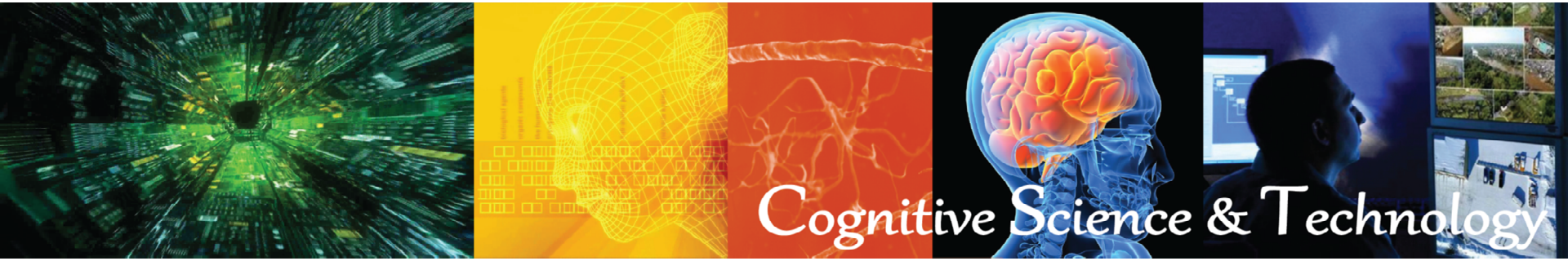
SAND2017-8824C



Effects of Task on Eye Movements During Comprehension of Abstract Data Visualizations

Laura Matzen, Kristin Divis & Michael Haass

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

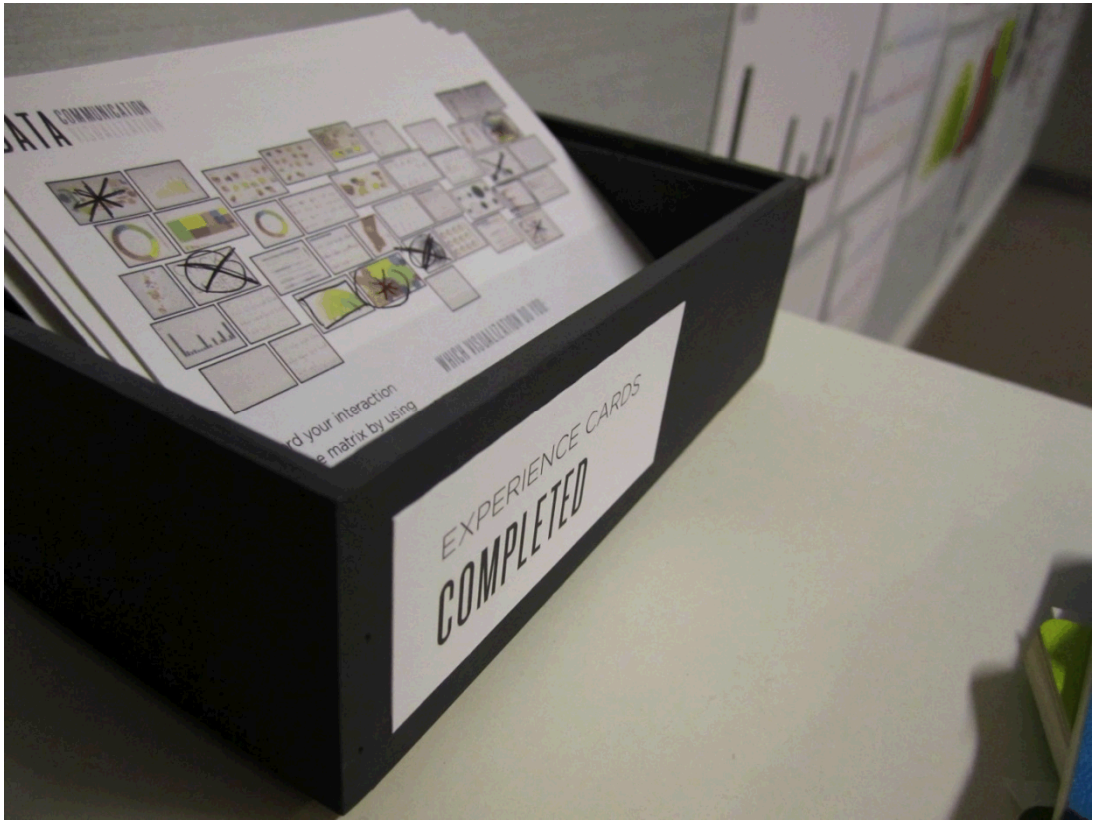
Background

- Analysts often rely on data visualizations when making high-consequence decisions, but little is known about how to evaluate a visualization's effectiveness for an end user

- The field of visual analytics is calling for the creation of models of human cognitive processing that can address this gap and advance our understanding of how humans reason about data visualizations.

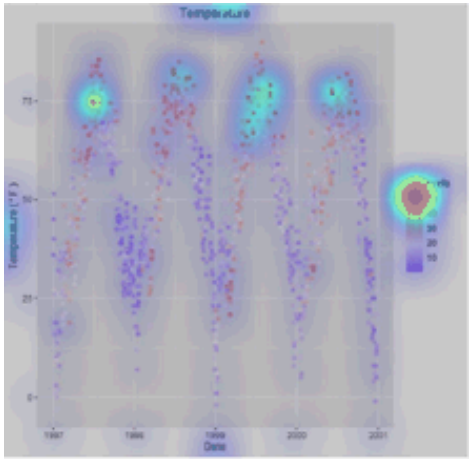
State of the art…

Lauren Manning, http://www.visualizing.org/stories/visualizing-food-40-ways

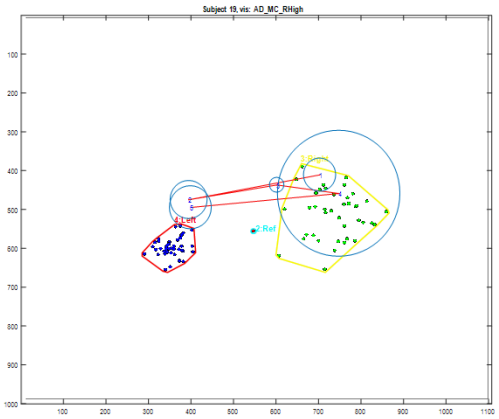


Project Goals

- **Develop models for assessing the bottom-up visual saliency of data visualizations.**
  - Ideally, a vis will draw the viewers' attention to the most important information for their task



- **Conduct experiments to characterize common top-down sensemaking strategies employed by users of visualizations.**
  - Studies using eye tracking to investigate how analysts navigate through abstract information
  - Expansion of the "Value of Vis" framework (Stasko, Georgia Tech)

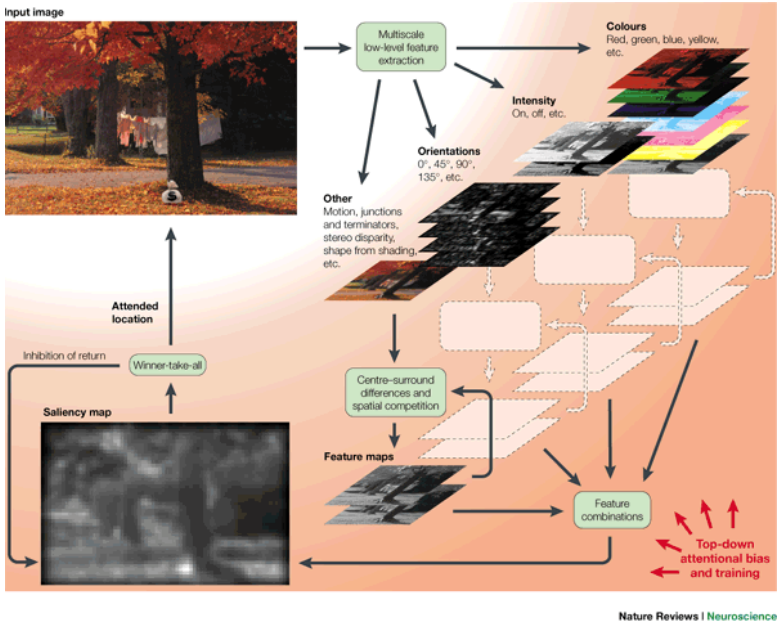


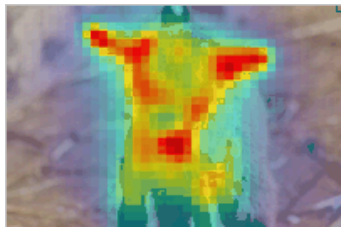
Models for assessing the bottom-up visual saliency of data visualizations

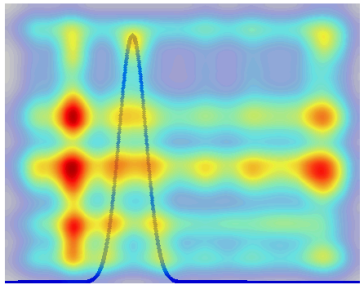
- Models of bottom-up saliency exist for natural scenes and can predict where people will look
  - A tool that could do the same for visualizations would be extremely useful for evaluations!

- These models fail for abstract visualizations
  - Inappropriate spatial scales and weighting
    - Visualizations have features that are very small relative to the extent of the image
  - Inadequate feature sets
    - Features used for natural scenes (orientation, intensity, color) don't capture key contrasts used in data visualizations
    - RGB color space does not correspond well to human color perception

(Itti & Koch, 2001)

