

# Visual Data Analysis Visualization Mapping I

Dr. Johannes Kehrer

# Disclaimer



These lecture slides and materials are for personal use by students and are intended for educational purposes only. Distribution or reproduction of the material is prohibited.

The views and opinions expressed by any individual during the course of these lecturers are their own and do not necessarily represent the views, opinions, or positions of Siemens or the Technical University of Munich.

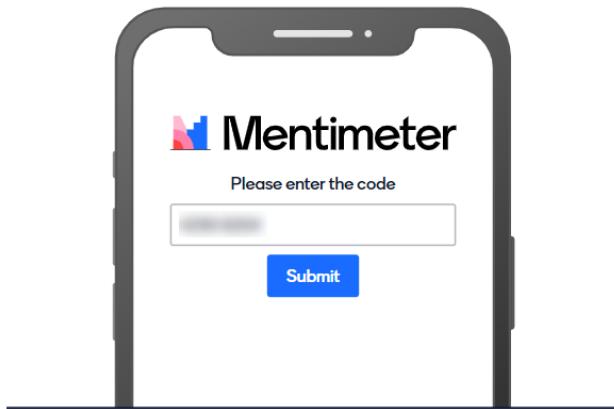
Although reasonable efforts have been made to ensure the information in these lecture slides is current and accurate, we do not assume any responsibility for the content of the posted material.

These slides also contain links to websites of third parties. As the content of these websites is not under our control, we cannot assume any liability for such external content.

# Any comments/questions?

Go to

**www.menti.com**



Enter the code

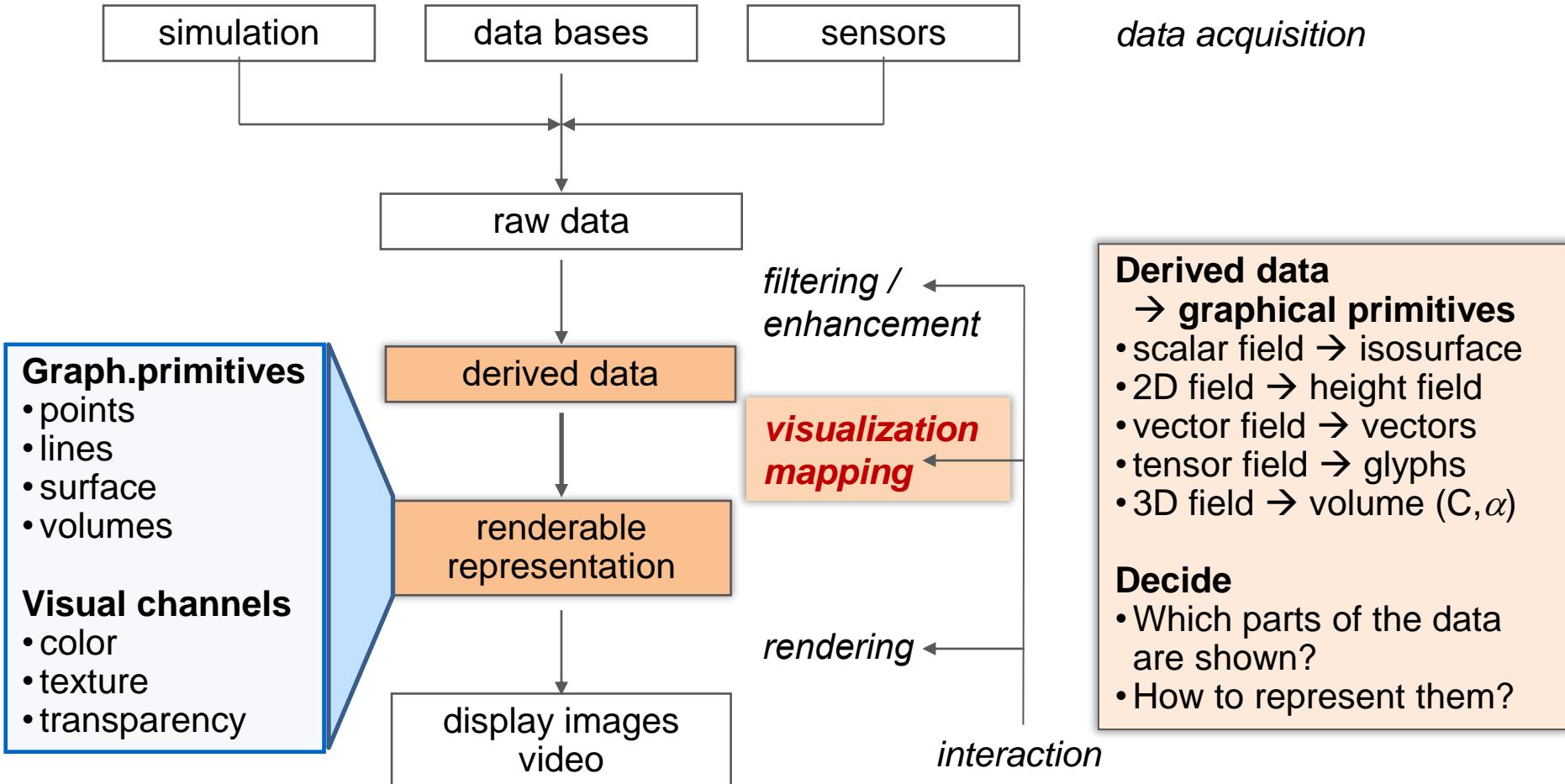
**3850 4109**



Or use QR code

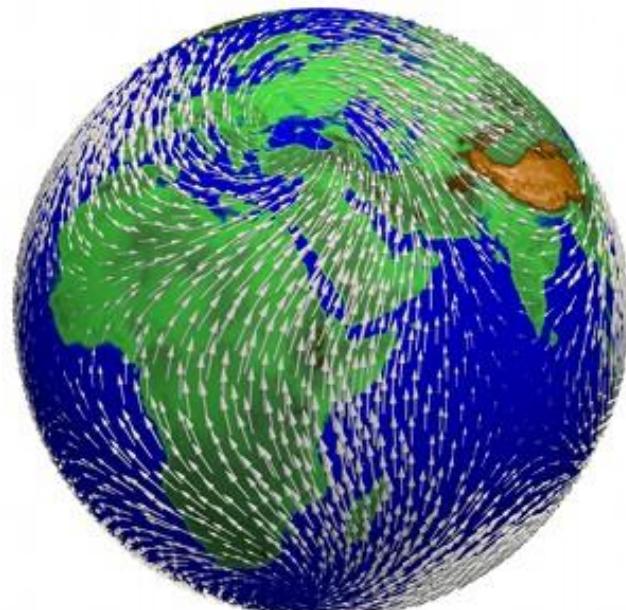
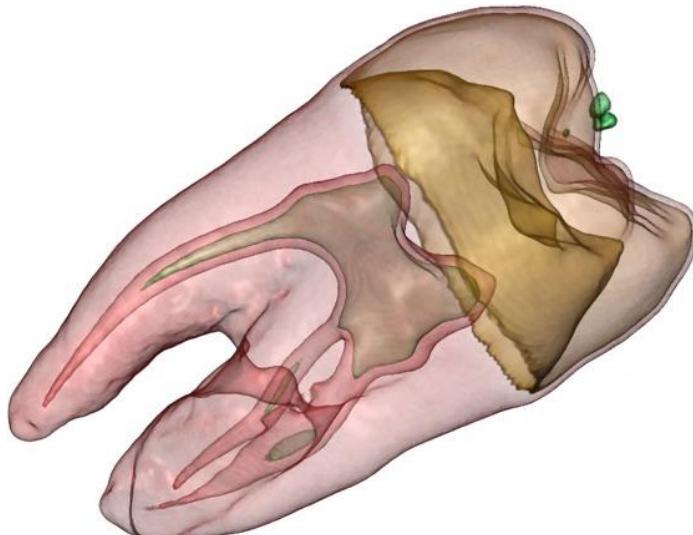
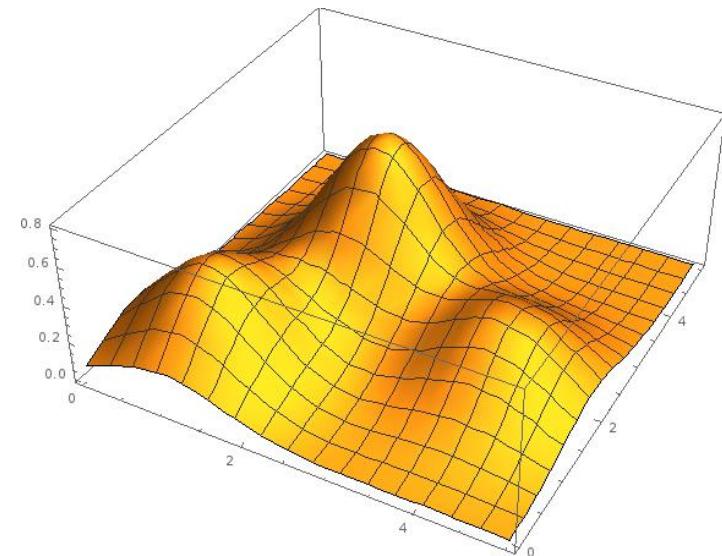
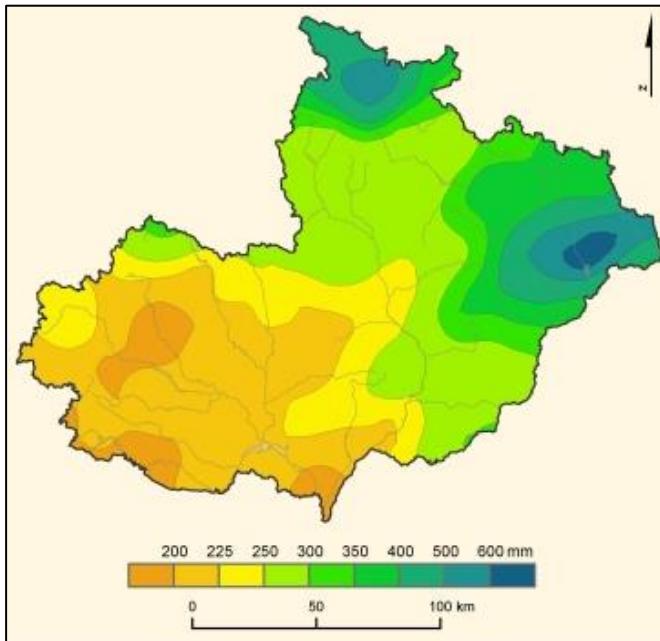
# Mapping techniques

- From derived data to a renderable representation



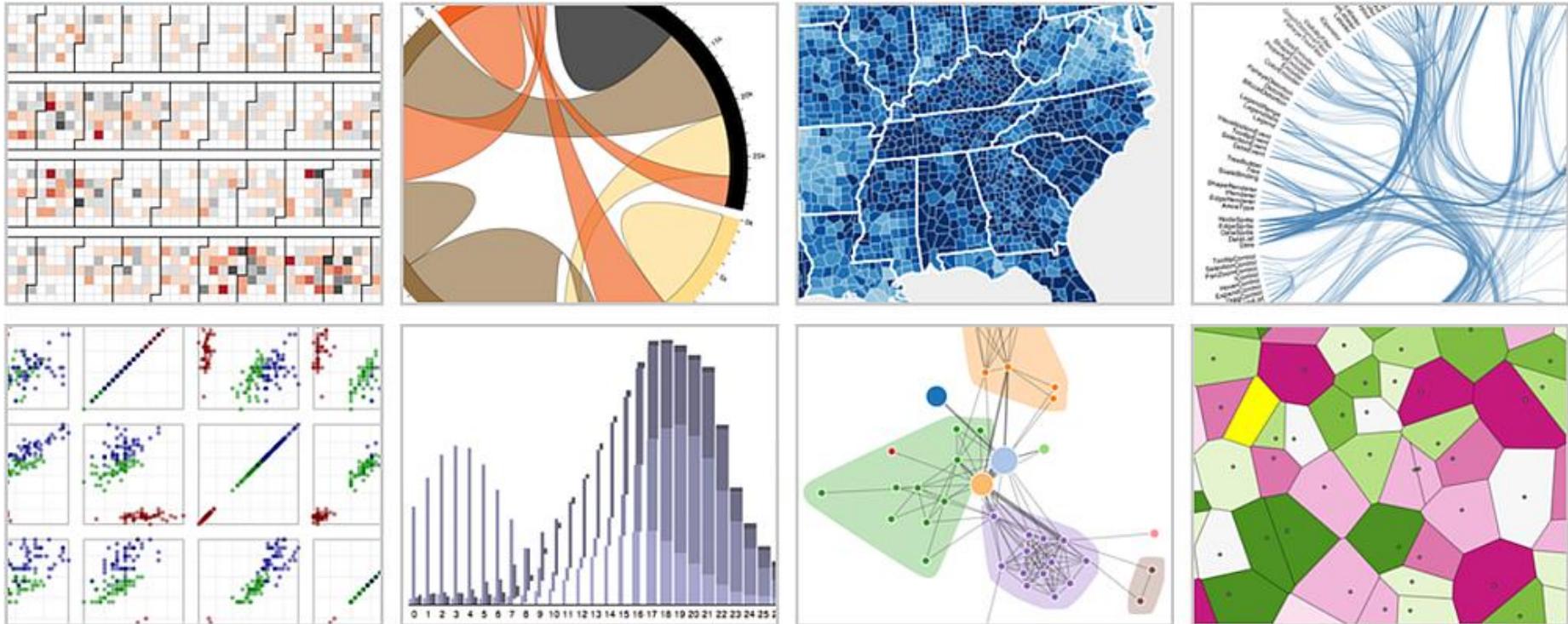
# Visualization Mappings – Examples

SIEMENS  
Ingenuity for life



# Visualization Mappings – Examples

SIEMENS  
Ingenuity for life

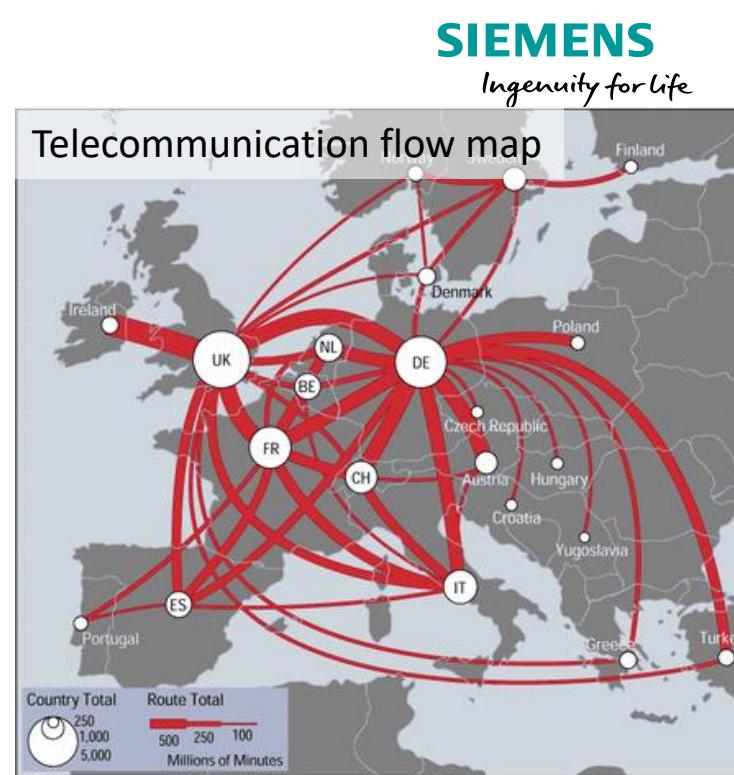
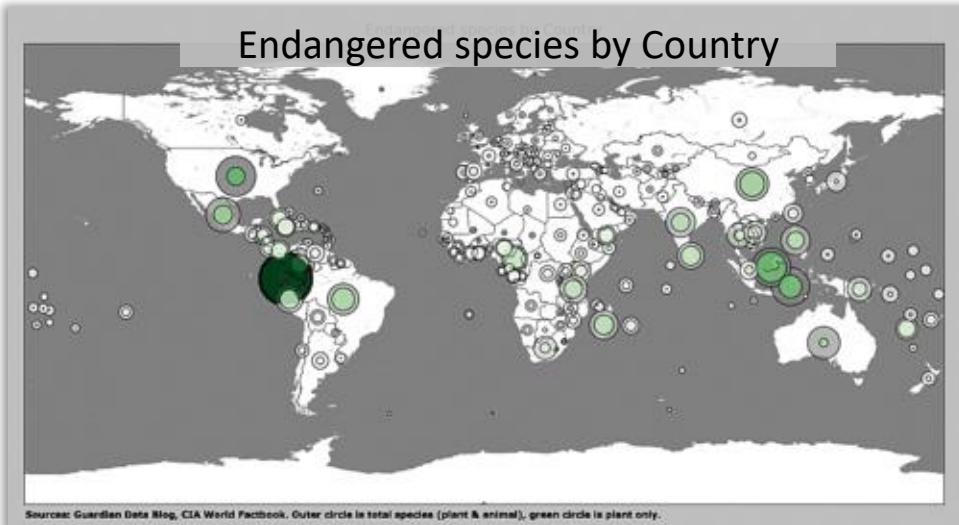


Interactive visualizations built with D3. From left to right: calendar view, chord diagram, choropleth map, hierarchical edge bundling, scatterplot matrix, grouped & stacked bars, force-directed graph clusters, Voronoi tessellation

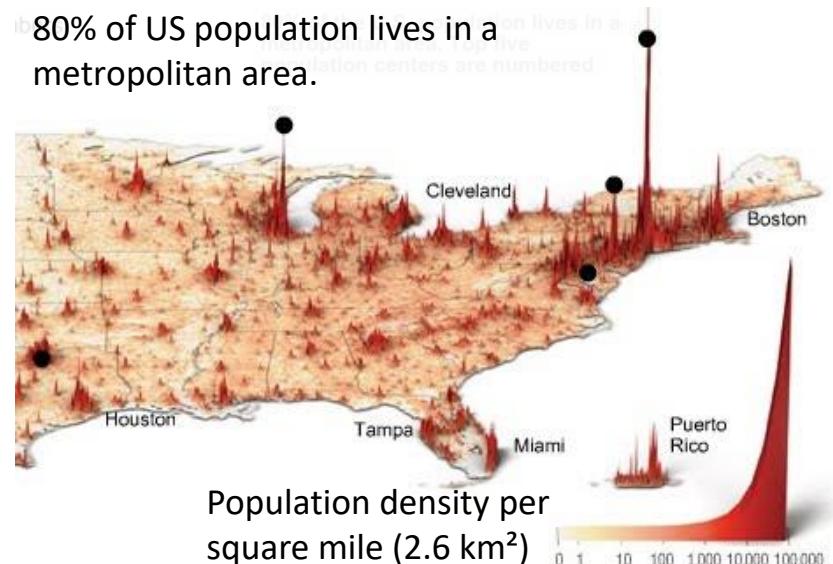
<https://d3js.org/>

# Mapping techniques

- Mapping of data to a visual (renderable) representation

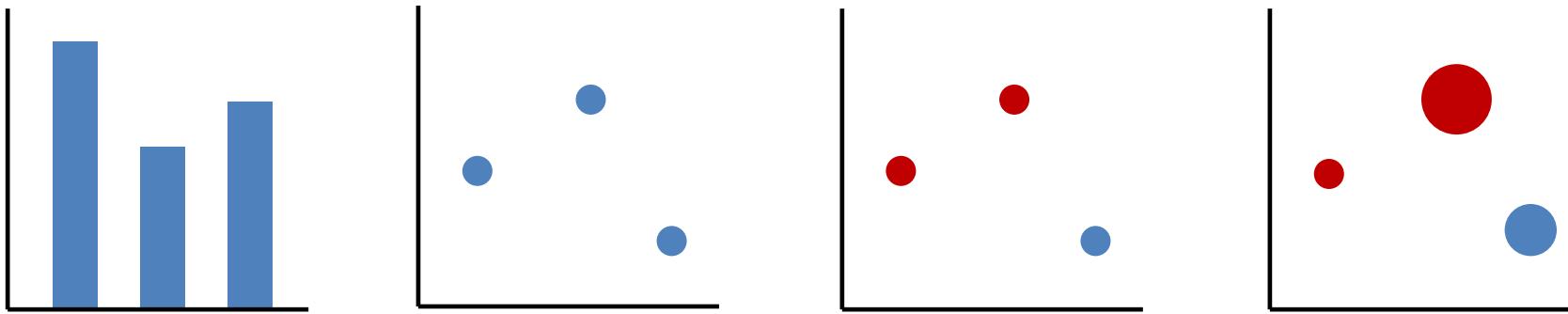


80% of US population lives in a metropolitan area. The top five population centers are numbered.



# Mapping techniques

How to systematically analyze visual mappings?



