# Data Visualization (CSCI 627/490)

Tabular Data

Dr. David Koop

Northern Illinois University

# 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?

# Ranking Channels by Effectiveness



**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)

Same

Same

Most

Effectiveness

Least

**Identity** Channels: **Categorical** Attributes

- Spatial region
- Color hue
- Motion
- Shape

[Munzner (ill. Maguire), 2014]

# Perception Studies Summary



Cleveland & McGill's Results

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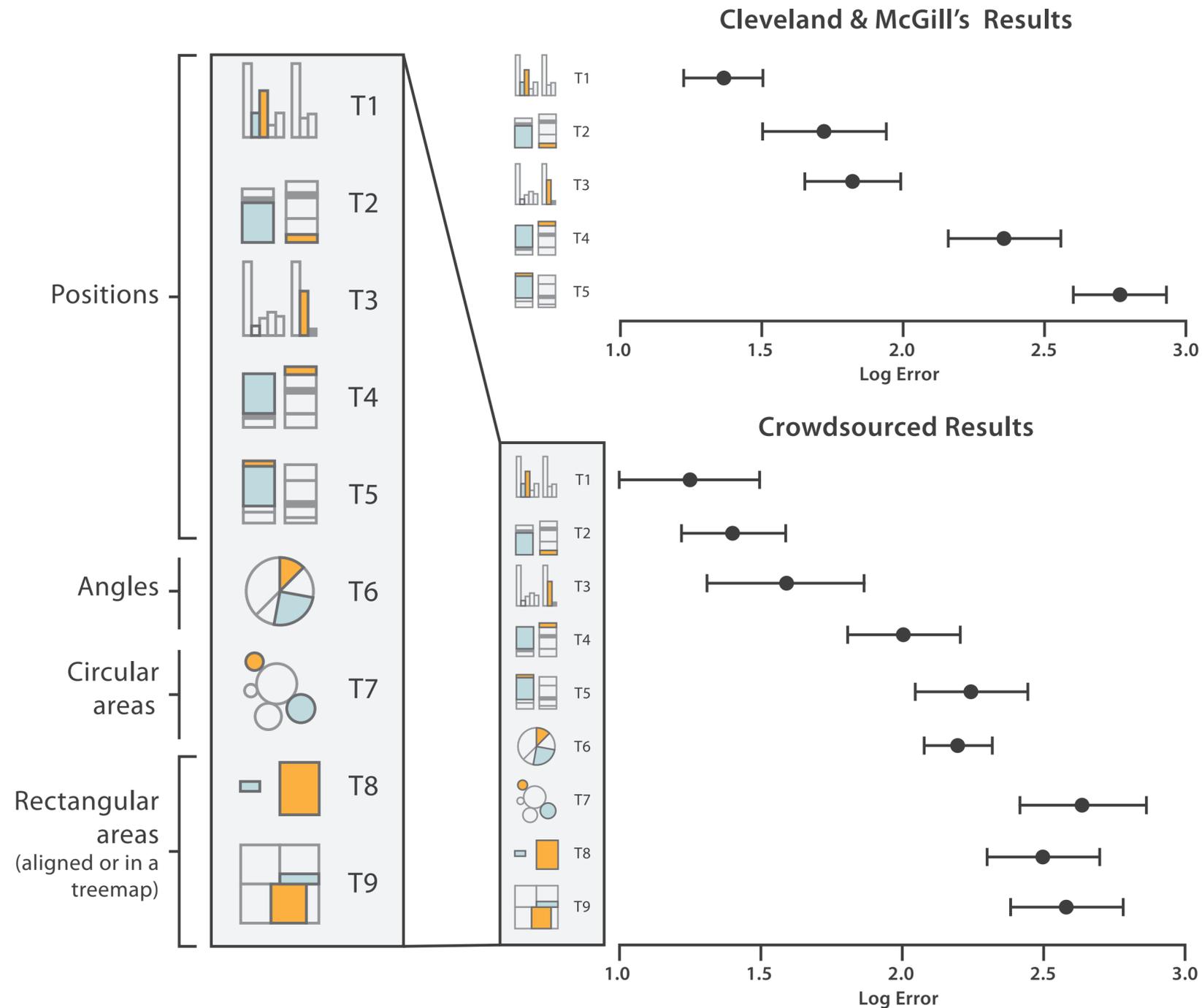
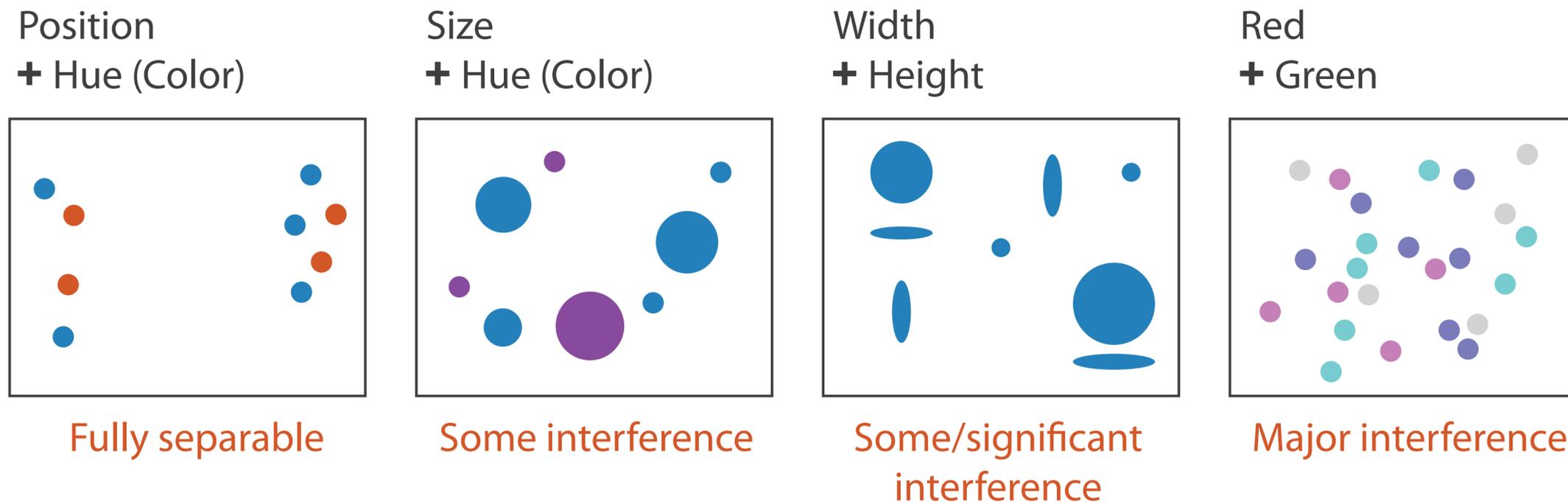[Munzner (ill. Maguire) based on Heer & Bostock, 2014]

# Discriminability

- Width encodes count of number of networks with a particular link.
- What is problematic here?
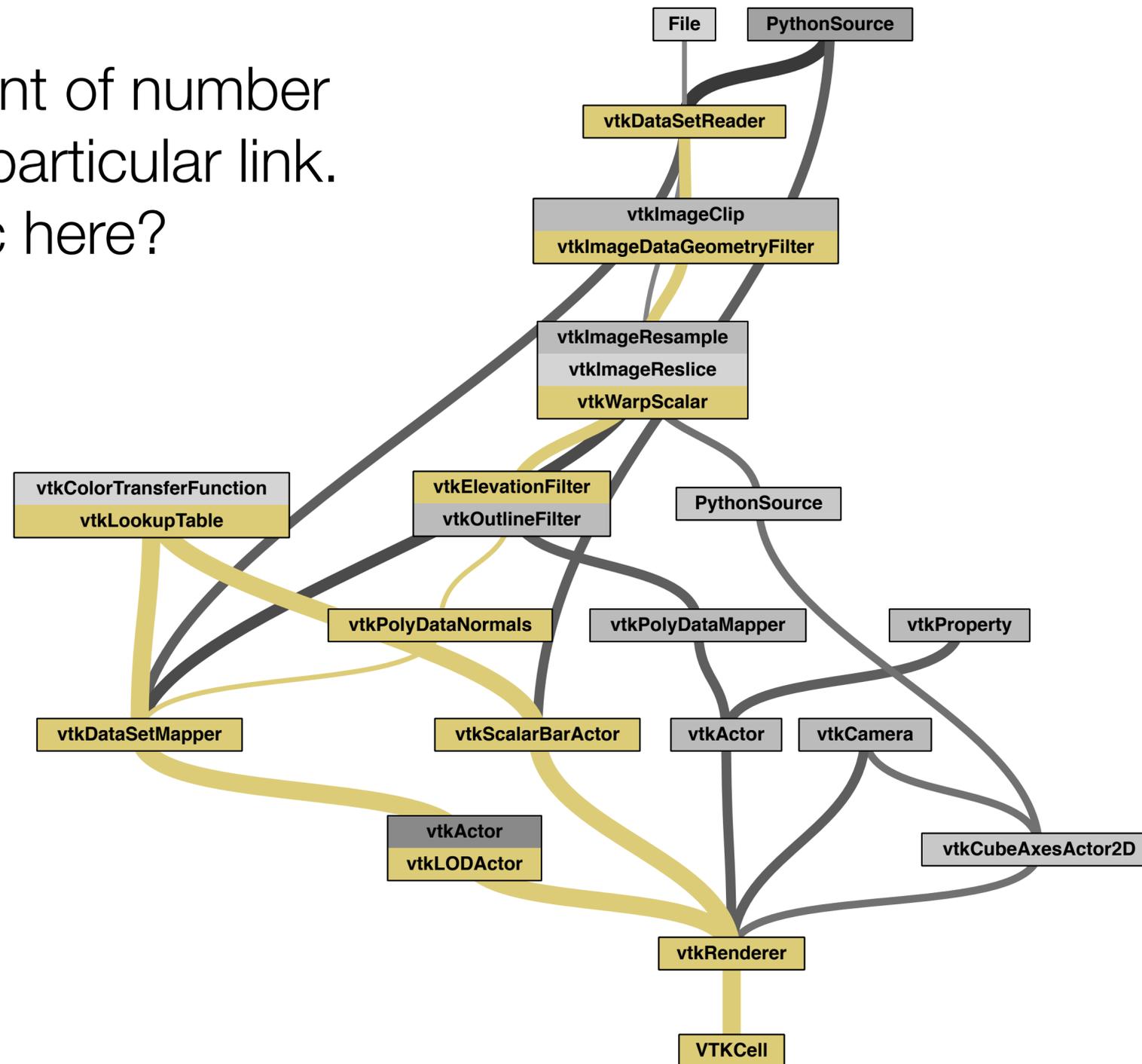


[Koop et al., 2013]

# Separability

- Cannot treat all channels as independent!
- **Separable** means each individual channel can be distinguished
- **Integral** means the channels are perceived together



| Position + Hue (Color) | Size + Hue (Color) | Width + Height | Red + Green |
|---|---|---|---|
| Fully separable | Some interference | Some/significant interference | Major interference |

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

# Project Proposal

- Due today

- Explore the listed project datasets

- Consider a group

- Submit proposal

  - Dataset including attributes with classifications

  - Tasks (make sure they are visualization tasks)

  - Initial sketch

- Goal: Interesting questions that allow different designs

- If you are doing research and can tie this project in, please contact me

# Assignment 3



Chicago Traffic Crashs 2024-2025
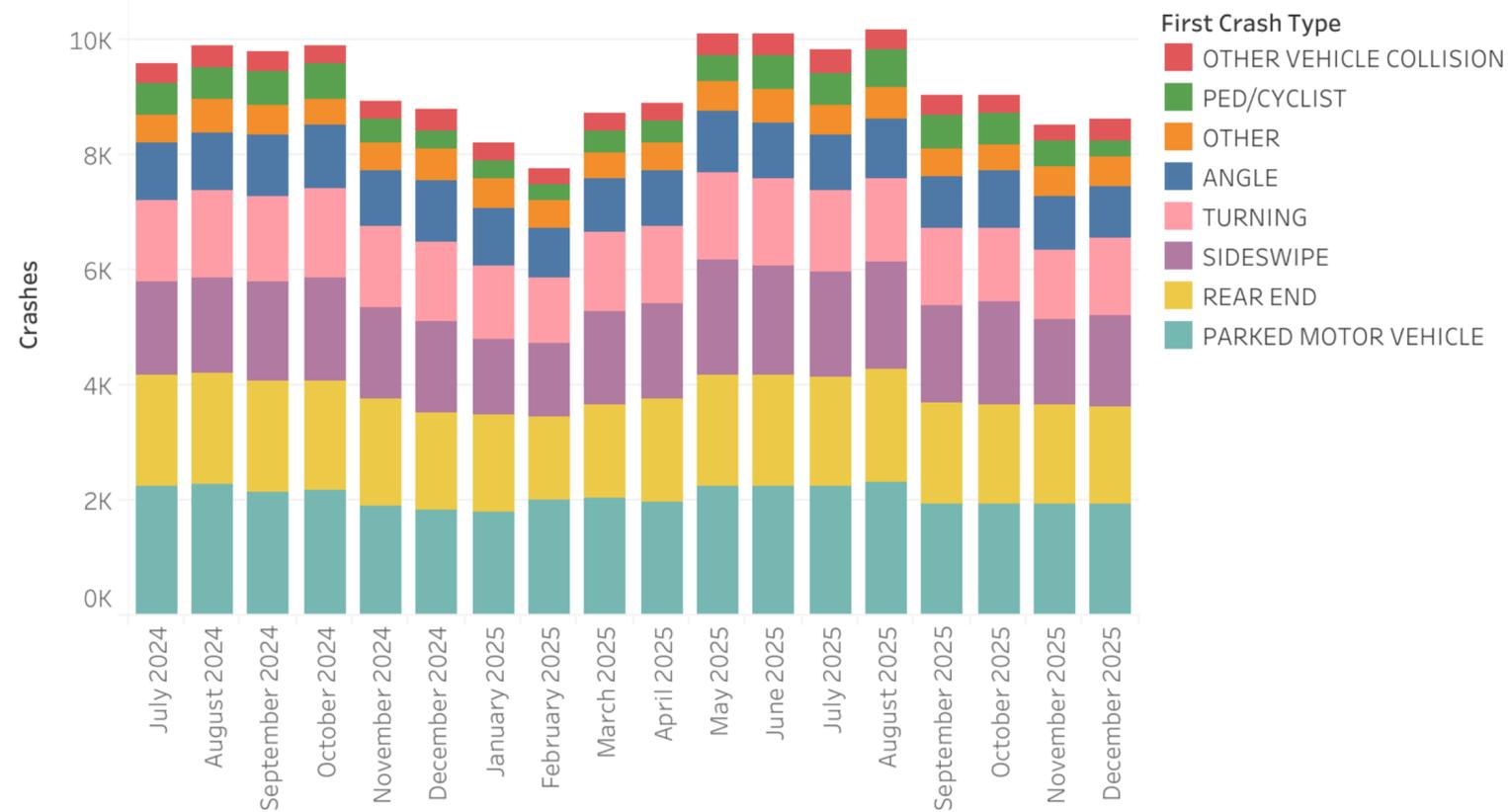
- Due next week
- Chicago Traffic Crashes Data
- Create same stacked bar chart with
  - Tableau Public
  - Observable Plot
  - D3
- D3 Stacked Bar Chart:
  - Required for CSCI 627 students
  - Extra credit for CSCI 490 students

# Courselets

- Educational resources for visualization using notebooks

- Reviewed charts over the last couple of classes, how do we construct them?

- How do we use visualization libraries, including those in other contexts like Python?

  - matplotlib: charts-matplotlib.ipynb

  - pyobsplot: charts-obsplot.ipynb

# Midterm

- In-class, Wednesday, March 4, 12:30-1:45pm

- Only need writing utensil (+eraser)

- Format:

  - Multiple Choice

  - Free Response

- CSCI 627 students will have an extra double-sided page with more research-focused questions

# Arrange Tables

**➔ Express Values**



**➔ Separate, Order, Align Regions**
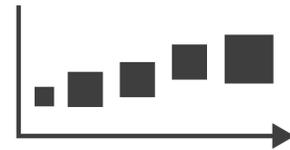
➔ Separate     ➔ Order     ➔ Align



**➔ Layout Density**
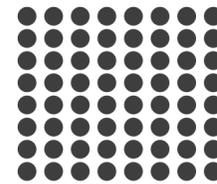
➔ Dense     ➔ Space-Filling



➔ *1 Key*
*List*

➔ *2 Keys*
*Matrix*
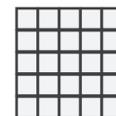
➔ *3 Keys*
*Volume*
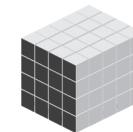
➔ *Many Keys*
*Recursive Subdivision*



**➔ Axis Orientation**
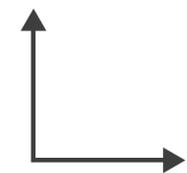
➔ Rectilinear     ➔ Parallel     ➔ Radial
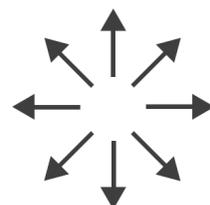


[Munzner (ill. Maguire), 2014]

