

CSE 442 - Data Visualization

Visual Encoding Design



Jeffrey Heer University of Washington

A Design Space of Visual Encodings

Mapping Data to Visual Variables

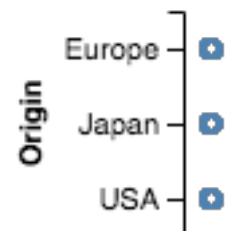
Assign **data fields** (e.g., with N , O , Q types) to **visual channels** (x , y , $color$, $shape$, $size$, ...) for a chosen **graphical mark** type ($point$, bar , $line$, ...).

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

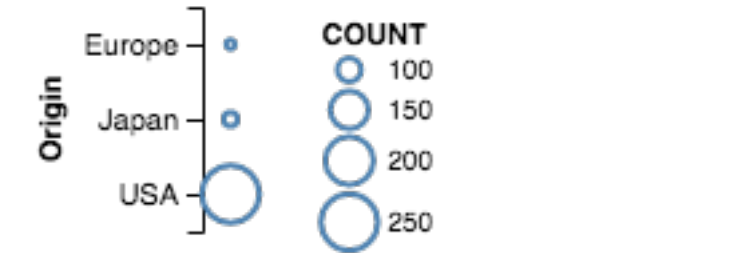
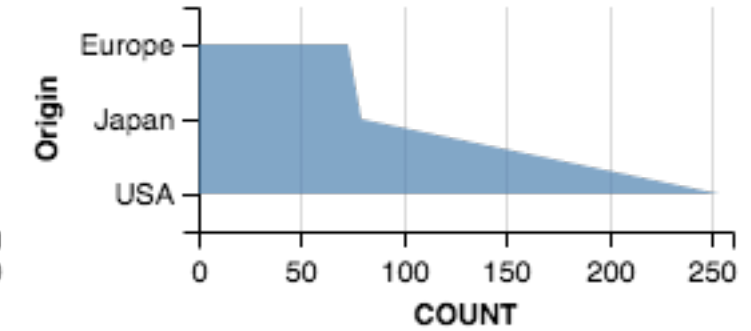
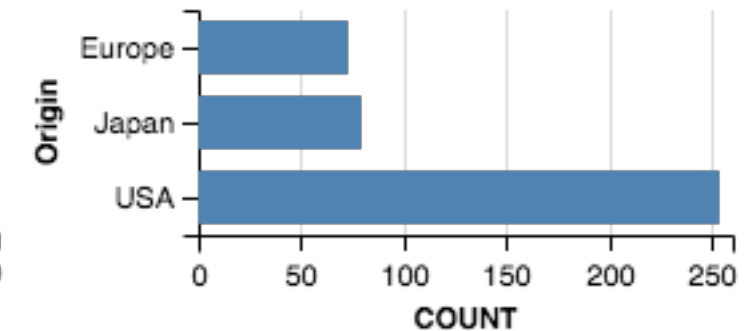
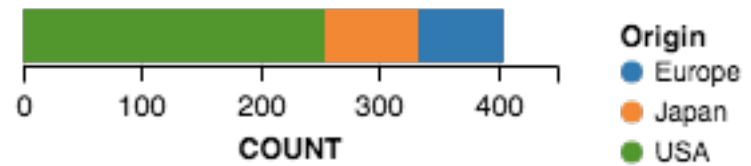
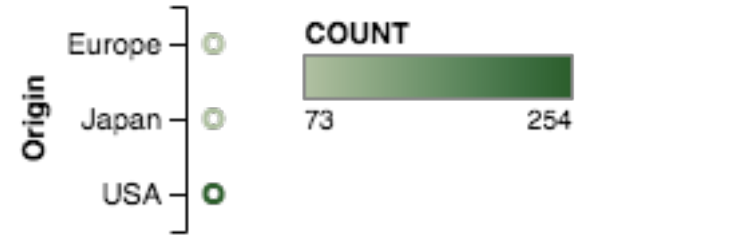
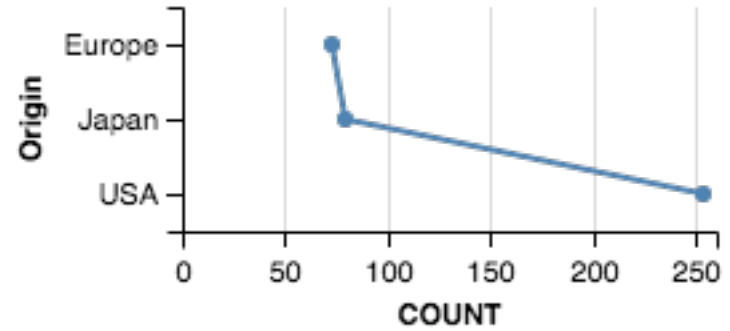
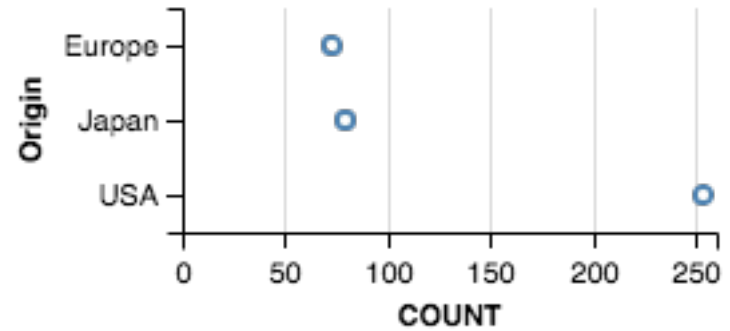
These options define a large combinatorial space, containing both useful and questionable charts!

1D: Nominal

Raw



Aggregate (Count)

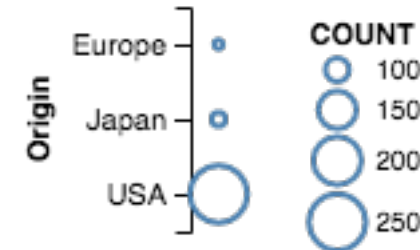
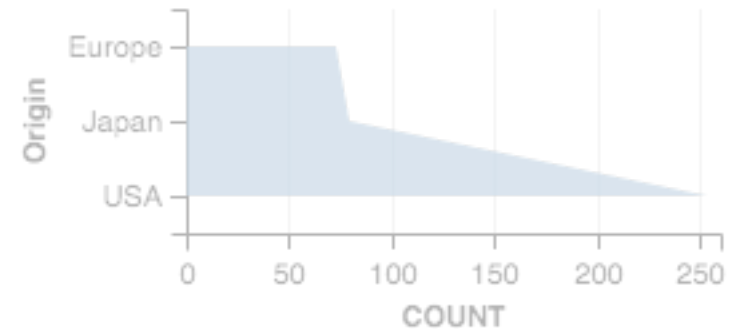
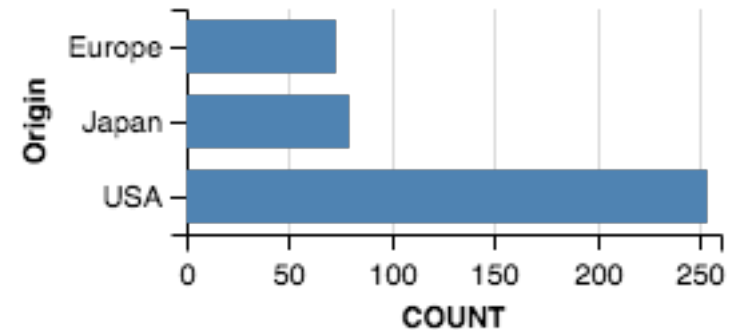
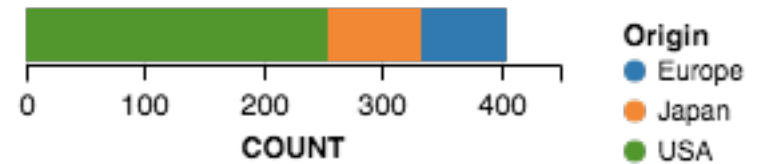
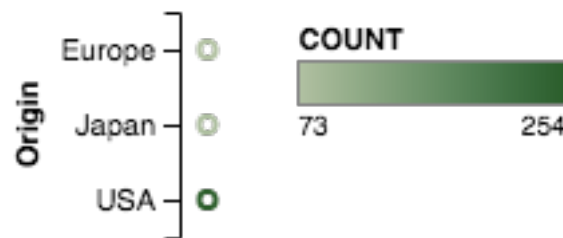
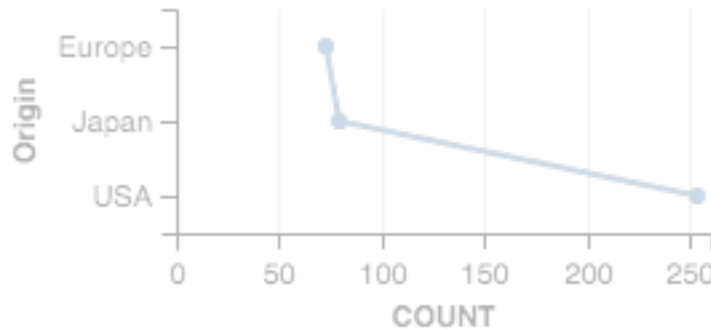
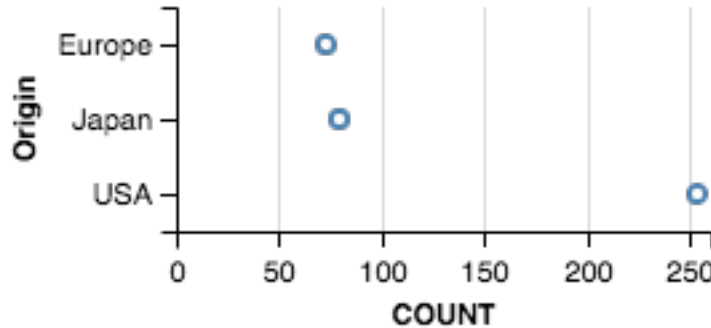


Expressive?

Raw

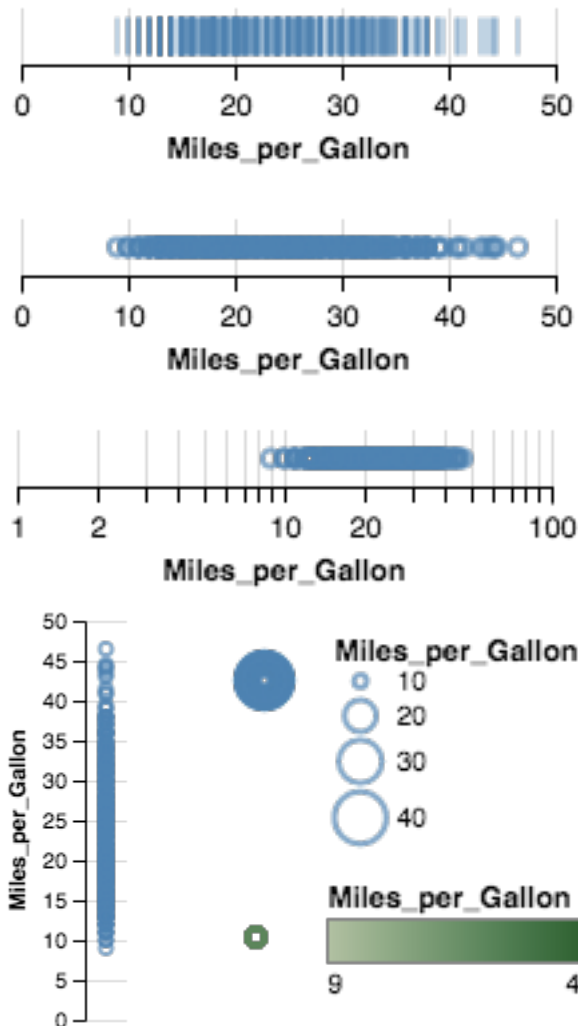


Aggregate (Count)

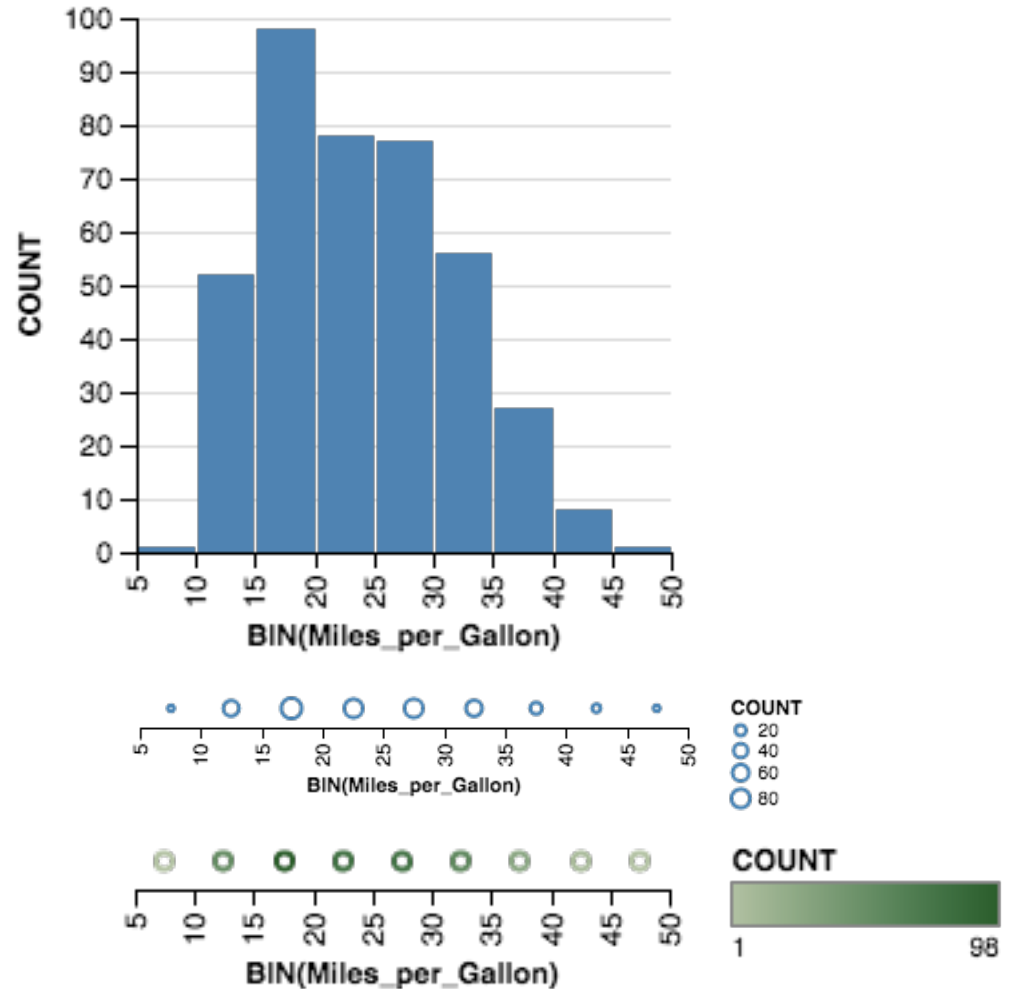


1D: Quantitative

Raw

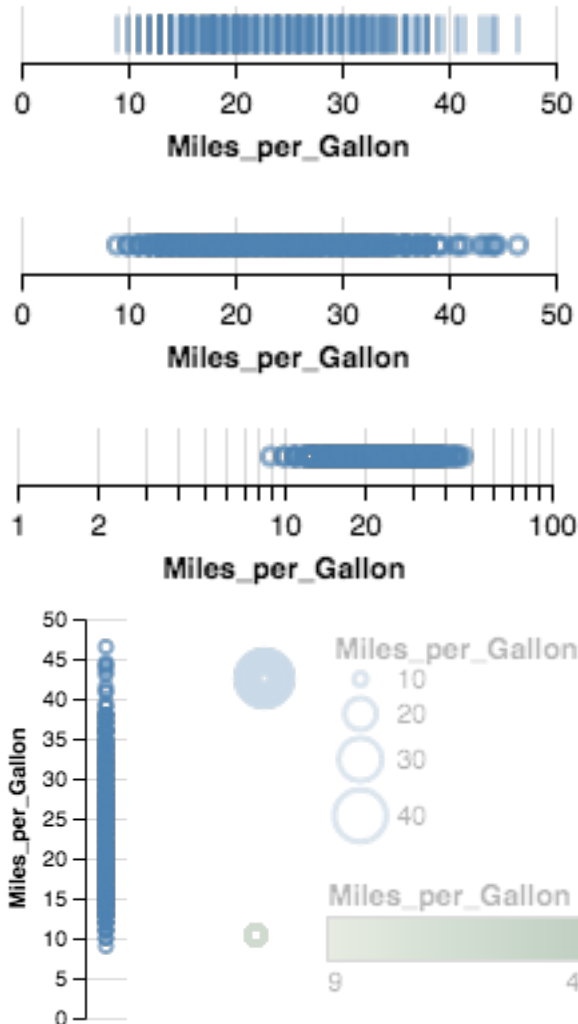


Aggregate (Count)

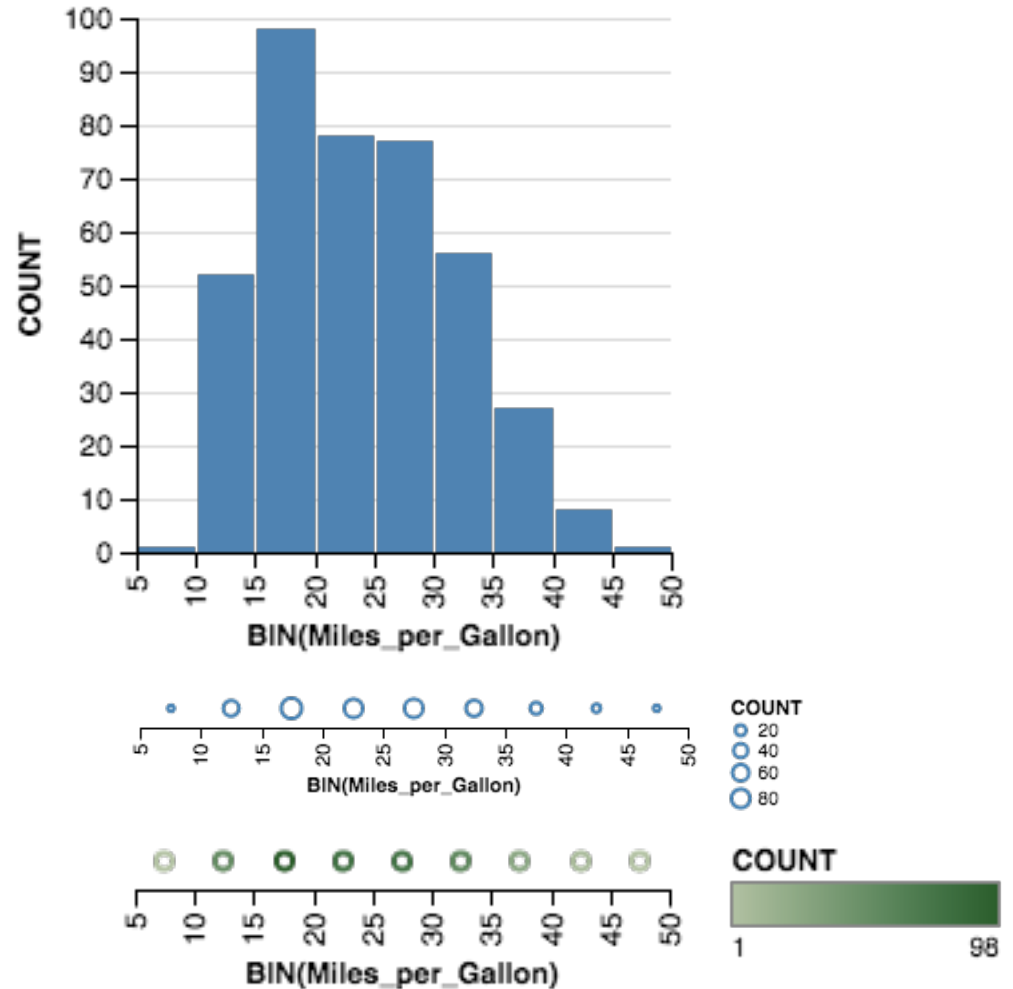


Expressive?