Mapping of data to a visual representation consists of:

- **Graphical primitives:** represent data items or links
- **Visual channels:** control appearance of graph. primitives based on data attributes

# Mapping techniques

- Mapping of data to a visual representation consists of:

## – Graphical primitives

that represent data items



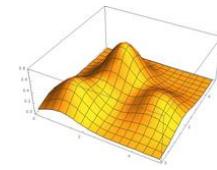
## points



## lines



## areas

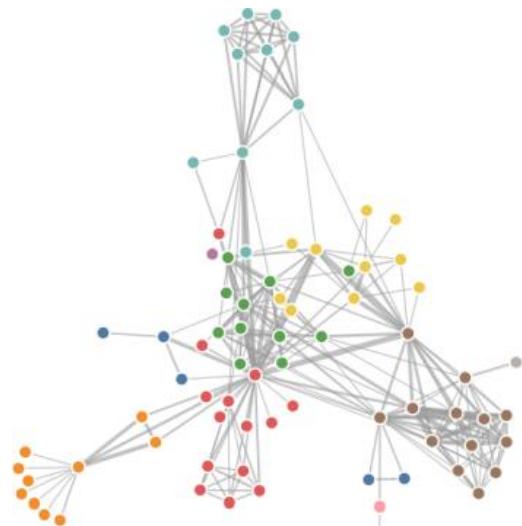


# surfaces

# Representation of links between data items



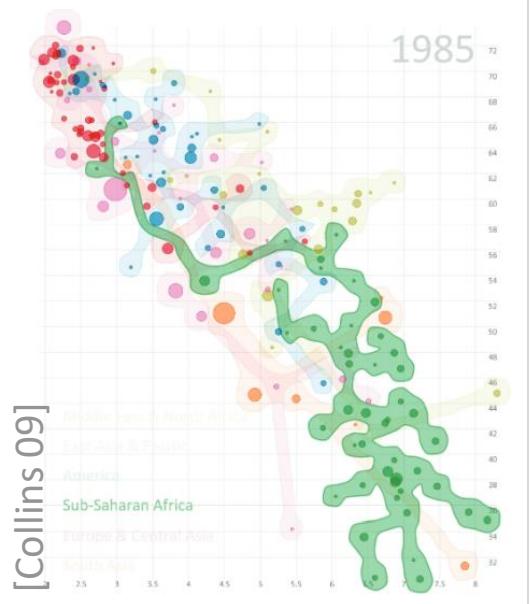
# Connection



[observablehq.com/@d3/force-directed-graph](http://observablehq.com/@d3/force-directed-graph)



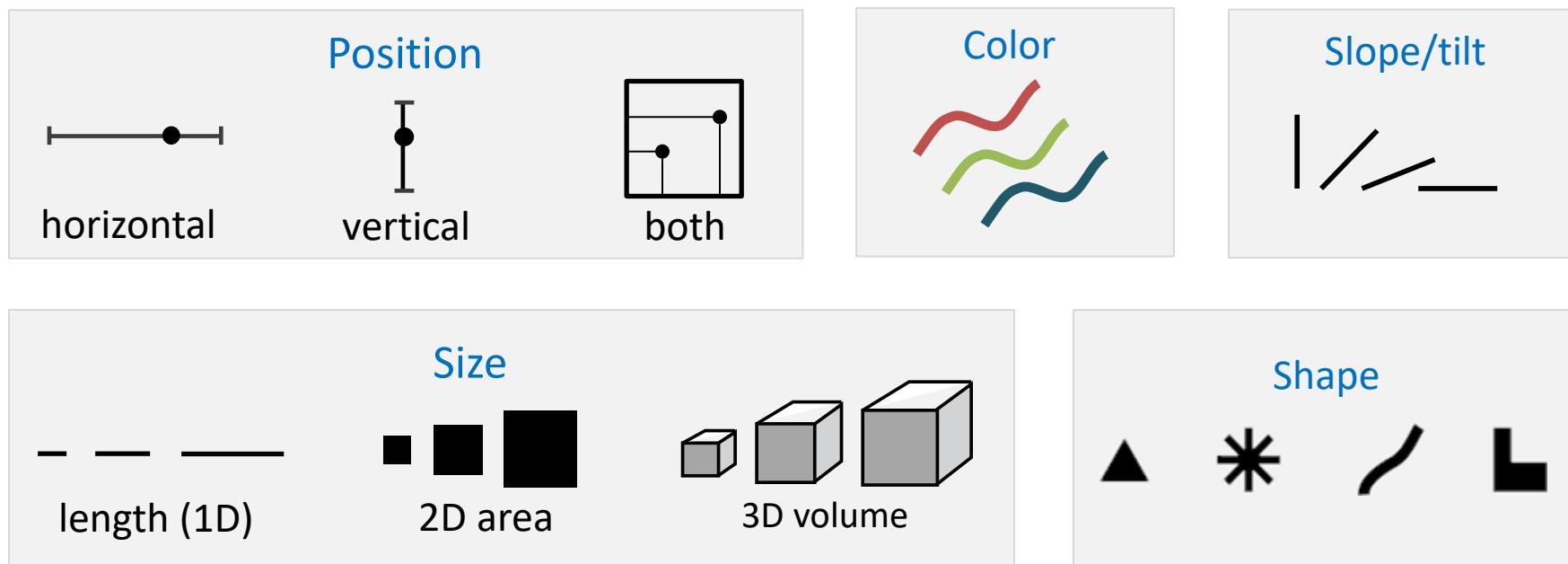
## Containment



[Collins 09]

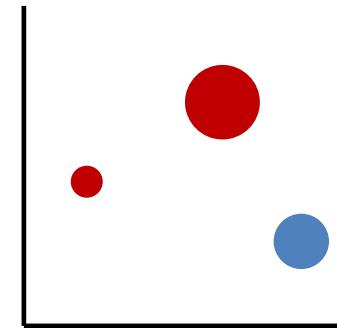
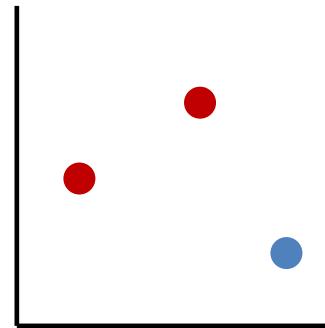
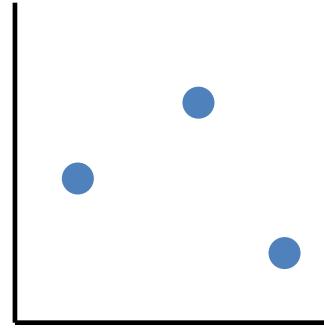
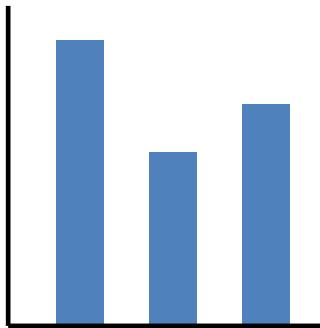
# Mapping techniques

- Mapping of data to a visual representation consists of:
  - **Visual channels** that control appearance of graph primitives based on data attributes



# Mapping techniques

Combination of graphical primitives and visual channels



## Visual channel

1:  
length/position

2:  
vertical position  
horizontal position

3:  
vertical position  
horizontal position  
color hue

4:  
vertical position  
horizontal position  
color hue  
size (area)

## Graphical primitive

line

point

point

point

# Mapping techniques

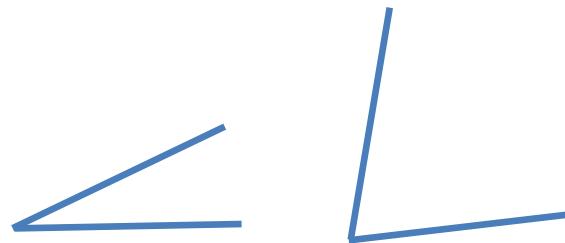
- Effectiveness principle
  - Some visual channels are better than others
  - Encode most important data attributes with most effective/accurate channels
- Properties of visual channels
  - Pop-out (emphasize important information)
  - Discriminability (how many usable steps?)
  - Separability (judge each channel independently)
  - Relative vs. absolute judgement

# Mapping techniques

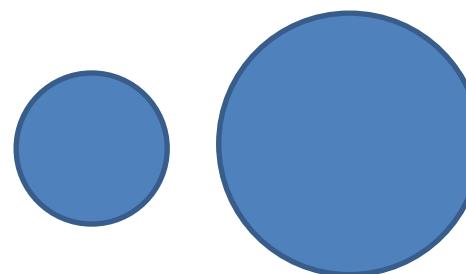
- Some visual channels can be compared more accurately



How much longer?



How much larger?

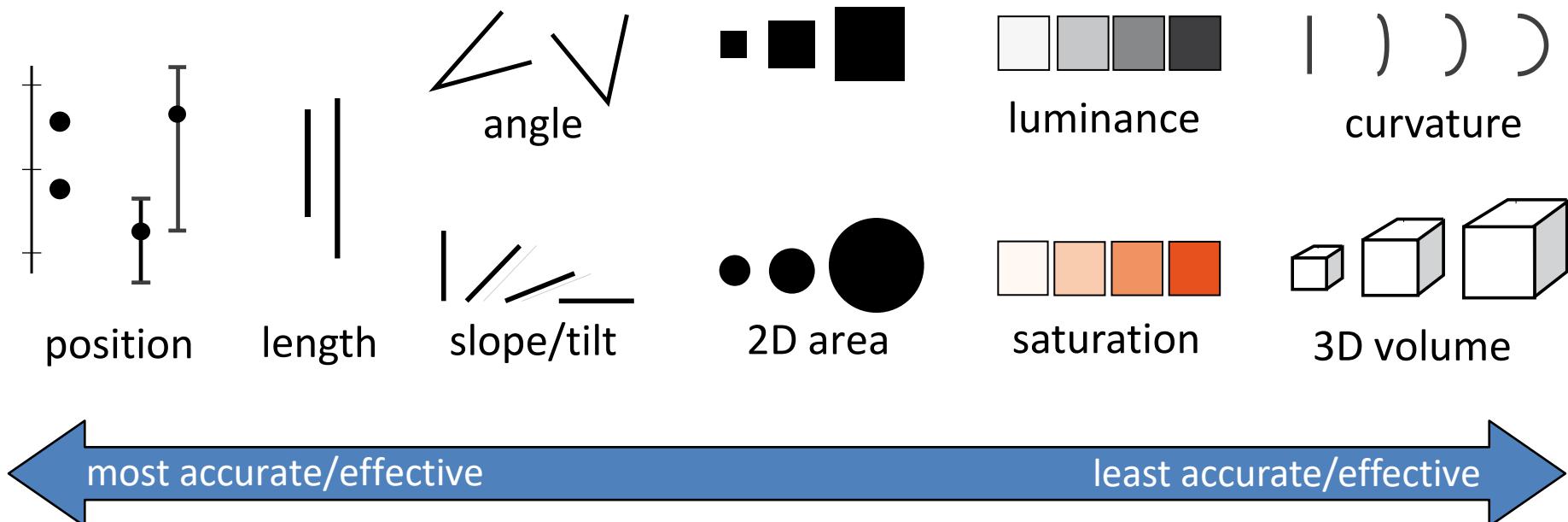


How much darker?

<http://woodgears.ca/eyeball/>

# Mapping techniques

- Accuracy/effectiveness (quantitative data)
    - Ranking of vis. channels derived from empirical studies



[Cleveland & McGill 84,  
Heer and Bostock 10,  
Munzner 14]

## Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT MC-GIL

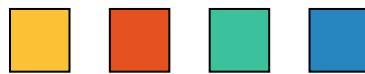
The subject of graphical methods for data analysis and for presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. The theory is based on graphical perception, which is defined as "the process of interpreting graphs—and it includes both theory and experience to test the theory. The theory deals with a small but important class of graphical displays, namely, perception. The first part is an identification of a set of elementary perceptual tasks that are carried out when people interpret graphical information from graphs. The second part is an ordering of the tasks in the basis of how accurately people perform them. The theory is tested by experimentation in which subjects respond to various graphical displays and to various graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guide for graph construction. Figure 1 illustrates the first two parts.

# Mapping techniques

- Accuracy/effectiveness (categorical data)
  - Ranking of vis. channels derived from empirical studies



Spatial  
region



Color hue



Motion



Shape

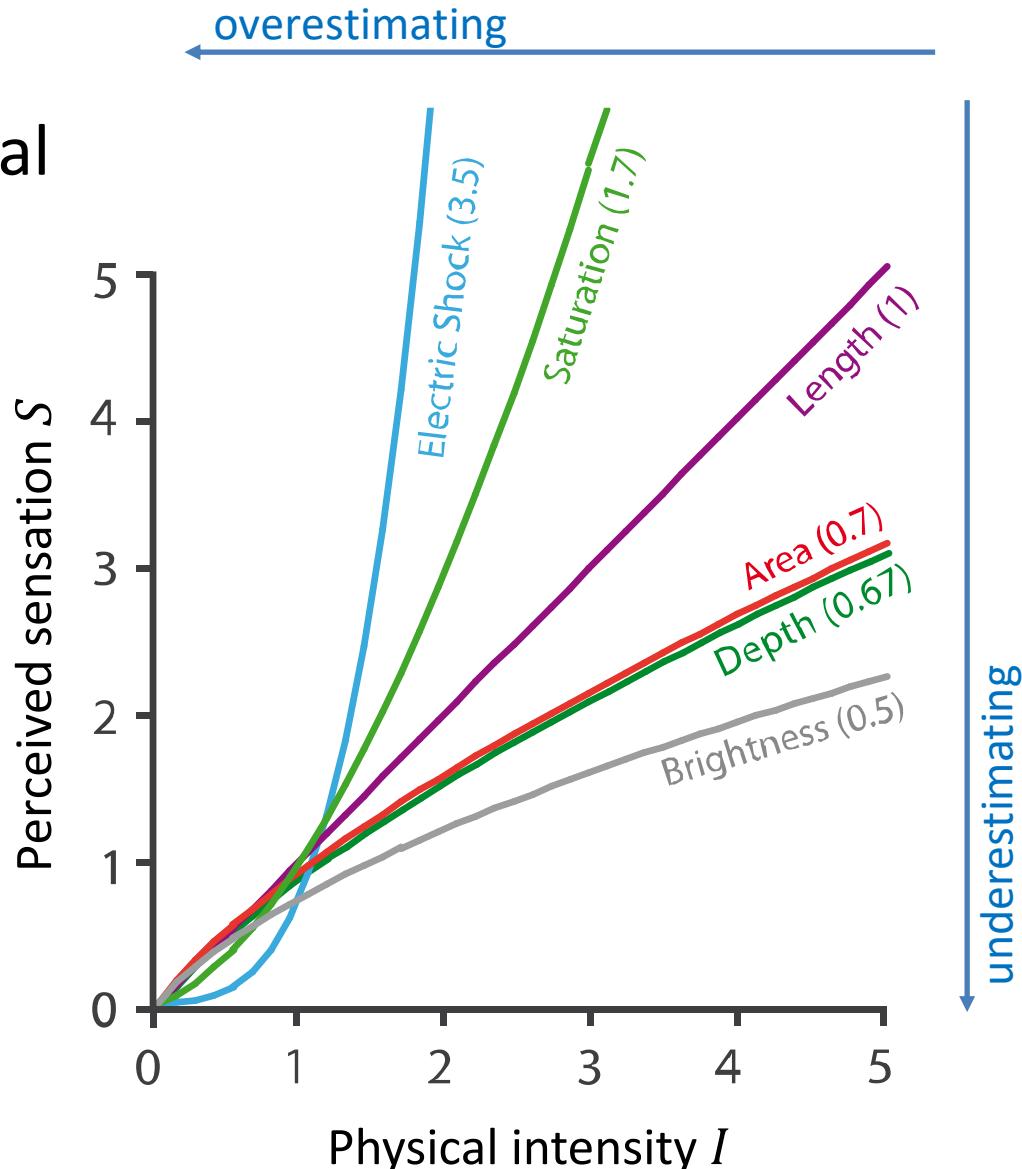


**Expressive mapping:**  
Match type of visual  
channel to data type

# Mapping techniques

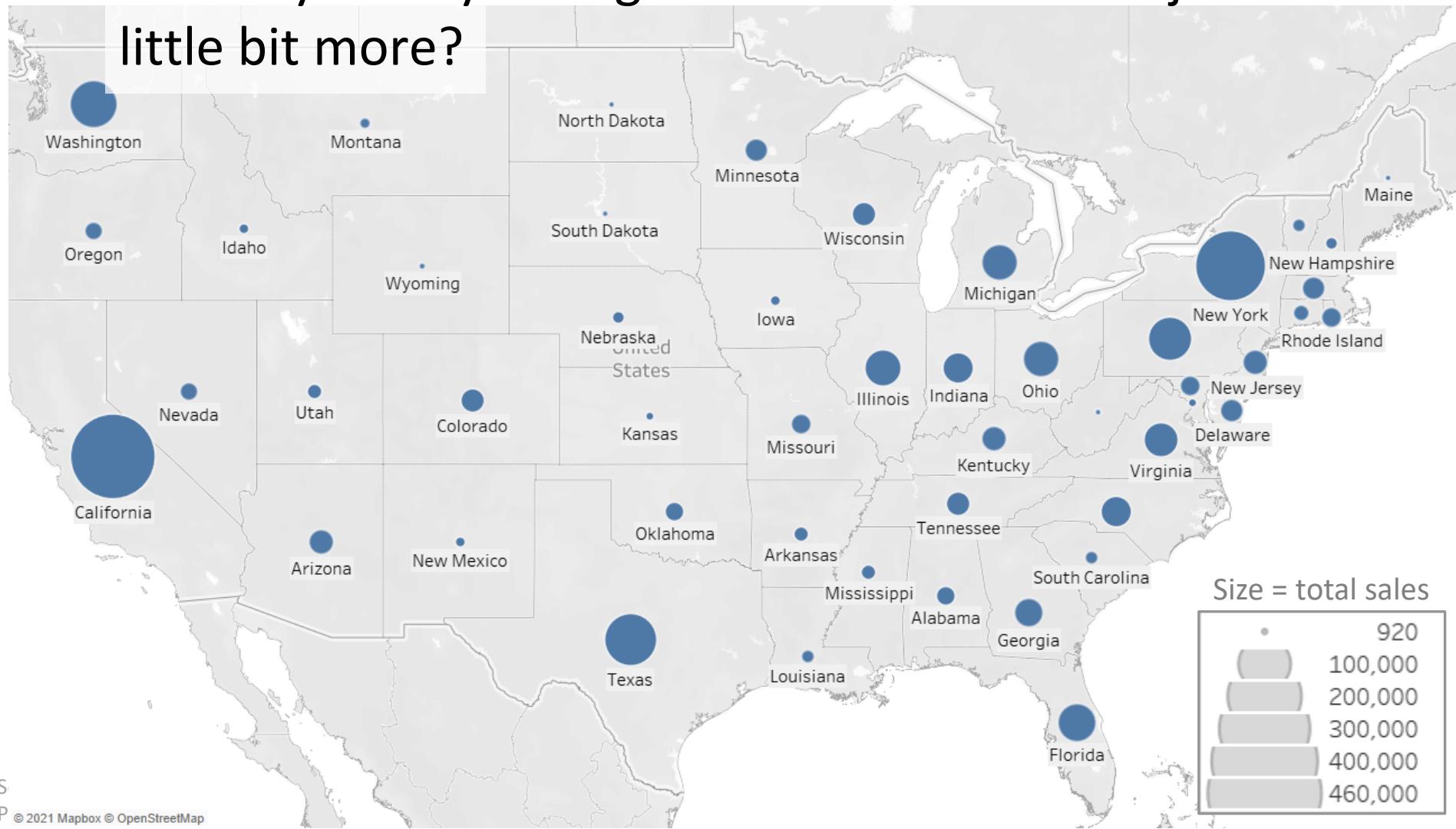
- Accuracy/effectiveness
  - Steven's psychophysical power law:  $S = I^\gamma$

- Length is accurate (linear)
- Other channels are magnified or compressed



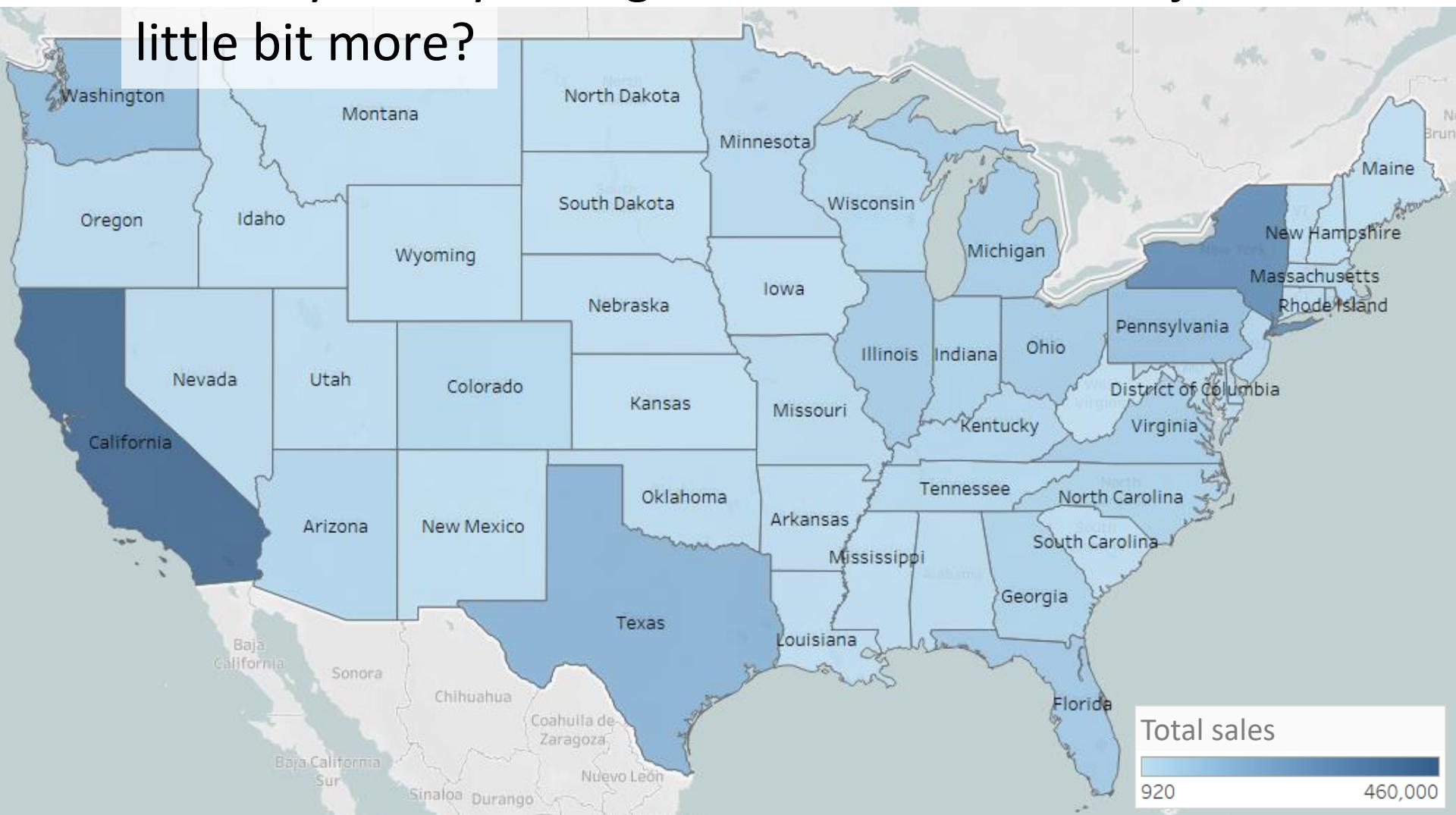
# Mapping techniques

- Which are the top five states in terms of sales?
- Are they clearly selling more than the rest or just a little bit more?



