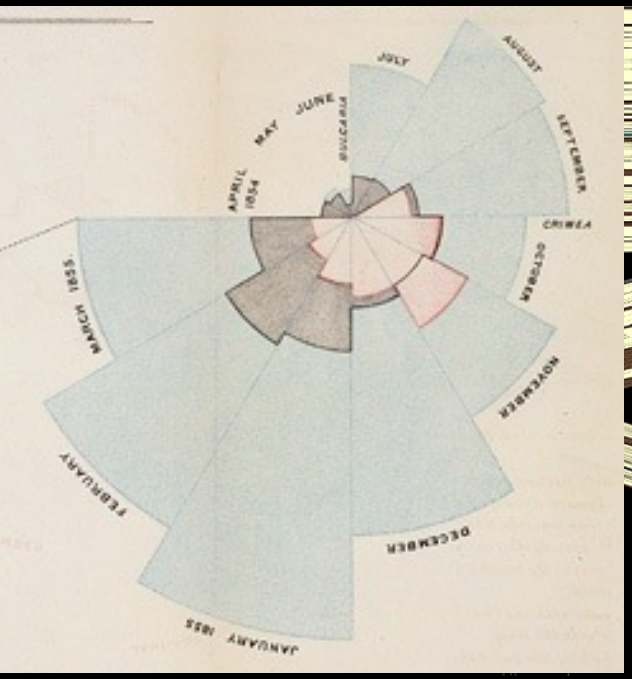
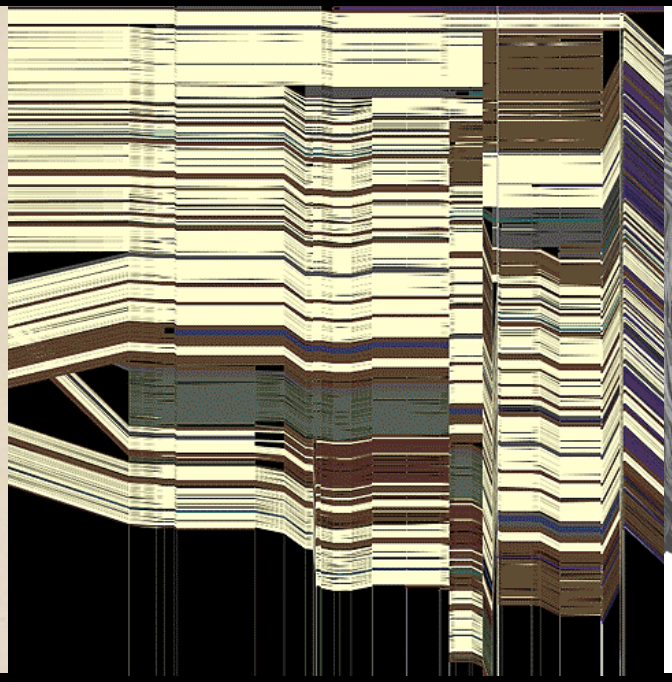
CSE 412 - Intro to Data Visualization

Visual Encoding Design





Jane Hoffswell University of Washington

A Design Space of Visual Encodings

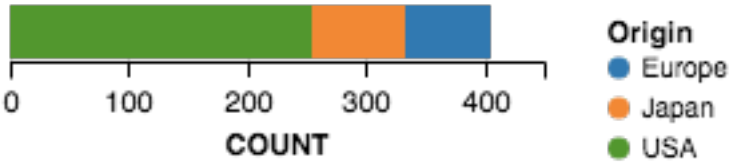
Mapping Data to Visual Variables

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, …*).

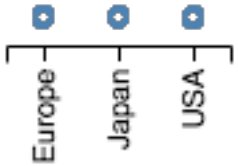
Additional concerns include choosing appropriate **encoding parameters** (*log scale, sorting, …*) and **data transformations** (*bin, group, aggregate, …*).

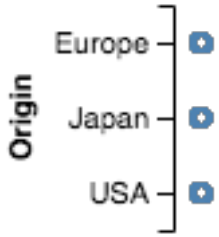
These options define a large combinatorial space, containing both useful and questionable charts!

1D: Nominal

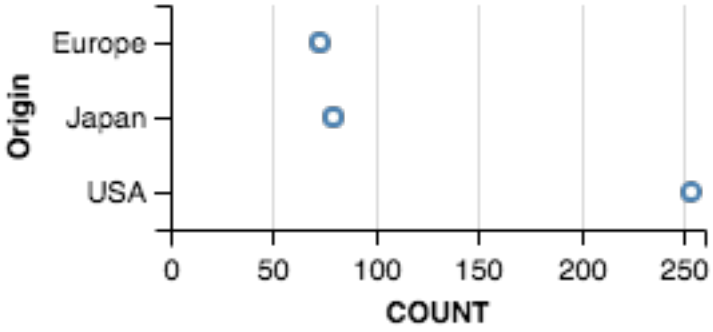


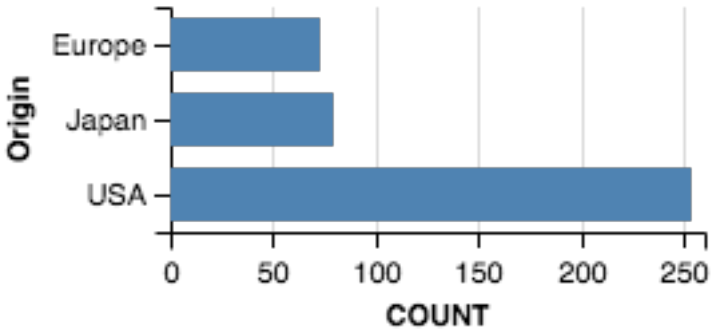
Raw

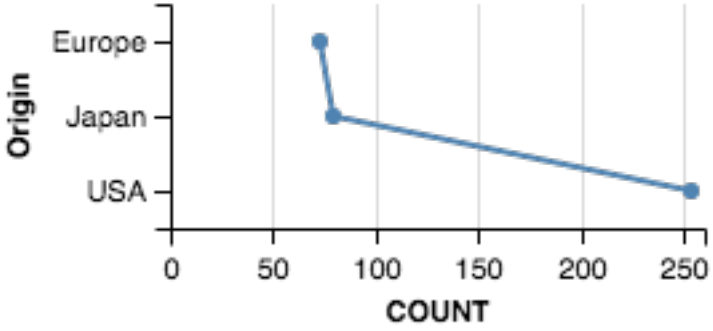


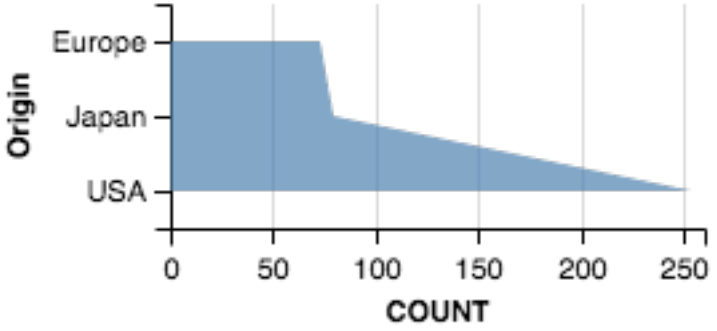


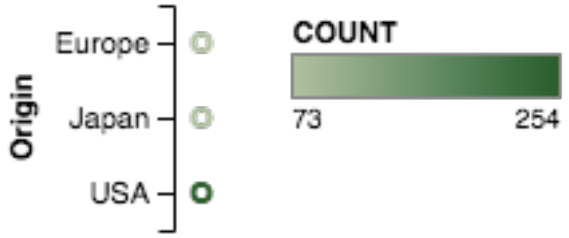
Aggregate (Count)

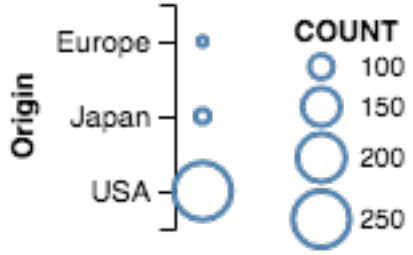








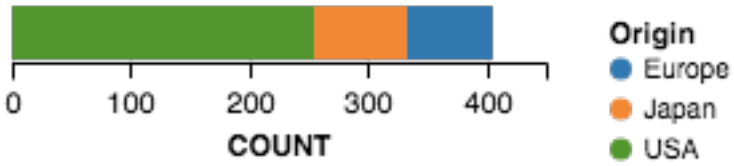


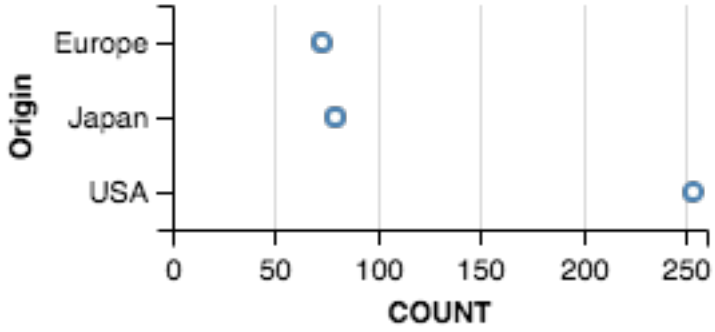


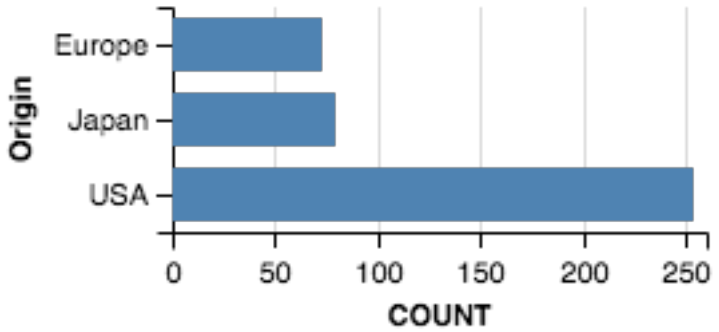
Expressive?

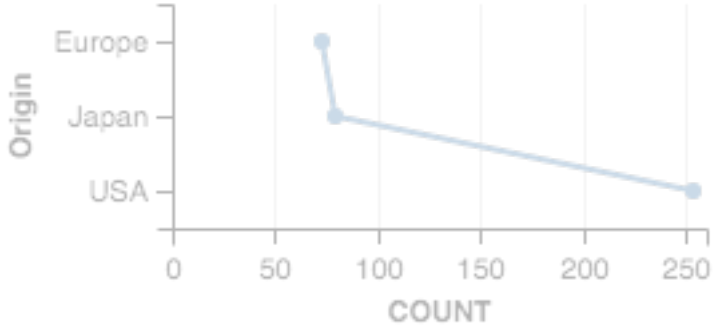
Raw

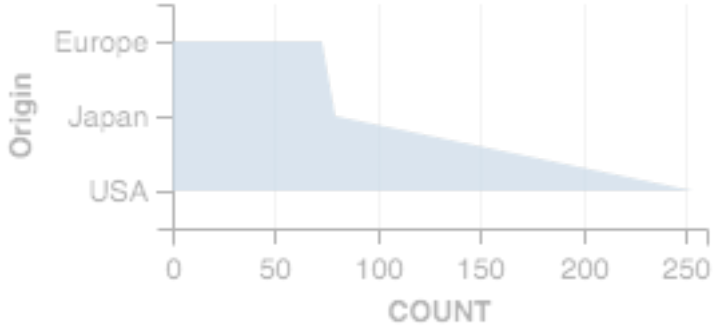
Aggregate (Count)

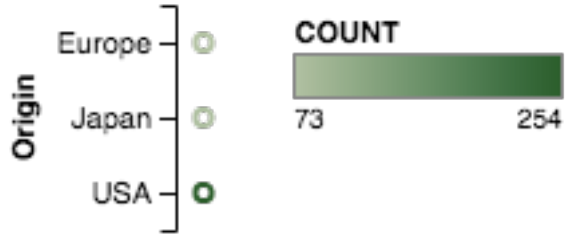


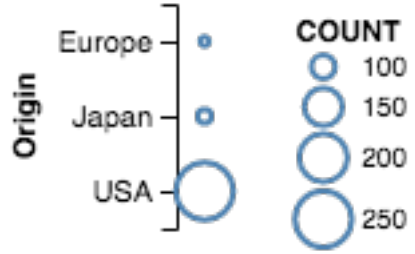






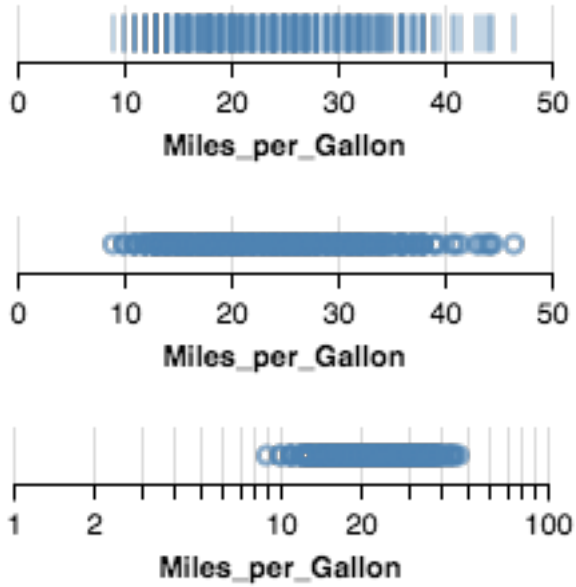


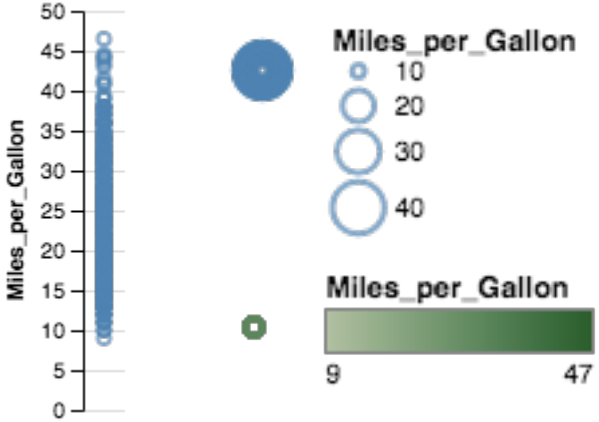




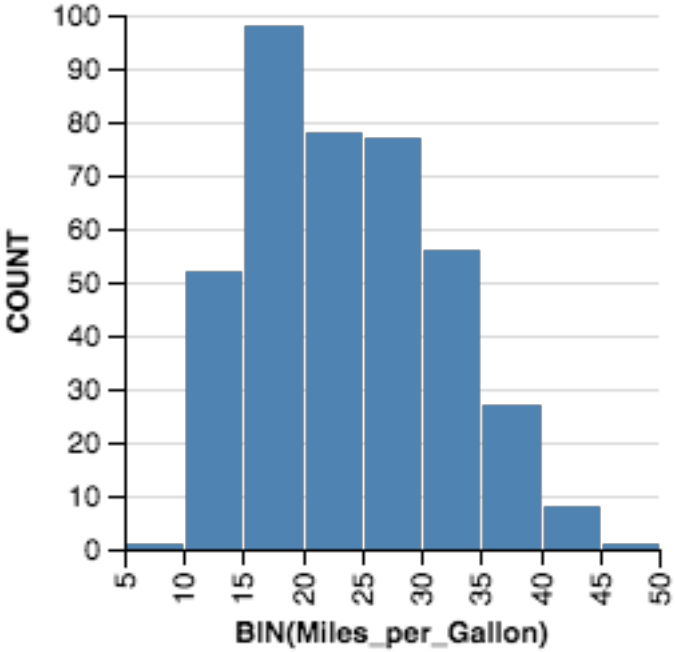
1D: Quantitative

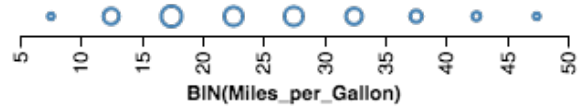
Raw





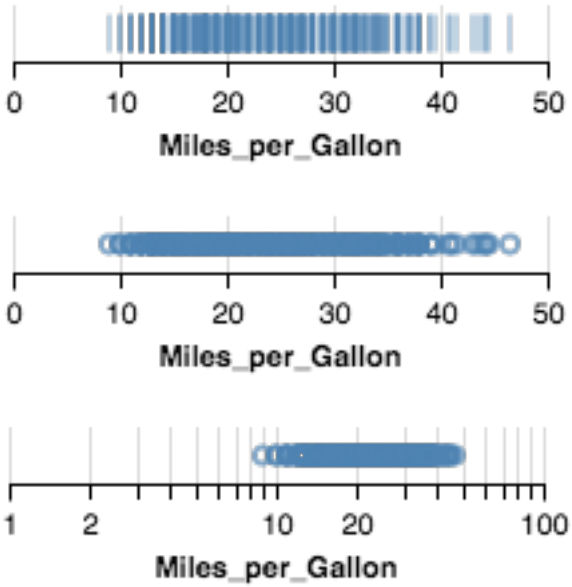
Aggregate (Count)

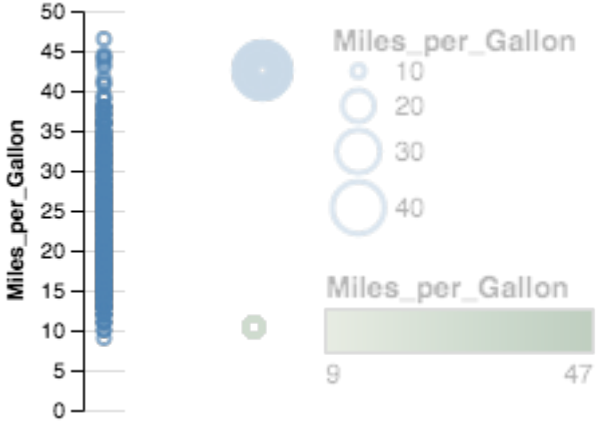




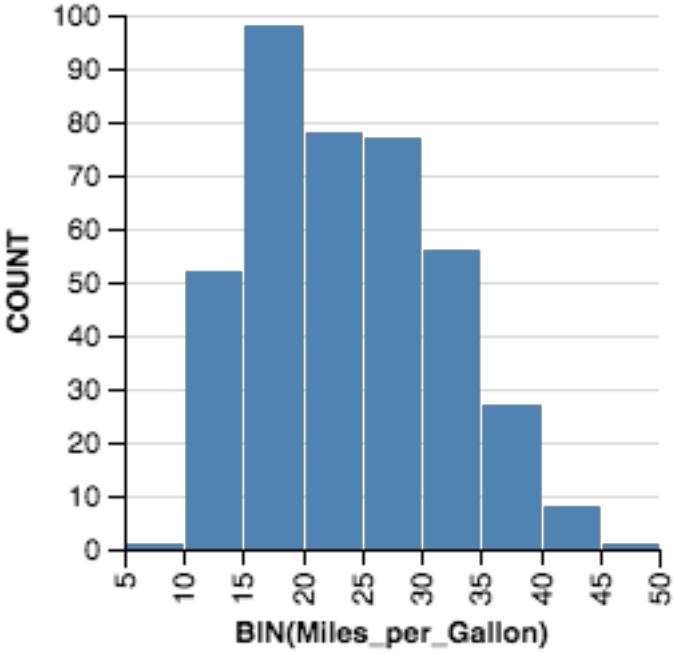
Expressive?

