Information Visualization

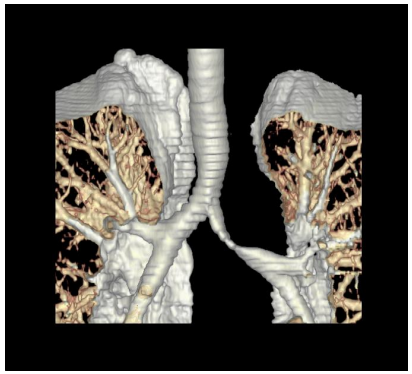
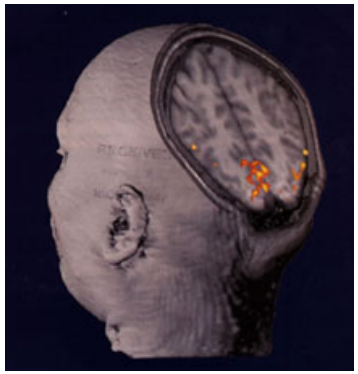
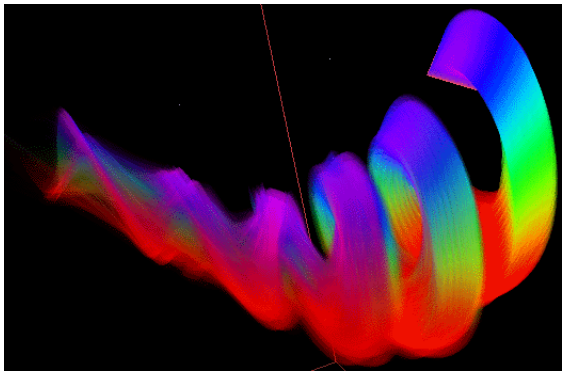
서진욱 교수
Human-Computer Interaction Lab
Department of Computer Science and Engineering
Seoul National University

What is it?

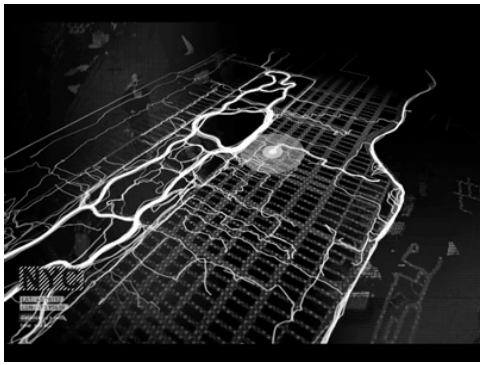
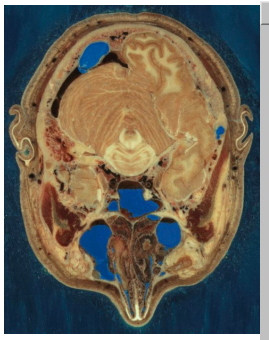
- Scientific Visualization vs. Information Visualization
- Data Visualizations – its power
- Data Visualization – definition

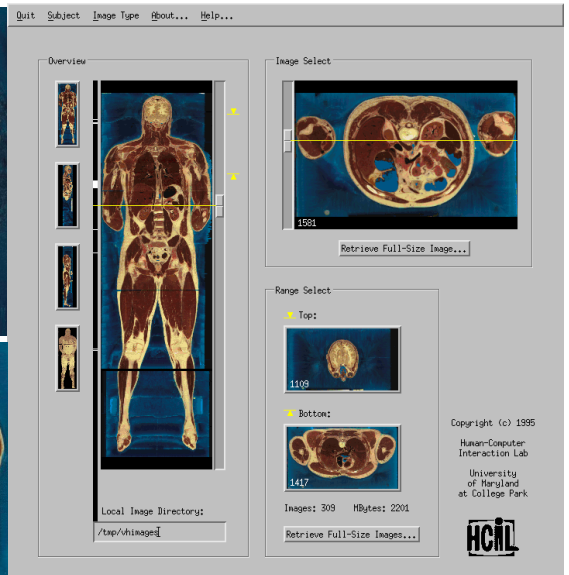
Visualization

- Scientific Visualization

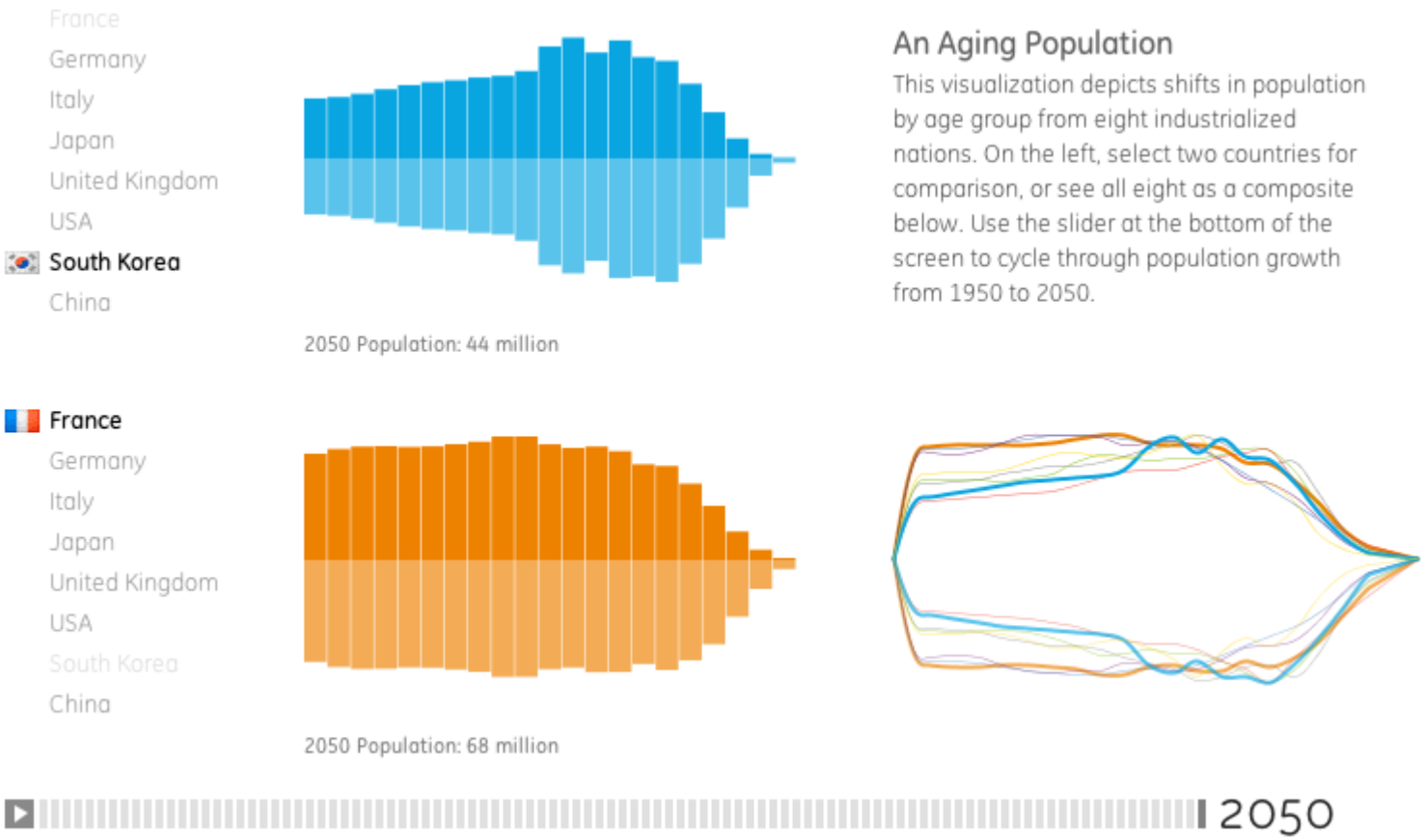




- Information Visualization
  - abstract data



Aging



http://fathom.info/aging

Text Visualization

- Aesthetic Visual Summary of Large Texts

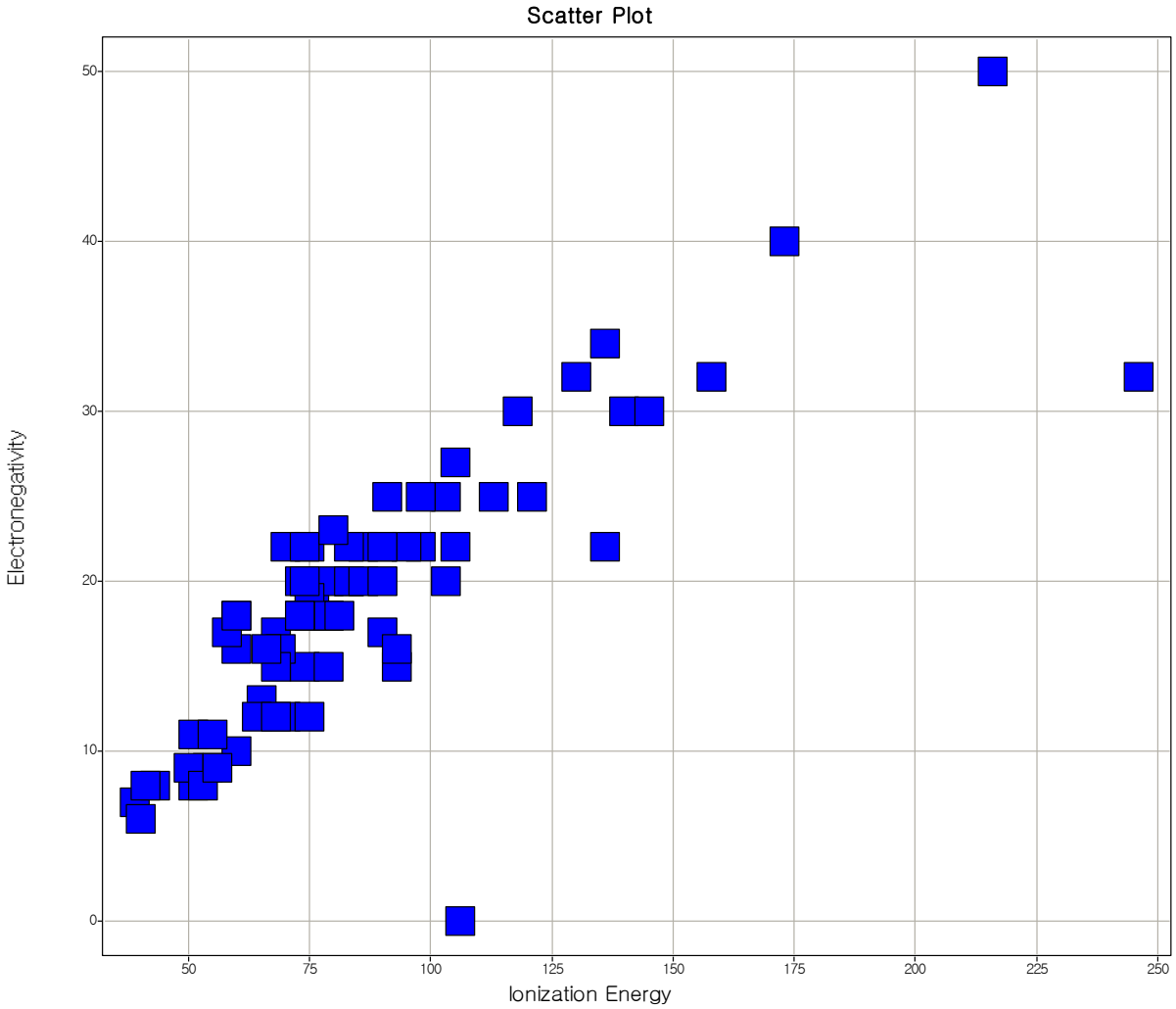
http://hcil.snu.ac.kr/research/maniwordle
http://hcil.snu.ac.kr/AniWordle

Anything interesting?

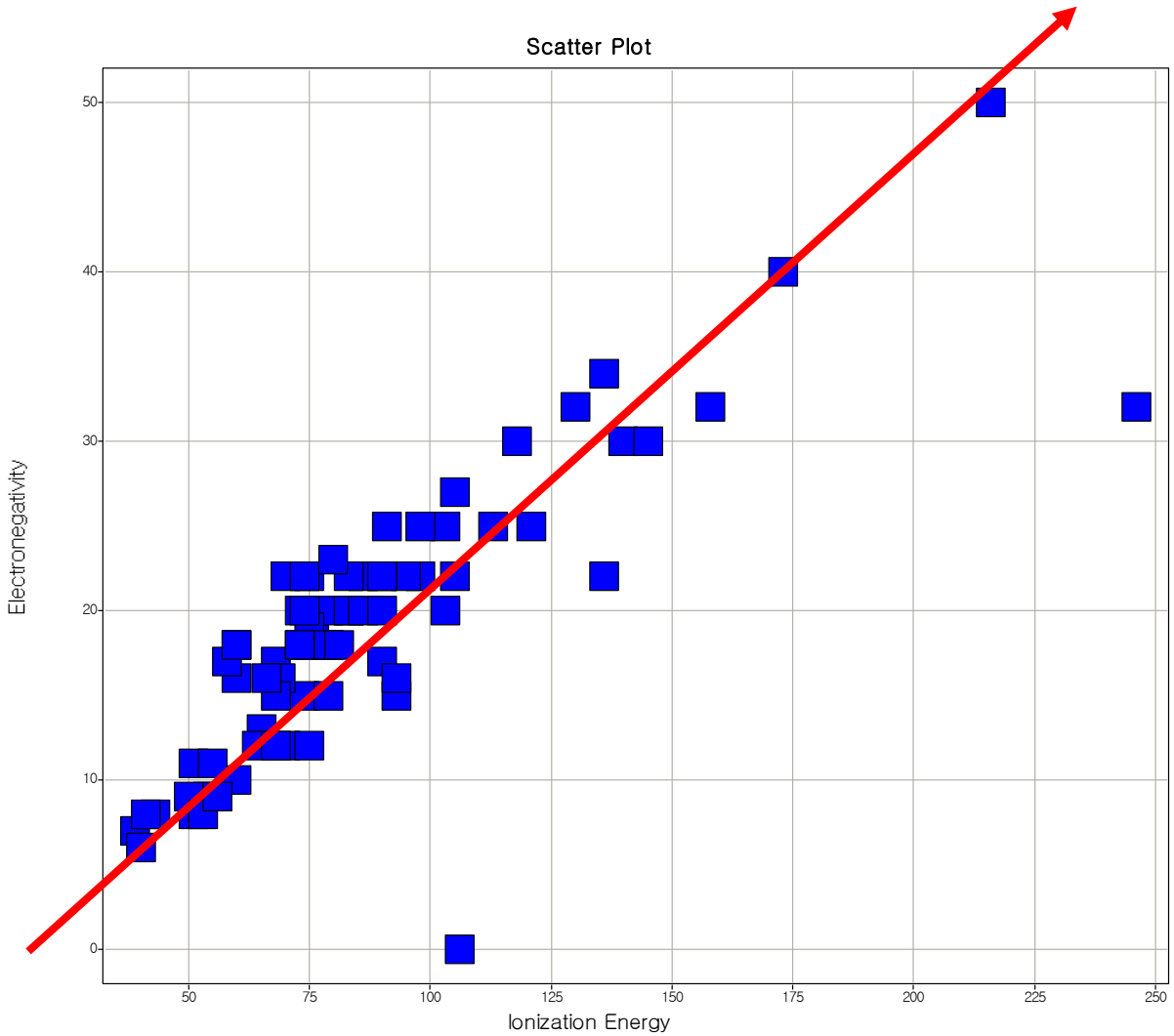
	A	B	C	D	E	F	G	H	I	J	K
1	Element	*P1	*P2	Atomic Num	Atomic Mas	Atomic Radi	Ionic Radius	Ionization E	Electronega	*C1	*C2
2	Ac	140	0	89	227	200	126	51	11	62	56
3	Ag	630	80	47	107	144	129	75	18	124	40
4	Al	750	160	13	27	143	67	60	16	28	25
5	Ar	1050	160	18	39	98	154	158	32	176	51
6	As	870	120	33	75	120	72	98	22	115	33
7	At	990	40	85	210	140	76	95	22	119	22
8	Au	630	40	79	197	144	99	91	25	131	22
9	B	750	200	5	10	85	41	83	20	101	8
10	Ba	80	40	56	137	222	149	51	8	46	56
11	Be	80	200	4	9	112	59	93	15	82	15
12	Bi	870	40	83	209	150	117	73	20	140	27
13	Br	990	120	35	79	114	182	118	30	161	44
14	C	810	200	6	12	77	30	113	25	82	1
15	Ca	80	120	20	40	197	114	60	10	70	51
16	Cd	690	80	48	112	151	109	90	17	113	43
17	Cl	990	160	17	35	100	167	130	32	173	47
18	Co	500	120	27	59	125	83	79	18	120	30
19	Cr	320	120	24	52	128	75	68	17	91	28
20	Cs	20	40	55	132	265	181	39	7	7	56
21	Cu	630	120	29	63	128	87	76	19	118	32
22	F	990	200	9	19	72	119	173	40	39	1
23	Fe	440	120	26	55	126	83	79	18	115	32
24	Fr	20	0	87	223	269	194	40	6	1	56
25	Ga	750	120	31	69	135	76	60	18	89	31
26	Ge	810	120	32	72	122	87	79	20	118	33
27	H	20	240	1	1	32	0	136	22	40	1
28	He	1050	240	2	4	31	93	246	32	1	1
29	Hf	200	40	72	178	159	85	70	12	95	44
30	Hg	690	40	80	200	151	116	103	20	147	27
31	I	990	80	53	126	133	206	105	27	153	44
32	In	750	80	49	114	167	94	58	17	93	42
33	Ir	500	40	77	192	136	82	90	22	116	25
34	K	20	120	19	39	227	152	43	8	37	56
35	Kr	1050	120	36	83	112	169	140	30	163	47

Periodic Table Data

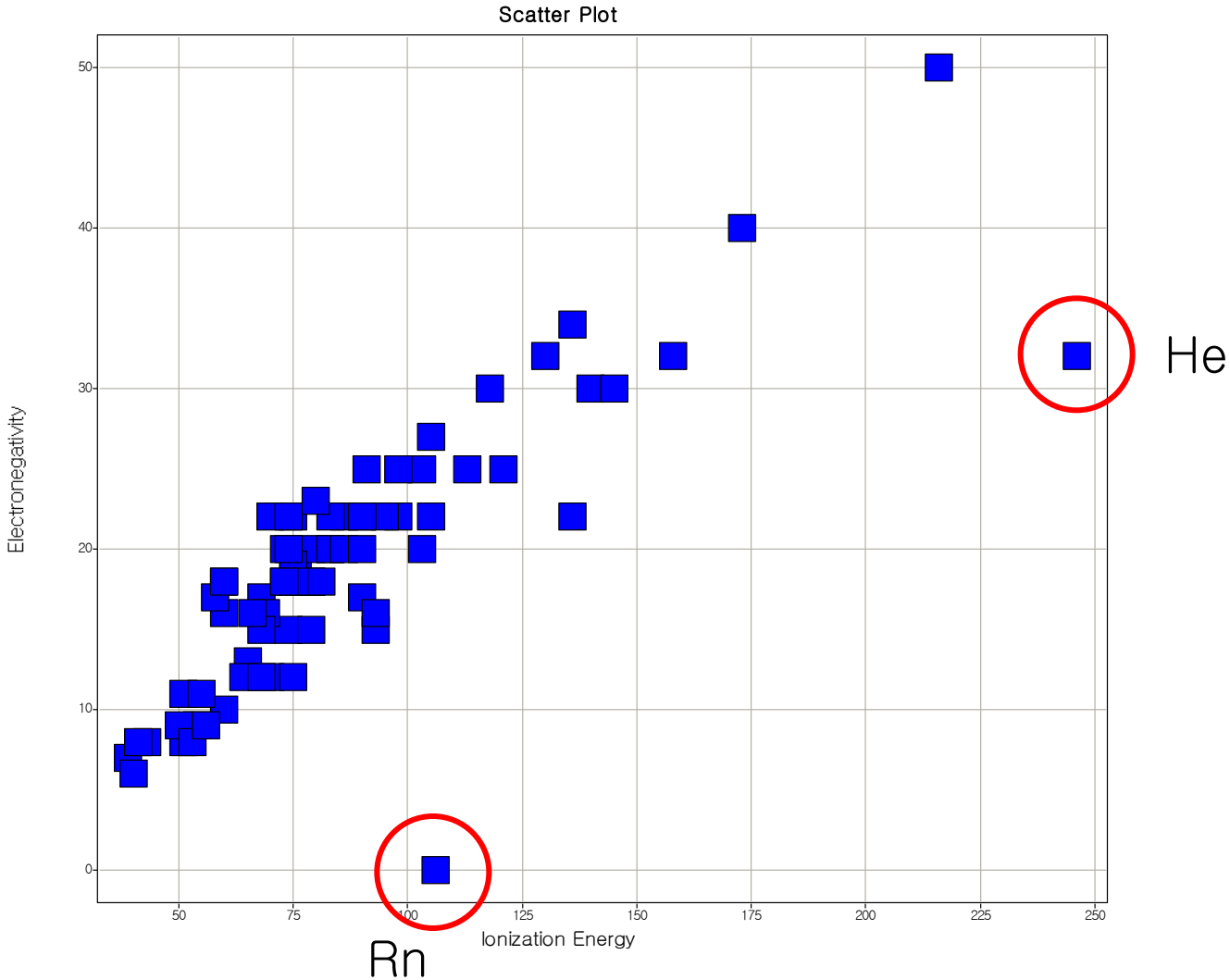
Anything interesting?



Correlation…What else?

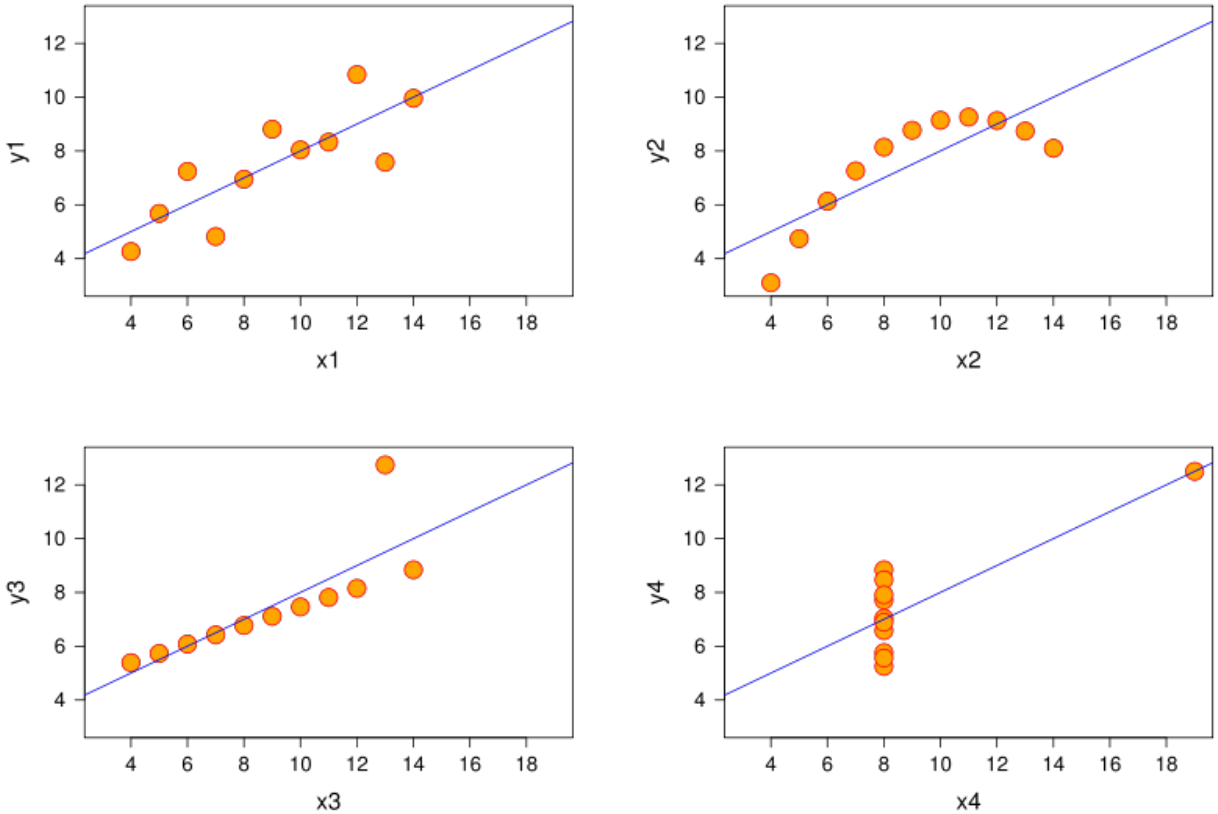


Outliers


Scatter Plot
Electronegativity
Ionization Energy
0
10
20
30
40
50
50
75
100
125
150
175
200
225
250
He
Rn

Visualizations Reveal the Data

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



Property	Value
Mean of each x variable	9.0
Variance of each x variable	10.0
Mean of each y variable	7.5
Variance of each y variable	3.75
Correlation between each x and y variable	0.816
Linear regression line	$y = 3 + 0.5x$

