Data Exploration & Visualization

Module 6

Basic Charts

Dr. ZENG Wei

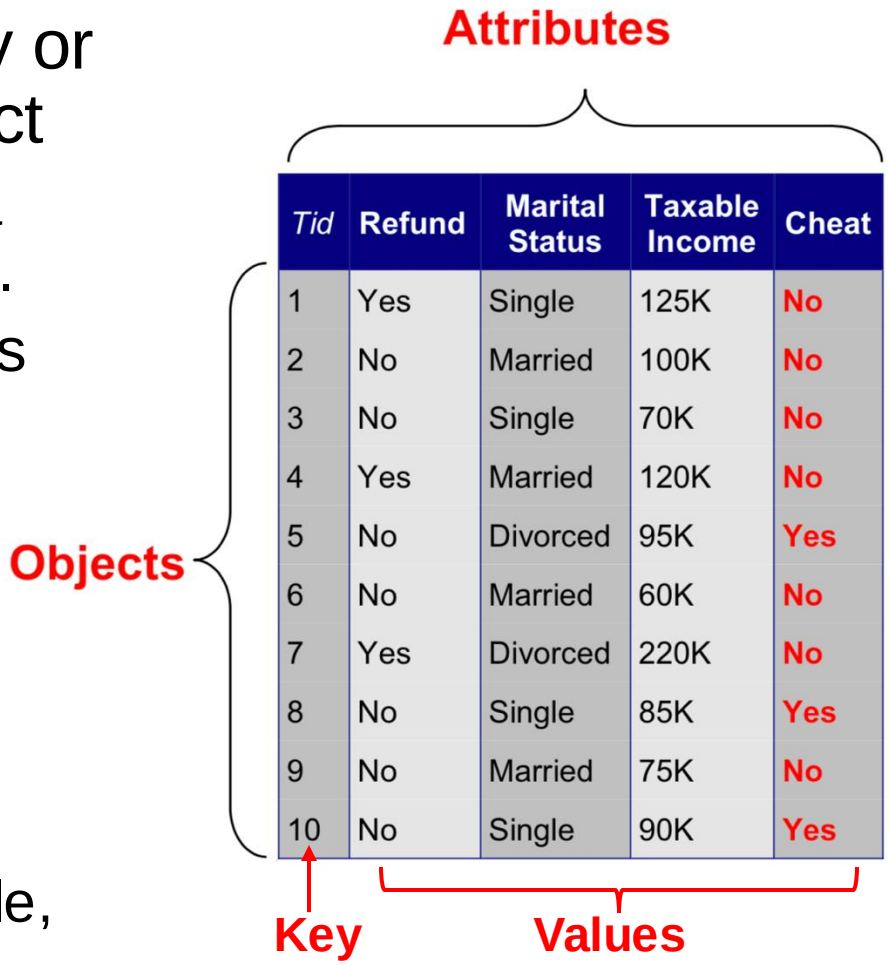
DSAA 5024

The Hong Kong University of Science and Technology (Guangzhou)

Tabular Data

- An attribute is a property or characteristic of an object
  - Examples: eye color of a person, temperature, etc.
  - Attribute is also known as variable, field, feature or characteristic

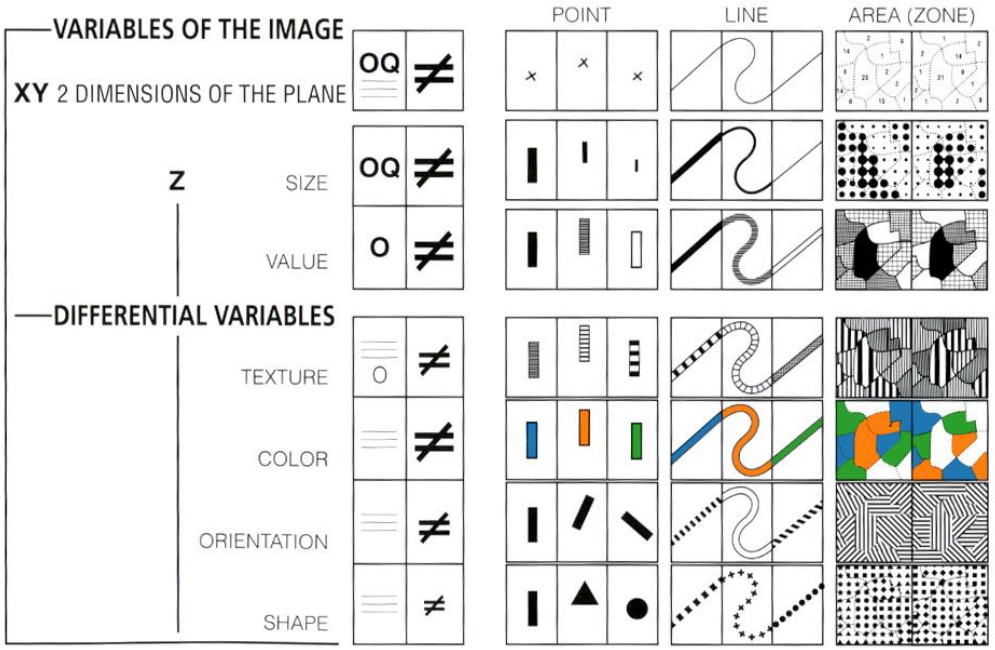
- A collection of attributes describe an object
  - Object is also known as record, point, case, sample, entity, or instance



Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Bertin's Semiology of Graphics

- Geometric primitives: **marks**
  - Points, lines, areas, volumes
- Attributes: visual/retinal variables (**channels**)
  - Parameters control mark appearance
  - [x, y]
    - Position
  - [z]
    - Size, shape,
    - Greyscale, color
    - Texture, orientation
- Data types
  - Nominal, ordinal, quant



Taxonomy

- Taxonomy of chat types

Category	subtypes
Area	Area chart, proportional area chart…
Bar	Bar chart, circular bar chart, grouped bar chart, stacked bar chart…
Circle	Belt chart, donut chart, pie chart, sector graph…
Diagram	Flow chart, sankey diagram, timeline…
Distribution	Histogram, distribution curve, box-plot…
Tree and network	Graph, matrix, hive graph, tree, treemap…
Grid / matrix	heat map

M. A. Borkin *et al.*, 2013.

Taxonomy

- Taxonomy of chat types

Category	subtypes
Line	Contour graph, density graph, line graph…
Map	Flow map, geographic map, street map, statistic map, choropleth map
Point	Dot plot, scatter plot…
Table	Table, text chart
Text	Phrase net, word cloud, word tree
SciVis	Surface rendering, volume rendering…

M. A. Borkin *et al.*, 2013.

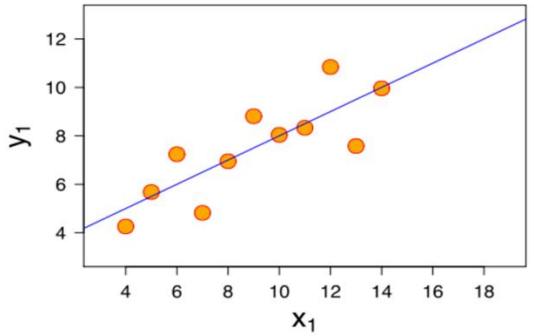
Data Exploration & Visualization

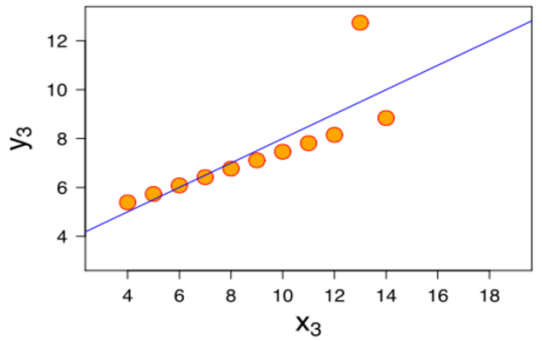
Module 6: Basic Charts

- Point charts
  - scatterplot, multi-class scatterplot
- Line charts
  - line chart, bar chart, stacked bar chart
- Area charts
  - bubble chart, pie chart
- Composite charts

Scatterplot

- Data
  - 2 quantitative attributes
  - No keys, only values
- Mark
  - points
- Channels
  - Horizontal + vertical positions
- Tasks
  - ***Express*** values
    - Quantitative attributes
  - Find trends, outliers, distributions, correlations, clusters

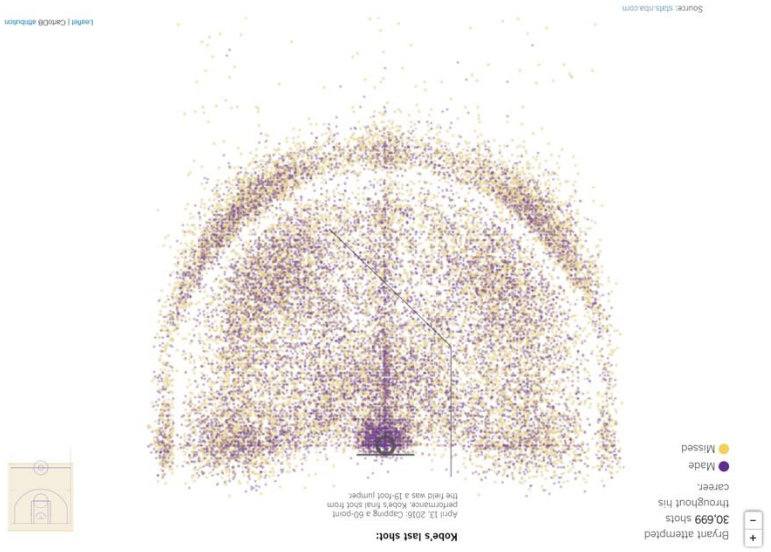


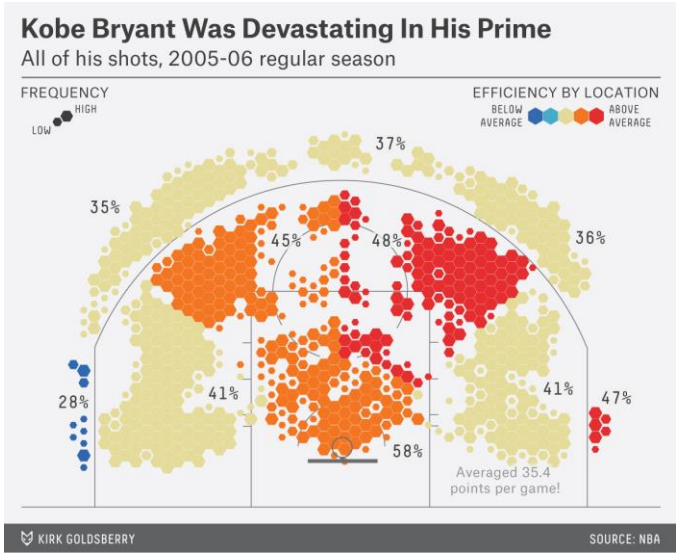


Credit: Tamara Munzner

Scatterplot: Aggregation

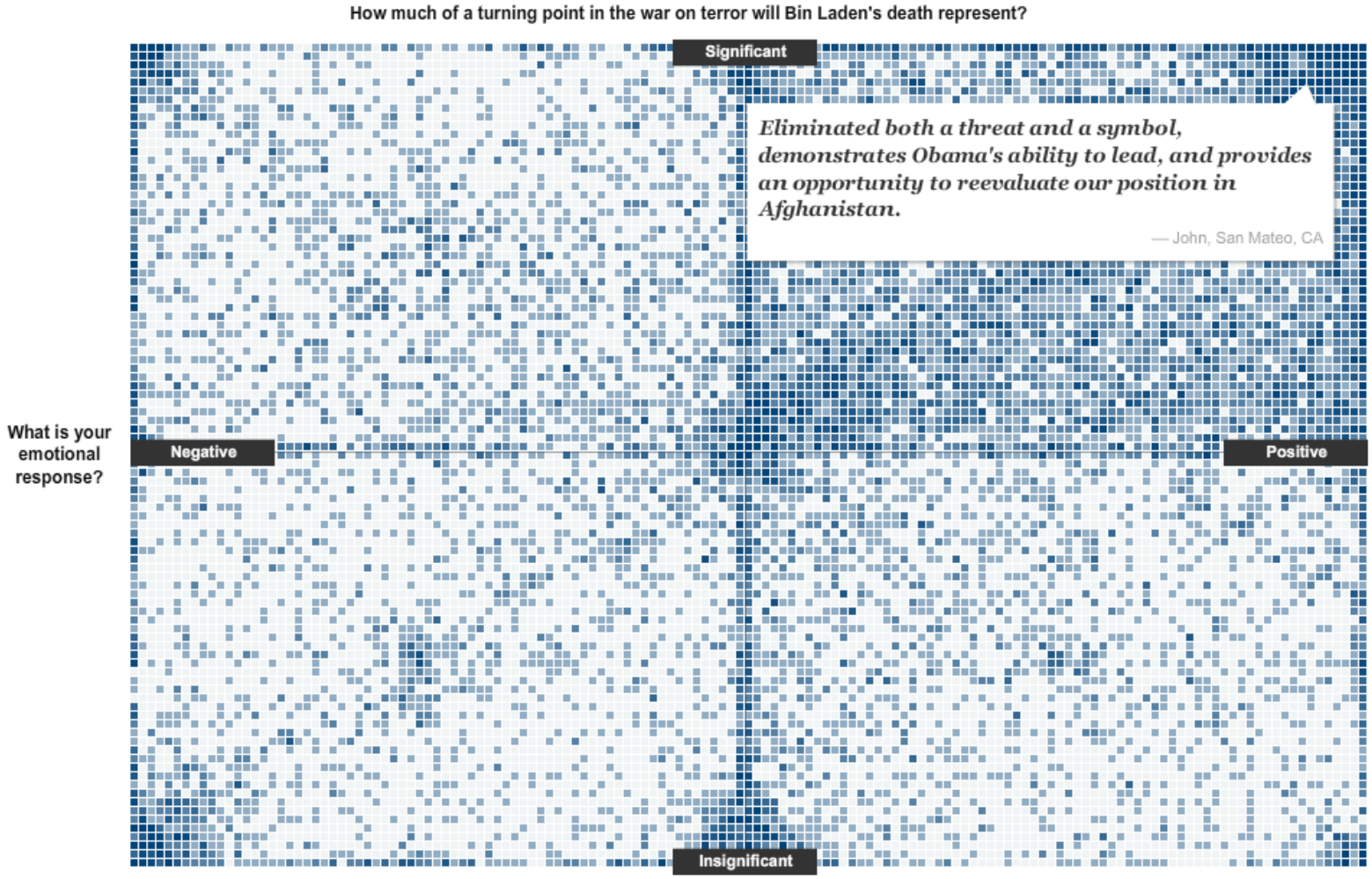
- Scalability (in terms of visual clutter)
  - hundreds of items
- What if too many items?
  - Aggregation
    - Drawback 1: loss of details – applicable when individual data item is not important
    - Drawback 2: Not feasible when there are multi-class attributes