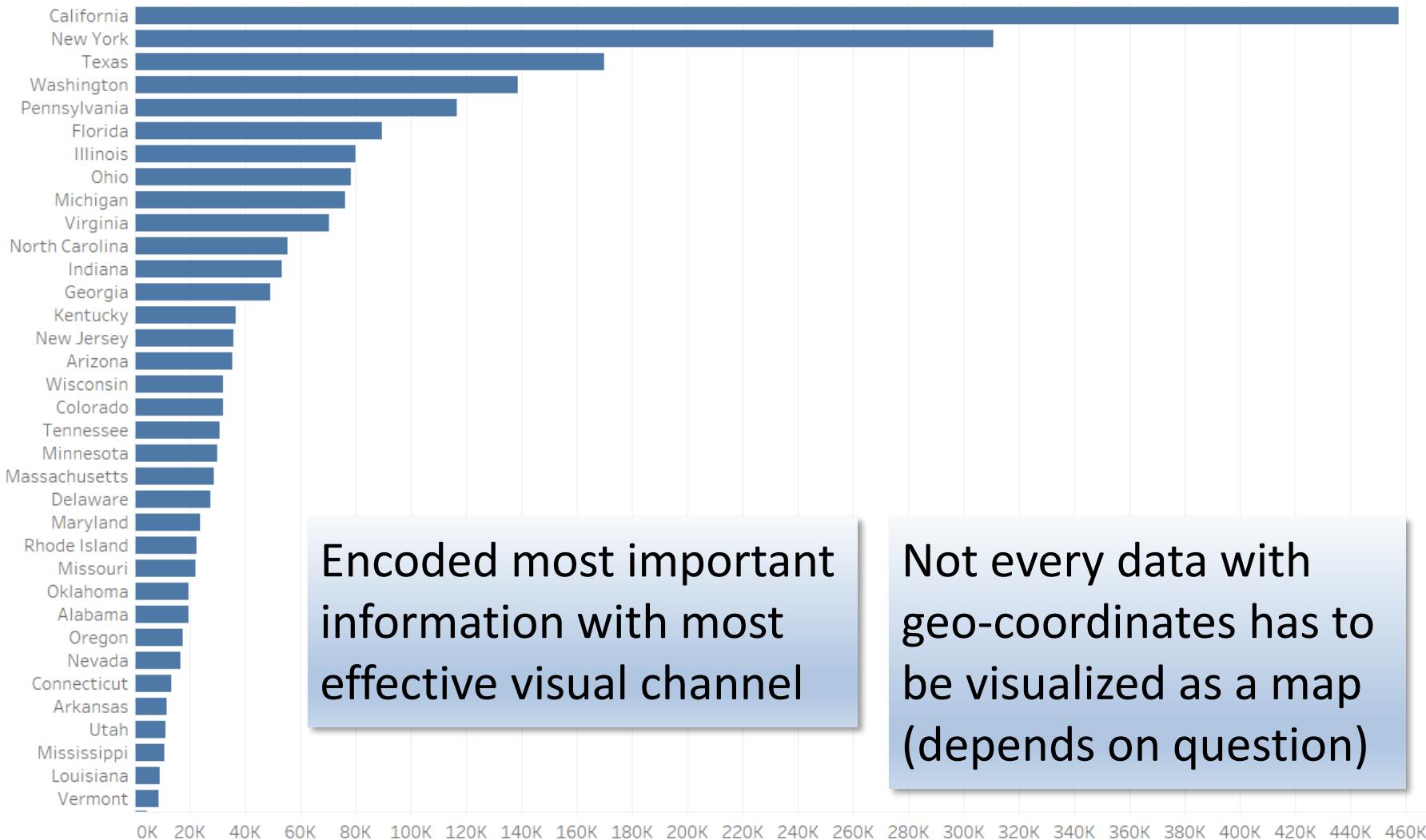
# Mapping techniques

- Which are the top five states in terms of sales?
  - Are they clearly selling more than the rest or just a little bit more?



- Which are the top five states in terms of sales?

- Are they clearly selling more than the rest or just a little bit more?



Encoded most important information with most effective visual channel

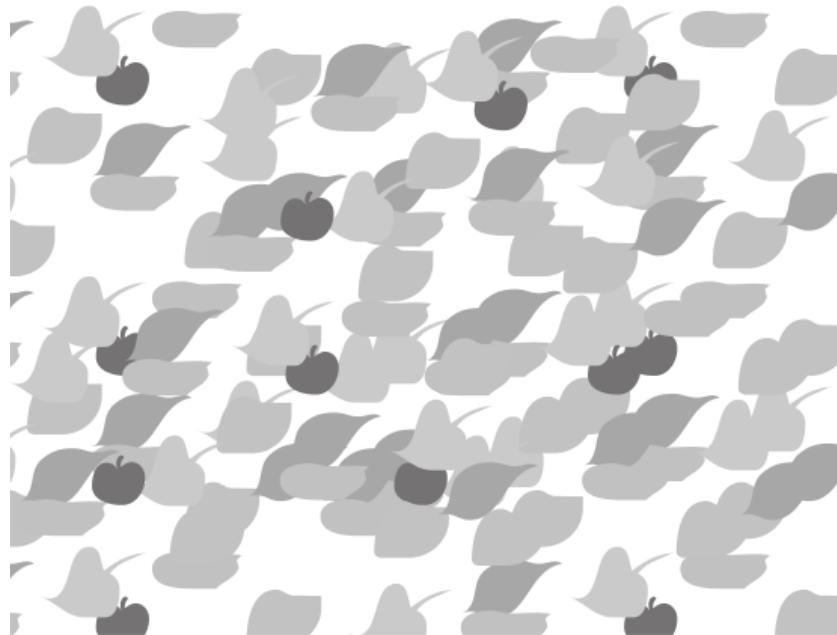
Not every data with geo-coordinates has to be visualized as a map (depends on question)

# Mapping techniques

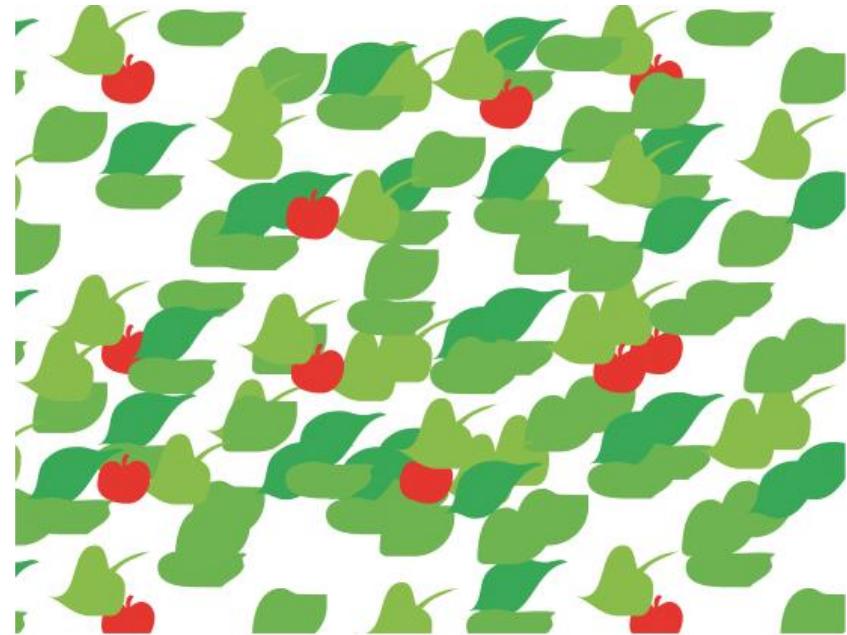
- Effectiveness principle
  - Some visual channels are better than others
  - Encode most important data attributes with most effective/accurate channels
- Properties of visual channels
  - Pop-out (emphasize important information)
  - Discriminability (how many usable steps?)
  - Separability (judge each channel independently)
  - Relative vs. absolute judgement

# Mapping techniques

- Some visual channels “pop out”



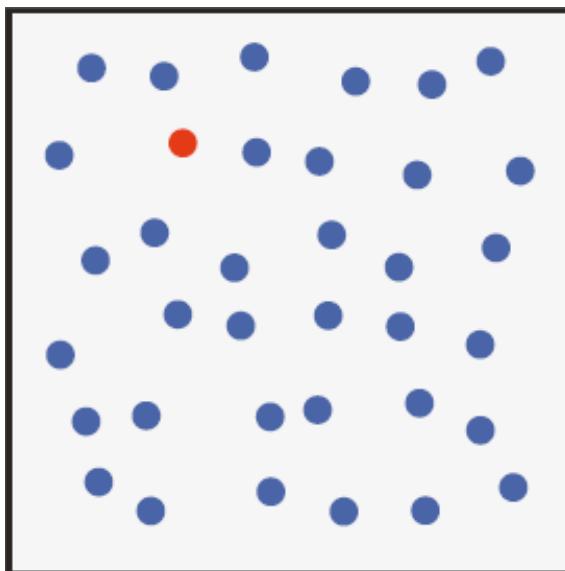
Where are the cherries?



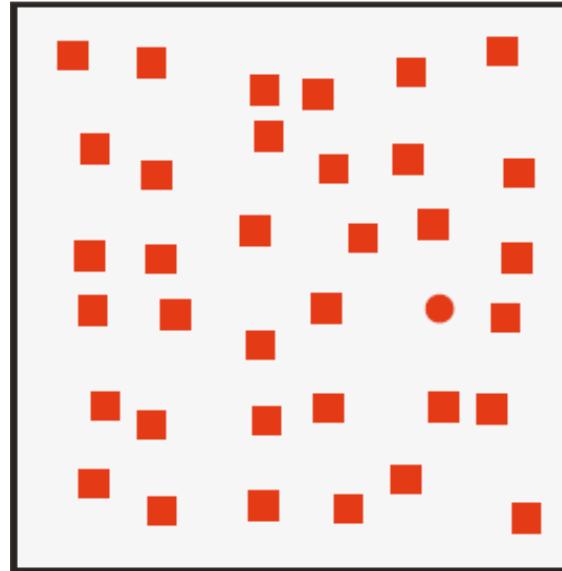
Hunters & gatherers

# Mapping techniques

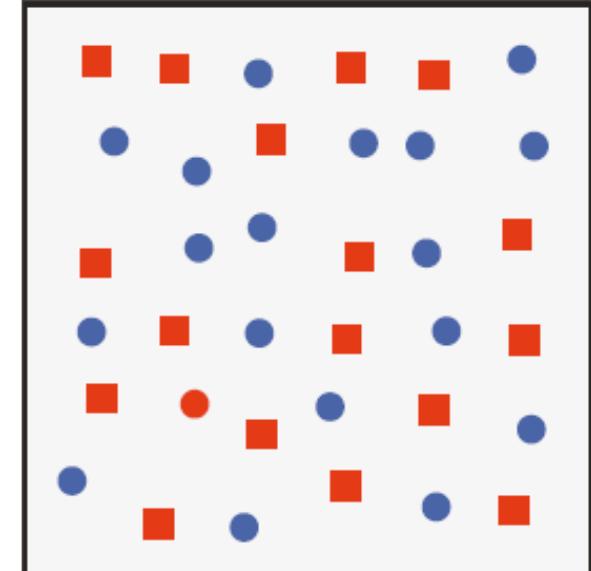
- Pop-out
  - Preattentive processing: automatic and parallel detection of basic features in visual information (200-250 msec)
  - Speed independent of distractor count
  - Works on many individual channels



Color



Shape



Combination of channels  
usually requires serial search

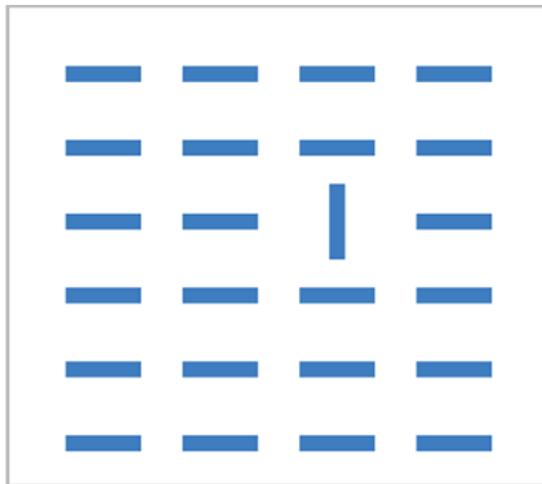
# The value of visualization

2	1	4	3	9	5	6	7	8	2	3	6	5	9	4	0	1
6	7	9	3	4	9	0	5	6	2	5	8	4	0	5	2	6
9	8	2	6	3	5	9	3	2	9	3	7	2	6	3	4	8
8	1	6	2	3	8	7	9	5	0	2	3	9	2	8	4	3
0	9	1	8	5	4	2	9	4	7	4	6	8	4	0	2	9
3	9	2	7	3	6	6	5	2	9	4	0	4	9	4	8	6
5	2	4	3	6	4	8	1	0	3	9	4	8	4	7	3	2
8	6	2	3	0	8	7	3	6	2	5	4	4	8	3	5	0

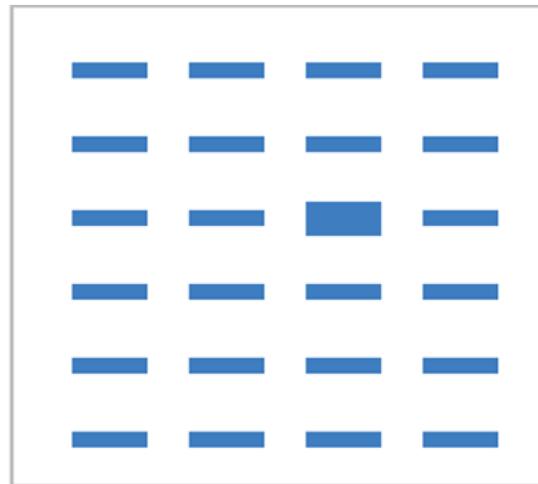
<https://vimeo.com/29684853>

# Mapping techniques

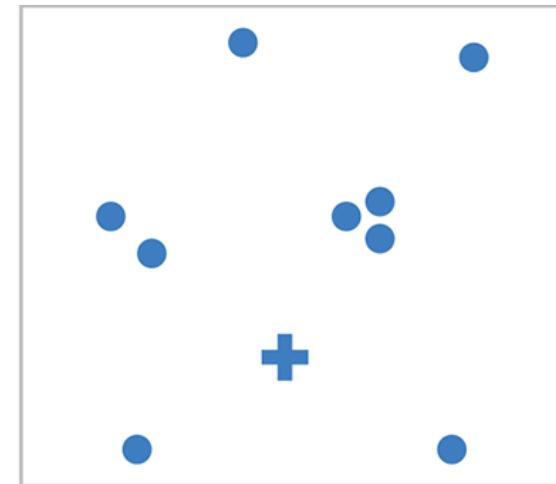
- Pop-out



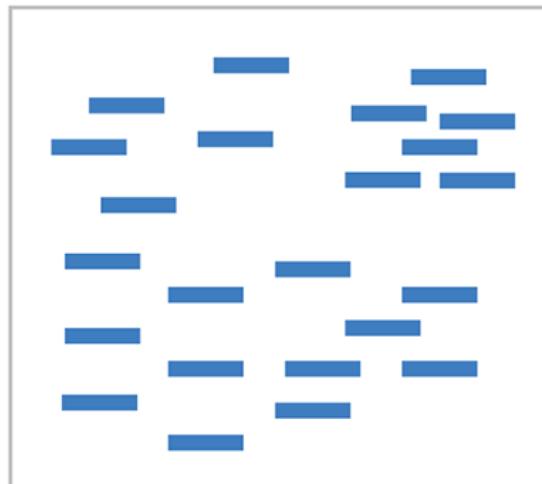
Tilt



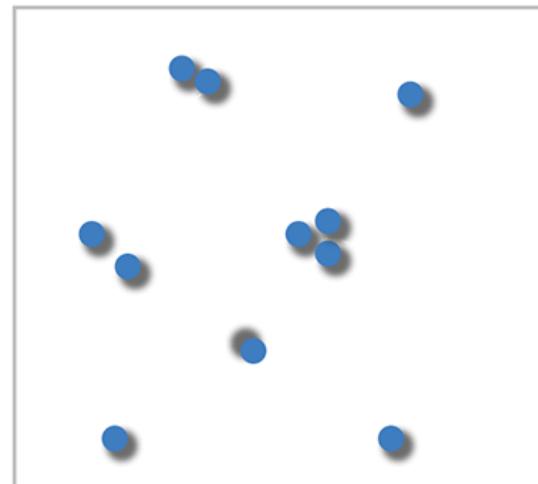
Size



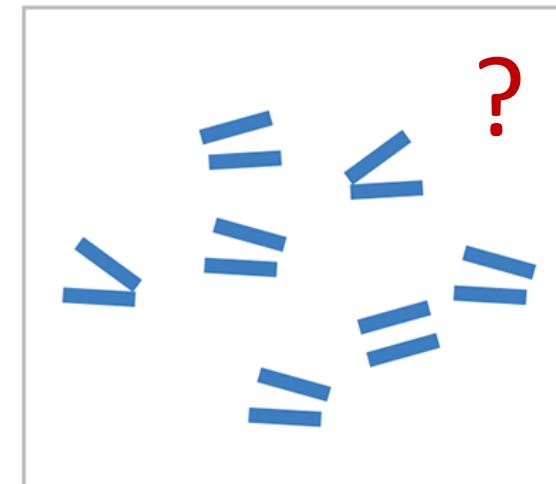
Shape



Proximity



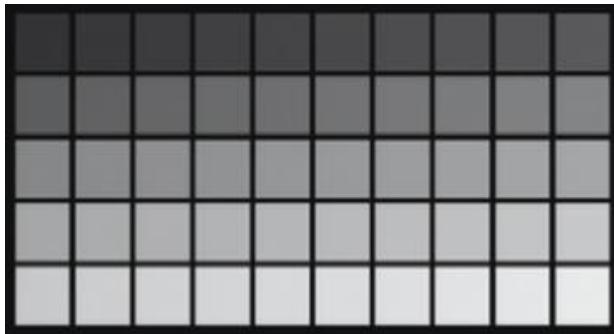
Shadow direction



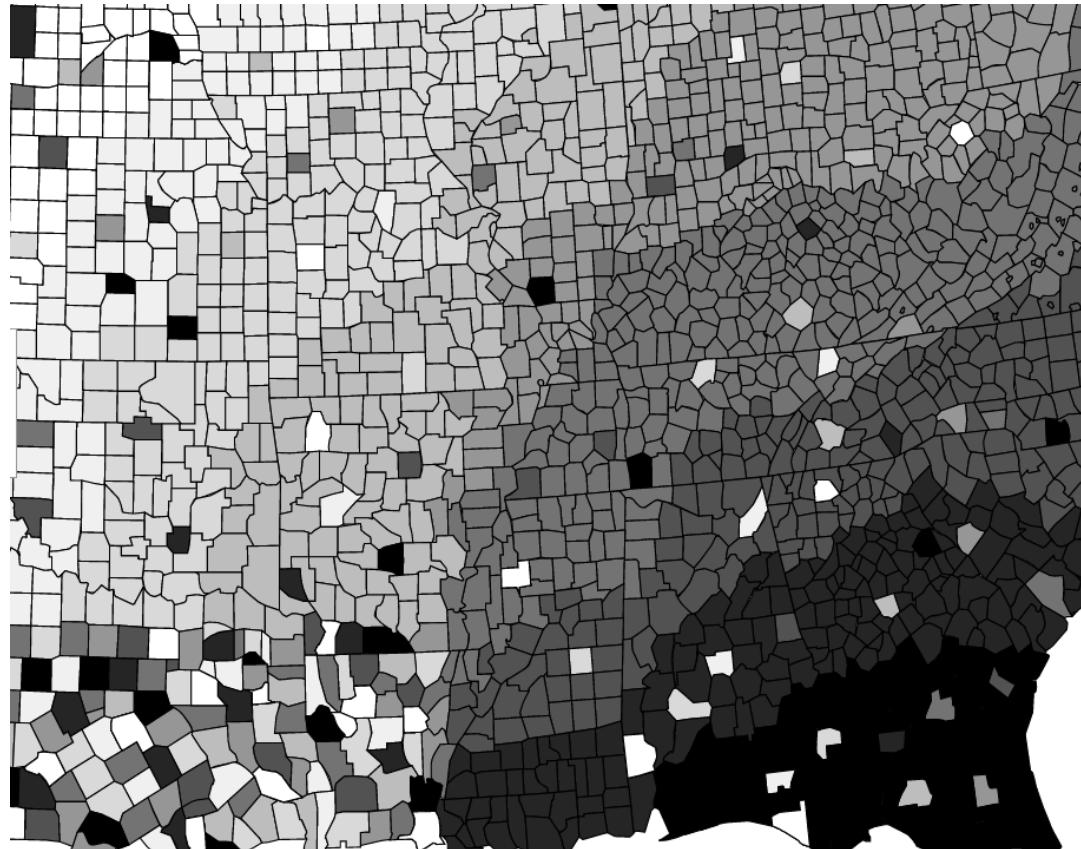
Parallel lines (serial search)

