color hue was best for nominal information because color perception is highly selective and known to be processed in the preattentive stage of the visual process. This enables an accurate distinction among nominal values.

Combining the results of Cleveland and McGill [3] and Mackinlay [4], an effectiveness ranking for different approaches for three types of information (quantitative, ordinal, and nominal) can be created (See Figure 1). It should be noted that in the original rankings, the two studies placed *position* as the best approach across all three information types. However, this approach is not considered in the present analysis because the position approach is known to be processed by a different system and this study wanted to control for this factor. That is, there are two separate systems for optic processing: visual and spatial information processing [5]. While visual information processing follows a projection from the occipital to the temporal cortex, spatial information processing follows a projection from the occipital to the parietal cortex. Accordingly, the present study focused on graphical presentations that are processed by the visual information system.

2.3 Represented and Representing Dimensions Study

Zhang [1] suggested that the mapping between the type of display and type of data can be studied effectively by the distributed cognition framework because cognitive activity in RID related tasks is distributed into two different representations: an internal representation (propositions, schemas, productions, mental images) and an external representation (physical symbols, graphs, objects). For a review on distributed cognition, see Zhang & Norman [6].

In an example of a RID task (see Figure 3), the density of the circle codes the value it represents. Viewers are told that the denser the circle, the more value it represents. By visual inspection of both external representations, the perceptual difference that (a) has more density than (b) can easily be identified. However, the decision of whether (a) represents more value than (b) can be made only by referring to the internal representation (i.e., a rule that has been learned) that the denser the circle, the more value it represents. Thus, RID tasks relate two distributed components: an internal representation and an external representation.

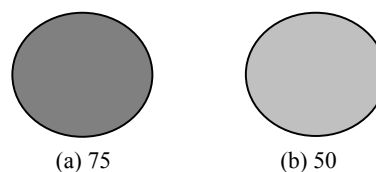


Figure 3. An example of a RID task. Each display represents an assigned number and follows the given rule that the denser the circle, the more value it represents.

One of the most promising methodologies that has been adopted in distributed cognition is *representational analysis* which investigates representational effect. The representational effect refers to the phenomenon that different representations of a common formal structure can cause dramatically different cognitive behaviors.

Zhang [1] argued that representational analysis could

¹ Size indicates elementary tasks such as length, angle, slope, area, and volume collectively.

effectively investigate RIDs as well. For each type of data (quantitative, ordinal, nominal) different features (length, density, color) conveying isomorphic information can be compared to see how they lead to differences in cognitive processing. Accordingly, he provided a theoretical framework that accounts for the representational effects of all RIDs (See Figure 4). The framework suggests that although there are no optimal displays that are efficient for all types of tasks, there can be a correct or incorrect mapping between the representation of a display and the structure of a task.

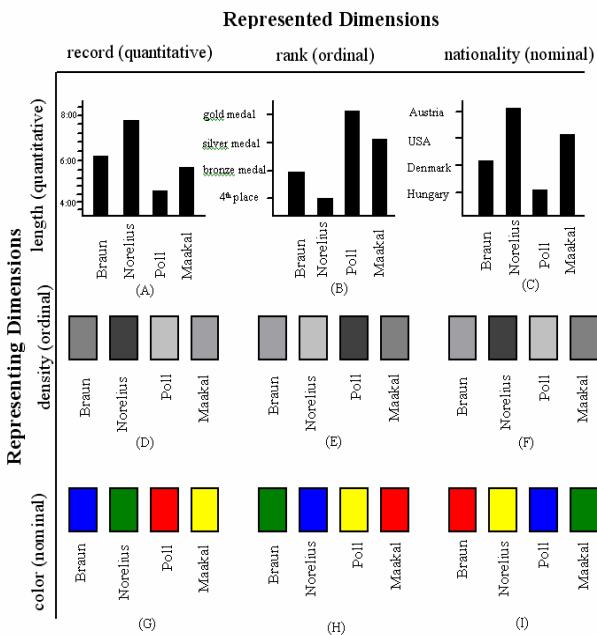


Figure 4. The mapping between the represented and representing dimensions. The scale type of each entity is inside the parentheses. Records are from a fictional women's 400 meter freestyle swimming competition.

