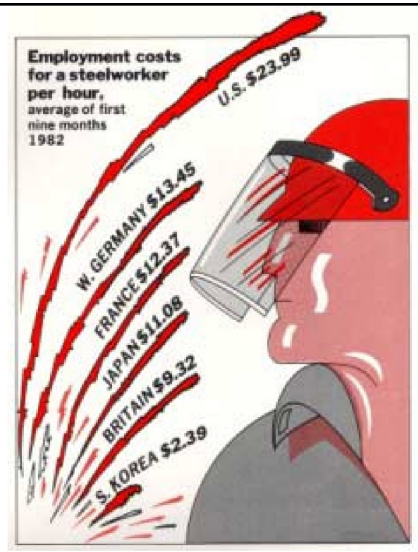
CHART DESIGN

CS 448B | Fall 2024

MANEESH AGRAWALA

1



2

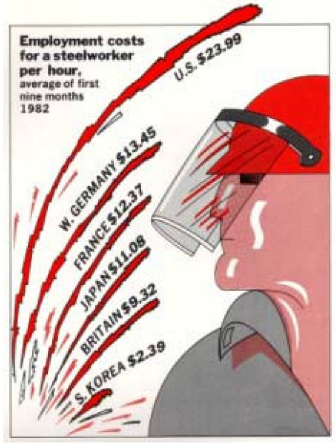
READING RESPONSE: QUESTIONS/THOUGHTS

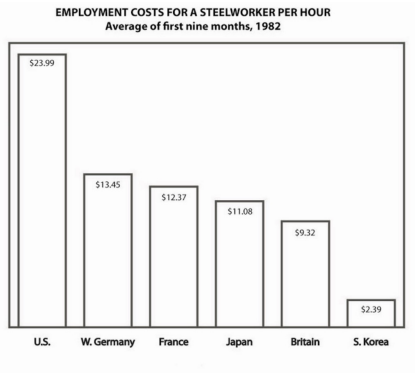
- In our interconnected, information-overloaded, and busy lives it's easy to see the value of erasing unnecessary information, and in general, I agree with Tufte. However, I think here ***it's critical to know your audience, rather than just create the highest data-ink ratio***. Tufte does say to erase "within reason", but says also the reason is "nearly always" that that ink presents new information. I believe consistency and context should also be valuable reasons to use extra ink.

- 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?***

3

CHARTJUNK: IS IT USEFUL? from [Bateman 2010]





4

LAST TIME: EXPLORATORY DATA ANALYSIS

5

LESSON: EDA IS AN ITERATIVE PROCESS

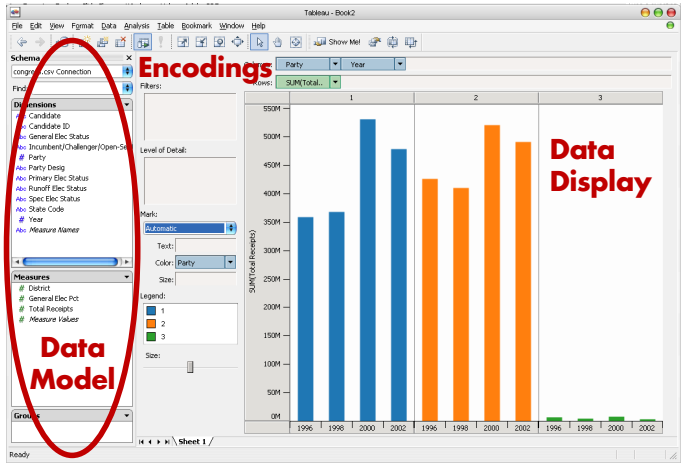
1. **Construct graphics to address questions**
2. **Inspect "answer" and assess new questions**
3. **Repeat!**

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

"Show data variation, not design variation" -Tufte

6

TABLEAU



7

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)

8

ANNOUNCEMENTS

9

ASSIGNMENT 2: EXP. DATA ANALYSIS

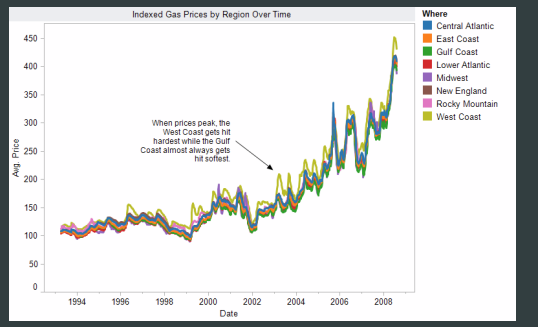
Due 10/14 10:30am

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



10

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.

11

DESIGN SPACE OF VISUAL ENCODINGS

13

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!

14

EXPRESSIVENESS CRITERIA [Mackinlay 1986]

Expressiveness

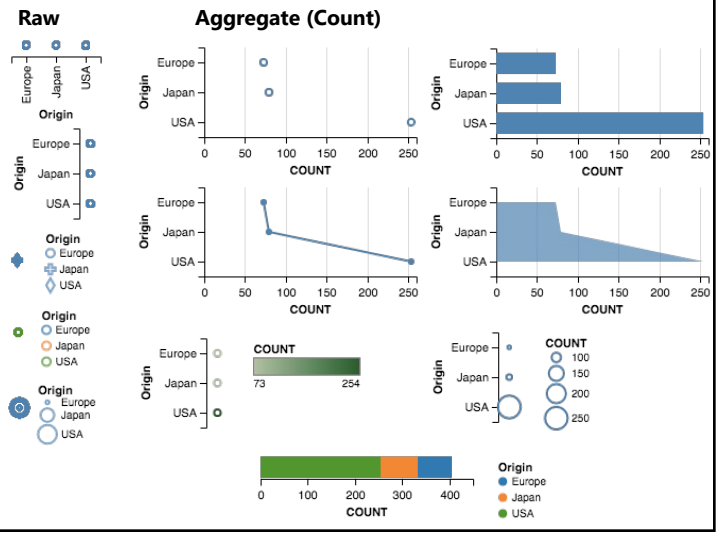
A set of facts is expressible in a visual language if the sentences (i.e., the visualizations) in the language express ***all*** the facts in the set of data, and ***only*** the facts in the data.

