Using Space Effectively: 2D

Maneesh Agrawala

CS 294-10: Visualization
Fall 2007

Assignment 3: Visualization Software

Create an interactive visualization application – you choose data domain and visualization technique.

1. Describe data and storyboard interface
   due Oct 3 (before class)
2. Implement interface and produce final writeup
   due Oct 15 (before class)
3. Submit the application and a final writeup on the wiki

Can work alone or in pairs

Final write up due before class on Oct 15, 2007

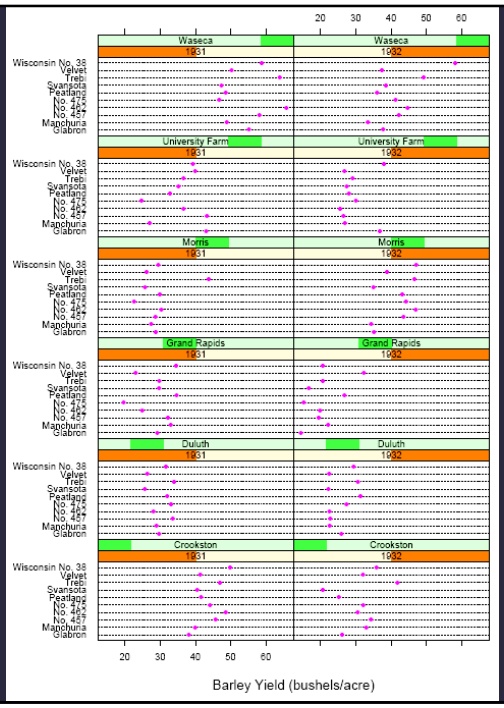
Topics

Rearrangements/Reorderable spaces
Displaying data in graphs
Banking to 45 degrees
Fitting data and depicting residuals
Displaying multidimensional data
Graphical calculations
Zooming and distortion

Reorderable Spaces

Trellis

[Becker, Cleveland, and Shyu 96]

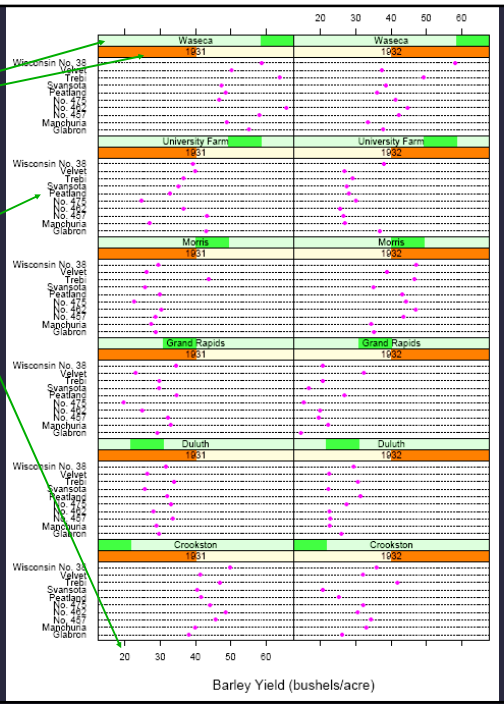


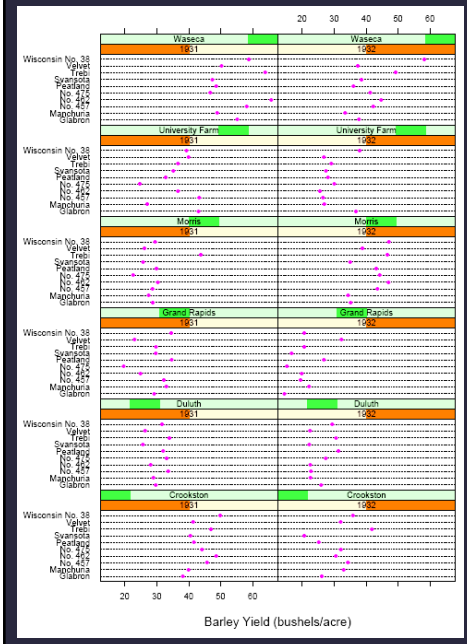
Trellis

[Becker, Cleveland, and Shyu 96]

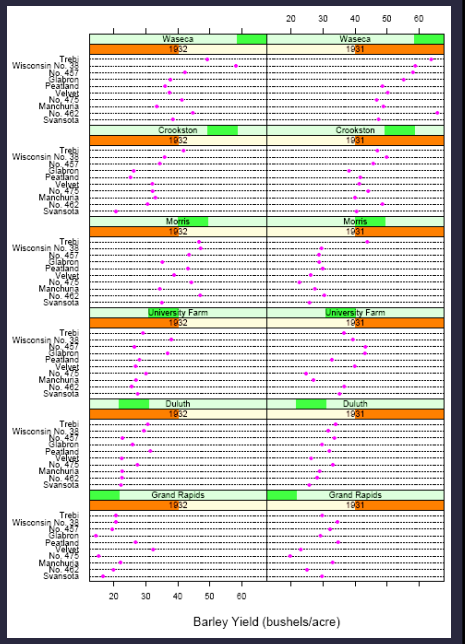
Condition variables
location, year

Panel variables
type, yield

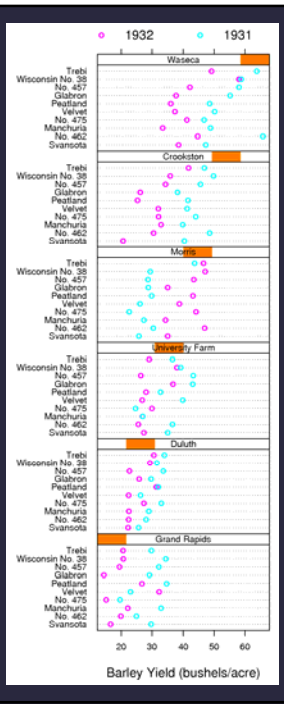




Alphabetical ordering



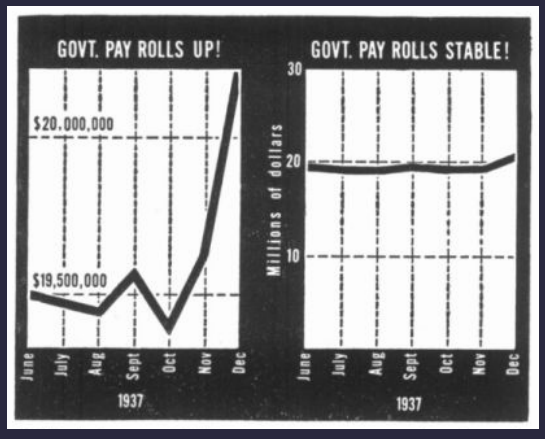
Main-effects ordering



Graphs and Lines

Effective use of space

Which graph is better?

