Data Visualization (CIS 490/680)

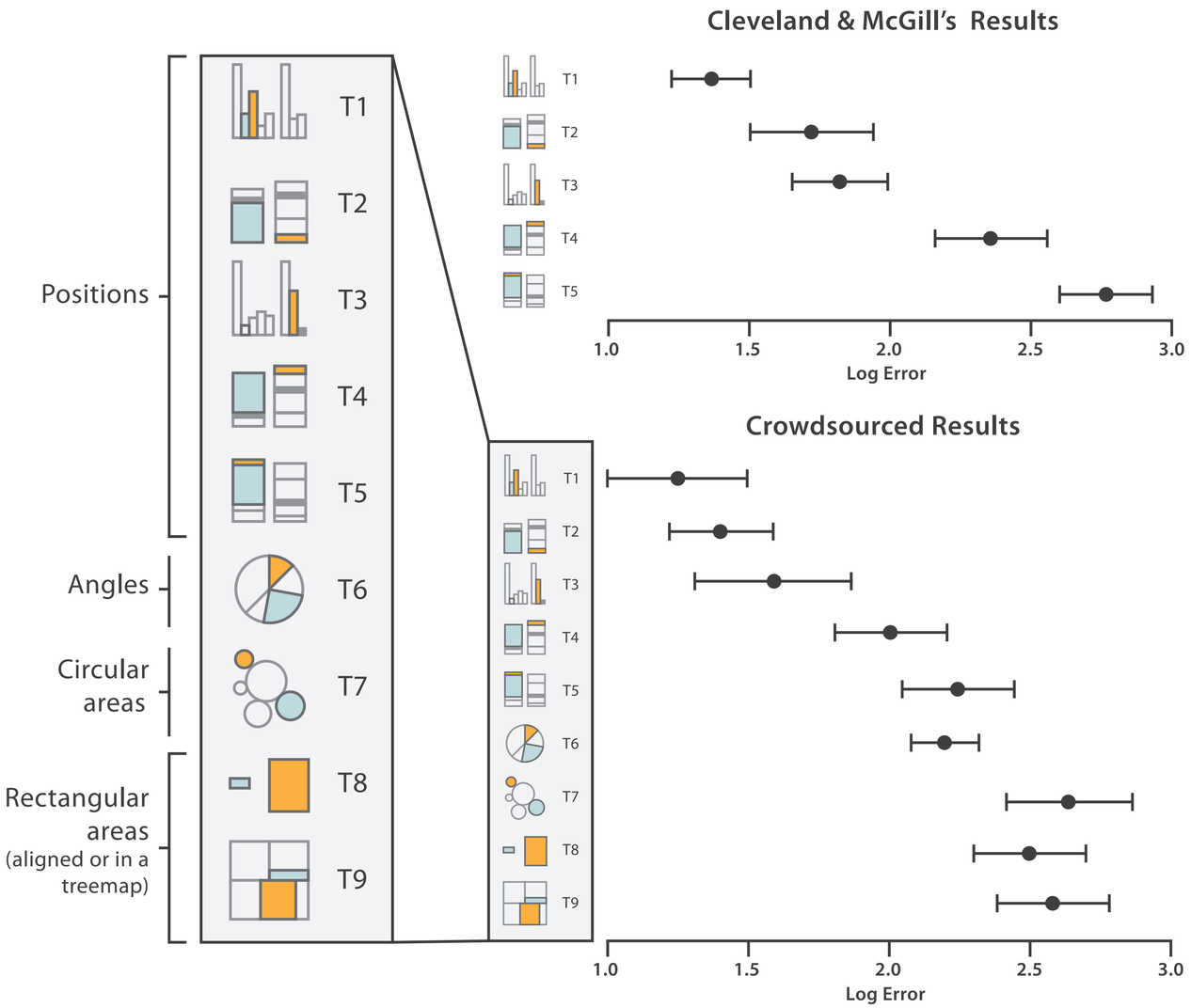
Tables & Color

Dr. David Koop

Expressiveness and Effectiveness

- Expressiveness Principle: all data from the dataset and nothing more should be shown
  - Do encode ordered data in an ordered fashion
  - Don’t encode categorical data in a way that implies an ordering
- Effectiveness Principle: the most important attributes should be the most **salient**
  - Saliency: how noticeable something is
  - How do the channels we have discussed measure up?

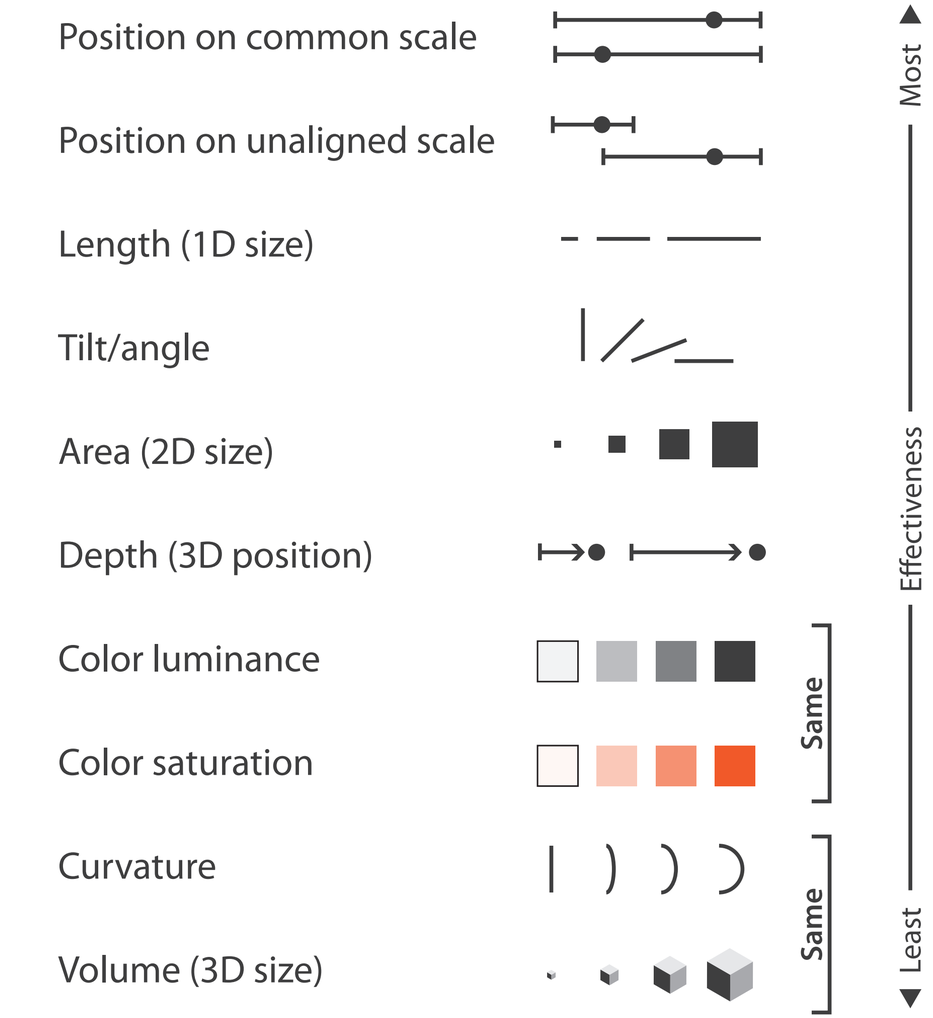
Perception Studies Summary



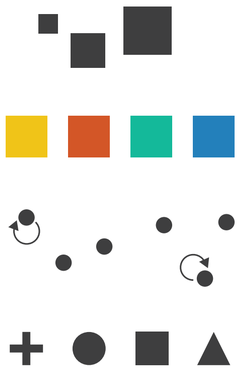
[Munzner (ill. Maguire) based on Heer & Bostock, 2014]

Ranking Channels by Effectiveness

Magnitude Channels: Ordered Attributes



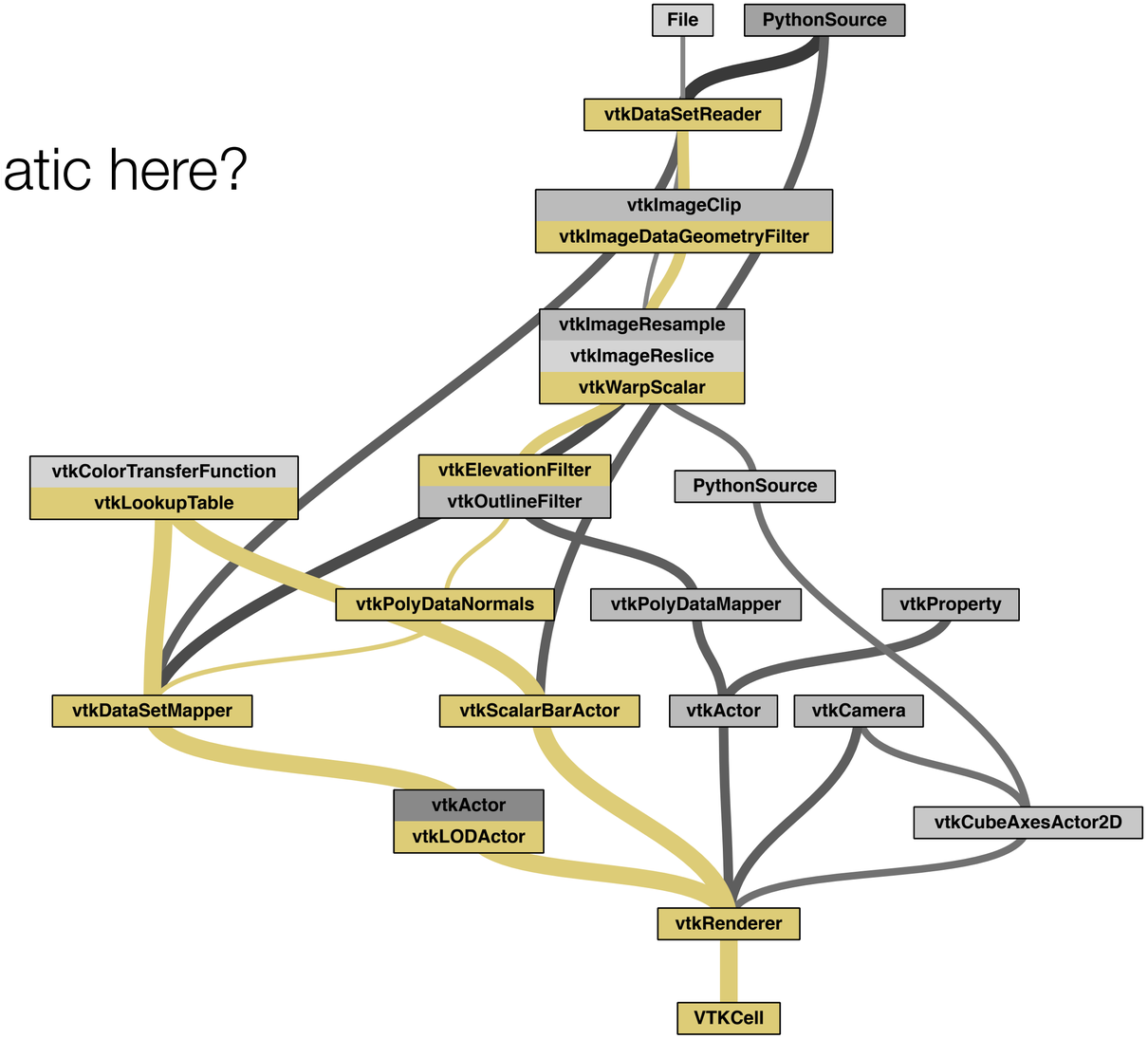
Identity Channels: Categorical Attributes



[Munzner (ill. Maguire), 2014]

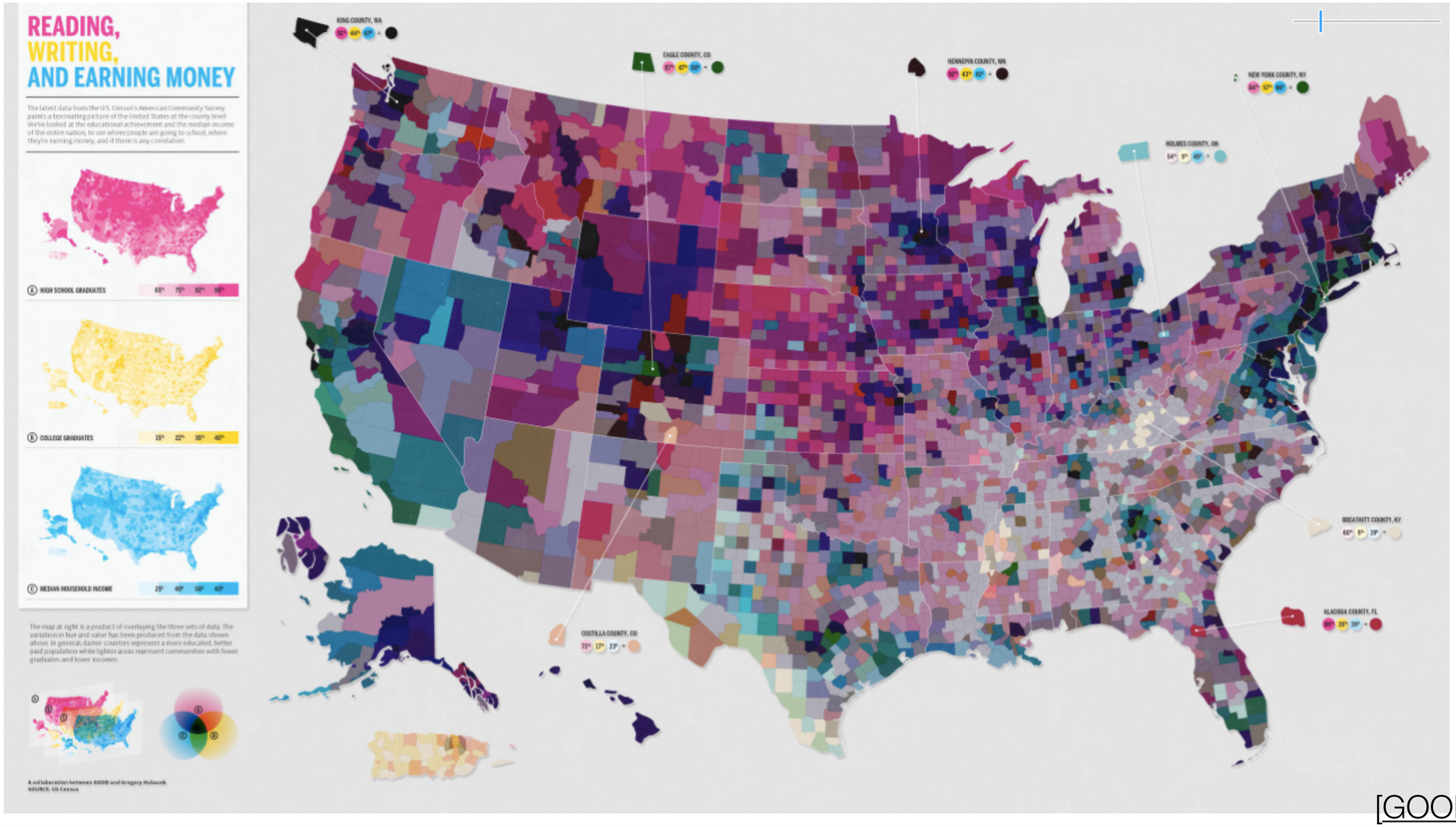
Discriminability

What is problematic here?



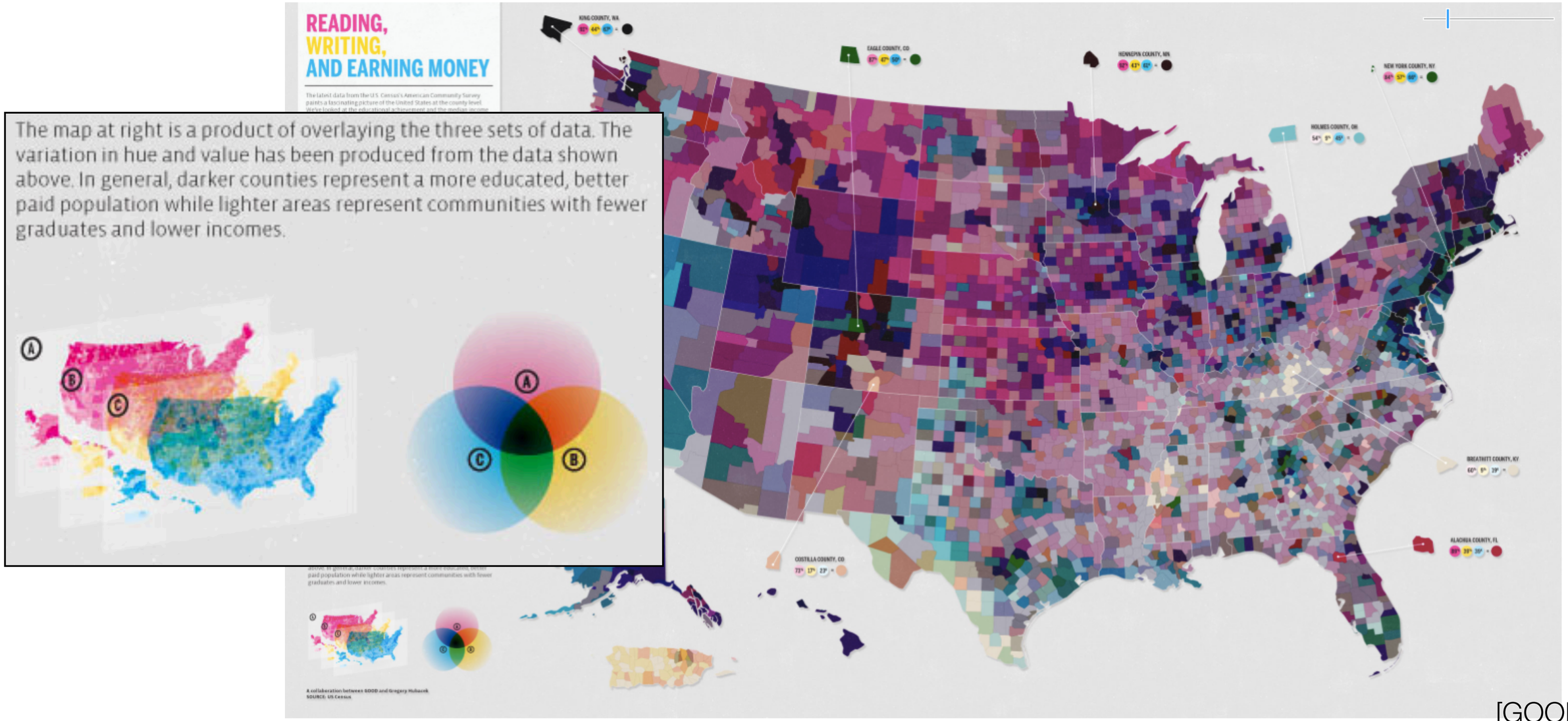
[Koop et al., 2013]

Separable or Integral?



[GOOD]

Separable or Integral?



