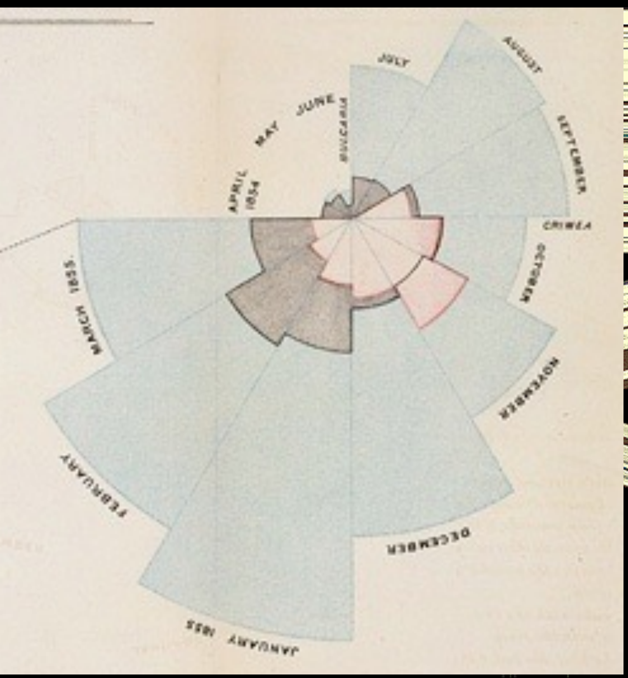
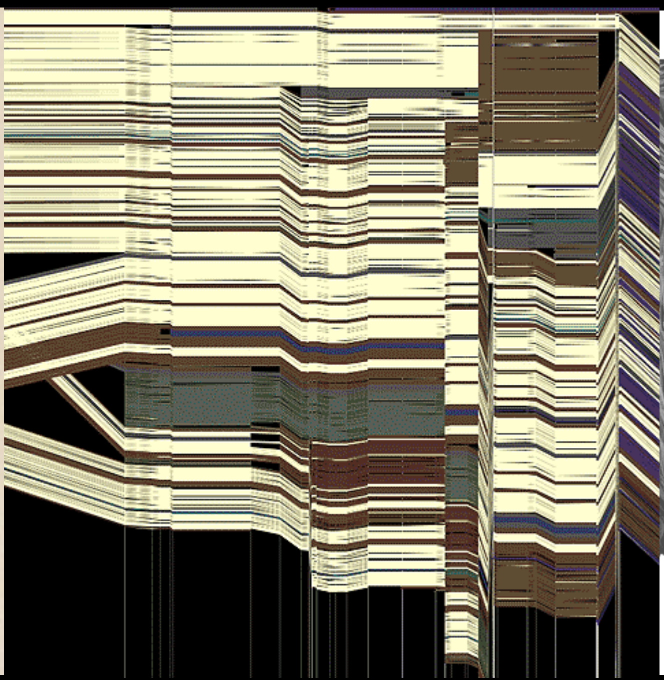
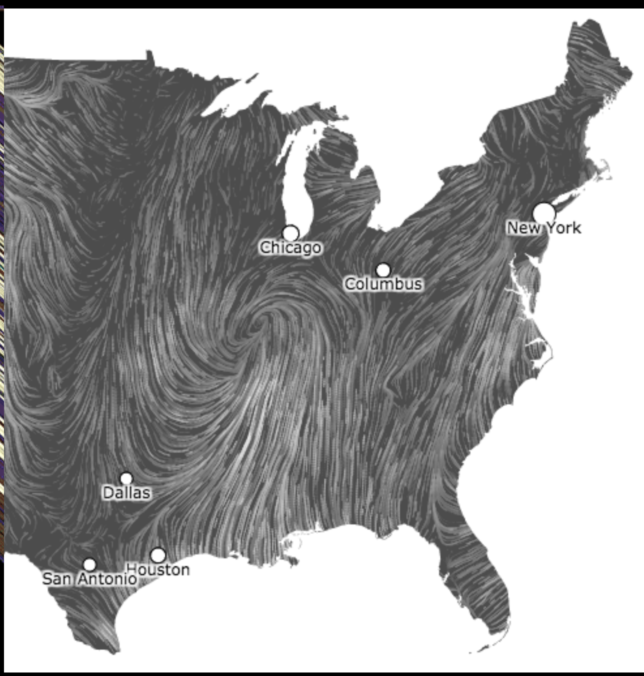
CSE 442 - Data Visualization

Using Space Effectively







Jeffrey Heer University of Washington

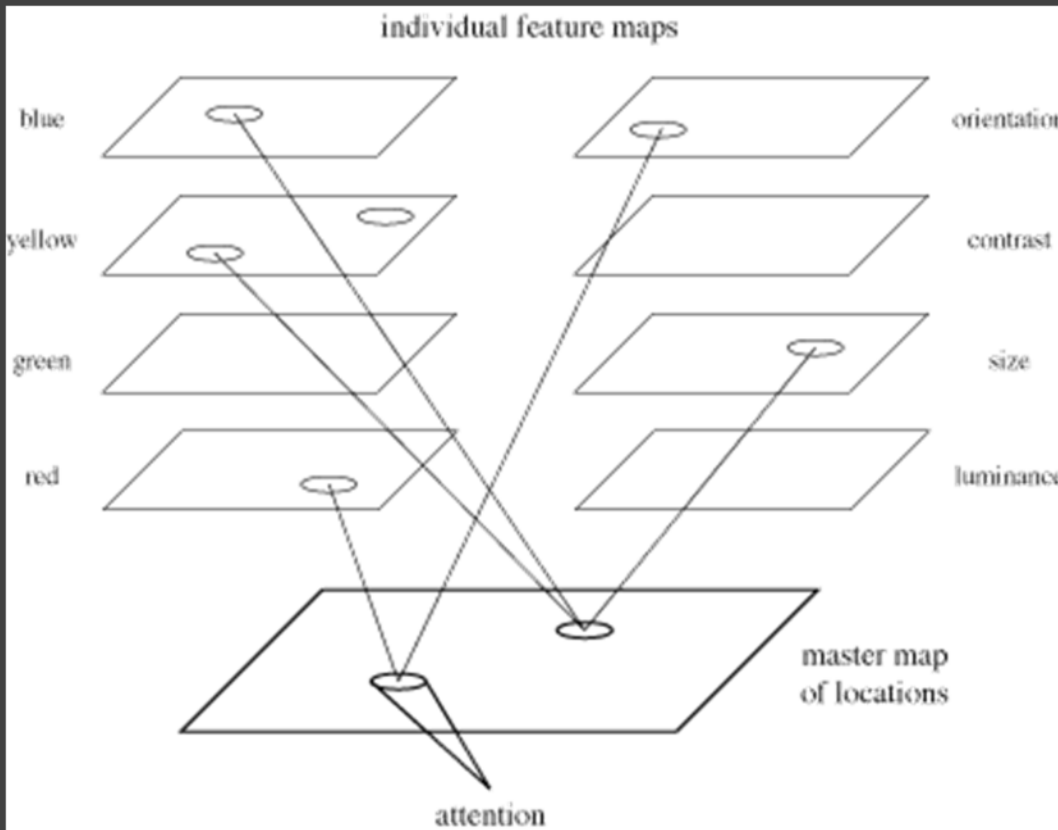
Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

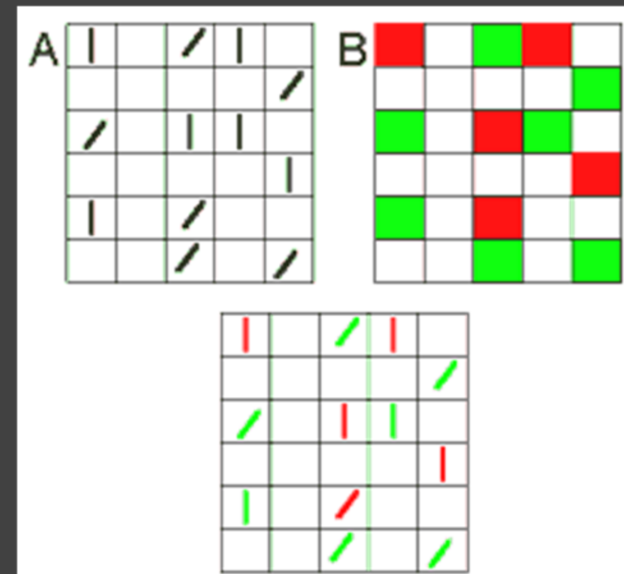
Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE	ORDINAL	NOMINAL
Position ·············	**Position** ·············	**Position**
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Feature Integration Theory



Treisman's feature integration model [Healey 04]



Feature maps for orientation & color [Green]

Indexed by *Position!*

Space (x, y) is the most important encoding channel.

But are you in the *right* space?

Topics

Scales & Axes

Data Space, Model Space

Optimizing Chart Design

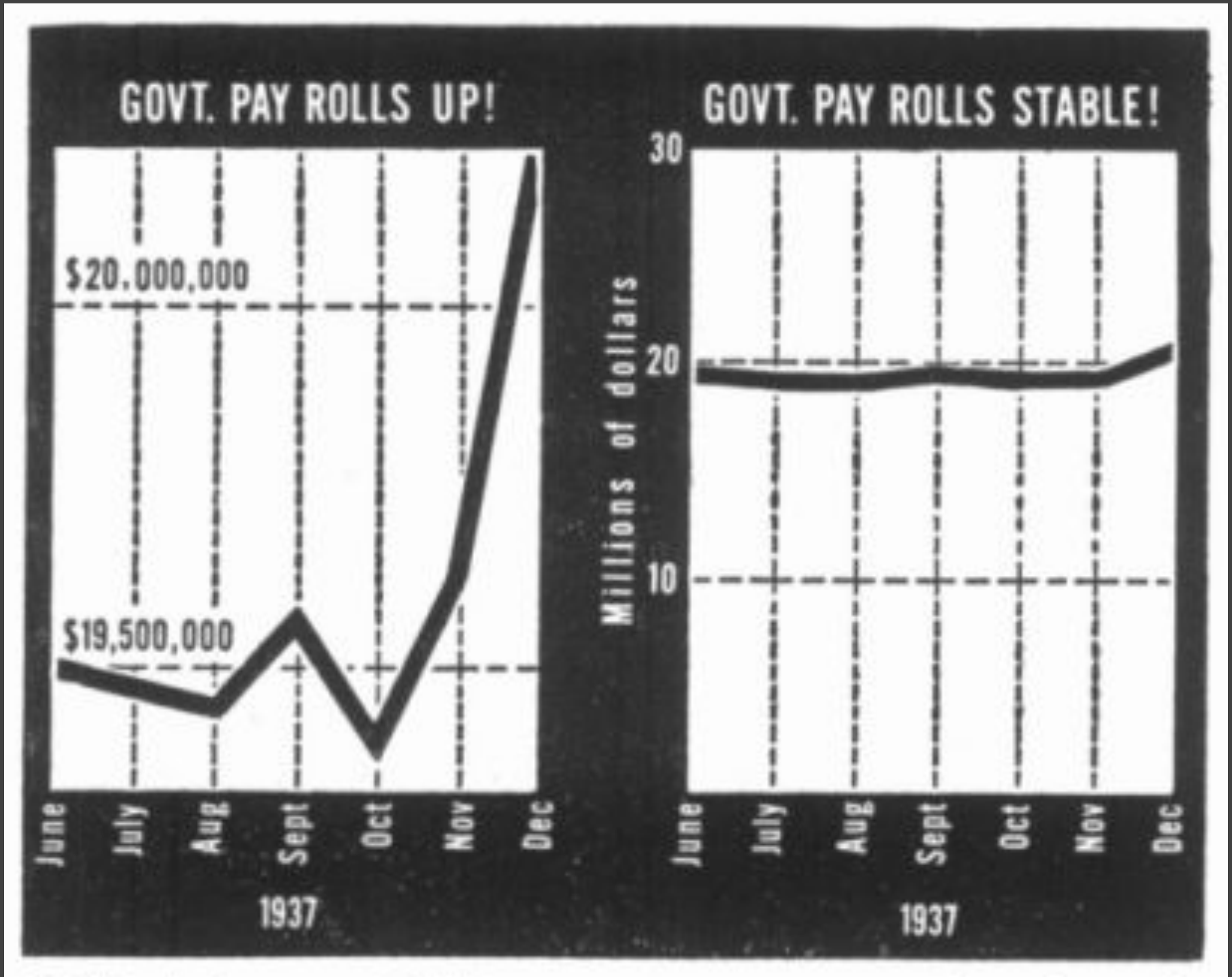
Zooming

Focus + Context

Dimensionality Reduction

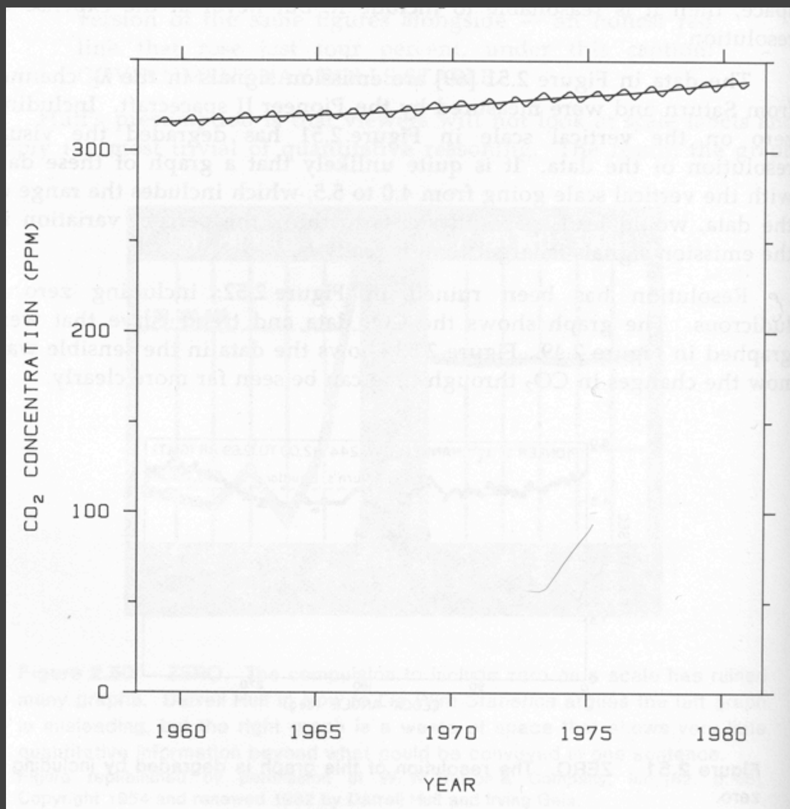
Scales & Axes

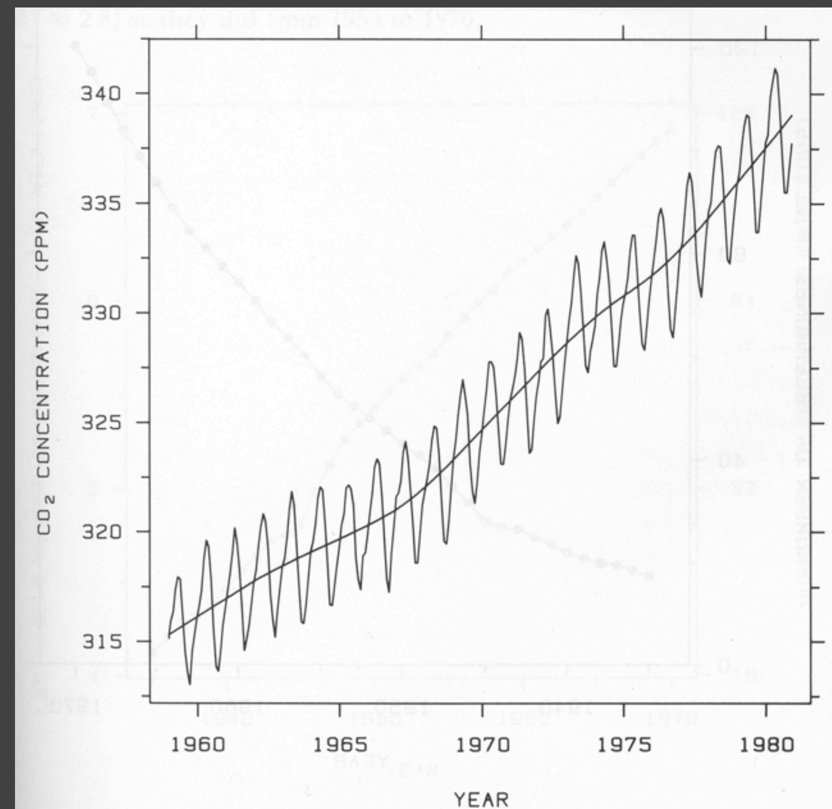
Include Zero in Axis Scale?



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

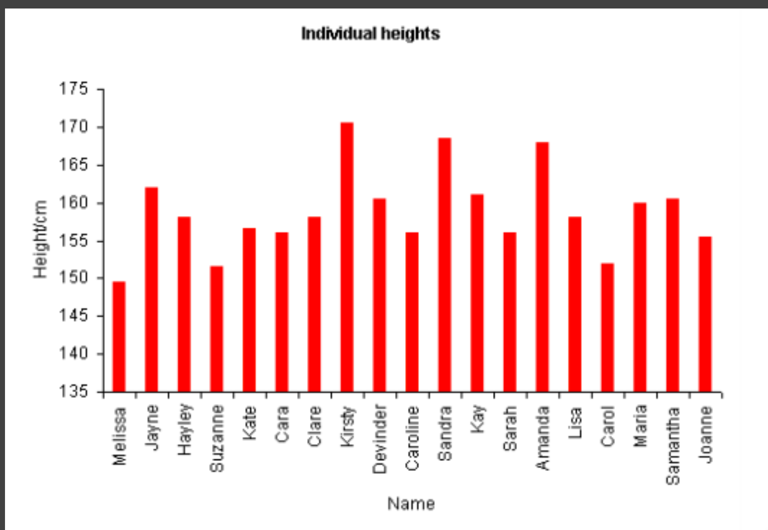
Include Zero in Axis Scale?





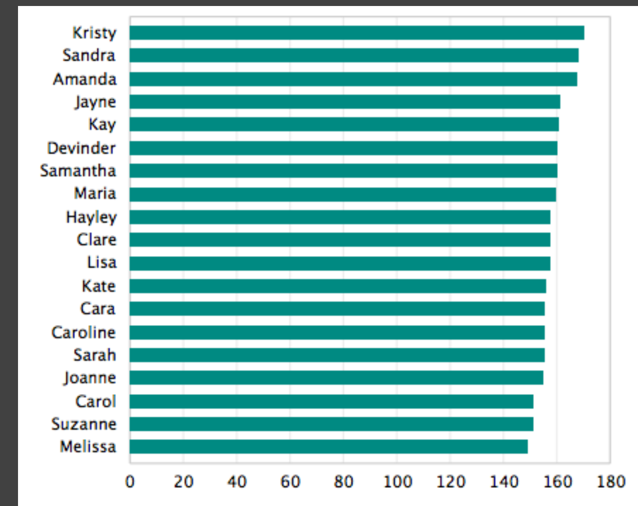
Yearly CO_2 concentrations [Cleveland 85]

Include Zero in Axis Scale?

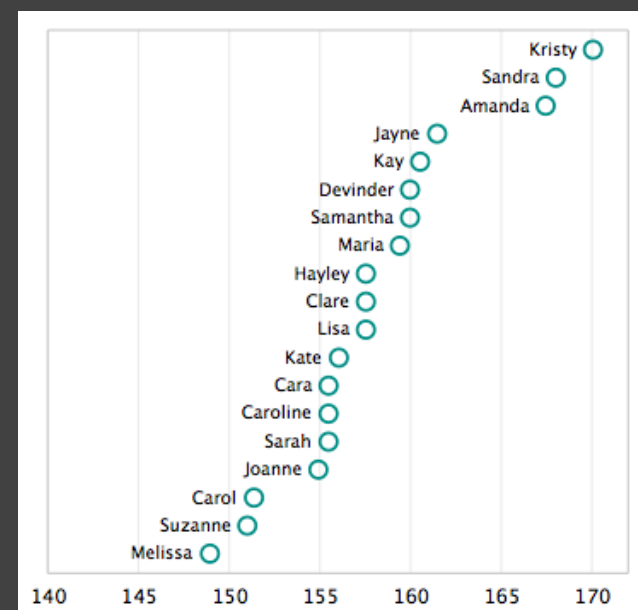


Violates Expressiveness Principle!

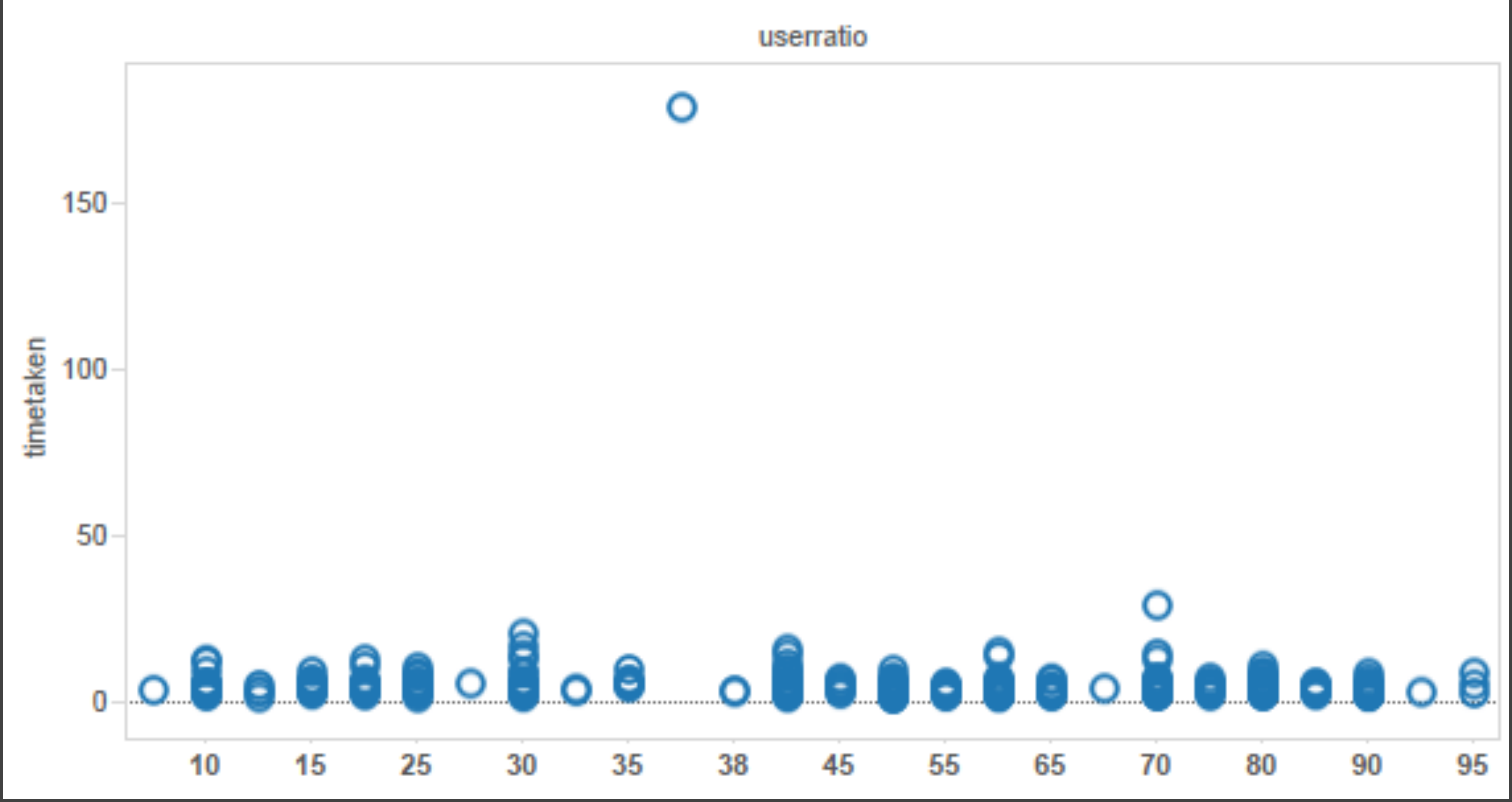
Compare Proportions (Q-Ratio)



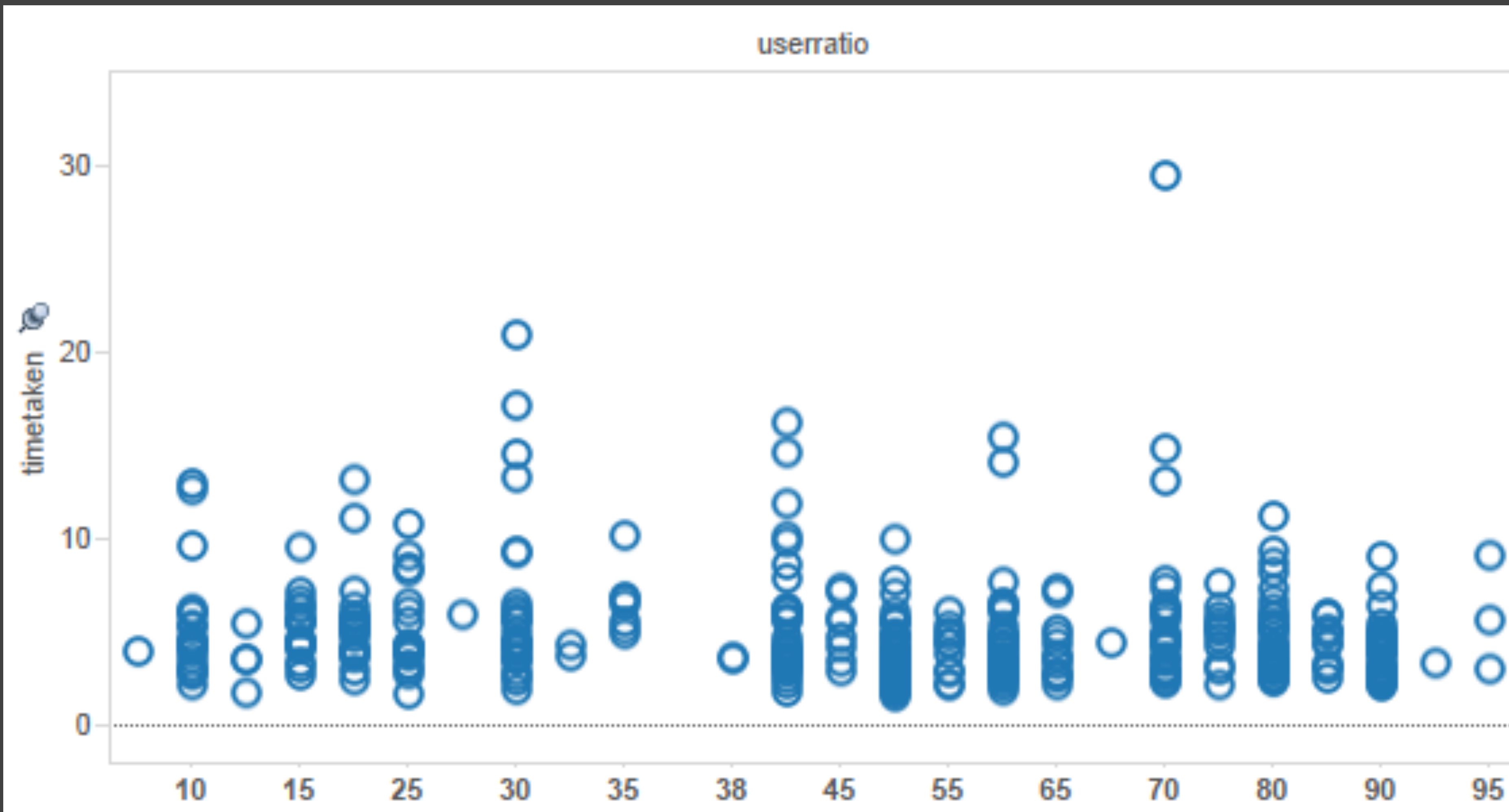
Compare Relative Position (Q-Interval)



How to Scale the Axis?

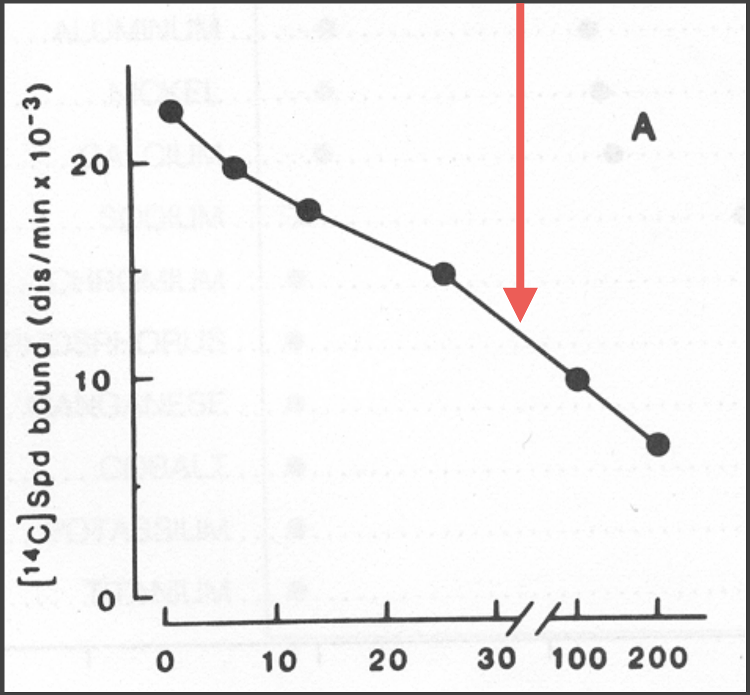


One Option: Clip Outliers

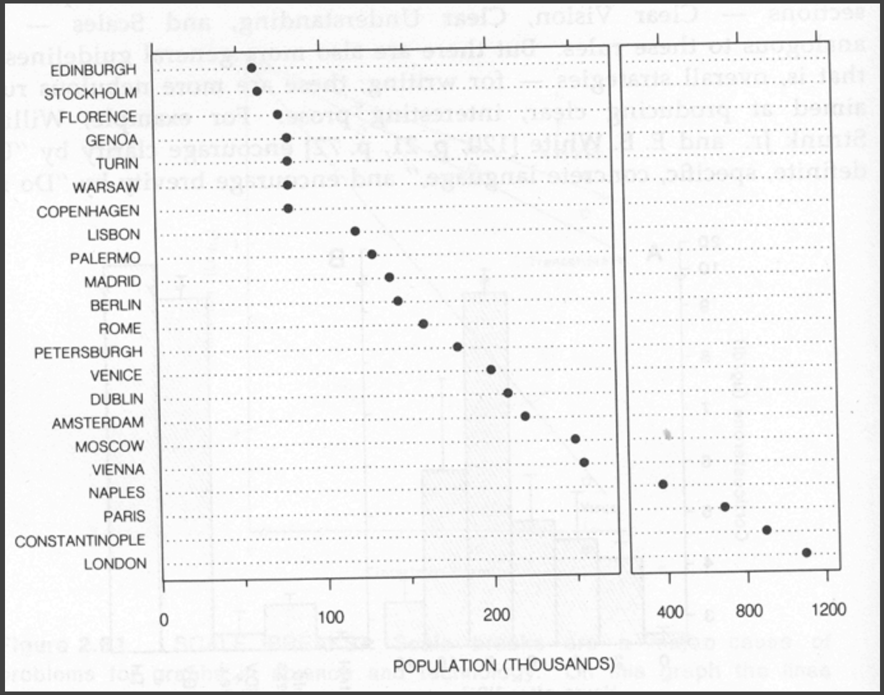


Clearly Mark Scale Breaks

Violates Expressiveness Principle!

