Evaluation of Trend Localization with Multi-Variate Visualizations

Mark A. Livingston and Jonathan W. Decker

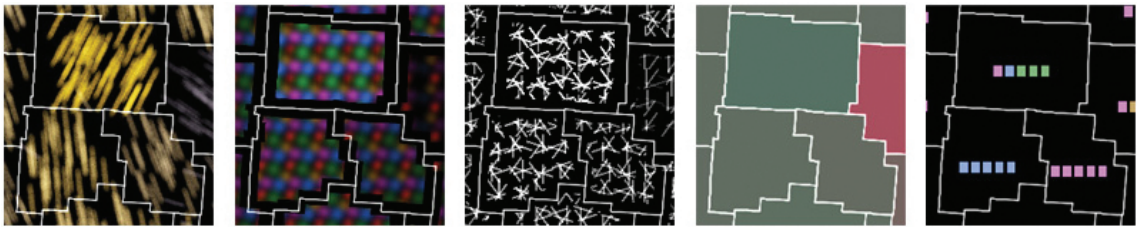

Fig. 1. Multi-variate visualization techniques we evaluated in our study, from left to right: Brush Strokes, Data-Driven Spots, Oriented Slivers, Color Blending, and Dimensional Stacking. These images depict a tri-county area in central Ohio. The encoded information is generated from a synthetic data set generated for the purposes of the study. See Section 3 for a explanation of the encoding.

Abstract—Multi-valued data sets are increasingly common, with the number of dimensions growing. A number of multi-variate visualization techniques have been presented to display such data. However, evaluating the utility of such techniques for general data sets remains difficult. Thus most techniques are studied on only one data set. Another criticism that could be levied against previous evaluations of multi-variate visualizations is that the task doesn't require the presence of multiple variables. At the same time, the taxonomy of tasks that users may perform visually is extensive. We designed a task, trend localization, that required comparison of multiple data values in a multi-variate visualization. We then conducted a user study with this task, evaluating five multi-variate visualization techniques from the literature (Brush Strokes, Data-Driven Spots, Oriented Slivers, Color Blending, Dimensional Stacking) and juxtaposed grayscale maps. We report the results and discuss the implications for both the techniques and the task.

Index Terms—User study, multi-variate visualization, visual task design, visual analytics.

1 Introduction

Multi-valued data sets are increasingly common in a diverse set of applications. Improved sensors acquire new or more fine-grained measurements. Meta-data such as uncertainty about such measurements constitute another data channel which may be helpful in conducting analysis of the data. Even initial data analysis techniques may produce synthesized measurements to include in the analysis.

Thus an increasingly common problem is that there is not sufficient time to analyze all the acquired data in the time available to make a decision. One classic response to this difficulty is the use of statistical graphics to paint the "big picture" of the data. While such summaries undoubtedly yield basic insights, more complex patterns and trends do not easily emerge from simple techniques. For example, geographic information systems (GIS) produce data that has a spatial component. A table of summary statistics and a scatterplot are unlikely to yield insight to geographic patterns in the way that a map-based visualization could [3]. Such summaries, however, do reduce the risk of overwhelming the analyst's capacity to understand the data presentation.

Another response to the explosion of variables in data sets has been the creation of multi-variate visualization techniques, in which a collection of variables may be viewed simultaneously. In these techniques, the details of any particular variable may be visible. Potential problems with integrated presentation of multiple values include that the analyst could be overwhelmed by the volume of data and that the number of usable perceptual channels is exceeded. The potential to discover combinations of perceptual cues that enable simultaneous understanding of multiple data layers (i.e. variables encoded with a visualization technique) helps to fuel exploration and innovation of such integrated techniques. One long-term goal in our work is to determine how many variables can be present in a visualization and still allow users to discover new insights in the data. We hope that these relationships need not be explicitly represented, thus avoiding the need to fully understand the data in order to present it in the best manner for a task that may or may not be itself fully understood a priori.

Thus a number of multi-variate visualization techniques have been devised to display such data. However, evaluation of the utility of such techniques for general data sets remains difficult. Thus most techniques are studied on only one data set and task. Another criticism that may be offered for some evaluations of multi-variate visualizations is that the task doesn't require the presence of multiple variables; the analyst would be best served by focusing only on one variable at a time. At the same time, the taxonomy of tasks that users may perform visually is extensive. We designed a task in the category of visual comparison of multiple data values in a multi-variate visualization. Specifically, we asked users to localize trends among the data layers. We synthesized five years of demographic surveys (e.g. percentage of residents under age 18); each year became a layer of data. We conducted a user study with this task, evaluating five multi-variate visualization techniques from the literature (Brush Strokes, Data-Driven Spots, Oriented Slivers, Color Blending, and Dimensional Stacking) and side-by-side grayscale maps. After reviewing (Section 2) multi-variate visualization techniques and evaluations of them, we describe the parametrization we used for the techniques (Section 3), detail our study design (Section 4), report the results (Section 5), and discuss the implications for the techniques, the task, and the field (Section 6).

2 Previous Work

Our work focuses on the evaluation of multi-variate visualization techniques; however, it is natural in the course of using such techniques to adapt them slightly in order that comparisons might be more fair to the techniques. Thus we review multi-variate visualization techniques as

- *Mark A. Livingston is with the Naval Research Laboratory, E-mail: mark.livingston@nrl.navy.mil.*
- *Jonathan W. Decker is with the Naval Research Laboratory, E-mail: jonathan.decker@nrl.navy.mil.*

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Evaluation of Trend Localization with Multi-Variate Visualizations	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,4555 Overlook Avenue SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Multi-valued data sets are increasingly common, with the number of dimensions growing. A number of multi-variate visualization techniques have been presented to display such data. However, evaluating the utility of such techniques for general data sets remains difficult. Thus most techniques are studied on only one data set. Another criticism that could be levied against previous evaluations of multi-variate visualizations is that the task doesn?t require the presence of multiple variables. At the same time, the taxonomy of tasks that users may perform visually is extensive. We designed a task, trend localization, that required comparison of multiple data values in a multi-variate visualization. We then conducted a user study with this task, evaluating five multivariate visualization techniques from the literature (Brush Strokes, Data-Driven Spots, Oriented Slivers, Color Blending, Dimensional Stacking) and juxtaposed grayscale maps. We report the results and discuss the implications for both the techniques and the task.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**10**	

well as evaluations of them. We note adaptations of the techniques for our user study in Section 3.

2.1 Multi-variate Visualizations

2.1.1 Brush Strokes

This technique creates a texture inspired by impressionist paintings; the multiple attributes of the Brush Strokes vary throughout the image to denote the data values [11, 12]. Strokes are randomly placed over the surface; they vary in the intensity and hue of their surface color, in orientation, and in their width and length (Figure 1, left). One difficulty presented by this method is that the parameters will not have the same resolution; for example, the intensity and hue will have more available output levels – both on the display device and in human perception – than the width of a stroke. In addition, we find that increasing stroke width can manifest itself as blurring in the image.

2.1.2 Data-Driven Spots (DDS)

This method encodes each data layer with Gaussian kernels ("spots") on a randomly jittered grid [2]. The spots for each layer differ in size and color; spot intensity encodes the data value (Figure 1, second from left). This technique may take advantage of other perceptual channels; layers may be animated by moving the spots across the surface, changing their intensity with the underlying data value. This feature (which may also apply to other visual representations) can synthesize greater resolution of the visualization than the initial sampling density, although it may conflict with the notion of designing a precise jitter pattern.

