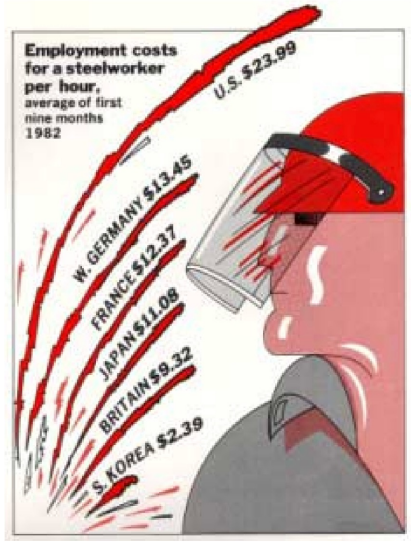
CHART DESIGN

CS 448B | Fall 2023

MANEESH AGRAWALA

1



2

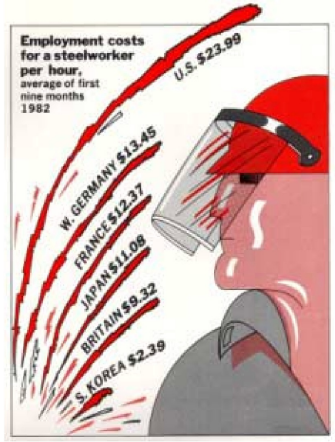
READING RESPONSE: QUESTIONS/THOUGHTS

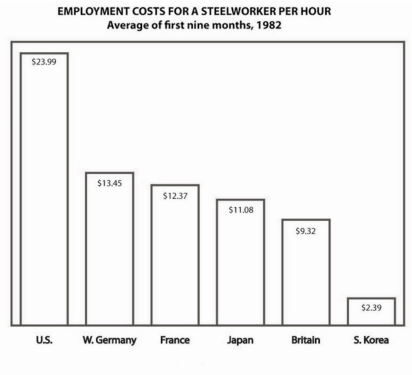
- I do wonder if there are instances in which ***exaggerated visuals are justified***: for example, public awareness campaigns, in which we play by the different rules of emotional psychology rather than statistical evidence in order to move an audience that traditionally bypasses statistics of growing issues.

- *"Cosmetic decoration, which frequently distorts the data, will never salvage an underlying lack of content."*: ***I don't agree*** with the generalized arguments made in the reading ***that lack of data cannot be enhanced by extra graphics and art***, even when those graphics become the central component of the visualization.

3

CHARTJUNK: IS IT USEFUL? from [Bateman 2010]





4

LAST TIME: EXPLORATORY DATA ANALYSIS

5

LESSON: EDA IS AN ITERATIVE PROCESS

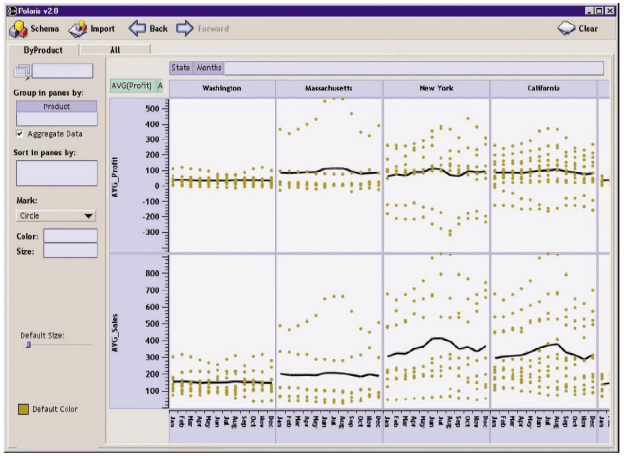
1. **Construct graphics to address questions**
2. **Inspect "answer" and assess new questions**
3. **Repeat!**

