上海科技大学
ShanghaiTech University

ARTS1422 Data Visualization

Lecture 4

Basics of Data Visualization (I)

Quan Li

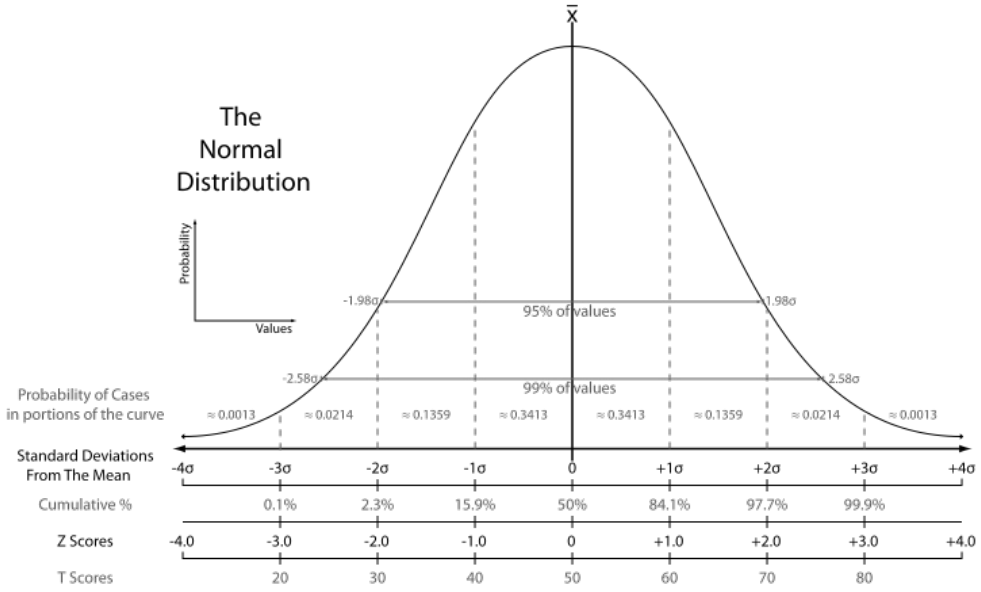
Spring 2024

2024. 03.07

Data Analysis

- Exploratory Data Analysis
- Data Mining

Something Basic: Statistics



- (Probably) Foundation of modern data analysis
- (Also) Foundation of machine learning, data mining, etc.

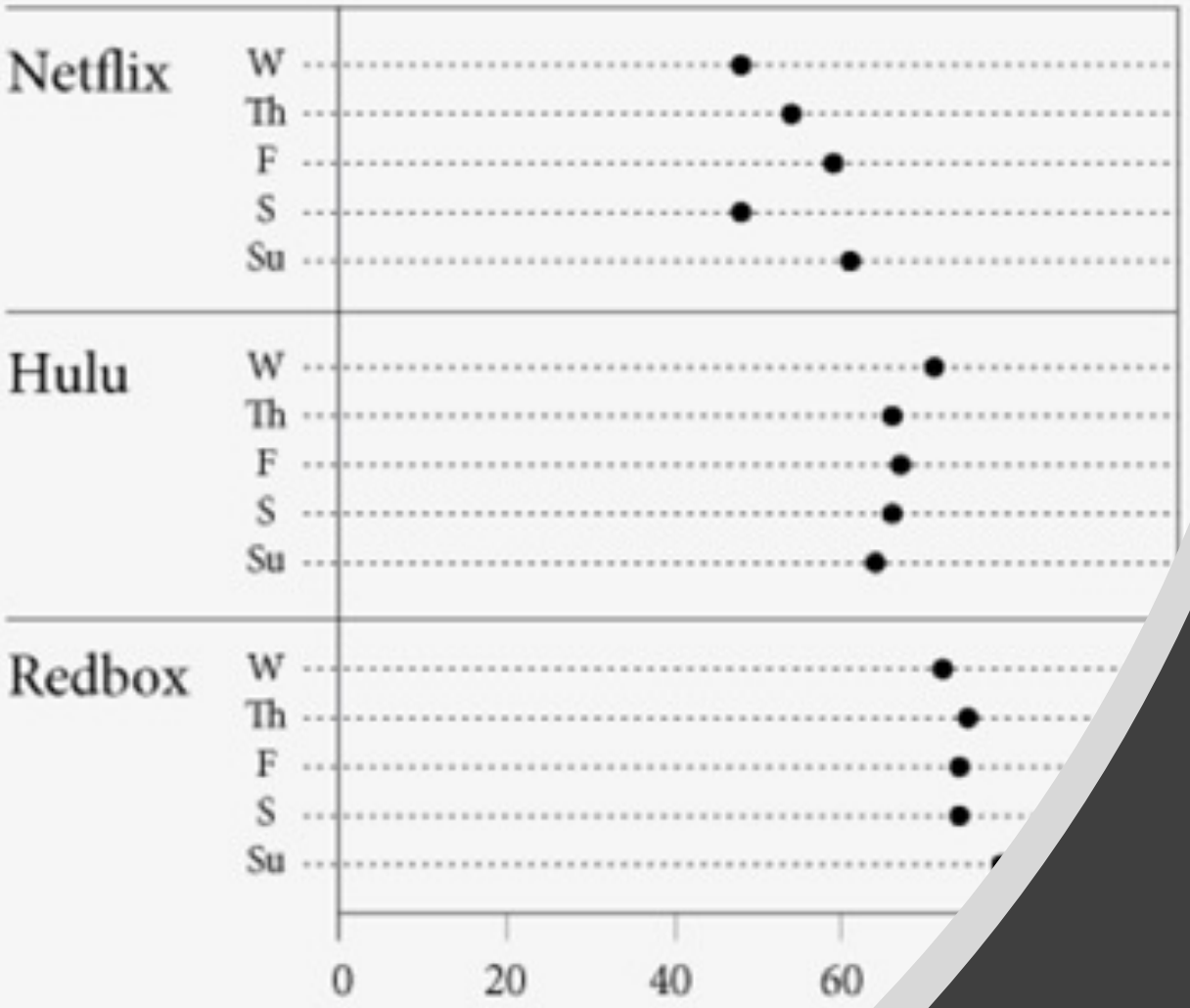
Exploratory Data Analysis (EDA)

- Based on statistics
- Data visualization-driven method
- Summary of main characteristics in easy-to-understand form

Visualization in EDA

- We will emphasize on visualization approaches in exploratory data analysis, but not statistical computing methods.

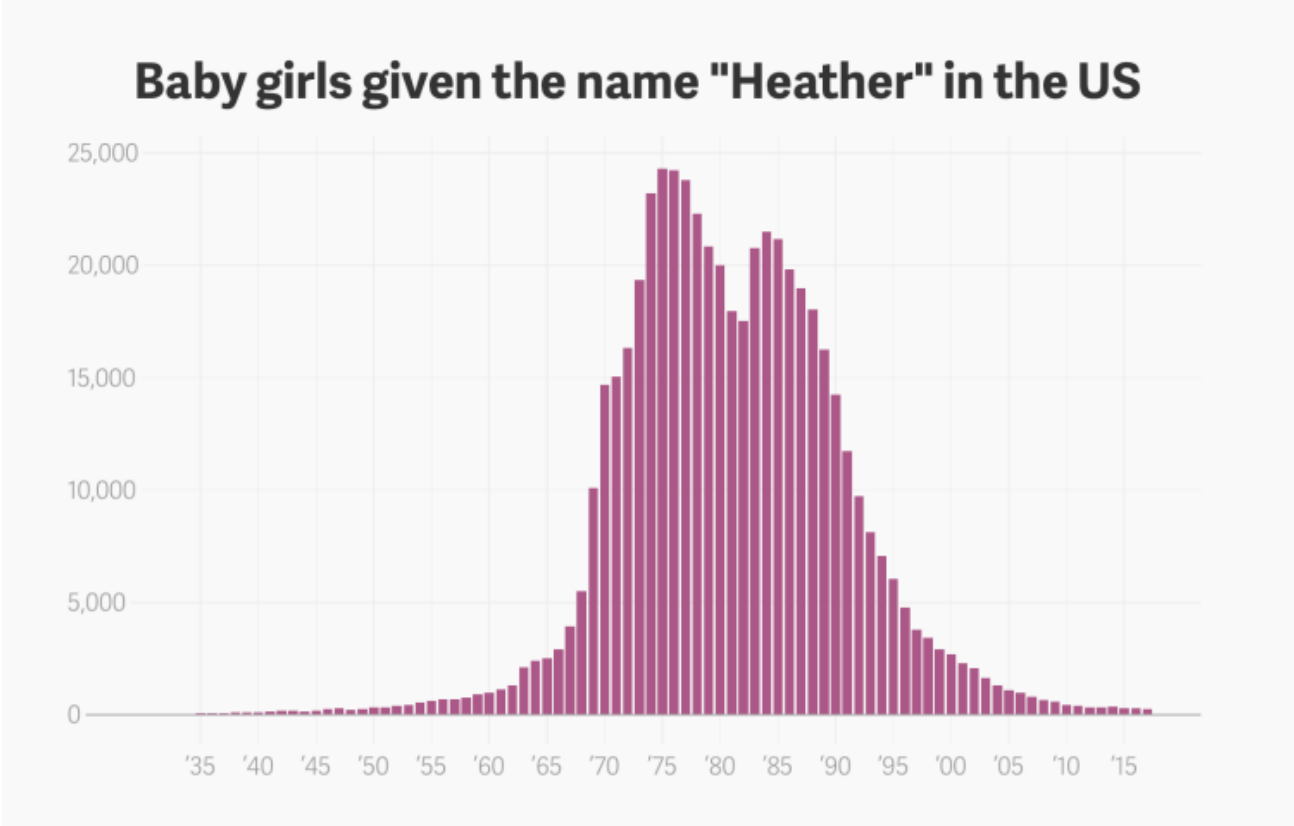
- Types of data visualization methods in EDA:
  - Plotting of raw data
  - Plotting of statistical values
  - Multiple coordinated views (Dashboard)



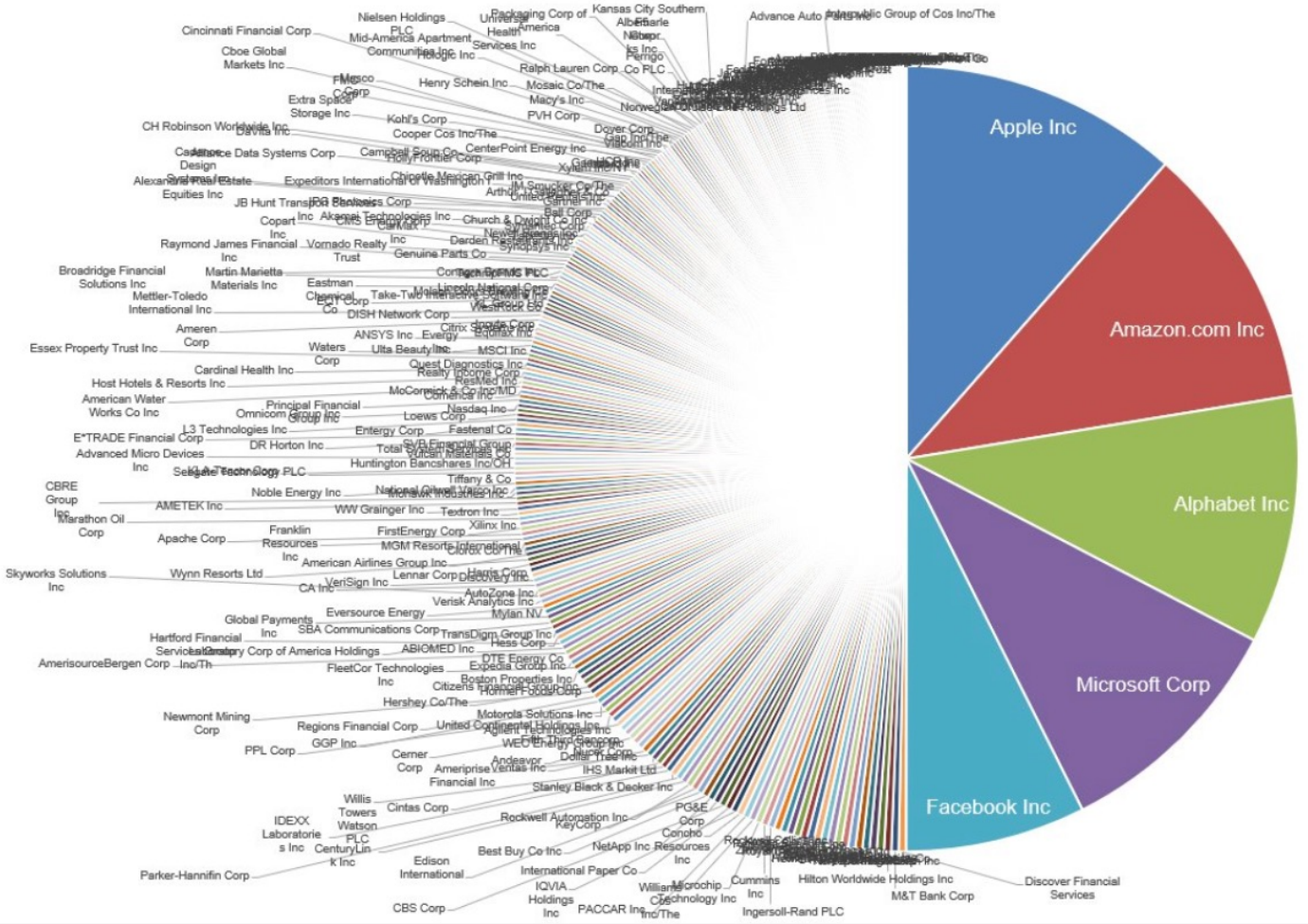
Original Data Graph: Data Trajectory

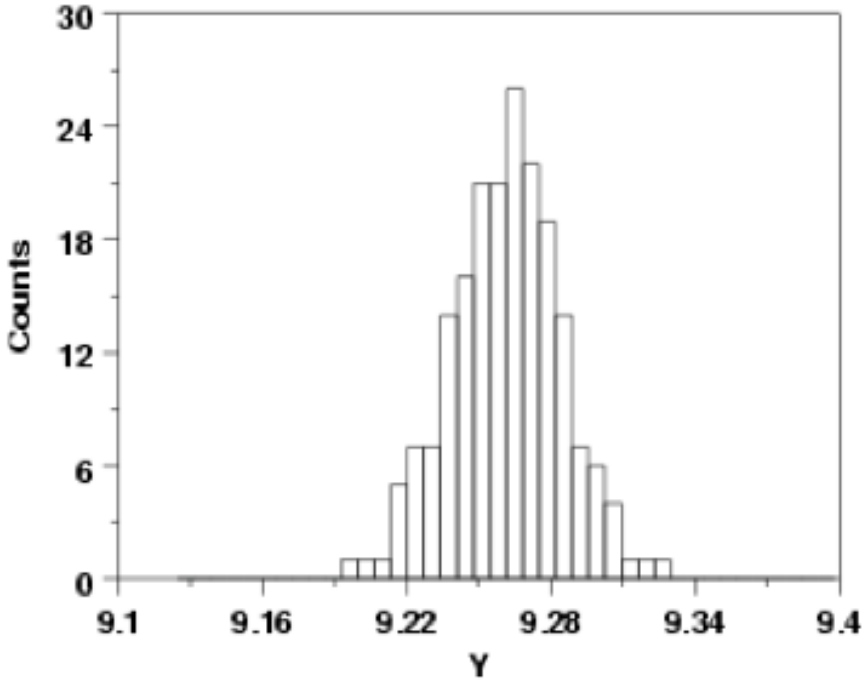
Plotting of raw data: Bar Chart

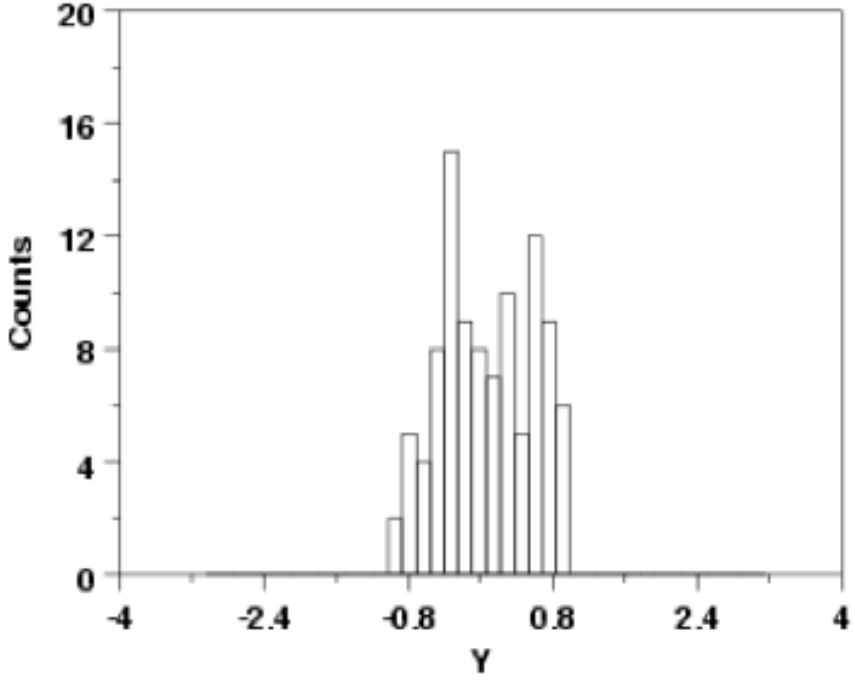
- https://flowingdata.com/2018/09/21/the-rise-and-sharp-plummet-of-the-name-heather/

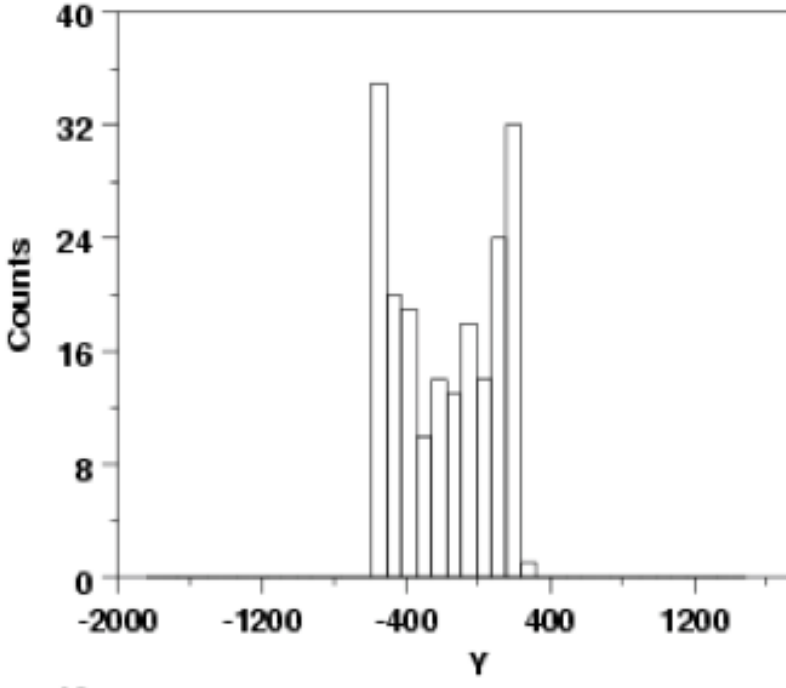


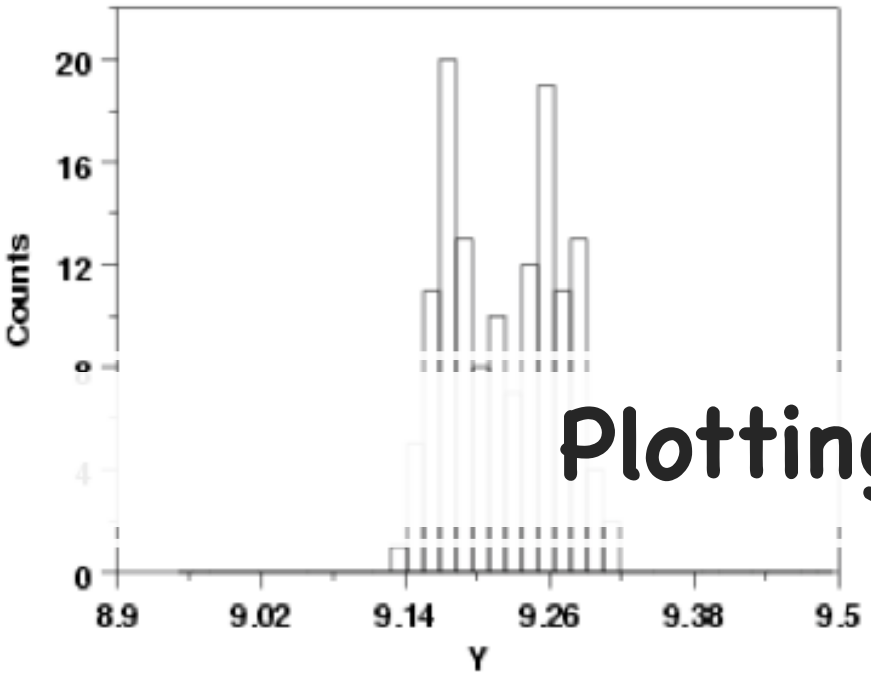
Plotting of raw data: Pie Chart

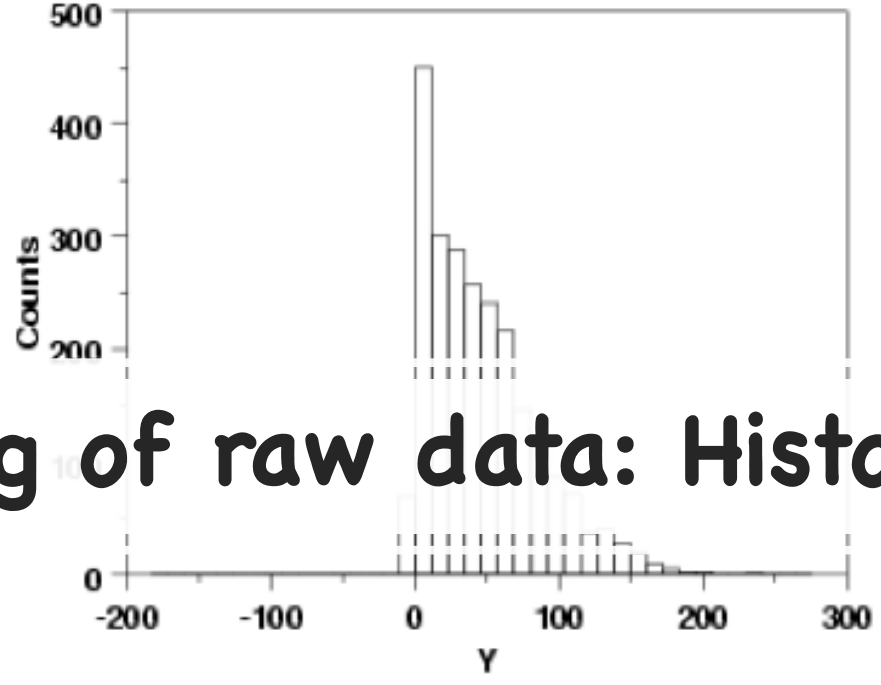


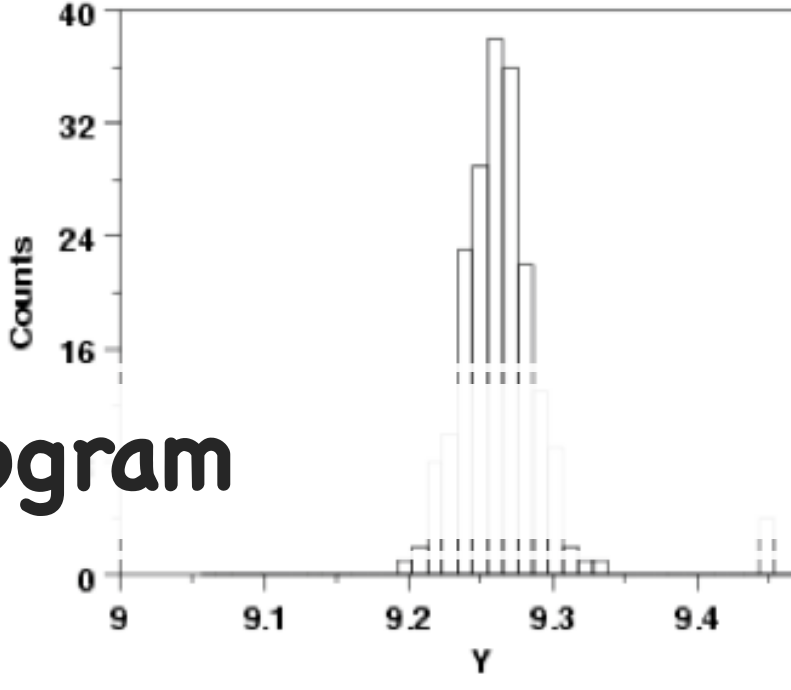




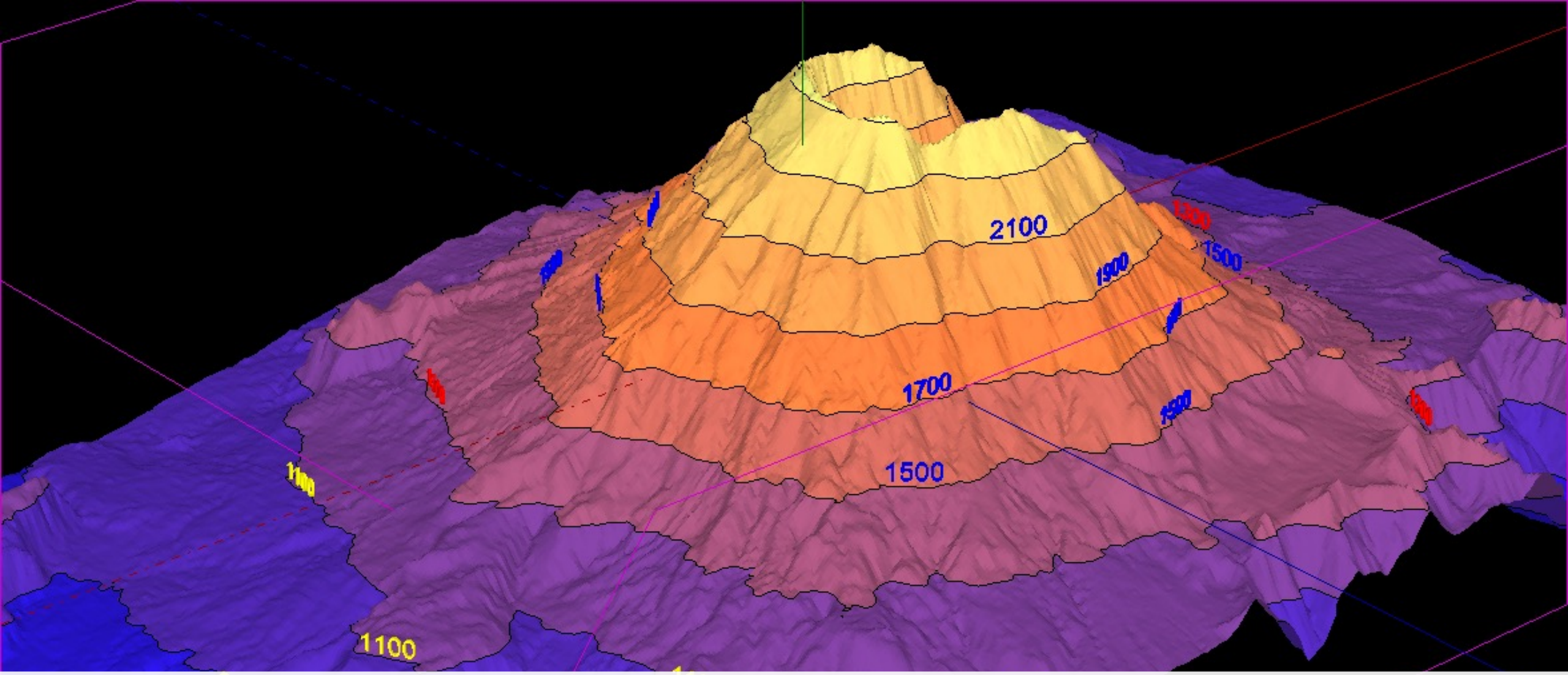








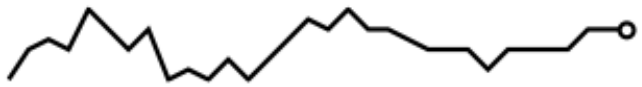
Plotting of raw data: Histogram

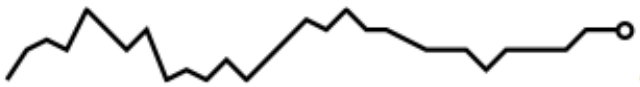


Plotting of raw data: Contour Map

Plotting of raw data: Sparkline

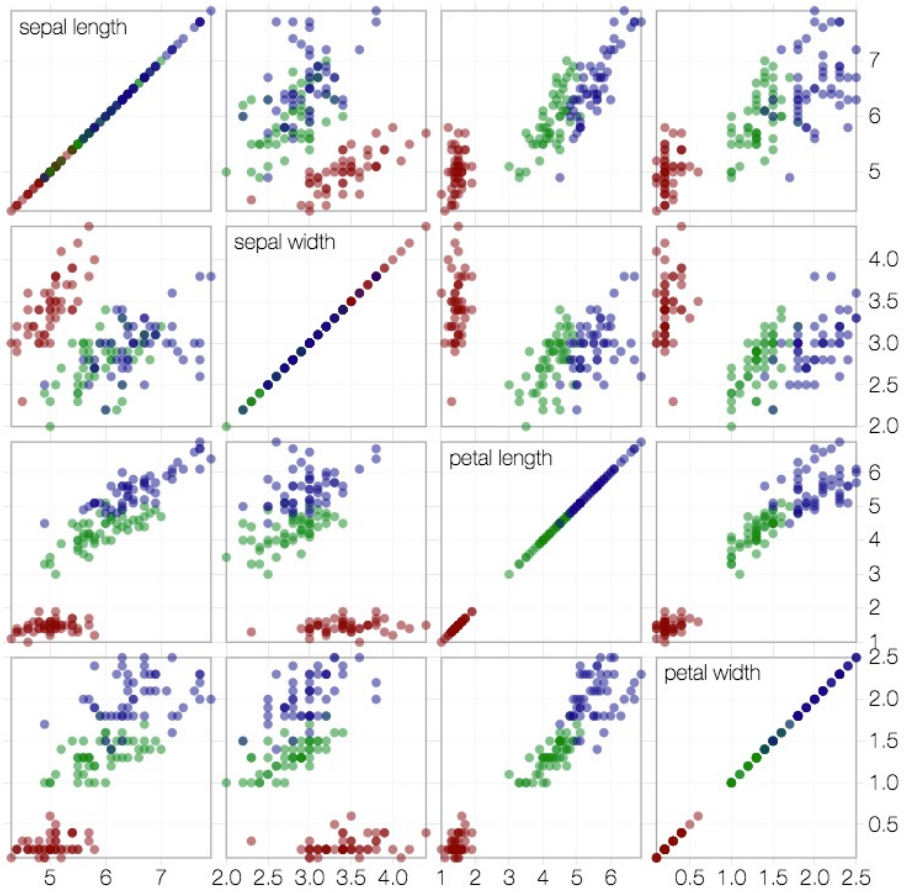
side of the flow of the text, i.e.:

 temperature **22.4° C**,

or be set inside it,  , to visualise evi-

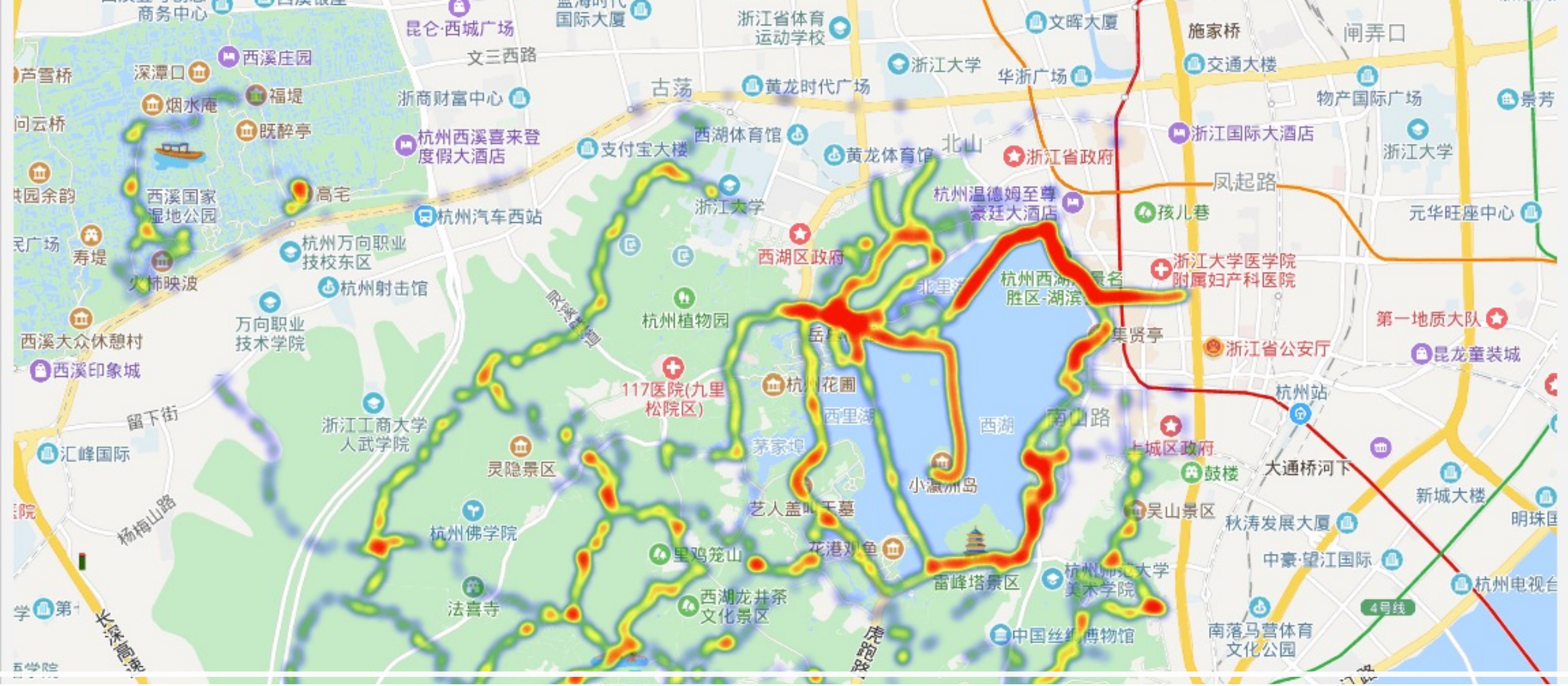
dence on the spot. In a stacked or a table form, such