[GOOD]

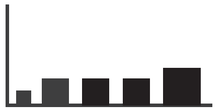
Arrange Tables

- **Express Values**

- **Separate, Order, Align Regions**
  - Separate

  - Order

  - Align



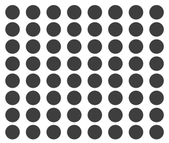
    - *1 Key*
      List

    - *2 Keys*
      Matrix

    - *3 Keys*
      Volume

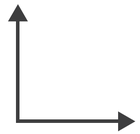
    - *Many Keys*
      Recursive Subdivision

- **Layout Density**
  - Dense



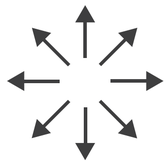
  - Space-Filling

- **Axis Orientation**
  - Rectilinear



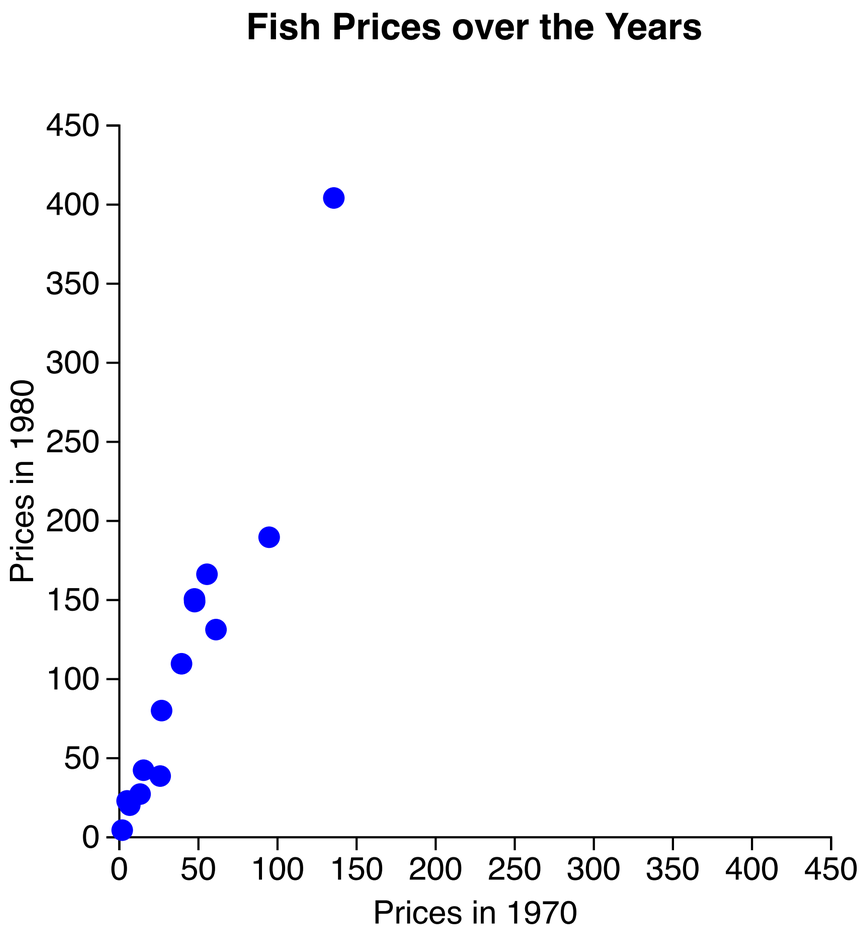
  - Parallel

  - Radial



[Munzner (ill. Maguire), 2014]

Express Values: Scatterplots

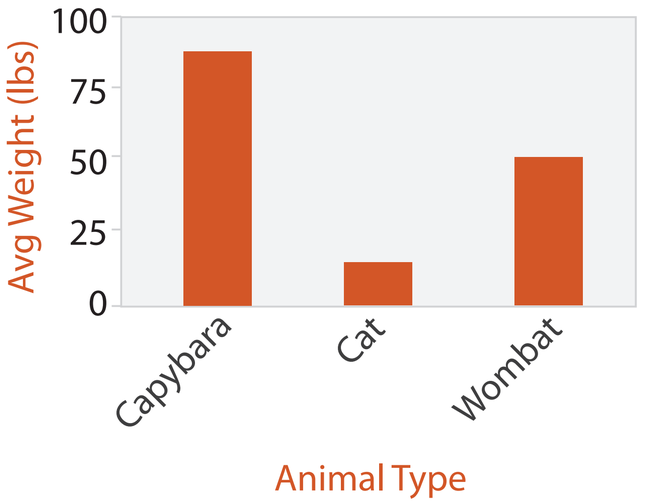


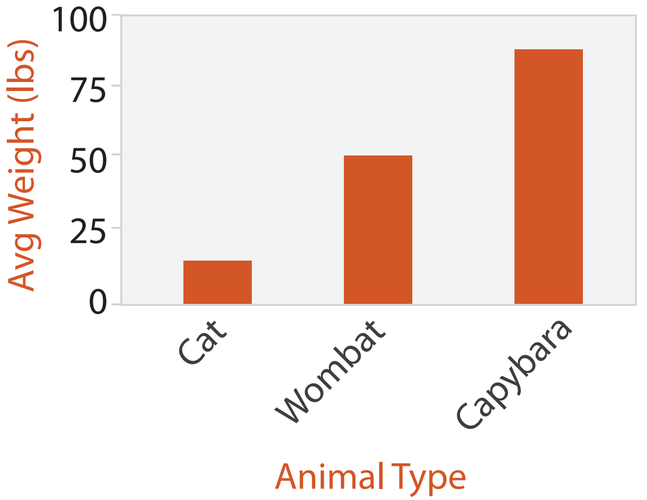
- Data: two quantitative values
- Task: find trends, clusters, outliers
- How: marks at spatial position in horizontal and vertical directions

- 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

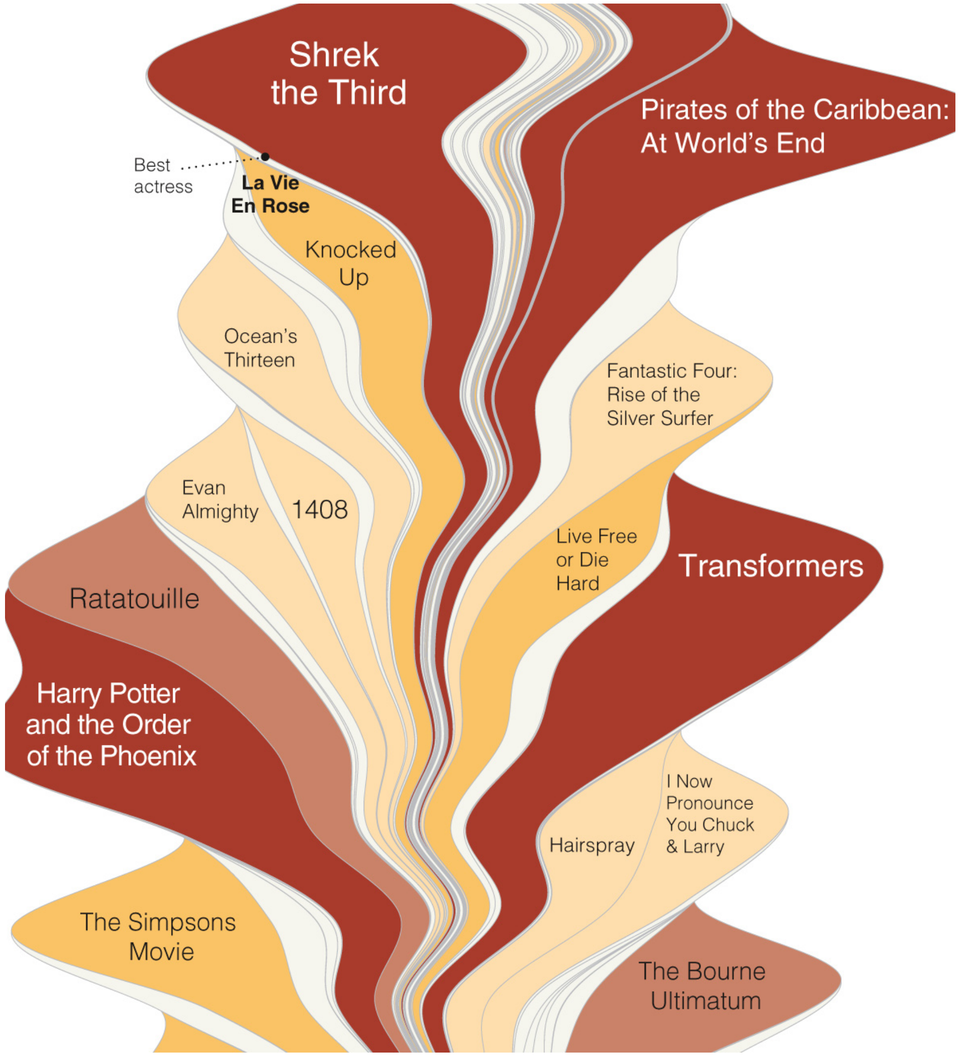




[Munzner (ill. Maguire), 2014]

Streamgraphs

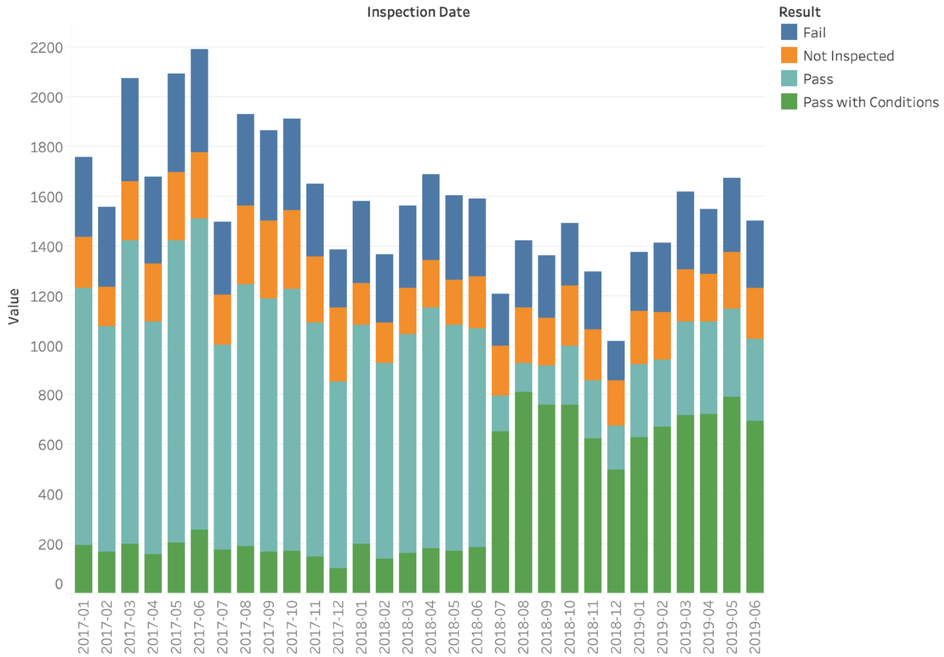
- Include a time attribute
- Data: multidimensional table, one quantitative attribute (count), one ordered key attribute (time), one categorical key attribute
- + derived attribute: layer ordering (quantitative)
- Task: analyze trends in time, find (maxmial) outliers
- How: derived position+geometry,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



[Byron and Wattenberg, 2012]

Assignment 3

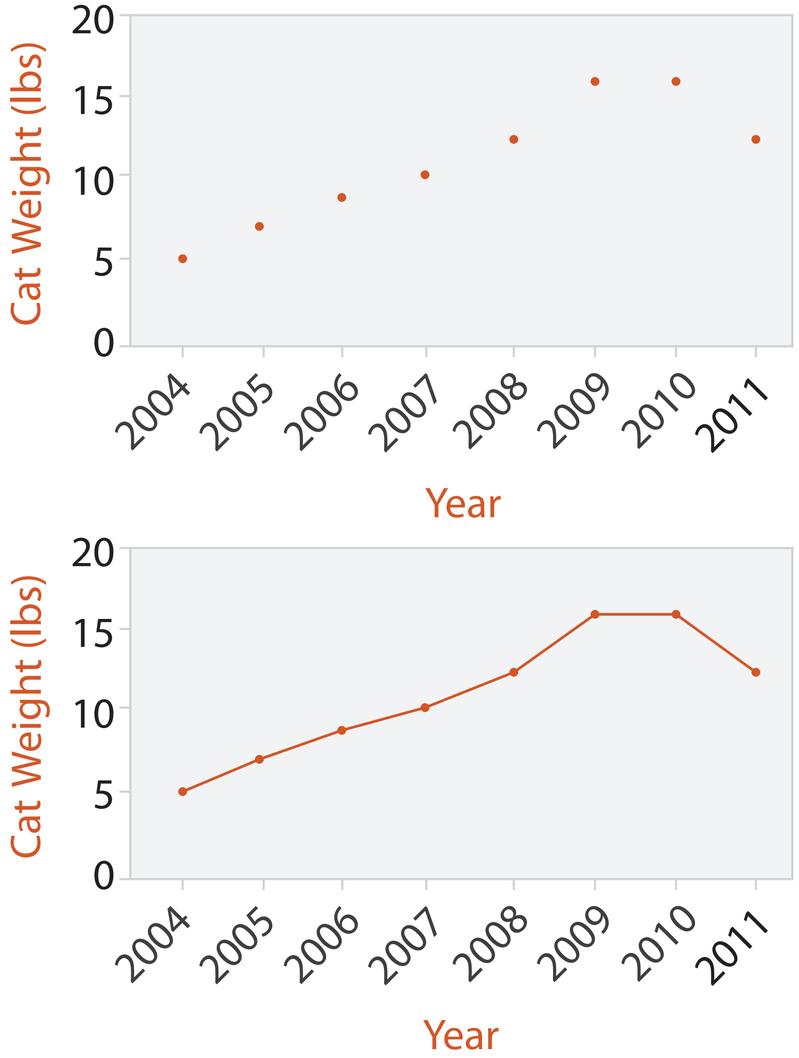
- Same stacked bar chart visualization
- Three tools
  - Tableau (free academic license)
  - Vega-Lite
  - D3
- For Vega-Lite, use the online editor
- For D3, use the template files so the data is properly loaded
- [CS 490] Only need to do a standard bar chart in D3



Project Proposal

- Find an interesting subject or dataset
  - see List of lists of datasets [B. Keegan]
- Understand the data available (format, types, semantics)
- Figure out some interesting questions and tasks
- Start brainstorming about visualizations and interactions
- Inspiration:
  - Information Is Beautiful Awards
  - MBTA Viz
- Due Friday, October 11, 2019

Dot and Line Charts



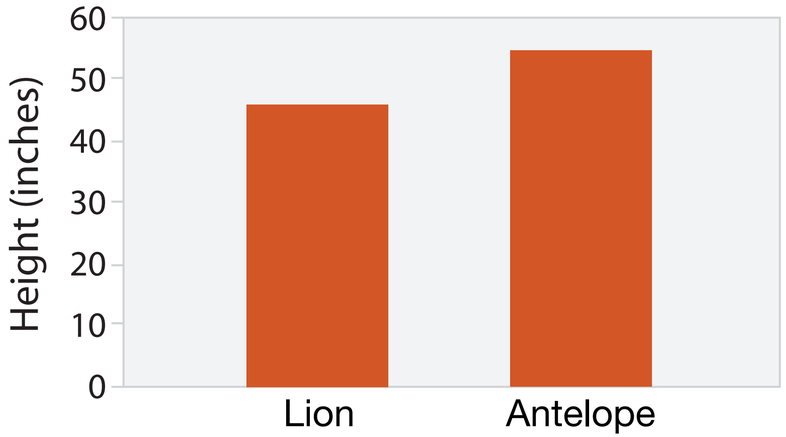
- Data: one quantitative attribute, one **ordered** attribute
- Task: lookup values, find outliers and trends
- How: point mark and positions

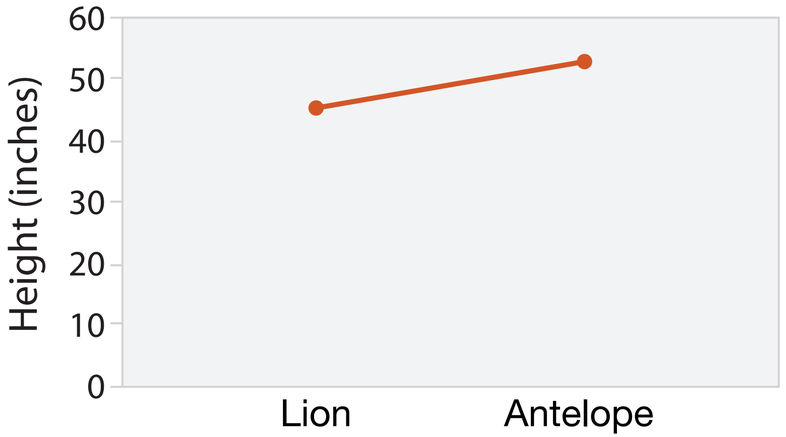
- Line Charts: add **connection mark** (line)

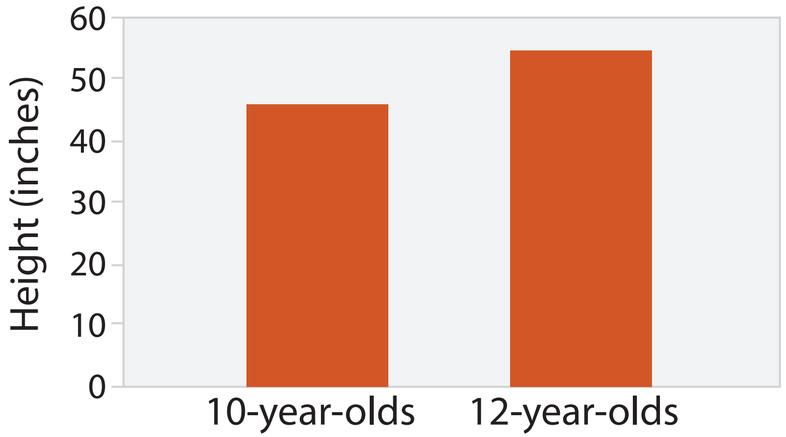
- Similar to scatterplots but allow ordered attribute

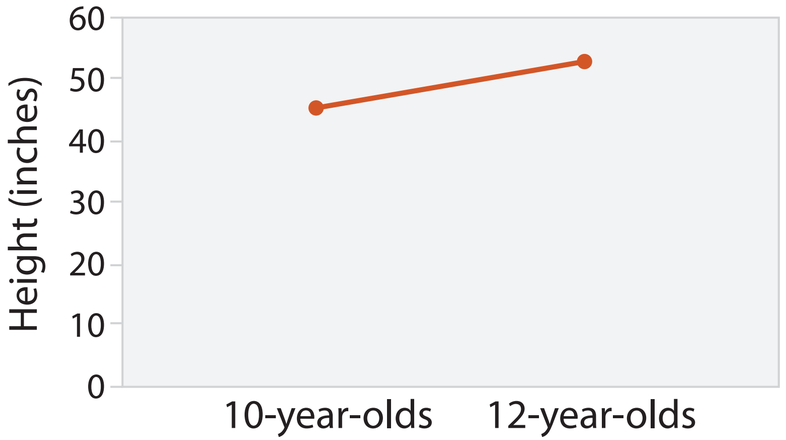
[Munzner (ill. Maguire), 2014]

