

Data Visualization

-

Visual Encoding

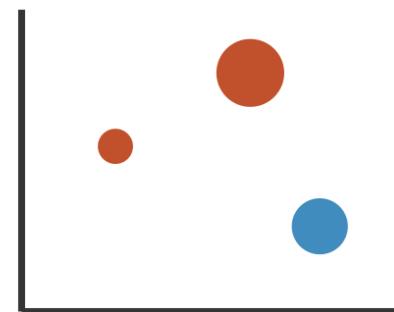
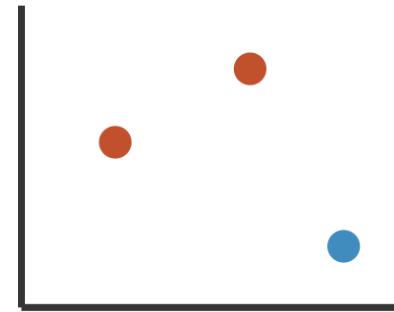
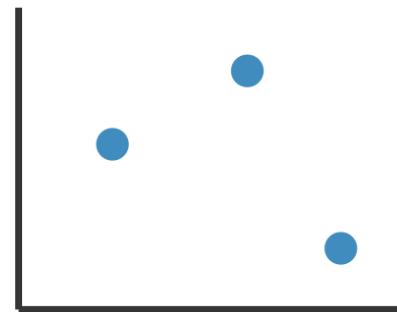
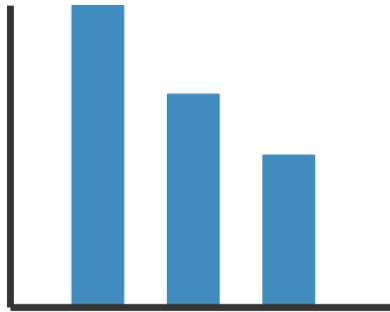
Dr. Claudius Zelenka

Kiel University

cze@informatik.uni-kiel.de

Visual encoding

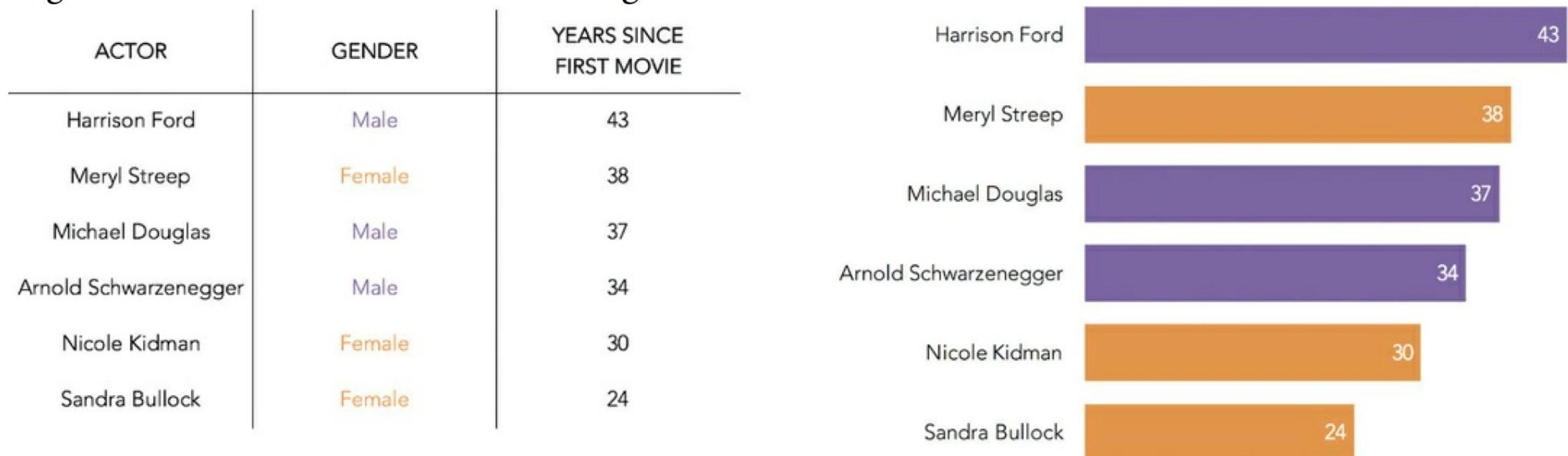
- Data abstraction always uses a **visual encoding**



- (a) Bar charts encode two attributes using a line mark with the vertical spatial position channel for the quantitative attribute, and the horizontal spatial position channel for the categorical attribute.
- (b) Scatterplots encode two quantitative attributes using point marks and both vertical and horizontal spatial position.
- (c) A third categorical attribute is encoded by adding color to the scatterplot.
- (d) Adding the visual channel of size encodes a fourth quantitative attribute as well.

Visual encoding - examples

Figure 6.1 Illustration of Visual Encoding



Visual encoding - examples



Figure 1.4 Two graphical methods for showing the same set of relationships between entities.

How to systematically analyze idiom structure?

- Marks are basic geometric elements that depict items or links
- Channels control their appearance

The effectiveness of a channel for encoding data depends on its type

Marks

→ Points



→ Lines



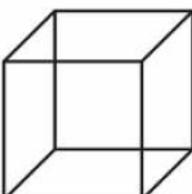
→ Areas



Figure 5.2. Marks are geometric primitives.

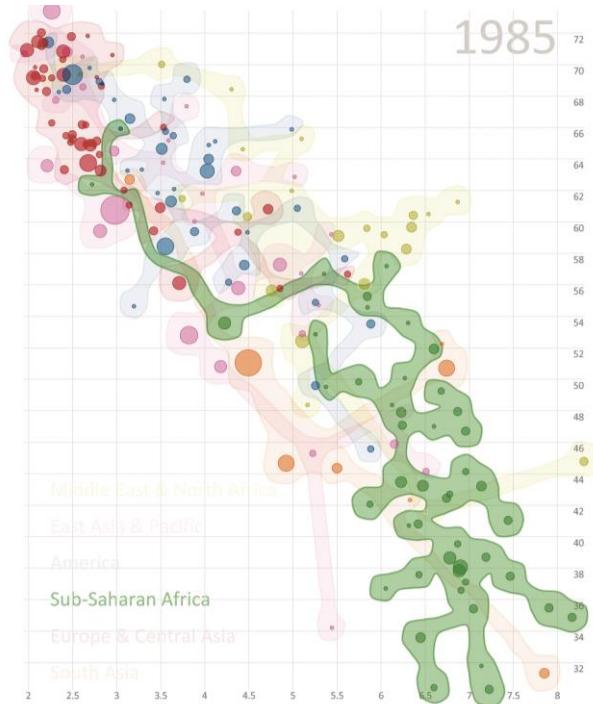
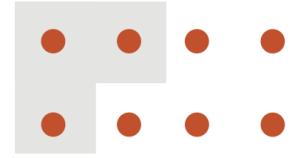
a zero-dimensional (0D) mark is a point, a one-dimensional (1D) mark is a line, and a two-dimensional (2D) mark is an area. A three-dimensional (3D) volume mark is possible, but they are not frequently used.

Marks

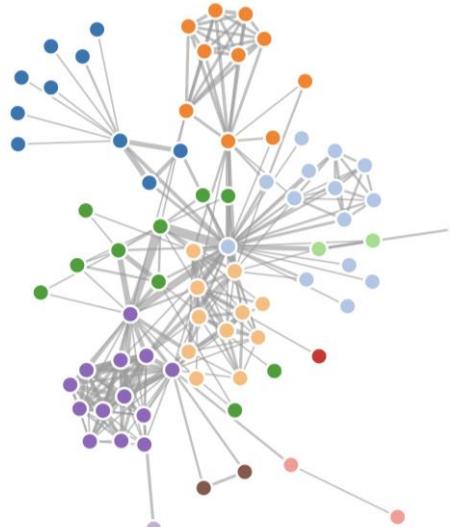
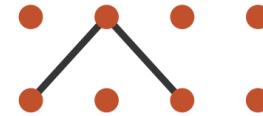
MARK	EXAMPLE	DESCRIPTION
Point		The <i>point</i> mark has no variation ('constant') in the spatial dimension. It is largely a placeholder commonly used to represent a quantity through position on a scale, forming the basis of, for example, scatter plots.
Line		The <i>line</i> mark has one ('linear') spatial dimension. It is commonly used to represent quantitative value through variation in size, forming the basis of, for example, the bar chart.
Area		The <i>area</i> mark has two ('quadratic') spatial dimensions. It is commonly used to represent quantitative values through variation in size and position, forming the basis of, for example, bubble plots.
Form		The <i>form</i> mark has three ('cubic') spatial dimensions. It might be used to represent quantitative values through variation in size (specifically, through volume), forming the basis of, for example, a 3D proportional shape chart.

Marks

→ Containment

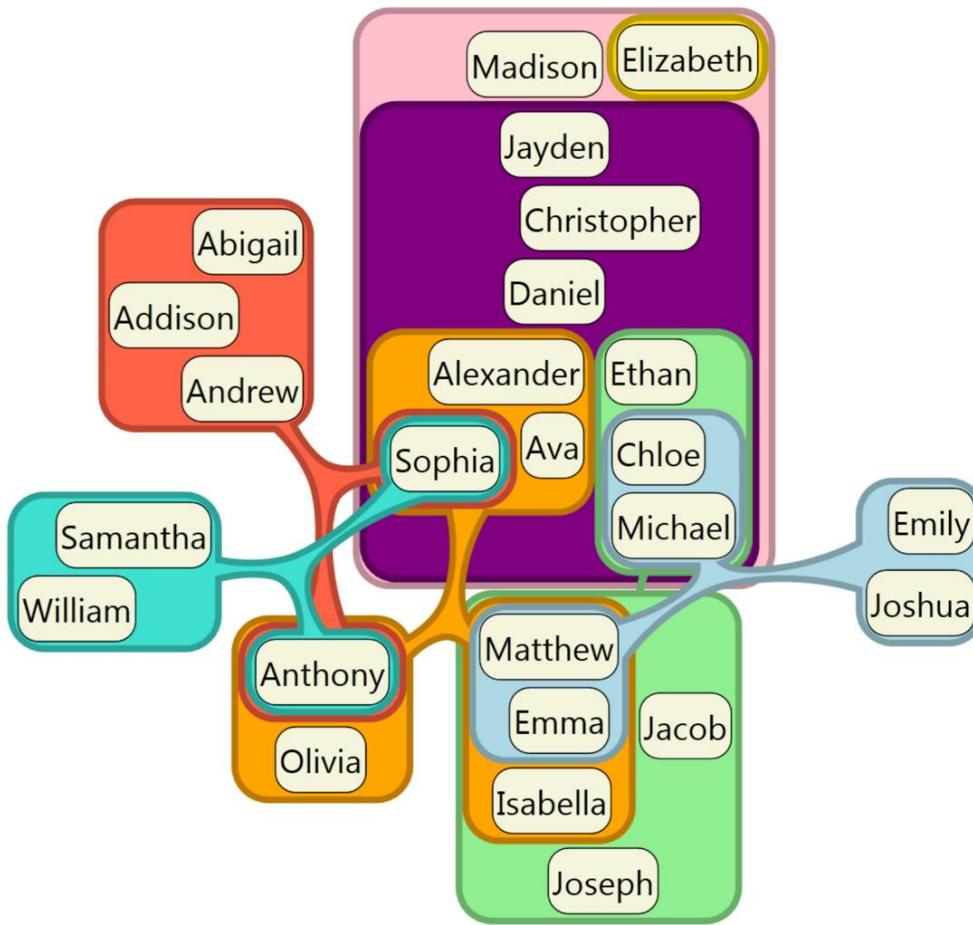


→ Connection



Marks can represent individual items or the links in between

Containment can be nested



Channel

- Many different words for a common concept
- channel, attribute, dimension, variable, feature, and carrier

Expressiveness

- The expressiveness principle dictates that the visual encoding should express all of, and only, the information in the dataset attributes.
- The identity channels are the correct match for the categorical attributes that have no intrinsic order.
- The magnitude channels are the correct match for the ordered attributes, both ordinal and quantitative.

Effectiveness

- The **effectiveness** principle dictates that the importance of the attribute should match the salience of the channel; that is, its noticeability.
- the most important attributes should be encoded with the most effective channels in order to be most noticeable
- decreasingly important attributes can be matched with less effective channels

Channels

- The term channel is independent of the dimensionality of the geometric primitive.
- channel properties differ by type and amount of information that can be conveyed to human perceptual system

Identity (what or where)

Magnitude (how much)

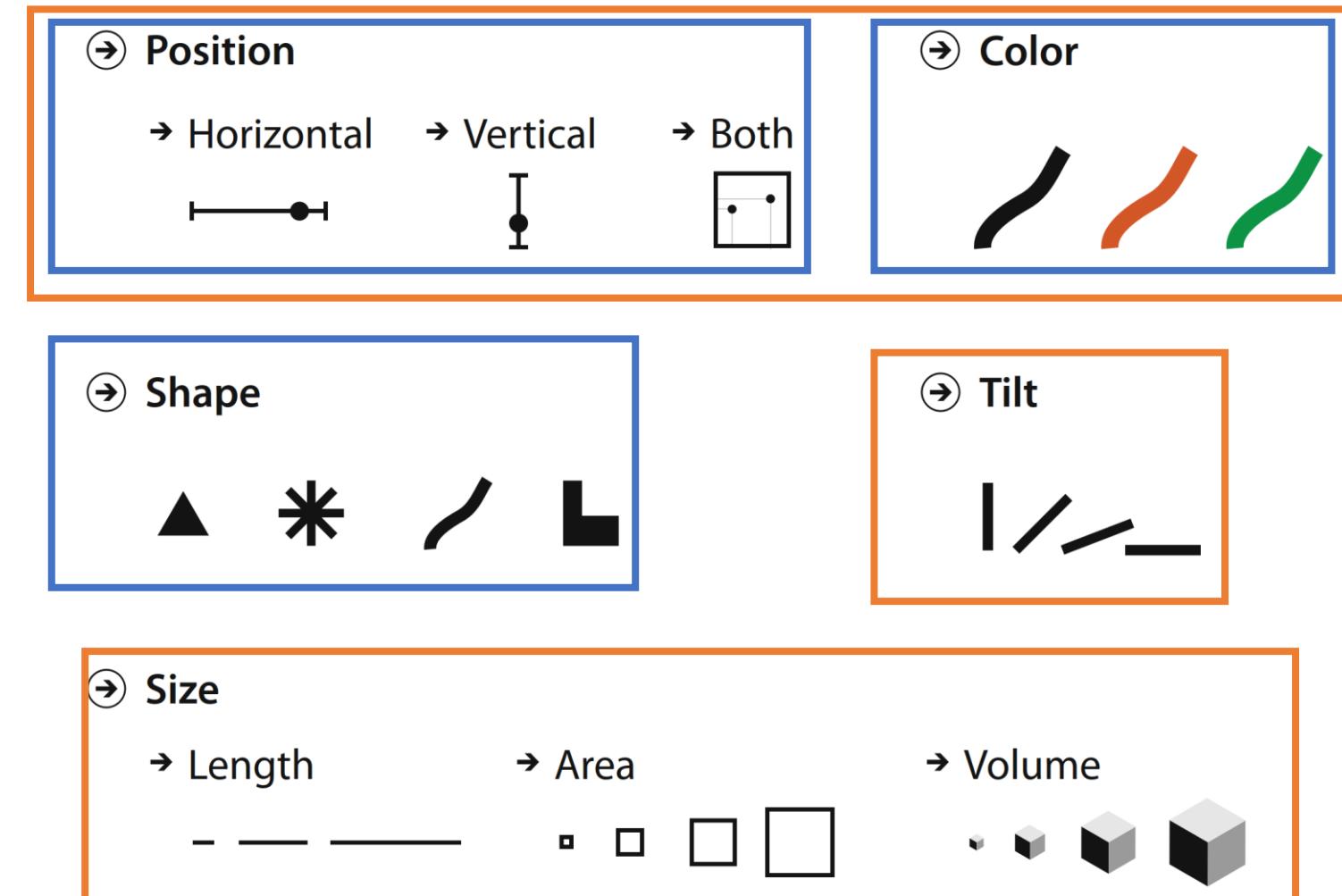
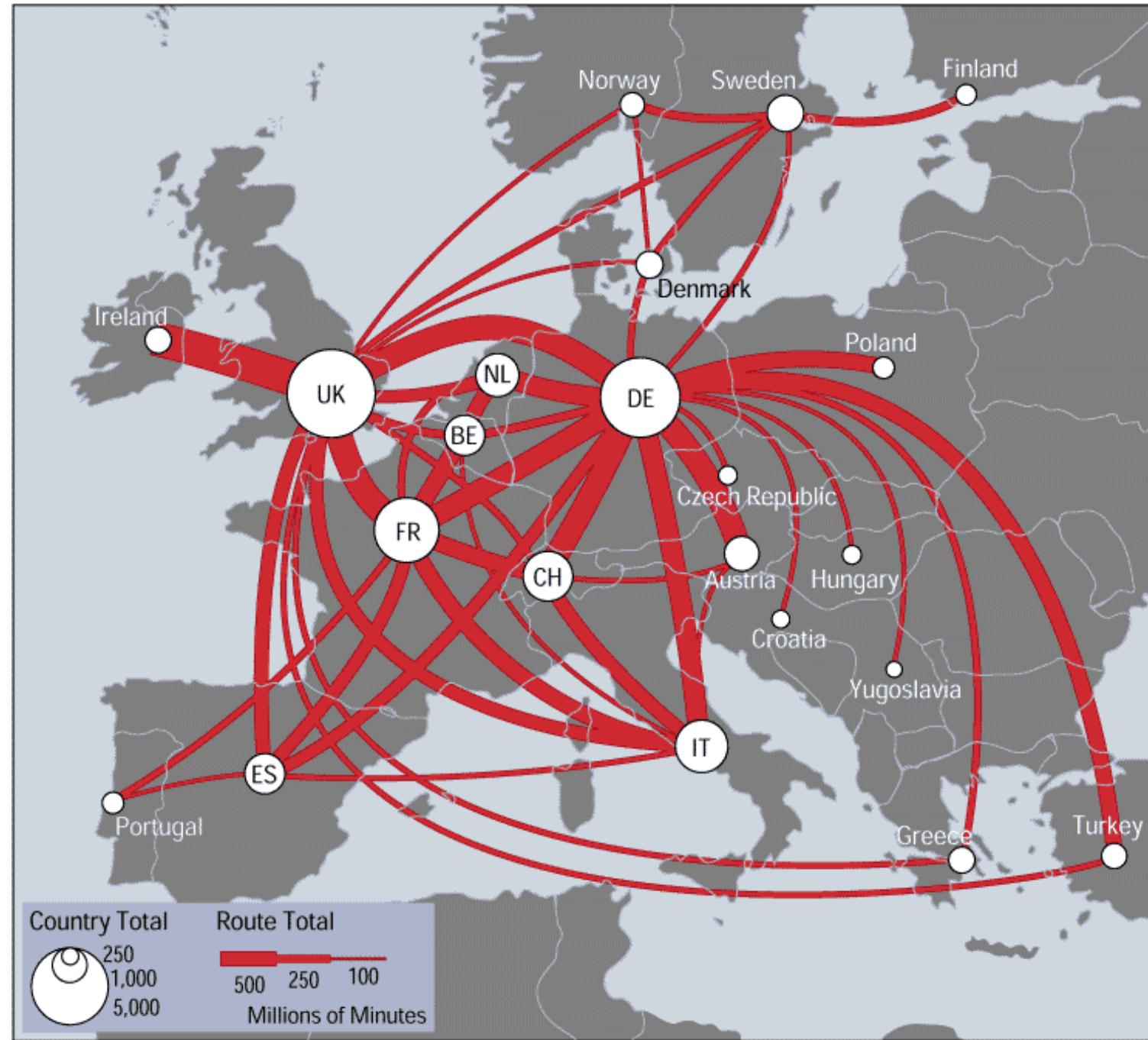


Figure 5.3. Visual channels control the appearance of marks.



Visual grammar of relationship

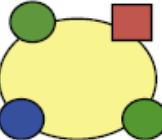
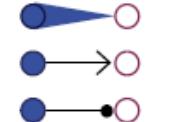
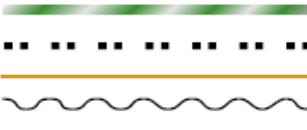
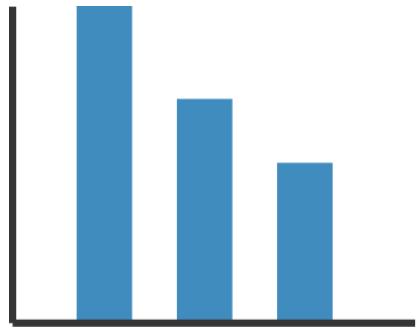
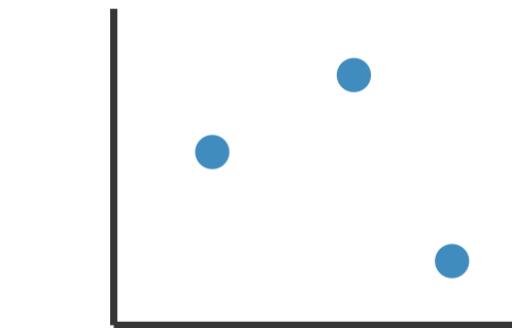
Graphical code	Visual instantiation	Semantics	Graphical code	Visual instantiation	Semantics
1. Partitioned region		Entity partitions	5. Linking line		Relationship between entities
2. Attached shapes		Part-of relationships	6. Asymmetrical connecting graphic		Asymmetrical relationship
3. Enclosed shapes		Contained entities Part-of relationships	7. Line style		Type of relationship
4. Sequence of shapes		Sequence of entities	8. Line weight		Strength of relationship
			9. Tab shapes with matching receptacles		A fit between components
			10. Proximity groupings		Groups of components

Figure 6.62 The visual grammar of relationship representations.