2.1.3 Oriented Slivers

Similar in concept to DDS, one may place a pattern of short, grayscale lines ("slivers") at randomly jittered grid positions [24]. The orientation of these slivers denotes which data layer is represented, and the intensity denotes the data value in that layer. Thus a data layer could be seen on the surface of one oriented set of slivers (Figure 1, middle). One critical design issue is the density of the slivers. Few slivers implies a sparse sampling of the data, which may be inappropriate for high-frequency data. Too many slivers can cause the slivers to overlap and become indistinct with regard to intensity (data value) and even orientation (data layer). Other parameters, such as sliver width and length, may cause similar perceptual problems. One advantage of this technique is that it can convey multi-dimensional data while occupying relatively few perceptual channels.

2.1.4 Color Blending

This classic technique assigns one source color for each layer of data. The composite data visualization shows the weighted sum of the source colors, with the weight derived from the data values. In this way, the dominant hue of the pixel or region in the visualization indicates the greatest component value among the values at that location (Figure 1, second from right). This technique has the advantage of using each pixel as an independent visual identifier for the underlying data (as opposed to the other techniques we describe, in which a multi-pixel region is required to show a single domain point's values). However, as we can display (with most modern displays) and perceive only three color channels, this technique is limited in its effectiveness for data sets with more than three values.

2.1.5 Dimensional Stacking

Early multi-variate visualizations used simple shapes or shape patterns [1, 14, 15], such as small, adjacent blocks or sectors of a circle (Figure 1, right). Each shape represents one value in the data; a cluster of such shapes represented a multi-valued sample point. The individual values were typically depicted through the intensity (in a grayscale implementation) or hue (in a color implementation). One critical design decision is how to convey the resolution of the data. The data range is often divided into bins, with each gray level or hue assigned to a particular bin. Given the limited resolution of human perception of intensity and color, this technique may be more valuable for showing gross differences than for fine details. This technique exhibits a problem in that a finite region is required to show the multiple data values at a single point of the domain.

2.2 Evaluations

Numerous authors have contributed to the body of anecdotal, theoretical, and quantitative evidence for the design quality of a multi-variate visualization. We concentrate our review of this body of work on the last two contributions.

One may begin by analyzing capabilities of human perception to derive design guidelines that may be applied to visualization tasks [23]. Among the type of tasks described were the "perception of emergent properties" made evident by a visual presentation of the data. With the advent of the conceptual framework of visual analytics [20] and its emphasis on analytical reasoning through visual interfaces, the importance of clarity of presentation for complex data sets was further stressed. These concepts are fundamental to our approach. More directly informing our work are approaches that begin with understanding the data and then examine visual properties. Healey et al. [8] identified four pieces of information by which a user and visualization system may architect a visualization: the importance of each attribute, the spatial frequency, whether it is continuous or discrete, and the task (if any) the user wishes to perform on the attribute. The authors then discussed how this information may be used in combination with understanding of human perception, mixed-initiative interaction, and automated search strategies to create a mapping from data attributes to visual features. Features employed included luminance, hue, size (height of bars), density, orientation, and regularity to a grid. Earlier, Zhou and Feiner [25] characterized data in order that an automated method might craft visualizations. The dimensions in the taxonomy were type (divisible or atomic), domain (semantic, e.g. physical or abstract), attributes (e.g. shape), relations (connections between data), role (with respect to user goals), and sense (user visual preferences). These taxonomies sparked our thinking about what aspects of the data created difficulties for given visualization techniques.

Urness et al. [21] applied overlay and embossing to composite textures which encoded multiple 2D vector fields. By adding colors and altering texture properties, such as line thickness or orientation, in line-integral convolution, they created effective visualizations for multiple flow fields, as assessed by domain experts. Laidlaw et al. [12] visualized seven-layer diffusion tensor images using ellipsoid glyphs and Brush Strokes. They showed significant differences between healthy and unhealthy spinal cords in mice. The glyphs were effective at showing tensor structure everywhere within the images, whereas layered Brush Strokes encoded field values and enabled users to understand relationships between layers. The difficulty in this method was the potential for cluttered images. This was not a serious problem because their application displayed a number of dependent variables (data layers).

Several user studies have examined the utility of individual techniques. Healey et al. [7] found that height and density of vertical bars over a 2D domain could be easily identified, but that certain combinations with background elements (such as salience or regularity of samples in a dense field) made it hard to understand the data. They validated their results on weather data. The introduction of Brush Strokes [9] (specifically, color, texture, and feature hierarchies among luminance, hue, and texture) enabled verification that guidelines for perception during visualization [23] applied to non-photorealistic visualizations as well. The authors also noted the aesthetic quality of such visualizations.

Oriented Slivers [24] enabled users to perceptually separate layers within a data set. To get the best performance on identifying the presence of a constant rectangular target in a constant background field, they found a minimum separation of 15° between layers necessary. Data-Driven Spots [2] enabled users to discern boundaries amongst as many as eight layers of data. Joshi [10] visualized time-varying fluid data using art-inspired techniques such as pointillism, speed lines, opacity, silhouettes, and boundary enhancement for weather and other data. Users were able to track a feature over time more accurately and expressed preferences for the illustration-inspired techniques.

Other studies have compared multiple, diverse visualization techniques. Laidlaw et al. [13] compared six techniques for 2D vector data, asking users to locate critical points, identify types of critical points, and advect a particle. Users performed better when the visualization explicitly represented the solution to the tasks – i.e. showed the sign of vectors in the field, represented integral curves, and showed critical point locations. Experts and non-experts did not show significant differences. Hagh-Shenas et al. [5] compared Color Blending (Section 2.1.4) and Color Weaving. Color Weaving refers to the pointillism techniques, such as DDS, discussed earlier. The name comes from the flow field color method [22], which works on the same concept of separating colored elements so that multiple, overlapping features can be identified in the same spatial region. Maintaining the original colors as in Color Weaving outperformed Color Blending [5]; this difference increased with the number of data layers. Color selection for the various scales was a critical issue in the blending methods. Tang et al. [19] developed multi-layer texture synthesis for weather data visualization, varying scale, brightness, orientation, and regularity. Users in their study performed as well with this technique as with one using the Brush Strokes technique proposed by Healey et al. [9]. In our own previous study [16], we found that the parameterized patterns of Data-Driven Spots and Oriented Slivers helped users perform a critical (maximum) point detection more accurately and faster than the glyph representations of Brush Strokes and Stick Figures [18]. We also found that some techniques were sensitive to the brightness and contrast settings of the monitor.

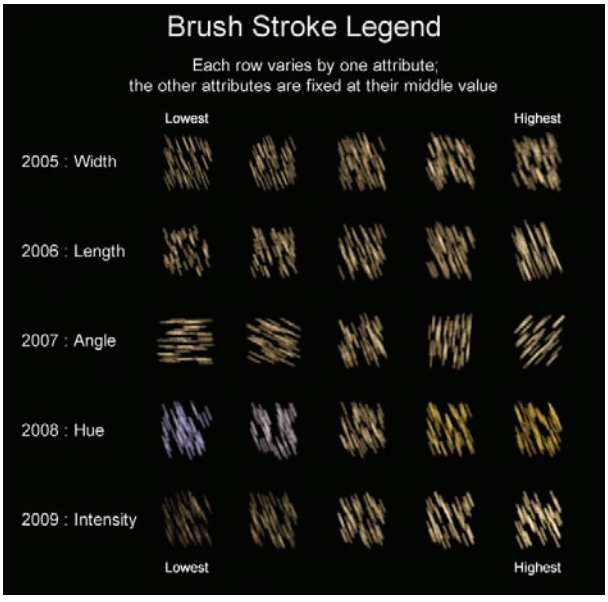

Fig. 2. The legend for the Brush Strokes indicated the styles that mapped to the range of data values.