# Express Values: Scatterplots
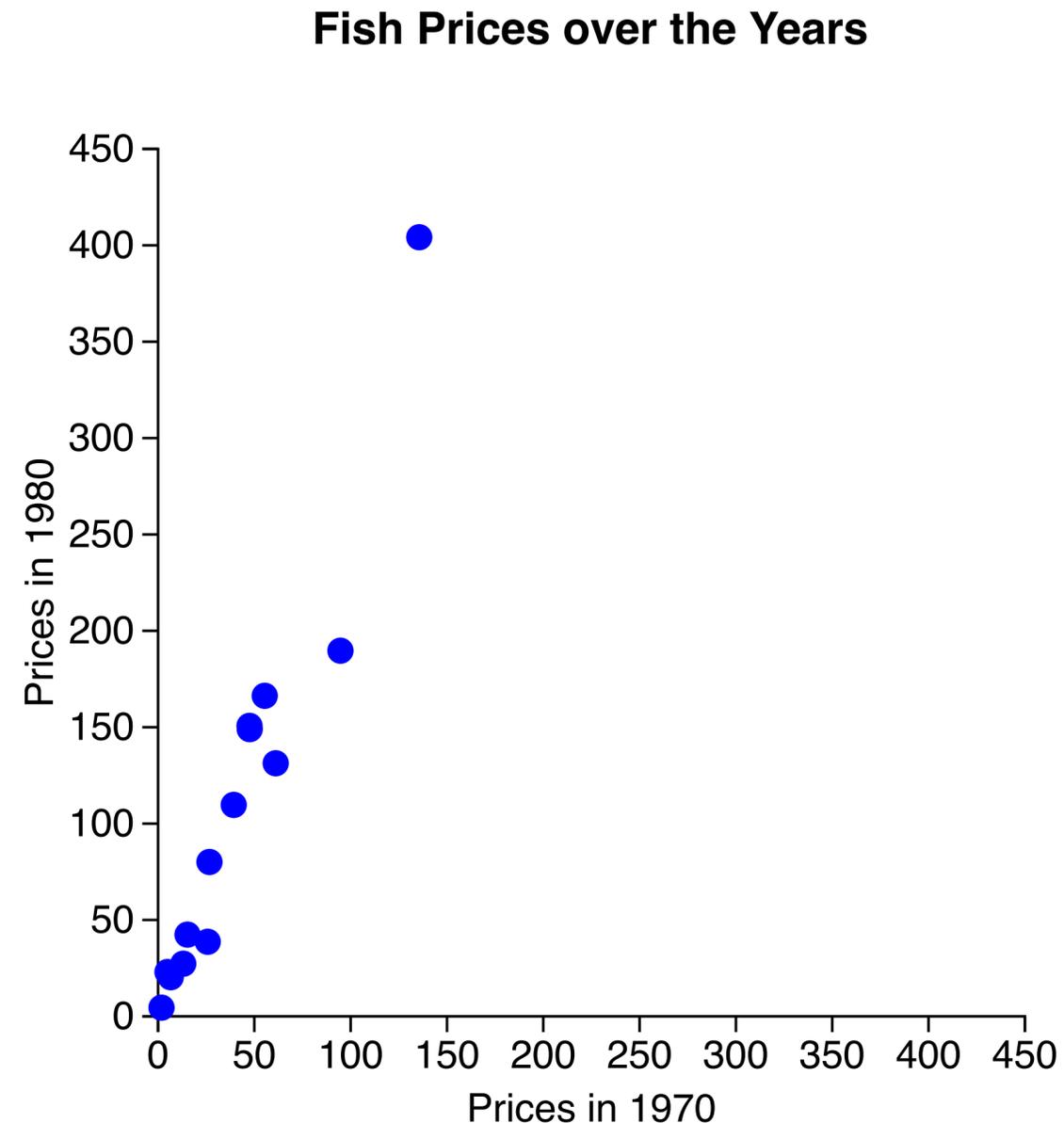
**Fish Prices over the Years**



- 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
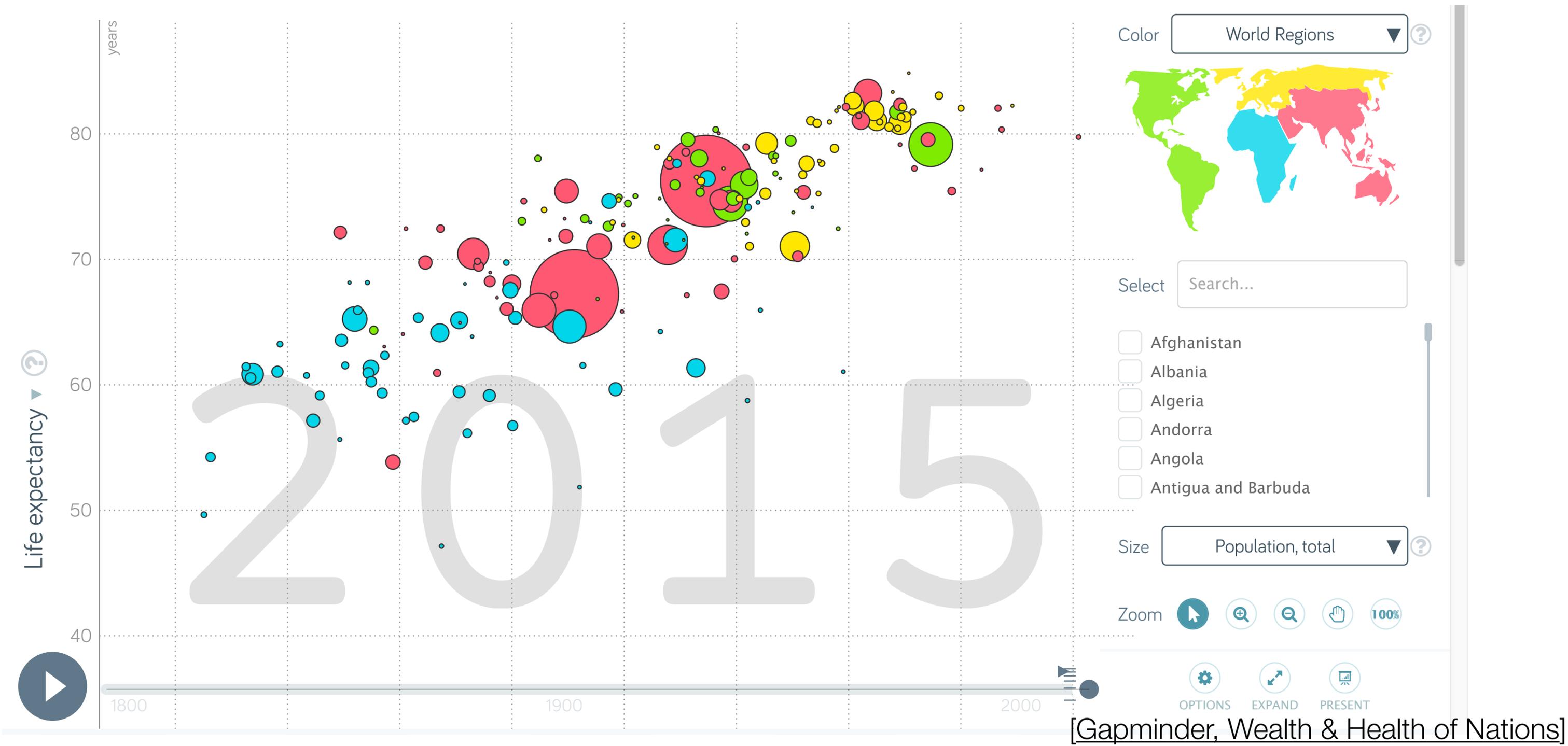- Coordinate system (axes) and labels are important!

# Coordinate Systems
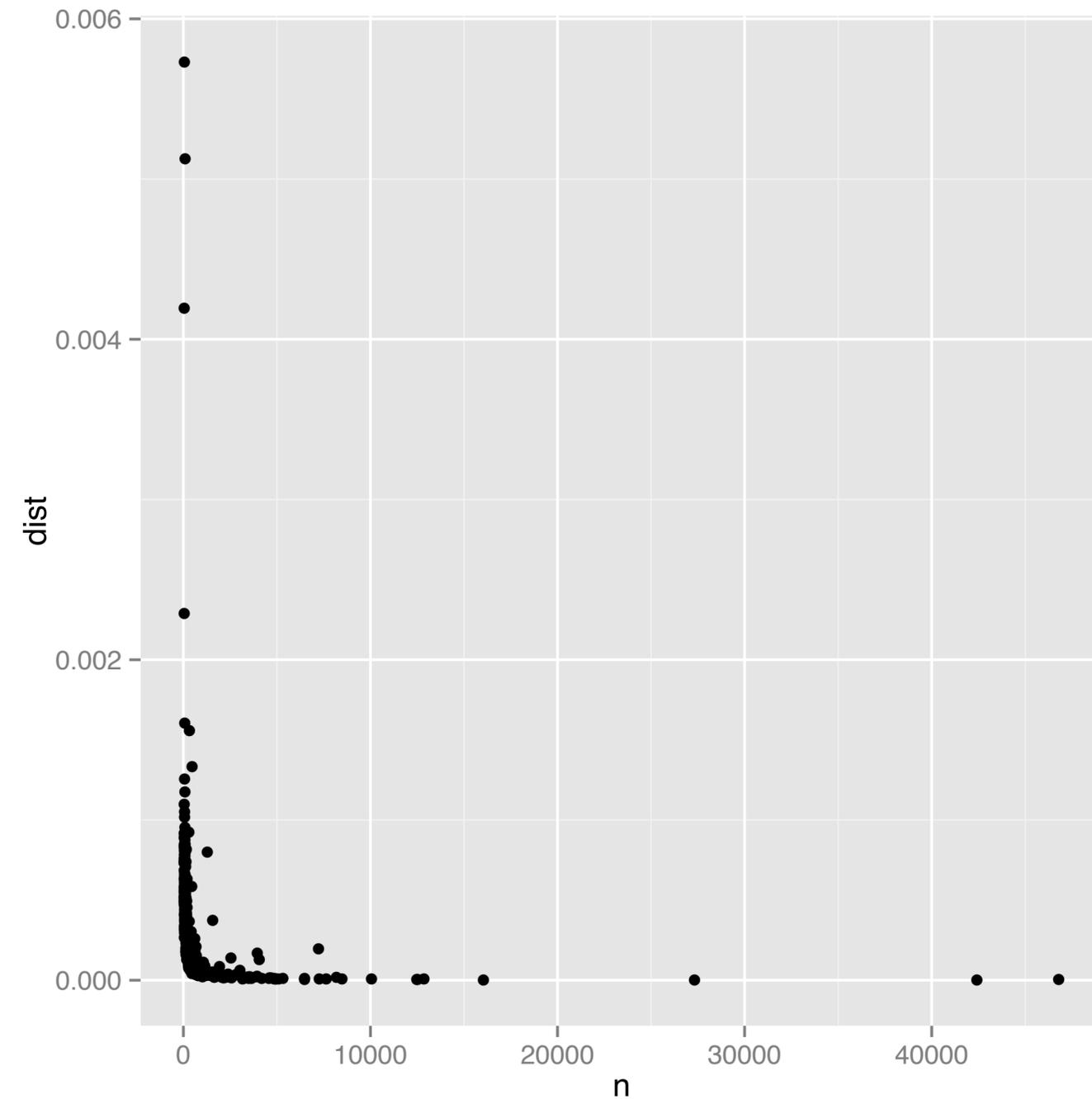


[Wickham, 2014]

# Coordinate Systems
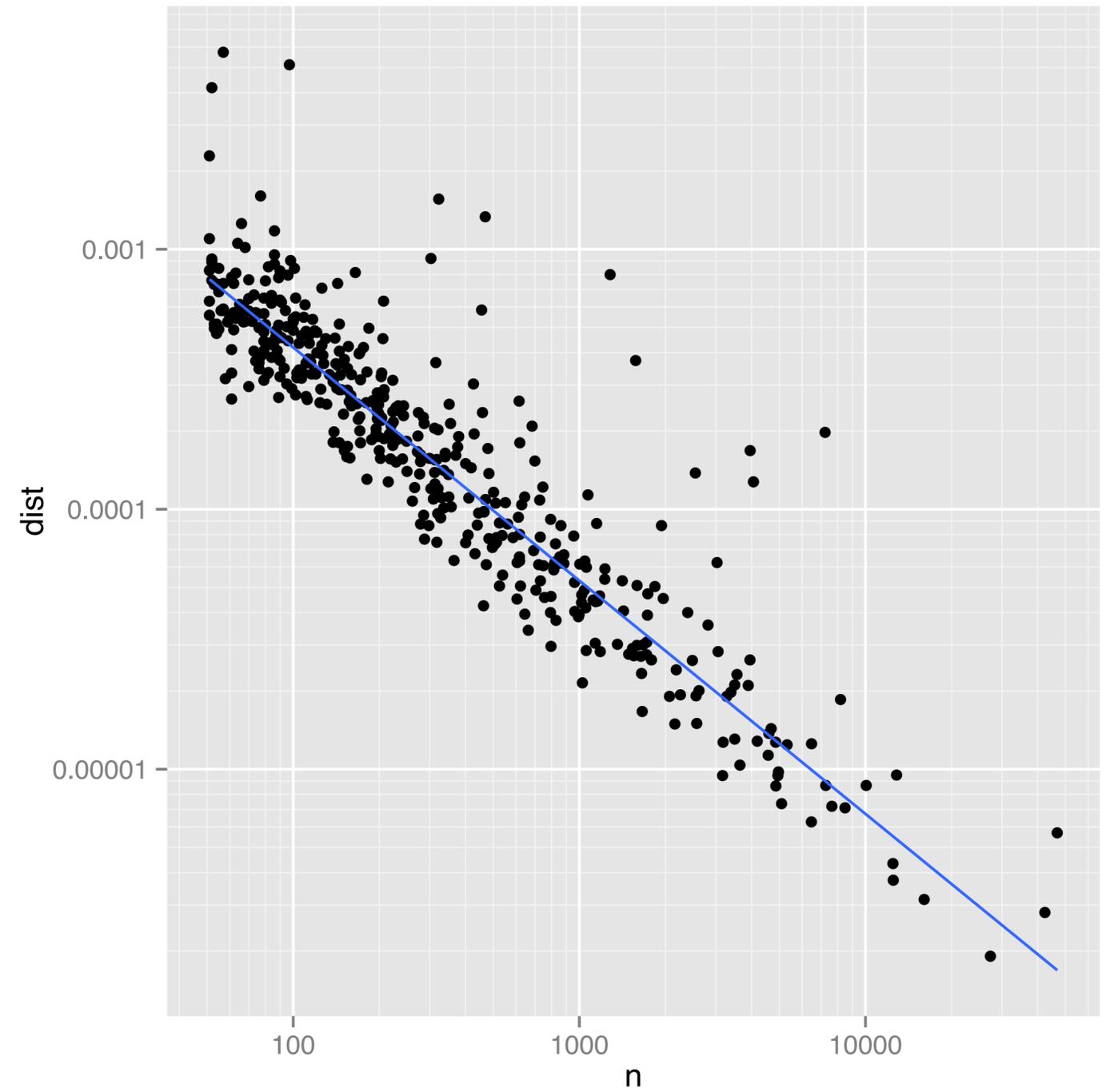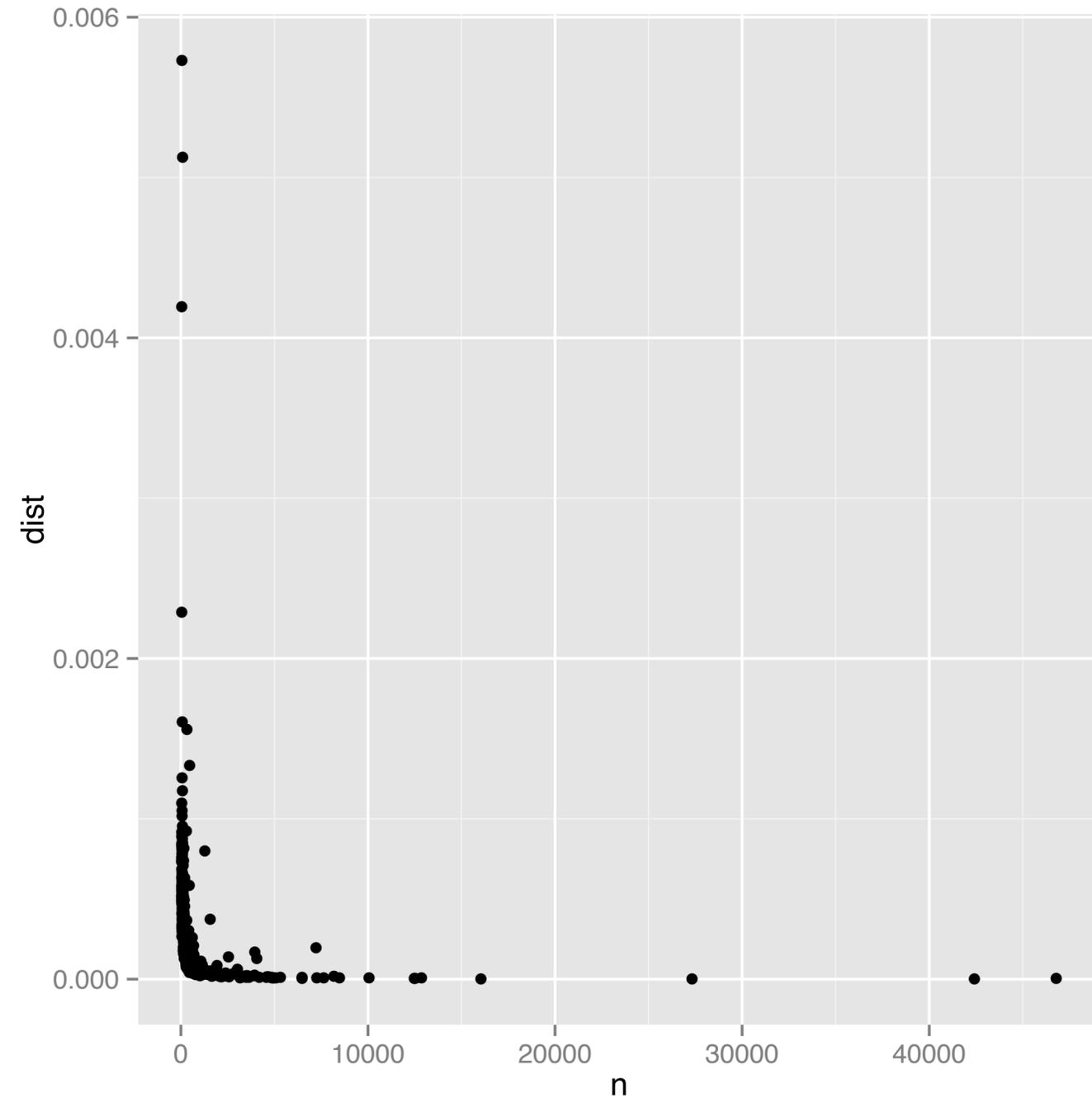


[Wickham, 2014]

# Bubble Plot



[Gapminder, Wealth & Health of Nations]

# Scatterplot

- Data: two quantitative values

- Task: find trends, clusters, outliers

- How: marks at spatial position in horizontal and vertical directions

- **Scalability**: hundreds of items

- "Ranking Visualizations of Correlation Using Weber's Law", 2014:

  - Correlation perception can be modeled via Weber's Law

  - Scatterplots are one of the best visualizations for both positive and negative correlation

  - Further analysis: M. Kay and J. Heer, "Beyond Weber's Law", 2015

# Separate, Order, and Align: Categorical Regions

- Categorical: =, !=
- Spatial position can be used for categorical attributes
- Use **regions**, distinct contiguous bounded areas, to encode categorical attributes
- Three operations on the regions:
  - Separate (use categorical attribute)
  - Align
  - Order
  (use some other ordered attribute)
- Alignment and order can use same or different attribute

# 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
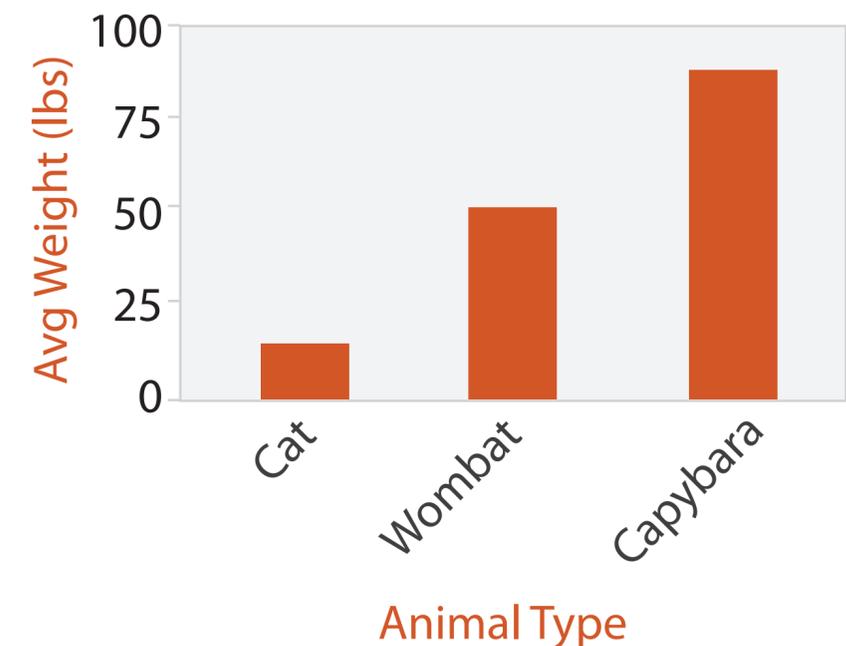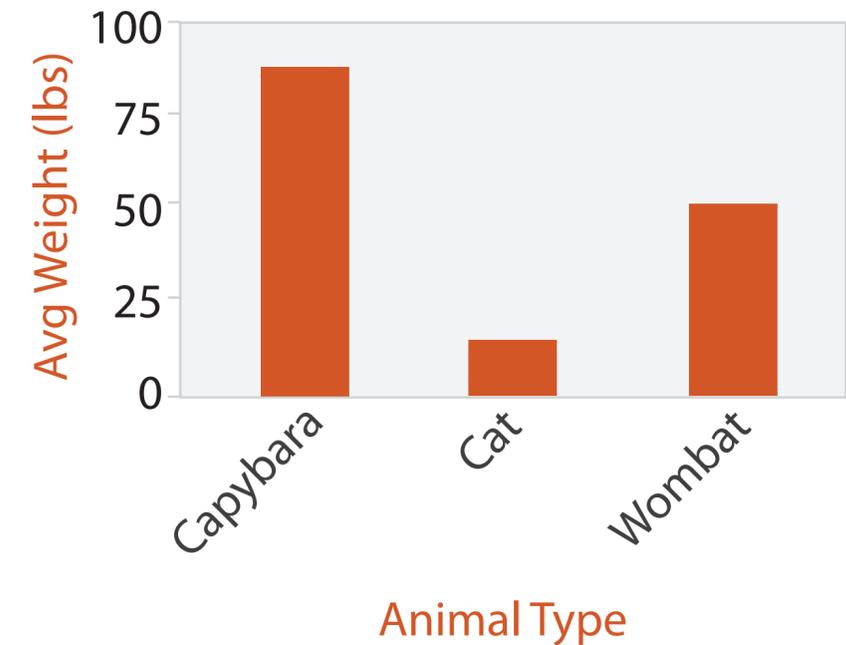


