Using Space Effectively

Maneesh Agrawala

CS 448B: Visualization
Winter 2020

1

Last Time: EDA

2

Data “Wrangling”

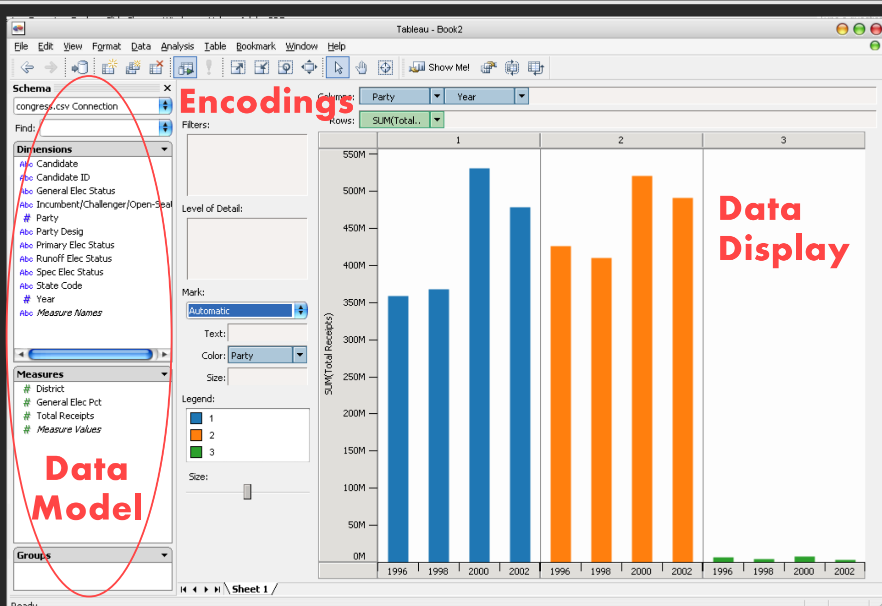
One often needs to manipulate data prior to analysis. Tasks include reformatting, cleaning, quality assessment, and integration

Some approaches:

- Writing custom scripts
- Manual manipulation in spreadsheets
- Trifacta Wrangler: http://trifacta.com/products/wrangler/
- Open Refine: http://openrefine.org

3

Tableau



4

Specifying Table Configurations

Operands are names of database fields

Each operand interpreted as a set {…}

Data is either O or Q and treated differently

Three operators:

- concatenation (+)
- cross product (x)
- nest (/)

6

Table Algebra

The operators (+,x,/) and operands (O,Q) provide an algebra for tabular visualization

Algebraic statements are mapped to

- **Visualizations** – trellis partitions, visual encodings
- **Queries** – selection, projection, group-by

In Tableau, users make statements via drag-and-drop

- Users specify operands NOT operators!
- Operators are inferred by data type (O,Q)

13

Table Algebra: Operands

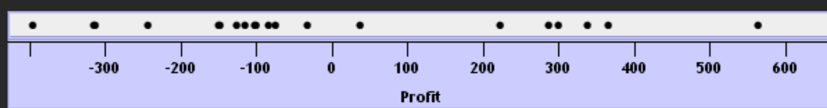
Ordinal fields: interpret domain as a set that partitions table into rows and columns

Quarter = {(Qtr1),(Qtr2),(Qtr3),(Qtr4)} →

Qtr1	Qtr2	Qtr3	Qtr4
95892	101760	105282	98225

Quantitative fields: treat domain as single element set and encode spatially as axes

Profit = {(Profit[-410,650])} →



14

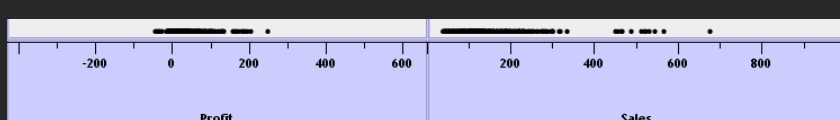
Concatenation (+) Operator

Ordered union of set interpretations

Quarter + Product Type
= {(Qtr1),(Qtr2),(Qtr3),(Qtr4)} + {(Coffee), (Espresso)}
= {(Qtr1),(Qtr2),(Qtr3),(Qtr4),(Coffee),(Espresso)}

Qtr1	Qtr2	Qtr3	Qtr4	Coffee	Espresso
48	59	57	53	151	21

Profit + Sales = {(Profit[-310,620]),(Sales[0,1000])}



15

Cross (x) Operator

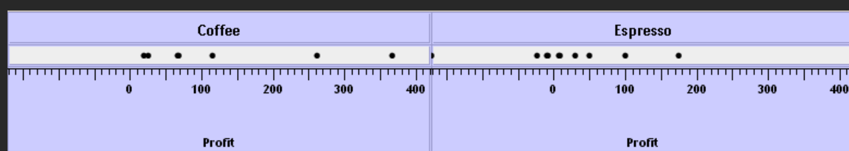
Cross-product of set interpretations

Quarter x Product Type

= {(Qtr1,Coffee), (Qtr1, Tea), (Qtr2, Coffee), (Qtr2, Tea), (Qtr3, Coffee), (Qtr3, Tea), (Qtr4, Coffee), (Qtr4,Tea)}

Qtr1		Qtr2		Qtr3		Qtr4	
Coffee	Espresso	Coffee	Espresso	Coffee	Espresso	Coffee	Espresso
131	19	160	20	178	12	134	33

Product Type x Profit =



Nest (/) Operator

Cross-product filtered by existing records

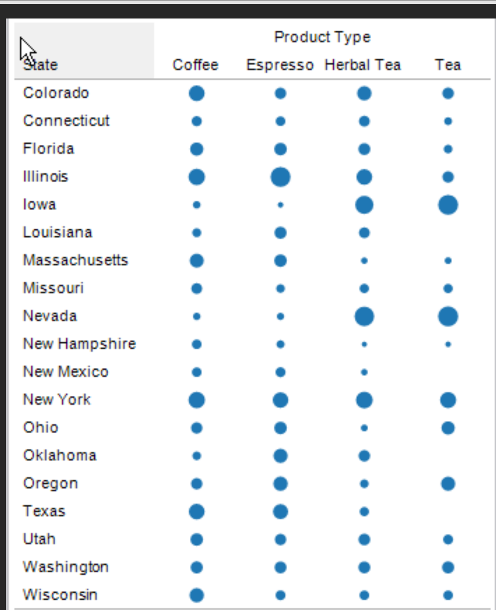
Quarter x Month

creates 12 entries for each qtr. i.e., (Qtr1, Dec)

Quarter / Month

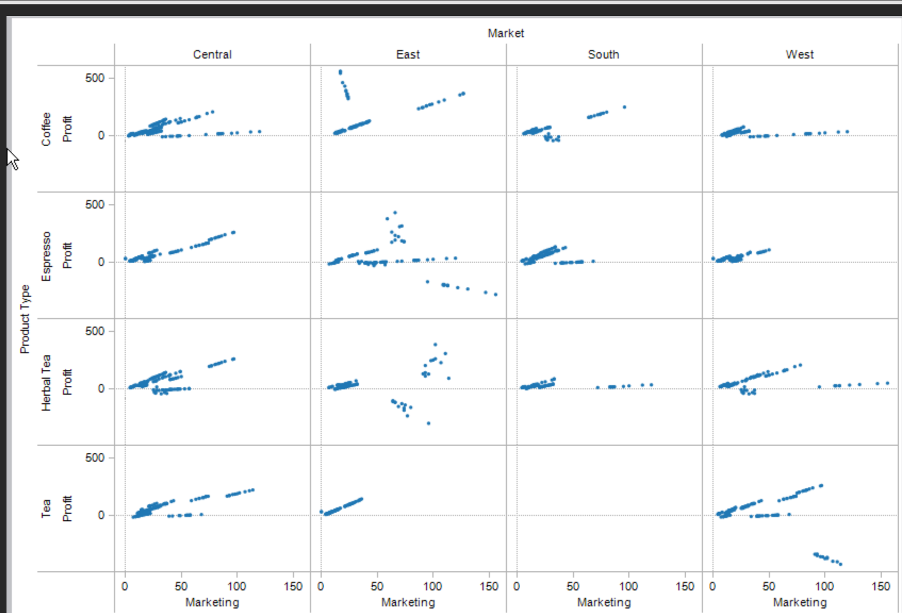
creates three entries per quarter based on tuples in database (not semantics)

Ordinal - Ordinal



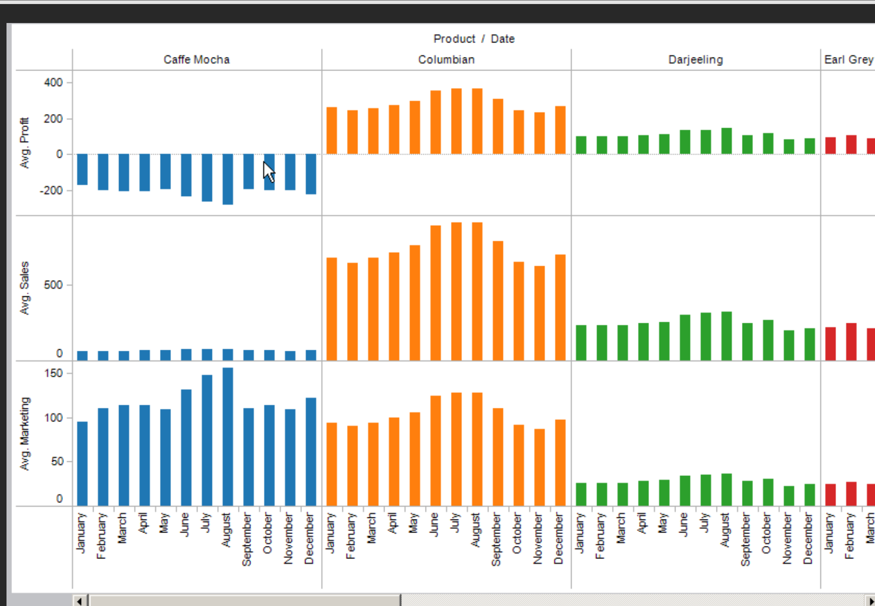
18

Quantitative - Quantitative



19

Ordinal - Quantitative



20

Summary

Exploratory analysis may combine graphical methods, and statistics

Use questions to uncover more questions

Interaction is essential for exploring large multidimensional datasets

21

Announcements

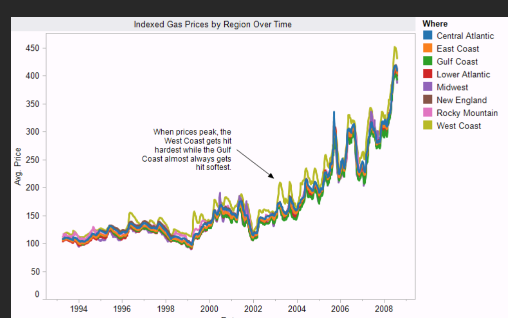
22

A2: Exploratory Data Analysis

Use Tableau to formulate & answer questions

First steps

Step 1: Pick domain & data
Step 2: Pose questions
Step 3: Profile data
Iterate as needed