Anscombe's Quartet

tables and images from *Wikipedia*

Definitions

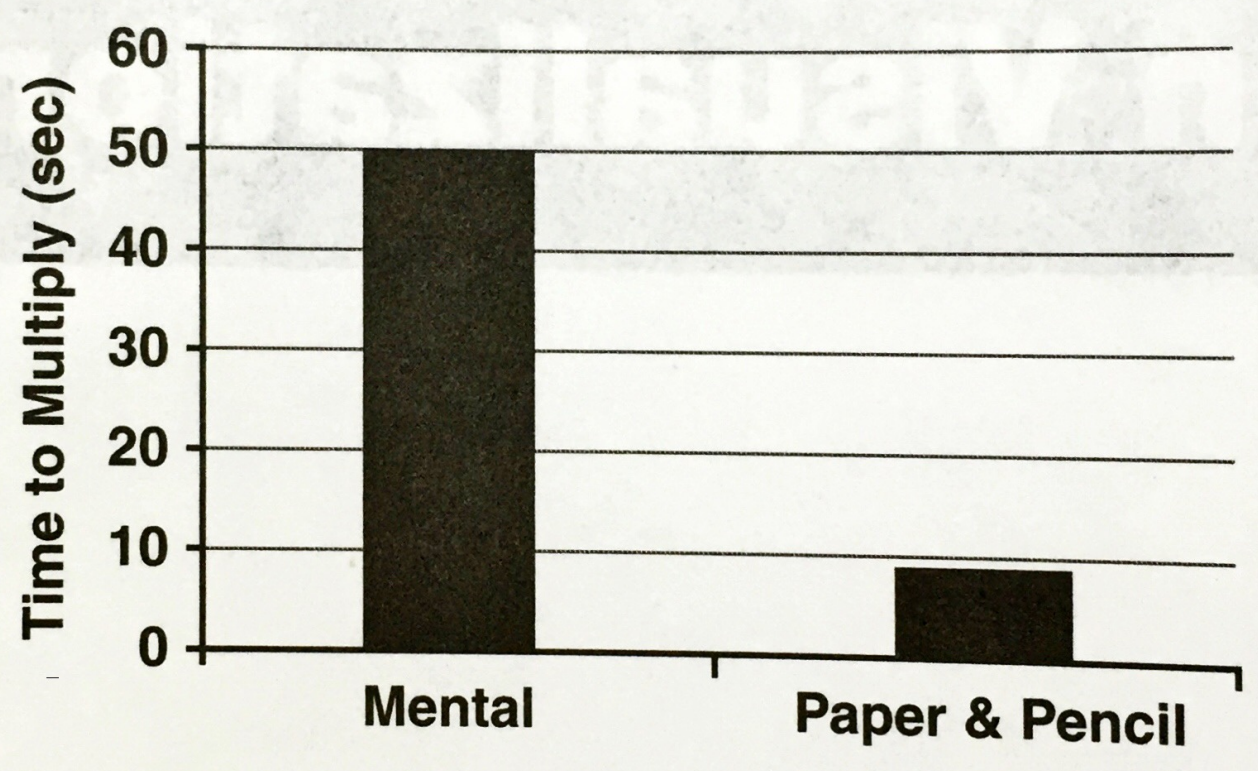
- The use of computer-supported, interactive, visual representations of *abstract* data to amplify cognition
  -- Stuart Card, Jock Mackinlay, Ben Shneiderman, 1999

- Finding the *artificial memory* that best supports our natural means of perception
  -- Bertin, 1983

- Provide tools that present data in a way to help people understand and gain insight from it

Extending Working Memory

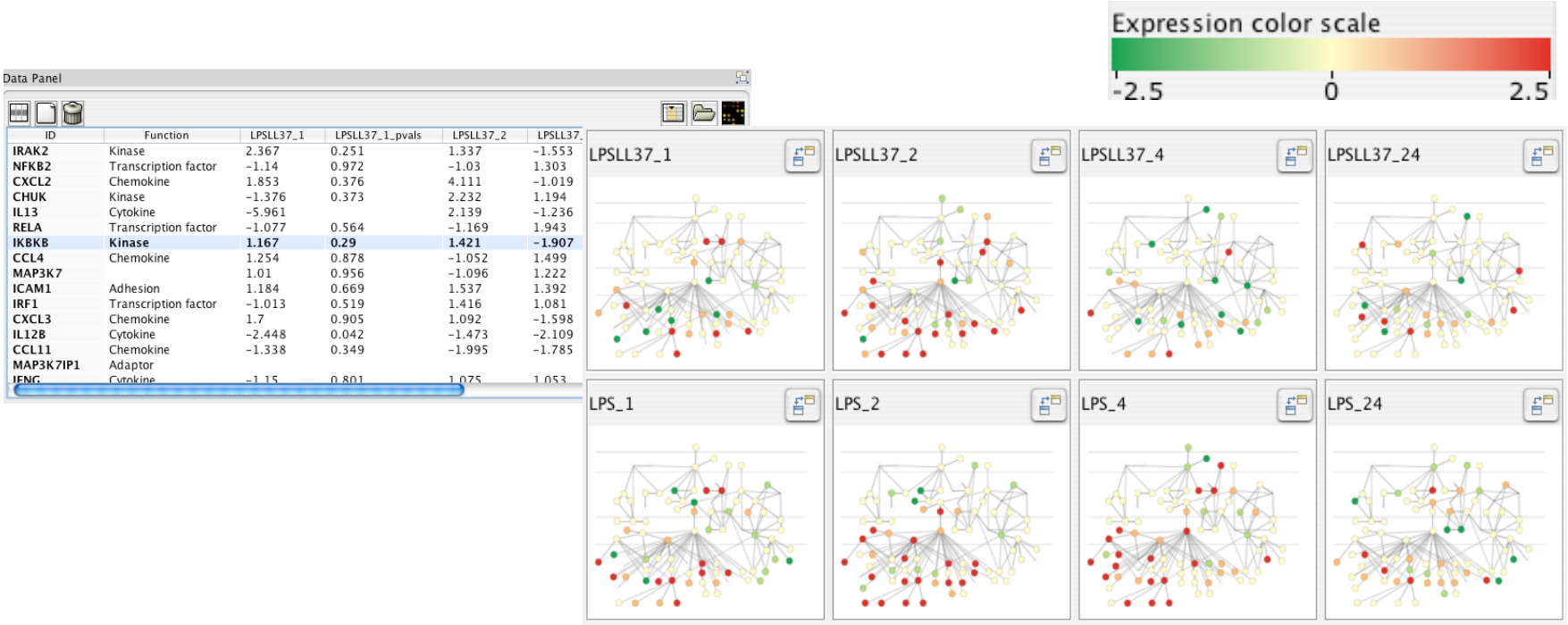
$$
\begin{array}{r}
34 \\
\times\, 72 \\
\hline
68 \\
23^{2}80 \\
\hline
24^{1}48
\end{array}
$$



Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- external representation: **replace cognition with perception**



ID	Function	LPSLL37_1	LPSLL37_1_pvals	LPSLL37_2	LPSLL37_
IRAK2	Kinase	2.367	0.251	1.337	-1.553
NFKB2	Transcription factor	-1.14	0.972	-1.03	1.303
CXCL2	Chemokine	1.853	0.376	4.111	-1.019
CHUK	Kinase	-1.376	0.373	2.232	1.194
IL13	Cytokine	-5.961		2.139	-1.236
RELA	Transcription factor	-1.077	0.564	-1.169	1.943
IKBKB	Kinase	1.167	0.29	1.421	-1.907
CCL4	Chemokine	1.254	0.878	-1.052	1.499
MAP3K7		1.01	0.956	-1.096	1.222
ICAM1	Adhesion	1.184	0.669	1.537	1.392
IRF1	Transcription factor	-1.013	0.519	1.416	1.081
CXCL3	Chemokine	1.7	0.905	1.092	-1.598
IL12B	Cytokine	-2.448	0.042	-1.473	-2.109
CCL11	Chemokine	-1.338	0.349	-1.995	-1.785
MAP3K7IP1	Adaptor				
IFNG	Cytokine	-1.15	0.801	1.075	1.053

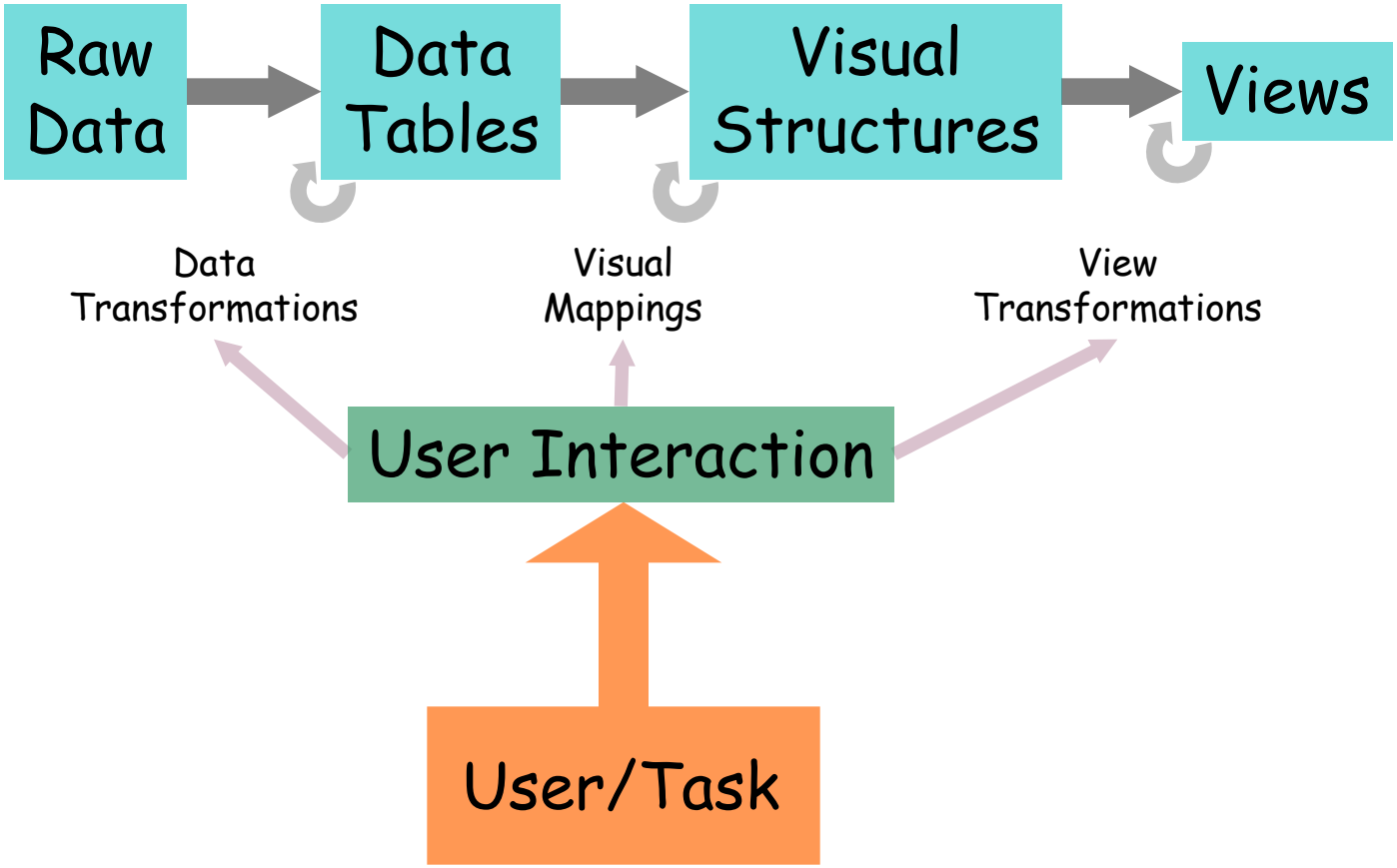
[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Resource limitations

Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

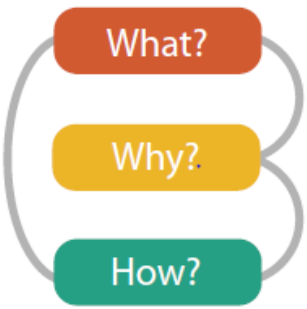
- computational limits
  - processing time
  - system memory
- human limits
  - human attention and memory
- display limits
  - pixels are precious resource, the most constrained resource
  - **information density**: ratio of space used to encode info vs unused whitespace
    - tradeoff between clutter and wasting space, find sweet spot between dense and sparse

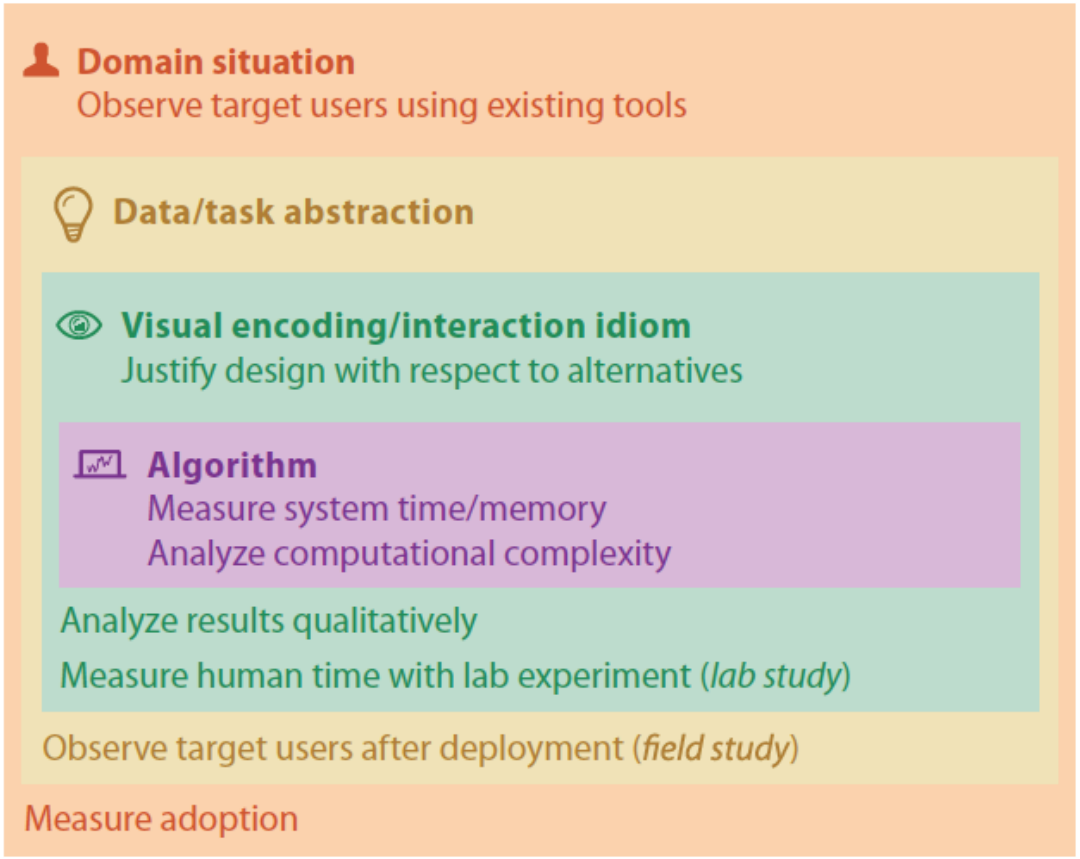
InfoVis Reference Model



Model for Design and Validation Vis System

- Four nested levels of vis design
- Threats to validating each level
  - Why validate? "Design space is large, and most designs are ineffective"



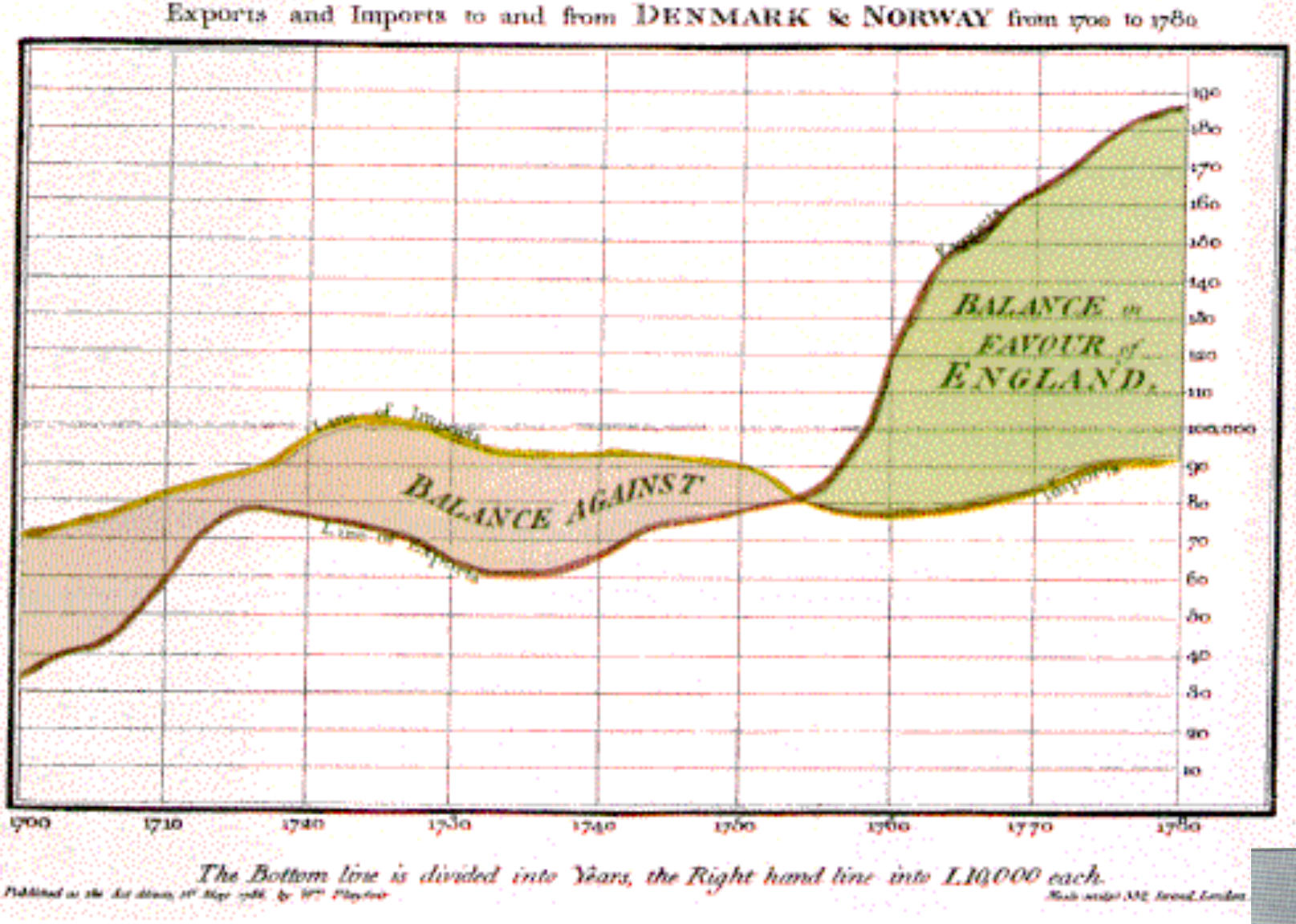


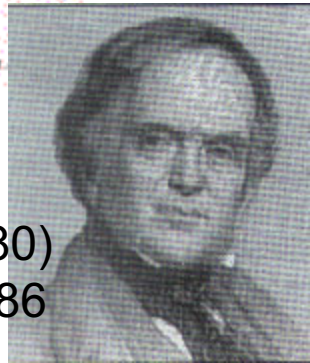
InfoVis is Interdisciplinary

- Graphics: drawing in real time (<100 ms)
- Cognitive psychology: appropriate representation
- **HCI**: using <u>users</u> and <u>tasks</u> to guide design and evaluation

Historical Aspect

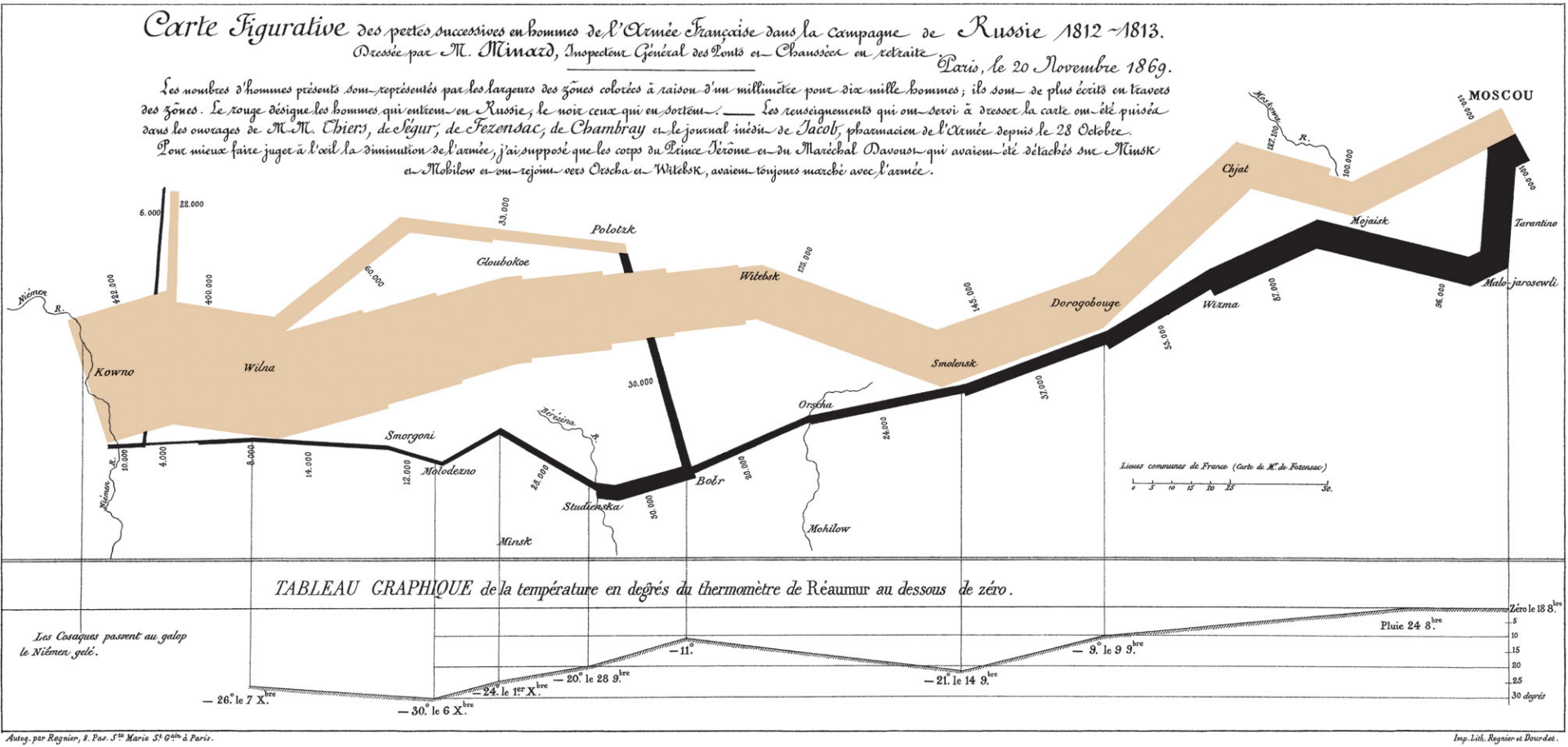
- Data Visualizations – historical examples
- Power of Visualization – scatterplots examples





Exports and Imports of England to and from Denmark & Norway (1700~1780)
The Commercial and Political Atlas, **William Playfair** (French civil engineer), 1786

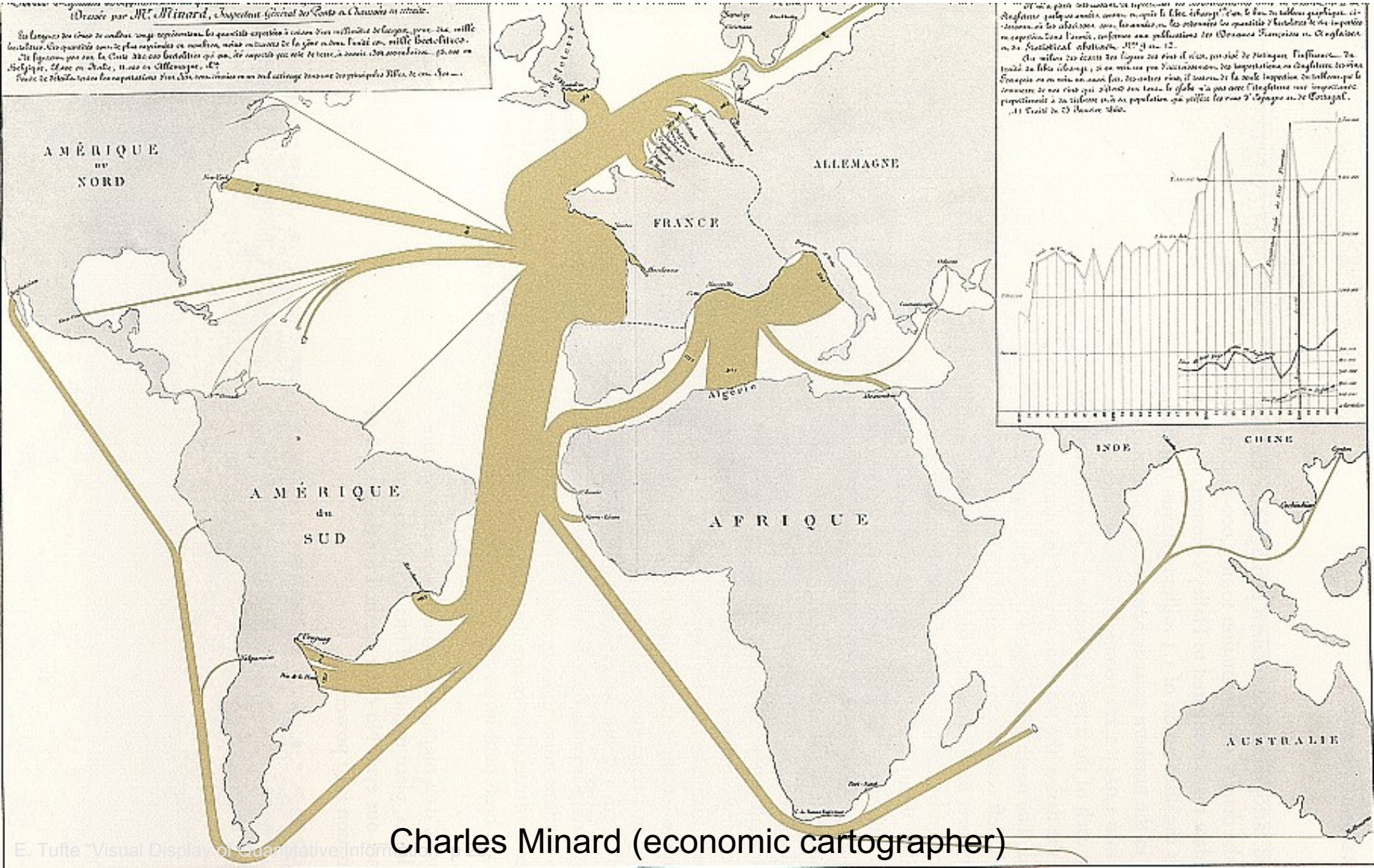
from *Microsoft Excel 2007*



Advance of Napoleon's Grande Armée into Russia in 1812
Charles Joseph Minard, 1861

