Data Visualization (CSCI 627/490)

Tabular Data

Dr. David Koop

Visual Encoding

- How do we encode data visually?
  - **Marks** are the basic graphical elements in a visualization
  - **Channels** are ways to control the appearance of the marks
- Marks classified by dimensionality:

Points **Lines** **Areas**

- Also can have surfaces, volumes
- Think of marks as a mathematical definition, or if familiar with tools like Adobe Illustrator or Inkscape, the path & point definitions

Channel Types

- Identity => what or where, Magnitude => how much

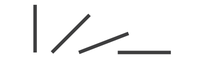
Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle



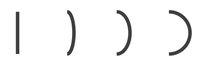
Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature



Volume (3D size)

Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

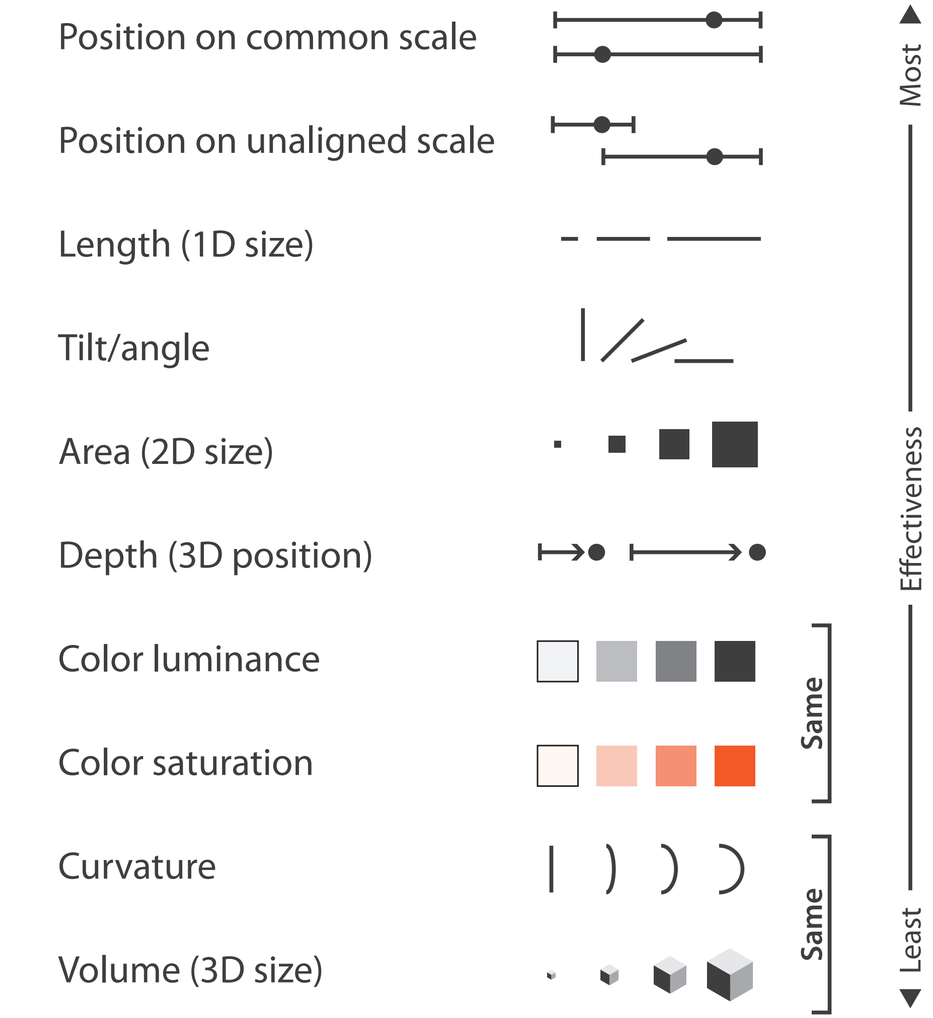
[Munzner (ill. Maguire), 2014]

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

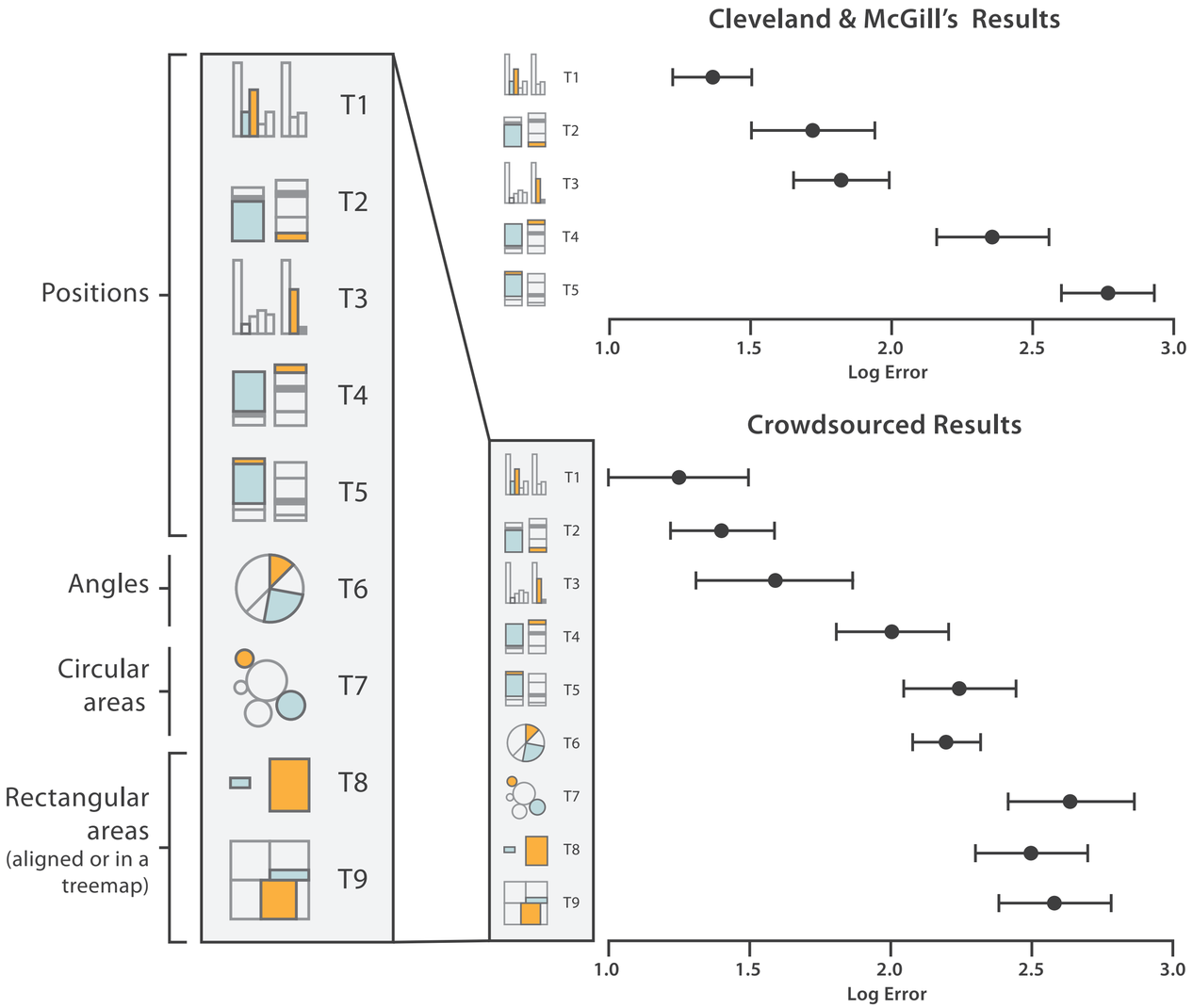


Identity Channels: Categorical Attributes

[Munzner (ill. Maguire), 2014]

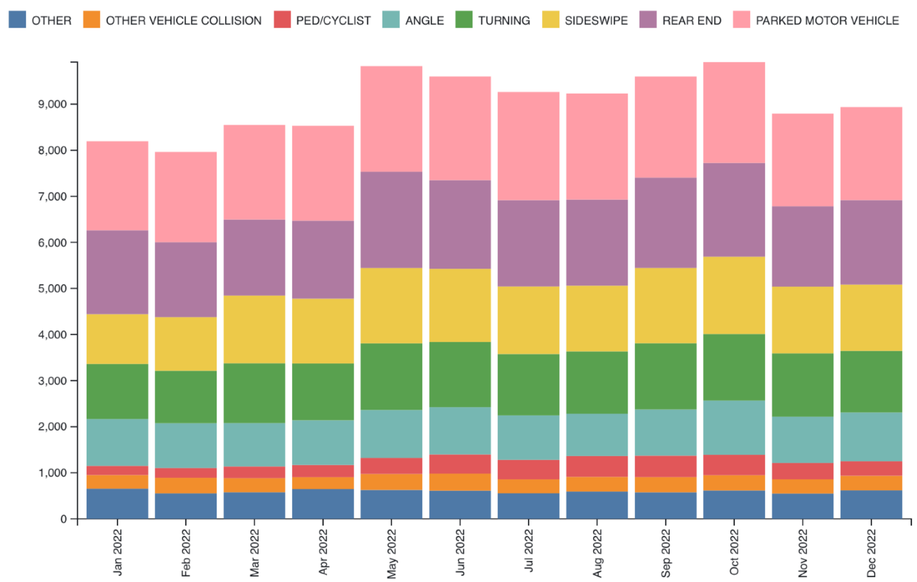
How was this determined?

Perception Studies Summary



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

Emergency Alert Test at 1:20pm

Project

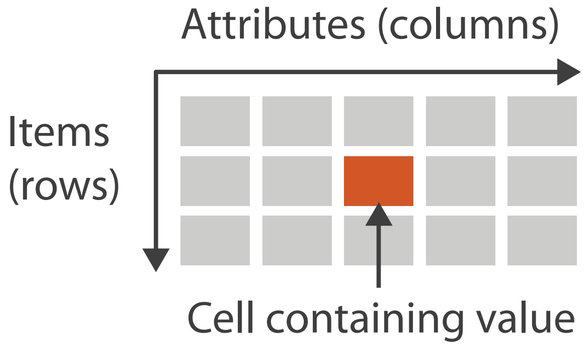
- Start thinking about project dataset and questions
- Working on posting some example datasets
- Goal: Less explored datasets (more opportunity for design/questions)
- If you are doing research and can tie this project in, please talk with me

Tables

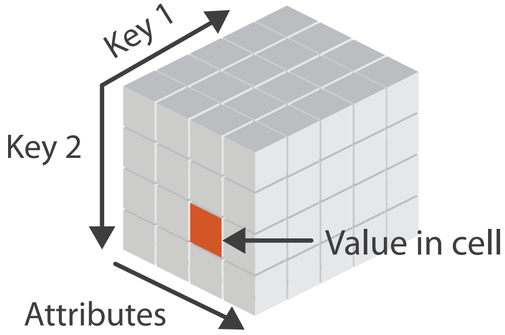
	REMOTE	STATION	FF	SEN/DIS	7-D AFAS UNL	D AFAS/RMF I	JOINT RR TKT	7-D UNL	30-D UNL
1	R011	42ND STREET & 8TH AVENUE	00228985	00008471	00000441	00001455	00000134	00033341	00071255
2	R170	14TH STREET-UNION SQUARE	00224603	00011051	00000827	00003026	00000660	00089367	00199841
3	R046	42ND STREET & GRAND CENTRAL	00207758	00007908	00000323	00001183	00003001	00040759	00096613
4	R012	34TH STREET & 8TH AVENUE	00188311	00006490	00000498	00001279	00003622	00035527	00067483
5	R293	34TH STREET - PENN STATION	00168768	00006155	00000523	00001065	00005031	00030645	00054376
6	R033	42ND STREET/TIMES SQUARE	00159382	00005945	00000378	00001205	00000690	00058931	00078644
7	R022	34TH STREET & 6TH AVENUE	00156008	00006276	00000487	00001543	00000712	00058910	00110466
8	R084	59TH STREET/COLUMBUS CIRCLE	00155262	00009484	00000589	00002071	00000542	00053397	00113966
9	R020	47-50 STREETS/ROCKEFELLER	00143500	00006402	00000384	00001159	00000723	00037978	00090745
10	R179	86TH STREET-LEXINGTON AVE	00142169	00010367	00000470	00001839	00000271	00050328	00125250
11	R023	34TH STREET & 6TH AVENUE	00134052	00005005	00000348	00001112	00000649	00031531	00075040
12	R029	PARK PLACE	00121614	00004311	00000287	00000931	00000792	00025404	00065362
13	R047	42ND STREET & GRAND CENTRAL	00100742	00004273	00000185	00000704	00001241	00022808	00068216

Visualization of Tables

- Items and attributes
- For now, attributes are not known to be positions
- Keys and values
  - **key** is an independent attribute that is unique and identifies item
  - **value** tells some aspect of an item
- Keys: categorical/ordinal
- Values: categorical/ordinal/quantitative
- Levels: unique *values* of categorical or ordered attributes



➔ *Multidimensional Table*

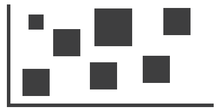


[Munzner (ill. Maguire), 2014]

Arrange Tables

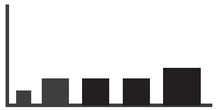
- **Express Values**

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

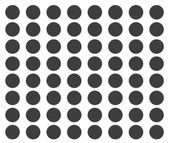


  - Order

  - Align



- **Layout Density**
  - Dense



  - Space-Filling

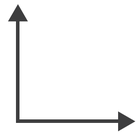
  - *1 Key*
    List

  - *2 Keys*
    Matrix

  - *3 Keys*
    Volume

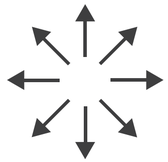
  - *Many Keys*
    Recursive Subdivision