Create visualizations

Interact with data
Refine questions

Author a report

Screenshots of most insightful views (10+)
Include titles and captions for each view

Due before class on Jan 27, 2020

23

Using Space Effectively

26

Topics

Graphs and lines
Selecting aspect ratio
Fitting data and depicting residuals
Graphical calculations
Cartographic distortion

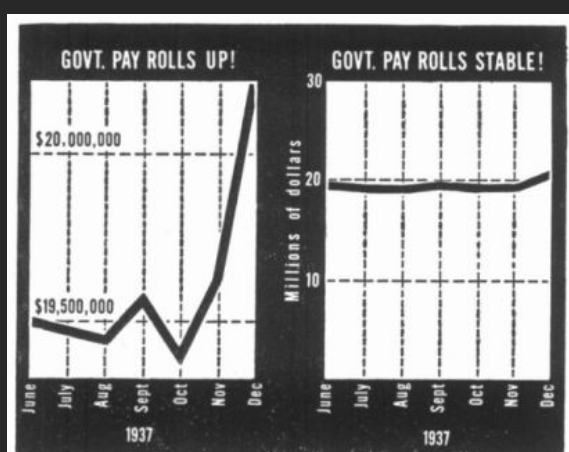
27

Graphs and Lines

28

Effective use of space

Which graph is better?



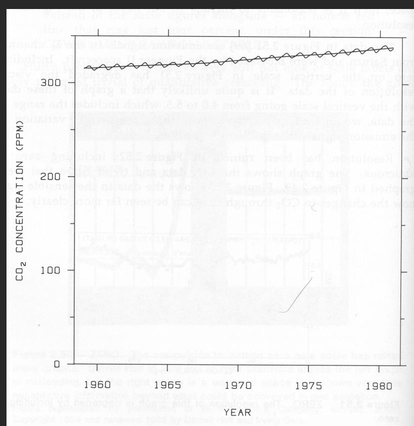
Government payrolls in 1937 [Huff 93]

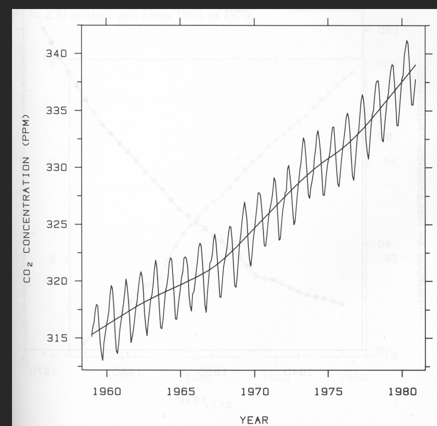
29

Aspect ratio

Fill space with data

Don't worry about showing zero

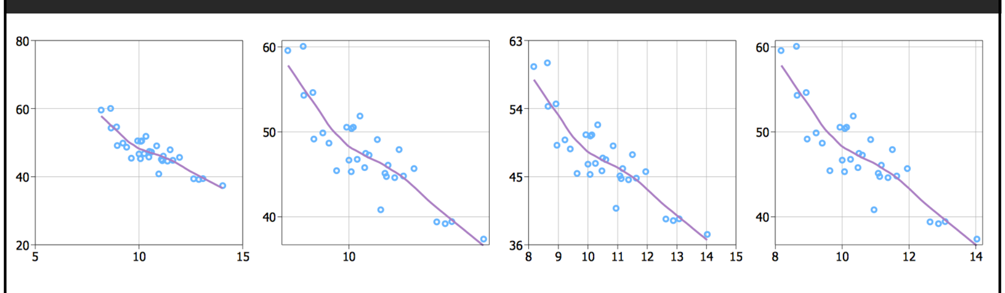




Yearly CO2 concentrations [Cleveland 85]

30

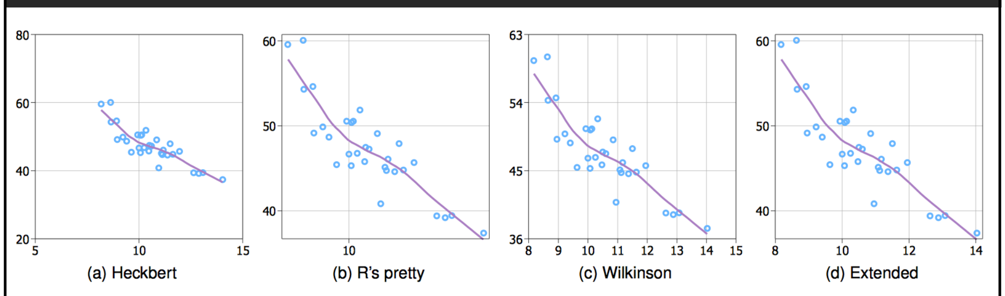
Axis Tick Mark Selection



What are some properties of "good" tick marks?

31

Axis Tick Mark Selection



Simplicity - numbers are multiples of 10, 5, 2

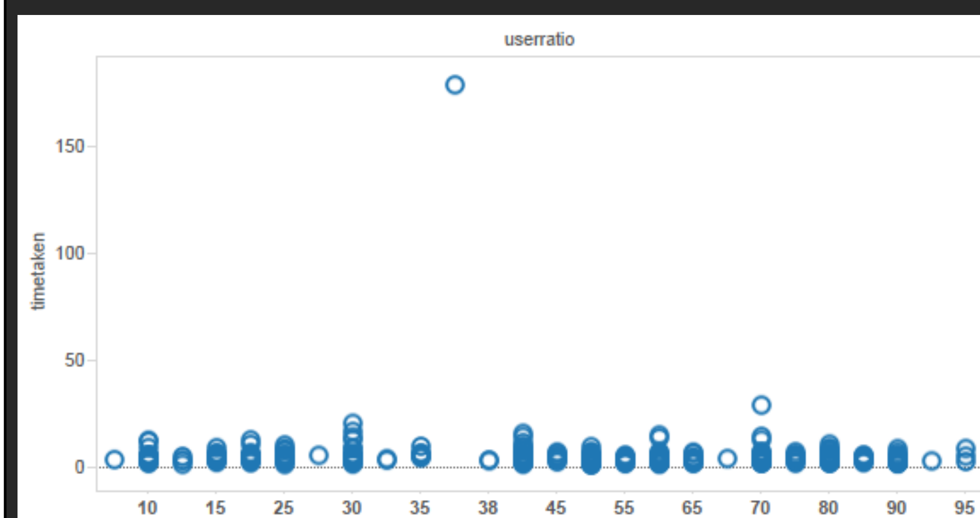
Coverage - ticks near the ends of the data

Density - not too many, nor too few

Legibility - whitespace, horizontal text, size

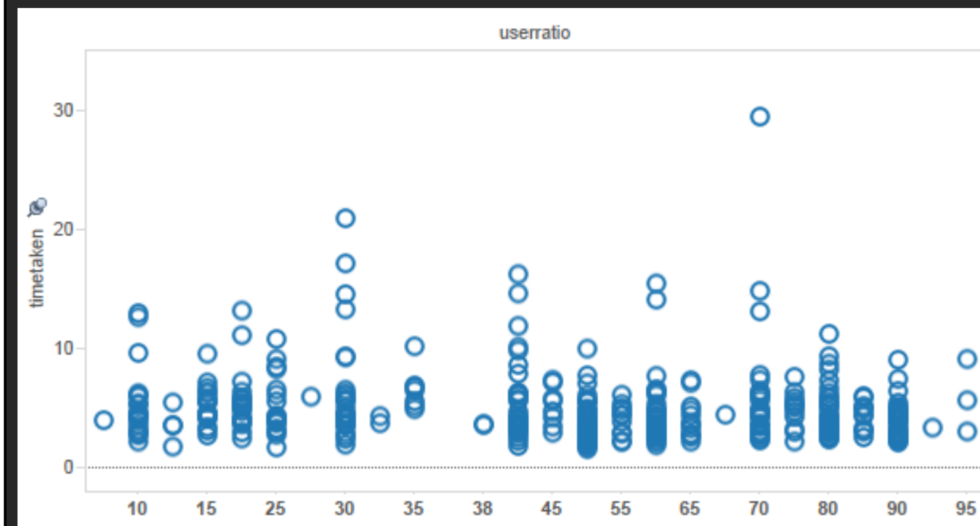
32

How to Scale the Axis?