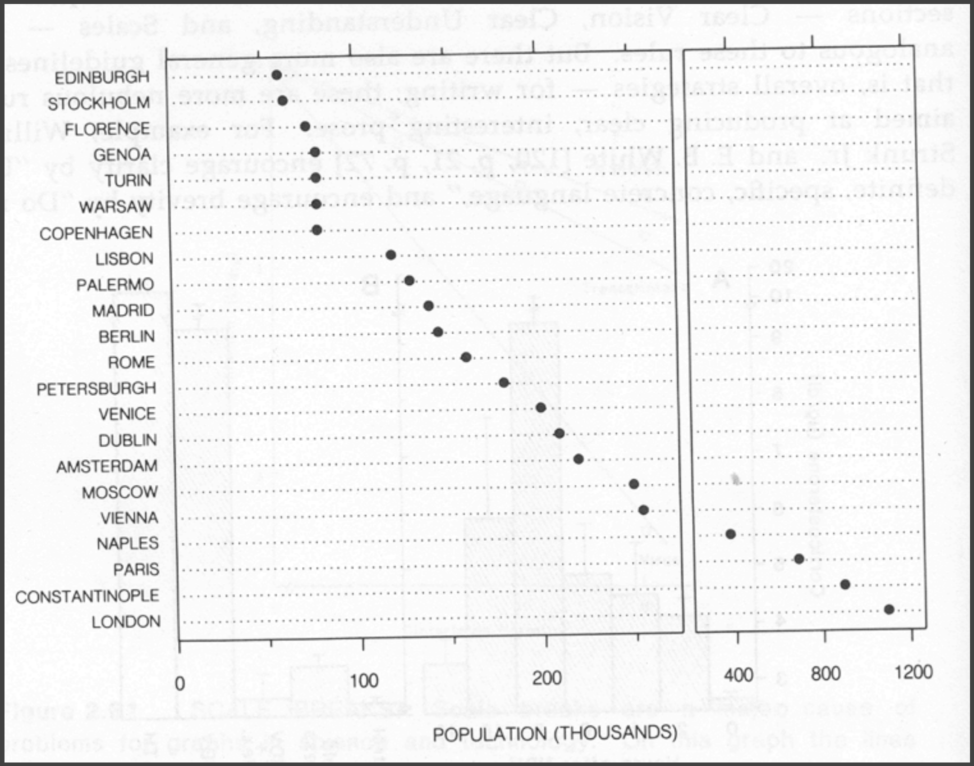


Poor scale break [Cleveland 85]

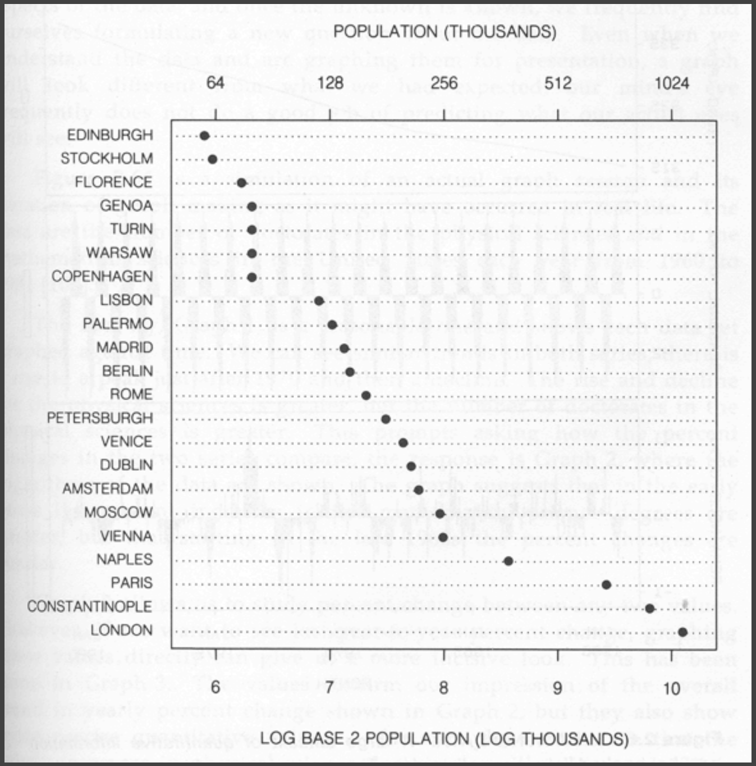


Well-marked scale break [Cleveland 85]

Scale Break vs. Log Scale



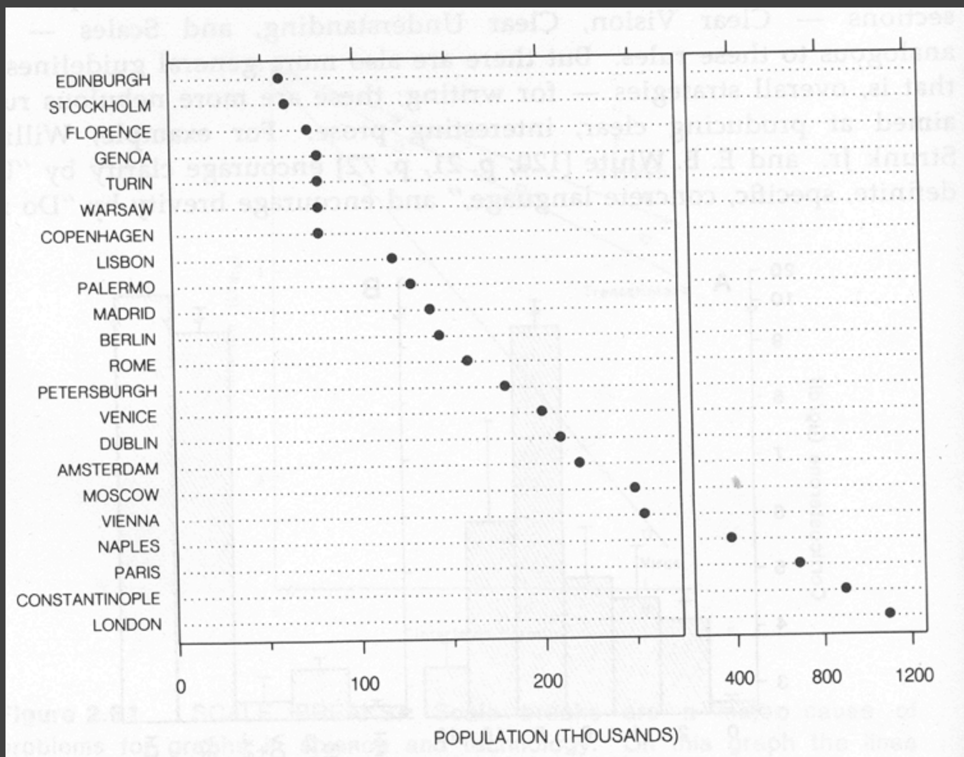
Scale Break

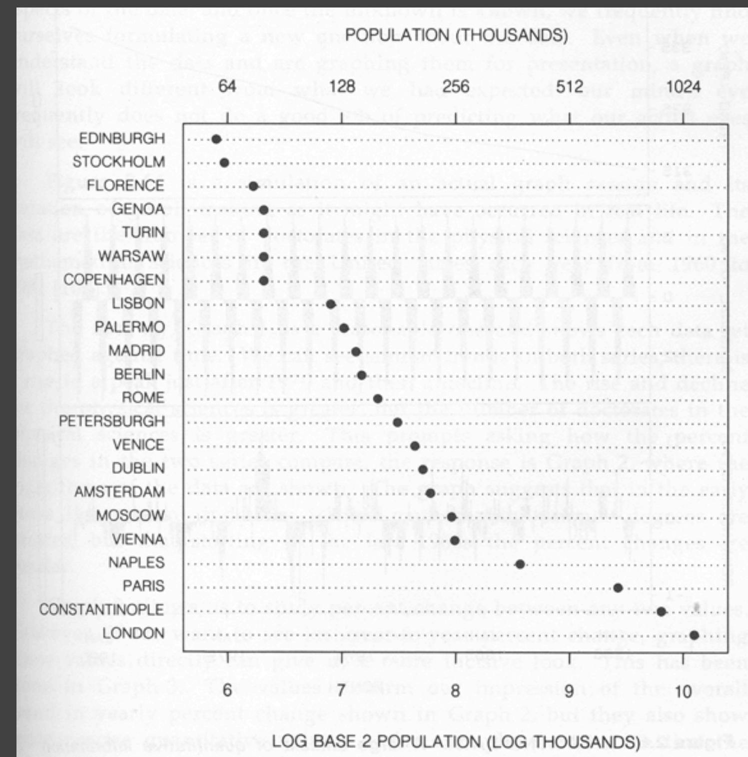


Log Scale

[Cleveland 85]

Scale Break vs. Log Scale





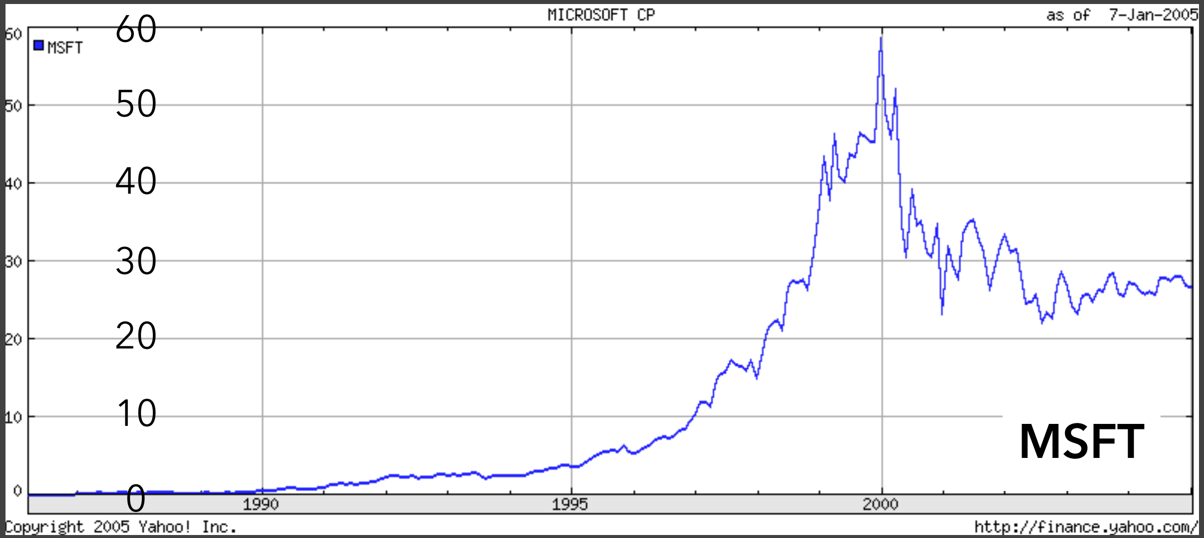
Both increase visual resolution

Scale break: difficult to compare (*cognitive* – not *perceptual* – work)

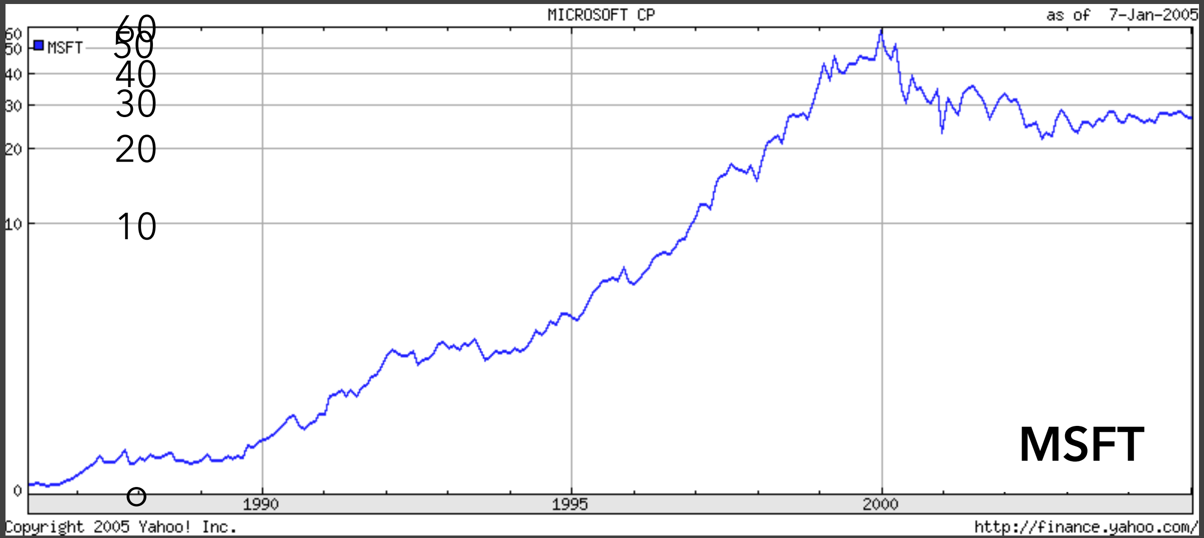
Log scale: direct comparison of all data

Linear Scale vs. Log Scale

Linear Scale



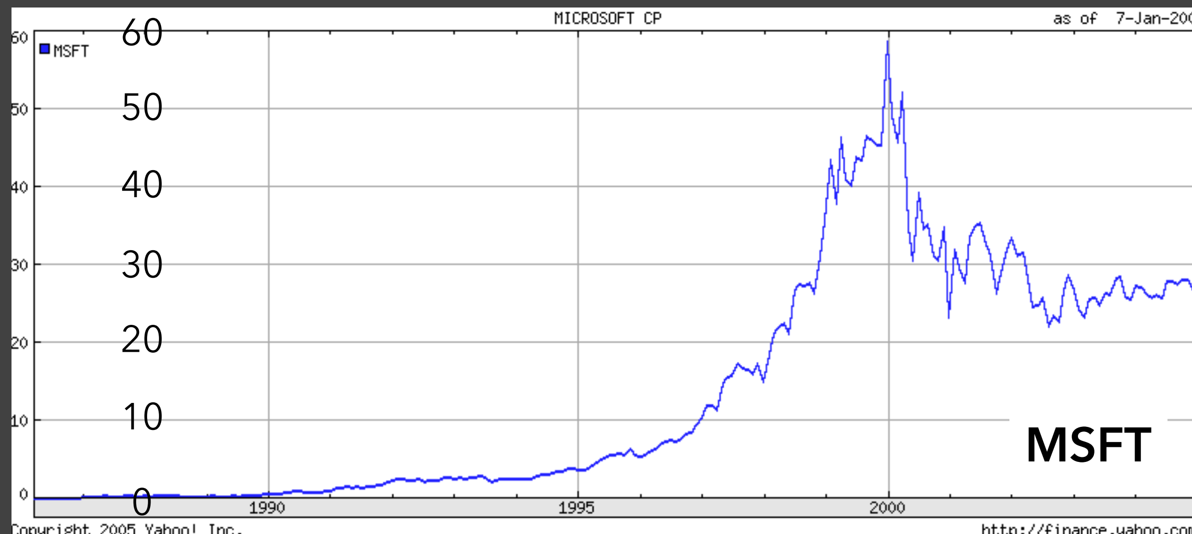
Log Scale



Linear Scale vs. Log Scale

Linear Scale

Absolute change

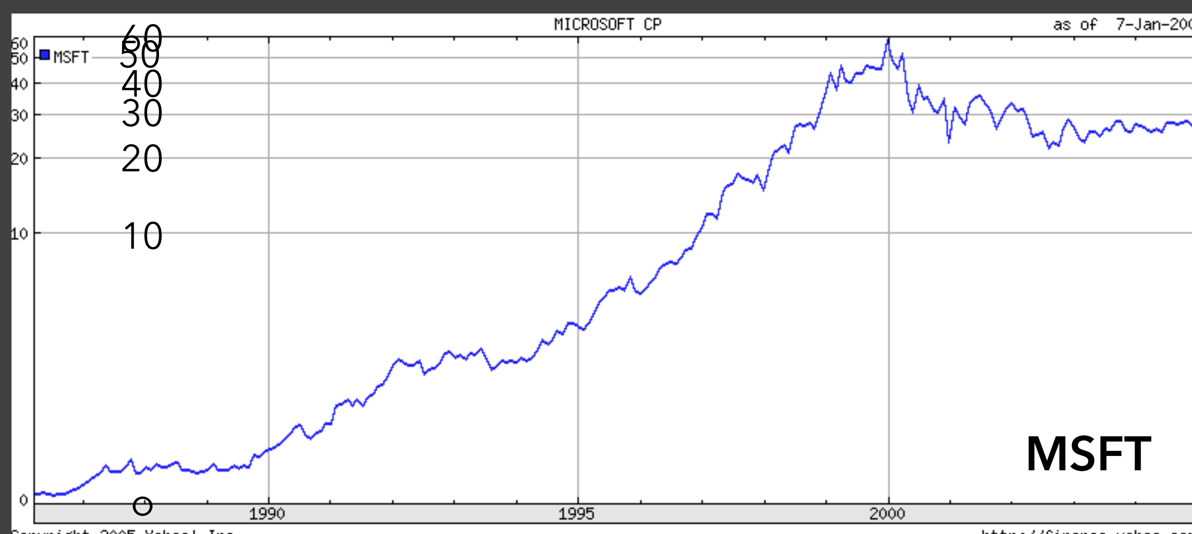


Log Scale

Small fluctuations

Percent change

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



When To Apply a Log Scale?

Address data skew (e.g., dominant outliers)

Enables comparison within and across multiple orders of magnitude.

Focus on multiplicative factors (not additive)

Recall that the logarithm transforms **×** to **+**!

Percentage change, not absolute value.

Constraint: **positive, non-zero values**

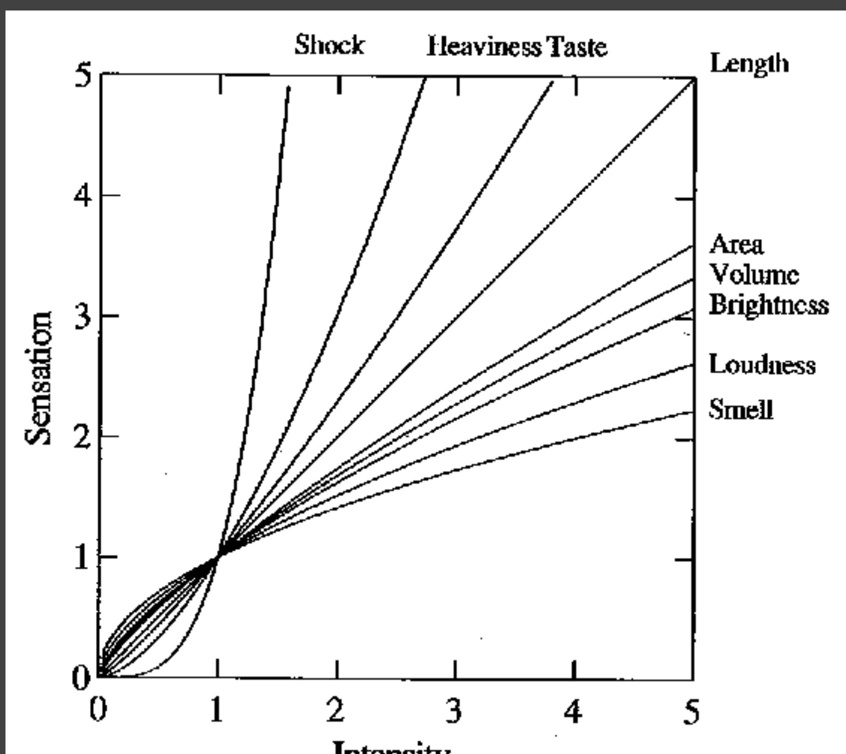
Constraint: **audience familiarity?**

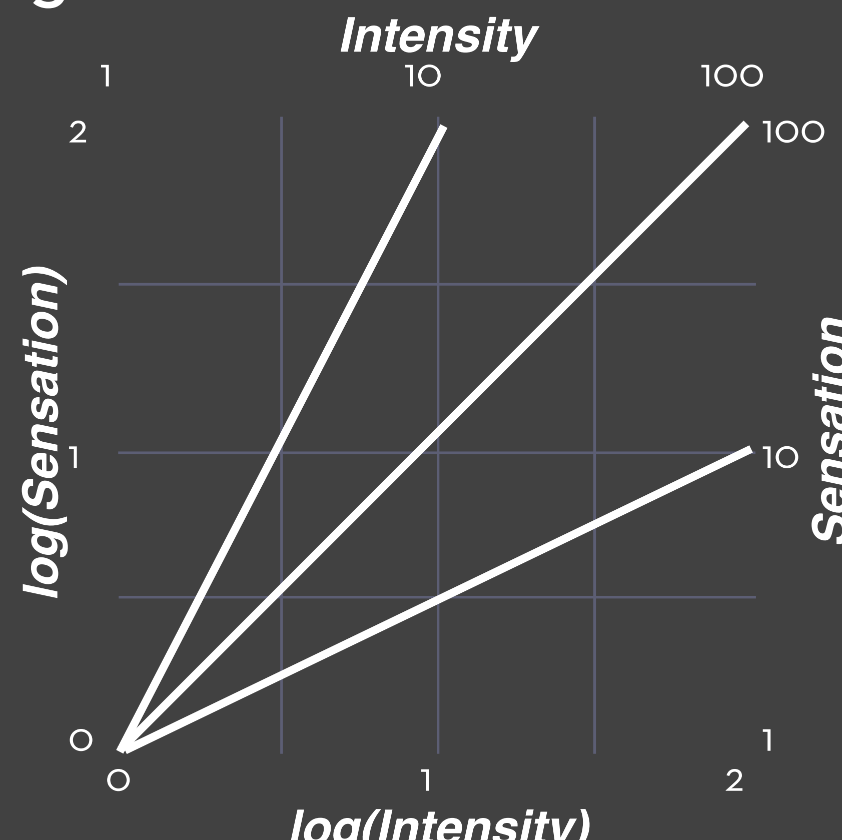
Log-Log Plot

Power functions ($\mathbf{y = kx^a}$) transform into lines

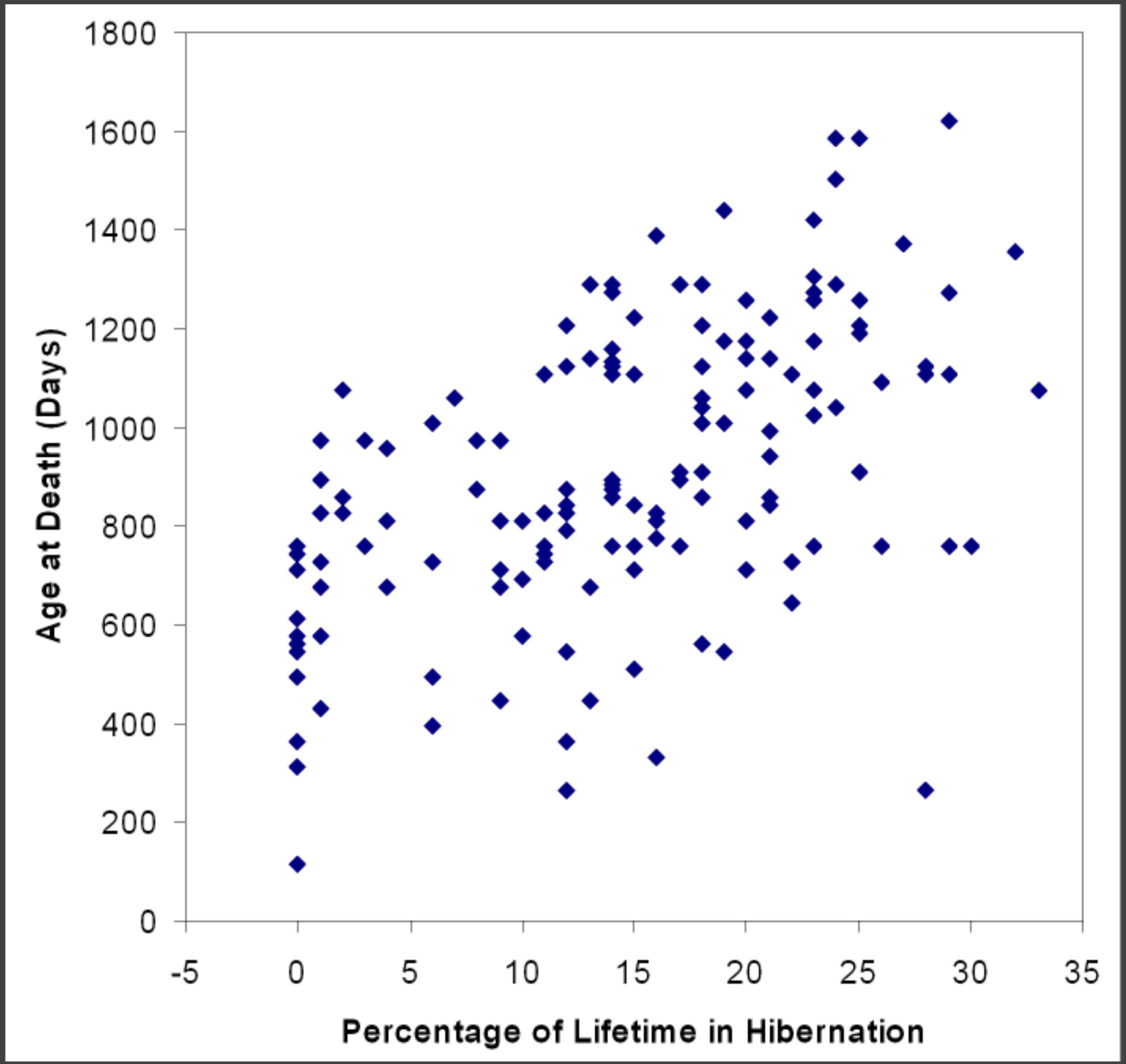
Example - Steven's Power Law:

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

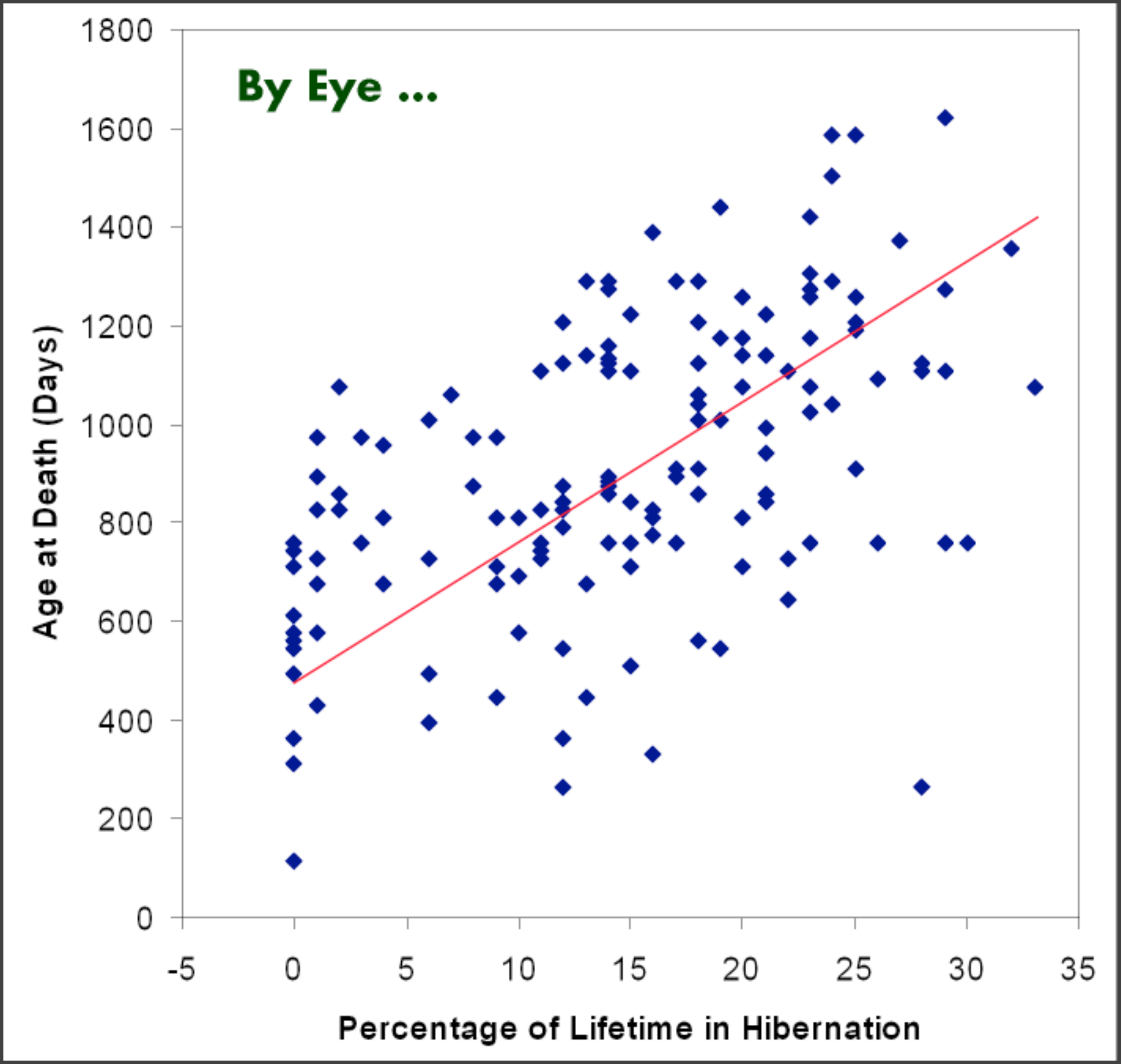




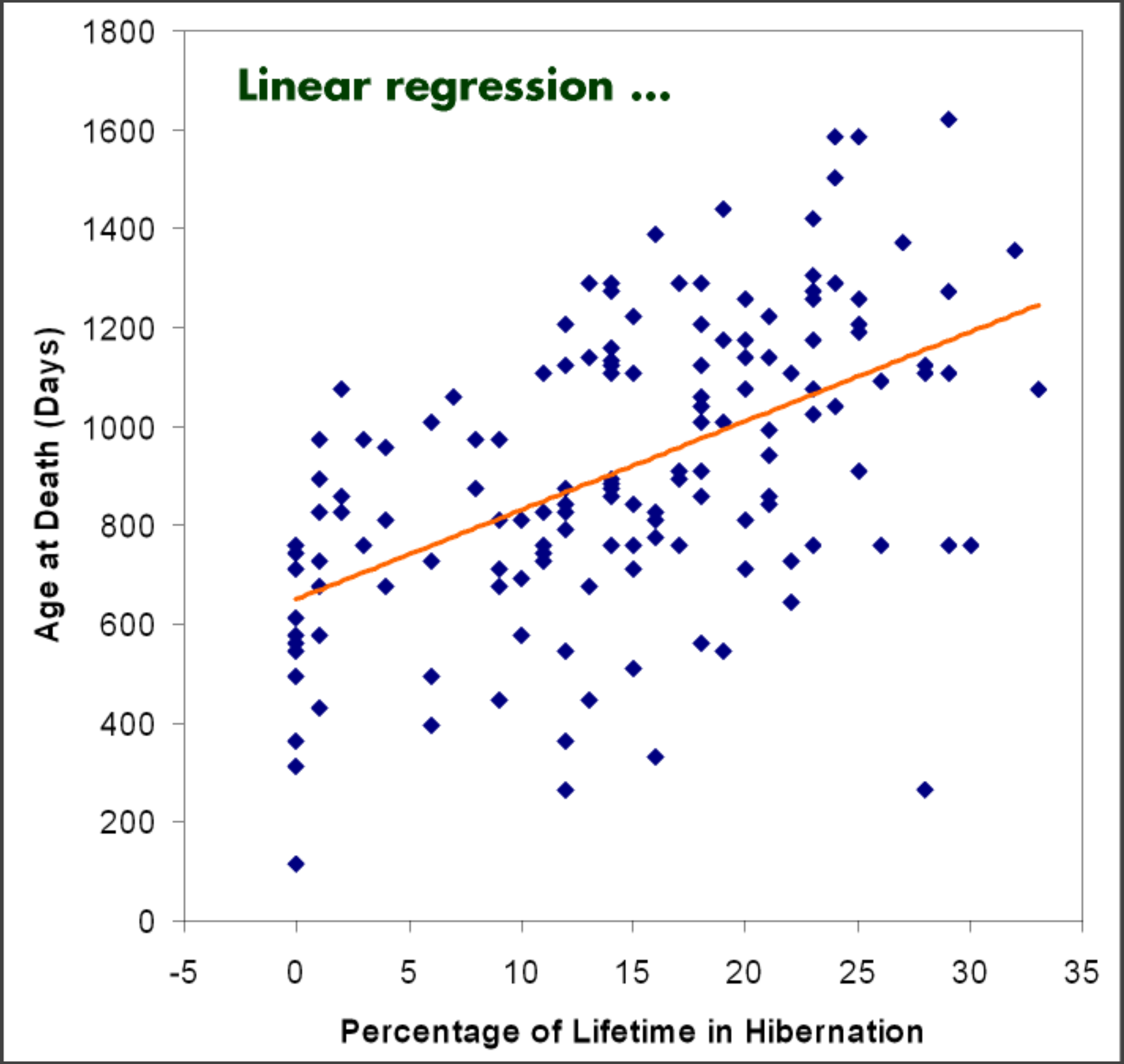
Data Space, Model Space



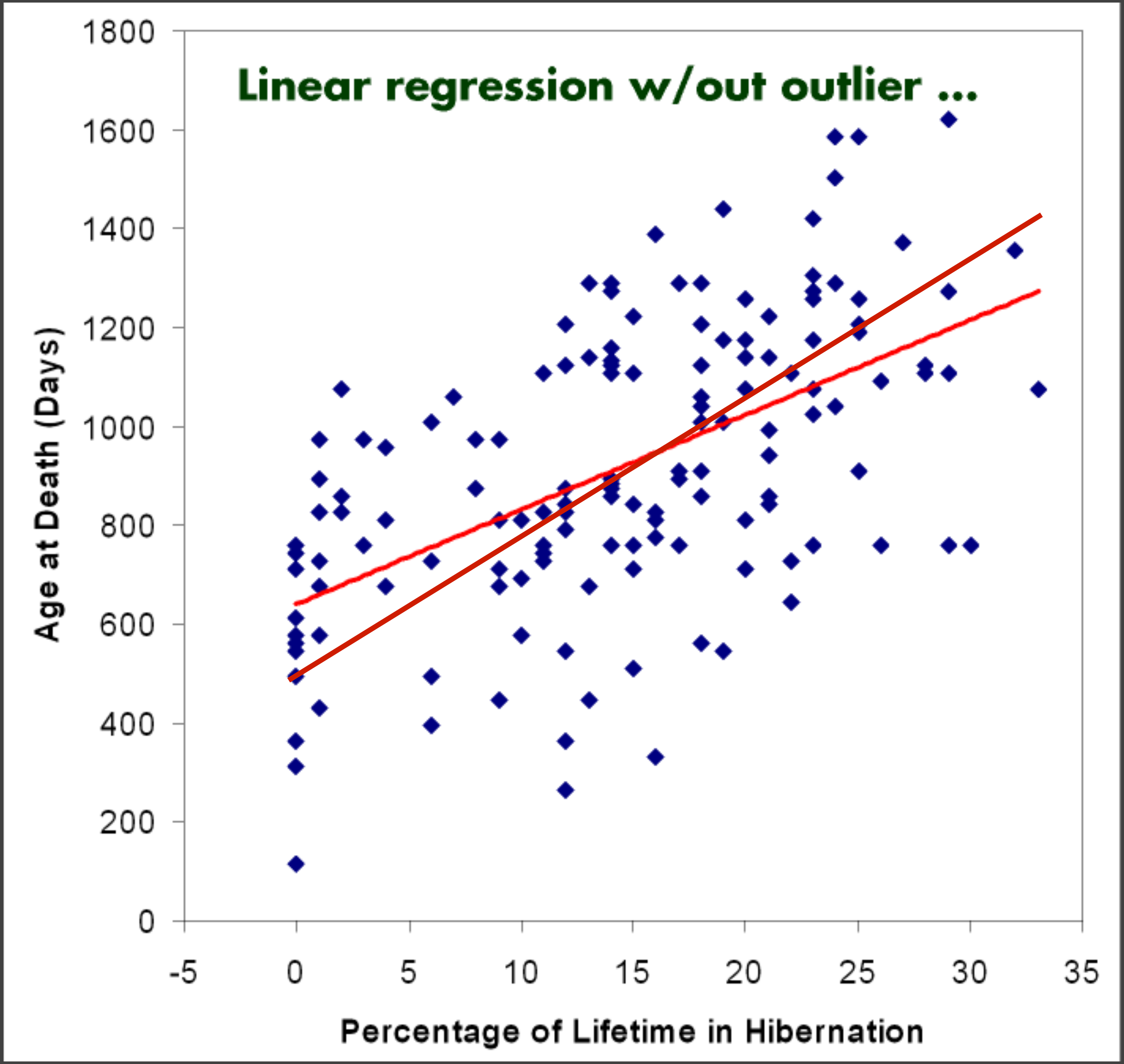
[The Elements of Graphing Data. Cleveland 94]



[The Elements of Graphing Data. Cleveland 94]



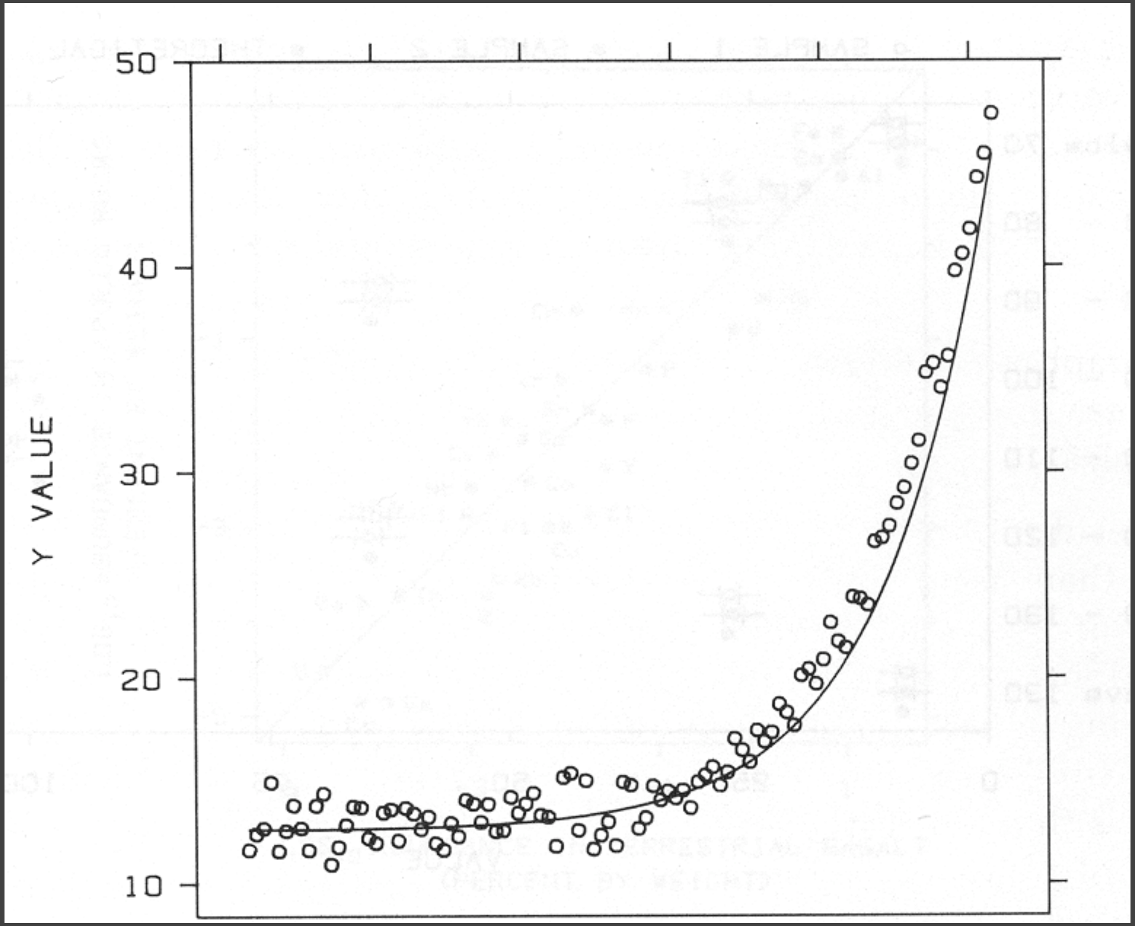
[The Elements of Graphing Data. Cleveland 94]



[The Elements of Graphing Data. Cleveland 94]

Transforming Data

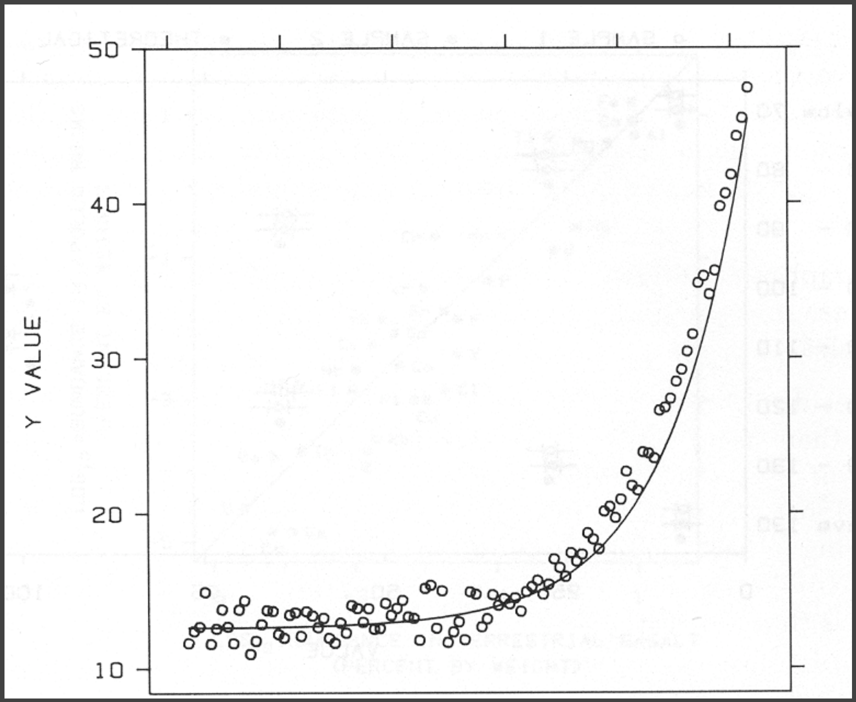
How well does the curve fit the data?

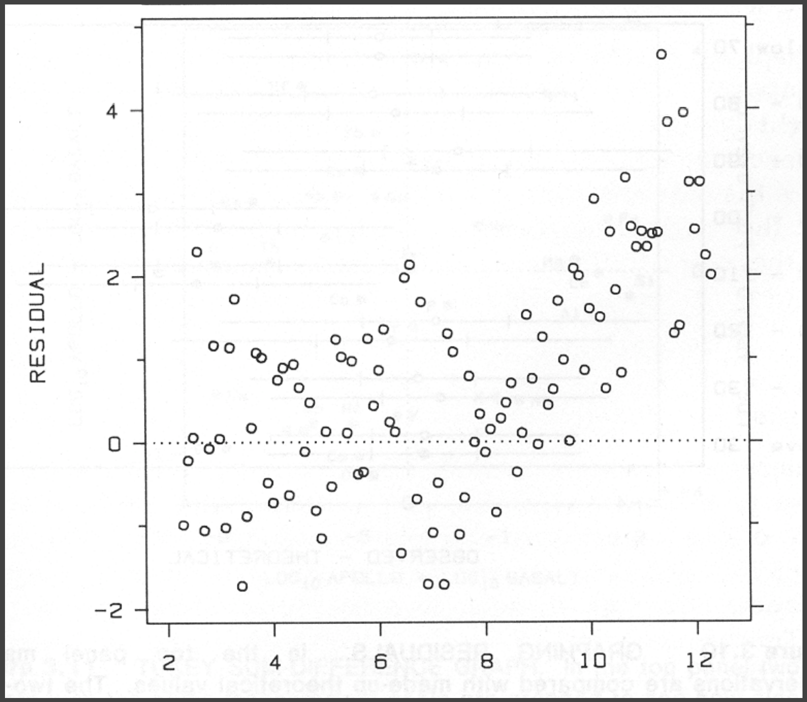


[Cleveland 85]

Plot the Residuals

Plot vertical distance from best fit curve
Residual graph shows accuracy of fit

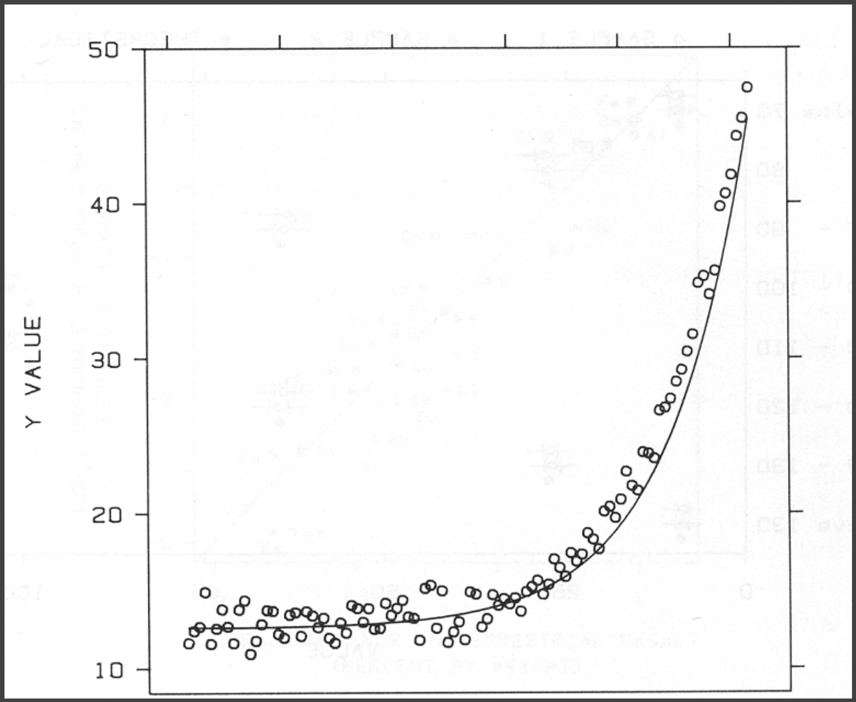




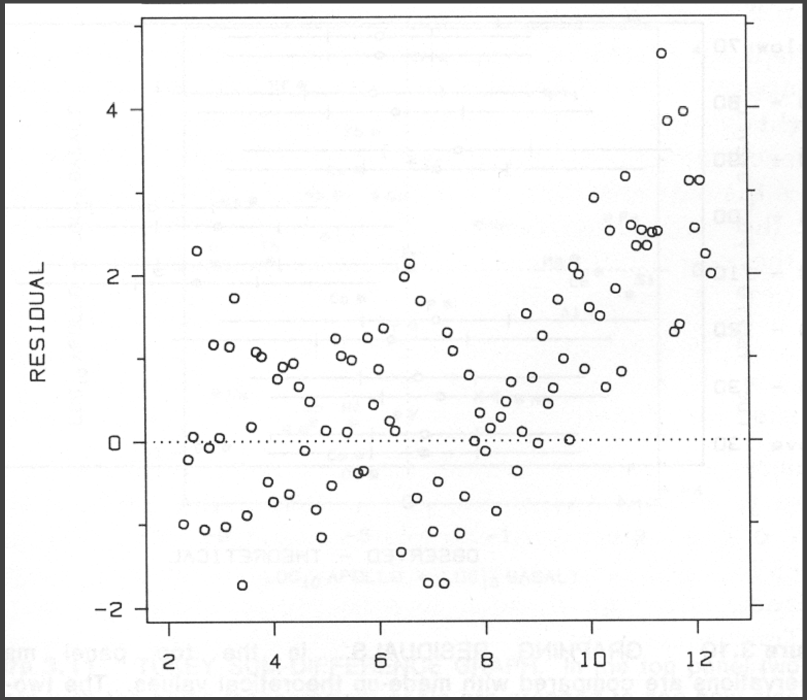
[Cleveland 85]

Multiple Plotting Options

Plot model in data space



Plot data in model space



[Cleveland 85]

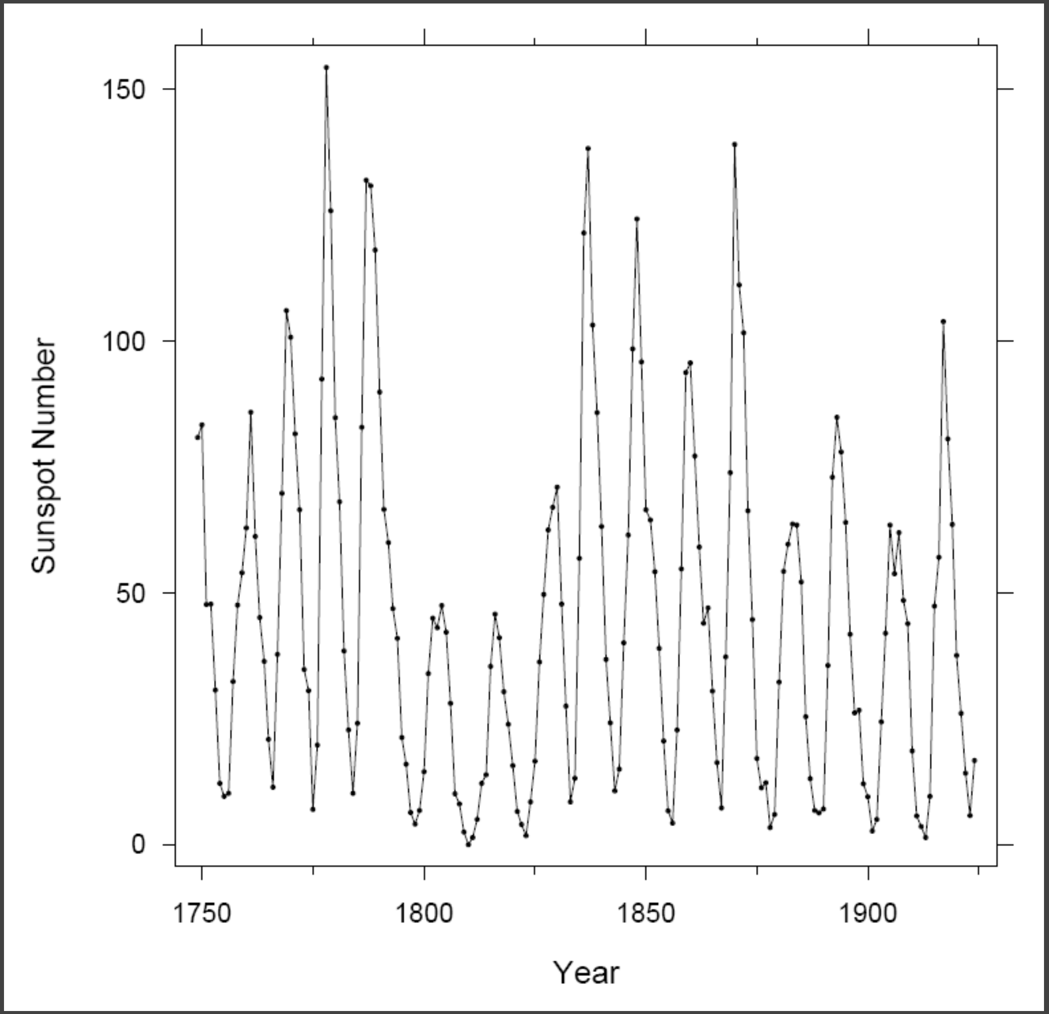
Optimizing Chart Design

Chart Design Parameters

Given a visual encoding (e.g., line chart), what aspects might affect graphical perception?

Physical Size
Aspect Ratio
Ticks, Labels, Gridlines
Line Width
Data Points (e.g., dots)

How might we determine optimized choices?



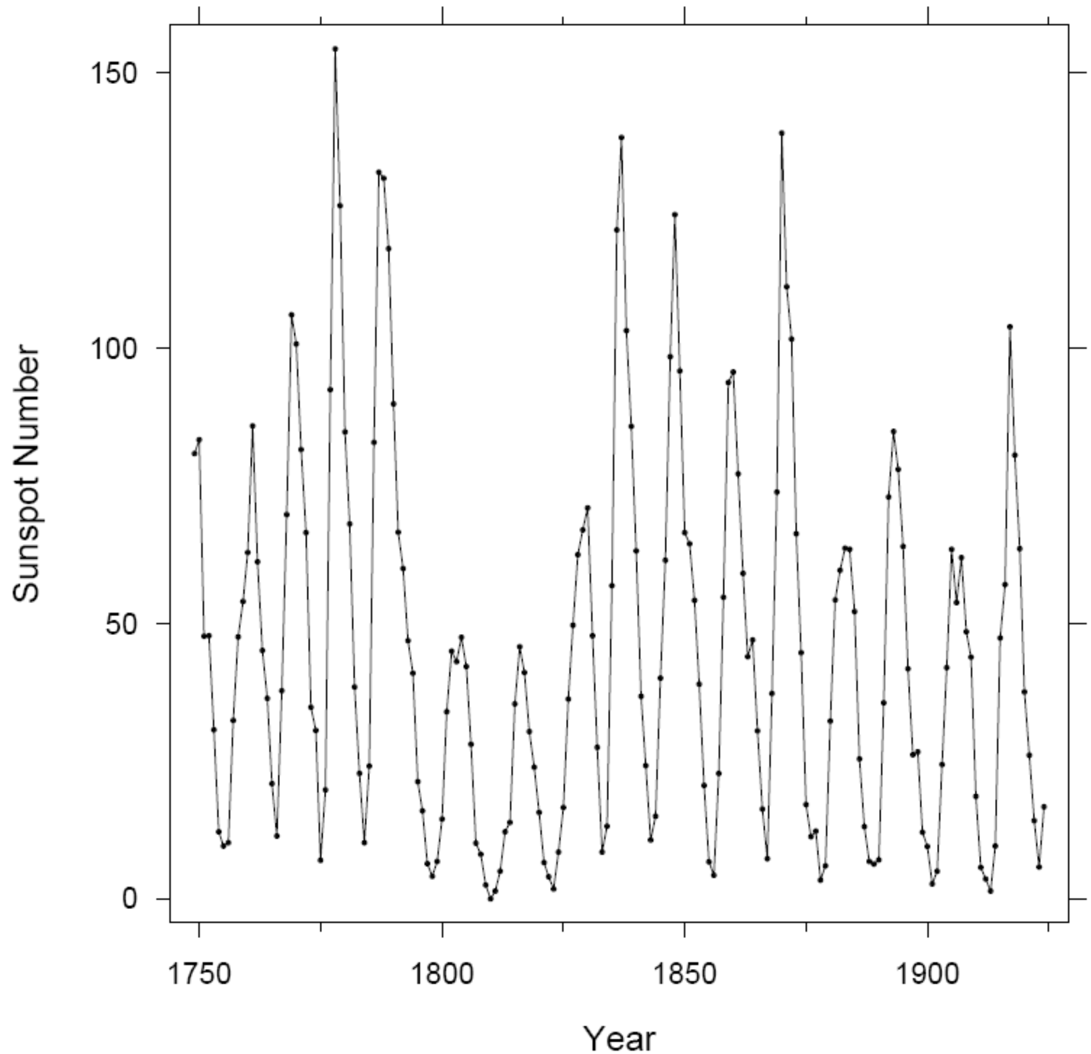
Optimization-Based Design

Determine *error* or *energy* functions for measuring the "quality" of a visualization.

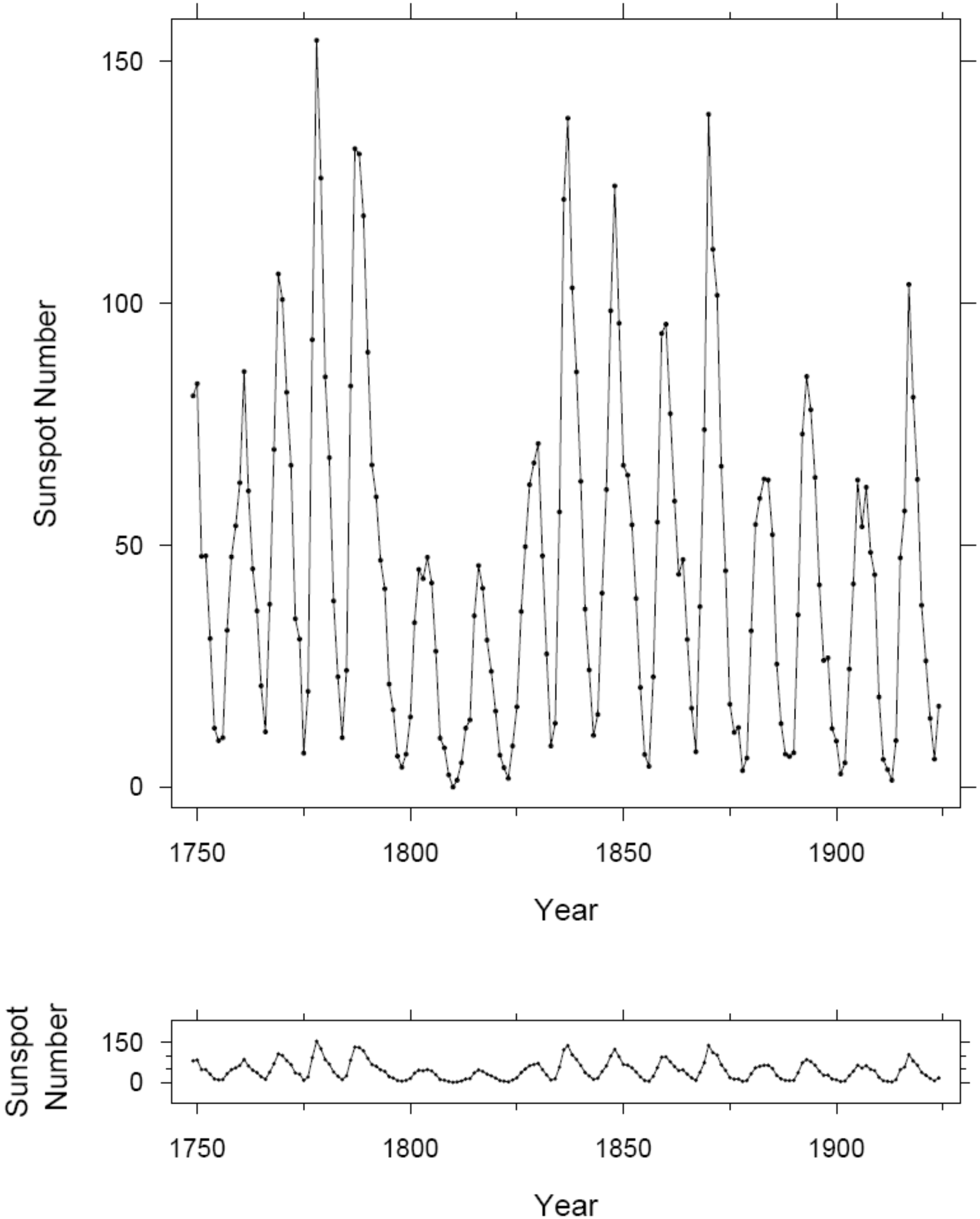
Treat as an optimization *objective* and then *solve* (or *search*) for better chart parameters.