Plotting of raw data: Scatter Plot/Scatter Plot Matrix



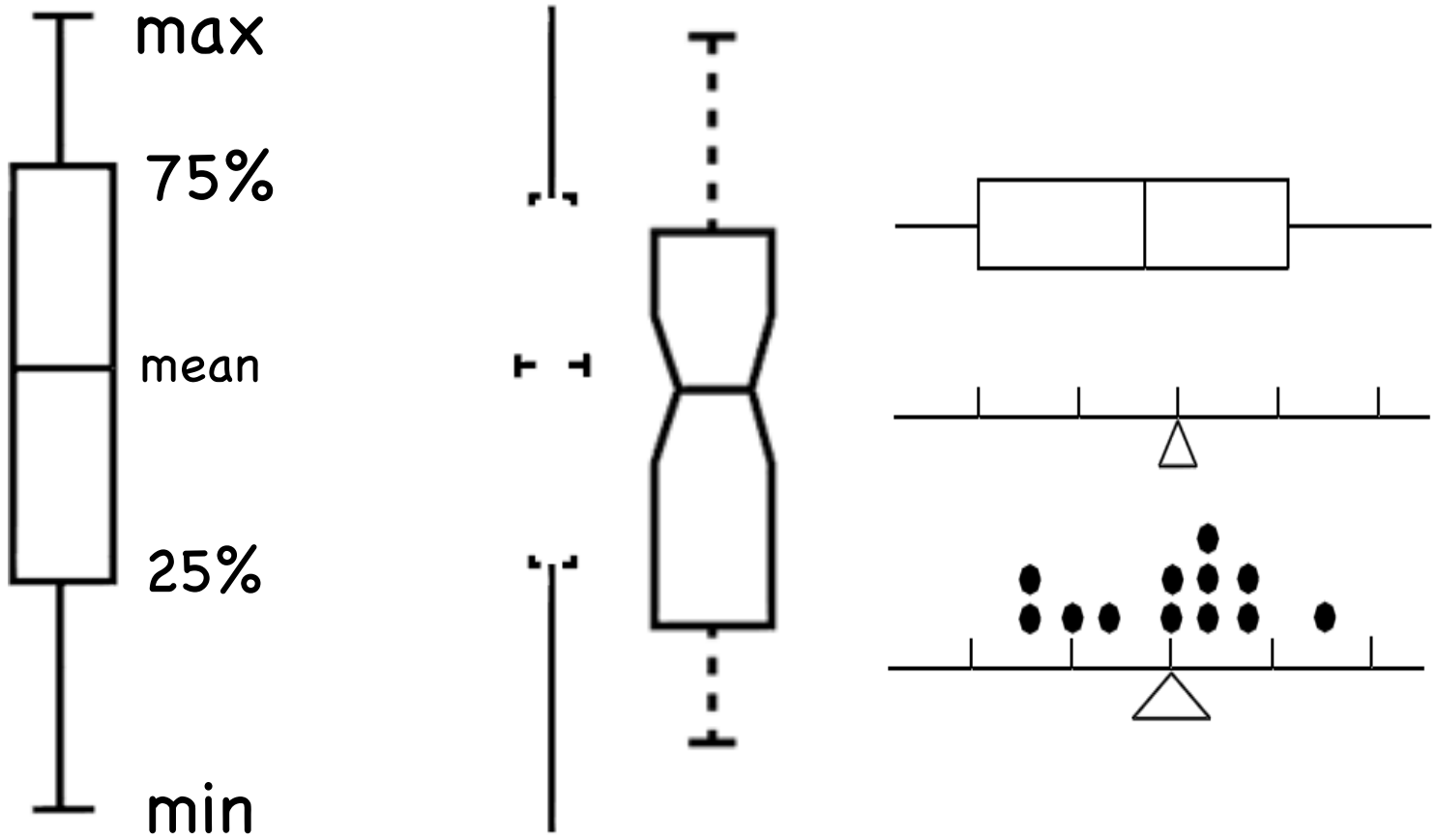
- *Iris setosa*
- *Iris versicolor*
- *Iris virginica*

Edgar Anderson's *Iris* data set
scatterplot matrix

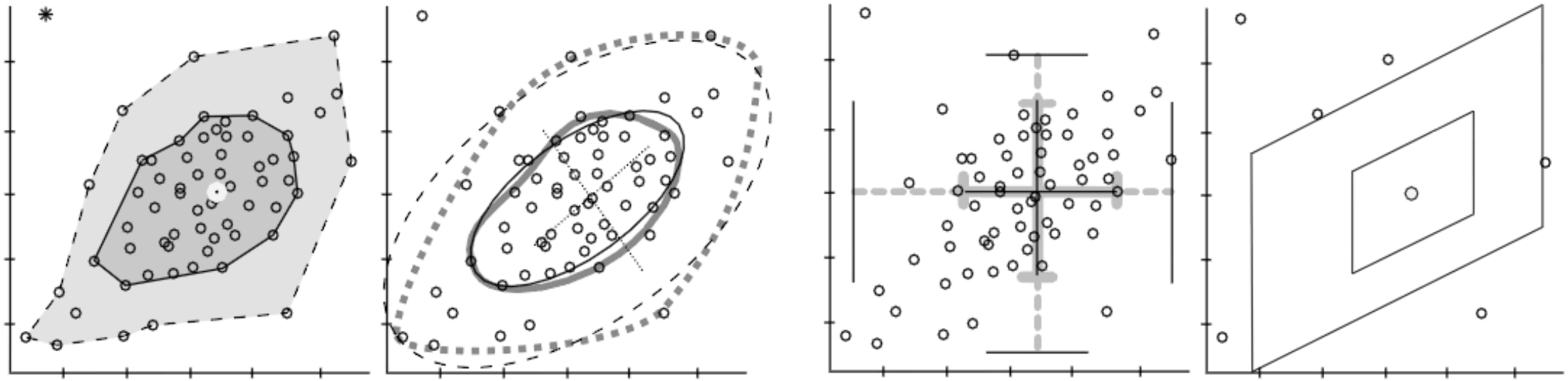


Plotting of raw data: Heatmap

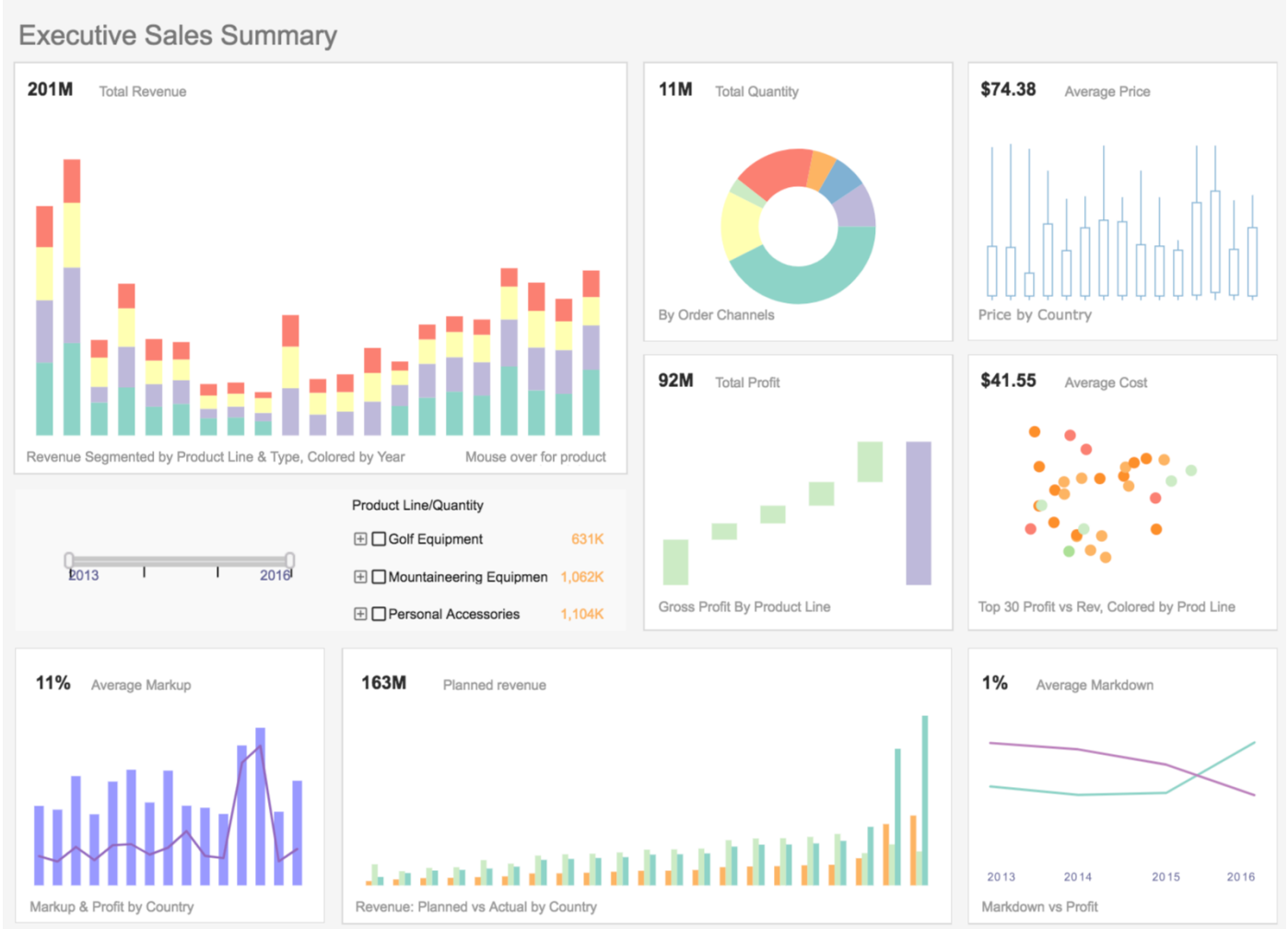
Plotting of statistical values: Box Plot (1-D)



Plotting of statistical values: Box Plot (2-D)



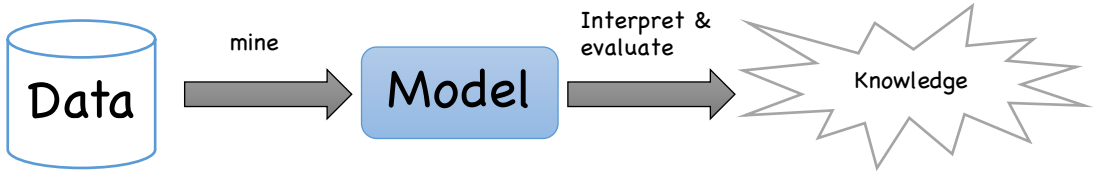
Multiple Coordinated Views (Dashboard)



Data Mining (DM)

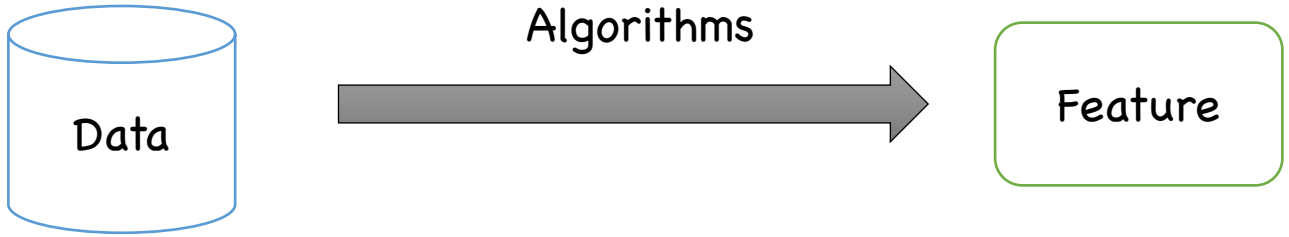
"Data Mining, also popularly referred to as knowledge discovery from data (KDD), is the automated or convenient extraction of patterns representing knowledge implicitly stored or captured in large databases, data warehouses, the Web, other massive repositories, or data streams."

-- H. Jiawei and M. Kamber, "Data Mining: Concepts and Techniques", 3rd ed., 2011.

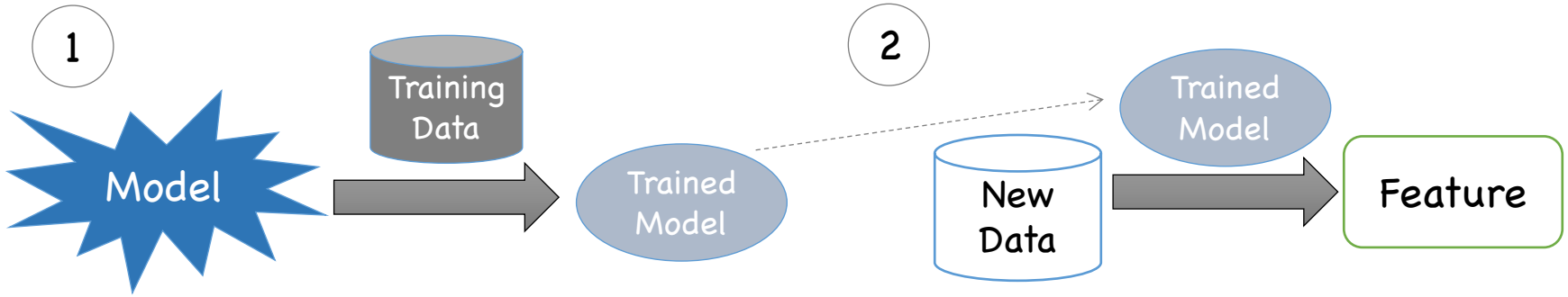


Tasks in DM

Descriptive Tasks



Predictive Tasks



Descriptive Tasks

Concept Description	Describe features of data directly
Association Analysis	Analyze "feature-value" pairs that occur frequently in data
Clustering	Group data on the principle of maximizing the intra-class similarity and minimizing the inter-class similarity
Outlier Mining	Analyze objects that do not comply with the general behavior or model of the data

Predictive Tasks

Classification

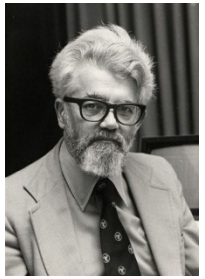
Find a model/function that describes and distinguish data classes or concepts based on analysis of a set of training data

Evolution Analysis

Analyze temporal and spatial patterns in dataset, model these patterns and predict data in unknown spatio-temporal positions

Approaches in DM

Statistics
(regression; parameter estimation)



Machine Learning
(decision tree; neural network; pattern recognition)

Statistical Learning
(statistical model; Bayesian network)

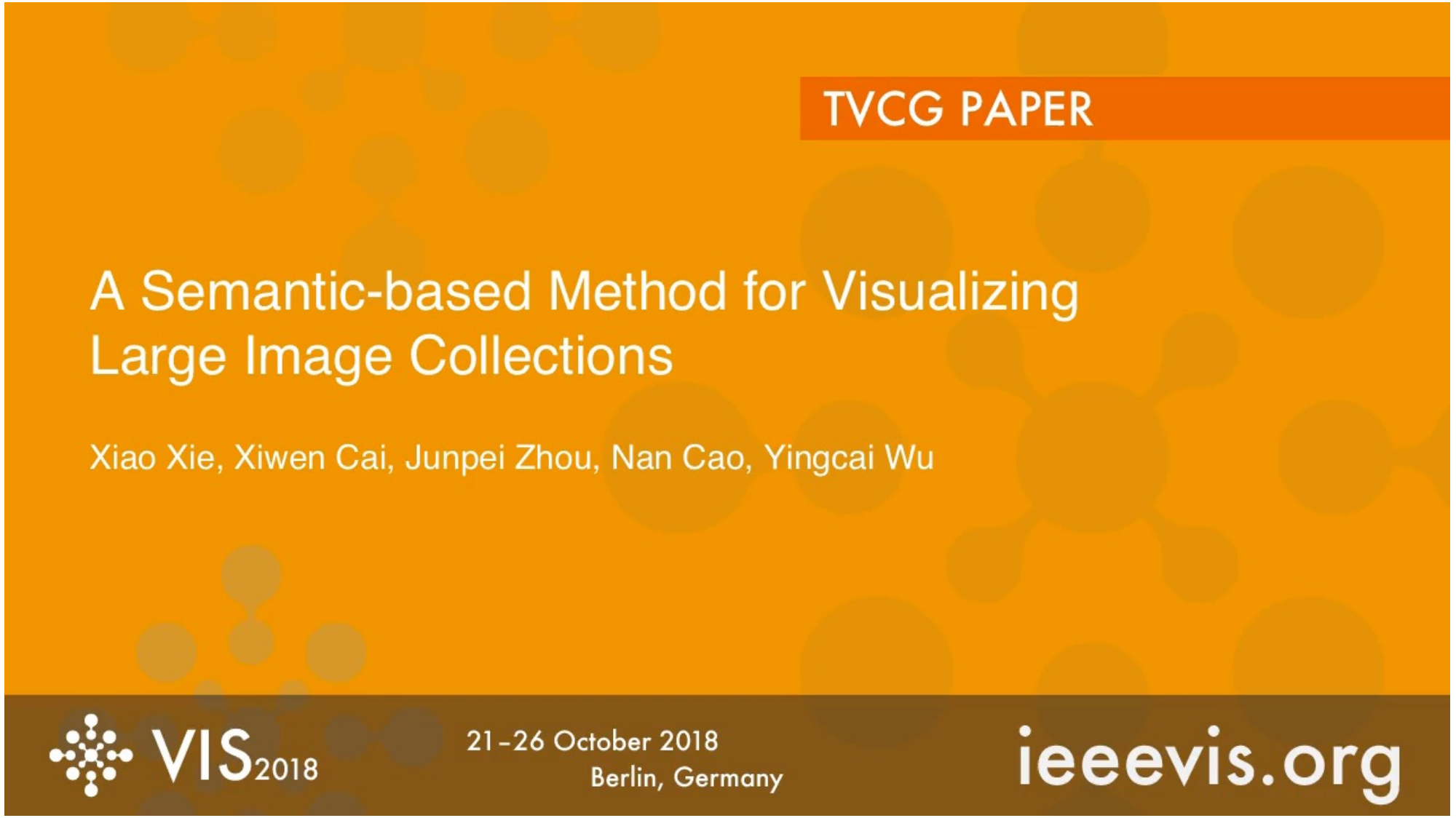
Algorithmic Methods
(K-means; k nearest neighbors)

What's More: Visual Data Mining

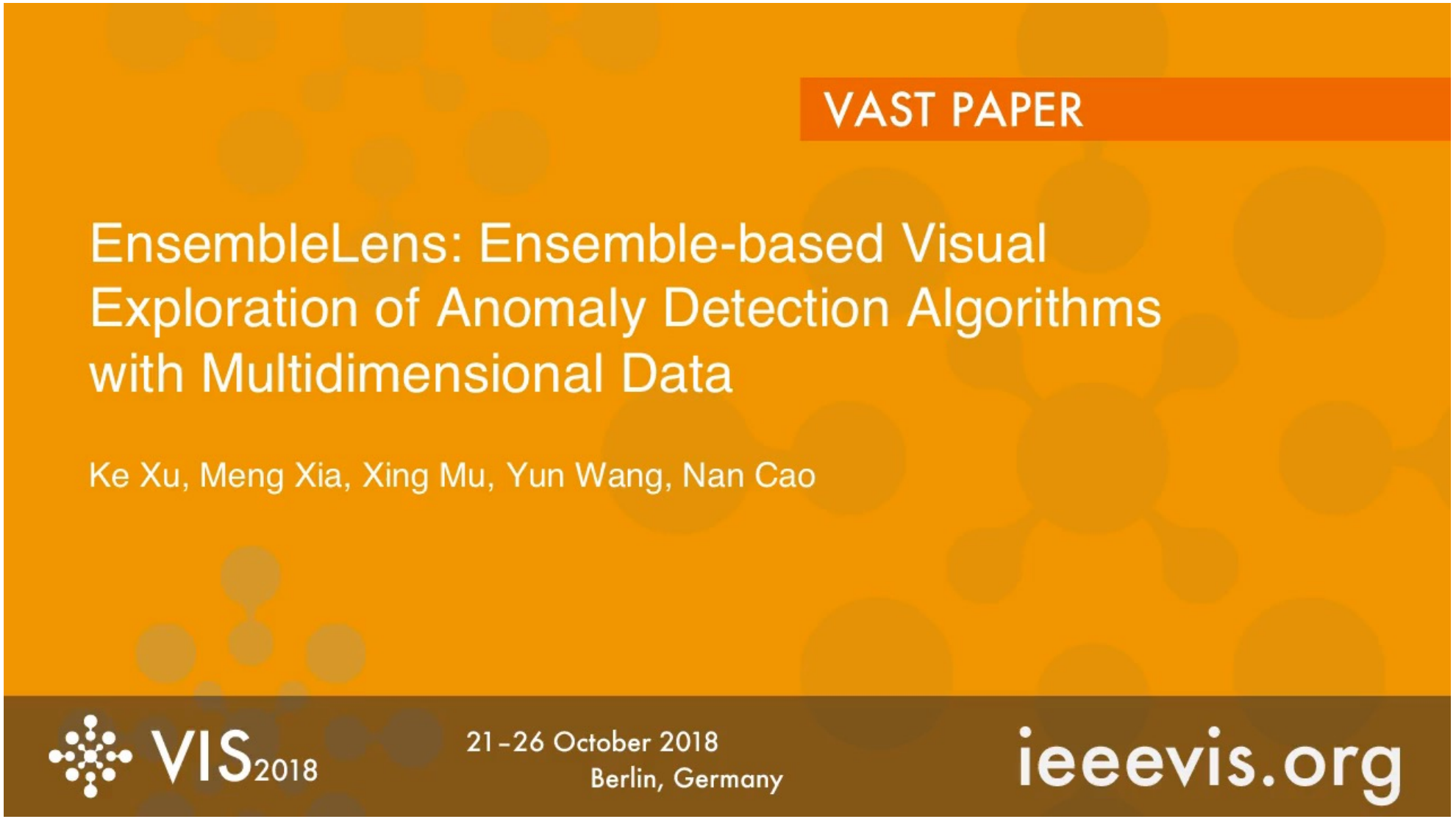
- User is involved in the process of data mining
- Some tasks are hard to accomplish via automated methods, e.g.:
  - Clustering result verification
  - Outliers inside dataset

Examples in Visual Data Mining

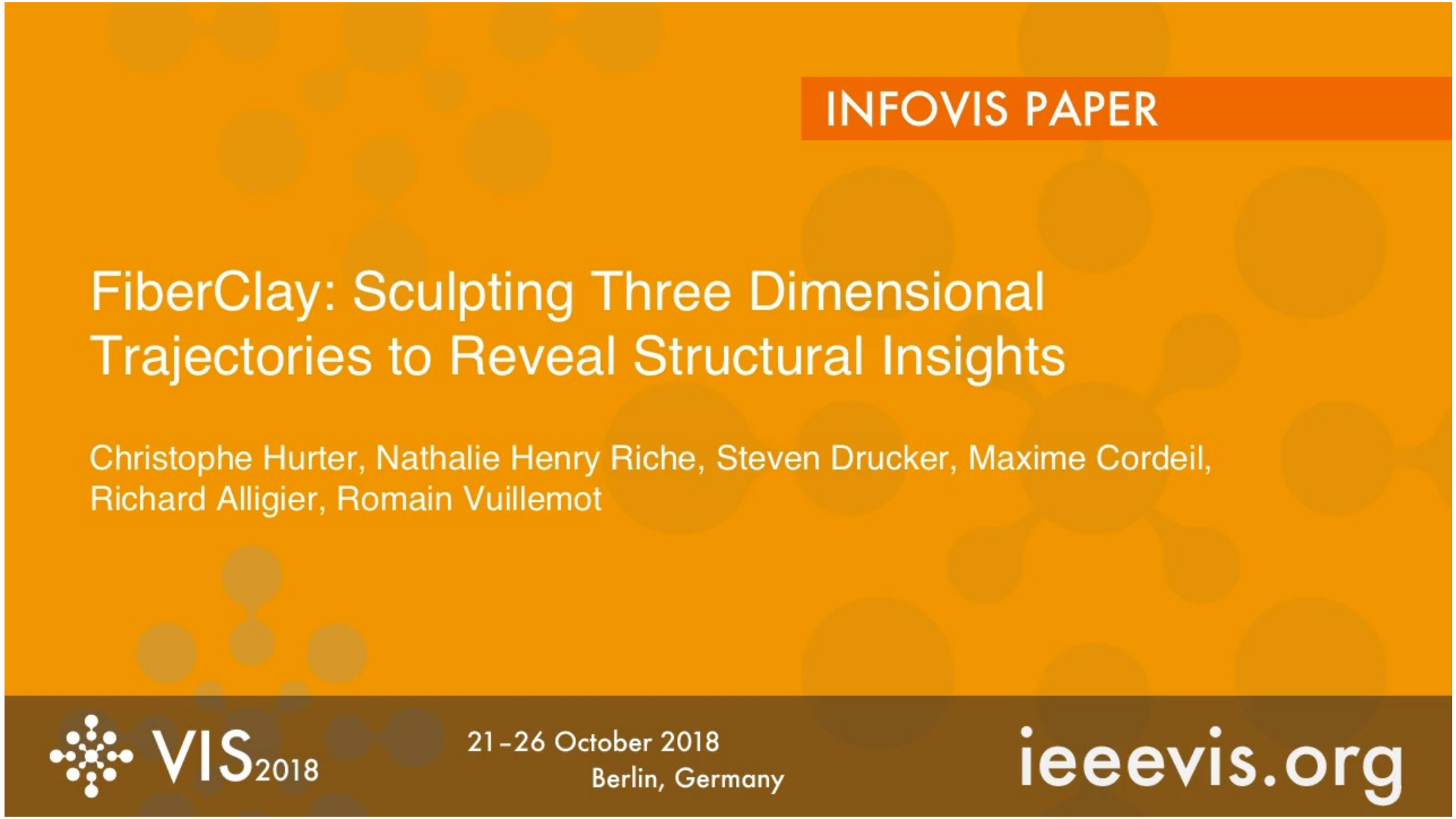
Examples in Visual Data Mining



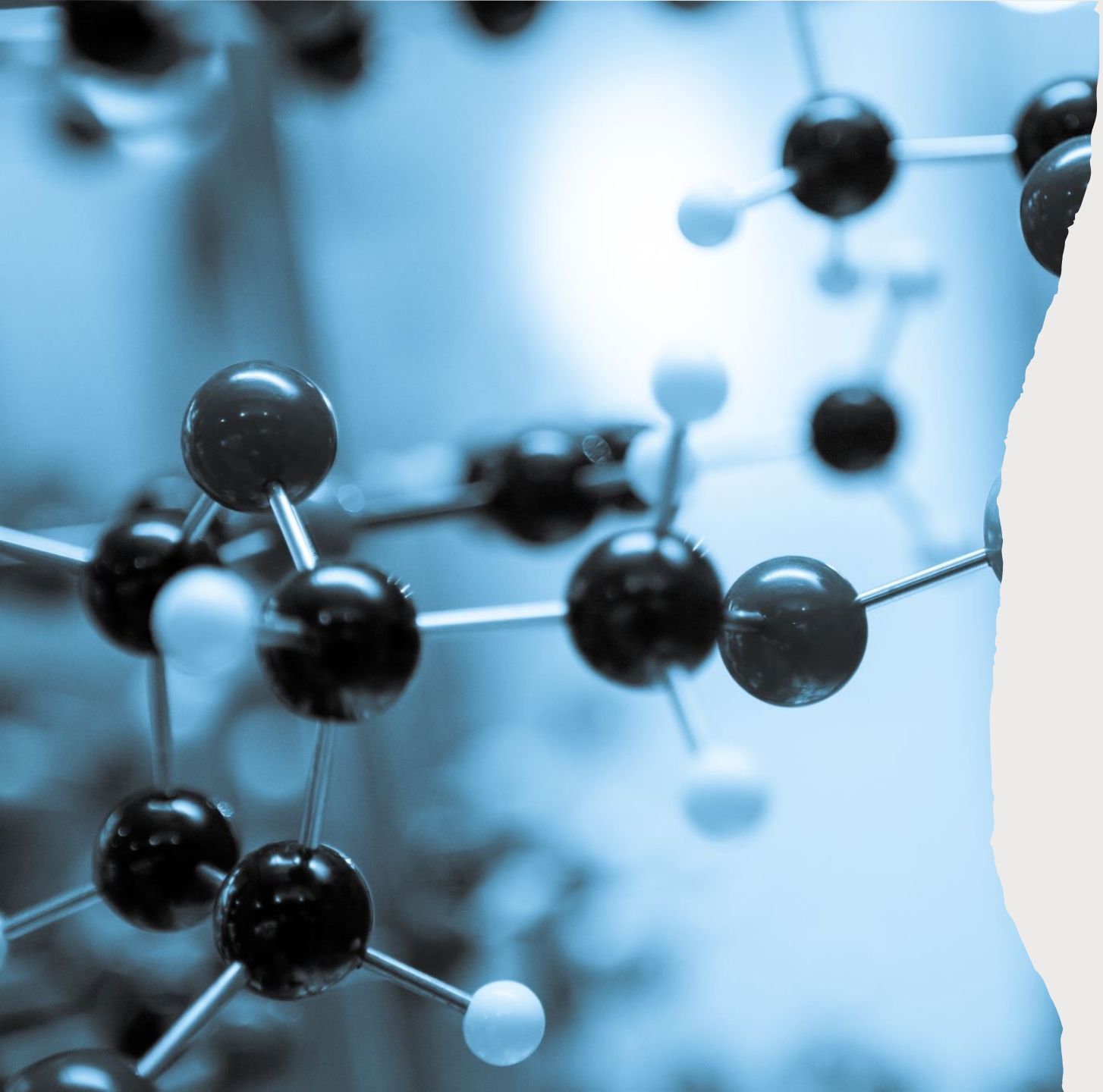
Examples in Visual Data Mining



Examples in Visual Data Mining



Examples in Visual Data Mining



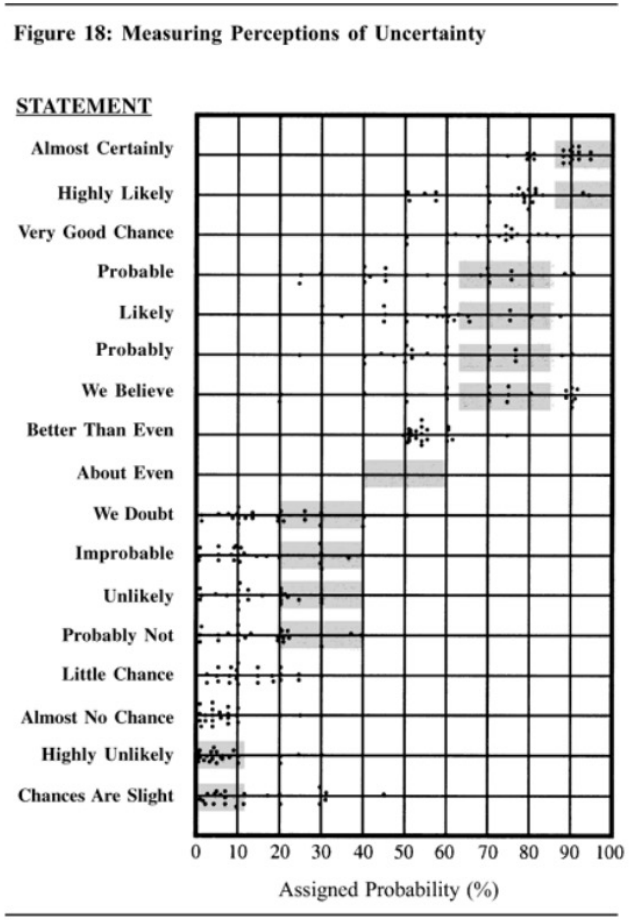
Conclusion: Why Data?

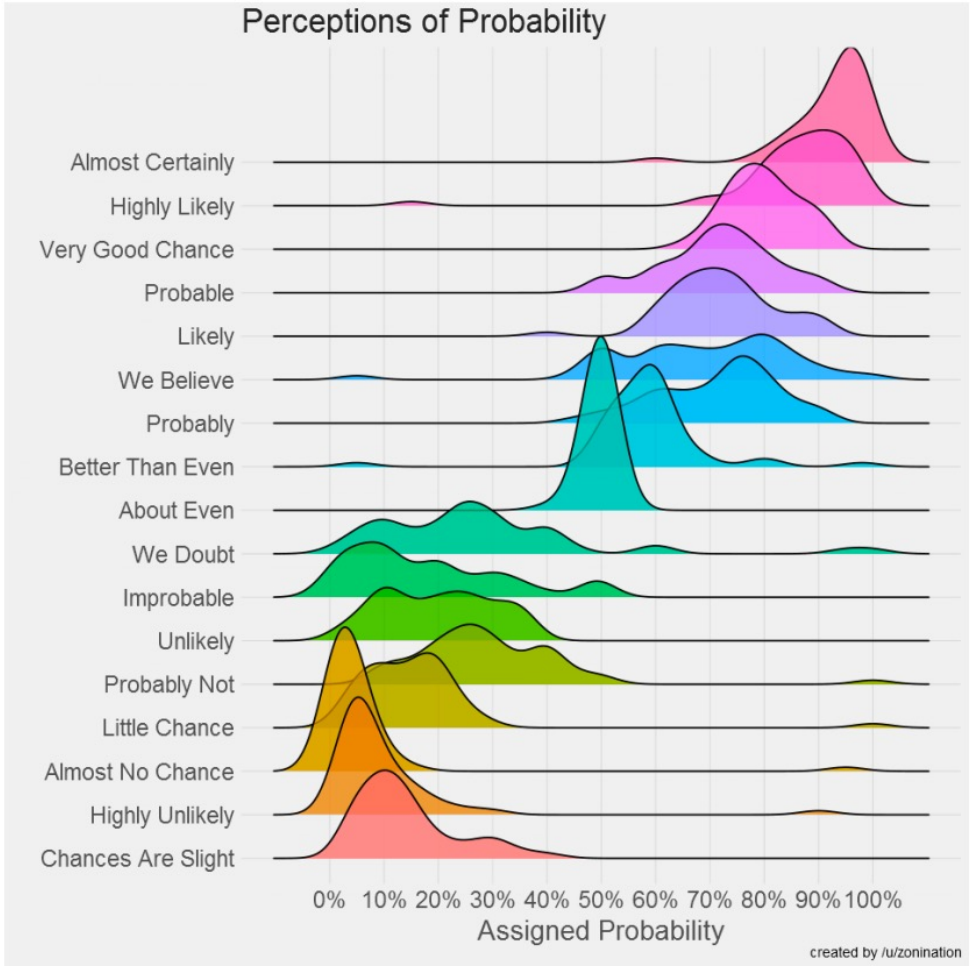
- No data, no visualization
- Visualization can boost every steps in data science

Graphical Excellence

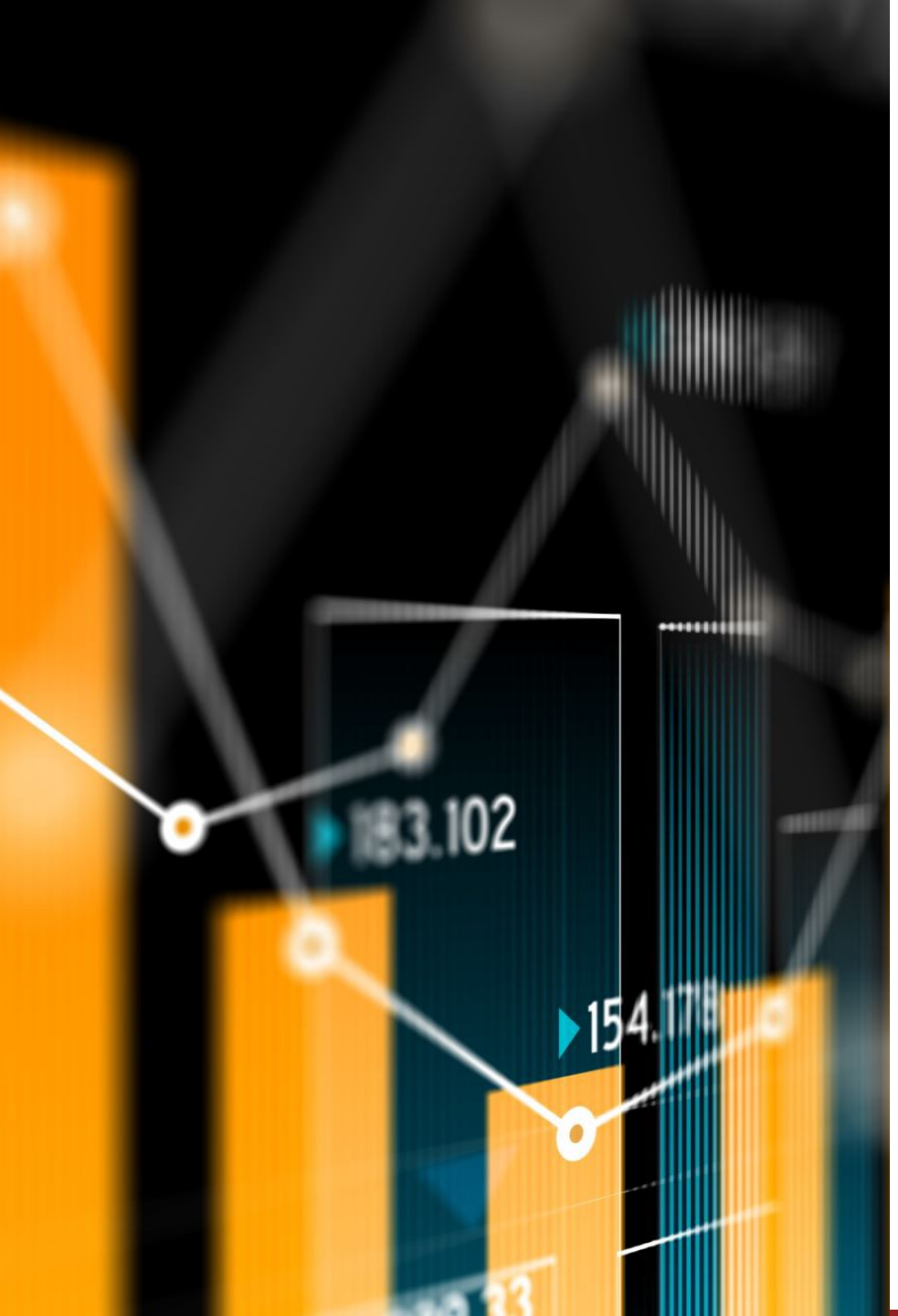
- Why are you showing these figures?
  - Well-designed presentation of interesting data, a combination of substance, statistics and design
- Why do they matter?
  - Greatest number of ideas in the shortest time with the least ink in the smallest place
- Am I conveying the information in the right way?
  - Tell the truth about the data
    - Design the right visualization
    - Design the visualization right

- This graph reports the results of a survey by Sherman Kent for the CIA with the question: what [probability/number] would you assign to the phrase "[phrase]"
- The goal was to understand how intelligence analysts use these terms.