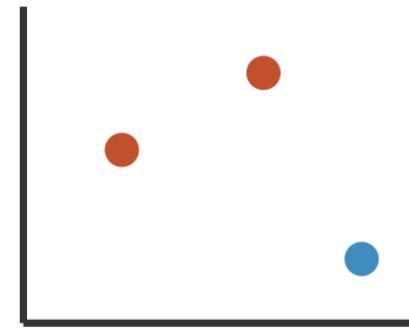
Combination of mark and channels



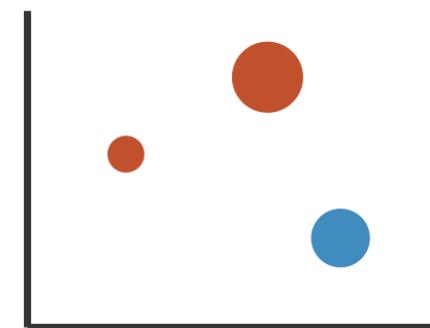
1:
vertical position
mark: line



2:
vertical position
horizontal position
mark: point



3:
vertical position
horizontal position
color hue
mark: point



4:
vertical position
horizontal position
color hue
size (area)
mark: point

Charts

- marks and attributes are the **ingredients**
- a chart ‘type’ is the **recipe** offering a predefined template for displaying data
- Different chart types offer different ways of representing data
- unique combinations of marks and attributes onto which specific types of data can be mapped

Bertin, 1967

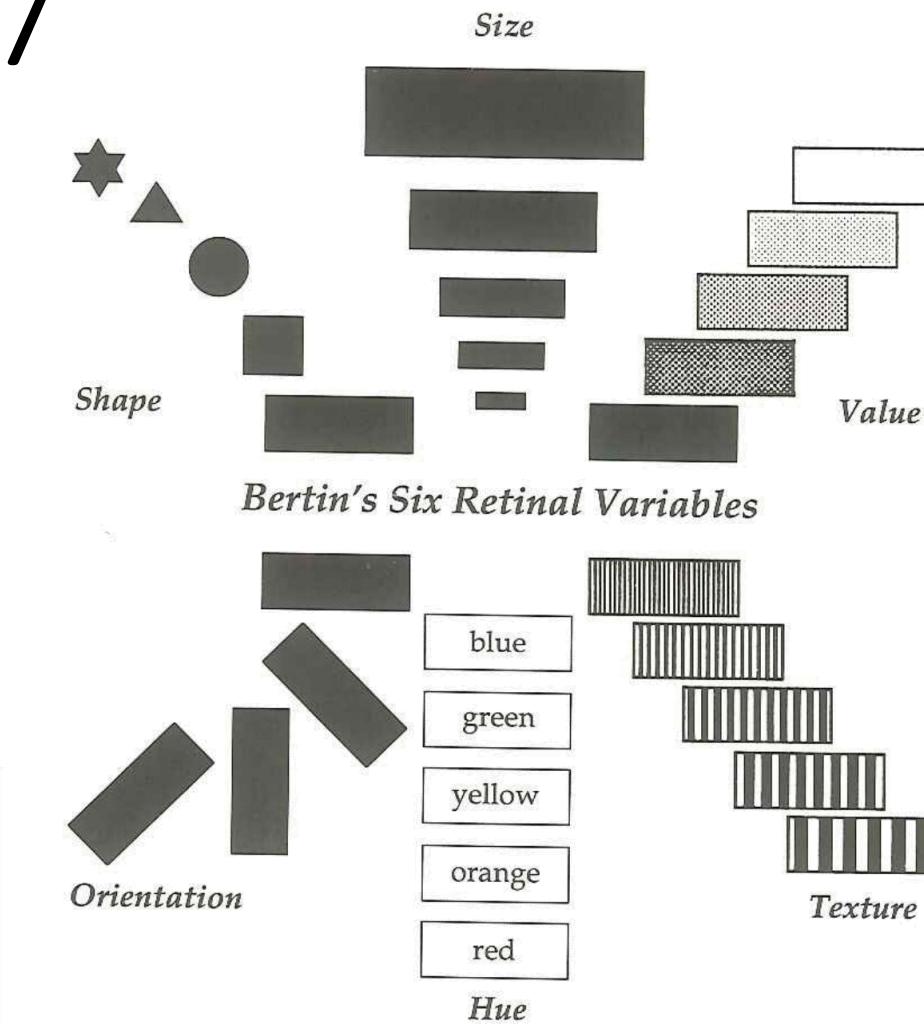
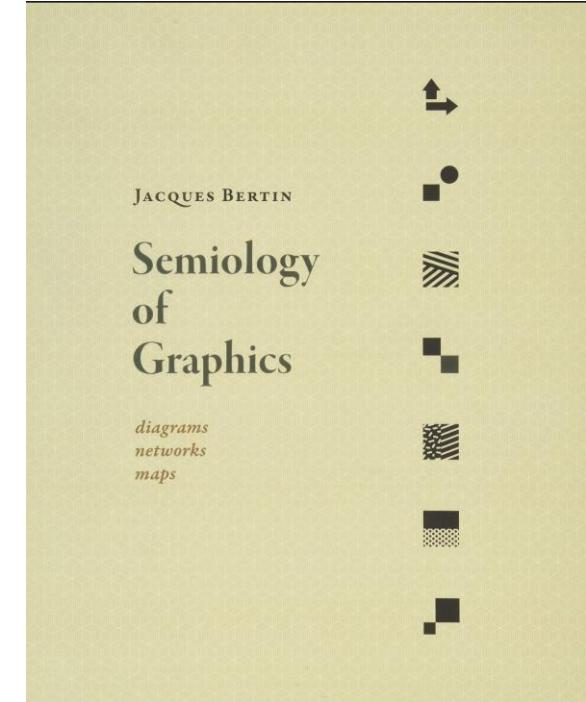
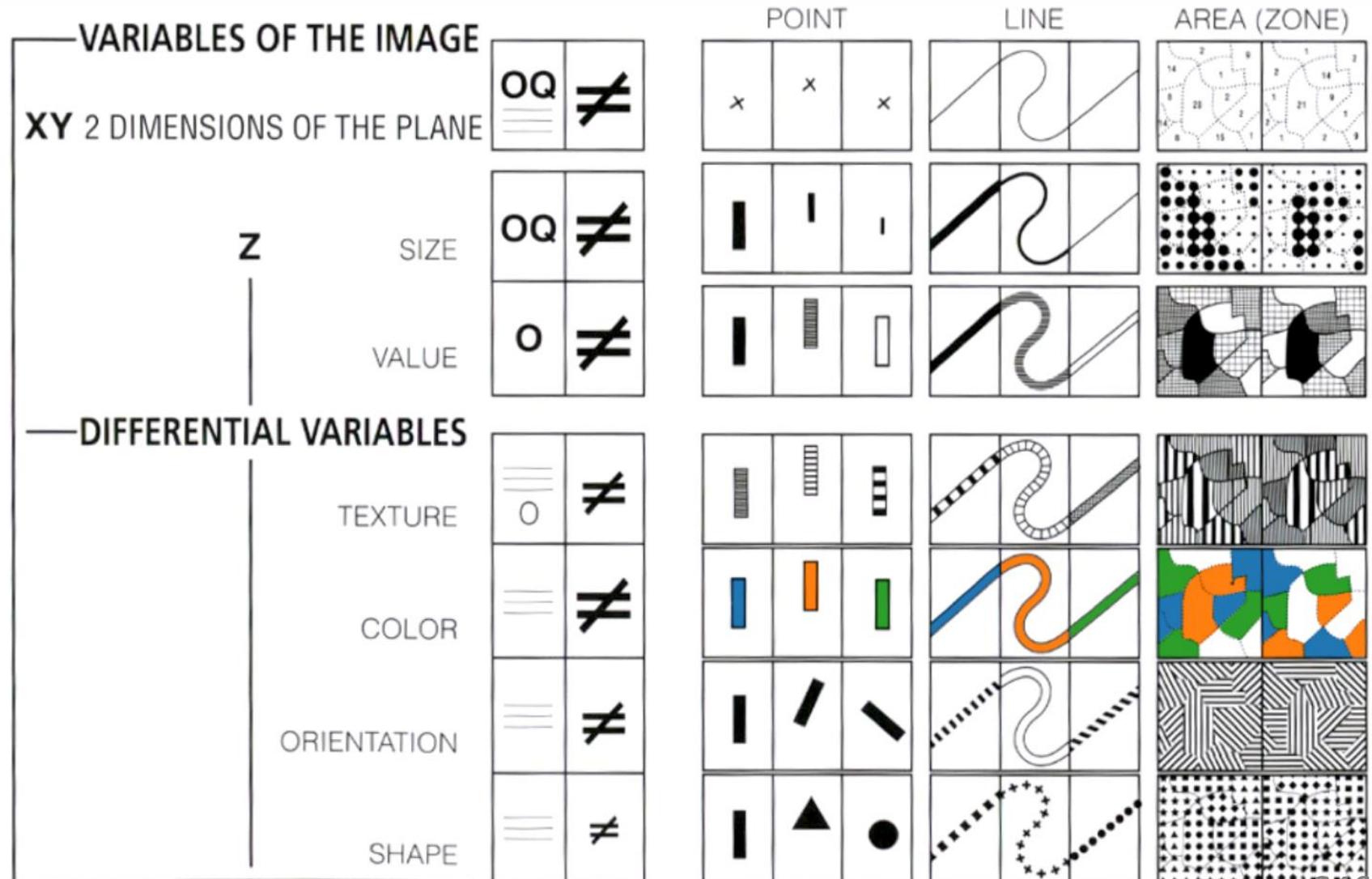


FIGURE 3.1. Bertin's six retinal variables.



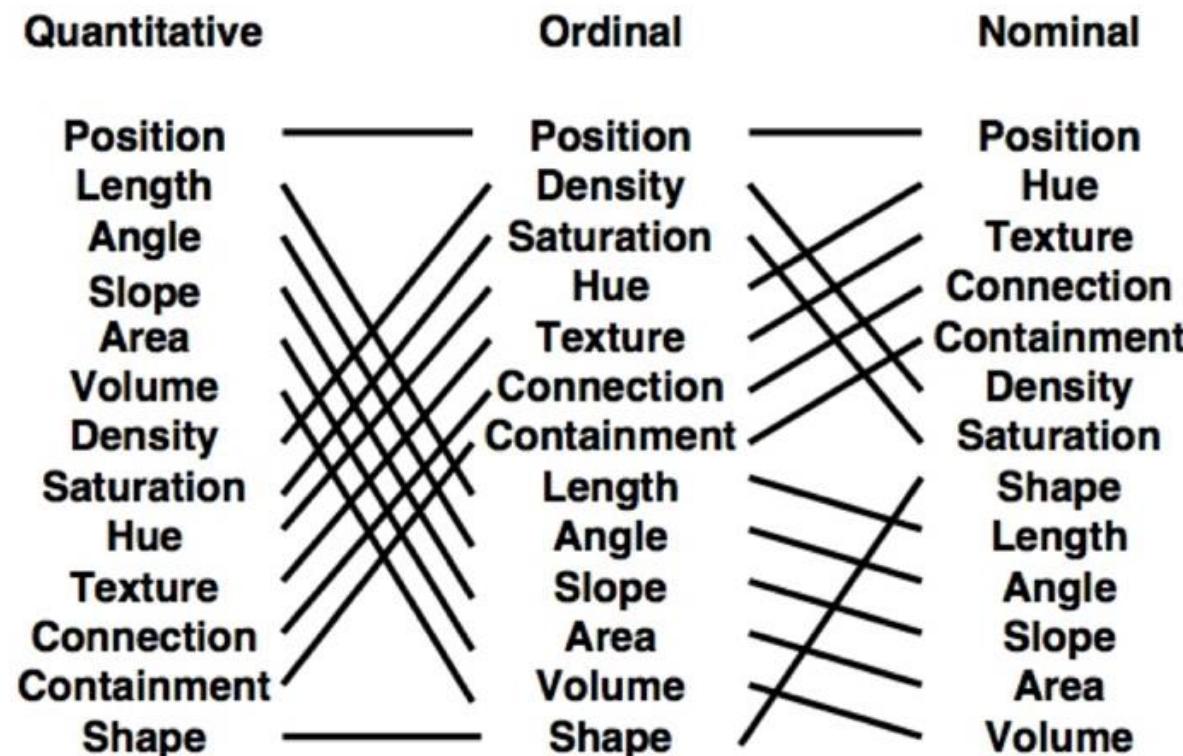
Bertin, 1967

O = Ordinal, Q = Quantitative
 ≠ = Differences = = Similarities



Mackinlay, 1986

Mackinlay, 1986



Cleveland / McGill, 1984

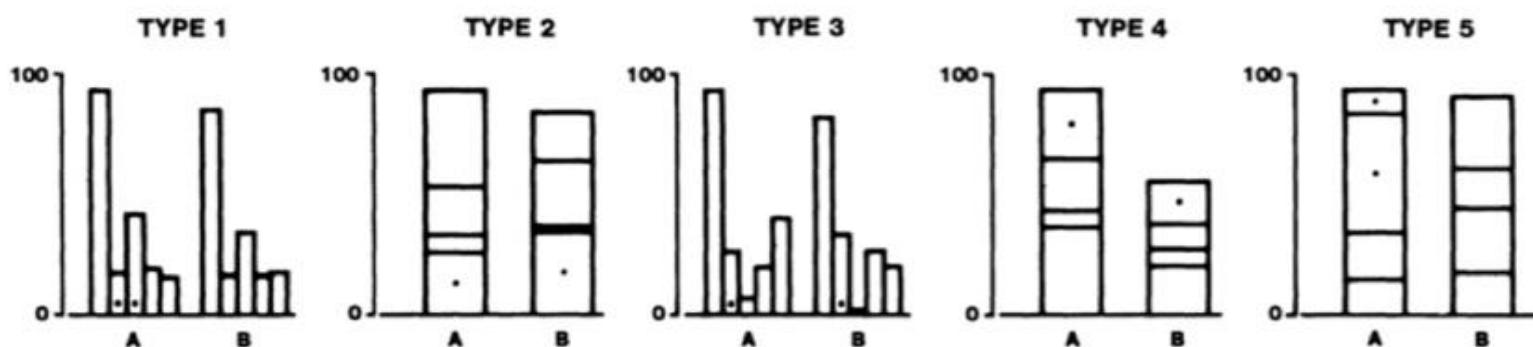


Figure 4. Graphs from position-length experiment.

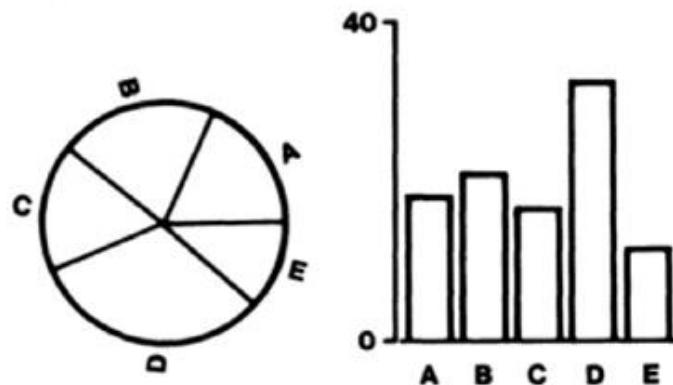


Figure 3. Graphs from position-angle experiment.

William S. Cleveland; Robert McGill. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." 1984

Heer & Bostock, 2010

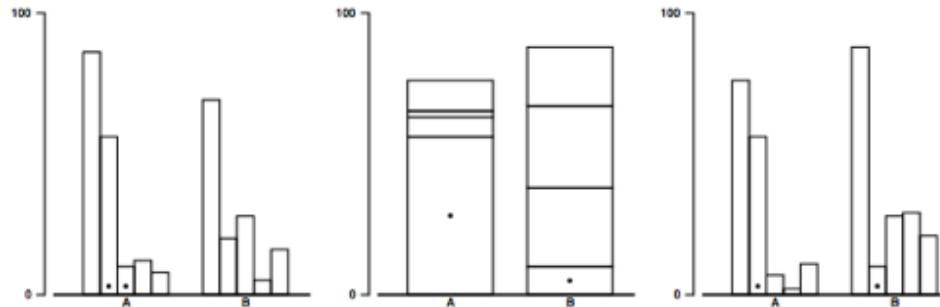


Figure 1: Stimuli for judgment tasks T1, T2 & T3. Subjects estimated percent differences between elements.

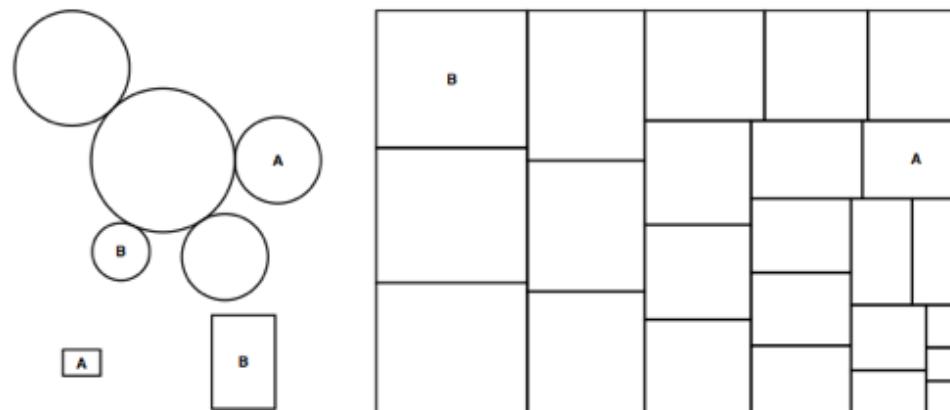


Figure 2: Area judgment stimuli. Top left: Bubble chart (T7), Bottom left: Center-aligned rectangles (T8), Right: Treemap (T9).

Accuracy of channels

- obvious way to quantify effectiveness is accuracy: how close is human perceptual judgement to some objective measurement of the stimulus?

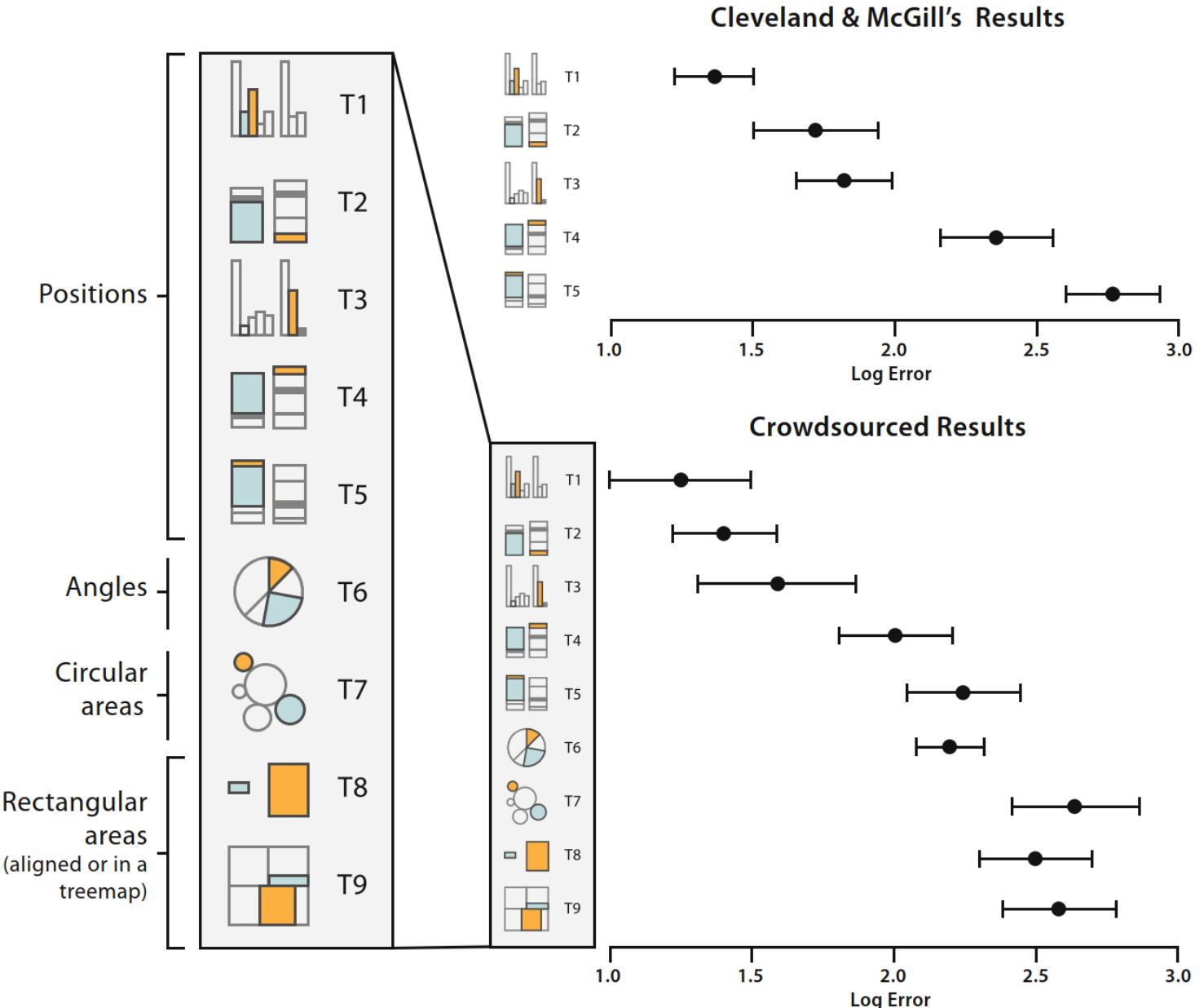


Figure 5.8. Error rates across visual channels, with recent crowdsourced results replicating and extending seminal work from Cleveland and McGill [Cleveland and McGill 84a]. After [Heer and Bostock 10, Figure 4]. 250

Properties and Best Uses of Visual Encodings

<u>Example</u>	<u>Encoding</u>	<u>Ordered</u>	<u>Useful values</u>	<u>Quantitative</u>	<u>Ordinal</u>	<u>Categorical</u>	<u>Relational</u>
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional (alphabetical or numbered)	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium/few	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (< 20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