# Mapping techniques

- Discriminability: How many usable steps?
  - Must be sufficient for number of discriminable bins



We can only distinguish a limited number of colors / brightness levels

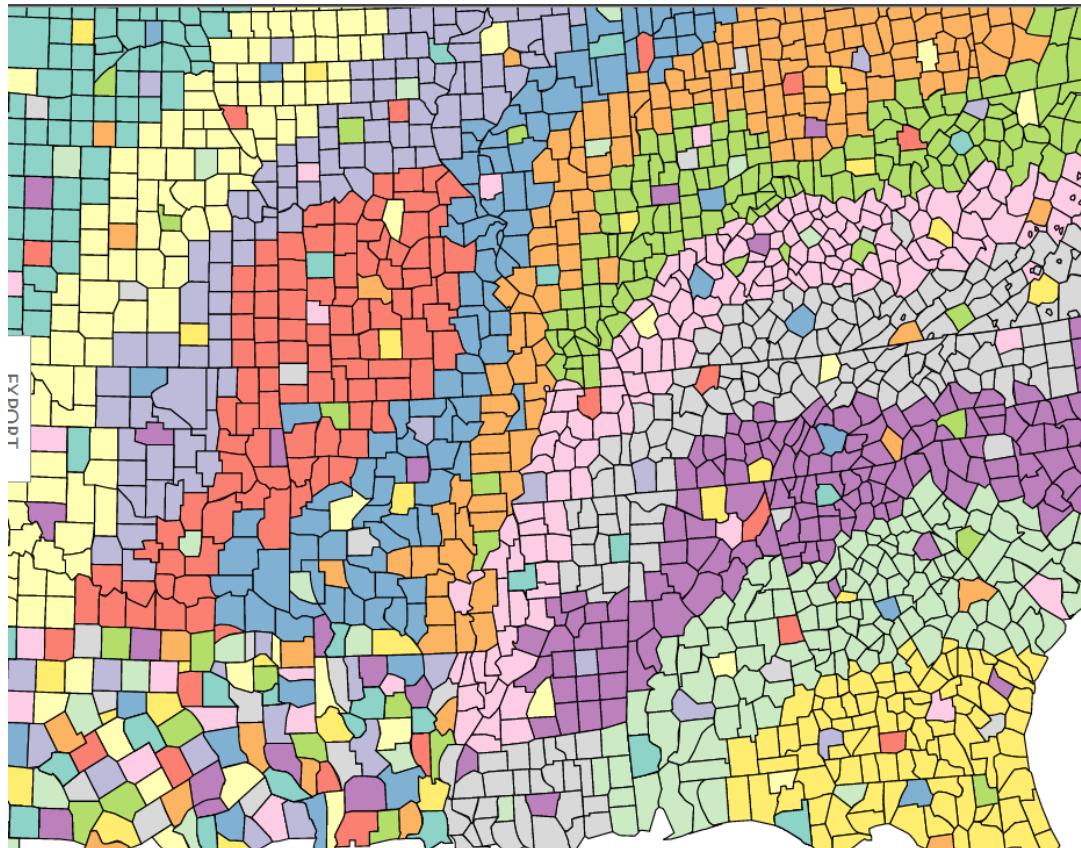
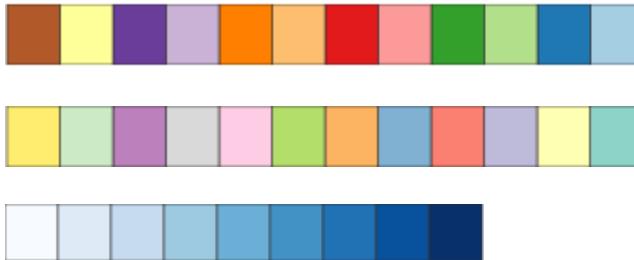


[colorbrewer2.org](http://colorbrewer2.org)

# Mapping techniques

- Discriminability: How many usable steps?
  - Must be sufficient for number of discriminable bins

9-12 color hues

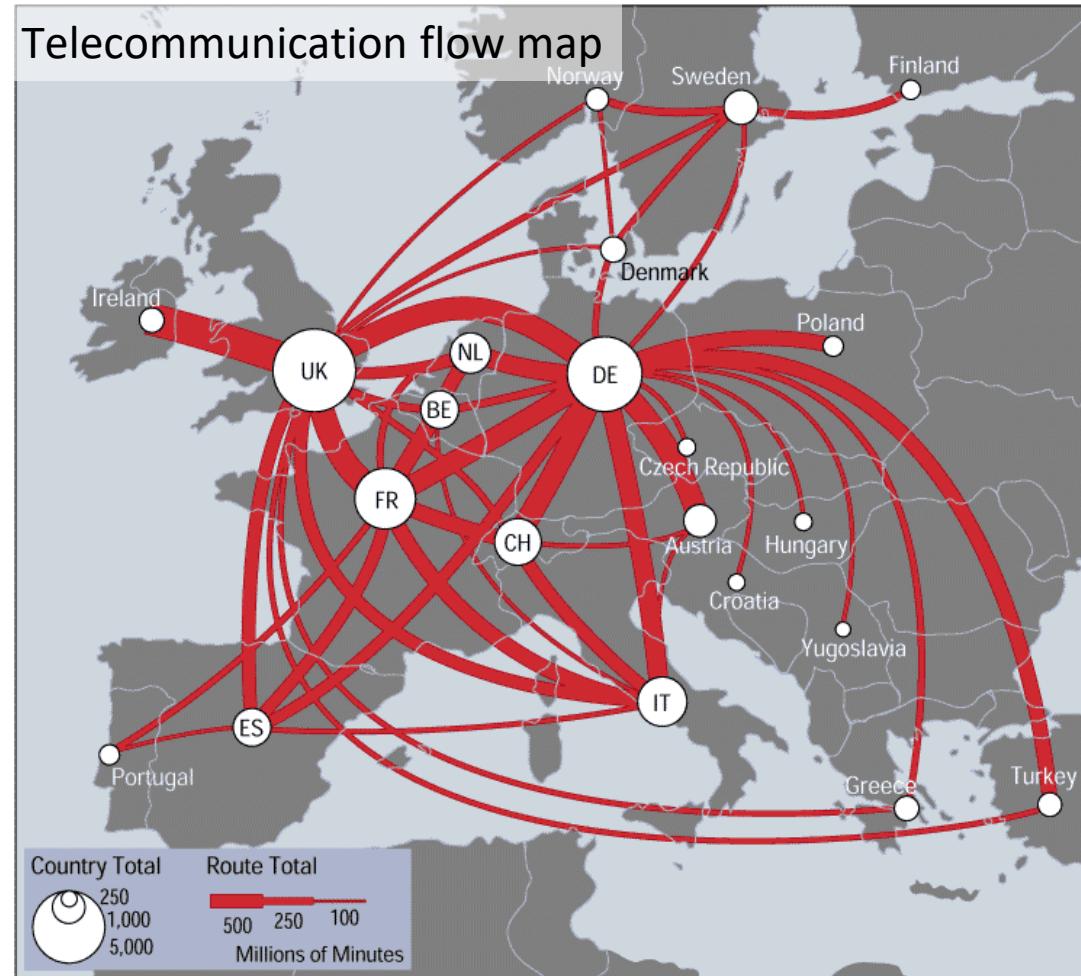


[colorbrewer2.org](http://colorbrewer2.org)

# Mapping techniques

- Discriminability: How many usable steps?
  - Must be sufficient for number of discriminable bins

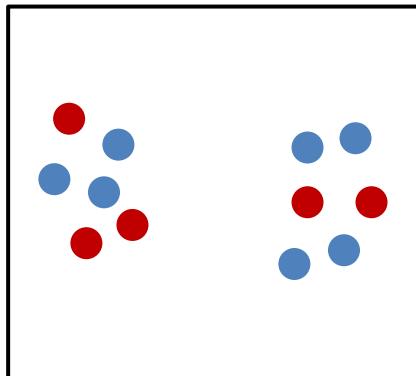
~5 size bins



# Mapping techniques

- Separable vs. integral visual channels

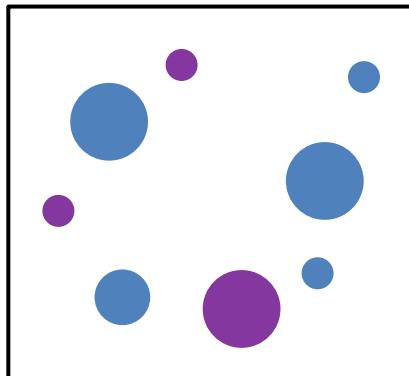
Color + position



Fully separable

2 groups each

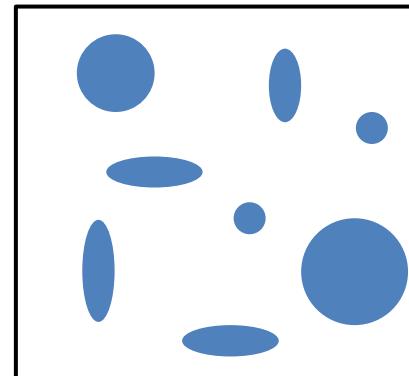
Color + size



Some interference

2 groups each

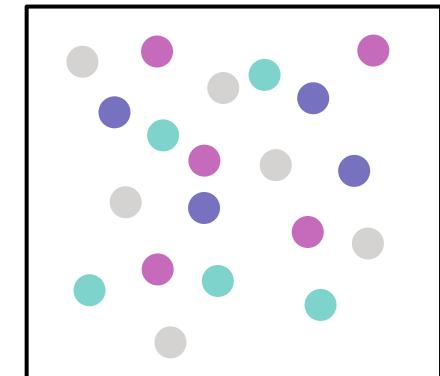
Width + height



Some/significant interference

3 groups total:  
integral area

Red + green



Major interference

4 groups total:  
integral color

Separable visual channels

We are able to judge each  
visual channel independently

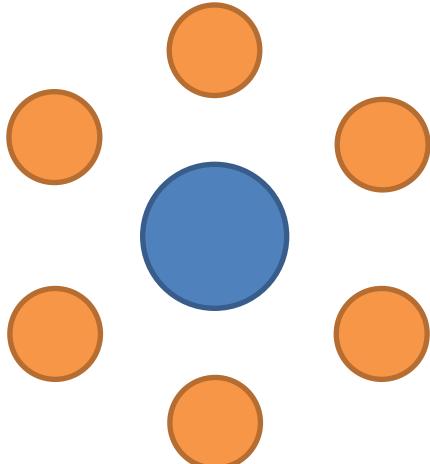
Integral visual channels

Channels are viewed holistically

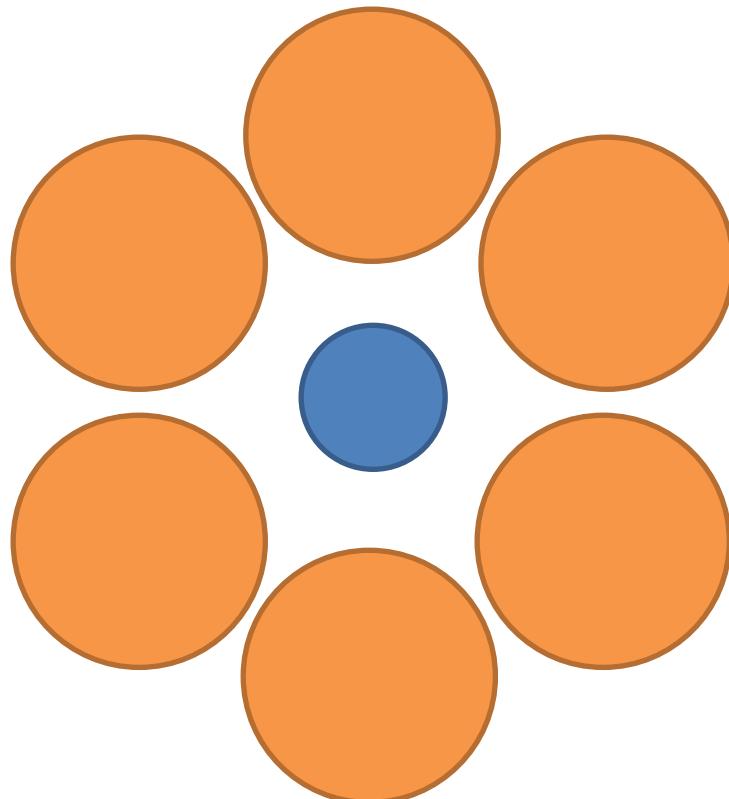
[Ware 13, Munzner 14]

# Mapping techniques

- Relative vs. absolute judgments
  - Perception highly context-dependent
  - Perceptual system mostly operates with **relative judgments**, not absolute ones

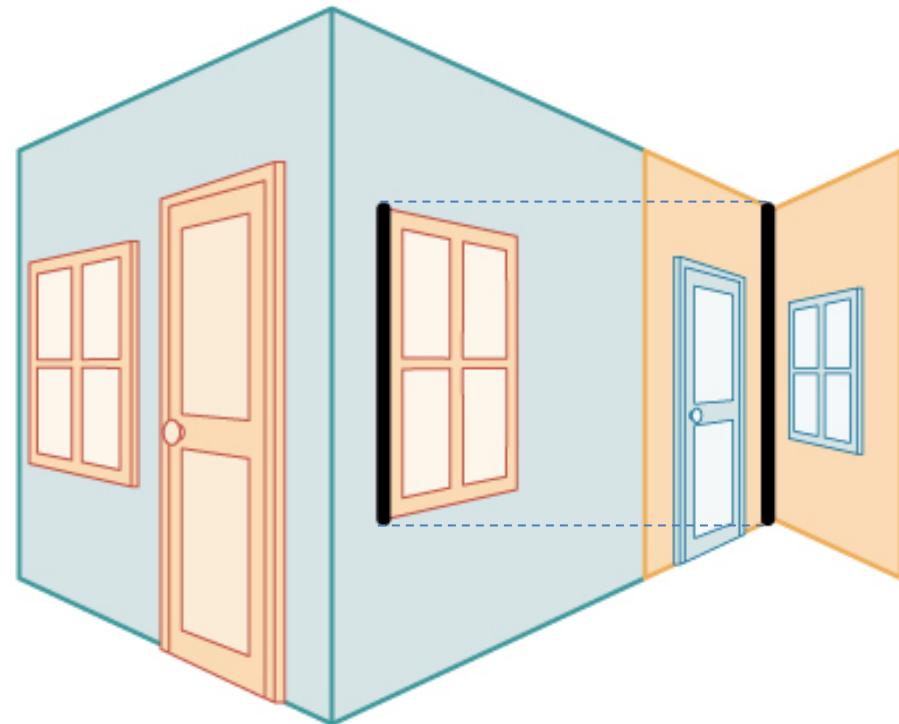
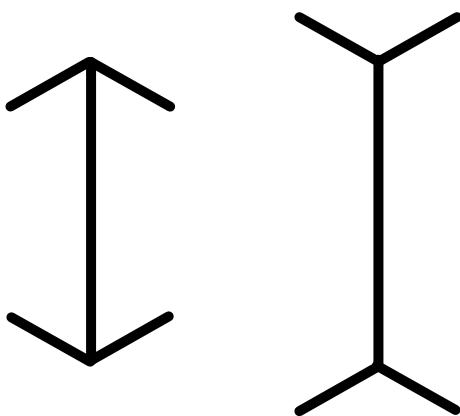


Which blue circle is bigger?



# Mapping techniques

- Relative vs. absolute judgments



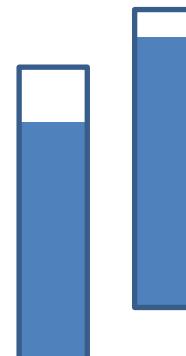
Müller-Lyer illusion

# Mapping techniques

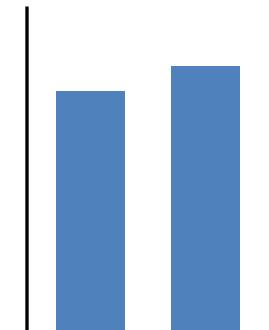
- Relative vs. absolute judgments
  - Perceptual system mostly operates with relative judgments, not absolute ones
  - [Weber's Law](#): just-noticeable difference is a fixed percentage of the magnitude of the stimuli (e.g., bar length)
    - filled rectangles differ in length by 1:9 → difficult judgment
    - white rectangles differ in length by 1:2 → easy judgment



length



position along  
unaligned  
common scale

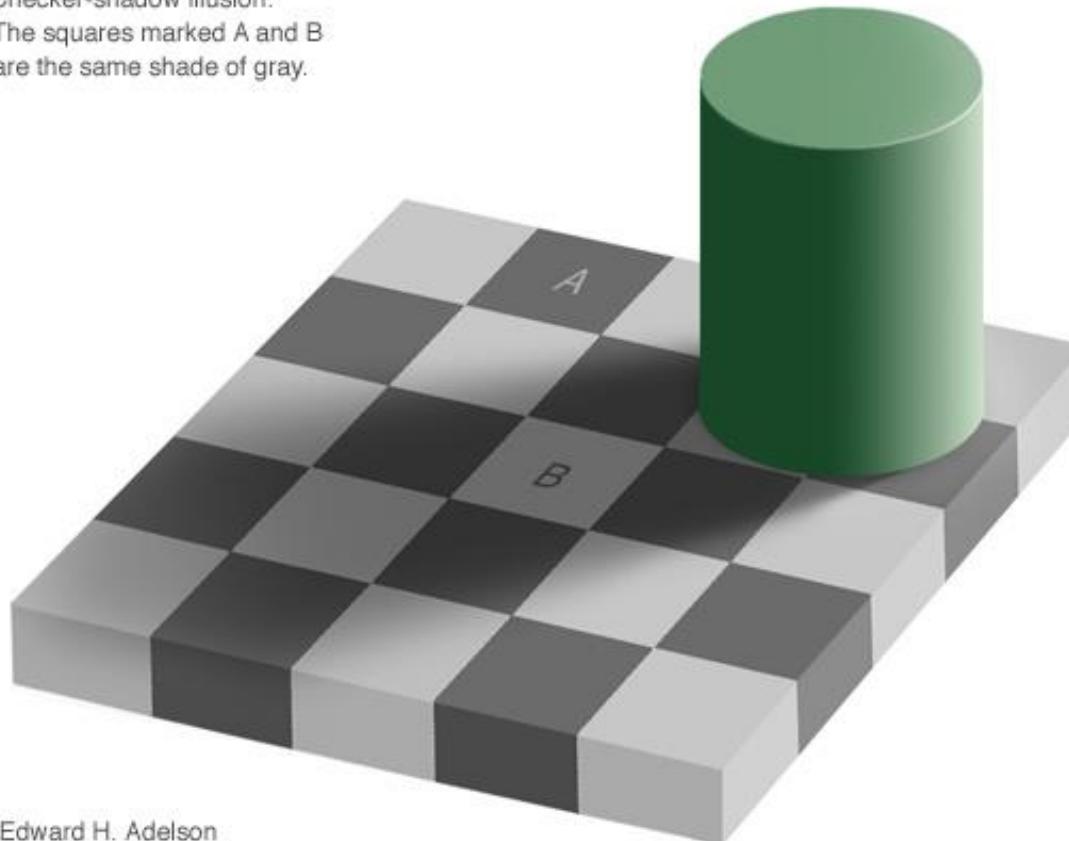


position along  
aligned scale

# Mapping techniques

- Perceived color is highly context dependent

Checker-shadow illusion:  
The squares marked A and B  
are the same shade of gray.



