Using Space Effectively

Maneesh Agrawala

CS 448B: Visualization
Fall 2021

1

Reading Response Questions/Thoughts

How do we know which type of visualization to use? Are there some general principals that lead us to choose a bar chart over a pie chart? What is the psychology of different mark types and visual encodings?

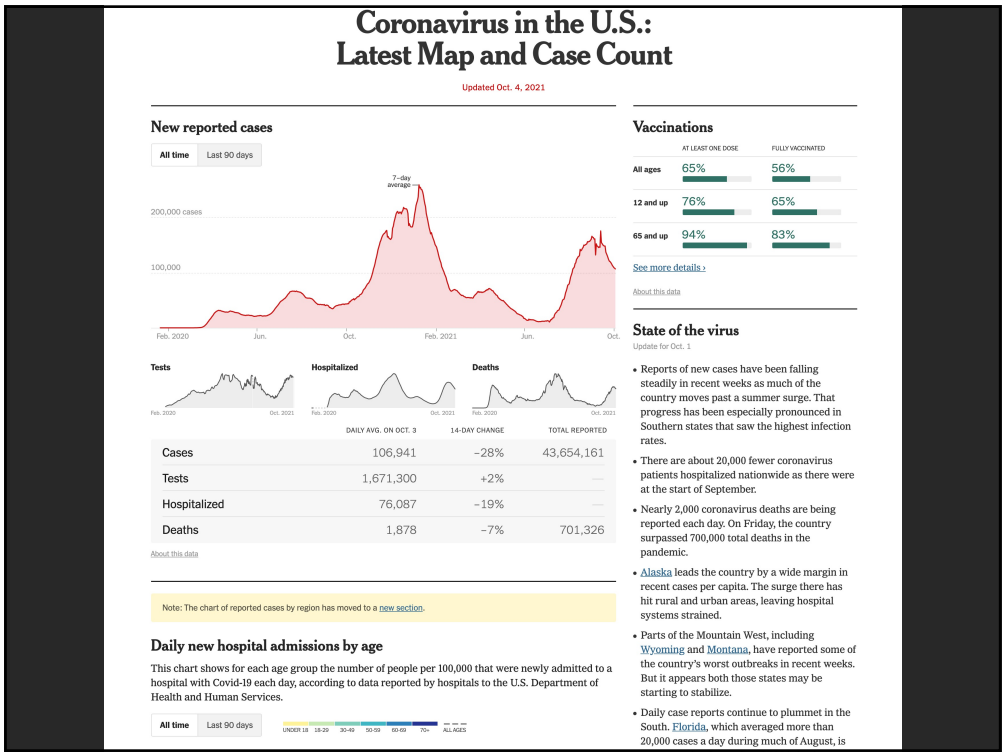
Is there a standard/scientific method of sorts by which graphic designers are supposed to explore, iterate, and finalize their designs?

In reference to the social network graph from Wednesday lecture with the node-link, linkage-sorted matrix, and non-sorted matrix views, "Are there other algorithms that can help bring out specific patterns in your data?"

In reference to public (Twitter) vs. private (academic) data visualization critiques and how people have paid more attention to data visualizations during the COVID-19 pandemic: "Do readers' goals align with designers' goals and if they don't how does that impact the insights that users walk away with as well as the redesign process?

Is it fair to leave it solely up to the experts? Furthermore, how do authors communicate their goals to users?"

2



Coronavirus in the U.S.: Latest Map and Case Count

Updated Oct. 4, 2021

New reported cases

All time | Last 90 days

7-day average

200,000 cases

100,000

Feb. 2020 Jun. Oct. Feb. 2021 Jun. Oct.

Tests

Hospitalized

Deaths

	DAILY AVG. ON OCT. 3	14-DAY CHANGE	TOTAL REPORTED
Cases	106,941	−28%	43,654,161
Tests	1,671,300	+2%	—
Hospitalized	76,087	−19%	—
Deaths	1,878	−7%	701,326

About this data

Note: The chart of reported cases by region has moved to a new section.

Daily new hospital admissions by age

This chart shows for each age group the number of people per 100,000 that were newly admitted to a hospital with Covid-19 each day, according to data reported by hospitals to the U.S. Department of Health and Human Services.

All time | Last 90 days

UNDER 18 18-29 30-49 50-59 60-69 70+ ALL AGES

Vaccinations

	AT LEAST ONE DOSE	FULLY VACCINATED
All ages	65%	56%
12 and up	76%	65%
65 and up	94%	83%

See more details ›

About this data

State of the virus

Update for Oct. 1

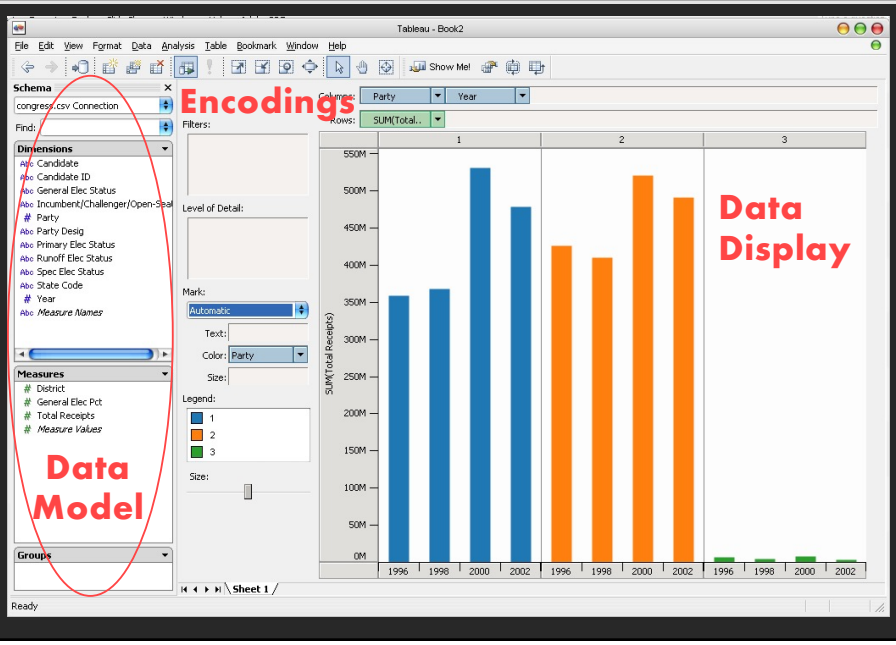
- Reports of new cases have been falling steadily in recent weeks as much of the country moves past a summer surge. That progress has been especially pronounced in Southern states that saw the highest infection rates.
- There are about 20,000 fewer coronavirus patients hospitalized nationwide as there were at the start of September.
- Nearly 2,000 coronavirus deaths are being reported each day. On Friday, the country surpassed 700,000 total deaths in the pandemic.
- Alaska leads the country by a wide margin in recent cases per capita. The surge there has hit rural and urban areas, leaving hospital systems strained.
- Parts of the Mountain West, including Wyoming and Montana, have reported some of the country's worst outbreaks in recent weeks. But it appears both those states may be starting to stabilize.
- Daily case reports continue to plummet in the South. Florida, which averaged more than 20,000 cases a day during much of August, is

3

Last Time: EDA

4

Tableau



6

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 (/)

7

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)

14

Table Algebra: Operands

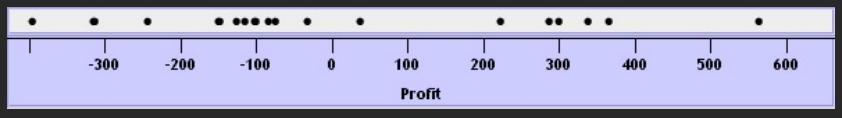
Ordinal fields: interpret domain as a set that partitions table into rows and columns