Raw

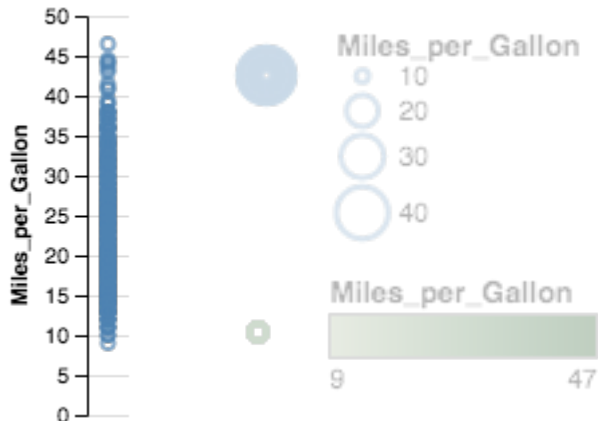
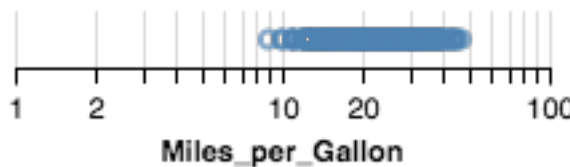
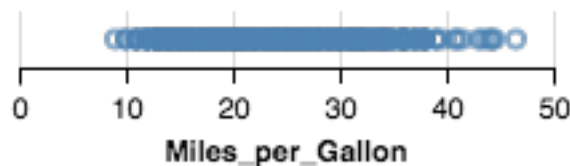
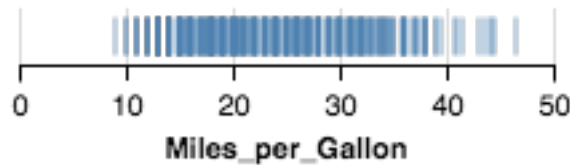


Aggregate (Count)

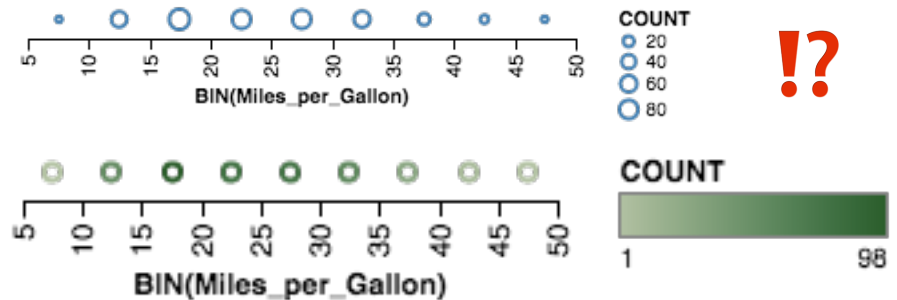
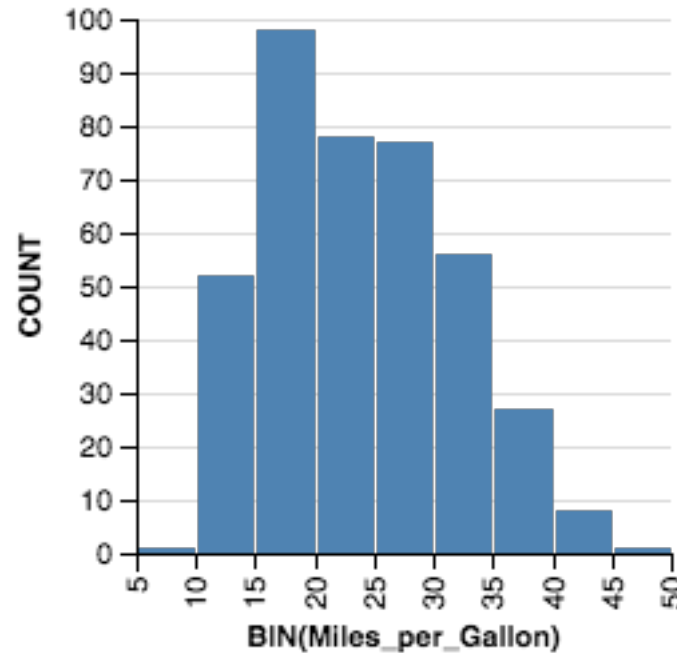


Effective?

Raw



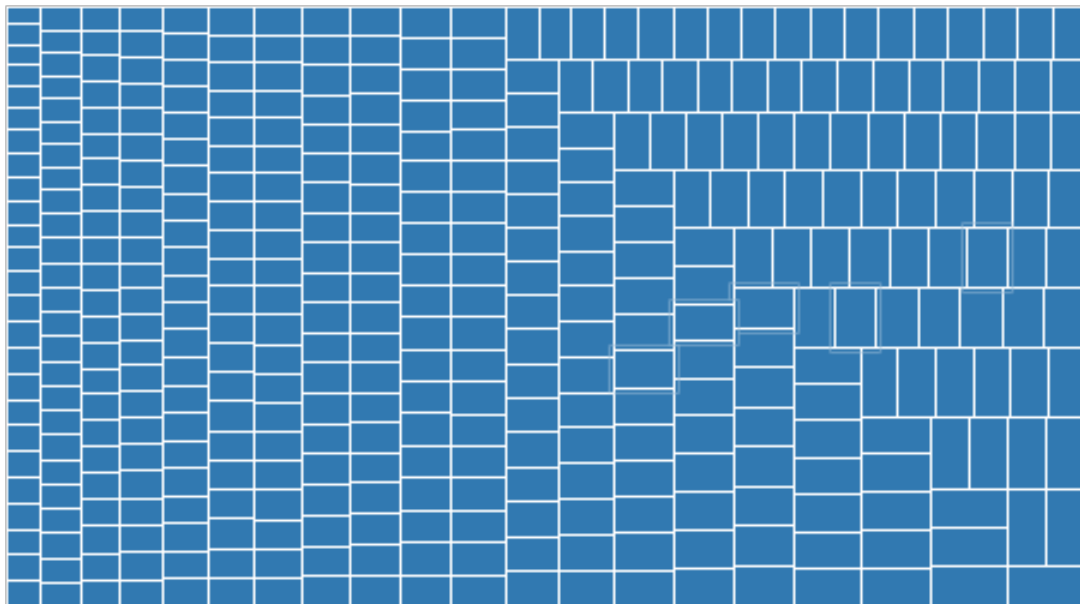
Aggregate (Count)



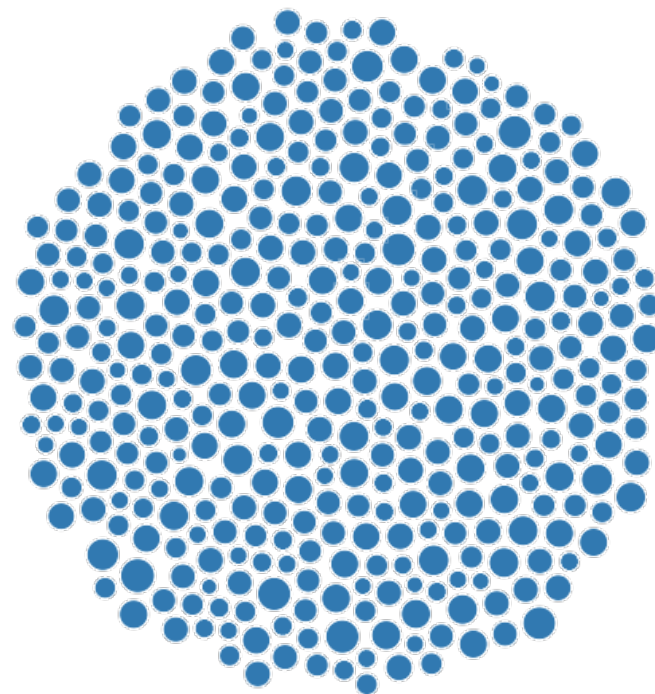
!?



Raw (with Layout Algorithm)

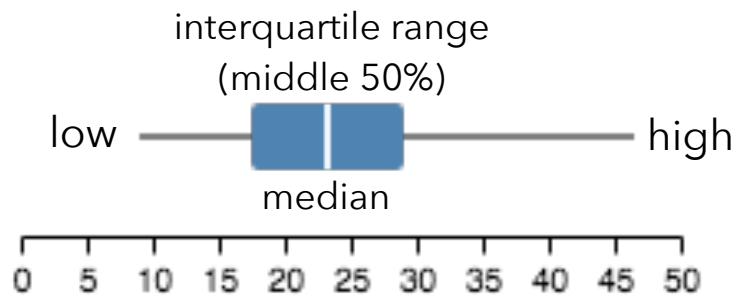


Treemap

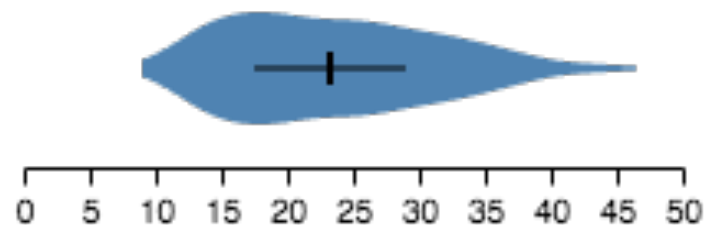


Bubble Chart

Aggregate (Distributions)



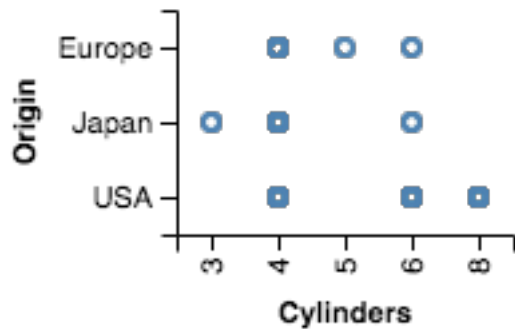
Box Plot



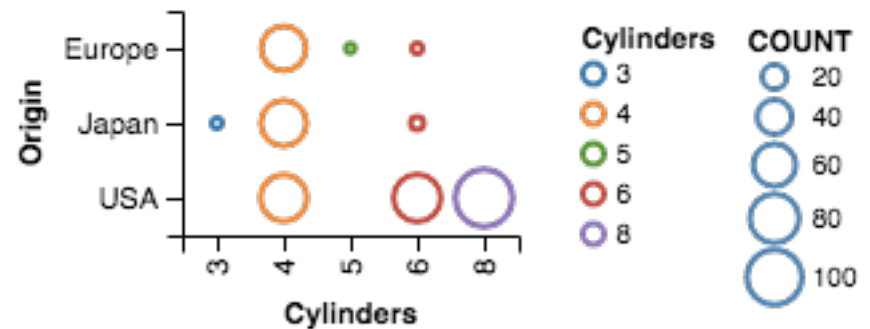
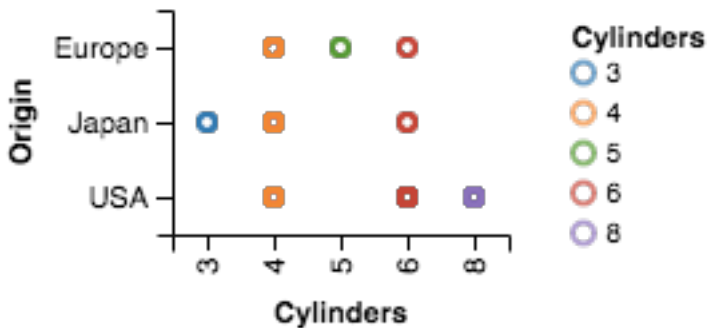
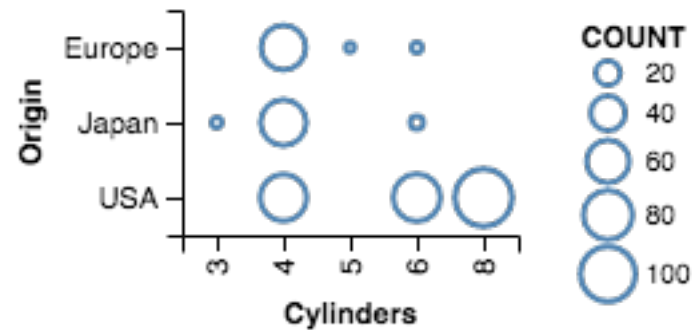
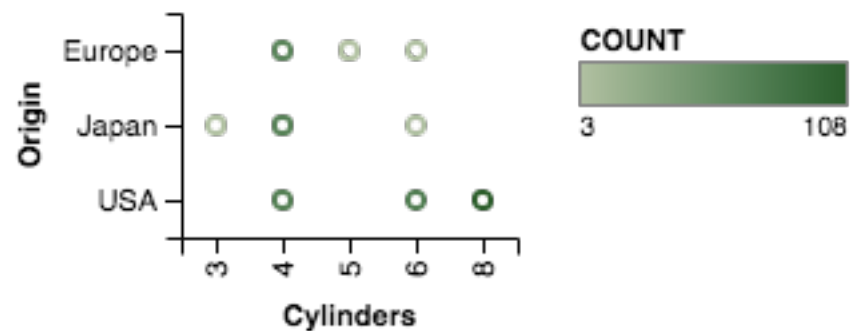
Violin Plot

2D: Nominal x Nominal

Raw

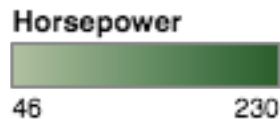
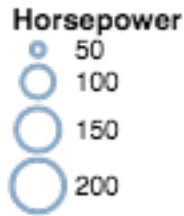
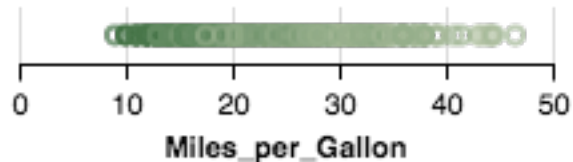
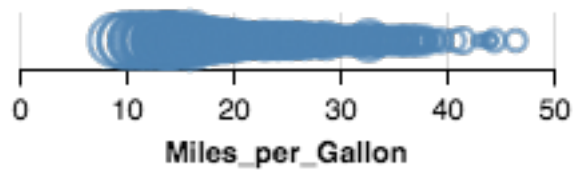
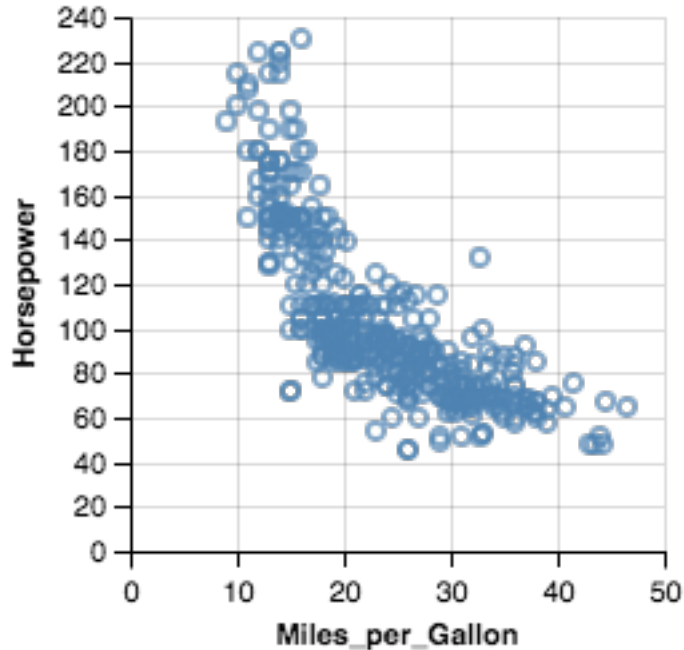


