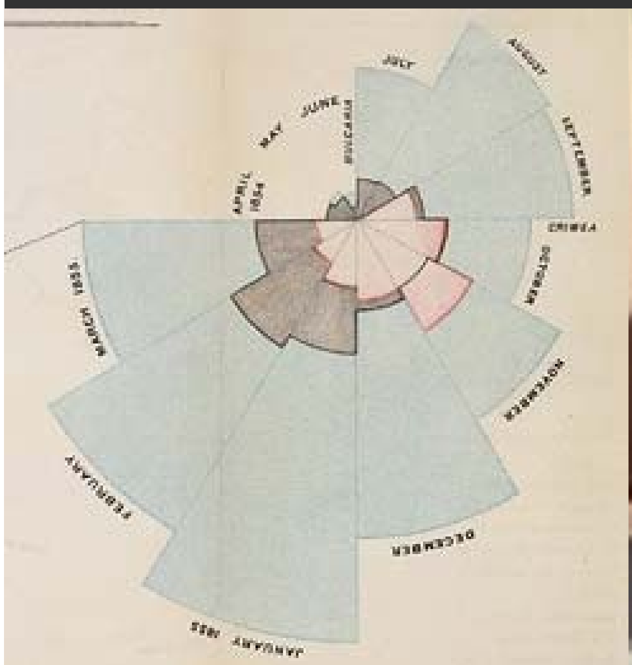
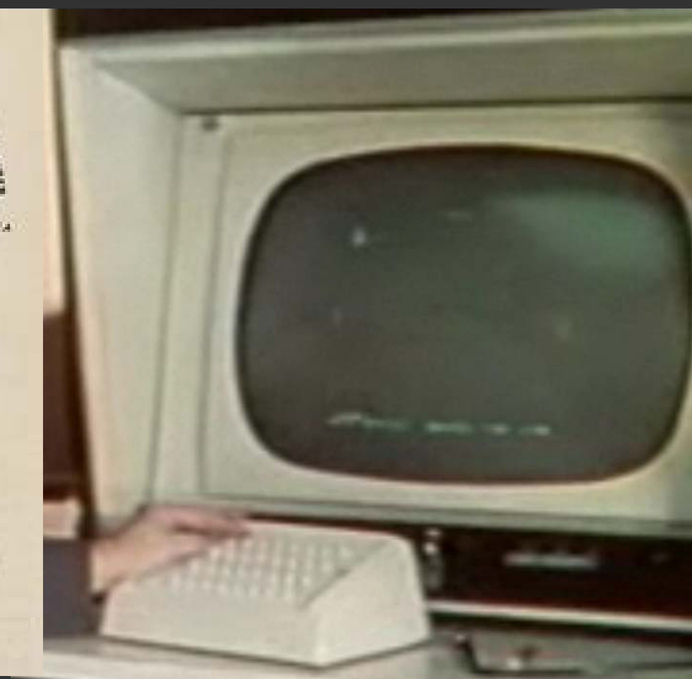
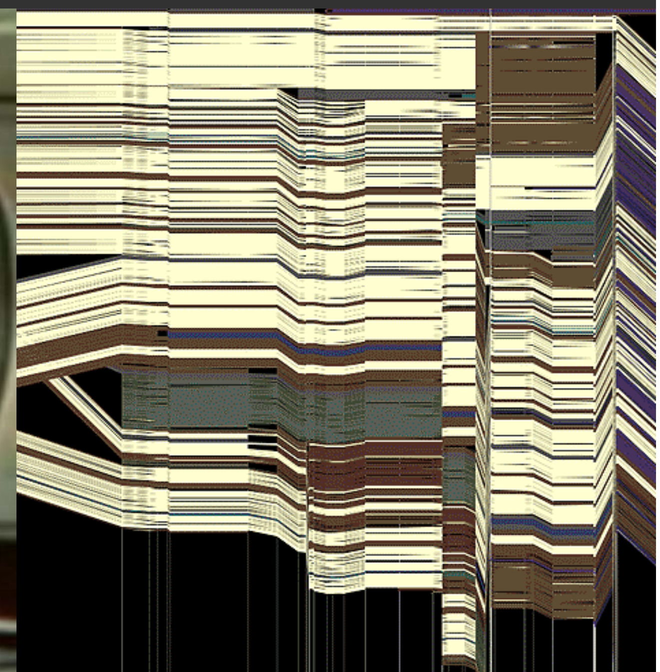
a brief introduction to

Data Visualization



Jeffrey Heer Stanford University

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

$u_X = 9.0$ $\sigma_X = 3.317$

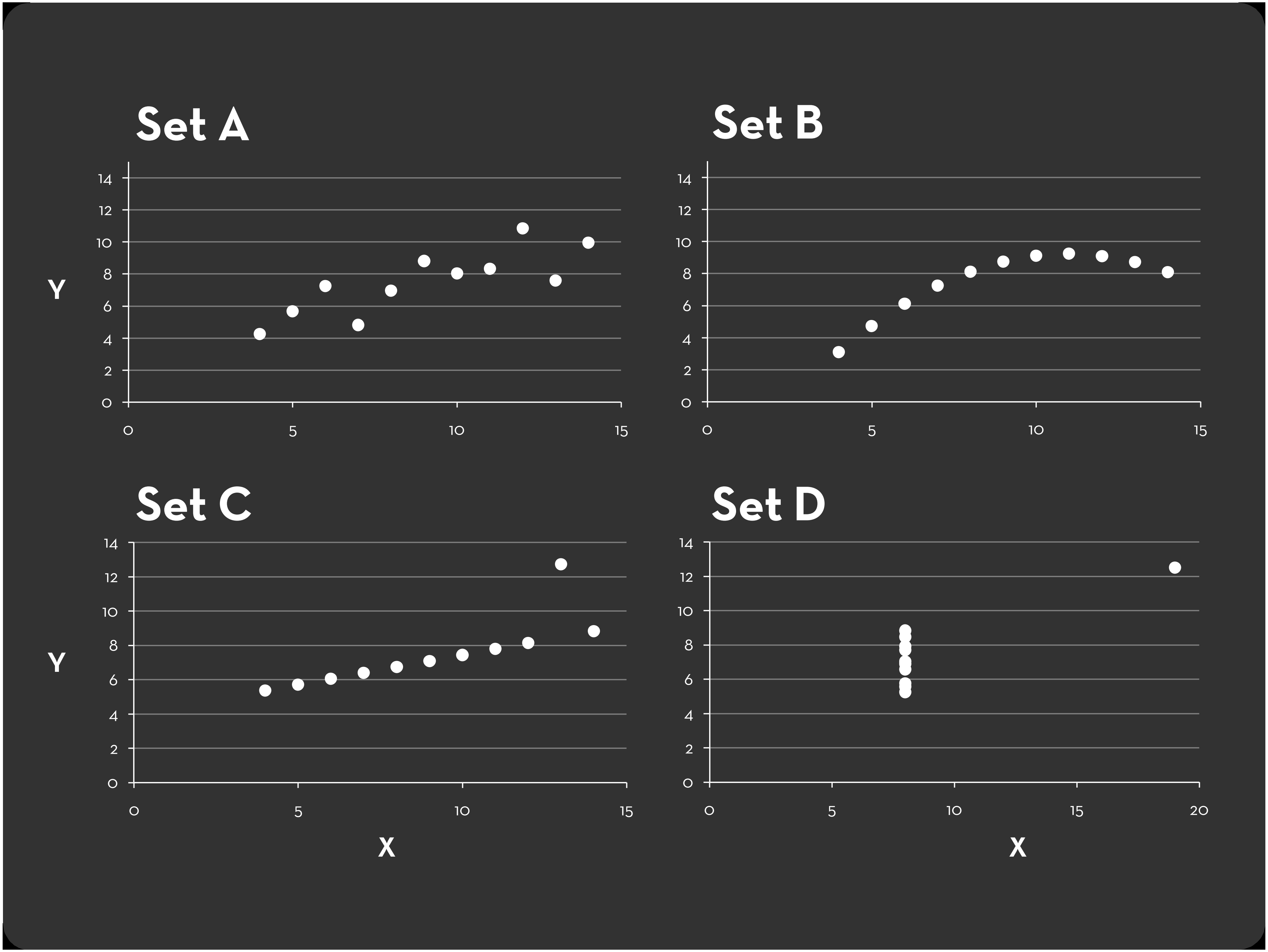
$u_Y = 7.5$ $\sigma_Y = 2.03$

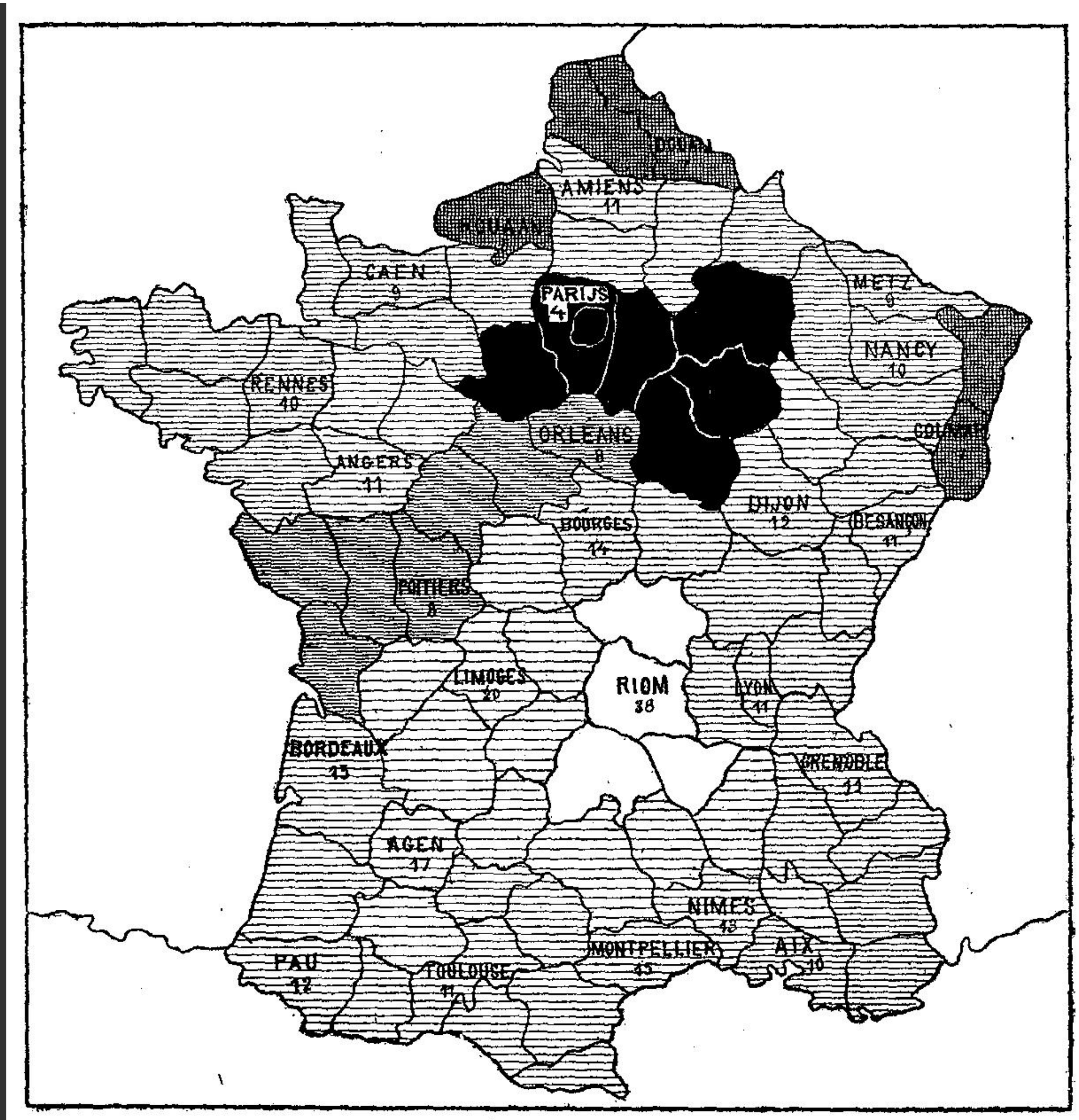
Linear Regression

$Y = 3 + 0.5\,X$

$R^2 = 0.67$

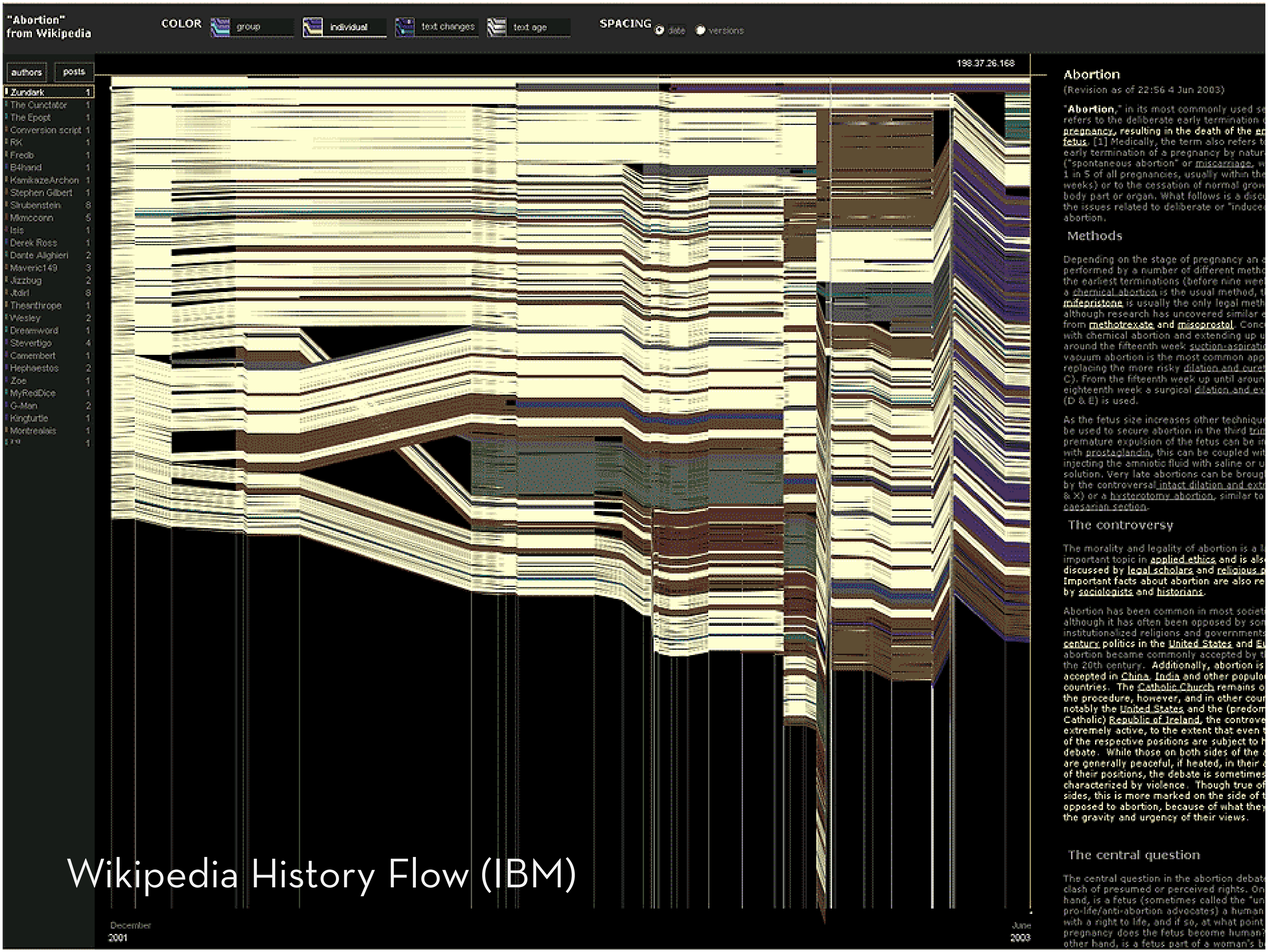
[Anscombe 73]


Set A
Set B
Set C
Set D
Y
Y
X
X
0
2
4
6
8
10
12
14
0
5
10
15
20

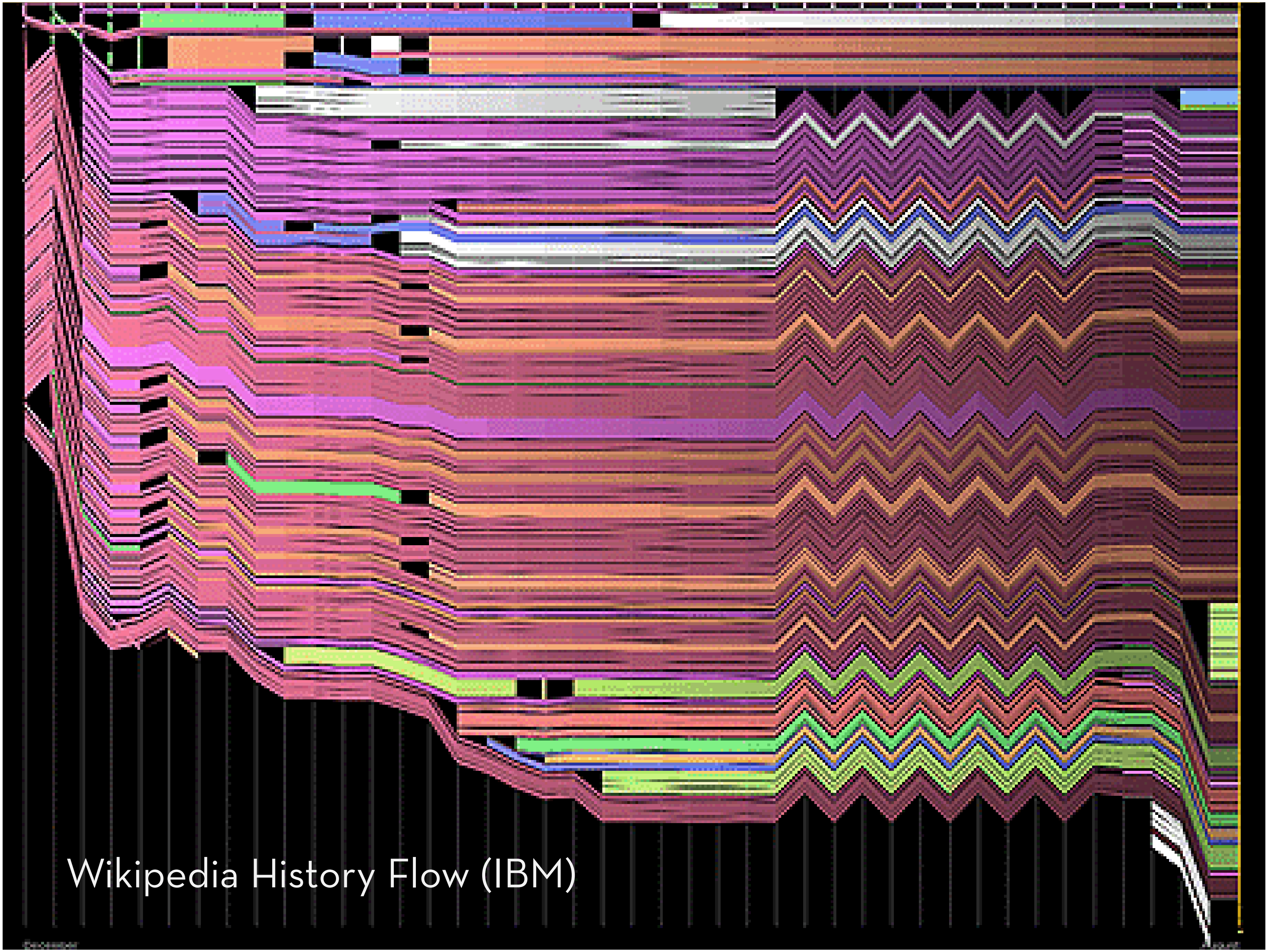


1826(?) Illiteracy in France, Pierre Charles Dupin


cabspotting.org



Wikipedia History Flow (IBM)



Wikipedia History Flow (IBM)

Questions of the Day

1. *Why* create visualizations?
2. *What* makes a visualization effective?
3. *How* does one design effective visualizations?
4. *What* visual analysis tools are available?
5. *Where* to look for further resources?

Why do we create visualizations?

Why do we create visualizations?

Why do we create visualizations?

- Answer questions (or discover them)
- Make decisions
- See data in context
- Expand memory
- Support graphical calculation
- Find patterns
- Present argument or tell a story
- Inspire

Attention

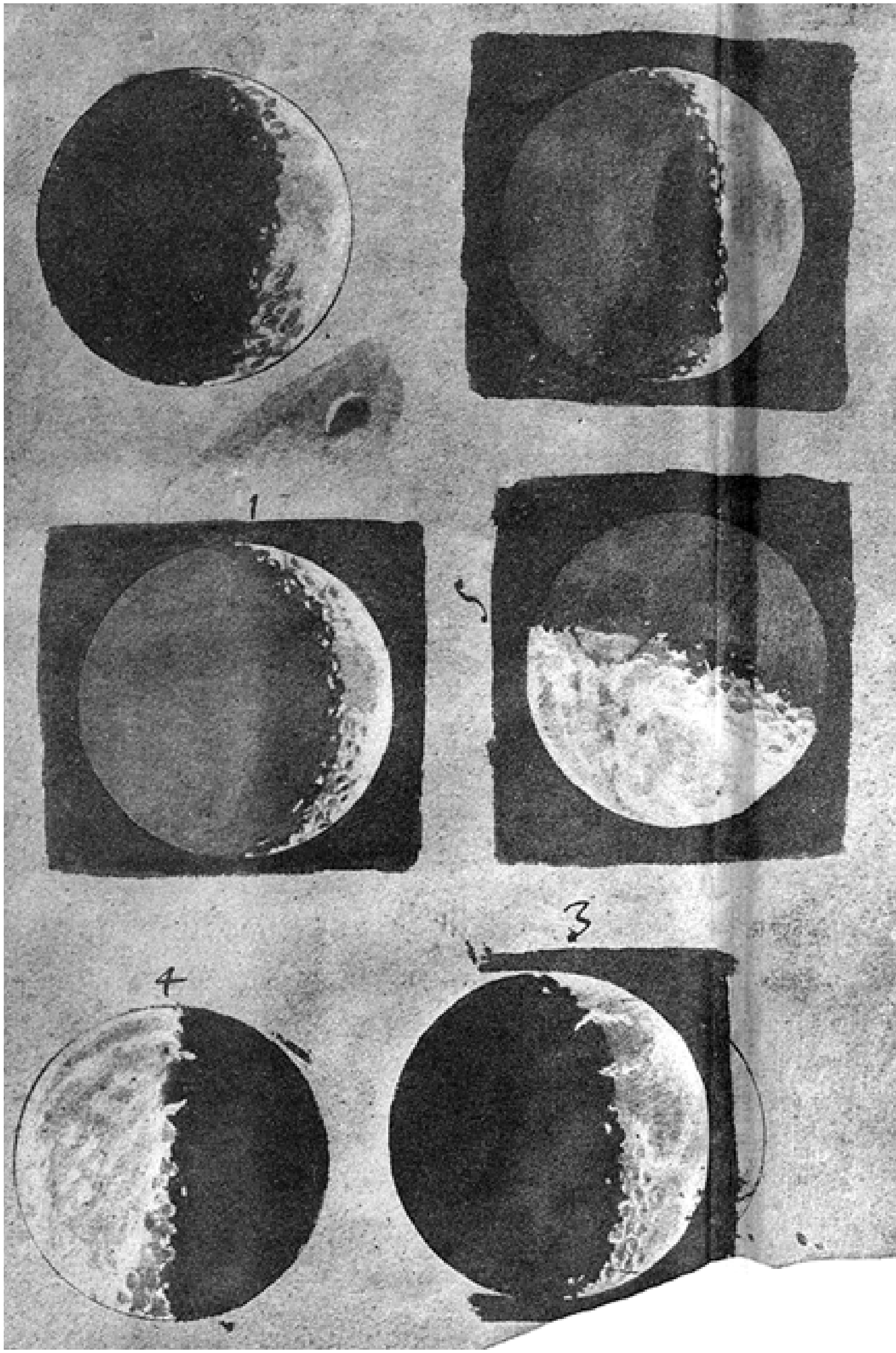
"What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."

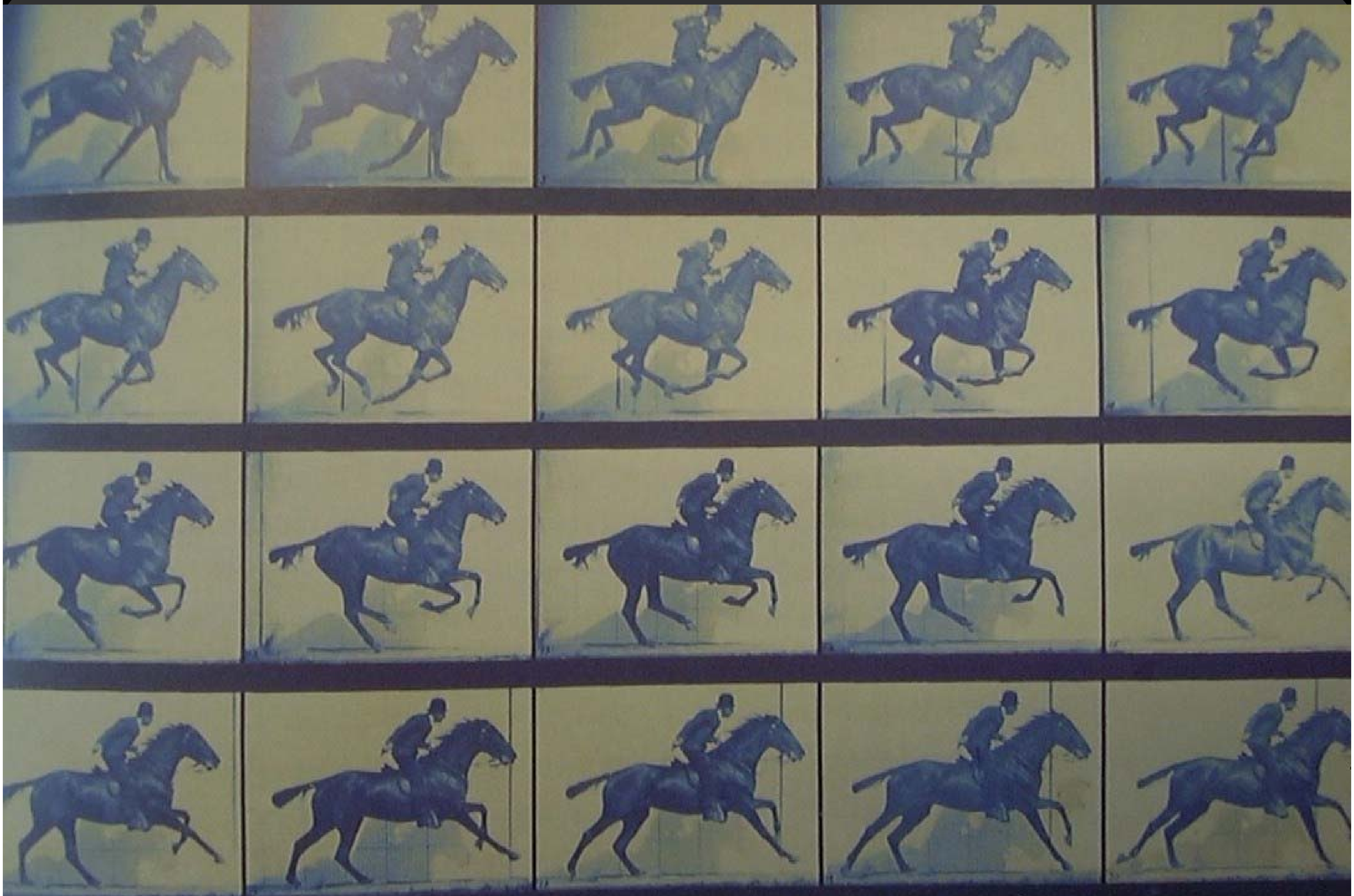
~Herb Simon
as quoted by Hal Varian
Scientific American
September 1995

Three functions of visualizations

Record: store information

- Photographs, blueprints, …


1
3
4



Gallop, Bay Horse "Daisy" [Muybridge 1884-86]

Three functions of visualizations

Record: store information

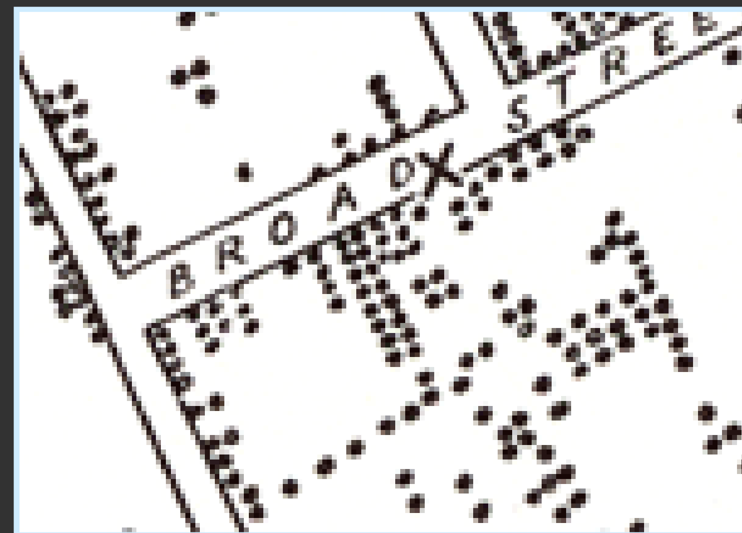
- Photographs, blueprints, …

Analyze: support reasoning about information

- Process and calculate
- Reason about data
- Develop models and hypotheses

In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]

Cholera outbreak



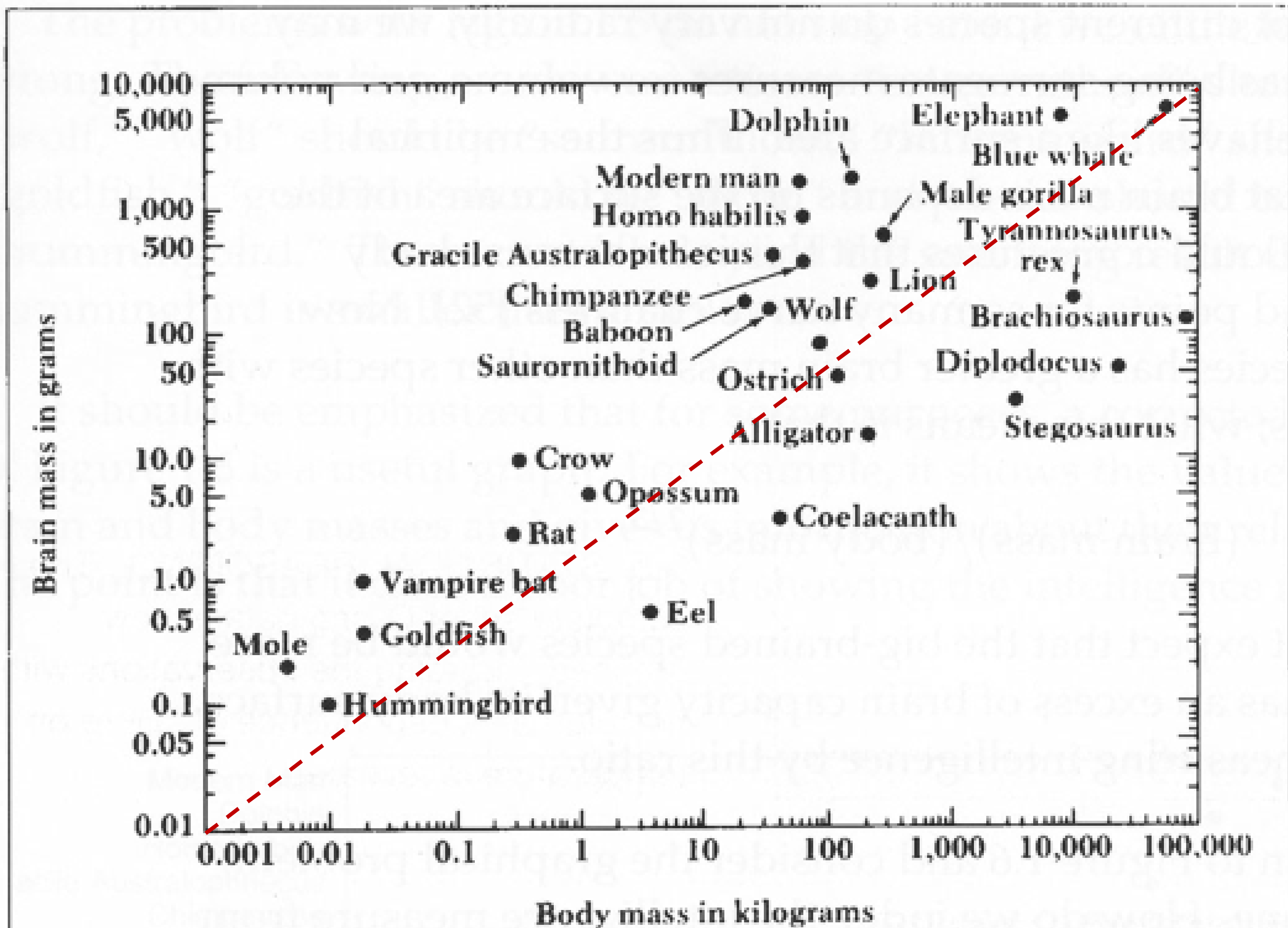
Used map to hypothesize that pump on Broad St. was the cause. [from Tufte 83]

The most powerful brain?

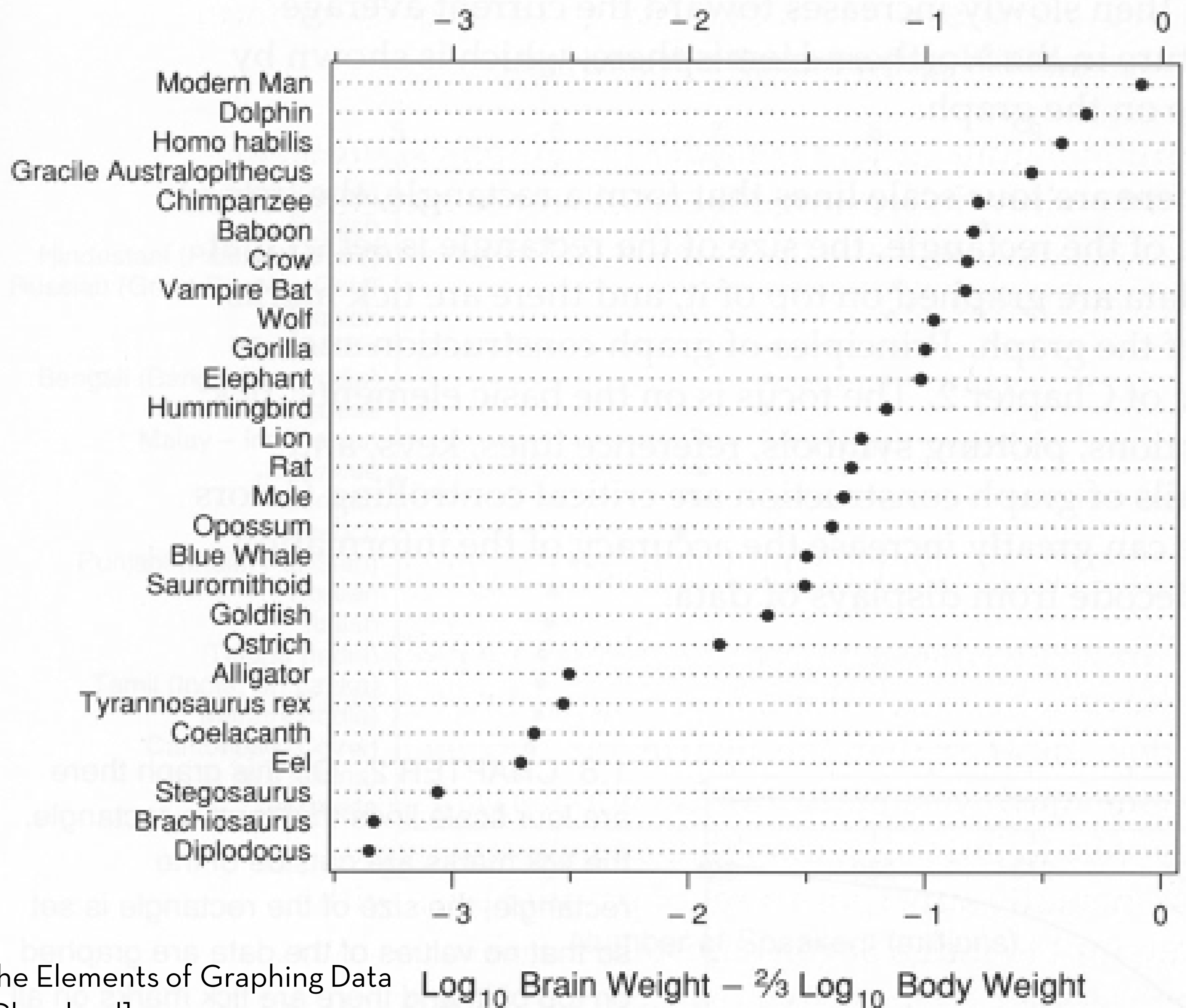
Microsoft Excel - animal.xls

ID	Name	Body Weight	Brain Weight
1	Lesser Short-tailed Shrew	5	0.14
2	Little Brown Bat	10	0.25
3	Mouse	23	0.3
4	Big Brown Bat	23	0.4
5	Musk Shrew	48	0.33
6	Star Nosed Mole	60	1
7	Eastern American Mole	75	1.2
8	Ground Squirrel	101	4
9	Tree Shrew	104	2.5
10	Golden Hamster	120	1
11	Mole Rate	122	3
12	Galago	200	5
13	Rat	280	1.9
14	Chinchilla	425	6.4
15	Desert Hedgehog	550	2.4
16	Rock Hyrax (a)	750	12.3
17	European Hedgehog	785	3.5
18	Tenrec	900	2.6
19	Arctic Ground Squirrel	920	5.7
20	African Giant Pouched Rat	1000	6.6
21	Guinea Pig	1040	5.5
22	Mountain Beaver	1350	8.1
23	Slow Loris	1400	12.5
24	Genet	1410	17.5
25	Phalanger	1620	11.4

animal



The Dragons of Eden [Carl Sagan]



The Elements of Graphing Data
[Cleveland]

Three functions of visualizations

Record: store information

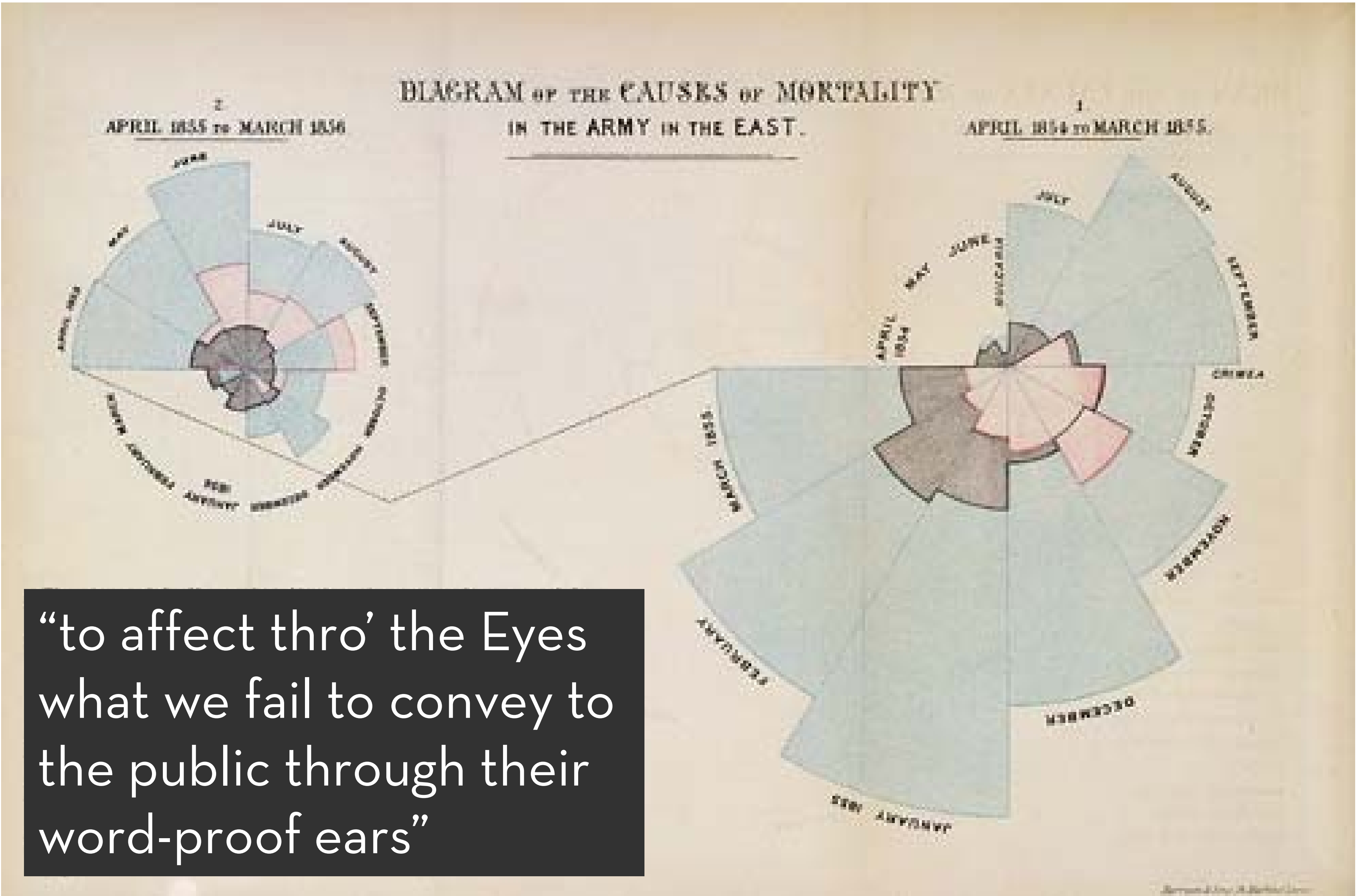
- Photographs, blueprints, …

Analyze: support reasoning about information

- Process and calculate
- Reason about data
- Feedback and interaction

Communicate: convey information to others

- Share and persuade
- Collaborate and revise
- Emphasize important aspects of data



"to affect thro' the Eyes what we fail to convey to the public through their word-proof ears"

1856 "Coxcomb" of Crimean War Deaths, Florence Nightingale

How do we design effective visualizations?

Goals of visualization research

1 **Understand** how visualizations convey information

What do people perceive/comprehend?

