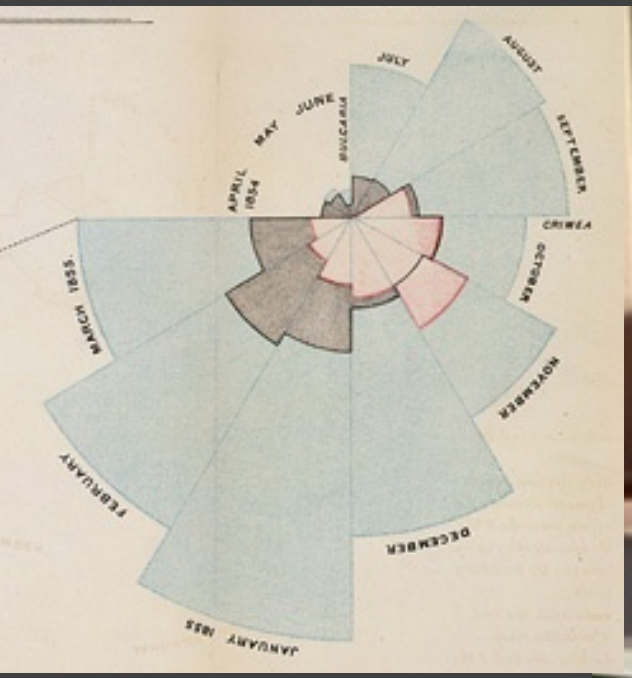
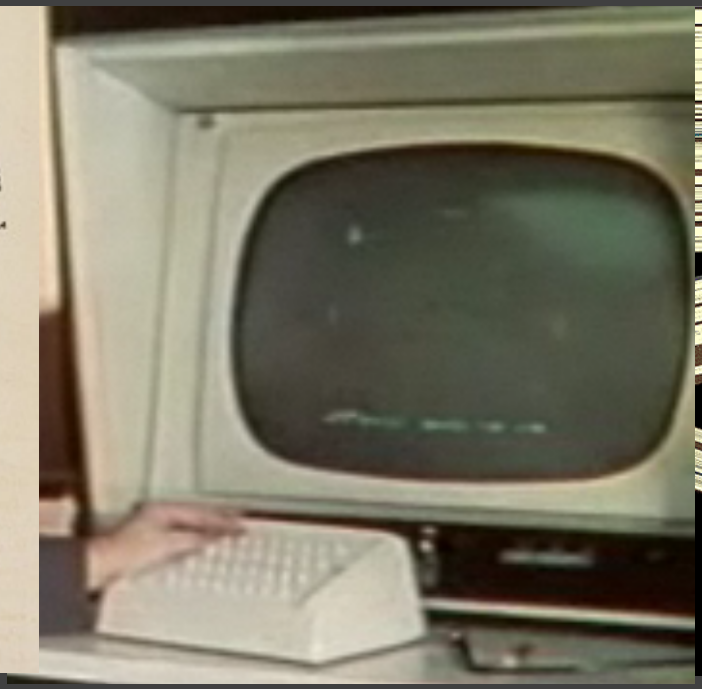
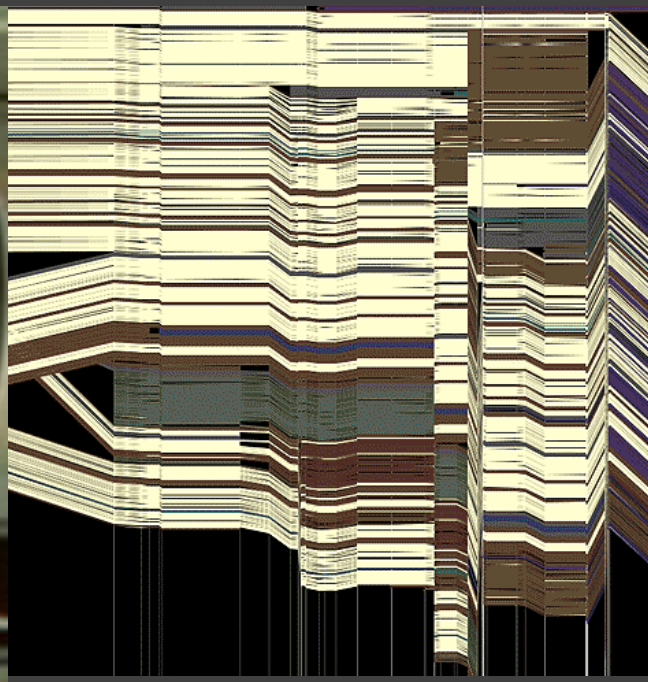
CS448B :: 8 Nov 2012

Using Space Effectively







Jeffrey Heer UW
Justin Talbot Tableau

Rank	Aspect judged
1	Position along a common scale
2	Position on identical but nonaligned scales
3	Length
4	Angle
	Slope (with θ not too close to 0, $\pi/2$, or π radians)
5	Area
6	Volume
	Density
	Color saturation
7	Color hue

[Cleveland & McGill 85]

Space is the most important encoding.
How can we use it effectively?

Strategies

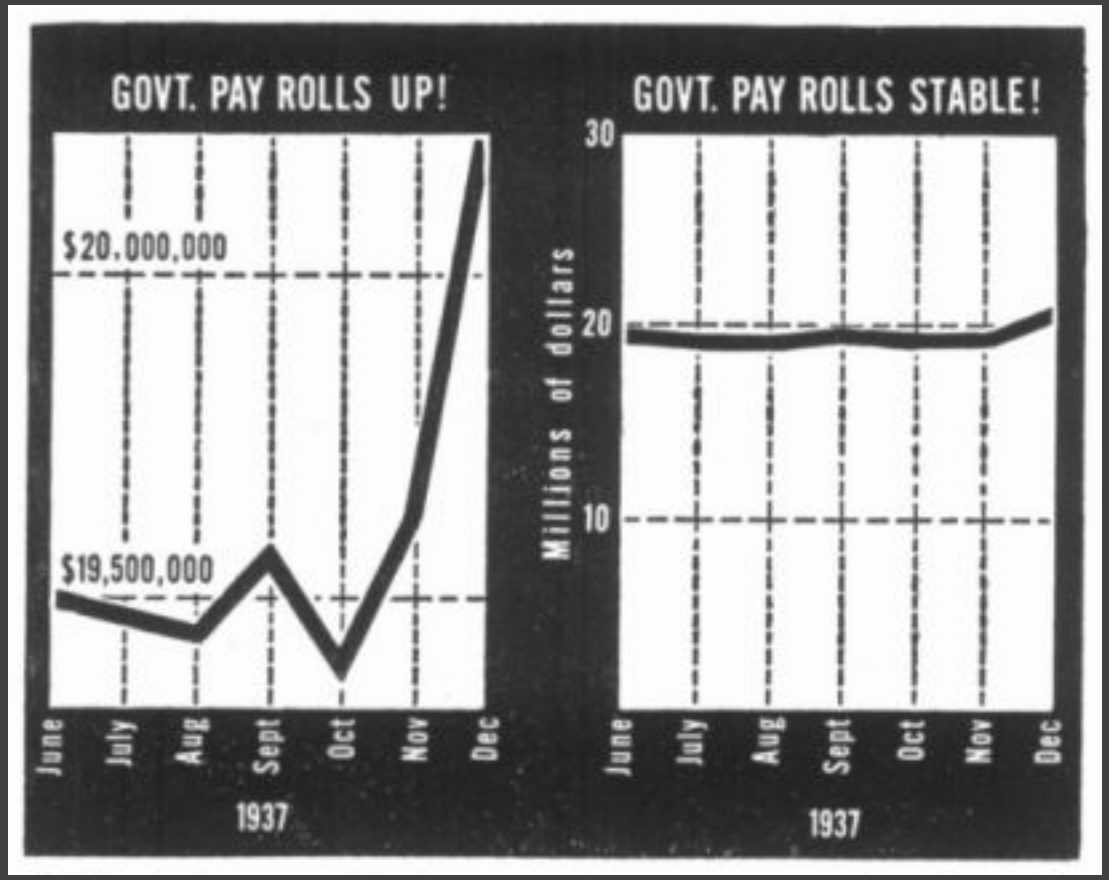
1. Focus + context
2. Transform data
3. Transform viz (aspect ratio)
4. Optimize layout (streamgraph)

Lots of tradeoffs.
Lots of history to learn from.

1. Focus + context

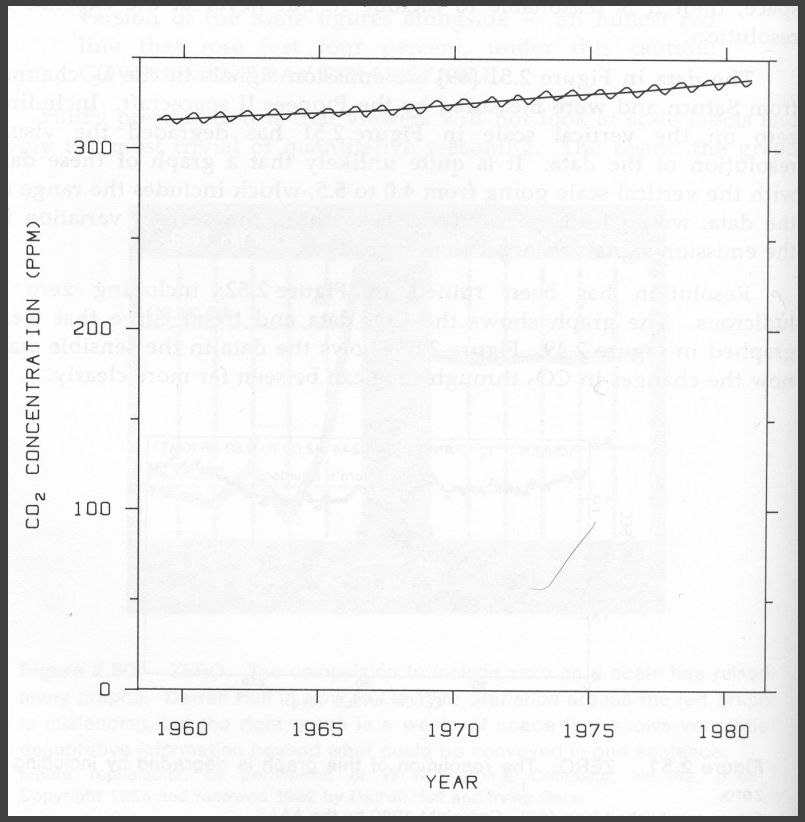
Focus: carefully pick what to show

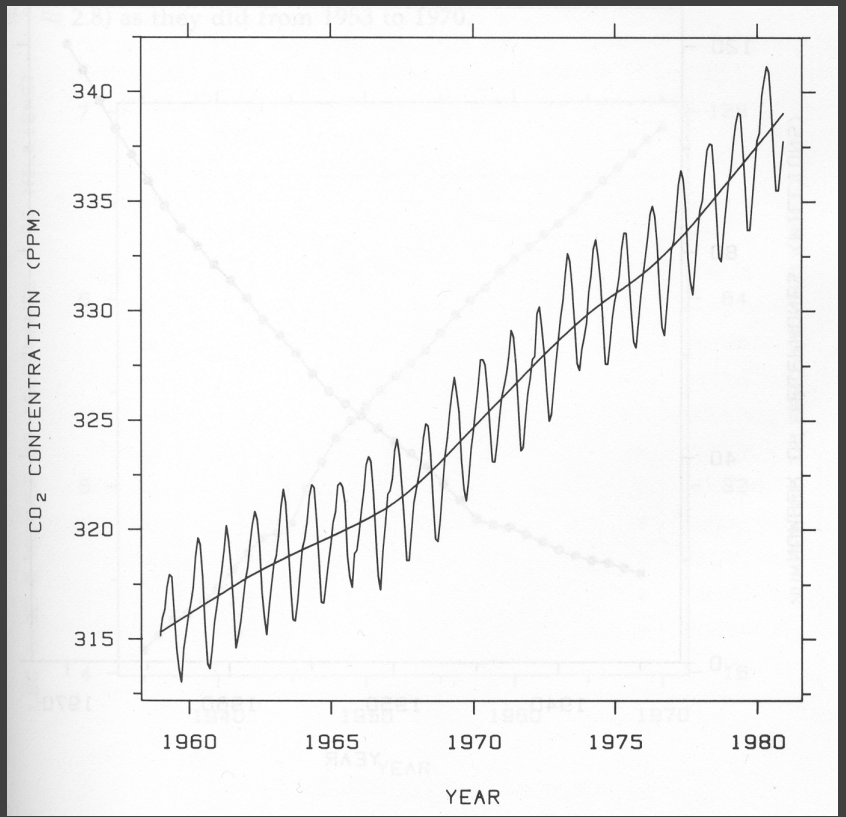
Should you include 0?



Government payrolls in 1937 [How To Lie With Statistics. Huff]

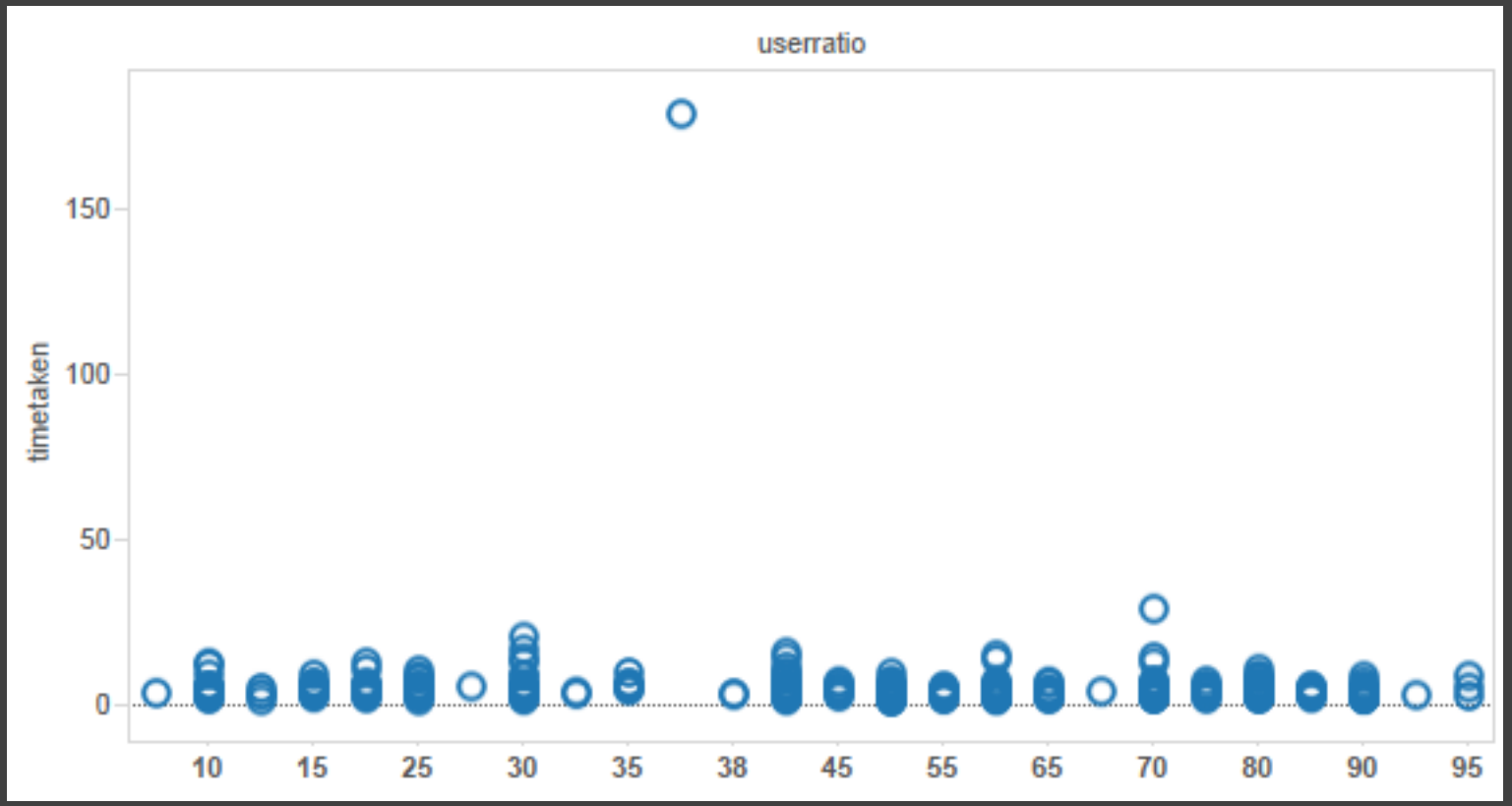
Should you include 0?



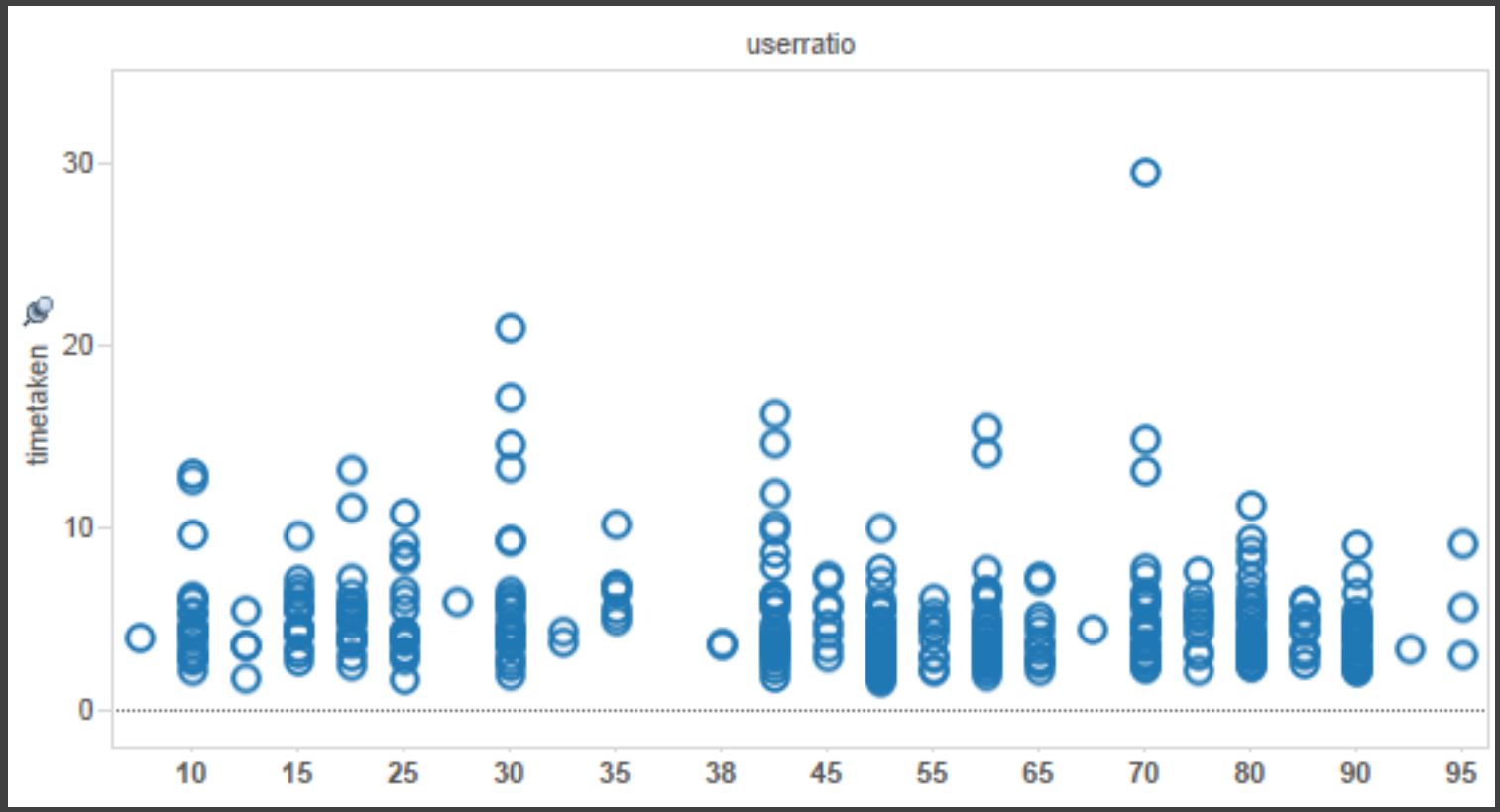


Yearly CO_2 concentrations [Cleveland 85]

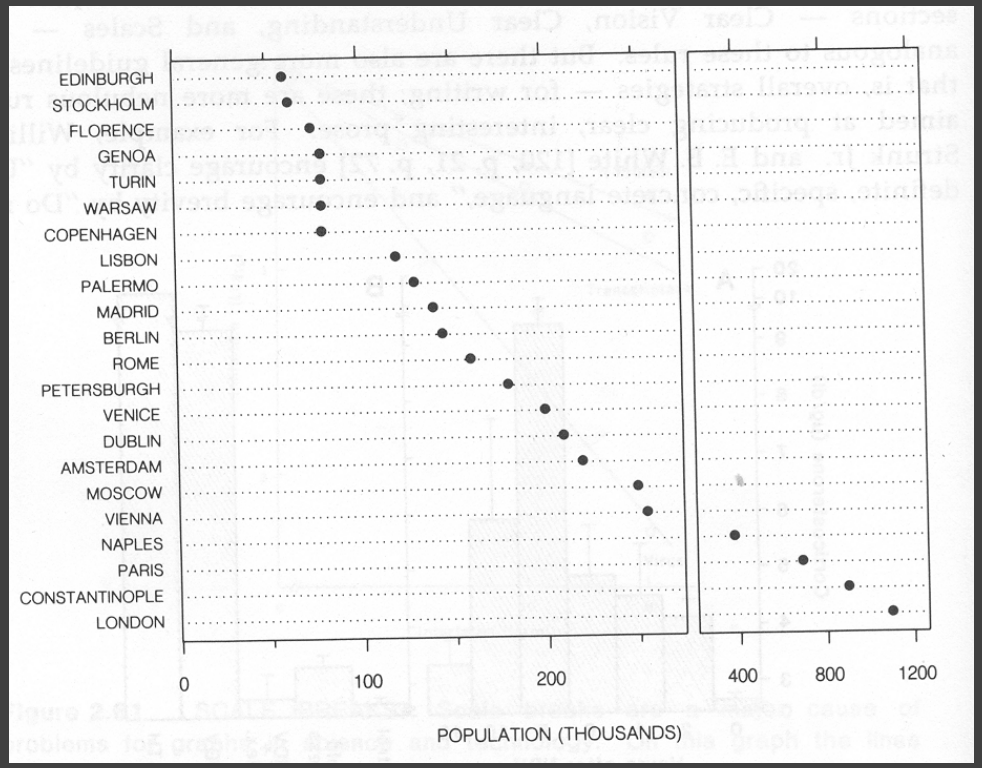
Outliers


userratio
timetaken
150
100
50
0
10
15
25
30
35
38
45
55
65
70
80
90
95

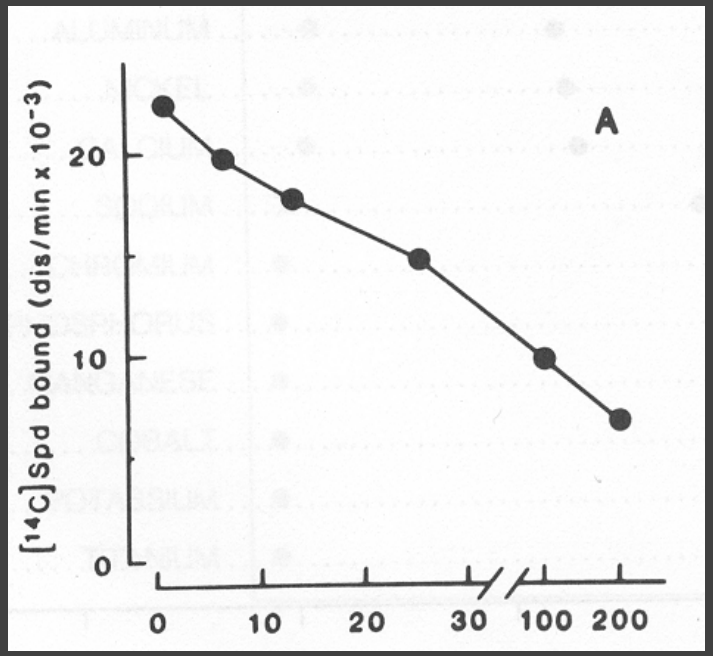
Outliers


userratio
timetaken
30
20
10
0
10
15
25
30
35
38
45
55
65
70
80
90
95

Scale breaks



Well marked scale break [Cleveland 85]



Poor scale break [Cleveland 85]

Caveat: whitespace changes viz!

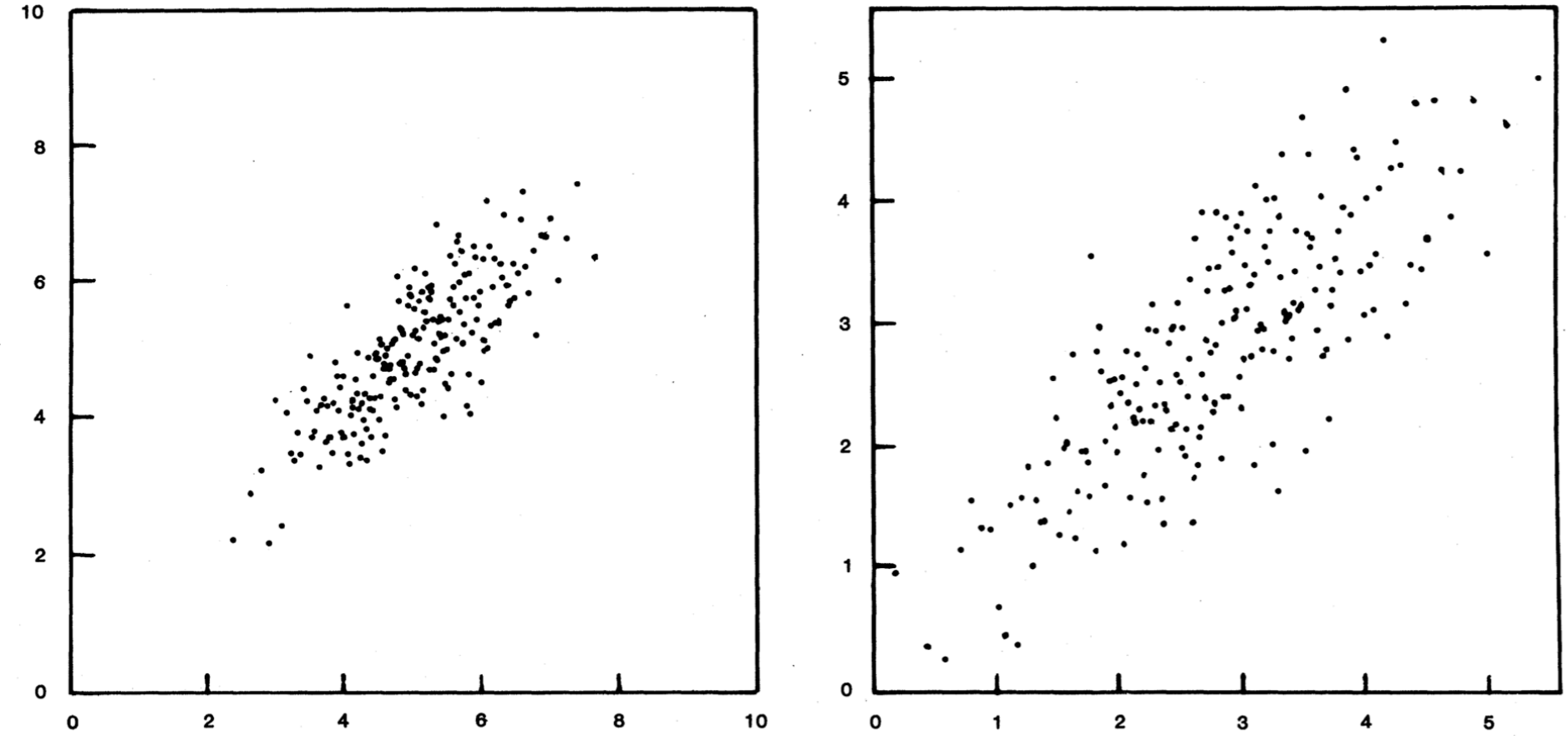


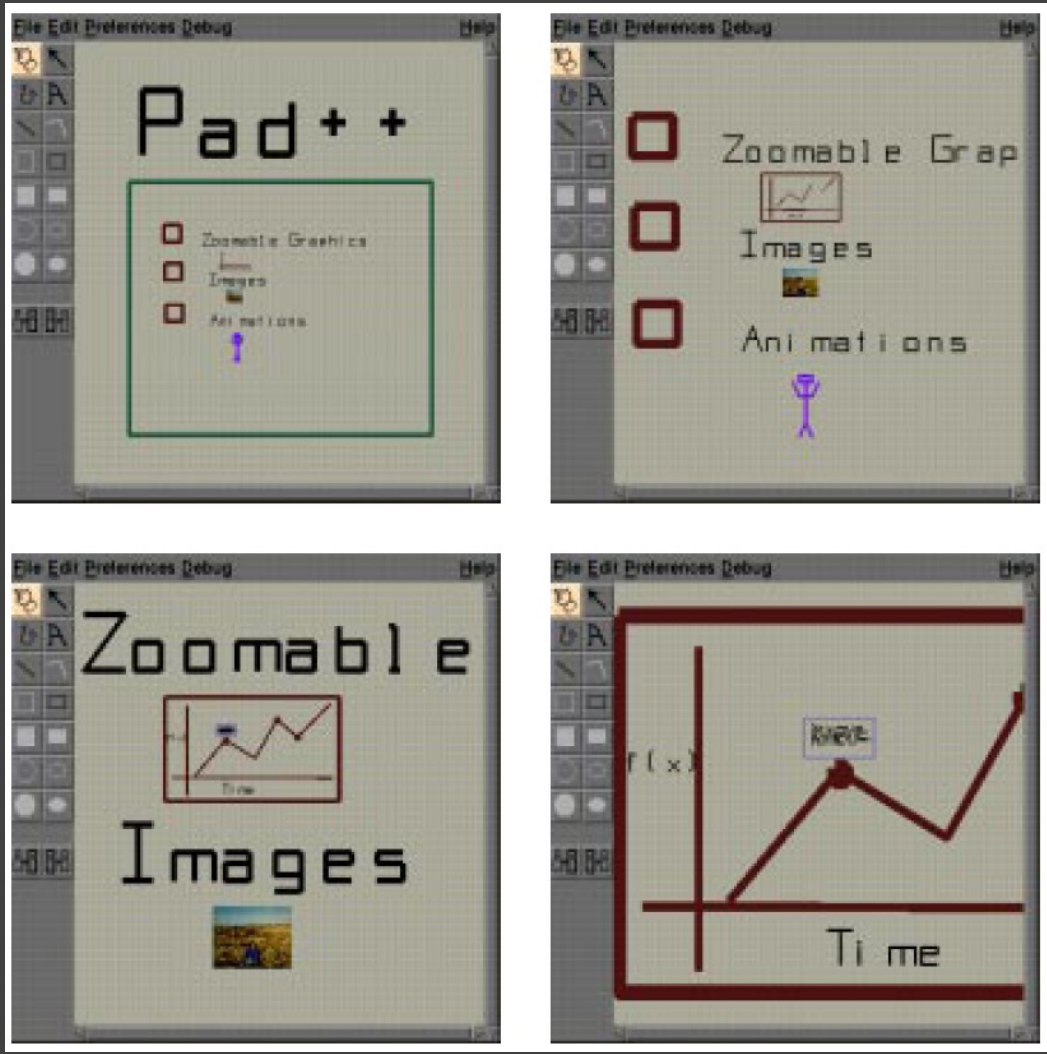
Fig. 1. Reductions of two scatterplots used in the three types of experiments. The left panel is point-cloud size 2 and the right panel is point-cloud

[Cleveland et al. 82]

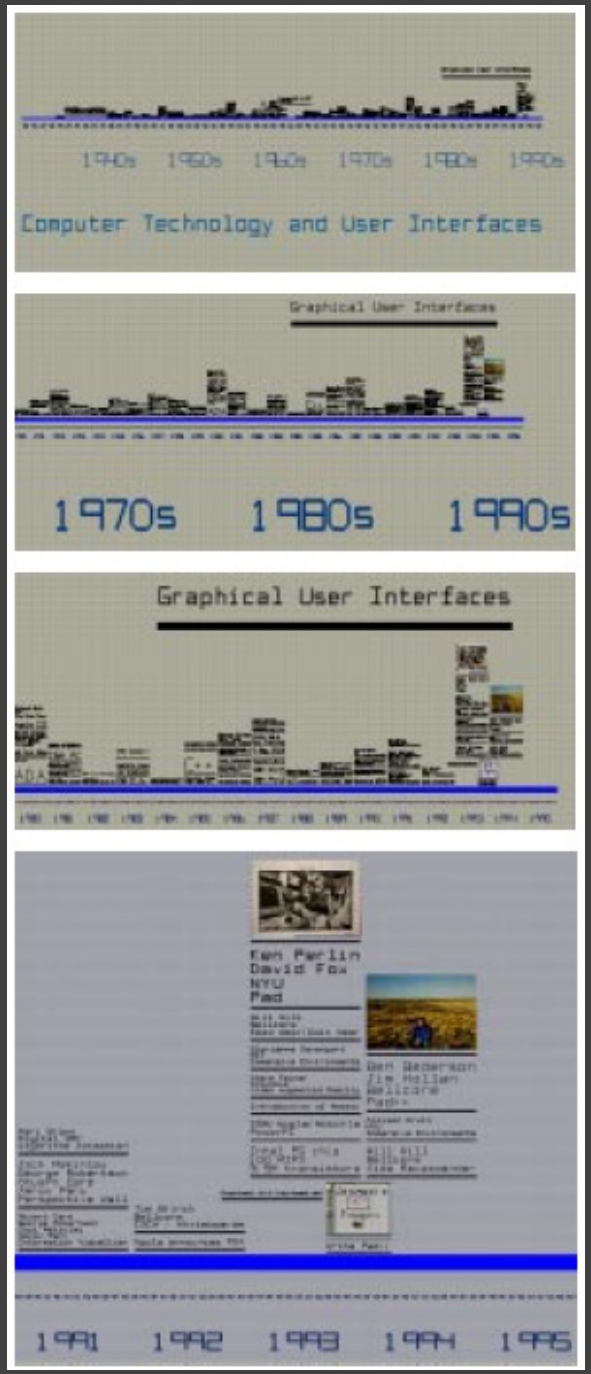
Zooming – Dynamic Focus

Eames' Powers of Ten [http://www.powersof10.com/]

Interactive Zooming



Pad++ [Bederson and Hollan 94]



Pad++

- Play video

Semantic Zooming

Change representations as zoom level changes

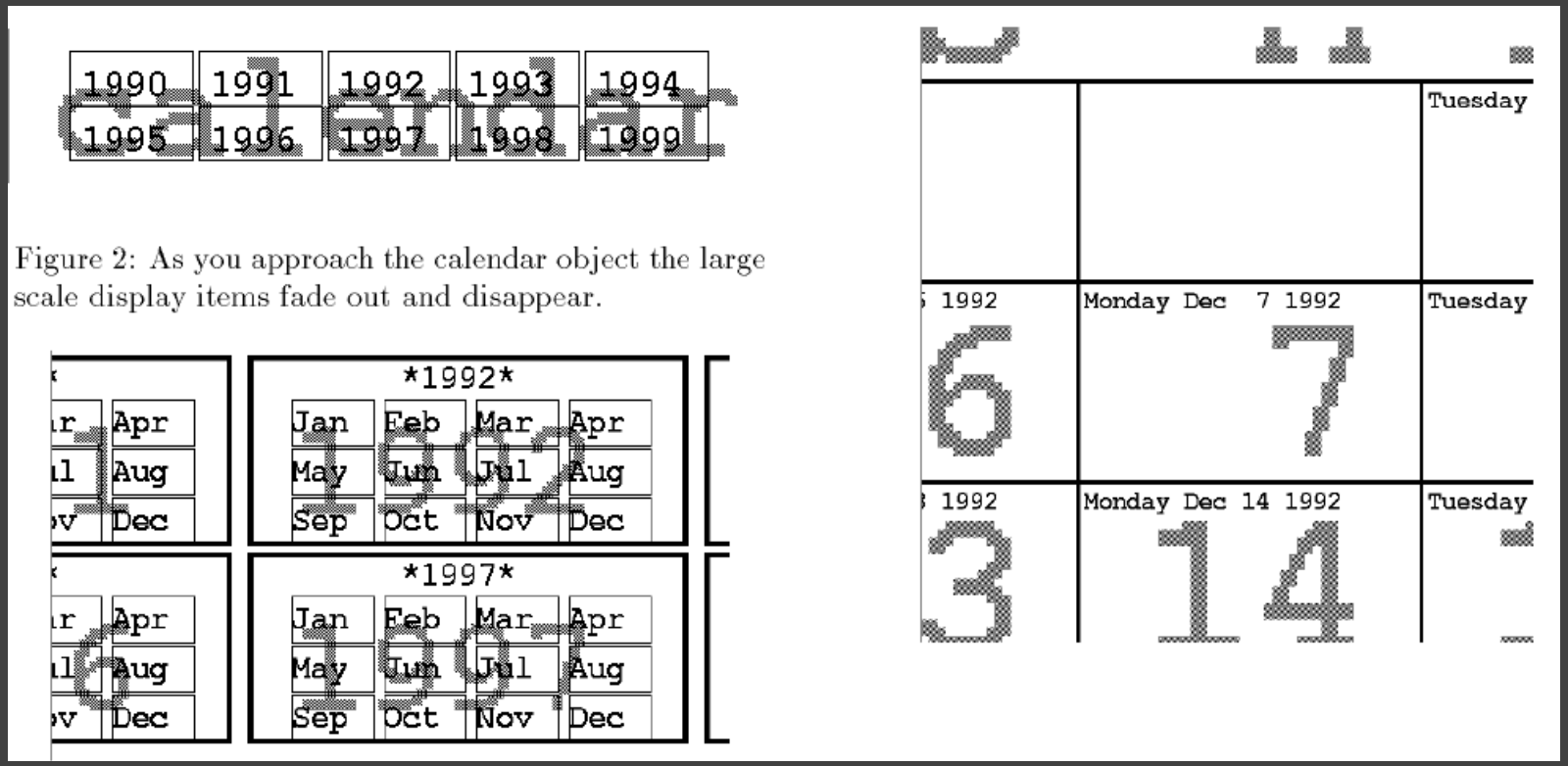


Figure 2: As you approach the calendar object the large scale display items fade out and disappear.