Proper Use of Line and Bar Charts



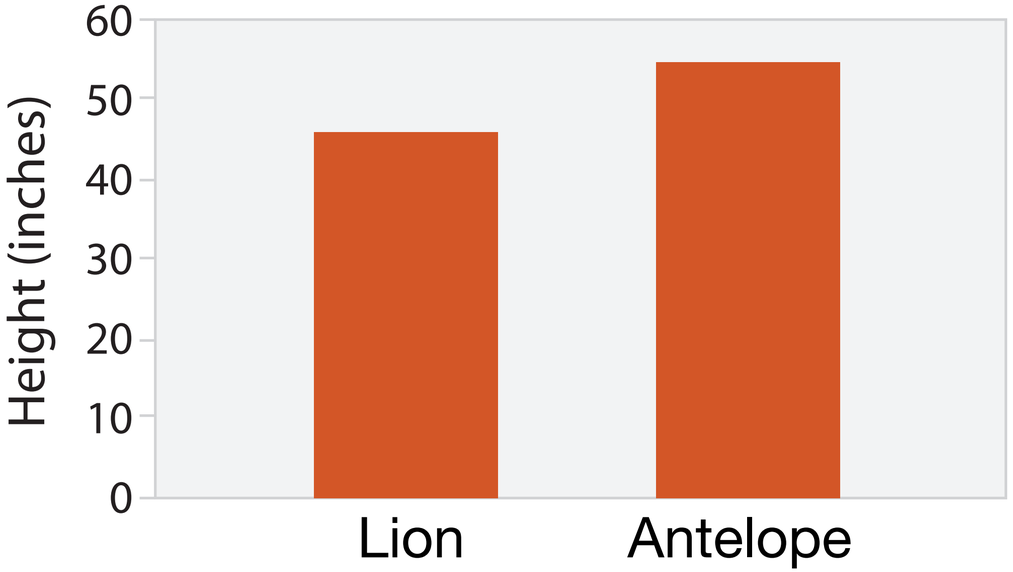


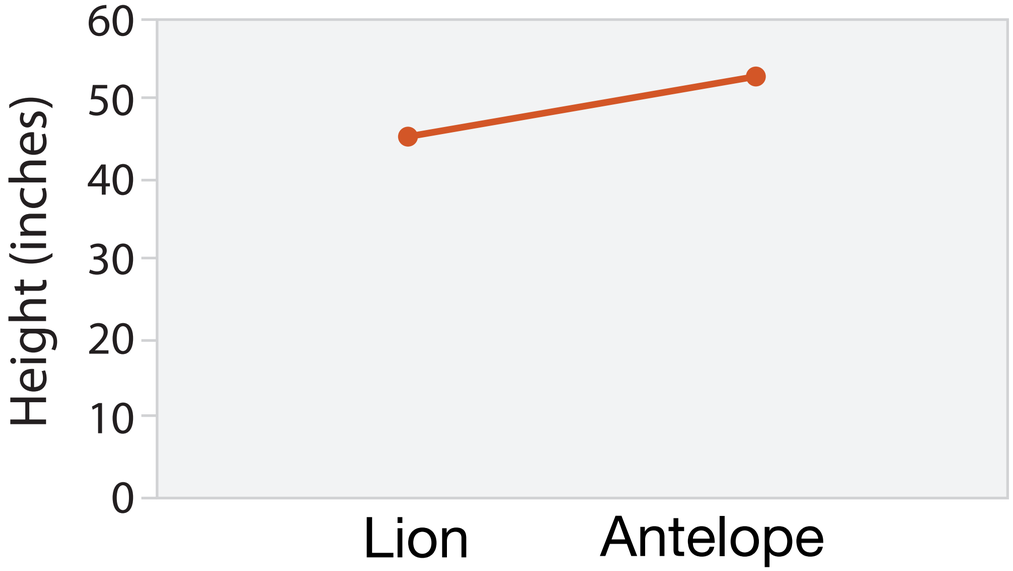




[Adapted from Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]

Proper Use of Line and Bar Charts





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

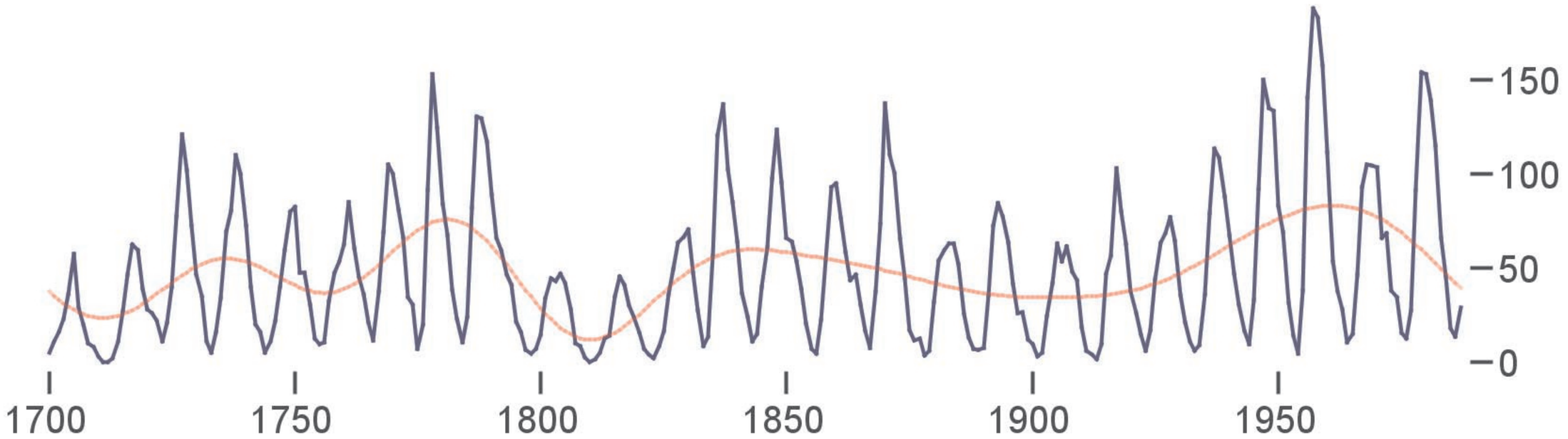
[Adapted from Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]

Aspect Ratio

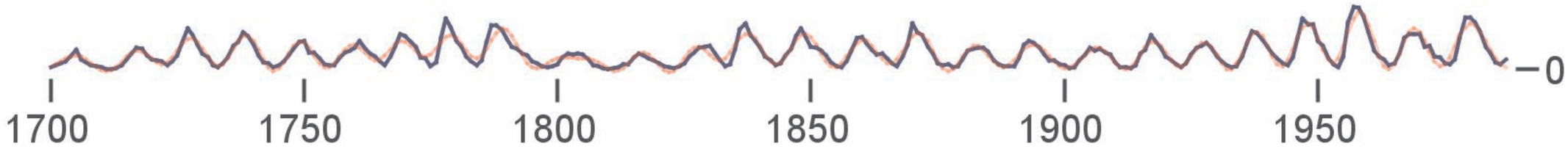
- Trends in line charts are more apparent because we are using angle as a channel
- Perception of angle (and the **relative difference** between angles) is important
- Initial experiments found people best judge differences in **slope** when angles are around 45 degrees (Cleveland et al., 1988, 1993)

Multiscale Banking

Aspect Ratio = 3.96



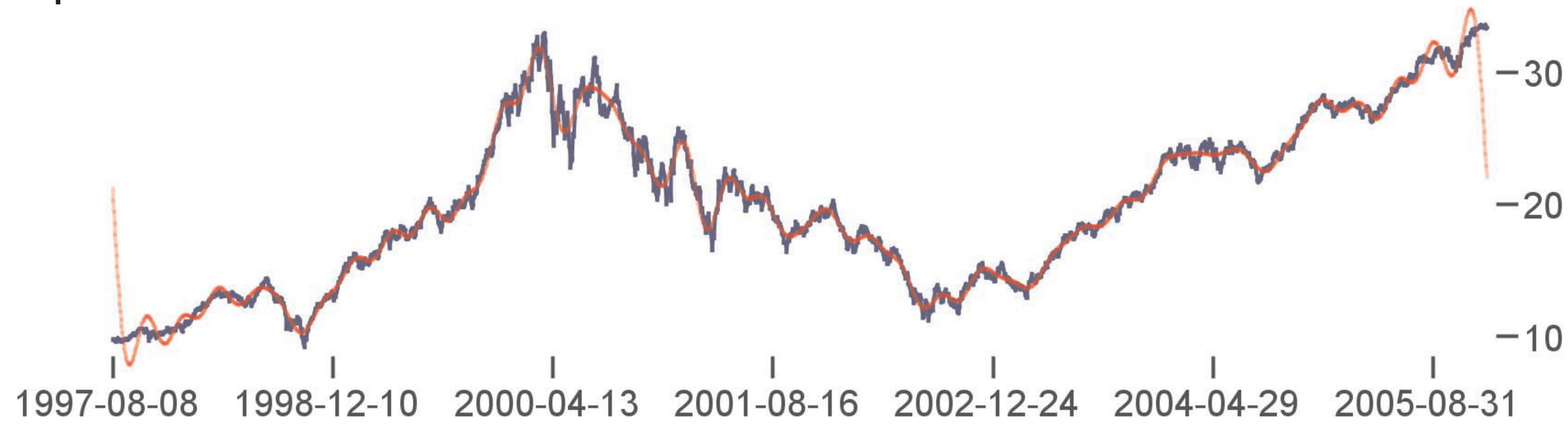
Aspect Ratio = 22.35



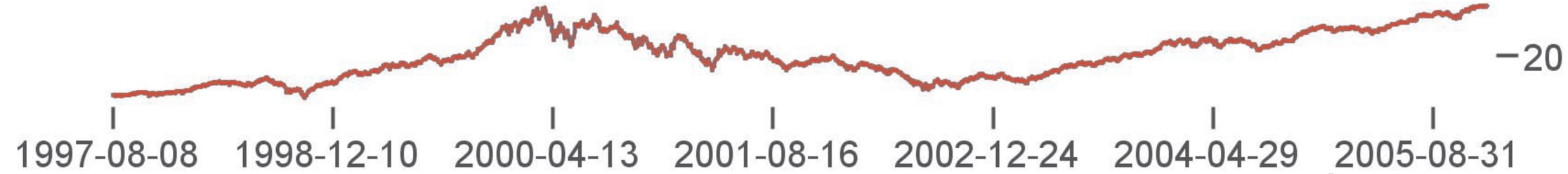
[Heer and Agrawala, 2006]

Multiscale Banking

Aspect Ratio = 4.23



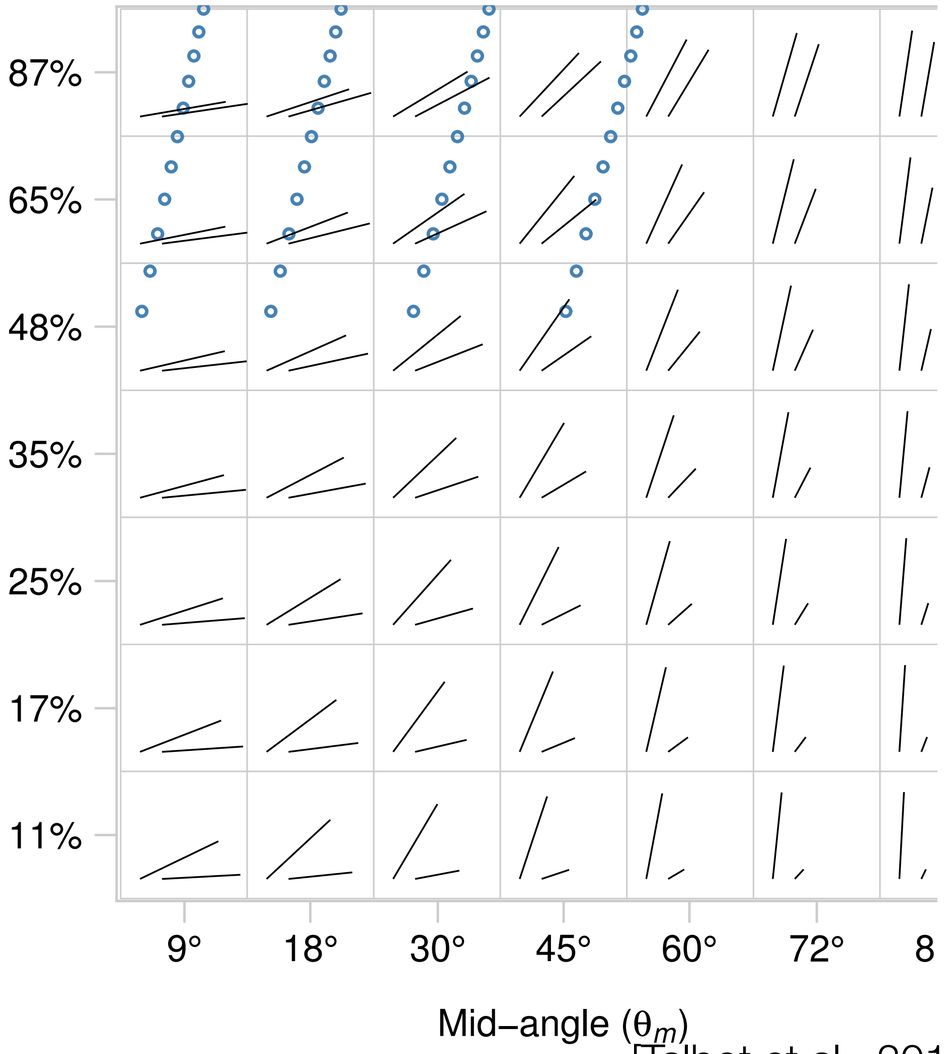
Aspect Ratio = 14.55



[Heer and Agrawala, 2006]

Expanding the Study

- Cleveland et al. did not study the entire space of slope comparisons and 45 degrees was at the low end of their study (blue marks on right)
- Talbot et al. compared more slopes and found that people do better with smaller slopes
- Baselines may aid with this

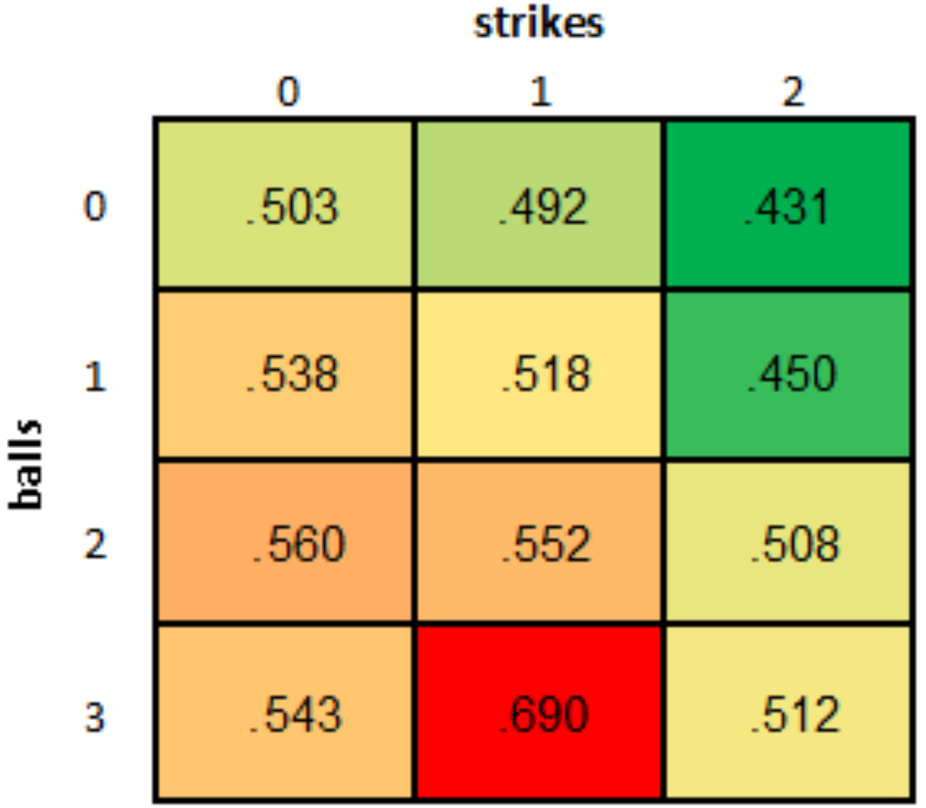


[Talbot et al., 2013]

Heatmaps

- Data: Two keys, one quantitative attribute
- Task: Find clusters, outliers, summarize
- How: area marks in grid, color encoding of quantitative attribute
- Scalability: number of pixels for area marks (millions), <12 colors
- Red-green color scales often used
  - Be aware of colorblindness!

Fast-Pitch Softball
Slugging Percentage



[fastpitchanalytics.com]

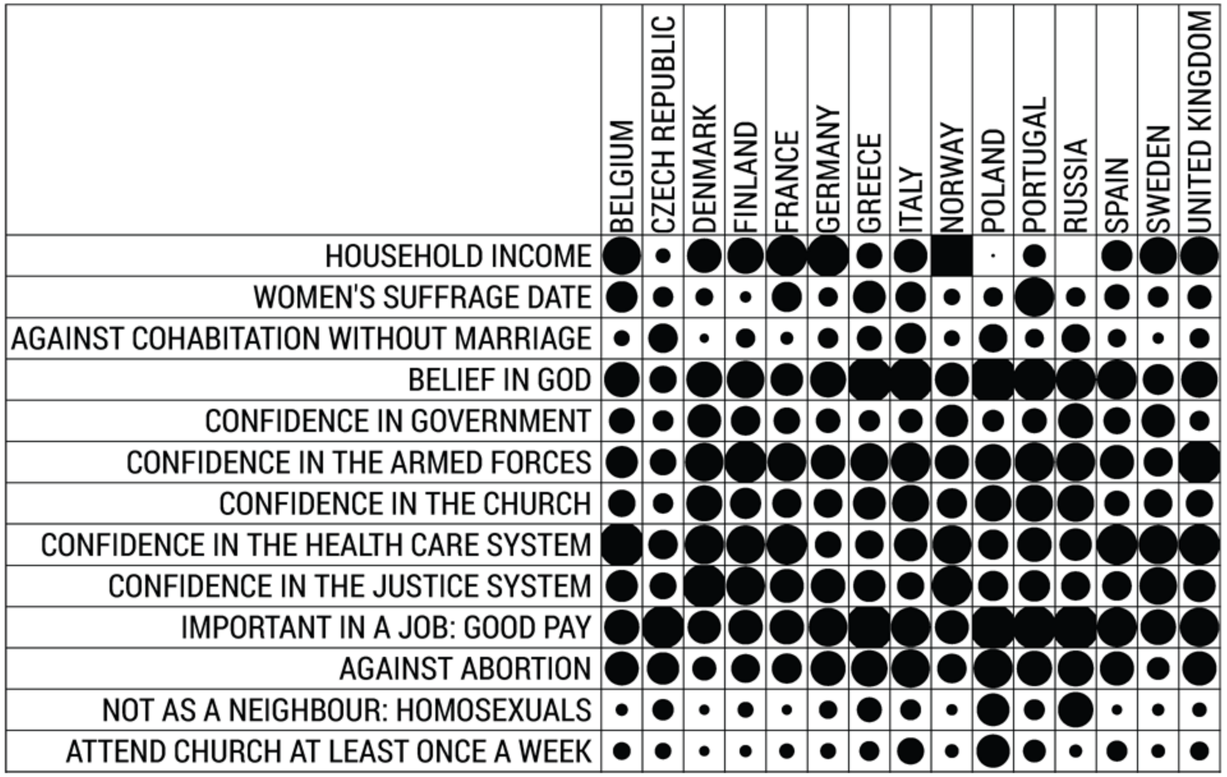
Bertin Matrices

- Must we only use color?
  - What other marks might be appropriate?