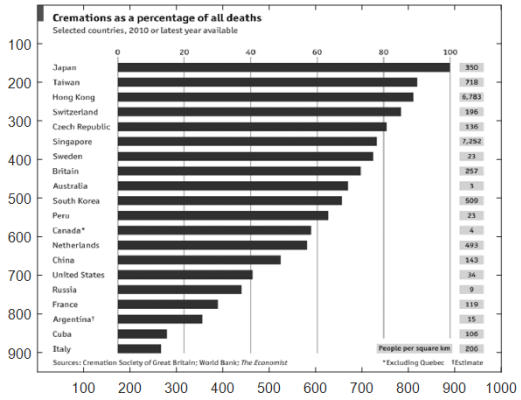




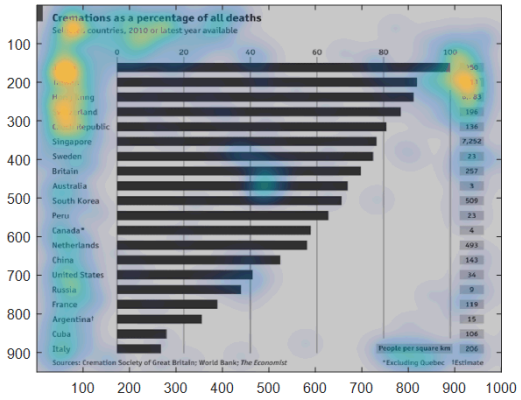


Sandia National Laboratories

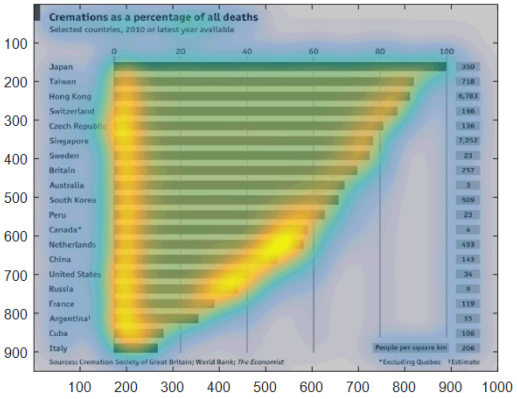
Economist Daily Visualization



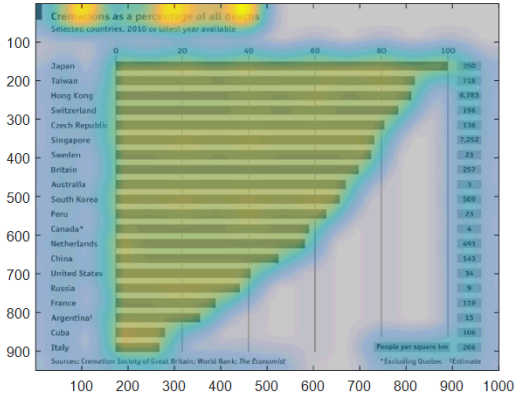
Human Subjects Fixation Map



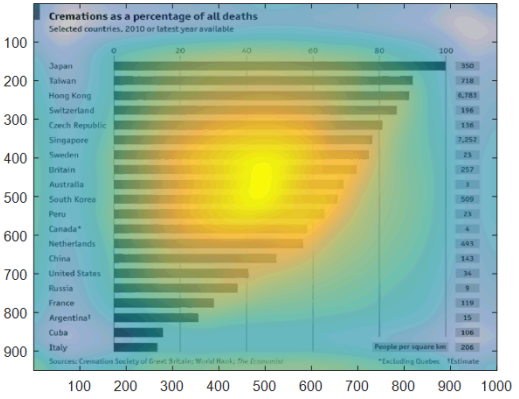
Itti & Koch



BMS



eDN



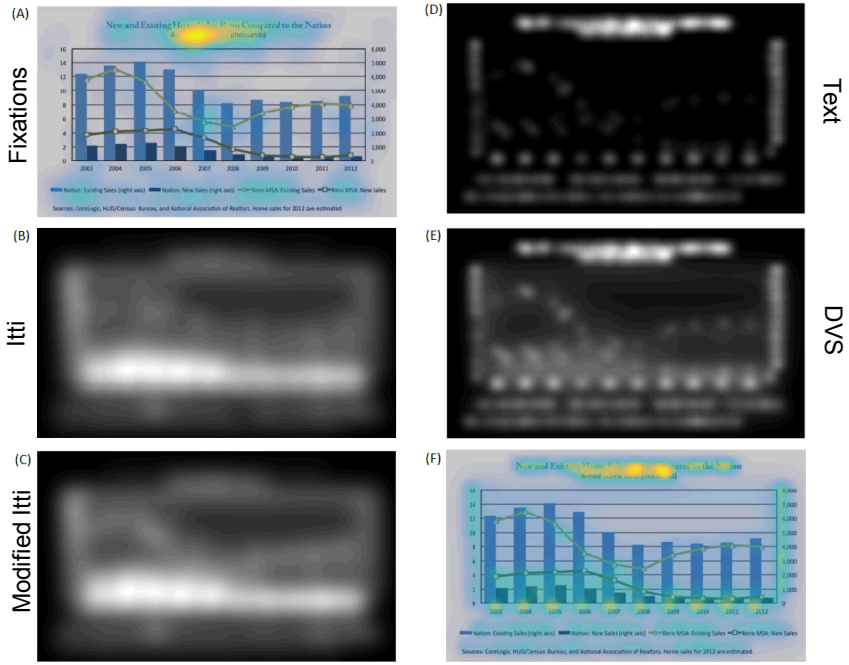
6

Data Visualization Saliency (DVS) Model

- Use existing model (Itti and Koch, 2001) as basis
  - Inspired by V1
  - Uses 3 feature channels: color, intensity, and orientation
  - Performed best of current models on data visualizations (but still not great)

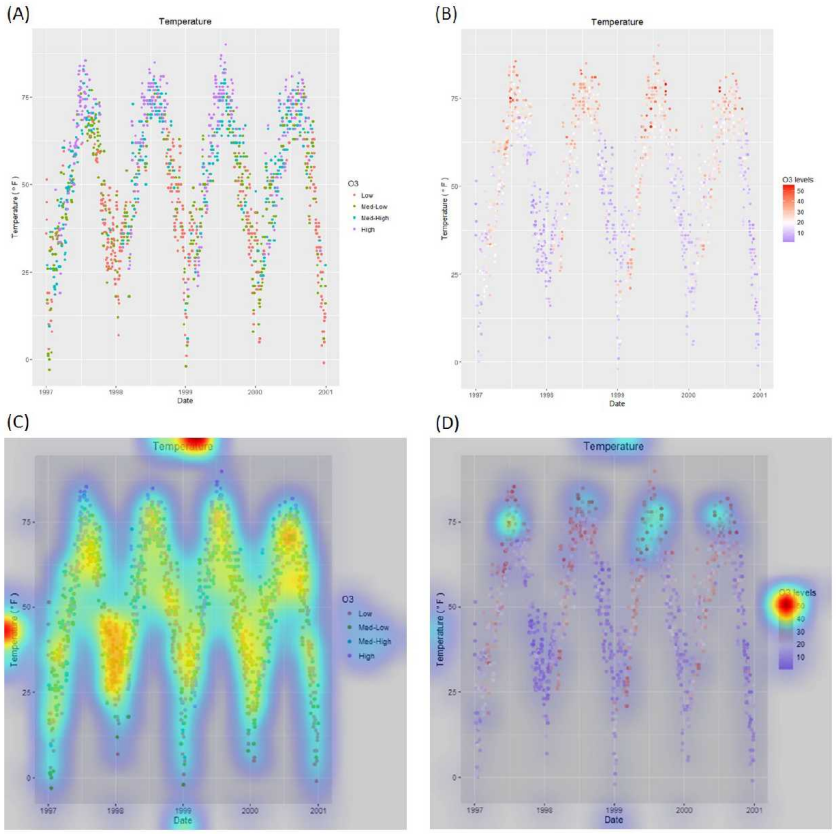
- DVS: Improve for data visualizations
  - Update from RGB to CIE LAB color space
  - Add text detector feature

- DVS model provides a significantly better match to human fixation data than other saliency models



Example of applying DVS

- Same data plotted two ways
  - Default ggplot2
  - Diverging color scheme
- Which to choose?
  - Want all points to be about equally salient?
  - Want user to focus on extreme upper points?

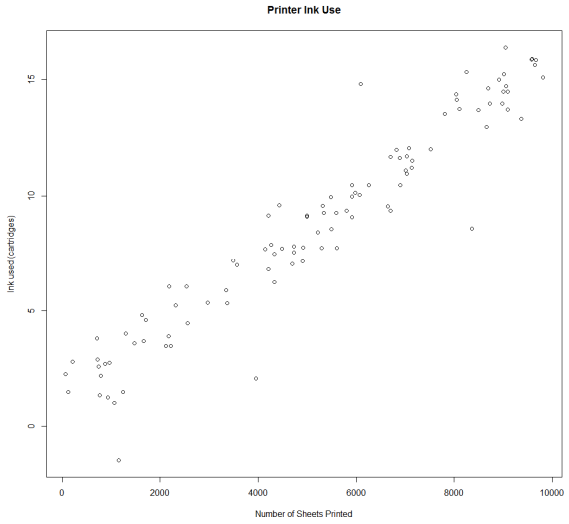


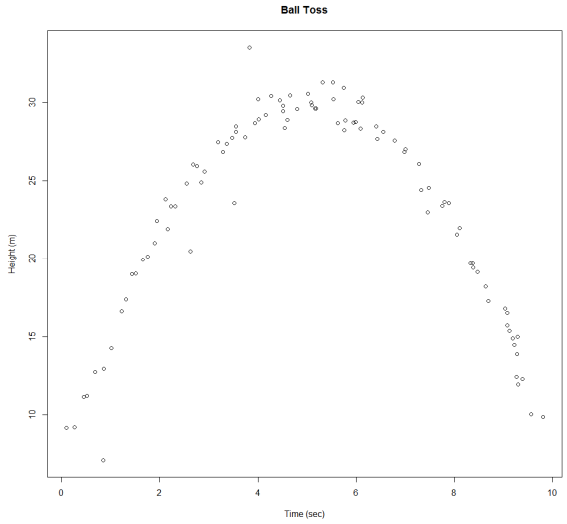
Next Steps

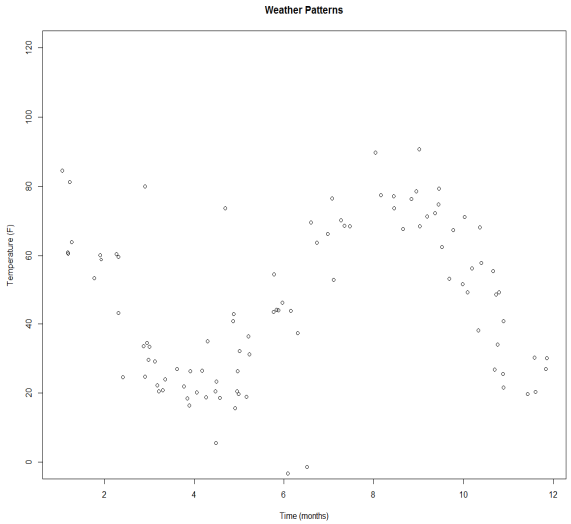
- Incorporate more top-down characteristics
  - Including text was **very** helpful
  - How about other characteristics?
- Identify important characteristics through human subjects research
  - How do tasks (goals) influence attention?
  - How does attention change for different types of data visualization?
- Approach
  - Start with relatively simple, controlled tasks and work our way toward more complex applications
    - Trend vs. Outlier
    - Cluster Membership vs. Cluster Characteristics

Trend vs. Outlier

- 30 Participants from the University of Illinois
- Stimuli: 32 scatterplots with different data patterns
  - Linear, cyclical, asymptotic, flat, quadratic
  - 2 or 4 outliers per plot
- Task: Participants asked to **describe the trend** or to **describe the outliers**
  - Counterbalanced across participants (15 participants did each task for each vis)







Trend vs. Outlier

- Which trends were participants better at describing?
  - Two independent raters scored responses to trend description task (inter-rater reliability >.95)

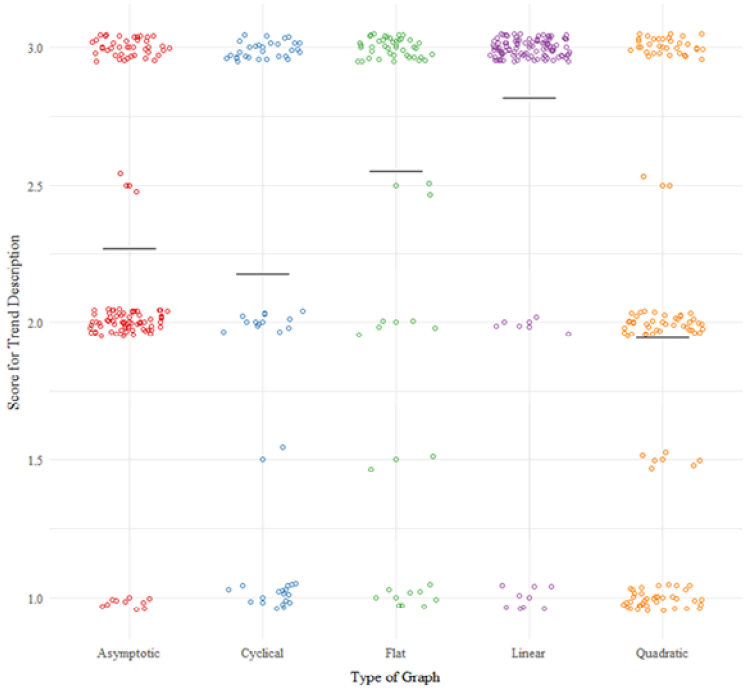


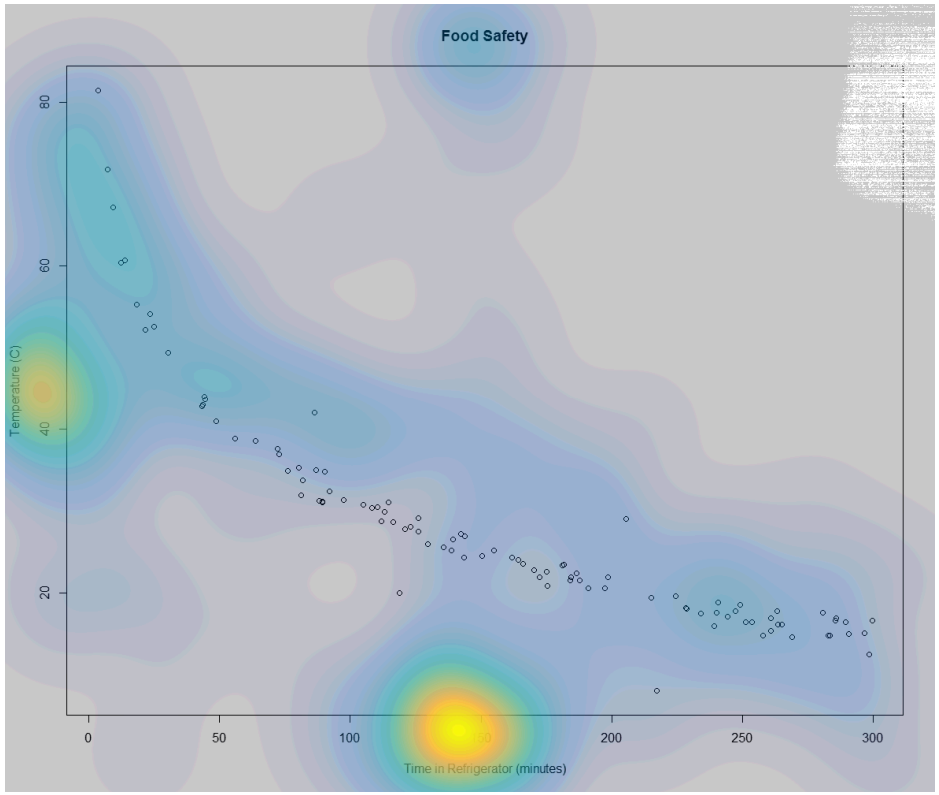
Figure XX. Average score in the trend description task as a function of type of graph. Points are jittered to avoid overplotting. The black line represents the mean for each type of graph.