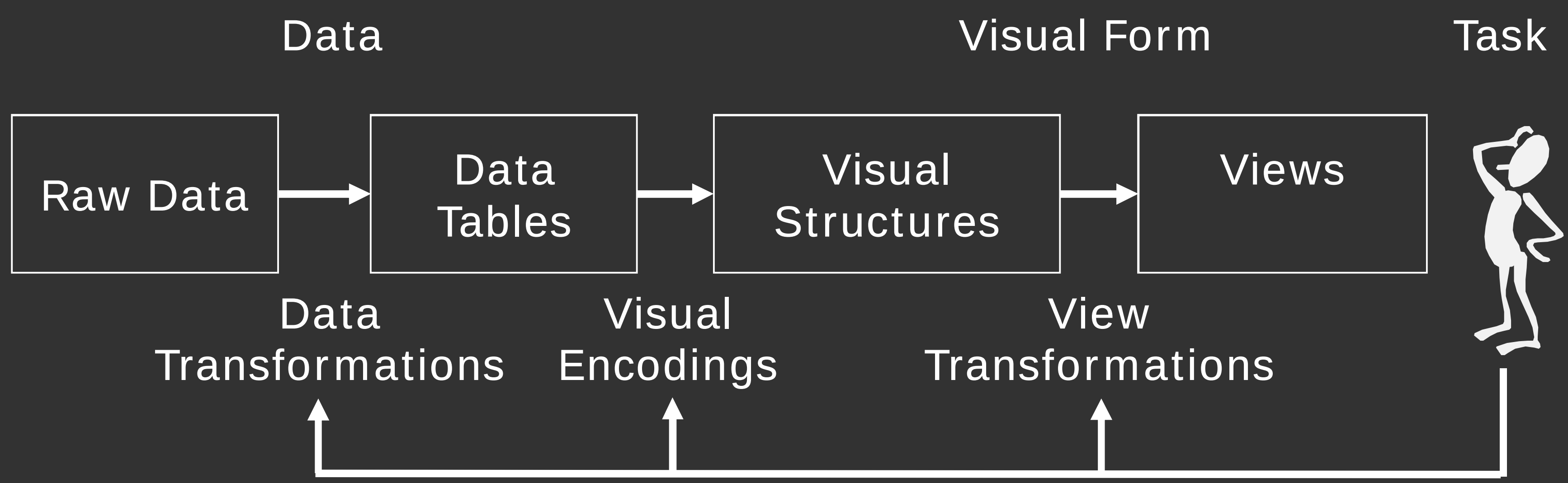
How do visualizations correspond with mental models?

2 **Develop principles and techniques** for creating effective visualizations and supporting analysis

Amplify perception and cognition

Strengthen tie between visualization and mental models

Visualization Reference Model



Taxonomy

- 1D (sets and sequences)
- Temporal
- 2D (maps)
- 3D (shapes)
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)
- Are there others?

The eyes have it: A task by data type taxonomy for information visualization [Shneiderman 96]

Nominal, Ordinal and Quantitative

N - Nominal (labels)

- Fruits: Apples, oranges, ...

O - Ordinal (rank-ordered)

- Quality of meat: Grade A, AA, AAA

Q - Interval (location of zero arbitrary)

- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Like a geometric point. Cannot compare directly
- Only differences (i.e. intervals) may be compared

Q - Ratio (zero fixed)

- Physical measurement: Length, Mass, Temp, ...
- Counts and amounts
- Like a geometric vector, origin is meaningful

S. S. Stevens, On the theory of scales of measurements, 1946

Example: U.S. Census Data

People: # of people in group

Year: 1850 – 2000 (every decade)

Age: 0 – 90+

Sex: Male, Female

Marital Status: Single, Married, Divorced, …

Example: U.S. Census

People

Year

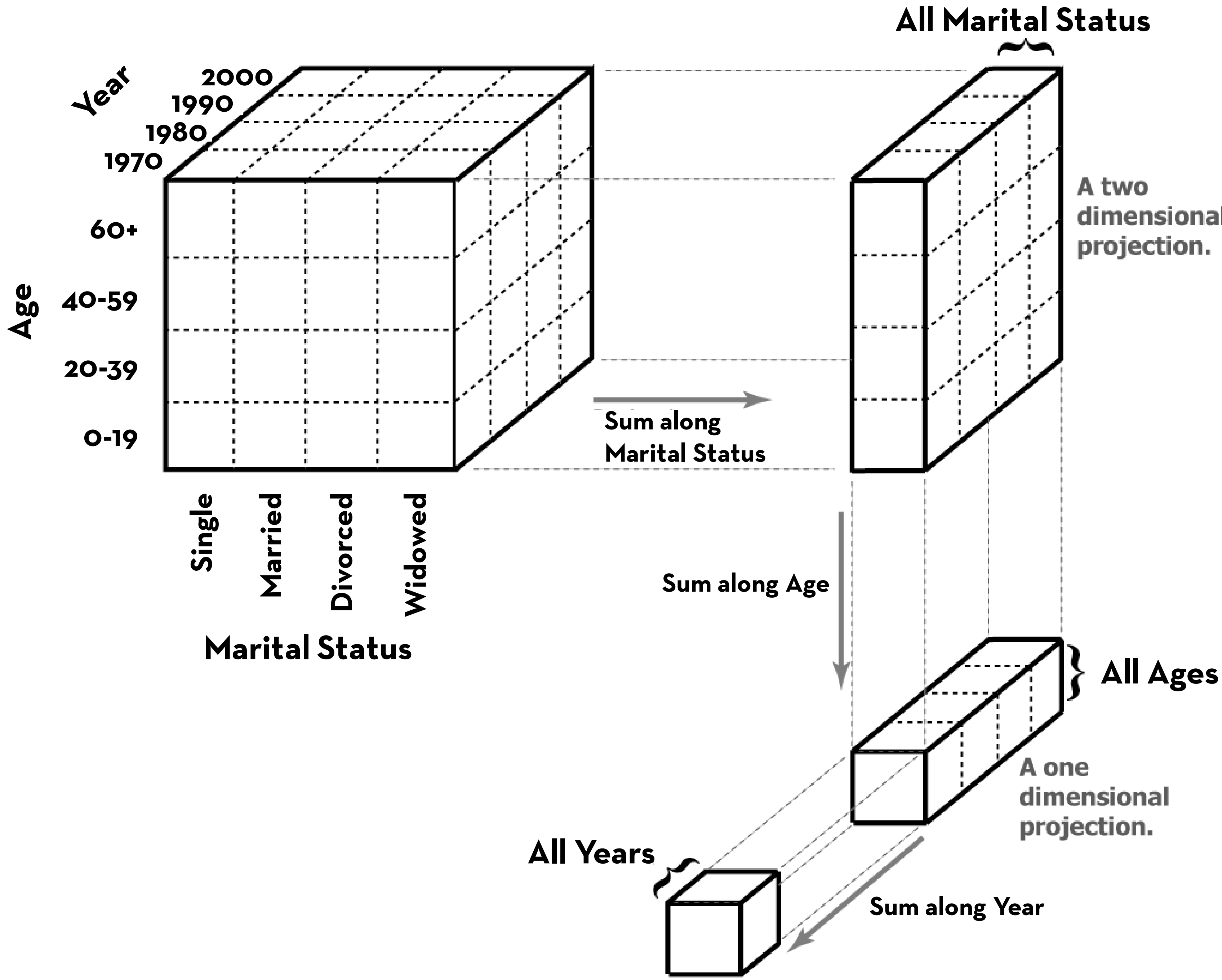
Age

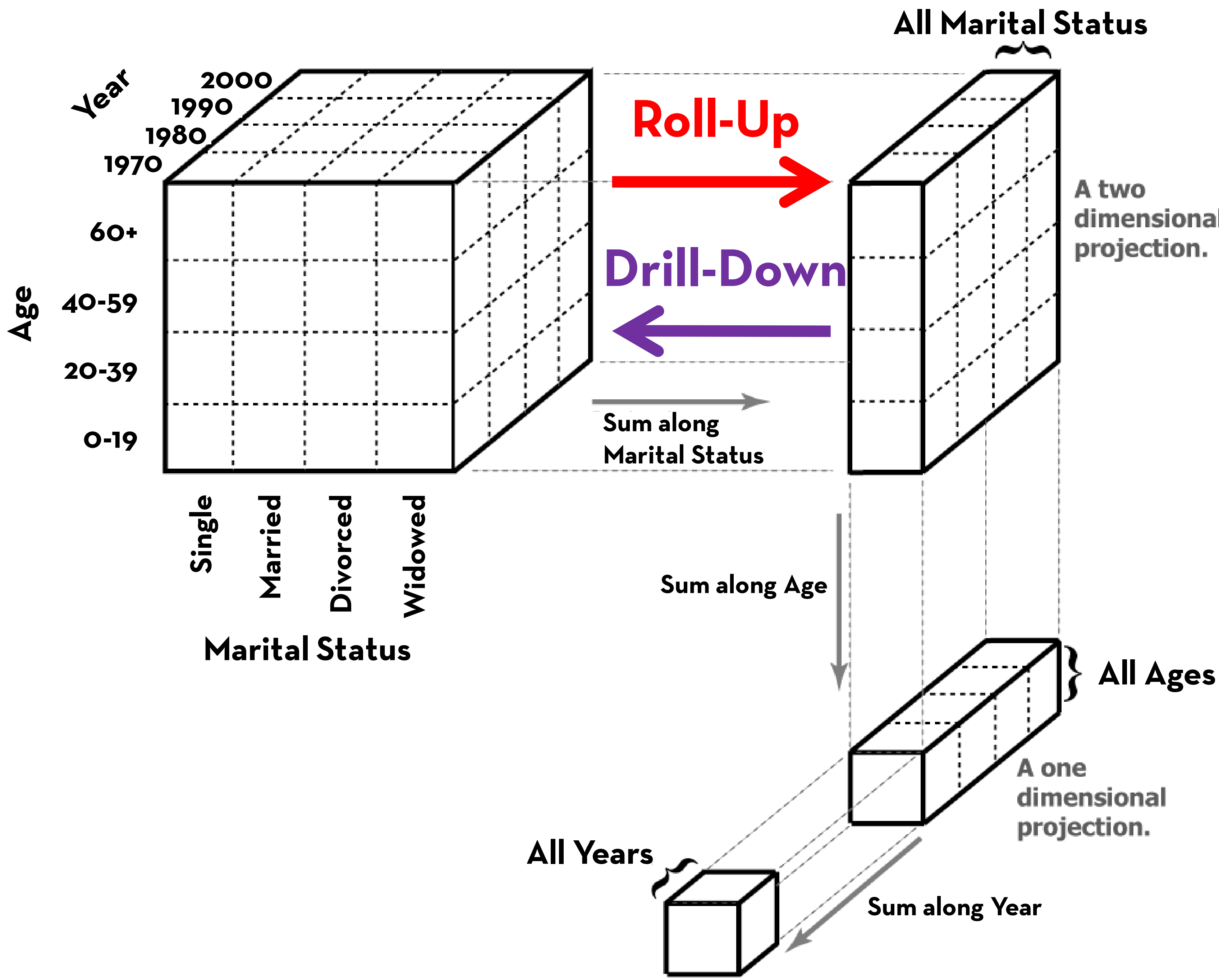
Sex

Marital Status

2348 data points

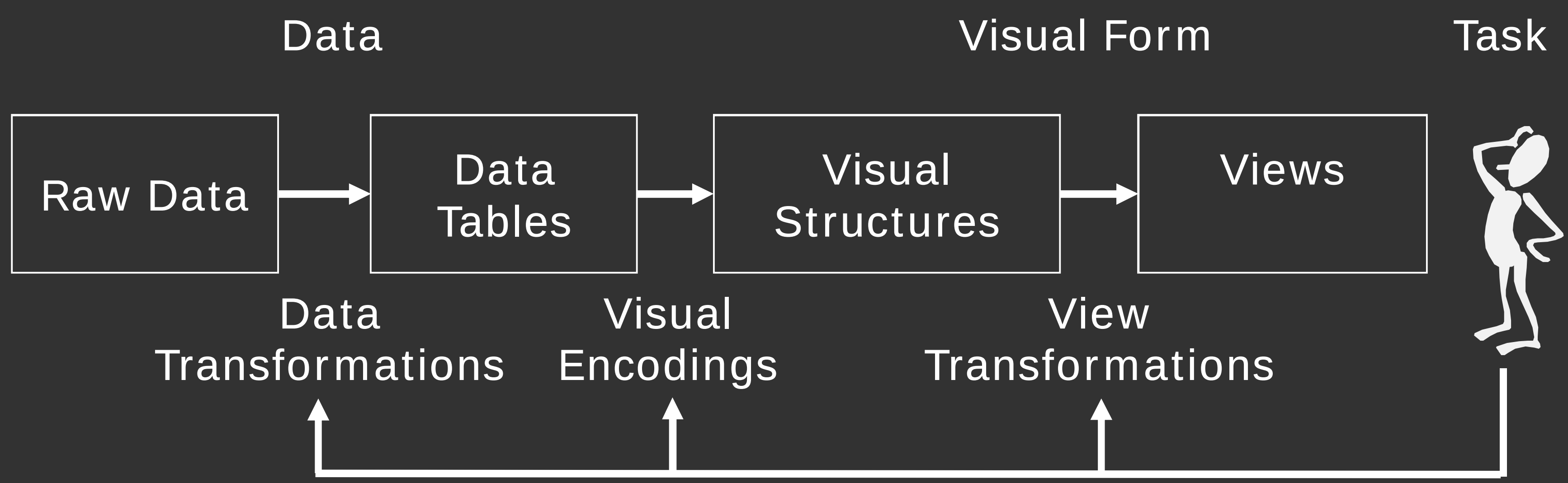
	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343
23	1850	50	0	2	286580
24	1850	55	0	1	194080
25	1850	55	0	2	187208
26	1850	60	0	1	174976
27	1850	60	0	2	162236
28	1850	65	0	1	106827
29	1850	65	0	2	105534
30	1850	70	0	1	73677
31	1850	70	0	2	71762
32	1850	75	0	1	40834
33	1850	75	0	2	40229
34	1850	80	0	1	23449
35	1850	80	0	2	22949
36	1850	85	0	1	8186
37	1850	85	0	2	10511
38	1850	90	0	1	5259
39	1850	90	0	2	6569
40	1860	0	0	1	2120846
41	1860	0	0	2	2092162

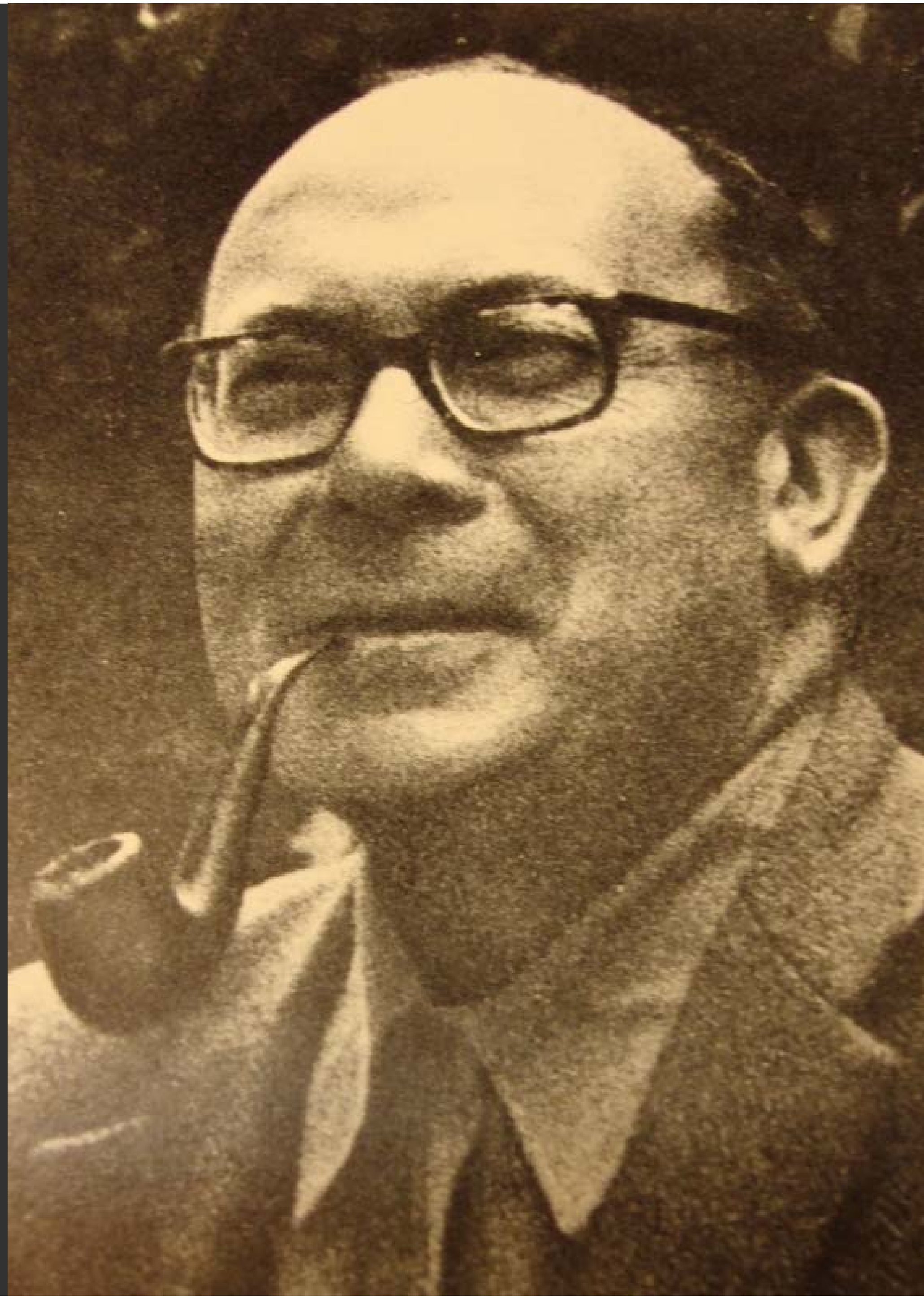

All Marital Status
Year
2000
1990
1980
1970
Age
60+
40-59
20-39
0-19
Single
Married
Divorced
Widowed
Marital Status
A two dimensional projection.
Sum along Marital Status
Sum along Age
All Ages
A one dimensional projection.
All Years
Sum along Year



All Marital Status
Year
2000
1990
1980
1970
Age
60+
40-59
20-39
0-19
Single
Married
Divorced
Widowed
Marital Status
Roll-Up
Drill-Down
Sum along
Marital Status
A two dimensional projection.
Sum along Age
All Ages
A one dimensional projection.
All Years
Sum along Year

Visualization Reference Model





Visual language is a sign system

Jacques Bertin

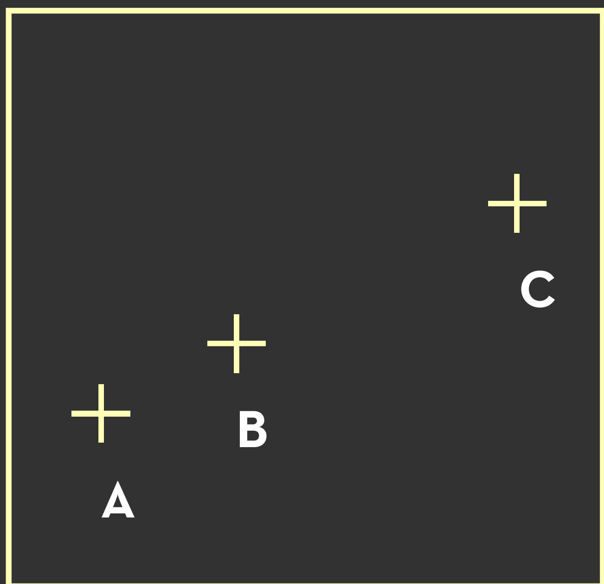
Images perceived as a set of signs

Sender encodes information in signs

Receiver decodes information from signs

Sémiologie Graphique, 1967

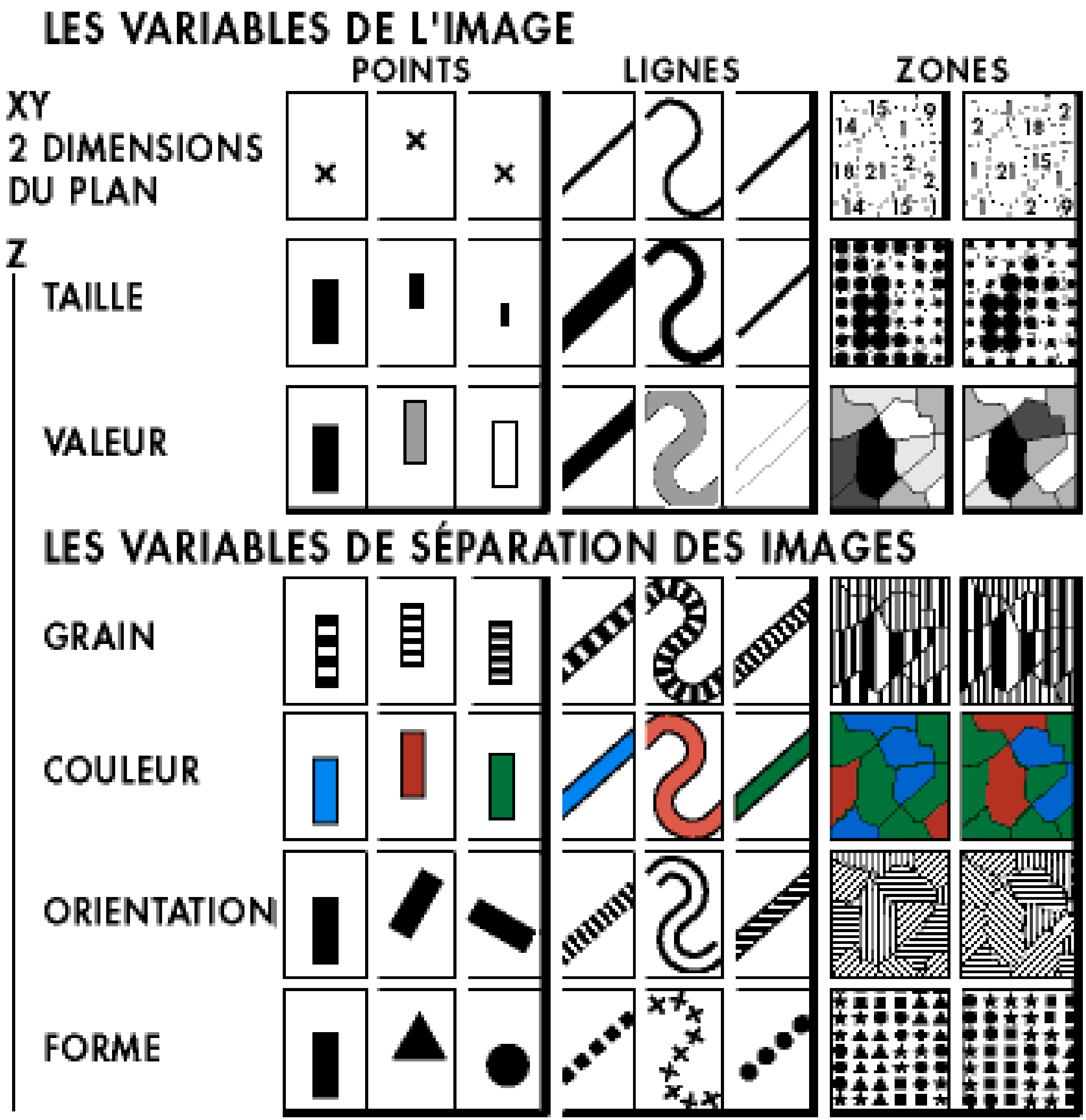
Bertin's Semiology of Graphics



1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.

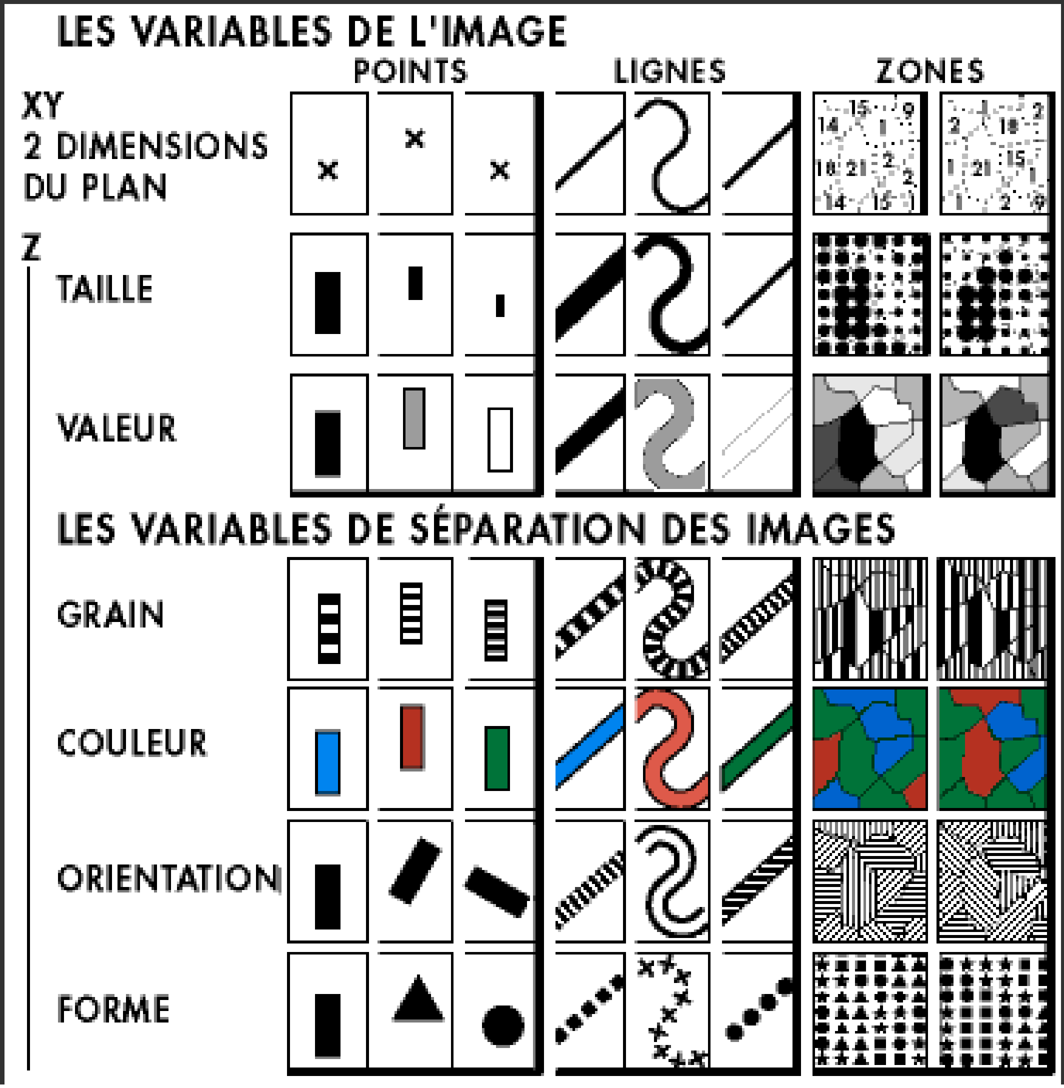
∴ Encode quantitative variables

"Resemblance, order and proportion are the three signifieds in graphics." - Bertin


LES VARIABLES DE L'IMAGE
POINTS
LIGNES
ZONES
XY
2 DIMENSIONS
DU PLAN
Z
TAILLE
VALEUR
LES VARIABLES DE SÉPARATION DES IMAGES
GRAIN
COULEUR
ORIENTATION
FORME

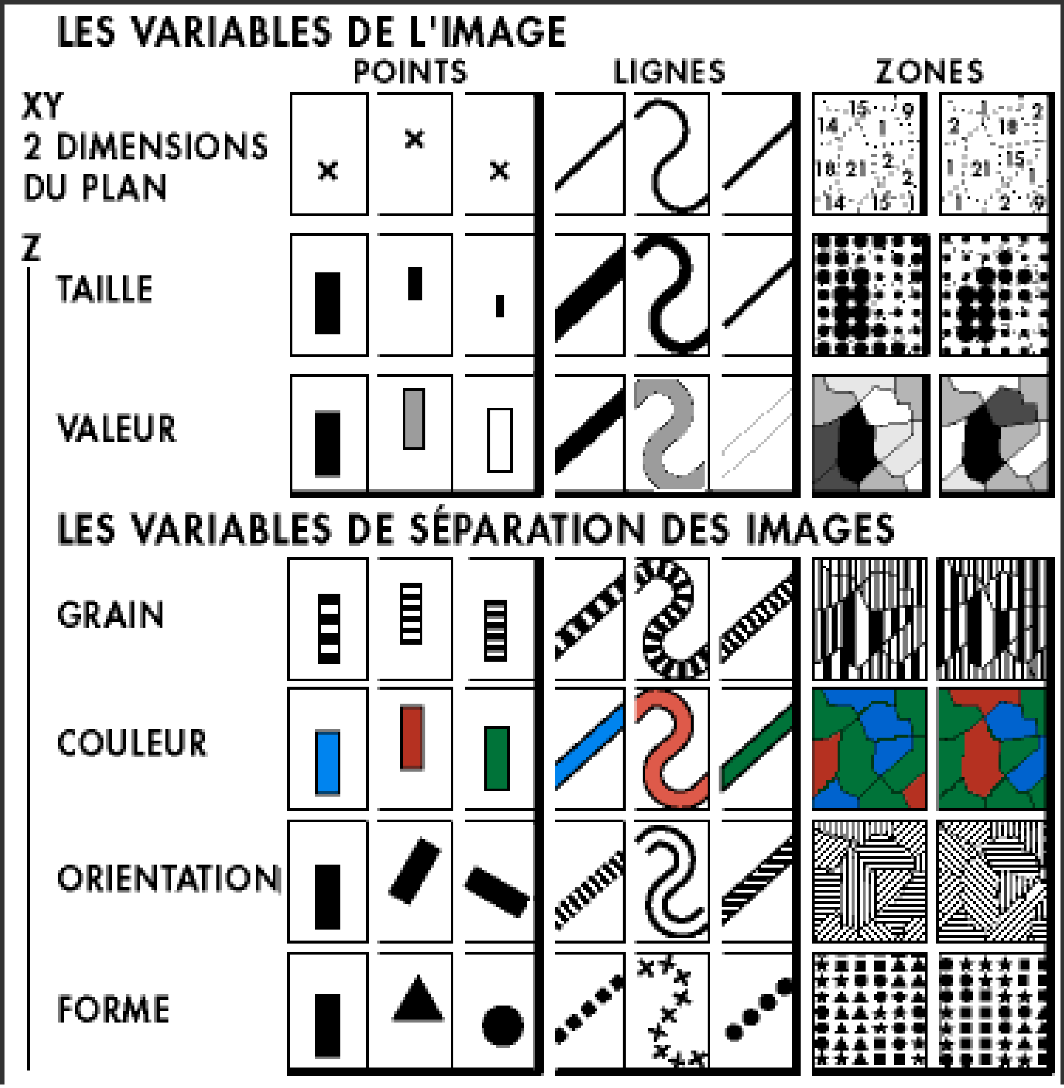
Visual encoding variables

Position (x 2)
Size
Value
Texture
Color
Orientation
Shape

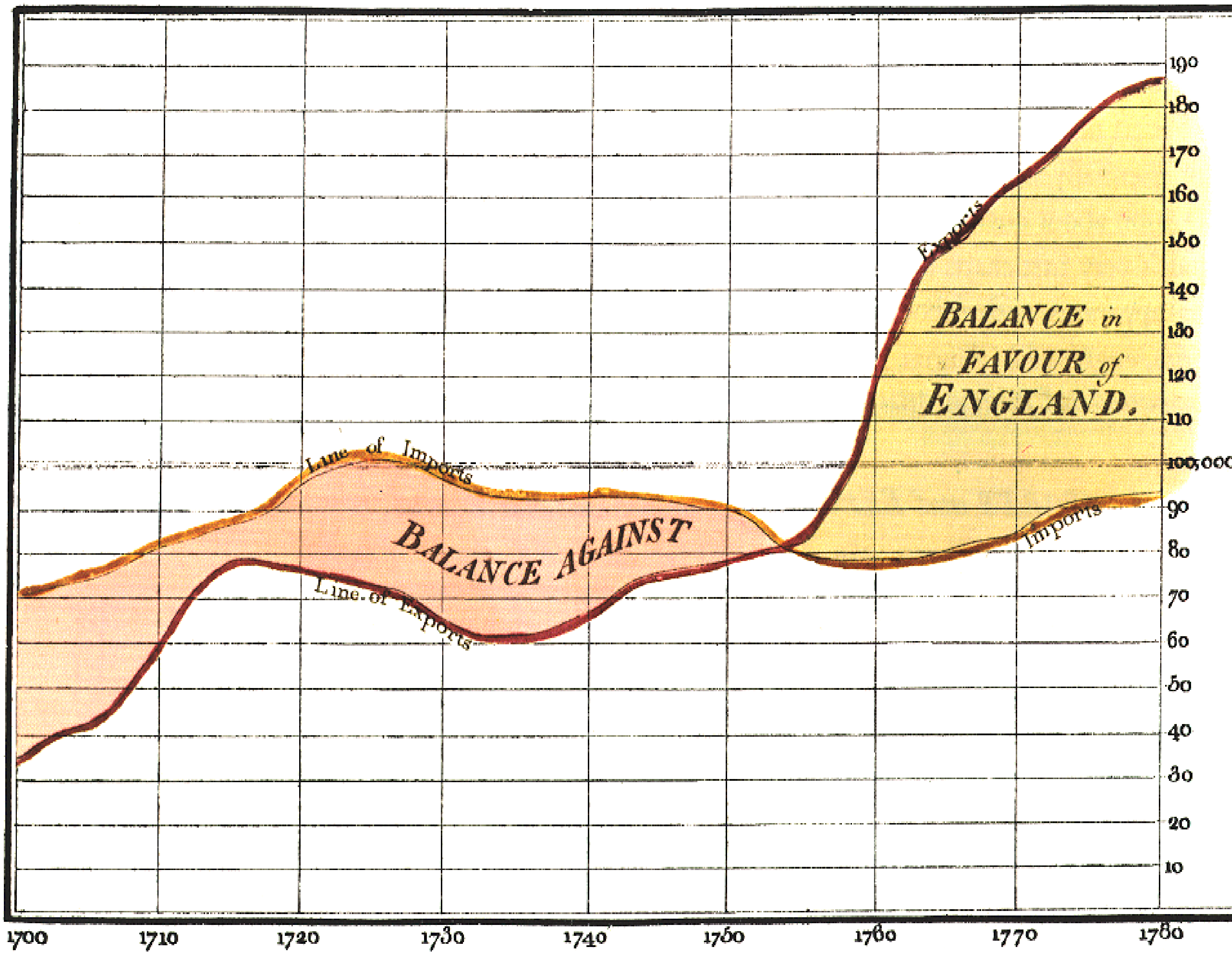


Visual encoding variables

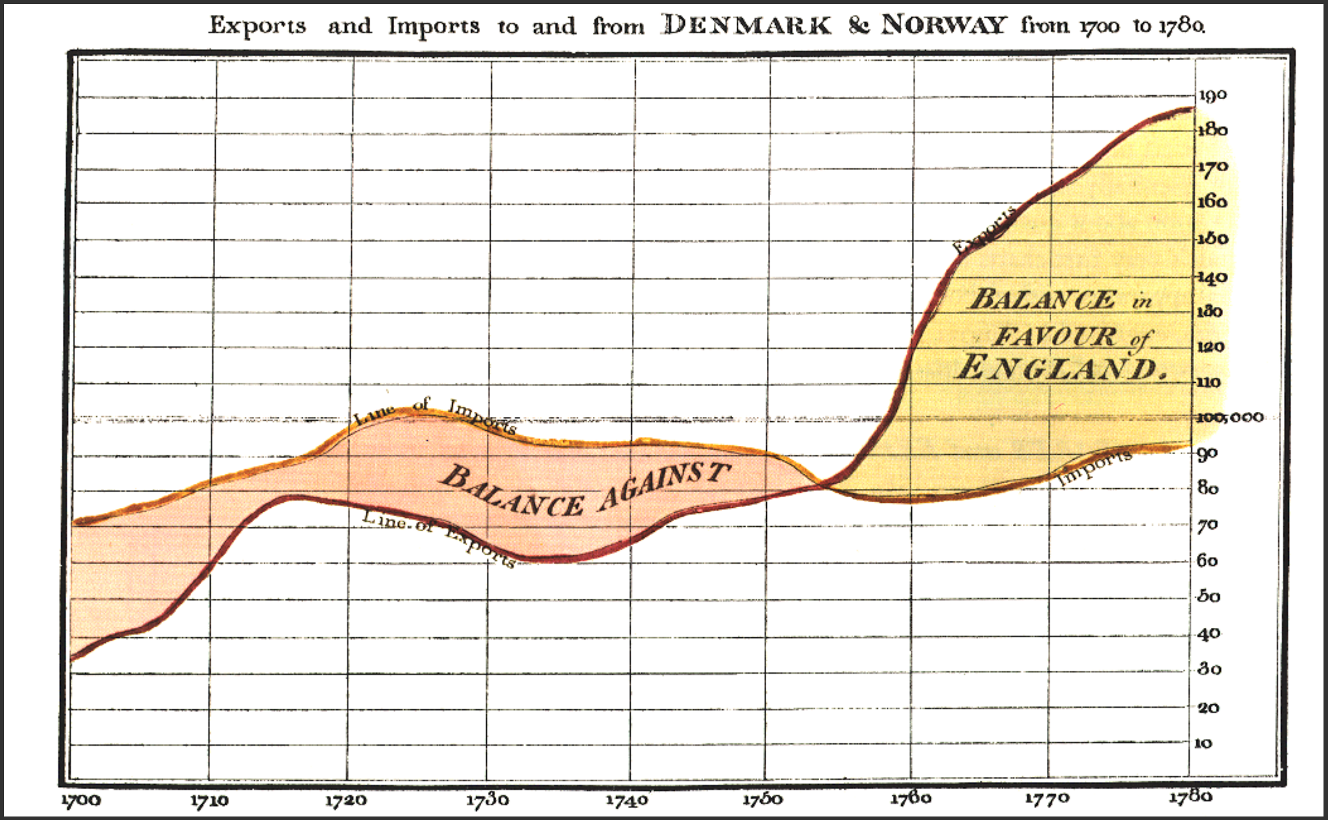
Position
Length
Area
Volume
Value
Texture
Color
Orientation
Shape
Transparency
Blur / Focus …



Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



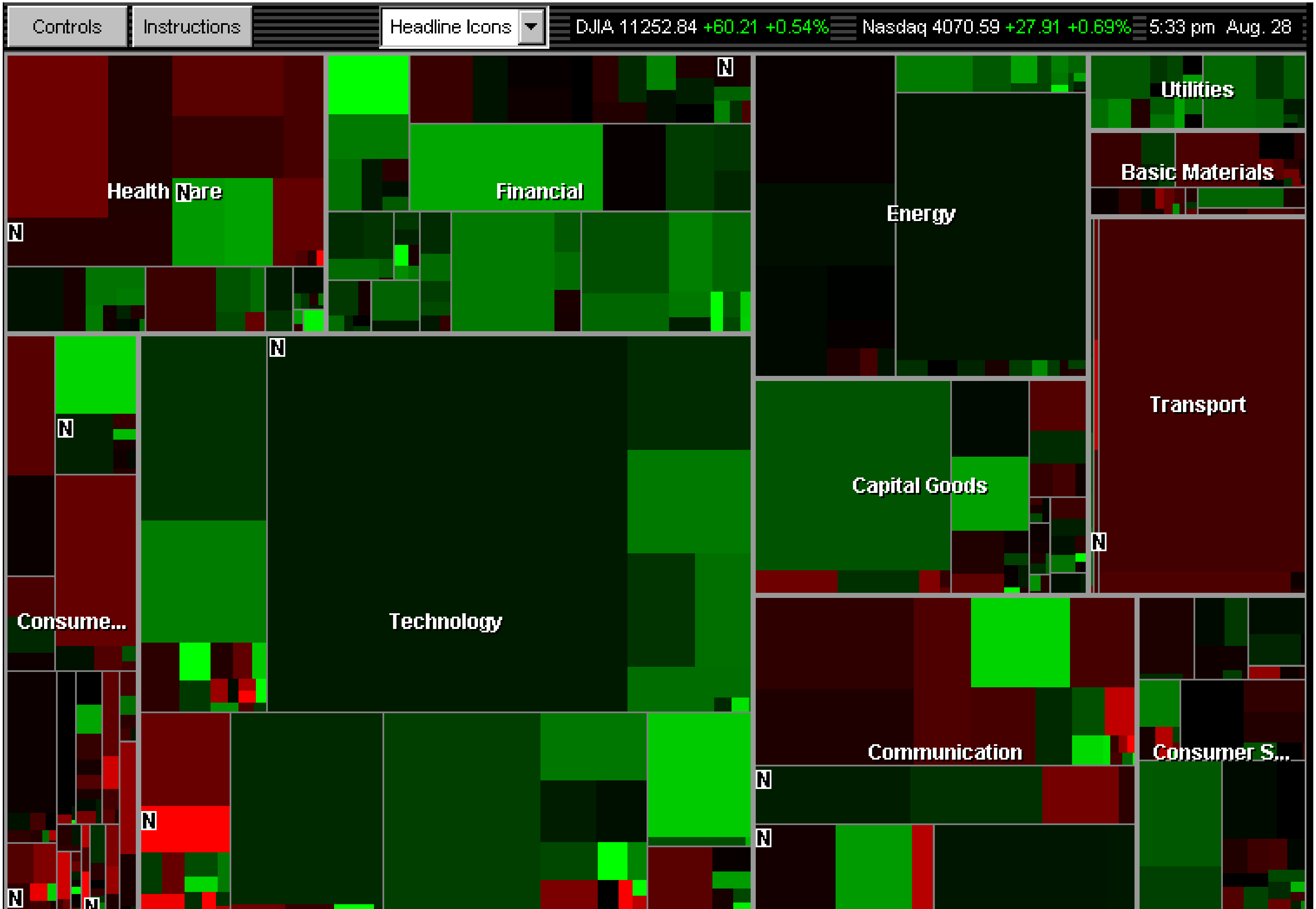
Playfair 1786



x-axis: year (Q)