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

"Show data variation, not design variation" -Tufte

6

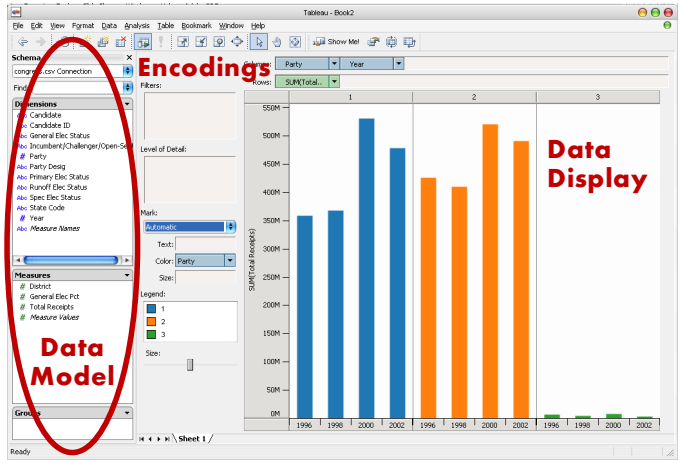
POLARIS [Stolte 2002]



Started as a Stanford research project by C. Stolte, D. Tang & P. Hanrahan

7

TABLEAU



8

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)

9

TABLEAU DEMO

Dataset:

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

10

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

11

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?

12

TABLEAU DEMO

13

SPECIFYING TABLE CONFIGURATIONS

Operands are names of database fields
Each operand interpreted as a set {…}
Data is either O or Q and are treated differently

Three operators:
concatenation (+)
cross product (x)
nest (/)

14

15

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
- **Data 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)**

16

ANNOUNCEMENTS

17

ASSIGNMENT 2: EXP. DATA ANALYSIS

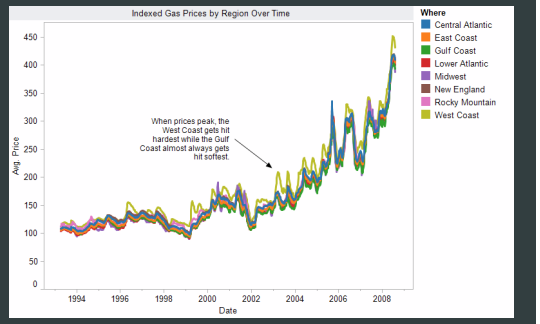
Due 10/16 11:30am

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



18

OFFICE HOURS: MANEESH

This week (11-noon tomorrow) will hold via Zoom.

Link will be posted on class Slack.

Come introduce yourself!

19

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.

20

DESIGN SPACE OF VISUAL ENCODINGS

22

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!

23

EXPRESSIVENESS CRITERIA [Mackinlay 1986]

Expressiveness

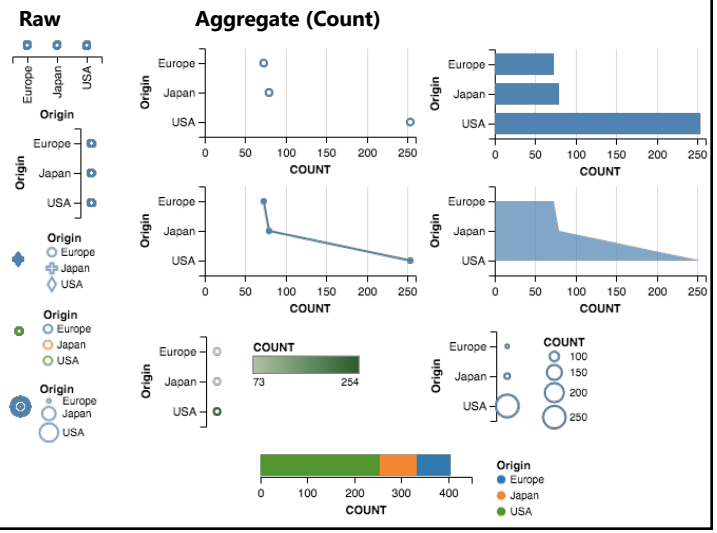
A set of facts is expressible in a visual language if the sentences (i.e., the visualizations) in the language express ***all*** the facts in the set of data, and ***only*** the facts in the data.