[Munzner (ill. Maguire), 2014]

# Stacked Bar Charts
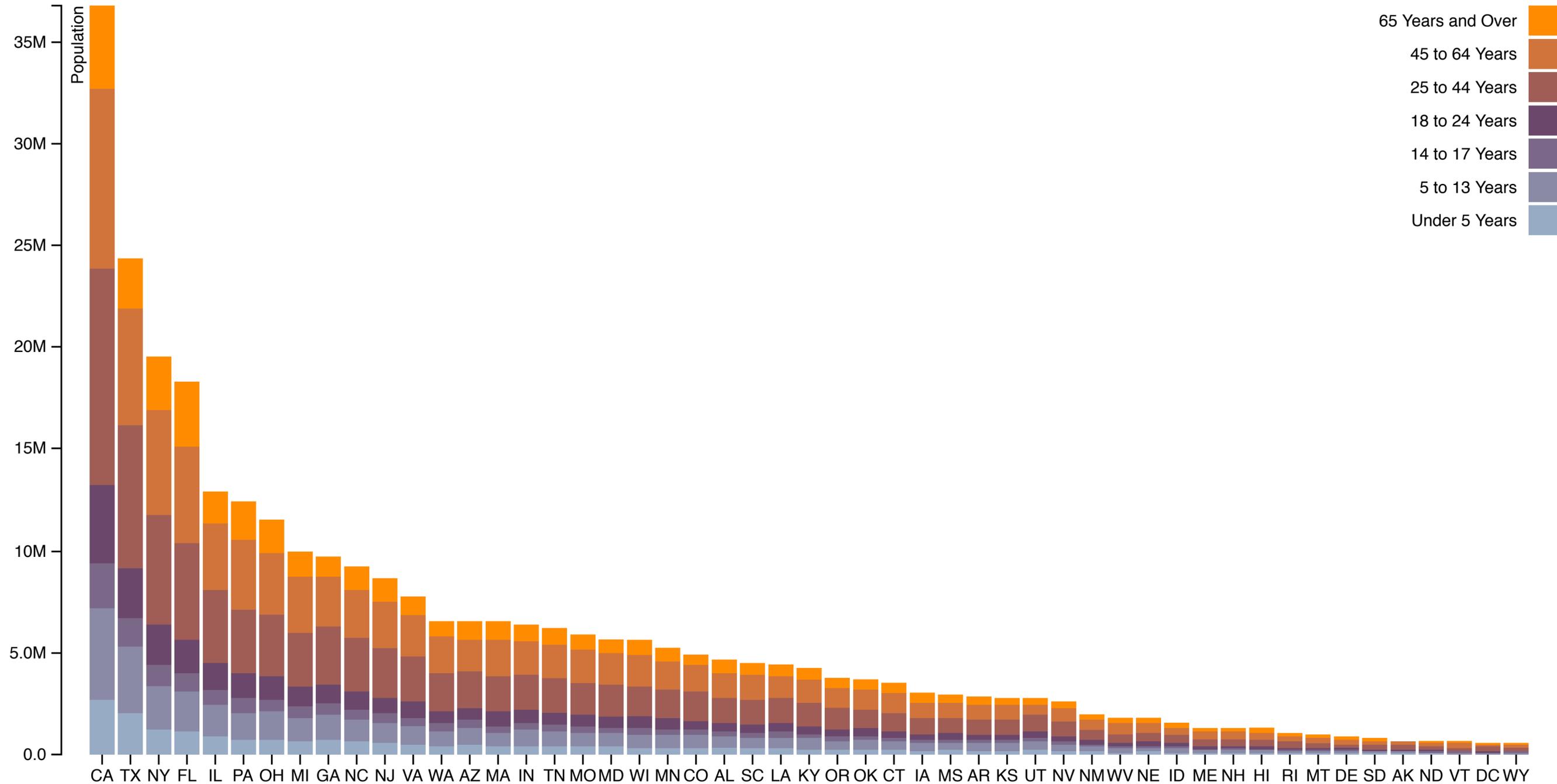


[Stacked Bar Chart, M. Bostock, 2017]

# Grouped Bar Chart
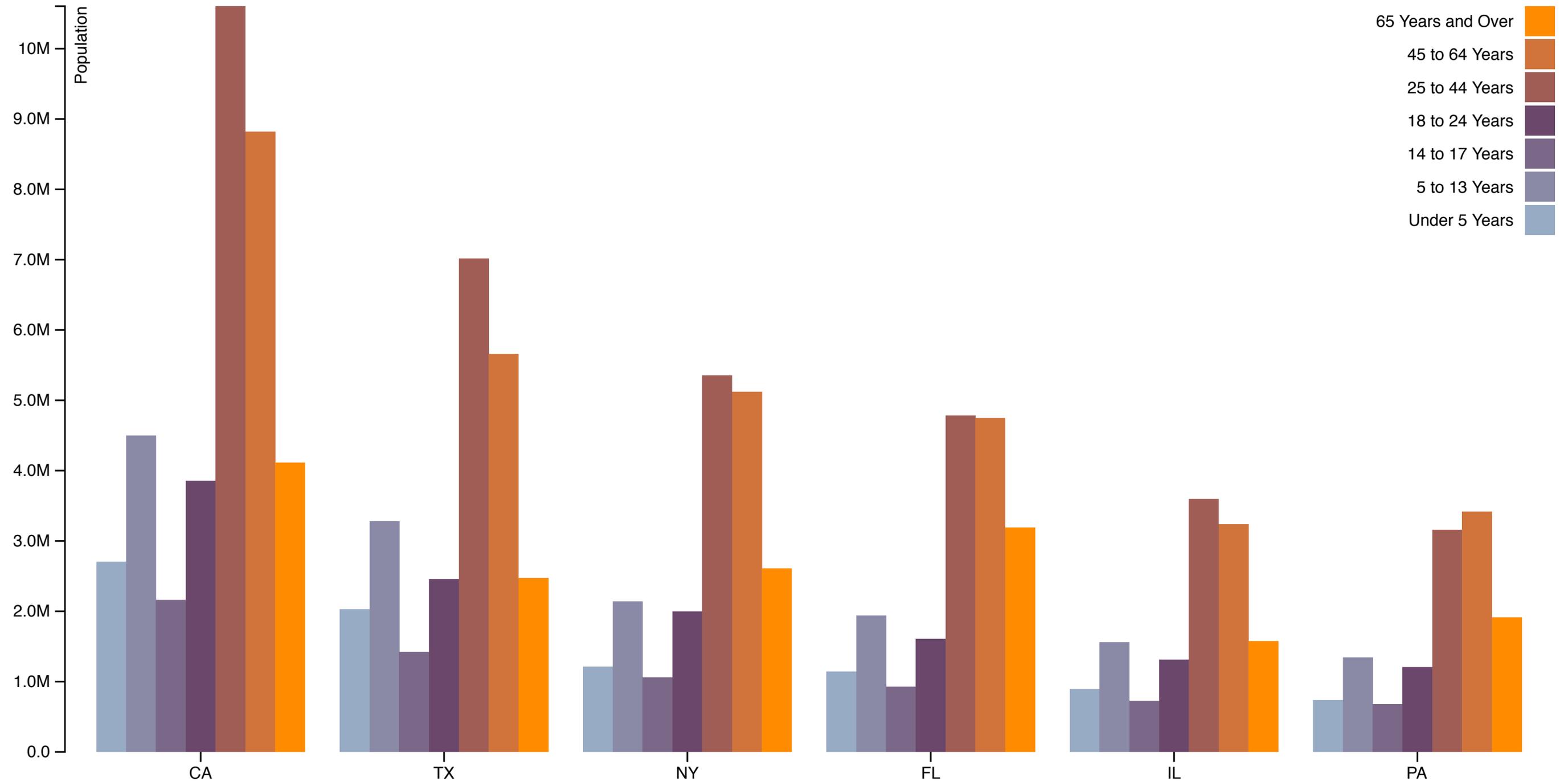


[Grouped Bar Chart, M. Bostock, 2017]

# Stacked Bar Charts

- Data: multidimensional table: one quantitative, **two** categorical

- Task: lookup values, part-to-whole relationship, trends

- How: line marks: position (both horizontal & vertical), subcomponent line marks: length, color

- Scalability: main axis (hundreds like bar chart), bar classes (<12)

- Orientation: vertical or horizontal (swap how horizontal and vertical position are used.

# 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
- Scalability: more categories than stacked bar charts
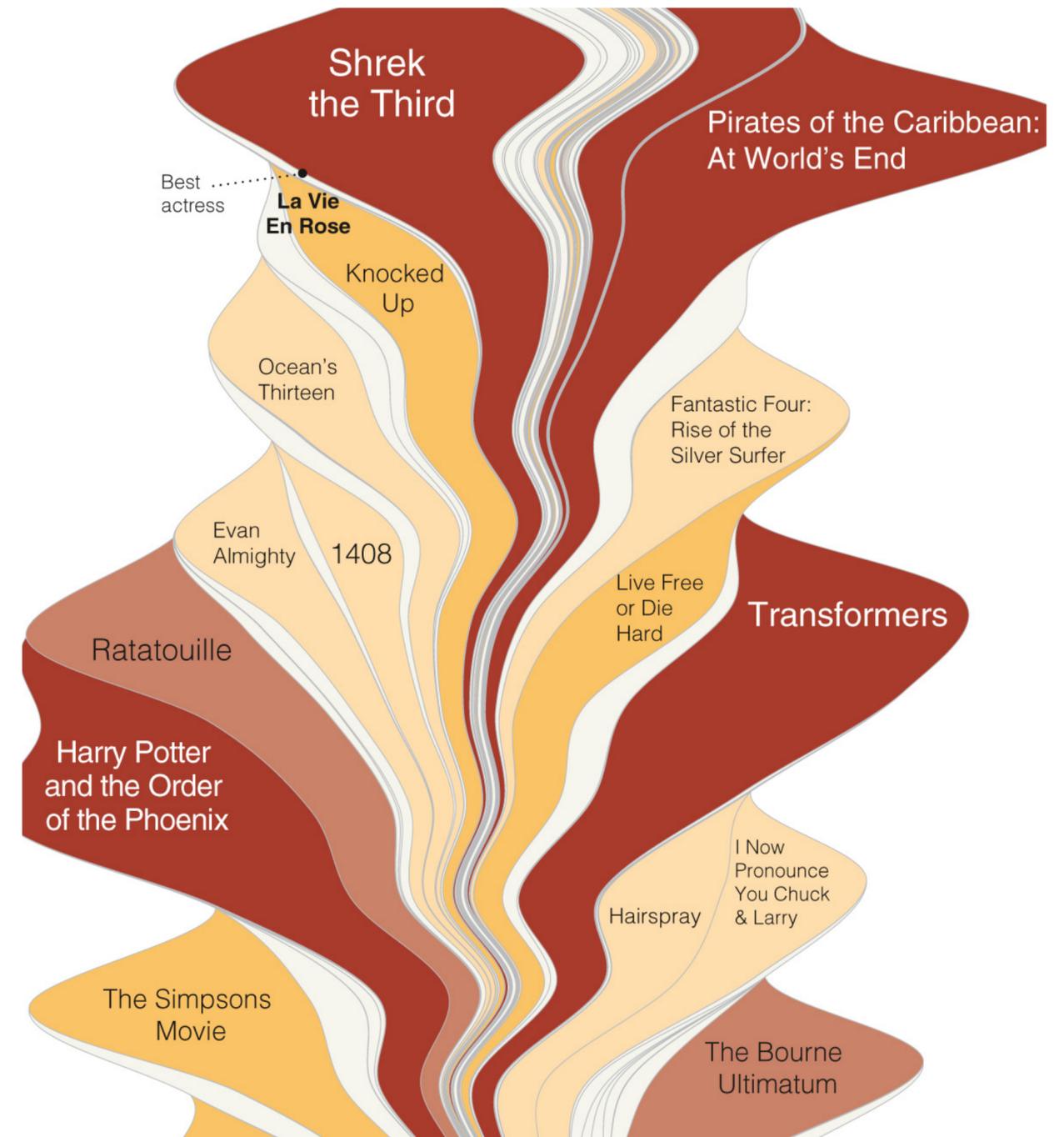


[Byron and Wattenberg, 2012]

# Streamgraphs
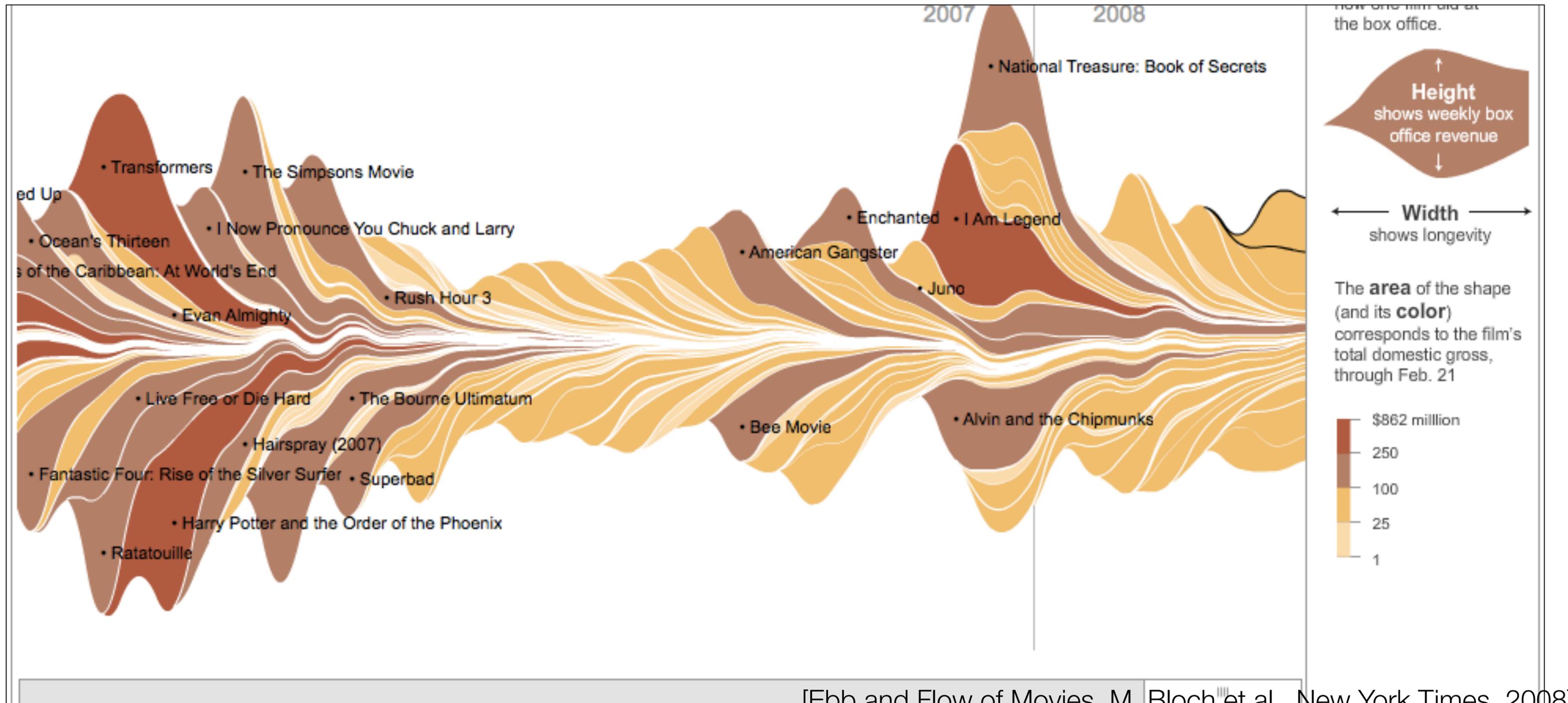


[Ebb and Flow of Movies, M. Bloch et al., New York Times, 2008]