- **Axis Orientation**
  - Rectilinear



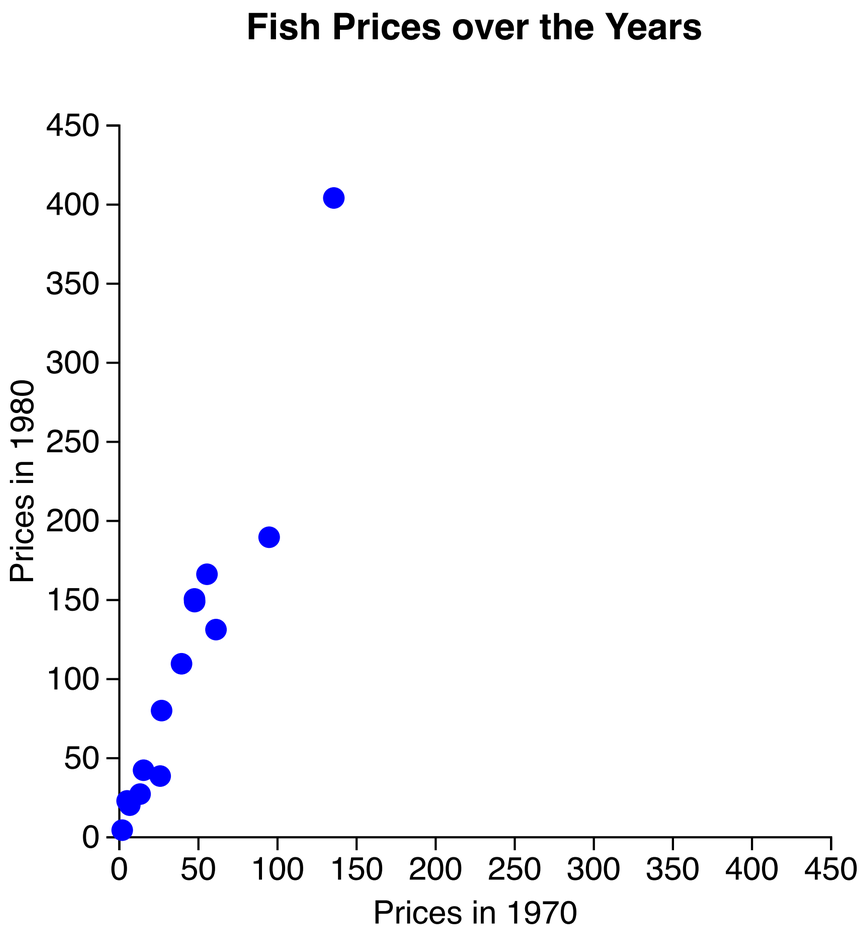
  - Parallel

  - Radial



[Munzner (ill. Maguire), 2014]

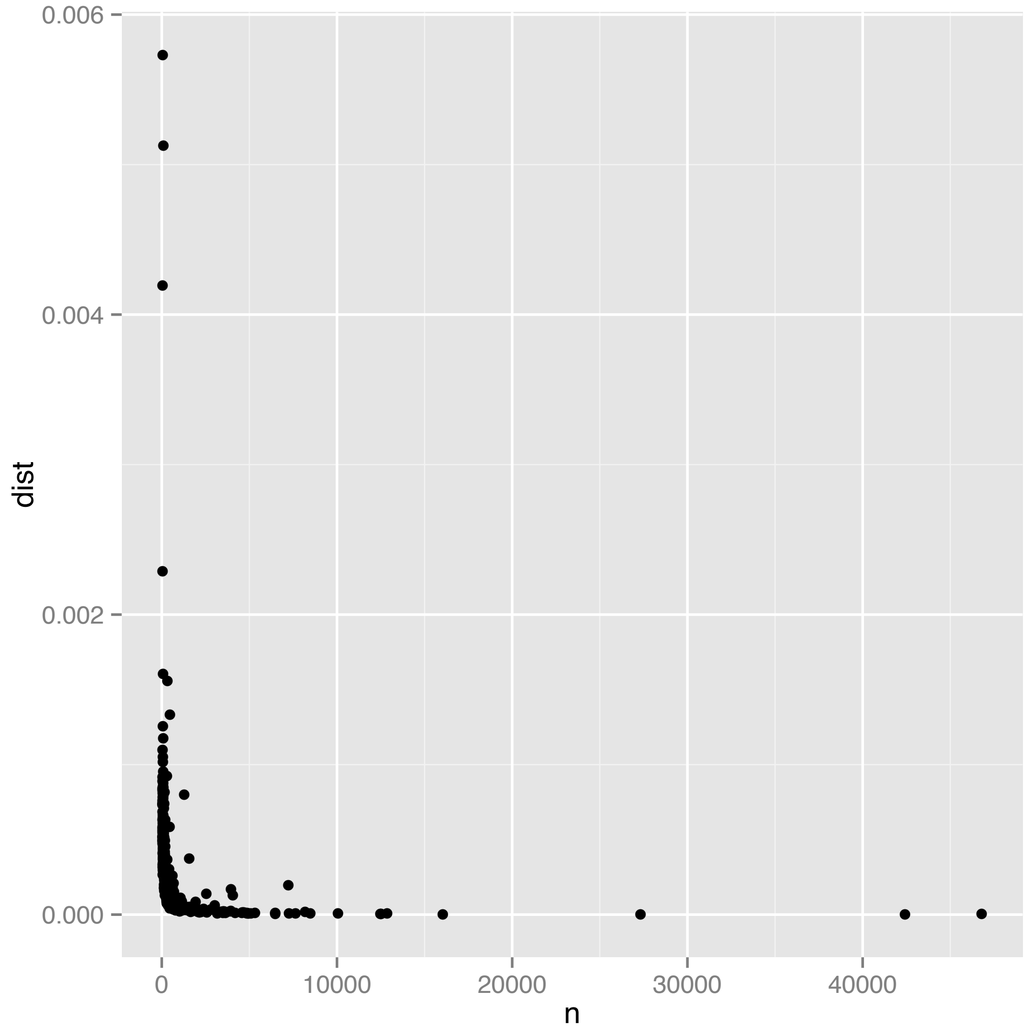
Express Values: Scatterplots



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

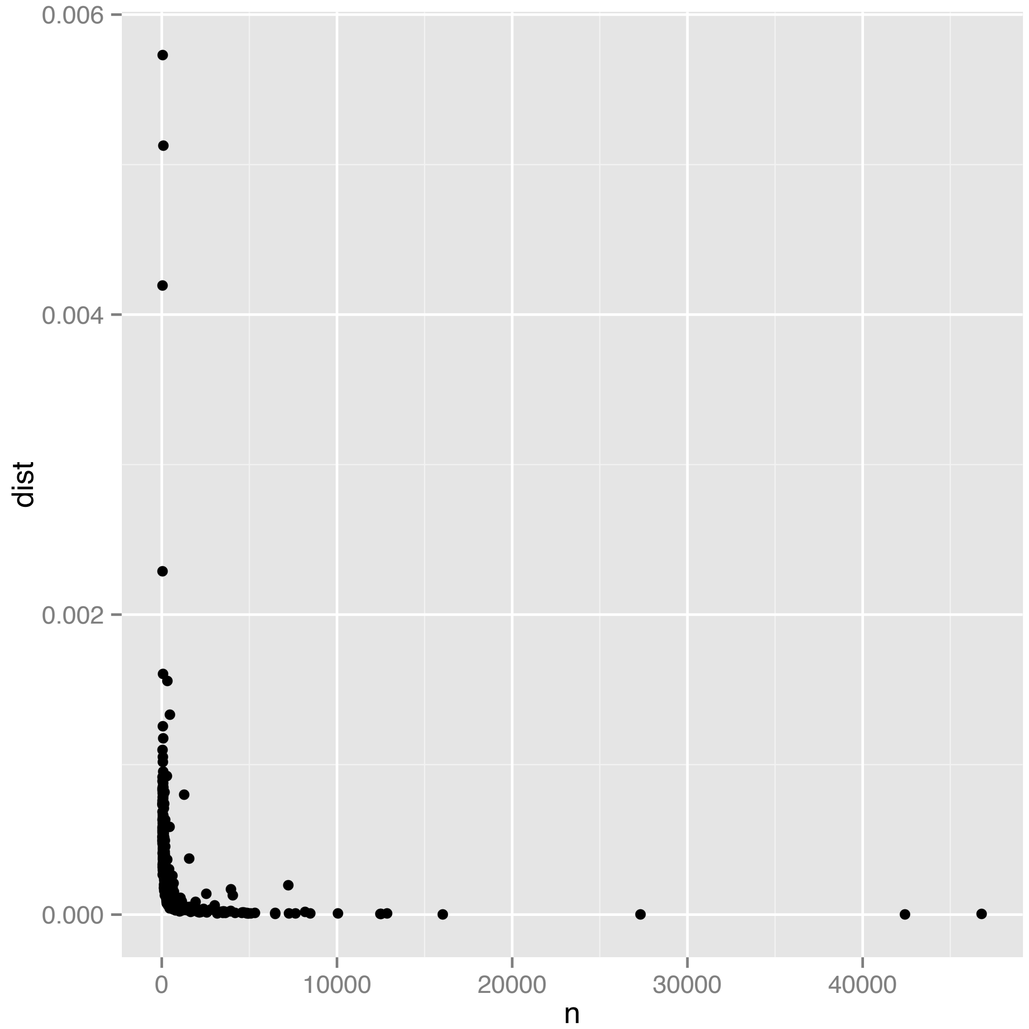
- Correlation: dependence between two attributes
  - Positive and negative correlation
  - Indicated by lines
- Coordinate system (axes) and labels are important!

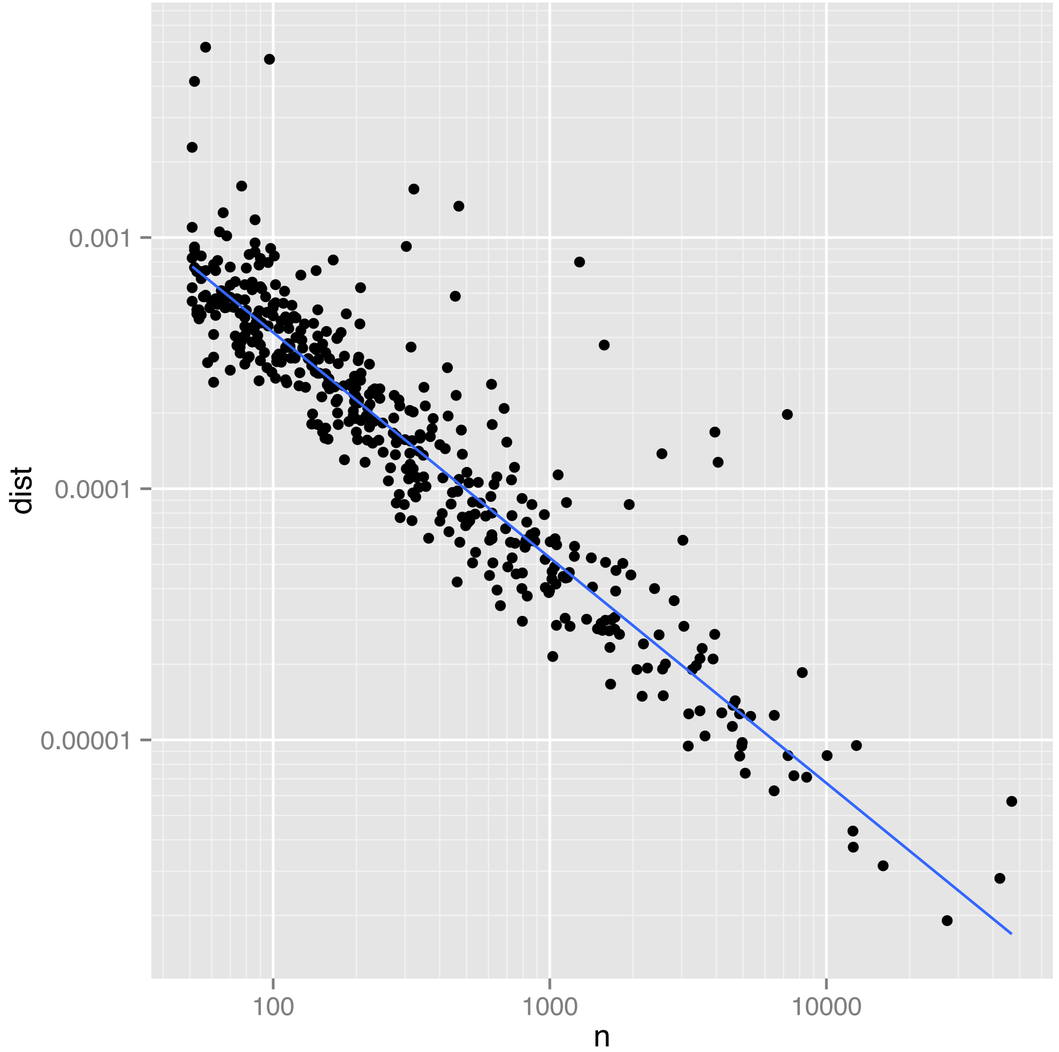
Coordinate Systems



[Wickham, 2014]