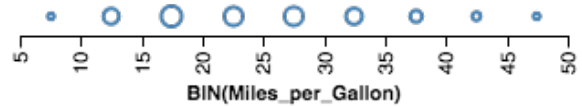
Raw

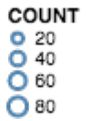




Aggregate (Count)

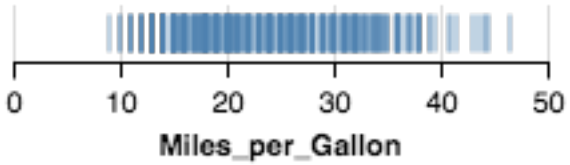


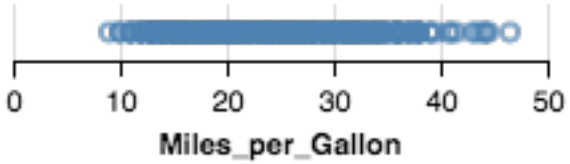


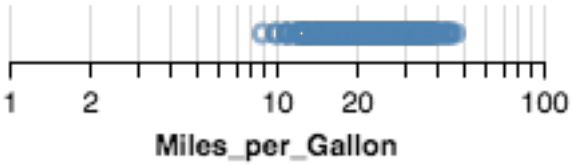


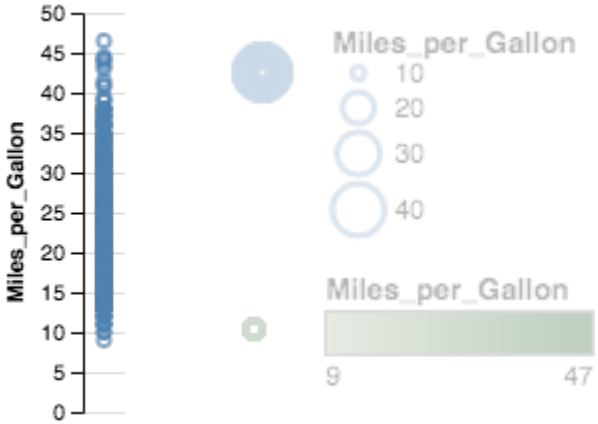
Effective?

Raw

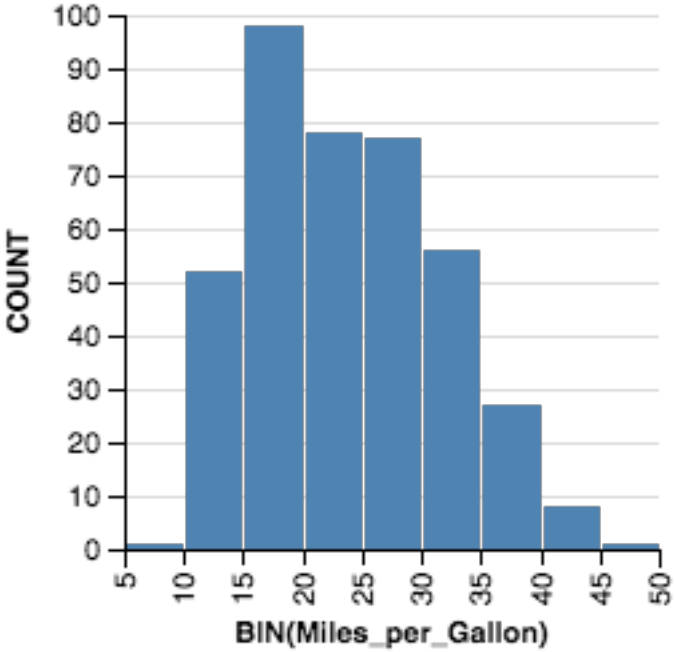


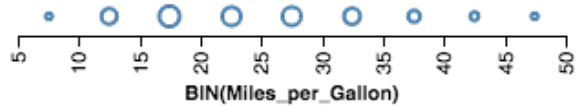


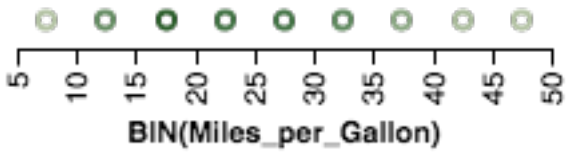




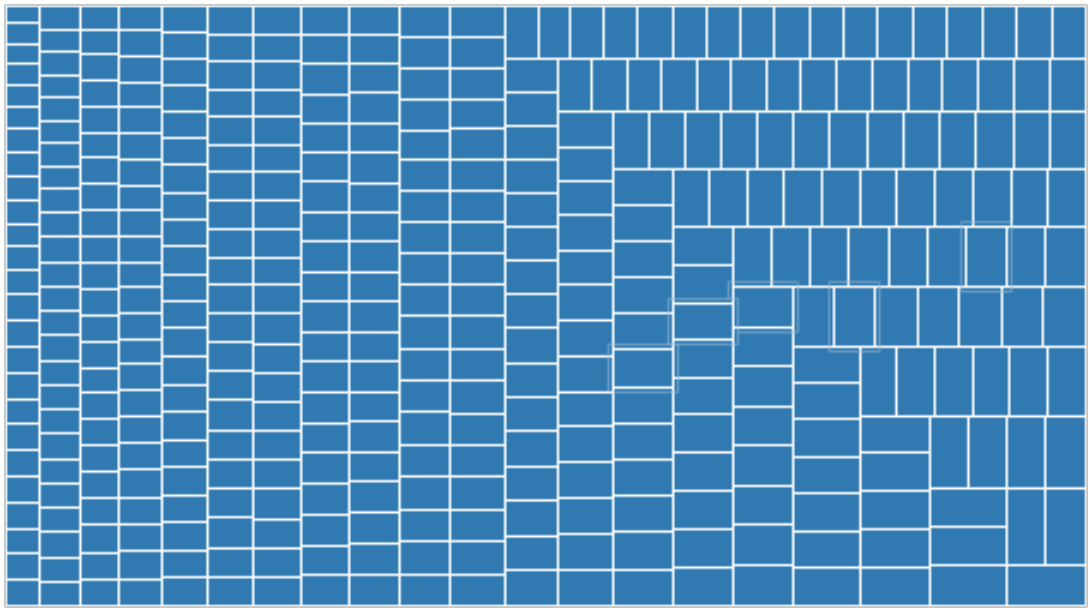
Aggregate (Count)



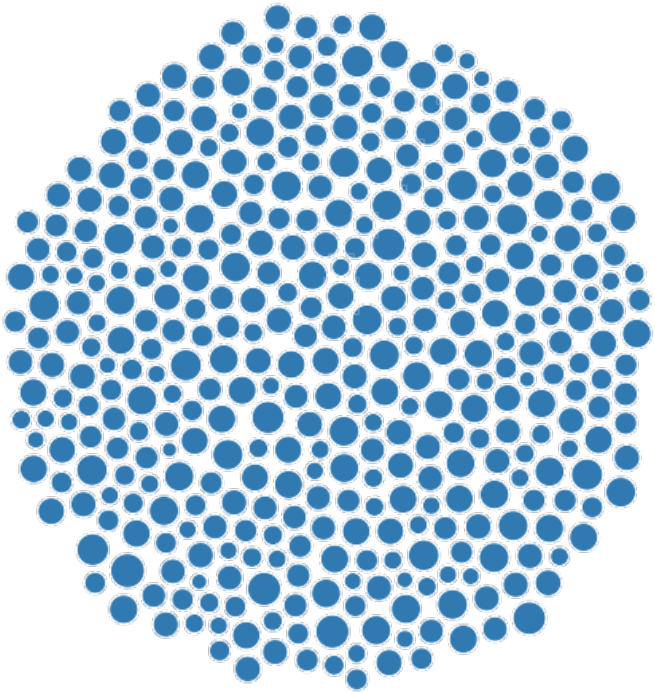




Raw (with Layout Algorithm)

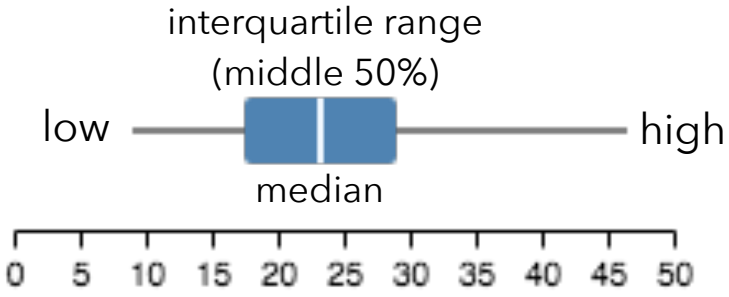


Treemap

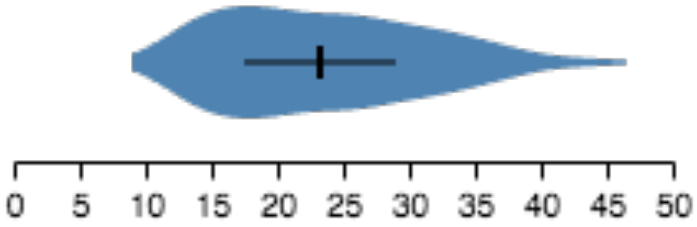


Bubble Chart

Aggregate (Distributions)



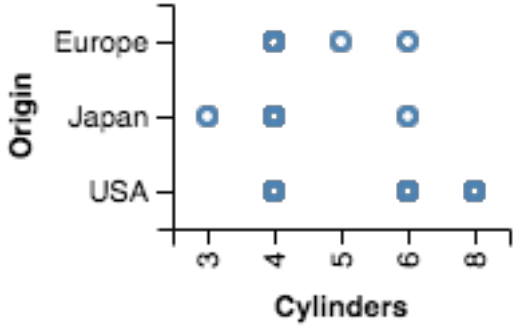
Box Plot

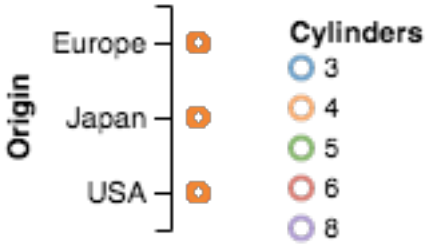


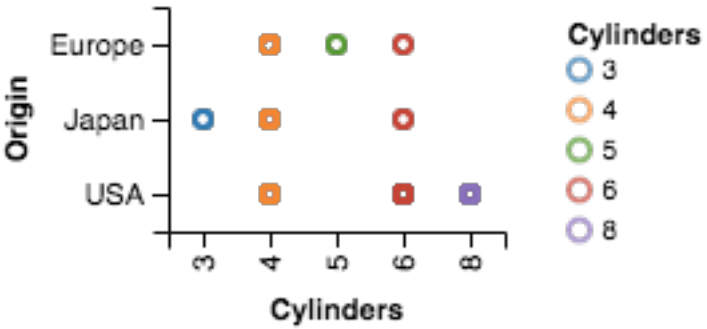
Violin Plot

2D: Nominal x Nominal

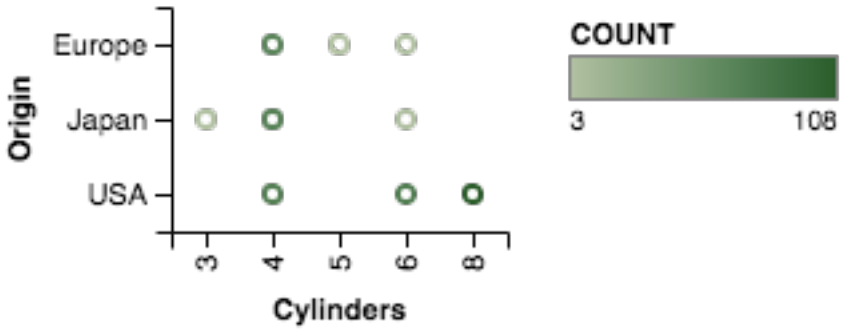
Raw

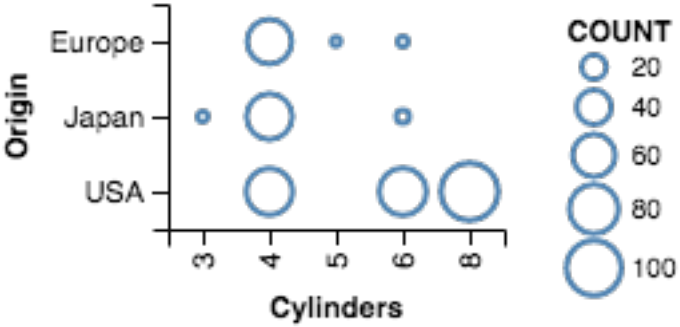


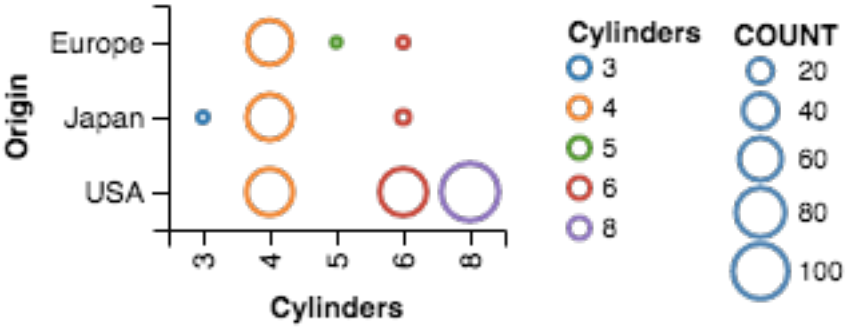




Aggregate (Count)

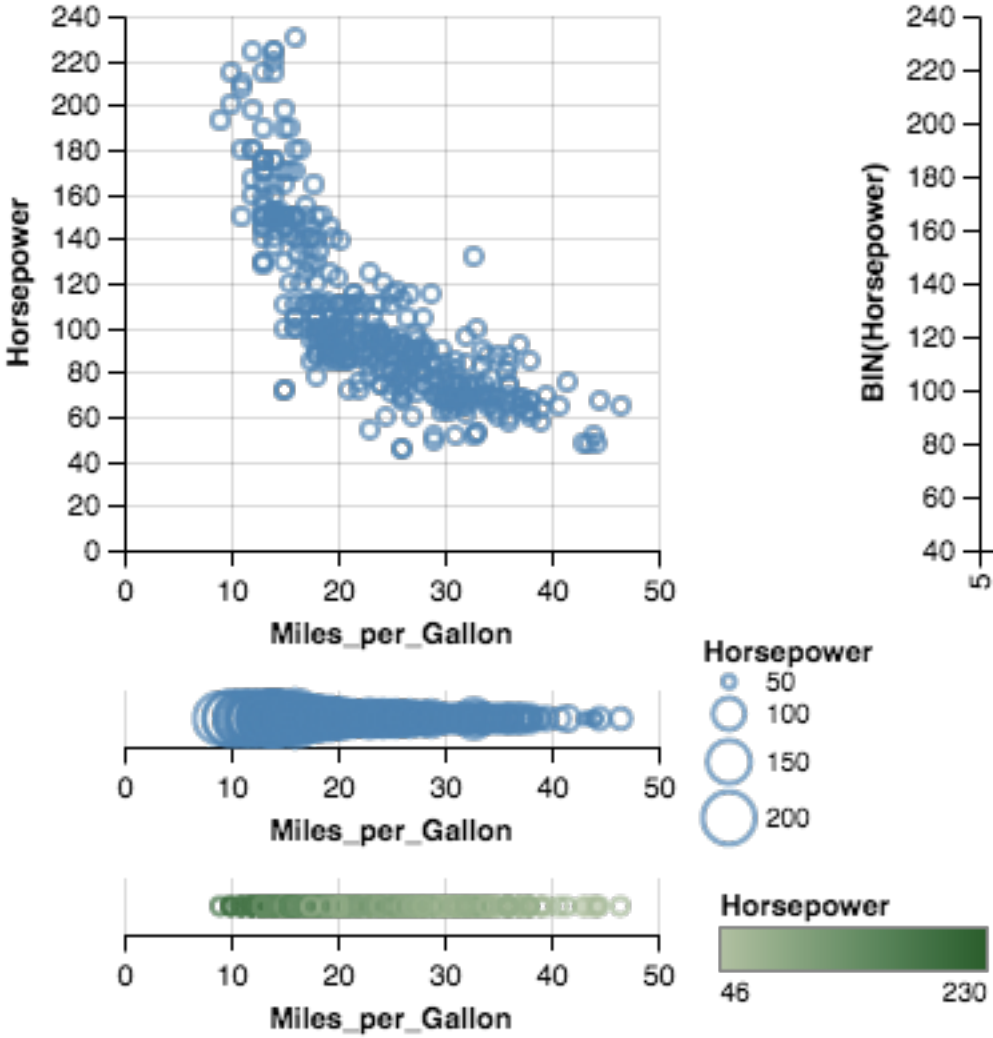




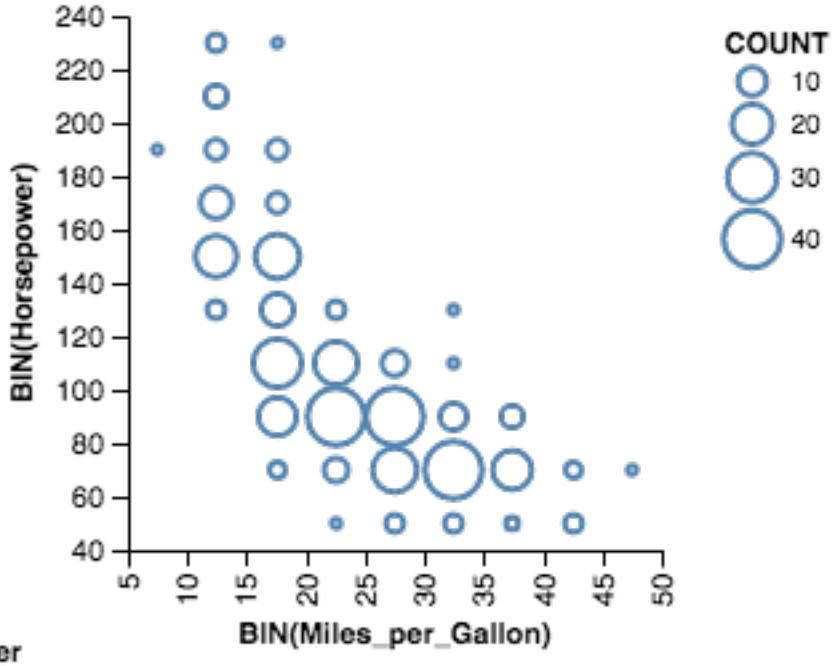


2D: Quantitative x Quantitative

Raw

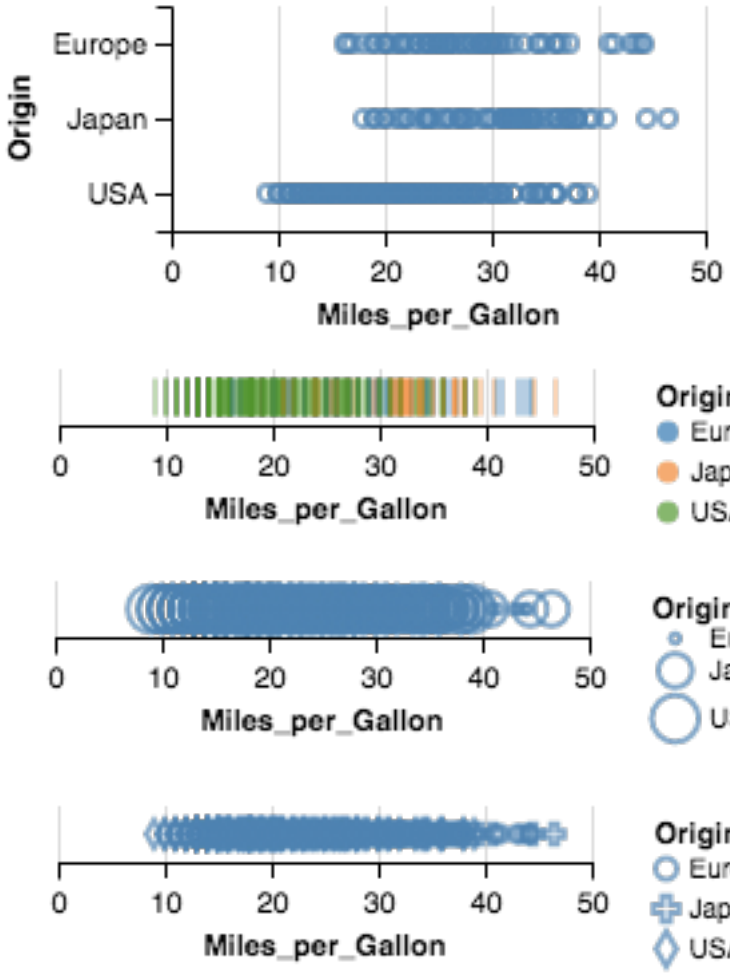


Aggregate (Count)