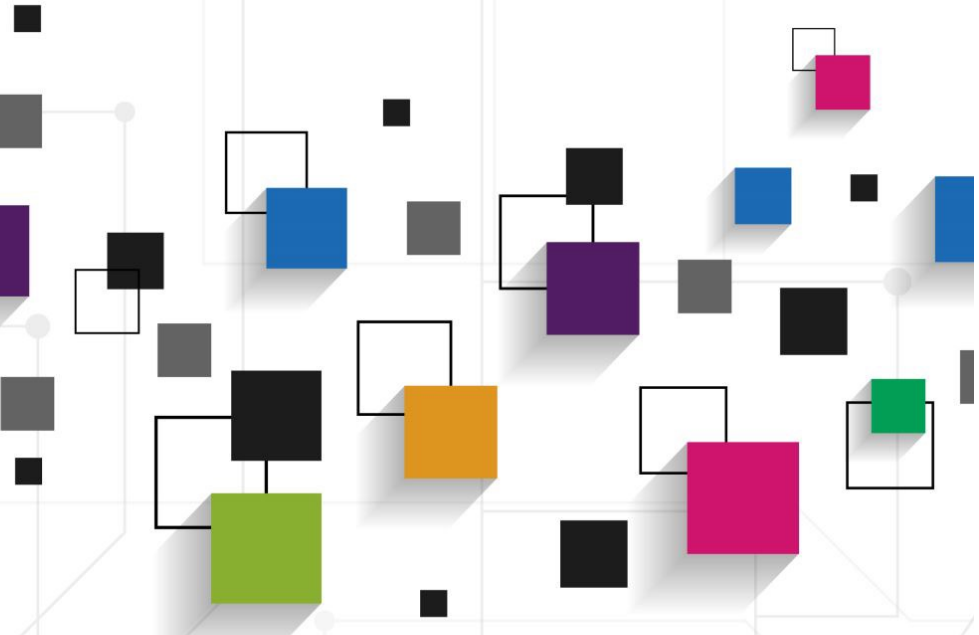


https://github.com/zonination/perceptions



Five Key Principles of Data Visualization

- **Graphical Integrity:** Visual representations of data must tell the truth
- **Data-Ink Ratio:** Data Ink is the ink on a graph that represents data. Good graphical representations maximize data-ink and erase as much non-data-ink as possible
- **Avoid Chart Junk:** Chart Junk is the excessive and unnecessary use of graphical effects in graphs
- **Data Density:** Proportion of total size of the graph that is dedicated to displaying data. Maximize data density and data matrix within reason
- **Small Multiples:** Series of the same small graph repeated in one visual. Small multiples are a great tool to visualize large quantities of data and with a high number of dimensions

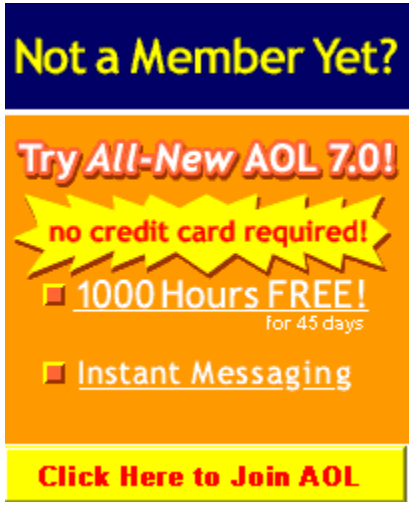


Principles of Graphical Integrity

- **Minimize the Lie Factor**
  - Tell the truth about the data – above all else show the data
- **Use Consistent Scale**
  - Show data variation, not design variation
  - Make large data sets coherent
- **Present Data in Context**
  - Reveal the data at several levels of detail from broad overview to the fine structure

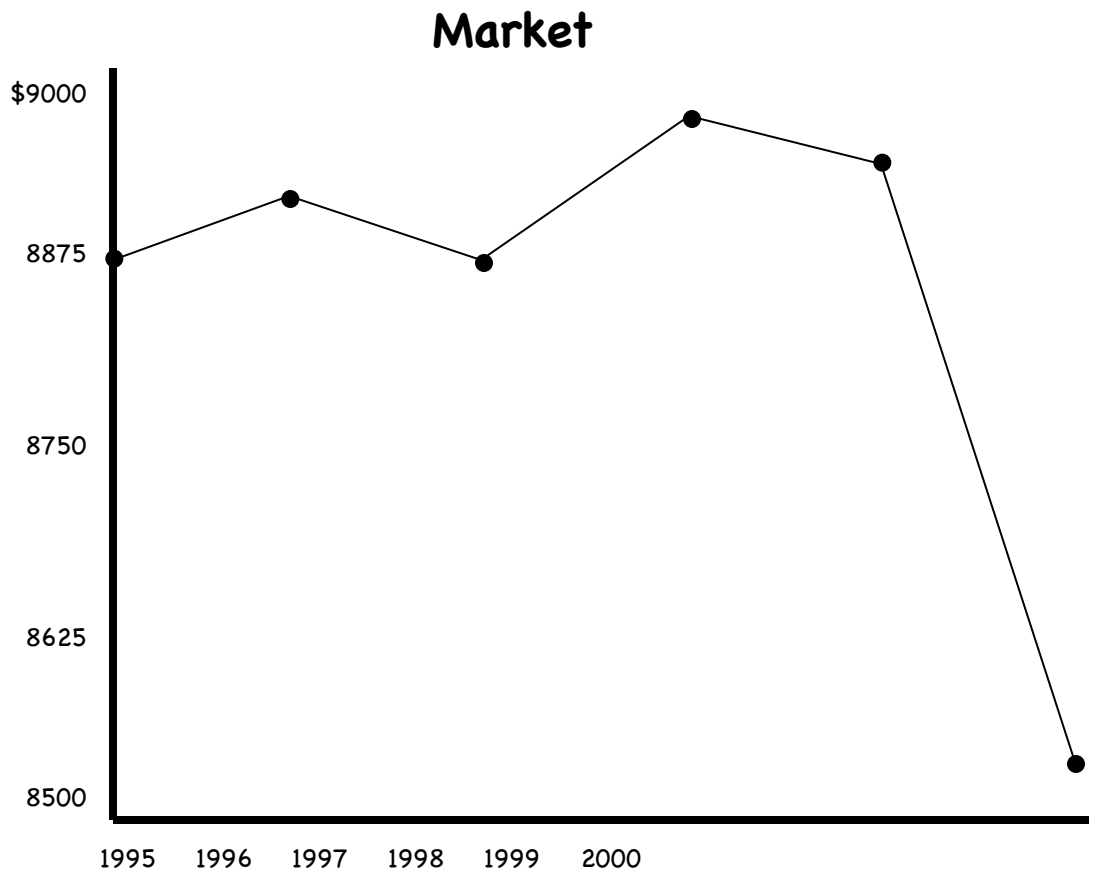
How (not) to lie with visualization

- Show and tell

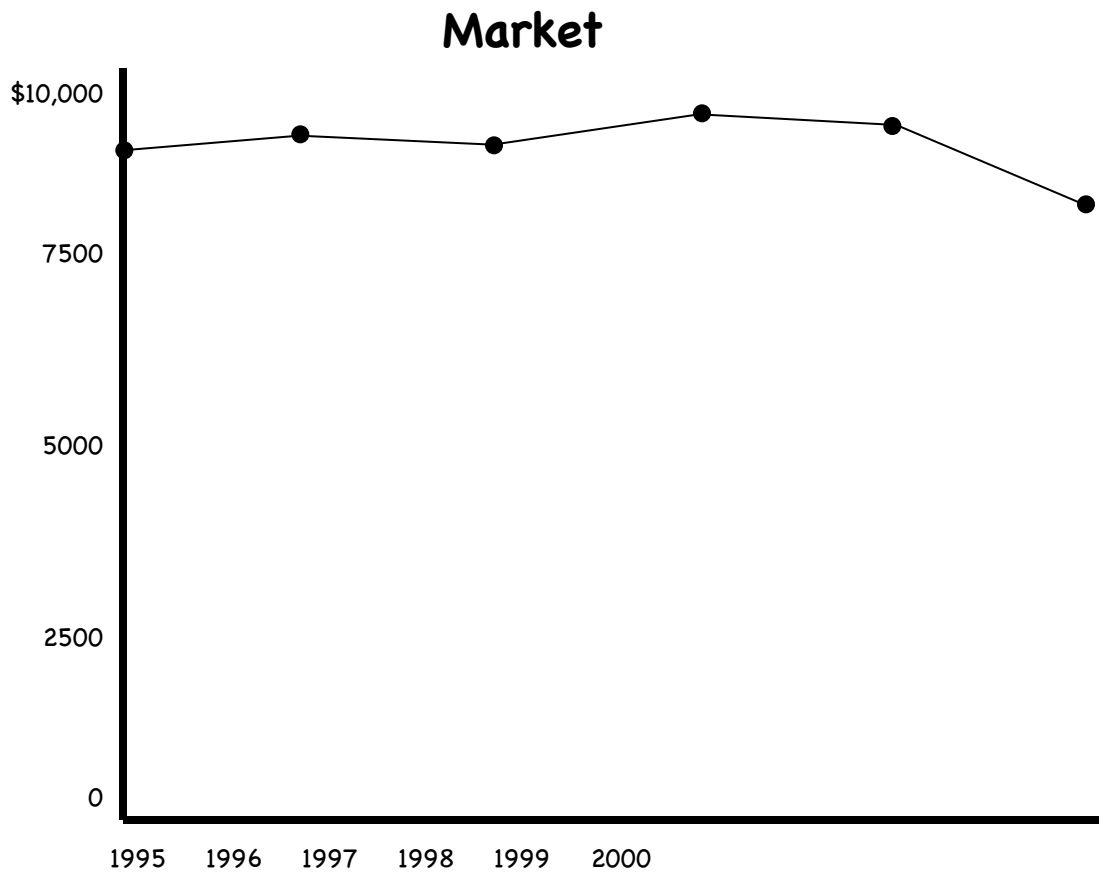


- "USA Today" graphs…

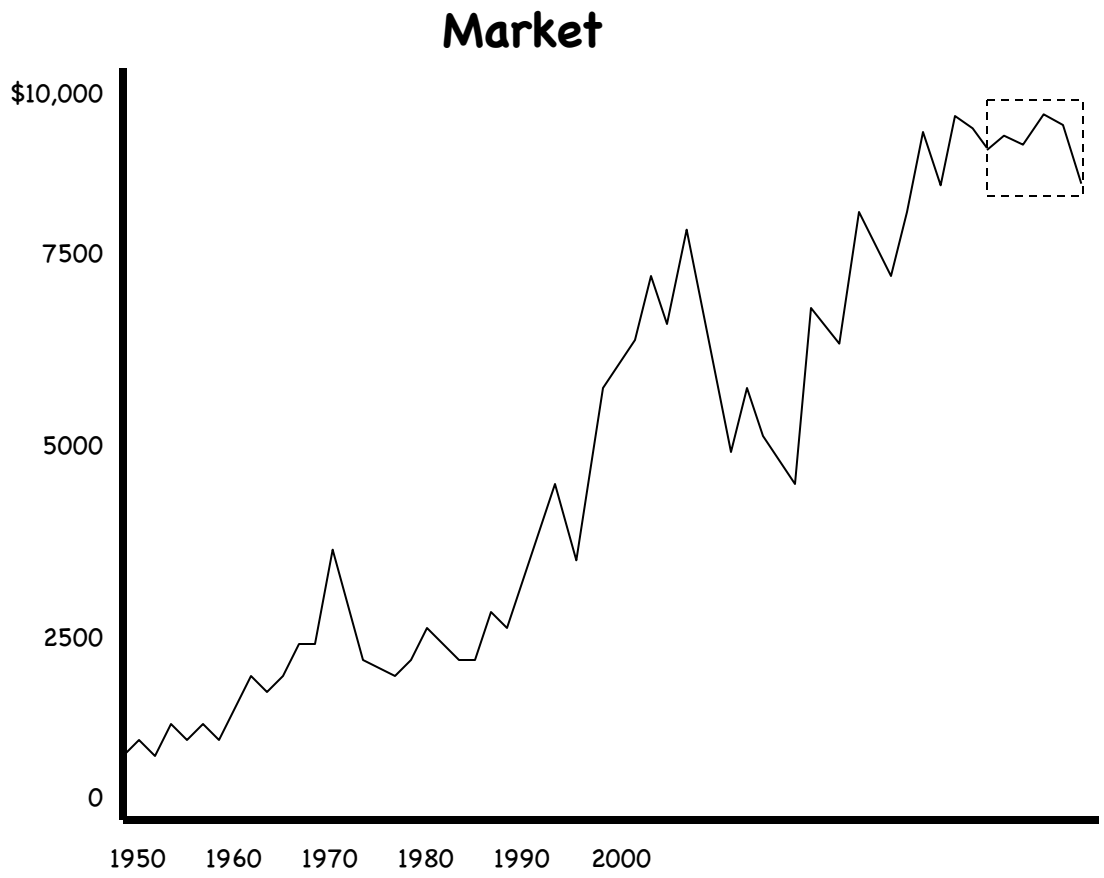
Stock Market Crash?!



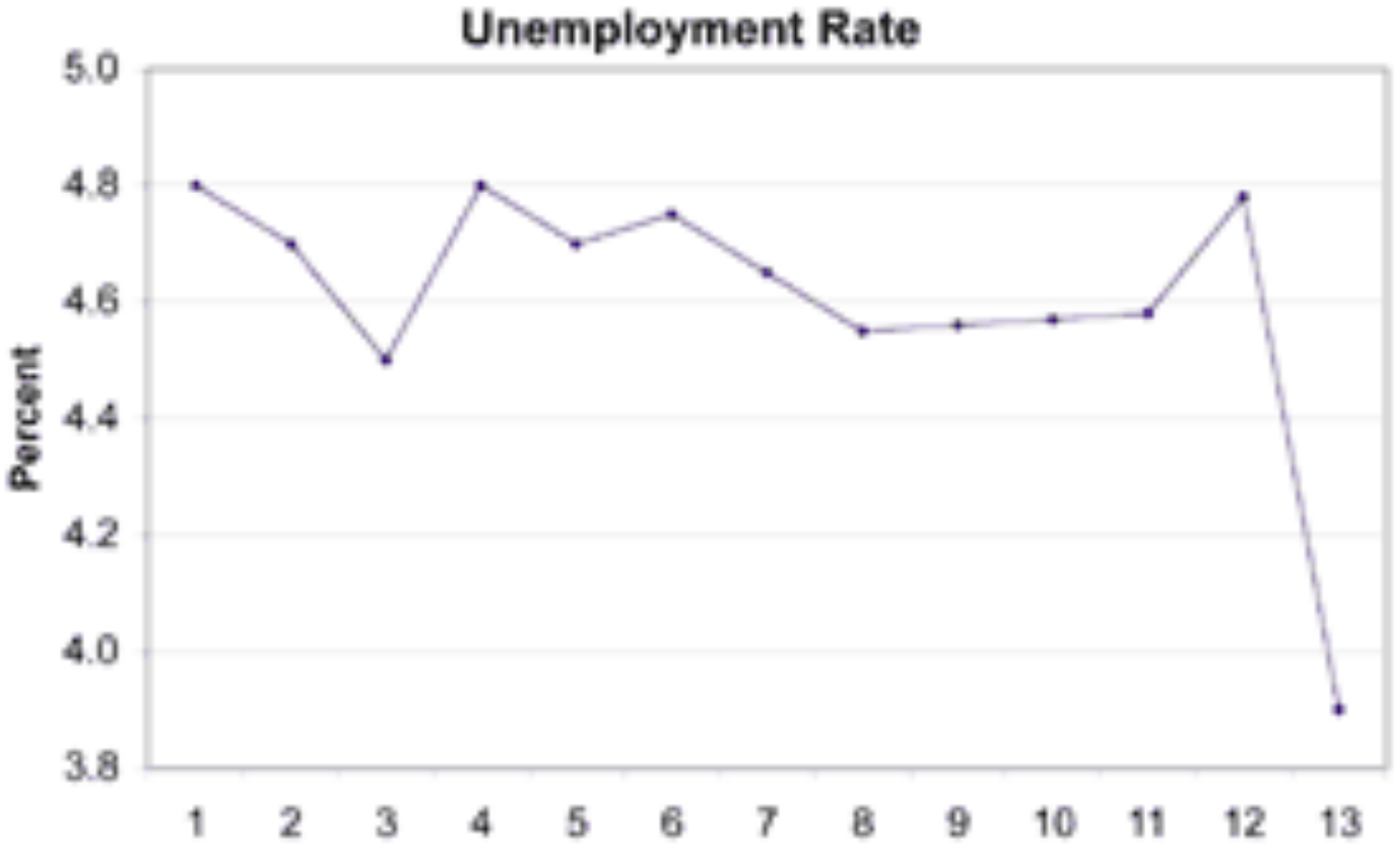
Showing entire scale



Shown in context

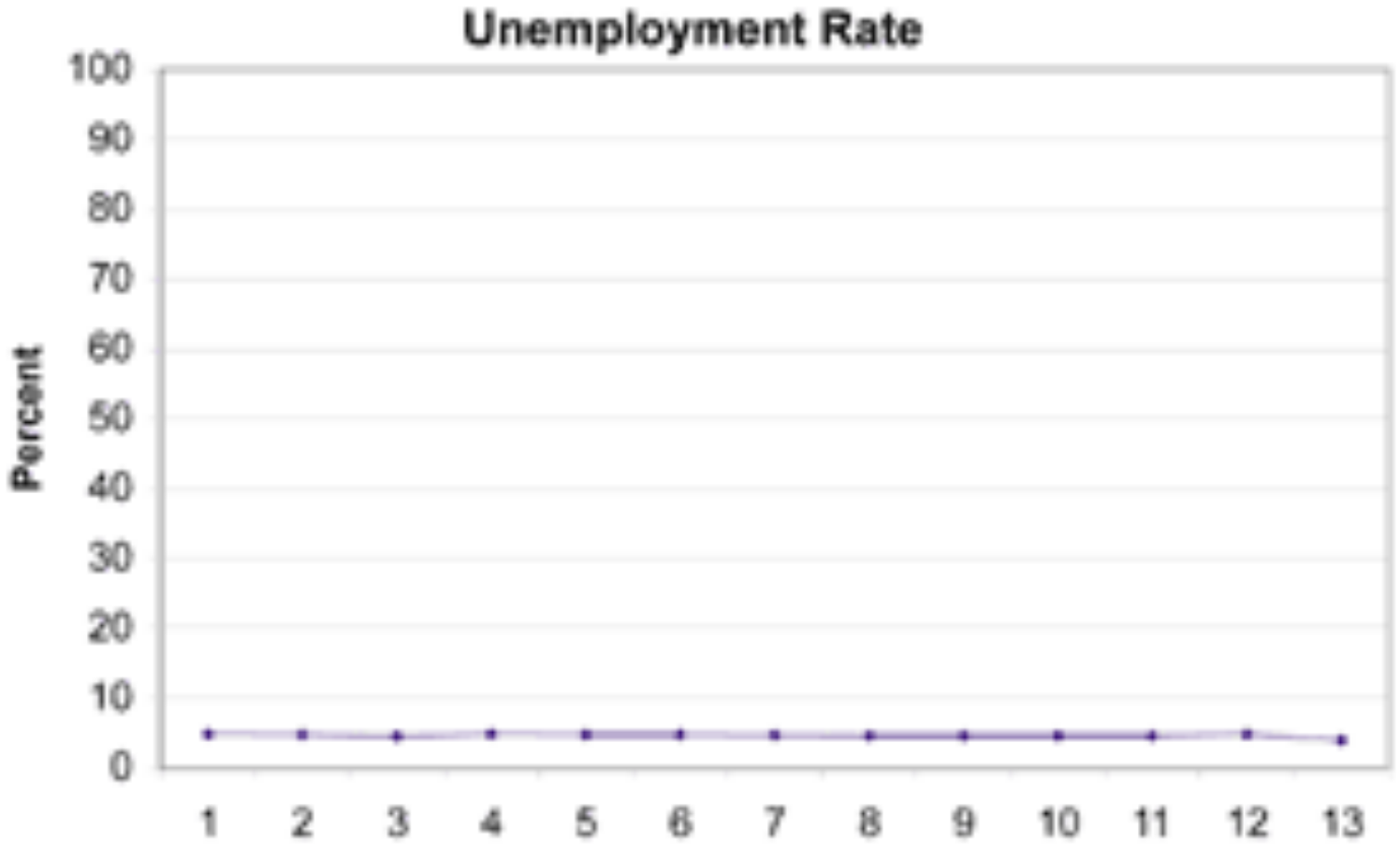


Another example



Percentages: 0% – 100%

Employment rate = 100 – unemployment rate



Tufte's Rule: Lie Factor

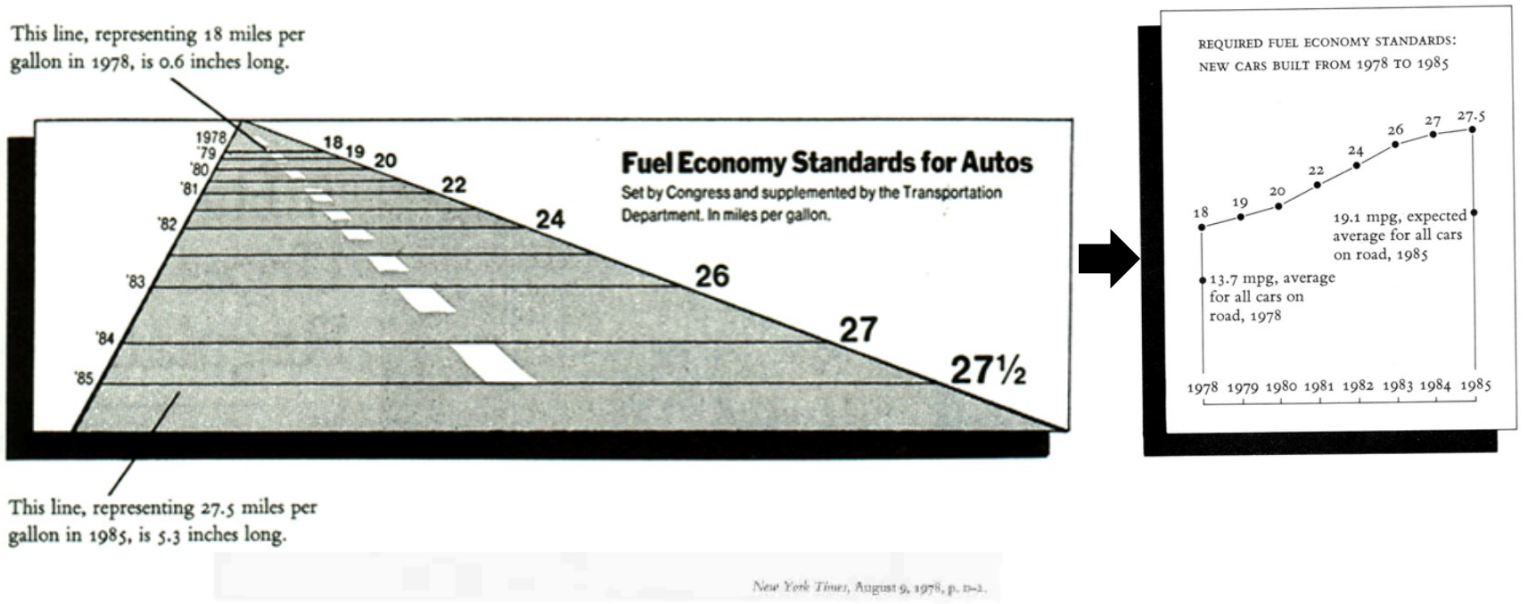
- Visual attribute value should be directly proportional to data attribute value

$$Lie\ Factor = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

$$size\ of\ effect = \frac{|\text{second value - first value}|}{\text{first value}}$$

- Truth: lie factor = 1.0
- Graph overstating: lie factor > 1.0
- Integrity threshold: lie factor between 0.95 and 1.05

Lie Factor

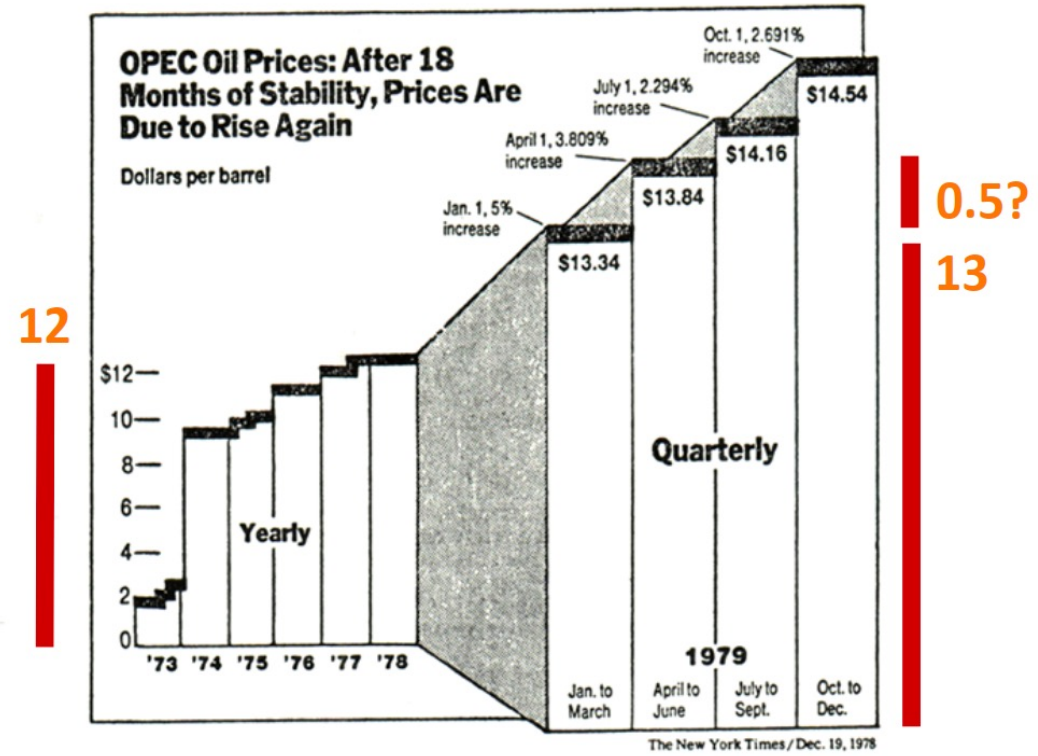


1985: the line representing 27.5 m/g is 5.3 inches long

1978: the line representing 18.0 m/g is 0.6 inches long

$$\text{Lie Factor} = \frac{\frac{5.3 - 0.6}{0.6}}{\frac{27.5 - 18}{18}} = 14.8$$

Changing Scale



Five different vertical scales show the price:

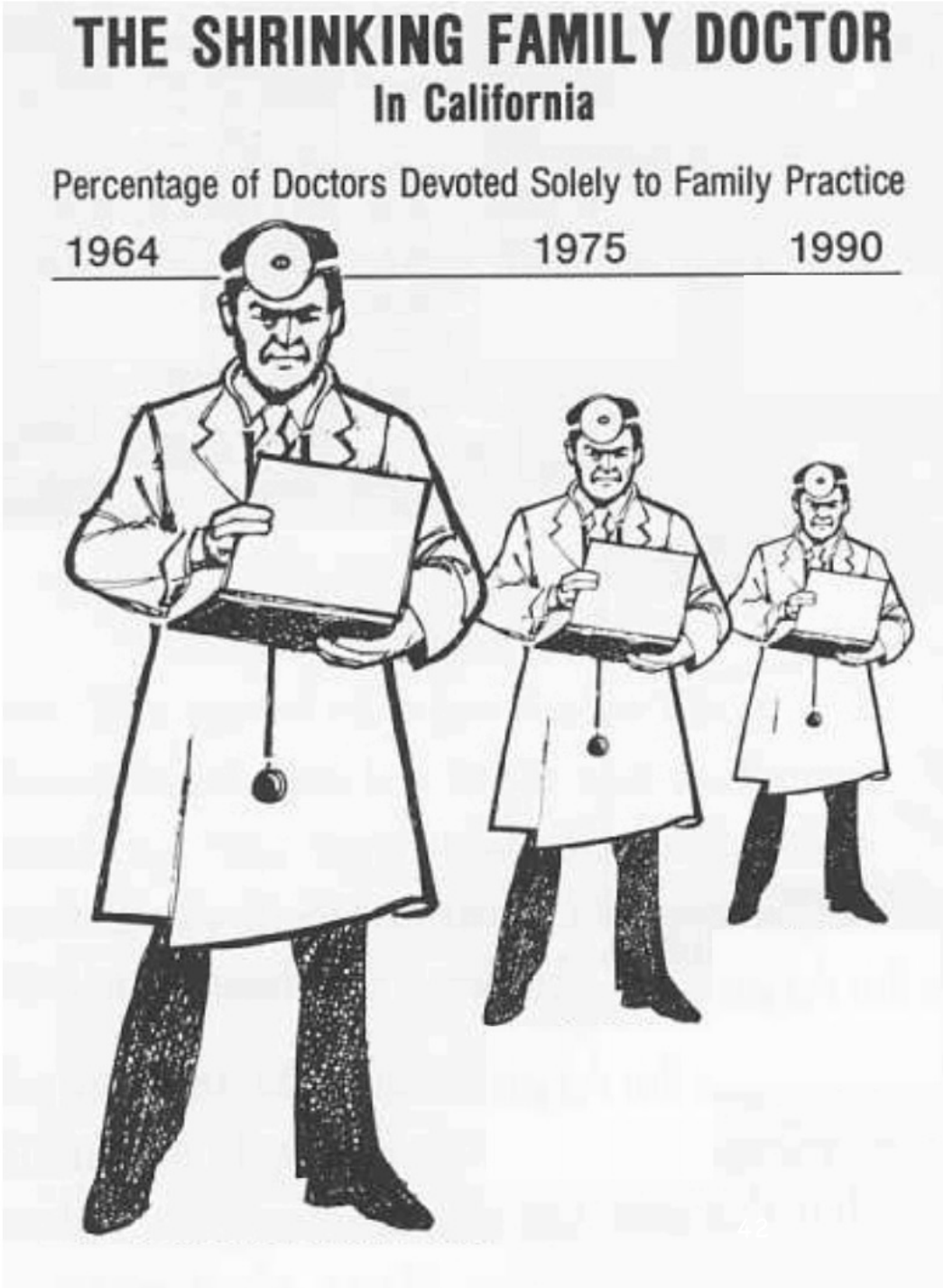
During this time	one vertical inch equals
1973–1978	$8.00
January–March 1979	$4.73
April–June 1979	$4.37
July–September 1979	$4.16
October–December 1979	$3.92

And two different horizontal scales show the passage of time:

During this time	one horizontal inch equals
1973–1978	3.8 years
1979	0.57 years

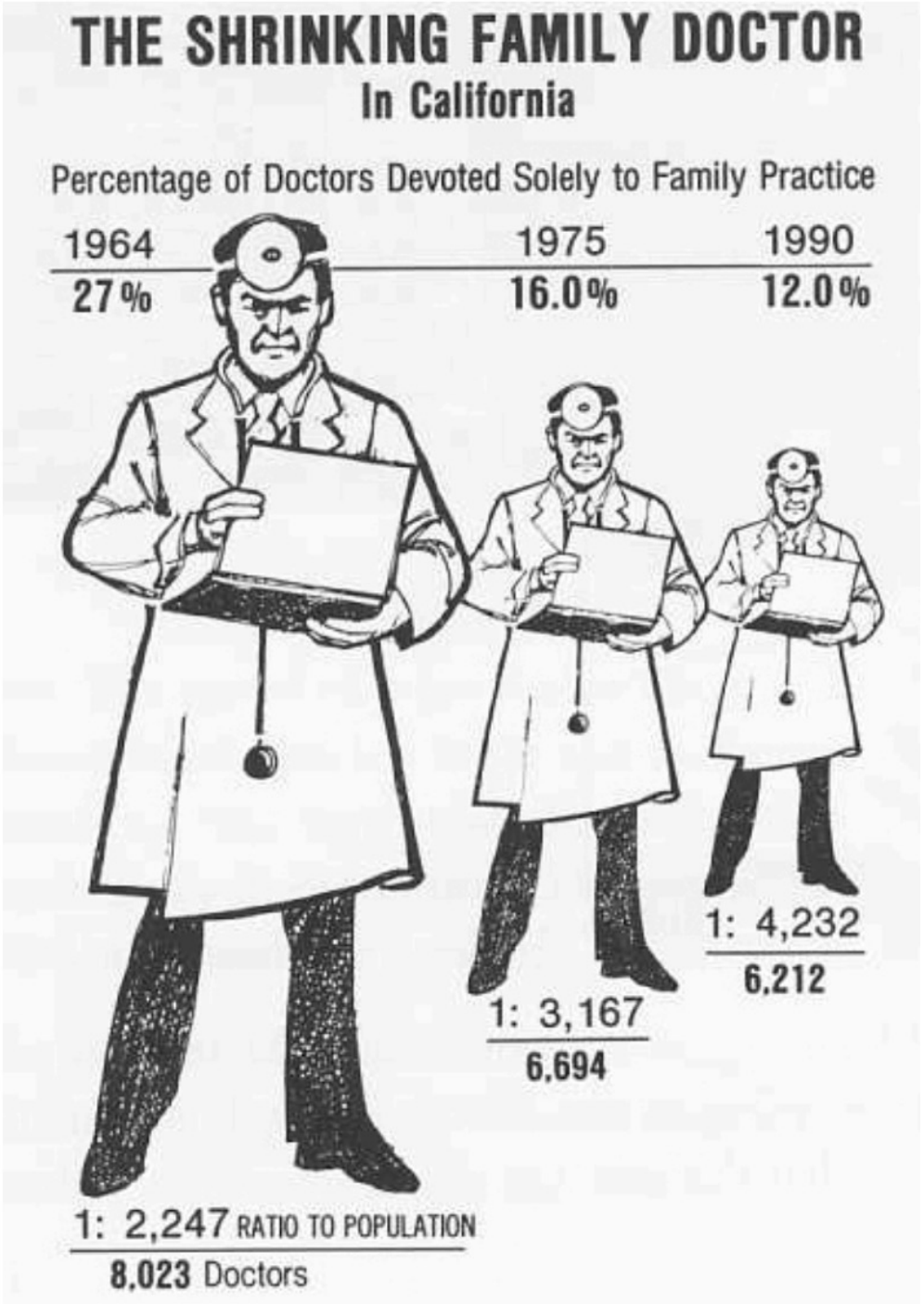
Size Encoding

- The percentage in 1964 is how many times of that in 1990?
- What is your basis?



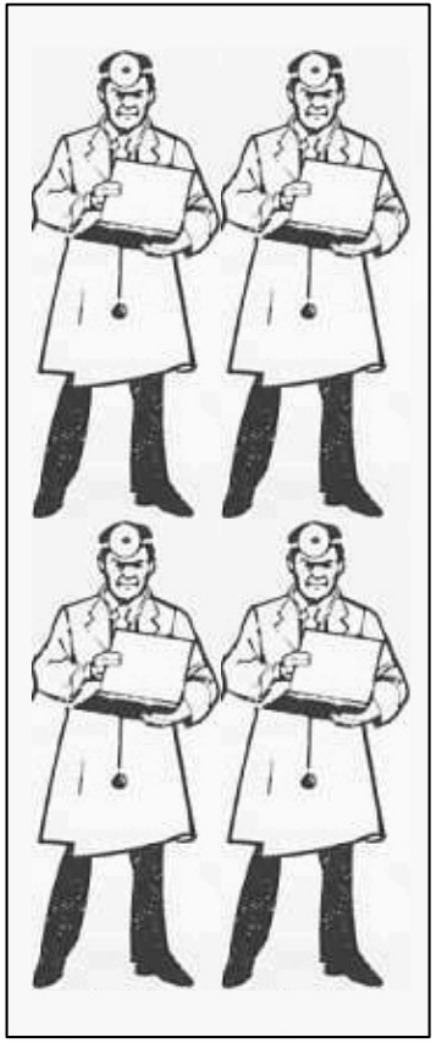
Size Encoding

- What is the lie factor in this case?