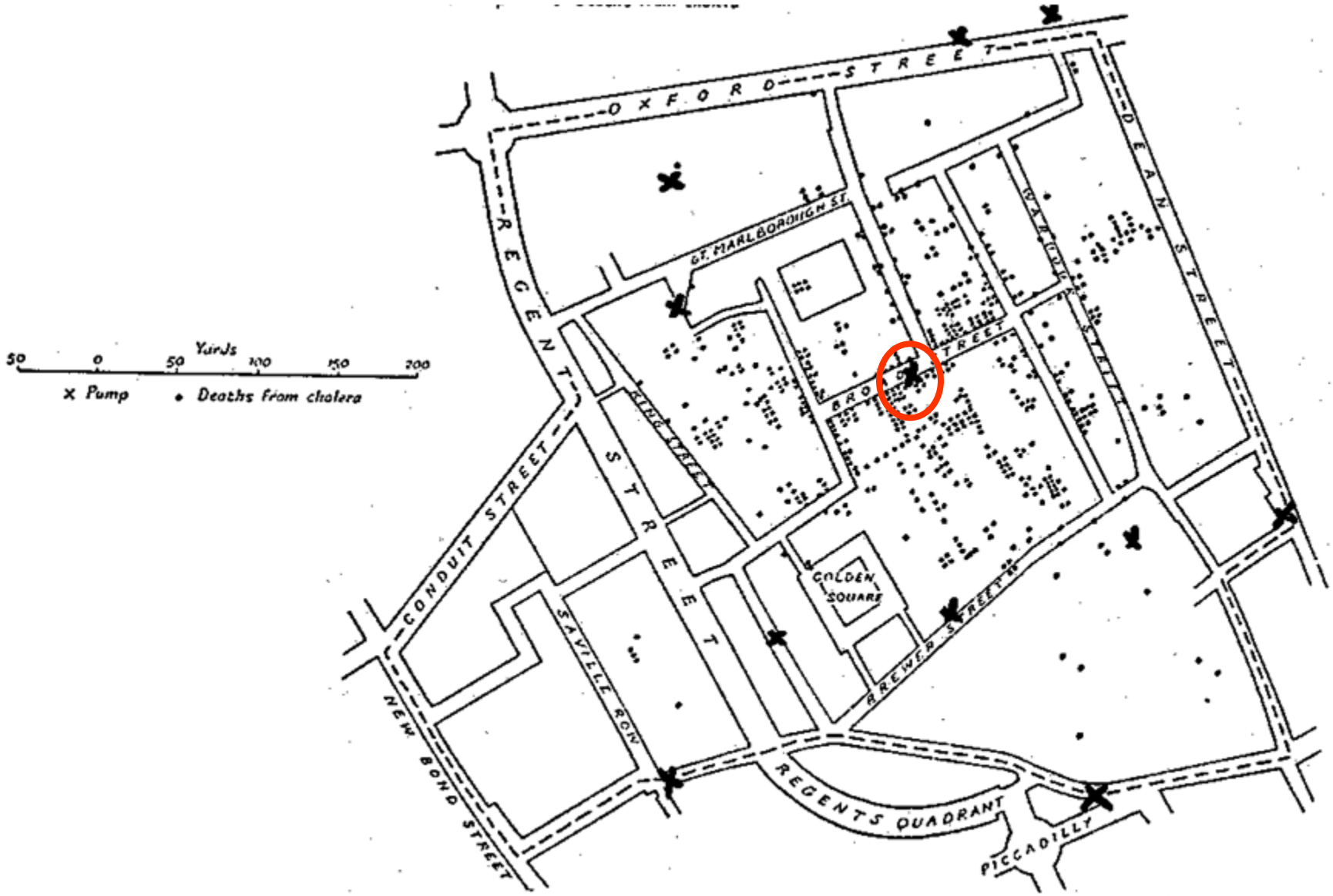
Size of army
Position
Direction of movement
Temperature
Time

1864 Exports of French Wine



Charles Minard (economic cartographer)

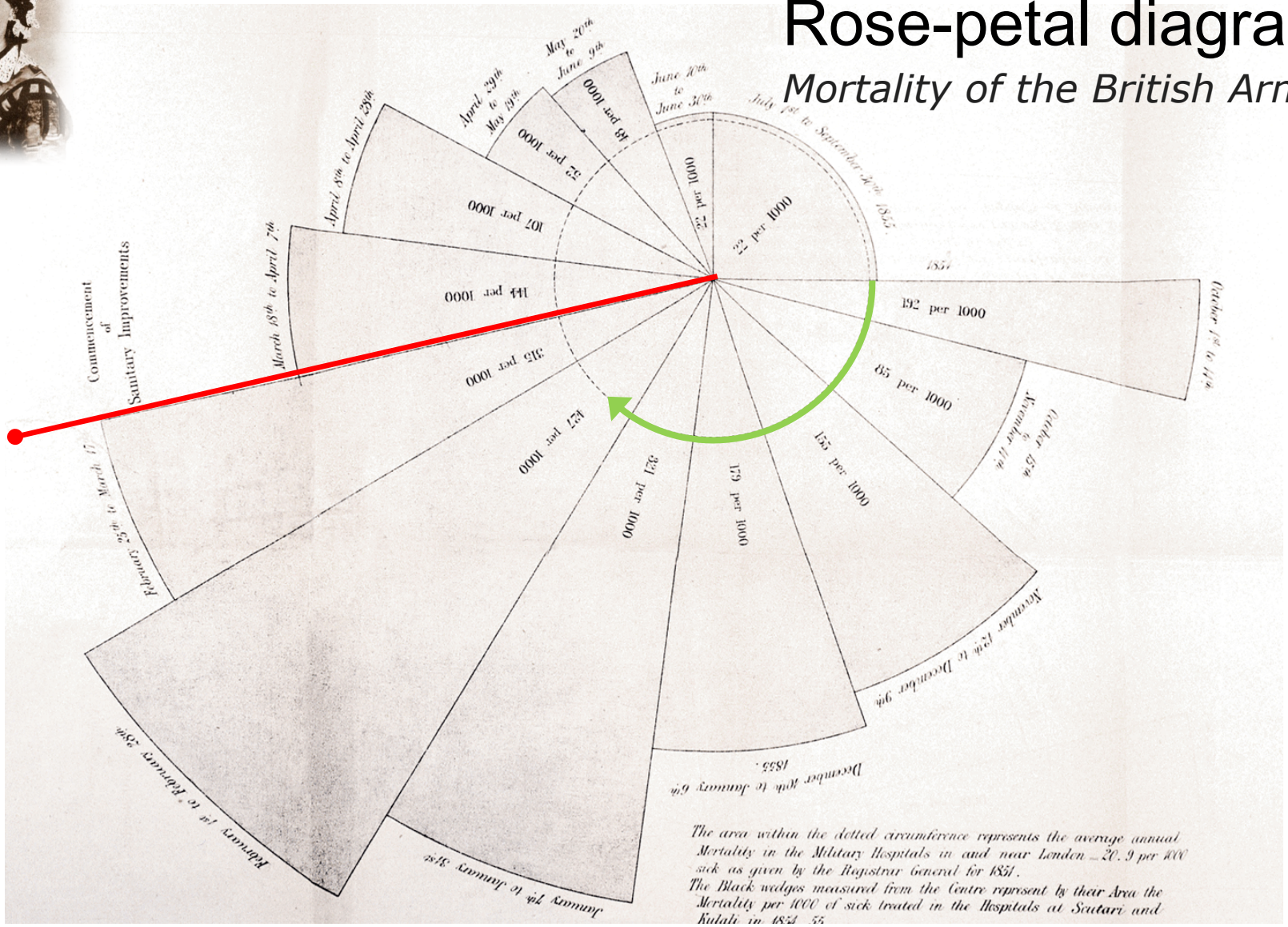
E. Tufte "Visual Display of Quantitative Inform



The 1854 London Cholera Epidemic
James Snow

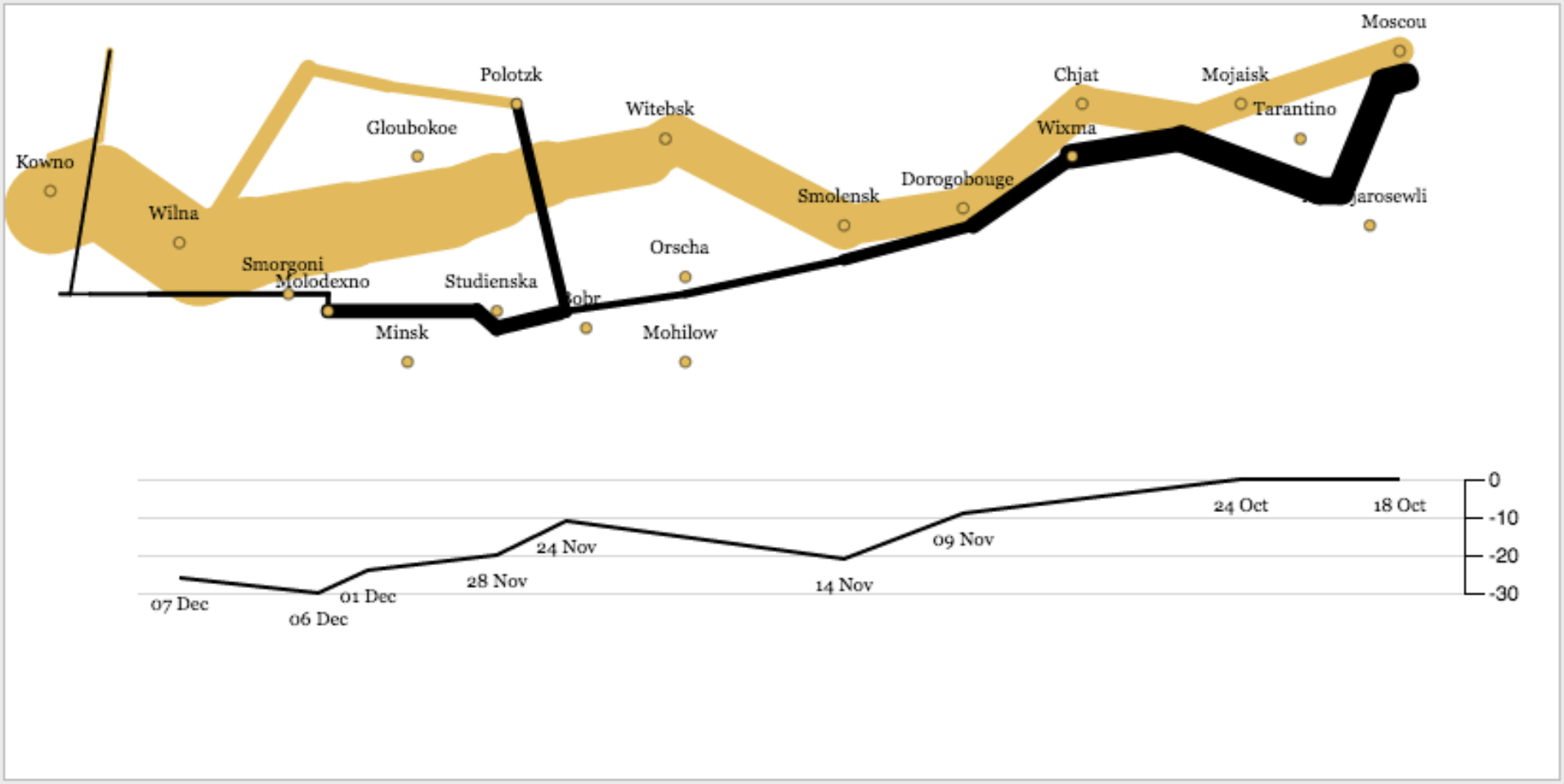
Rose-petal diagram

Mortality of the British Army

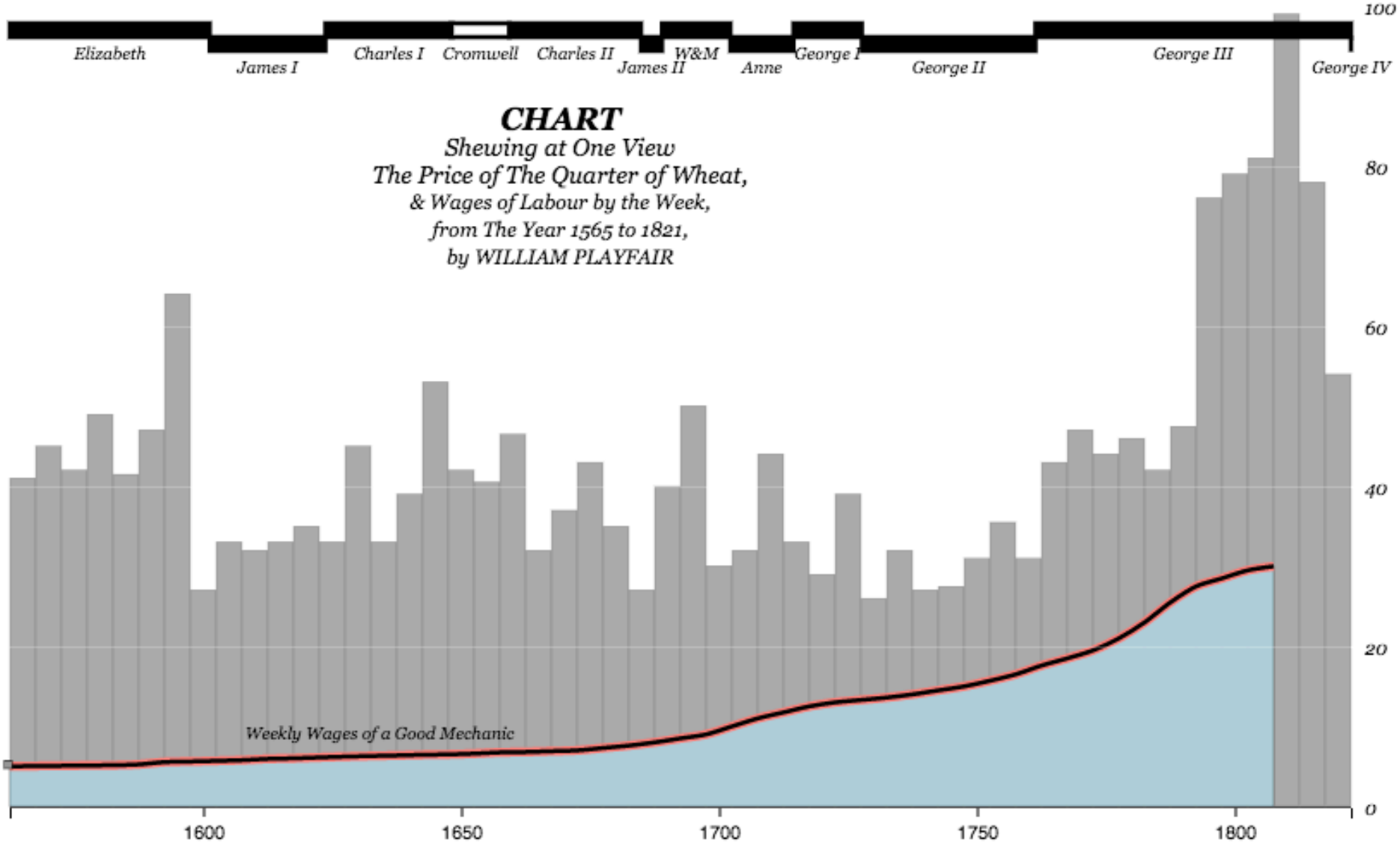


Florence Nightingale's diagram showing the dramatic reduction in death rates in the hospitals of Scutari following the changes she introduced

Source: Nightingale (1858)



http://idl.cs.washington.edu/projects/lyra/app/#?example=napoleon

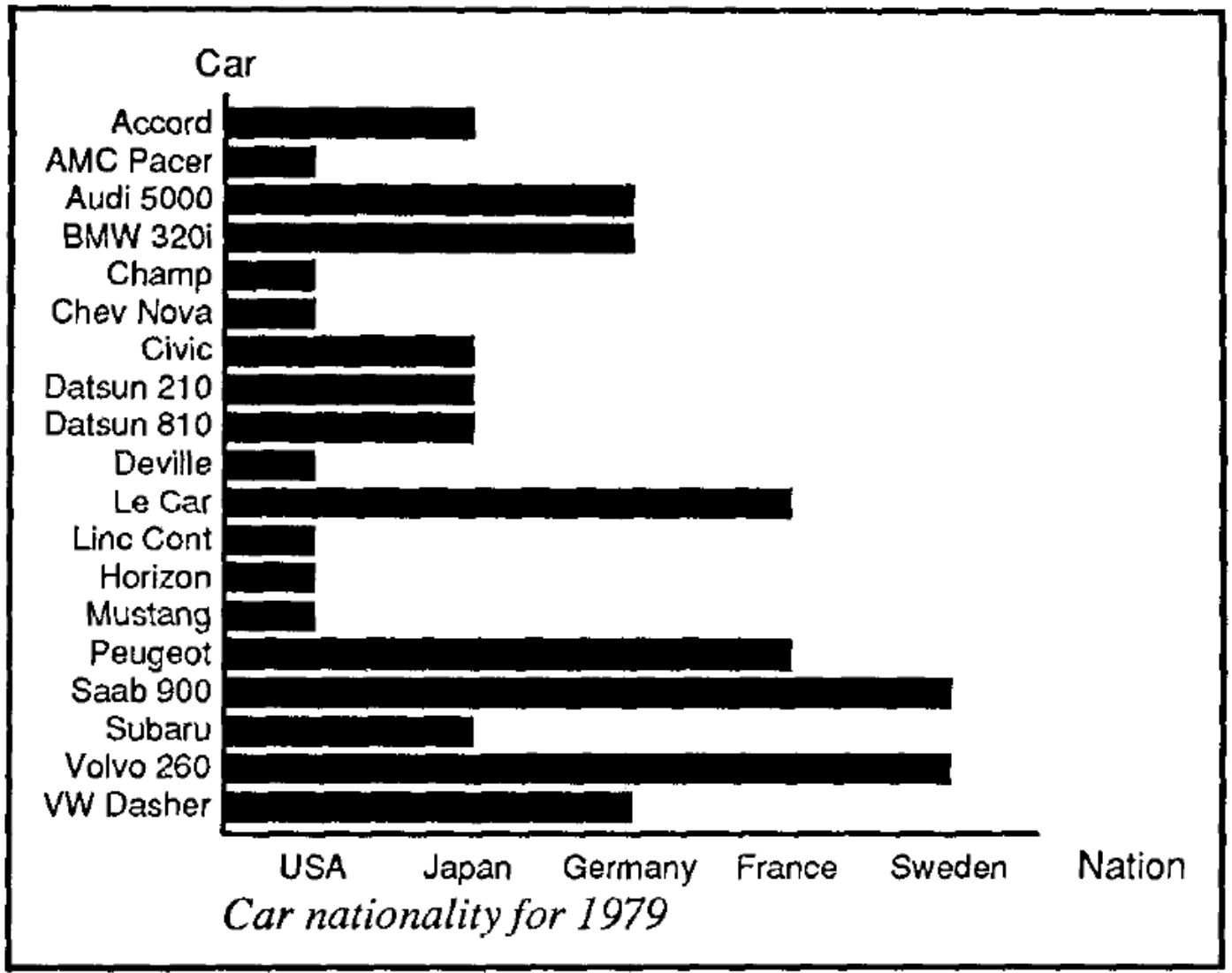


http://idl.cs.washington.edu/projects/lyra/app/#?example=playfair

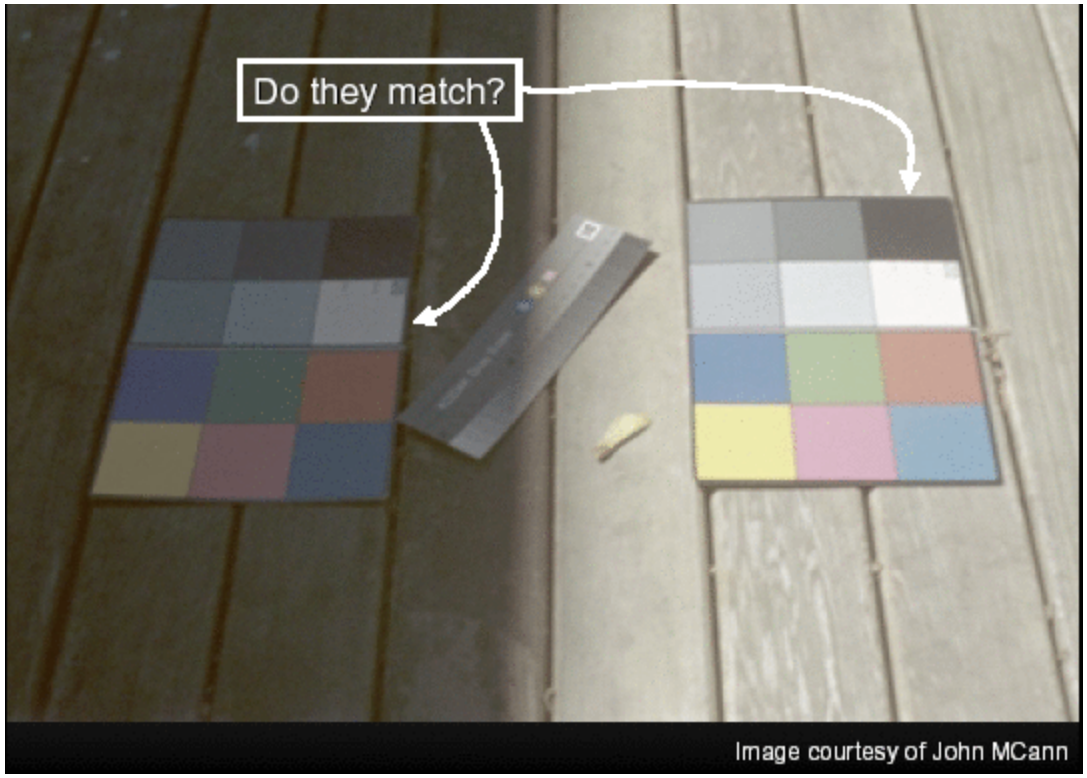
Perception for InfoVis

- Visual Perception
- Visual Encodings of Quantitative Data
- Data Types and Tasks for InfoVis

Any better encodings?

