



UNDERSTANDING HUMAN FACTORS OF BIG DATA VISUALIZATION

Mark A. Livingston, Ph.D.

Jonathan W. Decker

Zhuming Ai, Ph.D.

United States

Naval Research Laboratory

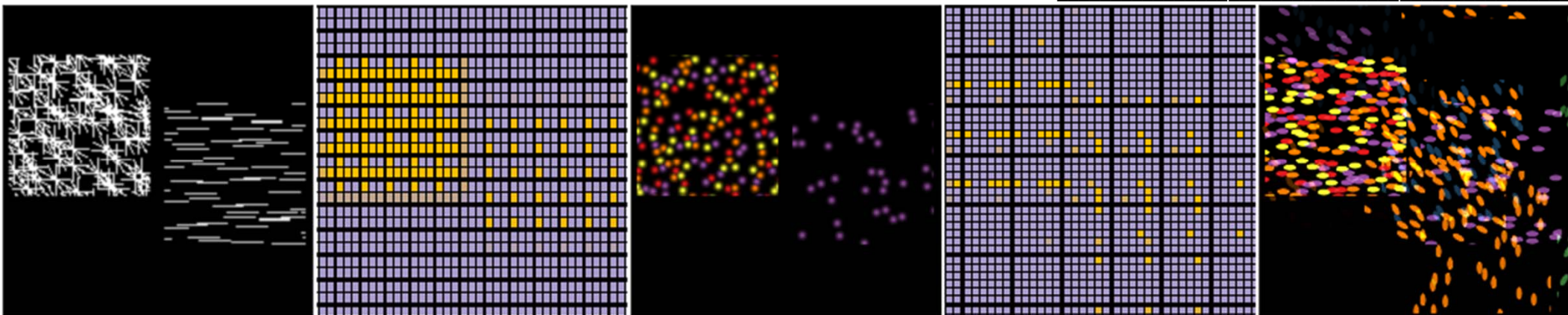
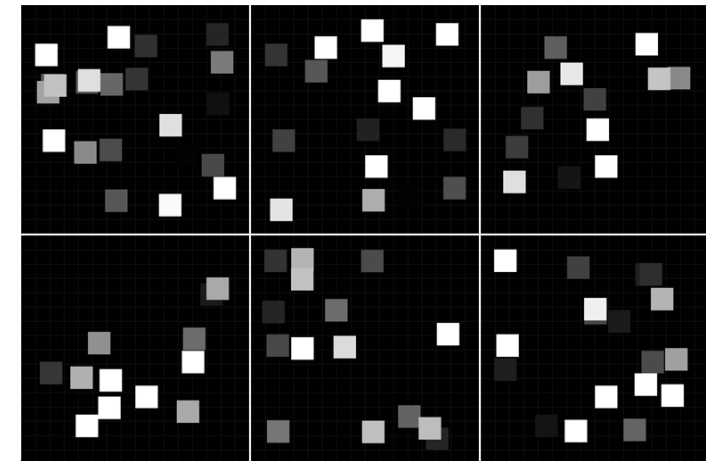
Washington, DC



Research Interests

1

- 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)

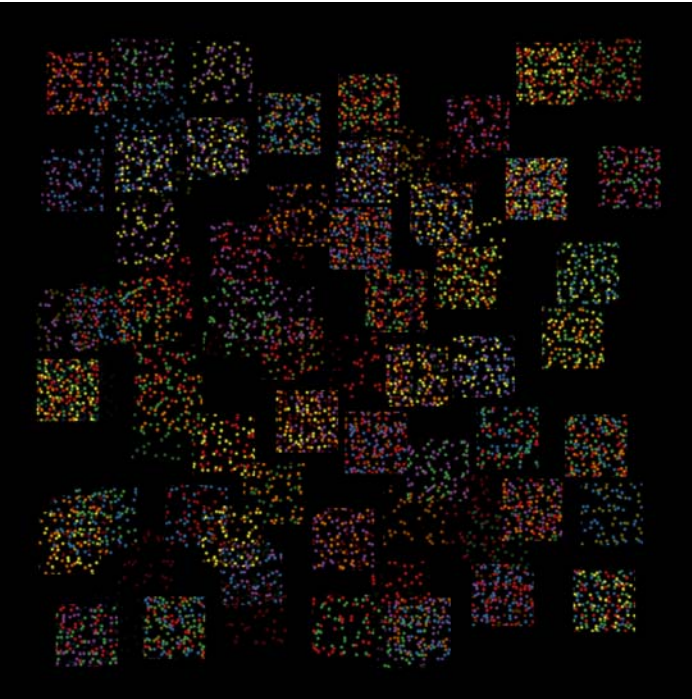




MVV Techniques

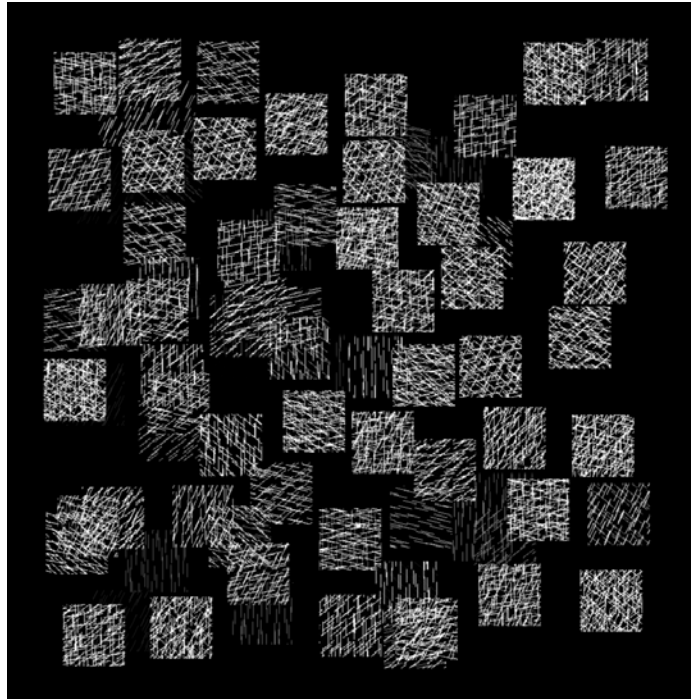
2

Data-driven Spots



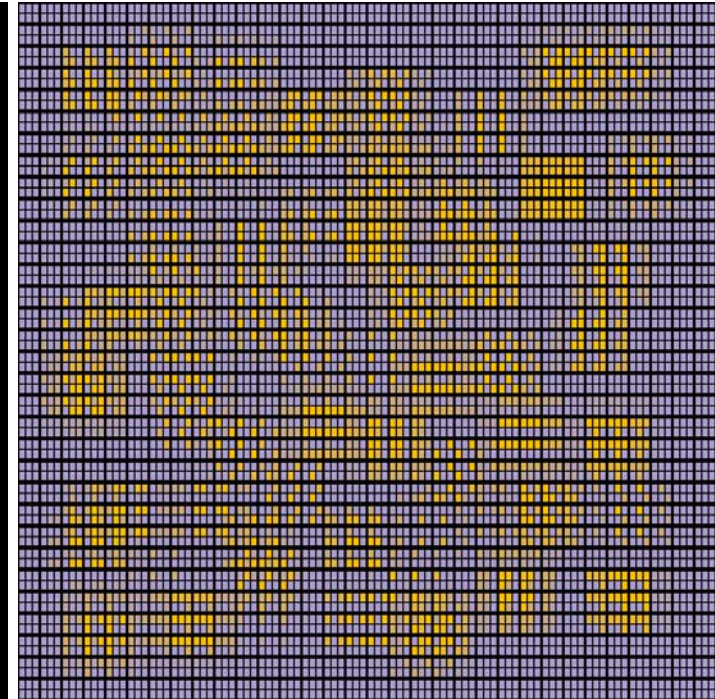
Color denotes variable
Intensity encodes value

Oriented Slivers



Orientation denotes variable
Intensity encodes value

Attribute Blocks



Position denotes variable
Heat map encodes value



MVV Techniques

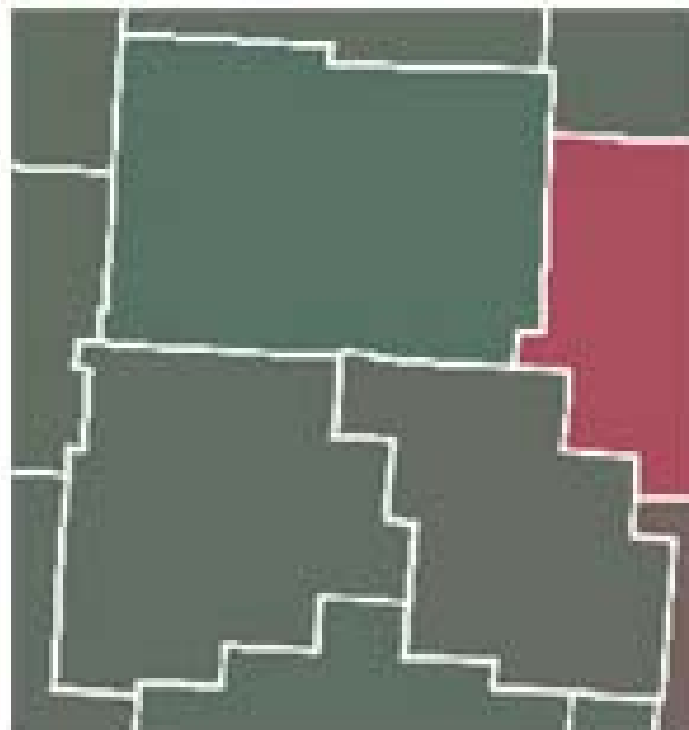
3

Brush Strokes



Stroke properties encode variable
Intensity, length, width, hue, and orientation
encode value

Color Blending



Hue encodes variable
Weighted average of hues
encodes value

Dimensional Stacking



Boxes encode variable
Discretized hue encodes
value range; extensions
use heat maps



Past Comparisons of MVVs

4

Techniques	Task(s)	Findings	Citation
Gridded icons, Jittered grid icons, Brush strokes, Line integral convolution, Image-guided streamlines, Grid-seeded streamlines	Localize critical values, Identify critical point type, Advect particle	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	[Laidlaw <i>et al.</i> , 2005]
Color weaving, color blending	Read sequence of data values	Weaving better by 4-11 percentage points	[Hagh-Shenas <i>et al.</i> , 2006]
Multi-layer texture synthesis, Brush strokes	Localize critical values	Texture synthesis better by 7.5 percentage points	[Tang <i>et al.</i> , 2006]
DDS, Oriented slivers, Brush strokes, Color blending, Stick figures	Localize critical values	#1 DDS: 39-58% of mean error of other techniques; #2 Slivers: 44-65%; of mean error of others; Monitor sensitivity	[Livingston <i>et al.</i> , VDA 2011]
DDS, Oriented slivers, Brush strokes, Color blending, Dimensional Stacking, baseline	Detect greatest trends	#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"	[Livingston <i>et al.</i> , TVCG 2011]
DDS, Oriented slivers, Brush strokes, Color blending, Attribute blocks, baseline	Detect greatest trends	Attribute blocks better than dimensional stacking in previous test; main effect of distance to distraction	[Livingston <i>et al.</i> , VDA 2012]
DDS, Oriented slivers, Attribute blocks, baseline	Detect multi-way overlap	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	[Livingston <i>et al.</i> , TVCG 2012]
DDS, Oriented slivers, Attribute blocks (3x2), Oriented DDS, Attribute blocks (6x6)	Detect multi-way overlap	Slivers and Attribute (3x2) best; new techniques were fast but inaccurate; confirmed density gain as key perceptual feature to predict error	[Livingston <i>et al.</i> , VDA 2013]