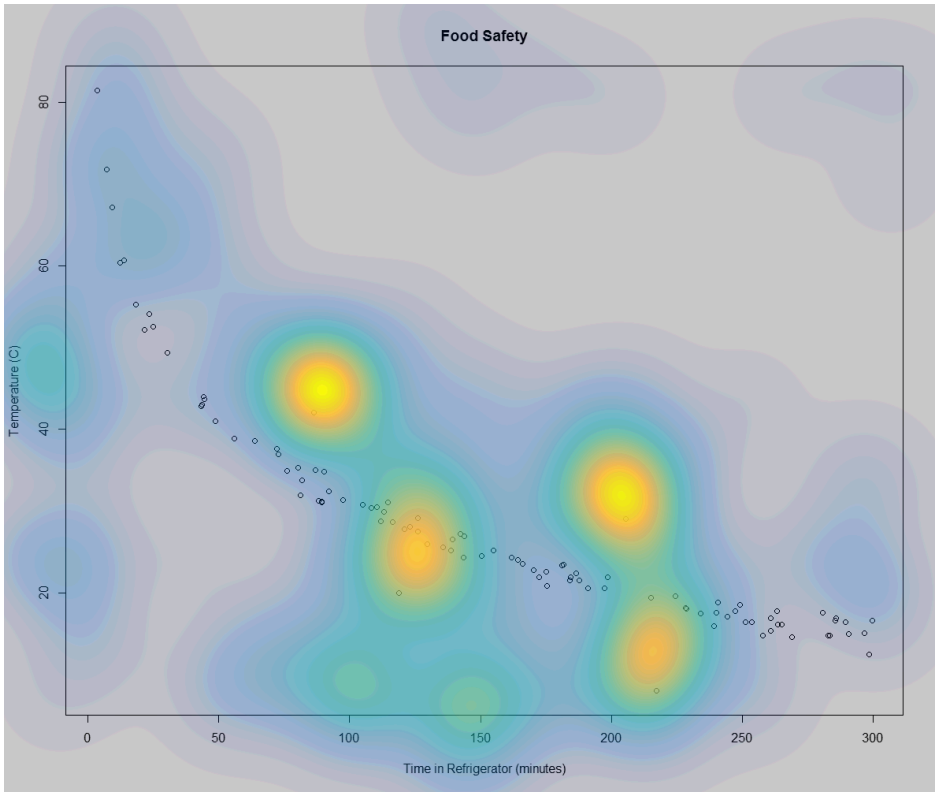
Trend vs. Outlier

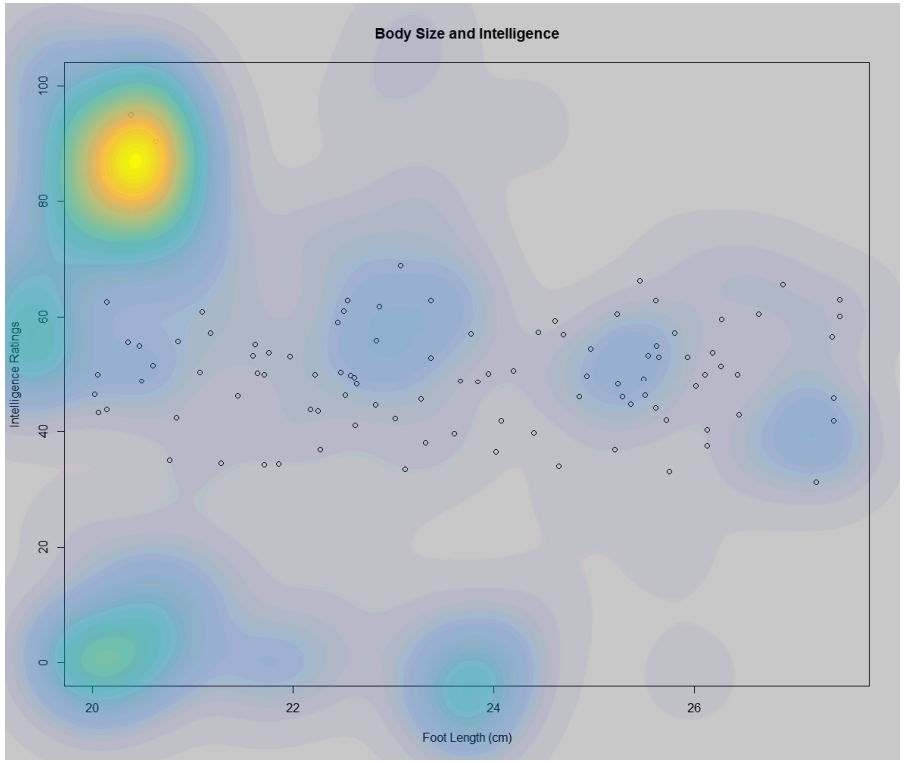
- How were they at outlier detection?
  - 2 or 4 outliers in every scatterplot
  - Two independent raters scored responses to outlier detection task (inter-rater reliability >.95)
  - Errors made on **55.3%** of the trials
    - More likely to **miss** an outlier (90.9%) than falsely identify an outlier

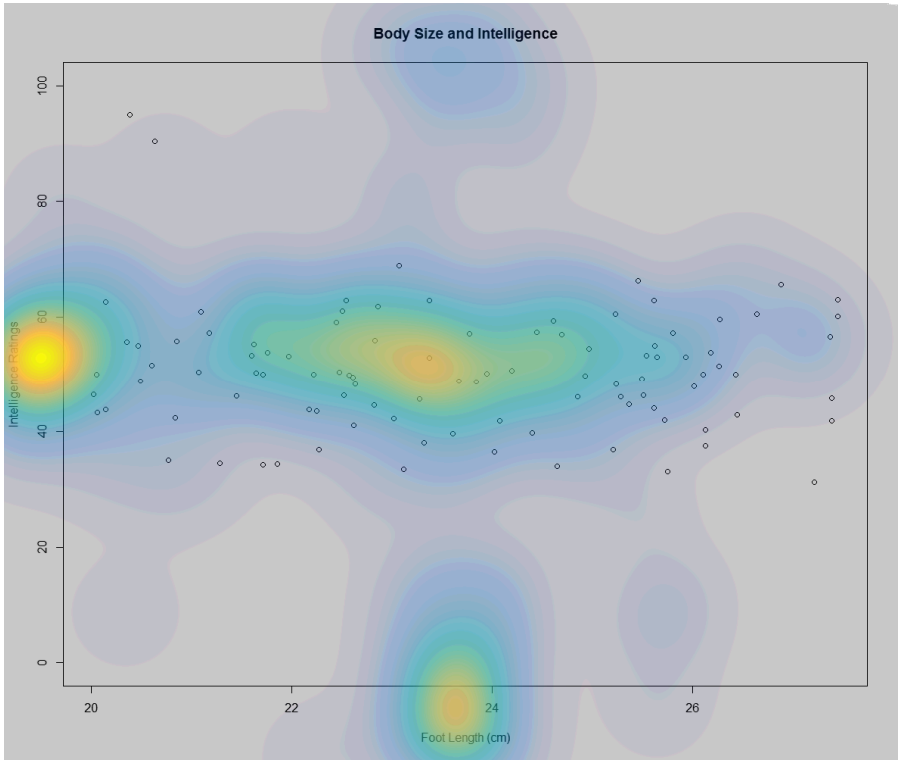
Figure YY. Absolute value of the error in outlier counts as a function of type of graph. Points are jittered to avoid overplotting. The black line represents the mean for each type of graph.