24

1D NOMINAL

Cars Data

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



25

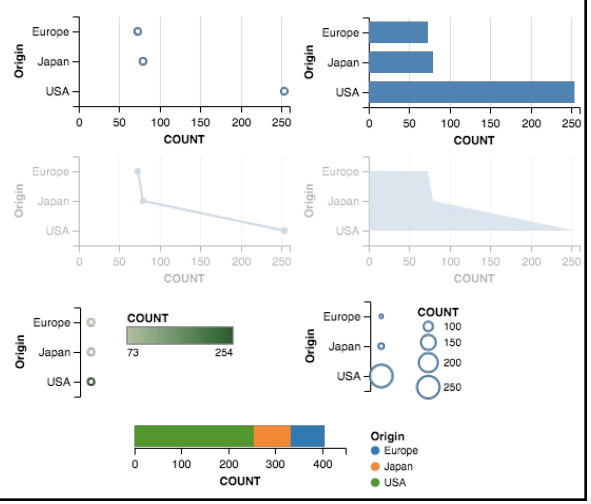
EXPRESSIVE?

Cars Data

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

Raw

Aggregate (Count)



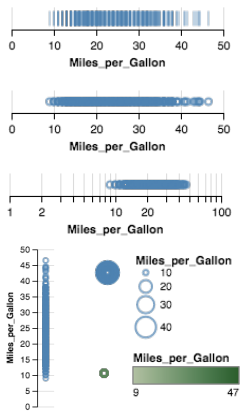
26

1D QUANTITATIVE

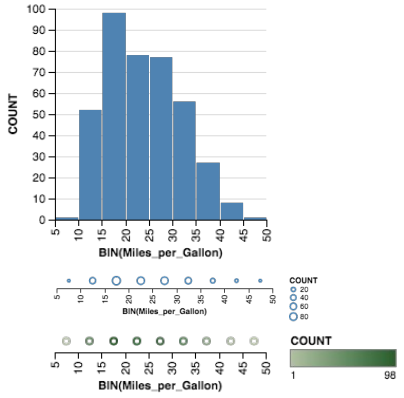
Cars Data

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

Raw



Aggregate (Count)



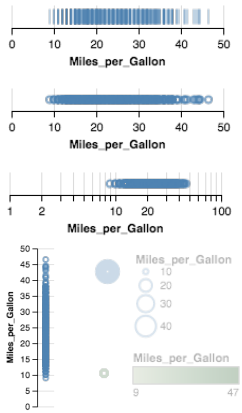
27

EXPRESSIVE?

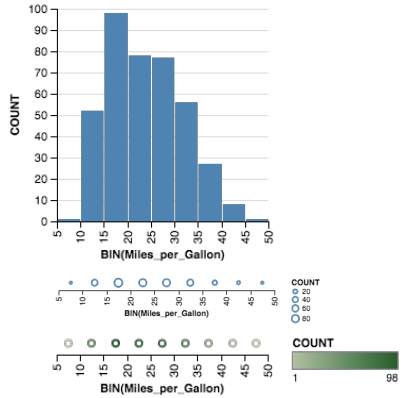
Cars Data

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

Raw

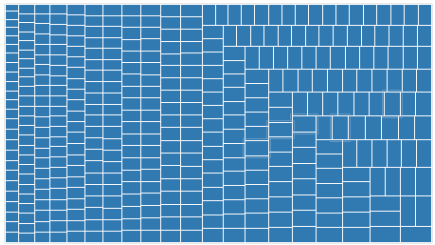


Aggregate (Count)

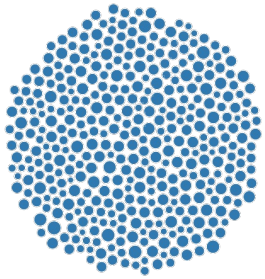


28

Raw (with Layout Algorithm)

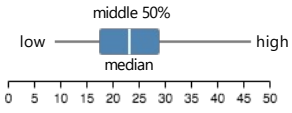


Treemap

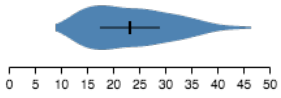


Bubble Chart

Aggregate (Distributions)



