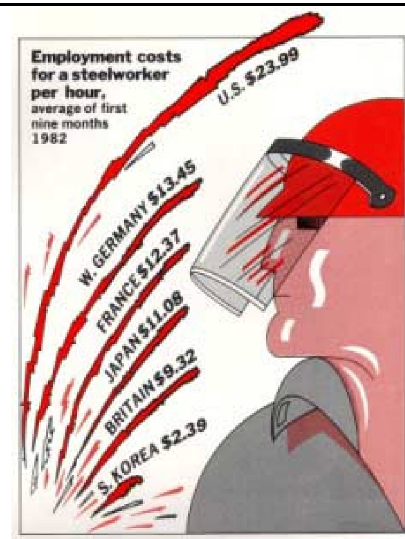
CHART DESIGN

CS 448B | Fall 2025

MANEESH AGRAWALA

1



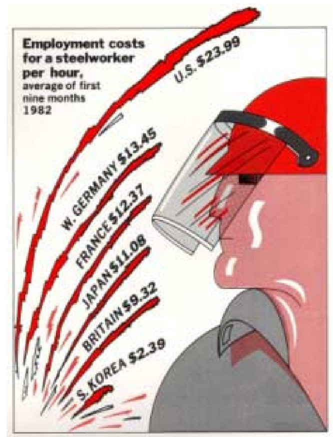
2

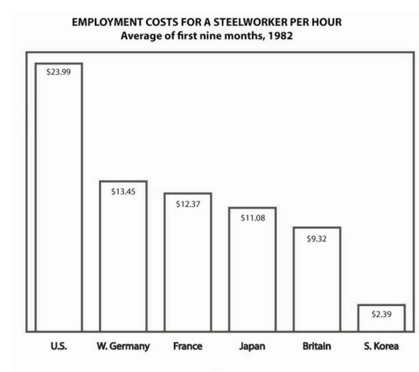
READING RESPONSE: QUESTIONS/THOUGHTS

- Tufte is ***fairly rigid*** in his ***stance against "decoration",*** but my instinct is that decoration, when not brash or overly distracting, might be able to aid engagement and interest, which I thought could then increase rates of recall and memory. ***… if all we use is "data ink", are we limiting how engaging and easy to remember our data visualizations are?***

3

CHARTJUNK: IS IT USEFUL? from [Bateman 2010]





4

READING RESPONSE: QUESTIONS/THOUGHTS

- Additionally, I found the discussion on the ***Polaris system*** particularly compelling, as it ***integrates interactivity*** into the visualization process. While traditional design principles focus on static visuals, Polaris allows users to manipulate data dynamically. I'm curious about how the process of redesign changes when users themselves control the interaction—***how much flexibility should be given to avoid overwhelming users, and does this shift the responsibility of good design from the creator to the end user?***

5

LAST TIME:
EXPLORATORY DATA ANALYSIS

6

LESSON: EDA IS AN ITERATIVE PROCESS

1. **Construct graphics to address questions**
2. **Inspect "answer" and assess new questions**
3. **Repeat!**

Transform the data appropriately (e.g., invert, log)

"Show data variation, not design variation" -Tufte

7

TABLEAU DEMO

Dataset:

Federal Elections Commission Receipts
Every Congressional Candidate from 1996 to 2002
4 Election Cycles
9216 Candidacies

9

DATA TYPES

Year (Qi)
Candidate Code (N)
Candidate Name (N)
Incumbent / Challenger / Open-Seat (N)
Party Code (N) [1=Dem,2=Rep,3=Other]
Party Name (N)
Total Receipts (Qr)
State (N)
District (N)

This is a subset of the larger data set available from the FEC, but should be sufficient for the demo

10

HYPOTHESES

What might we learn from this data?

- Have receipts increased over time?
- Do democrats or republicans spend more?
- Candidates from which state spend the most money?

11

POLARIS/TABLEAU APPROACH

Insight: simultaneously specify both database queries & visualization

Choose data, then visualization, not vice versa

Use **smart defaults** for visual encodings (Like APT)

Can also suggest more encodings upon request (ShowMe)

12

ANNOUNCEMENTS

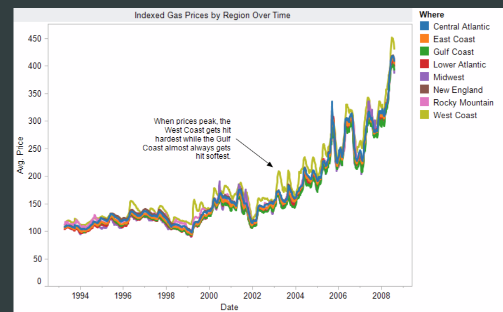
13

ASSIGNMENT 2: EXP. DATA ANALYSIS

Due 10/13 10:30am

Use Tableau or Vega-Lite to formulate & answer data questions

- First steps
  - Step 1: Pick domain & data
  - Step 2: Pose questions
  - Step 3: Profile data
  - Iterate as needed
- Create visualizations
  - See different views of data
  - Refine questions
- Author a report
  - Screenshots of most insightful views (8+)
  - Include titles and captions for each view



14

TODAY

Learning Objectives

1. How to choose good visual encodings from the large set of possibilities.
2. How scales, axes, aspect ratios, fitting and sorting can emphasize different aspects of the data.

