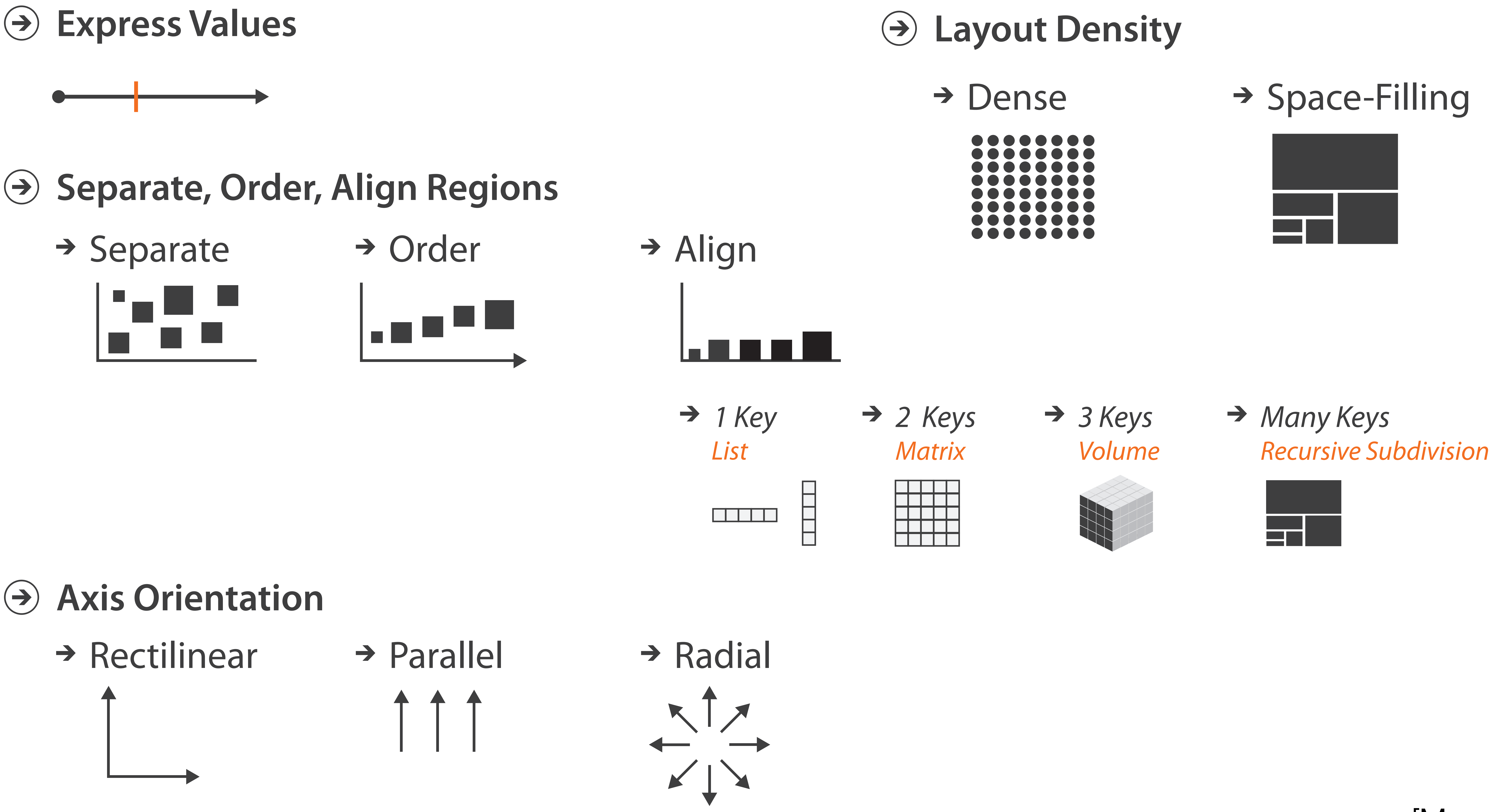
Data Visualization (CSCI 627/490)

Color

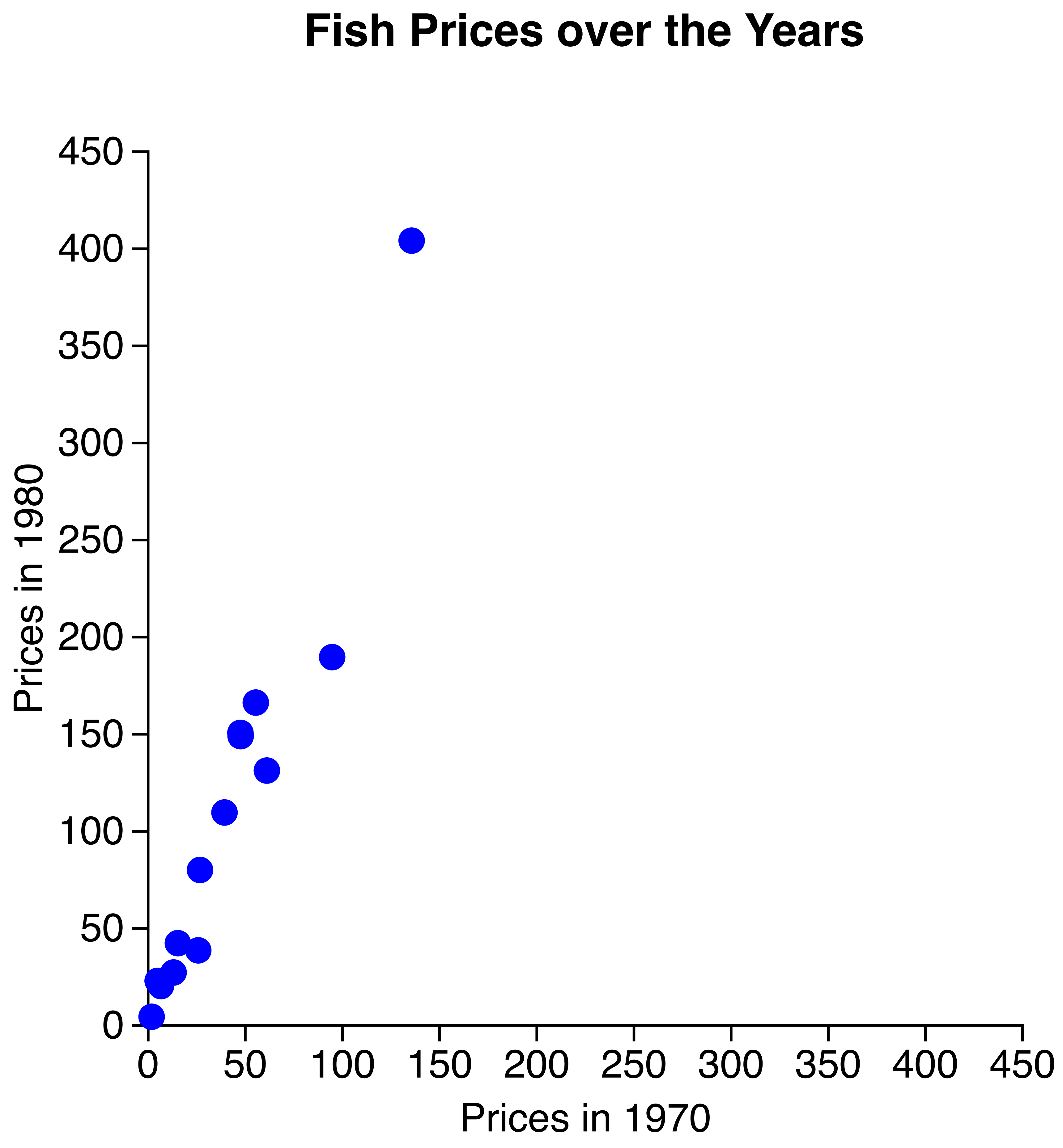
Dr. David Koop

Arrange Tables



[Munzner (ill. Maguire), 2014]

Express Values: Scatterplots

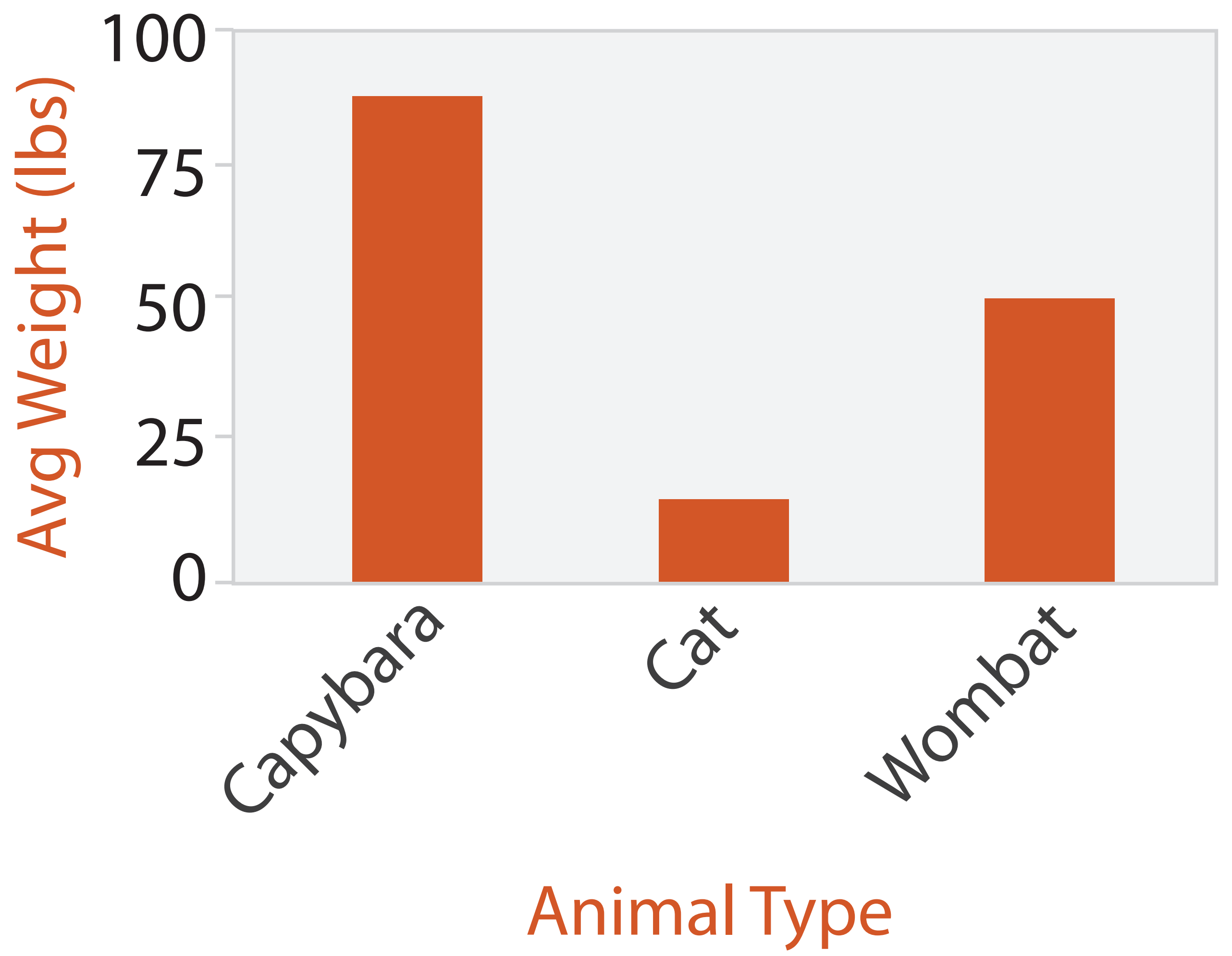


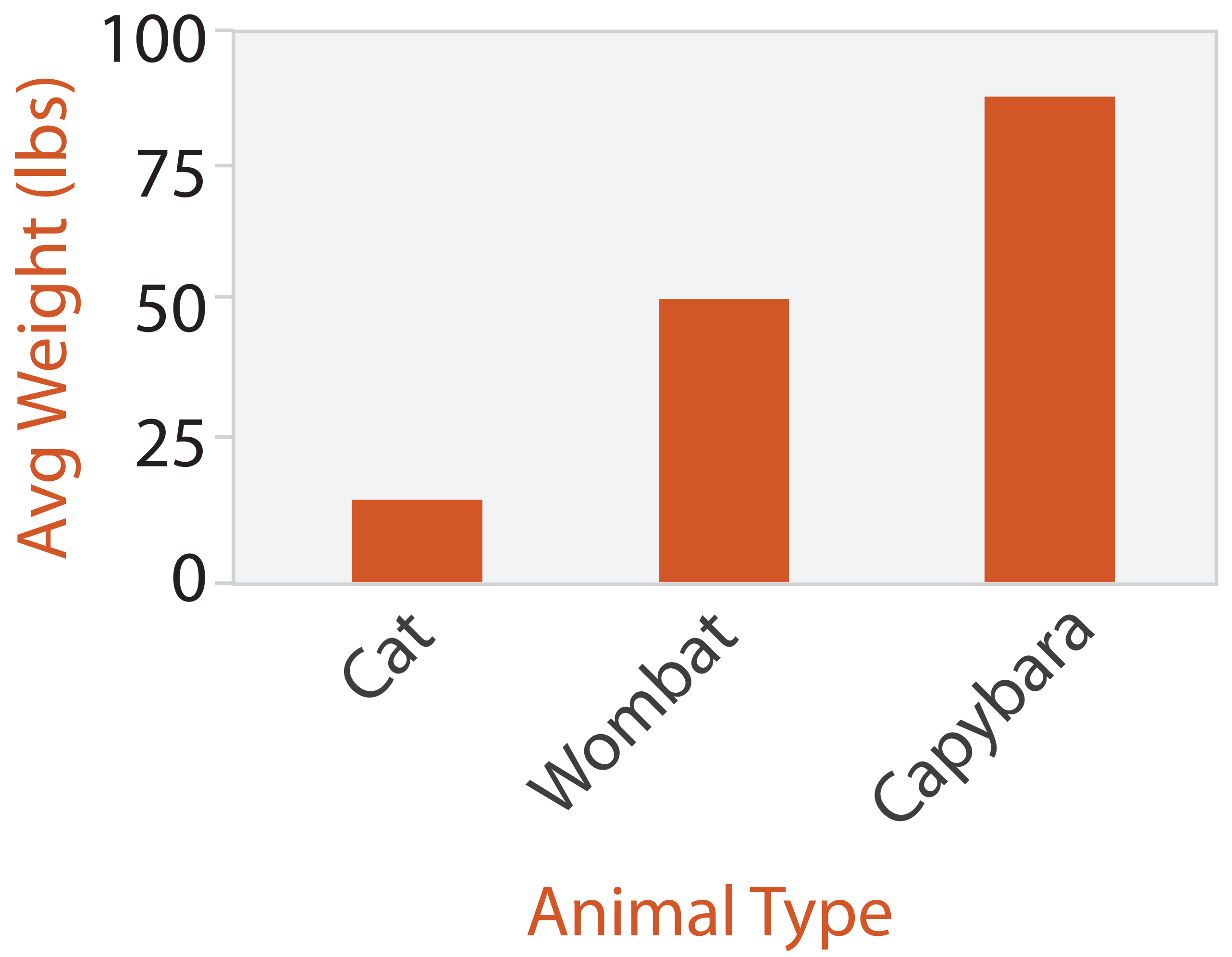
- Data: two quantitative values
- Task: find trends, clusters, outliers
- How: marks at spatial position in horizontal and vertical directions

- Correlation: dependence between two attributes
  - Positive and negative correlation
  - Indicated by lines
- Coordinate system (axes) and labels are important!

List Alignment: Bar Charts

- Data: one quantitative attribute, one categorical attribute
- Task: lookup & compare values
- How: line marks, vertical position (quantitative), horizontal position (categorical)
- What about **length**?
- Ordering criteria: alphabetical or using quantitative attribute
- Scalability: distinguishability
  - bars at least one pixel wide
  - hundreds

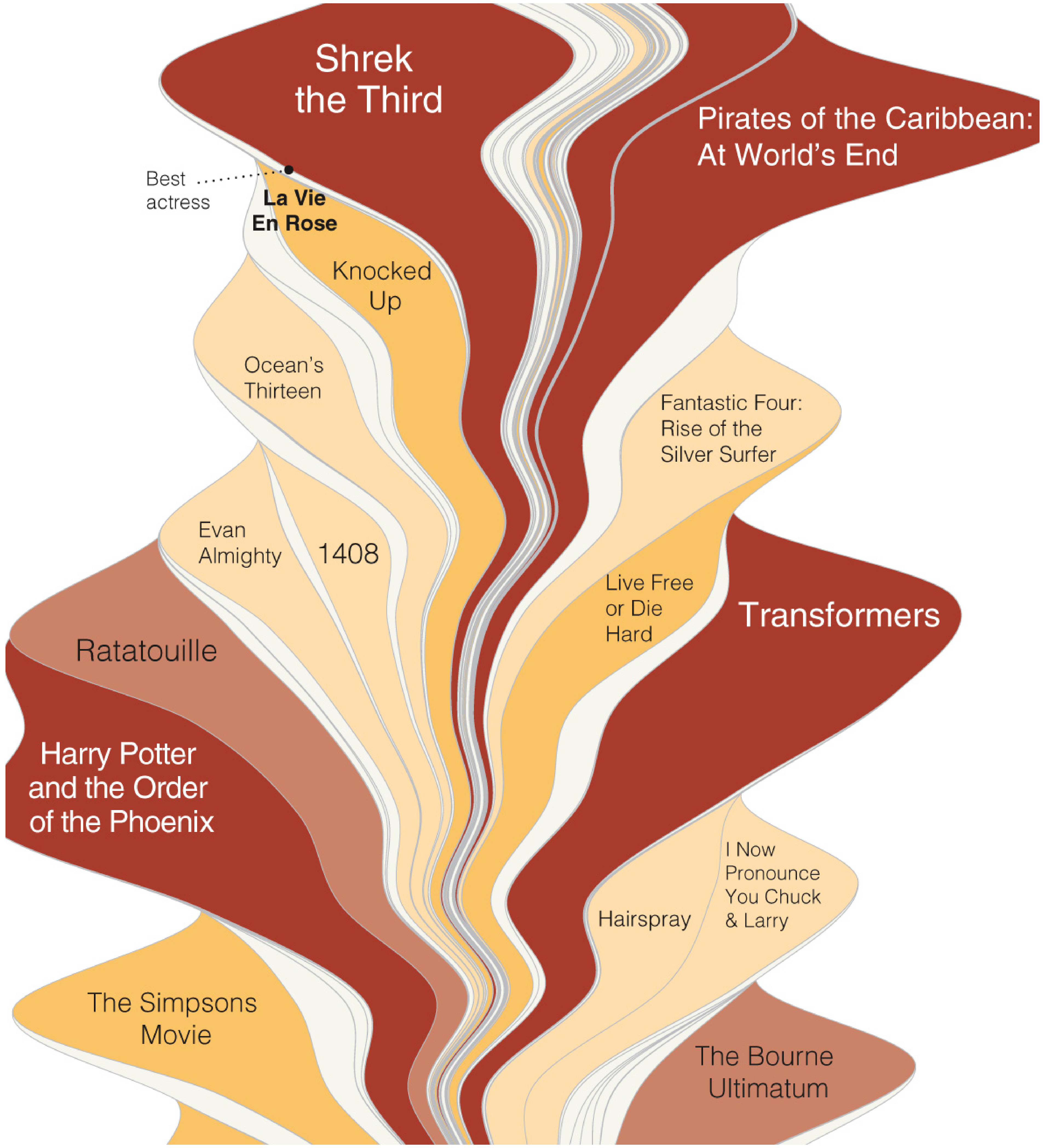




[Munzner (ill. Maguire), 2014]

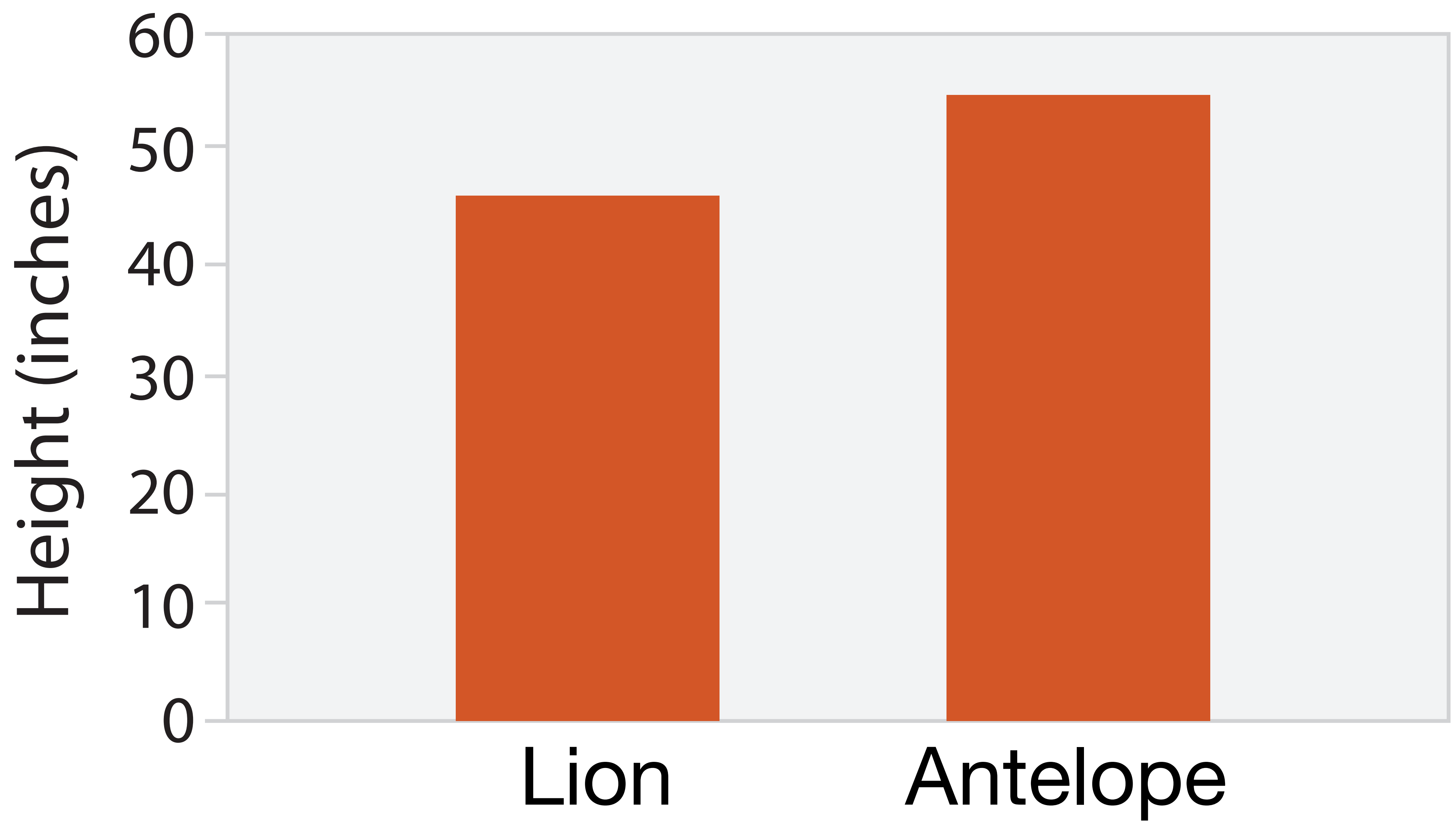
Streamgraphs

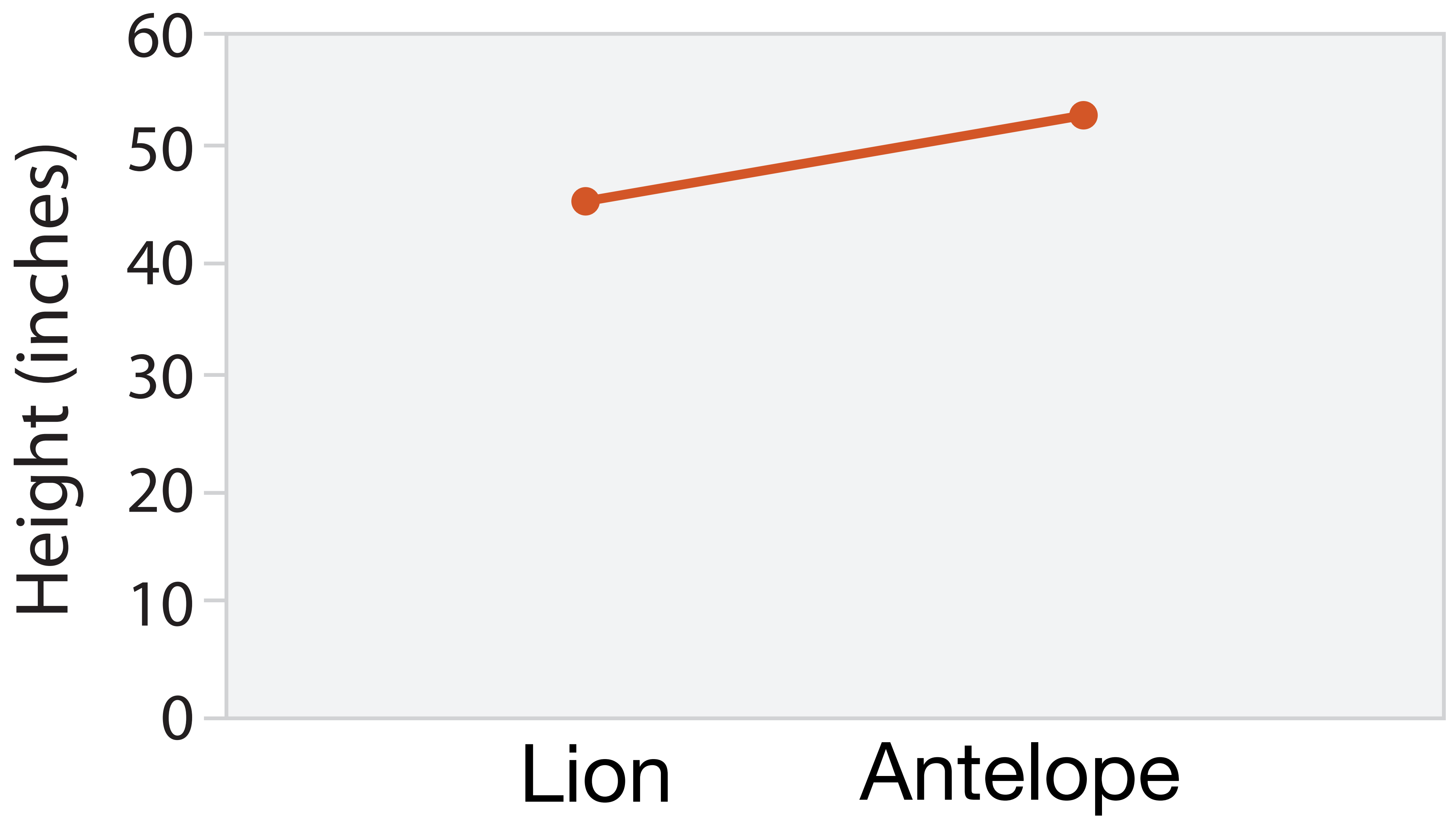
- Include a time attribute
- Data: multidimensional table, one quantitative attribute (count), one ordered key attribute (time), one categorical key attribute
- + derived attribute: layer ordering (quantitative)
- Task: analyze trends in time, find (maxmial) outliers
- How: derived position+geometry, length, color



[Byron and Wattenberg, 2012]

Proper Use of Line and Bar Charts



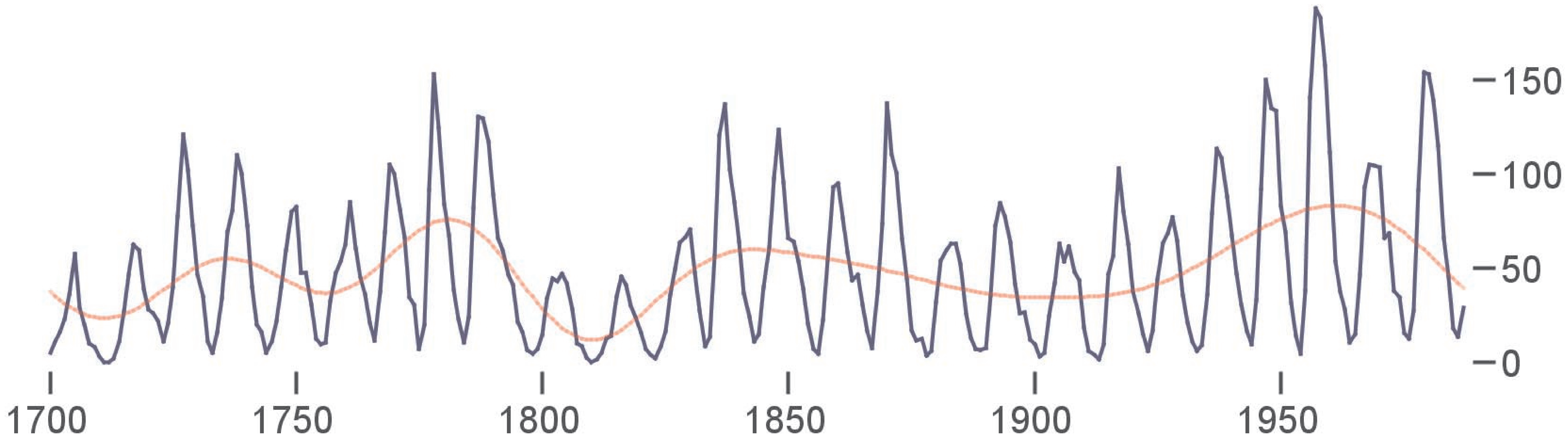


- What does the line indicate?
- Does this make sense?