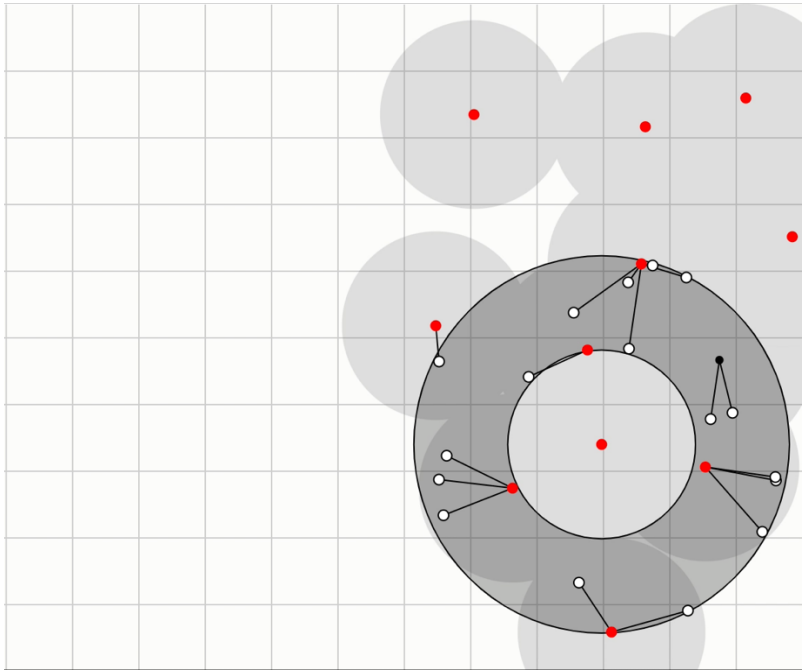
Scatterplot: Aggregation

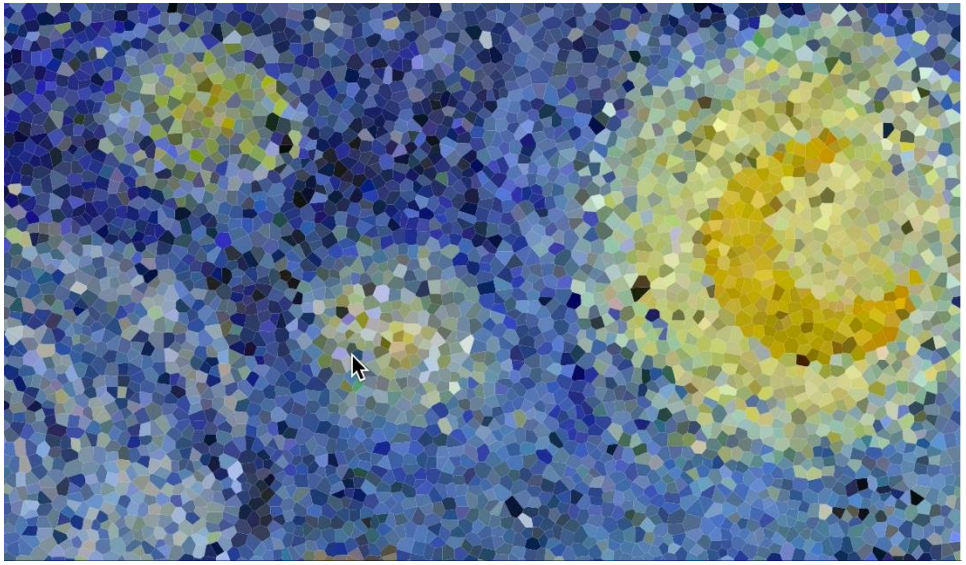
- ‘The Death of a Terrorist: A Turning Point?’ New York Times



Scatterplot: Sampling

- Scalability (in terms of visual clutter)
  - hundreds of items
- What if too many items?
  - Sampling
    - Be careful - choose appropriate sampling methods

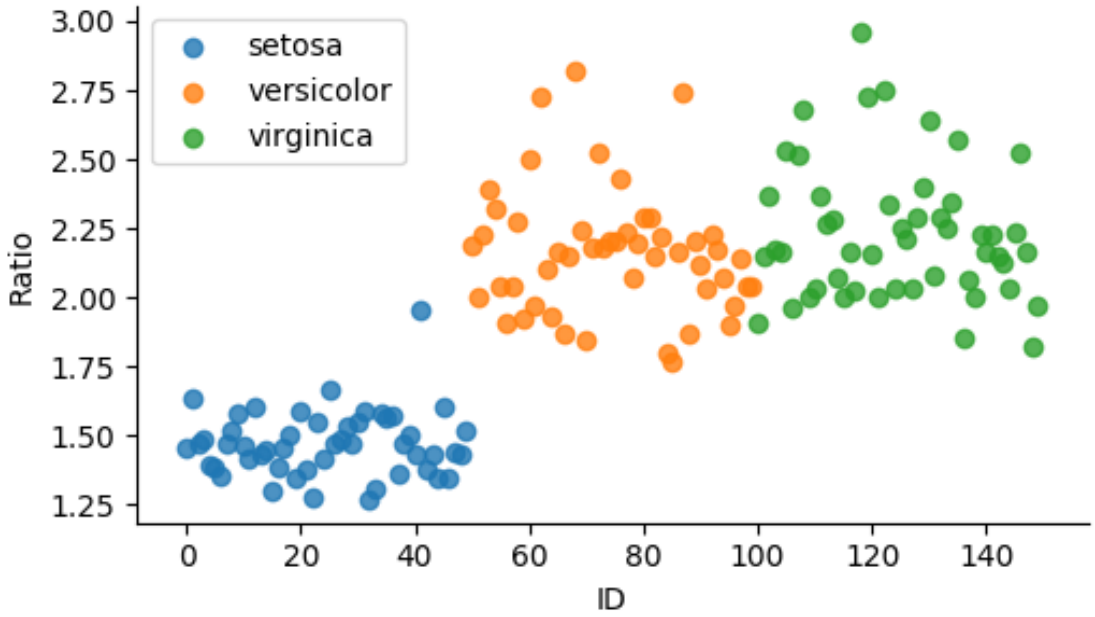




Visualizing Algorithms: Poisson-disc sampling
Credit: Mike Bostock https://bost.ocks.org/mike/algorithms/

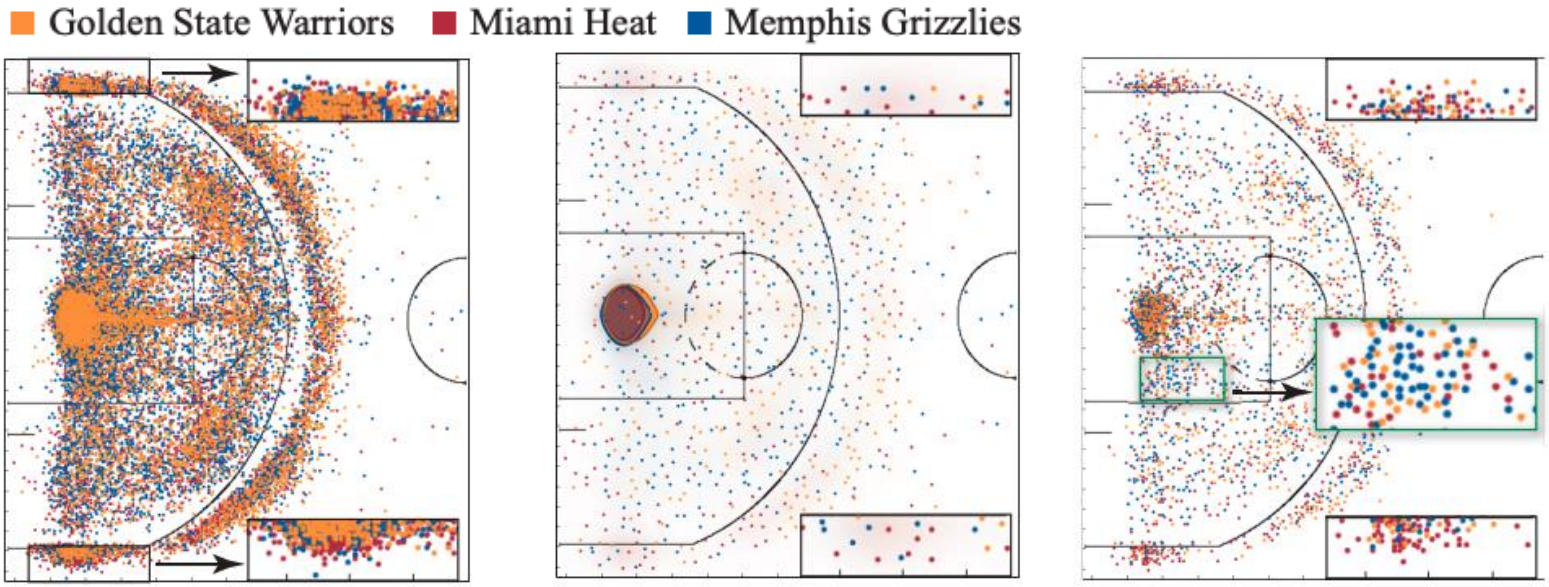
Multi-class Scatterplot

- Data
  - 2 quantitative attributes
  - 1 categorical attribute
  - One key, two values
- Mark: points
- Channels
  - Horizontal + vertical positions
  - Color
- Tasks
  - Clusters, comparison



Multi-class Scatterplot

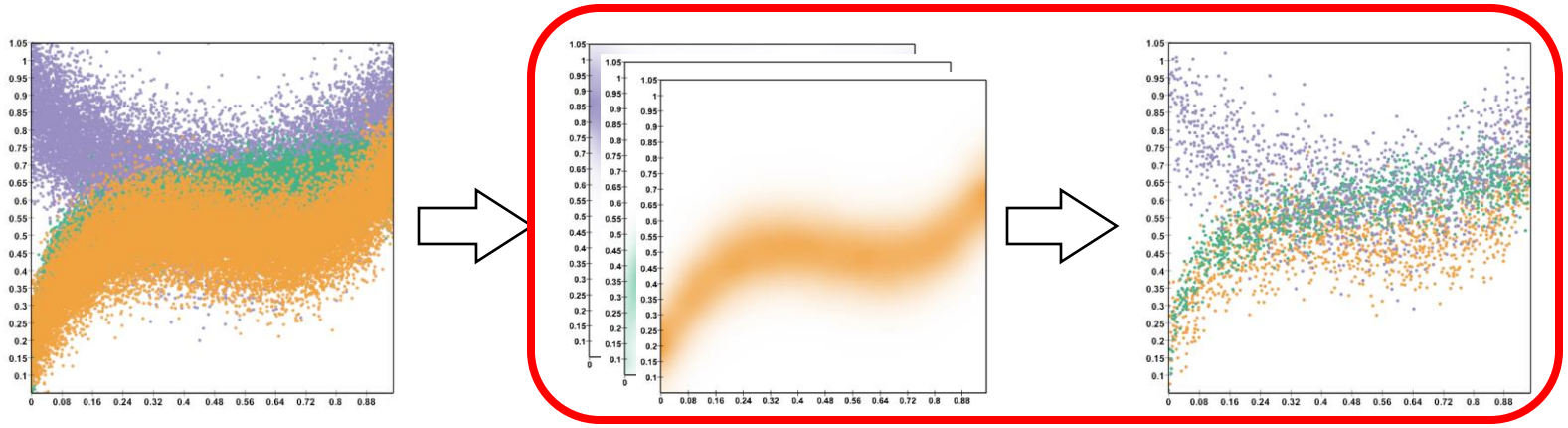
- Scalability
  - Hundreds of items
- What if too many items?
  - Splatterplot (sampling + density map)
  - Multi-class Sampling



Splatterplots 2013

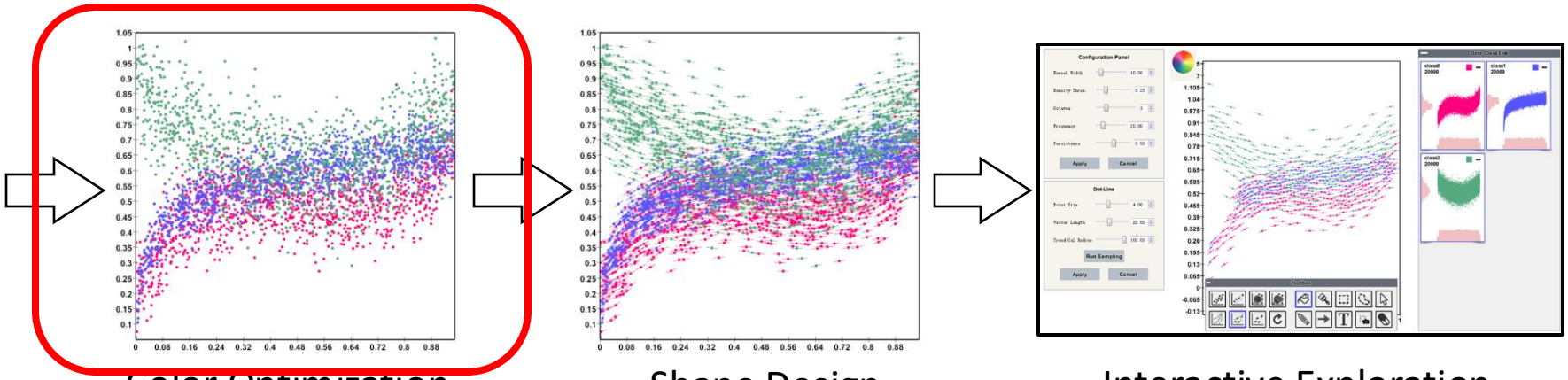
Multi-class sampling 2014

Multi-class Scatterplot: Sampling



Point Density Estimation

Point Resampling



Color Optimization

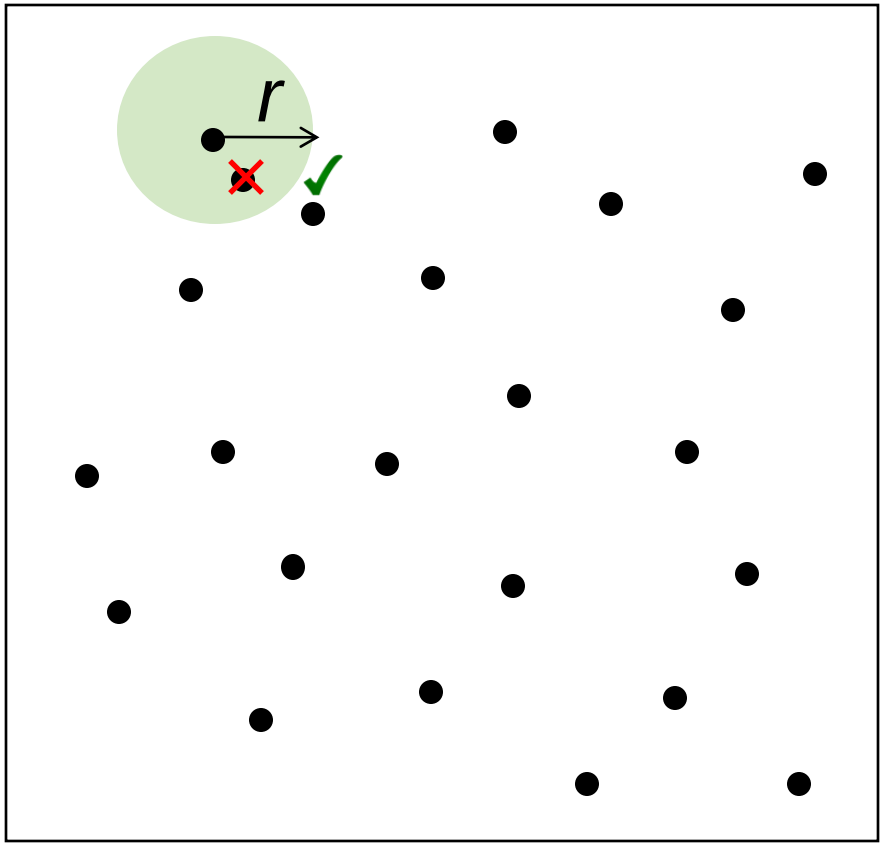
Shape Design

Interactive Exploration

H. Chen et al., "Visual Abstraction and Exploration of Multi-class Scatterplots," IEEE TVCG, vol. 20, no. 12, pp. 1683-1692, 2014.