Aggregate (Count)

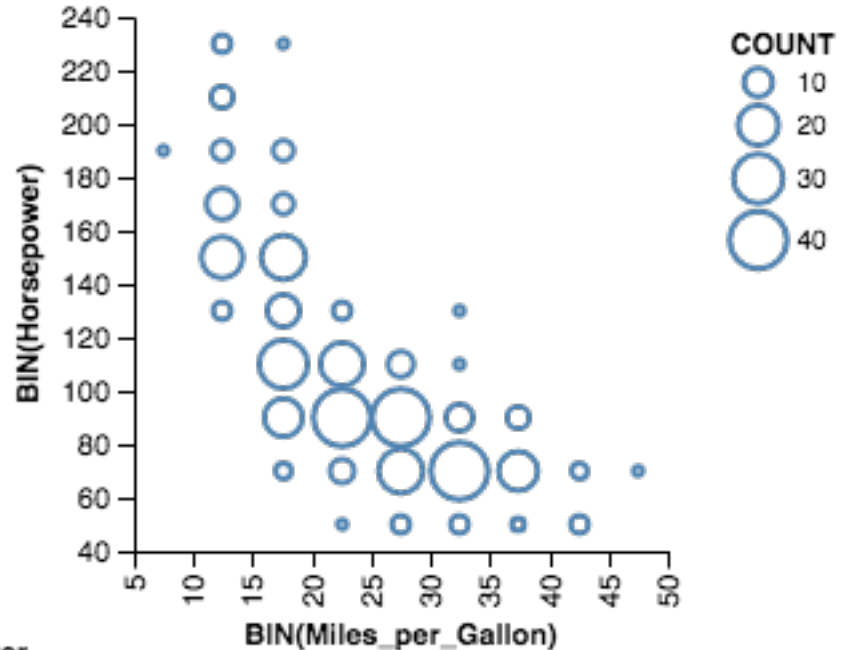


2D: Quantitative x Quantitative

Raw

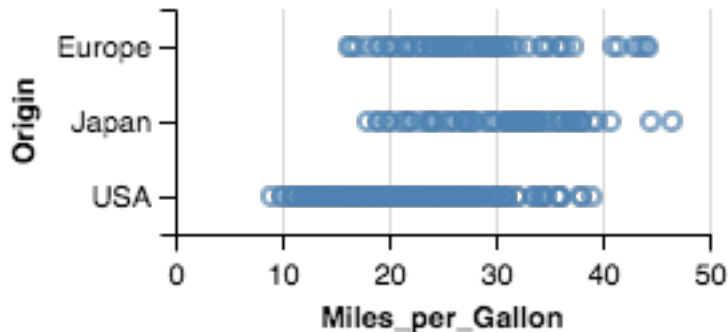


Aggregate (Count)

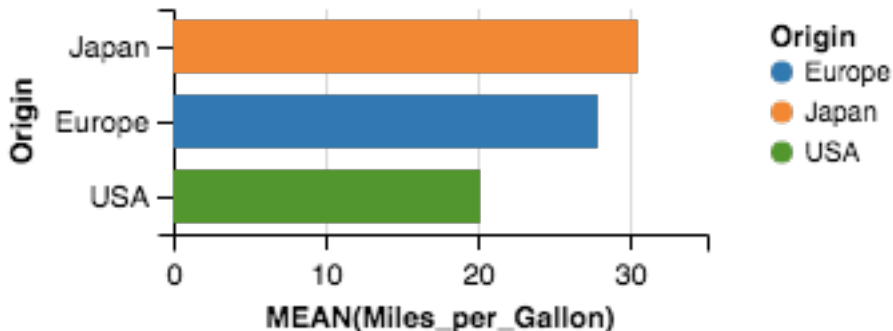
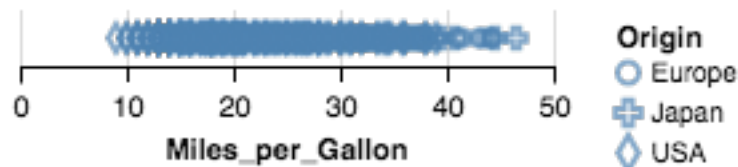
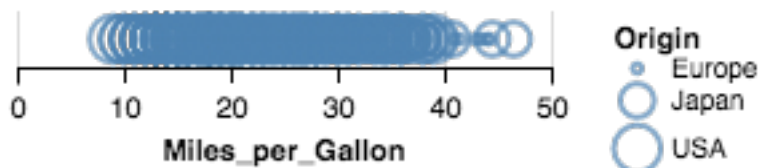
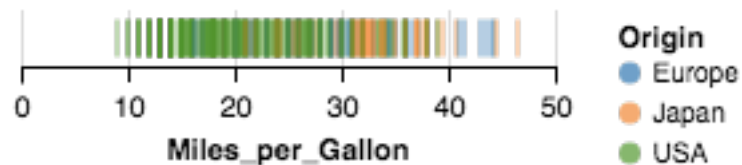
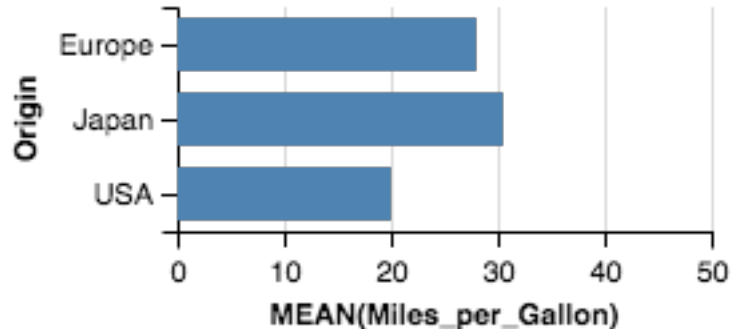


2D: Nominal x Quantitative

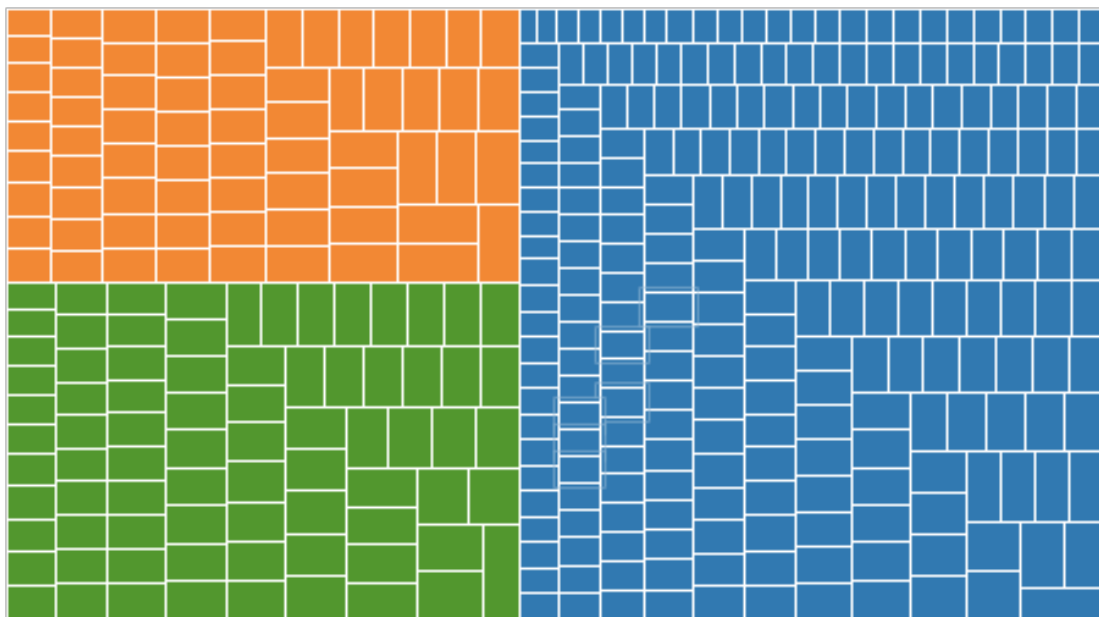
Raw



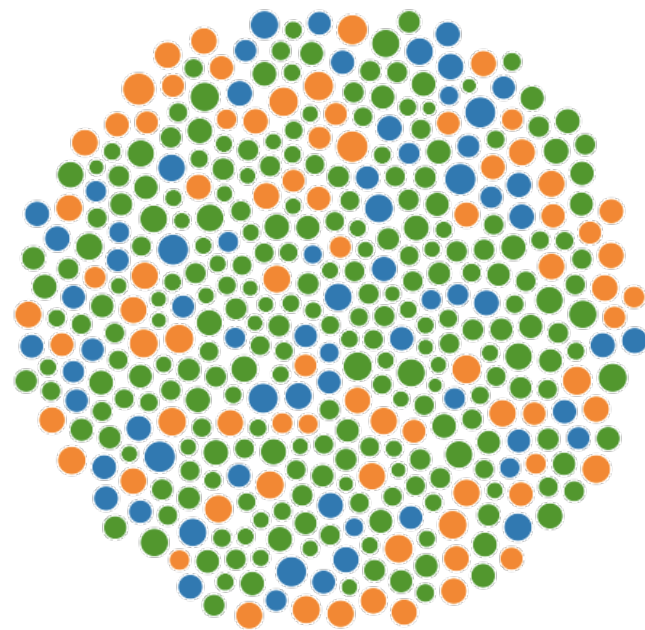
Aggregate (Mean)



Raw (with Layout Algorithm)

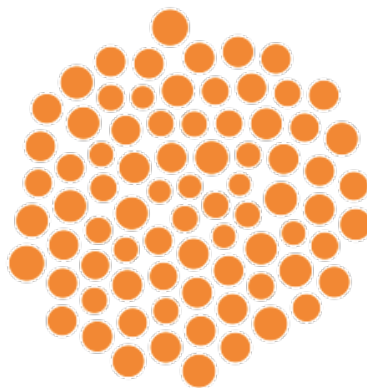
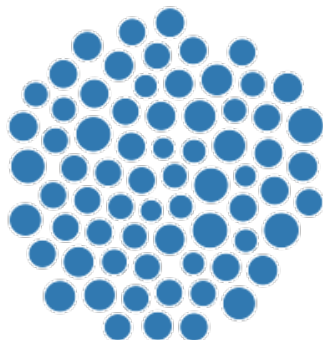


Treemap

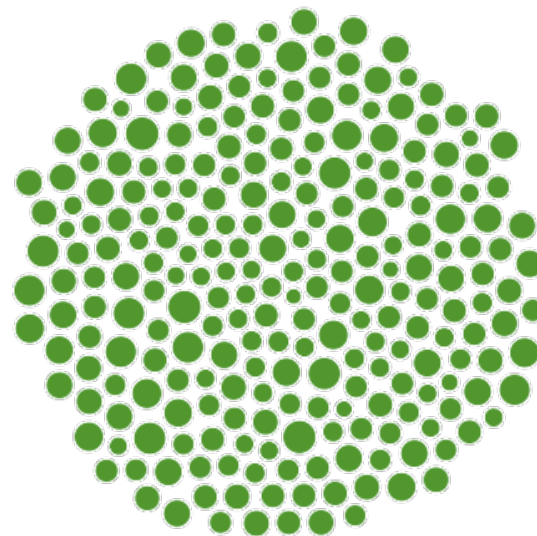


Bubble Chart

Origin
● Europe
● Japan
● USA



Beeswarm Plot



3D and Higher

Two variables $[x, y]$

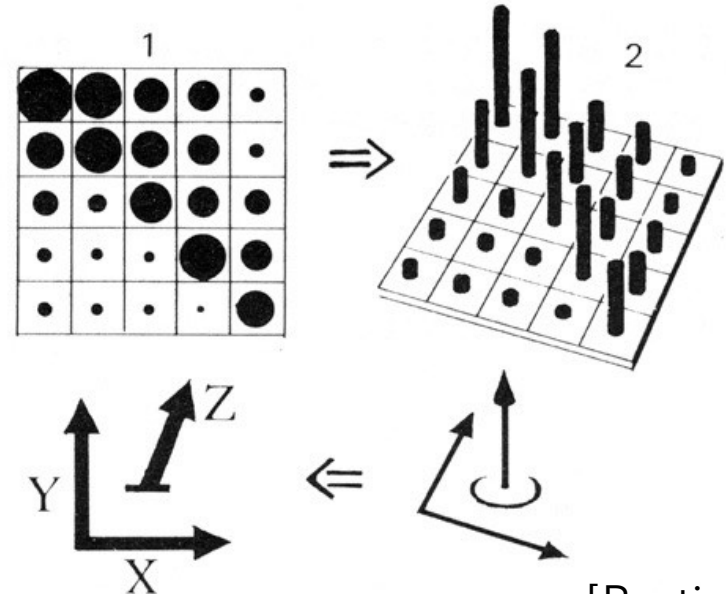
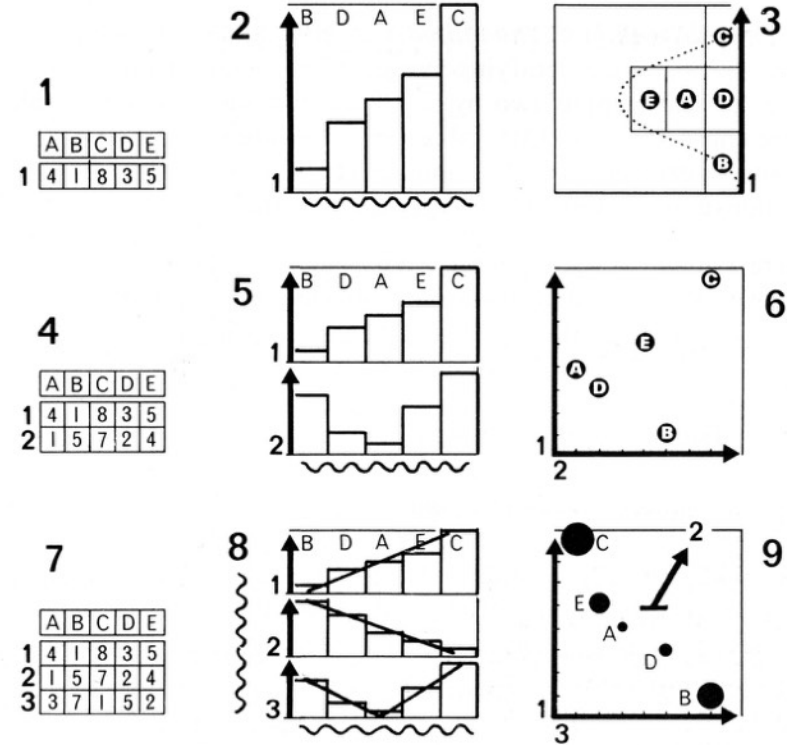
Can map to 2D points.

Scatterplots, maps, ...

Third variable $[z]$

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

What about 3D rendering?



Other Visual Encoding Channels?

wind map

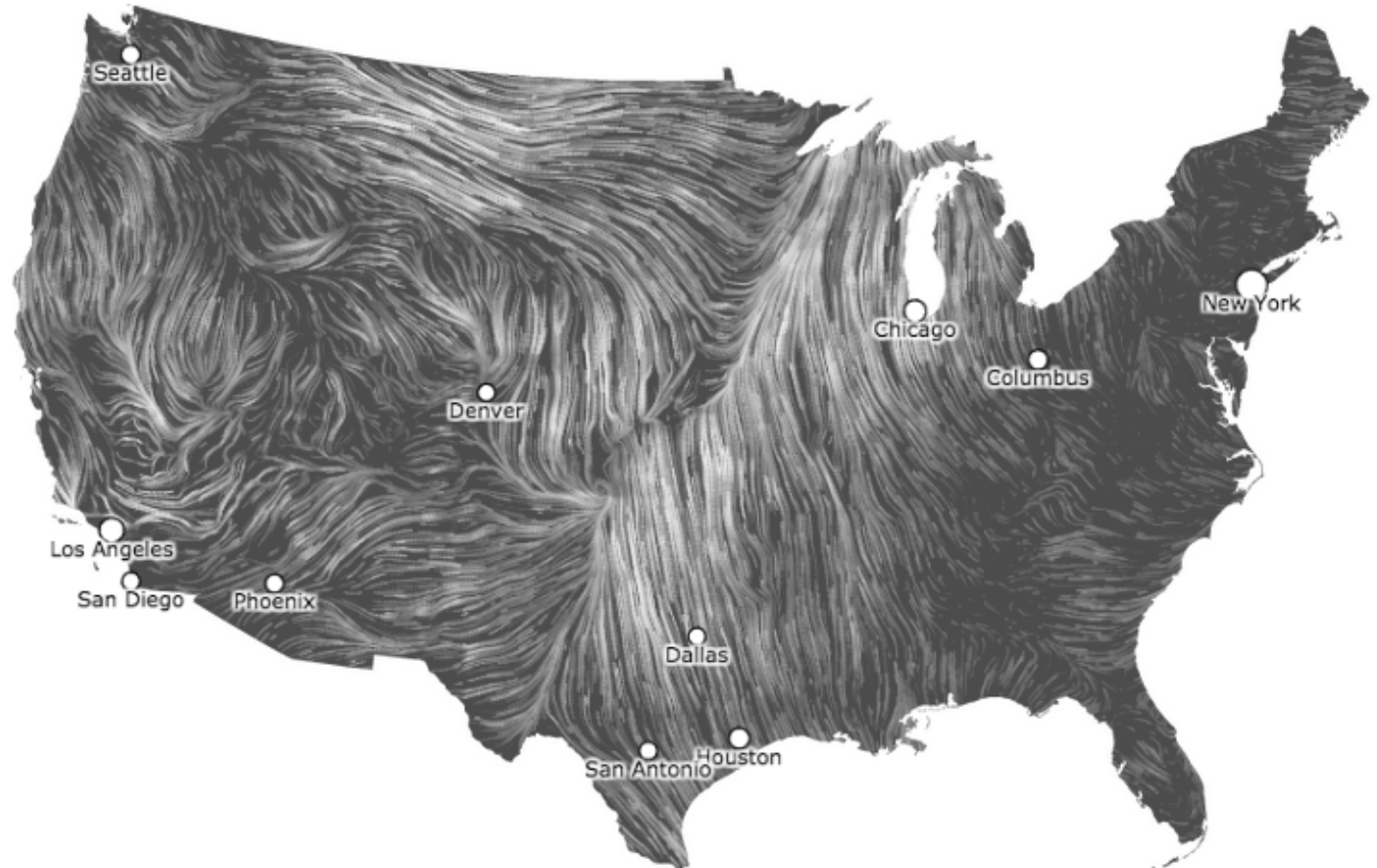
April 1, 2015

11:35 pm EST

(time of forecast download)

top speed: **30.5 mph**

average: **10.2 mph**



Encoding Effectiveness

Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE

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

ORDINAL

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

NOMINAL

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

Effectiveness Rankings [Mackinlay 86]