PAD [Perlin and Fox 93]

Semantic Zooming

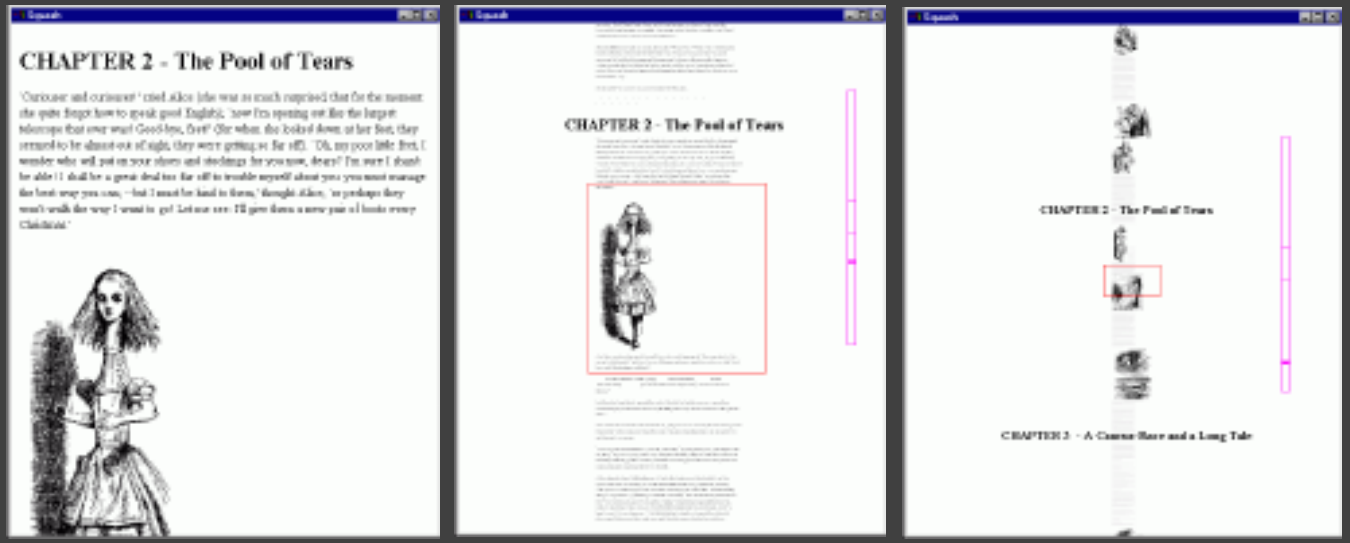
Windows 8 video

Speed-Dependent Zooming

Integrate Pan and Zoom into single interation

Automatically zoom to maintain optical flow

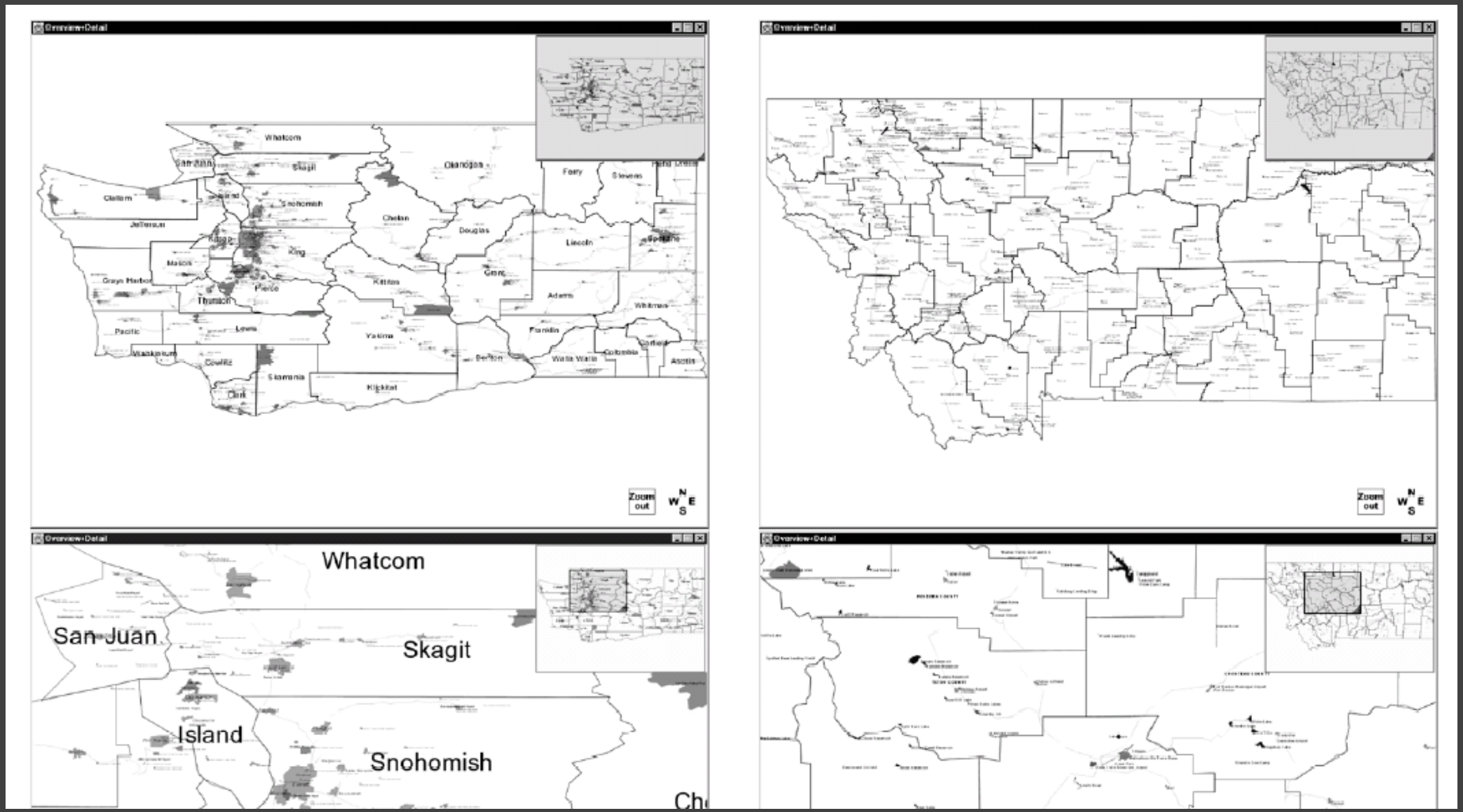
Semantic zooming can simplify zoomed-out view



http://www-ui.is.s.u-tokyo.ac.jp/~takeo/java/autozoom/autozoom.htm [Igarashi 00]

Context: hint at what you're not showing

Maps



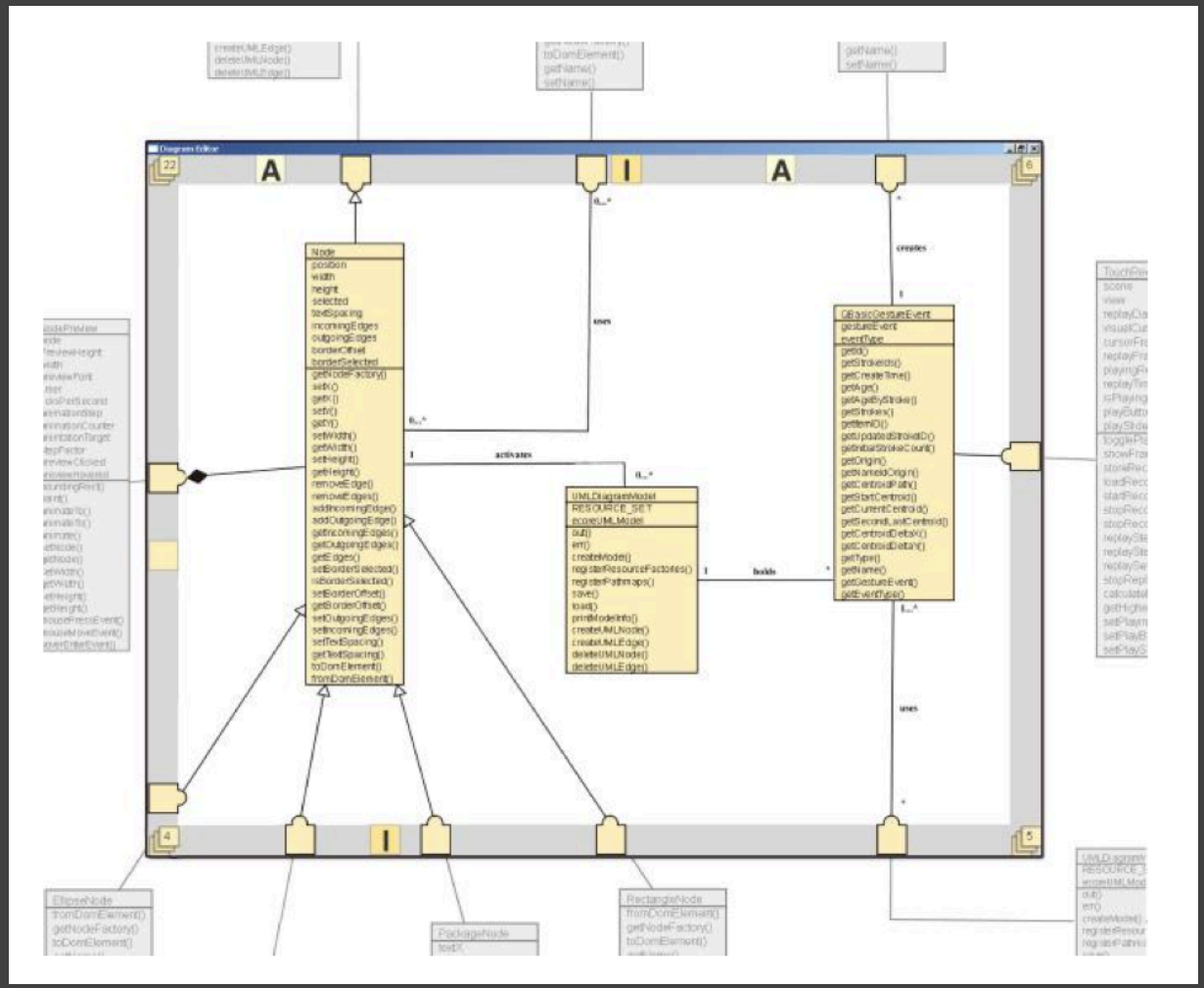
[Hornbaek et al. 2002]

Google Finance

DATA

Halo [Baudisch 03]

Node-link [Frisch & Dachselt 2013]



Focus + Context?

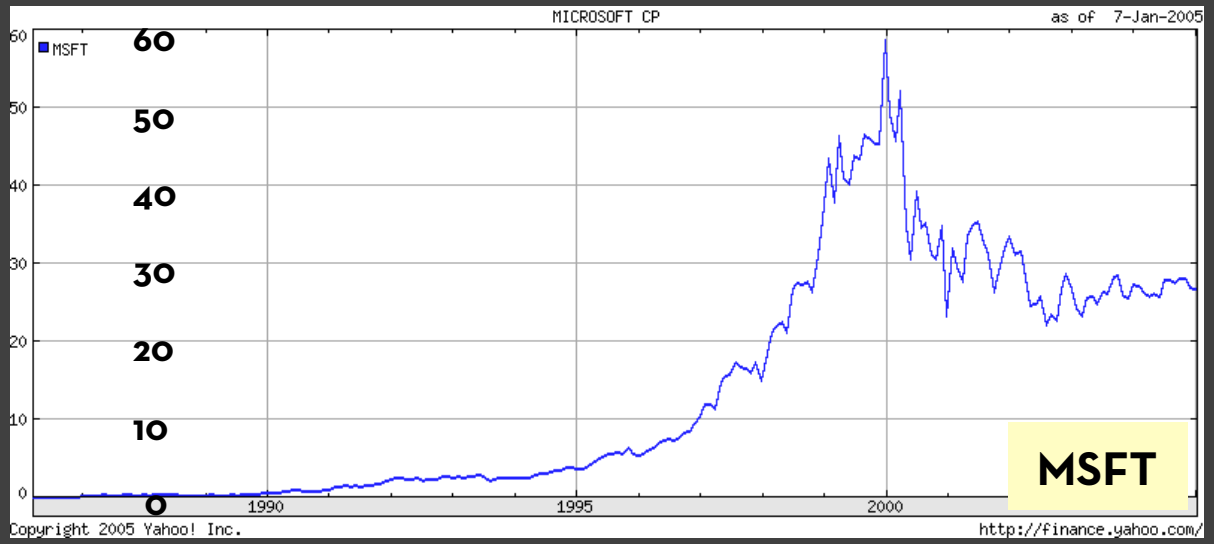
Axis Bounds

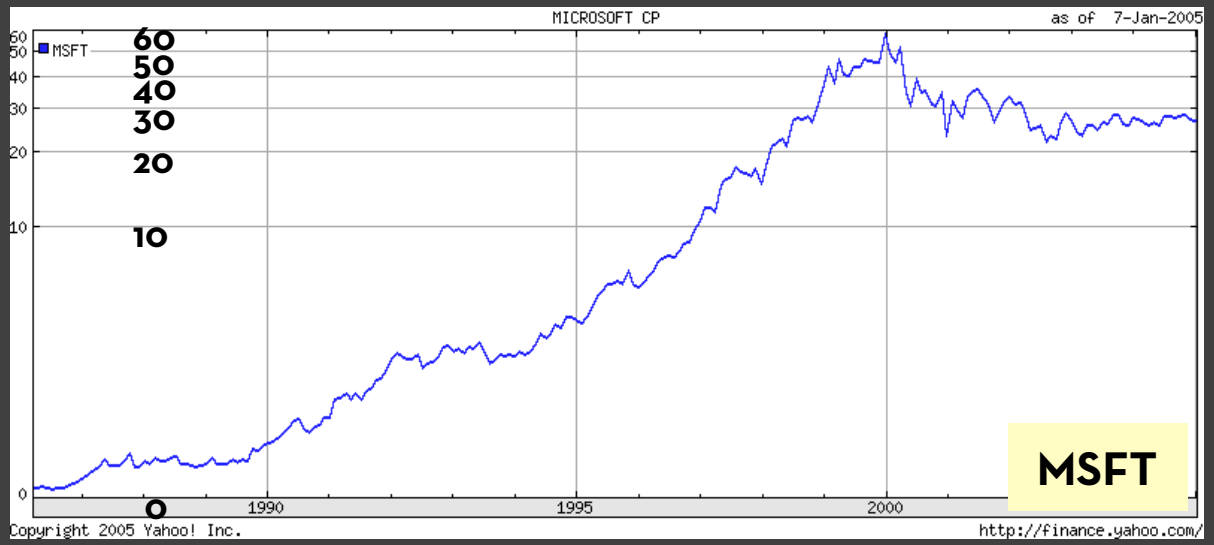
Zoom

Context

2. Transform Data

Linear scale vs. Log scale

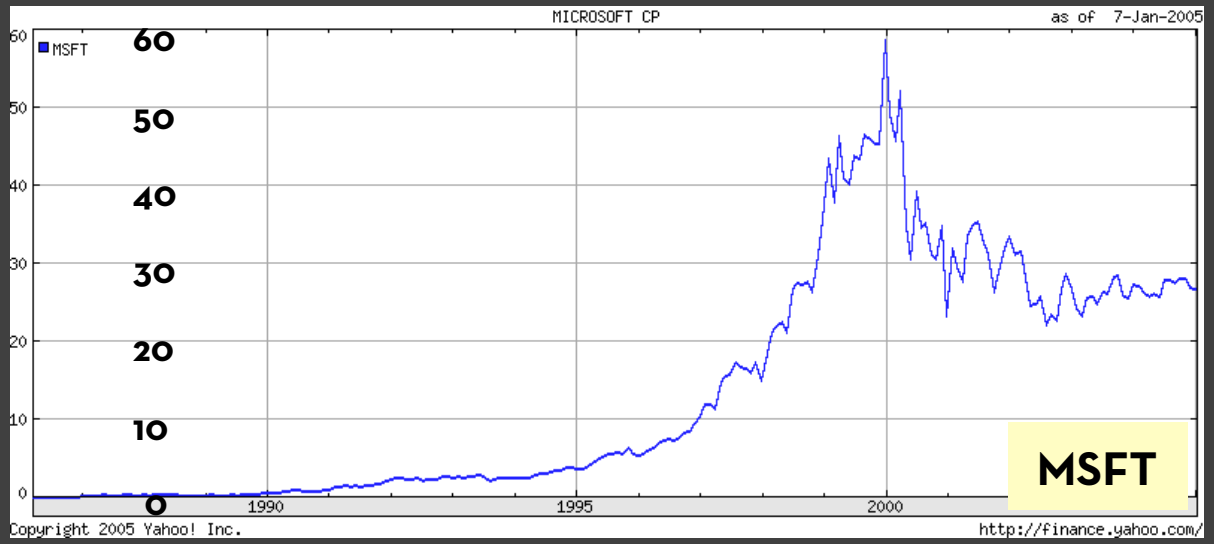




Linear scale vs. Log scale

Linear scale

Absolute change

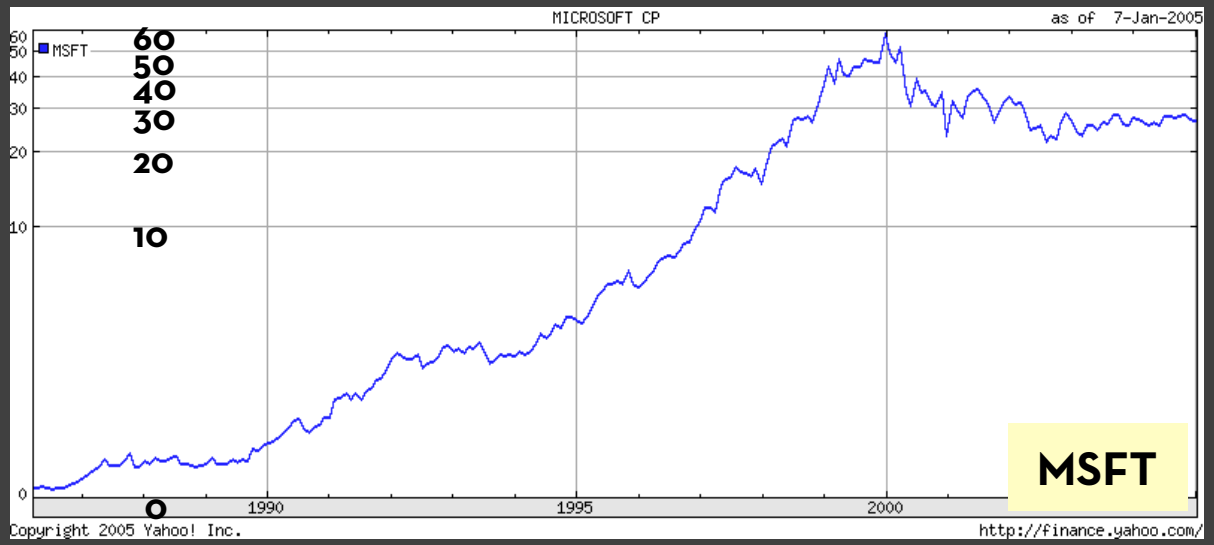


Log scale

Small fluctuations

Percent change

d(10,20) = d(30,60)

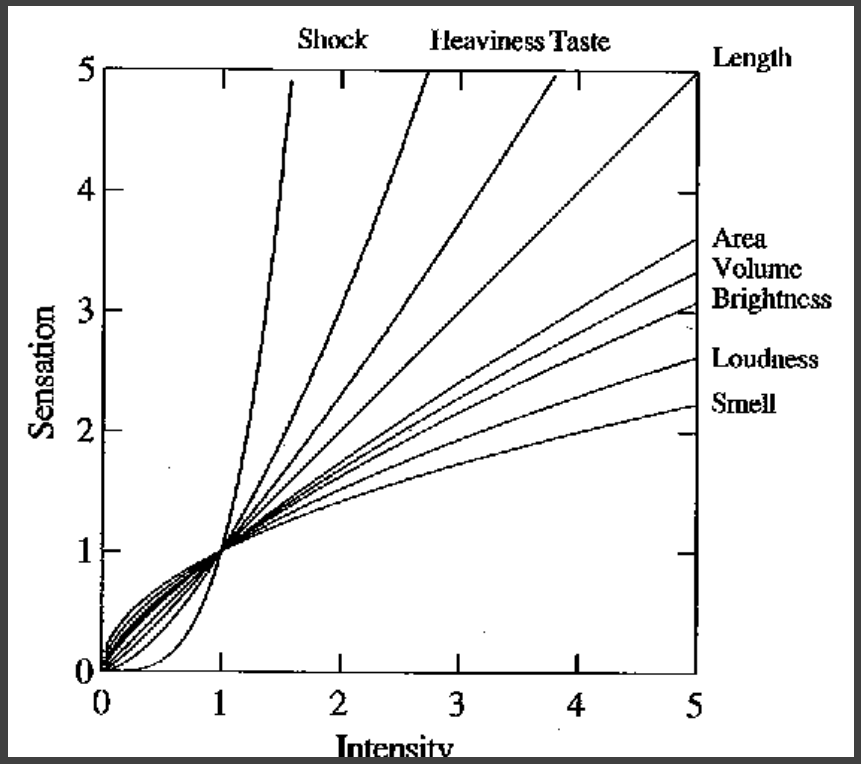


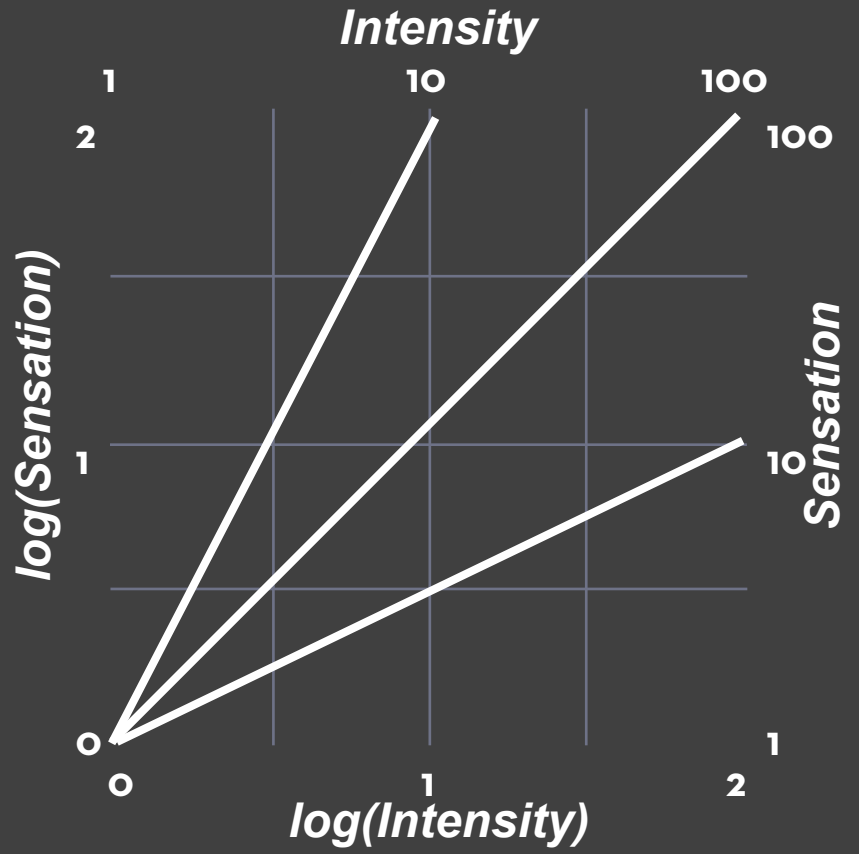
Log-Log graph

Power functions ($y = kx^a$) transform into lines

Example - Steven's power laws:

$$S = kI^p \rightarrow \log S = \log k + p \log I$$





Other non-linear scales

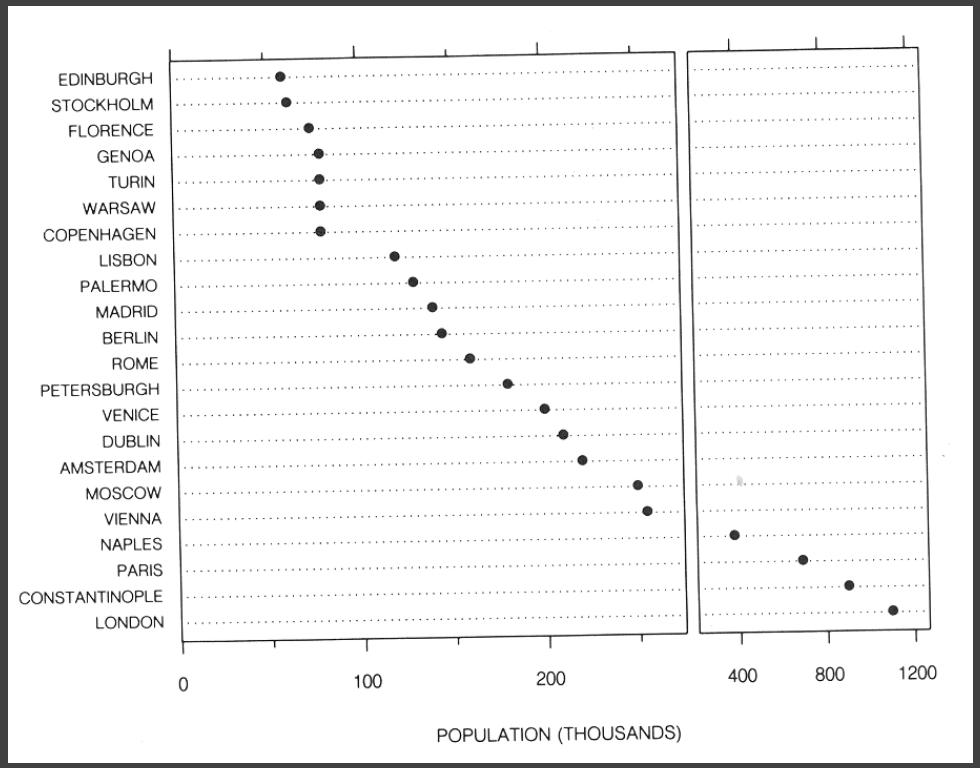
square root: area -> length

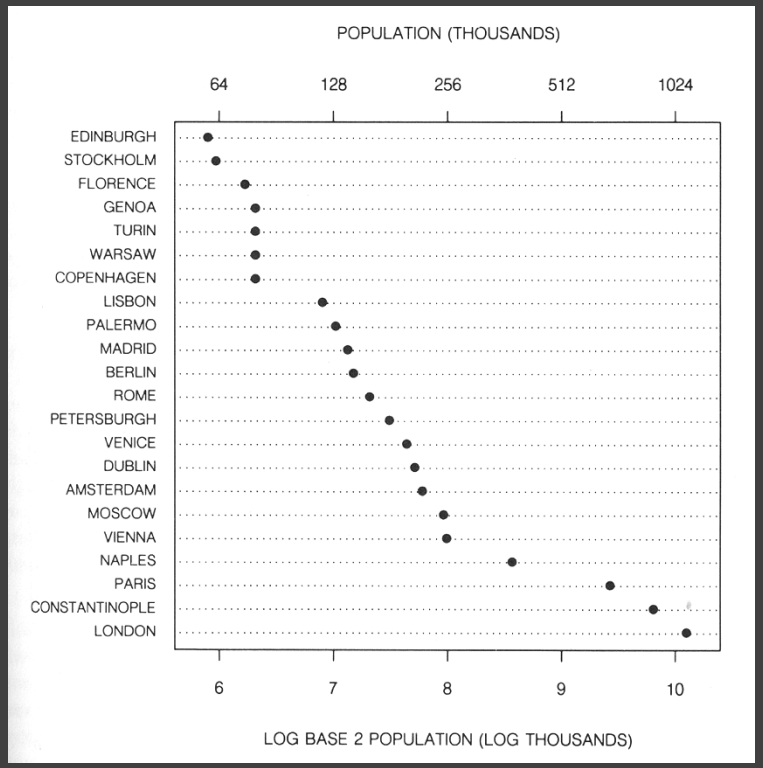
cube root: volume -> length

square: length -> area

cube: length -> volume

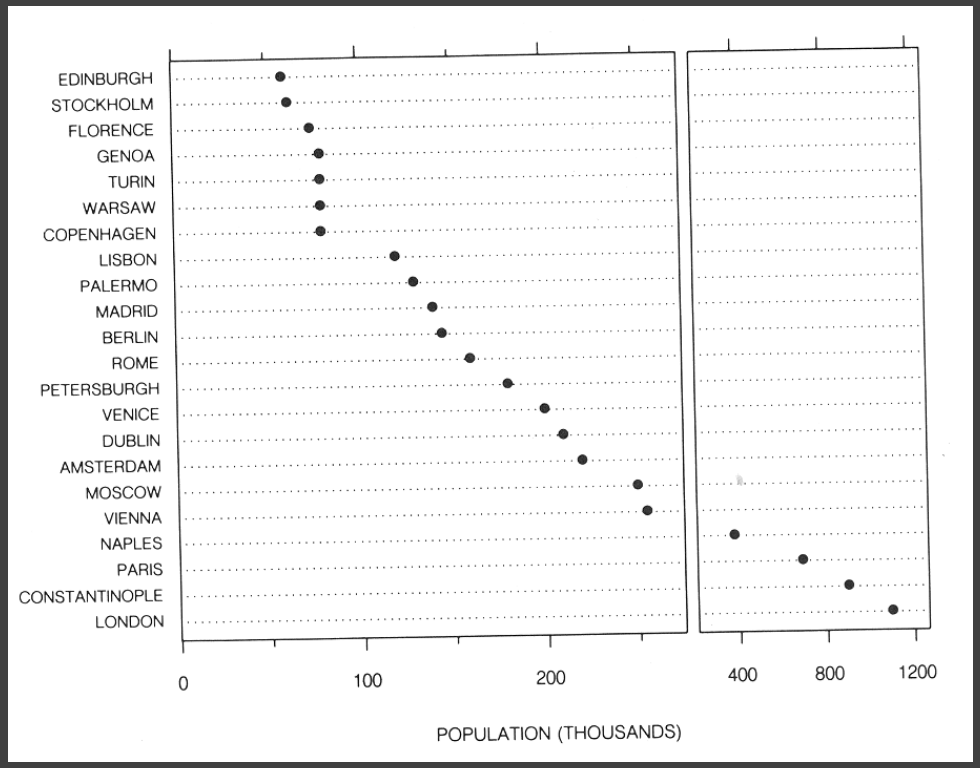
Compare: Scale break vs. log?

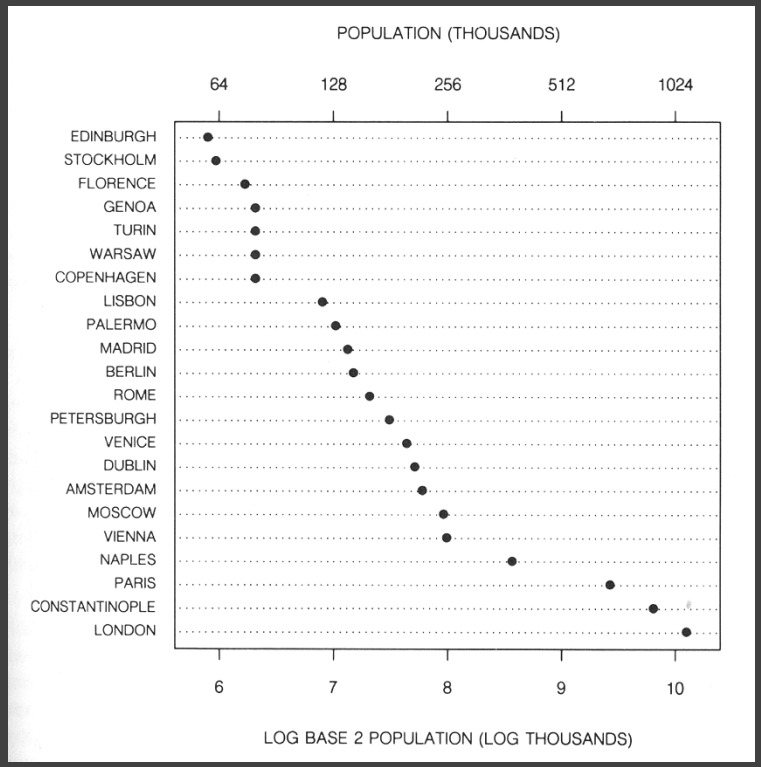




[Cleveland 85]

Compare: Scale break vs. log?