15

1D NOMINAL

Cars Data

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

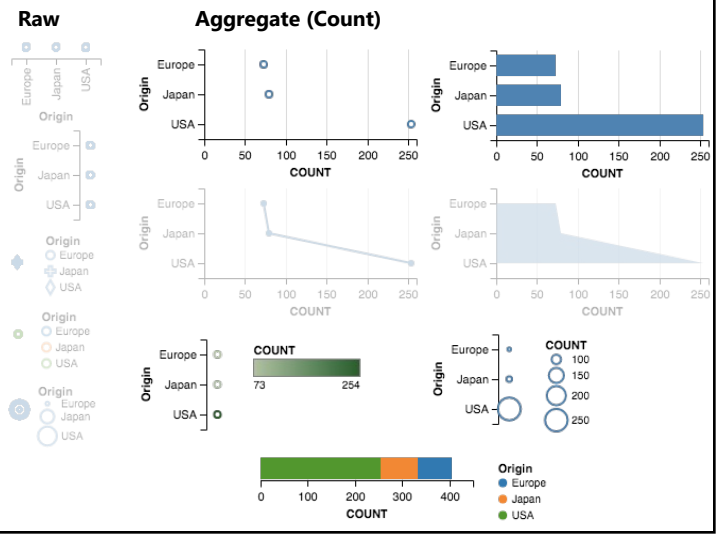


16

EXPRESSIVE?

Cars Data

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



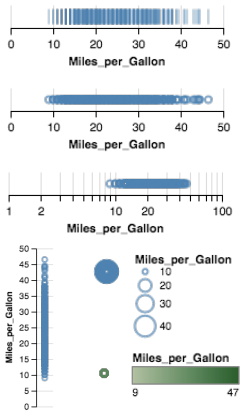
17

1D QUANTITATIVE

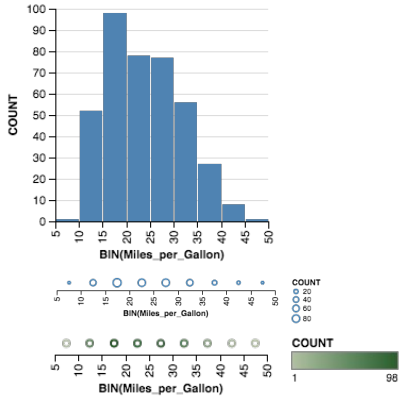
Cars Data

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

Raw



Aggregate (Count)



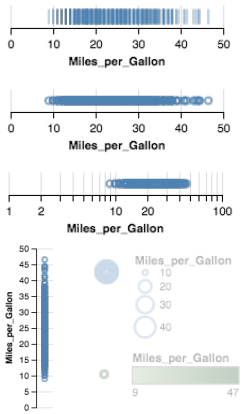
18

EXPRESSIVE?

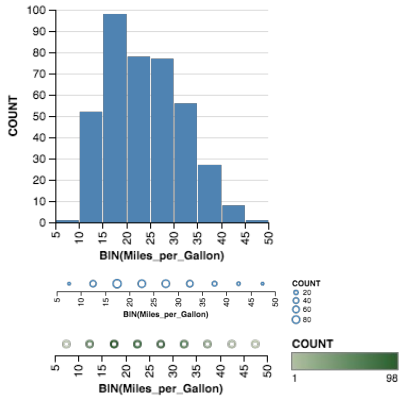
Cars Data

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

Raw

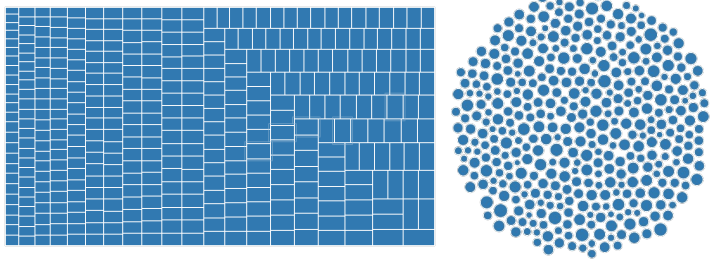


Aggregate (Count)



19

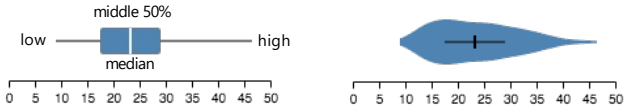
Raw (with Layout Algorithm)



Treemap

Bubble Chart

Aggregate (Distributions)



Box Plot

Violin Plot

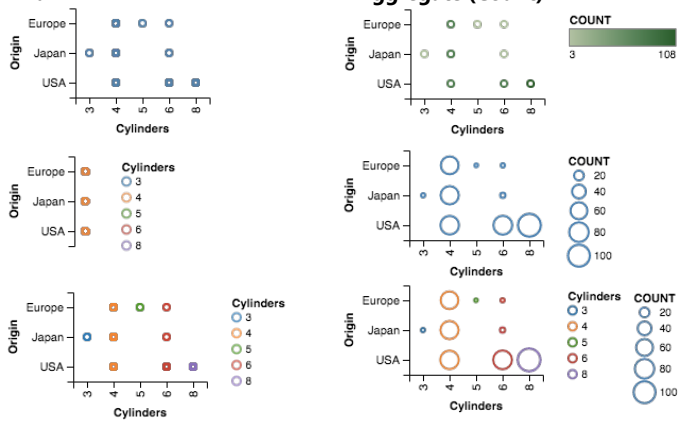
20

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)