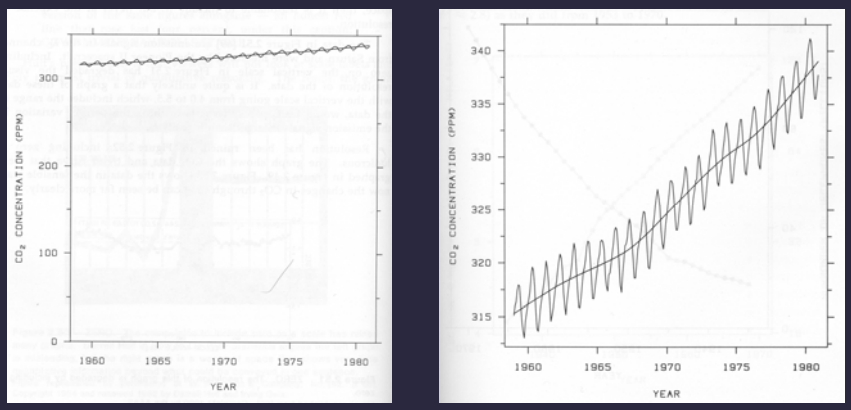


Government payrolls in 1937 [How To Lie With Statistics. Huff 93]

Aspect ratio

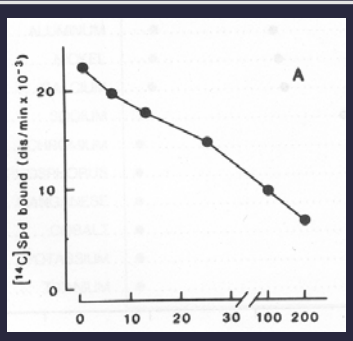
Fill space with data

Don't worry about showing zero

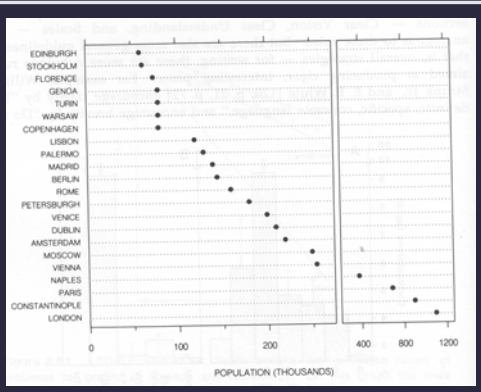


Yearly CO2 concentrations [Cleveland 85]

Clearly mark scale breaks

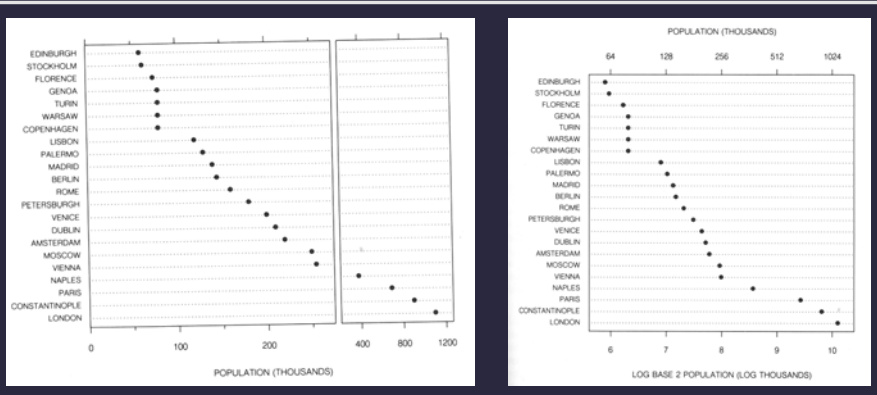


Poor scale break [Cleveland 85]



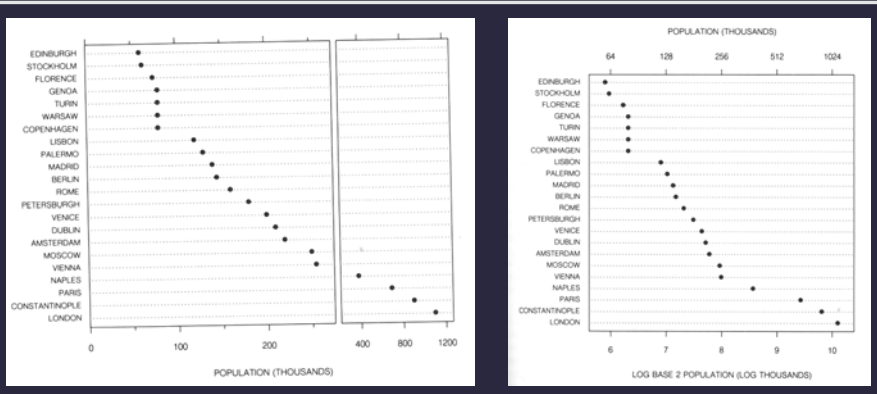
Well marked scale break [Cleveland 85]

Scale break vs. Log scale



[Cleveland 85]