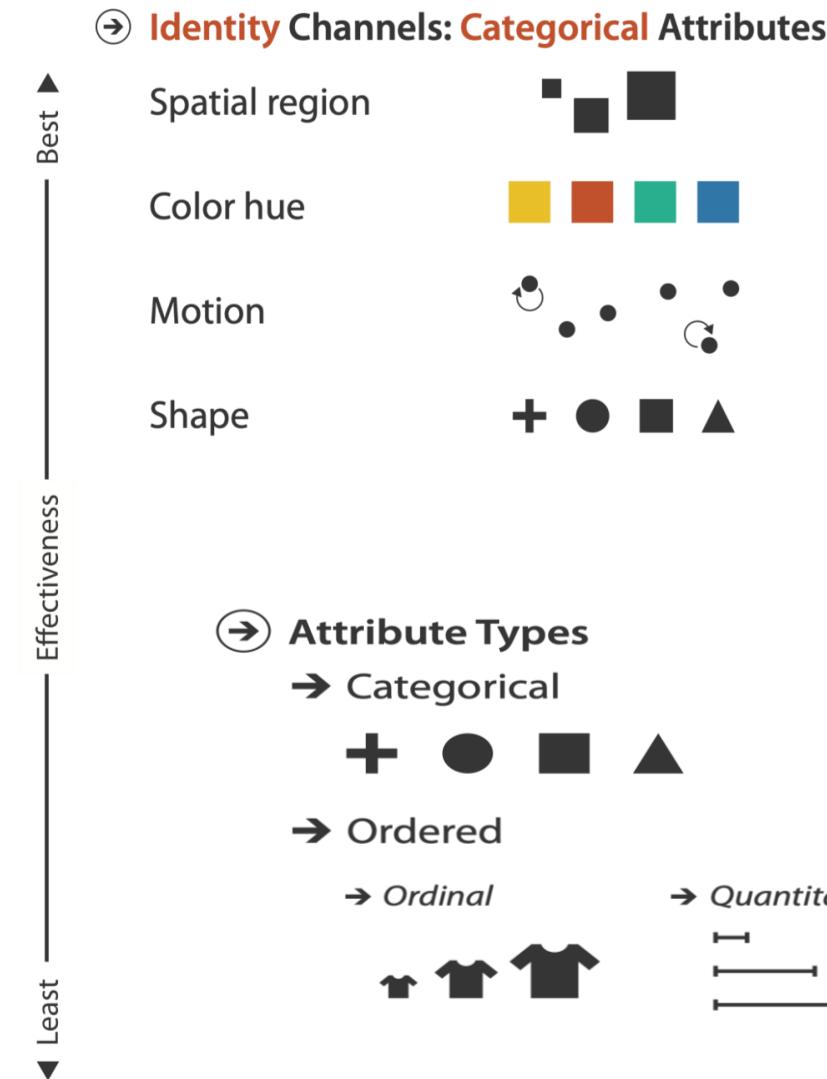
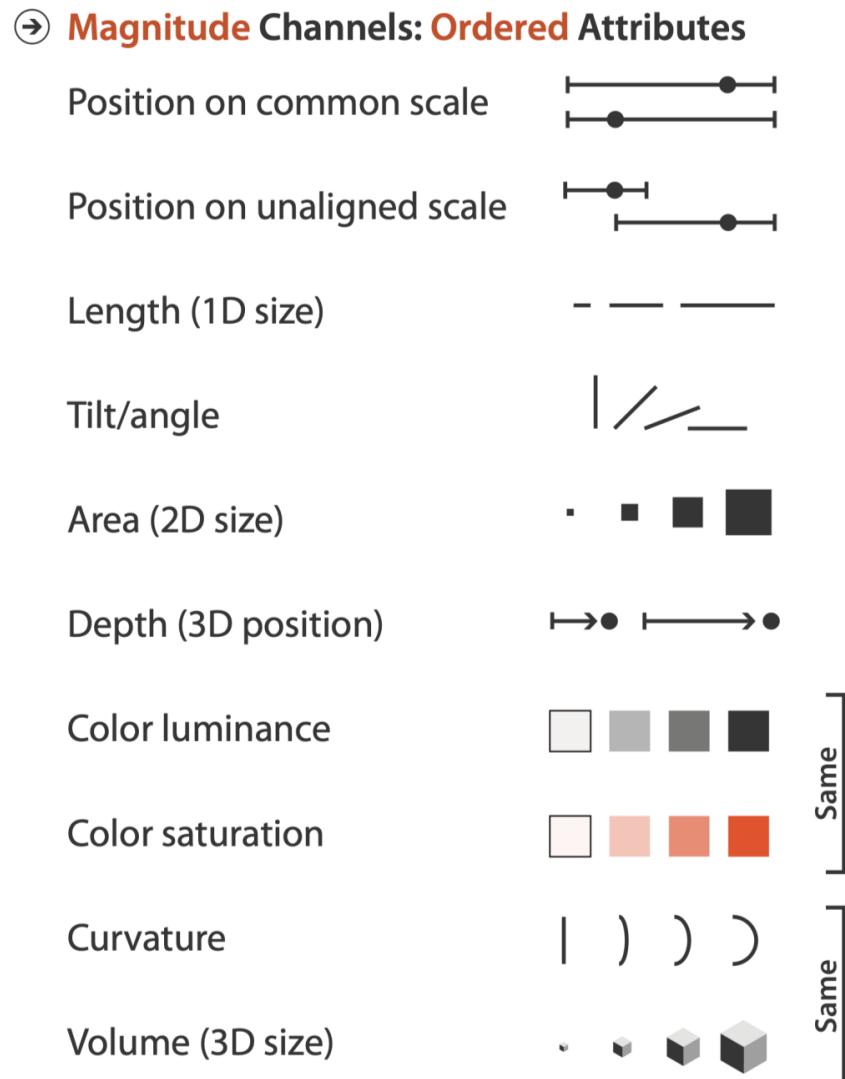


Table of Visual Attributes

Richard Brath
v. Sept 2013

		Information Visualization Researchers									Vision Rsch	Shape Rsch
		Bertin 1967	Cleveland 1985	Mackinlay 1986	MacEachren 1995	Wilkinson 1999	Ware 2000	Mazza 2009	Iliinsky 2012	Chen, Floridi 2013	Preattentive Perception	Brath 2009/2011
Transform	Position	X	X	X	X	X	X	X	X	X		
	Length		X	X			X	X	X	X	X	
	Size (Area)	X	X	X	X	X	X	X	X	X	X	
	Orientation	X		X	X	X	X	X	X	X	X	
	Volume		X	X		X						
Shape	Shape	X		X	X	X	X	X	X	X		X
	Angle		X	X					X			X
	Curvature										X	X
	Mark										X	X
	Line Ending						X	X	3		X	X
	Closure								X		X	X
	Local Warp											X
	Edge Type								1,2			X
	Corner Type								3			X
Colour	Icon, glyph, etc									4		
	Brightness	X		X	X	X	X	X	X	X	X	
	Hue	X	X	X	X	X	X	X	X	X	X	
	Saturation		X	X	X	X	X	X	X	X		
Texture	Granularity	X	X	X	X	X	X	X	X	X		
	Pattern				X	X	X	X				
	Orientation				X	X						
Relation	Connection			X			X	X	X			
	Containment			X			X	X				
Optics	Blur				X	X			X			
	Transparency				X	X			X			
	Stereo Depth										X	
	Concavity								X		X	
	Light Direction								X		X	
	Shadow								X			
	Partial occlusion								X			
Movement	Flicker					X			X		X	
	Speed					X			X		X	
	Direction								X		X	
Misc	Numerosity								X		X	
	Spatial Grouping								X		X	
	Arrangement				X							
	Resolution				X							
	Artistic Effects									X		
	Text Labels						X	X	X			

Effectiveness and expressiveness of channels



Perception intensity

- Weber's Law
- Not all perception/ senses are equal
- Exponents:

Loudness 0.6

Brightness 0.33

Smell 0.55 (Coffee) - 0.6 (Heptane)

Taste 0.6 (Saccharine) -1.3 (Salt)

Temperature 1.0 (Cold) – 1.6 (Warm)

Vibration 0.6 (250 Hz) – 0.95 (60 Hz)

Duration 1.1

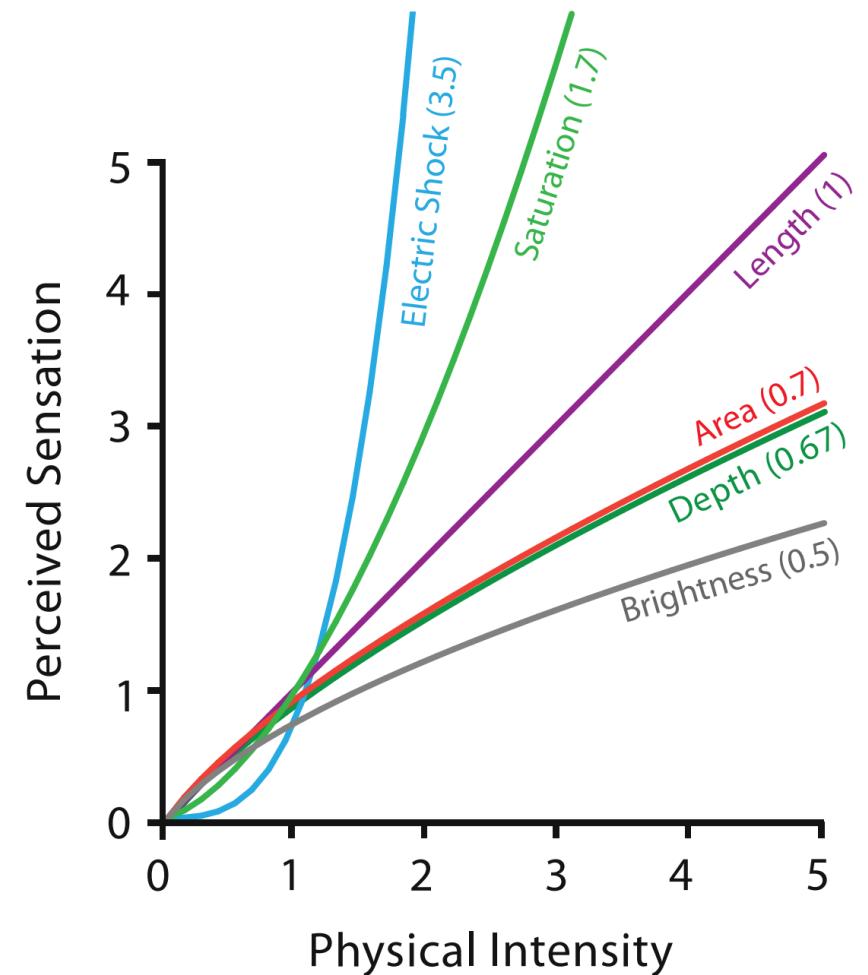
Pressure 1.1

Heaviness 1.45

Electric Shock 3.5

[Psychophysics of Sensory Function, Stevens 61]

Steven's Psychophysical Power Law: $S = I^N$



Absolute values

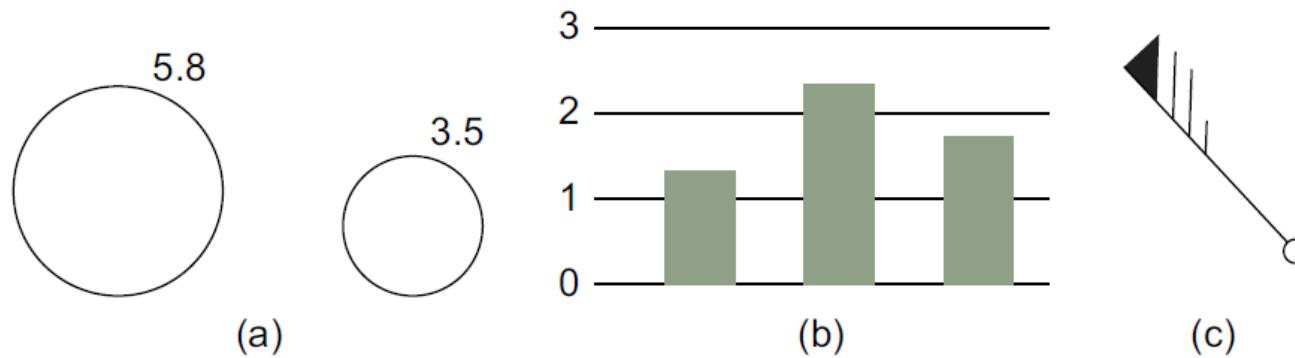


Figure 5.28 Three different ways that more exact numerical values can be read from a diagram. Bar and line graphs should be given faint horizontal lines extending the scale into the background of the chart.

Subitizing

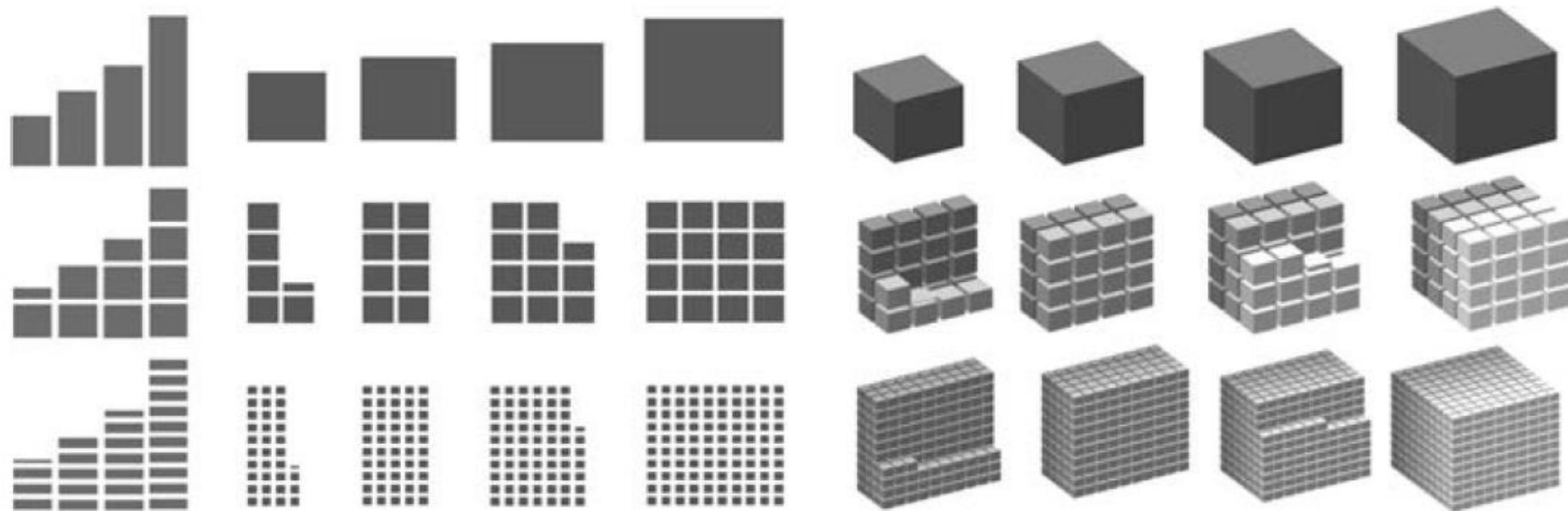
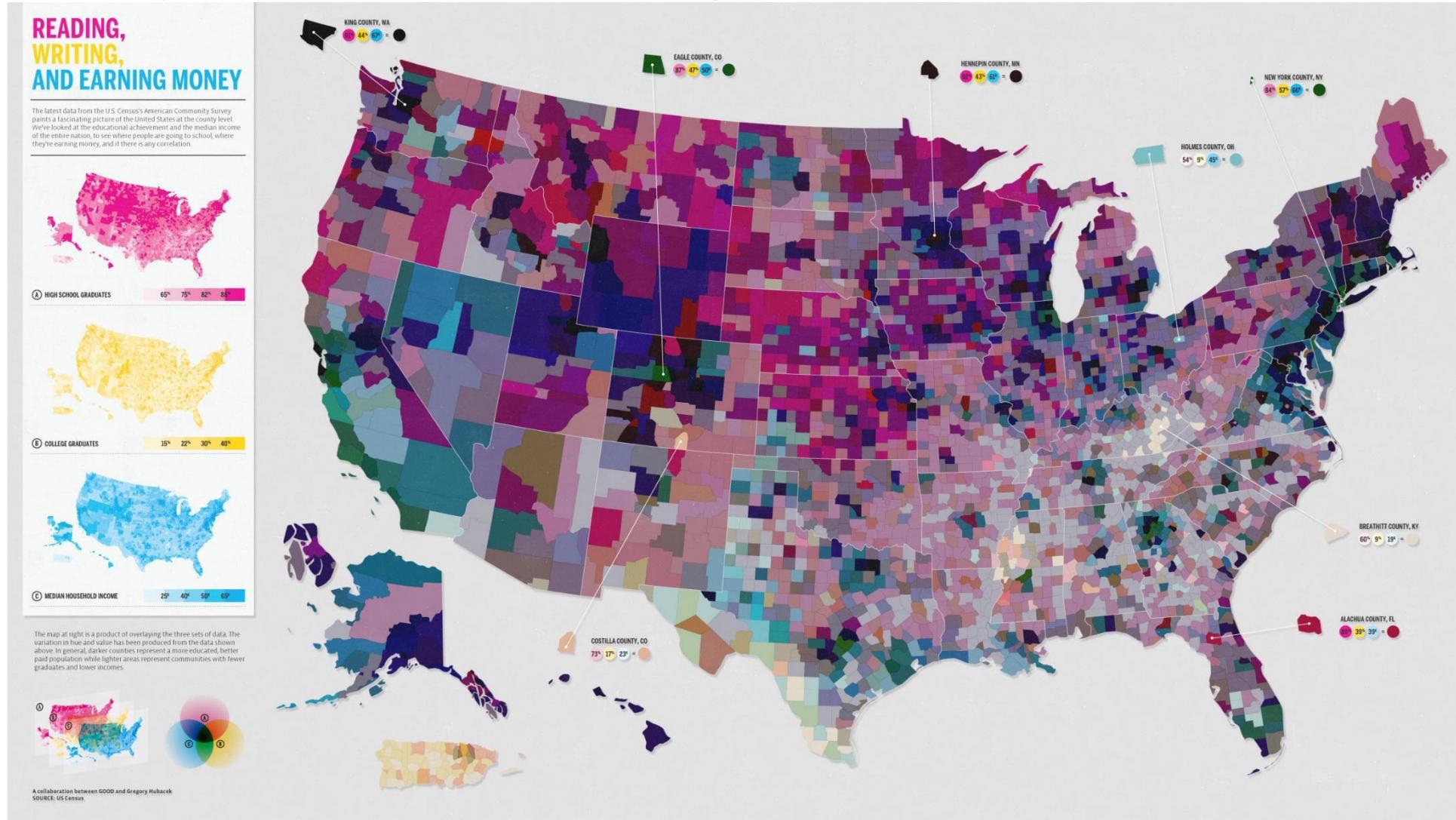


Figure 5.27 Greater accuracy can be obtained if area and volume glyphs are broken into discrete parts. However, if this is done, they cannot be made small.

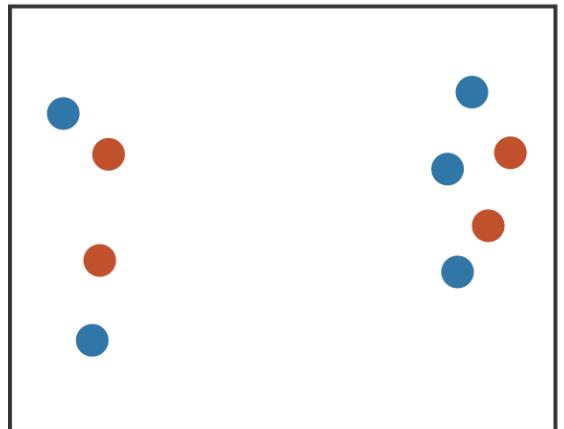
Separable vs. Integral



Separability vs. Integrality

Position

+ Hue (Color)

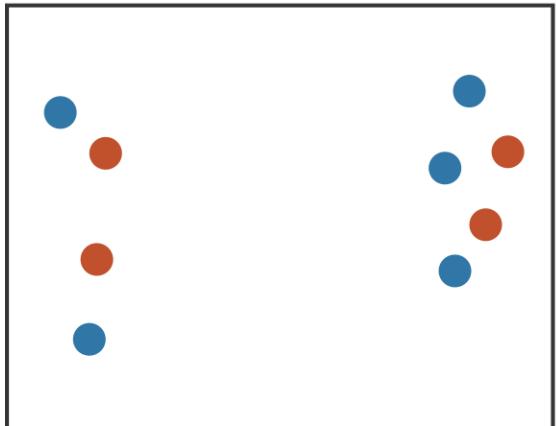


Fully separable

2 groups each

Separability vs. Integrality

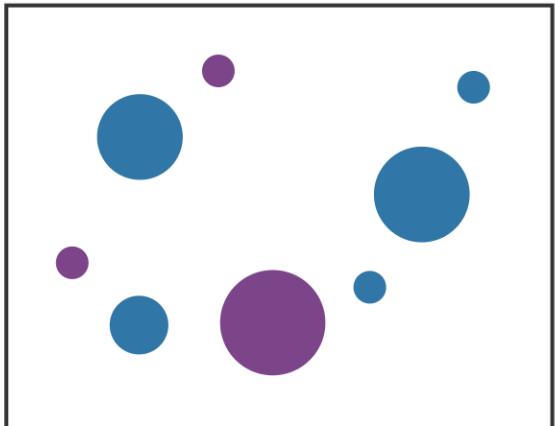
Position
+ Hue (Color)



Fully separable

2 groups each

Size
+ Hue (Color)

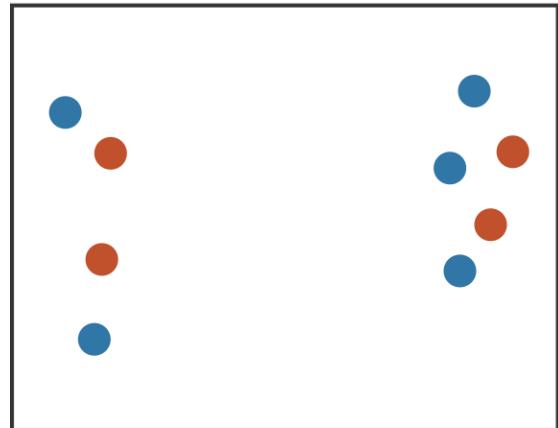


Some interference

2 groups each

Separability vs. Integrality

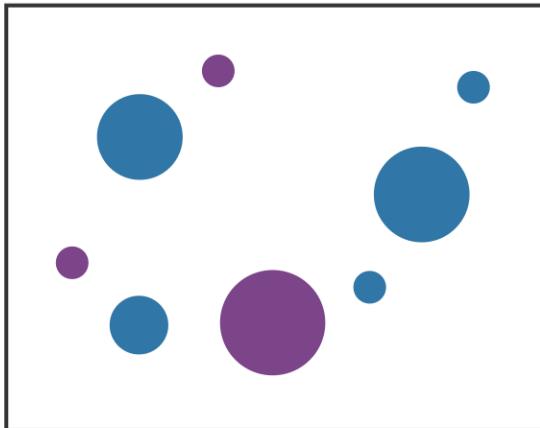
Position
+ Hue (Color)



Fully separable

2 groups each

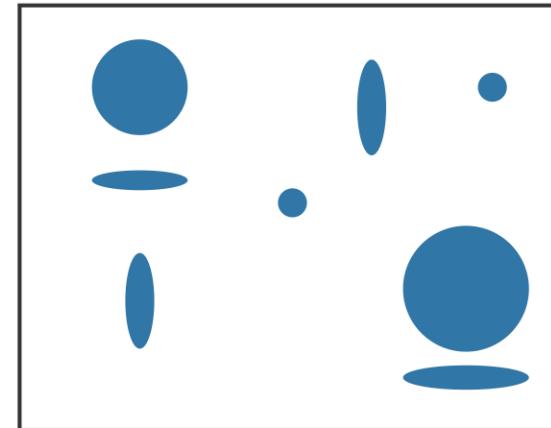
Size
+ Hue (Color)