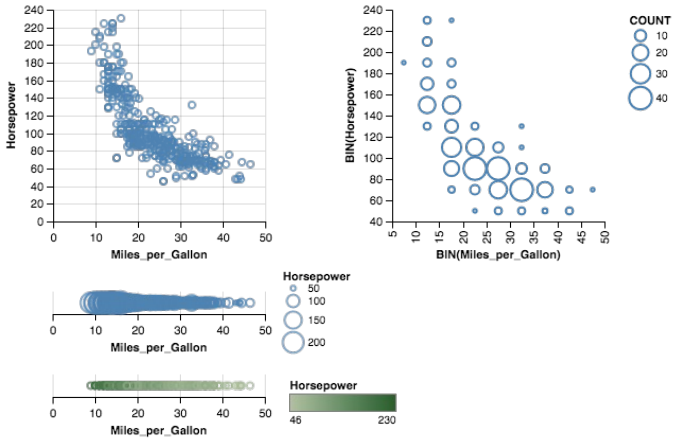
21

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)

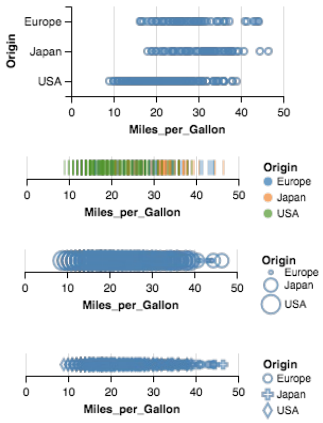
22

2D NOMINAL x QUANTITATIVE

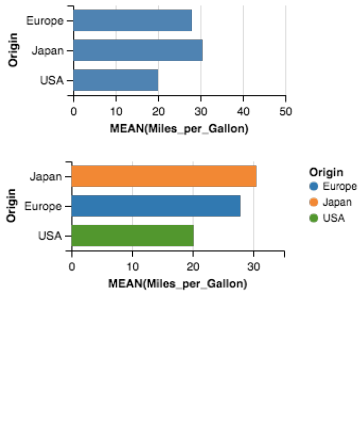
Cars Data

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

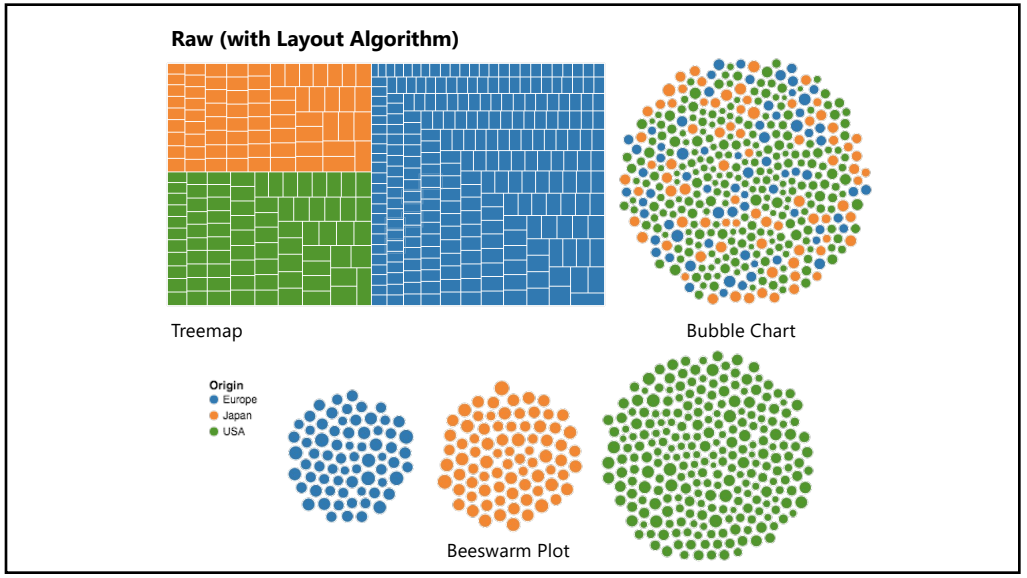
Raw



Aggregate (Mean)



23



24

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.

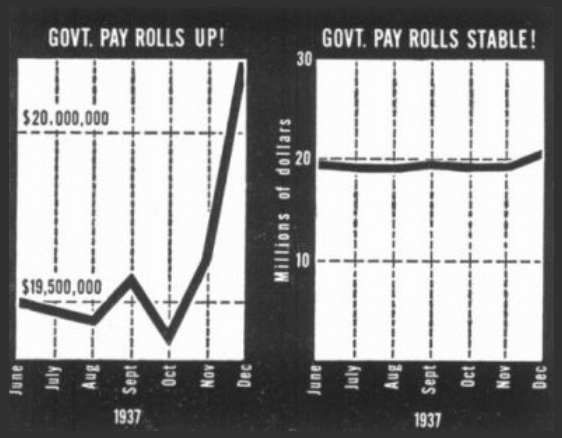
Subject of the Perception Lecture

25

SCALES AND AXES

31

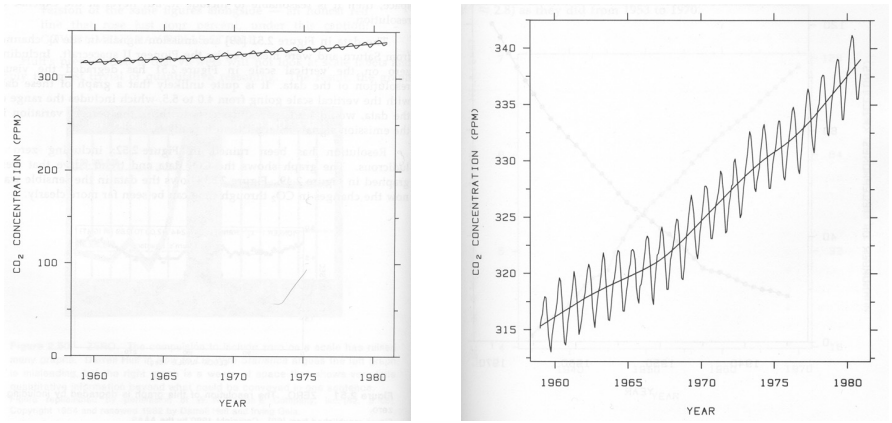
WHICH GRAPH IS BETTER



Government payrolls in 1937 [Huff 93]

32

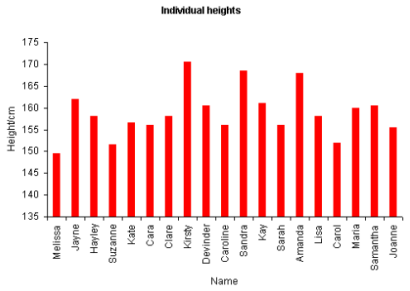
INCLUDE ZERO IN AXIS SCALE?



