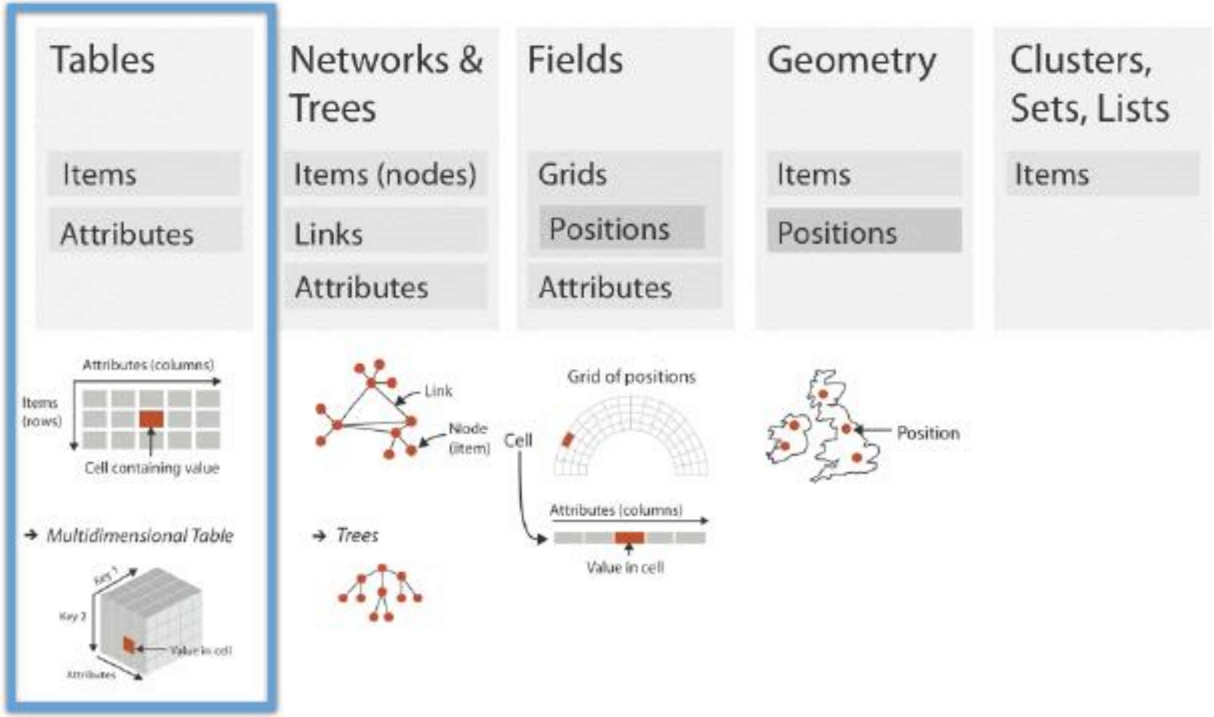
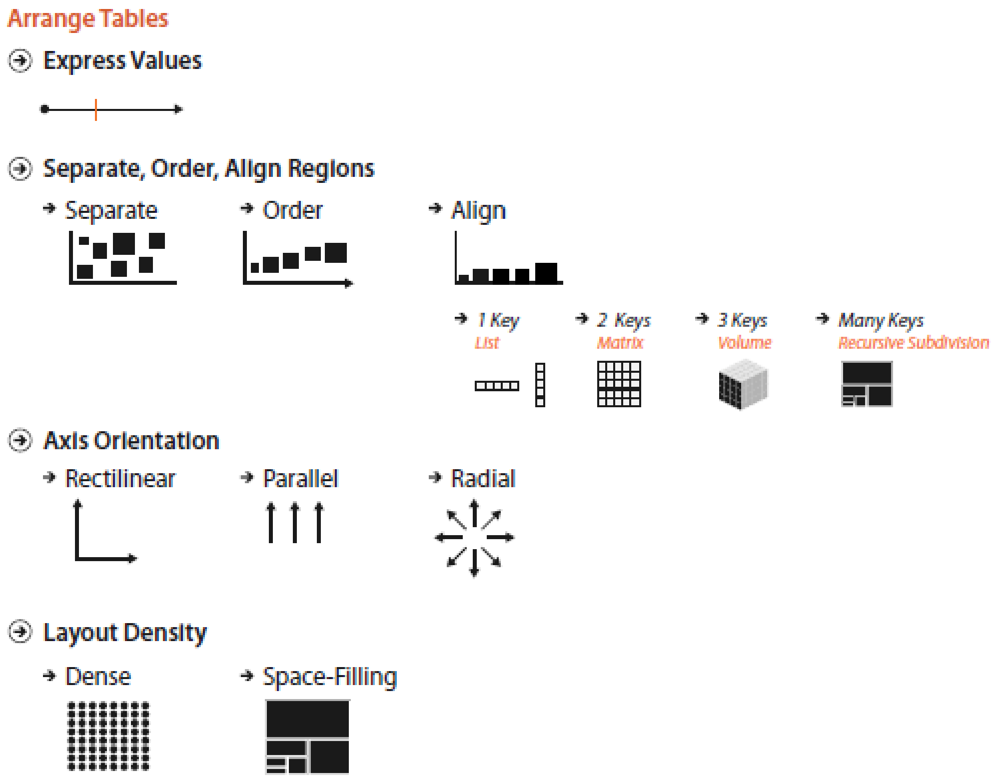
Chapter 7