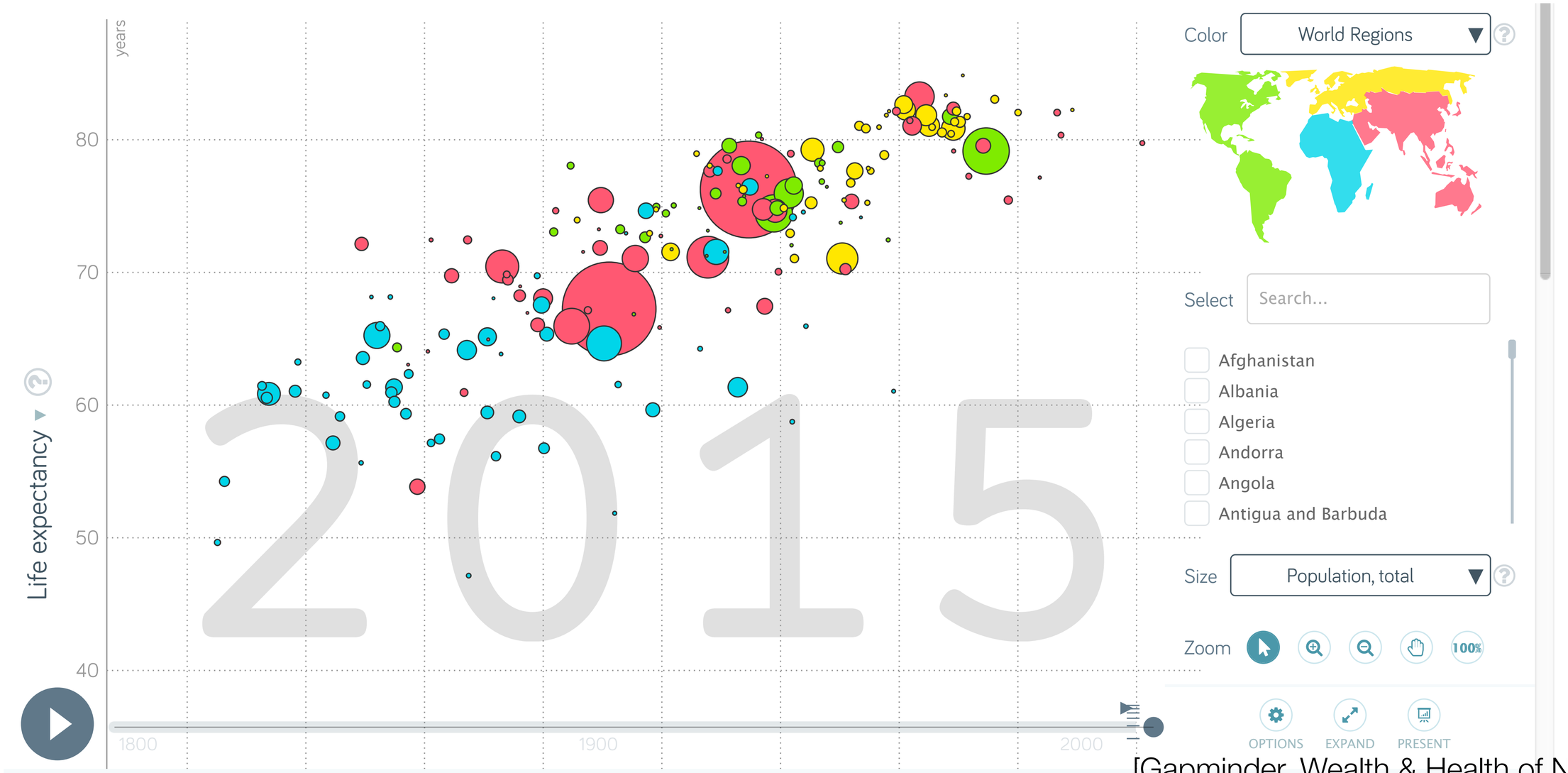
Coordinate Systems





[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

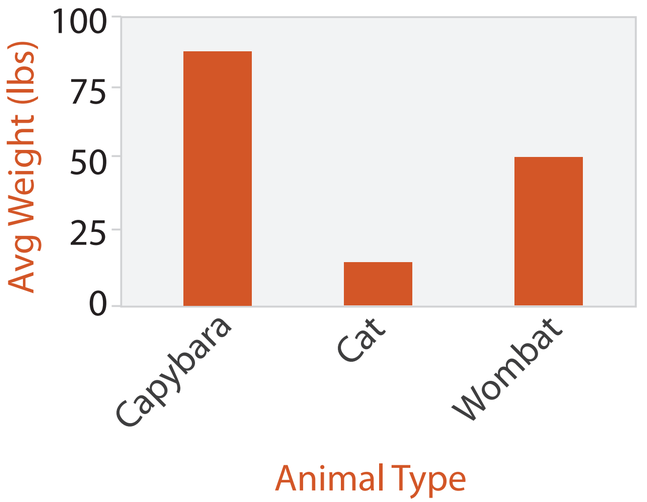
- "Ranking Visualizations of Correlation Using Weber's Law", 2014:
  - Correlation perception can be modeled via Weber's Law
  - Scatterplots are one of the best visualizations for both positive and negative correlation
  - Further analysis: M. Kay and J. Heer, "Beyond Weber's Law", 2015

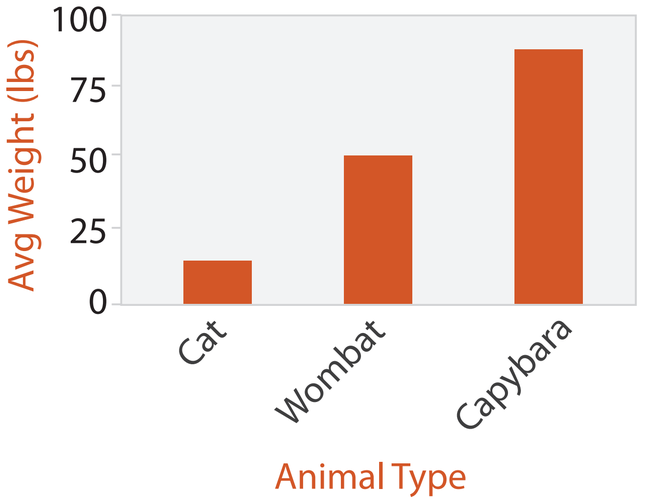
Separate, Order, and Align: Categorical Regions

- 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 (use some other ordered attribute)
  - Order
- Alignment and order can use same or different attribute

List Alignment: Bar Charts

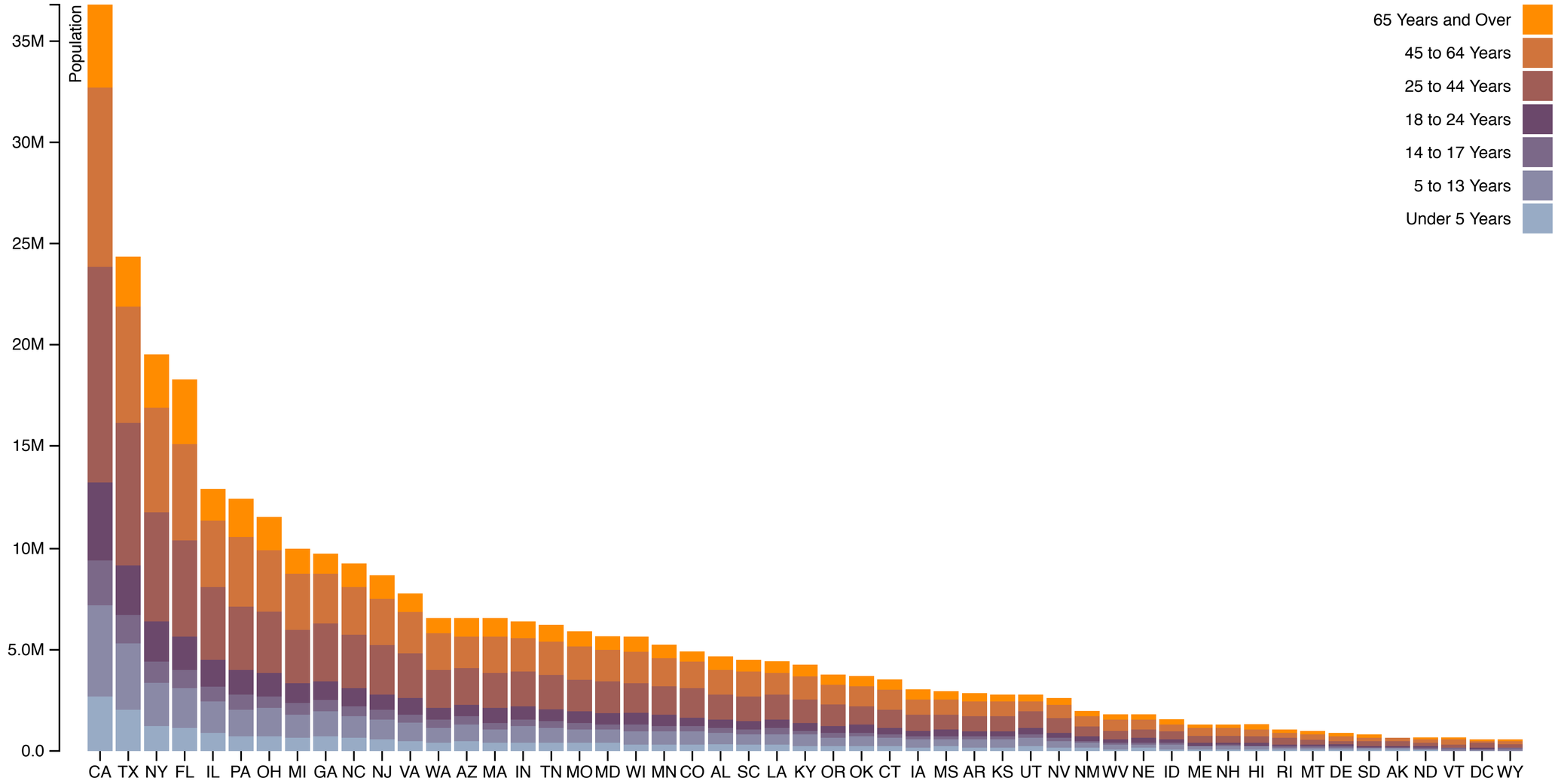
- Data: one quantitative attribute, one categorical attribute
- Task: lookup & compare values
- How: line marks, vertical position (quantitative), horizontal position (categorical)
- What about **length**?
- Ordering criteria: alphabetical or using quantitative attribute
- Scalability: distinguishability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





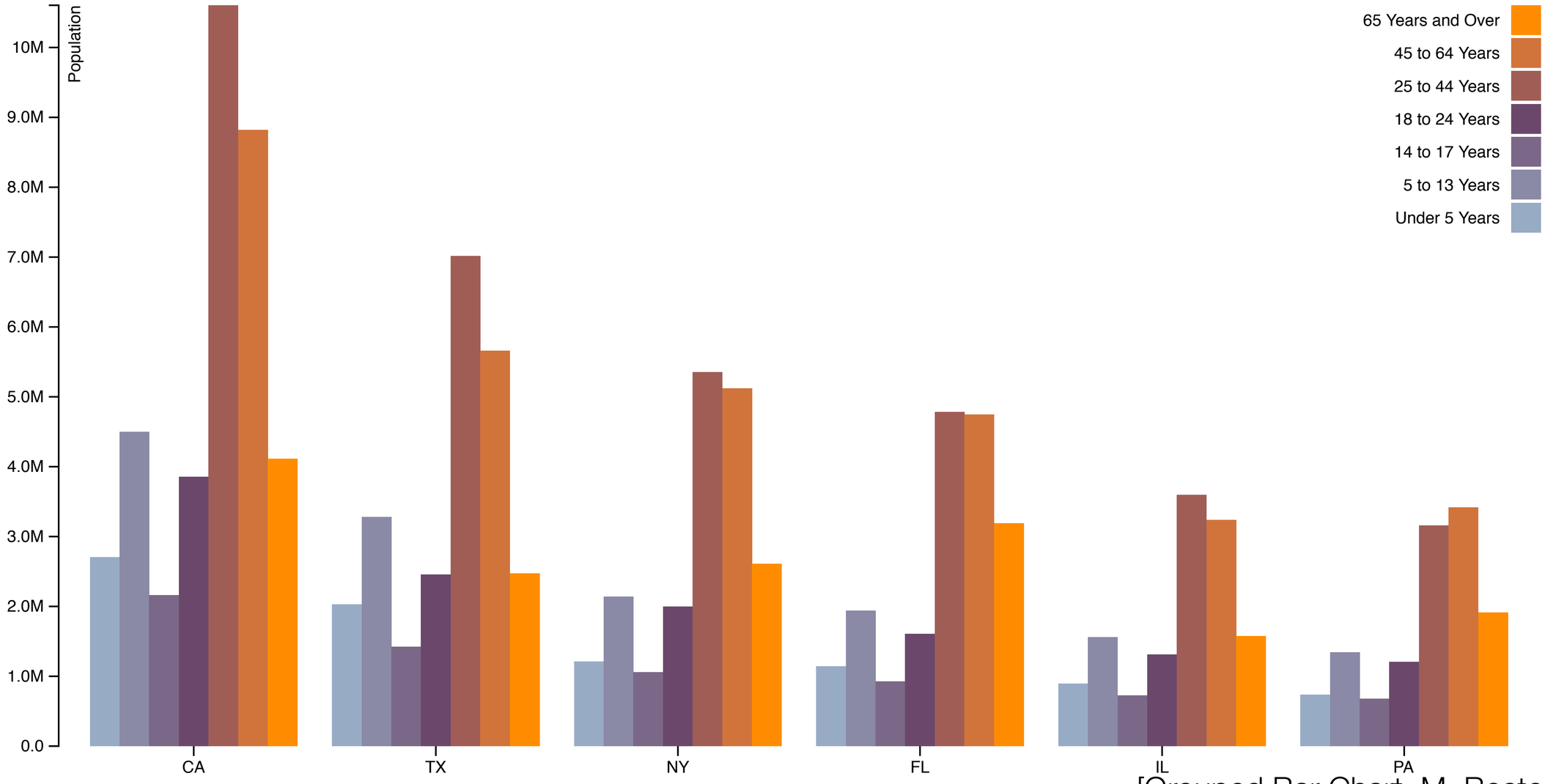
[Munzner (ill. Maguire), 2014]

Stacked Bar Charts



[Stacked Bar Chart, M. Bostock, 2017]