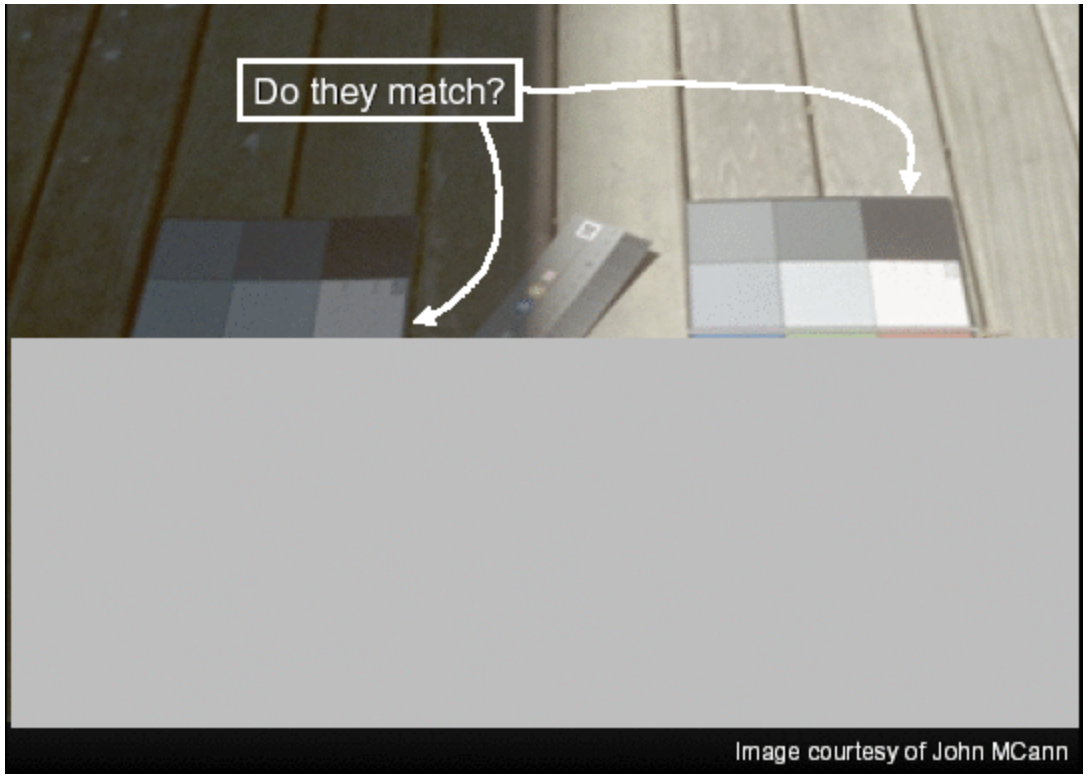


Relative Perception



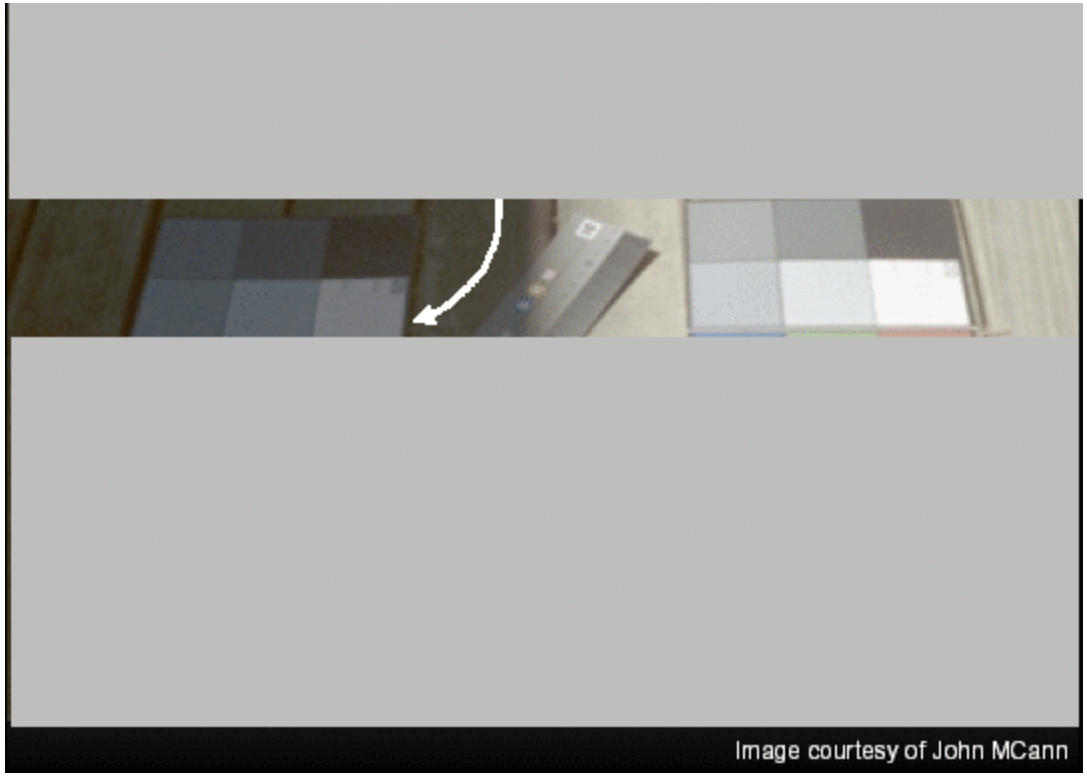
courtesy of John McCann, from Stone 2001 SIGGRAPH course graphics.stanford.edu/courses/cs448b-02-spring/04cdrom.pdf

Relative Perception



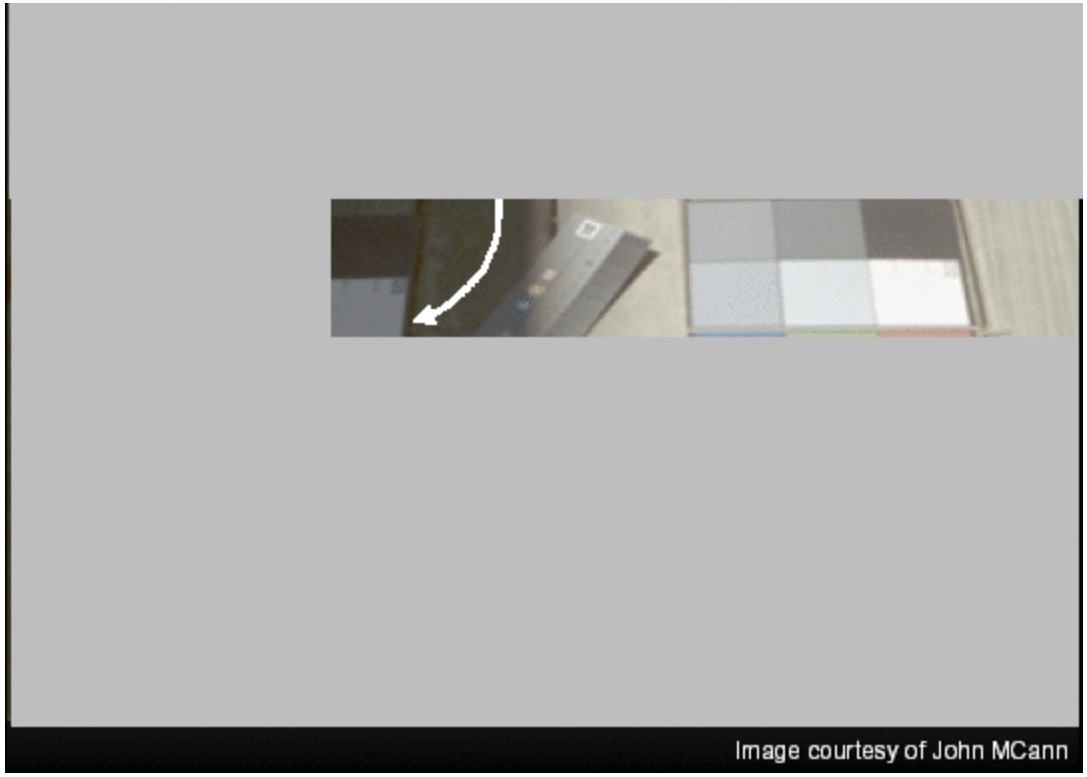
courtesy of John McCann, from Stone 2001 SIGGRAPH course graphics.stanford.edu/courses/cs448b-02-spring/04cdrom.pdf

Relative Perception



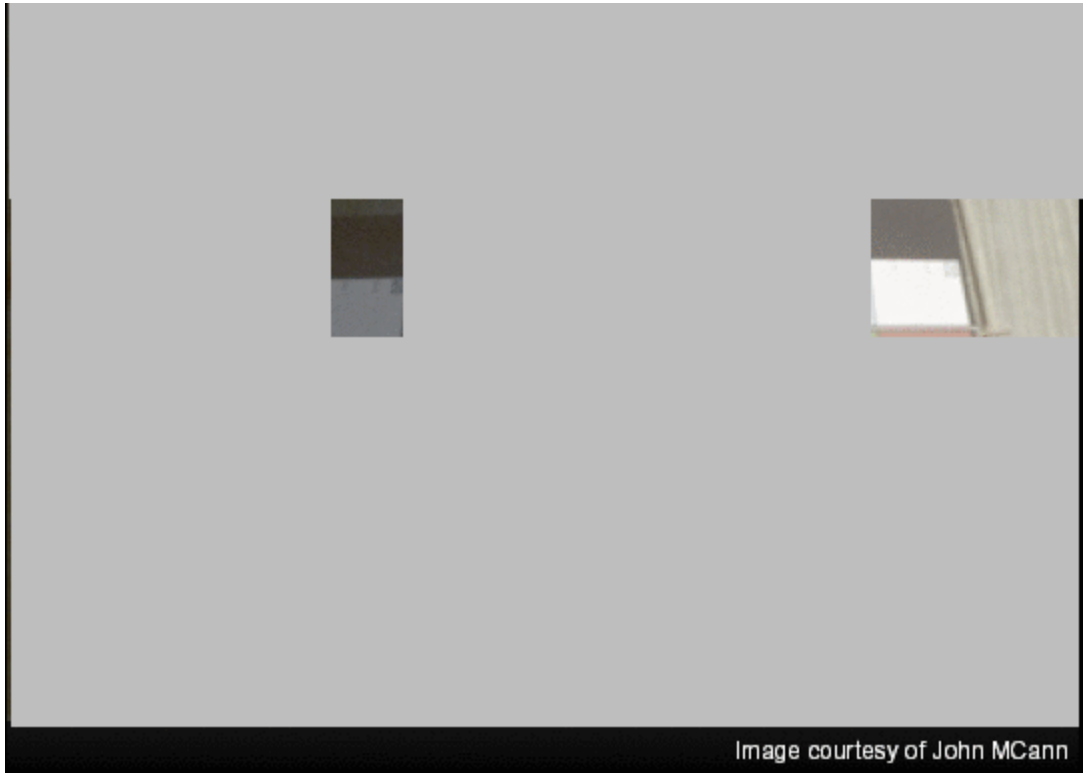
courtesy of John McCann, from Stone 2001 SIGGRAPH course graphics.stanford.edu/courses/cs448b-02-spring/04cdrom.pdf

Relative Perception



courtesy of John McCann, from Stone 2001 SIGGRAPH course graphics.stanford.edu/courses/cs448b-02-spring/04cdrom.pdf

Relative Perception



courtesy of John McCann, from Stone 2001 SIGGRAPH course graphics.stanford.edu/courses/cs448b-02-spring/04cdrom.pdf

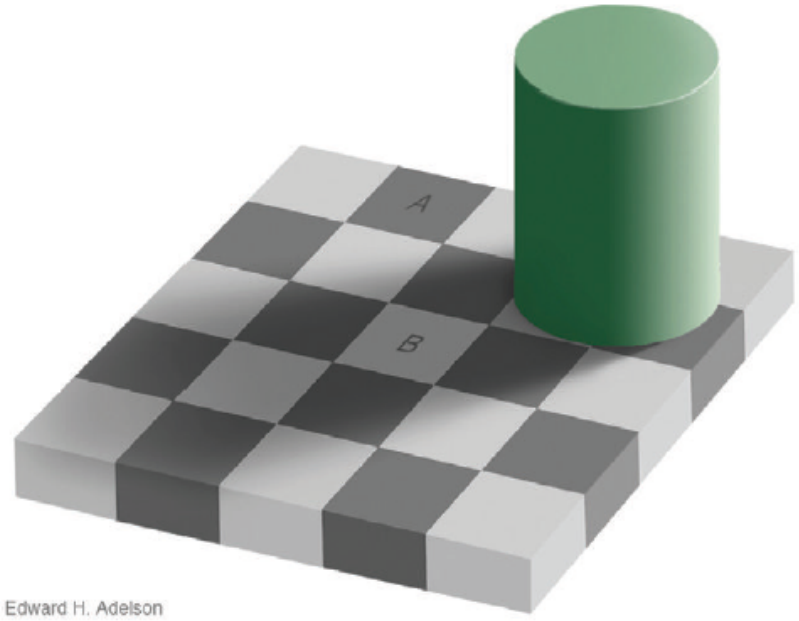
Relative Perception

courtesy of John McCann, from Stone 2001 SIGGRAPH course graphics.stanford.edu/courses/cs448b-02-spring/04cdrom.pdf

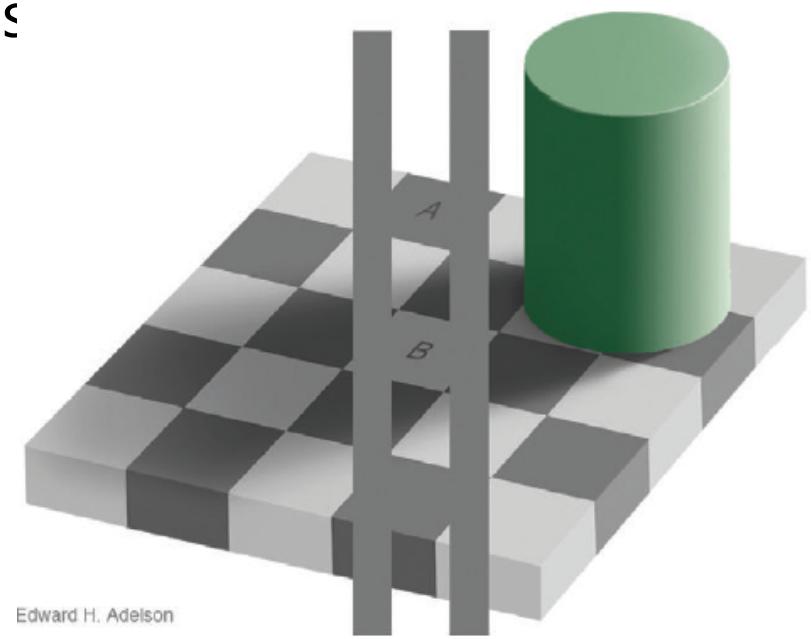
Relative versus Absolute Judgements

- Luminance contrast – Simultaneous Brightness Contra st

as



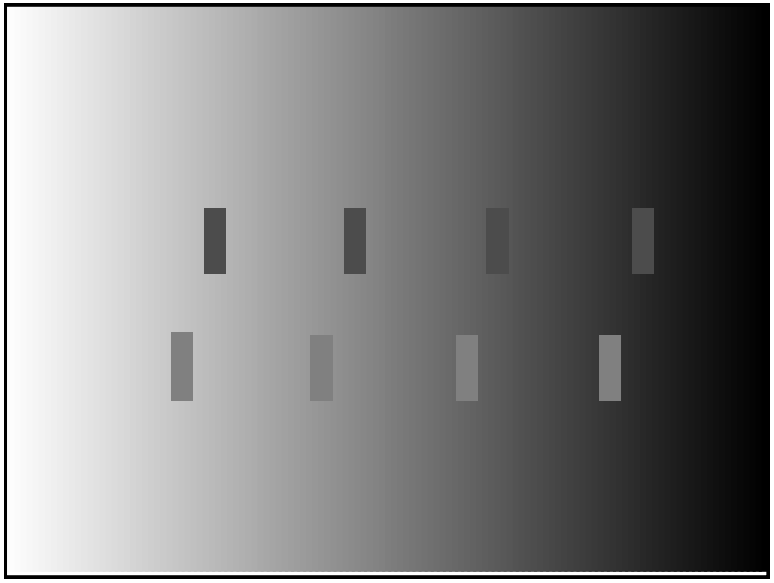
(a)

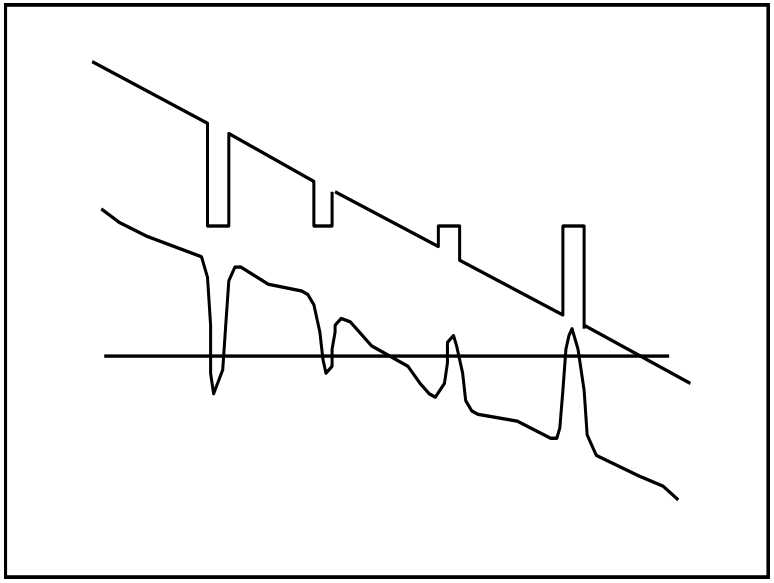


(b)

Luminance contrast
– Simultaneous Brightness Contrast

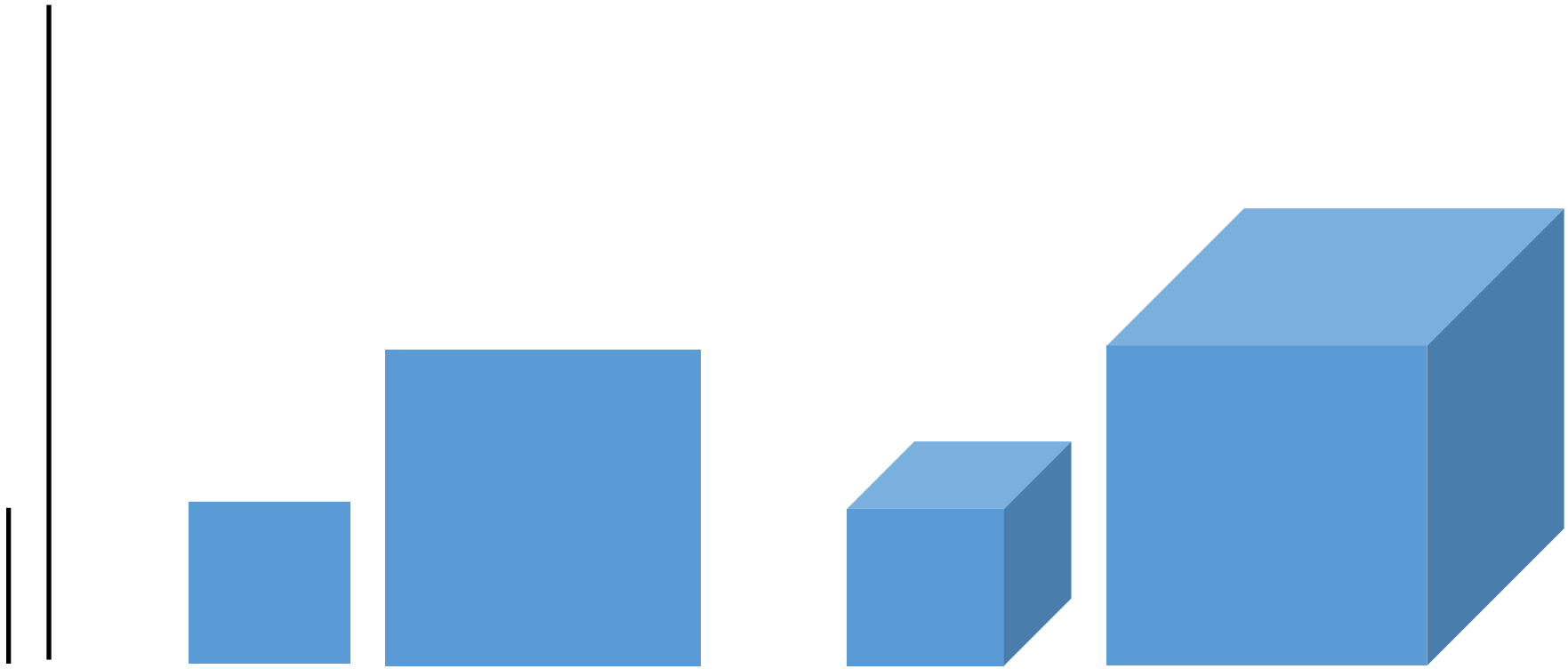
- A gray patch placed on a dark back ground looks lighter than the same gray patch on a light background





1D, 2D, 3D

- size ratio for each pair ➔ 1:4



INTERACTIVE DATA VISUALIZATION: FOUNDATIONS, TECHNIQUES, AND APPLICATIONS
Matthew O. Ward; Georges Grinstein; Daniel Keim, A K Peters Ltd (July 1, 2010)

Stevens' Power Law

- p: perceived magnitude
- a: actual magnitude

- $p = ka^{\alpha}$

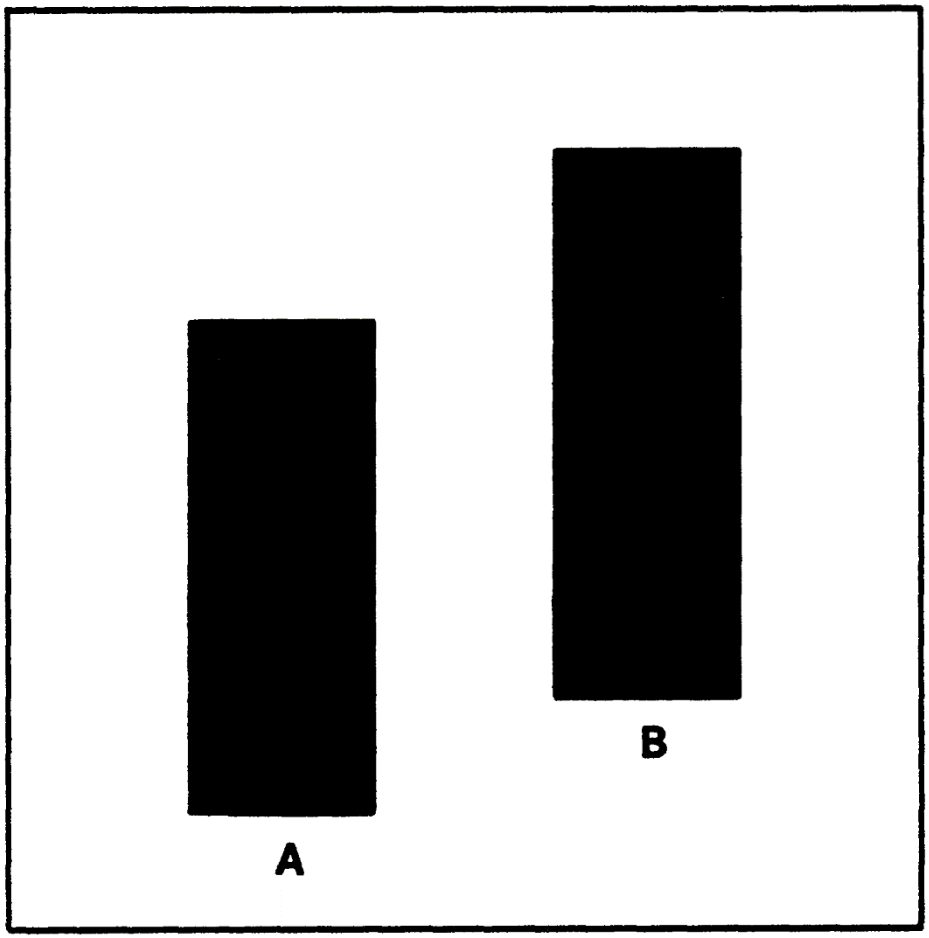
- $p_1/p_2 = (a_1/a_2)^{\alpha}$

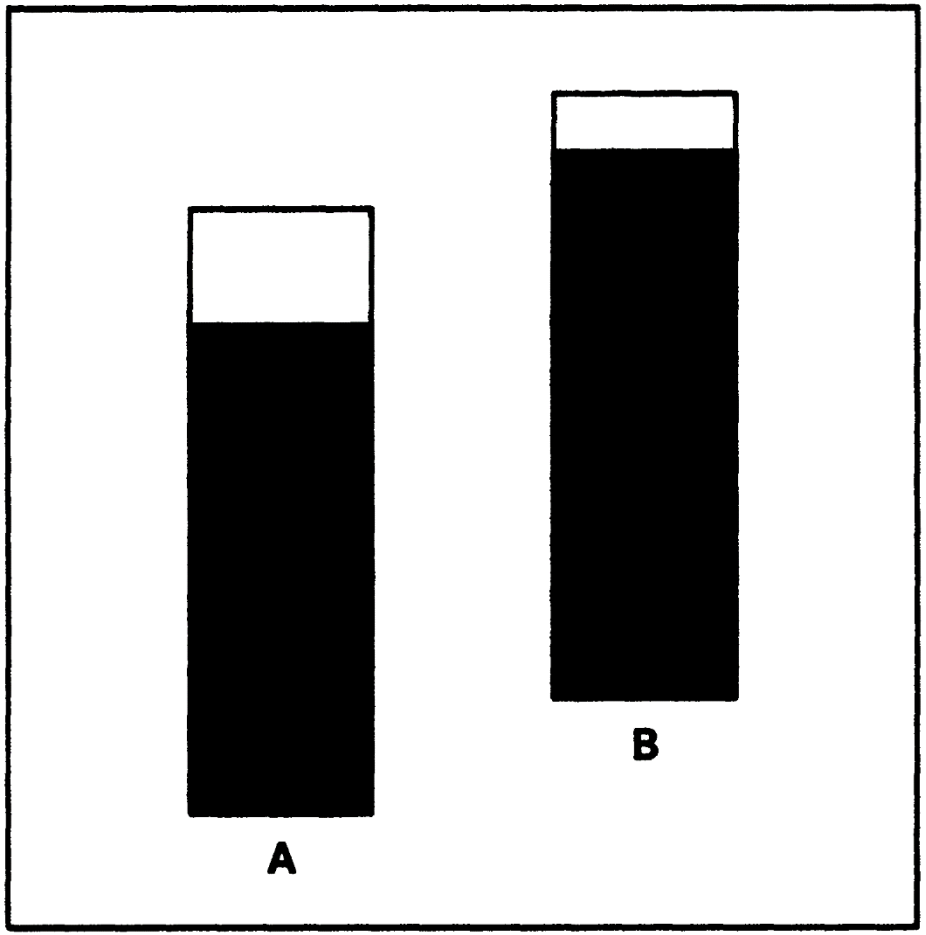
- length judgment: $\alpha \approx 1$
- area judgment: $\alpha < 1$
- volume judgment: $\alpha \ll 1$

http://en.wikipedia.org/wiki/Stevens'_power_law

Continuum	Exponent (a)	Stimulus condition
Loudness	0.67	Sound pressure of 3000 Hz tone
Vibration	0.95	Amplitude of 60 Hz on finger
Vibration	0.6	Amplitude of 250 Hz on finger
Brightness	0.33	5° target in dark
Brightness	0.5	Point source
Brightness	0.5	Brief flash
Brightness	1	Point source briefly flashed
Lightness	1.2	Reflectance of gray papers
Visual length	1	Projected line
Visual area	0.7	Projected square
Redness (saturation)	1.7	Red-gray mixture
Taste	1.3	Sucrose
Taste	1.4	Salt
Taste	0.8	Saccharin
Smell	0.6	Heptane
Cold	1	Metal contact on arm
Warmth	1.6	Metal contact on arm
Warmth	1.3	Irradiation of skin, small area
Warmth	0.7	Irradiation of skin, large area
Discomfort, cold	1.7	Whole body irradiation
Discomfort, warm	0.7	Whole body irradiation
Thermal pain	1	Radiant heat on skin
Tactual roughness	1.5	Rubbing emery cloths
Tactual hardness	0.8	Squeezing rubber
Finger span	1.3	Thickness of blocks
Pressure on palm	1.1	Static force on skin
Muscle force	1.7	Static contractions
Heaviness	1.45	Lifted weights
Viscosity	0.42	Stirring silicone fluids
Electric shock	3.5	Current through fingers
Vocal effort	1.1	Vocal sound pressure
Angular acceleration	1.4	5 s rotation
Duration	1.1	White noise stimuli

Which is longer?