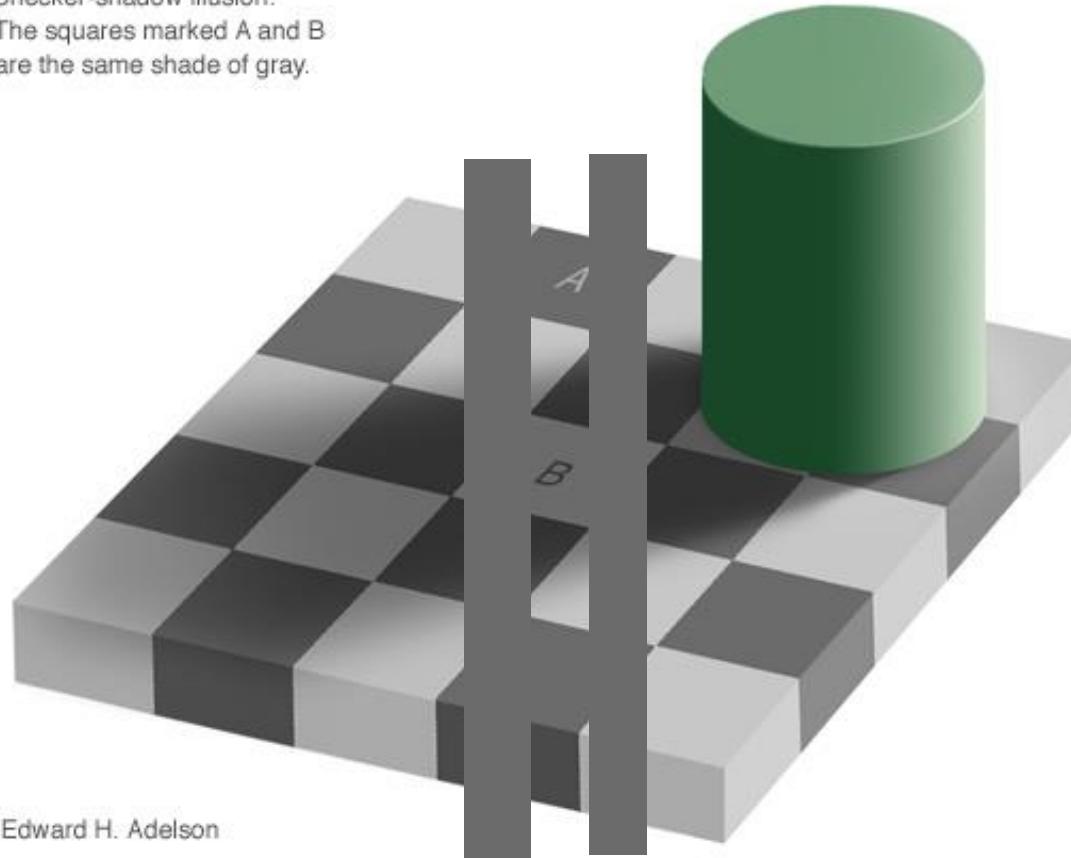
Edward H. Adelson

[http://web.mit.edu/persci/people/adelson/checkershadow\\_illusion.html](http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html)

# Mapping techniques

- Perceived color is highly context dependent

Checker-shadow illusion:  
The squares marked A and B  
are the same shade of gray.

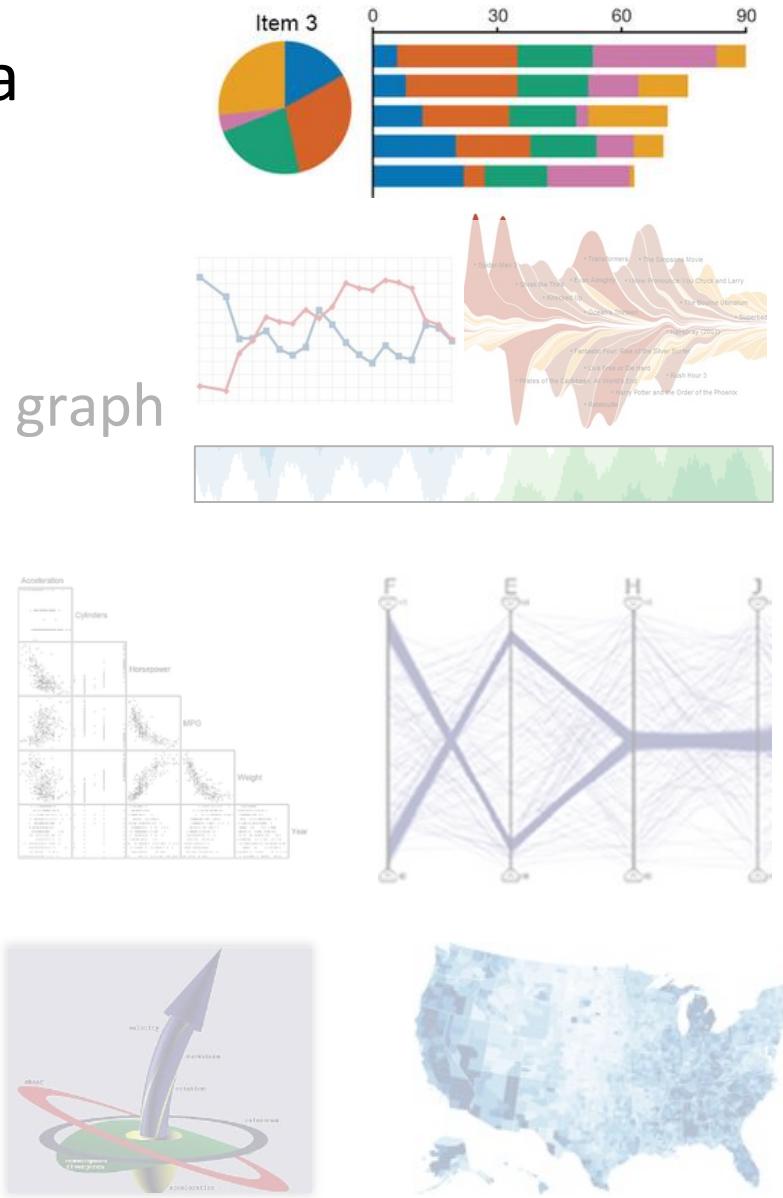


Edward H. Adelson

[http://web.mit.edu/persci/people/adelson/checkershadow\\_illusion.html](http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html)

# Diagram techniques

- Categorical + quantitative data
  - Bar/pie chart, stacked bars
- Time-dependent data
  - Line graph, ThemeRiver, Horizon graph
- Single and multiple variables
  - Histogram, scatterplot, parallel coordinates
  - Glyphs, color mapping



# Recap: Attribute types

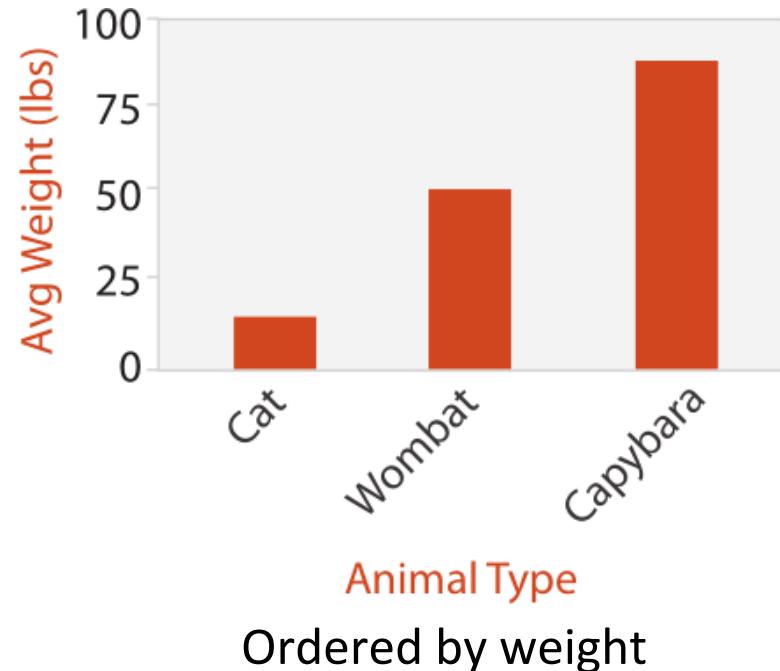
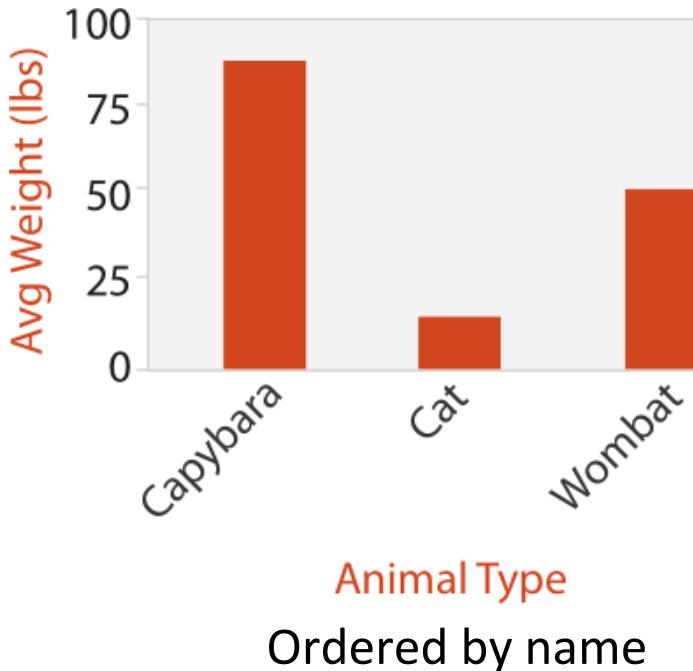
- **Quantitative** (numerical, measurable)
  - Objective data produced through a systematic process, not subject to interpretation (e.g., length, mass, temperature)
  - Metric scale – allows measure of distance
  - **Continuous** (real) or **discrete** (distinct & separate values)
- **Qualitative** (categorical, not measurable)
  - No metric scale; cannot be measured
  - Requires a subjective decision in order to be categorized
  - Discrete



**Expressive mapping:**  
Match type of visual  
channel to data type

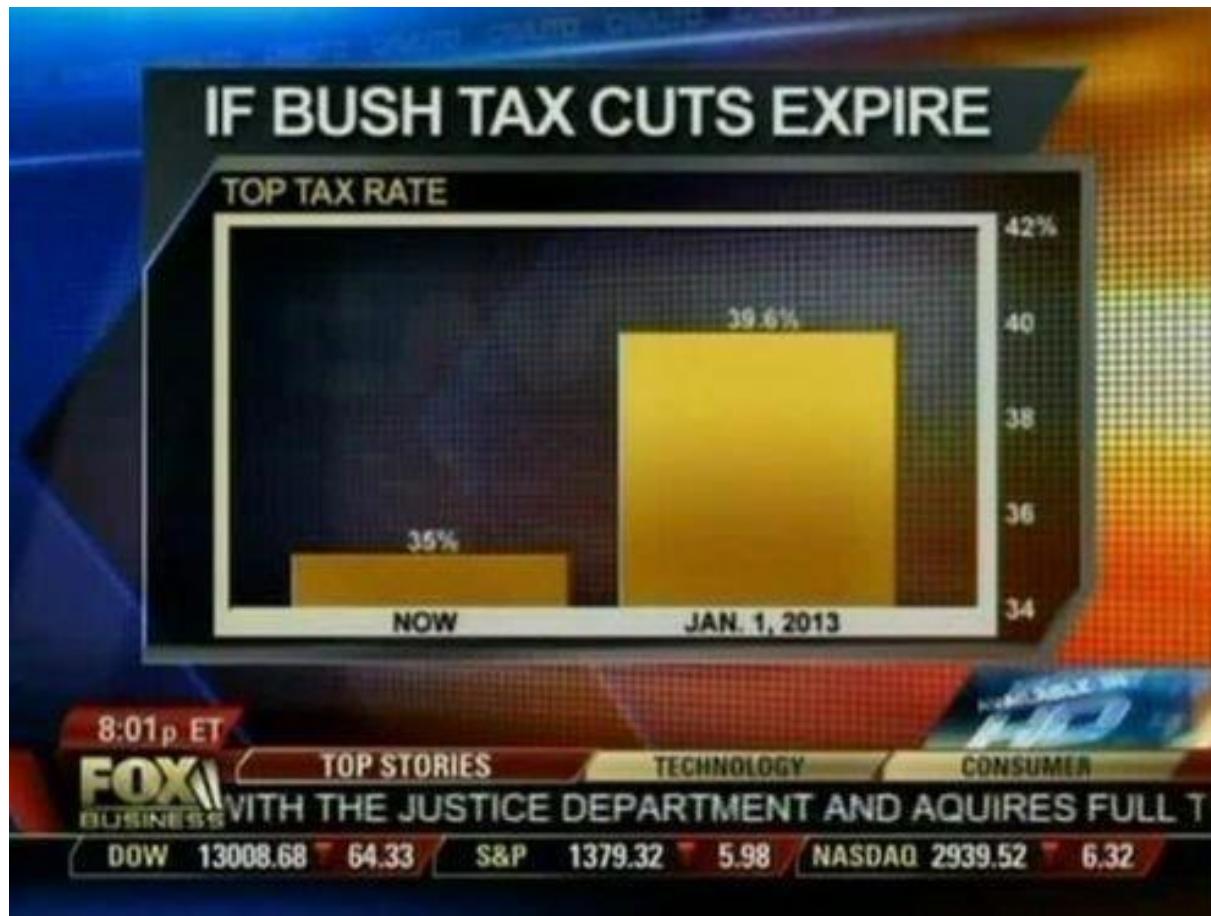
# Diagram techniques

- Bar chart
  - Attrib. 1: categorical → horizontal position
  - Attrib. 2: quantitative (dependent) → length/vertical position



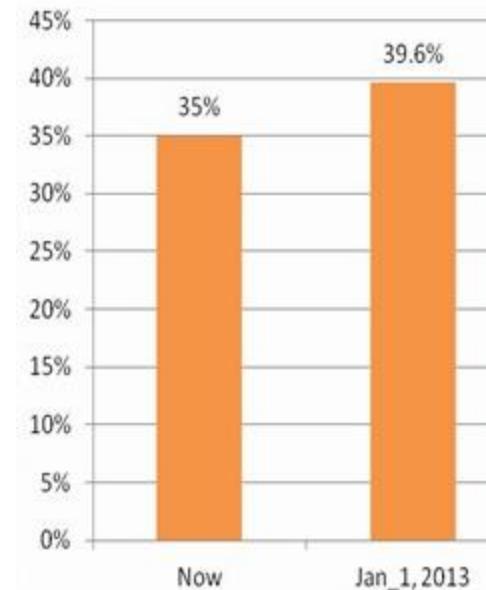
# Diagram techniques

- Bars should always start at zero!



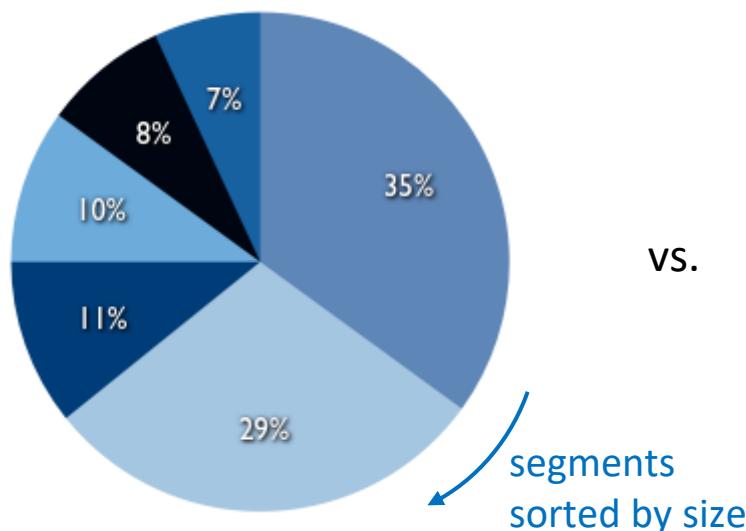
Taking misleading statistics to a new level, KDnuggets, 2012.

<http://www.kdnuggets.com/2012/12/taking-misleading-statistics-to-a-new-level.html>

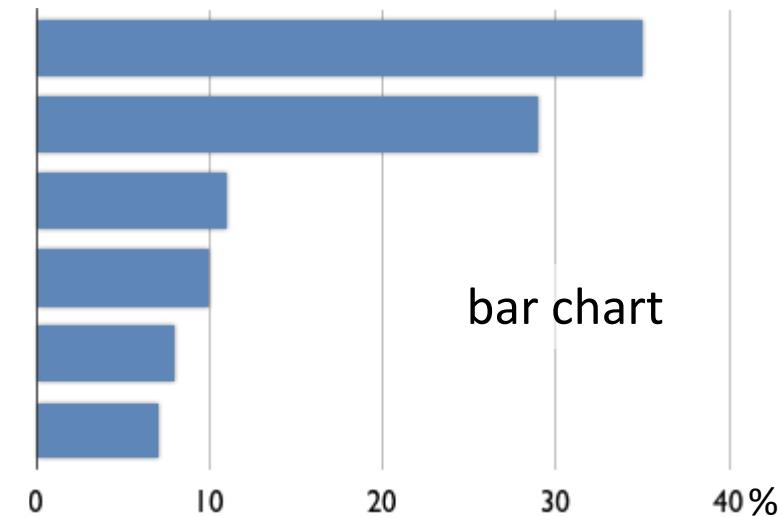


# Diagram techniques

- Pie chart
  - Pie chart splits population (100%) into parts
  - Attrib. 1: categorical → color
  - Attrib. 2: quantitative (dependent) → angle
  - However, angle/area less accurate than bar length

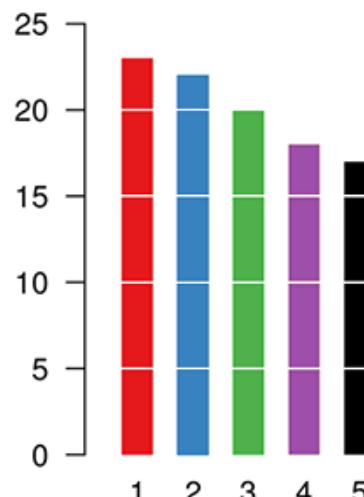
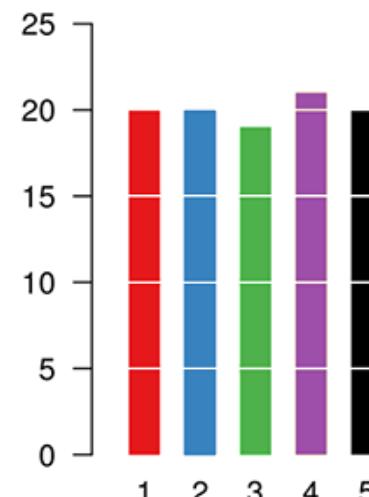
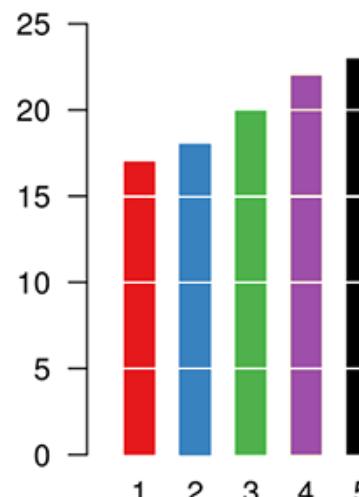
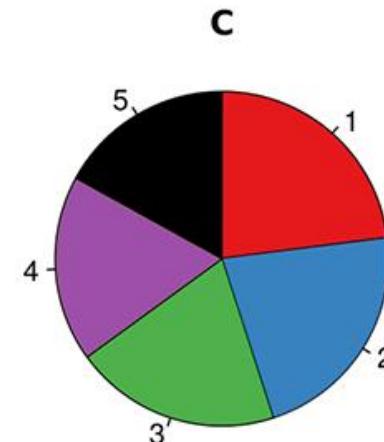
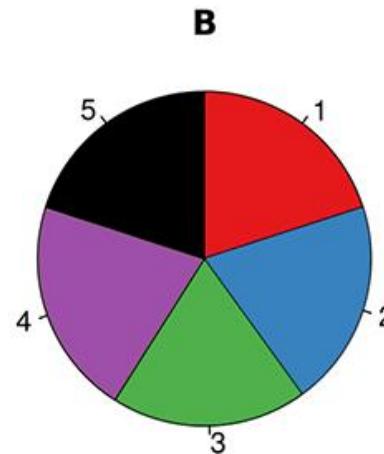
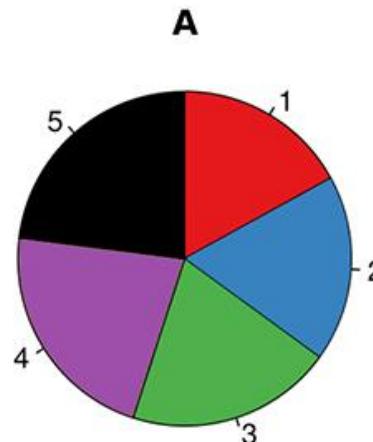


vs.