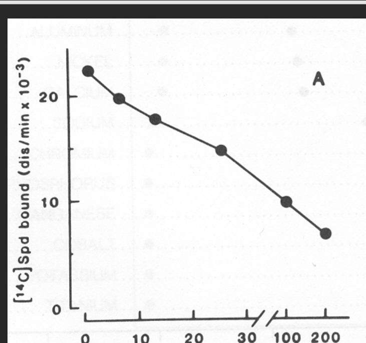
33

One Option: Clip Outliers

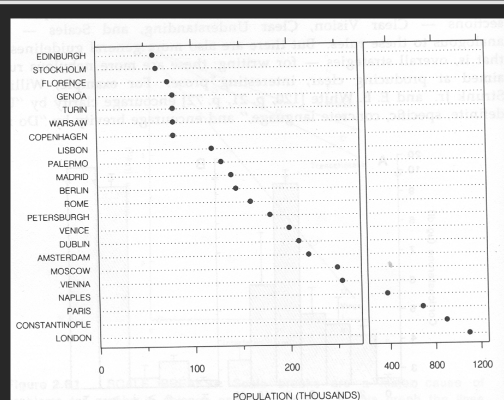


34

Clearly mark scale breaks



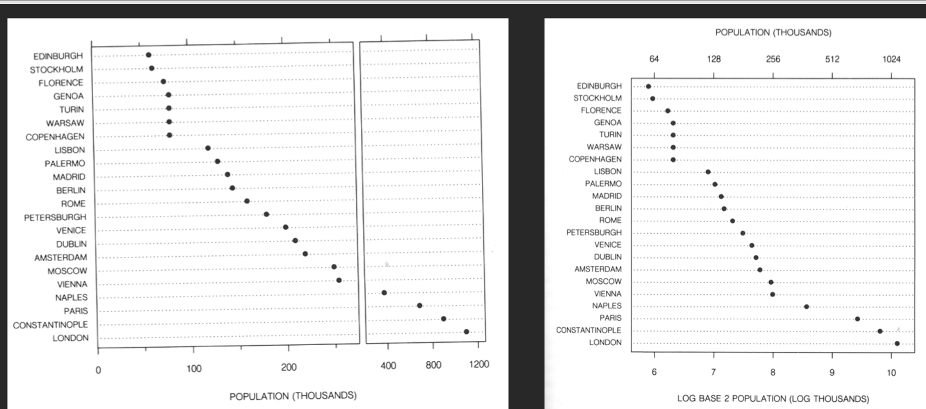
Poor scale break [Cleveland 85]



Well marked scale break [Cleveland 85]

35

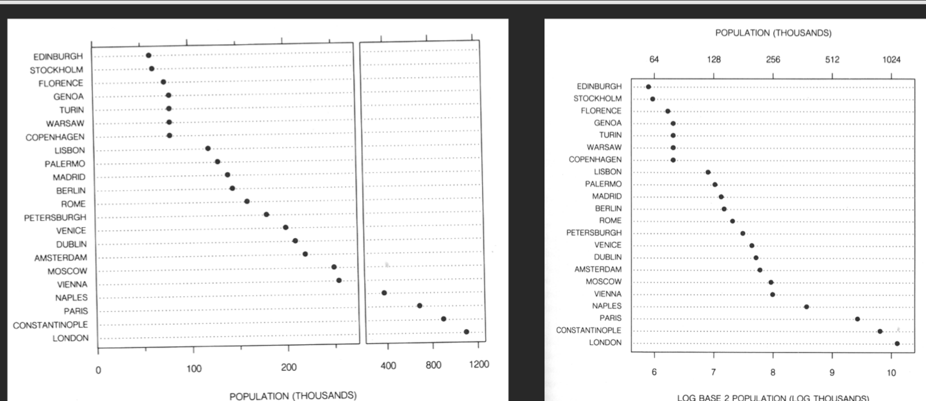
Scale break vs. Log scale



[Cleveland 85]

36

Scale break vs. Log scale



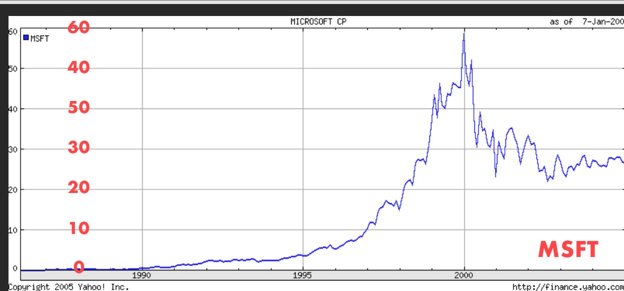
[Cleveland 85]

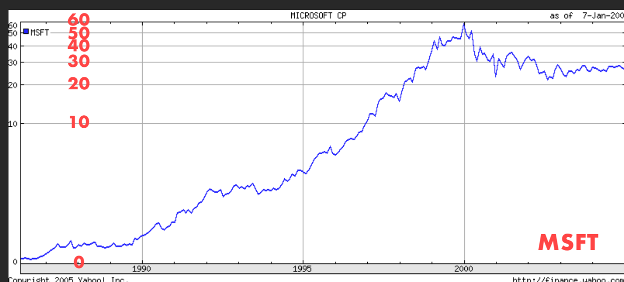
Both increase visual resolution

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

37

Linear scale vs. Log scale

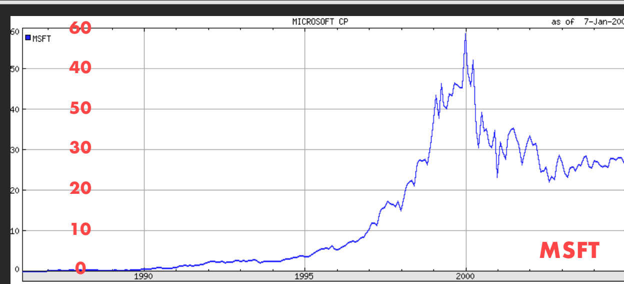




Linear scale vs. Log scale

Linear scale

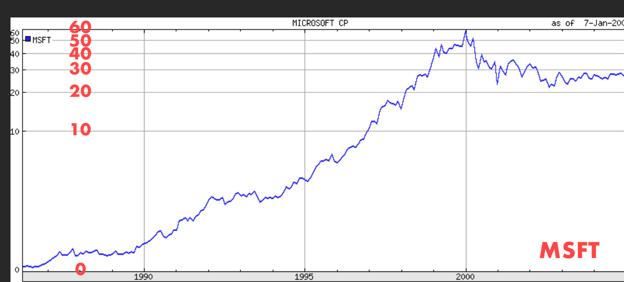
- Absolute change



Log scale

- Small fluctuations
- Percent change

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



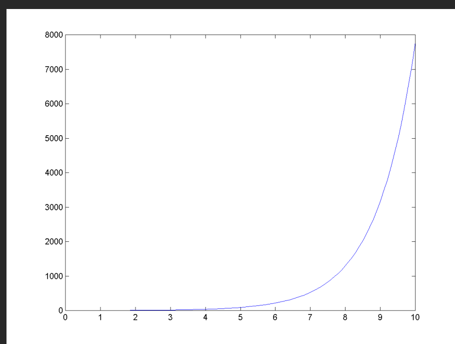
Semilog graph: Exponential growth

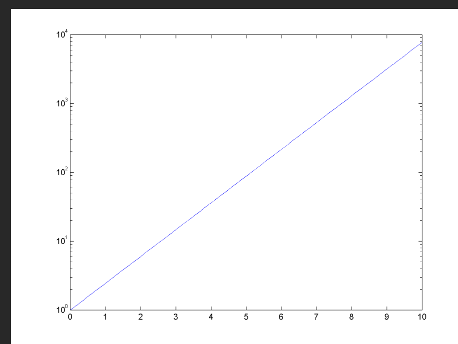
Exponential functions ($y = ka^{mx}$) transform into lines

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

Intercept: $\log(k)$

Slope: $\log(a)m$





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

40

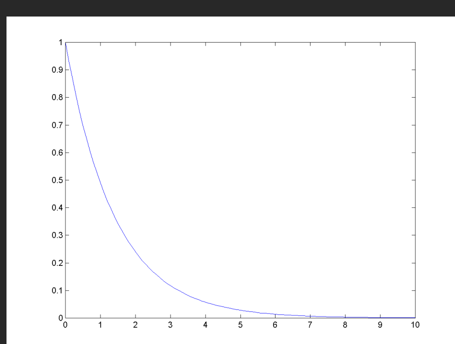
Semilog graph: Exponential decay

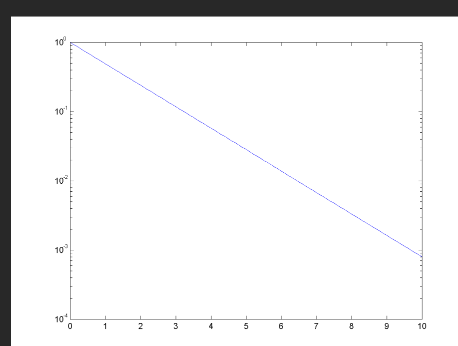
Exponential functions ($y = ka^{mx}$) transform into lines

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

Intercept: $\log(k)$

Slope: $\log(a)m$





$y = 0.5^{2x}$, **slope in semilog space:** $\log(0.5) * 2 = -0.602$

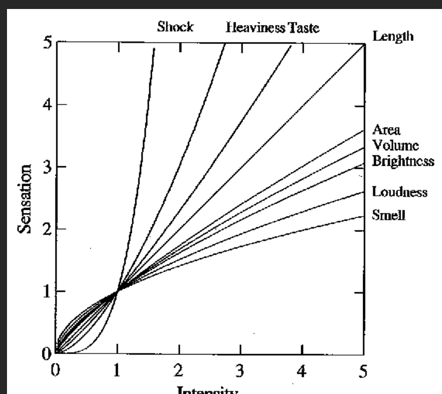
41

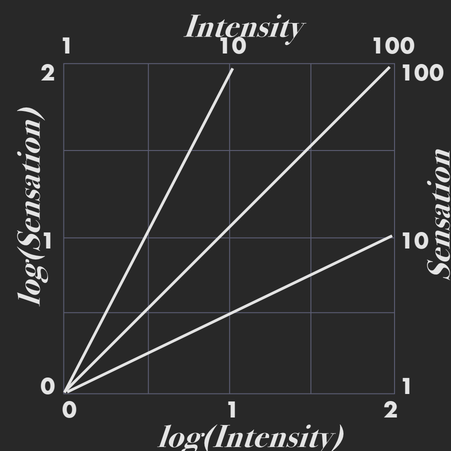
Log-Log graph

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

Example - Steven's power laws:

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





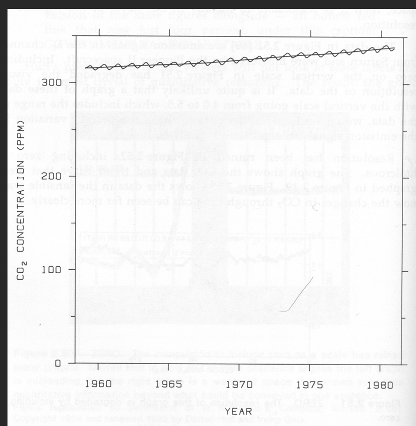
44

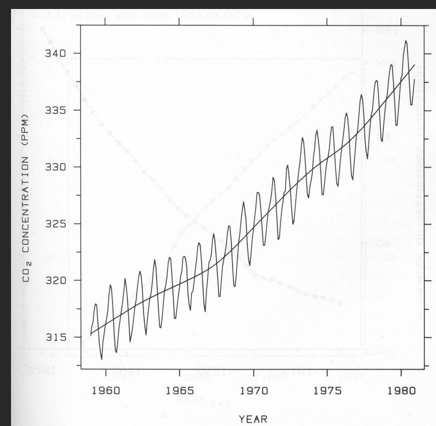
Selecting Aspect Ratio

45

Aspect ratio

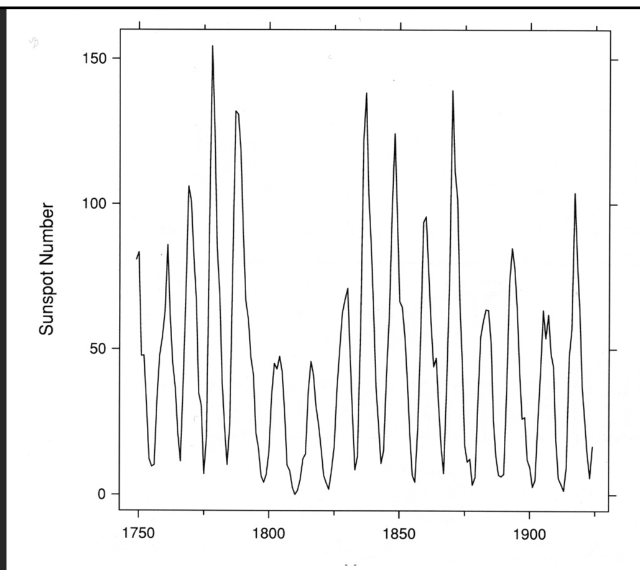
Fill space with data
Don't worry about showing zero





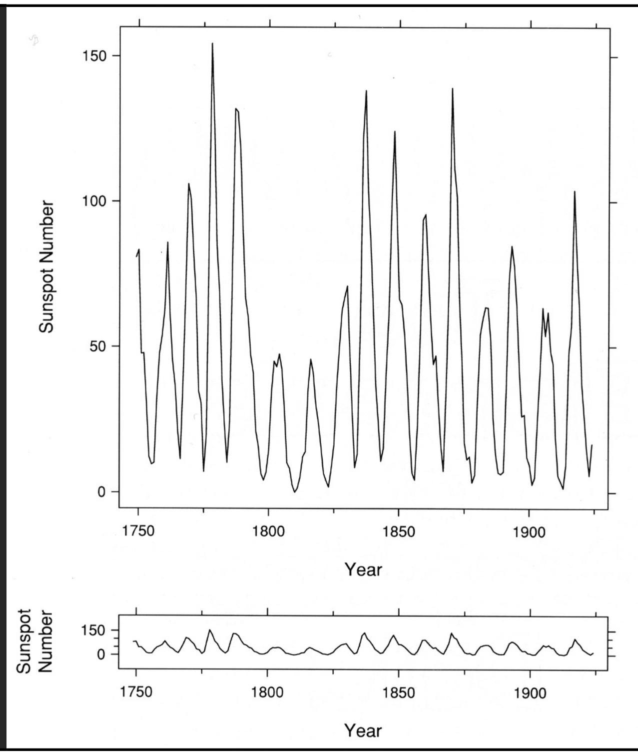
Yearly CO2 concentrations [Cleveland 85]

46



William S. Cleveland
The Elements of Graphing Data

47

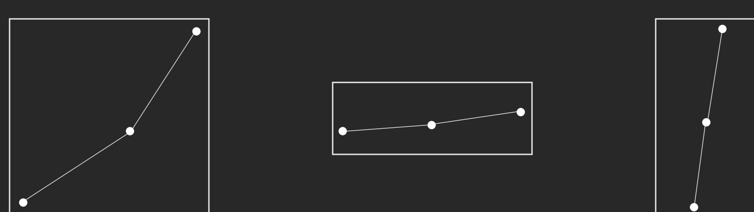


William S. Cleveland
The Elements of Graphing Data

48

Banking to 45° [Cleveland]

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



Two line segments are maximally discriminable when avg. absolute angle between them is 45°

Optimize the *aspect ratio* to bank to 45°

49

Aspect-ratio banking techniques

Median-Absolute-Slope

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

Average-Absolute-Slope

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

Has Closed Form Solution

Average-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)

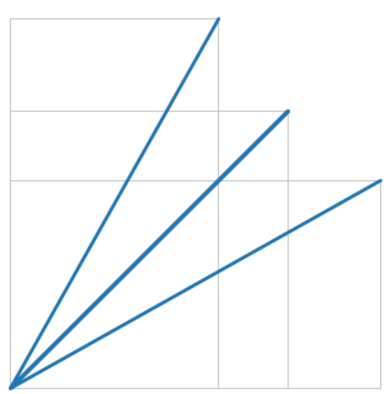
$$\sum_i |\theta_i(\alpha) - \theta_{i+1}(\alpha)|^2$$

Requires Iterative Optimization

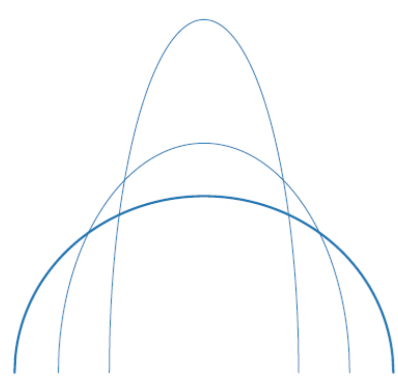
50

An alternate approach:

Minimize arc length (hold area constant)



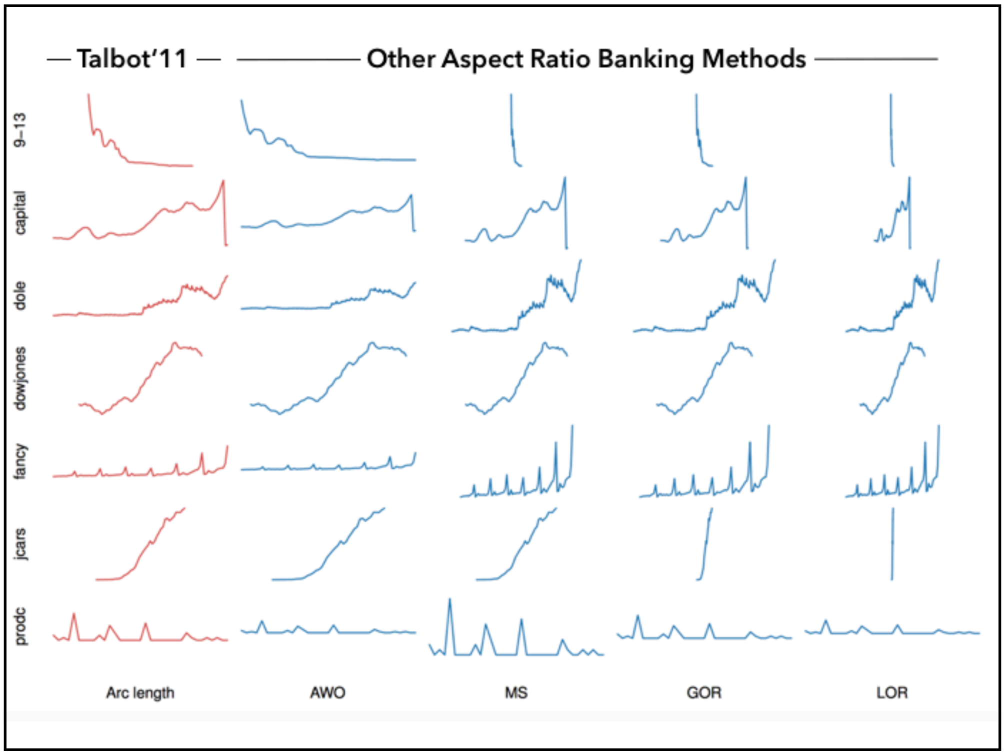
Straight line -> 45 deg



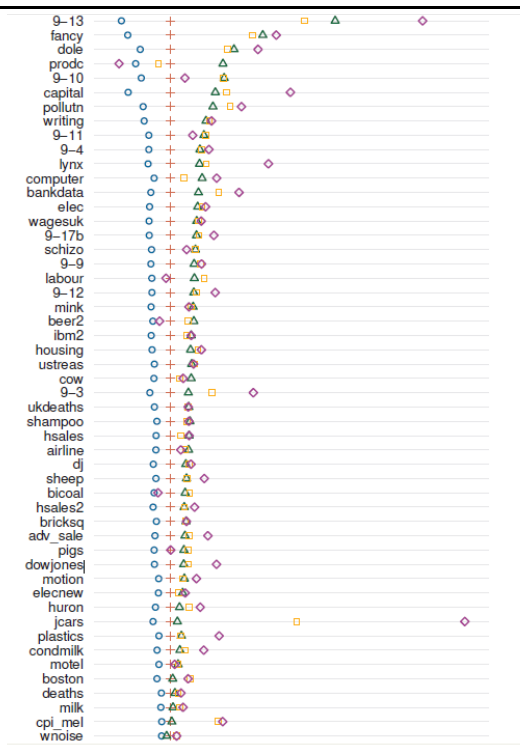
Ellipse -> Circle

[Talbot et al, 2011]

55



56

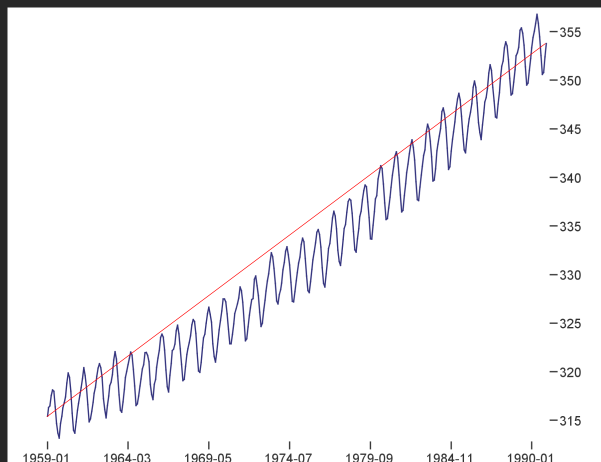


Compromise

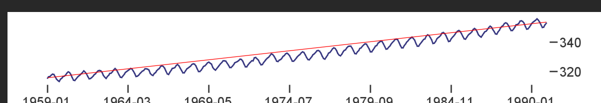
Arc-length banking produces aspect ratios in-between those produced by other methods.