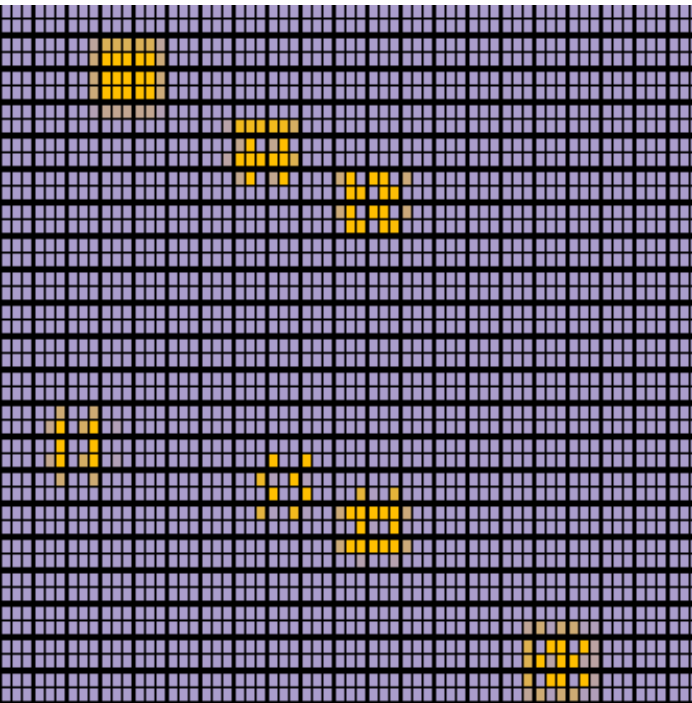
Discussion

Perceptual and Cognitive Interpretations

1. Use salient cues to direct exogenous attention

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

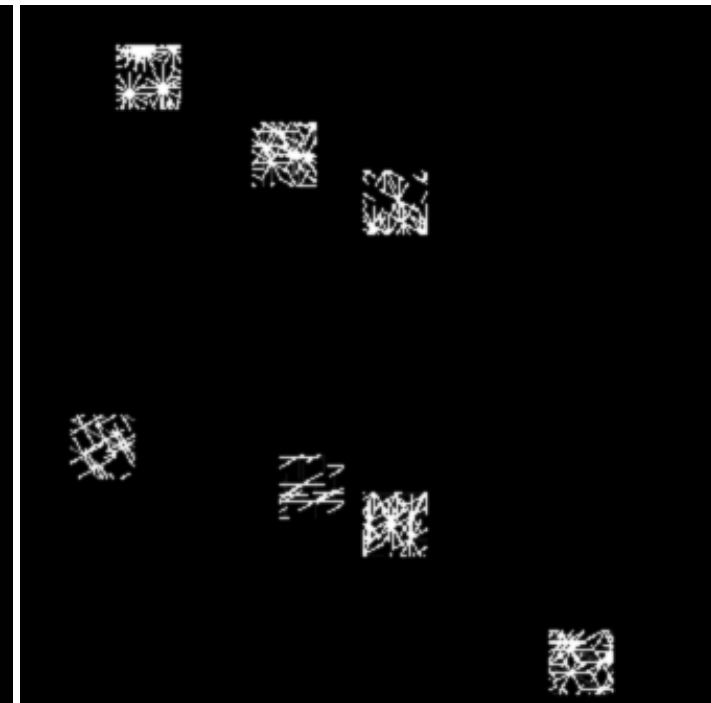
Attribute blocks



Data-driven spots



Oriented slivers

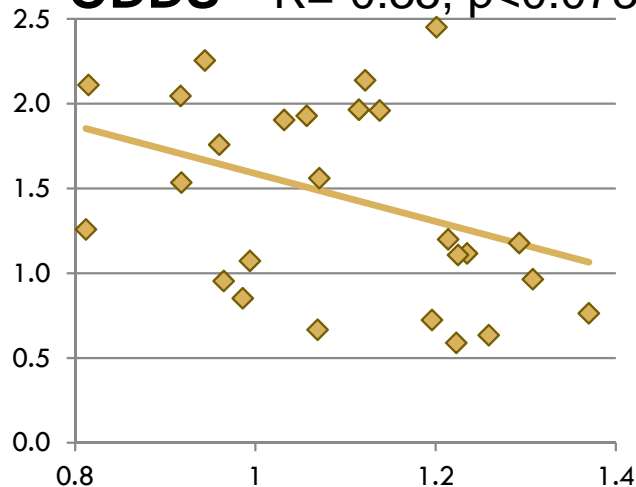




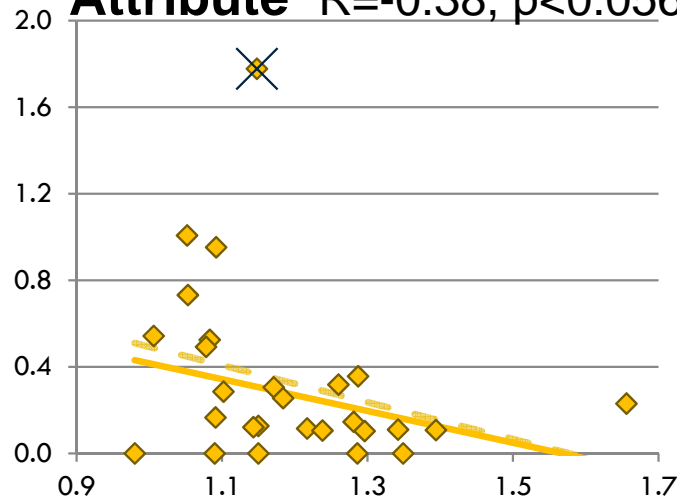
Relationship of Density to Error

7

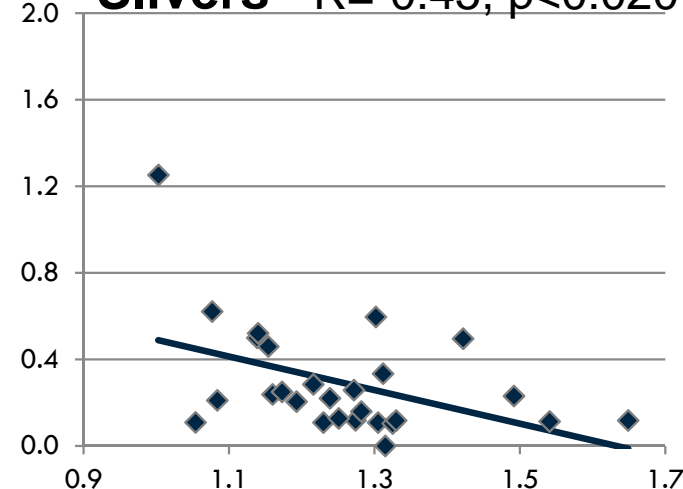
ODDS $R=-0.35, p<0.073$



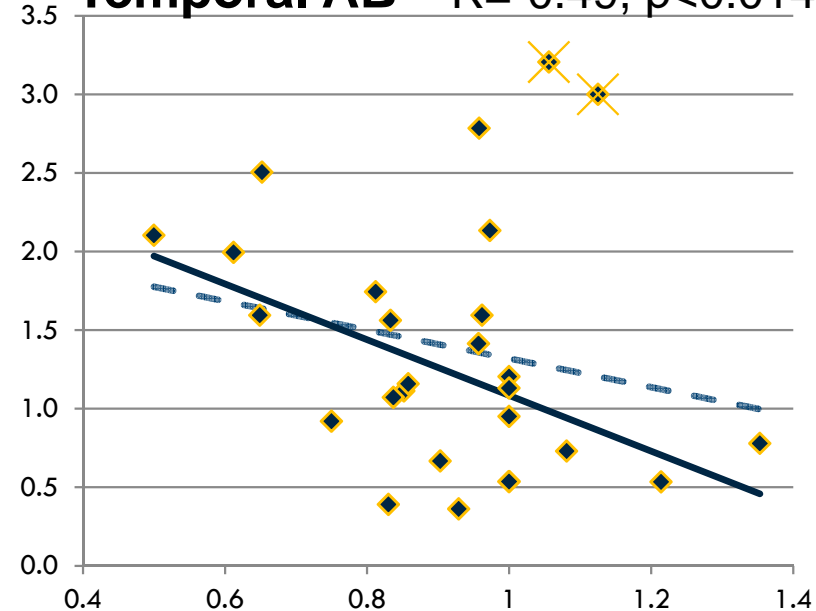
Attribute $R=-0.38, p<0.056$



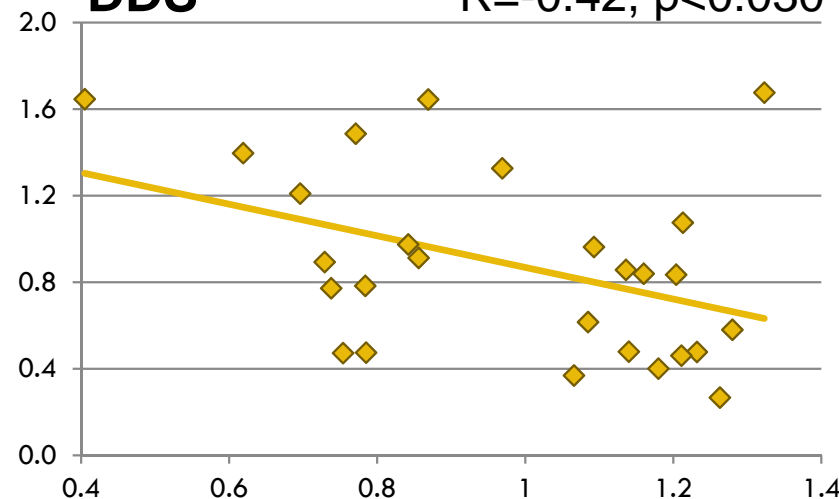
Slivers $R=-0.45, p<0.020$



Temporal AB $R=-0.49, p<0.014$



DDS $R=-0.42, p<0.030$

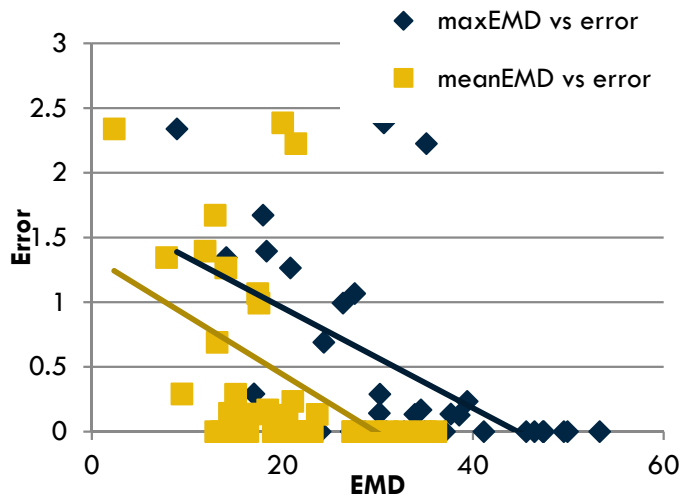




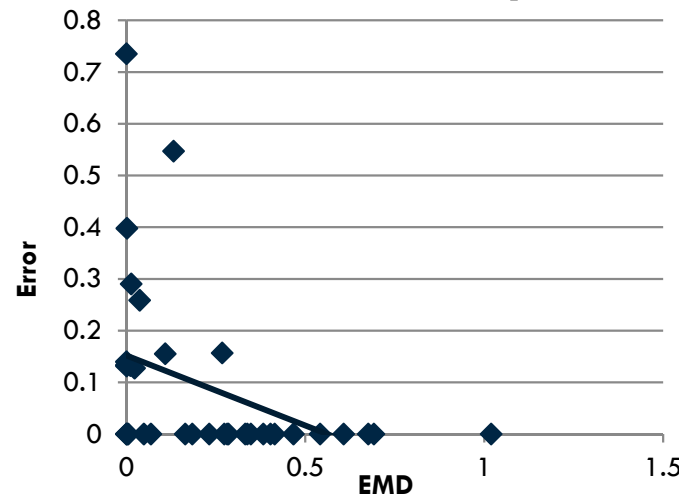
Earth Mover's Distance Results

8

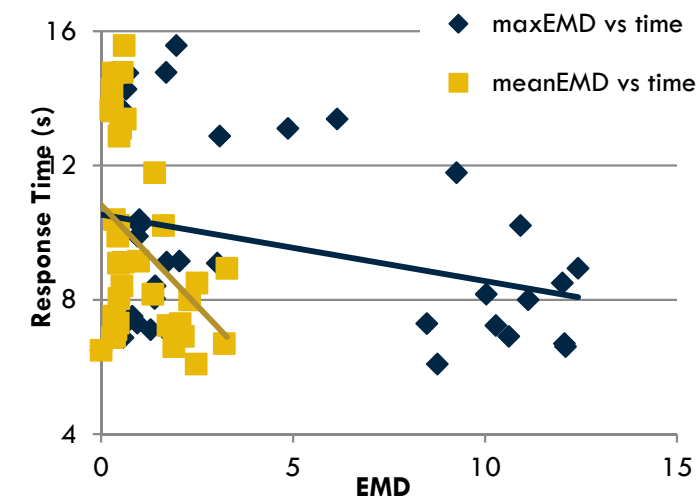
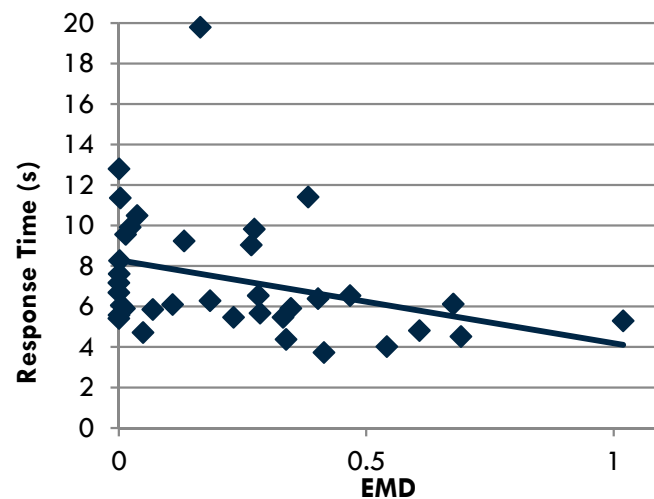
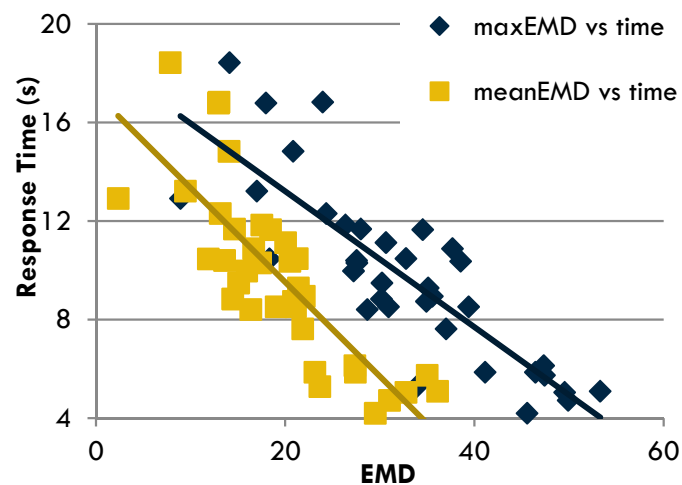
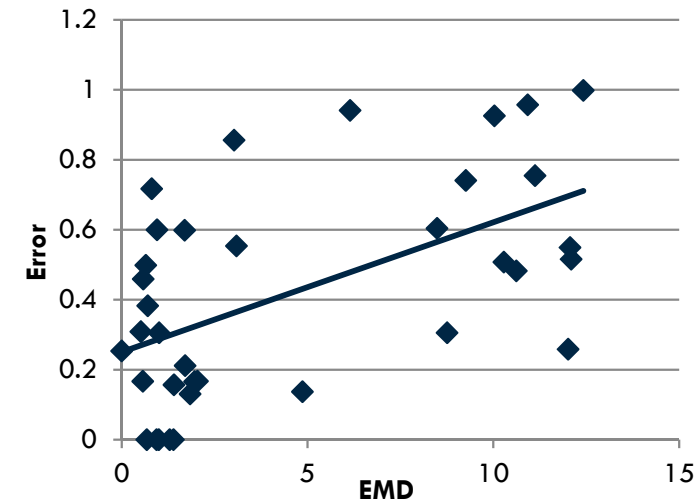
Attribute blocks