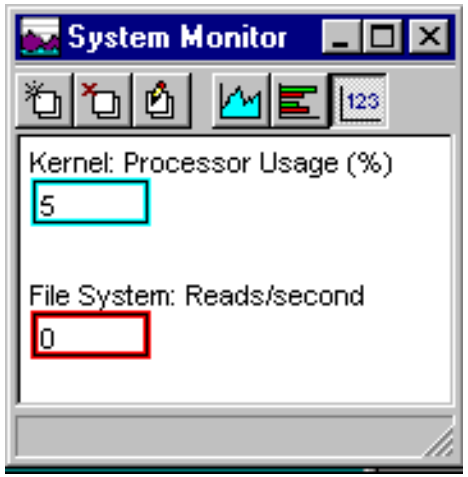




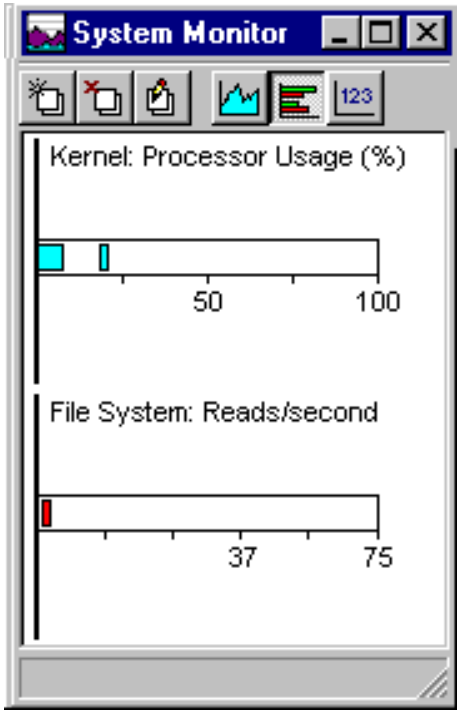
Graphical Perception: Theory, Experimentation and the Application to the Development of Graphical Models. William S. Cleveland, Robert McGill, J. Am. Stat. Assoc. 79:387, pp. 531–554, 1984.

Which representation is best?

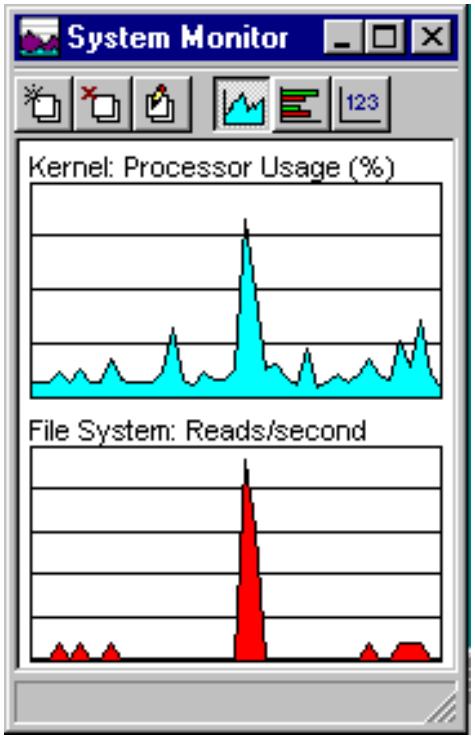
- depends heavily on task



What is precise value?



How does the performance now compared to its peak?

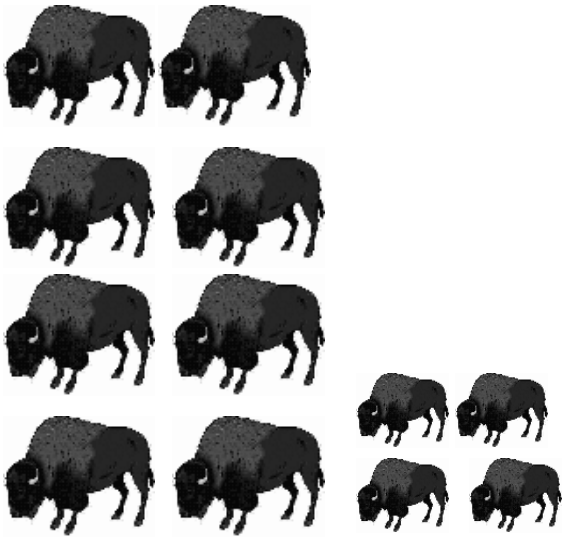


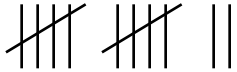
How does performance change over time?

Windows 95 System Monitor

Slide idea from Saul Greenberg

How many buffalo?





 # Buffalo

 # Buffalo

 # Adults

 # calfs

8

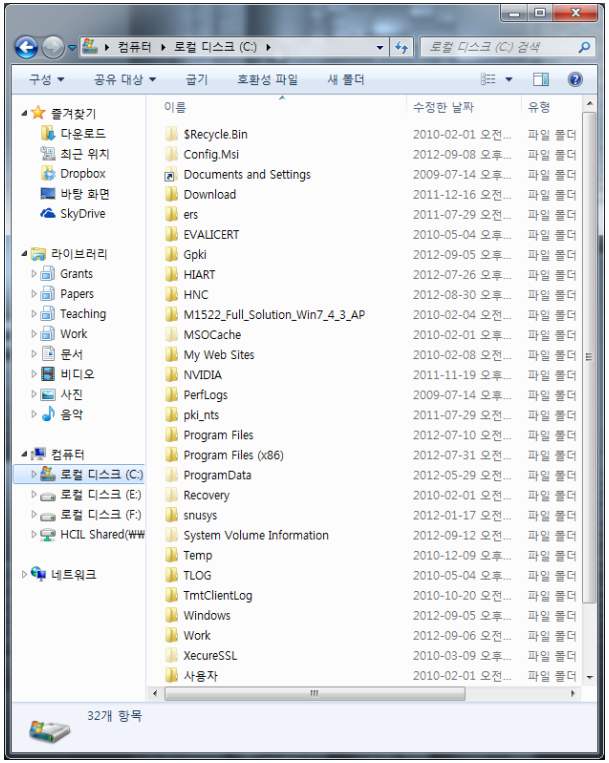
 # Adults

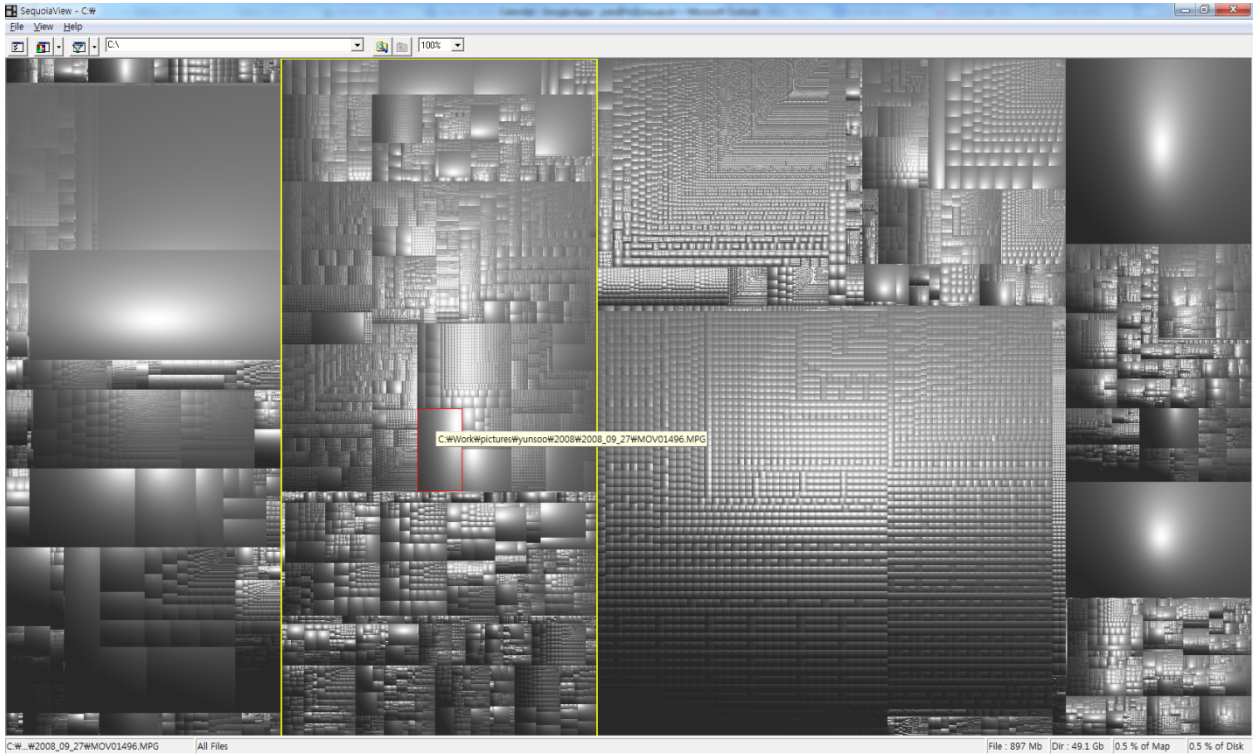
4

 # calfs

Slide idea from Saul Greenberg

Am I waiting my hdd space?





http://w3.win.tue.nl/nl/onderzoek/onderzoek_informatica/visualization/sequoiaview//

Mapping

- input
  - data semantics
  - use domain knowledge
- output
  - visual encoding
    - visual/graphical/perceptual/retinal
    - channels/attributes/dimensions/variables
  - use human perception
- processing
  - algorithms
  - handle computational constraints

Preattentive Processing

- Cognitive operations done preattentively, without the need for focused attention
  - less than 200-250 ms
    - eye movements take 200 ms
    - minimum time to initiate eye movement
    - involves only information available in a single glance

- Popout effects
- Segmentation effects

How many threes?

```
897390570927940579629765098294
080280850808308028098 50-802808
567847298872ty458202094757720 0
2178984389 0r455790456099272188
897594797902855892594573979209
```

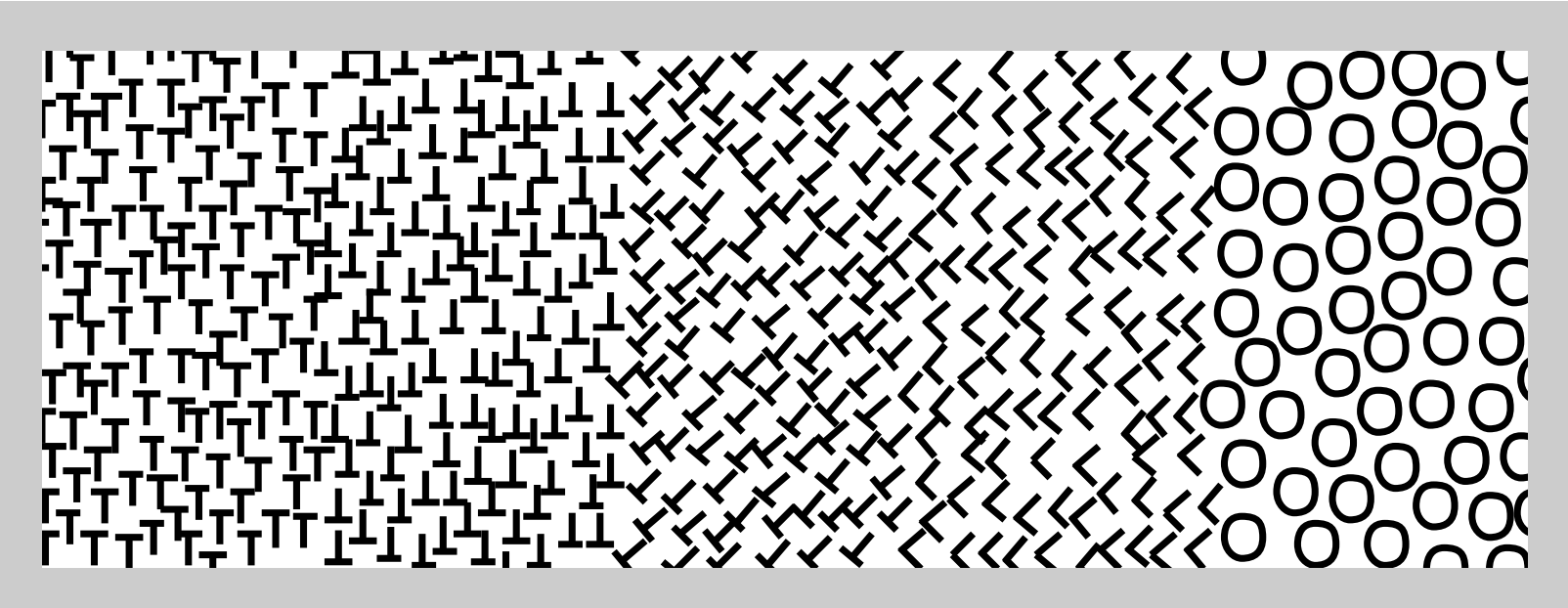
Colin Ware

Color Makes Them Pop Out

```
897390570927940579629765098294
080280850808308028098 50-802808
5678472988 72ty458202094757720 0
2178984 3890r455790456099272188
897594797902855892594573979209
```

Colin Ware

Segmentation

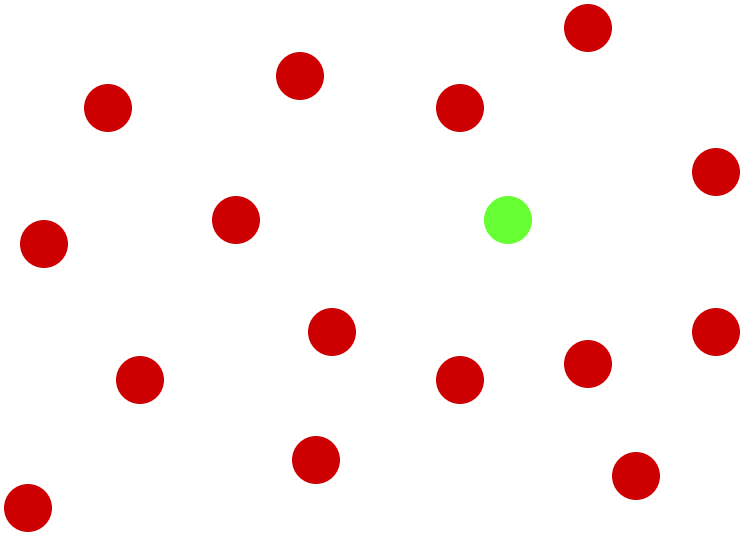


Colin Ware

What Kinds of Tasks?

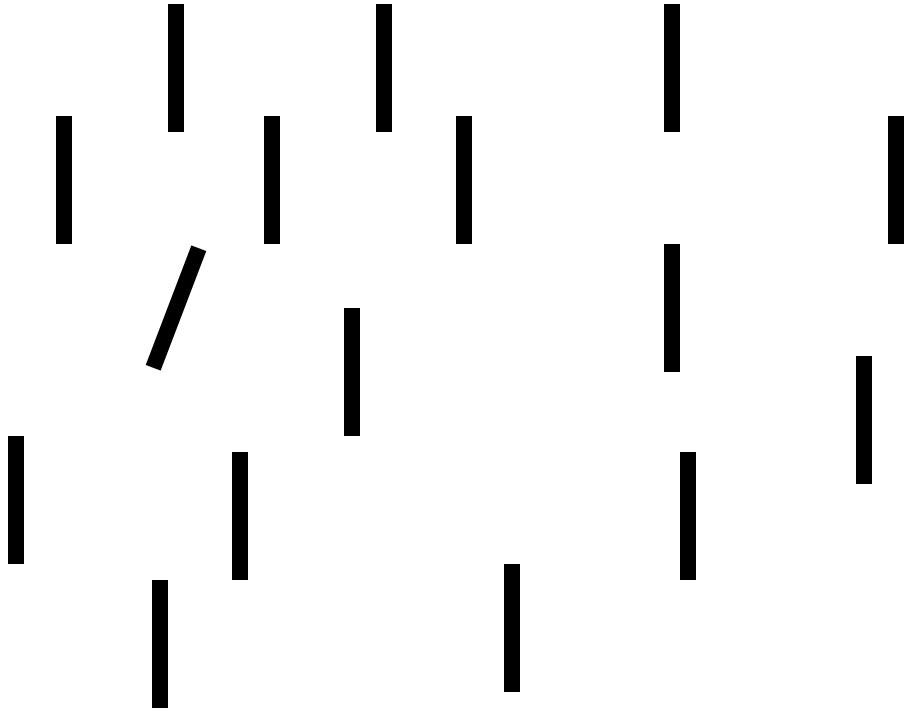
- Target detection
  - Is something there?
- Boundary detection
  - Can the elements be grouped?
- Region tracking
  - Can a distinctive moving group be traced?
- Counting
  - How many elements of a certain type are present?