[Talbot et al, 2011]

60



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]

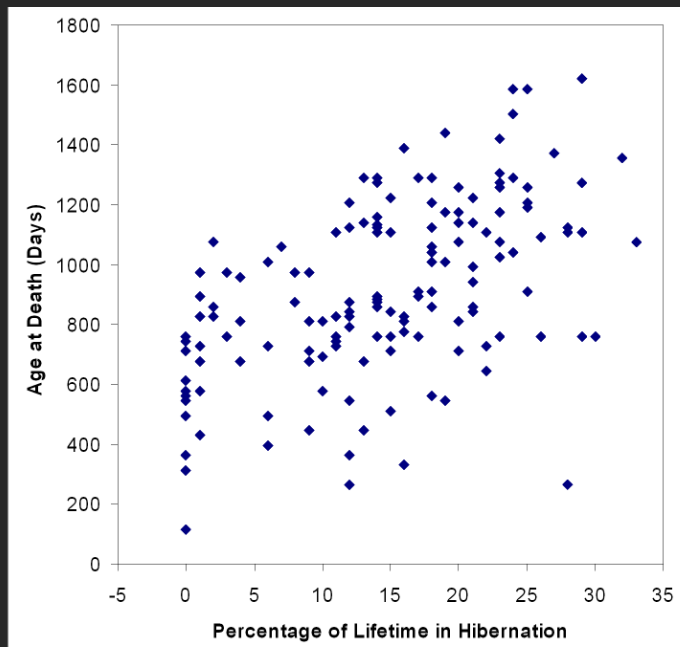
CO_2 Measurements

William S. Cleveland
Visualizing Data

64

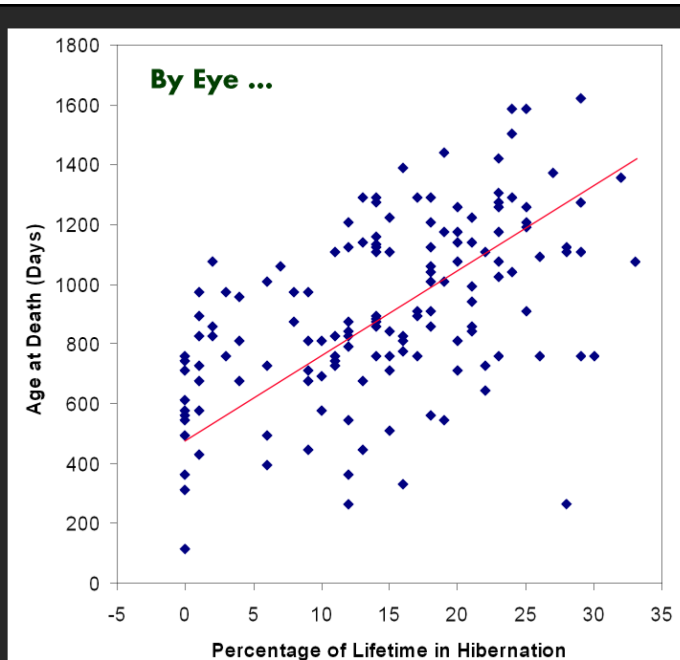
Fitting the Data

76



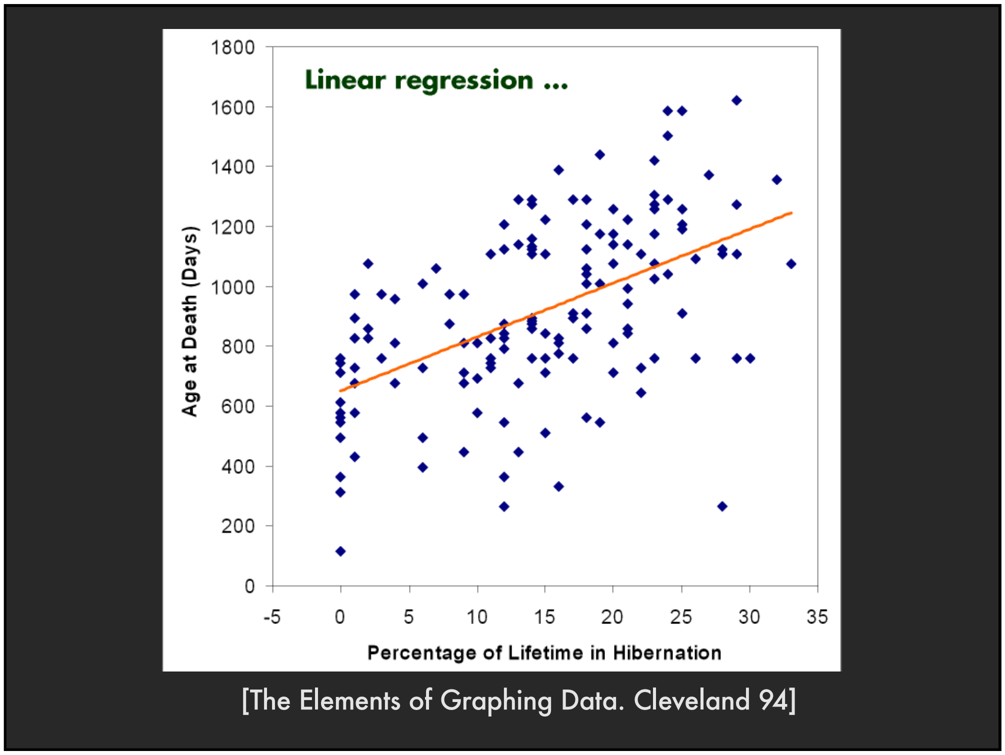
[The Elements of Graphing Data. Cleveland 94]

77

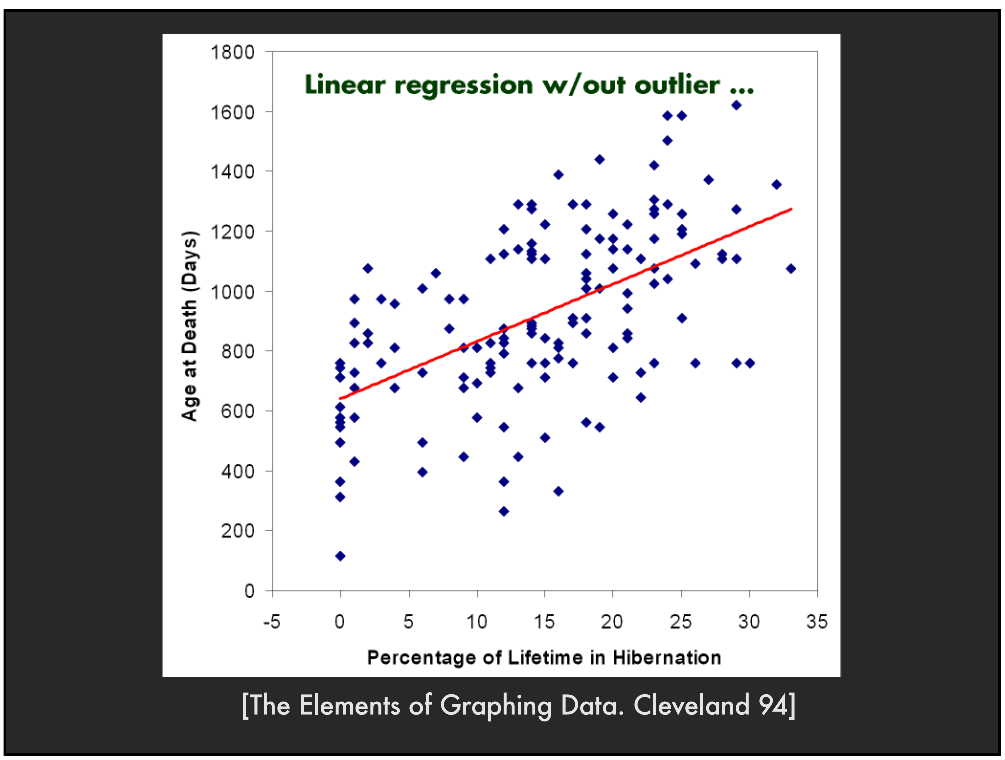


[The Elements of Graphing Data. Cleveland 94]

78


Linear regression ...
1800
1600
1400
1200
1000
800
600
400
200
0
-5
0
5
10
15
20
25
30
35
Age at Death (Days)
Percentage of Lifetime in Hibernation
[The Elements of Graphing Data. Cleveland 94]

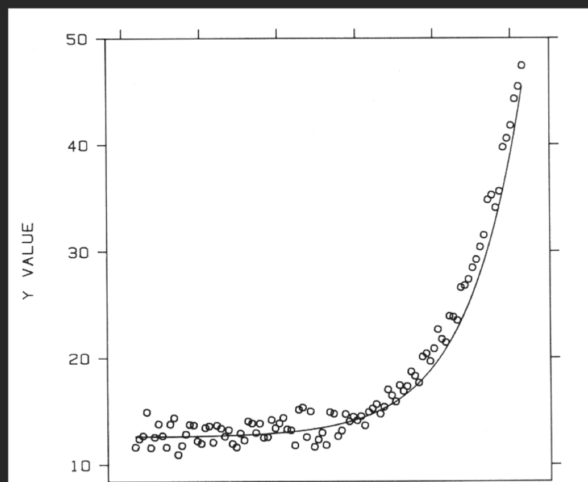
79


Linear regression w/out outlier ...
1800
1600
1400
1200
1000
800
600
400
200
0
-5
0
5
10
15
20
25
30
35
Age at Death (Days)
Percentage of Lifetime in Hibernation
[The Elements of Graphing Data. Cleveland 94]

80

Transforming data

How well does curve fit data?