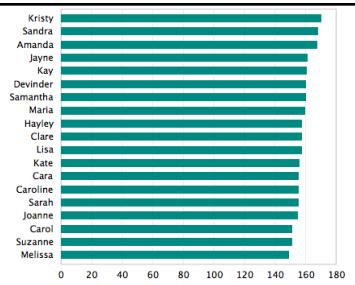
Yearly CO2 concentrations [Cleveland 85]

33

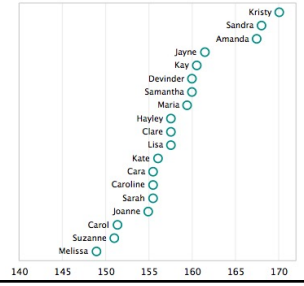
INCLUDE ZERO IN AXIS SCALE?



Compare Proportions (Q-Ratio)

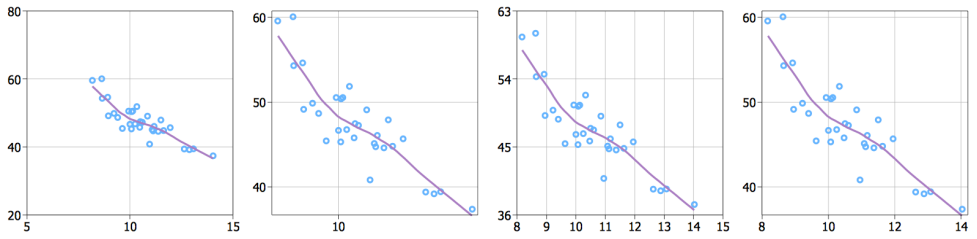


Compare Relative Position (Q-Interval)



34

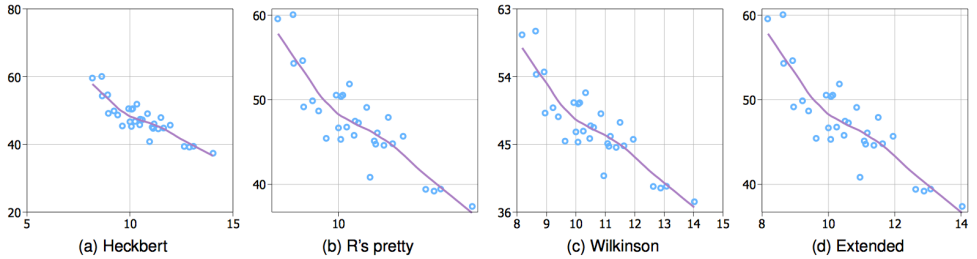
AXIS TICK MARK SELECTION



What are some properties of "good" tick marks?

35

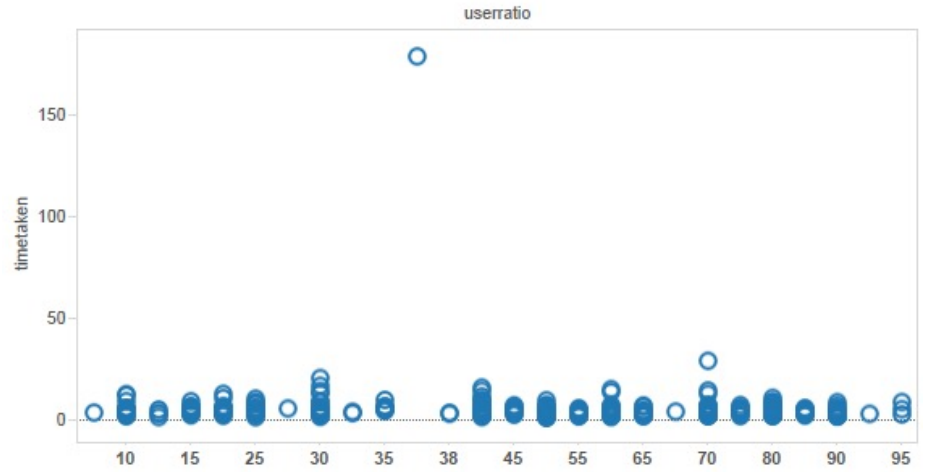
AXIS TICK MARK SELECTION



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

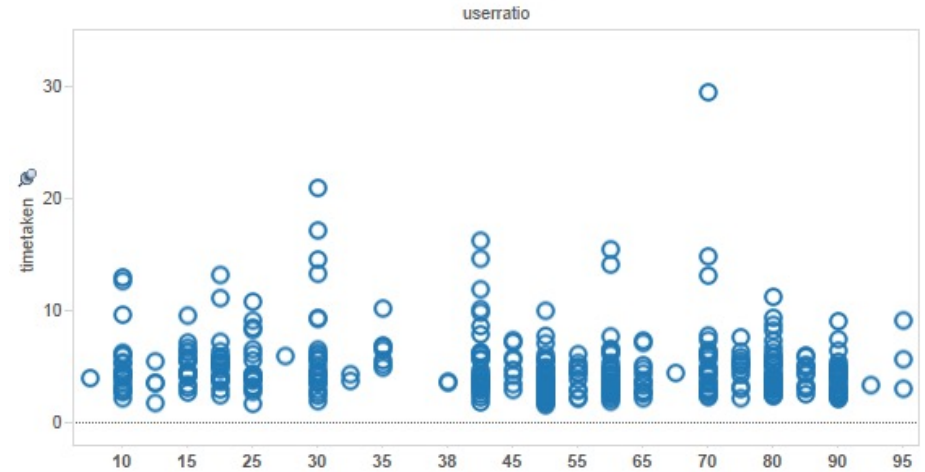
36

HOW TO SCALE THE AXIS?