Examples:
Determining chart aspect ratio
Selecting axis ticks
Streamgraph layout

Aspect Ratio



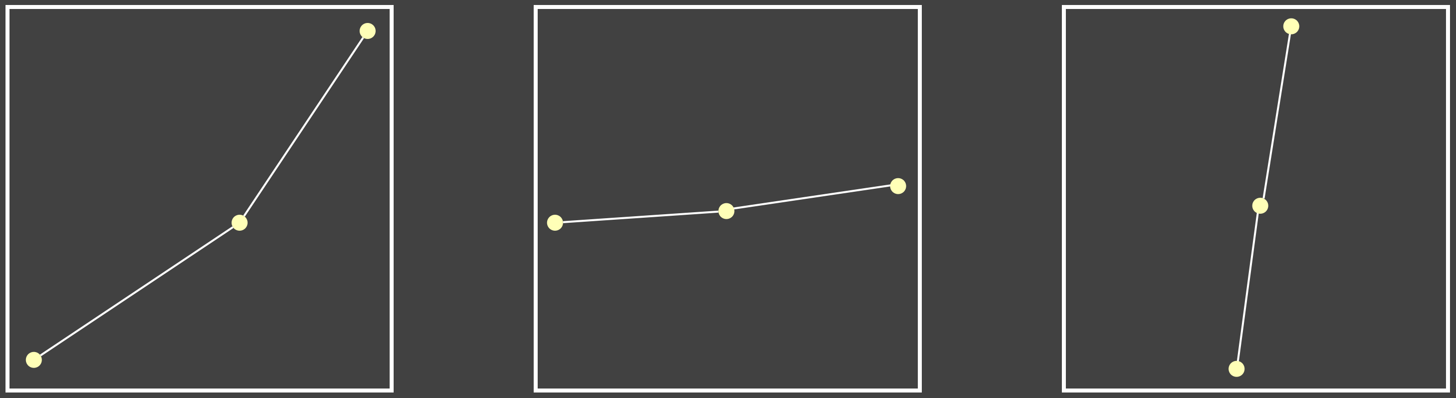
William S. Cleveland
The Elements of Graphing Data



William S. Cleveland
The Elements of Graphing Data

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°

Insight: to optimize the aspect ratio, bank to 45°

Aspect Ratio Banking Methods

Median-Absolute-Slope

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

Average-Absolute-Slope

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

Avg-Absolute-Orientation

Unweighted

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

Weighted

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

Max-Orientation-Resolution

Global (over all i, j s.t. i≠j)

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

Local (over adjacent segments)

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

Closed Form Solutions

Median-Absolute-Slope

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

Average-Absolute-Slope

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

Avg-Absolute-Orientation

Unweighted

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

Weighted

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

Max-Orientation-Resolution

Global (over all i, j s.t. i≠j)

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

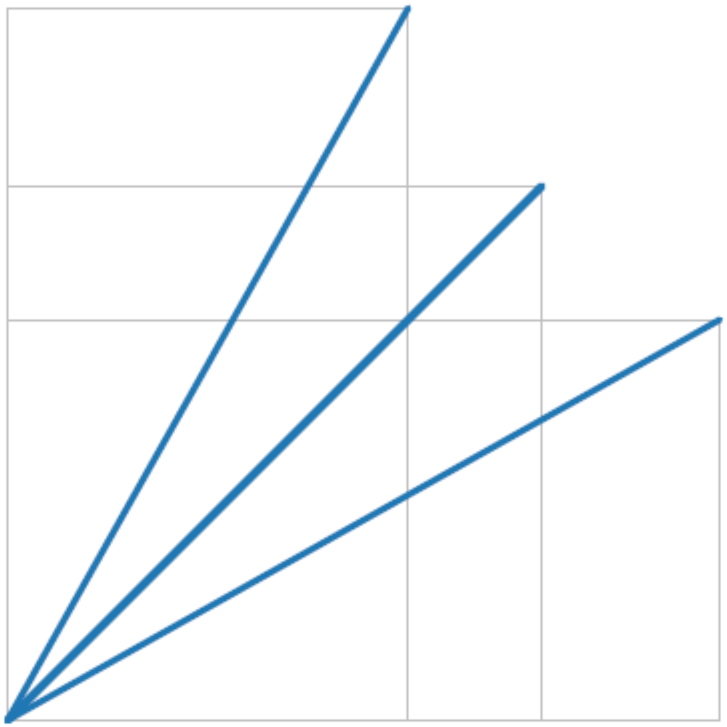
Local (over adjacent segments)

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

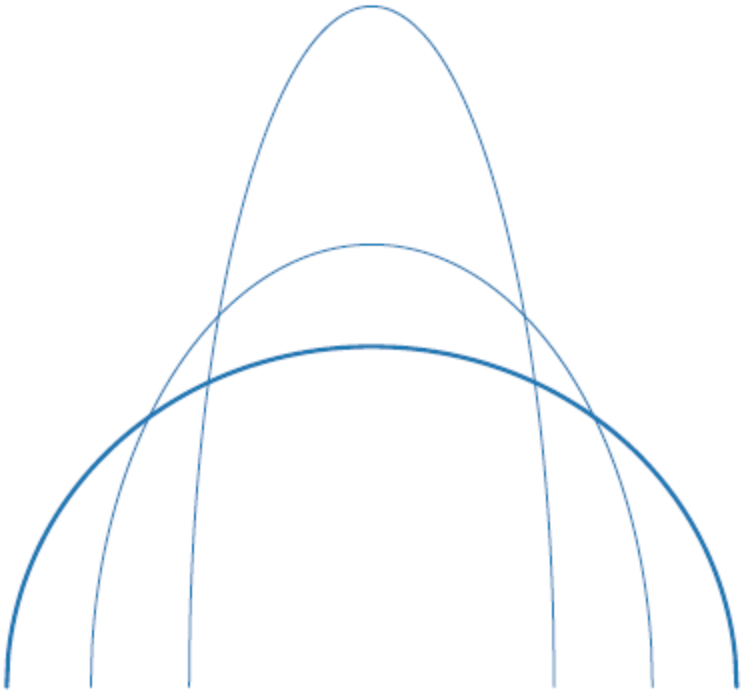
Requires Iterative Optimization

Alternative: Minimize Arc Length

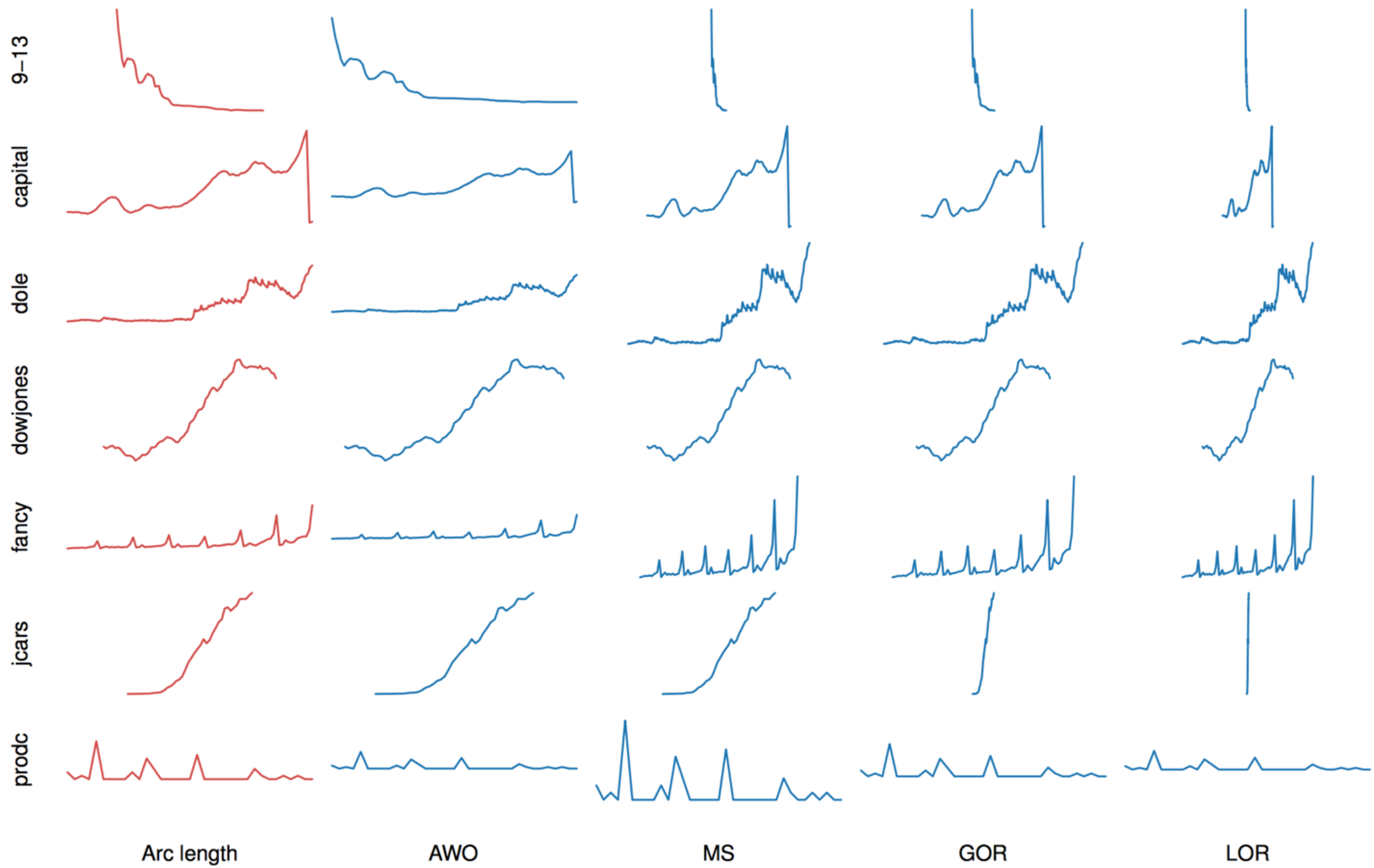
while holding area constant [Talbot et al. 2011]



Straight line -> 45°

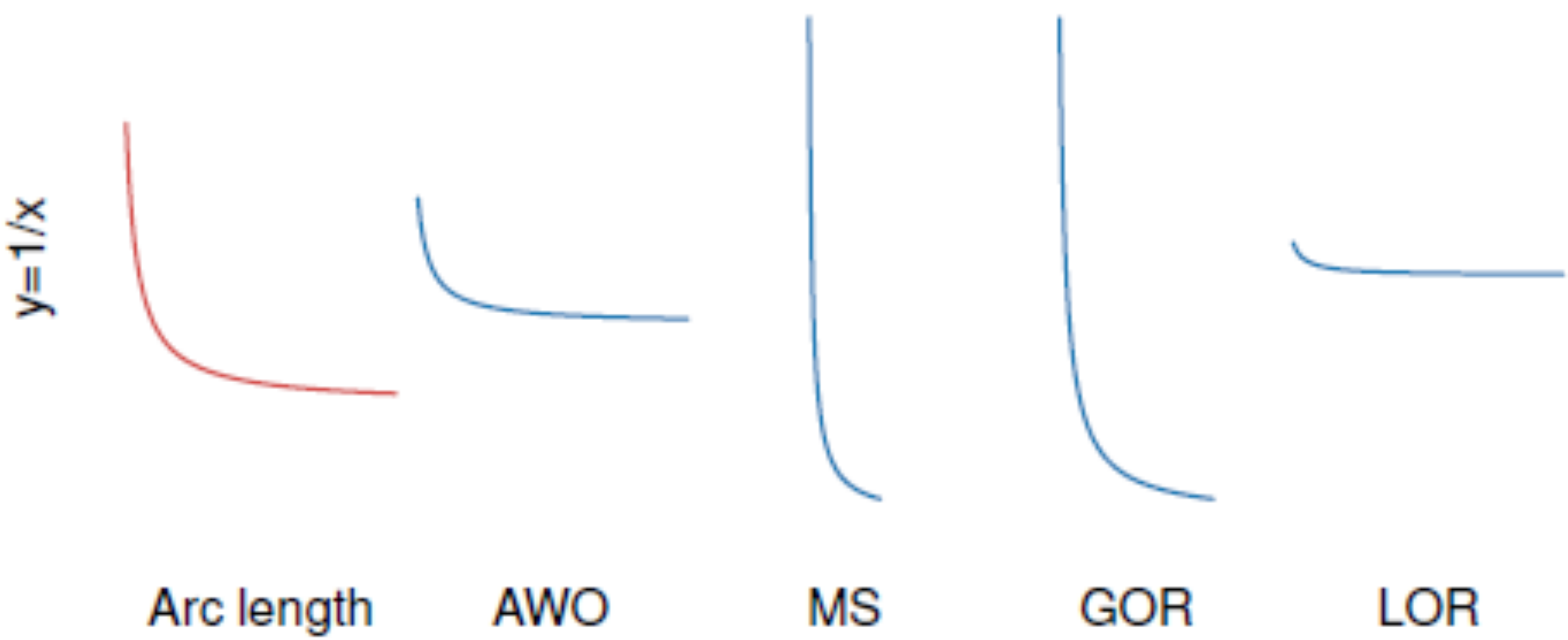


Ellipse -> Circle

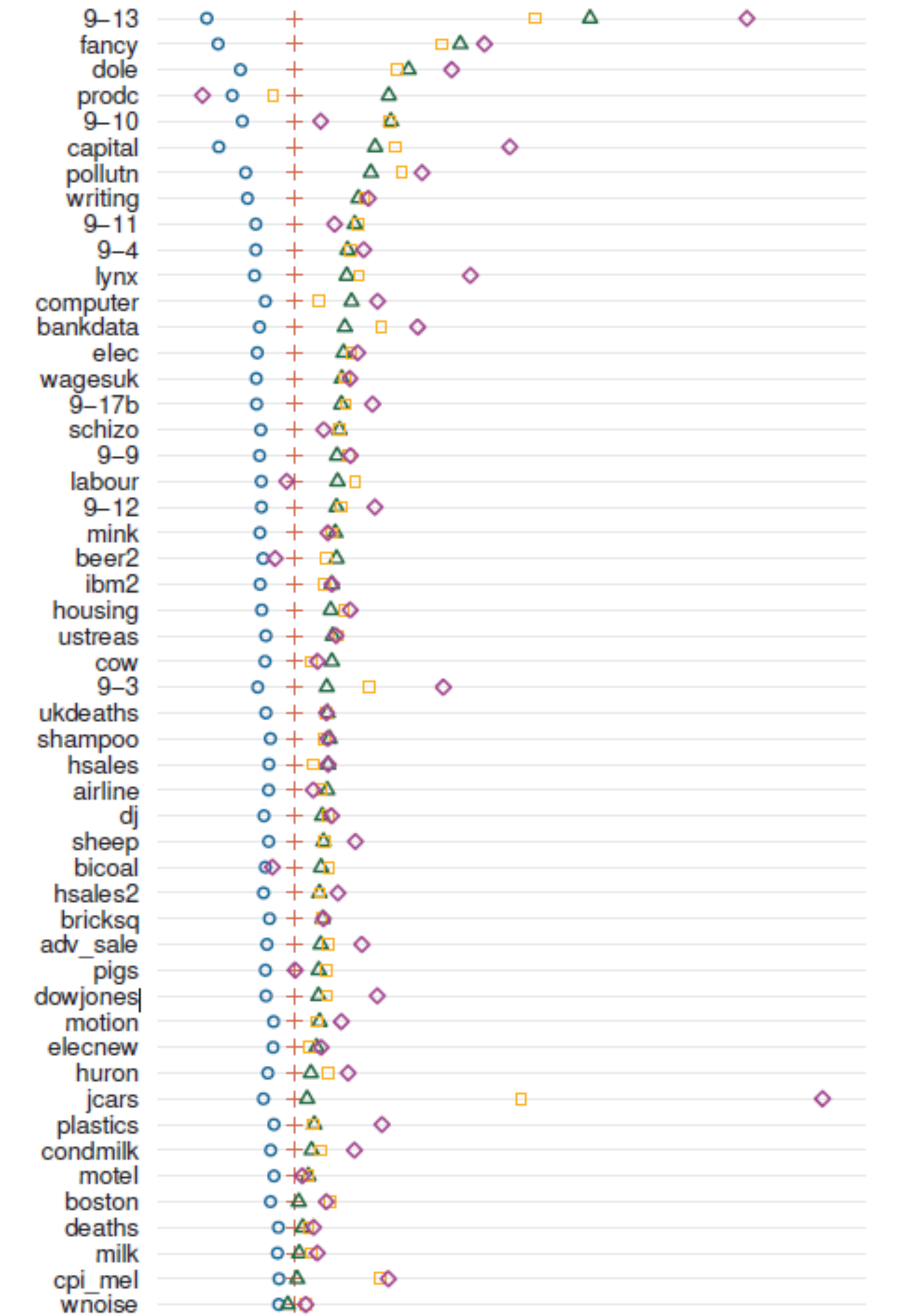


[Talbot et al. 2011]

Robustness: Banking y = 1/x



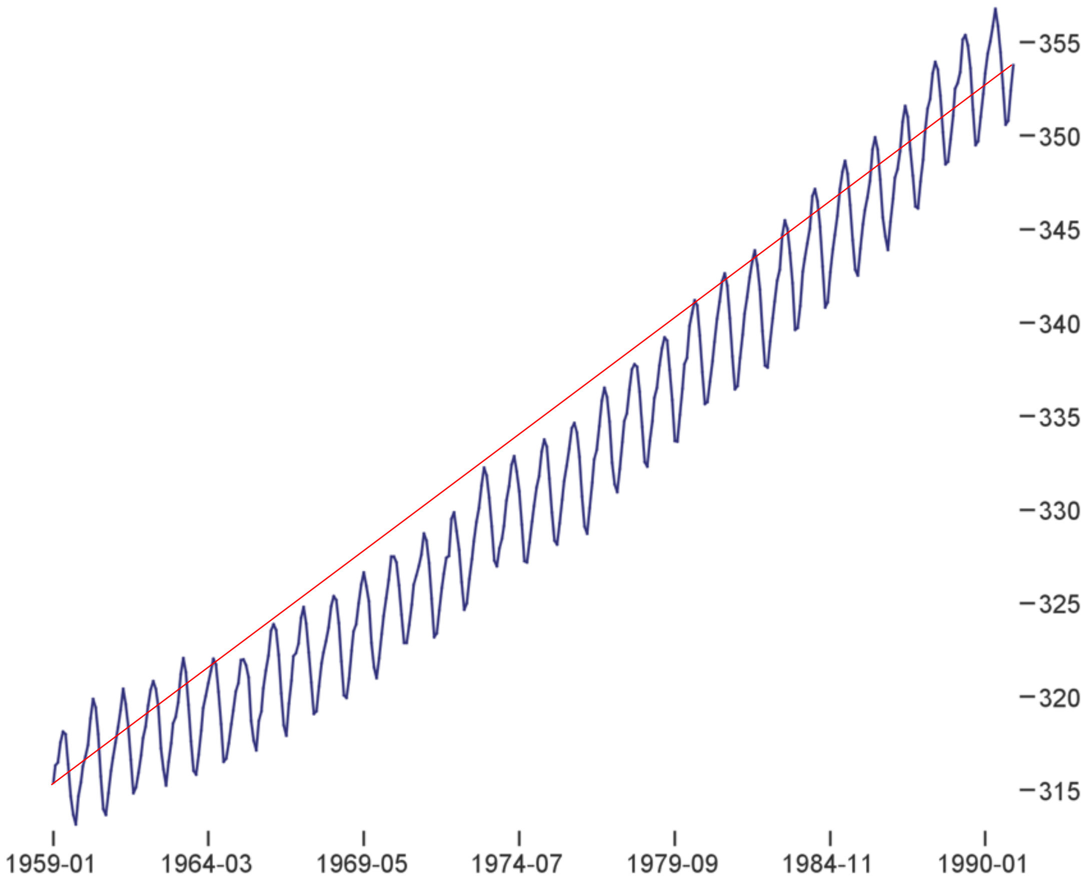
[Talbot et al. 2011]



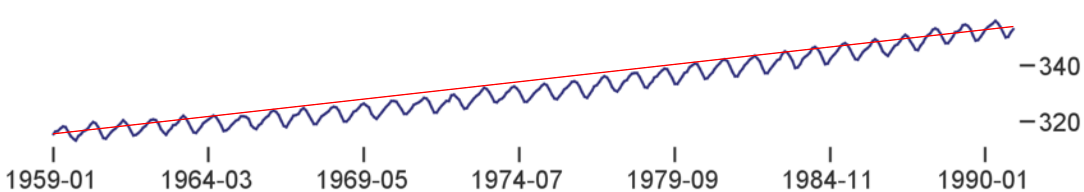
A Good Compromise

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

[Talbot et al. 2011]



Aspect Ratio = 1.17



Aspect Ratio = 7.87

Trends may occur at different scales!

Apply banking to the original data *or* to fitted trend lines.

[Heer & Agrawala '06]

CO_2 Measurements

William S. Cleveland
Visualizing Data

Discussion

Arc-length banking preferable to prior methods

Parameterization invariant

Robust (corner cases, compromise results)

Applicable to plotted curves and contour lines

Fast-converging iterative optimization

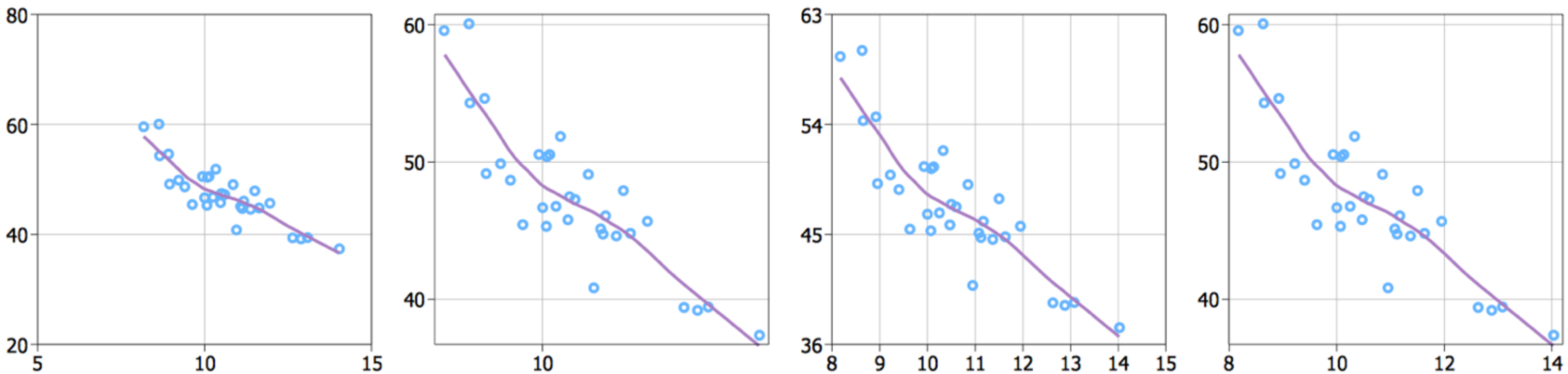
But what about perceptual effectiveness?

We lack theory to motivate aspect ratio selection.

Perceptual experiments needed to assess?

Axis Ticks

Tick Mark Selection



(a) Heckbert (b) R's pretty (c) Wilkinson (d) Extended

What are some properties of "good" tick marks?

Tick Mark Criteria

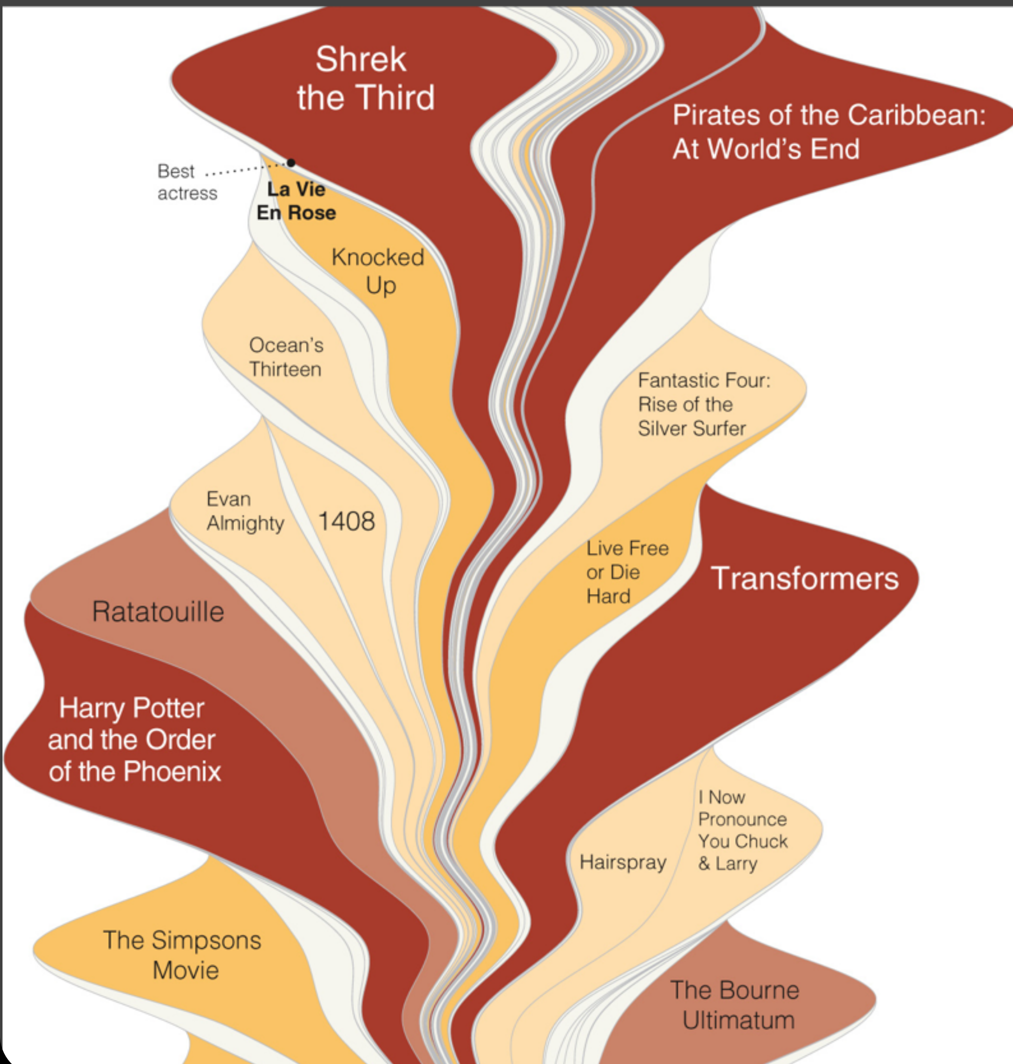
Simplicity - numbers are multiples of 10, 5, 2
Coverage - ticks near the ends of the data
Density - not too many, nor too few
Legibility - whitespace, horizontal text, size

Optimization
Talbot et al '10 use a search procedure that optimizes criteria in turn (e.g., find simple numbers first, then adjust coverage, *etc.*).

$S = 0.2\ simplicity + 0.25\ coverage + 0.5\ density + 0.05\ legibility$

Streamgraph Layout

Streamgraph Layout [Byron & Wattenberg '08]



Thickness encodes values, stack extent encodes sum of values.

Issues

Layers bias each other

Perceived thickness

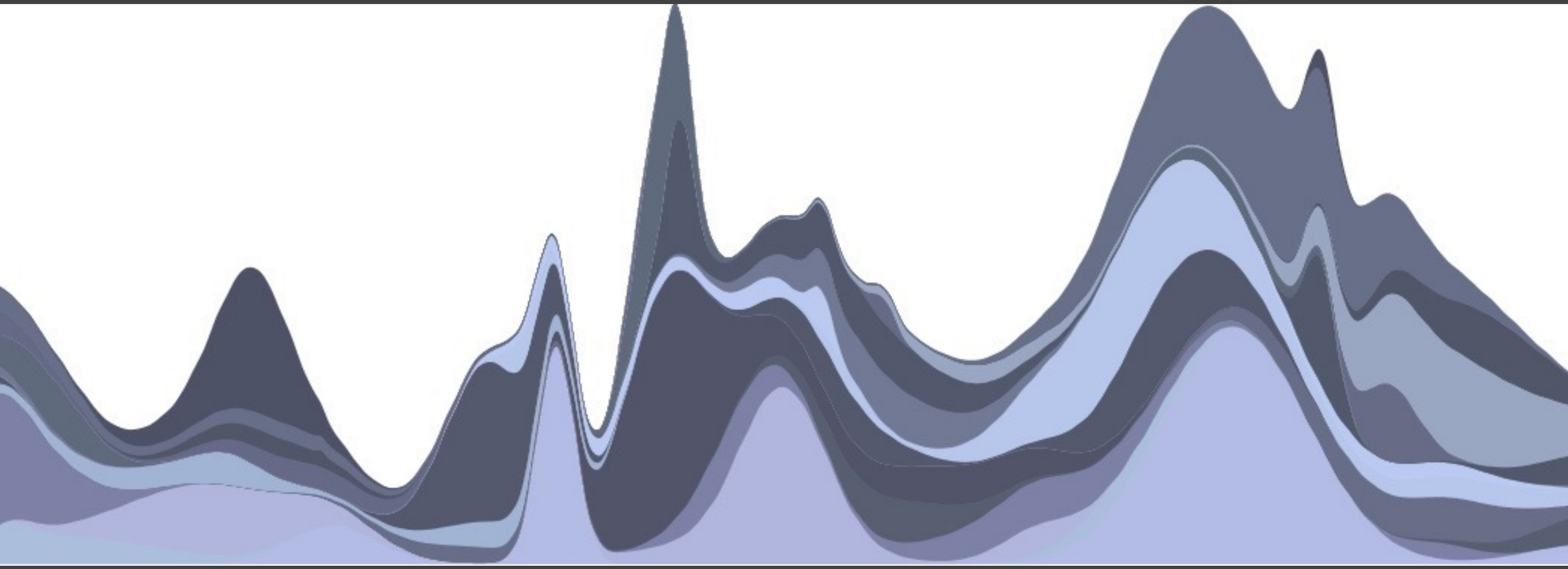
Slope (bank to 45?)

Design Parameters

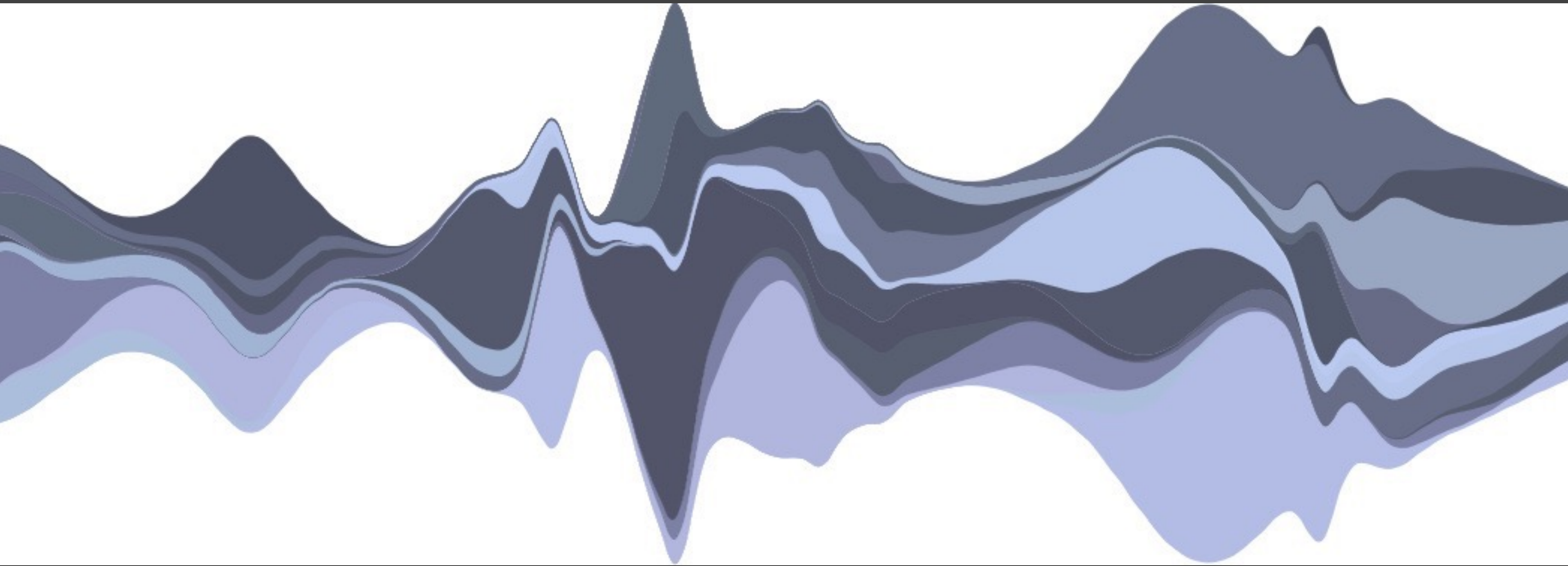
Layer baseline

Layer order

Zero Baseline

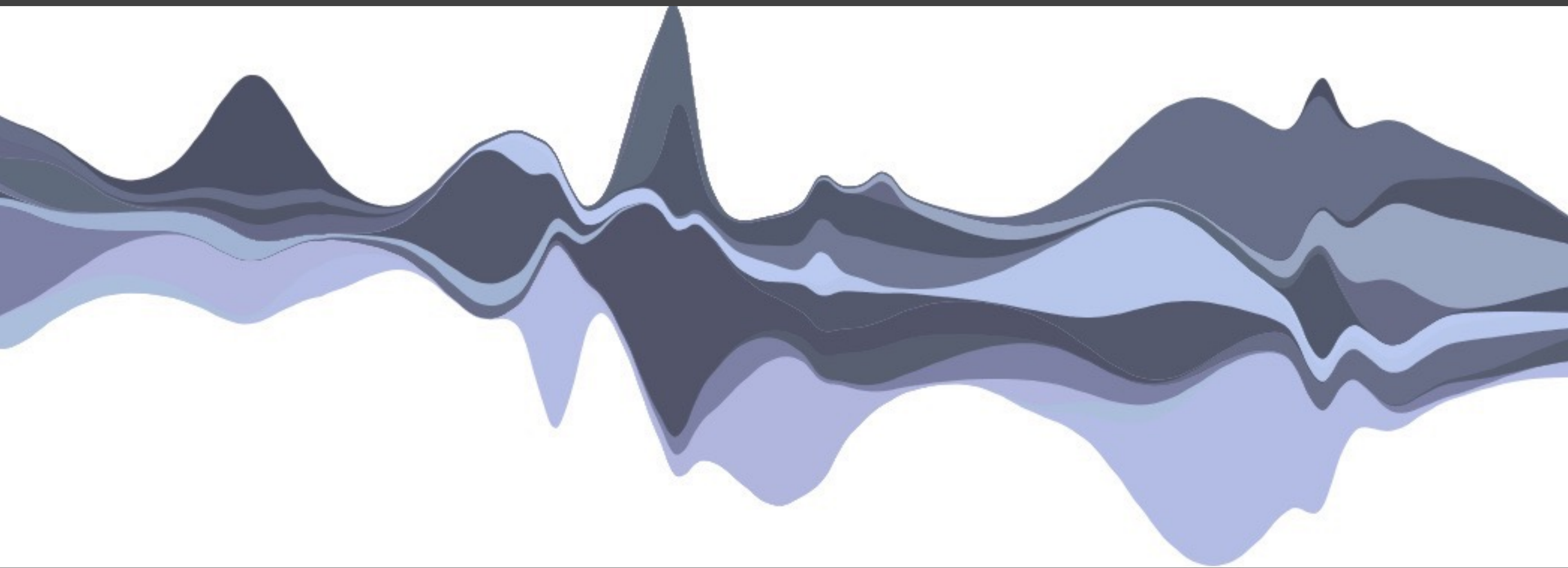


Centered Baseline



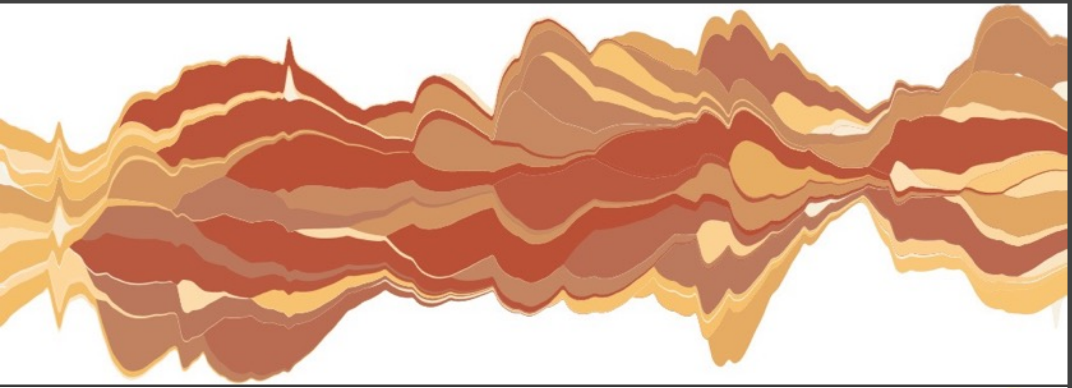
Pros: reduces "spikiness", minimizes slope (in the least squares sense) for bottom and top layers.
Cons: removes fixed baseline for sum comparison.