10/20/20

Dataset Types: Tables

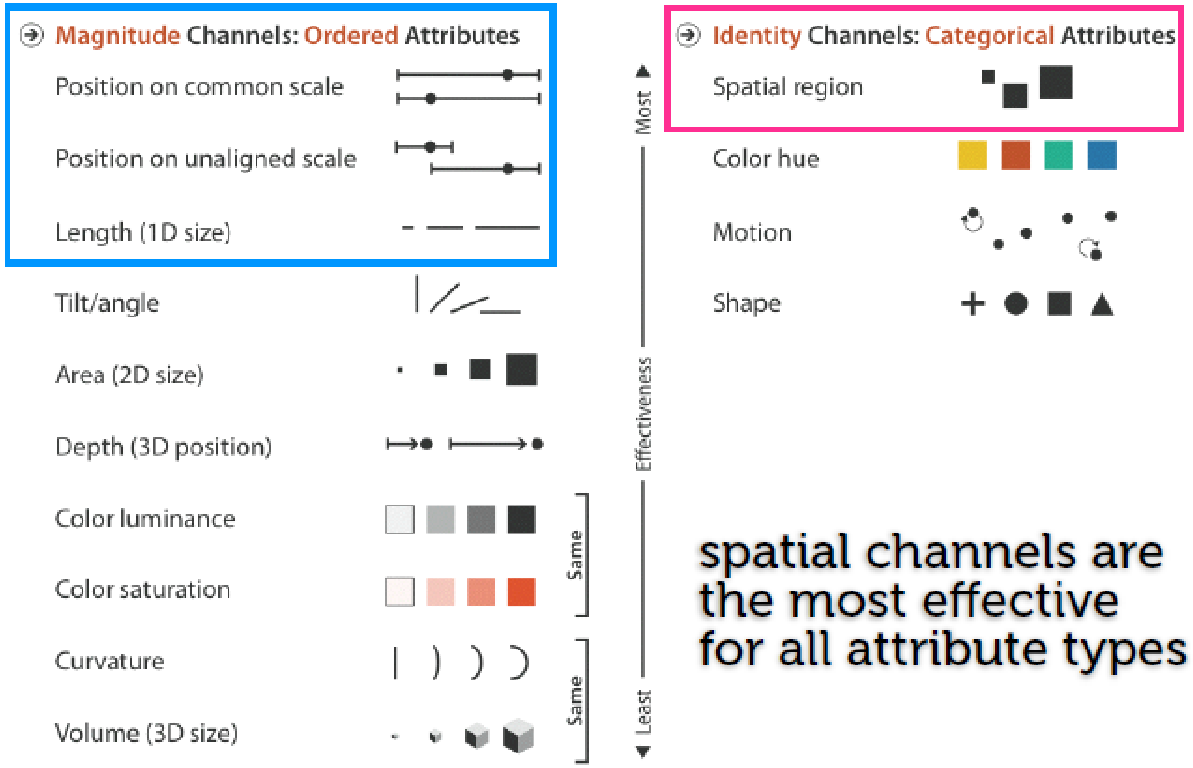


Design choices for arranging tables



Arrange

- Focus of all four design choices for tabular data



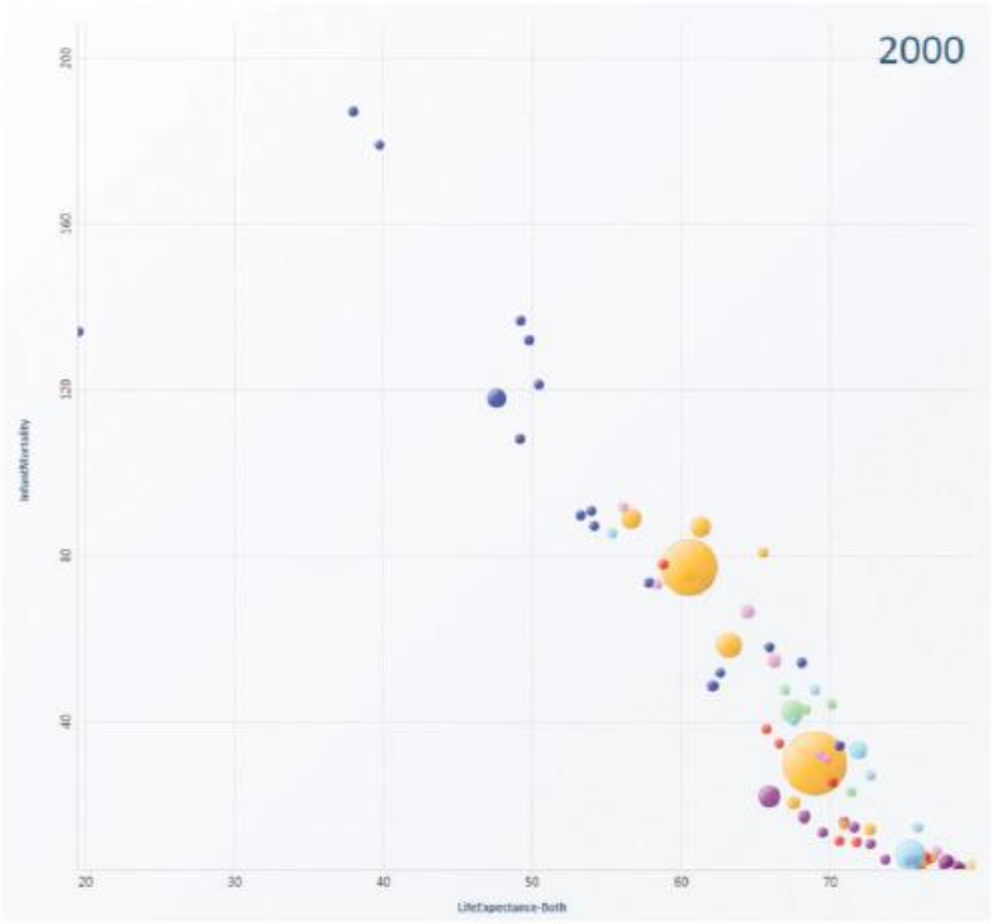
spatial channels are
the most effective
for all attribute types

Arrange by Keys and Values

- Key
  - Independent attribute that can be used as a unique index to look up items in a table
  - Categorical or ordinal
- Value
  - Dependent attribute; the value of a cell in a table
  - Can be all three of the types: categorical, ordinal, quantitative
  - Levels - unique values for a categorical or ordered attribute
- Examples:
  - Scatterplot – two value attributes
  - Bar charts – one key and one value attribute
  - Heatmap – two keys and one value attribute

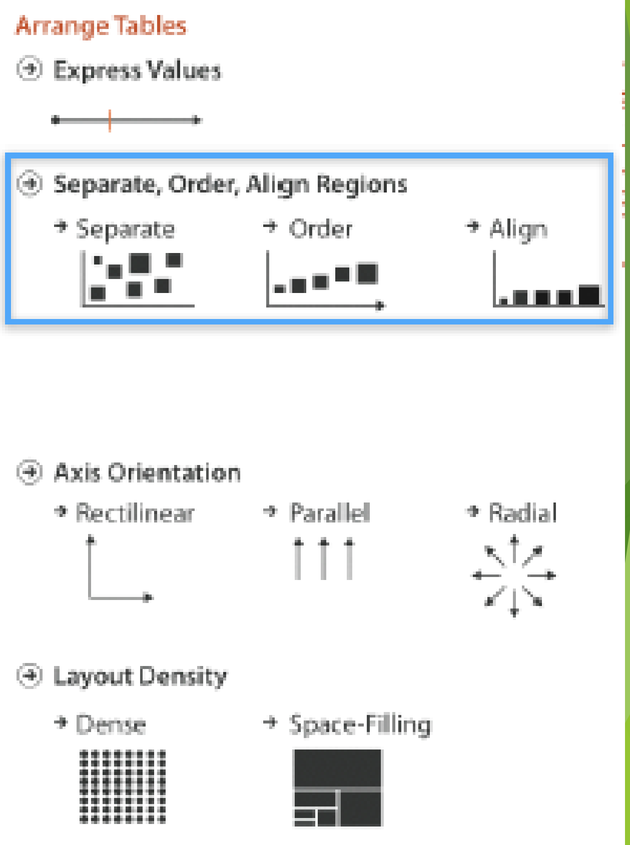
Express: Quantitative Values

- Plot example: Scatterplots
- Data: two quantitative values
- Task: find trends, clusters, outliers
- How: express values as point marks with horizontal and vertical spatial positioning
- Scalability: hundreds of points (items)
- Correlation: dependence between two attributes
  - Positive and negative correlation; indicated by lines
- Coordinate system (axes) and labels are important