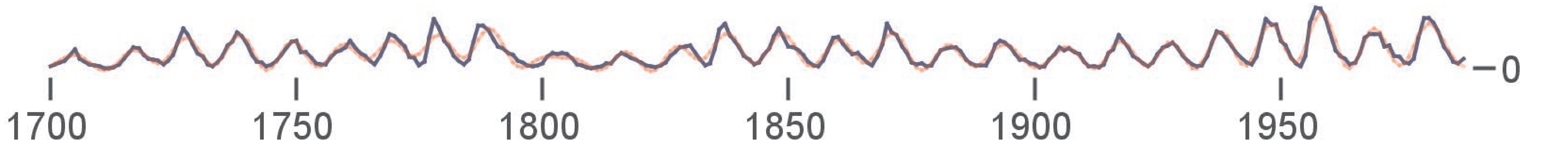
[Adapted from Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]

Multiscale Banking

Aspect Ratio = 3.96

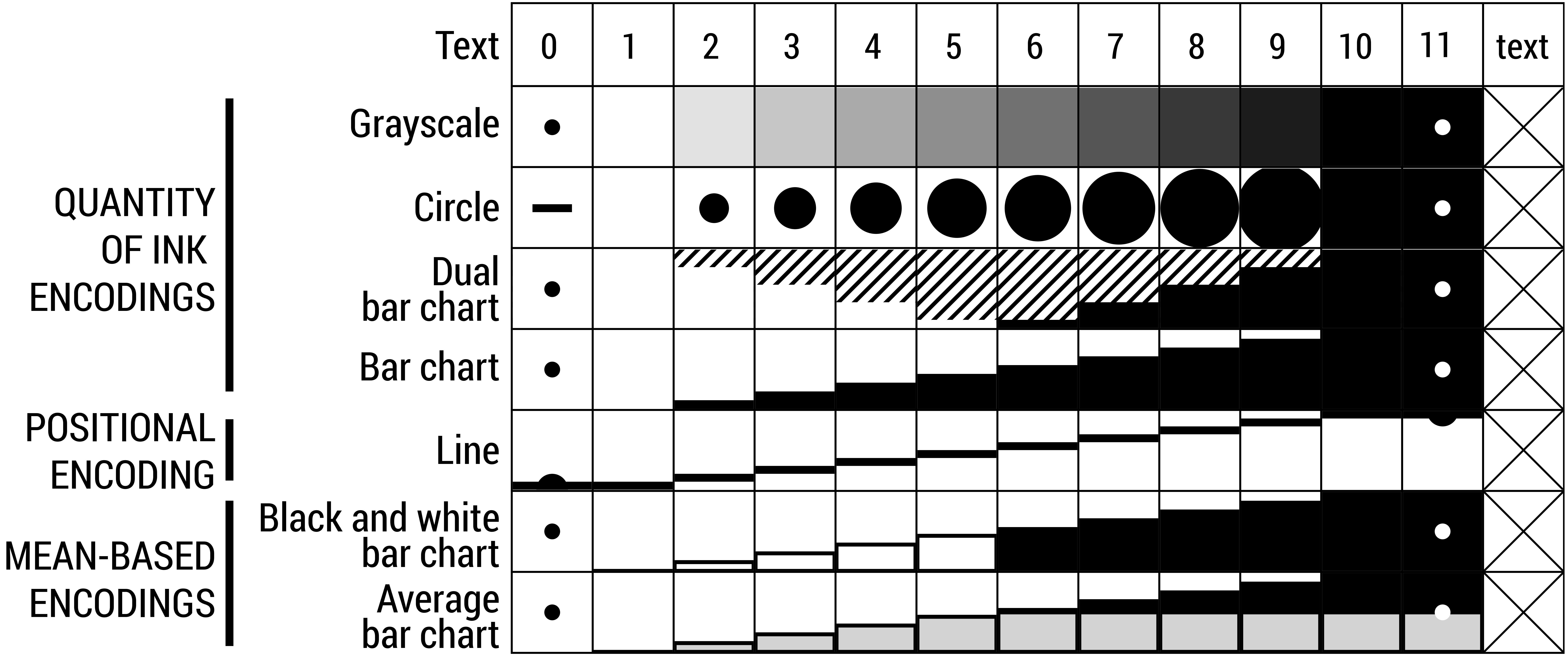


Aspect Ratio = 22.35



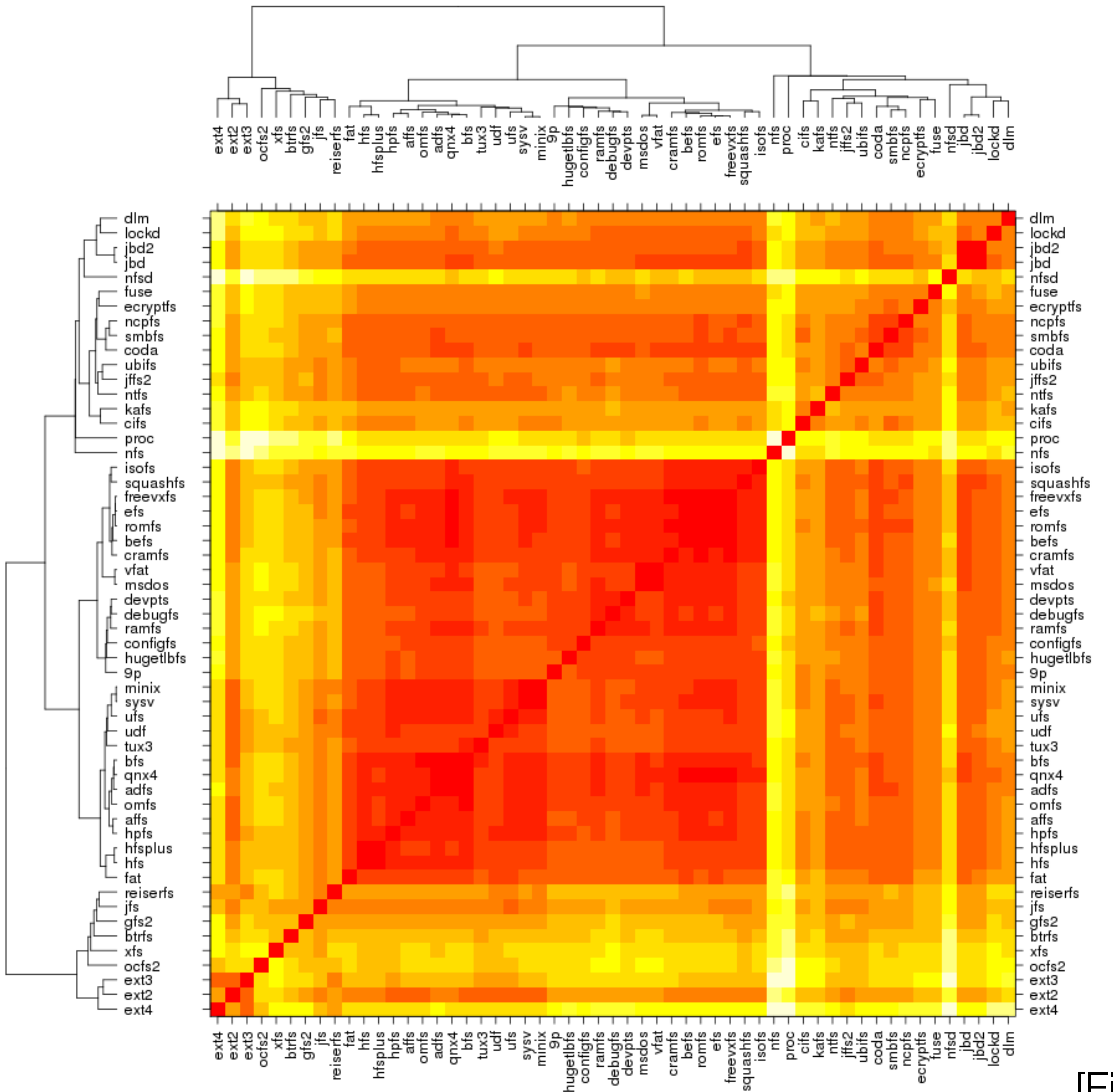
[Heer and Agrawala, 2006]

Bertin’s Matrix Encodings



[C.Perrin et al., 2014]

Cluster Heatmap

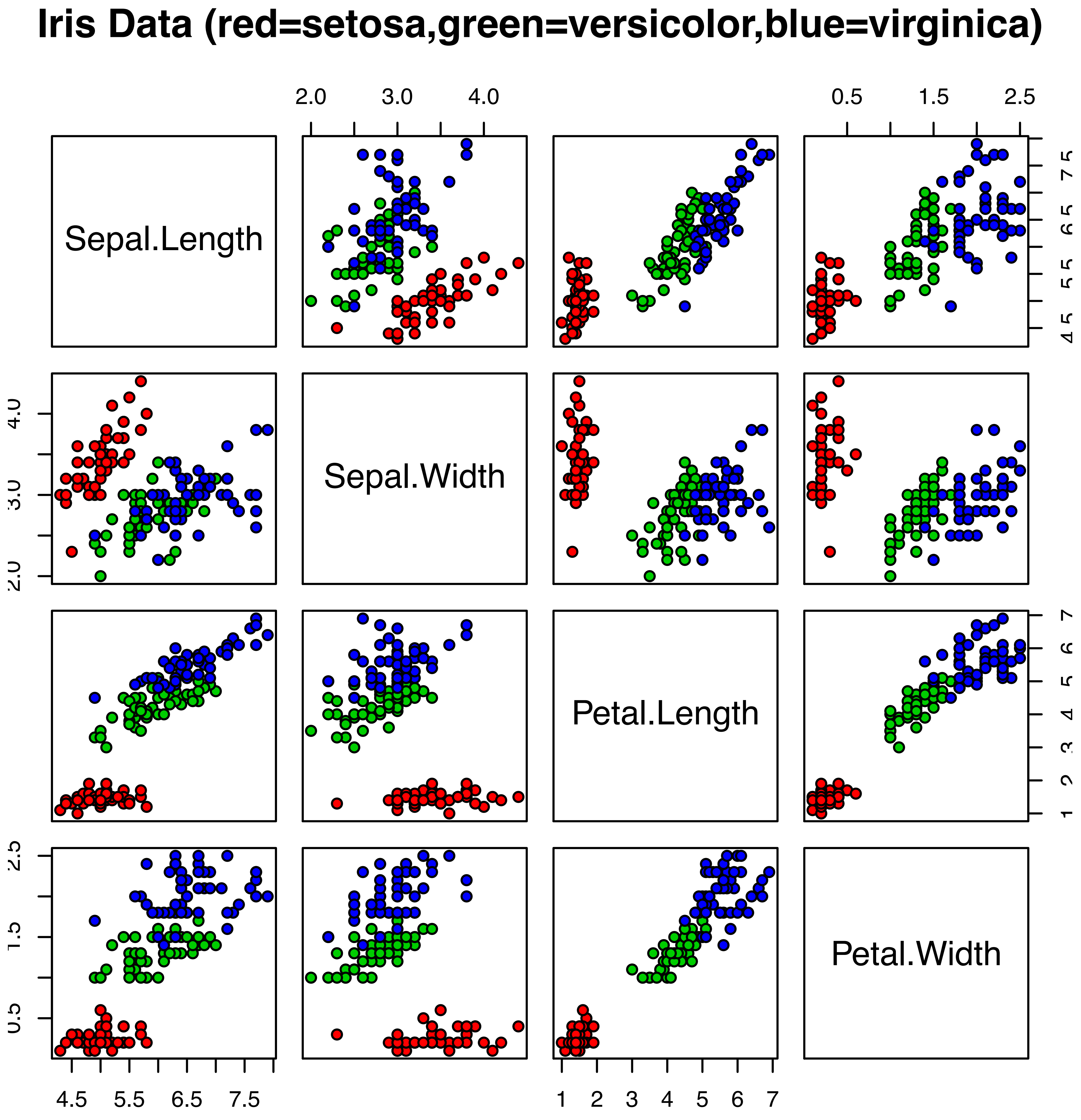


[File System Similarity, R. Musăloiu-E., 2009]

Scatterplot Matrix (SPLOM)

- Data: Many quantitative attributes
- Derived Data: names of attributes
- Task: Find correlations, trends, outliers
- How: Scatterplots in matrix alignment
- Scale: attributes: ~12, items: hundreds?

- Visualizations in a visualization: at high level, marks are themselves visualizations…



[Wikipedia]

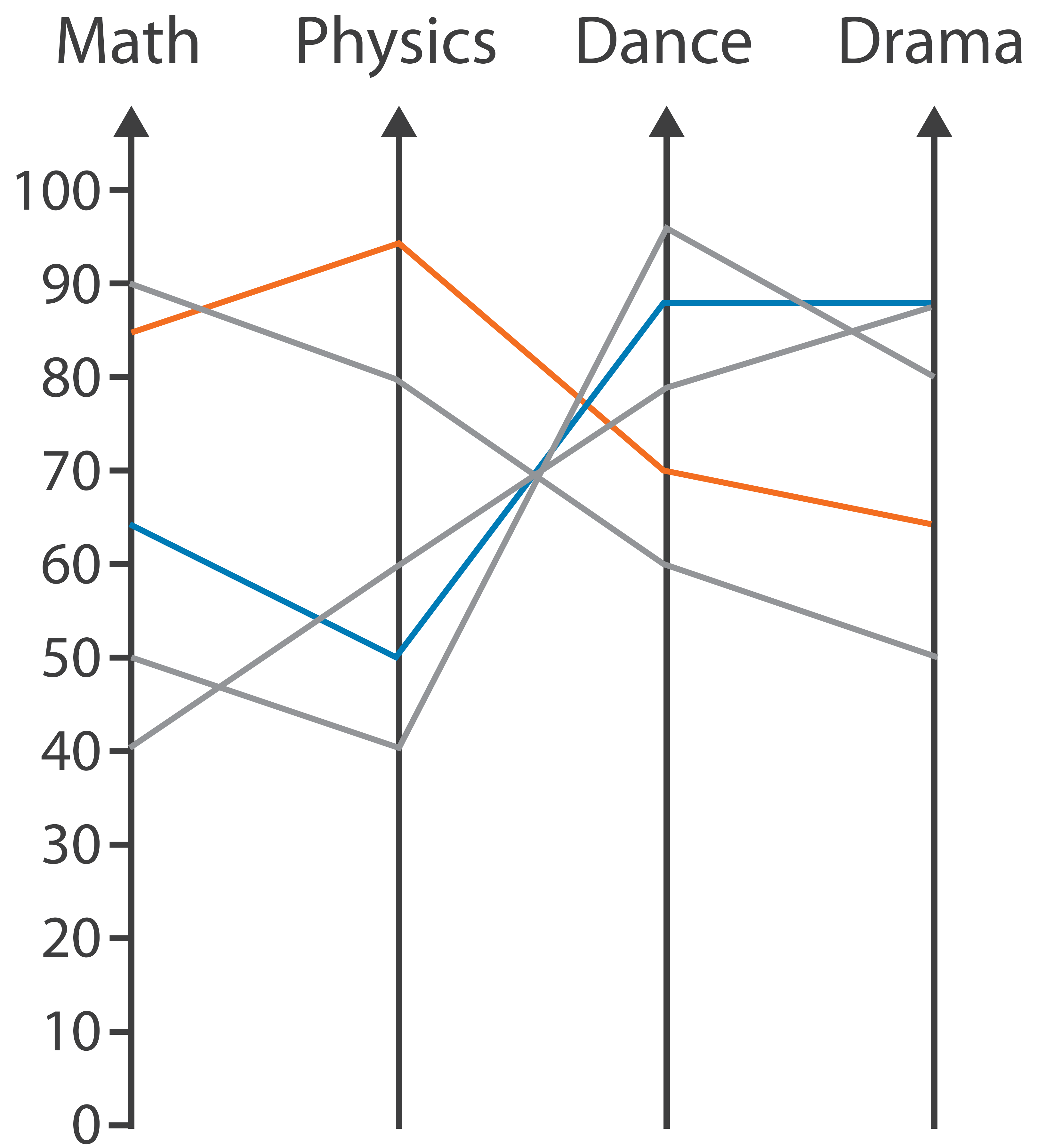
Spatial Axis Orientation

- So far, we have seen the vertical and horizontal axes (a **rectilinear** layout) used to encode almost everything
- What other possibilities are there for axes?

[Munzner (ill. Maguire), 2014]

Spatial Axis Orientation

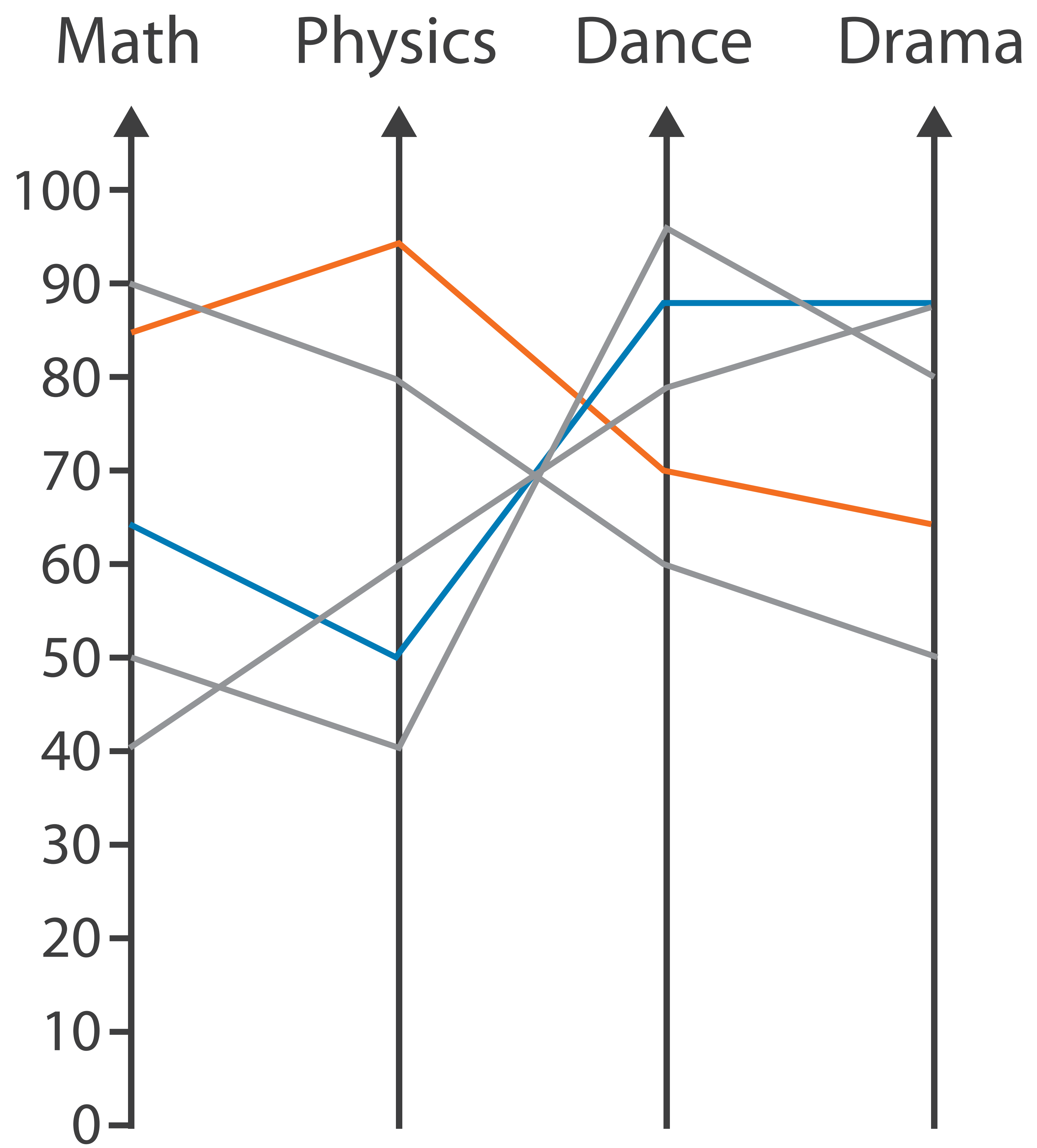
- So far, we have seen the vertical and horizontal axes (a **rectilinear** layout) used to encode almost everything
- What other possibilities are there for axes?
  - Parallel axes



[Munzner (ill. Maguire), 2014]

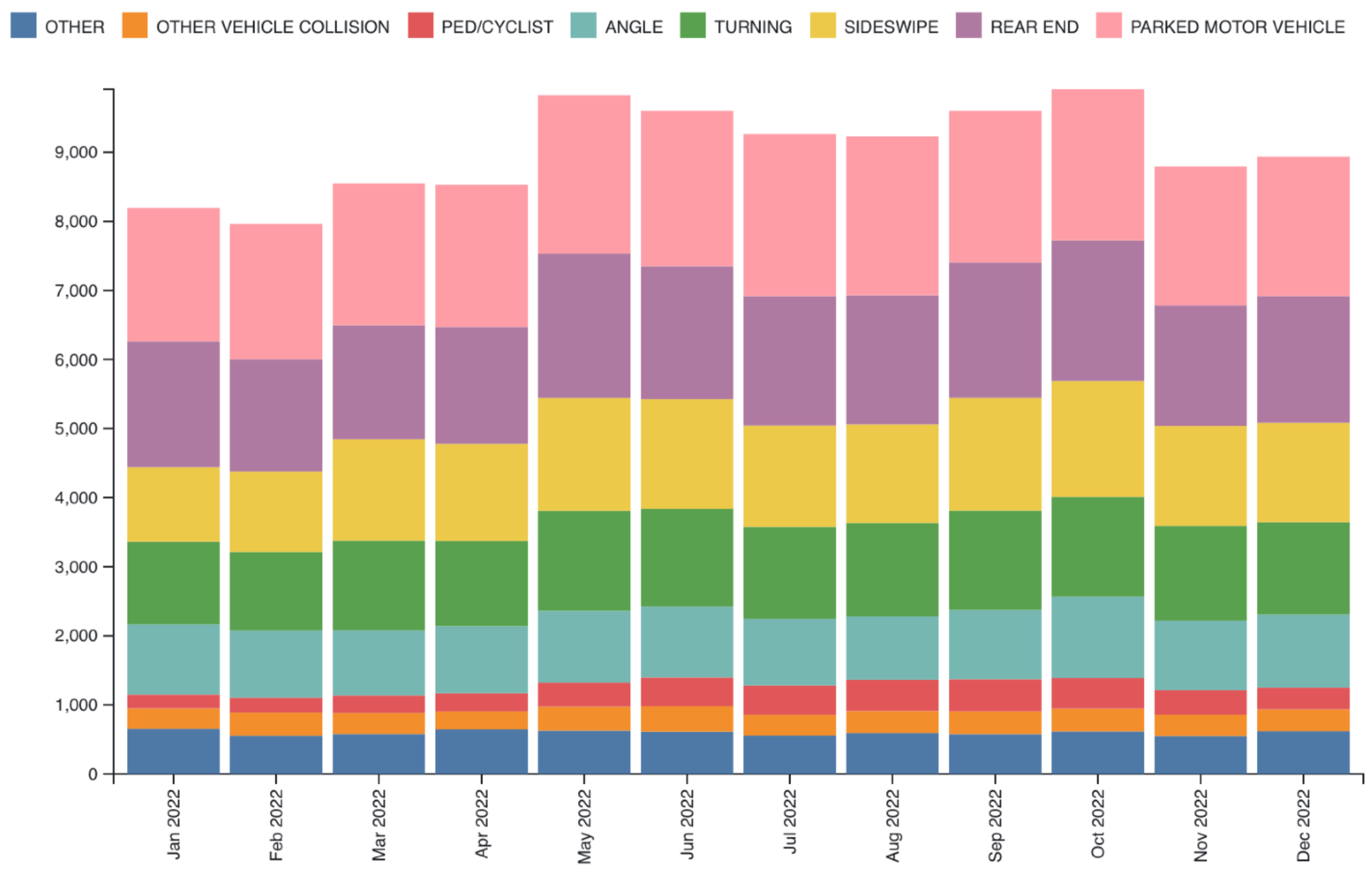
Spatial Axis Orientation

- So far, we have seen the vertical and horizontal axes (a **rectilinear** layout) used to encode almost everything
- What other possibilities are there for axes?
  - Parallel axes
  - Radial axes



[Munzner (ill. Maguire), 2014]

Assignment 3

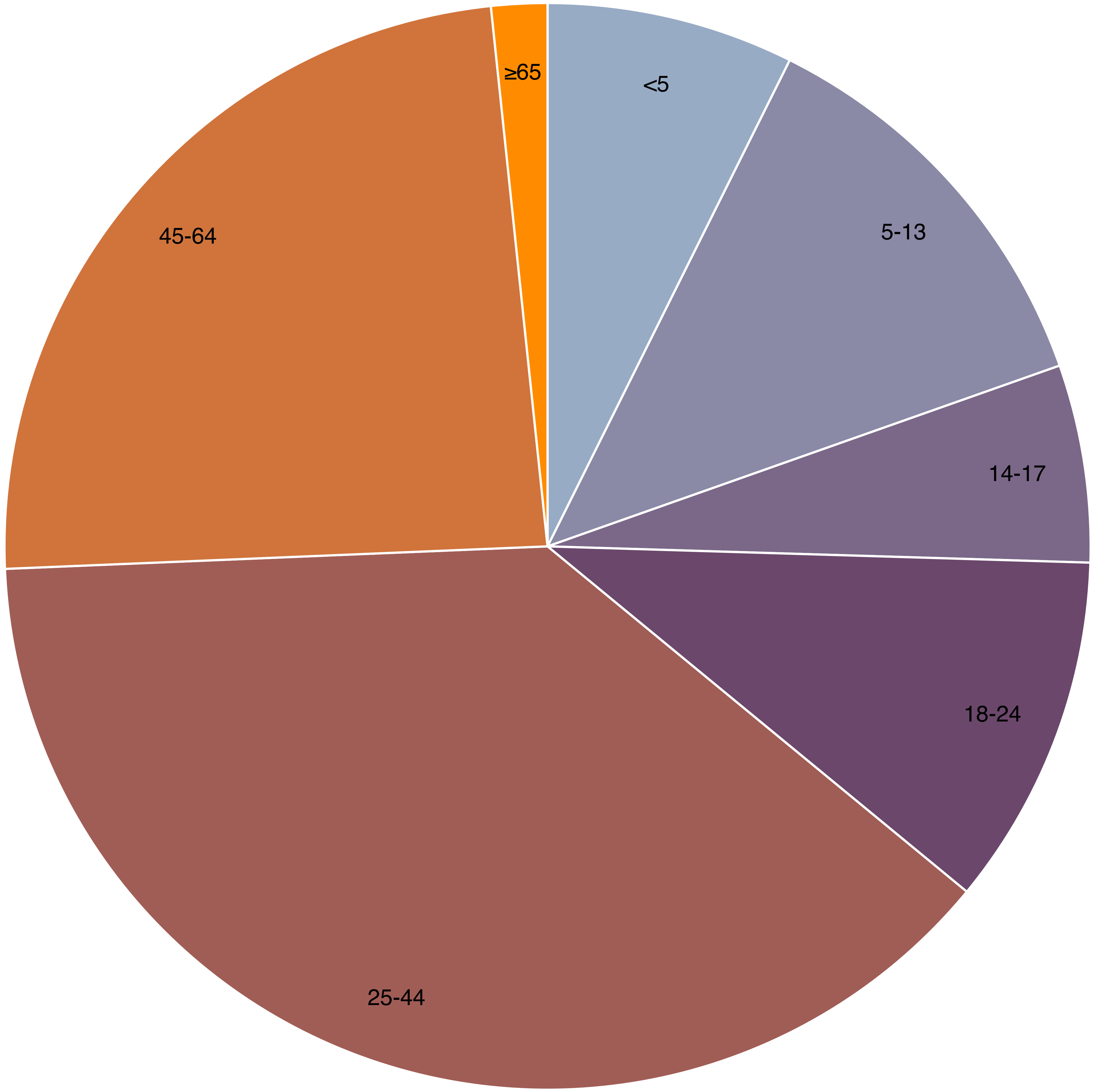


- Chicago Traffic Crashes
- Create the same stacked bar chart using
  - Tableau Public
  - Observable Plot
  - D3
- D3 Stacked Bar Chart:
  - Required for CSCI 627 students
  - CSCI 490 students can just do counts

Project

- Start thinking about project dataset and questions
- Working on posting some example datasets
- Goal: Less explored datasets (more opportunity for design/questions)
- If you are doing research and can tie this project in, please talk with me