Color - Hue



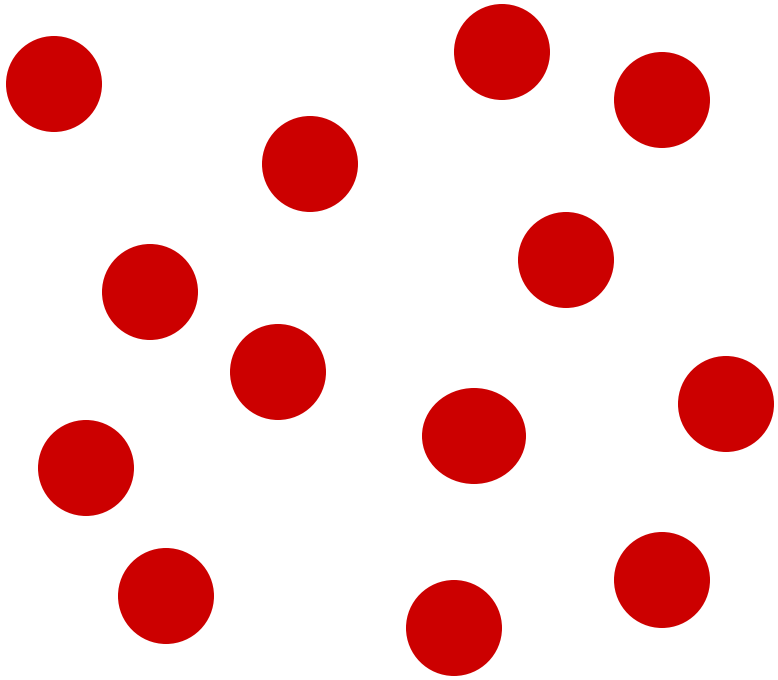
Colin Ware

Orientation



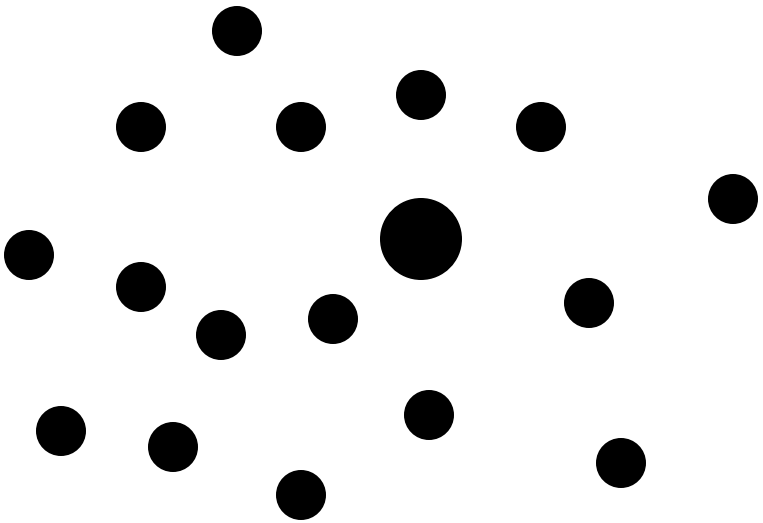
Colin Ware

Motion

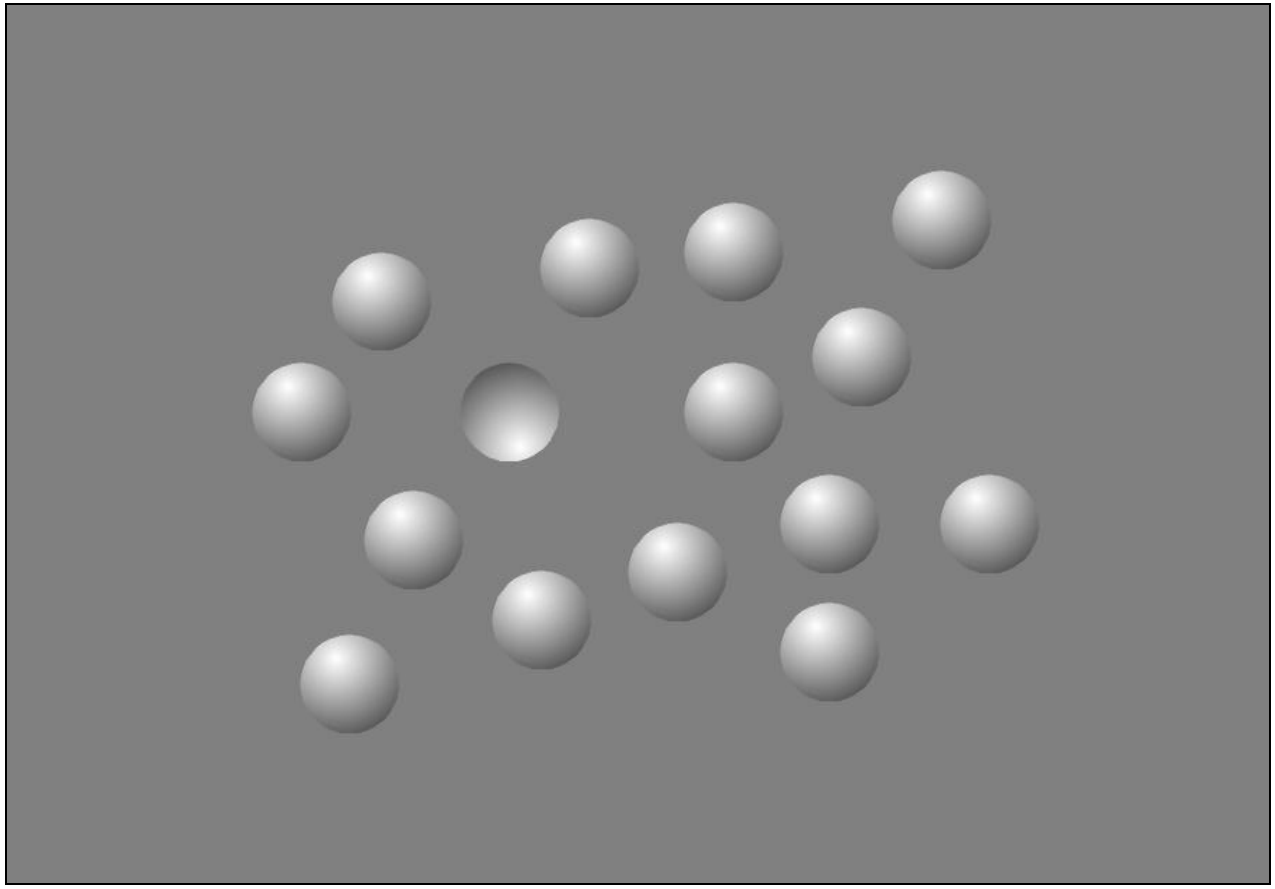


Colin Ware

Size

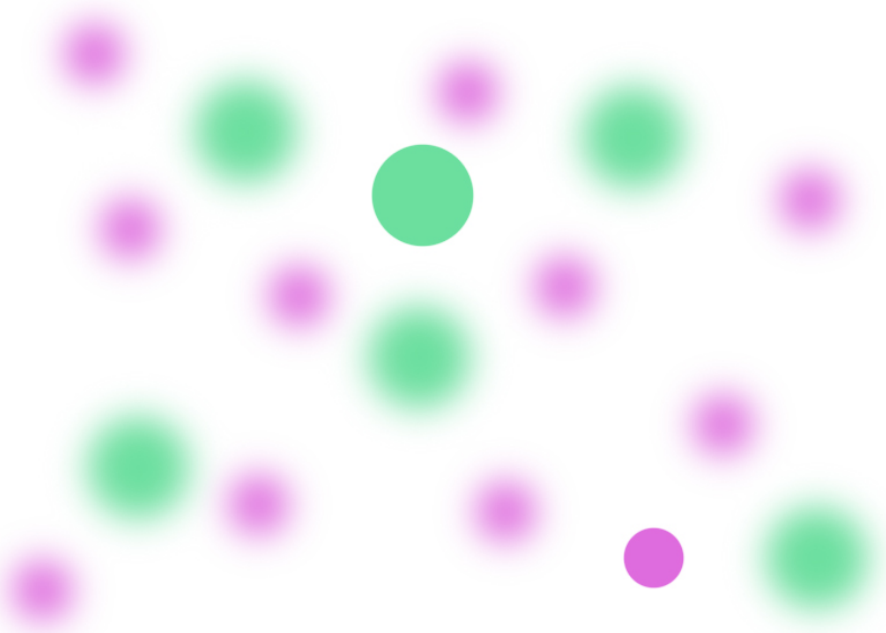


Simple shading



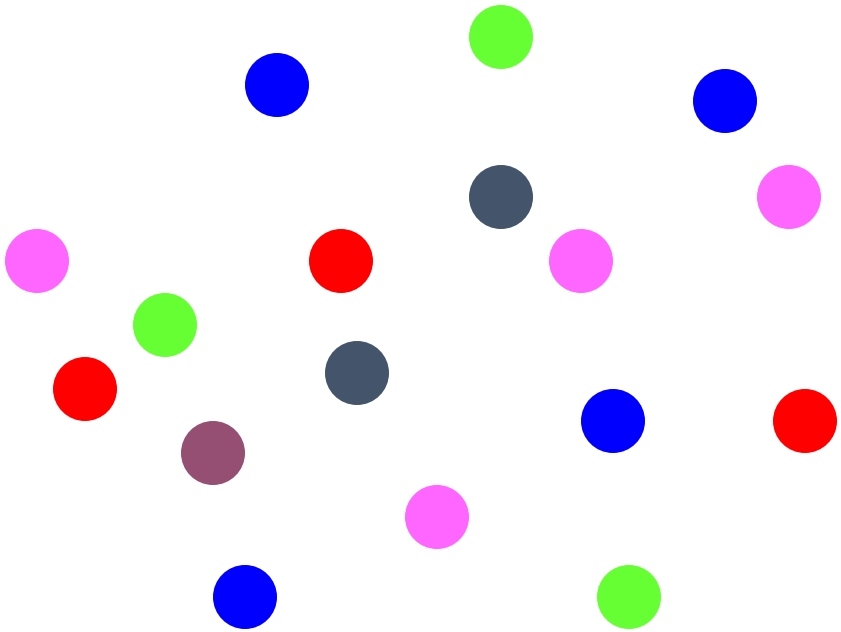
Colin Ware

Semantic Depth of Field



Colin Ware

Surrounded colors do not pop out



Colin Ware

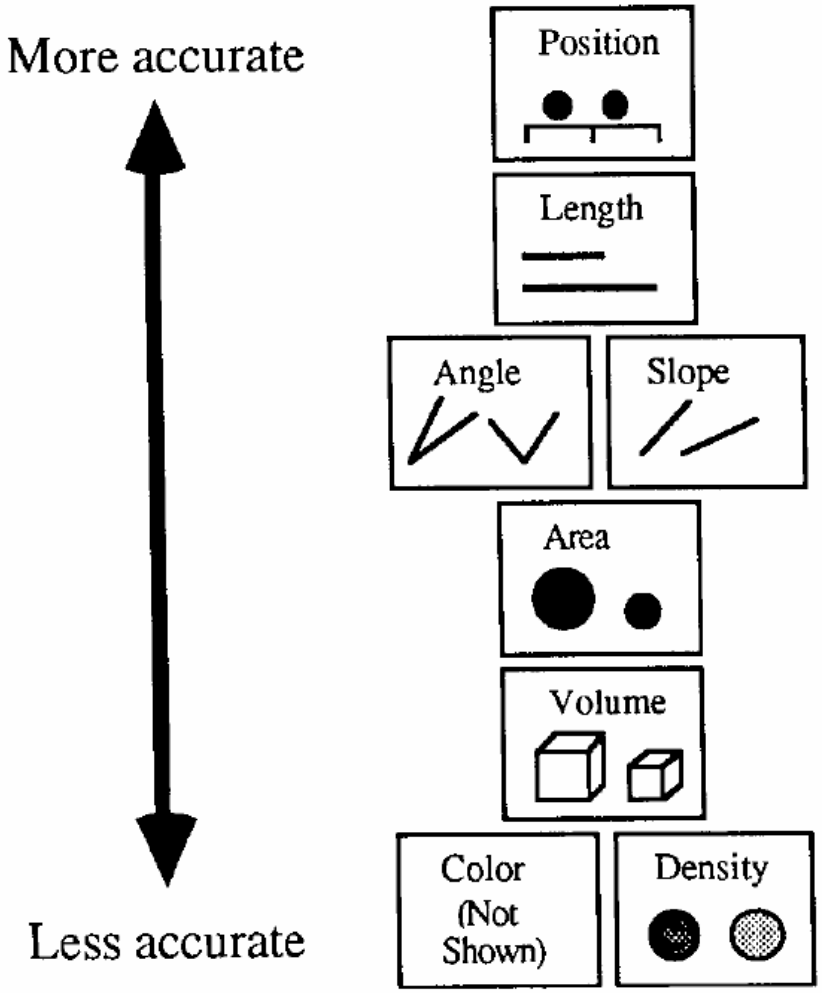
Laws of preattentive display

- Must stand out on some simple dimension
  - color,
  - simple shape = orientation, size
  - motion,
  - depth

- Lessons for highlighting – one of each

Visual Encoding

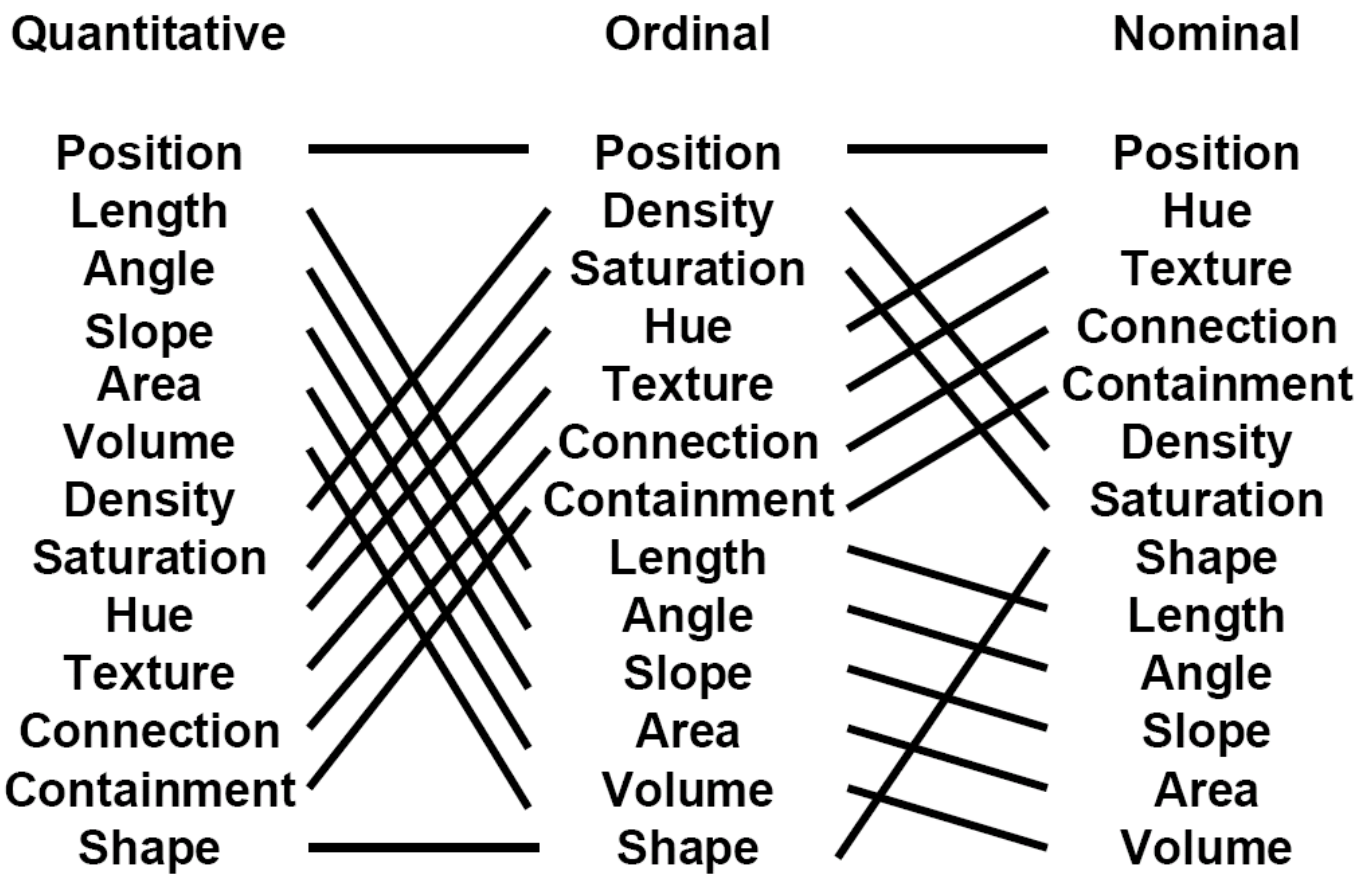
Accuracy of Quantitative Perceptual Tasks



Cleveland & McGill 1984

Visual Encoding Principles

- Channel Ranking Varies by Data Type

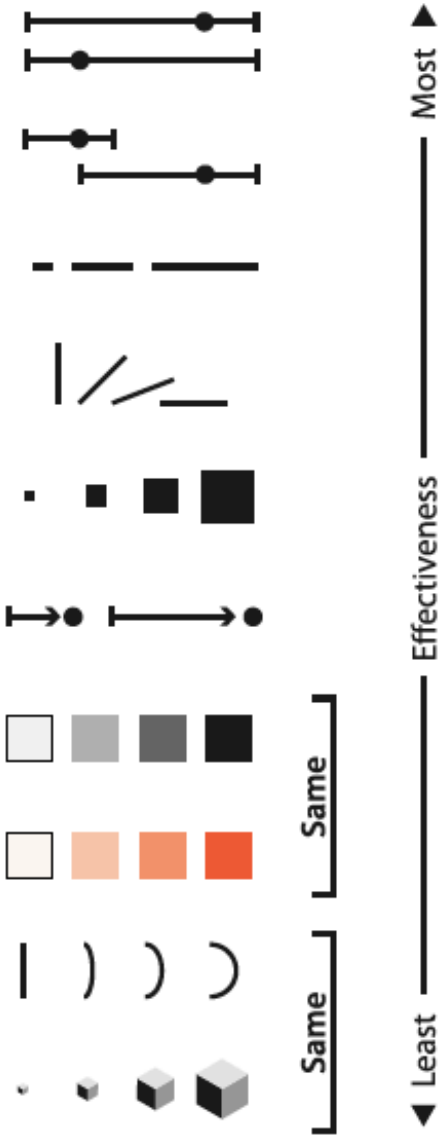


Automating the Design of Graphical Presentations of Relational Information ,Jock Mackinlay, ACM Transaction on Graphics, 1986

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered Attributes

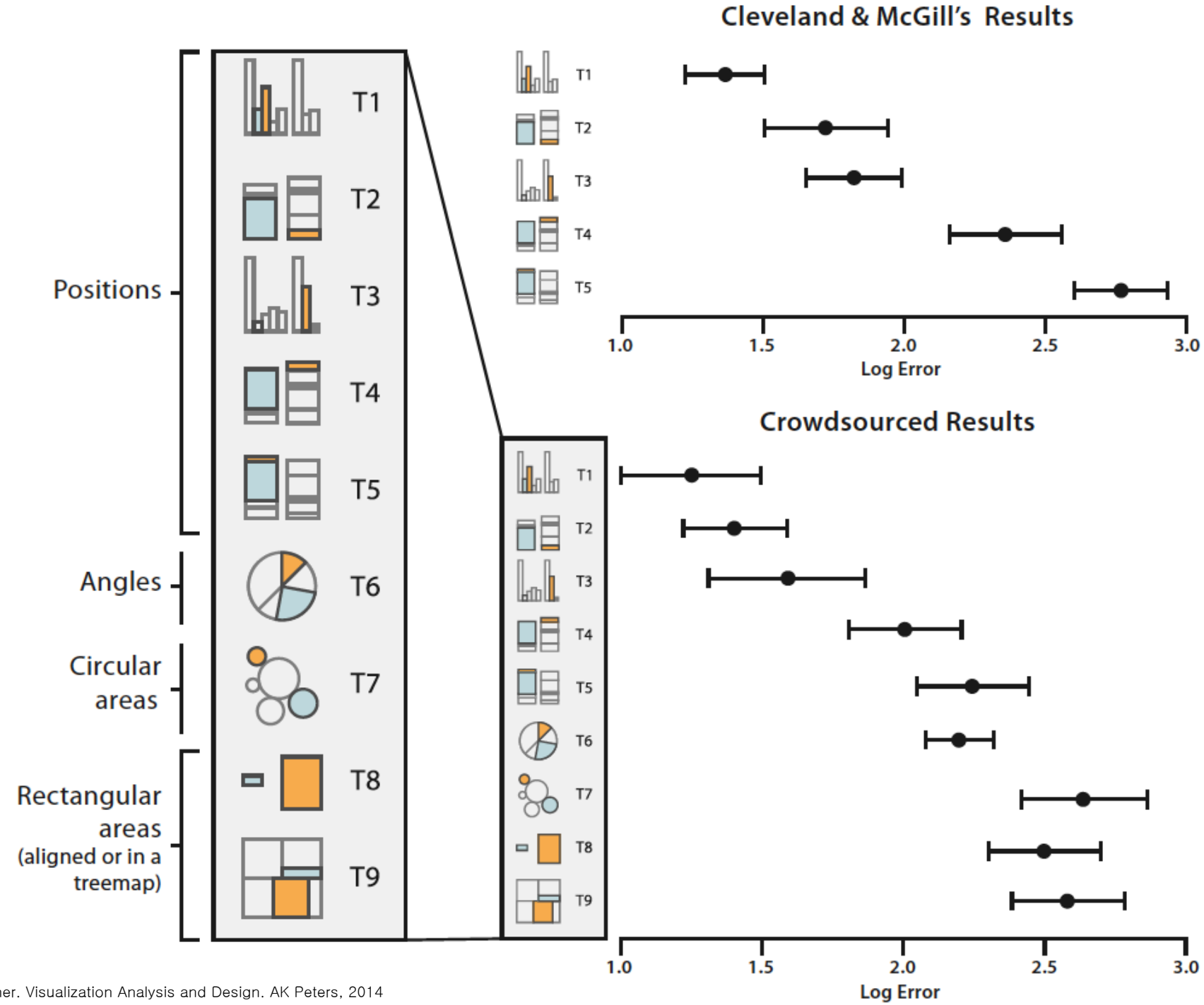
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)



Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

T. Munzner. Visualization Analysis and Design. AK Peters, 2014



T. Munzner. Visualization Analysis and Design. AK Peters, 2014

Data Types

- **1-D Linear** Document Lens, SeeSoft
- **2-D Map** GIS, ArcView, Google Map
- **3-D World** CAD, Medical
- **Multi-Dim** Parallel Coordinates, Spotfire, XGobi, Visage, Influence Explorer, TableLens
- **Temporal** Perspective Wall, LifeLines
- **Tree** Cone/Cam/Hyperbolic, Treemap
- **Network** Netmap, netViz, SeeNet, Butterfly, Multi-trees

Tasks

- **Overview** Gain an overview of the entire collection
- **Zoom** Zoom in on items of interest
- **Filter** Filter out uninteresting items
- **Details-on-demand**
  Select an item or group and get details when needed
- **Relate** View relationships among items
- **History** Keep a history of actions to support undo, replay, and progressive refinement
- **Extract** Allow extraction of sub-collections and of the query parameters

Design Principles

- Shneiderman's Guidelines
- Tufte's Design Principles
- The Feynman-Tufte Principle

Design Guidelines/Principles

- Visual presentation of query components
- Visual presentation of results

- Rapid, incremental and reversible actions
- Immediate and continuous feedback
- Selection by pointing (not typing)

- Reduces errors
- Encourages exploration

- Visual Information Seeking Mantra
  - Overview first, zoom and filter, details on demand

Tufte's Design Principles

- Tell the truth
  - Graphical integrity
- Do it effectively with clarity, precision…
  - Design principles/aesthetics

- "**simple design, intense content**"
  - The Feynman-Tufte Principle, April 2005 *Scientific American*

E. Tufte, *The Visual Display of Quantitative Information* (1983)
E. Tufte, *Envisioning Information* (1990)
E. Tufte, *Visual Explanations* (1997)
E. Tufte, *Beautiful Evidence*(2006)

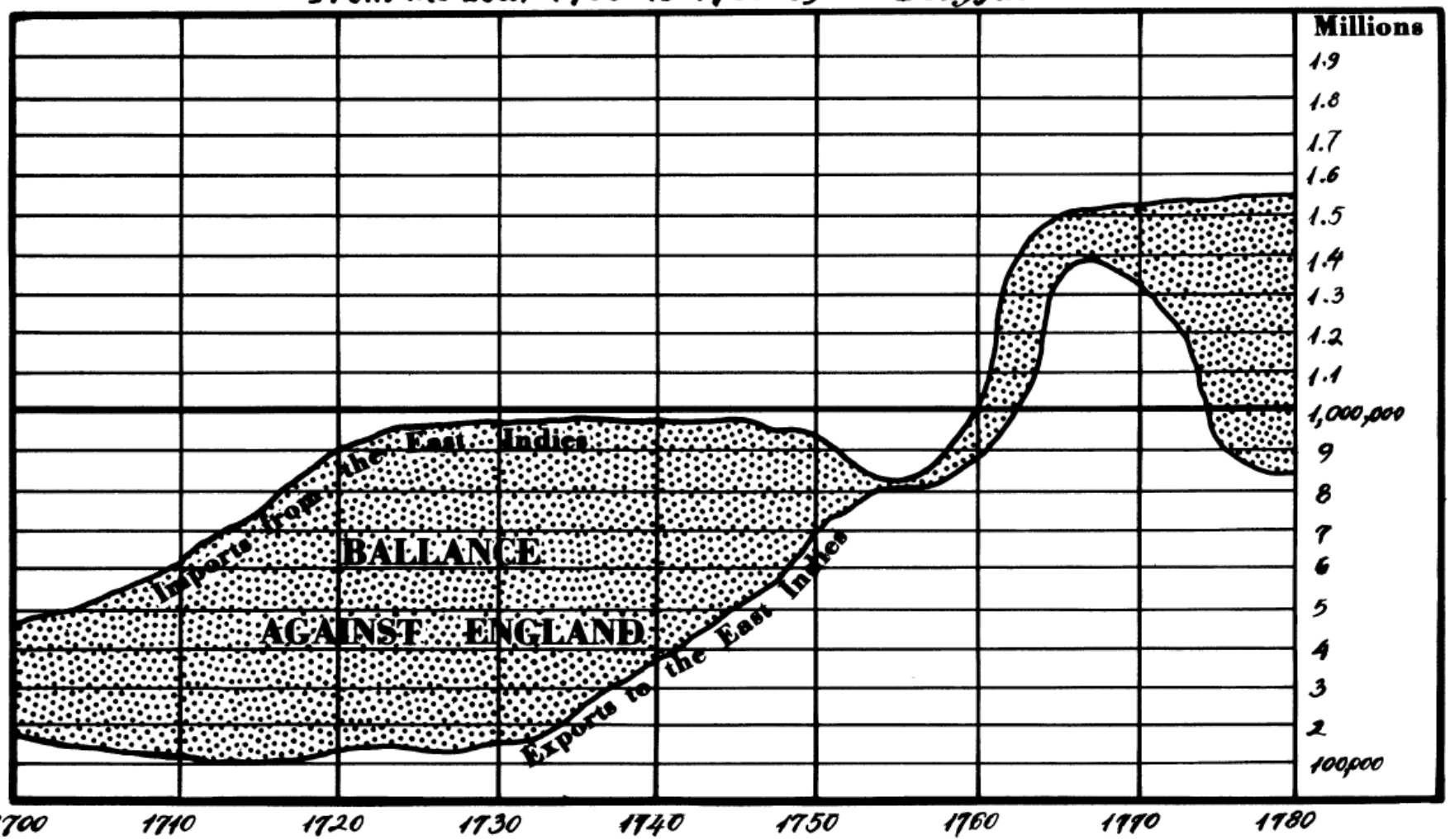
Graphical Integrity

- Your graphic should tell the truth about your data

- Consider visual perception
- User-centered vs. designer-centered

CHART *of* **EXPORTS** *and* **IMPORTS** *to and from the* **EAST INDIES**

From the Year 1700 to 1780 by W. Playfair

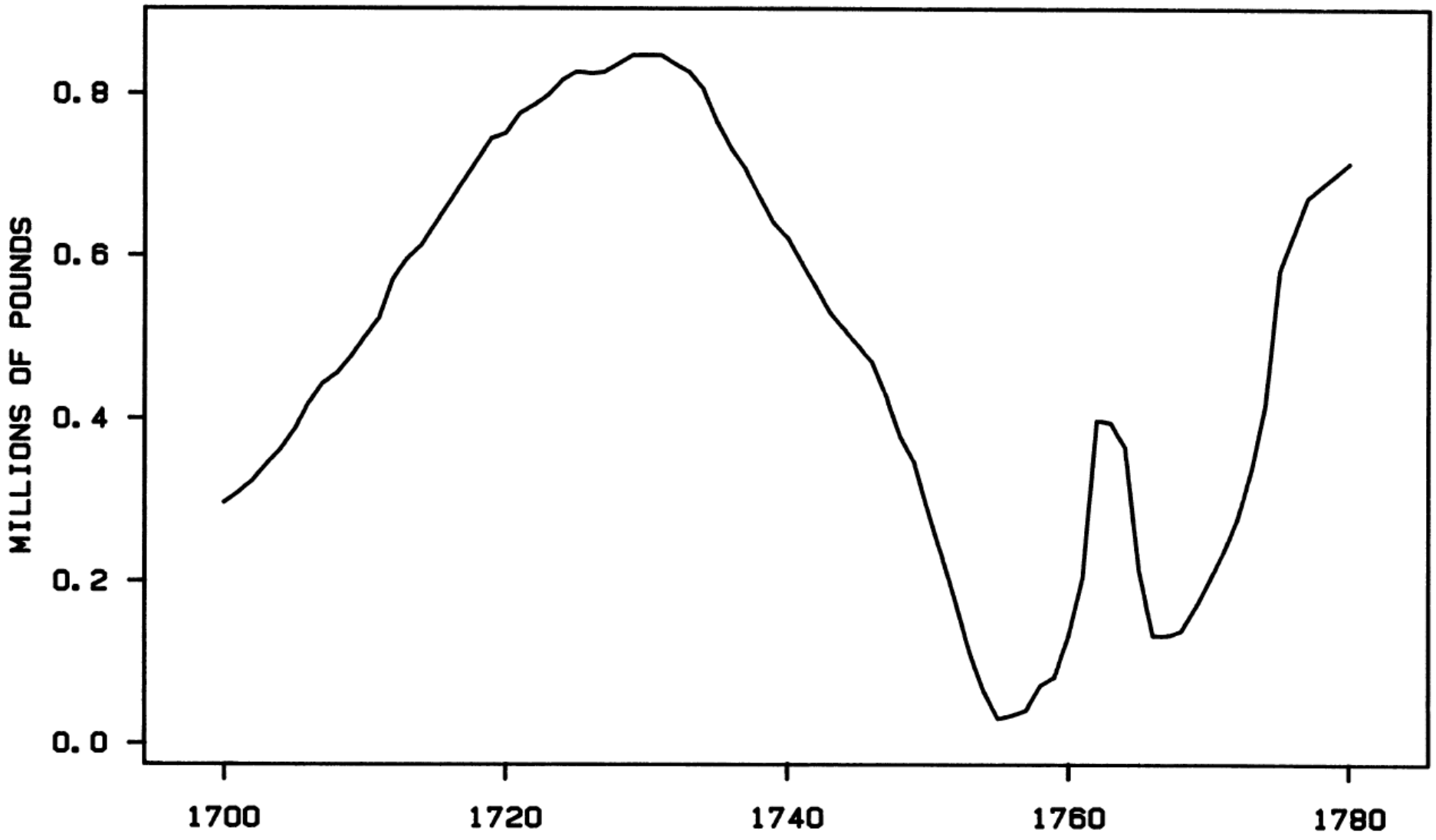


The Bottom Line is Divided into Years the Right hand Line into **HUNDRED THOUSAND POUNDS**

Sing the Knight

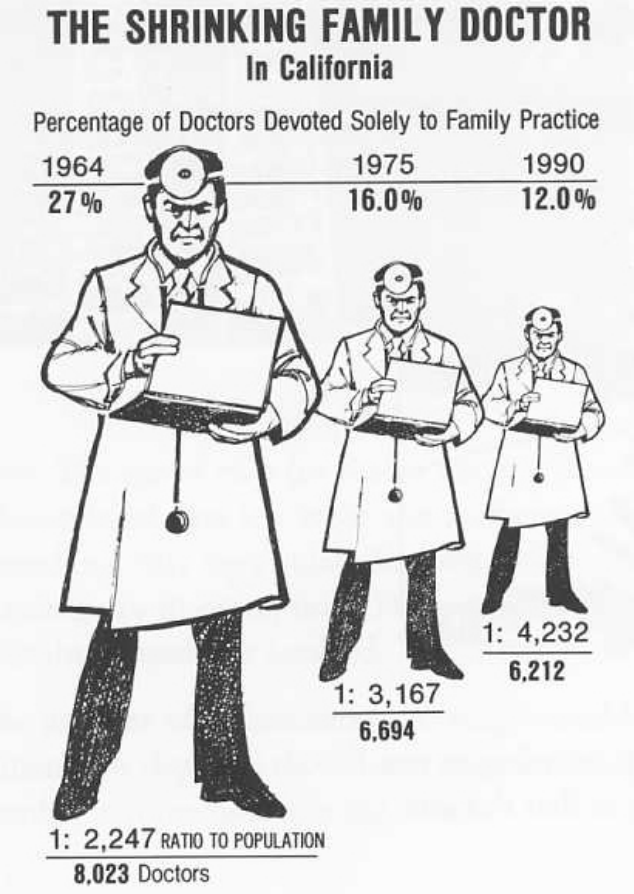
is face Page 31st

Published in the Act Divine 16th Aug. 1785


MILLIONS OF POUNDS
0.0
0.2
0.4
0.6
0.8
1700
1720
1740
1760
1780

Measuring Misrepresentation

- Visual attribute value should be directly proportional to data attribute value
- Height/width vs. area vs. volume