Scale break vs. Log scale

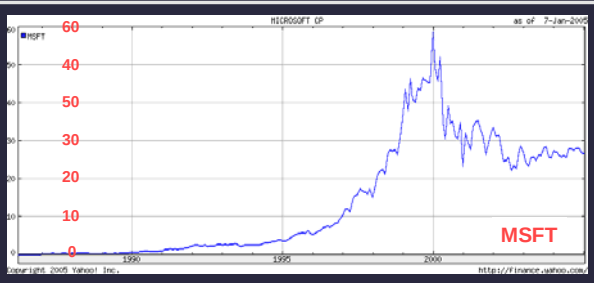


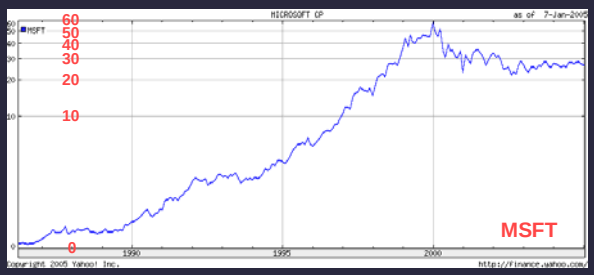
[Cleveland 85]

Both increase visual resolution

- **Log scale - easy comparisons of all data**
- **Scale break – more difficult to compare across break**

Linear scale vs. Log scale

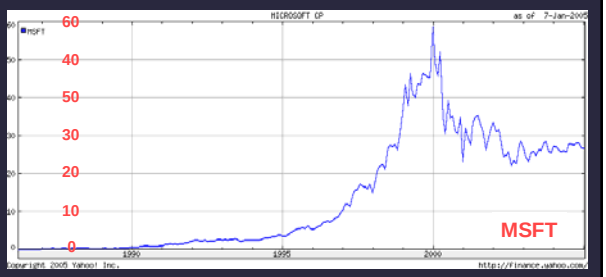




Linear scale vs. Log scale

Linear scale

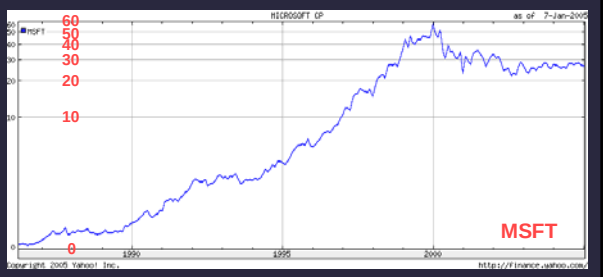
- Absolute change



Log scale

- Small fluctuations
- Percent change

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



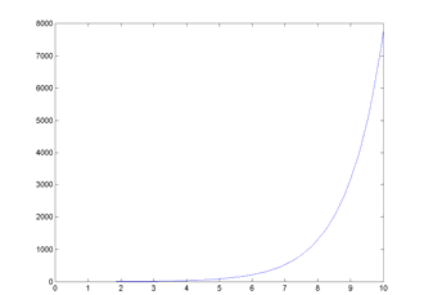
Semilog graph: Exponential growth

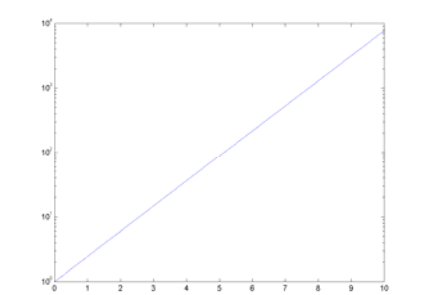
Exponential functions ($y = ka^{mx}$) transform into lines

$\log(y) = \log(k) + \log(a)mx$

Intercept: $\log(k)$

Slope: $\log(a)m$





$y = 6^{0.5x}$, slope in semilog space: $\log(6)*0.5 = 0.3891$

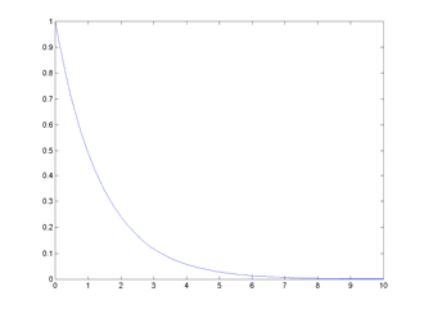
Semilog graph: Exponential decay

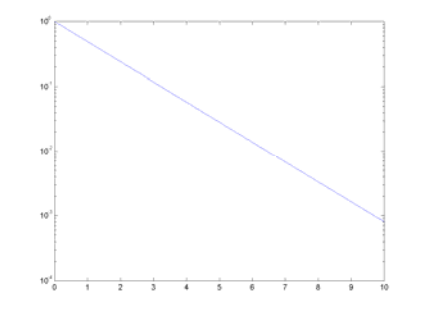
Exponential functions ($y = ka^{mx}$) transform into lines

$\log(y) = \log(k) + \log(a)mx$

Intercept: $\log(k)$

Slope: $\log(a)m$





$y = 0.5^{2x}$, slope in semilog space: $\log(0.7)*2 = -0.3098$

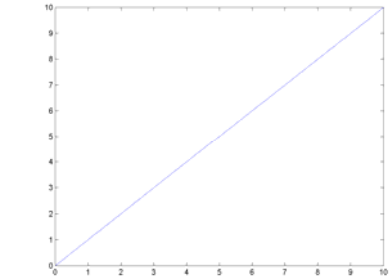
Semilog graph: Lines

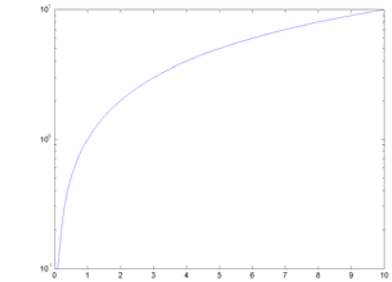
Exponential functions ($y = ka^{mx}$) transform into lines