Multi-class Scatterplot: Sampling

- Single-class sampling

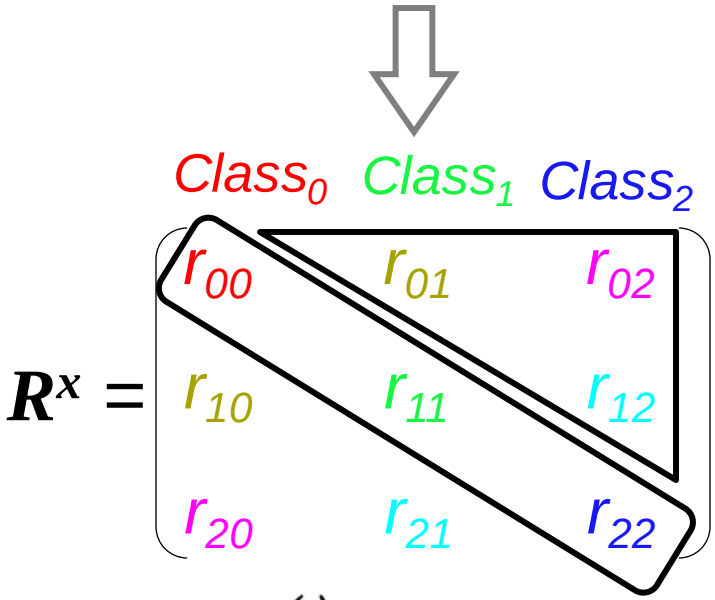


r : minimum distance away from each other

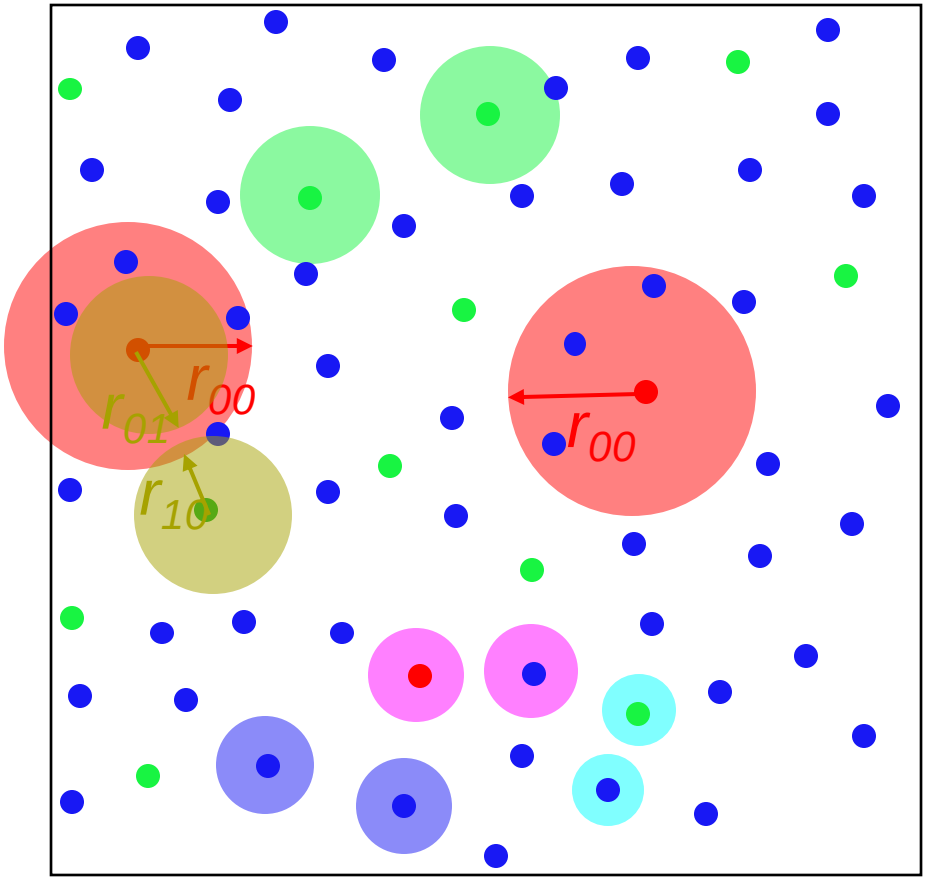
Multi-class Scatterplot: Sampling

- Multi-class blue noise sampling

r : minimum distance away from each other



$$r_{ii} = \frac{\omega}{\hat{f}_i(x)} \qquad \omega = \frac{r}{\varphi}\bar{f}$$



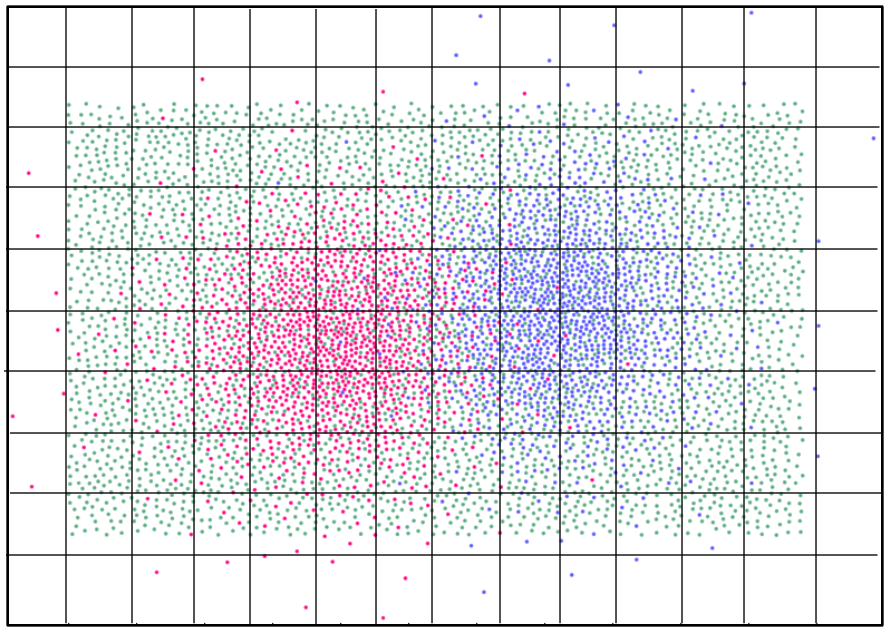
[Wei et al., 2010]

Multi-class Scatterplot: Sampling

- Point color optimization

Color set in the CIELAB color space:

$$C = \{C_1, C_2, ..., C_n\}$$



Objective function: $E_{\cos t} = \sum_{m=1}^{M} \beta_m \sum_{i,j<n,i<j} \alpha_{m,i,j} \left| C_i - C_j \right|$

M : the account of the divided blocks

β_m : the *inter-class weight* for the *m*-th block

$$\beta = \sum_{i=0}^{n} \overline{f_i}$$

$\alpha_{m,i,j}$: the *intra-class weight* for the *m*-th block

$$\alpha_{i,j} = e^{-\left|\overline{f_i} - \overline{f_j}\right|}$$

Color distance constraint:

$$E_{penalty}(C_i, C_j) = \max(0, 1 - \frac{\left|C_i - C_j\right|}{d})$$

$$\arg\min_{C} \left[-E_{\cos t} + k \sum_{i,j<n,i<j} E_{penalty}(C_i, C_j) \right]$$

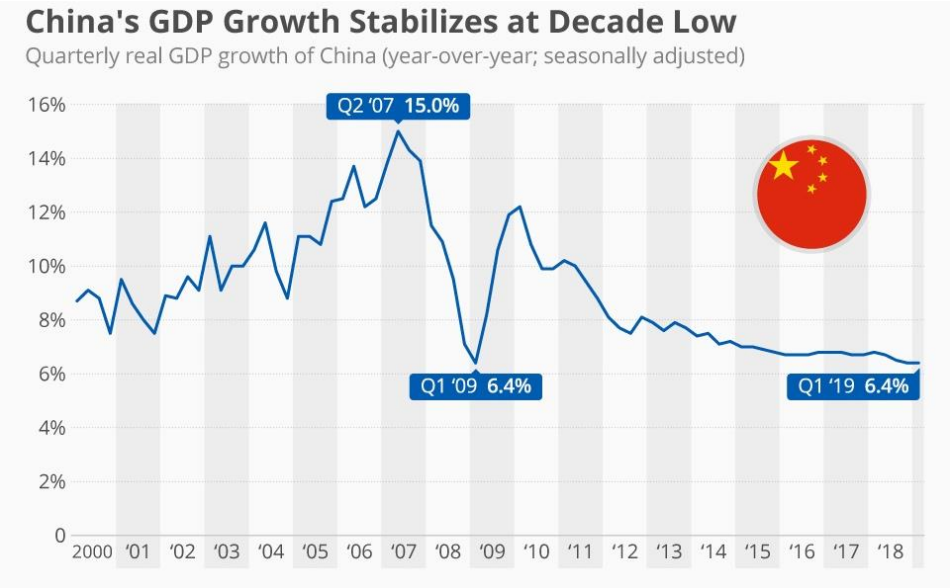
Data Exploration & Visualization

Module 6: Basic Charts

- Point charts
  - scatterplot, multi-class scatterplot
- Line charts
  - line chart, bar chart, stacked bar chart
- Area charts
  - bubble chart, pie chart
- Composite charts

Line chart

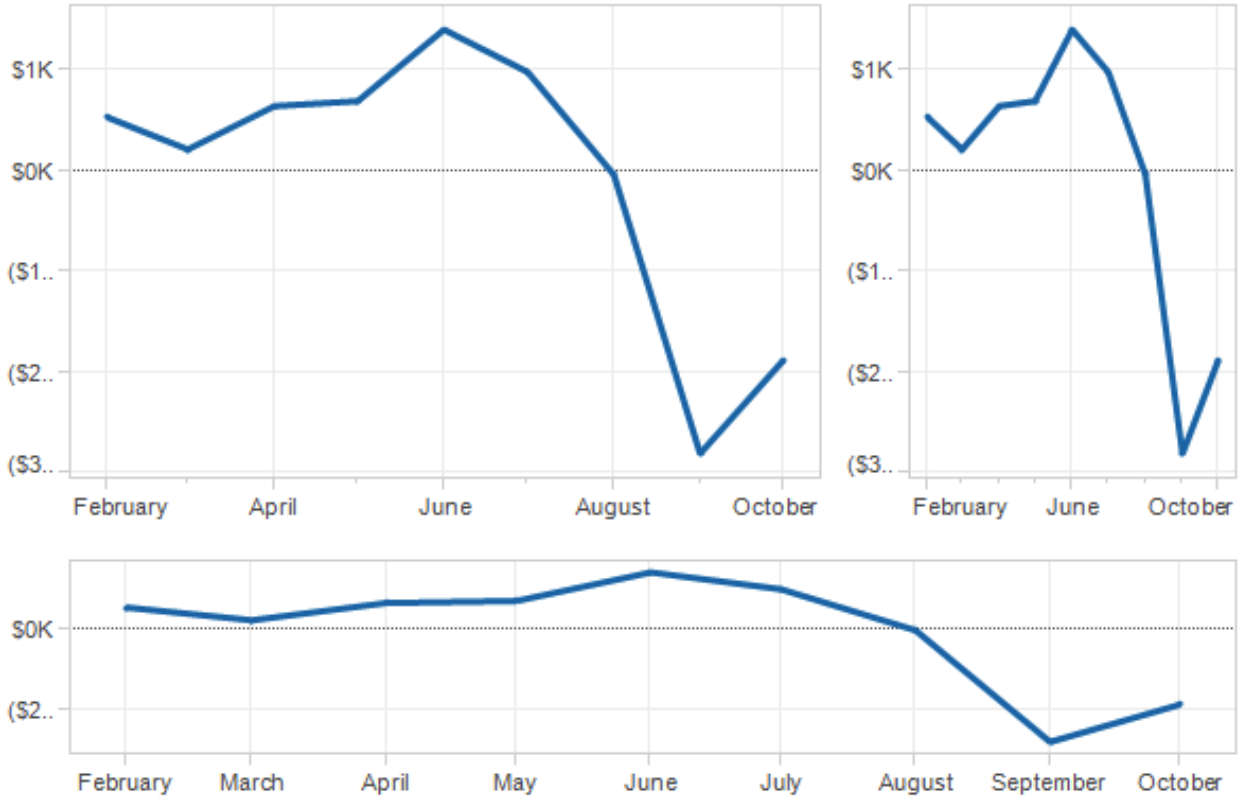
- Data
  - 2 quantitative attributes
  - one key, one value
- Mark
  - Points
  - Line connects marks between them



- Channels
  - Aligned lengths to express quantitative value
  - Separated and ordered by key attribute into horizontal regions
- Tasks
  - Find trend
- Scalability
  - hundreds of key and value levels

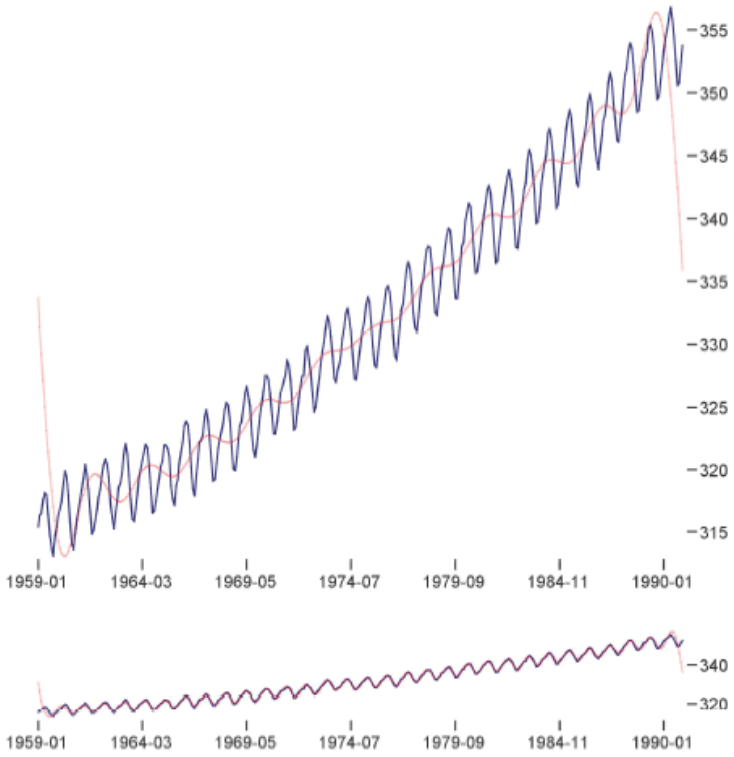
Line chart

- The same data can look very different
  - Aspect ratio influences the perception of the data