QUANTITATIVE

Position

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

ORDINAL

Position

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

NOMINAL

Position

Color Hue
Texture
Connection
Containment
Density (Value)
Color Sat
Shape
Length
Angle
Slope
Area
Volume

Effectiveness Rankings

[Mackinlay 86]

QUANTITATIVE

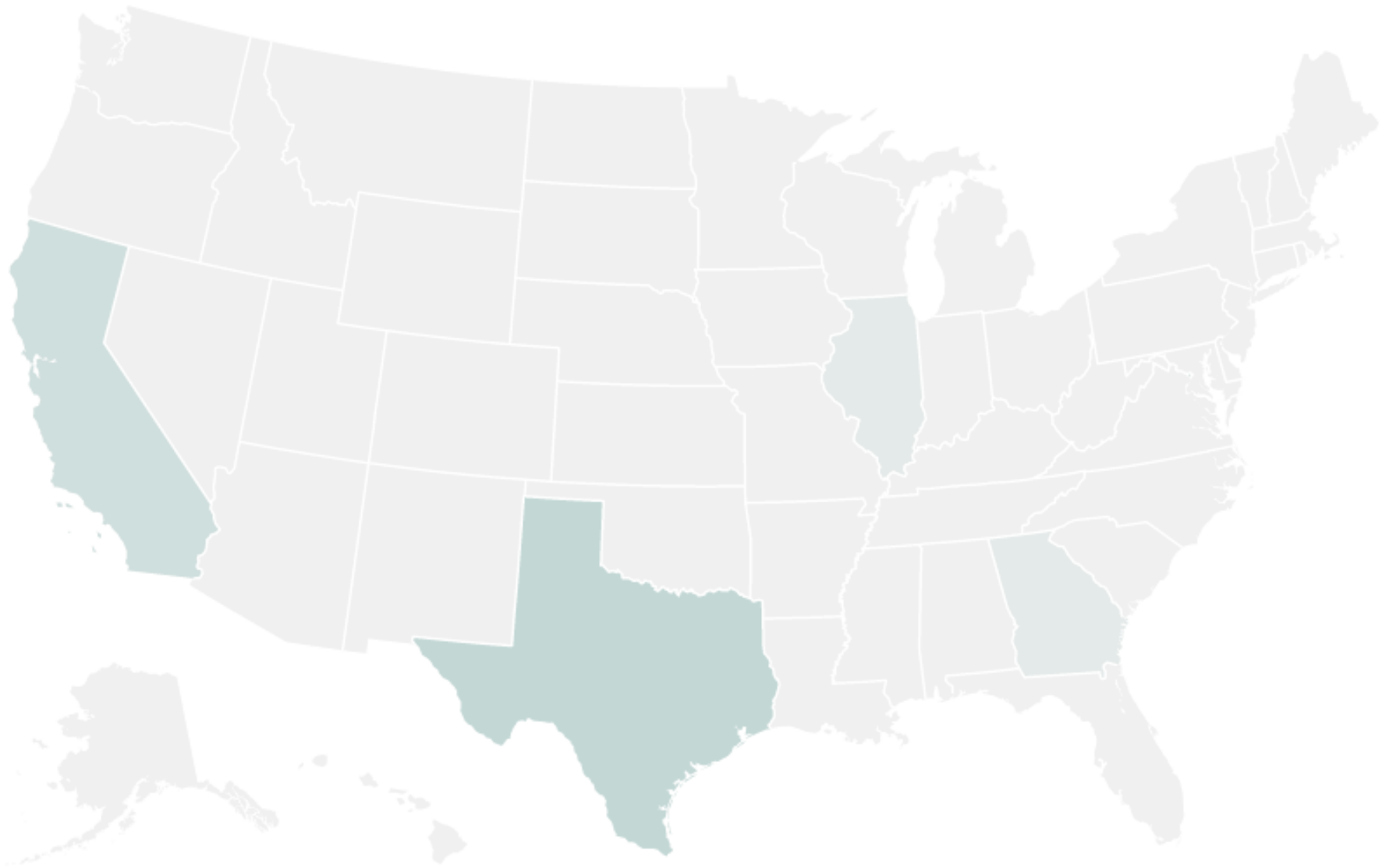
Position
Length
Angle
Slope
Area (Size)
Volume
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Shape

ORDINAL

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

NOMINAL

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



Color Encoding (Choropleth Map)

Effectiveness Rankings

QUANTITATIVE

Position

Length

Angle

Slope

Area (Size)

Volume

Density (Value)

Color Sat

Color Hue

Texture

Connection

Containment

Shape

ORDINAL

Position

Density (Value)

Color Sat

Color Hue

Texture

Connection

Containment

Length

Angle

Slope

Area (Size)

Volume

Shape

NOMINAL

Position

Color Hue

Texture

Connection

Containment

Density (Value)

Color Sat

Shape

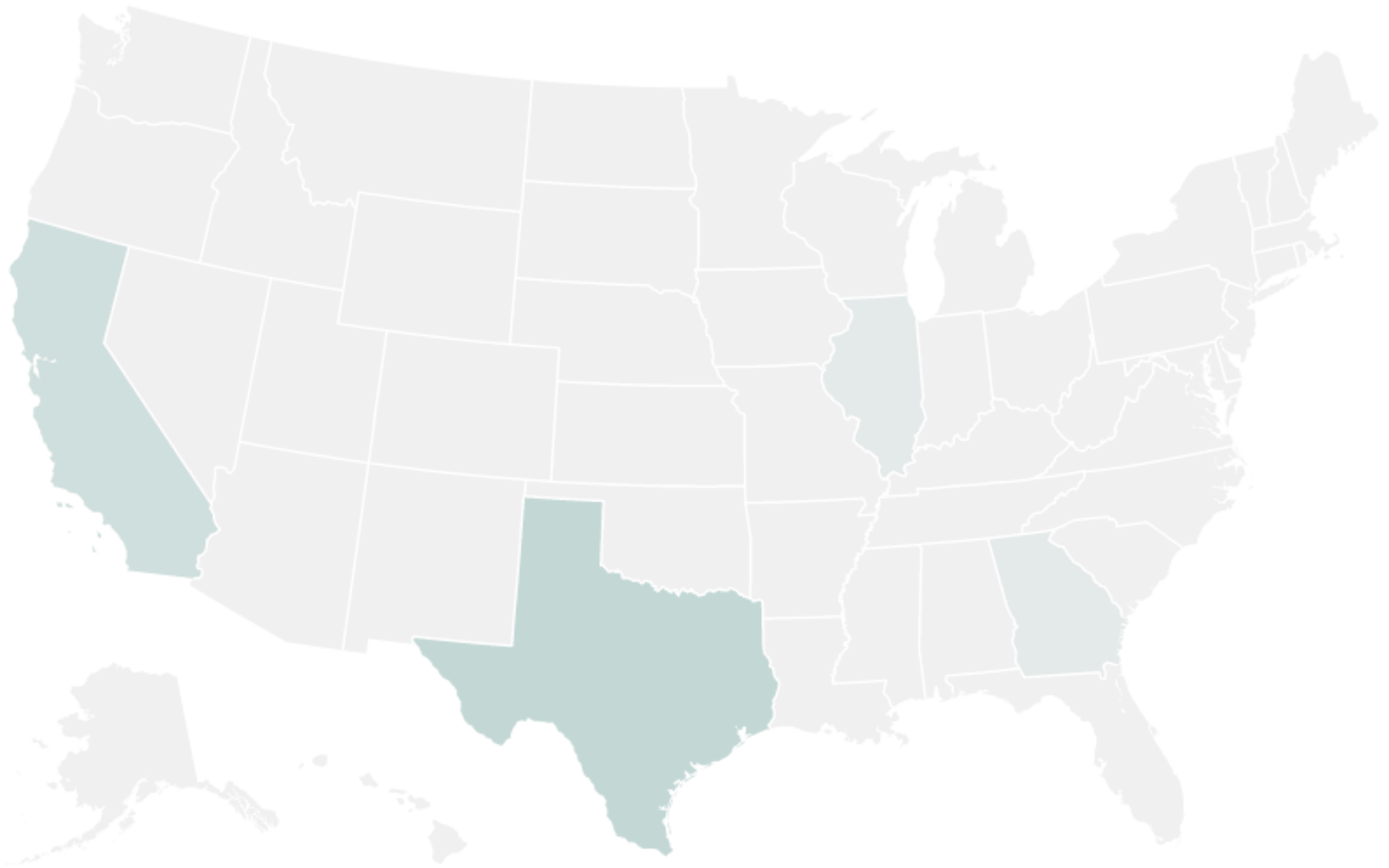
Length

Angle

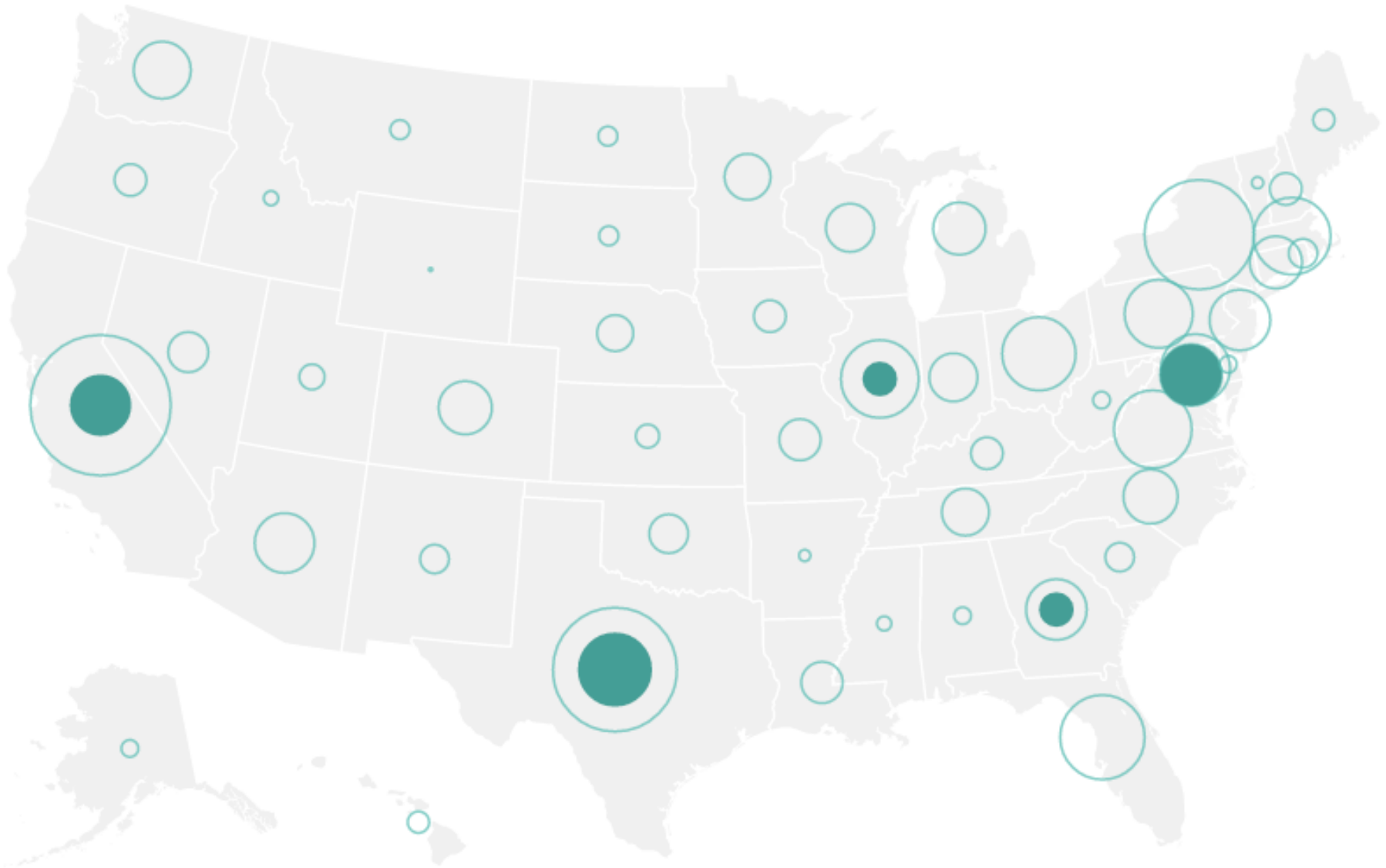
Slope

Area

Volume



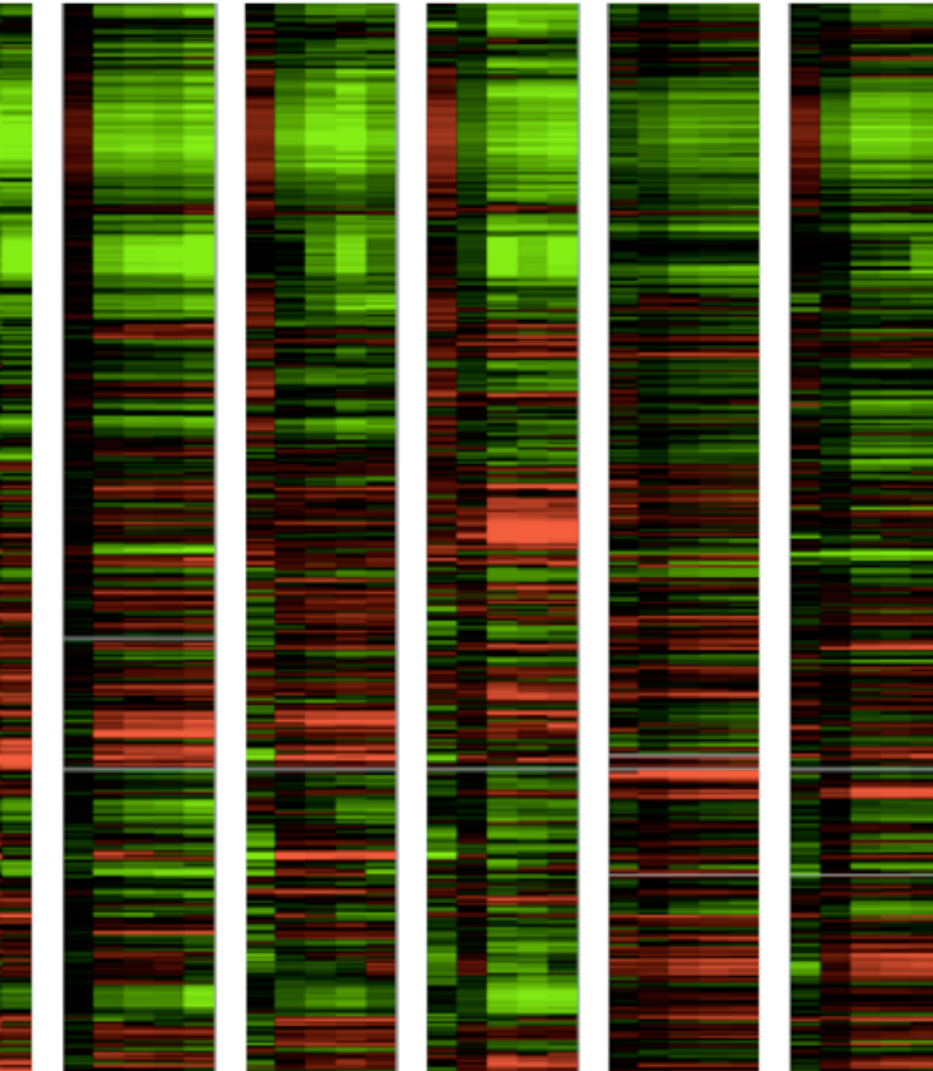
Color Encoding (Choropleth Map)



Area Encoding (Symbol Map)

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

Color Encoding



Effectiveness Rankings

QUANTITATIVE

Position

Length

Angle

Slope

Area (Size)

Volume

~~Density (Value)~~

Color Sat

~~Color Hue~~

Texture

Connection

Containment

Shape

ORDINAL

Position

Density (Value)

Color Sat

Color Hue

Texture

Connection

Containment

Length

Angle

Slope

Area (Size)

Volume

Shape

NOMINAL

Position

Color Hue

Texture

Connection

Containment

Density (Value)