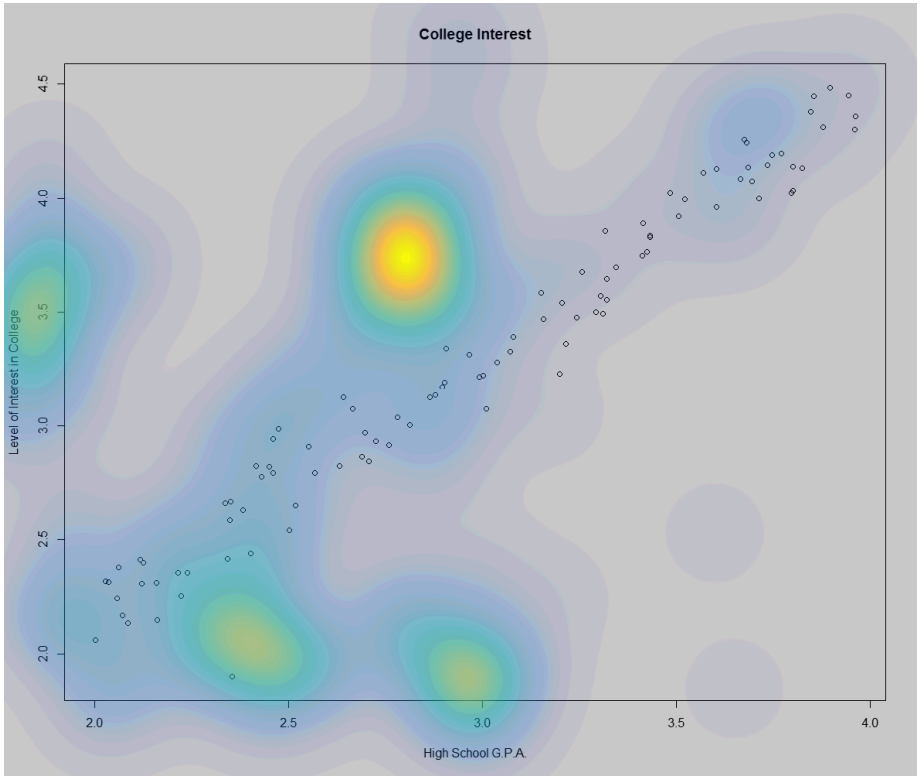
Trend vs. Outlier

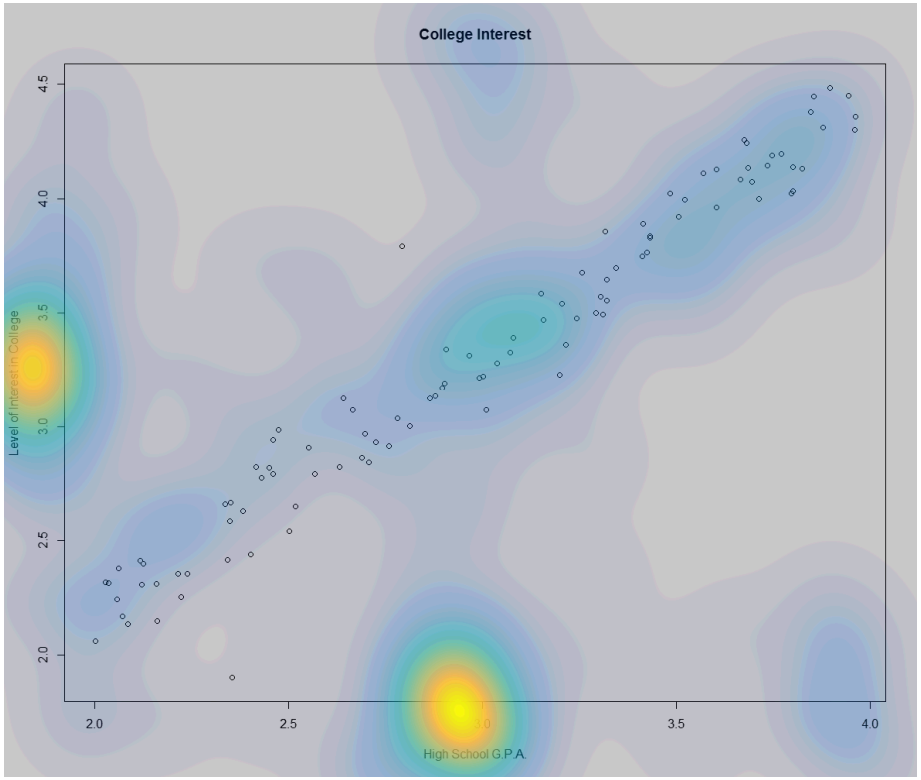


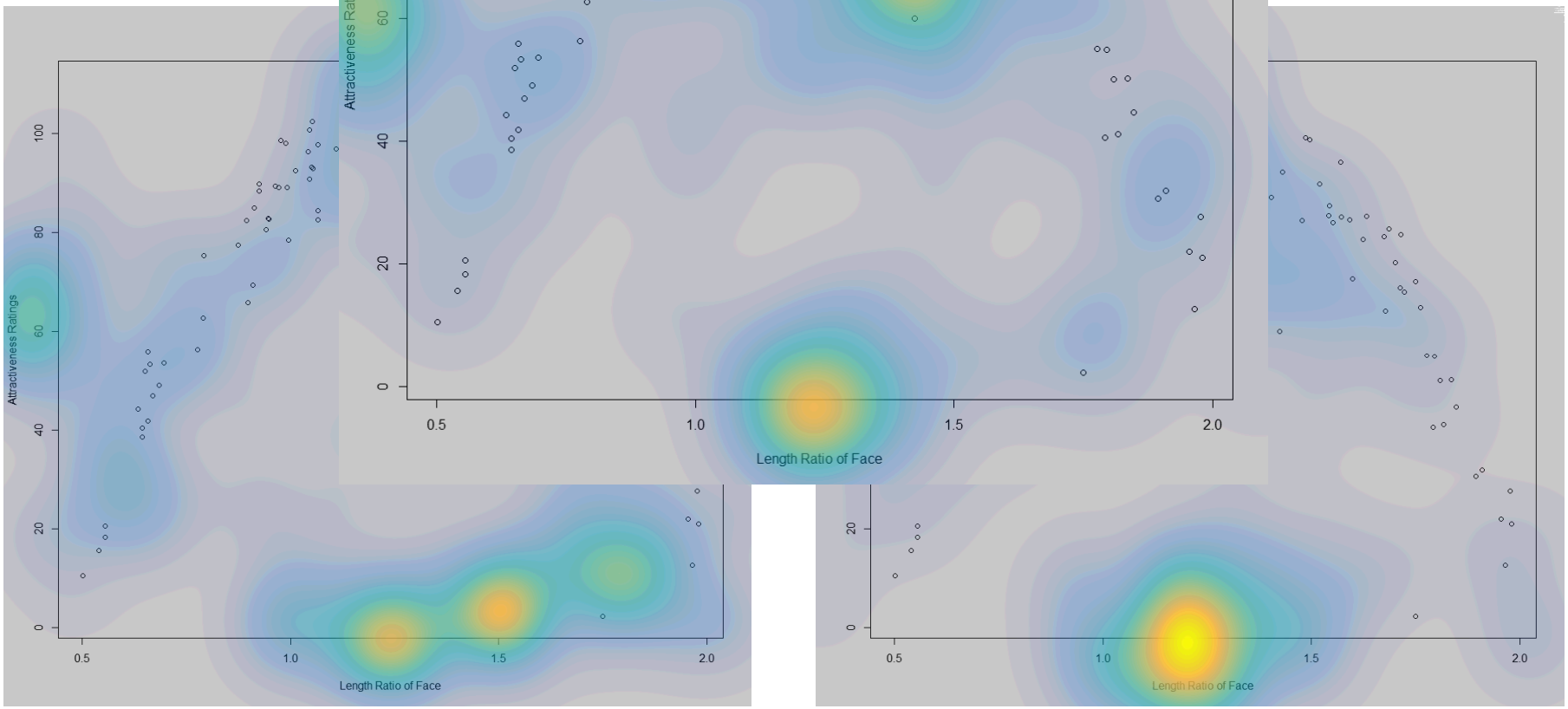



Body Size and Intelligence
Intelligence Ratings
0
20
40
60
80
100
20
22
24
26
Foot Length (cm)


Body Size and Intelligence
Intelligence Ratings
0
20
40
60
80
100
20
22
24
26
Foot Length (cm)


College Interest
Level of Interest in College
High School G.P.A.
2.0
2.5
3.0
3.5
4.0
4.5


College Interest
Level of Interest in College
High School G.P.A.
2.0
2.5
3.0
3.5
4.0
4.5


Beauty Metrics
Attractiveness Ratings
Length Ratio of Face
0
20
40
60
80
100
0.5
1.0
1.5
2.0

Trend vs. Outlier

- Preliminary behavioral results
  - Relatively good at describing trends for +/- or flat linear trends
  - Likely to miss outliers
  - More complex patterns lead to poorer performance

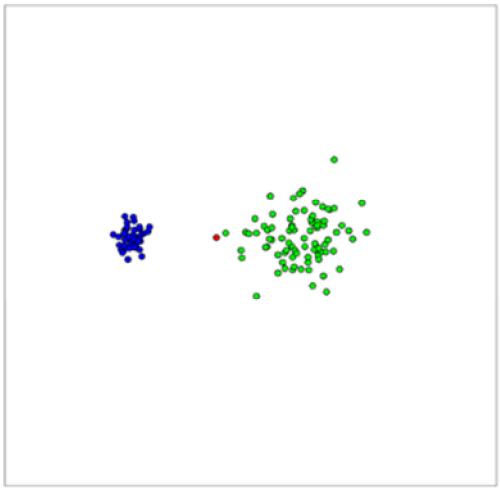
- Eye movement data
    - Significant differences in patterns of attention depending on task
      - In the Outlier Task, participants had significantly more fixations to the outliers
        - » Fixations to the outliers were also significantly longer
      - In the Trend Task, participants had significantly more fixations to the trend, title, Y-axis, and axis labels
    - Pattern of fixations also differs by task
      - Outliers viewed earlier in Outlier task, Trend viewed earlier in Trend Task

Cluster Perception

- Question: how do people interact with and understand slightly more complex data visualizations (e.g., clusters in two dimensional plane)

- Design
  - Monitor **eye movements** while viewing two clusters in two dimensional plane with counterbalanced properties
  - Ask participants to indicate to which cluster **reference point** belonged or which cluster had **highest mean**
  - 80 images; 30 participants

- Stimuli properties counterbalanced



Exp 2: Block B Design

- Stimuli properties counterbalanced
  - Cluster properties
    - Sparsity (low or high)
      - How many points
    - Dispersion (low or high)
      - How spread out the points are
  - Reference point: centered based on (horizontal) **mean** or **standard deviation** of two clusters
  - Cluster height: even or raised
  - Correct answer: left or right
  - Color: blue or green (randomized)

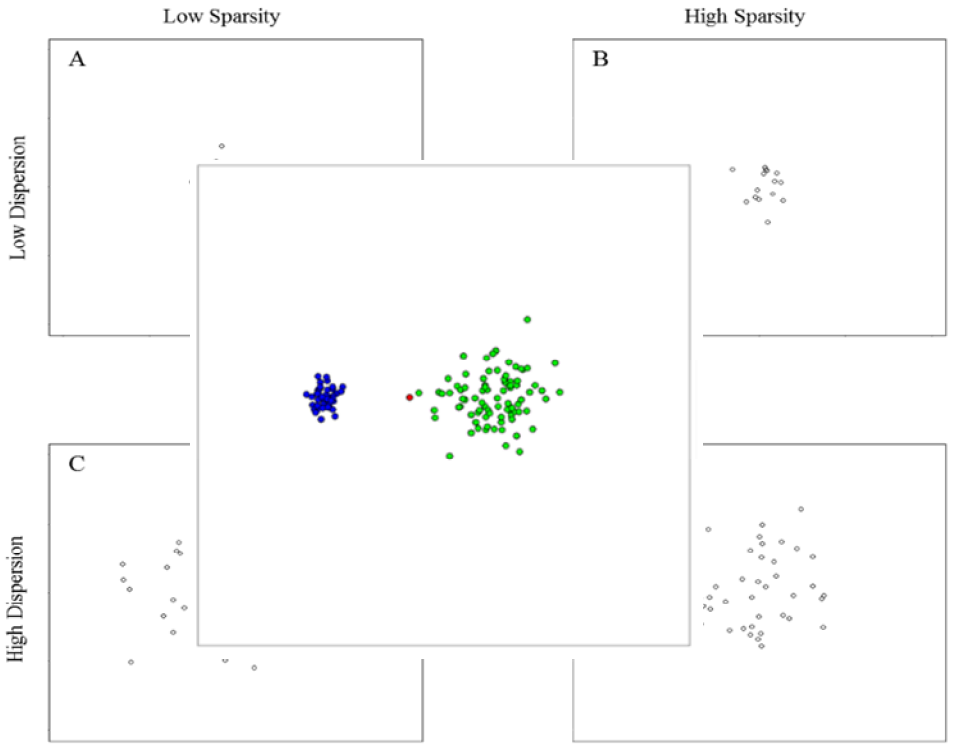


Figure B1. Examples of cluster types: (A) low sparsity, low dispersion, (B) high sparsity, low dispersion, (C) low sparsity, high dispersion, and (D) high sparsity, high dispersion.

Exp 2: Block B Preliminary Results (Reference Point Task)

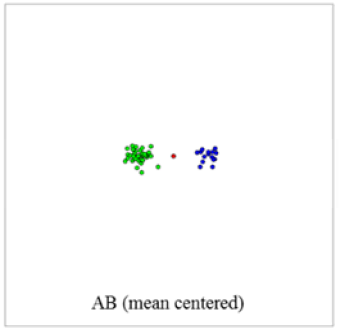
Reference point task: which cluster does the reference point belong to?

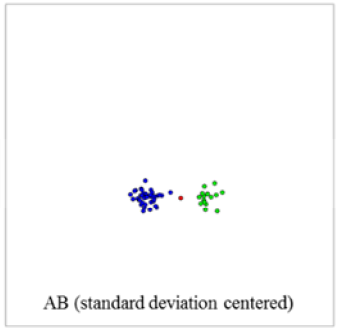
Exp 2: Block B Preliminary Results (Reference Point Task)

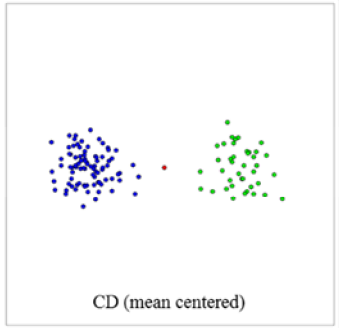
- Reference point task: centering method was important when dispersions **differed**

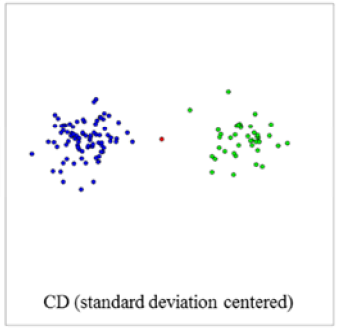
Dispersion \ Sparsity	Low	High
Low	A	B
High	C	D

Little influence when dispersions similar:



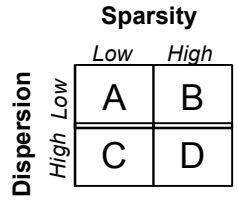






Exp 2: Block B Preliminary Results (Reference Point Task)

…but big influence for clusters with dissimilar dispersions!



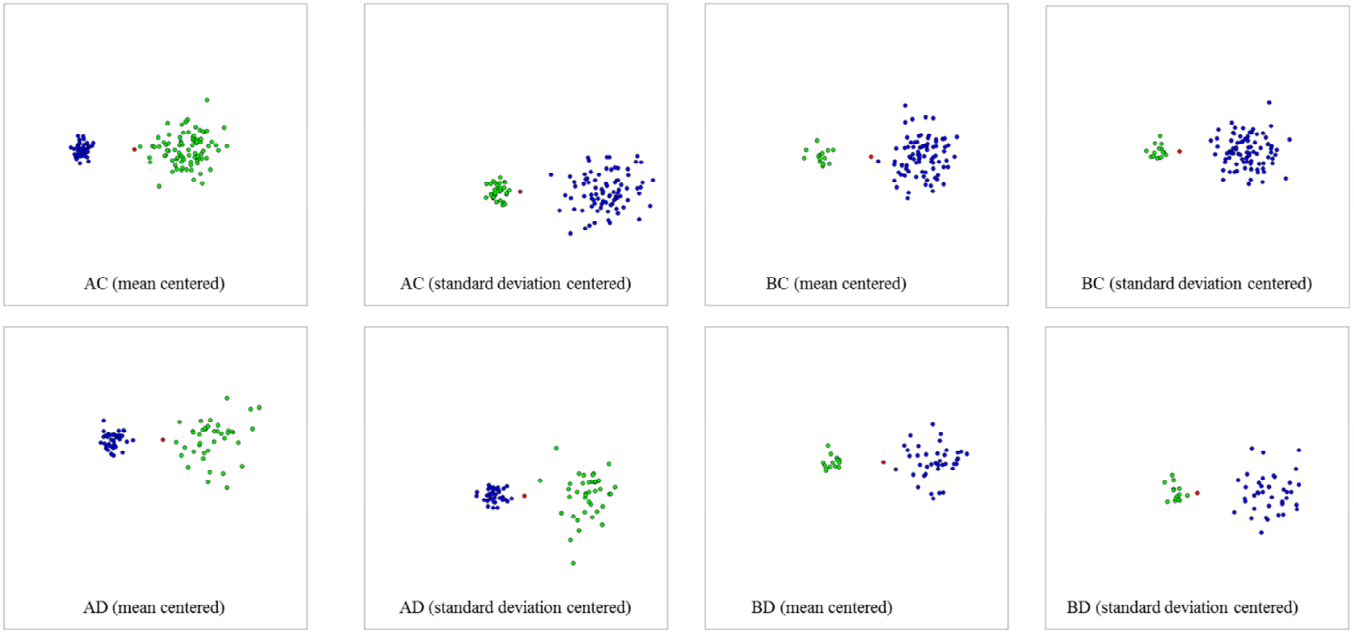
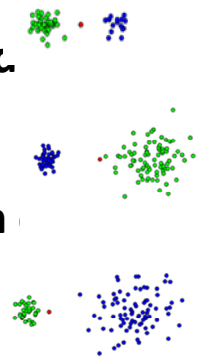


Figure B3. Examples of cluster pairing by type for stimuli when values are drawn from normal distributions with different standard deviations.

Exp 2: Block B Preliminary Results (Reference Point Task)

- Reference point task preliminary results:
  - Reference point was perfectly centered between the two clusters
  - Yet participants tended to indicate that the reference point most often belonged to the cluster with:
    - **Lower sparsity** (more data points)
      - 78.8% of the time (*CI* [73.0%, 83.7%], $p < .001$)
      - Regardless of dispersity or centering method
    - **Higher dispersion** when the reference point was **mean centered**
      - 91.7% of the time (*CI* [85.2%, 95.9%], $p < .001$)
      - Regardless of sparsity
    - **Lower dispersion** when reference point was **standard deviation**
      - 87.5% of the time (*CI* [80.2%, 92.8%], $p < .001$)
      - Regardless of sparsity
  - Take away: dispersion + centering method drove decision about reference point membership (sparsity was secondary)



Exp 2: Block B Preliminary Results (Reference Point Task)

- …but what else might be going on?