# Dot and Line Charts
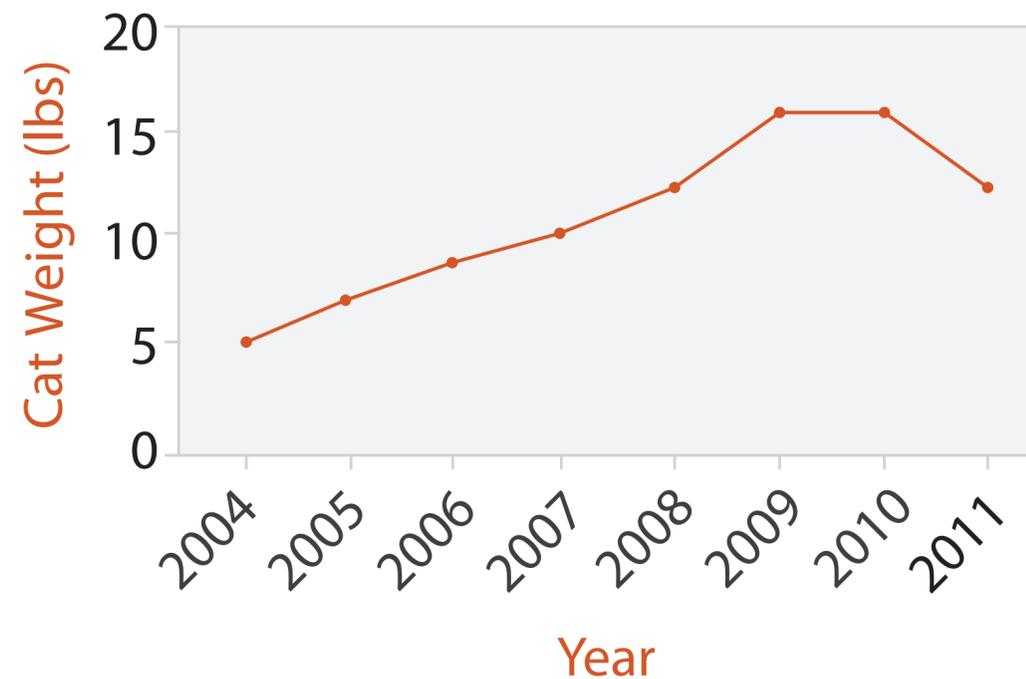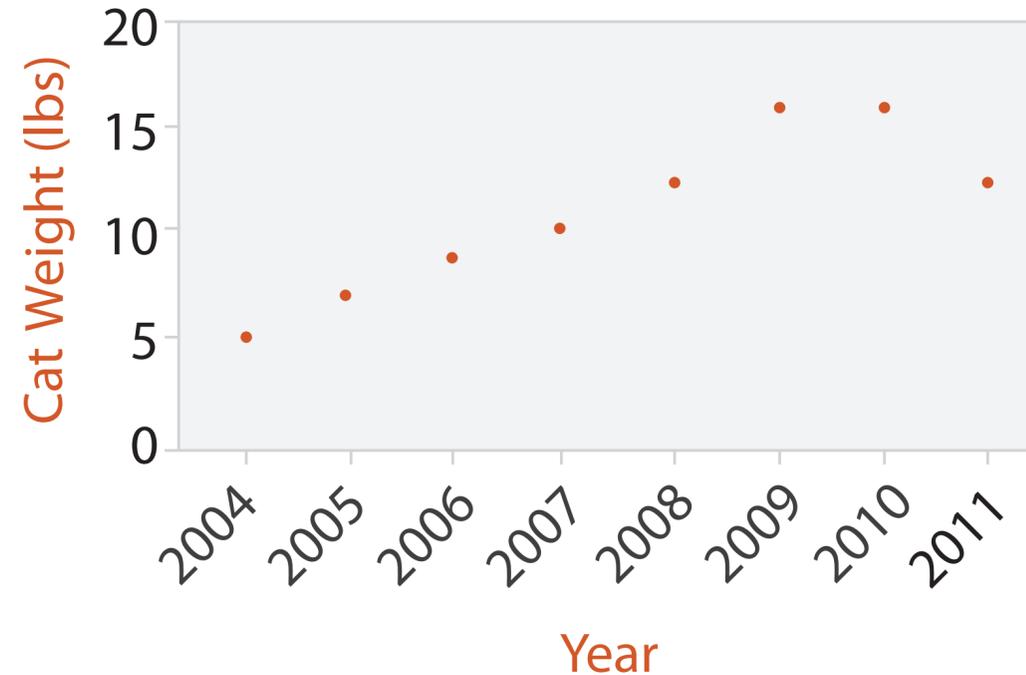


- Data: one quantitative attribute, one **ordered** attribute

- Task: lookup values, find outliers and trends

- How: point mark and positions

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

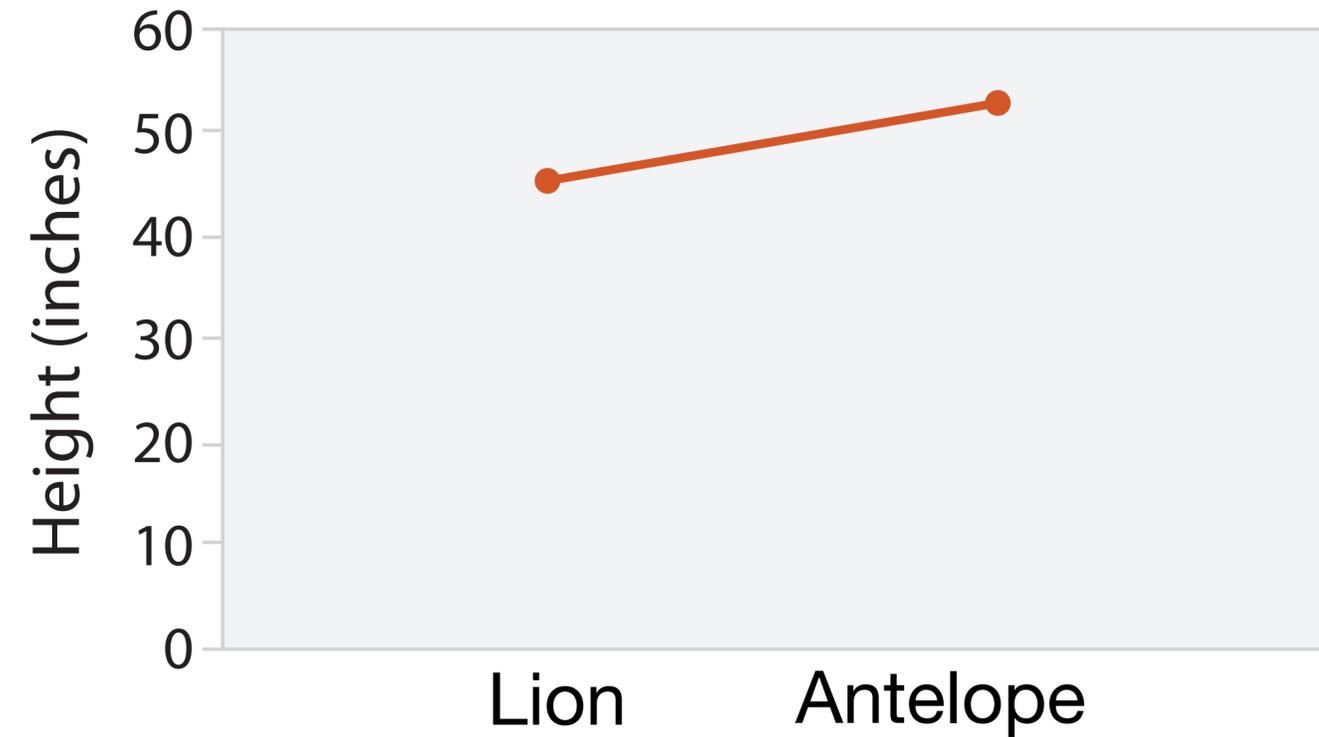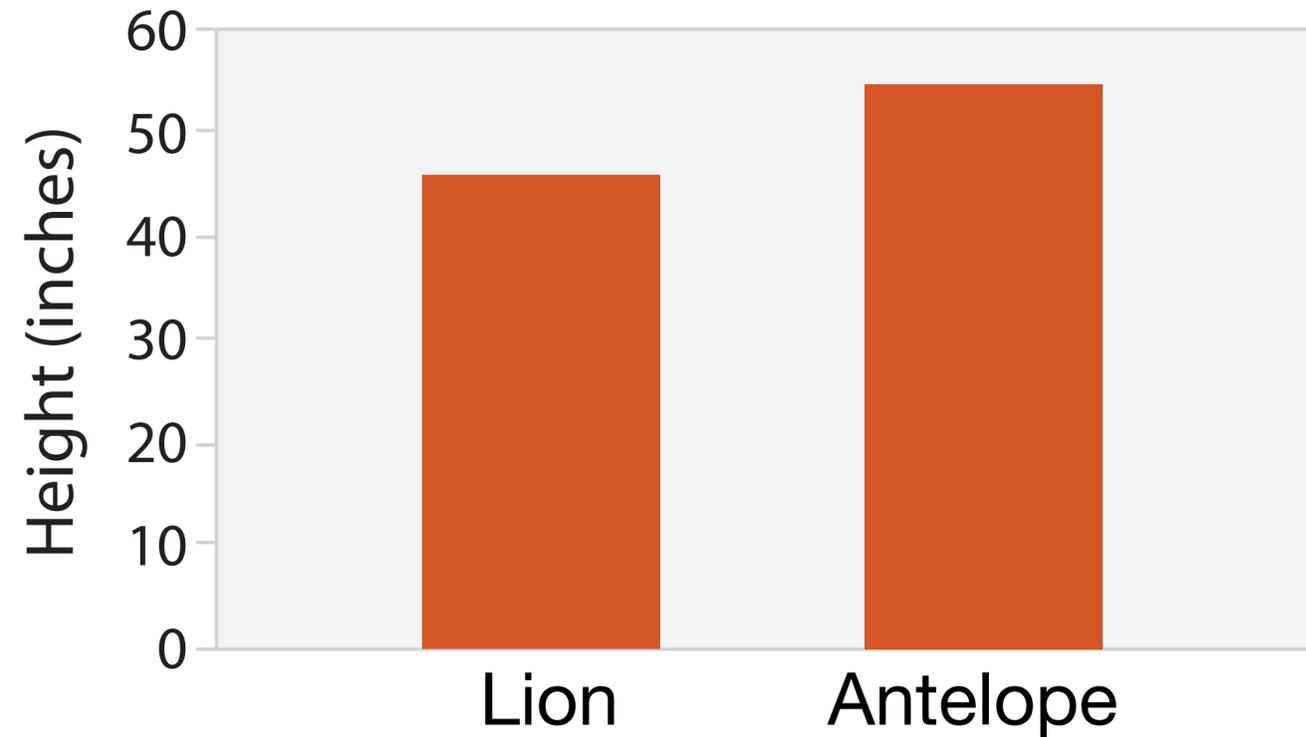- Similar to scatterplots but allow ordered attribute

[Munzner (ill. Maguire), 2014]
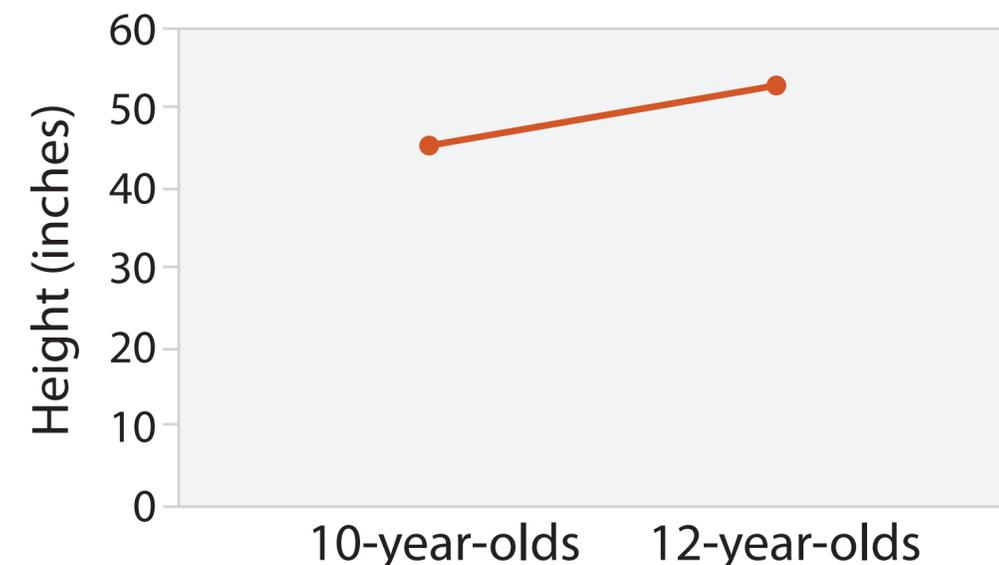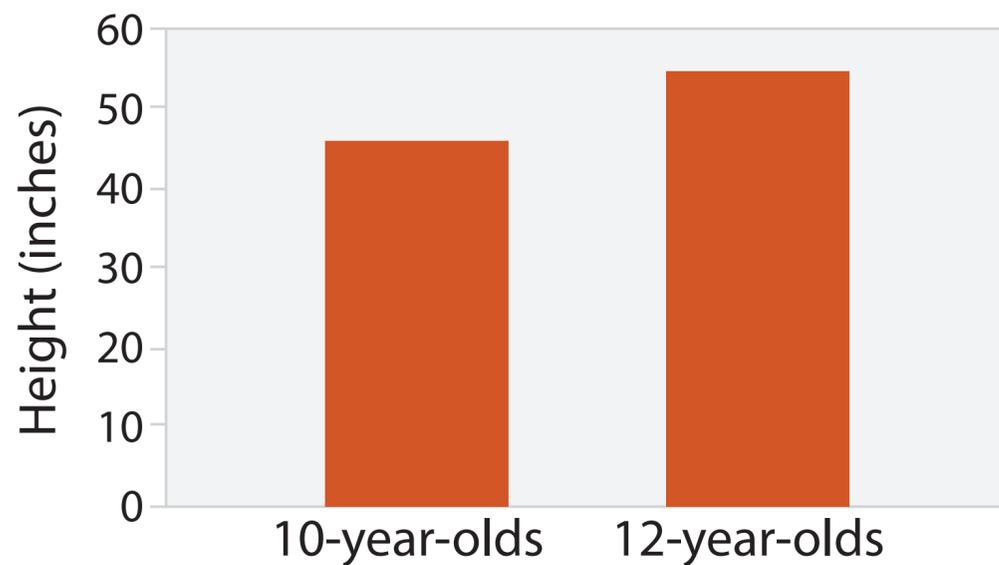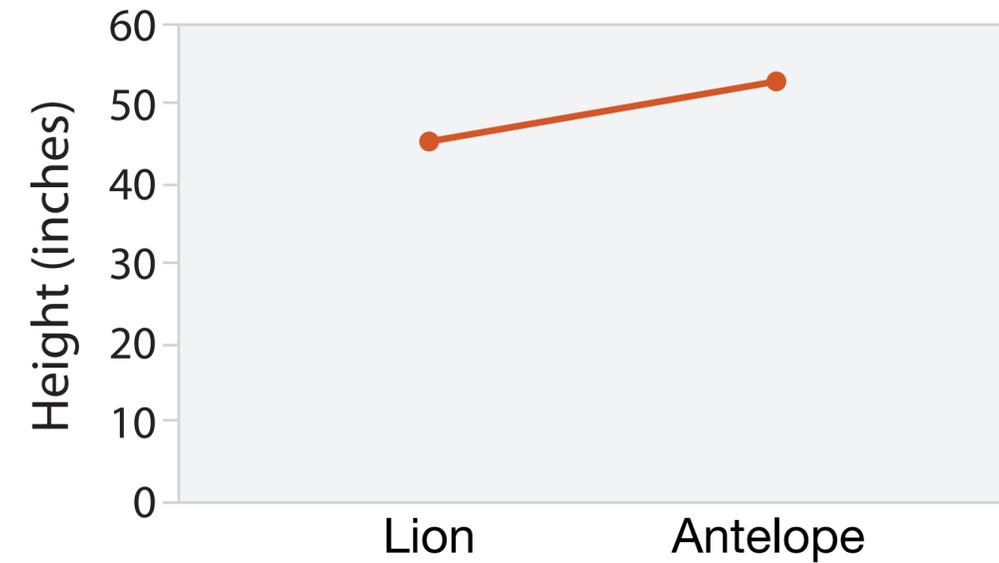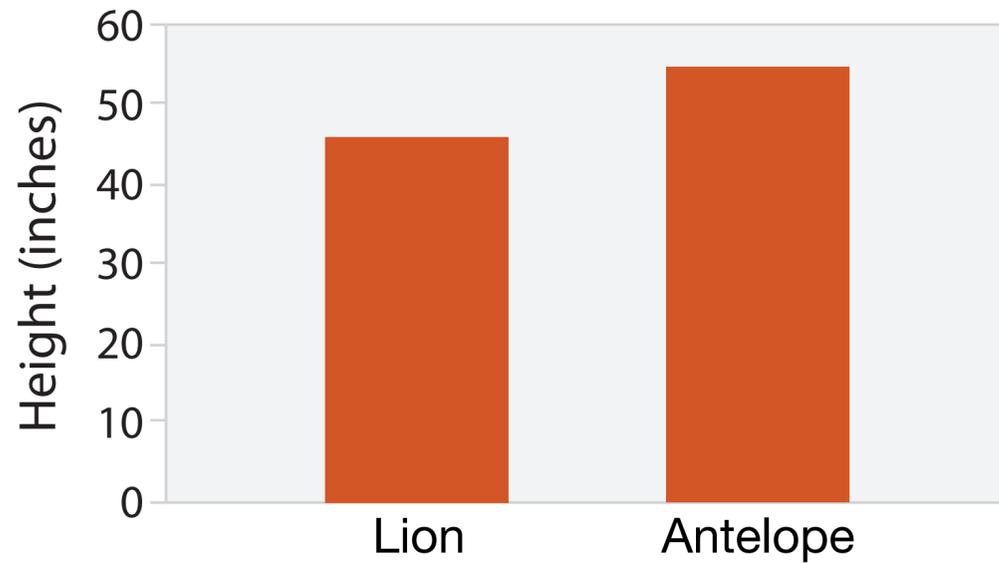
# Proper Use of Line and Bar Charts



[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
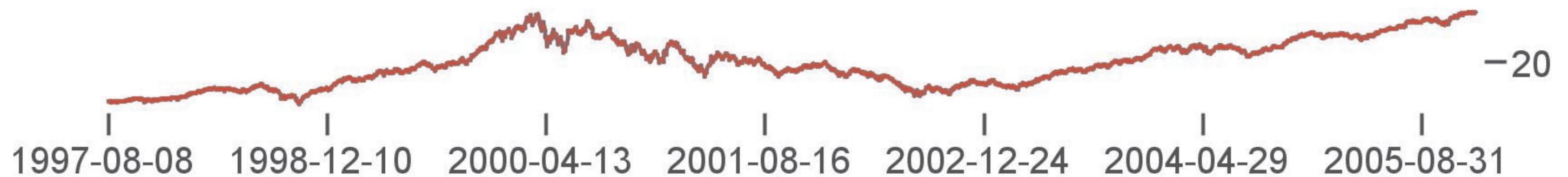
Aspect Ratio = 3.96
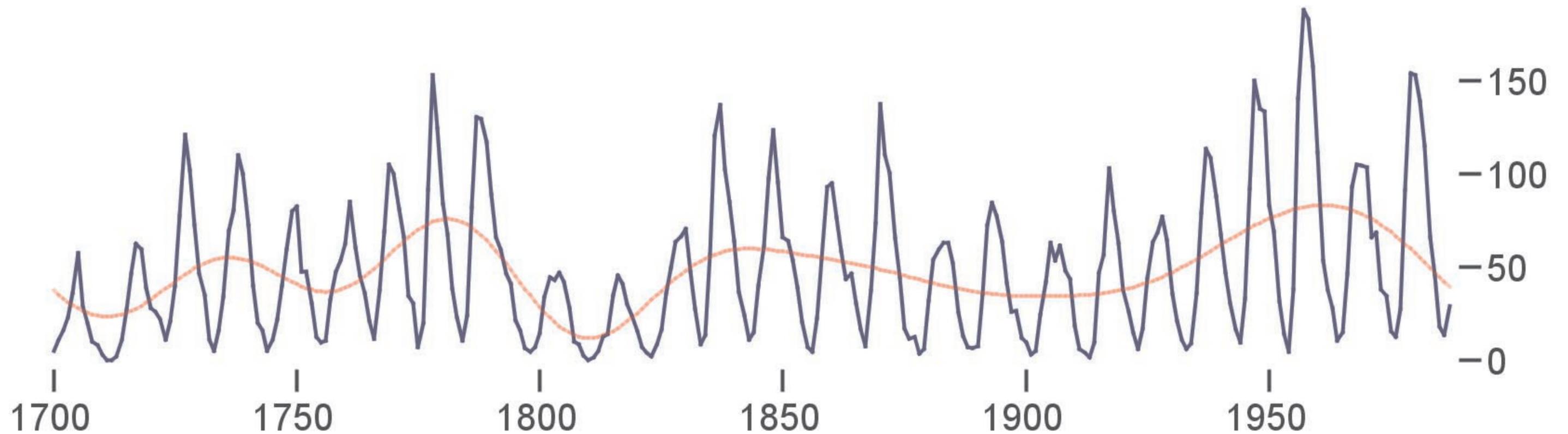