There are two dimensions: represented and representing dimensions. The represented dimensions of a RID are the data (e.g., times) while the representing dimensions are the display (e.g., length of bar). Zhang [1] argued that in order for a representation to be efficient and accurate, the represented and representing dimensions should match in scale (A, E, I in Figure 4).

In an attempt to examine the implications of Zhang's framework, Park and Catrambone [7] conducted an empirical study. The following three predictions were tested to verify the key hypothesis that when the represented and representing dimensions match, a person's performance will be best:

- Performance in retrieving quantitative information is optimal when the display has the same quantitative scale property.
- Performance in retrieving ordinal information is optimal when the display has the same ordinal scale property.
- Performance in retrieving nominal information is optimal when the display has the same nominal scale property.

While the first and third hypotheses were supported, the second was only partially supported in that there was not a significant difference in performance between line graph (length) and density when an ordinal task was given. Some potential problems that might have affected the results can be identified by thoroughly investigating the task set that was used in their experiment.

2.4 Issues in Comparative Studies on RIDs

Zhang's [1] original framework focused on the information retrieval and interpretation of the viewer at the cognitive level rather than the perceptual level. This is because he was interested in how the match or the mismatch of the scale between display and data would affect the viewer's cognitive behavior. Park and Catrambone [7] attempted to capture cognitive behavior through two measures, correctness (whether or not participants produced the right answer) and reaction time.

Consider the tasks represented in Figures 5 and 6. Both tasks are ordinal tasks with different types of display. This was a direct comparison experiment because in the analysis, Park and Catrambone [7] compared the mean values (correctness and reaction time) of the two displays. Figure 5 shows a flaw though: the two lengths representing Chris and Alexis are somewhat difficult to discriminate. Obviously, this is a case where distance effect of length occurs, which might have affected the results as an extraneous variable.

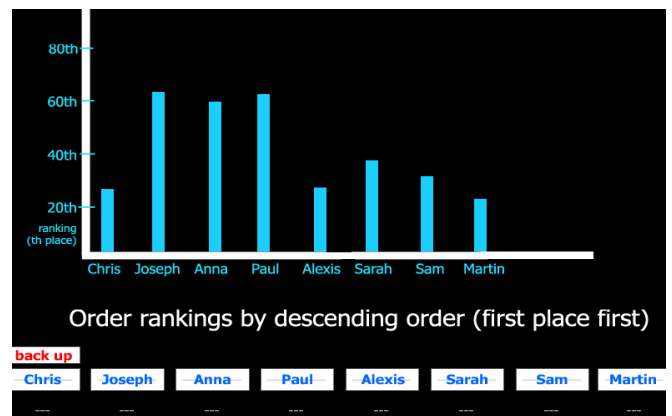


Figure 5. Ordinal task with Line Graph (length)



Figure 6. Ordinal task with Density

To avoid this problem when conducting an analysis of cognitive process in comparative research on RIDs, each element (e.g., a single line that represents Alexis in Figure 5) has to be readily discernable. However, the word “readily” is not discrete and definitely would not assist in designing definite tasks or stimuli. As a result, a need to quantify the interval among the elements arises (e.g., how large does the interval between the lines or the interval between levels of density need to be). It is also important that the ease of discriminating the elements in each task should be held constant. (i.e., the level of difficulty to discriminate lines in Figure 5 and density in Figure 6 should be the same). Thus, the first proposition in our framework is:

- 1) When conducting an analysis of cognitive process in comparative research on RIDs, each element in different graphs being compared by researchers has to be readily discernable and should have the same level of discrimination difficulty.

When looking at Figures 5 and 6, the number of graphical elements in the two graphs is set to 8. The number of elements in two graphs being compared should be held constant because the numbers affect participants’ searching process. The question is then how many elements can be used.