Line chart

- What is the optimal aspect ratio?
  - Banking to 45 degrees [Cleveland, 1988]
  - Multi-scale banking to 45 degrees [Heer & Agrawala, 2006]
  - Local orientation resolution [Wang et al., 2018]



$$\sum_i \frac{|\theta_i(\alpha)|}{n} = 45^\circ$$

where α : aspect ratio of the chart

$\theta_i(\alpha) = tan^{-1}(s_i/\alpha)$

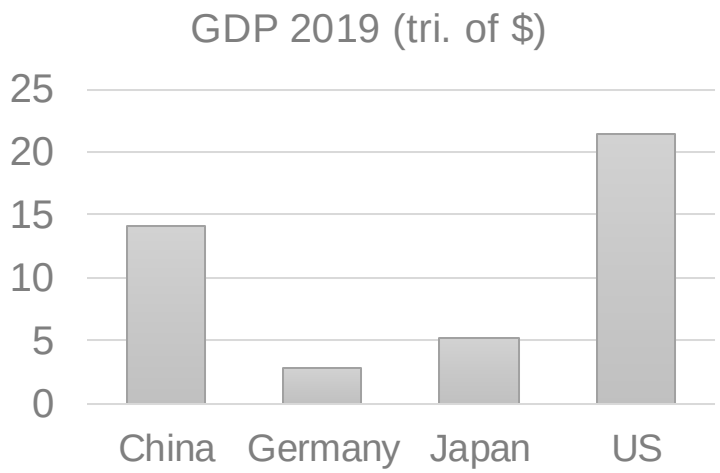
s_i: a line segment

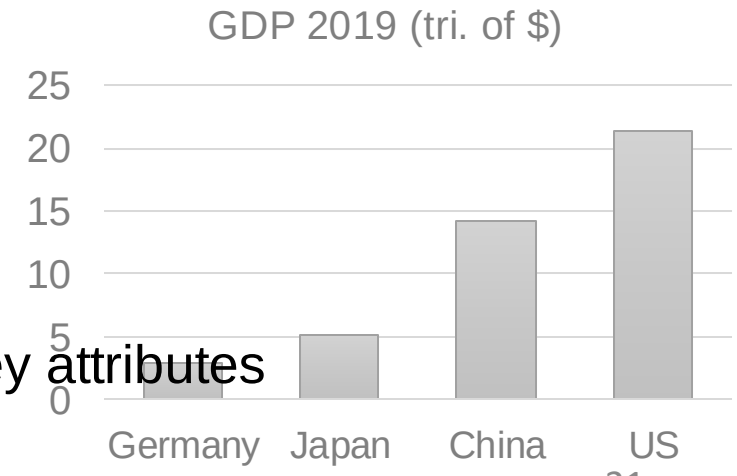
⇩ Take line length into account

$$\frac{\sum_i |\theta_i(\alpha)| l_i(\alpha)}{\sum_i l_i(\alpha)} = 45^\circ$$

Bar chart

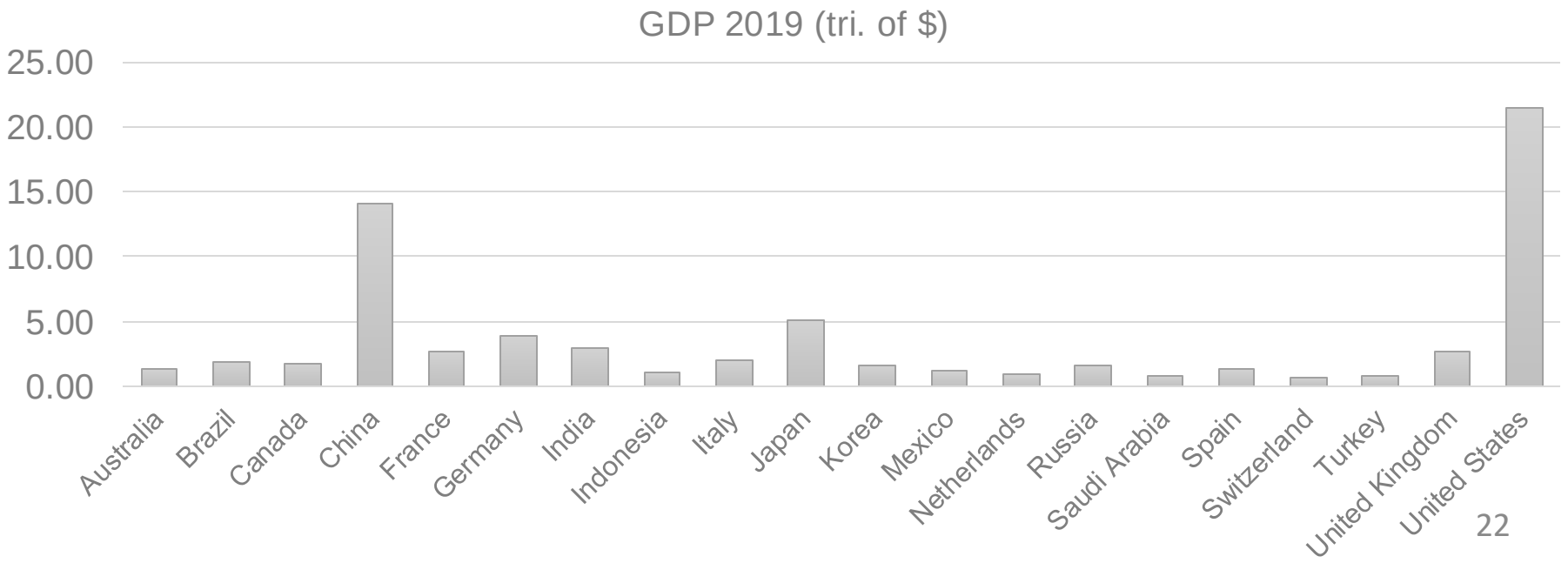
- Data
  - 1 categorical, 1 quantitative
  - One key, one value
- Mark
  - Line
- Channels
  - Aligned lengths
  - Ordered by label or quantitative attribute
- Tasks
  - Compare
  - Lookup values
- Scalability
  - dozens to hundreds of levels for key attributes





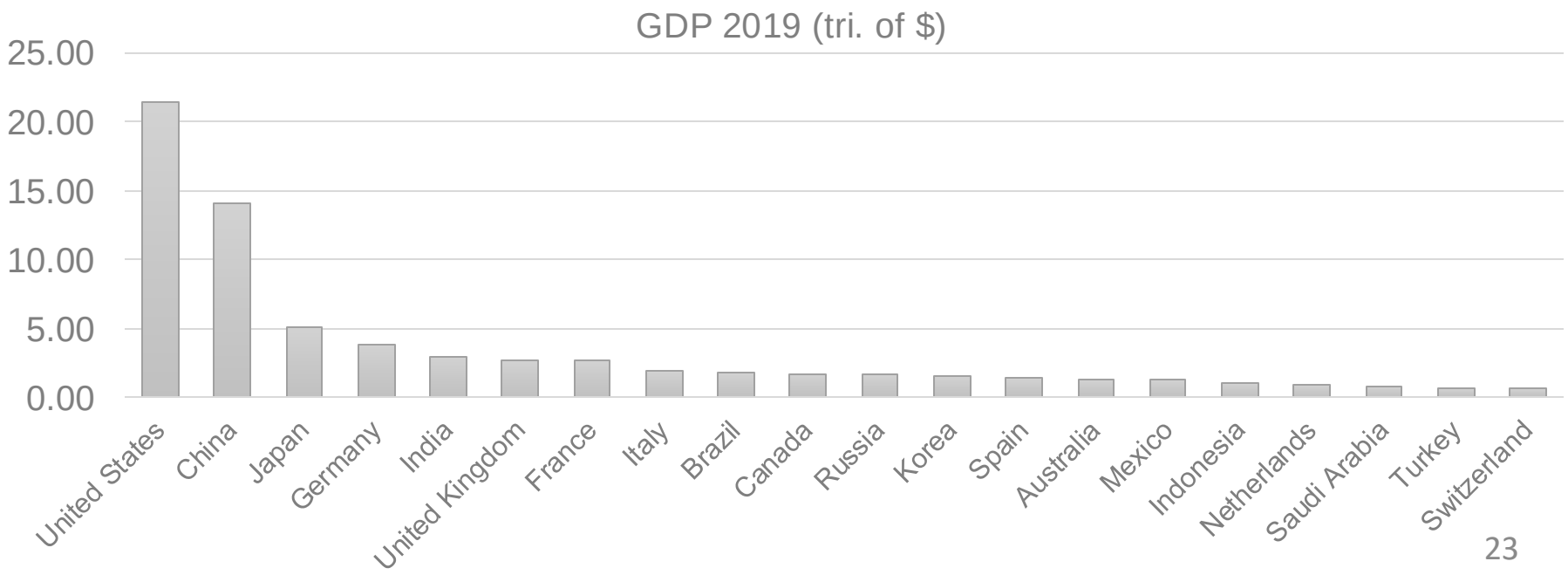
Bar chart

- Sort by labels
  - Separated and aligned but **NOT** ordered
  - Hard to know rank
    - What's the 5th highest? The 18th?



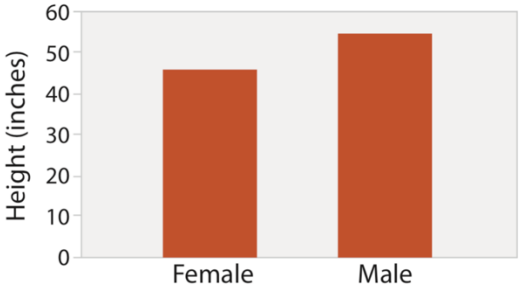
Bar chart

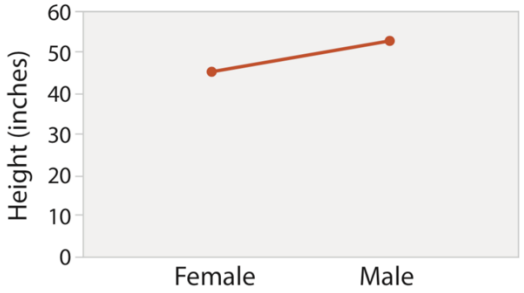
- Sort by data attributes
  - Easy to know rank
    - What's the 5th highest? The 18th?
  - More difficult to look up



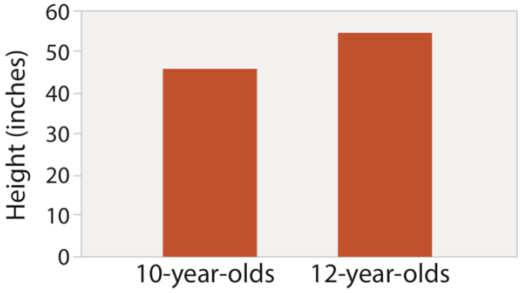
Bar or line chart

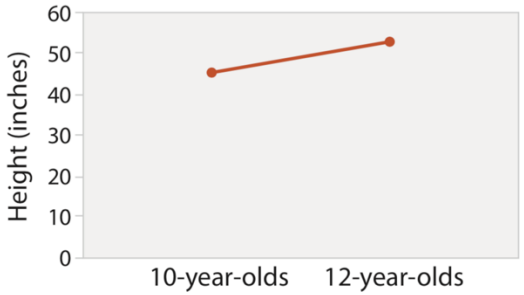
- Depends on type of *key* attributes
  - Bar chart if categorical
  - Line chart if ordered
- Do not use line chart for categorical key attributes
  - Violates expressiveness principle
    - implication of trend so strong that it overrides semantics!





"The more male a person is, the taller he/she is"

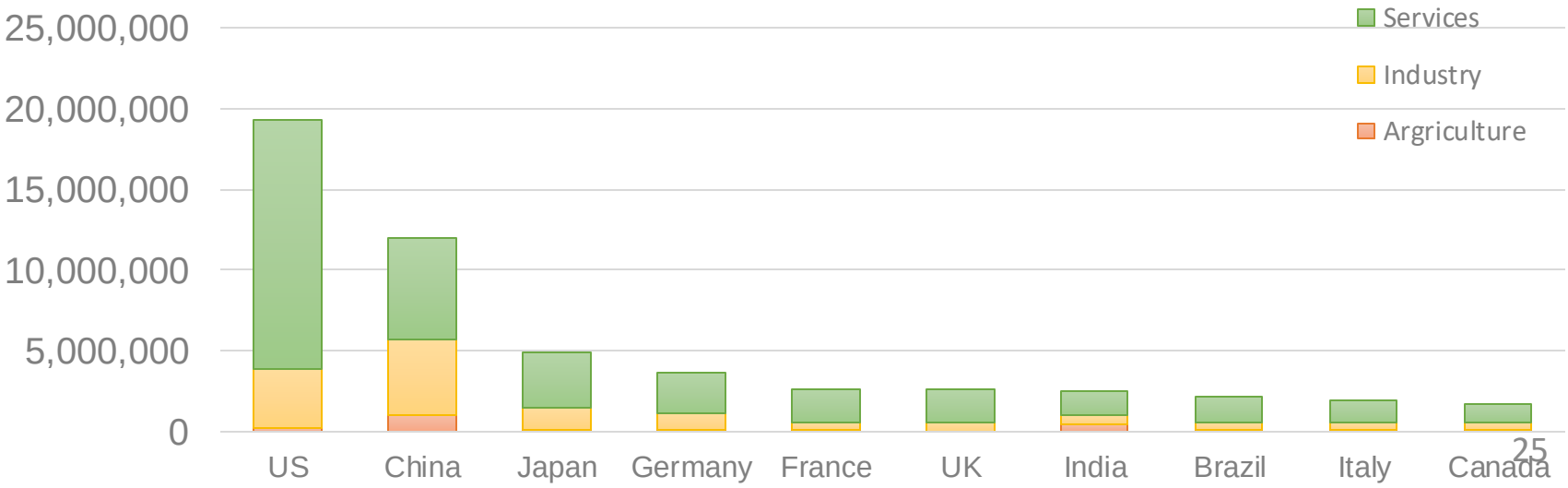




Credit: Tamara Munzner

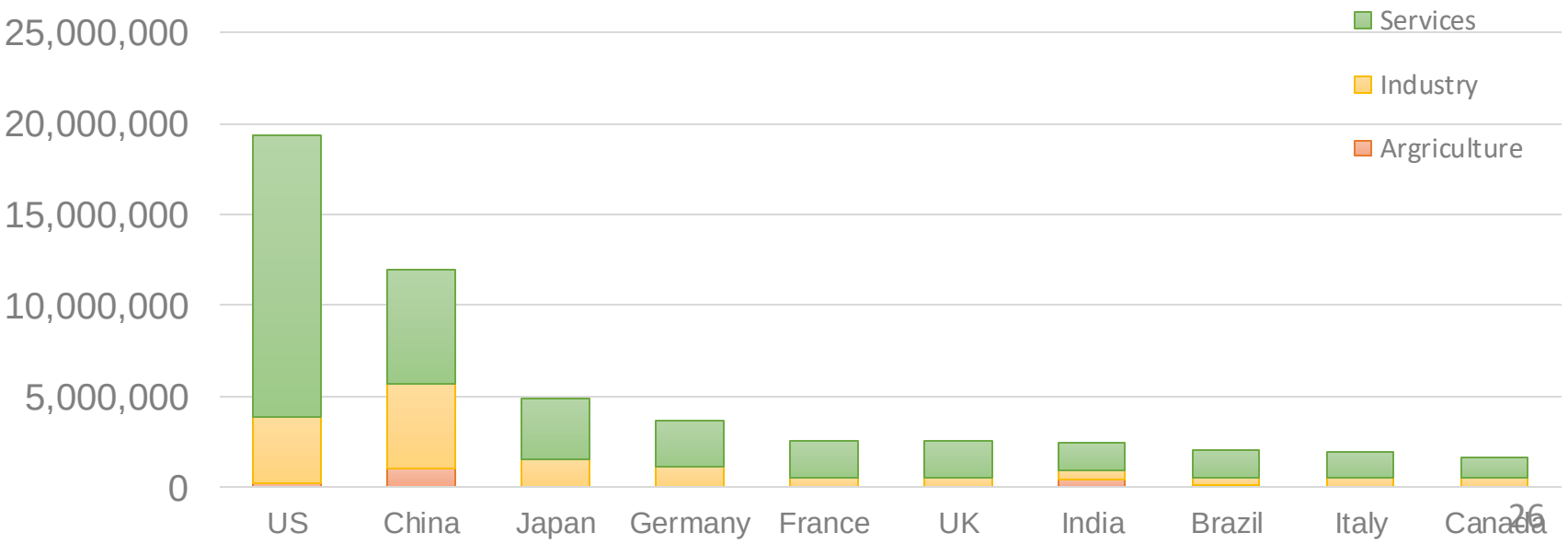
Stacked bar chart

- One more key
  - Data
    - 2 categorical attribute, 1 quantitative attribute
  - Mark: vertical stack of line marks
  - Channels
    - Length and color hue
    - Spatial regions: one per glyph