Pie Chart

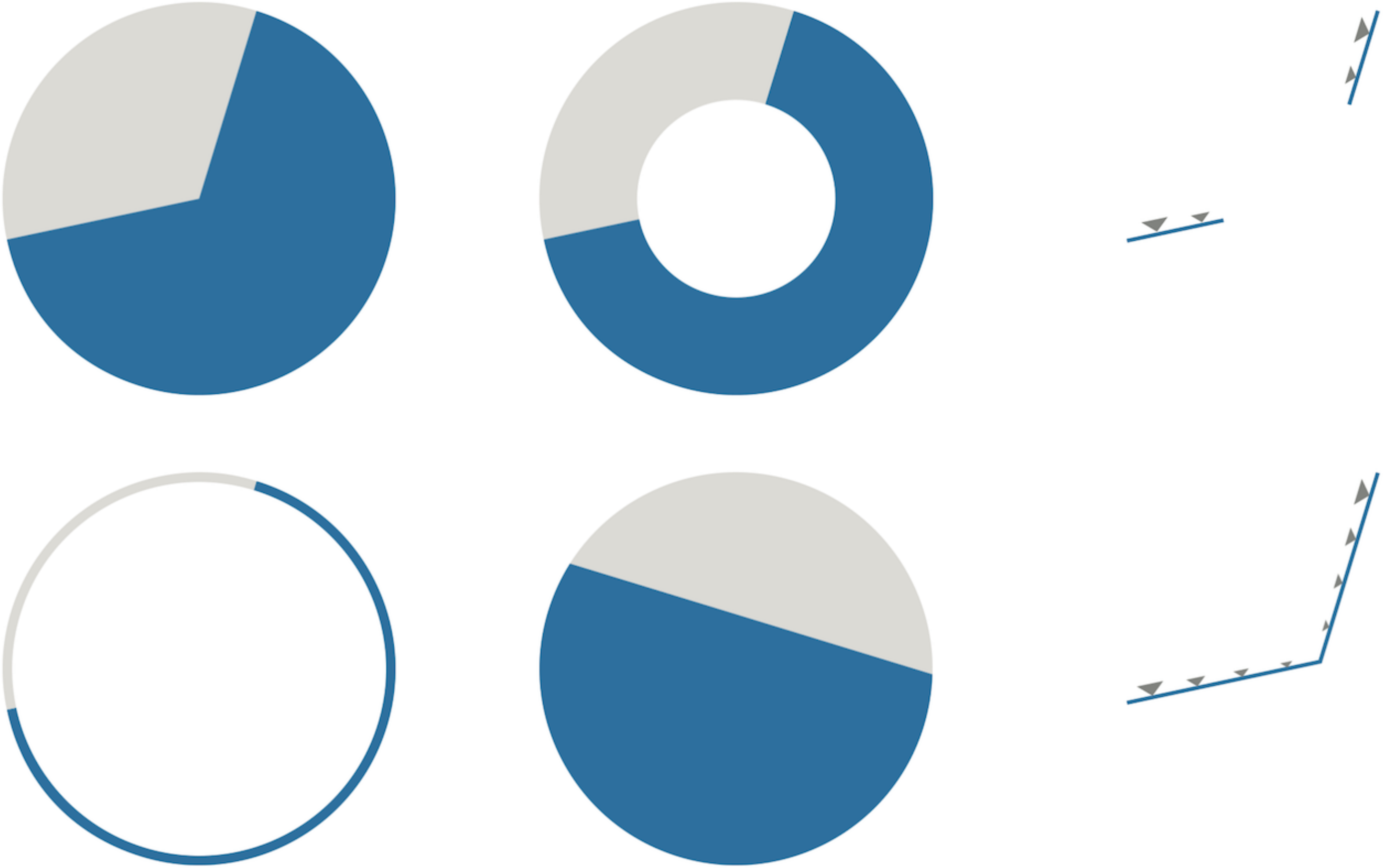


[Pie Chart, Bostock, 2017]

Pie Charts

- vs. bar charts [Munzner's Textbook, 2014]
  - Angle channel is lower precision then position in bar charts
- What about donut charts?
- Are we judging angle, or are we judging area, … or arc length?
  - "Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts", D. Skau and R. Kosara, 2016
  - "Judgment Error in Pie Chart Variations", R. Kosara and D. Skau, 2016
  - Summary: "An Illustrated Study of the Pie Chart Study Results"

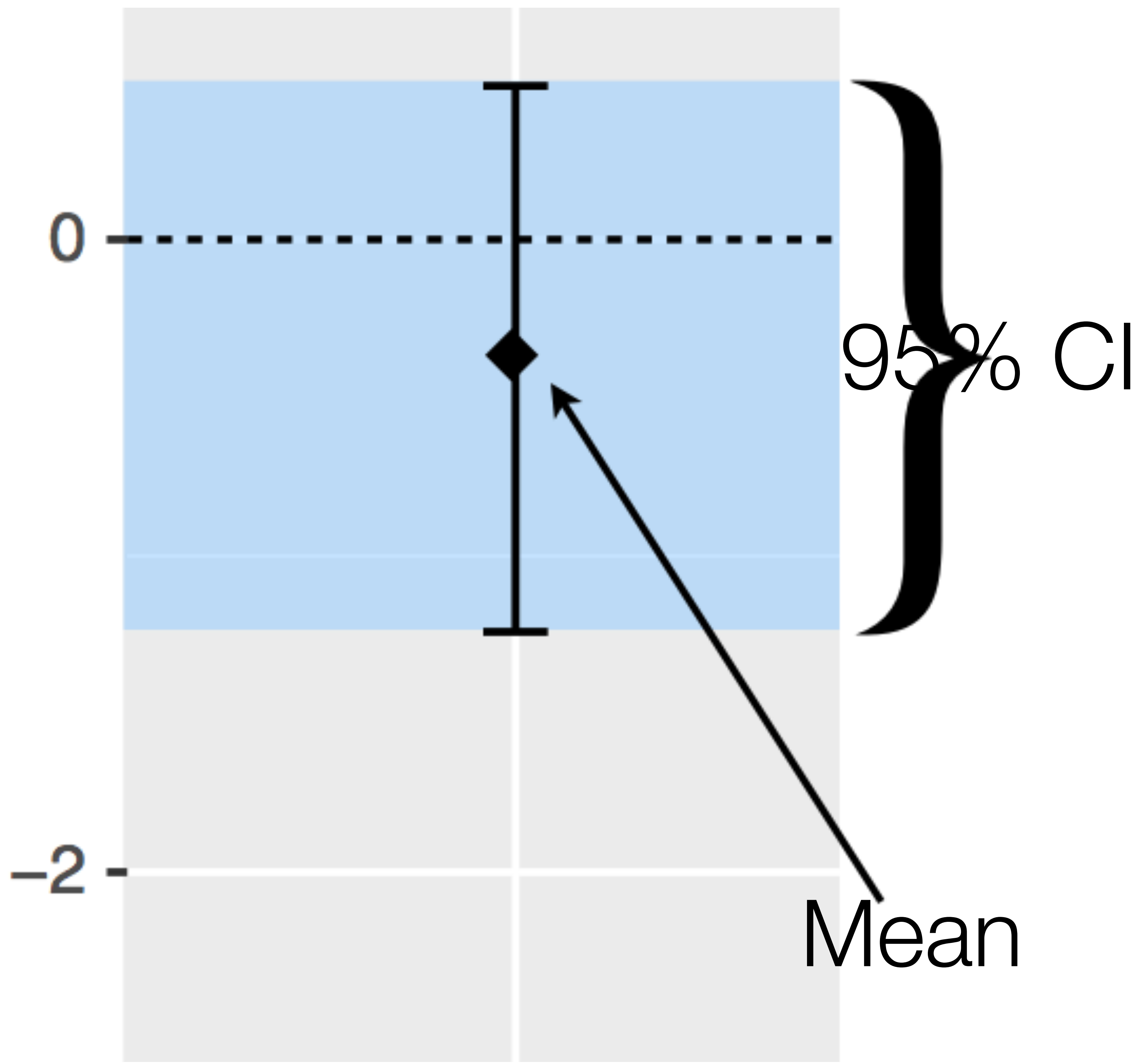
Arcs, Angles, or Areas?



[R. Kosara and D. Skau, 2016]

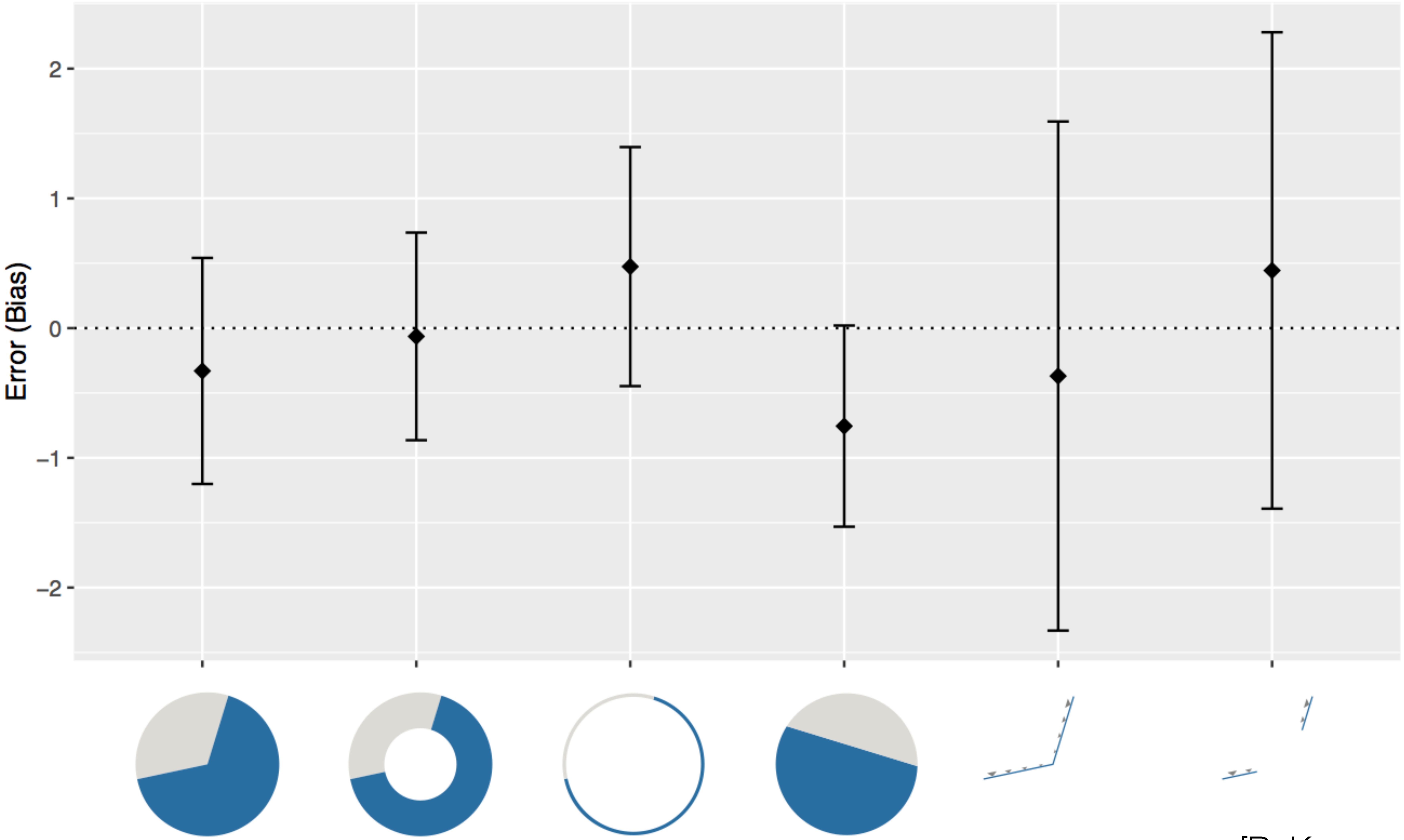
Study Setup

- Three studies
- 80-100 participants each
- Each answered ~60 questions
- Computed results using 95% Confidence Intervals



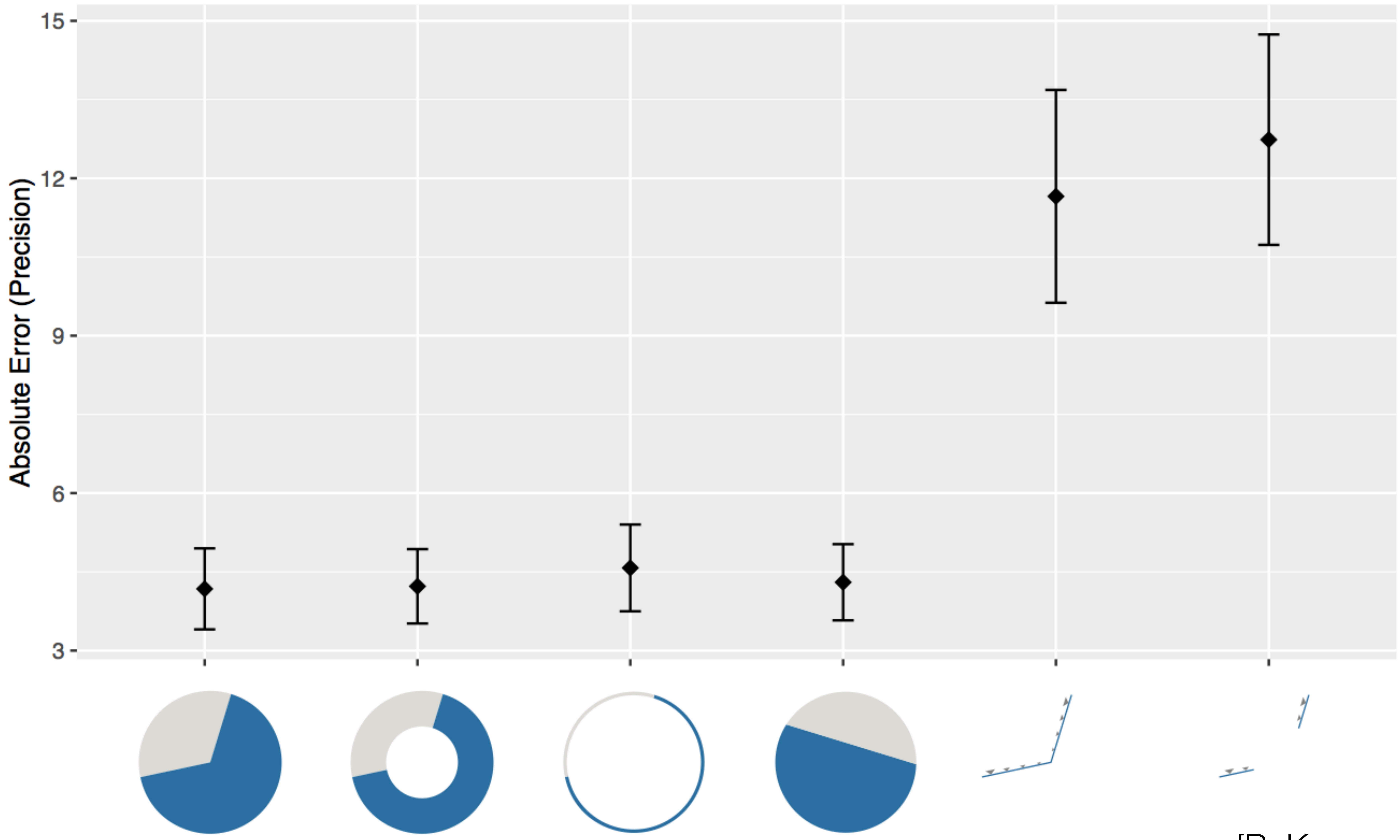
[R. Kosara and D. Skau, 2016]

Signed Error



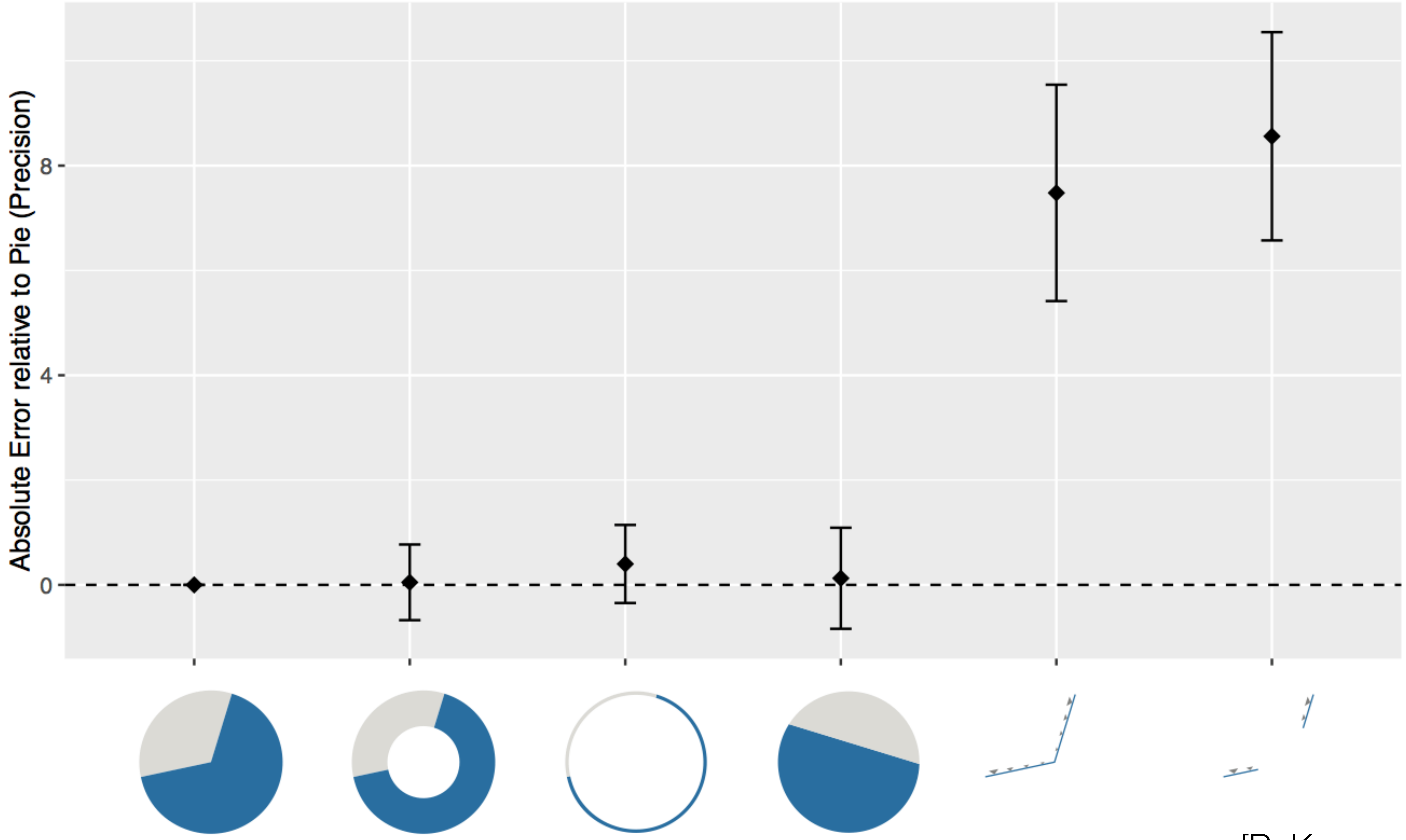
[R. Kosara and D. Skau, 2016]

Absolute Error



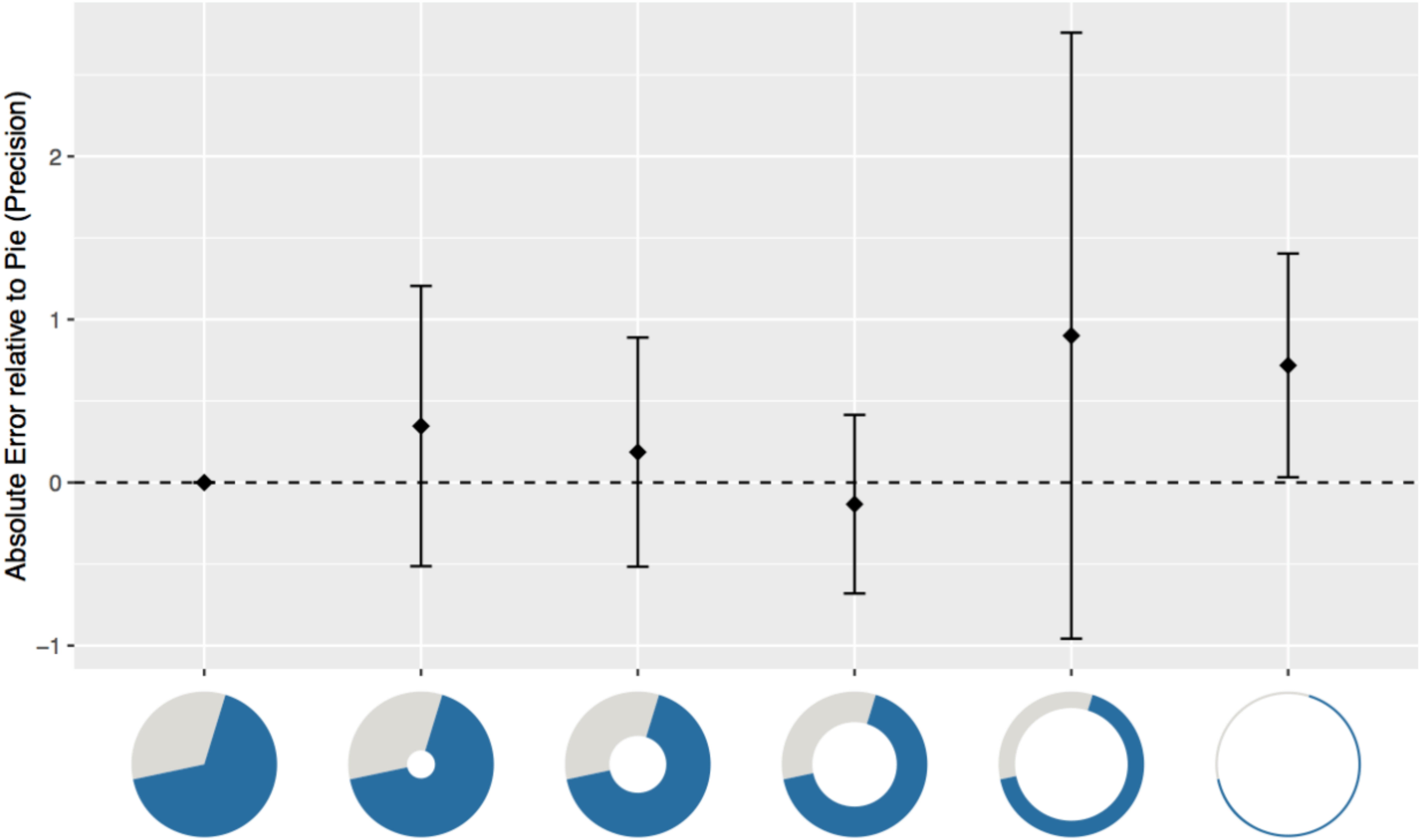
[R. Kosara and D. Skau, 2016]

Absolute Error Relative to Pie Chart



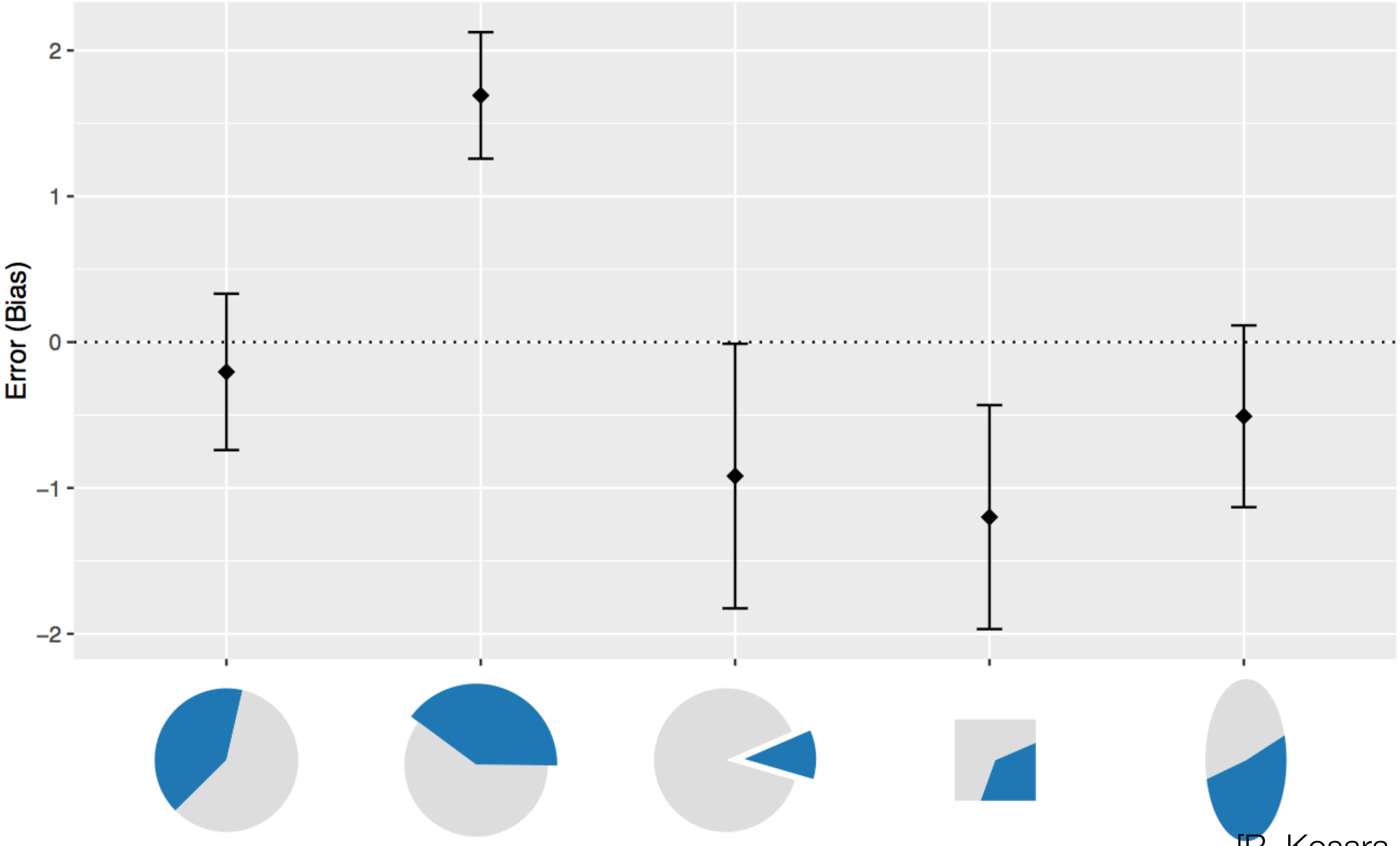
[R. Kosara and D. Skau, 2016]

Donut Charts Width



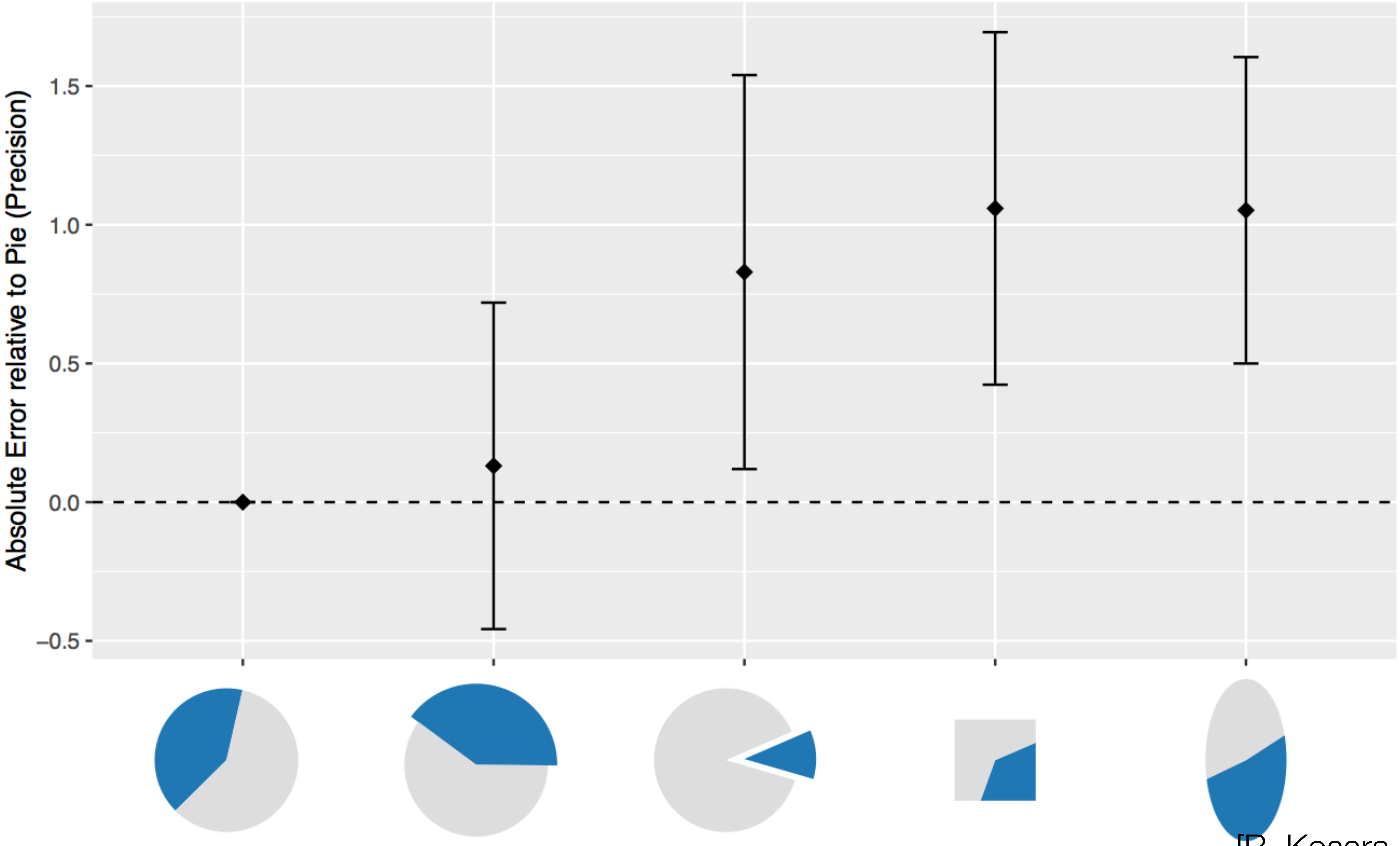
[R. Kosara and D. Skau, 2016]

Pie Chart Variations



[R. Kosara and D. Skau, 2016]

Pie Chart Variations



[R. Kosara and D. Skau, 2016]

Conclusion: We do not read pie charts by angle

[R. Kosara and D. Skau, 2016]

Pies vs. Bars

- …but area is still harder to judge than position
- Screens are usually not round

