Modeling Color Difference for Visualization Design

Danielle Albers Szafir
University of Colorado Boulder
@dalbersszafir

cmci

VisuaLab

Mode
for Vi

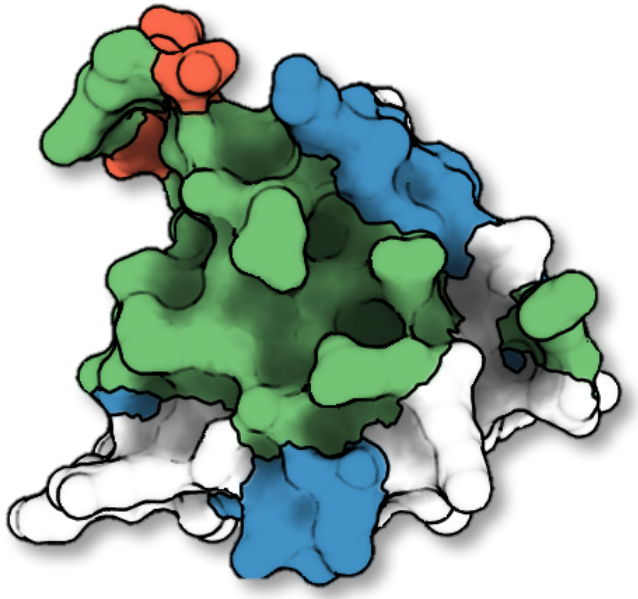
Warning: Colors and shadows on projectors ahead!

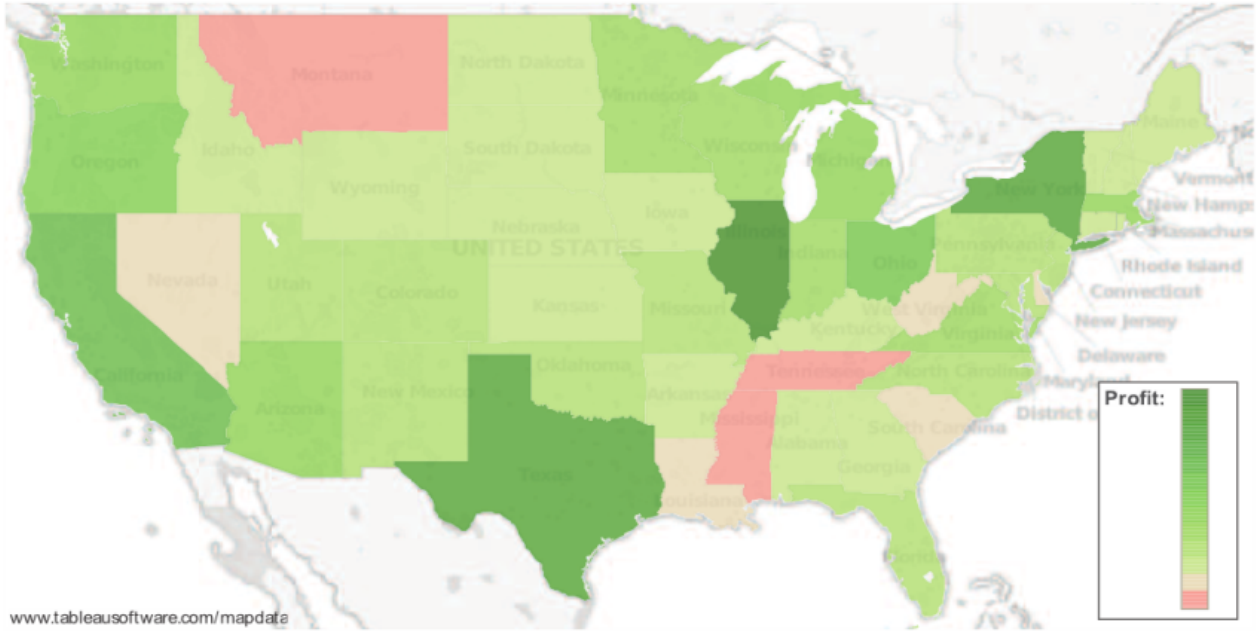
Follow along at: **https://goo.gl/rQDWU7**

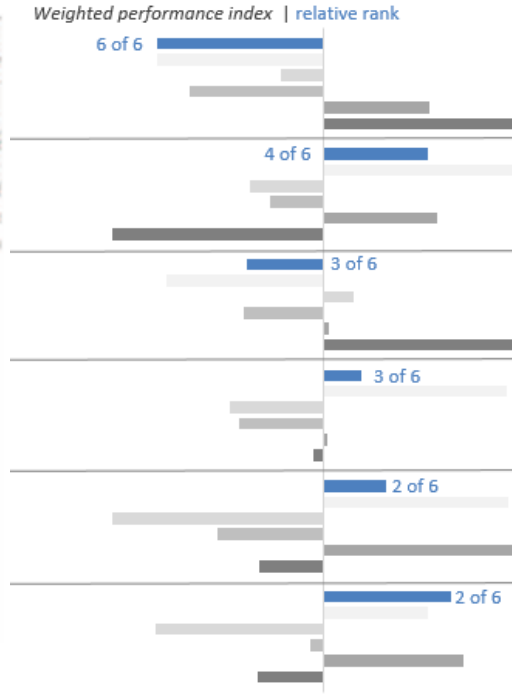
Danielle Albers Szafir
University of Colorado Boulder
@dalbersszafir

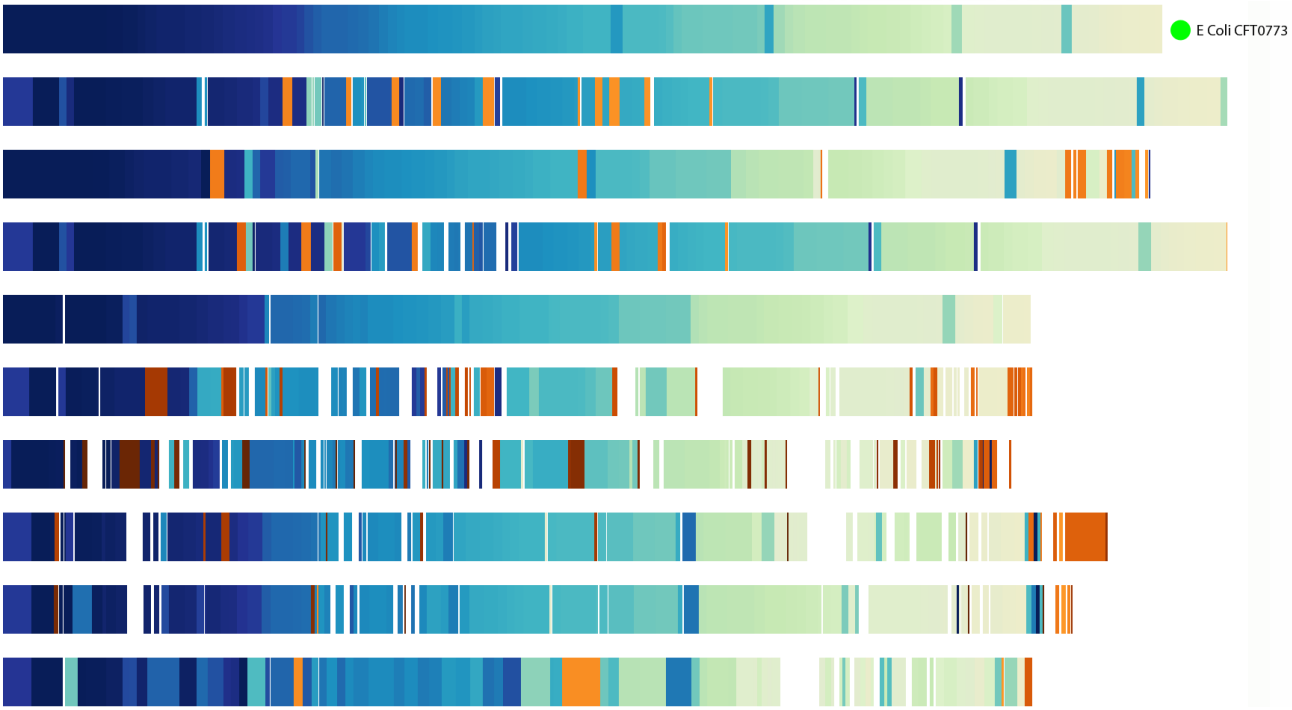
cmci

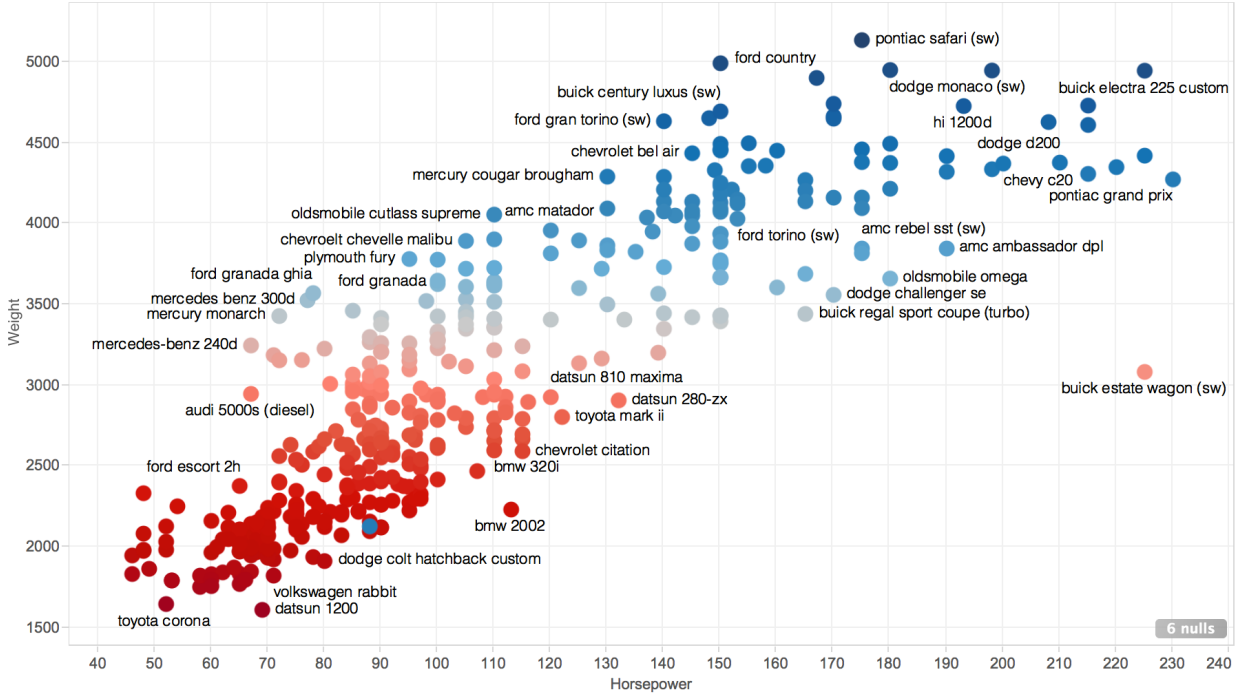
VisuaLab

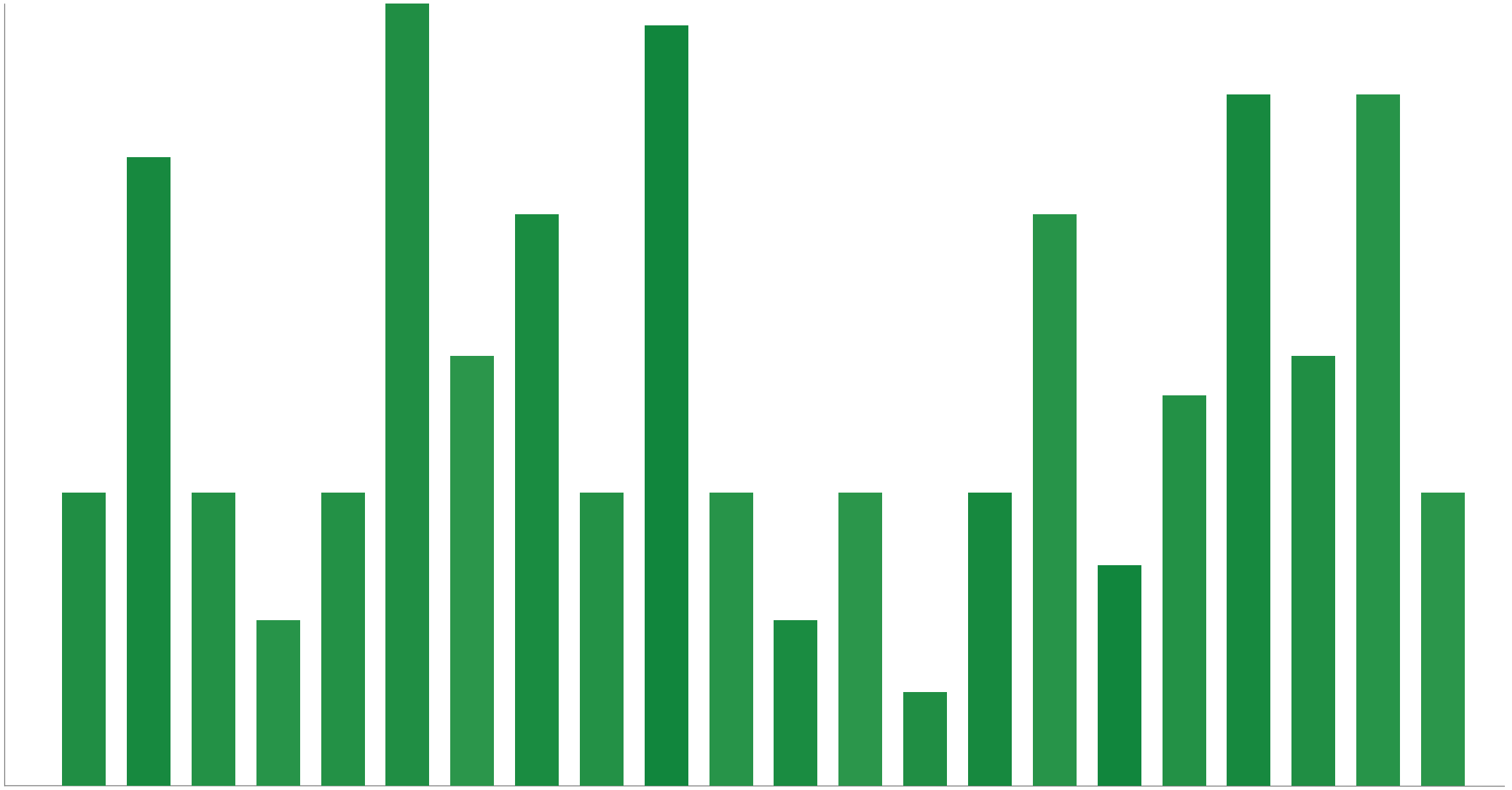



Profit:
www.tableausoftware.com/mapdata

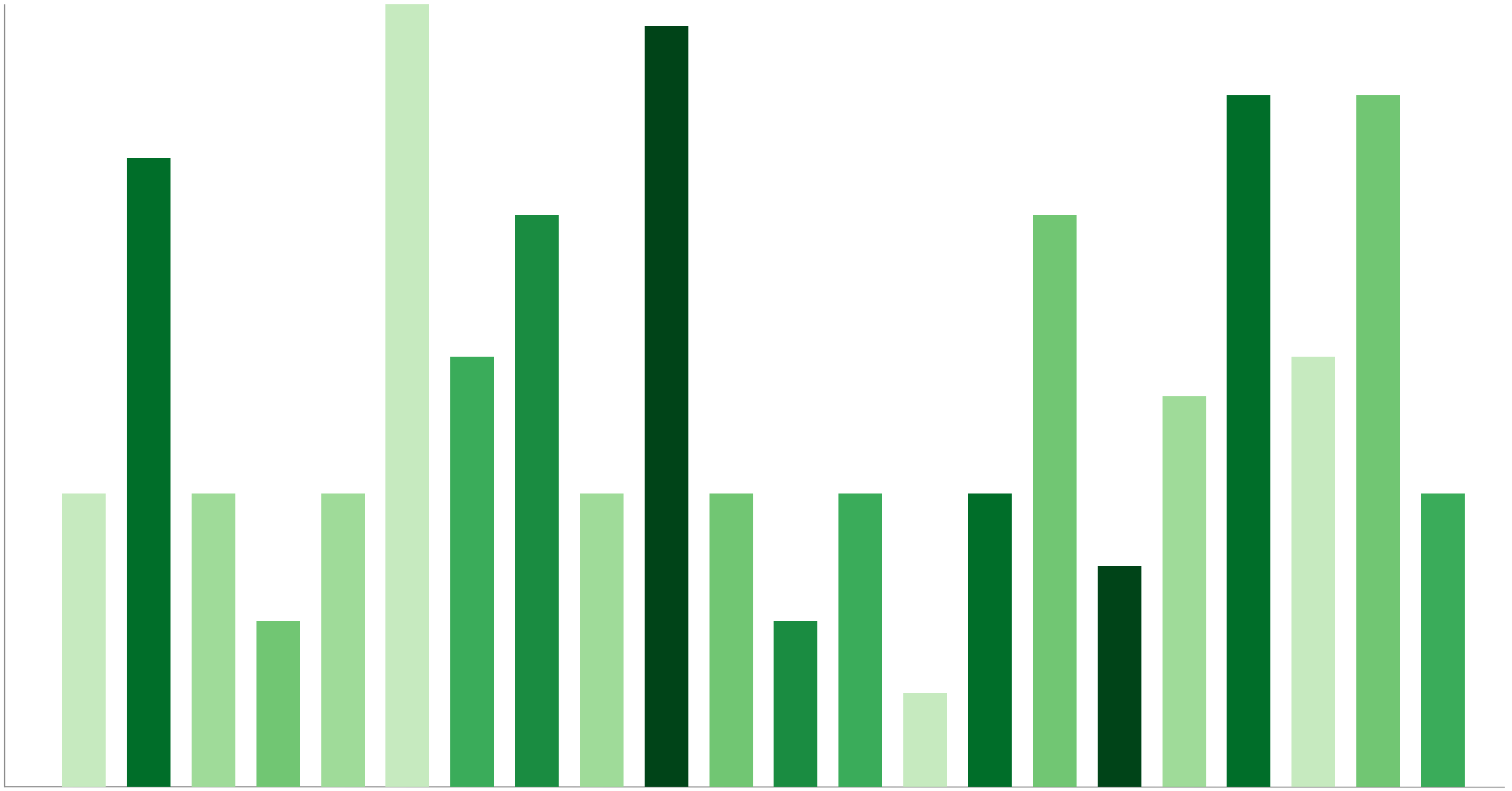

Weighted performance index | relative rank
6 of 6
4 of 6
3 of 6
3 of 6
2 of 6
2 of 6