Data-driven Spots

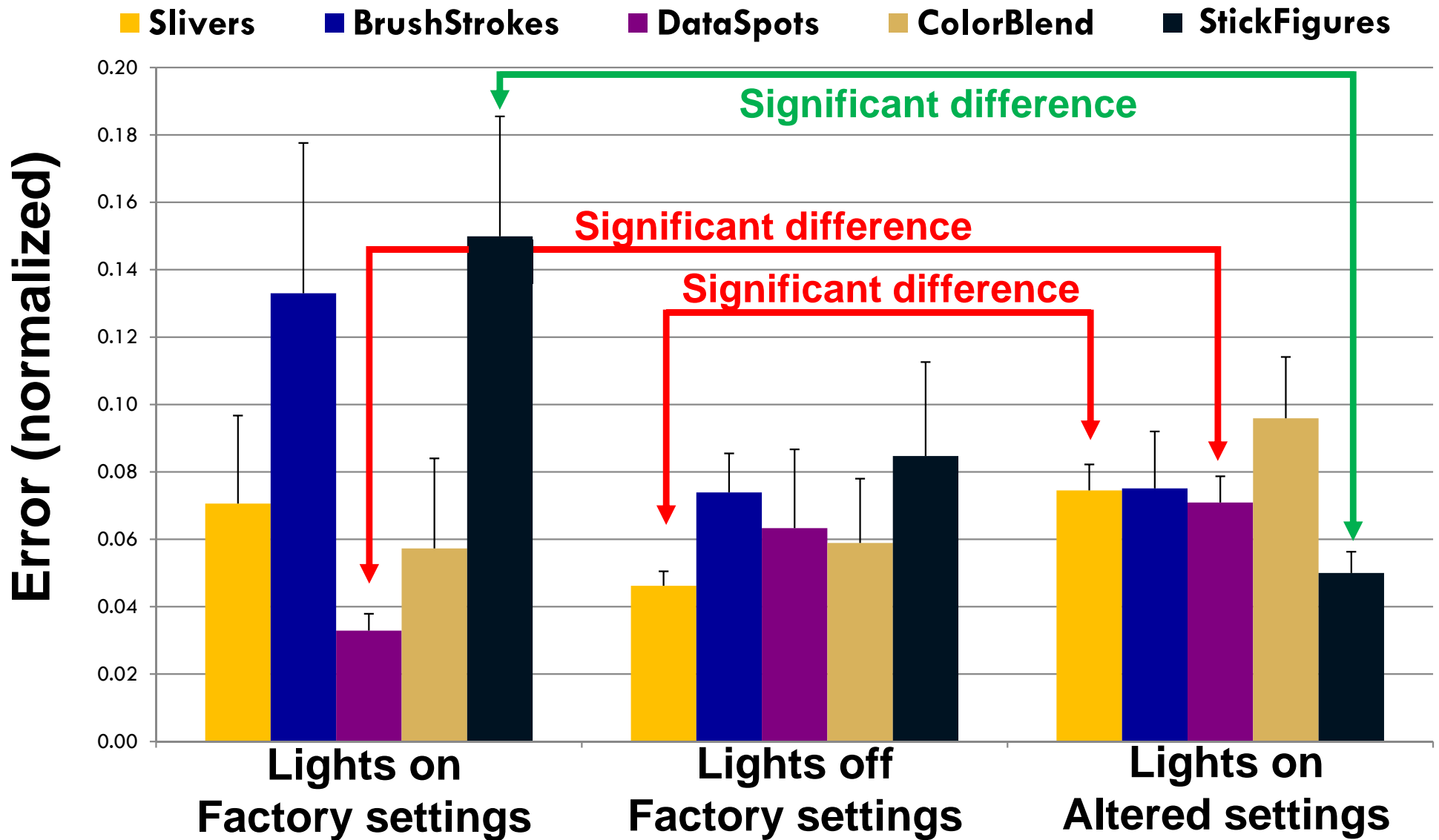


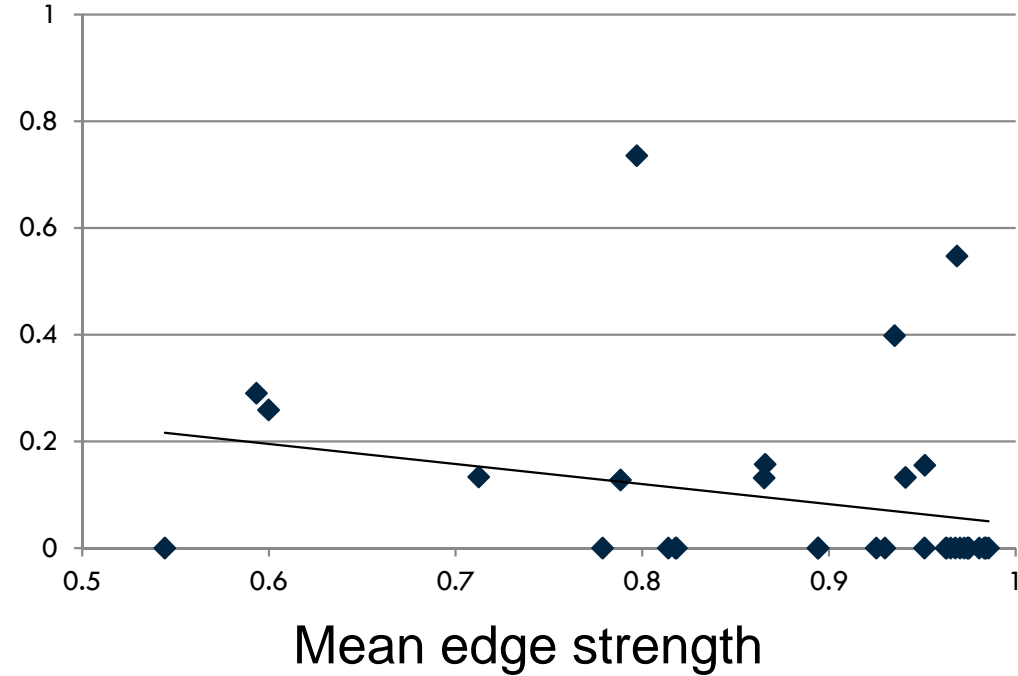
Oriented Slivers



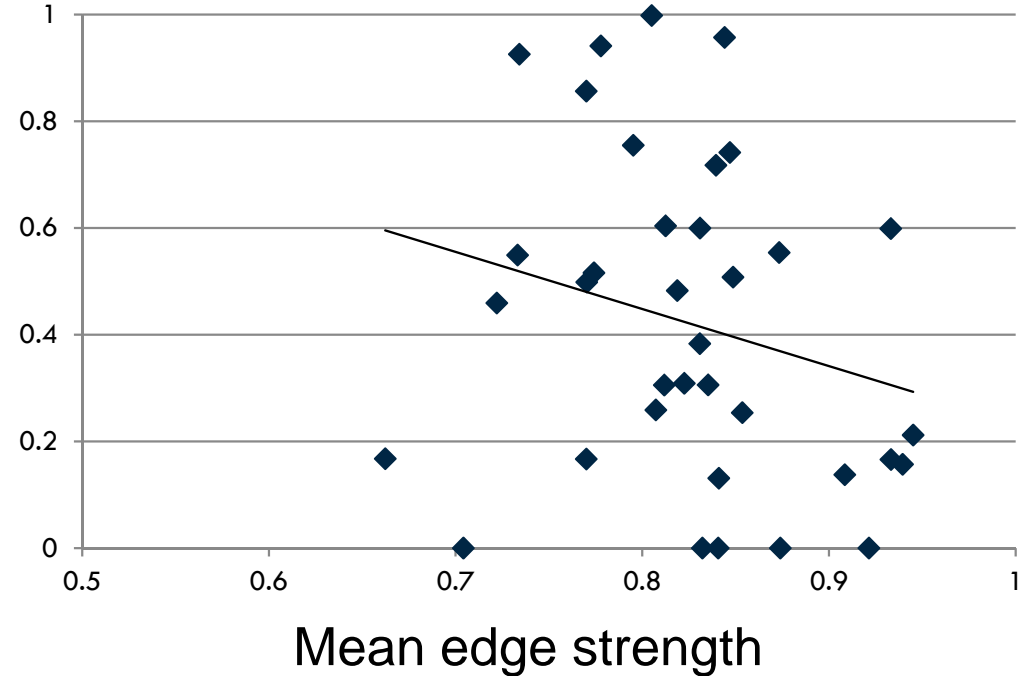


Sensitivity to Monitor Settings





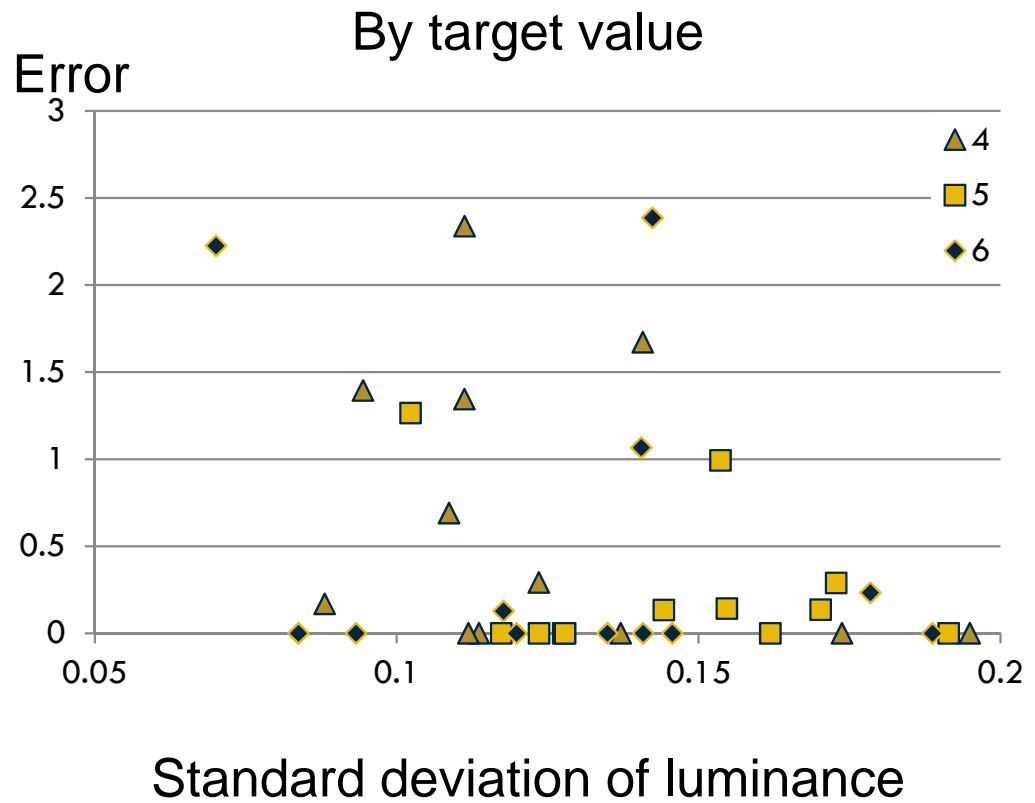
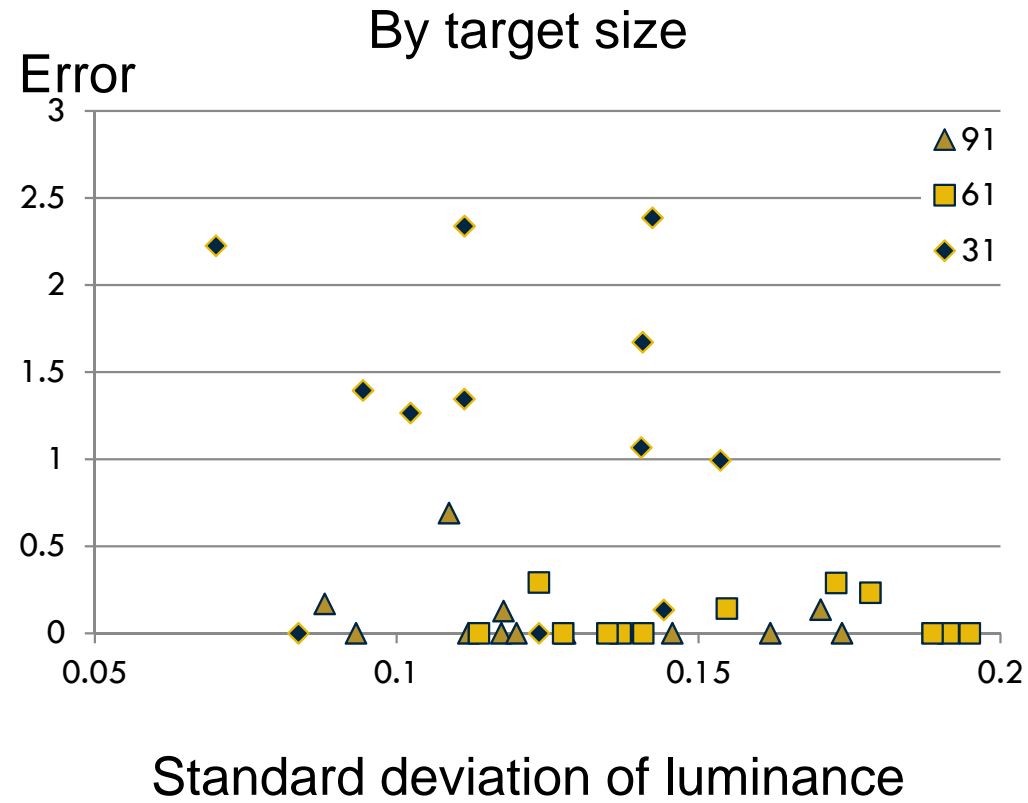
Oriented Slivers





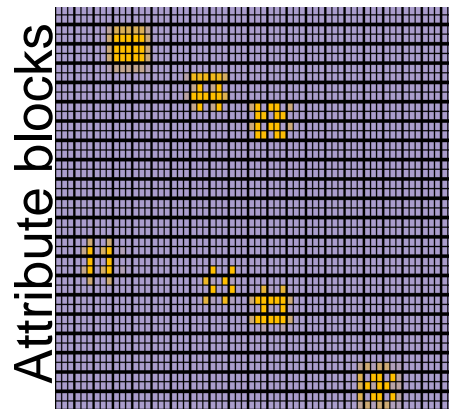
Luminance, Target Size, & Value

11

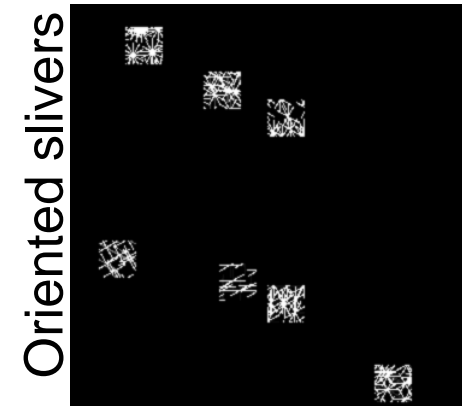