15

DESIGN SPACE OF VISUAL ENCODINGS

16

MAPPING DATA TO VISUAL CHANNELS

Assign **data fields** (e.g., with N, O, Q types) to **visual channels** *(x, y, color, shape, size, …)* for a chosen **graphical mark** type (point, bar, line, …)

Additional concerns include choosing appropriate **encoding parameters** *(log scale, sorting, …)* and **data transformations** *(bin, group, aggregate, …)*

These options define a large combinatorial space, containing both useful and questionable charts!

17

EXPRESSIVENESS CRITERIA [Mackinlay 1986]

Expressiveness

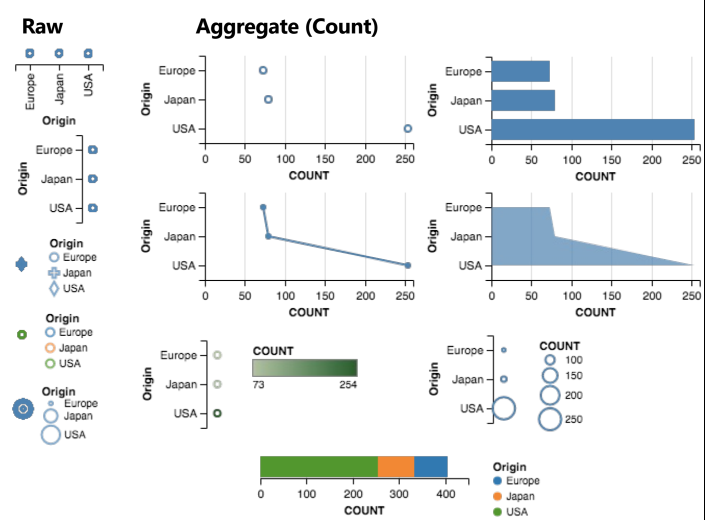
A set of facts is expressible in a visual language if the sentences (i.e., the visualizations) in the language express ***all*** the facts in the set of data, and ***only*** the facts in the data.

18

1D NOMINAL

Cars Data

Price	MPG	Origin	Make
13,500	22	Japan	Honda
7,200	31	Europe	BMW
11,300	12	USA	Ford
...	...	...	...



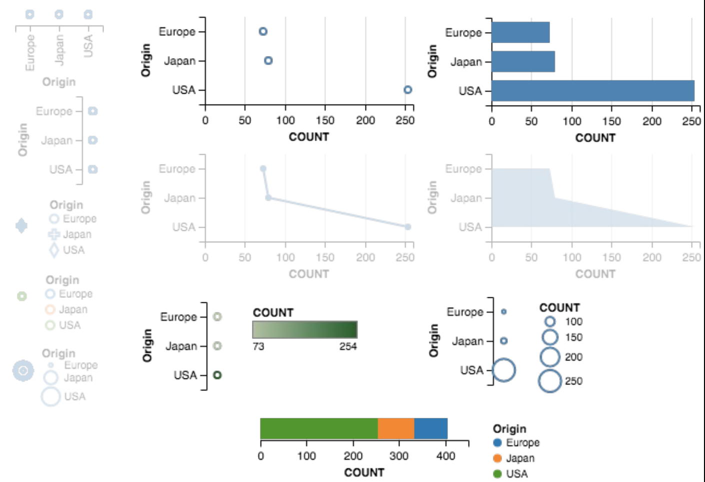
19

EXPRESSIVE?

Cars Data

Price	MPG	Origin	Make
13,500	22	Japan	Honda
7,200	31	Europe	BMW
11,300	12	USA	Ford
...	...	...	...

Raw **Aggregate (Count)**



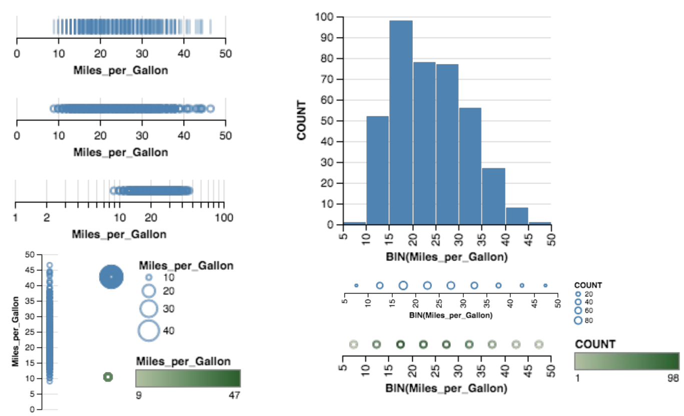
20

1D QUANTITATIVE

Cars Data

Price	MPG	Origin	Make
13,500	22	Japan	Honda
7,200	31	Europe	BMW
11,300	12	USA	Ford
...	...	...	...

Raw **Aggregate (Count)**



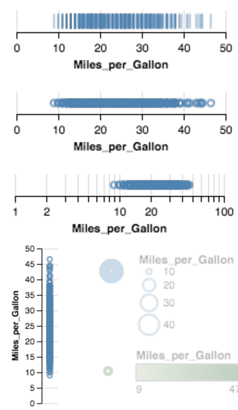
21

EXPRESSIVE?