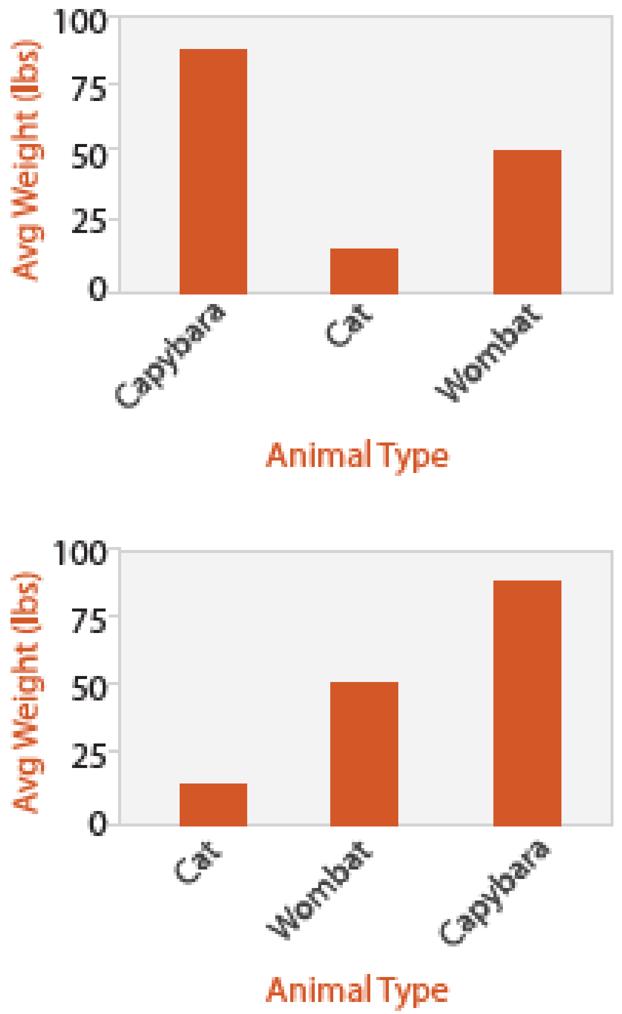
Separate, Order, and Align: Categorical Regions

- Spatial position – ordered magnitude visual channel
- Categorical attributes – unordered identity semantics
- Use spatial region, distinct contiguous bounded areas, to encode categorical attributes
- Break down the distribution of regions into three operations:
  - Separate (required) – done according to a categorical attribute
  - Align (optional) – done by some other attribute that is ordered
  - Order (required) – done by some other attribute that is ordered



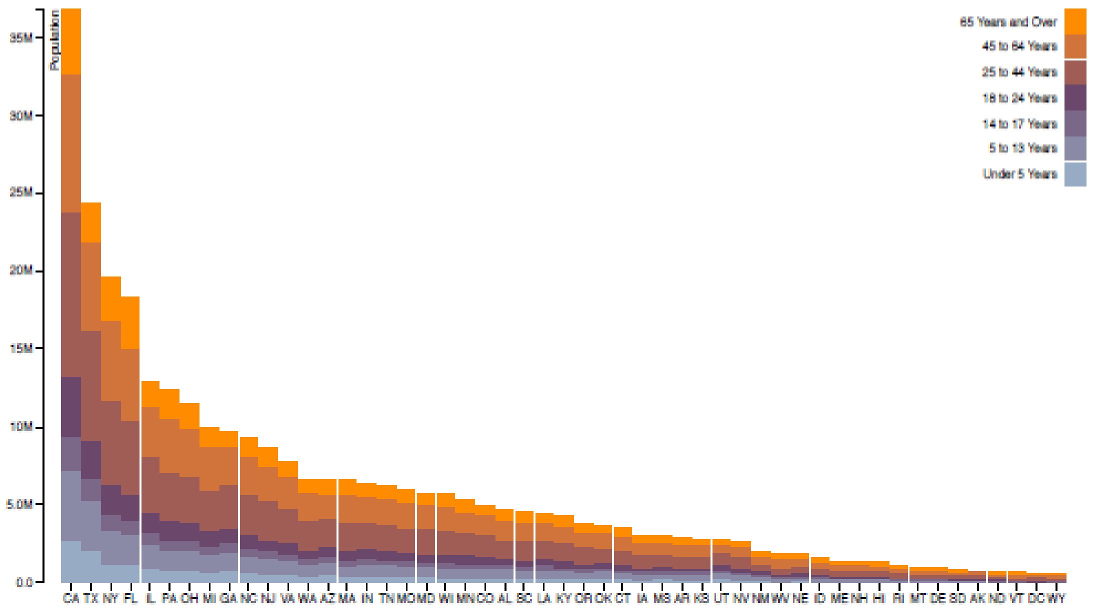
List Alignment: One Key Attribute

- Example: Bar Chart
- Data: one quantitative attribute, one categorical attribute
- Task: lookup and compare values
- How: line marks, horizontal position (categorical), vertical position (quantitative)
- Ordering criteria: alphabetical OR using quantitative attribute
- Scalability: distinguishability; hundreds



List Alignment: One Key Attribute

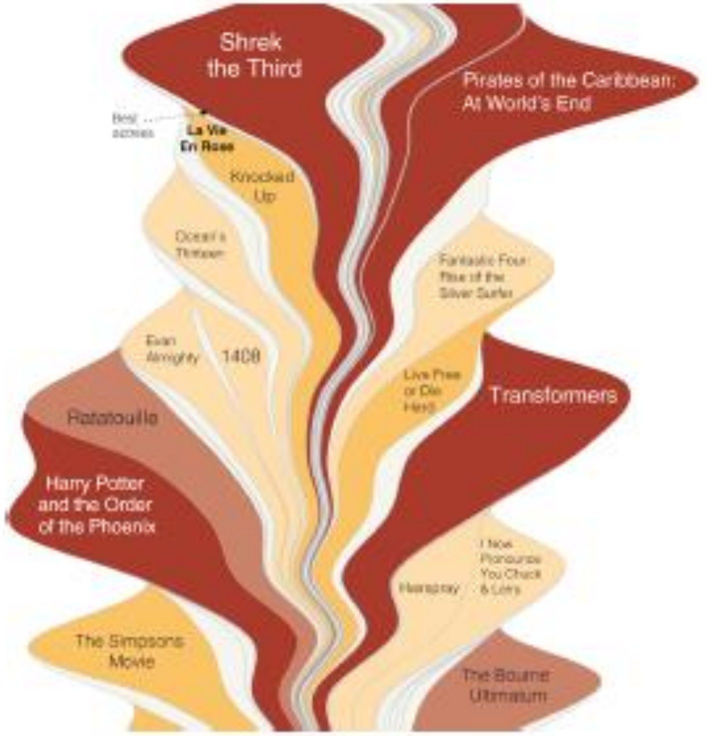
- Example: Stacked Bar Charts
- Data: multidimensional table: one quantitative attribute, two categorical attributes
- Task: lookup values, part-to-whole relationship, trends
- How:
  - line marks: position (both horizontal based on categorical and vertical (based on quantitative)
  - subcomponent line marks: length, color based on value attribute for each category of secondary key attribute
- Scalability: main axis (dozen to hundreds of levels), bar classes (stacked) (<12)



[Bostock, 2012]