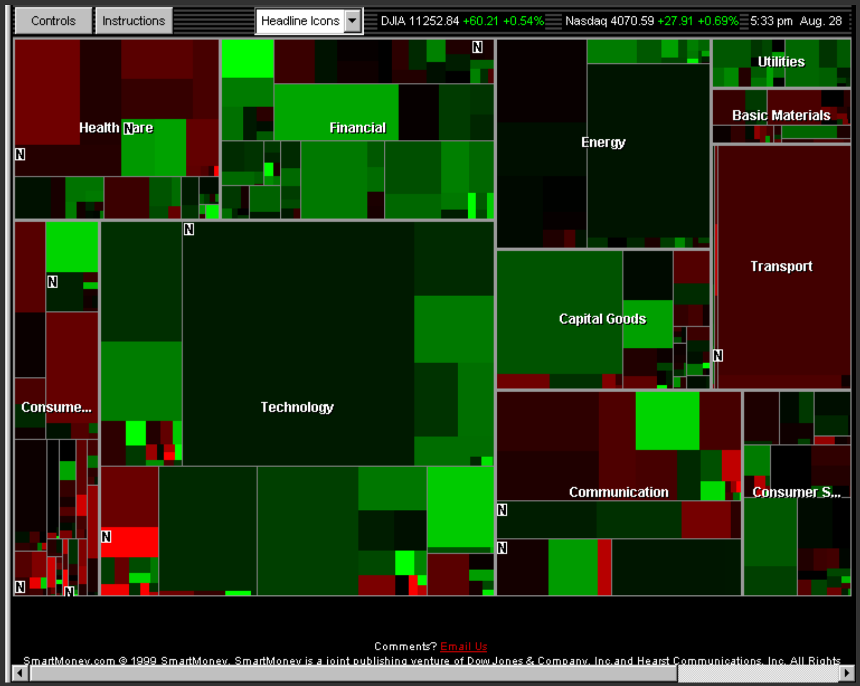
y-axis: currency (Q)

color: imports/exports (N, O)



http://www.smartmoney.com/marketmap/

Wattenberg 1998



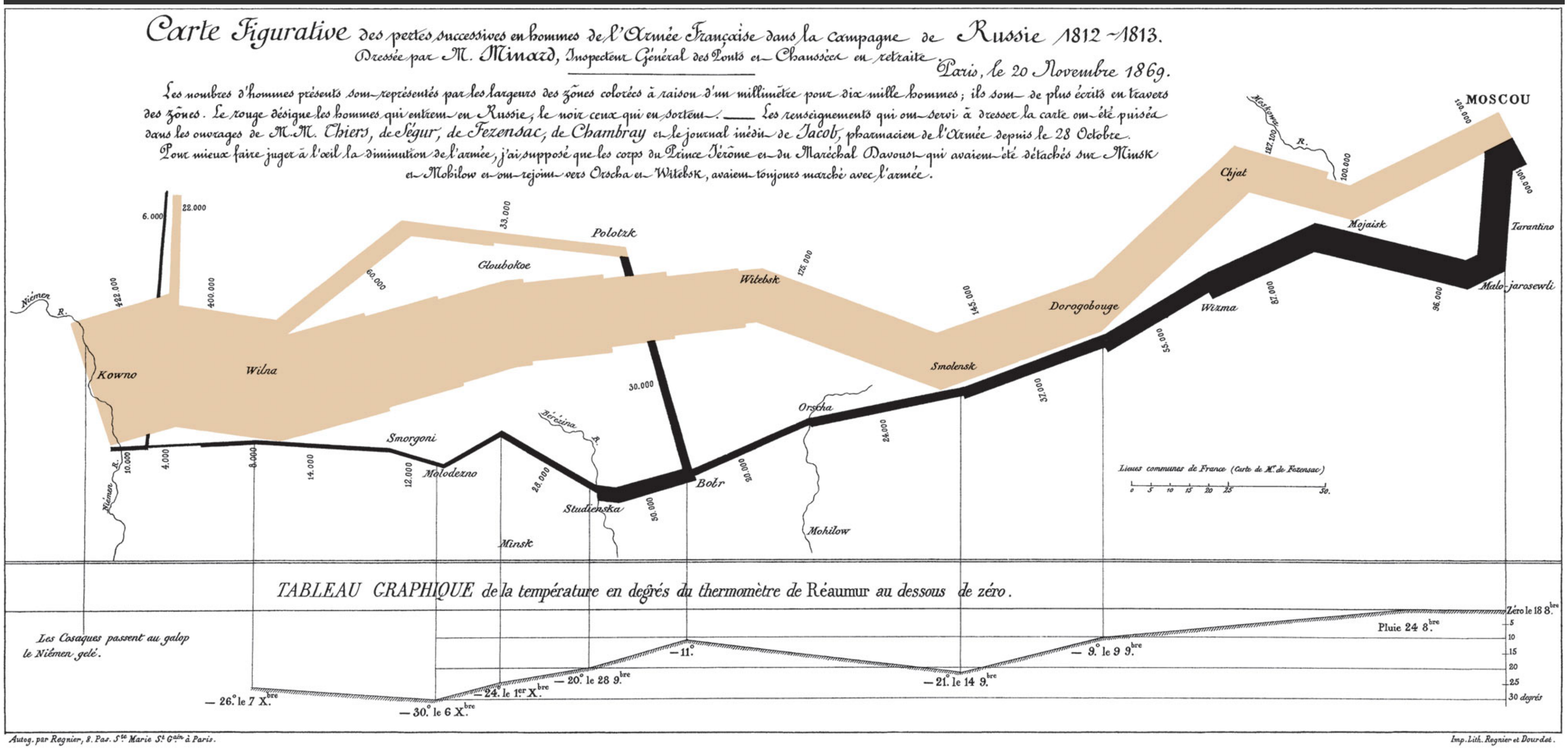
rectangle size: market cap (Q)

rectangle position: market sector (N), market cap (Q)

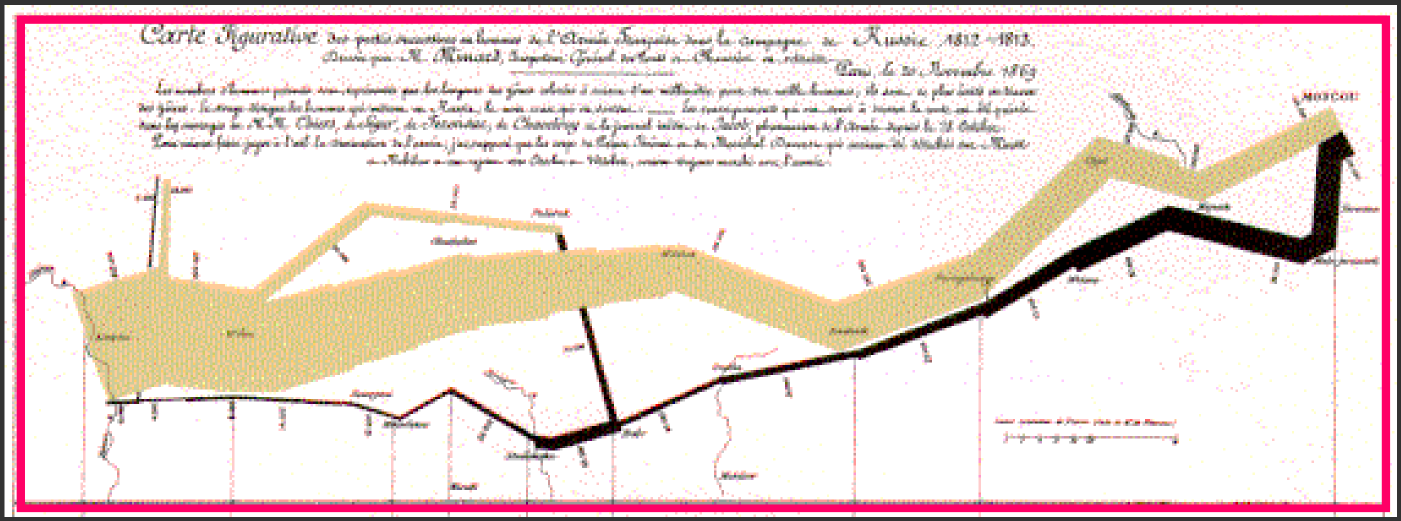
color hue: loss vs. gain (N, O)

color value: magnitude of loss or gain (Q)

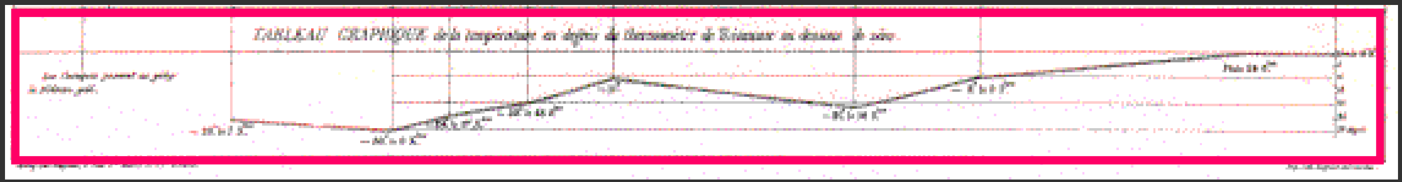
Minard 1869: Napoleon's march



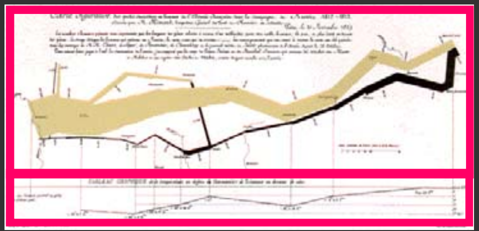
Single axis composition



+



=



Mark composition

y-axis: temperature (Q)

\+ x-axis: longitude (Q) / time (O)

= 

temp over space/time (Q x Q)

Mark composition

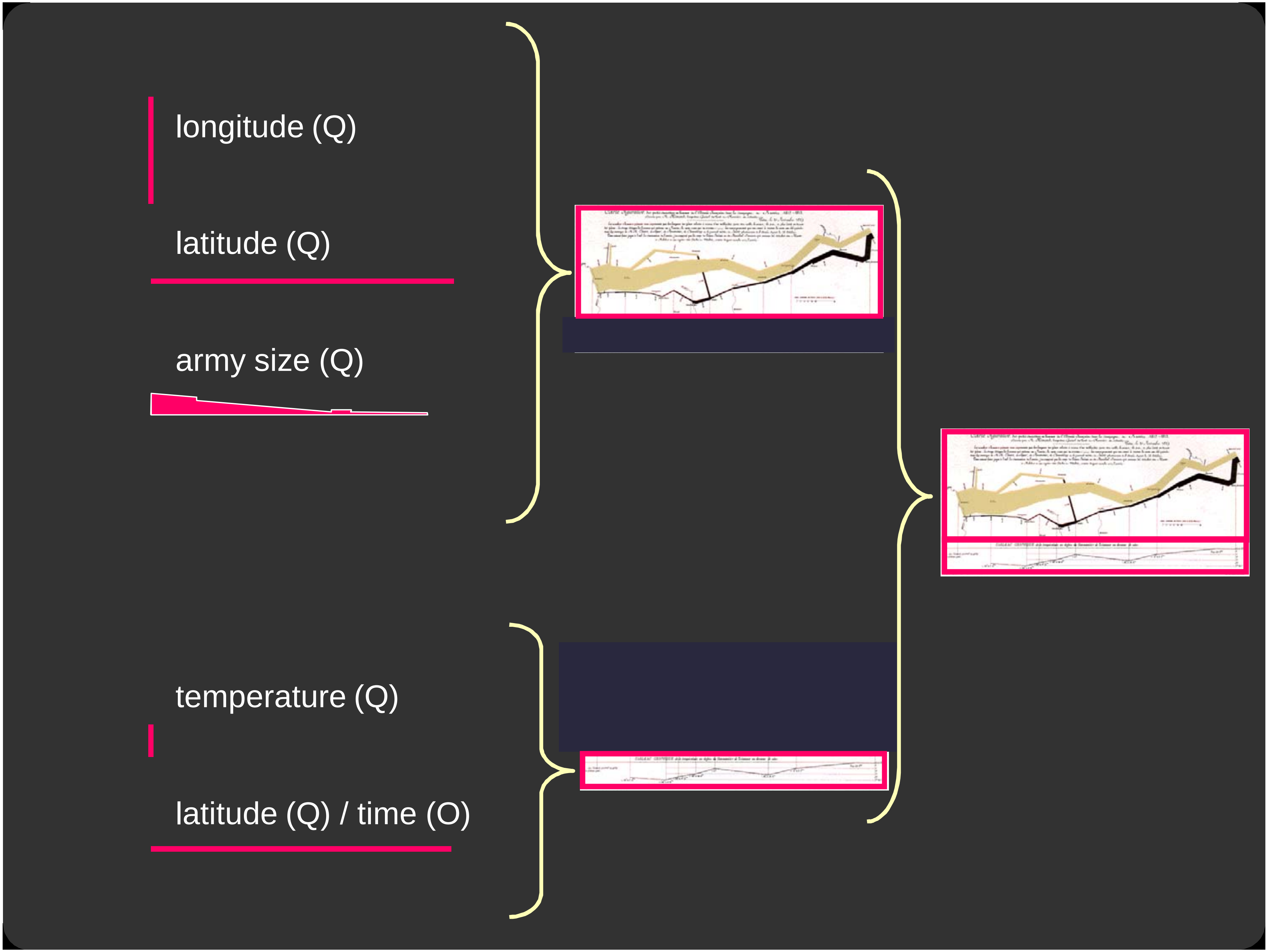
y-axis: longitude (Q)

\+ x-axis: latitude (Q)

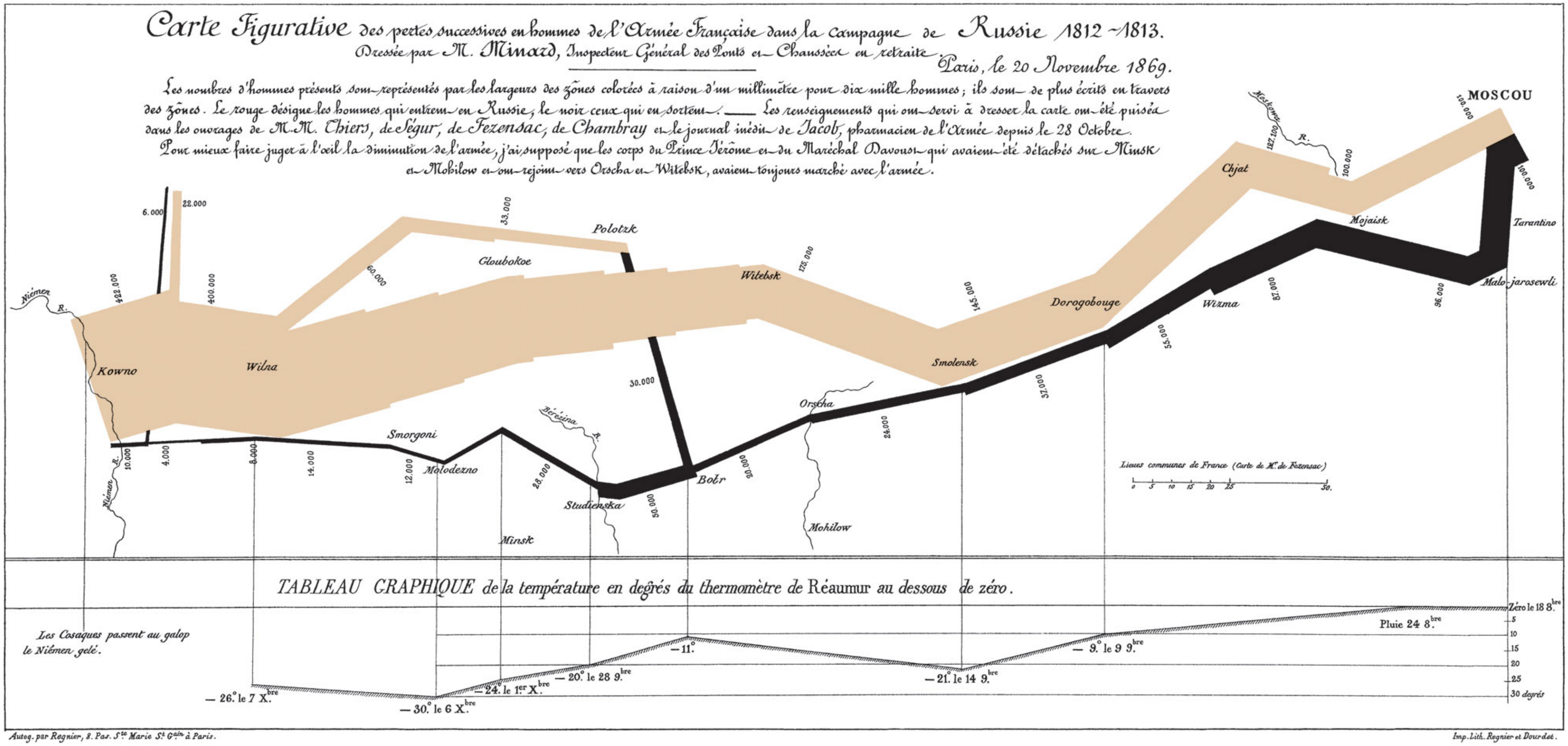
\+ width: army size (Q)

=

army position (Q x Q) and army size (Q)


longitude (Q)
latitude (Q)
army size (Q)
temperature (Q)
latitude (Q) / time (O)

Minard 1869: Napoleon's march



Depicts at least 5 quantitative variables. Any others?

Graphical Perception

Which best encodes quantities?

Position

Length

Area

Volume

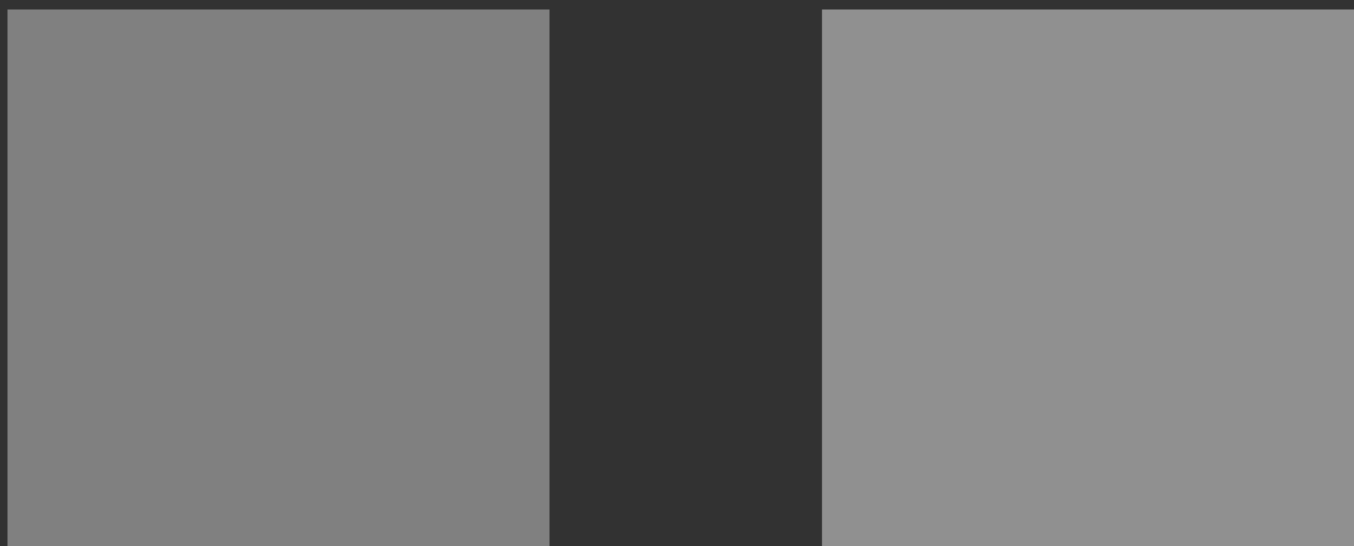
Value (Brightness)

Color Hue

Orientation (Angle)

Shape

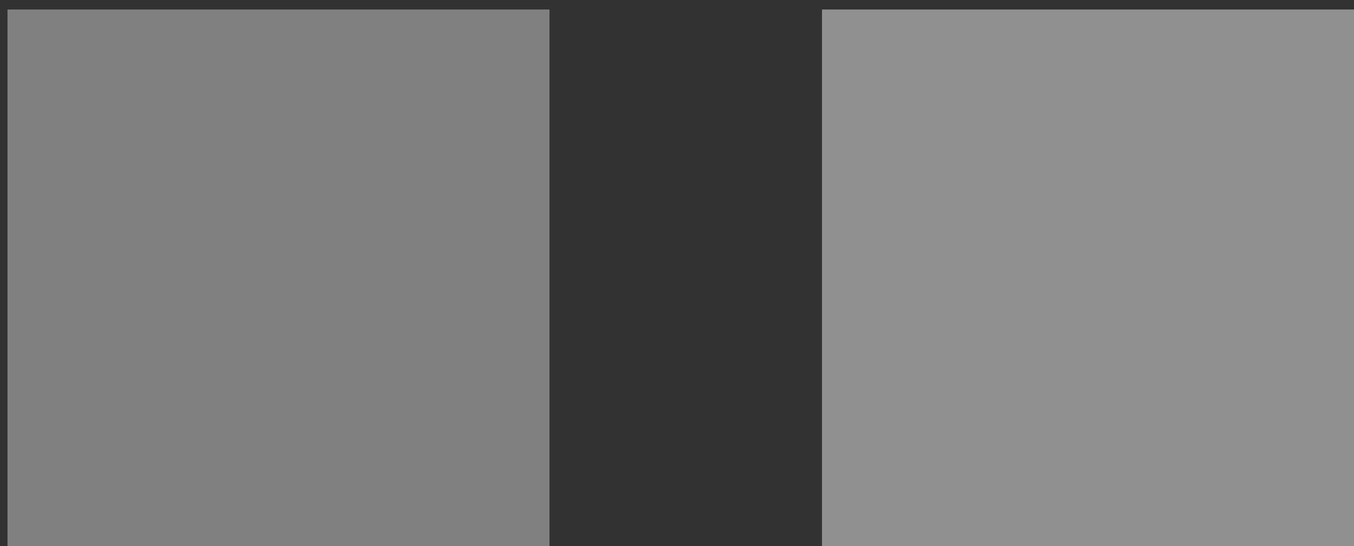
Detecting Brightness



Which is brighter?

Detecting Brightness

(128, 128, 128) (144, 144, 144)



Which is brighter?

Detecting Brightness

Which is brighter?

Detecting Brightness

(134, 134, 134) (128, 128, 128)

Which is brighter?

Just Noticeable Difference

JND (Weber's Law)

$$\Delta S = k \frac{\Delta I}{I}$$

Ratios more important than magnitude

Most continuous variation in stimuli perceived in discrete steps

Steps in font size

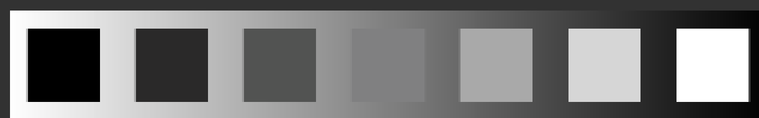
Sizes standardized in 16th century

a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
6	7	8	9	10	11	12	14	16	18	21	24	36	48	60	72

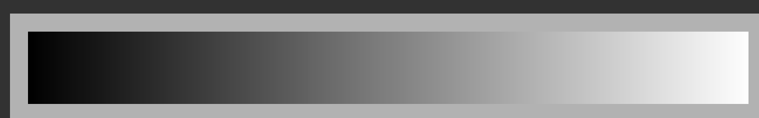
Information in color and value

Value is perceived as ordered

∴ Encode ordinal variables (O)

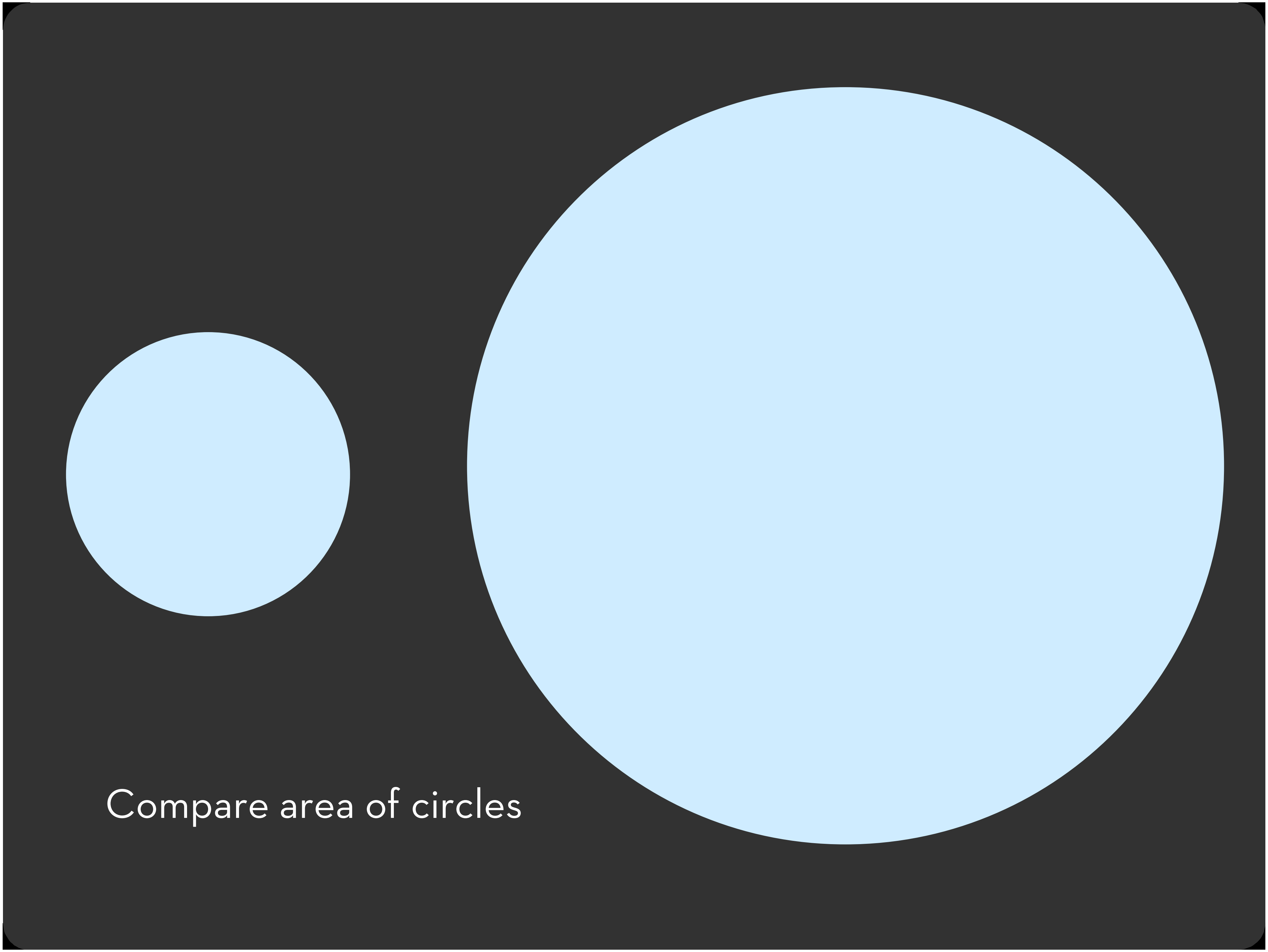


∴ Encode continuous variables (Q) [not as well]



Hue is normally perceived as unordered

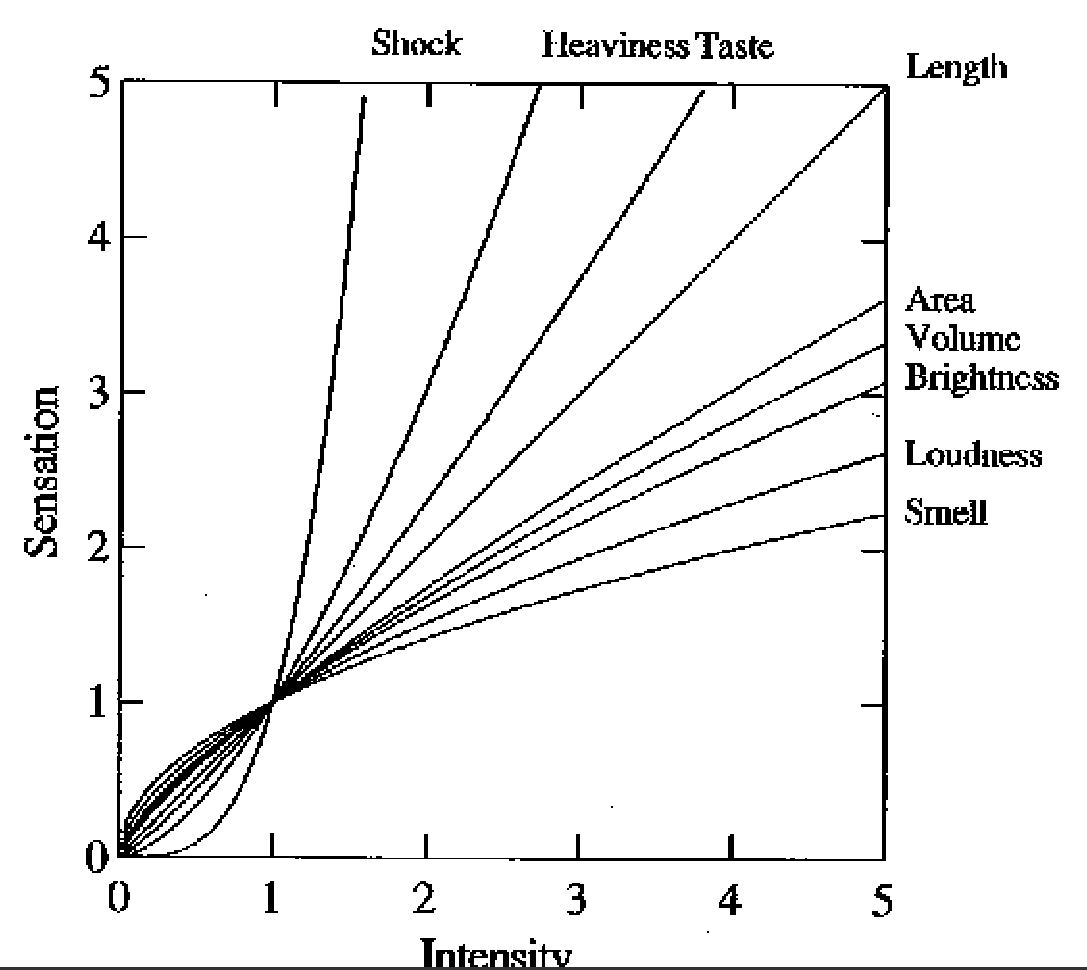
∴ Encode nominal variables (N) using color


Compare area of circles

Steven's Power Law

$$S = I^p$$

p < 1 : underestimate
p > 1 : overestimate



[graph from Wilkinson 99, based on Stevens 61]

Graduated sphere map

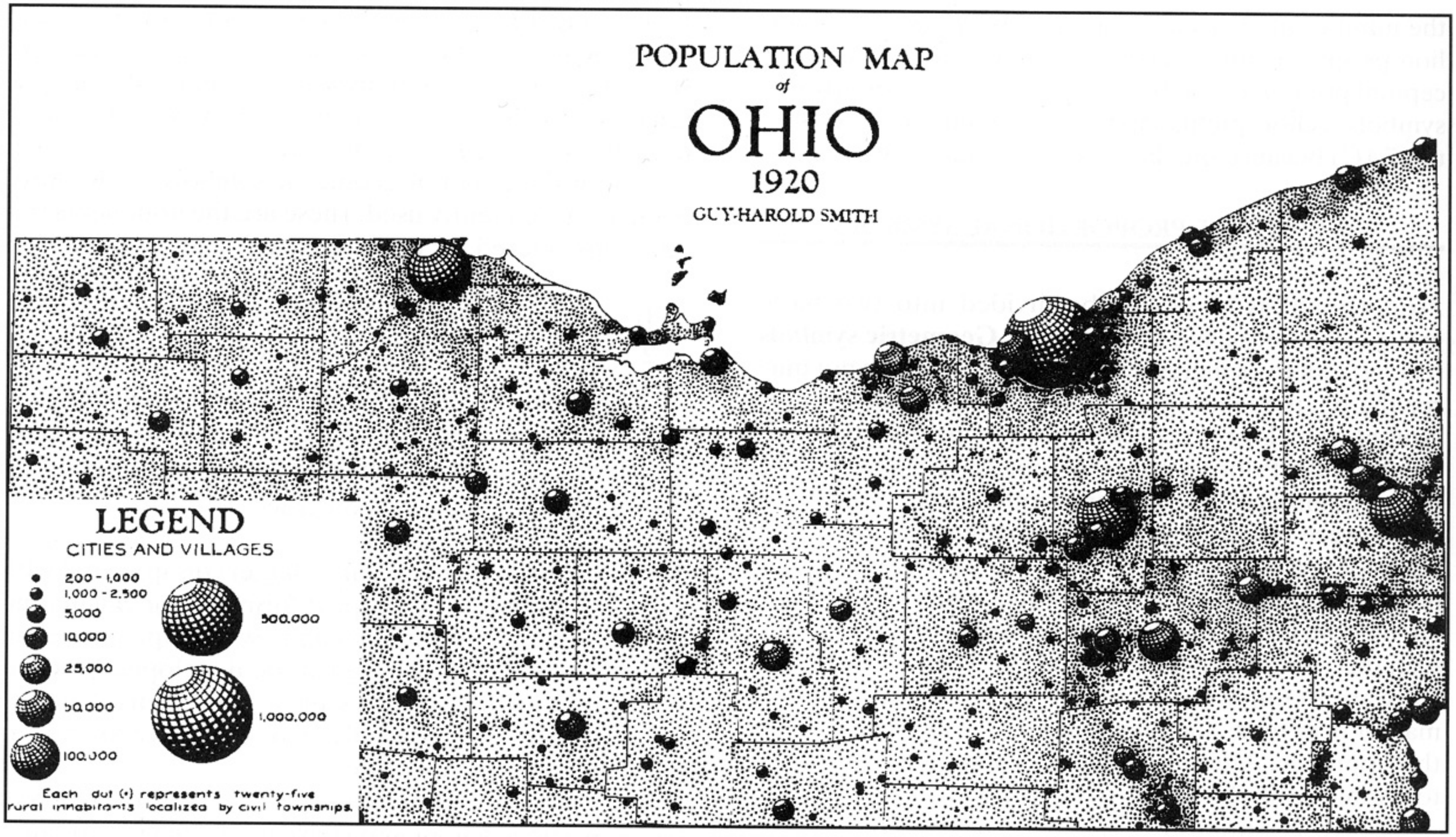
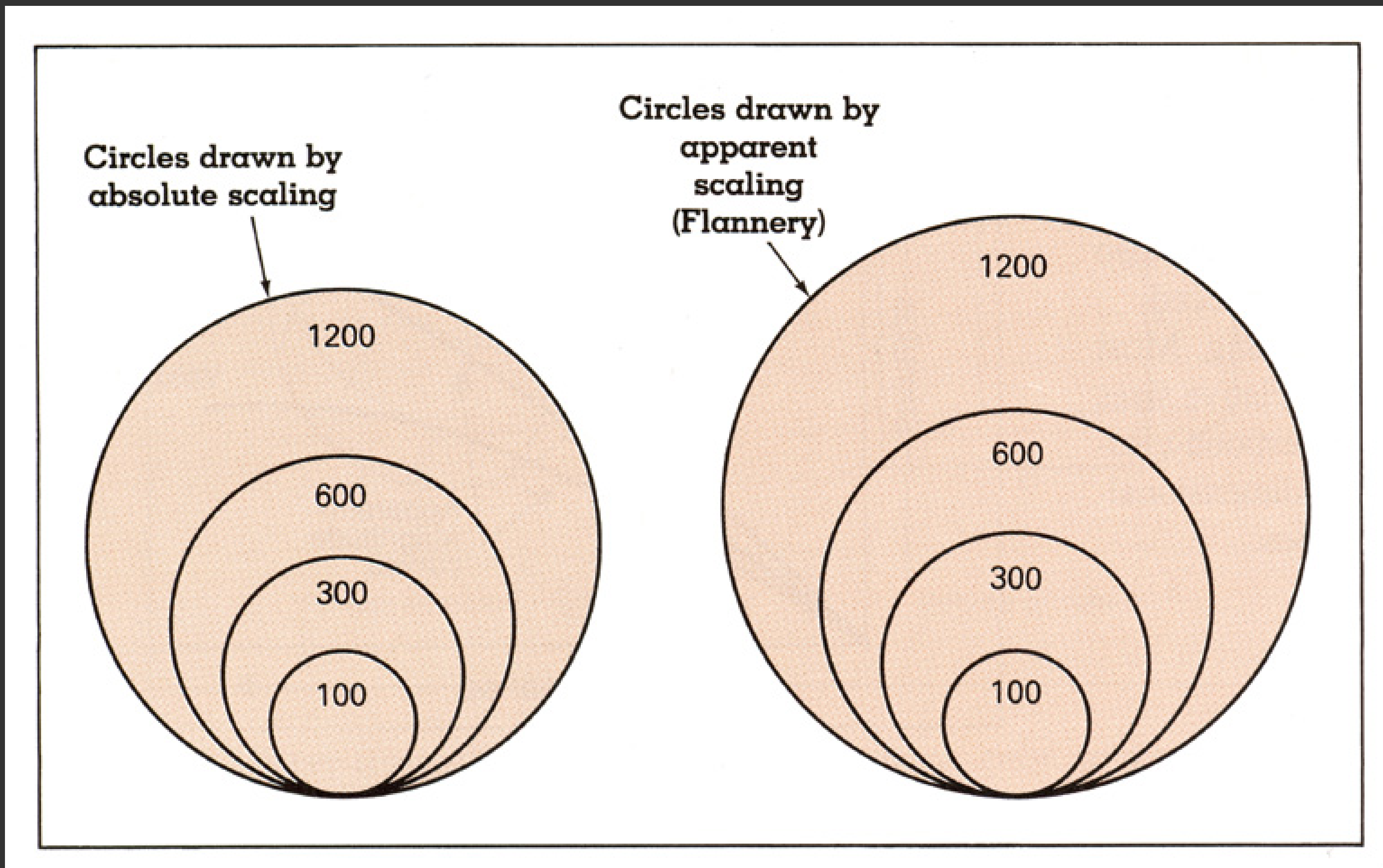


FIGURE 7.4. An eye-catching map created using three-dimensional geometric symbols. (After Smith, 1928. First published in *The Geographical Review,* 18(3), plate 4. Reprinted with permission of the American Geographical Society.)

Apparent magnitude scaling



[Cartography: Thematic Map Design, Figure 8.6, p. 170, Dent, 96]

$S = 0.98A^{0.87}$ [from Flannery 71]

How many 3’s

1281768756138976546984506985604982826762
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

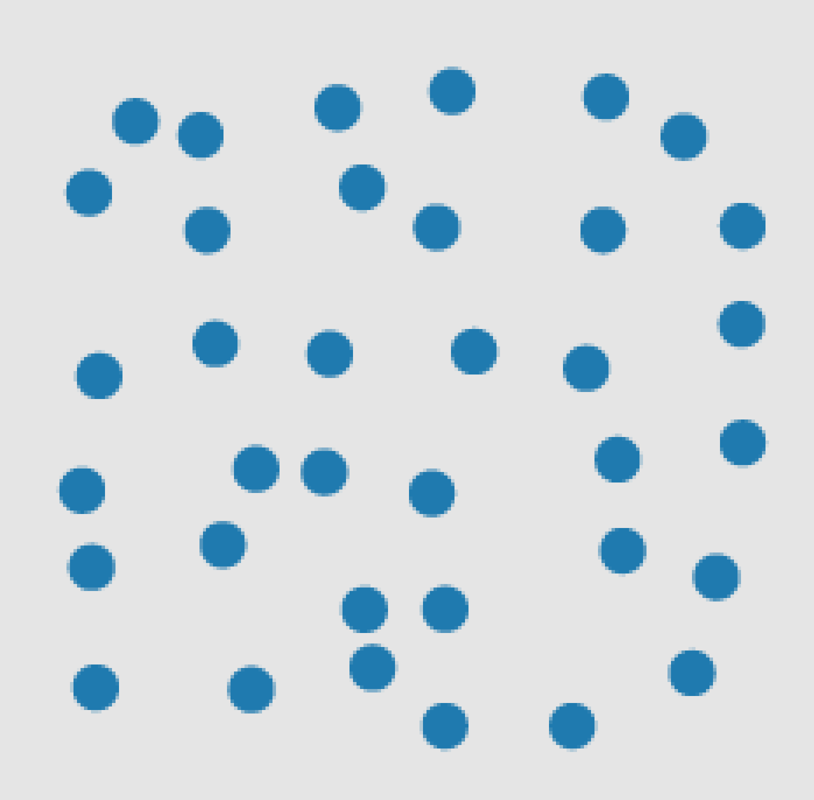
[based on slide from Stasko]

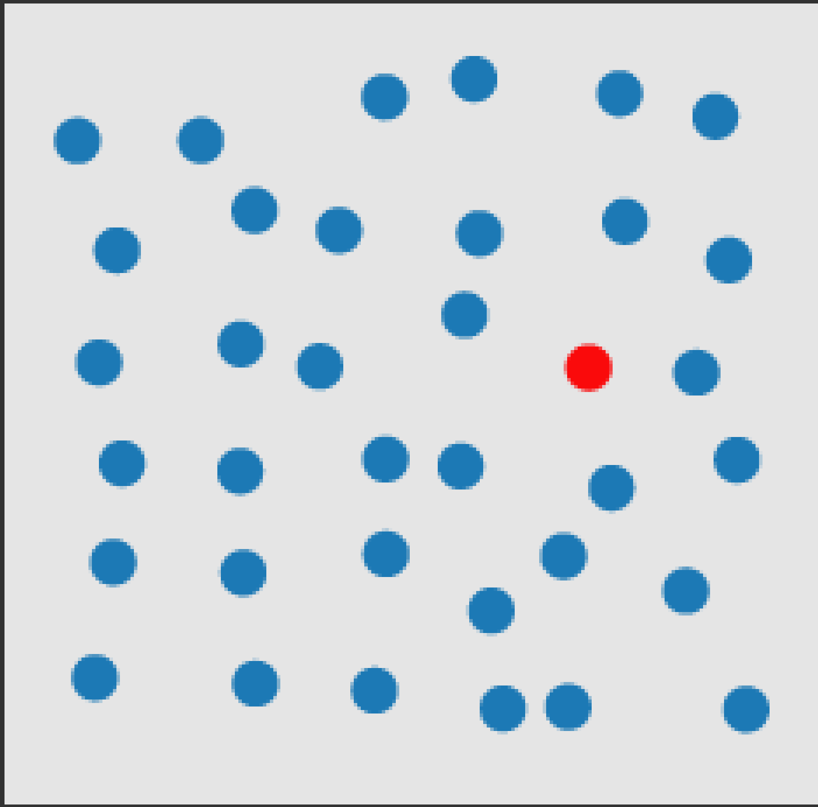
How many 3's

1281768756138976546984506985604982826762
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

[based on slide from Stasko]