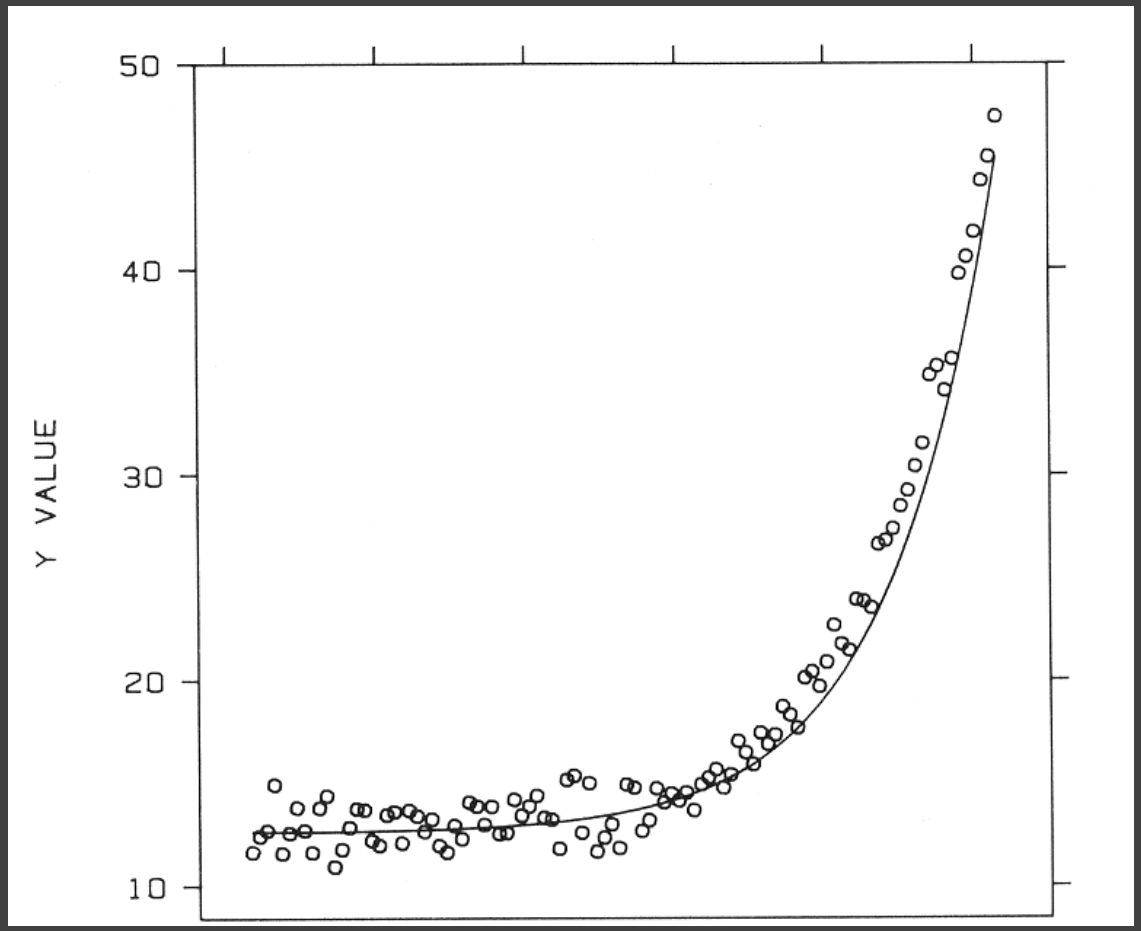
Both increase visual resolution

Log scale - easy comparisons of all data

Scale break – more difficult to compare across break

Residuals

How well does curve fit data?

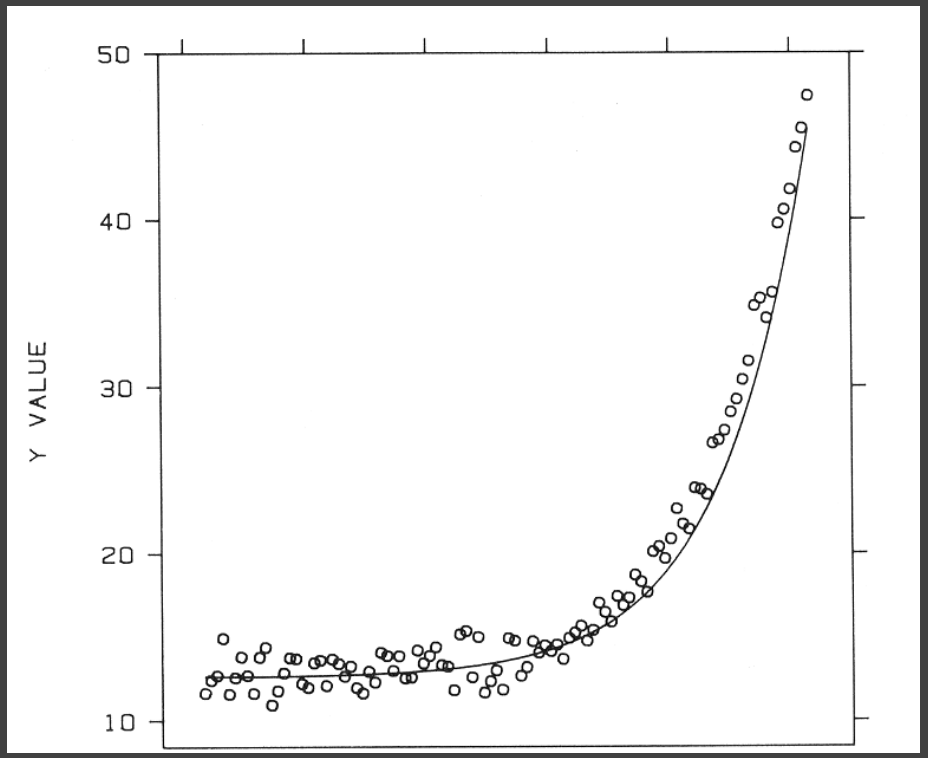


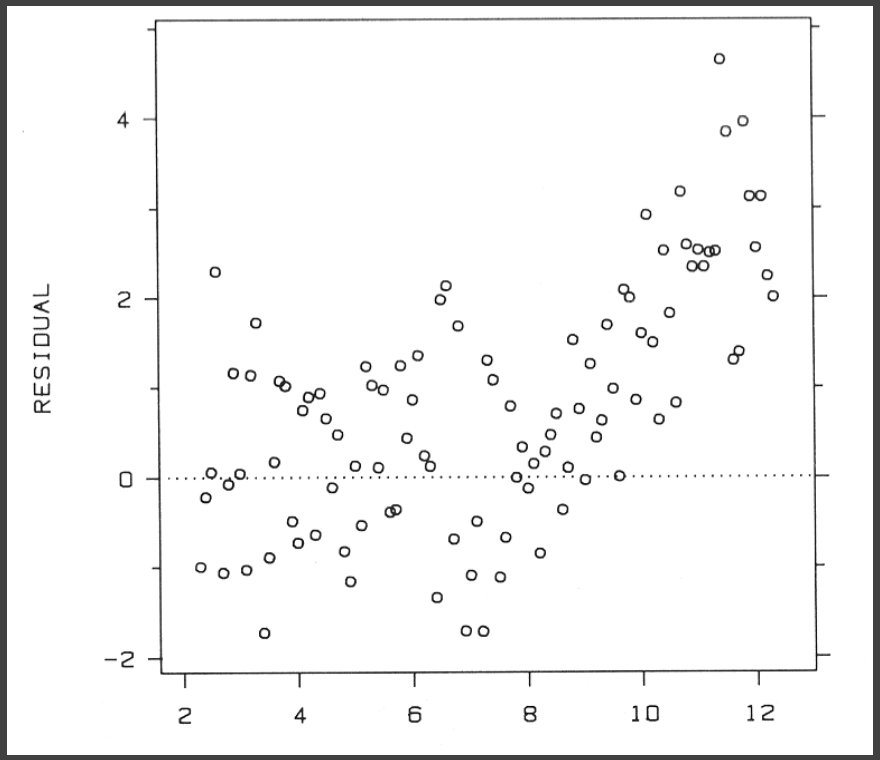
[Cleveland 85]

Residuals

Plot vertical distance from best fit curve

Residual graph shows accuracy of fit





[Cleveland 85]

Residuals

Compare data to expected/baseline

Transform Data?

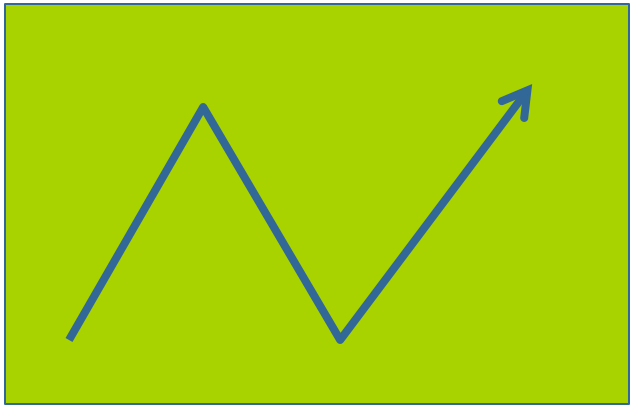
Non-linear scales

Residuals

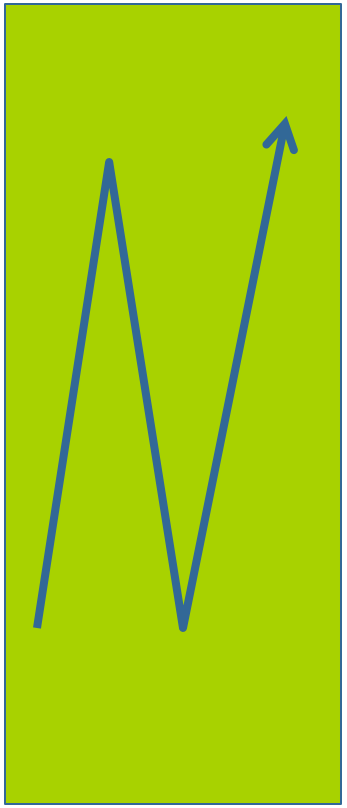
3. Transform Viz

Aspect Ratio Selection

My Start Up

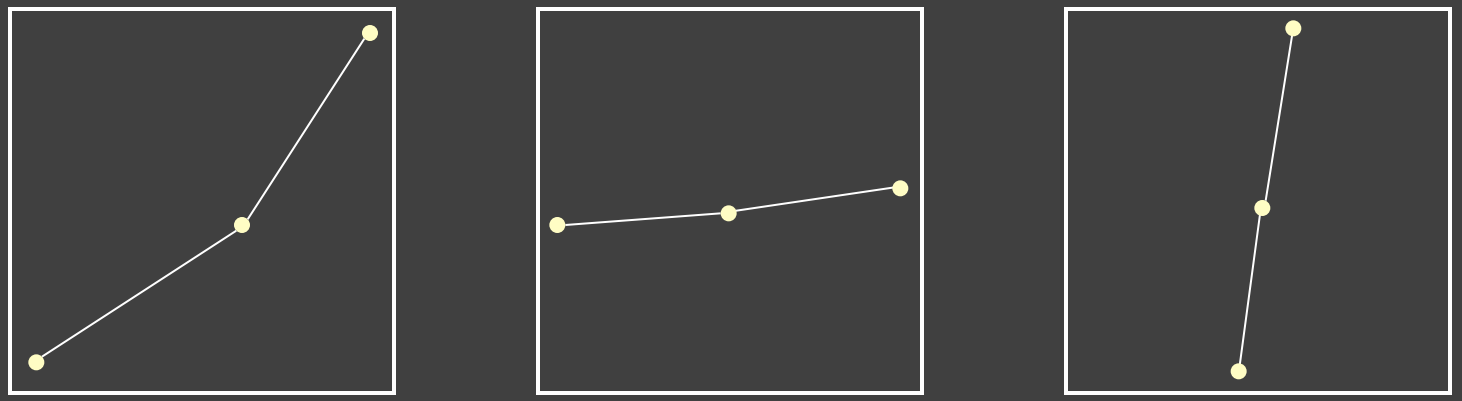


My Start Up



Banking to 45° [Cleveland]

To facilitate perception of trends, maximize the discriminability of line segment orientations



Two line segments are maximally discriminable when their average absolute angle is 45°

Optimize the *aspect ratio* to bank to 45°

Aspect-Ratio Banking Techniques

Has Closed Form Solution

Median-Absolute-Slope

$$\alpha = \text{median} \mid s_i \mid R_x / R_y$$

Average-Absolute-Slope

$$\alpha = \text{mean} \mid s_i \mid R_x / R_y$$

Average-Absolute-Orientation

Unweighted

$$\sum_i \frac{\mid \theta_i(\alpha) \mid}{n} = 45°$$

Weighted

$$\frac{\sum_i \mid \theta_i(\alpha) \mid l_i(\alpha)}{\sum_i l_i(\alpha)} = 45°$$

Max-Orientation-Resolution

Global (over all i, j s.t. i≠j)

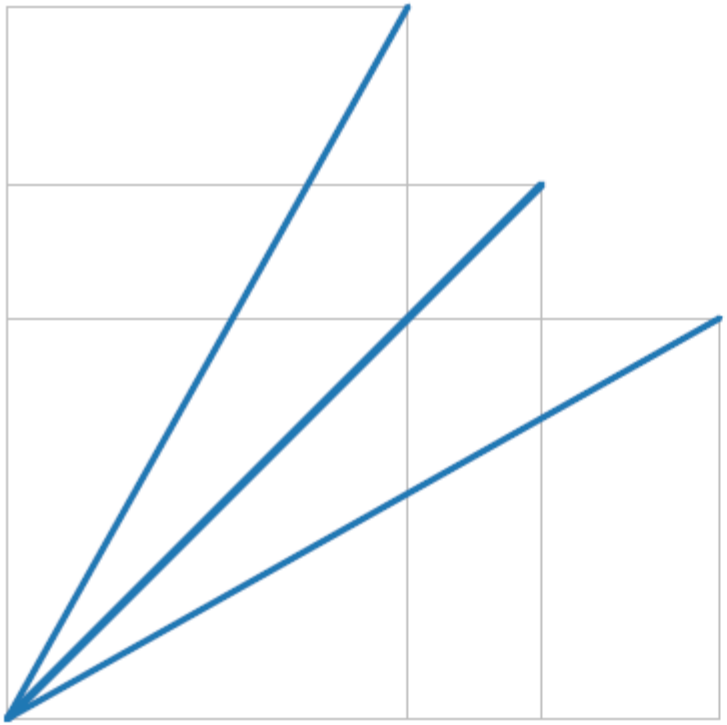
$$\sum_i \sum_j \mid \theta_i(\alpha) - \theta_j(\alpha) \mid^2$$

Local (over adjacent segments)

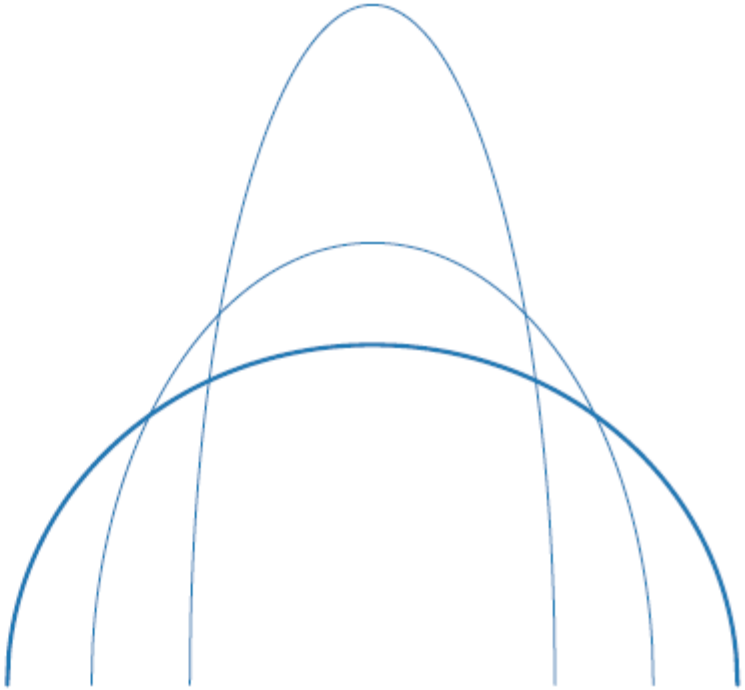
$$\sum_i \mid \theta_i(\alpha) - \theta_{i+1}(\alpha) \mid^2$$

Requires Iterative Optimization

An alternate approach:
Minimize arc length (hold area constant)



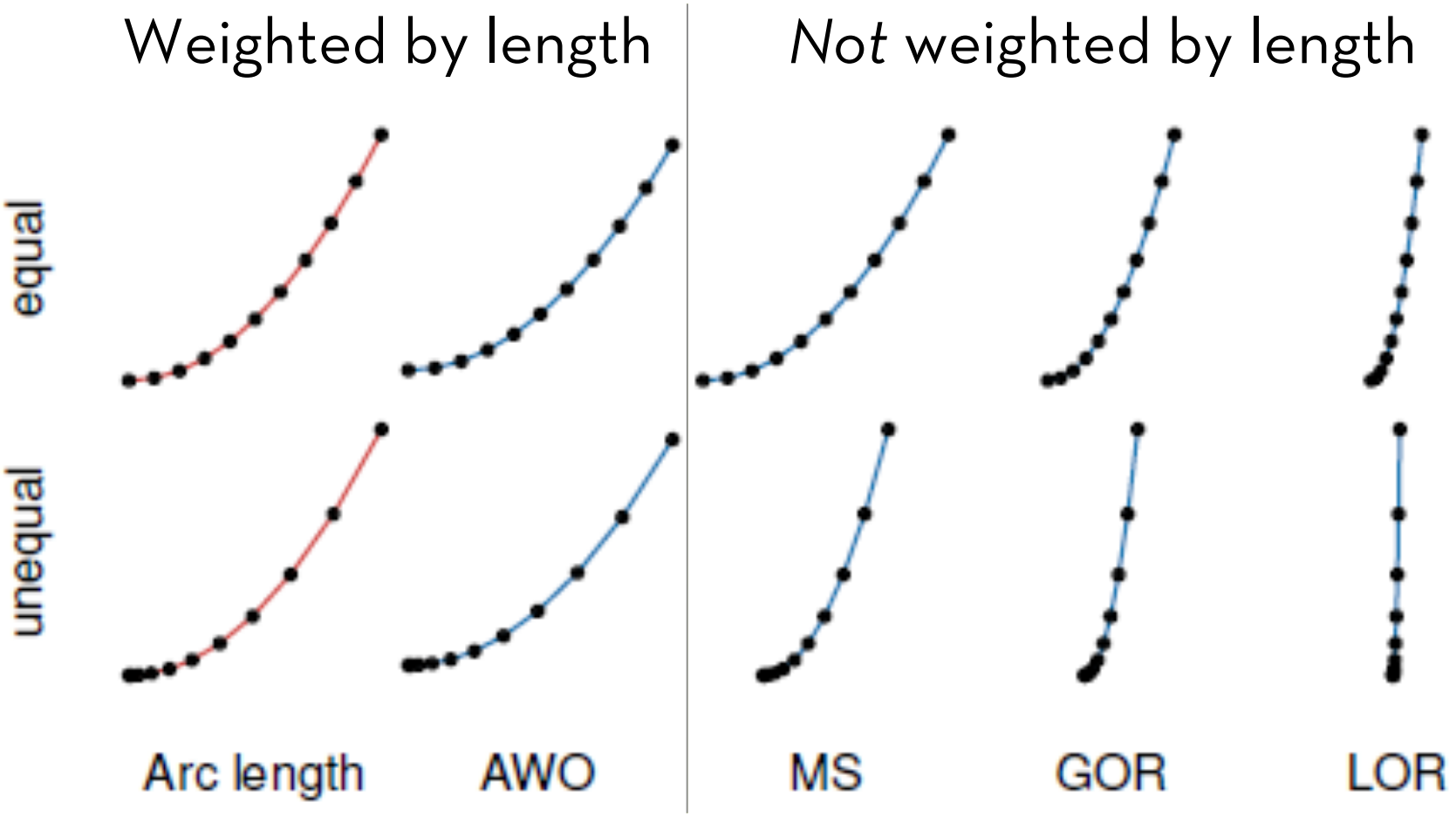
Straight line -> 45 deg



Ellipse -> Circle

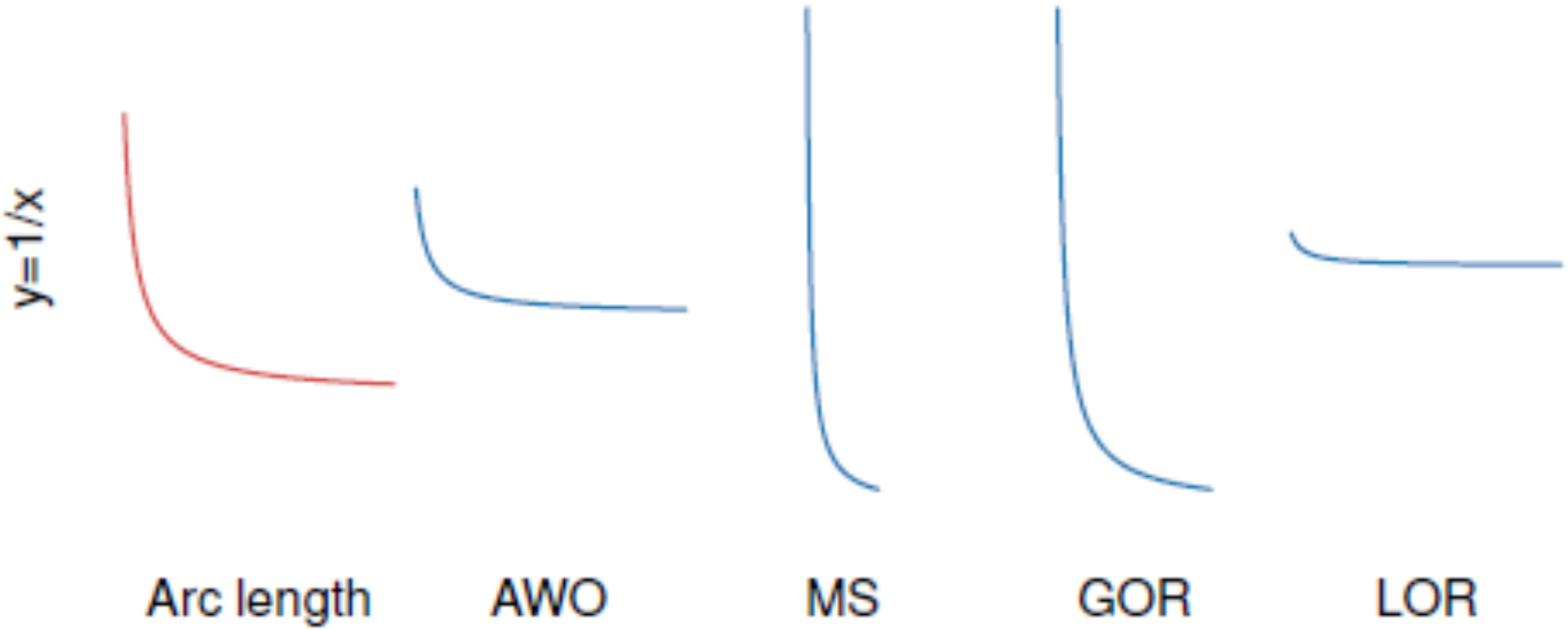
[Talbot et al, 2011]

Parameterization invariance

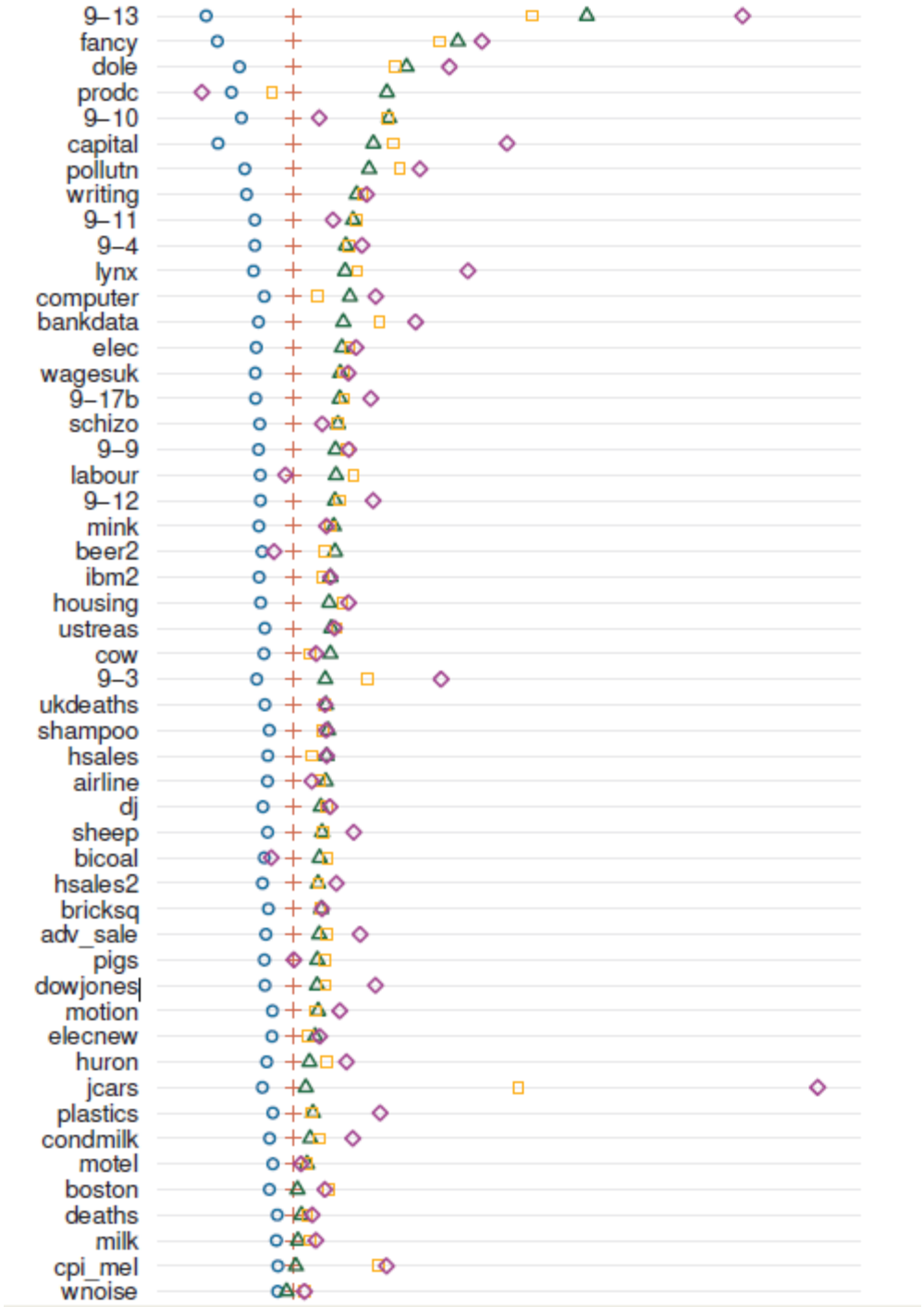


[Talbot et al, 2011]

Robustness: banking y = 1 / x



[Talbot et al, 2011]



Compromise

Arc-length banking produces aspect ratios in-between those produced by other methods.

[Talbot et al, 2011]

Discussion

Discussion

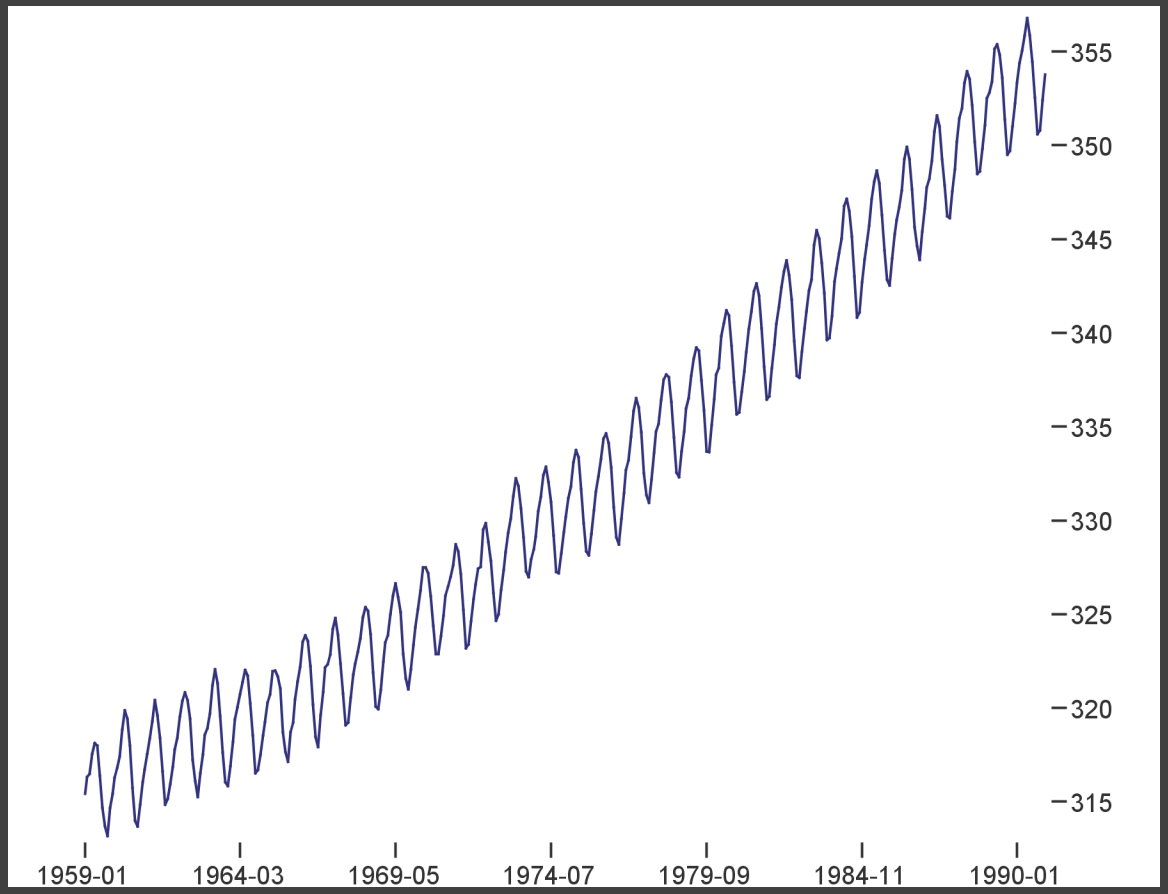
Arc-length banking preferable to prior methods

- Parameterization invariant
- Robust (handles corner cases, gives compromise results)
- Applicable to both plotted curves and contour lines
- Fast to compute (fast-converging iterative optimization)

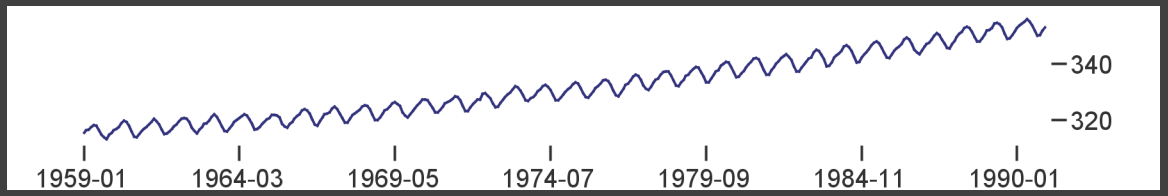
But what about perceptual effectiveness?

We lack theory to deeply motivate aspect ratio selection

Perceptual experiments needed to assess?



Aspect Ratio = 1.17



Aspect Ratio = 7.87

CO_2 Measurements

William S. Cleveland
Visualizing Data

Multi-Scale Banking

Goal

- Optimized aspect ratios for varying data scales

Approach

- Identify Scales of Interest
- Generate Scale-Specific Trend Lines
- Bank Trend Lines to Optimized Aspect Ratios
- Filter Redundant Aspect Ratios

Multi-Scale Banking [Heer 2006]

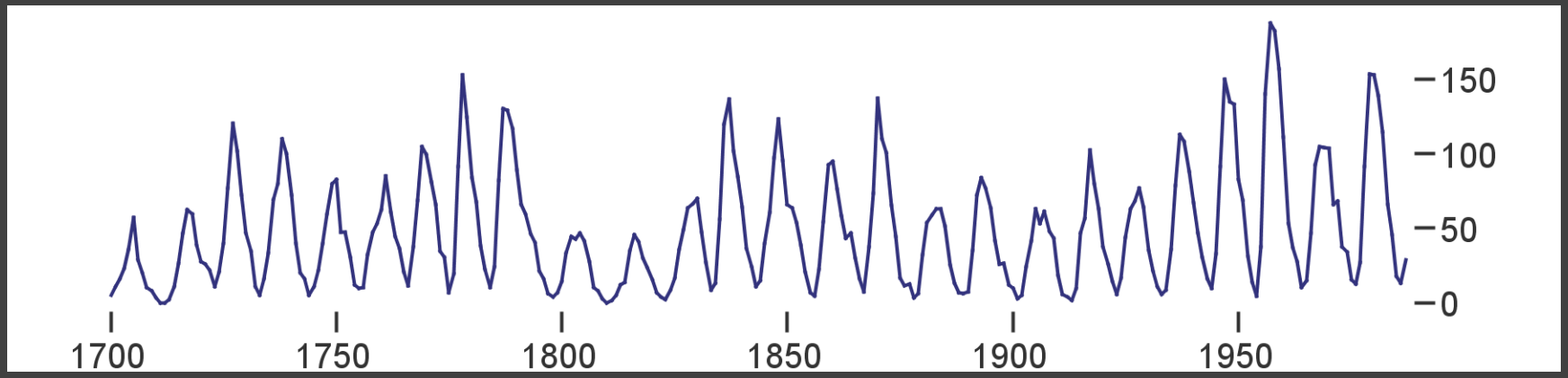
Idea: Use Spectral Analysis to identify trends

- Find strong frequency components
- Lowpass filter to create trend lines

Other trend-identification methods?