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])} →



15

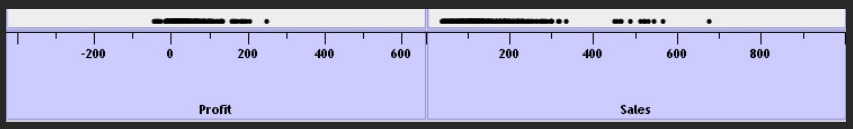
Concatenation (+) Operator

Ordered union of sets

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])}



16

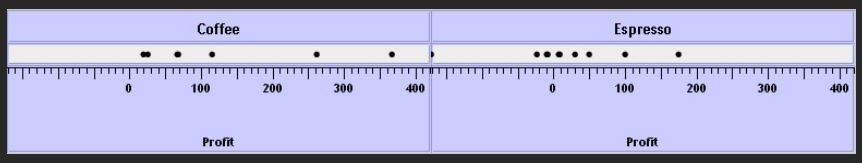
Cross (x) Operator

Cross-product of sets

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 =



17

Nest (/) Operator

Cross-product filtered by existing records

Quarter x Month

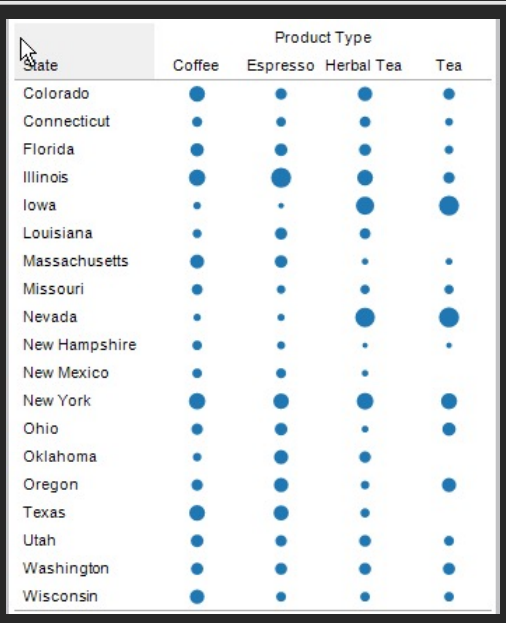
creates 12 entries for each qtr. i.e., (Qtr1, Dec)

Quarter / Month

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

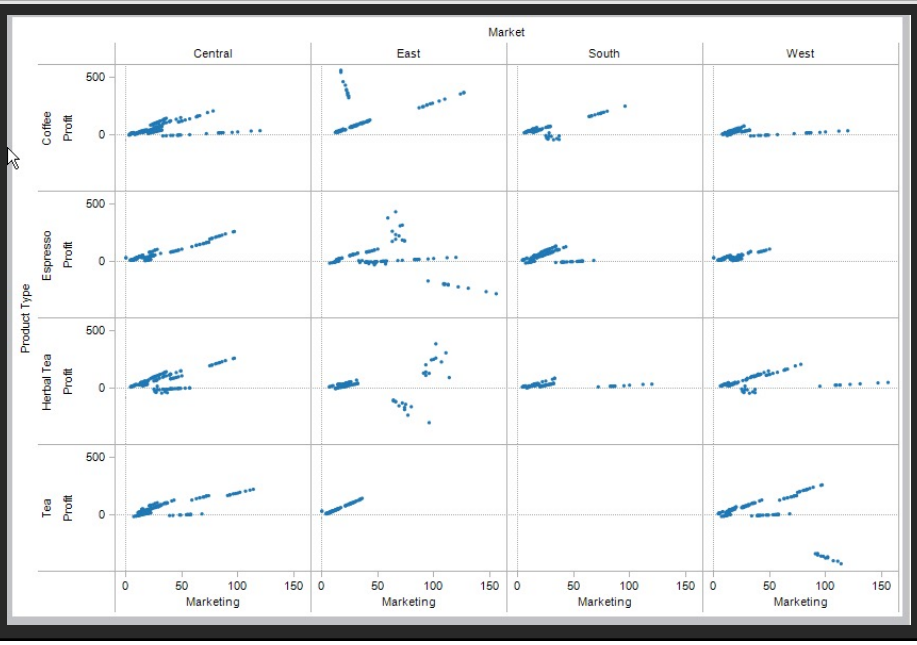
18

Ordinal - Ordinal



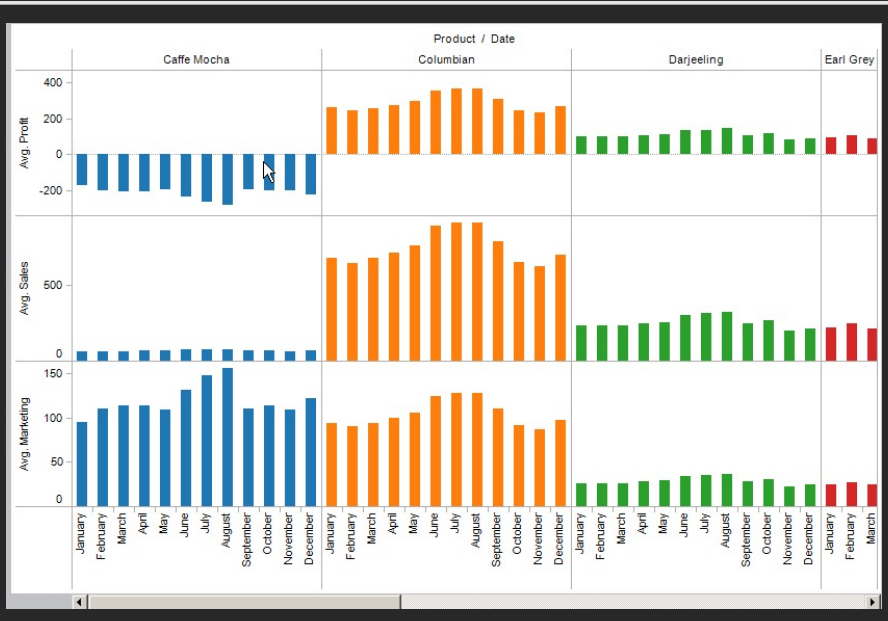
19

Quantitative - Quantitative



20

Ordinal - Quantitative



21

Summary

Exploratory analysis may combine graphical methods, and statistics

Use questions to uncover more questions

Interaction is essential for exploring large multidimensional datasets

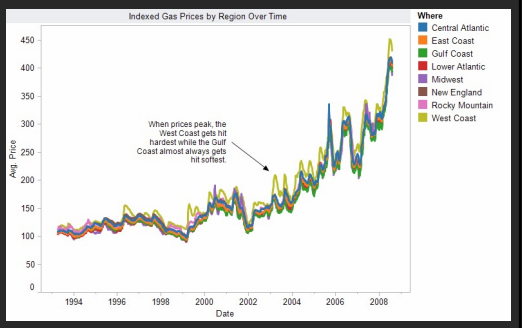
22

Announcements

23

A2: Exploratory Data Analysis

Use Tableau or Vega-Lite to formulate & answer 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

Due before class on Oct 11, 2021

24

Using Space Effectively

27

Topics

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

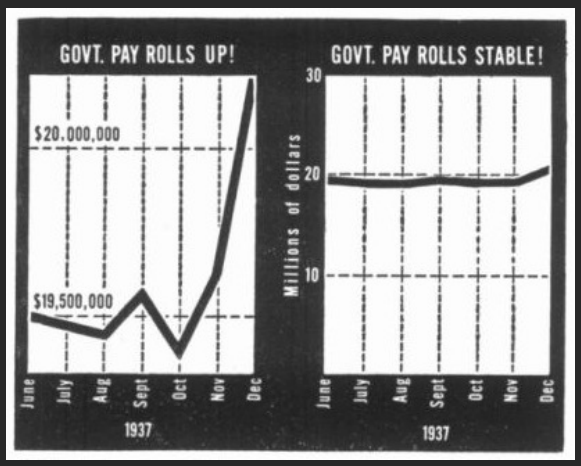
28

Graphs and Lines

29

Effective use of space

Which graph is better?