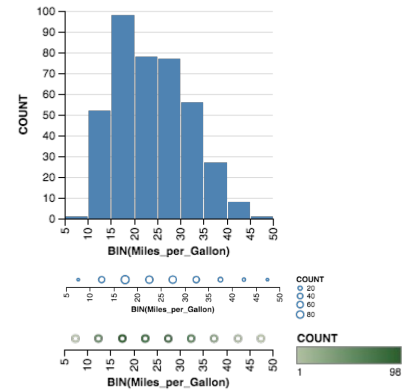
Cars Data

Price	MPG	Origin	Make
13,500	22	Japan	Honda
7,200	31	Europe	BMW
11,300	12	USA	Ford
...	...	...	...

Raw

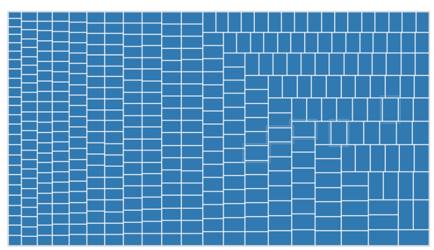


Aggregate (Count)

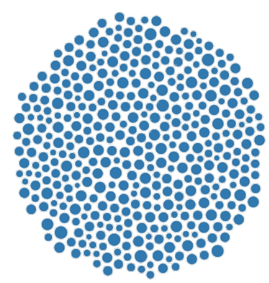


22

Raw (with Layout Algorithm)

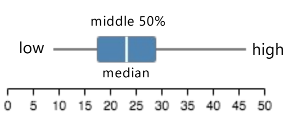


Treemap

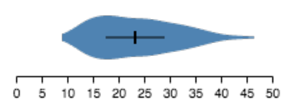


Bubble Chart

Aggregate (Distributions)



Box Plot



Violin Plot

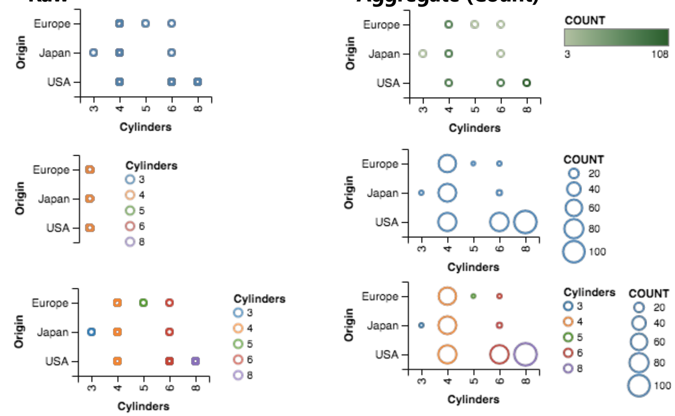
23

2D NOMINAL x NOMINAL

Cars Data

Price	MPG	Origin	Make
13,500	22	Japan	Honda
7,200	31	Europe	BMW
11,300	12	USA	Ford
...	...	...	...

Raw **Aggregate (Count)**



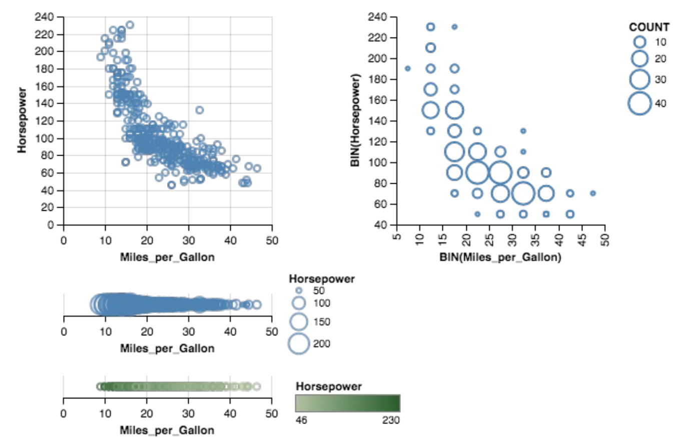
24

2D QUANTITATIVE x QUANTITATIVE

Cars Data

Price	MPG	Origin	Make
13,500	22	Japan	Honda
7,200	31	Europe	BMW
11,300	12	USA	Ford
...	...	...	...

Raw **Aggregate (Count)**



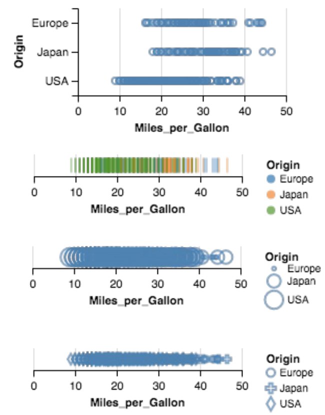
25

2D NOMINAL x QUANTITATIVE

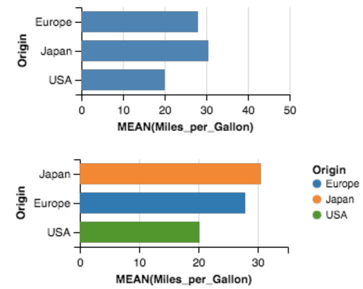
Cars Data

Price	MPG	Origin	Make
13,500	22	Japan	Honda
7,200	31	Europe	BMW
11,300	12	USA	Ford
...	...	...	...