Color

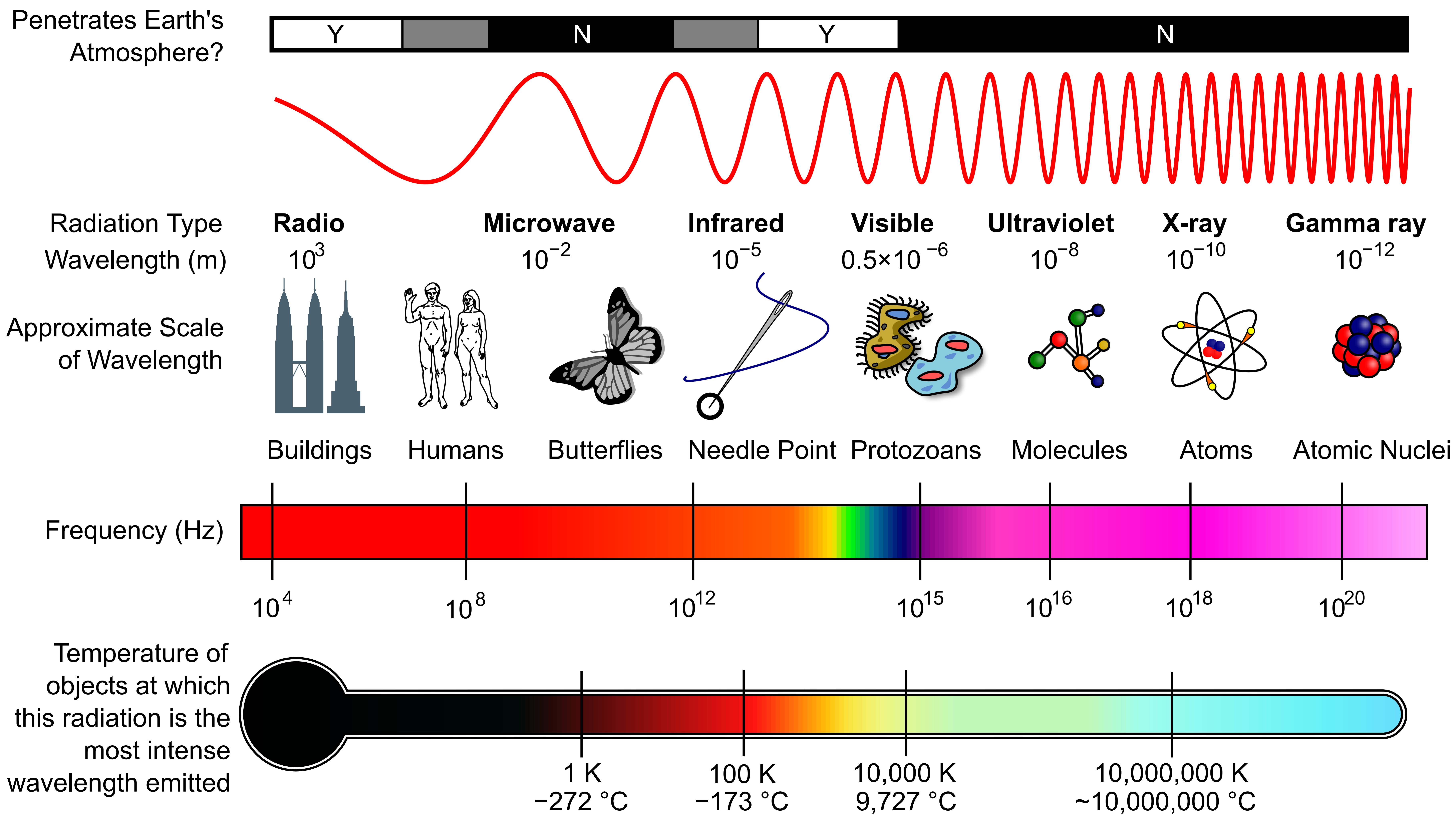
Color

Color

Color and Light

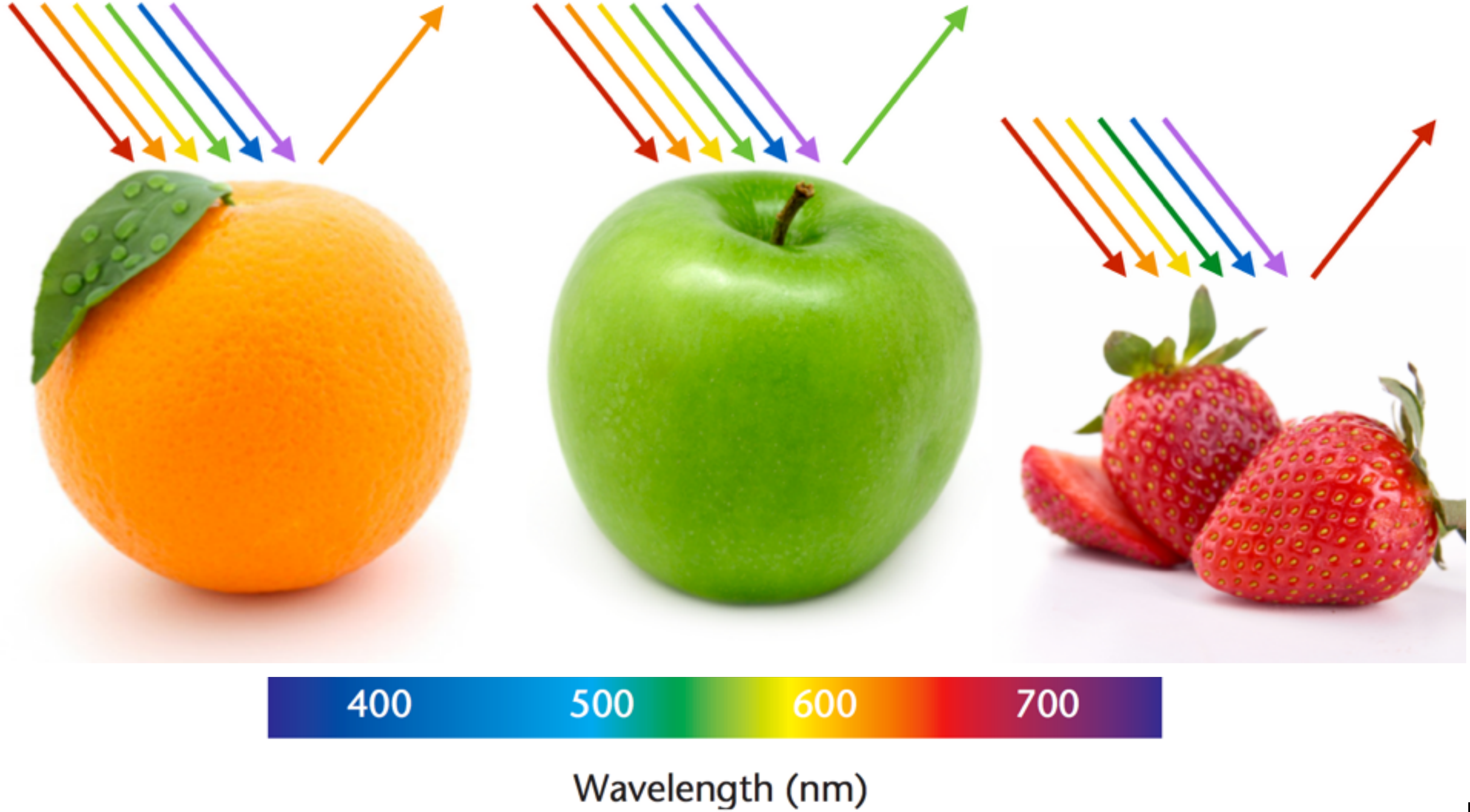
- Color is a **perceptive** property: color depends on the eyes and brain
- Visible light is a small portion of the **electromagnetic spectrum** which is composed of waves that at various frequencies (wavelengths), all traveling at the speed of light

Electromagnetic Spectrum



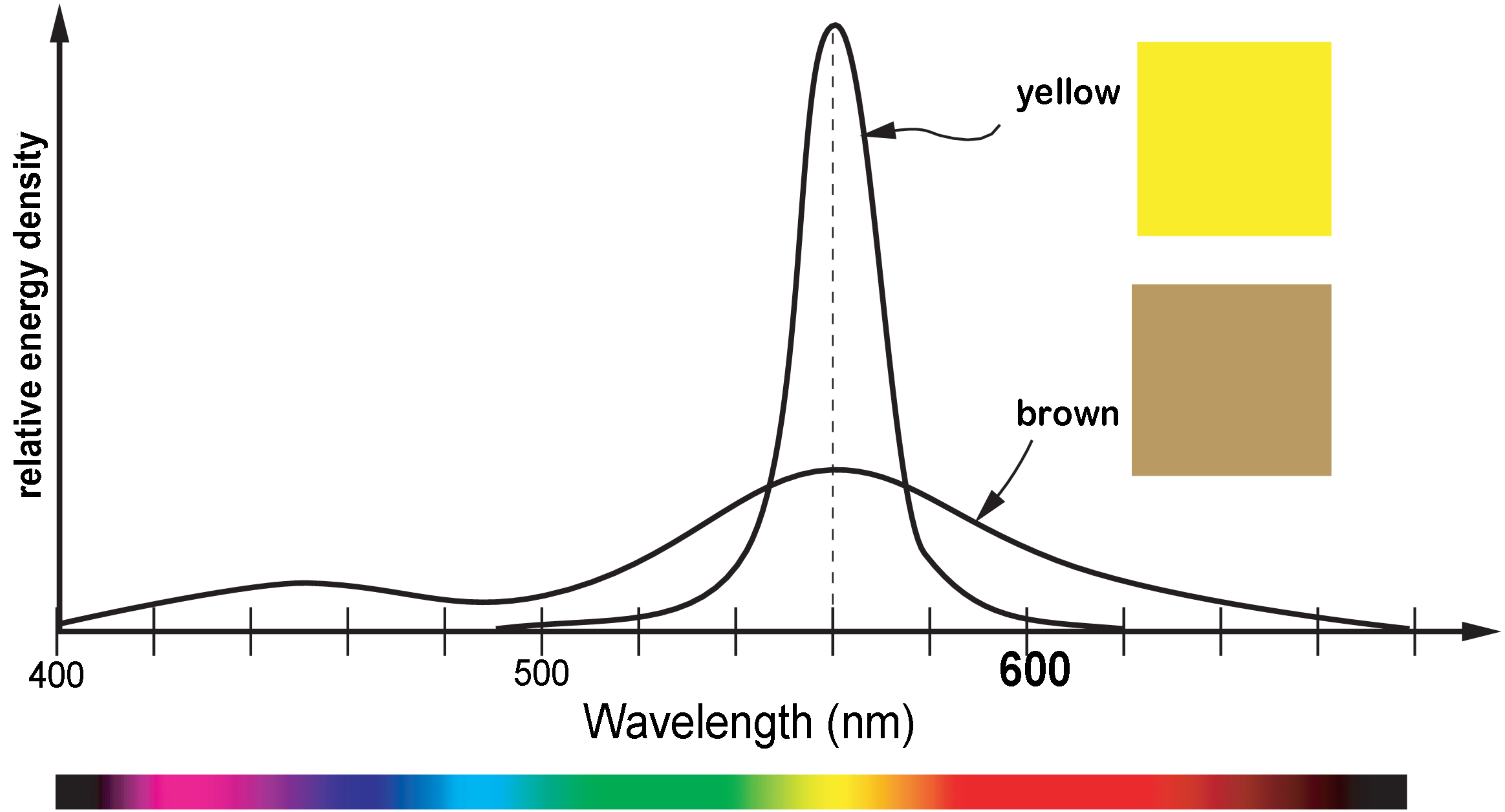
[Wikimedia, NASA]

Light Reflection & Absorption



[via M. Meyer]

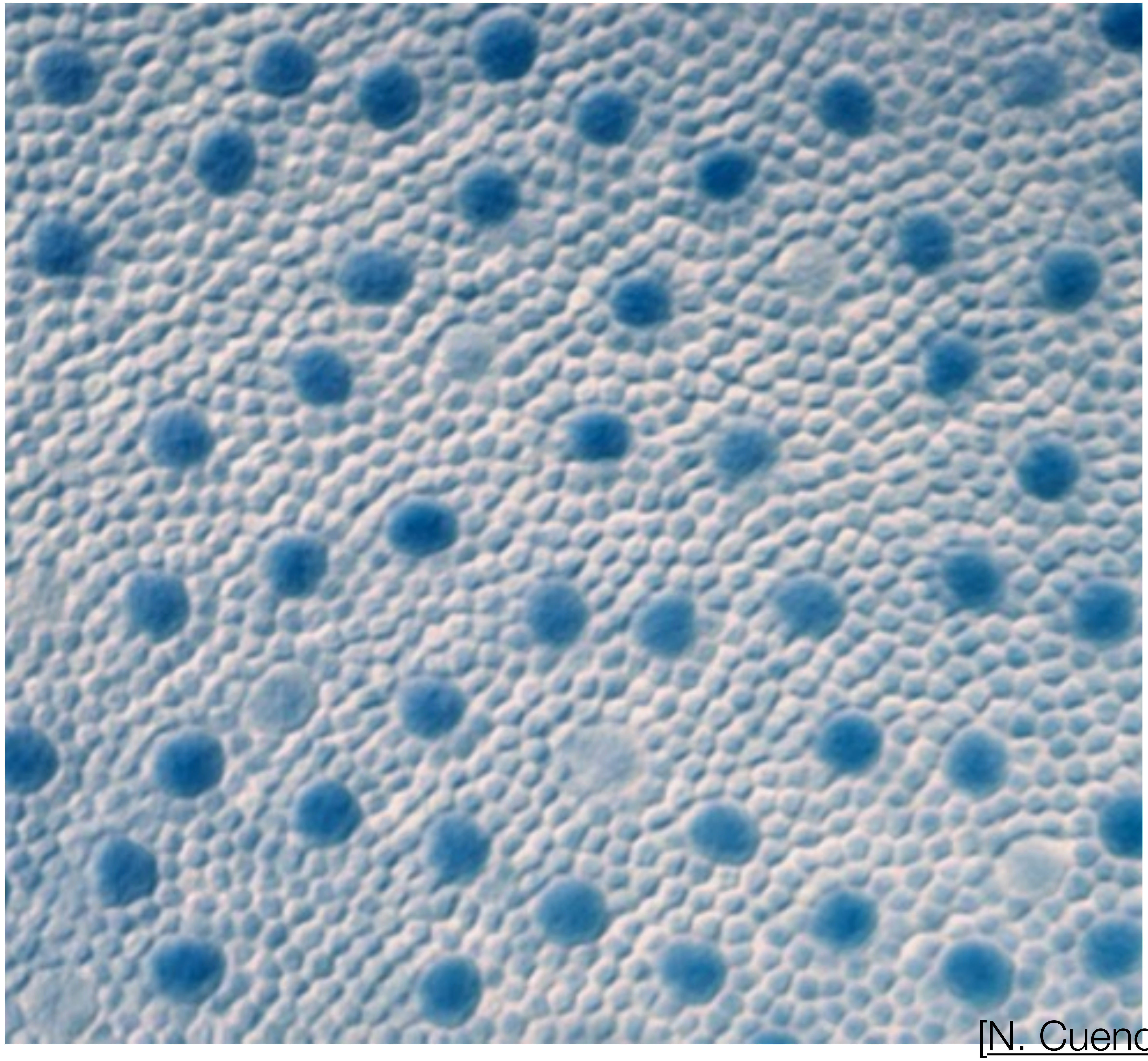
Color != Wavelength



[via M. Meyer]

Human Color Perception

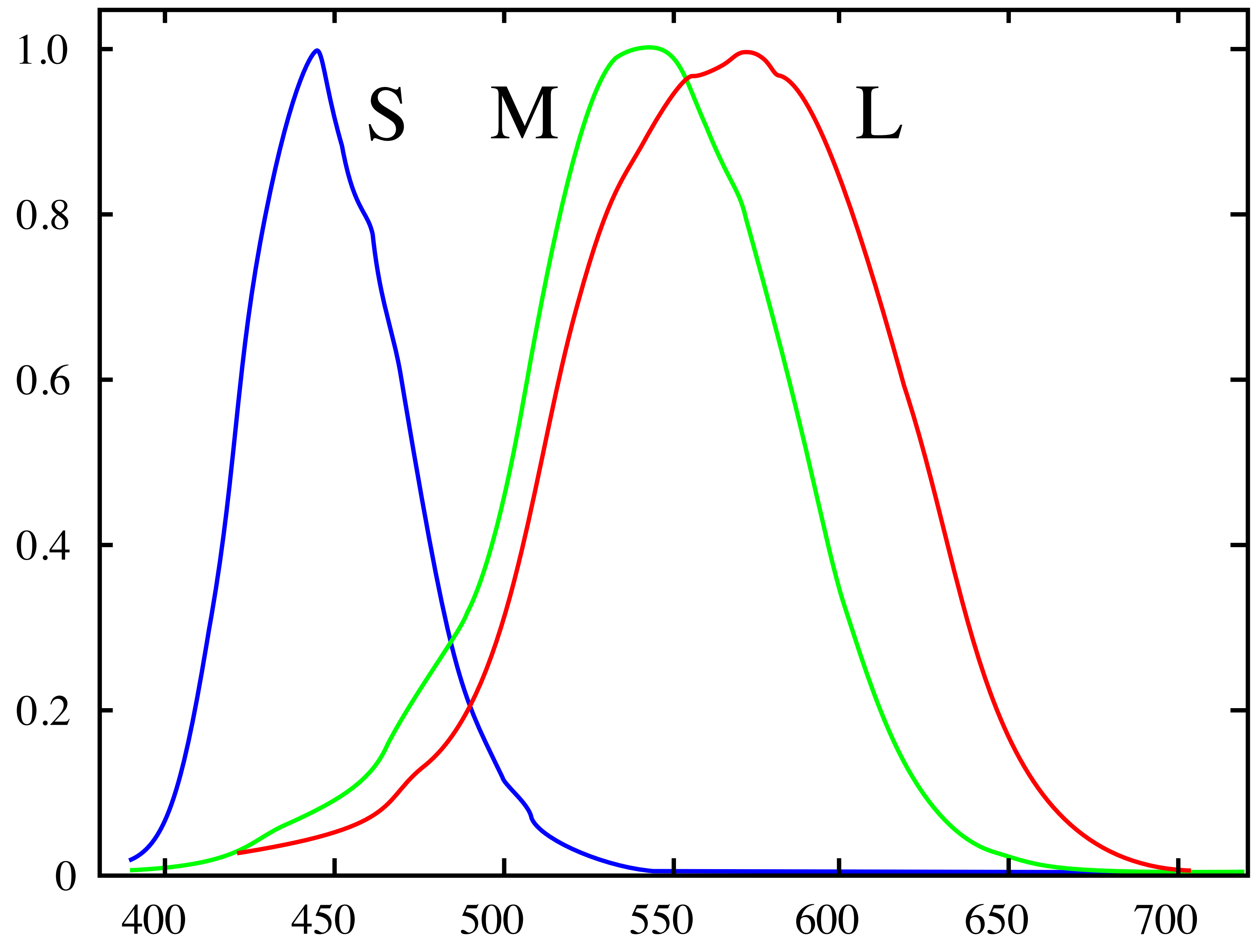
- Humans **do not** detect individual wavelengths of light
- Use **rods** and **cones** to detect light
  - rods capture intensity
  - cones capture color



[N. Cuenca]

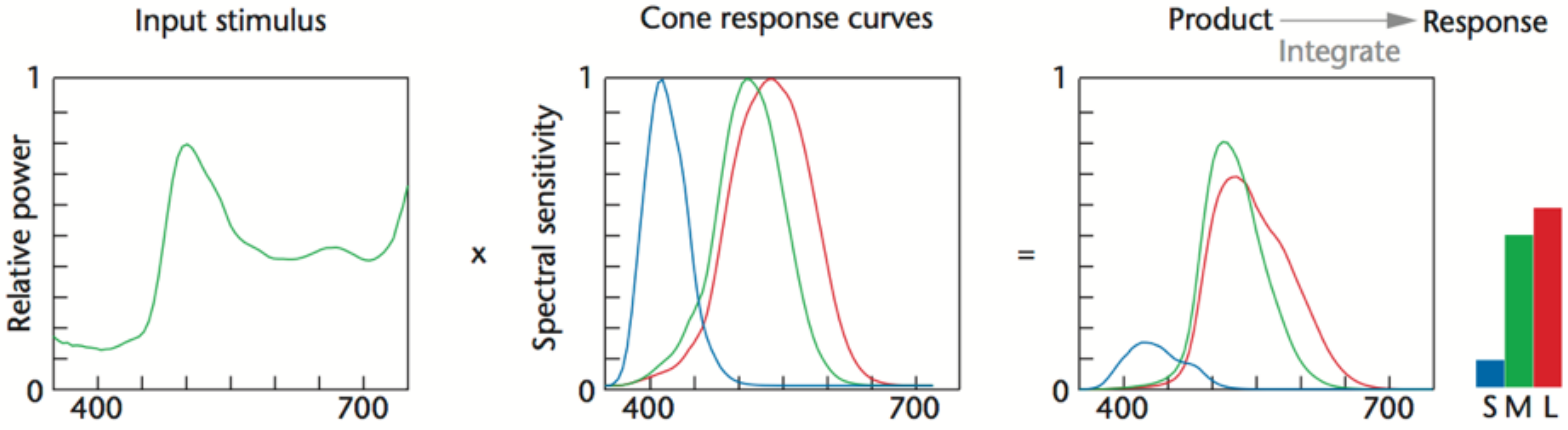
Human Color Perception

- Humans are **trichromatic**—we have three different types of cones
  - S (430nm): blue
  - M (540nm): green
  - L (570nm): "red"
- Note that the response curves **overlap**
- Spectra of visible light are "covered" by these responses
- Three numbers -> color



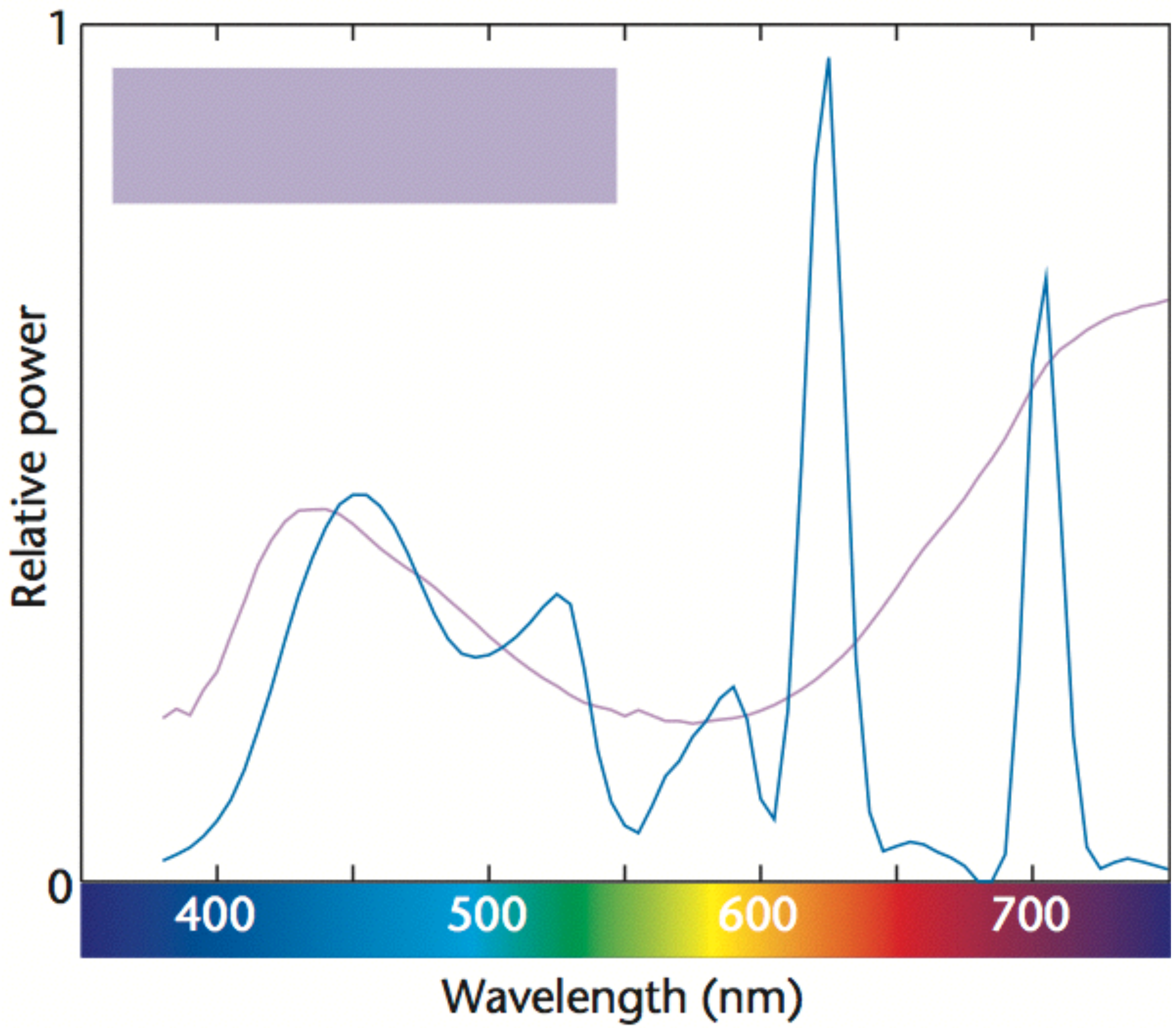
[Vanessaezekowitz at en.wikipedia]

Human Color Perception



[via M. Meyer]

Metamerism



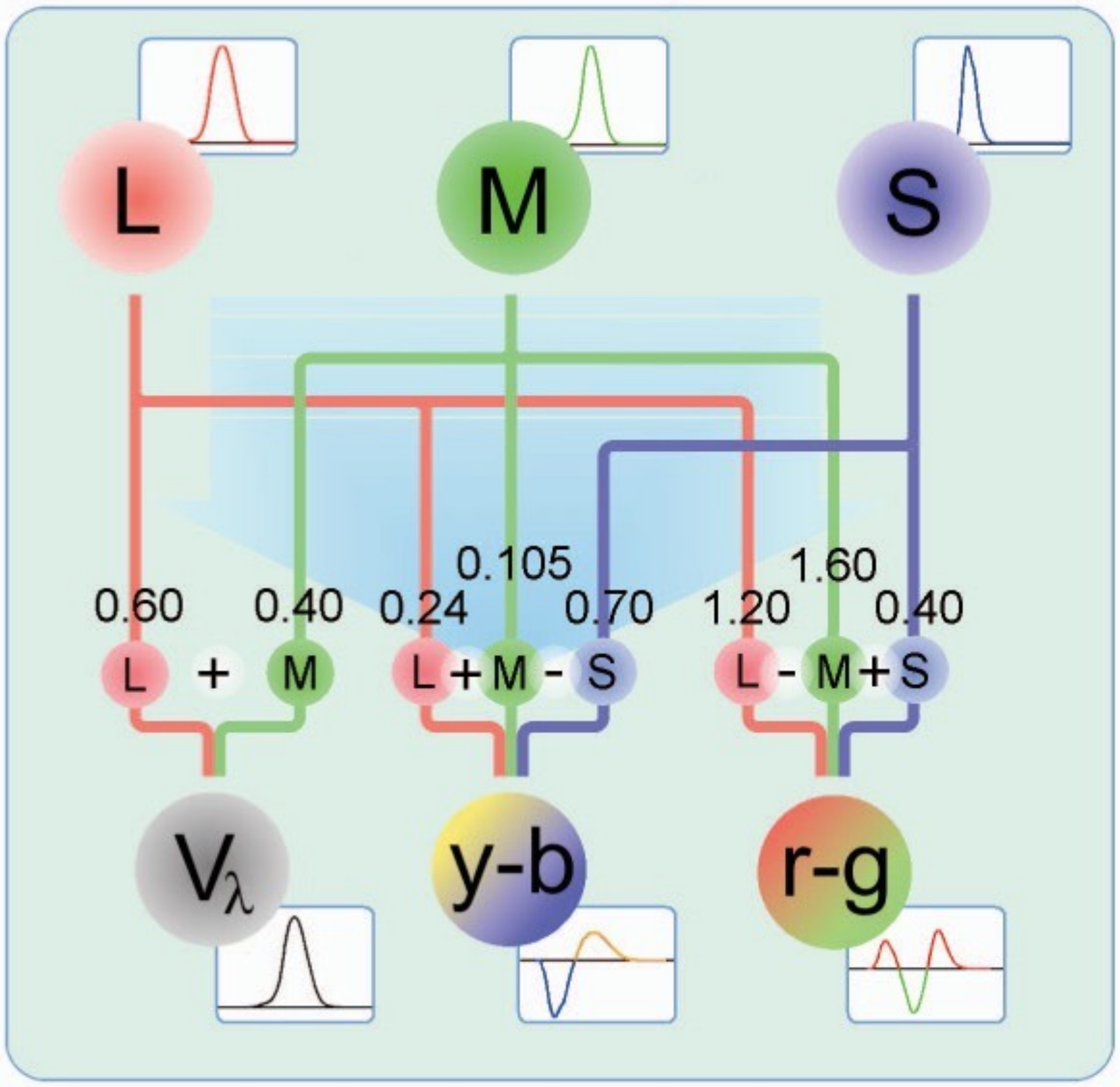
- Same responses == same color
- Humans are not spectrometers
- Do not get the whole function
- Three responses

[via M. Meyer]