Box Plot



Violin Plot

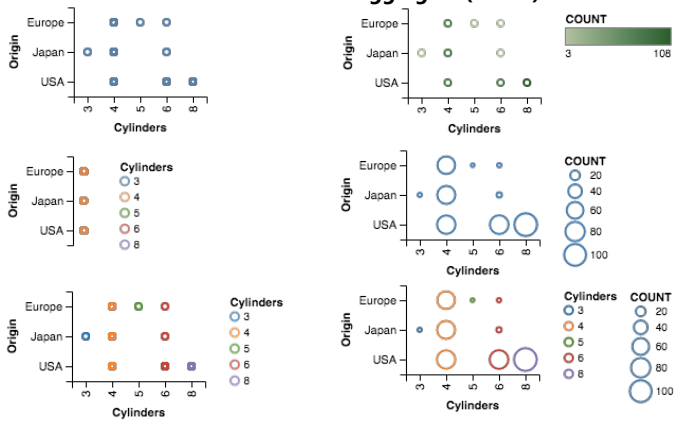
29

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)

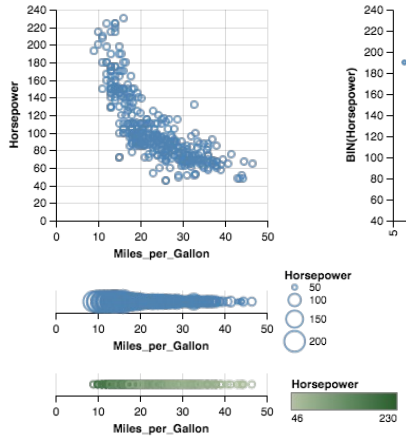
30

2D QUANTITATIVE x QUANTITATIVE

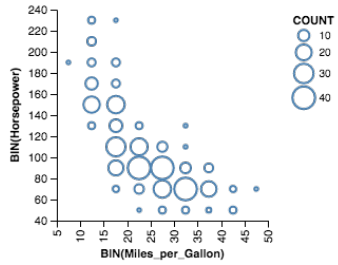
Cars Data

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

Raw



Aggregate (Count)



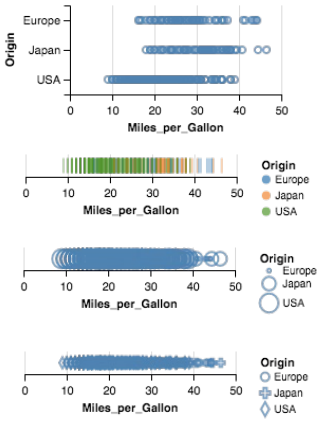
31

2D NOMINAL x QUANTITATIVE

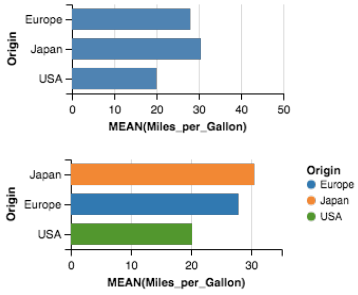
Cars Data

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

Raw

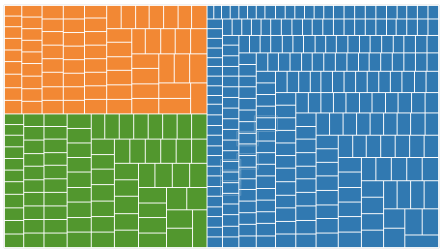


Aggregate (Mean)

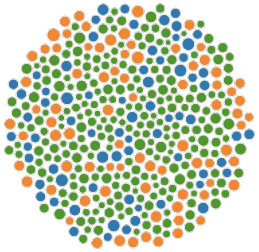


32

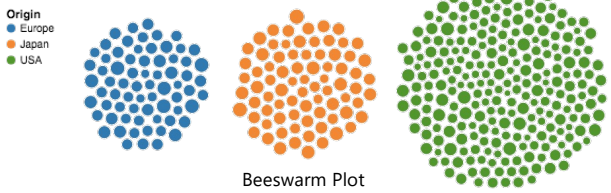
Raw (with Layout Algorithm)



Treemap



Bubble Chart



Beeswarm Plot

33

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.