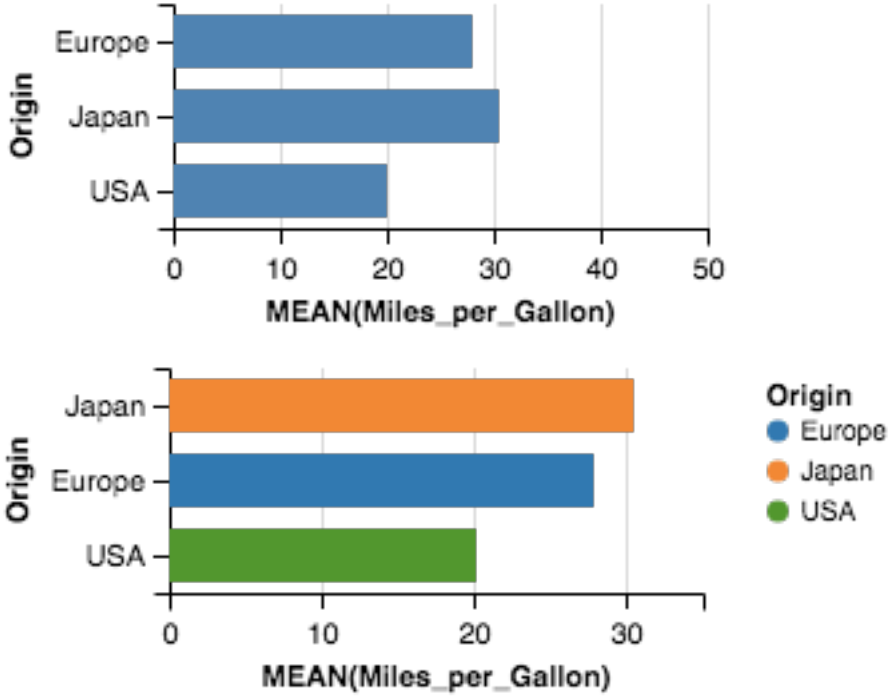


2D: Nominal x Quantitative

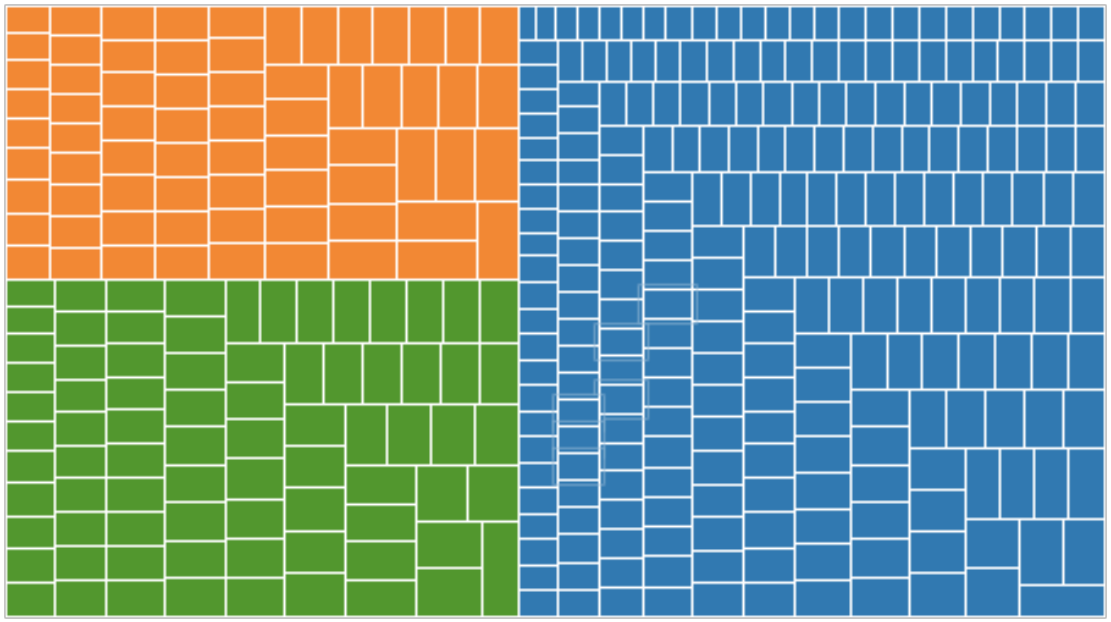
Raw



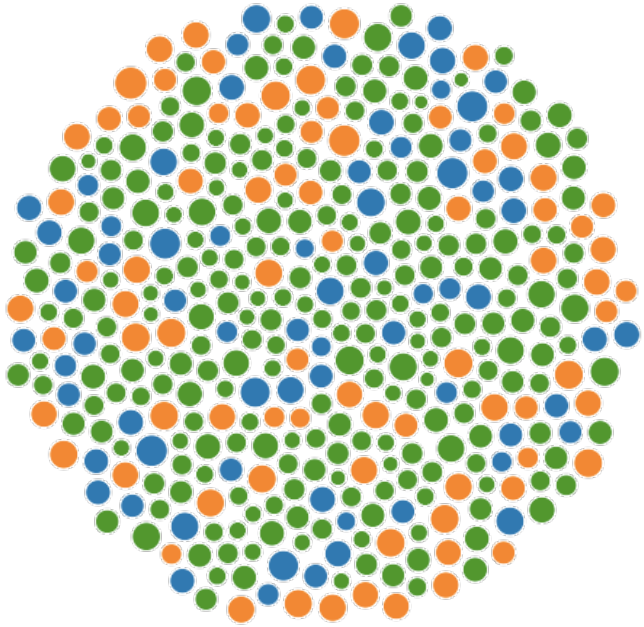
Aggregate (Mean)



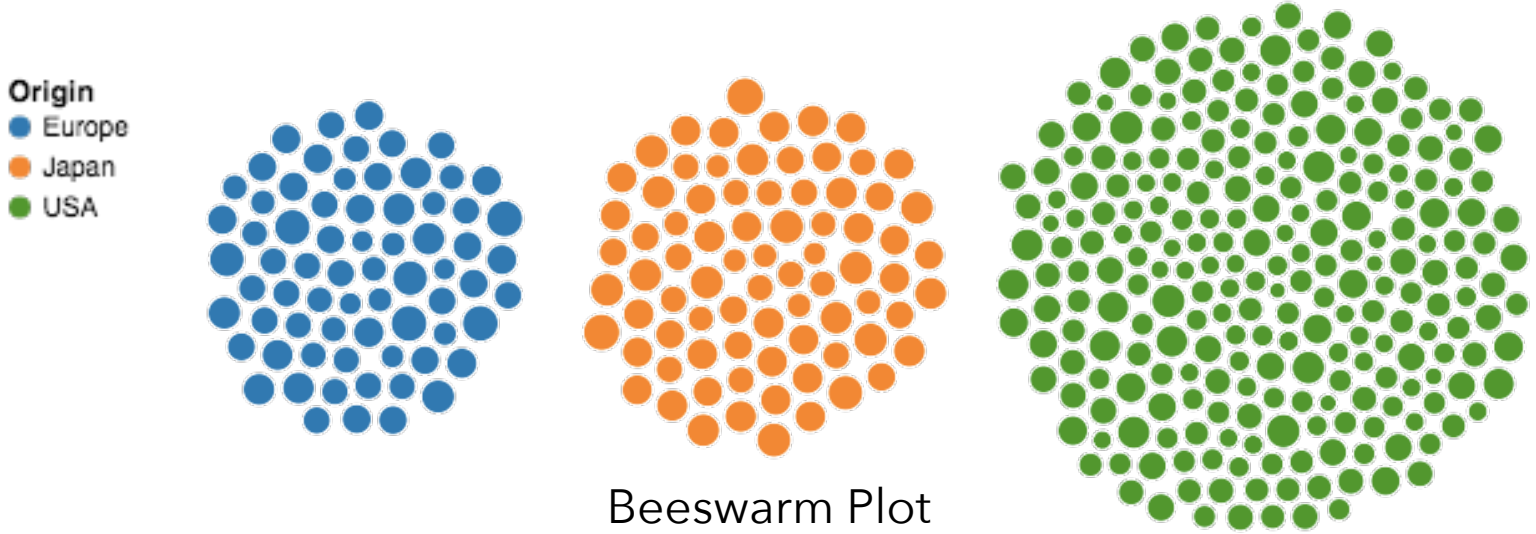
Raw (with Layout Algorithm)



Treemap



Bubble Chart



Beeswarm Plot

3D and Higher

Two variables [x,y]

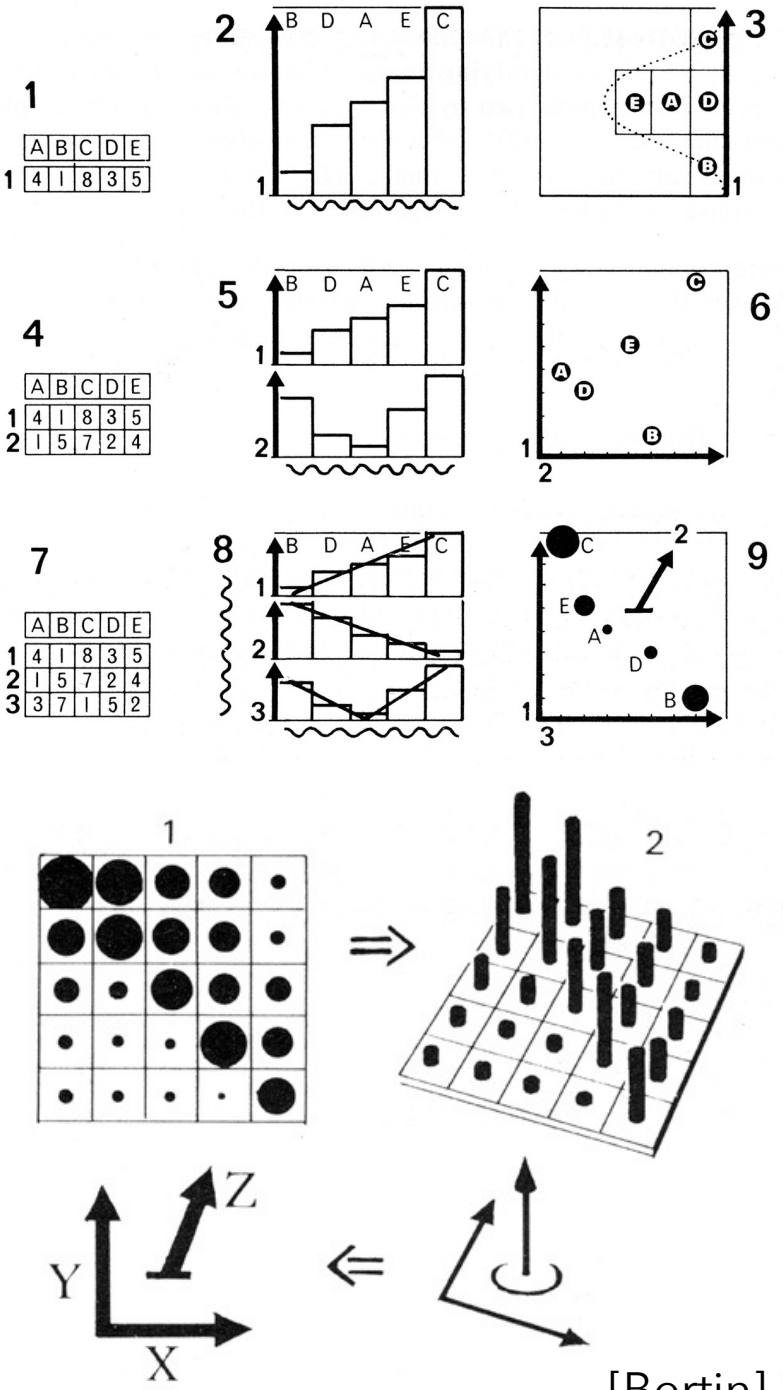
Can map to 2D points.

Scatterplots, maps, …

Third variable [z]

Often use one of size, color, opacity, shape, *etc*. Or, one can further partition space.

What about 3D rendering?



[Bertin]

Other Visual Encoding Channels?

Encoding Effectiveness

Effectiveness Rankings [Mackinlay 86]

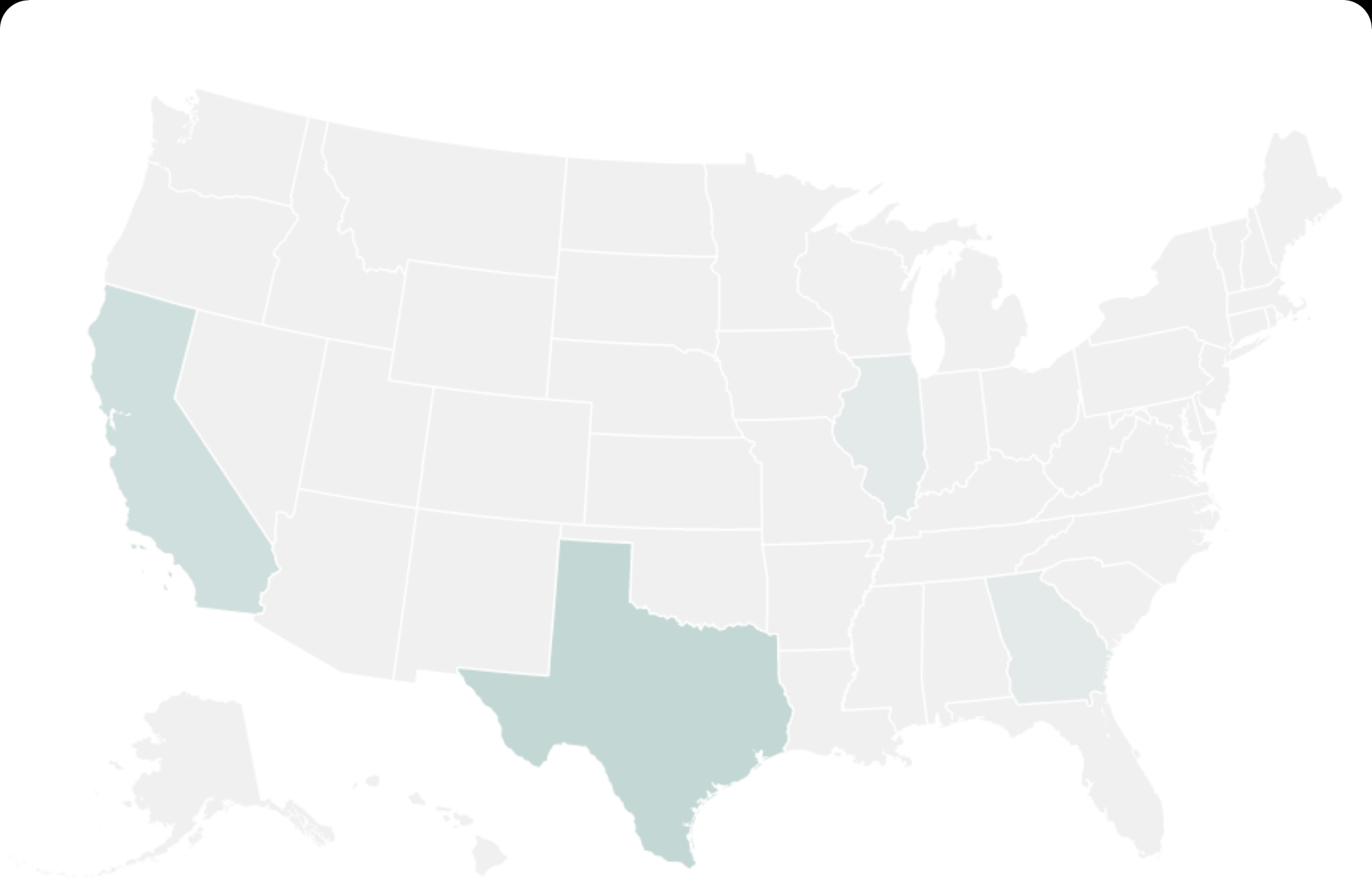
QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE	ORDINAL	NOMINAL
Position	**Position**	**Position**
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Effectiveness Rankings [Mackinlay 86]