Color

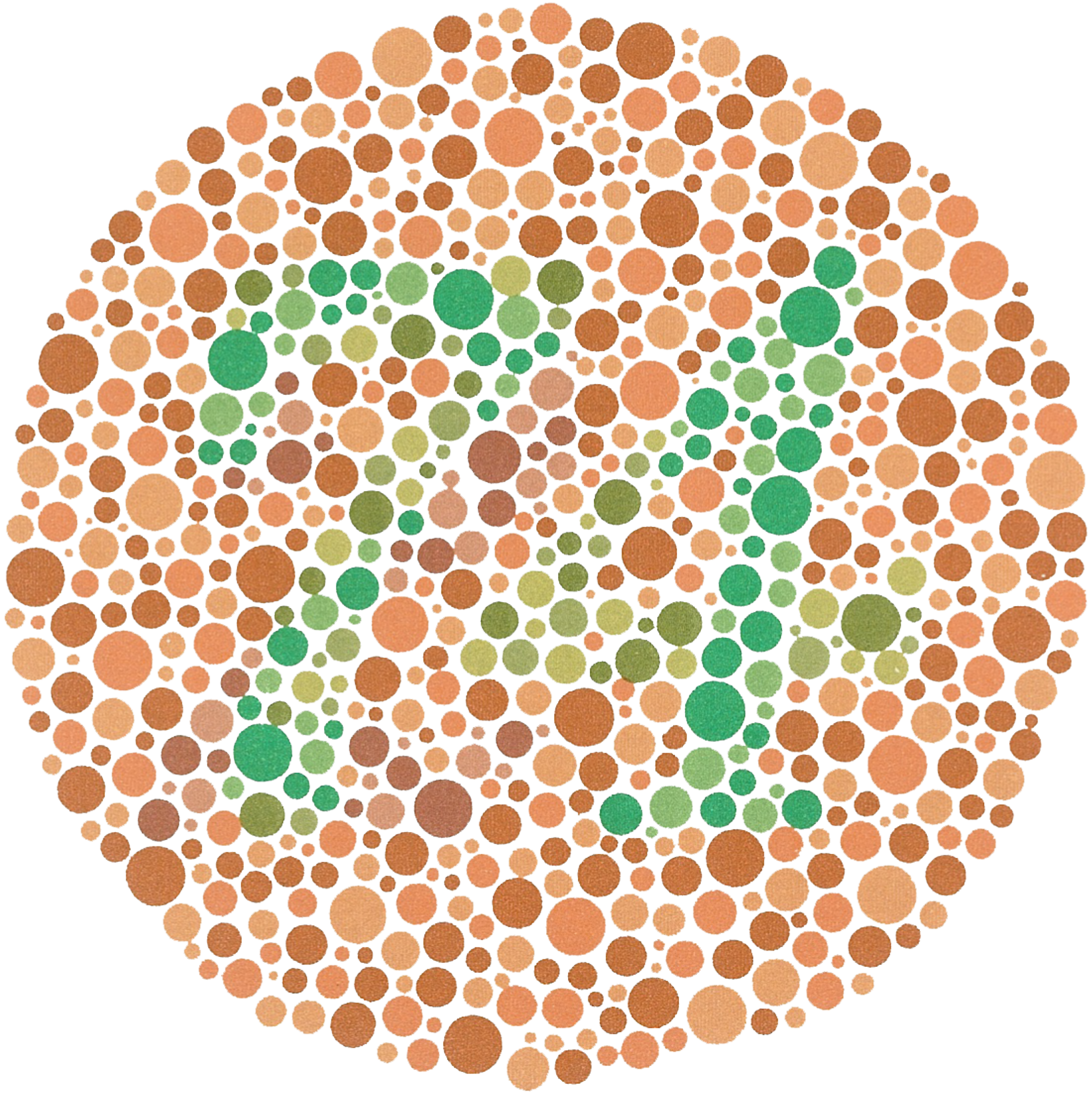
- Cones respond to different areas of the visible light spectrum
- Cover all wavelengths but certain wavelengths generate greater responses
- Color is determined by calculations based on the responses from the different cones
- Opponent Process Theory: three "opponent" channels
  - Light/Dark
  - Blue/Yellow
  - Red/Green
- Opposite colors are not perceived together

Opponent Process Theory



[Machado et. al, 2009]

Color Blindness

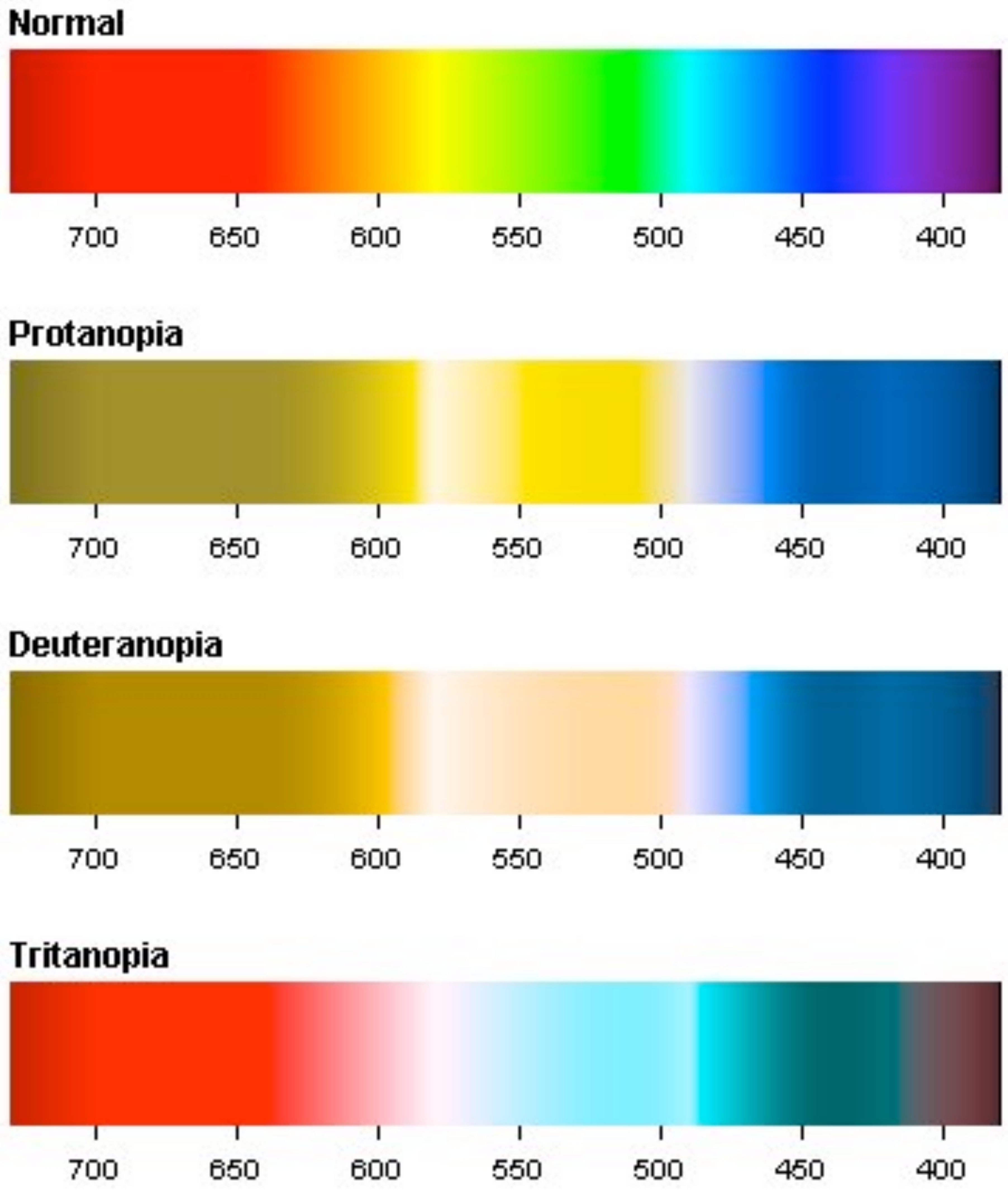


[Ishihara (Plate 9) via Wikipedia]

Color Blindness

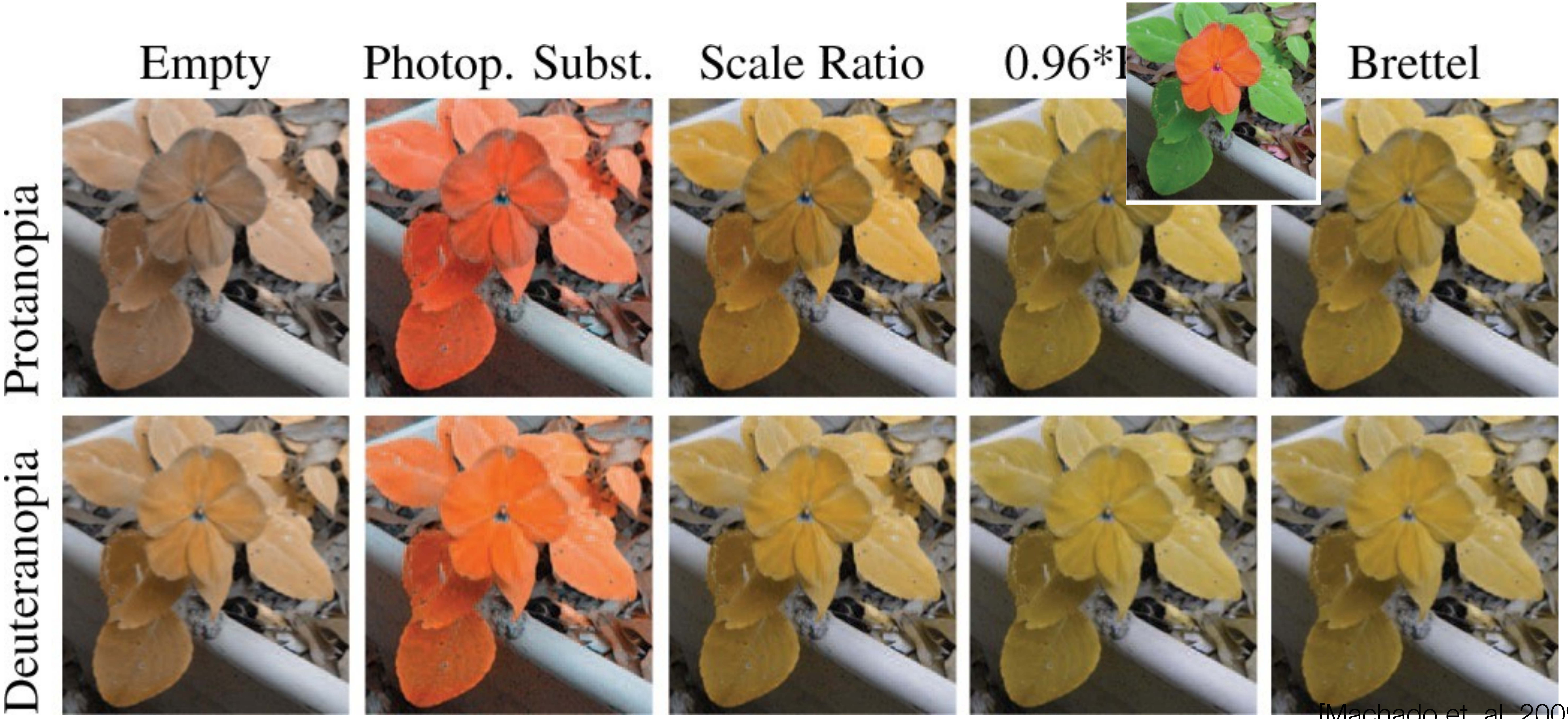
- Sex-linked: 8% of males and 0.4% of females of N. European ancestry
- Abnormal distribution of cones (e.g. missing the S, M, or L types)
- Either dichromatic (only two types of cones) or anomalous trichromatic (one type of cones has a defect)
    - Protanopia (L missing), Protanomaly (L defect)
    - Deuteranopia (M missing), Deuteranomaly (M defect) [Most Common]
    - Tritanopia (S missing), Tritanomaly (S defect) [Rare]
- Dichromacy is rarer than anomalous trichromacy
- Opponent process model explains why colors cannot be differentiated

Color Blindness



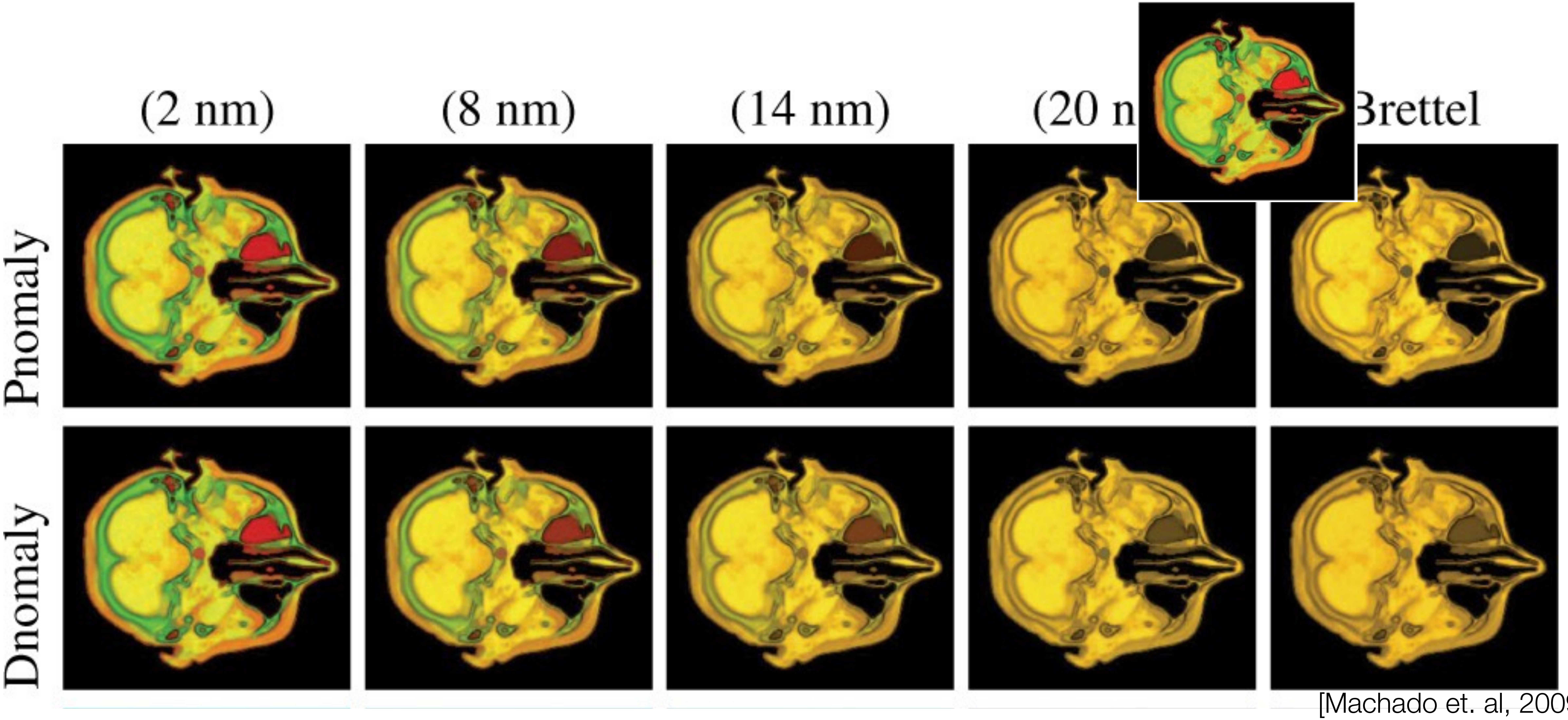
[via M. Meyer]

Simulating Color Blindness



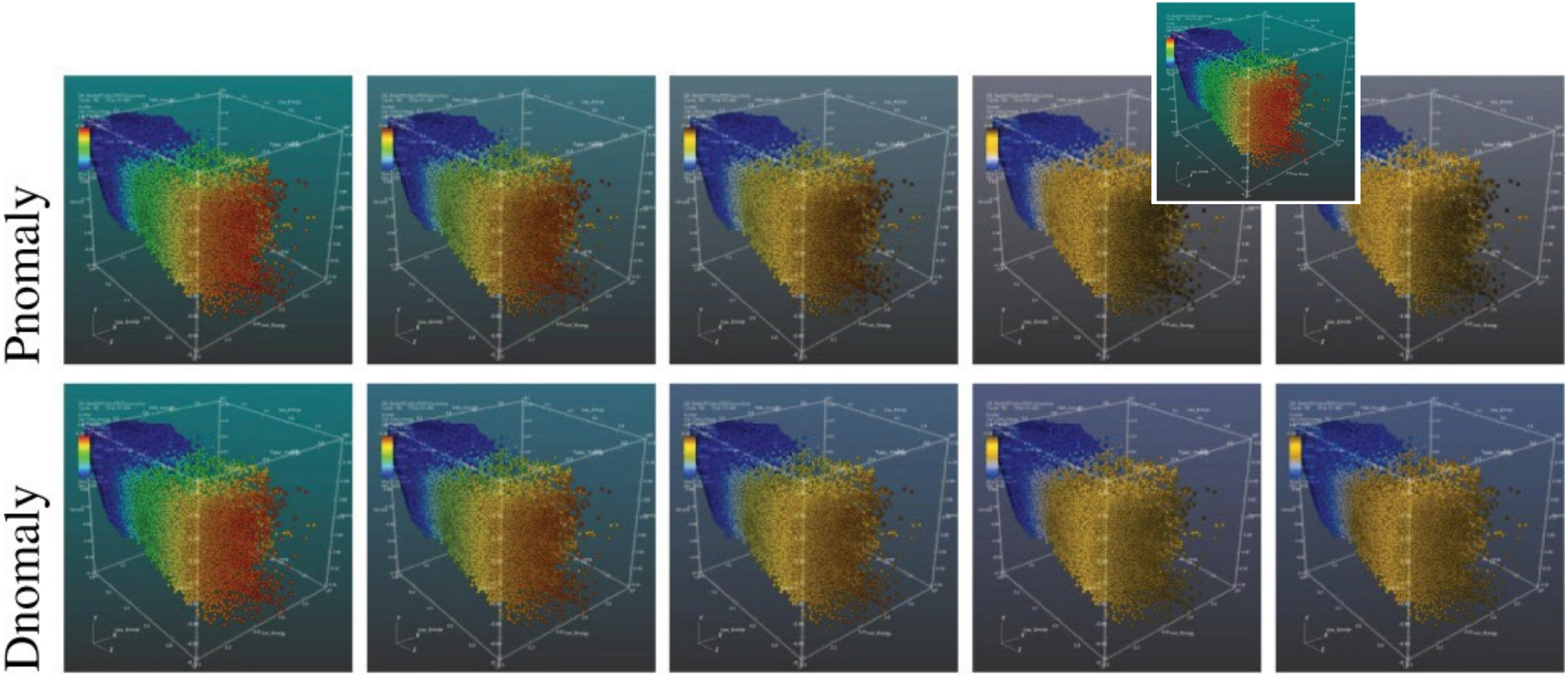
[Machado et. al, 2009]

Simulating Color Blindness



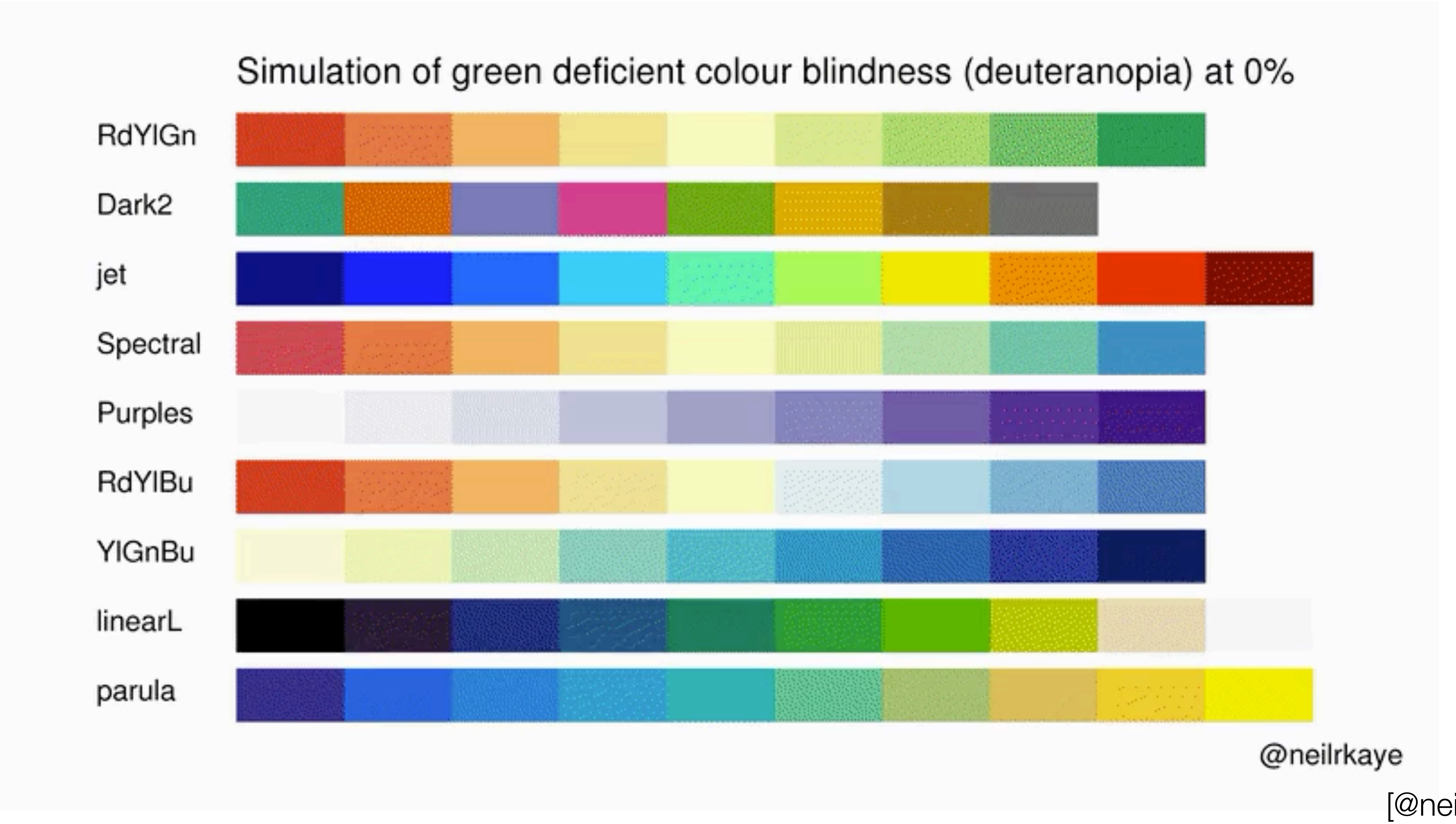
[Machado et. al, 2009]

Simulating Color Blindness



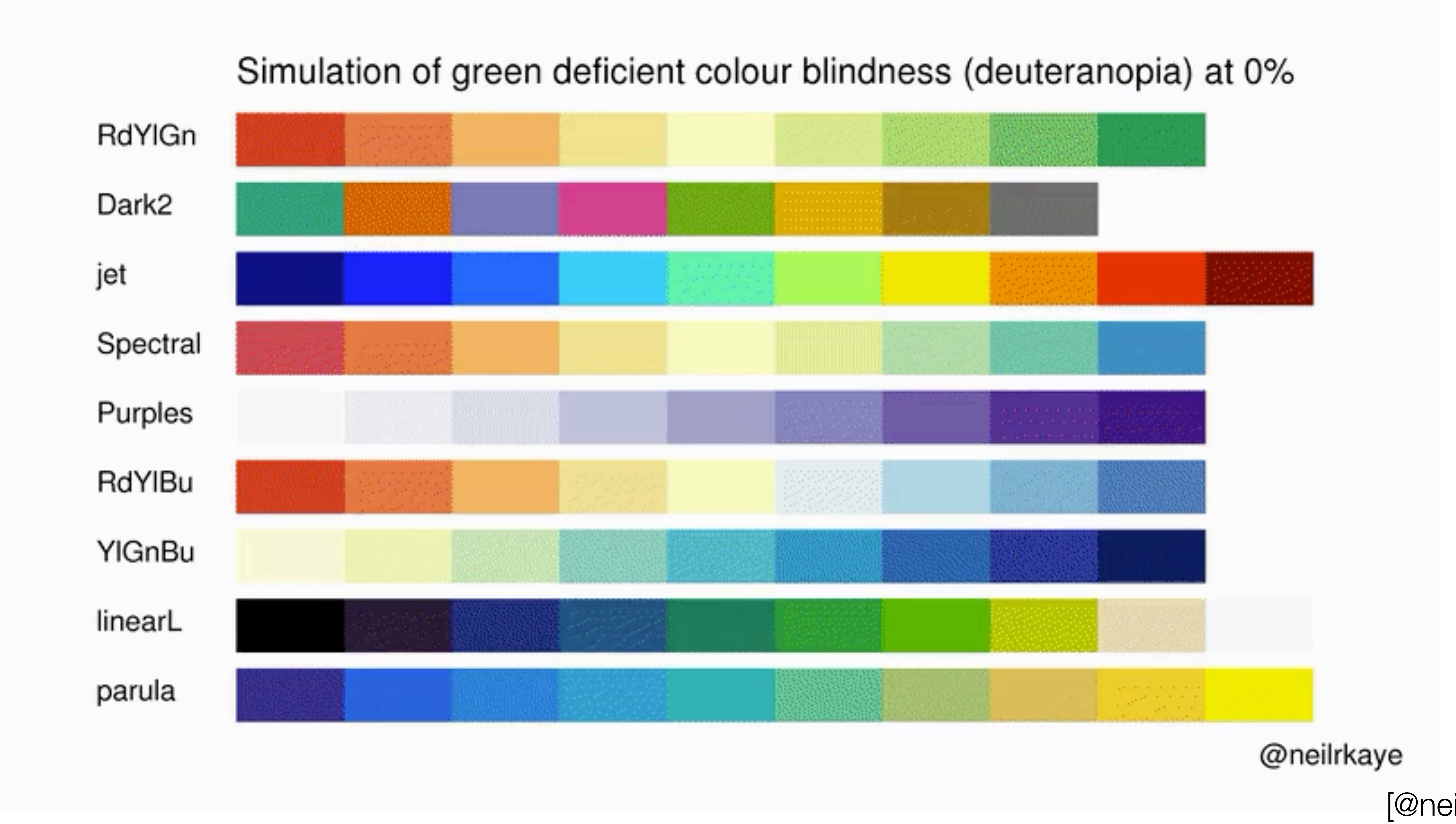
[Machado et. al, 2009]

Simulating Deuteranopia (Colormaps)



[@neilrkaye, reddit]

Simulating Deuteranopia (Colormaps)



[@neilrkaye, reddit]

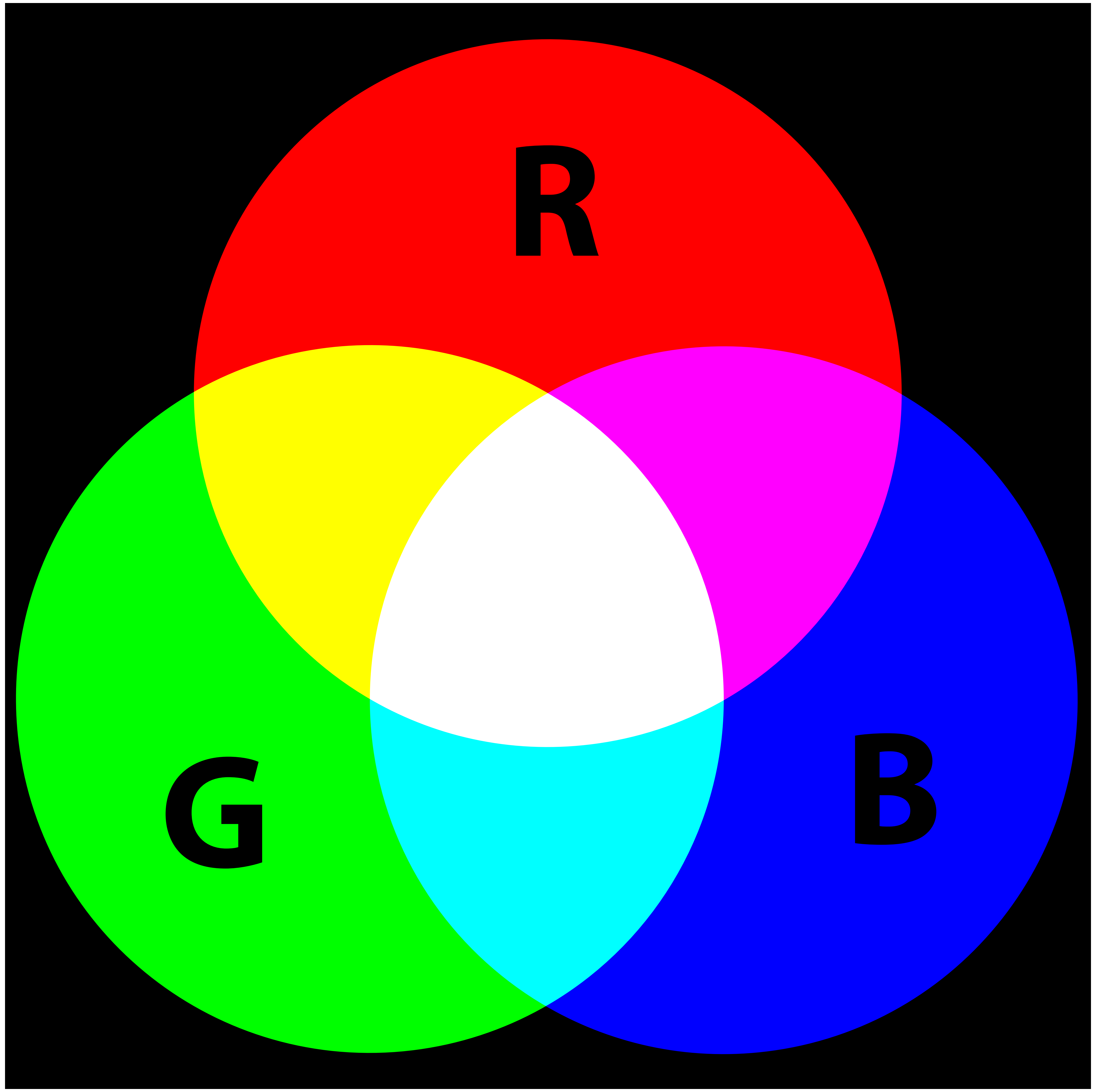
Primary Colors?