[C.Perrin et al., 2014]

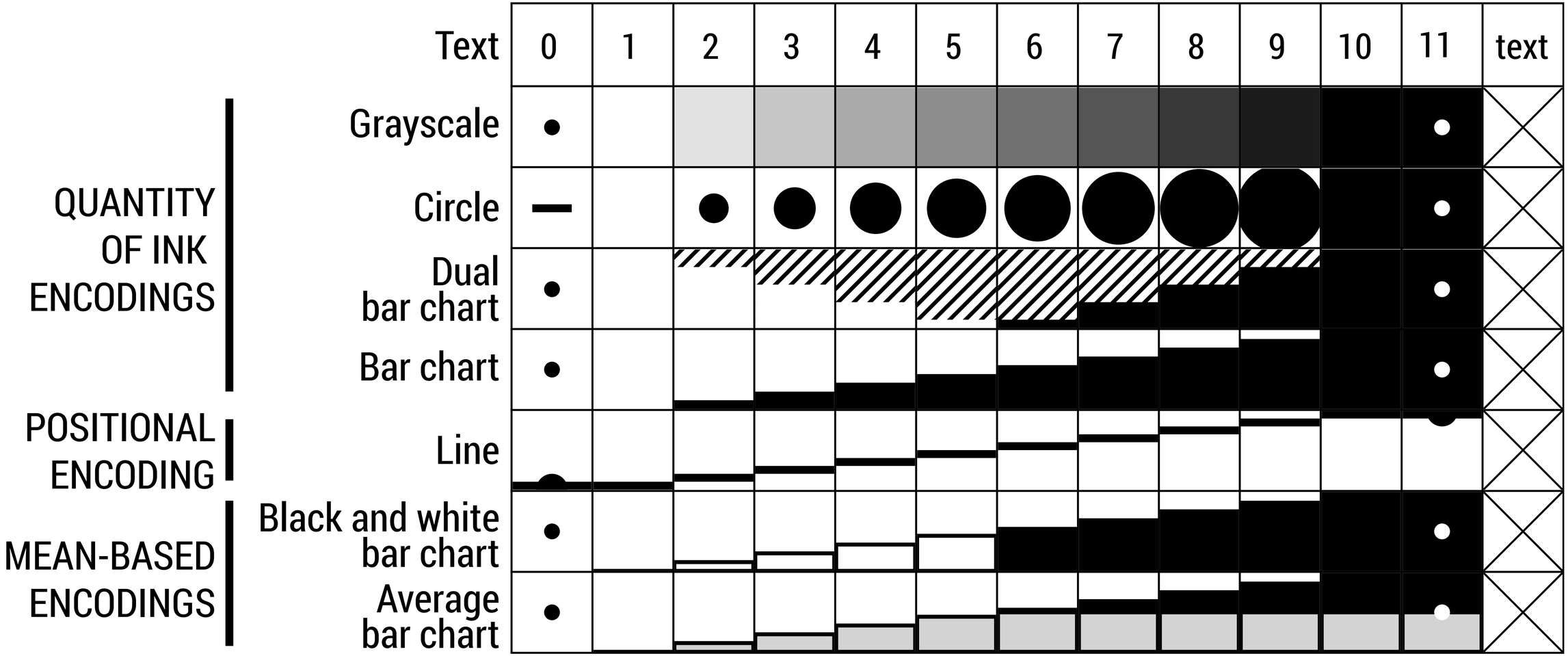
Bertin Matrices

- Must we only use color?
  - What other marks might be appropriate?



[C.Perrin et al., 2014]

Bertin's Encodings

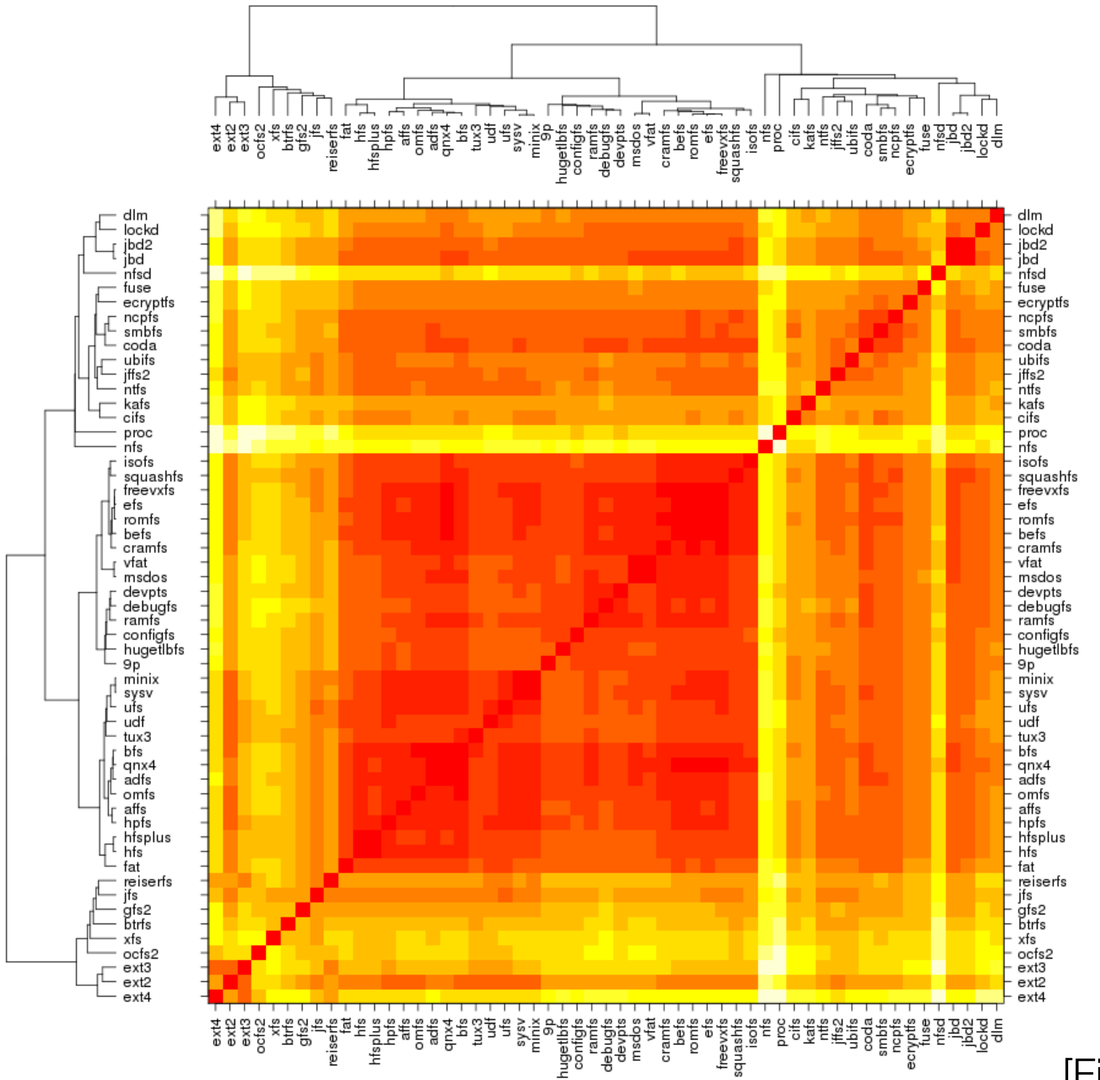


[C.Perrin et al., 2014]

Matrix Reordering

[Bertin Exhibit (INRIA, Vis 2014), Photo by Robert Kosara]

Cluster Heatmap



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

Cluster Heatmap

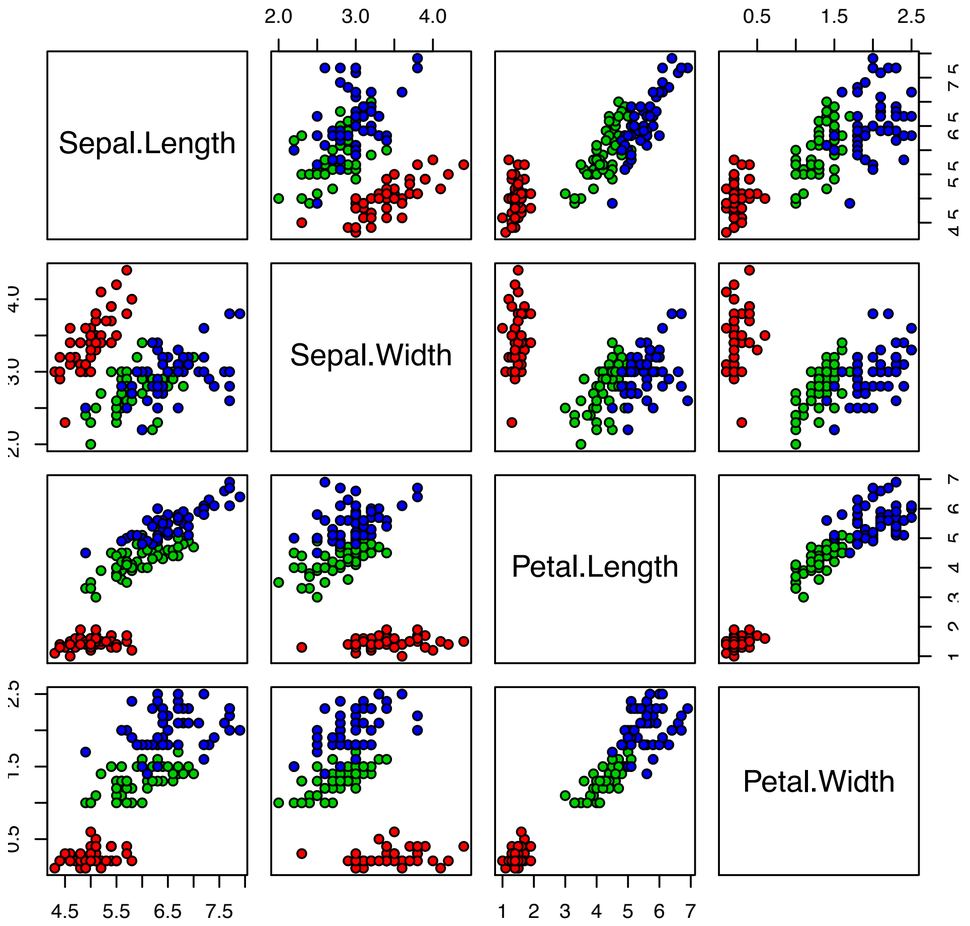
- Data & Task: Same as Heatmap
- How: Area marks but matrix is ordered by cluster hierarchies
- Scalability: limited by the cluster dendrogram

- Dendrogram: a visual encoding of tree data with leaves aligned

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]

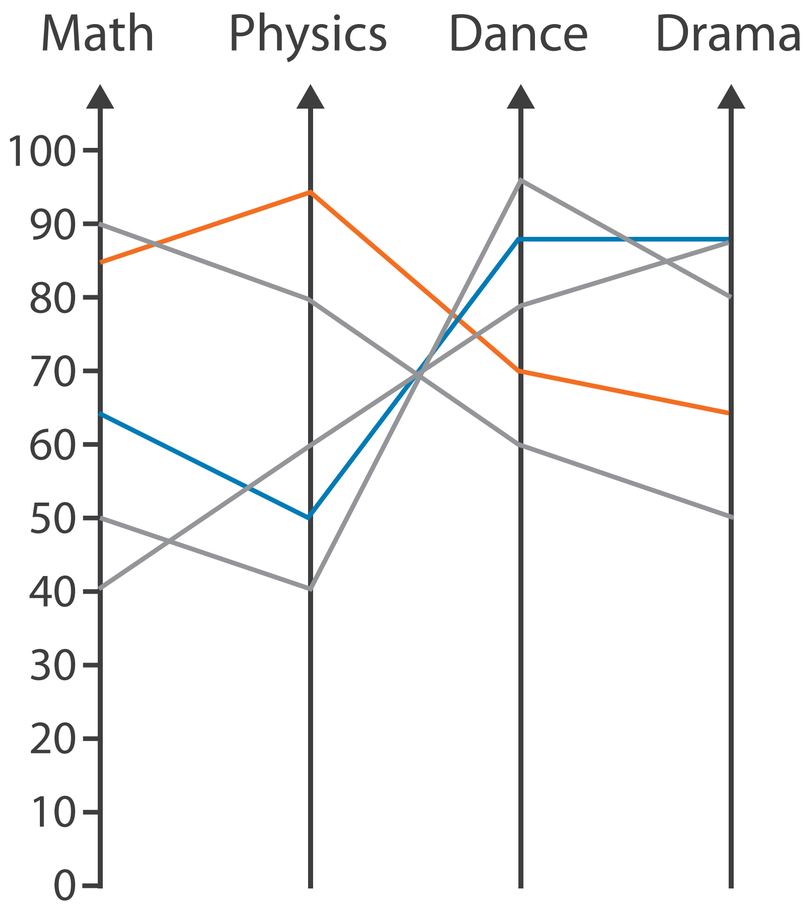
Spatial Axis Orientation

- 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

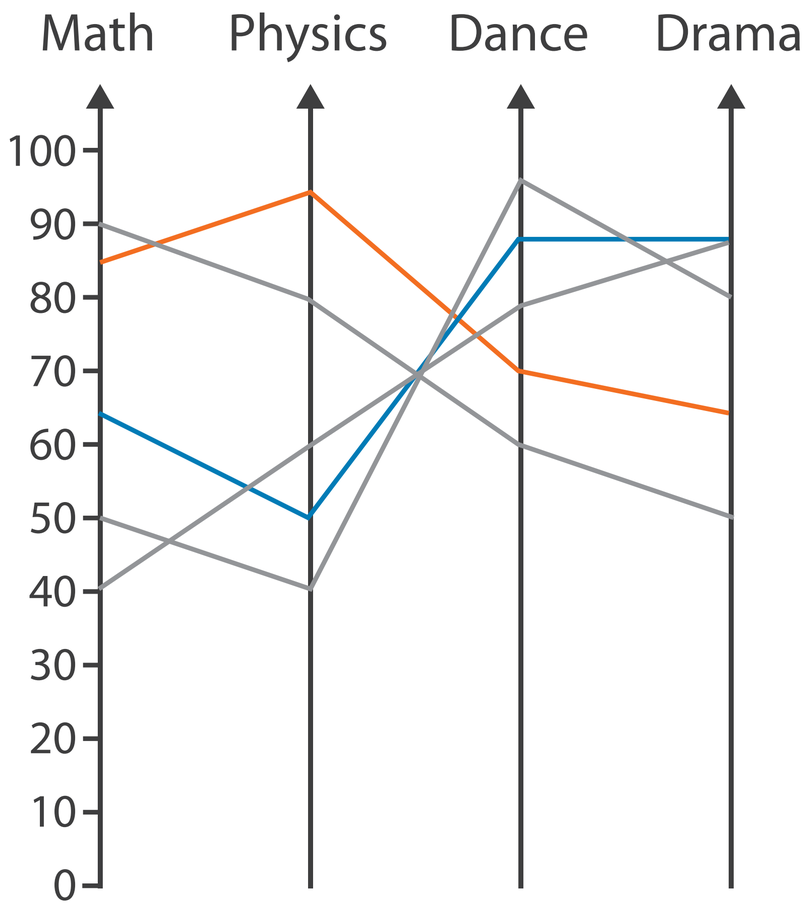
- So far, we have seen the vertical and horizontal axes (a **rectilinear** layout) used to encode almost everything
- What other possibilities are there for axes?
  - Parallel axes



[Munzner (ill. Maguire), 2014]

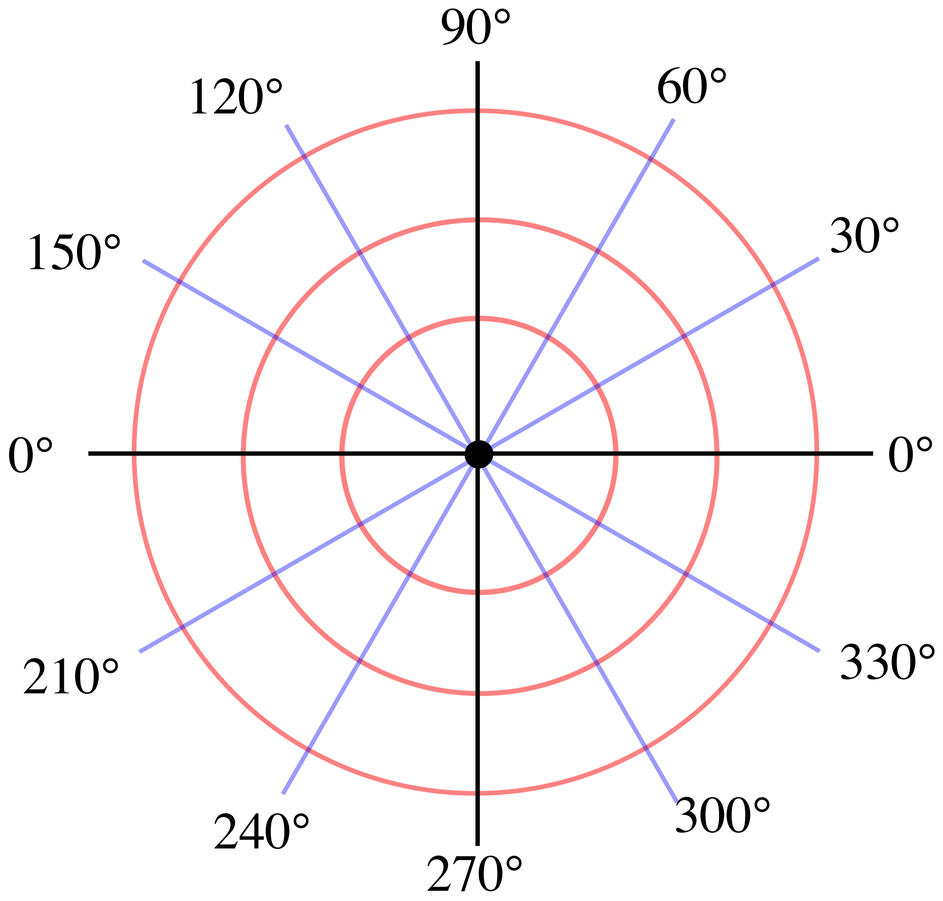
Spatial Axis Orientation

- 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

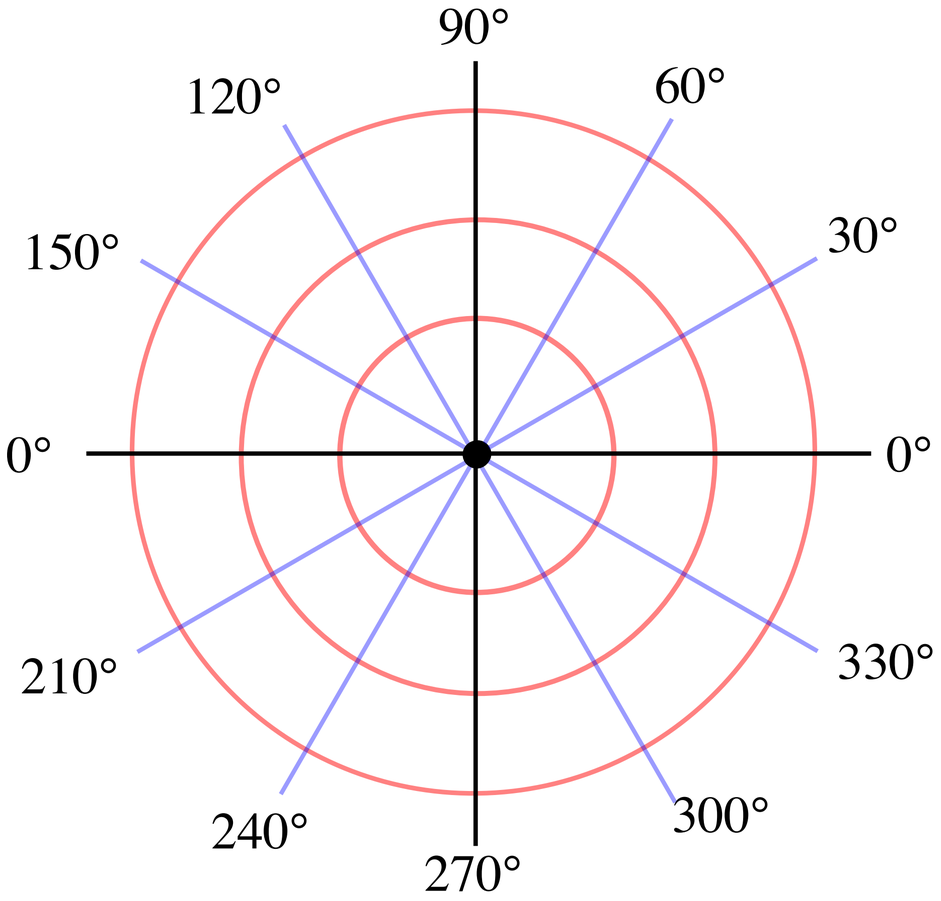


[Munzner (ill. Maguire), 2014]