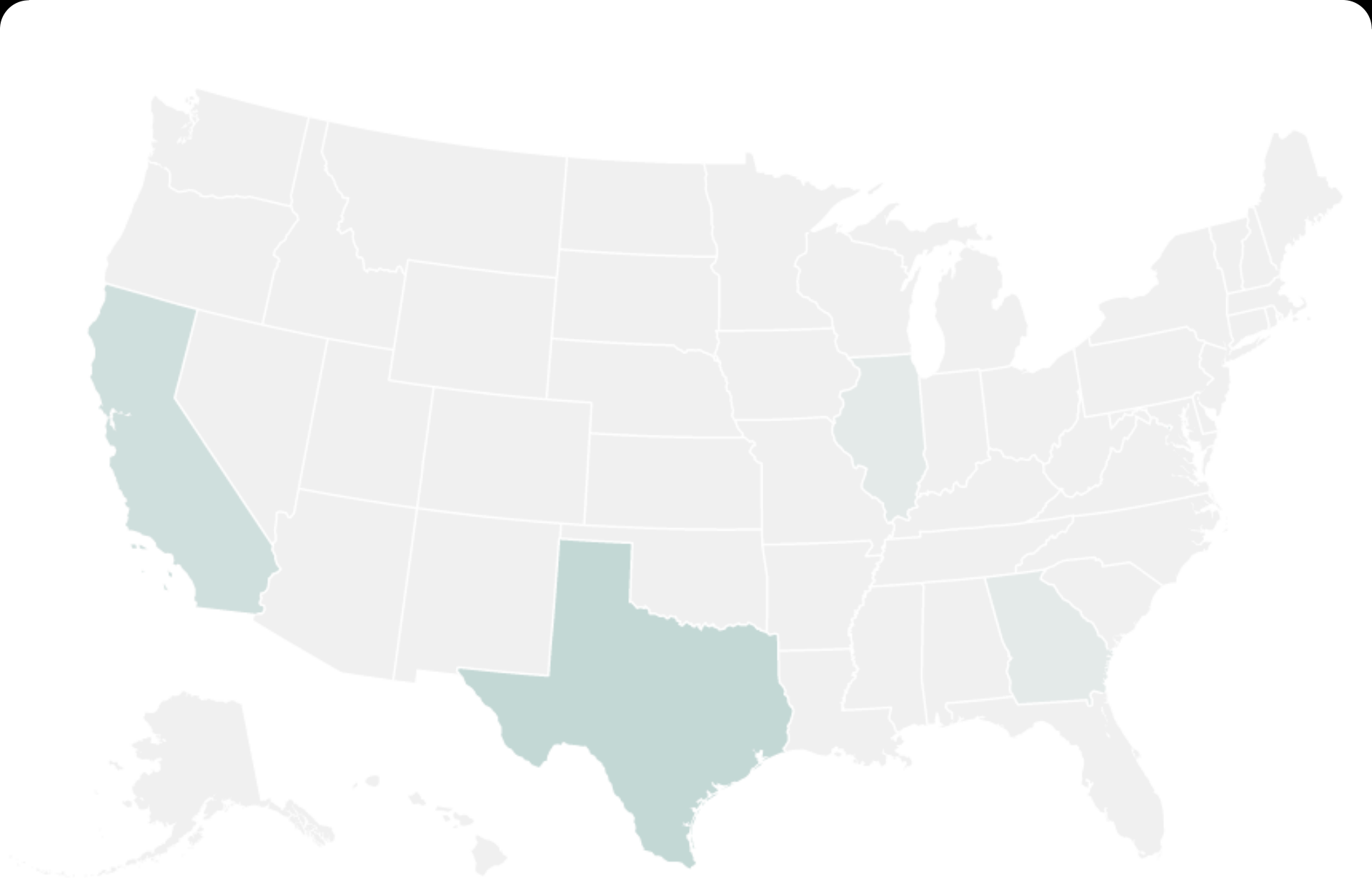
QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	**Color Hue**
Angle	Color Sat	Texture
Slope	**Color Hue**	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume



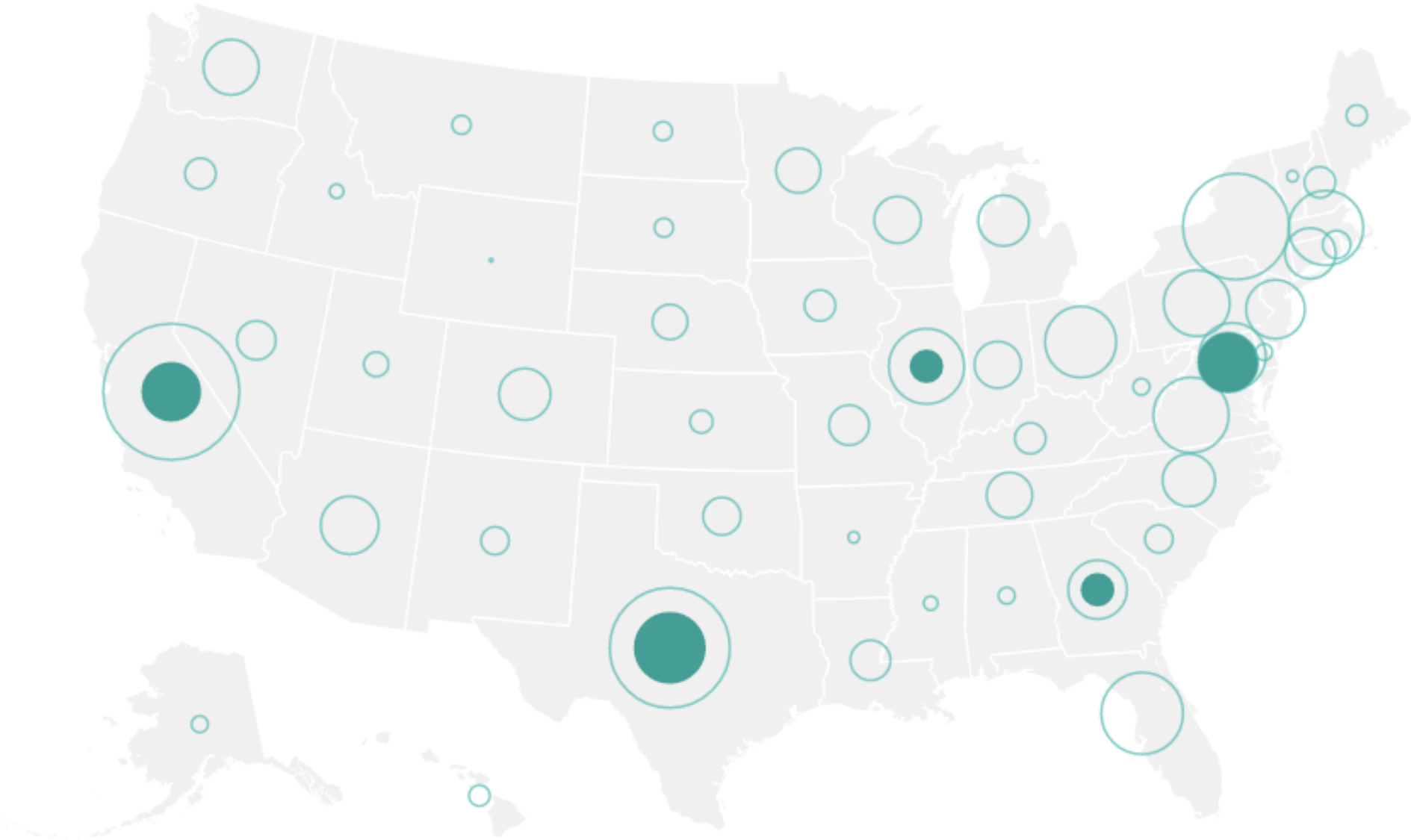
Color Encoding (Choropleth Map)

Effectiveness Rankings

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
~~**Density (Value)**~~	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume



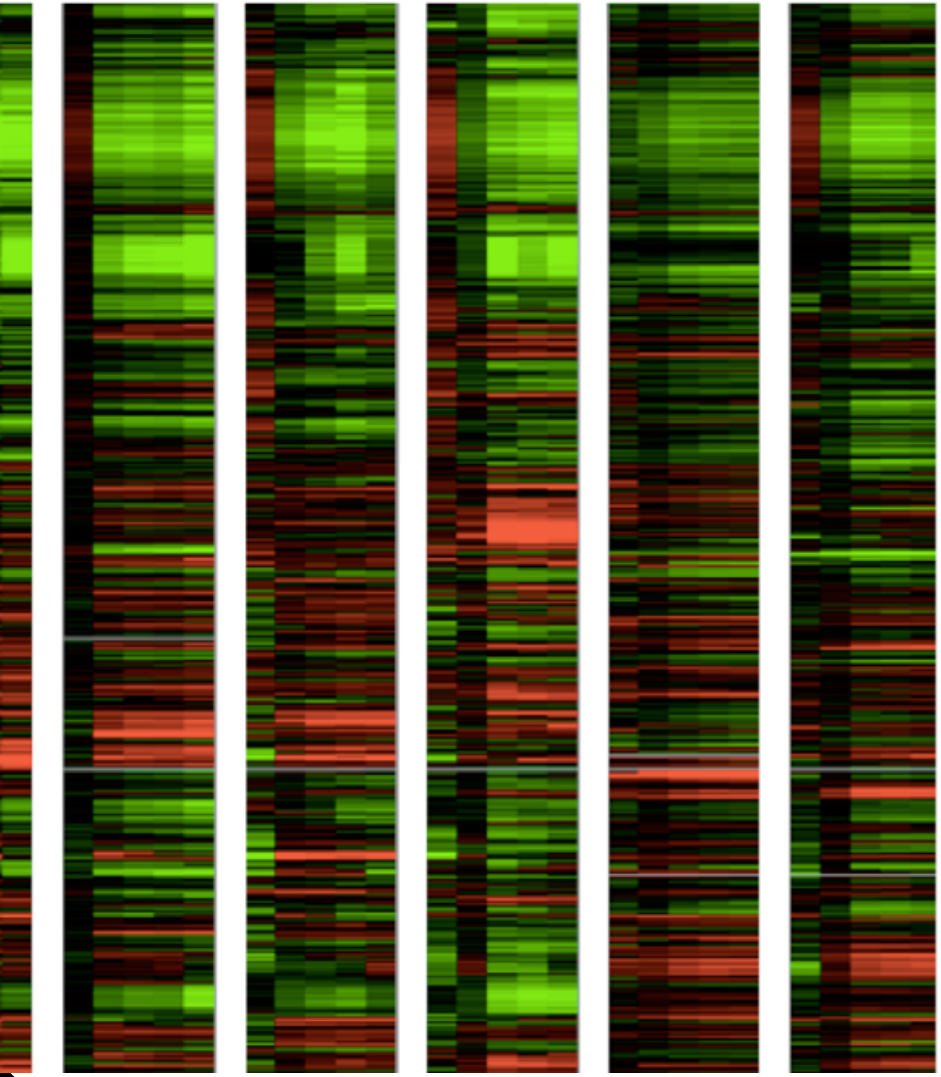
Color Encoding (Choropleth Map)



Area Encoding (Symbol Map)

Gene Expression Time-Series [Meyer et al '11]

Color Encoding

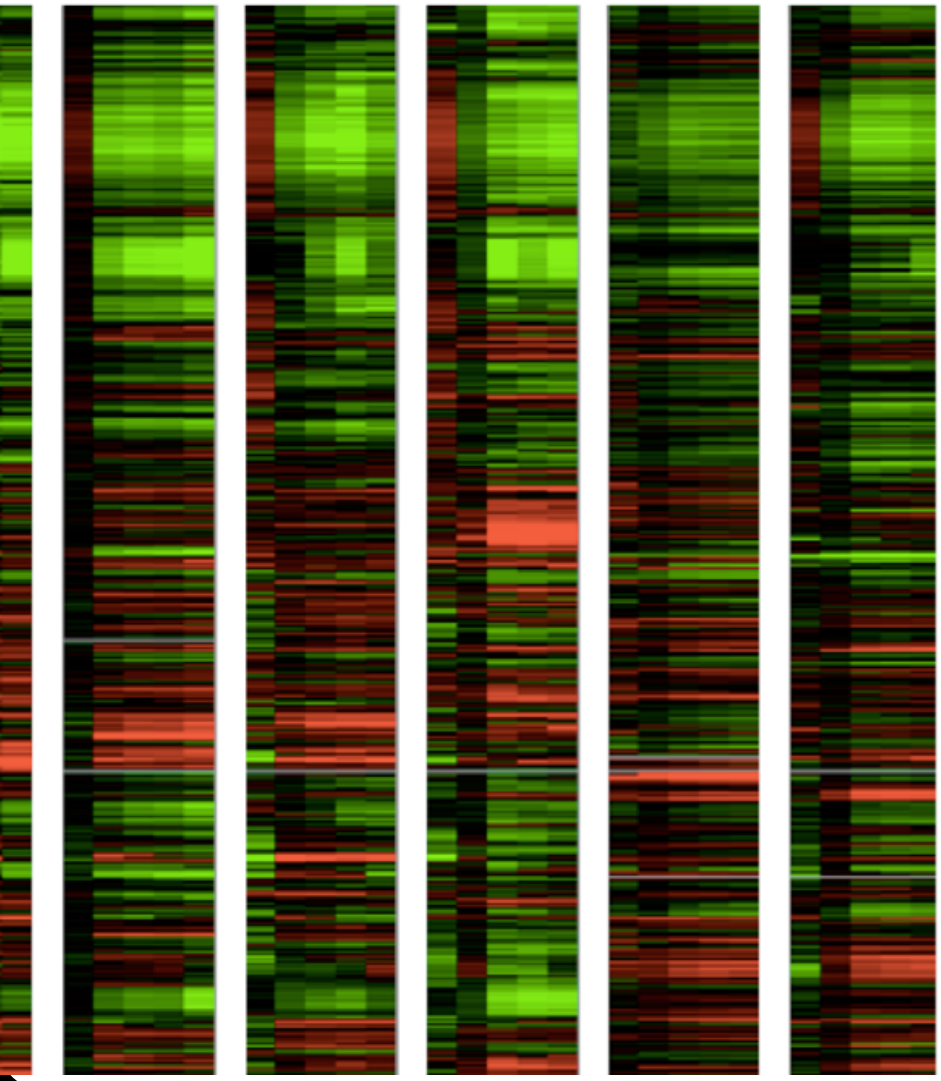


Effectiveness Rankings

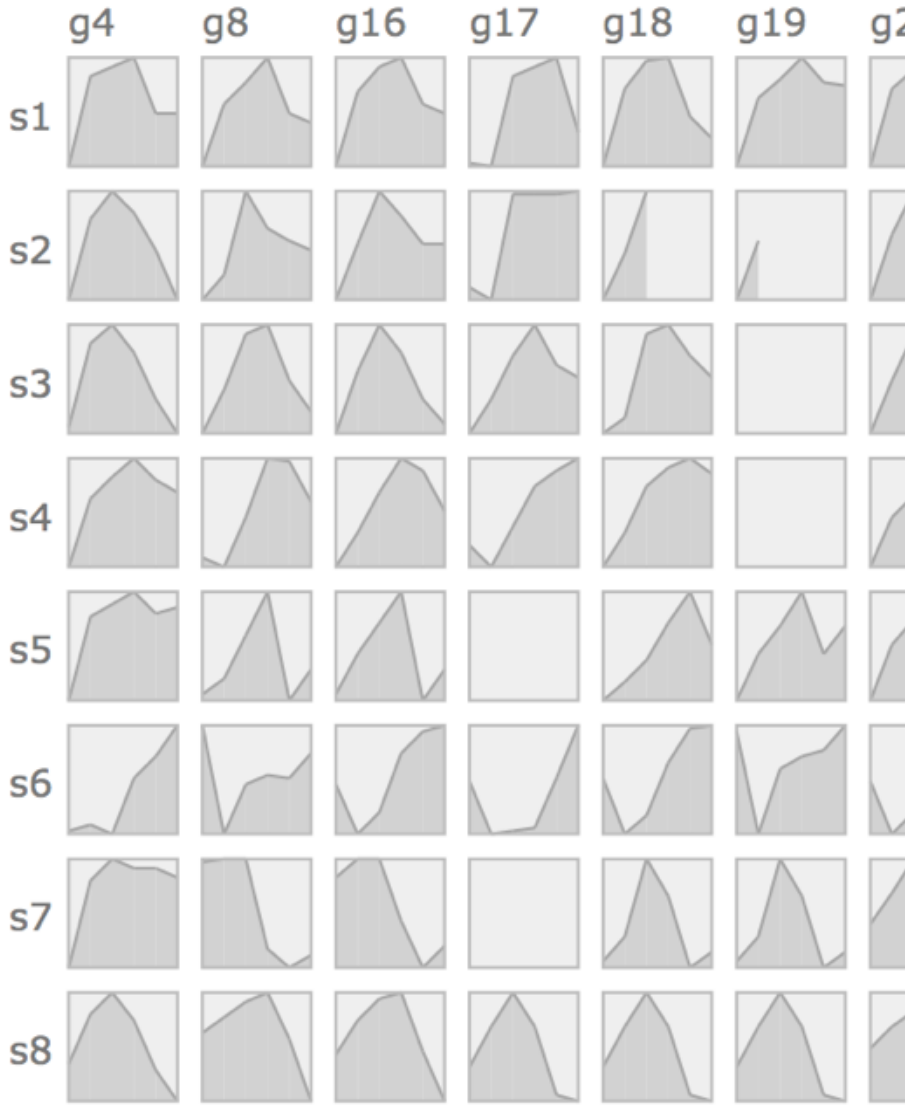
QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Value)
~~**Density (Value)**~~	Containment	Color Sat
Color Sat	Length	Shape
~~**Color Hue**~~	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Gene Expression Time-Series [Meyer et al '11]

Color Encoding



Position Encoding



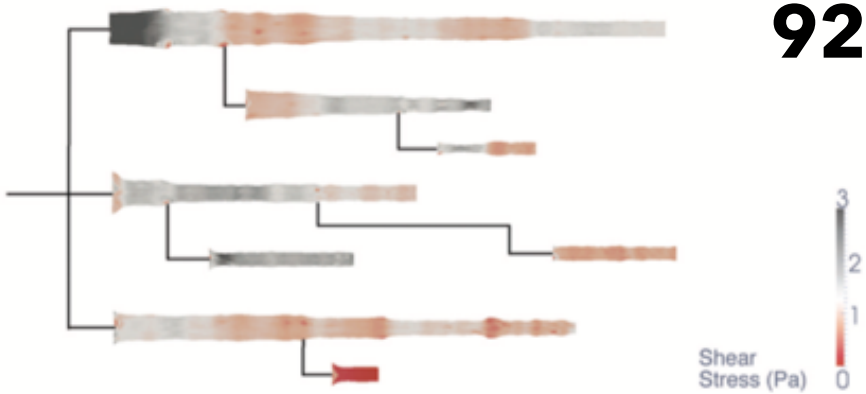
Artery Visualization [Borkin et al '11]