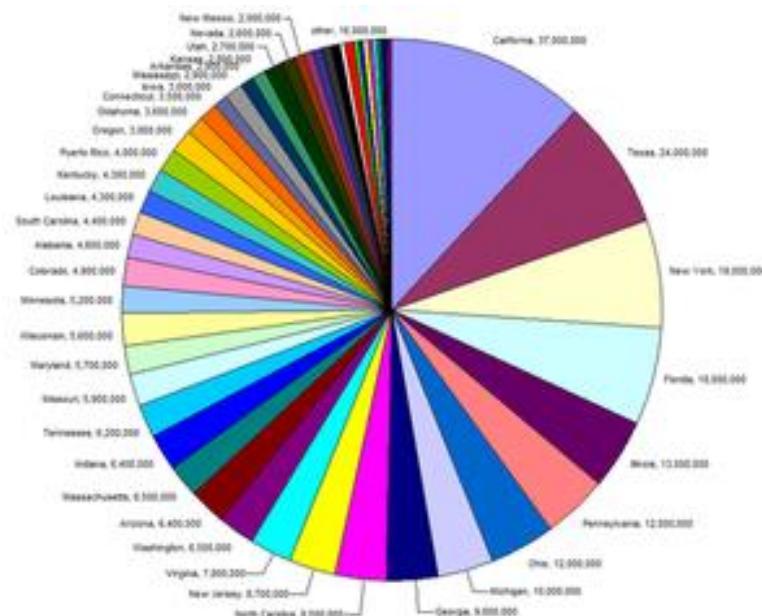
# Diagram techniques

- Pie chart vs. Bar chart
  - Angle/area judgment less accurate than bar length



# Diagram techniques

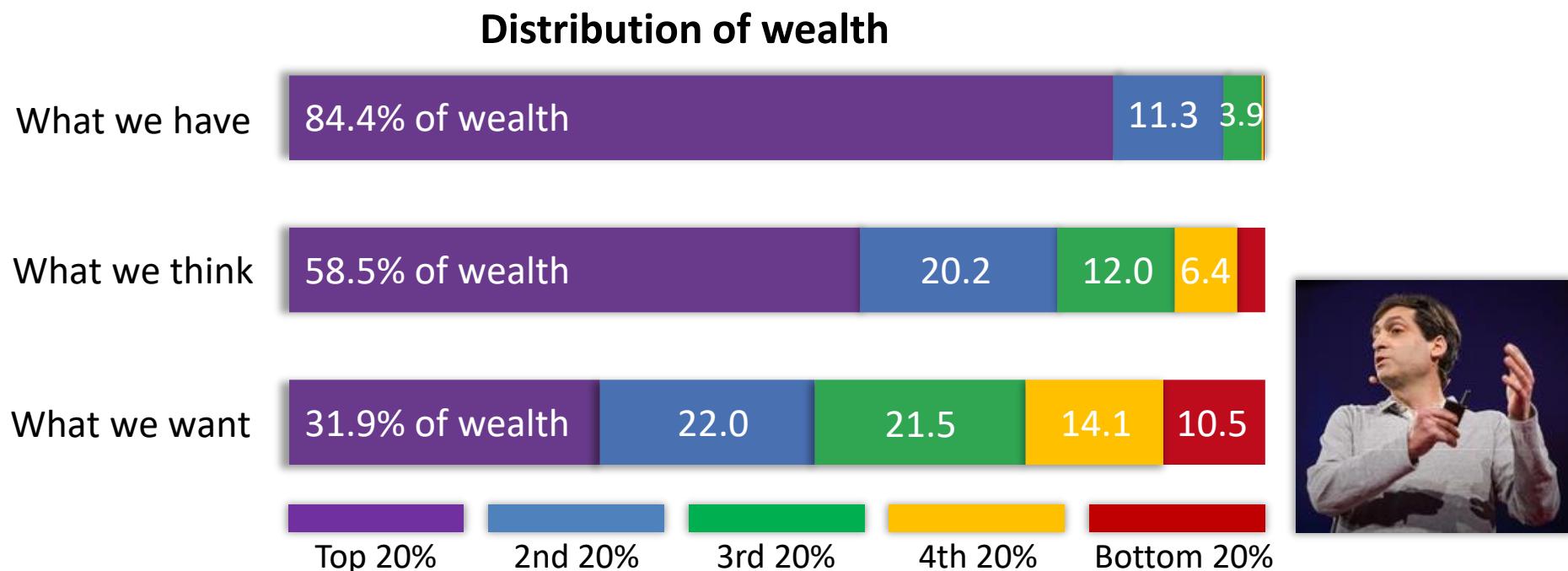
- When to use a pie chart?
  - Often bar chart is a better choice!
  - Do the parts make up a meaningful whole?
  - Are the parts mutually exclusive?
  - Do you want to compare the parts to each other or the parts to the whole?
  - How many parts do you have?



R. Kosara, [eagereyes.org/techniques/pie-charts](http://eagereyes.org/techniques/pie-charts)

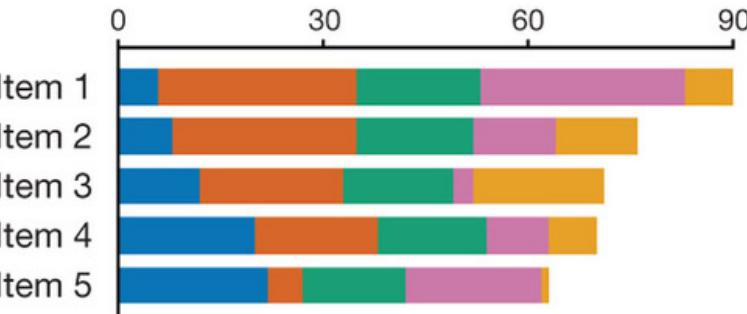
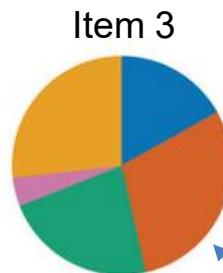
# Diagram techniques

- Stacked bar chart
  - Quantitative data wrt 2 categorical vars (horizontal & vertical)
  - Investigate part-to-whole relationship (100%)
  - Length and color hue



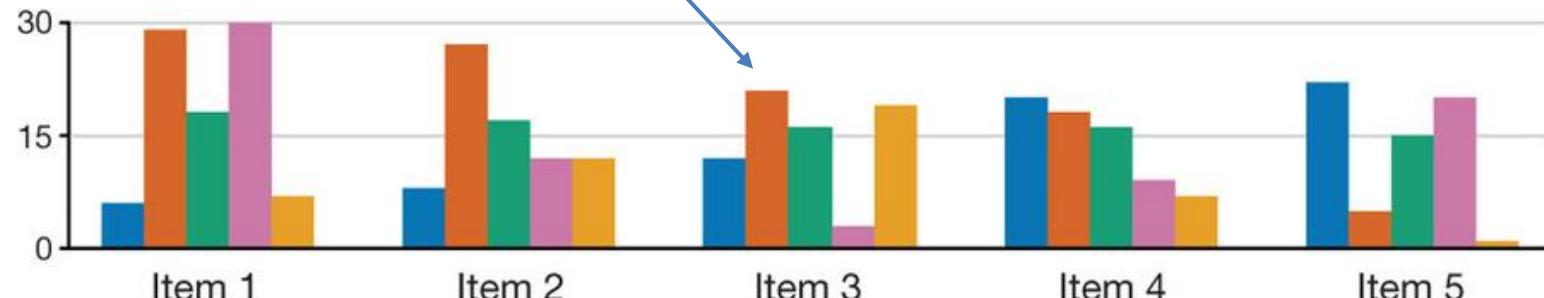
# Variations of pie & bar charts

- Category 1
- Category 2
- Category 3
- Category 4
- Category 5

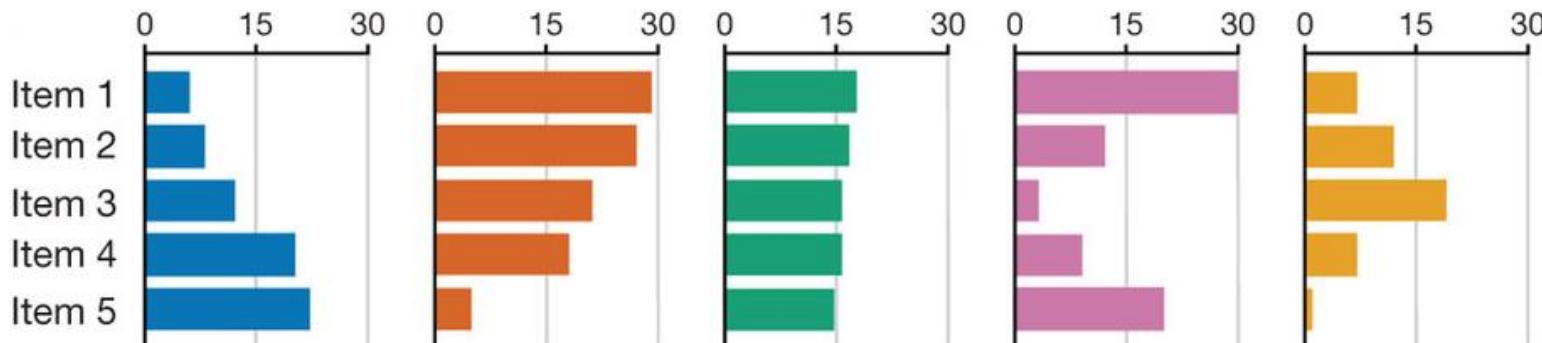


**Pie chart:** compare values in different categories

**Stacked bar charts:** compare overall values across items, but also show contribution per category



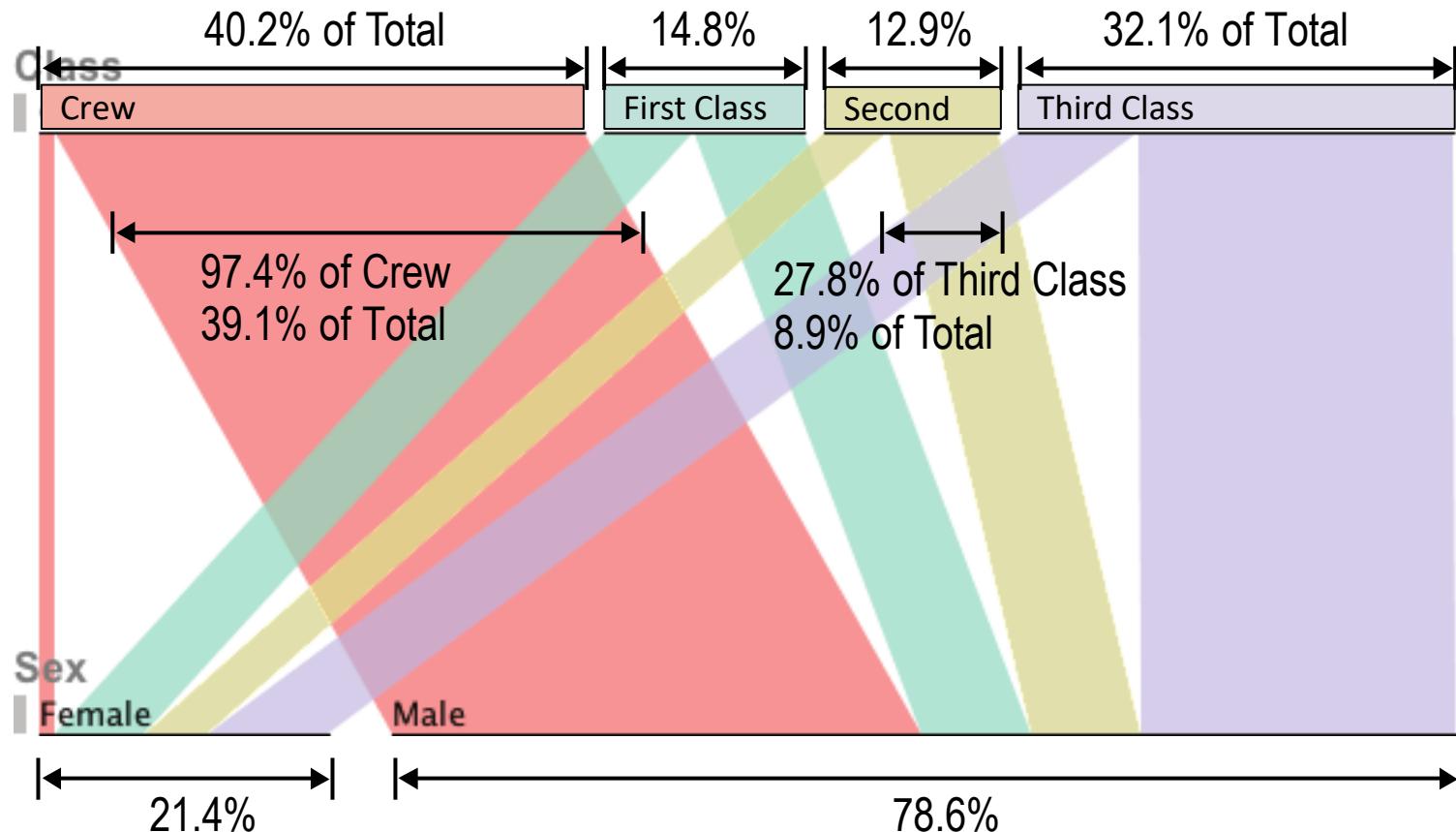
**Grouped bar charts:** compare values across categories within each item



**Layered bar charts:** compare values within categories

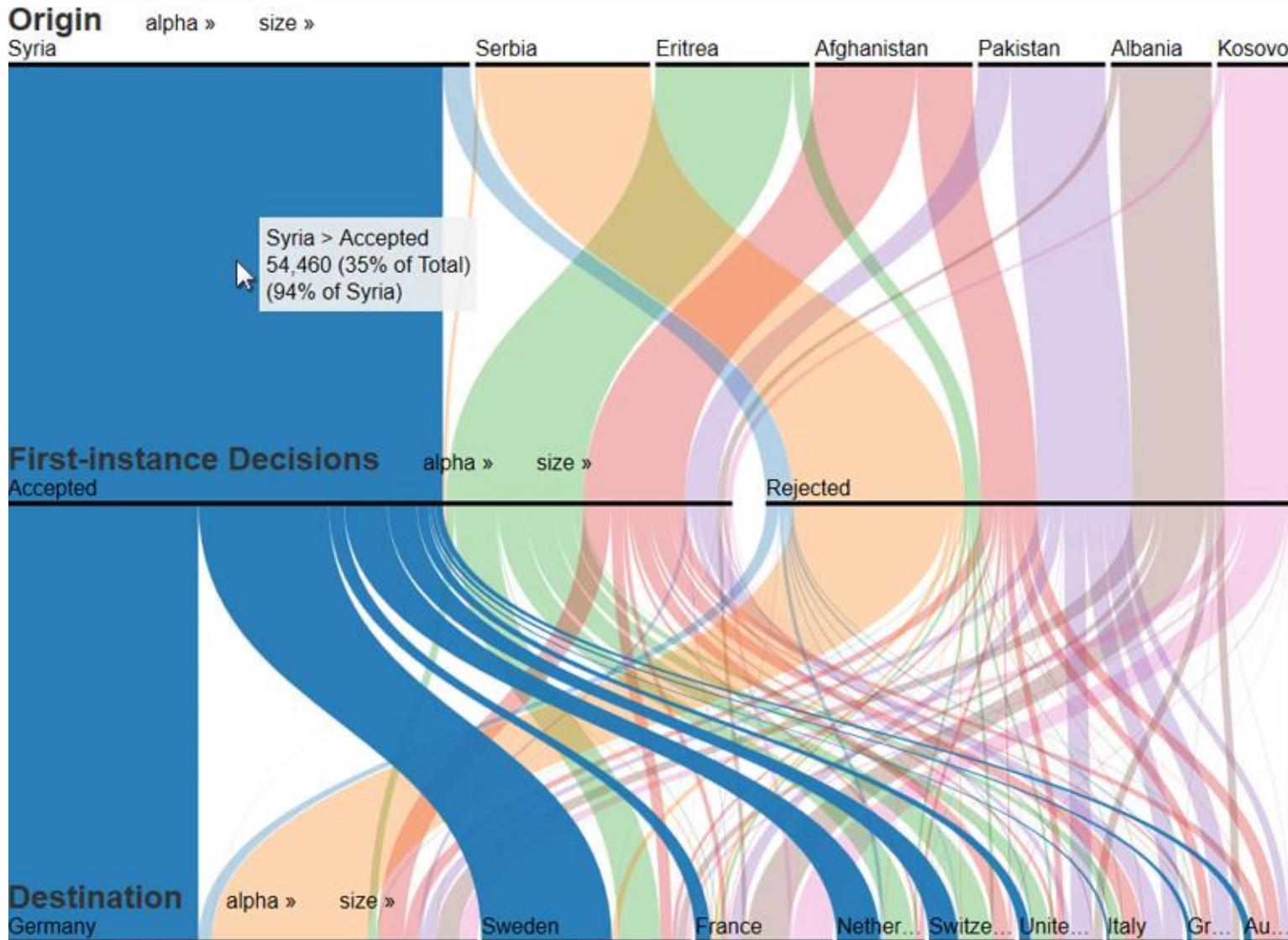
# Diagram techniques

- Parallel sets
  - Quantitative data wrt. multiple categorical attributes
  - Shows connections and proportions



# Diagram techniques

## European asylum decisions in 2014



Based on the infographics "[Seeking safety](#)," The Economist, 2015

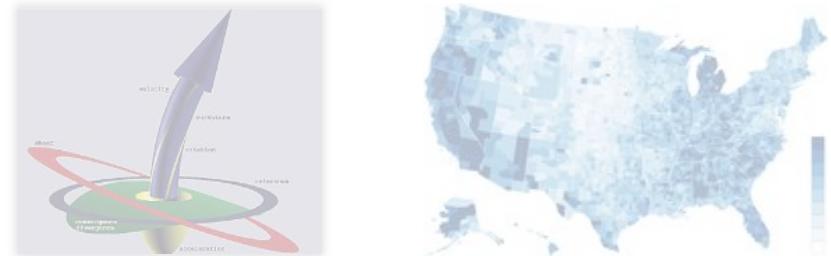
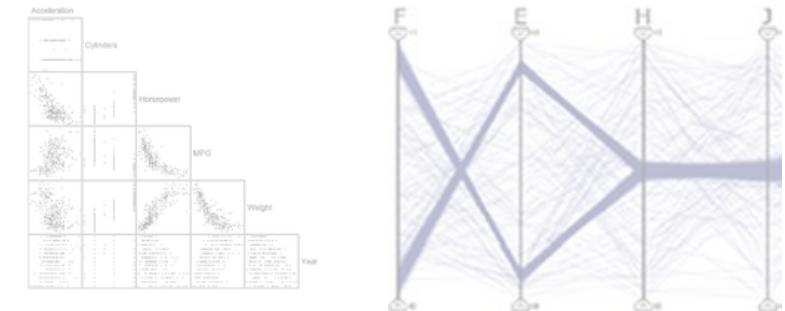
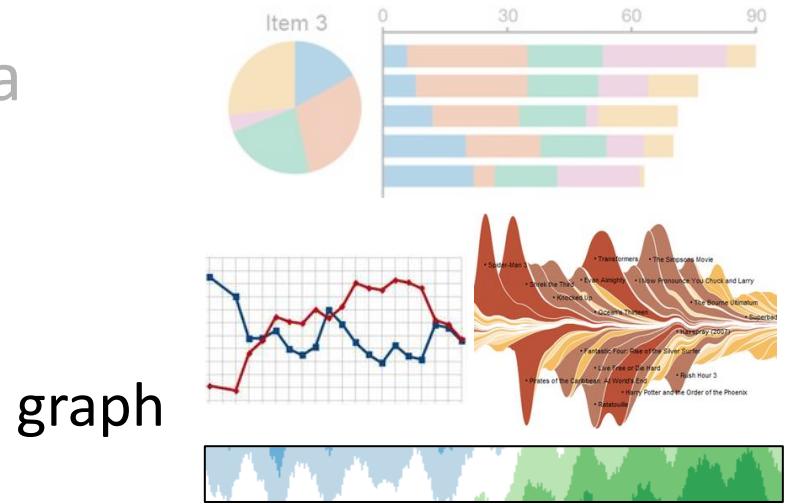
Data: [Eurostat](#)

\*Austria: data from 2013

[multivis.net/lecture/parallel-sets.htm](http://multivis.net/lecture/parallel-sets.htm)

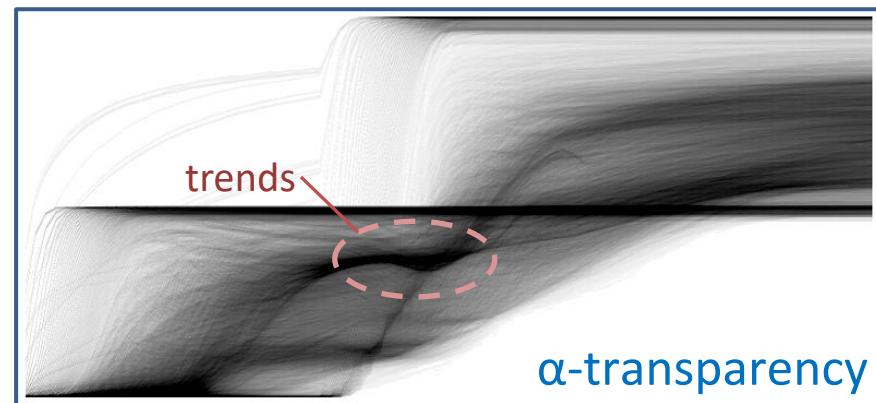
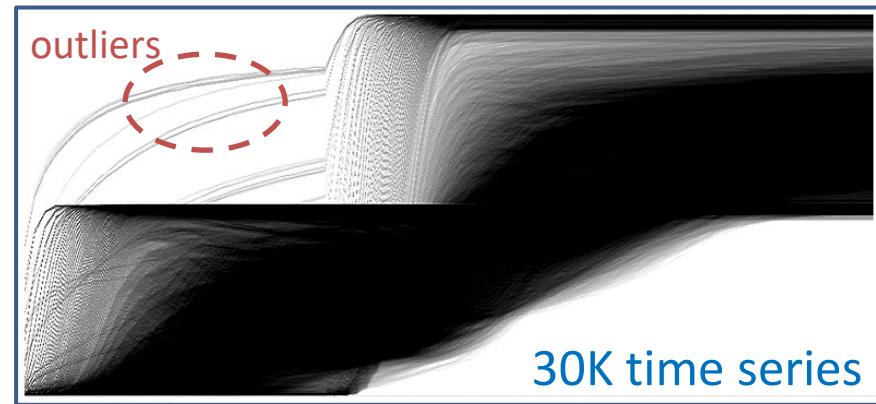
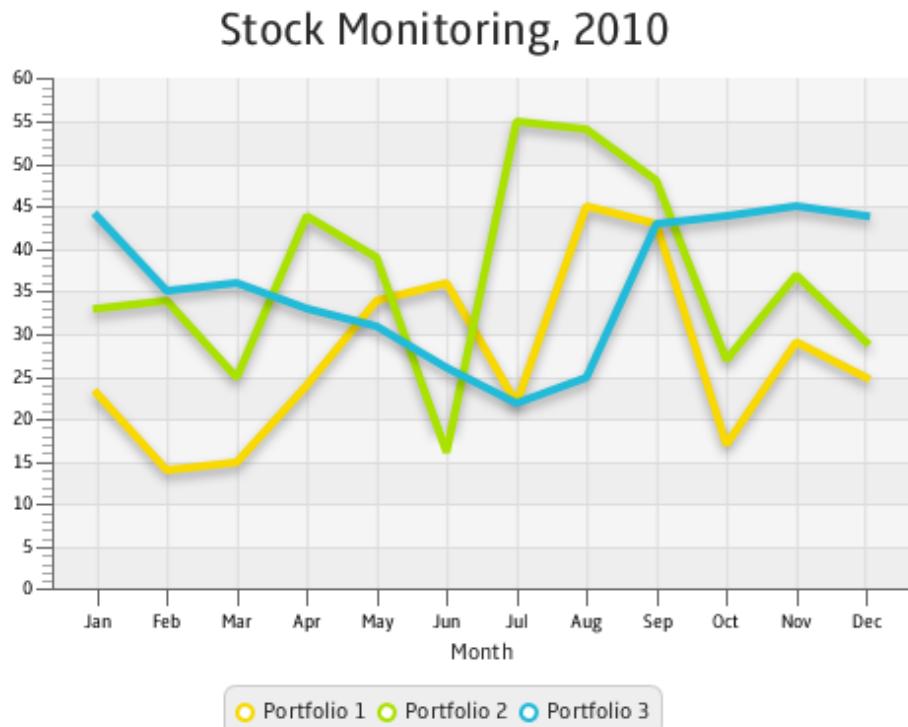
# Diagram techniques