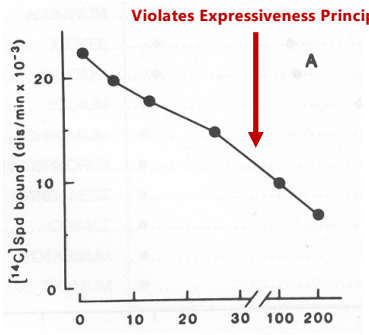
37

OPTION 1 : CLIP OUTLIERS

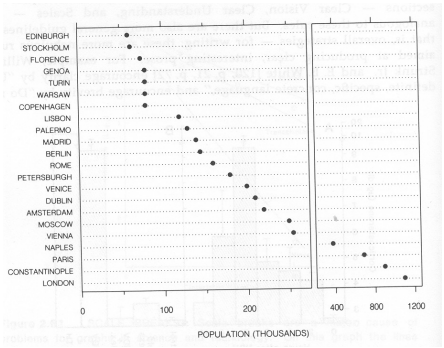


38

OPTION 2: SCALE BREAKS – CLEARLY MARKED



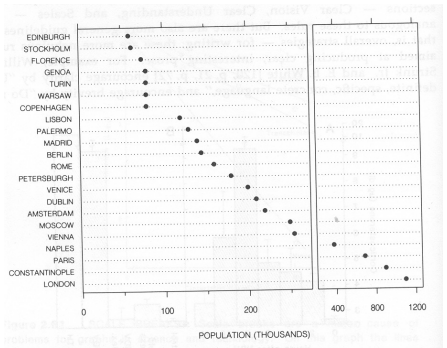
Poor scale break [Cleveland 85]



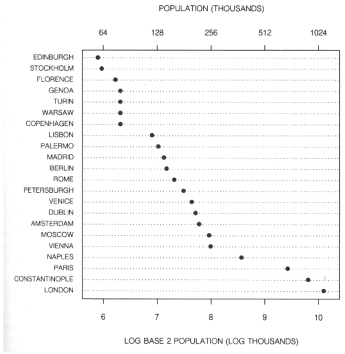
Well marked scale break [Cleveland 85]

39

OPTION 3: LOG SCALE



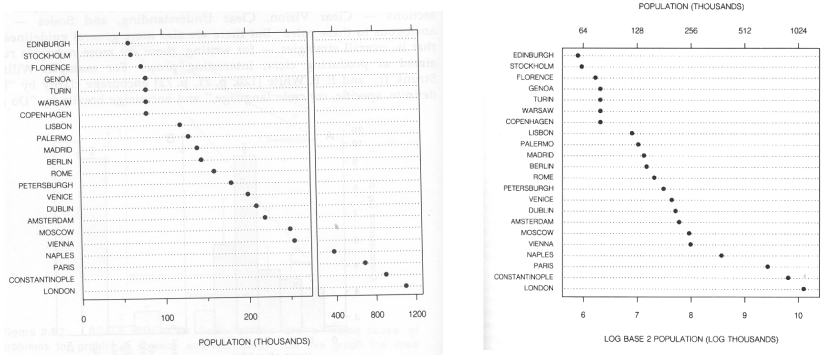
Scale break [Cleveland 85]



Log scale[Cleveland 85]

40

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

41

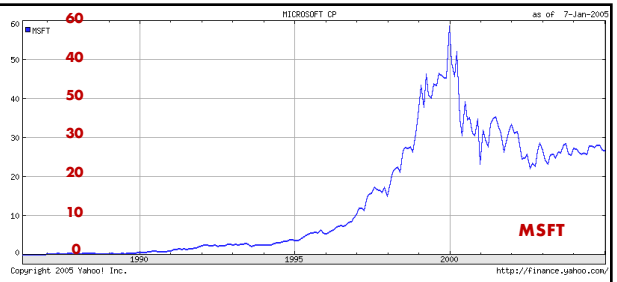
Logarithms turn ***multiplications*** into ***additions***

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

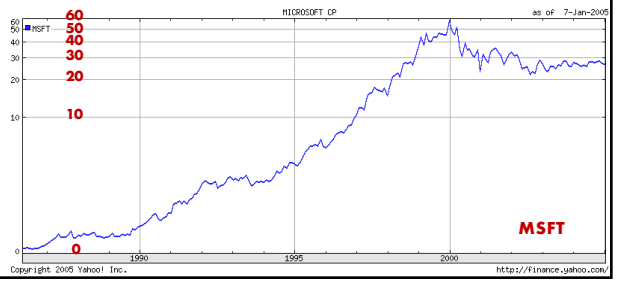
Equal steps on a log scale correspond to equal changes to a multiplicative scale factor