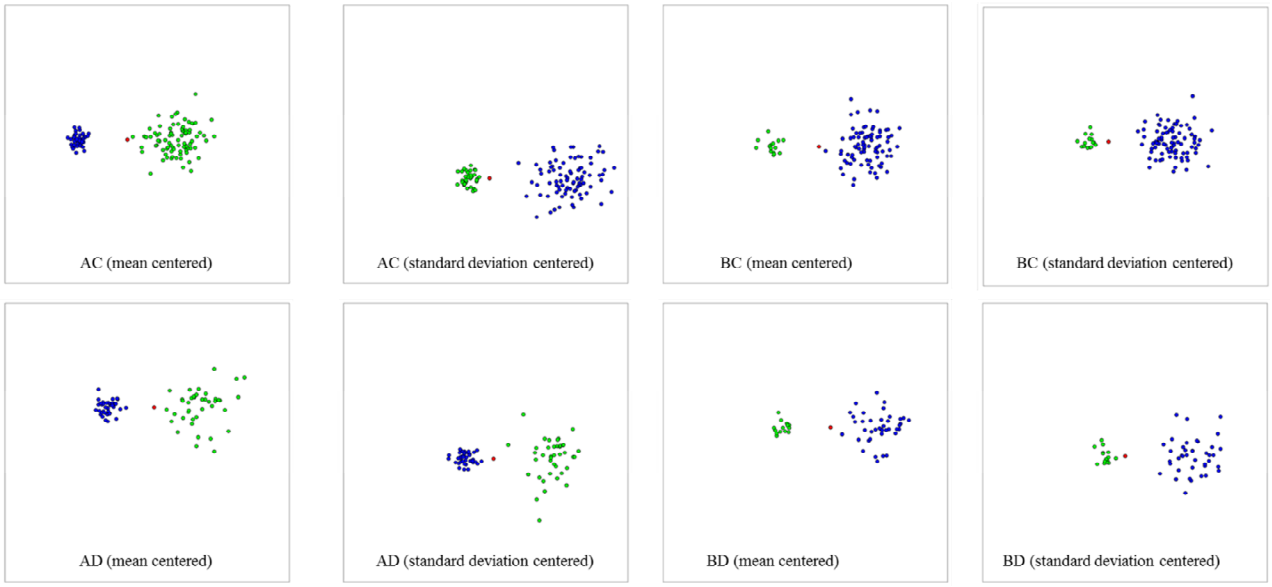


Figure B3. Examples of cluster pairing by type for stimuli when values are drawn from normal distributions with different standard deviations.

Just relying on nearest neighbor?

Exp 2: Block B Preliminary Results (Reference Point Task)

- Nearest neighbor?
  - Chose cluster with nearest neighbor 91.3% of the time
  - Nearest neighbor is highly correlated with dispersion/centering
  - Future work needed to further tease apart influence!

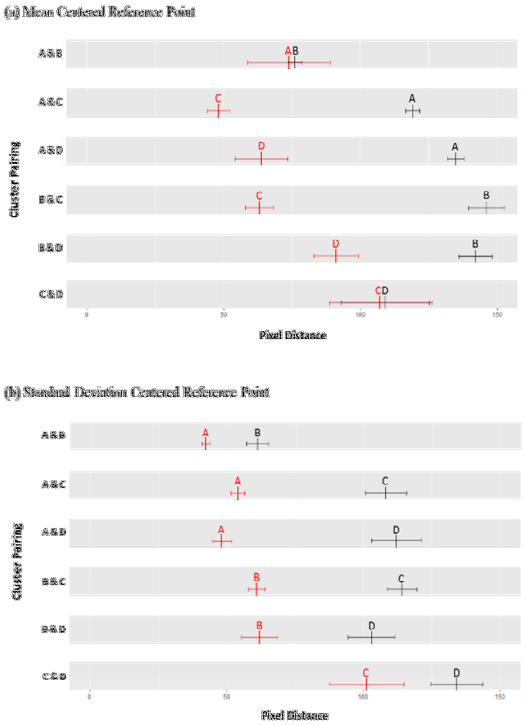


Figure 22. Average pixel distance between nearest neighbor and reference point when reference point was (a) mean centered and (b) standard deviation centered. The cluster chosen most often as being closest to the reference point is highlighted in red. Error bars represent standard deviation of the mean.

Exp 2: Block B Preliminary Results (Reference Point Task)

- What do these results tell us?
  - We can find drastically different interpretations for which cluster a reference point belongs to based on what method was used to center it, given two clusters with different dispersion levels
    - Seemingly low-level, simple decisions or properties of the data can have a big influence on decision makers!
  - …further research will help better tease apart the driving influencers, especially if looking at a particular domain

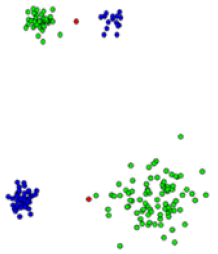
Exp 2: Block B Preliminary Results (Highest Cluster Task)

Sandia National Laboratories

Highest cluster task: which cluster has the highest vertical mean?

Exp 2: Block B Preliminary Results (Highest Cluster Task)

- Highest cluster task preliminary results:
  - Analyzed decision for clusters with same mean height
  - Participants tended to indicate that the highest cluster was the one with:
    - **Lower sparsity** (more data points)
      - 79.2% of the time (*CI* [70.8%, 86.0%], $p < .001$)
      - Regardless of dispersity or centering method
    - **Higher dispersion** (points more spread out)
      - 70.8% of the time (*CI* [61.8%, 78.8%], $p < .001$)
      - Regardless of sparsity or centering method
      - Note: Centering method had no effect on this task!
  - Take away: dispersion drove decision (sparsity was secondary)
    - Centering method had little influence—but reference point was not very important for this task



Exp 2: Block B Preliminary Results (Highest Cluster Task)

…but once again, what else might have been driving these effects?

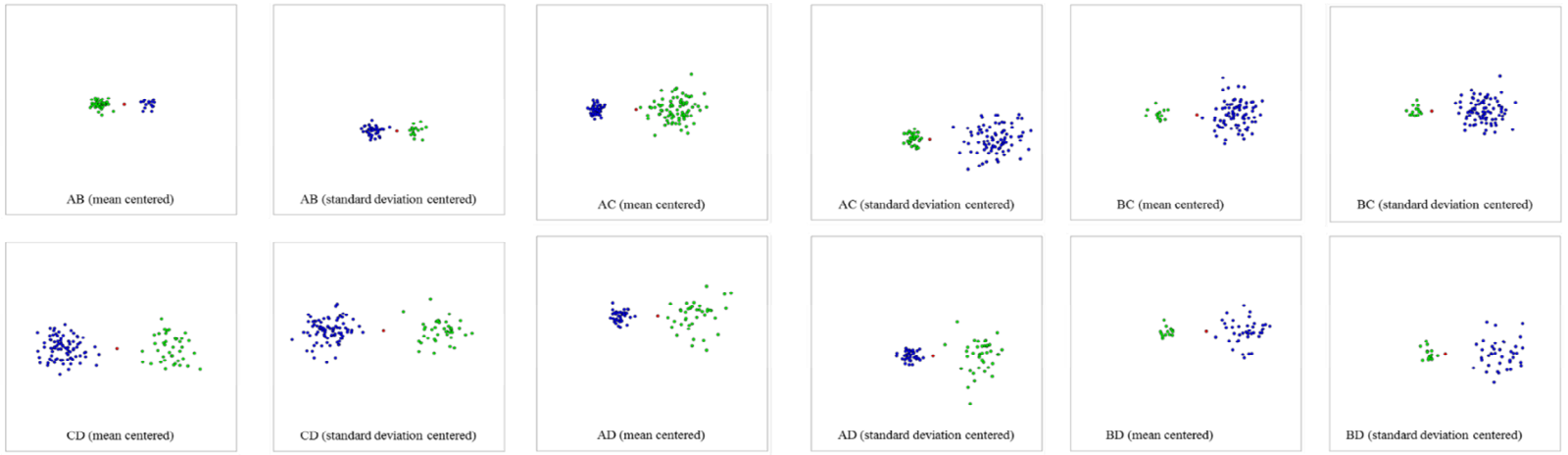


Figure B2. Examples of cluster pairings by type for stimuli whose values are drawn from normal distributions with the same standard deviation.

Figure B3. Examples of cluster pairings by type for stimuli whose values are drawn from normal distributions with different standard deviations.

Just relying on highest overall point?

…chose cluster with highest overall point 85% of the time

Exp 2: Block B Preliminary Results (Highest Cluster Task)

- What do these results tell us?
  - Factors (e.g., centering method) that influence some tasks may **not** influence others (e.g., reference point membership vs. highest point)
    - Top-Down influence (goals and expectations)
  - Dispersion and sparsity were strongly associated with decisions, but correlated with highest overall point
    - May need to be particularly careful with outliers—could have a strong influence on perception of underlying properties of data

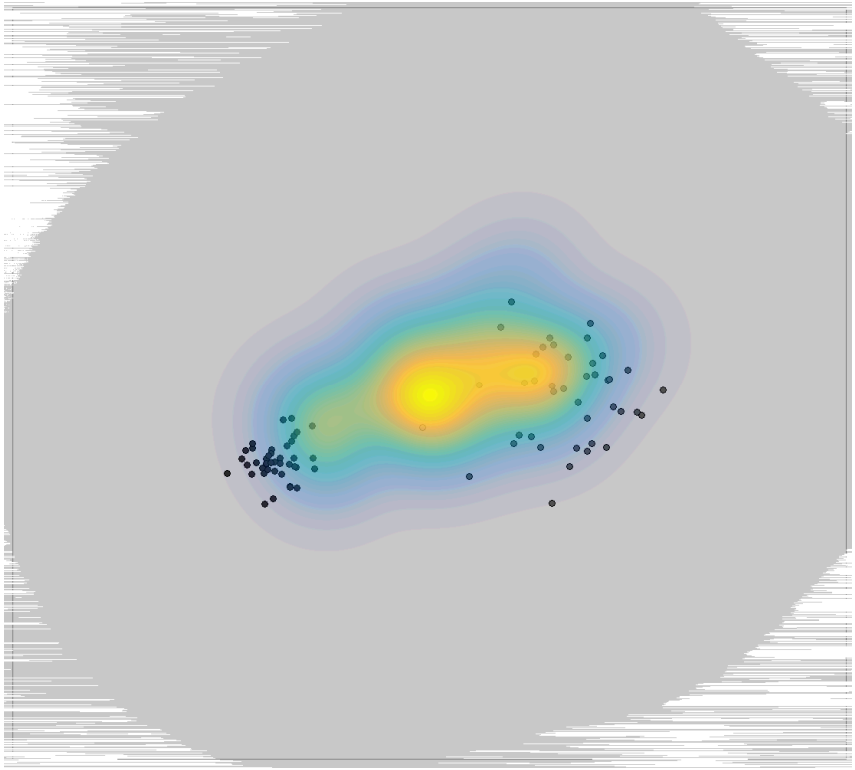
Experiment 2 – Eye tracking results

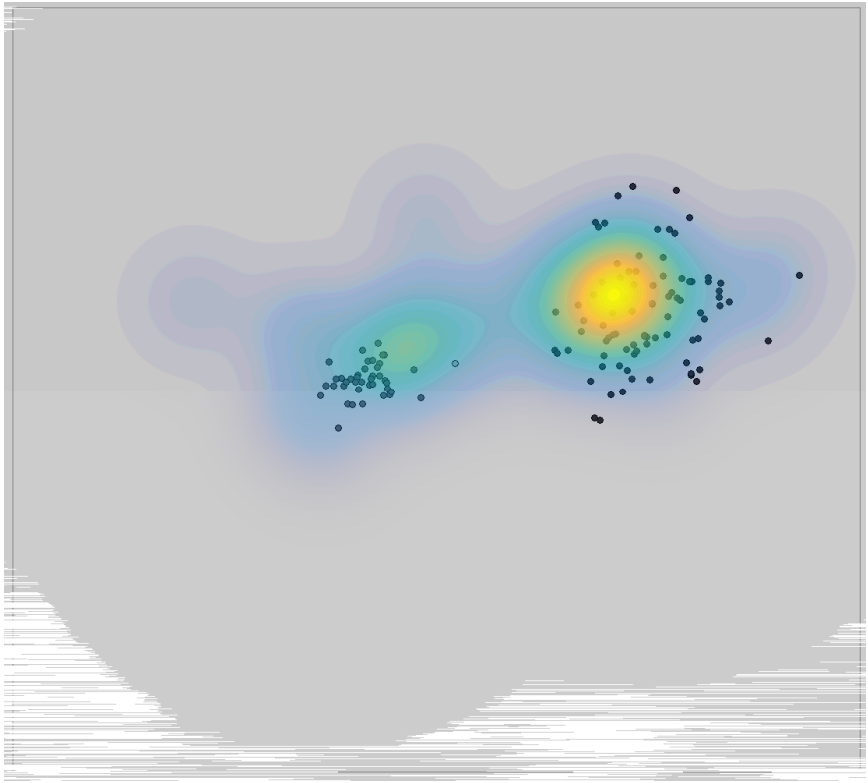
- When comparing fixations to the same image, but for different tasks…
- No difference in overall number of fixations based on task
  - Significantly more fixations to the reference point in the Reference Task
  - Significantly more fixations to the clusters themselves in the Mean Height Task
- Significantly longer fixations overall in Reference Task
  - Significantly longer fixations to the right cluster in Mean Height Task
- Different patterns of fixations based on task
  - Right cluster used as “baseline” in Mean Height Task
  - Repeated looks to reference point in Reference Task