Raw

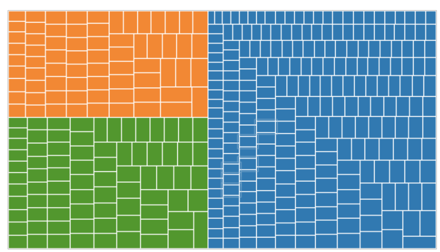


Aggregate (Mean)

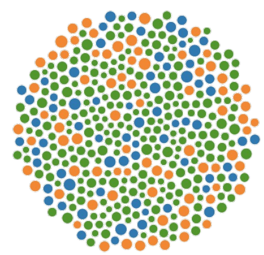


26

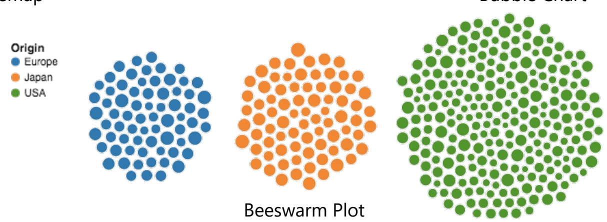
Raw (with Layout Algorithm)



Treemap



Bubble Chart



Beeswarm Plot

27

EFFECTIVENESS CRITERIA [Mackinlay 1986]

Effectiveness

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

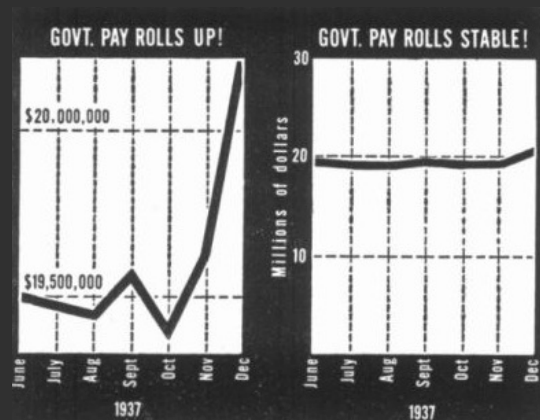
Subject of the Perception Lecture

28

SCALES AND AXES

34

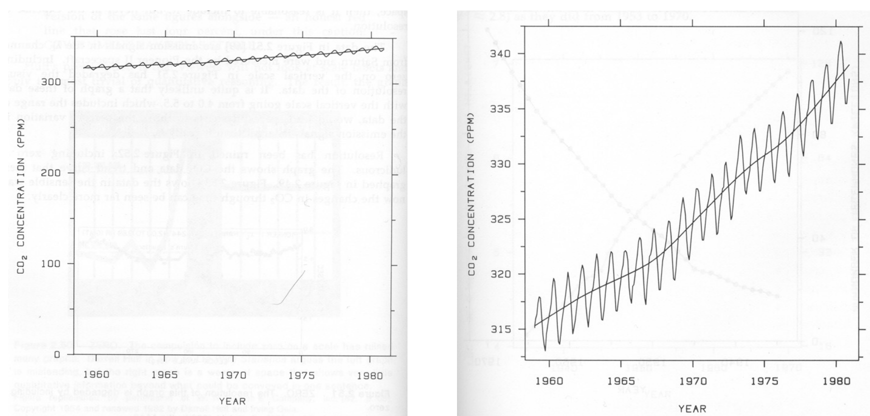
WHICH GRAPH IS BETTER



Government payrolls in 1937 [Huff 93]

35

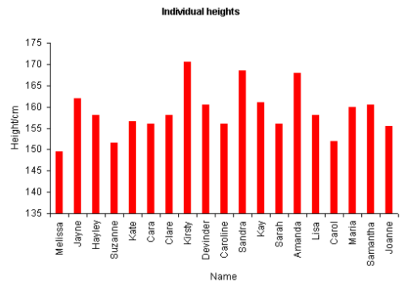
INCLUDE ZERO IN AXIS SCALE?



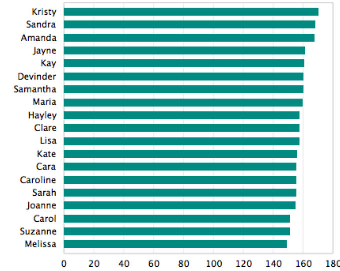
Yearly CO2 concentrations [Cleveland 85]

36

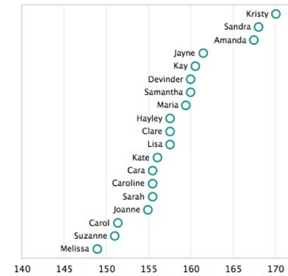
INCLUDE ZERO IN AXIS SCALE?