Rainbow Palette

Diverging Palette

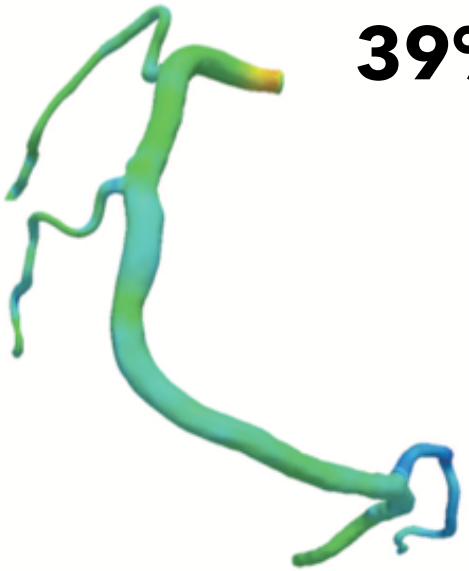
2D

62%

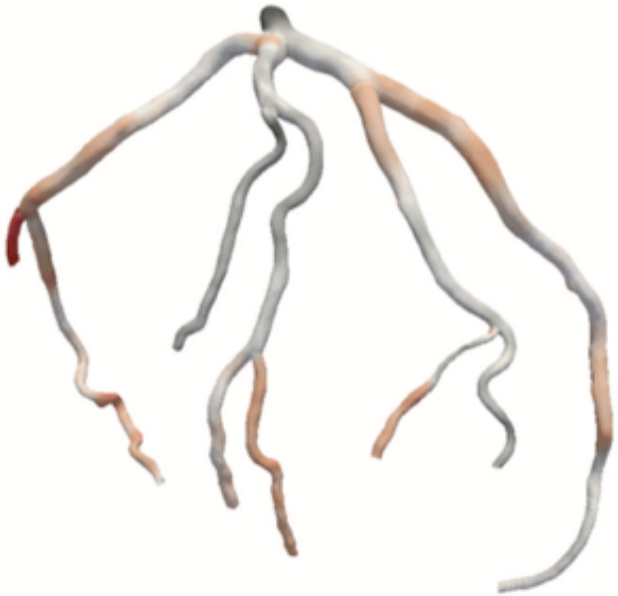


92%

3D



39%



71%

Effectiveness Rankings

QUANTITATIVE	ORDINAL	NOMINAL
Position ↻	Position	Position
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	**Connection**	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat	Length	Shape
~~**Color Hue**~~	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Administrivia

A2: Exploratory Data Analysis

Use visualization software to form & answer questions

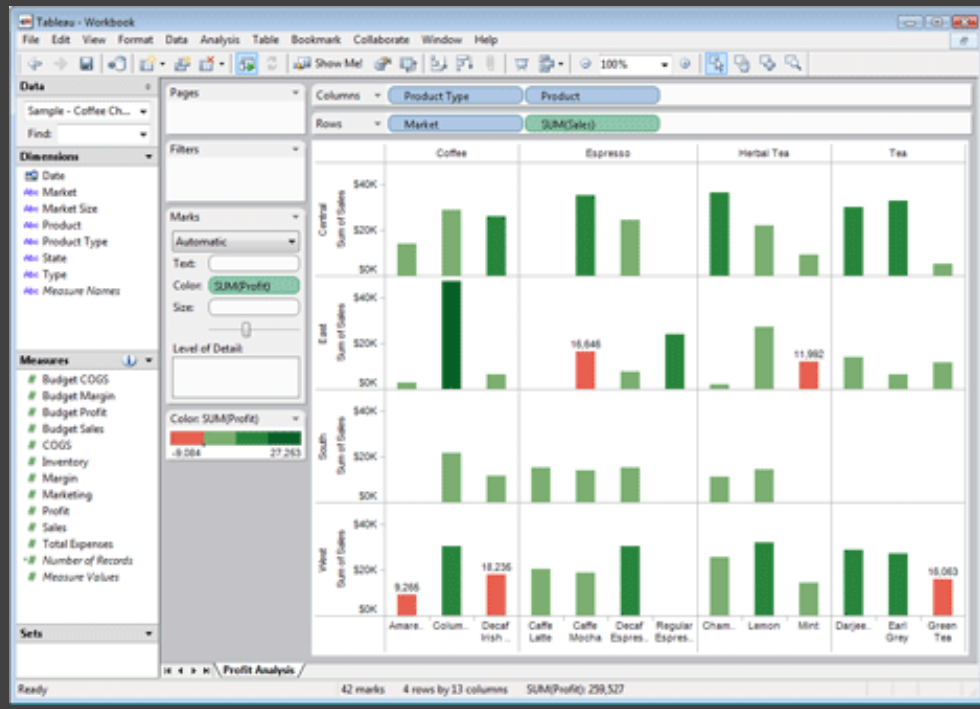
First steps:

Step 1: Pick domain & data

Step 2: Pose questions

Step 3: Profile the data

Iterate as needed



Create visualizations

Interact with data

Refine your questions

Author a report

Screenshots of most insightful views *(8+)*

Include titles and captions for each view

Due by 11:59pm

Monday, Apr 19

Required Readings for Wed 4/14

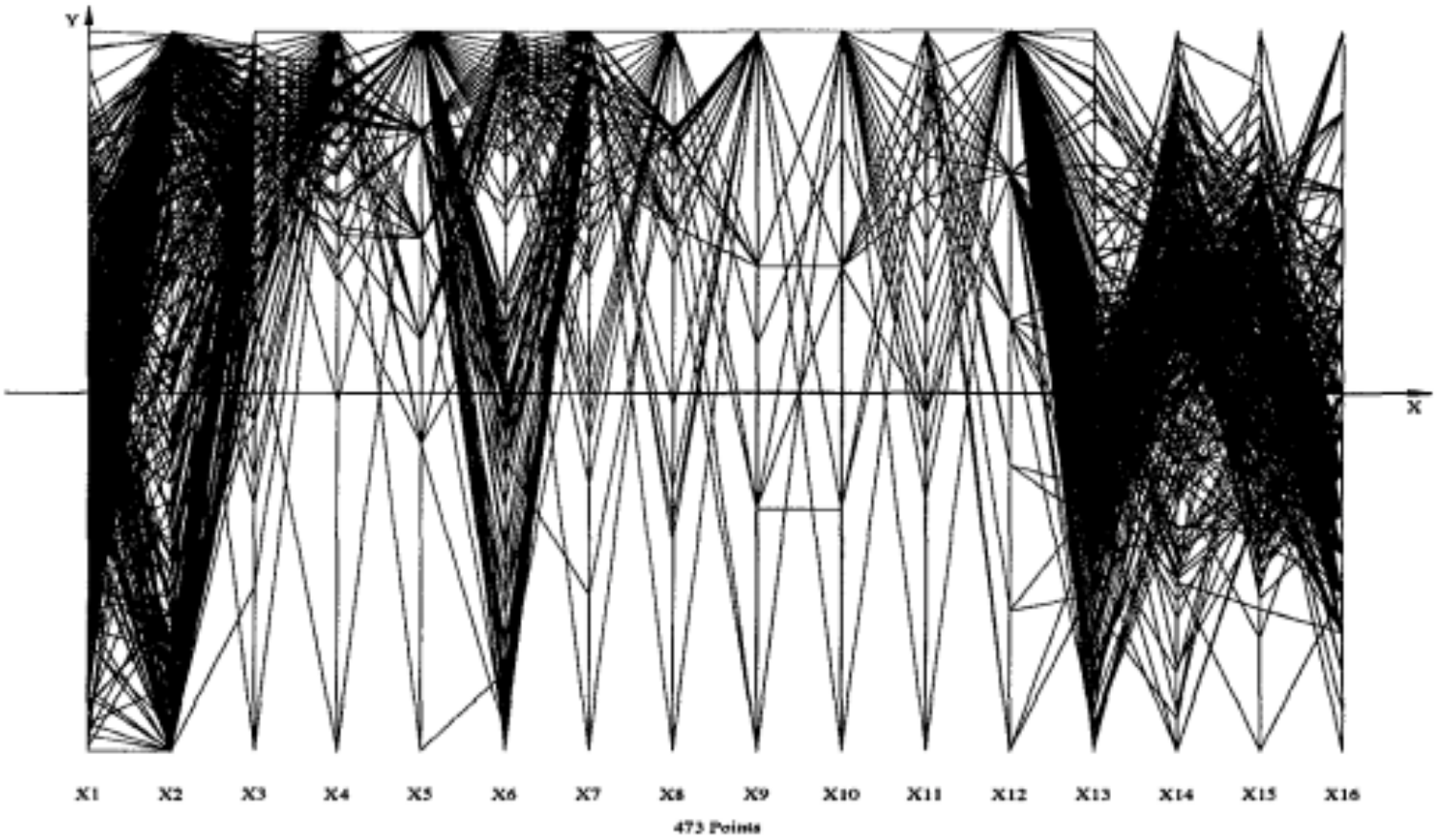
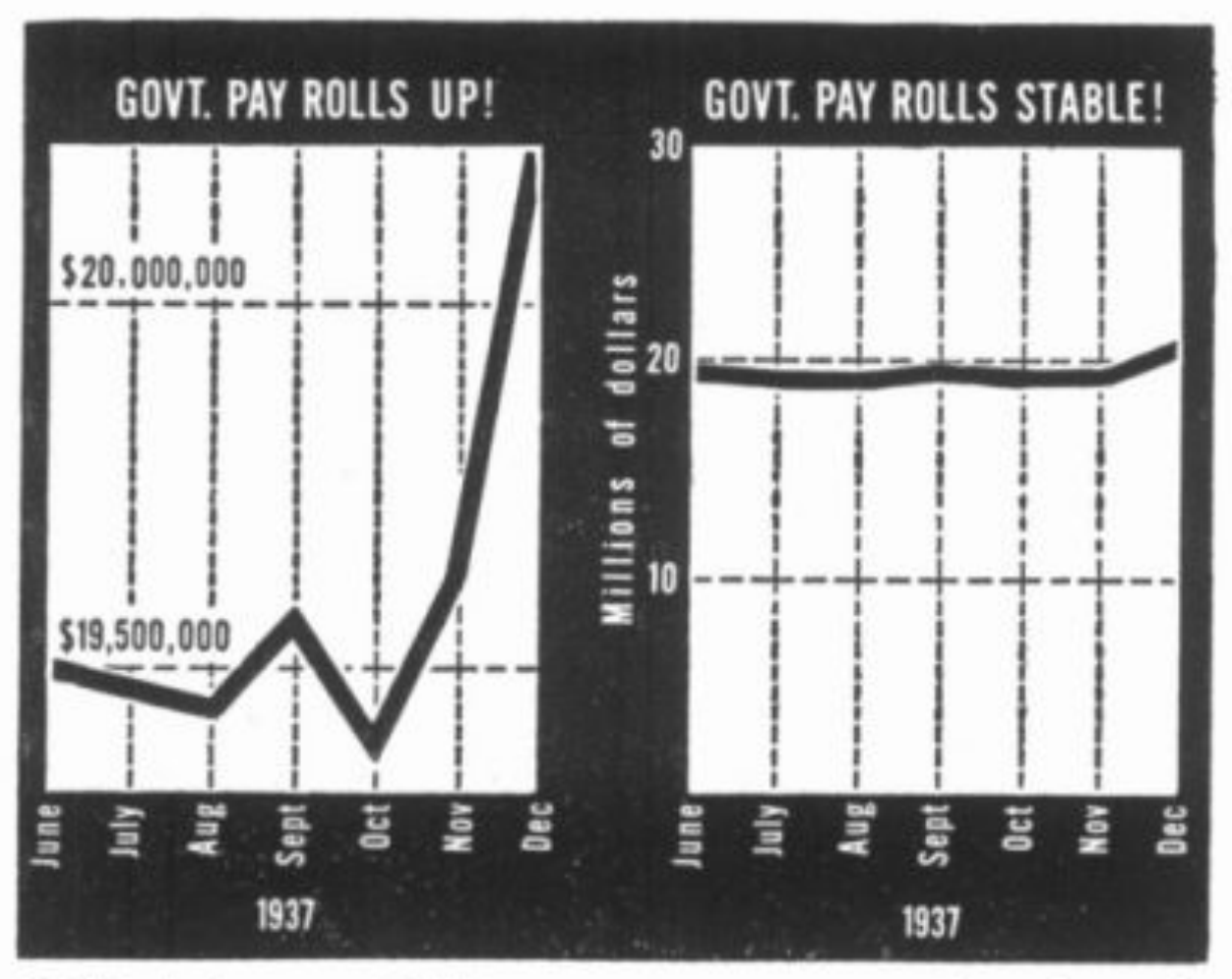


Figure 1: The full dataset consisting of 473 batches

Multidimensional Detective. Alfred Inselberg. IEEE InfoVis. 1997.

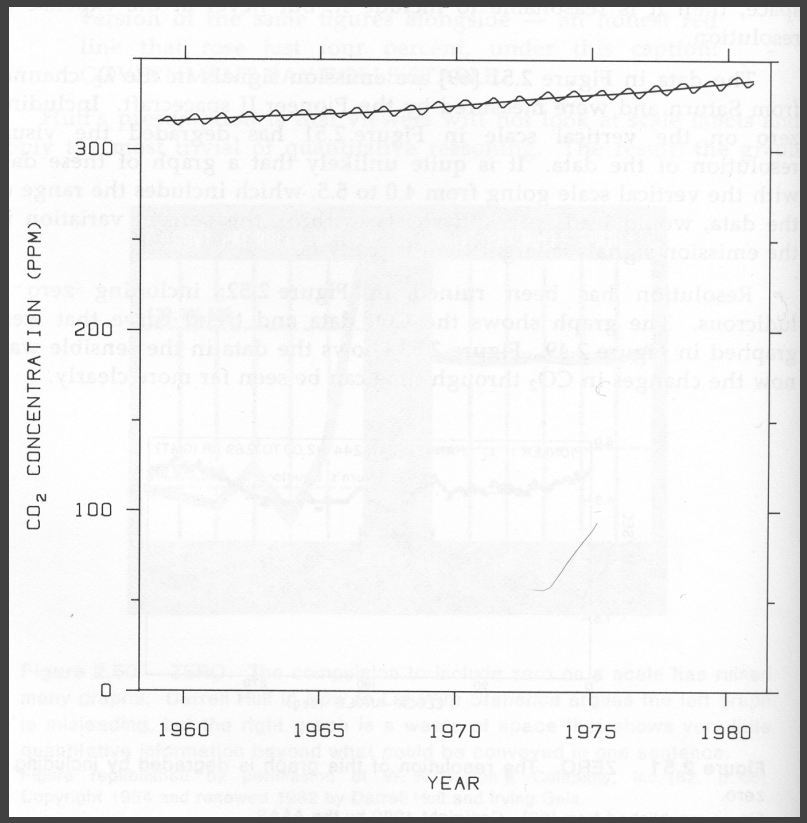
Scales & Axes

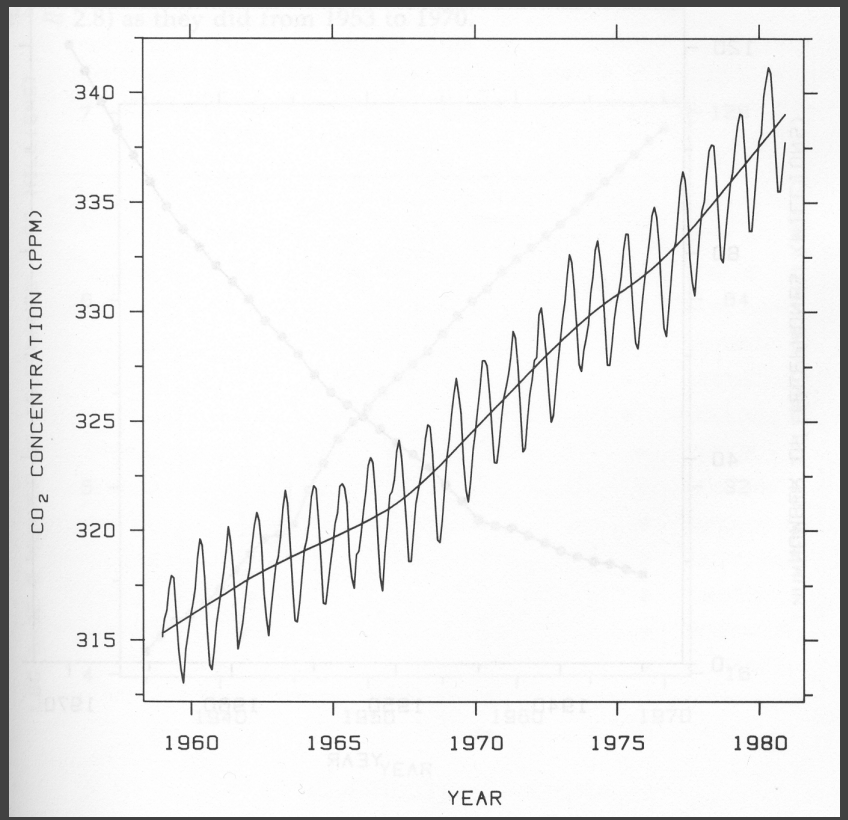
Include Zero in Axis Scale?



Government payrolls in 1937 [How To Lie With Statistics. Huff]

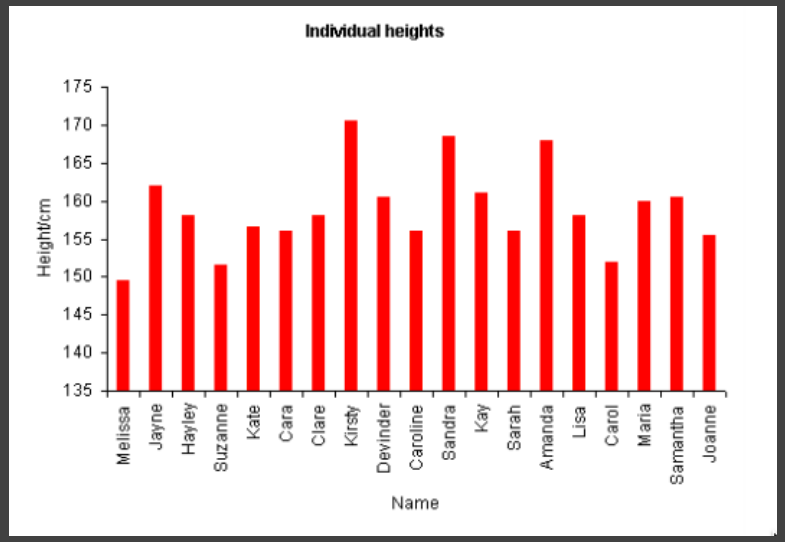
Include Zero in Axis Scale?





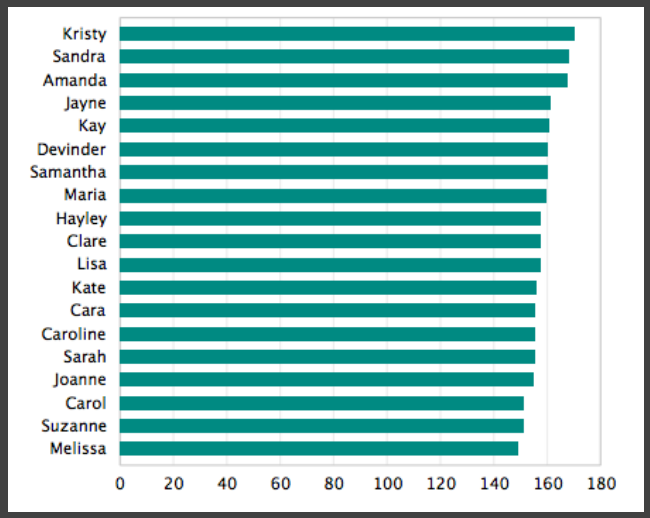
Yearly CO_2 concentrations [Cleveland 85]

Include Zero in Axis Scale?