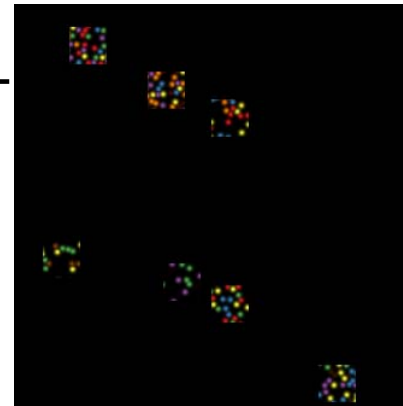


2. Provide strong grouping cues for information

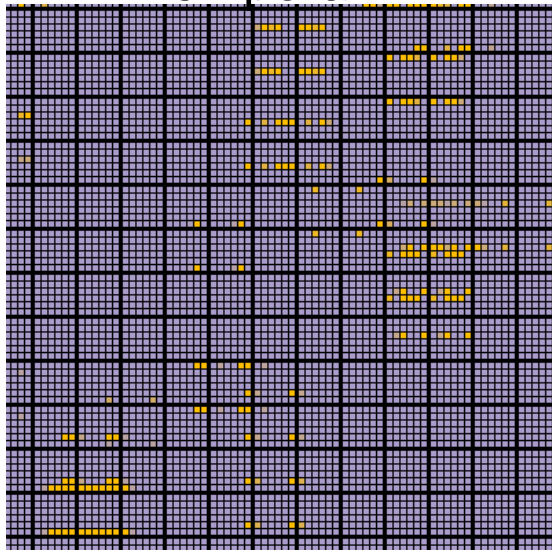
Working memory has a limited capacity, but the contents can be facts or concepts. Knowledge structures, such as patterns in a visualization that may be associated with data properties, can be retrieved from long-term memory through cues that give rise to similar encoding.