42

Linear Scale

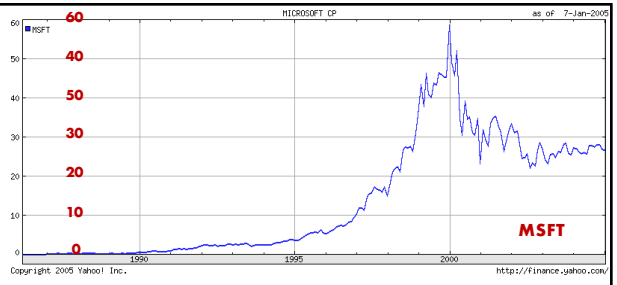


Log Scale

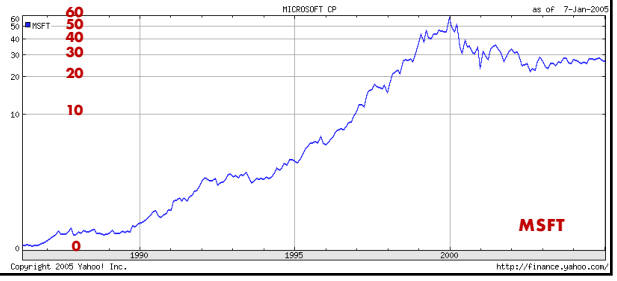


43

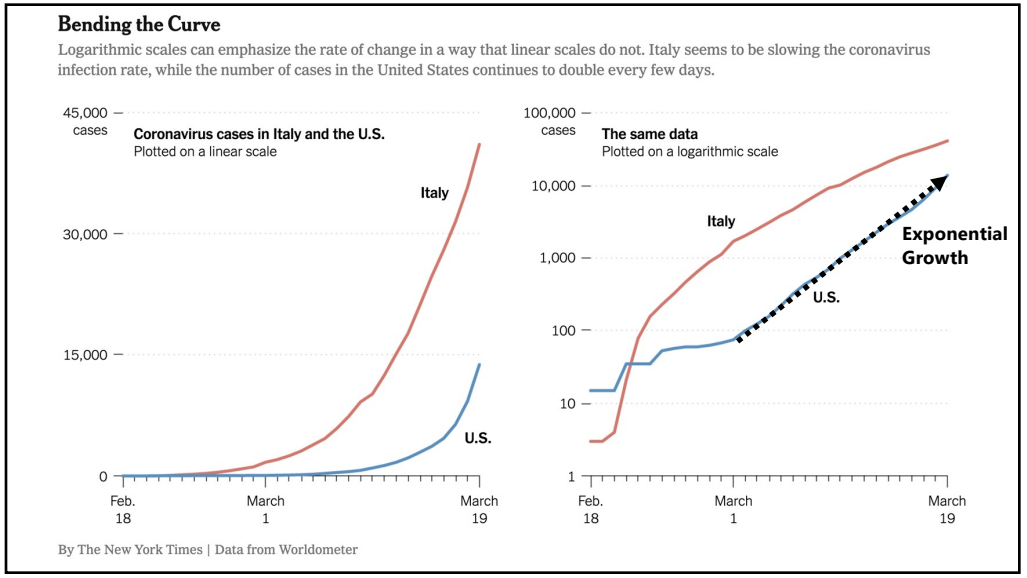
Linear Scale
- Absolute change



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



44



45

WHEN TO APPLY LOG SCALE?

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

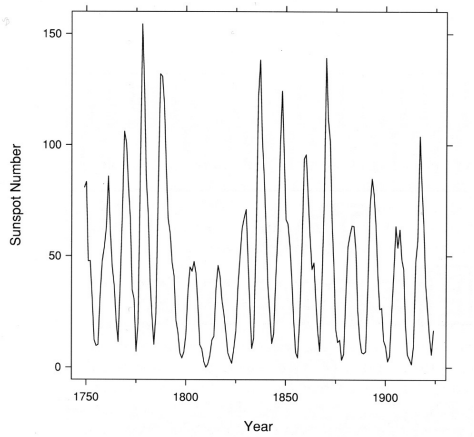
Focus on multiplicative factors (not additive)
Recall that the logarithm transforms **×** to **+** !
Percentage change, not linear difference

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

46

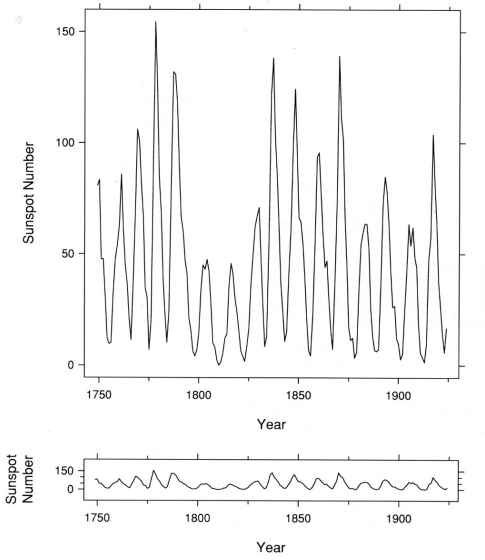
SELECTING ASPECT RATIO

52



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

53

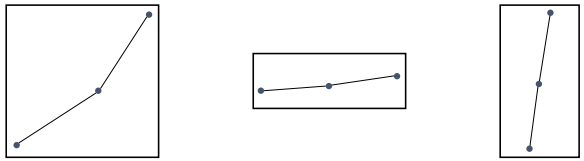


Sunspots
from William S. Cleveland
The Elements of Graphing Data, 1985

54

BANKING TO 45° [Cleveland 1985]

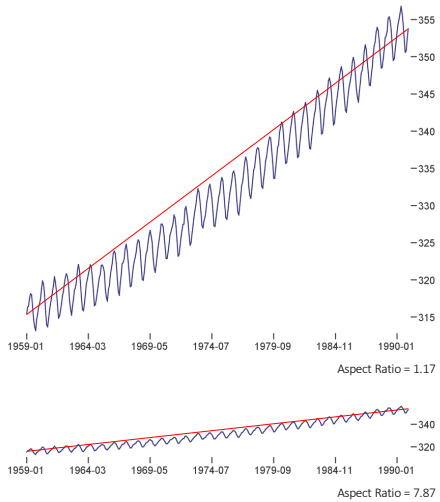
To facilitate perception of trends and maximize the discriminability of line segment orientations



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°

55



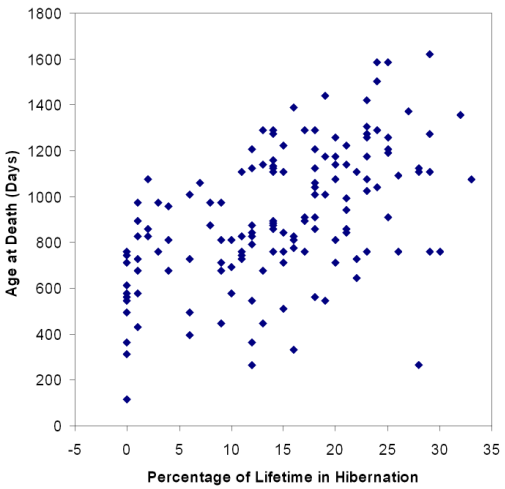
CO_2 Measurements
from William S. Cleveland
Visualizing Data, 1993