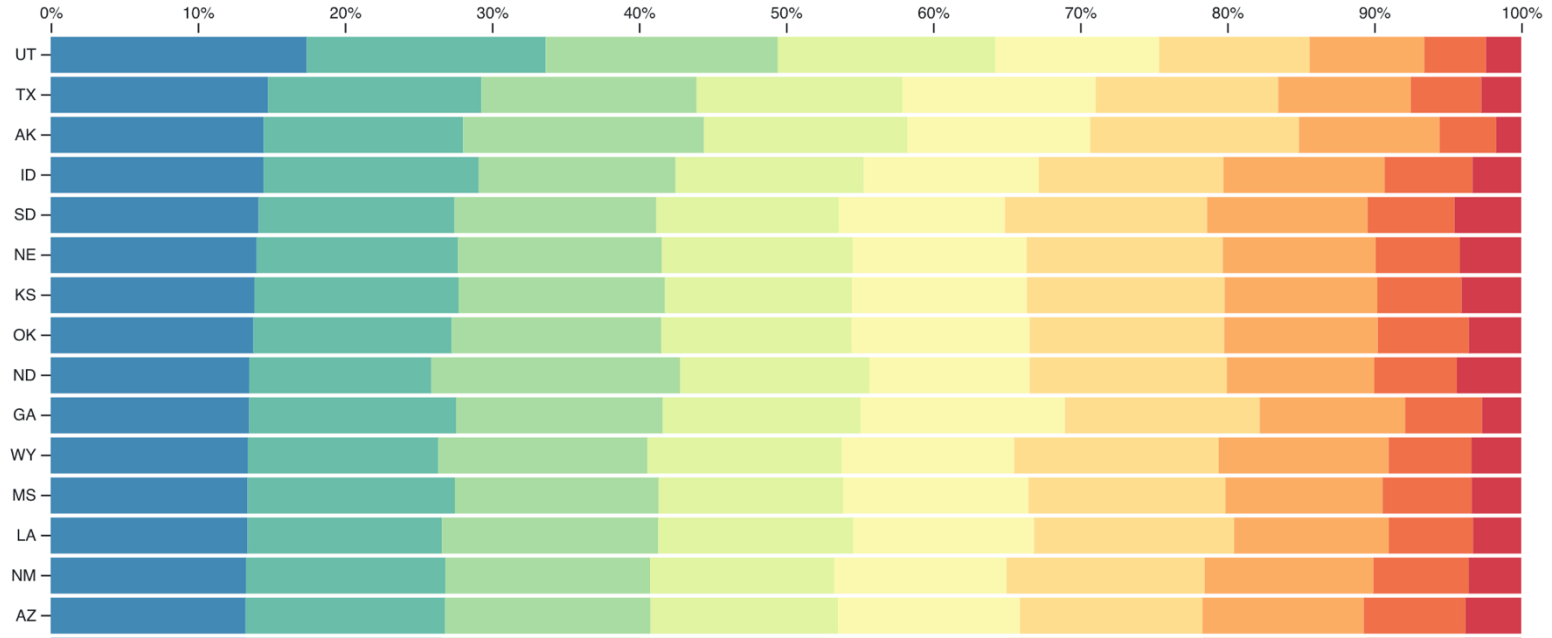
Stacked bar chart

- One more key
  - Task
    - Part-to-whole relationship
  - Scalability
    - Several to one dozen levels for stacked attributes



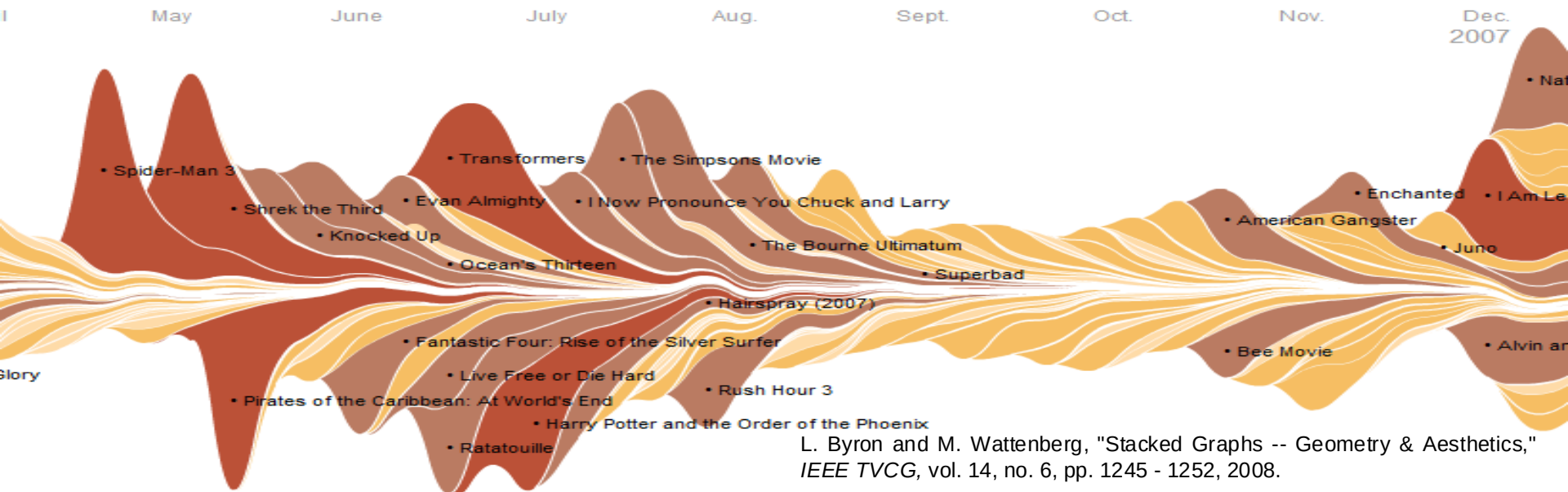
Normalized stacked bar chart

- One more key
  - Task
    - Part-to-whole judgement
  - Mark
    - Line marks normalized to full vertical height
    - Single stacked bar equivalent to full pie



Streamgraph

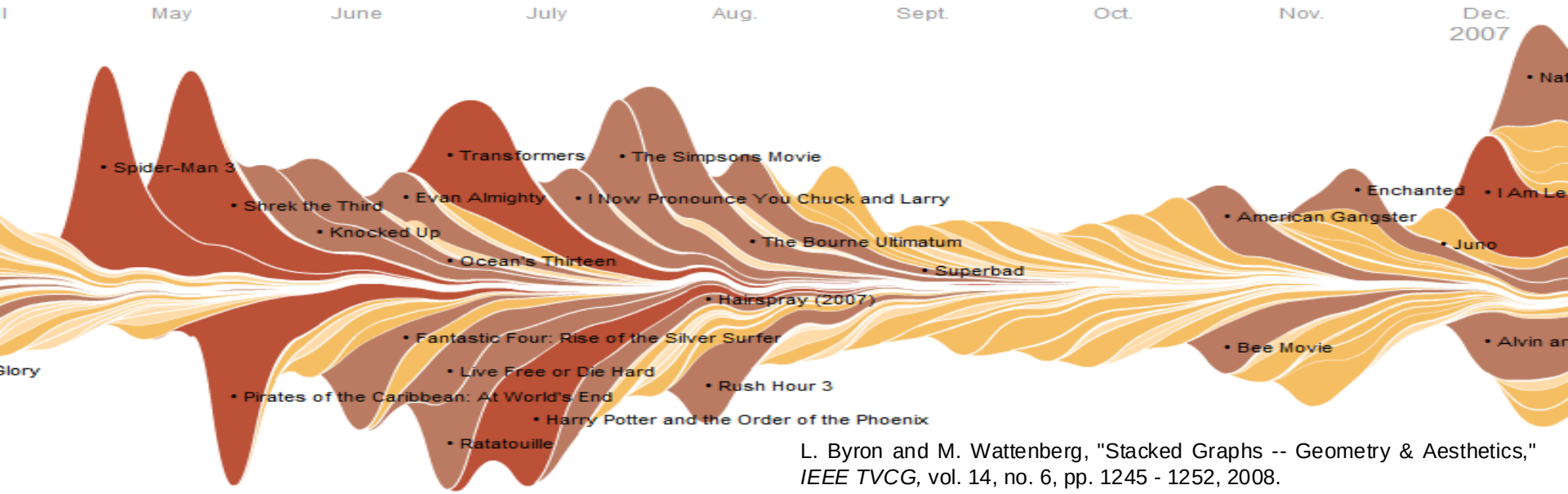
- Data
  - 1 categorical key attribute
  - 1 ordered key attribute (usually time)
  - 1 quantitative value attribute
- Task
  - Emphasizing horizontal continuity



L. Byron and M. Wattenberg, "Stacked Graphs -- Geometry & Aesthetics," *IEEE TVCG,* vol. 14, no. 6, pp. 1245 - 1252, 2008.

Streamgraph

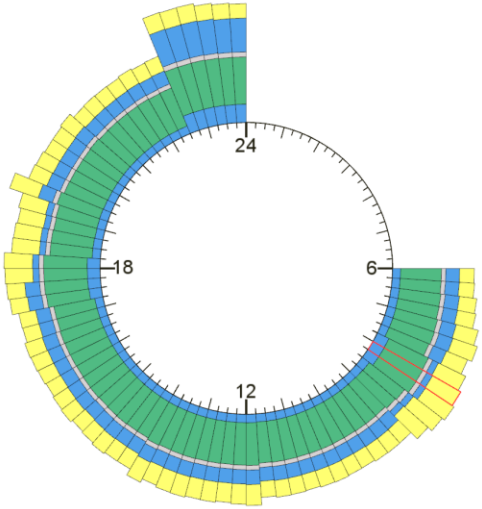
- Channels
  - Geometry: layers, where height encodes counts
  - Layer ordering
- Scalability
  - Hundreds of time keys
  - Dozens to hundreds of categories



L. Byron and M. Wattenberg, "Stacked Graphs -- Geometry & Aesthetics," *IEEE TVCG,* vol. 14, no. 6, pp. 1245 - 1252, 2008.

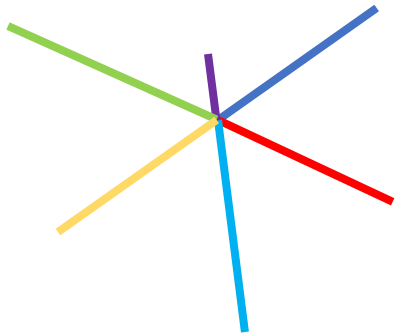
Bar chart in radial layout

- Radial layout
  - **Radial bar chart:** radial axes meet at central ring, line mark
  - **Star plot:** radial axes, meet at central point, line mark



Zeng et al., 2014

- Accuracy
  - less accurate than aligned with rectilinear



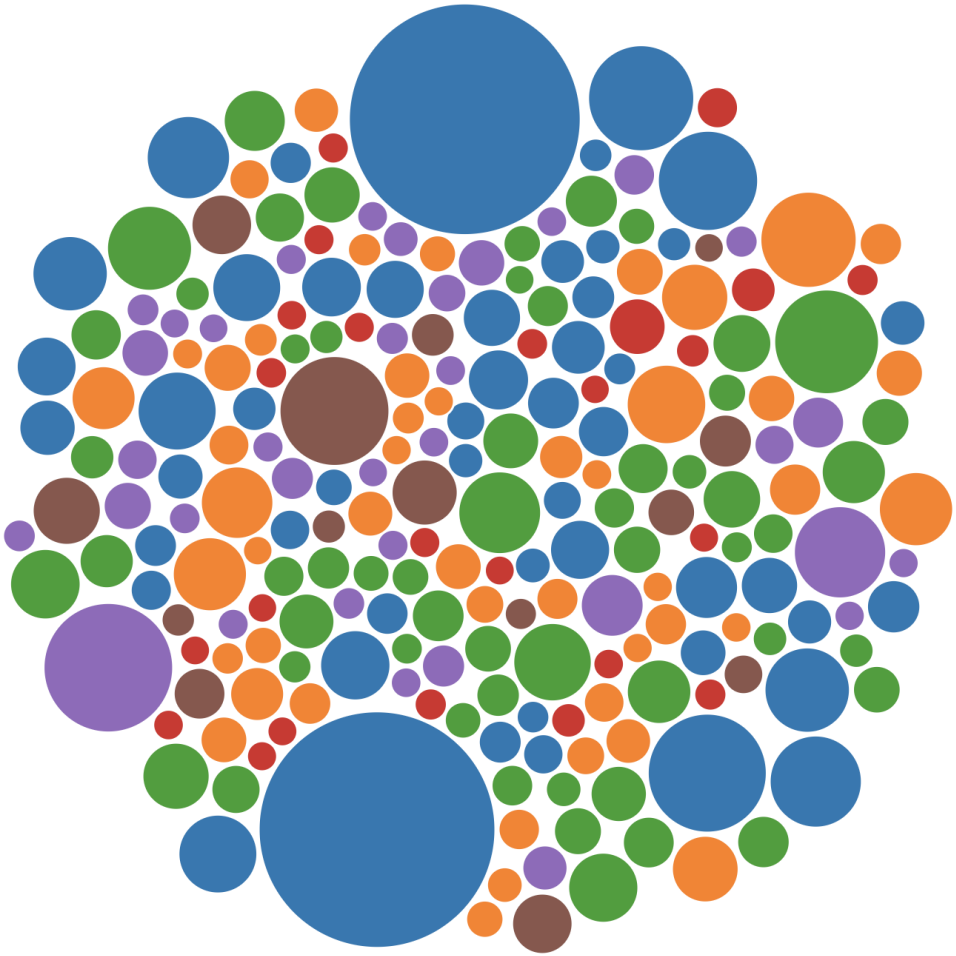
Data Exploration & Visualization

Module 6: Basic Charts

- Point charts
  - scatterplot, multi-class scatterplot
- Line charts
  - line chart, bar chart, stacked bar chart
- Area charts
  - bubble chart, pie chart
- Composite charts