Some interference

2 groups each

Width
+ Height

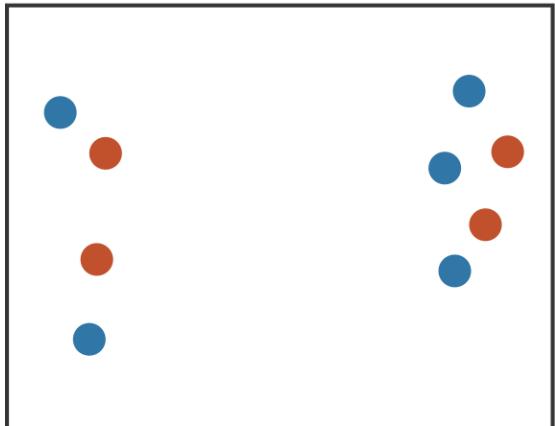


Some/significant
interference

3 groups total:
integral area

Separability vs. Integrality

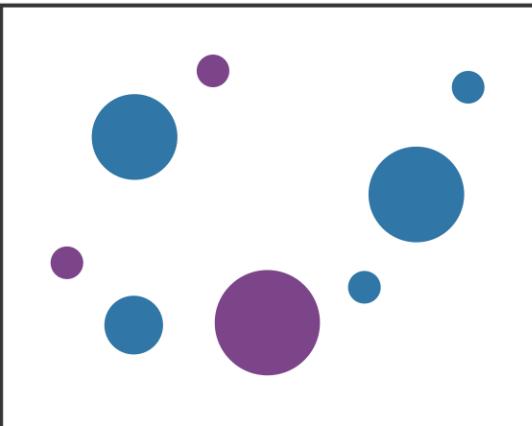
Position
+ Hue (Color)



Fully separable

2 groups each

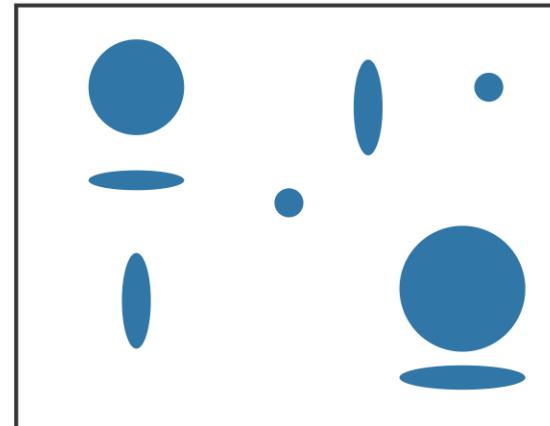
Size
+ Hue (Color)



Some interference

2 groups each

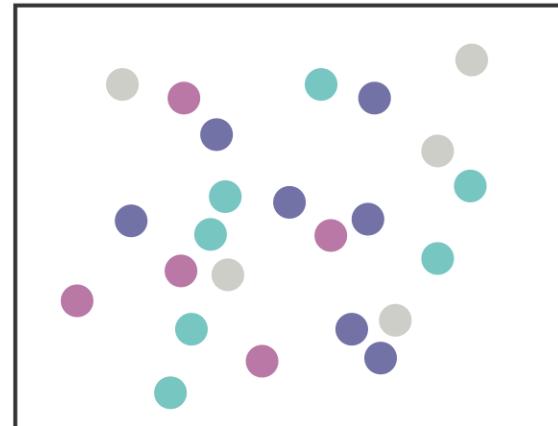
Width
+ Height



Some/significant
interference

3 groups total:
integral area

Red
+ Green



Major interference

4 groups total:
integral hue

Separable or integral

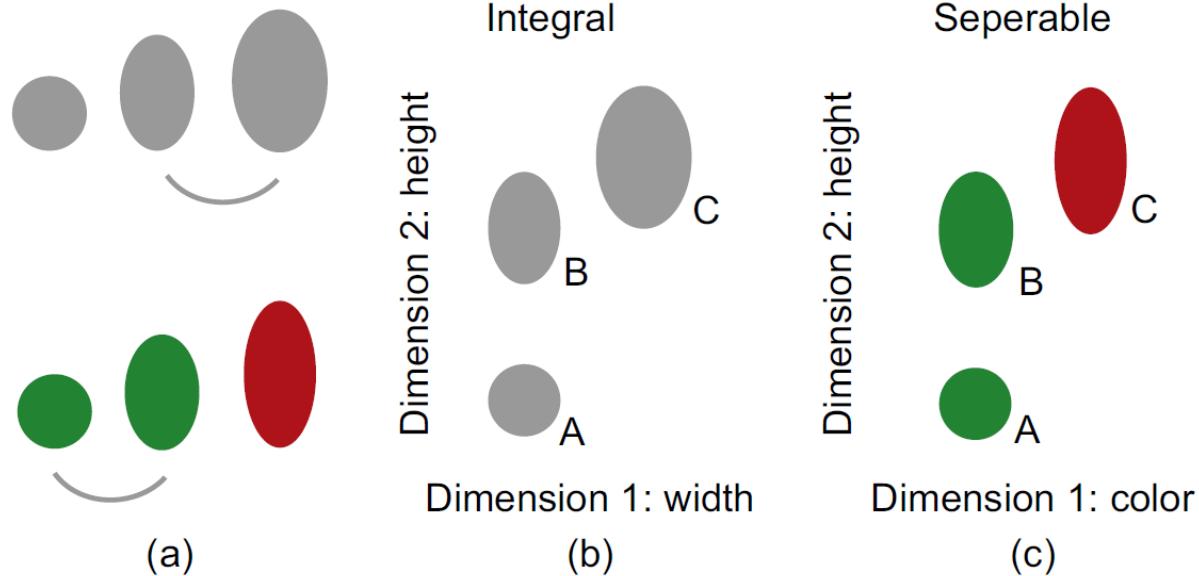
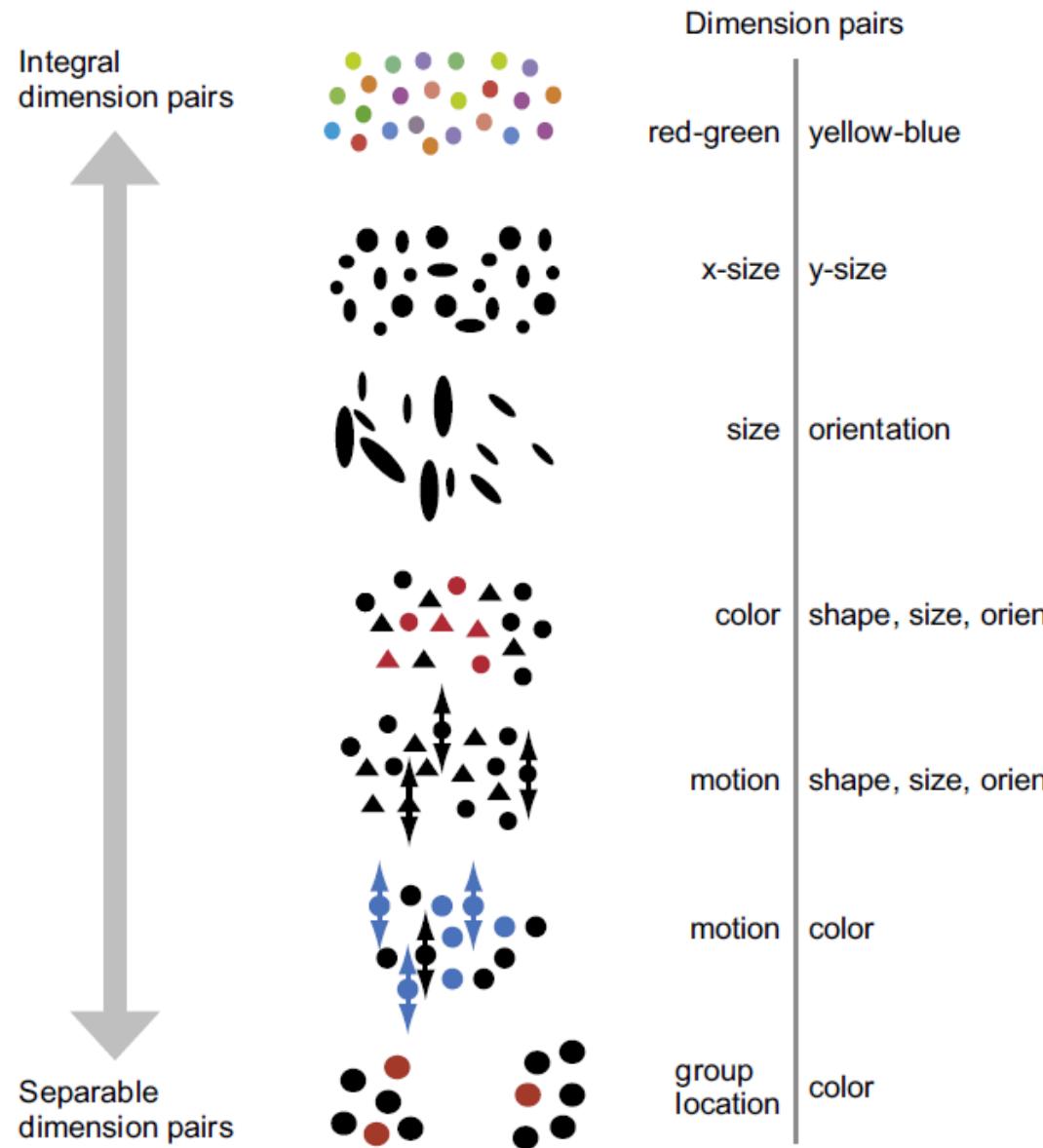


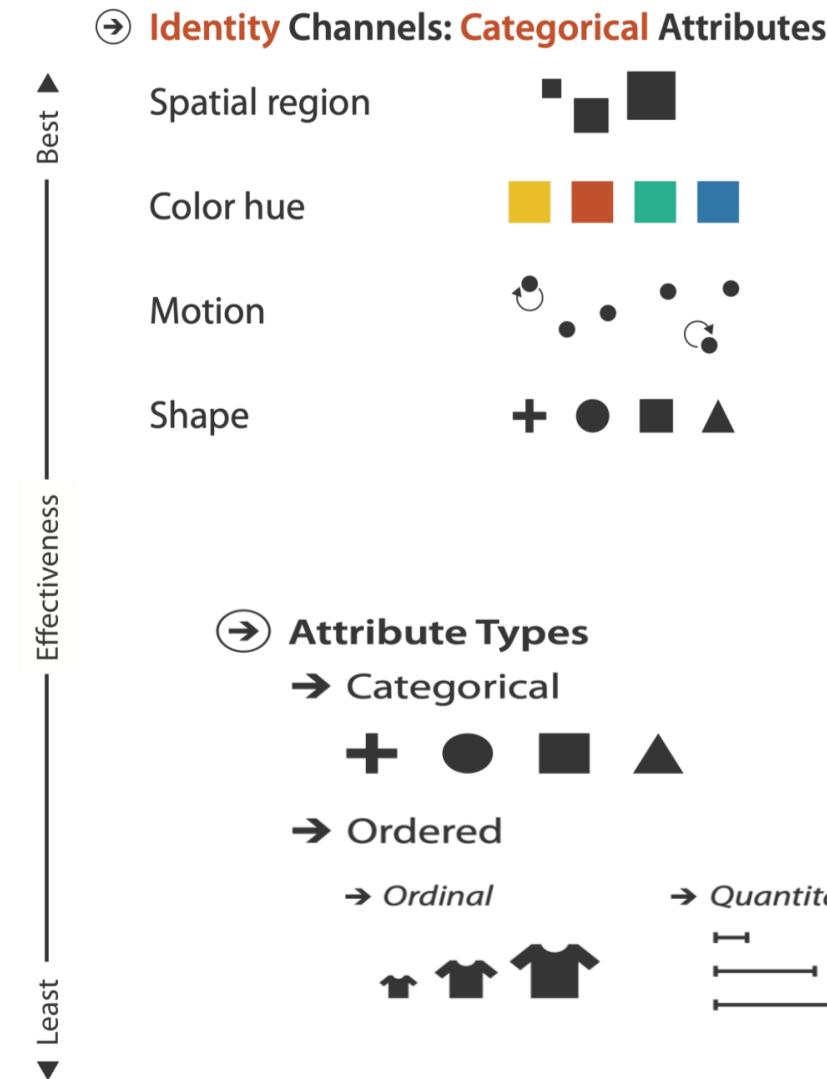
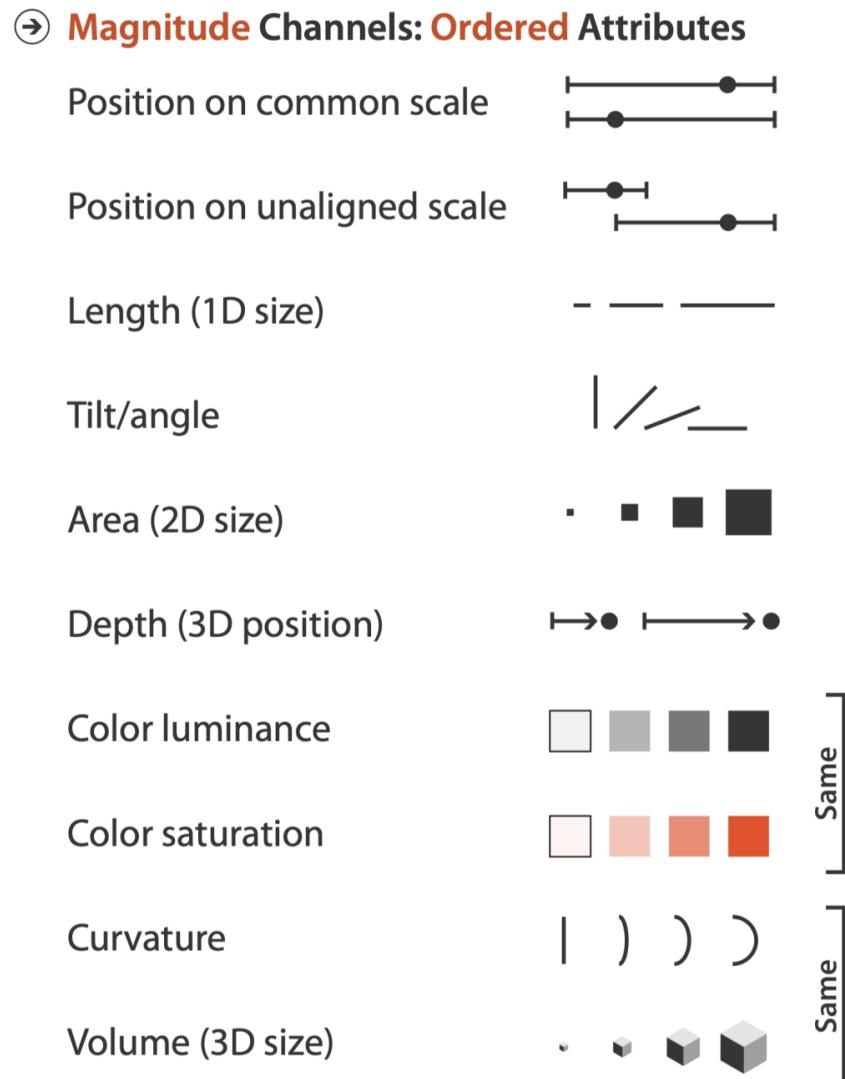
Figure 5.20 (a) The width and height of an ellipse are perceived integrally, so the ellipses are seen as more similar to each other (because they have the same shape) than the pair having the same width. The color and height of a shape are perceived separably, so the two green shapes are seen as most similar. (b, c) Space plots of the two examples.



Separable vs. integral

Figure 5.24 Examples of glyphs coded according to two display attributes. At the top are more integral coding pairs. At the bottom are more separable coding pairs.

Effectiveness and expressiveness of channels



Data Visualization – Color

Dr. Claudius Zelenka

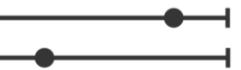
Kiel University

cze@informatik.uni-kiel.de

Channels: What's up with color?

→ Magnitude Channels: Ordered Attributes

Position on common scale



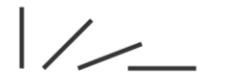
Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



→ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion



Shape



▲ Best

Effectiveness

▼ Least

Same

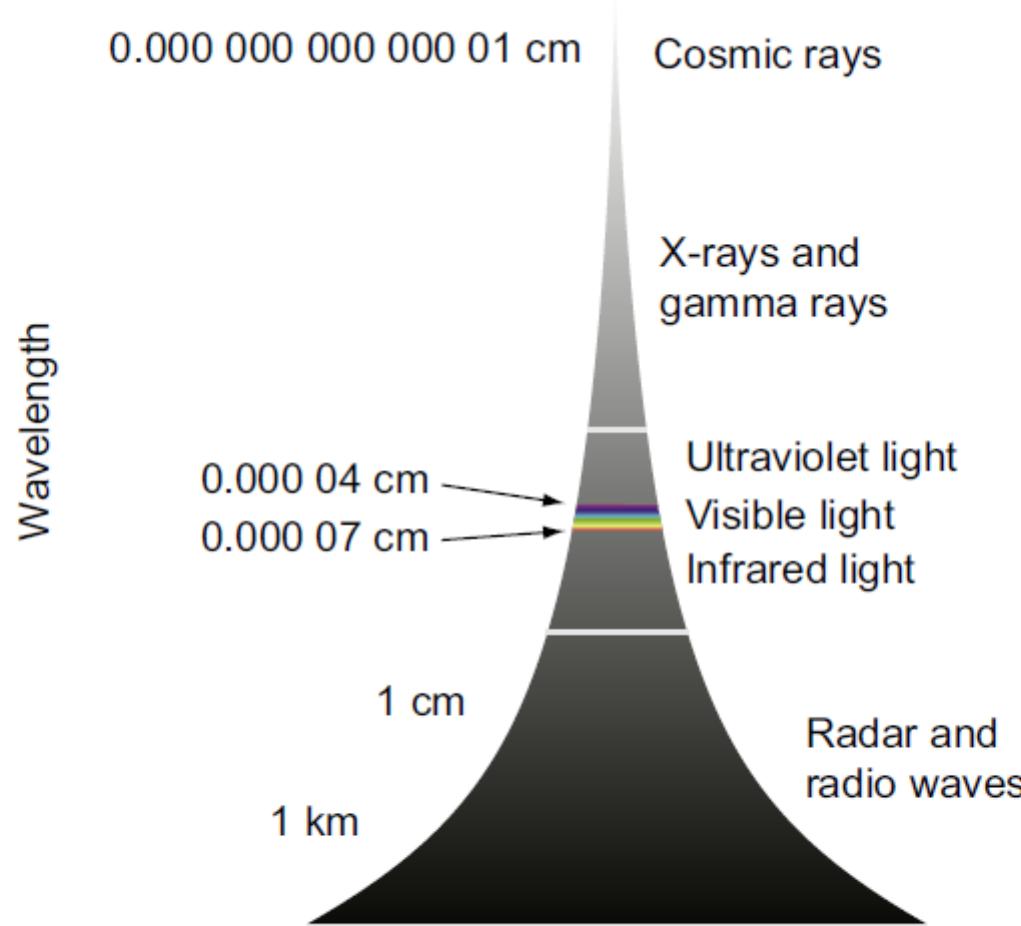
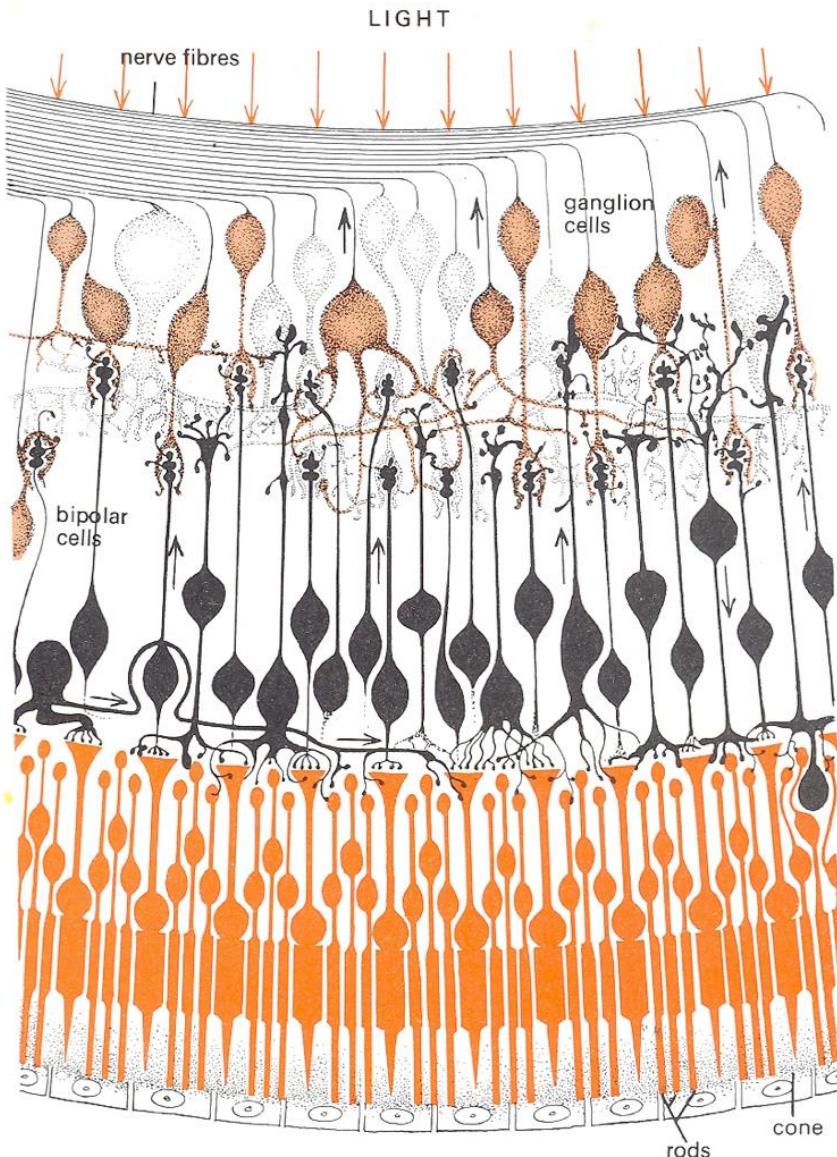


Figure 2.1 The visible light spectrum is a tiny part of a much larger spectrum of electromagnetic radiation.