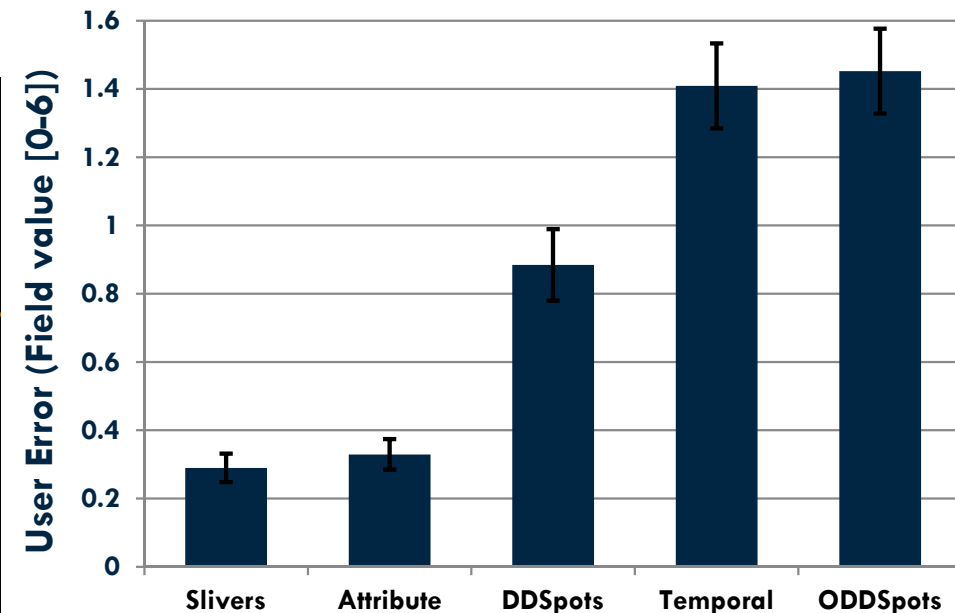
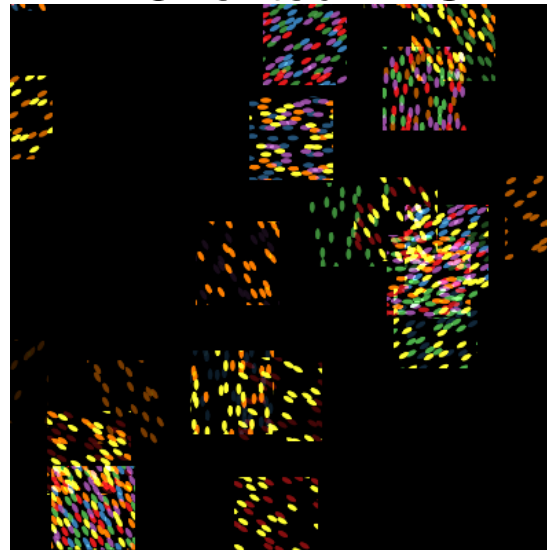
Data-driven spots



Temporal AB



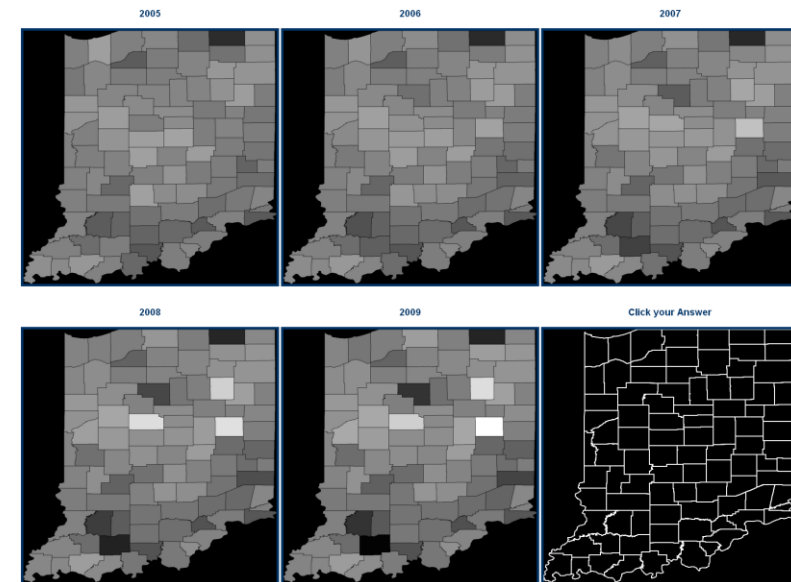
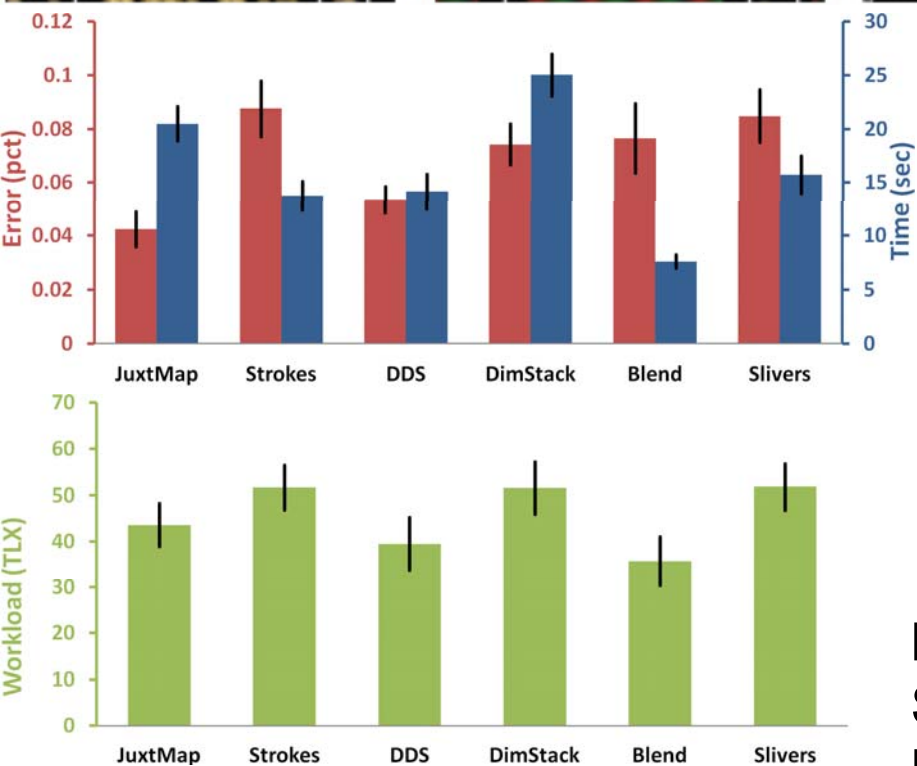
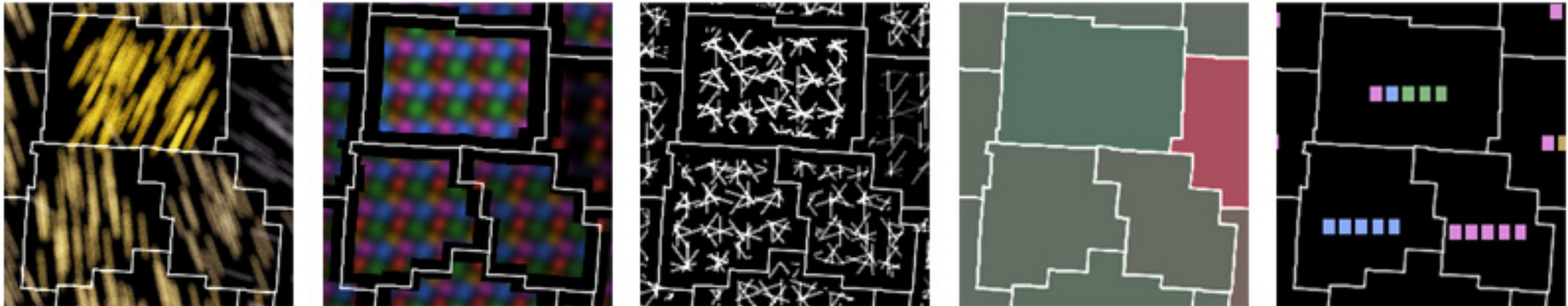
Oriented DDS