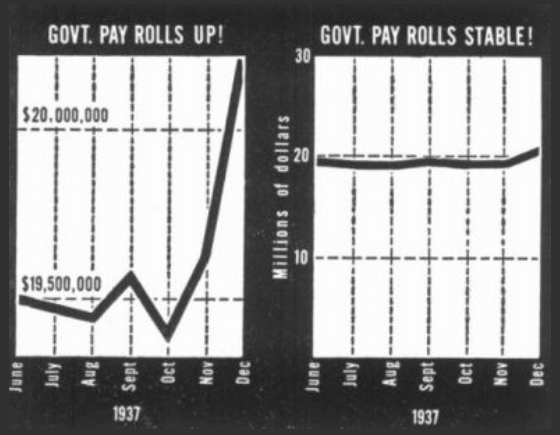
Subject of the Perception Lecture

34

SCALES AND AXES

40

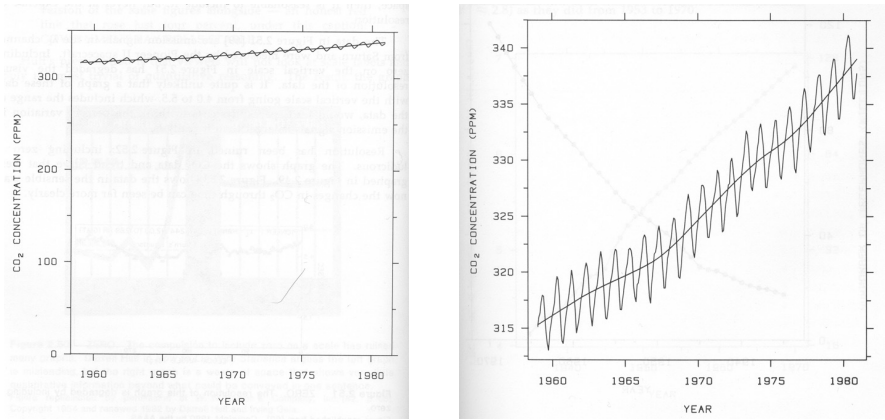
WHICH GRAPH IS BETTER



Government payrolls in 1937 [Huff 93]

41

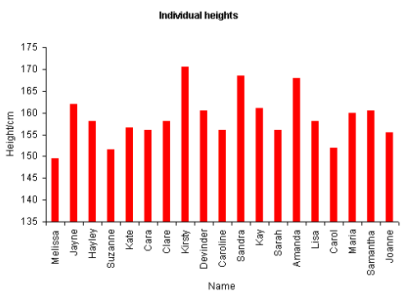
INCLUDE ZERO IN AXIS SCALE?



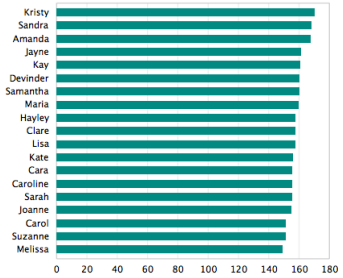
Yearly CO2 concentrations [Cleveland 85]

42

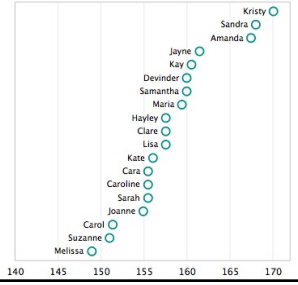
INCLUDE ZERO IN AXIS SCALE?