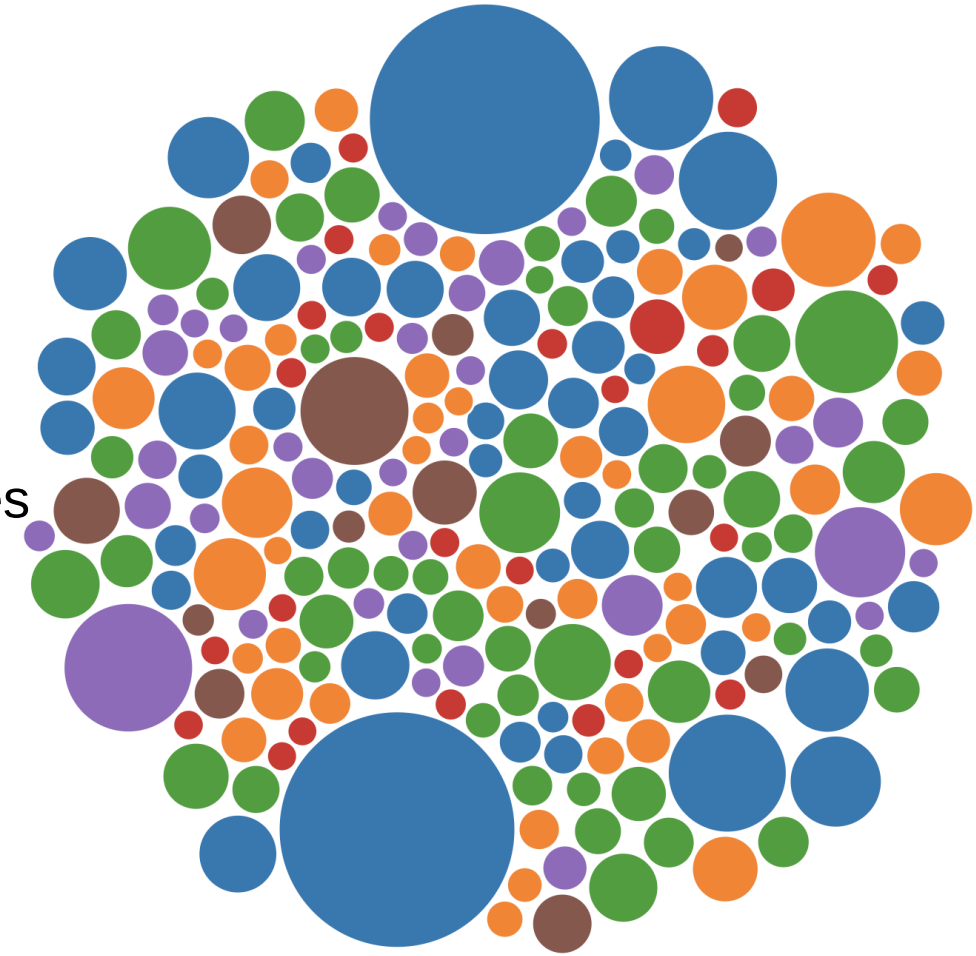
Bubble chart

- Data
  - 1 categorical, 1 quantitative
- Mark
  - Area
- Channels
  - Size
  - Separated but not ordered or aligned



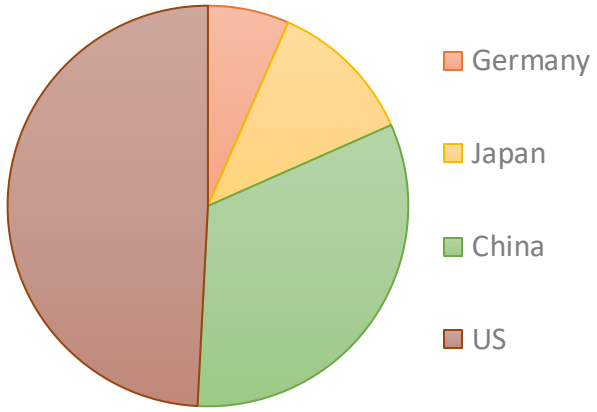
Bubble chart

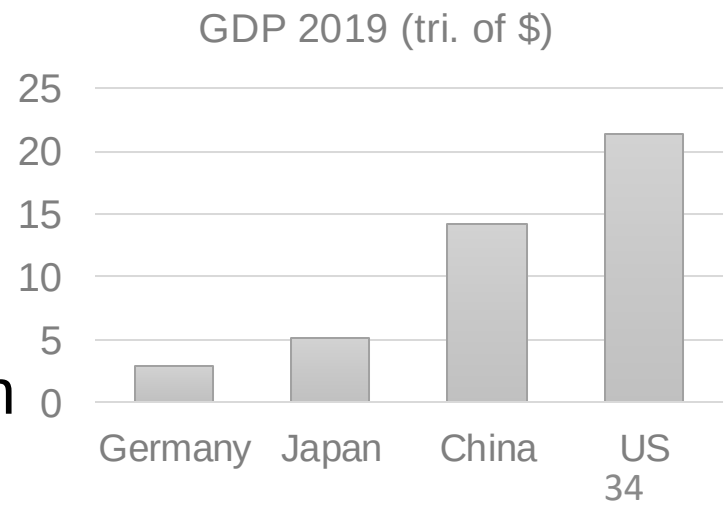
- Limitation
  - Hard to make comparisons
    - World population: China? India?
    - Size is less accurate for quantitative values
  - Seldomly used in visual analytics



Pie chart

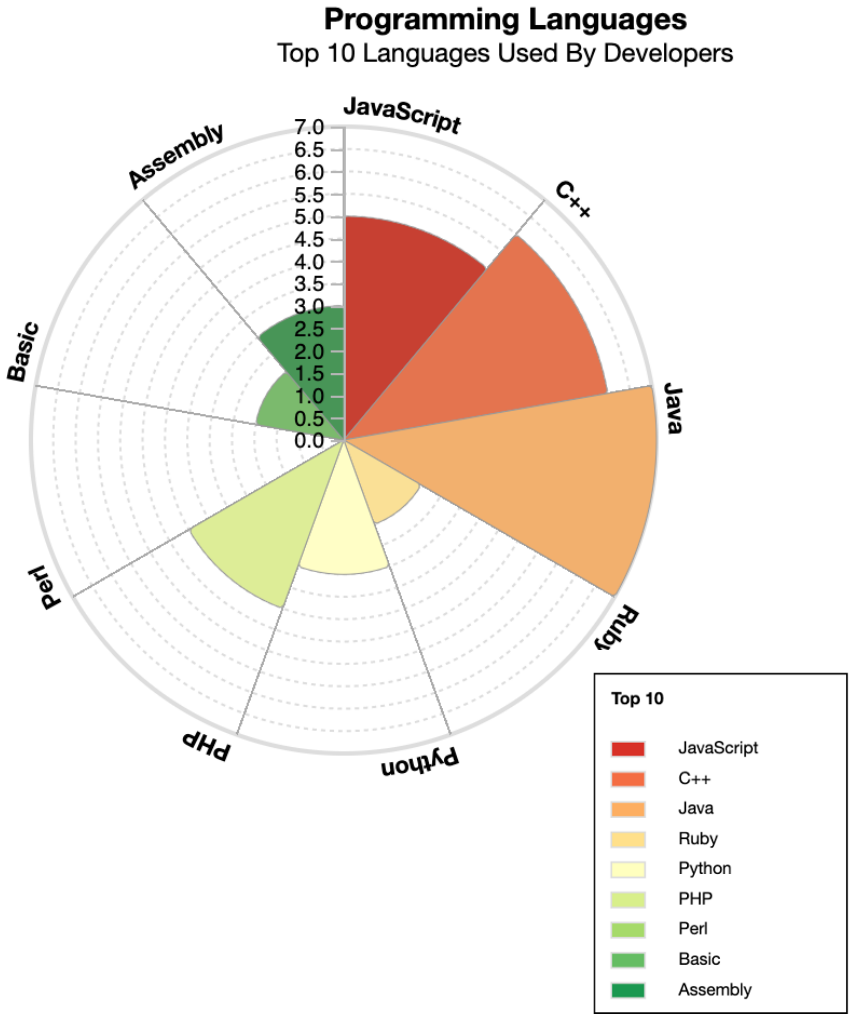
- Data
  - 1 categorical, 1 quantitative
- Mark
  - Area
- Channel
  - Angle
- Task
  - Part-to-whole judgements
- Limitation
  - angle/area less accurate than line length





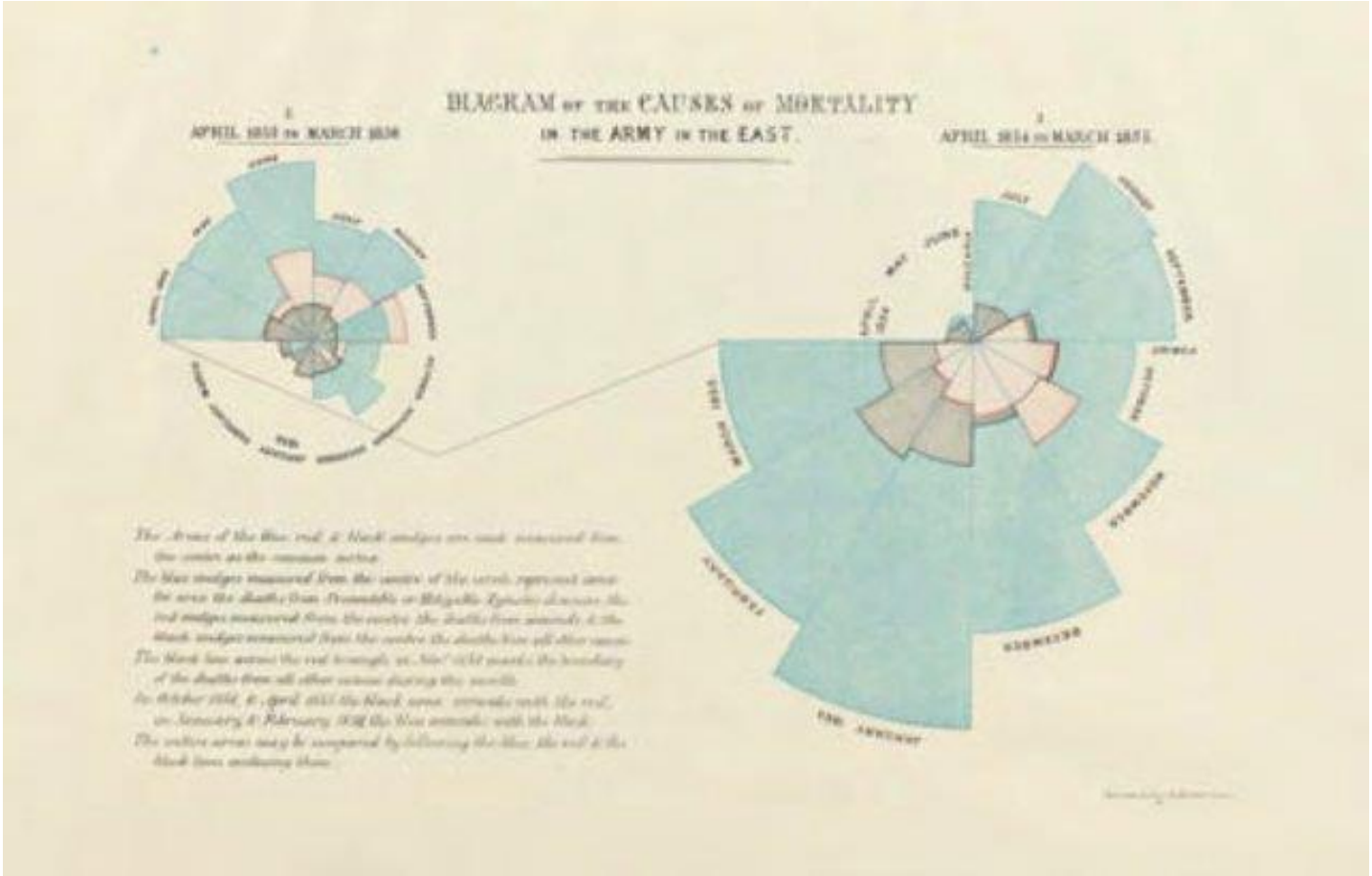
Polar area chart

- Data
  - 1 categorical, 1 quantitative
- Mark
  - Area
- Channel
  - Length
  - more direct analog to bar charts



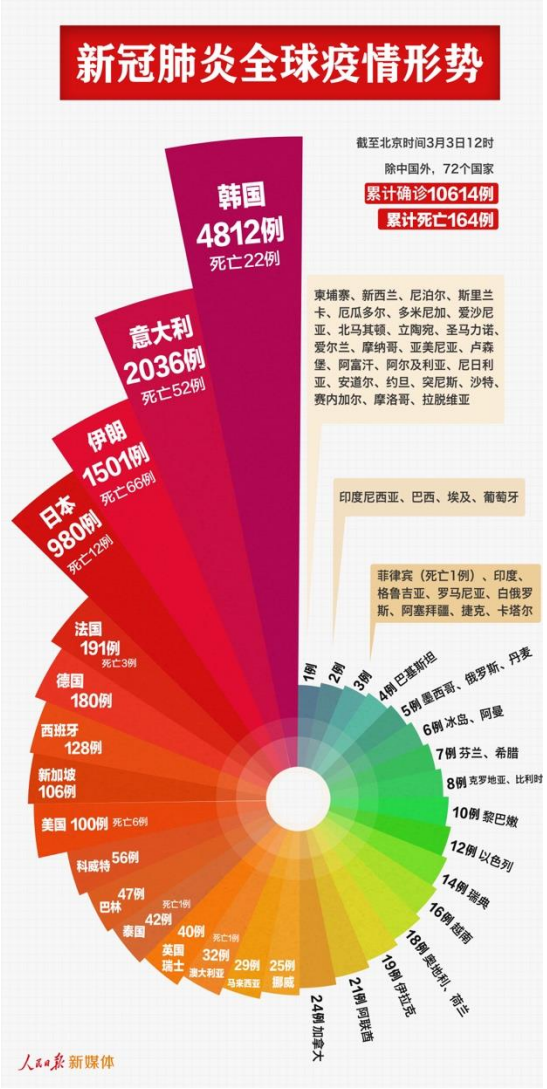
Polar area chart

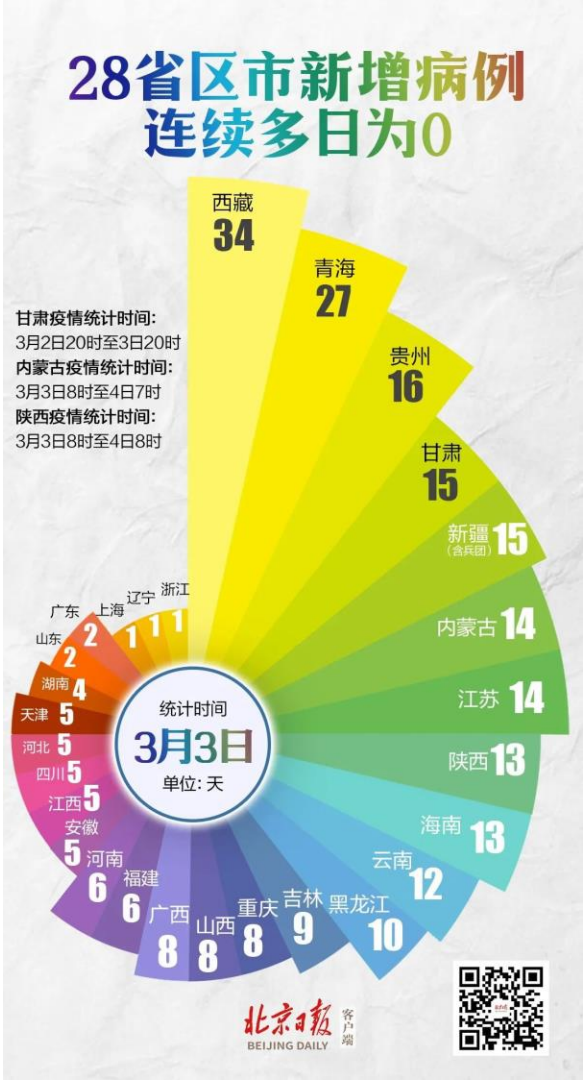
- Florence Nightingale's rose chart
  - Interactive version: http://bl.ocks.org/kgryte/raw/5926740/