- Categorical + quantitative data
  - Bar/pie chart, stacked bars
- Time-dependent data
  - Line graph, ThemeRiver, Horizon graph
- Single and multiple variables
  - Histogram, scatterplot, parallel coordinates
  - Glyphs, color mapping



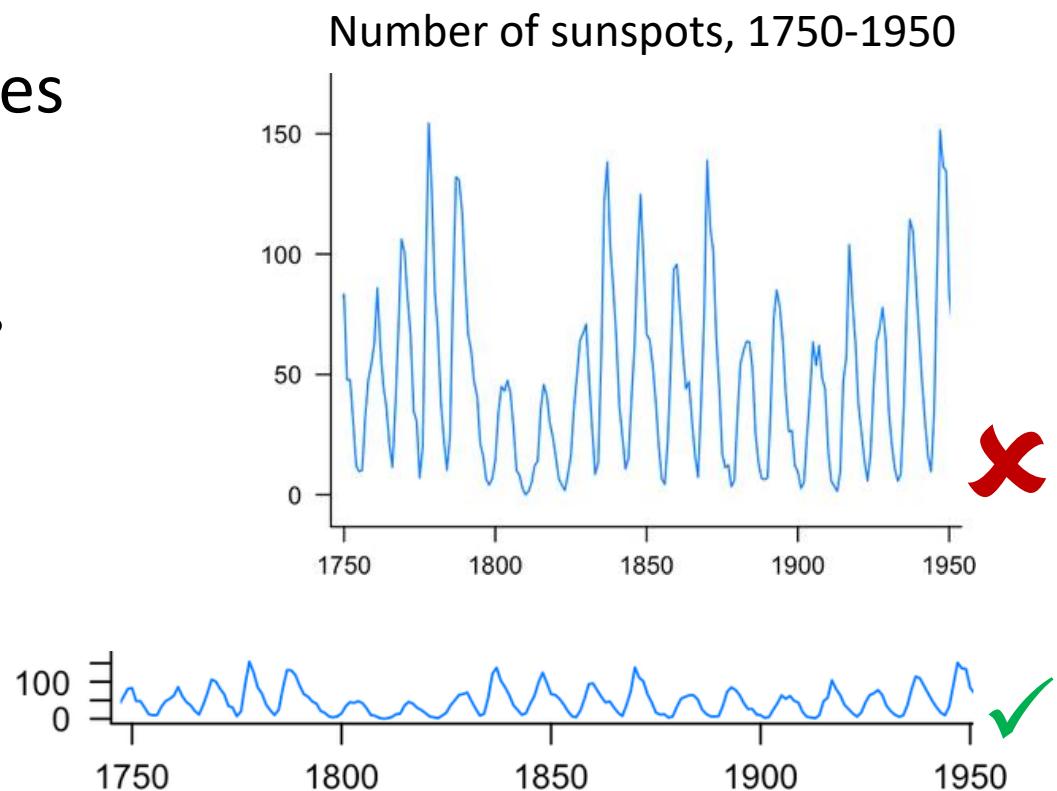
# Diagram techniques

- Line graph
  - Quantitative data on common scale(s) wrt. time
  - Connection between points – trends, structures, groups



# Diagram techniques

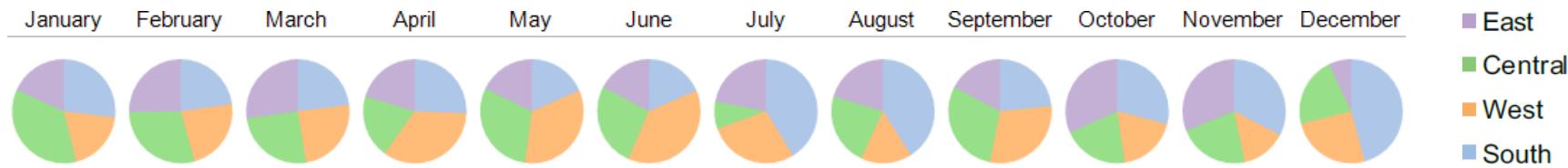
- Line graph
  - Banking to 45 degrees
    - Perceptual principle: most accurate angle judgment around 45°
    - Pick aspect ratio (height/width) accordingly



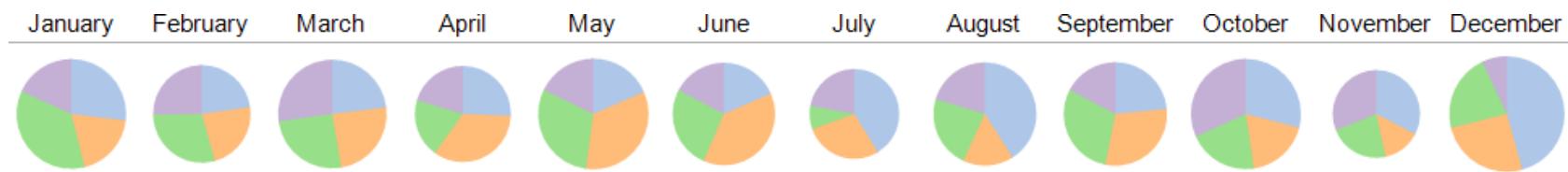
# Example: Analysis of Sales Data

Region	January	Februa..	March	April	May	June	July	August	Sept..	October	Novem..	Decem..
East	64,162	70,172	93,657	55,056	63,631	53,040	51,974	63,272	54,110	112,284	69,697	27,437
Central	125,392	80,851	87,645	52,996	107,159	80,349	19,907	70,431	92,568	72,961	49,881	85,084
West	67,364	63,742	83,856	92,412	120,284	116,618	66,692	49,671	92,920	65,971	31,516	99,000
South	94,572	63,234	79,491	68,963	65,868	56,659	97,101	126,879	73,240	102,589	73,044	177,943
Total (\$)	351,490	277,999	344,649	269,427	356,942	306,666	235,674	310,253	312,838	353,805	224,138	389,464

**Pie chart:** compare sales in different regions, but total sales are not shown



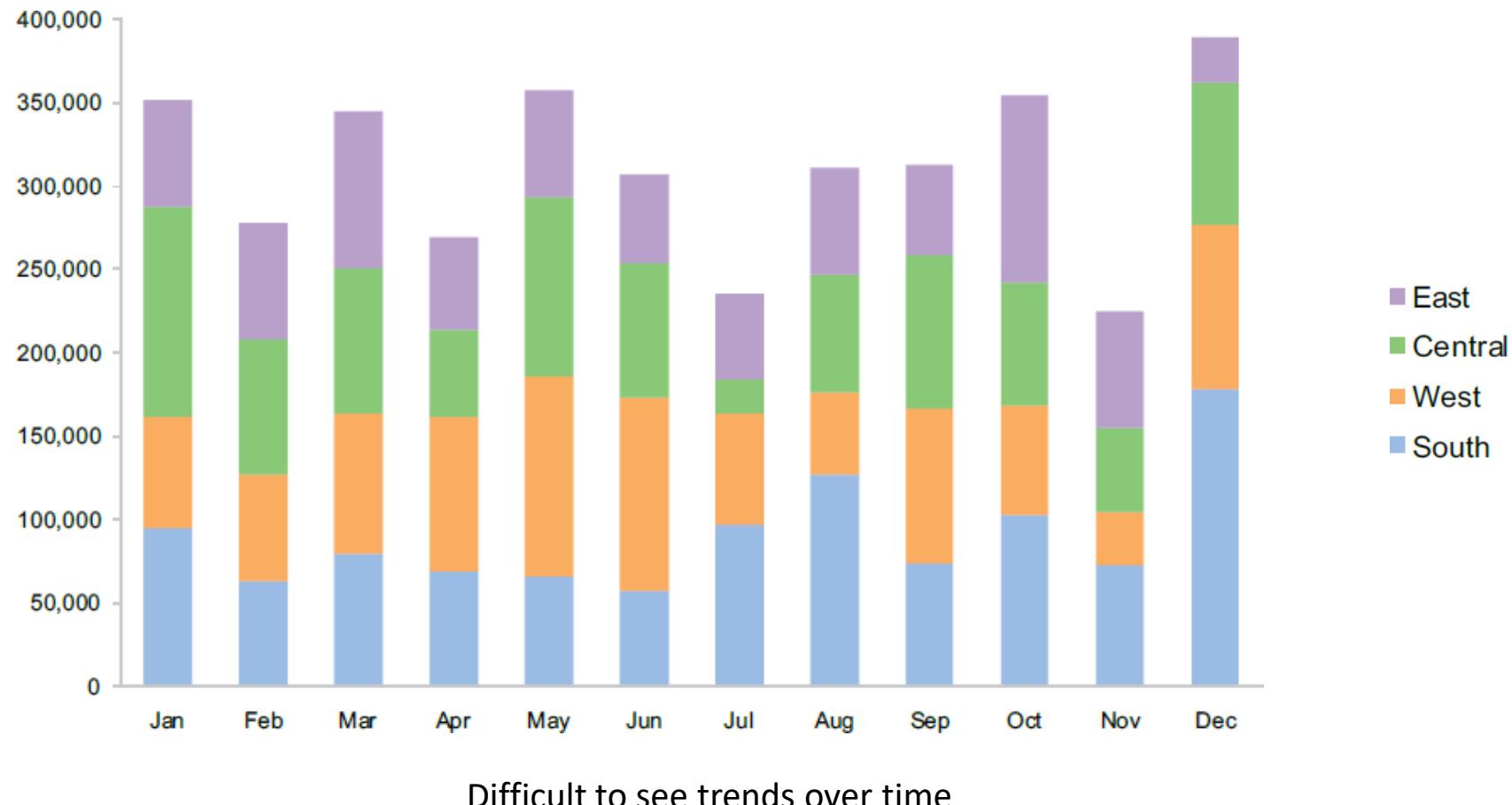
**Pie chart:** Total sales → Size



Hard to compare angles and/or sizes

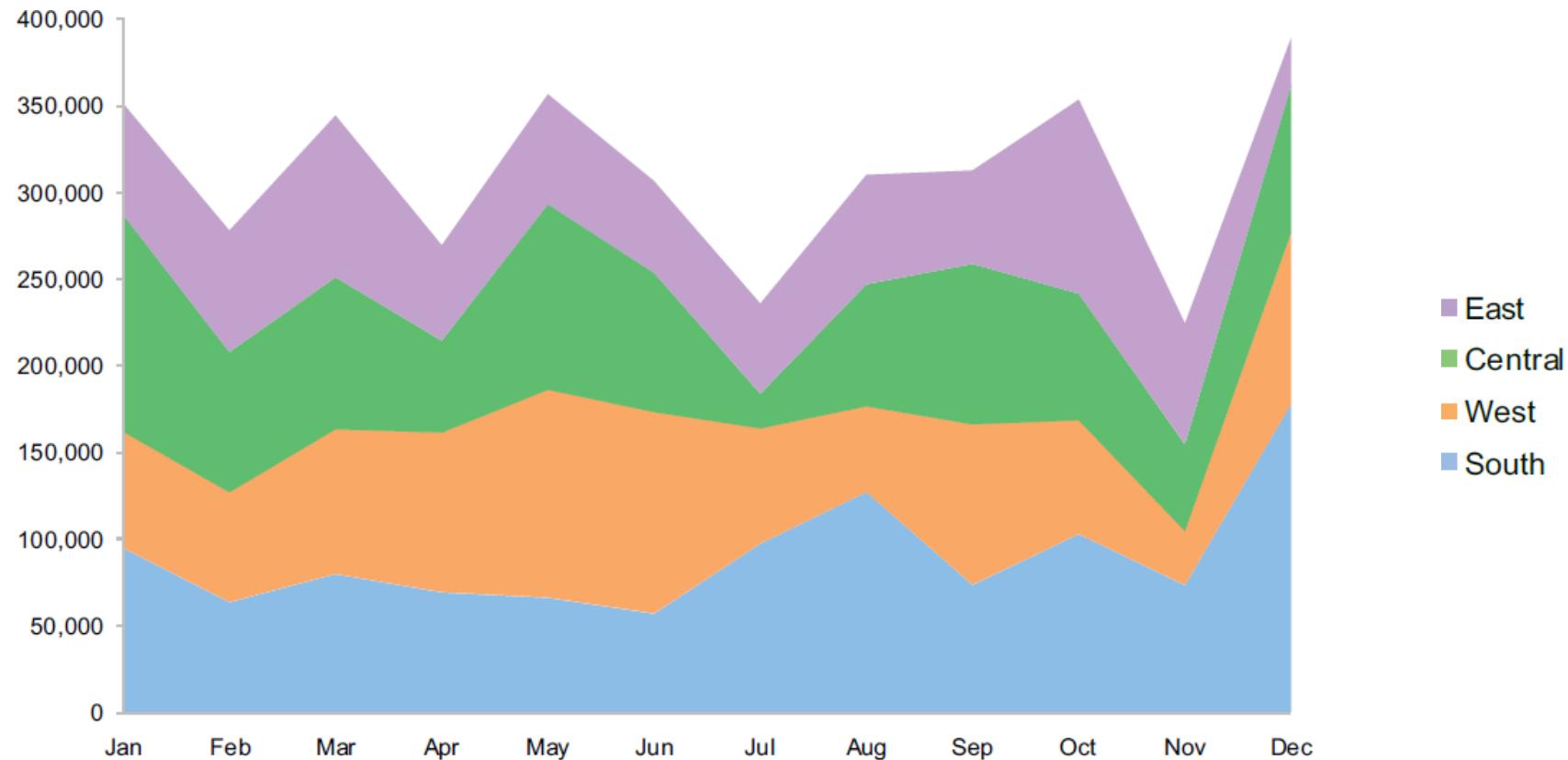
# Example: Analysis of Sales Data

Stacked bar chart



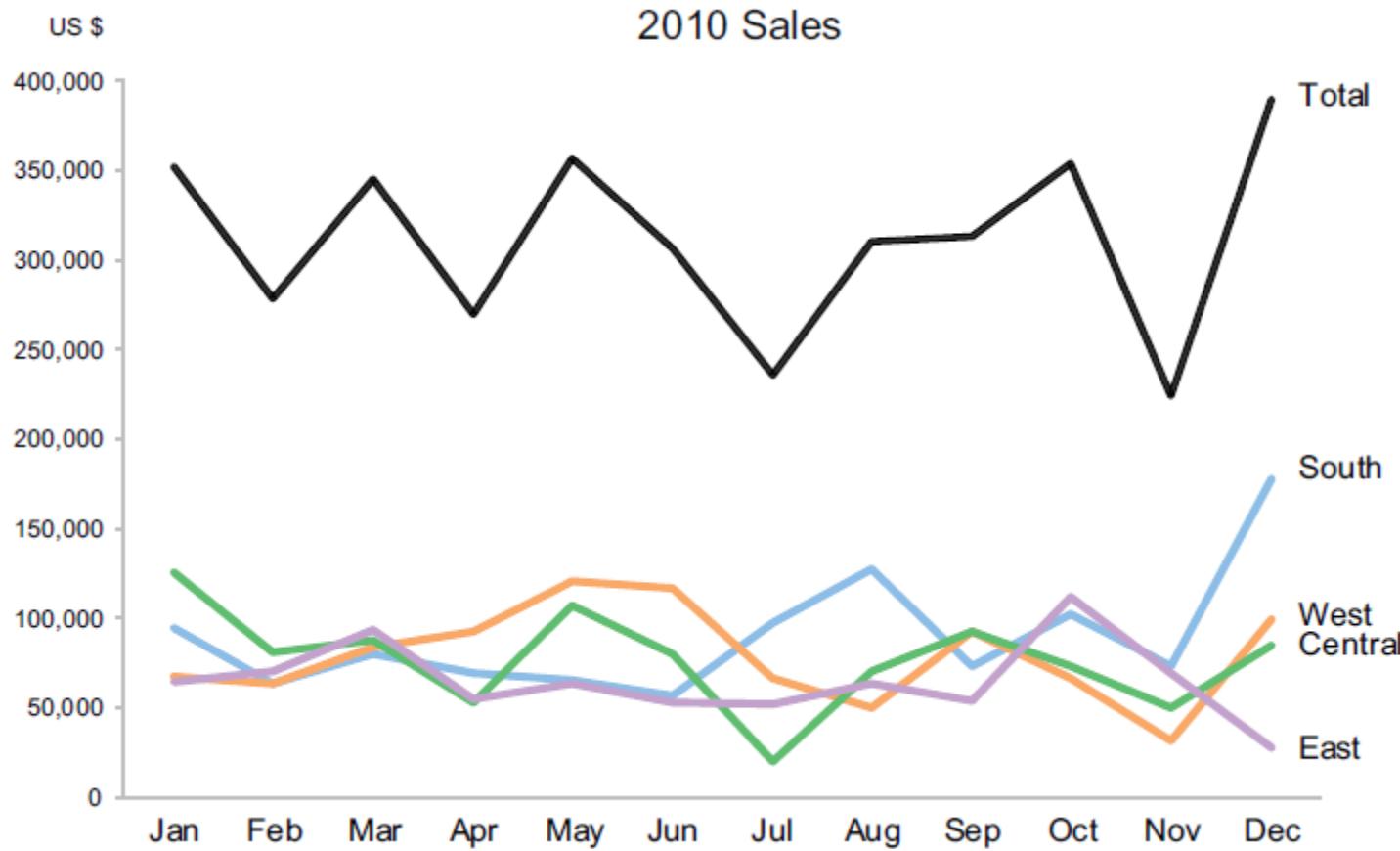
# Example: Analysis of Sales Data

Stacked area chart



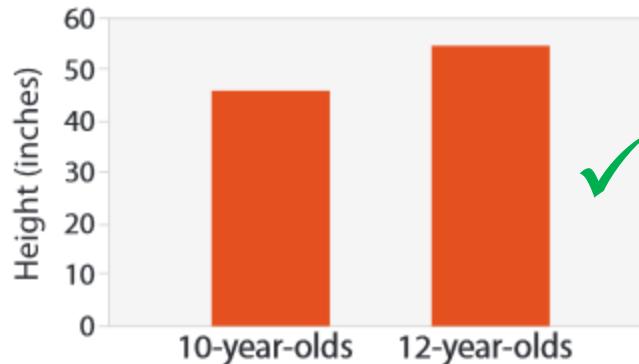
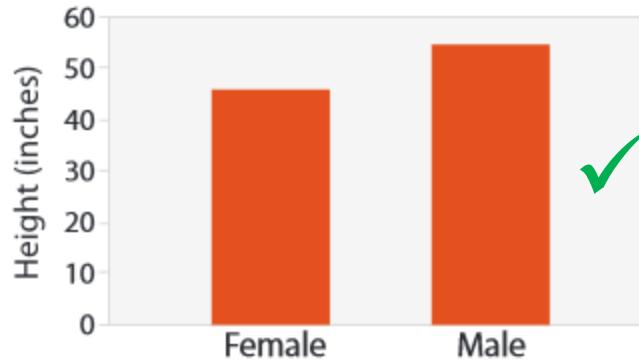
Exercise: Sketch development of orange curve