Color Sat

Shape

Length

Angle

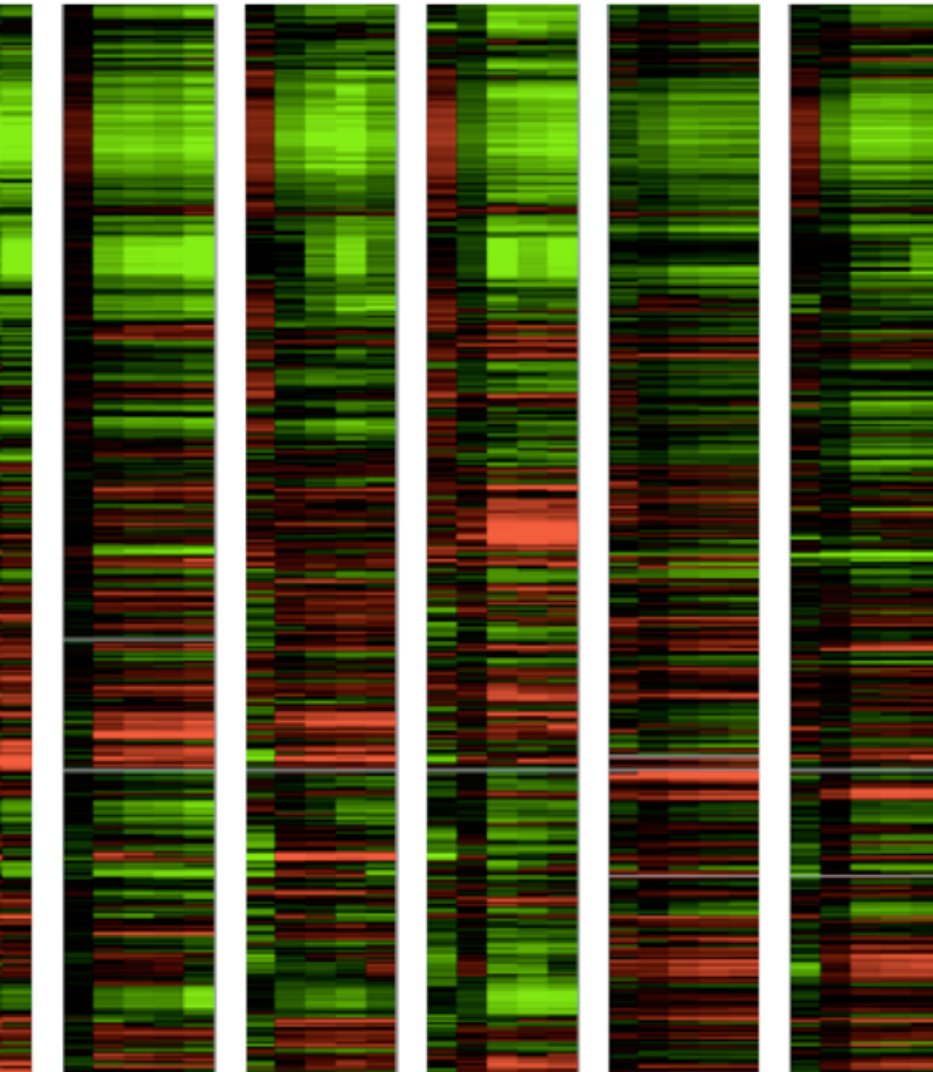
Slope

Area

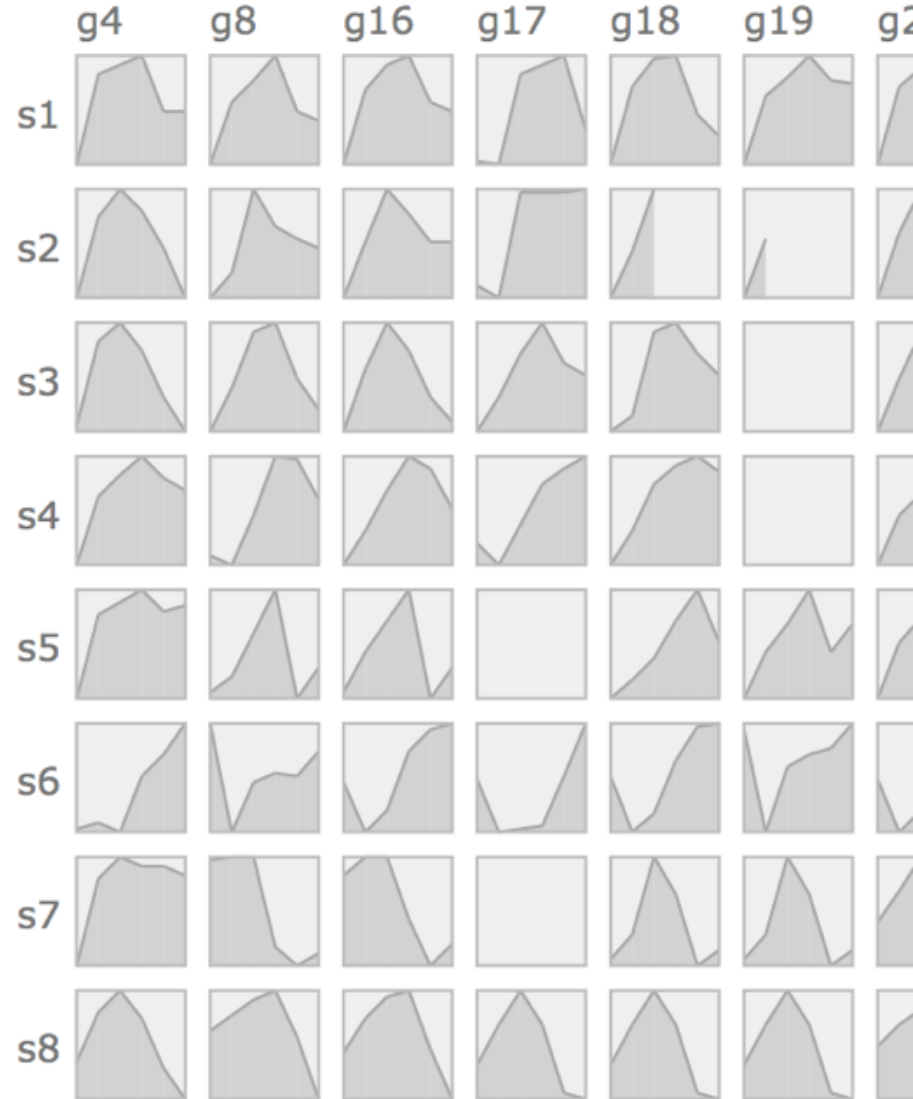
Volume

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

Color Encoding



Position Encoding

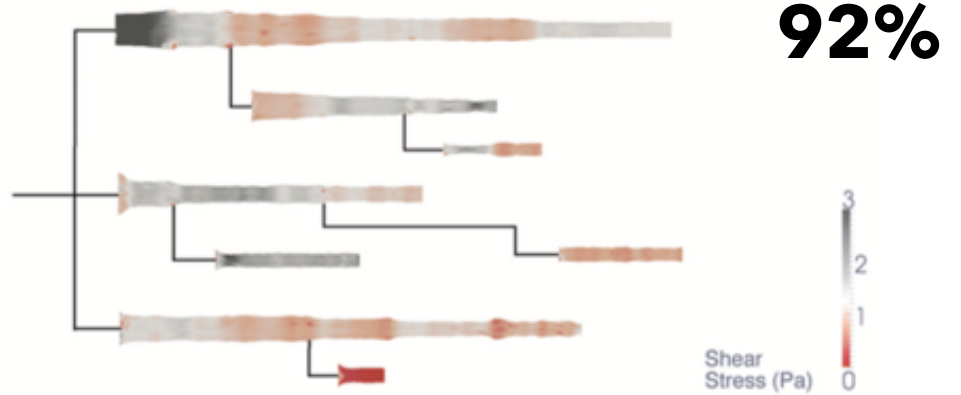


Artery Visualization [Borkin et al '11]

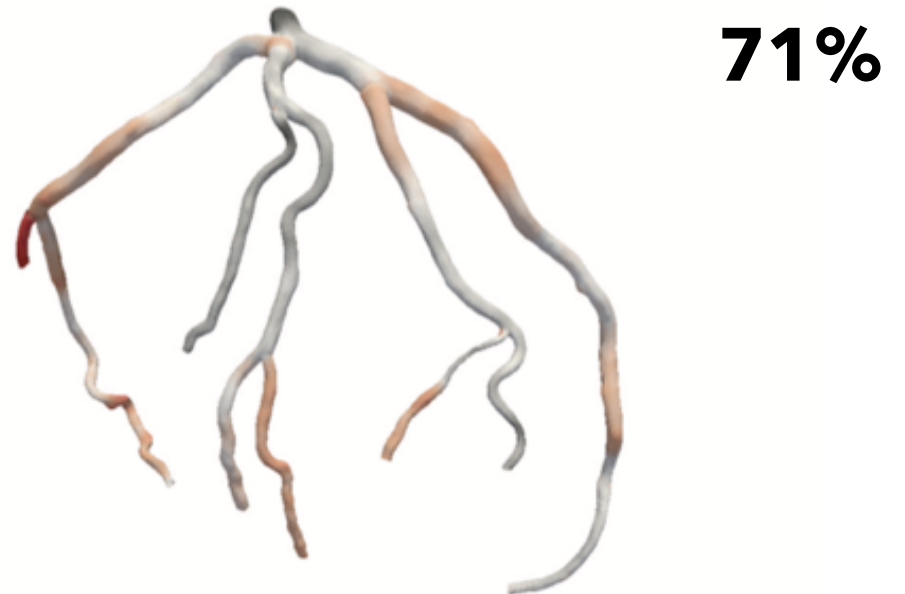
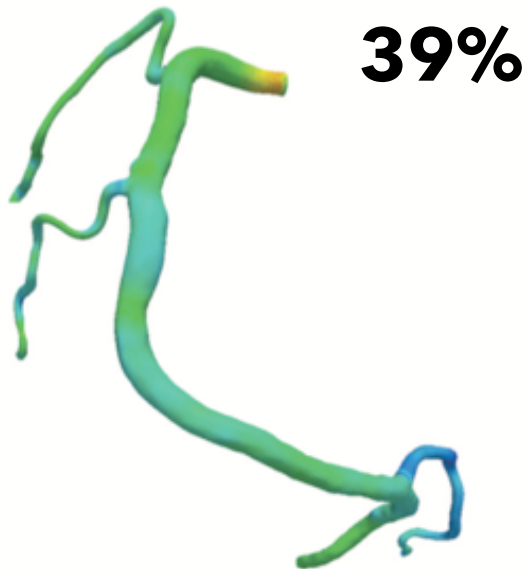
Rainbow Palette

Diverging Palette

2D



3D



Effectiveness Rankings

QUANTITATIVE

Position ↻

Length

Angle

Slope

Area (Size)

Volume

Density (Value)

Color Sat

Color Hue

Texture

Connection

Containment

Shape

ORDINAL

Position

Density (Value)

Color Sat

Color Hue

Texture

Connection

Containment

Length

Angle

Slope

Area (Size)

Volume

Shape

NOMINAL

Position

Color Hue

Texture

Connection

Containment

Density (Value)

Color Sat

Shape

Length

Angle

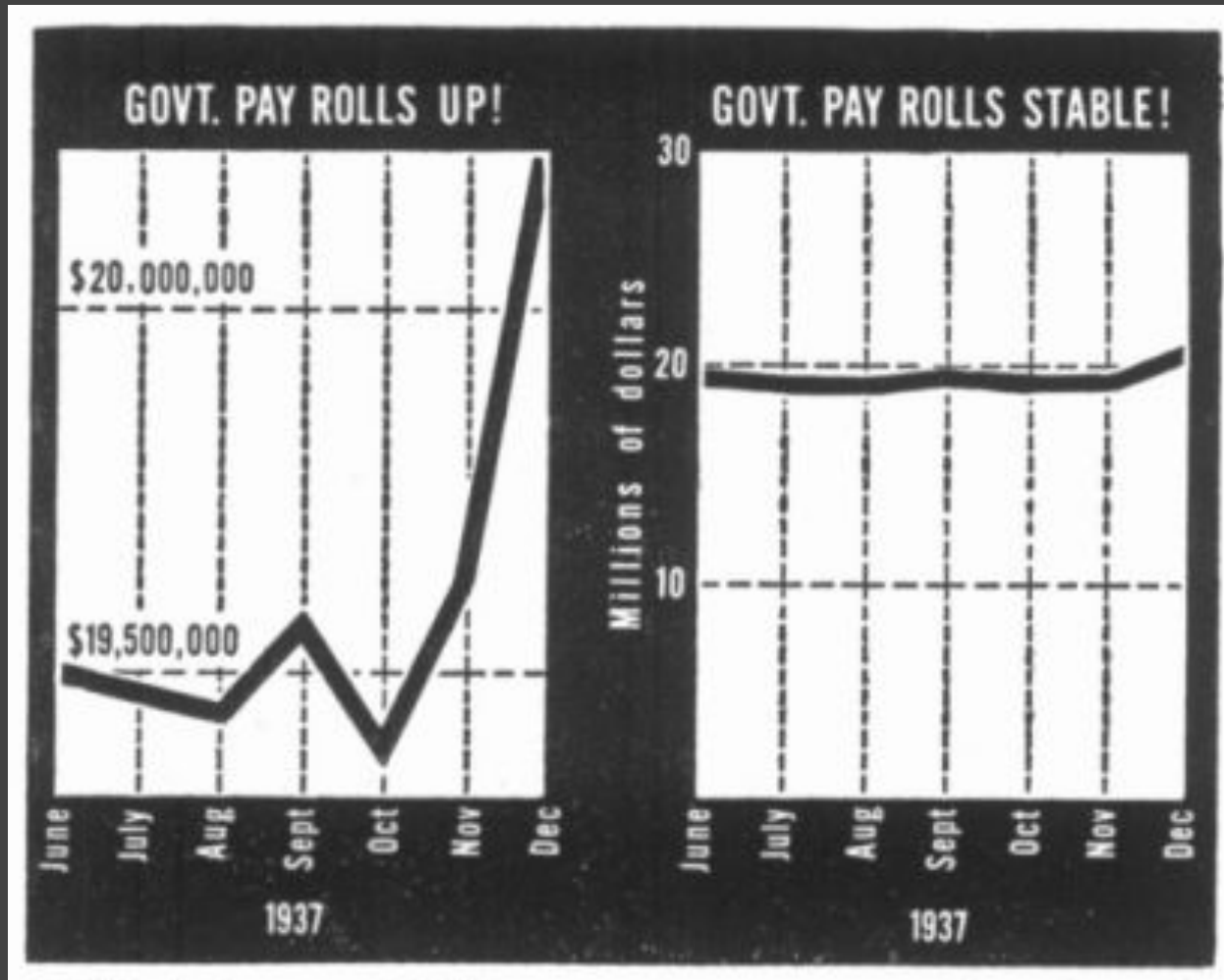
Slope

Area

Volume

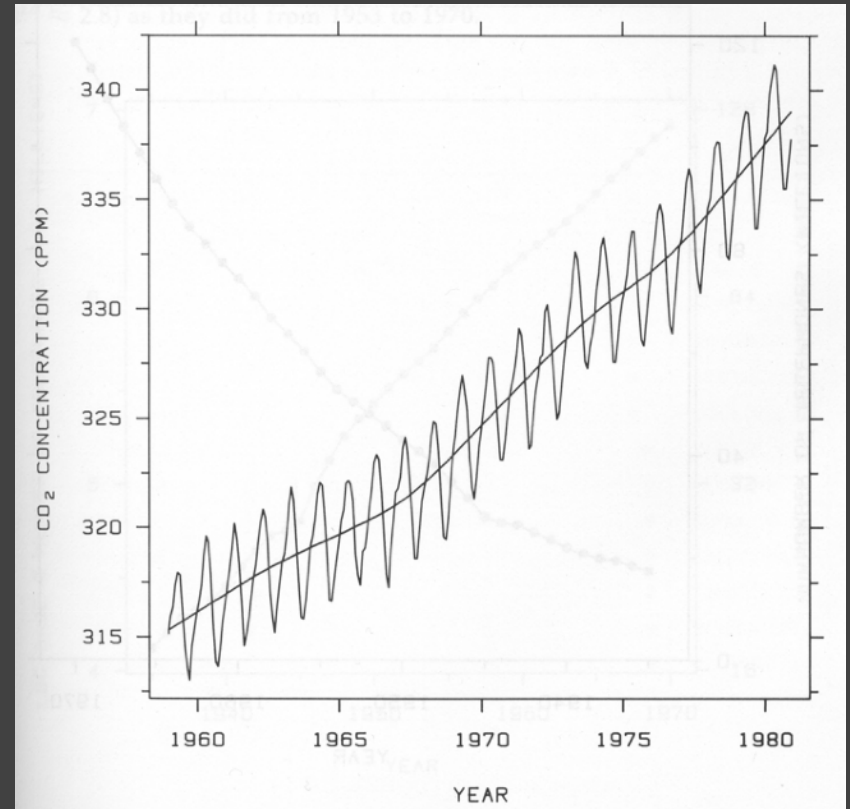
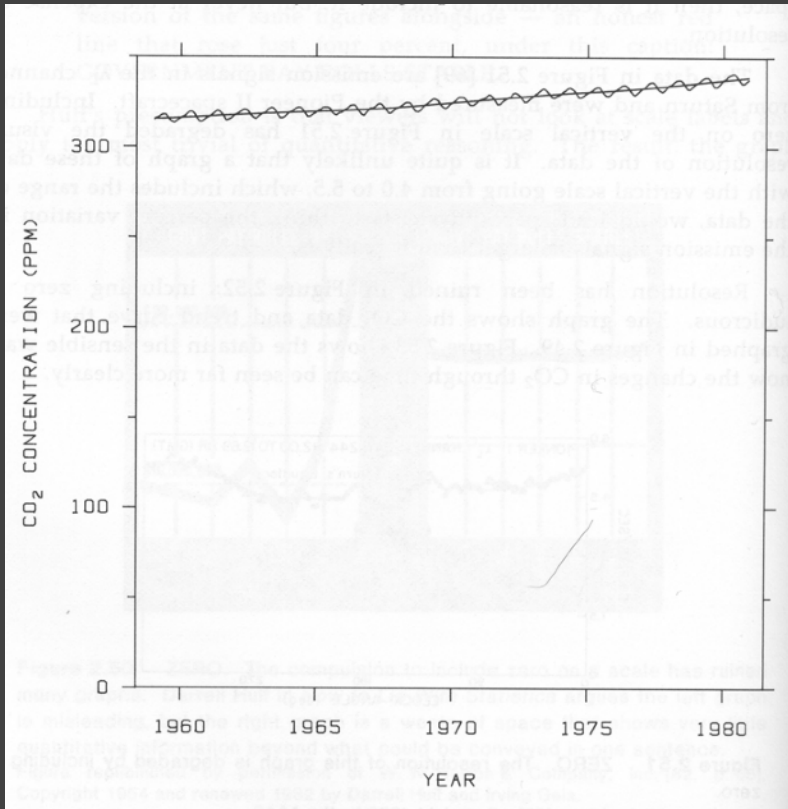
Scales & Axes

Include Zero in Axis Scale?