Trend Detection Study

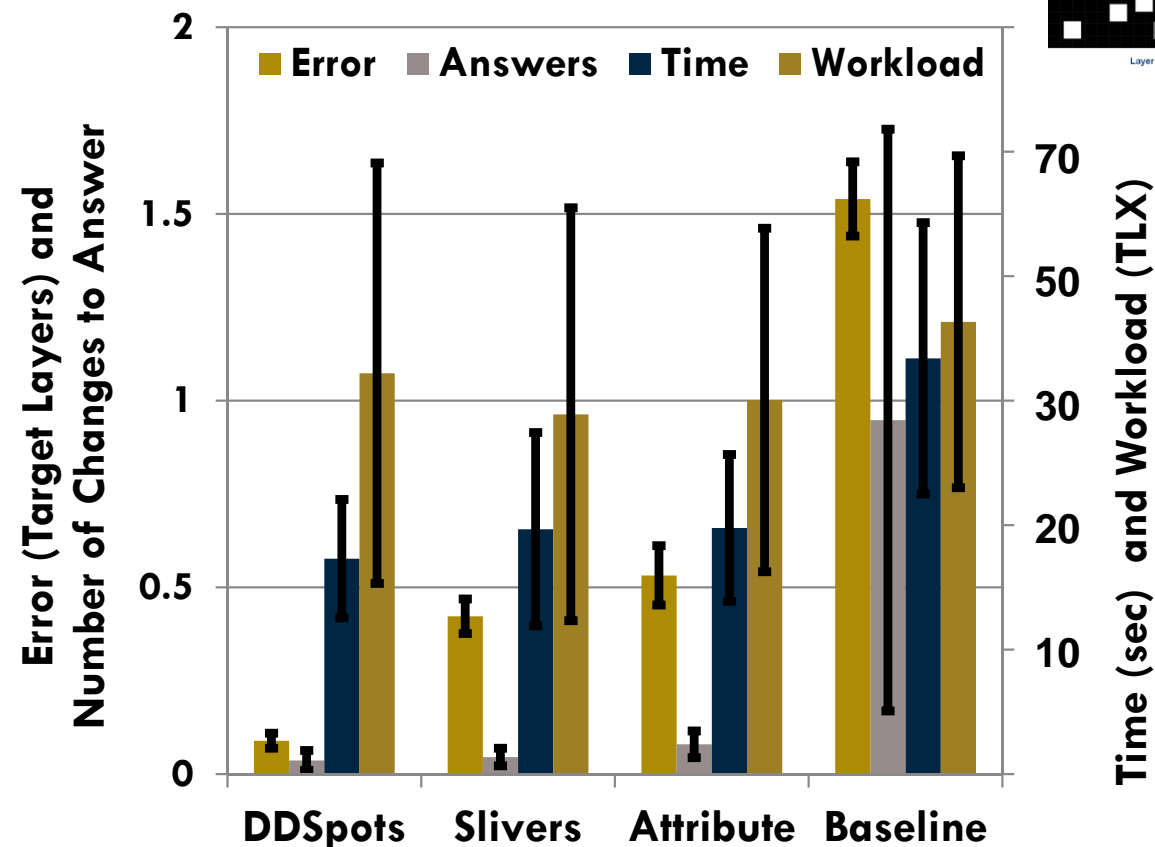
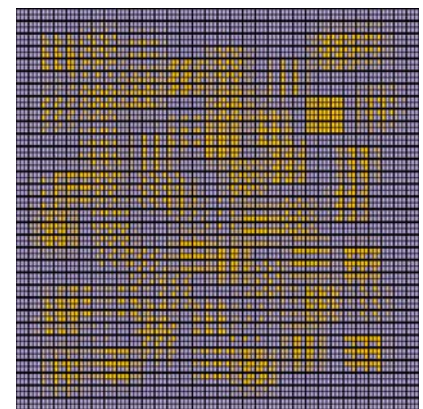
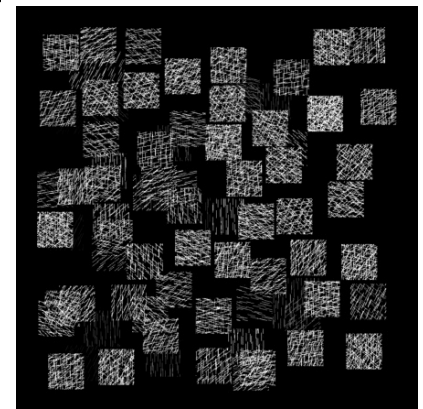
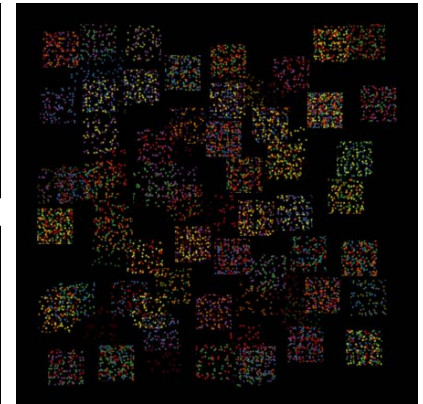
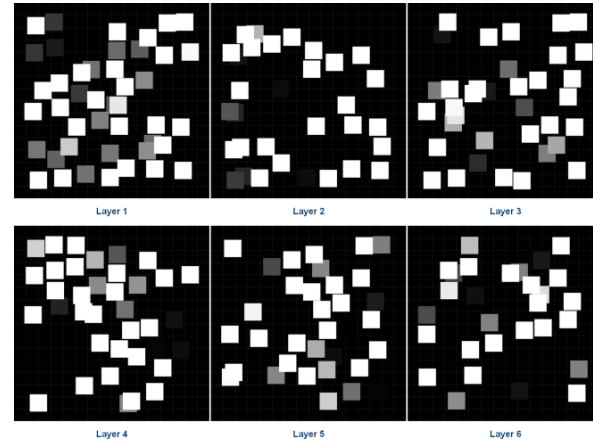
13



Baseline visualization had least error
Some techniques had ambiguous representations
Needed two variables to complete task

3. Provide strong grouping cues to facilitate chunking

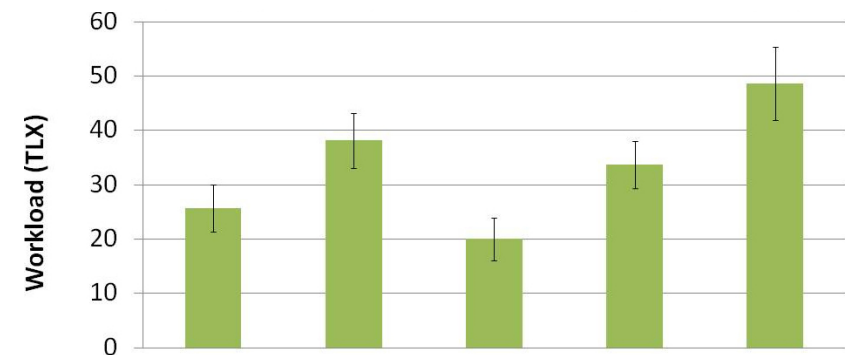
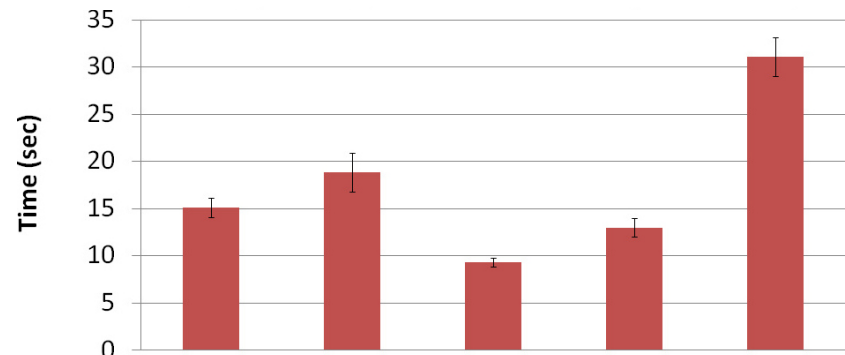
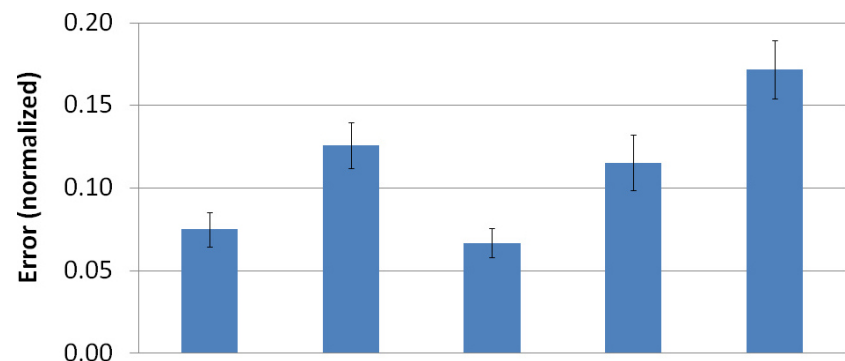
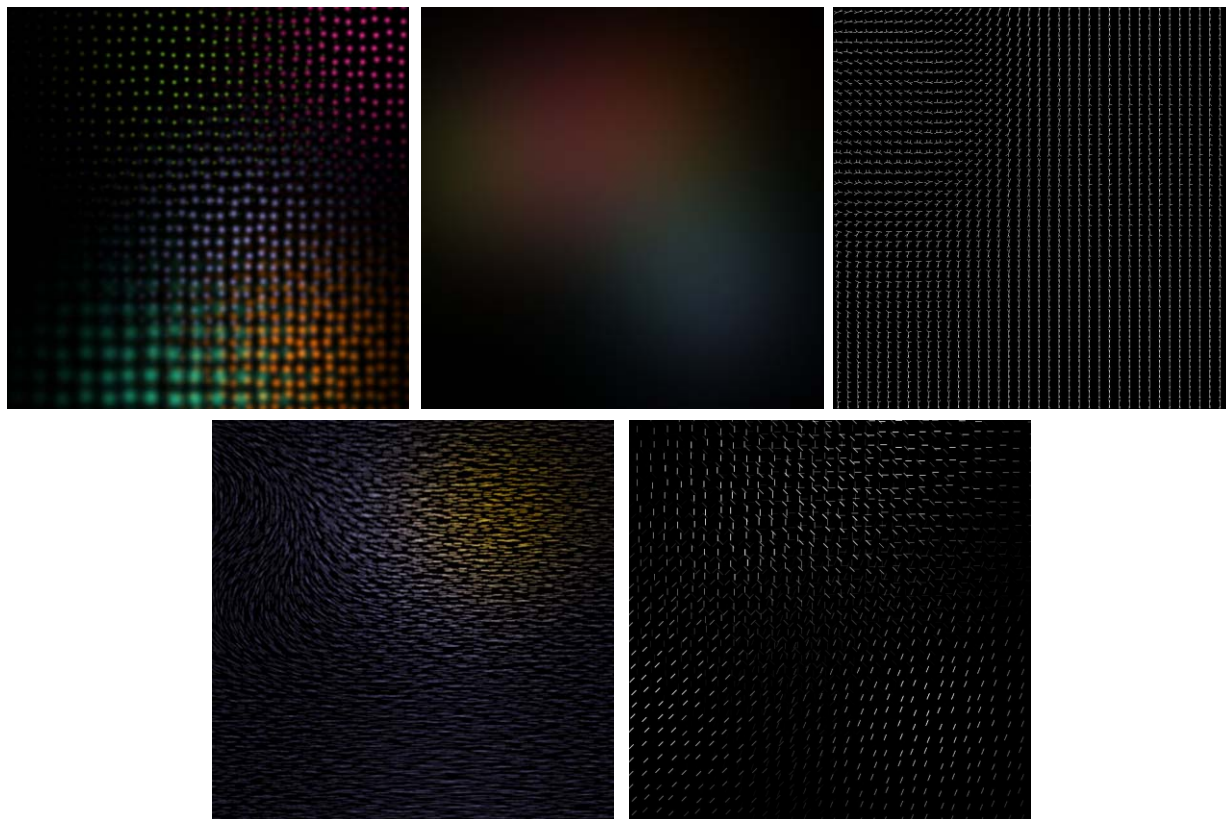
Reduce competition for working memory capacity by helping user to group pieces of information into larger “chunks.” Assist this by perceptual cues.