- Regression methods
- Wavelet analysis

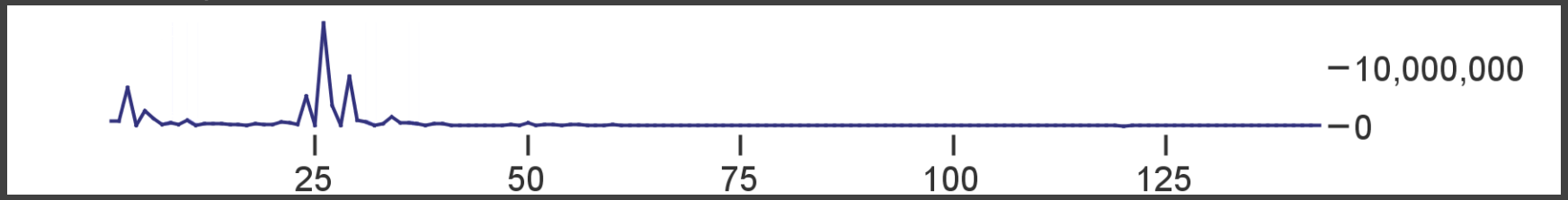
Compute Power Spectrum



Take Discrete Fourier Transform

Compute squared magnitudes

Power Spectrum

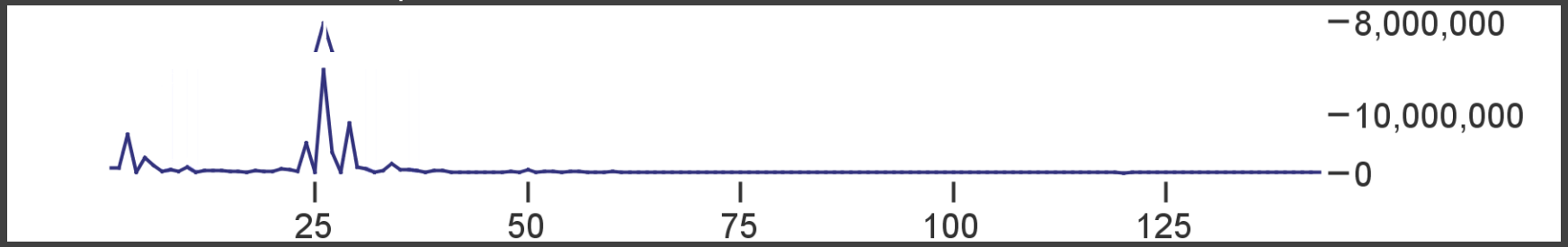


Smooth the Spectrum

Convolve with Gaussian filter

window size = 3, $\sigma = 1$

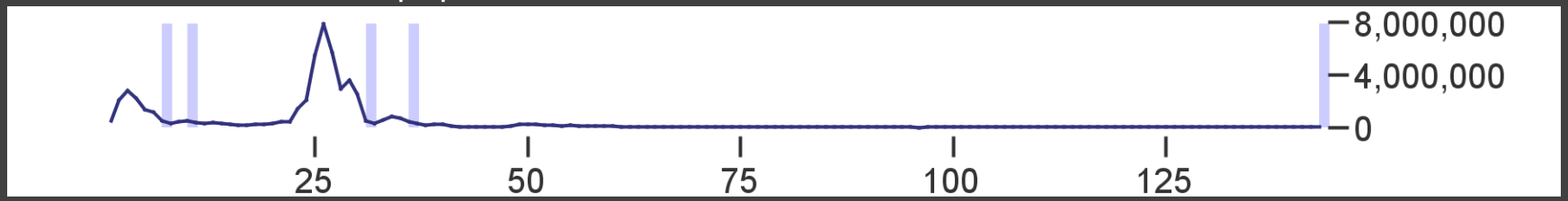
Smoothed Power Spectrum



Threshold the Spectrum

Threshold at $\mu + k\sigma$ (k=0 by default)

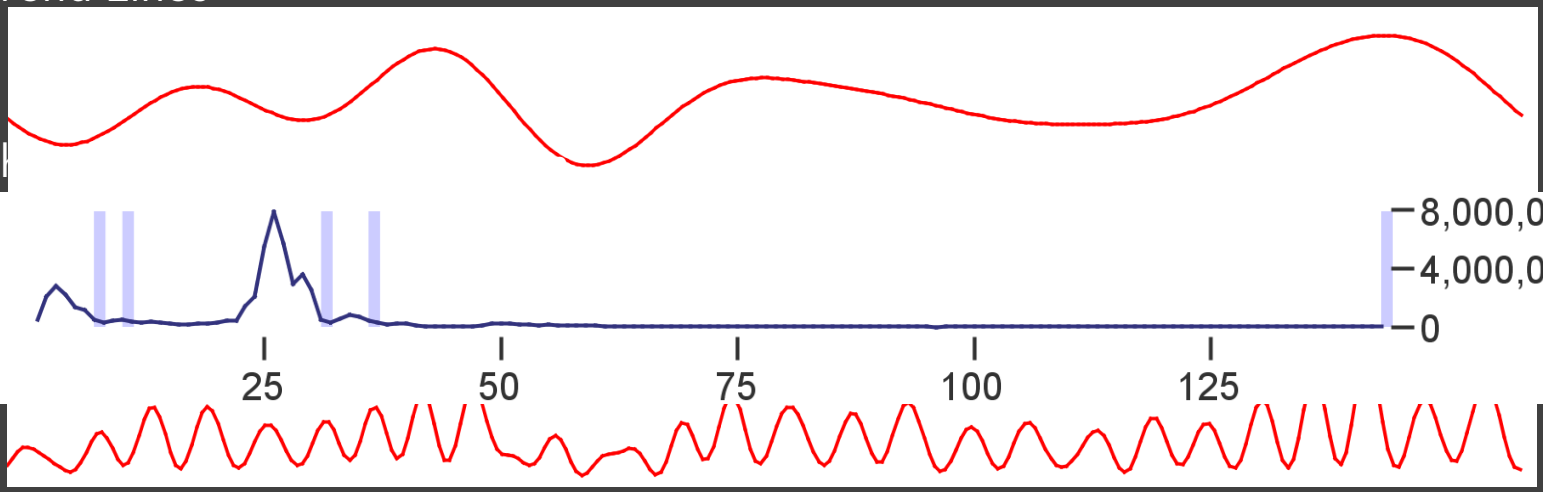
Retain last values of contiguous runs



Generate Trend Lines

Generate trend lines with lowpass filter

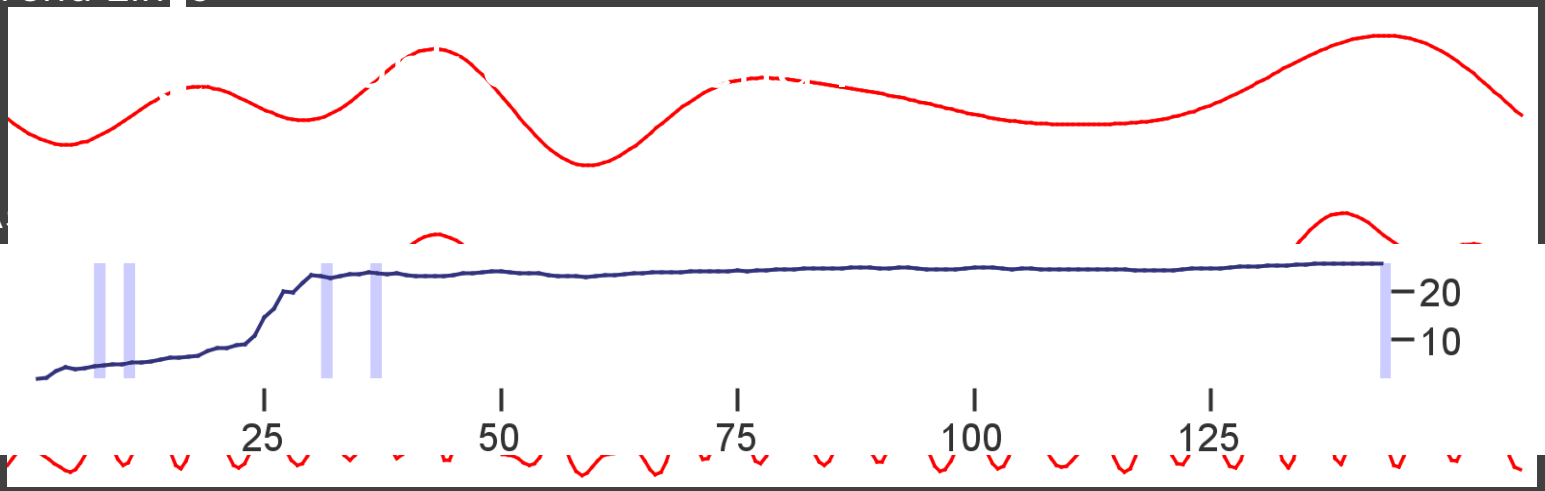
Trend Lines



Bank Trend Lines to 45°

Trend Lines

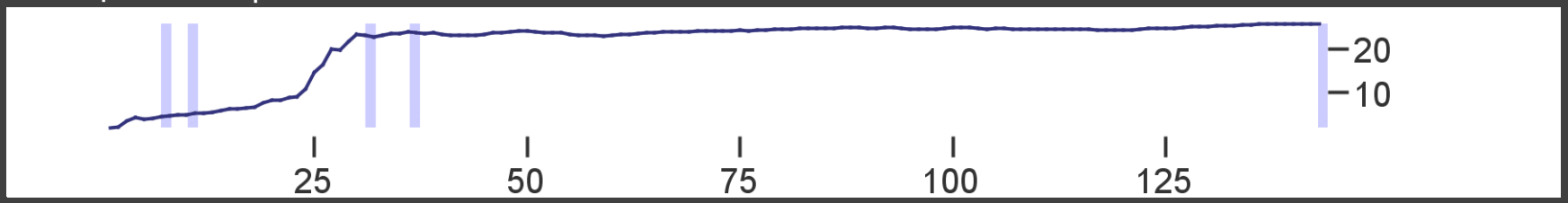
A



Filter Aspect Ratios

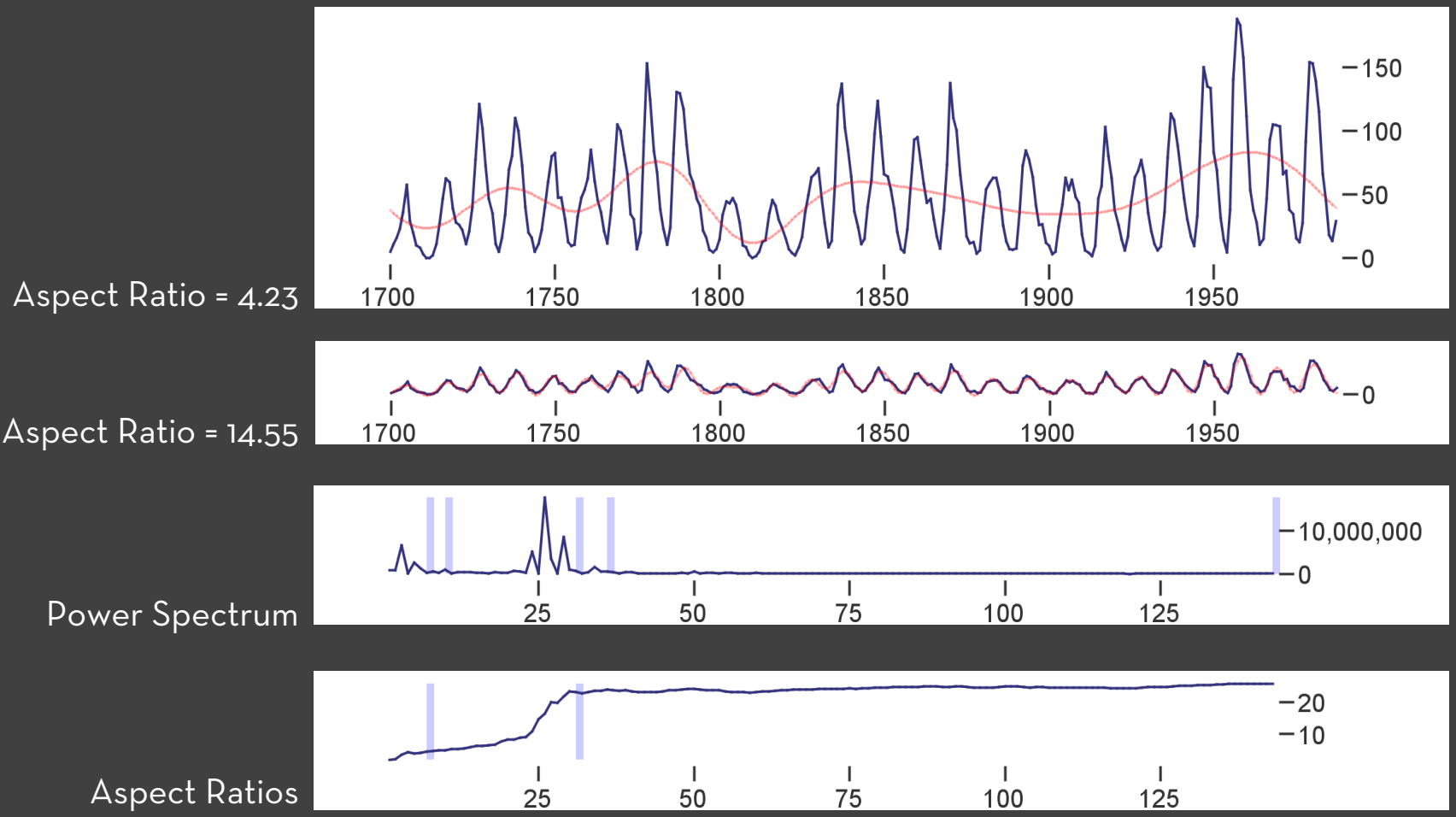
Filter similar aspect ratios

Keep if $\alpha_{i+1} > c\alpha_i$ (c=1.25 by default)



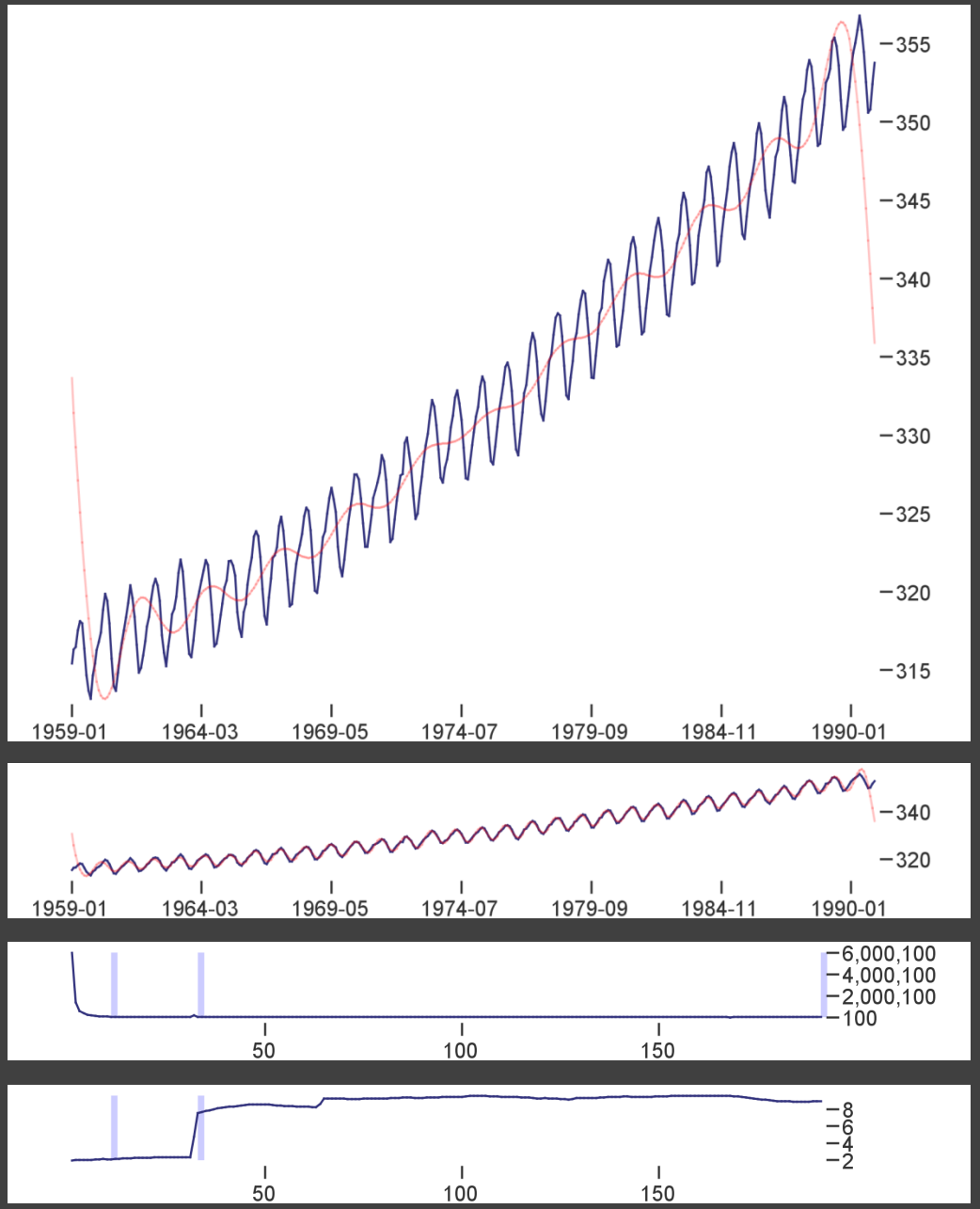
Sunspot Cycles

Yearly values 1700-1987

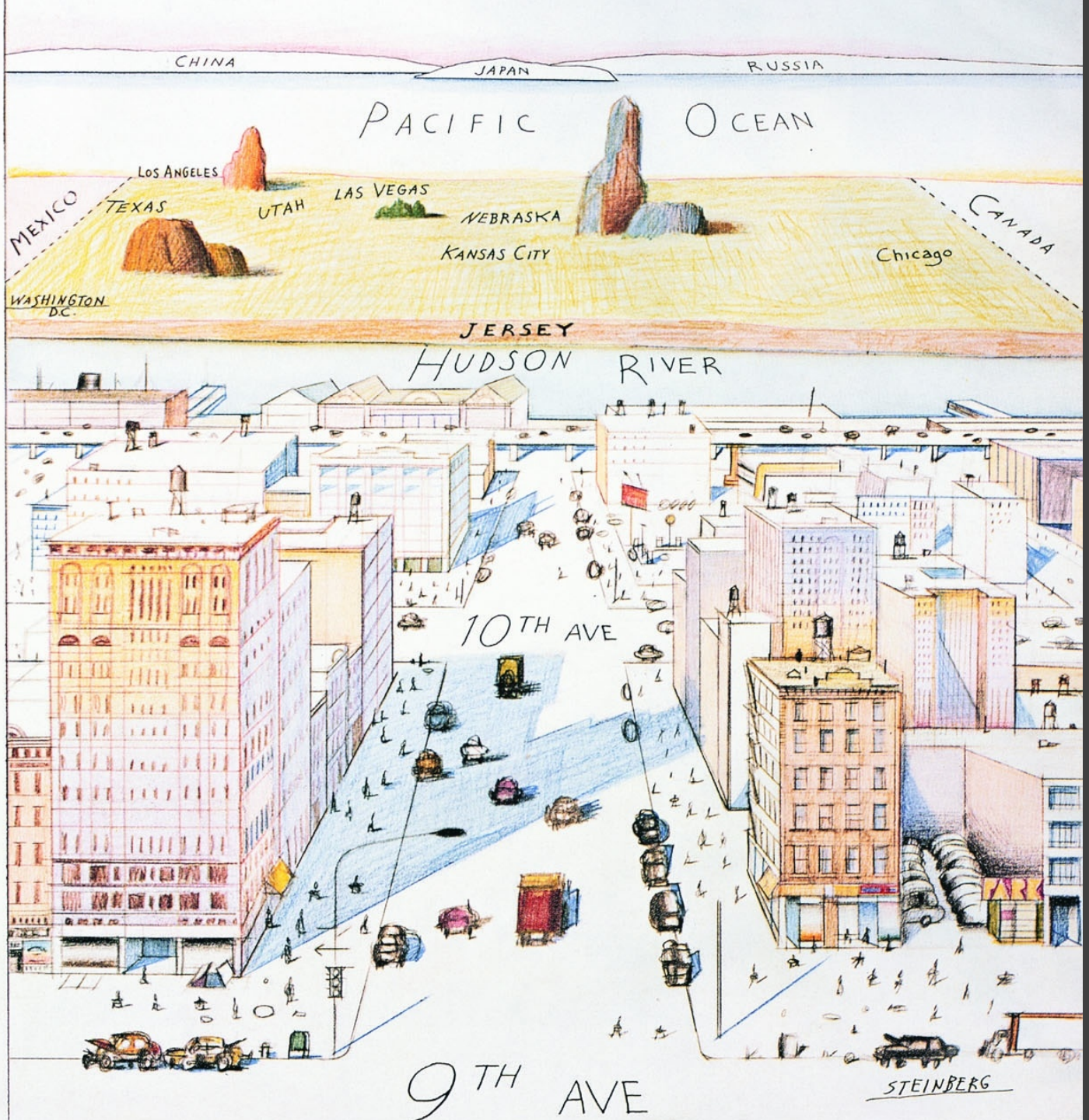


CO_2

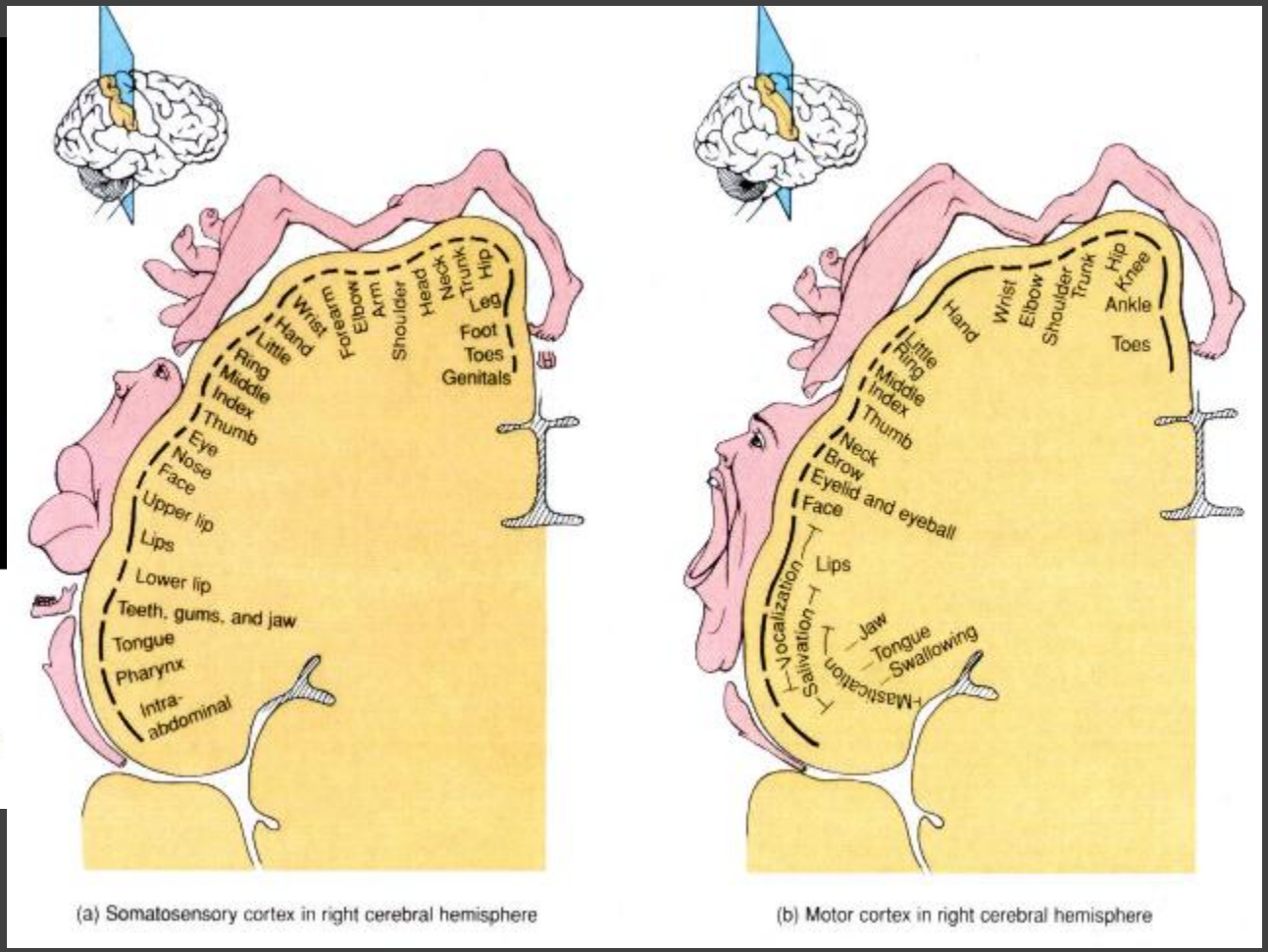
Monthly concentrations from the Mauna Loa Observatory, 1950-1990



Distortion


CHINA
JAPAN
RUSSIA
PACIFIC OCEAN
LOS ANGELES
MEXICO
TEXAS
UTAH
LAS VEGAS
NEBRASKA
KANSAS CITY
CANADA
Chicago
WASHINGTON DC.
JERSEY
HUDSON RIVER
10TH AVE
9TH AVE
PARK
STEINBERG

This model shows what a man's body would look like if each part grew in proportion to the area of the cortex of the brain concerned with its sensory perception. The hands and lips dominate — but the feet are also disproportionately large, indicating their sensory importance.



(a) Somatosensory cortex in right cerebral hemisphere

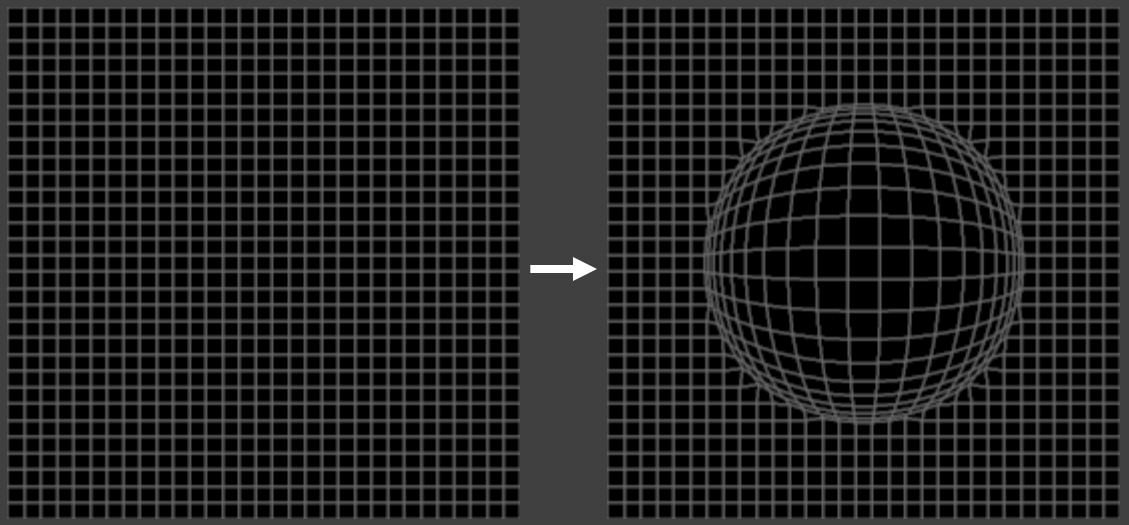
(b) Motor cortex in right cerebral hemisphere

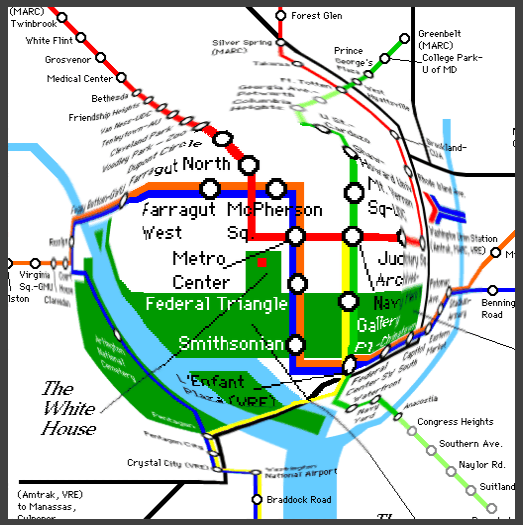
Single focus distortion

Focus area – local details

De-magnified area – surrounding context

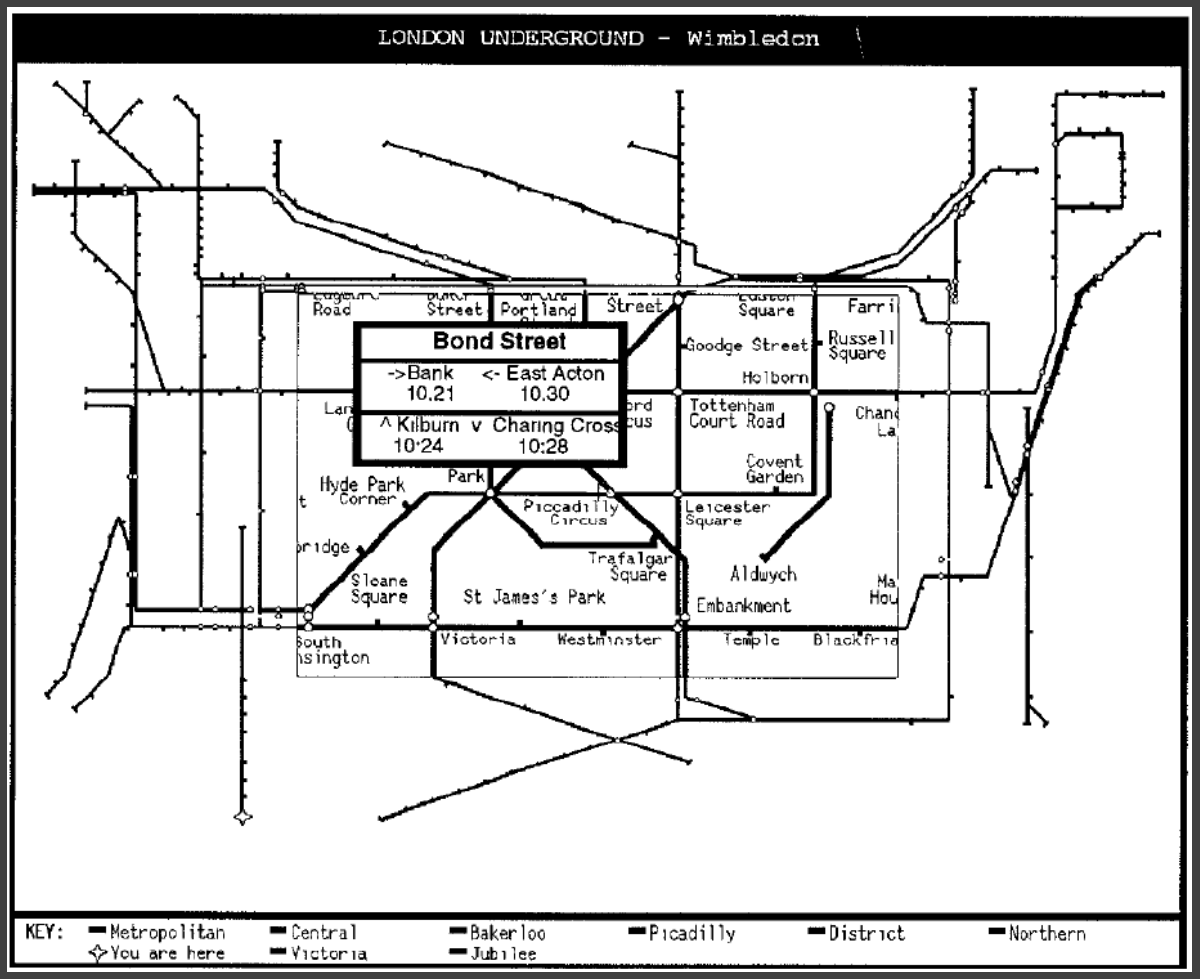
Like a rubber sheet with borders tacked down

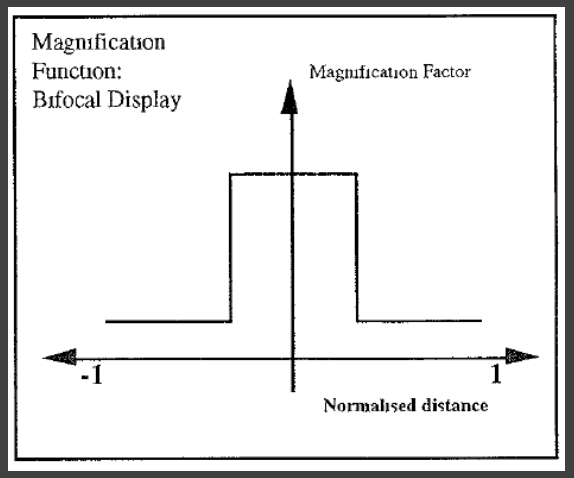


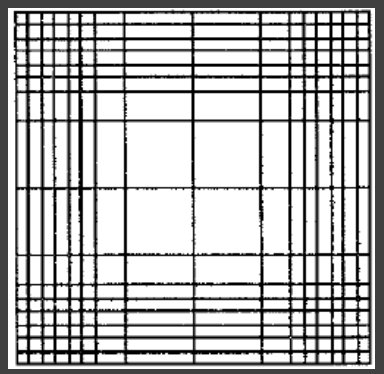


Nonlinear Magnification Infocenter [http://www.cs.indiana.edu/~tkeahey/research/nlm/nlm.html]

Bifocal display [Leung and Apperley 94]

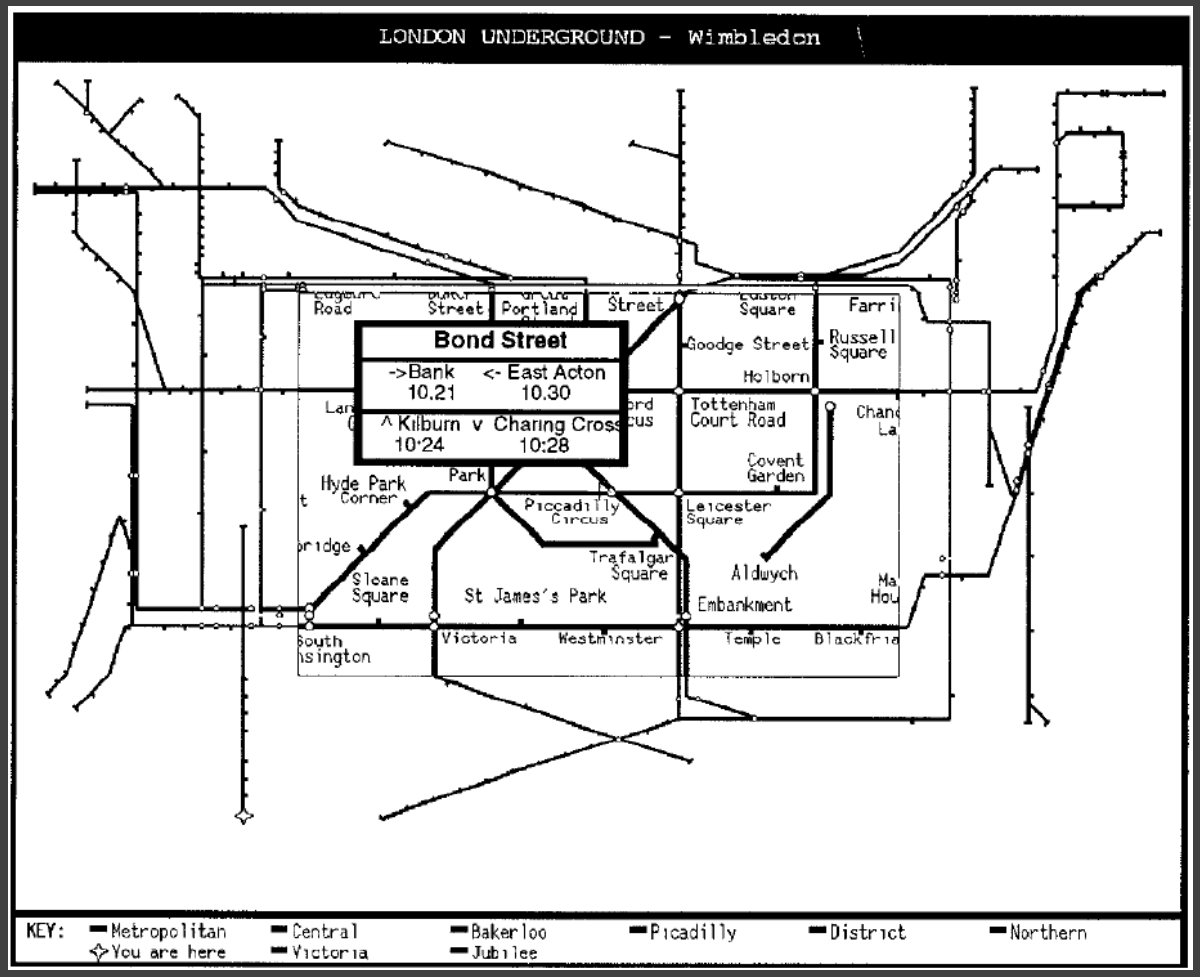


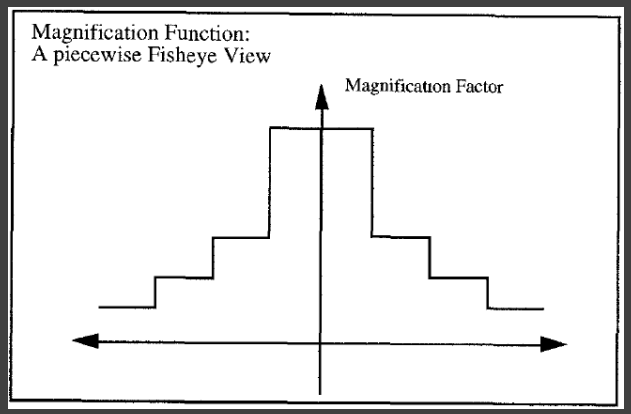




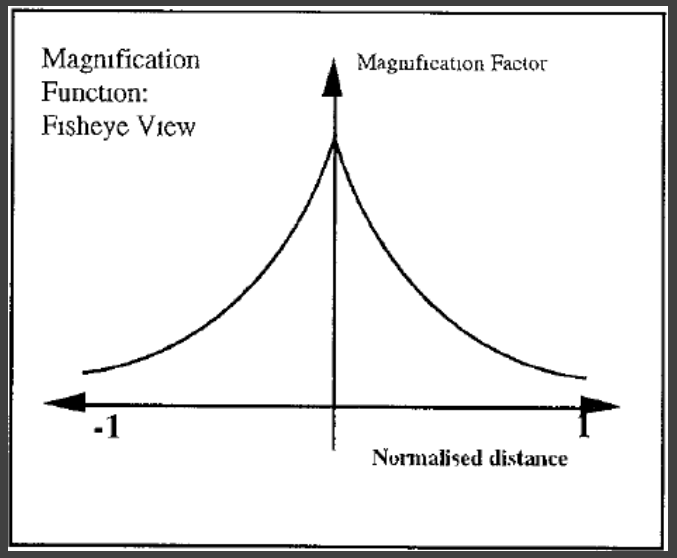
2D distortion

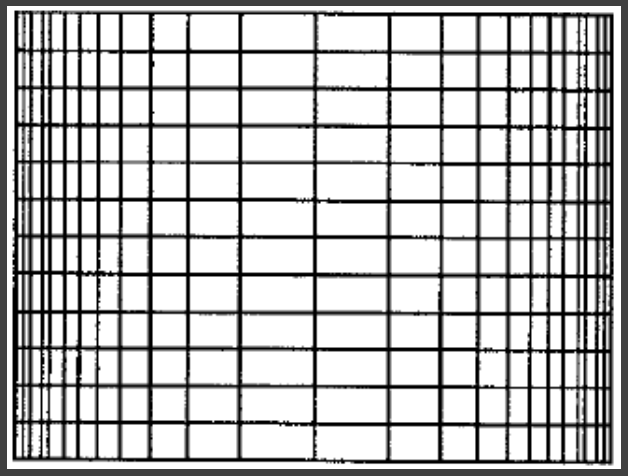
Multifocal display [Leung and Apperley 94]