$\log(y) = \log(k) + \log(a)mx$

Intercept: $\log(k)$

Slope: $\log(a)m$



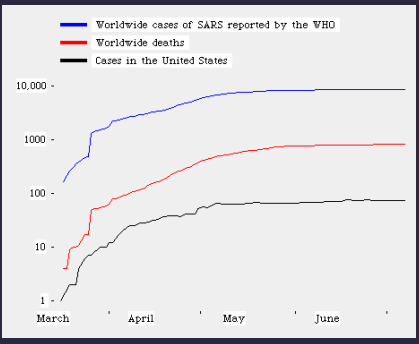


$y = x$, slope in semilog gives instantaneous : $\log(a)m$

Semilog graph

Exponential functions ($y = ka^{mx}$) transform into lines

$\log(y) = \log(k) + \log(a)mx$



SARS cases up March – July 7, 2003

http://www.squeak.org/us/ted/sars-graph.html

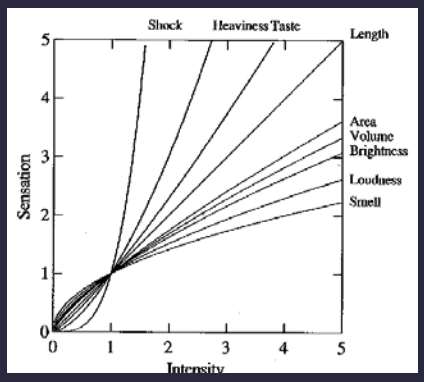
AIDS Cases: http://www.righto.com/java/statsgraph.html

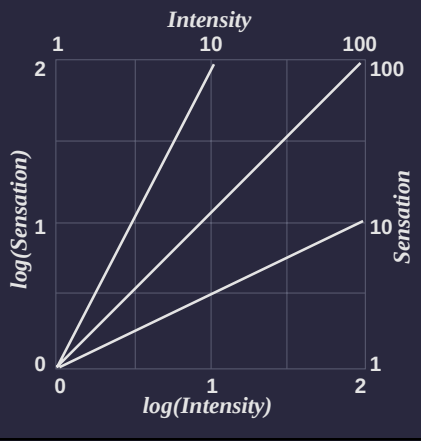
Log-Log graph

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

Example - Steven's power laws:

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



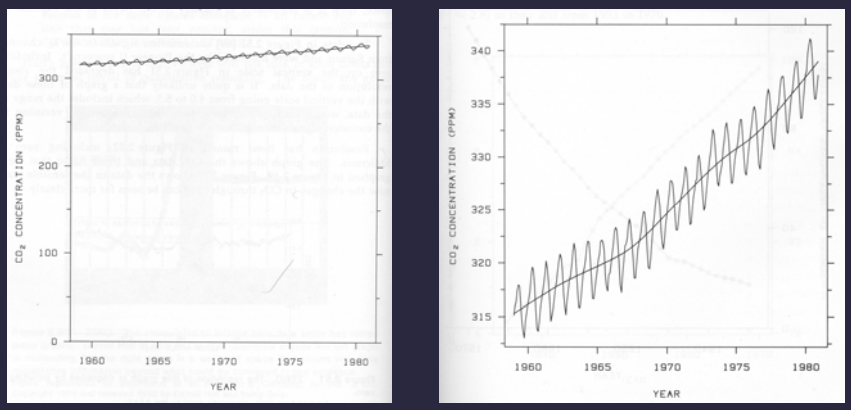


Banking to 45 Degrees

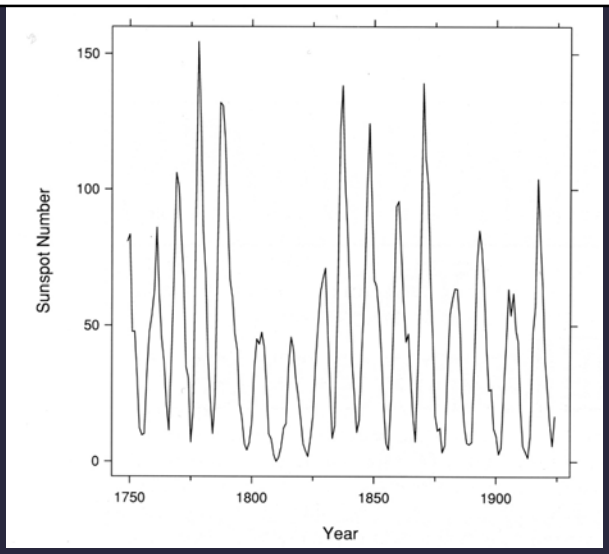
Aspect ratio

Fill space with data

Don’t worry about showing zero



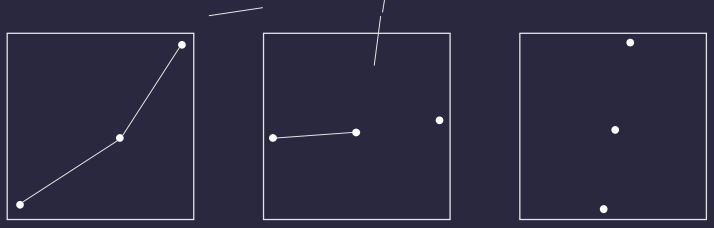
Yearly CO2 concentrations [Cleveland 85]



William S. Cleveland
The Elements of Graphing Data

Banking to 45 degrees

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

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

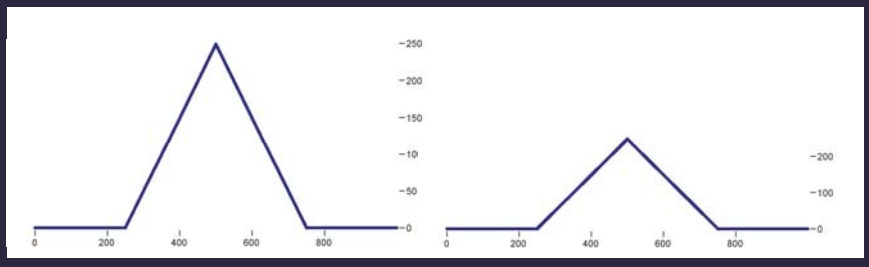
Global (over all i, j s.t. i≠j)