3 Techniques

The following section summarizes how we applied a technique to our data and gives descriptions and hints on the trend localization task that our subjects read. The core of the task was to find the county (region) exhibiting the greatest increase or greatest decrease in a variable within a five-year time span. A technique legend (if applicable) was provided to the subjects along with each question for reference. Compare the description and key to the cropped images provided in Figure 1 in addition to the Figures in this section.

3.1 Study Encodings

Juxtaposed Maps Our baseline technique used a series of grayscale maps, each of which corresponded to a single data layer. To localize the trend, the subject had to scan the maps. The intensity (brightness) of the gray value indicated the data value; brighter pixels indicated higher values. So if the county's brightness increased across the maps, the trend was increasing. If the county's brightness decreased, the trend was decreasing. Subjects selected their answers by clicking on a sixth, empty map at the lower right of the array of maps (Figure 6).

Brush Strokes The legend (Figure 2) illustrated the mapping of the properties of Brush Strokes. Since we felt that the final values were most helpful, we mapped hue and intensity to the fourth and fifth data values. Notice that characteristics such as length and width were more subtle than hue and intensity; the range of (horizontal) stroke width was six to twelve pixels (equal[1] to 0.51°–1.02°) and of stroke length, 25 to 49 pixels (2.13°–4.16°). Thus users may have found the initial value harder to interpret than the final value; this could be exacerbated by county boundaries cutting off strokes. Stroke orientation (third value) was horizontal for zero; a stroke tilted 135° clockwise from horizontal (slanting down to the left) represented the maximum value for any county in the current map. Strokes that were dim and blue, but long and wide, indicated decreasing trends. Strokes that were bright and gold, but short and narrow, indicated increasing trends. But since a value could start in the middle and either increase or decrease from there, such trends would have a slightly different pattern. (See Figure 1, left.)

[1]All angular sizes of features are given using the hardware configuration described in Section 4.

Data-Driven Spots For DDS, we created a set of jitter patterns that avoided the possibility of spots from different layers overlapping each other (Figure 3). Our implementation in other respects is no more restrictive or permissive than the original specification. The standard deviation of the Gaussian kernels were 25 (red, initial), 25, 17, 17, and 13 (green, final) pixels, for the five data layers (2.13°, 1.45°, and 1.11°, respectively). The technique legend (Figure 4(a)) mapped the size and color of spots to years. The spots were dim if the value was low, while the spots were bright if the value was high. So if the red spots were the brightest, followed by the brown, purple, blue, and then green, the trend was decreasing. Conversely, if red spots were dimmest, but the brown, purple, blue, and green got progressively brighter, then the trend was increasing.

Oriented Slivers With Oriented Slivers, we did not prevent overlap of slivers from different years; thus, our implementation is neither more restrictive nor more permissive than the original guidelines for the technique. In this study, the initial year (2005) was specified by the vertical sliver. Each subsequent year was represented by a sliver rotated slightly more clockwise than the previous year (Figure 4(b)). The slivers were three pixels wide and twenty pixels long ($0.26° \times 1.70°$) in the vertical orientation; any change from this is due to anti-aliasing or county boundaries. If the slivers for a particular county got progressively brighter, then the county's value was increasing. If they got progressively dimmer, then the county's value was decreasing.

Color Blending We used the same color set for Color Blending as for DDS (Figure 4(d)). The colors are blended by the vector

$$w_i \cdot v_i \; ; \; w = \{0.1111, 0.1111, 0.1111, 0.2222, 0.4445\}, \; v = \text{data}.$$

The colors {red, orange, purple, blue, green} are specified in CIELab by $L = 50$, $b = 50$, and $a = \{-95, -45, 0, 45, 95\}$. The weights $w_i v_i$ were renormalized so that the L and b parameters were unchanged. In each region, the dominant color corresponded to the highest value at those pixels. Thus, if the tint were more towards the red and brown and less towards the blue and green, then the trend was decreasing. If the tint were more towards green and less towards red, then the trend was increasing (Figure 1, second from right).

Dimensional Stacking We selected a color version of Dimensional Stacking with five bins; thus we could use the same color set as for DDS and Color Blending, although it was now mapped (Figure 4(c)) to data value (bin) rather than data layer (year). Each bin

Fig. 4. Four of the technique keys used in the study.

(a) The Data-Driven Spots legend indicated the spots' size and color for the data layers.

(b) The Oriented Slivers legend indicated the orientations assigned to each data layer.

(c) The Dimensional Stacking legend indicated the colors used to denote the value in each data layer

(d) The Color Blending legend indicated the colors that were used for each data layer.

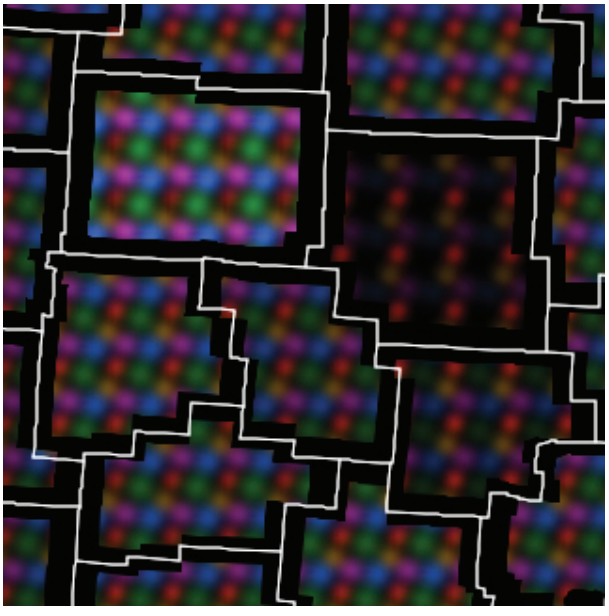

Fig. 3. An example of Data-Driven Spots centered on a region in central Ohio. Dot intensity patterns indicated trends in the DDS technique.

contained 20% of the values over the five-valued data for a trial; red for the lowest values, then brown, purple, blue, and finally green for the highest values. Our blocks were 9×11 pixels ($0.77^\circ \times 0.94^\circ$), and the set was arranged horizontally. So if the blocks progressed from red at the left through brown, purple, blue and then green at the right, the trend was increasing. If the blocks progressed from green at the left to red at the right, the trend was decreasing. But since the absolute values were not the focus of the task, but rather the trend, a strong increasing trend could have started at brown or purple and finished with green. Likewise, a strong decreasing trend may have started with blue or purple and finished with red. If the blocks did not change colors, then there was no strong trend in either direction (Figure 1, right).