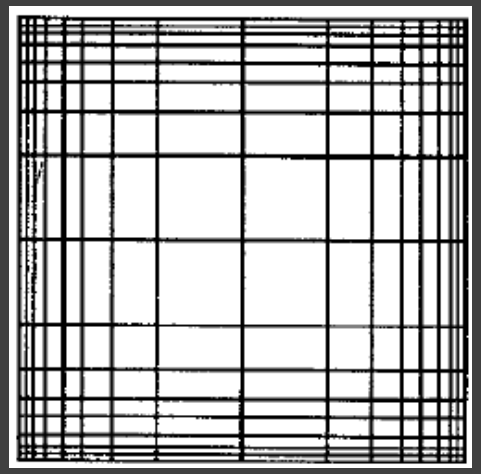


Fisheye [Leung and Apperley 94]

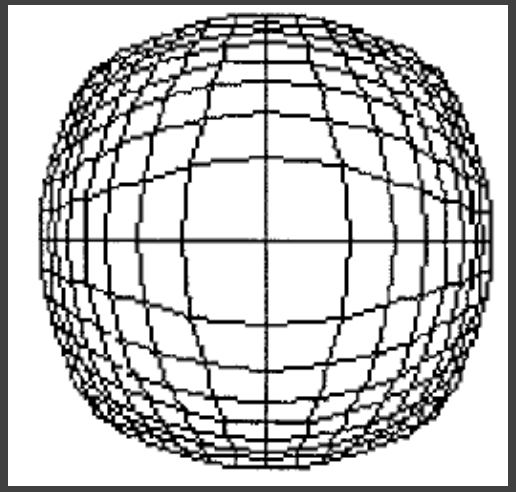




1D

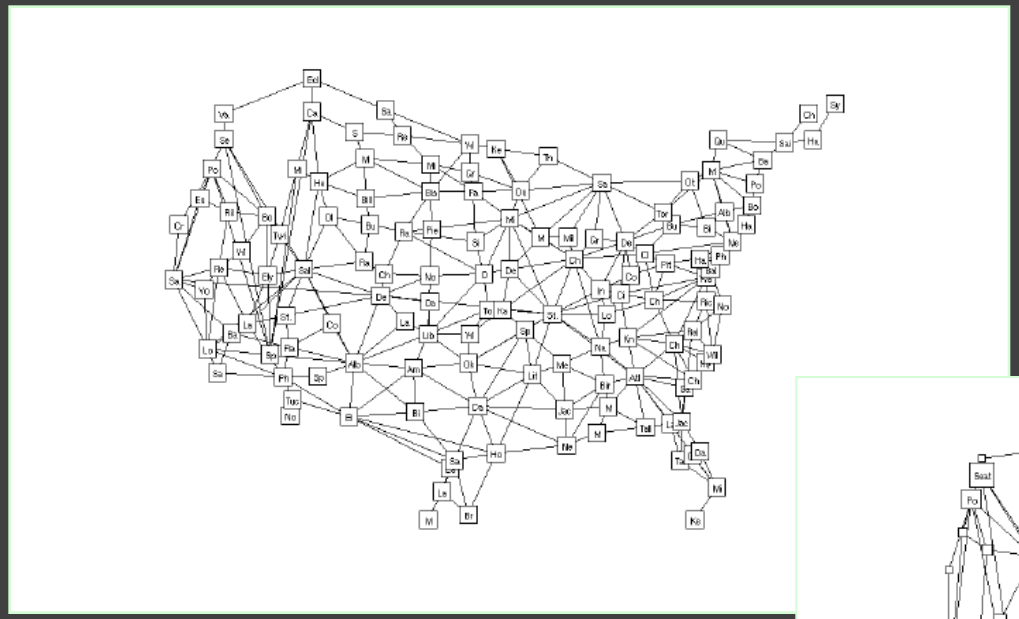


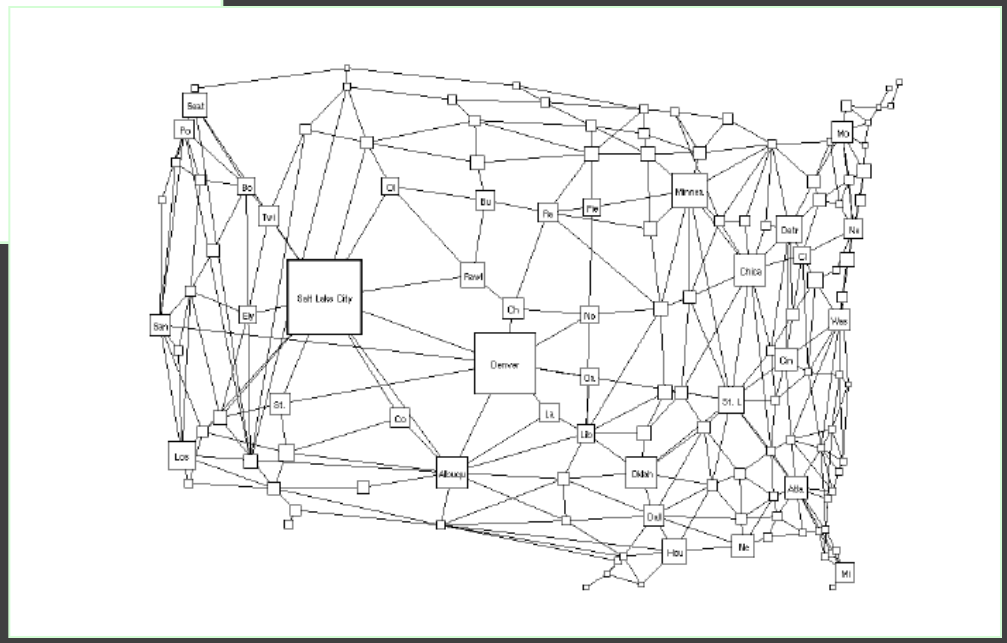
2D



Polar

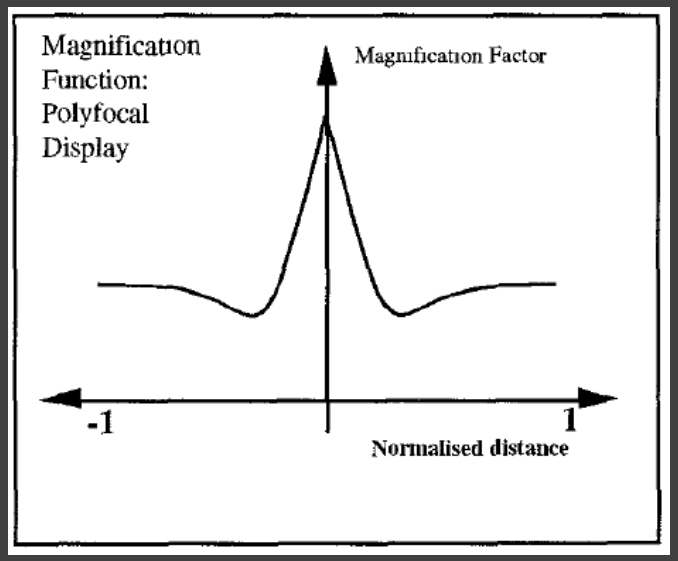
Fisheye graph

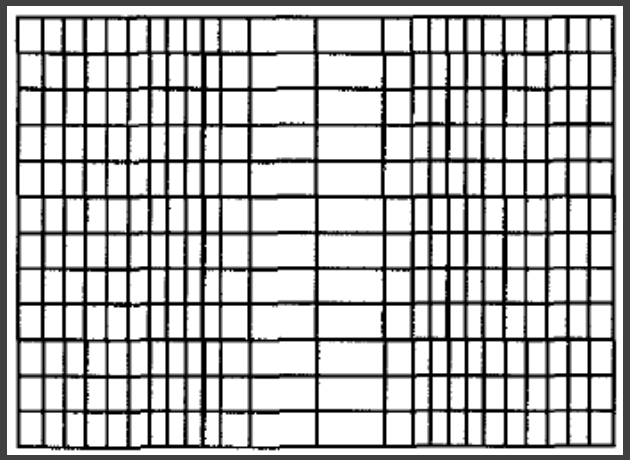




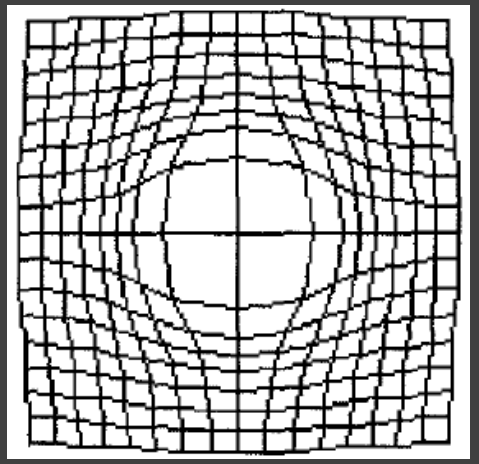
Graphical fisheye views of graphs [Sarkar & Brown 92]

Nonlinear magnification [Leung and Apperley 94]

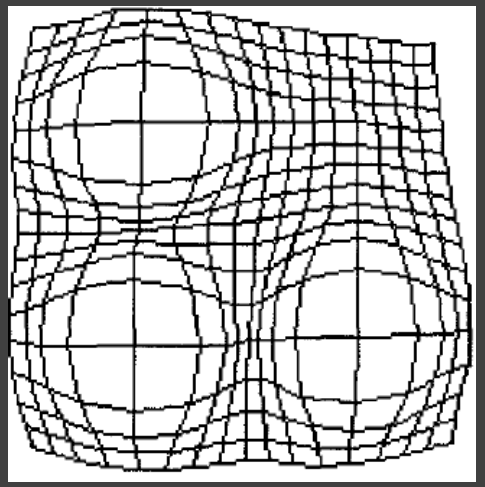


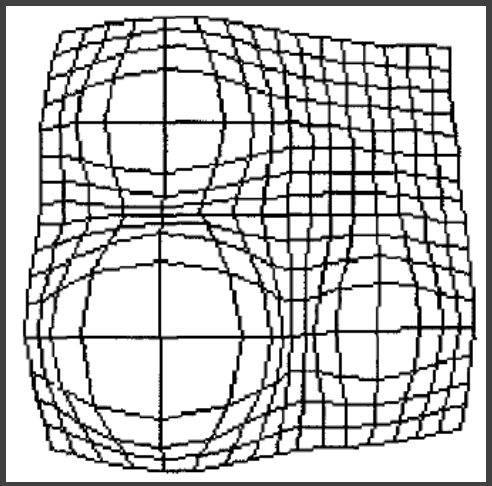


1D



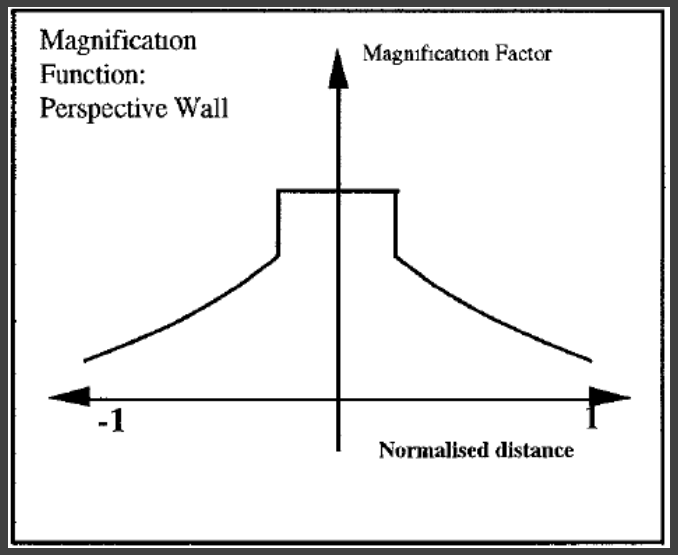
2D

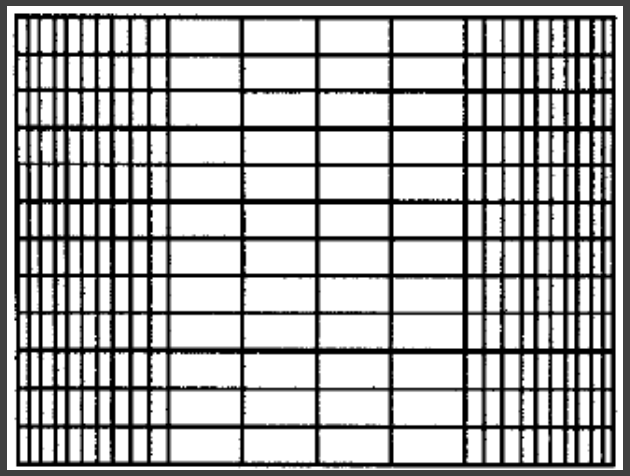


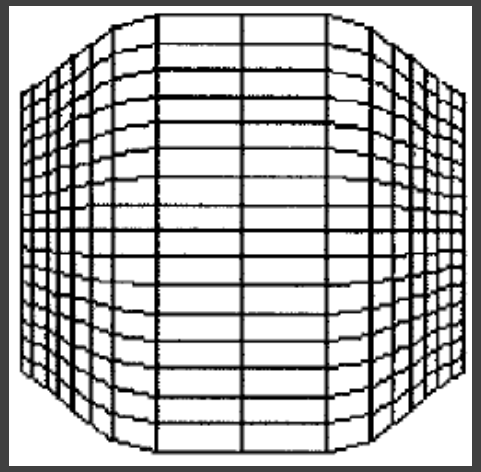


Multifocal

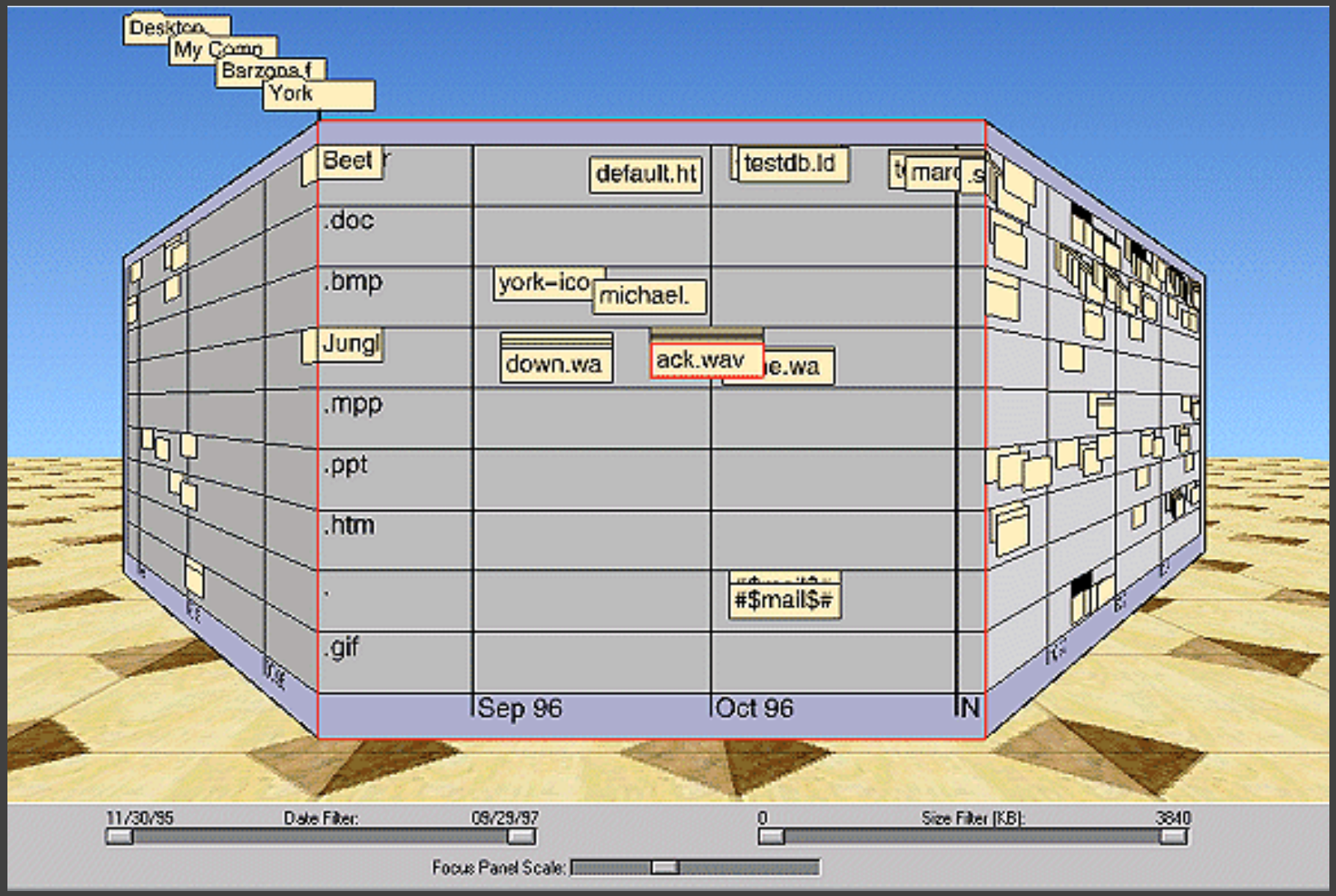
Perspective wall





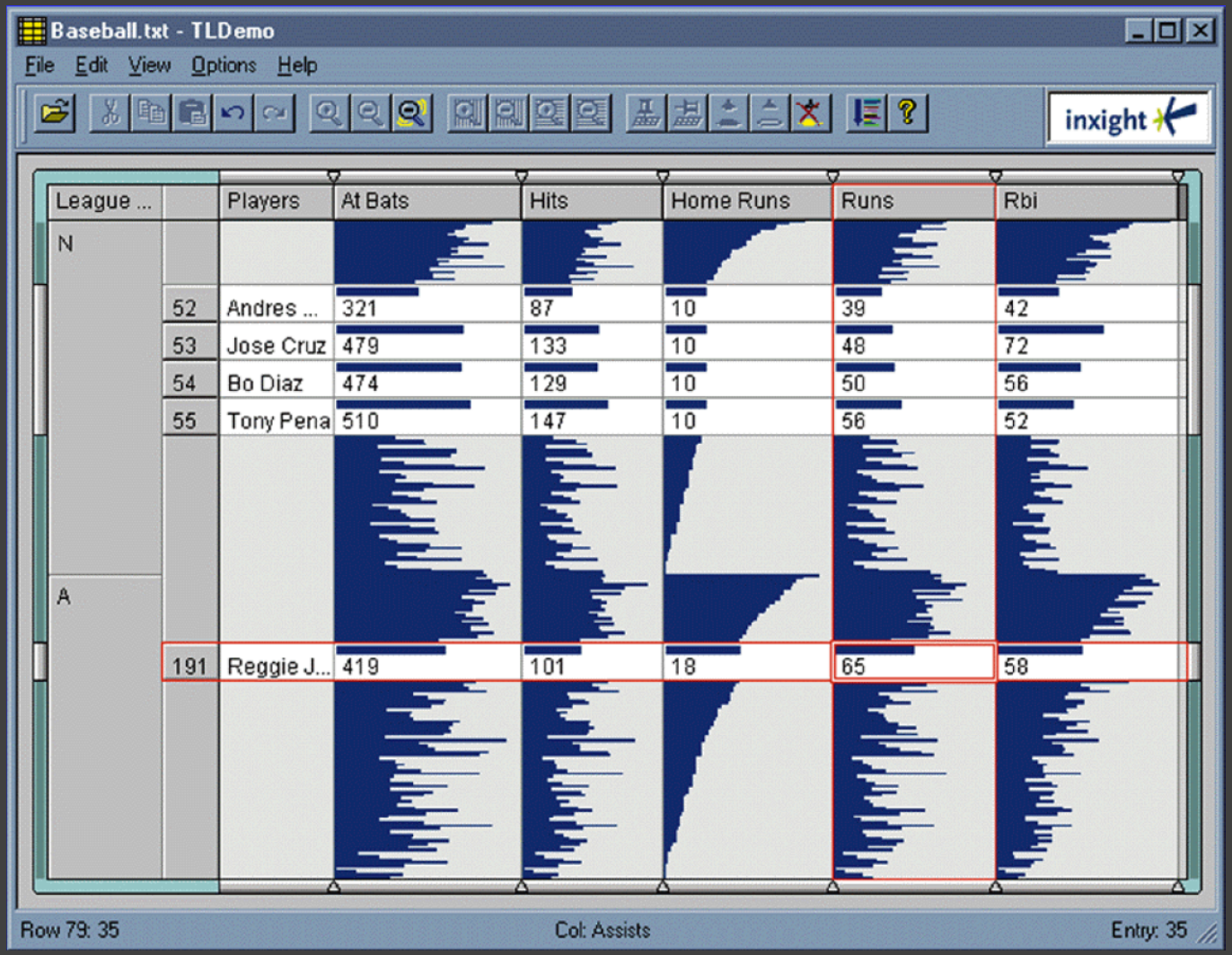


Perspective allows more context

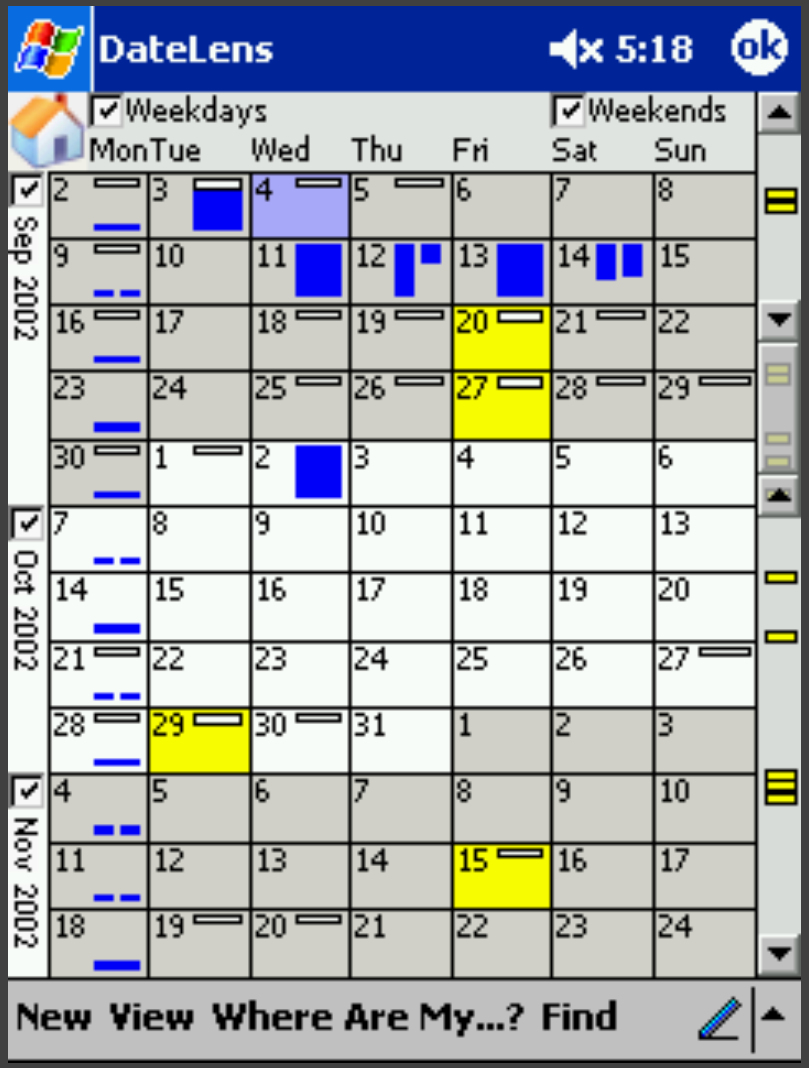


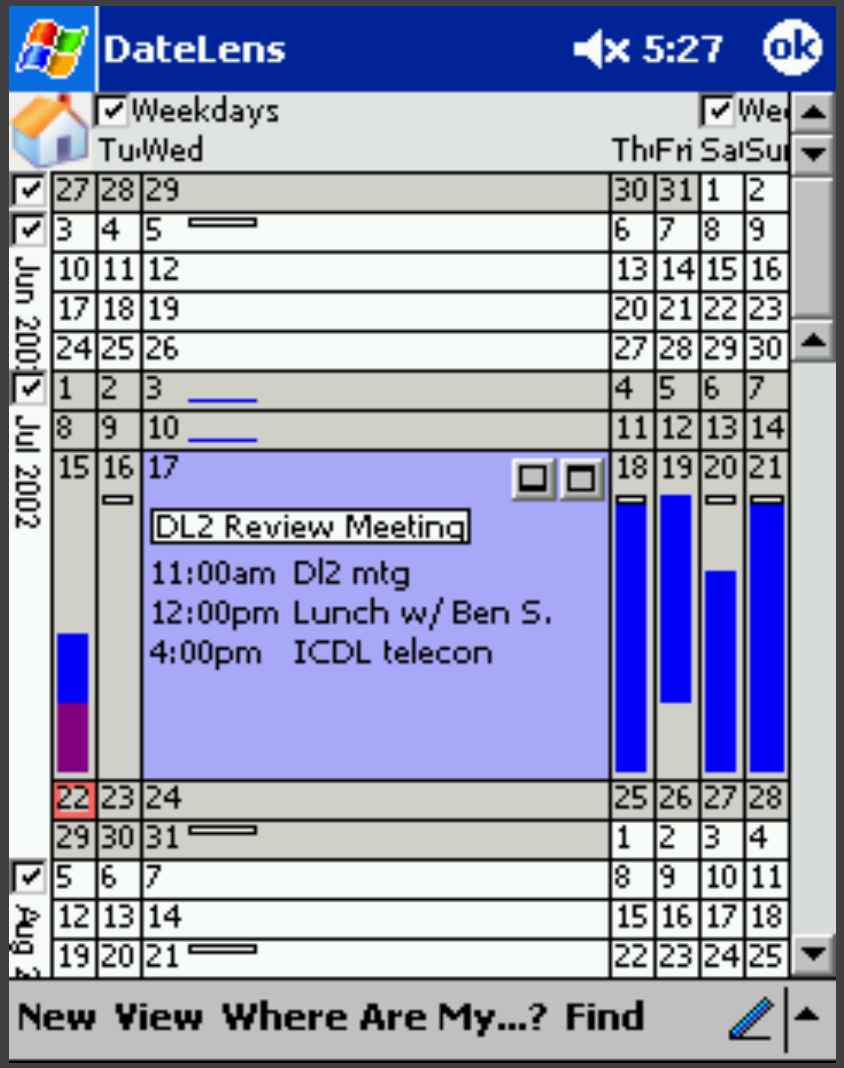
Perspective Wall [Mackinlay et al. 91]

TableLens [Rao & Card 94]



DateLens





[Bederson et al. 04]

Degree-of-Interest [Furnas 81, 06]

Estimate the saliency of information to display

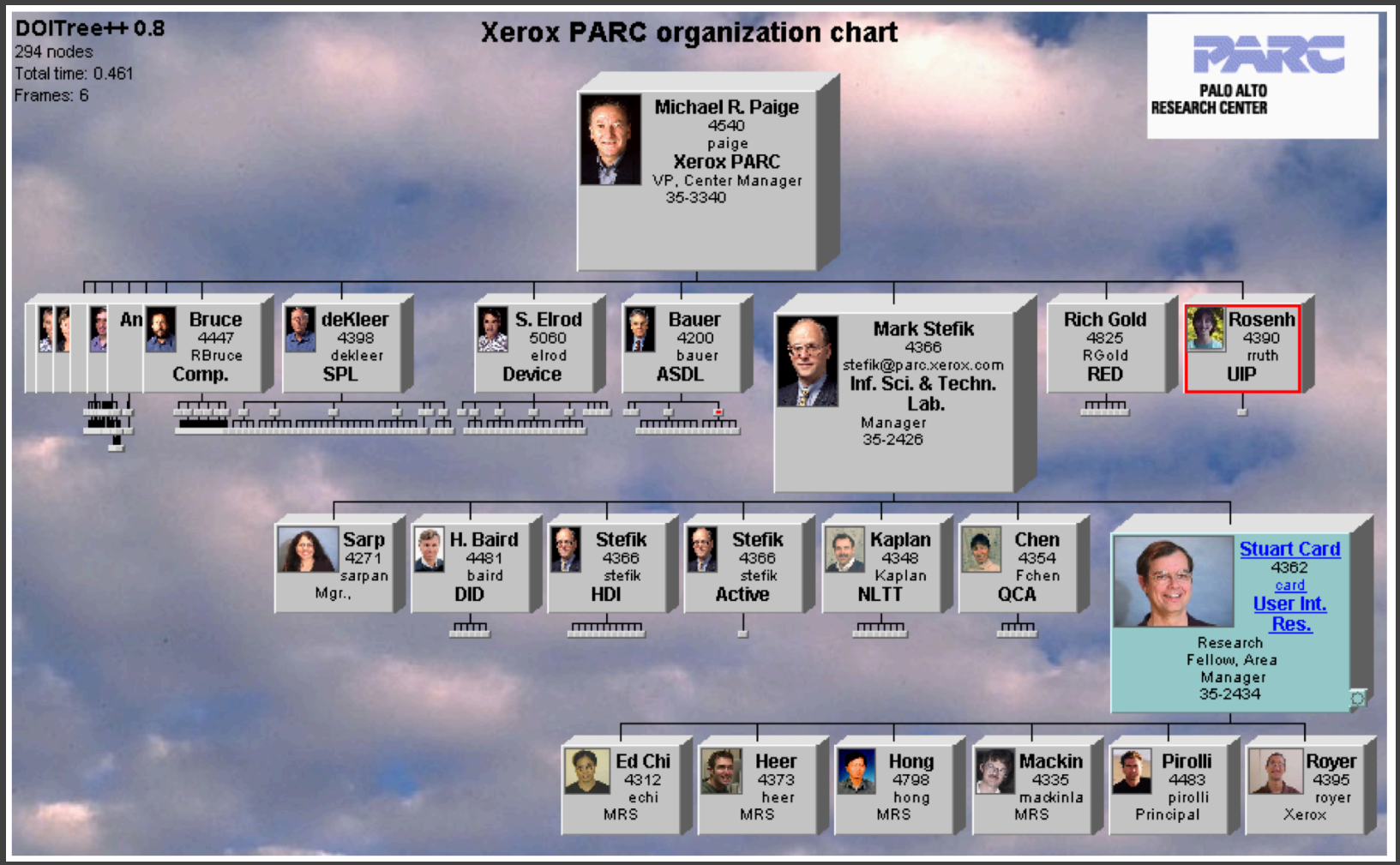
Can affect *what* is shown and/or *how* to show it

DOI ~ f(Current Focus, A Priori Importance)

Example: Google Search

- Current Focus = Query Hits (e.g., TF.IDF score)
- A Priori Importance = PageRank
- *What*: Top N results, *How*: List

DOI Tree [Card 2002]



Uses (and abuses) of distortion

Often more harm than help, unless

- Builds on experience (e.g., perspective wall) and enables a particular task
- Intended to elicit response, capture attention
  - In which case it should draw attention directly to the phenomenon of interest.

Pan and zoom more familiar—and visually stable—than “rubber sheet”

Transform Viz?