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

Local (over adjacent segments)

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

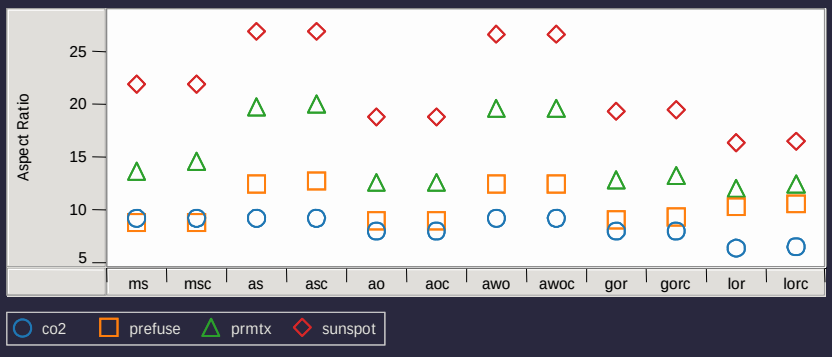
Slopeless line culling



Standard, Aspect Ratio = 1.97 Culled, Aspect Ratio = 4.00

Exclude line segments with zero or infinite slope

Comparison (Results)



Discussion

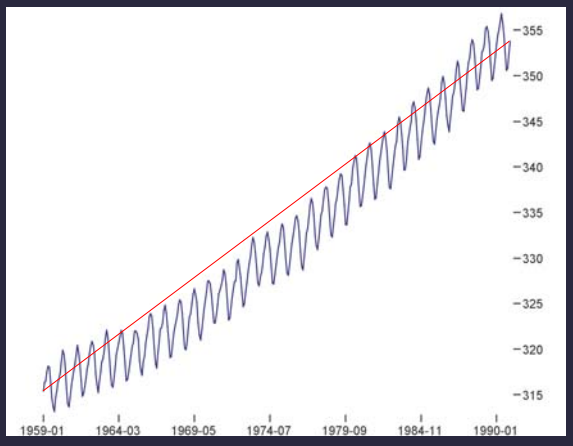
Due to computational complexity…

- **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



Aspect Ratio = 1.17

Aspect Ratio = 7.87

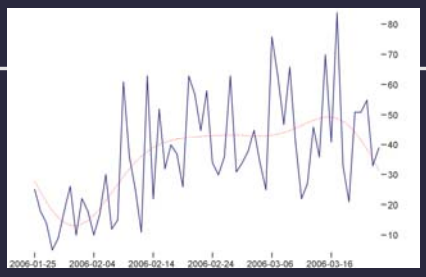
CO_2 Measurements

William S. Cleveland
Visualizing Data

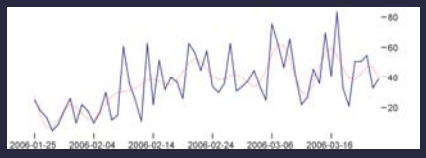
Downloads

Two months of daily prefuse downloads

Aspect Ratio = 1.44



Aspect Ratio = 2.89



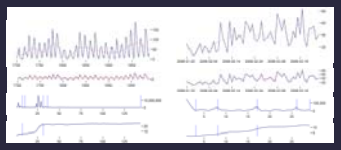
Aspect Ratio = 8.81

Power Spectrum

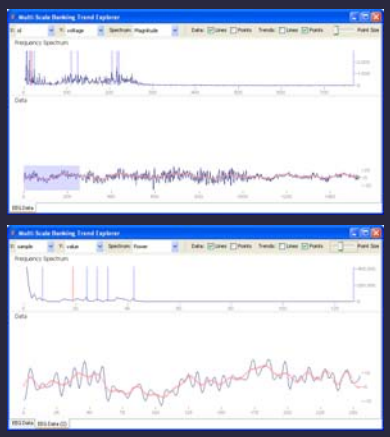
Aspect Ratios

Applications