Compare Proportions (Q-Ratio)

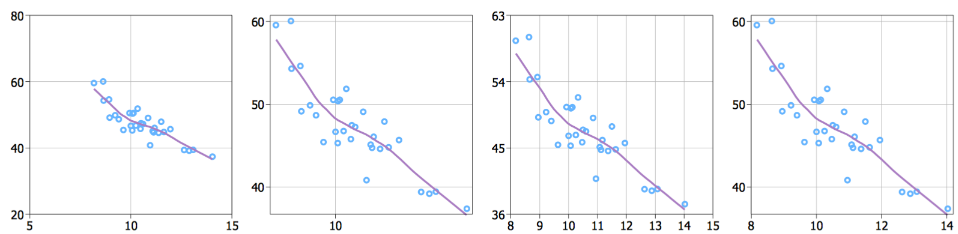


Compare Relative Position (Q-Interval)



37

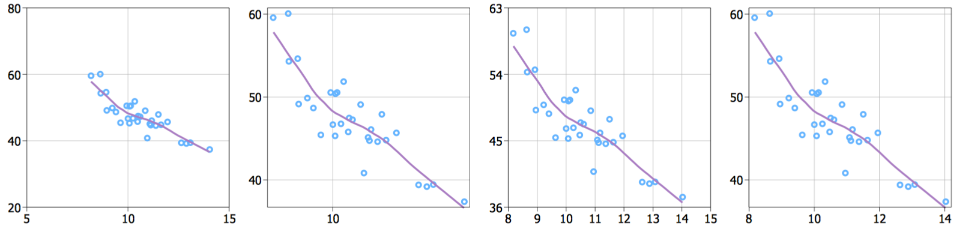
AXIS TICK MARK SELECTION



What are some properties of “good” tick marks?

38

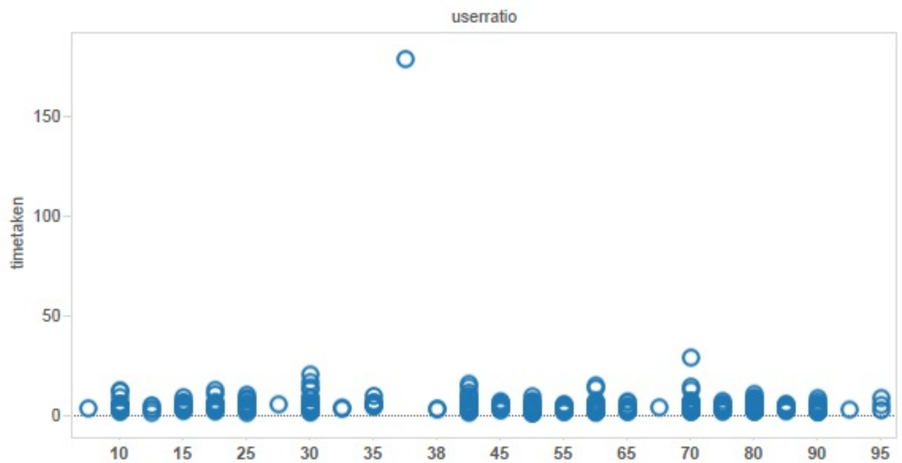
AXIS TICK MARK SELECTION



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

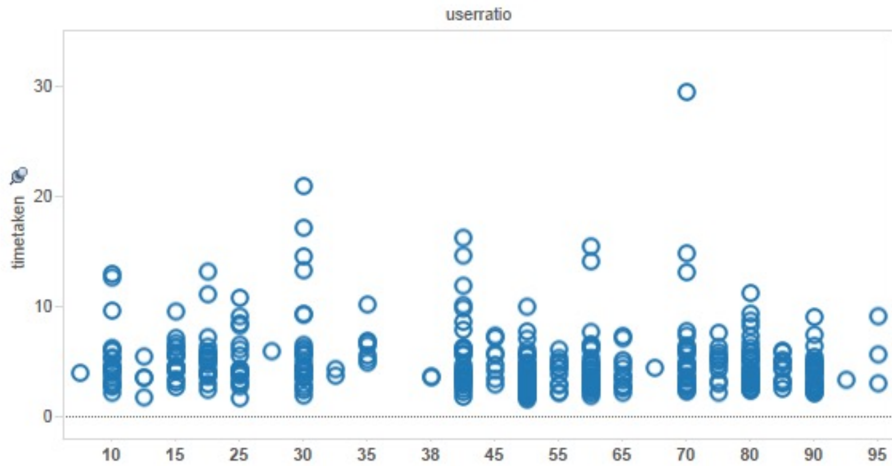
39

HOW TO SCALE THE AXIS?