E Coli CFT0773

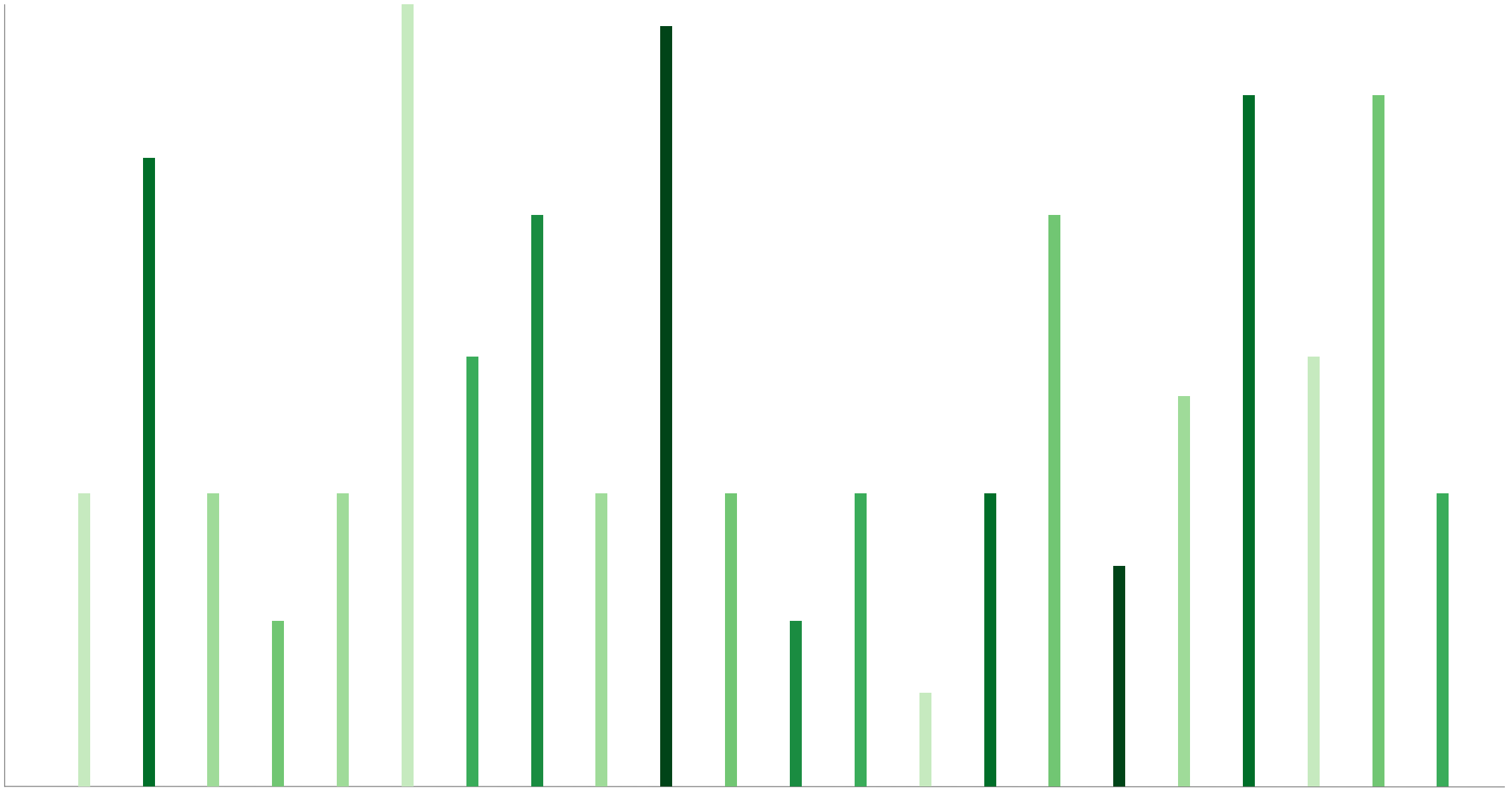

Weight
Horsepower
pontiac safari (sw)
ford country
dodge monaco (sw)
buick electra 225 custom
buick century luxus (sw)
ford gran torino (sw)
hi 1200d
dodge d200
chevrolet bel air
chevy c20
pontiac grand prix
mercury cougar brougham
oldsmobile cutlass supreme
amc matador
ford torino (sw)
amc rebel sst (sw)
chevrolet chevelle malibu
amc ambassador dpl
plymouth fury
ford granada ghia
ford granada
oldsmobile omega
mercedes benz 300d
dodge challenger se
mercury monarch
buick regal sport coupe (turbo)
mercedes-benz 240d
datsun 810 maxima
buick estate wagon (sw)
datsun 280-zx
audi 5000s (diesel)
toyota mark ii
chevrolet citation
ford escort 2h
bmw 320i
bmw 2002
dodge colt hatchback custom
volkswagen rabbit
datsun 1200
toyota corona
6 nulls