Aspect Ratio = 22.35
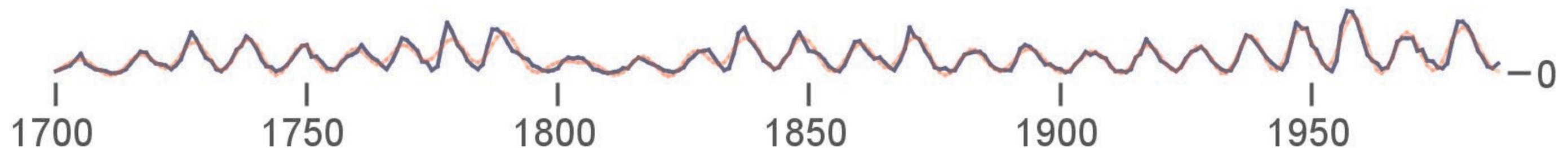


[Heer and Agrawala, 2006]

# Multiscale Banking

Aspect Ratio = 4.23
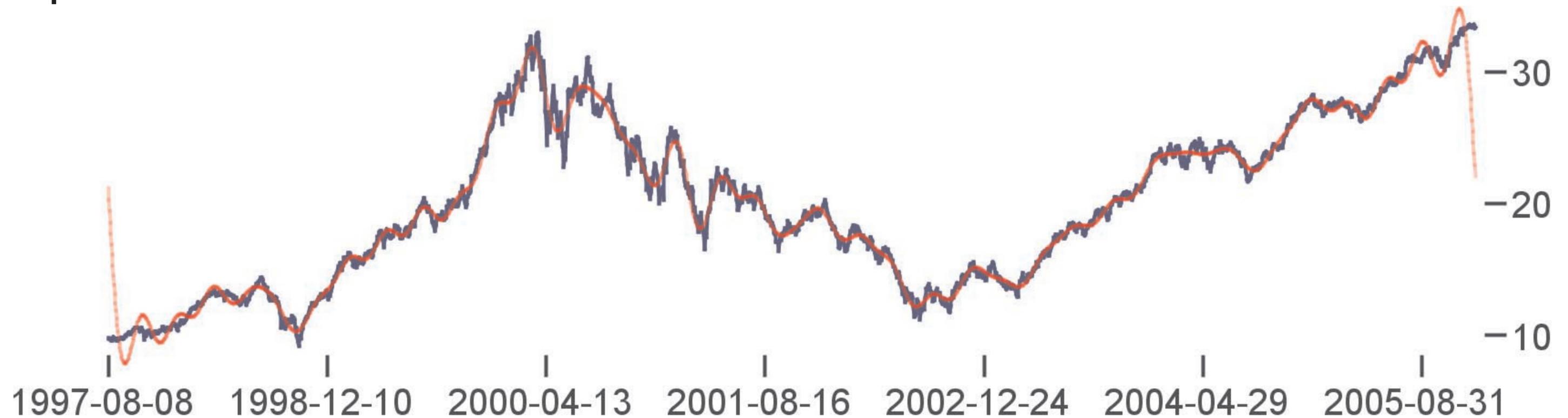


Aspect Ratio = 14.55



[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



Mid−angle ($\theta_m$)

[Talbot et al., 2013]
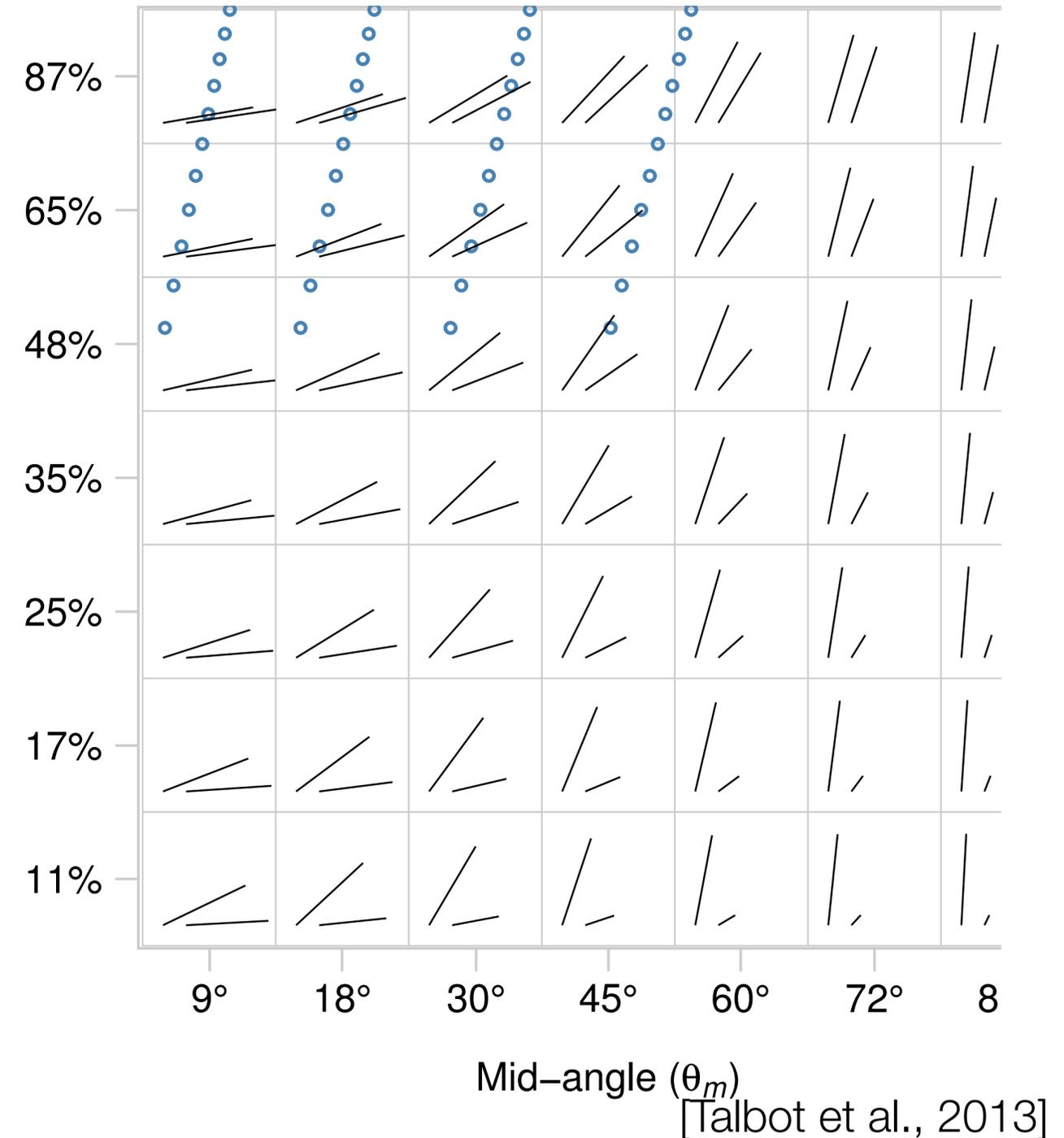
# Heatmaps

- 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)
- Red-green color scales often used
  - Be aware of colorblindness!

Fast-Pitch Softball
Slugging Percentage
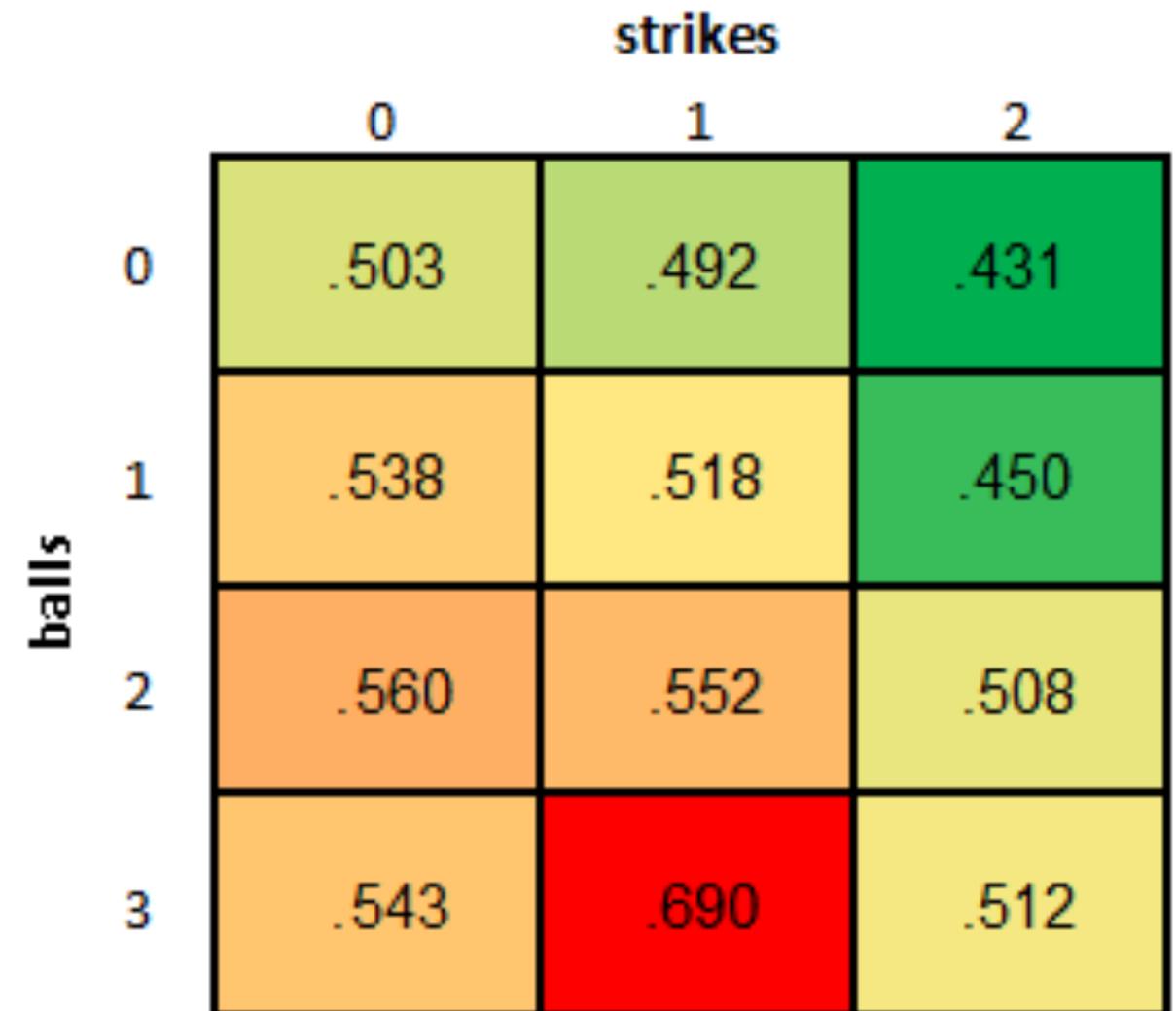


[fastpitchanalytics.com]

# Bertin Matrices

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

[C.Perrin et al., 2014]

Northern Illinois University

# Bertin Matrices

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



[C.Perrin et al., 2014]

# Bertin's Encodings



| Text | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | text |
|------|---|---|---|---|---|---|---|---|---|---|----|----|------|
| Grayscale | | | | | | | | | | | | | |
| Circle | | | | | | | | | | | | | |
| Dual bar chart | | | | | | | | | | | | | |
| Bar chart | | | | | | | | | | | | | |
| Line | | | | | | | | | | | | | |
| Black and white bar chart | | | | | | | | | | | | | |
| Average bar chart | | | | | | | | | | | | | |

QUANTITY OF INK ENCODINGS

POSITIONAL ENCODING

MEAN-BASED ENCODINGS

[C.Perrin et al., 2014]

# Matrix Reordering



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

# Cluster Heatmap
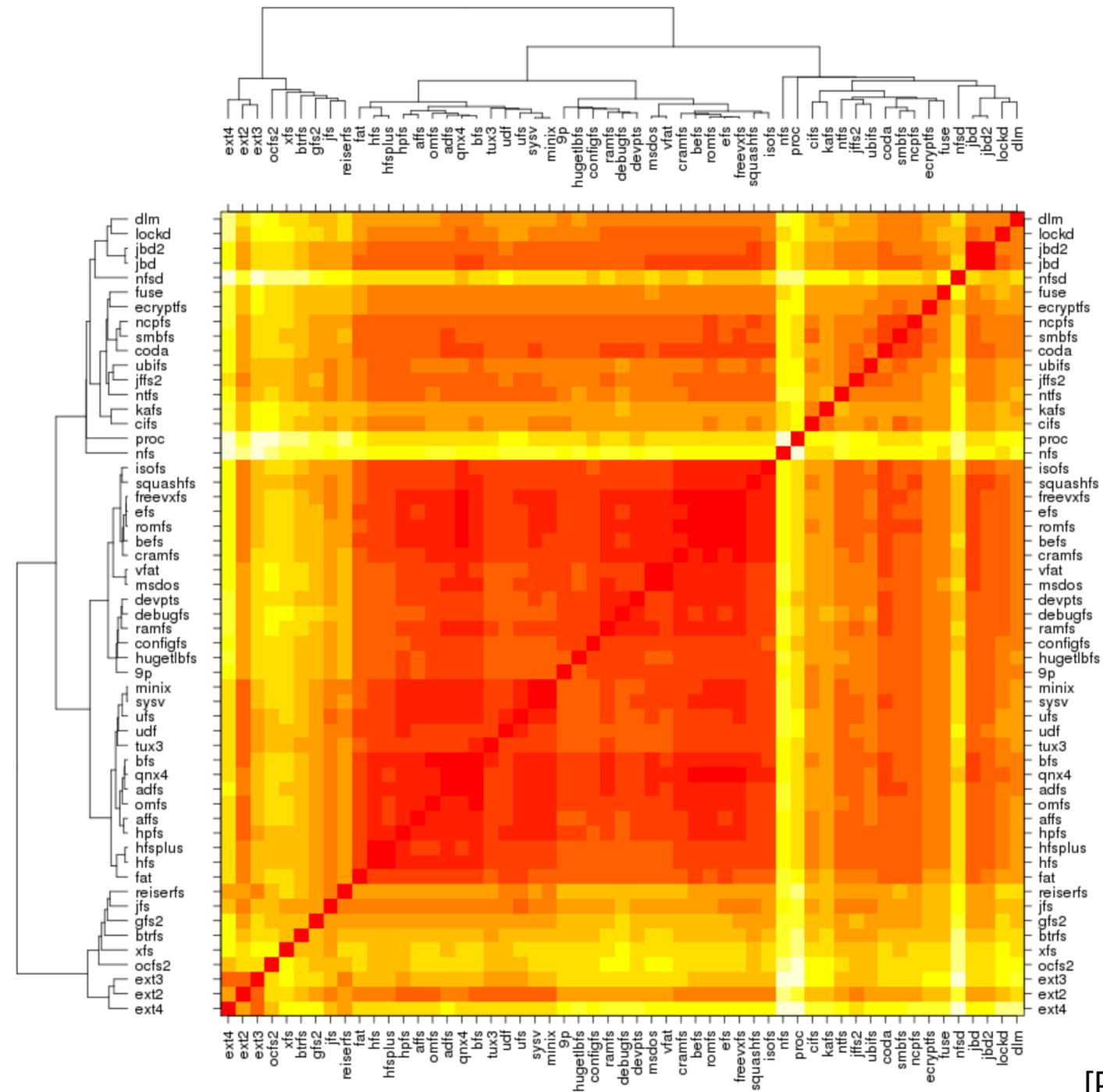


[File System Similarity, R. Musăloiu-E., 2009]