Compare Proportions (Q-Ratio)

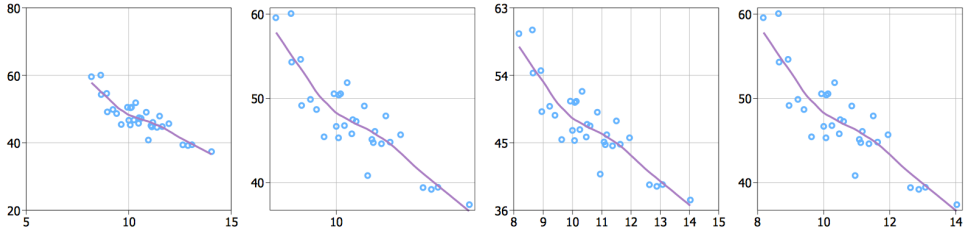


Compare Relative Position (Q-Interval)



43

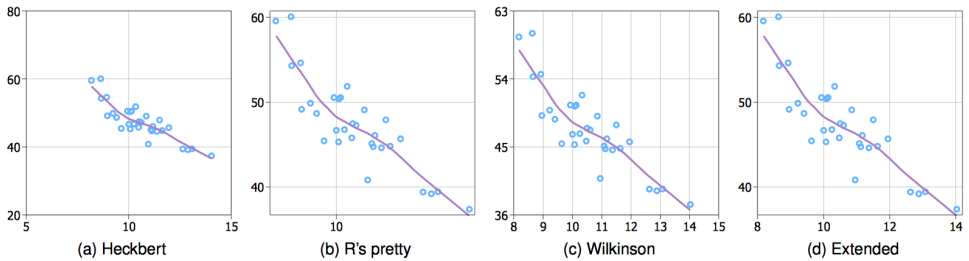
AXIS TICK MARK SELECTION



What are some properties of "good" tick marks?

44

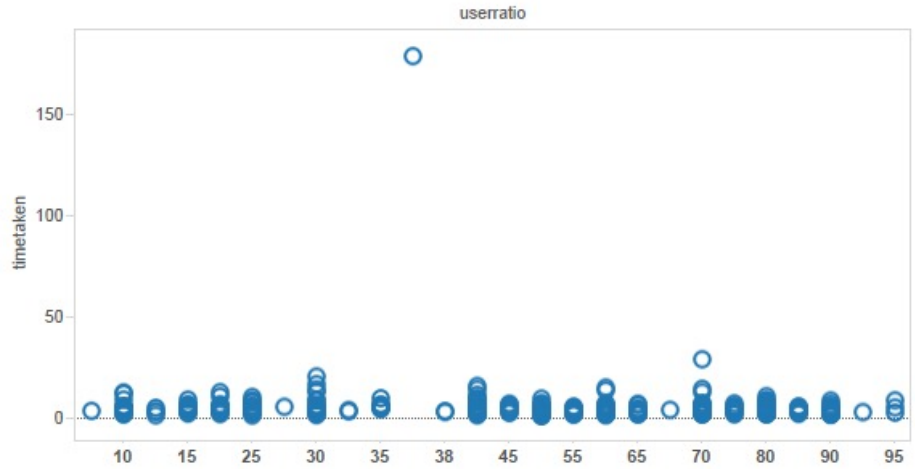
AXIS TICK MARK SELECTION



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

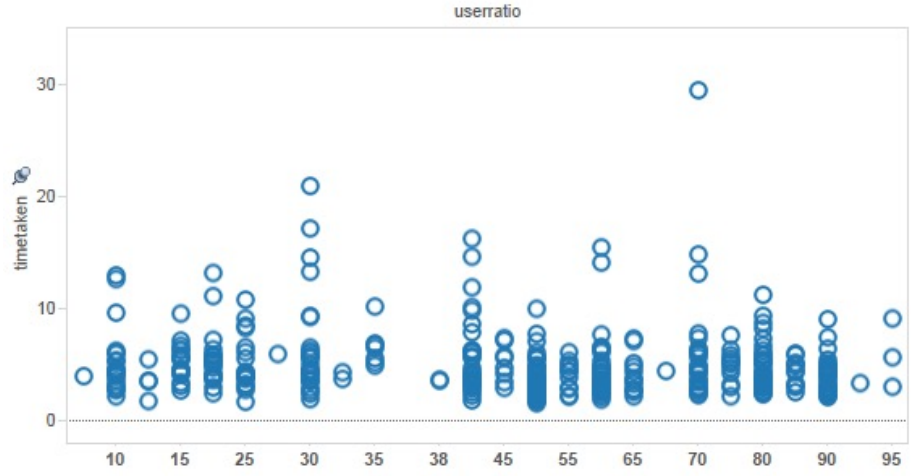
45

HOW TO SCALE THE AXIS?