Size Encoding: height or area?

=



?

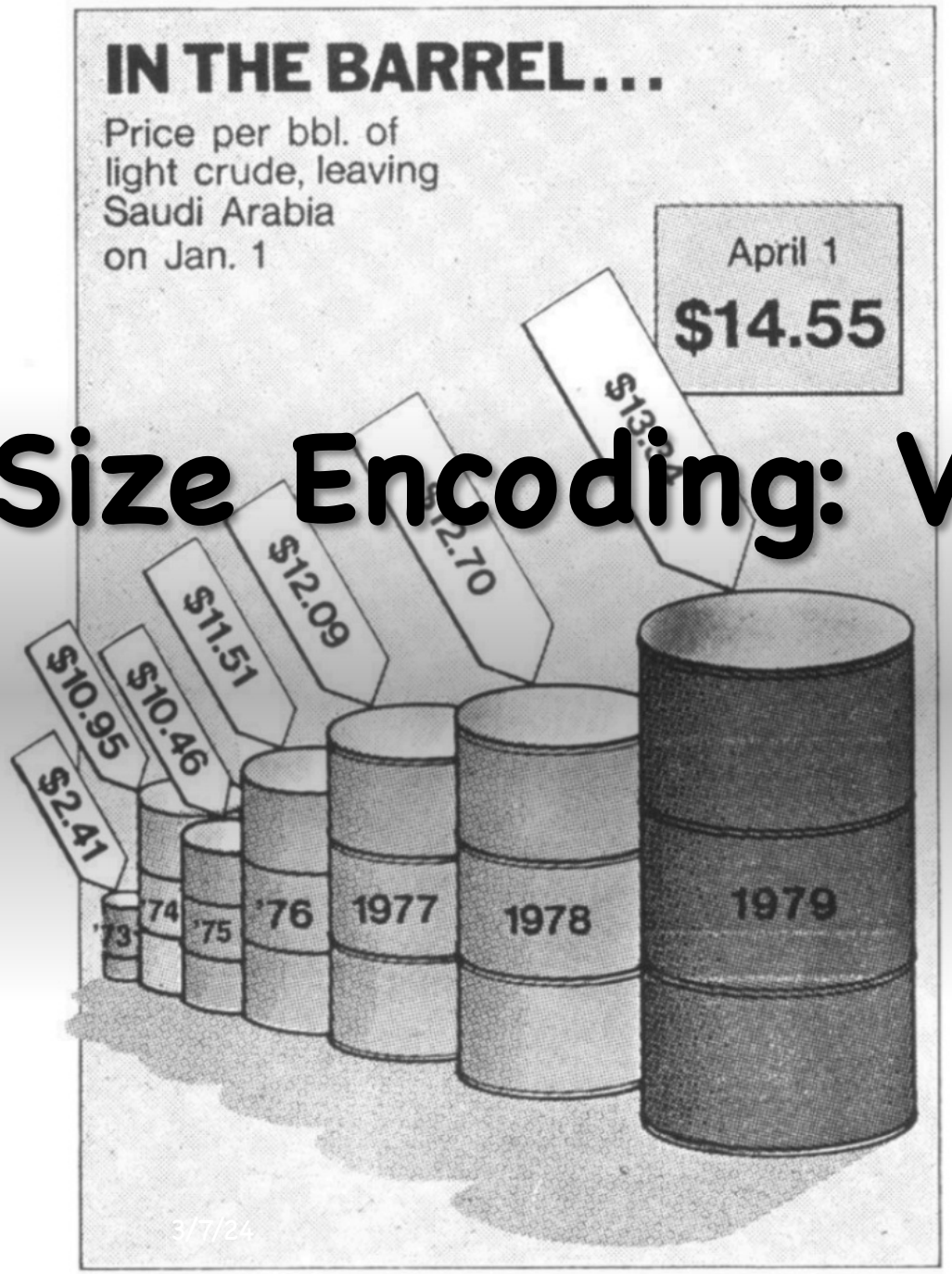
Size Encoding: height or area?

- Height = value
  Width = value
  Area = value^2

  or

- Area = value
  height × width = value
  height = width = value $^{0.5}$

Problem:
Using 2D image to represent 1D value



Height?

Diameter?

Surface area?

Volume?

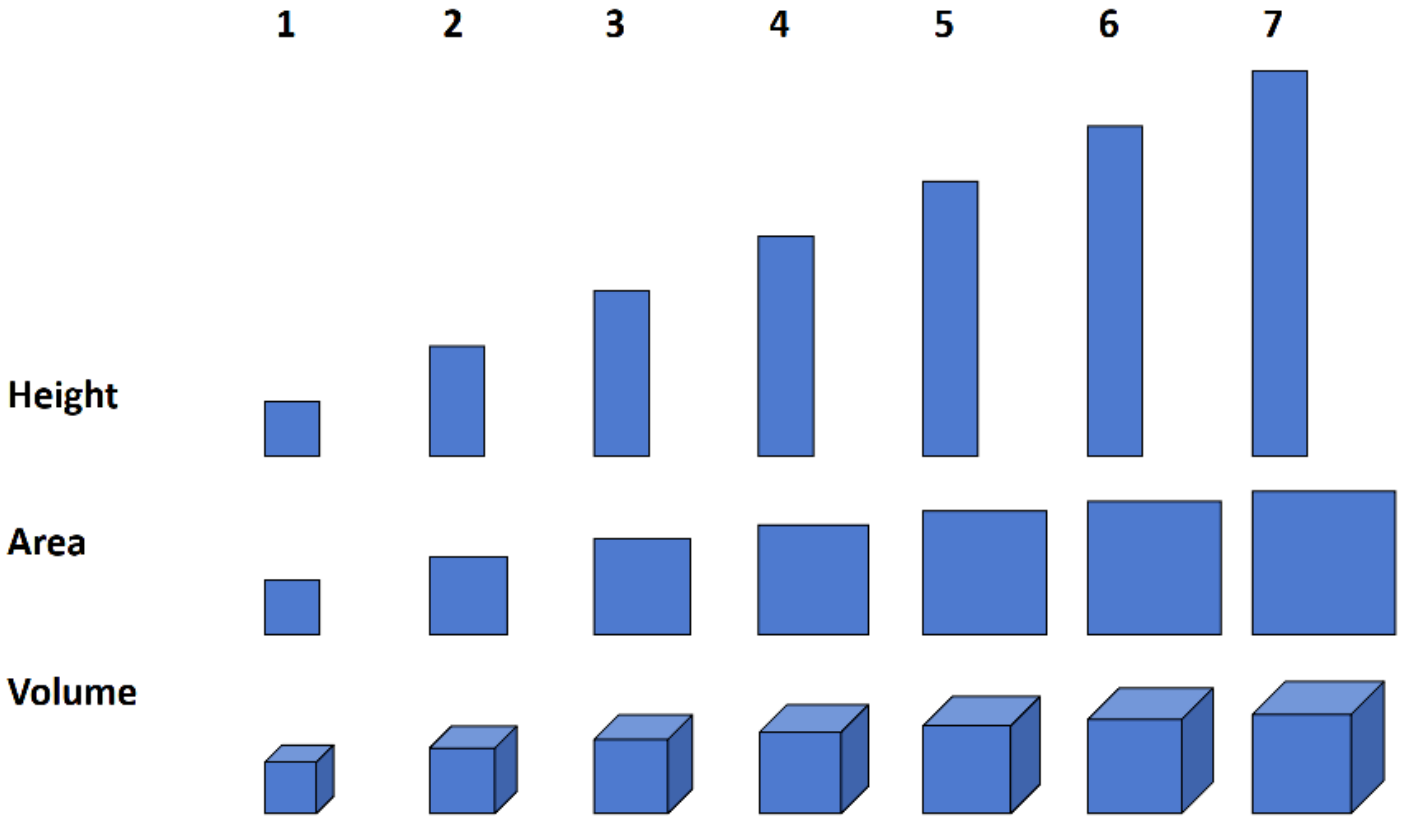
Size Encoding: Volume? (Lie Factor = ?)

73 – 79 data difference = 5.5x

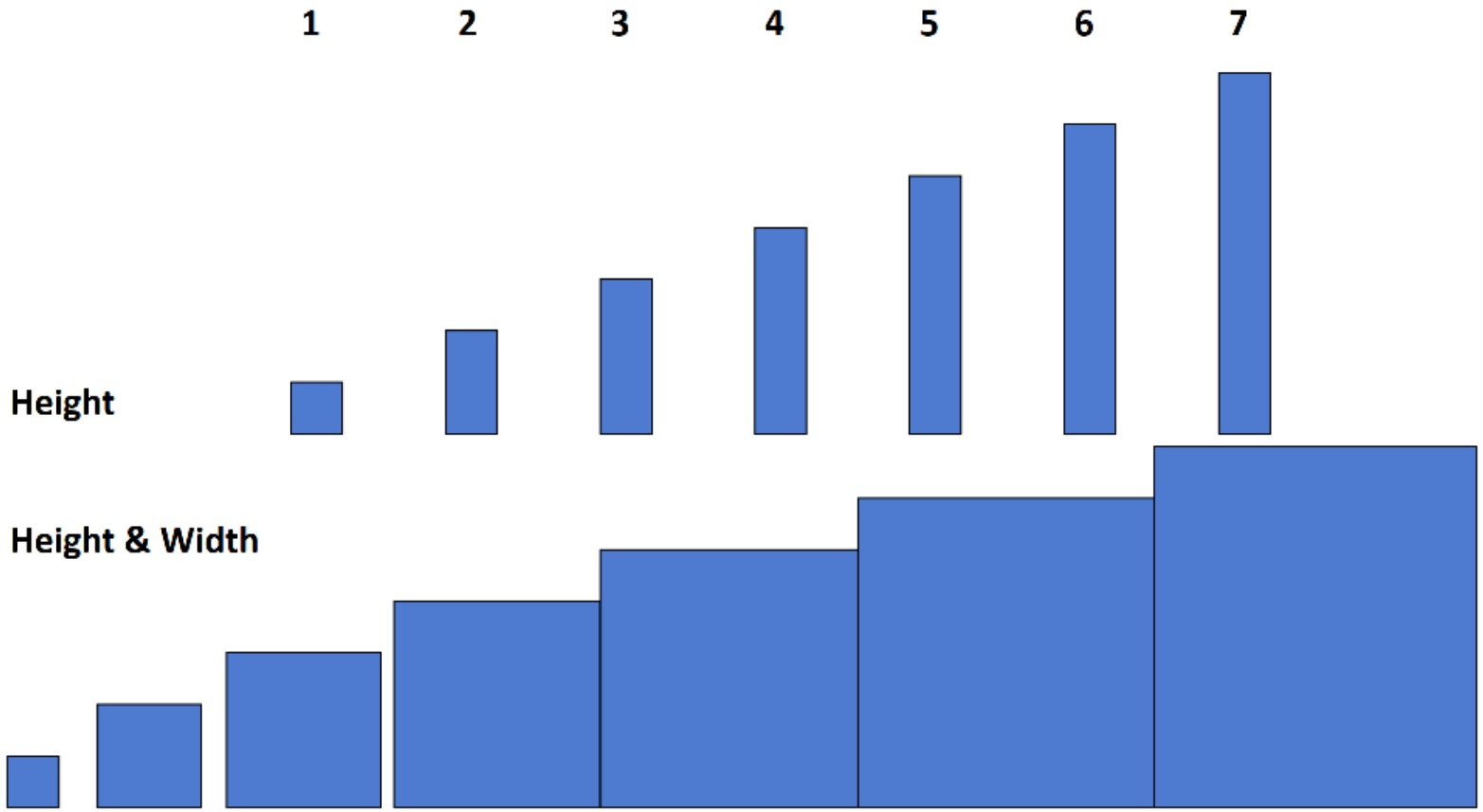
73 – 79 volume difference = 270x

Lie factor = 9.4 (2D) and 59.4 (3D)

Problem with Size Encoding



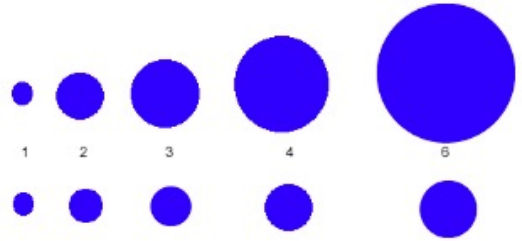
Height & Width Encoding



Perception of Area

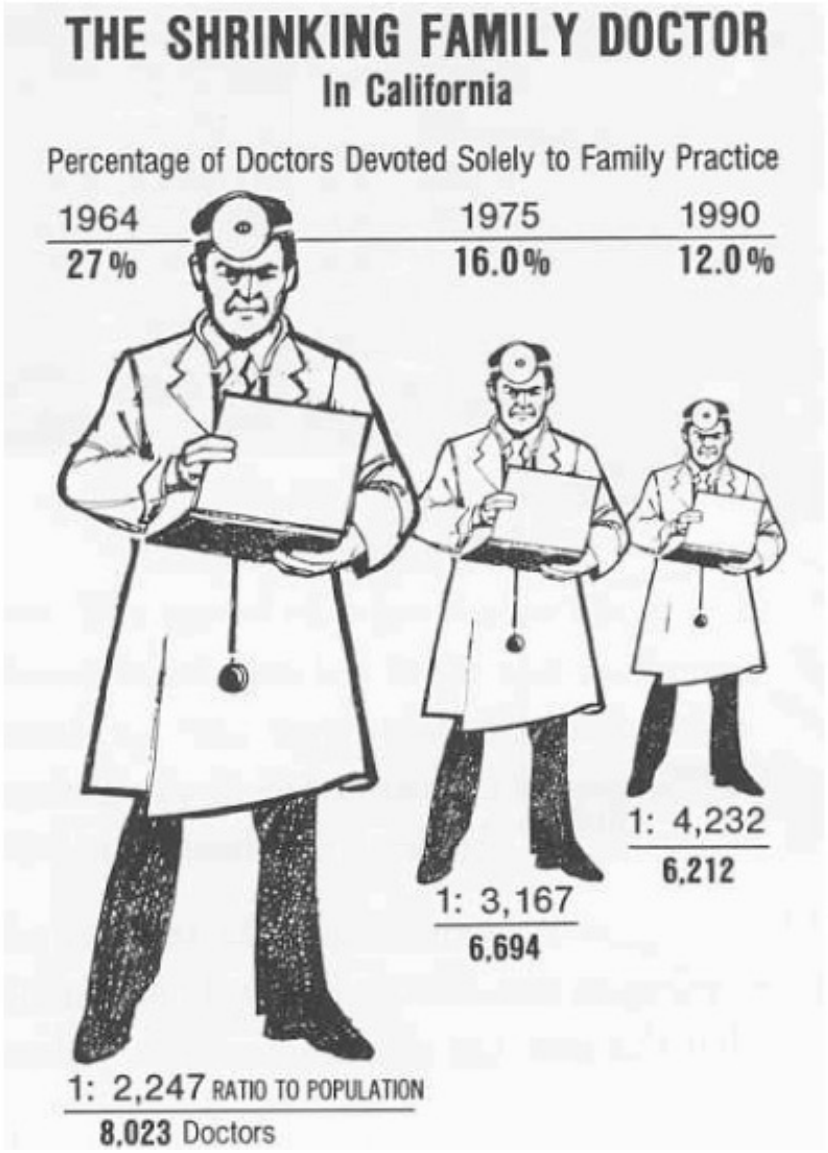
Perception of area versus magnitude varies per person. The perceived area of a circle grows more slowly than the actual:

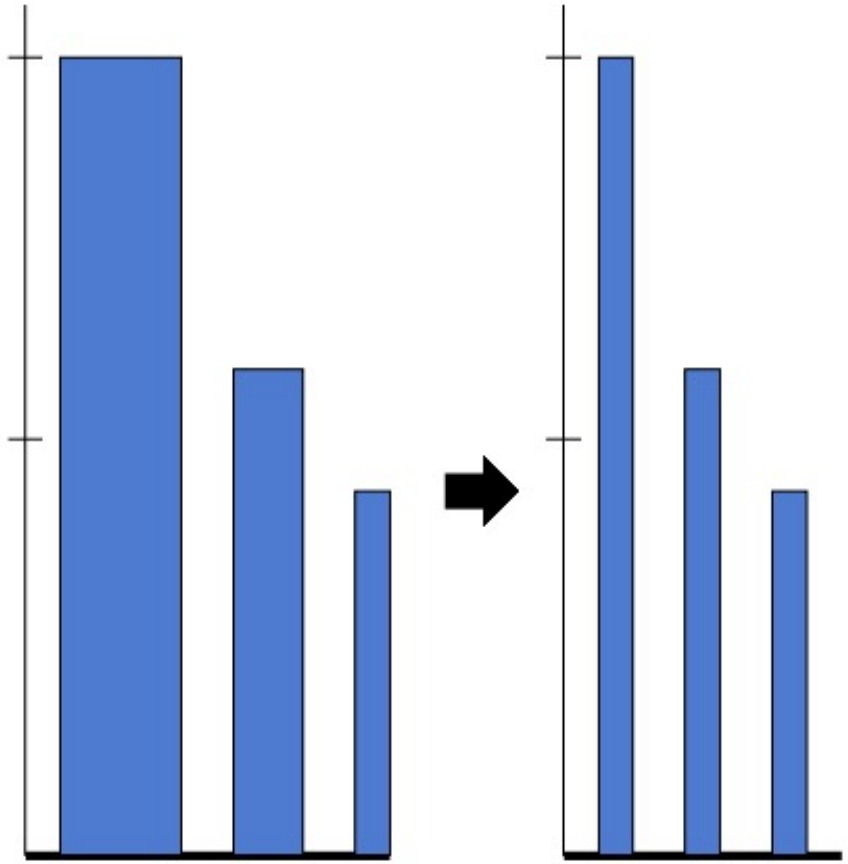
perceived area = (actual area)$^{(.8\pm.3)}$ So if the area is 4, the range of perceived area ranges from 2 =$4^{(.5)}$ to 6 =$4^{(1.1)}$ and if the circle's area grows to 8 then the perceived area could be from 3 =$8^{(.5)}$ to 15 =$8^{(1.1)}$.



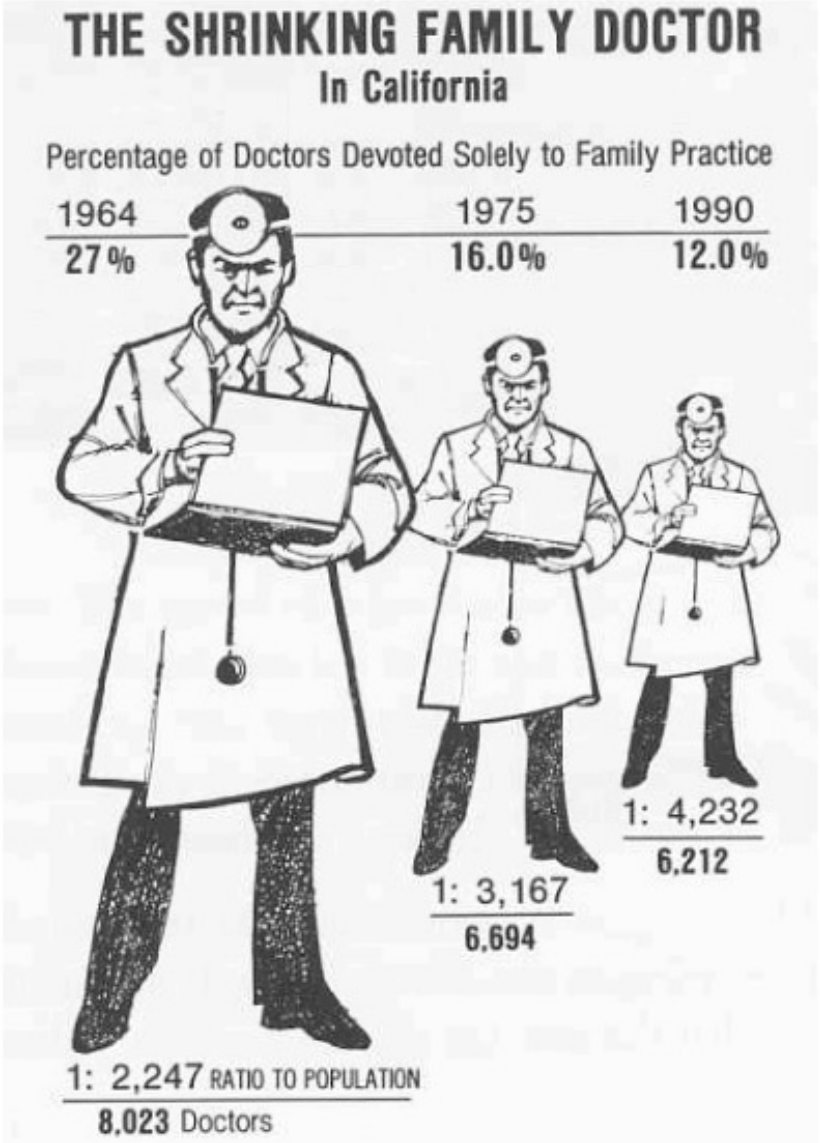
That is, if the area doubled, some might only see a 50% increase where others might see a 150% increase in the same visual.
Using two dimensional objects to represent scalars is naturally misleading--especially if you use the diameter as proportional to the scalars to view. Growth of a scalar value becomes perceived to be squared. Be sure the areas are what you really represent if you use such objects.

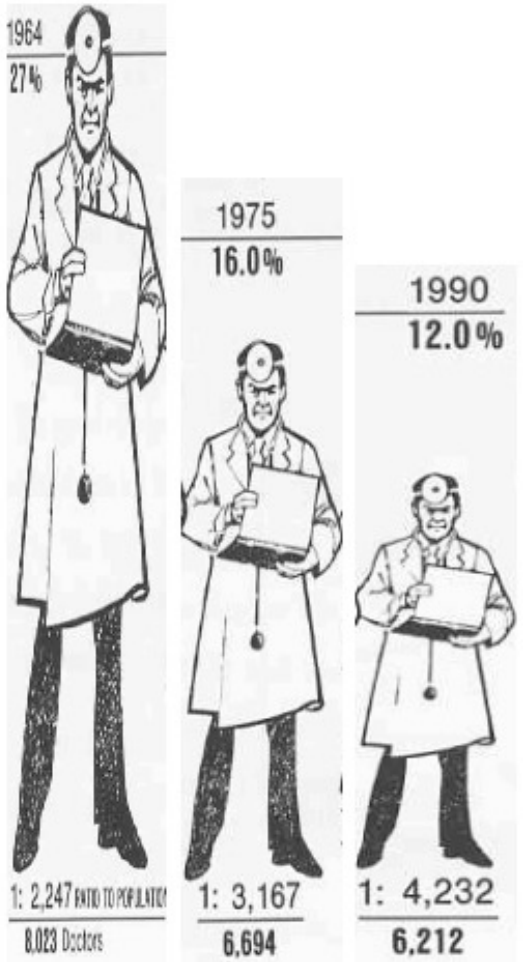
Solution: Just Use Height

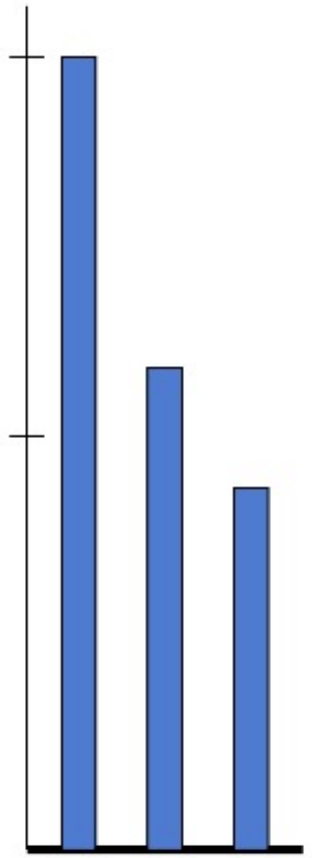




Solution: Just Use Height







How NOT to Lie

- Show entire scale
- Show data in context
- Consistent, linear scale
  - Log scale for log data
- Up vs. down: indicate direction of improvement
- Avoid size encoding
  - Use height OR width
  - Do not use both for same data attribute
  - Avoid area, volume encoding

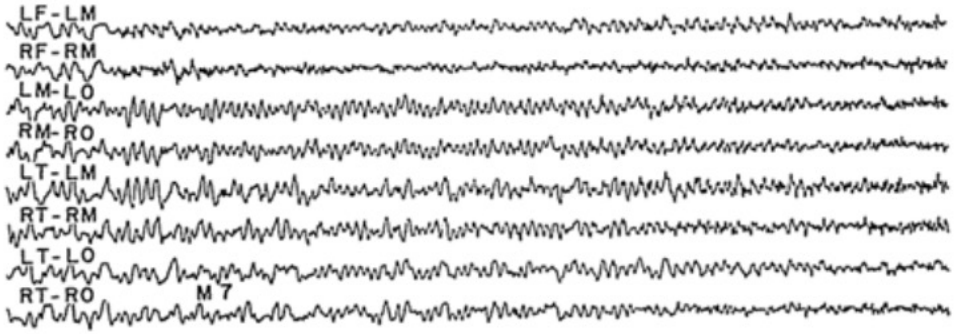
Data-Ink Ratio

- Ink = non-white data-pixels. (White or background pixels are not counted.) We want a high ratio of data presentation to the pixel/ink used.

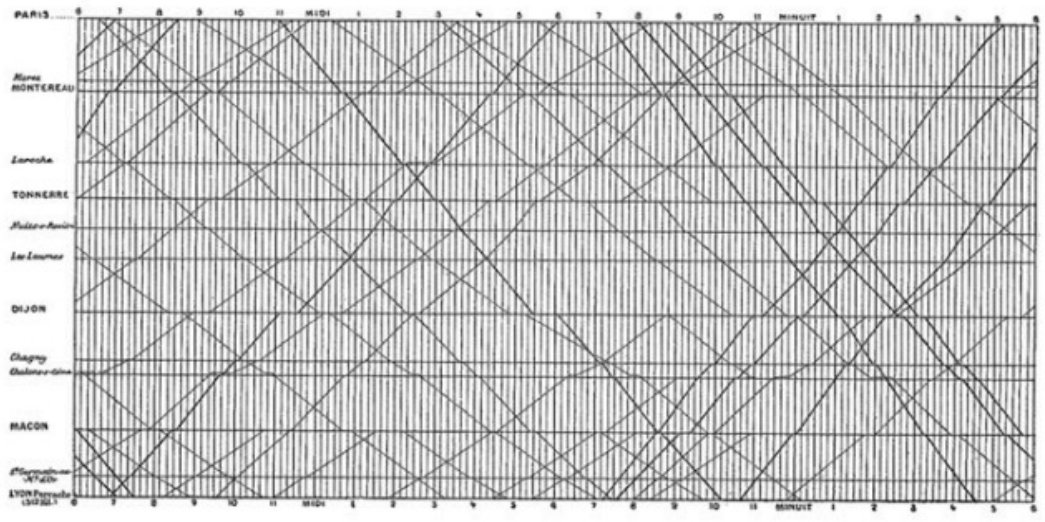
$$\text{Data-ink ratio} = \frac{\text{data-ink (pixels used directly for data)}}{\text{total ink used in graphic (total non-background pixels)}}$$

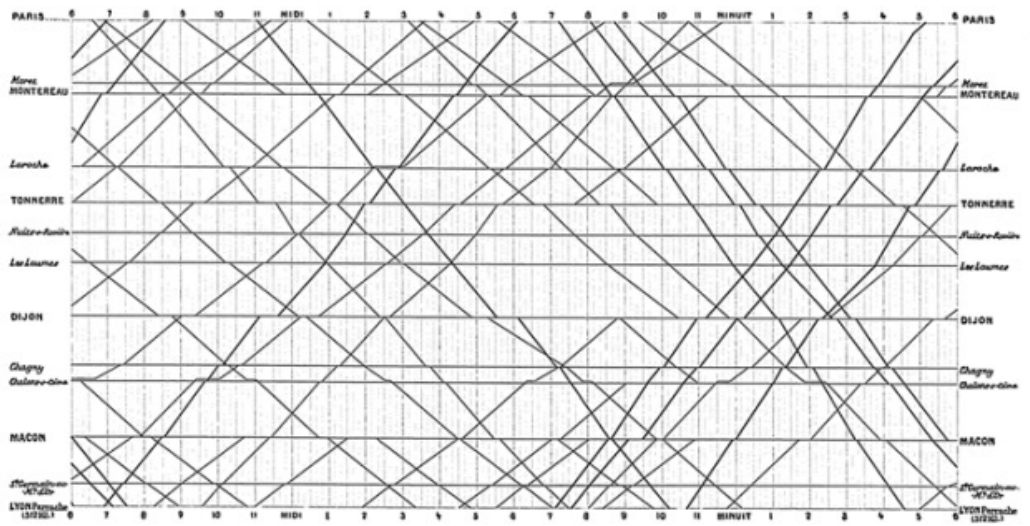
= proportion of ink devoted to the non-redundant display of information

= 1 - proportion of graphic that can be erased without loss of information



Electroencephalogram: Every pixel represents data, except labels



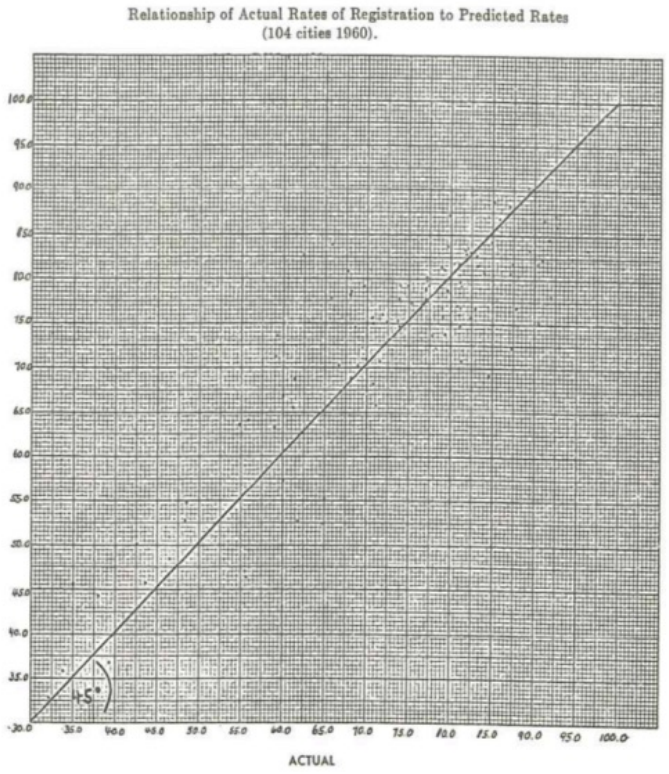


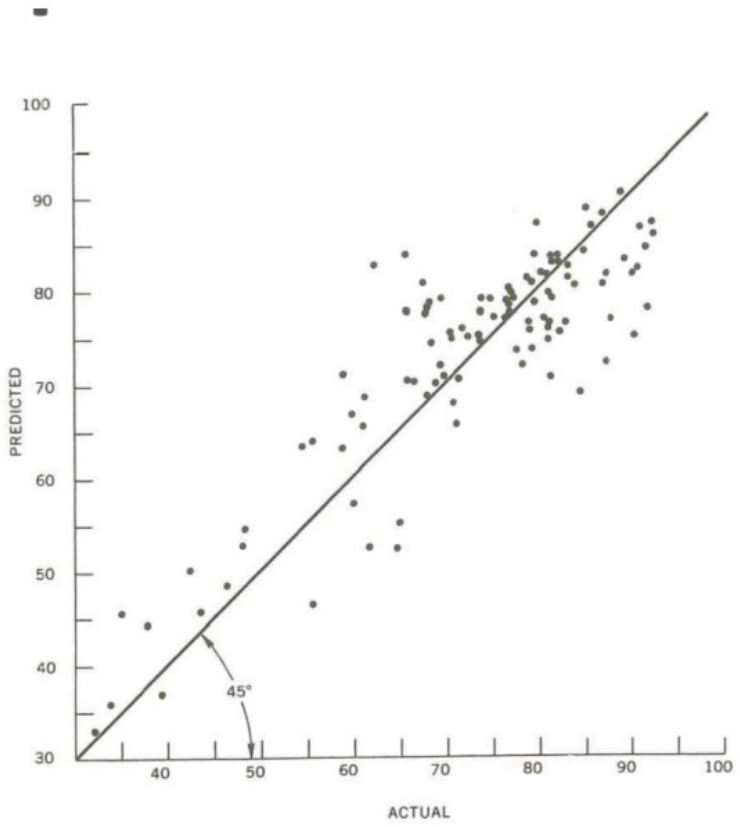
Train schedule in France (one of Tufte's book cover).