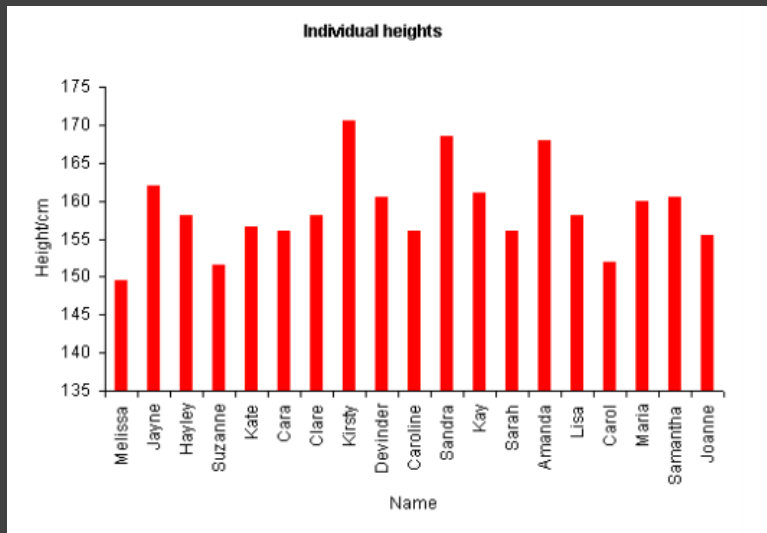
Government payrolls in 1937 [How To Lie With Statistics. Huff]

Include Zero in Axis Scale?



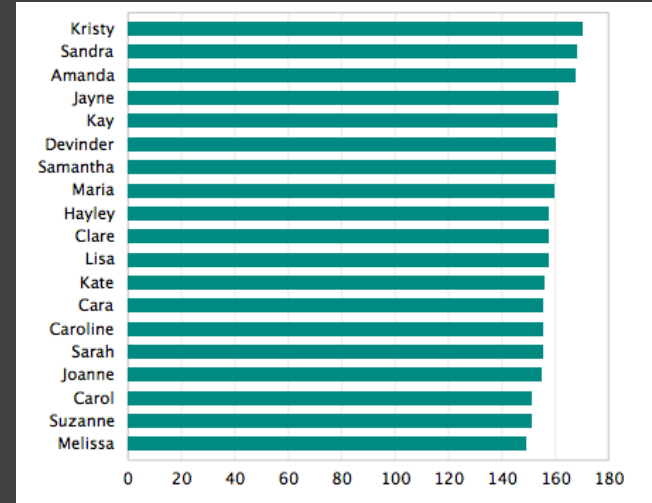
Yearly CO₂ concentrations [Cleveland 85]

Include Zero in Axis Scale?

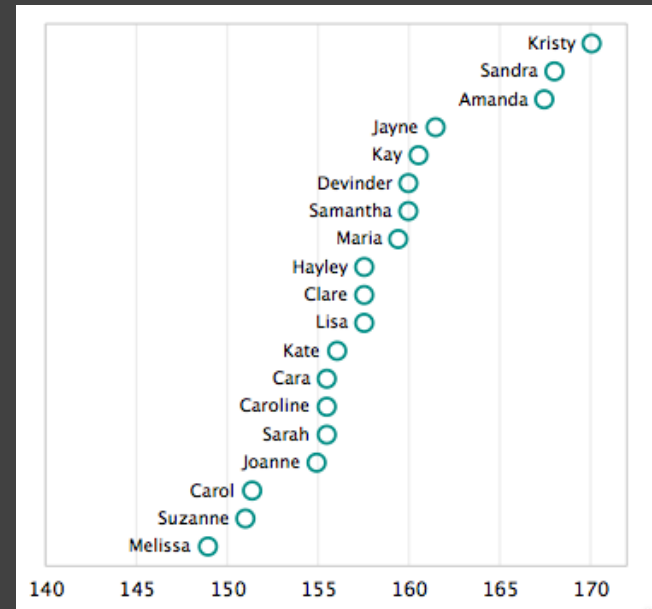


Violates Expressiveness Principle!

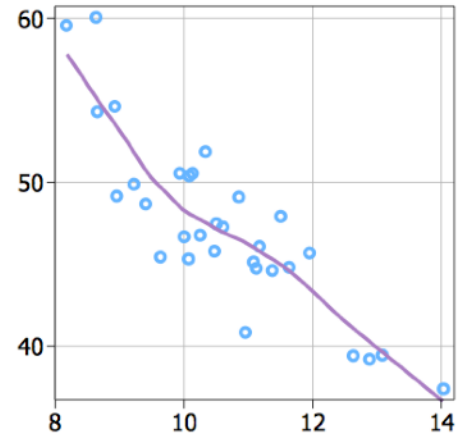
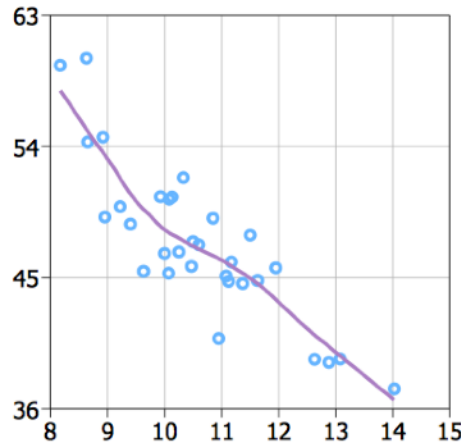
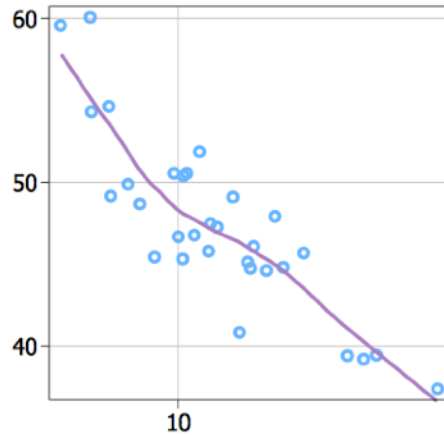
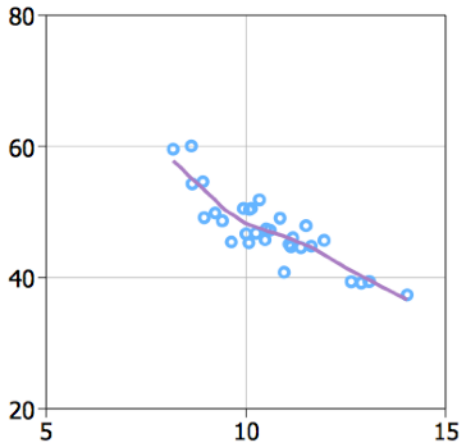
Compare Proportions (Q-Ratio)



Compare Relative Position (Q-Interval)

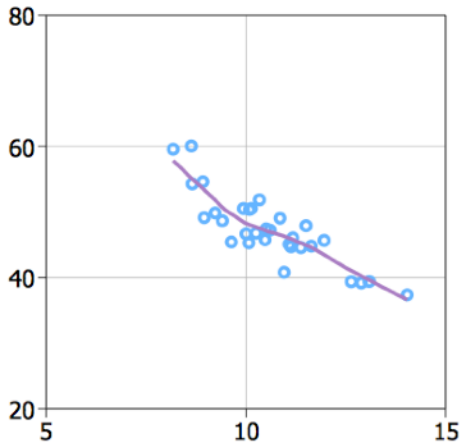


Axis Tick Mark Selection

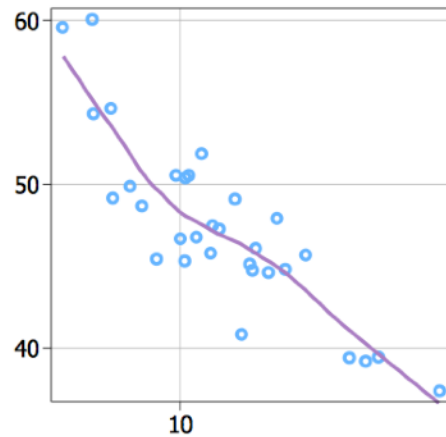


What are some properties of "good" tick marks?

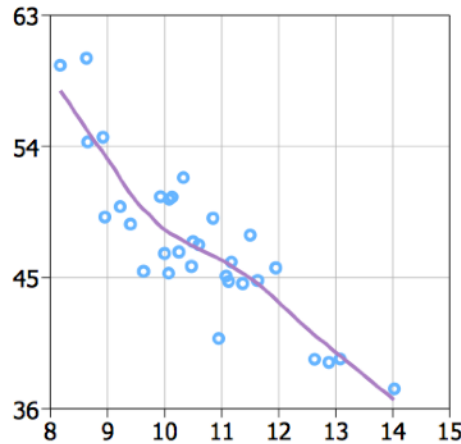
Axis Tick Mark Selection



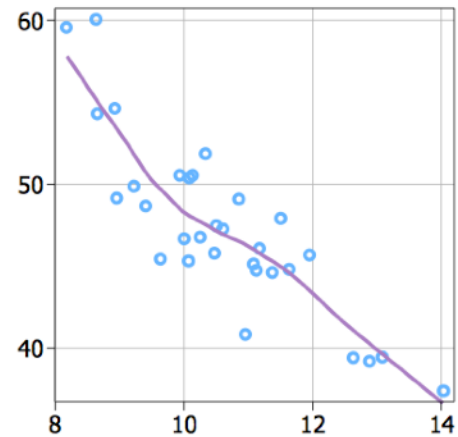
(a) Heckbert



(b) R's pretty



(c) Wilkinson



(d) Extended

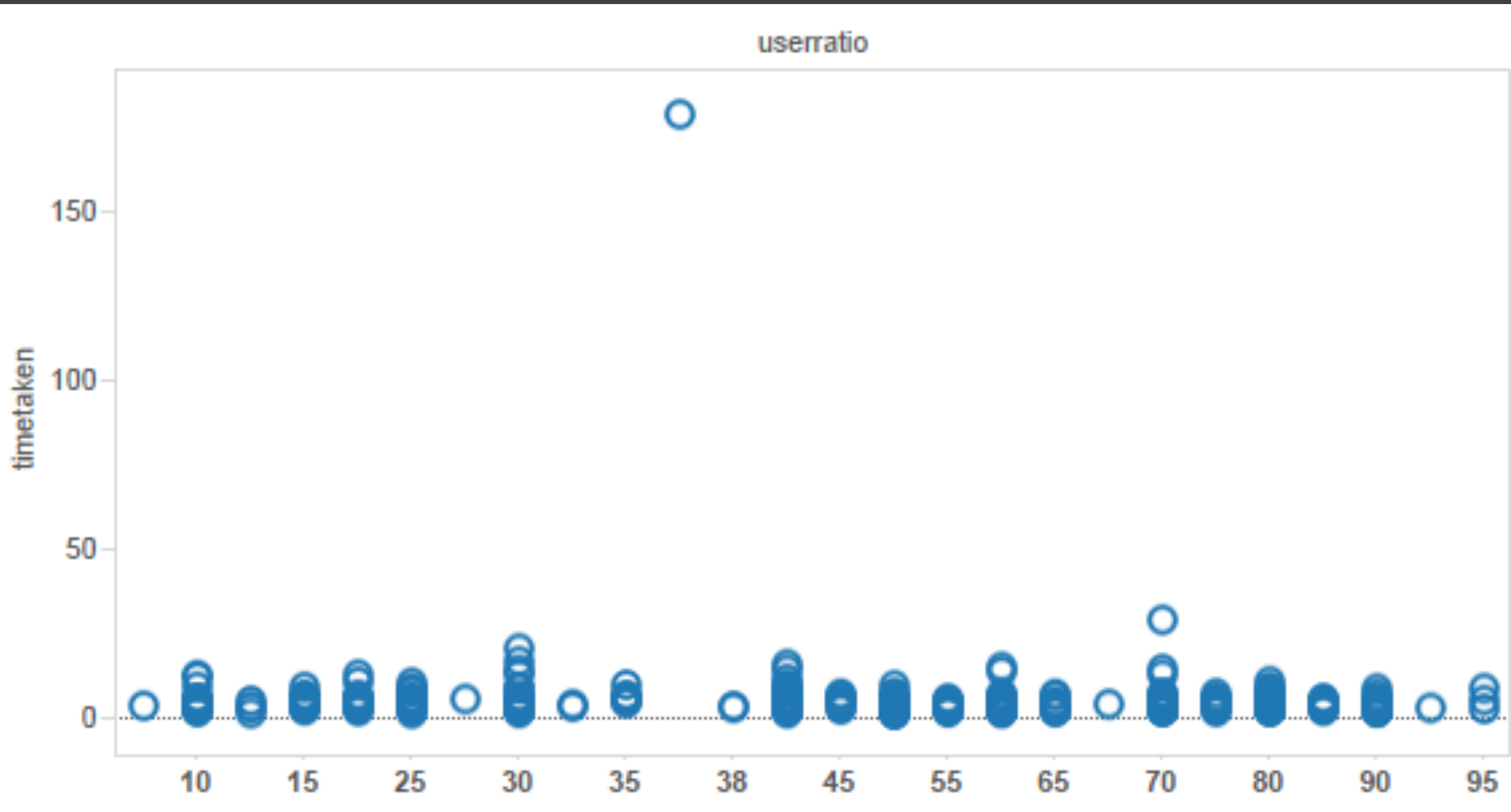
Simplicity - numbers are multiples of 10, 5, 2

Coverage - ticks near the ends of the data

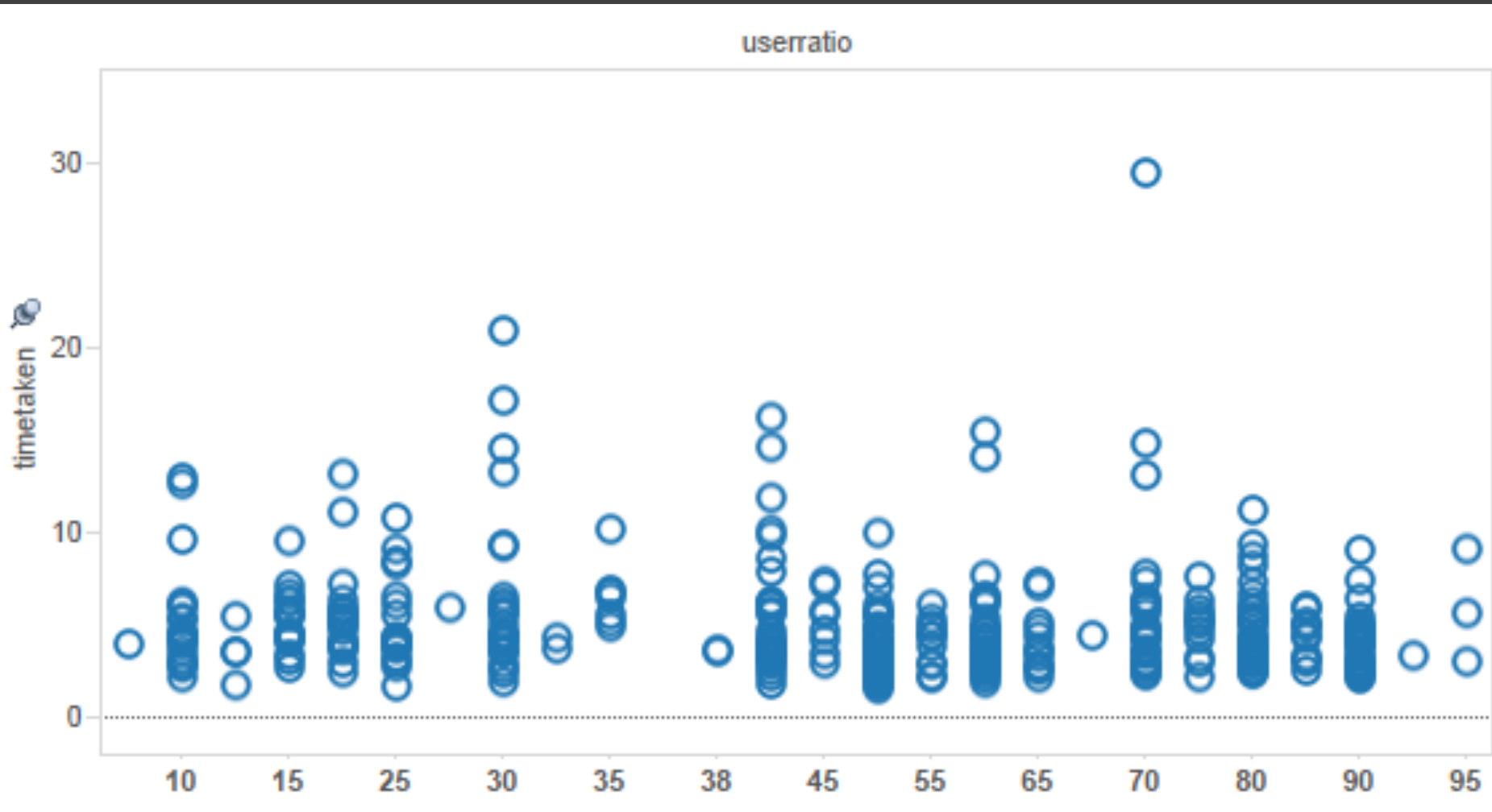
Density - not too many, nor too few

Legibility - whitespace, horizontal text, size

How to Scale the Axis?