Aspect Ratio
Distortion

Discussion

Compare/contrast
transforming data vs. transforming viz

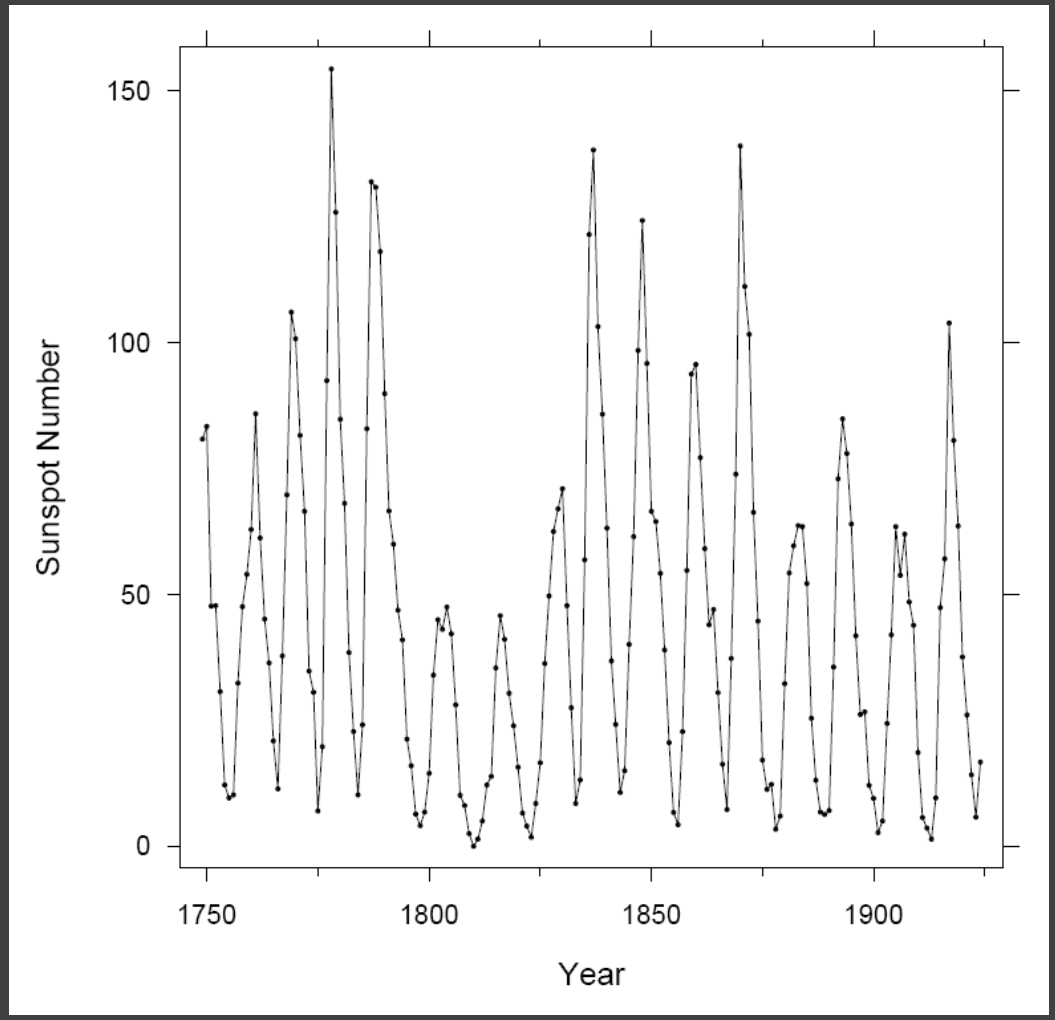
4. Optimize Layout

Layout

Given a visualization, what layout aspects might affect graphical perception?

- *Axis bounds*
- *Aspect Ratio*
- Ticks, Labels, Gridlines
- Line Width
- Data Points (e.g., dots)
- ...

How might we determine "optimal" choices?



Layout Optimization

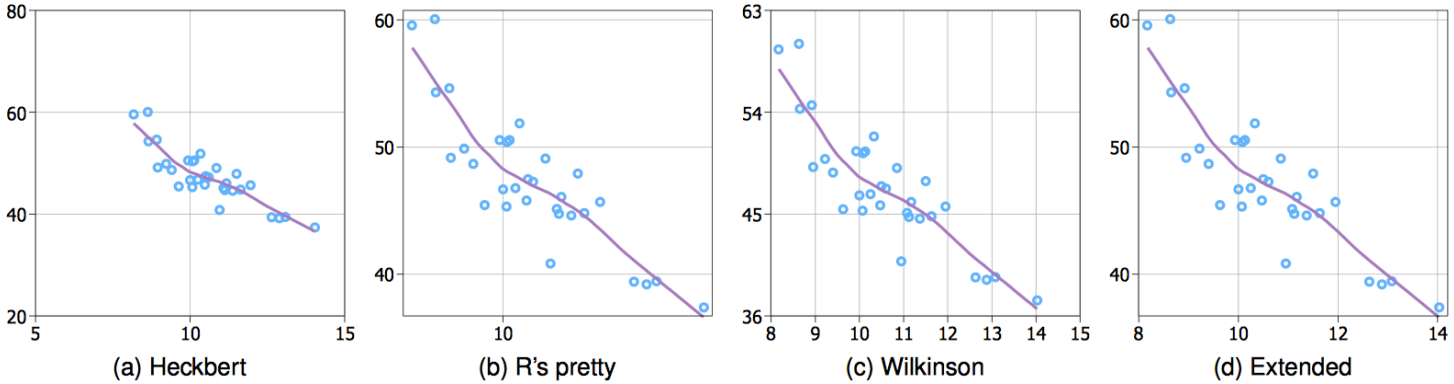
Define:

1. Search space
2. Optimization criteria
3. Algorithm

Layout Optimization

1. Tick marks
2. Labels
3. Streamgraph (stacked area charts)

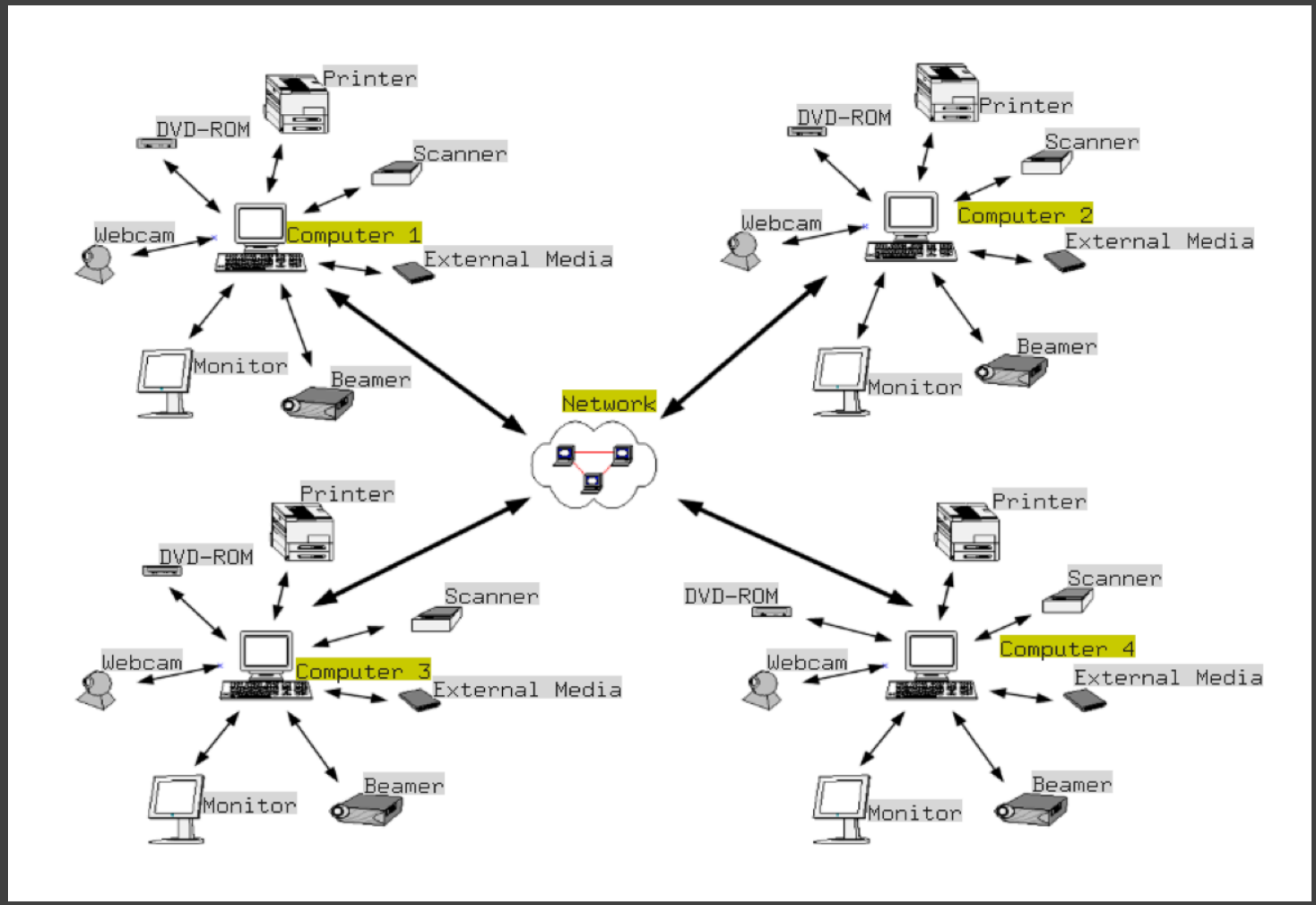
Tick Marks [Talbot 2010]



Tick Mark Optimization Criteria

1. Simple numbers (e.g. multiples of 10, 5, 2)
2. Coverage (ticks near the ends of the data)
3. Density (not too many or too few)
4. Legibility (whitespace, horizontal text, size)

Labels [Luboschik 2008]



Labels Optimization Criteria

No overlap

Close to labeled point

Label as many as possible

Label layout

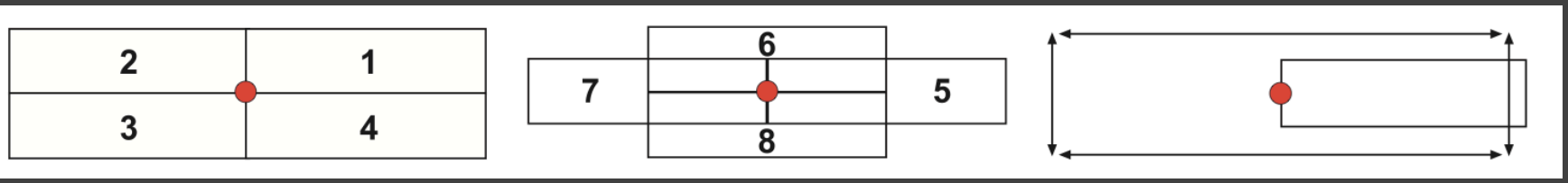
Known to be NP-hard

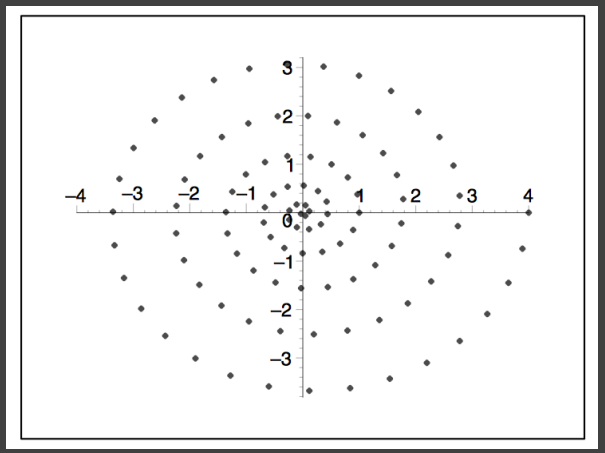
So use heuristics

Most approaches use backtracking/annealing.
Slow...

Label layout

1. Labeling with the 4-position model,
2. Labeling with positions 5–8 of the 8-position model,
3. Labeling with the 4-slider model,
4. Labeling with distant positions.

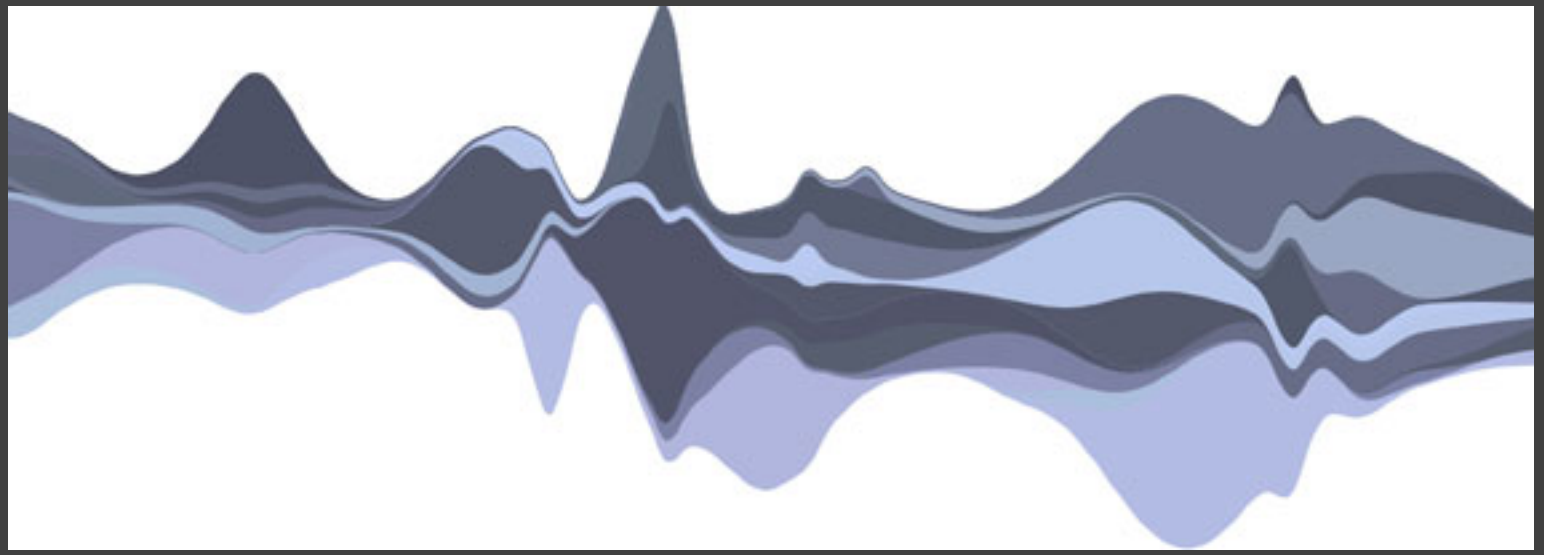


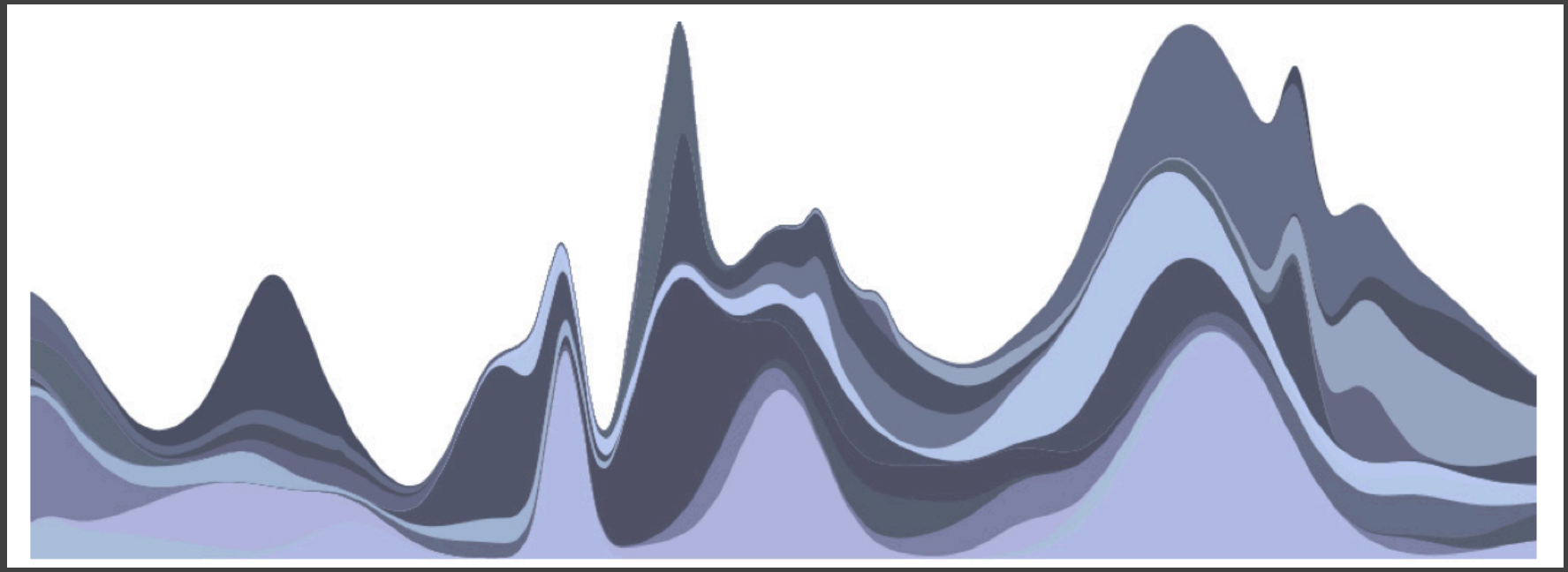


Label layout

Video: http://www.informatik.uni-rostock.de/~malub/pub/08/labeling/08_LSC_labeling.avi

Streamgraphs [Byron 2008]





Optimization criteria

Minimize distance from x-axis

Minimize variation in slope

Weight bigger layers more

“Inside-out” ordering (tails in the middle)

Discussion

What is good/bad about their resulting layout?

What other optimization criteria might they have picked?

Why do you think people responded positively to this viz?

Why have we not seen anything similar since?

Optimize Layout?

Lots of open research directions!

Summary

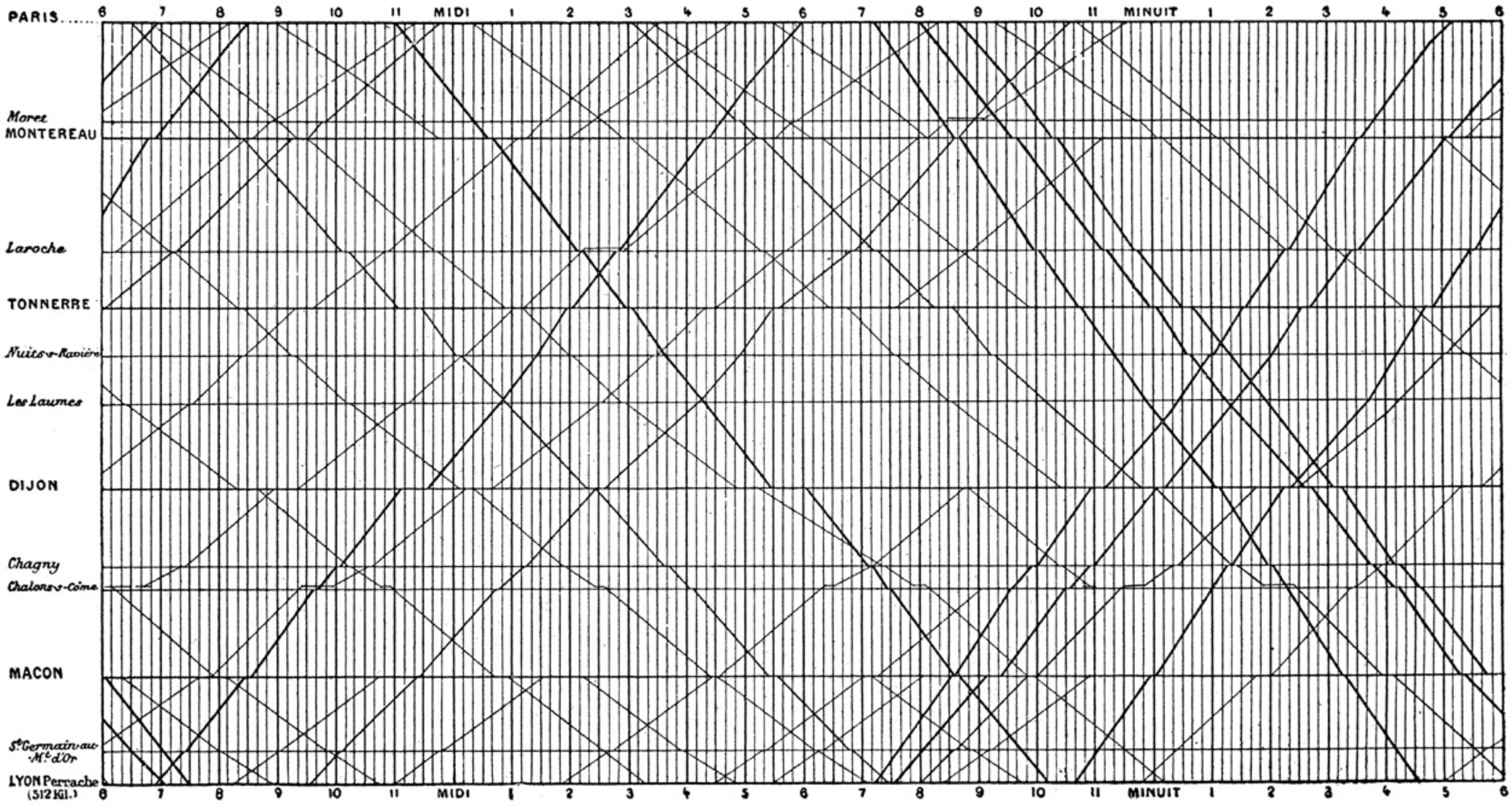
Spatial layout is the most important encoding

Strategies

1. Focus + Context
2. Transform Data
3. Transform Viz
4. Layout Optimization

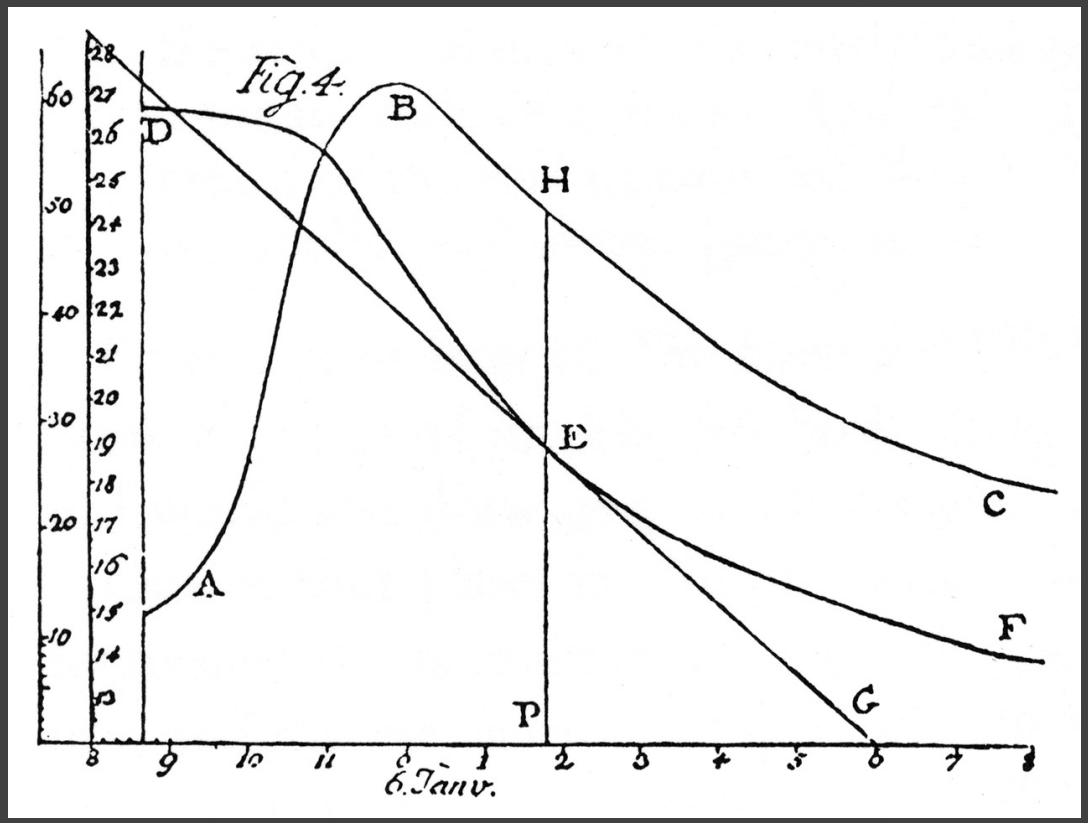
Tradeoffs, so consider carefully

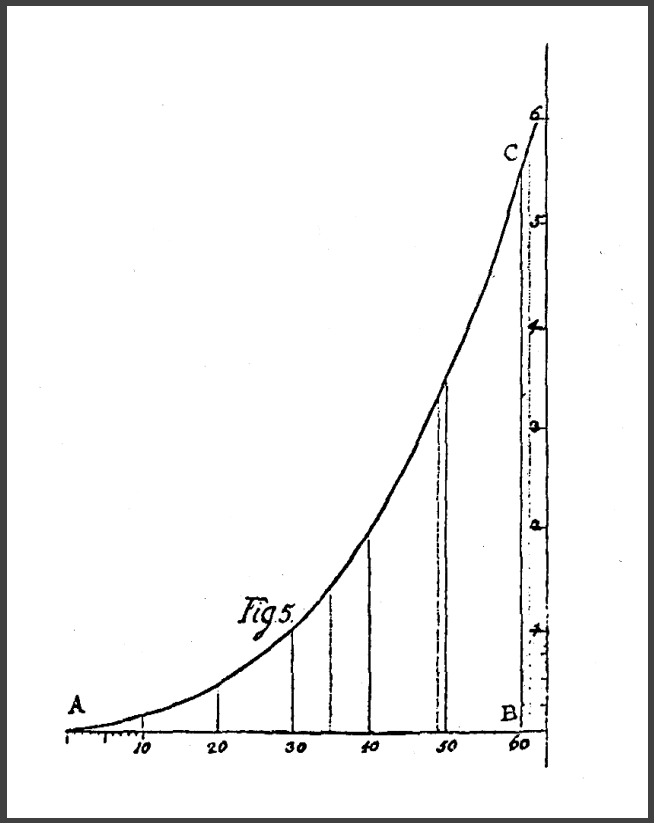
Bonus Topic: Graphical Calculation



E. J. Marey, *La Méthode Graphique* (Paris, 1885), p. 20. The method is attributed to the French engineer, Ibry.

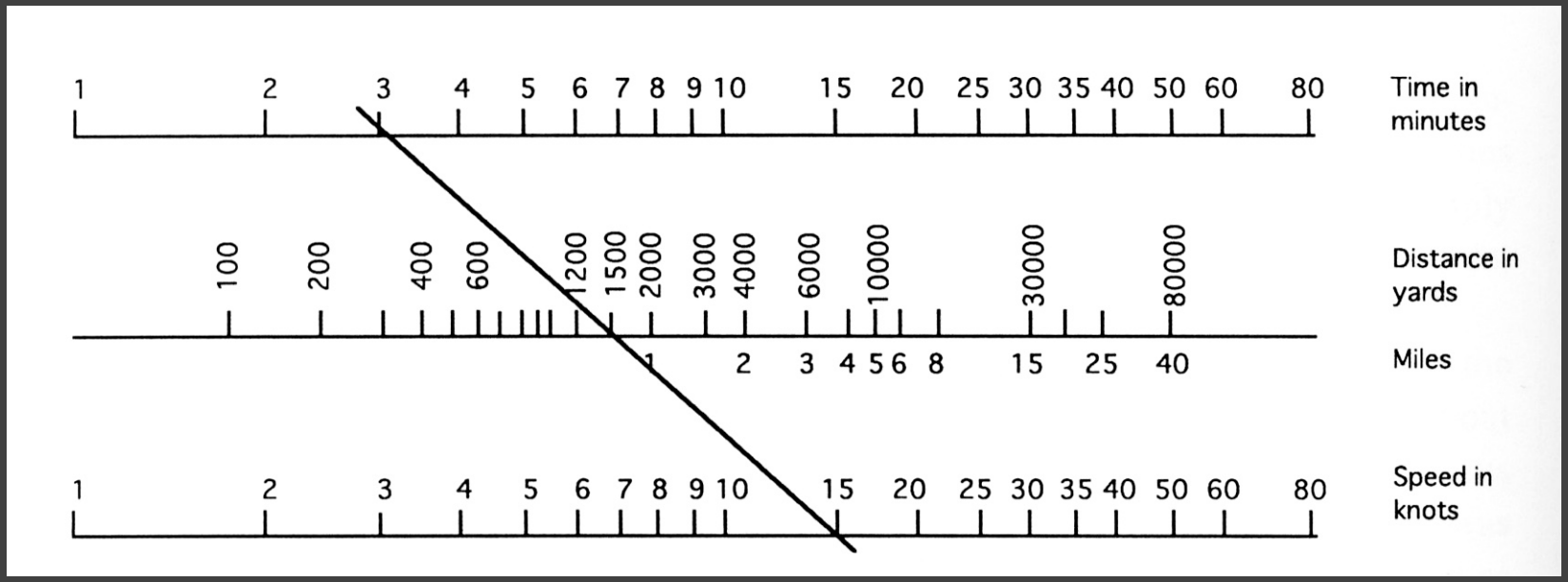
Lambert's graphical construction





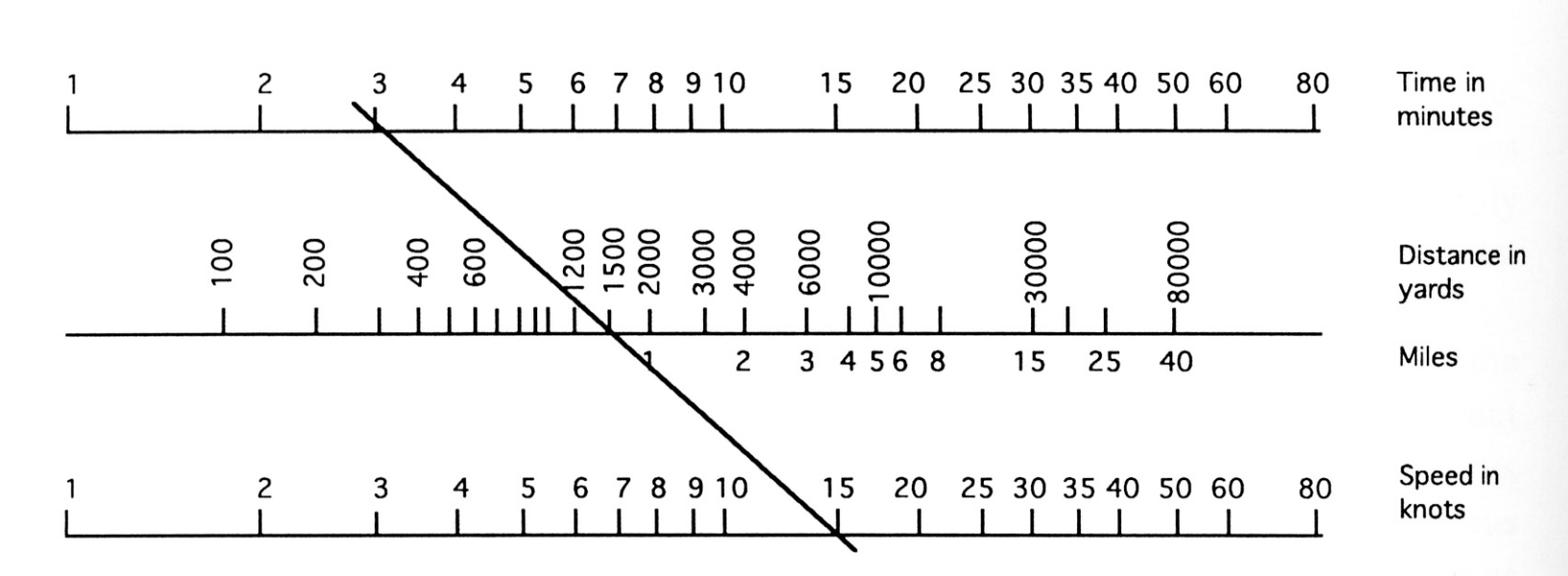
Johannes Lambert used graphs to study the rate of water evaporation as function of temperature [from Tufte 83]

Nomograms



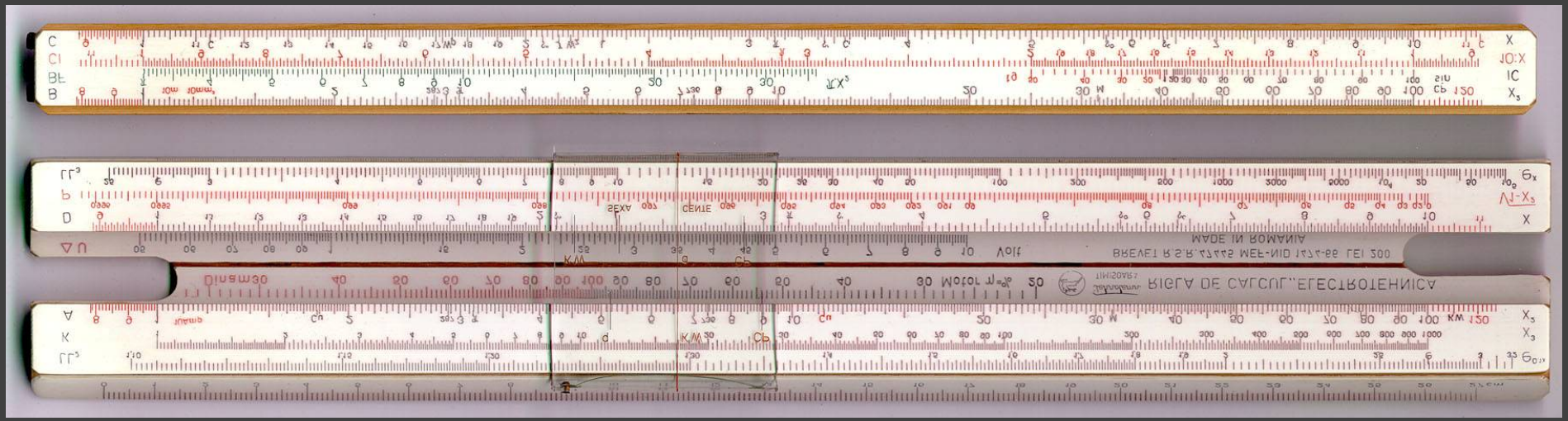
Sailing: The Rule of Three

Nomograms



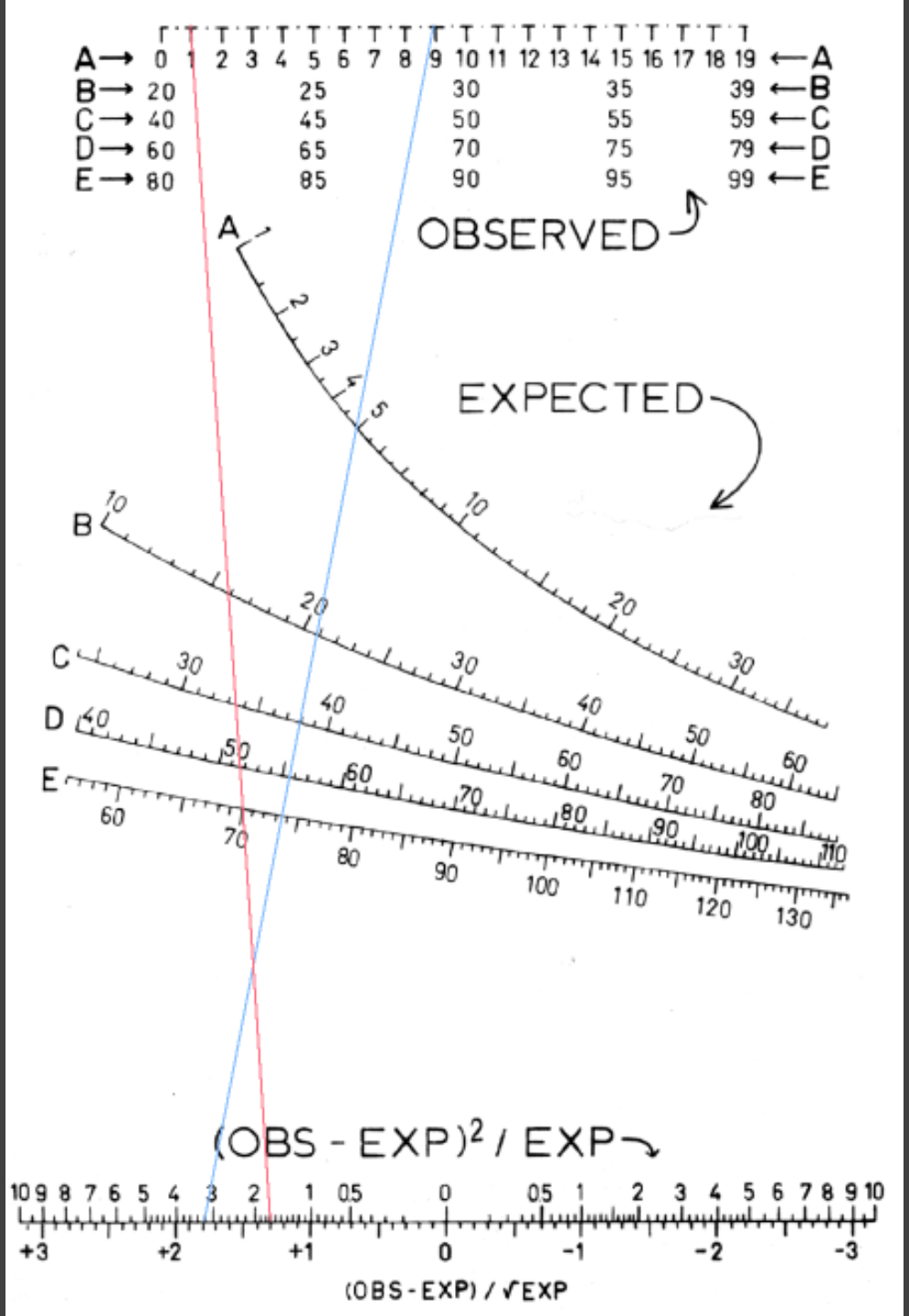
1. Compute in any direction; fix n-1 params and read nth param
2. Illustrate sensitivity to perturbation of inputs
3. Clearly show domain of validity of computation