Grouped Bar Chart



[Grouped Bar Chart, M. Bostock, 2017]

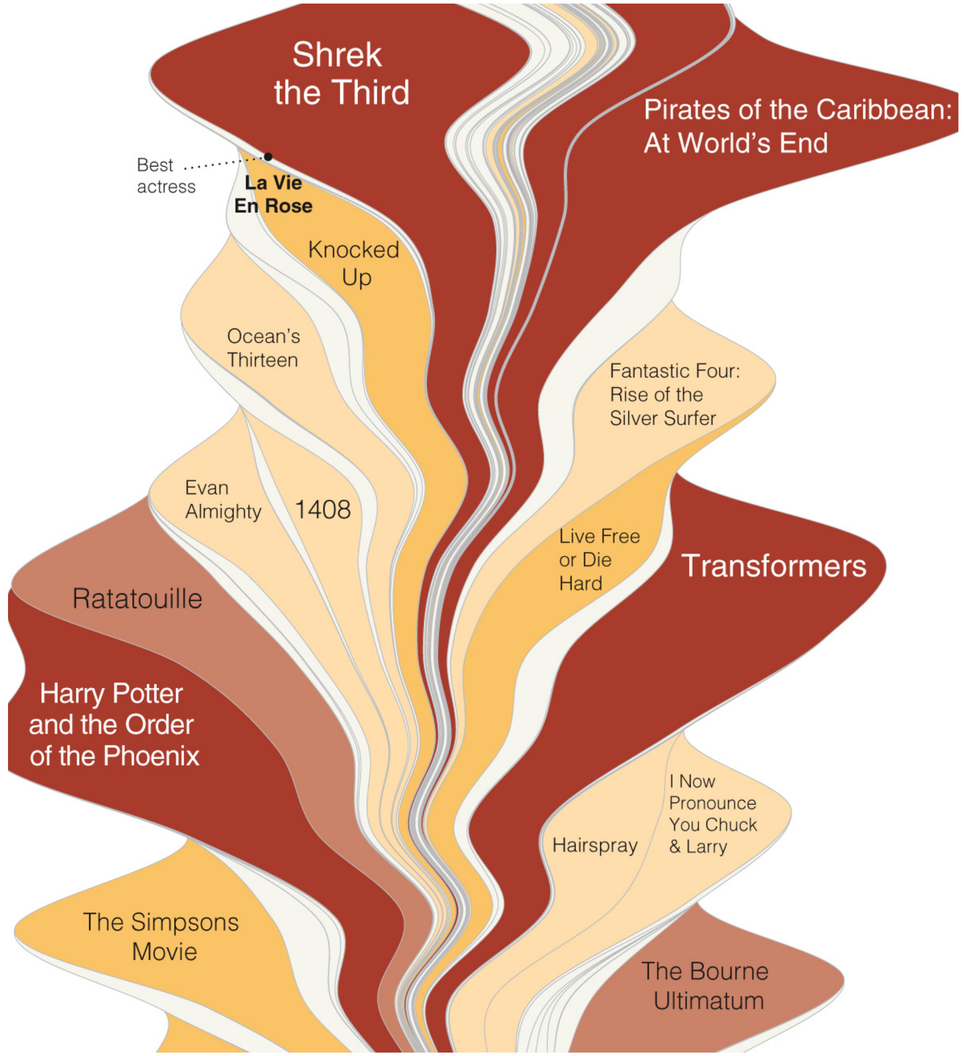
Stacked Bar Charts

- 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.

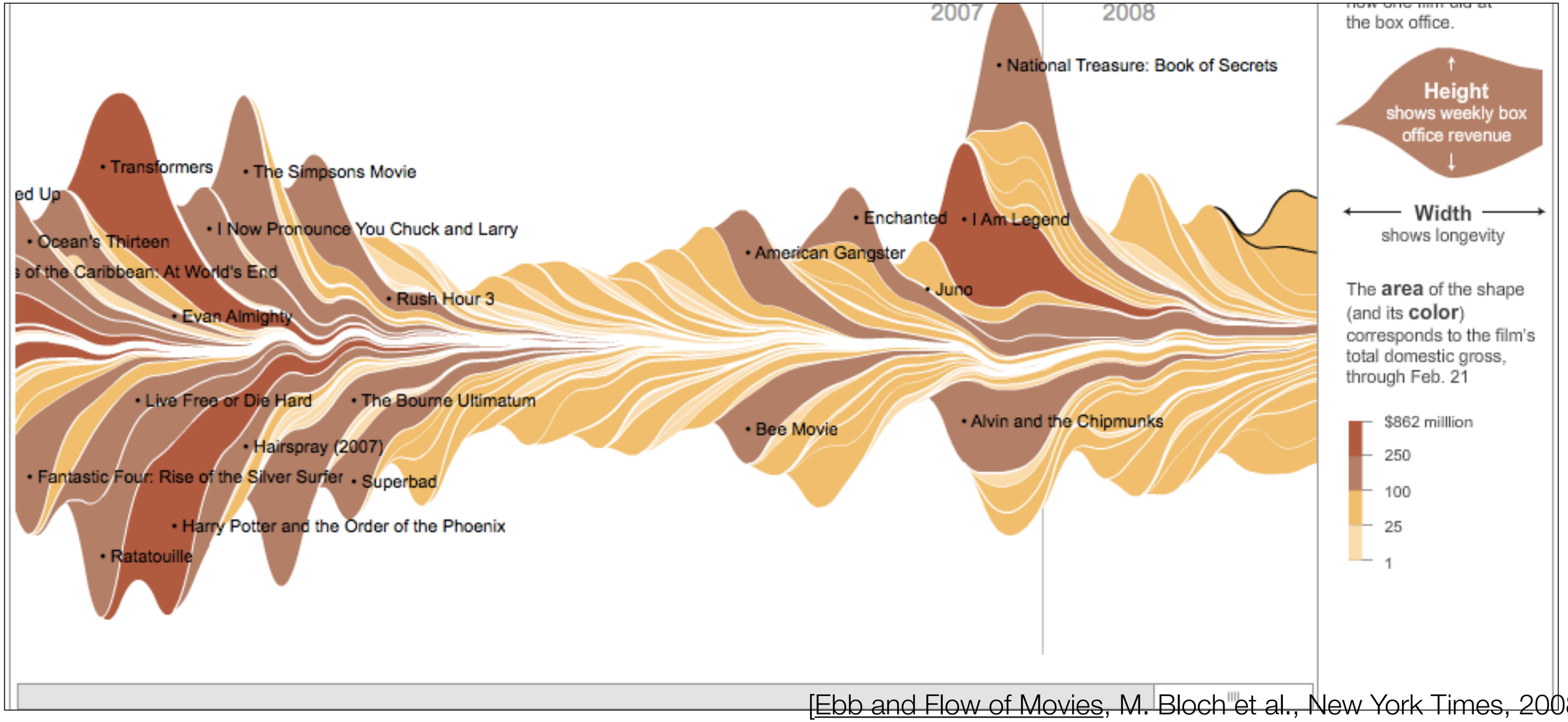
Streamgraphs

- 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



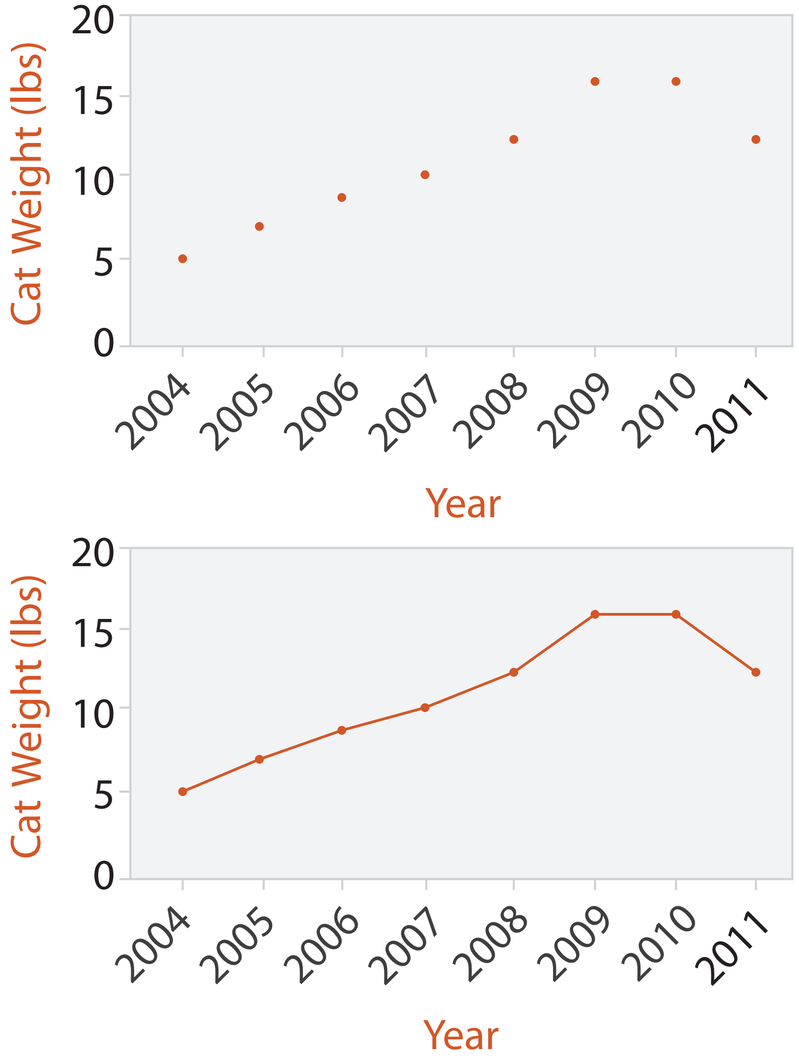
[Byron and Wattenberg, 2012]

Streamgraphs



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

Dot and Line Charts



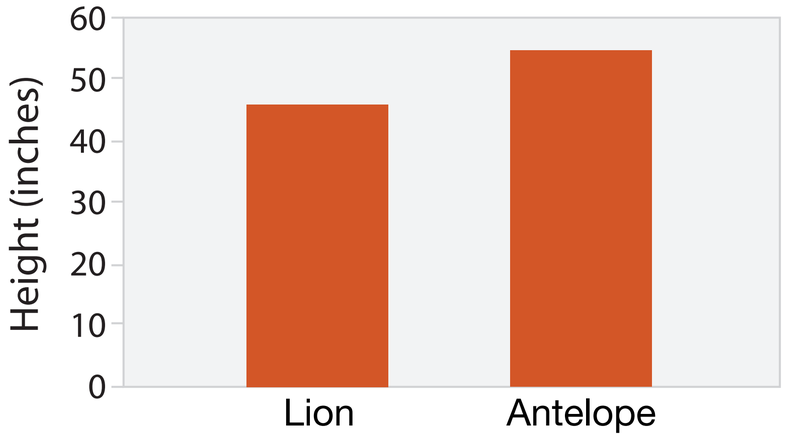
- Data: one quantitative attribute, one **ordered** attribute
- Task: lookup values, find outliers and trends
- How: point mark and positions

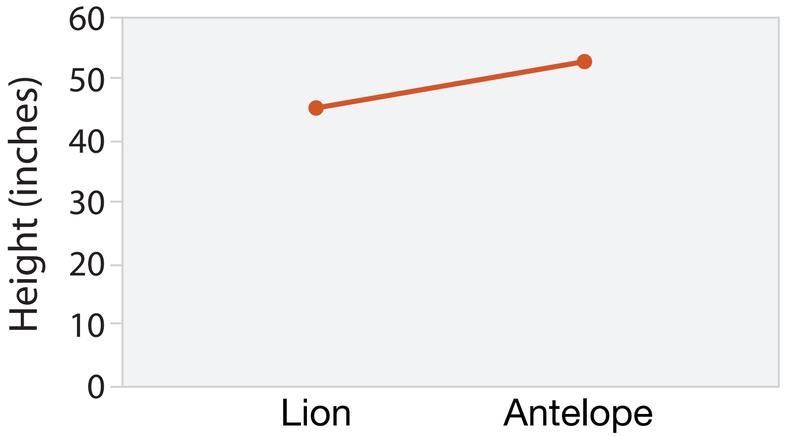
- Line Charts: add **connection mark** (line)

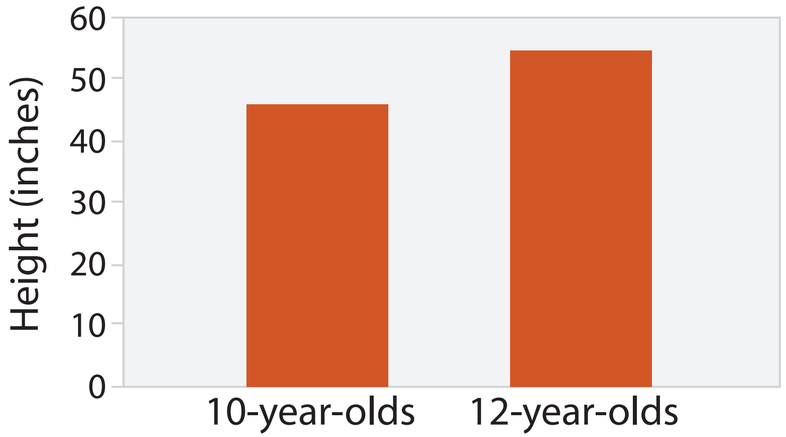
- Similar to scatterplots but allow ordered attribute