List Alignment: One Key Attribute

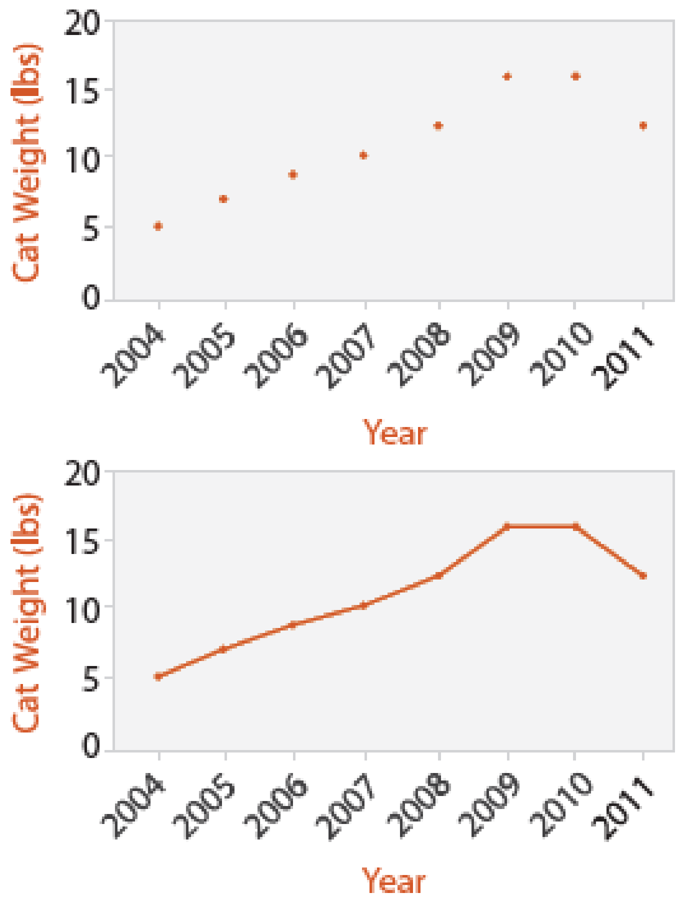
- Example: Steamgraphs
- Data: multidimensional table: one quantitative attribute (counts), one ordered key attribute (time), one categorical key attribute
- Task: analyze trends in time, find outliers
- How: derived position (height layer encodes counts) and geometry (categorical attribute across time)
- Scalability: more categories than stacked bar charts



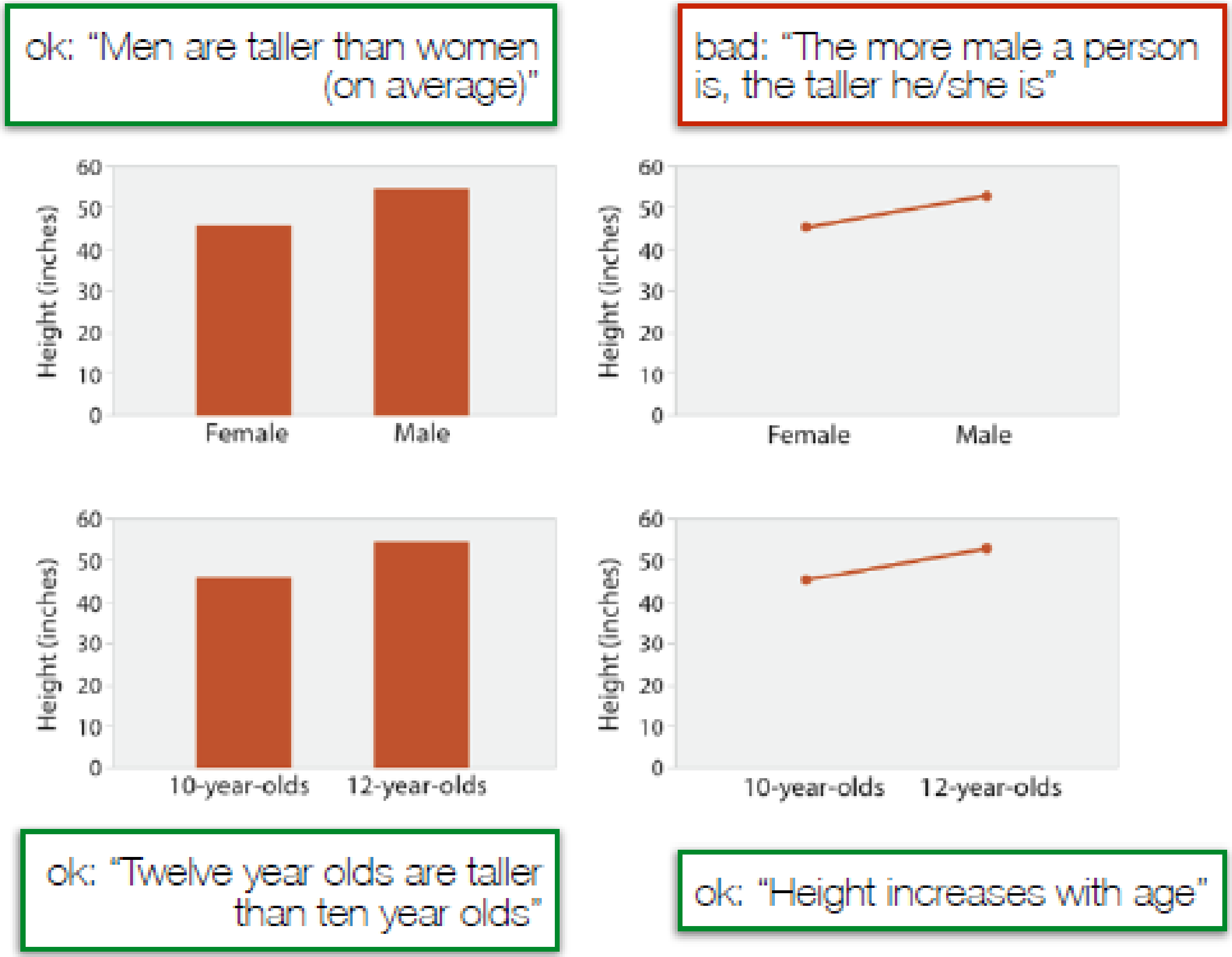
[Byron and Wattenberg, 2012]

List Alignment: One Key Attribute

- Example: Dot and Line Charts
- Data: one quantitative attribute, one ordered attribute
  - similar to scatterplots, but allows ordered attribute
- Task: lookup values, find outliers and trends
- How:
  - Dot charts: point mark and positions
  - Line charts: add connection mark (line)
- Scalability: key attribute can have hundreds of levels



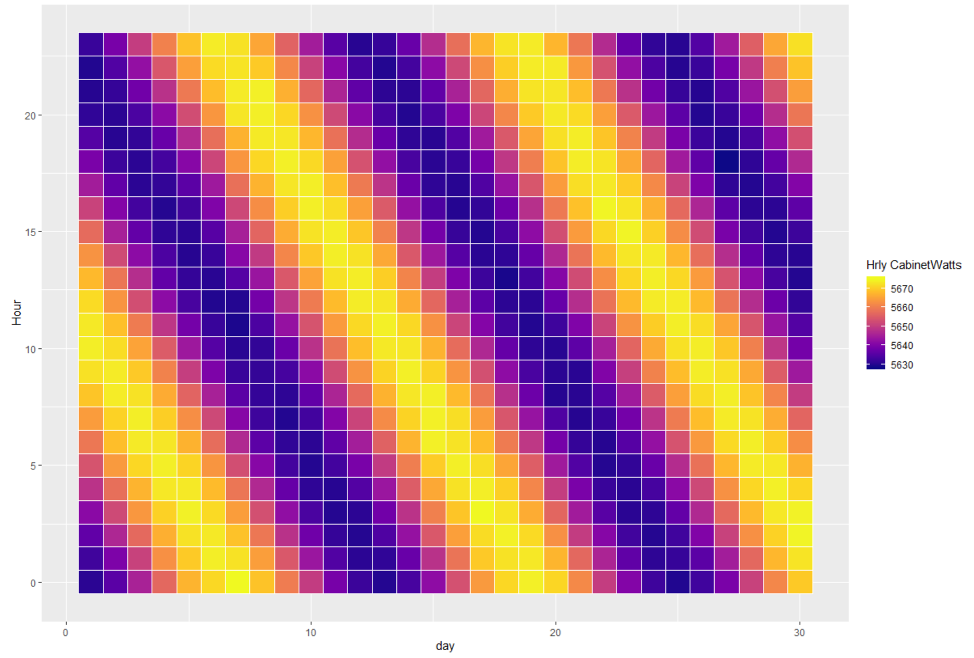
Proper Use of Line and Bar Charts



Don't use line charts for categorical attributes!