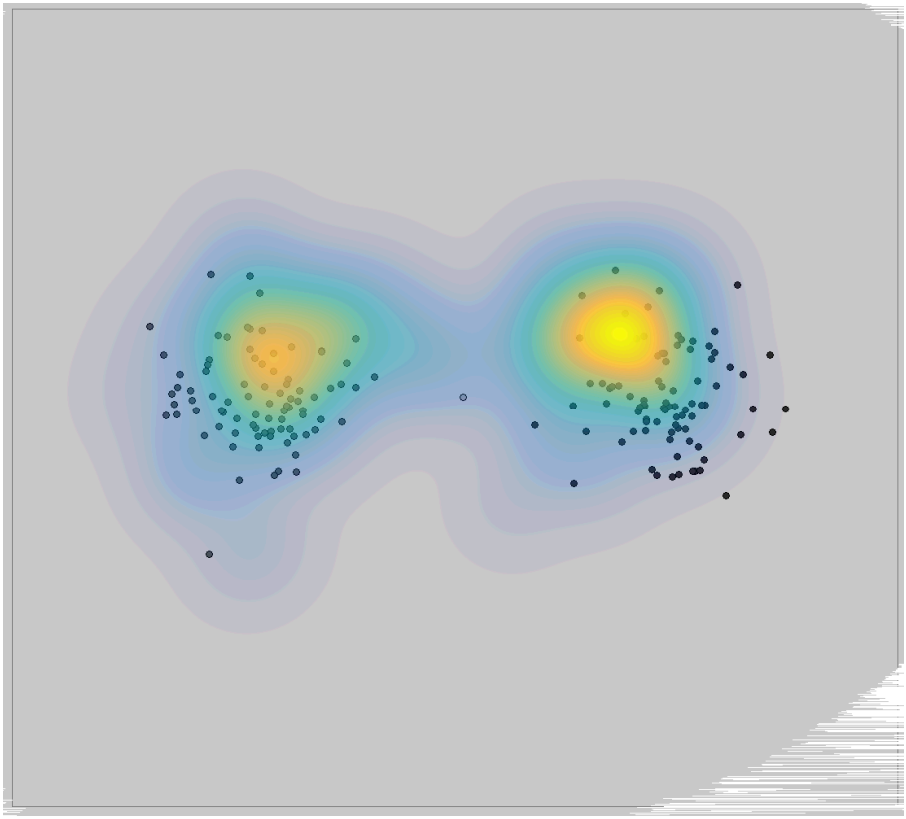
Exp 2: Summary

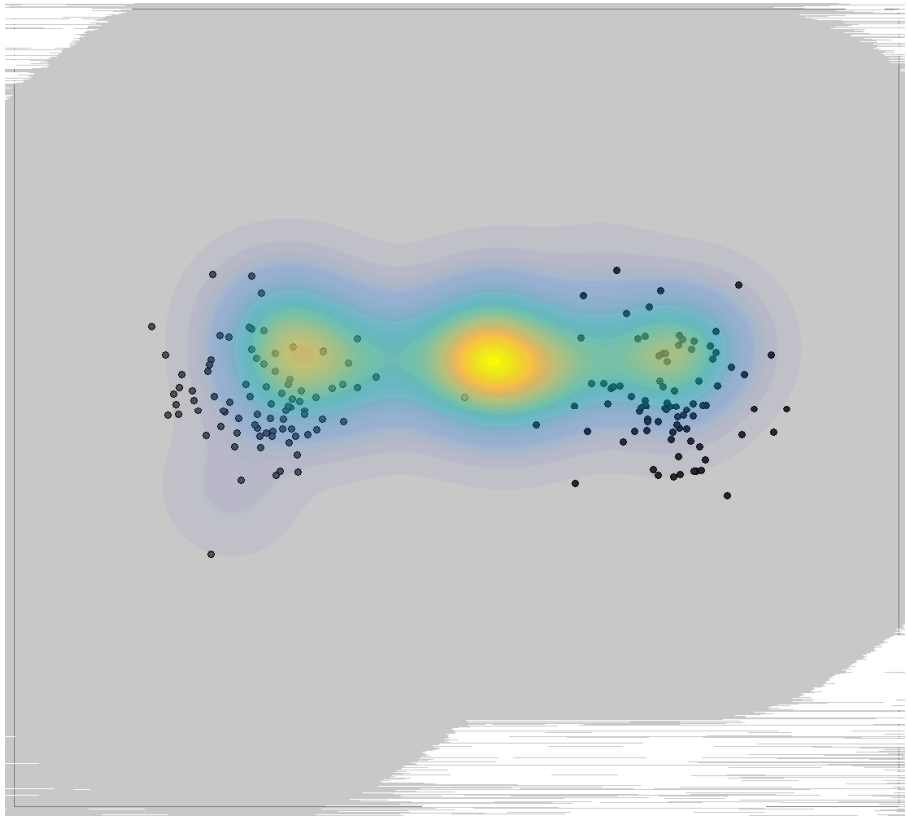
- What do the results of Experiment 2, taken together, tell us?
  - Decisions you make in how to represent your data are **important**
  - Users may not interpret visualizations how you expect
    - E.g., does a data point count as an outlier? Which cluster does it belong to?
      - Likely not just using standard statistical definitions (e.g., >3 stdev away or closest mean)
      - Heuristics and/or combination of properties
  - Important to keep in mind for relatively **simple** visualizations
    - Likely just as or more important for more **complex** visualizations
  - The more complex/large the data you want to condense down to an abstract data visualization, the more you should think about how the user will interpret it!

- …and more to come from eye movement data & DVS model









Thank you!

- Questions?

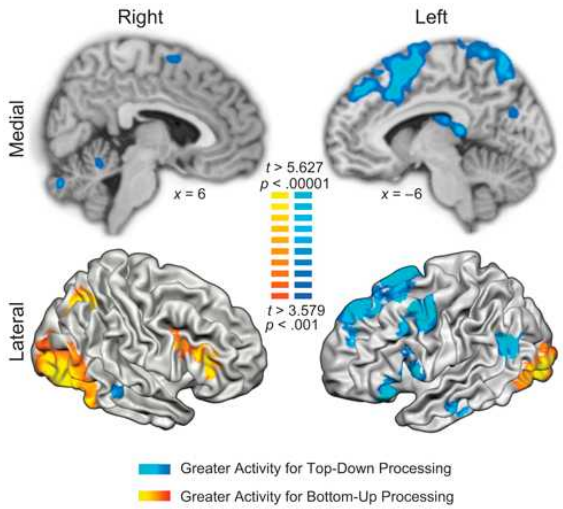
- Contact:
  - lematze@sandia.gov

Backup Slides

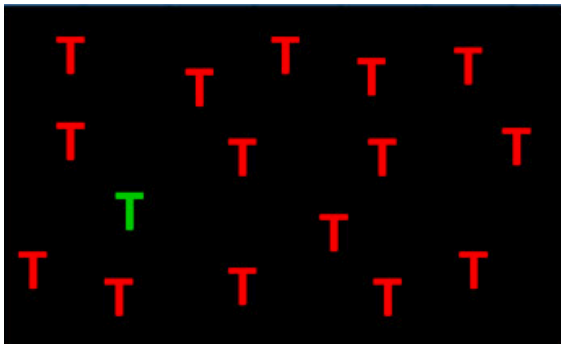
Bottom-up versus top-down visual processing

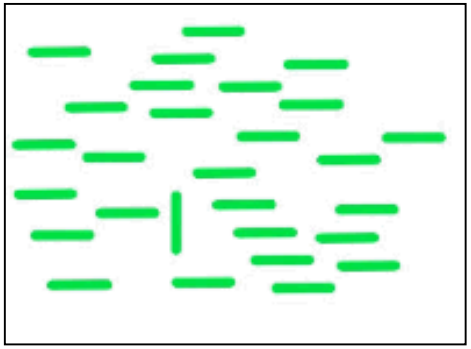
- Two parallel neural processes that guide visual processing
  - Bottom-up = stimulus-driven visual attention
  - Top-down = goal-oriented visual attention

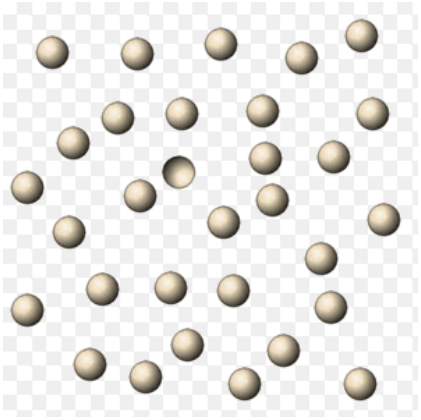
Ochsner et al. Psych Science, 20, 1322-1331.



- Bottom-up attention is captured *automatically* by the physical properties of a stimulus
  - Color, shape, orientation, motion



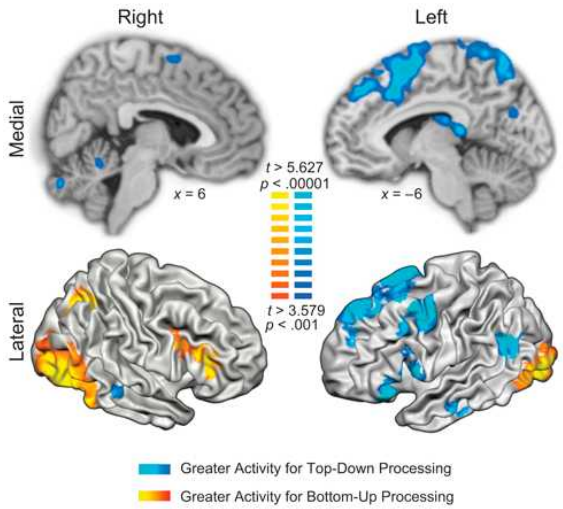




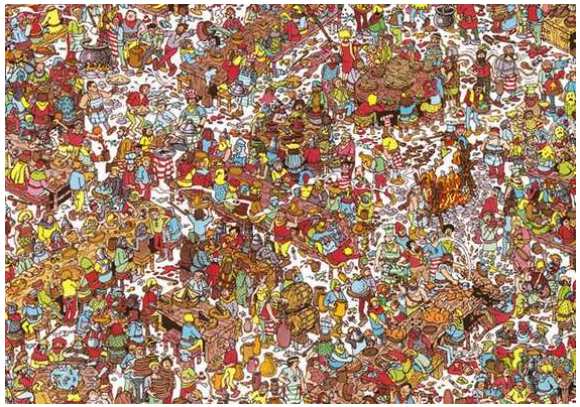
Bottom-up versus top-down visual processing

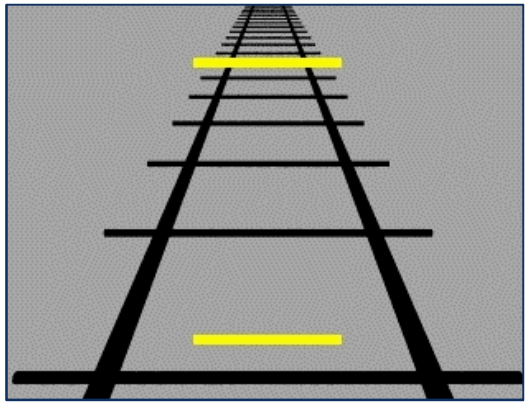
- Two parallel neural processes that guide visual processing
  - Bottom-up = stimulus-driven visual attention
  - Top-down = goal-oriented visual attention

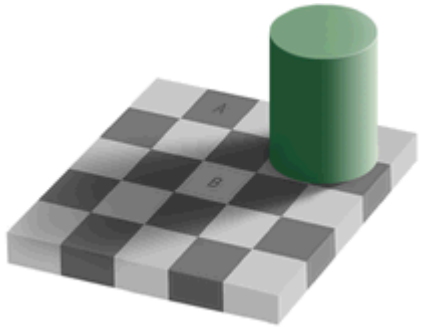
Ochsner et al. Psych Science, 20, 1322-1331.