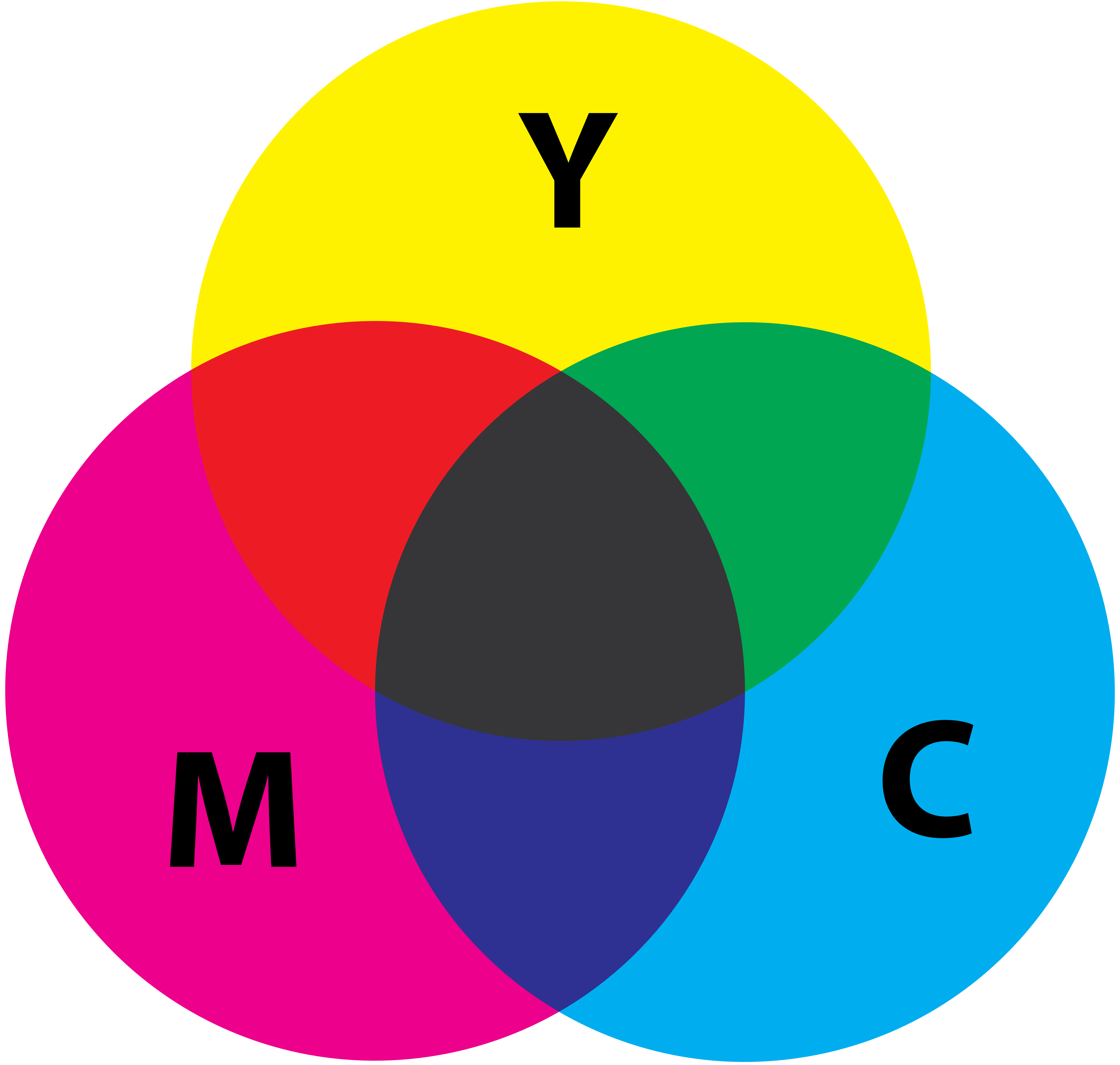
- Red, Green, and Blue
- Red, Yellow, and Blue
- Orange, Green, and Violet
- Cyan, Magenta, and Yellow

Primary Colors?

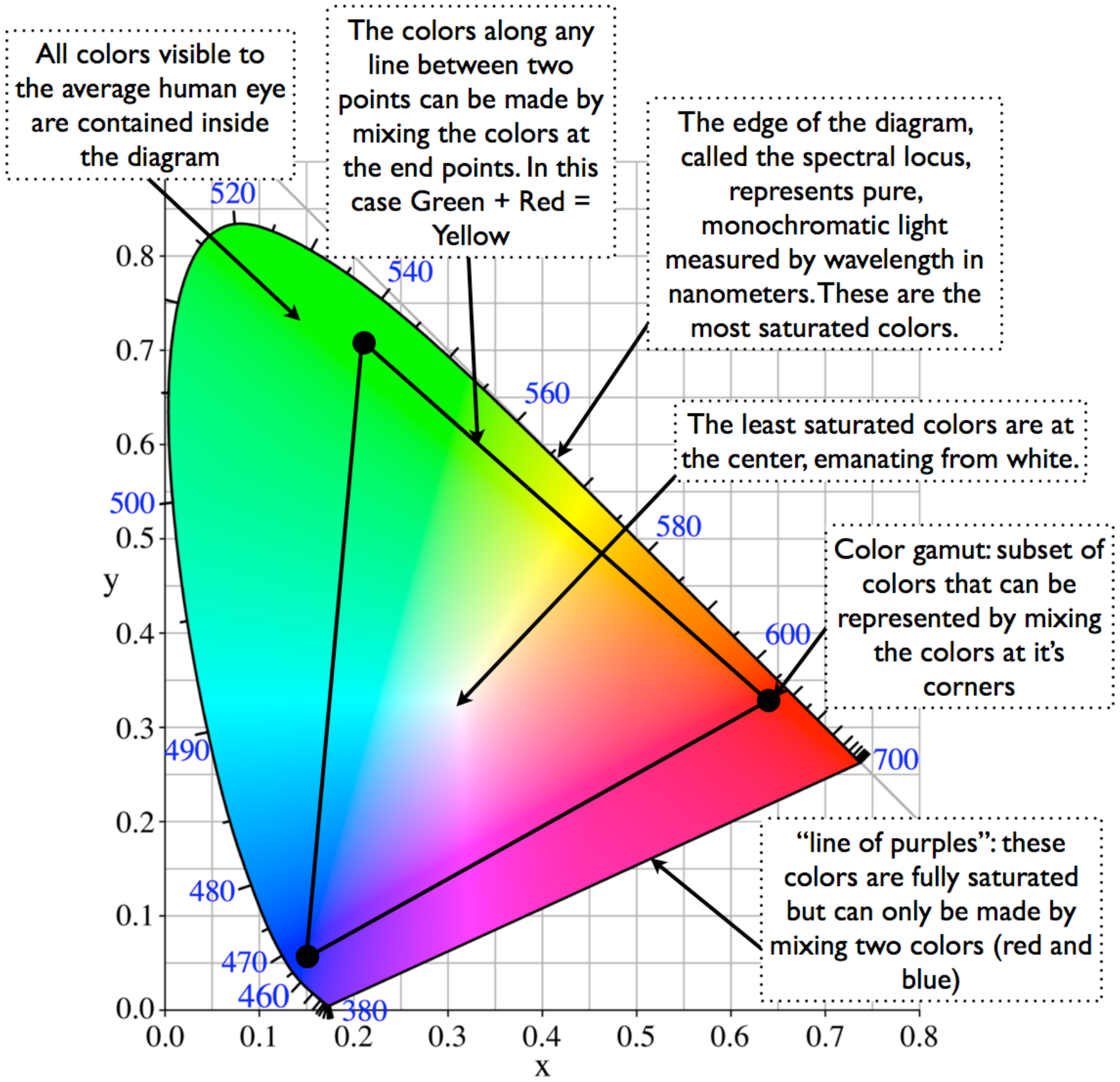
- Red, Green, and Blue
- Red, Yellow, and Blue
- Orange, Green, and Violet
- Cyan, Magenta, and Yellow
- **All of the above!**

Color Addition and Subtraction





Color Spaces and Gamuts



- **Color space**: the organization of all colors in space
  - Often human-specific, what we can see (e.g. CIELAB)
- **Color gamut**: a subset of colors
  - Defined by corners of color space
  - What can be produced on a monitor (e.g. using RGB)
  - What can be produced on a printer (e.g. using CMYK)
  - The gamut of your monitor != the gamut of someone else's or a printer

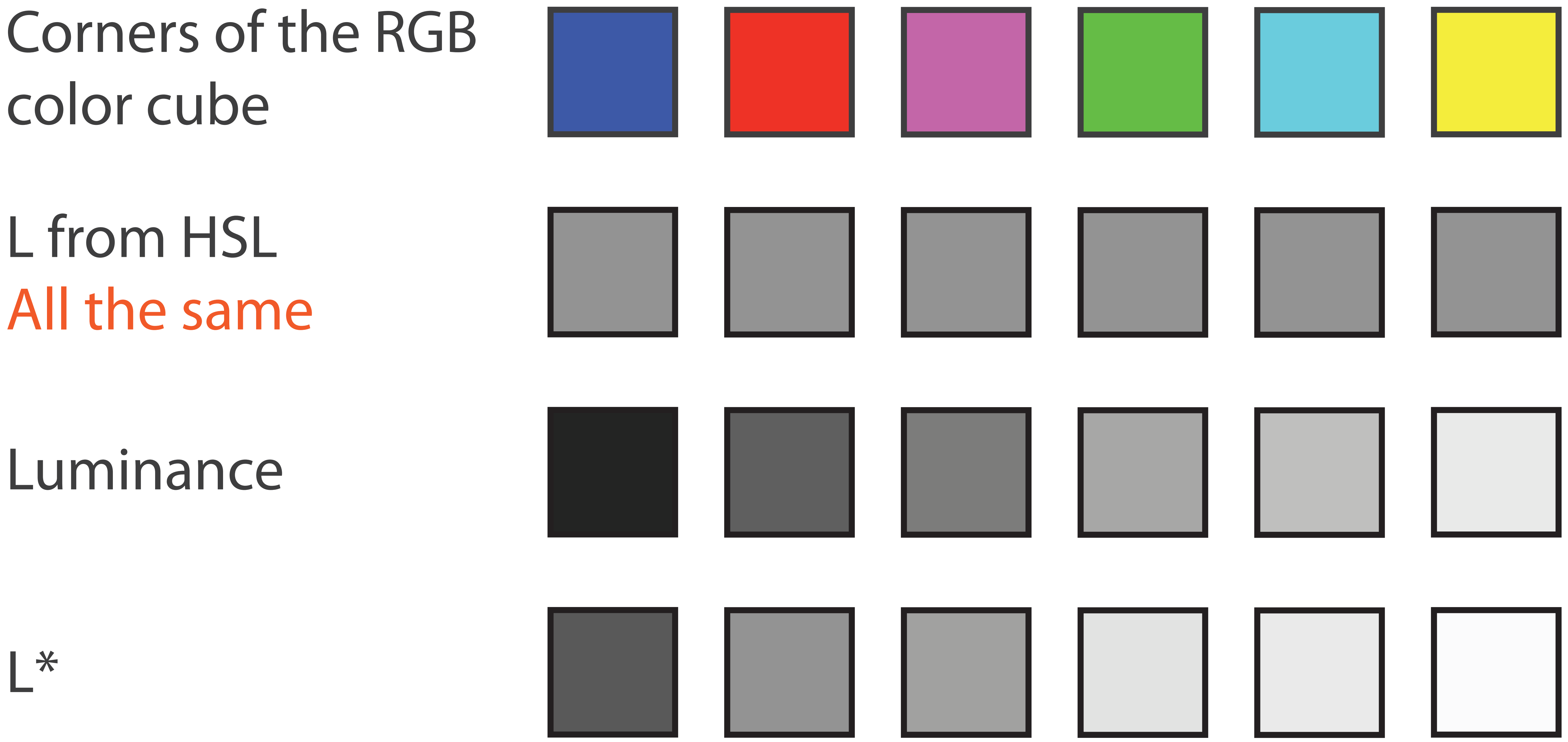
[Anatomy of a CIE Chromaticity Diagram]

Color Models

- A **color model** is a representation of color using some basis
- RGB uses three numbers (red, blue, green) to represent color
- Color space ~ color model, but there can be many color models used in the same color space (e.g. OGV)
- Hue-Saturation-Lightness (HSL) is more intuitive and useful
  - Hue captures pure colors
  - Saturation captures the amount of white mixed with the color
  - Lightness captures the amount of black mixed with a color
  - HSL color pickers are often circular
- Hue-Saturation-Value (HSV) is similar (swap black with gray for the final value), linearly related

Luminance

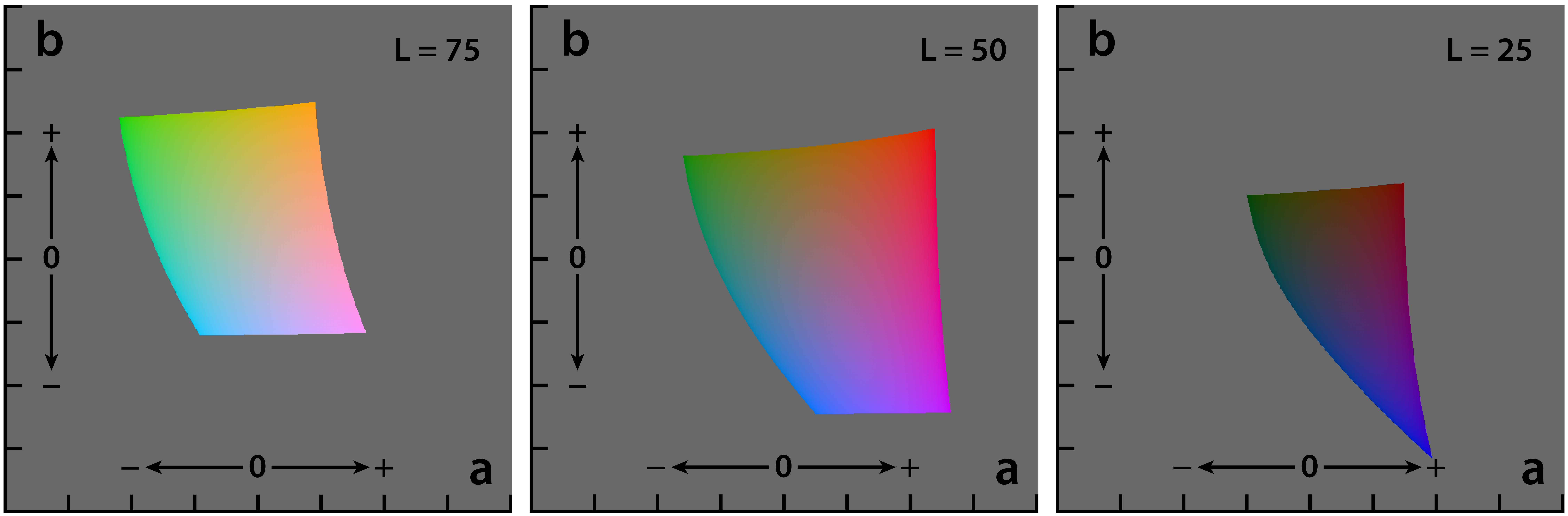
- HSL does not truly reflect the way we perceive color
- Even though colors have the same lightness, we perceive their luminance differently
- Our perception (L*) is **nonlinear**



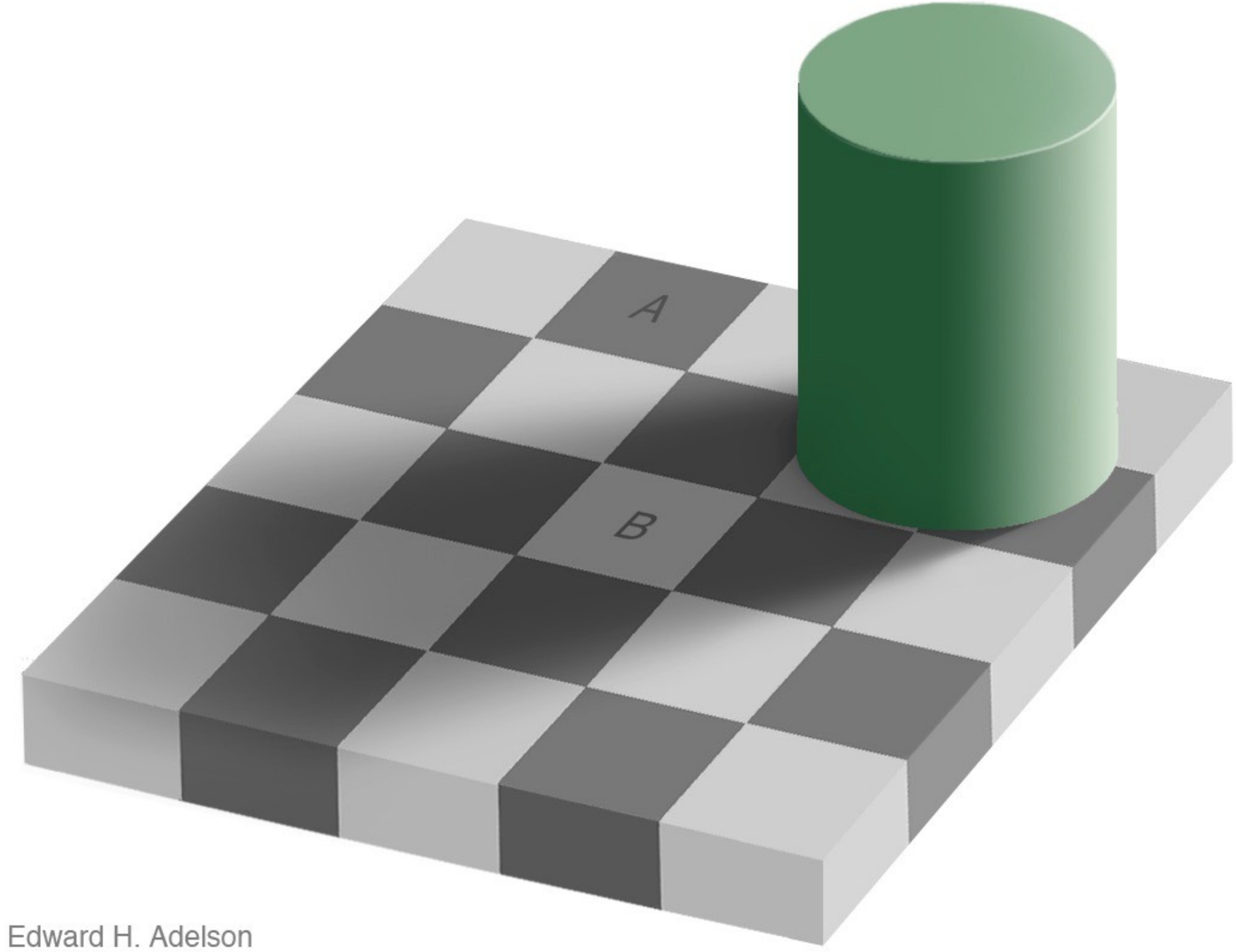
[Munzner (ill. Maguire), 2014 (based on Stone, 2006)]

Perceptually Uniform Color Spaces

- L*a*b* allows perceptually accurate comparison and calculations of colors

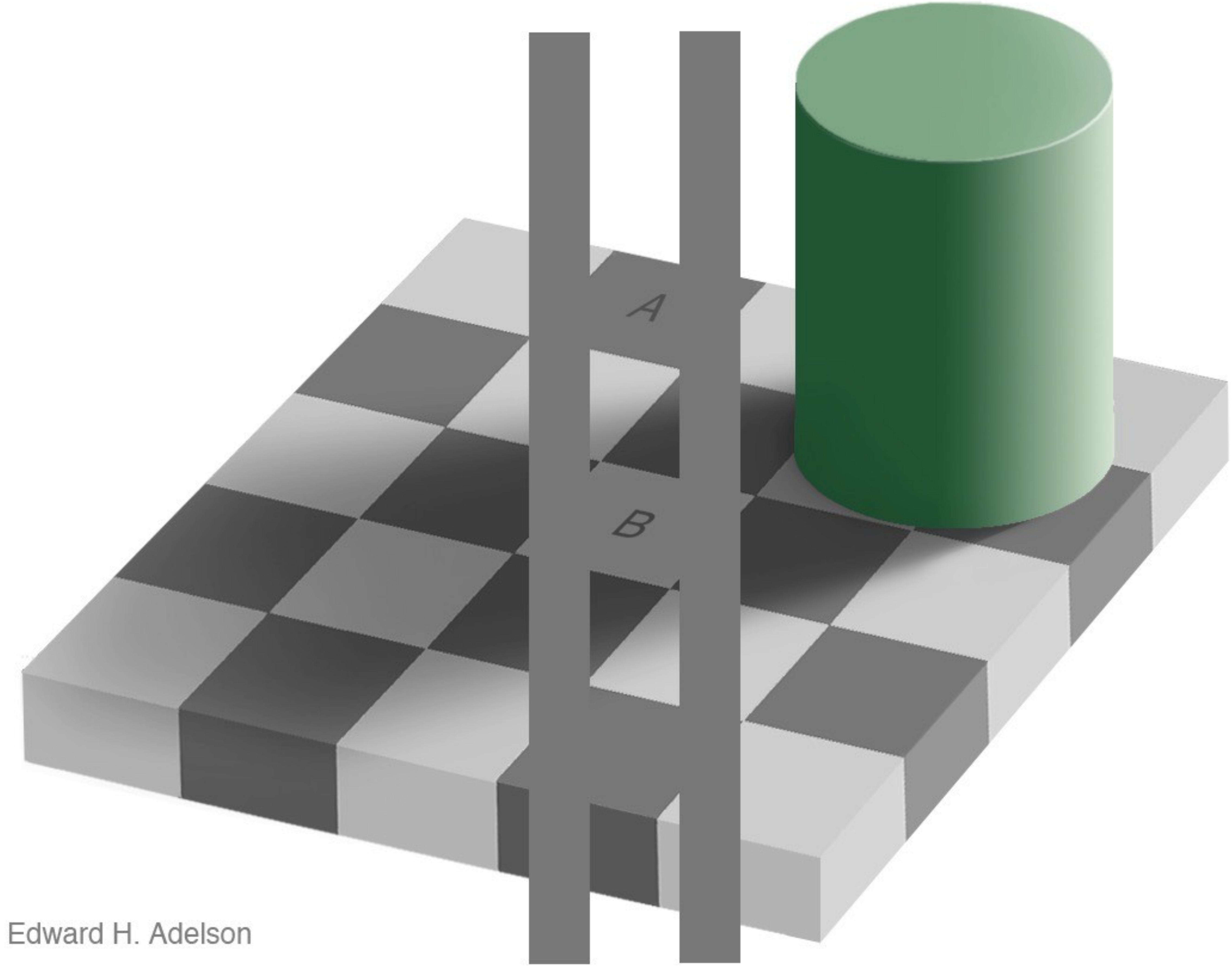


Luminance Perception (Spatial Adaption)



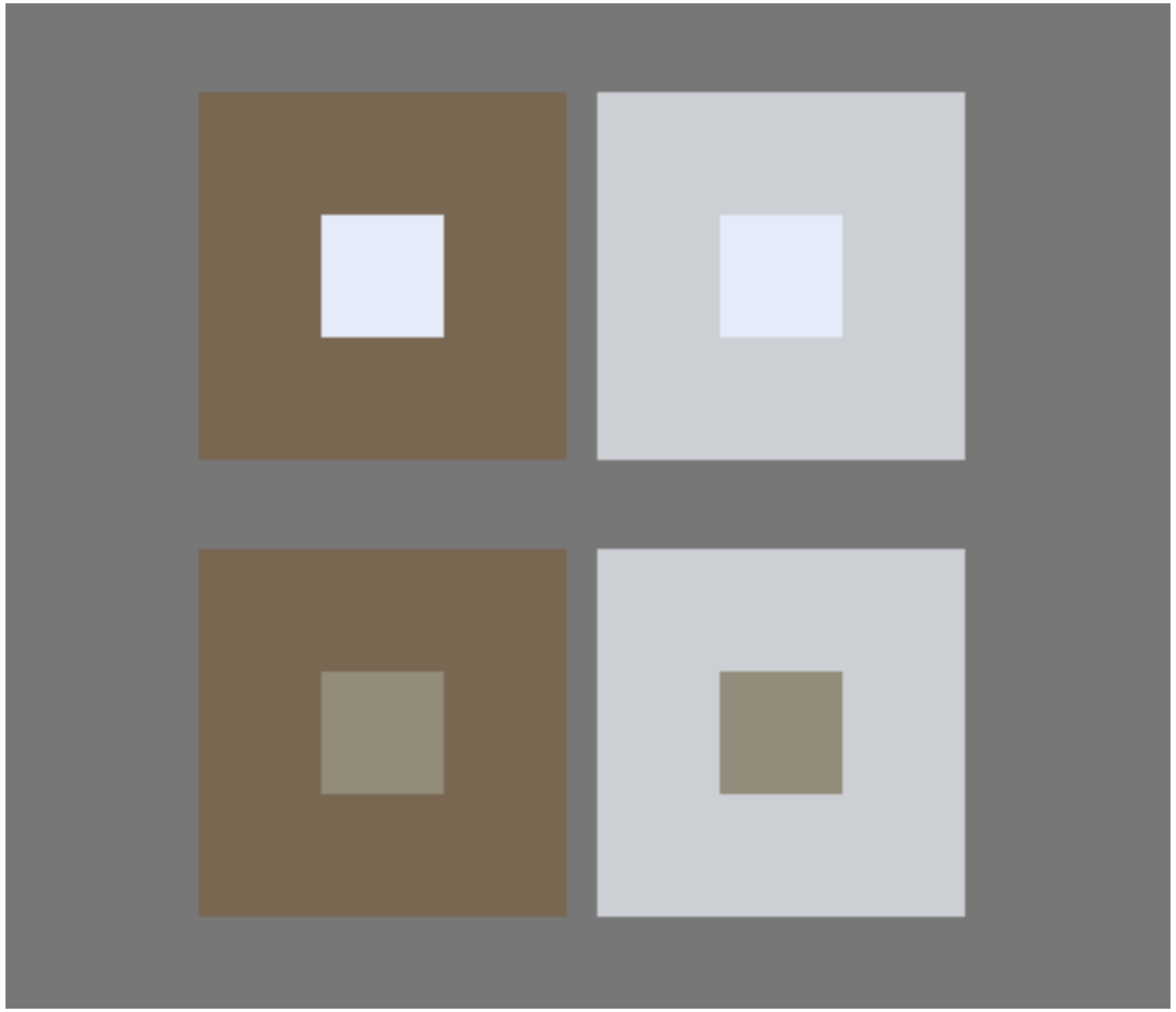
[E. H. Adelson, 1995]

Luminance Perception (Spatial Adaption)

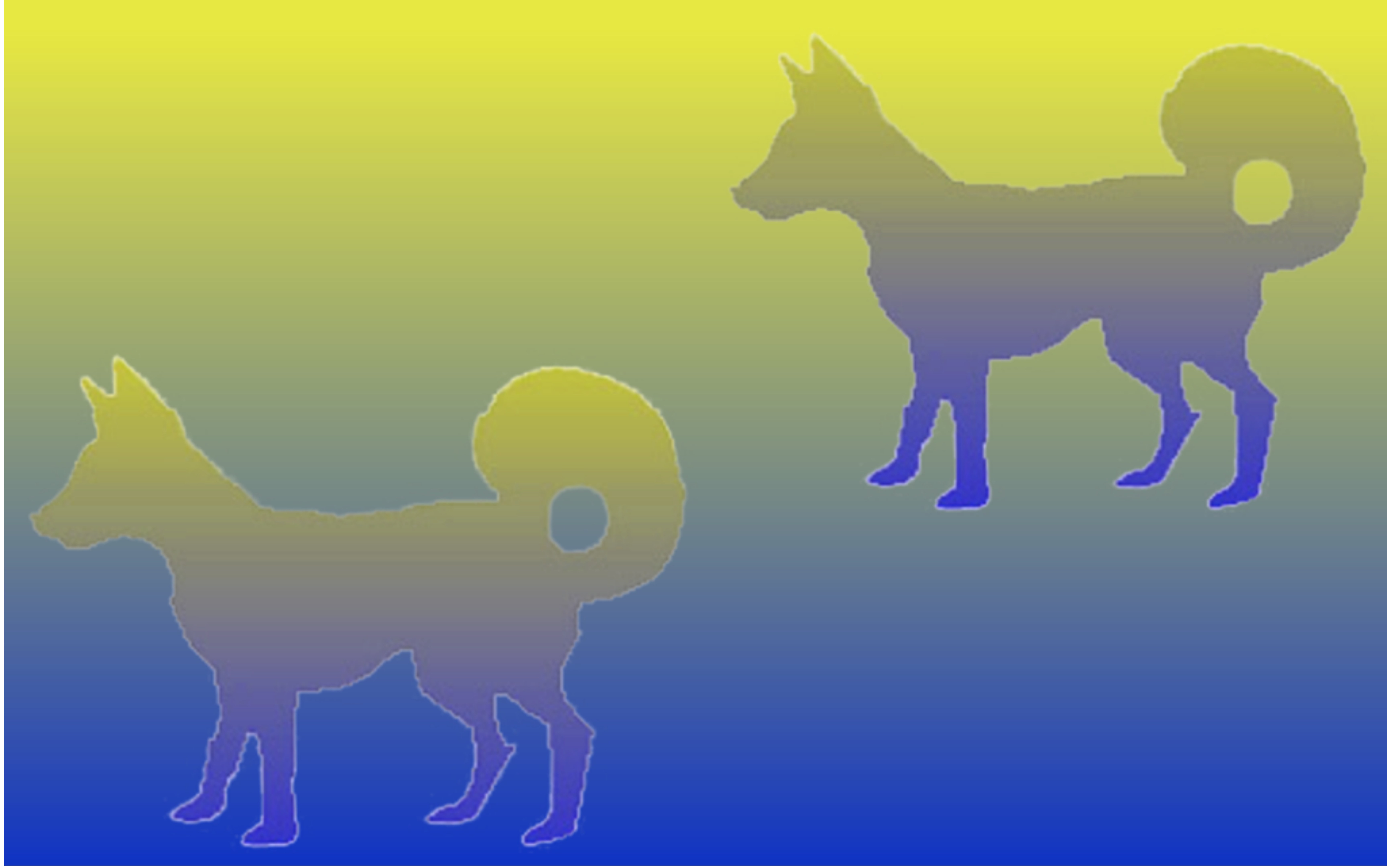


[E. H. Adelson, 1995]