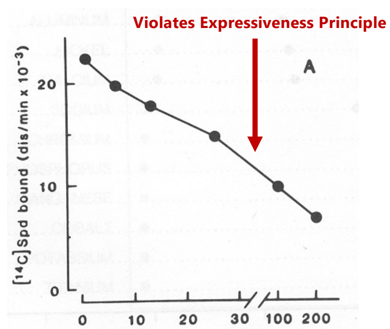
40

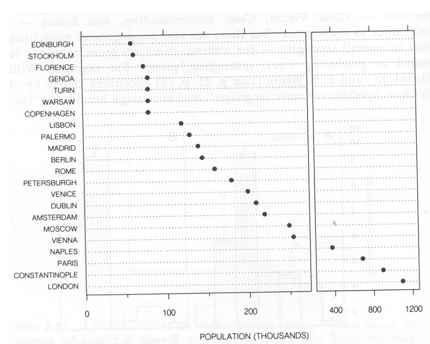
OPTION 1 : CLIP OUTLIERS



41

OPTION 2: SCALE BREAKS – CLEARLY MARKED

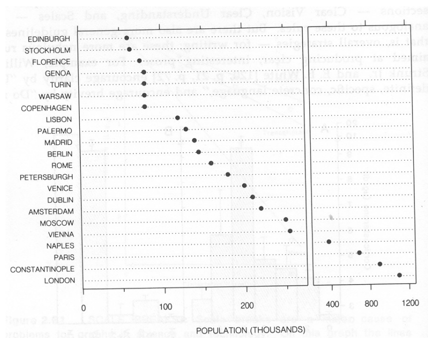


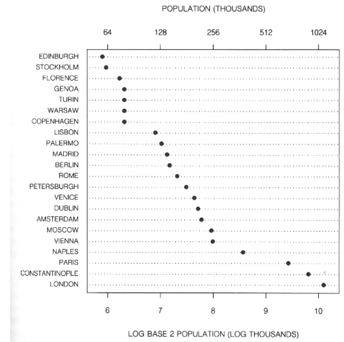


Poor scale break [Cleveland 85] **Well marked scale break** [Cleveland 85]

42

OPTION 3: LOG SCALE

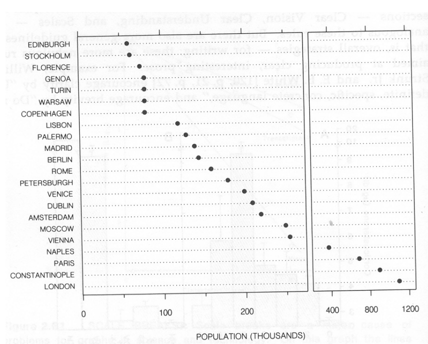


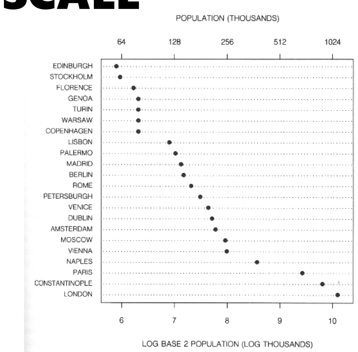


Scale break [Cleveland 85] **Log scale**[Cleveland 85]

43

SCALE BREAK VS. LOG SCALE





Both increase visual resolution

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

Log scale - direct comparisons of all data

44

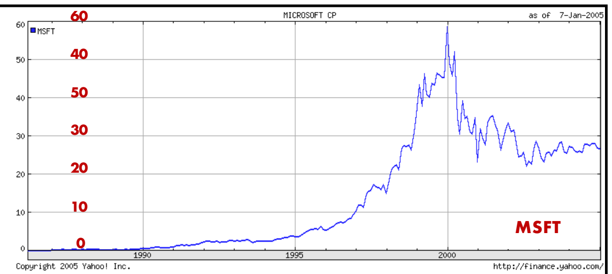
Logarithms turn ***multiplications*** into ***additions***

$$\log(xy) = \log(x) + \log(y)$$

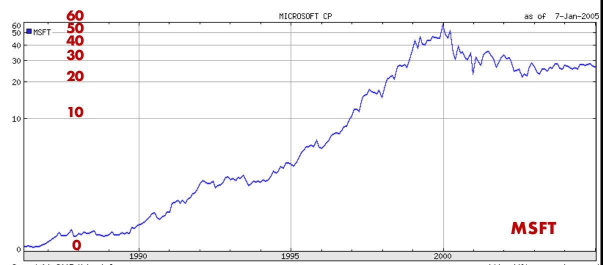
Equal steps on a log scale correspond to equal changes to a multiplicative scale factor

45

Linear Scale

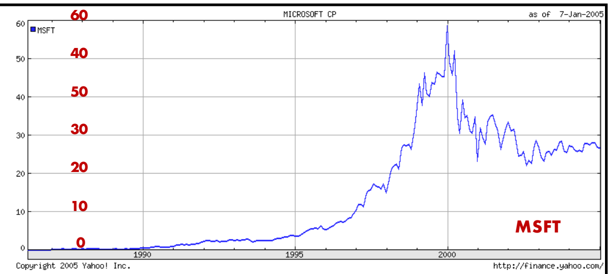


Log Scale

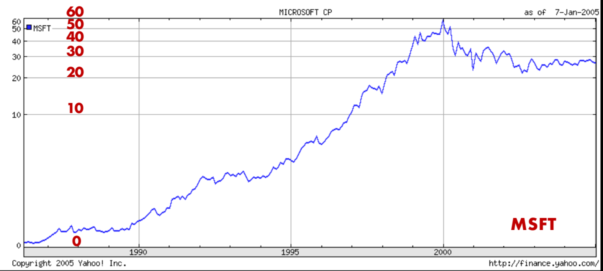


46

Linear Scale
- Absolute change



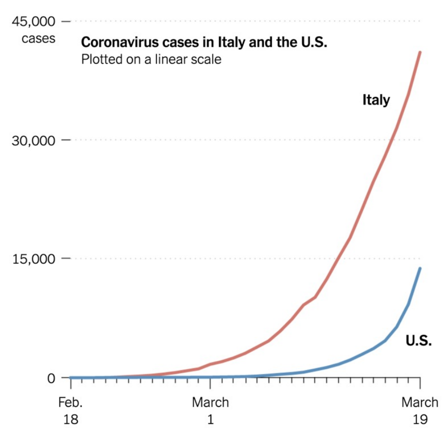
Log Scale
- Small fluctuations
- Percent change
  d(10,30) > d(30, 60)