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

70

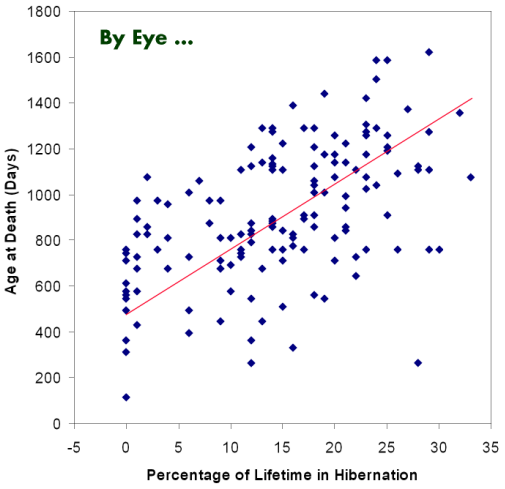
FITTING THE DATA

71



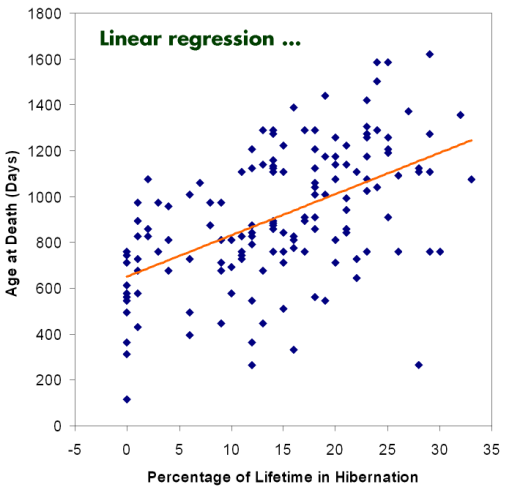
from The Elements of Graphing Data [Cleveland 1994]

72



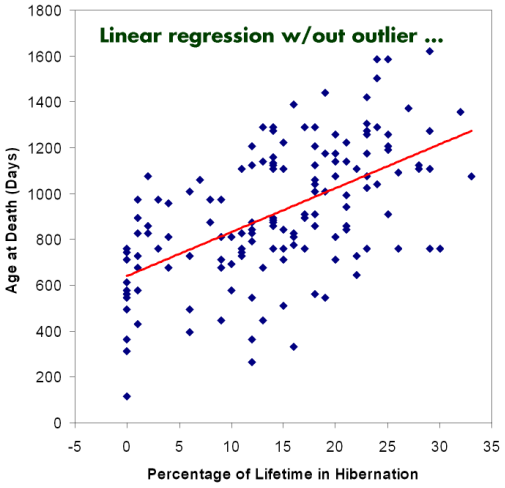
from The Elements of Graphing Data [Cleveland 1994]

73



from The Elements of Graphing Data [Cleveland 1994]

74

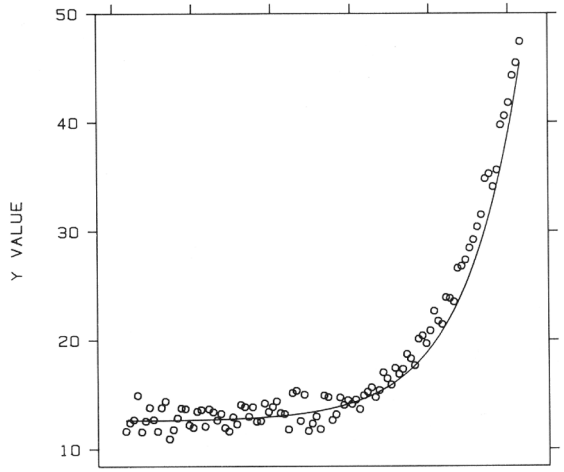


from The Elements of Graphing Data [Cleveland 1994]

75

TRANSFORMING DATA

How well does curve fit data?



[Cleveland 85]

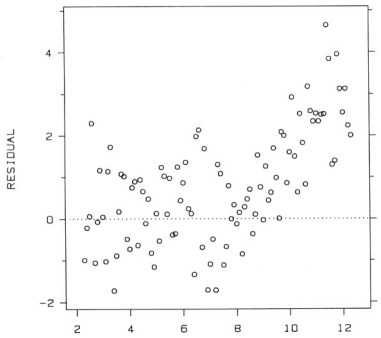
76

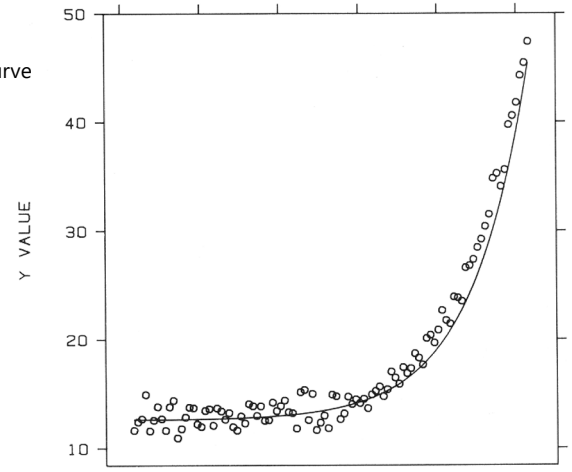
TRANSFORMING DATA

Residual graph

Plot vertical distance from best fit curve

Residual graph shows accuracy of fit



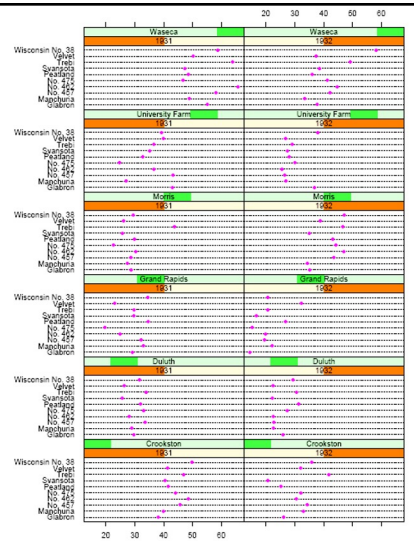


[Cleveland 85]