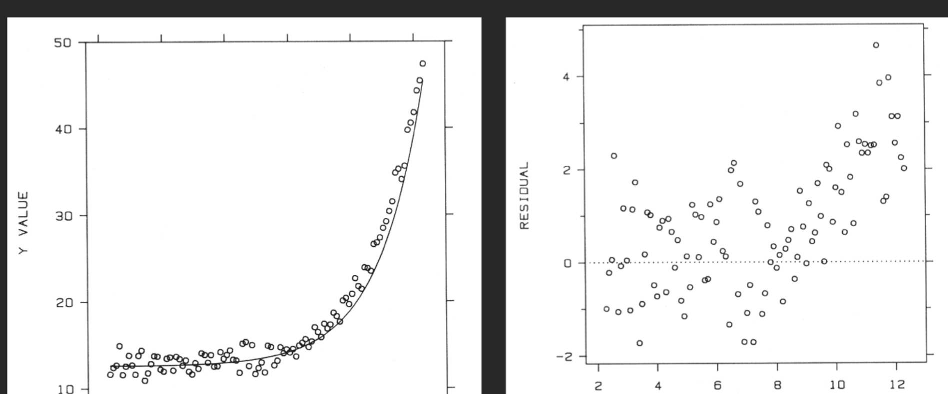
[Cleveland 85]

81

Transforming data

Residual graph

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



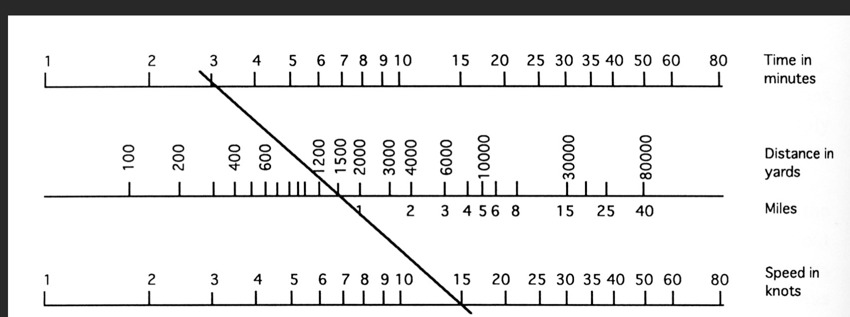
[Cleveland 85]

82

Graphical Calculations

90

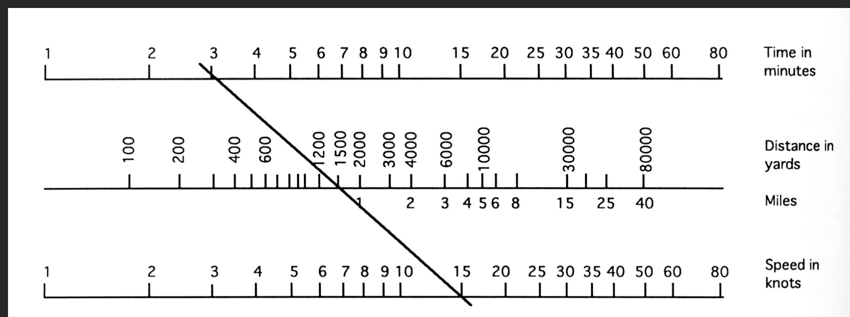
Nomograms



Sailing: The Rule of Three

91

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**

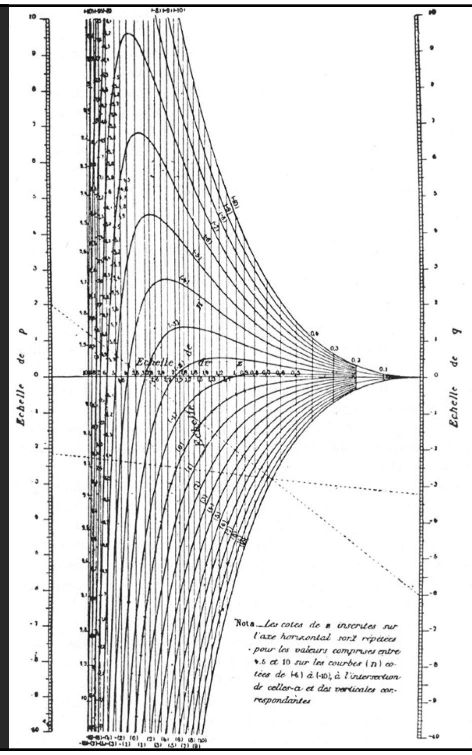
92

Slide rule

Model 1474-66 Electrotechnica 18 Scales

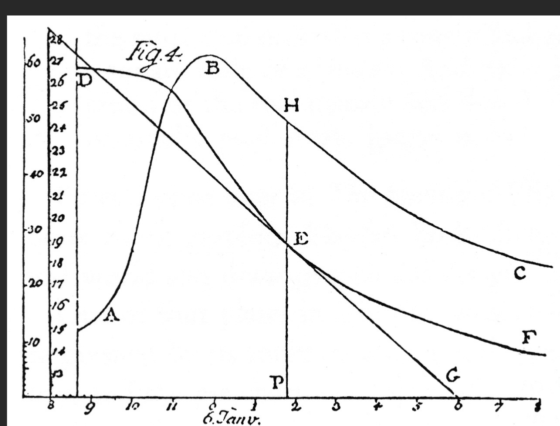
Tehnolemn Timisoara Slide Rule Archive
http://pubpages.unh.edu/~jwc/tehnolemn/

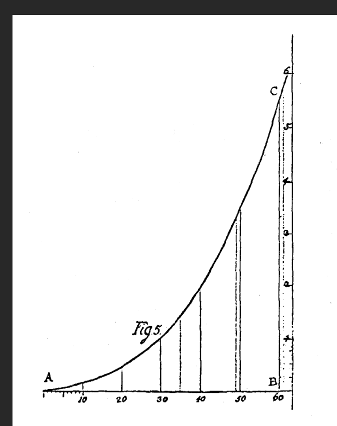
94



95

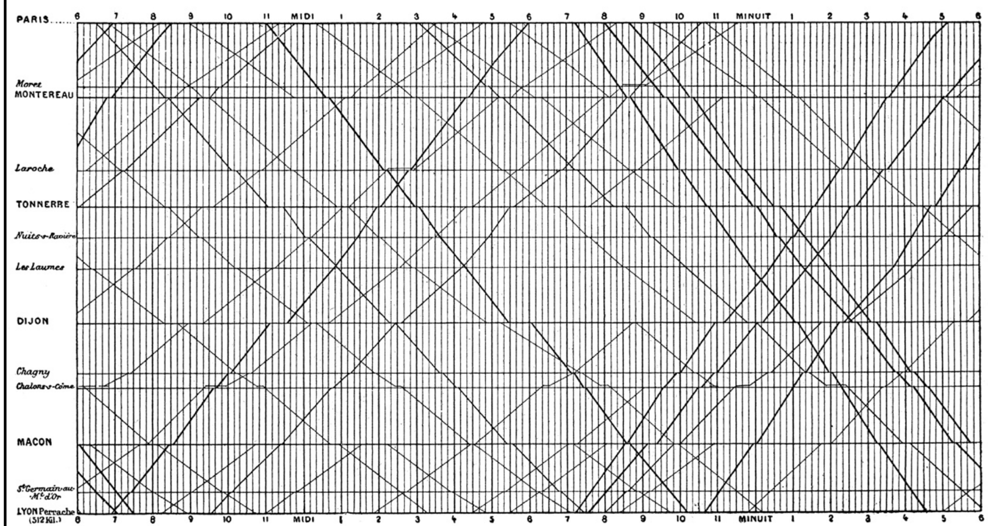
Lambert's graphical construction





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

97



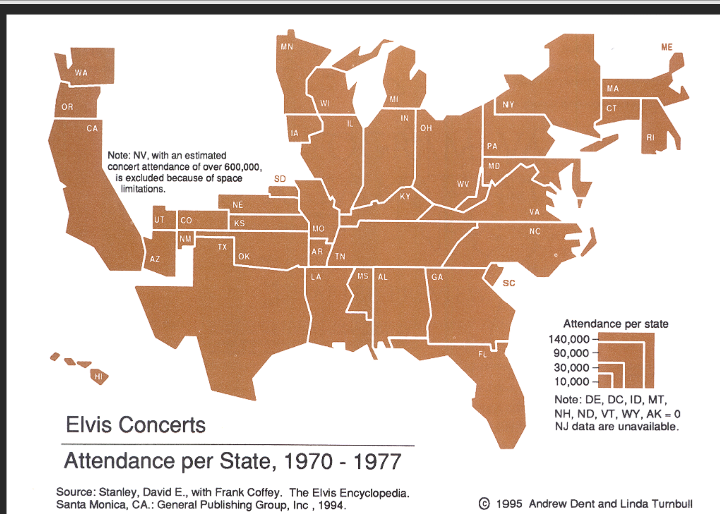
E. J. Marey, *La Méthode Graphique* (Paris, 1885), p. 20. The method is attributed to the French engineer, Ibry.