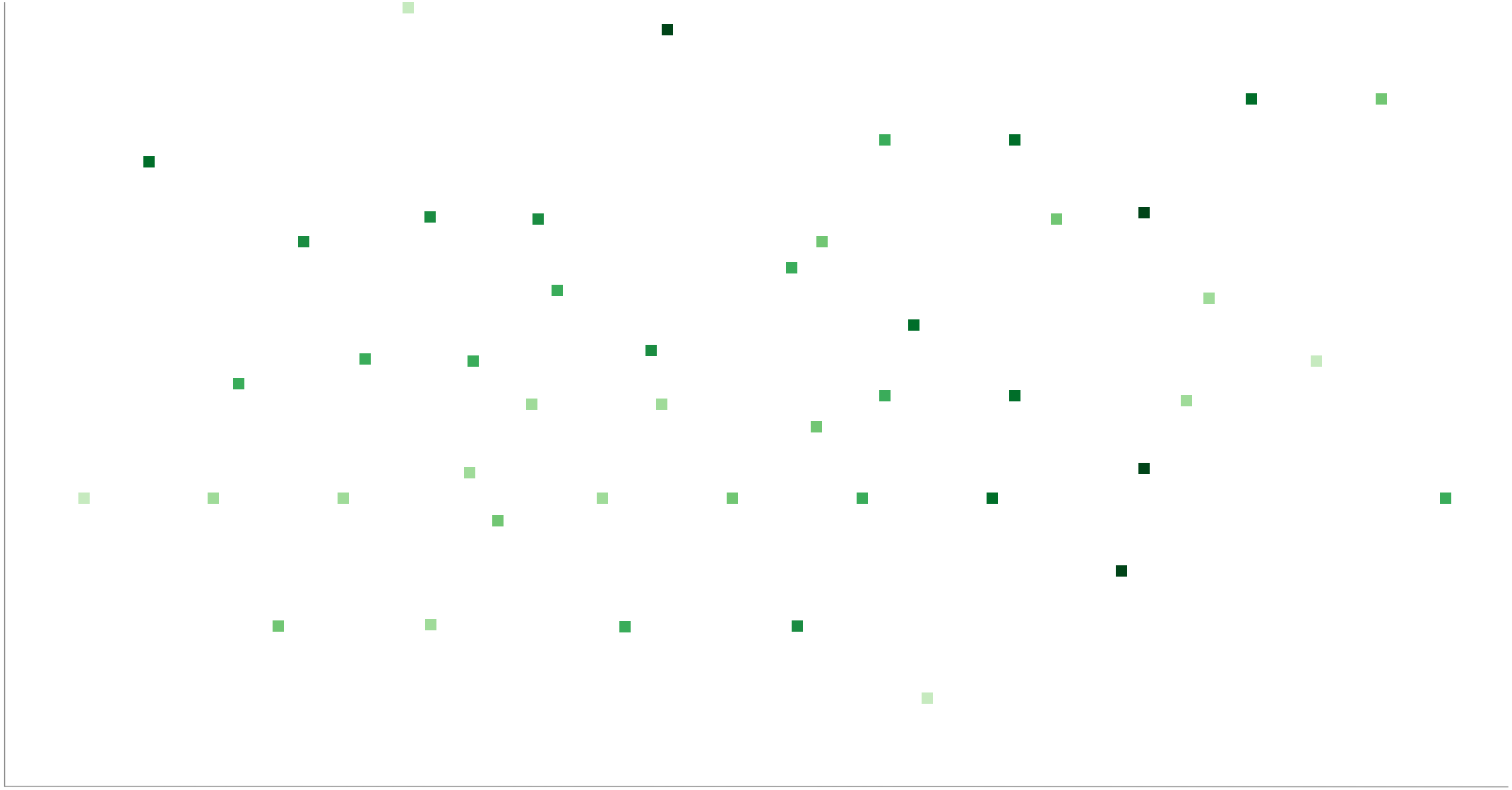
Seven Step Encoding in CIELAB



Seven Step Encoding From ColorBrewer



Seven Step Encoding From ColorBrewer



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

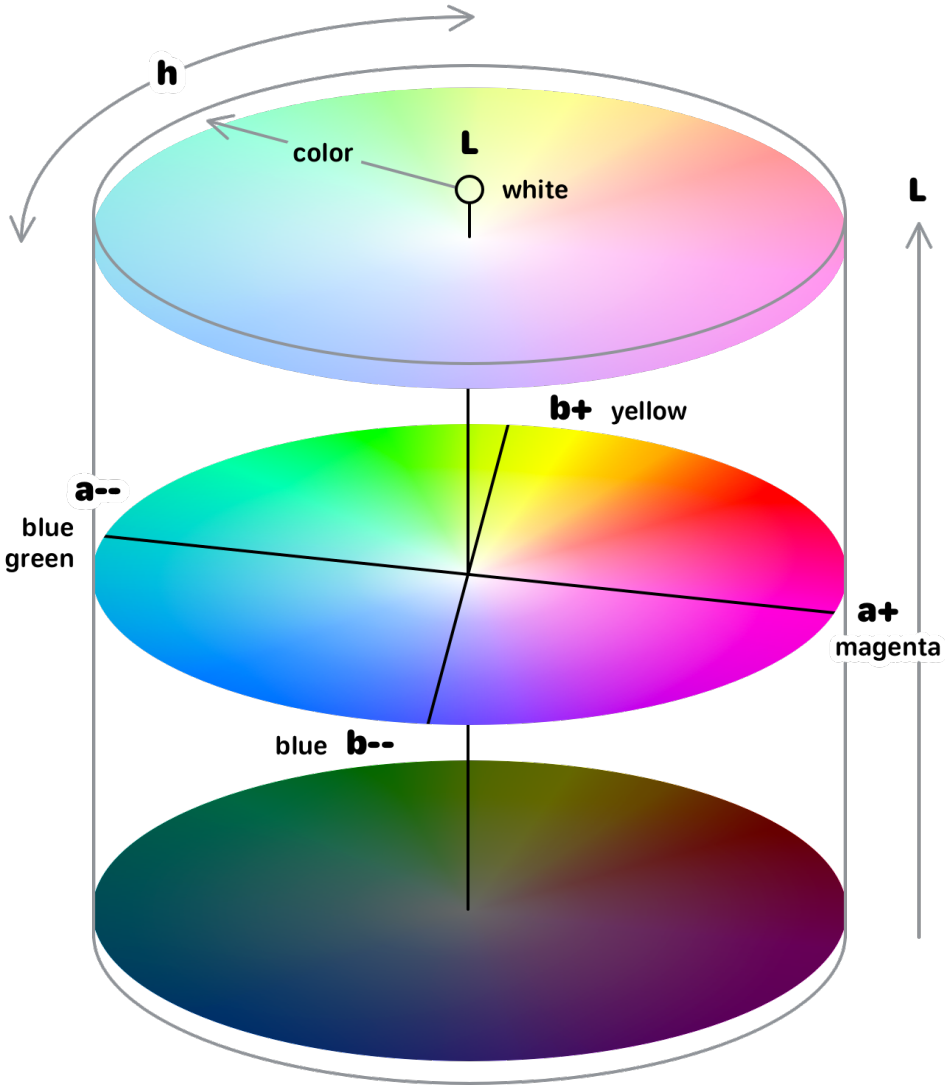
Summary & Applications

CIELAB

Commonly used in visualizations

Approximately perceptually linear

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



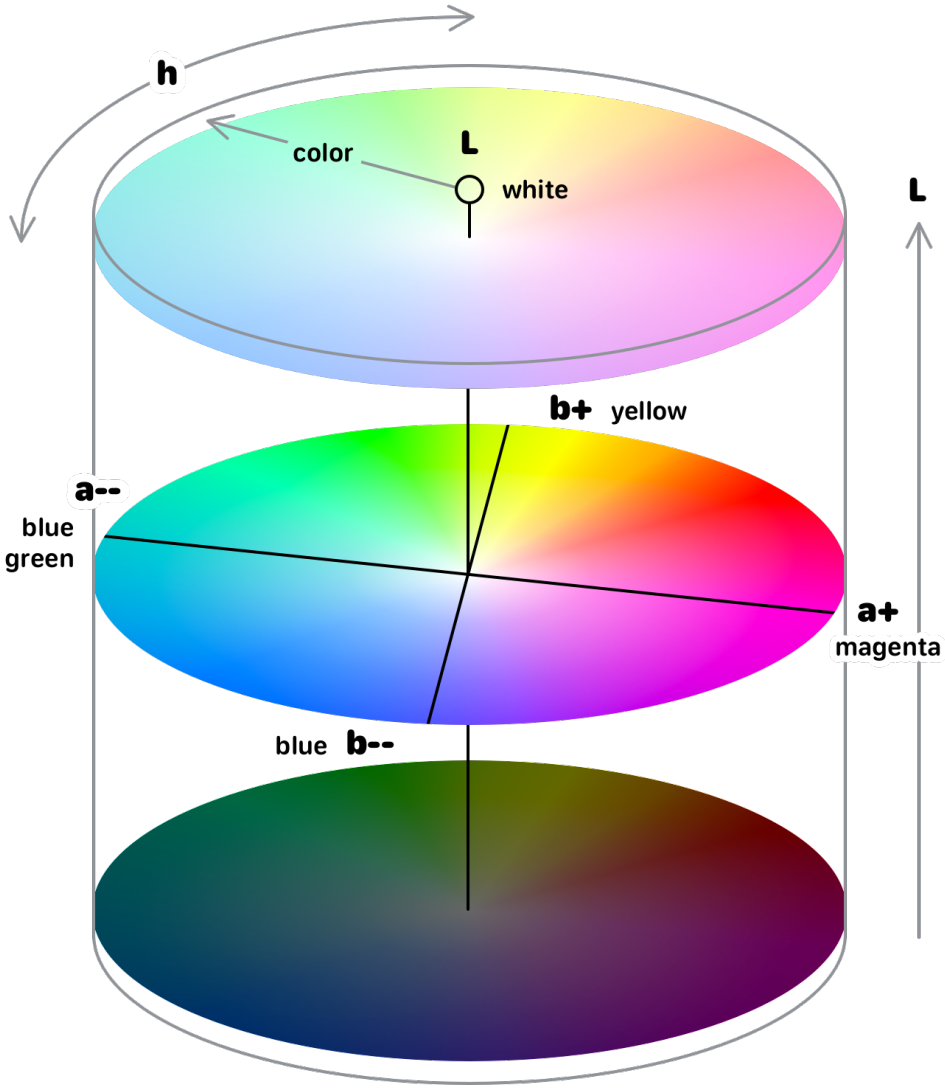
CIELAB

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

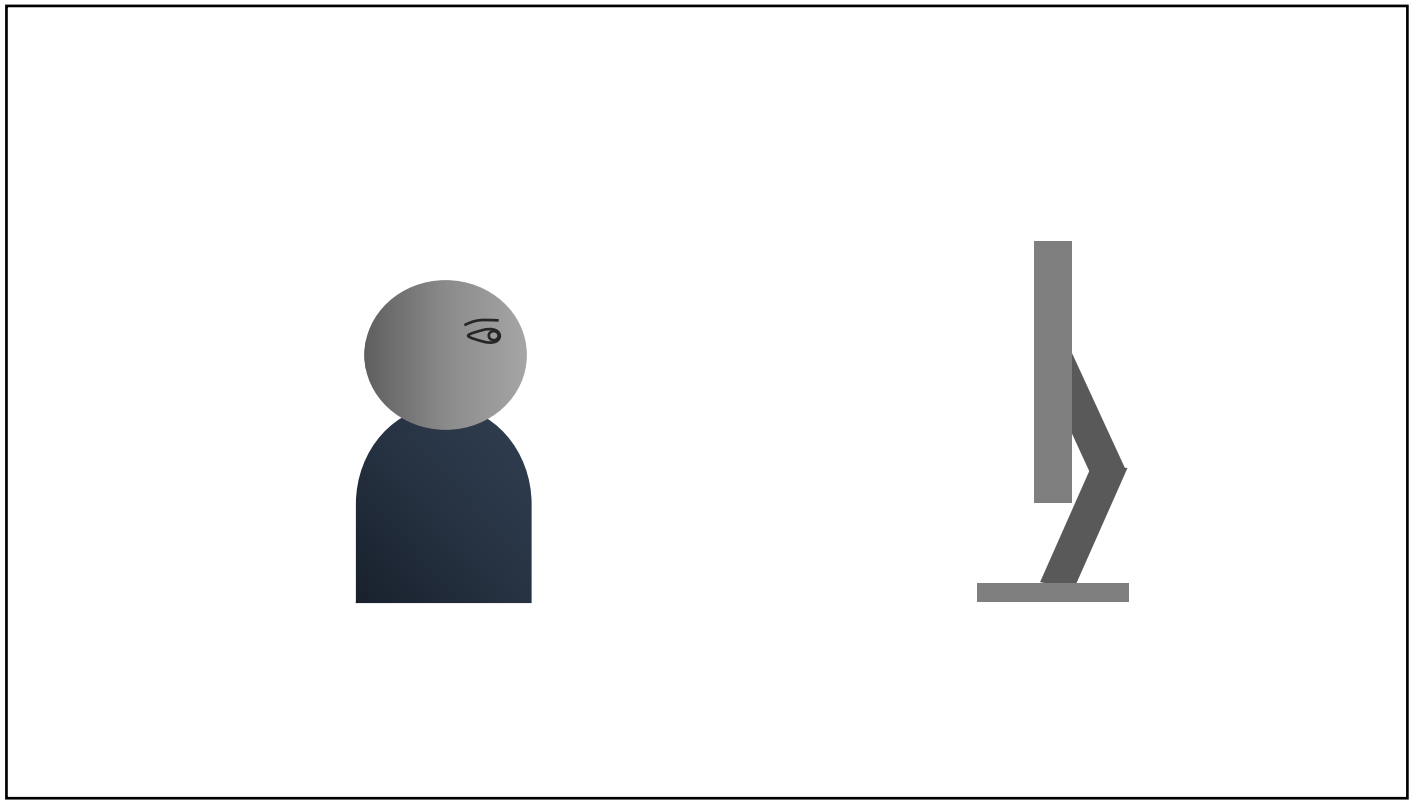
Geometric Assumption

Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption

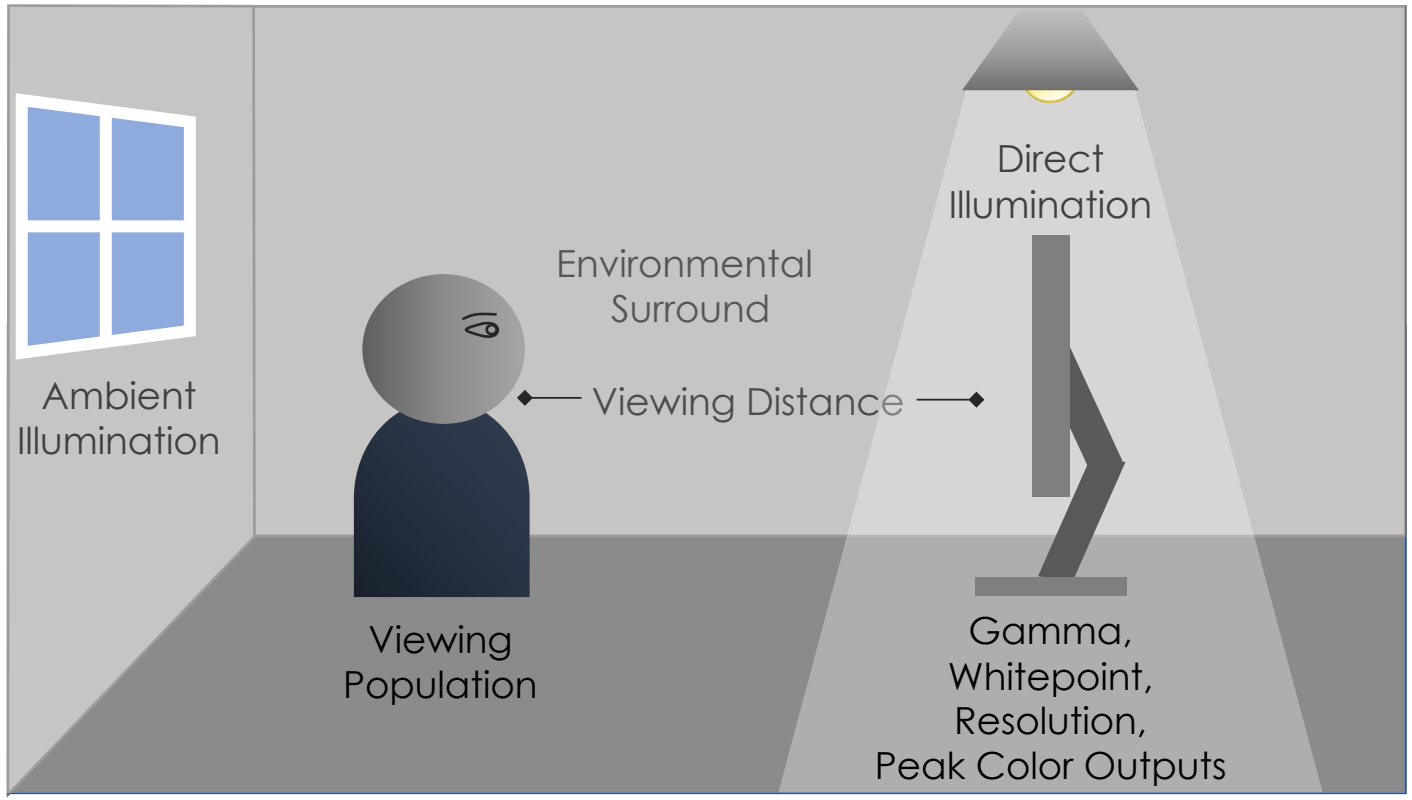


Visualizations violate three CIELAB assumptions

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

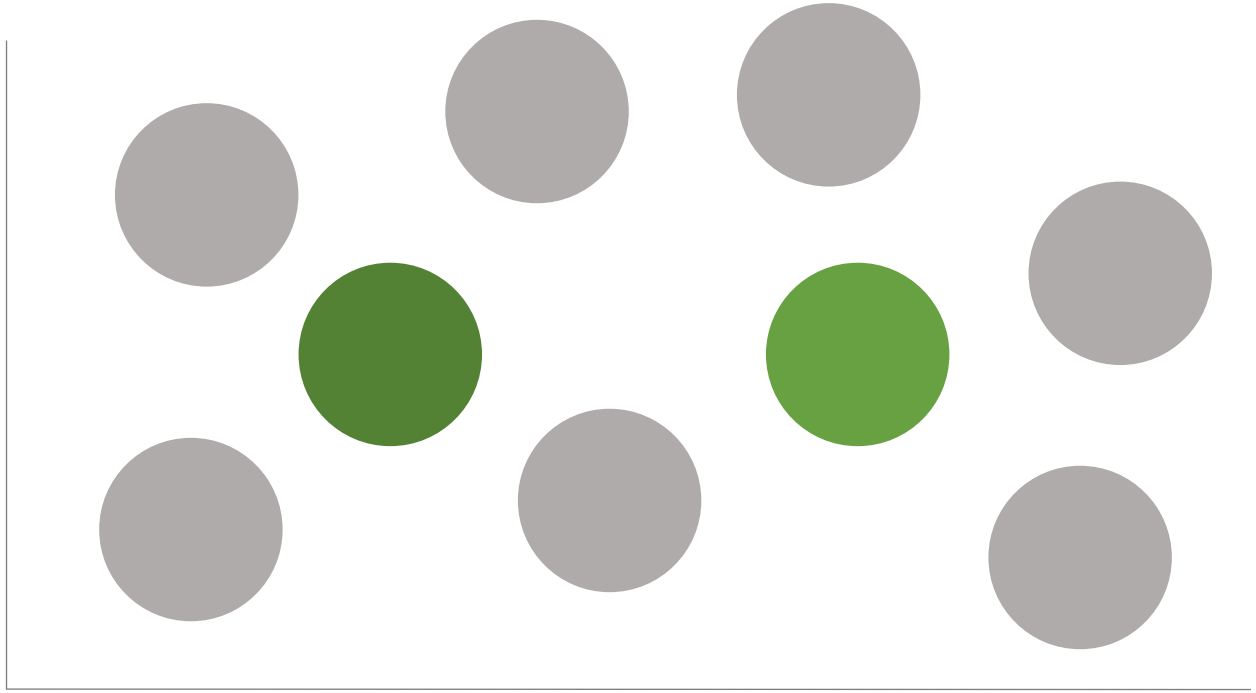
Geometric Assumption

Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption

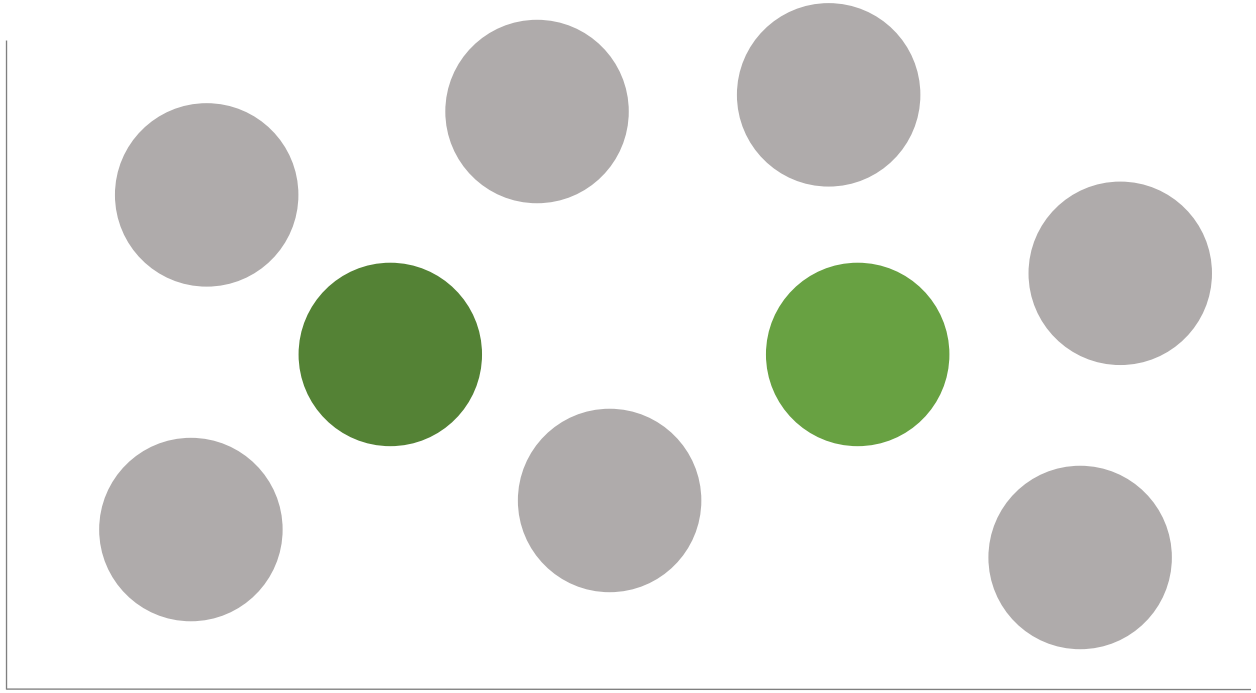


Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption

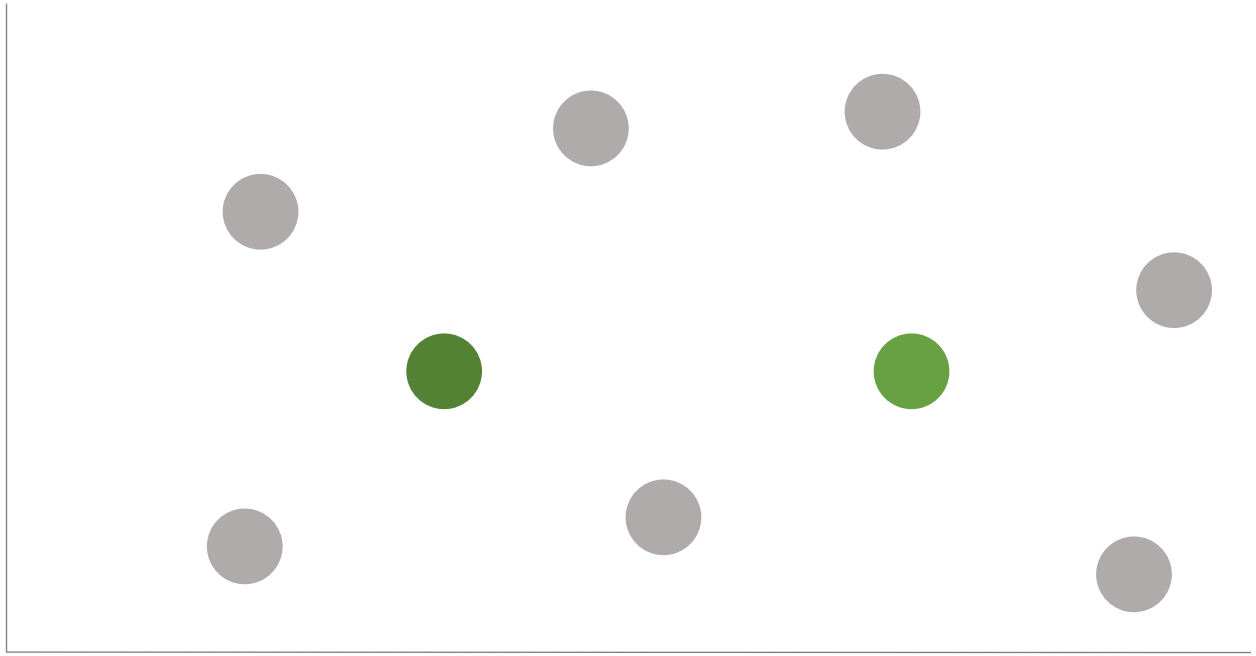


Visualizations violate three CIELAB assumptions

Simple World Assumption

Isolation Assumption

Geometric Assumption



Size-Based Sampling

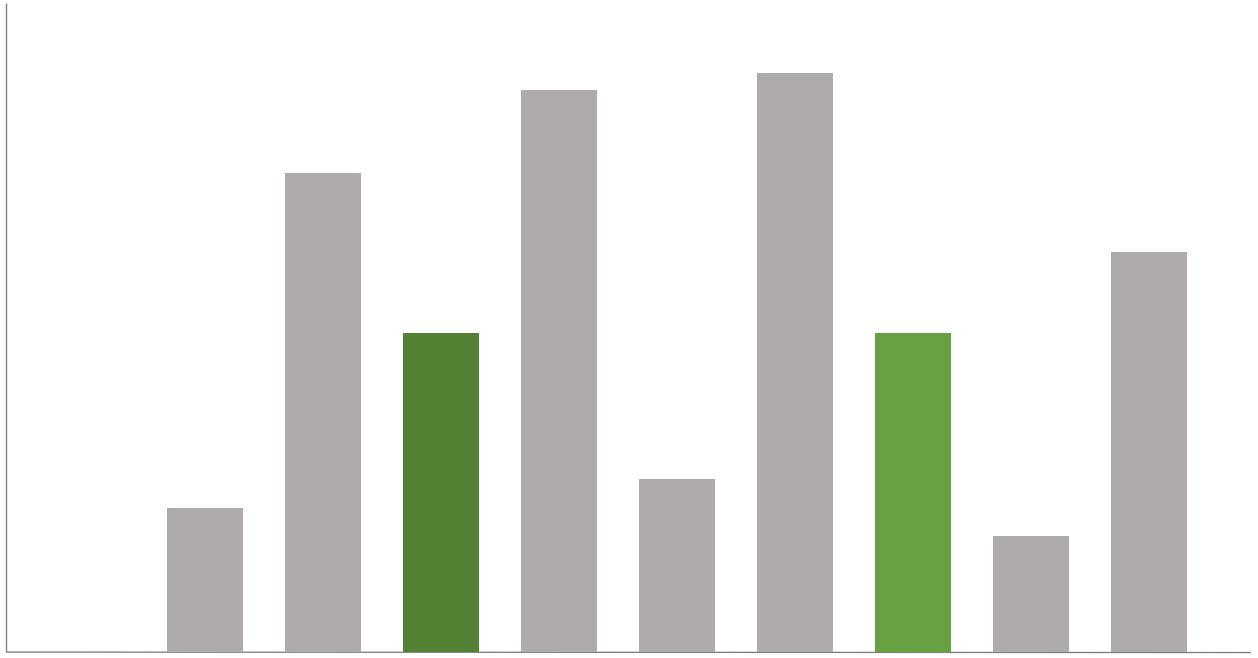
Carter & Silverstein, 2010
Stone, Szafir, & Setlur, 2014