While the lower boundary is open and a task can be scaled down to decrease difficulty (i.e., in Figures 5 and 6 there can be fewer than 8 elements), the upper boundary might be less flexible. In Park and Catrambone’s [7] experiment, they stated that when levels go from black to white in density, there are a limited number of separate levels that can be shown while keeping them discriminable. Each RGB level for gray saturation increases together from 0 to 255 in an IBM PC (e.g. R = 0, G = 0, B = 0 for black and R = 255, G = 255, B = 255 for white). Informal pilot testing in their study suggested that when the interval between RGB levels becomes less than about 36, the visual distinction among levels starts to break down. Therefore, their experiment used 8 density levels (i.e., a set of 8 choices) that had an interval greater than 36. It is interesting to note that the line graph (length) suffers less severely than density from this limitation meaning that the number of elements in line graph (Figure 5) ought to be greater than the maximum number for density. This suggests the second proposition in the framework:

- 2) When conducting an analysis of cognitive process in comparative research on RIDs, the number of elements in different graphs being compared by researchers has to be the same and the maximum number of elements has to be derived from the graphic presentation that has the least discriminable levels (e.g., density).

Another important notion to consider is Steven’s Law that the perceived magnitude of a difference should increase as the absolute physical magnitude does. In other words, the interval of 36 at the lower end and the higher end of density might not provide an identically perceived magnitude difference.

While the main goal of the experiment is to study the differential thresholds across levels of density, the secondary goal is to test the hypothesis of whether Steven’s Law affects the interval as density increases.

3 METHOD

3.1 Participants

Participants were 40 undergraduate students who were enrolled in introductory psychology courses at the Georgia Institute of Technology who volunteered for the experiment to earn course credit.

3.2 Materials

Participants viewed a pair of density levels. A Java application was created to implement the condition sets on IBM PCs.

3.3 Procedure

Each participant had two density levels presented side by side on the screen. One was a "standard" density and the other was a "comparison" density. Participants were asked to determine which density level was denser. They pressed ‘Z’ for the left stimulus and ‘/’ for the right stimulus.

While the standard density remained constant in levels within each block, the comparison density changed in each trial. The comparison density changed by an interval of 59 from the RGB level of the standard density (e.g., if the standard density has a RGB level of 64, the range of comparison density will be from 35 to 93). The step size of the comparison density was two. This study adopted the method of constant stimuli in psychophysics so the comparison density randomly varied with the restriction of range and step size in each trial.

There were five blocks of trials with 94 trials each. Five blocks represented five different standard densities: RGB level of 0, 64, 128, 192, and 255. This was done to equally spread different density levels across a spectrum from 0 to 255. One goal was to test the hypothesis of whether the differential threshold is affected by Steven’s Law.

3.4 Design

The experiment was a one factor within subject design with five levels (RGB level of 0, 64, 128, 192, and 255) serving as the standards. A Latin square was used to determine the order of the levels. Participants were evenly distributed to each of the following task orders:

Table 1. Latin square orders

# Participants	Order of Standard Density
8	0, 64, 128, 192, 255
8	64, 128, 192, 255, 0
8	128, 192, 255, 0, 64
8	192, 255, 0, 64, 128
8	255, 0, 64, 128, 192

4 RESULTS AND DISCUSSION

The Latin square order had no effect and all results were collapsed over this variable. Density judgment accuracies involving the different standard densities as a function of the comparison densities are shown in Figures 7 to 11.