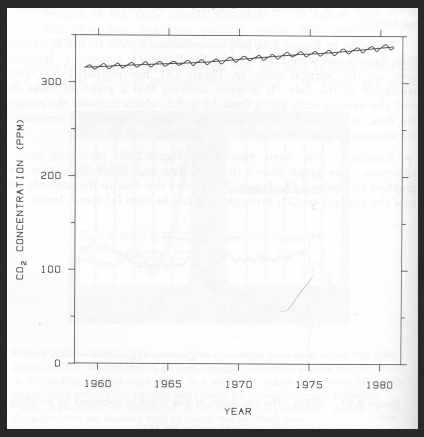


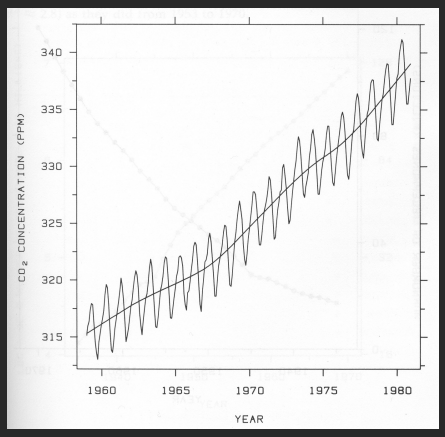
Government payrolls in 1937 [Huff 93]

30

Fill space

Show data with as much resolution as possible
Don't worry about showing zero

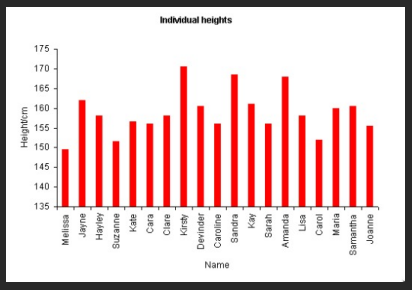




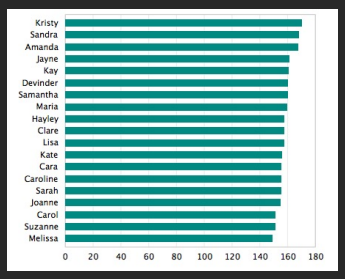
Yearly CO2 concentrations [Cleveland 85]

31

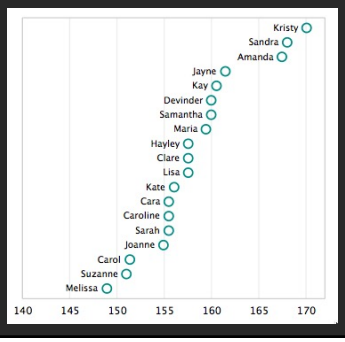
Include zero in axis scale?



Compare Proportions (Q-Ratio)

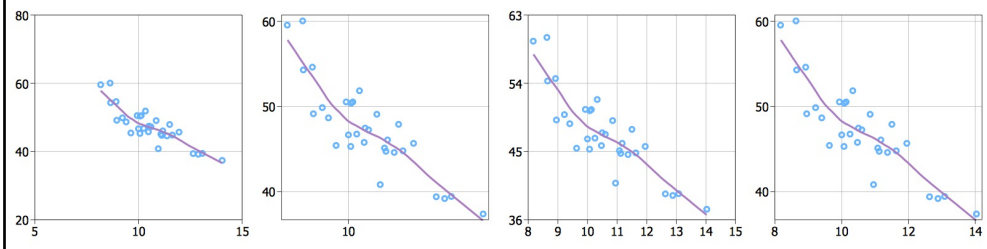


Compare Relative Position (Q-Interval)



32

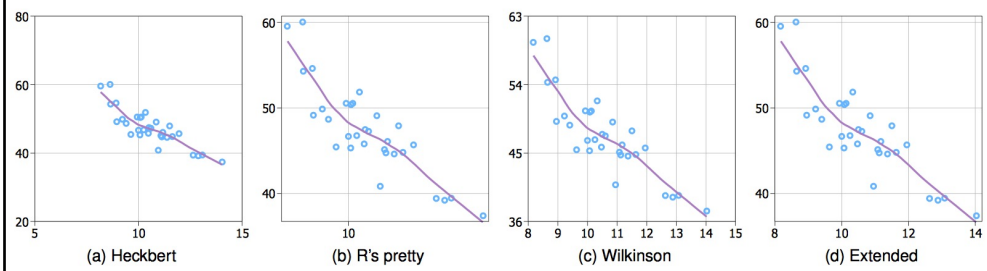
Axis Tick Mark Selection



What are some properties of "good" tick marks?

33

Axis Tick Mark Selection



Simplicity - numbers are multiples of 10, 5, 2

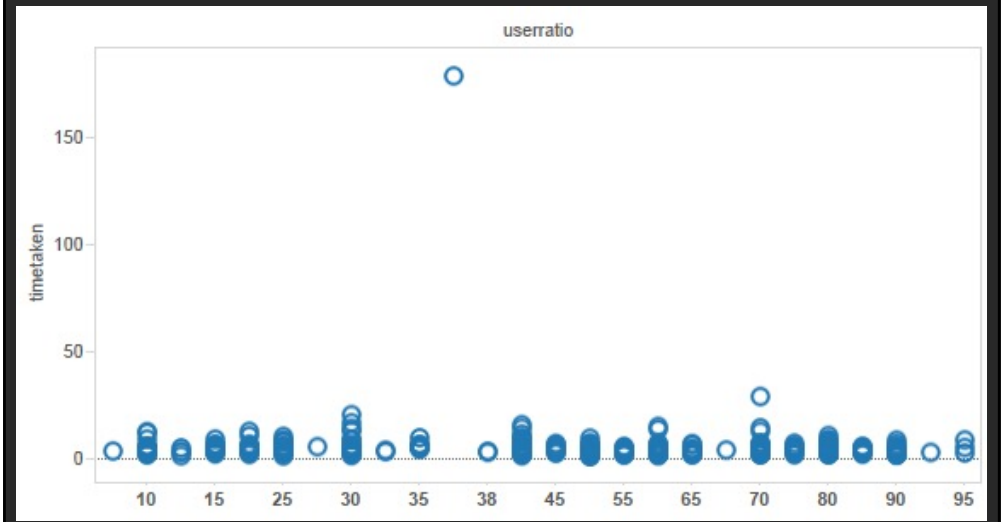
Coverage - ticks near the ends of the data