Radial Axes

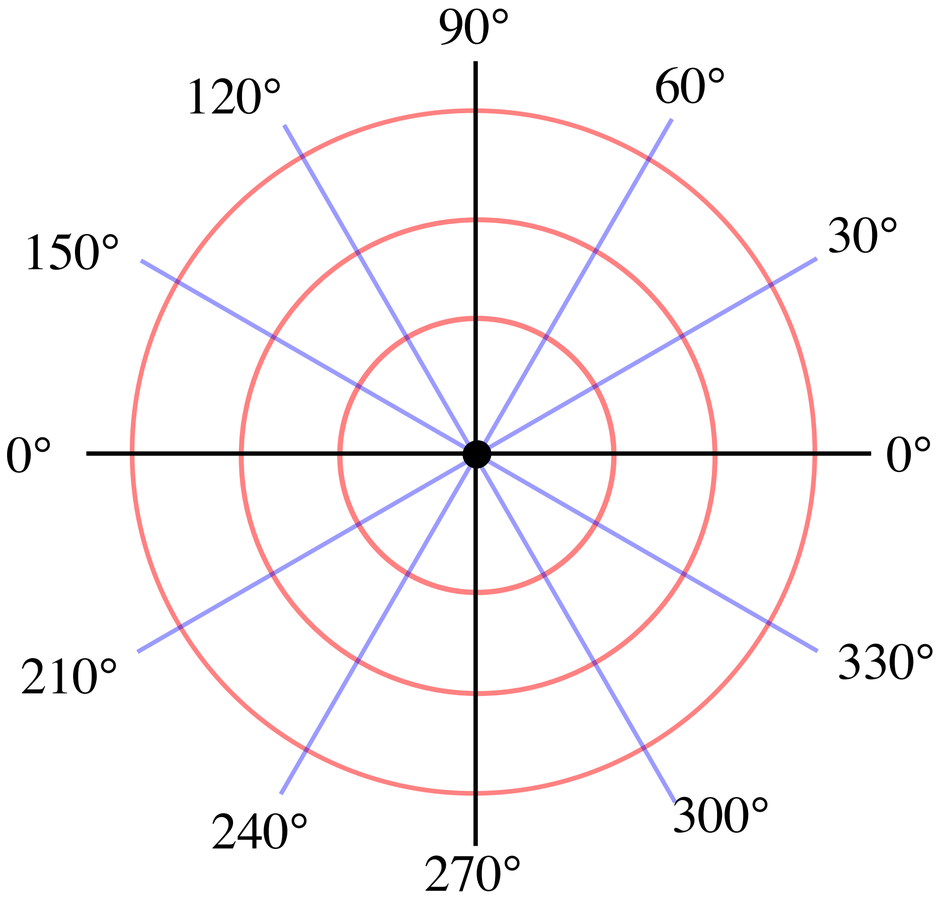


Radial Axes



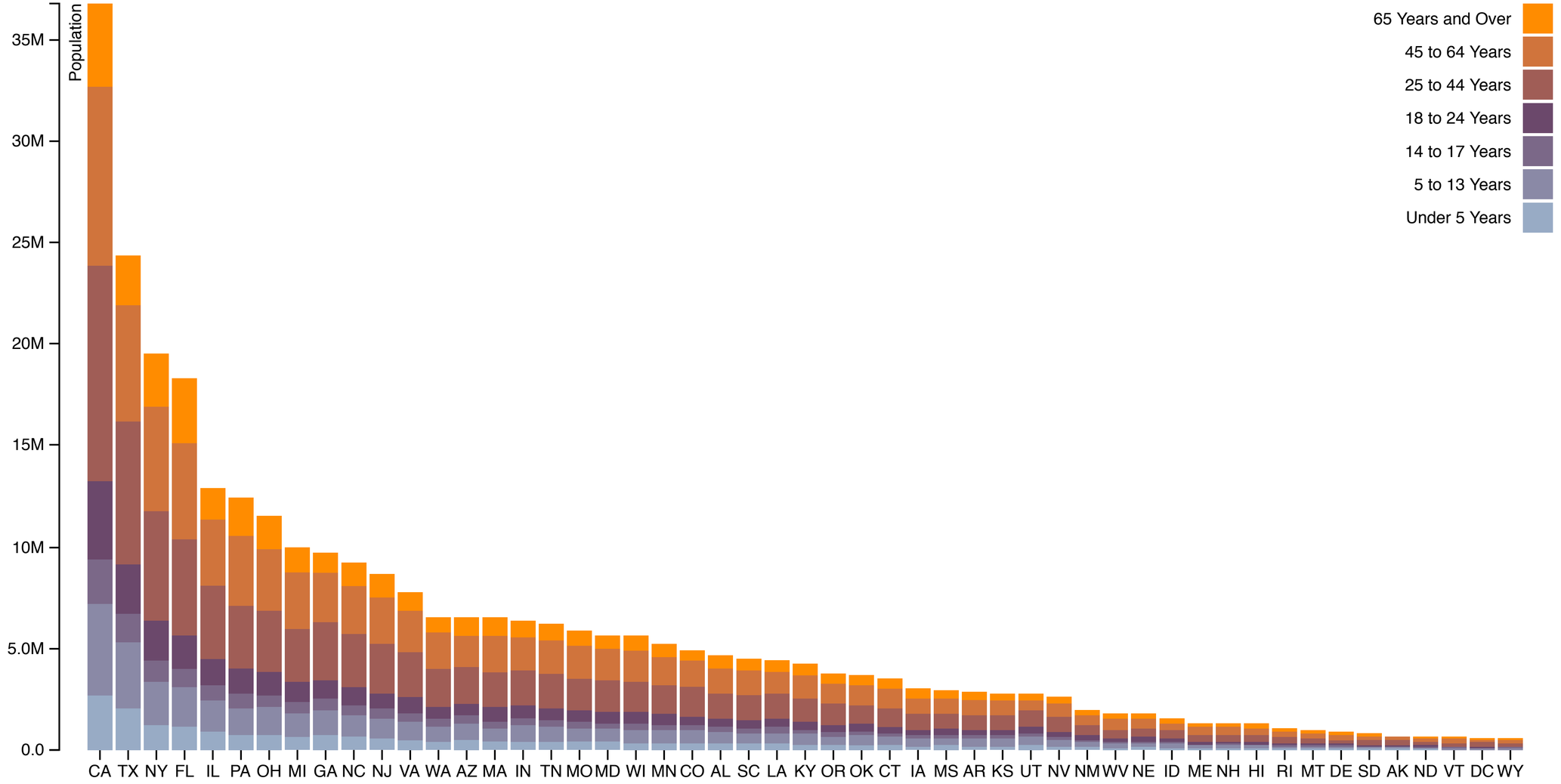
- Polar Coordinates (angle + position along the line at that angle)
- What types of encodings are possible for tabular data in polar coordinates?

Radial Axes



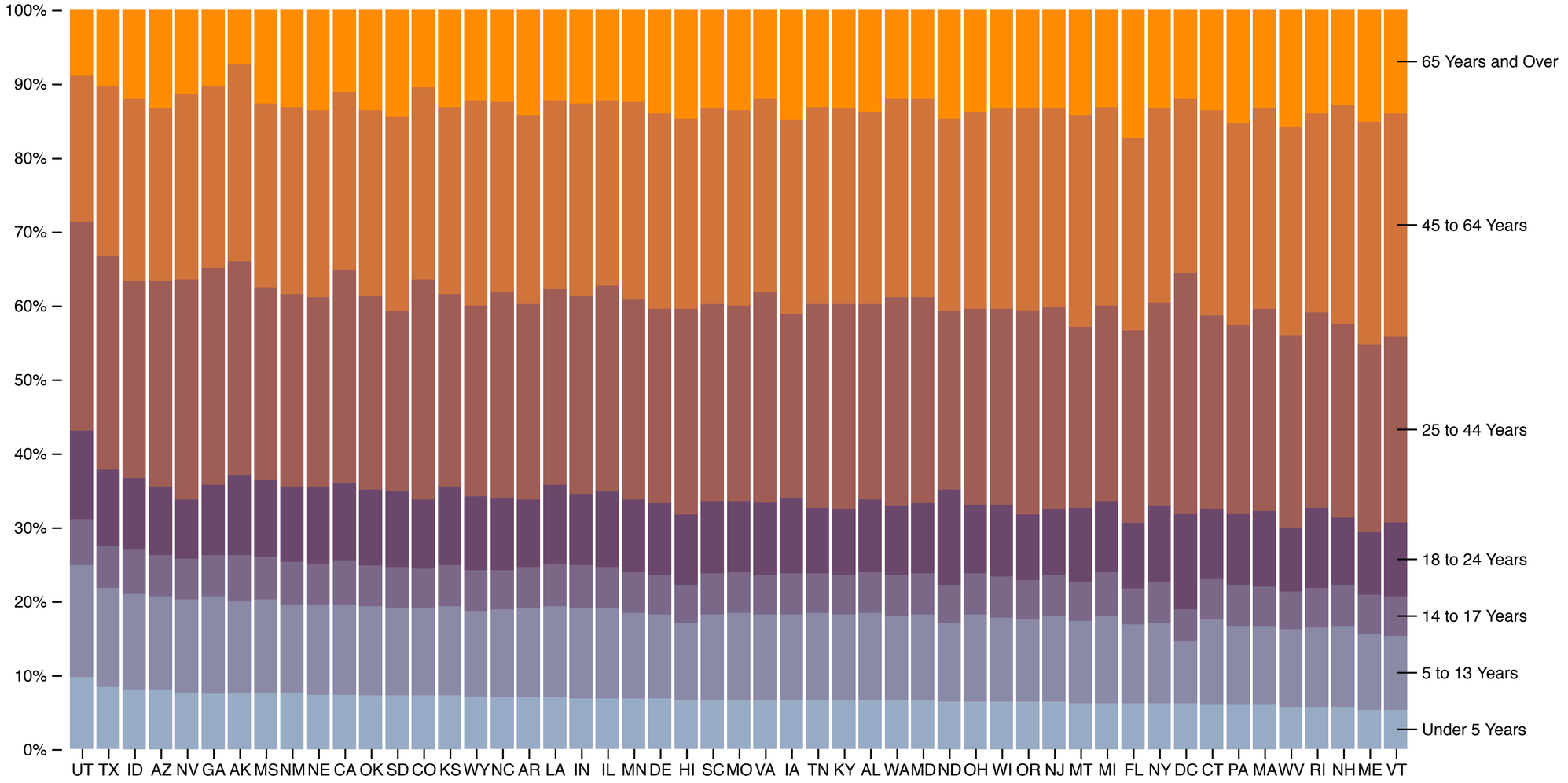
- Polar Coordinates (angle + position along the line at that angle)
- What types of encodings are possible for tabular data in polar coordinates?
  - Radial bar charts
  - Pie charts
  - Donut charts

Part-of-whole: Relative % comparison?



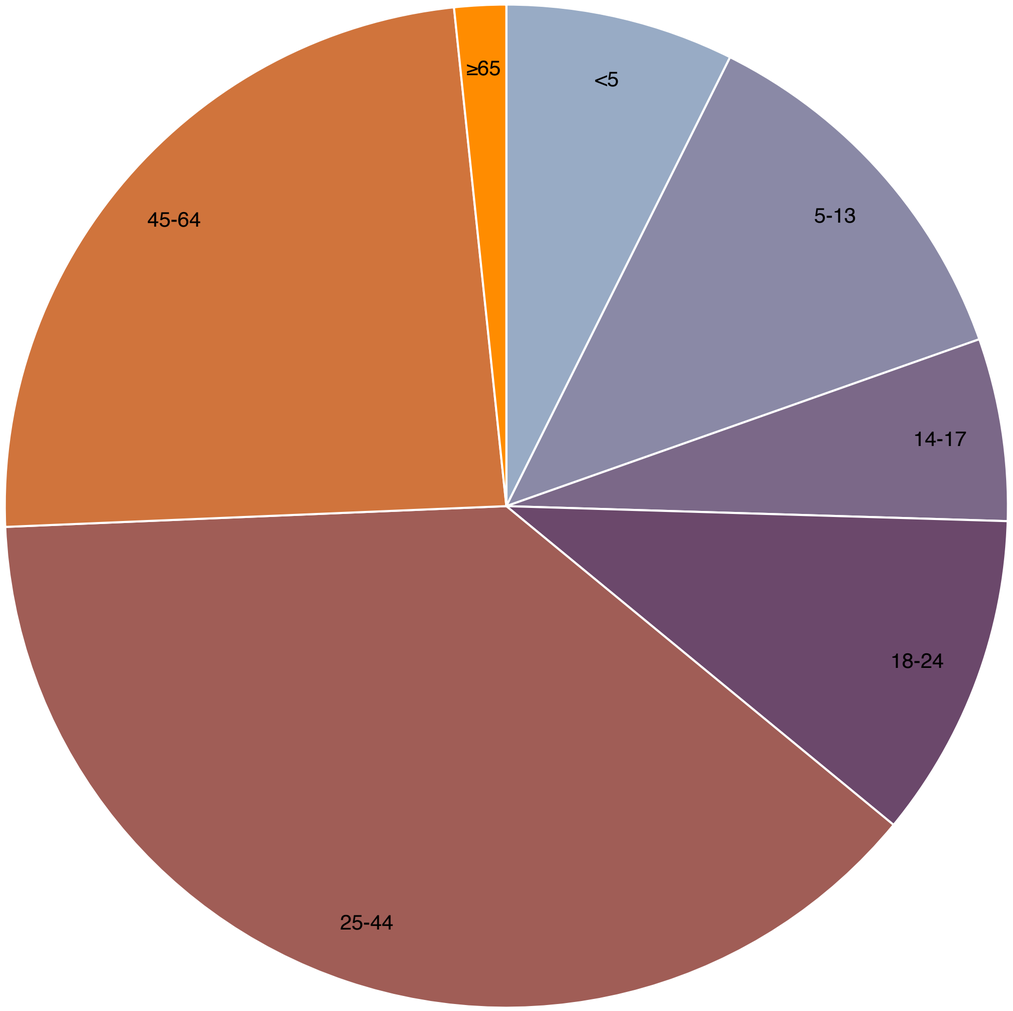
[Stacked Bar Chart, M. Bostock, 2017]

Normalized Stacked Bar Chart



[Normalized Stacked Bar Chart, Bostock, 2017]

Pie Chart

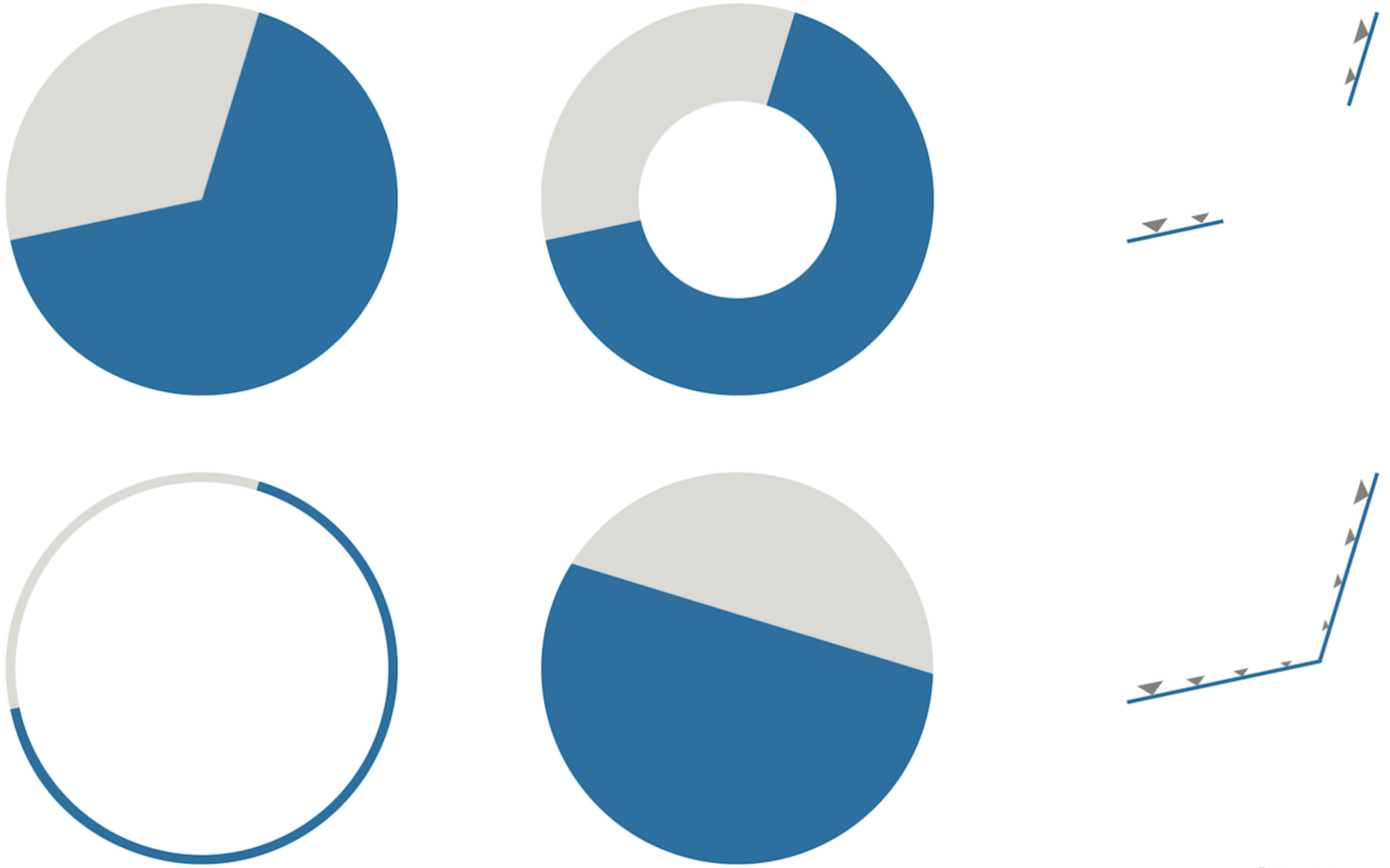


[Pie Chart, Bostock, 2017]