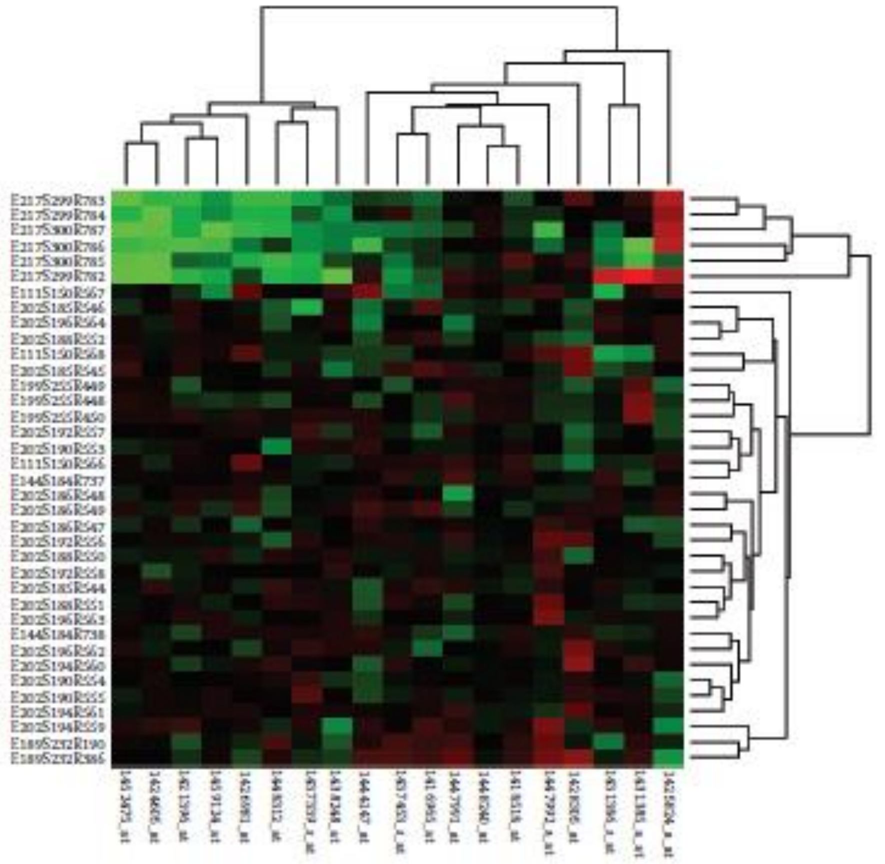
Matrix Alignment: Two Keys

- Example: Heatmaps
- Data: two categorical key attributes, one quantitative value attribute
- Task: find clusters, outliers, summarize
- How: area marks in grid, color encoding of quantitative attribute
- Scalability:
  - Items – one million
  - Categorical attribute levels – hundreds
  - Quantitative attribute levels – less than a dozen (colors)
- Note: be aware of colorblindness!



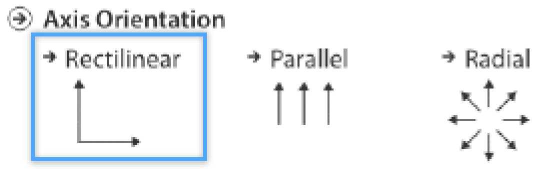
Matrix Alignment: Two Keys

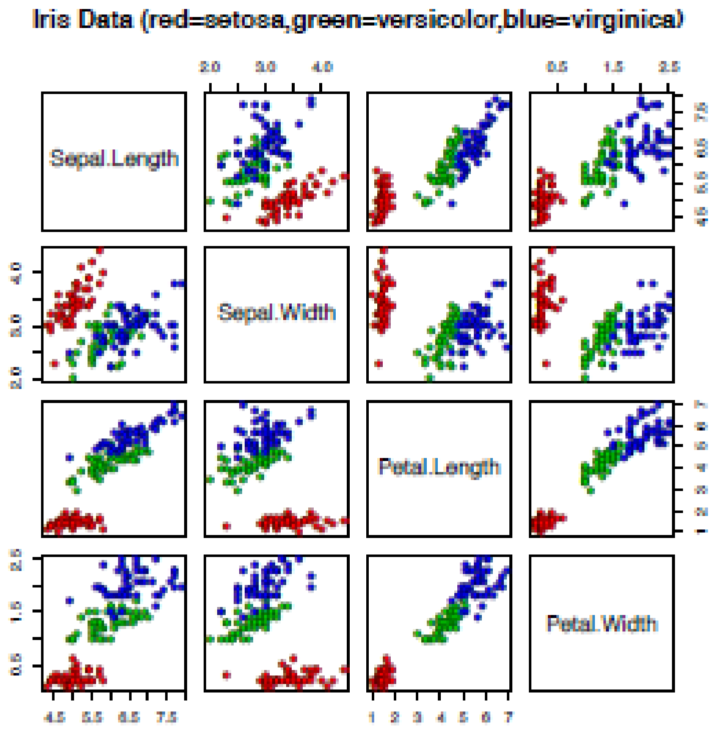
- Example: Cluster Heatmap
- Data: same as heatmap
- How: Area marks but matrix is ordered by two cluster hierarchies for table rows and columns
- Scalability: limited by the cluster dendrogram
- Dendrogram: a visual encoding of tree data with leaves aligned



Matrix Alignment: Small-Multiple Views

- Example: scatterplot matrix (SPLOM)
- Data: many quantitative attributes (table)
- Task: find correlations, trends, outliers
- How: scatterplot in 2D matrix alignment
- Scale:
  - Attributes: about one dozen
  - Items: dozens to hundreds
- Visualizations in a visualization: at a higher level, marks are themselves visualizations





https://commons.wikimedia.org/wiki/File:Anderson%27s_Iris_data_set.png#Creation

Spatial Axis Orientation

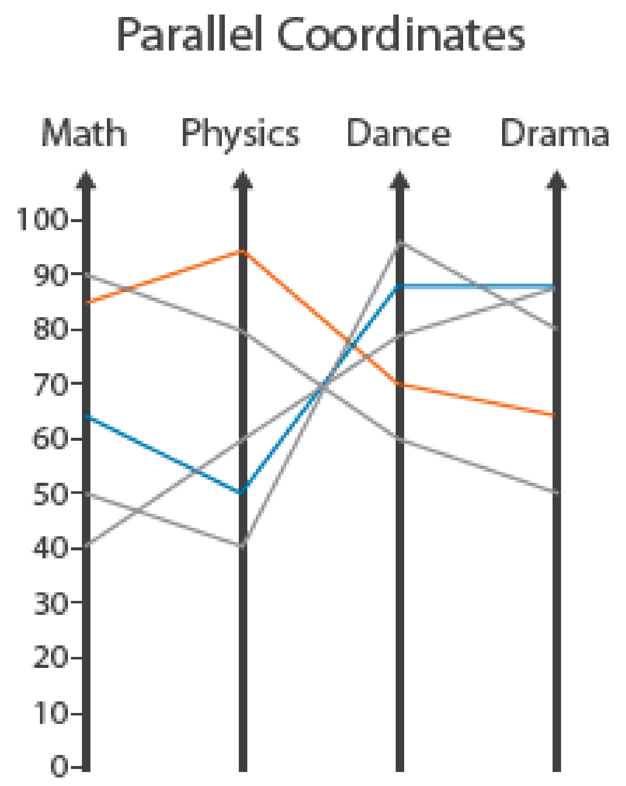
- Rectilinear layouts (horizontal and vertical axes) have been used so far to encode almost everything!
- Are there other possibilities?