Polar area chart

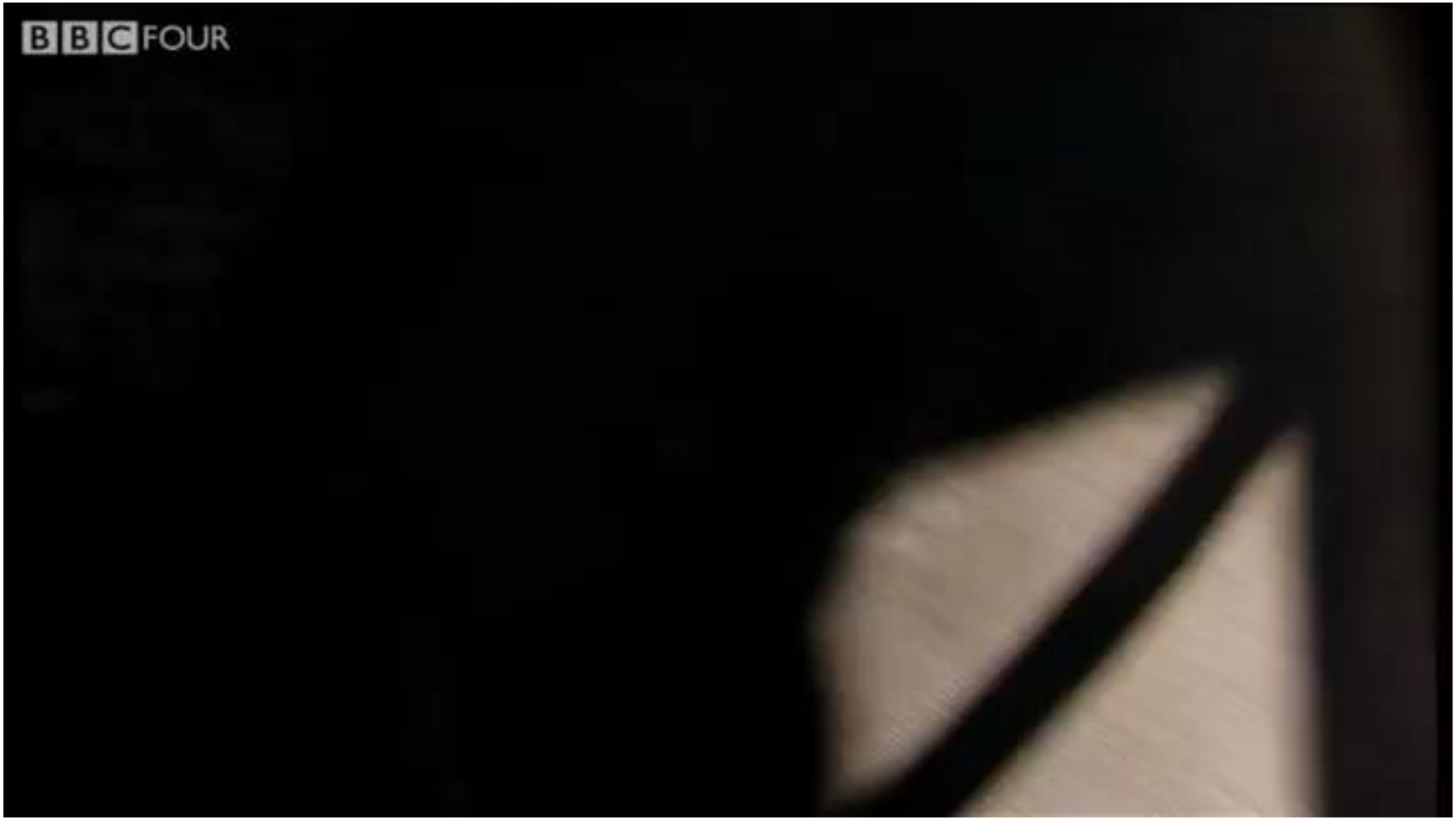
- Abuse of polar area charts
  - May want to fit to mobile screen
  - Can be replaced with vertical bar charts





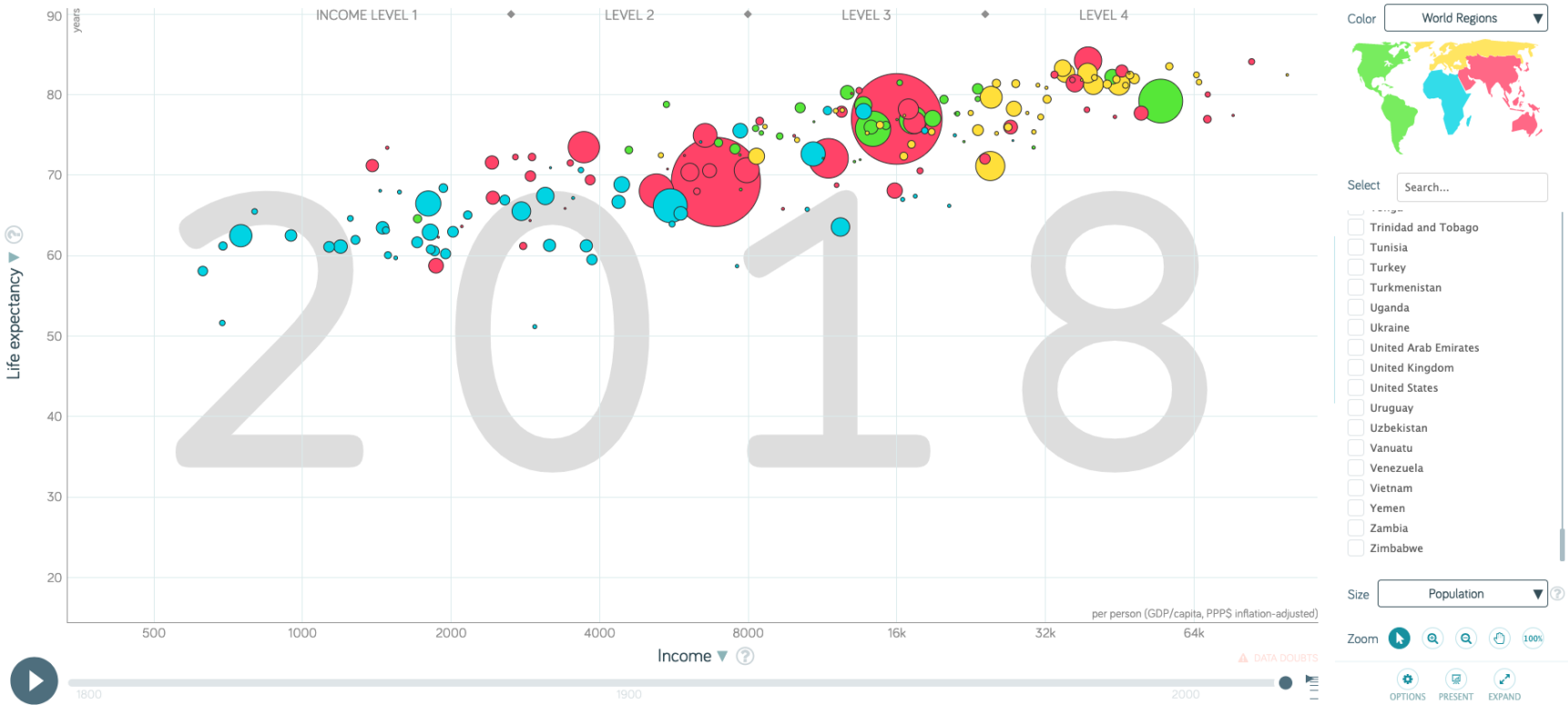
In-class exercise

- “Hans Rosling’s 200 Countries, 200 Years, 4 Minutes - The Joy of Stats”, BBC 2010.



In-class exercise

- Chart type?
- Visual mark and channels?
- Tasks?



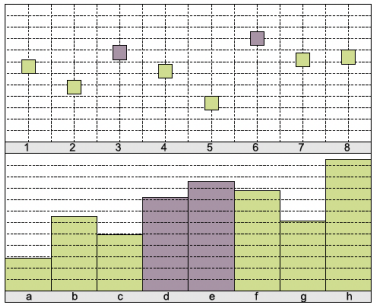
Data Exploration & Visualization

Module 6: Basic Charts

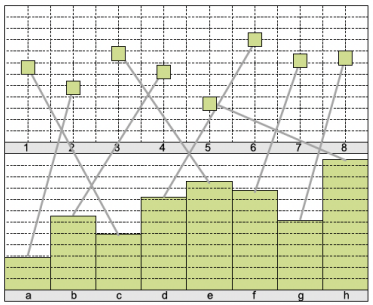
- Point charts
  - scatterplot, multi-class scatterplot
- Line charts
  - line chart, bar chart, stacked bar chart
- Area charts
  - bubble chart, pie chart
- Composite charts

Composite charts

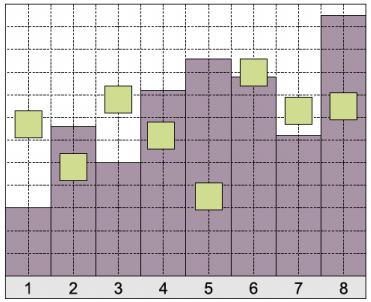
- Composite charts are popular
  - Difficult to design novel visual representations
  - All chart types have strengths and weakness
  - Combining different charts to balance their strengths and weakness
- Composite patterns
  - Juxtaposition: side-by-side
  - Superimposition: overlaying
  - Nesting
  - Integration: visual links
  - Overloading



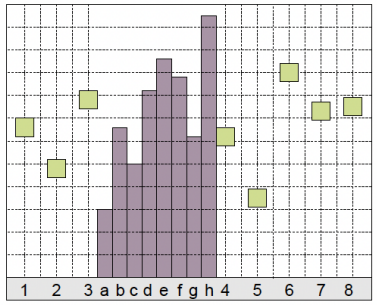
(a) Juxtaposed views.



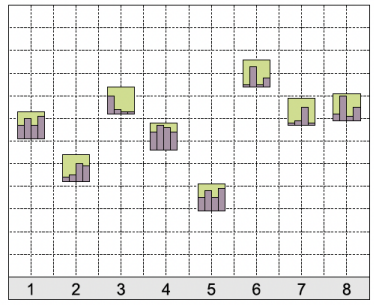
(b) Integrated views.



(c) Superimposed views.



(d) Overloaded views.

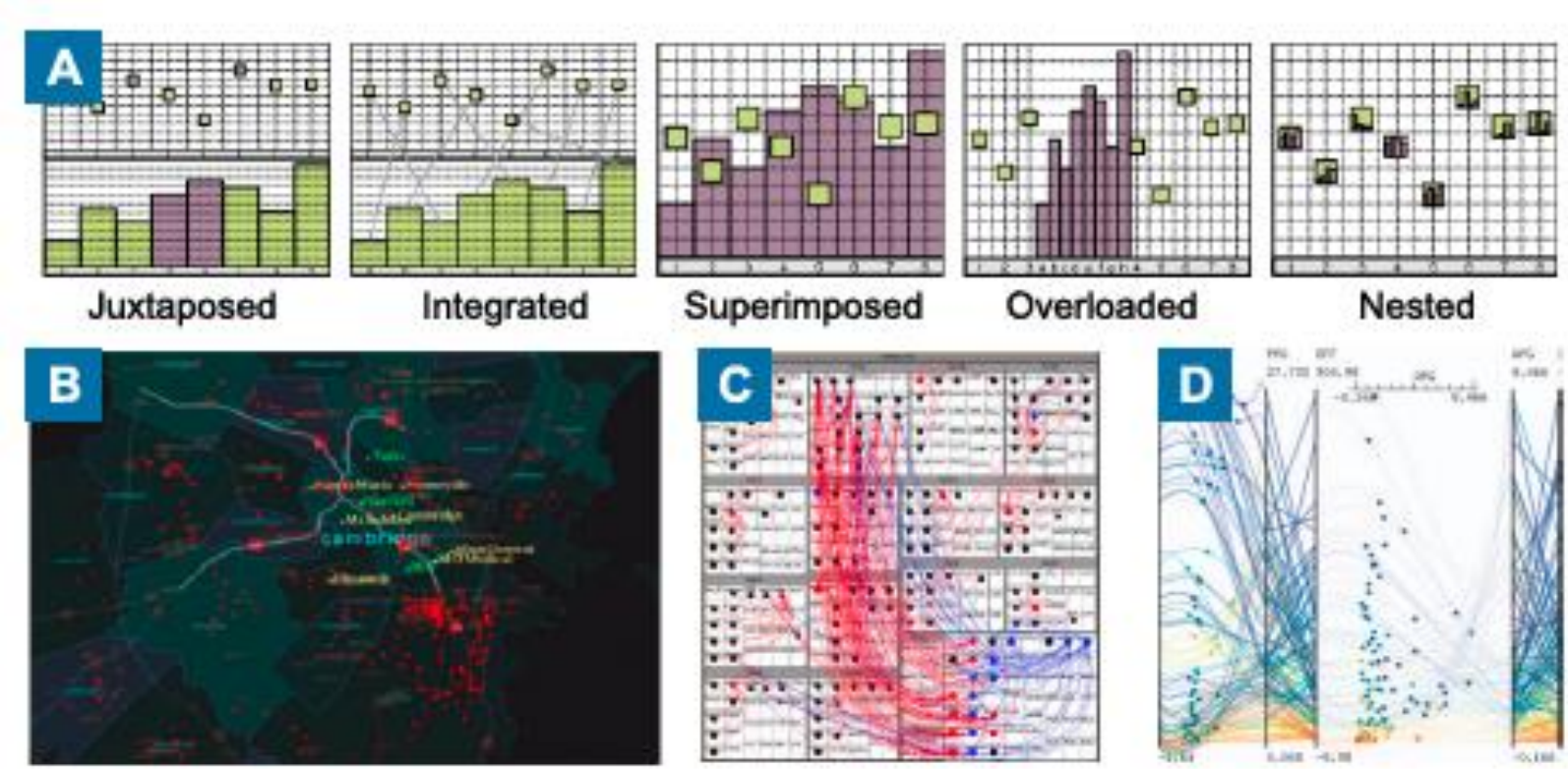


(e) Nested views.

Javed & Elmqvist, 2012

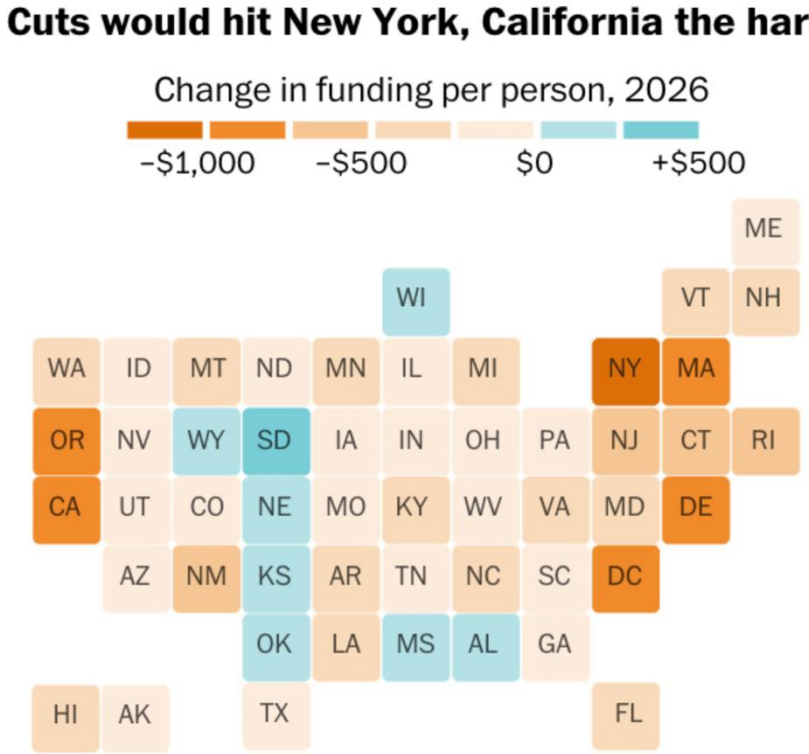
Composite charts

- Composition patterns
  - https://composite-visualizations.github.io/explorer/



Composite charts

- Grid maps: Small multiples of visualization, with the charts arranged in a map formation.