Photoreceptors



Rods

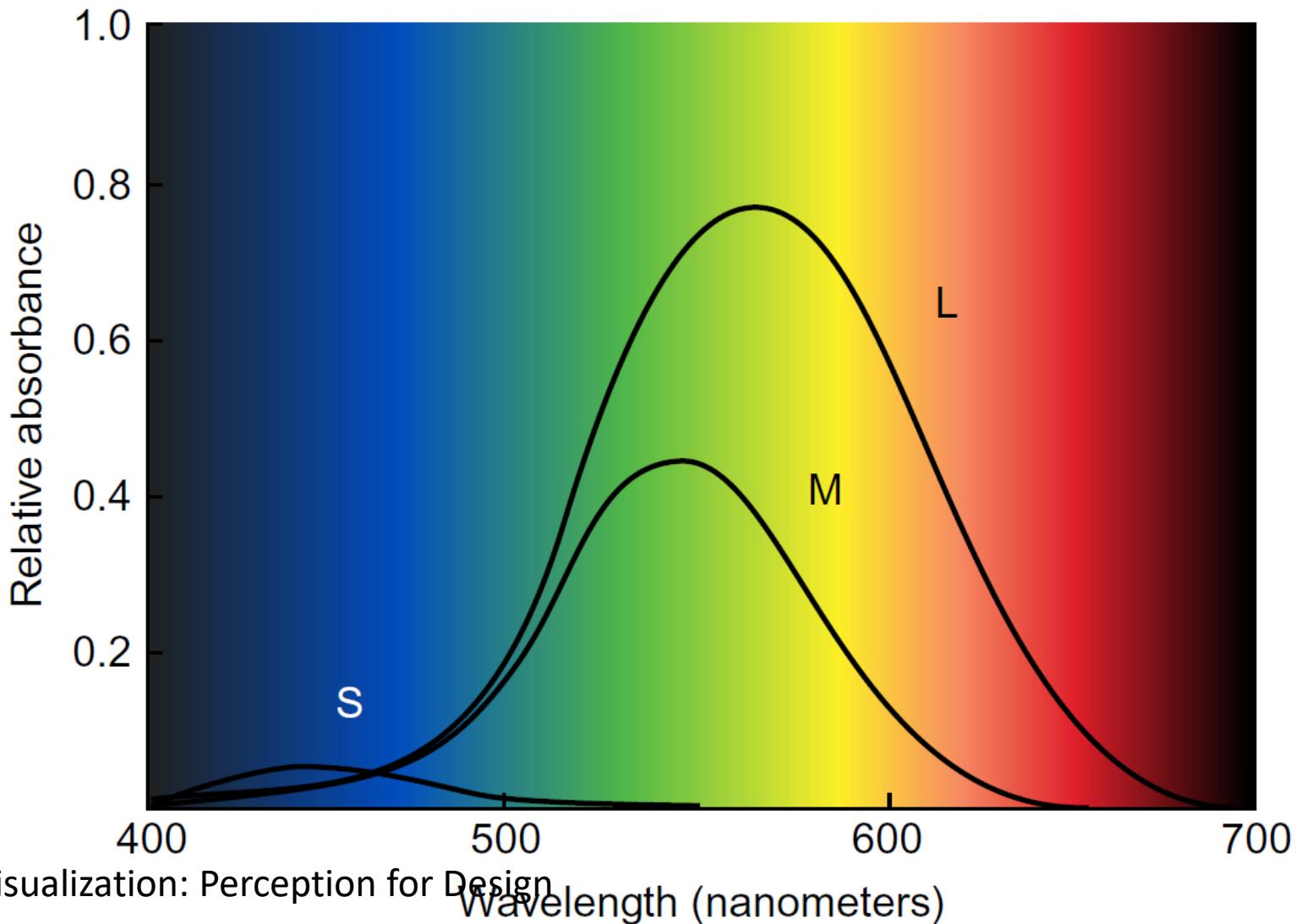
- Approximately 100-150 million rods.
- Non-uniform distribution across the retina
- Sensitive to low-light levels (scotopic vision)

Cones

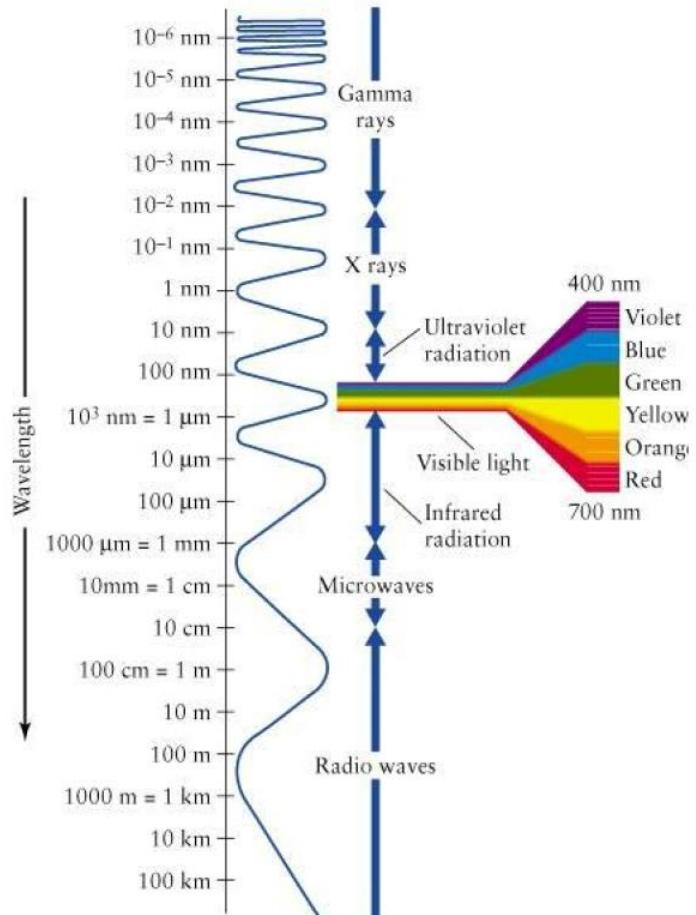
- Approximately 6-7 million cones.
- Sensitive to daytime-light levels (photopic vision)
- Detect color by the use of 3 different kinds:
- Red (L cone) : 564-580nm wavelengths (65% of all cones)
- Green (M cone) : 534-545nm (30% of all cones)
- Blue (S cone) : 420-440nm (5% of all cones)

Cones (SML)

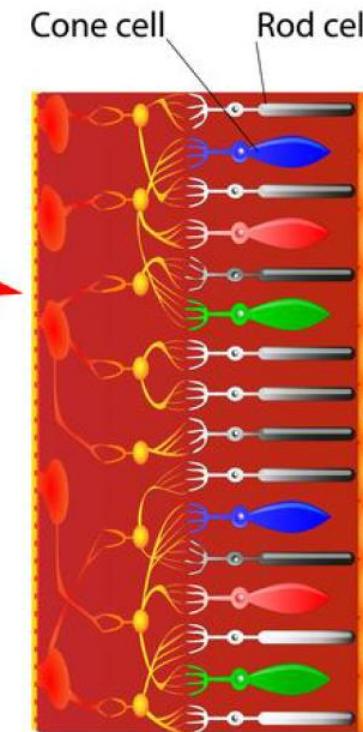
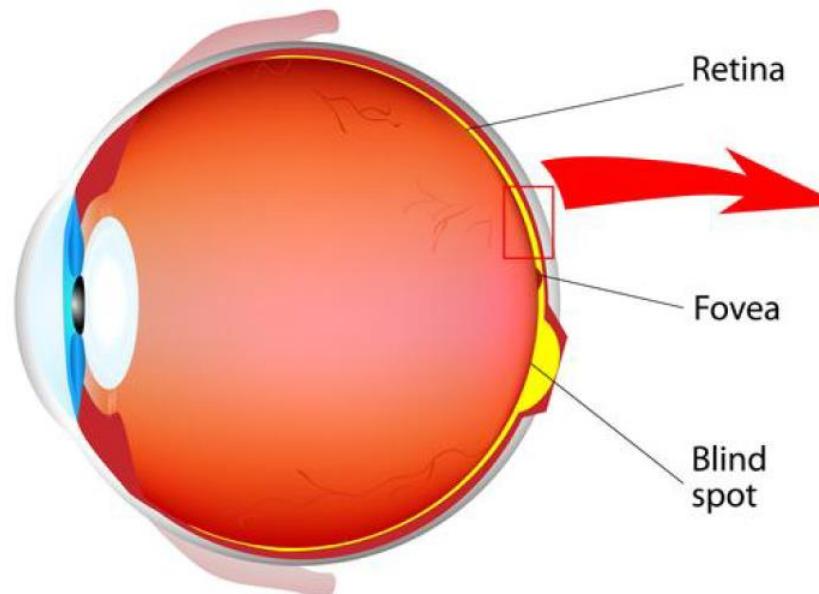
(short, medium, long)



Photoreceptor cell



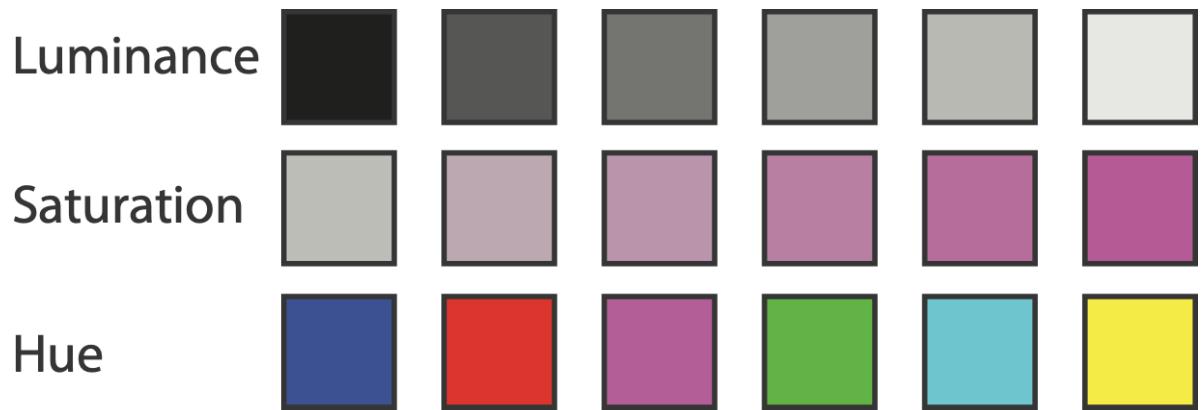
400-700nm



**~120 million rods
~6 million cones**

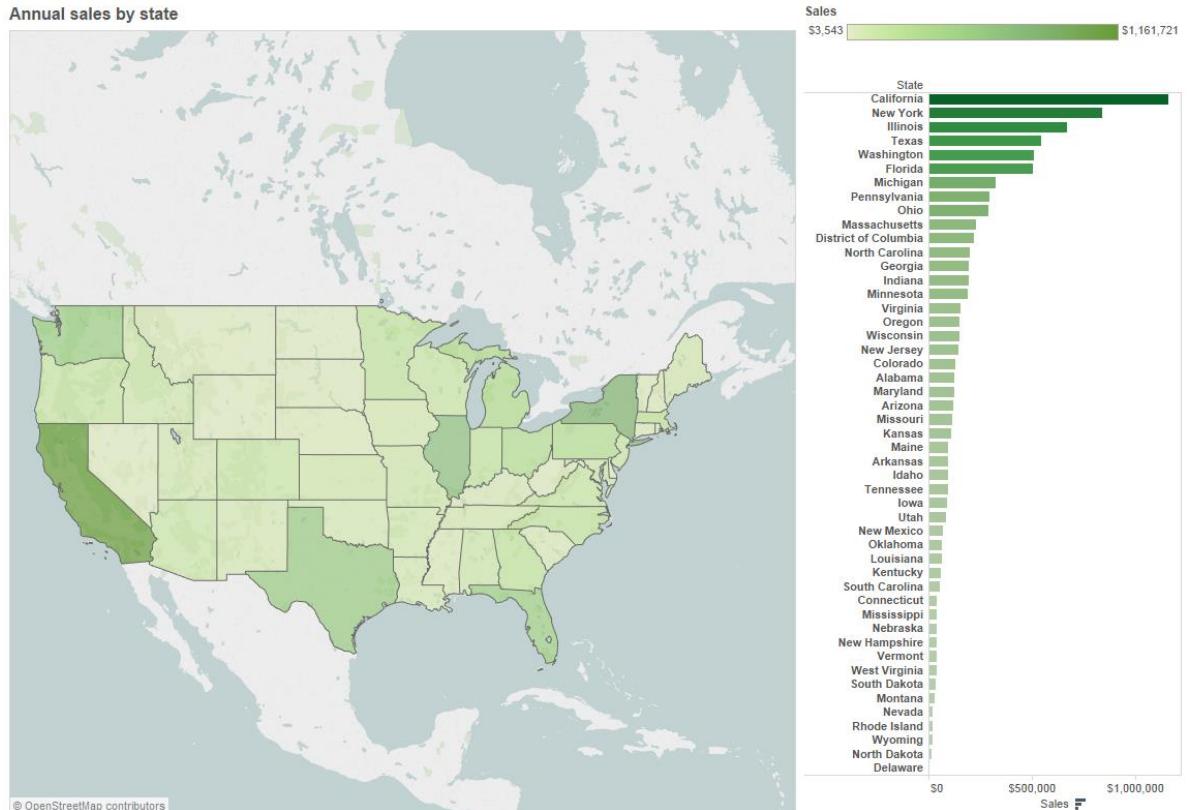
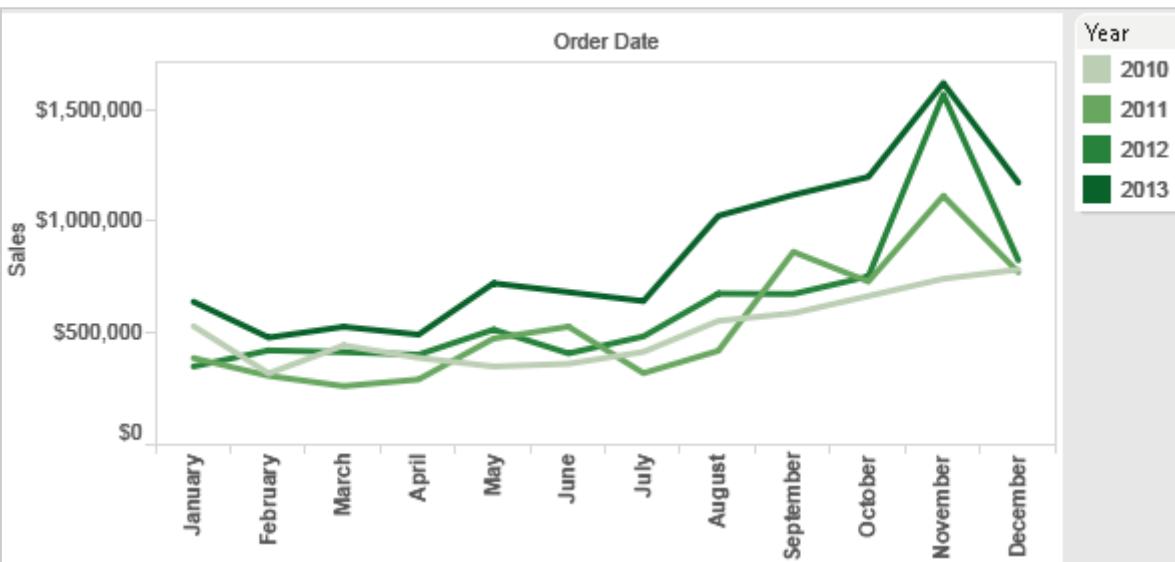
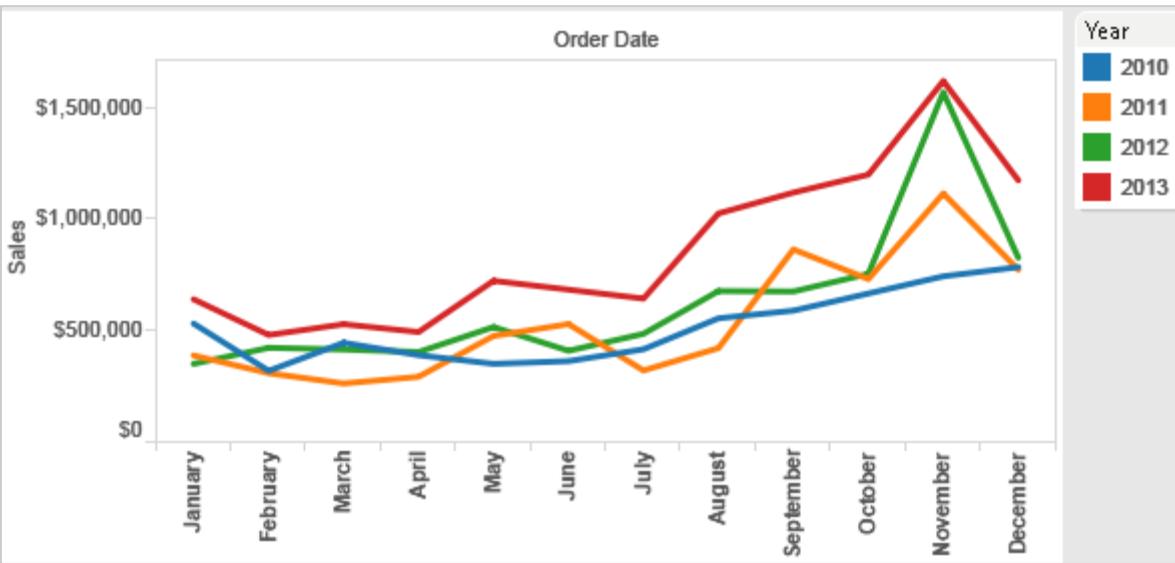
Decomposing color

- first rule of color: do not (just) talk about color!
 - color is confusing if treated as monolithic
- decompose into three channels
 - ordered can show magnitude
 - **luminance**: how bright (B/W)
 - **saturation**: how colorful
 - categorical can show identity
 - **hue**: what color
- channels have different properties
 - what they convey directly to perceptual system
 - how much they can convey
 - how many discriminable bins can we use?



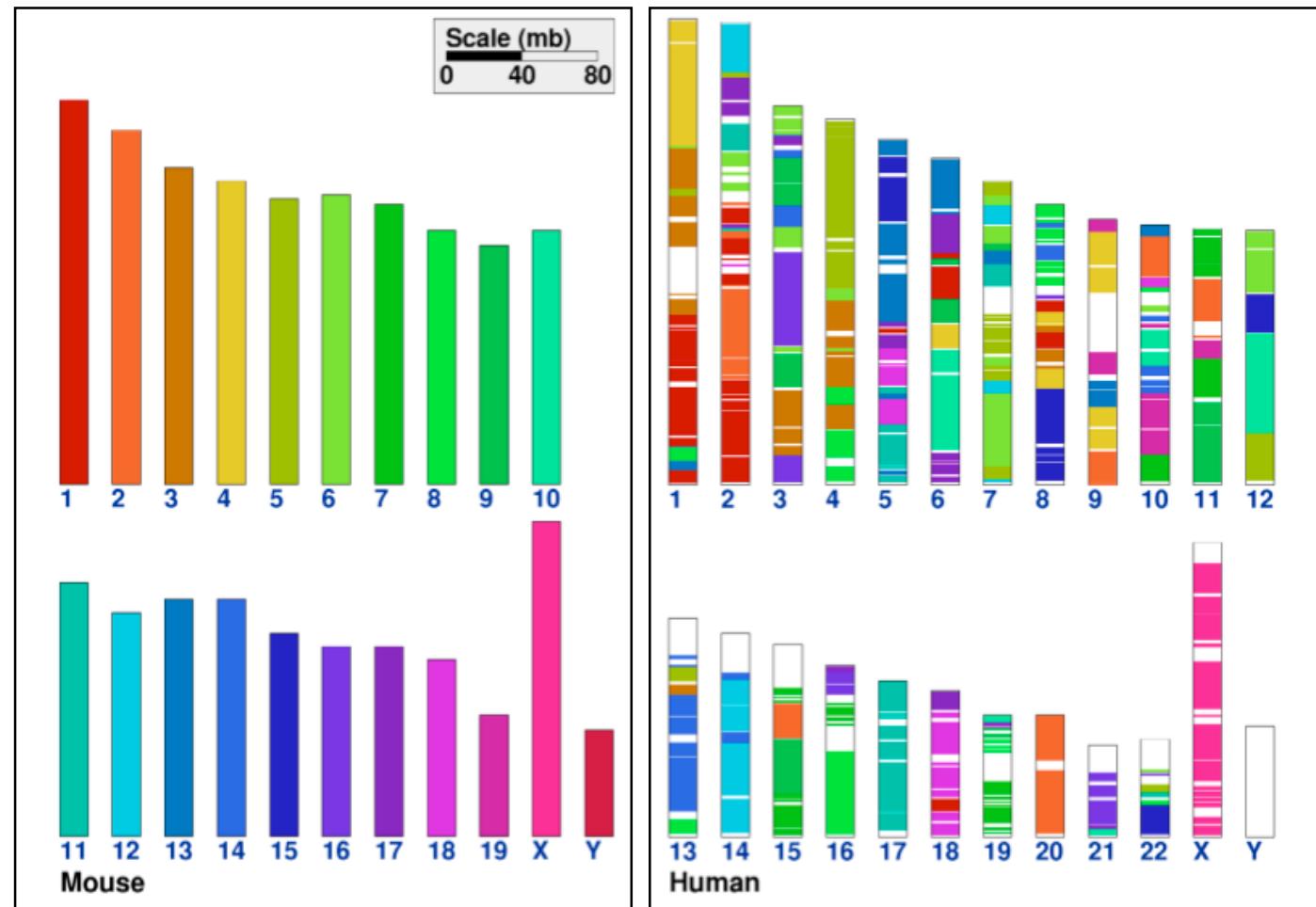
Color Channels in Visualization

Categorical vs ordered color



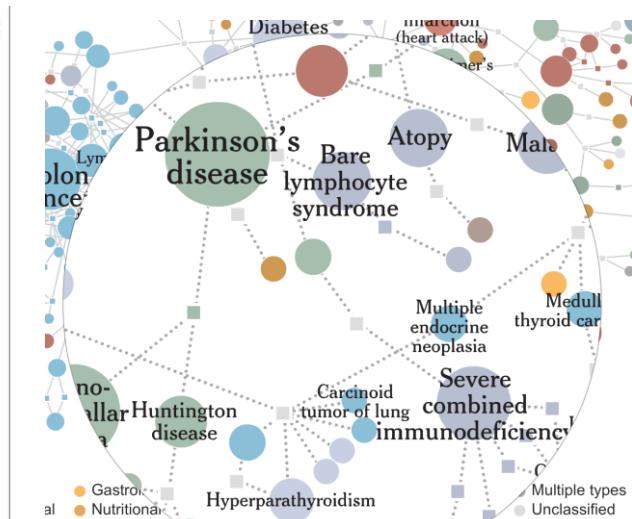
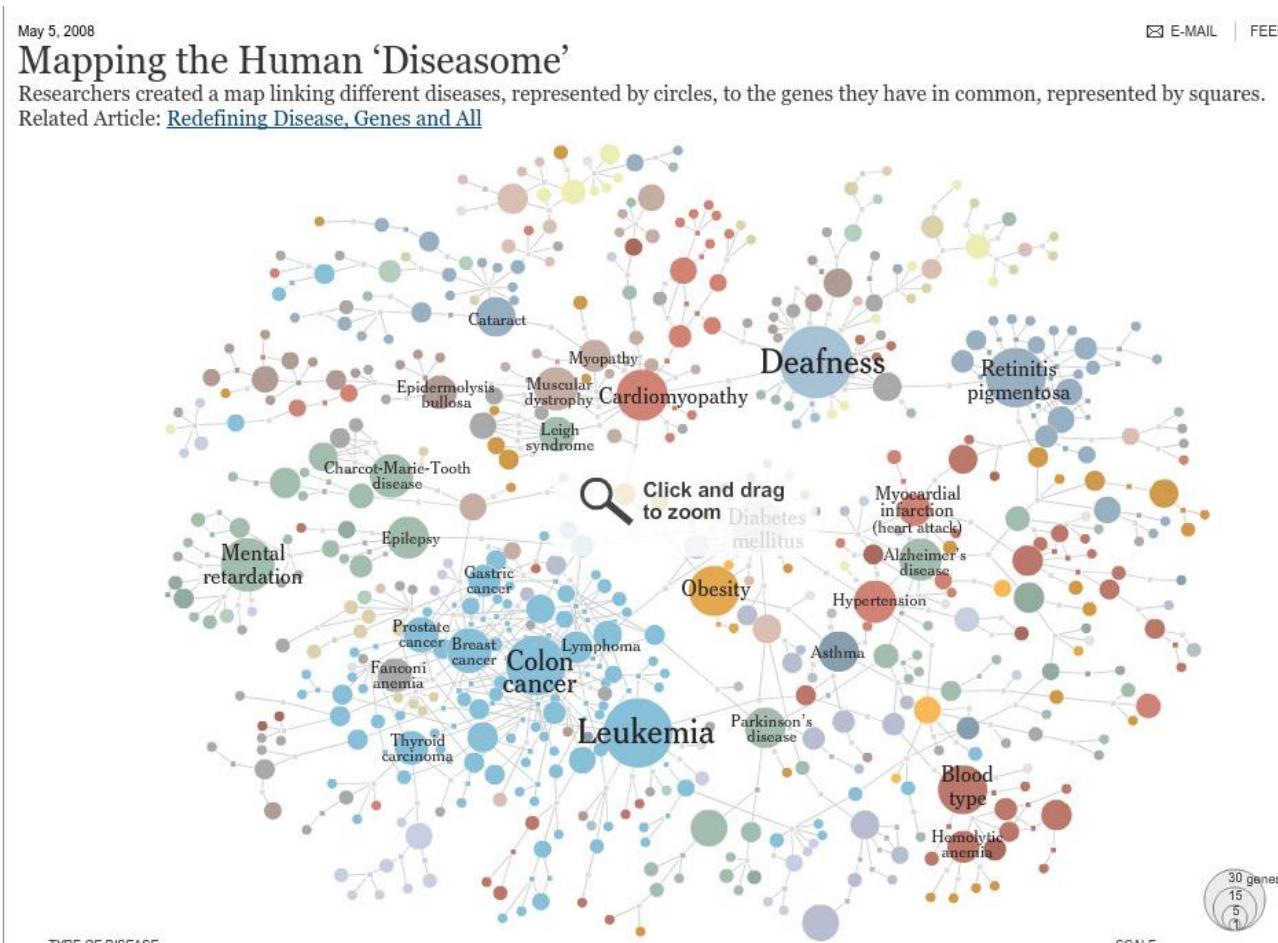
Categorical color: limited number of discriminable bins