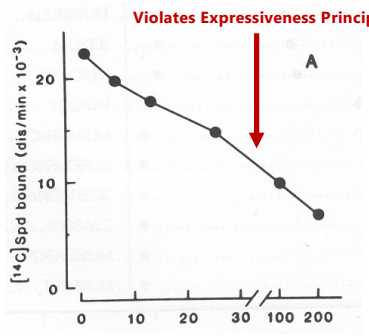
46

OPTION 1 : CLIP OUTLIERS

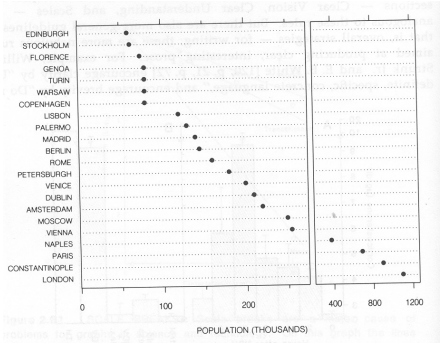


47

OPTION 2: SCALE BREAKS – CLEARLY MARKED



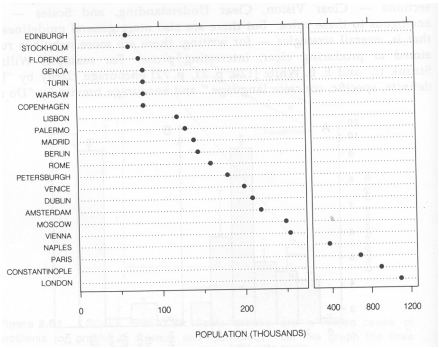
Poor scale break [Cleveland 85]



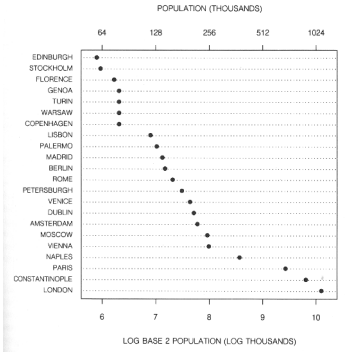
Well marked scale break [Cleveland 85]

48

OPTION 3: LOG SCALE



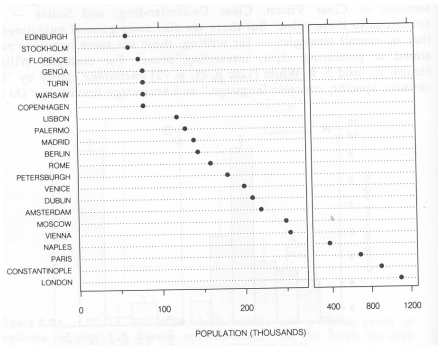
Scale break [Cleveland 85]

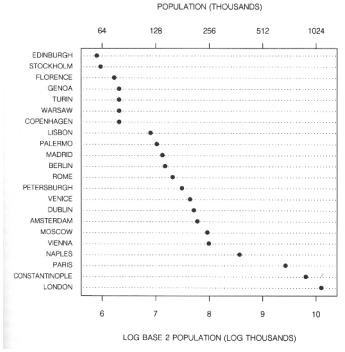


Log scale[Cleveland 85]

49

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

50

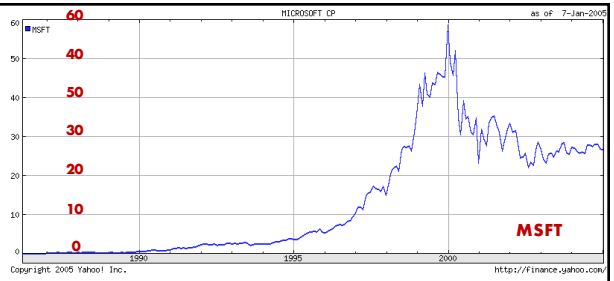
Logarithms turn ***multiplications*** into ***additions***