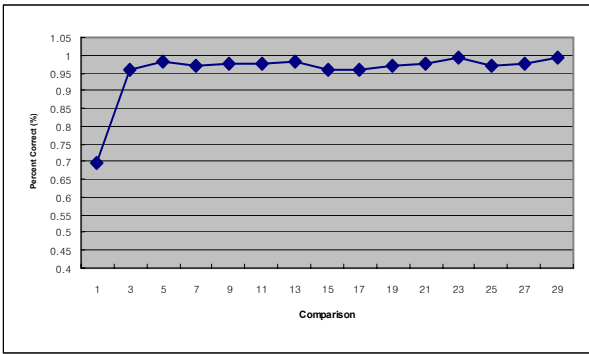


Figure 7. Percent Correct at standard density of 0

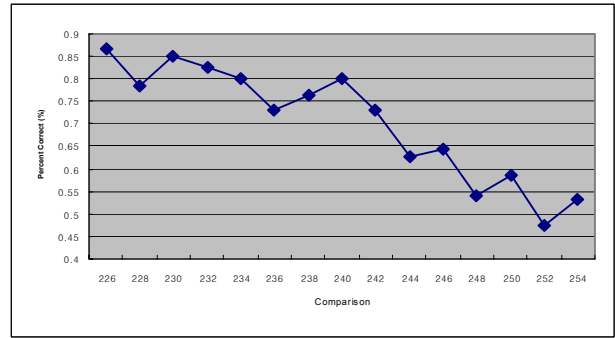


Figure 11. Percent Correct at standard density of 255

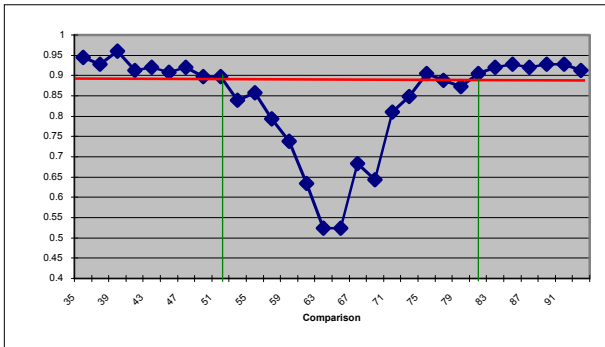


Figure 8. Percent Correct at standard density of 64

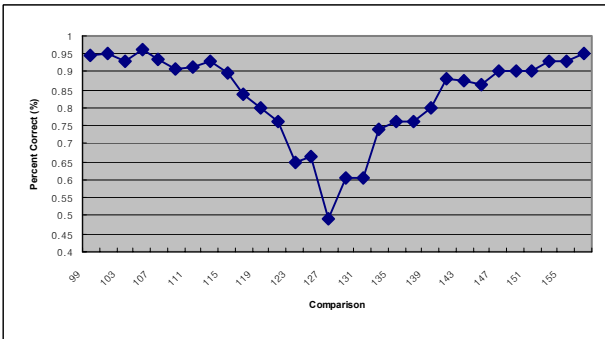


Figure 9. Percent Correct at standard density of 128

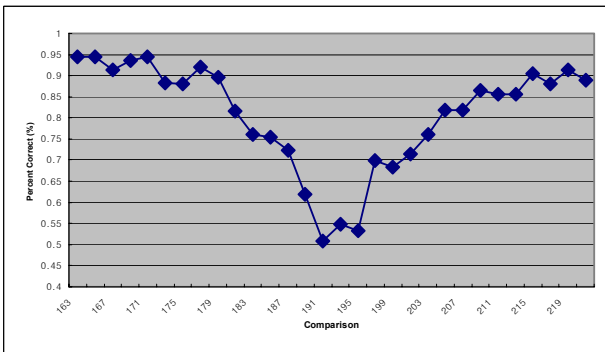


Figure 10. Percent Correct at standard density of 192

The upper and lower difference threshold was obtained by drawing a vertical line at 90% percent correct. While the traditional psychophysics study considers 50% as the differential threshold level, this study decided to set at 90% because as stated in proposition #1 in the framework, we are seeking a value that enables the viewer to “readily” discriminate the graph presentation. For example, in Figure 8, a density level of 51 or less (lower differential threshold) and 81 or more (upper differential threshold) can be readily discriminated when compared to with the standard density of 64.

It is interesting to note that in the lowest density (i.e., density of 0, black) only a slight increase in the comparison density enabled a clear discrimination (See Figure 7), whereas in the highest density (i.e., density of 255, white) the lower differential threshold was relatively far away at 226 (See Figure 11). This difference in threshold range is very noticeable when all the density levels are compared (See Figure 12).