"Lie factor" = 2.8

$$\text{Lie factor} = \frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$$

Design Principles

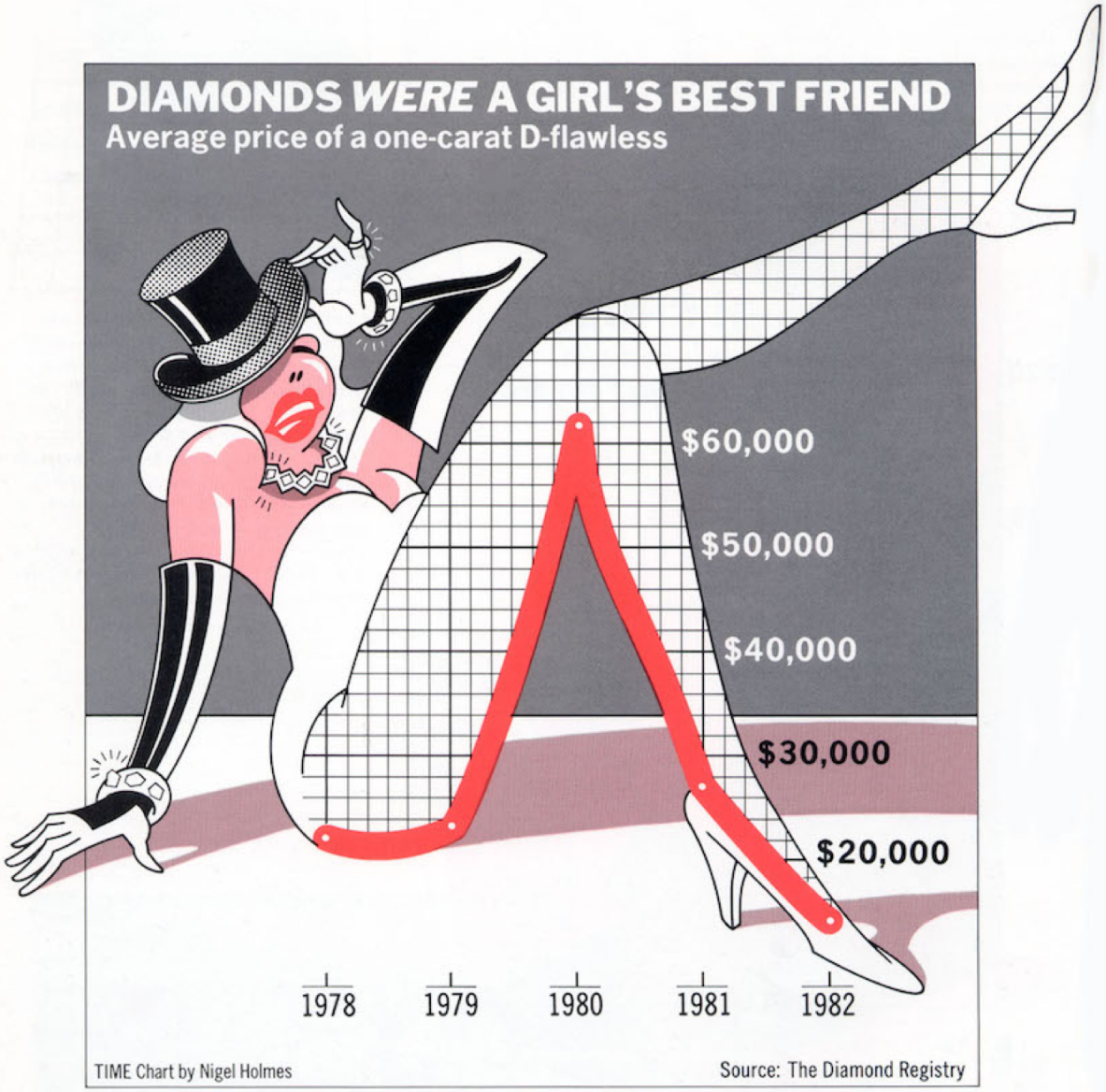
- Maximize data-ink ratio

$$\text{Data ink ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$

= proportion of graphic's ink devoted to the non-redundant display of data-information

Design Principles

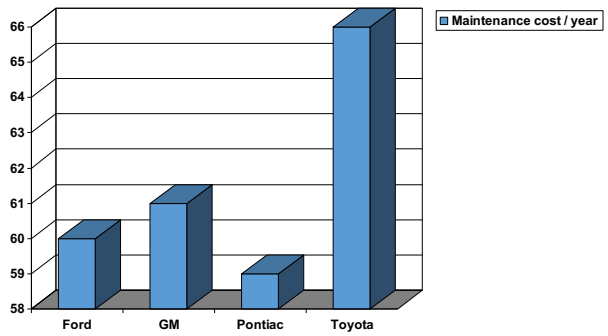
- Avoid chartjunk
  - Extraneous visual elements that detract from information

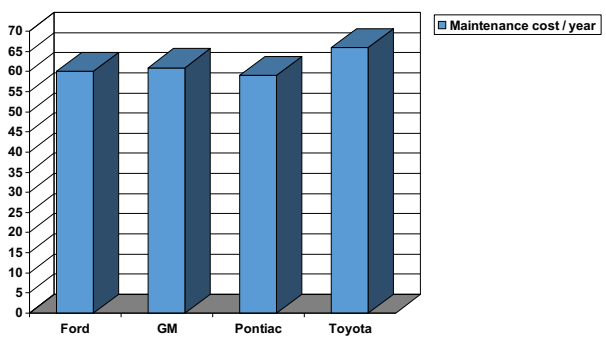


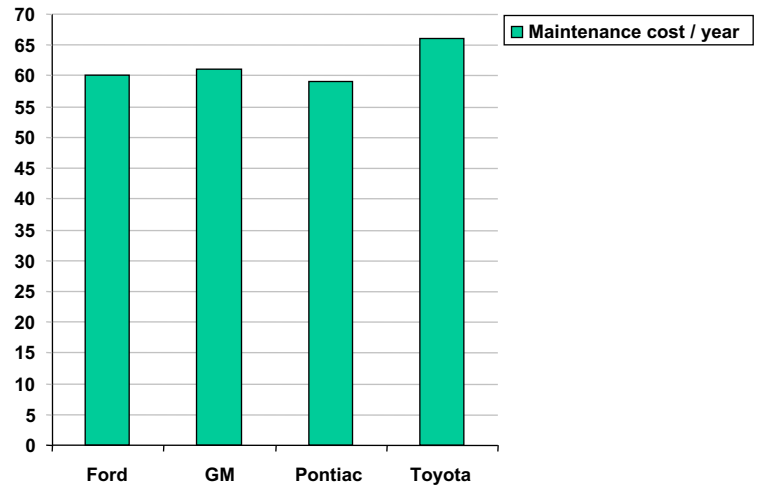
http://nigelholmes.com/

Chartjunk

all visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information

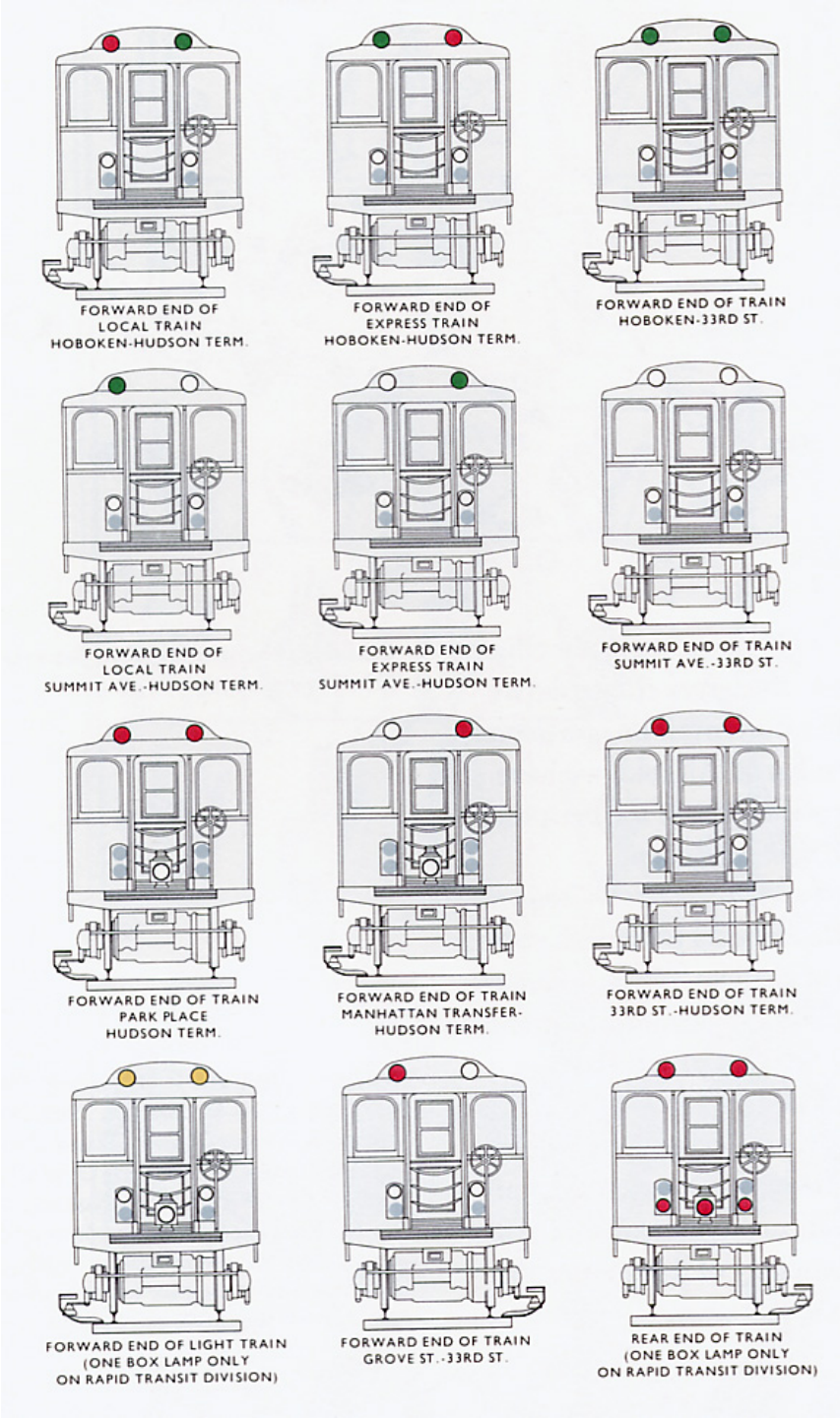






Design Principles

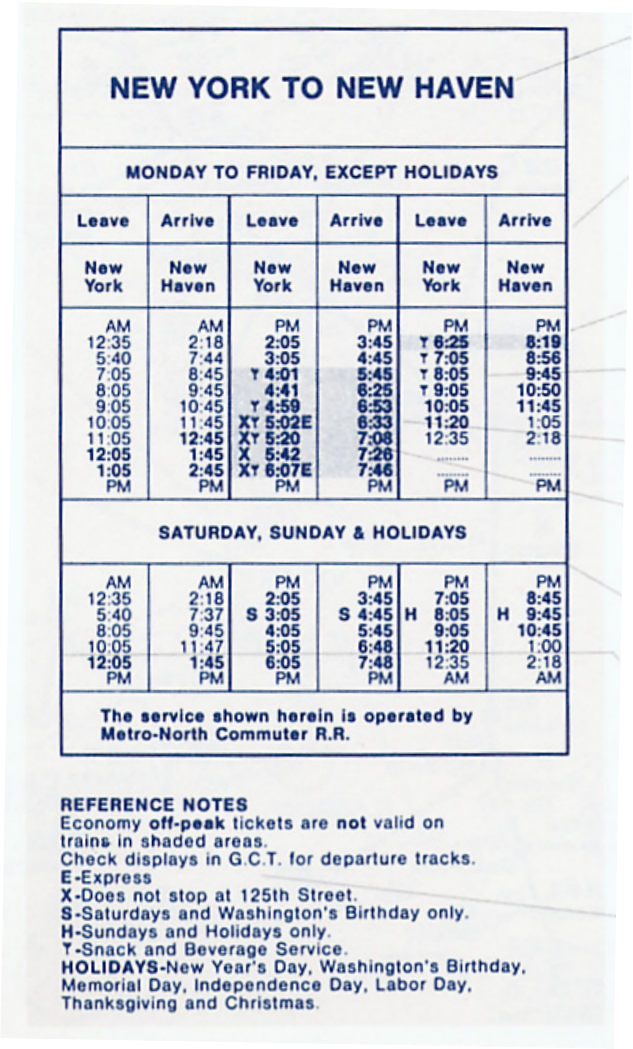
- Use **small multiples**
  - Repeat visually similar graphical elements nearby rather than spreading far apart
  - The same graphical design structure is repeated
  - Learn once and compare
    → invite comparisons
  - Reveal, all at once, a scope of alternatives, a range of options
    →overview



Rules and Regulations for the Government of Employees of the Operating Department of the Hudson & Manhattan Railroad Company, 1923

Design Principles

- Utilize narratives of space and time
  - Tell a story of position and chronology through visual elements



NEW YORK TO NEW HAVEN

MONDAY TO FRIDAY, EXCEPT HOLIDAYS

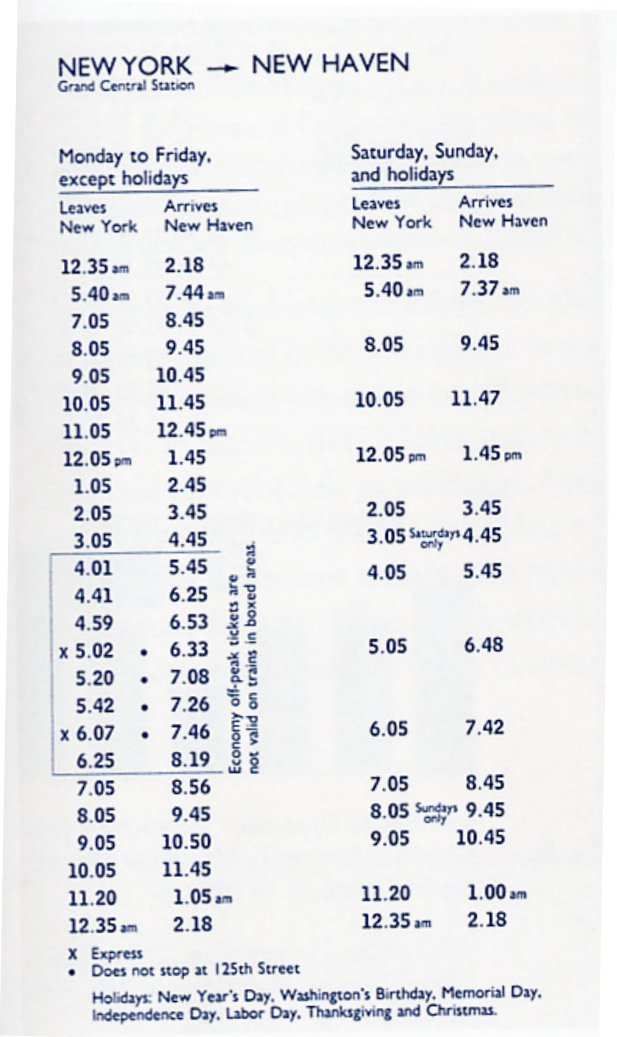
Leave New York	Arrive New Haven	Leave New York	Arrive New Haven	Leave New York	Arrive New Haven
AM	AM	PM	PM	PM	PM
12:35	2:18	2:05	3:45	T 6:25	8:19
5:40	7:44	3:05	4:45	T 7:05	8:56
7:05	8:45	T 4:01	5:45	T 8:05	9:45
8:05	9:45	4:41	6:25	T 9:05	10:50
9:05	10:45	T 4:59	6:53	10:05	11:45
10:05	11:45	XT 5:02E	6:33	11:20	1:05
11:05	12:45	XT 5:20	7:08	12:35	2:18
12:05	1:45	X 5:42	7:26	.....	.....
1:05	2:45	XT 6:07E	7:46	.....	.....
PM	PM	PM	PM	PM	PM

SATURDAY, SUNDAY & HOLIDAYS

AM	AM	PM	PM	PM	PM
12:35	2:18	2:05	3:45	7:05	8:45
5:40	7:37	S 3:05	S 4:45	H 8:05	H 9:45
8:05	9:45	4:05	5:45	9:05	10:45
10:05	11:47	5:05	6:48	11:20	1:00
12:05	1:45	6:05	7:48	12:35	2:18
PM	PM	PM	PM	AM	AM

The service shown herein is operated by Metro-North Commuter R.R.

REFERENCE NOTES
Economy **off-peak** tickets are **not** valid on trains in shaded areas.
Check displays in G.C.T. for departure tracks.
E-Express
X-Does not stop at 125th Street.
S-Saturdays and Washington's Birthday only.
H-Sundays and Holidays only.
T-Snack and Beverage Service.
HOLIDAYS-New Year's Day, Washington's Birthday, Memorial Day, Independence Day, Labor Day, Thanksgiving and Christmas.



NEW YORK → NEW HAVEN
Grand Central Station

Monday to Friday, except holidays: Leaves New York	Arrives New Haven	Saturday, Sunday, and holidays: Leaves New York	Arrives New Haven
12.35 am	2.18	12.35 am	2.18
5.40 am	7.44 am	5.40 am	7.37 am
7.05	8.45		
8.05	9.45	8.05	9.45
9.05	10.45		
10.05	11.45	10.05	11.47
11.05	12.45 pm		
12.05 pm	1.45	12.05 pm	1.45 pm
1.05	2.45		
2.05	3.45	2.05	3.45
3.05	4.45	3.05 Saturdays only	4.45
4.01	5.45	4.05	5.45
4.41	6.25		
4.59	6.53		
x 5.02 •	6.33	5.05	6.48
5.20 •	7.08		
5.42 •	7.26		
x 6.07 •	7.46	6.05	7.42
6.25	8.19		
7.05	8.56	7.05	8.45
8.05	9.45	8.05 Sundays only	9.45
9.05	10.50	9.05	10.45
10.05	11.45		
11.20	1.05 am	11.20	1.00 am
12.35 am	2.18	12.35 am	2.18

Economy off-peak tickets are not valid on trains in boxed areas.

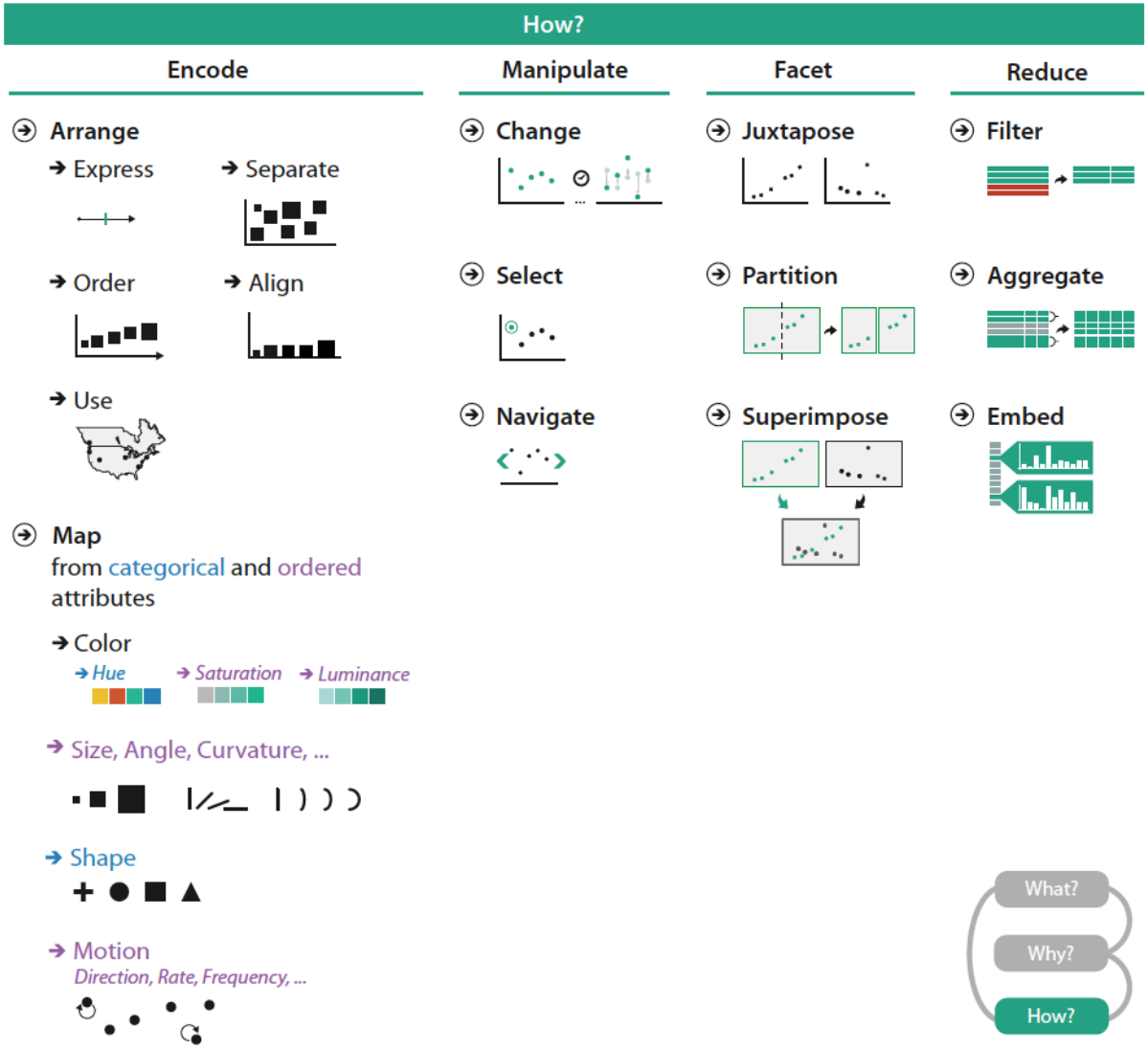
x Express
• Does not stop at 125th Street

Holidays: New Year's Day, Washington's Birthday, Memorial Day, Independence Day, Labor Day, Thanksgiving and Christmas.

Power of Negative Space

Power of Negative Space

How to Design Vis Idioms

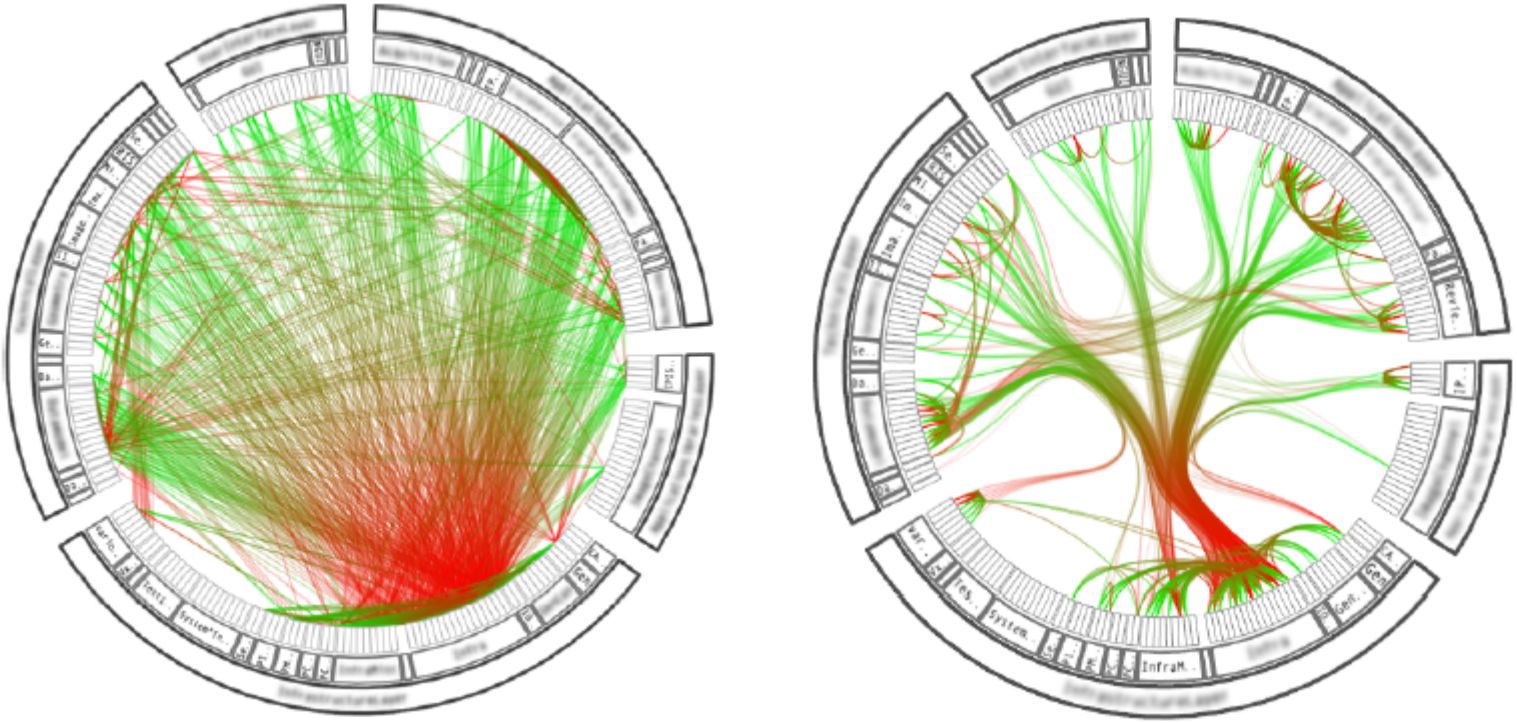


T. Munzner. Visualization Analysis and Design. AK Peters, 2014

Examples and Tools

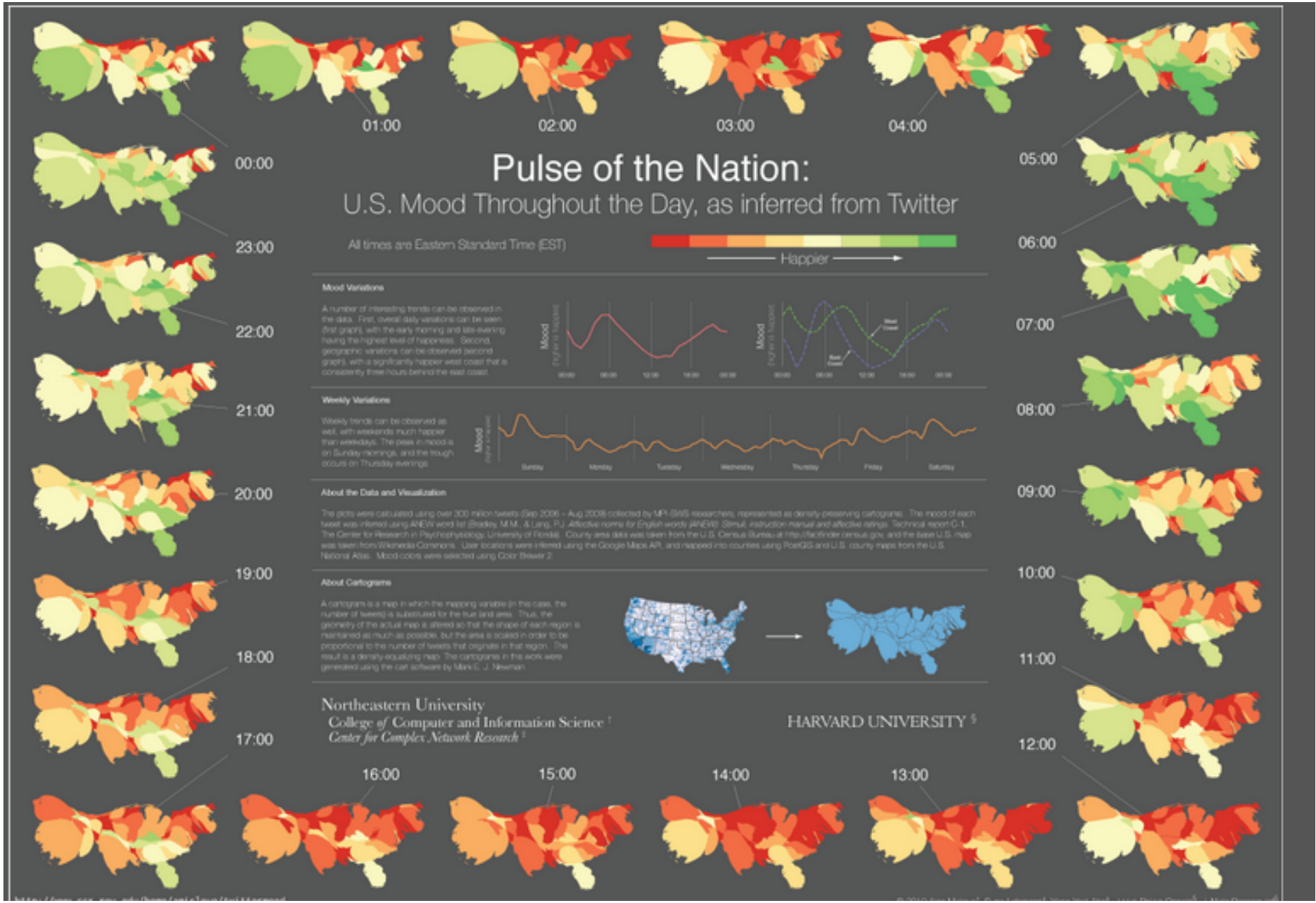
- Visualization Examples
- Visualization Tools