"Wiggle" Baseline [Byron & Wattenberg '08]

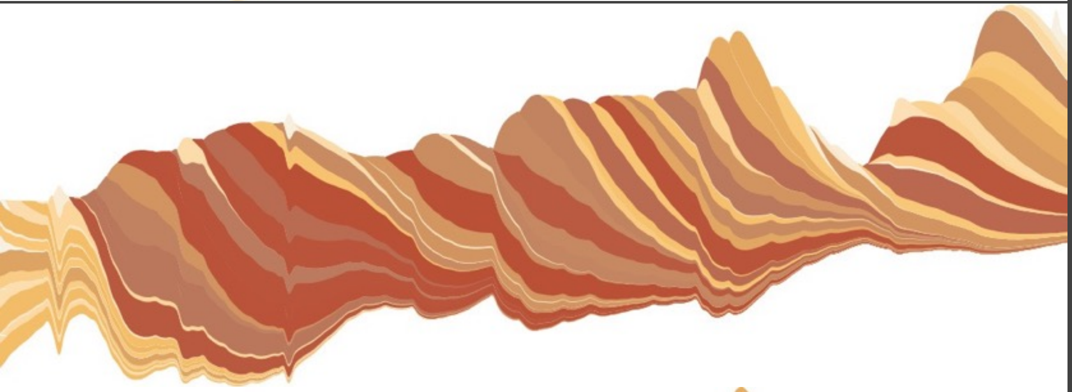


Pros: minimizes *overall* slope (in the least squares sense), weighted by layer thickness.
Cons: chart may require more total area.

Layer Sorting



Unsorted



Sort by First Non-Zero Value

Sort “Inside-Out”: Place layers alternately at top or bottom, in a thickness balancing manner.

Streamgraphs Redux [Di Bartolomeo & Hu '16]

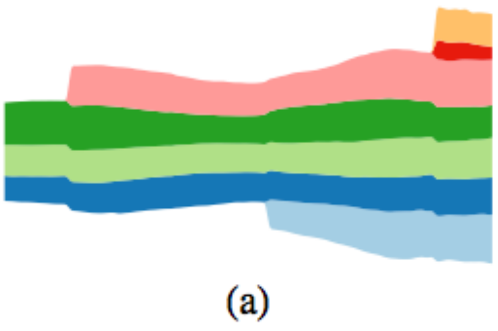


Figure 2: *Different baselines: (a) with 2-norm, distortions are present; (b) with 1-norm, all layers are smooth.*

Minimize slope using **L1-norm**, not (least-squares) **L2-norm**.

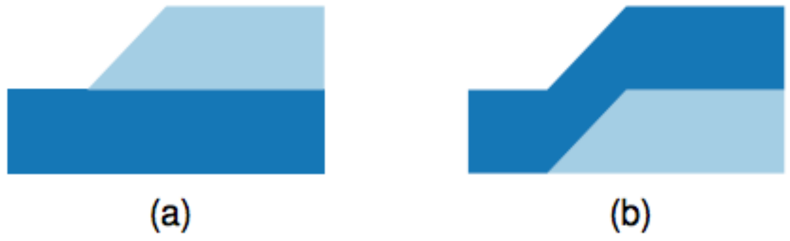
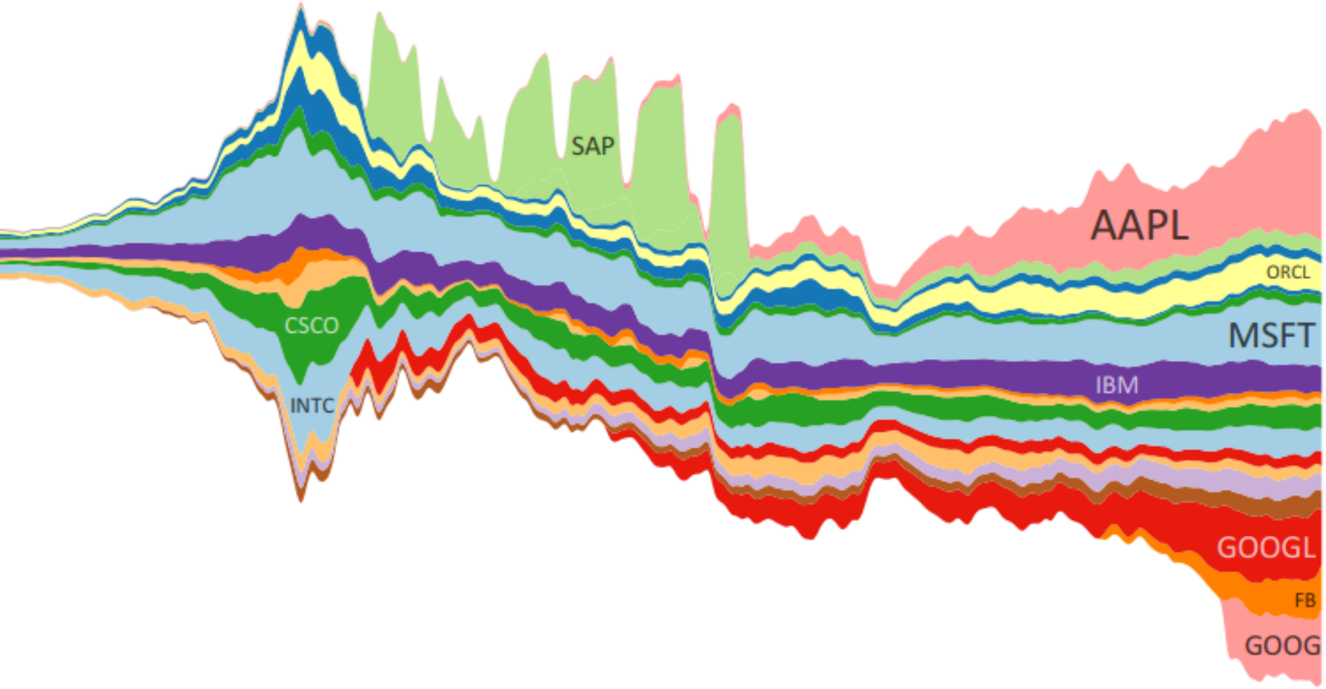


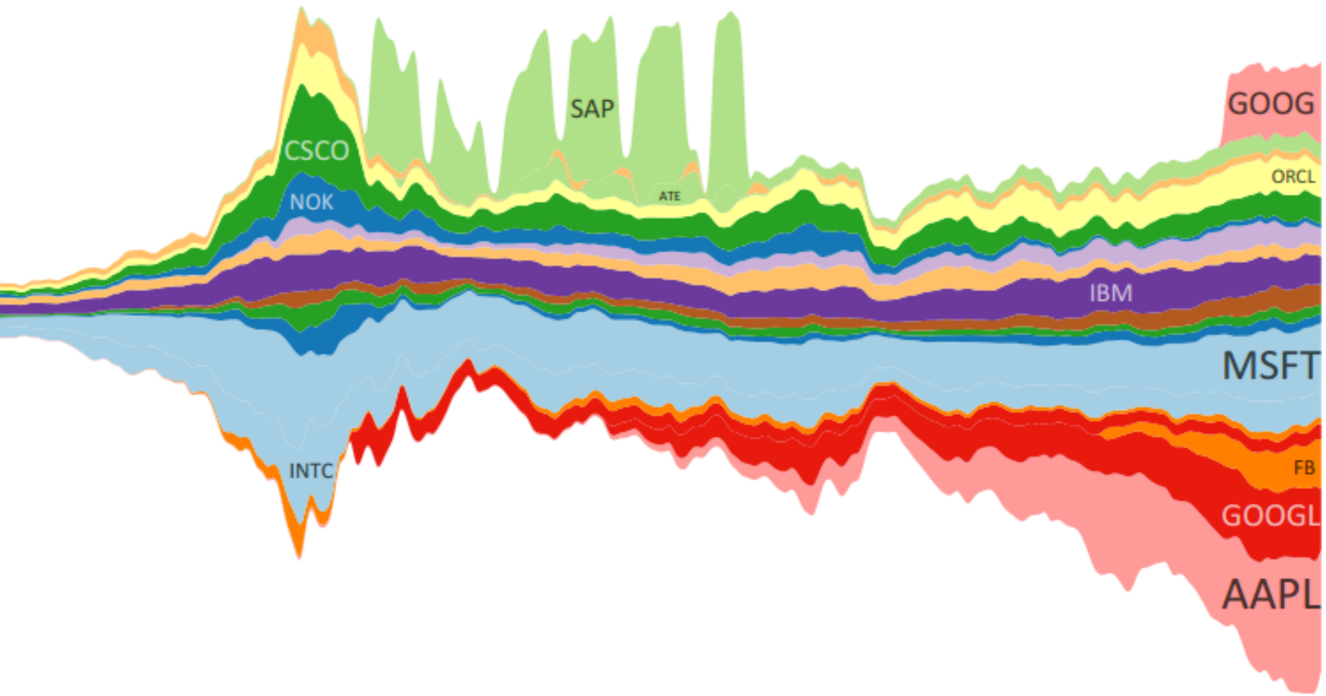
Figure 4: *Two orderings of two layers on a flat baseline. Case (b) produces a distortion, case (a) is preferred.*

Perform **iterative optimization**, swapping adjacent layers if it improves the overall layout.

Results



Byron & Wattenberg ’08
(via D3’s stacked layout)



Di Bartolomeo & Hu ’16

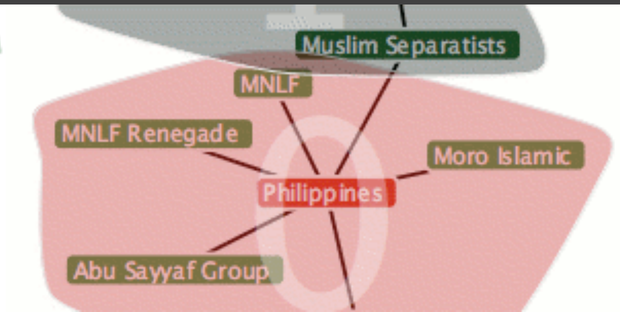
Administrivia

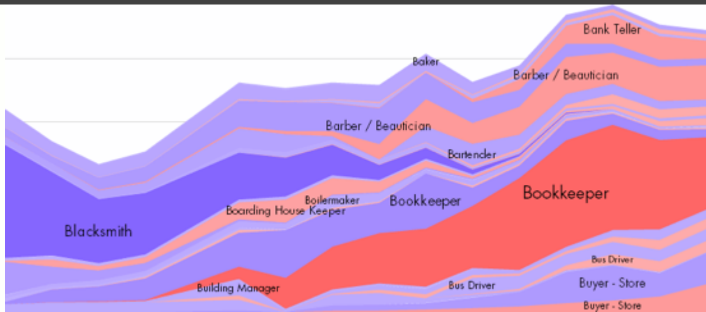
FP: Interactive Prototype

Create an interactive visualization. Choose a driving question in your topic area and develop an appropriate visualization + interaction techniques, then deploy your visualization on the web.

Due by *5pm* on **Wednesday, May 3**.

We will assign your GitHub repos shortly!



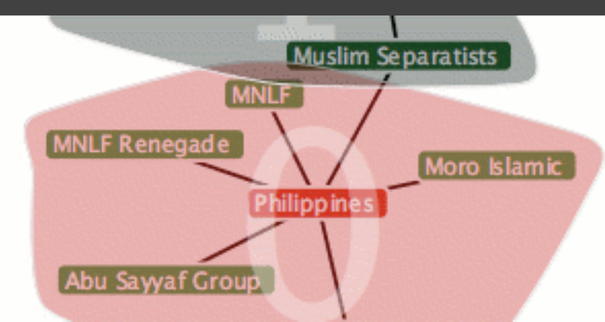


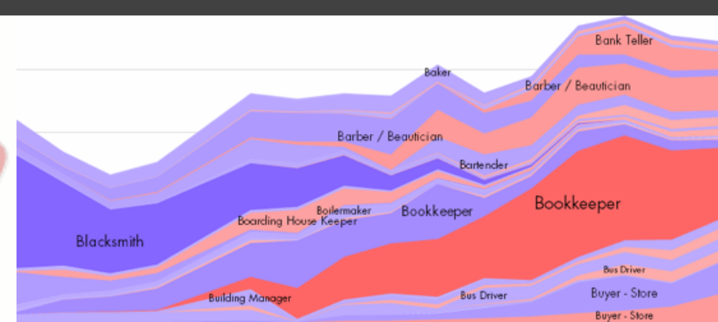
Requirements

Interactive. You must implement interaction methods! However, this is not only selection / filtering / tooltips. Also consider annotations or other narrative features to draw attention and provide additional context

Web-based. D3 is encouraged, but not required. Deploy your visualization using GitHub pages.

Write-up. Provide design rationale on your web page.



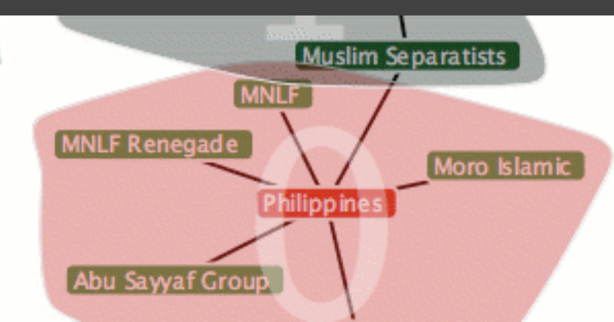


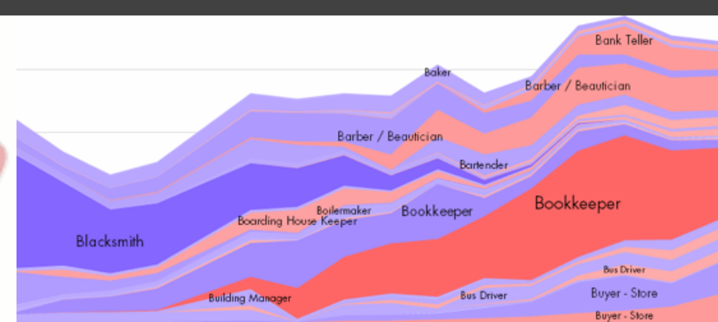
Interactive Prototype Tips

Start now. It will take longer than you think.

Keep it simple. Choose a minimal set of interactions that enables users to explore and generate interesting insights. Do not feel obligated to convey *everything* about the data: focus on a compelling subset.

Promote engagement. How do your chosen interactions reveal interesting observations?





Team Member Roles

We encourage you to structure team responsibilities!

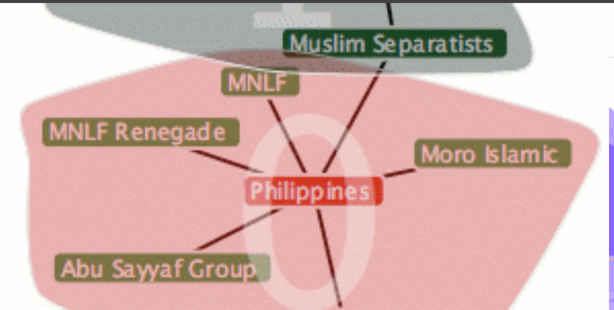
Coordinator: Organize meetings, track deadlines, *etc.*

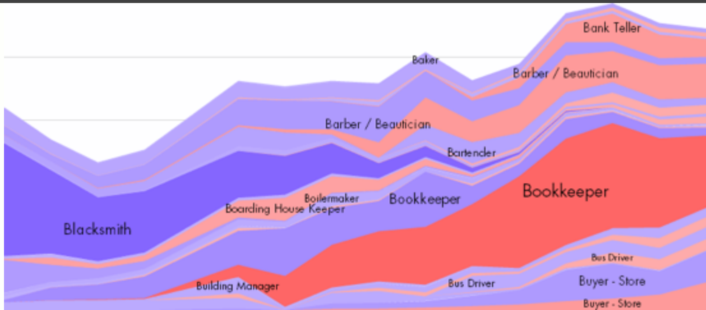
Data Lead: Data wrangling, management, distillation

Tech Lead: Manage code integration, GitHub repo

UX Lead: Visualization/interaction design & evaluation

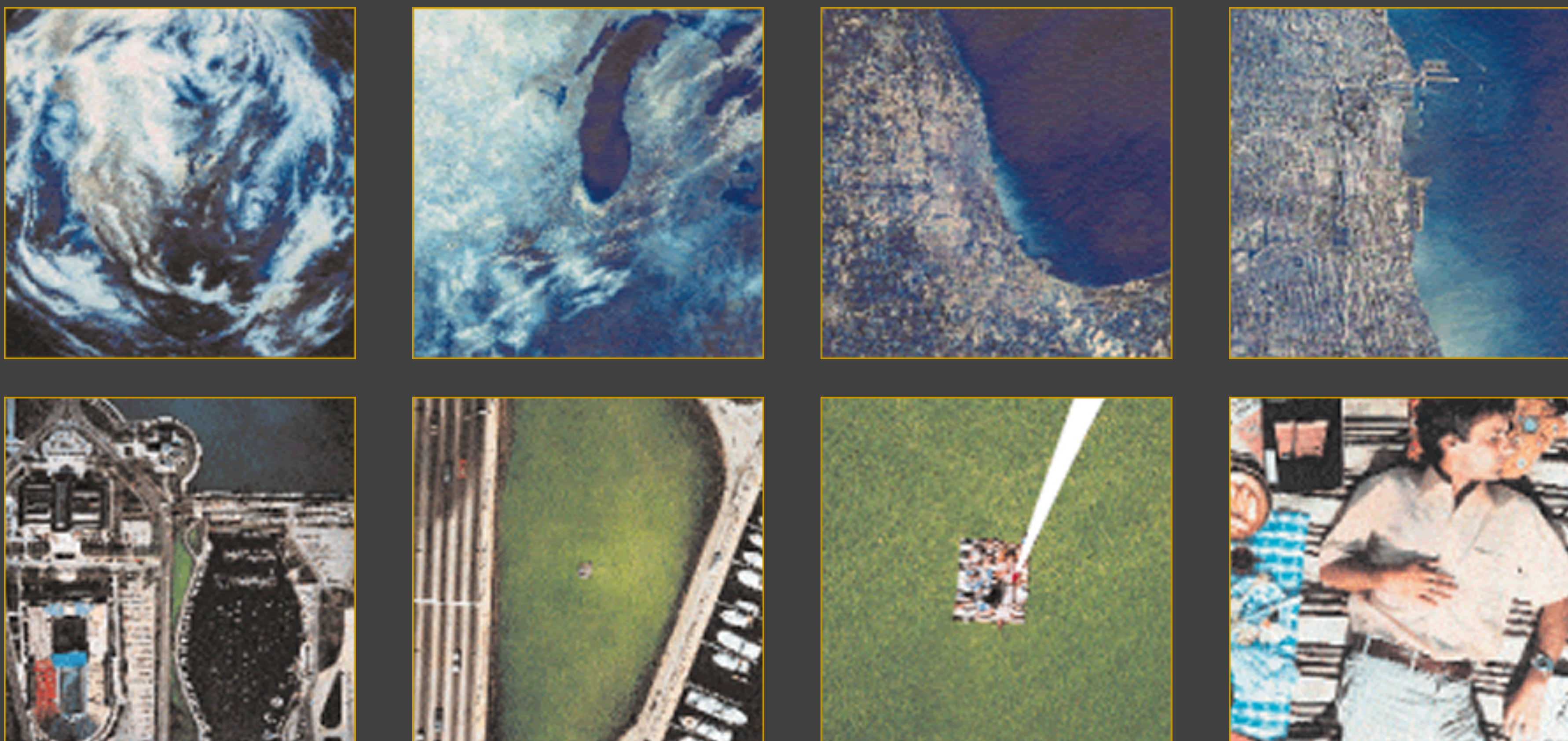
One may have multiple roles, share work across roles…





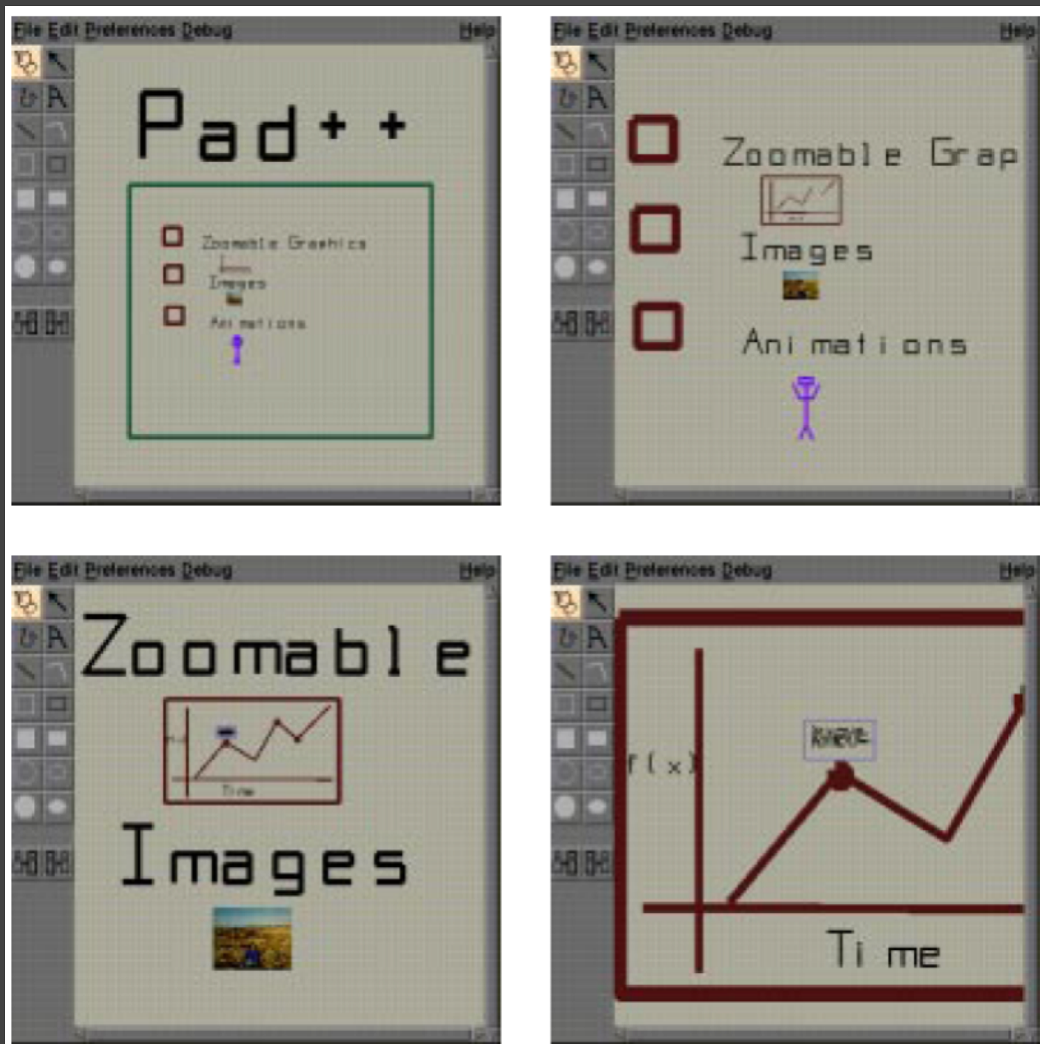
Zooming

Zooming

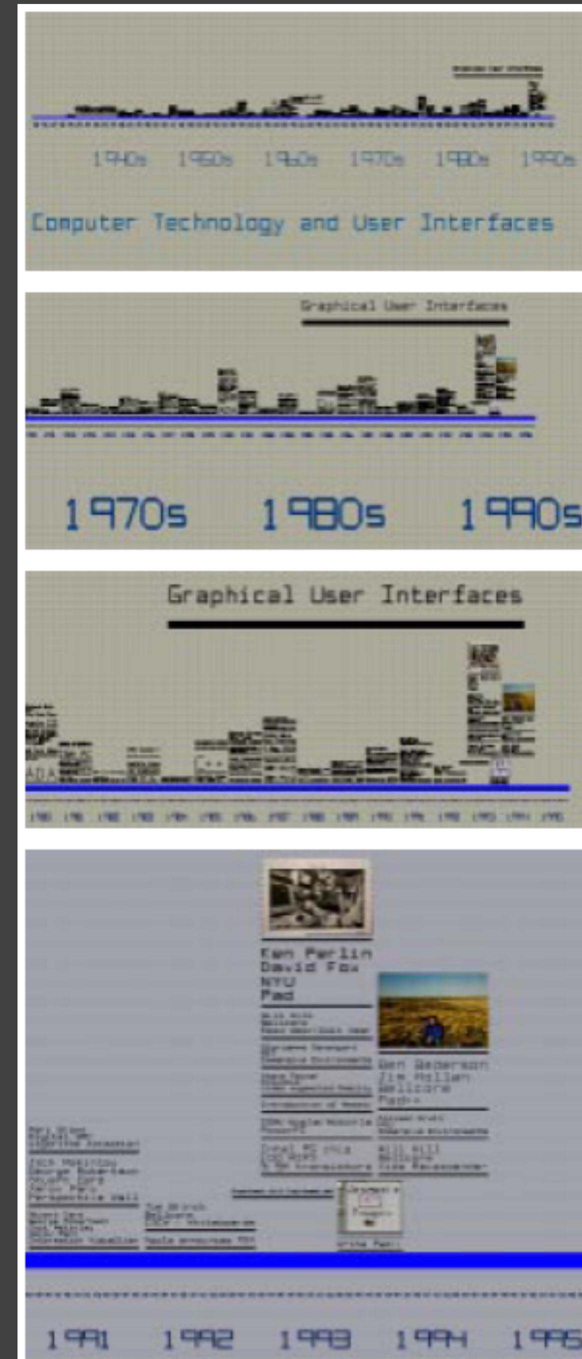


Eames’ Powers of Ten

Interactive Zooming

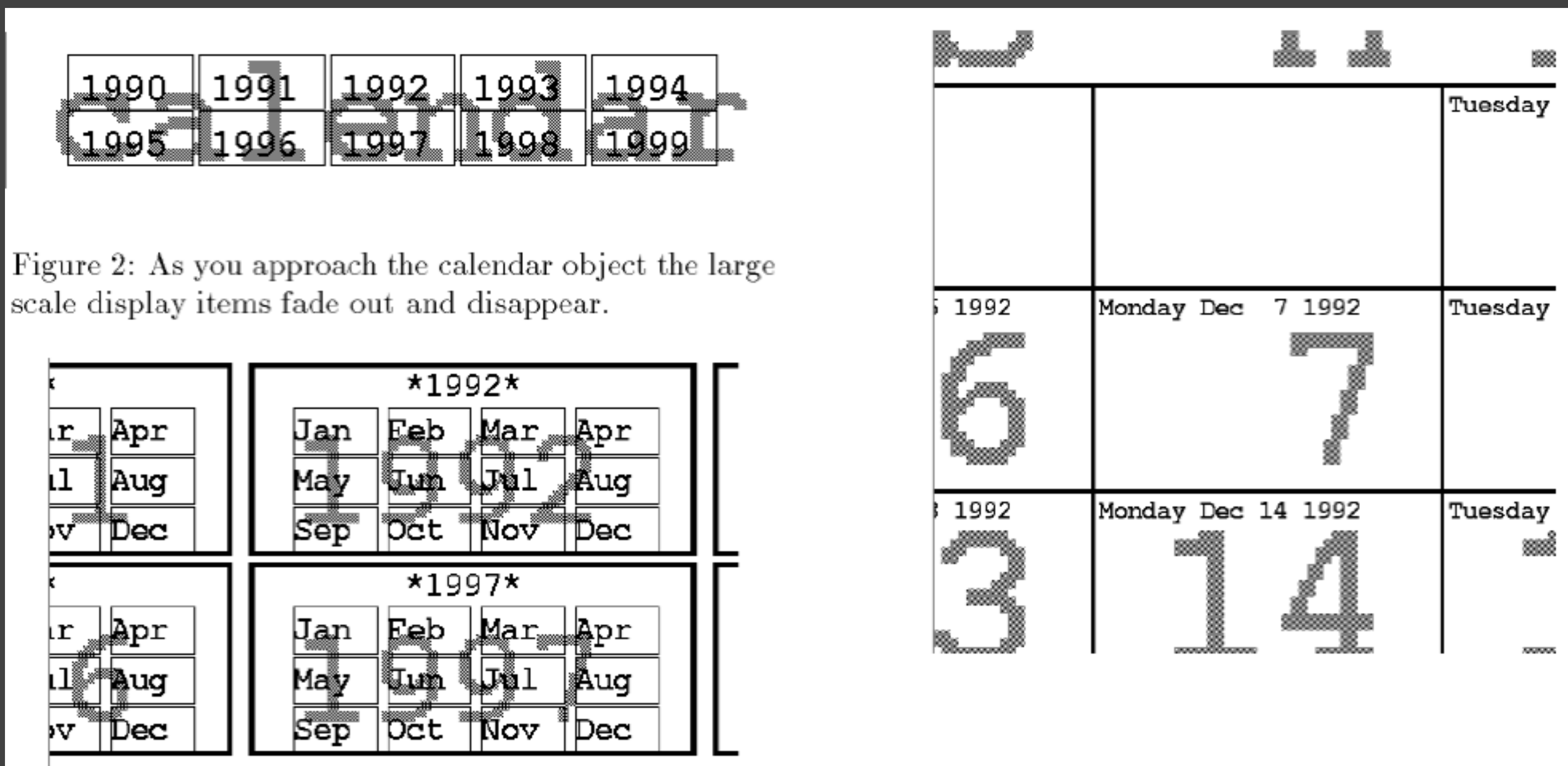


Pad++ [Bederson and Hollan 94]