77

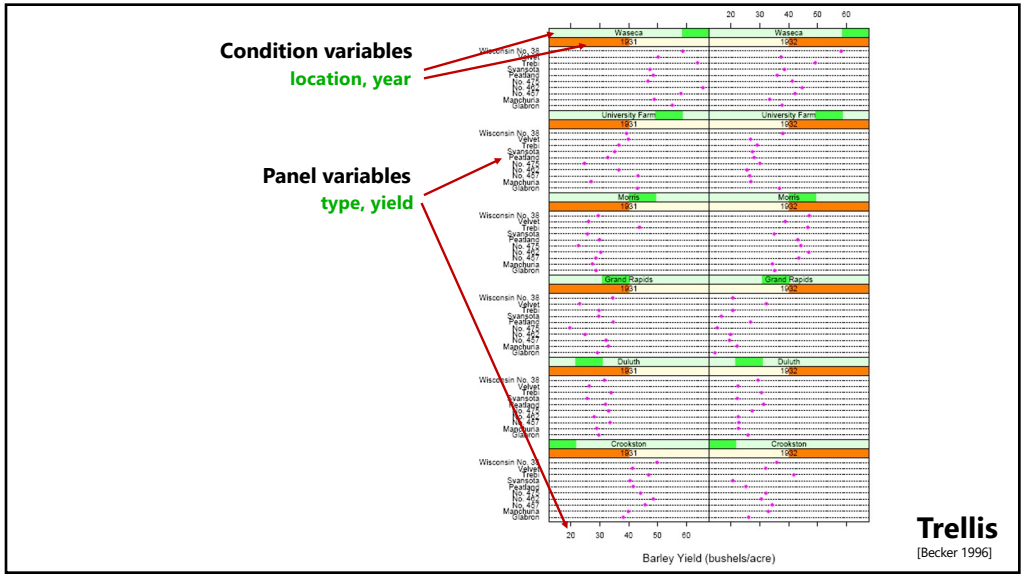
SORTING

85

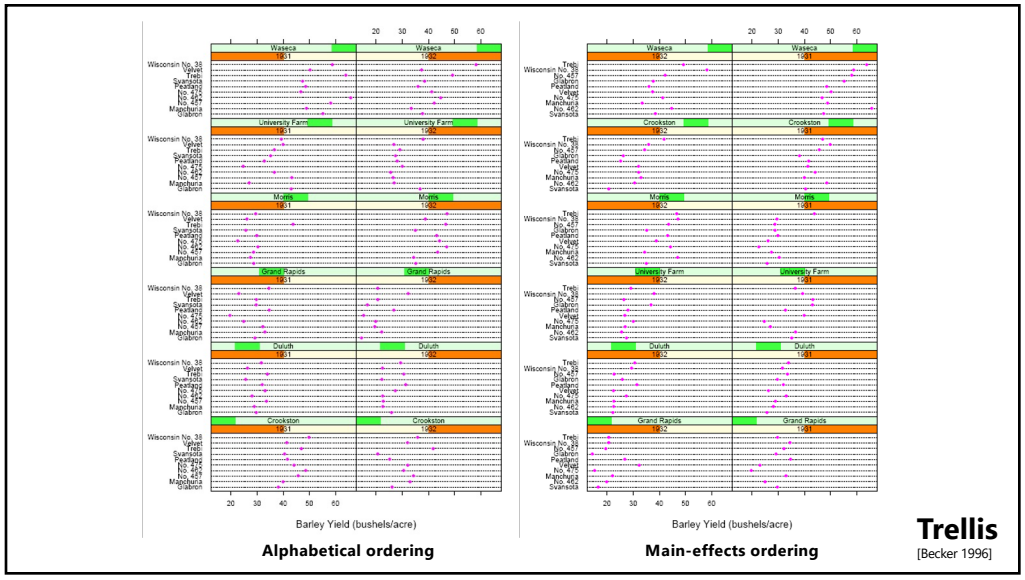


Trellis
[Becker 1996]

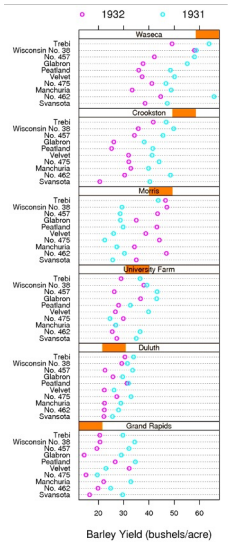
86


Condition variables
location, year
Panel variables
type, yield
Barley Yield (bushels/acre)
Trellis
[Becker 1996]

87


Barley Yield (bushels/acre)
Alphabetical ordering
Barley Yield (bushels/acre)
Main-effects ordering
Trellis
[Becker 1996]

88



Trellis
[Becker 1996]

89

SUMMARY

Well designed visualizations

Use *expressive* and *effective* encodings

Avoid *over-encoding*

Emphasize features of the data most relevant to the task

Rarely does a single visualization answer all questions.
Instead, the ability to generate appropriate visualizations quickly is critical!

90

DESIGN EXERCISE

91

13 8

How many visualizations can you think of for conveying these two numbers?
Feel free to invent tasks or contexts. **Sketch as many as you can!**

Don't stress over quality, go for quantity.

Time: ~5 minutes

92

13 8

Share your designs with fellow students. Introduce yourselves! Then compare your designs. How many ideas are the same? How many are different?

Capture your favorite images and post them on the Slack channel **#in-class-activity-10-07-24**

93