Critical Point Study Results

15



Oriented Slivers Brush Strokes Data-Driven Spots Color Blending Stick Figures

- 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

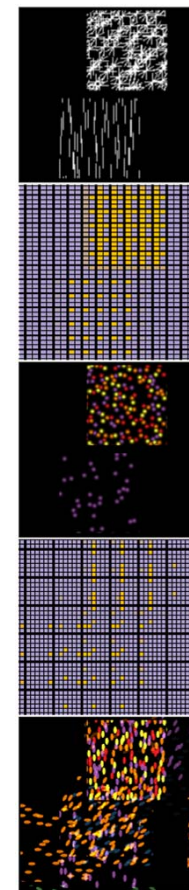
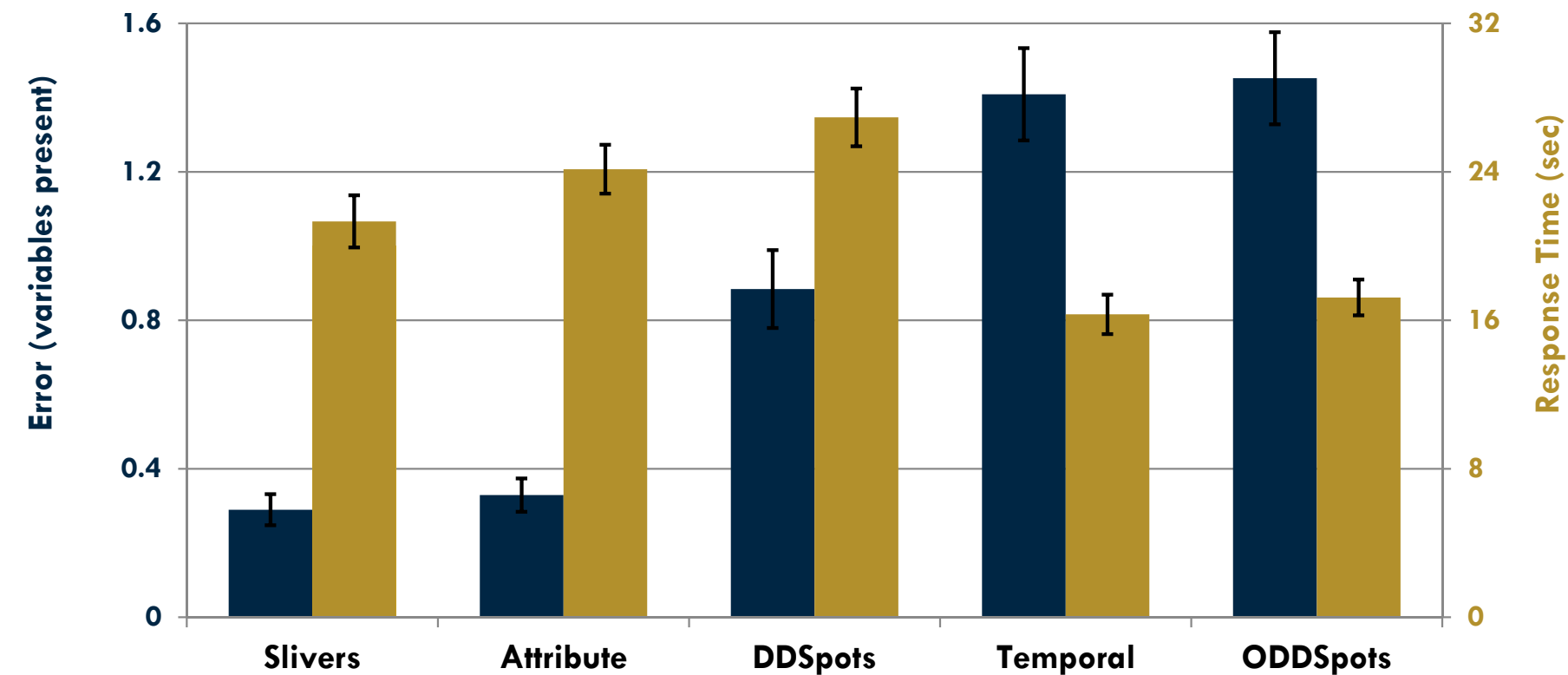
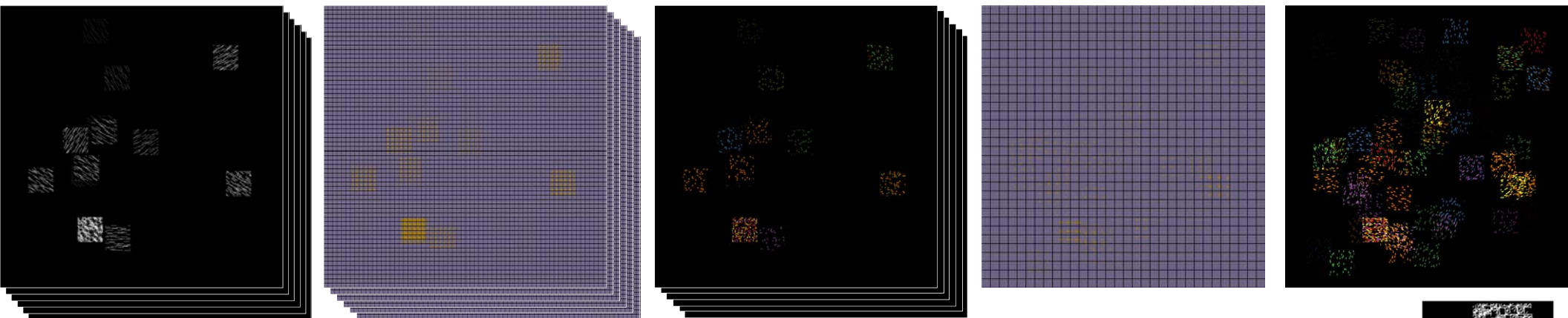
INFORMATION TECHNOLOGY DIVISION

INFORMATION MANAGEMENT AND DECISION ARCHITECTURES BRANCH



Overlap Detection Task

16

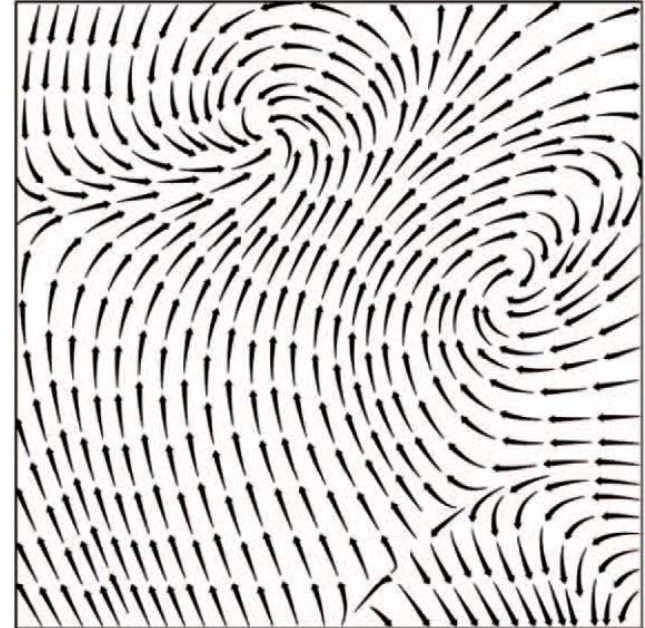


4. Organize information for knowledge structures

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.



Tasks in Multivariate Visualization User Studies

Identify critical point type
Advect particle

Localize critical values

Read sequence of data values

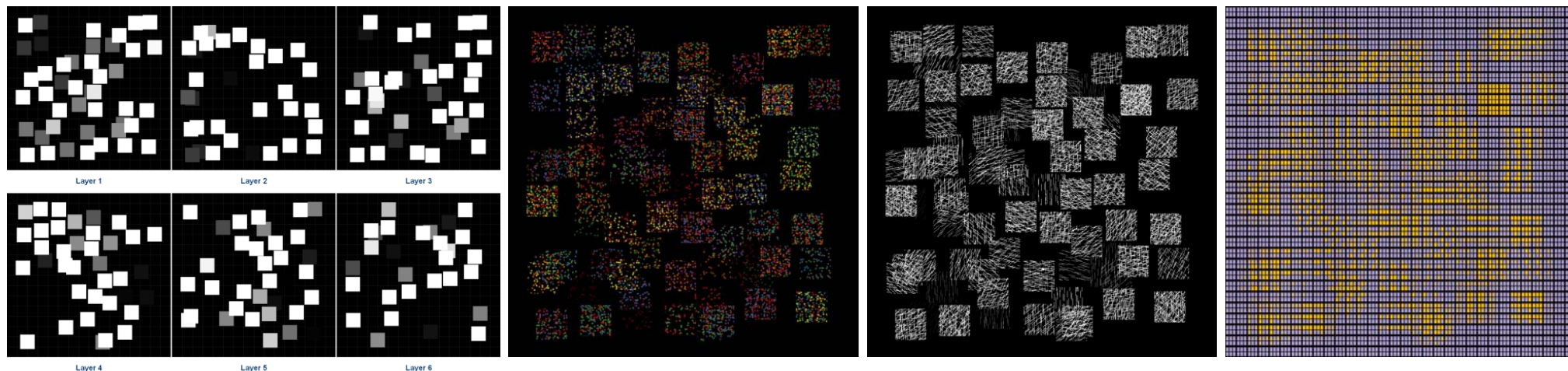
Detect greatest trends

Detect multi-way overlap

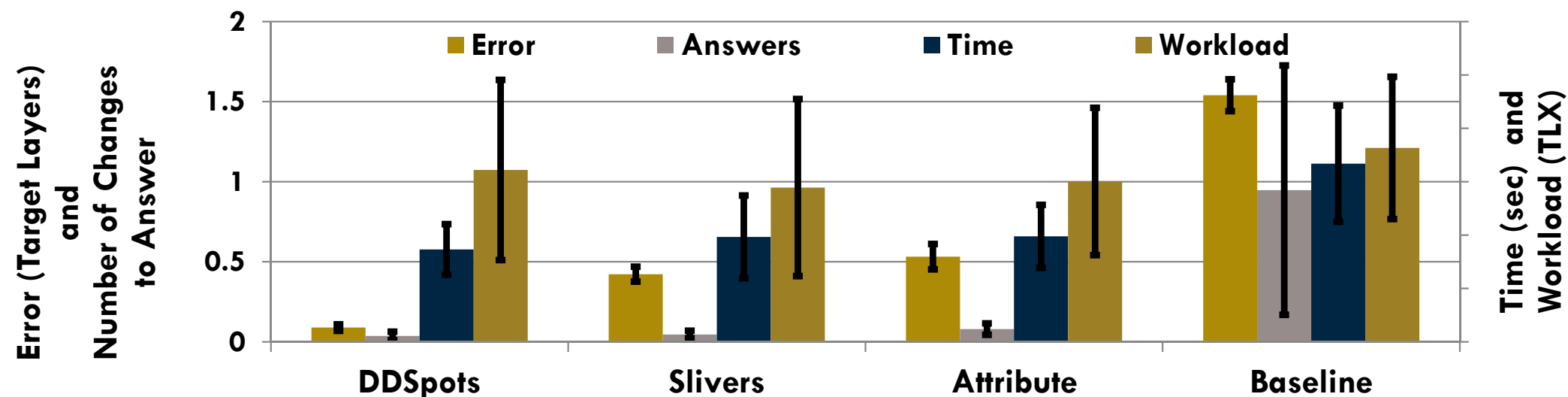


Overlap Detection Task

18



	Data-driven Spots	Oriented slivers	Attribute blocks	Juxtaposed layers
Error (layers)	0.09	0.42	0.47	1.54
Time (sec)	7.33	9.62	9.74	46.72

