# Redundant Coding Can Speed Up Segmentation in Multiclass Displays

Christine Nothelfer*
Northwestern University

Michael Gleicher†
University of Wisconsin – Madison

Steven Franconeri‡
Northwestern University

**ABSTRACT**

Multiclass data visualizations allow viewers to compare one dataset to another. The visual marks that represent these datasets, or *classes*, are visually distinguished from one another by easily perceived visual feature differences, such as color or shape. A designer of a graph or map might encode one class of marks as either red, or circular, and another class as either blue, or triangular. One common technique is to combine these cues in a redundant fashion, encoding one class as red *and* circular, and the other as blue *and* triangular, under the assumption that a larger difference (via *multiple* differing features) should help. Recent work [6] has empirically demonstrated strengthened grouping and improved accuracy in segmentation of redundantly coded objects. Does this redundancy benefit generalize to more realistic displays, and to other measures such as segmentation speed? We demonstrate in an experiment that redundant coding can lead to a small improvement in speed of visual differentiation in a simulated dataset in a crowded display.

**Keywords**: Human-centered computing—Empirical studies in visualization. Human-centered computing—Visualization techniques.

## 1 INTRODUCTION

Graphs and maps often depict multiple datasets, or *classes*, that are important to distinguish efficiently. These classes are typically designated by differences in easily perceived visual features, such as colors (e.g., red and blue data points) or shapes (e.g., circular and triangular data points). Such cues are often used in combination (e.g. shape *and* color, such as a graph with red circles and blue triangles) as a redundant coding of the class designations.

Earlier work has clashed in illustrating whether this technique improves a viewer's ability to isolate and compare data points from a given set of classes, or help link legends to data. Some might consider redundant coding to be inelegant in that it violates the rule that data displays should be as simple and unembellished as possible [3,7]. It might leave viewers confused – at least temporarily – about whether the independent dimensions reflect different aspects of the data or data types, or which dimension links data to legend terms. On the other hand, several studies [1,2,5] are often cited as showing an advantage for redundant coding.

Further, much of the literature against redundant coding does not demonstrate this empirically, and the studies in favor typically employ tasks that are only distantly related to class segmentation in data displays. Many of these latter studies require precise categorization of the value (e.g., color, size, or position) of an object along a dimension (is this the second reddest?), amid closely spaced alternative values from a predefined set of categories (e.g., this object is reddish-orange, not reddish; this object is the second biggest). While these examples are often cited in data visualization textbooks as a best available argument for their use (e.g., [8]), these tasks show an advantage for redundant coding for categorizing single objects, which does not reflect the demands of perceiving visual data displays. More commonly, visual data displays require observers to segment an entire collection of objects (pick out the bright ones), among widely spaced alternatives (red, green, or blue).

However, recent work has tested for a redundancy benefit in more realistic displays. One study tested the effectiveness of redundant coding in a task that better simulates the demands of viewing a data display (deciding which of two classes in a scatterplot had the higher average value), finding no evidence of an advantage for redundantly coded displays, even when the displays were crowded with objects [4]. While this single study casts doubt on the usefulness of redundant coding, it relied on an average value estimation task, which may be too noise-resistant to reveal an advantage of redundant coding.

Another study [6] tested for redundancy benefits in crowded displays, revealing that redundant coding greatly improves accurate detection of a global spatial pattern in briefly flashed displays, and that redundant coding can strengthen visual grouping among a set of visual objects. While this work is promising, only one of the two experiments in this study presented relatively realistic displays. Further, this experiment a) contained targets arranged in a fairly regular shape (a C-shaped partial ring), b) asked participants to indicate the quadrant of the screen where the target object ring was *completely missing* elements, c) presented test displays for only a fraction of a second, and d) tested for only segmentation accuracy. Together, these results may not be representative of typical distributions of data within a display (a, b) and viewing conditions (c), and the range of this redundancy benefit could be better understood (d).

It is surprising that there overall appears to be very little work demonstrating that redundant coding of visually presented information can improve visual processing of data displays, especially considering the ubiquity of this design technique. For example, redundant coding is a default setting for the creation of new graphs in Microsoft Excel.

Here we show that the redundancy benefit does indeed extend to more realistic displays and viewing conditions. Our task a) uses test displays that include randomly placed target objects, b) asks participants to locate the quadrant of the screen with the *smallest number* of target objects, c) presents test displays until the participant makes a response, and d) tests for segmentation speed (via participant response time), and whether segmentation difficulty modulates any redundancy benefit.

*cnothelfer@u.northwestern.edu
†gleicher@cs.wisc.edu
‡franconeri@northwestern.edu

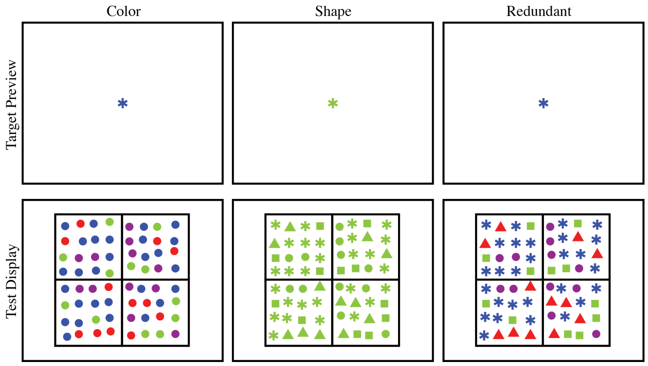


Figure 1: Participants saw a target preview screen (top row; until response), followed by the test display (bottom row). Trials concluded with the quadrant outlines until participants indicated which quadrant of the screen had the smallest number of target objects (bottom right quadrant in this example). Target objects differed from distractors either by color (left), shape (center), or color and shape redundantly (right).