- Top-down attention is allocated *voluntarily* according to the viewer's goals and expectations
  - Current goal, past experience, cognitive load

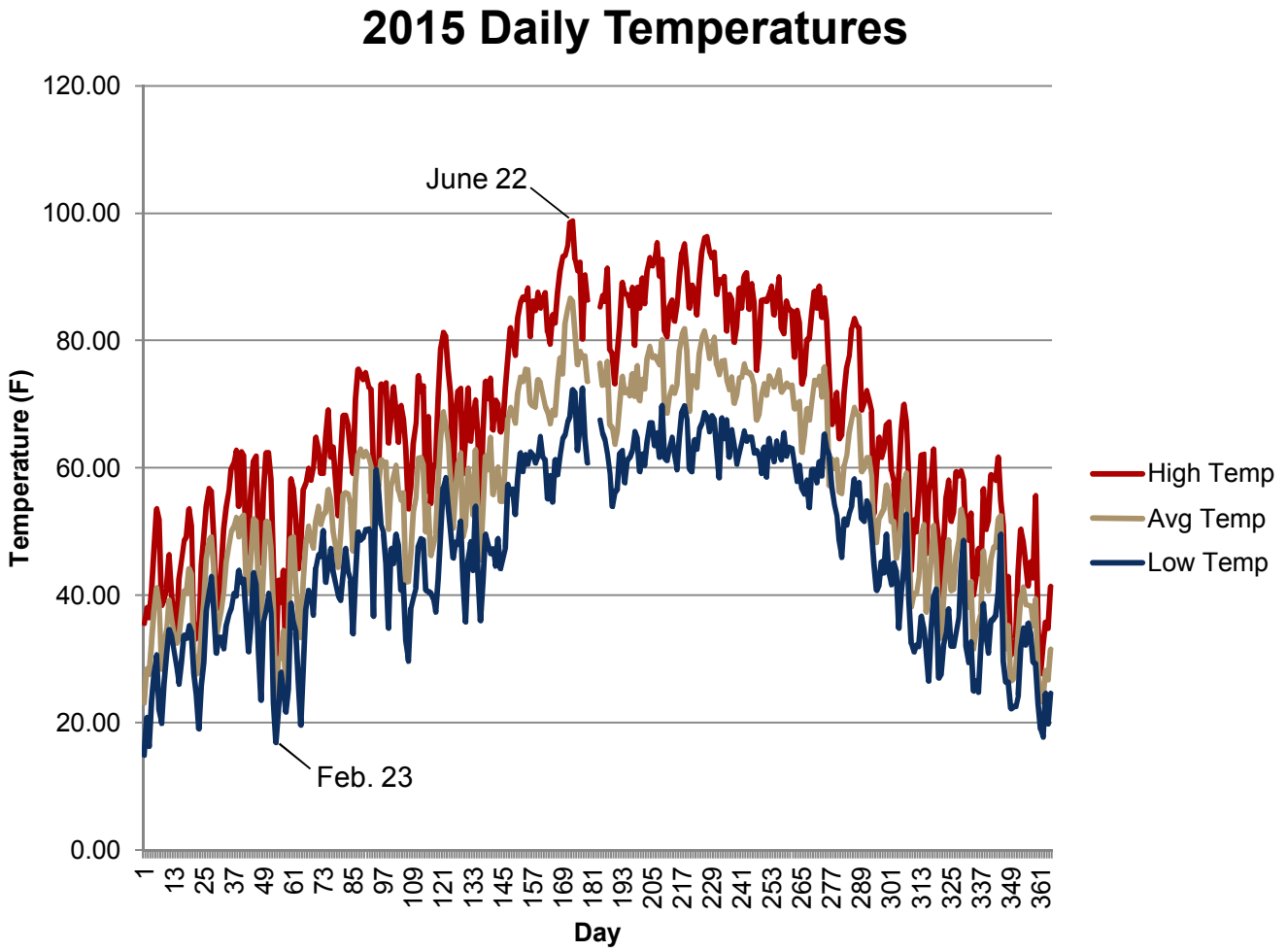






Regions of Interest (ROIs)

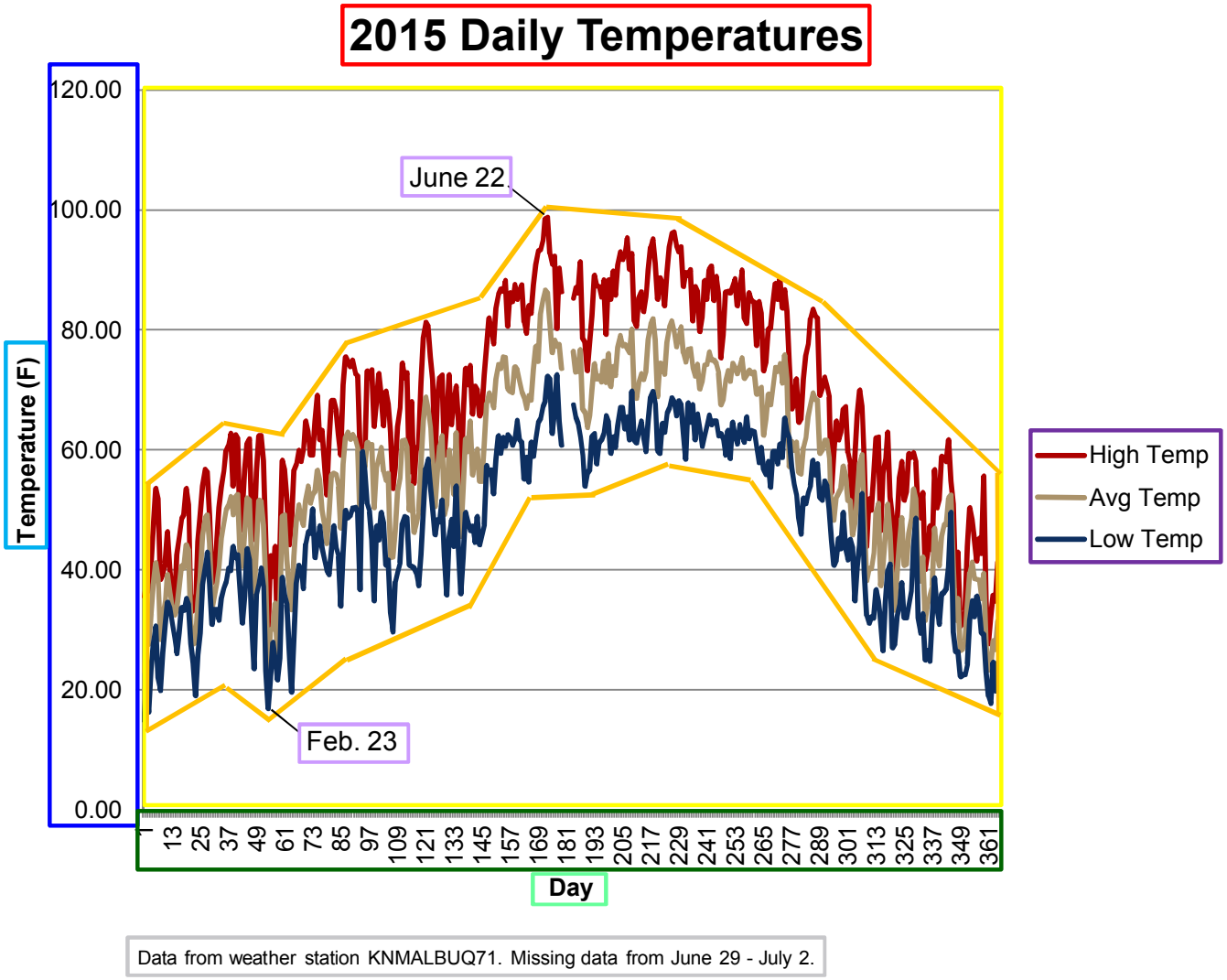
- Title
- Data
- Data Area
- X-Axis
- X-Axis Label
- Y-Axis
- Y-Axis Label
- Legend
- Data Labels
- Text



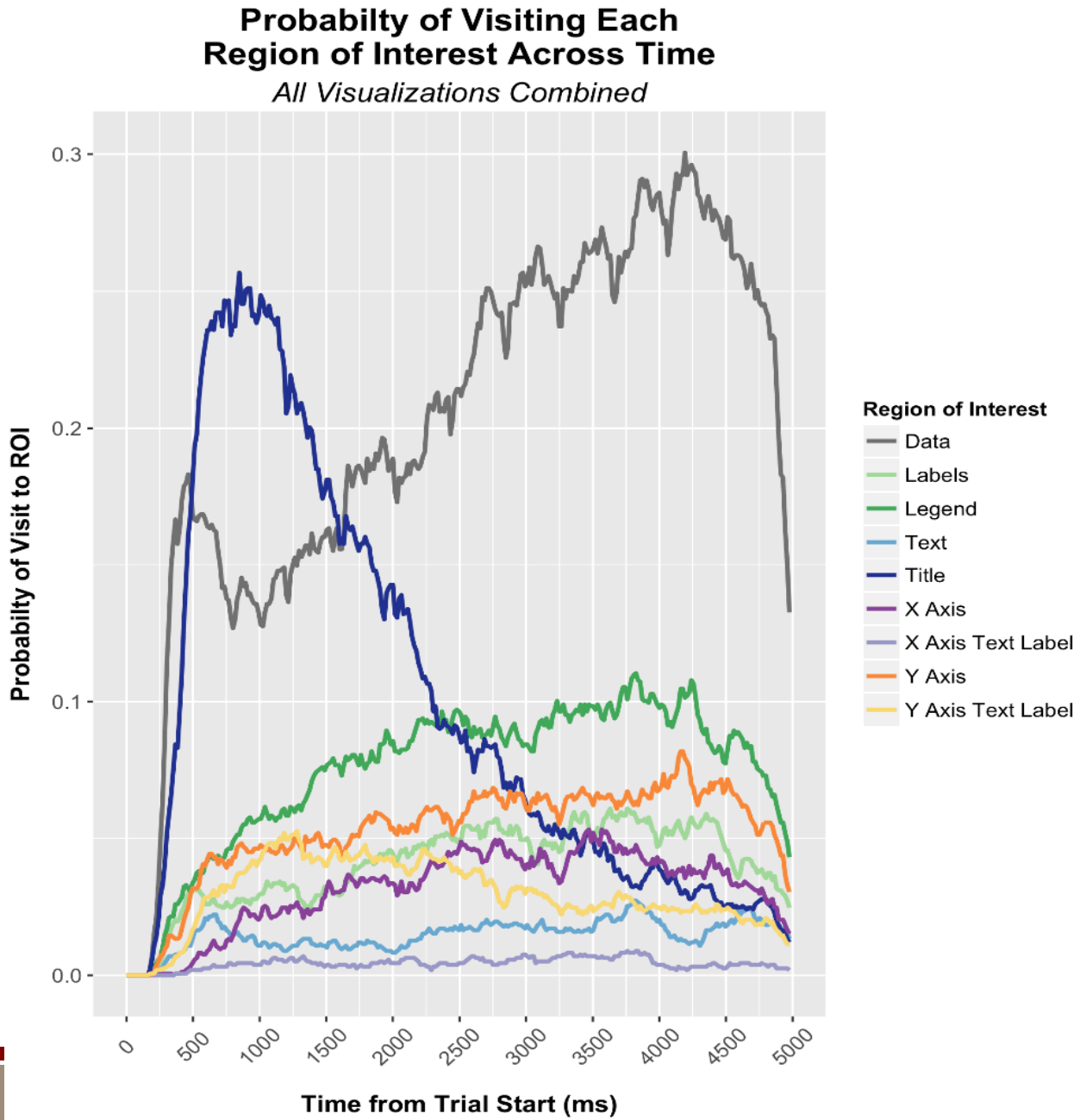
Data from weather station KNMALBUQ71. Missing data from June 29 - July 2.