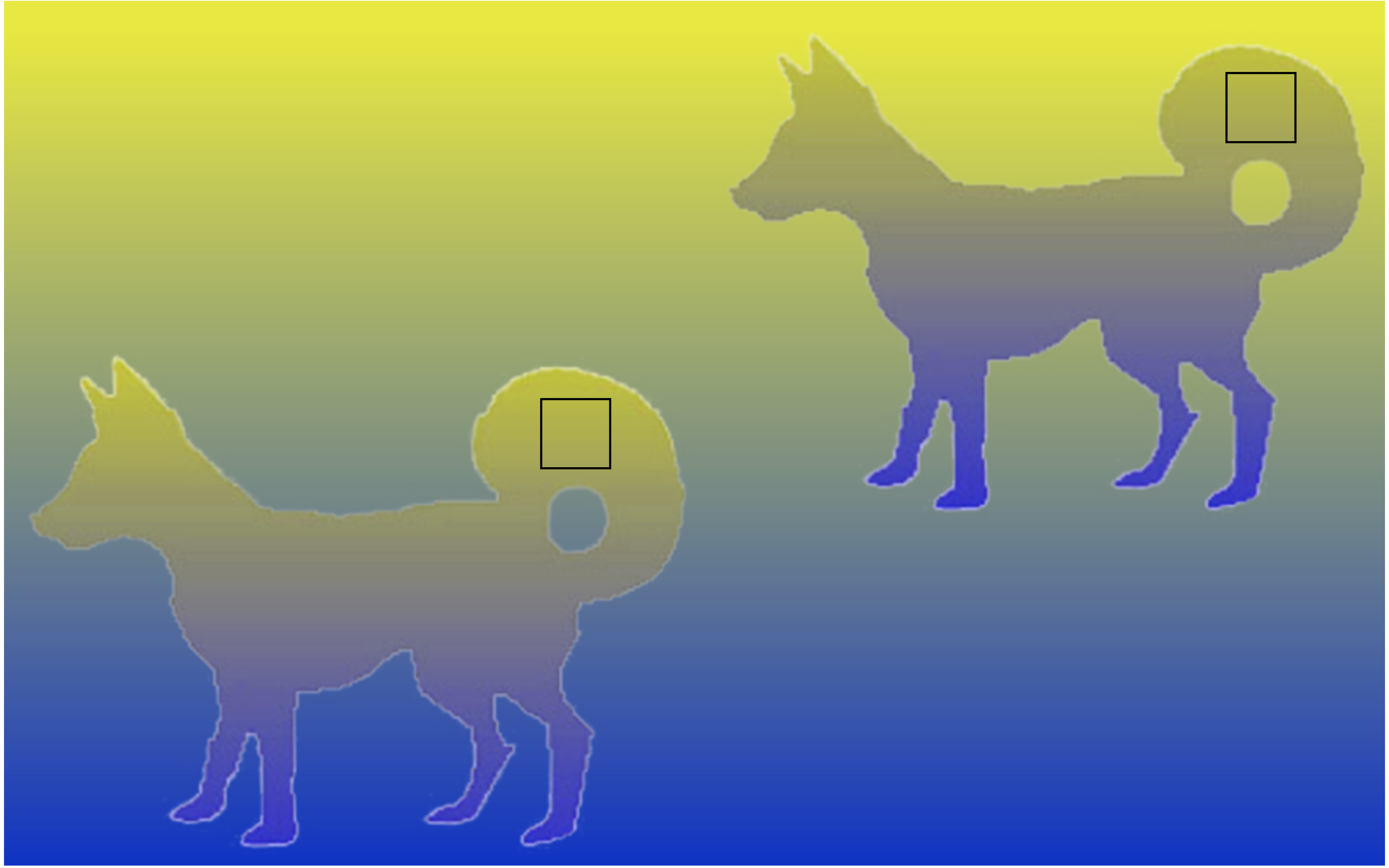
Simultaneous Contrast



Simultaneous Contrast

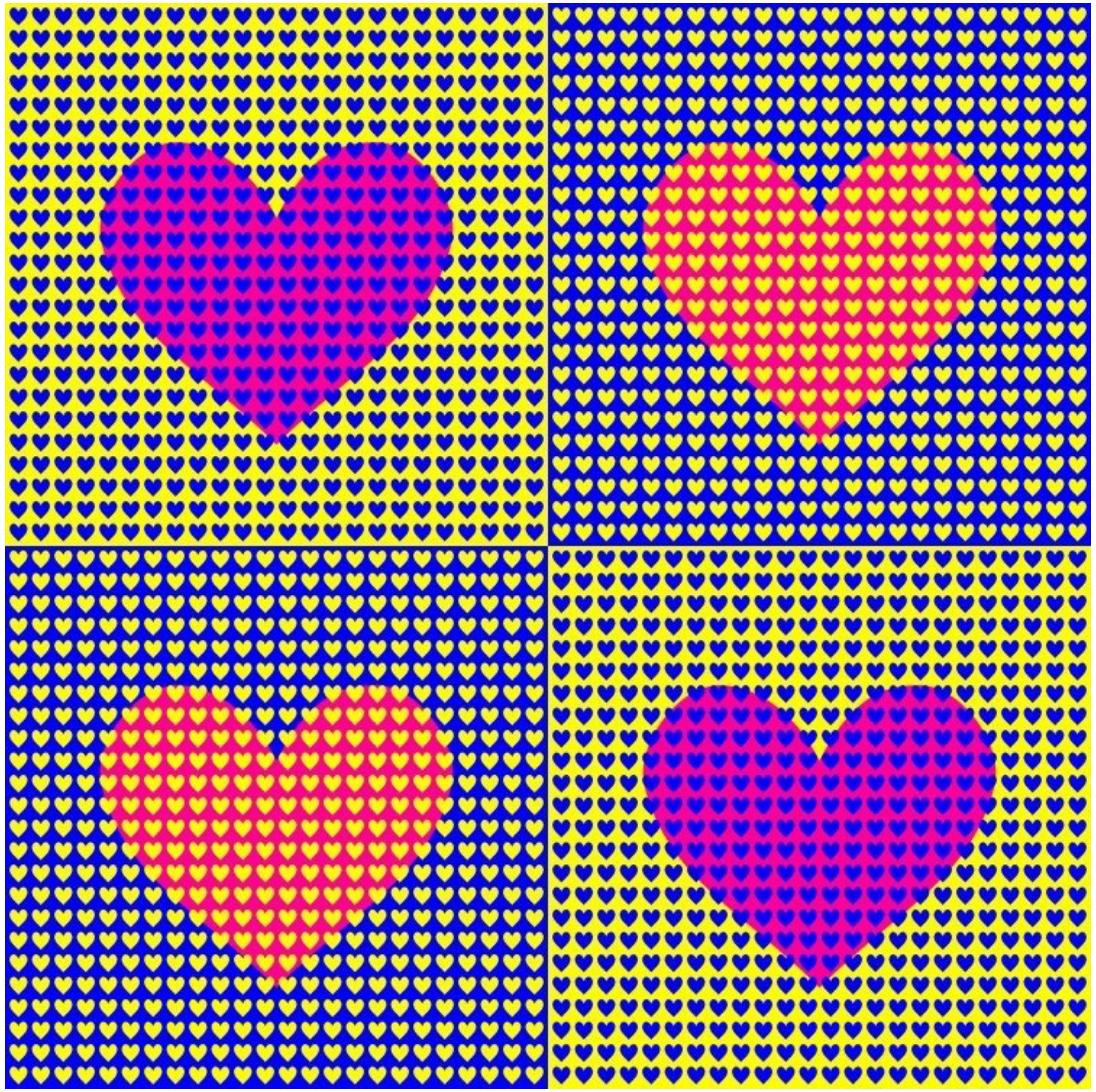


Simultaneous Contrast



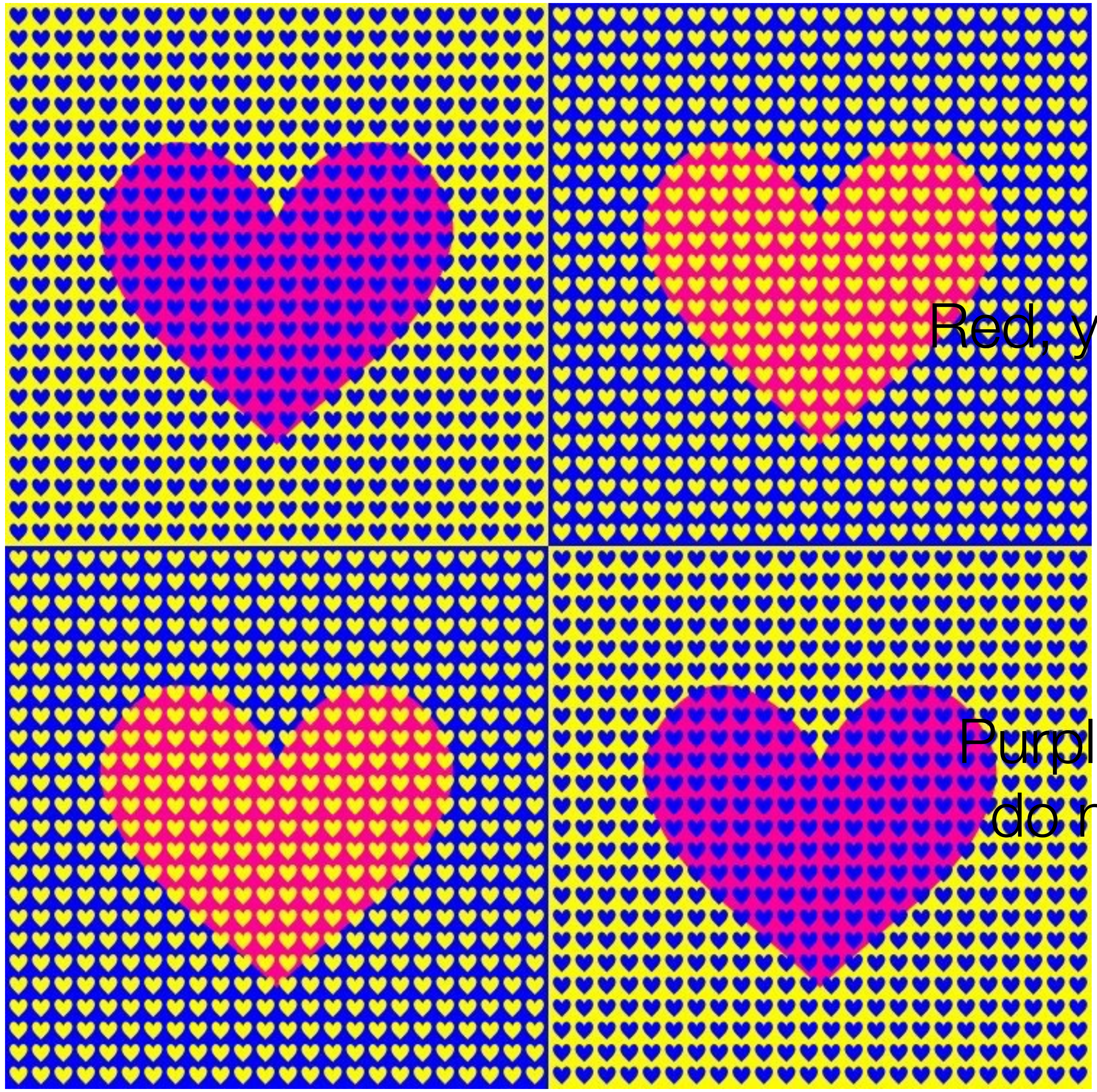
Simultaneous Contrast

What colors?



[A. Kitaoka]

What colors?



Red, yellow, blue

Purple, orange do not exist!

[A. Kitaoka]

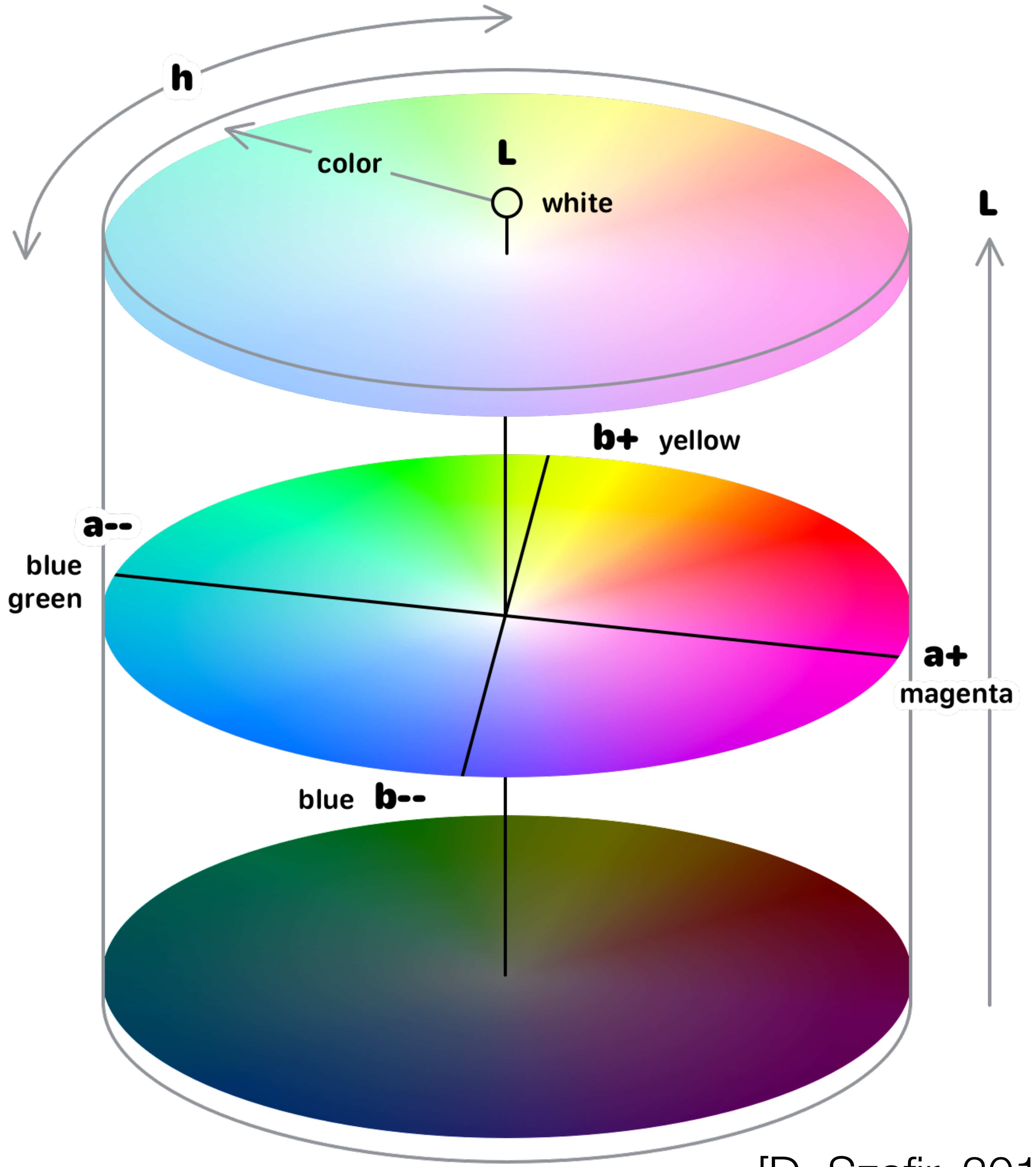
What does this mean for visualization?

What does this mean for visualization?

- We need to be aware of colorblindness when encoding via color
- Our brains may misinterpret color (surrounding colors matter!) even if we aren't colorblind
- Be careful! Don't assume that adding color always works the way you intended
- Use known colormaps when possible

Violations of CIELAB Assumptions

- CIELAB:
  - Approximately perceptually linear
  - 1 unit of Euclidean distance = 1 Just Noticeable Difference (JND)
  - JND: people detect change at least 50% of the time
- Assumptions CIELAB makes:
  - Simple world
  - Isolation
  - Geometric

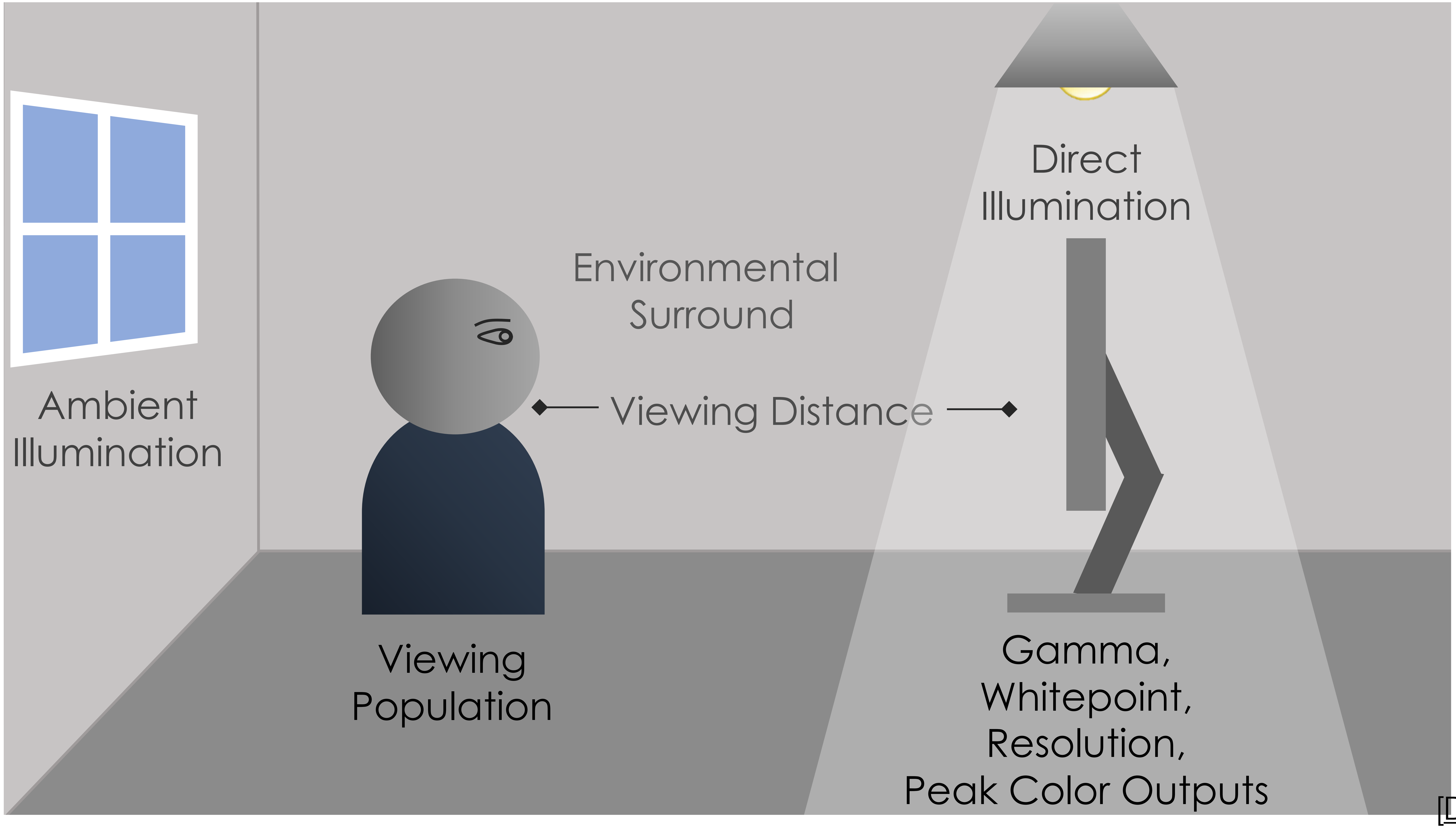


[D. Szafir, 2017]

Simple World Assumption

[D. Szafir, 2017]

Problems with Simple World Assumption

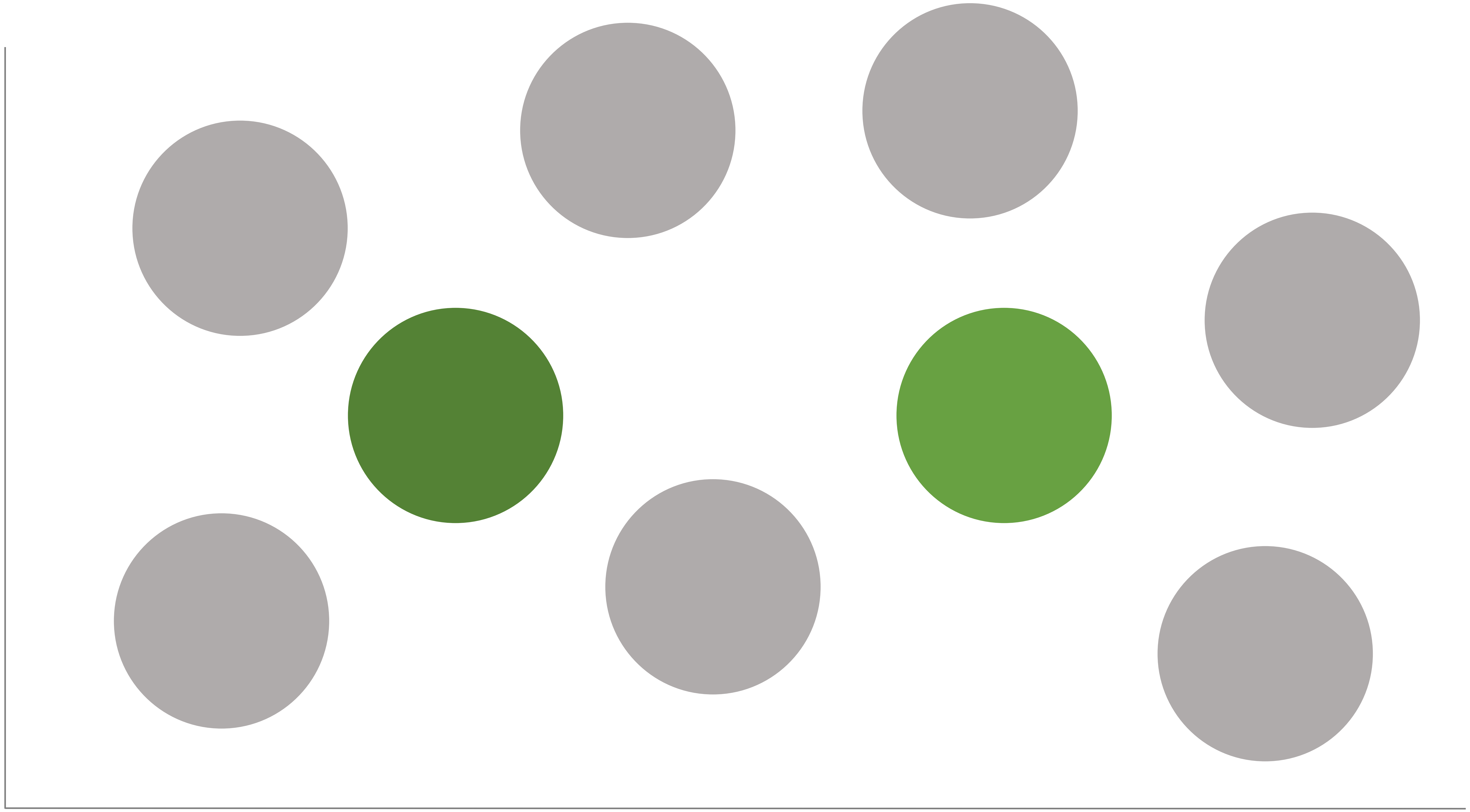


[D. Szafir, 2017]

Isolation Assumption

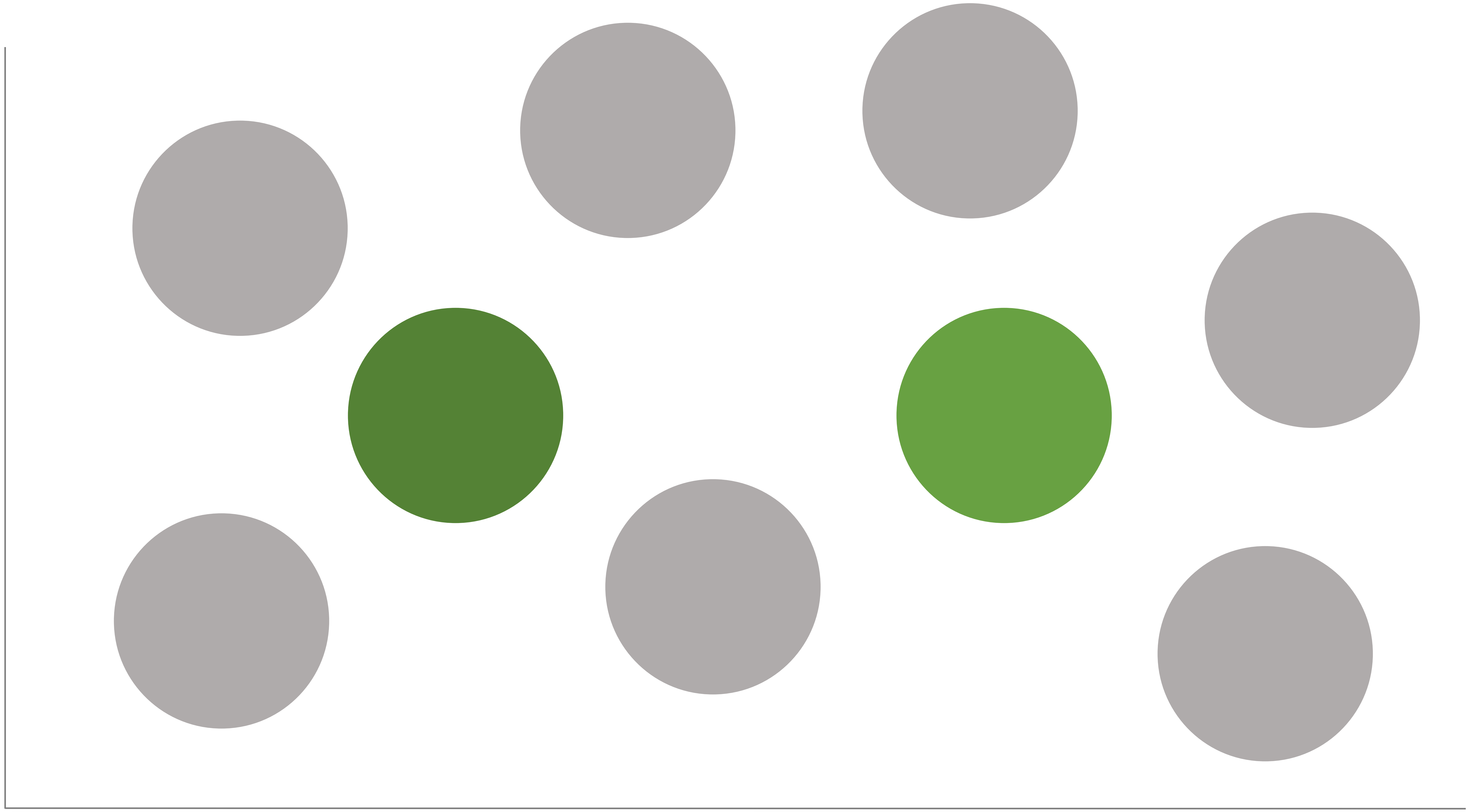
[D. Szafir, 2017]

Problems with Isolation Assumption



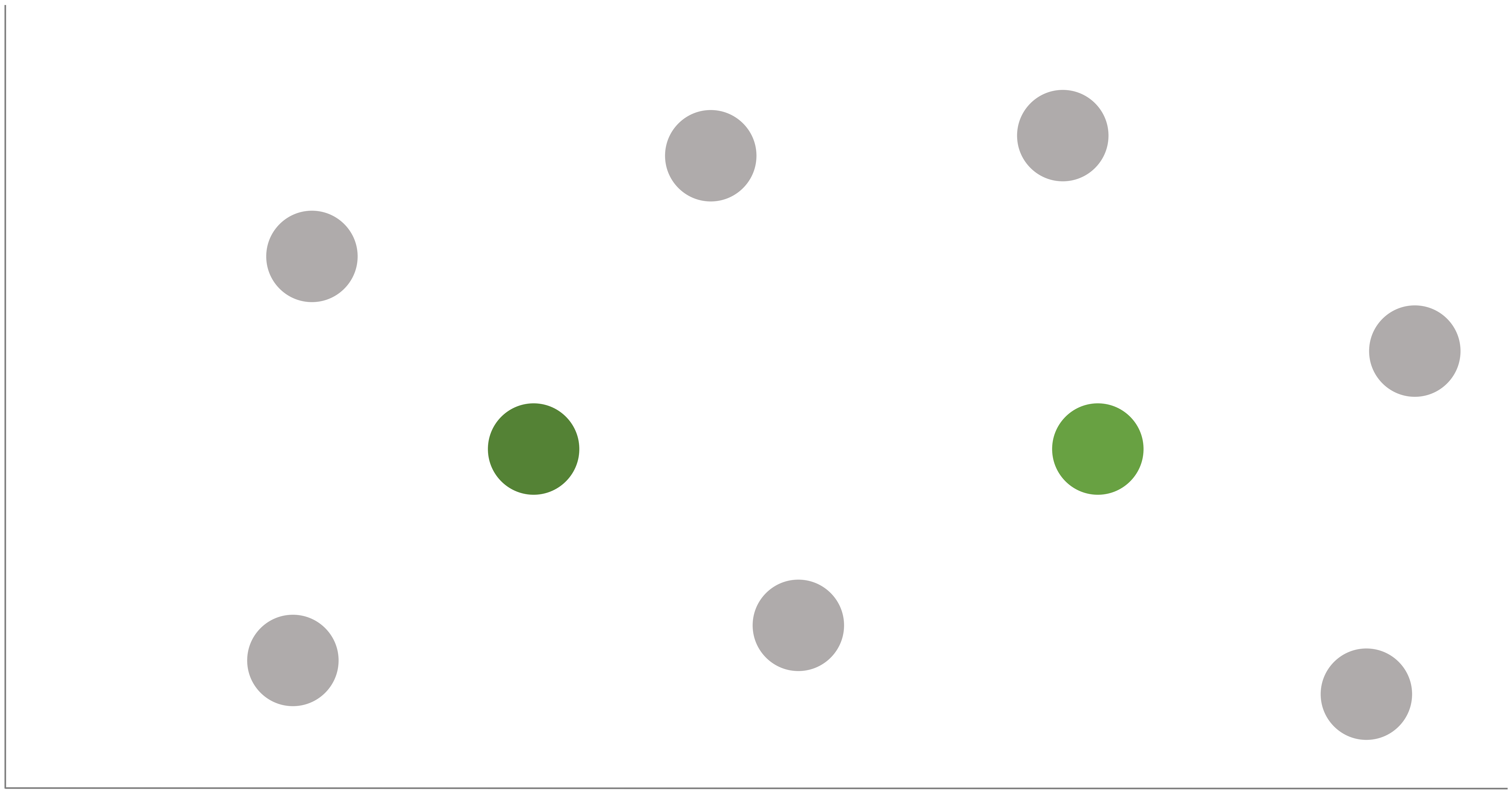
[D. Szafir, 2017]