Spatial Axis Orientation

- Rectilinear layouts (horizontal and vertical axes) have been used so far to encode almost everything!
- Are there other possibilities?
  - Parallel axes
  - Radial axes

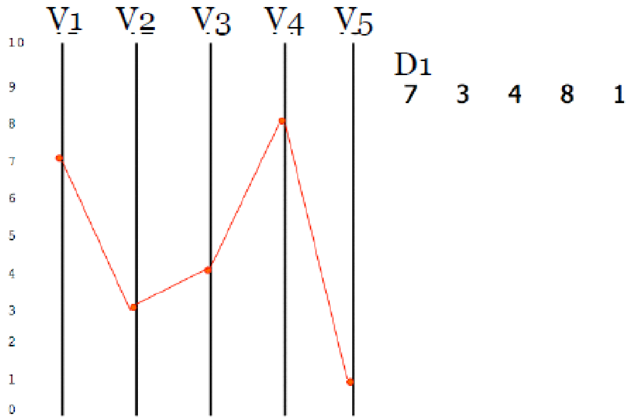
Parallel Coordinates

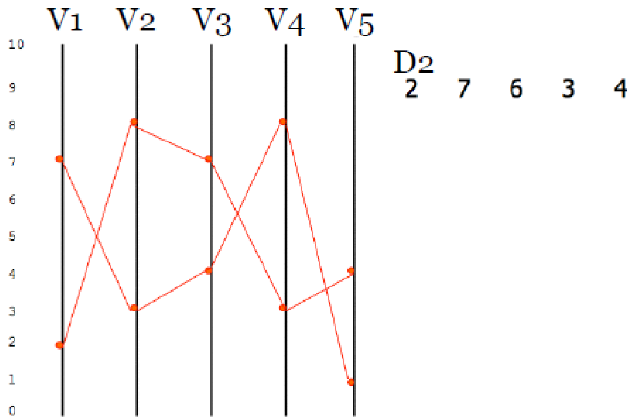
- Data: many quantitative attributes at once (using spatial position)
- Task: find trends, extremes, correlation
- How: vertical spatial position for each attribute, axis horizontally spaced (ordering horizontal axis is important), connection marks for identity (help visualize trends between particular values)
- Scalability: less than 40 attributes, hundreds of values
- Motivation: used for abstract task of checking for correlation between attributes

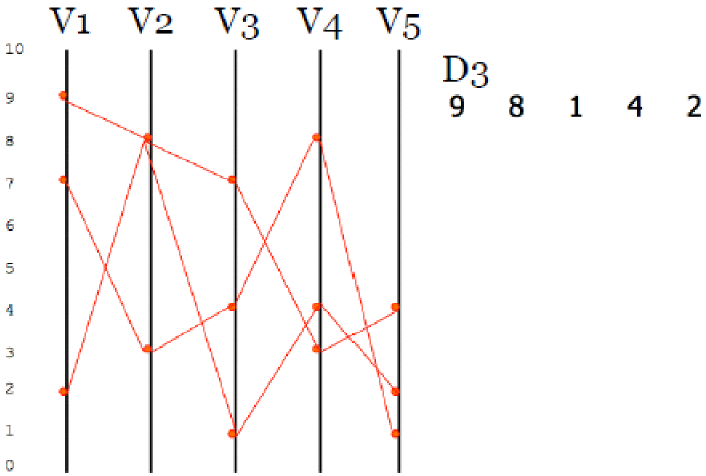


Example

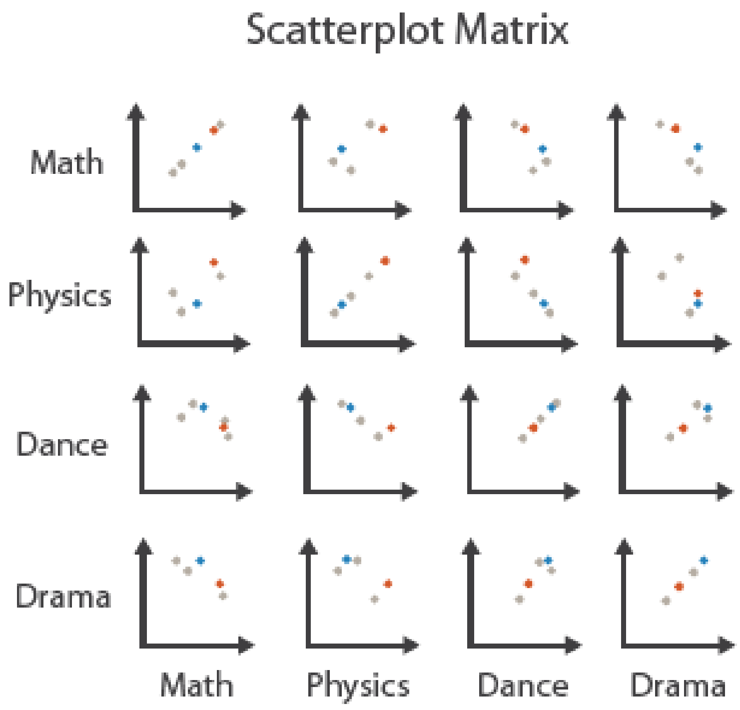
	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