Regions of Interest (ROIs)

- Title
- Data
- Data Area
- X-Axis
- X-Axis Label
- Y-Axis
- Y-Axis Label
- Legend
- Data Labels
- Text

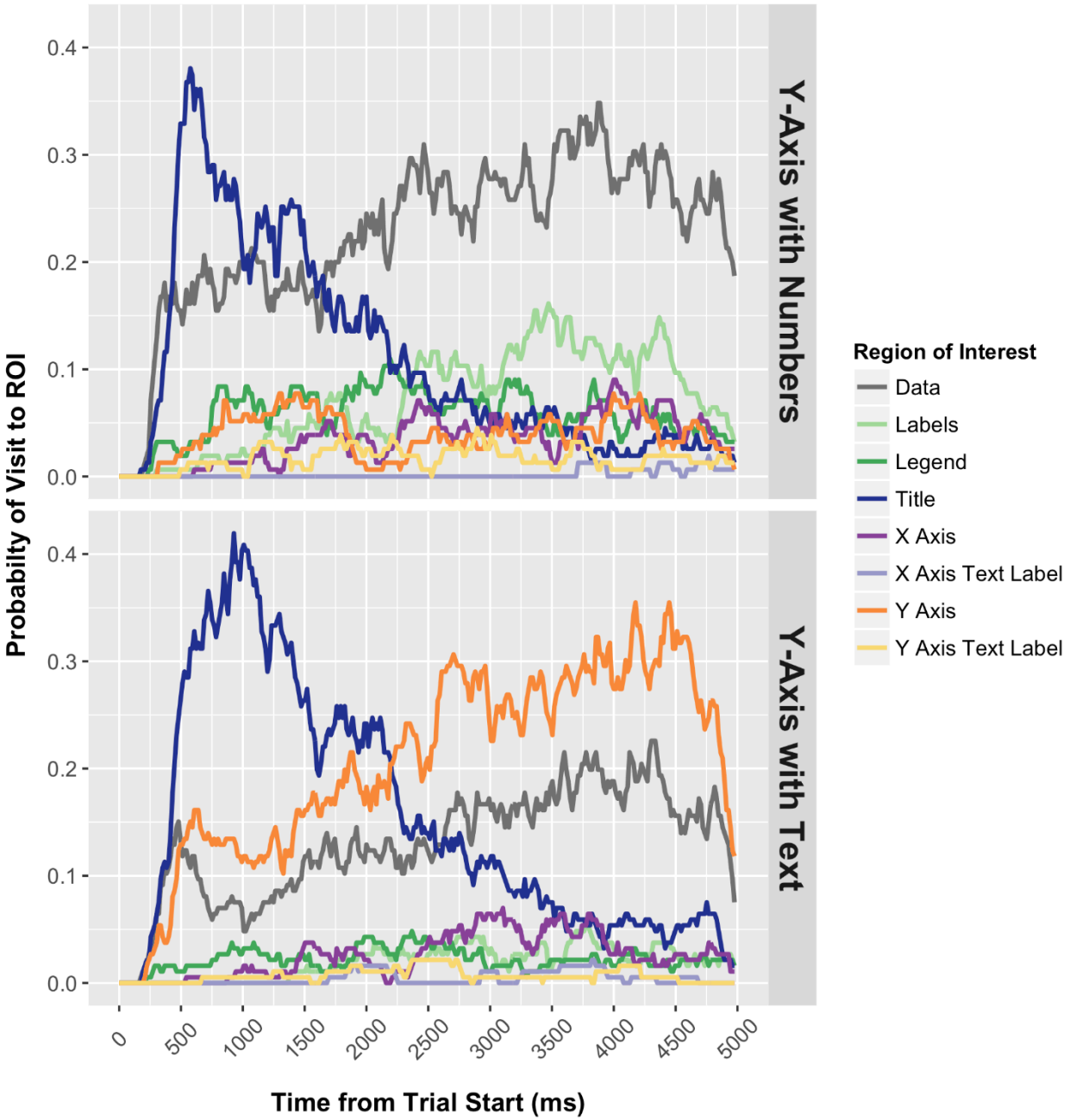


Text Influences Order of Attention



Probabilty of Visiting Each Region of Interest Across Time

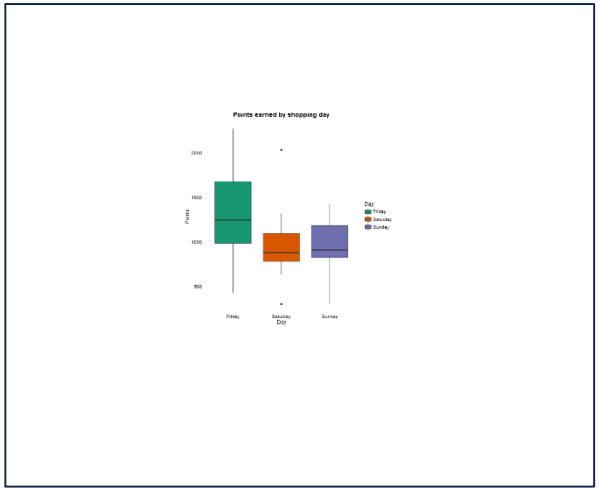
Visualizations with and without Y-Axis Text



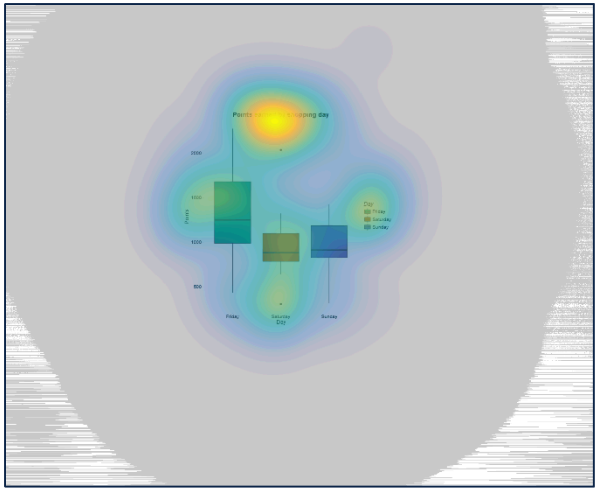
Subsequent Work

- Data Visualization Saliency Model (DVS)

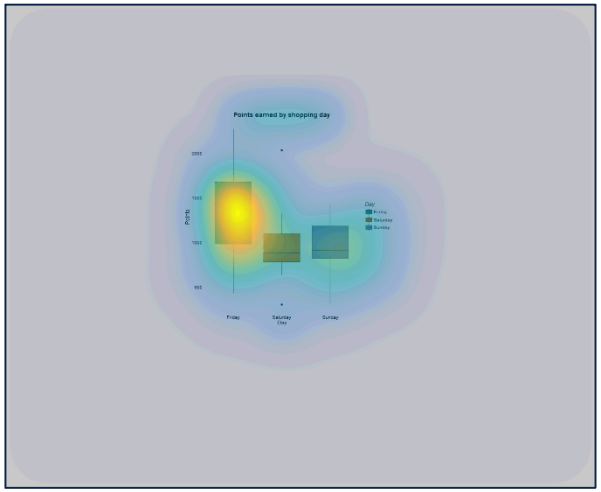
Original Vis



Fixation Map

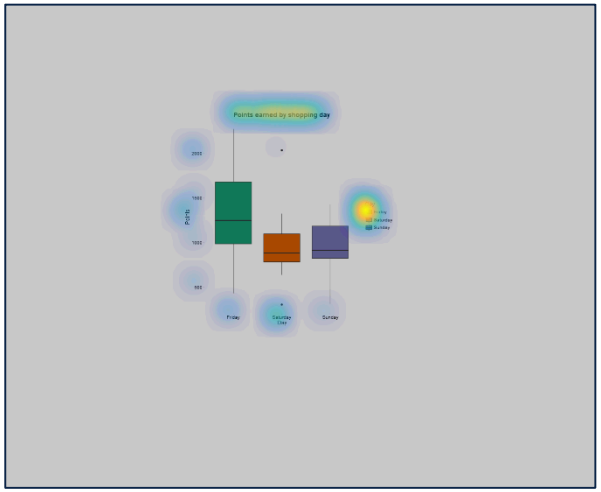


Modified Itti Saliency Map



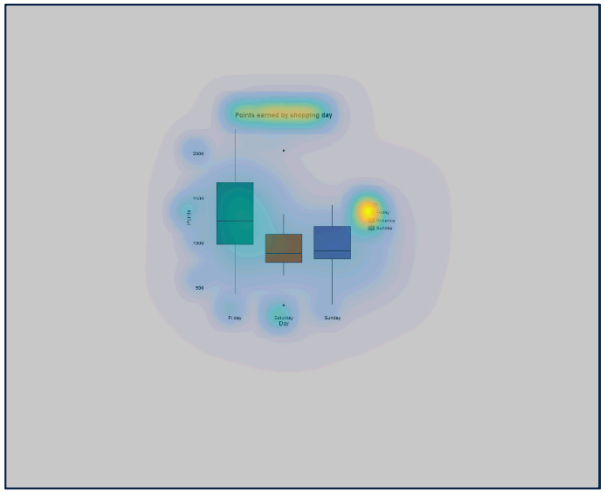
+

Text Saliency Map



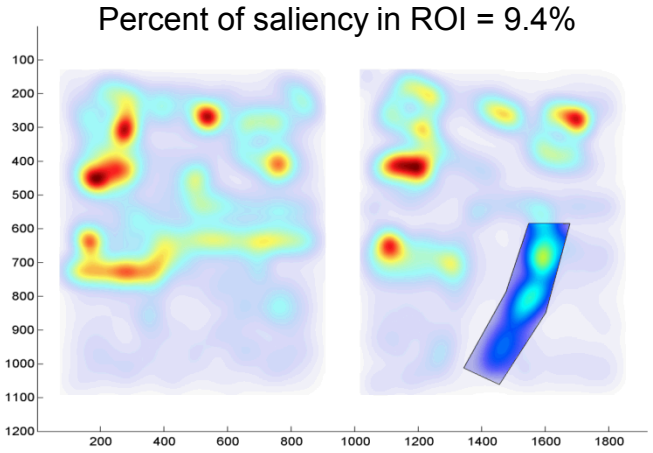
=

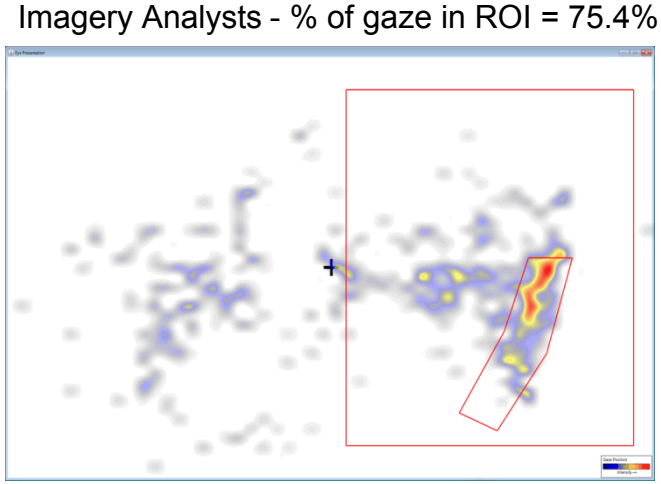
Data Visualization Saliency Map



Visual attention in data visualizations

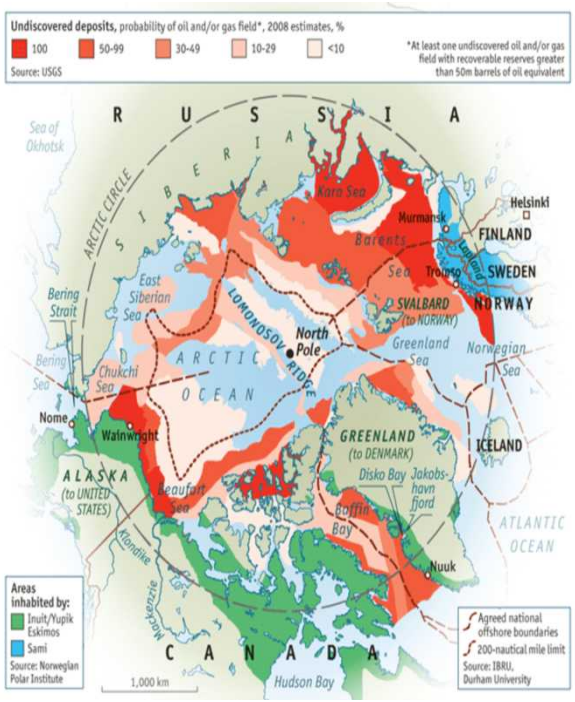
- Ideally, an image would draw the viewers' attention to the most important information for their task
  - Stimulus-driven (bottom-up) visual attention
    - Is the important information visually salient?
  - Goal-driven (top-down) visual attention
    - How do goals and prior experience influence navigation through imagery?

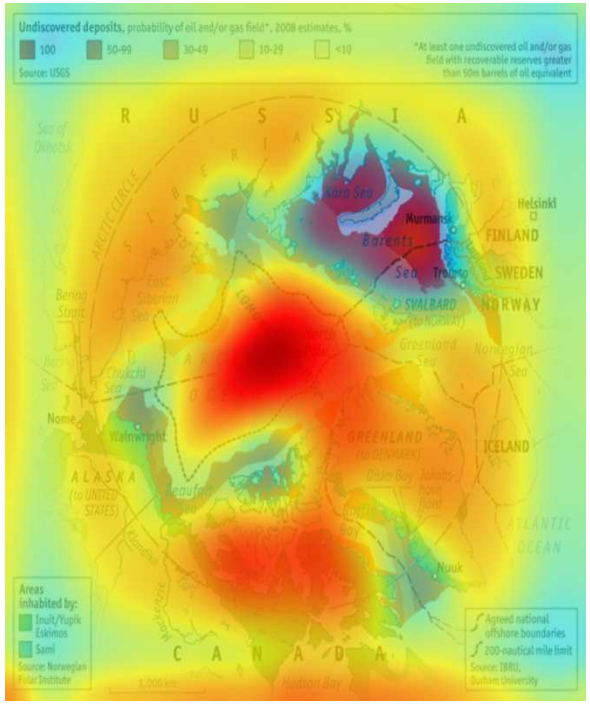




Findings

- Incomplete feature sets
  - RGB color space poorly calibrated to physiological response





Existing Saliency Models for Natural Scenes

- Models based on neural architecture can predict where people will look in natural scenes
  - Would be a useful tool for assessing vis designs – do they draw bottom-up attention as the designer intends?