Geometric Assumption



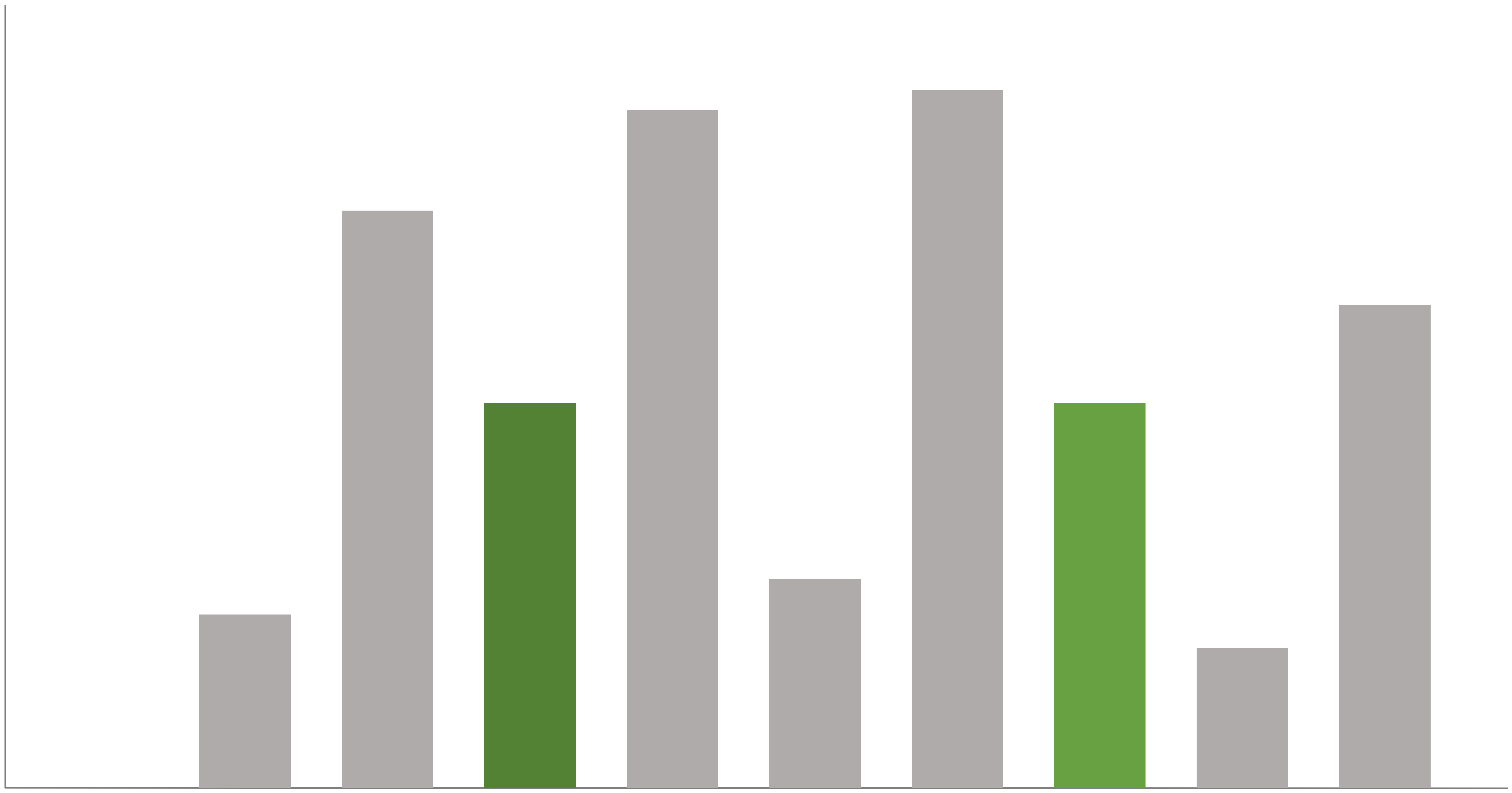
[D. Szafir, 2017]

Size Problem with Geometric Assumption



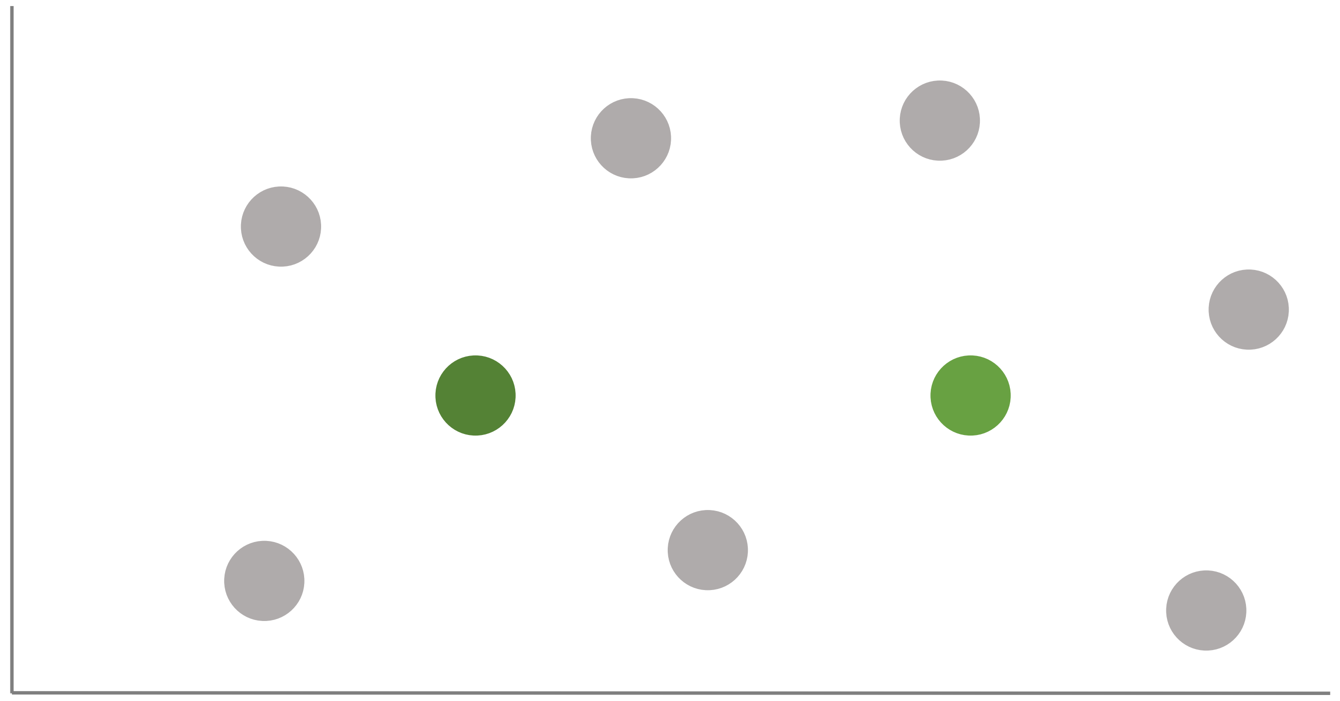
[D. Szafir, 2017]

Shape Problem with Geometric Assumption

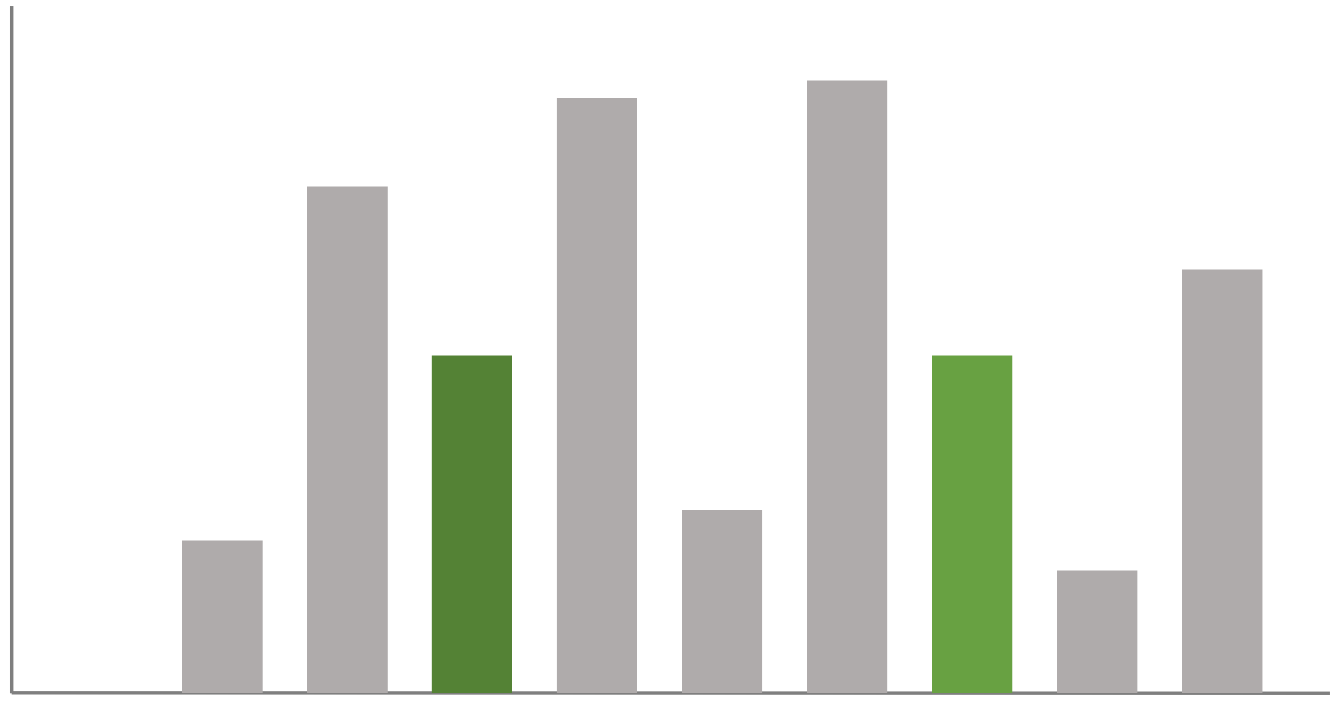


[D. Szafir, 2017]

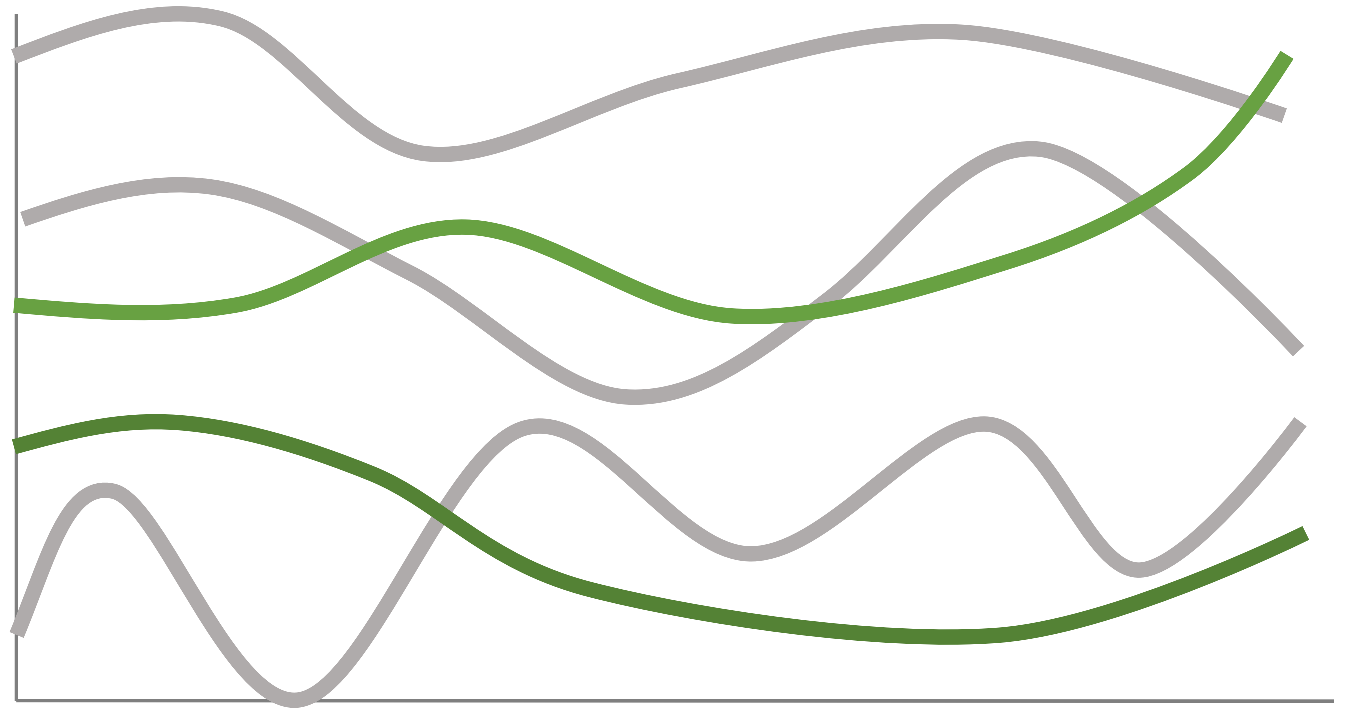
Types of Geometry



Diagonally Symmetric Marks



Elongated Marks



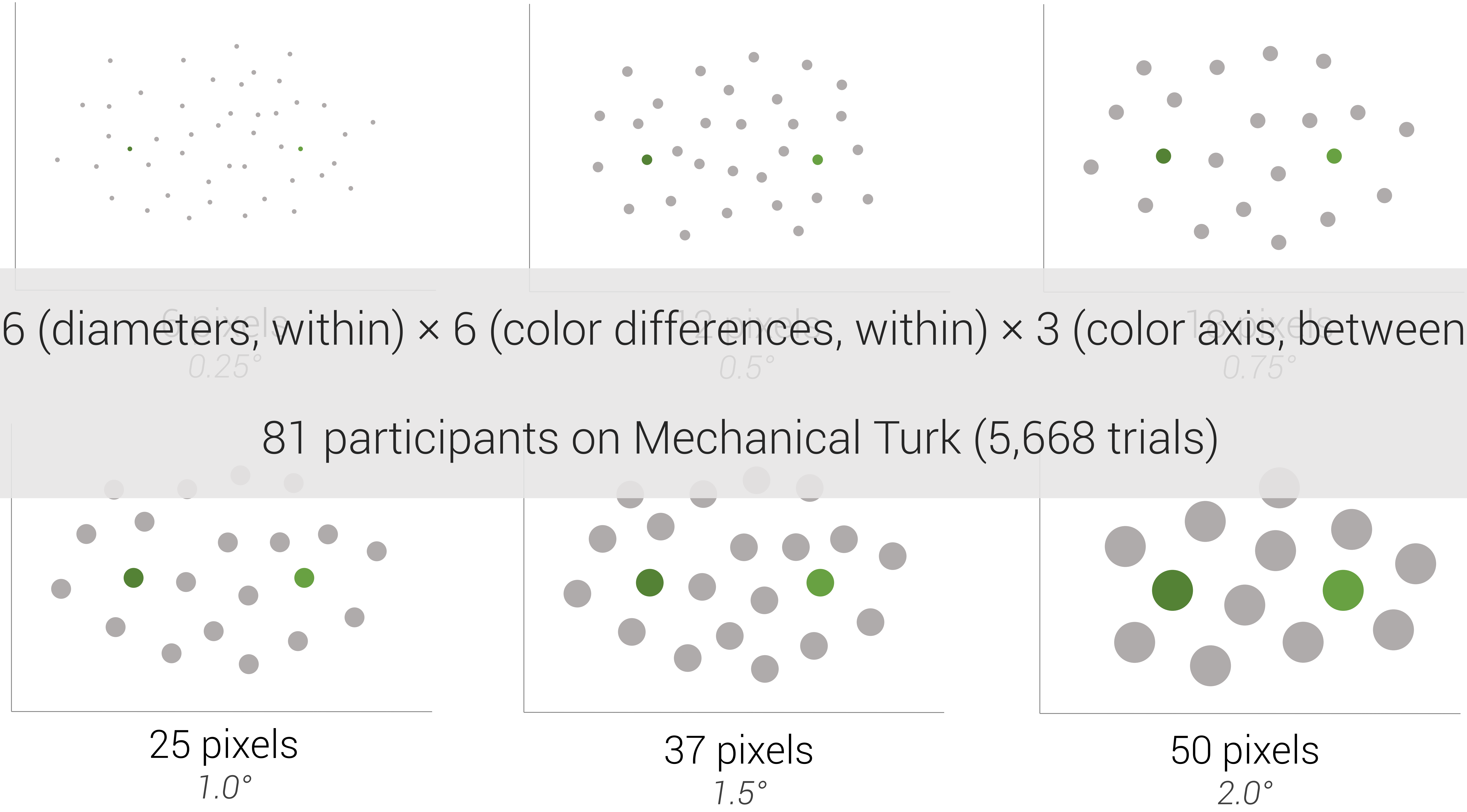
Asymmetric Marks

Area Marks

[D. Szafir, 2017]

Run the tests!

Color Study

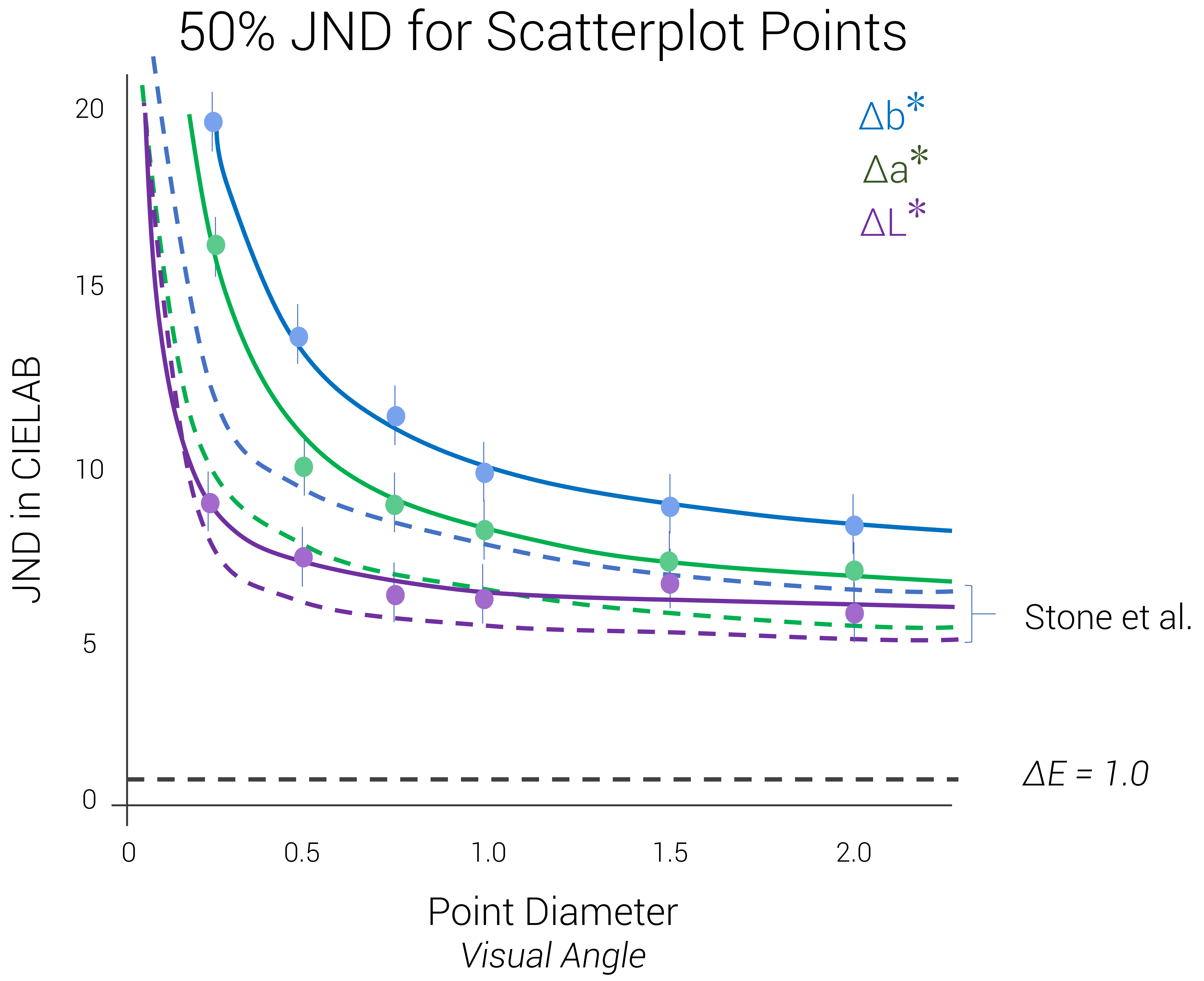


6 (diameters, within) × 6 (color differences, within) × 3 (color axis, between)

81 participants on Mechanical Turk (5,668 trials)

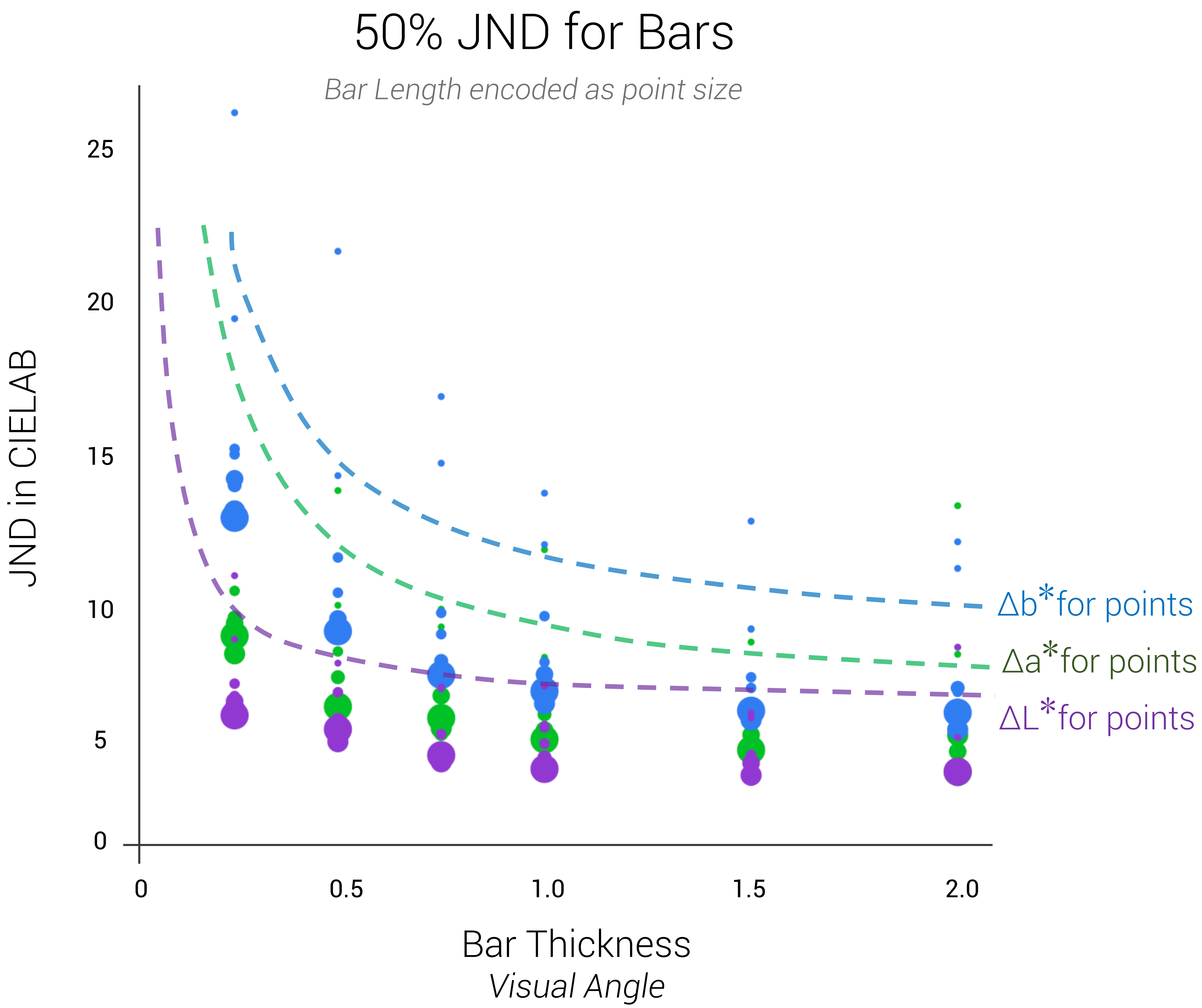
[D. Szafir, 2017]

Point Size: consistent with previous results



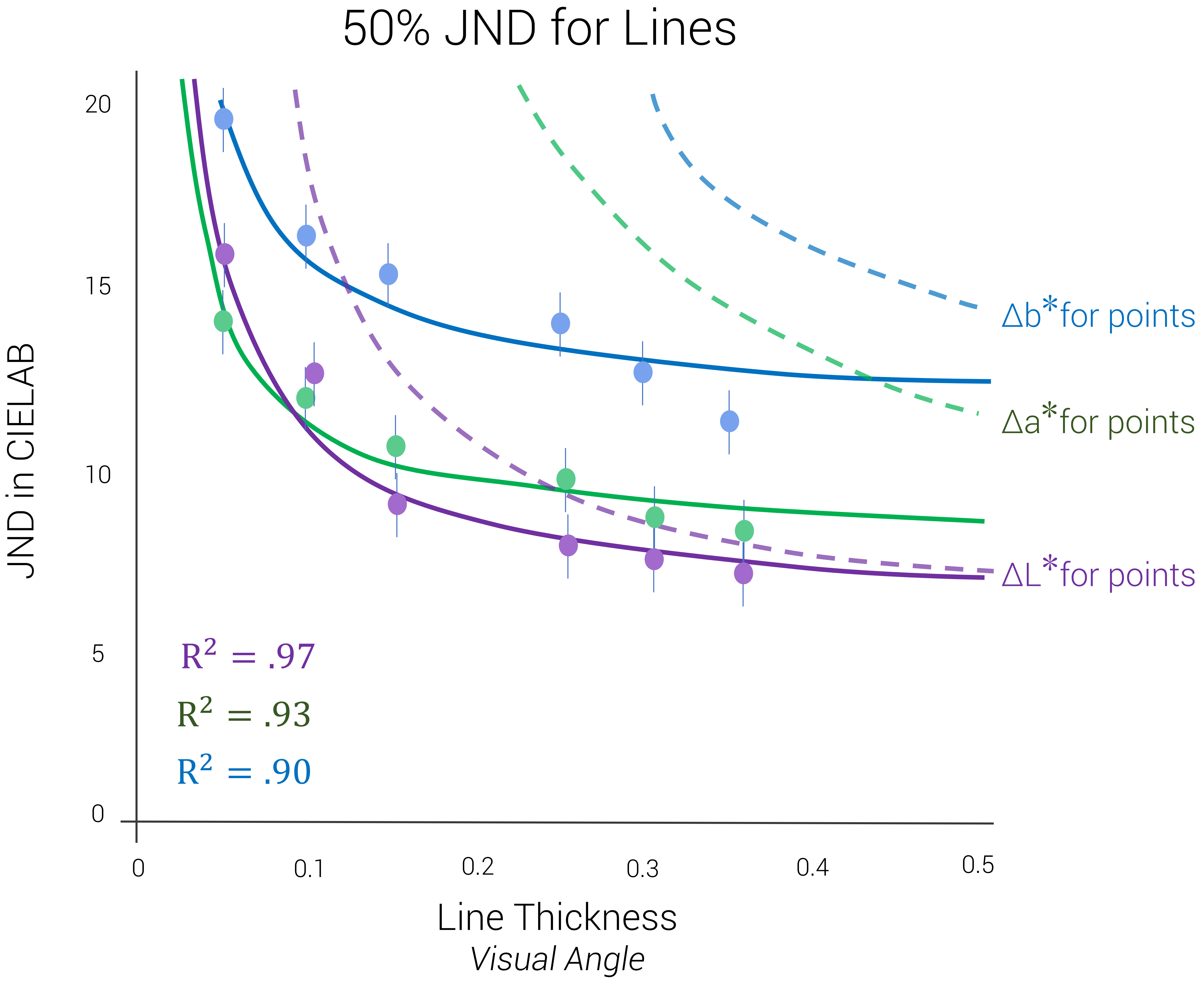
[D. Szafir, 2017]

Bar Thickness and Length: longer bars help



[D. Szafir, 2017]

Line Thickness: better than points



[D. Szafir, 2017]

Color perception in real-world visualizations is complicated

Akiyoshi Kitaoka's Illusion pages