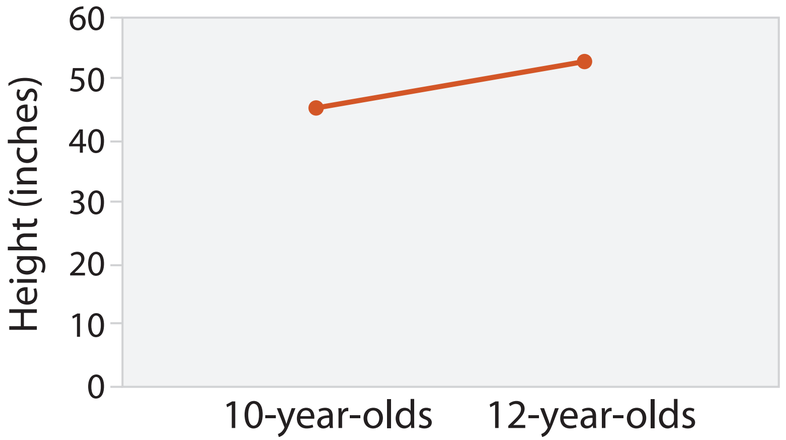
[Munzner (ill. Maguire), 2014]

Proper Use of Line and Bar Charts



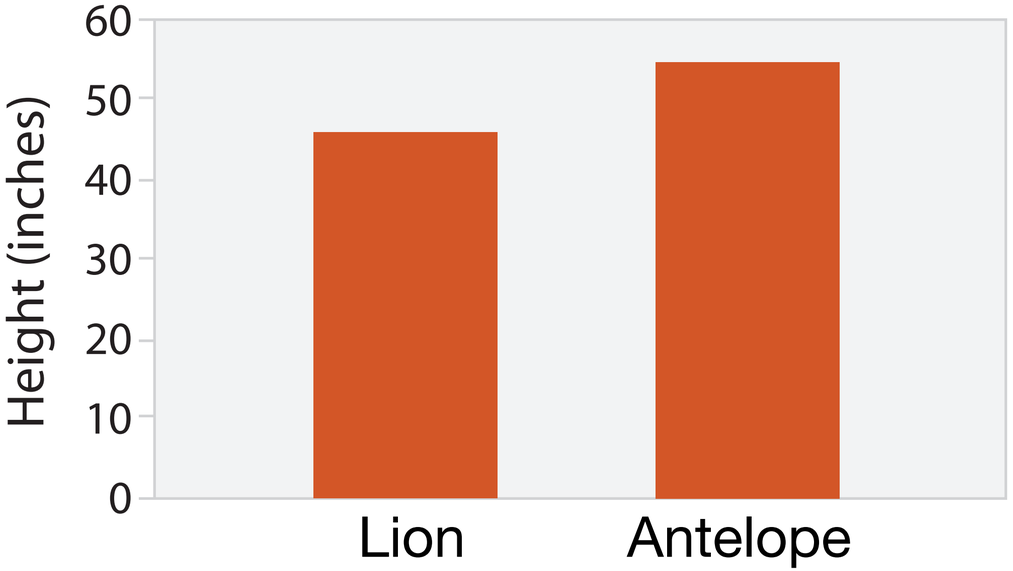


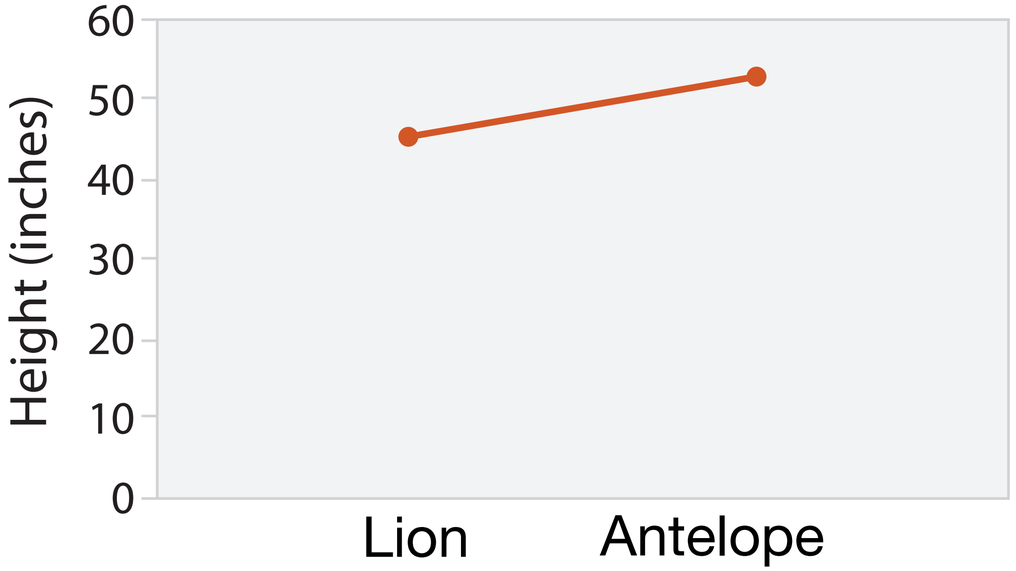




[Adapted from Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]

Proper Use of Line and Bar Charts





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

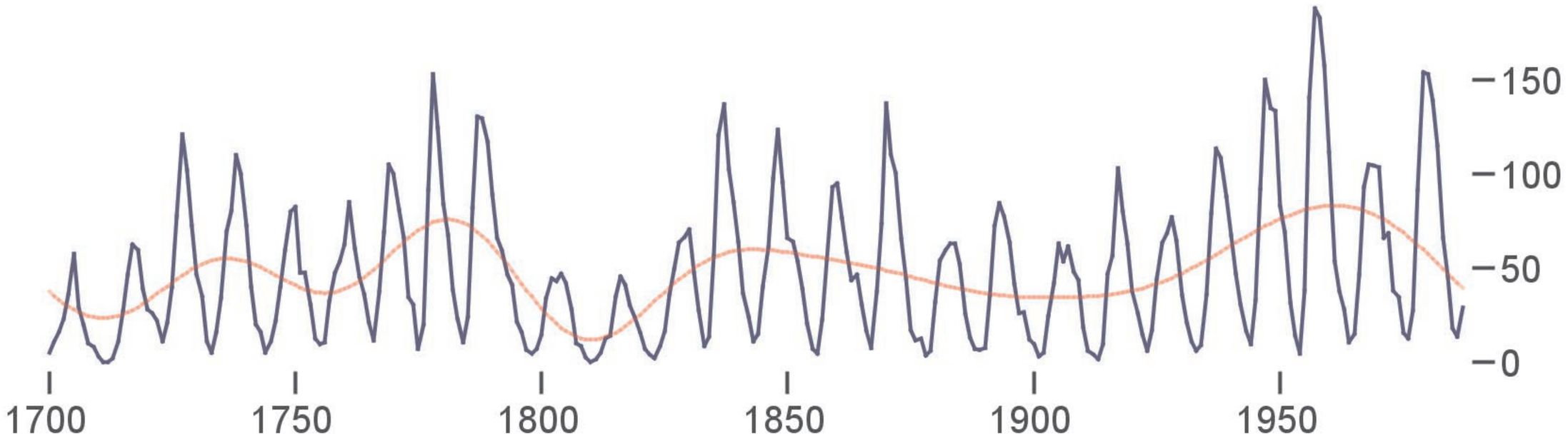
[Adapted from Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]

Aspect Ratio

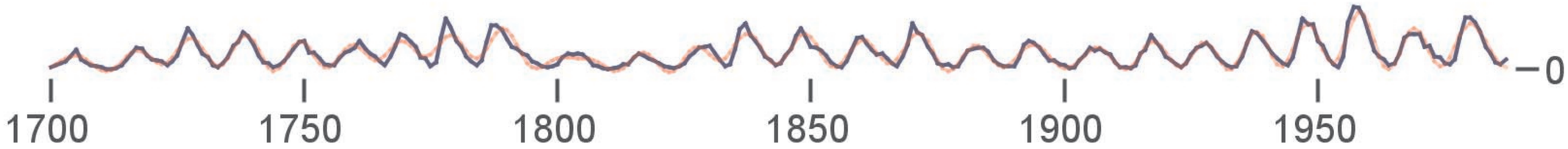
- Trends in line charts are more apparent because we are using angle as a channel
- Perception of angle (and the **relative difference** between angles) is important
- Initial experiments found people best judge differences in **slope** when angles are around 45 degrees (Cleveland et al., 1988, 1993)

Multiscale Banking

Aspect Ratio = 3.96



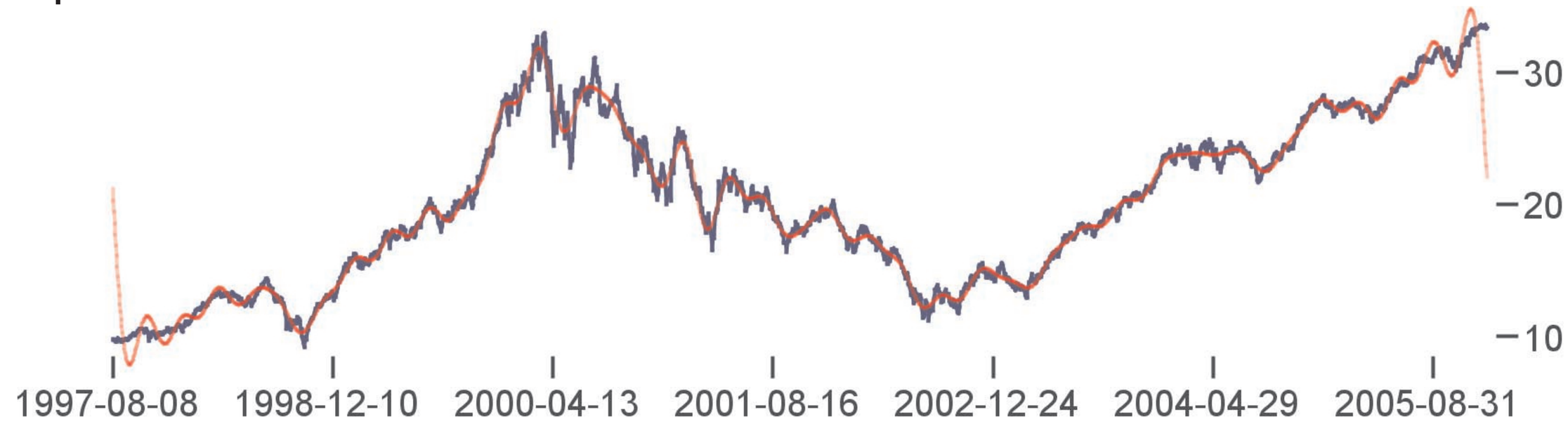
Aspect Ratio = 22.35



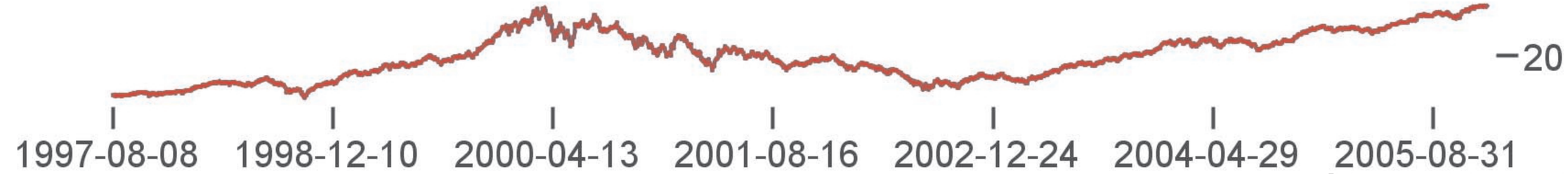
[Heer and Agrawala, 2006]