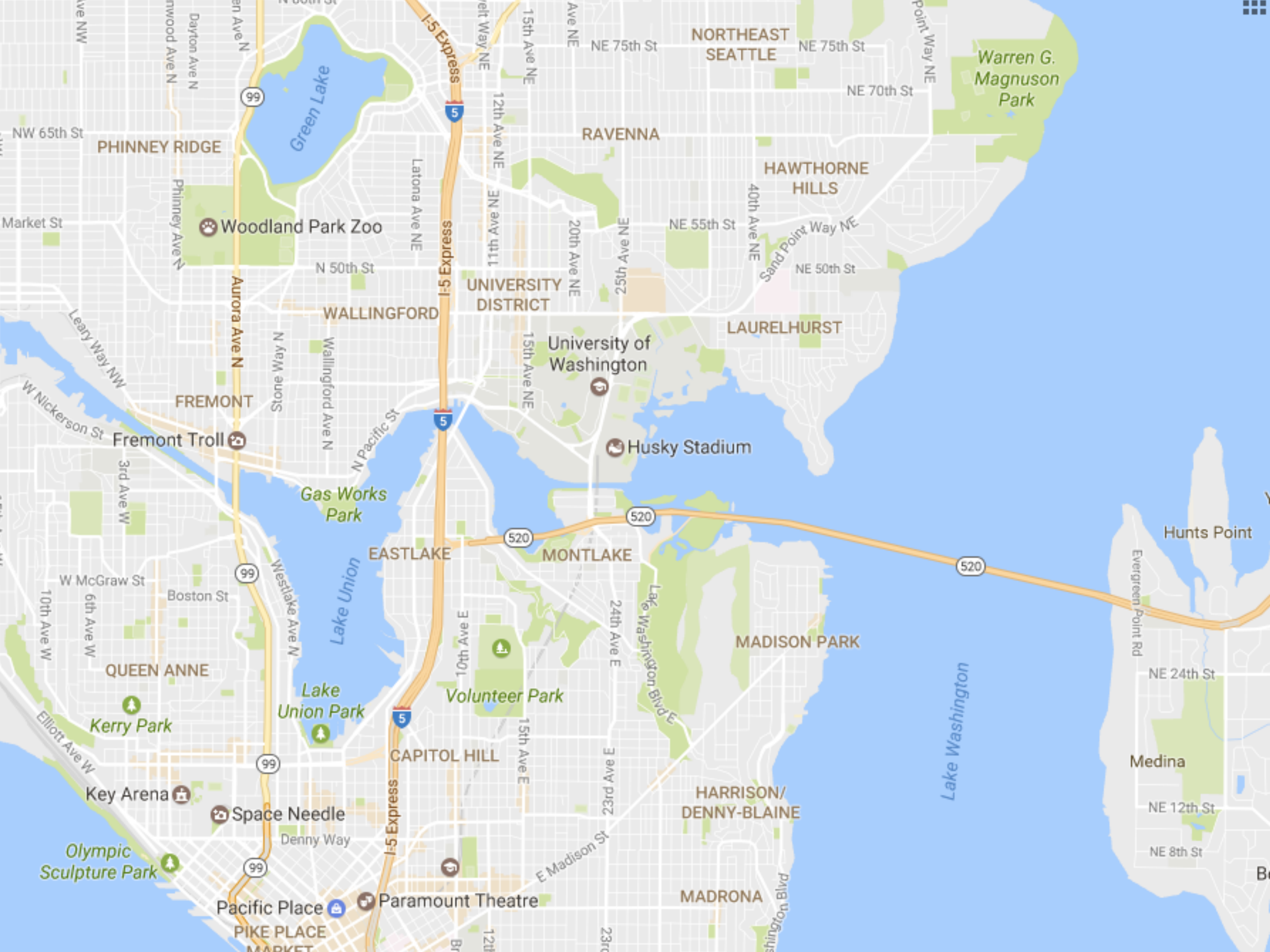


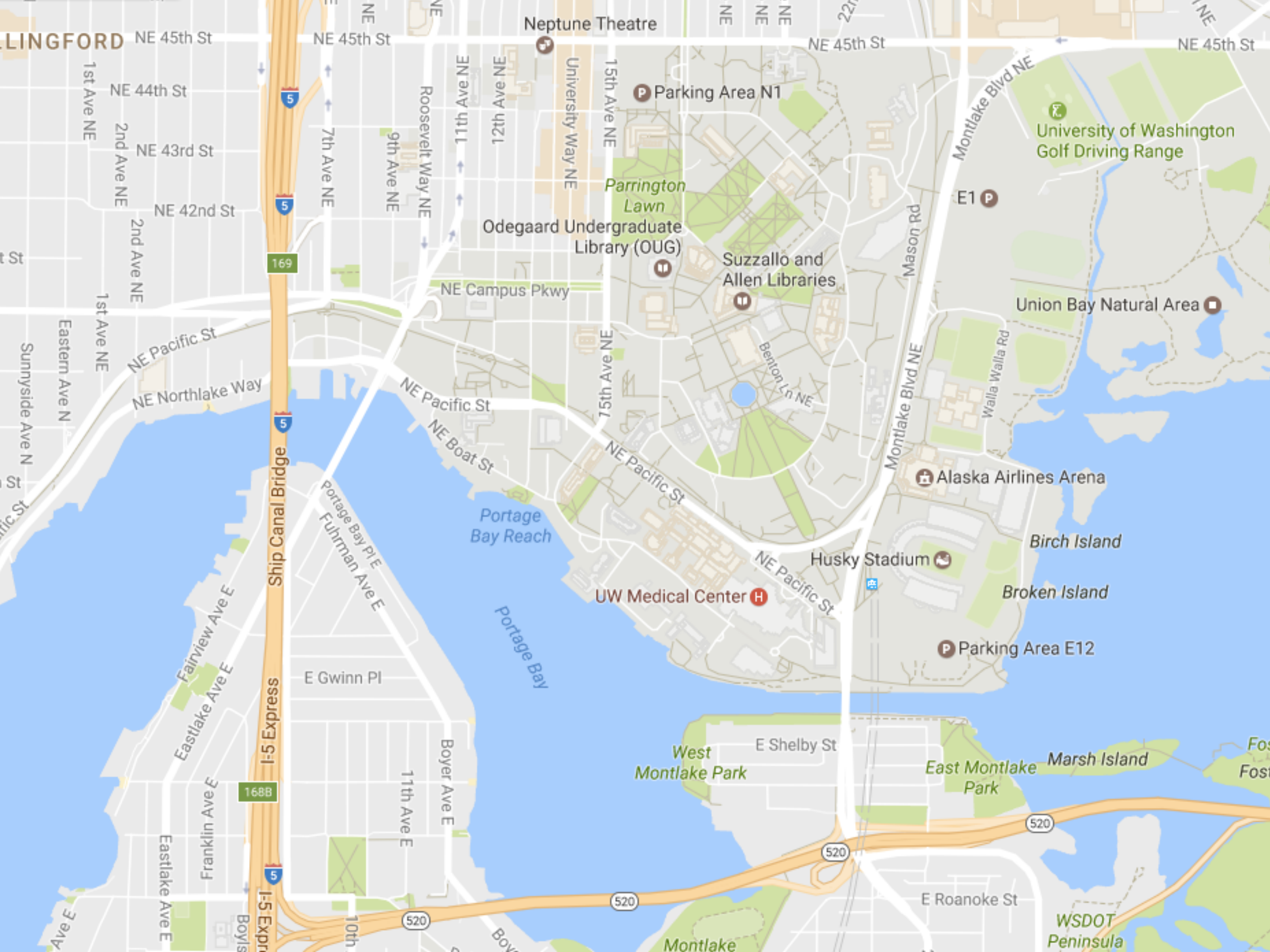
Semantic Zooming

Change representations as zoom level changes

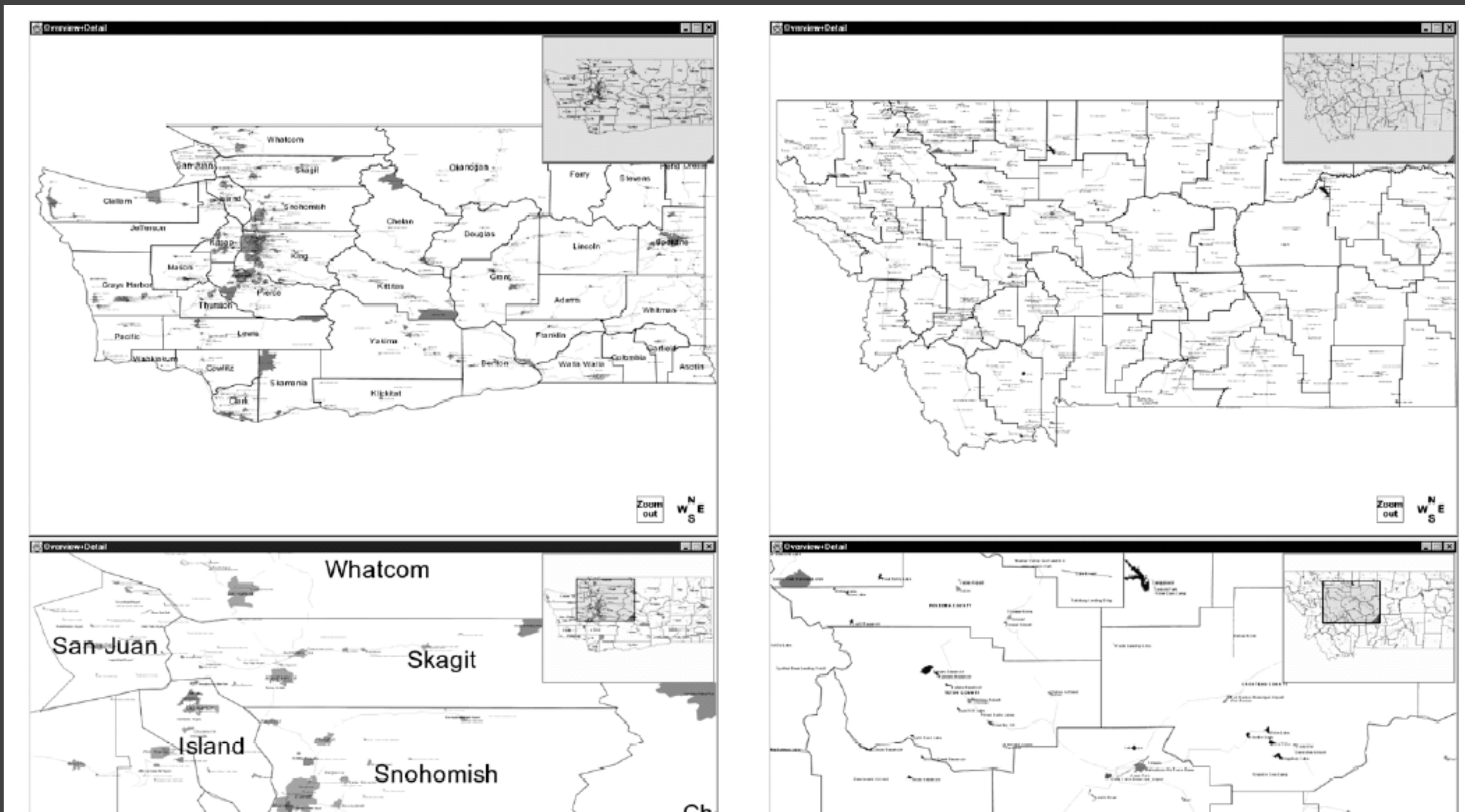


PAD [Perlin and Fox 93]


NORTHEAST SEATTLE
NE 75th St
NE 70th St
Warren G. Magnuson Park
Green Lake
PHINNEY RIDGE
NW 65th St
RAVENNA
HAWTHORNE HILLS
Woodland Park Zoo
Market St
NE 55th St
Sand Point Way NE
NE 50th St
N 50th St
UNIVERSITY DISTRICT
WALLINGFORD
LAURELHURST
University of Washington
Aurora Ave N
Leary Way NW
FREMONT
W Nickerson St
Fremont Troll
Husky Stadium
Gas Works Park
N Pacific St
EASTLAKE
MONTLAKE
Hunts Point
W McGraw St
Boston St
Lake Union
Westlake Ave N
Lake Washington Blvd E
MADISON PARK
QUEEN ANNE
Kerry Park
Lake Union Park
Volunteer Park
Lake Washington
Elliott Ave W
CAPITOL HILL
Medina
Key Arena
Space Needle
HARRISON/ DENNY-BLAINE
Denny Way
NE 24th St
NE 12th St
NE 8th St
Olympic Sculpture Park
E Madison St
Pacific Place
Paramount Theatre
MADRONA
PIKE PLACE MARKET
I-5 Express
Evergreen Point Rd


Neptune Theatre
LINGFORD
NE 45th St
NE 45th St
NE 45th St
NE 45th St
NE 44th St
NE 43rd St
NE 42nd St
1st Ave NE
2nd Ave NE
7th Ave NE
9th Ave NE
Roosevelt Way NE
11th Ave NE
12th Ave NE
University Way NE
15th Ave NE
Parking Area N1
Montlake Blvd NE
University of Washington Golf Driving Range
Parrington Lawn
Odegaard Undergraduate Library (OUG)
Suzzallo and Allen Libraries
E1
Mason Rd
NE Campus Pkwy
Union Bay Natural Area
NE Pacific St
NE Northlake Way
Eastern Ave N
Sunnyside Ave N
Benton Ln NE
Walla Walla Rd
Montlake Blvd NE
NE Pacific St
NE Boat St
NE Pacific St
Alaska Airlines Arena
Ship Canal Bridge
Portage Bay Pl E
Fuhrman Ave E
Portage Bay Reach
Birch Island
Husky Stadium
Broken Island
UW Medical Center
NE Pacific St
Fairview Ave E
Eastlake Ave E
Portage Bay
Parking Area E12
E Gwinn Pl
I-5 Express
E Shelby St
West Montlake Park
East Montlake Park
Marsh Island
Boyer Ave E
11th Ave E
Franklin Ave E
Eastlake Ave E
E Roanoke St
WSDOT Peninsula
Montlake
169
168B
5
520

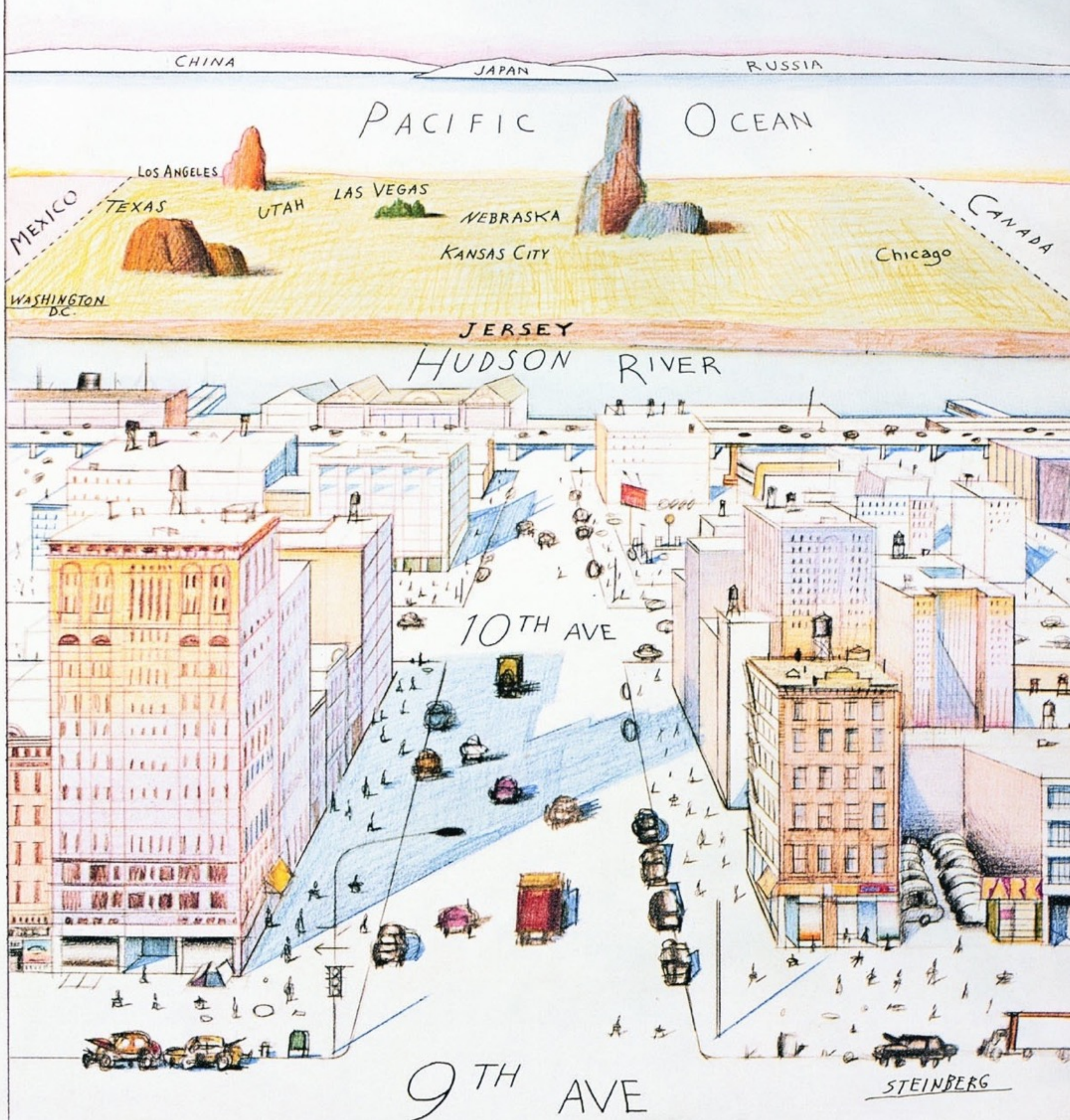
Overview + Detail



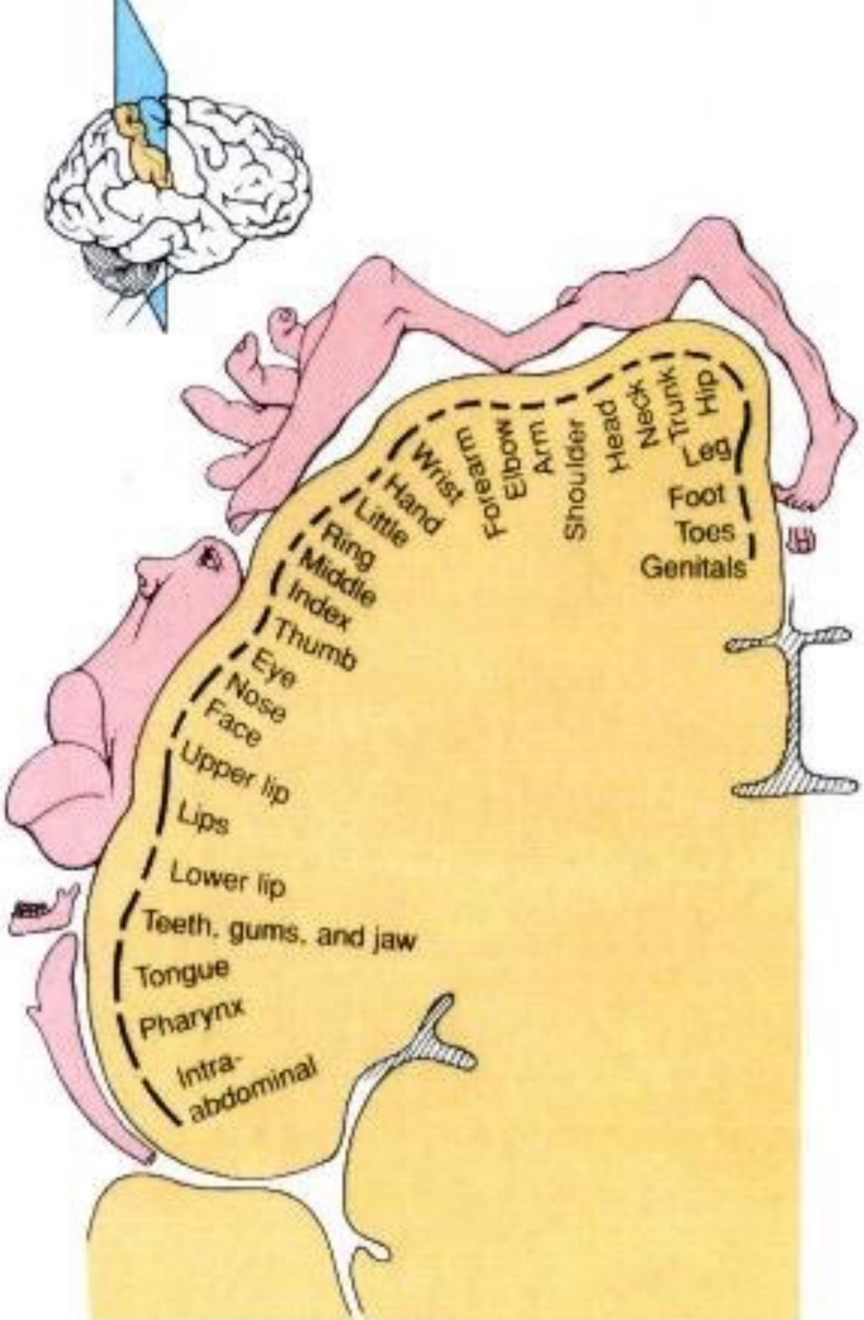
[Hornbaek et al. 2002]

Halo [Baudisch et al '03]

Focus + Context


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

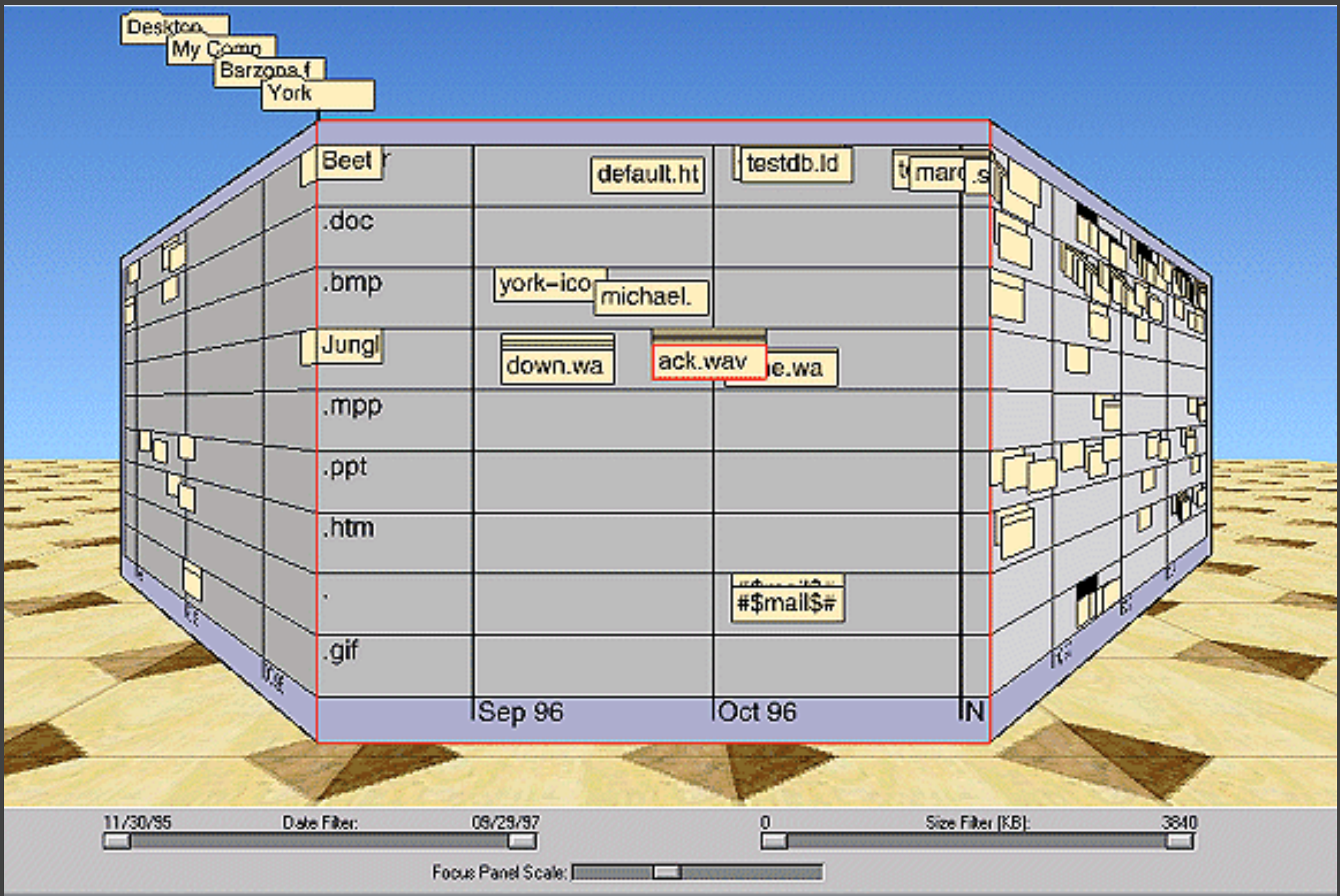
Degree-of-Interest [Furnas ’81, ’06]

Estimate the saliency of information to display.
Can affect **what** to show and/or **how** to show it!

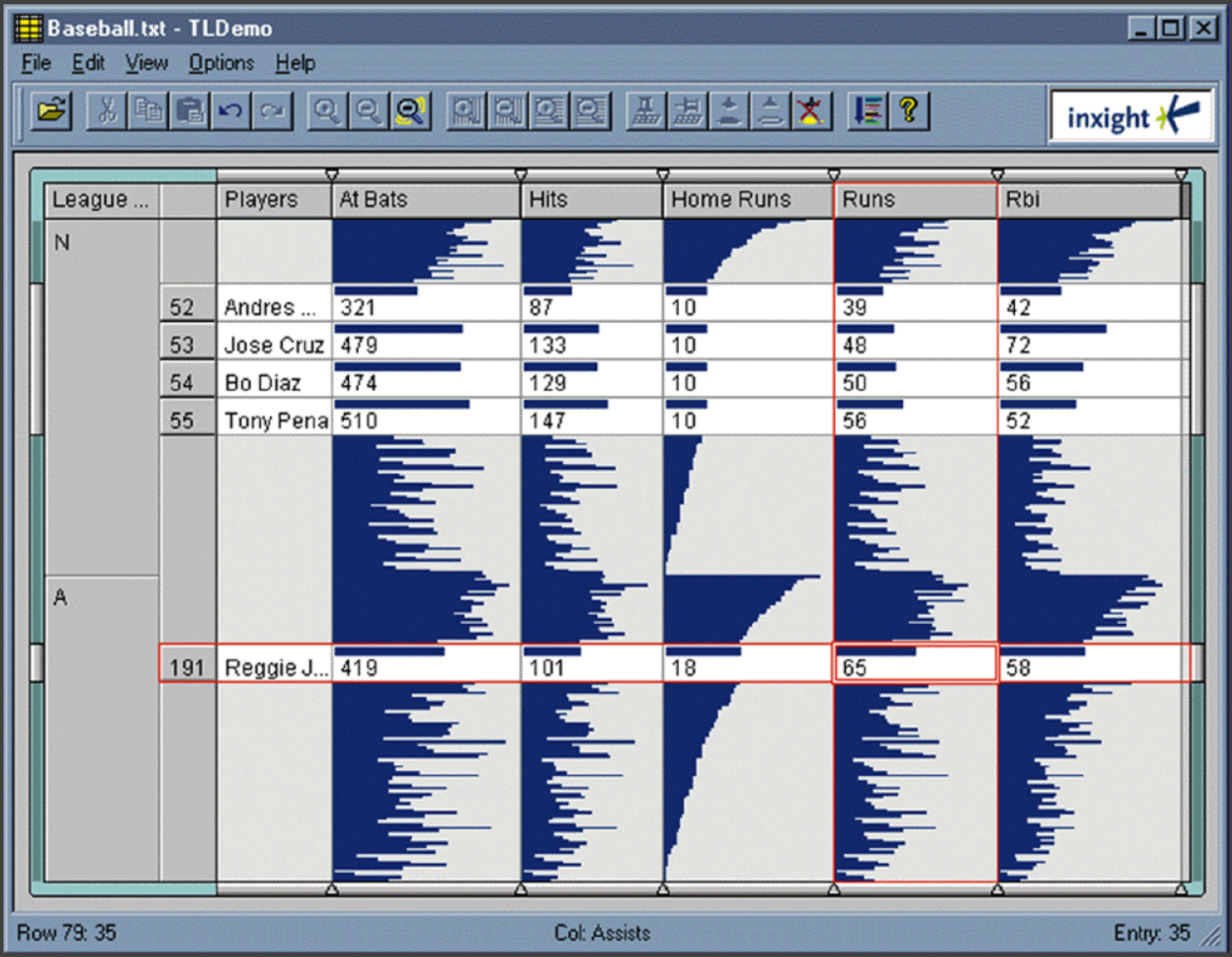
DOI ~ f (*Current Focus*, *A Priori Importance*)

Example: Google Search
Current Focus = Query Hits (e.g., text matching)
A Priori Importance = PageRank
What to show: Top N results
How to show it: List