Small Multiples Displays

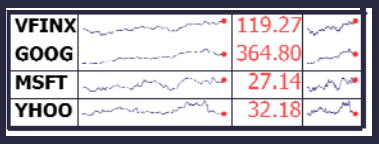


Trend Explorer

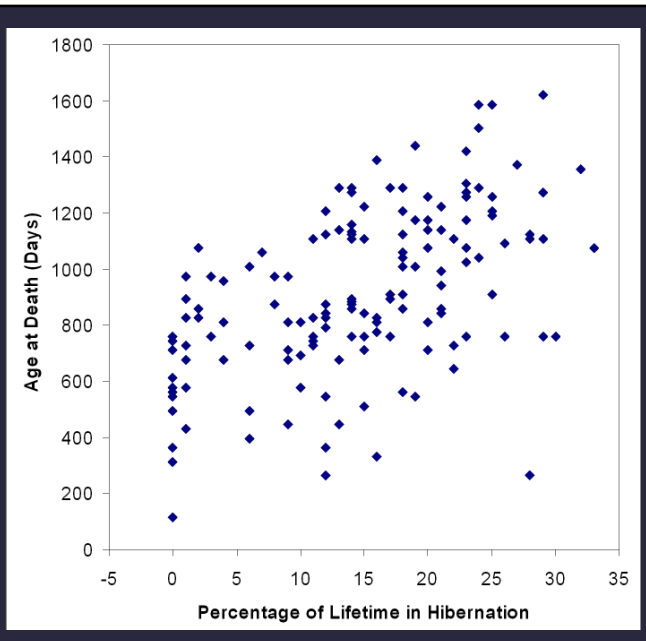


Sparklines

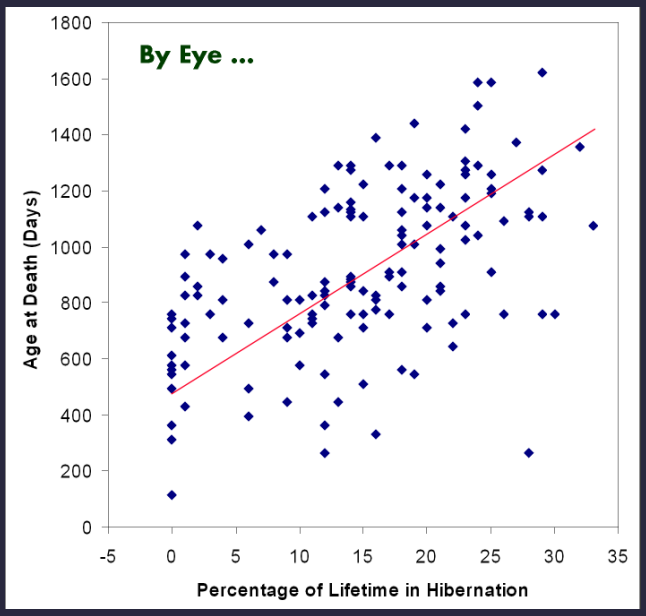
Banking can be applied to create *sparklines*, data-intense, word-sized graphics. A plot might be included inline  , supporting uninterrupted reading.



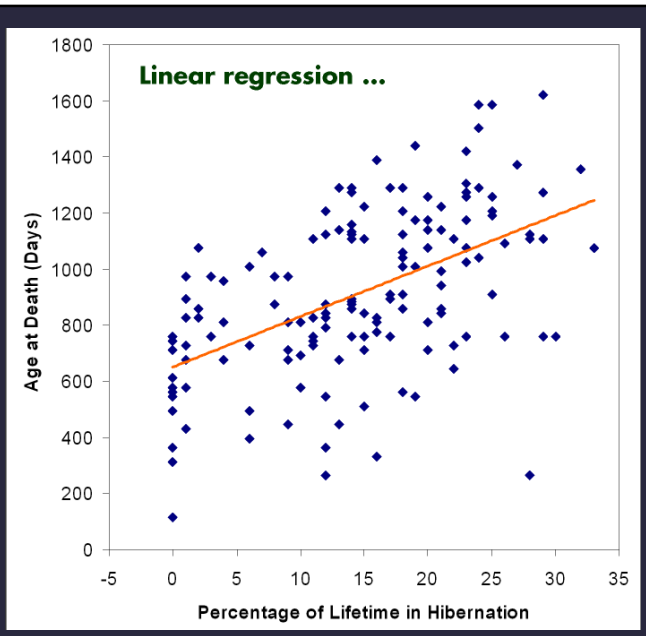
Fitting the Data



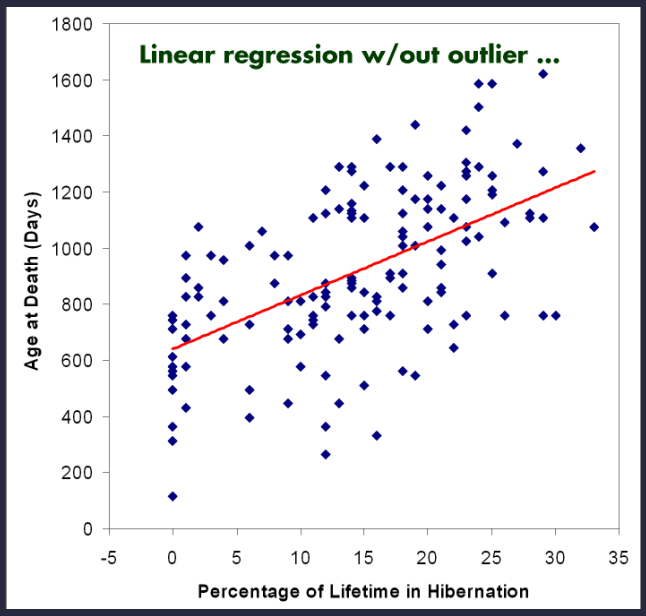
[The Elements of Graphing Data. Cleveland 94]



[The Elements of Graphing Data. Cleveland 94]



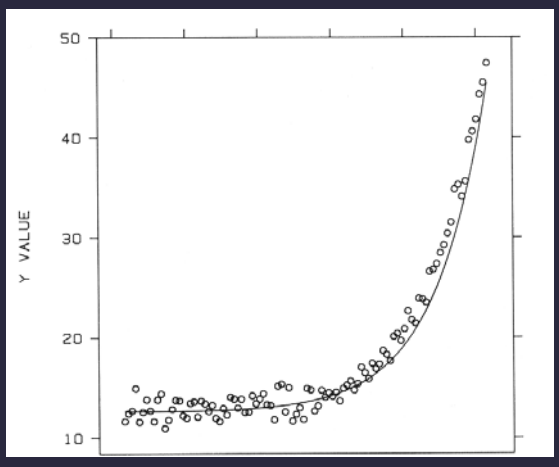
[The Elements of Graphing Data. Cleveland 94]



[The Elements of Graphing Data. Cleveland 94]

Transforming data

How well does curve fit data?

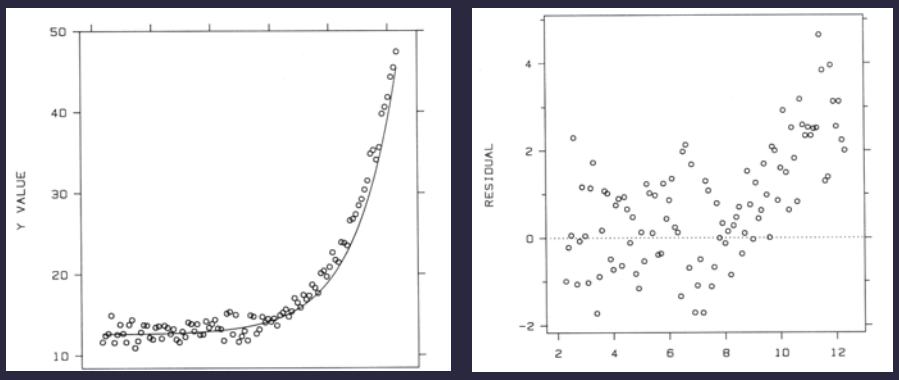


[Cleveland 85]

Transforming data

Residual graph

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



[Cleveland 85]

Most powerful brain?