Multiscale Banking

Aspect Ratio = 4.23



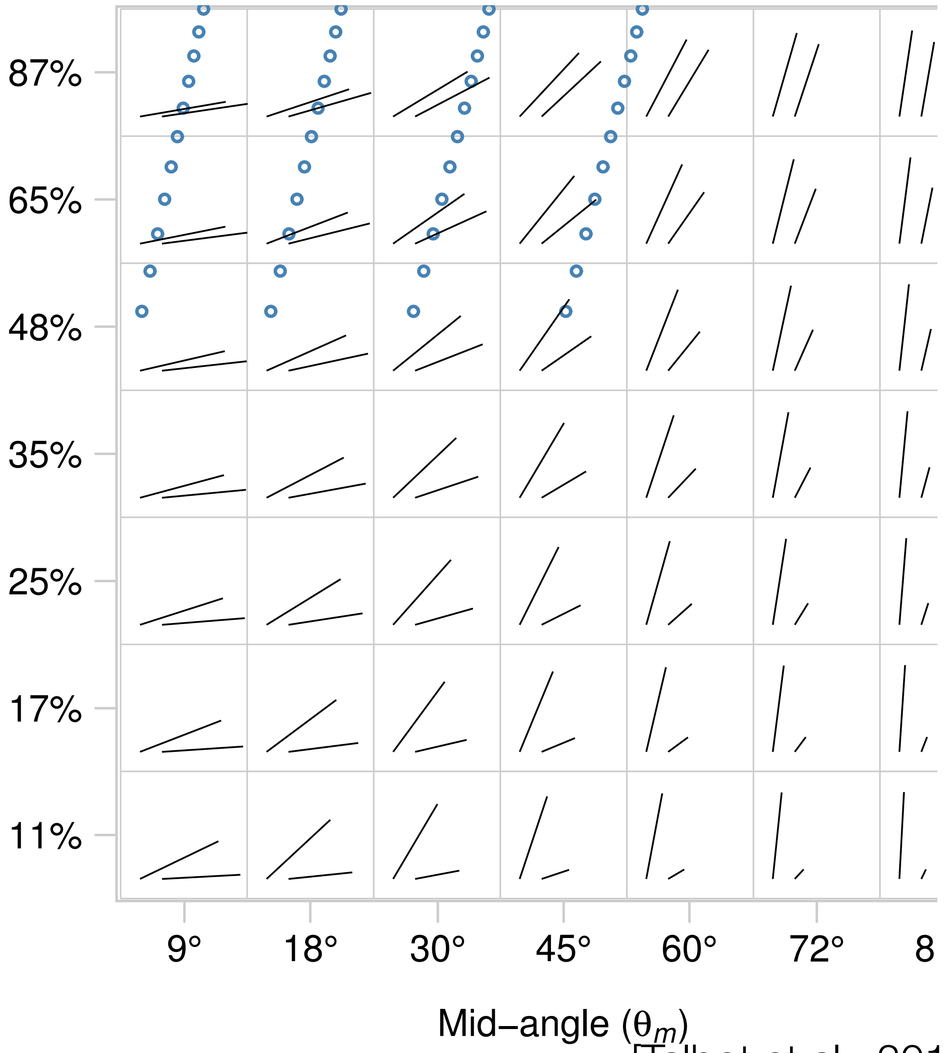
Aspect Ratio = 14.55



[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

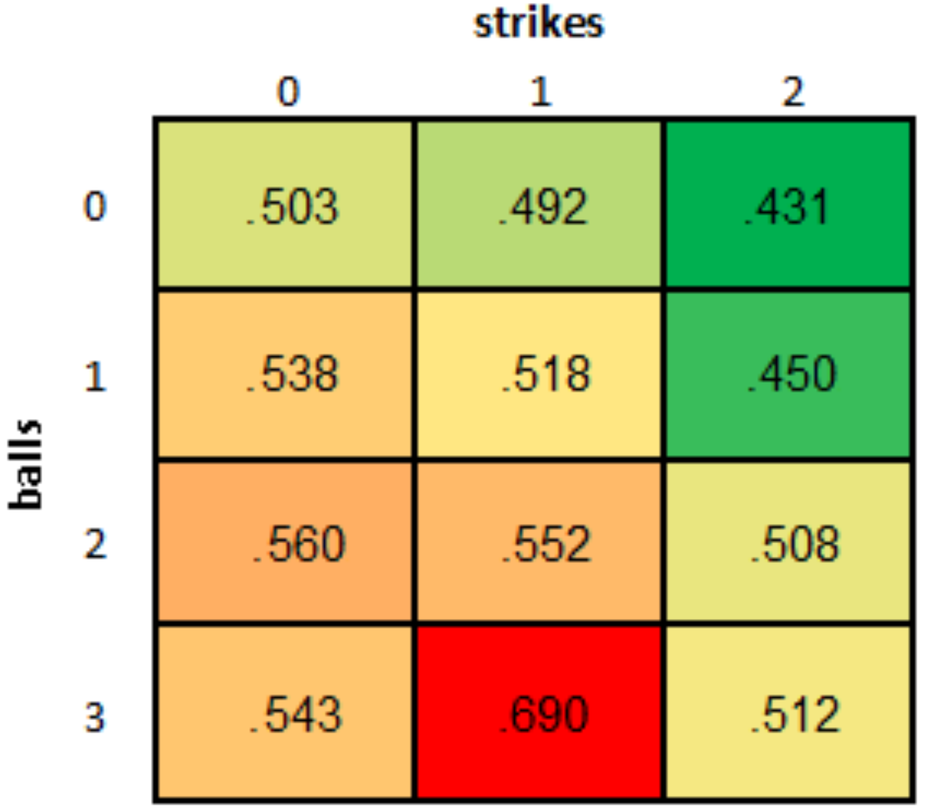


[Talbot et al., 2013]

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]

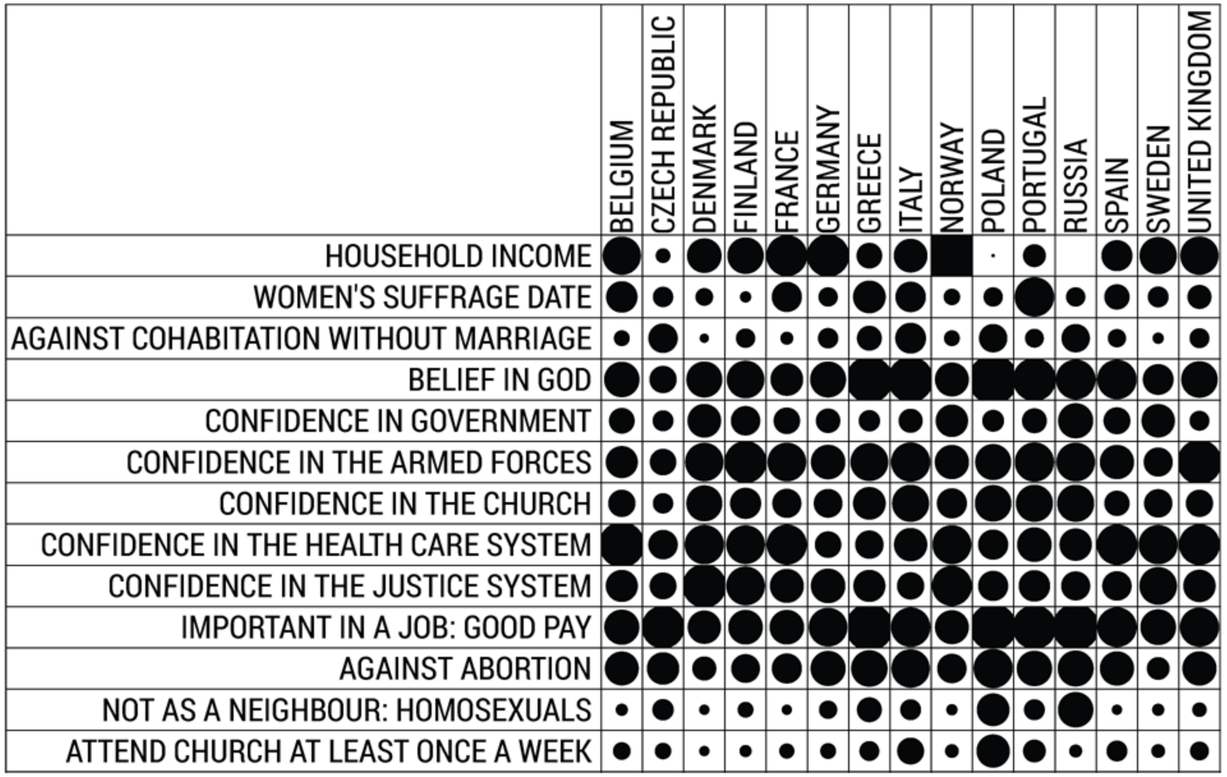
Bertin Matrices

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

[C.Perrin et al., 2014]

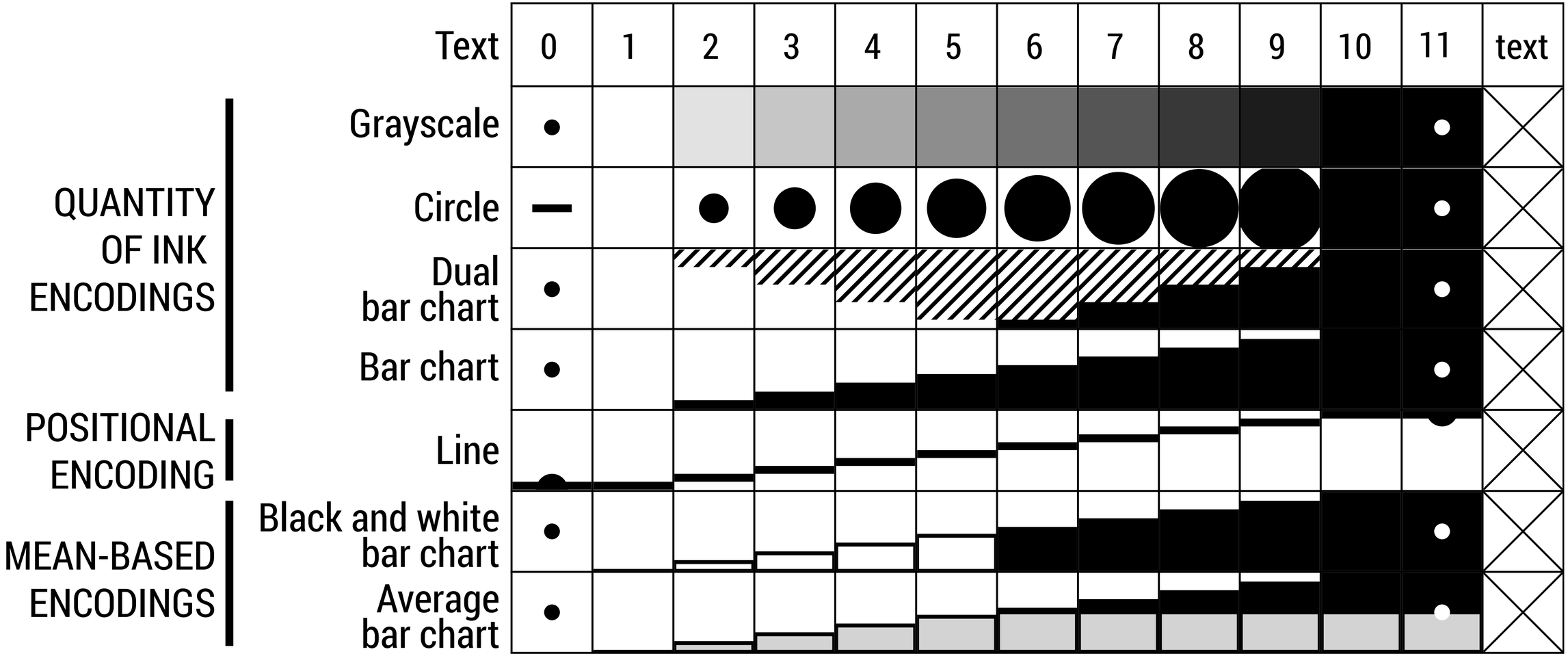
Bertin Matrices

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



[C.Perrin et al., 2014]

Bertin's Encodings

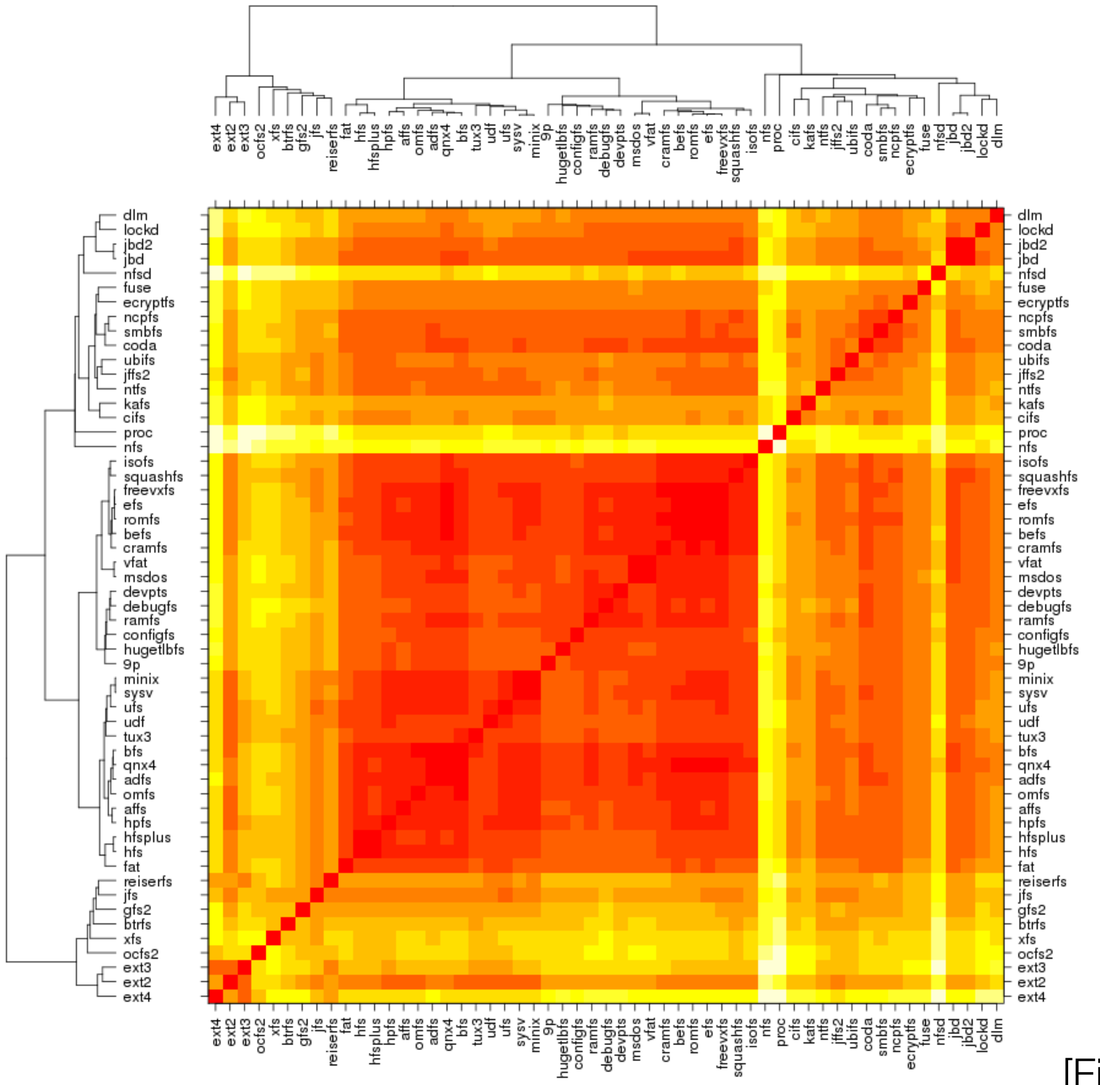


[C.Perrin et al., 2014]