3.2 Implementation and Design

Each of the techniques evaluated in this study was created using a custom, modular application which executed a sequence of layer filters and operations. This program was designed to allow us to quickly implement a variety of methods. It benefits these techniques to provide an interface for interactive parametrization [4]. However, the aforementioned application only accepts flat-file parameter strings, and time was not allocated to optimize these techniques to render at interactive rates. Therefore, devising the specific parameter set for each technique for the given application was a tedious endeavor. We highly recommend that any interested party looking to use these or similar techniques in their work consider taking the time to create interactive interfaces to allow their designers to modify the specific look of a technique and see the effect of their changes interactively.

We made several specific modifications to the techniques in order to aid user comprehension of the encoded information. Most noticeably, we centered a single Dimensional Stacking glyph in each county. In the original incarnation, glyphs were sampled over the entire surface of the combined layers. Since we knew that our data only varied between counties, we felt that texturing the entire area of each county would only detract from the information on display. Furthermore, image generation and compression artifacts existed along boundaries, which would have created observable misinformation in glyphs situated over county borders. Once we made this alteration in the formula for Dimensional Stacking, we noticed that this technique suddenly had a slight advantage over the other techniques which involved a varying pattern over the surface of the map. The set of affected techniques includes Data-Driven Spots, Oriented Slivers, and Brush Strokes. To level the playing field, we introduced black borders between the regions of interest. We feel this provided a visual experience compatible with using a single glyph to represent each county, allowing us to safely compare these techniques to our implementation of Dimensional Stacking.

In addition, a Dimensional Stacking glyph is usually a compact shape, with a grid of squares instead of a single sequence. This is because the glyph would represent a specific grid cell of a high-dimensional data set. Once we had decided to display only one glyph for each county, there was no longer any reason for the glyph to remain square. This is beneficial to the technique, since it helps clarify which cell corresponds to which layer.

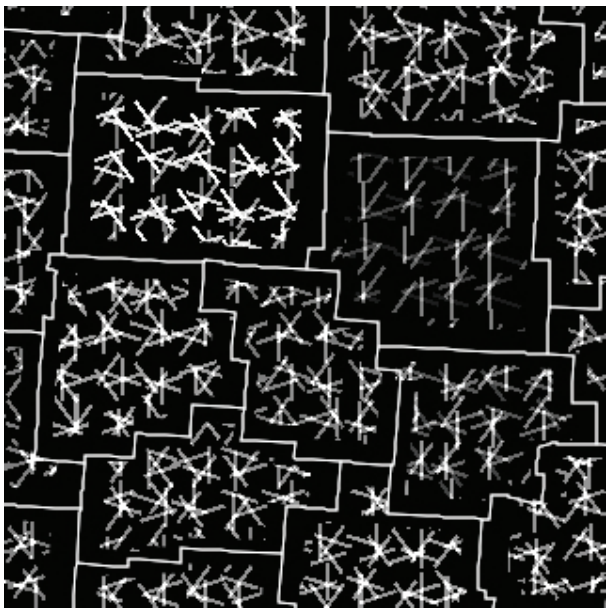

Fig. 4. An example of Oriented Slivers centered a region of Ohio shows the bright spots at the overlap of slivers.

Dimensional Stacking was limited in precision, since it was encoded using only five color values. There were some cases where an *identical stack* was represented with exactly the same set of blocks as the target, despite having a different trend size. This forced users to choose between two (or more) equally likely candidates.

We combined the layers of Oriented Slivers using weighted-sum blending with equal weights, which created some bright regions where the individual slivers within the patterns overlap (Figure 4). In many cases these points of overlap were the brightest features within the pattern, which could make it hard to read the layers. Blending by maximum intensity projection might mitigate this artifact.

The Ohio and Indiana county boundaries used in our study were obtained from the US Census website. We converted the ArcView Shapefiles to Scalable Vector Graphics (SVG). These SVG files were then used to raster grayscale maps from our synthetic data. They were also used to create dynamic boundaries which enabled our subjects to select their answer during the survey.

4 Experimental Design

We generated images for the study using an off-line application discussed in Section 3.2, then presented these images to the subjects in a browser-based survey on two identically-configured workstations in a controlled laboratory setting. Previously, we found that the brightness, contrast, and gamma settings affected performance [16]. Both workstations resided in the same room lit with standard fluorescent lights. Both workstations ran Windows XP (Service Pack 3) and the Google Chrome Web Browser (v10.648.204). We used 30in monitors (Dell 3008WFP) running at their native resolution of 2560×1600 using default factory settings: Brightest: 75, Contrast: 50, Sharpness: 50, Gamma "PC," Color Settings Mode "Graphics," Preset Mode "Desktop." The following sub-sections give specific details of our study design. We did not mandate a precise viewing distance; the desktop yielded a viewing distance of 67cm for a typical seated position (yielding pixel pitch of 0.25mm).

4.1 Trend Localization Task

We surveyed literature on visualization tasks, looking for a task that would require users to use multiple layers of data. One criticism we offered of our own previous study was that one data layer was the target layer, whereas all other layers of data present were merely distractors. The subjects would have been better served if those non-target layers were removed from the visualization. We desired a task for which multiple layers would be required for the users' success.

Fig. 5. The maps for Indiana and Ohio that were used in the study.

We found in studying use cases for micromaps [3] the task of discovering a trend in the data. This fits in the taxonomy of Zhou and Feiner [25] in the category of comparison of layers; it requires finding the difference between at least two layers and was used by Joshi in studying pointillism-based techniques [10]. A long-term goal of our research is to determine the utility of multi-variate visualization techniques in mitigating the difficulty of seeing relationships in high-dimensional data. Thus we decided to present five layers of data, an amount that seemed tractable in our previous study. However, we made the trend identifiable through only the first and last layers. That is, the difference between the first and last layers was set based on a region's status as a target, distractor, or noise. Then the intermediate layers were computed with linear interpolation of the initial and final values. Note that this does not imply that the greatest value indicated the correct answer for the greatest increase (and similarly, the lowest value did not necessarily belong to the greatest decrease). While this did occur, it was not always the case. This strategy in one sense made the task easier than that used by Joshi, who asked users to recognize five types of trends: increasing, decreasing, constant, increasing-then-decreasing, or decreasing-then-increasing. However, our task was more difficult in a different sense. We asked users to find the greatest increase or decrease across a visualization; thus users needed to compare different spatial locations, which was not a feature in Joshi's task.

We opted to simulate data as plausible responses to demographic questions. Due to the near similarity in county size, we opted to portray these as statistics collected from Indiana and Ohio counties (Figure 5). The narrow range of sizes used for the target reduced the possibility that the target size was an uncontrolled factor in the subjects' performance. We wanted subjects to be engaged in the questions but not to have preconceived notions of the answers. Thus questions included items such as the percentages (expressed in the range [0..1]) of people who thought a dishwasher, microwave, cell phone, or some other consumer electronics item was a necessity, and more traditional demographics such as the percentage of residents who were born in the county or were under age 18.

We created the initial year's data with a random number generator, keeping the numbers close enough to the center that any trend could be applied to any county, i.e. a range of [0.3,0.7]. The final year's data was then generated for all counties according to the categories of target, distractors, and noise. Finally, the internal years' data were interpolated linearly from the initial and final years.