Density - not too many, nor too few

Legibility - whitespace, horizontal text, size

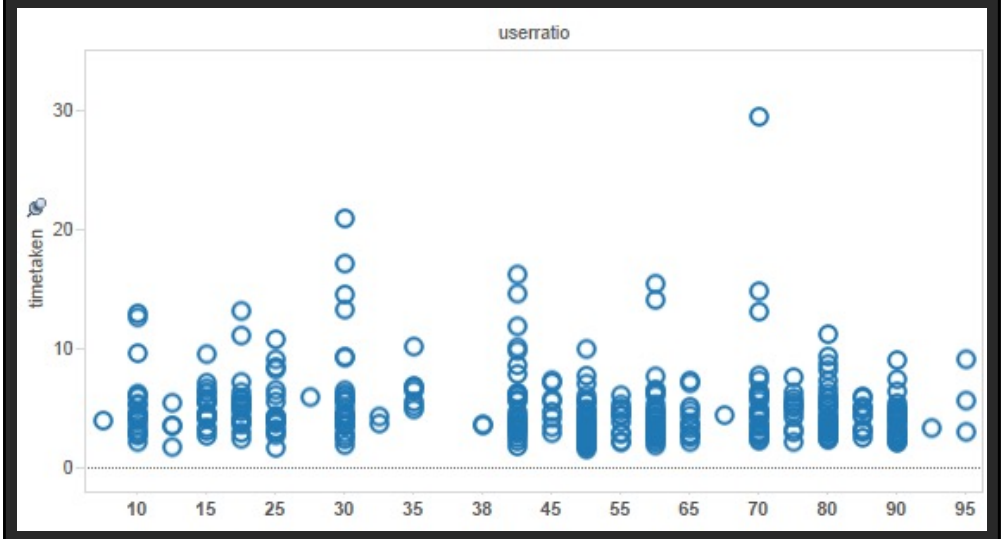
34

How to Scale the Axis?



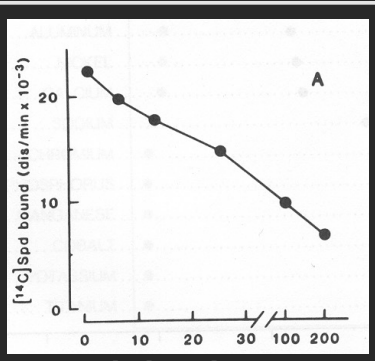
35

One Option: Clip Outliers

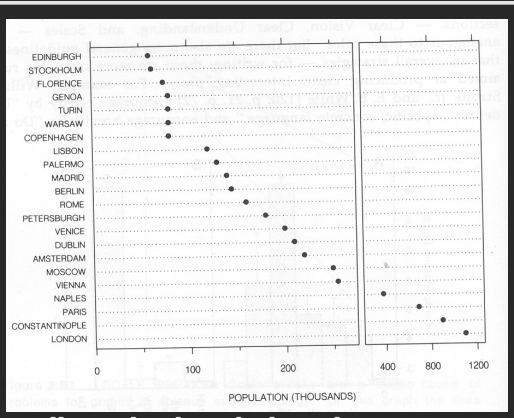


36

Clearly mark scale breaks



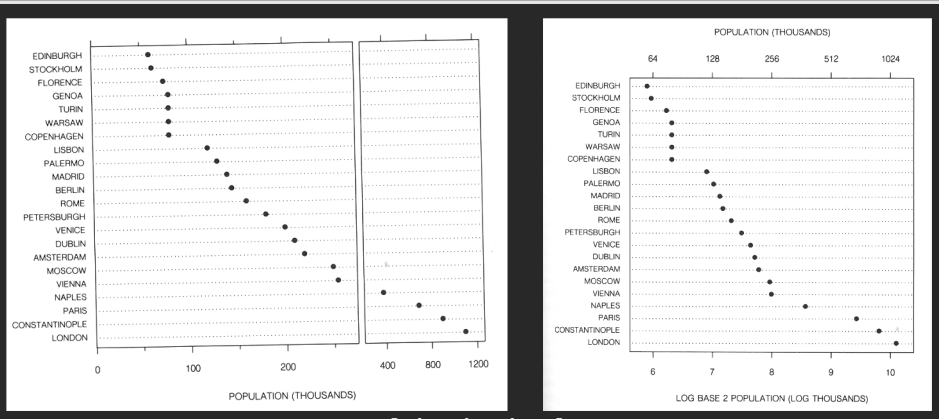
Poor scale break [Cleveland 85]



Well marked scale break [Cleveland 85]

37

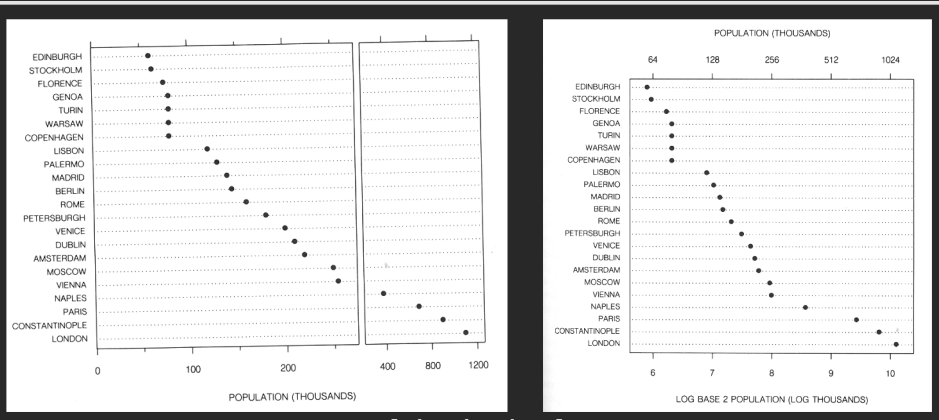
Scale break vs. Log scale



[Cleveland 85]

38

Scale break vs. Log scale



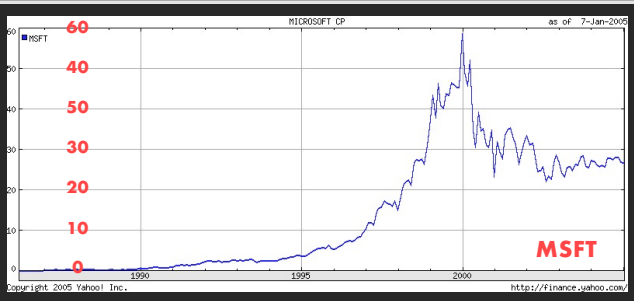
[Cleveland 85]

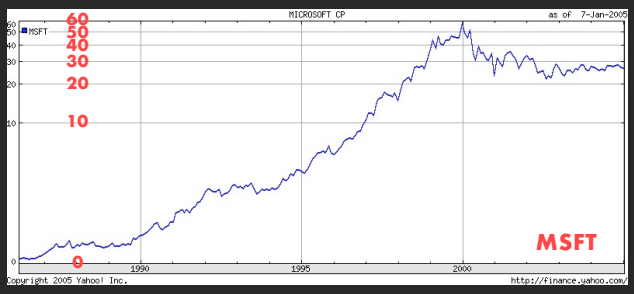
Both increase visual resolution

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

39

Linear scale vs. Log scale



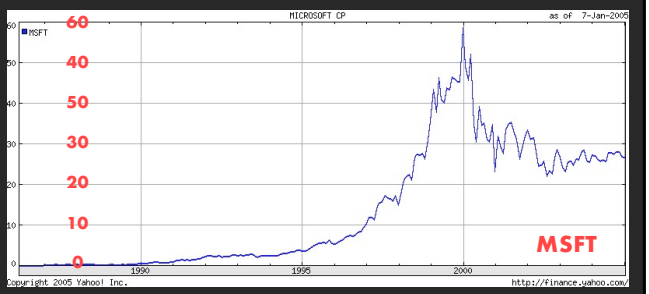


40

Linear scale vs. Log scale

Linear scale

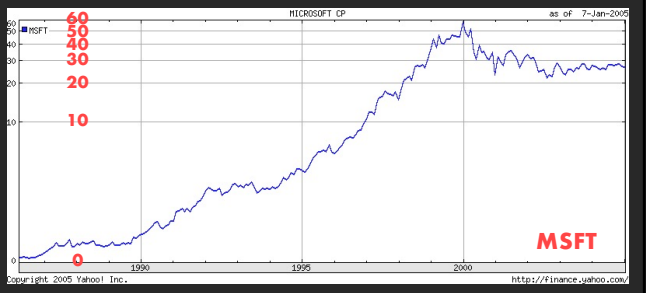
- Absolute change



Log scale

- Small fluctuations
- Percent change

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



41

Log scales

Logarithms turn multiplication into addition

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

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

42

When to apply log scale?