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"

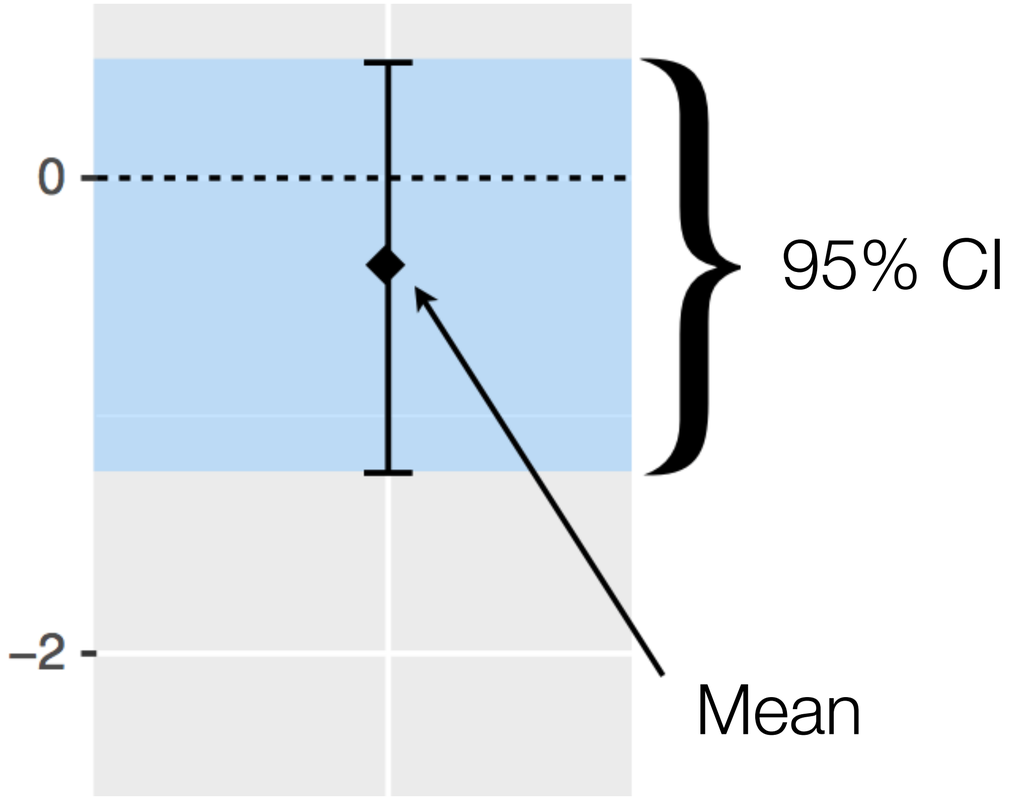
Arcs, Angles, or Areas?



[R. Kosara and D. Skau, 2016]

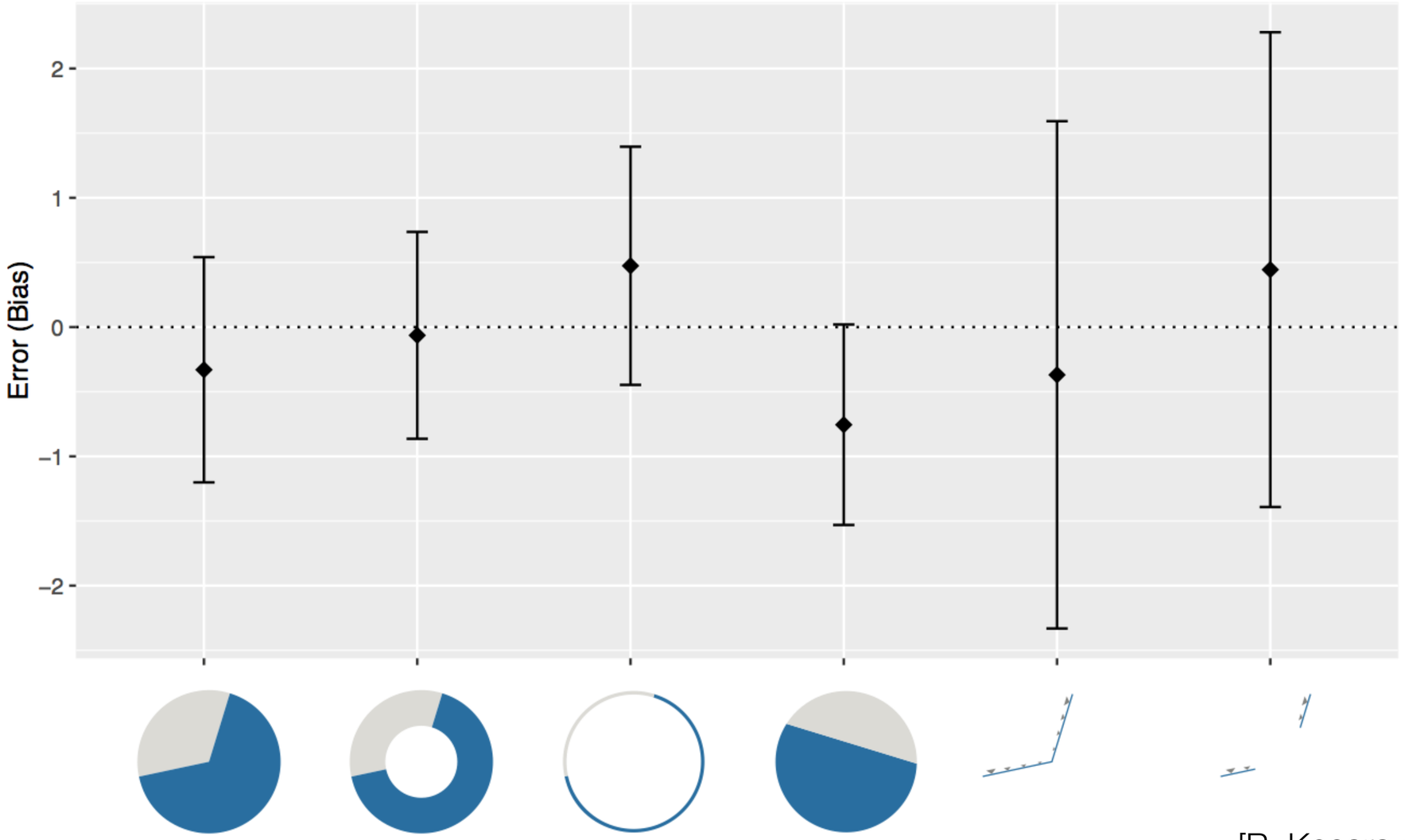
Study Setup

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



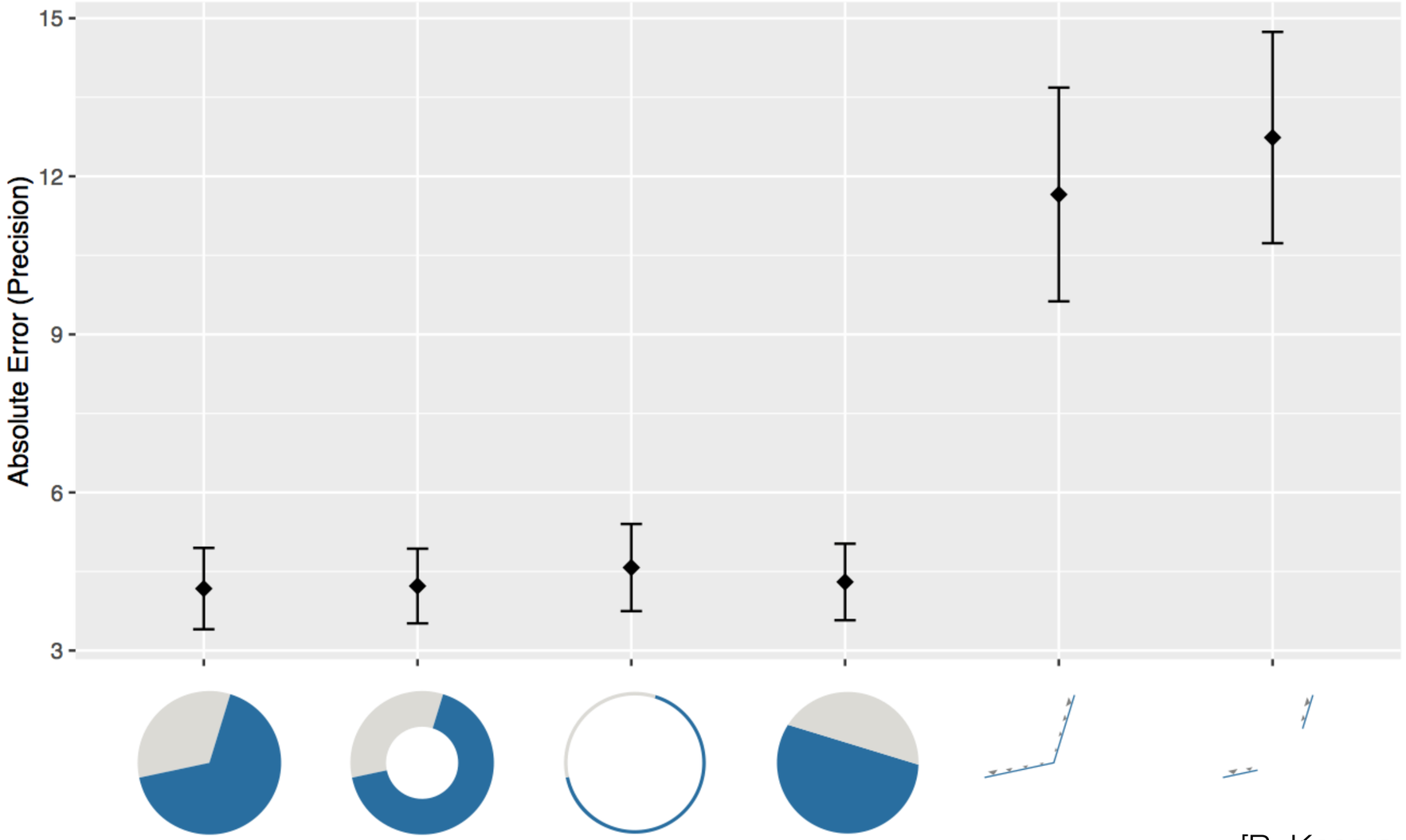
[R. Kosara and D. Skau, 2016]

Signed Error



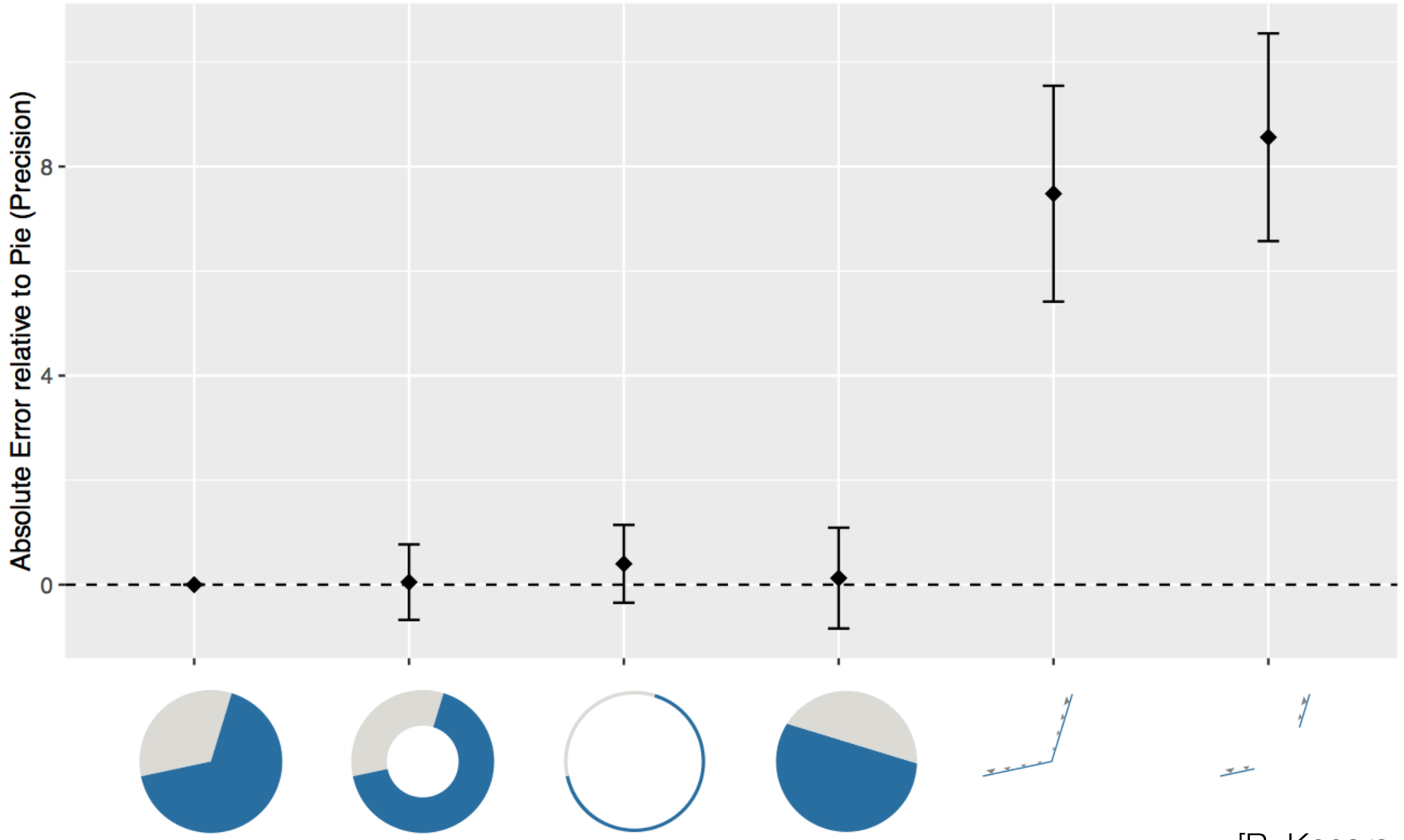
[R. Kosara and D. Skau, 2016]

Absolute Error



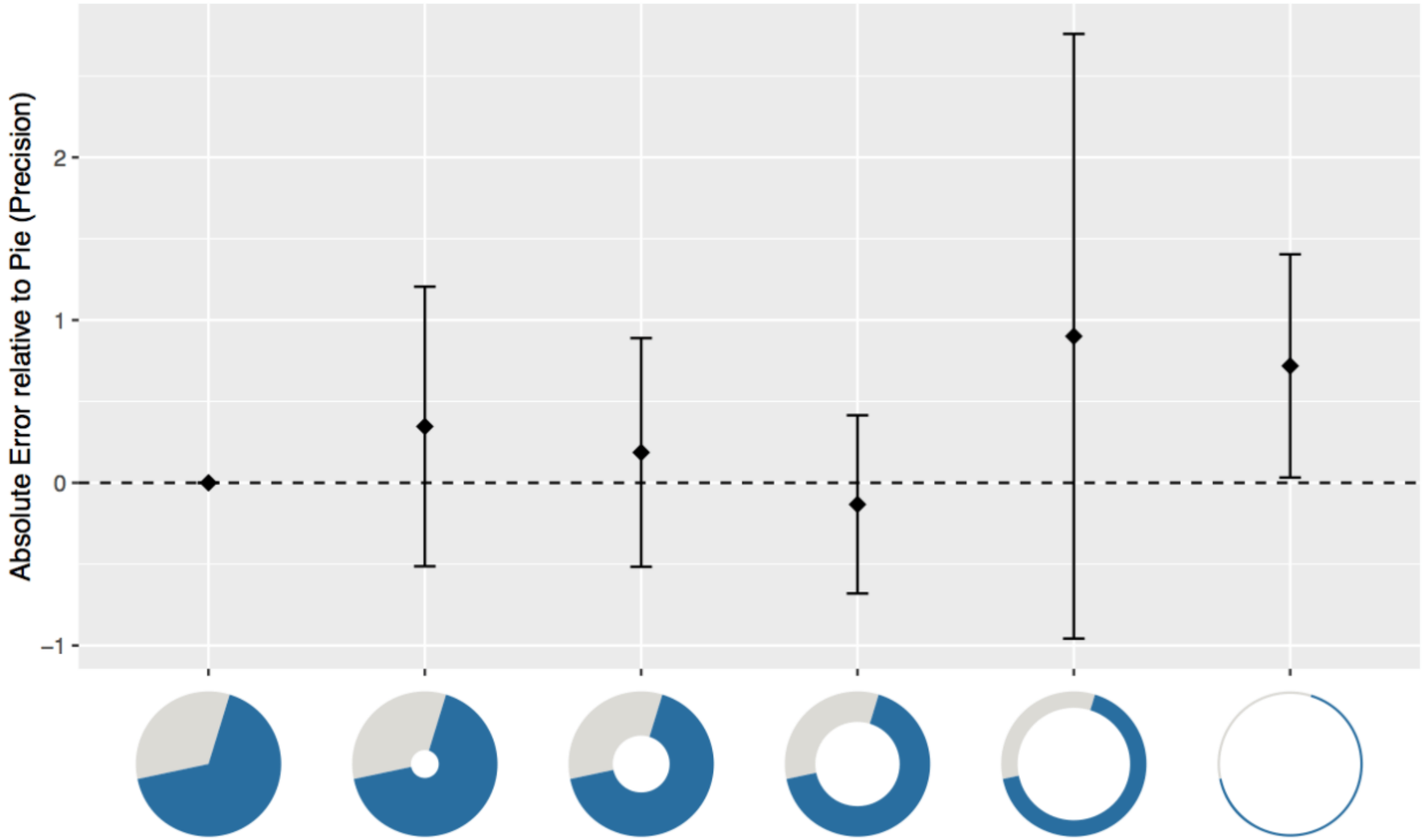
[R. Kosara and D. Skau, 2016]