# 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…
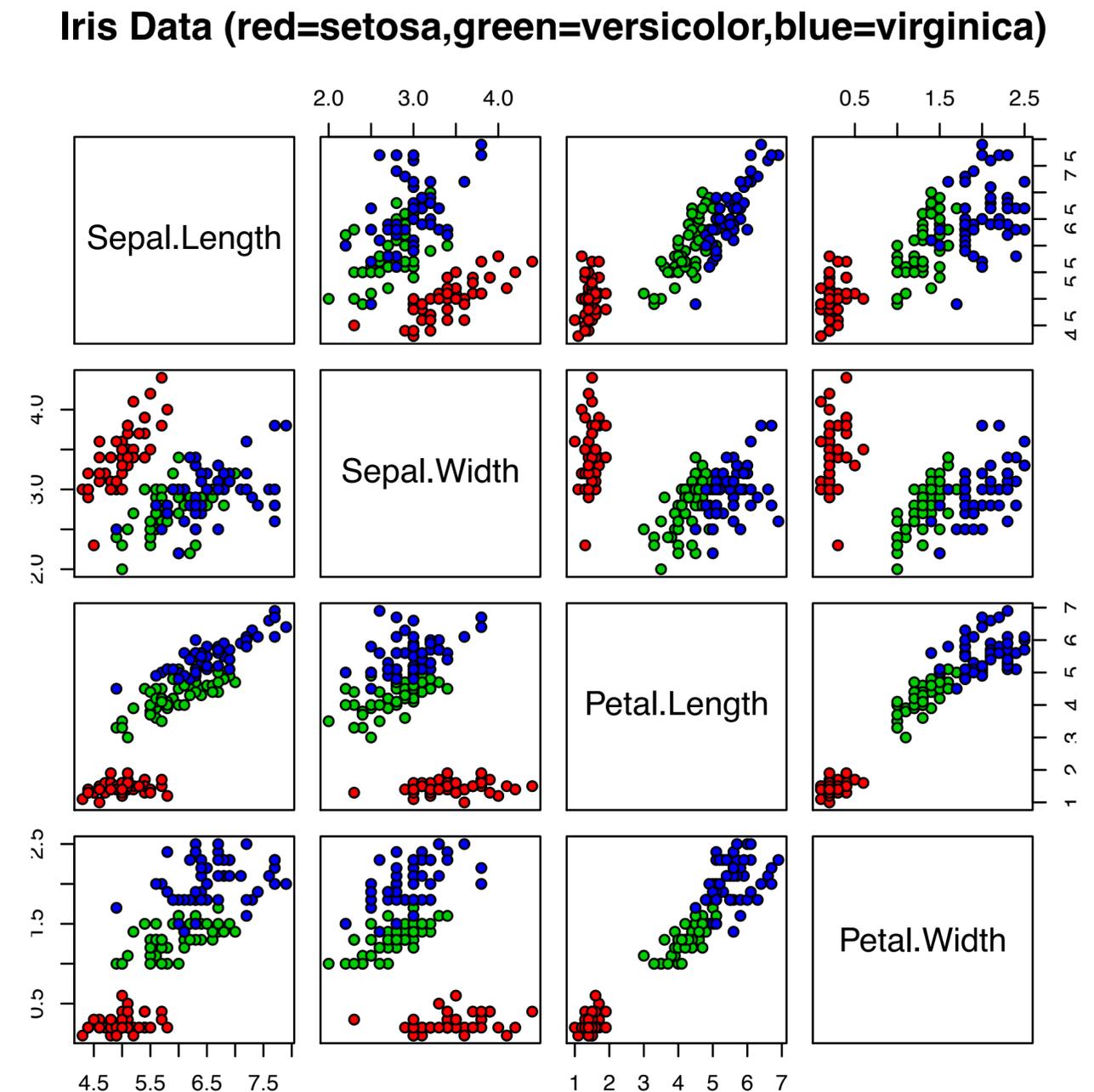


**Iris Data (red=setosa,green=versicolor,blue=virginica)**

# 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?

# 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
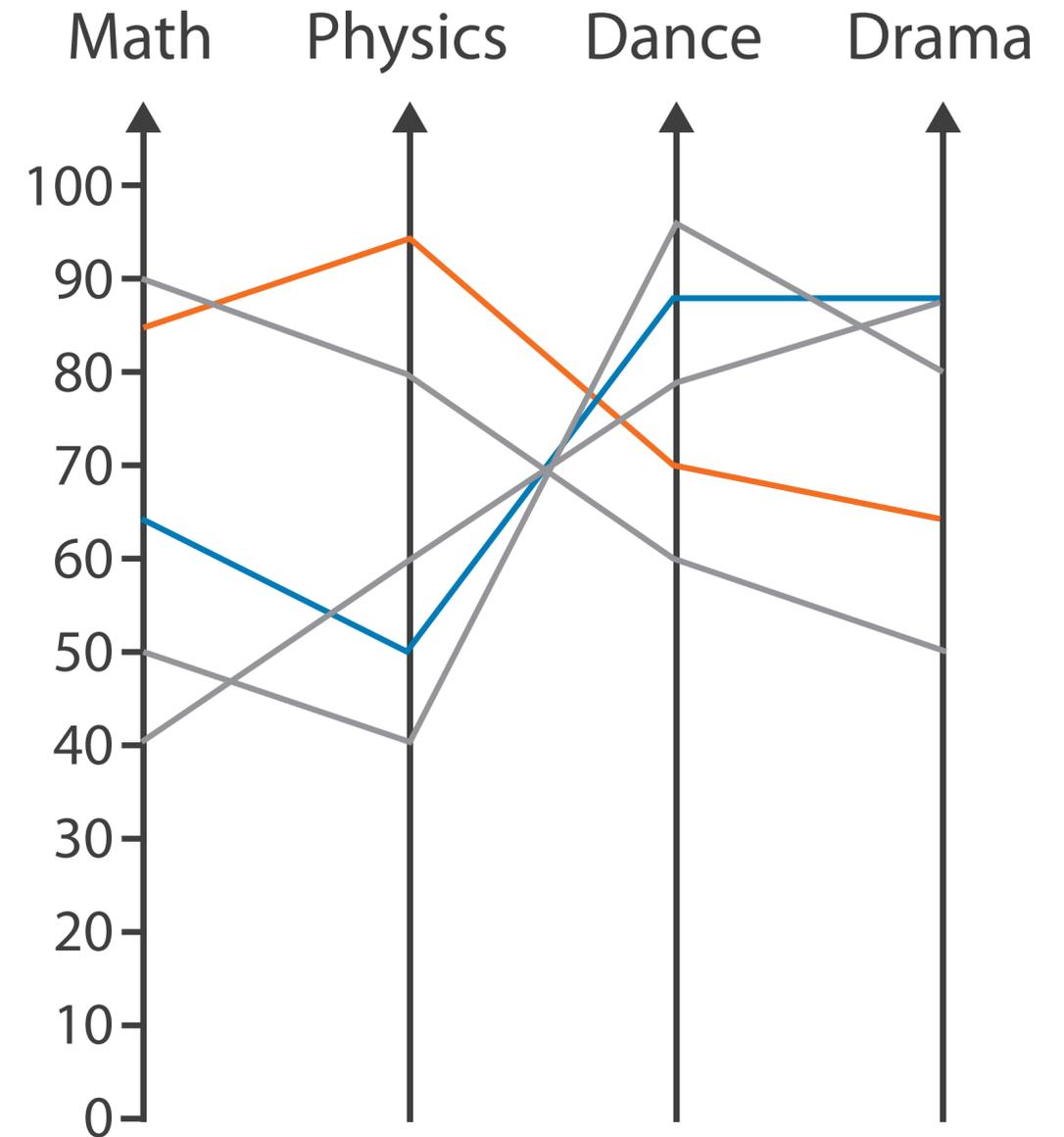


[Munzner (ill. Maguire), 2014]

# Radial Axes
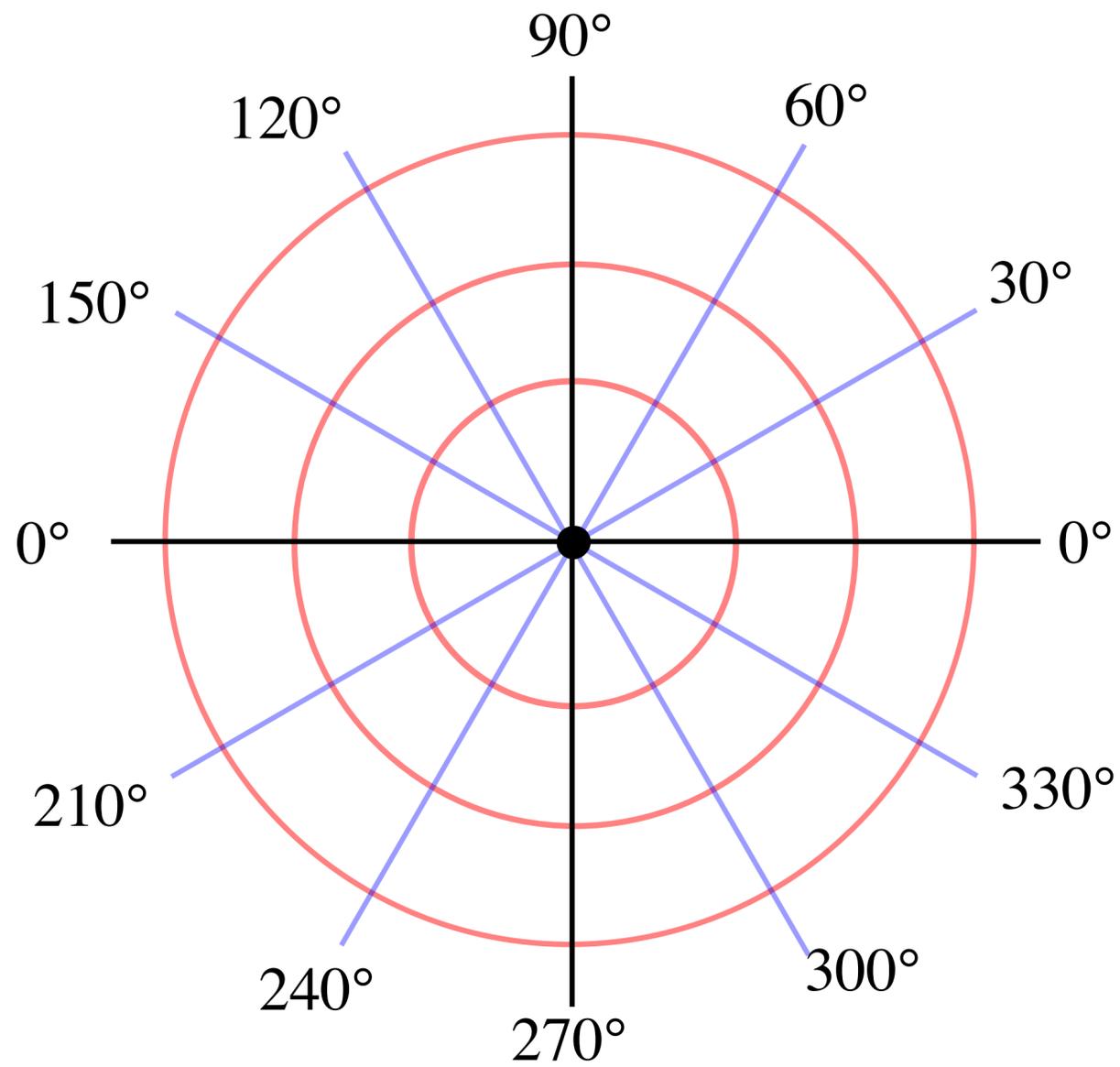
# Radial Axes
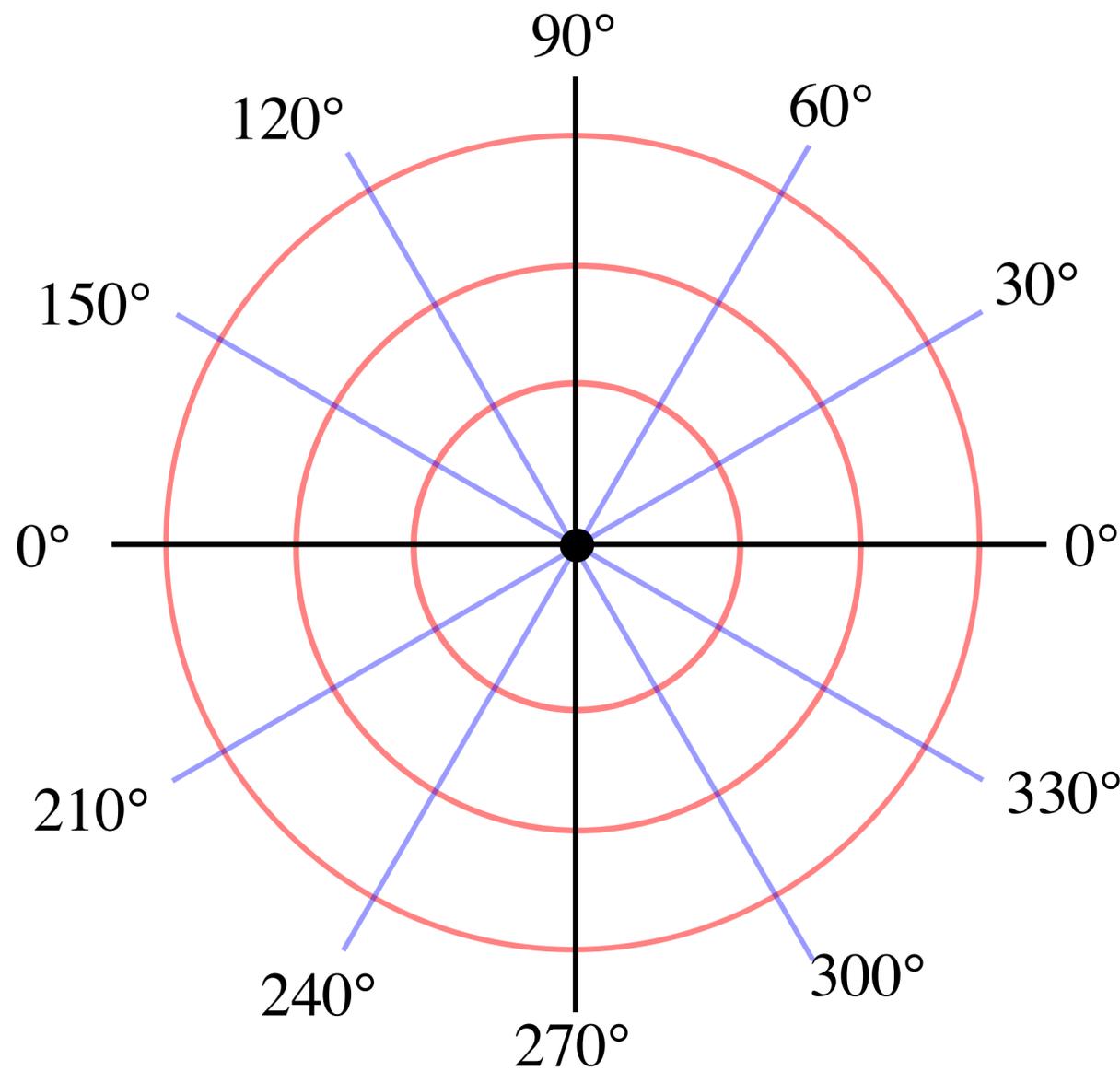


- Polar Coordinates (angle + position along the line at that angle)
- What types of encodings are possible for tabular data in polar coordinates?
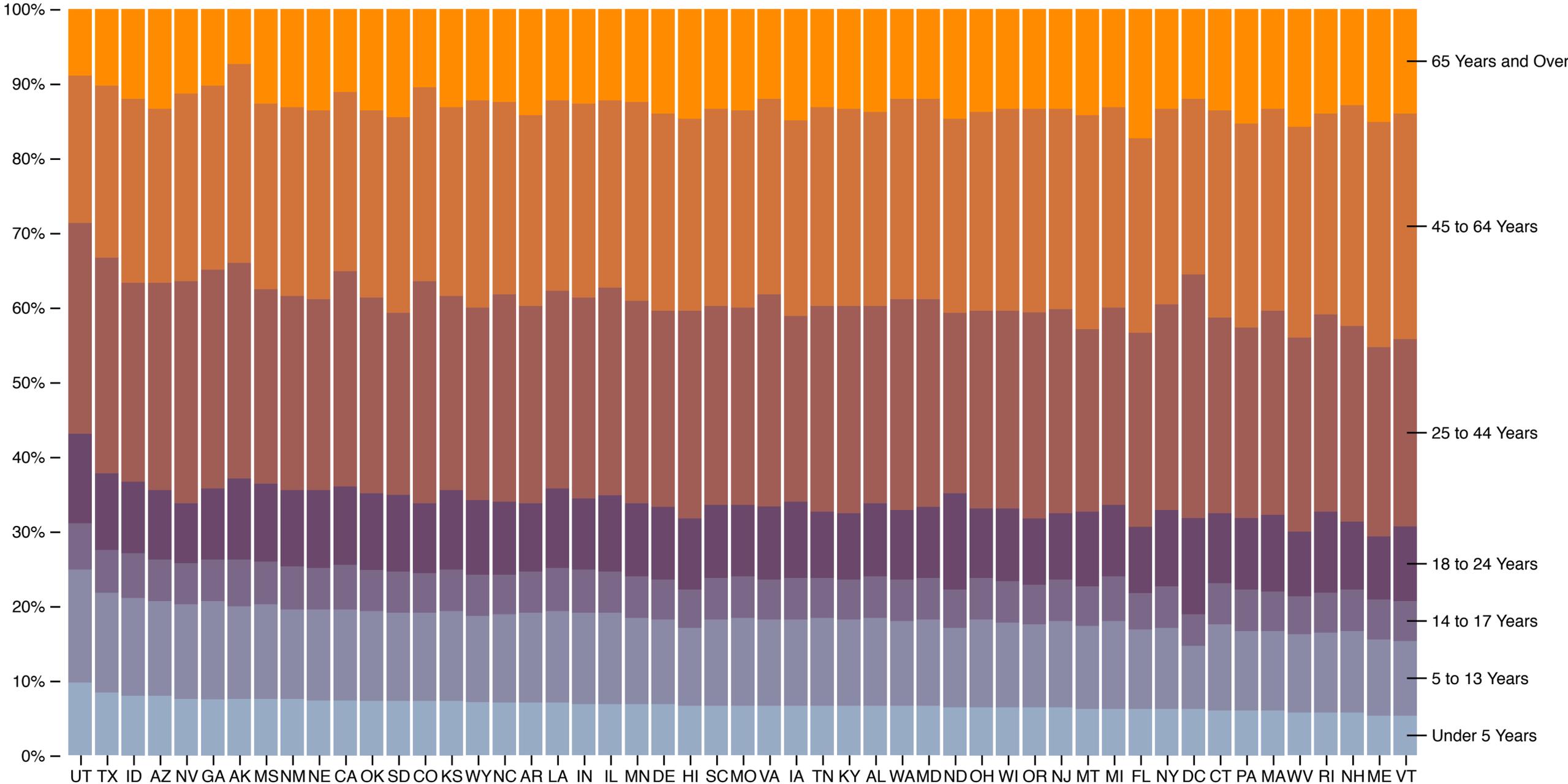
# Radial Axes
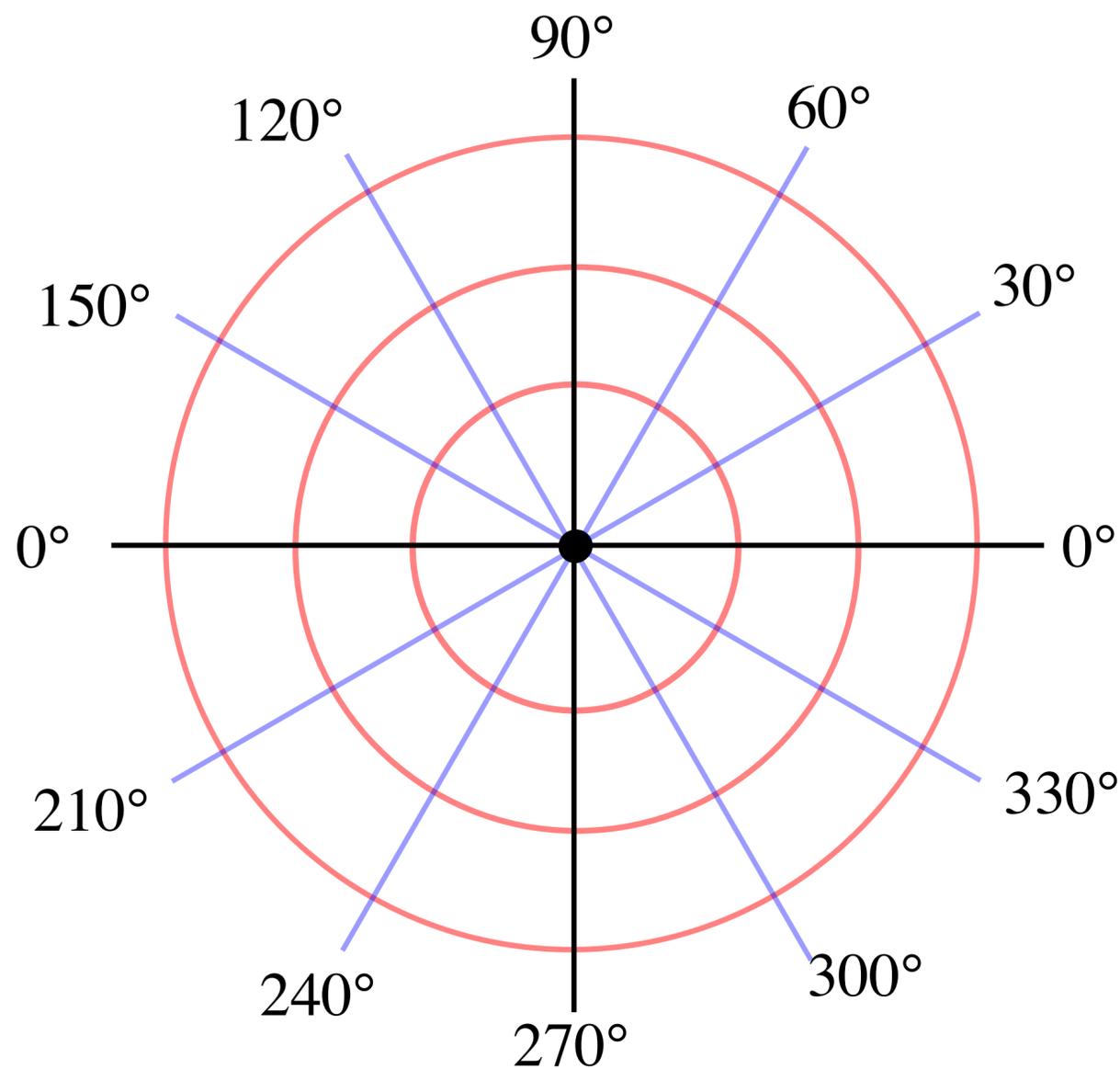


- Polar Coordinates (angle + position along the line at that angle)
- What types of encodings are possible for tabular data in polar coordinates?
  - Radial bar charts
  - Pie charts
  - Donut charts