Visualizations violate three CIELAB assumptions

Simple World Assumption

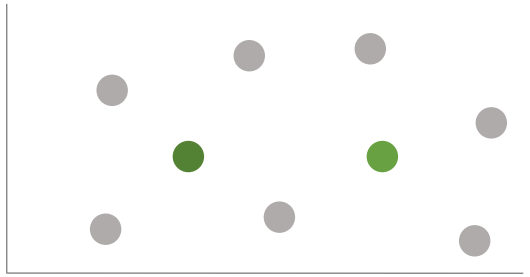
Isolation Assumption

Geometric Assumption

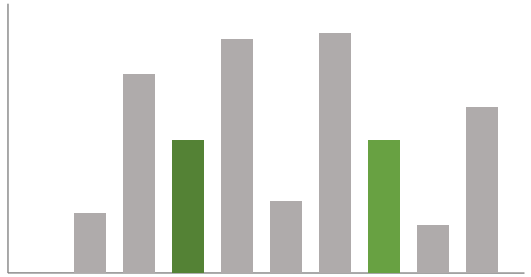


Visualizations violate three CIELAB assumptions

Simple World Assumption



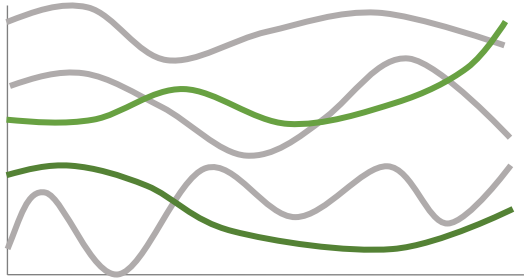
Diagonally Symmetric Marks



Elongated Marks

Isolation Assumption

Geometric Assumption

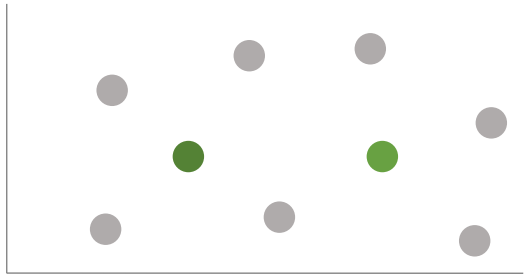


Asymmetric Marks

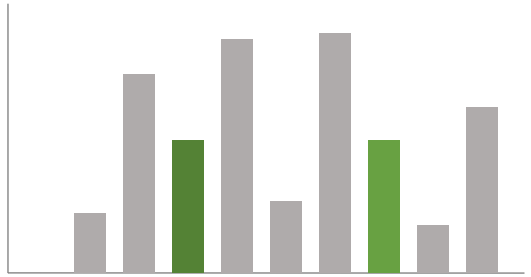
Area Marks

Visualizations violate three CIELAB assumptions

Simple World Assumption



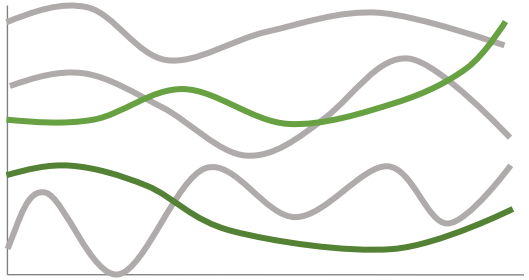
Diagonally Symmetric Marks



Elongated Marks

Isolation Assumption

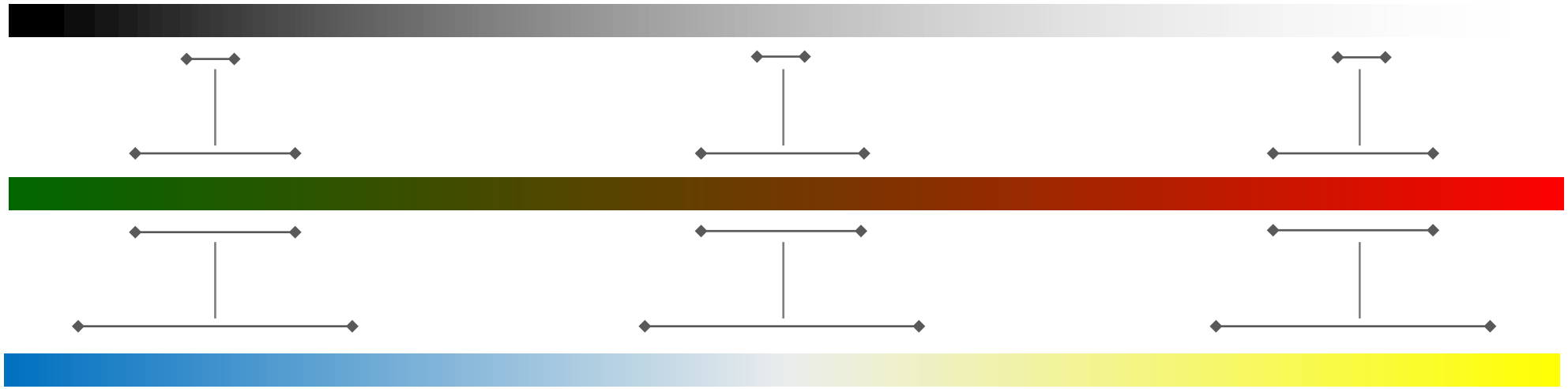
Geometric Assumption



Asymmetric Marks

Area Marks

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% JND***

Szafir, Stone, & Gleicher, 2014
Stone, Szafir, & Setlur, 2014

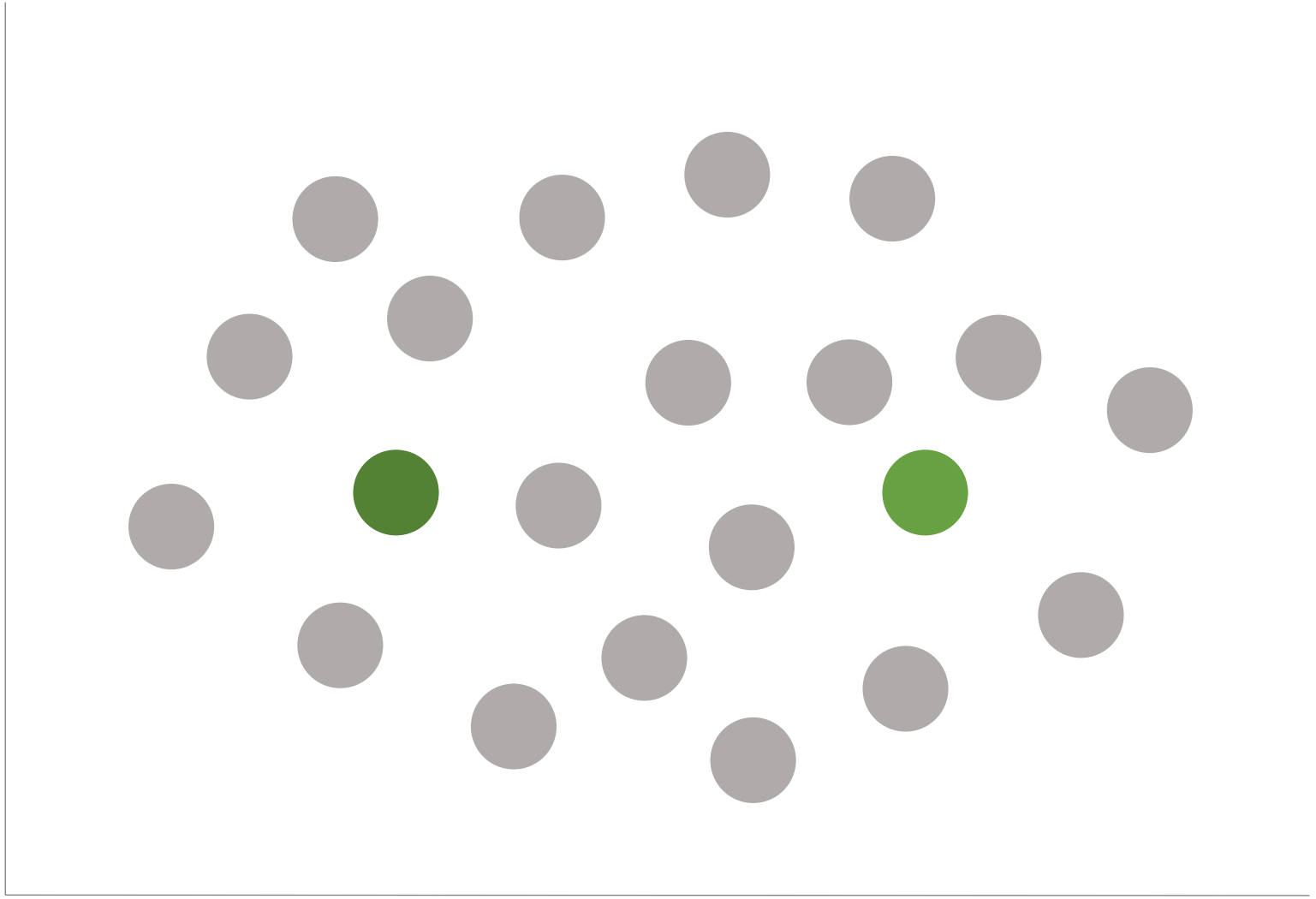
Factors for Color Difference in Visualization

Scatterplots – Diagonally Symmetric

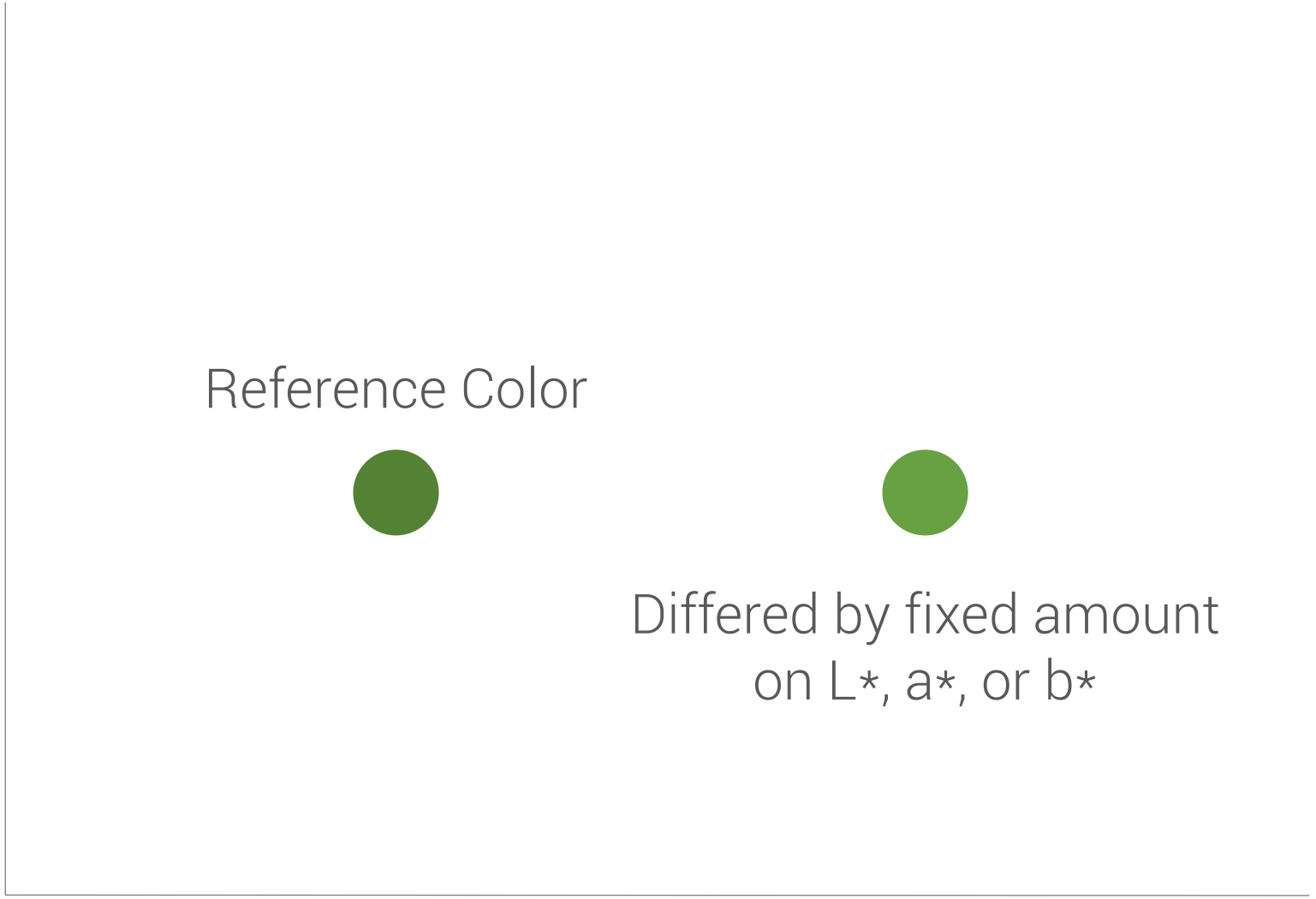
Bar Charts

Line Graphs

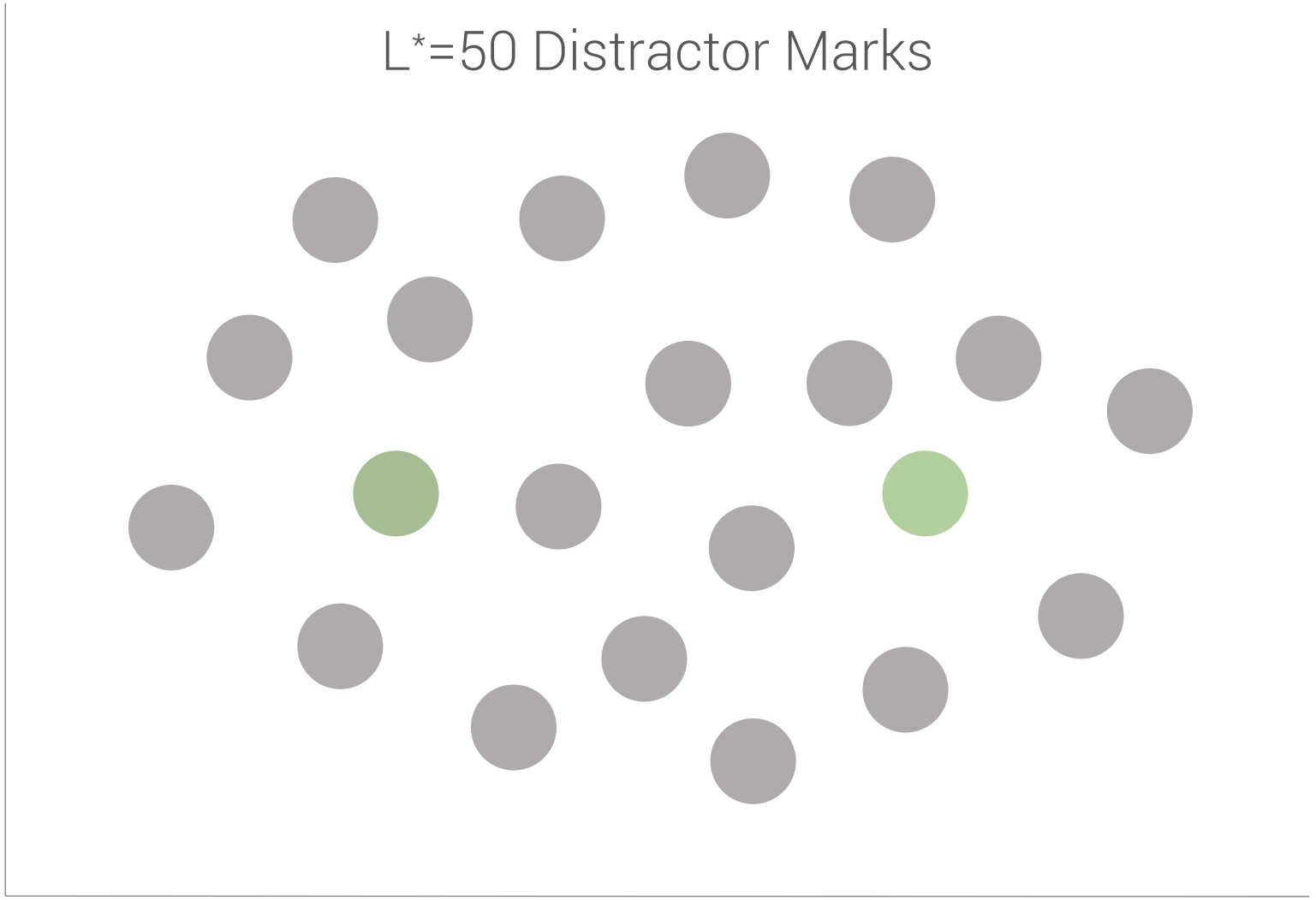
Summary & Applications



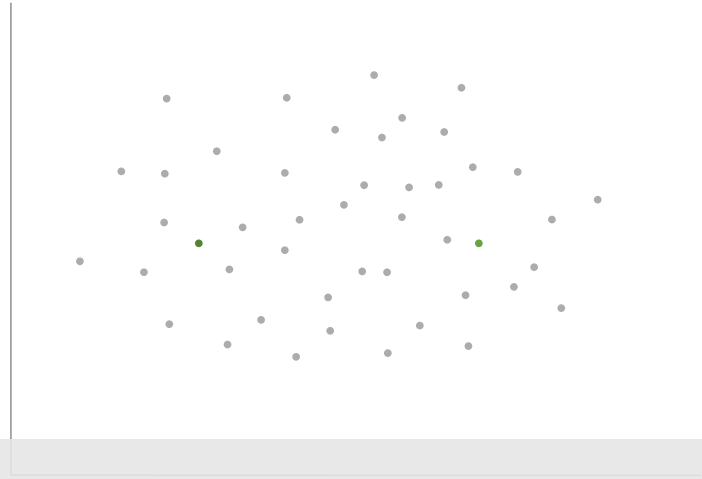
Do the colorful marks appear the same or different?