"Least Ink" Principles

- Above all else show data
- Maximize the data-ink ratio
- Erase non-data-ink
- Erase redundant data-ink
- Revise and edit

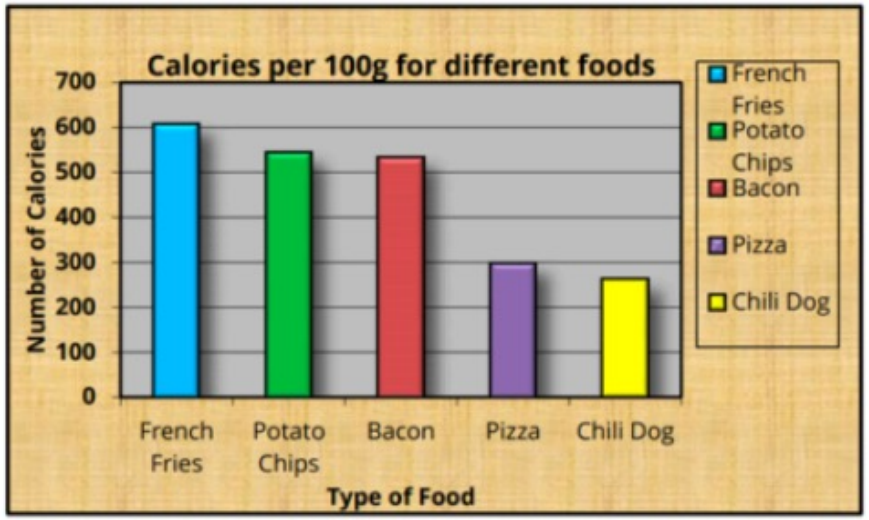
"Least Ink" (High Data-Ink Ratio)

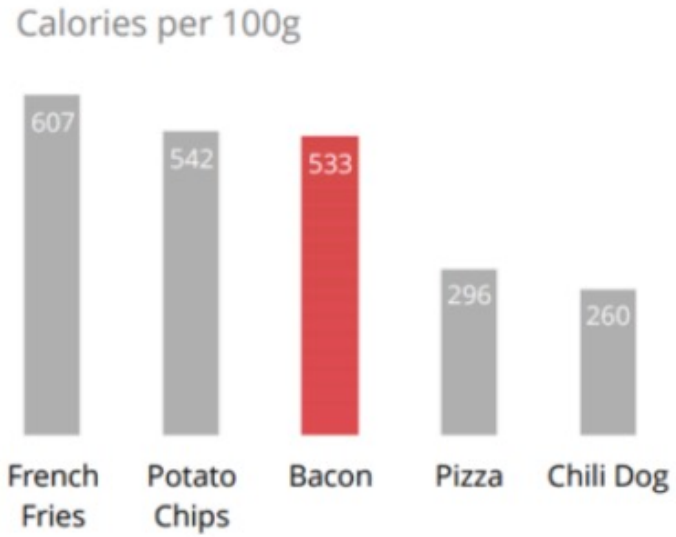




Avoid Chart Junk

- Definition: the excessive and unnecessary use of graphical effects in visualization for demonstrating the graphic ability of the design rather than displaying the data





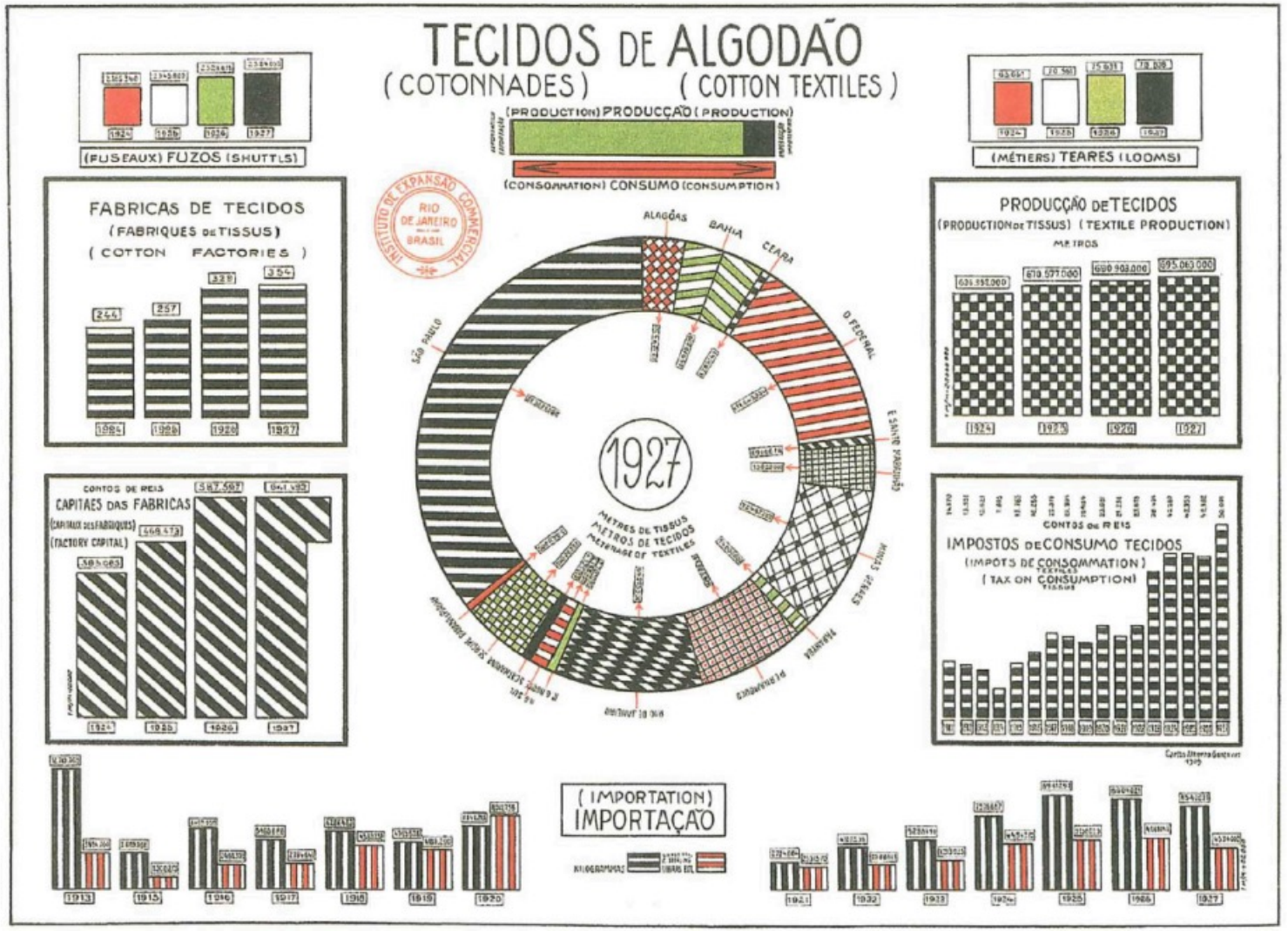
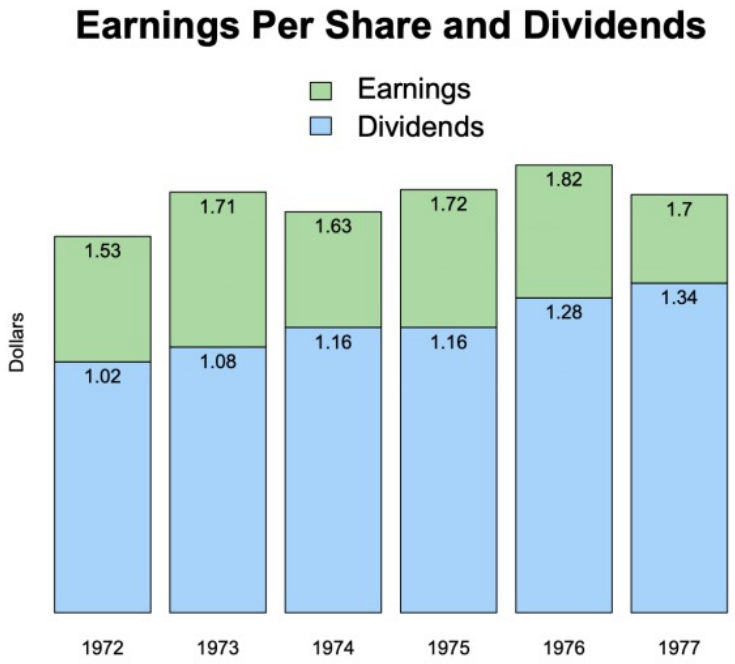
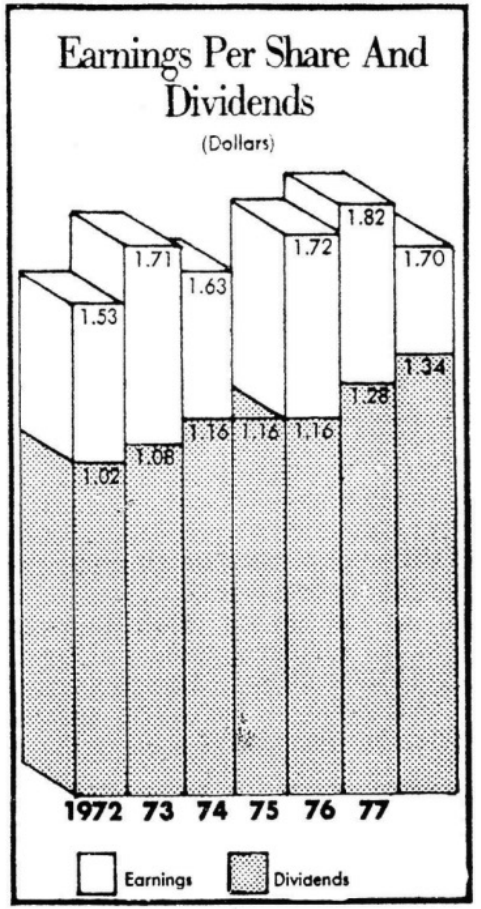


Chart Junk (1): Vibration

Too much texture and vibration in this graphic – a moire vibration

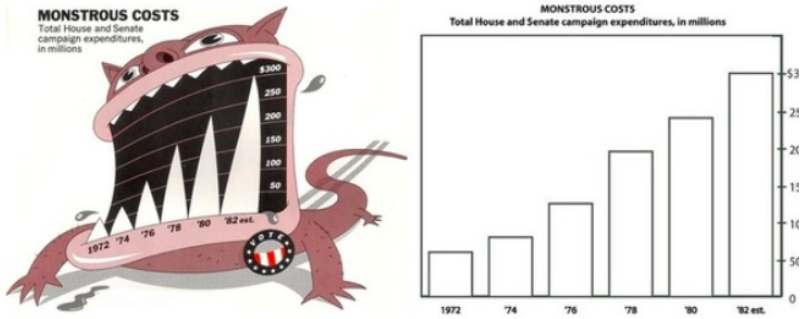
Chart Junk (2): Illusion

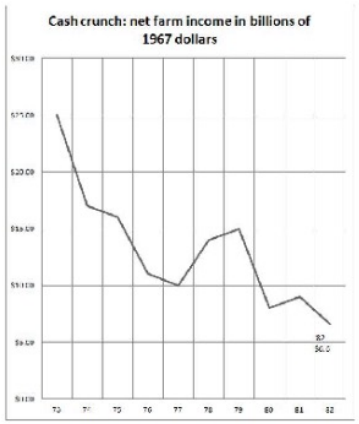


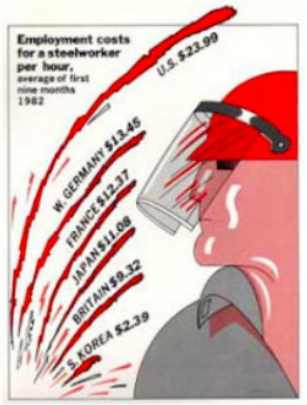


The extra dimension used in this graph has confused even the person who created it

Chart Junk (3): Visual Distortion







Rethink Axes and Markers: Ratio

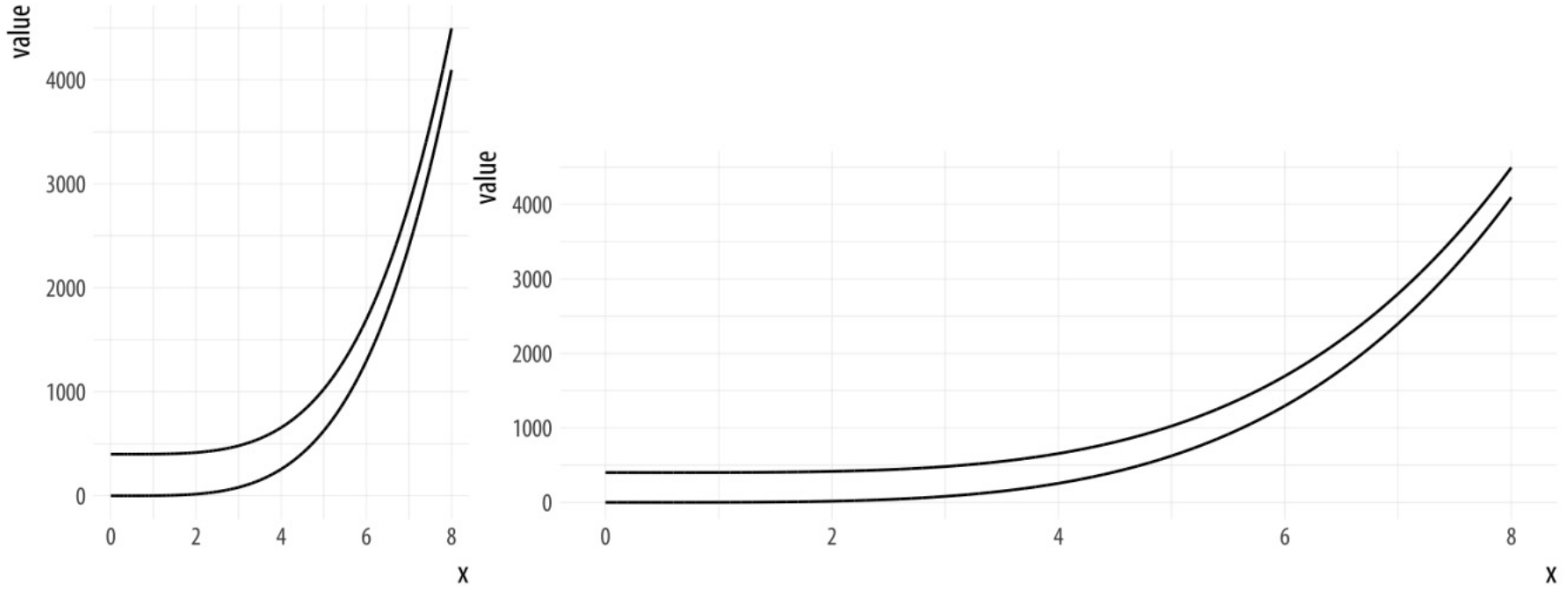


Figure: Aspect ratios affect our perception of rates of change. (After an example by William S. Cleveland.)

Rethink Axes and Markers: Avoid Dual Axis

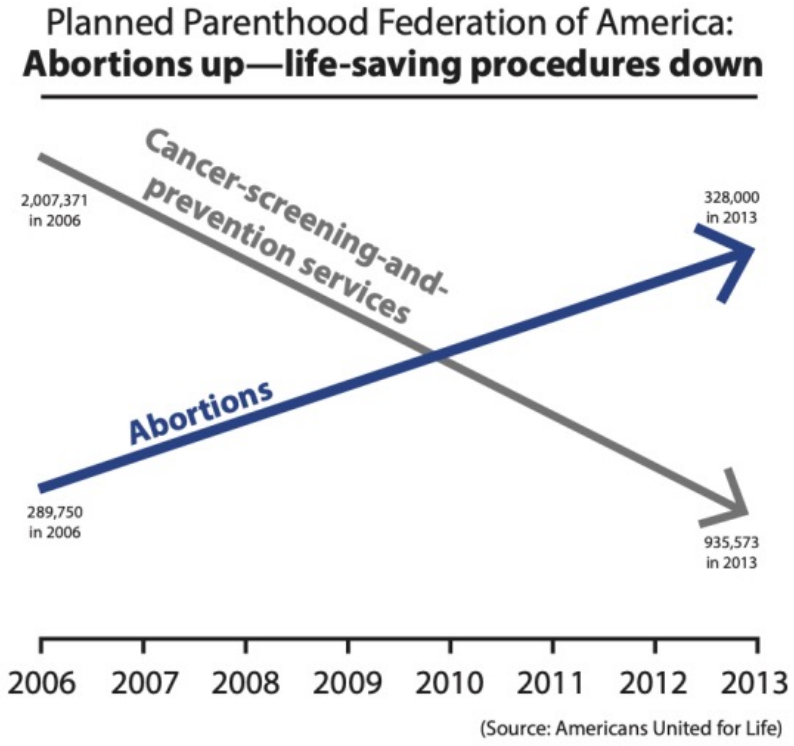
- Apples vs. oranges, which one is which?

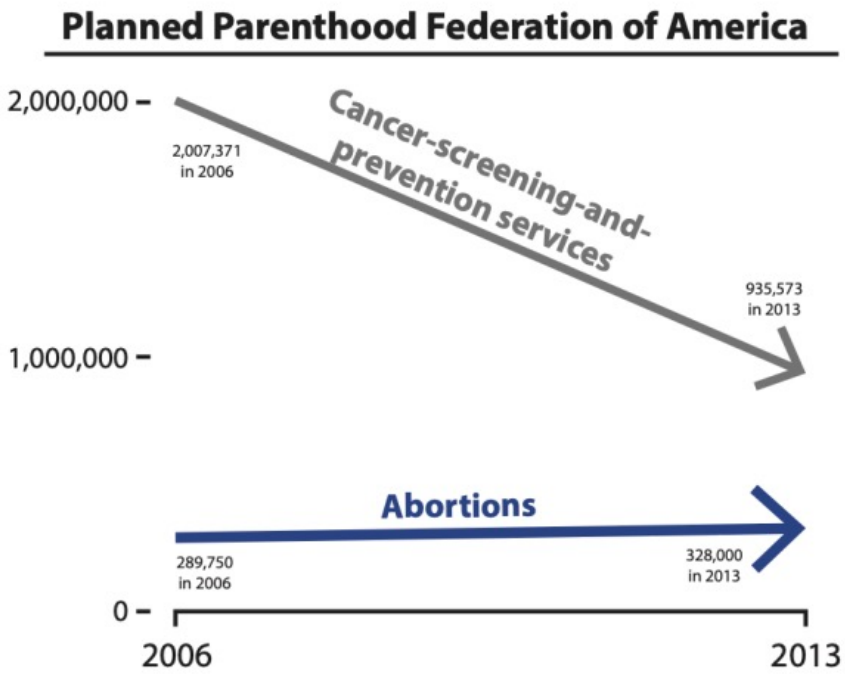
Original

Better

Rethink Axes and Markers: Normalization

- When comparing, using a common metric
- Percent is common, always stating percent to what



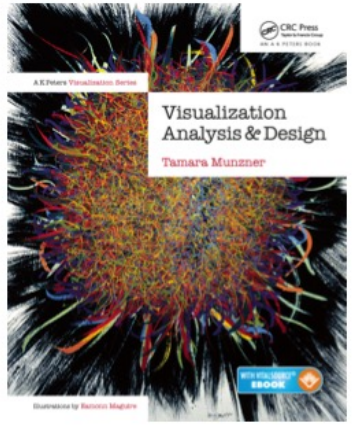


Aesthetics and Techniques

- **Have a properly chosen format and design**, e.g., a bivariate chart for bivariate data. Particular design will take work and multiple iterations/attempts to see what appeals.
- **Use words, numbers and drawing together**. Don't repeat the numbers in the words; words help understand the plan of the drawing if not obvious, labels.
- **Reflect a balance, a proportion, a sense of relevant scale**. All those examples of imbalance of area and quantity.
- **Display an accessible complexity of detail**. Add details as you can, but not to the point of obliterating information.
- **Often have a narrative quality**. There is a story to tell about the data, try to let it tell the story. Your narrative may not capture it all.
- **Are drawn in a professional manner**, with the technical details of production done with care (use available tools)
- **Avoid content-empty decoration, including chart junk**

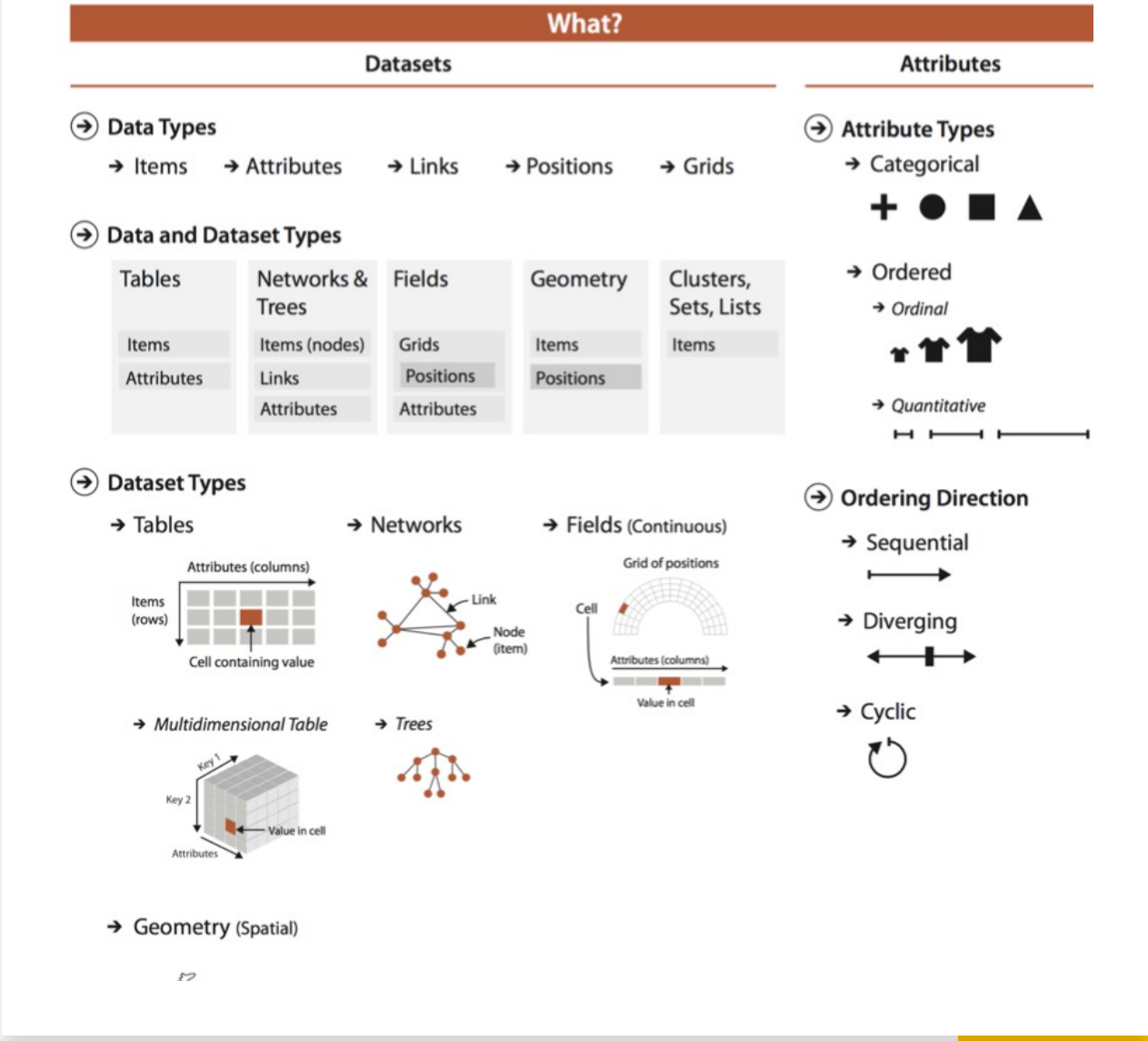

What?
Datasets
Attributes
Data Types
Items
Attributes
Links
Positions
Grids
Data and Dataset Types
Tables
Networks & Trees
Fields
Geometry
Clusters, Sets, Lists
Items
Attributes
Items (nodes)
Links
Attributes
Grids
Positions
Attributes
Items
Positions
Items
Dataset Types
Tables
Networks
Fields (Continuous)
Multidimensional Table
Trees
Geometry (Spatial)
Dataset Availability
Static
Dynamic
Attribute Types
Categorical
Ordered
Ordinal
Quantitative
Ordering Direction
Sequential
Diverging
Cyclic
Why?
Actions
Targets
Analyze
Consume
Discover
Present
Enjoy
Produce
Annotate
Record
Derive
Search
Target known
Target unknown
Location known
Lookup
Browse
Location unknown
Locate
Explore
Query
Identify
Compare
Summarize
All Data
Trends
Outliers
Features
Attributes
One
Many
Distribution
Dependency
Correlation
Similarity
Extremes
Network Data
Topology
Paths
Spatial Data
Shape
How?
Encode
Manipulate
Facet
Reduce
Arrange
Express
Separate
Order
Align
Use
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency ...
Change
Select
Navigate
Juxtapose
Partition
Superimpose
Filter
Aggregate
Embed

WHAT
WHY
HOW


CRC Press
Visualization Analysis & Design
Tamara Munzner

Data Abstraction

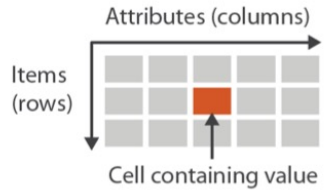
- Abstracting the data
- Types of datasets
  - Idiom design
- Types of attributes
  - Representations
- Ordering Direction
- Data aggregation
- Deriving new data



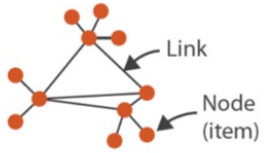
Data Types

Dataset Types

Tables

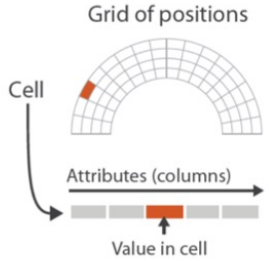


Networks



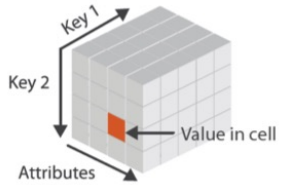
Spatial

Fields (Continuous)



Geometry (Spatial)

Multidimensional Table



Trees

Data File Format

Table

price	brand	model	title_status	mileage	color	lot
21500	ford	f-150	clean vehicle	76876	gray	167783132
7500	nissan	door	clean vehicle	58126	black	167598085
10700	chevrolet	trax	clean vehicle	44280	red	167792666
23100	dodge	charger	clean vehicle	12265	silver	167735423
33800	dodge	durango	clean vehicle	15003	black	167734879
4140	ford	door	clean vehicle	40747	white	167656519
13900	nissan	rogue	clean vehicle	38760	white	167762102
15700	ford	cutaway	clean vehicle	75862	white	167780452
1150	ford	door	salvage insurance	123349	red	167652717
26100	ford	f-150	clean vehicle	32149	white	167741409
14000	ford	fusion	clean vehicle	50513	white	167749355
22500	ford	door	clean vehicle	43646	black	167780692

Sample data from Kaggle's US Cars dataset

CSV

```
price,brand,model,title_status,mileage,color,lot
21500,ford,f-150,clean vehicle,76876,gray,16778313
7500,nissan,door,clean vehicle,58126,black,1675980
10700,chevrolet,trax,clean vehicle,44280,red,16779
23100,dodge,charger,clean vehicle,12265,silver,167
33800,dodge,durango,clean vehicle,15003,black,1677
4140,ford,door,clean vehicle,40747,white,167656519
13900,nissan,rogue,clean vehicle,38760,white,16776
15700,ford,cutaway,clean vehicle,75862,white,16778
1150,ford,door,salvage insurance,123349,red,167652
26100,ford,f-150,clean vehicle,32149,white,1677414
14000,ford,fusion,clean vehicle,50513,white,167749
22500,ford,door,clean vehicle,43646,black,16778069
```

JSON

```
[
  {
    "price": 21500,
    "brand": "ford",
    "model": "f-150",
    "title_status": "clean vehicle",
    "mileage": 76876,
    "color": "gray",
    "lot": 167783132
  },
  {
    "price": 7500,
    "brand": "nissan",
    "model": "door",
    "title_status": "clean vehicle",
    "mileage": 58126,
    "color": "black",
    "lot": 167598085
  },
  {
    "price": 10700,
    "brand": "chevrolet",
    "model": "trax",
    "title_status": "clean vehicle",
    "mileage": 44280,
    "color": "red",
    "lot": 167792666
  },
  {
    "price": 23100,
```

Attribute Types

- Attribute Types
  - Categorical

  - Ordered
    - *Ordinal*

    - *Quantitative*

- Ordering Direction
  - Sequential

  - Diverging

  - Cyclic

Data Types

Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

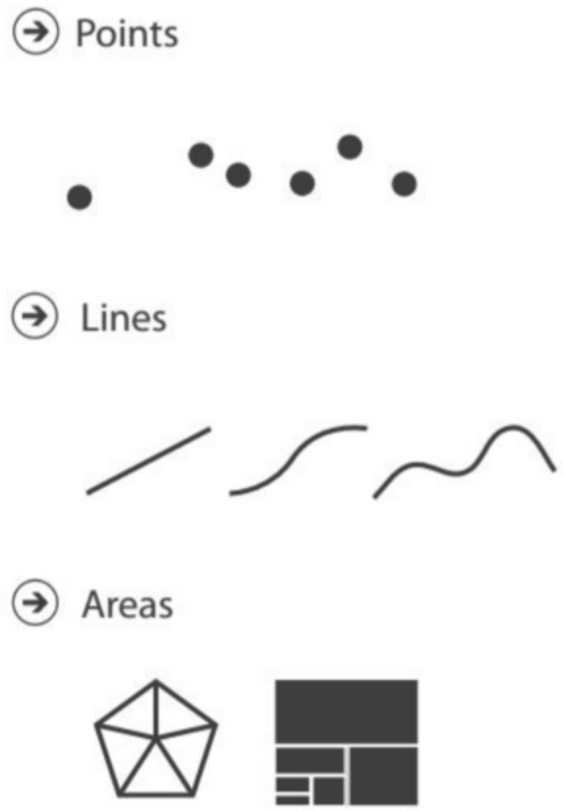
Data Types

→ Items → Attributes → Links → Positions → Grids

Dataset Availability

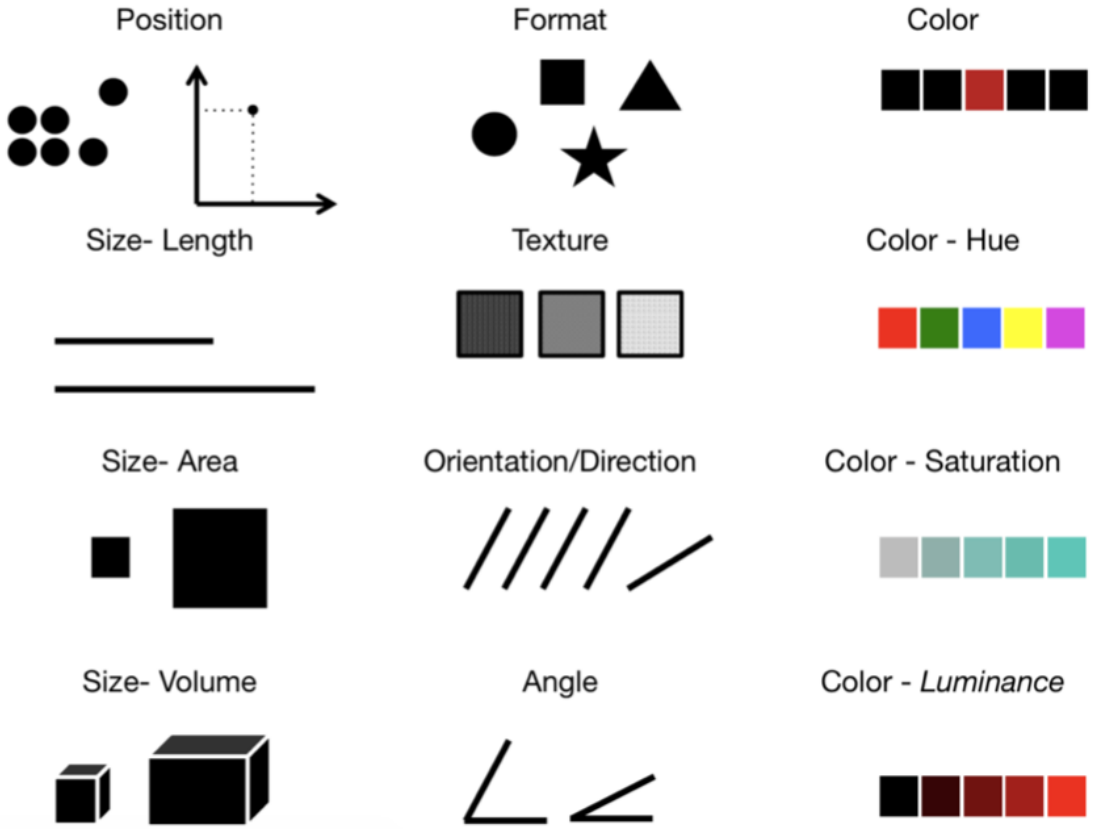
→ Static → Dynamic

Marks



Geometric Primitives

Visual Channels

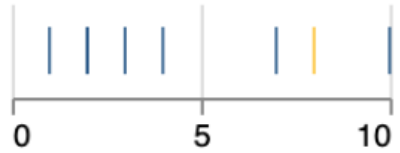


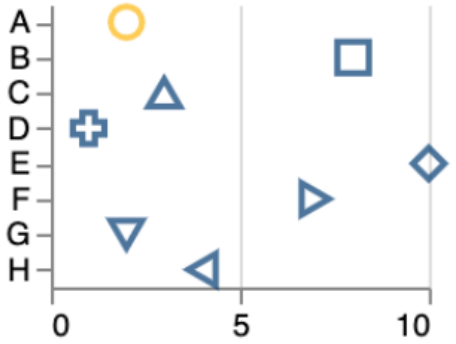
Control Appearance of Marks

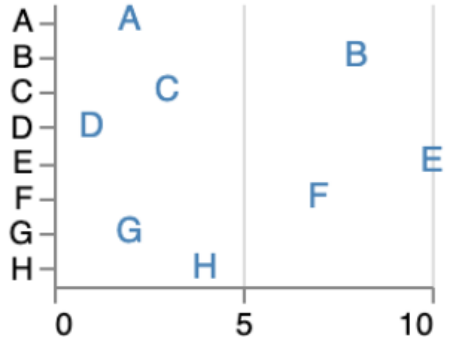
- Type and amount of info → human perception
- Qualitative (identity) vs. quantitative (magnitude)
- Discriminable and accuracy of perception

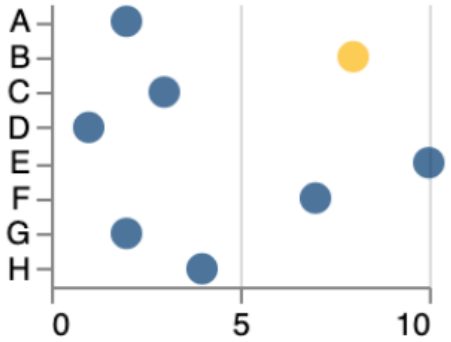
Mark: Points

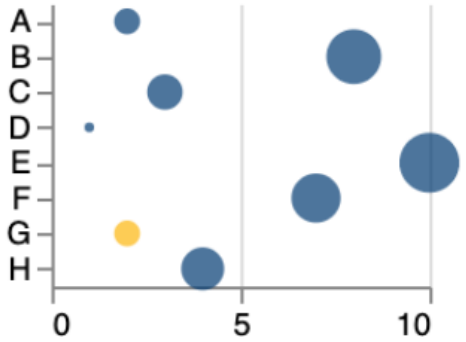
- Zero-dimensional
- Convey position only
- Can be size, shape-, color-coded

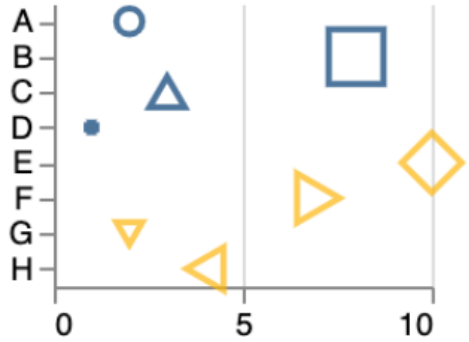






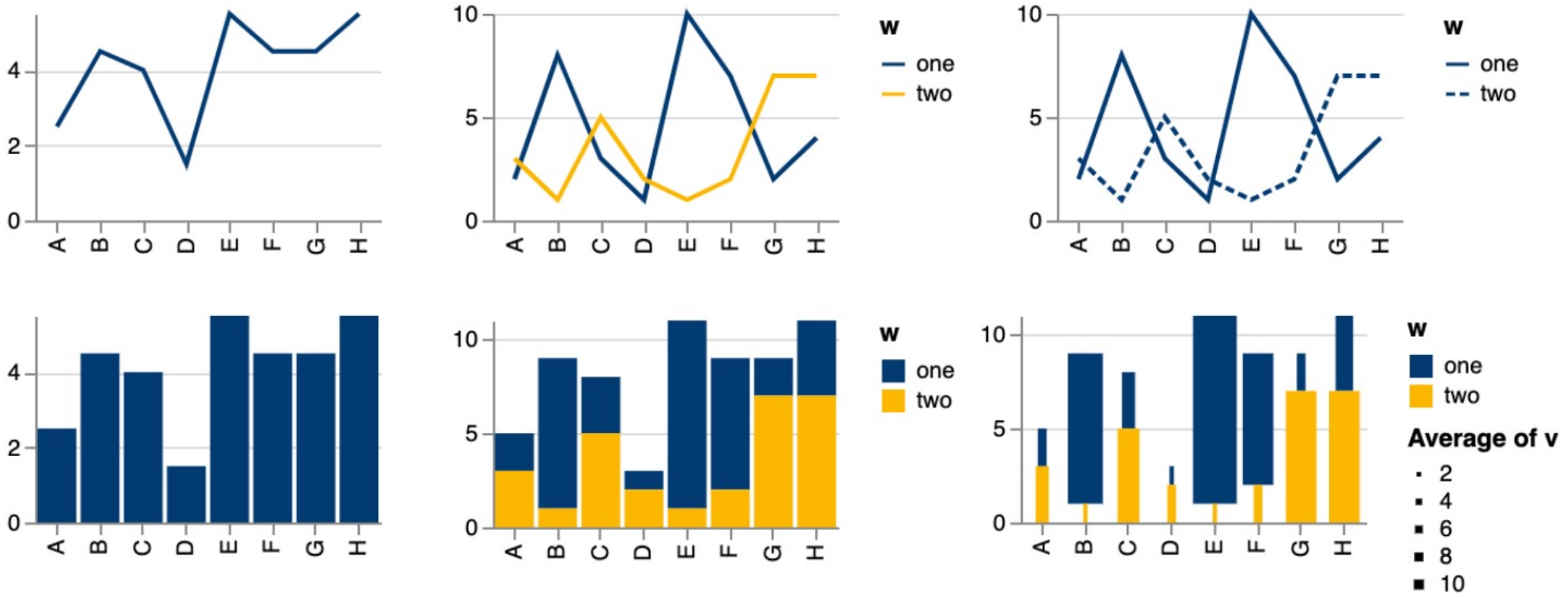






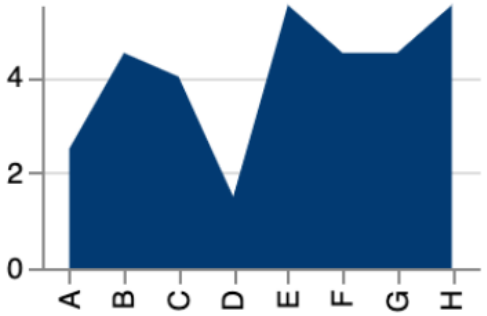
Mark: Lines

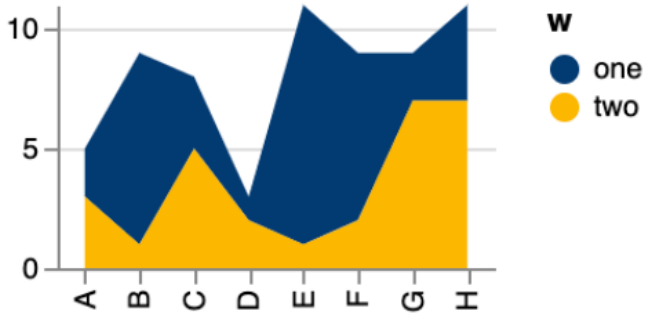
- One-dimensional
- Convey position and length
- Can be width-, color-, shape-coded