The target trend was size always 0.3 (on a scale of 0.0 to 1.0); for both the increasing and decreasing targets, a set of up to five distractors was selected to have a trend size of 0.2, another set of up to ten was selected to have a trend size of 0.1, and yet another set of up to twenty was selected to have a trend size of 0.05. The remaining counties were in the category of "noise," having a range of values between -0.03 and 0.03. The range of county sizes used for targets and distractors was 375–475 mi^2, which yielded 100 candidates for targets and distrac-

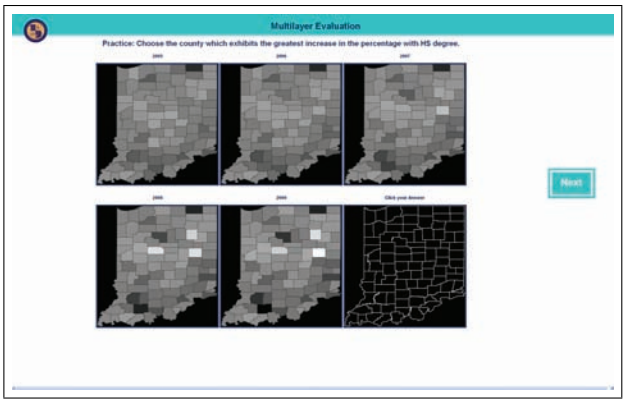

Fig. 6. The screen layout showed subjects the technique legend for multi-variate techniques, the visualization, question, and a "Next" button. This image shows Juxtaposed Maps, which did not need a legend.

tors out of the 180 total counties. The display selected a target and distractors, then showed the subject the state with the correct target.

4.2 Subject Procedures

We instructed users to identify the county in the presented map with the greatest increase or greatest decrease. After these general instructions, the trials were presented in blocks by technique. Each block was structured as follows. First, instructions specific to the technique were given. These instructions included hints for cues that would indicate trends in the upcoming technique (Section 3.1). Next, two tutorial questions were presented (consecutively). Upon the user indicating a response to the tutorial question, the control program immediately indicated the correct answer (whether the user response was correct or not). Finally, the users began the data trials for the technique. Twenty questions were presented for each technique; ten asked the user to find an increasing trend and ten asked for a decreasing trend. In the data trials, no feedback was given regarding the correct answer. However, users could change their answer in data trials by simply clicking on another county. Only when the user clicked a "Next" button did the final answer get entered as the user's response (Figure 6). The order of the visualization techniques was counterbalanced with a 6×6 Latin square. The trend type was counterbalanced by alternation, with half the subjects beginning with each type.

Eighteen subjects (twelve male, six female) completed ten questions of each trend type for each of the six techniques, for a total of 10*2*6*18=2160 trials. Subjects ranged in age from 26 to 66 (mean of 42). All reported being moderate or heavy computer users. One subject reported red-green color blindness, but we allowed this user to complete the study. This subject's performance on the four methods that used color ranked in the top half of the subject pool on the two techniques that depended most on red-green color perception (Color Blending and Dimensional Stacking), and in the top third of the pool for Brush Strokes (which required blue-yellow color perception), but in the bottom half with Data-Driven Spots, which in our implementation relied entirely on color perception to differentiate the layers. Thus we did not remove this subject from the study. No other users reported any color blindness, although we did not test users.

4.3 Independent Variables

The primary independent variable was the visualization technique. We introduced the variable Trend Type upon noticing that some techniques seemed more conducive for indicating one trend type than the other.

4.4 Dependent Variables

We recorded the county users selected and measured error as the difference between the trend size of the target (0.3) and the trend size of the selected county. Both types of trends could be present in both types of questions; thus error could in theory range (disregarding sign) between 0.0 for a correct answer to 0.6 if the largest trend in the wrong direction were selected. (Some errors of this magnitude exist in our data; we address this below.) We also measured response times from the onset of the stimulus until the user's final response for each question was selected on the screen. The time from the selection of the final response until the user confirmed the selection and moved to the next trial was not included. Users completed the NASA Task-load Index [6] to measure subjective workload of each technique.

4.5 Hypotheses

Based on our previous results, we expected that Data-Driven Spots and Oriented Slivers would lead to the best accuracy in the task; however, in pilot testing, we revised our expectation to only Data-Driven Spots performing well. We expected Dimensional Stacking to have the lowest accuracy due to the low resolution of the color squares to discern fine differences in the trend sizes. We expected the baseline of Juxtaposed Maps to perform well, as it was likely to be most familiar.

Similarly based on our past results, we expected Data-Driven Spots to exhibit the fastest response times. In our past study, Color Blending showed fast response times, but we concluded (given the low accuracy with Color Blending in our previous study) that this indicated that users were simply abandoning the task. However, we believed we had improved the technique's usability and would find fast user response times. We expected Juxtaposed Maps to lead to the slowest response times due to the increased scanning area it required from users. Finally, with regard to subjective workload, we expected that Data-Driven Spots would be judged to have the lowest workload, as it had in our previous study. We expected Juxtaposed Maps and Dimensional Stacking to have the highest workload ratings, due to the wider scanning area and potential for identical stacks, respectively.

5 Study Results

We analyzed the accuracy and response times with separate repeated-measures ANOVA, using a 6 (Visual Techniques) $\times$ 2 (Trend Type) design. We also checked for interactions between the visual techniques and the trend type of increasing or decreasing. We analyzed the subjective workload data with a 6-way ANOVA[2].

Visualization technique had a main effect on error – $F(5,85)=3.018$, $p=0.015$ (Figure 7). Subjects were more accurate with the Juxtaposed Maps and Data-Driven Spots than with the remaining techniques, confirming our hypotheses for high accuracy. Dimensional Stacking performed poorly, but not statistically worse than the remaining techniques, so we cannot accept our hypothesis with regard to the poorest accuracy. Visualization technique also had a main effect on response time – $F(5,85)=16.653$, $p=0.000$; users were fastest with Color Blending, while Brush Strokes, Data-Driven Spots, and Oriented Slivers exhibited nearly equivalent response times. This is clearly contrary to our hypothesis; reasons for this appear to be explained by the results for Trend Type, discussed below. Finally, Visualization had a main effect on subjective workload – $F(5,85)=3.661$, $p=0.005$ (Figure 8). Users found Color Blending, Data-Driven Spots, and Juxtaposed Maps to have the least workload (in that order). Again, our hypothesis was inaccurate for most of the techniques (save Data-Driven Spots).

Trend type had main effects on error – $F(1,17)=26.063$, $p=0.000$ – and response time – $F(1,17)=17.065$, $p=0.001$ (Figure 9). Users were both more accurate and faster in locating decreasing trends than increasing trends. Visualization and Trend Type exhibited an interaction for both error – $F(5,85)=9.384$, $p=0.000$ – and response time – $F(5,85)=2.775$, $p=0.023$. Color Blending led users to notably more accurate and faster responses with decreasing trends than increasing trends; this would appear to account for its fast response times and low workload rating. There was relatively little difference between trend types for the remaining techniques (on either accuracy or response time).

Although it was not a goal in our study design, the randomly varying number of "close" distractors in the visualization may be analyzed.

[2]Significance tests are given via the standard F-test with two degrees of freedom for explained and unexplained variances; the probability (p) is that of obtaining the F-value if the null hypothesis were true.