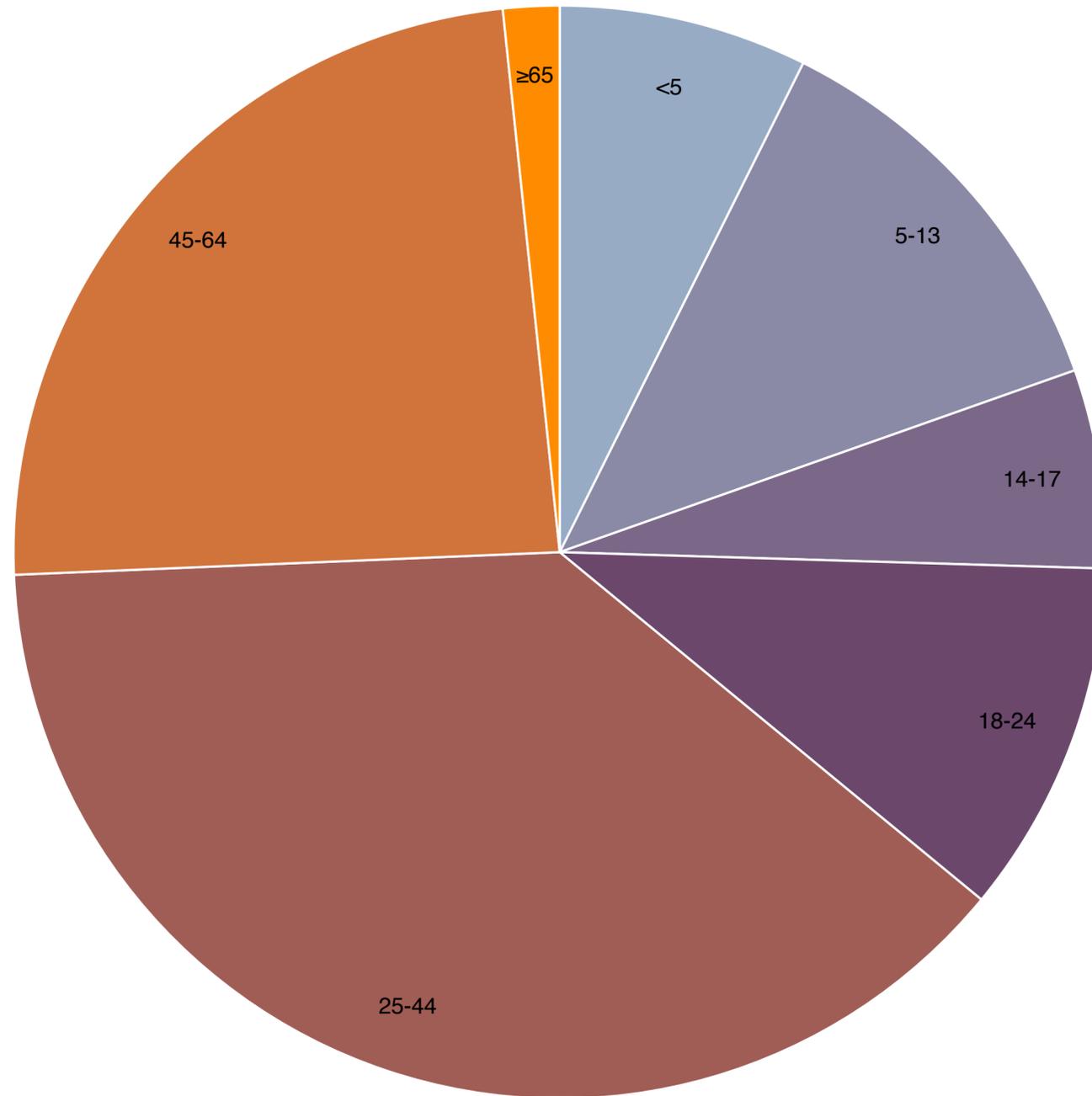# Part-of-whole: Relative % comparison?
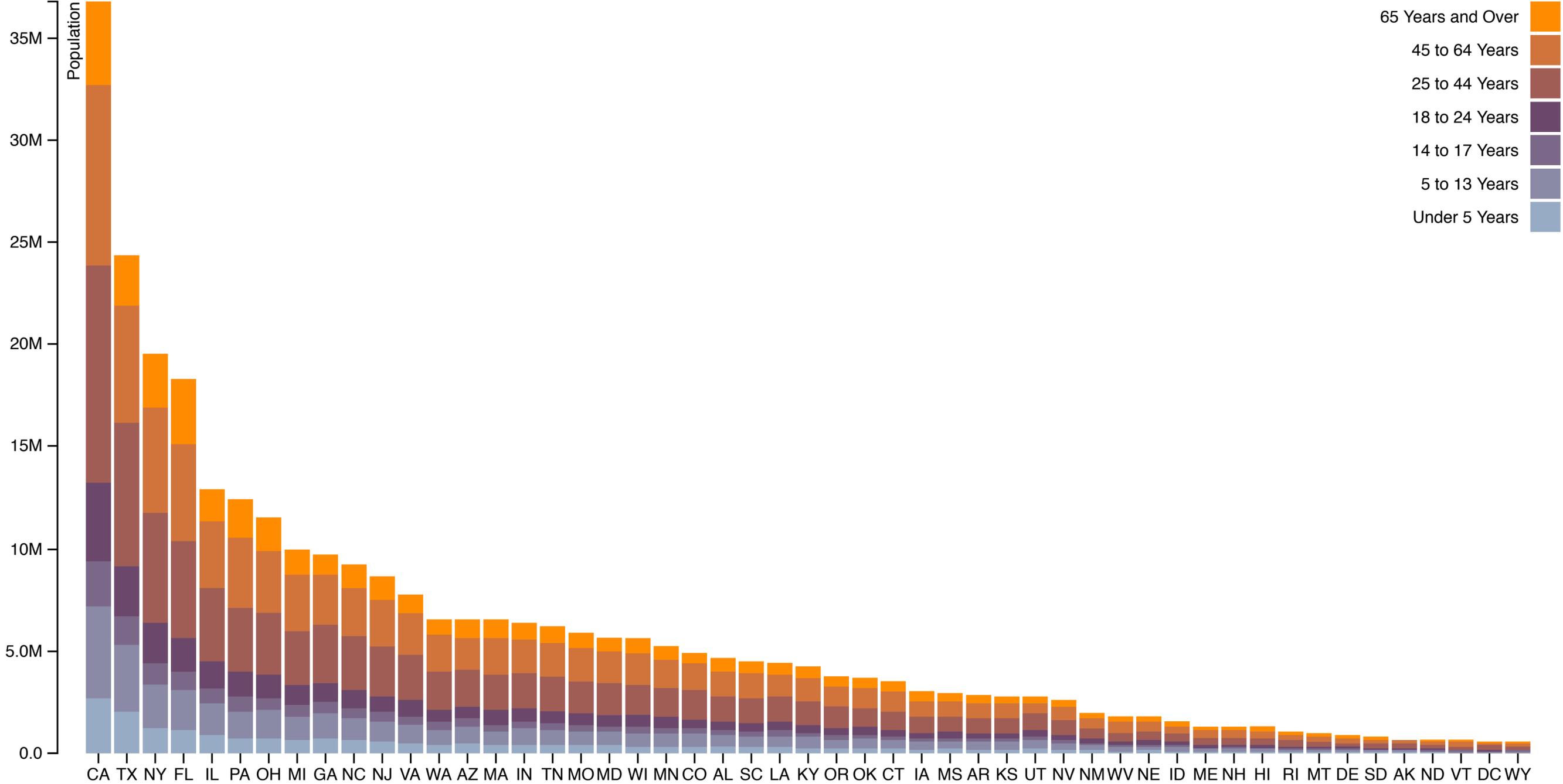


[Stacked Bar Chart, M. Bostock, 2017]

# 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
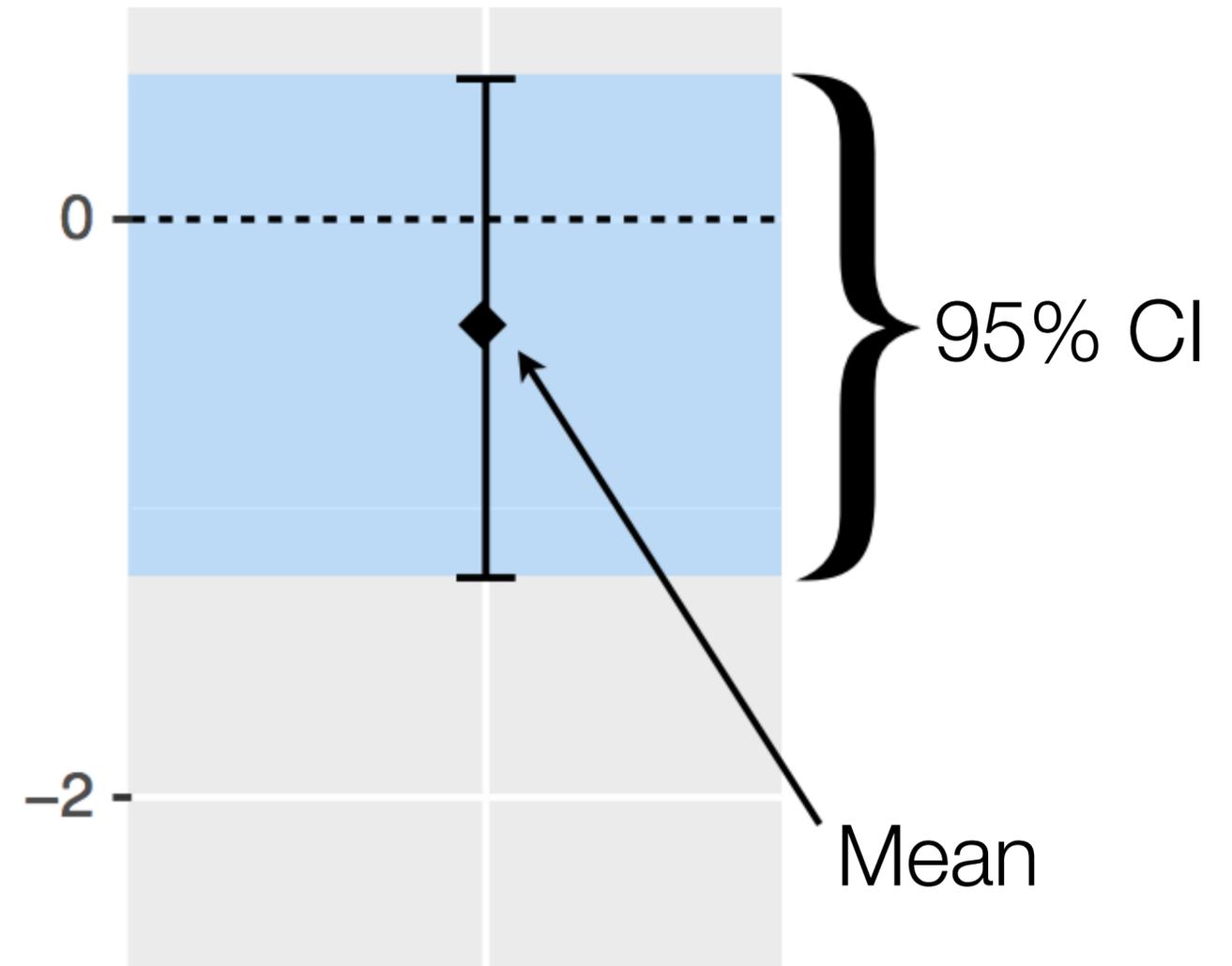  - Summary: "An Illustrated Study of the Pie Chart Study Results"

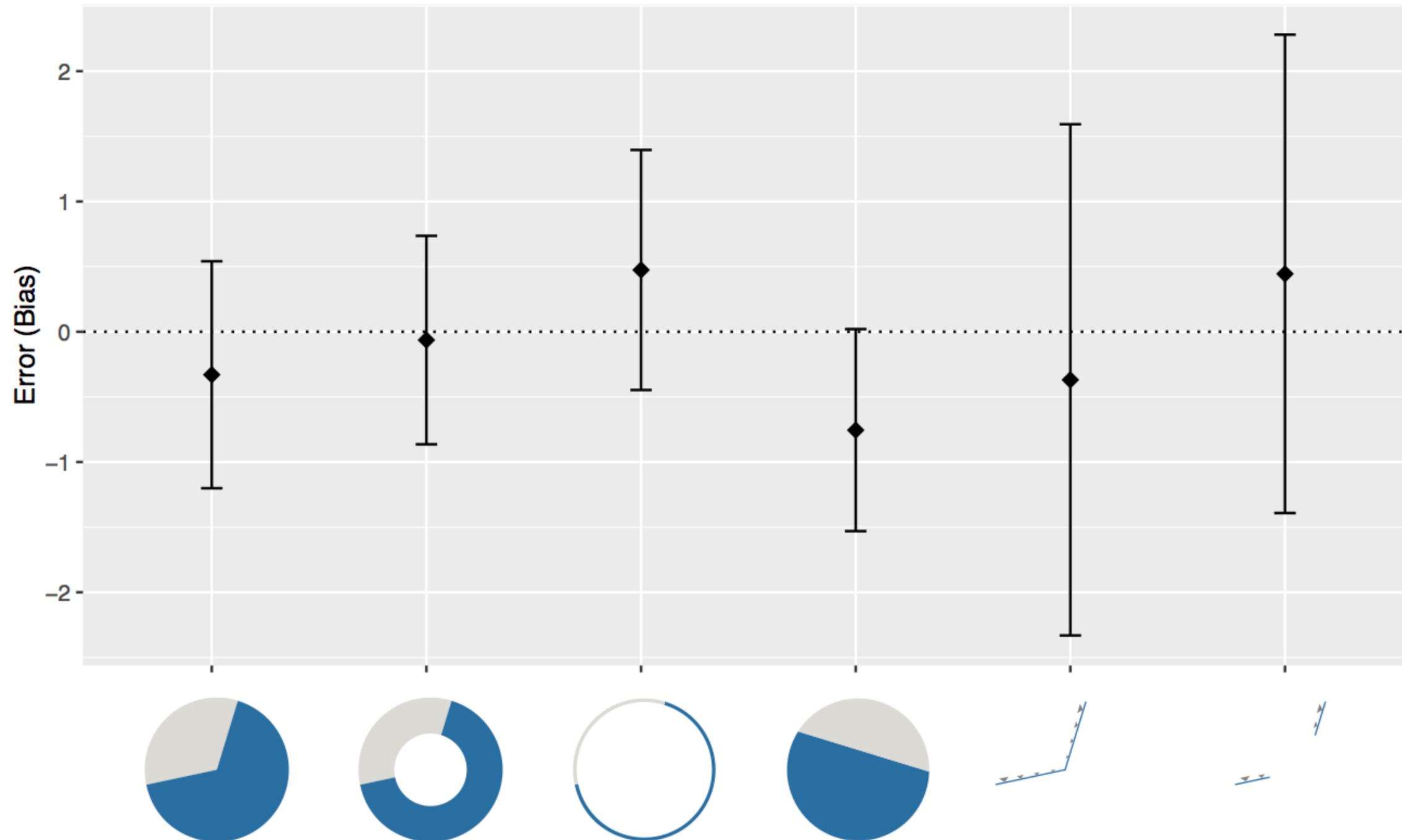# Arcs, Angles, or Areas?
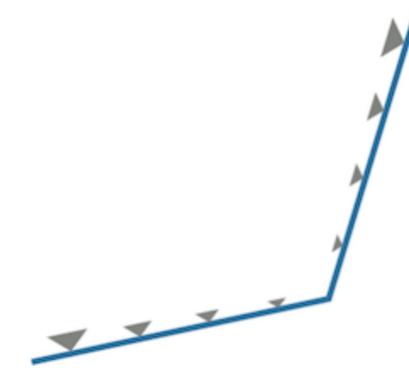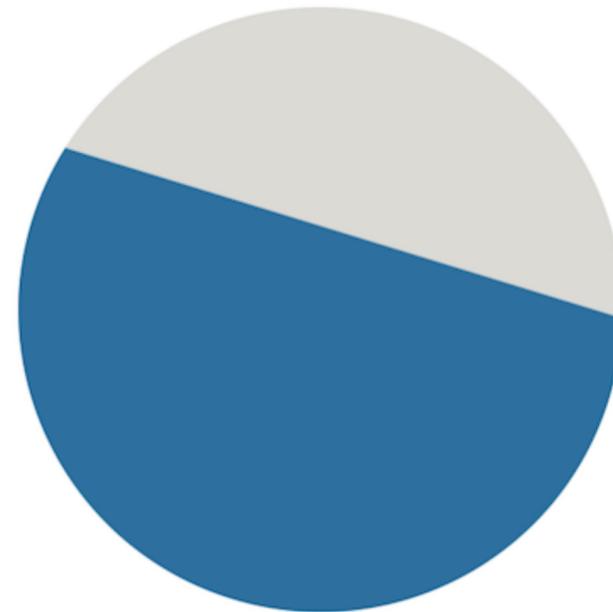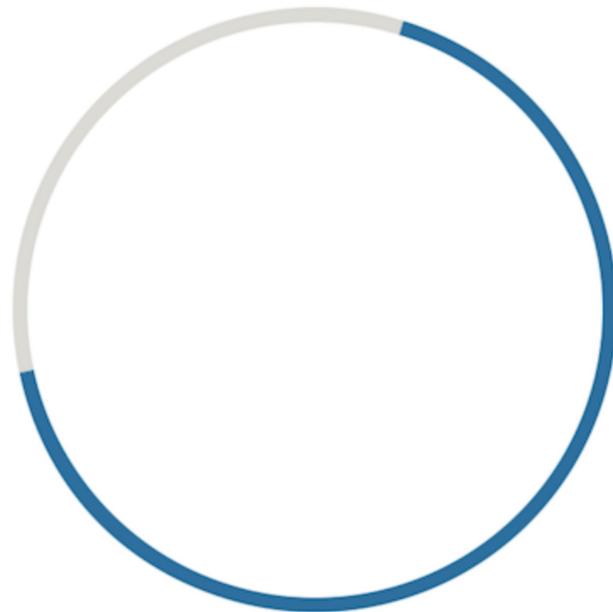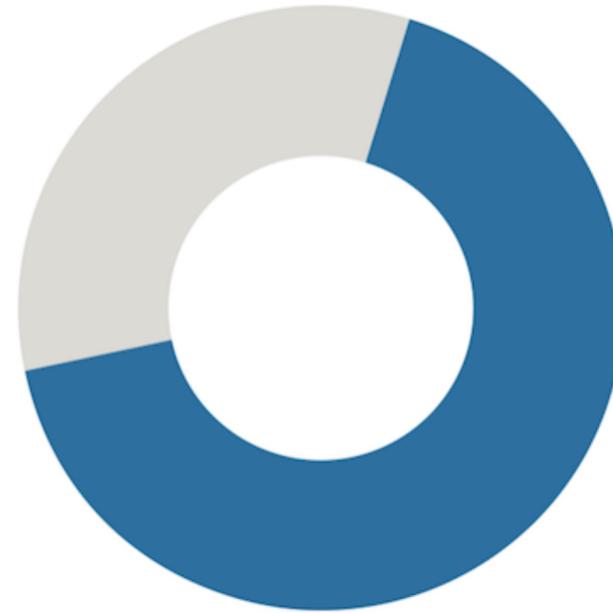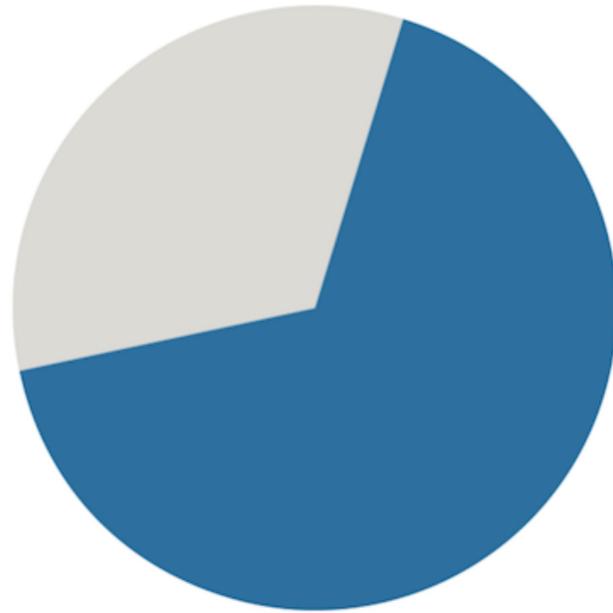