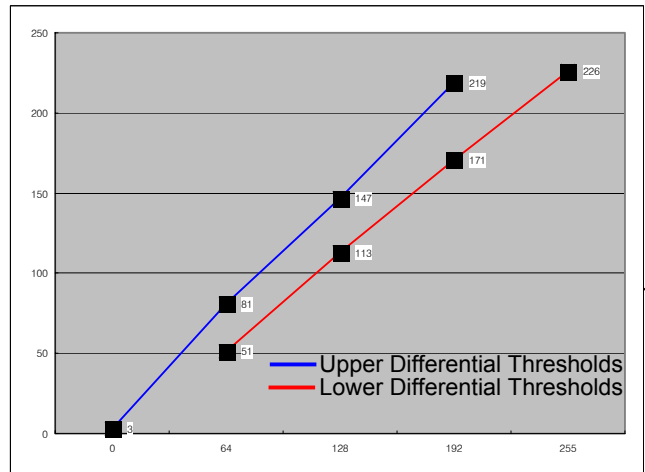


Figure 12. Upper and Lower Differential Threshold for the Standard Densities.

Notice that the interval between the lower and upper threshold increases as a function of standard density. For example, the interval at standard density of 64 is 31 (i.e., $81 - 51 + 1$) whereas the interval at standard density of 128 is 35 and of 192 is 49. This indicates that it becomes more difficult to discriminate as the

density level increases. This is consistent with Steven's law.

We discovered the lower and upper thresholds of various standard densities. We also found that the interval between the differential threshold and the standard density increases as a function of the standard density. This implies that the step size among different levels of density should be carefully determined when density graphs are compared with other types of graphs. For example, fixing the step size of the density graph to 20 is not a good idea. It might work for density levels below 200 but above this would no longer readily discriminable.

There are two ways to leverage this foundational data when conducting comparative research on RIDs. The first is to fix the step size of density graphs to the threshold that works for all levels. We learned that 226 is the lower differential threshold for the standard density of 255 (white). This interval (i.e., 30) is the highest interval across all density levels. The second is to vary the step sizes based on the density levels. This will require a precise control of the presentation but will potentially guarantee more elements (number of density levels in a graph) available for use.

5 CONCLUSION

This study outlined some critical issues in comparative research on RIDs. The assumption is that when conducting an analysis of the cognitive process of participants viewing different displays, the perceptual process of these participants has to be held constant so that it does not affect the results.

The framework suggested two propositions when conducting an analysis of cognitive process on different RIDs. First, each element in different graphs that are being compared for information display quality has to be readily discernable and should have the same level of discrimination difficulty. Second, the number of elements in the different graphs has to be the same and the maximum number of elements has to be derived from the graphic presentation that has the least discriminable levels.

The framework, in conjunction with the present study, particularly suggested that the number of elements has to be set equal to the density display because density has the fewest number of discriminable levels. Finally, a psychophysics experiment was conducted in order to determine differential thresholds across various density levels. We learned that Steven's power law plays a key role in deriving the differential thresholds and determining the number of elements that can be used in comparative research on relational information displays.

REFERENCES

- [1] Zhang, J. A representational analysis of relational information displays. In *International Journal of Human-Computer Studies*, 45, 59-74. 1996.
- [2] Carswell, C. M. & Wickens, C. D. Comparative graphics: History and applications of perceptual integrality theory and the proximity compatibility hypothesis (TR ARL-88-2/AHEL-88-1). Aviation Research Laboratory, University of Illinois at Urbana-Champaign.
- [3] Cleveland, W. S. & McGill, R. Graphical perception and graphical methods for analyzing scientific data. In *Science*, 828-833. 1985.
- [4] Mackinlay, J. D. Automating the design of graphical presentations of relational information. In *ACM Transactions on Graphics*, 5(2), 110-141. 1986.
- [5] Ungerleider, L. G. & Macko, K. A. Object vision and spatial vision: two cortical pathways. In *Trends in Neurosciences*, 6, 414-417.
- [6] Zhang, J. & Norman, D. A. Representations in distributed cognitive tasks. In *Cognitive Science*, 18, 87-122. 1994.
- [7] Park, S. & Catrambone, R. Represented and Representing

Dimensions in Relational Information Displays. In *Proceedings of 8th International Conference on Information Visualization*, 605-612. 2004.