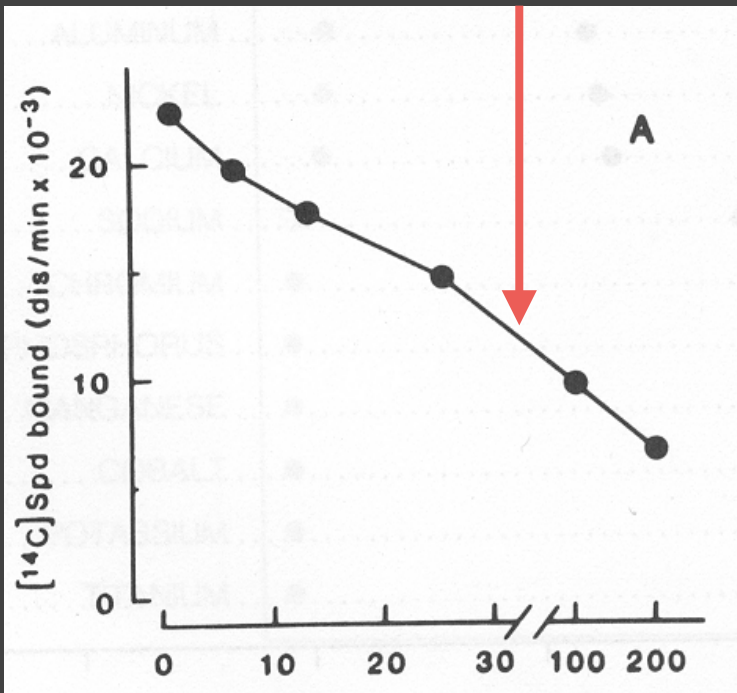


One Option: Clip Outliers

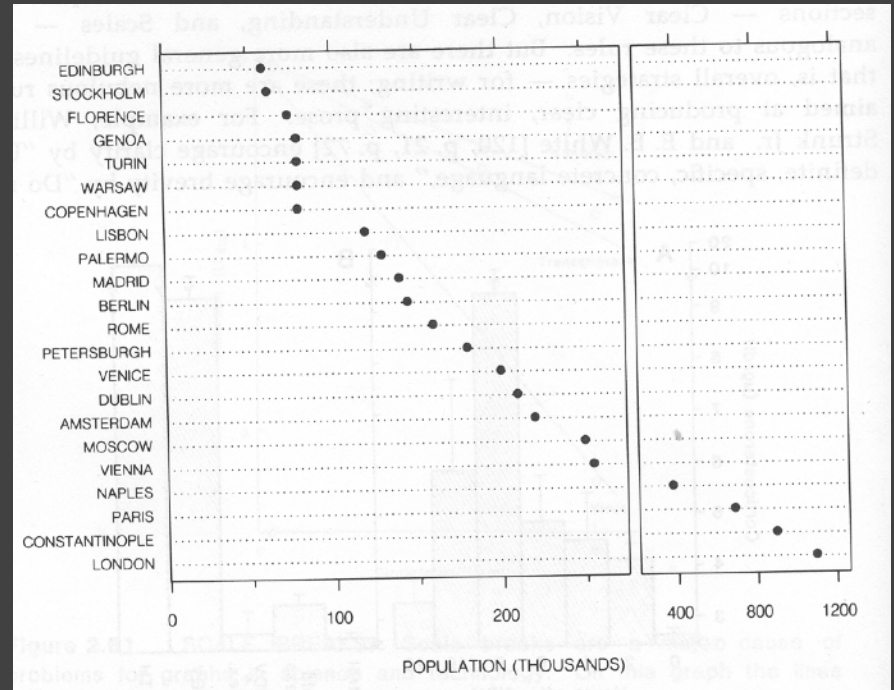


Clearly Mark Scale Breaks

Violates Expressiveness Principle!

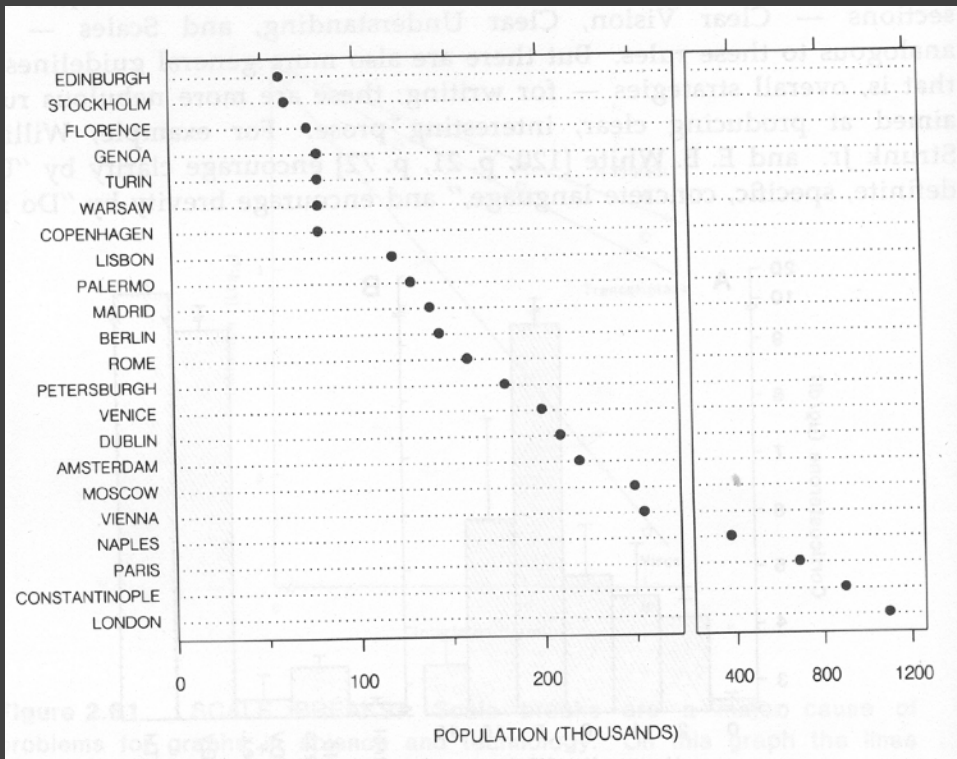


Poor scale break [Cleveland 85]

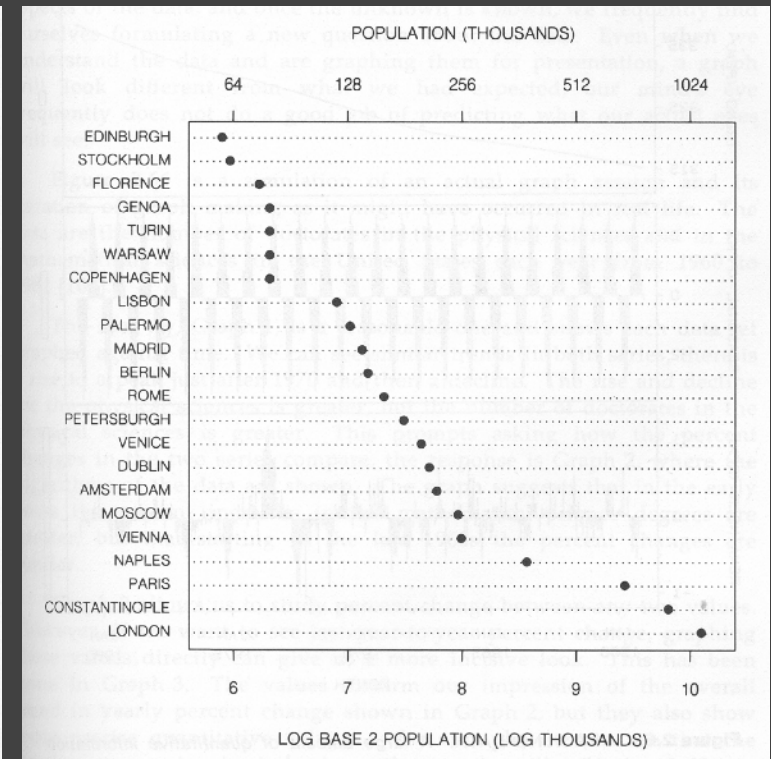


Well-marked scale break [Cleveland 85]

Scale Break vs. Log Scale

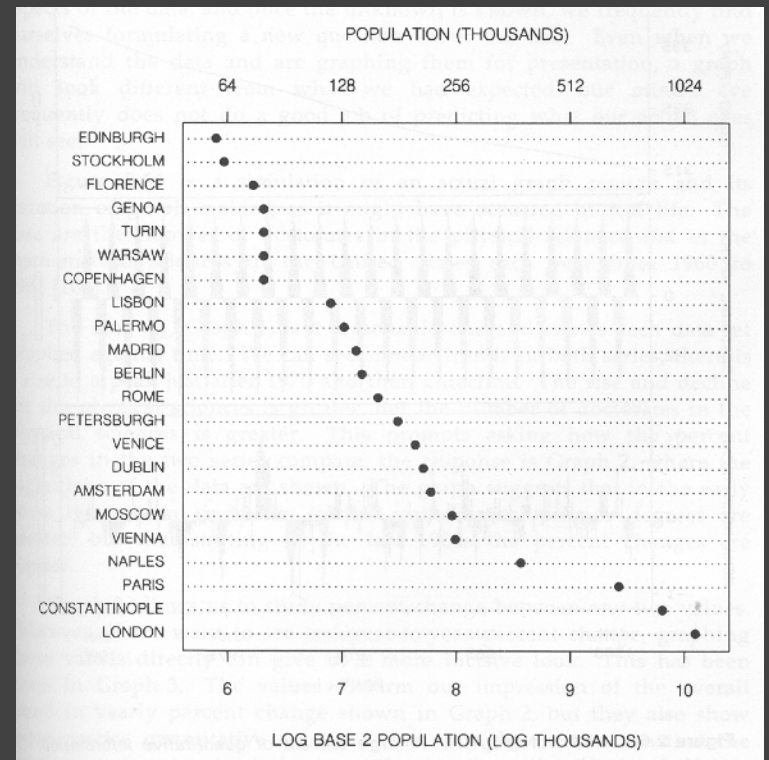
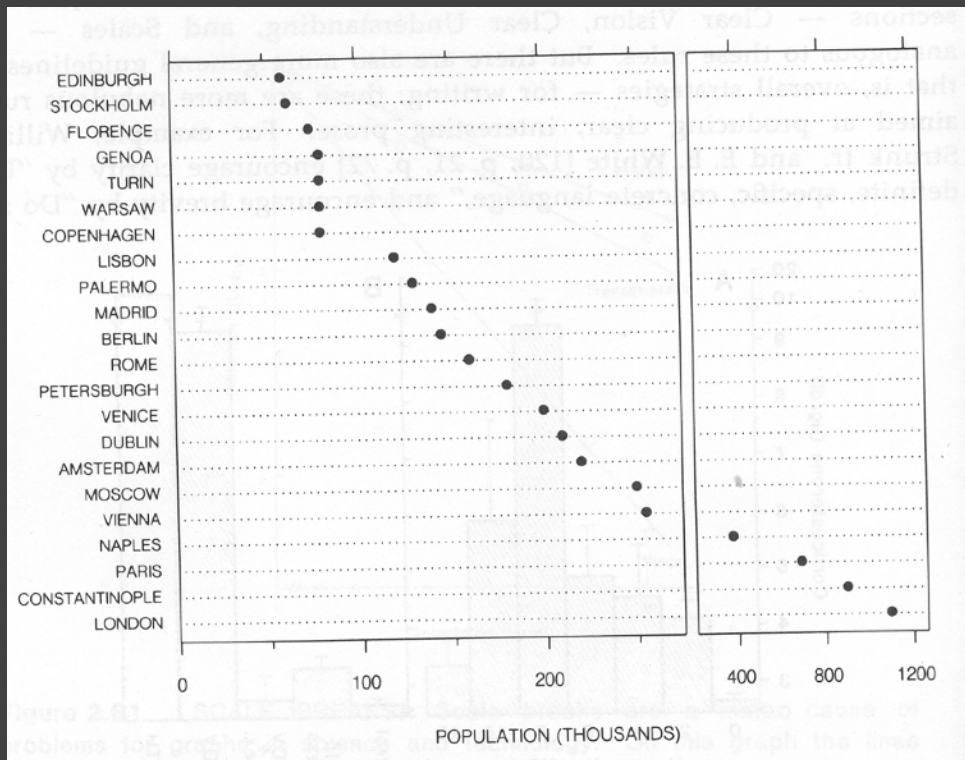


Scale Break



Log Scale

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

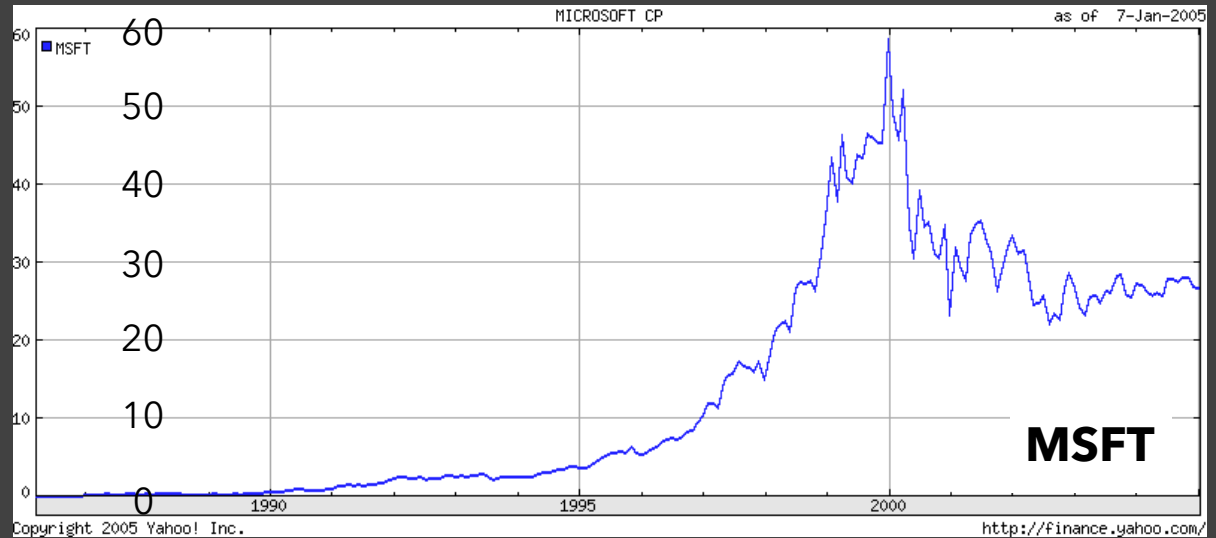
Logarithms turn *multiplication*
into *addition*.

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

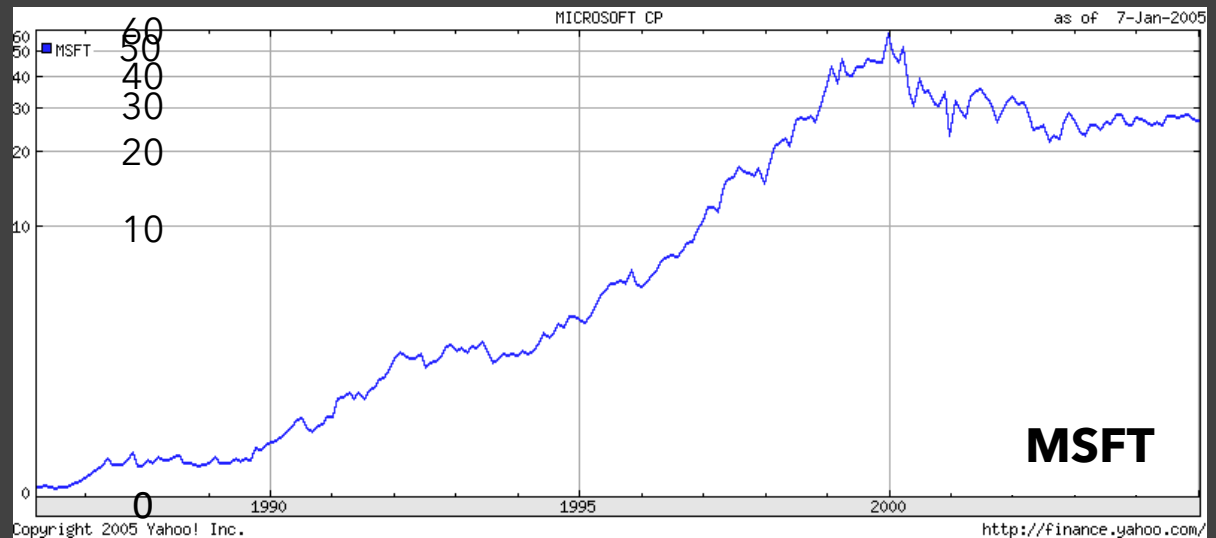
Equal steps on a log scale
correspond to equal changes to
a multiplicative scale factor.