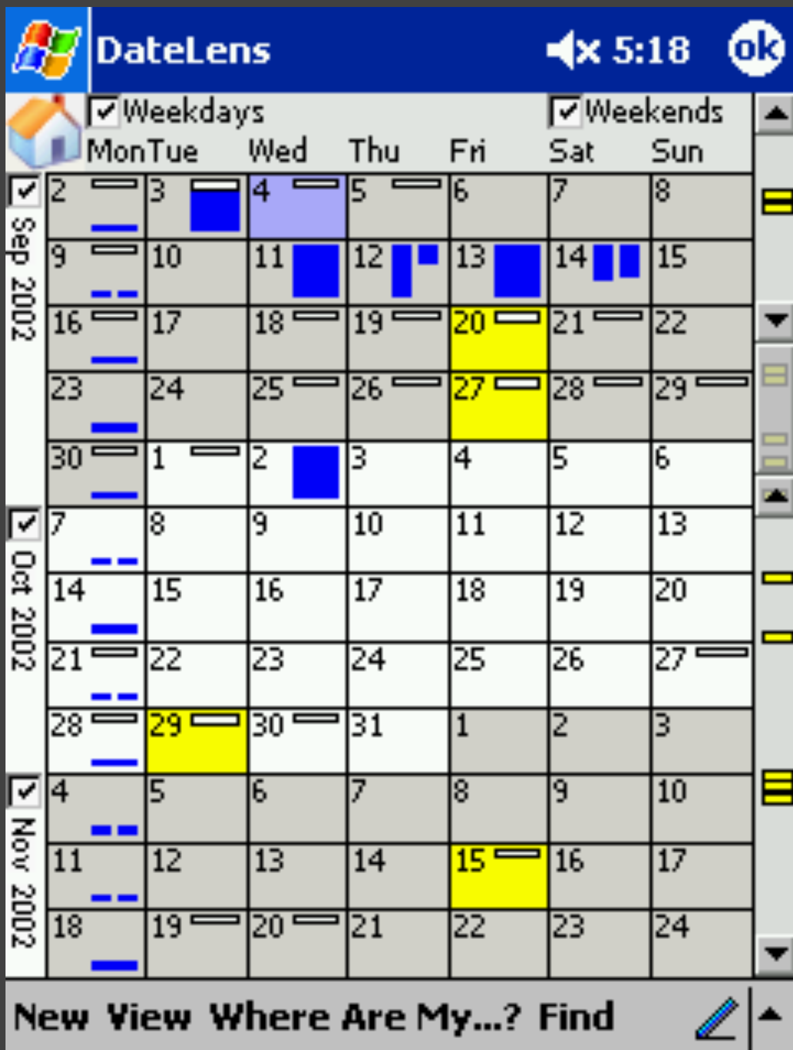
Perspective Wall [Mackinlay et al. '91]

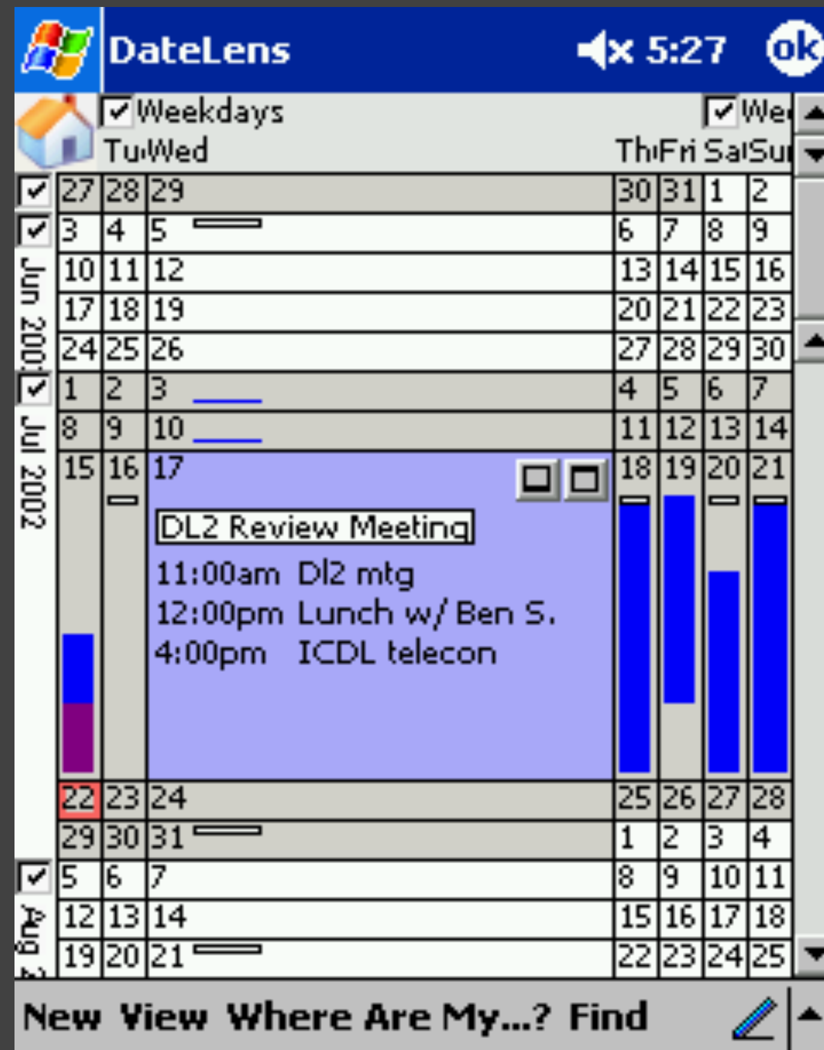


TableLens [Rao & Card '94]



DateLens [Bederson et al. '04]



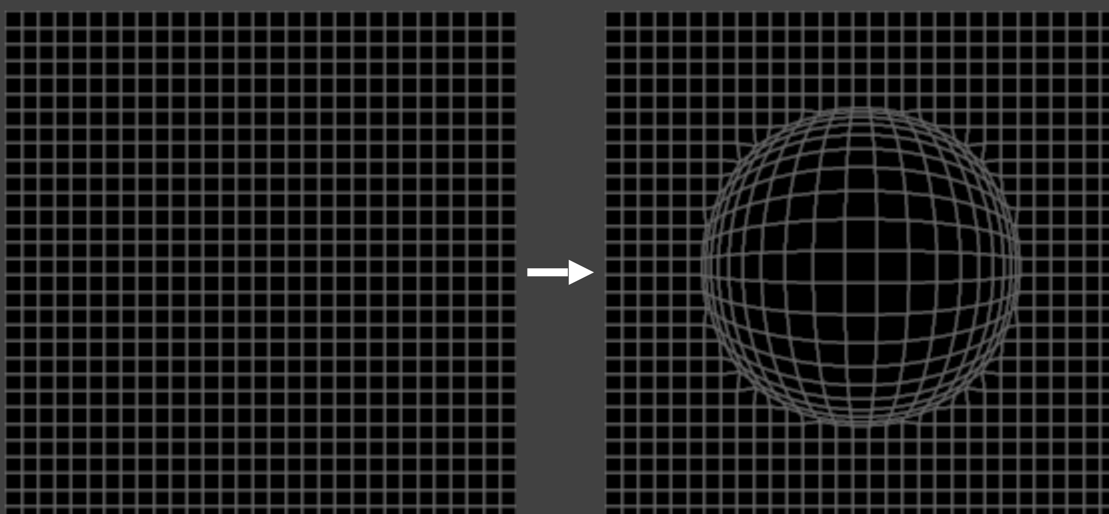


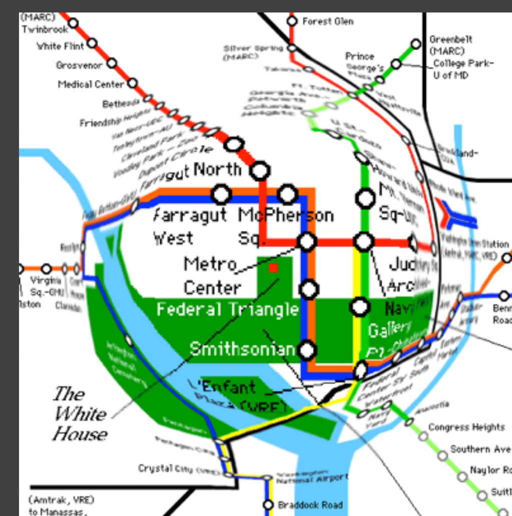
Focus + Context Distortion

Focus area – local details

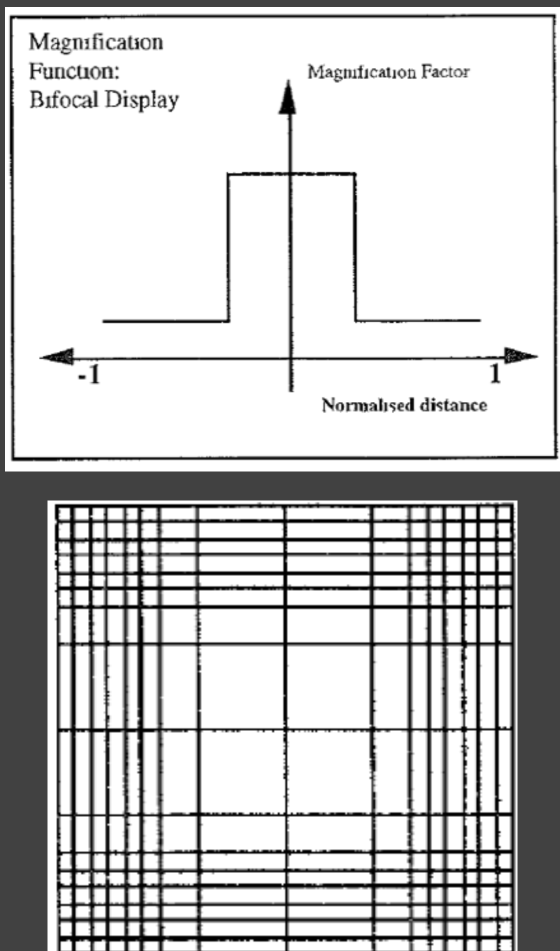
De-magnified area – surrounding context

Like a rubber sheet with borders tacked down

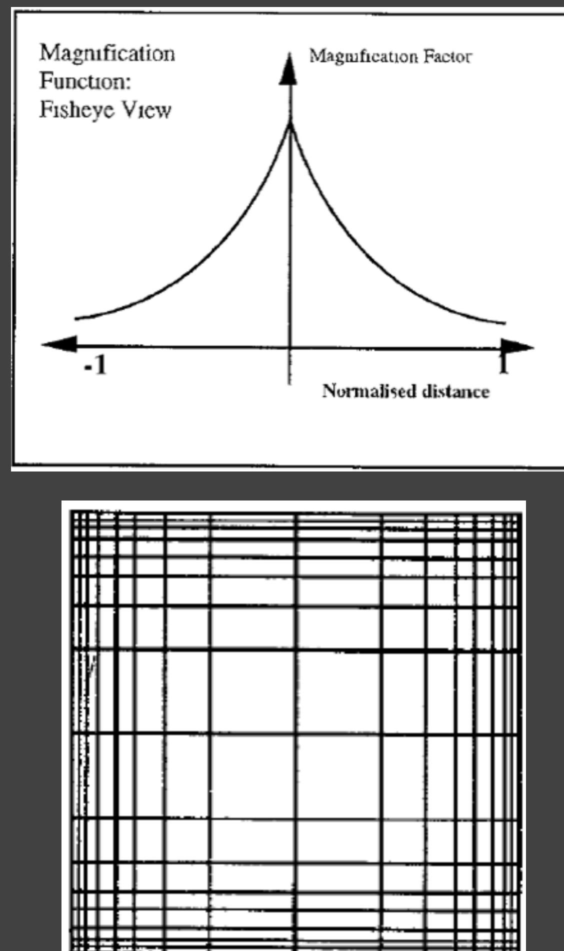




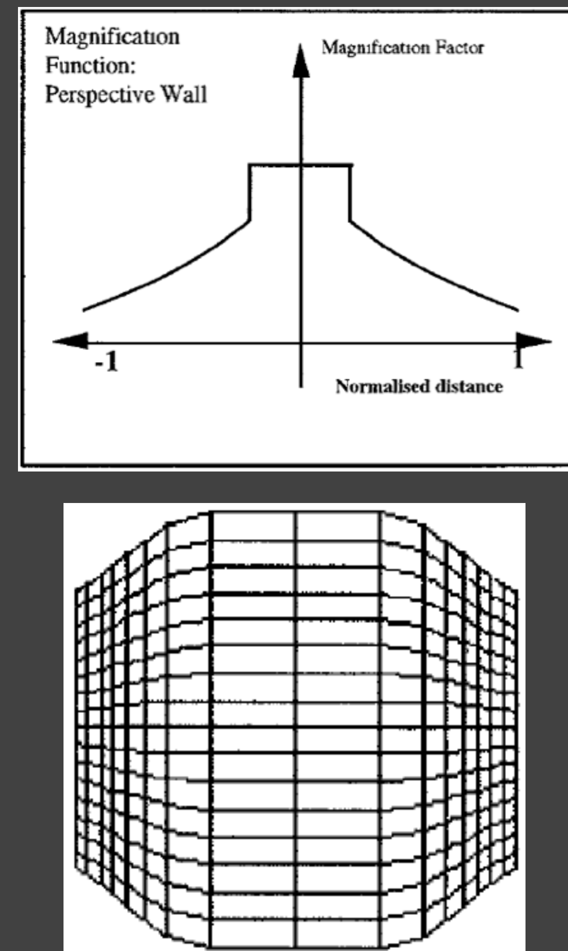
Distortion Functions



Bifocal



Fisheye



Perspective Wall

Uses (and Abuses) of Distortion

Often more harm than help…

Pan & zoom more familiar – and visually stable – than "rubber sheet" navigation.

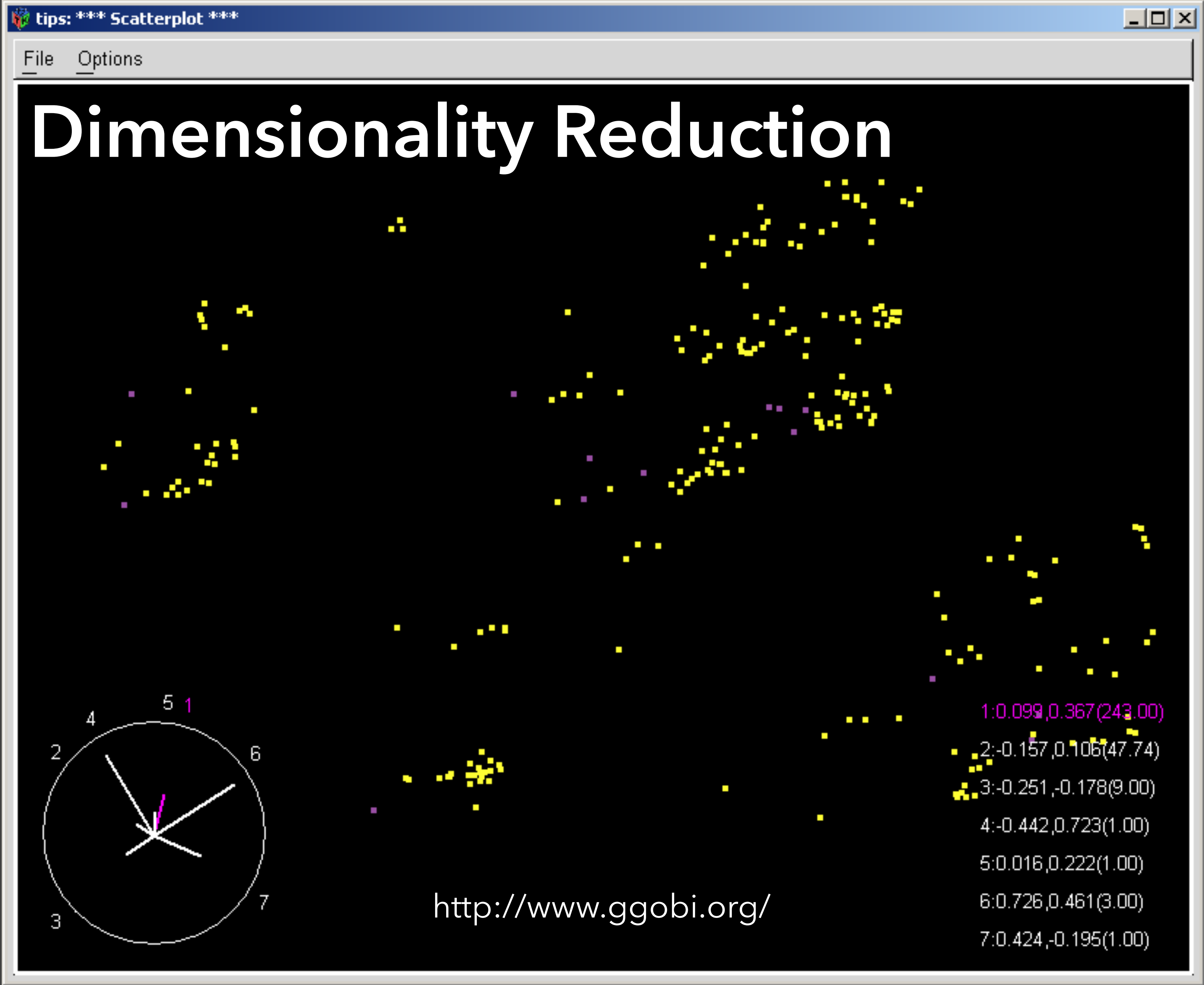
Consider F+C over **data** rather than just **view**

Dimensionality Reduction

Project from nD down to 2D.

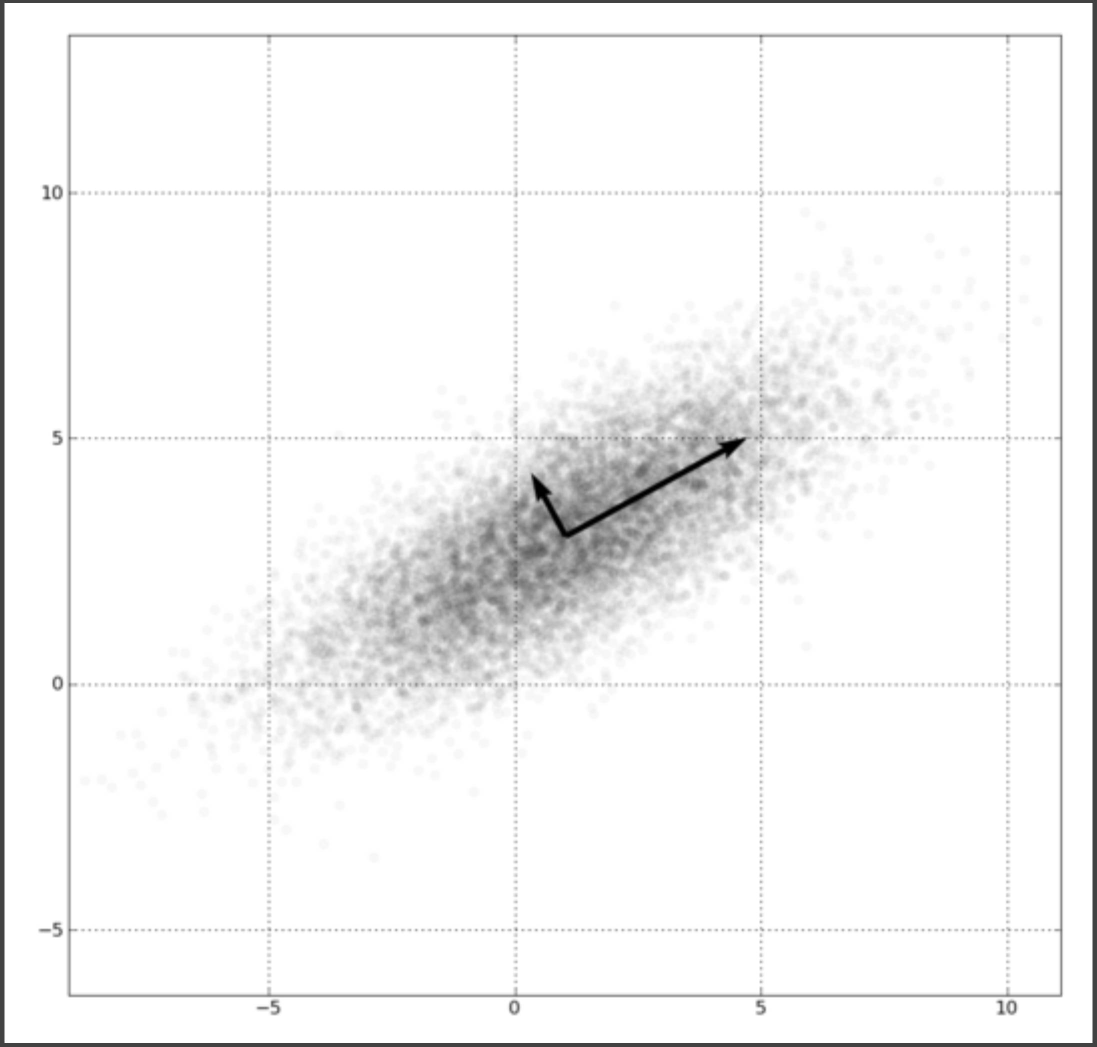
Spatial proximity –> data points are (approximately) more similar.

Dimensionality Reduction



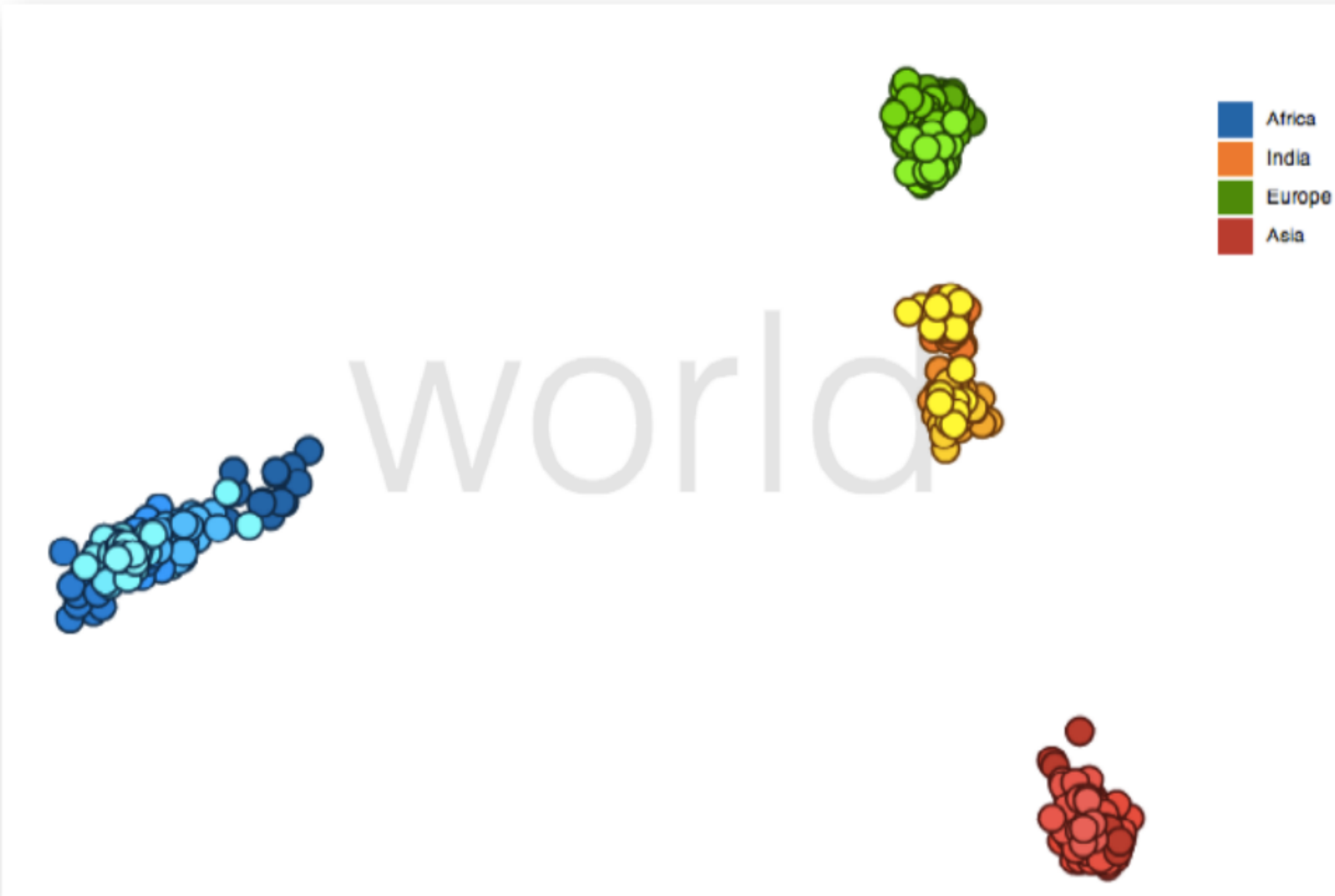
http://www.ggobi.org/

Principal Components Analysis

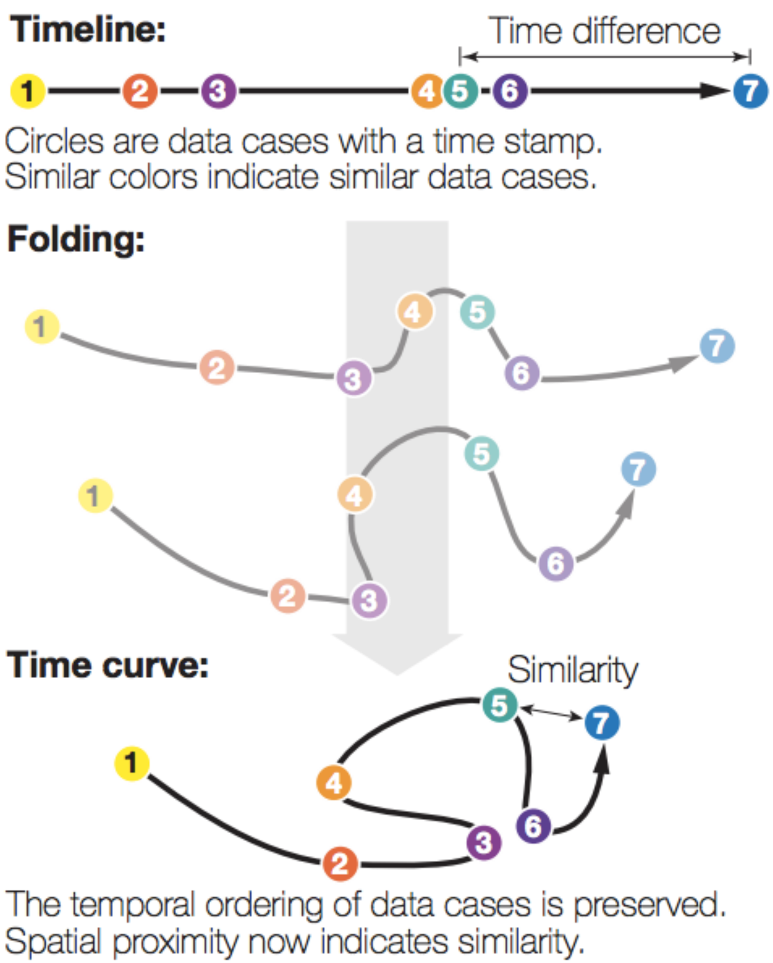


1. Mean-center the data.
2. Find ⊥ basis vectors that maximize the data variance.
3. Plot the data using the top vectors.

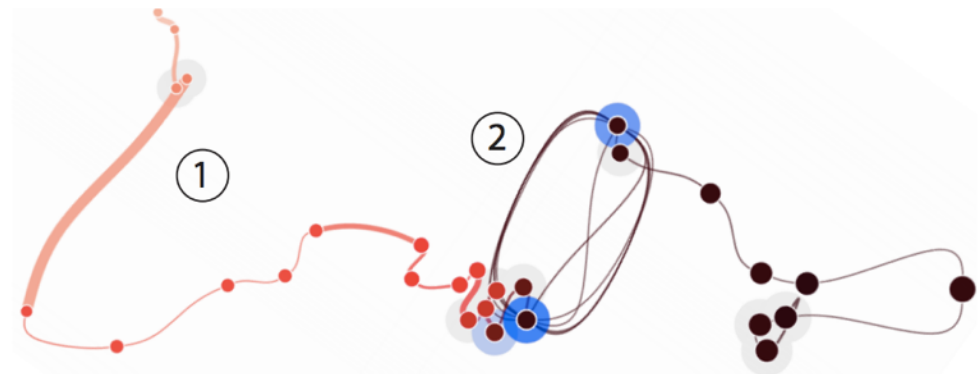
PCA of Genomes [Demiralp et al. '13]



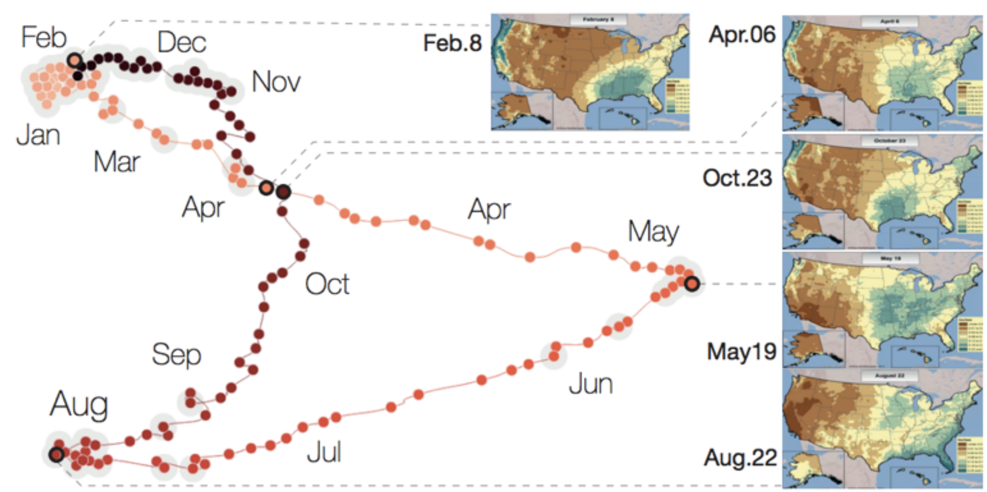
Time Curves [Bach et al. '16]



(a) Folding time



Wikipedia "Chocolate" Article



U.S. Precipitation over 1 Year

Many Reduction Techniques!

Principal Components Analysis (PCA)

Multidimensional Scaling (MDS)

Locally Linear Embedding (LLE)

t-Dist. Stochastic Neighbor Embedding (t-SNE)

Isomap

Auto-Encoder Neural Networks

Topological methods

…

Many Reduction Techniques!