Matrix Reordering

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

Cluster Heatmap



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

Cluster Heatmap

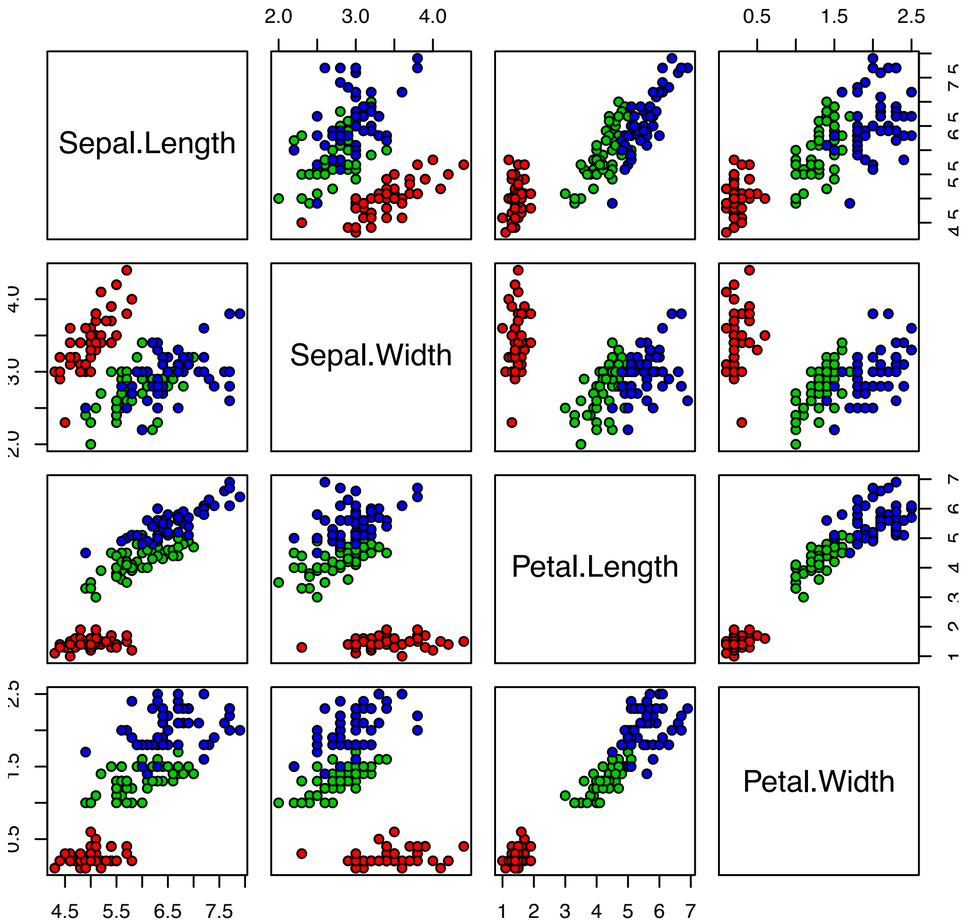
- Data & Task: Same as Heatmap
- How: Area marks but matrix is ordered by cluster hierarchies
- Scalability: limited by the cluster dendrogram

- 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]

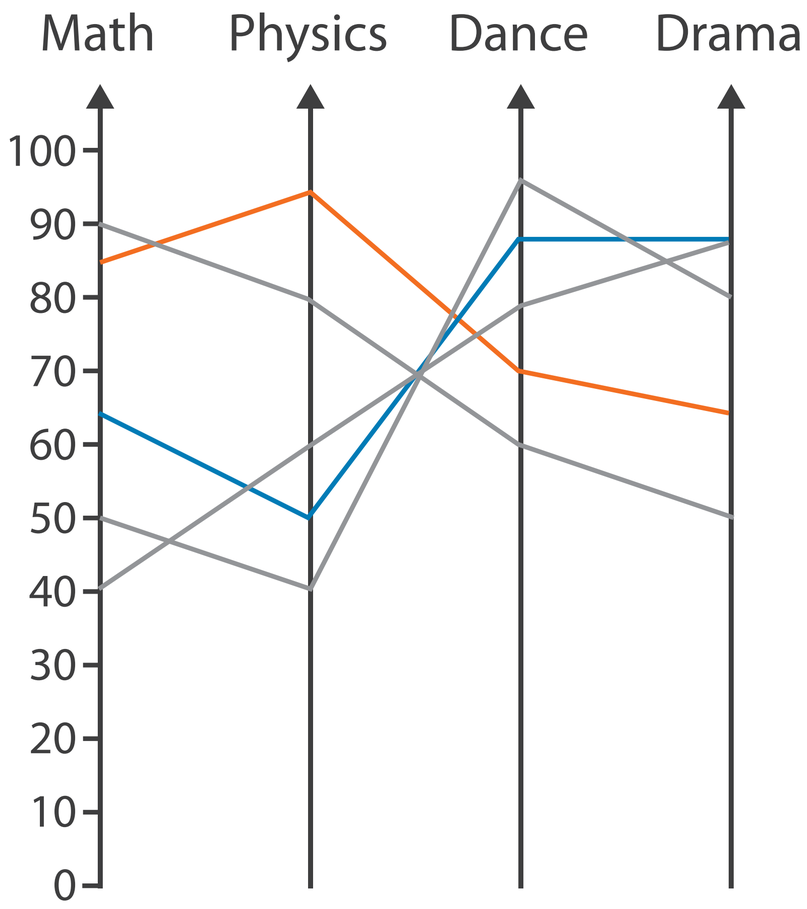
Spatial Axis Orientation

- 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

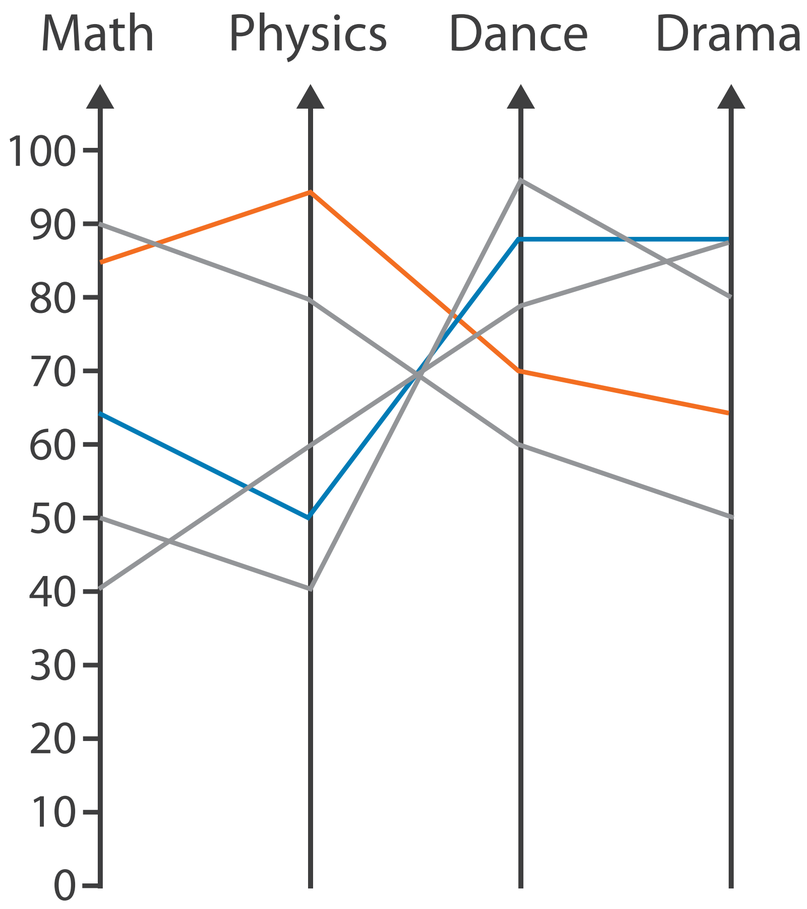
- So far, we have seen the vertical and horizontal axes (a **rectilinear** layout) used to encode almost everything
- What other possibilities are there for axes?
  - Parallel axes



[Munzner (ill. Maguire), 2014]

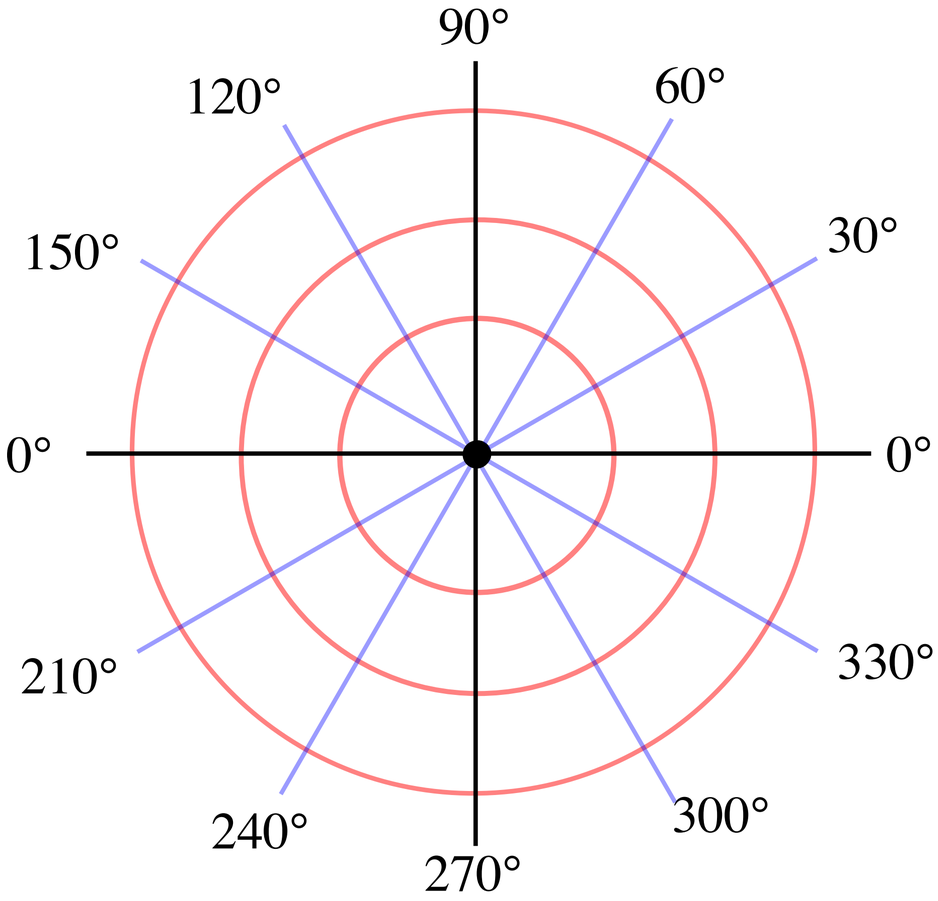
Spatial Axis Orientation

- 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

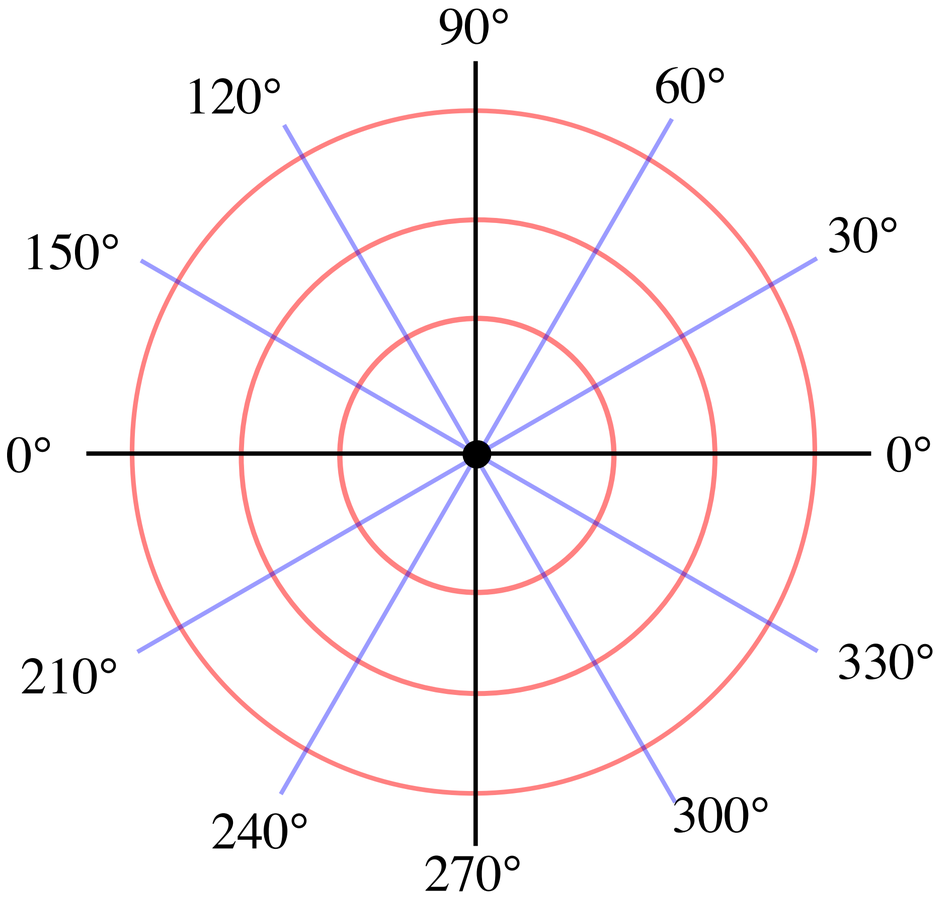


[Munzner (ill. Maguire), 2014]