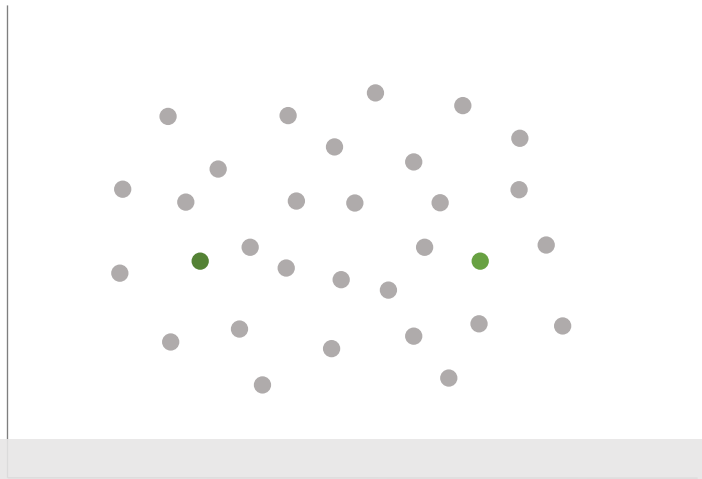


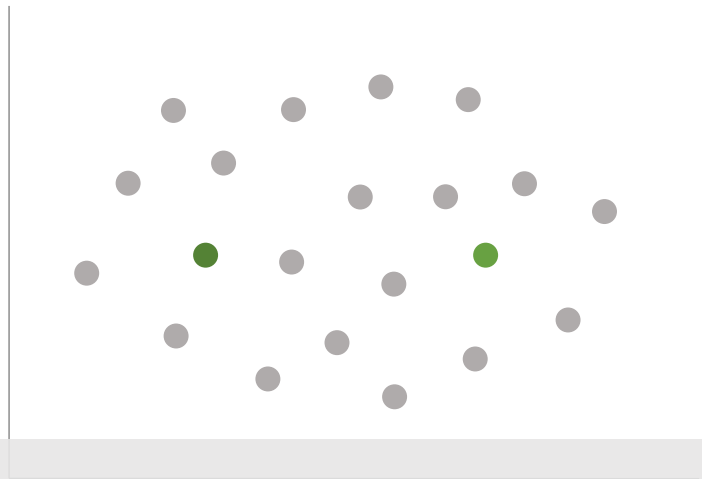
79 reference colors
36 total color differences



Random Gaussian with Overdraw Removed

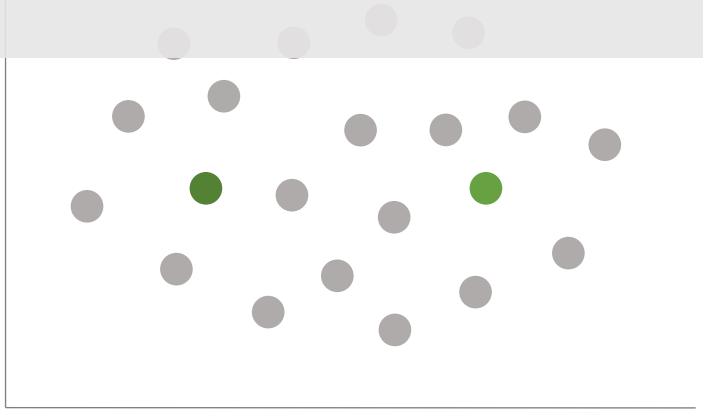




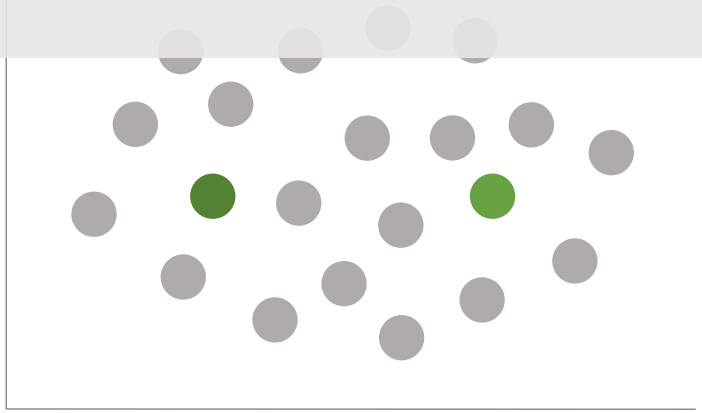


6 (diameters, within) × 6 (color differences, within) × 3 (color axis, between)

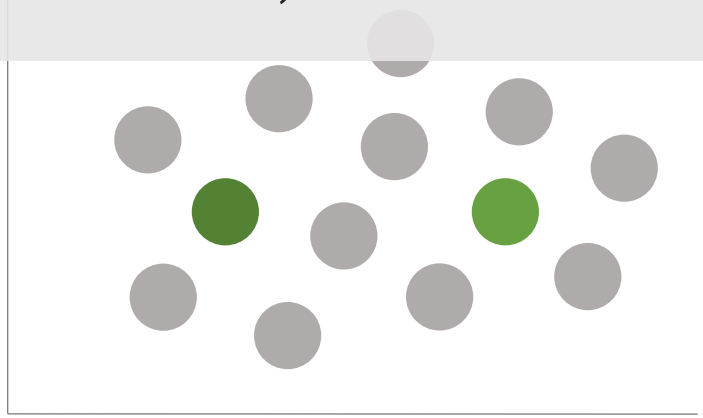
81 participants on Mechanical Turk (5,668 trials)



25 pixels
1.0°



37 pixels
1.5°



50 pixels
2.0°

Factor Analysis:

ANCOVA with question order and source color as covariates

Modeling Process:

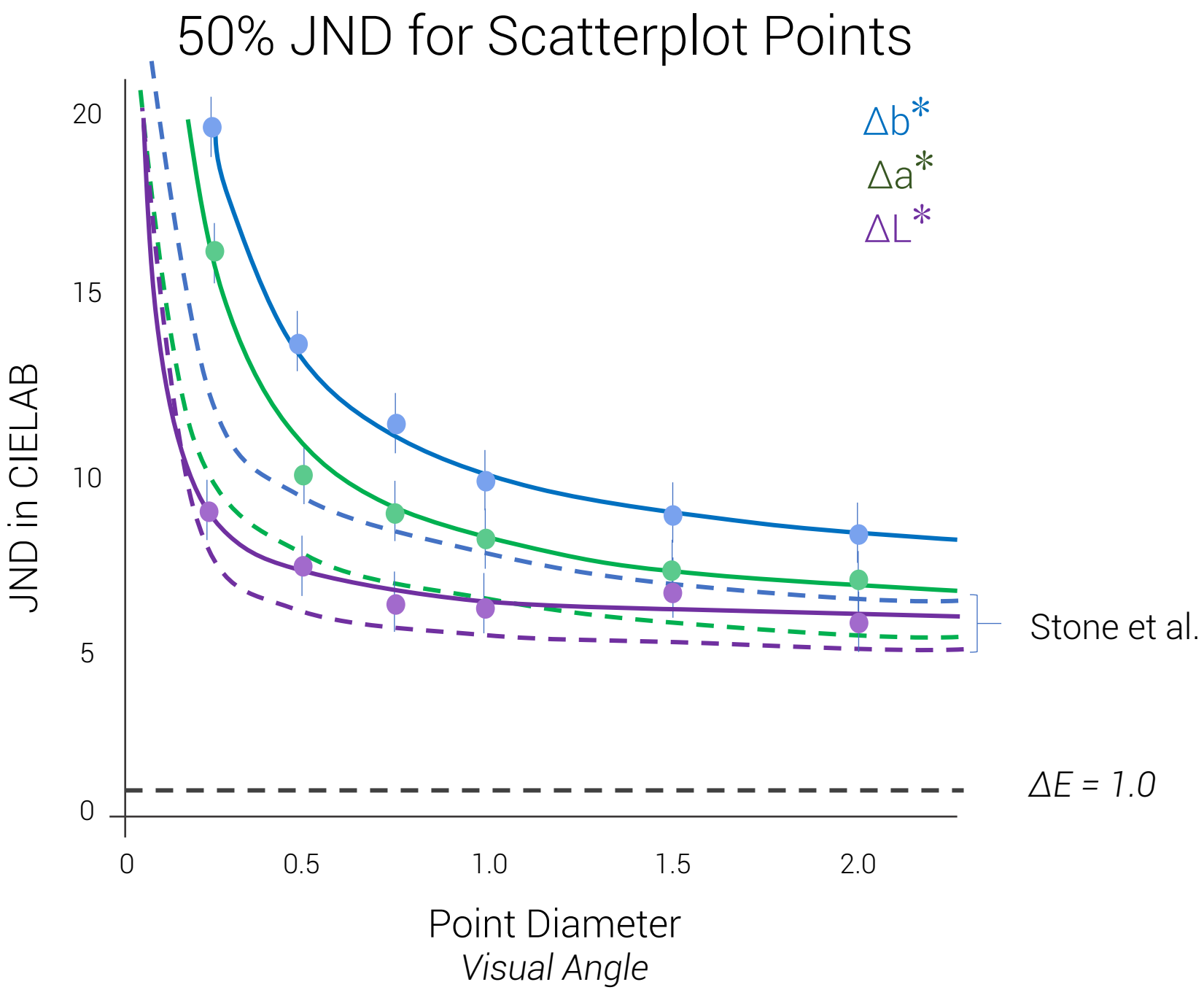
Compute rate of perceived differences for size x color difference

Model rates using linear regression fit to origin controlling for covariates

Sample linear models for desired probabilities

Fit logistic regression to the samples

Szafir, Stone, & Gleicher, 2014
Stone, **Szafir**, & Setlur, 2014


50% JND for Scatterplot Points
20
15
10
5
0
JND in CIELAB
Δb*
Δa*
ΔL*
Stone et al.
ΔE = 1.0
0
0.5
1.0
1.5
2.0
Point Diameter
Visual Angle

$$ND_L\,(50\%, s) = 5.095 + \frac{0.80}{s}, R^2 = .93$$

$$ND_a\,(50\%, s) = 5.089 + \frac{2.69}{s}, R^2 = .99$$

$$ND_b\,(50\%, s) = 6.786 + \frac{3.20}{s}, R^2 > .99$$

$$\Delta E_p = \sqrt{\left(\frac{\Delta L}{ND_L(p)}\right)^2 + \left(\frac{\Delta a}{ND_a(p)}\right)^2 + \left(\frac{\Delta b}{ND_b(p)}\right)^2}$$

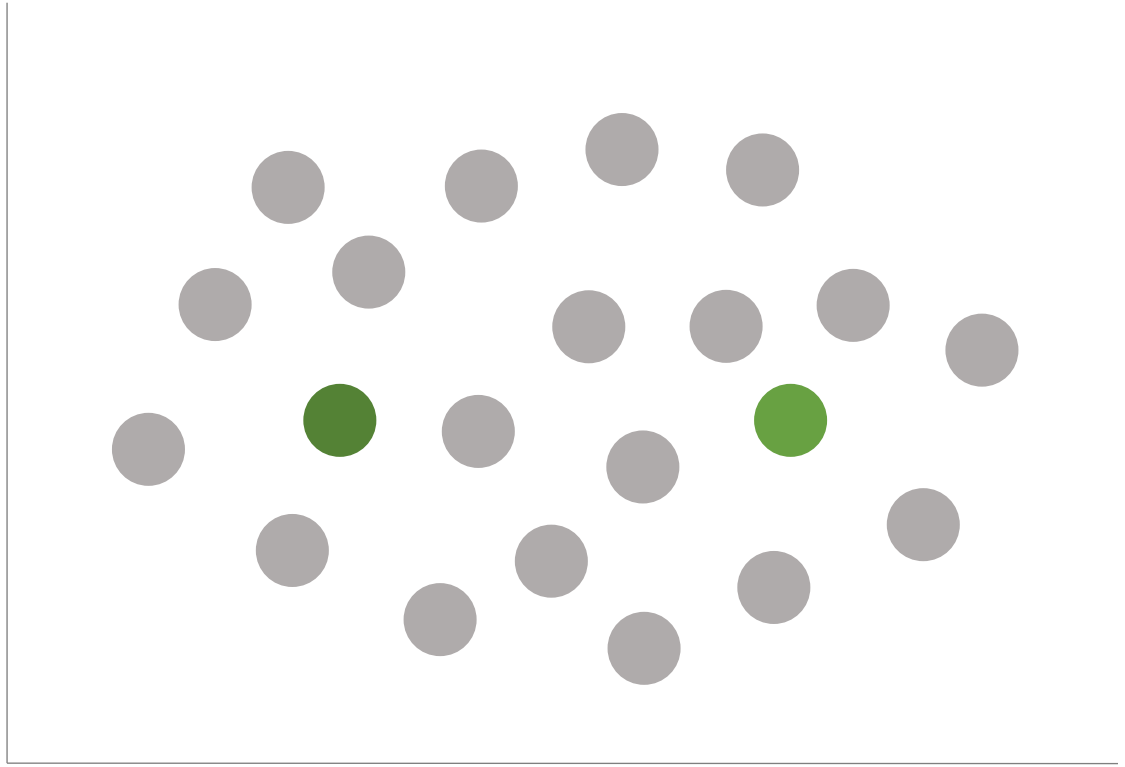
Takeaways—Diagonally Symmetric Points

7x larger than existing models

Vary with inversely size

Replicate Stone et al., 2014

Increased by distractor points



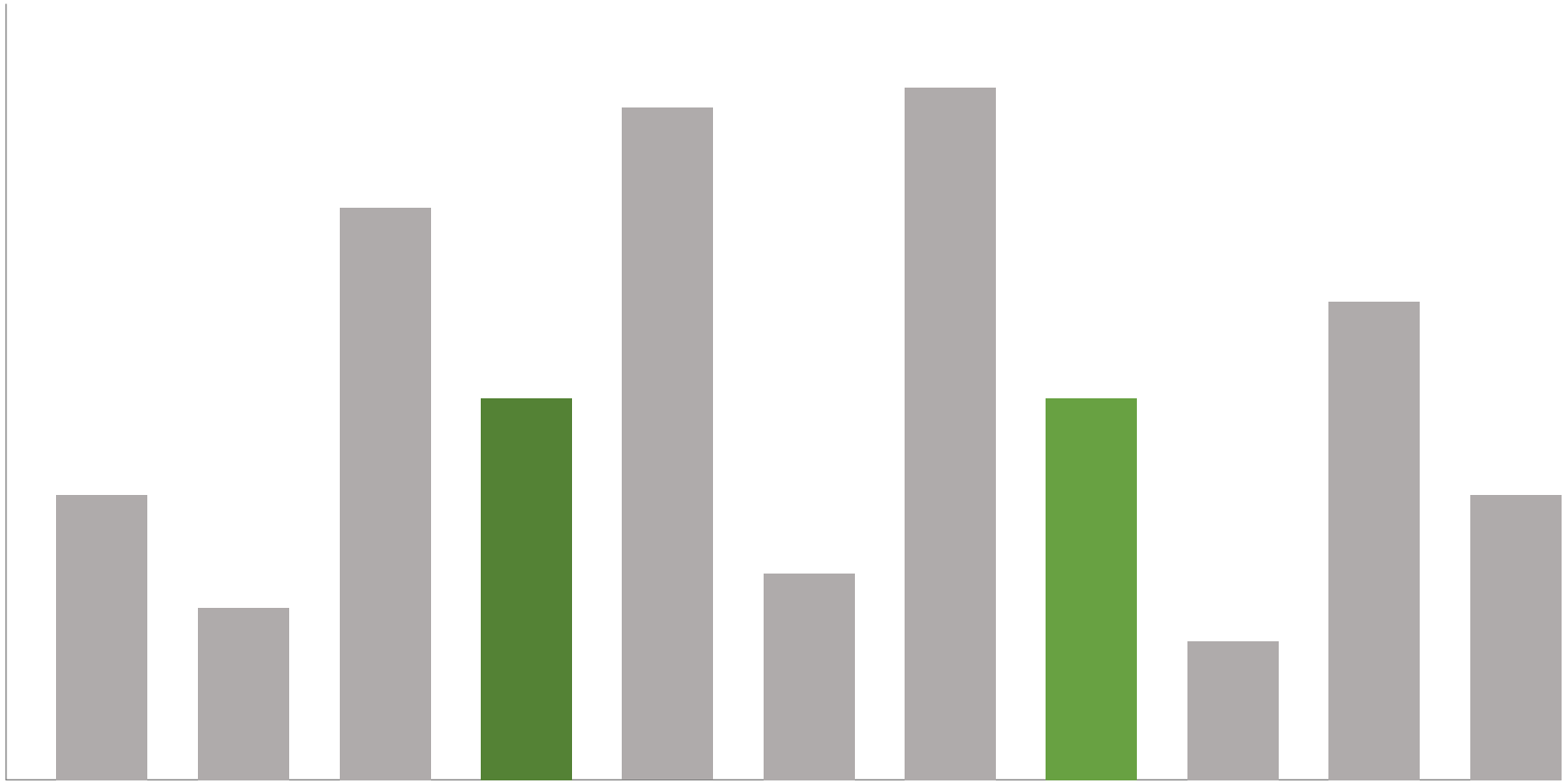
Factors for Color Difference in Visualization