Microsoft Excel - animal.xls

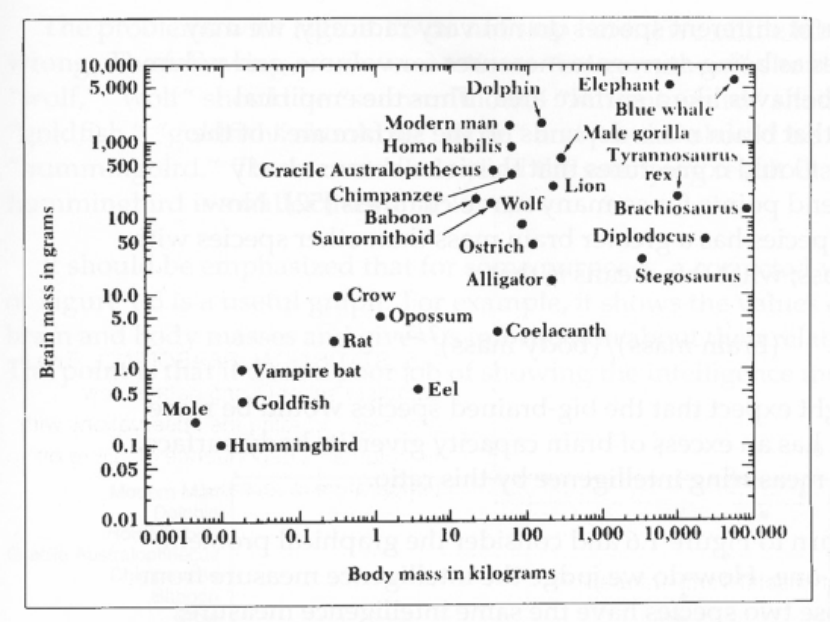
File Edit View Insert Format Tools Data Window Help

A1 ID

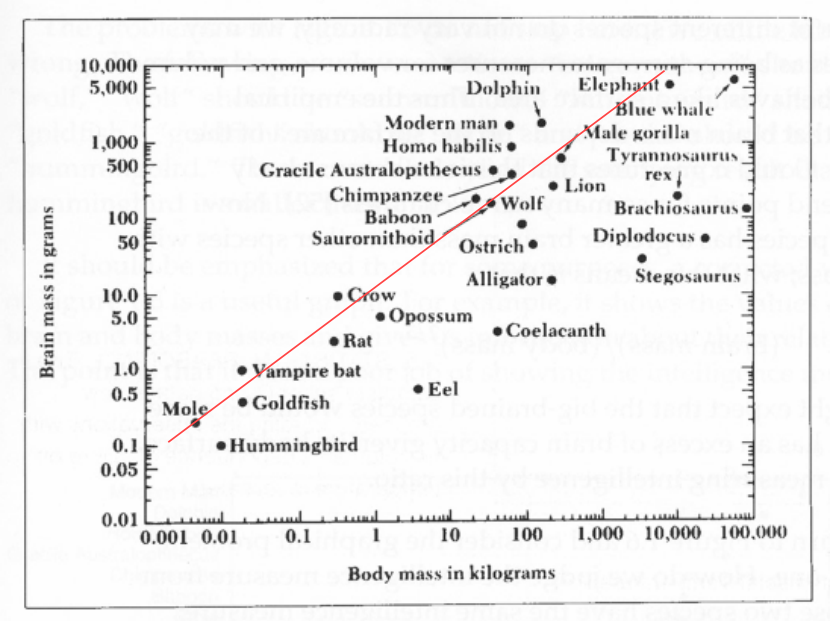
	A	B	C	D	E
1	ID	Name	Body Weight	Brain Weight	
2	1	Lesser Short-tailed Shrew	5	0.14	
3	2	Little Brown Bat	10	0.25	
4	3	Mouse	23	0.3	
5	4	Big Brown Bat	23	0.4	
6	5	Musk Shrew	48	0.33	
7	6	Star Nosed Mole	60	1	
8	7	Eastern American Mole	75	1.2	
9	8	Ground Squirrel	101	4	
10	9	Tree Shrew	104	2.5	
11	10	Golden Hamster	120	1	
12	11	Mole Rate	122	3	
13	12	Galago	200	5	
14	13	Rat	280	1.9	
15	14	Chinchilla	425	6.4	
16	15	Desert Hedgehog	550	2.4	
17	16	Rock Hyrax (a)	750	12.3	
18	17	European Hedgehog	785	3.5	
19	18	Tenrec	900	2.6	
20	19	Arctic Ground Squirrel	920	5.7	
21	20	African Giant Pouched Rat	1000	6.6	
22	21	Guinea Pig	1040	5.5	
23	22	Mountain Beaver	1350	8.1	
24	23	Slow Loris	1400	12.5	
25	24	Genet	1410	17.5	
26	25	Phalanger	1620	11.4	