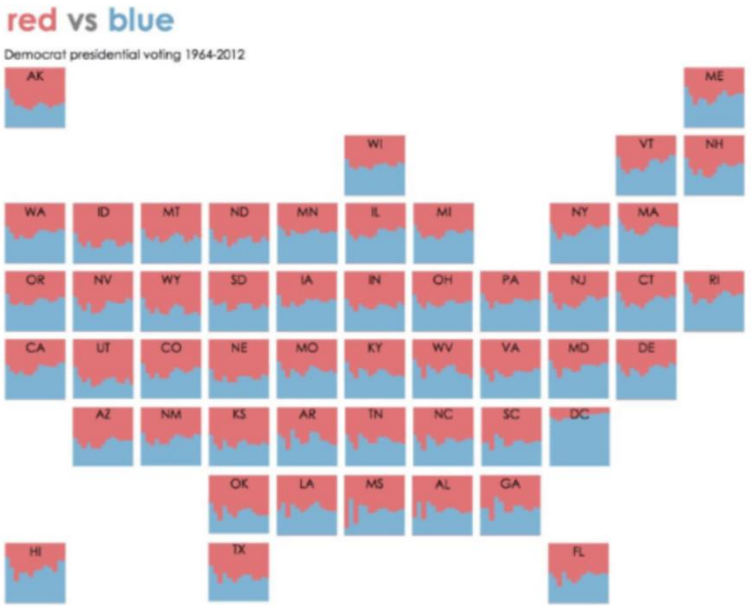


Heatmap + geographical information

Bar chart + geographical information

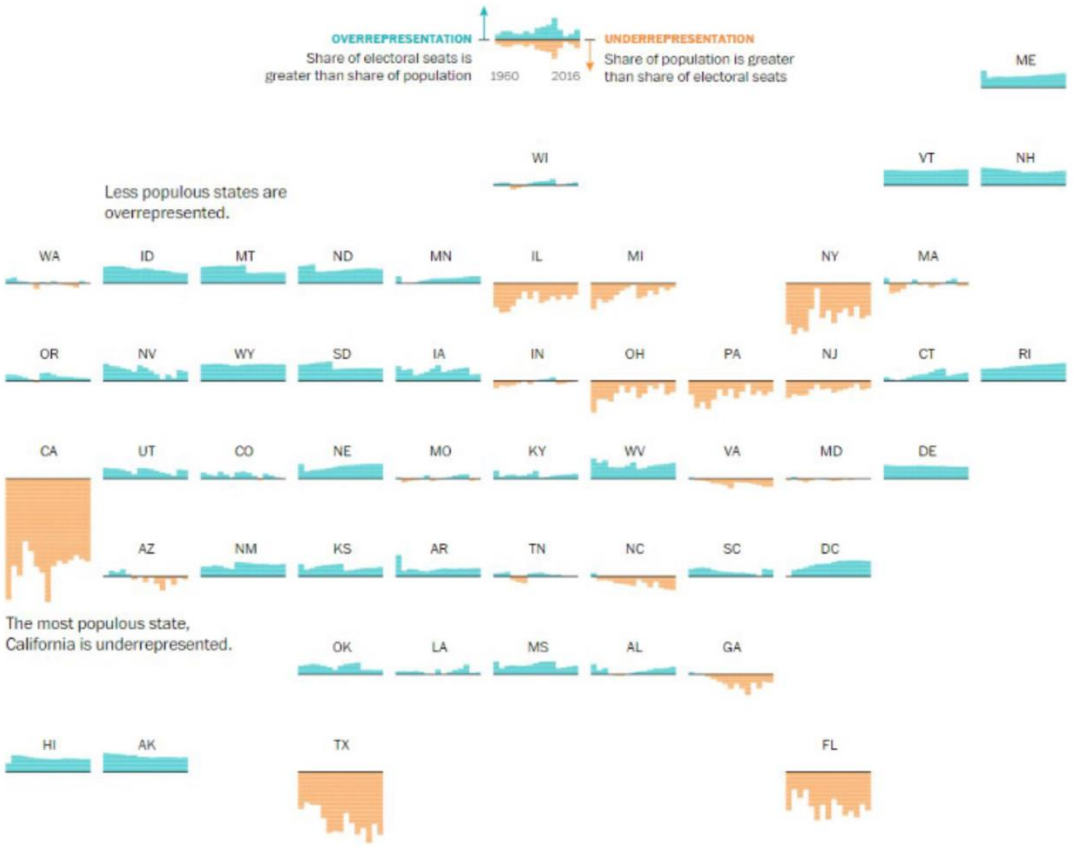
Composite charts

- Grid map + Bar chart
  - Republican and Democrat presidential voting.
- Grid map + Stacked histogram:
  - The ratio between Republican and Democrat presidential voting is visualized over time and across all 51 states.



Composite charts

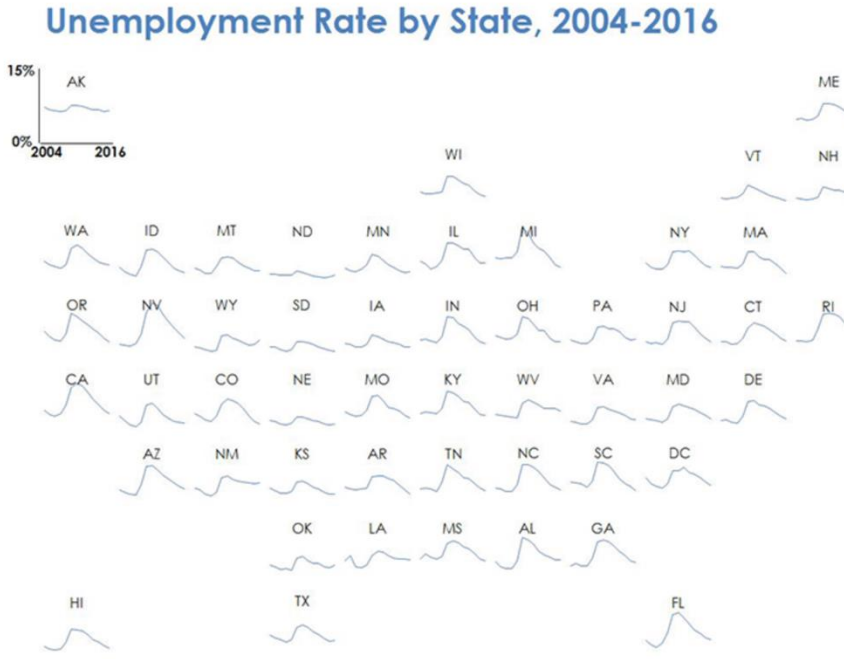
- Grid map + Histogram
  - The difference between the share of the national population and the state's share of votes in the electoral college between 1960 to 2016.
  - each bar represents an election year
  - Color: blue → greater share, yellow → lower share
- vote carries more 'weight' the less populated the state you vote in is it



The Washington Post

Composite charts

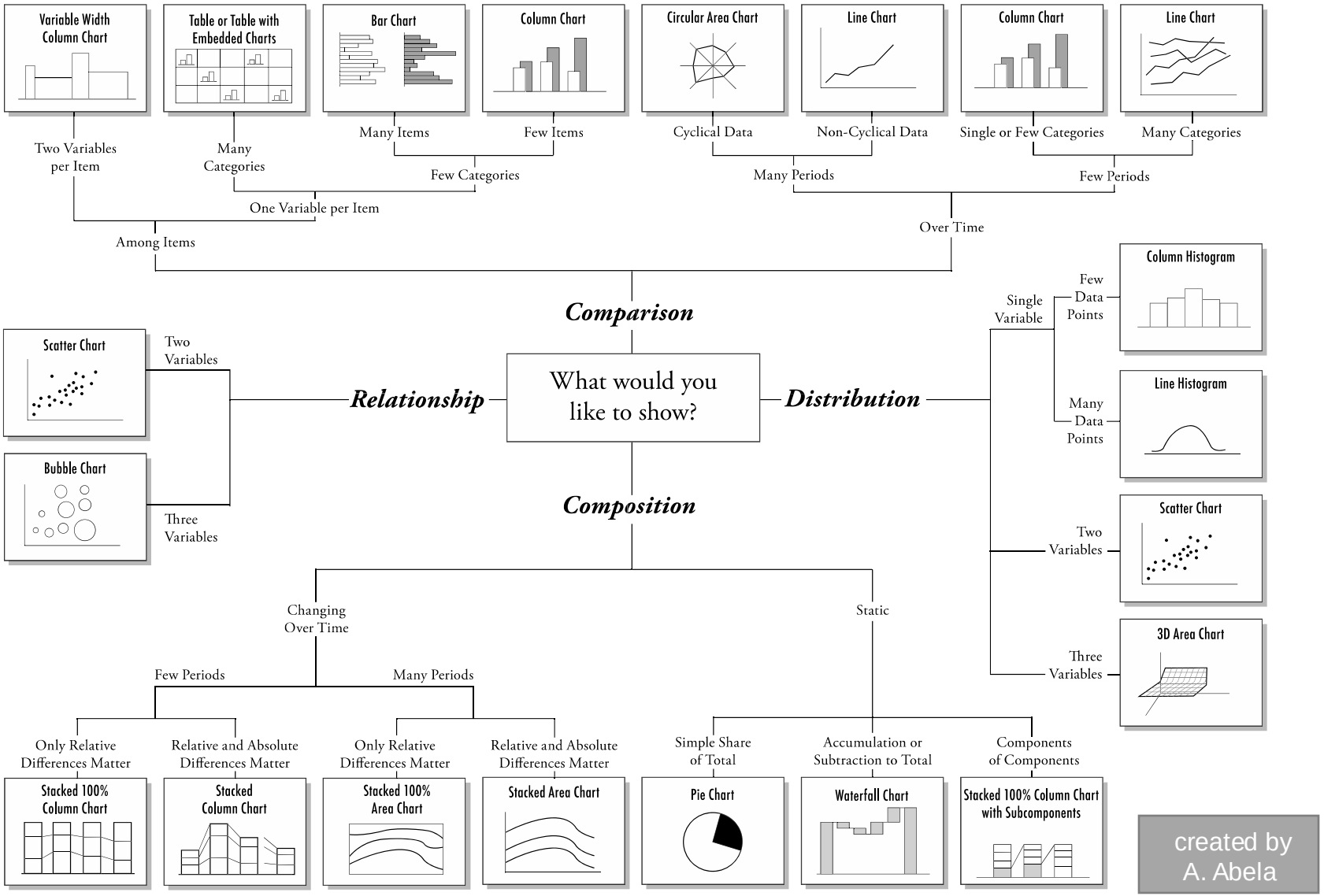
- Grid map + Line chart
  - the change in unemployment over time (2004 to 2016)



- Grid map + Pie chart
  - Limit the number of slices to two

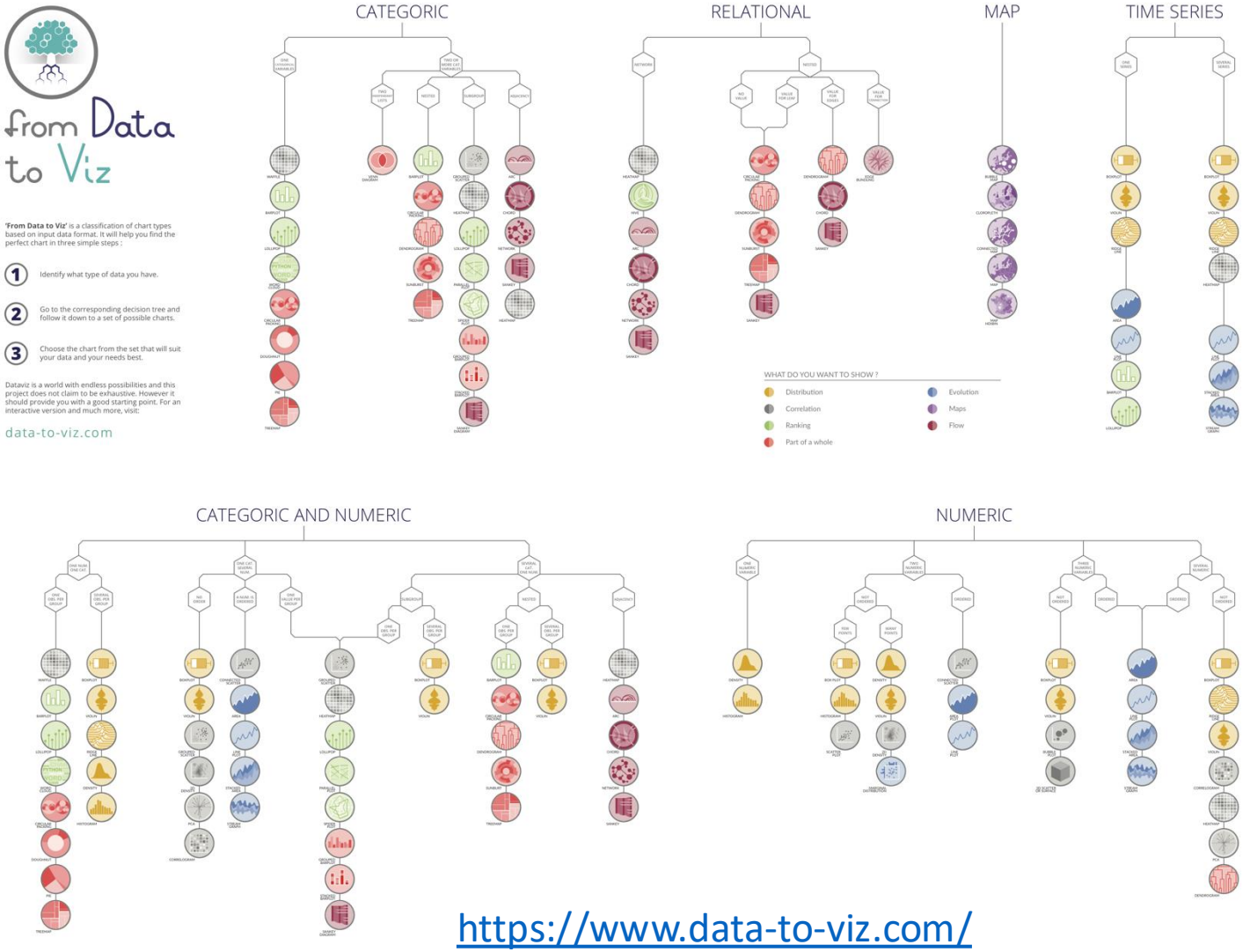
http://datavizcatalogue.com/blog/category/chart-combinations/

Summary



created by
A. Abela

Summary



https://www.data-to-viz.com/