Scatterplots

Bar Charts – Elongated

Line Graphs

Summary & Applications

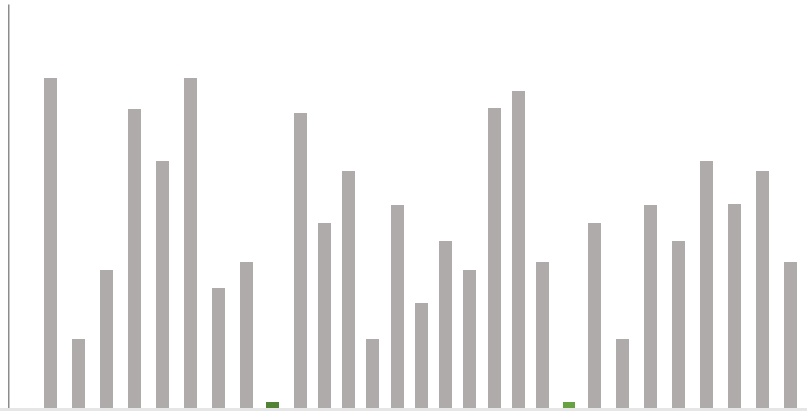


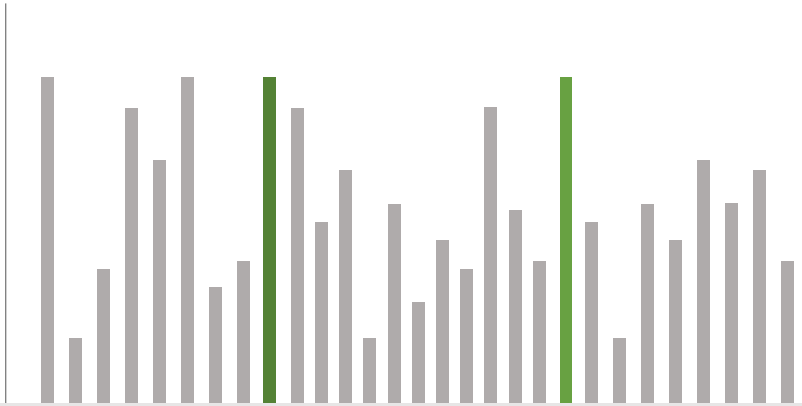
79 reference colors
36 total color differences

Mark size varies in two dimensions

6 thicknesses:

6 pixels … 50 pixels



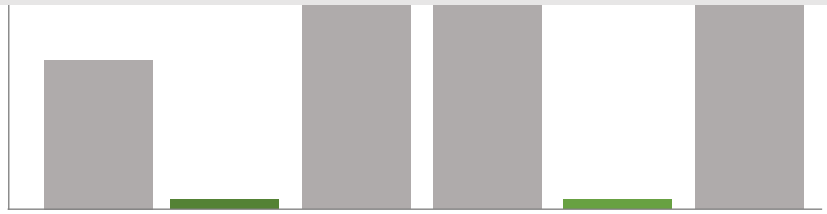


6 (thicknesses, blocked between) × 8 (lengths, blocked between) × 6 (color differences, within) × 3 (color axis, between)

301 participants on Mechanical Turk (22,752 trials)

3 pixels … 150 pixels

0.125° *6.0°*

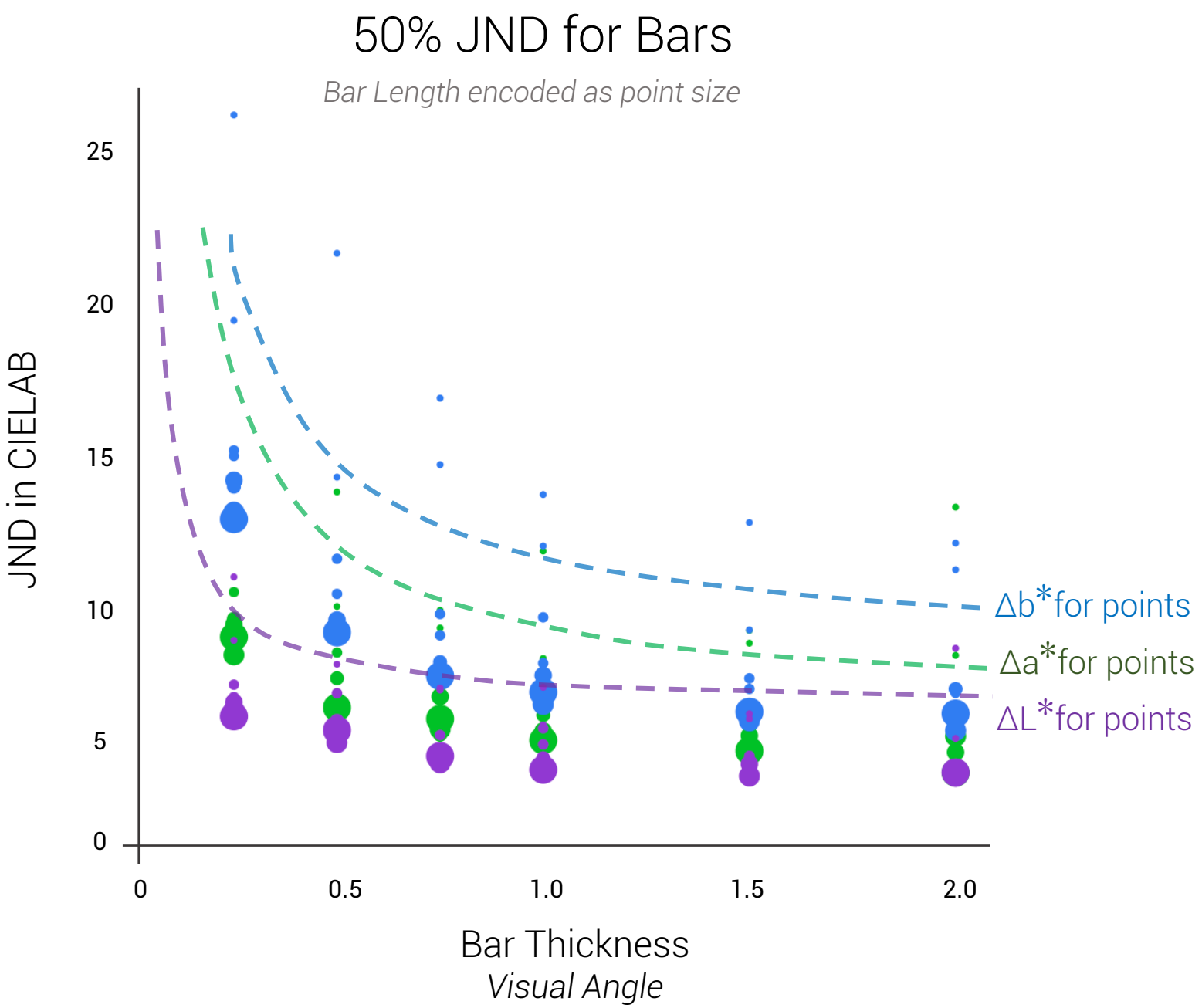


50 pixels x 3 pixels

2.0° *0.125°*

50 pixels x 150 pixels

2.0° *6.0°*



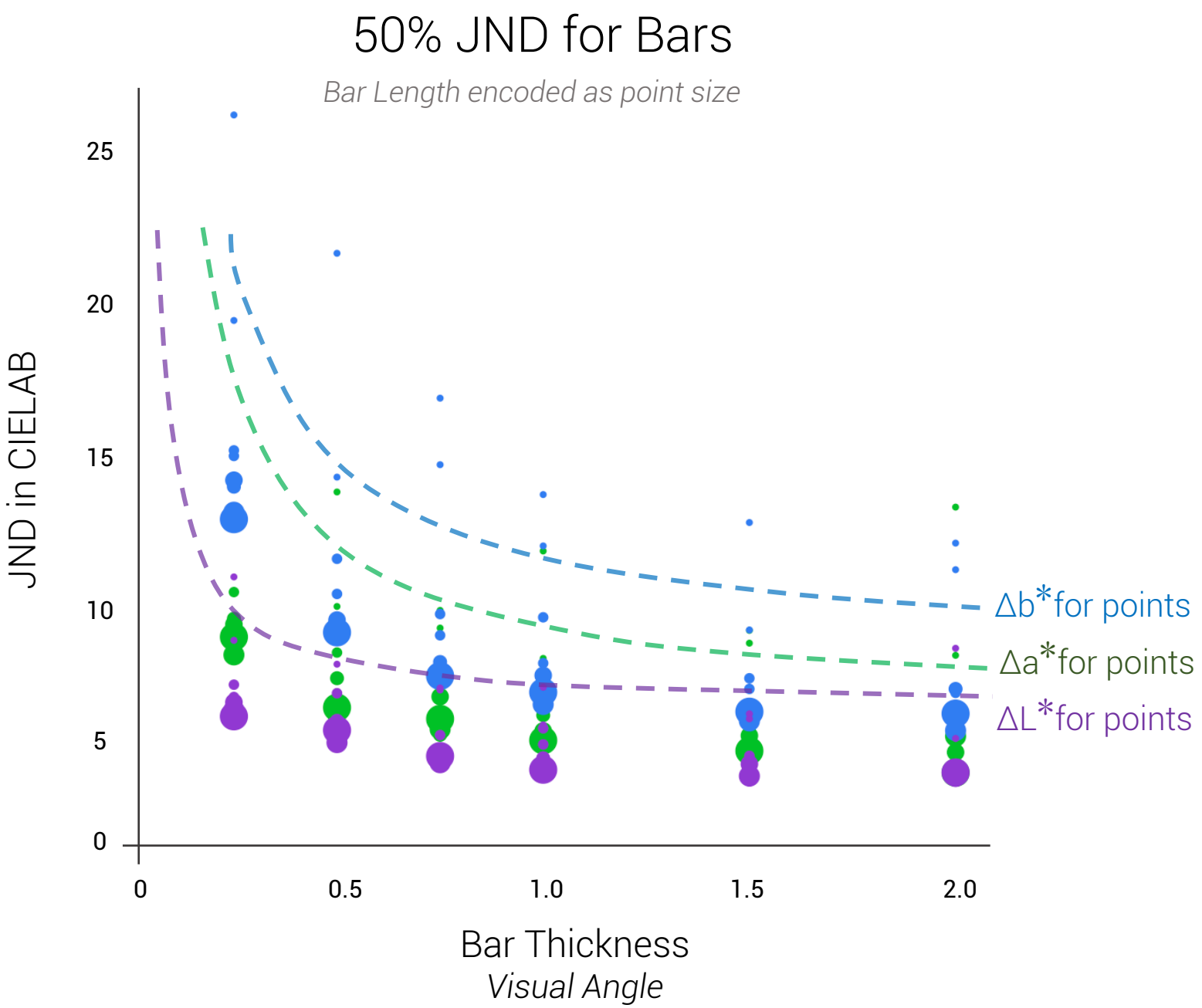
Accounting for Gains:

Longest Edge

Shortest Edge

Area = longest edge x shortest edge

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



Accounting for Gains:

Longest Edge

Shortest Edge

Area = longest edge x shortest edge

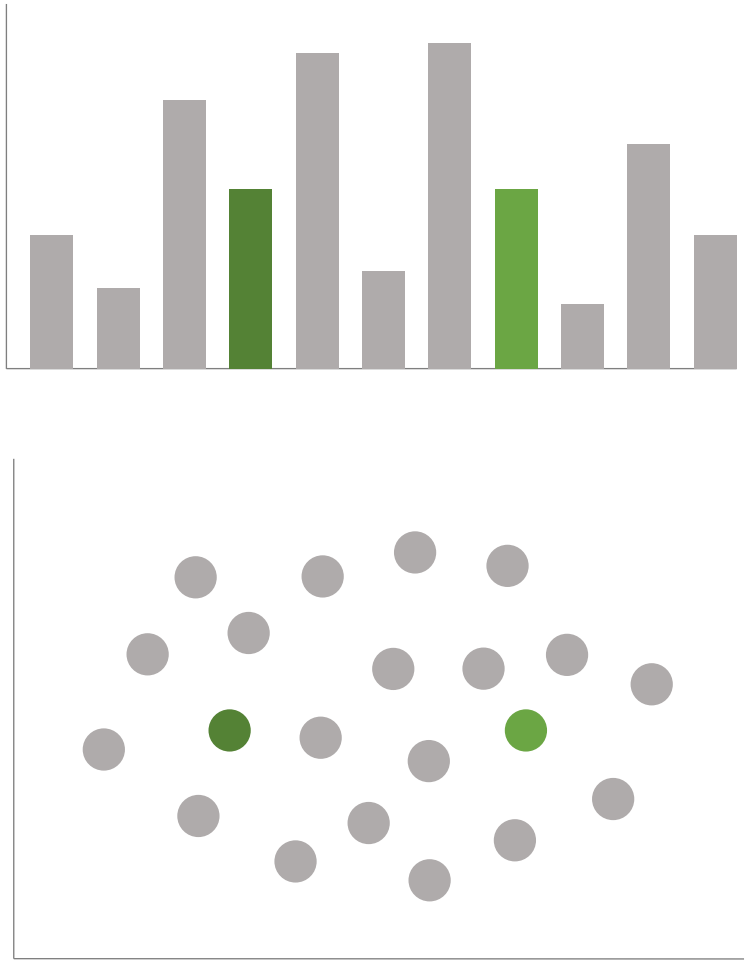
Elongation = $\frac{\textbf{longest } edge}{\textbf{shortest } edge}$

Takeaways—Elongated Marks

Vary with bar length & thickness

Predicting data perceptions by thickness gives conservative model

Gains over points are asymptotic based on elongation



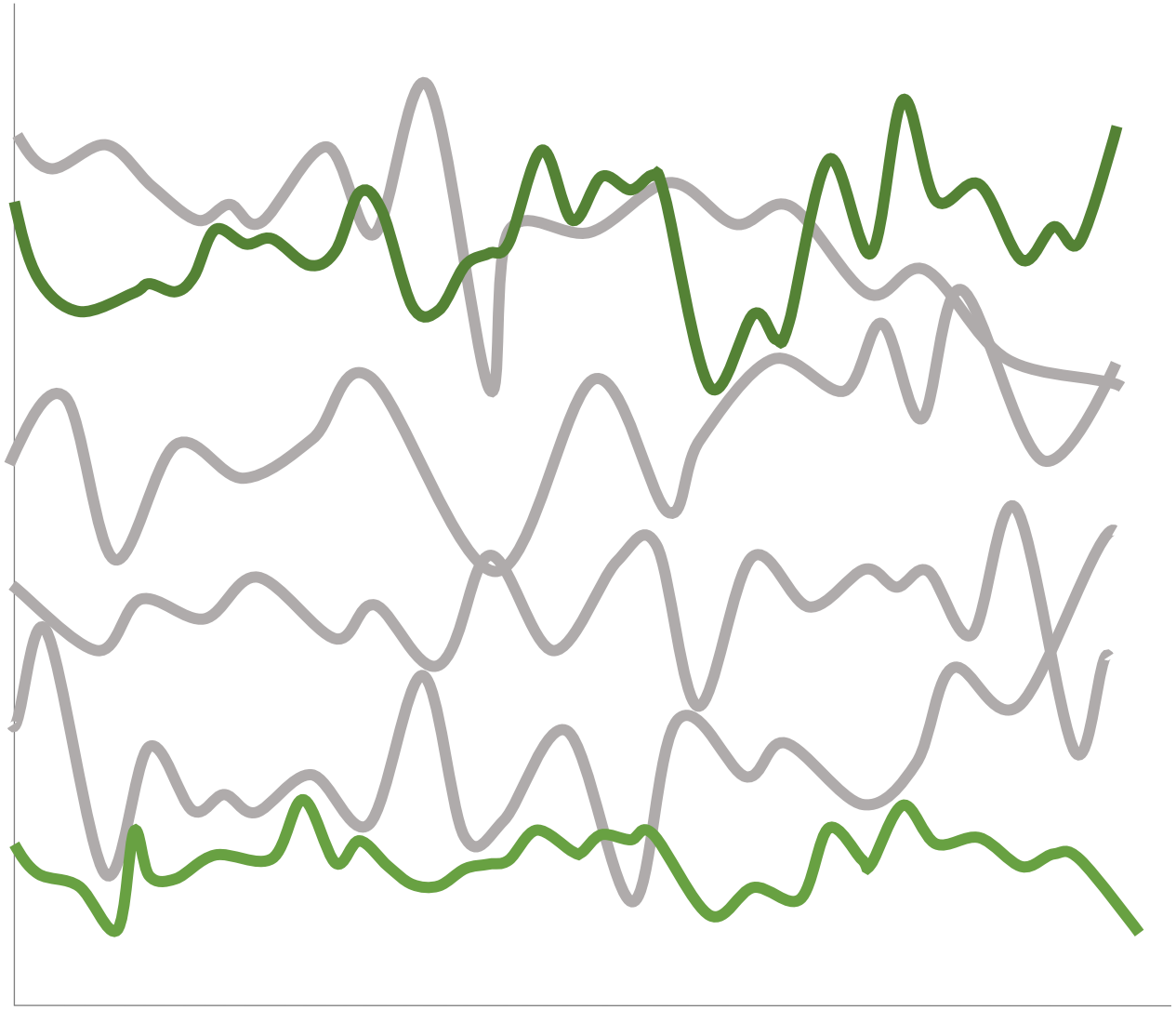
Factors for Color Difference in Visualization

Scatterplots

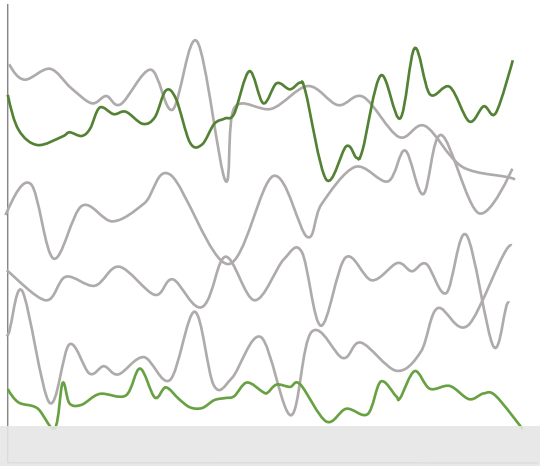
Bar Charts

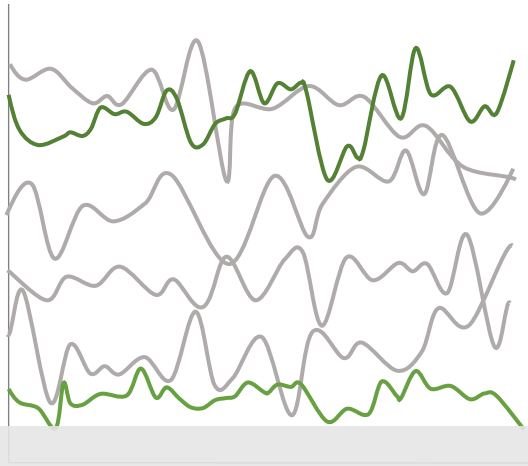
Line Graphs – Asymmetric

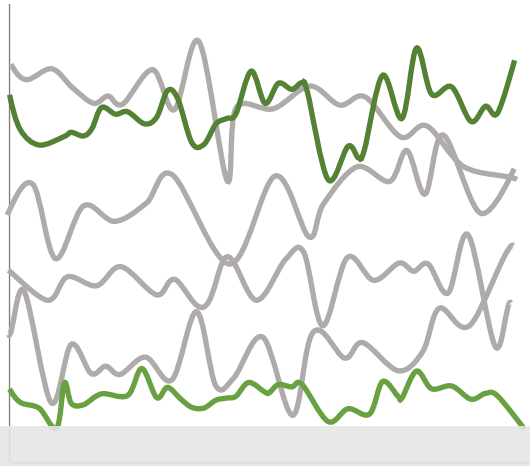
Summary & Applications



Do the colorful marks appear the same or different?