animal

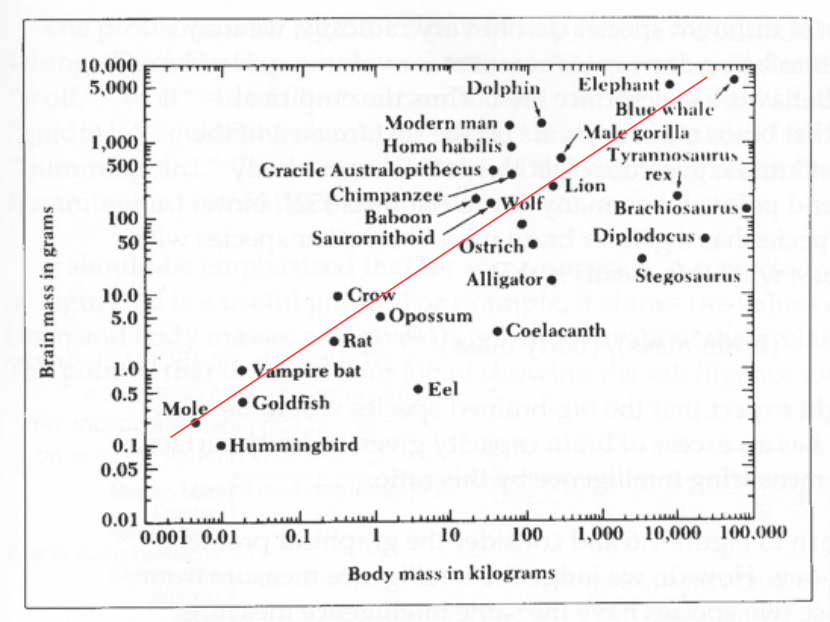
Ready



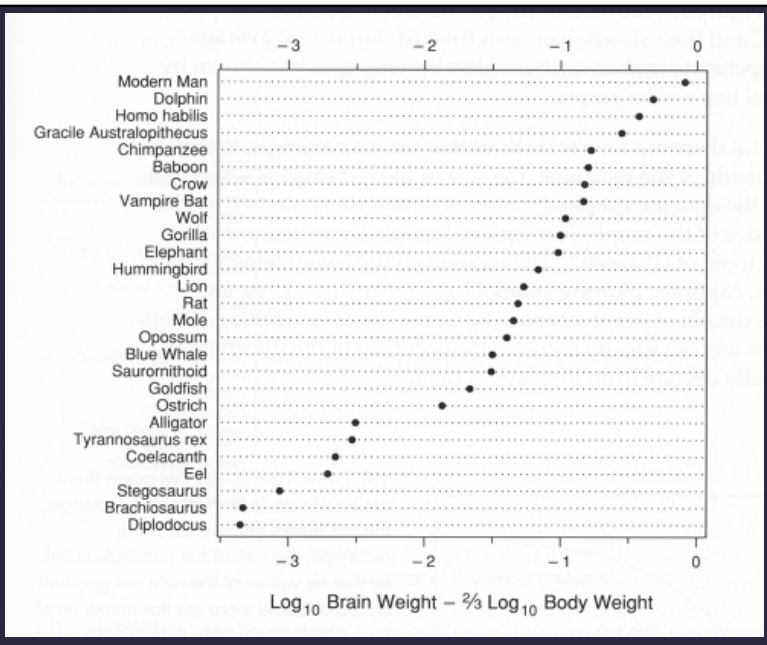
The Dragons of Eden [Carl Sagan]



The Dragons of Eden [Carl Sagan]



The Dragons of Eden [Carl Sagan]



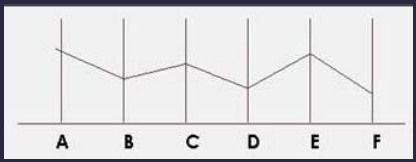
The Elements of Graphing Data [Cleveland]

Parallel Coordinates

Parallel coordinates

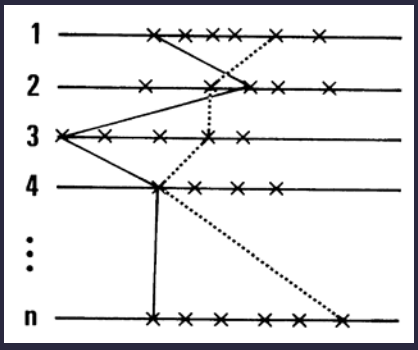
Visualizing nD in planar image

- Only 2 orthogonal axes
- Use parallel axes instead



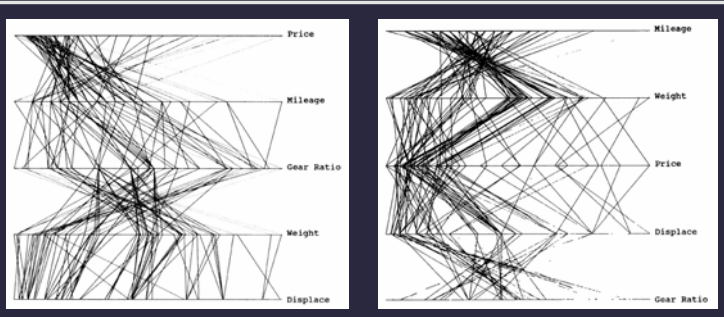
Plot each dimension of point x on separate axis

- $x = (a, b, c, d, \ldots)$



[Wegman 90]

Parallel coordinates: Axis ordering

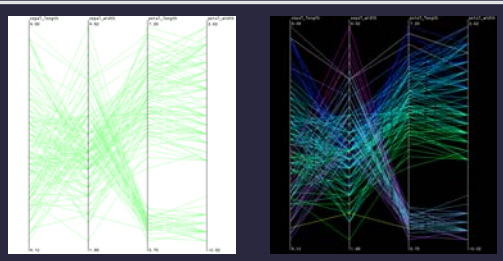


5D Automobile Data [Wegman 90]

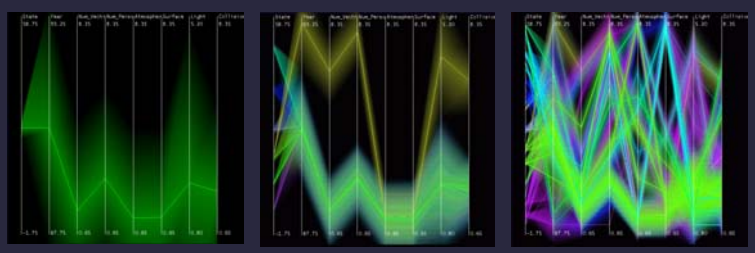
No intrinsic axis order

- Interactive axis swap
  - Bad: Relies on human examination
  - Good: Powerful interaction
- Machine learning of axis order [Inselberg 99]

Parallel coordinates



Proximity-based coloring



Visualizing hierarchical clusters [Fua et al. 99]