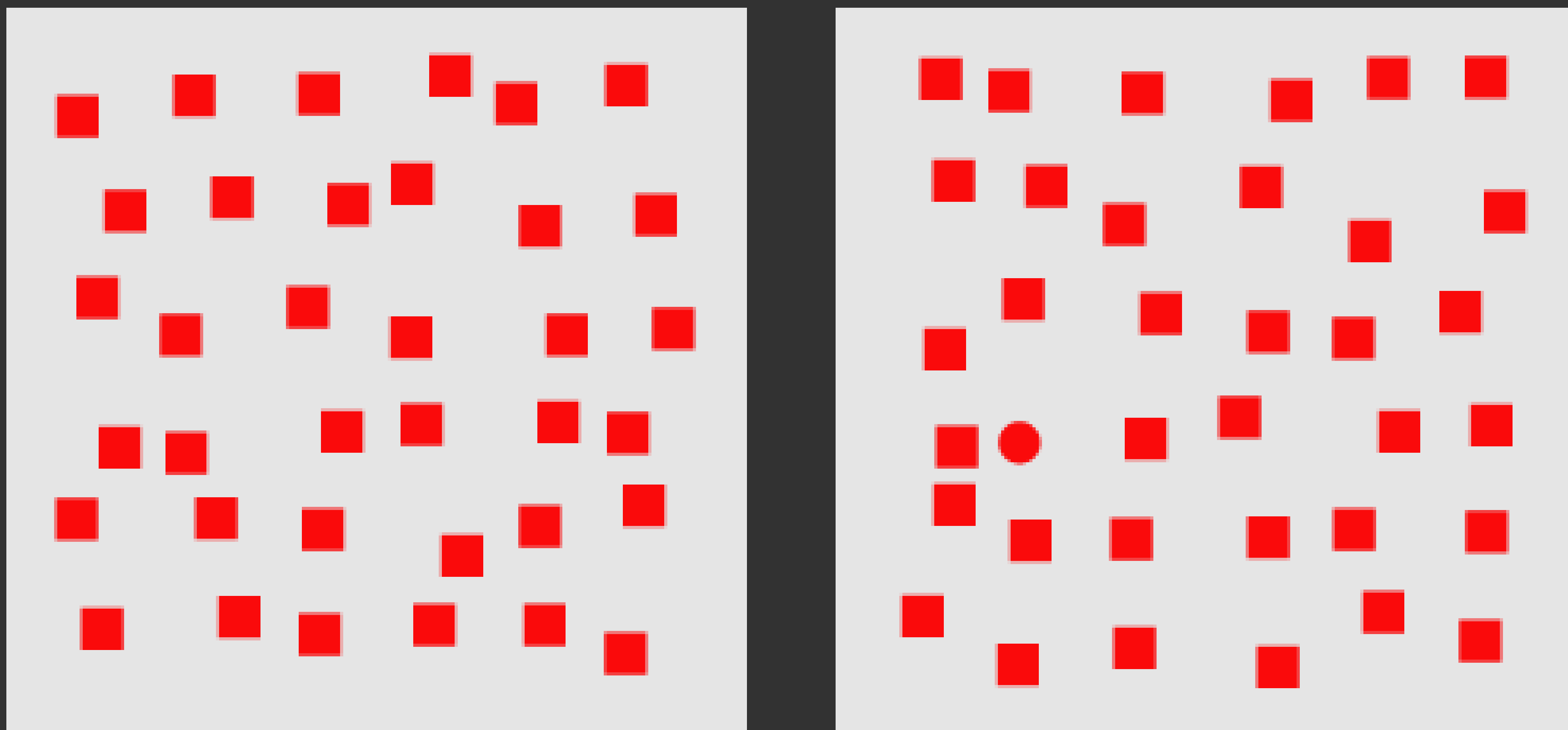
Visual pop-out: Color





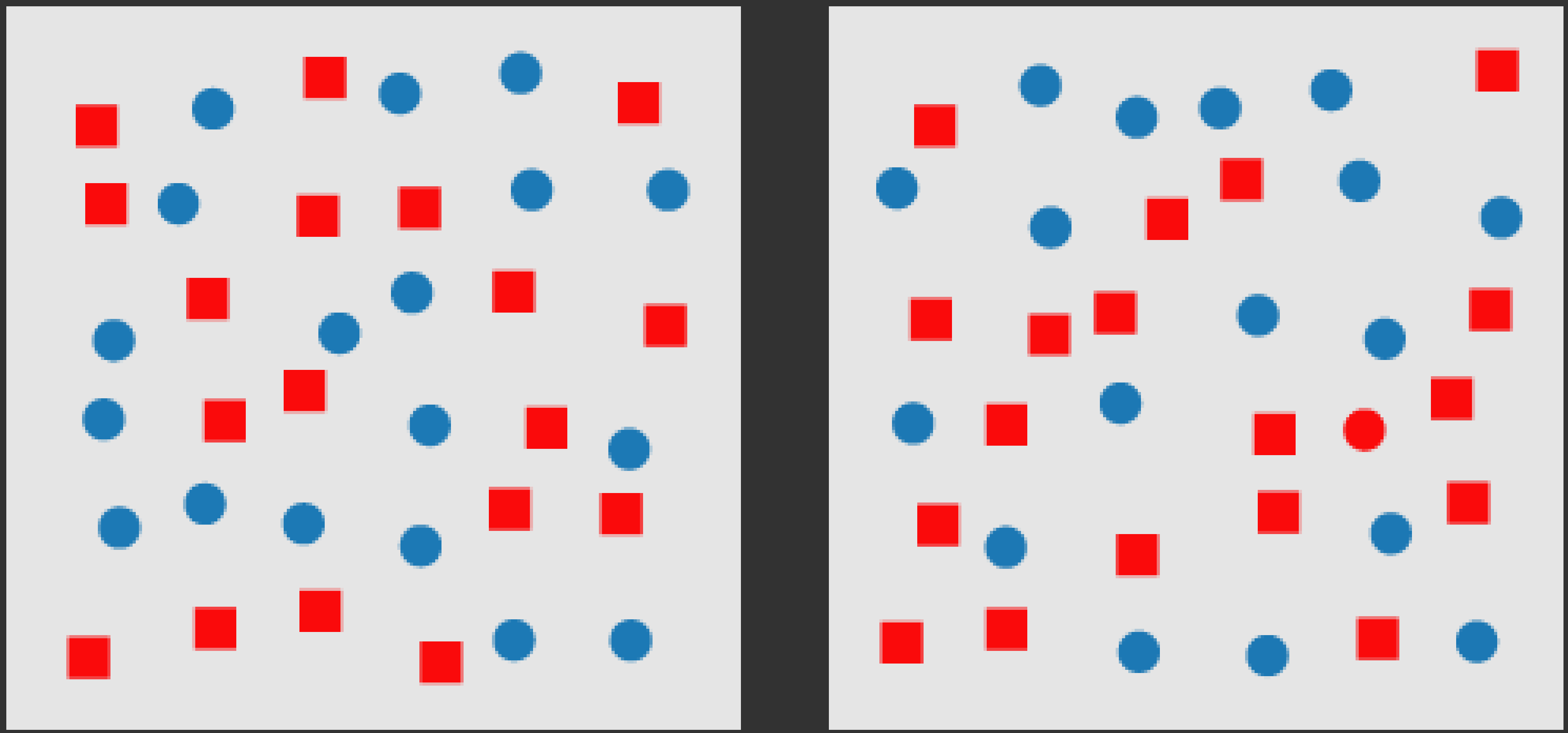
http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Visual pop-out: Shape



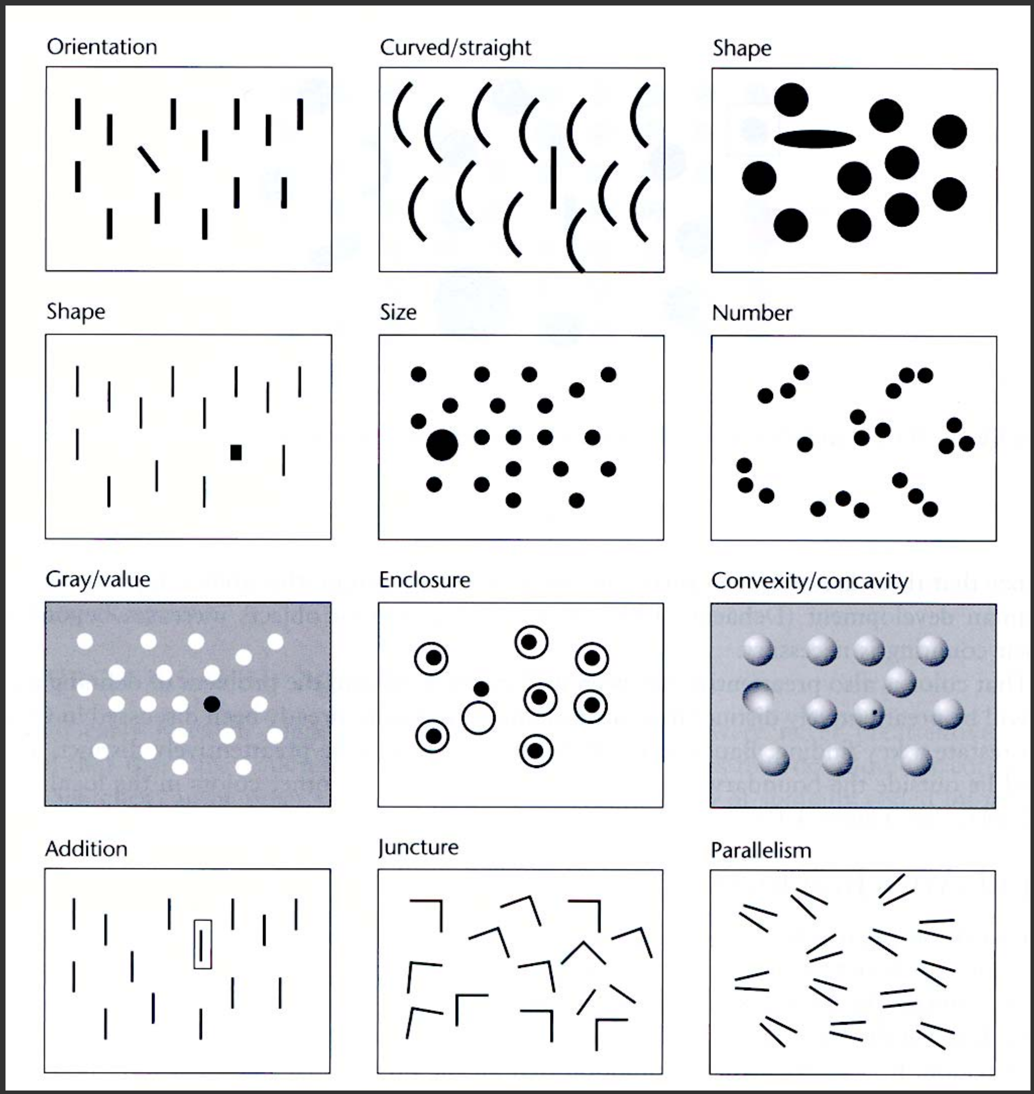
http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Feature Conjunctions



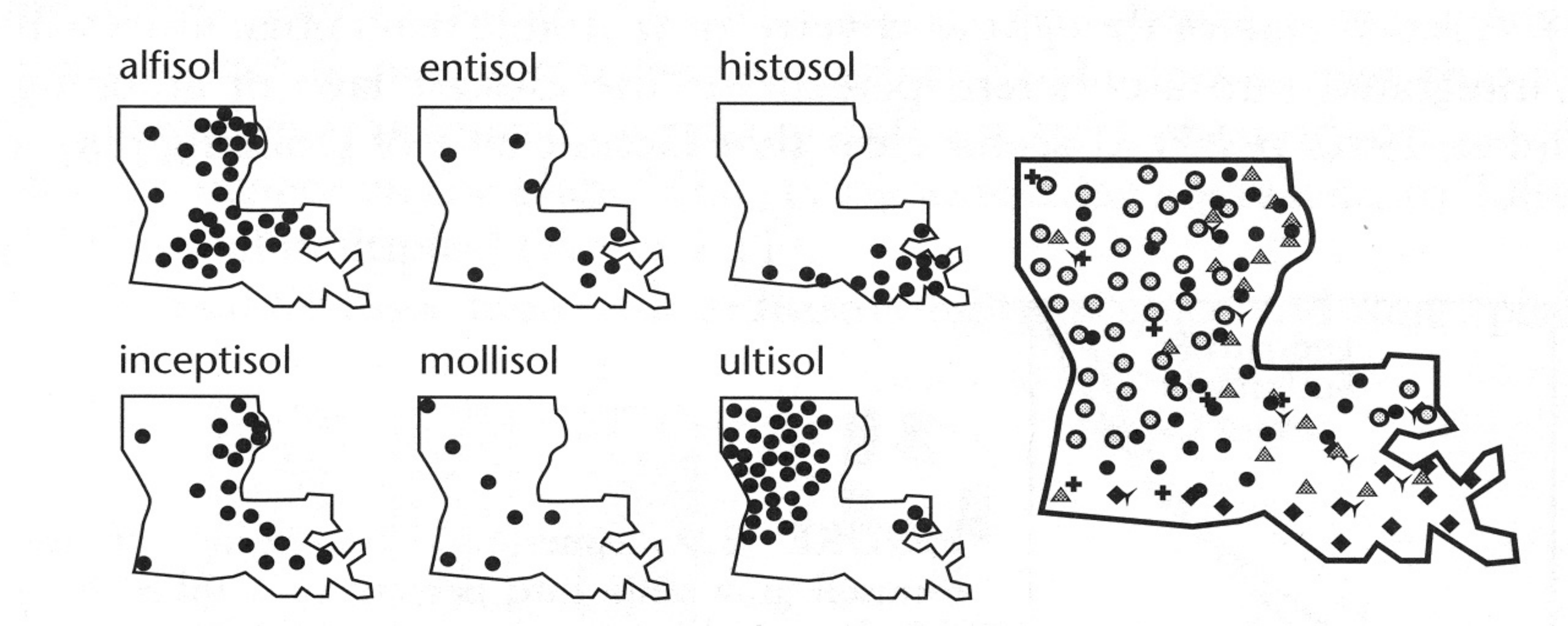
http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Pre-Attentive features



[Information Visualization. Figure 5. 5 Ware 04]

Small Multiples



[Figure 2.11, p. 38, MacEachren 95]

Which best encodes quantities?

Position

Length

Area

Volume

Value (Brightness)

Color Hue

Orientation (Angle)

Shape

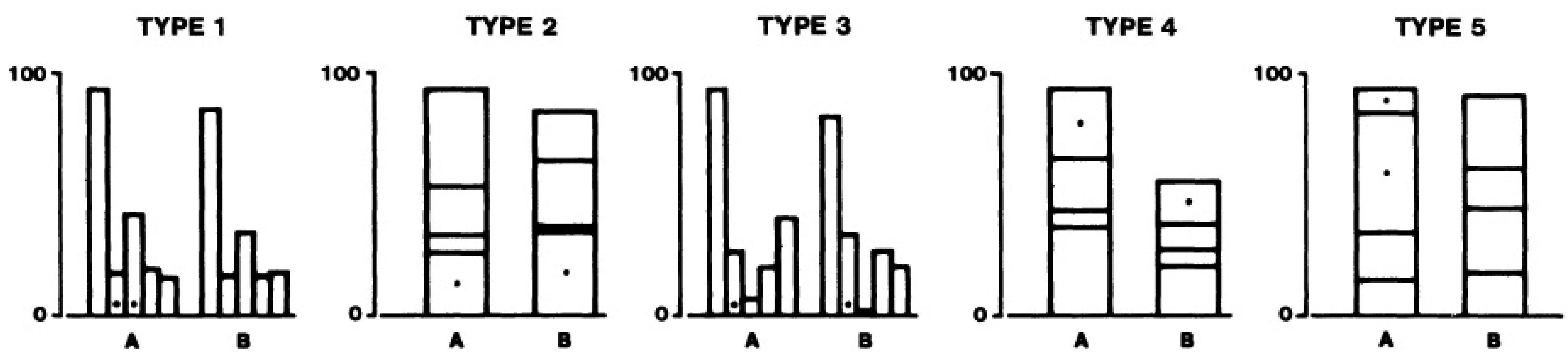


Figure 4. Graphs from position–length experiment.

Cleveland & McGill, Graphical Perception 1984

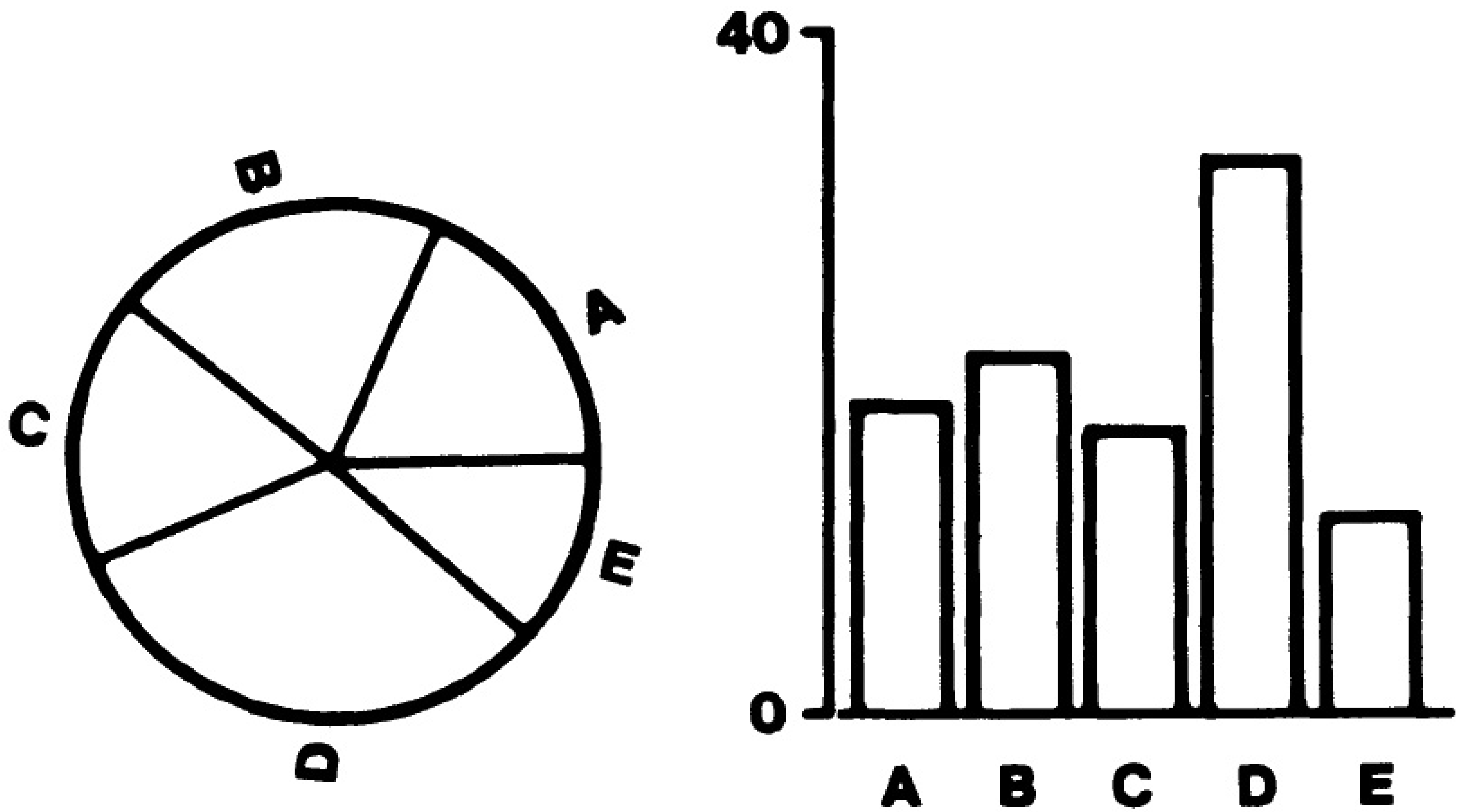


Figure 3. Graphs from position–angle experiment.

[Cleveland and McGill 84]

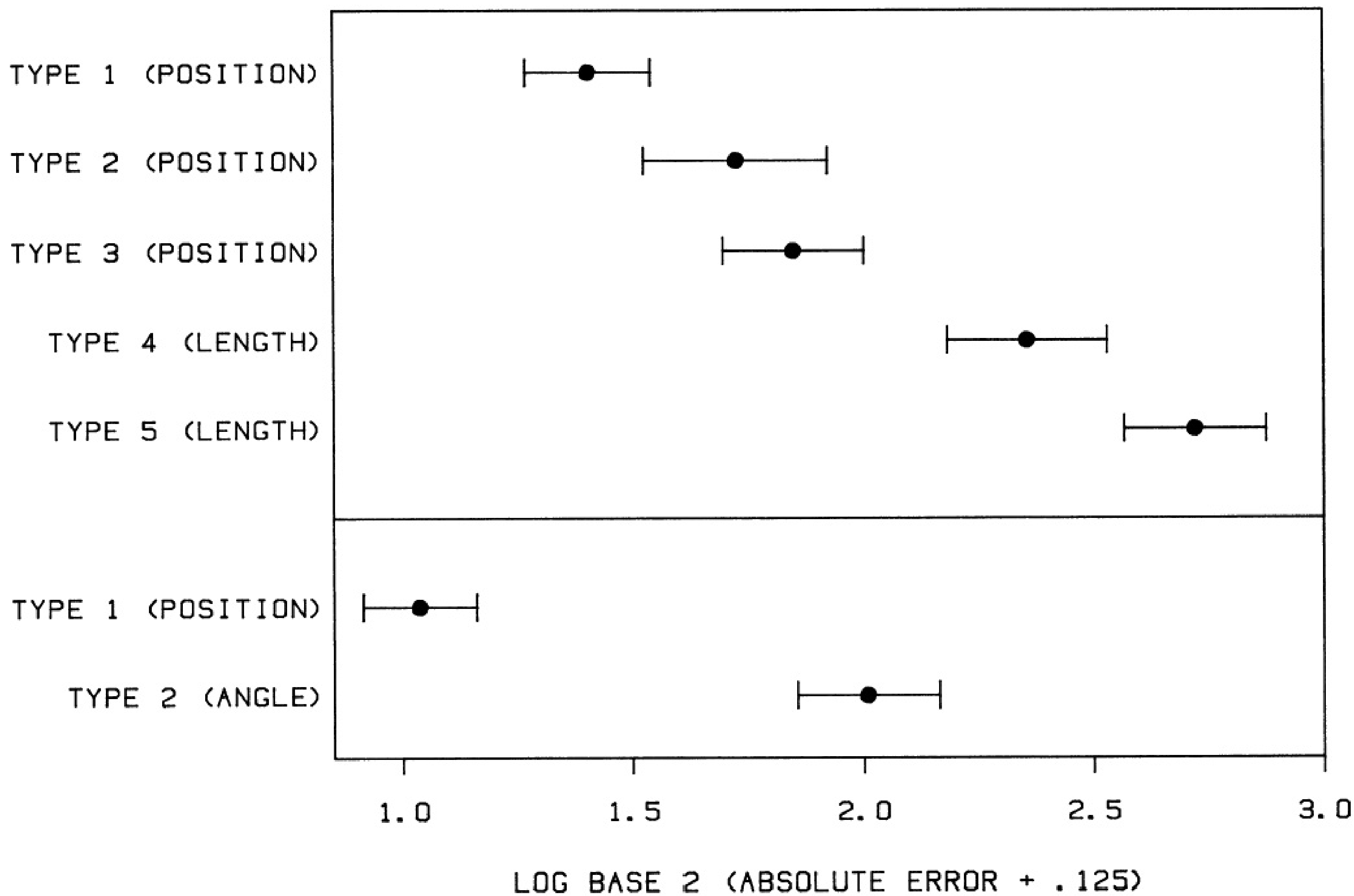
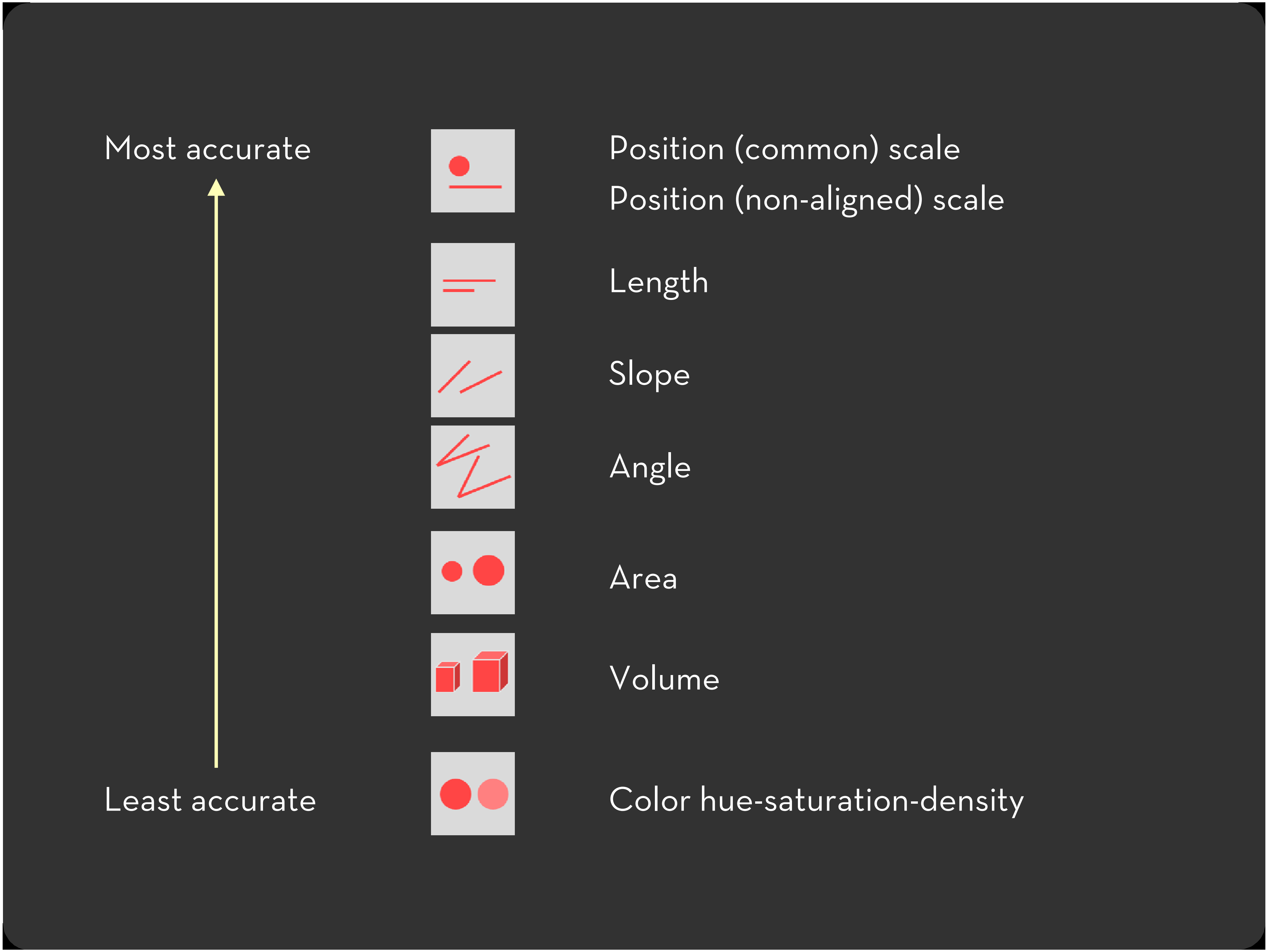


Figure 16. Log absolute error means and 95% confidence intervals for judgment types in position–length experiment (top) and position–angle experiment (bottom).

[Cleveland and McGill 84]


Most accurate
Least accurate
Position (common) scale
Position (non-aligned) scale
Length
Slope
Angle
Area
Volume
Color hue-saturation-density

Combinatorics of Encodings

Challenge:

Pick the best encoding from the exponential number of possibilities $(n+1)^8$

Principle of Consistency:

The properties of the image (visual variables) should match the properties of the data.

Principle of Importance Ordering:

Encode the most important information in the most effective way.

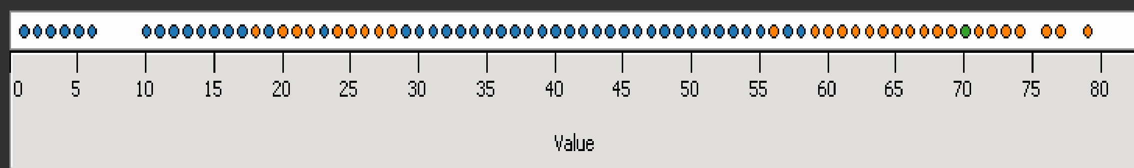
Design Criteria (Mackinlay)

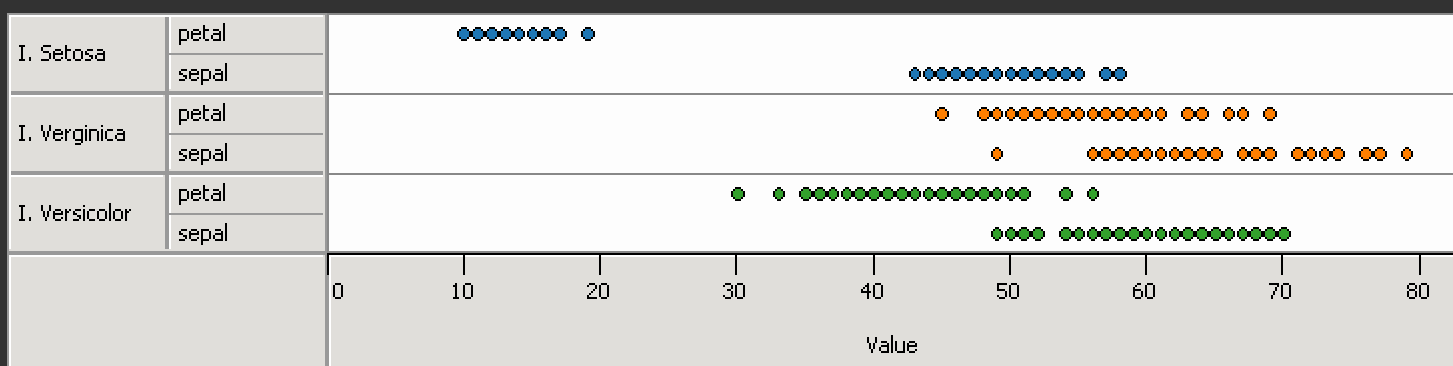
Expressiveness

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express *all* the facts in the set of data, and *only* the facts in the data.

Cannot express the facts

A one-to-many (1 → N) relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position





Expresses facts not in the data

A length is interpreted as a quantitative value;

∴ Length of bar says something untrue about N data

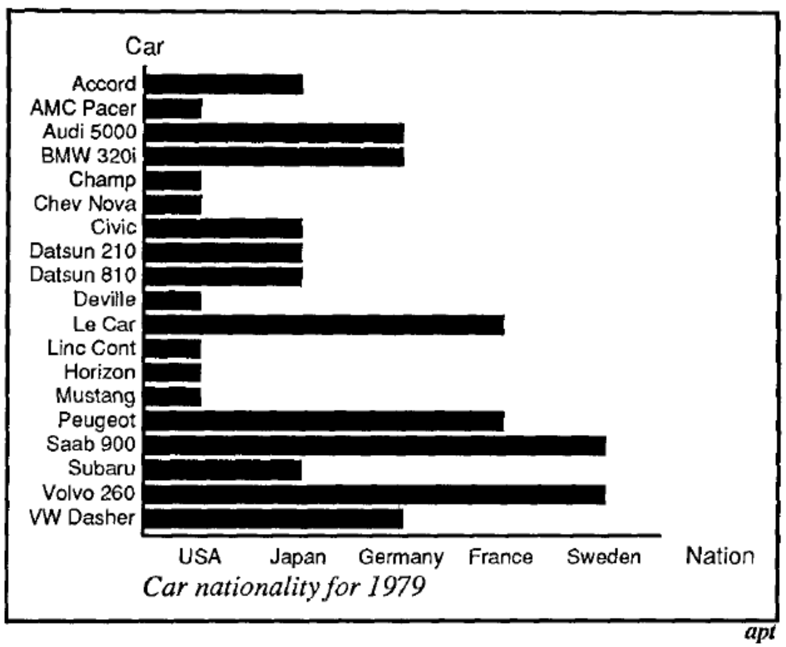


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, APT, 1986]

Design Criteria (Mackinlay)

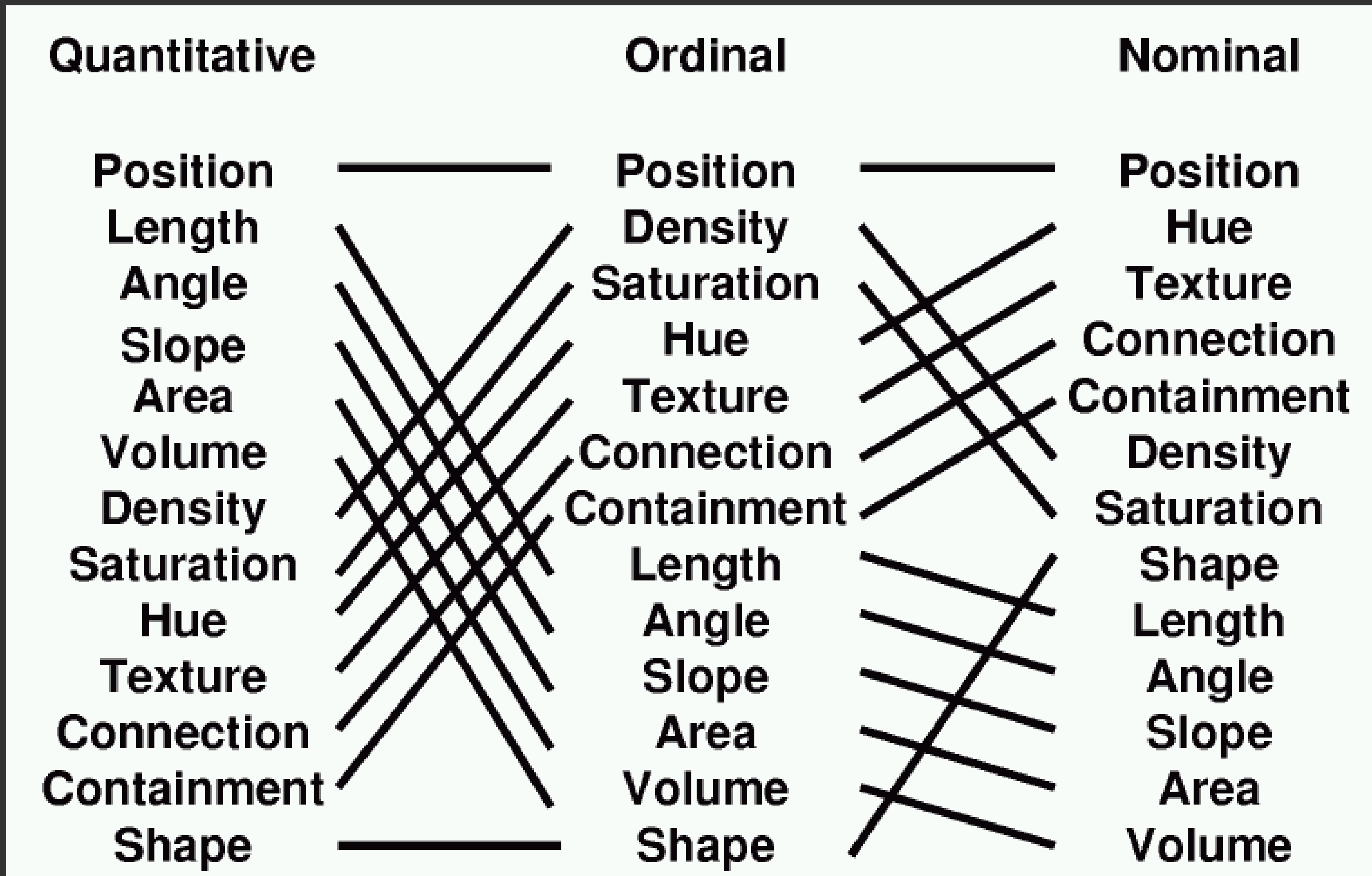
Expressiveness

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express *all* the facts in the set of data, and *only* the facts in the data.

Effectiveness

A visualization is more effective than another visualization if the information conveyed by one visualization is more readily *perceived* than the information in the other visualization.

Mackinlay's Ranking



Conjectured *effectiveness* of the encoding

Visualization
(Re-)Design Exercise

In-Class Design Exercise

Task: Analyze and Re-design visualization

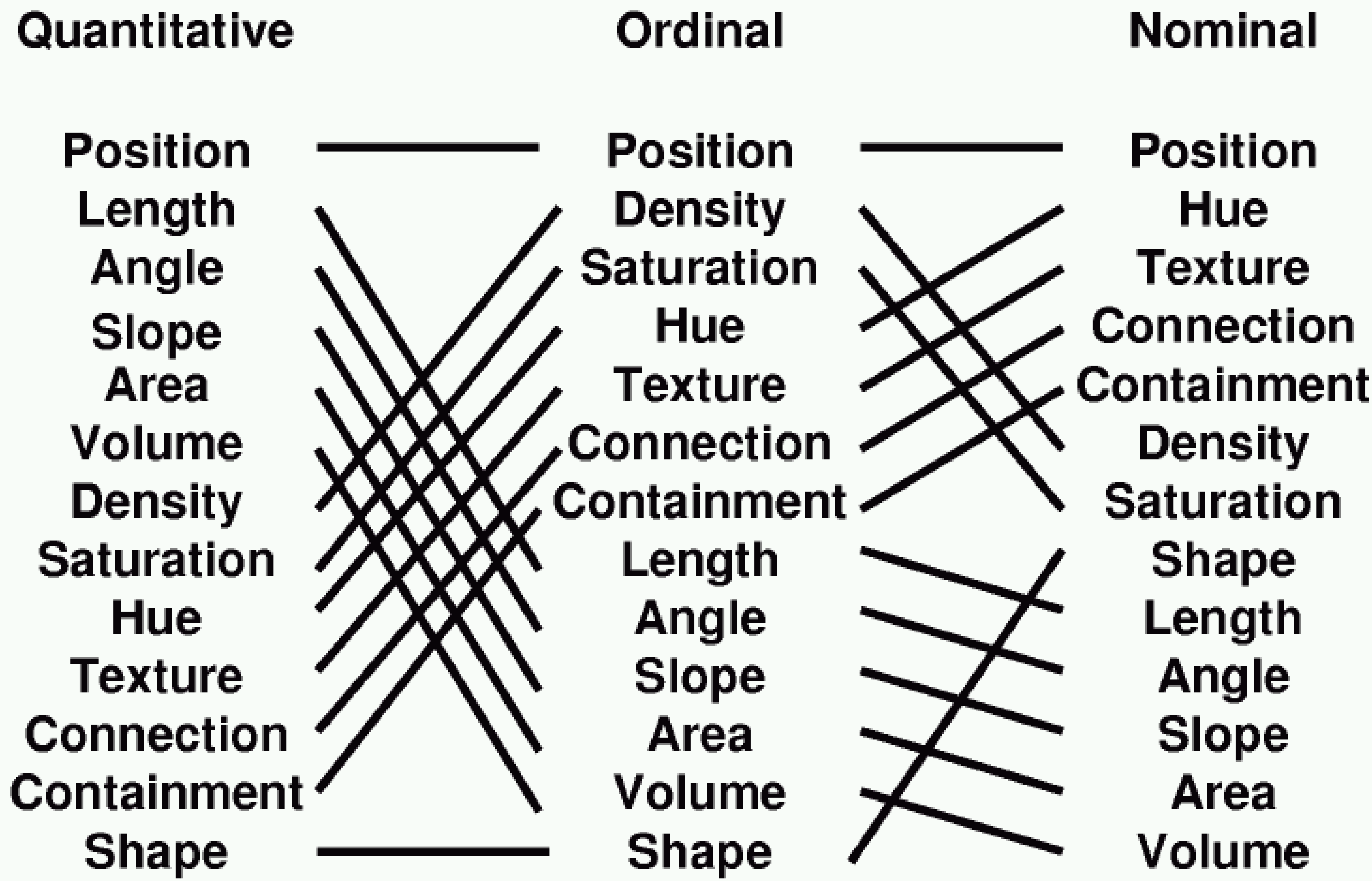
- Identify data variables (N,O,Q) and encodings
- Critique the design: what works, what doesn't
- Sketch a re-design to improve communication
- Present result to the class

Break into groups (~5 people per group)

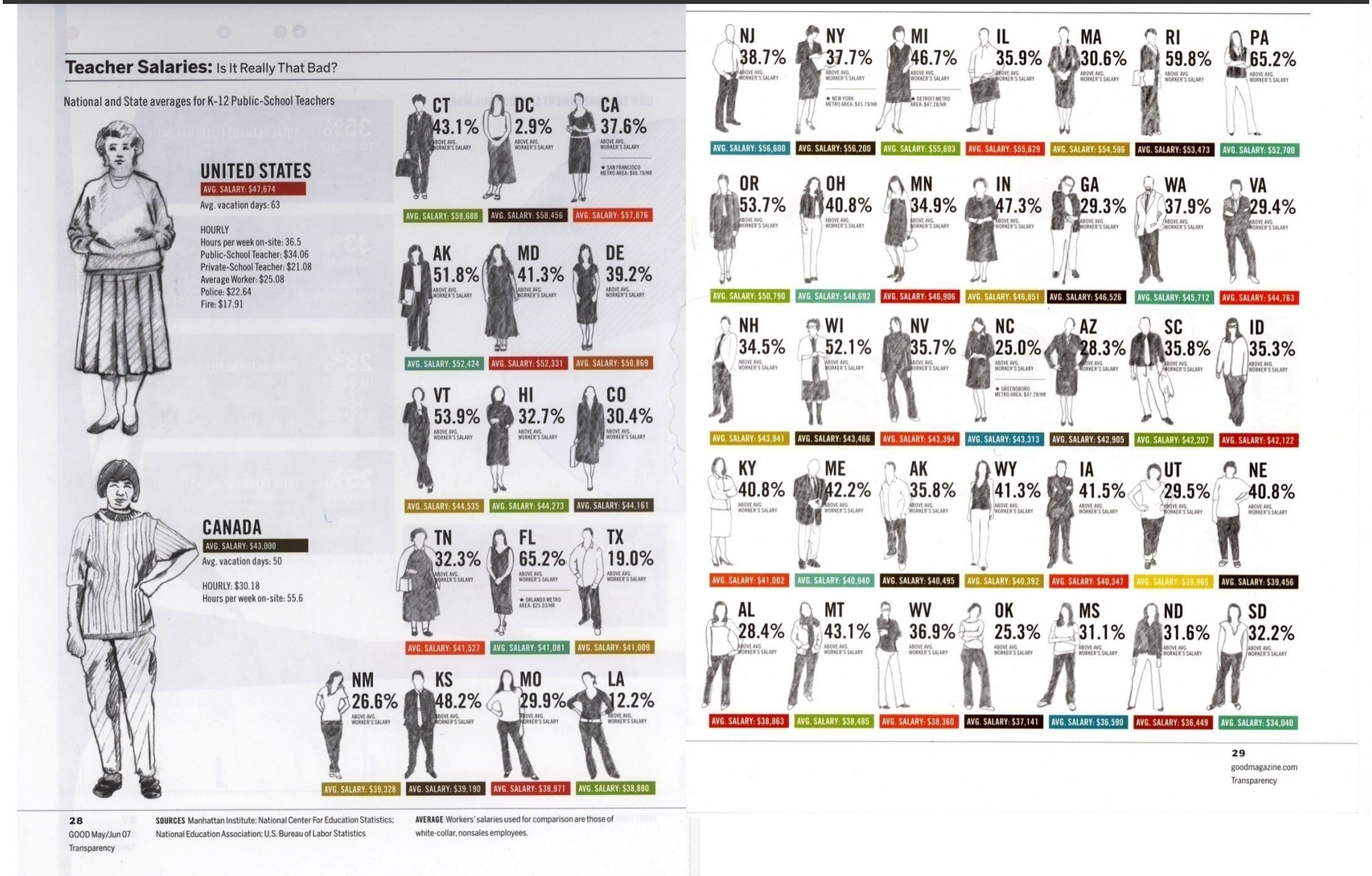
You have ~35 minutes.

Re-Design Presentation (~4 min)

1. Describe the data and visualization
2. Present your critique
3. Share your re-design



Mackinlay's conjectured *effectiveness* of visual encodings



Teacher Salaries: Is It Really That Bad?