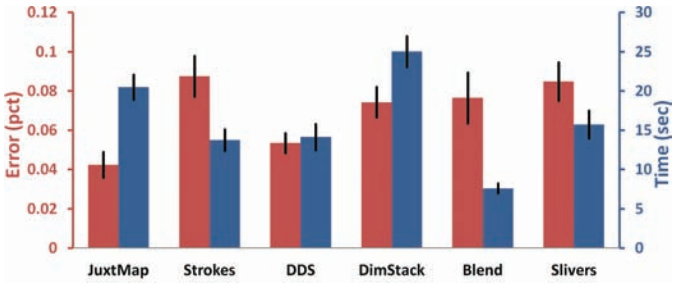

Fig. 7. Visualization had a main effect on both error (red) and time (blue). Juxtaposed Maps had the least error, but users were fastest with Color Blending. Data-Driven Spots seemed to offer a good compromise of the two objective measures. Error bars in this and all graphs are one standard error.

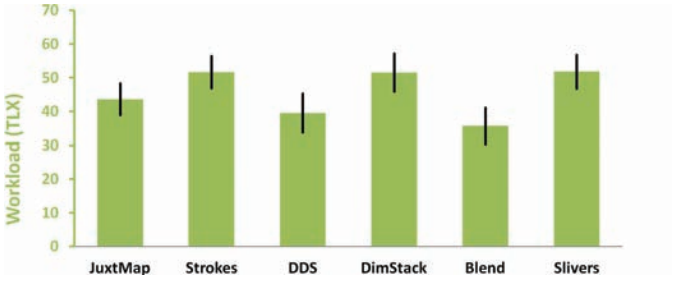

Fig. 8. Visualization had a main effect on workload; users judged Color Blending to have the lowest workload, followed by Data-Driven Spots.

We define these distractors as trends that were of size 0.2, one step (in our synthesized data) below the target trend size of 0.3 (for both increasing and decreasing trends). Because the number was randomized but not counterbalanced, we used a series of Welch's t-tests to determine statistically significant differences. (Recall from Section 4 that [0..5] distractors of size 0.2 could be present, [0..10] of size 0.1, and [0..20] of size 0.05.) We found that if no distractors of size 0.2 were present (only some number of distractors of sizes 0.1 and 0.05), users were most accurate; they were least accurate with three such "close" distractors present. The difference between zero and three distractors was significant – $t(66)=2.488$, $p=0.015$; we also saw trends for zero distractors to be more accurate than one, two, or four distractors, but not five distractors. This result is somewhat counter-intuitive. Minor differences across the techniques did not reveal any interesting findings, except that no errors were made by users in cases of zero distractors in the Juxtaposed Maps and Data-Driven Spots visualizations. We saw a significant difference based on the proximity of such distractors; if the distractor was adjacent to the target, users did better than if the nearest distractor was five counties away; using Welch's t-test, $t(9)=2.448$, $p=0.037$. Trends were noticed for other distances.

We did note the presence of some "catastrophic" errors by users, cases in which it would appear that the user searched for the wrong type of trend (increasing instead of decreasing, or vice-versa). There were 37 such errors overall, or 1.7% of the 2160 data trials. There was no apparent pattern amongst the visualization techniques, and these outlier points were not removed for the preceding analysis. We also tested for correlations between the time spent per trial and error, as well as the time spent on tutorials and error. Neither analysis lended support to the possibility that increased time spent on either tutorials or trials improved the subjects' performances. We found a standard practice effect on response time, but not for error. Users generally got faster with successive trials (but not monotonically so); there was no significant difference in the error.

One possible explanation for errors is that subjects mistakenly assumed that the extreme final value was achieved by the county with the greatest trend (maximum final value implying greatest increase or minimum final value implying greatest decrease). This was often the case with our simulated data, but not always. We found that this type of error accounted for 13.1% of the errors overall, but it happened in 35.5% of the cases where the extreme value resided with a county that did not exhibit the greatest trend. So it appears that this was a noteworthy source of error for our users. We analyzed this error using a binary dependent variable of occurrence and a 6 (Visualization) $\times$ 2 (Trend Type) $\times$ 10 (Repetition) repeated-measures ANOVA. Visualization tended to influence the occurrence of this error, but it was due to a low rate with Dimensional Stacking. Since other issues (discussed below) seem to have dominated the errors users made with Dimensional Stacking, we discount this as a trend. Trend type showed a main effect – $F(1,17)=17.958$, $p=0.001$ – and an interaction with Visualization – $F(5,85)=3.153$, $p=0.006$ on this type of error. Color Blending was most affected by Trend Type; Brush Strokes and Oriented Slivers were also affected. The interpretation of this result appears in the Discussion, below.

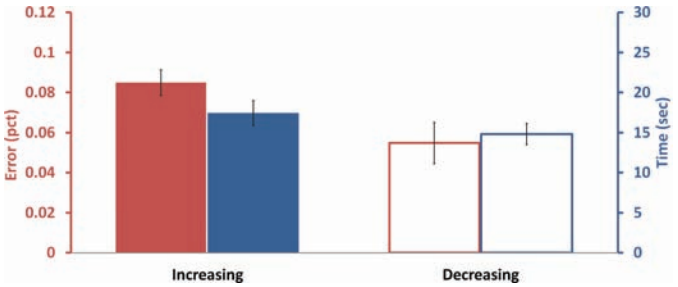

Fig. 9. The trend type – increasing or decreasing – had a main effect on both the error and time. Users were faster and more accurate in identifying the greatest decrease rather than the greatest increase.

6 Discussion

There are two ways in which we believe our results contribute to our ongoing effort to understand multi-variate visualizations, how to evaluate techniques in a task-based context with users and how to improve or understand individual multi-variate visualization techniques.

6.1 Evaluation of Multi-variate Visualizations

In comparing our results to previous multi-variate evaluations, we hesitate to draw many general conclusions because the task in our current study differs from previous tasks. As noted above, one critique of evaluations of multi-variate visualizations is that in many (but not all) cases, the task could be accomplished with only one layer of data; this was true for our previous study [16].

The most direct comparison for this study is the work of Joshi [10]. His trend recognition task at a point ranked pointillism-based techniques, such as Data-Driven Spots, as yielding better accuracy than a panorama of snapshots, akin to Juxtaposed Maps. We found no statistical difference between these two techniques for the spatial comparison of trends, although both led to better performance than the remaining techniques that we tested (Figure 7). For now, we can only presume that the difference in the task caused the different result and note this as an issue to be investigated in future studies.

We see some common issues emerging from the collected literature of studies. We commented previously [16] on the importance of sampling density and renew that concern here. We modified the techniques as described in Section 3 to avoid sampling in boundary regions out of concerns that such samples could confuse the users into incorrect analysis of the data. Previous authors who used sampling as a cue [7, 19] found similar issues with the clarity of the data presentation.

One could argue that our task is somewhat artificial in that if a user were concerned with finding such trends, then these trends should be explicitly represented by the data and the visualization; users will most likely perform the task better with explicit representations [10, 13]. While this is true for patterns such as the increasing or decreasing

trends over time we depicted, we contend that the possibility of bringing unforeseen patterns to the user's attention is worthy of investigating the impact of visualizations on such unexpected discoveries [20]. Further, we note that the direction of a trend was not inherently more or less difficult; differences that we saw with respect to trend resulted from Color Blending and the representation of each Trend Type. Its green hues (increasing) were quite low in saturation relative to the red (decreasing); thus, the error was higher for increasing trends, which indicates the need for the color to stand out more.

We again found that Color Blending led to higher error overall than most other techniques, as with our previous study. (This was in spite of the accuracy with the saturated red for decreasing trends.) Other authors have found pointillism techniques to out-perform Color Blending [5, 22] and noted the importance of color selection for the success of Color Blending. We find that this also applies to the color implementation of Dimensional Stacking we devised for this experiment. Clearly, these techniques can be improved for this and other fundamental tasks performed with visualization.