98

Cartographic Distortion

122

Cartograms: Distort areas

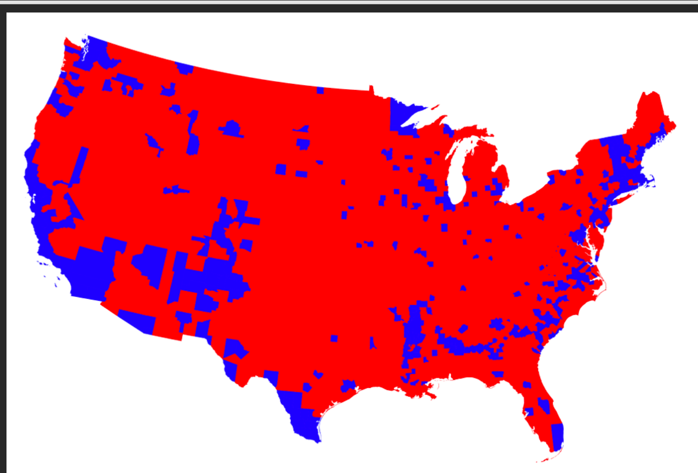


Scale area by data

[From *Cartography*, Dent]

124

Election 2016 map



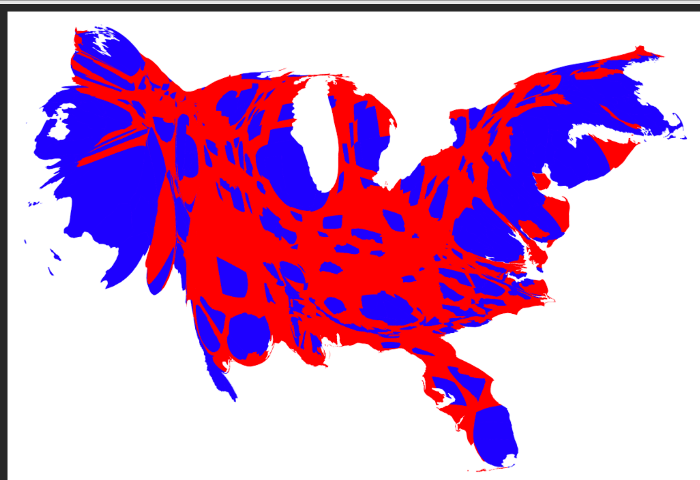
% voted democrat

% voted republican

http://www-personal.umich.edu/~mejn/election/

131

Election 2016 map

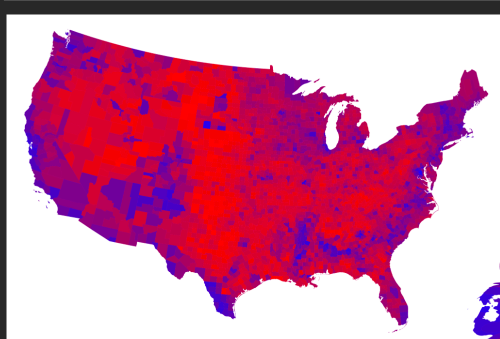


% voted democrat
% voted republican

http://www-personal.umich.edu/~mejn/election/

132

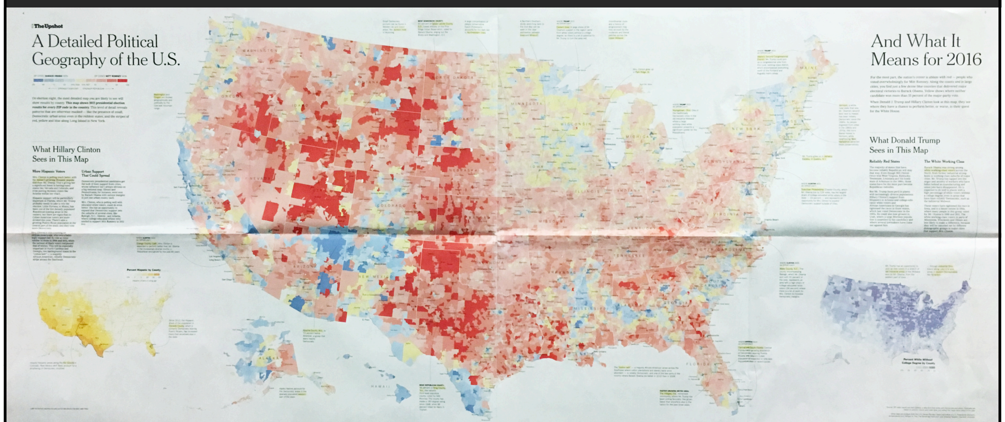
Election 2016 map



http://www-personal.umich.edu/~mejn/election/

133

NYT Election 2016 (based on 2012)



Statistical map with shading

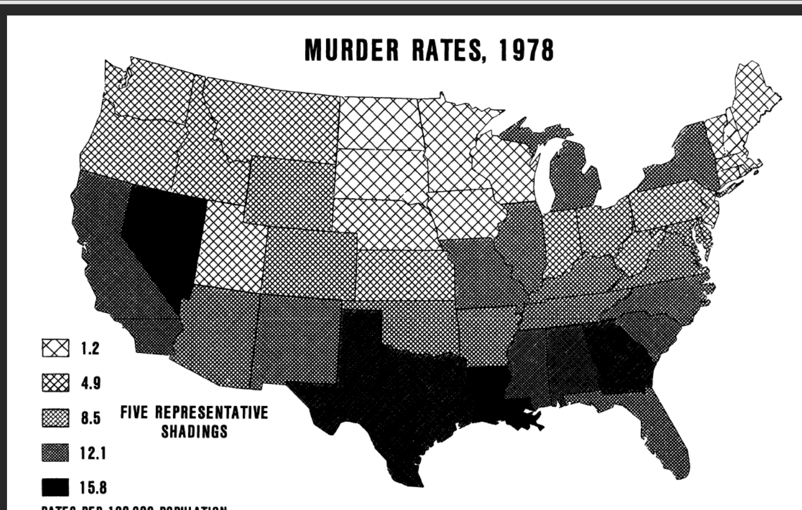
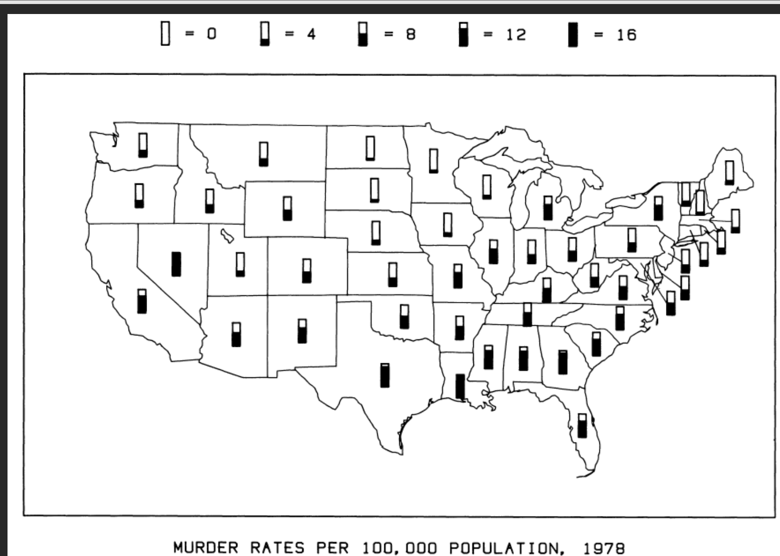


Figure 5. Statistical map with shading.

[Cleveland and McGill 84]

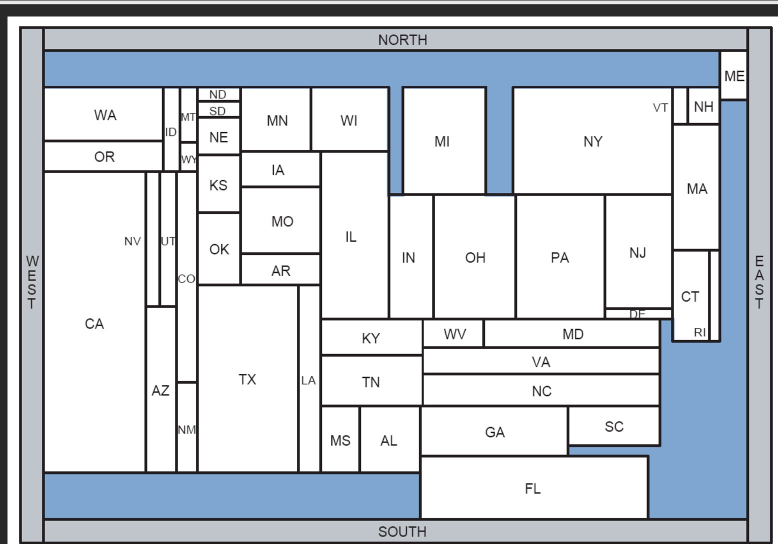
Framed rectangle chart



[Cleveland and McGill 84]

136

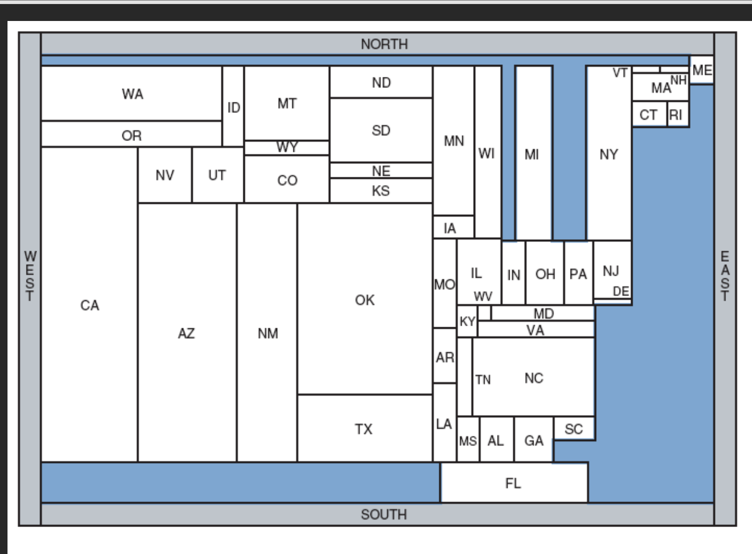
Rectangular cartogram



American population [van Kreveld and Speckmann 04]

137

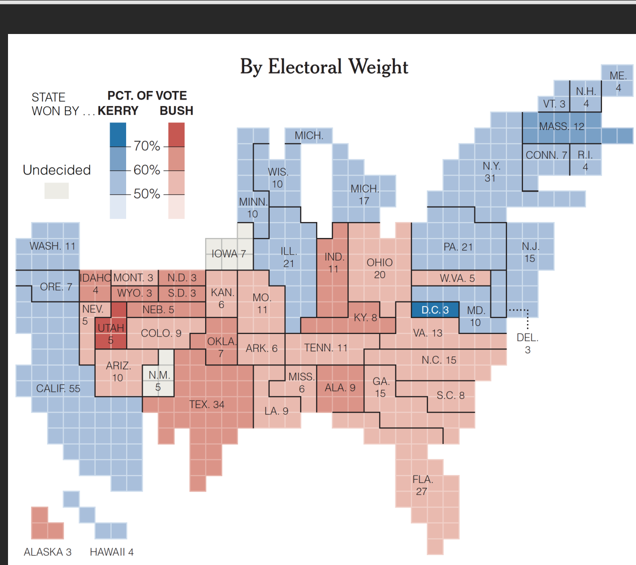
Rectangular cartogram



Native American population [van Kreveld and Speckmann 04]