Principal Components Analysis (PCA)

Multidimensional Scaling (MDS)

Locally Linear Embedding (LLE)

t-Dist. Stochastic Neighbor Embedding (t-SNE)

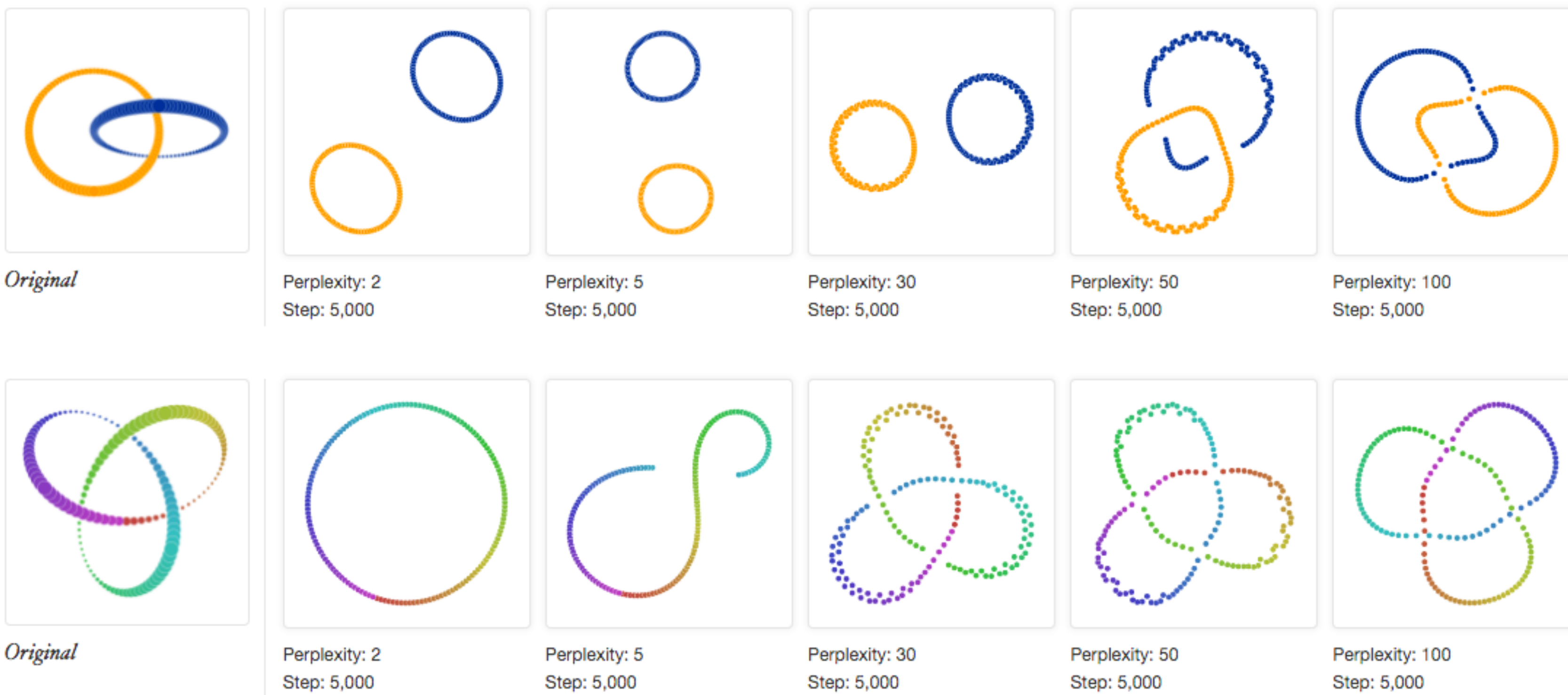
Isomap

Auto-Encoder Neural Networks

Topological methods

…

Visualizing t-SNE [Wattenberg et al. '16]



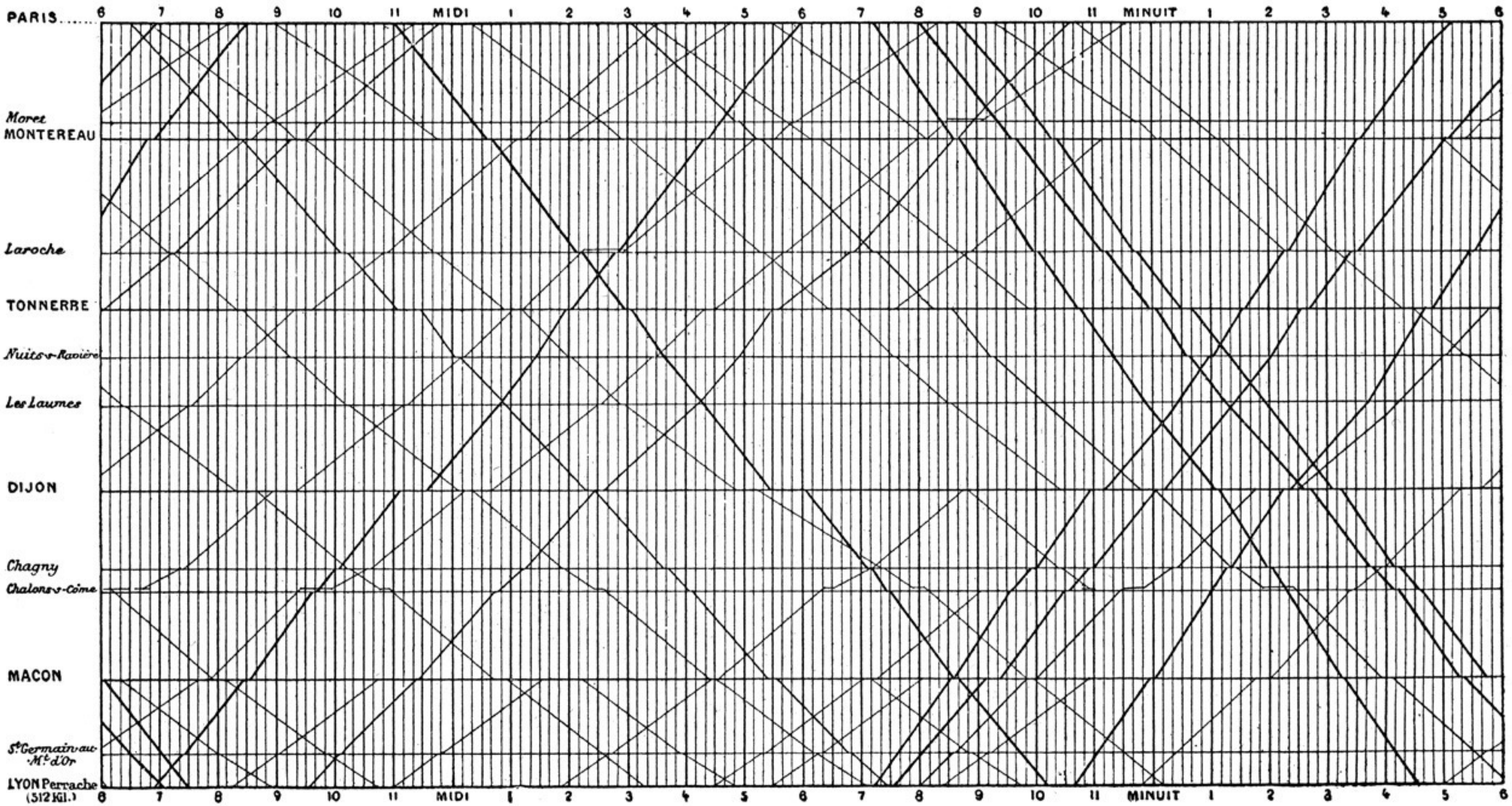
Summary

Spatial layout is the most important encoding
… but you need to be in the right space.

Geometric properties of spatial transforms support geometric reasoning

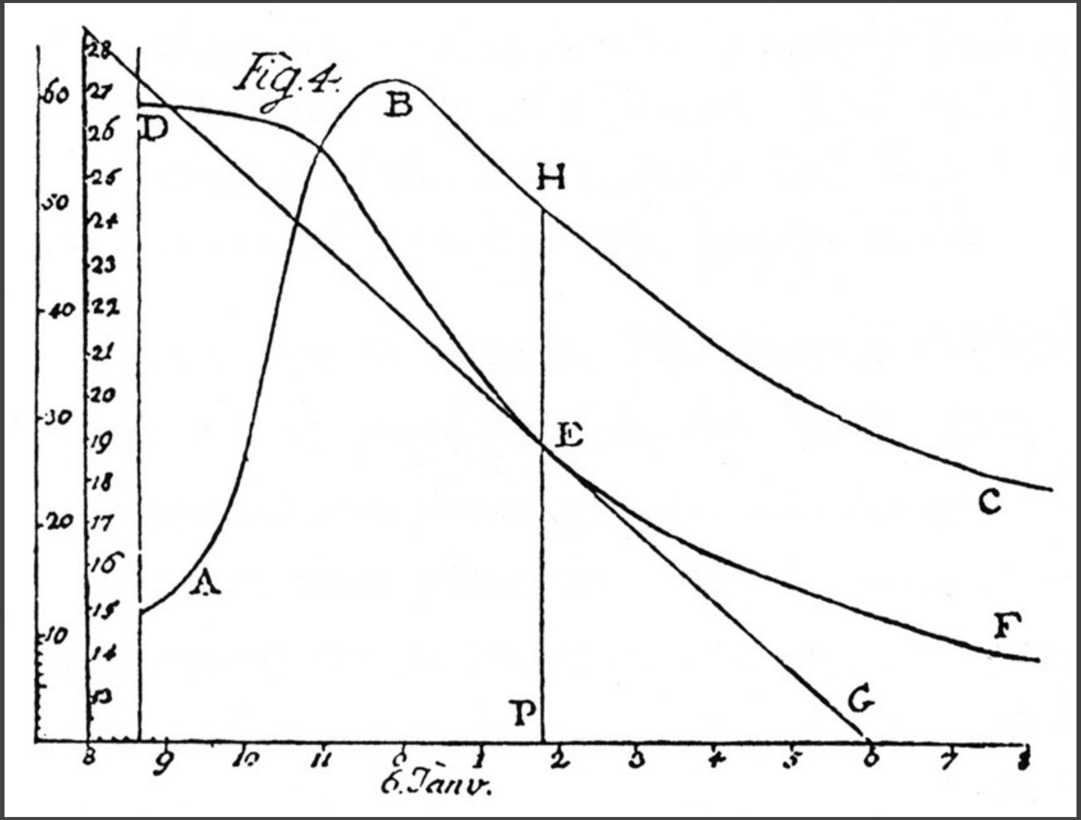
Emphasize important information
Consider *what* to show, not just *how*

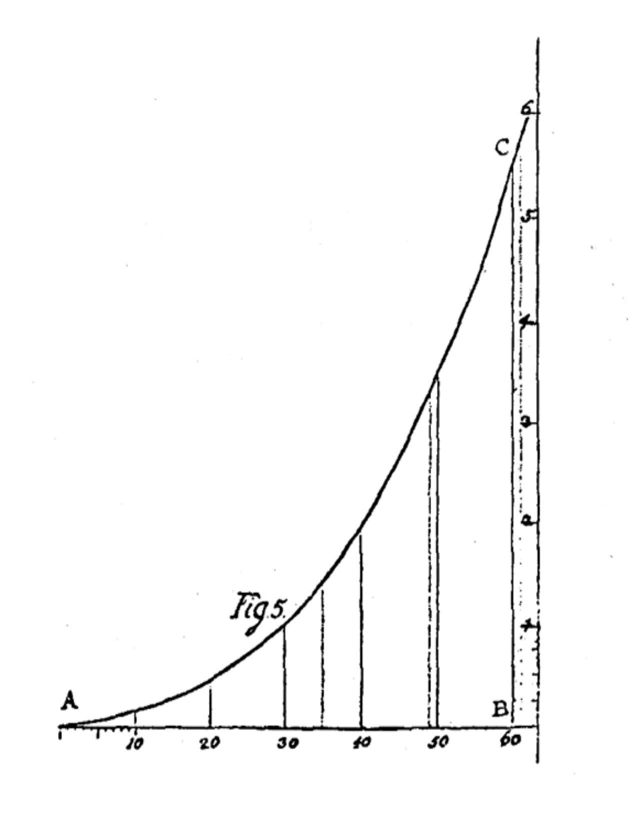
Graphical Calculation



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

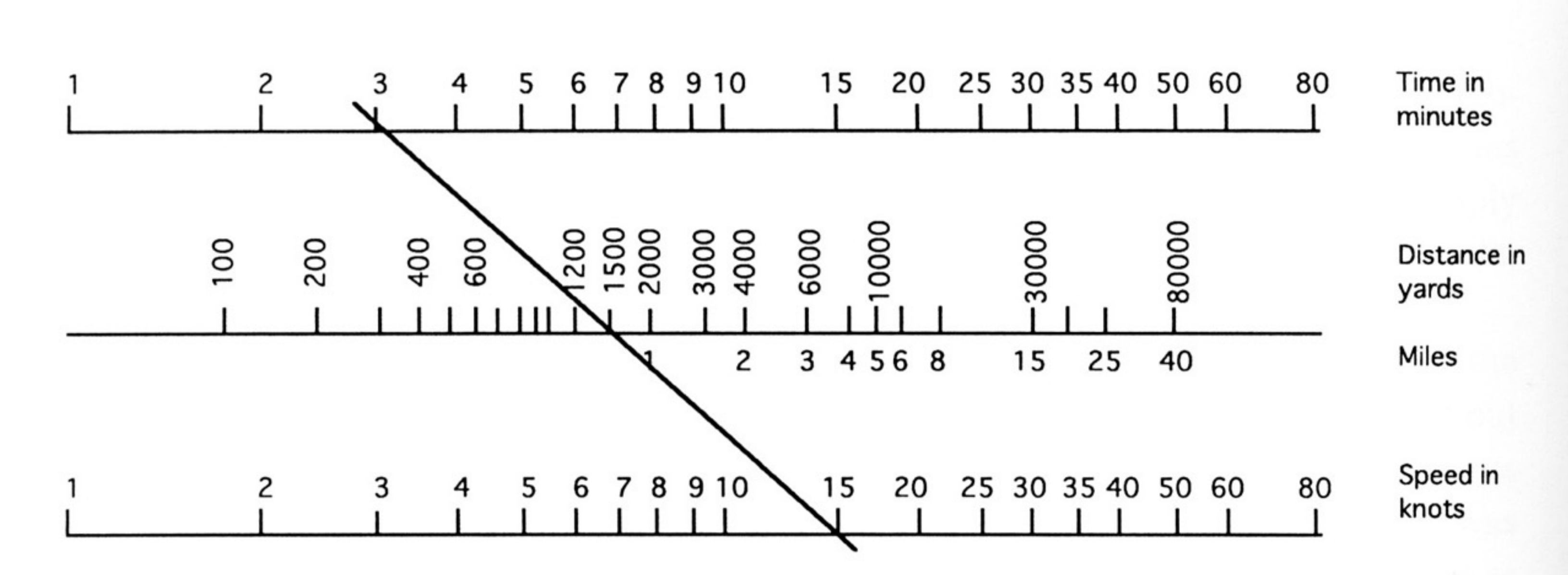
Lambert's Graphical Construction





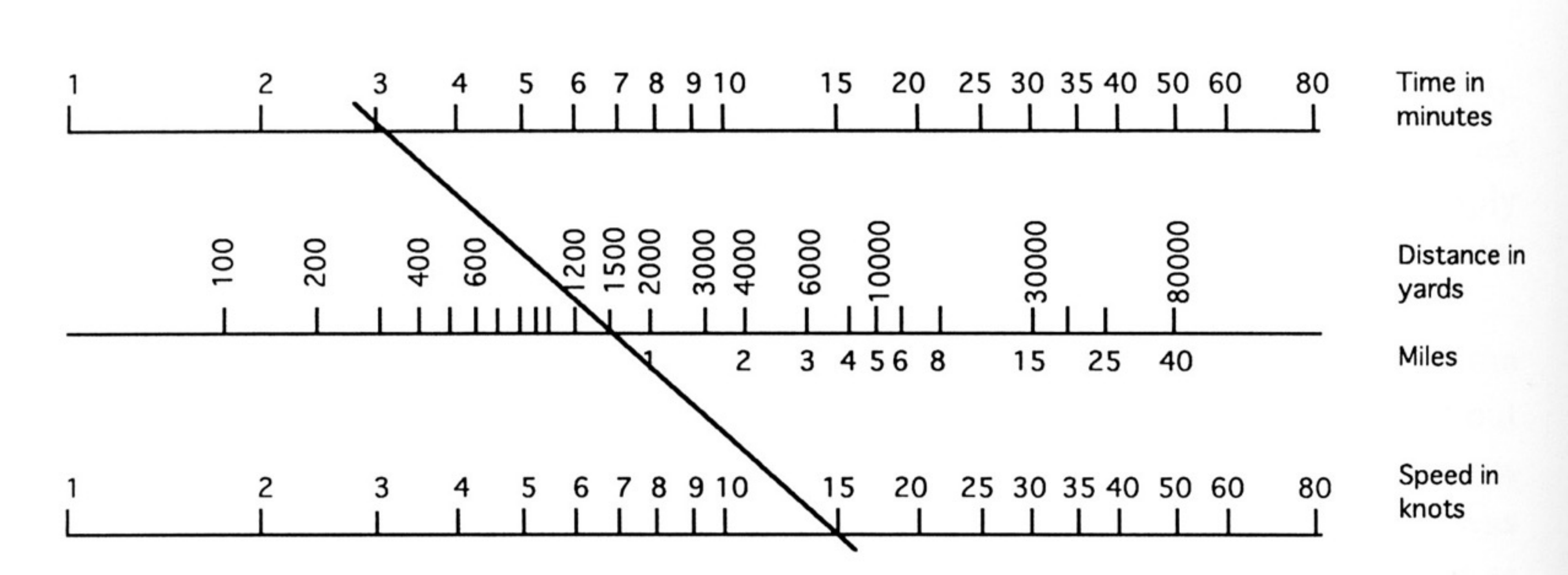
Johannes Lambert used graphs to study the rate of water evaporation as function of temperature [from Tufte 83].

Nomograms



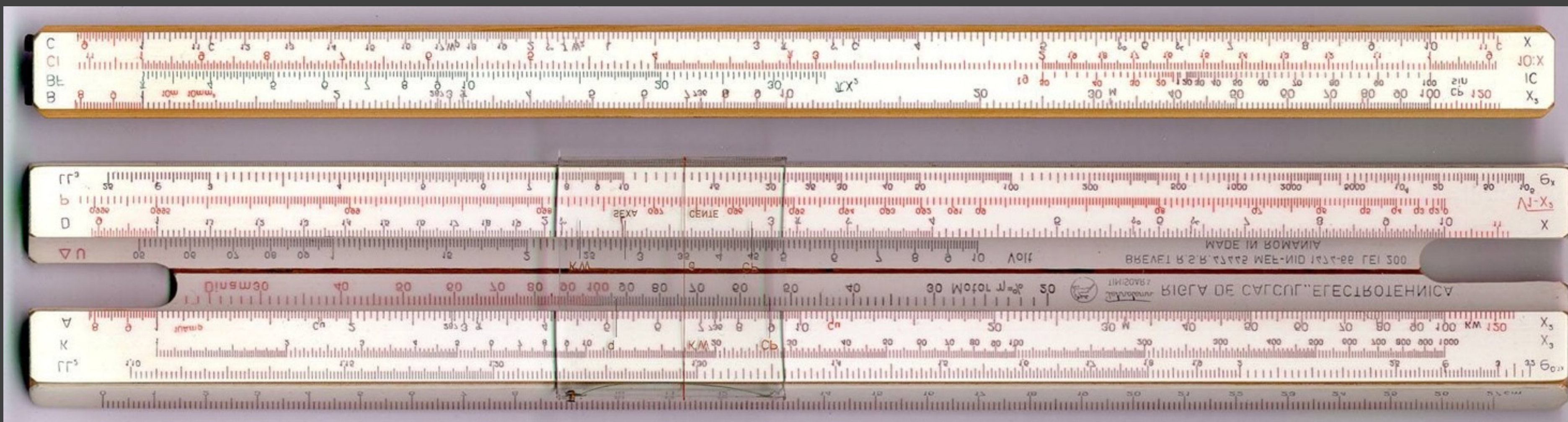
Sailing: The Rule of Three

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

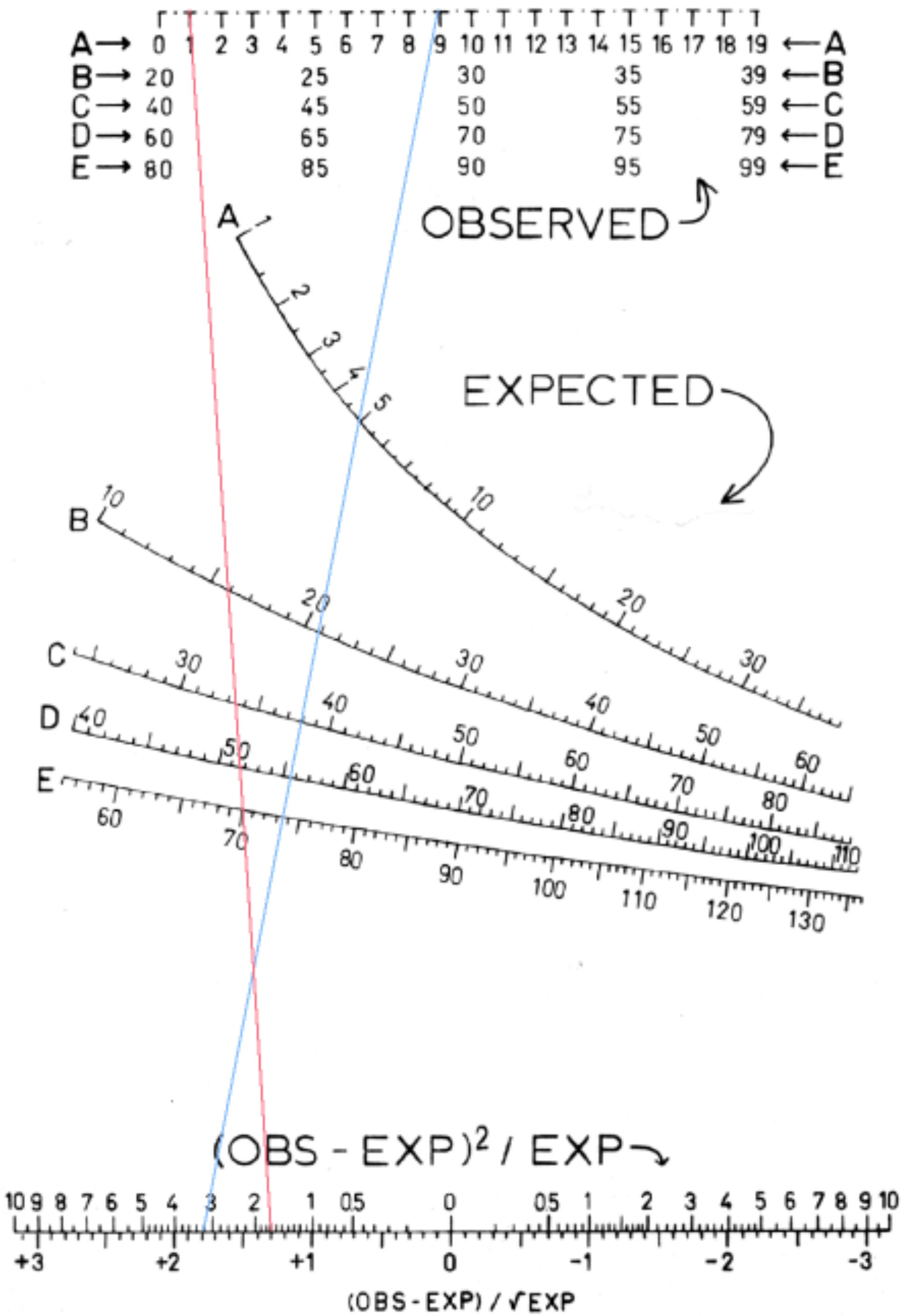
Slide Rule



Model 1474-66 Electrotechnica 18 Scales

Tehnolemn Timisoara Slide Rule Archive

http://pubpages.unh.edu/~jwc/tehnolemn/



Chi-Square Test

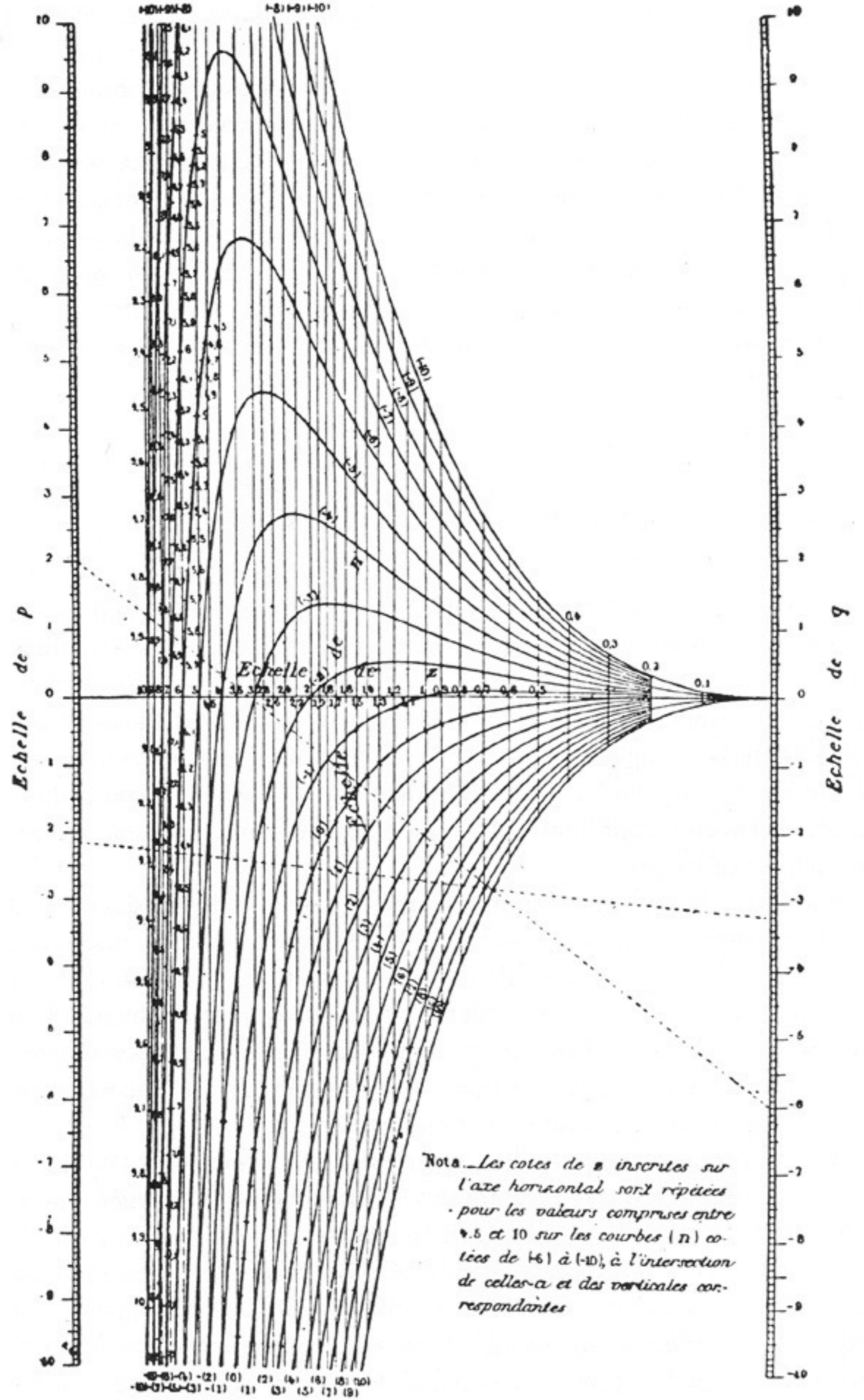
$(Obs - Exp)^2 / Exp$

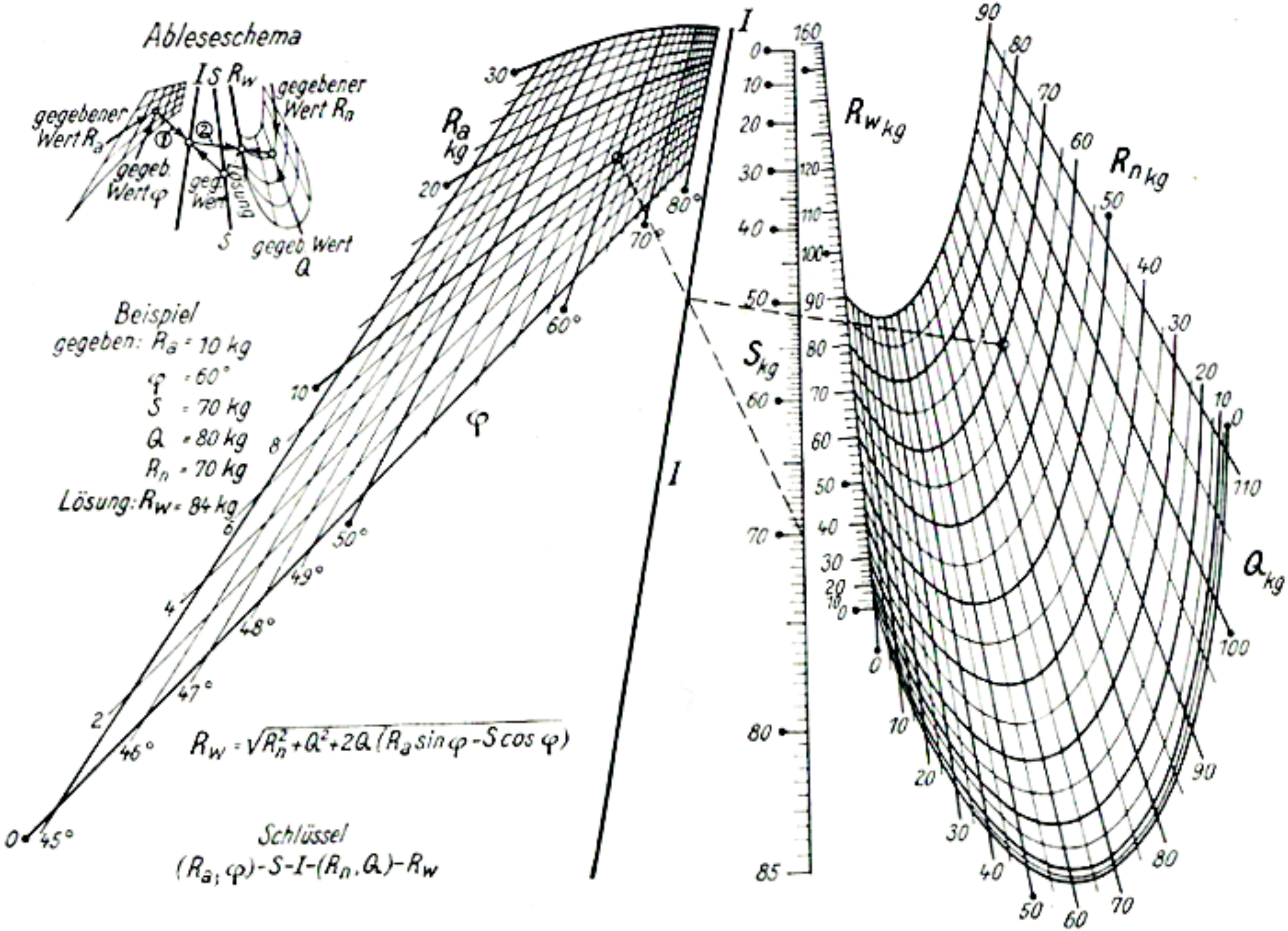
Blue line:

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

Red line:

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


Echelle de p
Echelle de q
Echelle de z
Echelle de n
Echelle
Nota. Les cotes de z inscrites sur l'axe horizontal sont répétées pour les valeurs comprises entre 4.5 et 10 sur les courbes (n) cotées de (-6) à (-10), à l'intersection de celles-ci et des verticales correspondantes



Ableseschema
I S Rw
gegebener Wert Ra
gegebener Wert Rn
gegeb. Wert φ
geg. Wert
Lösung
S
gegeb. Wert Q
Beispiel
gegeben: Ra = 10 kg
φ = 60°
S = 70 kg
Q = 80 kg
Rn = 70 kg
Lösung: Rw = 84 kg
$R_W = \sqrt{R_n^2 + Q^2 + 2Q(R_a \sin\varphi - S\cos\varphi)}$
Schlüssel
(Ra, φ)-S-I-(Rn, Q)-Rw
Ra kg
φ
I
Rw kg
S kg
Rn kg
Q kg