Address data skew (e.g., long tails, outliers)

Enables comparison 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?

43

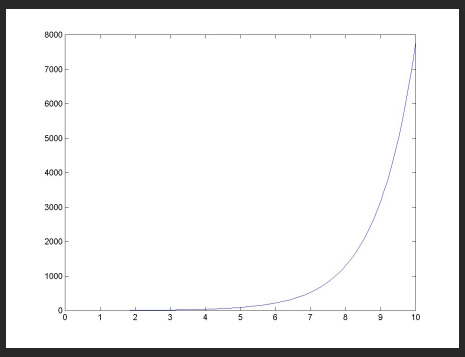
Semilog graph: Exponential growth

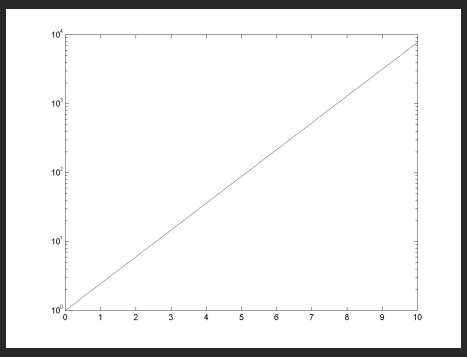
Exponential functions ($y = ka^{mx}$) transform into lines

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

Intercept: $\log(k)$

Slope: $\log(a)m$





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

44

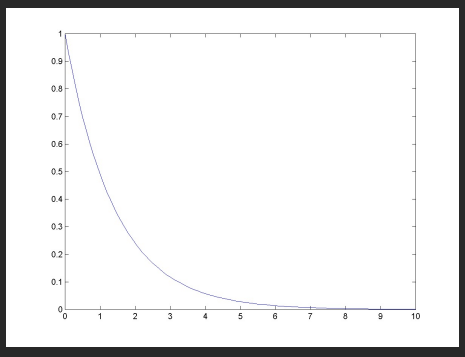
Semilog graph: Exponential decay

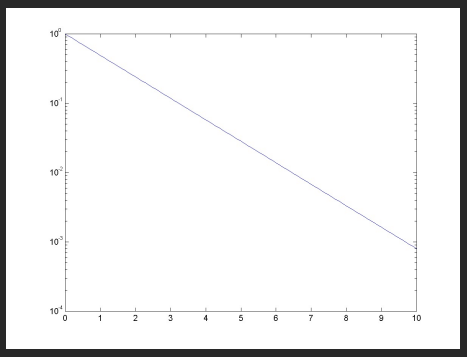
Exponential functions ($y = ka^{mx}$) transform into lines

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

Intercept: $\log(k)$

Slope: $\log(a)m$



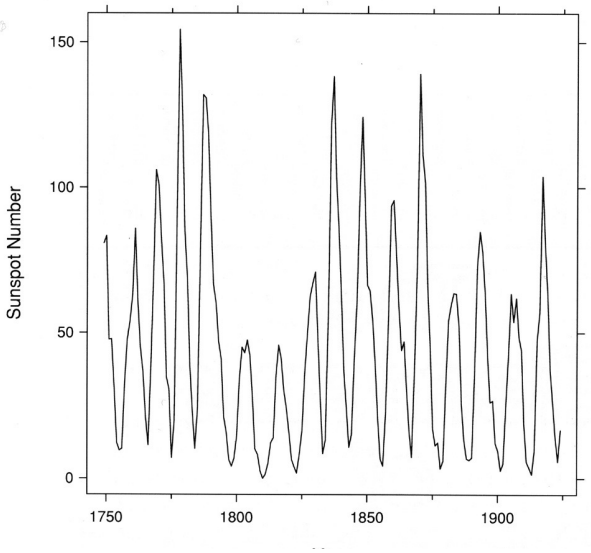


$y = 0.5^{2x}$, **slope in semilog space:** $\log(0.5) * 2 = -0.602$

45

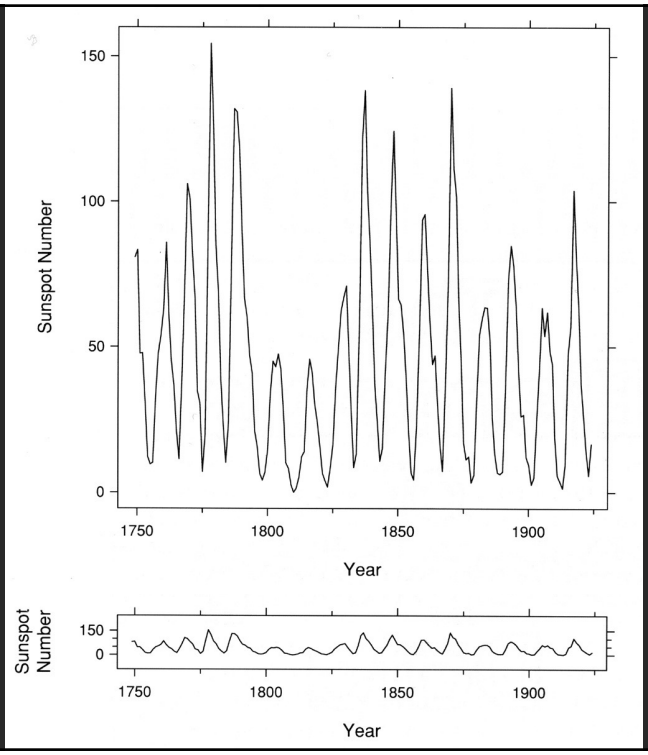
Selecting Aspect Ratio

49



William S. Cleveland
The Elements of Graphing Data

50

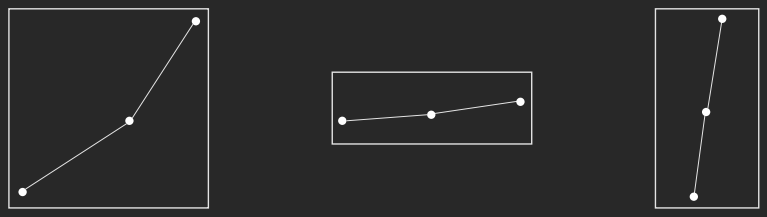


William S. Cleveland
The Elements of Graphing Data

51

Banking to 45° [Cleveland]

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

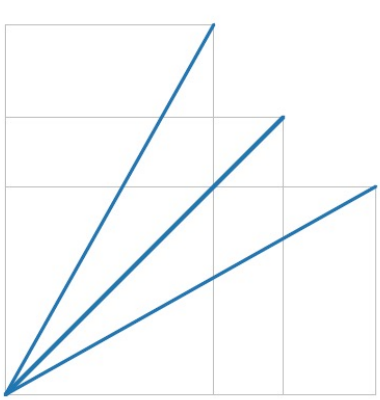


Two line segments are maximally discriminable when the absolute angle between them is 45°

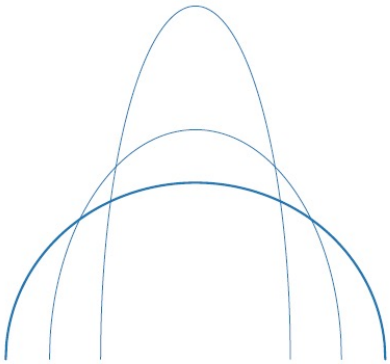
Method: Optimize the *aspect ratio* such that the average absolute angle between all segments is 45°