Graph Visualization - Edge Bundle



http://www.win.tue.nl/~dholten/papers/bundles_infovis.pdf

Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data, Danny Holten, InfoVis, 2006

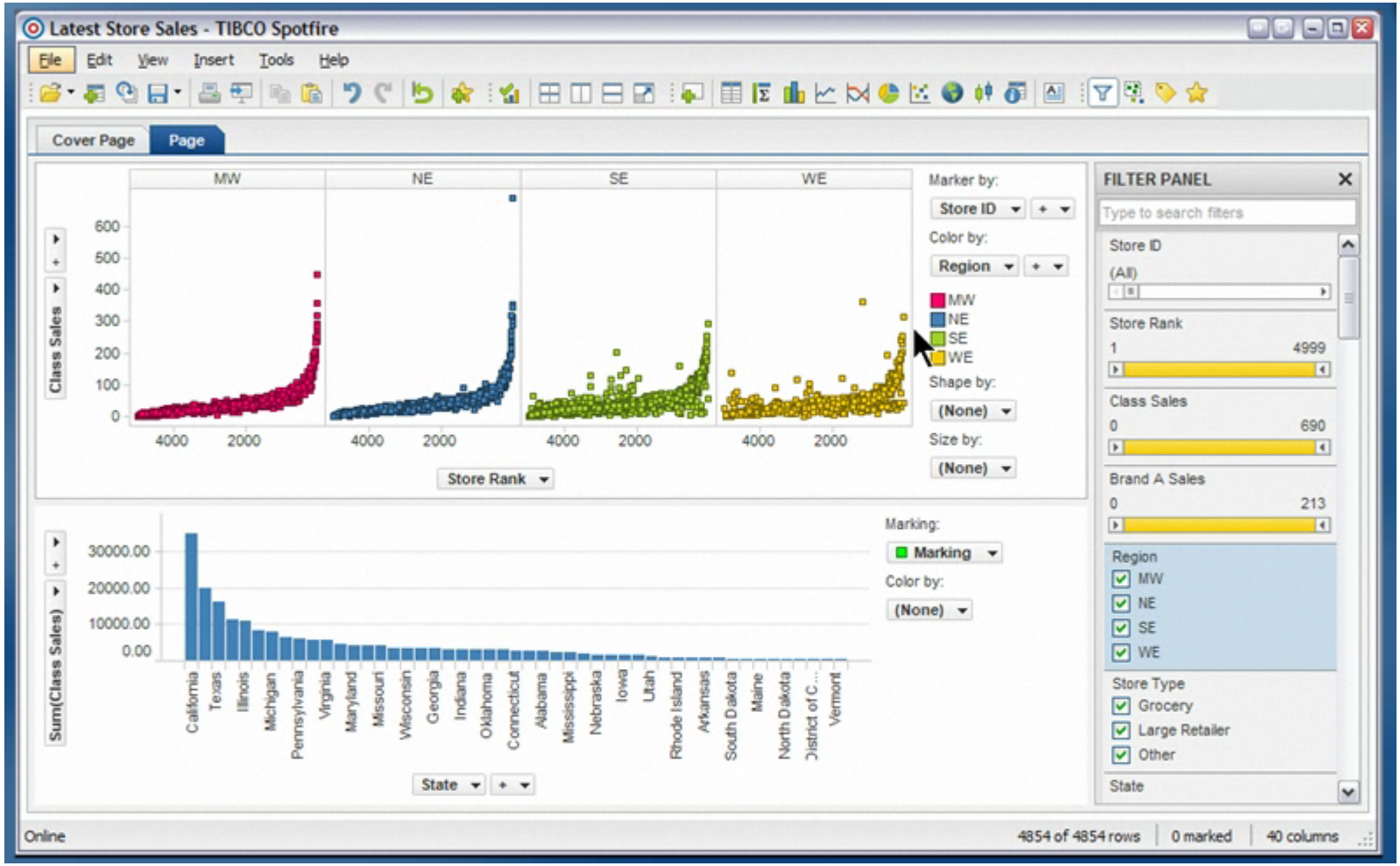


http://www.ccs.neu.edu/home/amislove/twittermood/

Treemap for Hierarchical Data

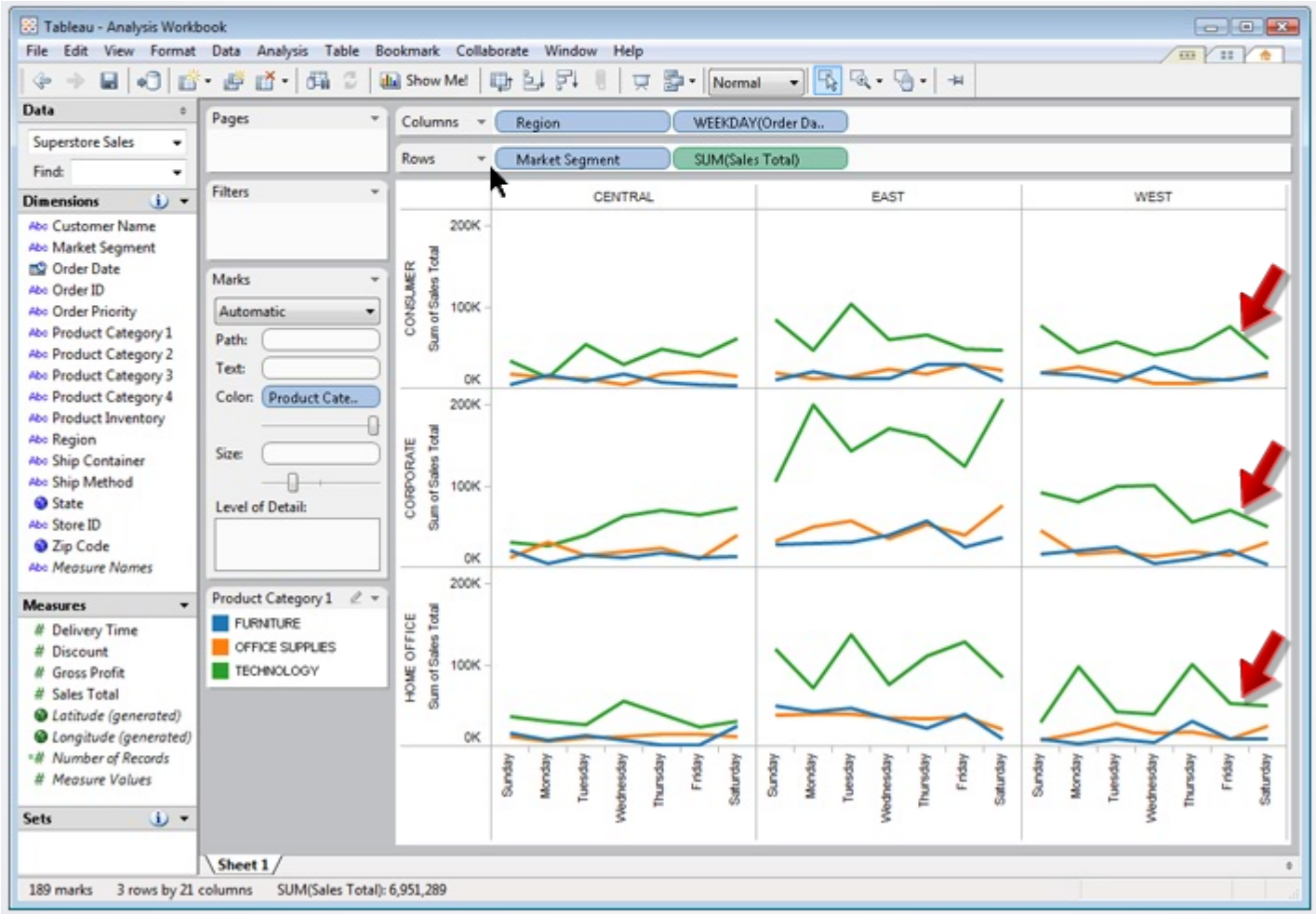
https://finviz.com/map.ashx

Spotfire by TIBCO



http://spotfire.tibco.com/tour/

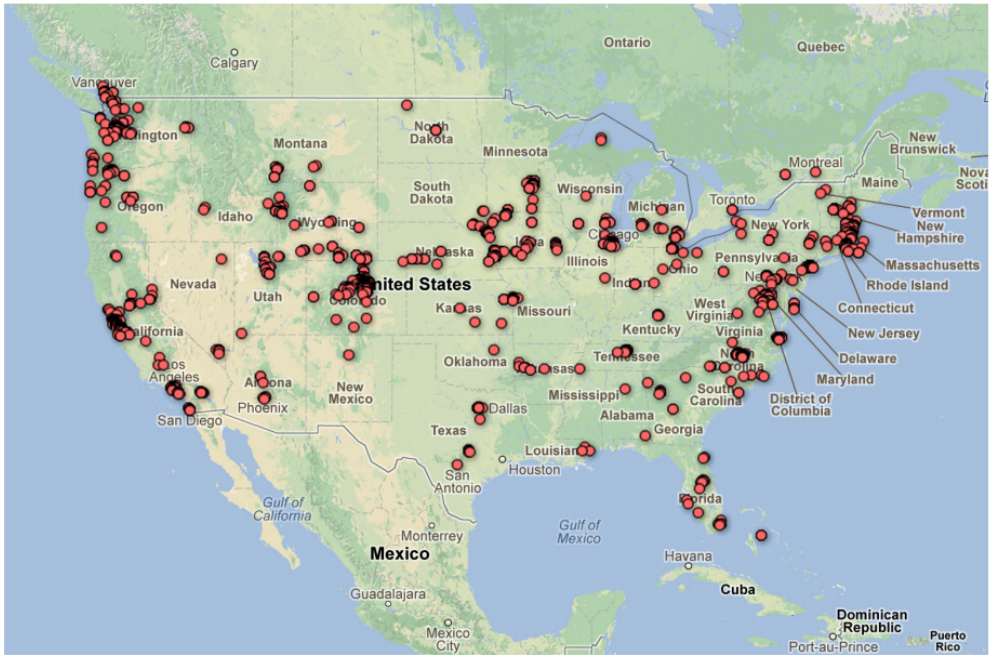
Tableau

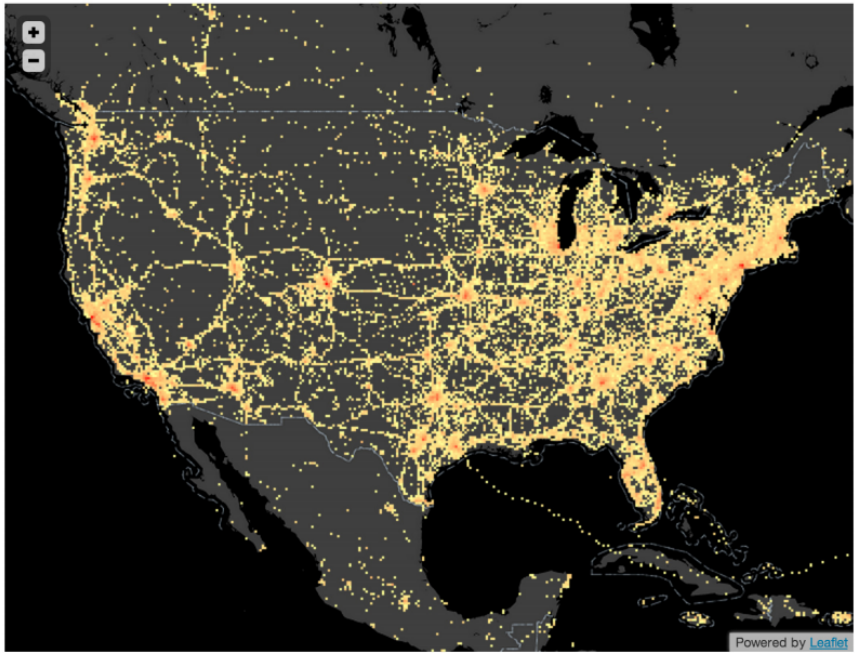


http://www.tableausoftware.com/products/tour

imMems

- Real-time Visual Querying of Big Data





imMens: Real-time Visual Querying of Big Data
Zhicheng Liu, Biye Jiang, Jeffrey Heer
Computer Graphics Forum (Proc. EuroVis), 32(3), 2013

imMems

- Decomposing a full cube into sub-cubes and data tiles

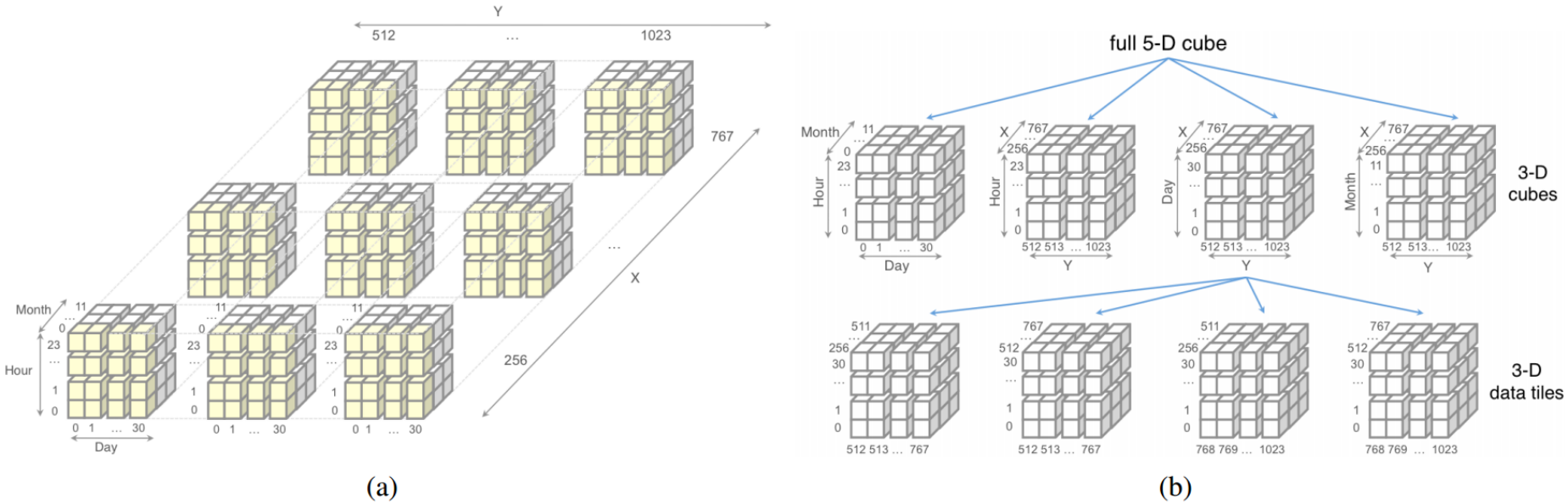
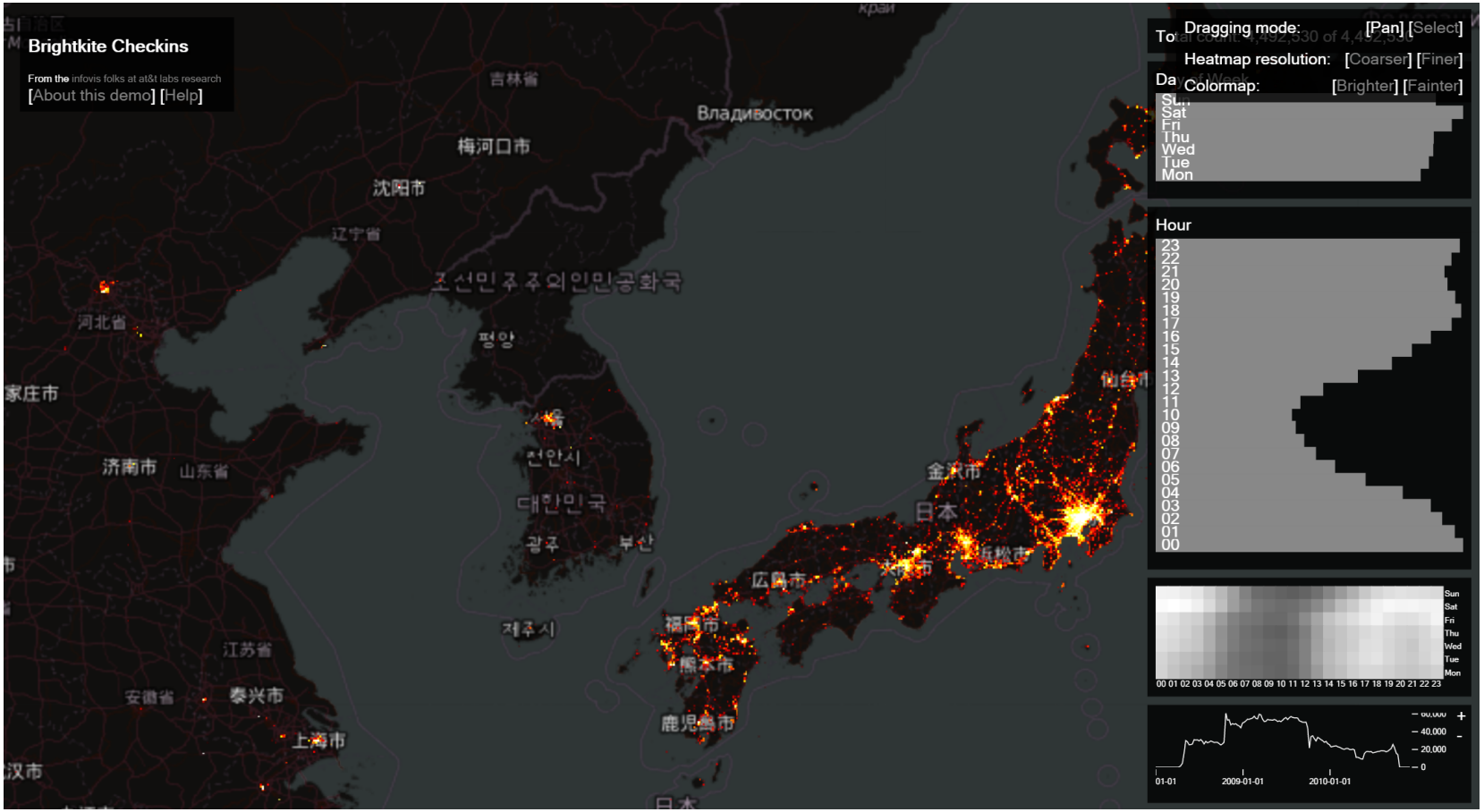


Figure 5: *(a) A 5-dimensional data cube of Brightkite check-ins; (b) Decomposing a full cube into sub-cubes and data tiles.*

imMens: Real-time Visual Querying of Big Data
Zhicheng Liu, Biye Jiang, Jeffrey Heer
Computer Graphics Forum (Proc. EuroVis), 32(3), 2013

Nanocubes

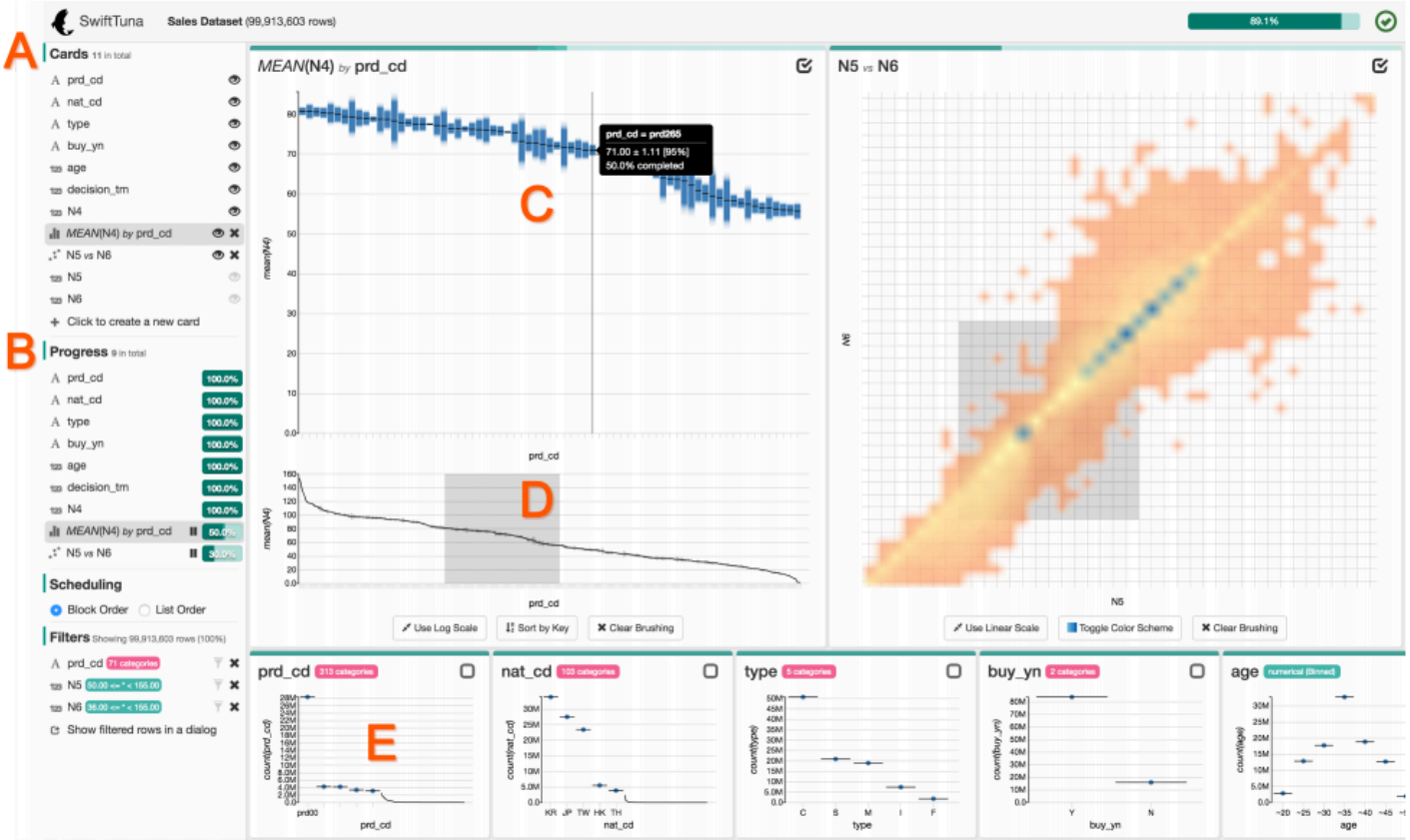
- Fast visualization of large spatiotemporal datasets
- hierarchical data cubes, WebGL → visualize millions of records



Lauro Lins, James T. Klosowski, and Carlos Scheidegger. Nanocubes for Real-Time Exploration of Spatiotemporal Datasets. Visualization and Computer Graphics, IEEE Transactions on 19, no. 12 (2013): 2456-2465.

SwiftTuna

- Incrementally Exploring Large-scale Multidimensional Data
- Without building precomputed data structures



Questions?