## 2 Experimental Design

We constructed a task designed to emulate situations in which observers judge the distribution of the data points that constitute a class (Figure 1). Sixty-four target and distractor objects were placed 4 squares. Thirteen participants were asked to indicate the quadrant of the screen that contained the smallest number of target objects. After previewing the target objects (until the Space bar was pressed), the test display remained on the screen until the participant pressed the Space bar to indicate that they knew the answer. Participants then saw a blank screen (1s), followed by the square outlines (until the participant answered with the specific quadrant), and concluded with a blank screen (0.25s). Participants were told when they were incorrect in order to maintain an overall high accuracy and yield enough correct trials for analysis.

Target objects (e.g., blue asterisks) were identical to each other, and differed from distractors in color only (*color* trials), shape only (*shape* trials), or in both color and shape dimensions (*redundant* trials). The number of target objects varied within each quadrant to produce different levels of difficulty; quadrants contained 3, 6, 7, and 9 targets, 3, 9, 11, and 12 targets, 4, 7, 8, and 9 targets, or 5, 8, 10, and 11 targets (or, 0.33:0.67:0.78:1, 0.42:0.75:0.92:1, 0.44:0.78:0.89:1, or 0.45:0.73:0.91:1 as difficulty ratios, respectively). Trials were presented randomly in this within-subject design. If attending to objects encoded by multiple dimensions yields better visual selection, then participants should be fastest in the *redundant* condition. Additionally, this may depend on task difficulty.

### 2.1 Results & Discussion

Response time medians to correct trials were analyzed after removing the fastest 1% and slowest 1% of the trials. Confirming a redundancy benefit, we found that performance in the *redundant* condition ($M$ = 0.967s, $SD$ = 0.148s) was approximately 0.122s faster overall (color: $M$ = 1.066s, $SD$ = 0.225s; shape: $M$ = 1.113s, $SD$ = 0.125s), and significantly faster than whichever condition – *color* or *shape* – was fastest for each participant (average speed for participants' best condition (*color* or *shape*) – $M$ = 1.020s, $SD$ = 0.133s), as indicated by a two-tailed t-test, $t(12) = -4.35$, $p$ = 0.001. These results were not affected by the target difficulty ratio.

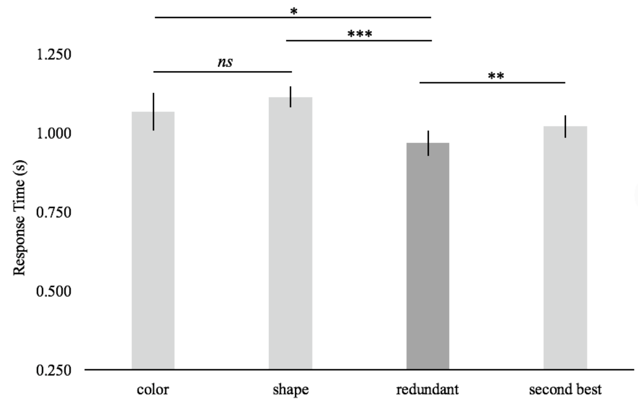


Figure 2: Results. Participants segment target objects faster when they are redundantly coded by color *and* shape (redundant), rather than color or shape alone. Participants are also faster on redundant trials than whichever condition – color or shape – was fastest (second best) for each participant.

## 3 Conclusion

Redundant coding can improve the speed of visual differentiation of classes in a crowded display. This benefit, however, is rather small (~11% faster), regardless of segmentation difficulty. The present results from our simple task are a building block in understanding more complex decisions when assessing data displays. It is unclear whether this benefit scales up to longer viewing times, more visually complex data displays (in terms of both number of objects and number of visual features), and different types of task (e.g., searching through a data display serially vs. attending to all data points simultaneously). More work will also be needed to explain why redundant coding leads to a small response time advantage, but a rather large accuracy boost (20-30%) [6].

## Acknowledgments

We thank Juju Miao for assistance in data collection. This work was funded by NSF awards IIS-1162037, IIS-1162067, and GRFP.

## References

[1] H. Egeth and R. Pachella. Multidimensional stimulus identification. *Perception & Psychophysics, 5*(6), 341-346, 1969.

[2] C. W. Eriksen and H. W. Hake. Multidimensional stimulus differences and accuracy of discrimination. *Journal of Experimental Psychology, 50*(3), 153, 1955.

[3] S. Few. *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Analytics Press, Oakland, CA, 2012.

[4] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *Visualization and Computer Graphics, IEEE Transactions on, 19*(12), 2316-2325, 2013.

[5] G. R. Lockhead. Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psych., 72*(1), 95, 1966.

[6] C. Nothelfer, M. Gleicher, and S. Franconeri. Redundant Coding Can Improve Segmentation in Multiclass Displays. Poster presentation at VIS (Chicago, IL), 2015.

[7] E. R. Tufte. *The Visual Display of Quantitative Information* (2nd ed.). Graphics Press, Romeoville, IL 2001.

[8] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, Burlington, MA, 2012.