52

An alternate approach:
Minimize arc length (hold area constant)



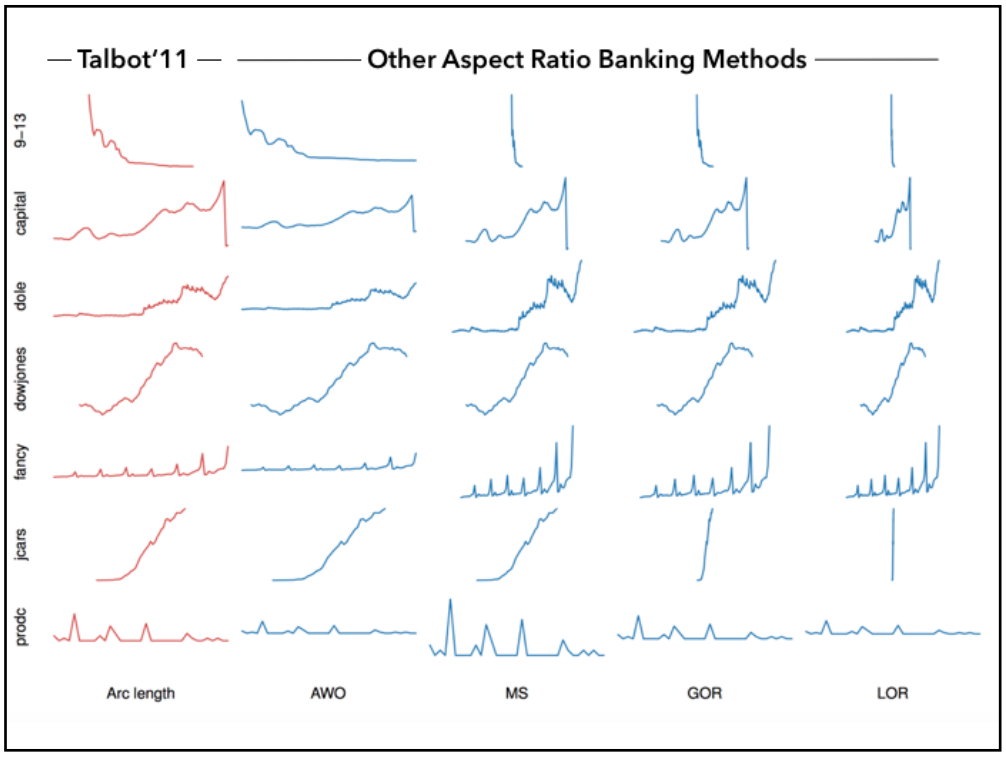
Straight line -> 45 deg



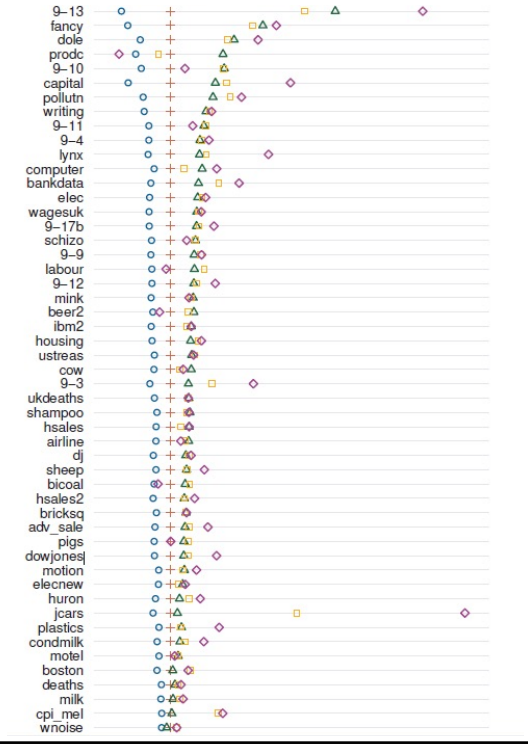
Ellipse -> Circle

[Talbot et al, 2011]

58



59

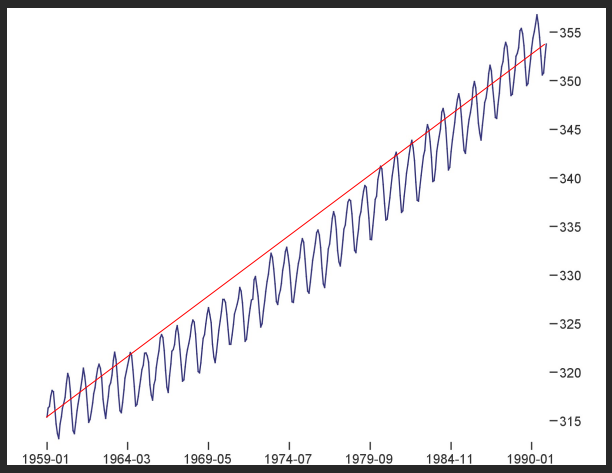


Good Compromise

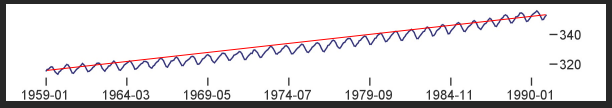
Arc-length banking produces aspect ratios in-between those produced by other methods.

[Talbot et al, 2011]

63



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]