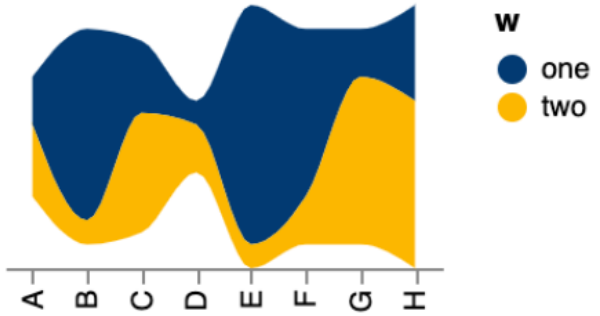


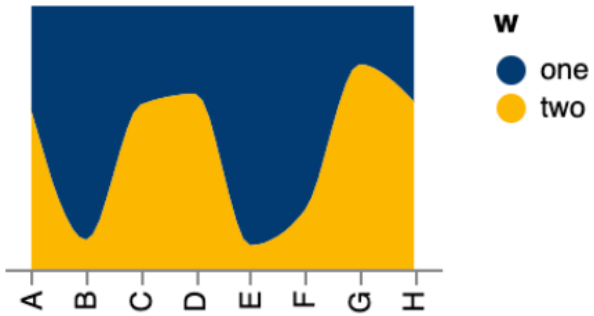
Mark: Areas

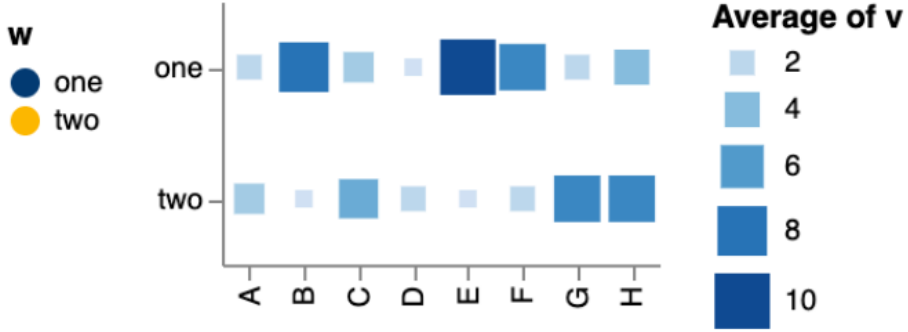
- Two-dimensional
- Are fully constrained
- Can be size- and color-coded



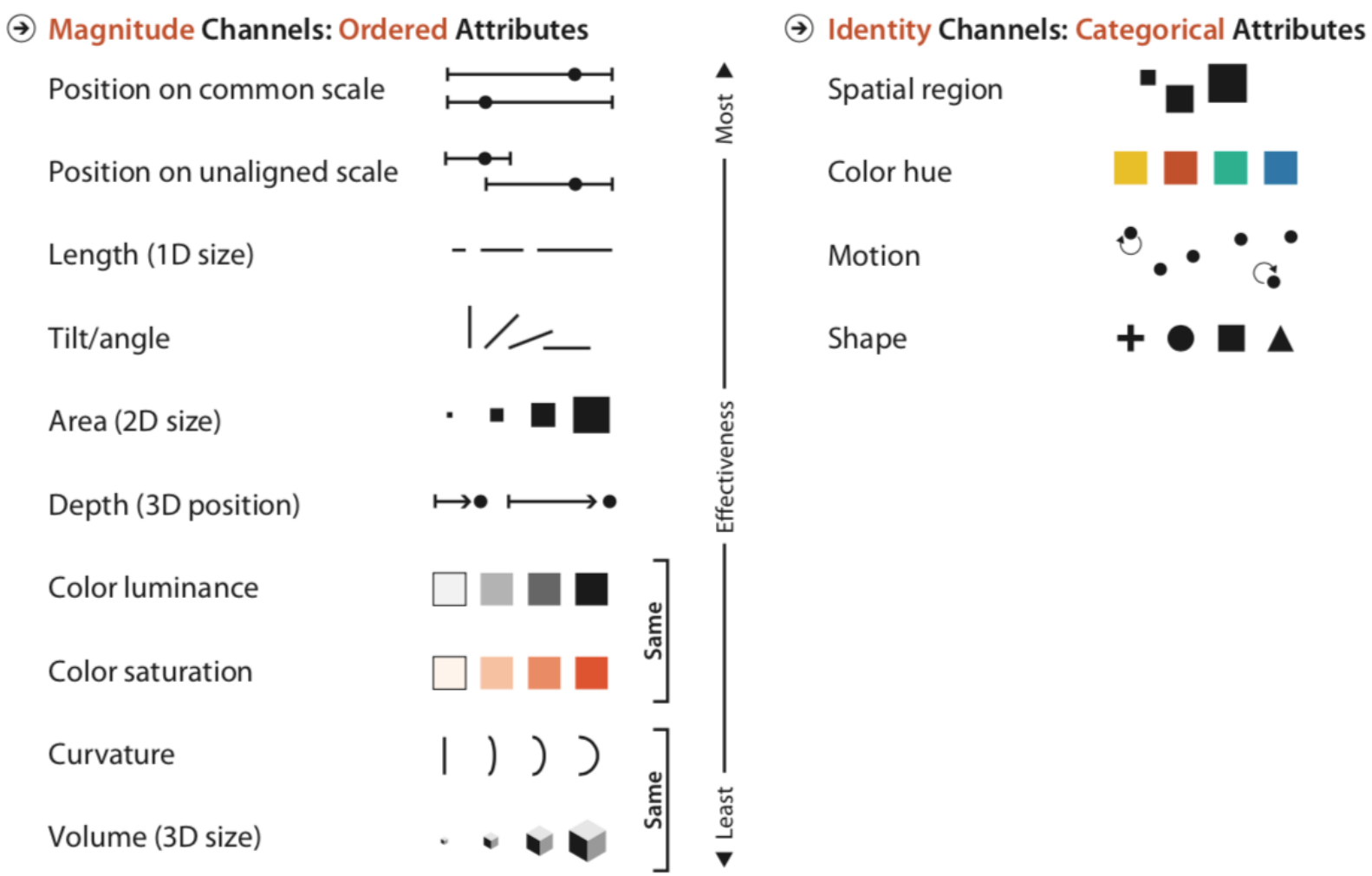




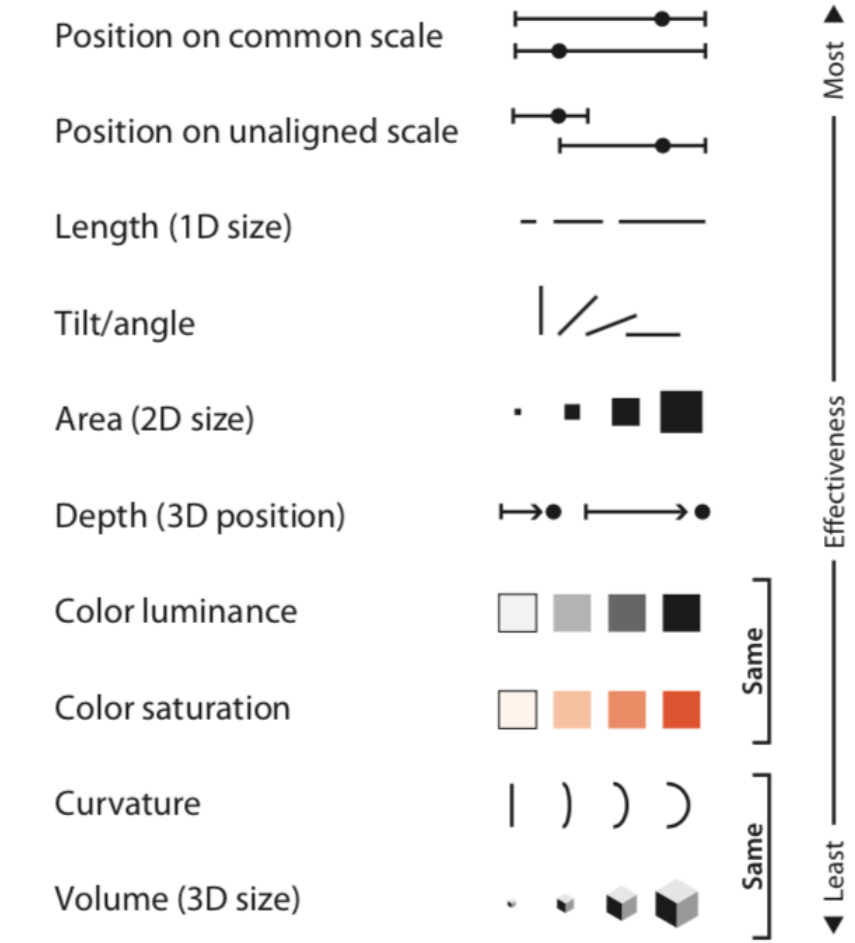




Channels: expressiveness types and effectiveness rankings



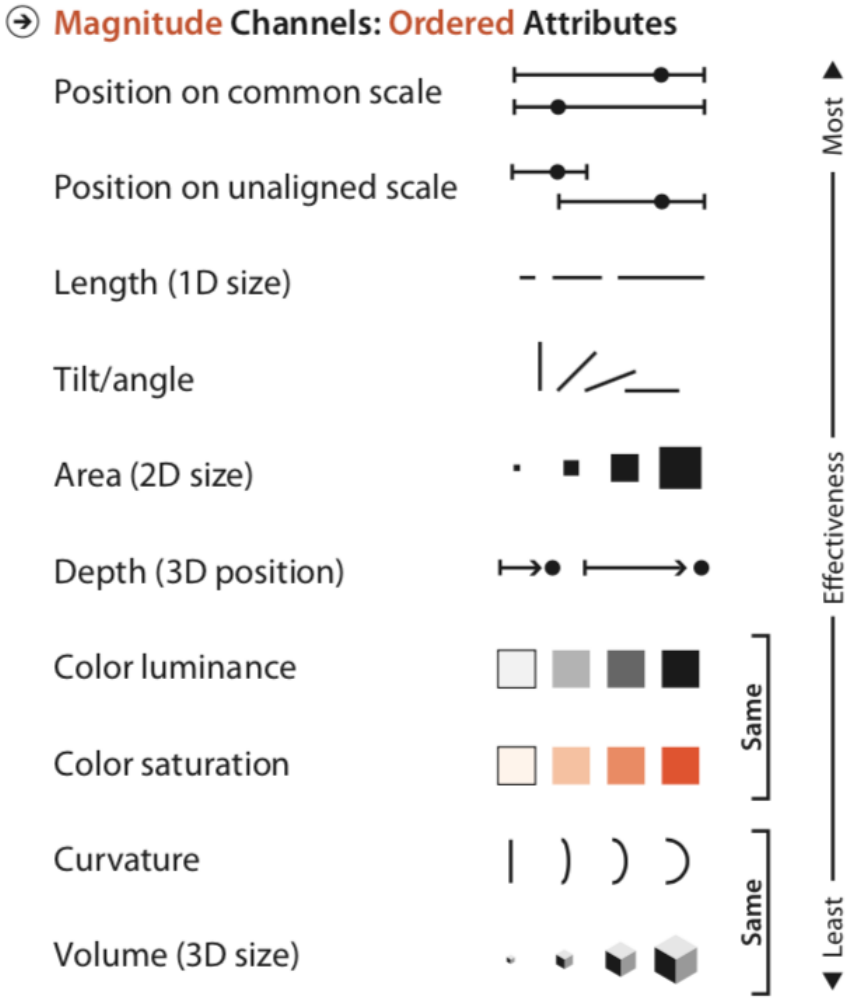
Channels: Matching Types

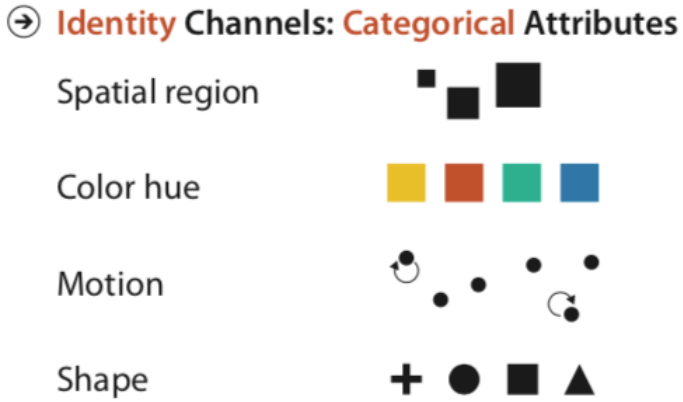


Expressiveness principle

- match channel and data characteristics

Channels: Rankings





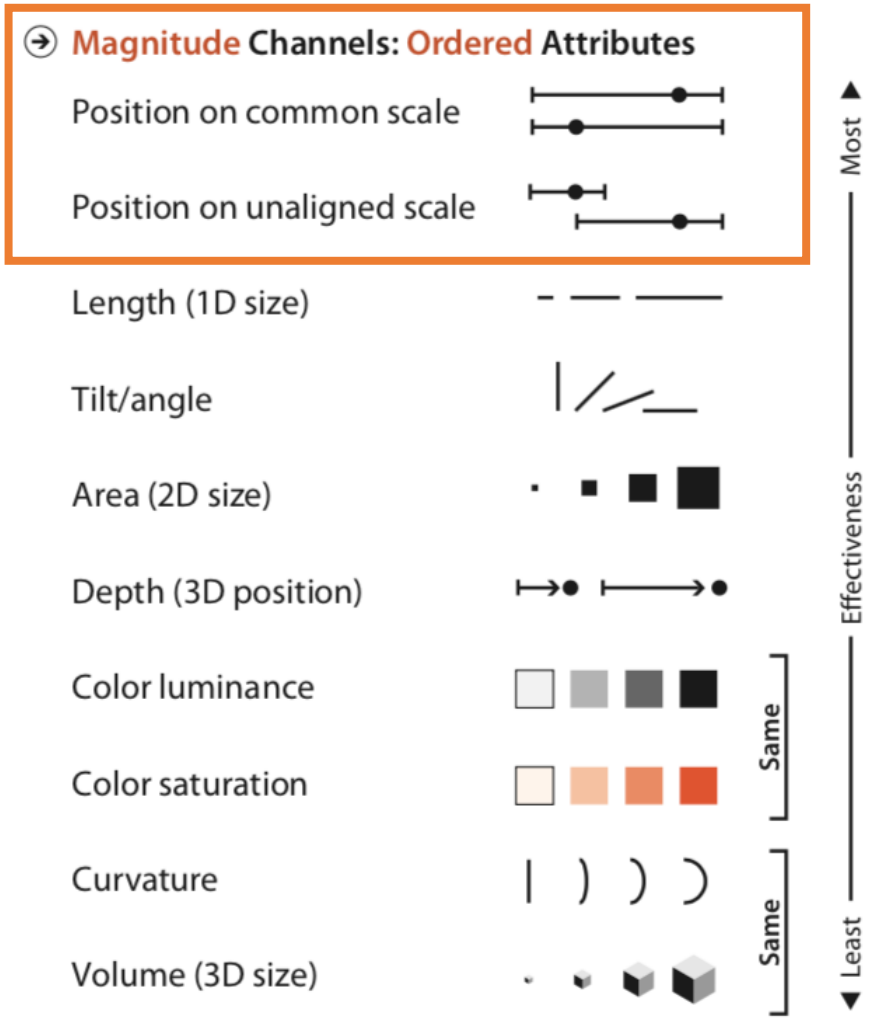
Expressiveness principle

- match channel and data characteristics

Effectiveness principle

- encode most important attributes with highest ranked channels

Channels: Expressiveness types and effectiveness rankings



Expressiveness principle

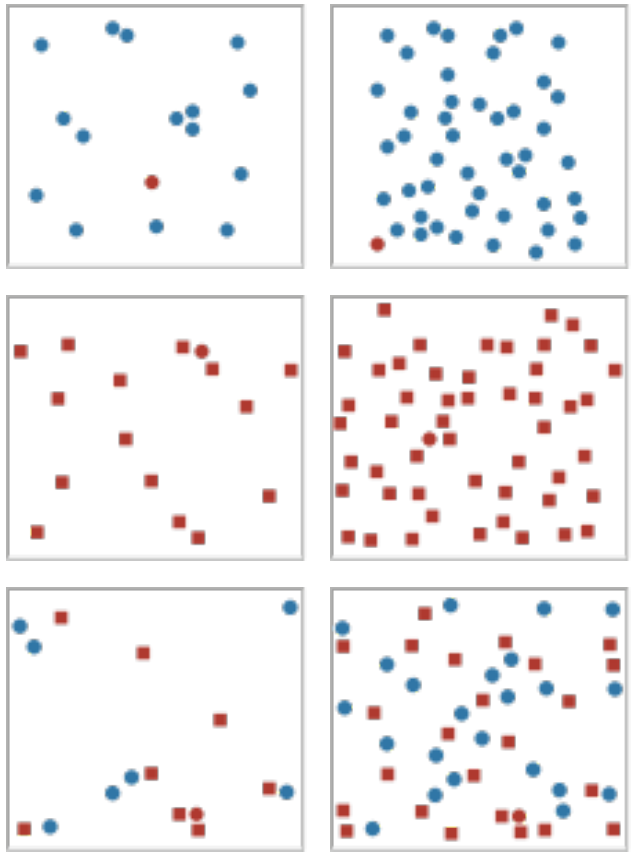
- match channel and data characteristics

Effectiveness principle

- encode most important attributes with highest ranked channels
- spatial position ranks high for both

Popout

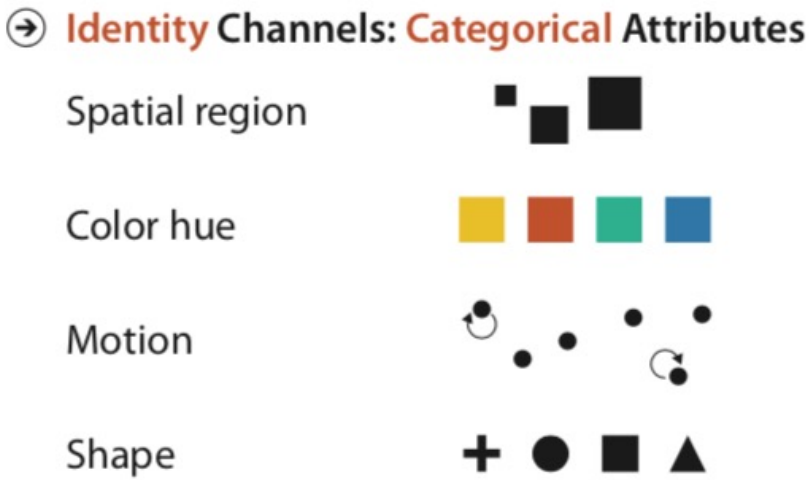
- Find the red dot
  - How long does it take?
- Parallel processing on many individual channels
  - Speed independent of distractor count
  - Speed depends on channel and amount of difference from distractors
- Serial search for (almost all) combinations
  - Speed depends on number of distractors



Grouping

- Containment
- Connection

- Proximity
  - Same spatial region
- Similarity
  - Same values as other categorical channels



Highlighting

	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
red	25.37	13.70	0.05	26.27	14.13	0.04	18.41	10.16	0.05	17.43	9.30	0.00
green	22.14	51.24	0.35	20.68	49.17	0.44	21.11	46.00	0.20	16.36	37.95	0.12
blue	13.17	3.71	74.89	15.38	5.20	86.83	11.55	3.37	65.53	9.96	3.44	56.14
gray	63.46	73.30	78.05	64.66	71.99	90.08	52.96	62.49	67.99	45.54	53.65	58.14
black	0.66	0.70	0.77	0.63	0.66	1.09	0.47	0.58	0.70	0.44	0.54	0.71

	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
red	25.37	13.70	0.05	26.27	14.13	0.04	18.41	10.16	0.05	17.43	9.30	0.00
green	22.14	51.24	0.35	20.68	49.17	0.44	21.11	46.00	0.20	16.36	37.95	0.12
blue	13.17	3.71	74.89	15.38	5.20	86.83	11.55	3.37	65.53	9.96	3.44	56.14
gray	63.46	73.30	78.05	64.66	71.99	90.08	52.96	62.49	67.99	45.54	53.65	58.14
black	0.66	0.70	0.77	0.63	0.66	1.09	0.47	0.58	0.70	0.44	0.54	0.71

Based on slide from Stone

Color in small regions is difficult to perceive, and bright colors in large areas appear bigger

Use bright, saturated colors for small regions, and use low saturation pastel colors for large regions and backgrounds.

Colors for Categories

Facts

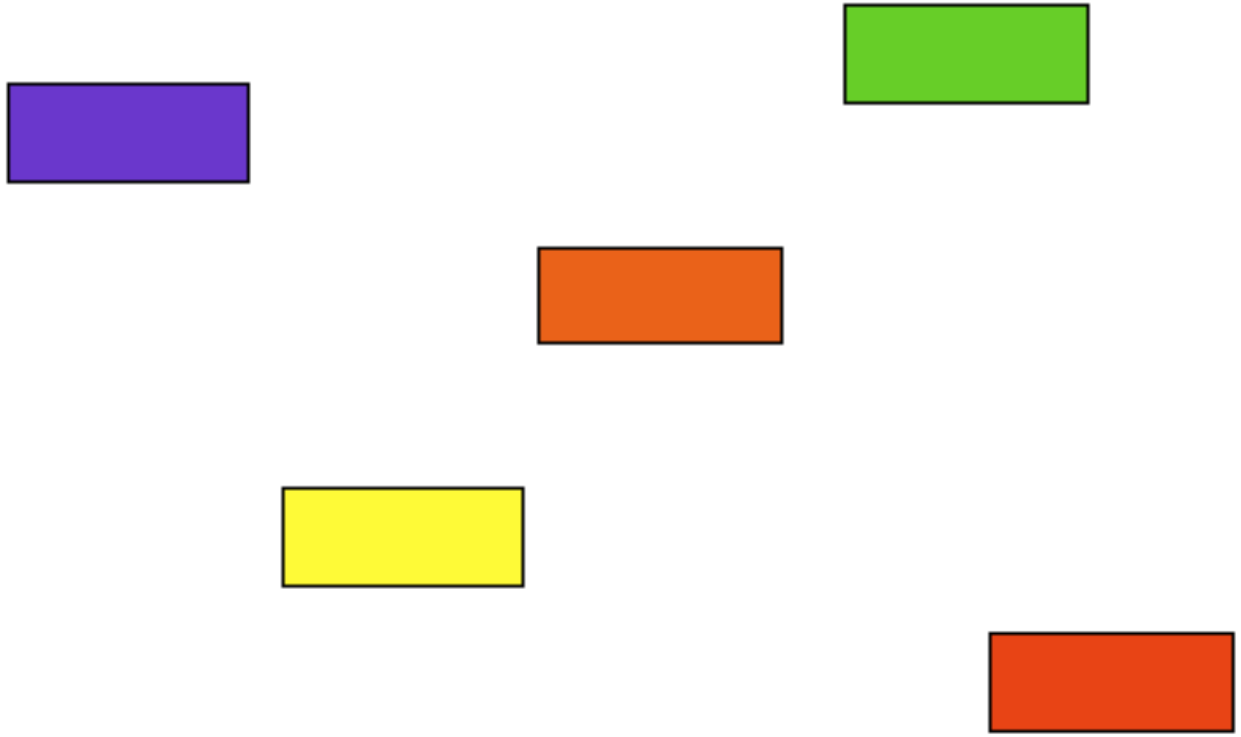
Only a small number of colors can be used effectively as nominal labels.

Keep the number of colors for nominal data to less than eight.

Use quiet medium grey backgrounds.

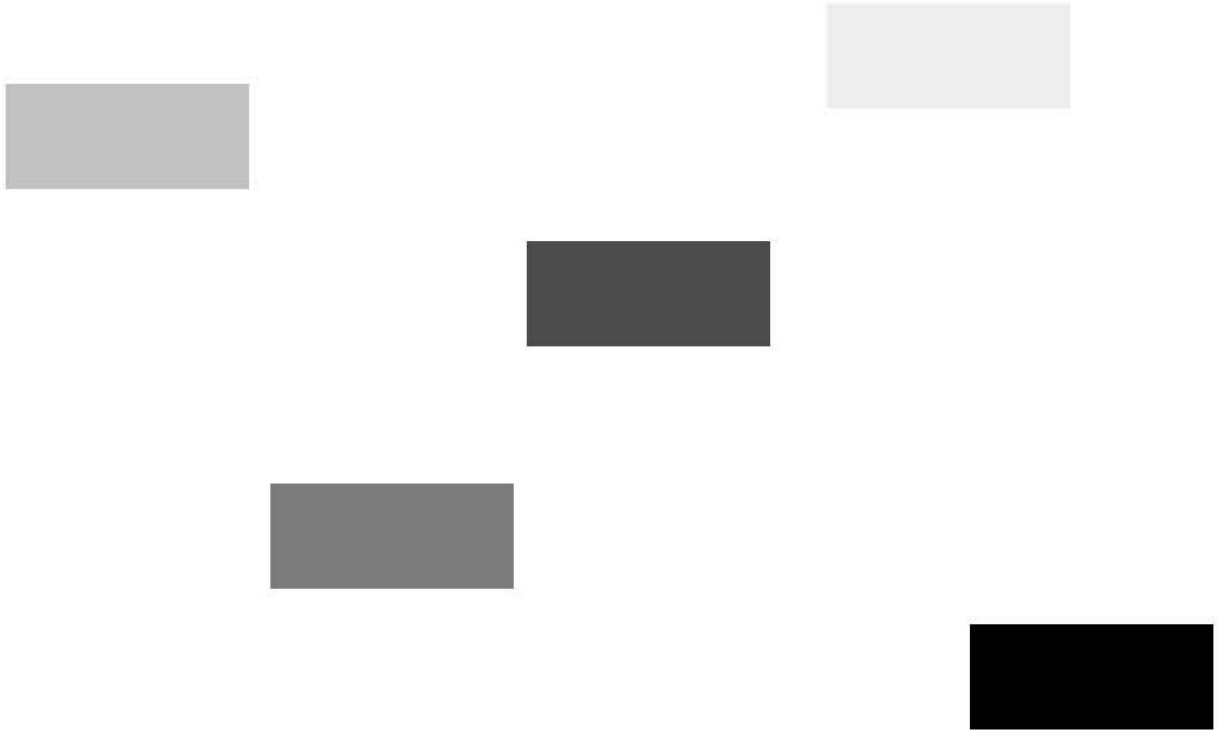
Color: Ordinal

Order These Colors



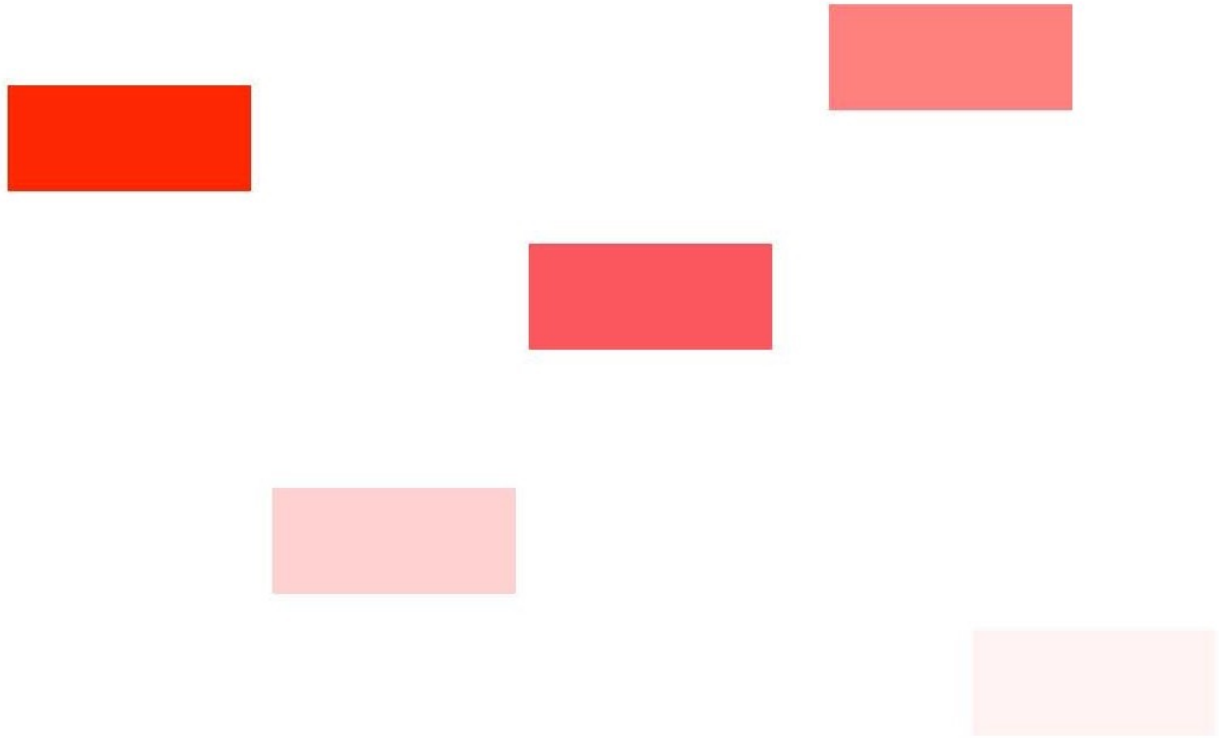
Based on slide from Stasko

Order These Colors



Based on slide from Stasko

Order These Colors

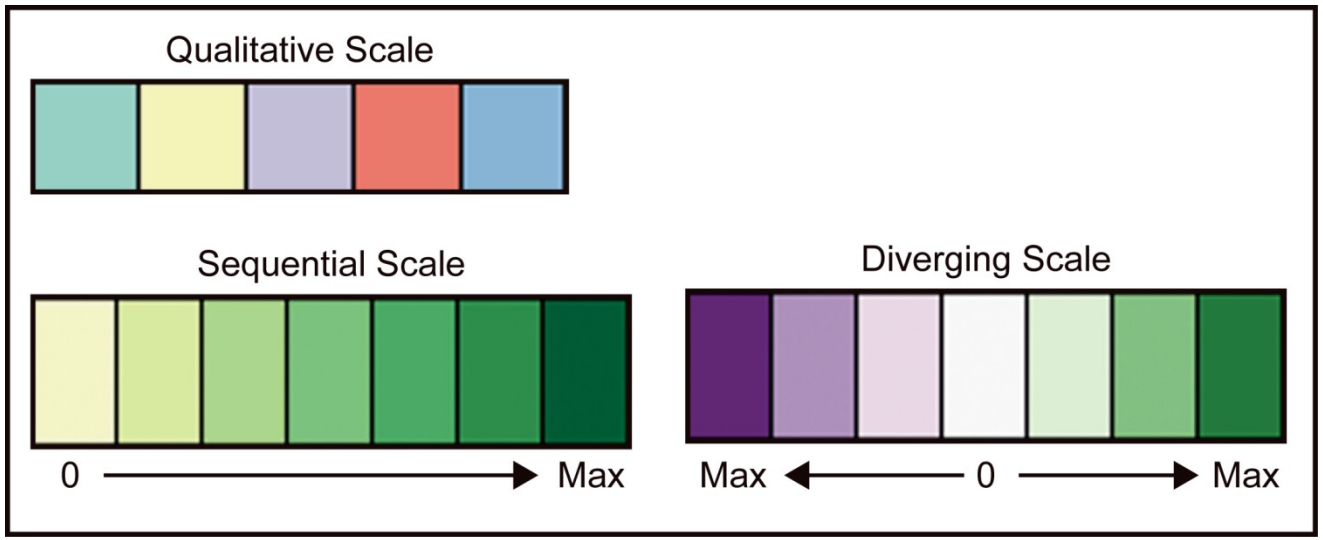


Based on slide from Stasko

Brewer Scales

Nominal

Ordinal



Cynthia Brewer, Color Use Guidelines for Data Representation

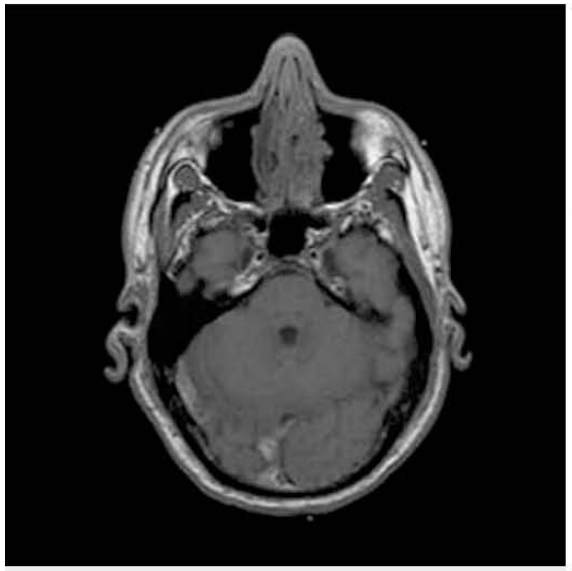
Facts

Lightness and saturation are effective for ordinal data because they have an implicit perceptual ordering

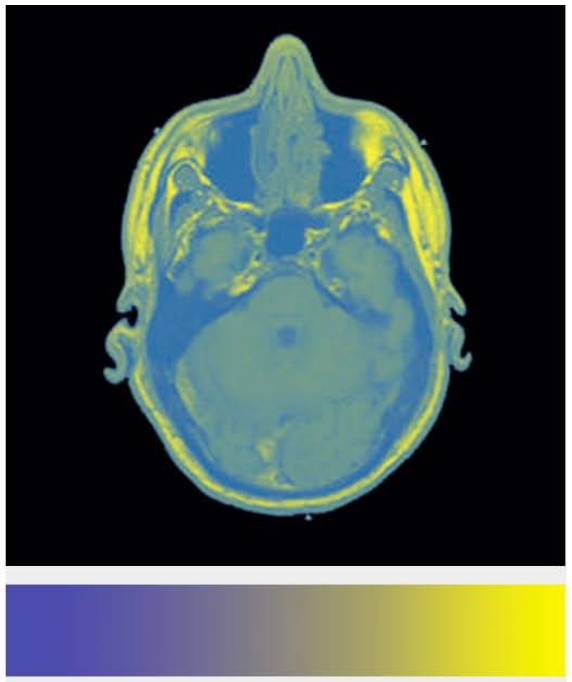
Show ordinal data with a discrete set of color values that change in lightness or saturation

Color: Quantitative

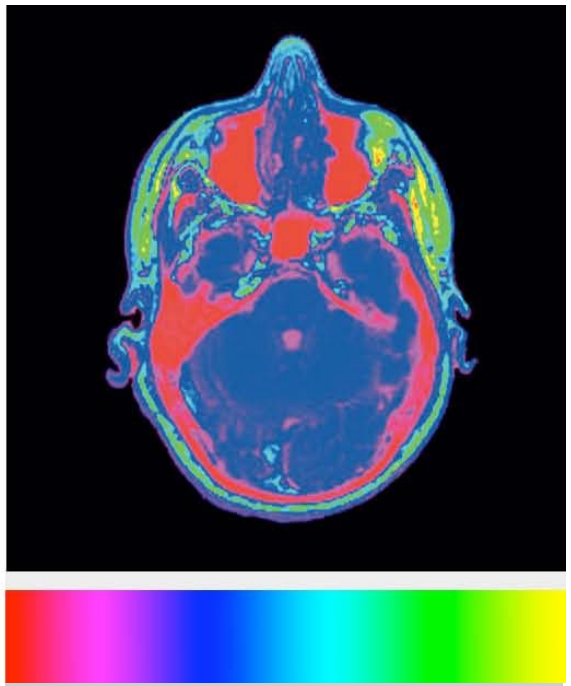
Colormaps



Lightness scale



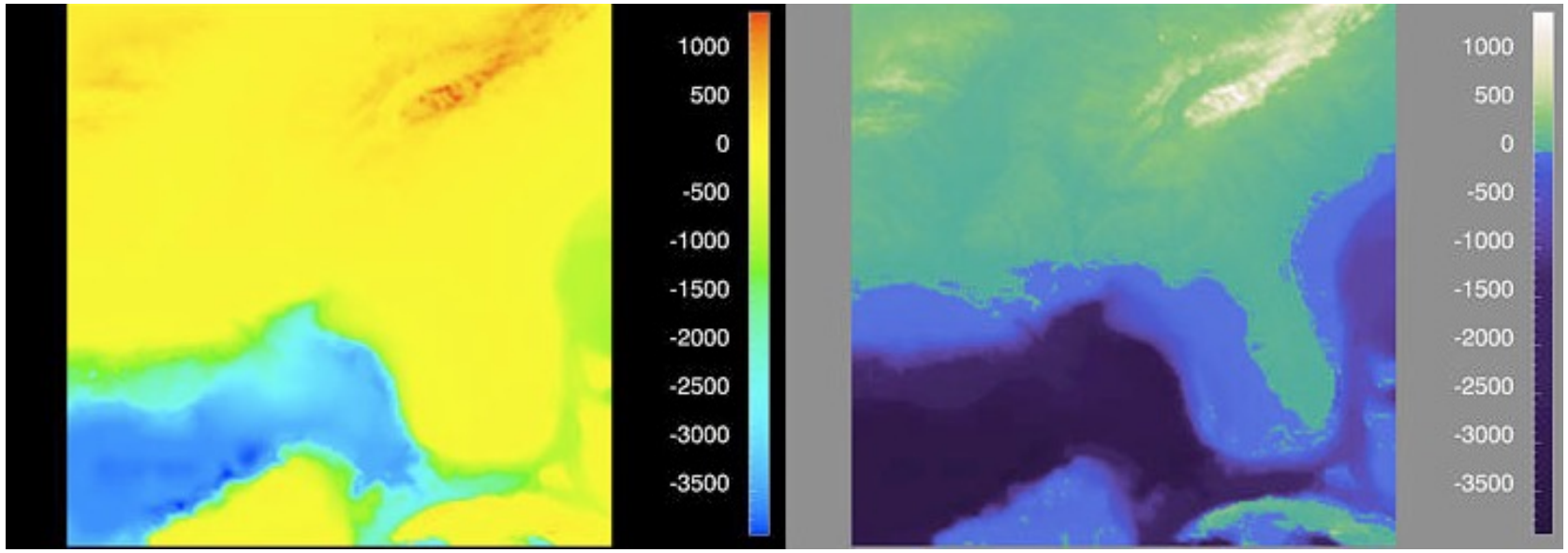
Lightness scale with hue and chroma variation



Hue scale with lightness variation

After slide from M. stone

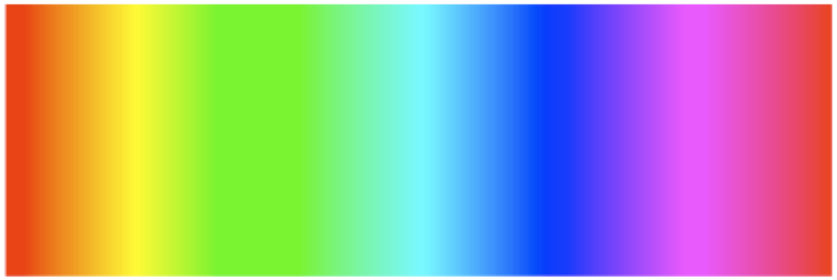
Rainbow Colormap



Rogowitz and Treinish, Why should engineers and scientists be worried about color?