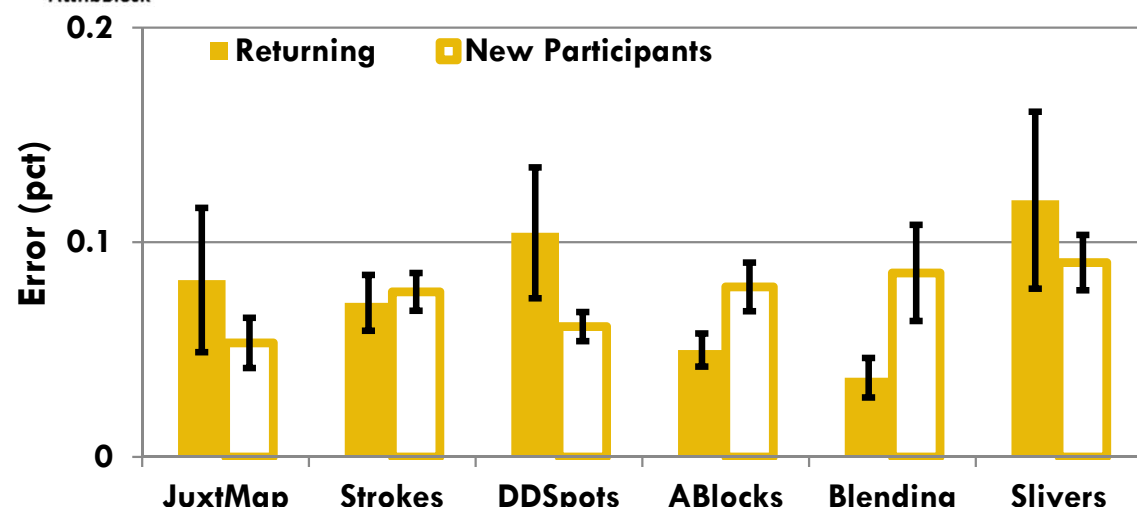
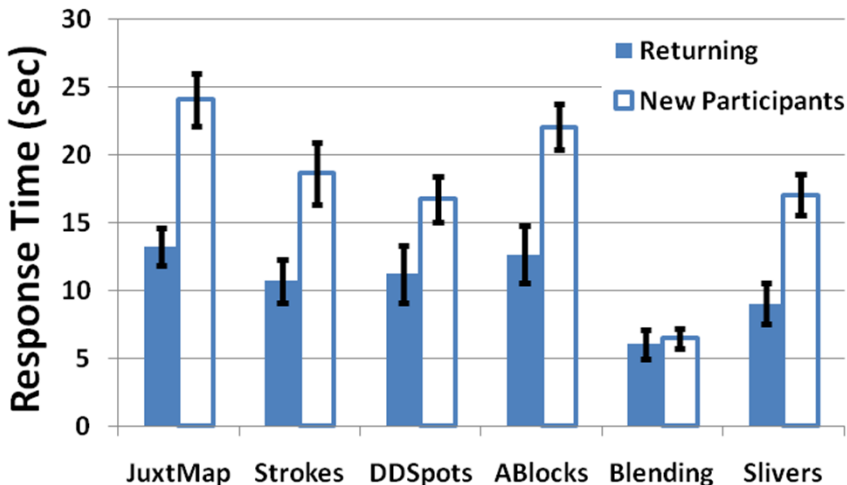
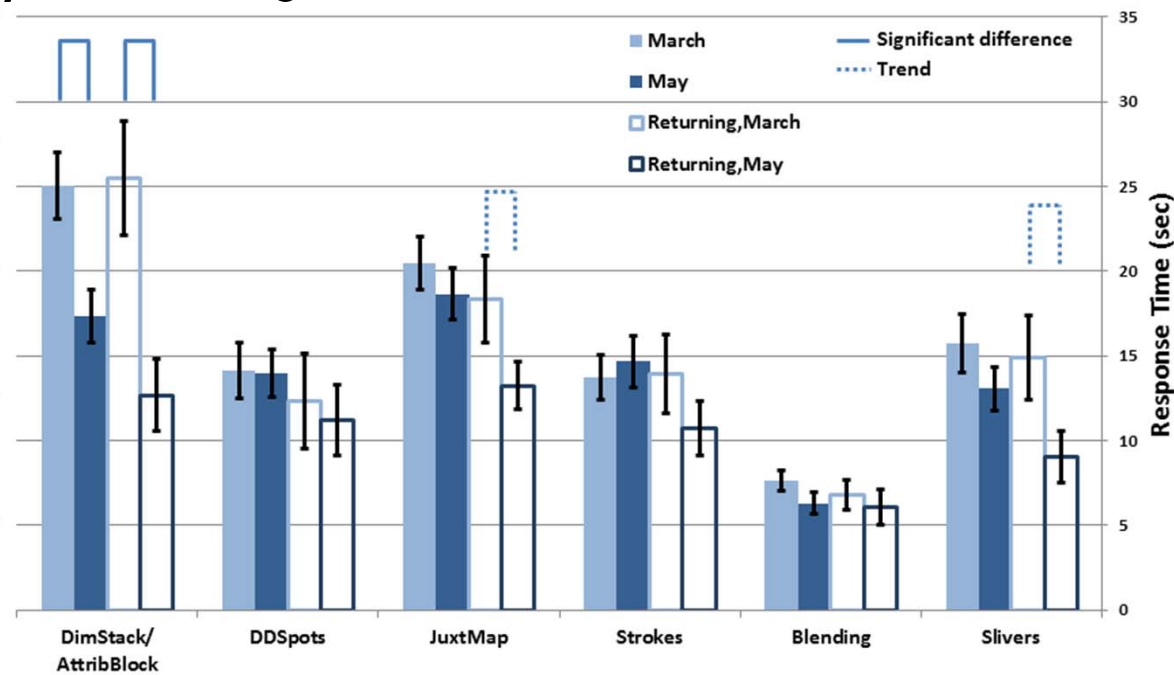
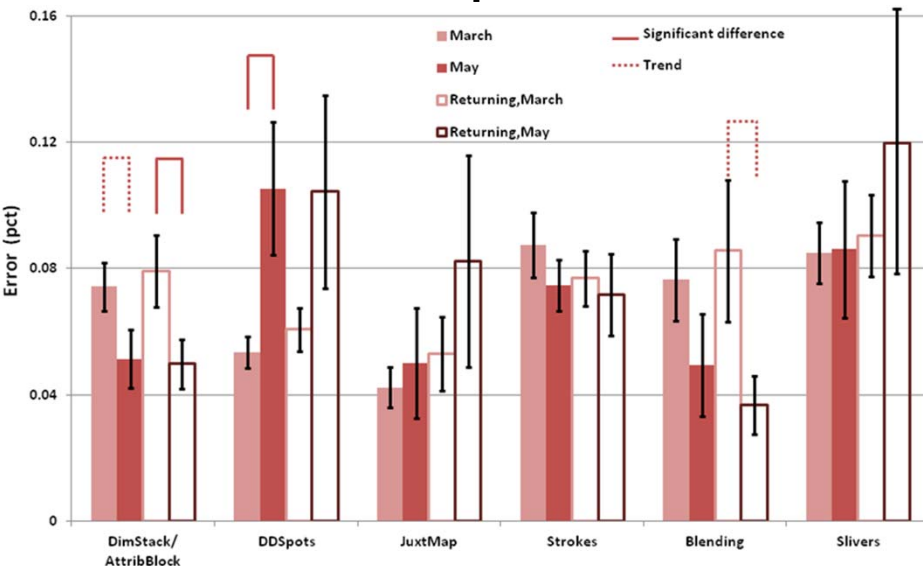


5. Structure information to provide strong retrieval cues

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

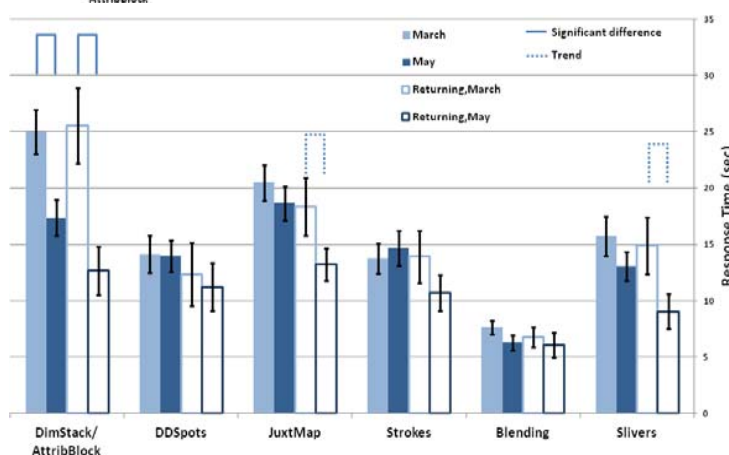
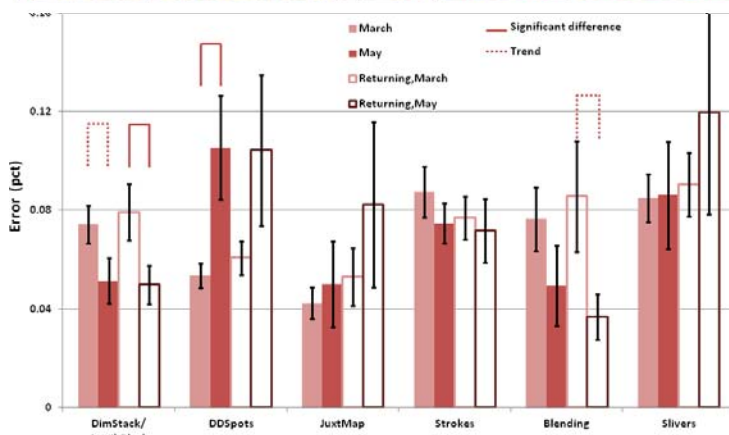
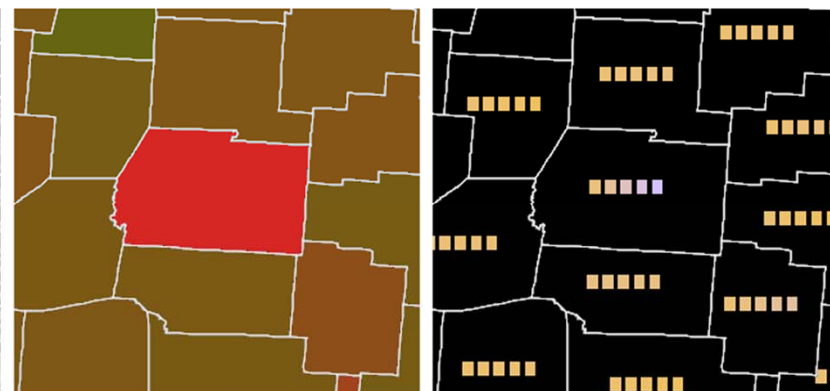
Statistical regularities may be learned implicitly; thus, visualizations could prime users to recognize certain data relationships through trained visual patterns that may be recognized.



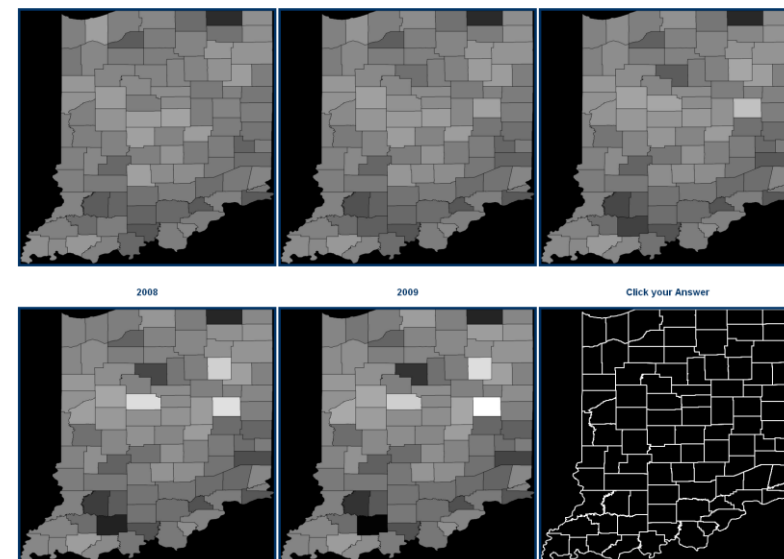


Trend Detection Follow-up Study

21



- 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

22

- 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



Thank you!

23

- Work supported by NRL Base Program
- Thanks to anonymous subjects and reviewers of papers