National and State averages for K-12 Public-School Teachers

UNITED STATES
AVG. SALARY: $47,674
Avg. vacation days: 63

HOURLY
Hours per week on-site: 36.5
Public-School Teacher: $34.06
Private-School Teacher: $21.08
Average Worker: $25.08
Police: $22.64
Fire: $17.91

CANADA
AVG. SALARY: $43,000
Avg. vacation days: 50

HOURLY: $30.18
Hours per week on-site: 55.6

State	Above avg. worker's salary	Avg. salary
CT	43.1%	$58,688
DC	2.9%	$58,456
CA	37.6%	$57,876
NJ	38.7%	$56,600
NY	37.7%	$56,200
MI	46.7%	$55,693
IL	35.9%	$55,629
MA	30.6%	$54,596
RI	59.8%	$53,473
PA	65.2%	$52,700
AK	51.8%	$52,424
MD	41.3%	$52,331
DE	39.2%	$50,869
OR	53.7%	$50,790
OH	40.8%	$48,692
MN	34.9%	$46,906
IN	47.3%	$46,851
GA	29.3%	$46,526
WA	37.9%	$45,712
VA	29.4%	$44,763
VT	53.9%	$44,535
HI	32.7%	$44,273
CO	30.4%	$44,161
NH	34.5%	$43,941
WI	52.1%	$43,466
NV	35.7%	$43,394
NC	25.0%	$43,313
AZ	28.3%	$42,905
SC	35.8%	$42,207
ID	35.3%	$42,122
TN	32.3%	$41,527
FL	65.2%	$41,081
TX	19.0%	$41,009
KY	40.8%	$41,002
ME	42.2%	$40,940
AK	35.8%	$40,495
WY	41.3%	$40,392
IA	41.5%	$40,347
UT	29.5%	$39,965
NE	40.8%	$39,456
NM	26.6%	$39,328
KS	48.2%	$39,190
MO	29.9%	$38,971
LA	12.2%	$38,880
AL	28.4%	$38,863
MT	43.1%	$38,485
WV	36.9%	$38,360
OK	25.3%	$37,141
MS	31.1%	$36,590
ND	31.6%	$36,449
SD	32.2%	$34,040

★ San Francisco metro area: $46.70/hr
★ New York metro area: $45.79/hr
★ Detroit metro area: $47.28/hr
★ Greensboro metro area: $47.28/hr
★ Orlando metro area: $25.03/hr

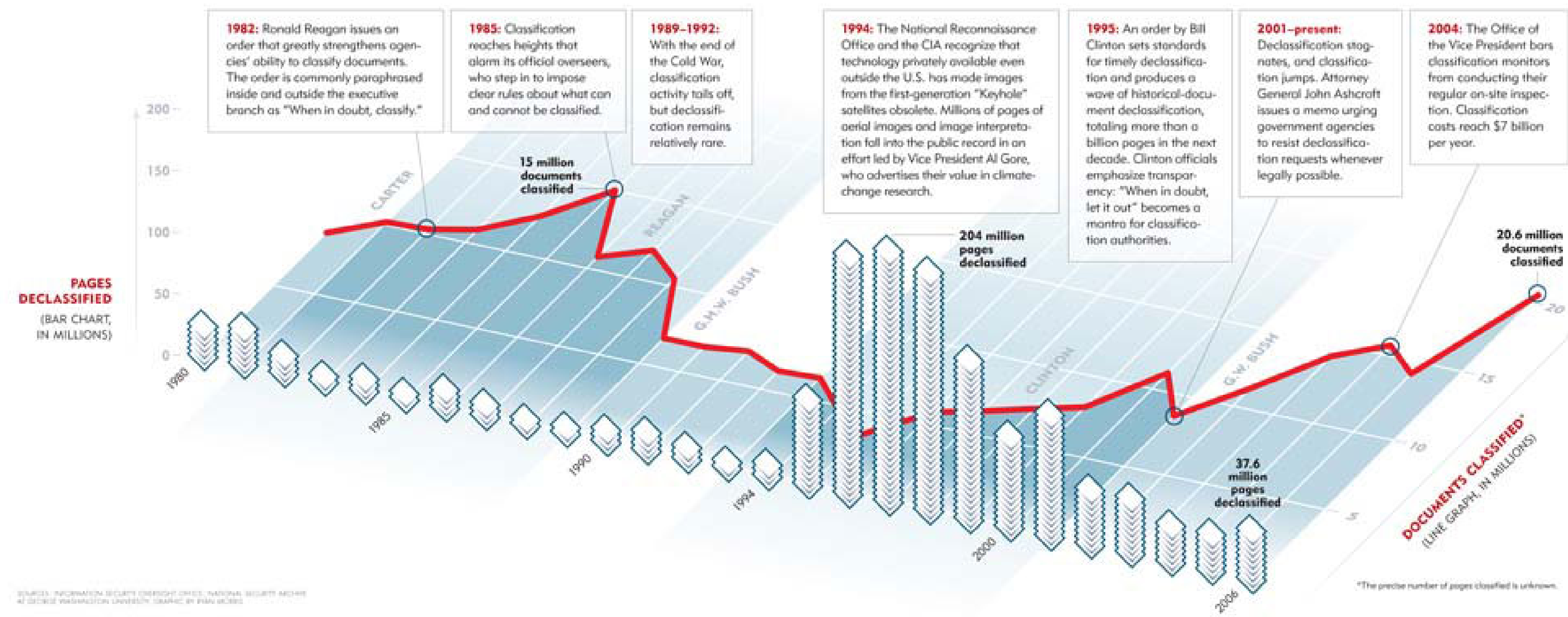
SOURCES Manhattan Institute; National Center For Education Statistics; National Education Association; U.S. Bureau of Labor Statistics

AVERAGE Workers' salaries used for comparison are those of white-collar, nonsales employees.

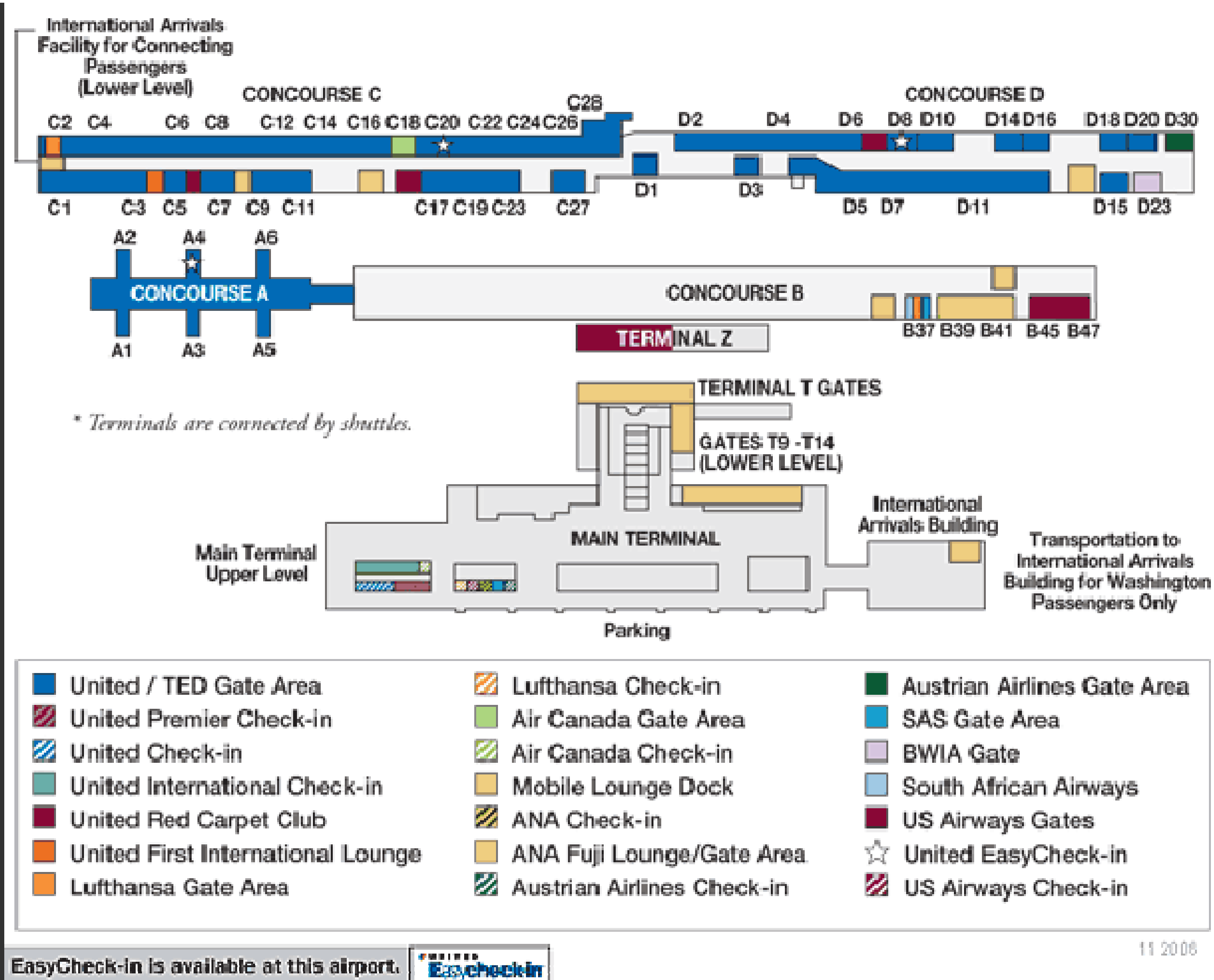
28 GOOD May/Jun 07 Transparency

29 goodmagazine.com Transparency

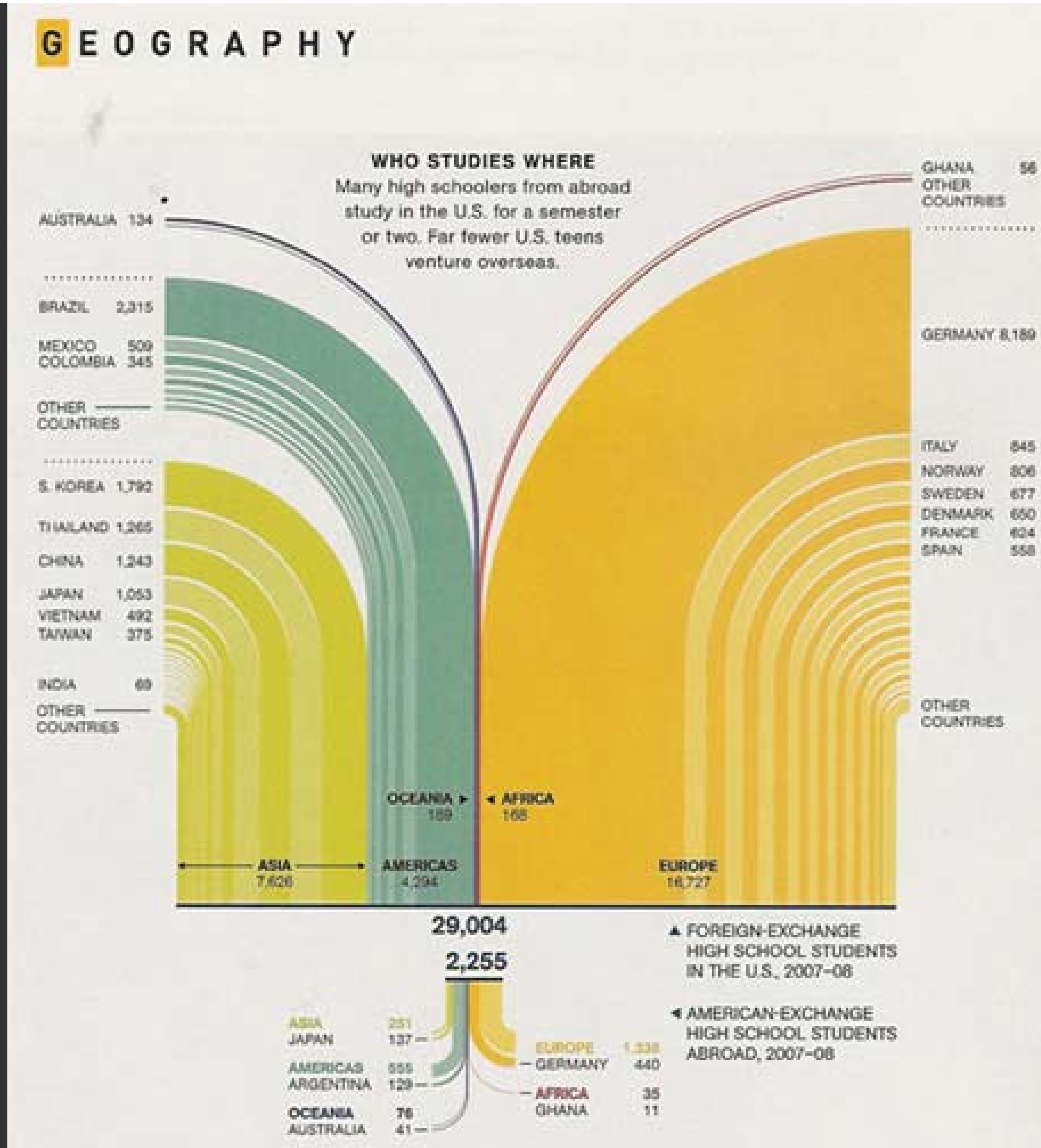
Source: *Good* Magazine



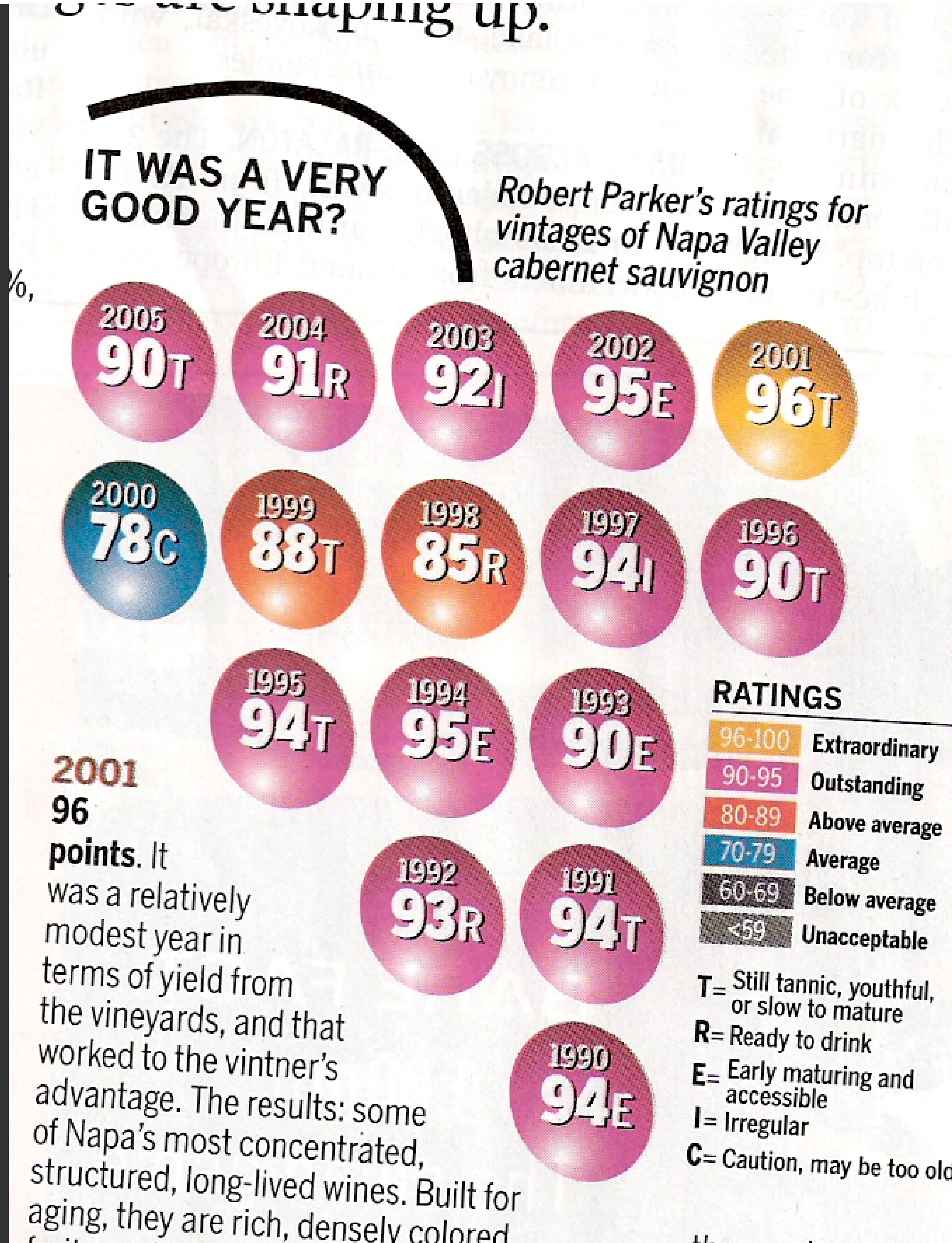
Source: ***The Atlantic*** **300 no. 2 (September 2007)**
Number of Classified U.S. Documents



Washington Dulles Airport Map
Source: United Airlines *Hemispheres*



Source: *National Geographic*, September, 2008, p. 22. Silver, Mark. "High School Give-and-Take."

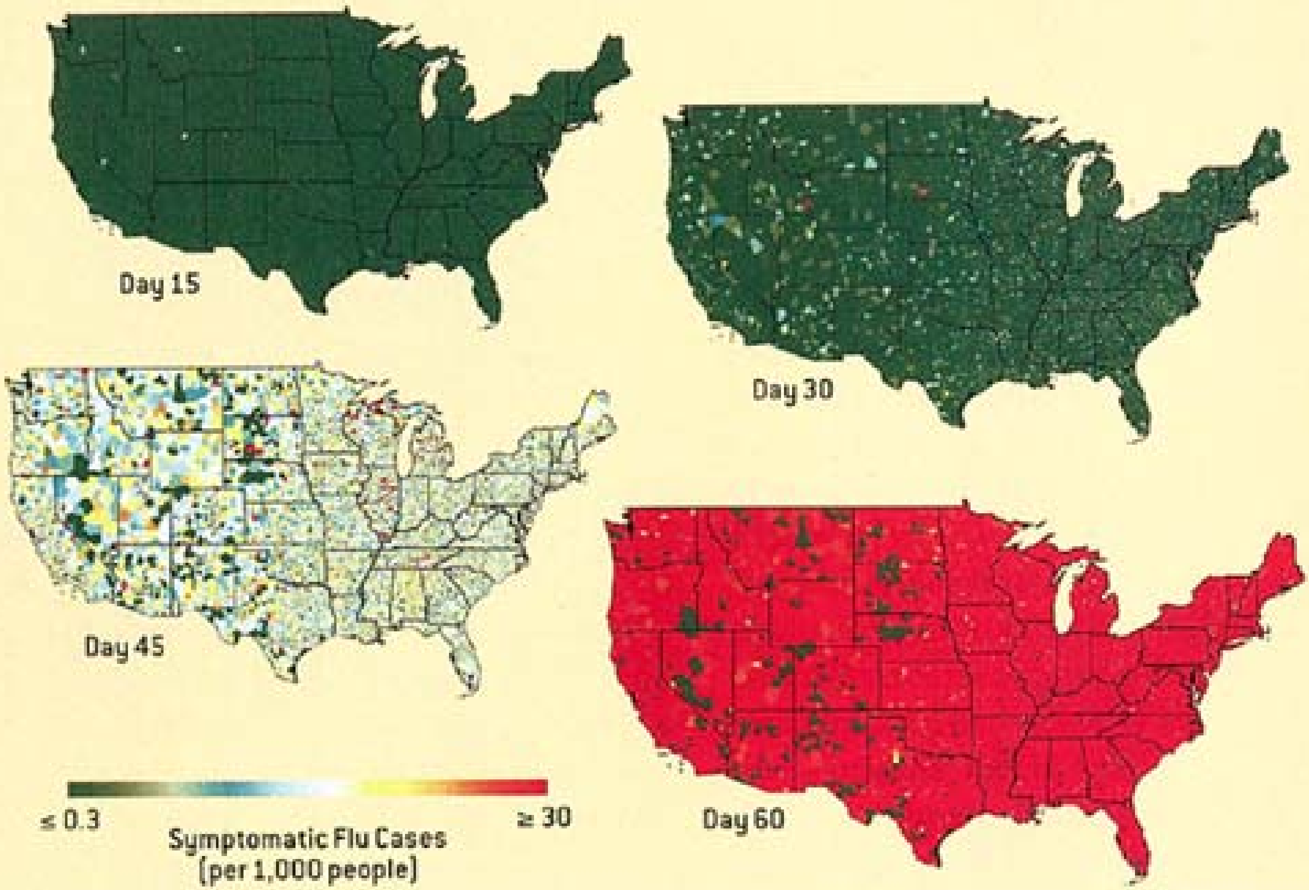


Source: *Business* Week, June 18, 2007

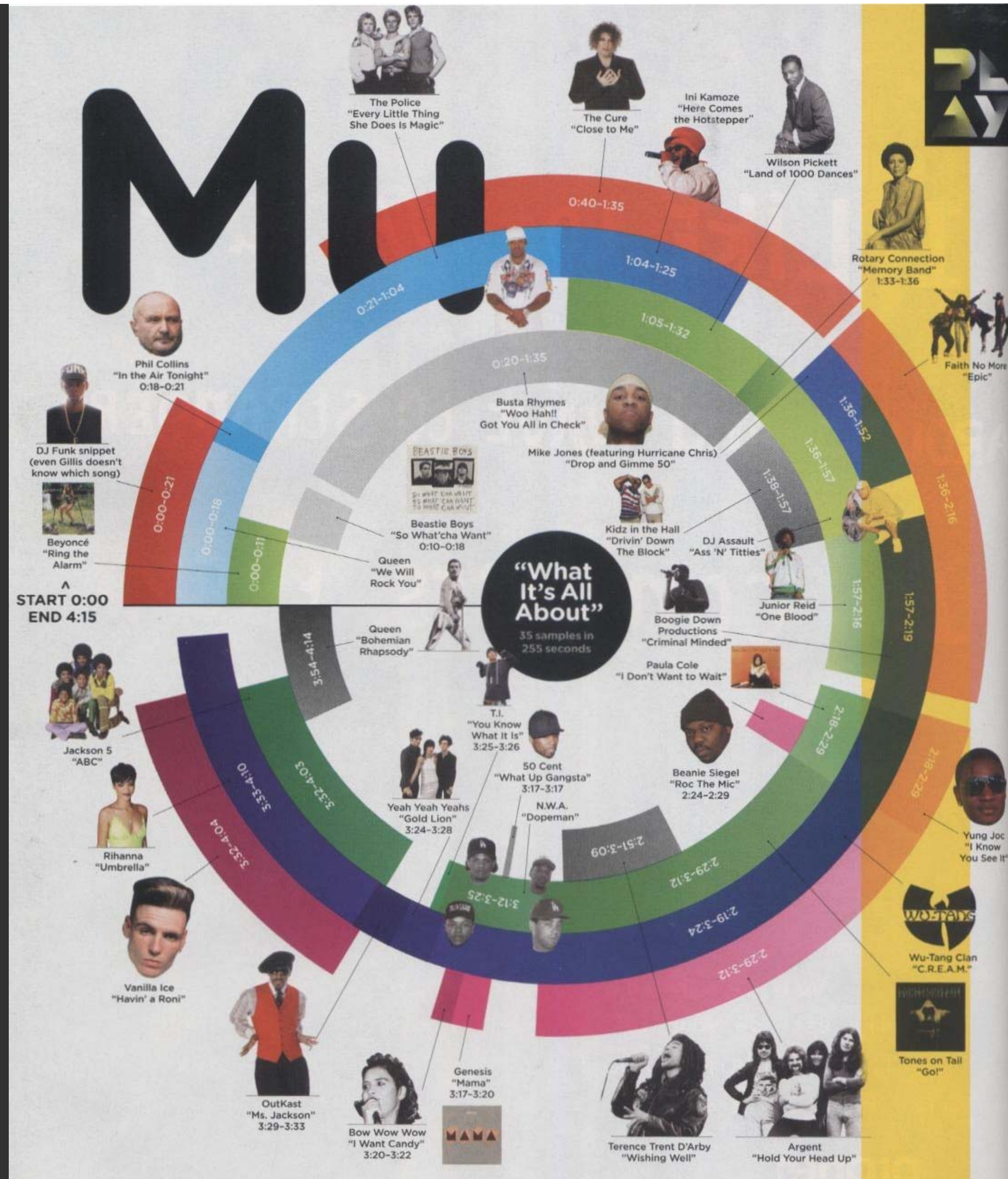
Source: *India Today*

Pandemic Flu Hits the U.S.

A simulation created by researchers from Los Alamos National Laboratory and Emory University shows the first wave of a pandemic spreading rapidly with no vaccine or antiviral drugs employed to slow it down. Colors represent the number of symptomatic flu cases per 1,000 people (*see scale*). Starting with 40 infected people on the first day, nationwide cases peak around day 60, and the wave subsides after four months with 33 percent of the population having become sick. The scientists are also modeling potential interventions with drugs and vaccines to learn if travel restrictions, quarantines and other disruptive disease-control strategies could be avoided.



Preparing for a Pandemic
Source: *Scientific American*, 293(5). November, 2005, p. 50



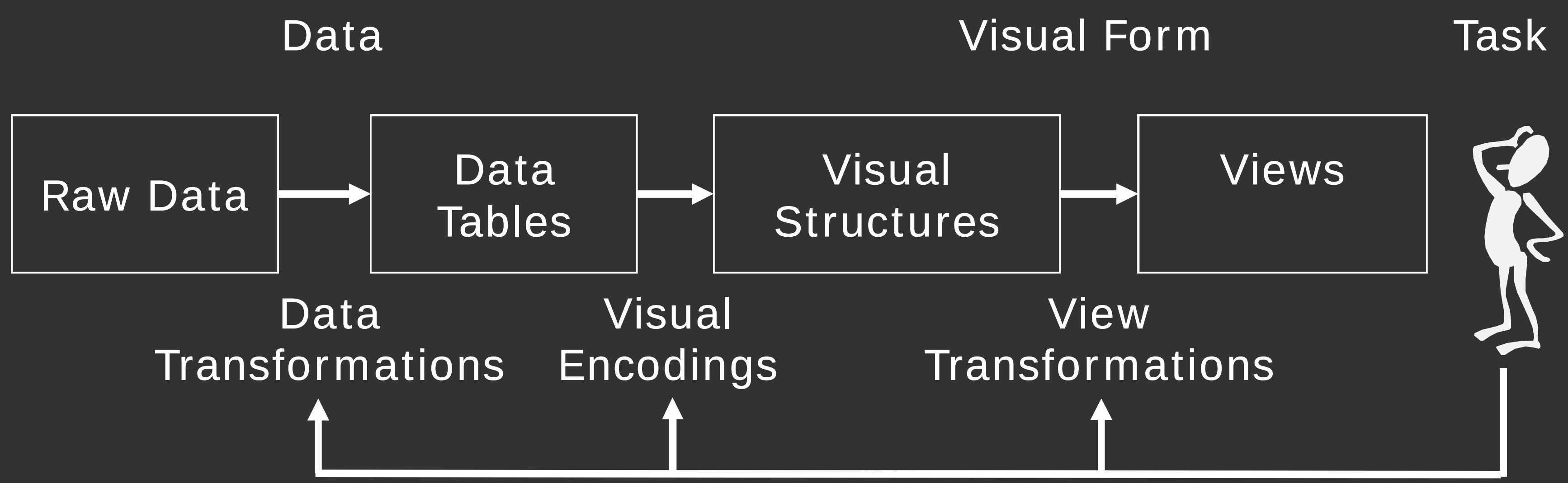
Source: ***Wired Magazine*****, September 2008 Edition**
Music: Super Cuts (page 92)

Re-Design Presentation (~4 min)

1. Describe the data and visualization
2. Present your critique
3. Share your re-design

Visual Analysis Software

Visualization Reference Model



Interaction Techniques

Dynamic Queries

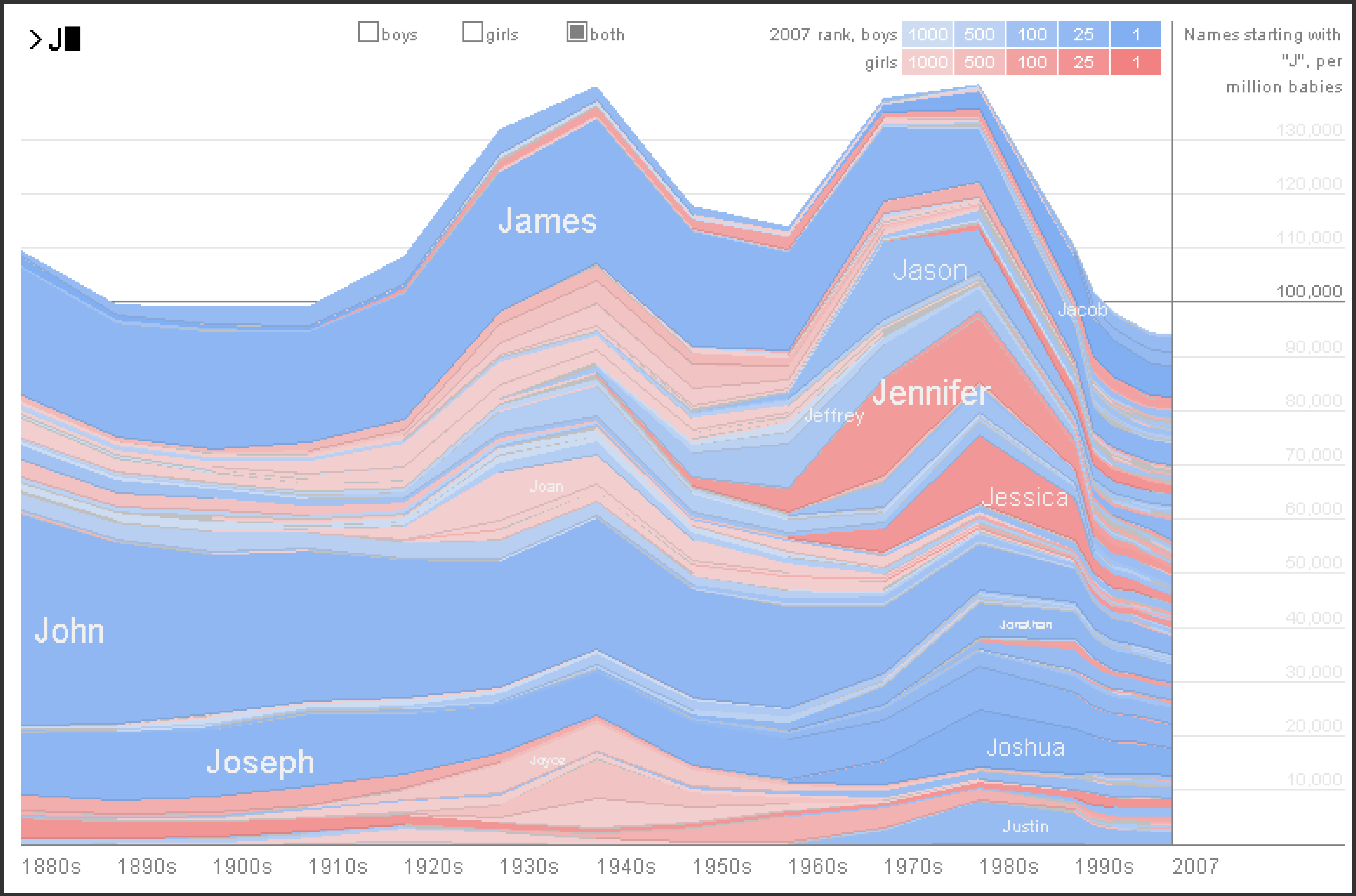
Filter a visualization through direct, reversable actions that avoid complex syntax.

Brushing and Linking

Highlight relationships between related items across multiple visualization views.

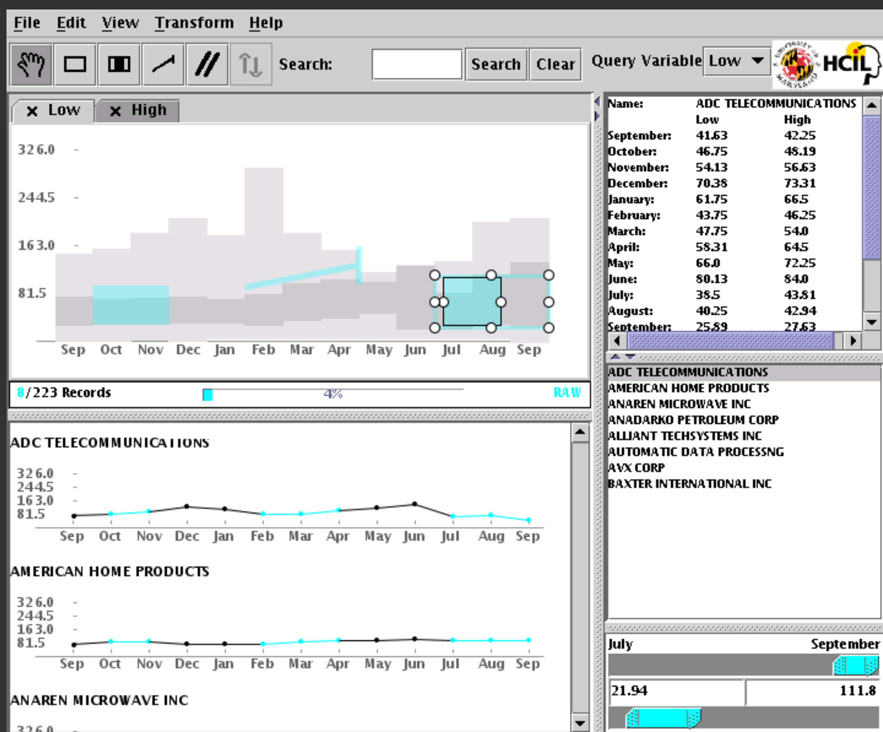
Time-Series Data

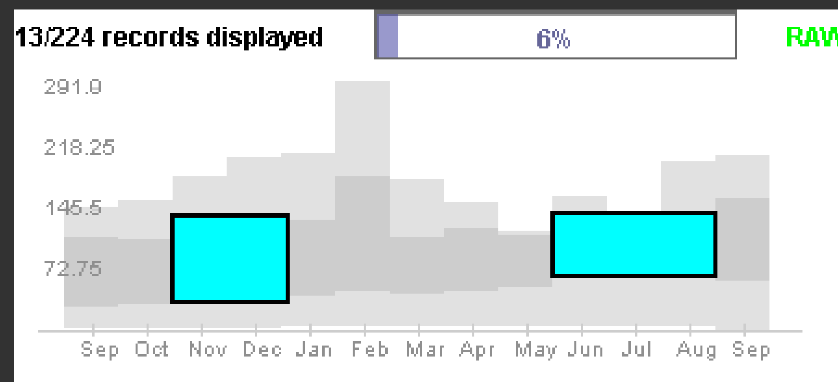
NameVoyager

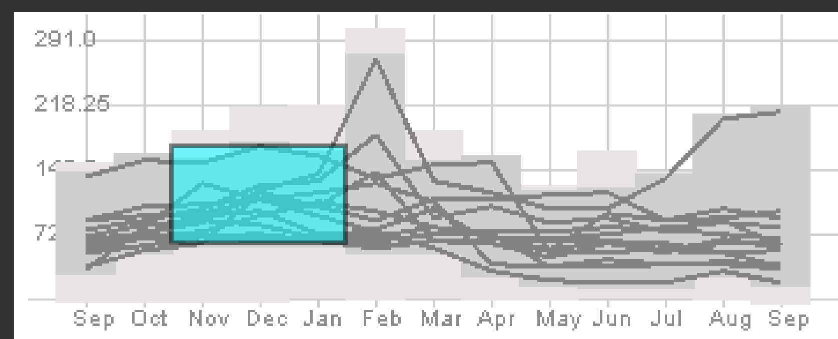


http://www.babynamewizard.com/voyager

TimeSearcher [Hochheiser & Shneiderman 02]







Based on Wattenberg's [2001] idea for sketch-based queries of time-series data.

Multivariate Data

Baseball Statistics [from Wills 95]

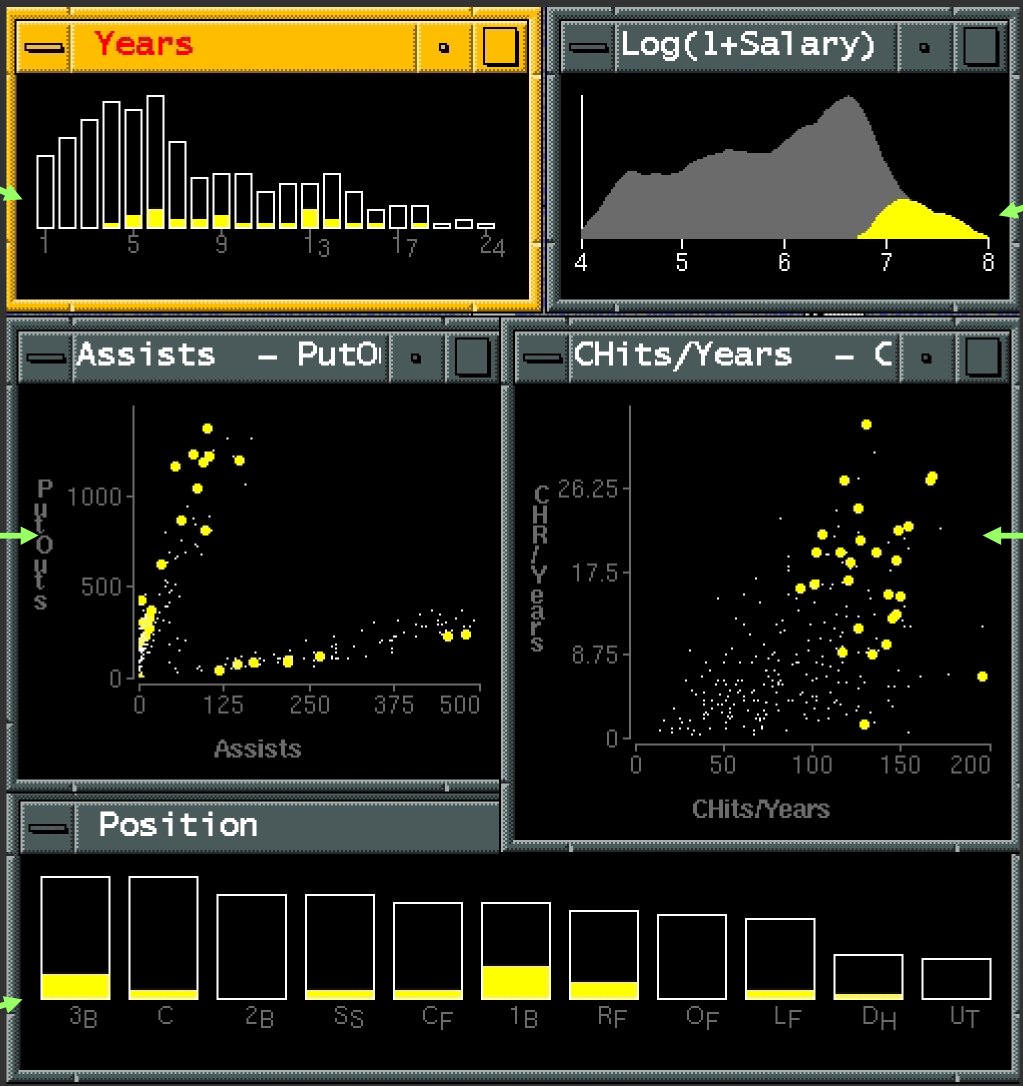
how long in majors

select high salaries

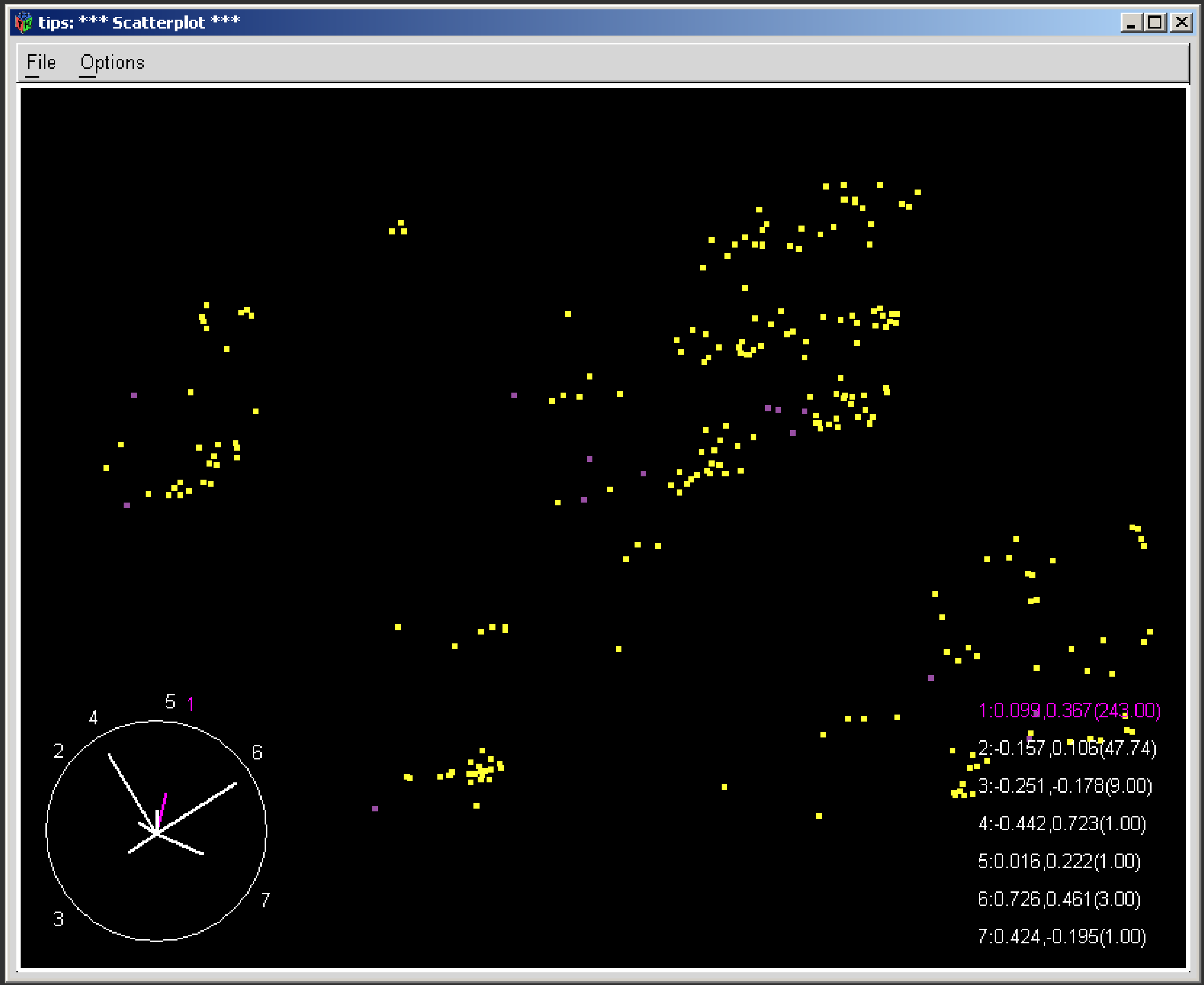
avg assists vs avg putouts (fielding ability)

avg career HRs vs avg career hits (batting ability)

distribution of positions played



GGobi: Projections of nD data



http://www.ggobi.org/

Parallel Coordinates

Parallel Coordinates [Inselberg]

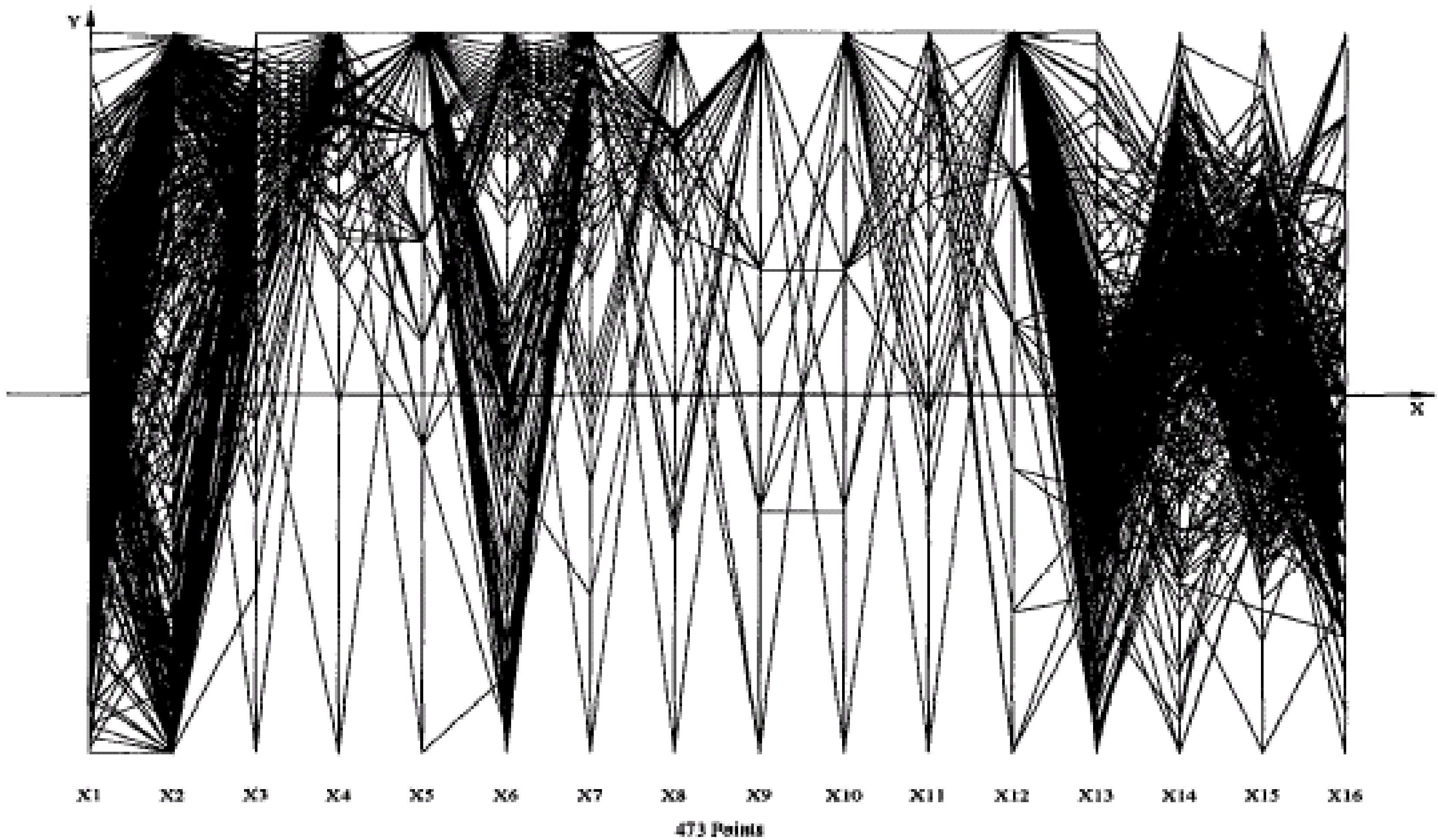


Figure 1: The full dataset consisting of 473 batches

The Multidimensional Detective

The Dataset:

- Production data for 473 batches of a VLSI chip
- 16 process parameters:

  X1: The yield: % of produced chips that are useful

  X2: The quality of the produced chips (speed)

  X3 … X12: 10 types of defects (zero defects shown at top)

  X13 … X16: 4 physical parameters

The Objective:

Raise the yield (X1) and maintain high quality (X2)

A. Inselberg, Multidimensional Detective, Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), 1997

Parallel Coordinates

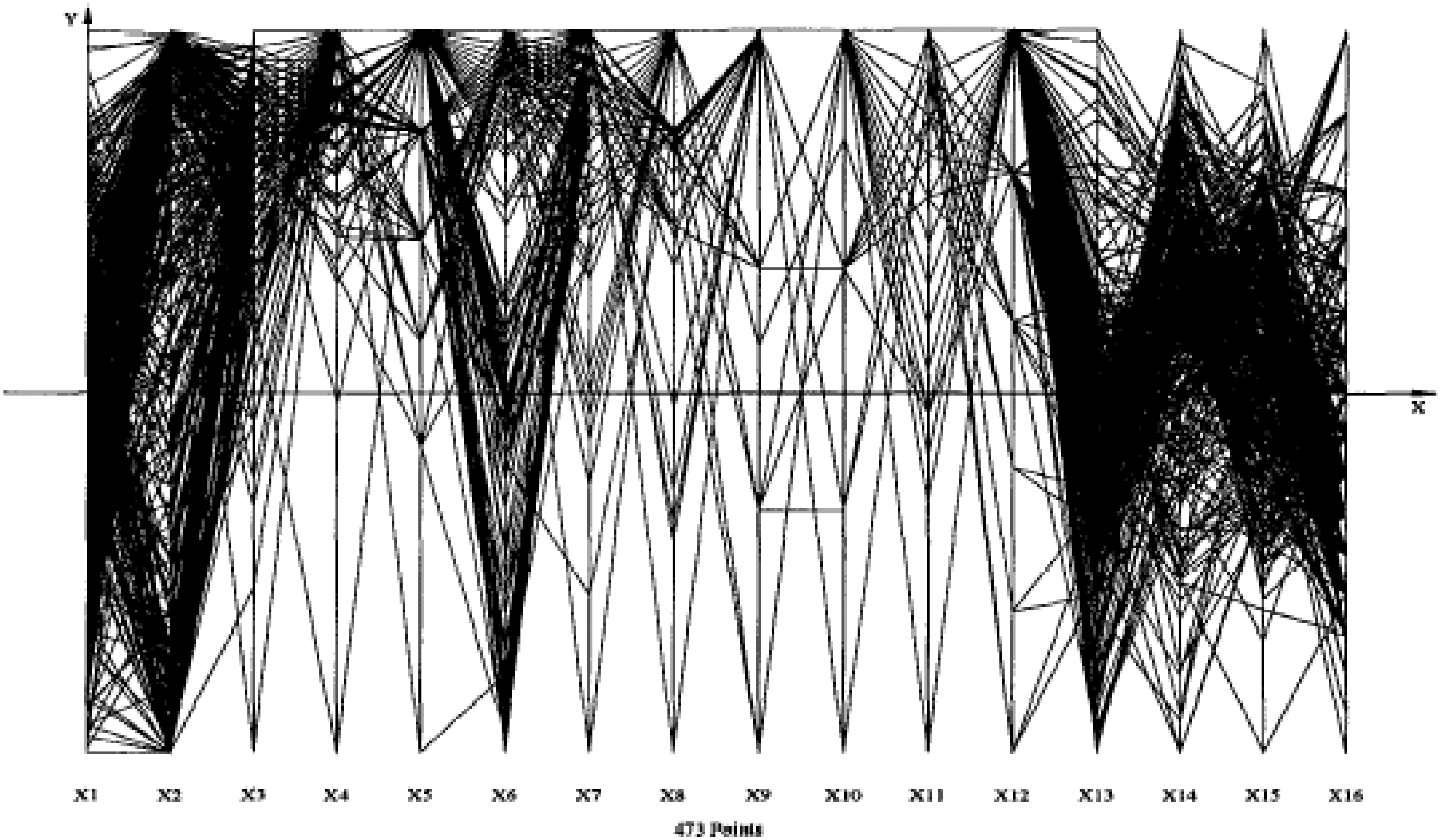


Figure 1: The full dataset consisting of 473 batches

Inselberg's Principles

1. Do not let the picture scare you
2. Understand your objectives
   - Use them to obtain visual cues
3. Carefully scrutinize the picture
4. Test your assumptions, especially the "I am really sure of's"
5. You can't be unlucky all the time!