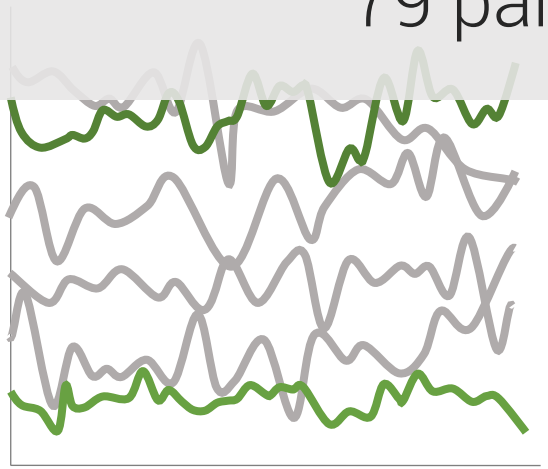




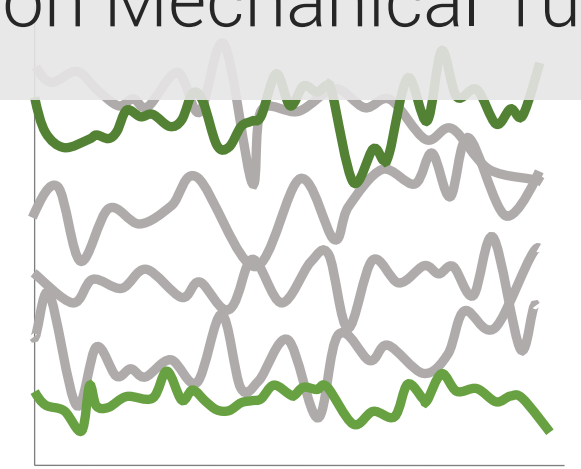


6 (thickness, within) × 6 (color differences, within) × 3 (color axis, between)

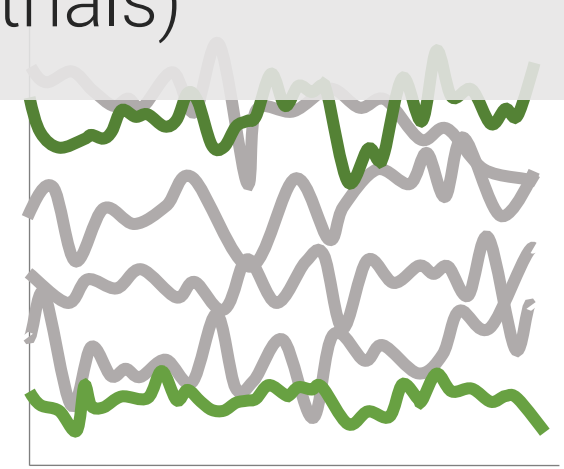
79 participants on Mechanical Turk (5,668 trials)



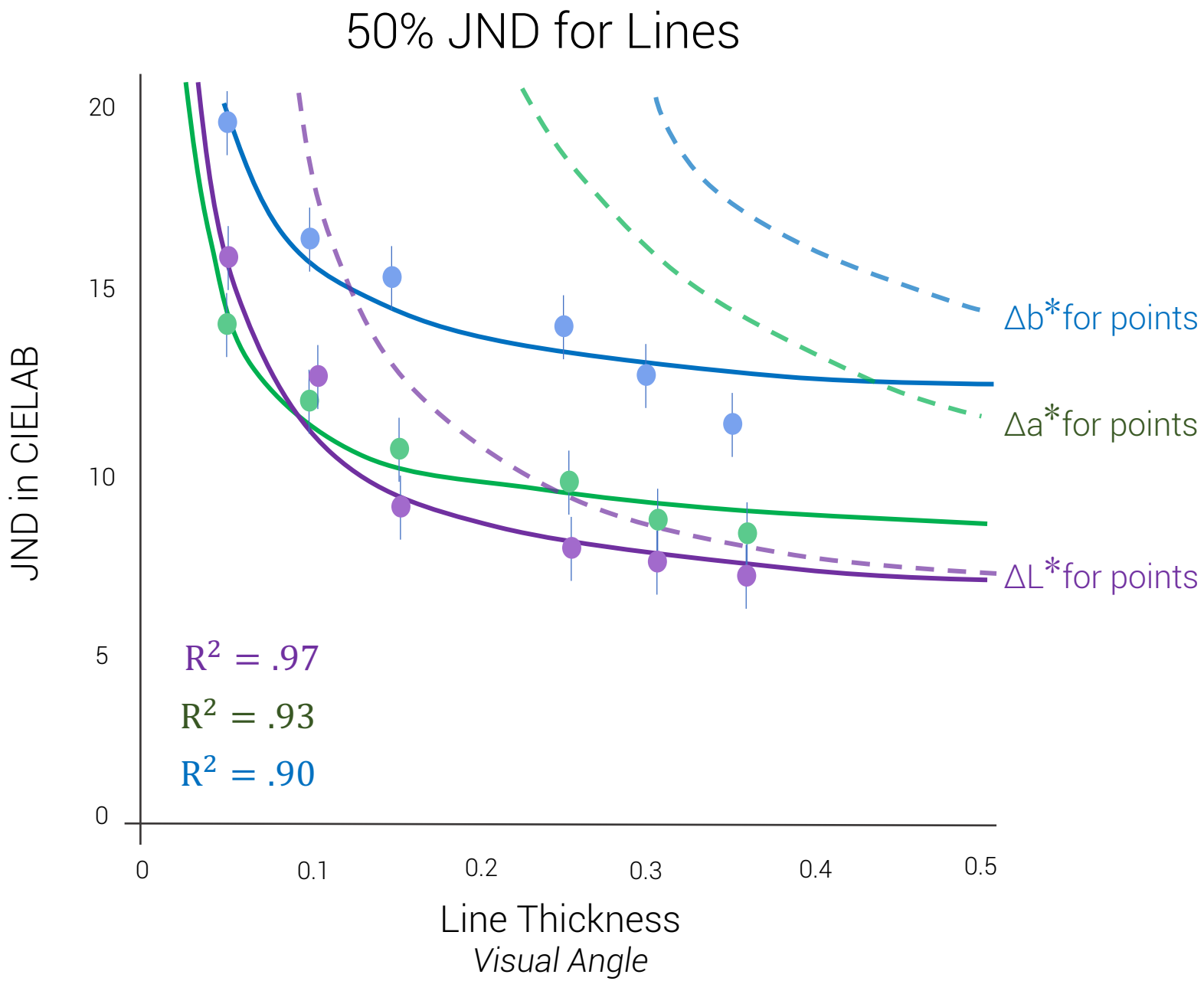
6 pixels
0.25°



7 pixels
0.30°



9 pixels
0.35°


50% JND for Lines
JND in CIELAB
20
15
10
5
0
0
0.1
0.2
0.3
0.4
0.5
Line Thickness
Visual Angle
Δb*for points
Δa*for points
ΔL*for points
R² = .97
R² = .93
R² = .90

Takeaways—Asymmetric Marks

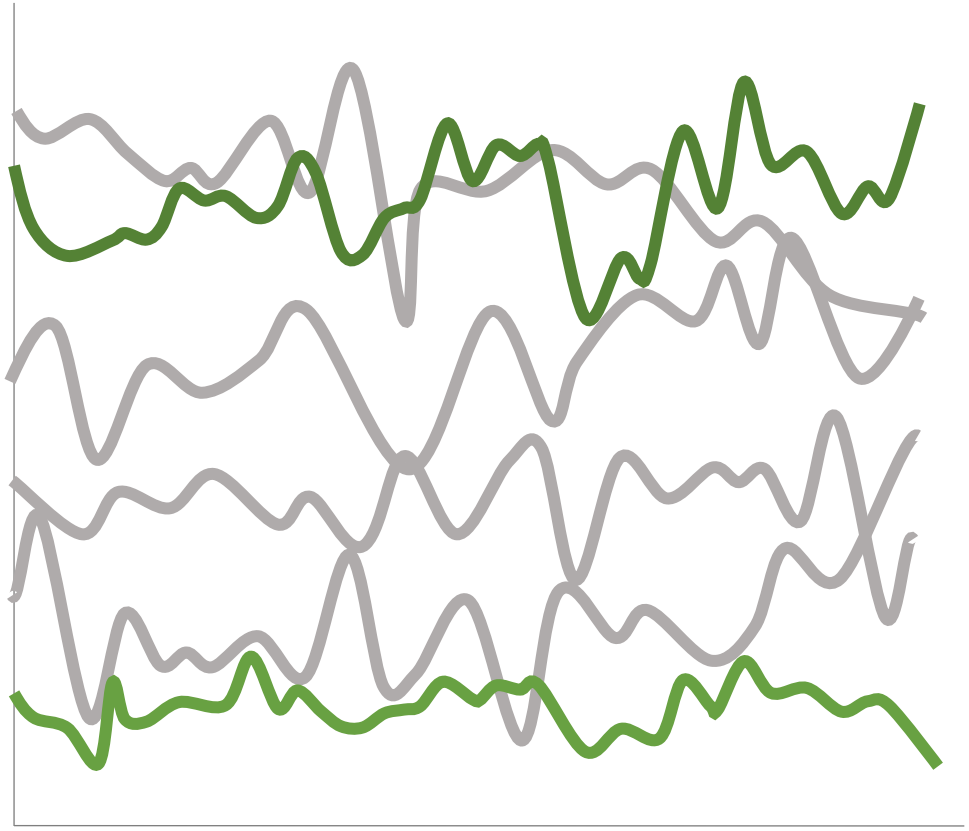
Vary with inversely with line thickness

Points are overly conservative for lines

Significant gains over points

16Δa JND for 6 pixel points*

9.4Δa JND for 6 pixel lines*



Factors for Color Difference in Visualization

Scatterplots

Bar Charts

Line Graphs

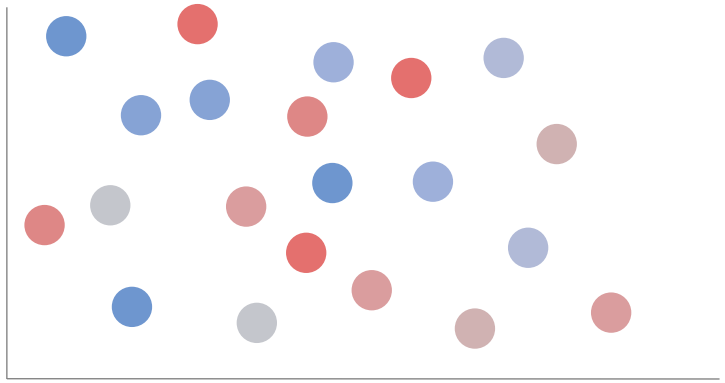
Summary & Applications

Factors Effecting Color Encodings

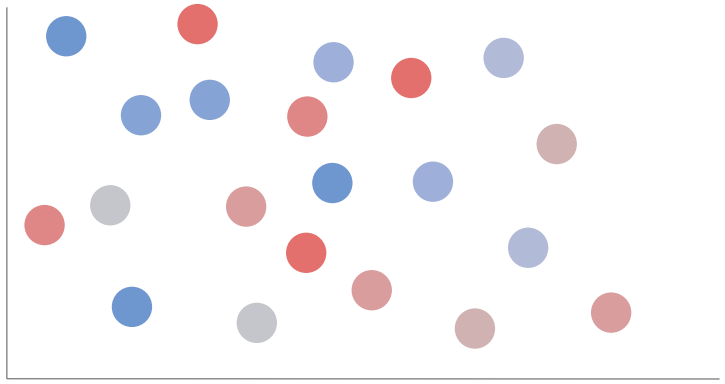
Simple World Assumption:
Viewing visualizations online introduces variation in data discrimination

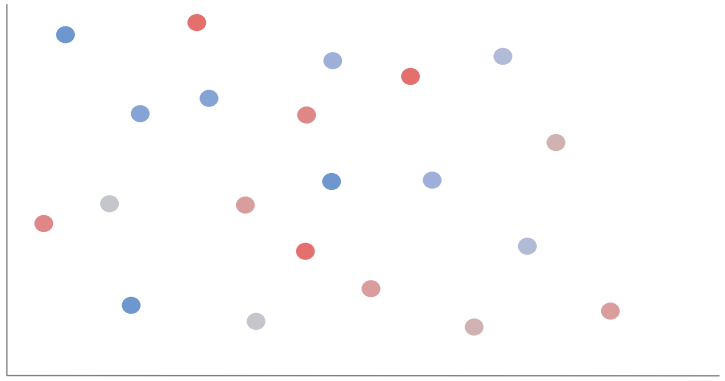
Isolation Assumption:
The presence of other points complicates data discrimination

Geometric Assumption:
Data discrimination varies inversely with mark size
Elongating marks increases data discrimination asymptotically

