UNDERSTANDING HUMAN FACTORS OF BIG DATA VISUALIZATION

Mark A. Livingston, Ph.D.
Jonathan W. Decker
Zhuming Ai, Ph.D.

United States
Naval Research Laboratory
Washington, DC
Research Interests

- How much data can we show before operator overload?
- Which multivariate visualization (MVV) is best for which task?
- Can users discover patterns without explicit representations?
- Does multivariate visualization solve big data visualization?
- Ultimate goal: explain why a visualization representation works (or doesn’t)
MVV Techniques

Data-driven Spots
- Color denotes variable
- Intensity encodes value

Oriented Slivers
- Orientation denotes variable
- Intensity encodes value

Attribute Blocks
- Position denotes variable
- Heat map encodes value
## MVV Techniques

<table>
<thead>
<tr>
<th><strong>Brush Strokes</strong></th>
<th><strong>Color Blending</strong></th>
<th><strong>Dimensional Stacking</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke properties encode variable</td>
<td>Hue encodes variable</td>
<td>Boxes encode variable</td>
</tr>
<tr>
<td>Intensity, length, width, hue, and orientation encode value</td>
<td>Weighted average of hues encodes value</td>
<td>Discretized hue encodes value range; extensions use heat maps</td>
</tr>
</tbody>
</table>

---

**Techniques**

- **Stroke properties:** encode variable properties such as intensity, length, width, hue, and orientation.
- **Hue encoding:** represents variable value using a weighted average of hues.
- **Boxes encoding:** discretized hue encodes value range, with extensions using heat maps.
<table>
<thead>
<tr>
<th>Techniques</th>
<th>Task(s)</th>
<th>Findings</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gridded icons, Jittered grid icons, Brush strokes, Line integral convolution, Image-guided streamlines, Grid-seeded streamlines</td>
<td>Localize critical values, Identify critical point type, Advect particle</td>
<td>Image-guided streamlines 1.5 standard deviations (SD) below mean error and 1.0 SD below mean response time (RT) for advection; also 1.0 SD below mean error and RT for identifying critical point type; LIC 1.0 SD below mean error and RT for critical point localization</td>
<td>[Laidlaw et al., 2005]</td>
</tr>
<tr>
<td>Color weaving, color blending</td>
<td>Read sequence of data values</td>
<td>Weaving better by 4-11 percentage points</td>
<td>[Hagh-Shenas et al., 2006]</td>
</tr>
<tr>
<td>Multi-layer texture synthesis, Brush strokes</td>
<td>Localize critical values</td>
<td>Texture synthesis better by 7.5 percentage points</td>
<td>[Tang et al., 2006]</td>
</tr>
<tr>
<td>DDS, Oriented slivers, Brush strokes, Color blending, Stick figures</td>
<td>Localize critical values</td>
<td>#1 DDS: 39-58% of mean error of other techniques; #2 Slivers: 44-65%; of mean error of others; Monitor sensitivity</td>
<td>[Livingston et al., VDA 2011]</td>
</tr>
<tr>
<td>DDS, Oriented slivers, Brush strokes, Color blending, Dimensional Stacking, baseline</td>
<td>Detect greatest trends</td>
<td>#1 Baseline: 48-57% of mean error of MVVs; #2 DDS: 61-72%; of mean error of remaining MVVs; Users fooled by extreme value, better if target was near “distraction”</td>
<td>[Livingston et al., TVCG 2011]</td>
</tr>
<tr>
<td>DDS, Oriented slivers, Brush strokes, Color blending, Attribute blocks, baseline</td>
<td>Detect greatest trends</td>
<td>Attribute blocks better than dimensional stacking in previous test; main effect of distance to distraction</td>
<td>[Livingston et al., VDA 2012]</td>
</tr>
<tr>
<td>DDS, Oriented slivers, Attribute blocks, baseline</td>
<td>Detect multi-way overlap</td>
<td>MVVs better than baseline; DDS best among MVVs; inconclusive evidence for density and for density gain as key perceptual feature to predict error with MVVs</td>
<td>[Livingston et al., TVCG 2012]</td>
</tr>
<tr>
<td>DDS, Oriented slivers, Attribute blocks (3x2), Oriented DDS, Attribute blocks (6x6)</td>
<td>Detect multi-way overlap</td>
<td>Slivers and Attribute (3x2) best; new techniques were fast but inaccurate; confirmed density gain as key perceptual feature to predict error</td>
<td>[Livingston et al., VDA 2013]</td>
</tr>
</tbody>
</table>
Discussion

Perceptual and Cognitive Interpretations
1. Use salient cues to direct exogenous attention

Exogenous attention may be directed by varying color or texture in MVV.