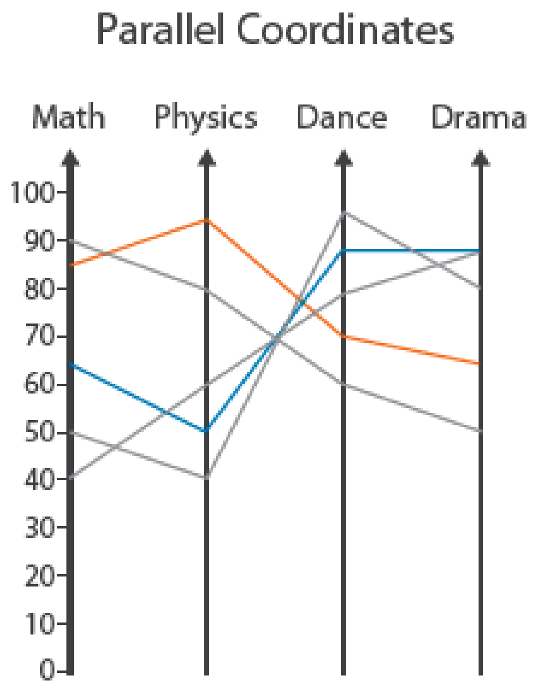




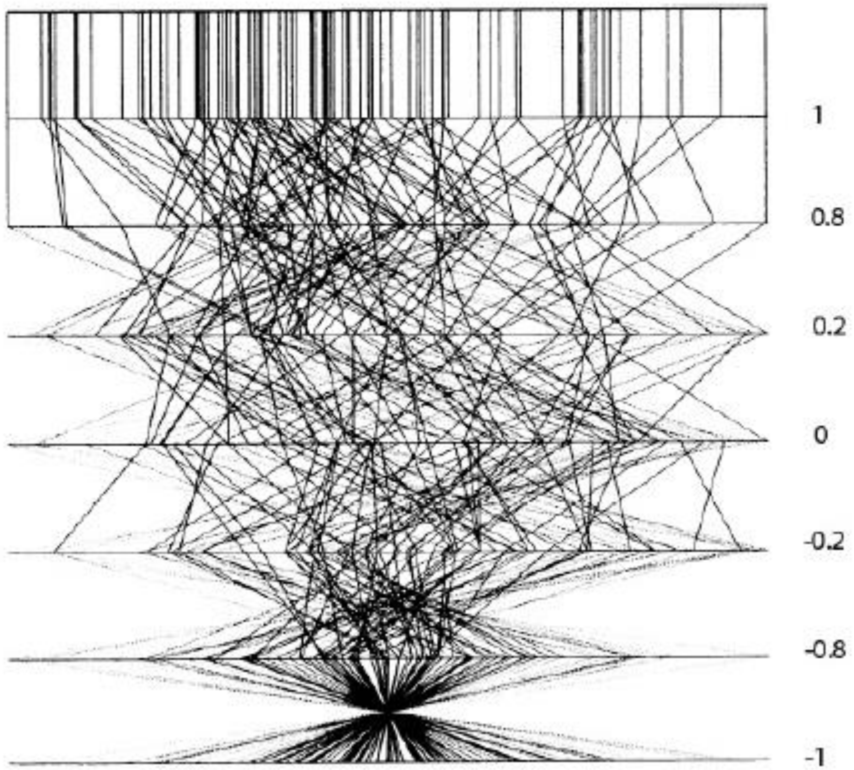


Comparing SPLOMs and Parallel Coordinates

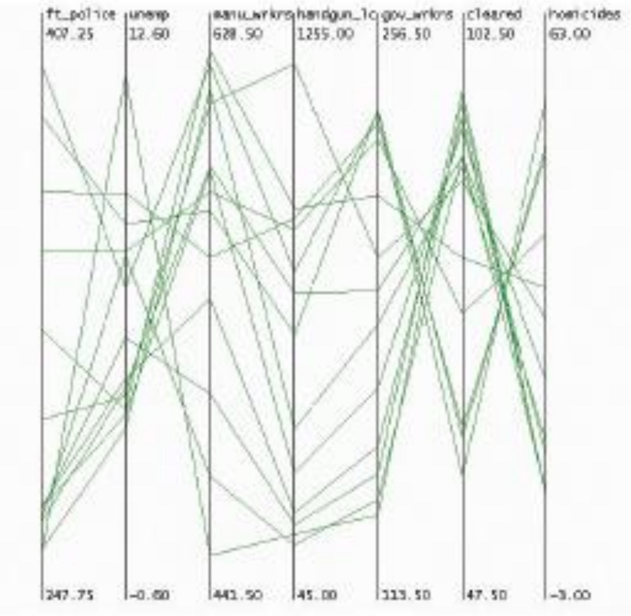


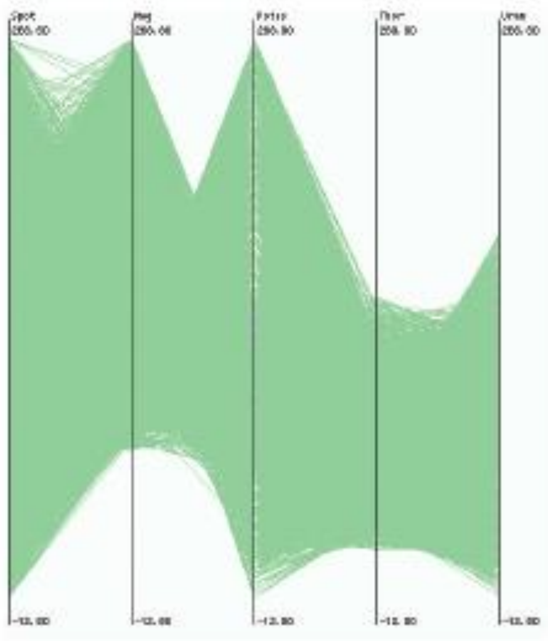


Correlation in Parallel Coordinates



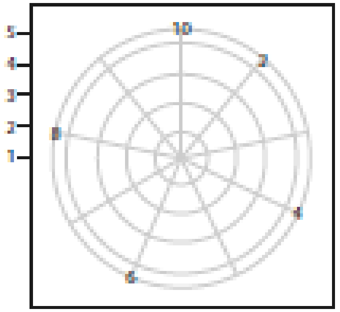
Overdraw in Parallel Coordinates

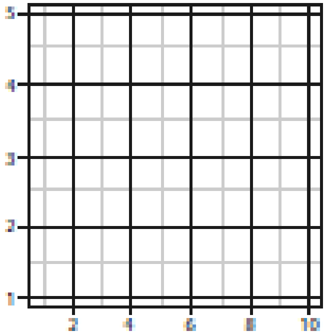


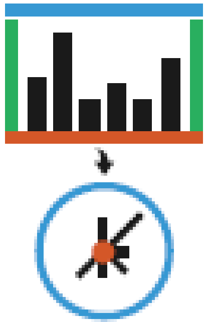


Radial Layouts

- Use Polar Coordinates
  - Angle and Length
  - Rectilinear and polar layouts same (mathematically)
  - Different, perceptually
  - What are the channel differences??

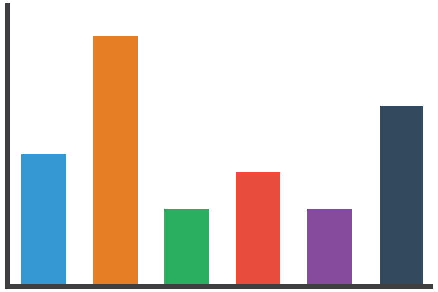


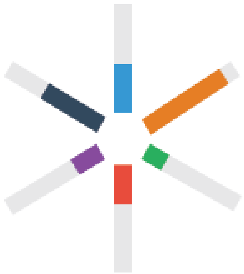




Radial Layouts

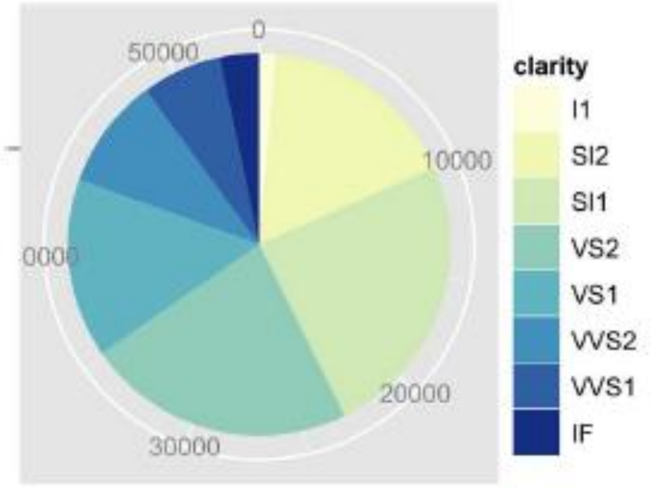
- Angle channel is less accurately perceived
- Angle channel is inherently cyclic
- More effective in visualizing periodicity of patterns!
- Misleading to encode non-periodic data with radial layouts!
- Misleading if two attributes have equal importance b/c radial layouts imply an asymmetry of importance between two attributes