As we continue in our progression of user studies of multi-variate visualizations, we gain some insights about the difficulties in conducting such evaluations and user behavior with the techniques. For example, we asked users to identify any strategies they may have used in the course of completing the tasks. Five users noted that they developed a notion of an "outlier" within data layers (which likely caused the response to the saturated red in Color Blending). Three of these users explicitly described working "backwards" by trying to identify outliers in the final year's data, then trying to determine the change from the initial year's data for that county. In this way, the users reduced the data on which they focused to the minimum needed in order to answer the question. It would be fair to say that our visualizations contained two layers of data critical to successful task completion (first and last year) and three layers (intermediate years) of distracting data. Thus another benefit of adopting Joshi's task would be to require all five data layers to be attended by the user; this would truly test the capabilities and benefits users receive from multi-variate visualizations.

One subject, perhaps more experienced in the use of statistical graphics as well as more advanced visualizations than others, wished that our tutorials were more extensive. Based on the user's comments as well as our own observations, we can offer the following critique of the tutorials. We gave an initial screen with hints (detailed in Section 3), but these hints were not available once the user saw the practice questions. Further, we did not provide images with the hints so that the user could immediately see illustrations of the potential cues. In our previous study, we offered one tutorial question per technique. We increased that to two, one of each Trend Type, but this is still clearly a minimalist tutorial for techniques, which were completely unfamiliar to most users. (Five subjects had participated in our previous study.) We could examine the effects of tutorials in future studies. We noted a standard practice effect on response time, but not on error. We saw no correlation between time spent on tutorials and error or between error and response time. Also, experts do not always perform better than non-experts [13]. One could argue, however, that domain experts are the likely users of the techniques we ultimately develop, and thus their assessments are of greater interest [21].

One point emphasized by a subject and supported by the "catastrophic" errors noted above is that we should have separated the increasing and decreasing trends into separate blocks. It appears that sixteen users made such an error at least once. Separating these types of questions would improve our study design.

6.2 Implications for Multi-variate Visualizations

We now discuss each technique in terms of its success enabling the task we presented to the subjects and what we could have improved about our encoding. In general, we conclude that Data-Driven Spots performed the best overall, Dimensional Stacking was hindered by a precision issue, and in general the tutorial sections of the survey could have been longer and more informative.

Juxtaposed Maps This method of providing the user with a series of grayscale images was included in the study as a baseline technique. Due to the well-known concept of spatial blindness, it is difficult to mentally overlap images. However, since the information is available, it is still possible to find the correct answer with patience. In light of this fact, it follows that Juxtaposed Maps was the most accurate technique but was among the most time-consuming (Figure 7).

Brush Strokes The Brush Strokes technique used five graphical attributes to encode the layers. To notice the trend, the subjects needed to follow the change of all five attributes, based on an arbitrary mapping, and thus needed to repeatedly consult the key (Figure 2). As stated in Section 3, we wrote hints in the tutorial which described the type of strokes to determine the trends in each county. While comparable in speed to the other techniques, it was one of the least effective in terms of accuracy. Subjects were able to find patterns that fit the criteria, but failed to differentiate between a variety of distractors and the correct answer. It is possible the subjects did not spend enough time reading the key or images before answering. It may also be that the encodings do not generally lend themselves to wide-range, continuous values. Length and width have low resolution for this data; orientation, while of sufficient resolution, is unconventional for most novice users; even the heat map is not as effective as intensity for continuous data. We experimented with using hue as the first value and intensity as the final value for our five years (data layers), but rejected it during pilot testing. The strokes were no smaller than the slivers, and both fared poorly with respect to error. One may suspect that this size was insufficient, although our user data can not offer any evidence to support this. Brush Strokes were affected negatively by a tendency to focus on the extreme value, especially in the decreasing trend; it seems likely that the dark blue representation of (near) minimum values lent itself to mis-interpretation of the trend.

Data-Driven Spots We settled on a set of similarly-sized, non-overlapping dot layers positioned on rigid grids. DDS fared quite well among the techniques, so no distinct issue arises within the approach. However, it is possible that a more organic approach, where the dots vary in size and are not confined to a grid, could create images where trends appear more salient than in the current rendering.

We noticed that subjects were faster, but not more accurate, in localizing decreasing trends. Although this was not a statistically significant result, we see a possible explanation. In Figure 3, the county in the upper left shows an increasing trend, while the county to its right shows a decreasing trend. This decreasing trend is much more salient, because the larger green and blue dots are completely invisible and the red dots are very bright. Meanwhile, the dots in the increasing county seem very similar, but the red color is still noticeably dimmer than its base value. This technique creates a field-of-view effect which allows the user to parse the trend from the pattern. These types of representational issues are the insights we seek.

Oriented Slivers Oriented Slivers performed much more poorly than we expected for this task. This technique seems more suited for distinguishing (segmenting) overlapping features with large variation in surface value. Here, the variation between layers was muted, and the boundaries for every layer were the same. For this reason, it was hard for the users to perceptually separate the individual layers. In addition, the blending of sliver layers created bright spots at the point of overlap. This may have misled users about some data values.

For this study, we oriented the slivers so that they spanned the full rotational range about their central location. In the previous study, we used cardinal directions and their divisions, which led to a fifth layer which was difficult to differentiate from another layer (since it had the recommended minimum separation of 15°). As noted above, the size of the slivers was no larger than the size of the strokes; we have no evidence to support a claim that the size of the strokes or slivers impeded performance, but it is a potential hypothesis for future studies. The Oriented Slivers technique was affected by the tendency to select the extreme value rather than the trend; it appears that the low intensity of the sliver for the final year led users astray. This may indicate a need for an improved intensity ramp (including, perhaps, a minimum value).

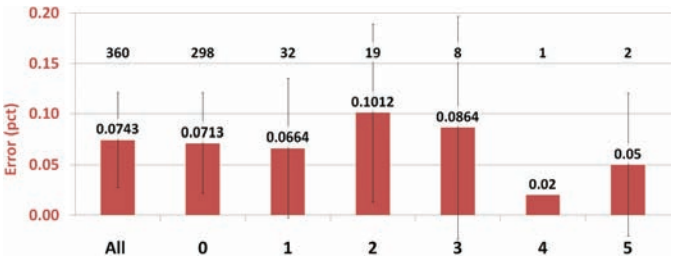

Fig. 10. The error rates for Dimensional Stacking do not appear to have been adversely affected by identical stacks. As this graph shows, the mean error in the 298 cases without a stack identical to the target (second bar) is not significantly different from the performance on all 360 trials of Dimensional Stacking (first bar). The remaining bars show the performance with each number of identical stacks that occurred. The mean errors for each case are shown at the top of the bar, and the number of such cases floats above each bar.

A possible improvement would be to implement the slivers as a single glyph at each sample instead of a series of repeating patterns. This would be a similar adaptation as we made for Dimensional Stacking, centering a non-repeating representation in each county.

Color Blending Color Blending was found to be faster and more accurate for decreasing trends over increasing trends. Our encoding equation and color set (Section 3.1) produced a strong red color when the trend was greatly decreasing, while it produced a less salient shade of green when the trend was greatly increasing. This fact clearly contributed to the poor performance on increasing trends. It appears to have caused the users to select the extreme value, especially for decreasing trends. It is possible that we could modify the equation so that both increasing and decreasing trends produce bright, qualitative colors. But separating the extreme value and trend would still be a difficult task. This underscores the difficulty of Color Blending for as many as five values.