Rainbow Colormap

- Hue is used to show ordinal data
- Not perceptually linear:
  - Equal steps in the continuous range are not perceived as equal steps
- Not good for colorblind people

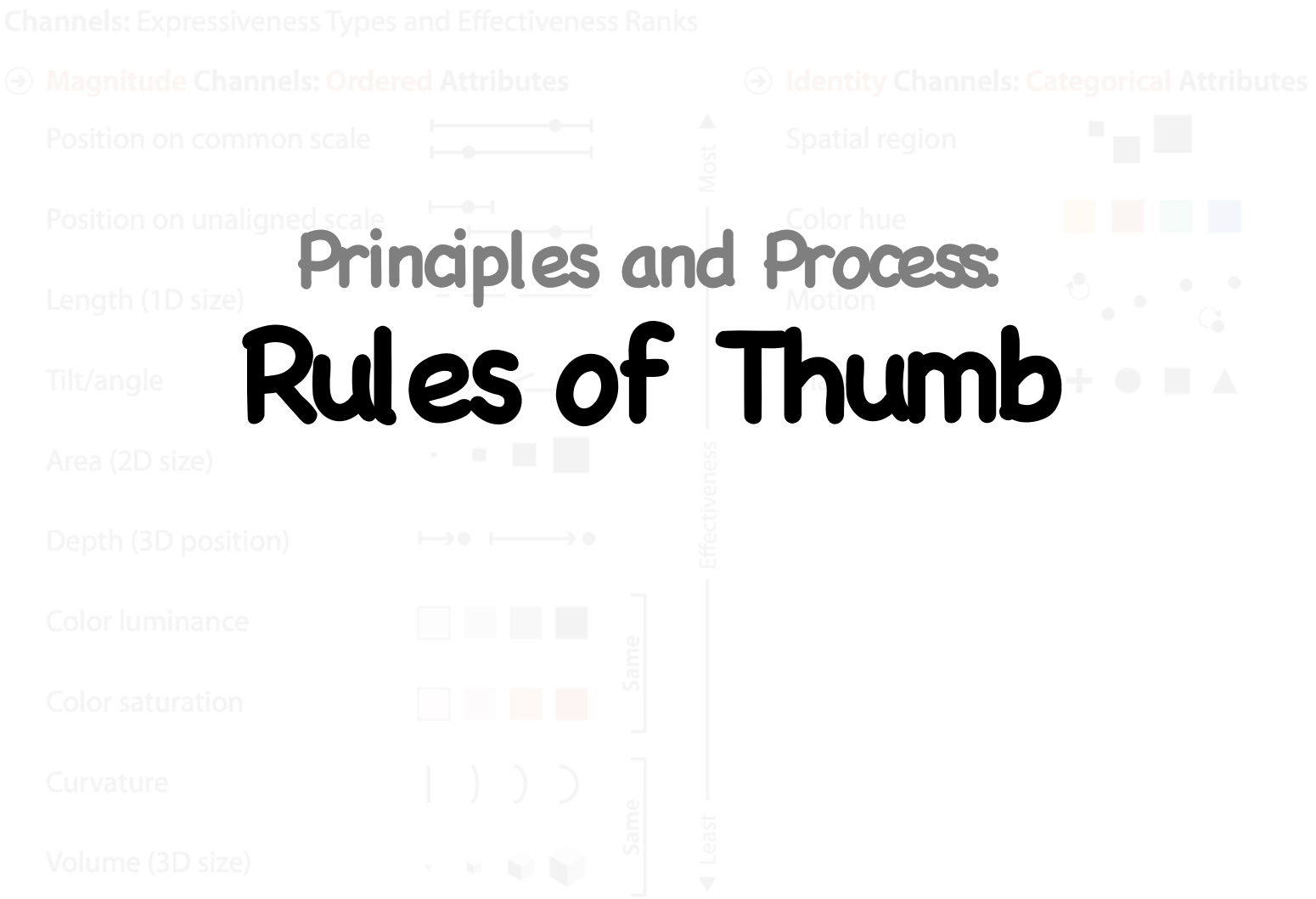


Facts

Quantitative data can be shown with a discrete or continuous colormap

Use colormaps with a limited hue palette and redundantly vary lightness and saturation.

Use discrete colormaps for accuracy.



Principles and Process:

Rules of Thumb

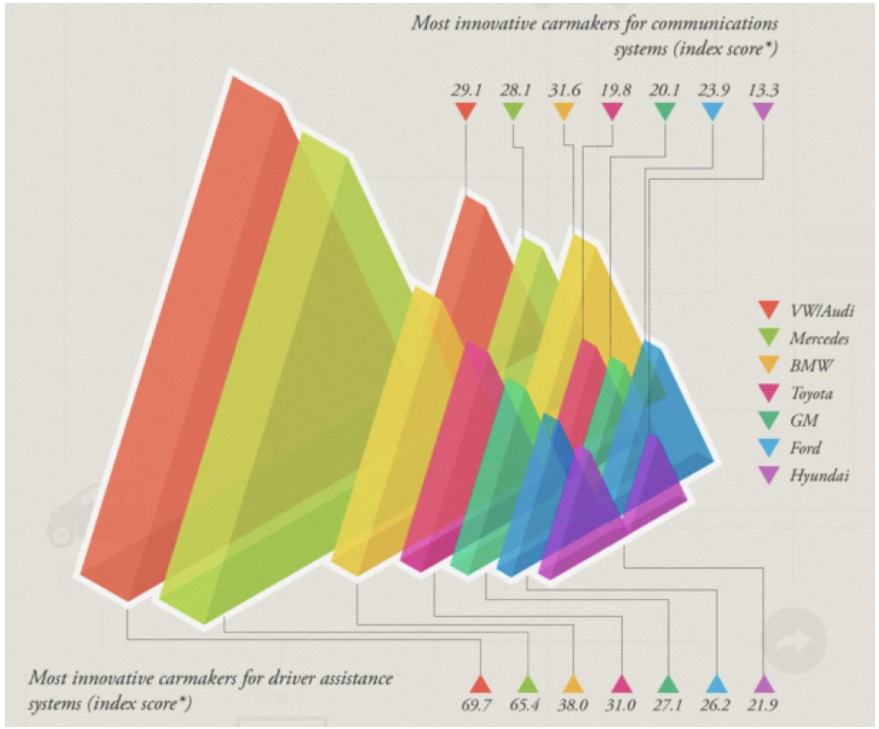
Outline

Courtesy of Tamara Munzner Department of Computer Science University of British Columbia

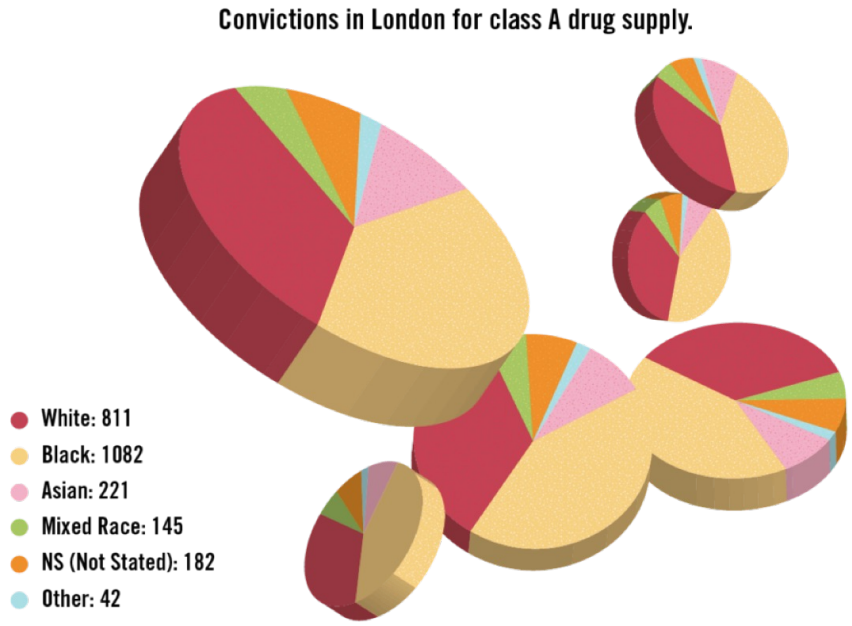
- No unjustified 3D
  - Power of the plane, dangers of depth
  - Occlusion hides information
  - Perspective distortion loses information
  - Tilted text isn't legible
- No unjustified 2D
- Eyes beat memory
- Resolution over immersion
- Visualization mantra
  - Overview first, zoom and filter, details on demand
- Function first, form next
- Responsiveness is required
- (Get it right in black and white)

https://john-guerra.github.io/MIDS_W209_Information_Visualization_Slides/05_Rules_of_thumb/index.html#/

(1) No Unjustified 3D

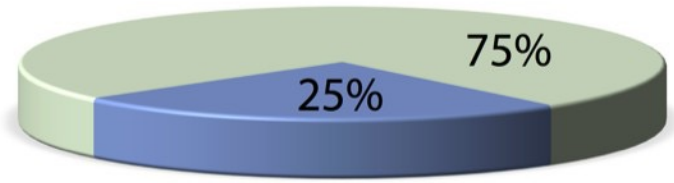


http://viz.wtf/post/137826497077/eye-popping-3d-triangles

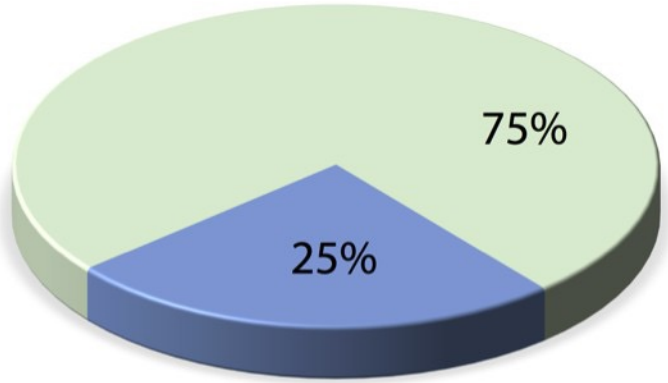


http://viz.wtf/post/139002022202/designer-drugs-ht-ducqn

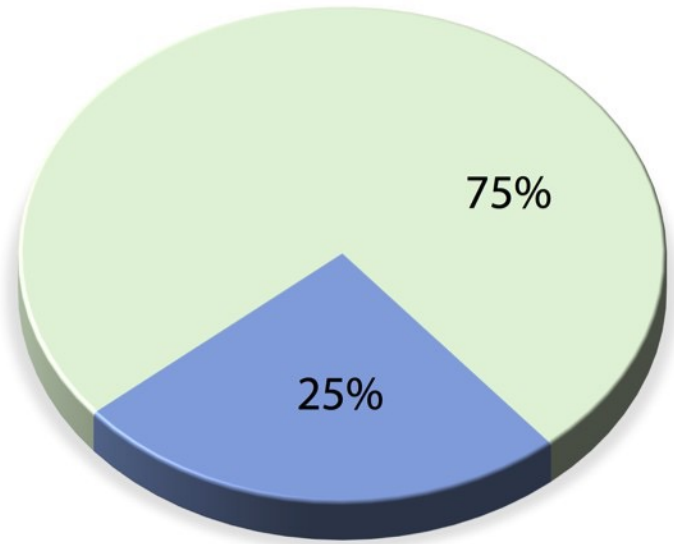
a



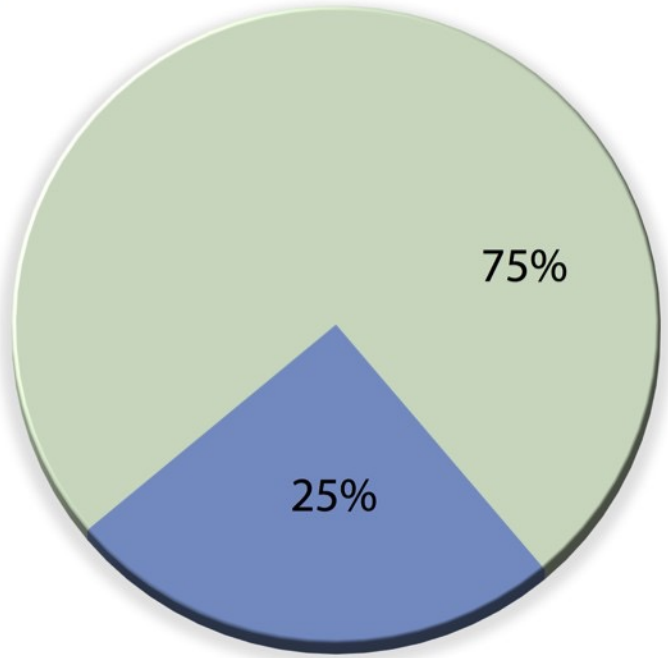
b



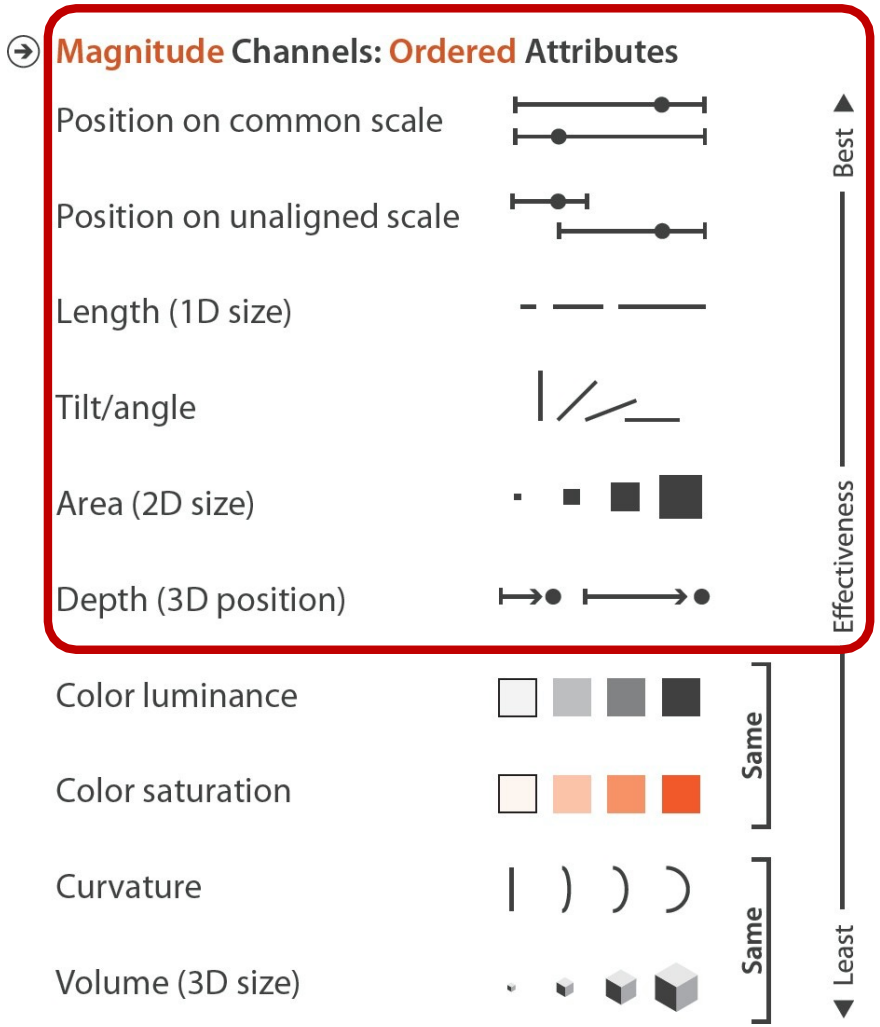
c



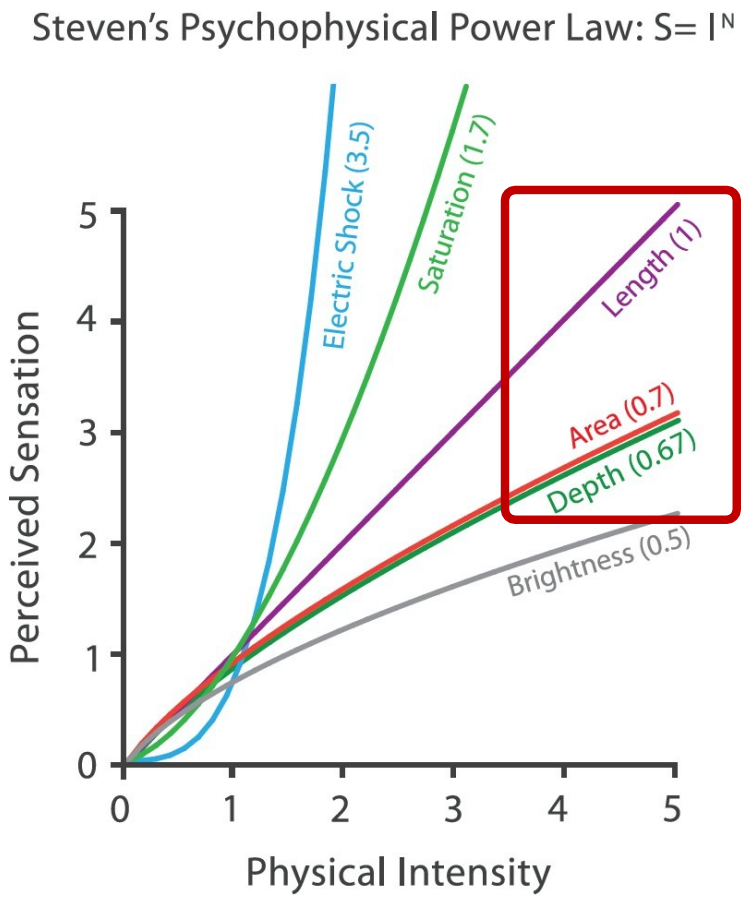
d



No Unjustified 3D: Power of the Plane

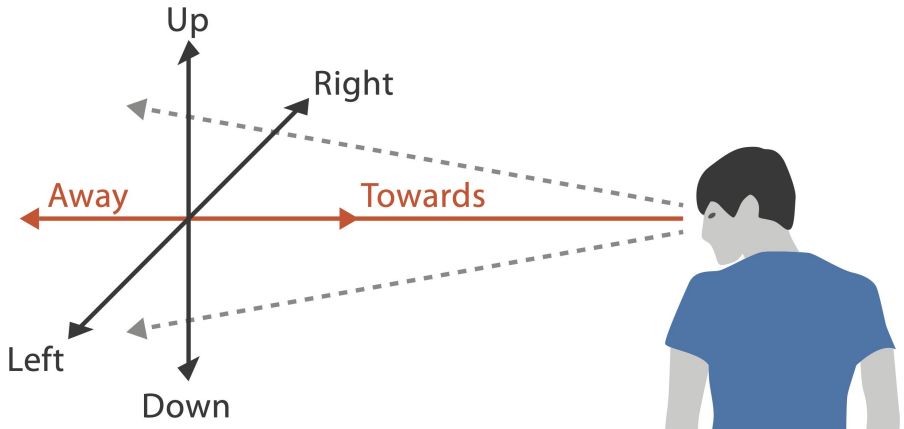


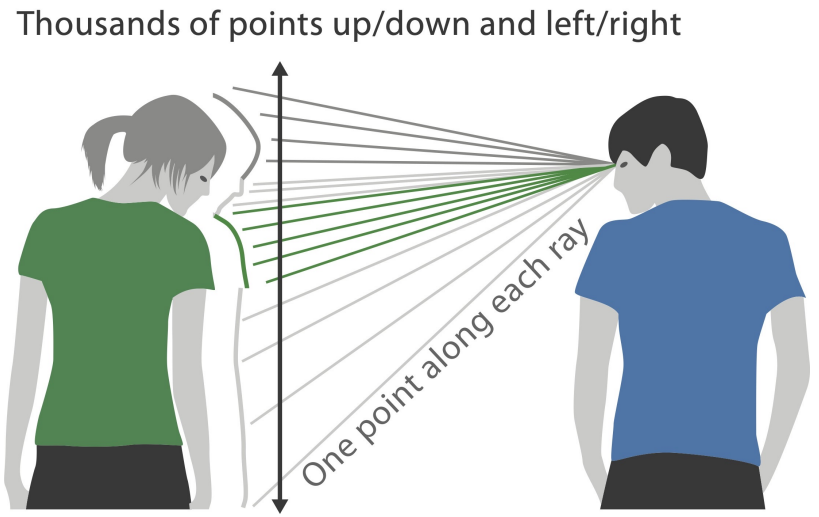
Planar Spatial Position >> Depth



No Unjustified 3D: Danger of Depth

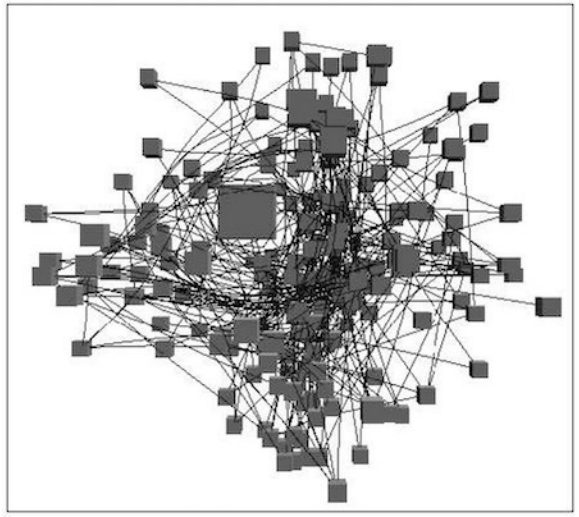
- We don't really live in 3D, We **see** in 2.05D
  - Acquire more info on image plane quickly from eye movements
  - Acquire more info for depth slower, from head/body motion



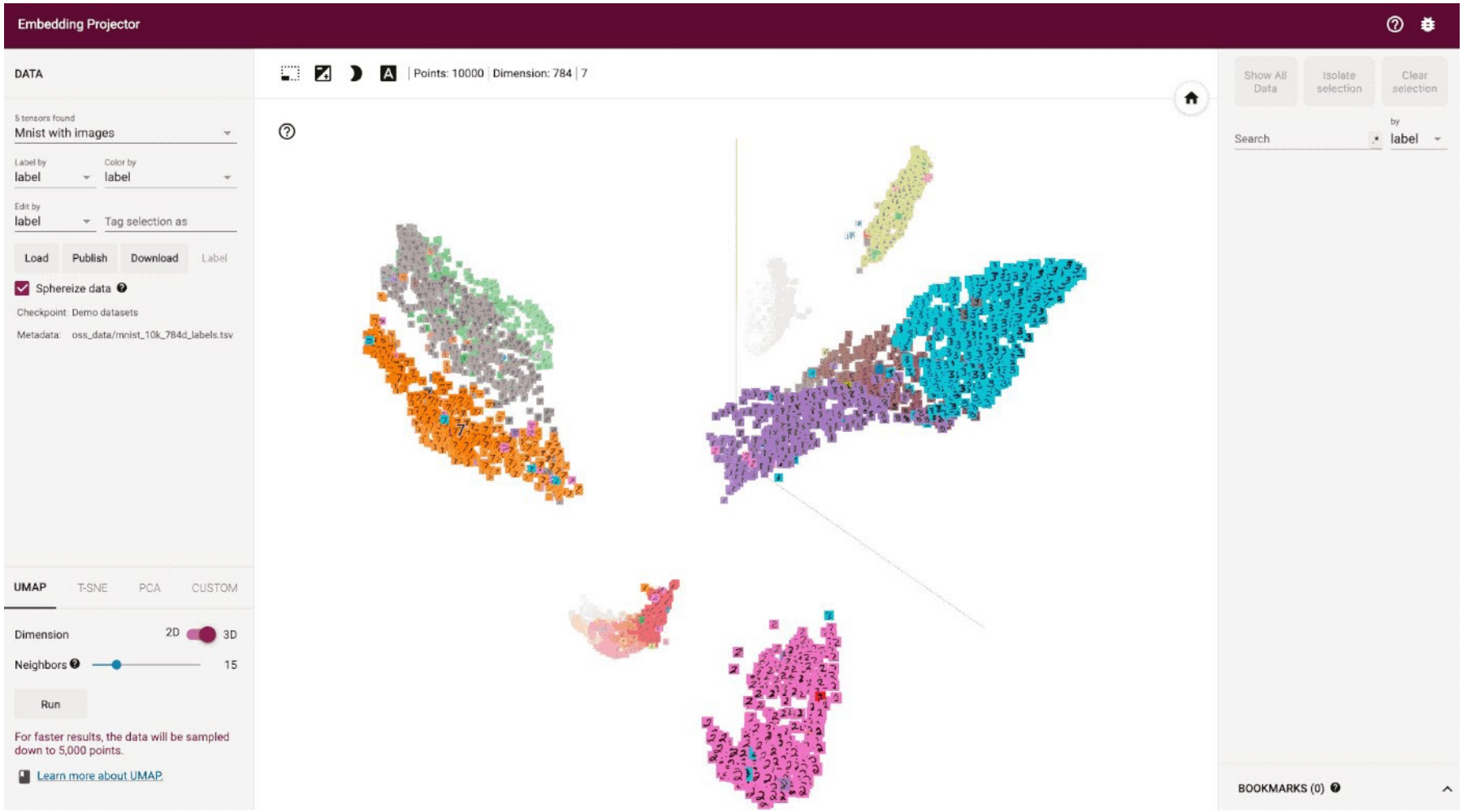


Occlusion Hides Information

- Interaction can resolve occlusion…… but at cost of time and cognitive load

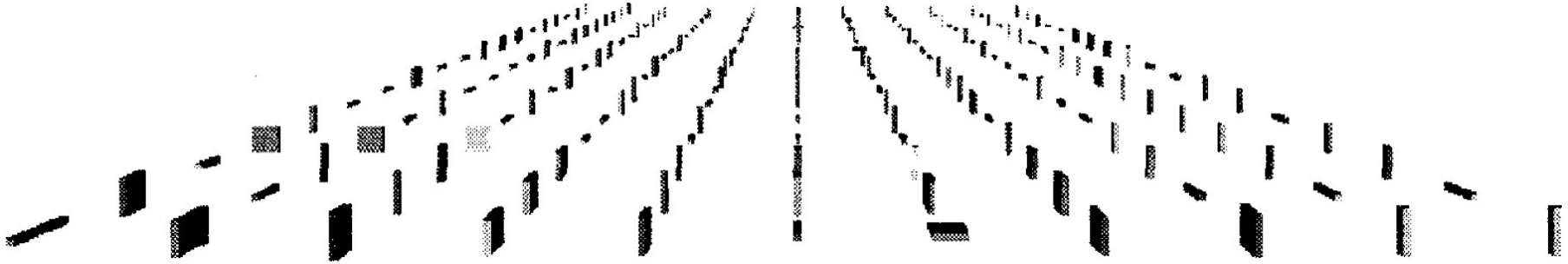


Distortion Viewing Techniques for 3D Data. Carpendale et al. InfoVis1996.


Embedding Projector
DATA
5 tensors found
Mnist with images
Label by
label
Color by
label
Edit by
label
Tag selection as
Load
Publish
Download
Label
Sphereize data
Checkpoint: Demo datasets
Metadata: oss_data/mnist_10k_784d_labels.tsv
UMAP
T-SNE
PCA
CUSTOM
Dimension
2D
3D
Neighbors
15
Run
For faster results, the data will be sampled down to 5,000 points.
Learn more about UMAP.
Points: 10000 | Dimension: 784 | 7
Show All Data
Isolate selection
Clear selection
Search
by
label
BOOKMARKS (0)

Perspective Distortion Loses Information

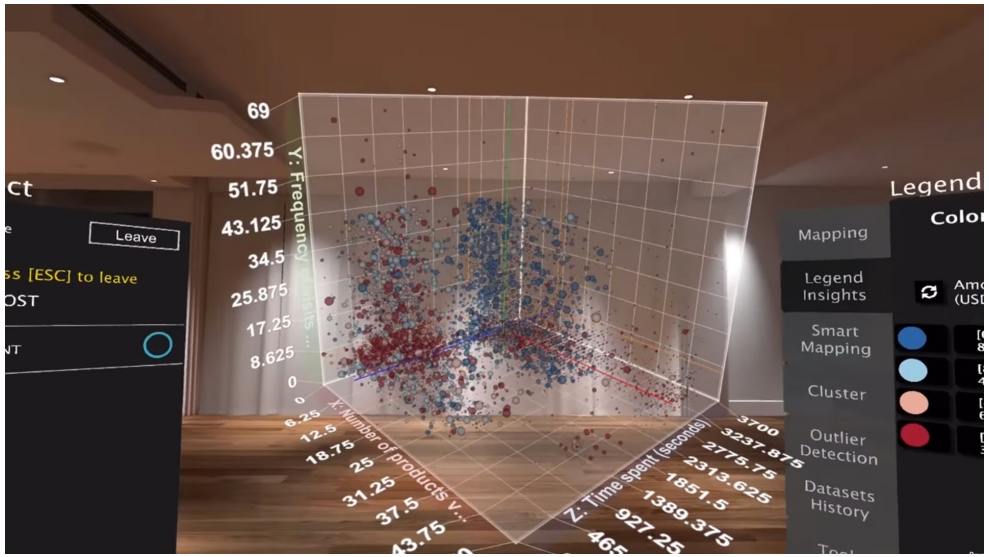
- Perspective Distortion
  - Interferes with all size channel encodings
  - Power of the plane is lost!

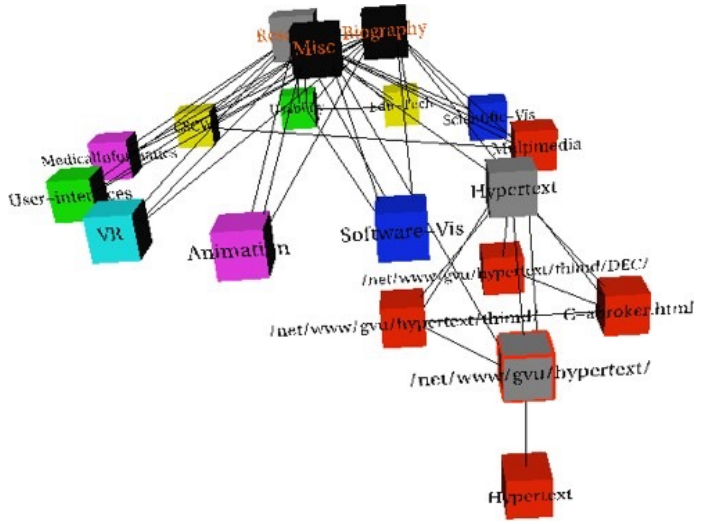


[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]

Tilted Text is not Legible

- Text Legibility [Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays. Grossman et al. CHI 2007]
  - Far worse when tilted from image plane





https://www.roadtovr.com/virtualitics-closes-7m-series-b-round-develop- vr-ar-data-visualization-platform/

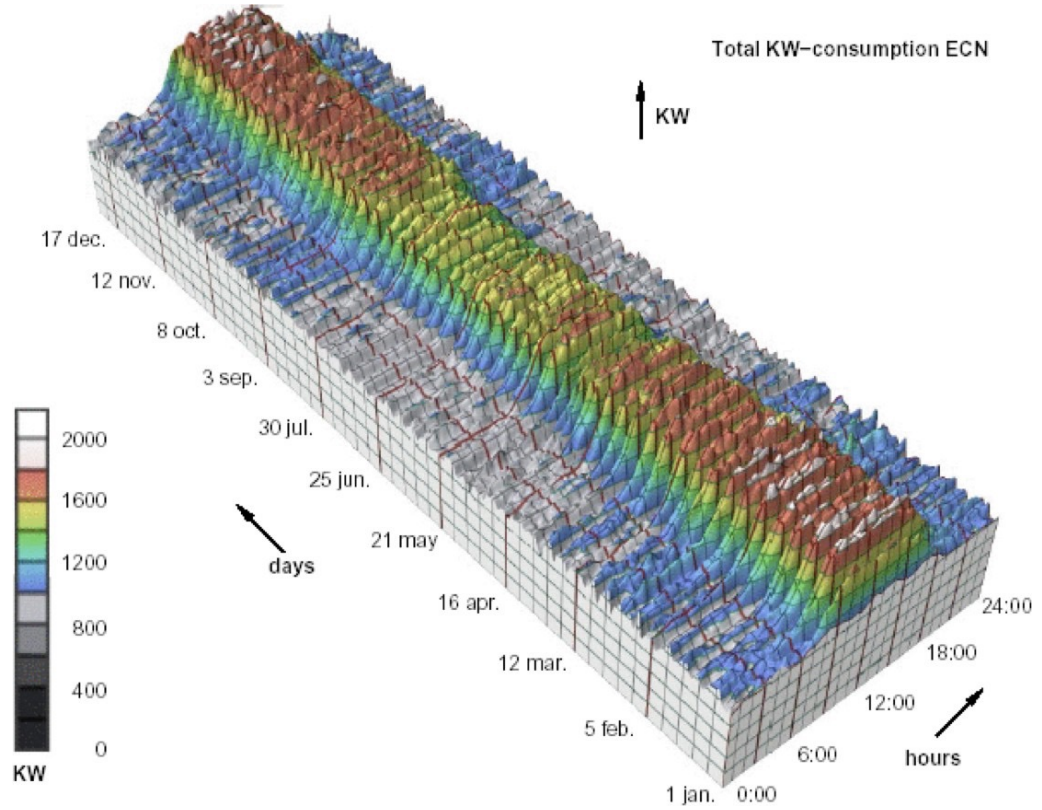
[Visualizing the World-Wide Web with the Navigational View Builder. Mukherjea and Foley. Computer Networks and ISDN Systems,1995.]