Each line represents a tuple (e.g., VLSI batch)

Filtered below for high values of X1 and X2

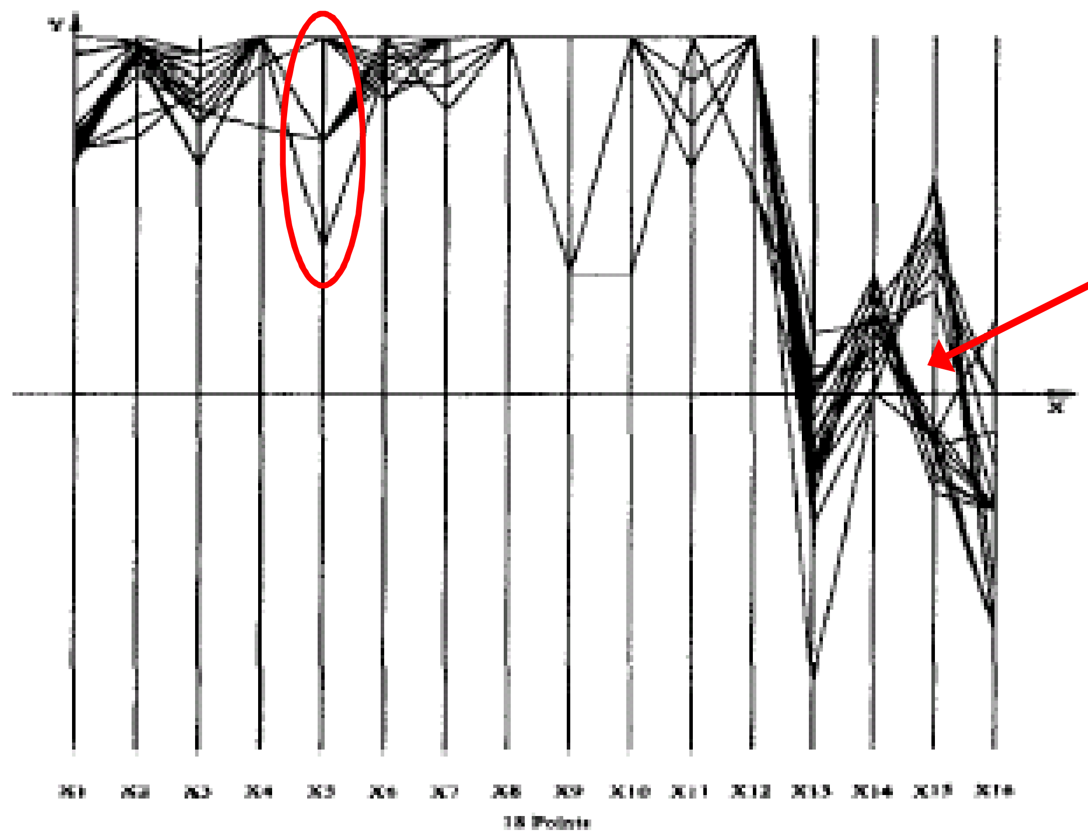
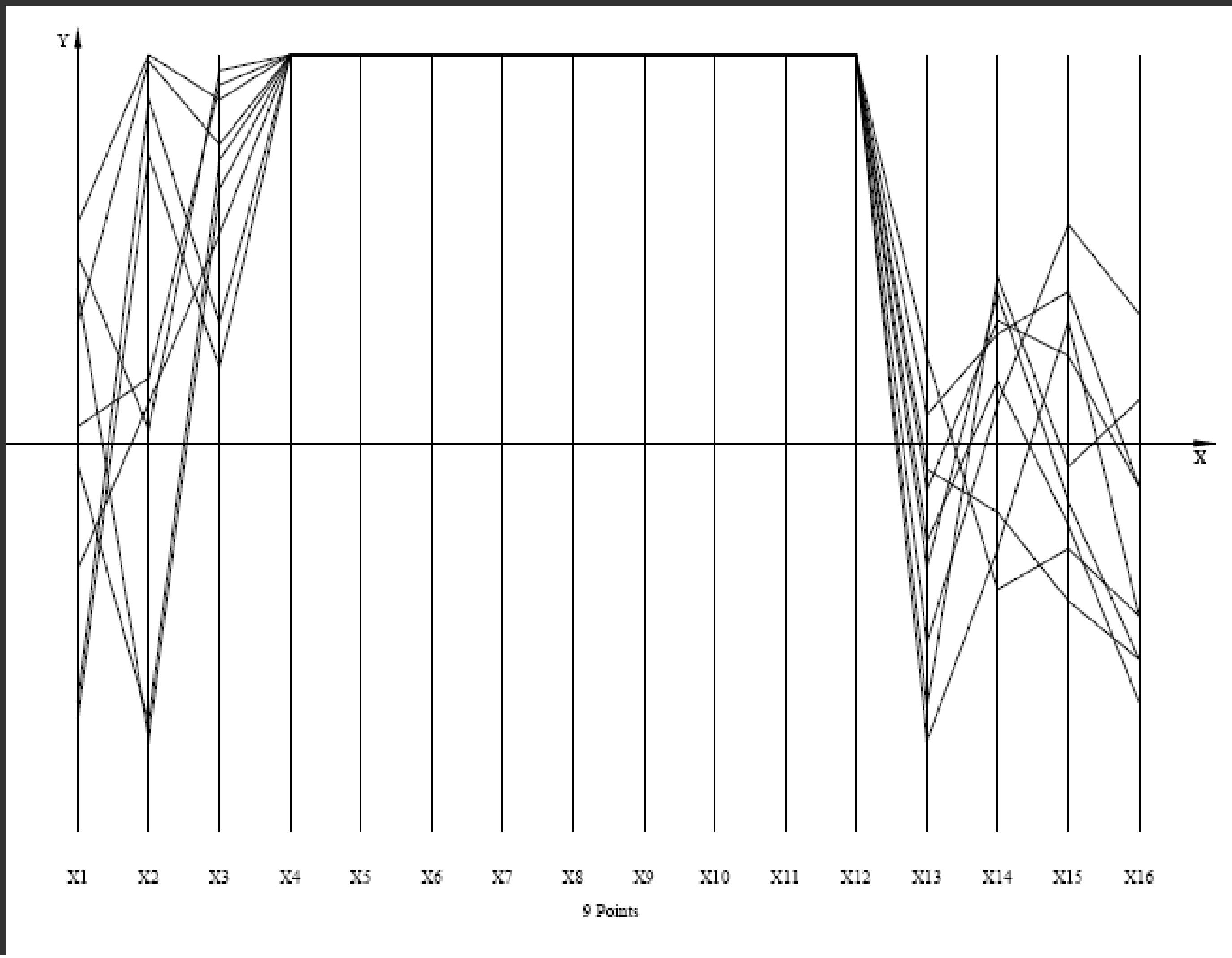


Figure 2: The batches high in Yield, $X1$, and Quality, $X2$.

Look for batches with *nearly* zero defects (9/10)

Most of these have low yields → defects OK.



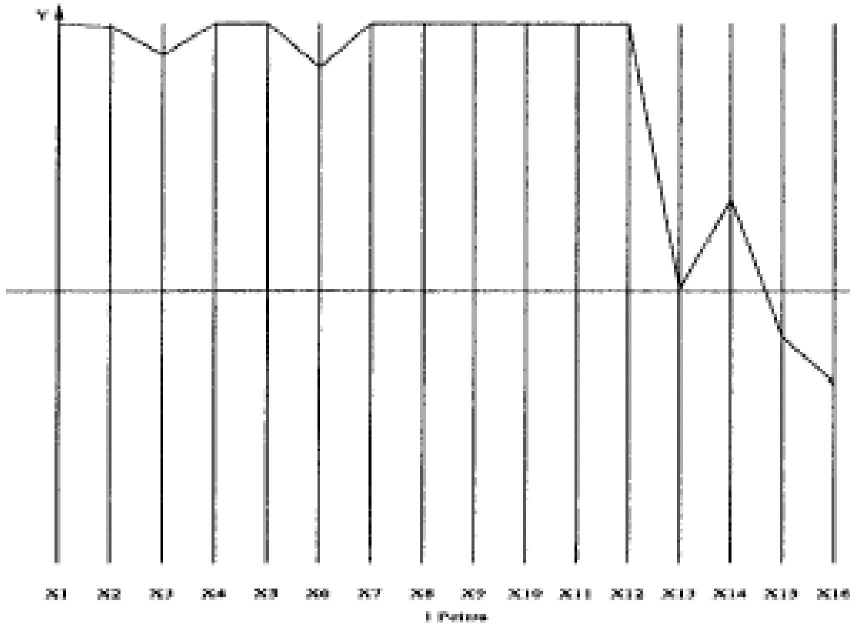


Figure 5: The best batch. Highest in Yield, $X1$, and very high in Quality, $X2$.

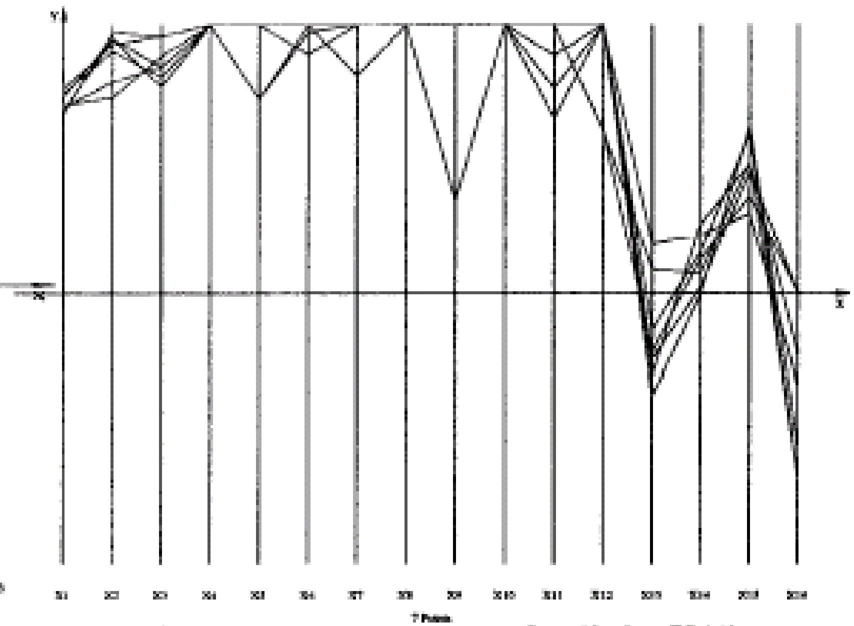
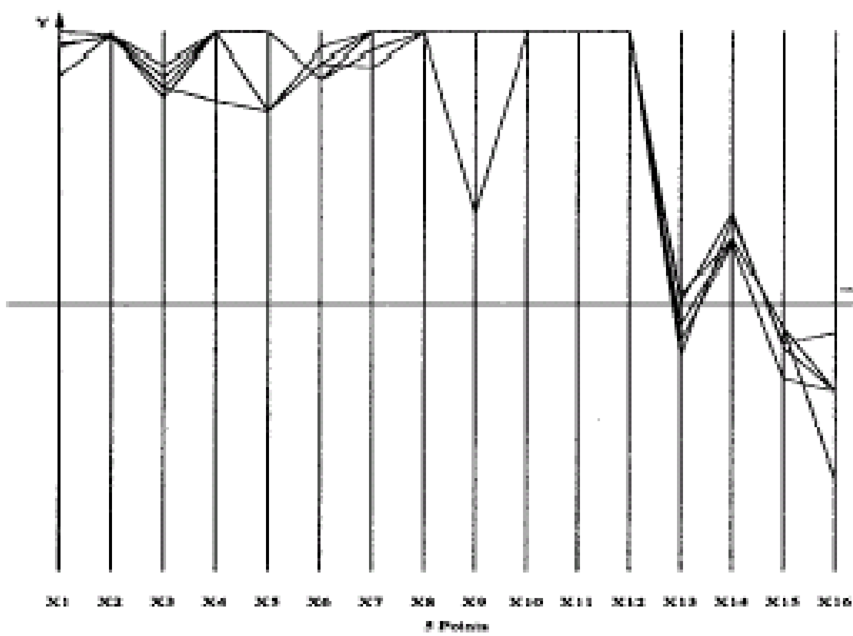
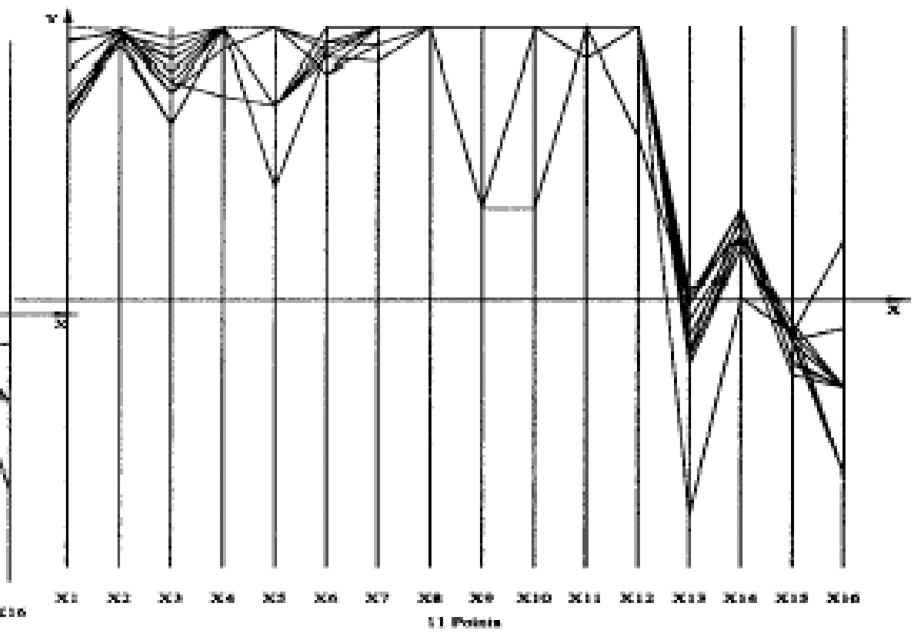


Figure 7: Upper range of split in $X15$





Notice that X6 behaves differently.

Allow 2 defects, including X6 → best batches

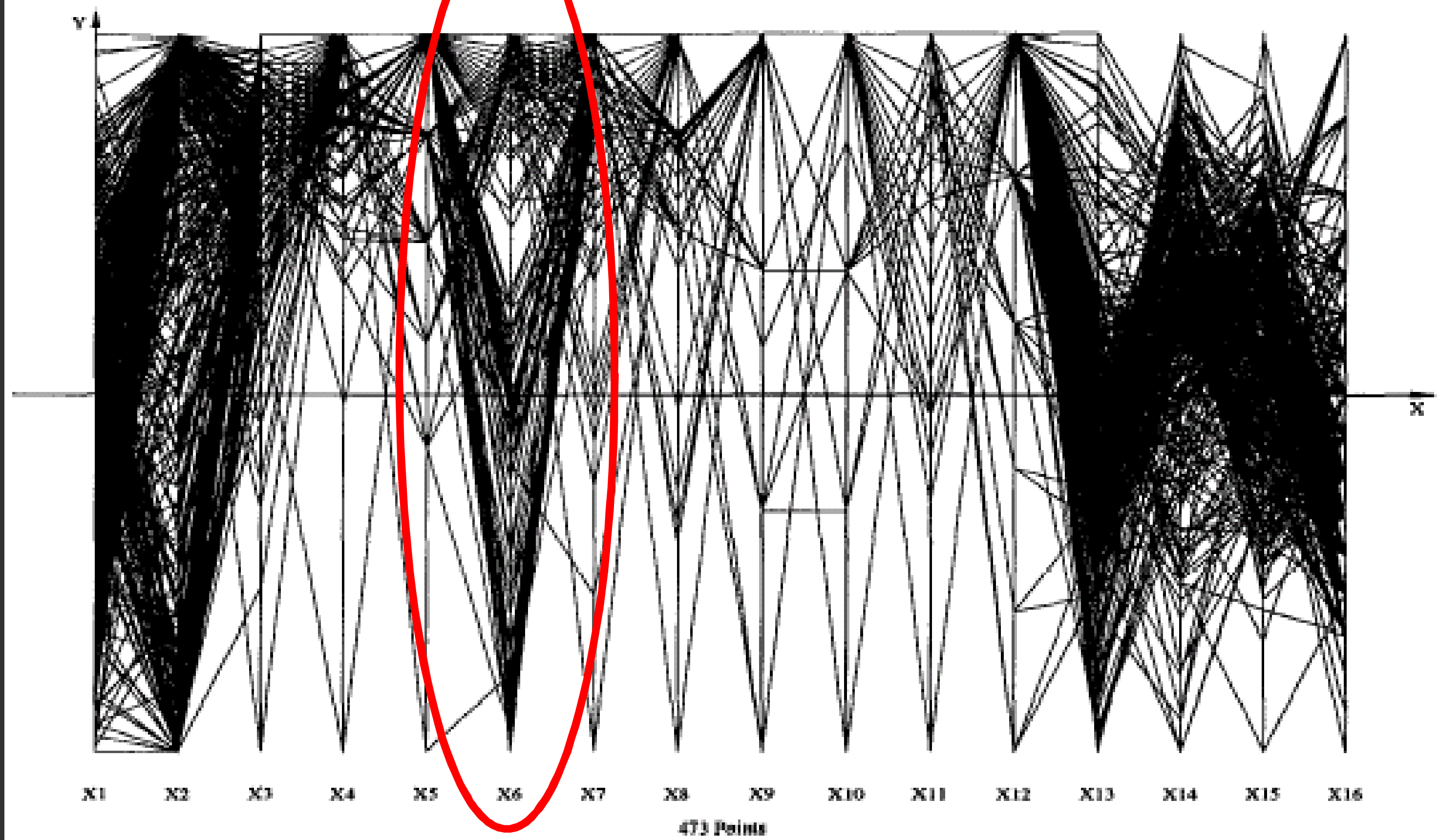


Figure 1: The full dataset consisting of 473 batches

Parallel Coordinates

Free implementation: Parvis by Ledermen

- http://home.subnet.at/flo/mv/parvis/

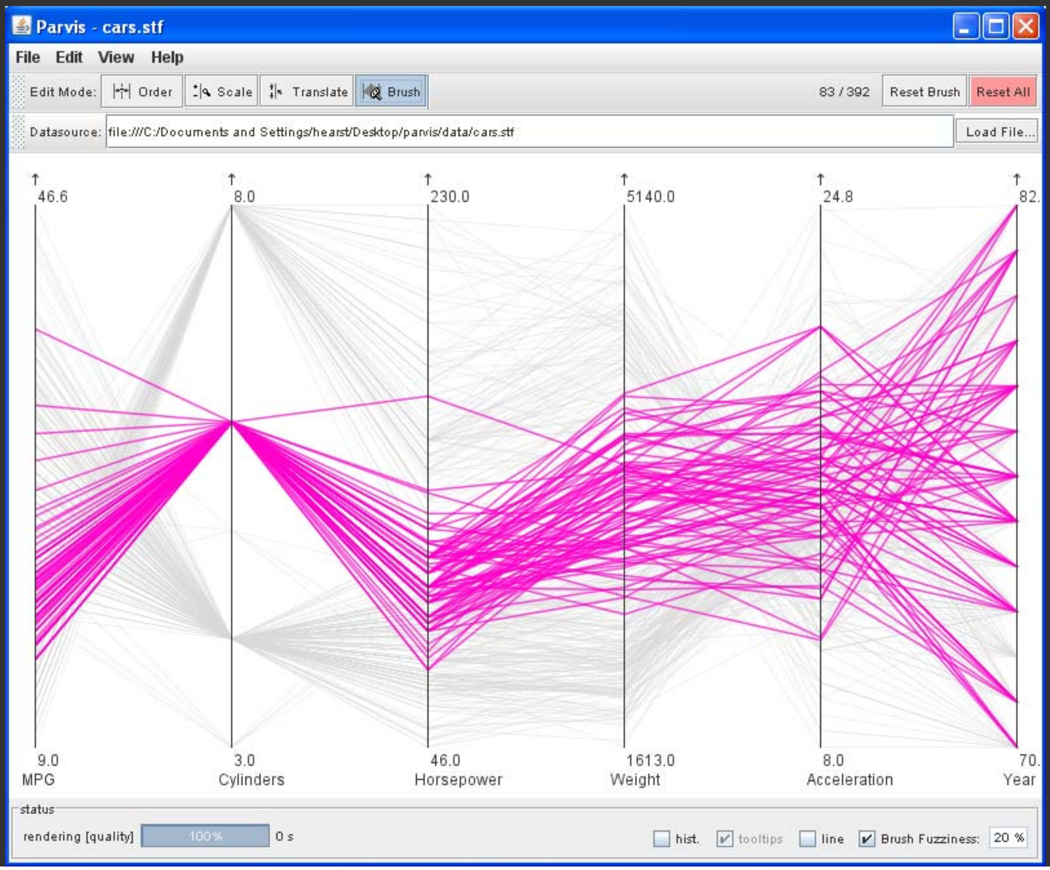
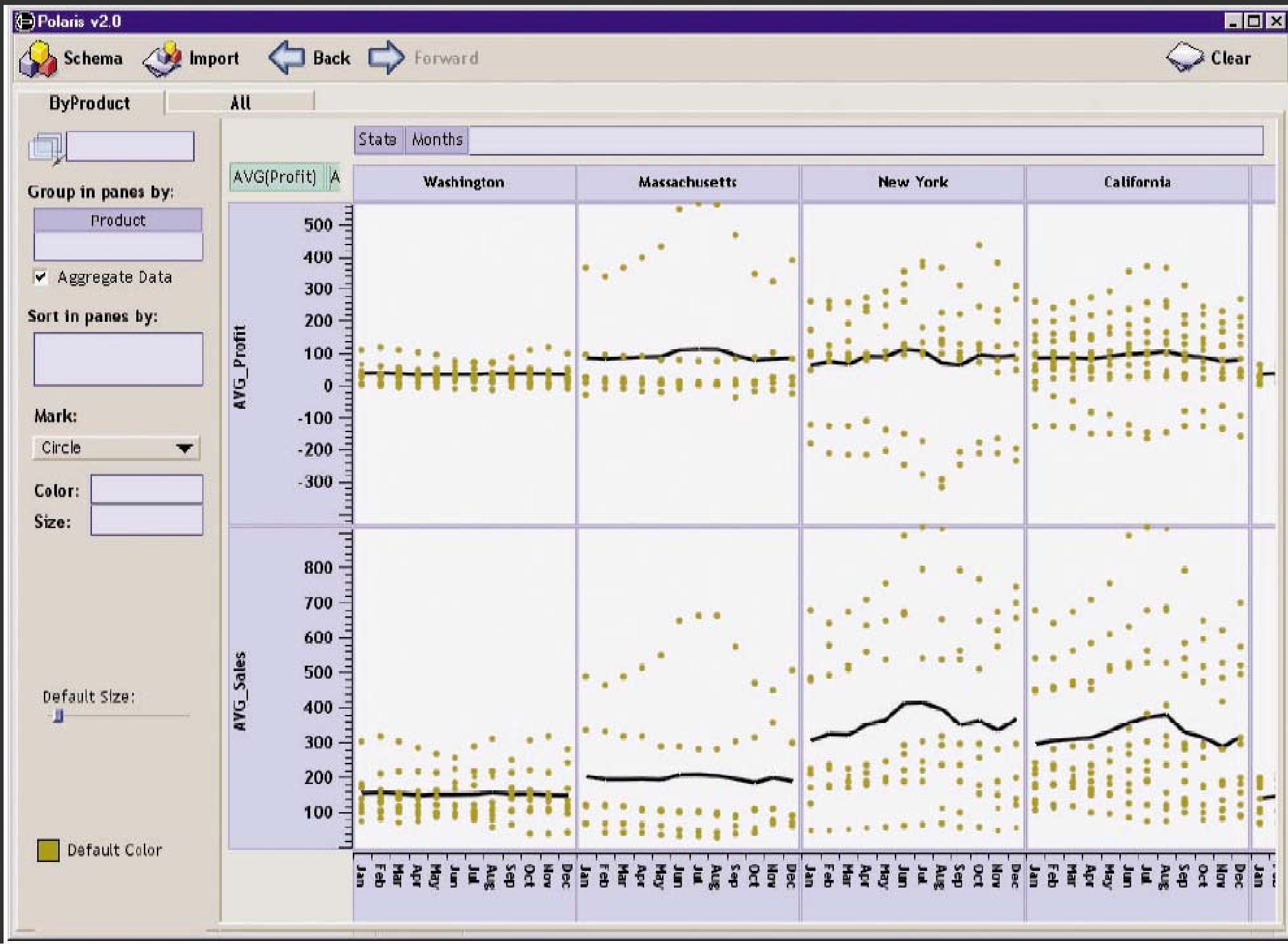


Tableau / Polaris

Polaris

Research at Stanford by Stolte, Tang, and Hanrahan.



Tableau

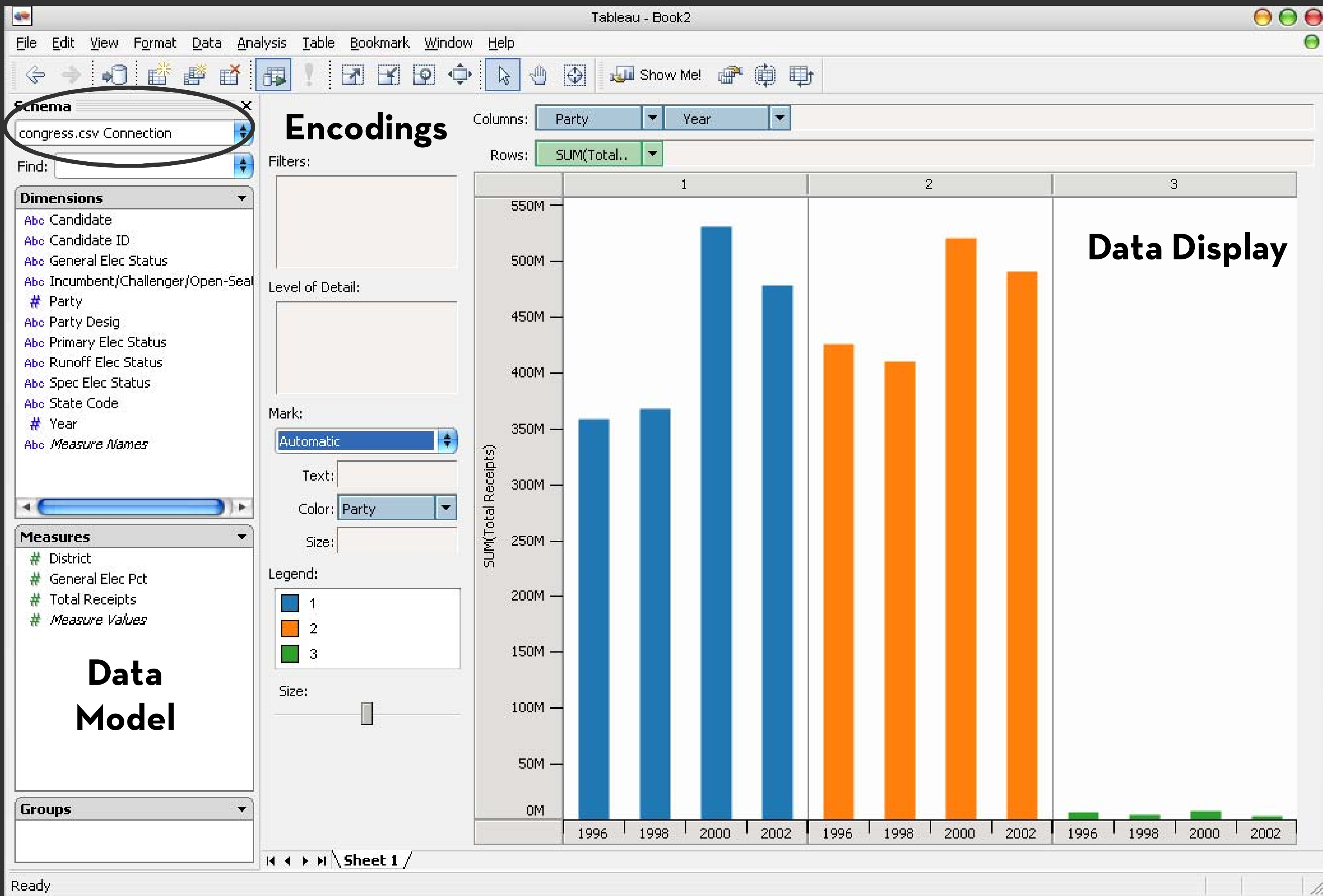


Tableau - Book2
File Edit View Format Data Analysis Table Bookmark Window Help
Show Me!
Schema
congress.csv Connection
Find:
Dimensions
Candidate
Candidate ID
General Elec Status
Incumbent/Challenger/Open-Seat
Party
Party Desig
Primary Elec Status
Runoff Elec Status
Spec Elec Status
State Code
Year
Measure Names
Measures
District
General Elec Pct
Total Receipts
Measure Values
Data Model
Groups
Encodings
Filters:
Level of Detail:
Mark:
Automatic
Text:
Color: Party
Size:
Legend:
1
2
3
Size:
Columns: Party Year
Rows: SUM(Total..
Data Display
SUM(Total Receipts)
550M
500M
450M
400M
350M
300M
250M
200M
150M
100M
50M
0M
1996 1998 2000 2002
Sheet 1
Ready

Tableau Demo

The dataset:

Federal Elections Commission Receipts

Every Congressional Candidate from 1996 to 2002

4 Election Cycles

9216 Candidacies

Hypotheses?

What might we learn from this data?

- ??

Hypotheses?

What might we learn from this data?

Correlation between receipts and winners?

Do receipts increase over time?

Which states spend the most?

Which party spends the most?

Margin of victory vs. amount spent?

Amount spent between competitors?

Tableau Demo

Polaris/Tableau Approach

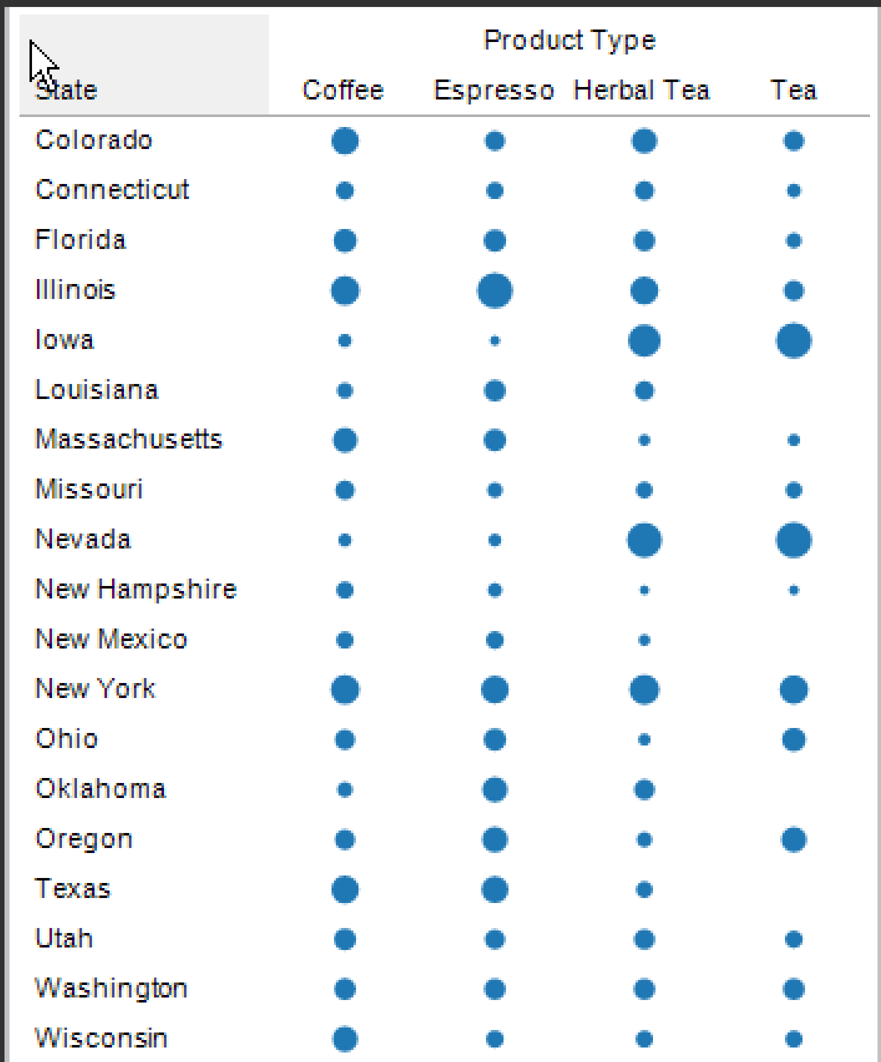
Insight: can simultaneously specify both database queries and visualization
(c.f., Leland Wilkinson's Grammar of Graphics)