Absolute Error Relative to Pie Chart



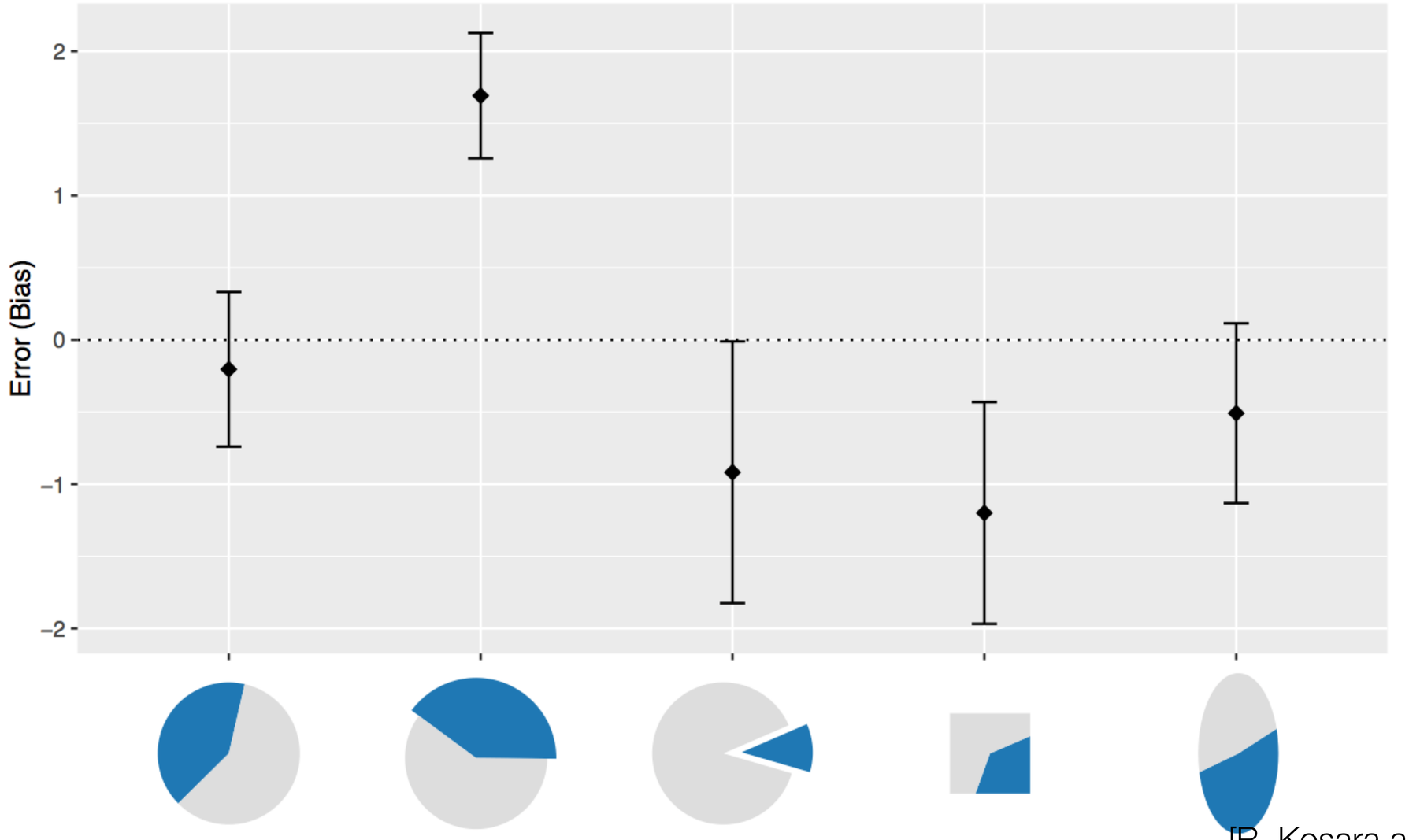
[R. Kosara and D. Skau, 2016]

Donut Charts Width



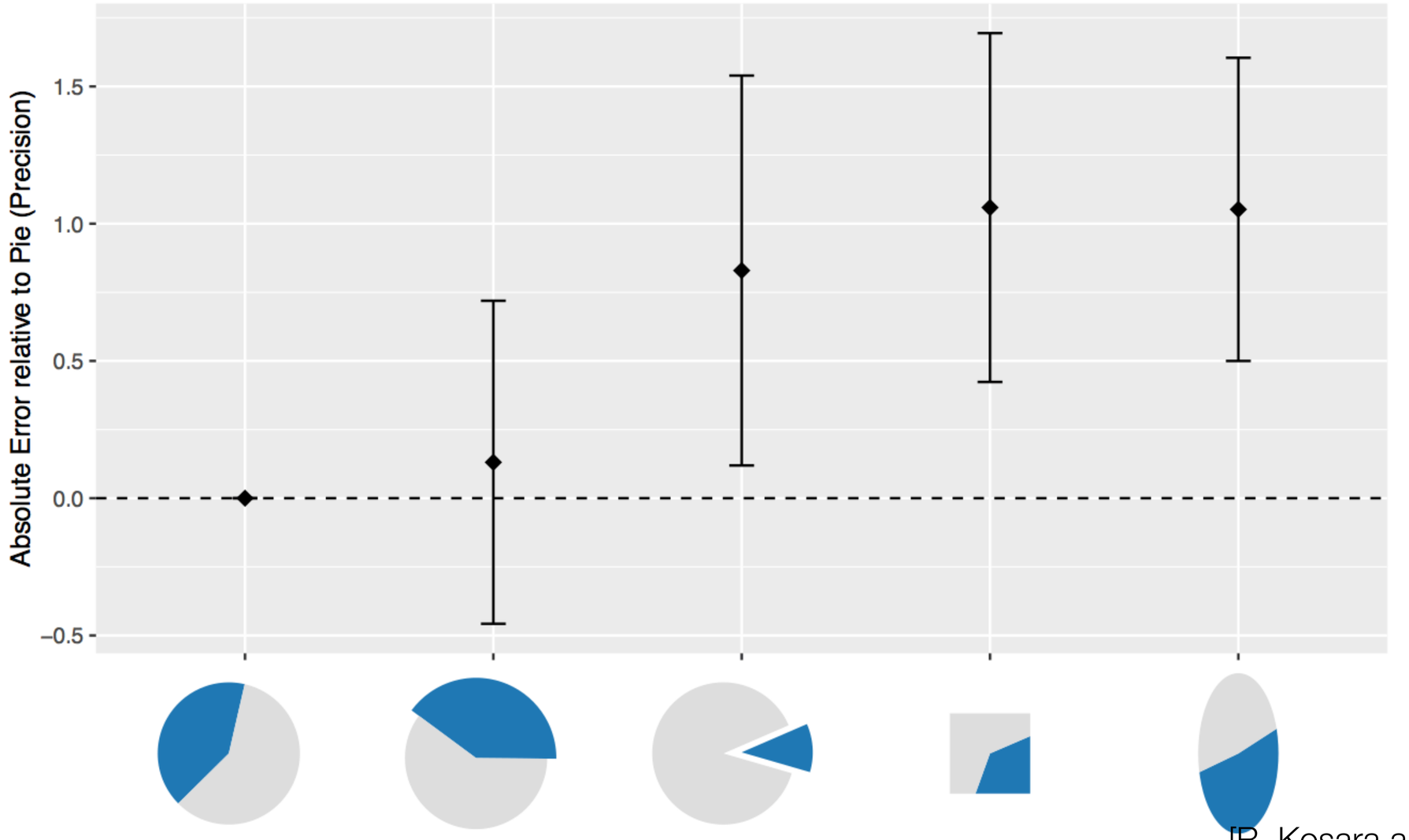
[R. Kosara and D. Skau, 2016]

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

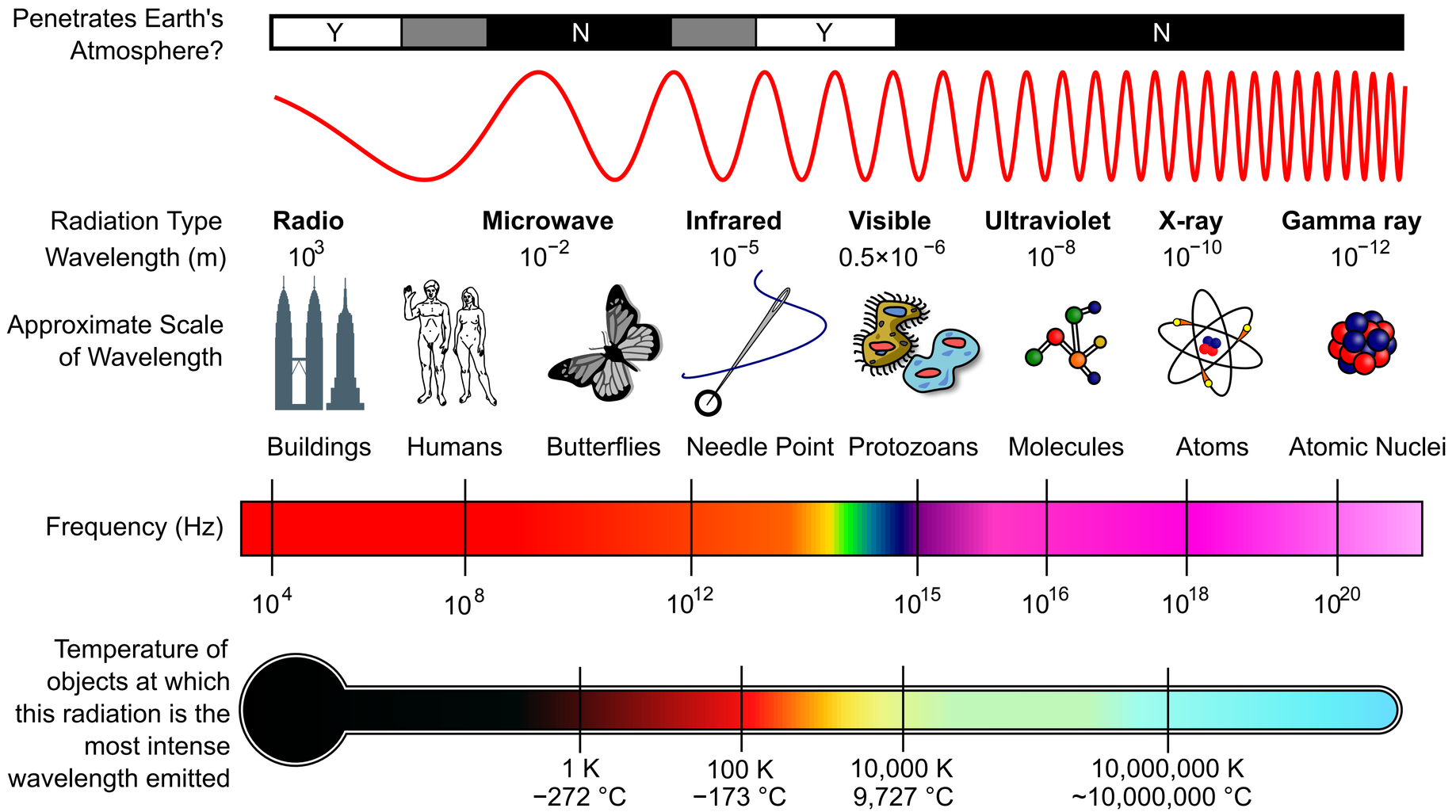
Color

Color

Color and Light

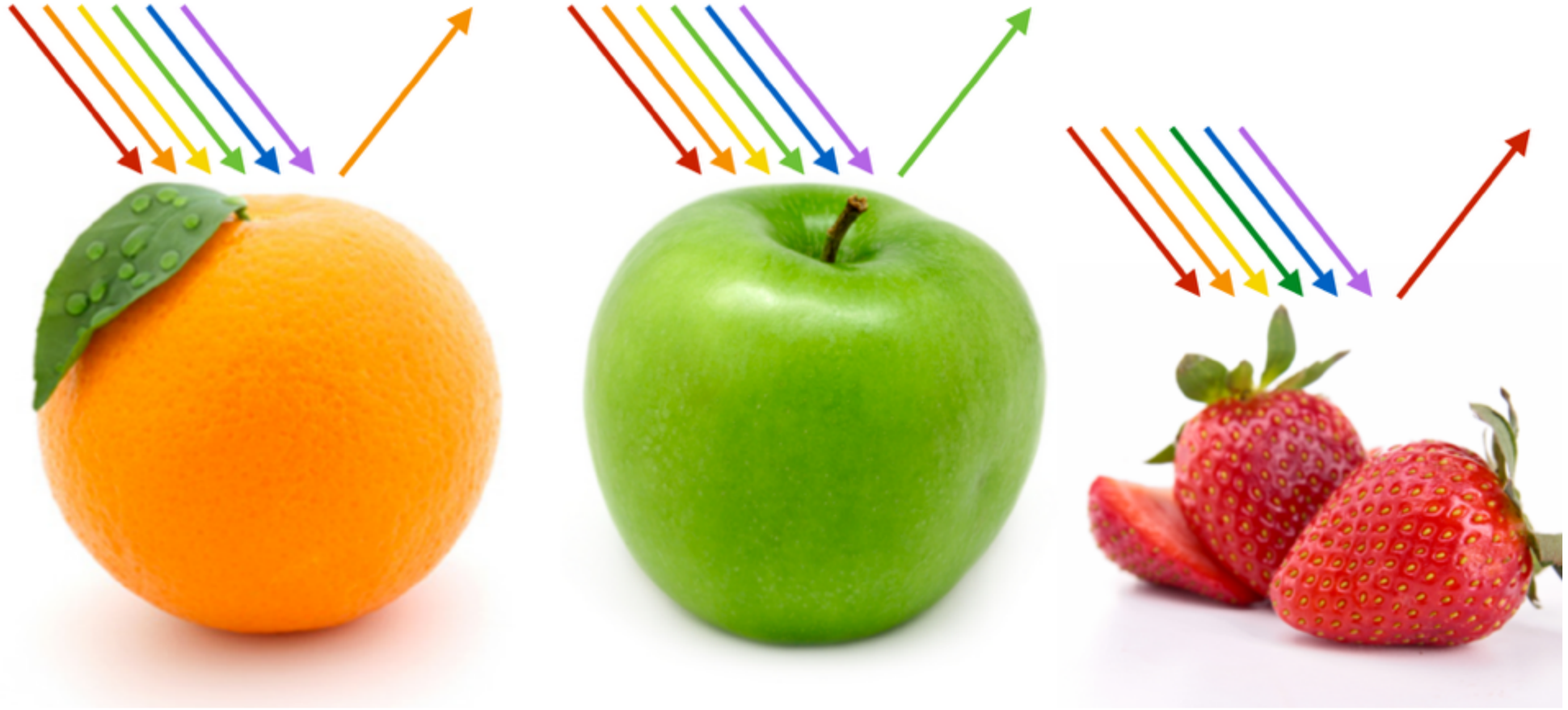
- Color is a **perceptive** property: color depends on the eyes and brain
- Visible light is a small portion of the **electromagnetic spectrum** which is composed of waves that at various frequencies (wavelengths), all traveling at the speed of light

Electromagnetic Spectrum



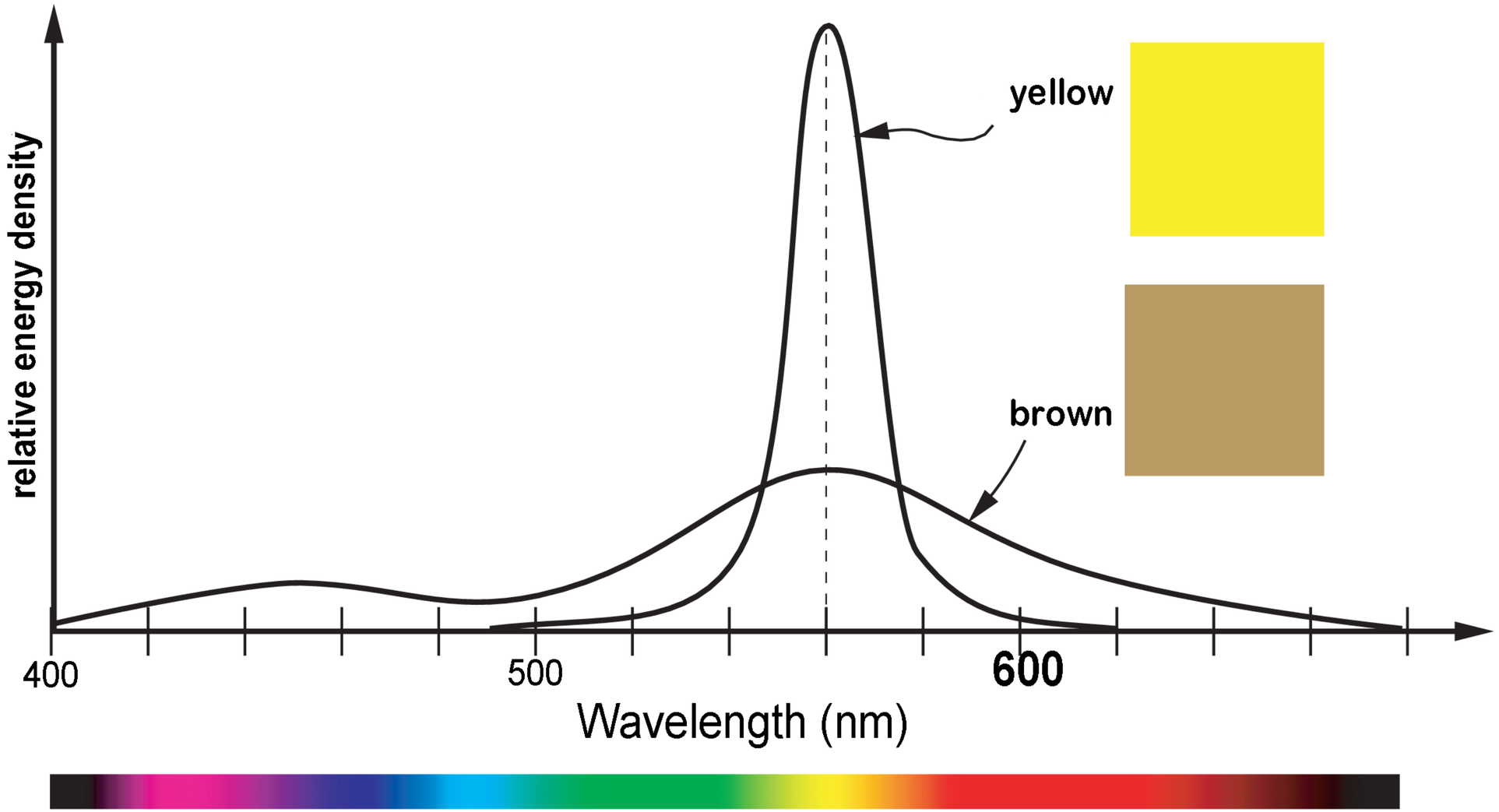
[Wikimedia, NASA]

Light Reflection & Absorption



[via M. Meyer]