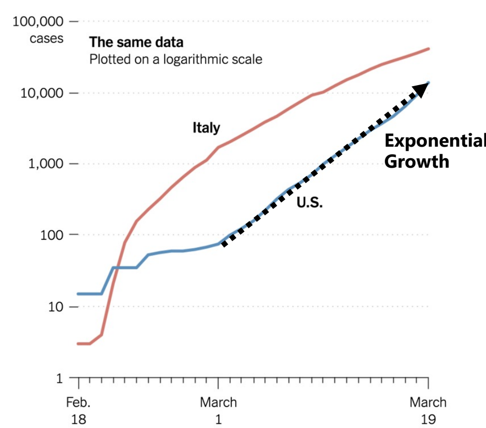


47

Bending the Curve

Logarithmic scales can emphasize the rate of change in a way that linear scales do not. Italy seems to be slowing the coronavirus infection rate, while the number of cases in the United States continues to double every few days.





By The New York Times | Data from Worldometer

48

WHEN TO APPLY LOG SCALE?

Address data skew (e.g., long tails, outliers)
Enables comparison within and across multiple orders of magnitude

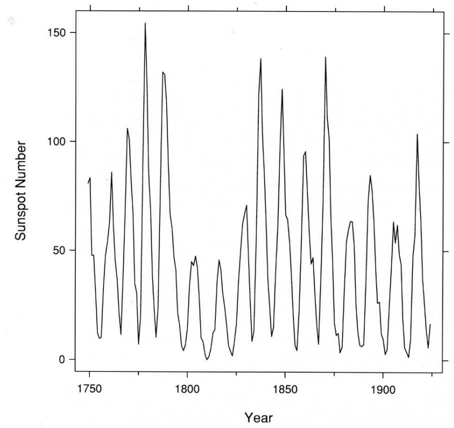
Focus on multiplicative factors (not additive)
Recall that the logarithm transforms **×** to **+** !
Percentage change, not linear difference

Constraint: **positive, non-zero values**
Constraint: **audience familiarity?**

49

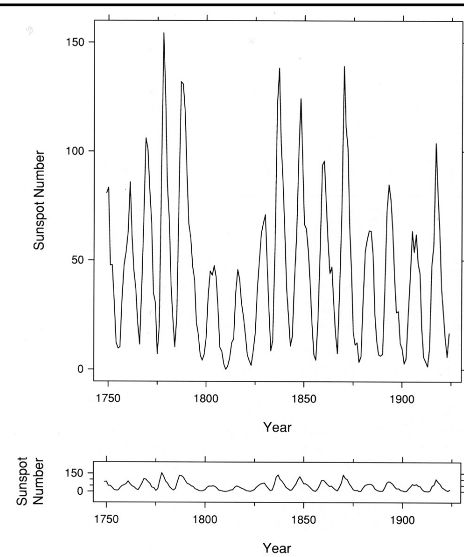
SELECTING ASPECT RATIO

55



Sunspots
from William S. Cleveland
The Elements of Graphing Data, 1985

56

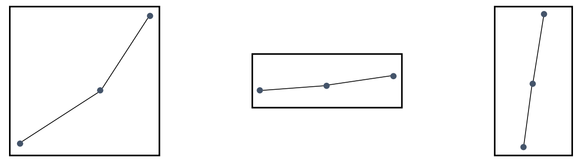


Sunspots
from William S. Cleveland
The Elements of Graphing Data, 1985

57

BANKING TO 45° [Cleveland 1985]

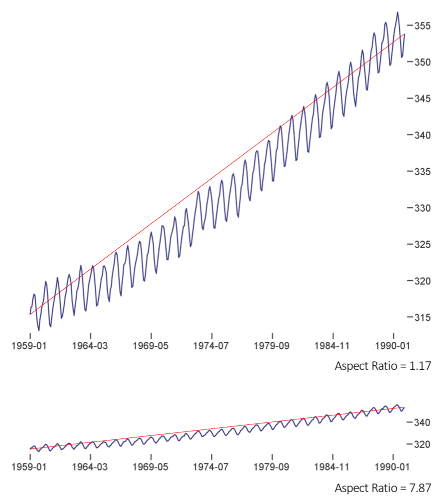
To facilitate perception of trends and maximize the discriminability of line segment orientations



Line segments are maximally discriminable when the absolute angle between them is 45°

Method: Optimize the *aspect ratio* such that the average absolute angle between all segments is 45°

58



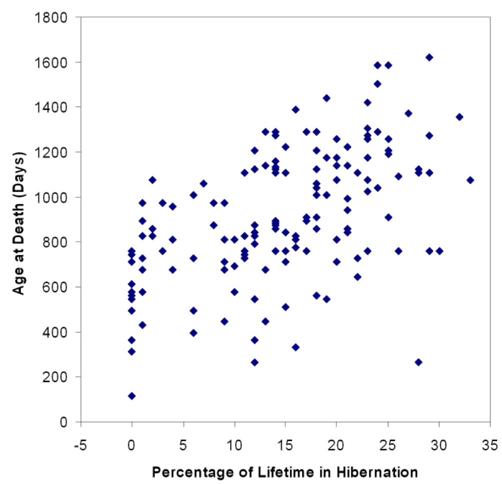
CO_2 Measurements
from William S. Cleveland
Visualizing Data, 1993