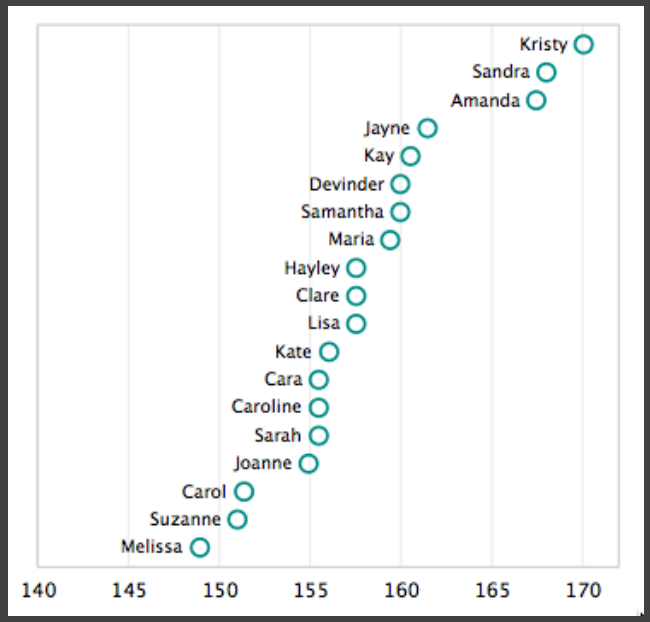


Violates Expressiveness Principle!

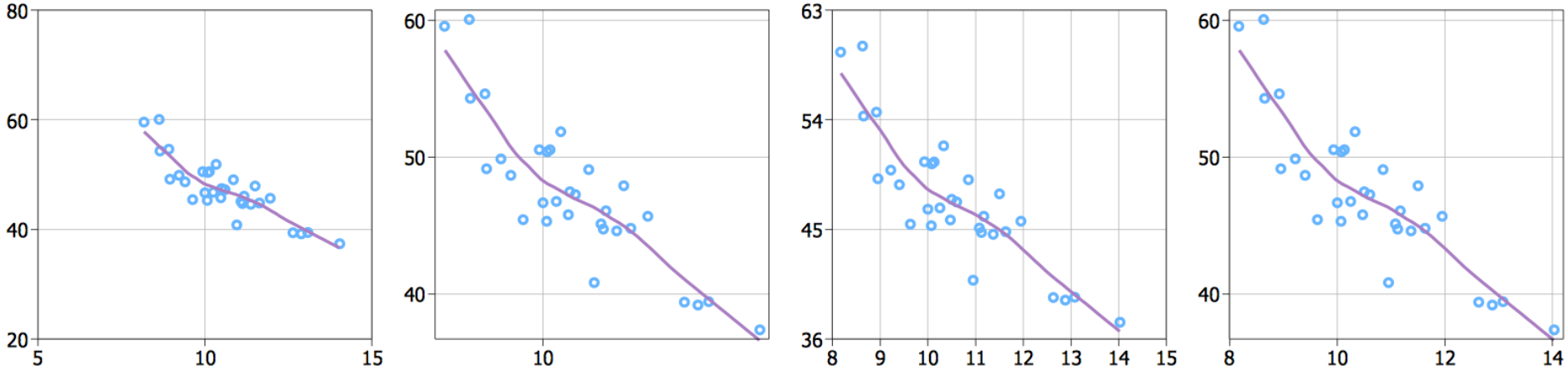
Compare Proportions (Q-Ratio)



Compare Relative Position (Q-Interval)

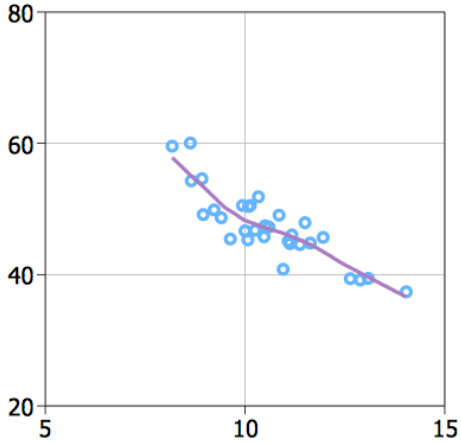


Axis Tick Mark Selection

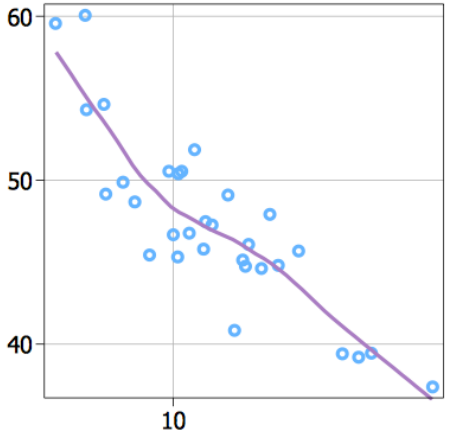


What are some properties of “good” tick marks?

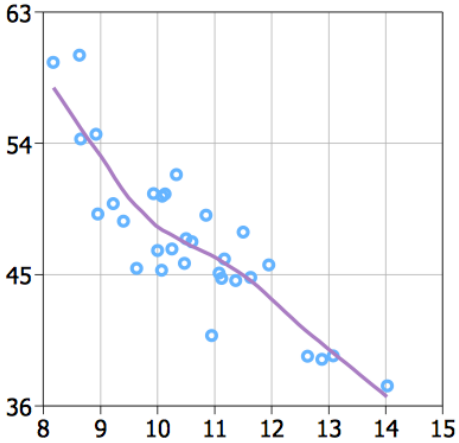
Axis Tick Mark Selection



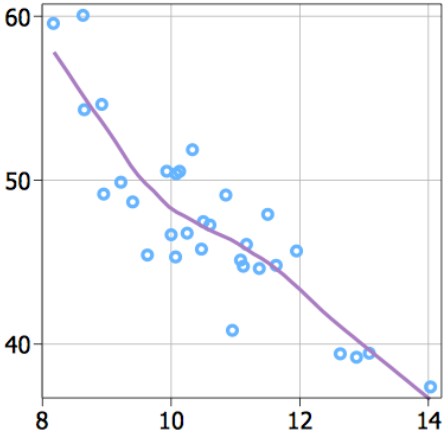
(a) Heckbert



(b) R's pretty



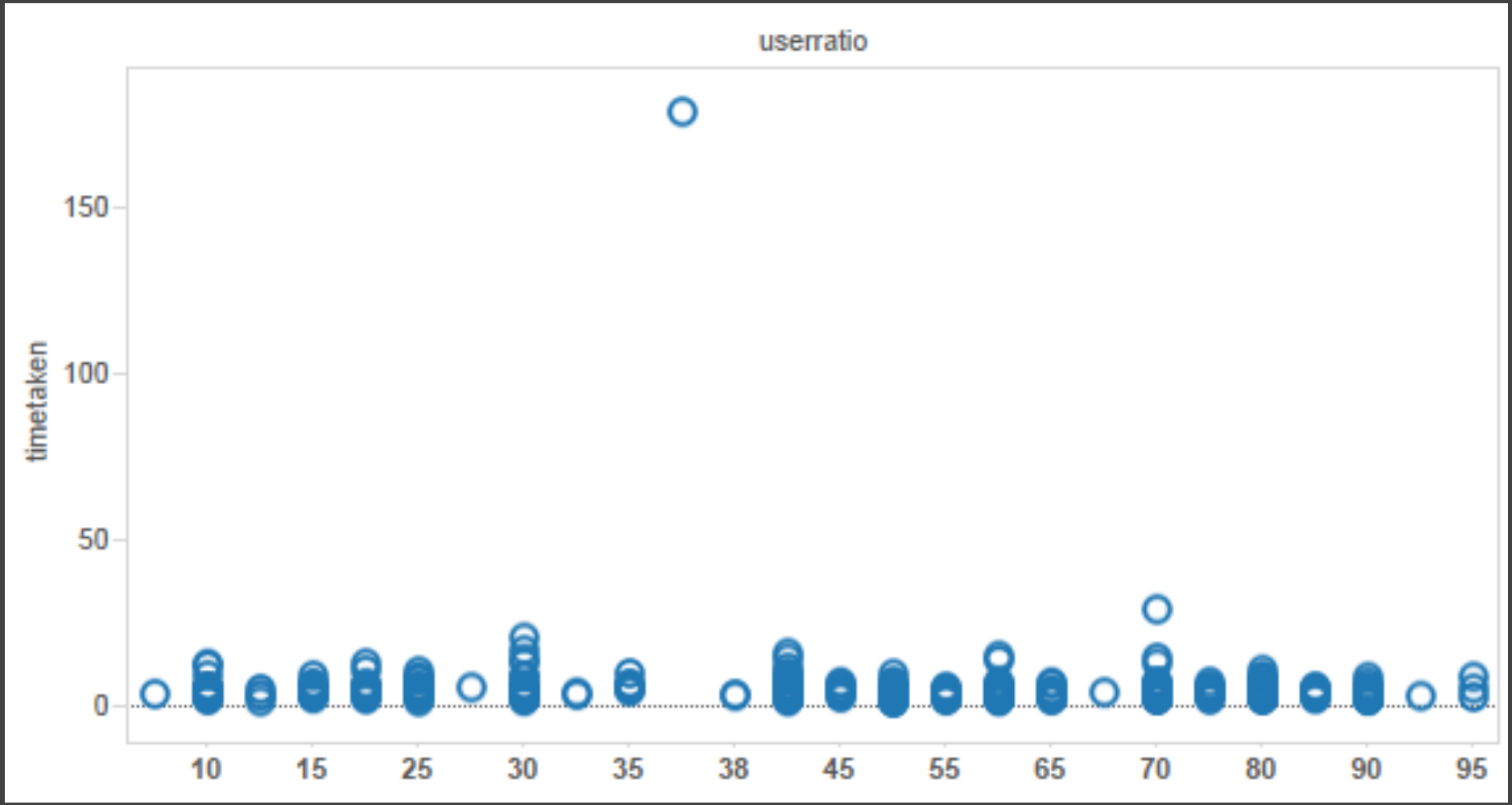
(c) Wilkinson



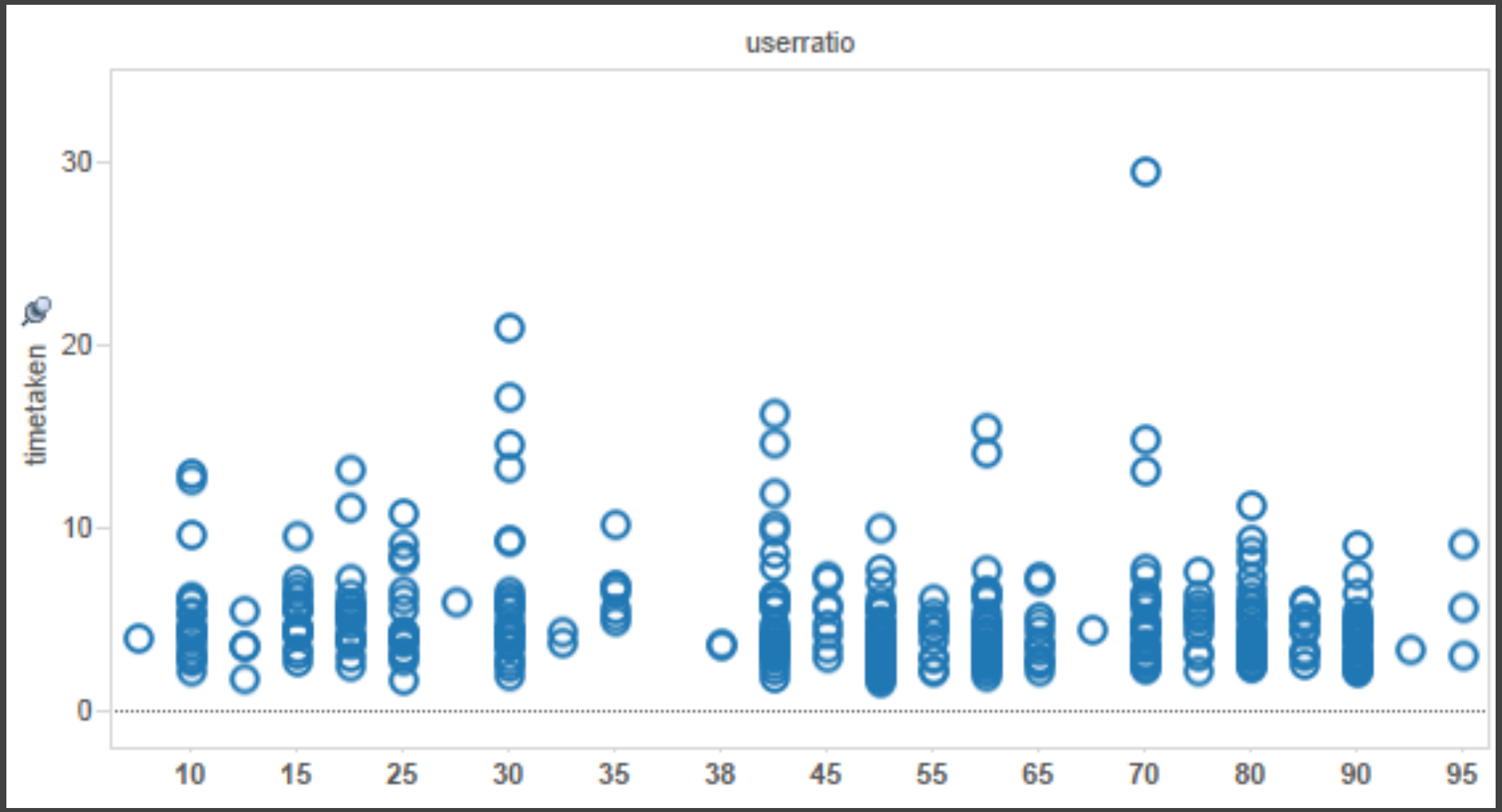
(d) Extended

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

How to Scale the Axis?

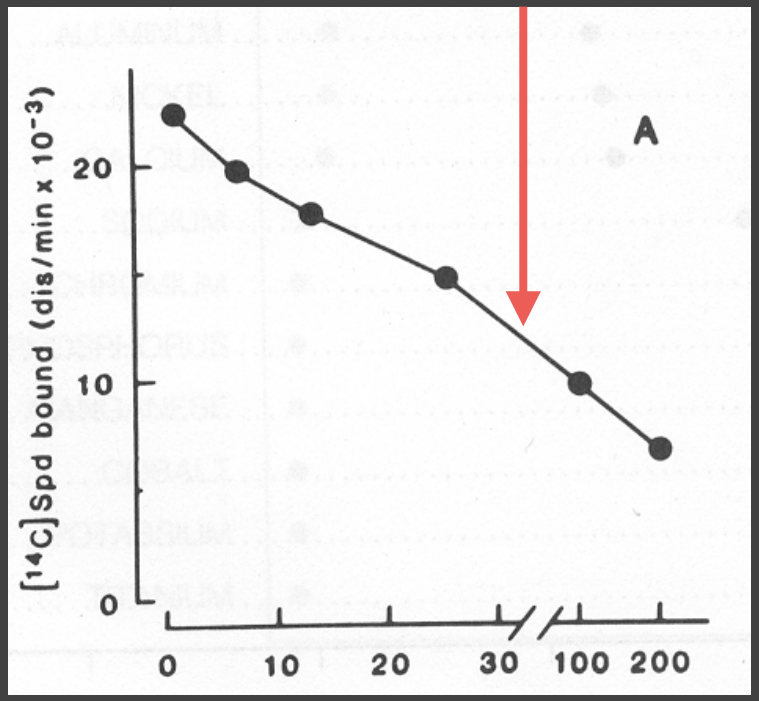


One Option: Clip Outliers

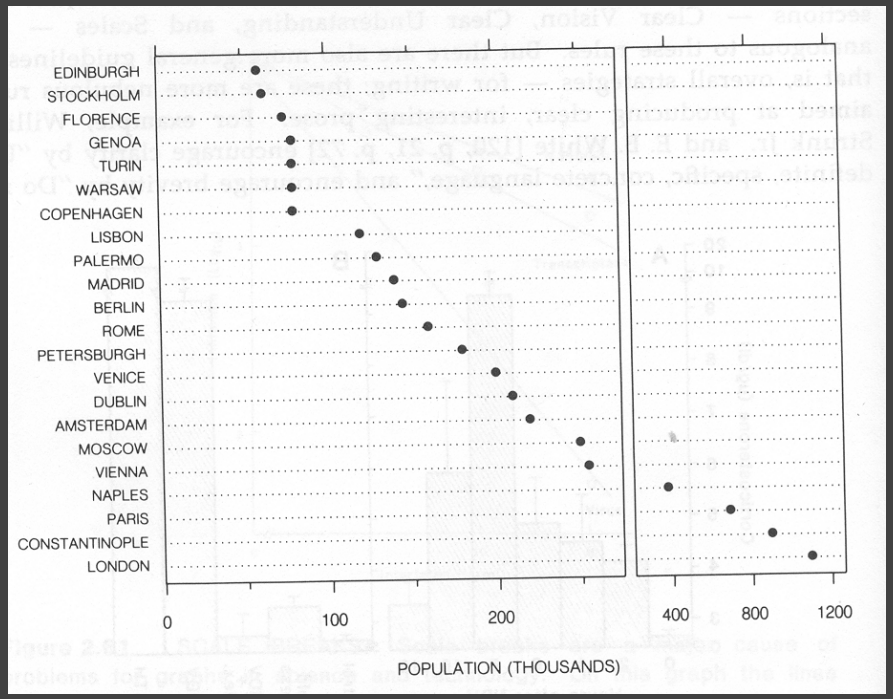


Clearly Mark Scale Breaks

Violates Expressiveness Principle!