No Unjustified 3D Example: Time-series Data

- Extruded curves: detailed comparisons impossible

POWER DEMAND OF A RESEARCH FACILITY

Task: identify/compare similar periods of time



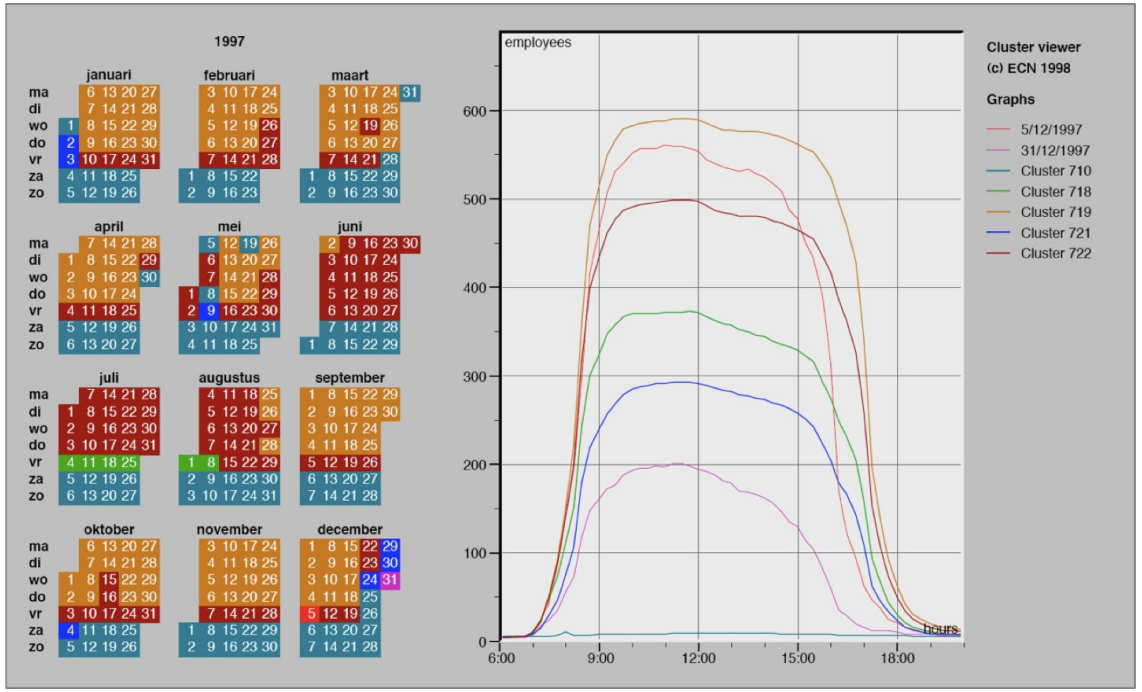
[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

No Unjustified 3D Example: Time-series Data

- Transform for new data abstraction: cluster hierarchy
  - Juxtapose multiple views (calendar)
  - Superimposed 2D curves (clusters)

POWER DEMAND OF A RESEARCH FACILITY

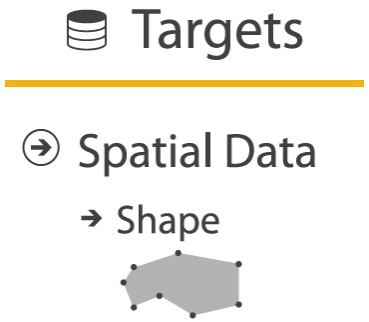
Task: identify/compare similar periods of time

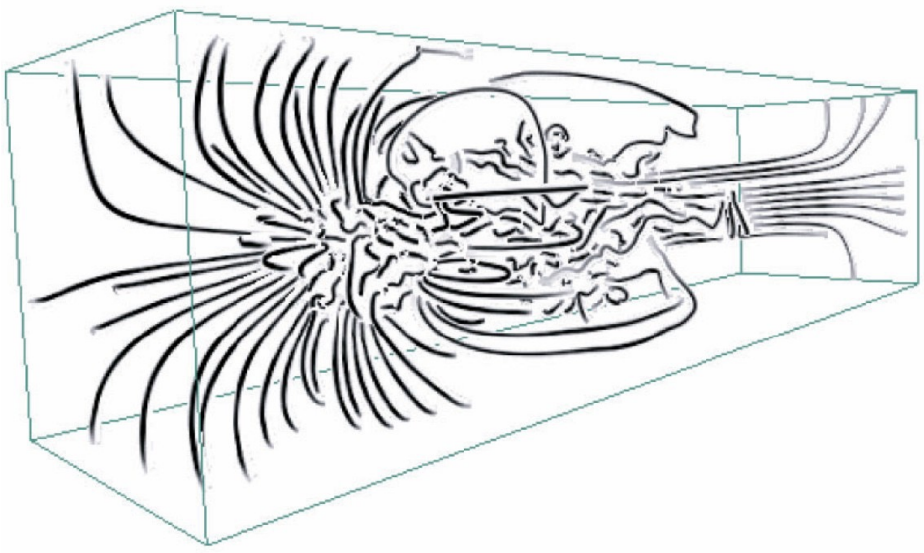


[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]

When to Use 3D: Shape Perception

- Benefits outweigh costs when **target** is shape perception for 3D spatial data
  - Interactive navigation supports synthesis across many viewpoints

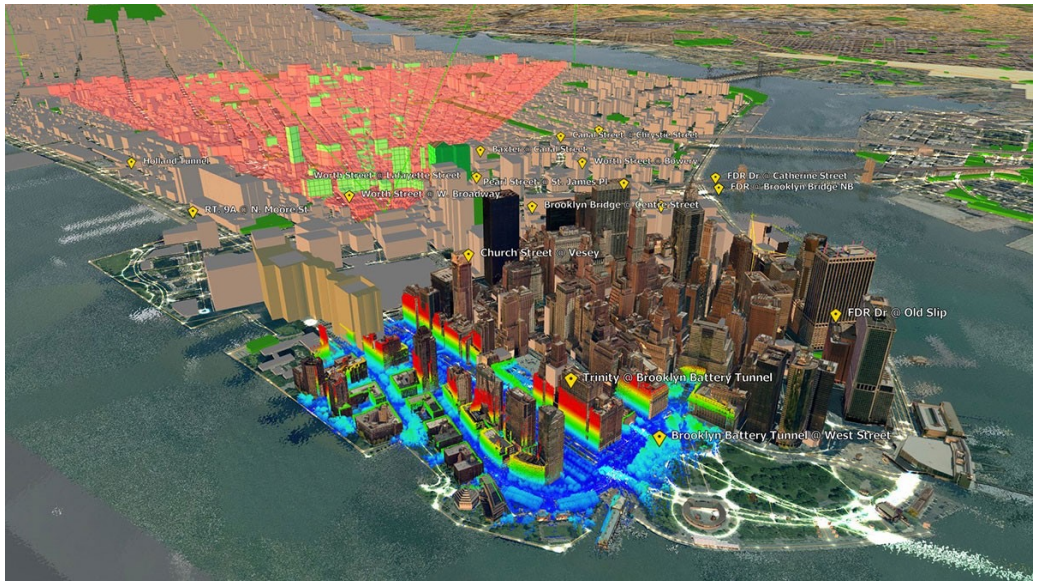




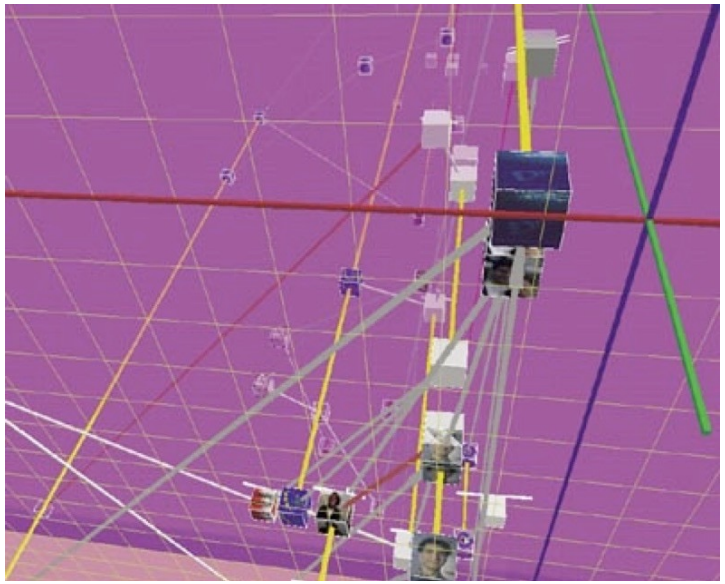
[Image-Based Streamline Generation and Rendering. Li and Shen. IEEE Trans. Visualization and Computer Graphics (TVCG) 13:3 (2007), 630-640.]

When to Use 3D

- 3D legitimate for true 3D spatial data
- 3D needs very careful justification for abstract data
  - Enthusiasm in 1990s, but now skepticism
  - Be especially careful with 3D for point clouds or networks



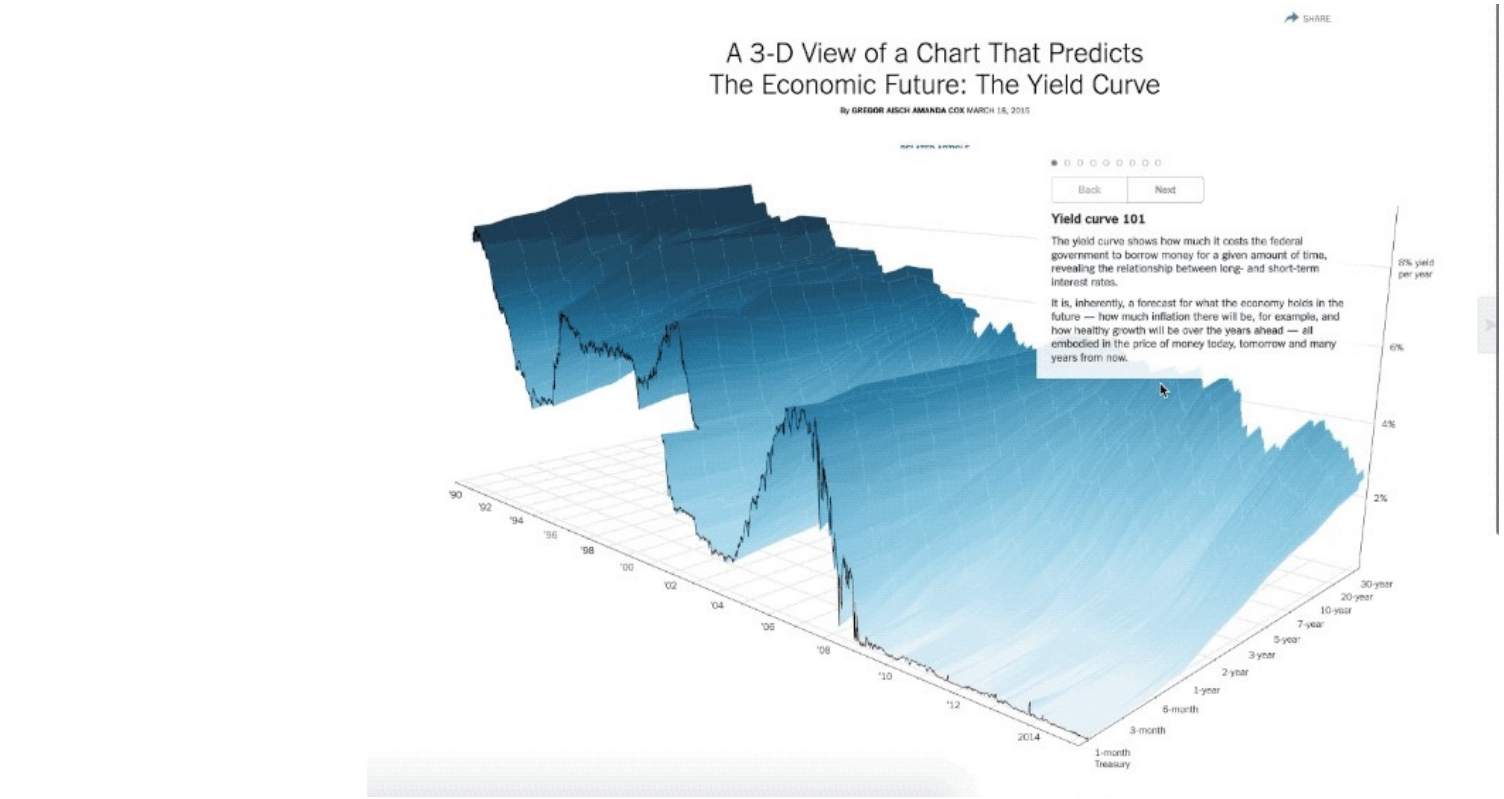
https://cityvis.io/project.php?id=83



[WEBPATH-a three dimensional Web history. Frecon and Smith. Proc. InfoVis 1999]

Good Use of 3D

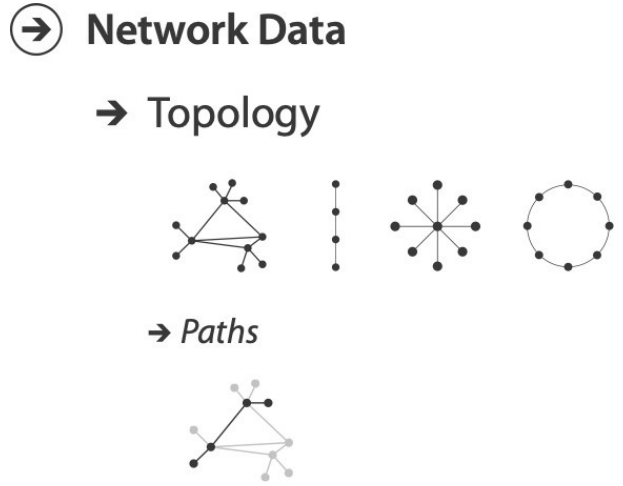
- Constrained navigation steps through carefully designed viewpoints



http://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html

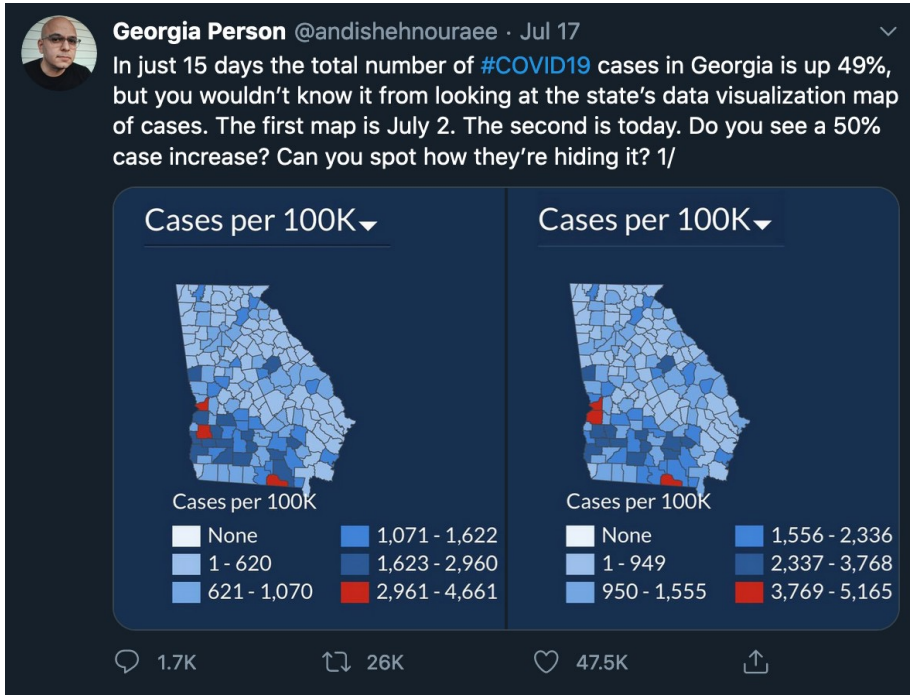
(2) No Unjustified 2D

- Consider whether network data requires 2D spatial layout
  - Especially if reading text is central to task!
  - Arranging as network means lower information density and harder label lookup compared to text lists
- Benefits outweigh costs when topological structure/context important for task
  - Be especially careful for search results, document collections, ontologies



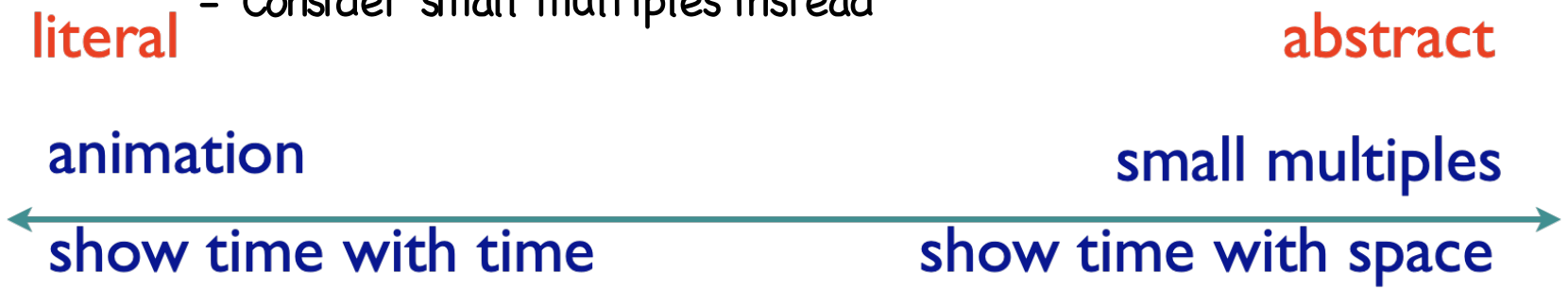
When to Use 2D: Map

- Don't use a map just because you have geo data
  - Use them for geographical tasks
  - Bad for comparison tasks



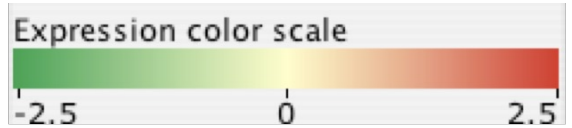
(3) Eyes Beat Memory

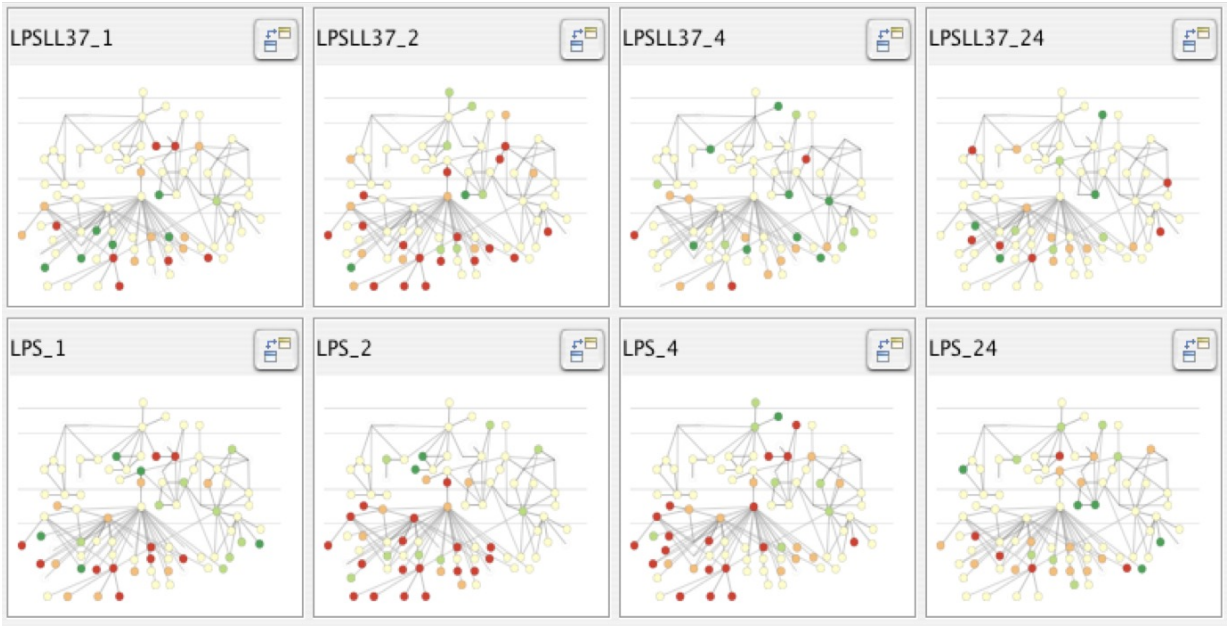
- Principle: external cognition vs. internal memory
  - Easy to compare by moving eyes between side-by-side views
  - Harder to compare visible item to memory of what you saw
- Implications for animation
  - Great for choreographed storytelling
  - Great for transitions between two states
  - Poor for many states with changes everywhere
    - Consider small multiples instead



Eyes Beat Memory Example

- Small multiples: one graph instance per condition
  - Same spatial layout
  - Color differently, by condition

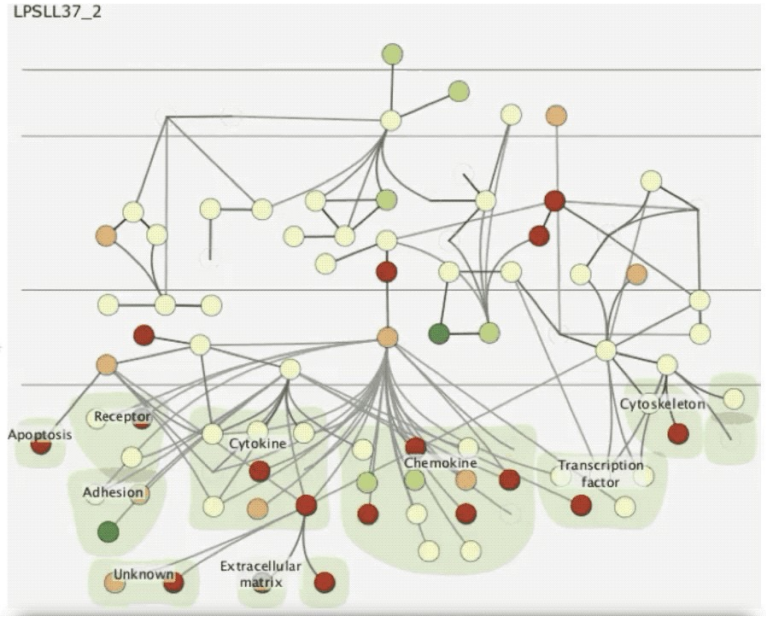




[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14:6 (2008), 1253–1260.]

Why not Animation?

- Disparate frames and regions: comparison difficult
  - Vs. contiguous frames
  - Vs. small region
  - Vs. coherent motion of group



- Safe special case
  - Animated transitions

- Change blindness
  - Even major changes difficult to notice if mental buffer wiped

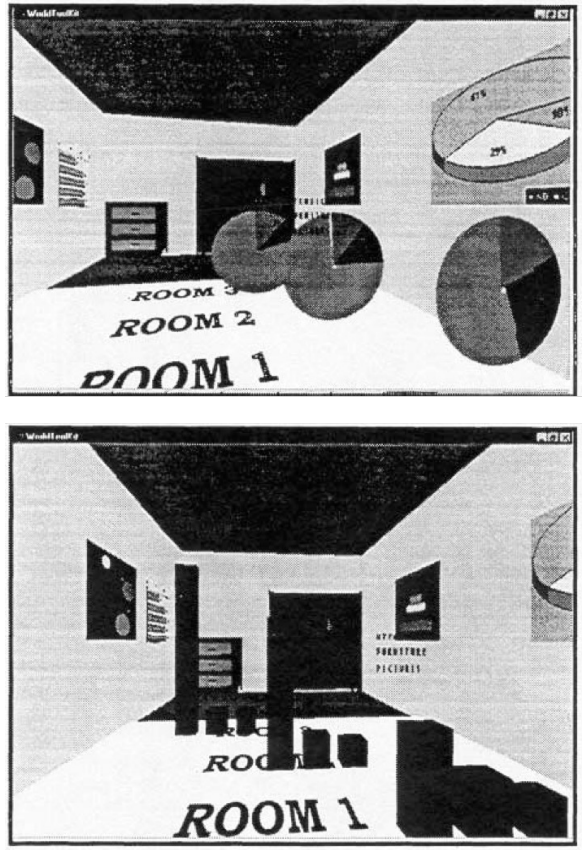
[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14:6 (2008), 1253-1260.]

Change Blindness

If attention is directed elsewhere, even drastic changes not noticeable

(4) Resolution Beats Immersion

- Immersion typically not helpful for abstract data
  - Do not need sense of presence or stereoscopic 3D
  - Desktop also better for workflow integration
- Resolution much more important
  - Pixels are the scarcest resource
- Virtual reality for abstract data difficult to justify thus far
  - But stay tuned with second wave



[Development of an information visualization tool using virtual reality. Kirner and Martins. Proc. Symp. Applied Computing 2000]

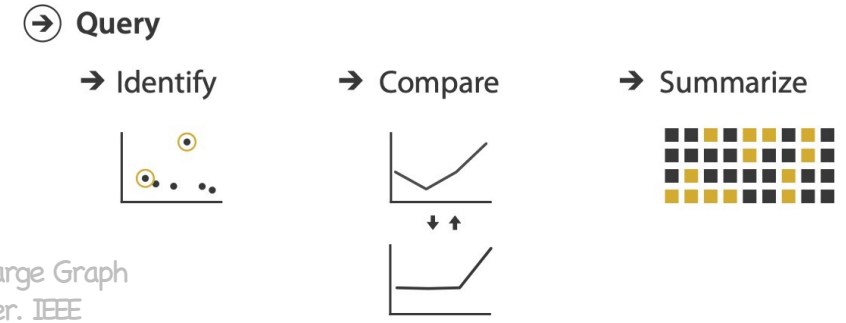
(5) Visualization Mantra (Shneiderman)

- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand
- Overview first, zoom and filter, then details on demand

[The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Shneiderman. Proc. IEEE Visual Languages, pp. 336-343, 1996.]

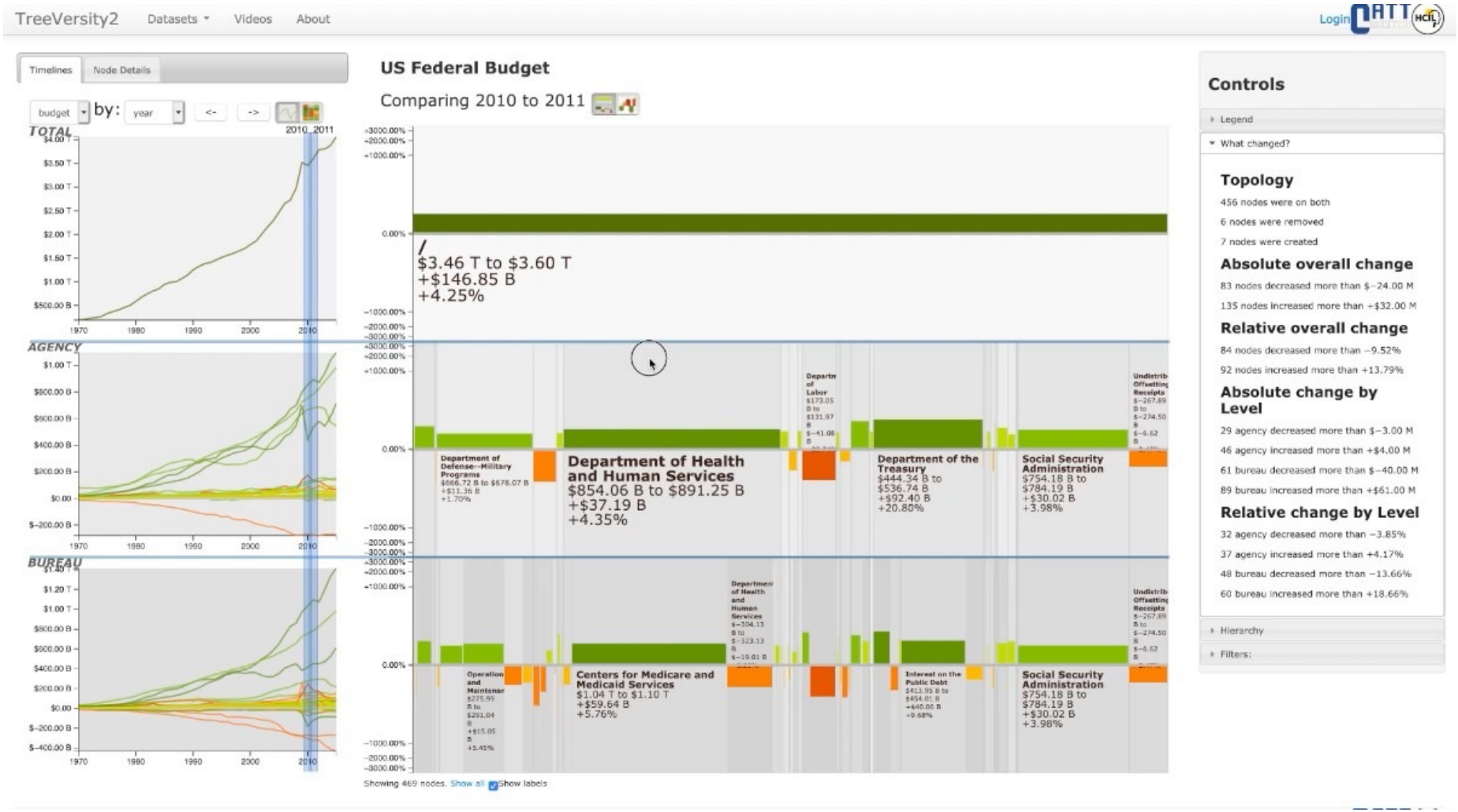
Overview First

- Overview = Summary
  - Microcosm of full visualization design problems
- Useful for exploration
- nuances
  - Beyond just two levels: multi-scale structure
  - Difficult when scale huge: give up on overview and browse local neighborhoods?



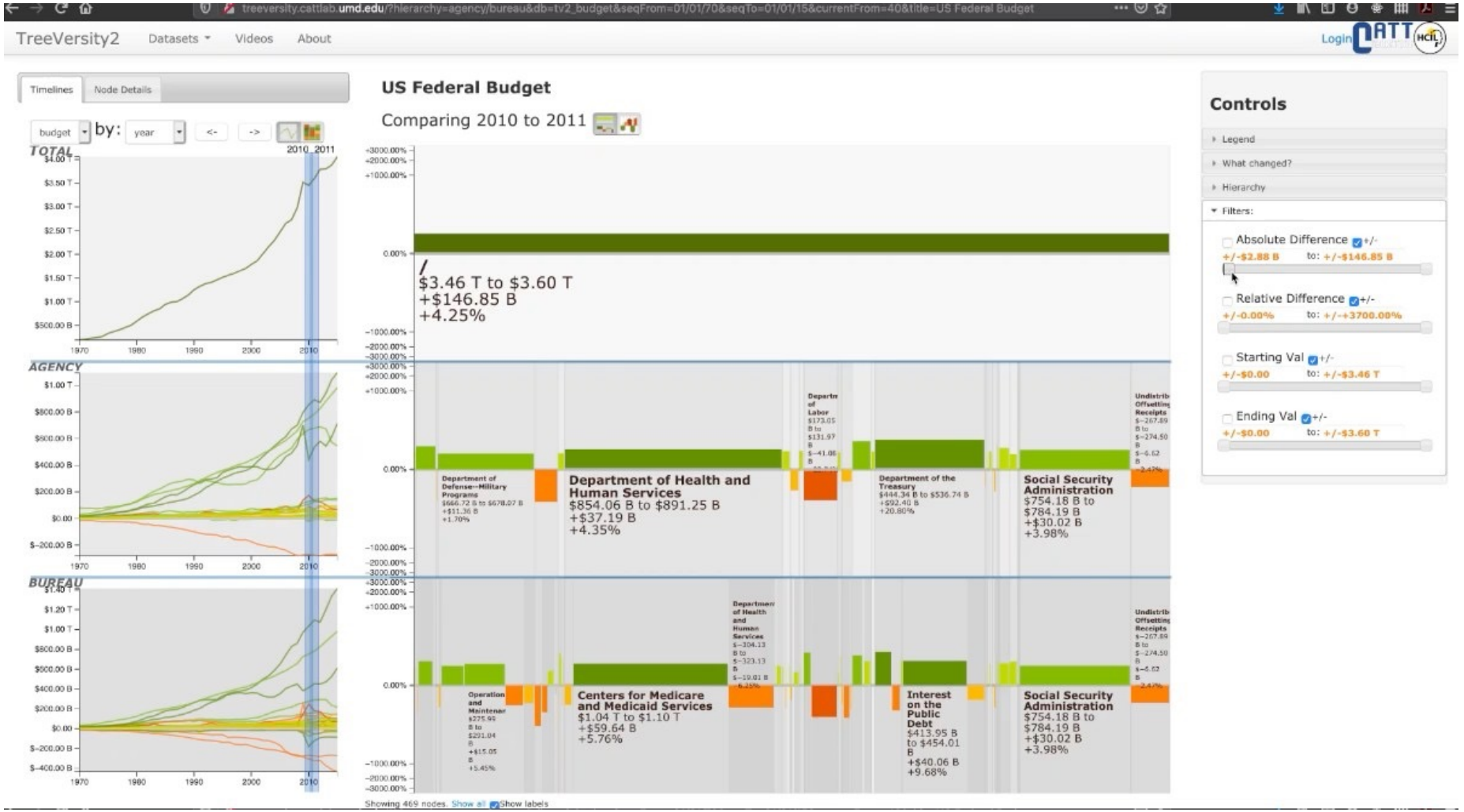
[Search, Show Context, Expand on Demand: Supporting Large Graph Exploration with Degree-of-Interest. van Ham and Perer. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2009) 15:6 (2009),

Overview → Zoom



http://treeversity.cattlab.umd.edu/?hierarchy=agency/bureau&db=tv2_budget&seqFrom=01/01/70&seqTo=01/01/15&currentFrom=40&title=US%20Federal%20Budget

Overview → Filter



http://treeversity.cattlab.umd.edu/?hierarchy=agency/bureau&db=tv2_budget&seqFrom=01/01/70&seqTo=01/01/15&currentFrom=40&title=US%20Federal%20Budget

When to Use Bottom-up?

- Sometimes makes sense to start small
- Useful for presentation
- Common in storytelling
- Helps explaining complex concepts

https://johnguerra.co/viz/resultadosSegundaVuelta2018/

(6) Function First, Form Next

- Start with focus on functionality
  - Possible to improve aesthetics later on, as refinement
  - If no expertise, seek for help from graphic designers
- Dangerous to start with aesthetics
  - Usually impossible to add function retroactively

Aesthetics DO Matter

- Another level of function
  - Visual hierarchy, alignment, flow
  - Gestalt principles in action
- Very important for the enjoy task

[The Non-Designer's Design Book. Robin Williams. 3rd edition. Peachpit Press, 2008.]

(7) Responsiveness is Required

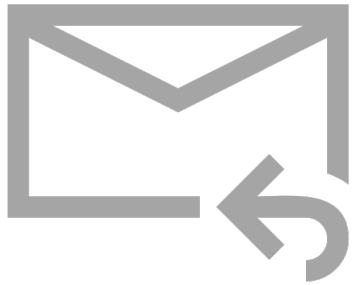
Visual feedback: three rough categories

- 0.1 seconds: perceptual processing
  - Subsecond response for mouseover highlighting: ballistic motion
- 1 second: immediate response
  - Fast response after mouse click, button press: Fitts's Law limits on motor control
- 10 seconds: brief tasks
  - Bounded response after dialog box: mental model of heavyweight operation (file load)

Scalability Considerations

- Highlight selection without complete redraw of view (graphics frontbuffer)
- Show for multi-second operations (check for cancel/undo)
- Show progress bar for long operations (process in background thread)
- Rendering speed when item count is large (guaranteed frame rate)

Questions?
Thank you ☺



Quan Li

liquan@shanghaitech.edu.cn