[R. Kosara and D. Skau, 2016]

# Study Setup

- Three studies

- 80-100 participants each

- Each answered ~60 questions

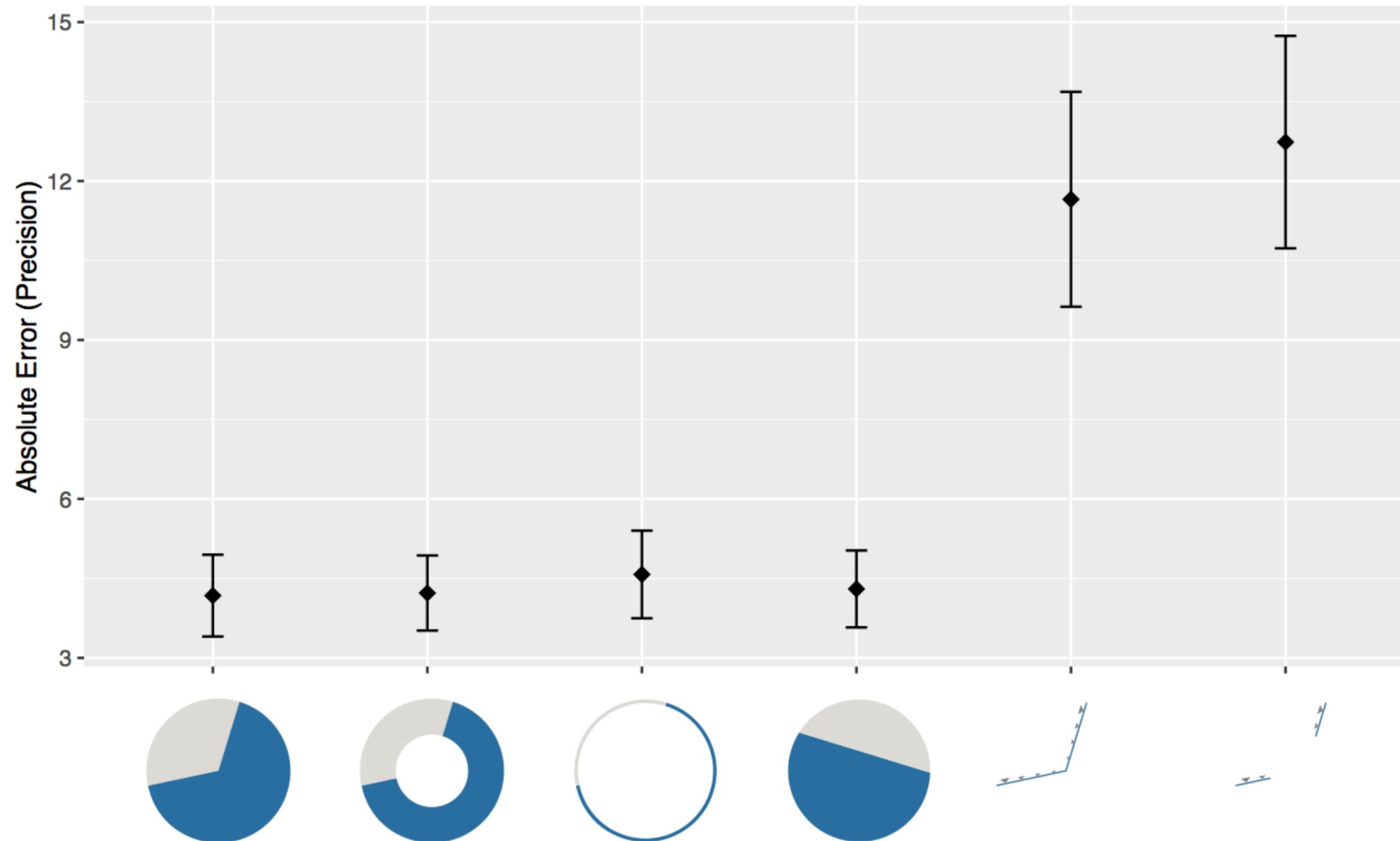- Computed results using 95% Confidence Intervals
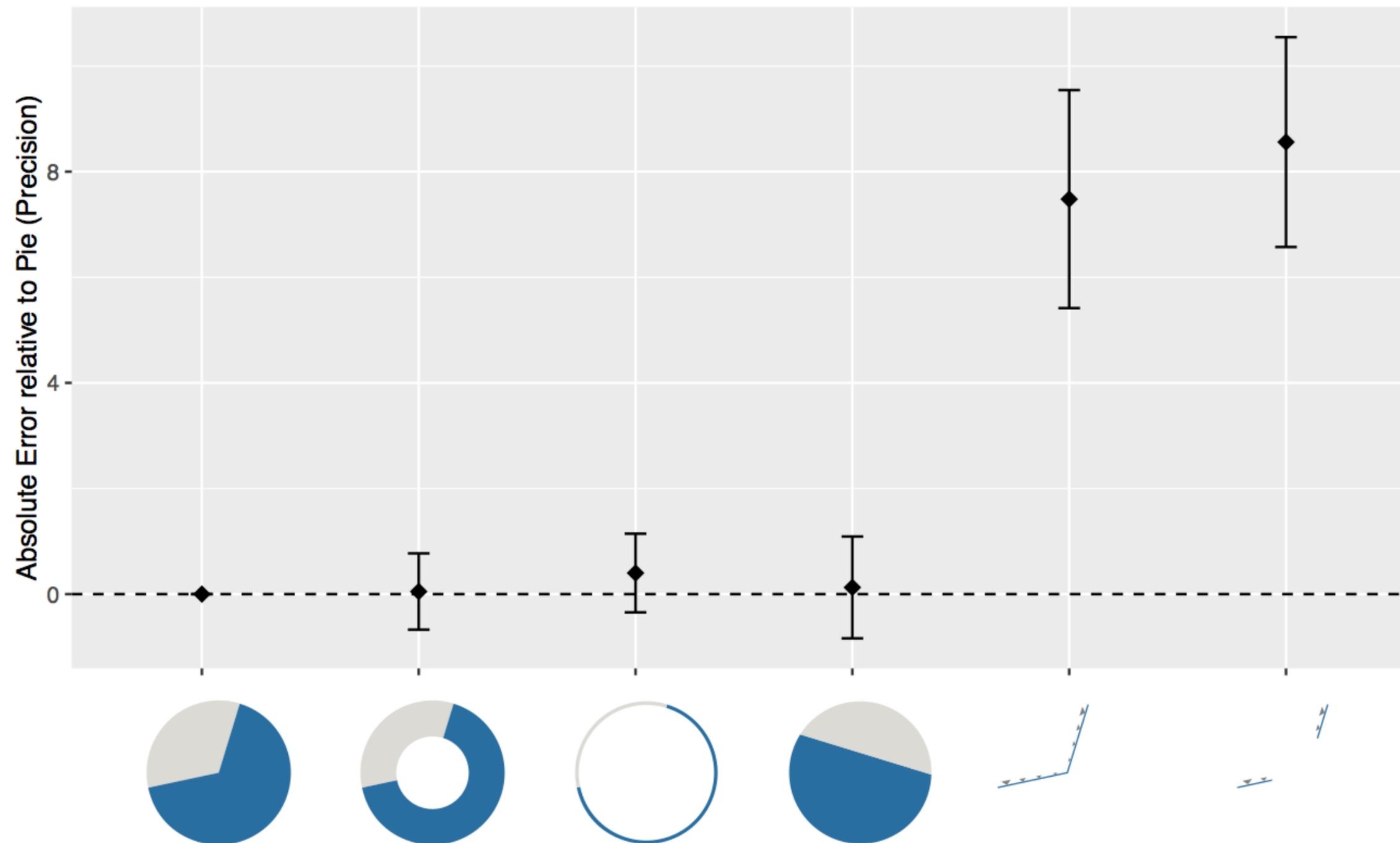


0

95% CI

−2

Mean

[R. Kosara and D. Skau, 2016]

# Signed Error

# Absolute Error
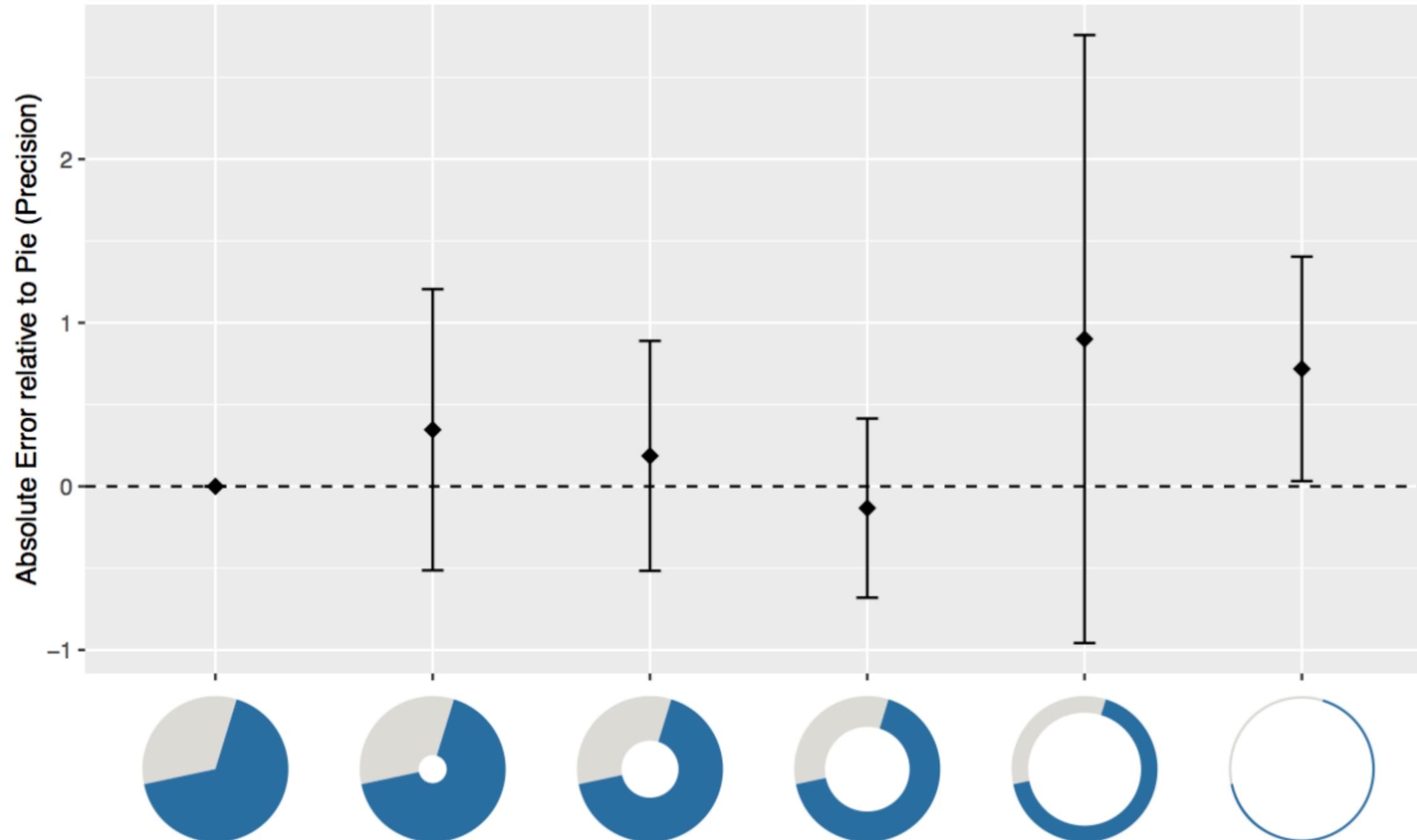
# Absolute Error Relative to Pie Chart
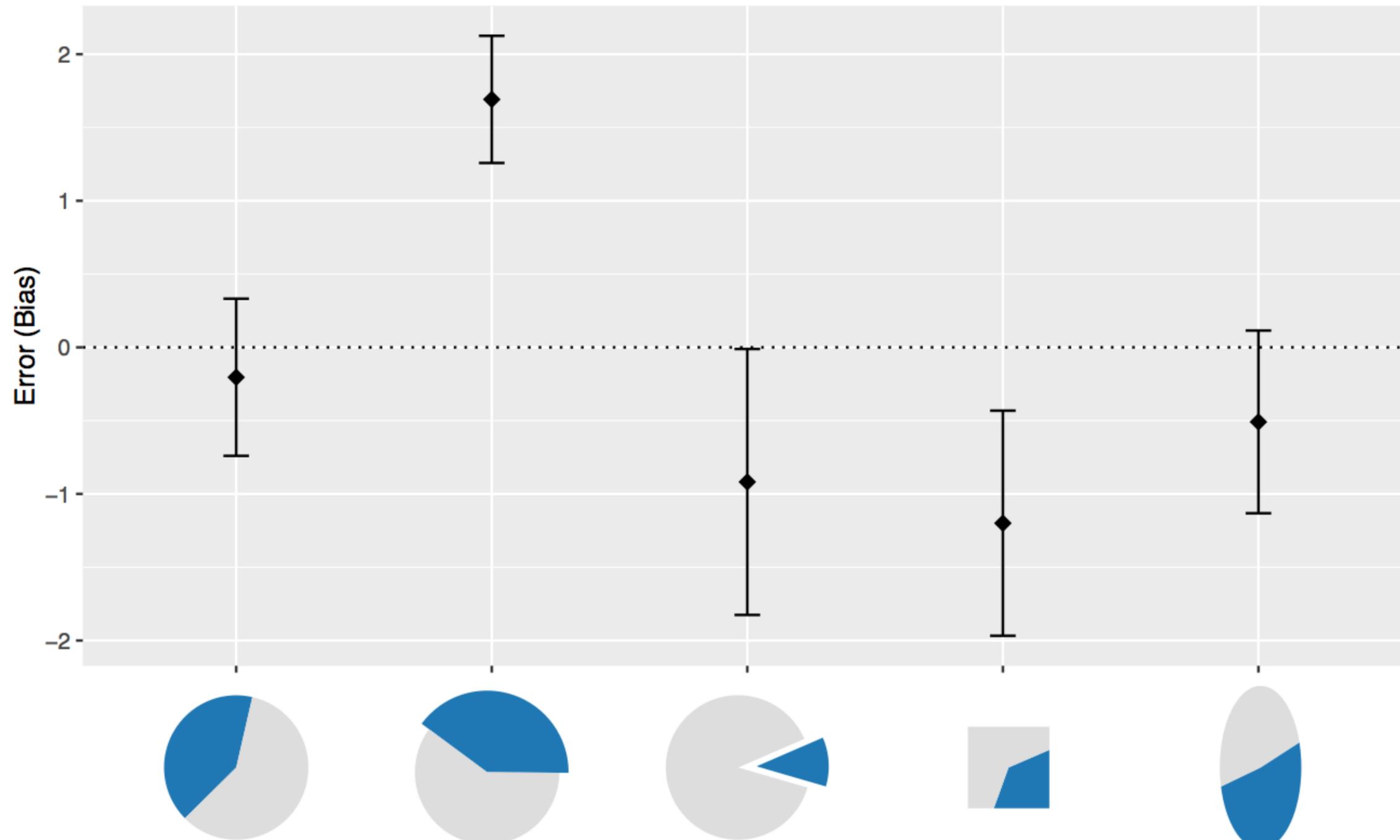


[R. Kosara and D. Skau, 2016]
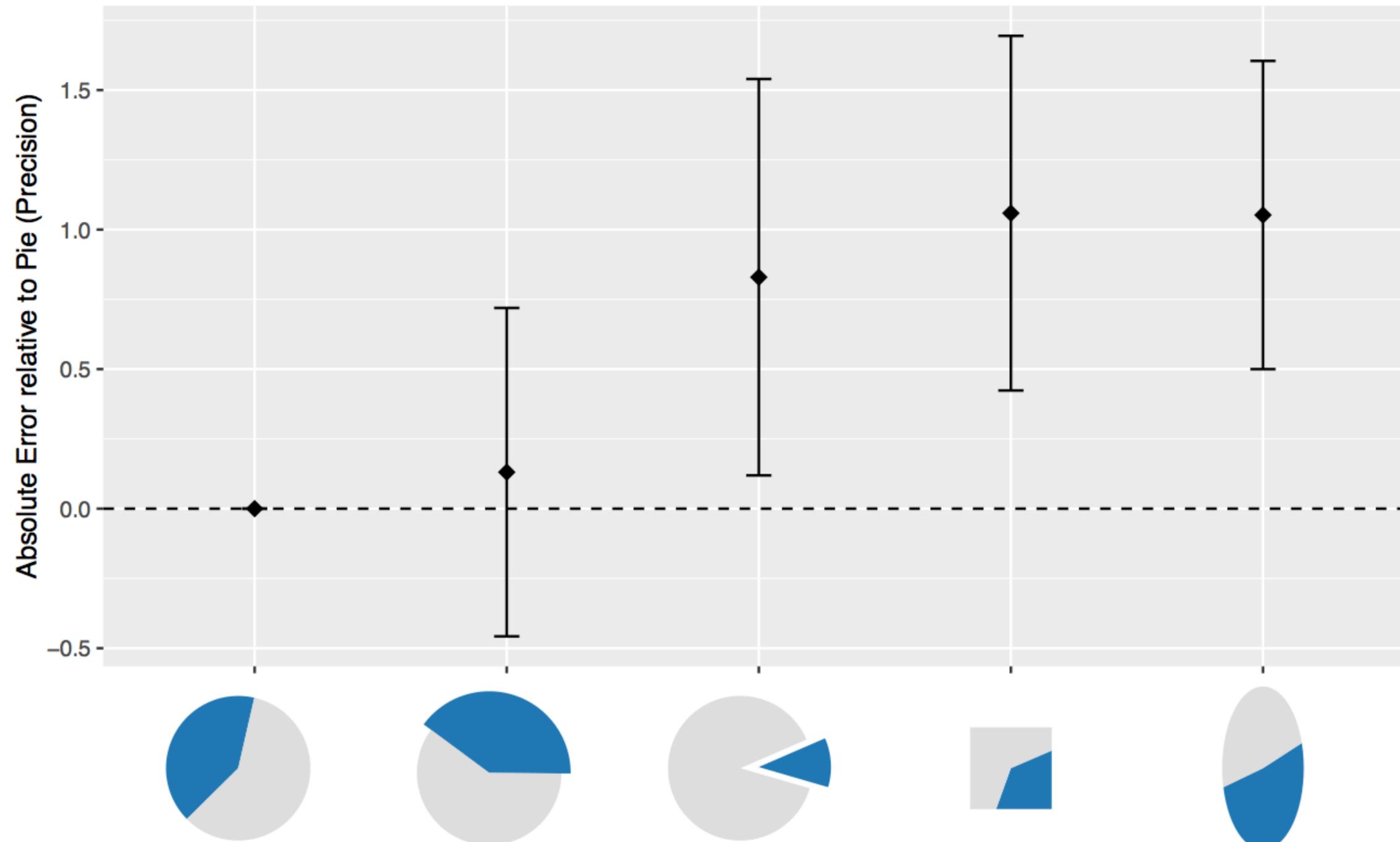
# Donut Charts Width



[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