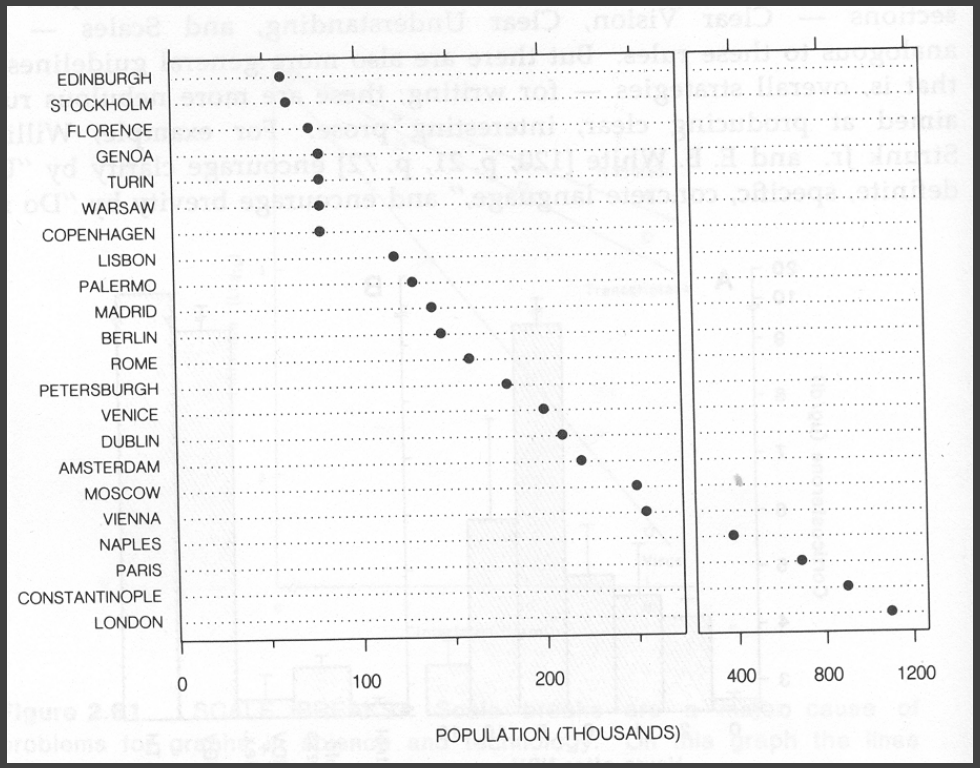


Poor scale break [Cleveland 85]

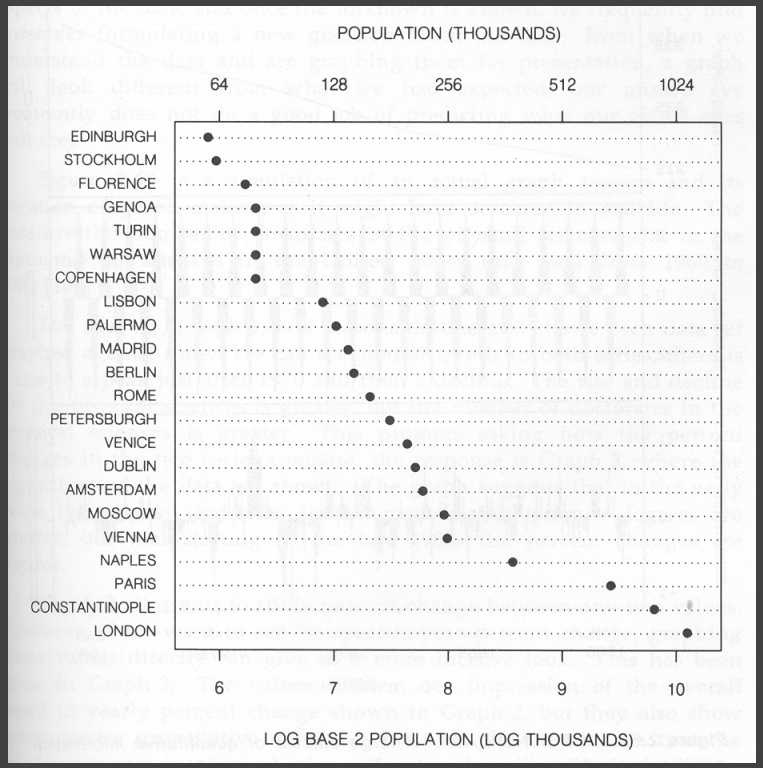


Well-marked scale break [Cleveland 85]

Scale Break vs. Log Scale



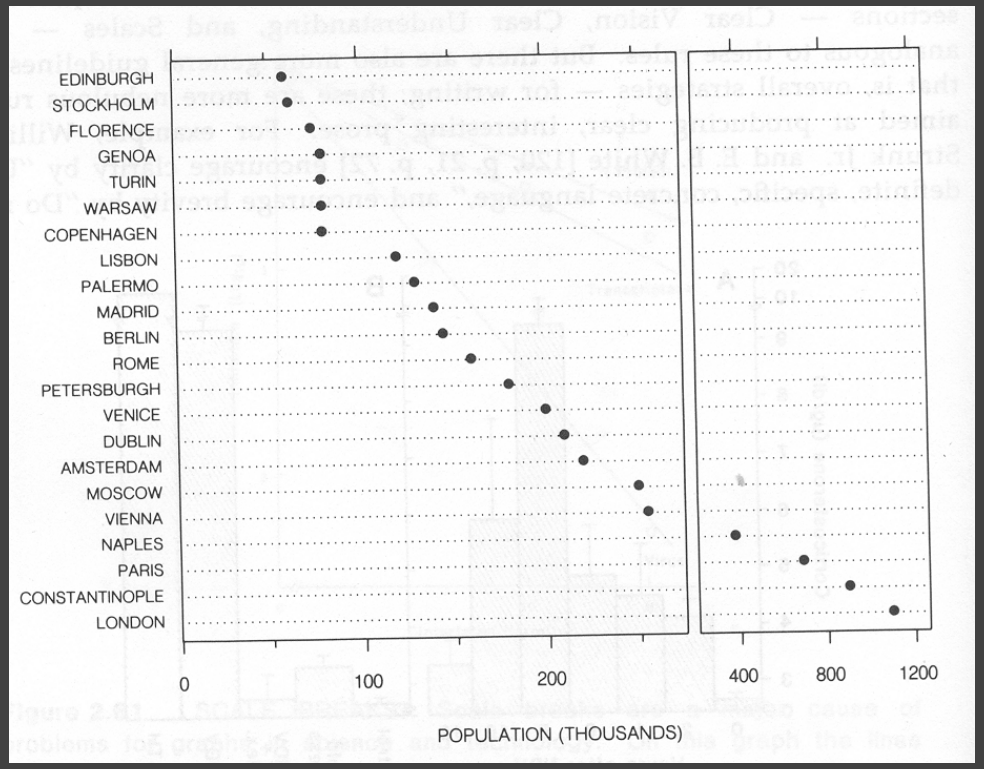
Scale Break

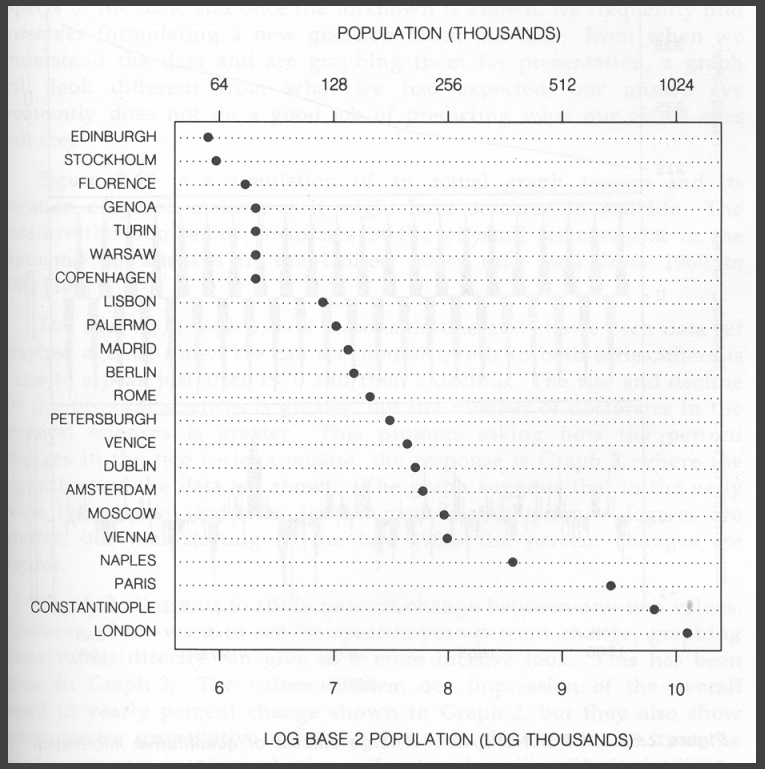


Log Scale

[Cleveland 85]

Scale Break vs. Log Scale





Both increase visual resolution

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

Log scale: direct comparison of all data

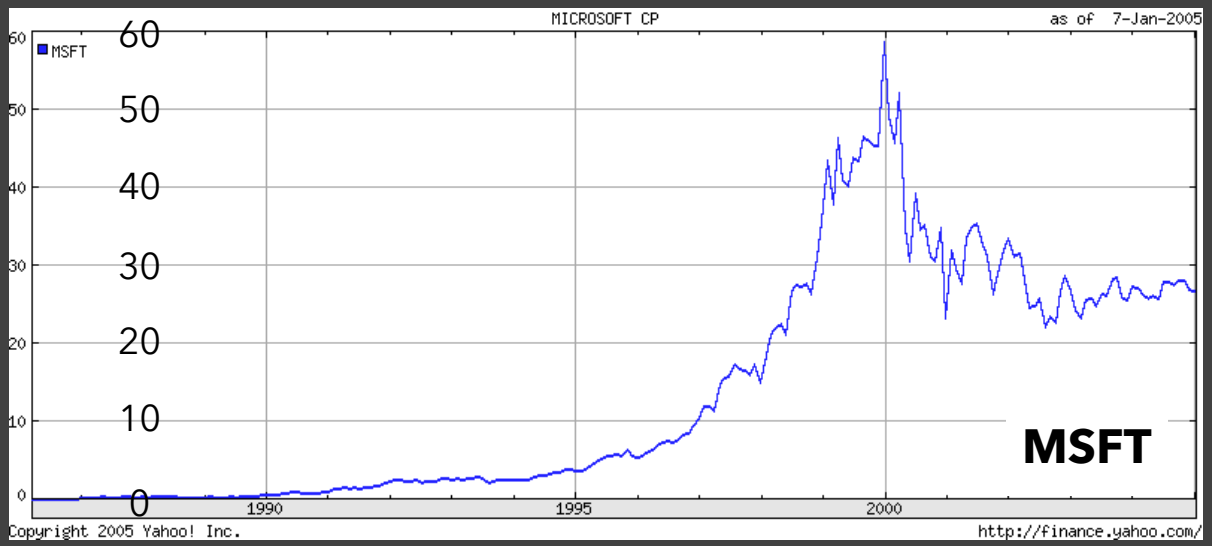
Logarithms turn *multiplication* into *addition*.

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

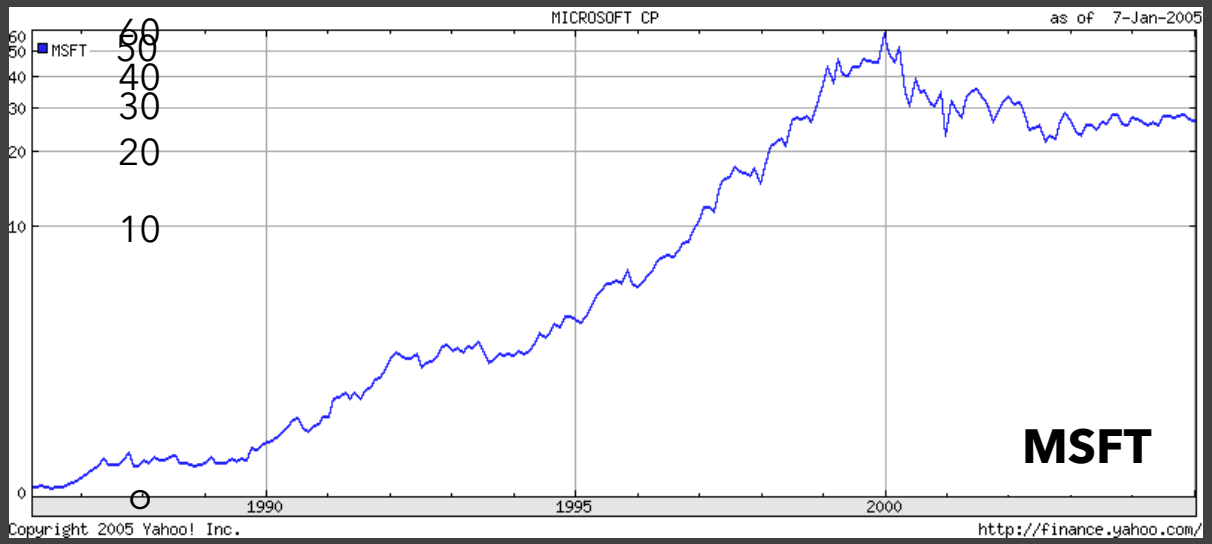
Equal steps on a log scale correspond to equal changes to a multiplicative scale factor.

Linear Scale vs. Log Scale

Linear Scale



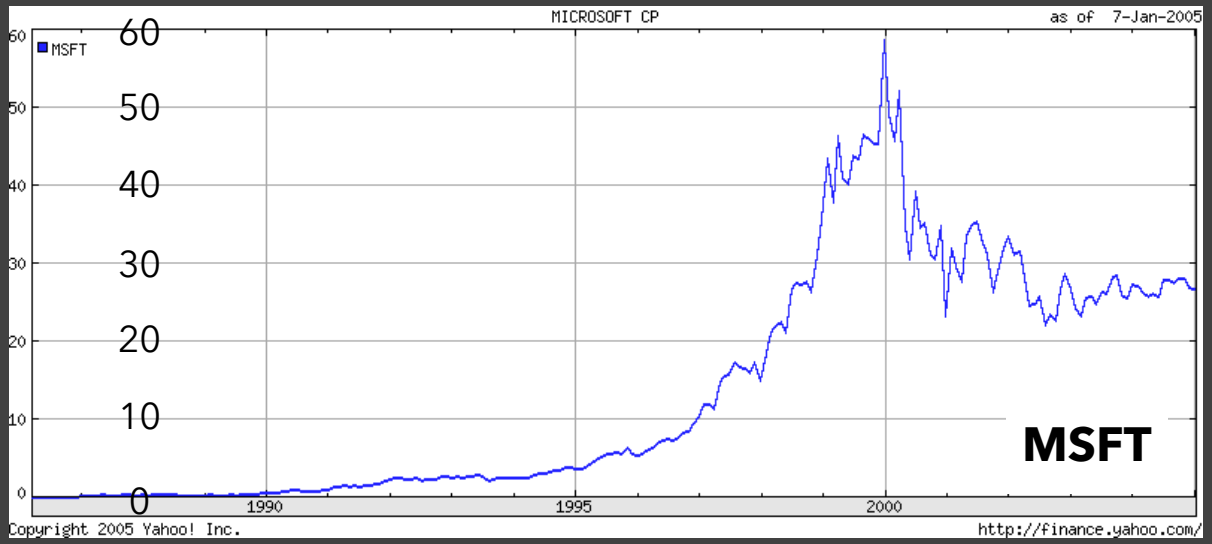
Log Scale



Linear Scale vs. Log Scale

Linear Scale

Absolute change

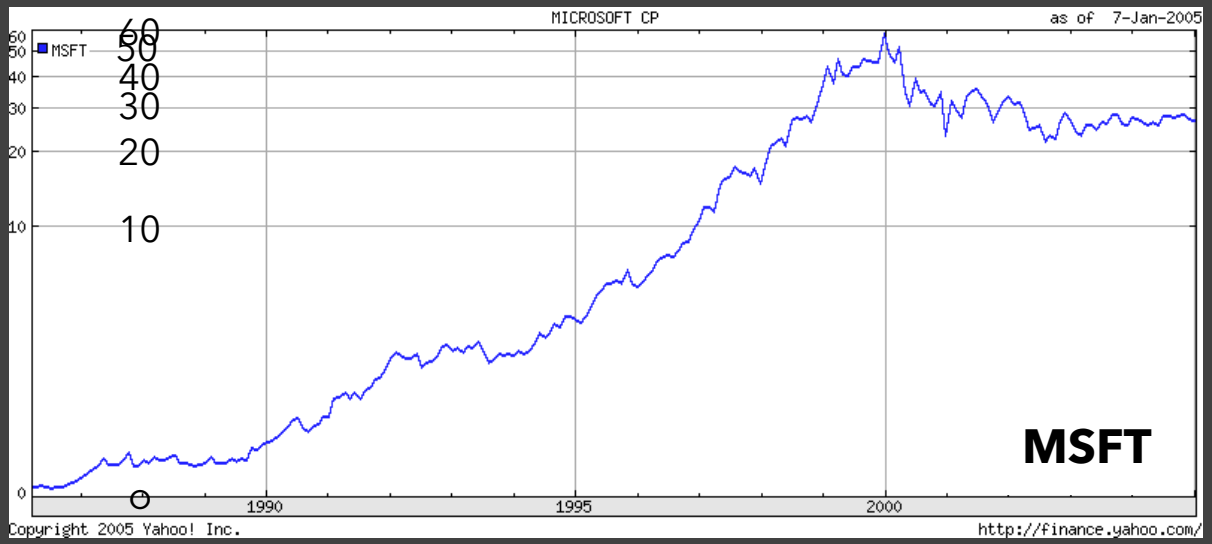


Log Scale

Small fluctuations

Percent change

d(10,30) > d(30,60)



When To Apply a 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 **+**!

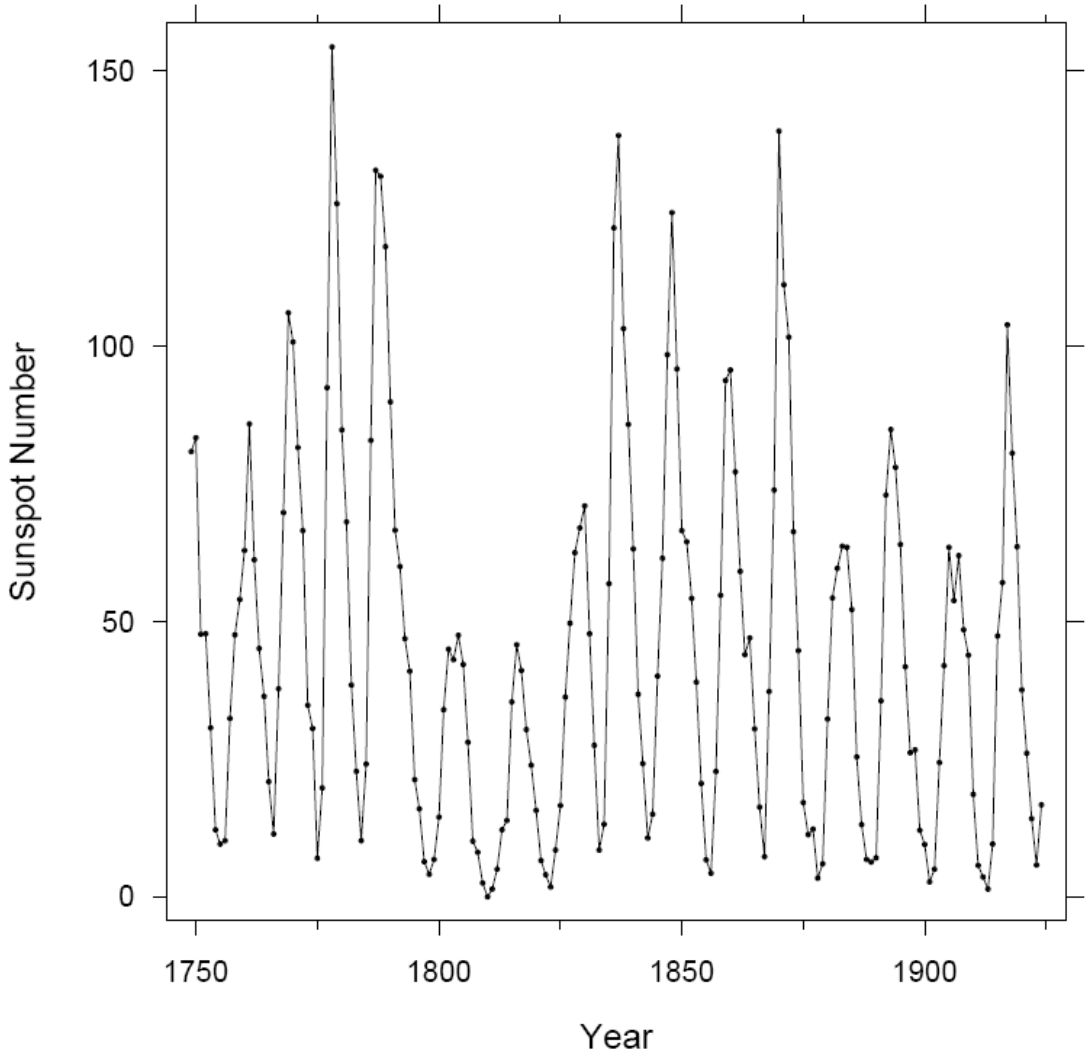
Percentage change, not linear difference.