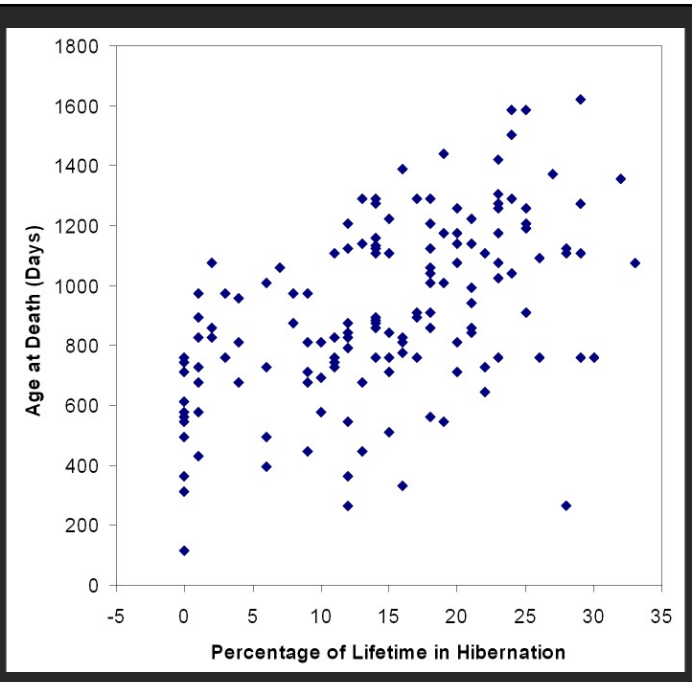
CO_2 Measurements

William S. Cleveland
Visualizing Data

67

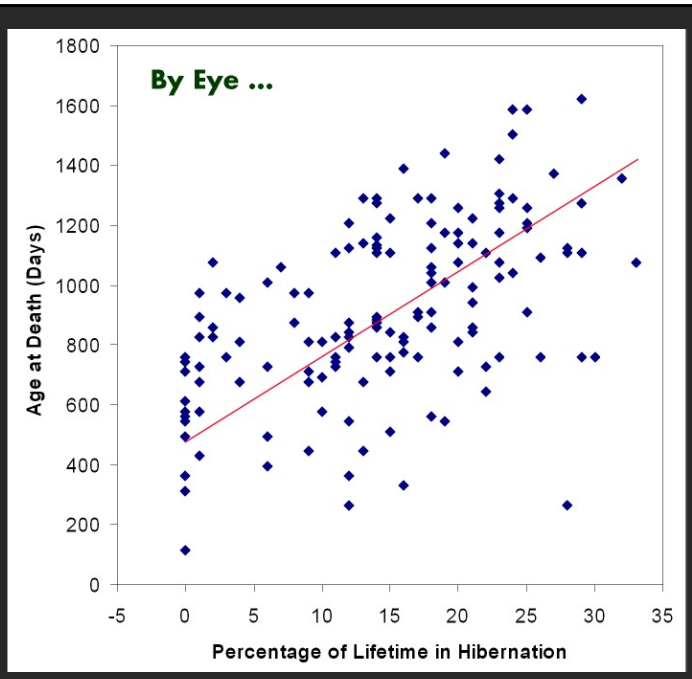
Fitting the Data

79



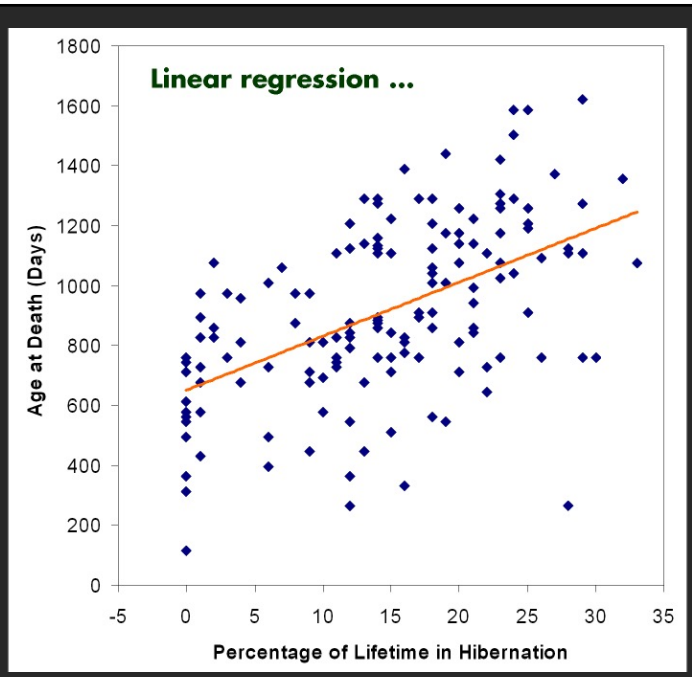
[The Elements of Graphing Data. Cleveland 94]

80



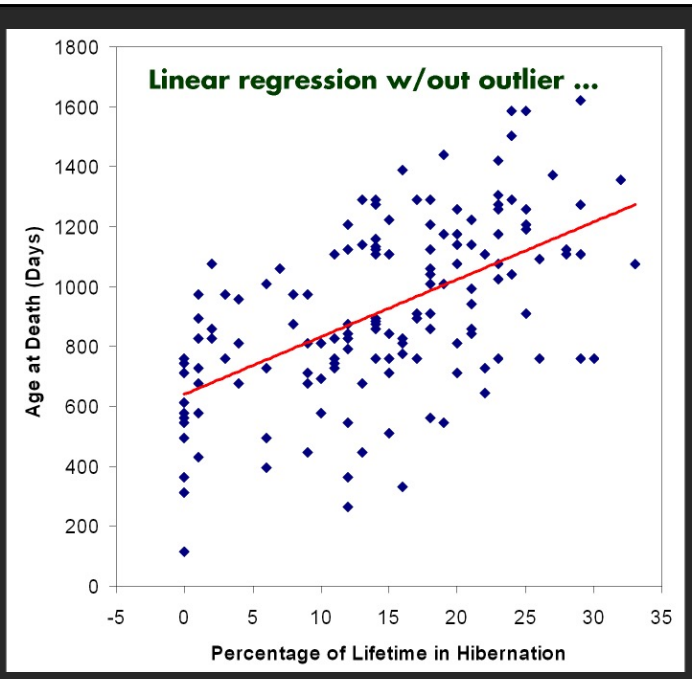
[The Elements of Graphing Data. Cleveland 94]

81



[The Elements of Graphing Data. Cleveland 94]

82

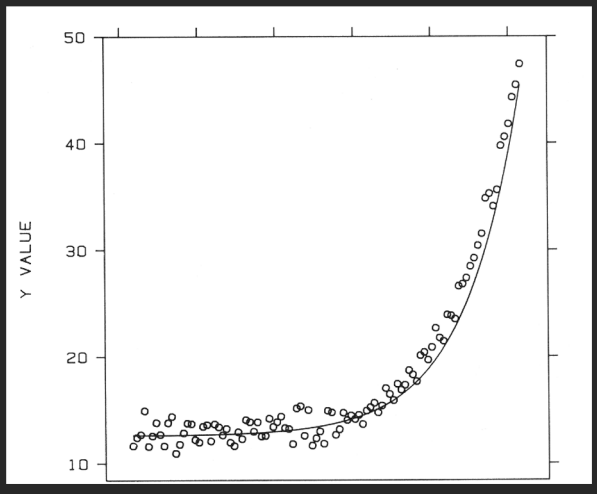


[The Elements of Graphing Data. Cleveland 94]

83

Transforming data

How well does curve fit data?



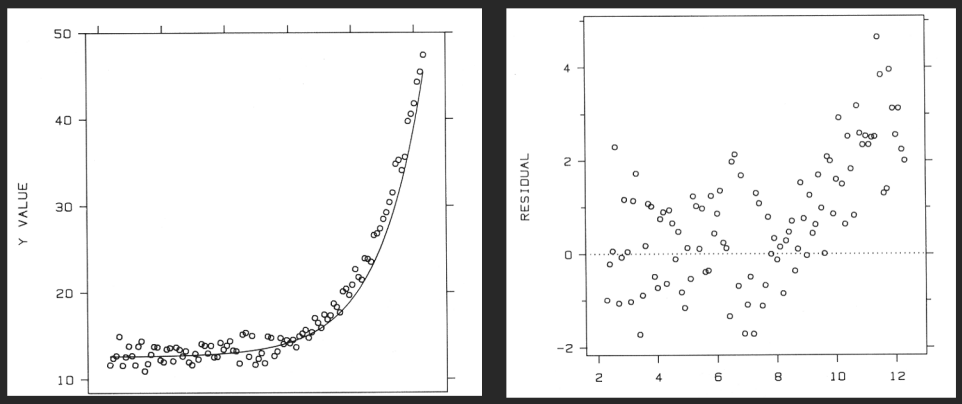
[Cleveland 85]

84

Transforming data

Residual graph

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



[Cleveland 85]

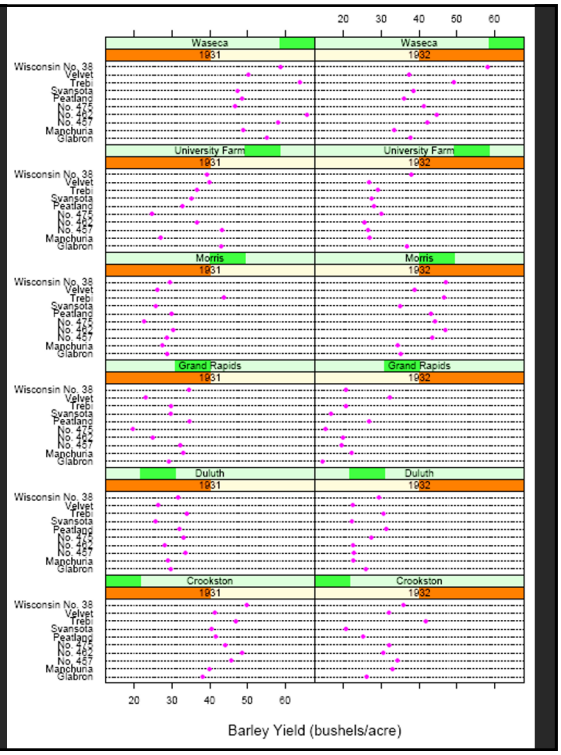
85

Sorting

93

Trellis

[Becker, Cleveland, and Shyu 96]