Dimensional Stacking The limitation of the precision in which the values were encoded created situations where there were distractors represented with the exact same glyph as the target (called "identical stacks" in Section 3.2). Since the subject could only choose one county, they were forced to make a random choice between the top contenders. Our major shortcoming here was not to demonstrate this limitation in the tutorial. However, we looked for evidence that this ambiguity caused increased error and failed to find it. As Figure 6.2 shows, the mean error for cases in which there were no identical stacks was not significantly different from the mean error for all trials of Dimensional Stacking. It could be that this problem caused slower responses and increased workload ratings, however.

Several subjects noted Dimensional Stacking mapped values to the color spectrum in the reverse order that it is often mapped (Figure 4(c)). The colors were arranged in order from red to green as they are in the visible color spectrum, but it transitioned from warm to cool as the value increased; two users mentioned this during debriefing. It is more frequently the case that cool colors are mapped to low values, and warm colors are mapped with high values. In fact, we did just that with the hue attribute of Brush Strokes (Figure 2). It is also well-known that color is not a metric quantity and thus does not lend itself well to continuous variables such as our data layers. This may have been a factor in causing Dimensional Stacking to yield lower performance than other the techniques (most of which use intensity to encode data values); as noted above, this may have limited performance with Brush Strokes as well.

With regard to the size of the blocks; they were at least as large as the Gaussian kernels of DDS (modulo the shapes). The performance of Dimensional Stacking seems to have been impaired by other factors, but we note the potential (as with size of Brush Strokes and Oriented Slivers) of this as a variable for future studies.

7 Future Work

In the discussion above, several issues were raised that will inform our future work. We have a list of potential improvements to the individual techniques, as well as improvements in the study design. We tried to confine our improvements in the current study to changes within the original definition of the respective techniques. One can certainly imagine an extension to Dimensional Stacking, such as a heat map for value [17], as we plan for a follow-up study. Extensions to the other techniques, such as color encoding for Oriented Slivers, are also an option. Our study design could be extended to explicitly include the number of distractors and/or proximity of distractors as independent variables.

We plan to look for insights in the data regarding performance benefits of extended exposure to the techniques; we could ask users to return for a future study and measure performance improvements in that way as well. We could also separate users into "novice" and "expert" categories and look for differences between these groups. Alternatively, we could provide feedback during the study to examine learning effects. Finally, we hope to add to our library of tasks, finding tasks that require users to focus on as many data layers as we can, in order to test the limits of insight from multi-variate visualizations.

Acknowledgments

The authors wish to thank Dan Carr, Georges Grinstein, Barry Haack, and the anonymous subjects. This work was supported by the NRL Base Program.

References

[1] J. Beddow. Shape coding of multidimensional data on a microcomputer display. In *Proceedings of IEEE Visualization*, pages 238–246, Oct. 1990.
[2] A. A. Bokinsky. *Multivariate Data Visualization with Data-driven Spots*. PhD thesis, The University of North Carolina at Chapel Hill, 2003.
[3] D. B. Carr and L. W. Pickle. *Visualizing Data Patterns with Micromaps*. CRC Press, 2010.
[4] J. W. Decker and M. A. Livingston. Poster: An interactive, visual composite tuner for multi-layer spatial data sets. In *IEEE Visualization*, 2010.
[5] H. Hagh-Shenas, V. Interrante, C. Healey, and S. Kim. Weaving versus blending: a quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, page 164, 2006.
[6] S. G. Hart and L. E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, pages 239–250. Elsevier Science Publishers, 1988.
[7] C. G. Healey and J. T. Enns. Building perceptual textures to visualize multidimensional datasets. In *IEEE Visualization*, pages 111–118, 1998.
[8] C. G. Healey, S. Kocherlakota, V. Rao, R. Mehta, and R. S. Amant. Visual perception and mixed-initiative interaction for assisted visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 14(2):396–411, March–April 2008.
[9] C. G. Healey, L. Tateosian, J. T. Enns, and M. Remple. Perceptually based brush strokes for nonphotorealistic visualization. *ACM Transactions on Graphics*, 23(1):64–96, 2004.
[10] A. Joshi. *Art-inspired techniques for visualizing time-varying data*. PhD thesis, The University of Maryland, Baltimore County, 2007.
[11] R. M. Kirby, H. Marmanis, and D. H. Laidlaw. Visualizing multivalued data from 2d incompressible flows using concepts from painting. In *Proceedings of IEEE Visualization '99*, pages 333–340, 1999.
[12] D. H. Laidlaw, E. T. Ahrens, D. Kremers, M. J. Avalos, R. E. Jacobs, and C. Readhead. Visualizing diffusion tensor images of the mouse spinal cord. In *Proceedings of IEEE Visualization '98*, pages 127–134, 1998.
[13] D. H. Laidlaw, R. M. Kirby, C. D. Jackson, J. S. Davidson, T. S. Miller, M. da Silva, W. H. Warren, and M. J. Tarr. Comparing 2D vector field visualization methods: A user study. *IEEE Transactions on Visualization and Computer Graphics*, 11(1):59–70, January/February 2005.
[14] J. LeBlanc, M. O. Ward, and N. Wittels. Exploring N-dimensional databases. In *Proc. of IEEE Visualization*, pages 230–237, Oct. 1990.
[15] H. Levkowitz. Color icons: Merging color and texture perception for integrated visualization of multiple parameters. In *Proceedings of IEEE Visualization*, pages 164–170, 420, Oct. 1991.

[16] M. A. Livingston, J. W. Decker, and Z. Ai. An evaluation of methods for encoding multiple, 2D spatial data. In *SPIE Visualization and Data Analysis*, Jan. 2011.

[17] J. R. Miller. Attribute blocks: Visualizing multiple continuously defined attributes. *IEEE Computer Graphics & Applications*, 27(3):57–69, May/June 2007.

[18] R. M. Pickett and G. G. Grinstein. Iconographic displays for visualizing multidimensional data. In *Proc. of the 1988 IEEE Intl. Conf. on Systems, Man, and Cybernetics*, pages 514–519, Aug. 1988.

[19] Y. Tang, H. Qu, Y. Wu, and H. Zhou. Natural textures for weather data visualization. In *Tenth International Conference on Information Visualization*, pages 741–750, July 2006.

[20] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005.

[21] T. Urness, V. Interrante, E. Longmire, I. Marusic, S. O'Neill, and T. W. Jones. Strategies for the visualization of multiple 2D vector fields. *IEEE Computer Graphics & Applications*, 26(4):74–82, 2006.

[22] T. Urness, V. Interrante, I. Marusic, E. Longmire, and B. Ganapathisubramani. Effectively visualizing multi-valued flow data using color and texture. In *IEEE Visualization*, pages 115–121, Oct. 2003.

[23] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2000.

[24] C. Weigle, W. Emigh, G. Liu, R. M. Taylor II, J. T. Enns, and C. G. Healey. Effectively visualizing multi-valued flow data using color and texture. In *Graphics Interface*, pages 153–162, 2000.

[25] M. X. Zhou and S. K. Feiner. Visual task characterization for automated visual discourse synthesis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 392–399, Apr. 1998.