- Attribute blocks
- Data-driven spots
- Oriented slivers
Relationship of Density to Error

**ODDS**  \( R = -0.35, p < 0.073 \)

**Attribute**  \( R = -0.38, p < 0.056 \)

**Slivers**  \( R = -0.45, p < 0.020 \)

**Temporal AB**  \( R = -0.49, p < 0.014 \)

**DDS**  \( R = -0.42, p < 0.030 \)
Earth Mover’s Distance Results

**Attribute blocks**
- **maxEMD vs error**
- **meanEMD vs error**

**Data-driven Spots**
- **maxEMD vs time**
- **meanEMD vs time**

**Oriented Slivers**
- **maxEMD vs time**
- **meanEMD vs time**
Sensitivity to Monitor Settings

Error (normalized)

- Slivers
- BrushStrokes
- DataSpots
- ColorBlend
- StickFigures

Significant difference

Lights on Factory settings

Lights off Factory settings

Lights on Altered settings
Edge Strength

Data-driven spots

Mean edge strength

Oriented Slivers

Mean edge strength
Luminance, Target Size, & Value

By target size

By target value

Standard deviation of luminance

Standard deviation of luminance
2. **Provide strong grouping cues for information**

Working memory has a limited capacity, but the contents can be facts or concepts. Knowledge structures, such as patterns in a visualization that may be associated with data properties, can be retrieved from long-term memory through cues that give rise to similar encoding.
Trend Detection Study

Baseline visualization had least error
Some techniques had ambiguous representations
Needed two variables to complete task
3. **Provide strong grouping cues to facilitate chunking**

Reduce competition for working memory capacity by helping user to group pieces of information into larger “chunks.” Assist this by perceptual cues.

![Graph showing error, answers, time, and workload](image)

- **Error (Target Layers) and Number of Changes to Answer**
- **Time (sec) and Workload (TLX)**
Critical Point Study Results

- **Error**: Data-driven Spots and Oriented Slivers best
  
  $F(4,56) = 9.8364, \ p = 0.000$  
  Outliers not removed – 7 for $Z>4$ or 24 for $Z>3$

- **Time**: Data-driven Spots and Color Blending best
  
  - Intuitiveness? Giving up? $F(4,56) = 34.0763, \ p = 0.000$

- **Workload (TLX)**
  
  - User confidence? $F(4,56) = 4.9599, \ p = 0.002$

Need only a single variable to complete task

![Graphs showing error, time, and workload results](image-url)
Overlap Detection Task

![Graph showing error (variables present) vs. response time (sec) for different categories: Slivers, Attribute, DDSpots, Temporal, ODDSpots.](image-url)
4. Organize information for knowledge structures

Help user recall mental models, which involve possibilities and projection; these are in turn similar to sense-making. This should help a user understand data and turn it into actionable information.

Unfortunately, tasks in most user studies of multivariate visualization are quite low-level and performed by novice users. Laidlaw et al. (2005) found that image-guided streamline placement assisted in advection of particles and identification of critical point type.

This follows from the leverage point, since streamlines guide even novice users to the proper mental model of flow through the field, which is critical to each task.

<table>
<thead>
<tr>
<th>Tasks in Multivariate Visualization User Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify critical point type</td>
</tr>
<tr>
<td>Advect particle</td>
</tr>
<tr>
<td>Localize critical values</td>
</tr>
<tr>
<td>Read sequence of data values</td>
</tr>
<tr>
<td>Detect greatest trends</td>
</tr>
<tr>
<td>Detect multi-way overlap</td>
</tr>
</tbody>
</table>
Overlap Detection Task

<table>
<thead>
<tr>
<th></th>
<th>Data-driven Spots</th>
<th>Oriented slivers</th>
<th>Attribute blocks</th>
<th>Juxtaposed layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error (layers)</td>
<td>0.09</td>
<td>0.42</td>
<td>0.47</td>
<td>1.54</td>
</tr>
<tr>
<td>Time (sec)</td>
<td>7.33</td>
<td>9.62</td>
<td>9.74</td>
<td>46.72</td>
</tr>
</tbody>
</table>

![Data-driven Spots](image1)

![Oriented slivers](image2)

![Attribute blocks](image3)

![Juxtaposed layers](image4)
5. Structure information to provide strong retrieval cues

Structure information to provide strong retrieval cues for mental models to help analogical reasoning. Enable user to apply prior knowledge to new data patterns.
6. Develop training regimes for implicit learning

Statistical regularities may be learned implicitly; thus, visualizations could prime users to recognize certain data relationships through trained visual patterns that may be recognized.
Trend Detection Follow-up Study

- Improved some multivariate techniques; failed to improve others
- Theories regarding saturation, total intensity, and color maps
- Techniques that did not integrate variables tightly considered for further studies
Future Directions

- Gather more data!
- New tasks for user studies
  - More applied tasks?
- Beginning to unlock the meaning of various components of multivariate visualizations for the difficulty they provide in accomplishing visual analysis tasks
  - Data-driven spots: density and color difference, then edge strength
  - Oriented slivers: edge strength variation, then density
  - Attribute blocks: color variation, then intensity variation
Work supported by NRL Base Program

Thanks to anonymous subjects and reviewers of papers

Thank you!