Modeling Color Difference for Visualization Design

Danielle Albers Szafir
University of Colorado Boulder
@dalbersszafir
Warning: Colors and shadows on projectors ahead!

Follow along at: https://goo.gl/rQDWU7

Danielle Albers Szafir
University of Colorado Boulder
@dalbersszafir
Seven Step Encoding in CIELAB
Seven Step Encoding From ColorBrewer
Seven Step Encoding From ColorBrewer
How does visualization design change how we perceive color encodings?
Factors for Color Difference in Visualization

Scatterplots

Bar Charts

Line Graphs

Summary & Applications
Factors for Color Difference in Visualization

Scatterplots

Bar Charts

Line Graphs

Summary & Applications
CIELAB

Commonly used in visualizations

Approximately perceptually linear

1 unit Euclidean difference equals 1 Just Noticeable Difference (JND)
CIELAB

Commonly used in visualizations

Approximately perceptually linear

1 unit Euclidean difference equals 1 Just Noticeable Difference (JND)
Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption
Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption
Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption

Crowdsourced Sampling
Szafir, Stone, & Gleicher, 2014
Reinecke, Flatla, & Brooks, 2016
Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption
Visualizations violate three CIELAB assumptions

- Simple World Assumption
- Isolation Assumption
- Geometric Assumption
Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption
Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption

Size-Based Sampling
Carter & Silverstein, 2010
Stone, Szafir, & Setlur, 2014
Visualizations violate three CIELAB assumptions

- Simple World Assumption
- Isolation Assumption
- Geometric Assumption
Visualizations violate three CIELAB assumptions

- Simple World Assumption
- Isolation Assumption
- Geometric Assumption

Diagonally Symmetric Marks

Elongated Marks

Asymmetric Marks

Area Marks
Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption
Renormalize CIELAB per Mark Type

Scale each axis such that p% of viewers will identify a difference at one unit of Euclidean distance – a \( p\% \text{ JND} \)

Szafir, Stone, & Gleicher, 2014
Stone, Szafir, & Setlur, 2014
Factors for Color Difference in Visualization

Scatterplots – Diagonally Symmetric

Bar Charts

Line Graphs

Summary & Applications
Do the colorful marks appear the same or different?
79 reference colors
36 total color differences

Reference Color

Differed by fixed amount on L*, a*, or b*
Random Gaussian with Overdraw Removed

L* = 50 Distractor Marks
6 (diameters, within) × 6 (color differences, within) × 3 (color axis, between)

81 participants on Mechanical Turk (5,668 trials)
Factor Analysis:
ANCOVA with question order and source color as covariates

Modeling Process:
Compute rate of perceived differences for size x color difference

Model rates using linear regression fit to origin controlling for covariates

Sample linear models for desired probabilities

Fit logistic regression to the samples
50% JND for Scatterplot Points

\[ \Delta E = 1.0 \]

Stone et al.

\[ \Delta b^* \]
\[ \Delta a^* \]
\[ \Delta L^* \]
\[
\begin{align*}
\text{ND}_L (50\%, s) &= 5.095 + \frac{0.80}{s}, \quad R^2 = .93 \\
\text{ND}_a (50\%, s) &= 5.089 + \frac{2.69}{s}, \quad R^2 = .99 \\
\text{ND}_b (50\%, s) &= 6.786 + \frac{3.20}{s}, \quad R^2 > .99
\end{align*}
\]

\[
\Delta E_p = \sqrt{\left(\frac{\Delta L}{\text{ND}_L(p)}\right)^2 + \left(\frac{\Delta a}{\text{ND}_a(p)}\right)^2 + \left(\frac{\Delta b}{\text{ND}_b(p)}\right)^2}
\]
Takeaways—Diagonally Symmetric Points

7x larger than existing models

Vary with inversely size

*Replicate Stone et al., 2014*

Increased by distractor points
Factors for Color Difference in Visualization

Scatterplots

Bar Charts – Elongated

Line Graphs

Summary & Applications
79 reference colors
36 total color differences
Mark size varies in two dimensions

- **thickess**
- **length**
6 thicknesses:
6 pixels - 50 pixels

6 (thicknesses, blocked between) \times 8 \text{(lengths, blocked between)} \times 6 \text{(color differences, within)} \times 3 \text{(color axis, between)}

301 participants on Mechanical Turk (22,752 trials)
50% JND for Bars

Bar Length encoded as point size

Accounting for Gains:

Longest Edge

Shortest Edge

Area = longest edge \times shortest edge

\text{Elongation} = \frac{\text{longest edge}}{\text{shortest edge}}

\Delta b^* \text{for points}

\Delta a^* \text{for points}

\Delta L^* \text{for points}
50% JND for Bars

Bar Length encoded as point size

Accounting for Gains:

Longest Edge
Shortest Edge
Area = longest edge \times shortest edge

Elongation = \frac{\text{longest edge}}{\text{shortest edge}}
Takeaways—Elongated Marks

Vary with bar length & thickness

Predicting data perceptions by thickness gives conservative model

Gains over points are asymptotic based on elongation
Factors for Color Difference in Visualization

Scatterplots

Bar Charts

Line Graphs – Asymmetric

Summary & Applications
Do the colorful marks appear the same or different?
$6 \times 6 \times 3 = 108$ conditions

79 participants on Mechanical Turk (5,668 trials)
50% JND for Lines

JND in CIELAB

Line Thickness

Visual Angle

Δb* for points

Δa* for points

ΔL* for points

$R^2 = .97$

$R^2 = .93$

$R^2 = .90$
Takeaways—Asymmetric Marks

Vary with inversely with line thickness

Points are overly conservative for lines

Significant gains over points

$16\Delta a^* \ JND \ for \ 6 \ pixel \ points$
$9.4\Delta a^* \ JND \ for \ 6 \ pixel \ lines$
Factors for Color Difference in Visualization

Scatterplots

Bar Charts

Line Graphs

Summary & Applications
Factors Effecting Color Encodings

Simple World Assumption:
Viewing visualizations online introduces variation in data discrimination

Isolation Assumption:
The presence of other points complicates data discrimination

Geometric Assumption:
Data discrimination varies inversely with mark size
Elongating marks increases data discrimination asymptotically
Guide Effective Designs
Guide Effective Designs
Guide Effective Designs
Guide Effective Designs
Guide Effective Designs
Encoding Validation

Nine-step sequential Brewer ramps; 4px lines & 10px points
Encoding Validation

Nine-step sequential Brewer ramps; 4px lines & 10px points
Encoding Validation

13 of 18 nine-step sequential Brewer ramps are not robust
Thanks!

Data available at http://cmci.colorado.edu/visualab/VisColors/

National Science Foundation
NSF CRII: CHS #1657599

Danielle.Szafir@Colorado.edu
@dalbersszafir