# Example: Analysis of Sales Data



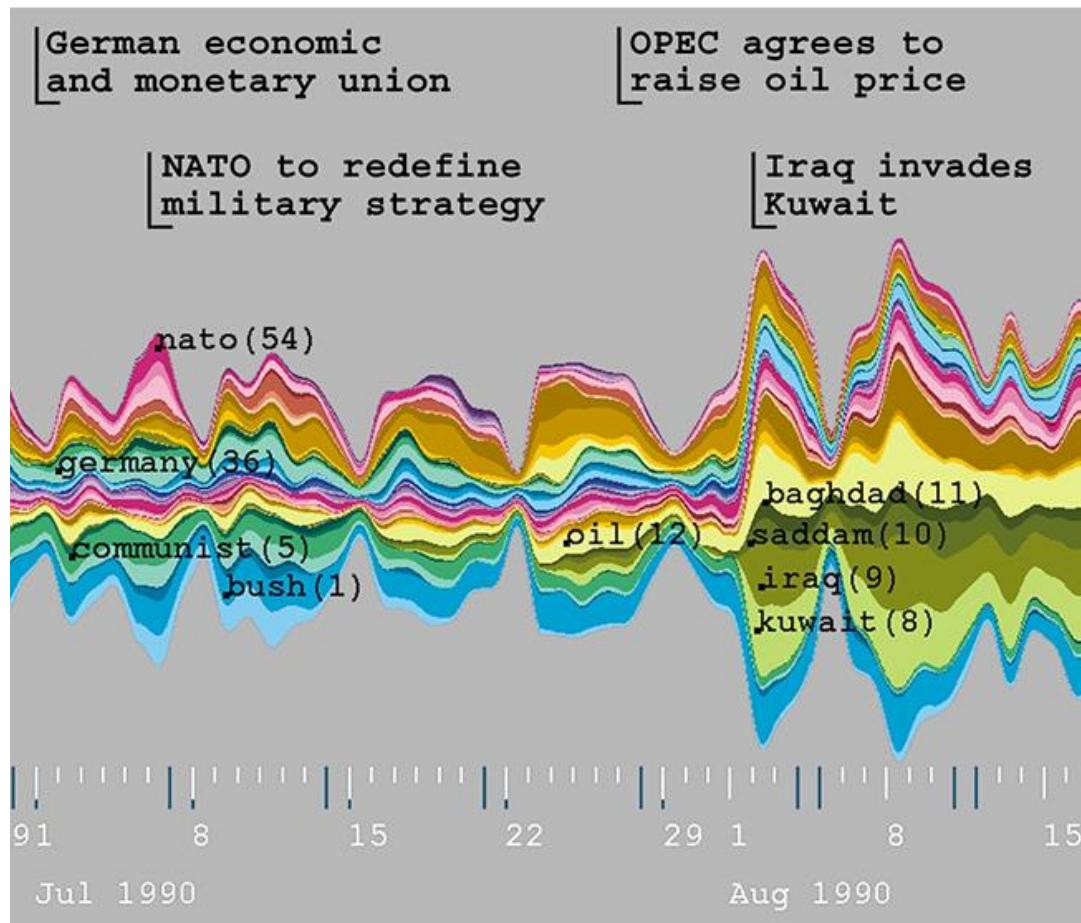
# Diagram techniques

- Bar vs. line chart
  - Bars support comparison
  - Lines imply trends



# Diagram techniques

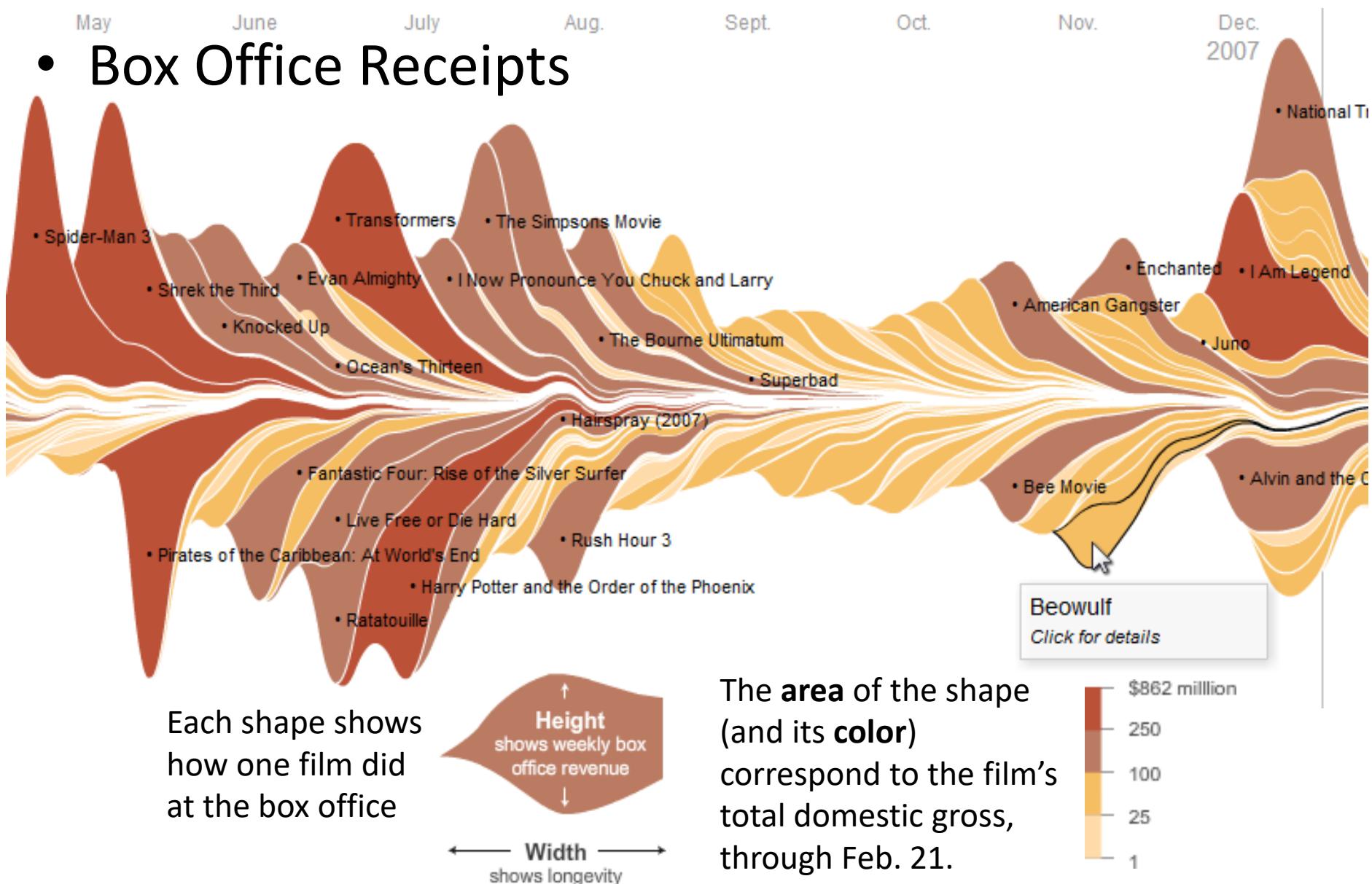
- ThemeRiver
  - Thematic changes in documents
  - Occurrence per topic/category mapped to width of river band
  - Less distorted around center
  - Rearranging bands



[Havre et al. 01]

# Diagram techniques

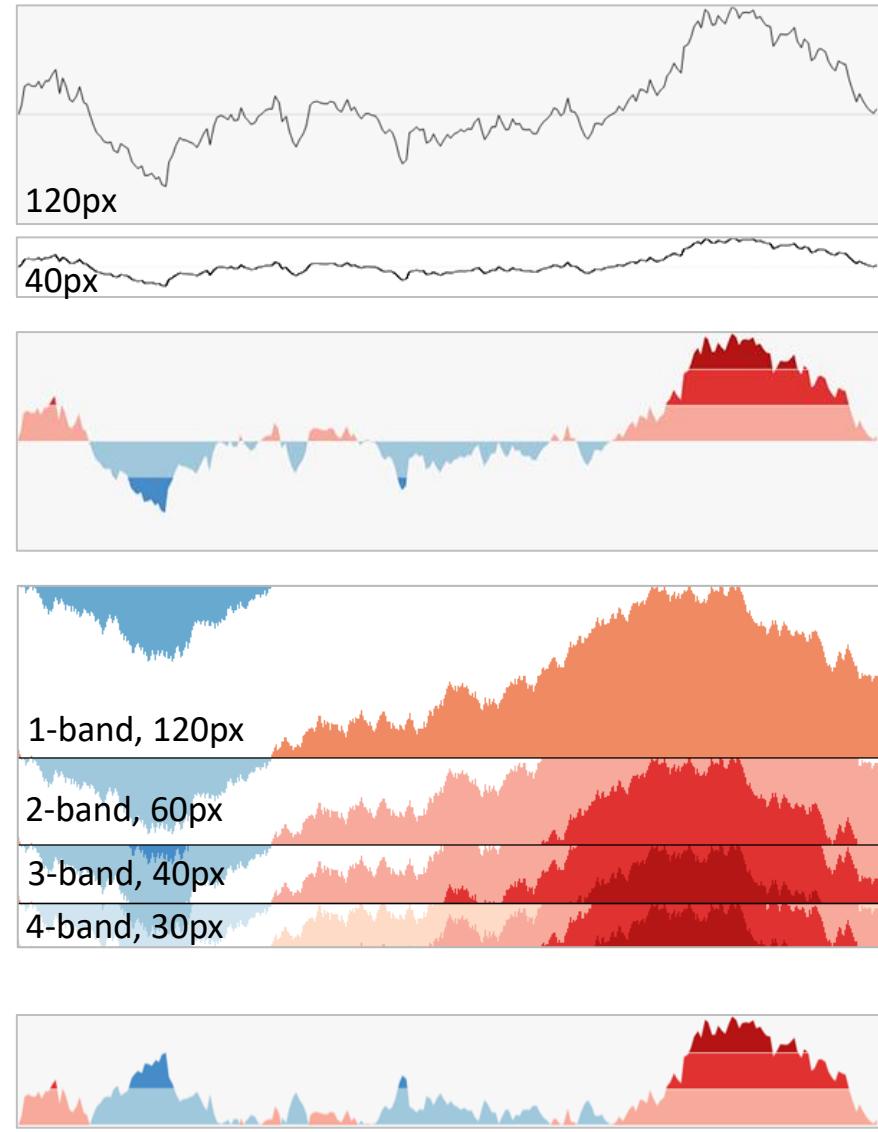
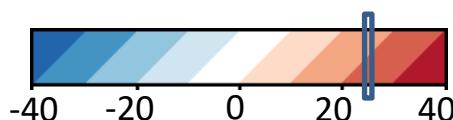
## • Box Office Receipts



Each shape shows how one film did at the box office

# Diagram techniques

- Horizon graph
  - Reduces vertical space without losing precision
  - Split vertically into layered bands
  - Collapse color bands to show values in less vertical space
  - Optional mirroring of negative values

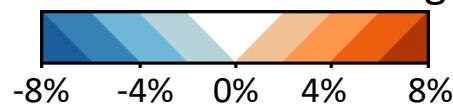


# Diagram techniques

## Unemployment rate (1982-2012)

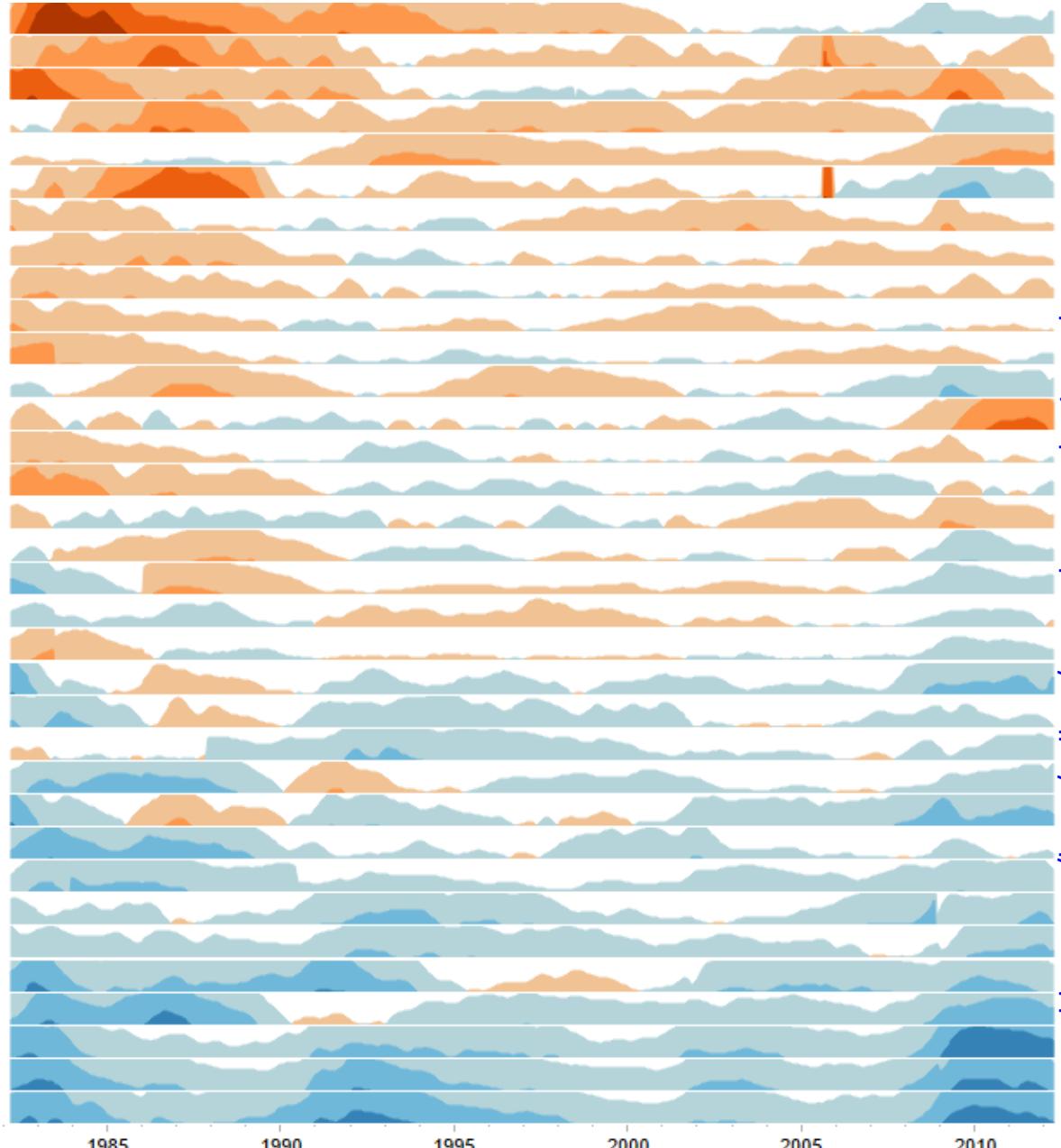
SIEMENS  
*Ingenuity for life*

Difference from average



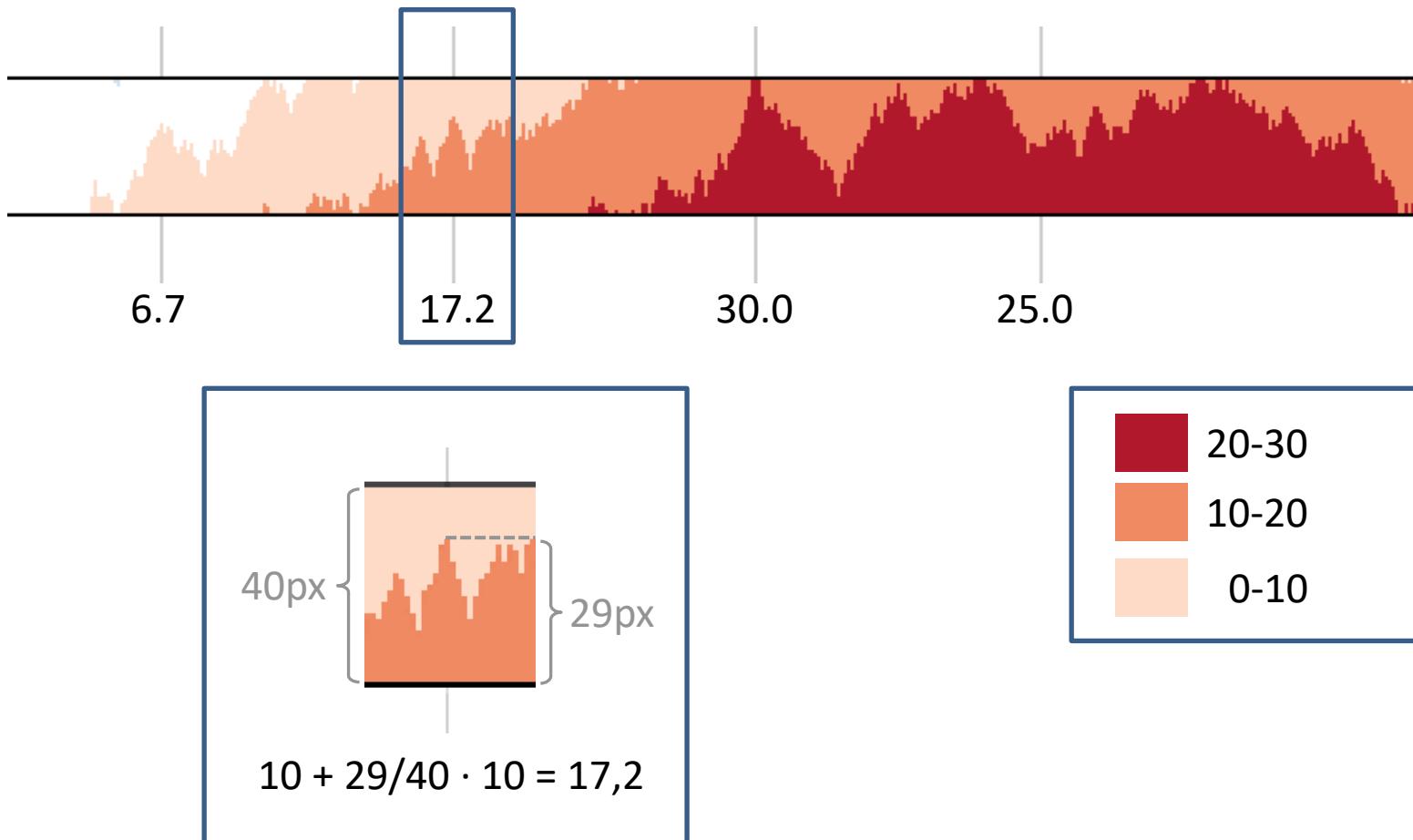
- 6% to 8%
- 4% to 6%
- 2% to 4%
- 0% to 2%
- -2% to 0%
- -4% to -2%
- -6% to -4%

West Virginia  
Mississippi  
Michigan  
Alaska  
California  
Louisiana  
Oregon  
Kentucky  
Illinois  
Washington  
Ohio  
New Mexico  
Nevada  
Tennessee  
Alabama  
South Carolina  
Arkansas  
Texas  
New York  
Pennsylvania  
Oklahoma  
Colorado  
Wisconsin  
Massachusetts  
Wyoming  
Connecticut  
Maryland  
Utah  
Minnesota  
Hawaii  
New Hampshire  
North Dakota  
South Dakota  
Nebraska

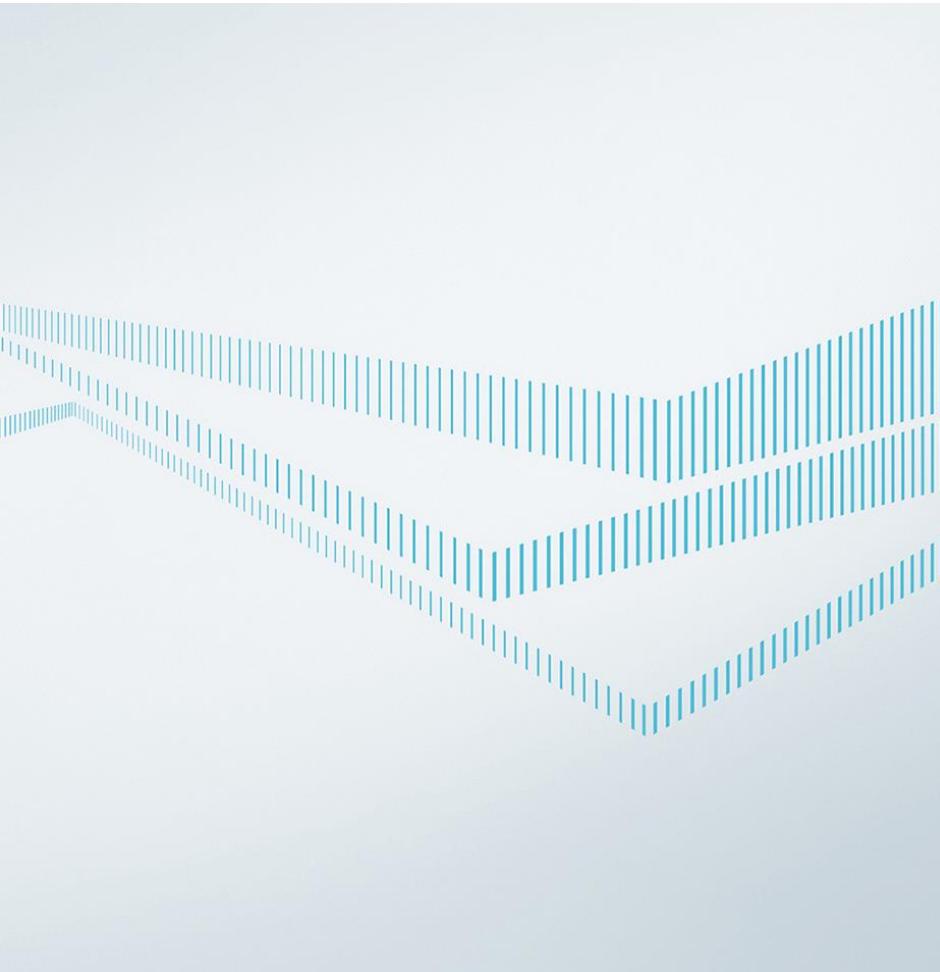


# Diagram techniques

- Example



# Contact information



**Dr. Johannes Kehrer**

Siemens Technology  
T RDA BAM IBI-DE  
Otto-Hahn-Ring 6  
81739 München, Deutschland

E-mail:  
[kehrer.johannes@siemens.com](mailto:kehrer.johannes@siemens.com)

Internet  
[siemens.com/innovation](http://siemens.com/innovation)

Intranet  
[intranet.ct.siemens.com](http://intranet.ct.siemens.com)