Constraint: **positive, non-zero values**

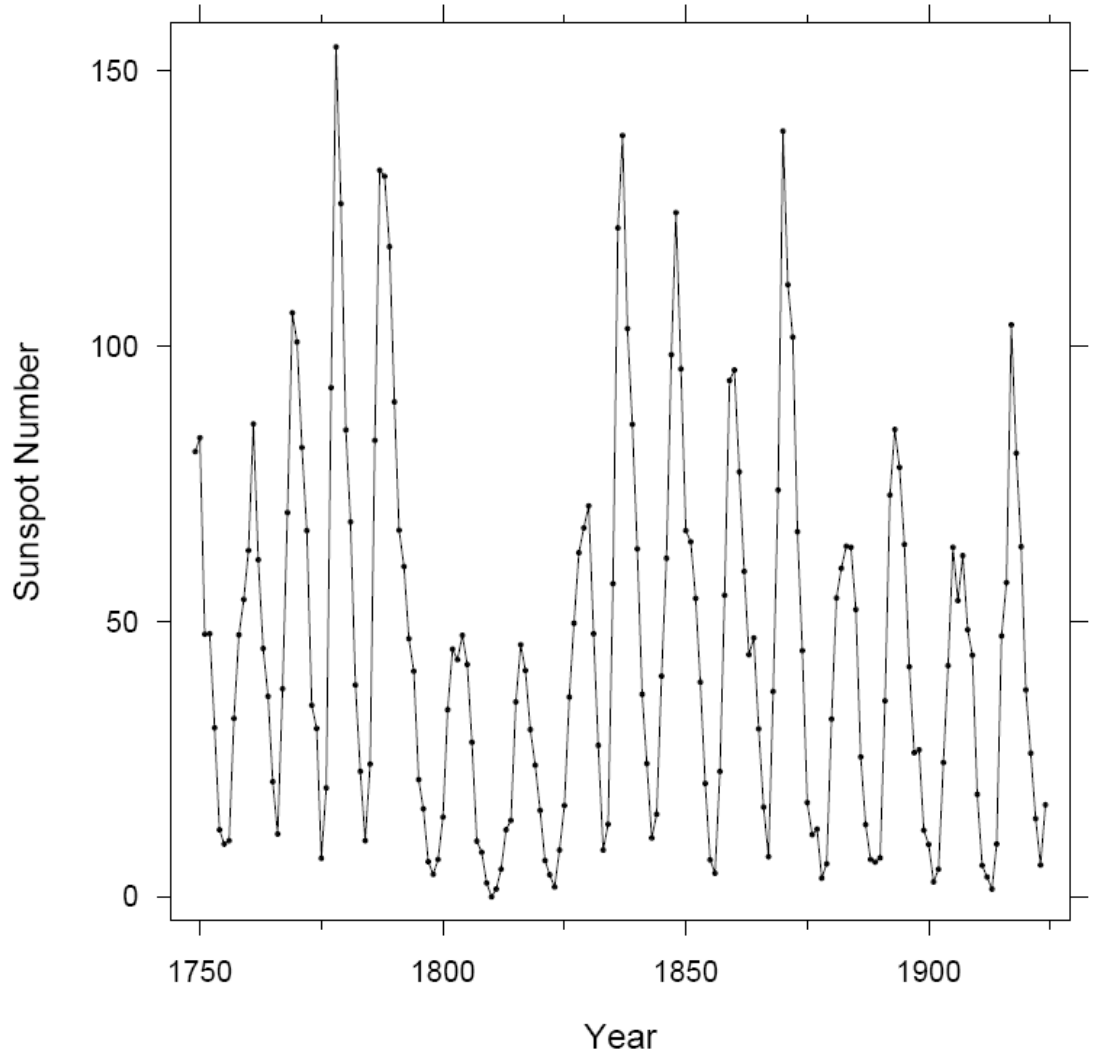
Constraint: **audience familiarity?**

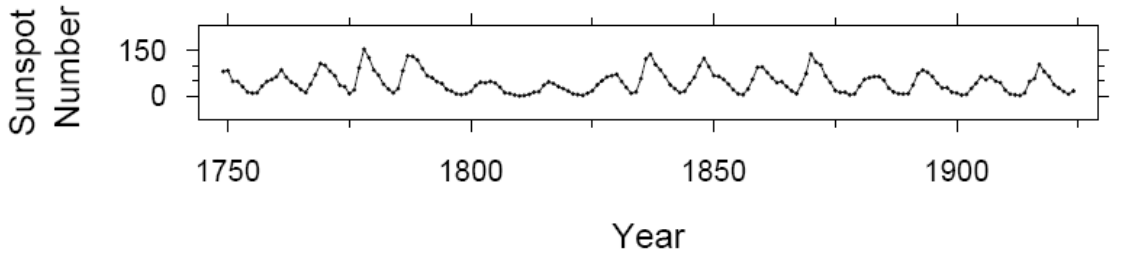
Aspect Ratio

(width : height)



William S. Cleveland
The Elements of Graphing Data

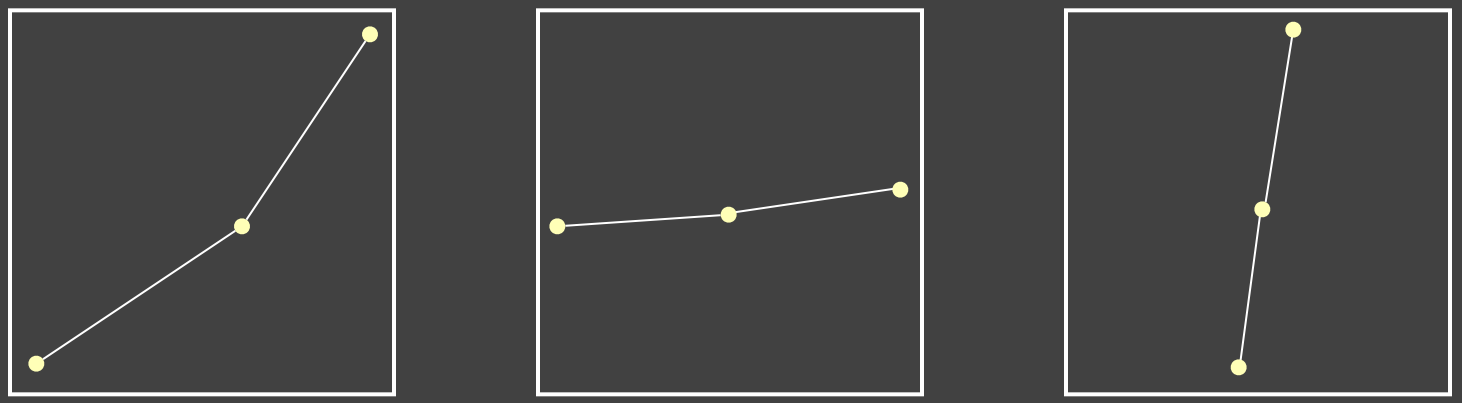




William S. Cleveland
The Elements of Graphing Data

Banking to 45° [Cleveland]

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

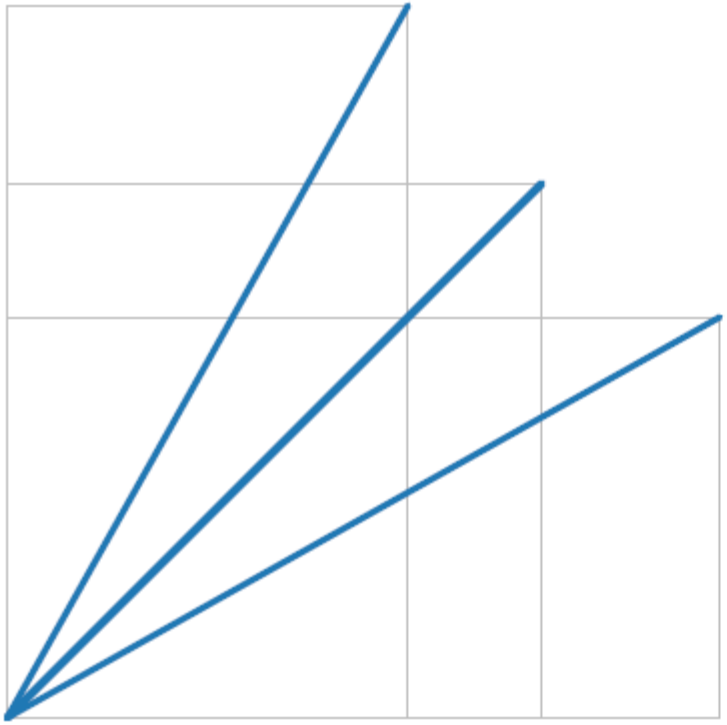


Two line segments are maximally discriminable when their average absolute angle is 45°

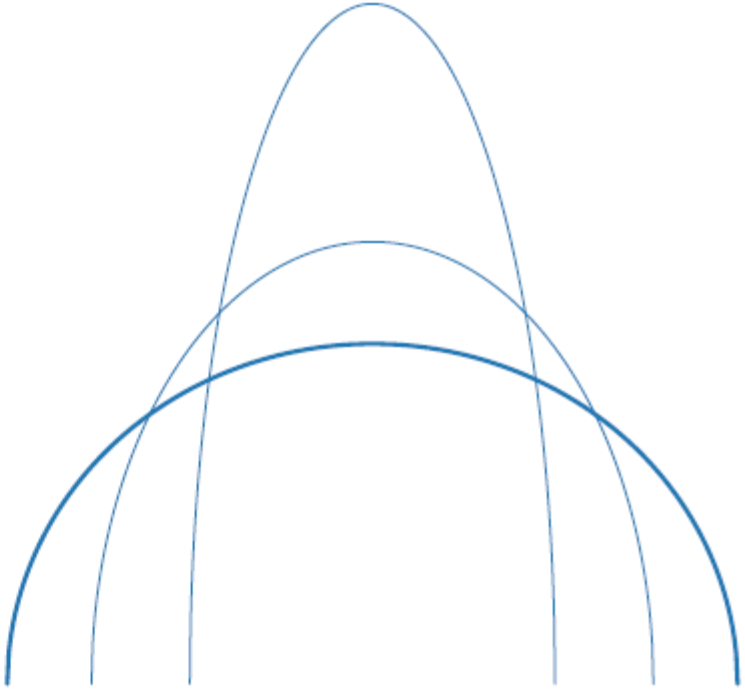
Method: optimize the aspect ratio such that the average absolute angle of all segments is 45°

Alternative: Minimize Arc Length

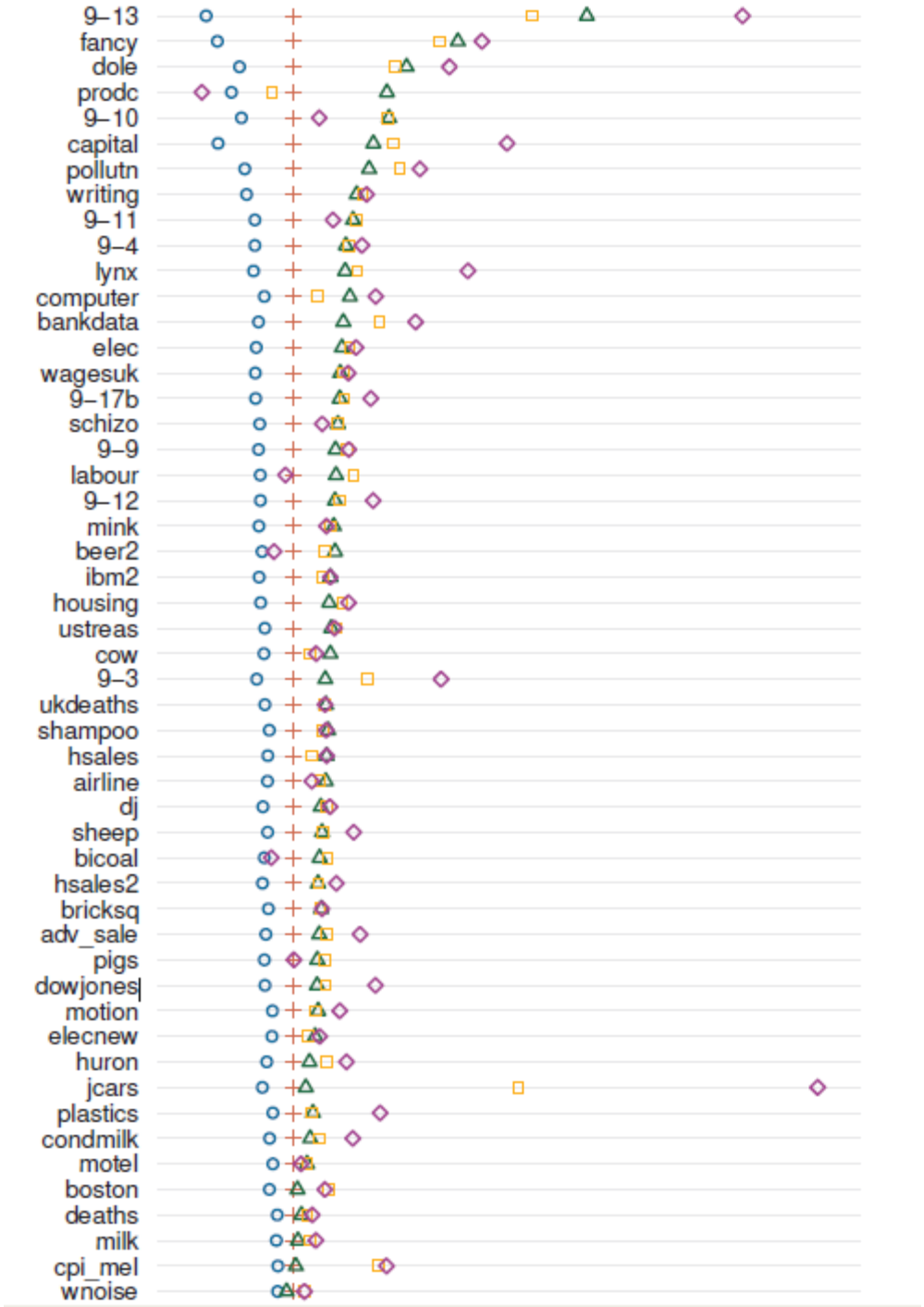
while holding area constant [Talbot et al. 2011]



Straight line -> 45°



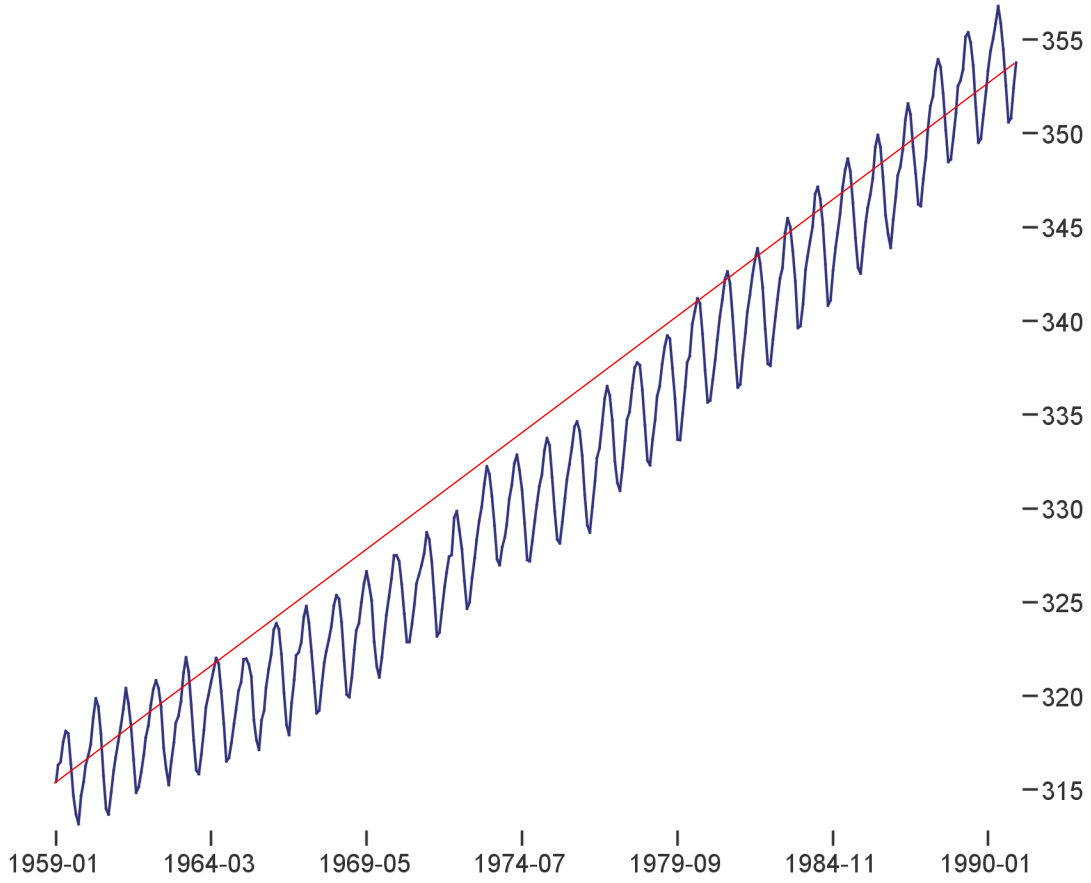
Ellipse -> Circle



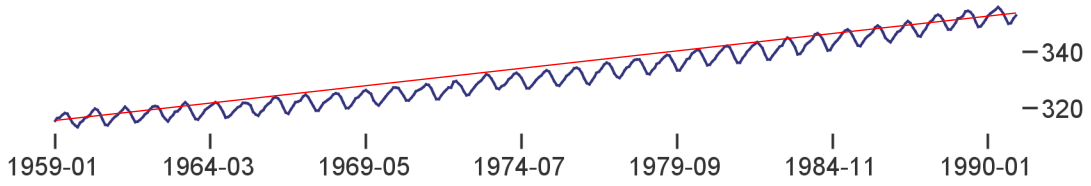
A Good Compromise

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

[Talbot et al. 2011]



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

Visual Encoding Design

Use **expressive** and **effective** encodings
Reduce the problem space
Avoid **over-encoding**
Use **space** and **small multiples** intelligently
Use **interaction** to generate *relevant* views

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

About the design process…

Visualization draws upon both science and art!

Principles like expressiveness & effectiveness are not hard-and-fast rules, but can assist us to guide the process and articulate alternatives.

They can lead us to think more deeply about our design rationale and prompt us to reflect.

It helps to know "the rules" in order to wisely bend (*or break*) them at the right times!