- human perception built on relative comparisons
 - great if color contiguous
 - surprisingly bad for absolute comparisons
- noncontiguous small regions of color
 - fewer bins than you want
 - rule of thumb: 6-12 bins, including background and highlights

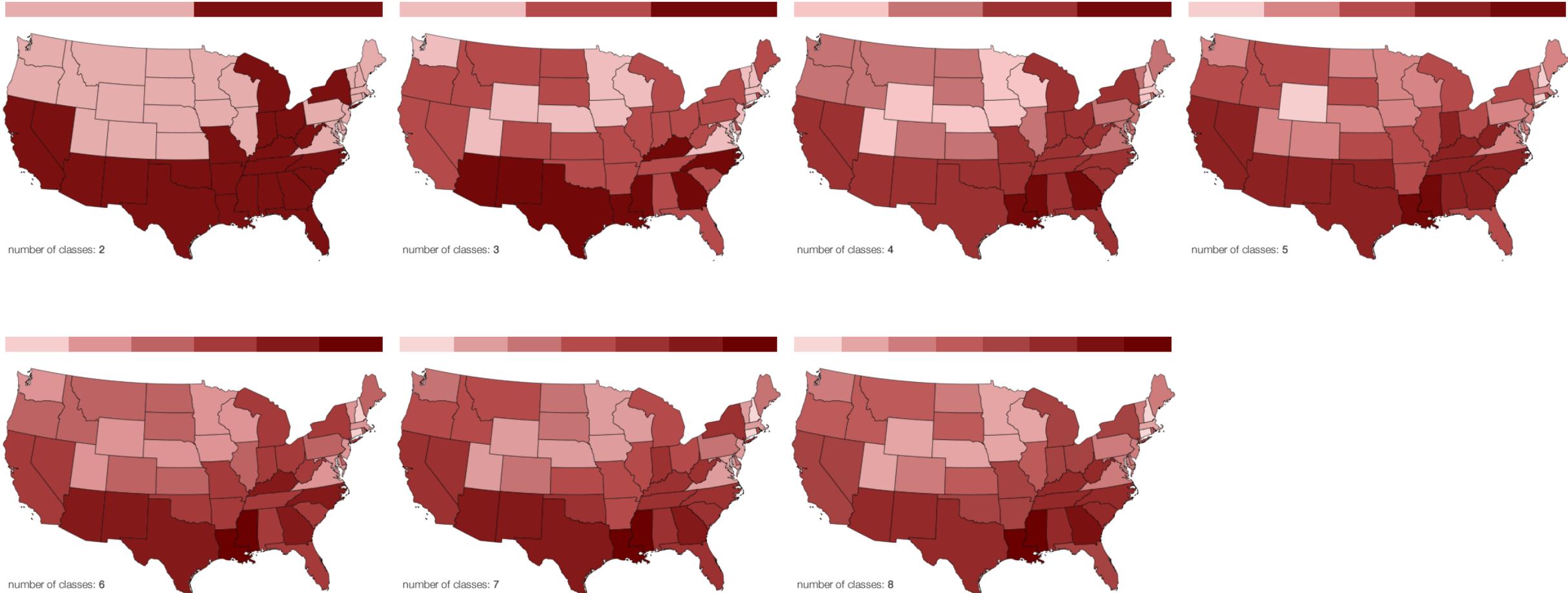


Categorical color: limited number of discriminable bins

-



Ordered color: limited number of discriminable bins



Ordered color: Rainbow is poor default

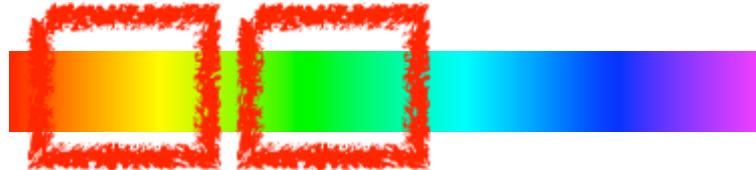
- problems



- perceptually unordered
- perceptually nonlinear

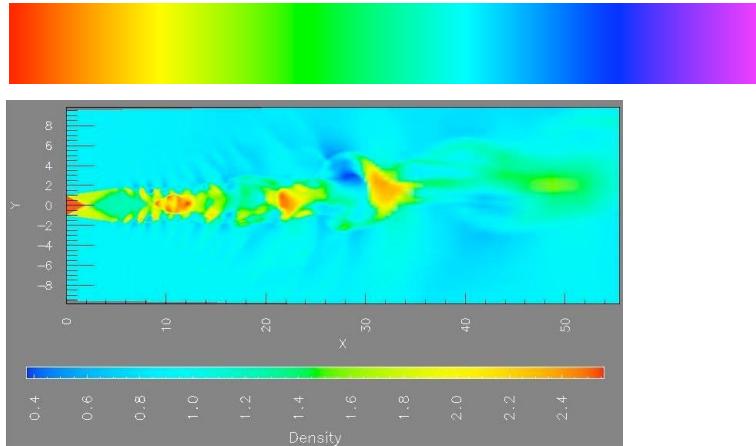
Ordered color: Rainbow is poor default

- problems
 - perceptually unordered
 - perceptually nonlinear

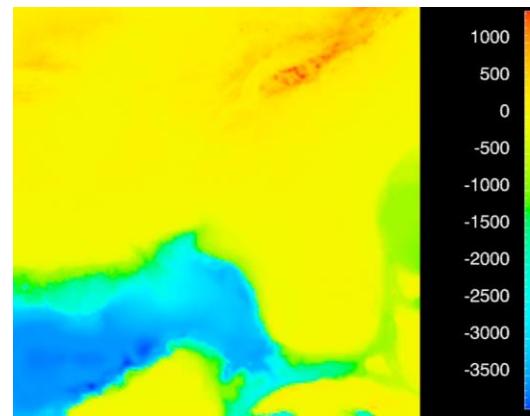


Ordered color: Rainbow is poor default

- problems
 - perceptually unordered
 - perceptually nonlinear
- benefits
 - fine-grained structure visible and nameable



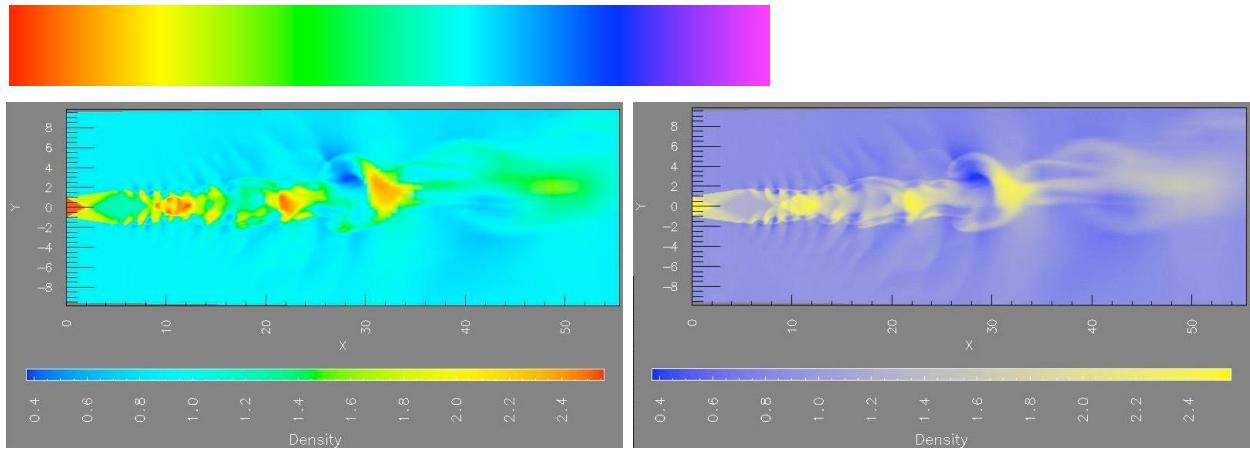
[A Rule-based Tool for Assisting Colormap Selection. Bergman., Rogowitz, and. Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



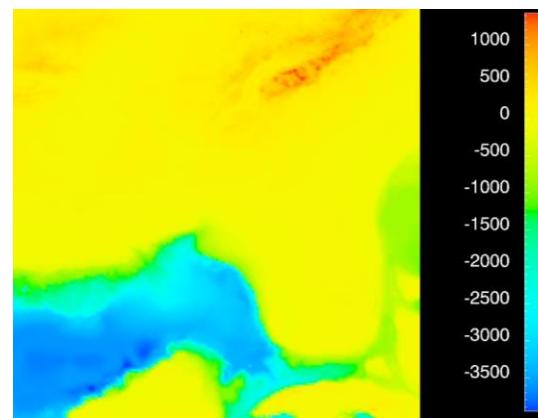
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998.
<http://www.research.ibm.com/people/l/lloyd/color/color.HTM>]

Ordered color: Rainbow is poor default

- problems
 - perceptually unordered
 - perceptually nonlinear
- benefits
 - fine-grained structure visible and nameable
- alternatives
 - large-scale structure: fewer hues



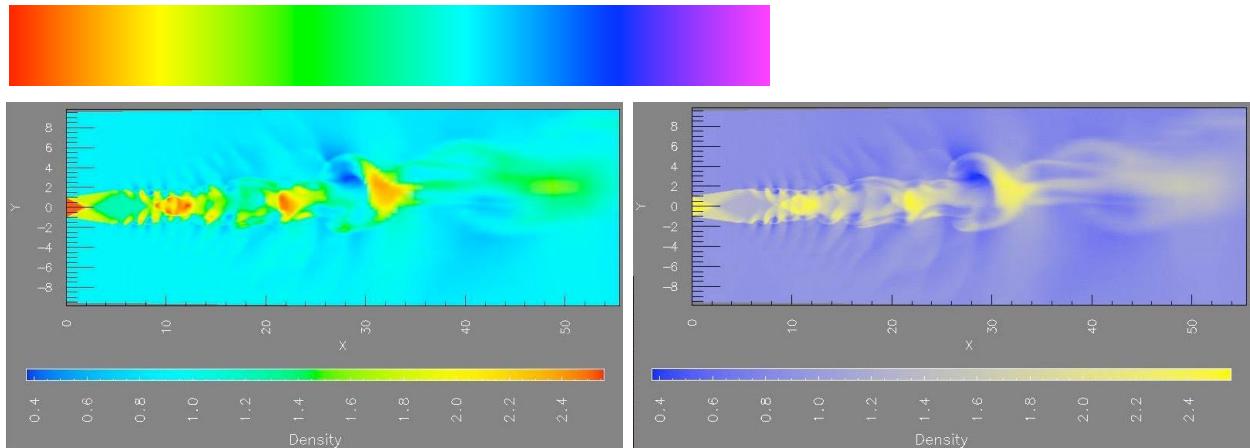
[A Rule-based Tool for Assisting Colormap Selection. Bergman, Rogowitz, and Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



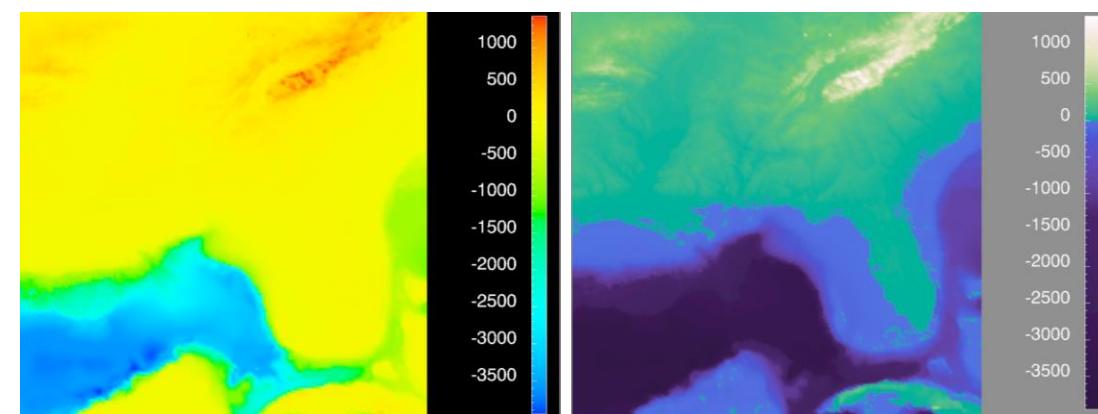
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998.
<http://www.research.ibm.com/people/l/lloyd/color/color.HTM>]

Ordered color: Rainbow is poor default

- problems
 - perceptually unordered
 - perceptually nonlinear
- benefits
 - fine-grained structure visible and nameable
- alternatives
 - large-scale structure: fewer hues
 - fine structure: multiple hues with monotonically increasing luminance [eg viridis]



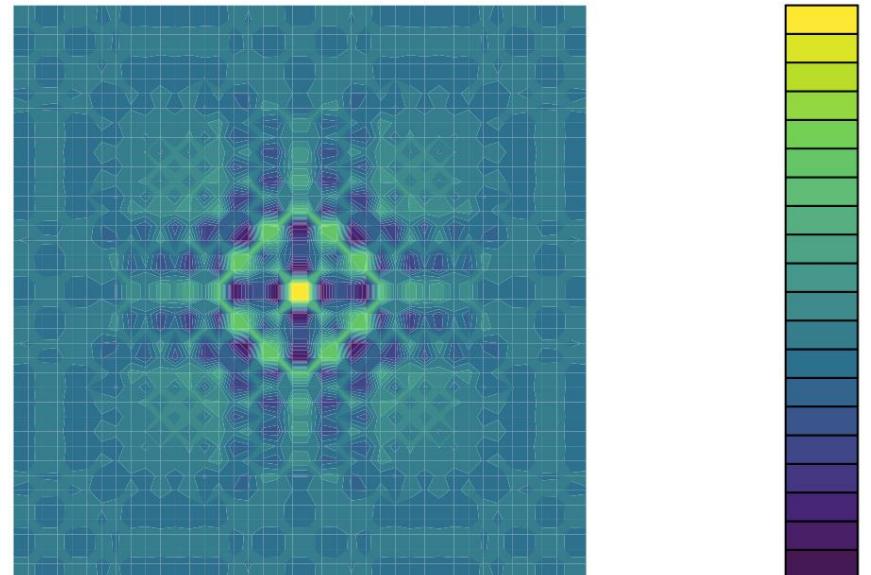
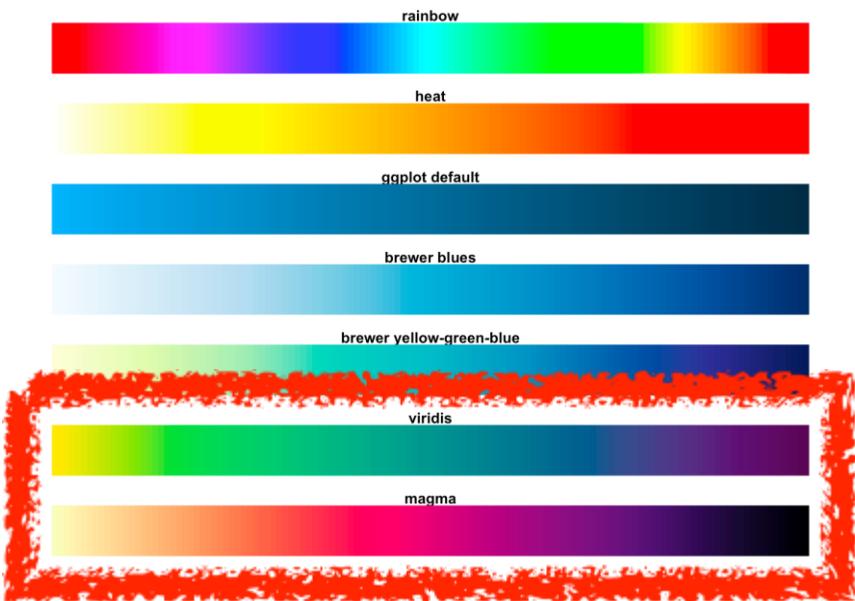
[A Rule-based Tool for Assisting Colormap Selection. Bergman., Rogowitz, and. Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998.
<http://www.research.ibm.com/people/l/lloyd/color/color.HTM>]

Viridis / Magma: sequential colormaps

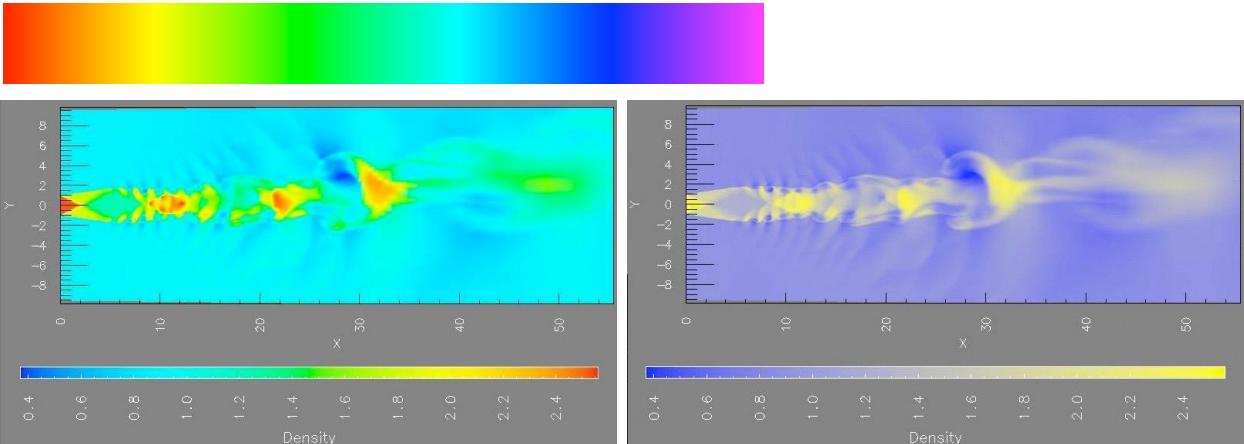
- monotonically increasing luminance, perceptually uniform
- colorful, colorblind-safe



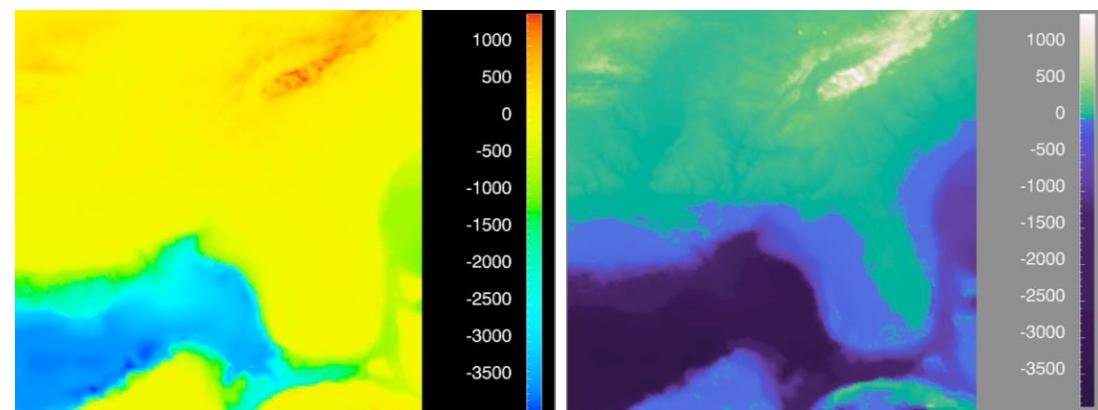
<https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html>

Ordered color: Rainbow is poor default

- problems
 - perceptually unordered
 - perceptually nonlinear
- benefits
 - fine-grained structure visible and nameable
- alternatives
 - large-scale structure: fewer hues
 - fine structure: multiple hues with monotonically increasing luminance [eg viridis]
- legit for categorical
 - segmented saturated rainbow is good!