Choose data, then visualization, **not vice versa**

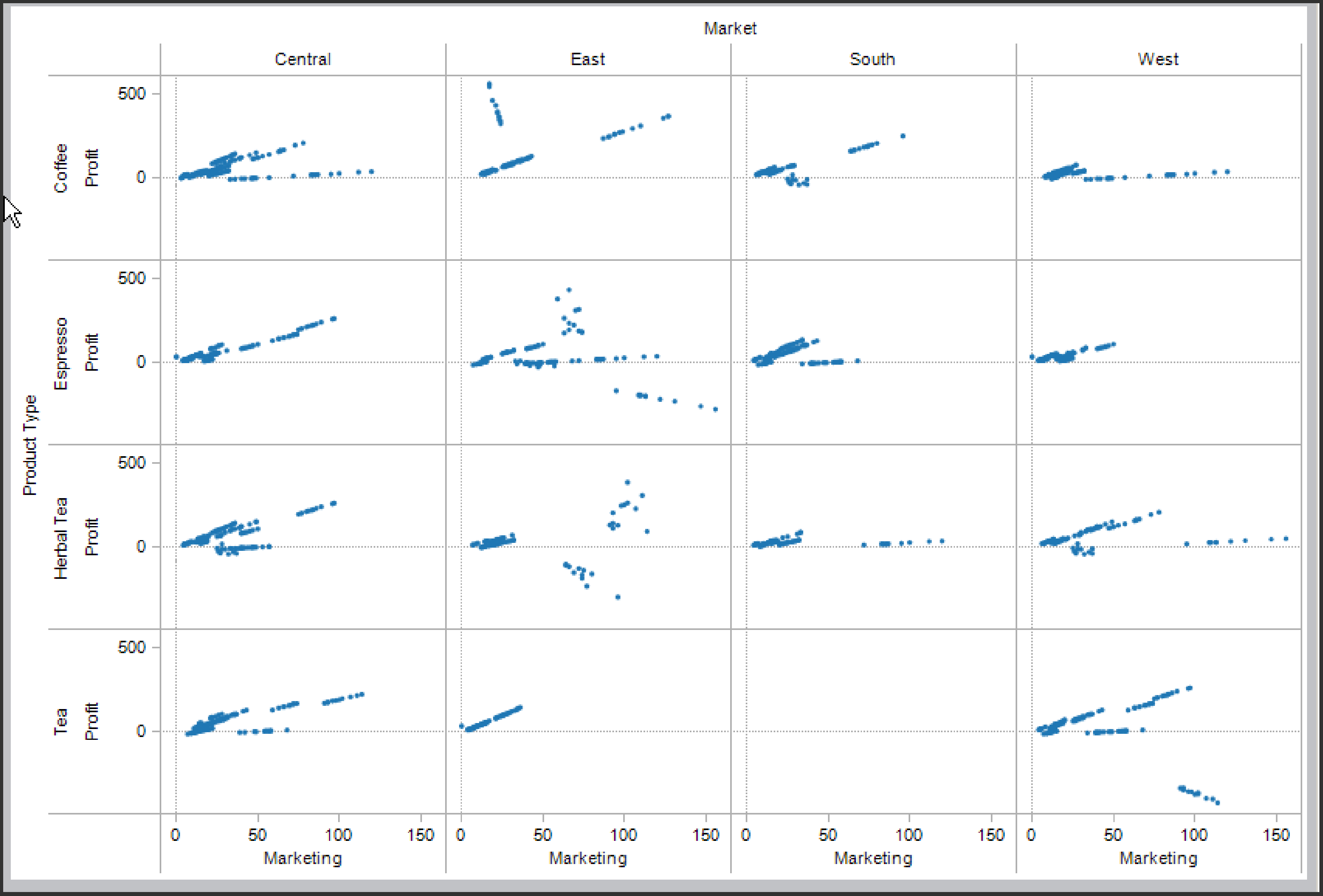
Use smart defaults for visual encodings

More recently: automate visualization design

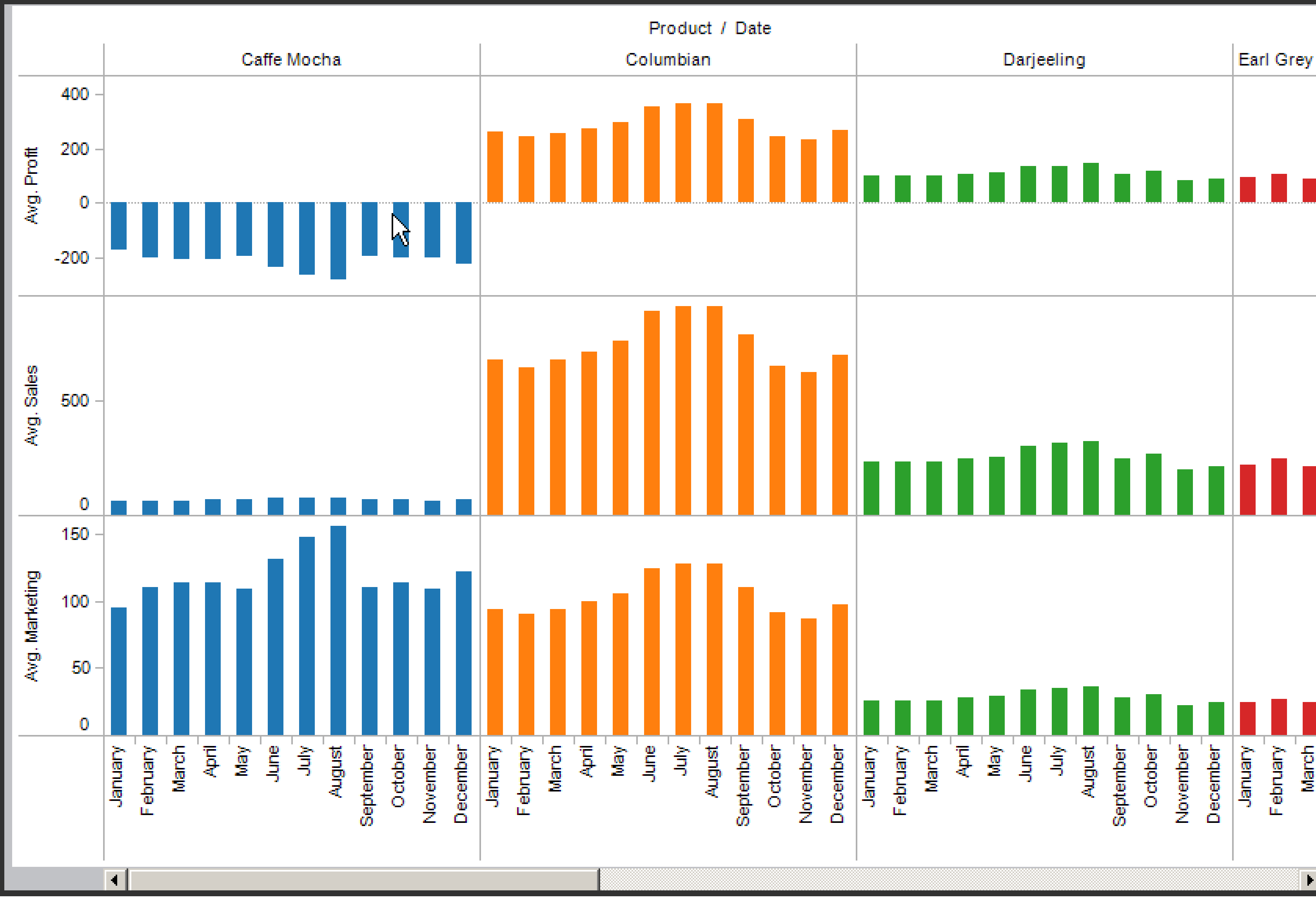
Ordinal - Ordinal



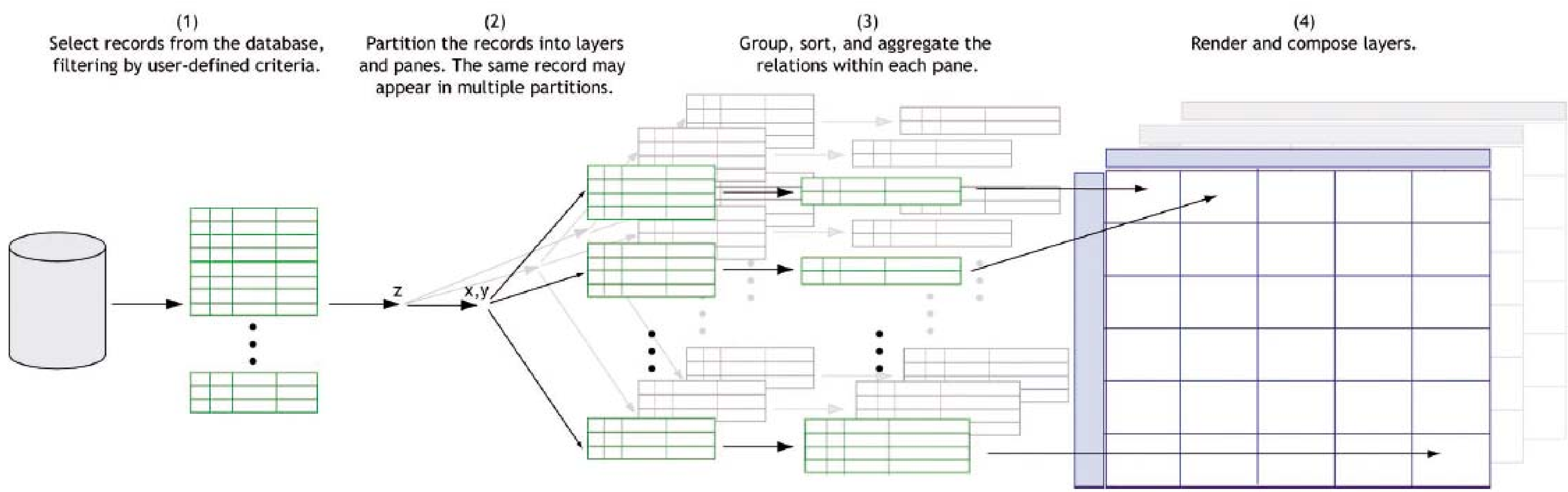
Quantitative - Quantitative

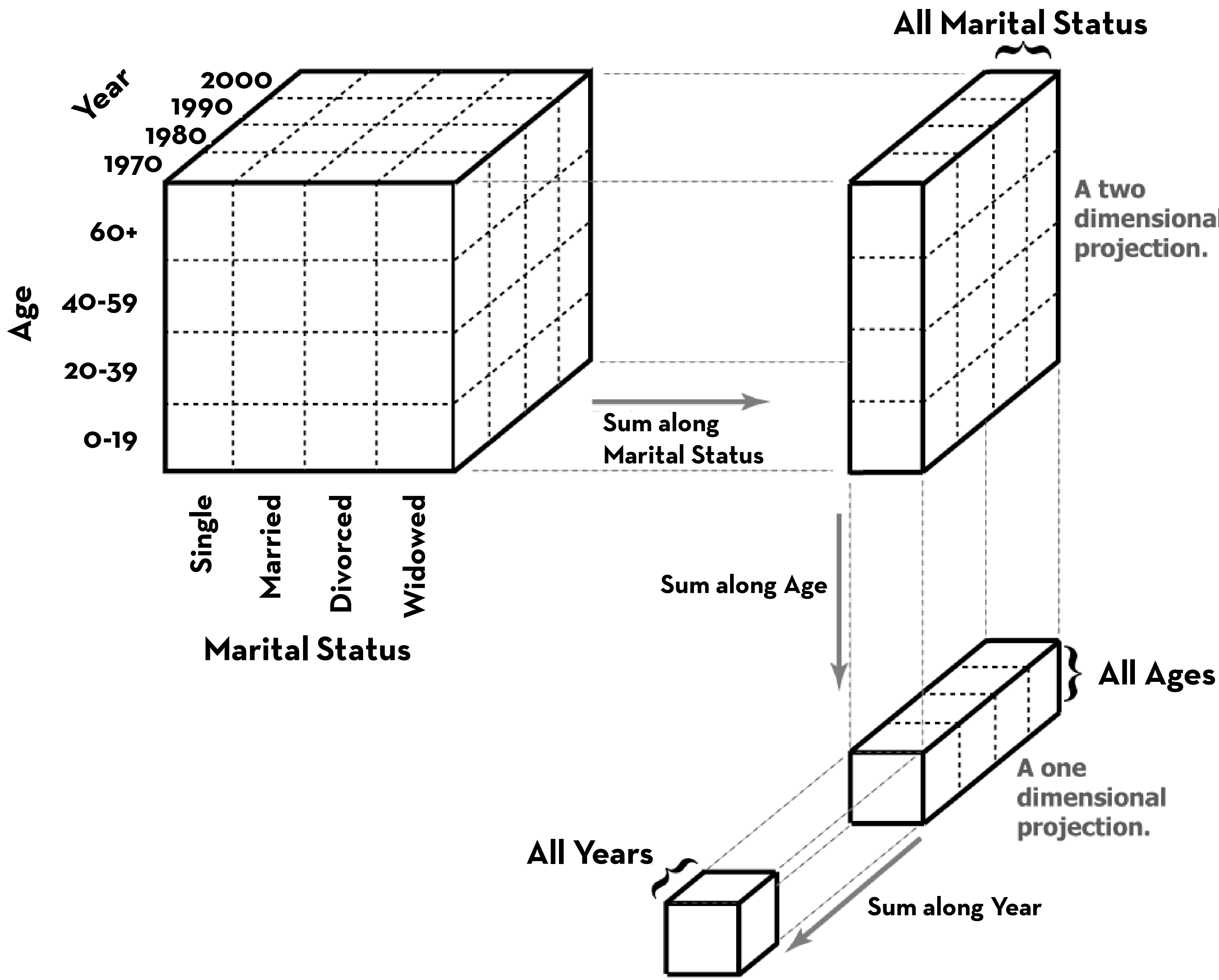


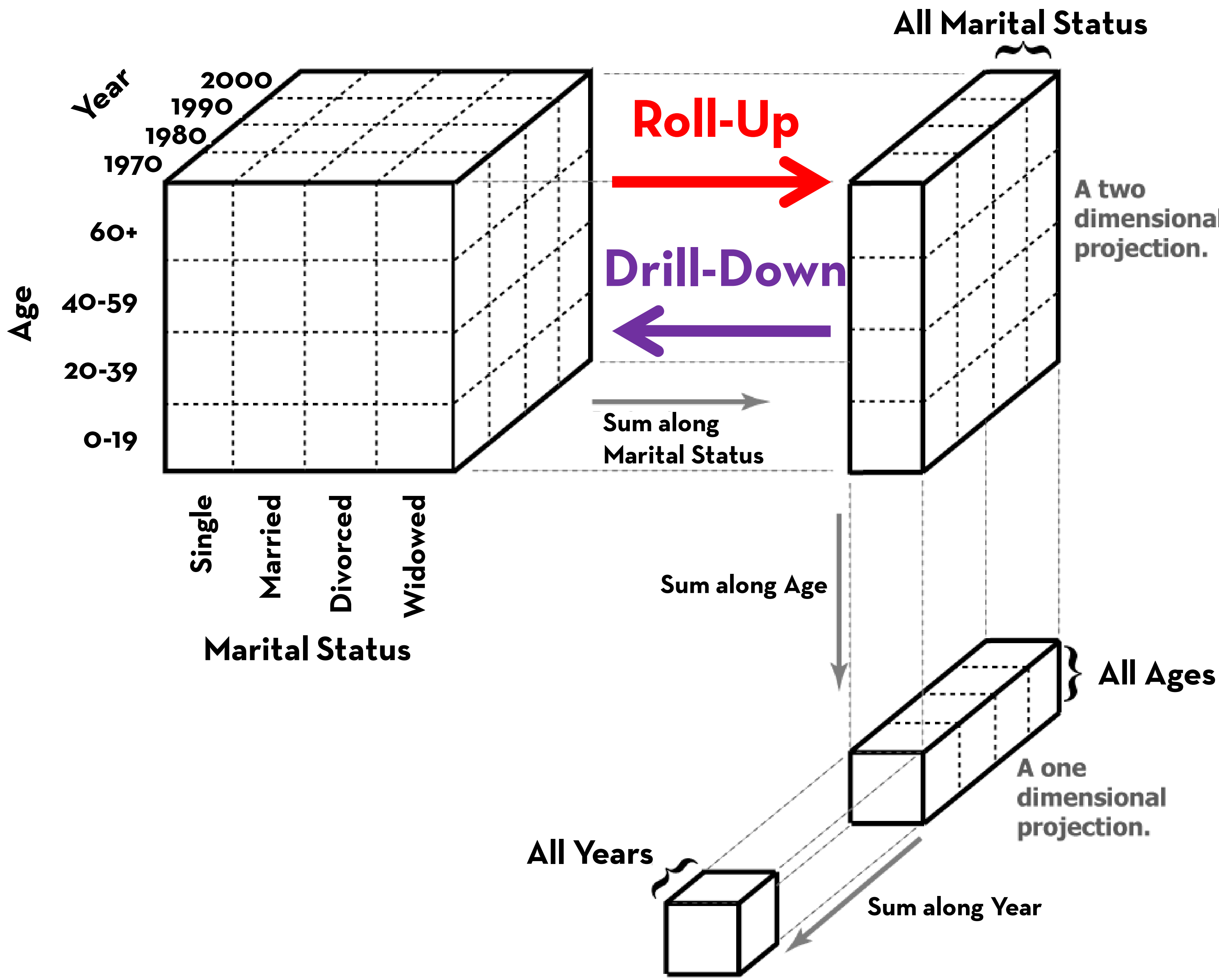
Ordinal - Quantitative



Querying the Database




All Marital Status
Year
2000
1990
1980
1970
Age
60+
40-59
20-39
0-19
A two dimensional projection.
Sum along Marital Status
Single
Married
Divorced
Widowed
Marital Status
Sum along Age
All Ages
A one dimensional projection.
All Years
Sum along Year



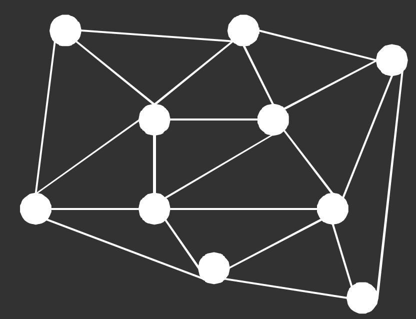
Year
2000
1990
1980
1970
Age
60+
40-59
20-39
0-19
Single
Married
Divorced
Widowed
Marital Status
Roll-Up
Drill-Down
Sum along Marital Status
All Marital Status
A two dimensional projection.
Sum along Age
All Ages
A one dimensional projection.
Sum along Year
All Years

Network Structures

Graphs and Trees

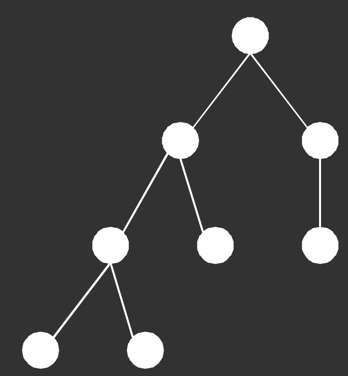
Graphs

- Model relations among data
- *Nodes* and *edges*



Trees

- Graphs with hierarchical structure
  - Connected graph with N-1 edges
- Nodes as *parents* and *children*



Spatial Layout

The primary concern of graph drawing is the spatial layout of nodes and edges

Often (but not always) the goal is to effectively depict the graph structure

- Connectivity, path-following
- Network distance
- Clustering
- Ordering (e.g., hierarchy level)

Tree Visualization

Indentation

- Linear list, indentation encodes depth

Node-Link diagrams

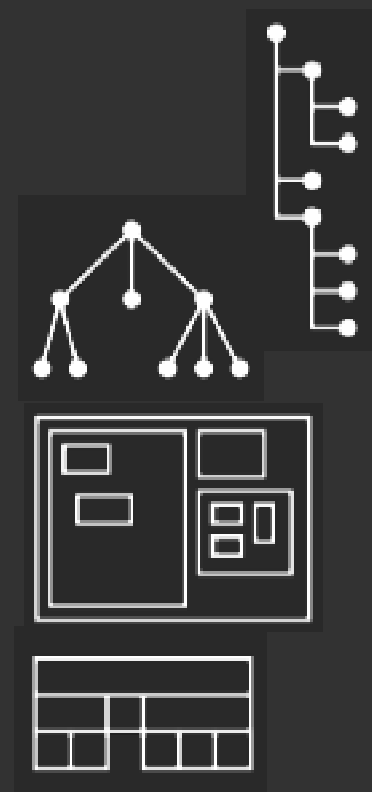
- Nodes connected by lines/curves

Enclosure diagrams

- Represent hierarchy by enclosure

Layering

- Layering and alignment



Examples: http://flare.prefuse.org/demo

Problems with Node-Link Diagrams

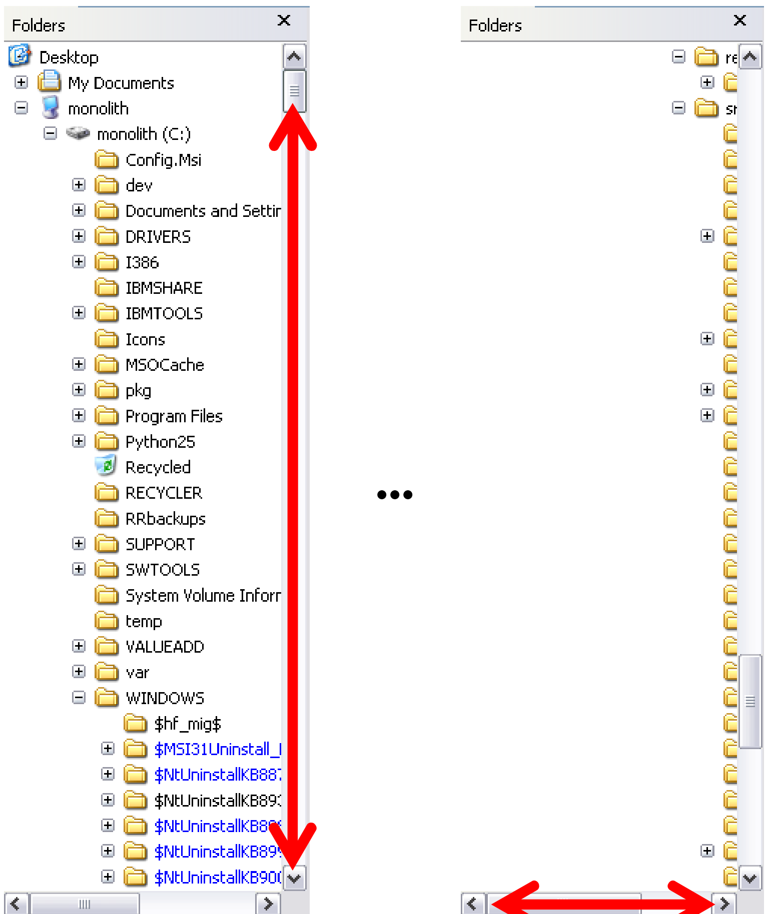
Scale

- Tree breadth often grows exponentially
- Even with tidier layout, quickly run out of space

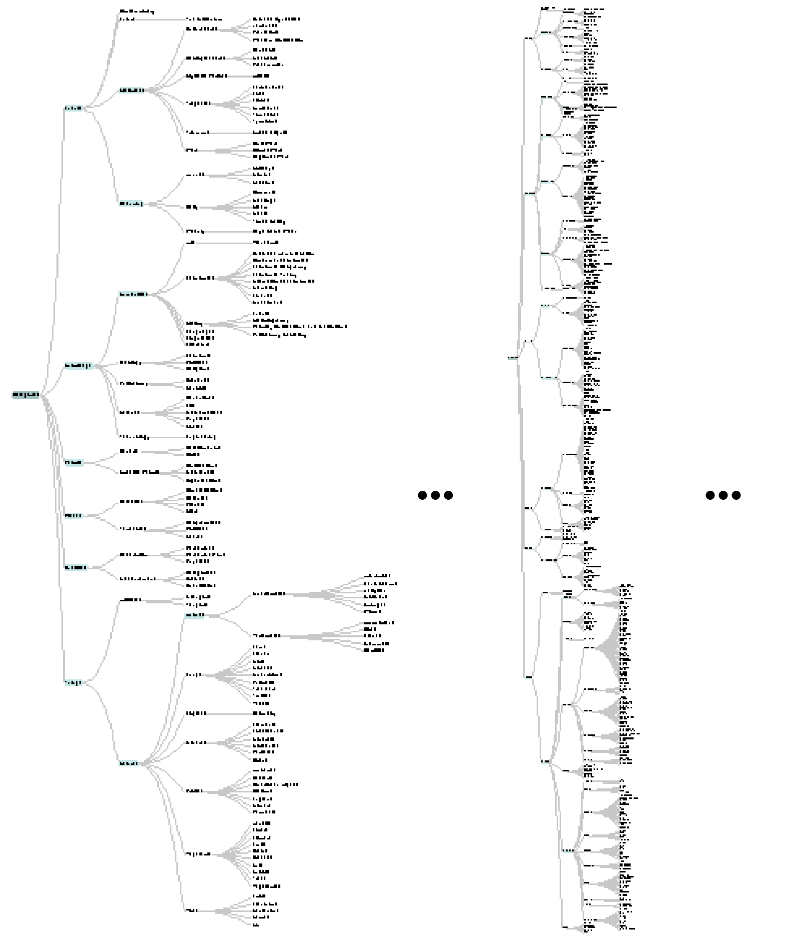
Possible solutions

- Filtering
- Focus+Context
- Scrolling or Panning
- Zooming
- Aggregation

Visualizing Large Hierarchies



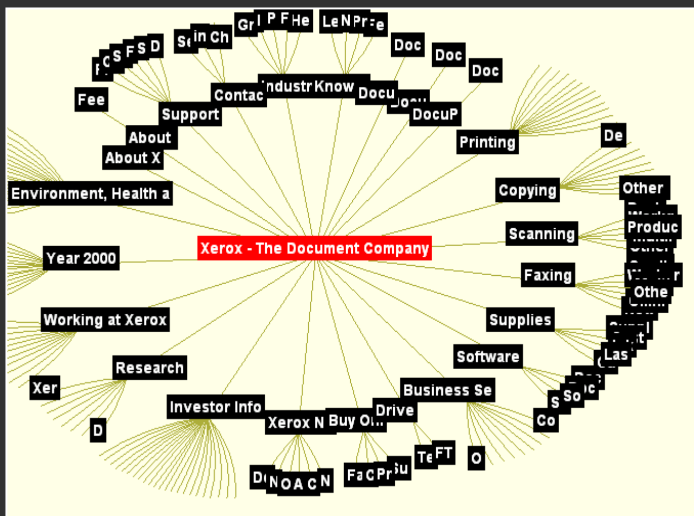
Indented Layout

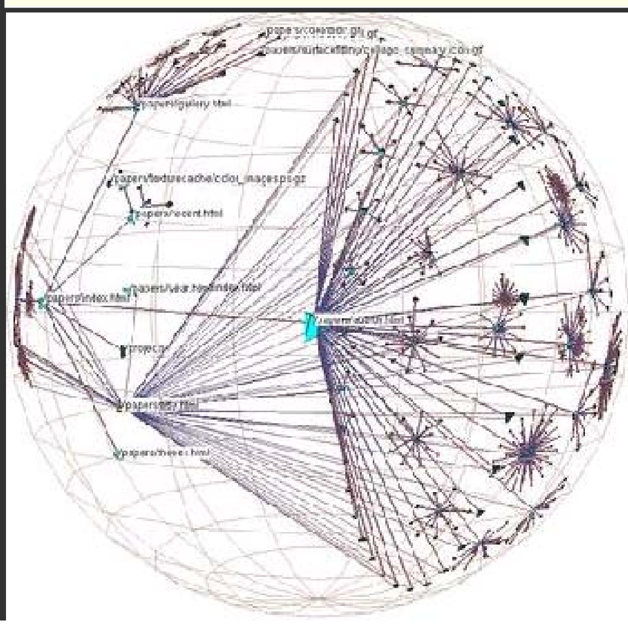


Reingold-Tilford Layout

MC Escher, *Circle Limit IV*

Hyperbolic Layout



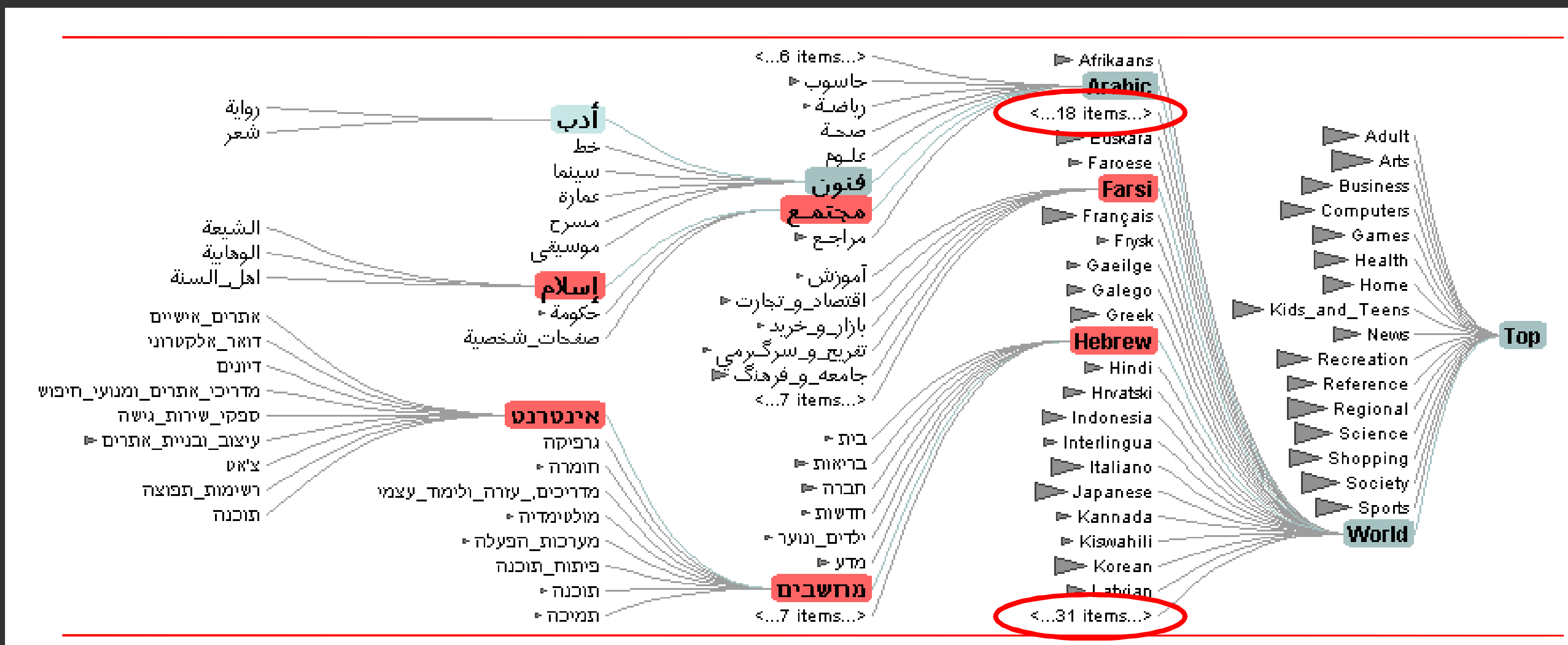


Perform tree layout in hyperbolic geometry, then project the result on to the Euclidean plane.

Why? Like tree breadth, the hyperbolic plane expands exponentially!

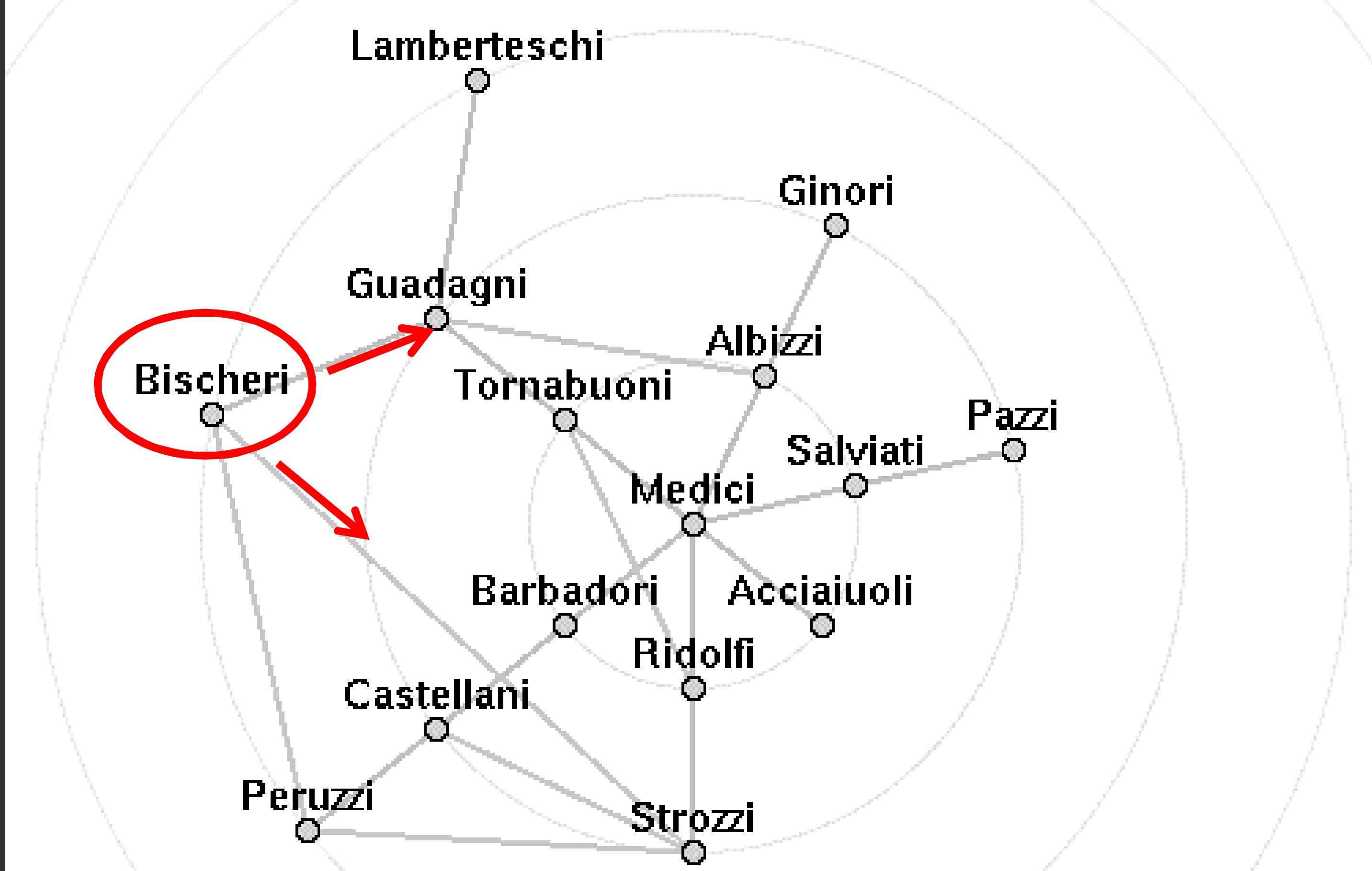
Also computable in 3D, projected into a sphere.

Degree-of-Interest Trees [AVI 04]



Space-constrained, multi-focal tree layout

Graph Visualization



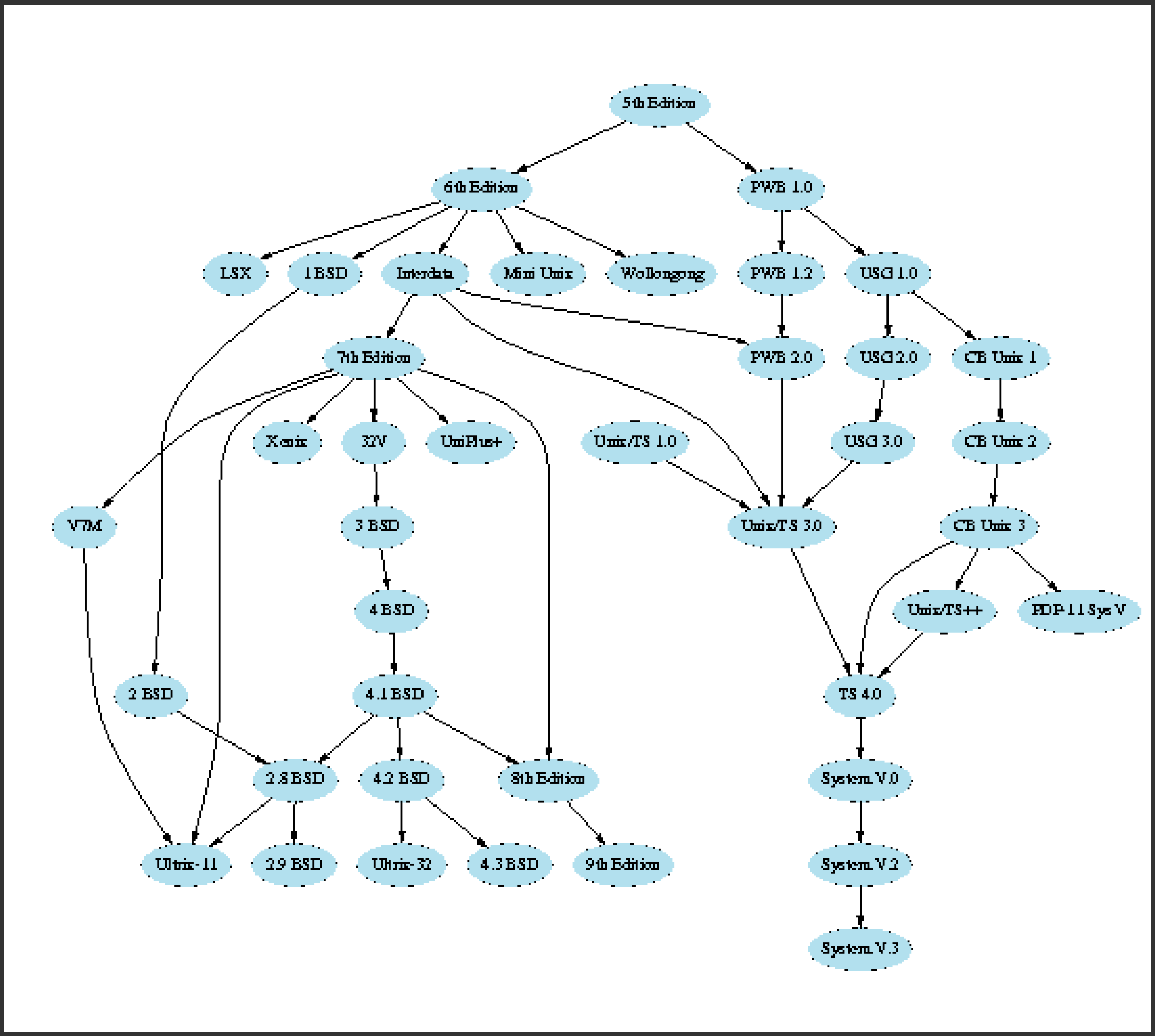
Use a tree layout based on a *spanning tree* of the graph

Sugiyama-style graph layout

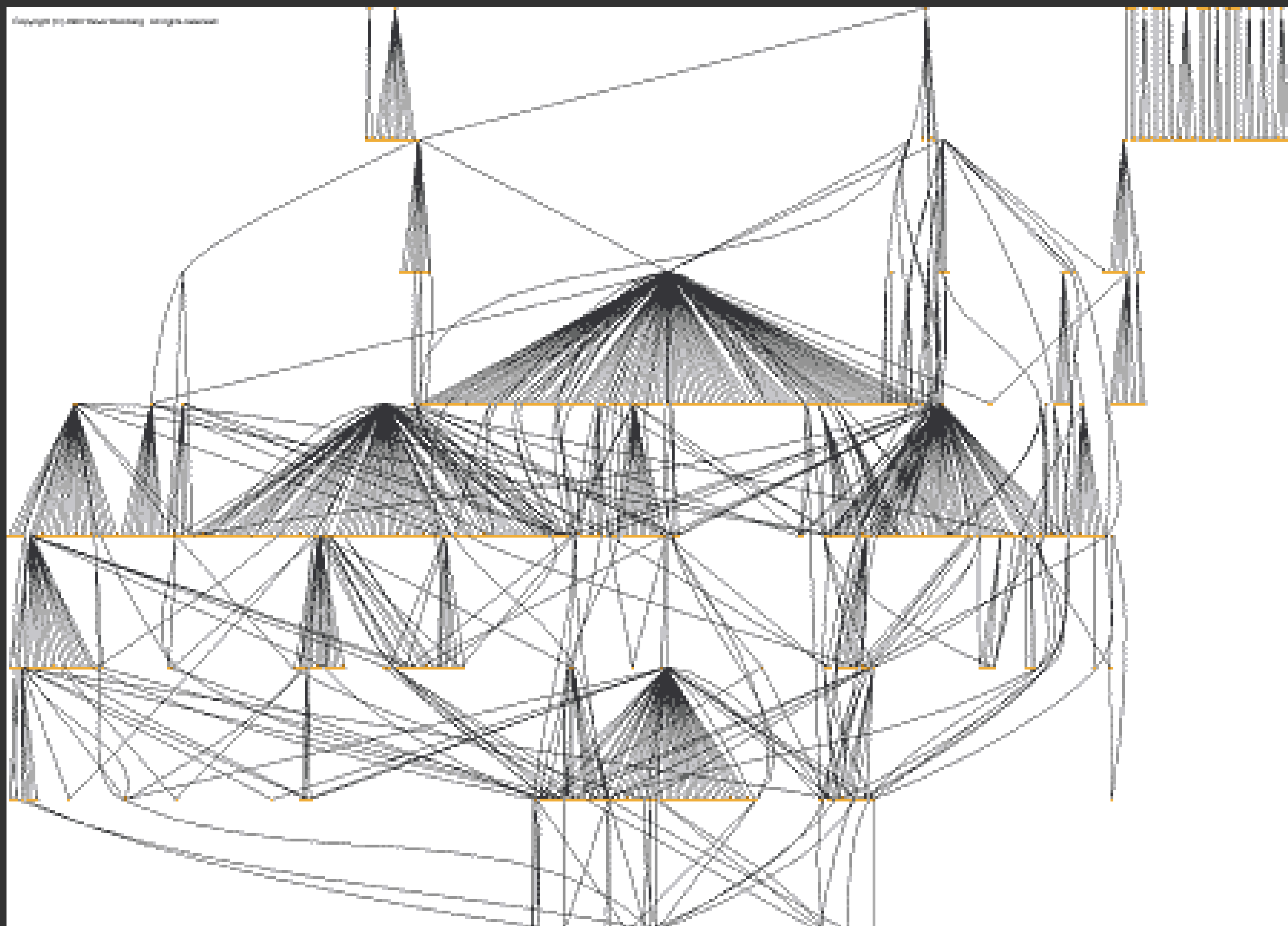
Evolution of the UNIX OS

Hierarchical layering based on descent

Approach used by GraphViz's "dot" layout



Hierarchical graph layout



Gnutella network

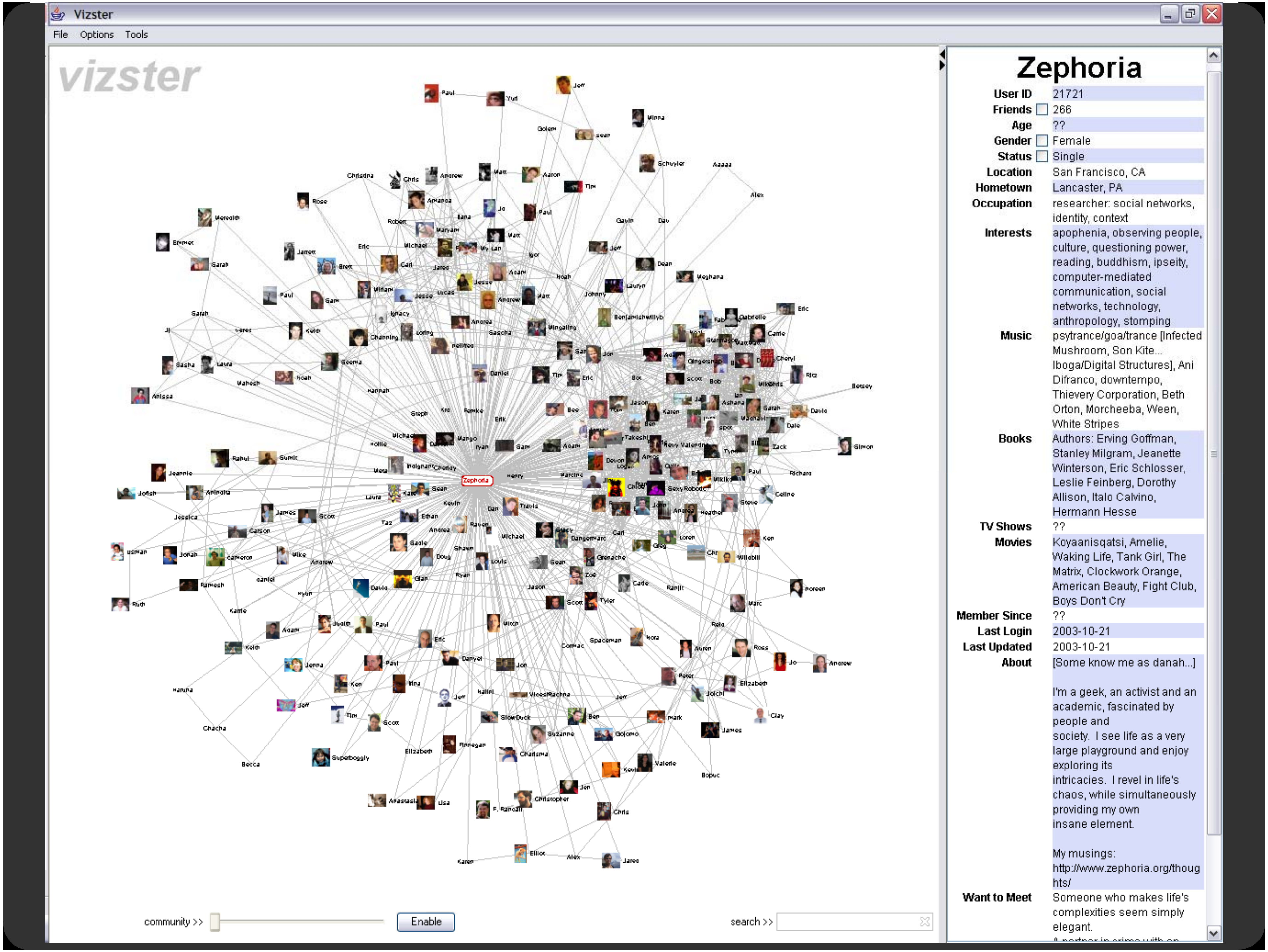
Force-Directed Layout

Edges = springs $F = -k * (x - L)$

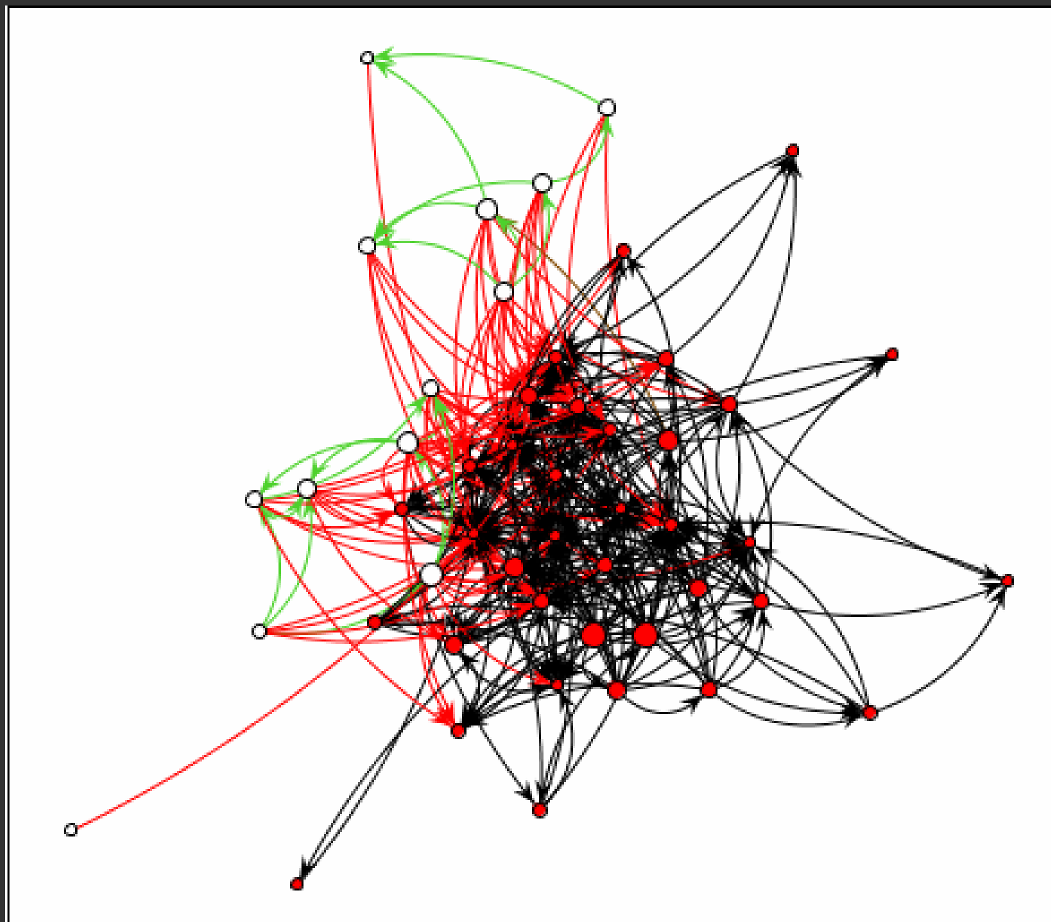
Nodes = charged particles $F = G*m_1*m_2 / x^2$