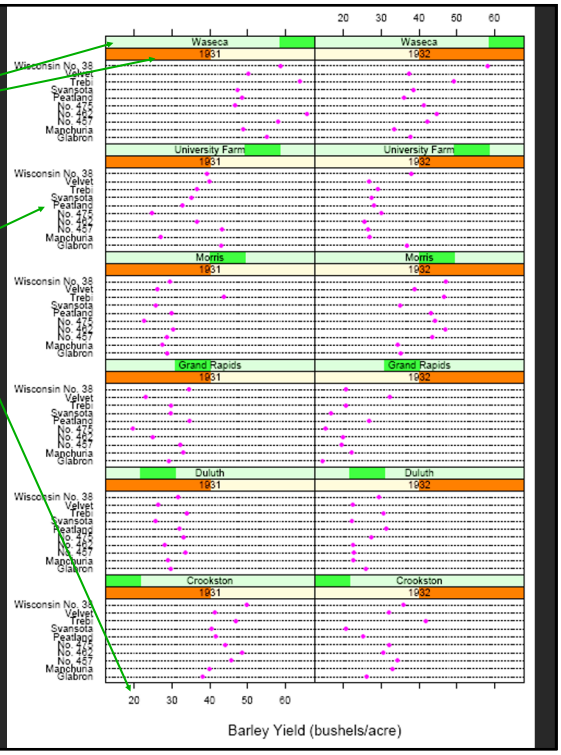
94

Trellis

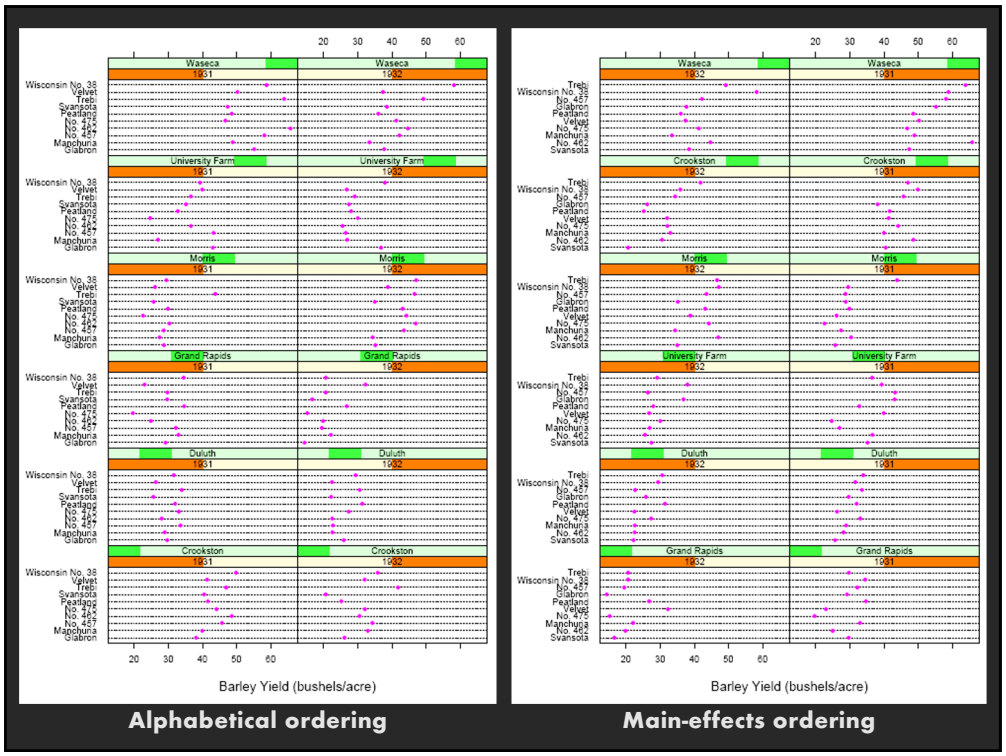
[Becker, Cleveland, and Shyu 96]

Condition variables
location, year

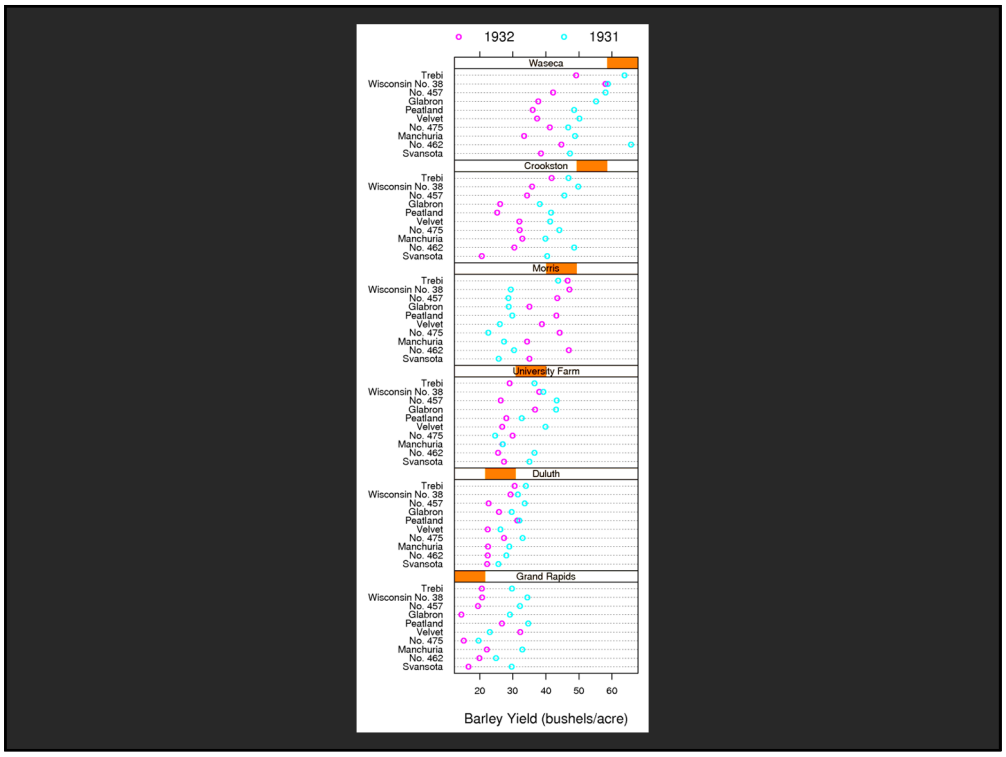
Panel variables
type, yield



95


Waseca
1931
1932
University Farm
Morris
Grand Rapids
Duluth
Crookston
Wisconsin No. 38
Velvet
Trebi
Svansota
Peatland
No. 475
No. 462
No. 457
Manchuria
Glabron
20
30
40
50
60
Barley Yield (bushels/acre)
Alphabetical ordering
Main-effects ordering

96


1932
1931
Waseca
Crookston
Morris
University Farm
Duluth
Grand Rapids
Trebi
Wisconsin No. 38
No. 457
Glabron
Peatland
Velvet
No. 475
Manchuria
No. 462
Svansota
20
30
40
50
60
Barley Yield (bushels/acre)

97