[A Rule-based Tool for Assisting Colormap Selection. Bergman., Rogowitz, and. Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998.
<http://www.research.ibm.com/people/l/lloyd/color/color.HTM>]



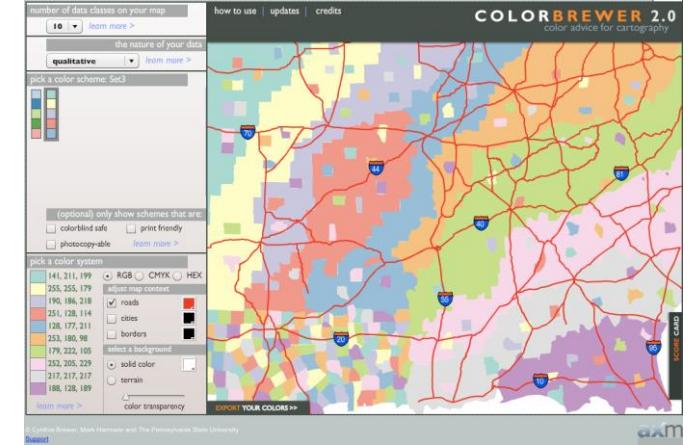
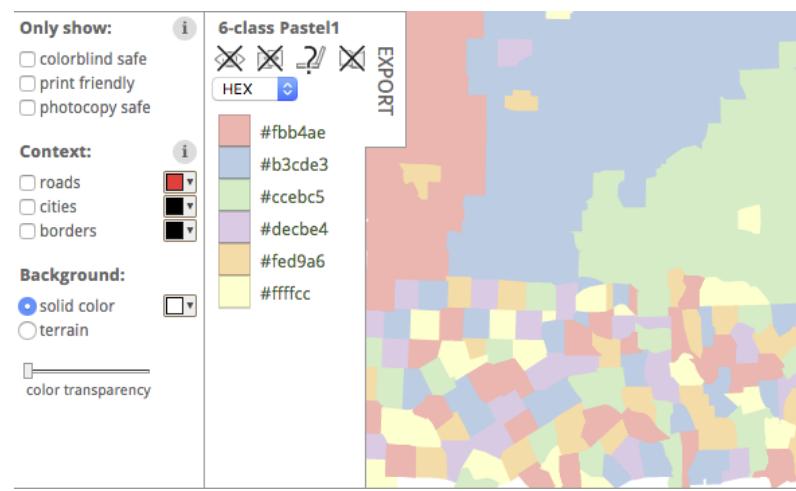
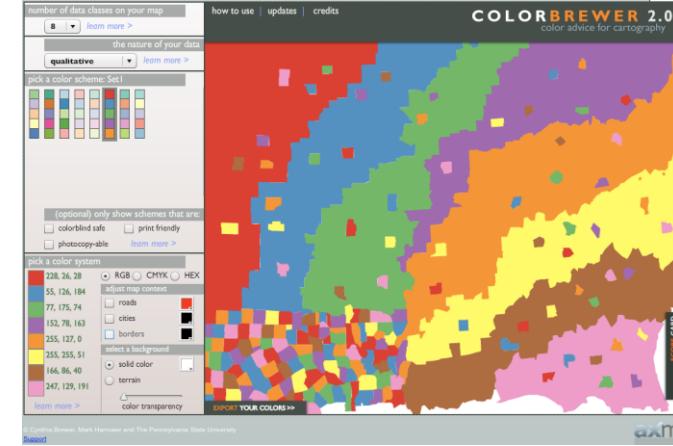
[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Interaction between channels: Not fully separable

- color channel interactions
 - size heavily affects salience
 - small regions need high saturation
 - large regions need low saturation
- saturation & luminance:
 - not separable from each

other!

- also not separable from transparency
- small separated regions: 2 bins safest (use only one of these channels), 3-4 bins max
- contiguous regions: many bins (use only one of these channels)



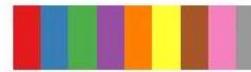
Color Palettes

Color palettes: univariate

→ Categorical



- categorical
 - aim for maximum distinguishability
 - aka *qualitative*, *nominal*



Color palettes: univariate

→ Categorical



→ Ordered

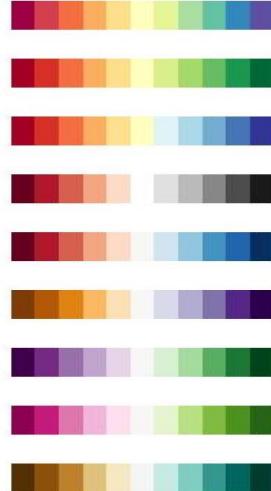
→ *Sequential*

→ *Diverging*



- diverging
 - useful when data has meaningful "midpoint"
 - use neutral color for midpoint
 - white, yellow, grey
 - use saturated colors for endpoints
- sequential
 - ramp luminance or saturation
 - if multi-hue, good to order by luminance

diverging



sequential



Cividis

Viridis

Inferno

Magma

Plasma

Warm

Cool

CubehelixDefault

Color palettes: univariate

→ Categorical



→ Ordered

→ *Sequential*



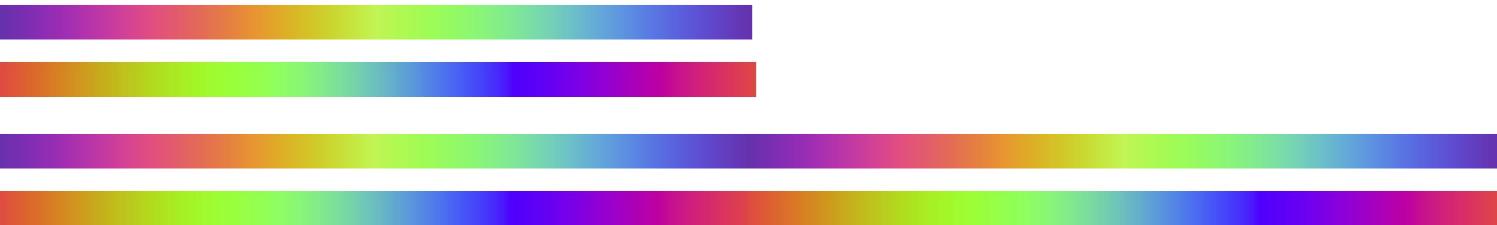
→ *Diverging*



→ Cyclic



cyclic multihue

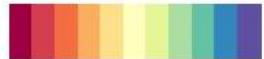


<https://github.com/d3/d3-scale-chromatic>

Color palette design considerations: univariate

segmented

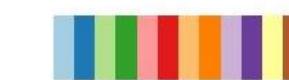
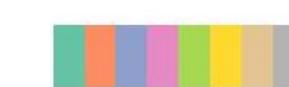
diverging



sequential



categorical



continuous

(a)

(b)

(c)

cyclic multihue



- segmented or continuous?
- diverging or sequential or cyclic?
- single-hue or two-hue or multi-hue?
- perceptually linear?
- ordered by luminance?
- colorblind safe?

sequential
single hue

diverging
two hue

sequential
multihue

Colormaps: bivariate

→ Categorical



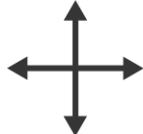
→ Ordered

→ Sequential

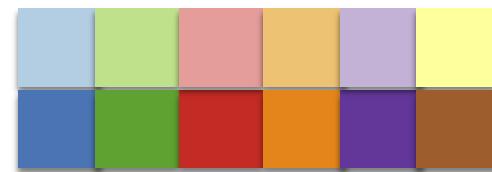
→ Diverging



→ Bivariate



- bivariate best case
 - binary in one of the directions



binary saturation

categorical hue

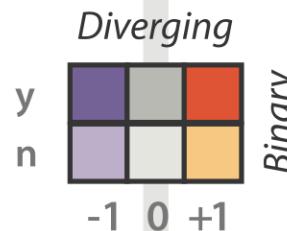
Binary



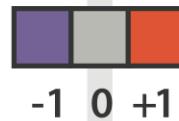
Categorical



Binary



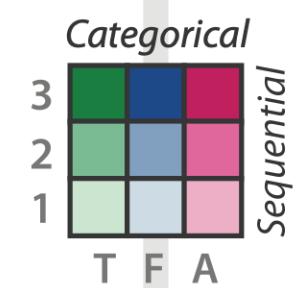
Diverging



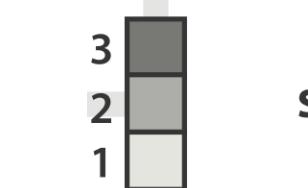
Binary



Categorical



Categorical



Sequential

Colormaps: bivariate

→ Categorical



→ Ordered

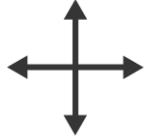
→ *Sequential*



→ Diverging



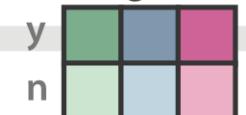
→ Bivariate



Binary

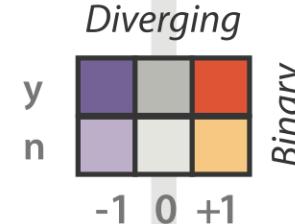


Categorical

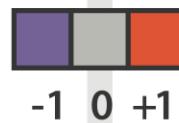


Categorical

Diverging



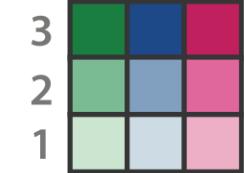
Diverging



y |



Categorical



Sequential

Sequential

Colormaps

→ Categorical



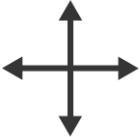
→ Ordered

→ Sequential

→ Diverging

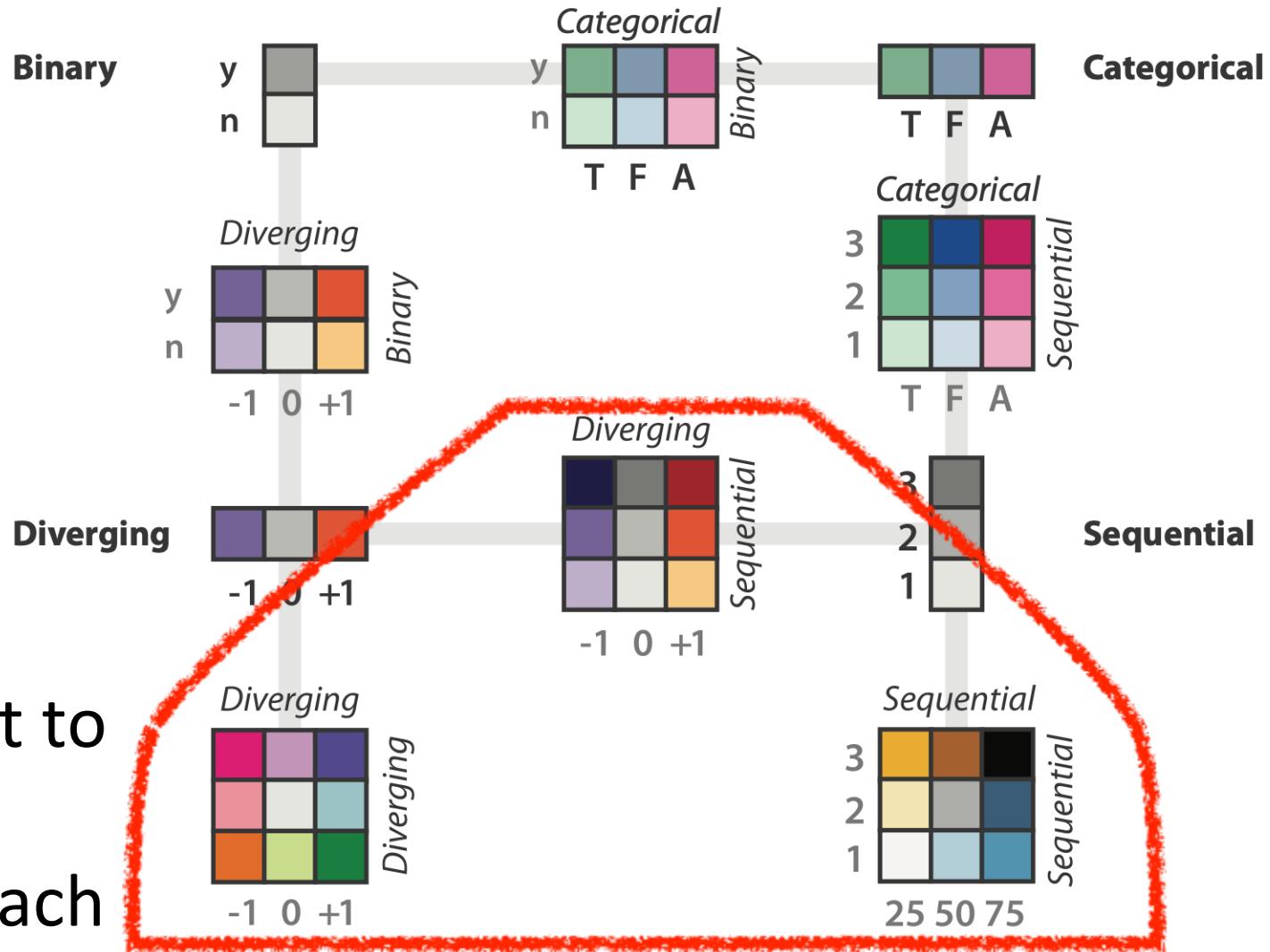


→ Bivariate



use with care!

- bivariate can be very difficult to interpret
 - when multiple levels in each direction



Number of data classes: 3

[how to use](#) | [updates](#) | [downloads](#) | [credits](#)

COLORBREWER 2.0

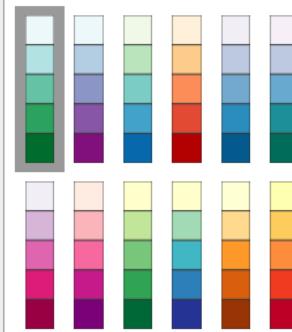
color advice for cartography

Nature of your data:

sequential diverging qualitative

Pick a color scheme:

Multi-hue:



Single hue:



Only show:

- colorblind safe
- print friendly
- photocopy safe

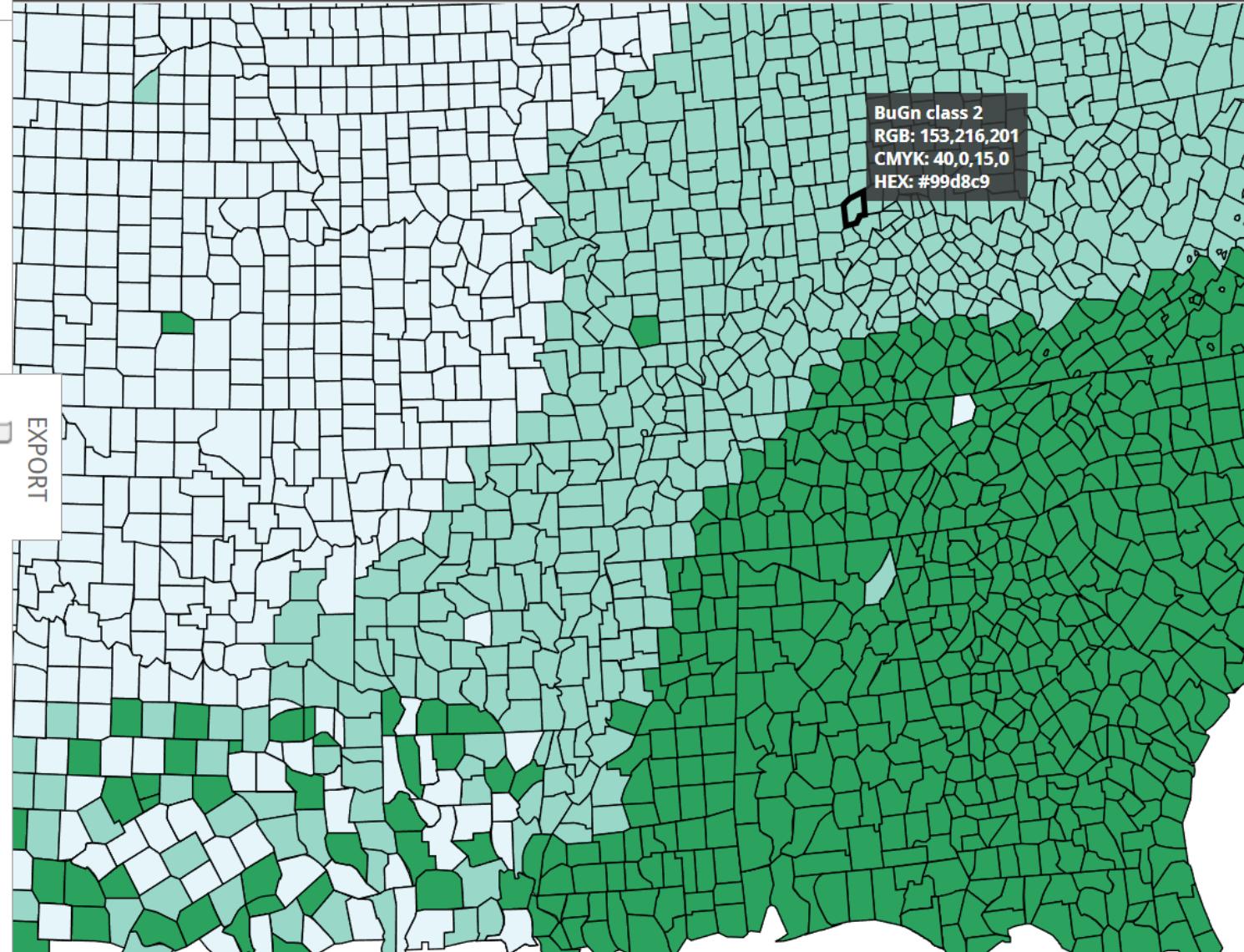
Context:

- roads
- cities
- borders

Background:

- solid color
- terrain

color transparency





i want hue

Colors for data scientists.

Generate and refine palettes of optimally distinct colors.

Color space

Default preset

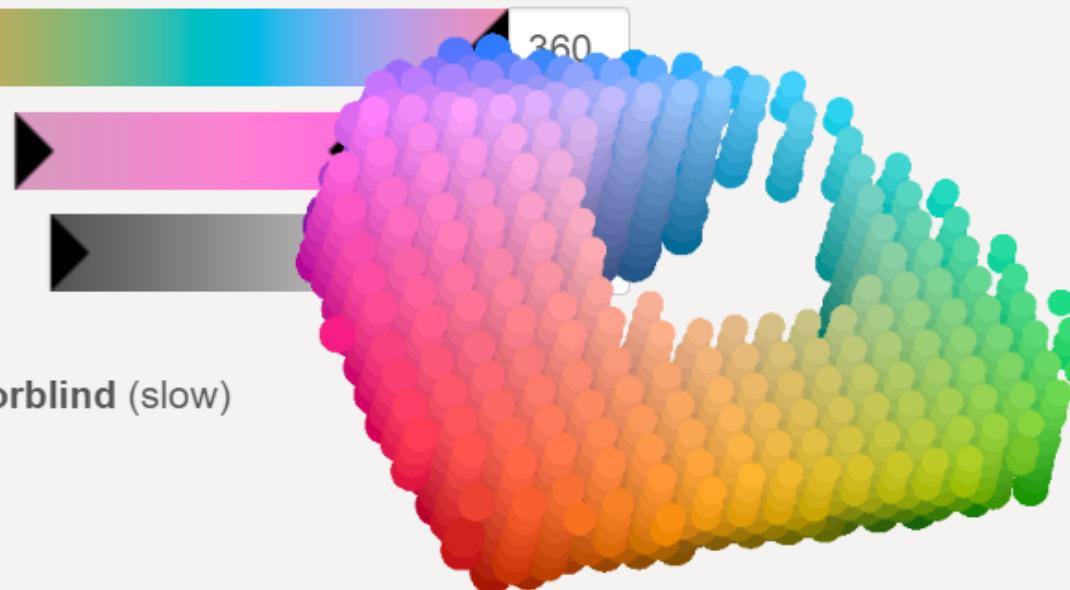
H 0

C 30

L 35

Improve for the **colorblind** (slow)

Dark background



Palette

5

cold

soft (k-Mea

Make a palette



<https://medialab.github.io/iwanthue/>

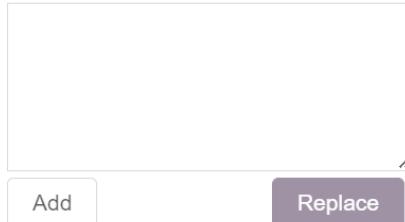
Data Visualization - Palettes

VIZ PALETTE

By: Elijah Meeks
& Susie Lu

PICK

Use Chroma.js



Add

Replace

Use Colorgorical

Use ColorBrewer

EDIT

7 Colors

- ≡ 1 ● #ffd700
- ≡ 2 ● #ffb14e
- ≡ 3 ● #fa8775
- ≡ 4 ● #ea5f94
- ≡ 5 ● #cd34b5
- ≡ 6 ● #9d02d7
- ≡ 7 ● #0000ff

Add

hex rgb

hsl

GET

hex rgb

hsl

- String quotes
- Object with metadata

```
[ "#ffd700",
  "#ffb14e",
  "#fa8775",
  "#ea5f94",
  "#cd34b5",
  "#9d02d7",
  "#0000ff" ]
```

COLORS IN ACTION

Background color: `#ffffff`

Font color: `● #000000`

Charts made with
[Semiotic](#)

Color Population:

No Color Deficiency - 96%

Deuteranomaly - 2.7%

Protanomaly - 0.66%

Protanopia - 0.59%

Deuteranopia - 0.56%

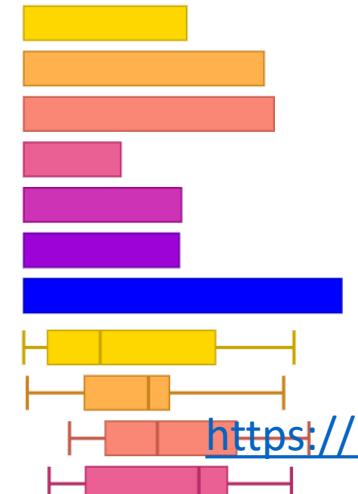
Greyscale

Sample font

Randomize Data

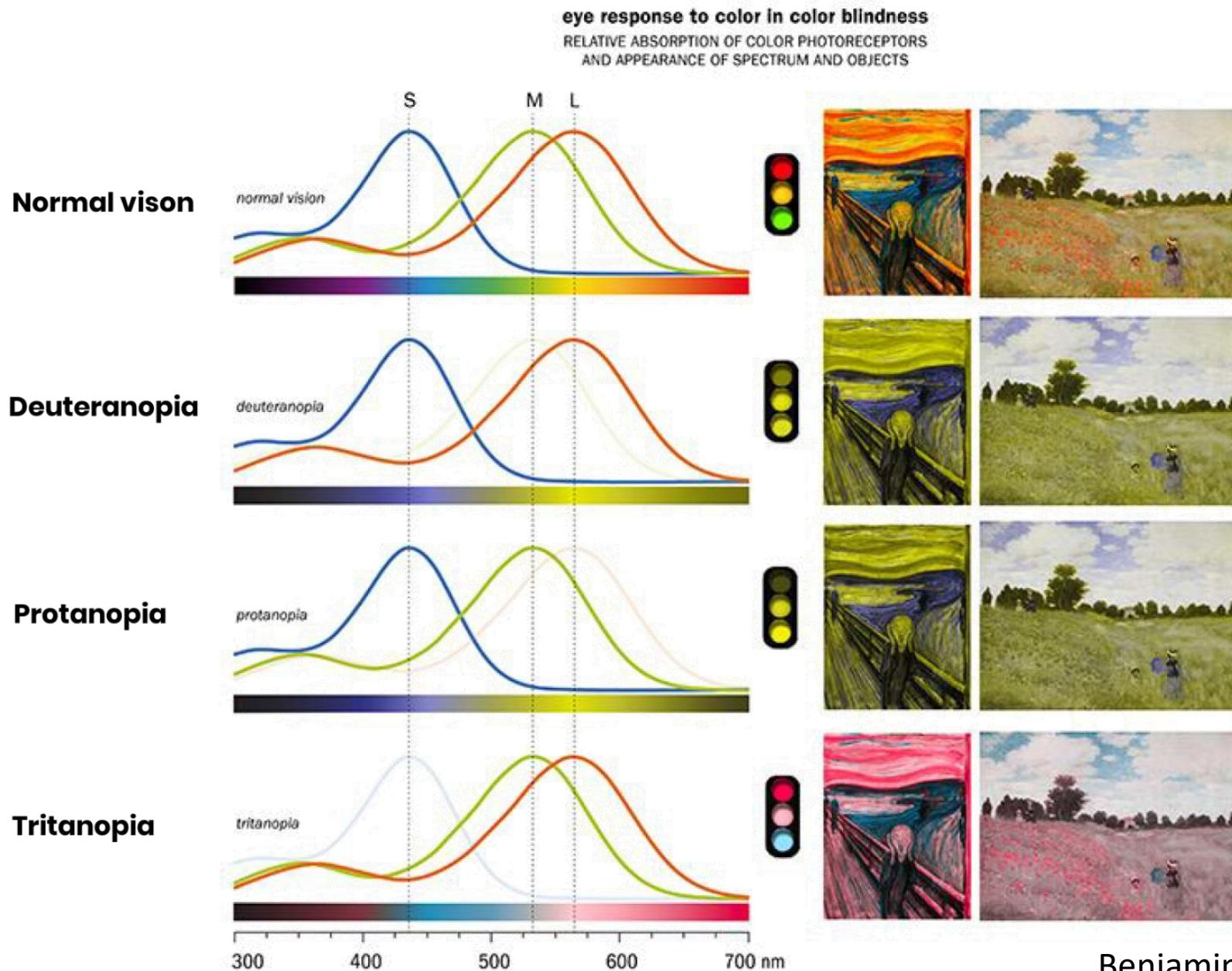
Stroke:

Dark None



<https://projects.susielu.com/viz-palette>

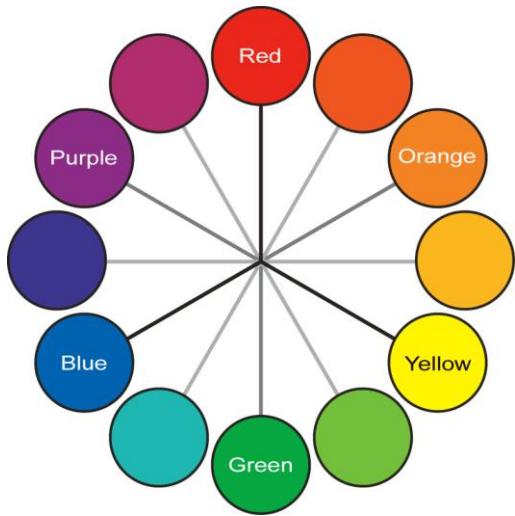
Color deficiency aka Color blindness or color vision deficiency (CVD)



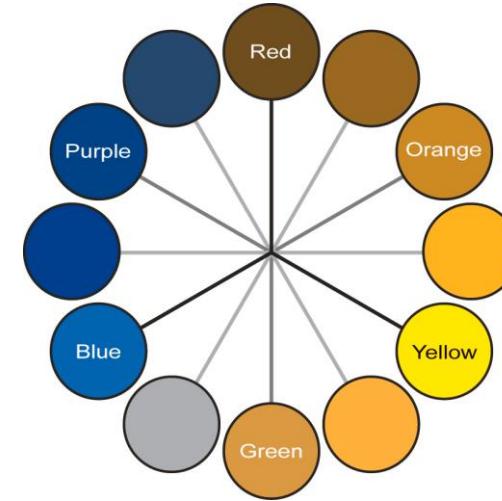
Benjamin Bach, University of Edinburgh

<http://mkweb.bcgsc.ca/colorblind>

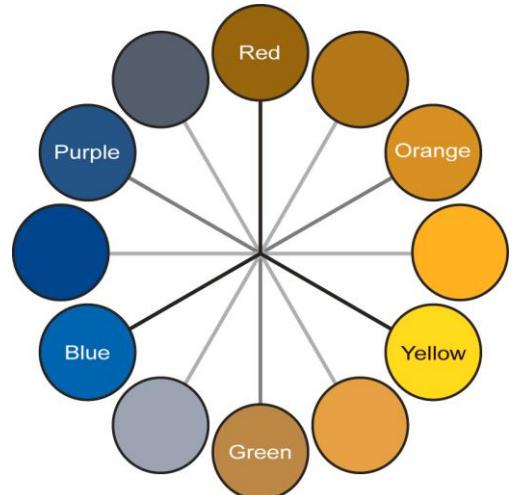
Color deficiency: Reduces color to 2 dimensions



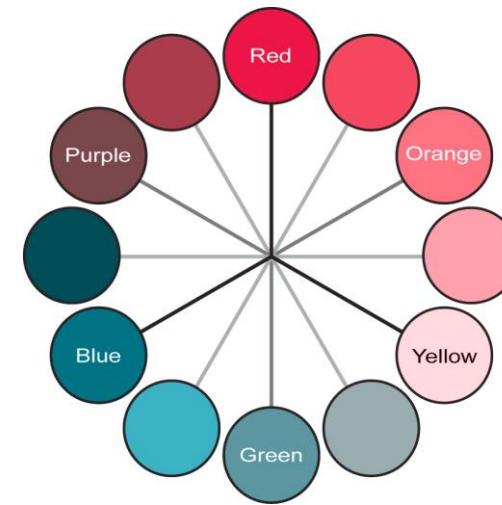
Normal



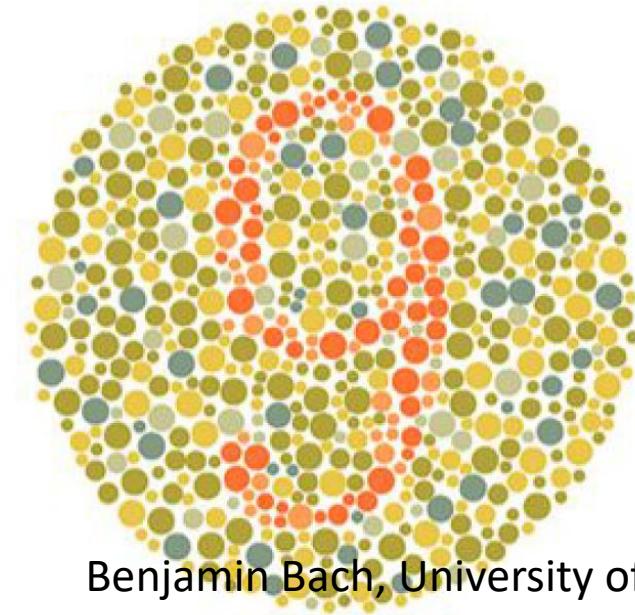
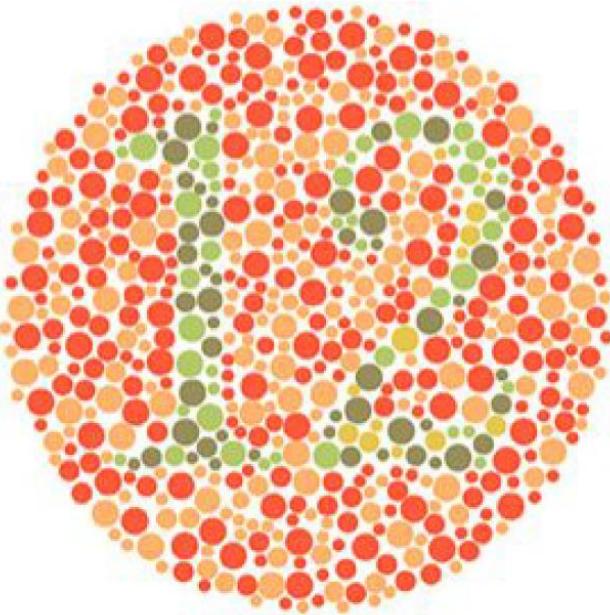
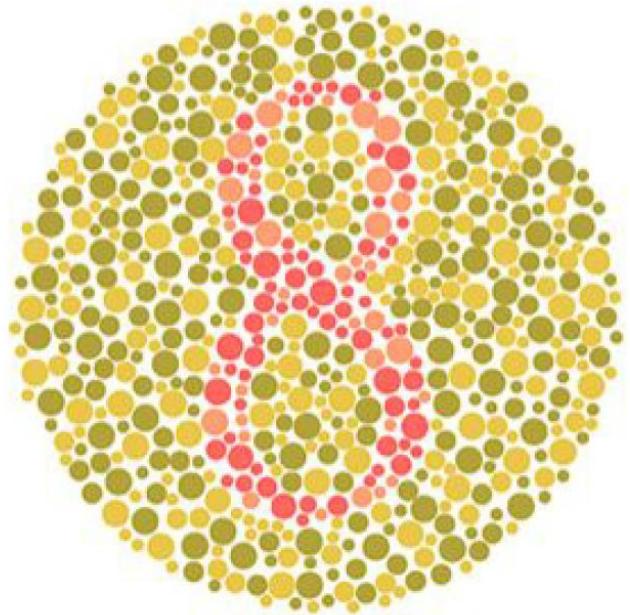
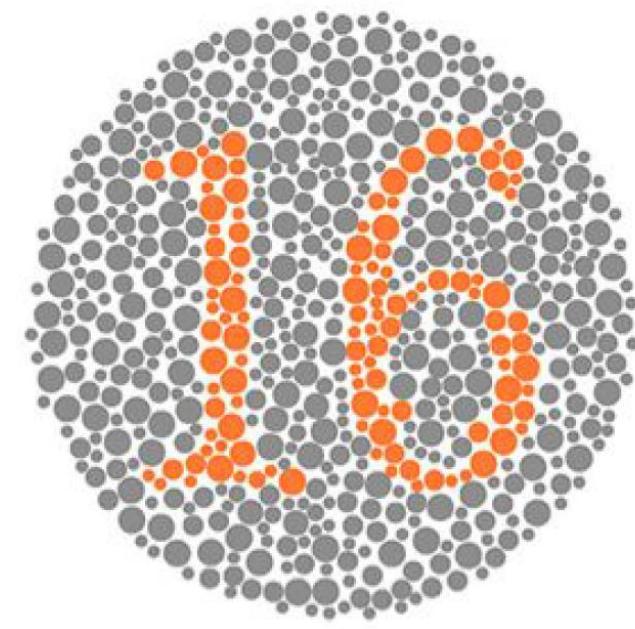
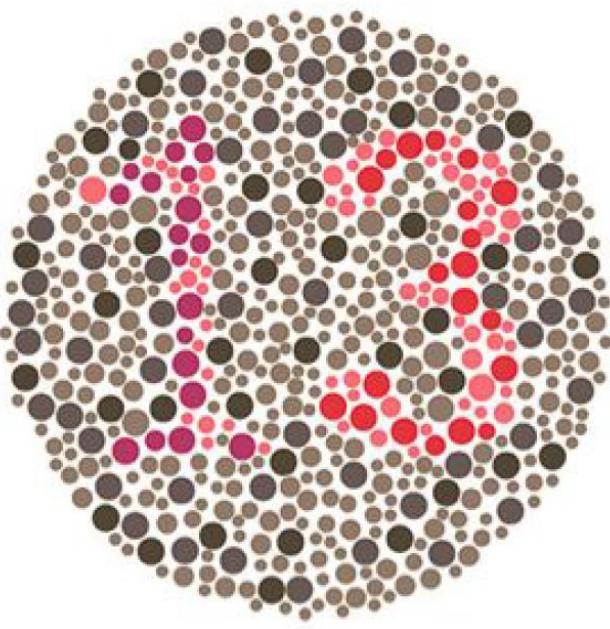
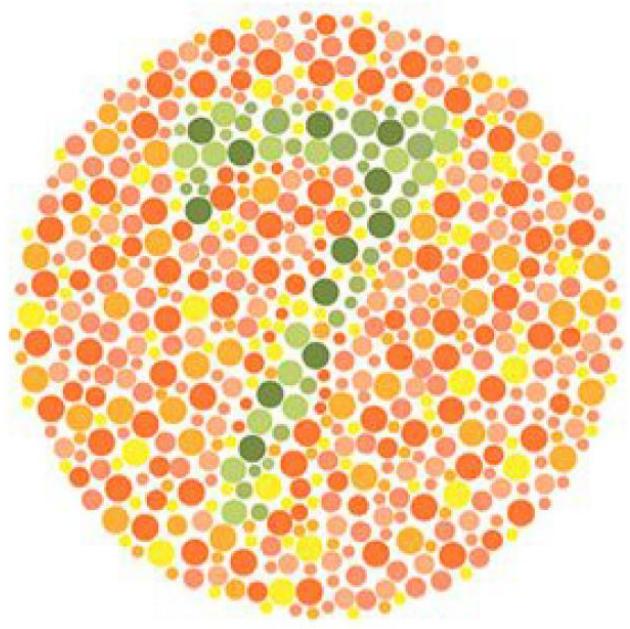
Protanope



Deutanope



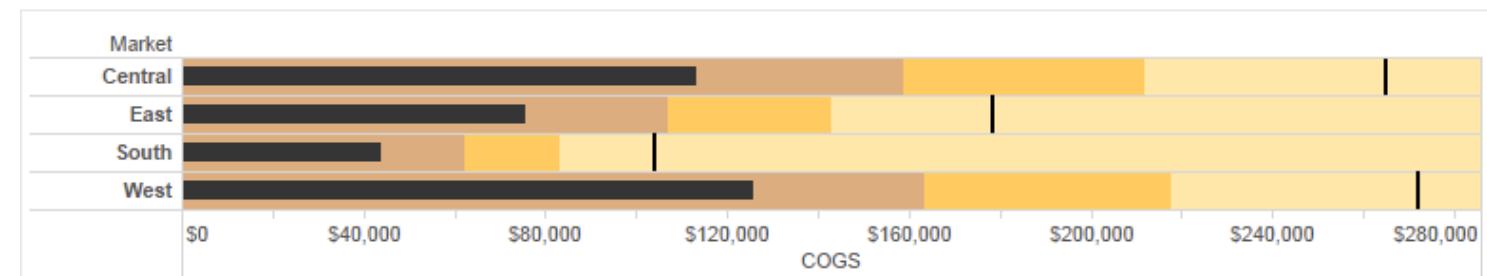
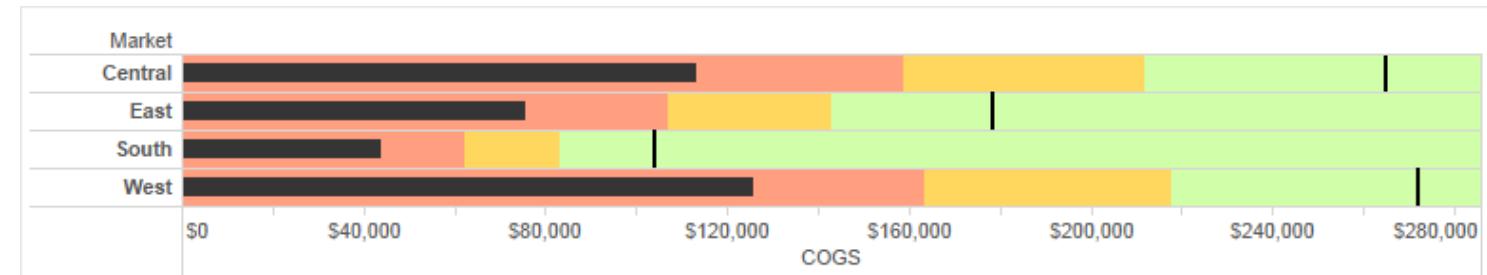
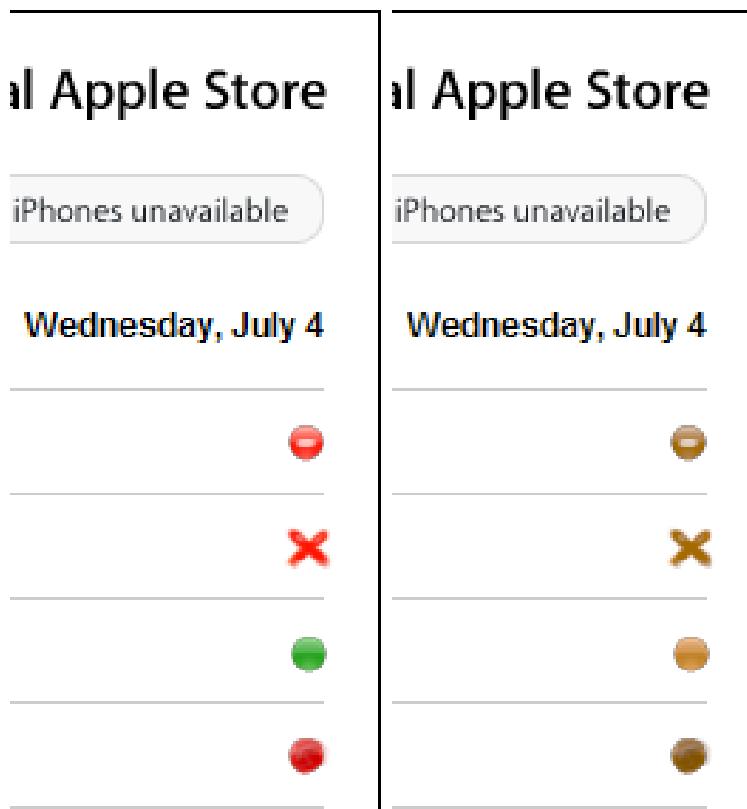
Tritanope



Benjamin Bach, University of Edinburgh

Designing for color deficiency: Avoid encoding by hue alone

- redundantly encode
 - vary luminance
 - change shape



Deuteranope simulation

Change the shape

Vary luminance

Color deficiency

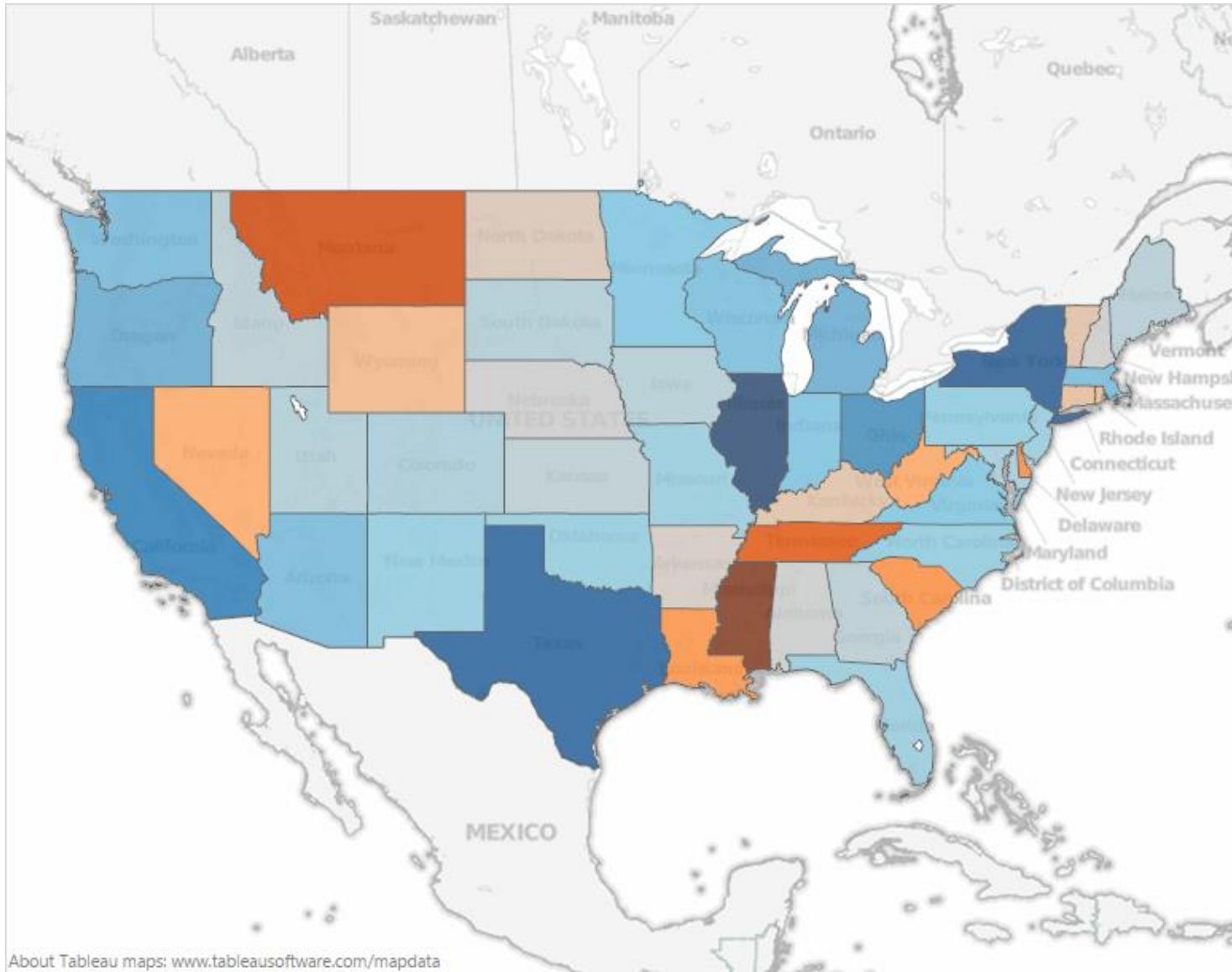
- Population of northern european descent

		Protanopia	Deutanopia	Tritanopia
Men	91.4%	2.45%	6.1%	0.011%
Women	99.6%	0.04%	0.36%	0.04%
Overall	95.5%	1.25%	3.24%	0.025%

The figure consists of four square color patches arranged horizontally. Each patch contains the same set of colored squares: Red, Orange, Yellow, Green, Blue, and Magenta. The colors are rendered differently in each patch to demonstrate how they appear to someone with specific color deficiencies:

- Protanopia:** Shows Red, Orange, and Yellow appearing darker or more muted compared to the original.
- Deutanopia:** Shows Red, Orange, and Yellow appearing darker or more muted, and Green appearing slightly darker than the original.
- Tritanopia:** Shows Red, Orange, and Yellow appearing darker or more muted, and Green appearing significantly darker than the original.

Designing for color deficiency: Blue-Orange is safe



[Seriously Colorful: Advanced Color Principles & Practices. Stone.Tableau Customer Conference 2014.]