Repeatedly calculate forces, update node positions

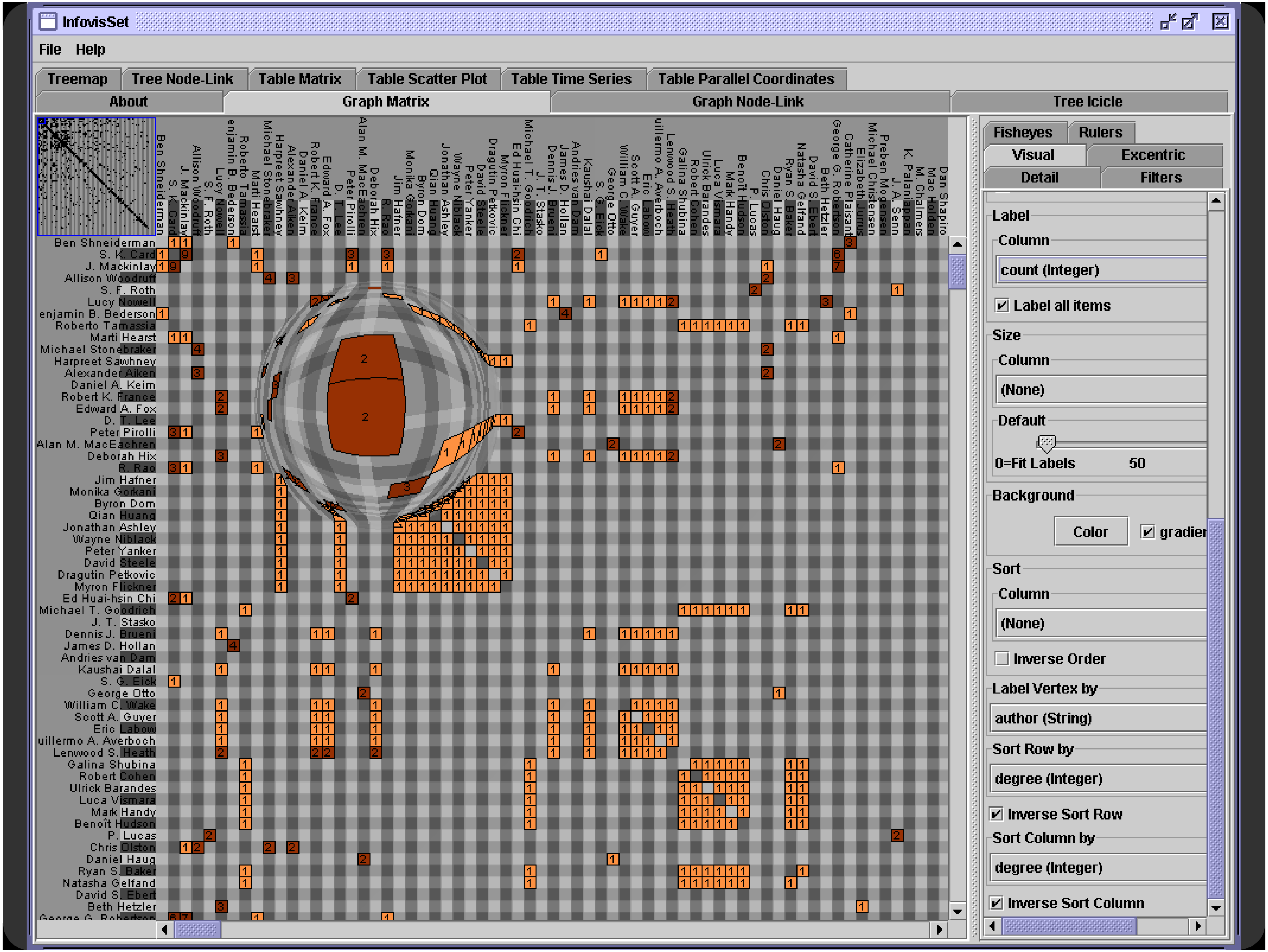
- Naïve approach $O(N^2)$, speed up to $O(N \log N)$ quadtree
- Numerical integration of forces at each time step

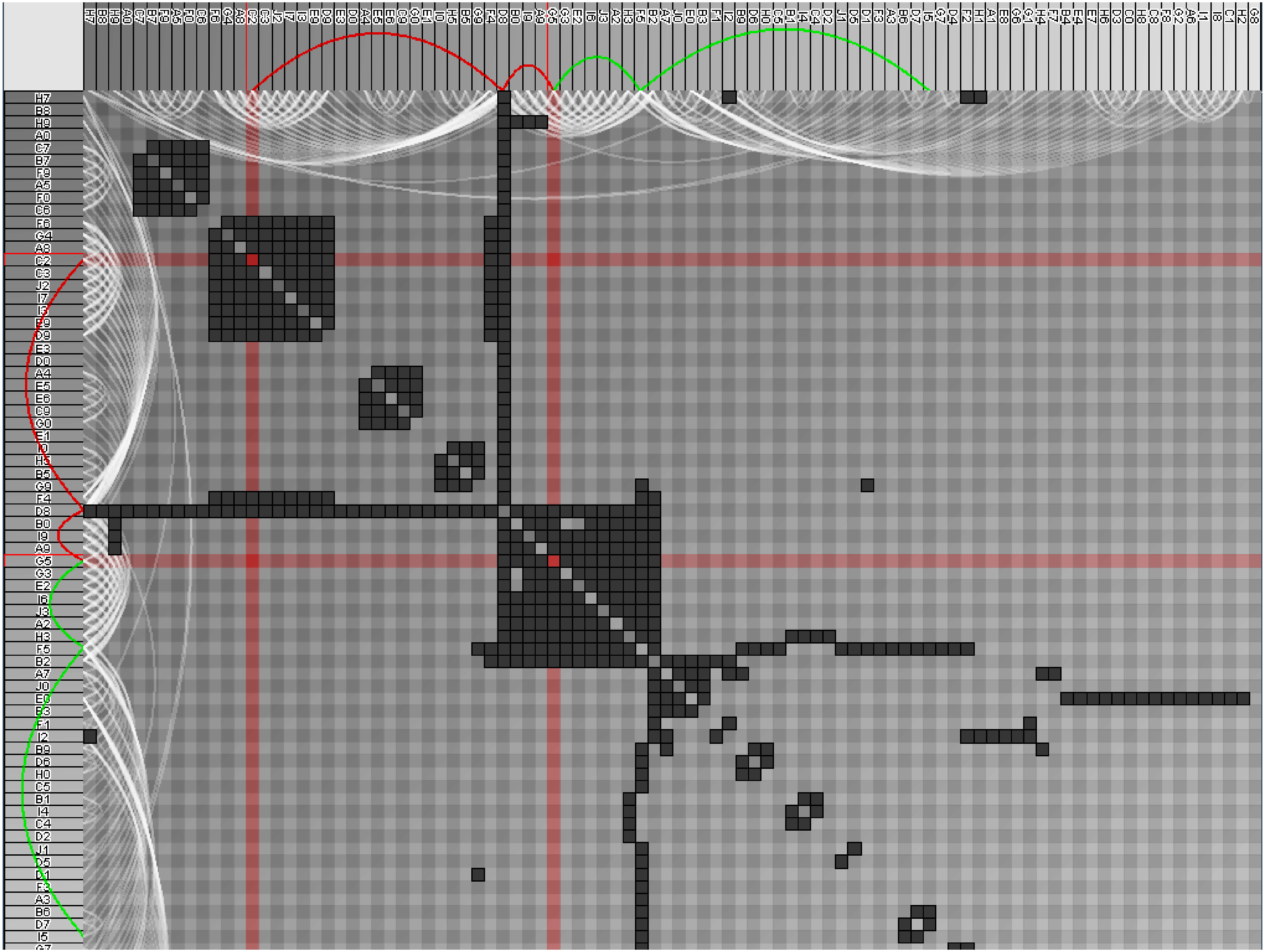

Vizster
File Options Tools
vizster
Zephoria
User ID 21721
Friends 266
Age ??
Gender Female
Status Single
Location San Francisco, CA
Hometown Lancaster, PA
Occupation researcher: social networks, identity, context
Interests apophenia, observing people, culture, questioning power, reading, buddhism, ipseity, computer-mediated communication, social networks, technology, anthropology, stomping
Music psytrance/goa/trance [Infected Mushroom, Son Kite... Iboga/Digital Structures], Ani Difranco, downtempo, Thievery Corporation, Beth Orton, Morcheeba, Ween, White Stripes
Books Authors: Erving Goffman, Stanley Milgram, Jeanette Winterson, Eric Schlosser, Leslie Feinberg, Dorothy Allison, Italo Calvino, Hermann Hesse
TV Shows ??
Movies Koyaanisqatsi, Amelie, Waking Life, Tank Girl, The Matrix, Clockwork Orange, American Beauty, Fight Club, Boys Don't Cry
Member Since ??
Last Login 2003-10-21
Last Updated 2003-10-21
About [Some know me as danah...]
I'm a geek, an activist and an academic, fascinated by people and society. I see life as a very large playground and enjoy exploring its intricacies. I revel in life's chaos, while simultaneously providing my own insane element.
My musings: http://www.zephoria.org/thoughts/
Want to Meet Someone who makes life's complexities seem simply elegant.
community >>
Enable
search >>

Limitations of Node-Link Layout



Edge-crossings and occlusion


InfovisSet
File
Help
Treemap
Tree Node-Link
Table Matrix
Table Scatter Plot
Table Time Series
Table Parallel Coordinates
About
Graph Matrix
Graph Node-Link
Tree Icicle
Fisheyes
Rulers
Visual
Excentric
Detail
Filters
Label
Column
count (Integer)
Label all items
Size
Column
(None)
Default
0=Fit Labels
50
Background
Color
gradier
Sort
Column
(None)
Inverse Order
Label Vertex by
author (String)
Sort Row by
degree (Integer)
Inverse Sort Row
Sort Column by
degree (Integer)
Inverse Sort Column


H7
B8
H9
A0
C7
B7
F9
A5
F0
C6
F6
G4
A8
C2
C3
J2
I7
I3
E9
D9
E3
D0
A4
E5
E6
C9
G0
E1
I0
H5
B5
G9
F4
D8
B0
I9
A9
G5
G3
E2
I6
J3
A2
H3
F5
B2
A7
J0
E0
B3
F1
I2
B9
D6
H0
C5
B1
I4
C4
D2
J1
D5
D1
F3
A3
B6
D7
I5
G7

Attribute-Driven Layout

Large node-link diagrams get messy!

Is there additional structure we can exploit?

Idea: Use data attributes to perform layout

- e.g., scatter plot based on node values

Dynamic queries and/or brushing can be used to explore connectivity

Attribute-Driven Layout

The “Skitter” Layout

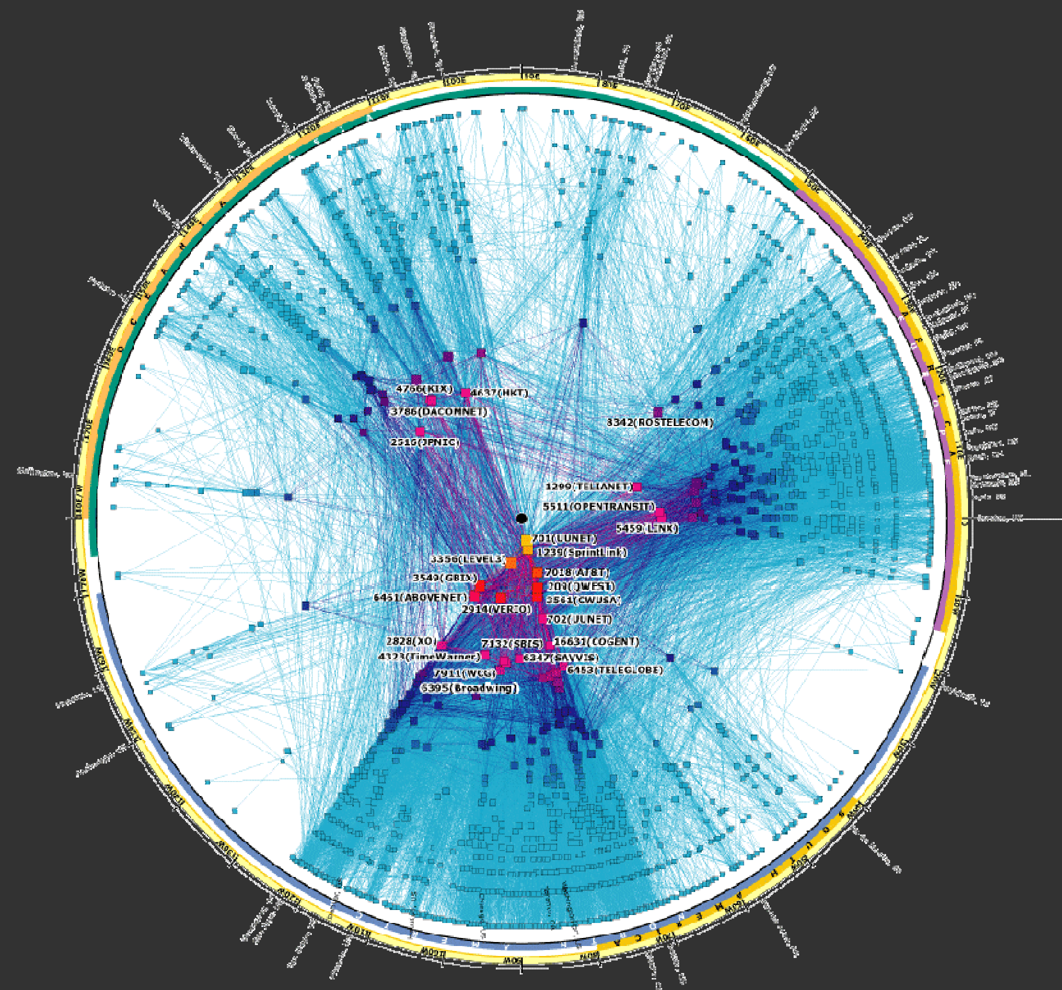
- Internet Connectivity
- Radial Scatterplot

Angle = Longitude

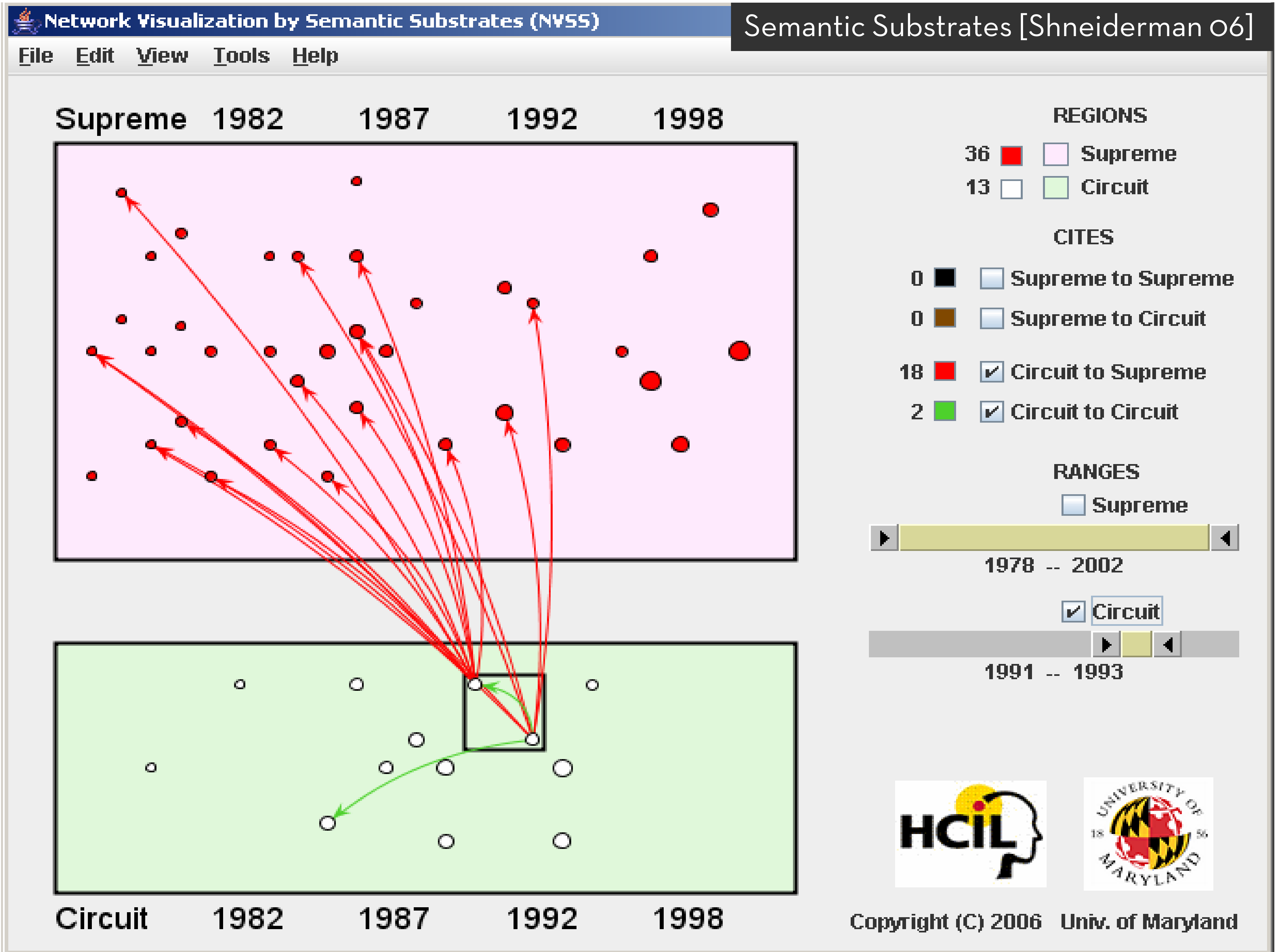
- Geography

Radius = Degree

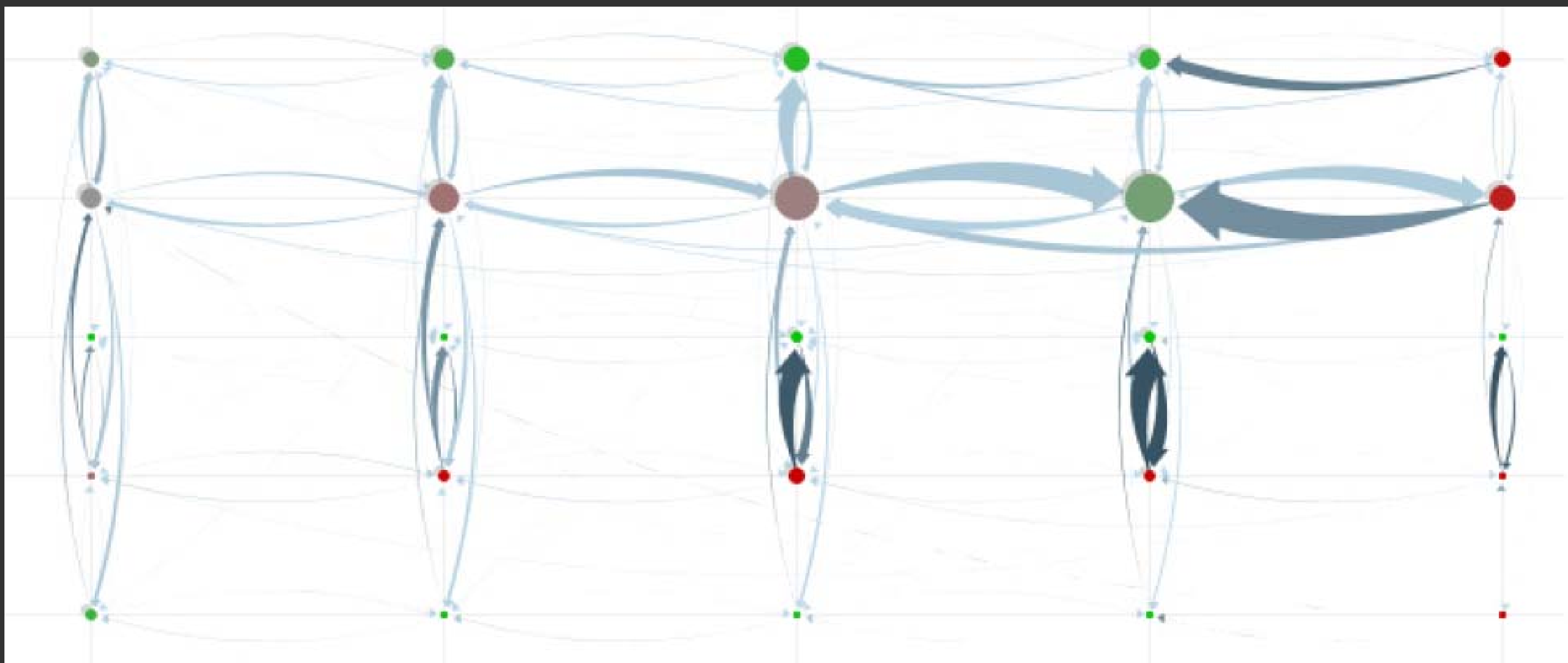
- # of connections
- (a statistic of the nodes)



Semantic Substrates [Shneiderman 06]

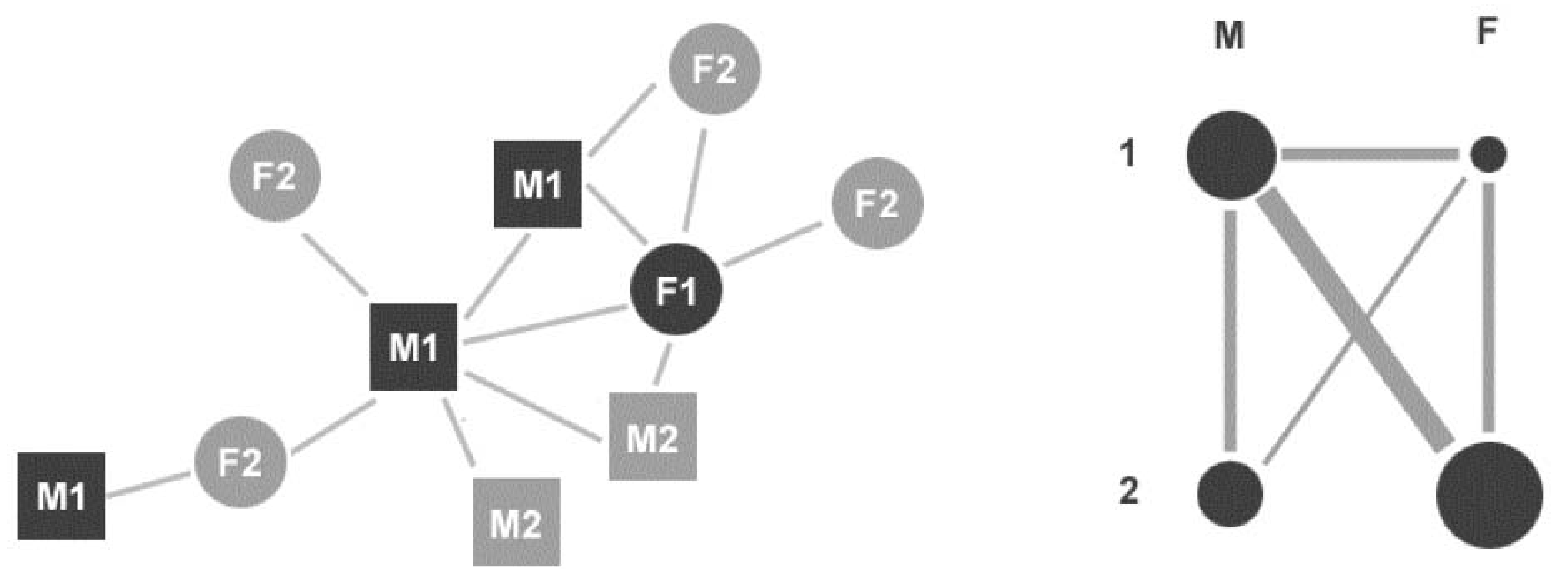


PivotGraph [Wattenberg 2006]



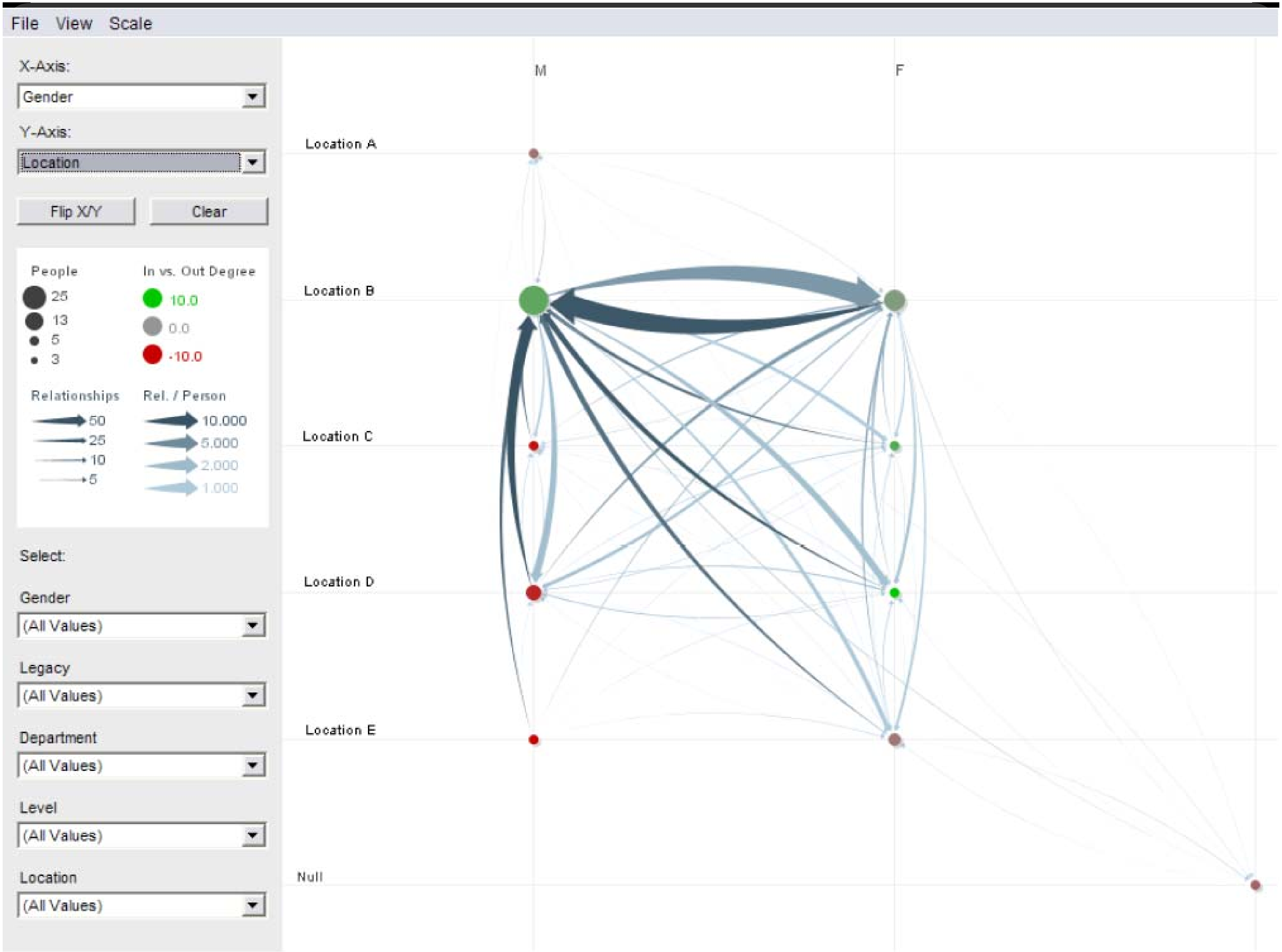
Layout aggregated graphs according to node attributes.
Analogous to pivot tables and trellis display.

PivotGraph



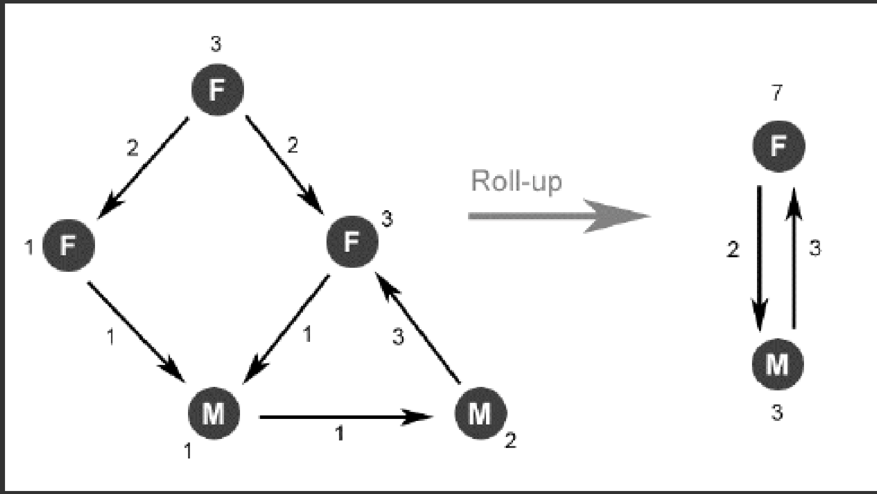
Node and Link Diagram

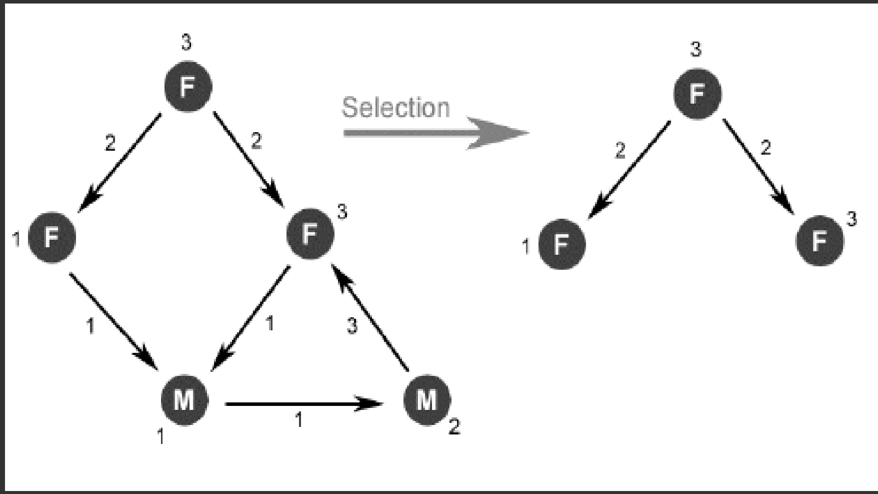
PivotGraph Roll-up



File View Scale
X-Axis:
Gender
Y-Axis:
Location
Flip X/Y
Clear
People
25
13
5
3
In vs. Out Degree
10.0
0.0
-10.0
Relationships
50
25
10
5
Rel. / Person
10.000
5.000
2.000
1.000
Select:
Gender
(All Values)
Legacy
(All Values)
Department
(All Values)
Level
(All Values)
Location
(All Values)
M
F
Location A
Location B
Location C
Location D
Location E
Null

Operators



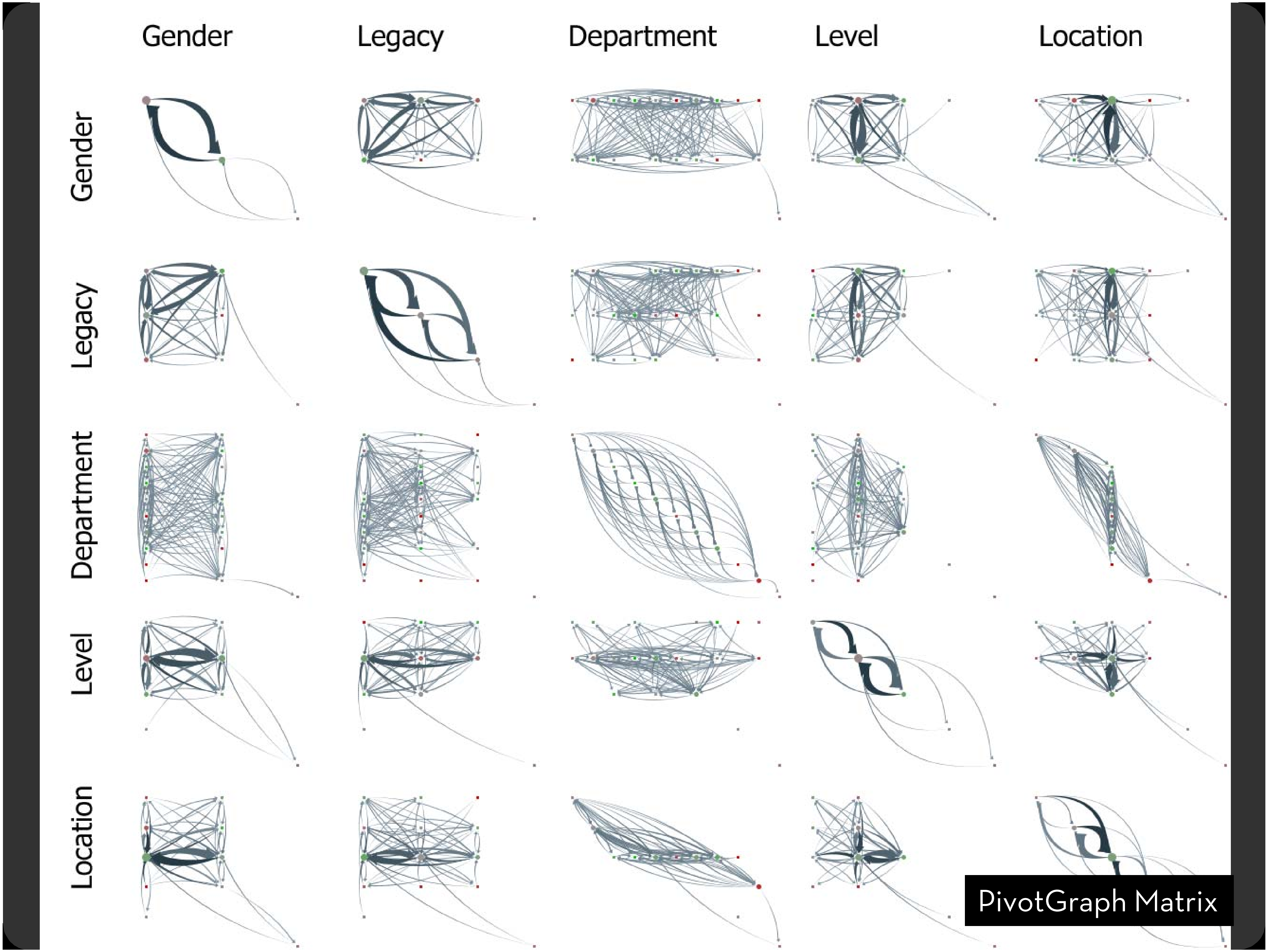


Roll-Up

Aggregate items with matching data values

Selection

Filter on data values


Gender
Legacy
Department
Level
Location
Gender
Legacy
Department
Level
Location
PivotGraph Matrix

Where can you find useful resources?

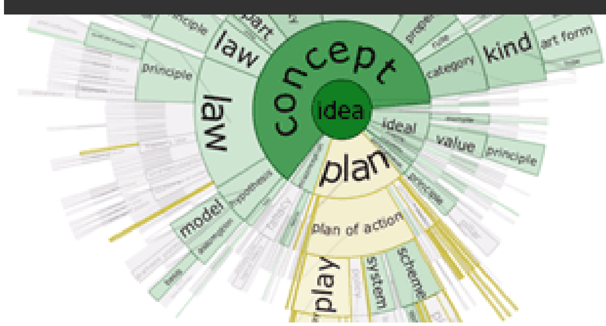
Free Visualization Software

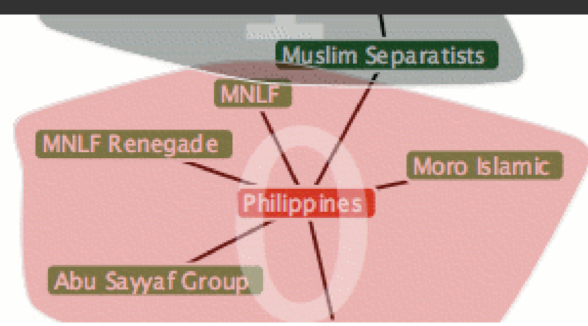
Tableau	http://tableausoftware.com/public
Many-Eyes	http://many-eyes.com
GGobi	http://ggobi.org
GGPlot2 (in R)	http://had.co.nz/ggplot2
NodeXL	http://nodexl.codeplex.com
Gephi	http://gephi.org
GraphViz	http://www.graphviz.org

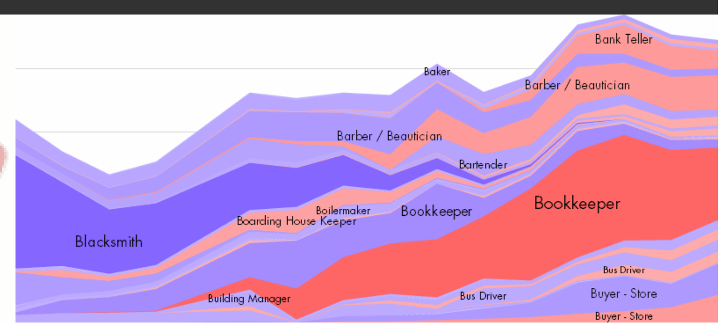
and many others...

Programming Tools

protovis.org	Visualization tools for JavaScript
flare.prefuse.org	Visualization tools for Flash
prefuse.org	Visualization tools for Java
modestmaps.com	Mapping tools for Flash/JavaScript
processing.org	A popular graphics language







Books

The Grammar of Graphics, *Leland Wilkinson*

Visualizing Data, *William S. Cleveland*

The Visual Display of Quantitative Information, *Edward Tufte*

Information Visualization: Perception for Design, *Colin Ware*

Show Me the Numbers: Designing Tables and Graphs to Enlighten, *Stephen Few*

cs448b **Data Visualization**

Fall 2010

Summary

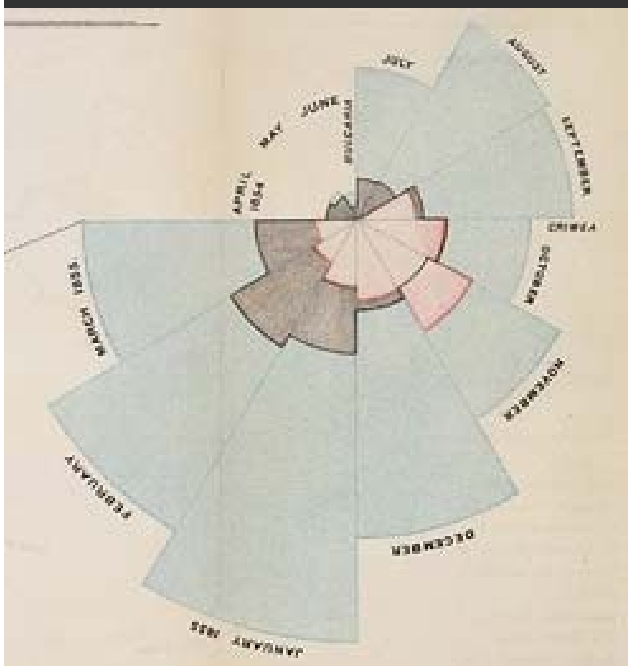
Effective visualizations can help us **record**, **analyze**, and **communicate** data

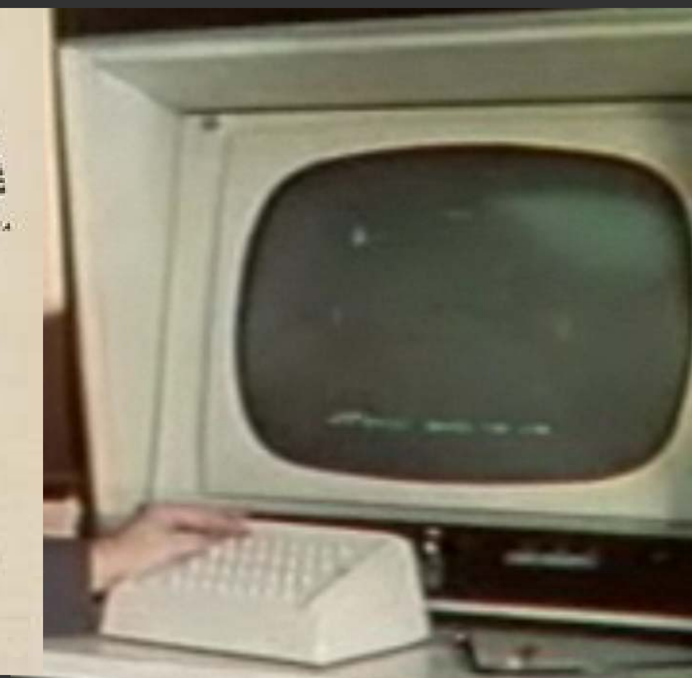
Creating visualizations requires **modeling data** and applying **appropriate visual encodings** and **interaction techniques**.

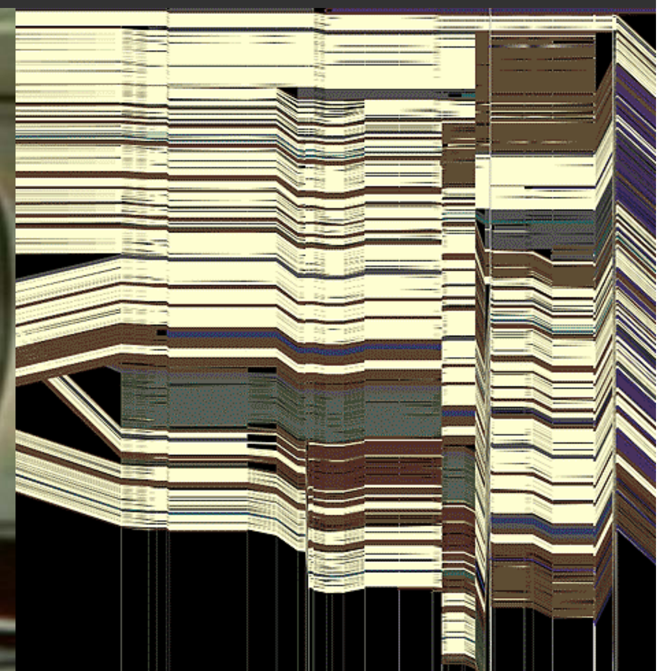
As you encounter visualizations in the world, try to **deconstruct what you see**. What are the underlying data types and encodings?

a brief introduction to

Data Visualization







Jeffrey Heer hci.stanford.edu/jheer