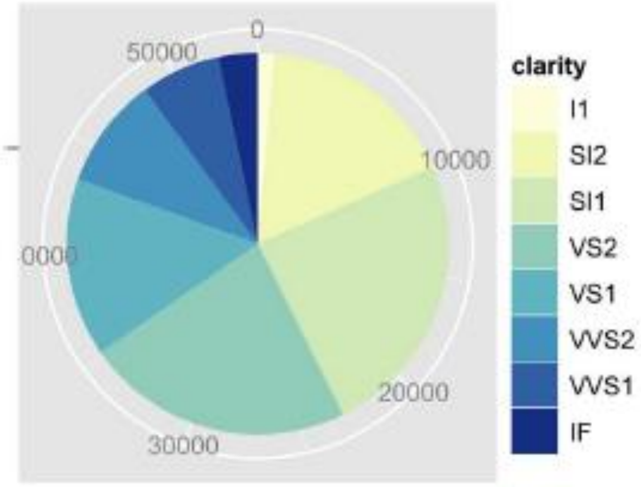




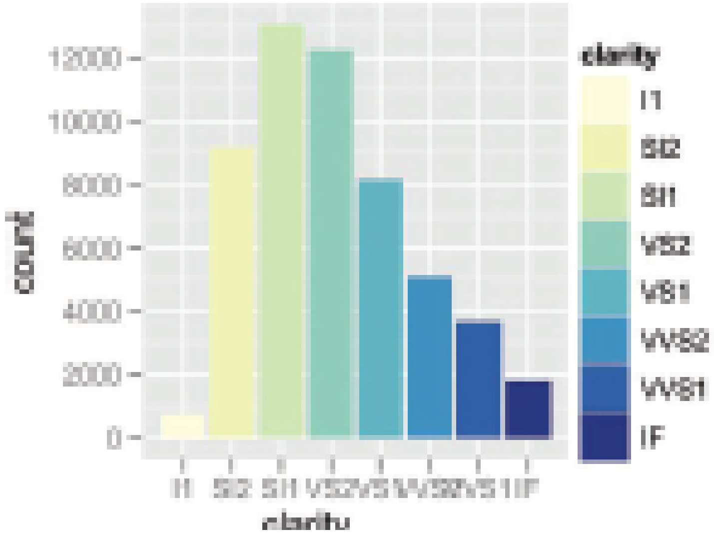
Pie Chart

- Data: one quantitative, one categorical
- Task: part-whole relationship
- How: area marks using angle channels, radial layout
- Scalability: dozens of categories
- Color channel?

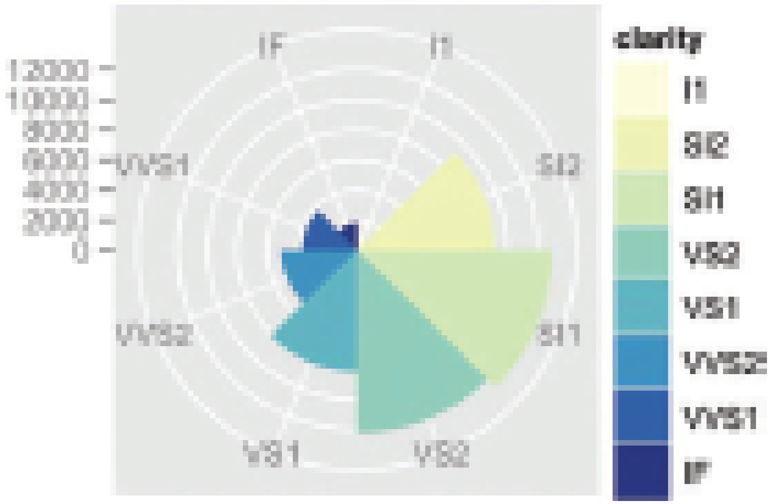




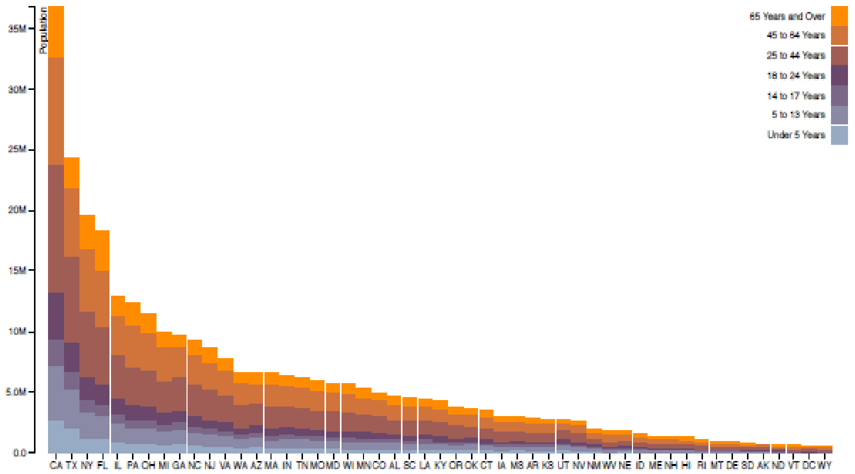
0
50000
10000
0000
20000
30000
clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF

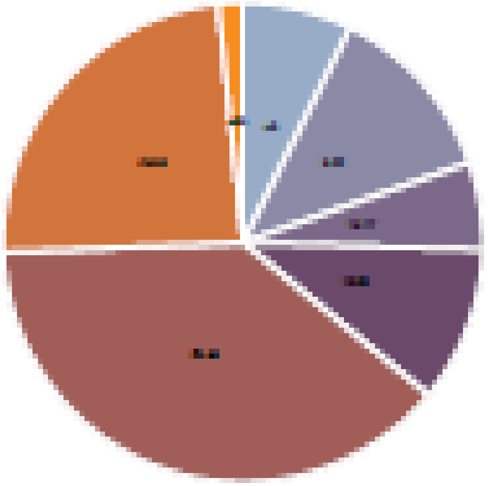


count
12000
10000
8000
6000
4000
2000
0
clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF

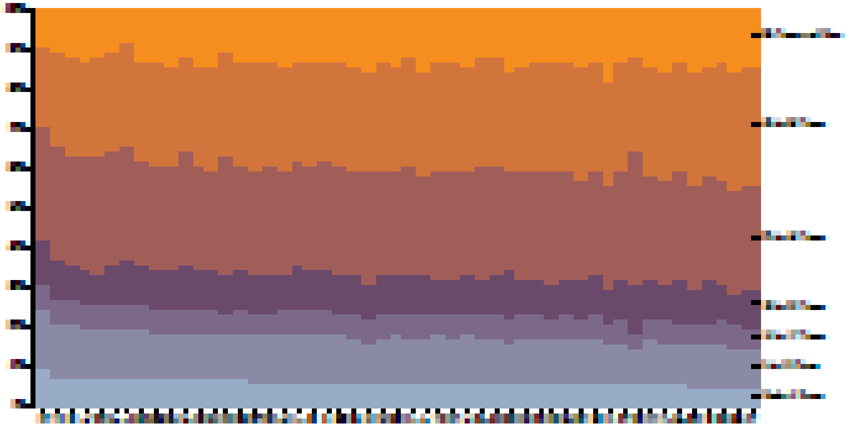


12000
10000
8000
6000
4000
2000
0
IF
I1
VVS1
SI2
VVS2
SI1
VS1
VS2
clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF





[Bostock, 2012]



Sources/Credits

- Tamara Munzner, Visualization Analysis & Design, A K Peters Visualization Series, CRC Press, 2014.
- Utah, Miriah Meyer, Visualization (2014).
- UMass Dartmouth, David Koop, Data Visualization (2015).