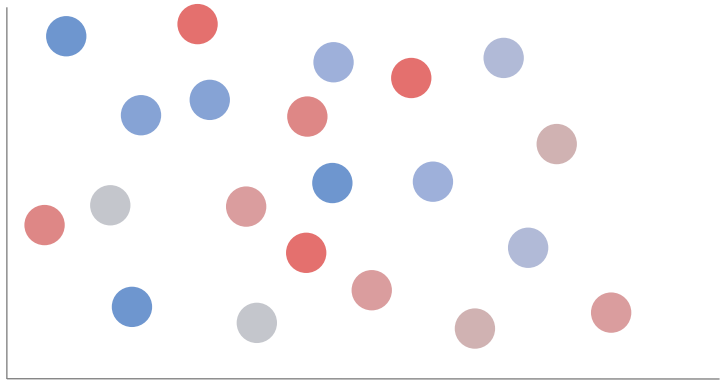


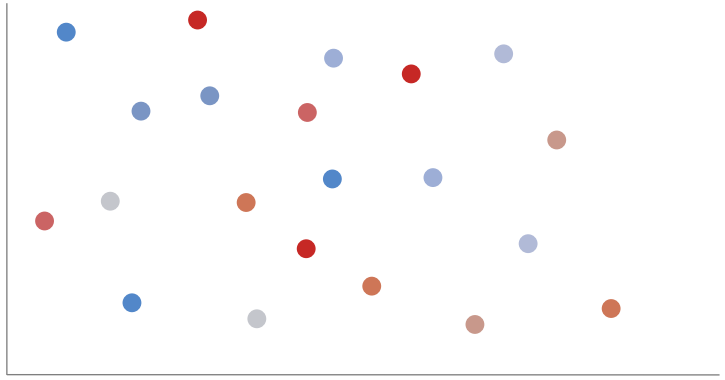
Guide Effective Designs



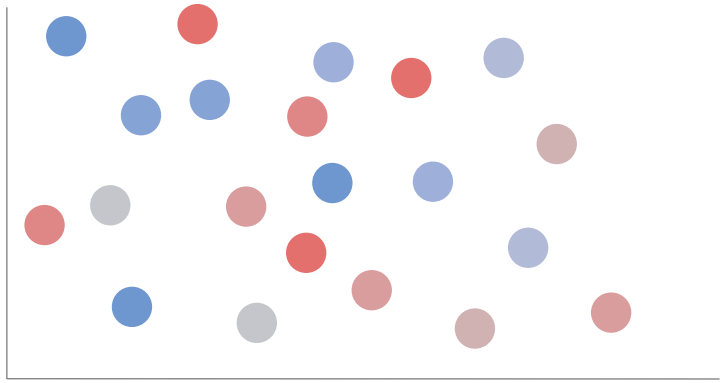


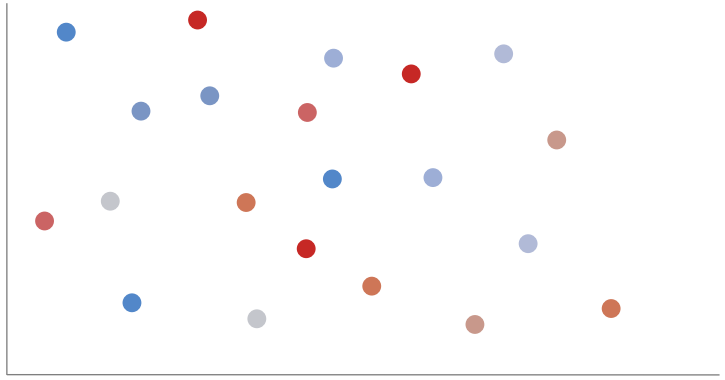
Guide Effective Designs



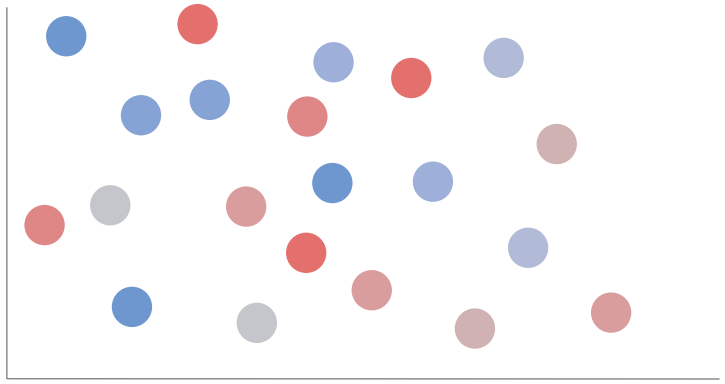


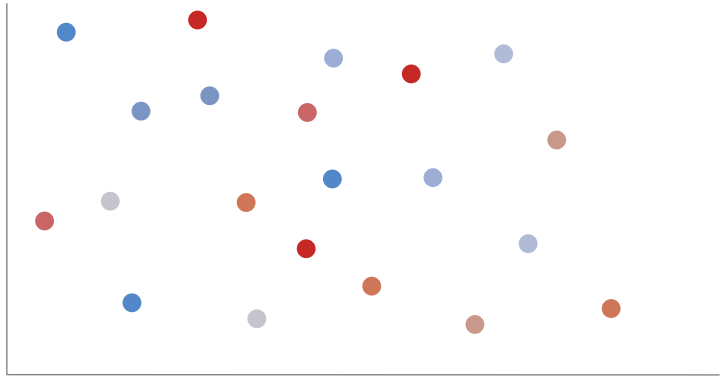
Guide Effective Designs

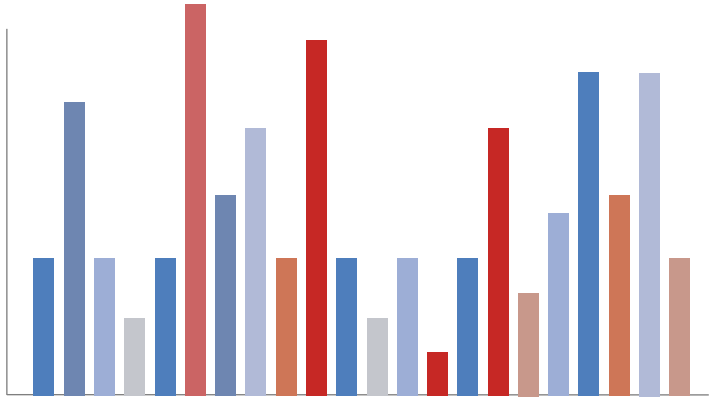




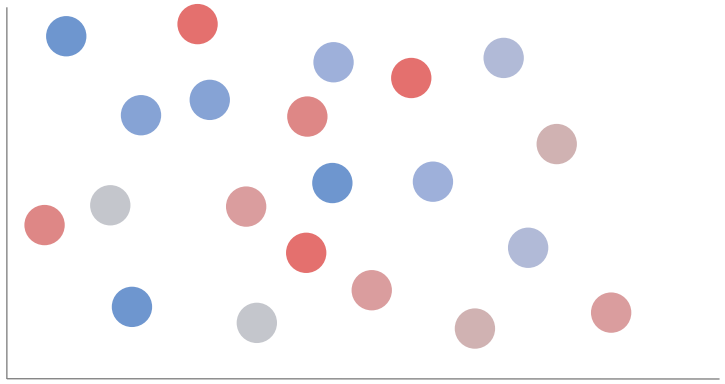
Guide Effective Designs

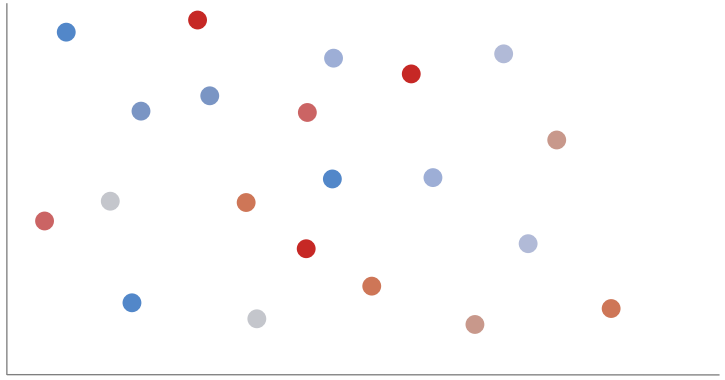


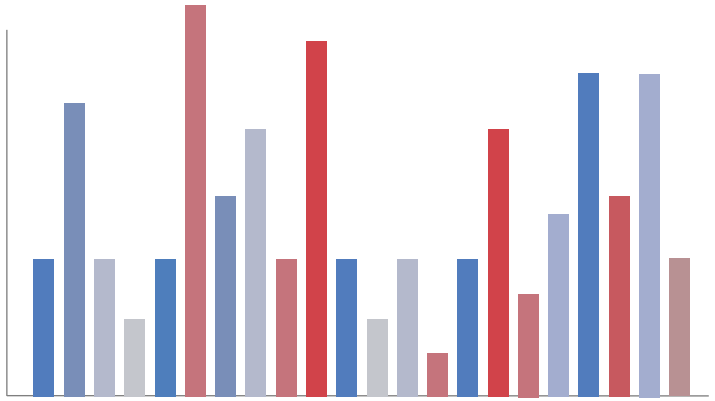




Guide Effective Designs

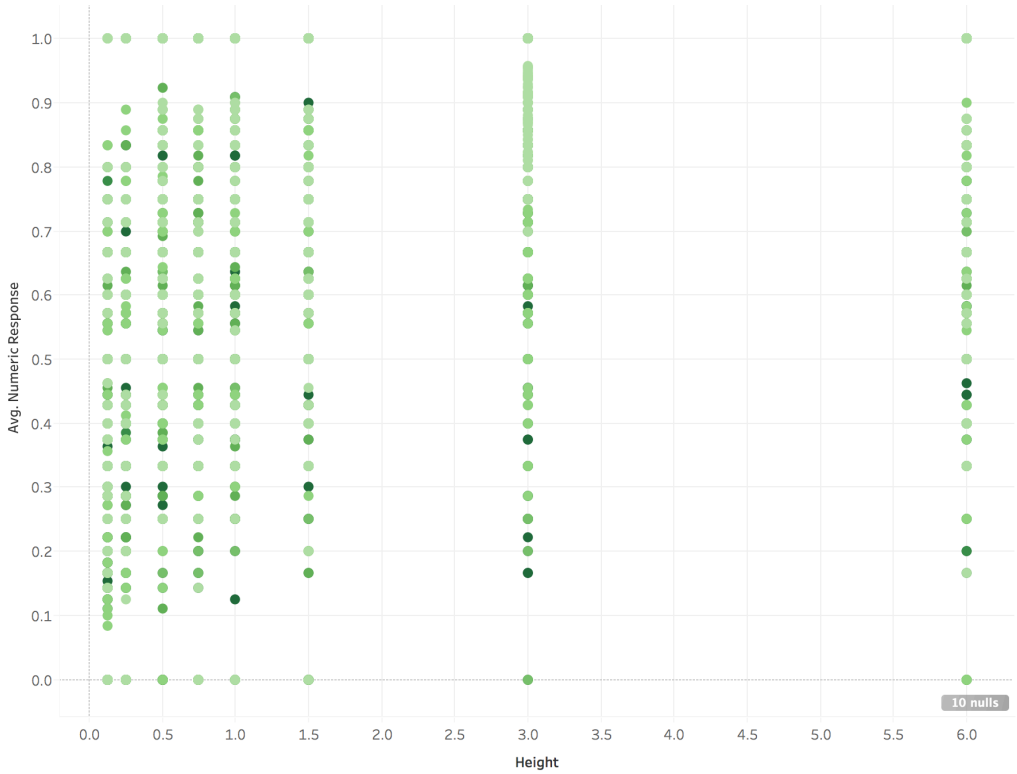


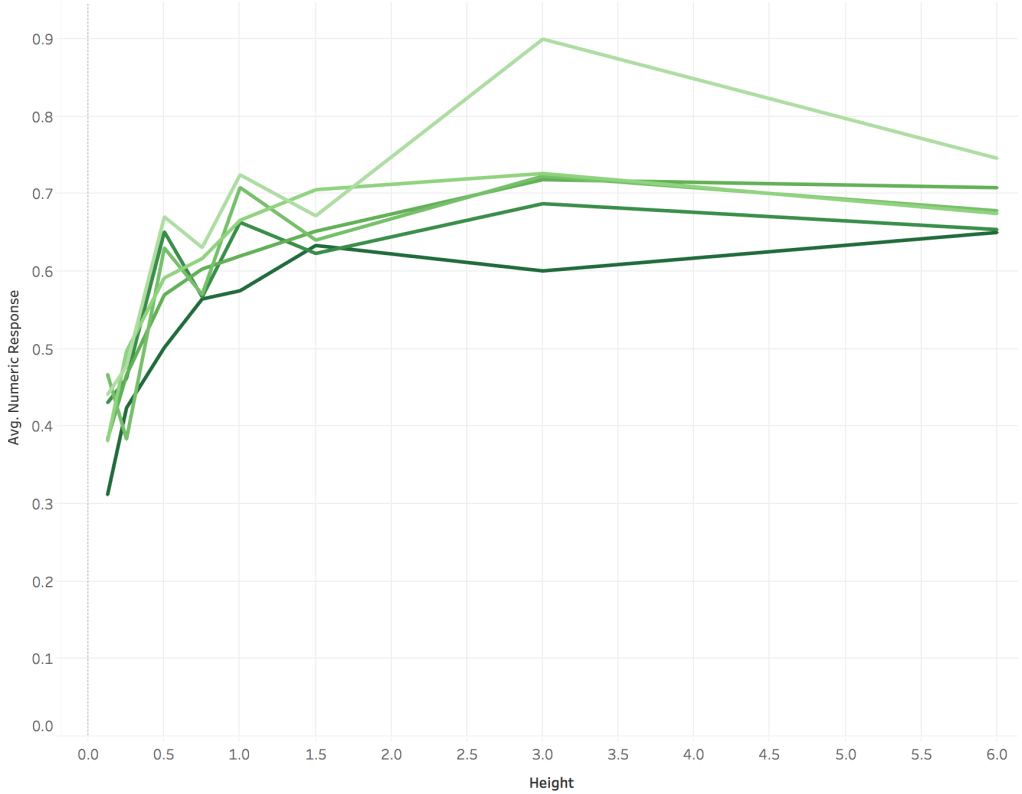




Guide Effective Designs

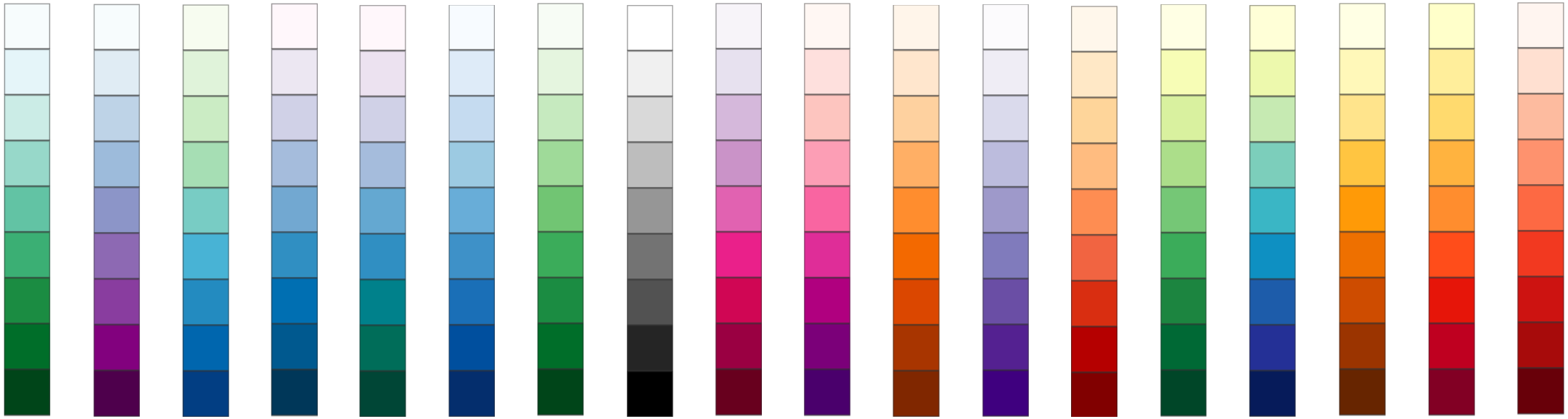
Encoding Validation





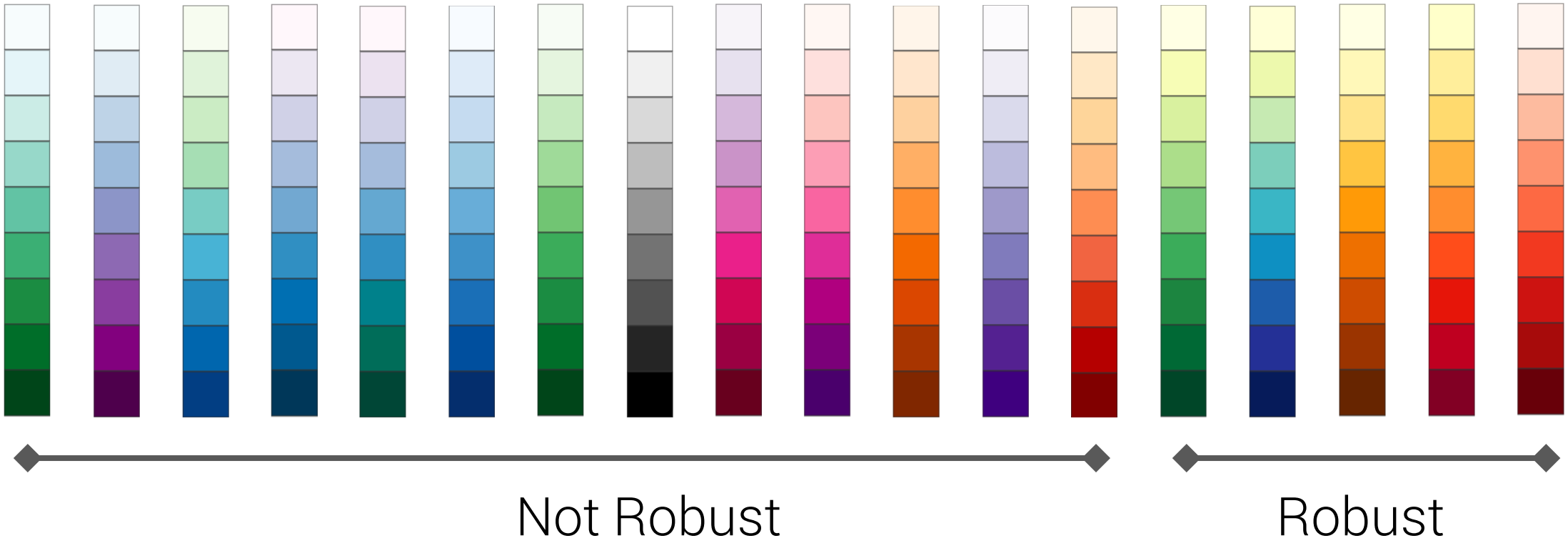
Nine-step sequential Brewer ramps; 4px lines & 10px points

Encoding Validation



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

Encoding Validation

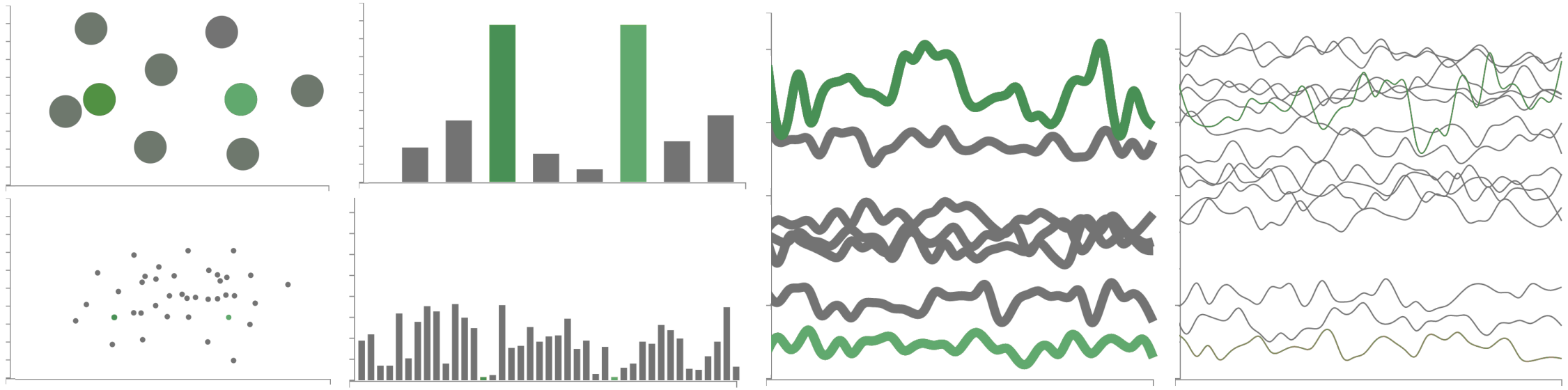


13 of 18 nine-step sequential Brewer ramps are not robust

Thanks!

VisualLab

cmci



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

National Science Foundation
NSF CRII: CHS #1657599

Danielle.Szafir@Colorado.edu
@dalbersszafir