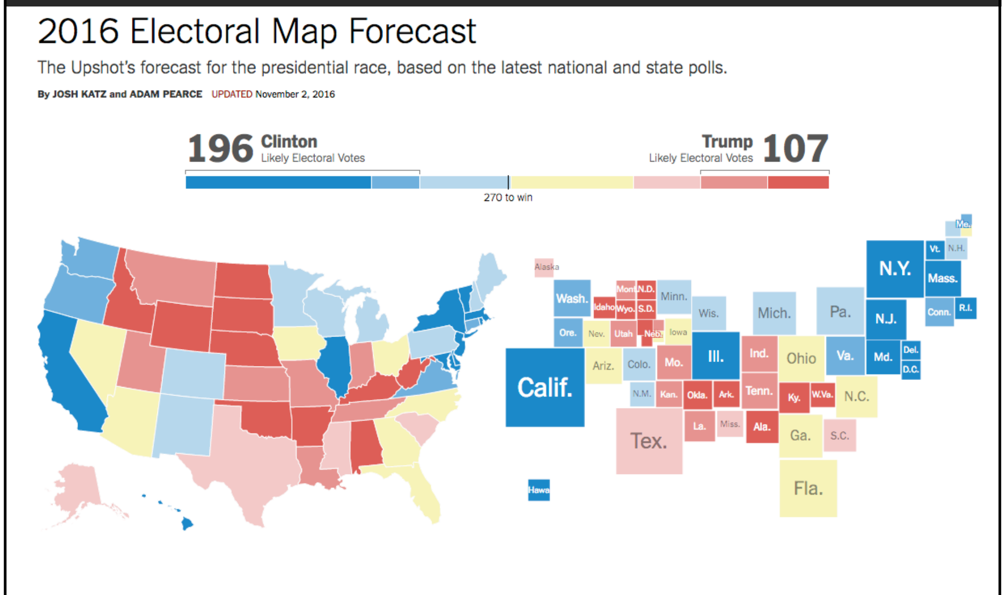
138

New York Times Election 2004



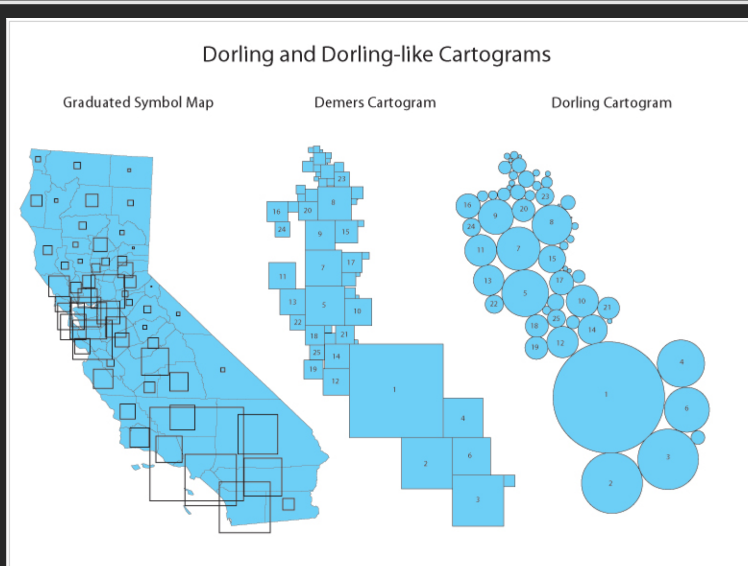
139

New York Times Election 2016



140

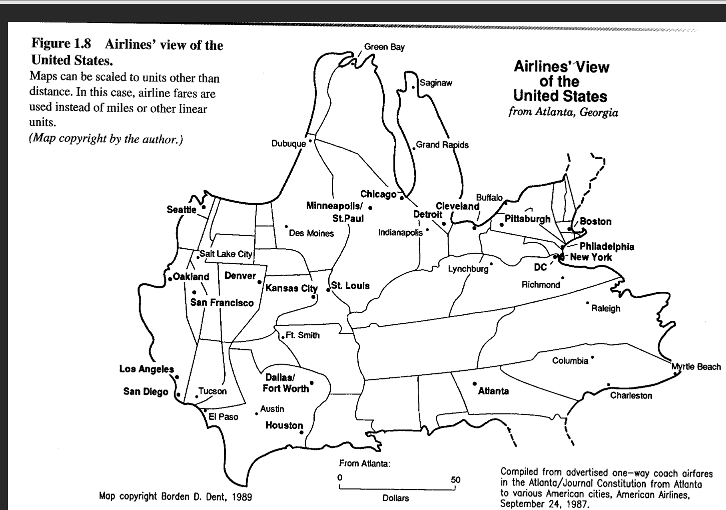
Dorling cartogram



http://www.ncgia.ucsb.edu/projects/Cartogram_Central/types.html

141

Distorting distances



Scale distance by data (airline fare)

[From *Cartography*, Dent]

142

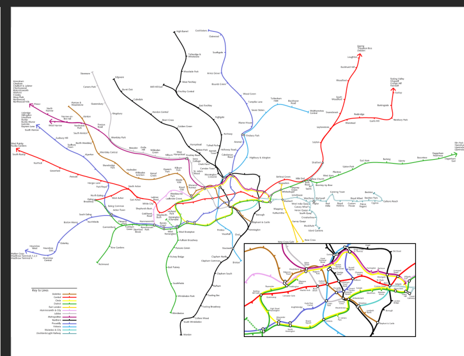
London underground

http://www.thetube.com/content/history/map.asp

144

Comparison to geographic map

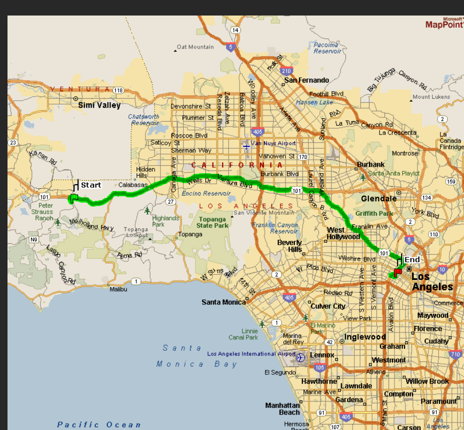
Distorted

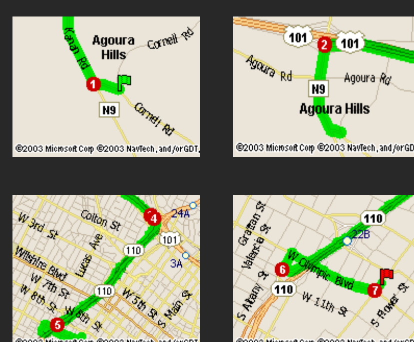


Undistorted

145

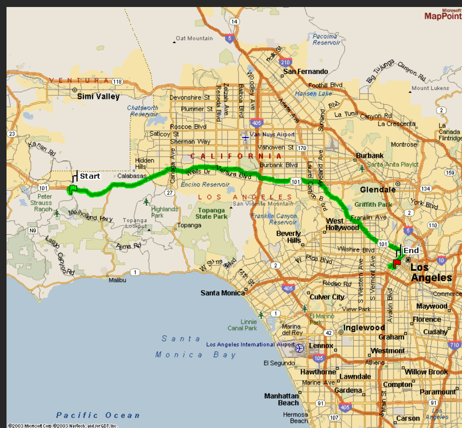
Visualizing Routes

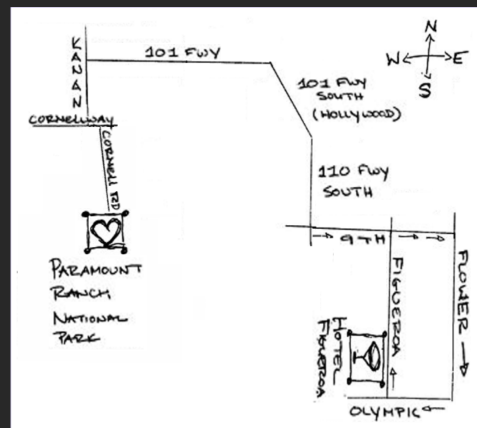




146

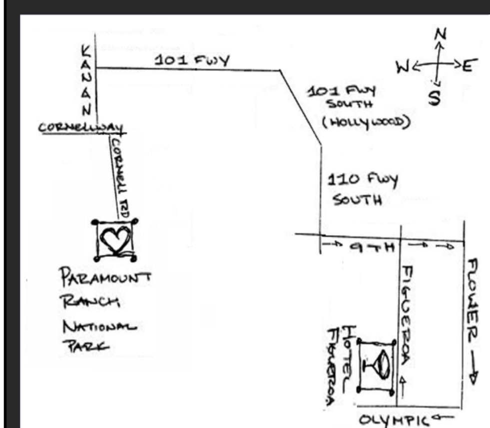
A Better Visualization



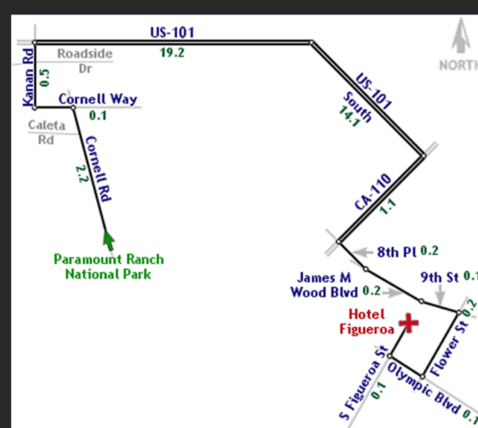


147

LineDrive [Agrawala & Stolte 2001]



Hand-drawn route map



LineDrive route map

148

Summary

- Space is the most important visual encoding
- Geometric properties of spatial transforms support geometric reasoning
- Show data with as much resolution as possible
- Use distortions to emphasize important information

149