Trends may occur at different scales!
Apply banking to the original data or to fitted trend lines. [Heer & Agrawala 2006]

73

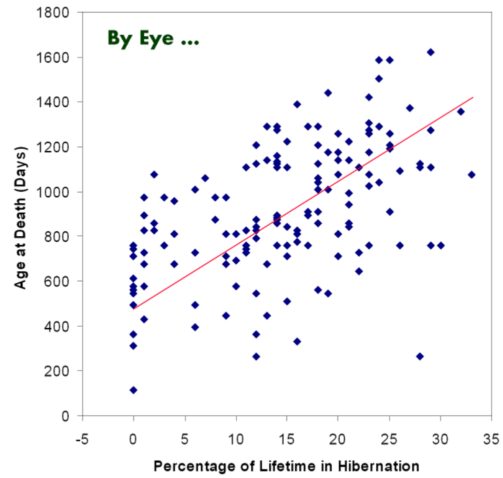
FITTING THE DATA

74



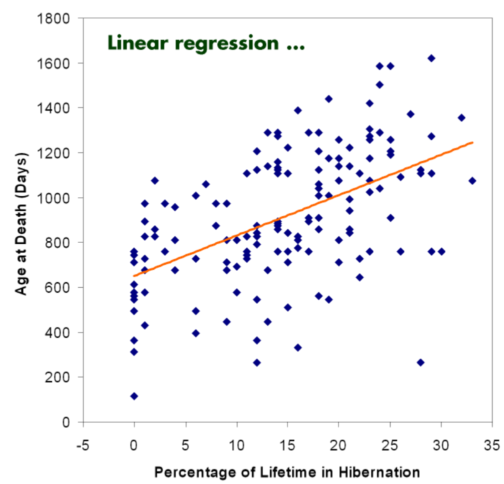
from The Elements of Graphing Data [Cleveland 1994]

75



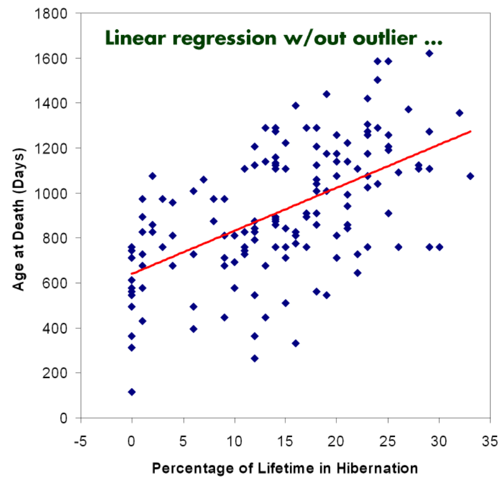
from The Elements of Graphing Data [Cleveland 1994]

76



from The Elements of Graphing Data [Cleveland 1994]

77

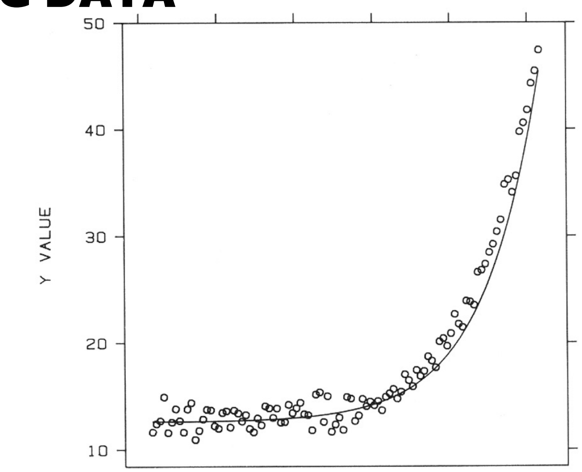


from The Elements of Graphing Data [Cleveland 1994]

78

TRANSFORMING DATA

How well does curve fit data?



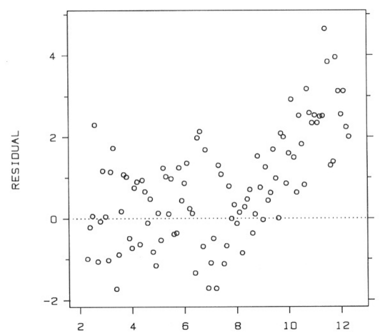
[Cleveland 85]

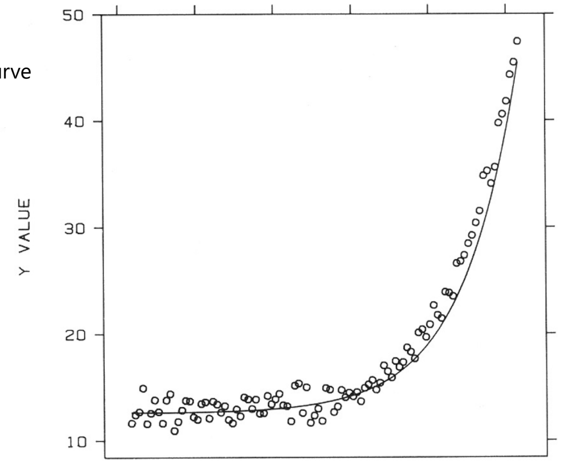
79

TRANSFORMING DATA

Residual graph

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





[Cleveland 85]

80

SORTING

88