Slide rule



Model 1474-66 Electrotechnica 18 Scales

Tehnolemn Timisoara Slide Rule Archive
http://pubpages.unh.edu/~jwc/tehnolemn/



Chi-square test

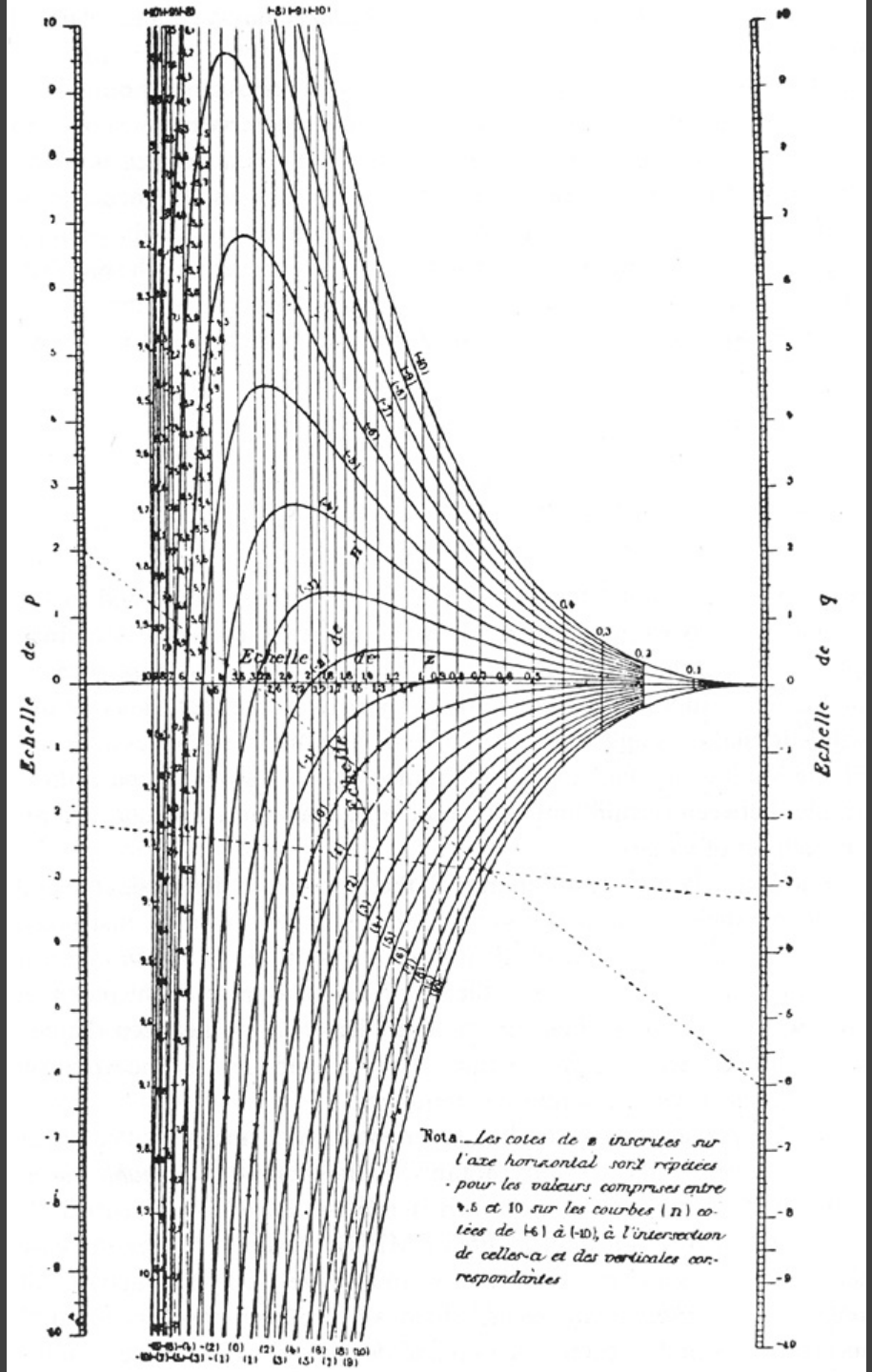
$(Obs - Exp)^2 / Exp$

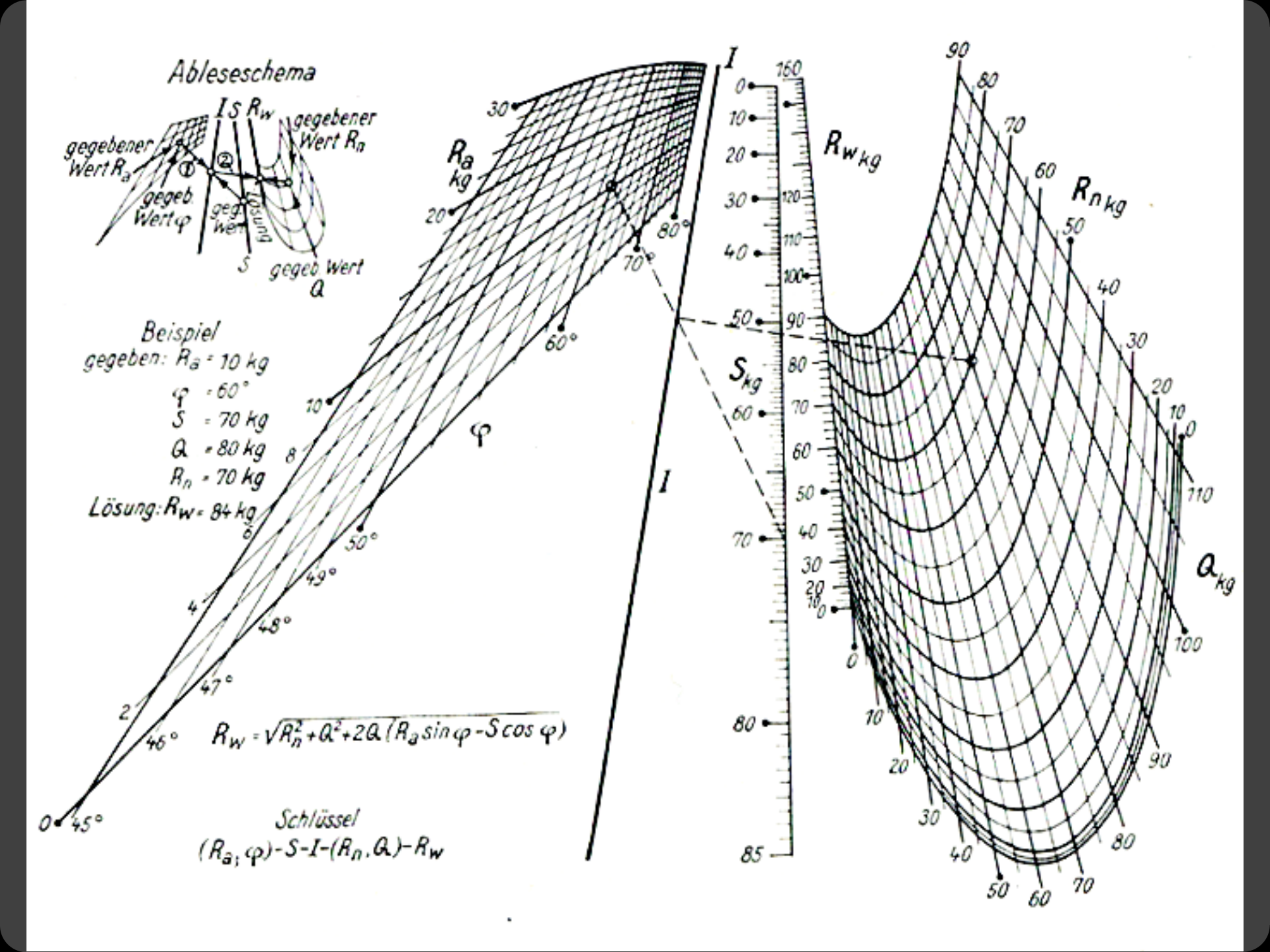
Blue line:

$(9 - 5)^2 / 5 = 3.2$

Red line:

$(81 - 70)^2 / 70 = 1.7$


Echelle de p
Echelle de q
Echelle de
Nota. Les cotes de n inscrites sur l'axe horizontal sont répétées pour les valeurs comprises entre 9.5 et 10 sur les courbes (n) cotées de (+6) à (-10), à l'intersection de celles-ci et des verticales correspondantes


Ableseschema
I S R_w
gegebener Wert R_a
gegebener Wert R_n
gegeb Wert φ
gegeb Wert
Lösung
S
gegeb Wert Q
Beispiel
gegeben: R_a = 10 kg
φ = 60°
S = 70 kg
Q = 80 kg
R_n = 70 kg
Lösung: R_w = 84 kg
$R_w = \sqrt{R_n^2 + Q^2 + 2Q(R_a \sin\varphi - S\cos\varphi)}$
Schlüssel
$(R_a; \varphi)$-S-I-(R_n, Q)-R_w
R_a kg
φ
I
$R_{w\,kg}$
S_{kg}
$R_{n\,kg}$
Q_{kg}

Summary

Spatial layout is the most important encoding

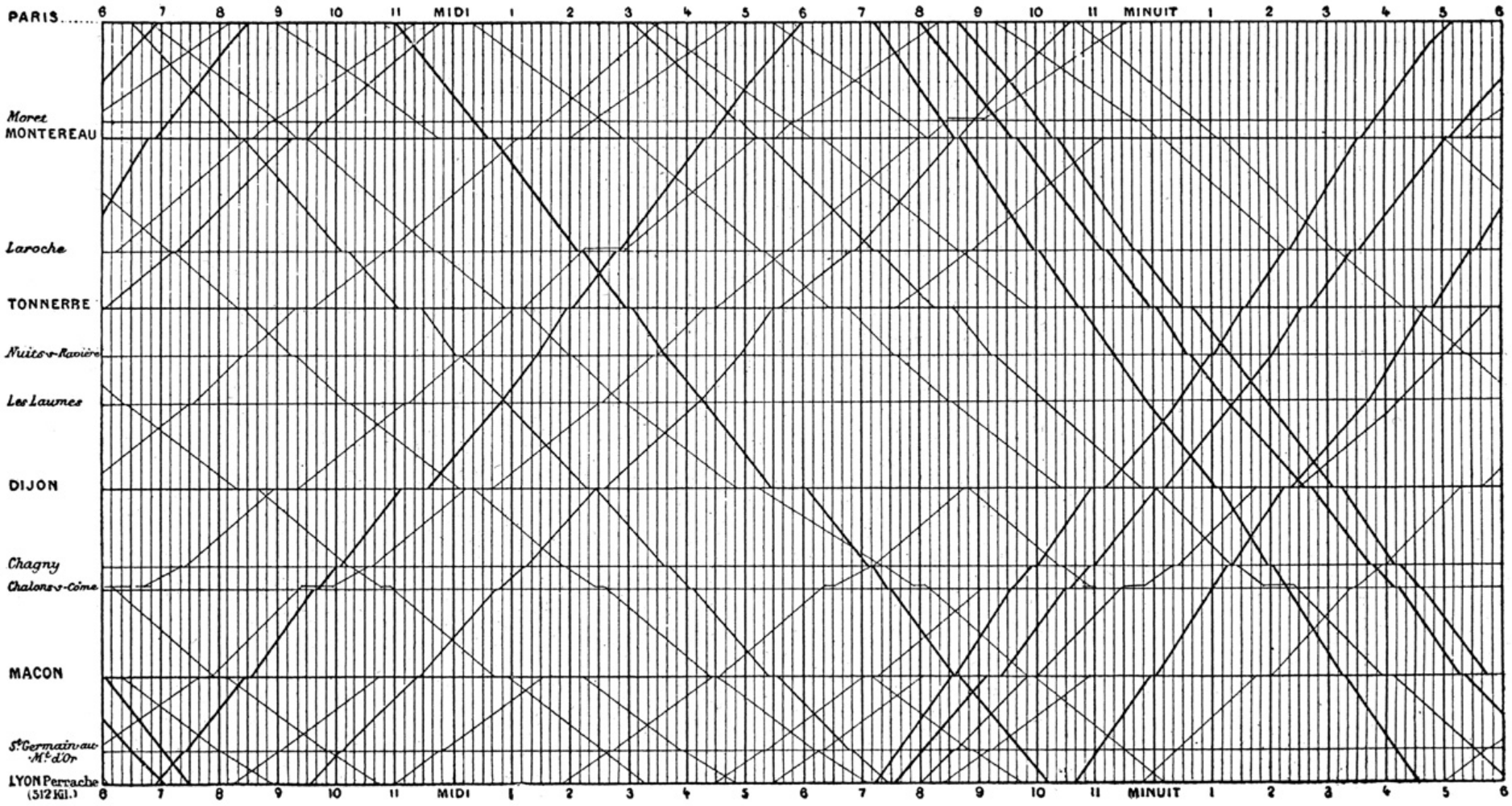
...but you need to be in the *right* space.

Geometric properties of spatial transforms support geometric reasoning

Emphasize important information

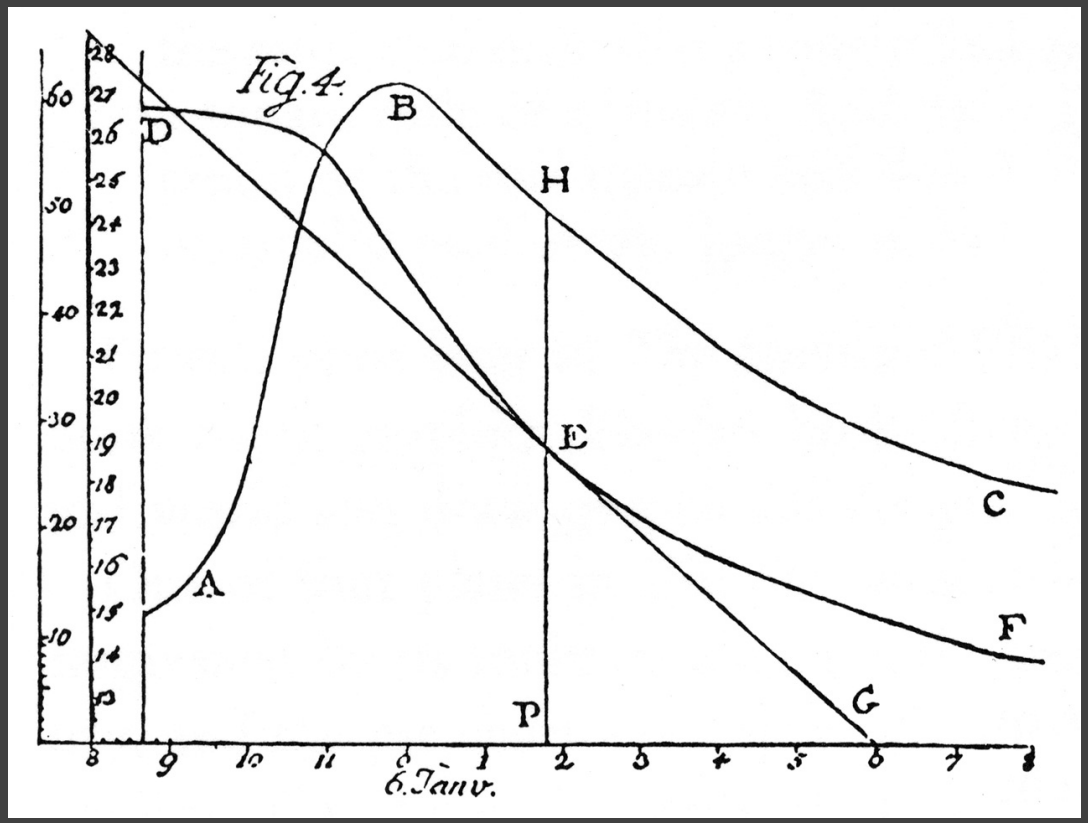
- Consider *what* to show, not just *how*

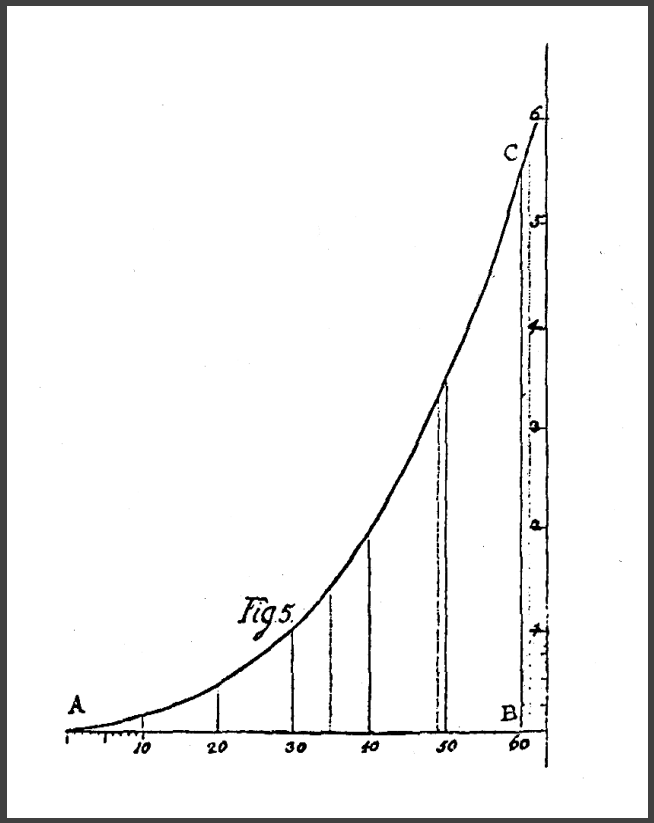
Graphical Calculations



E. J. Marey, *La Méthode Graphique* (Paris, 1885), p. 20. The method is attributed to the French engineer, Ibry.

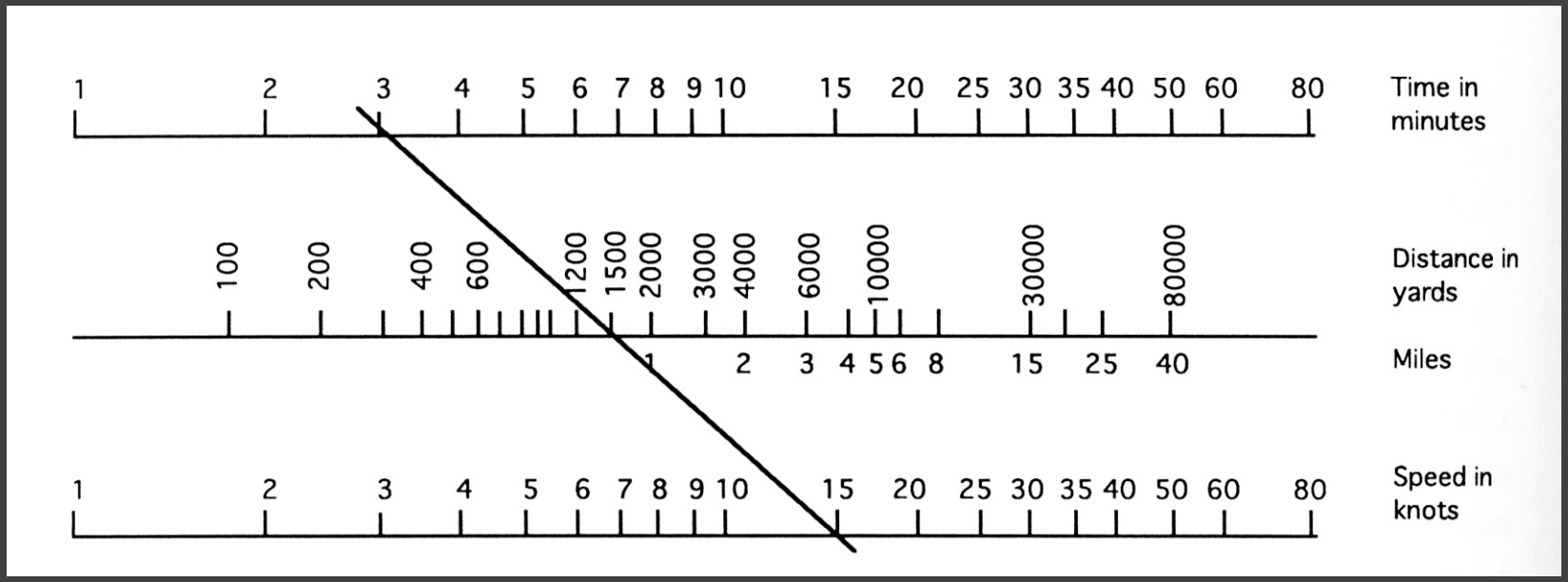
Lambert's graphical construction





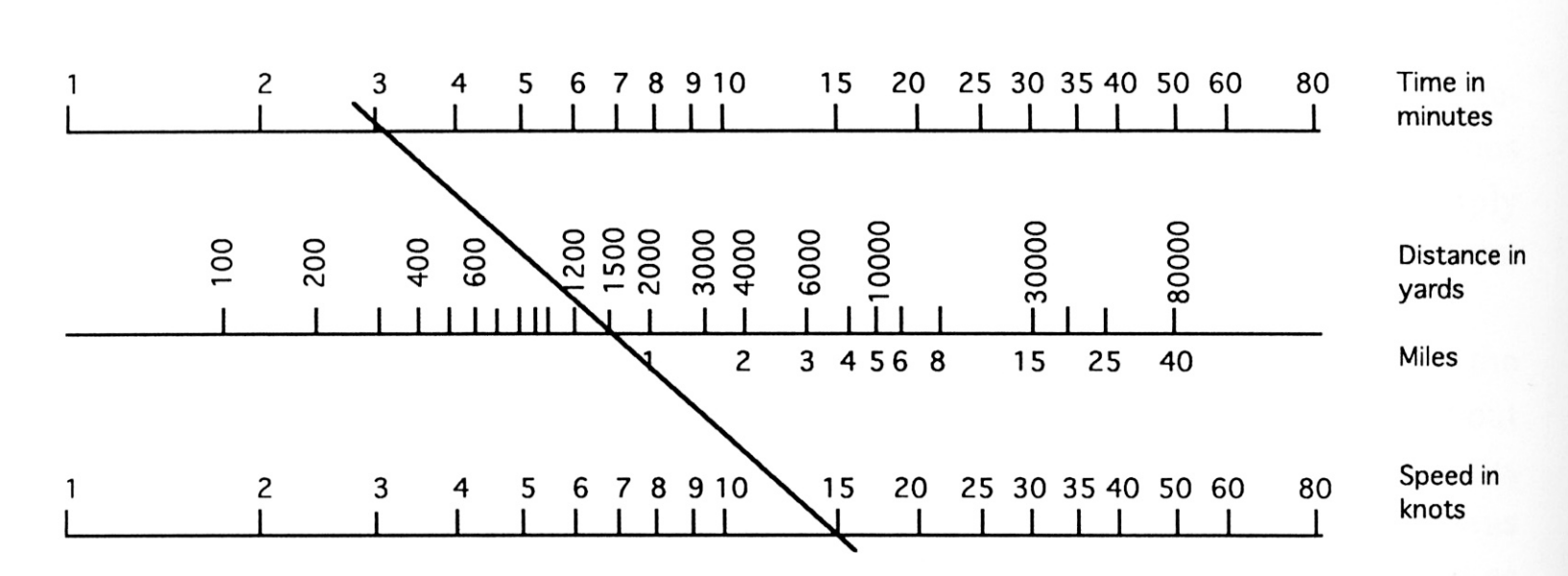
Johannes Lambert used graphs to study the rate of water evaporation as function of temperature [from Tufte 83]

Nomograms



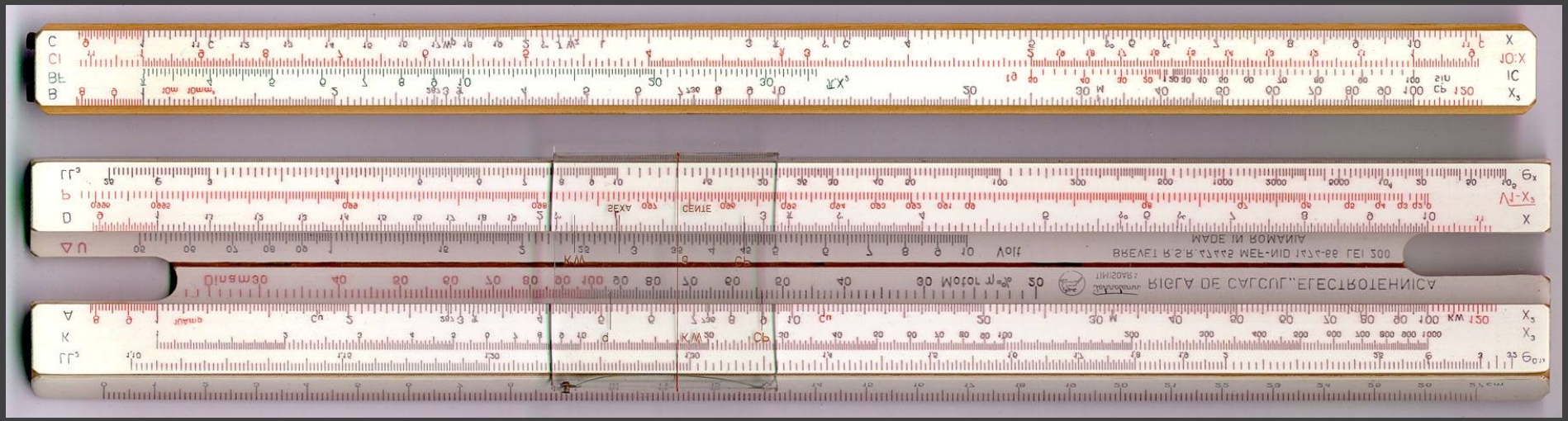
Sailing: The Rule of Three

Nomograms



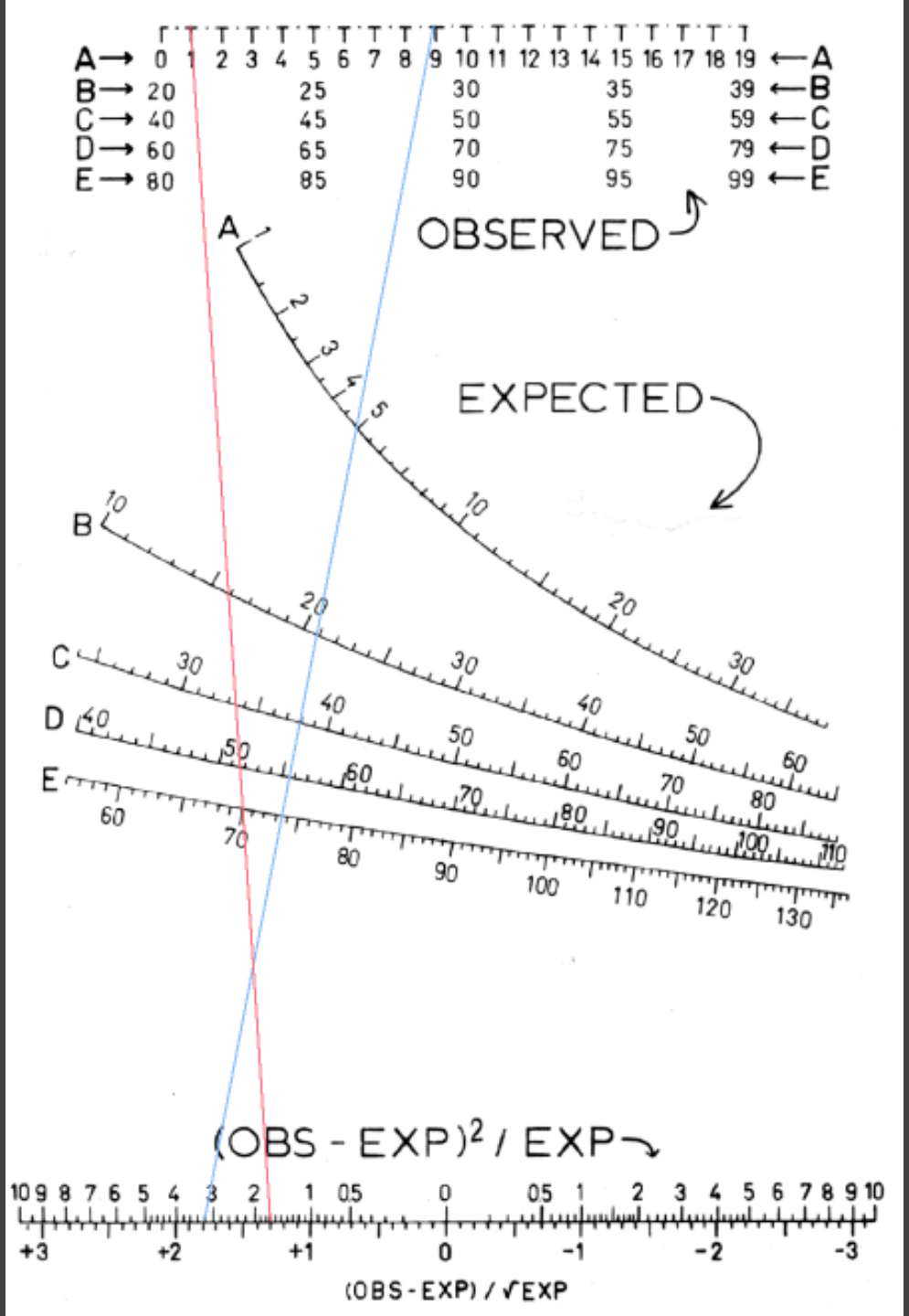
1. Compute in any direction; fix n-1 params and read nth param
2. Illustrate sensitivity to perturbation of inputs
3. Clearly show domain of validity of computation

Slide rule



Model 1474-66 Electrotechnica 18 Scales

Tehnolemn Timisoara Slide Rule Archive
http://pubpages.unh.edu/~jwc/tehnolemn/

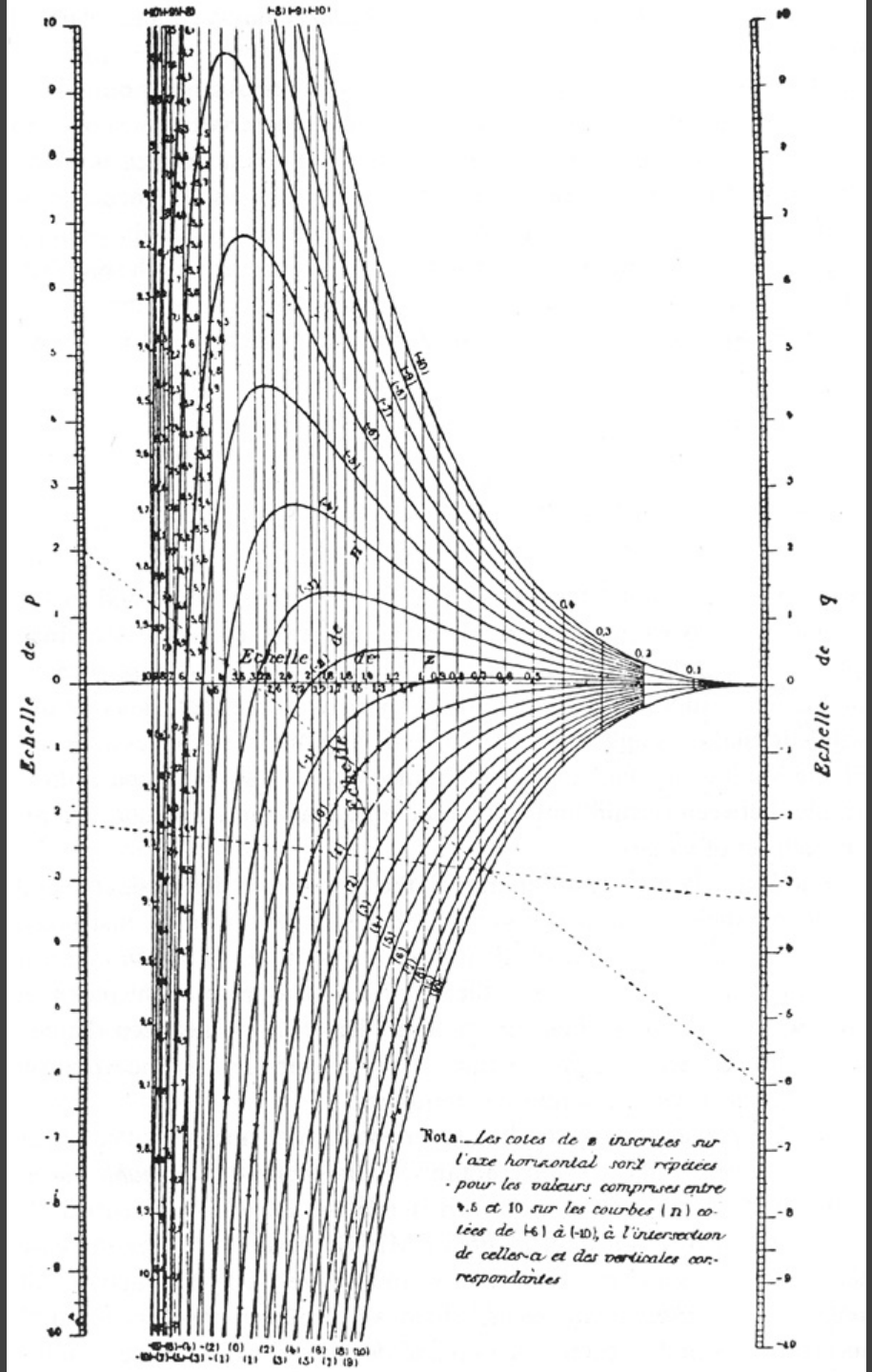


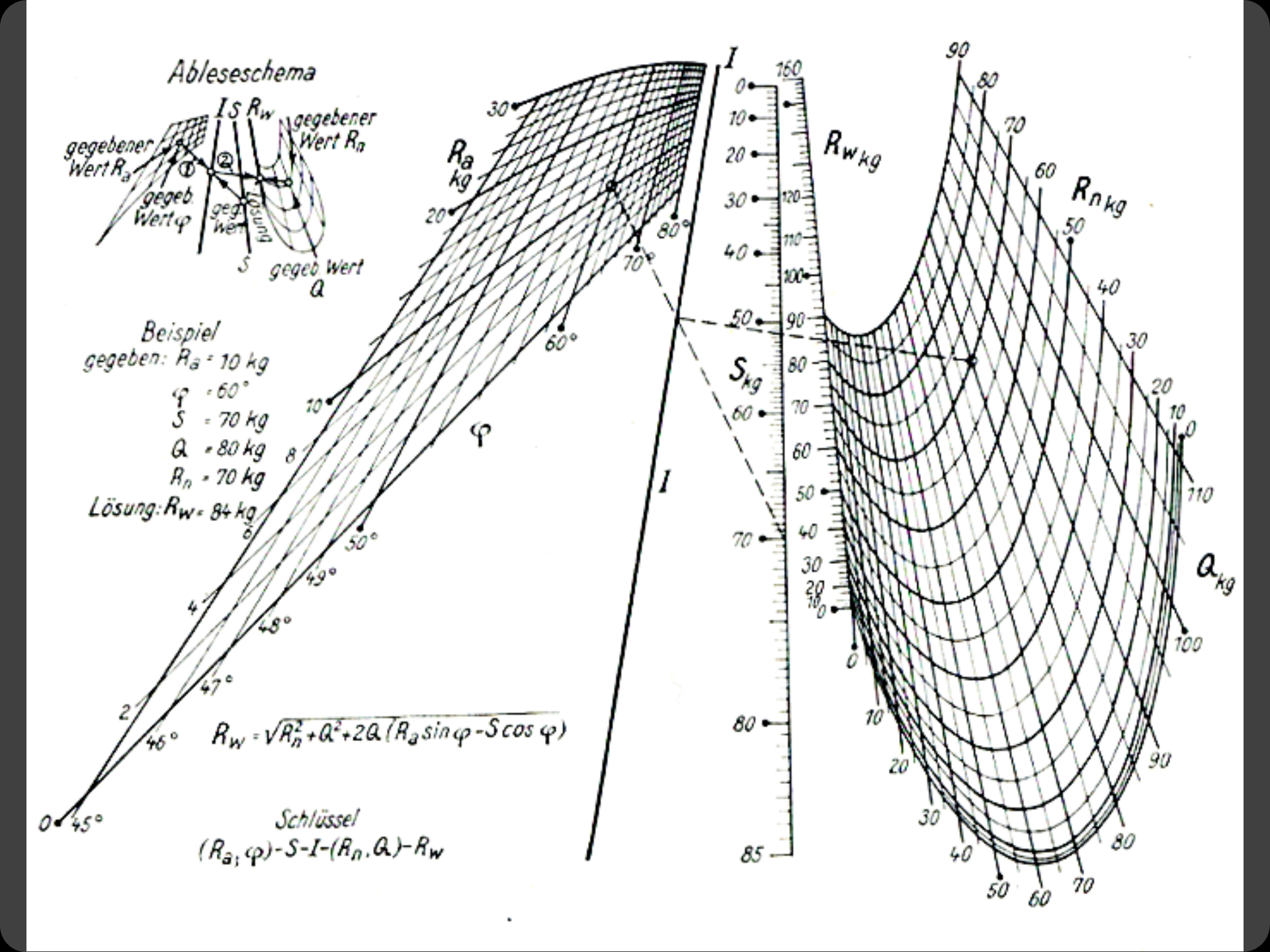
Blue line:

$(9 - 5)^2 / 5 = 3.2$

Red line:

$(81 - 70)^2 / 70 = 1.7$

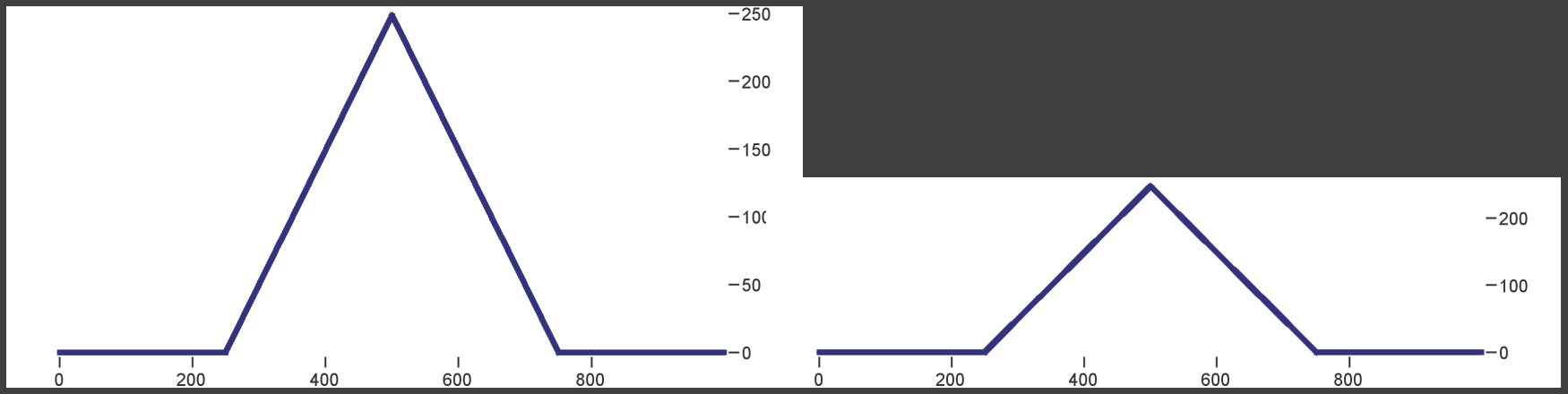

Echelle de p
Echelle de q
Echelle de
Nota. Les cotes de n inscrites sur l'axe horizontal sont répétées pour les valeurs comprises entre 9.5 et 10 sur les courbes (n) cotées de (+6) à (-10), à l'intersection de celles-ci et des verticales correspondantes


Ableseschema
I S R_w
gegebener Wert R_a
gegebener Wert R_n
gegeb Wert φ
gegeb Wert
Lösung
S
gegeb Wert Q
Beispiel
gegeben: R_a = 10 kg
φ = 60°
S = 70 kg
Q = 80 kg
R_n = 70 kg
Lösung: R_w = 84 kg
$R_w = \sqrt{R_n^2 + Q^2 + 2Q(R_a \sin\varphi - S\cos\varphi)}$
Schlüssel
$(R_a; \varphi)$-S-I-(R_n, Q)-R_w
R_a kg
φ
I
$R_{w\,kg}$
S_{kg}
$R_{n\,kg}$
Q_{kg}

Theory

$$\begin{vmatrix} x_1(u) & y_1(u) & w_1(u) \\ x_2(v) & y_2(v) & w_2(v) \\ x_3(s,t) & y_3(s,t) & w_3(s,t) \end{vmatrix} = 0$$

Slopeless Line Culling



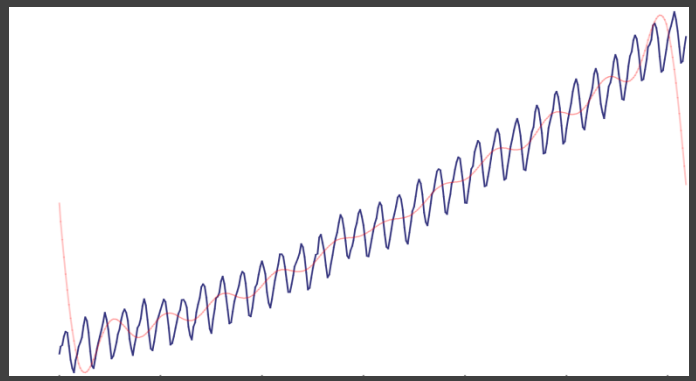
Standard, Aspect Ratio = 1.97

Culled, Aspect Ratio = 4.00

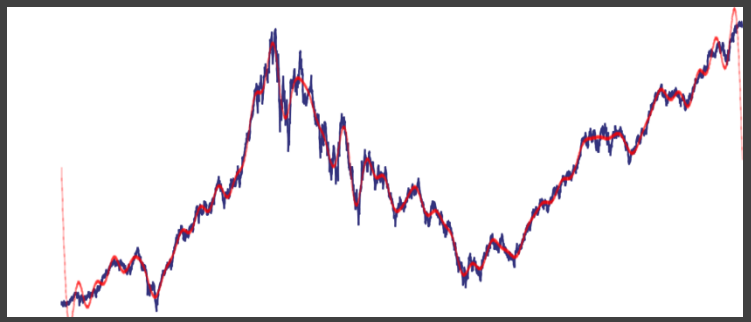
Exclude line segments with zero or infinite slope

Comparison (Data Sets)

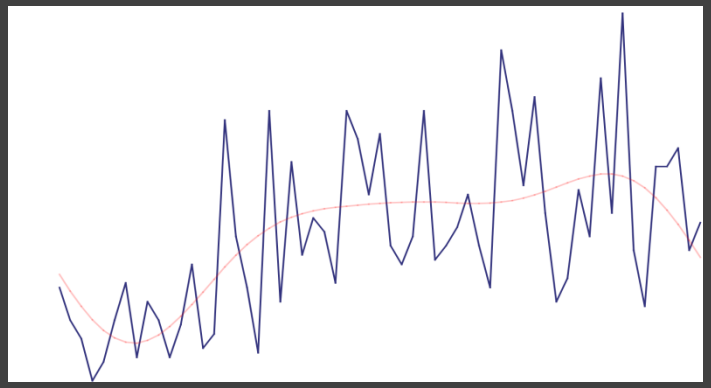
CO_2 Measurements (co2)



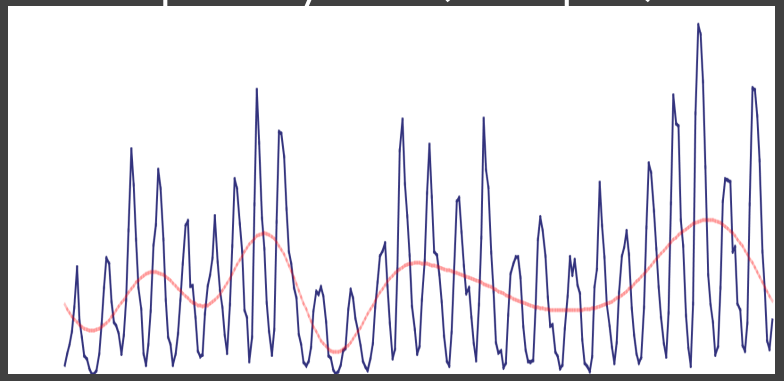
PRMTX Mutual Fund (prmtx)



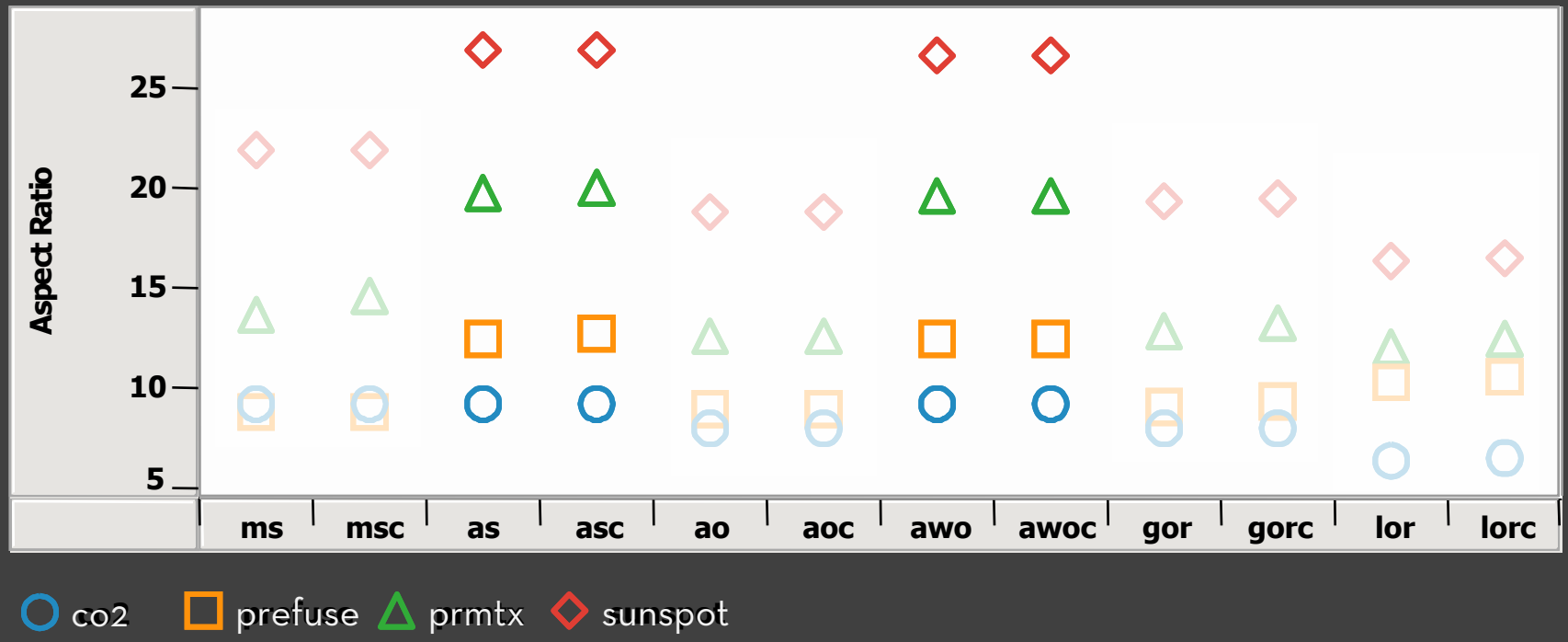
Prefuse Downloads (prefuse)



Sunspot Cycles (sunspot)

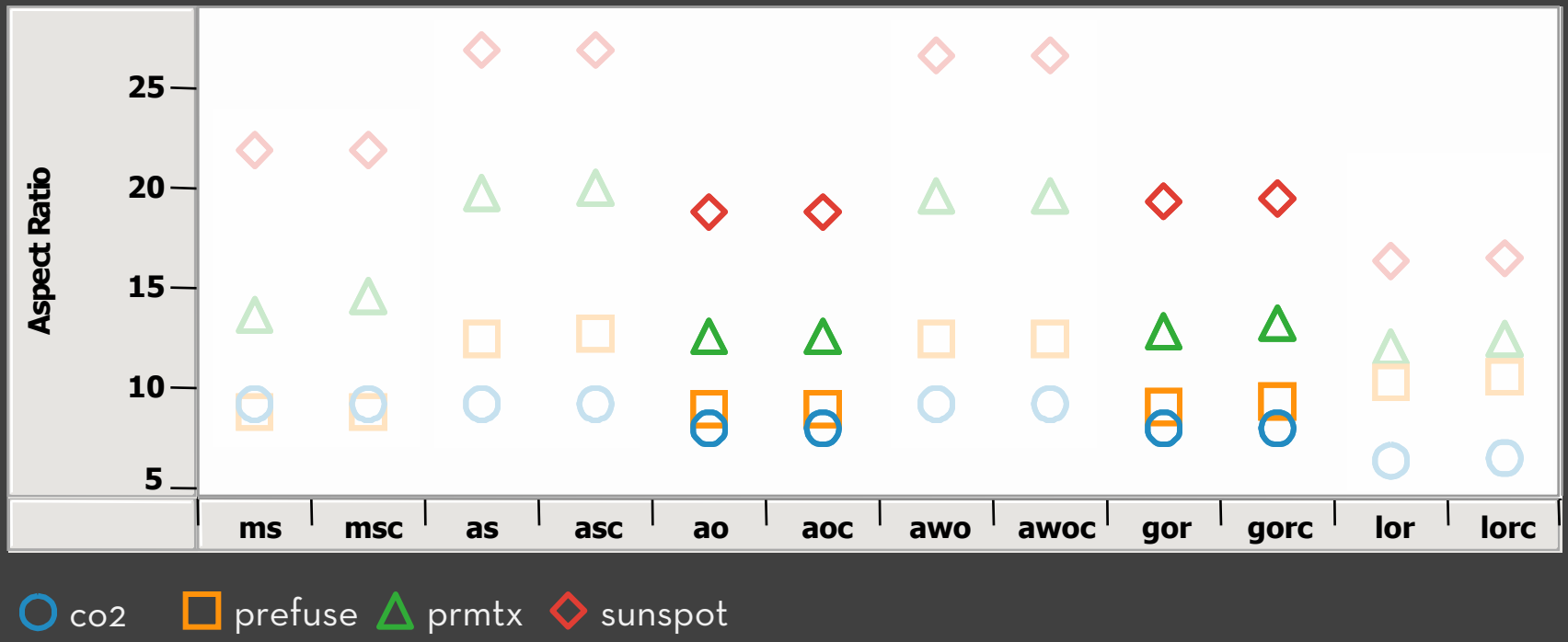


Comparison (Results)



Average-slope (as) and Average-weighted-orientation (awo) provide similar ratios

Comparison (Results)



Average-orientation (ao) and Global-orientation-resolution (gor) provide similar ratios

Discussion

Due to computational complexity...

Prefer avg-slope to avg-weighted-orient

Prefer avg-orient to global-orient-resolution

But due to perceptual effectiveness... ?

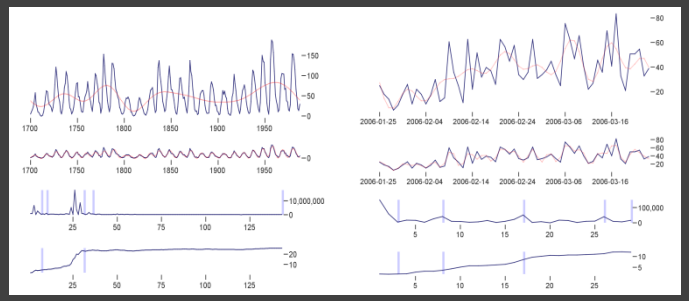
Cleveland recommends weighted-avg-orient

But, goal is to maximize discriminability

Perceptual experiments needed to clarify

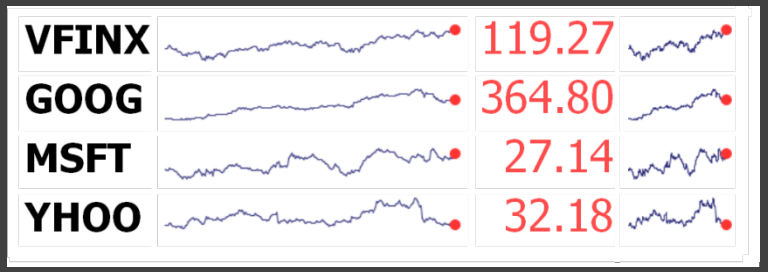
Applications

Small Multiples Displays

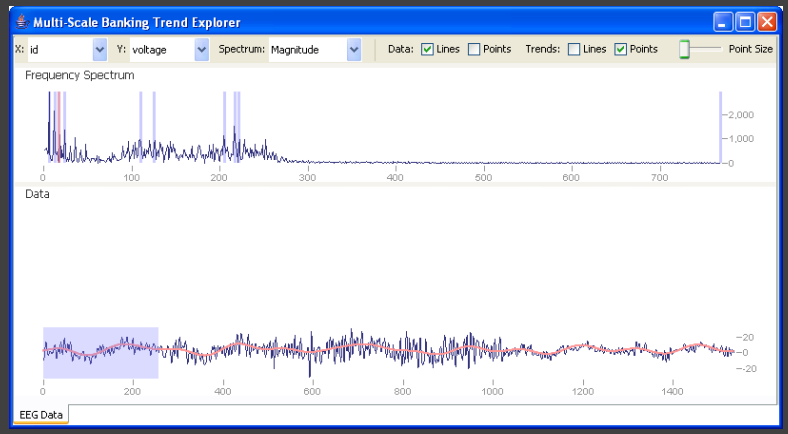


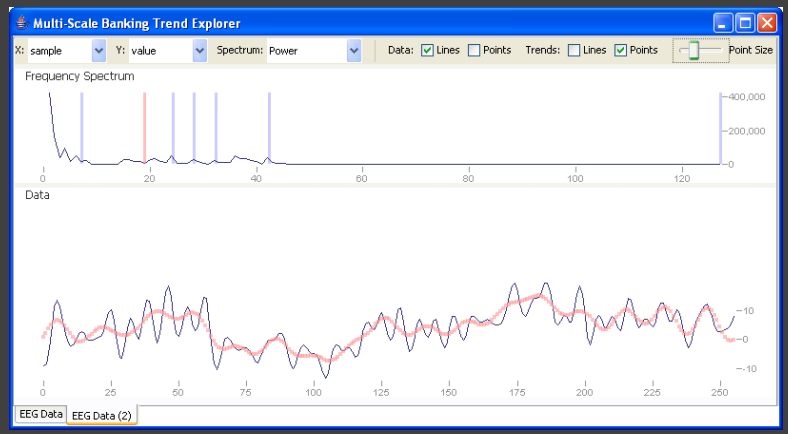
Sparklines

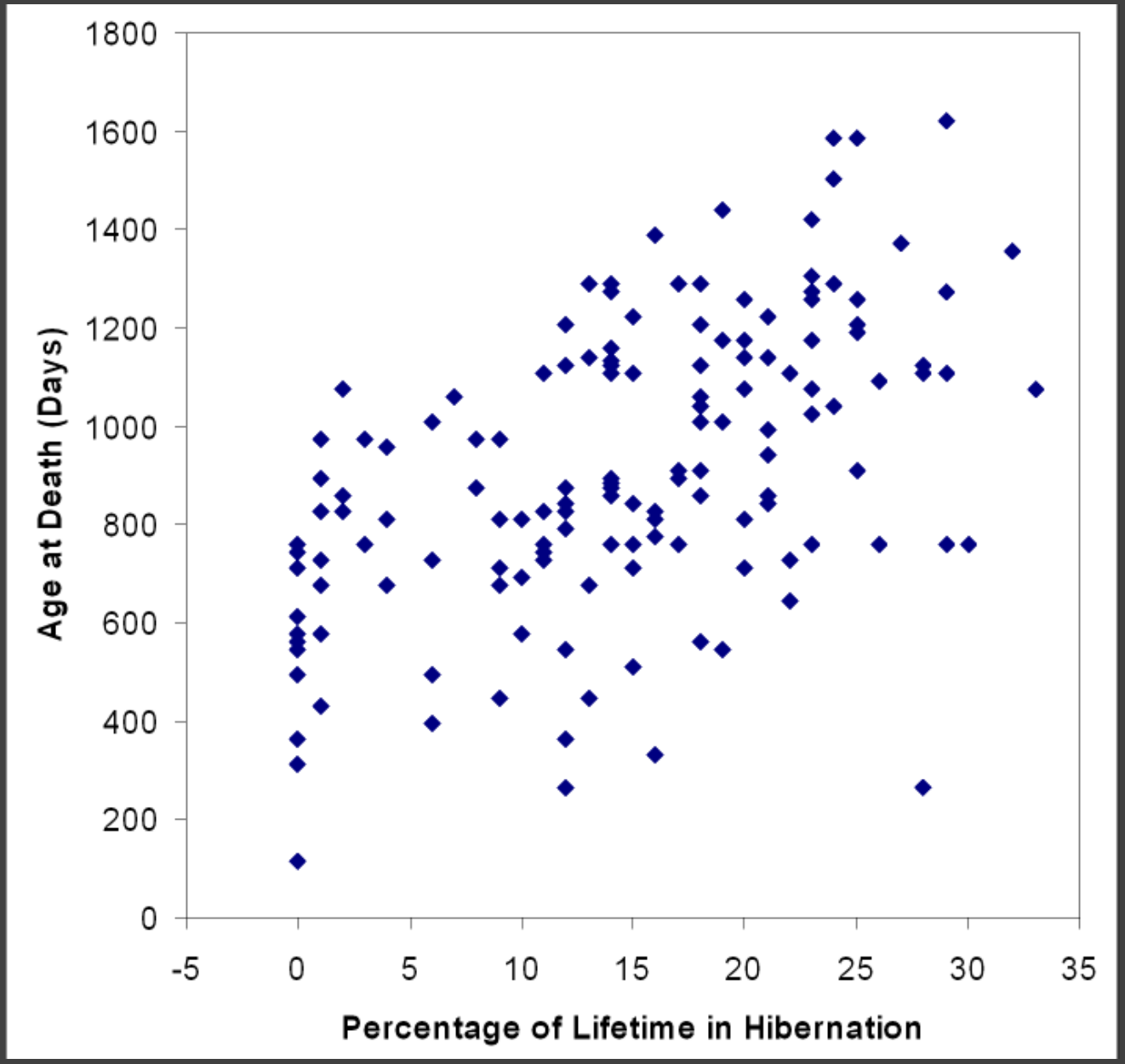
Banking can be applied to create *sparklines*, data-intense, word-sized graphics. A plot might be included inline  , supporting uninterrupted reading.



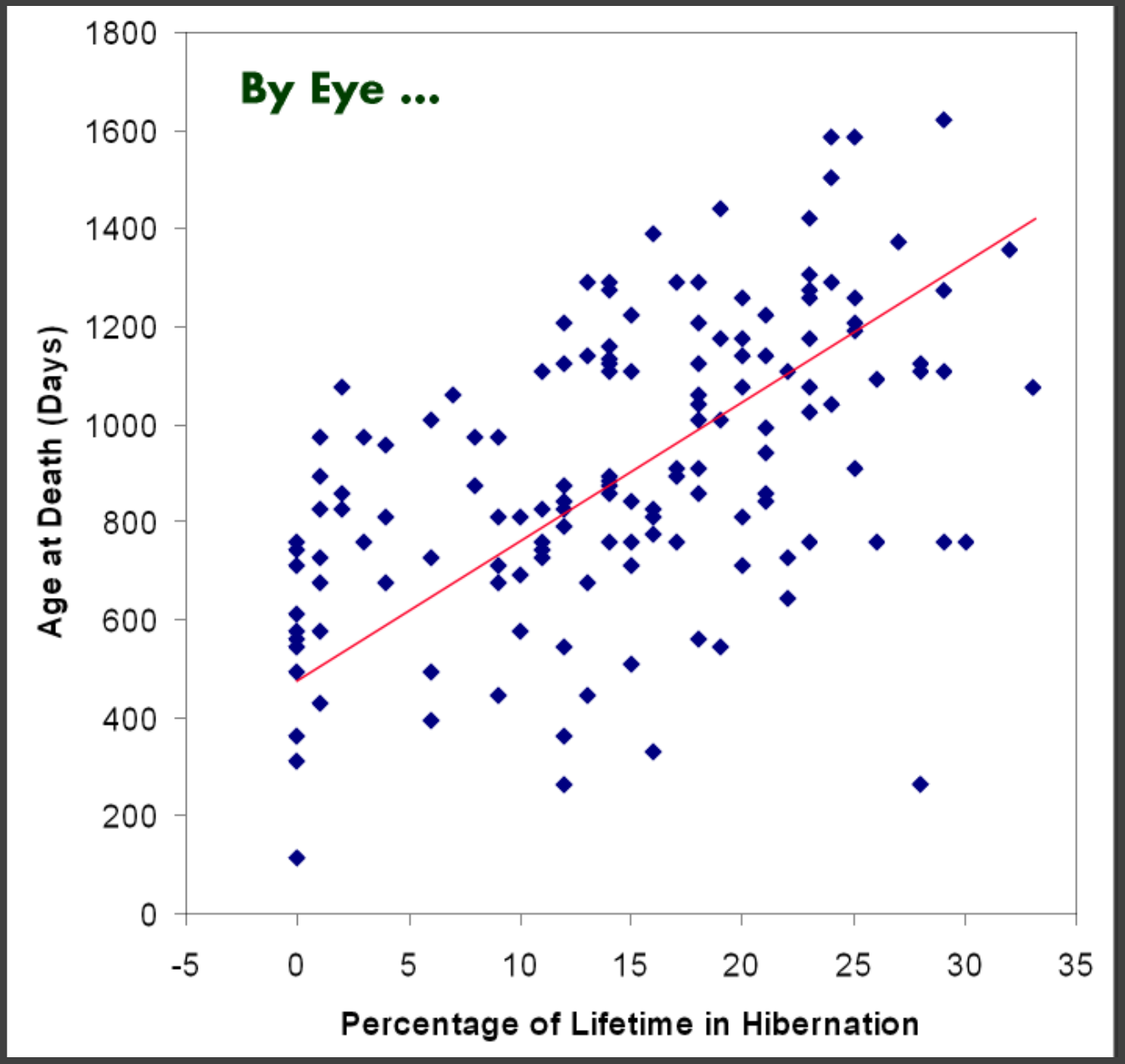
Trend Explorer



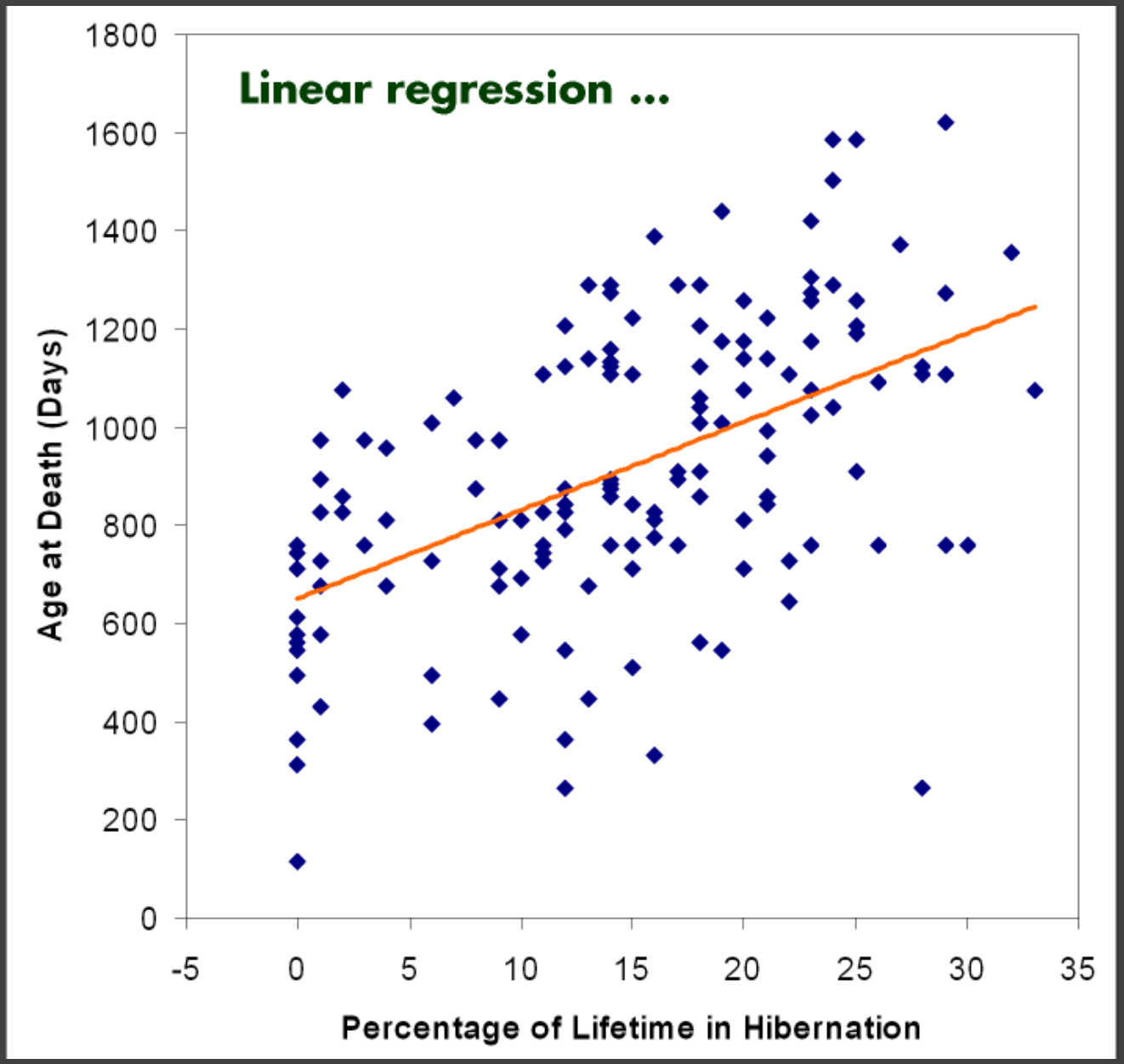




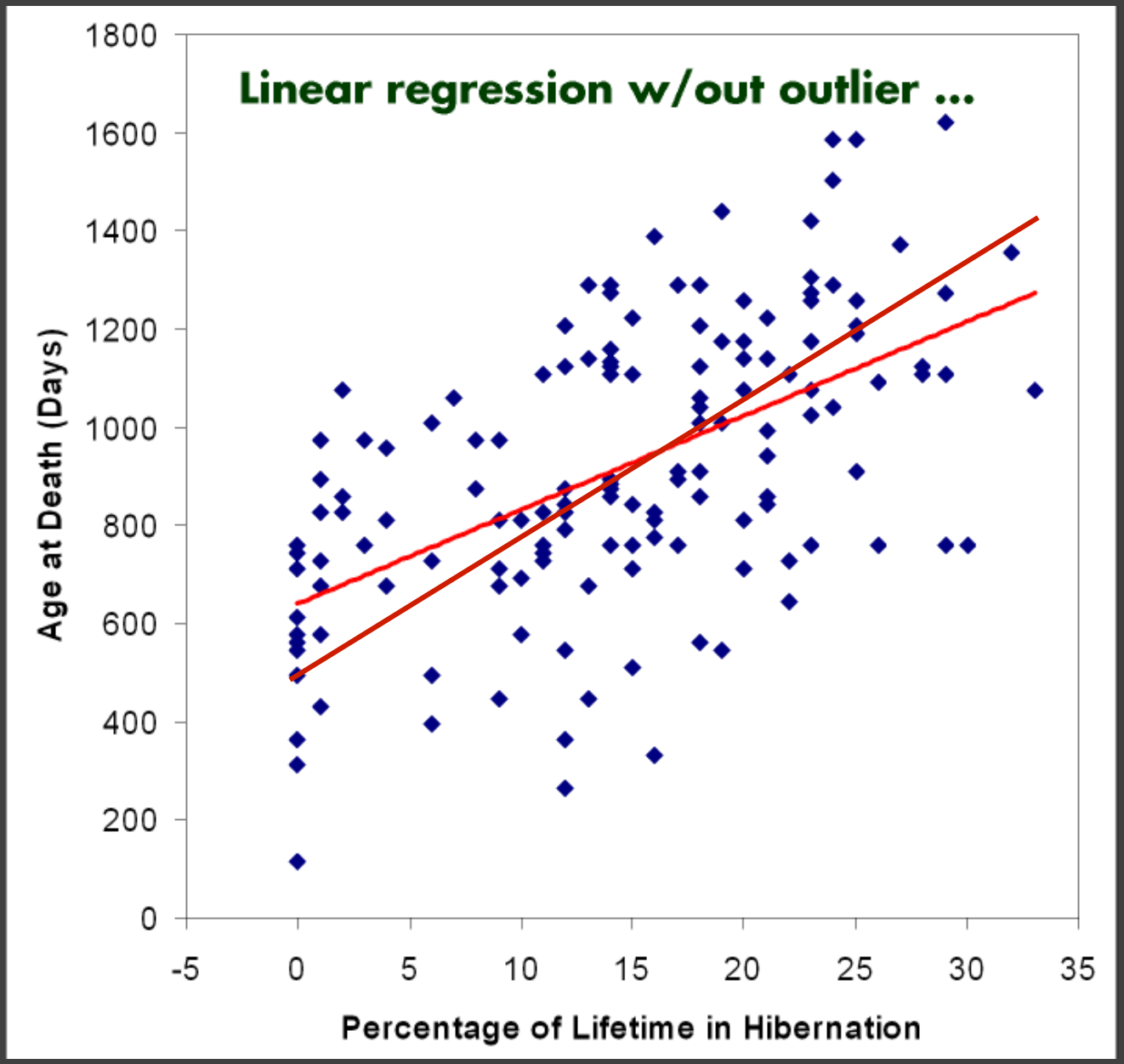
[The Elements of Graphing Data. Cleveland 94]



[The Elements of Graphing Data. Cleveland 94]



[The Elements of Graphing Data. Cleveland 94]



[The Elements of Graphing Data. Cleveland 94]