Radial Axes

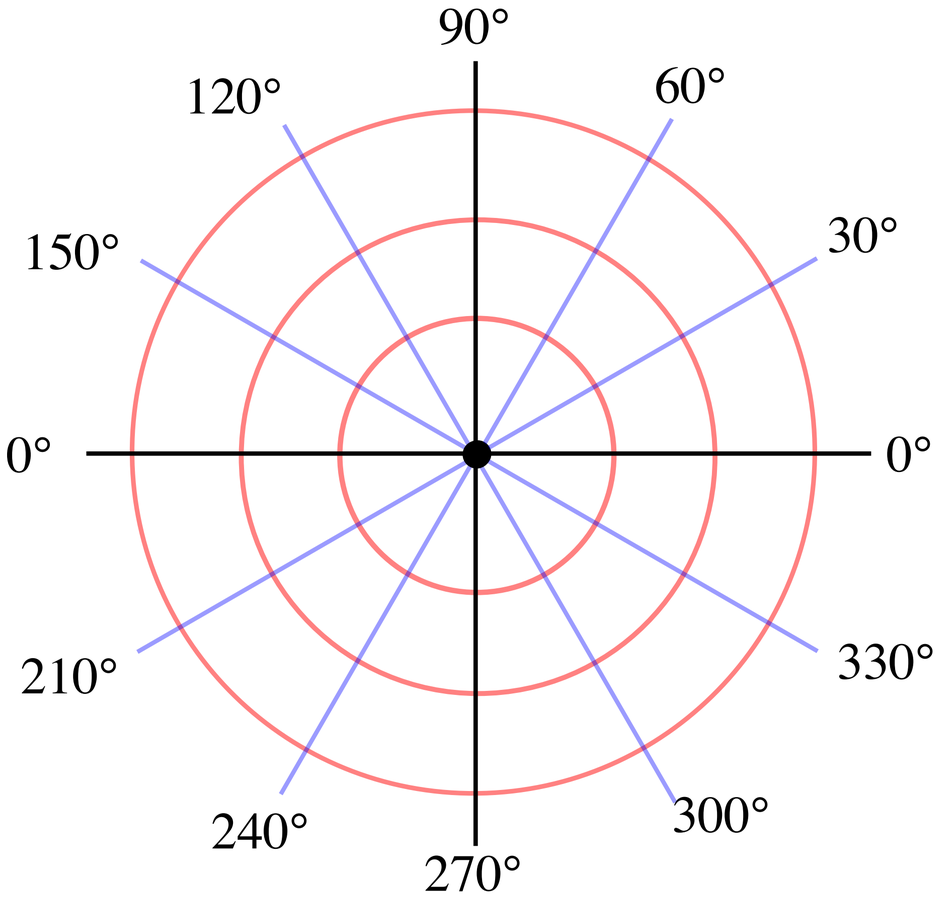


Radial Axes



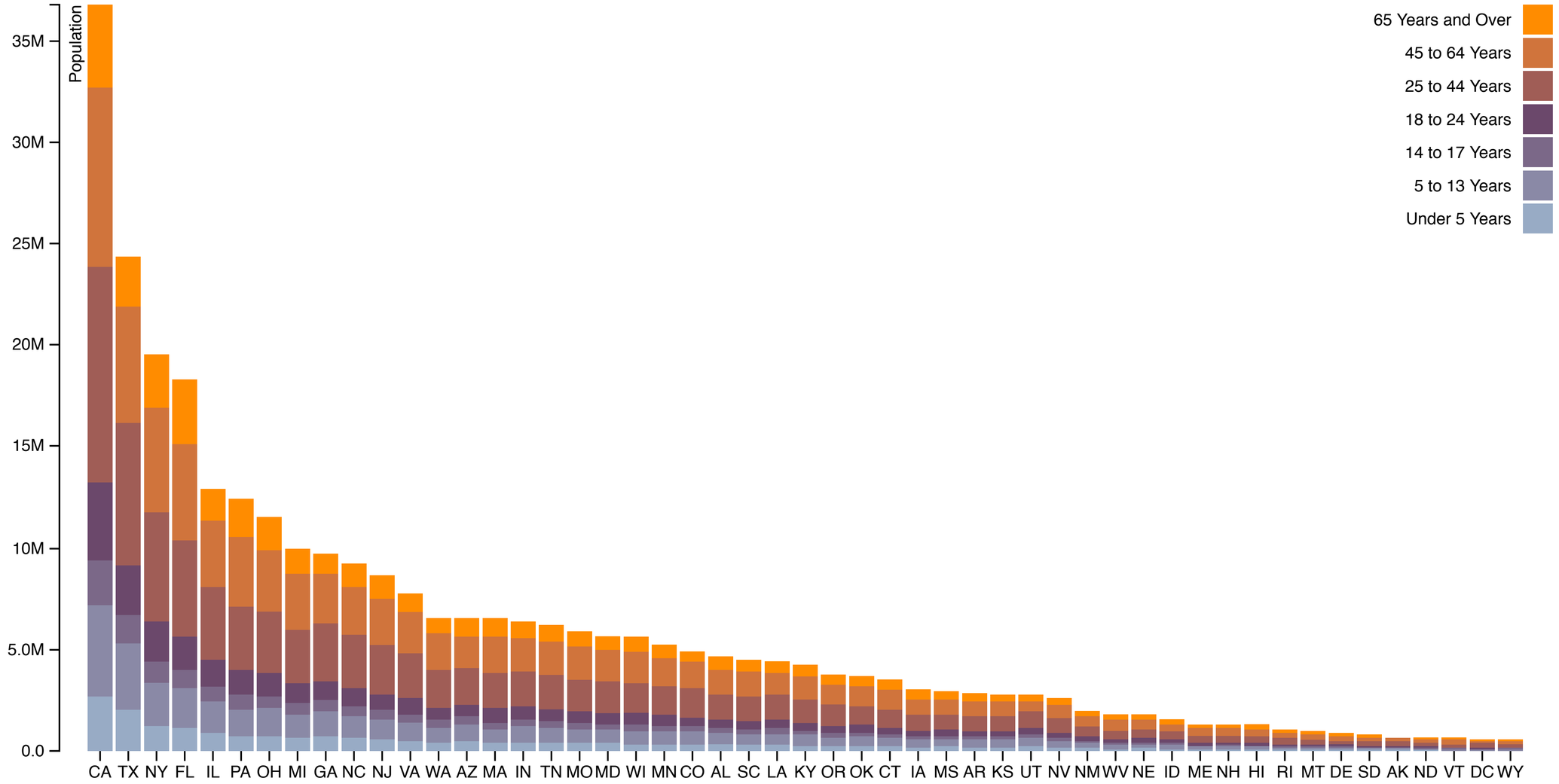
- Polar Coordinates (angle + position along the line at that angle)
- What types of encodings are possible for tabular data in polar coordinates?

Radial Axes



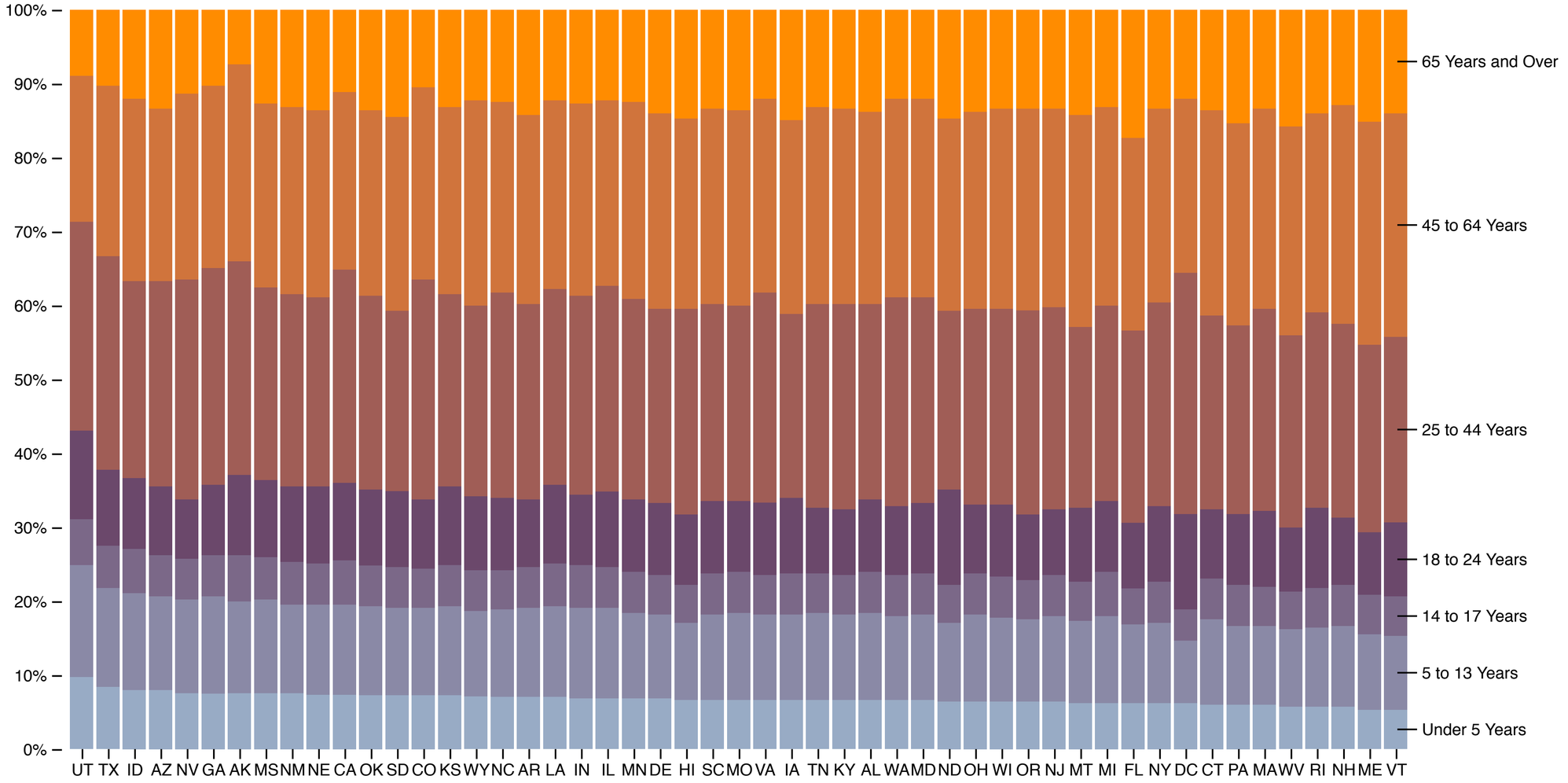
- Polar Coordinates (angle + position along the line at that angle)
- What types of encodings are possible for tabular data in polar coordinates?
  - Radial bar charts
  - Pie charts
  - Donut charts

Part-of-whole: Relative % comparison?



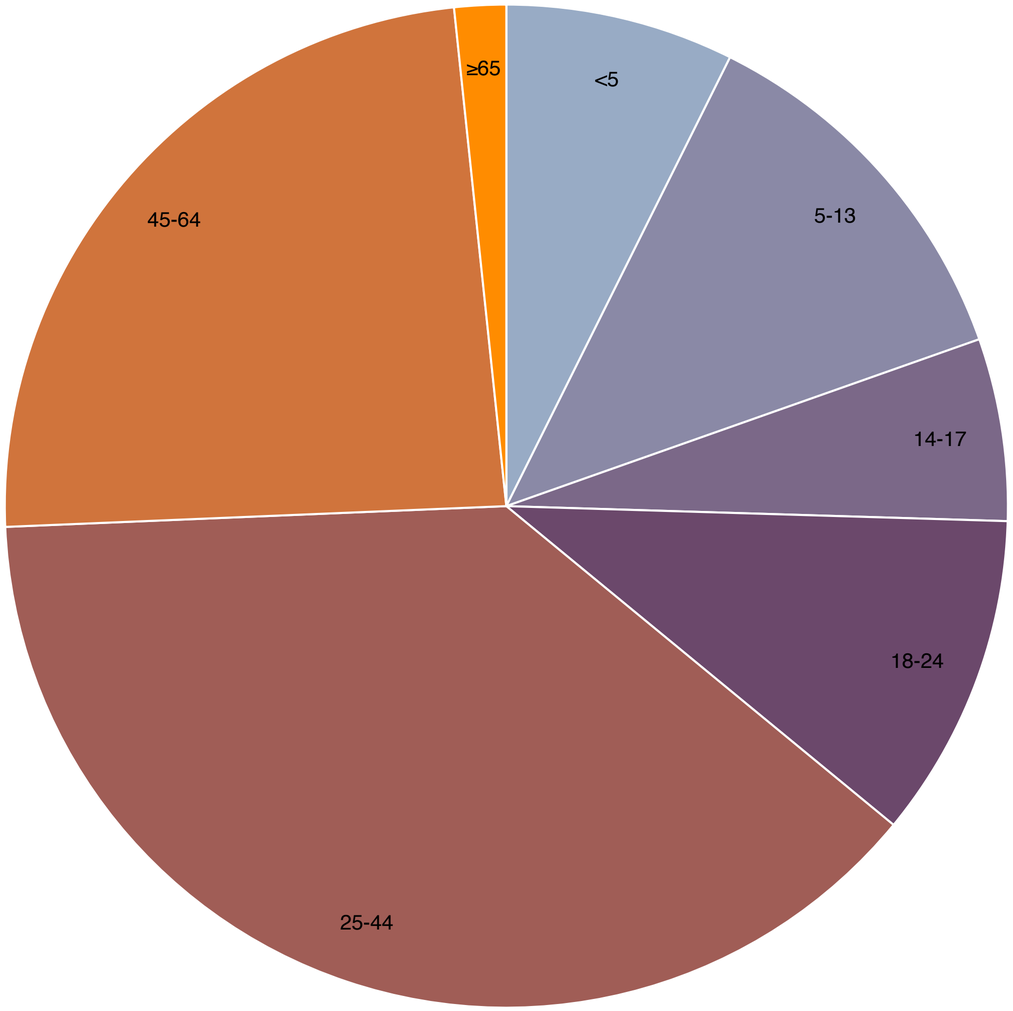
[Stacked Bar Chart, M. Bostock, 2017]

Normalized Stacked Bar Chart



[Normalized Stacked Bar Chart, Bostock, 2017]

Pie Chart



[Pie Chart, Bostock, 2017]