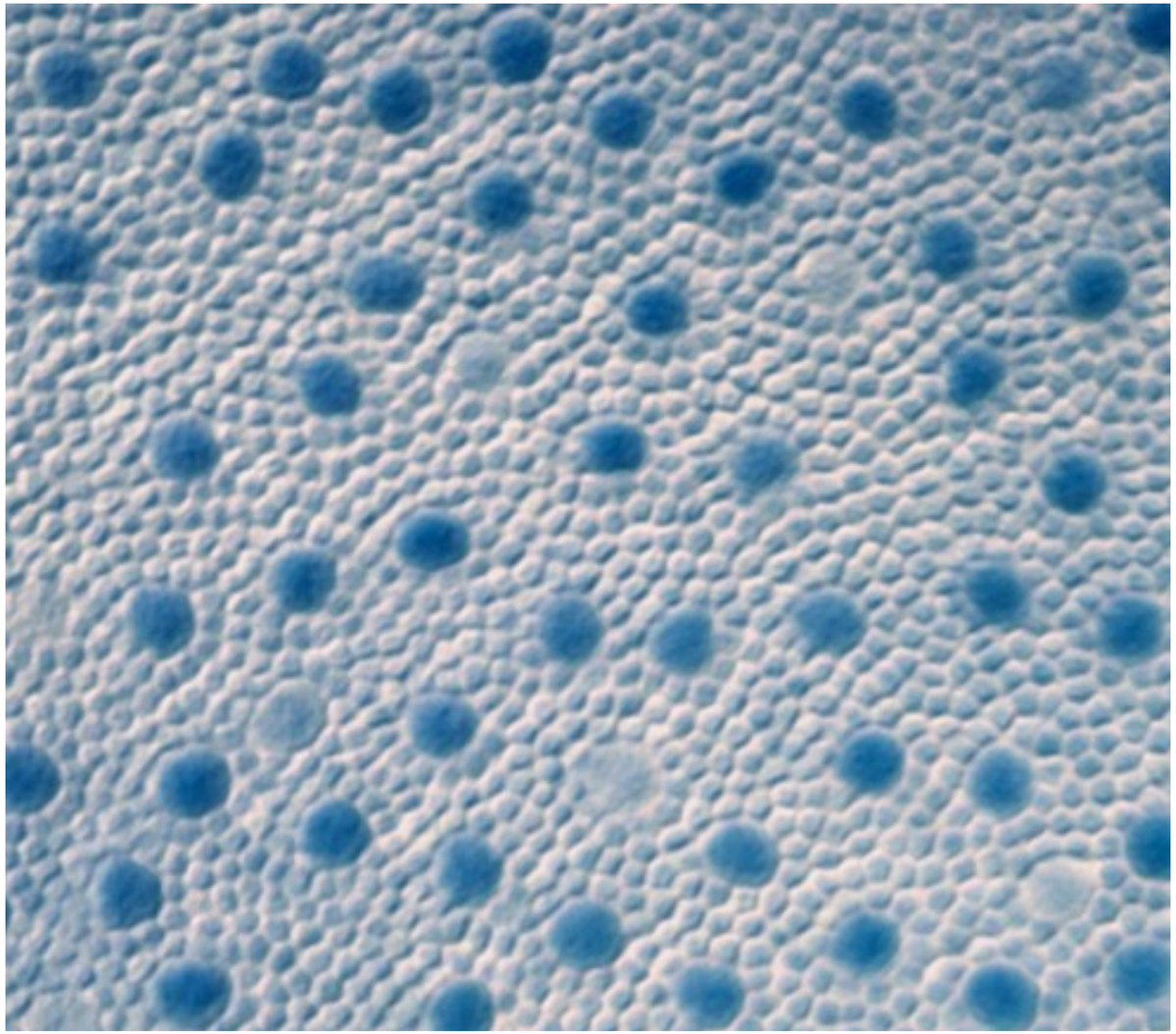
Color != Wavelength



[via M. Meyer]

Human Color Perception

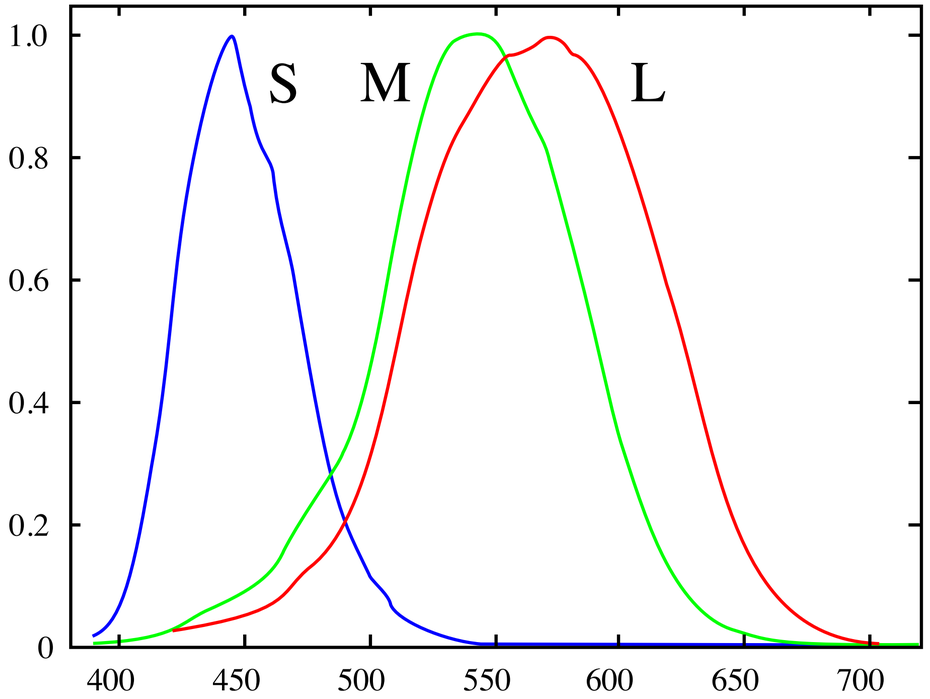
- Humans **do not** detect individual wavelengths of light
- Use **rods** and **cones** to detect light
  - rods capture intensity
  - cones capture color



[N. Cuenca]

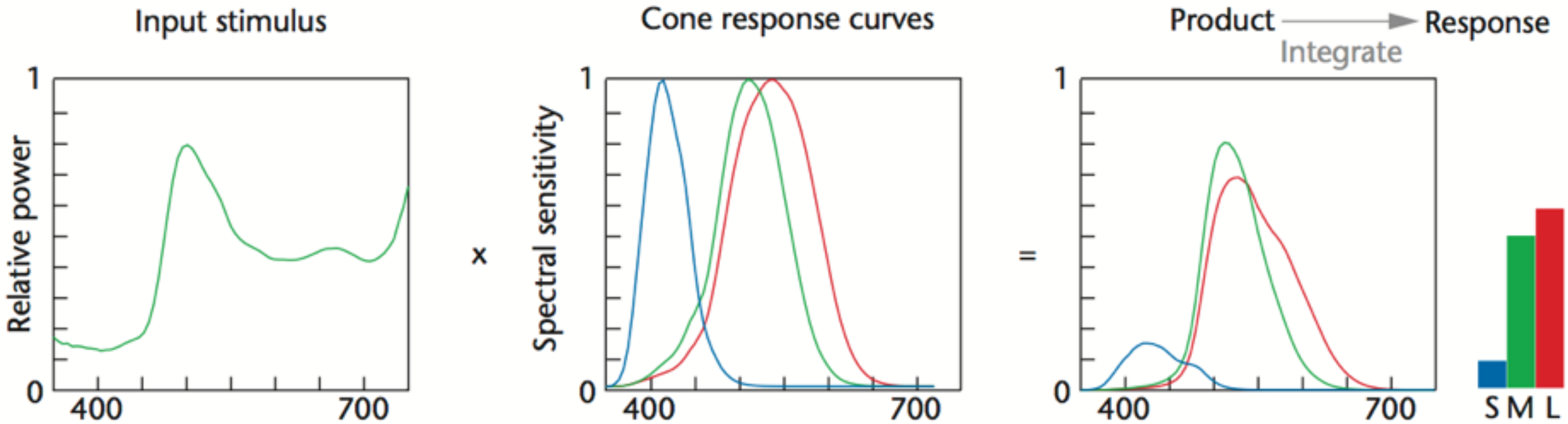
Human Color Perception

- 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



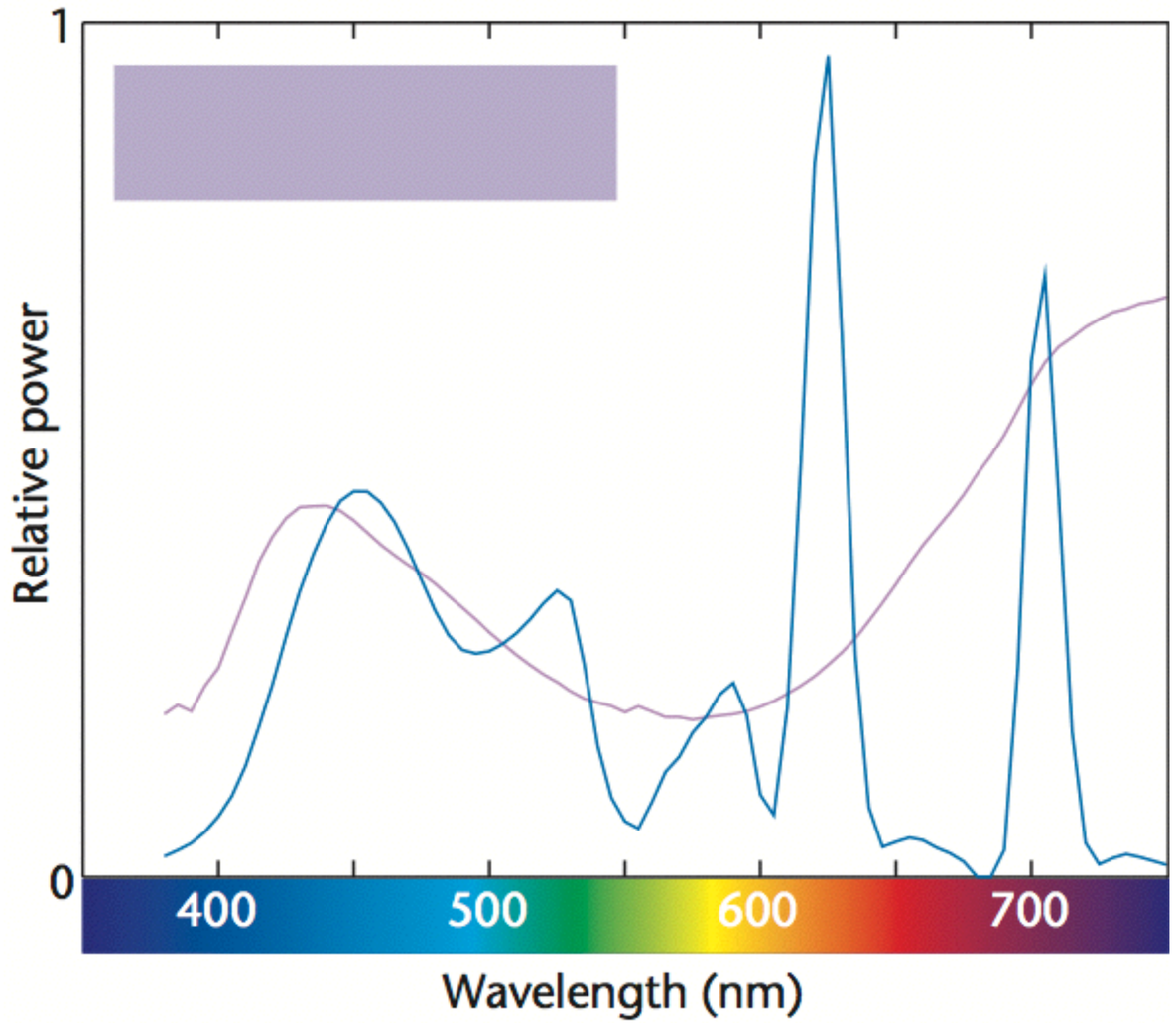
[Vanessaezekowitz at en.wikipedia]

Human Color Perception



[via M. Meyer]

Metamerism



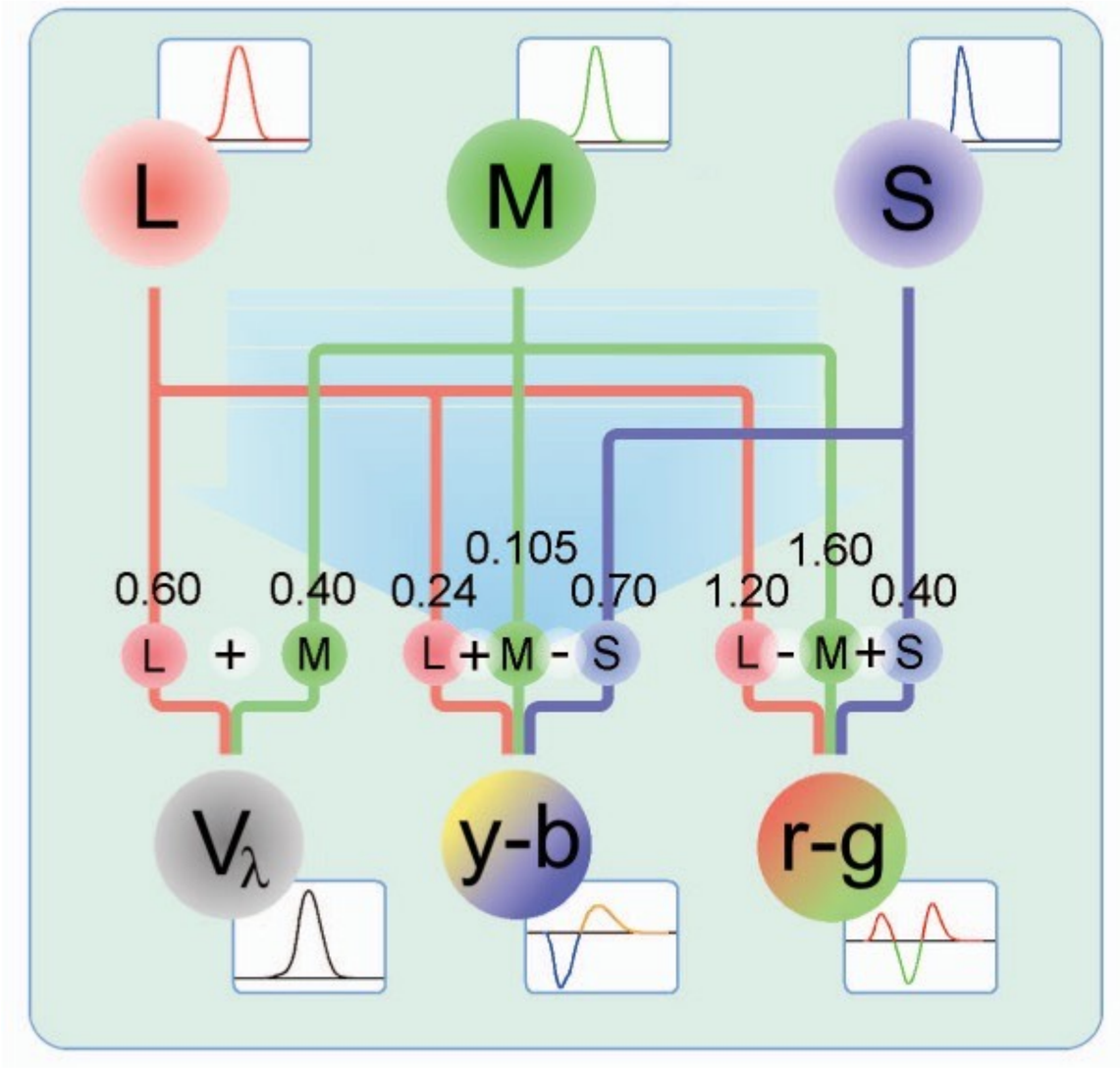
- Same responses == same color
- Humans are not spectrometers
- Do not get the whole function
- Three responses

[via M. Meyer]

Color

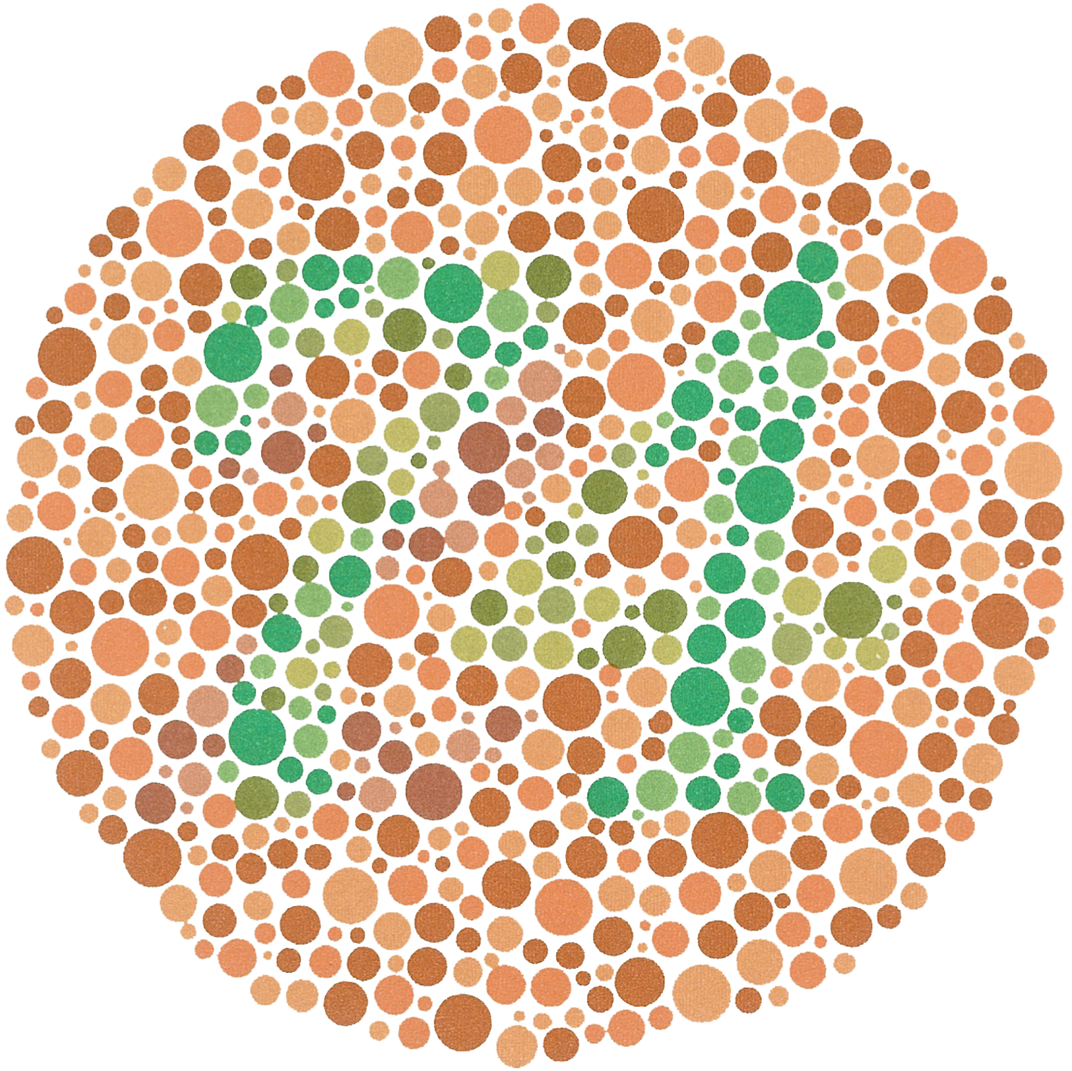
- Cones respond to different areas of the visible light spectrum
- Cover all wavelengths but certain wavelengths generate greater responses
- Color is determined by calculations based on the responses from the different cones
- Opponent Process Theory: three "opponent" channels
  - Light/Dark
  - Blue/Yellow
  - 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