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

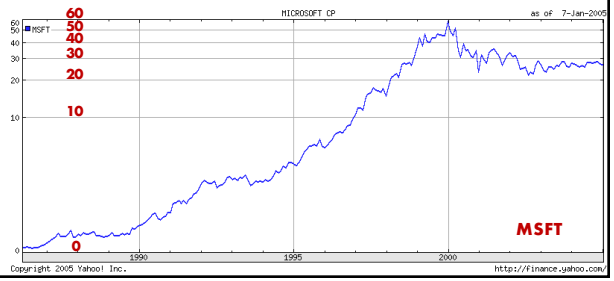
Equal steps on a log scale correspond to equal changes to a multiplicative scale factor

51

Linear Scale



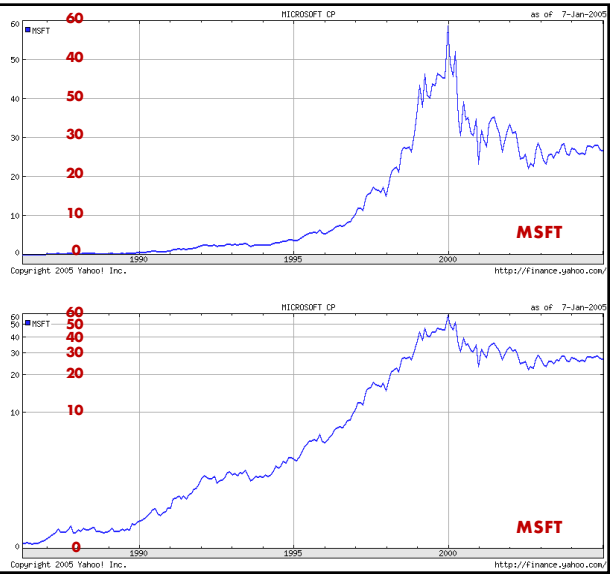
Log Scale



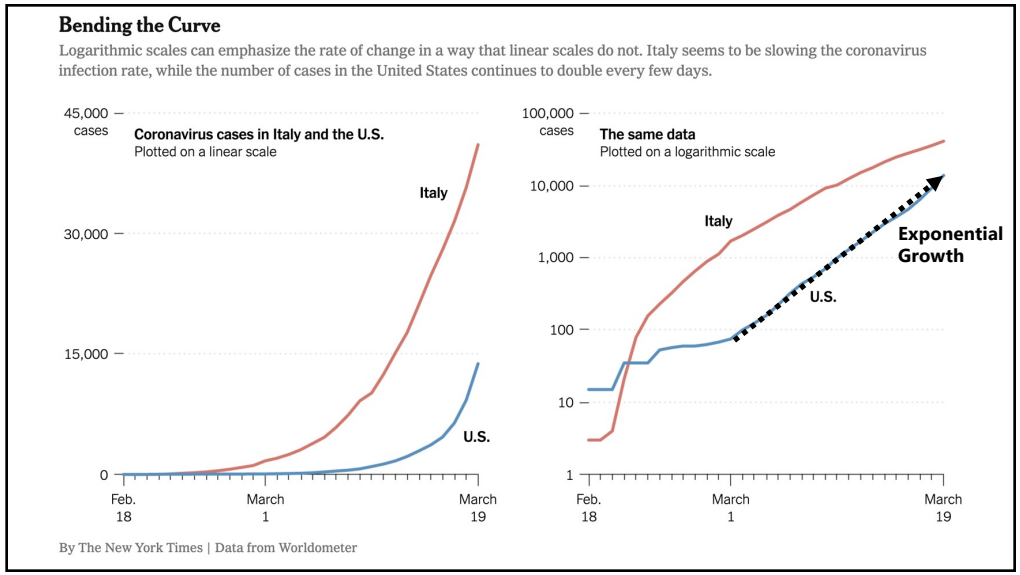
52

Linear Scale
Absolute change

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



53



54

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?**

55