Linear Scale vs. Log Scale

Linear Scale



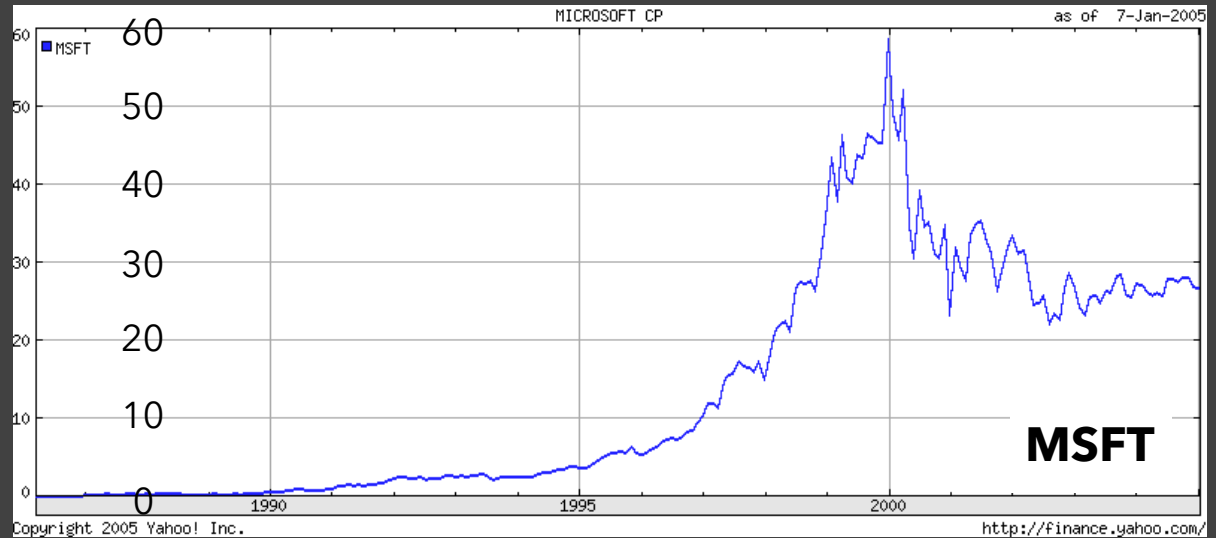
Log Scale



Linear Scale vs. Log Scale

Linear Scale

Absolute change

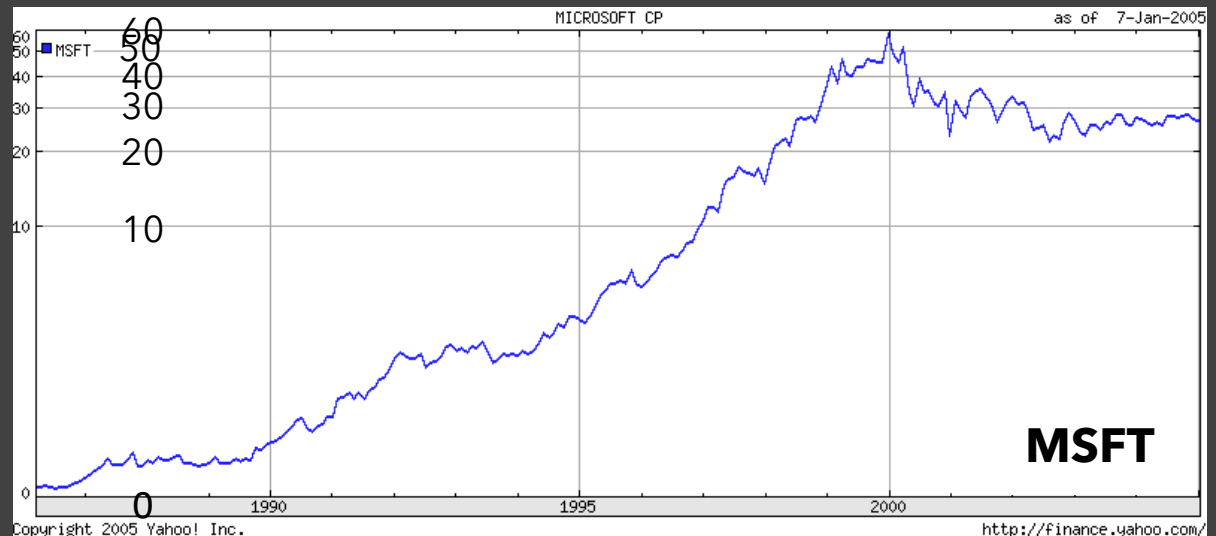


Log Scale

Small fluctuations

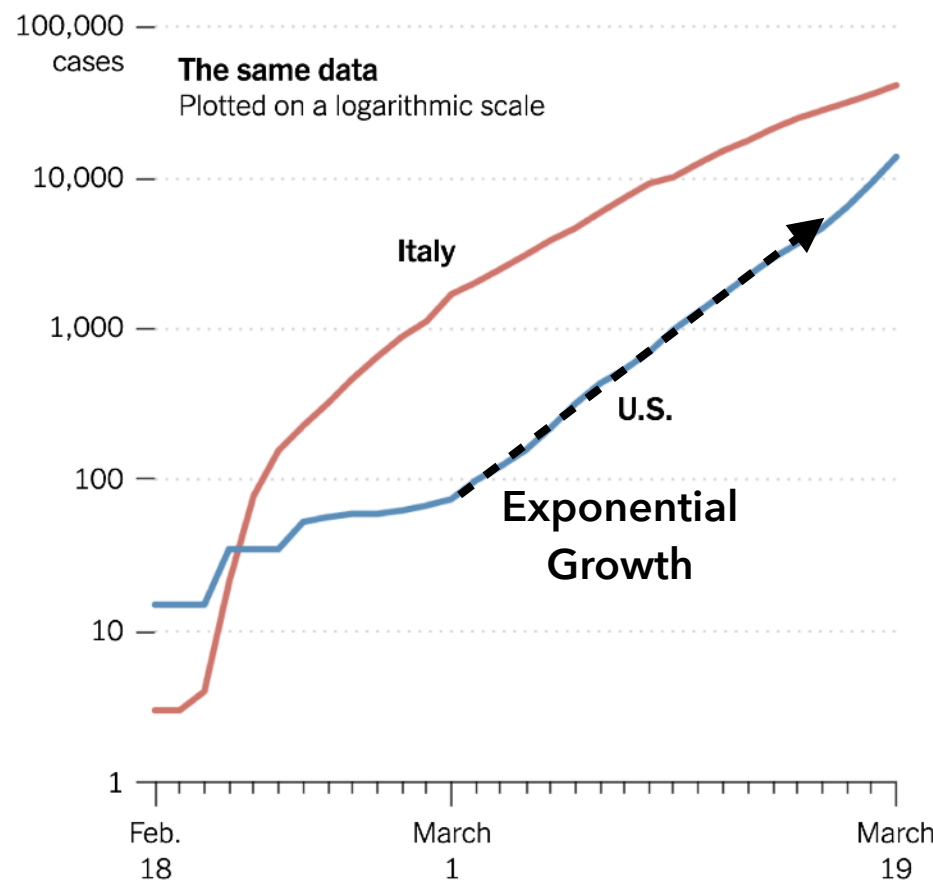
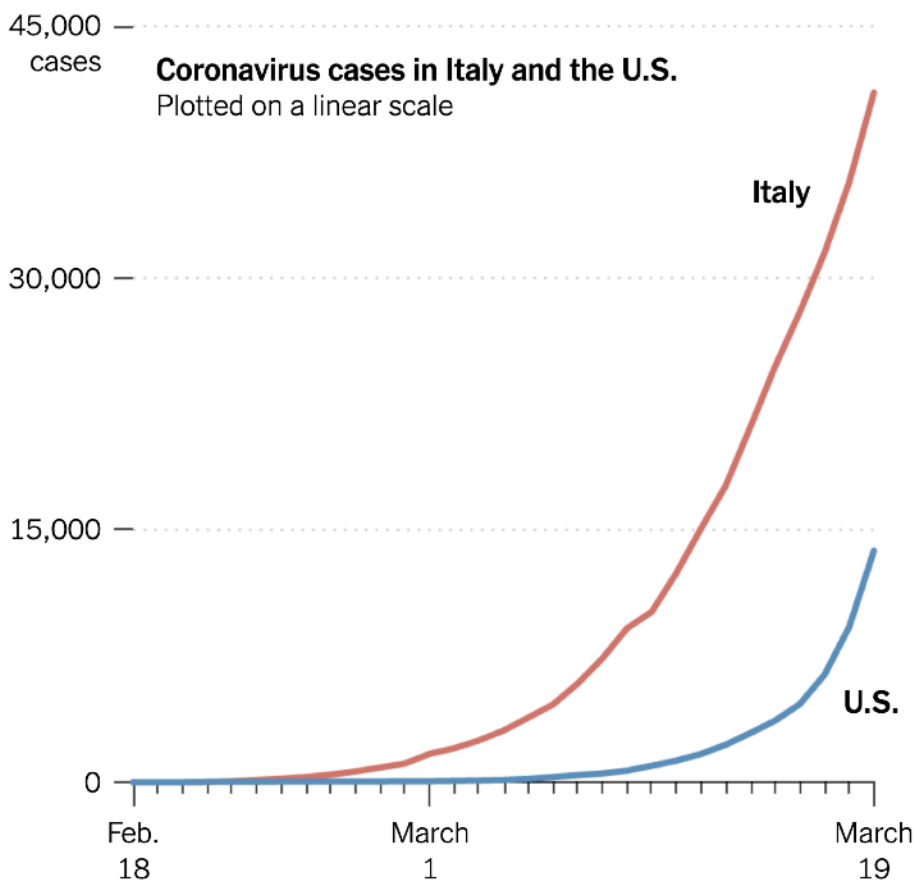
Percent change

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



Bending the Curve

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



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 \times to $+$!

Percentage change, not linear difference.

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

Constraint: **audience familiarity?**

Multidimensional Data

Visual Encoding Variables

Position (X)

Position (Y)

Area

Value

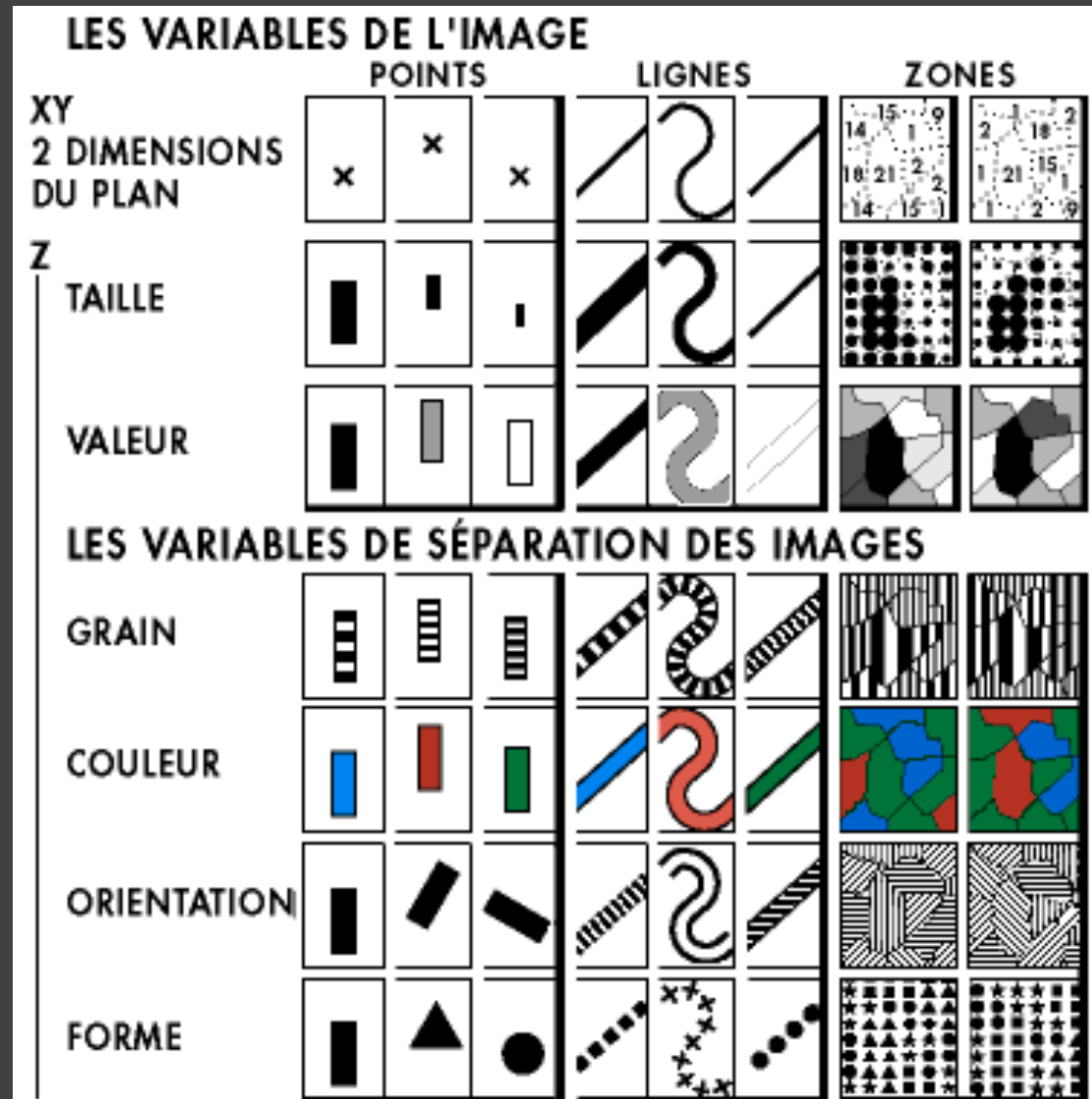
Texture

Color

Orientation

Shape

~8 dimensions?



Example: Coffee Sales

Sales figures for a fictional coffee chain

Sales	Q-Ratio
Profit	Q-Ratio
Marketing	Q-Ratio
Product Type	N {Coffee, Espresso, Herbal Tea, Tea}
Market	N {Central, East, South, West}

Filters

YEAR(Date): 2010

Marks

x+ Automatic

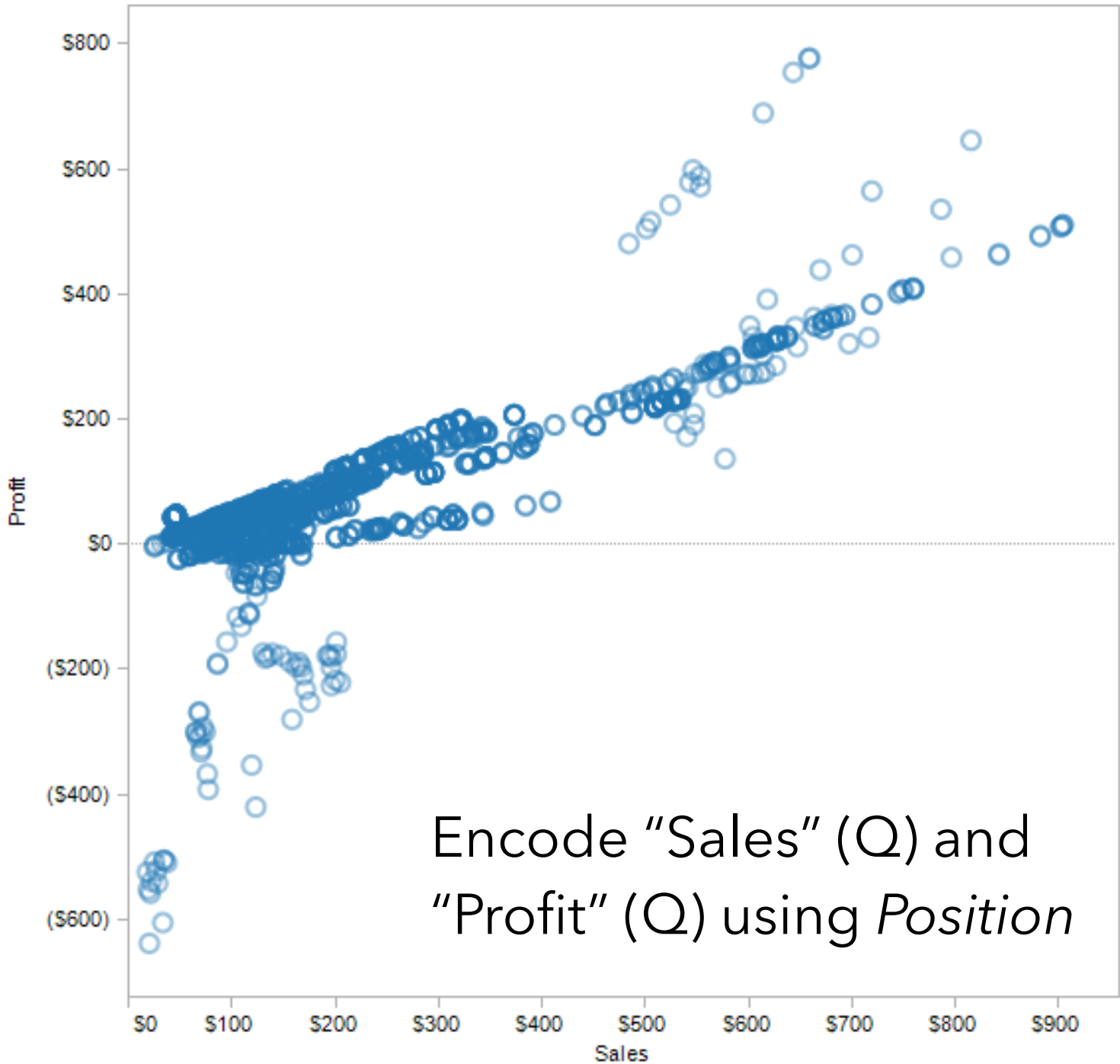
Shape Circle

Label

Color

Size

Level of Detail



Filters

YEAR(Date): 2010

Marks

x+ Automatic

Shape

Label

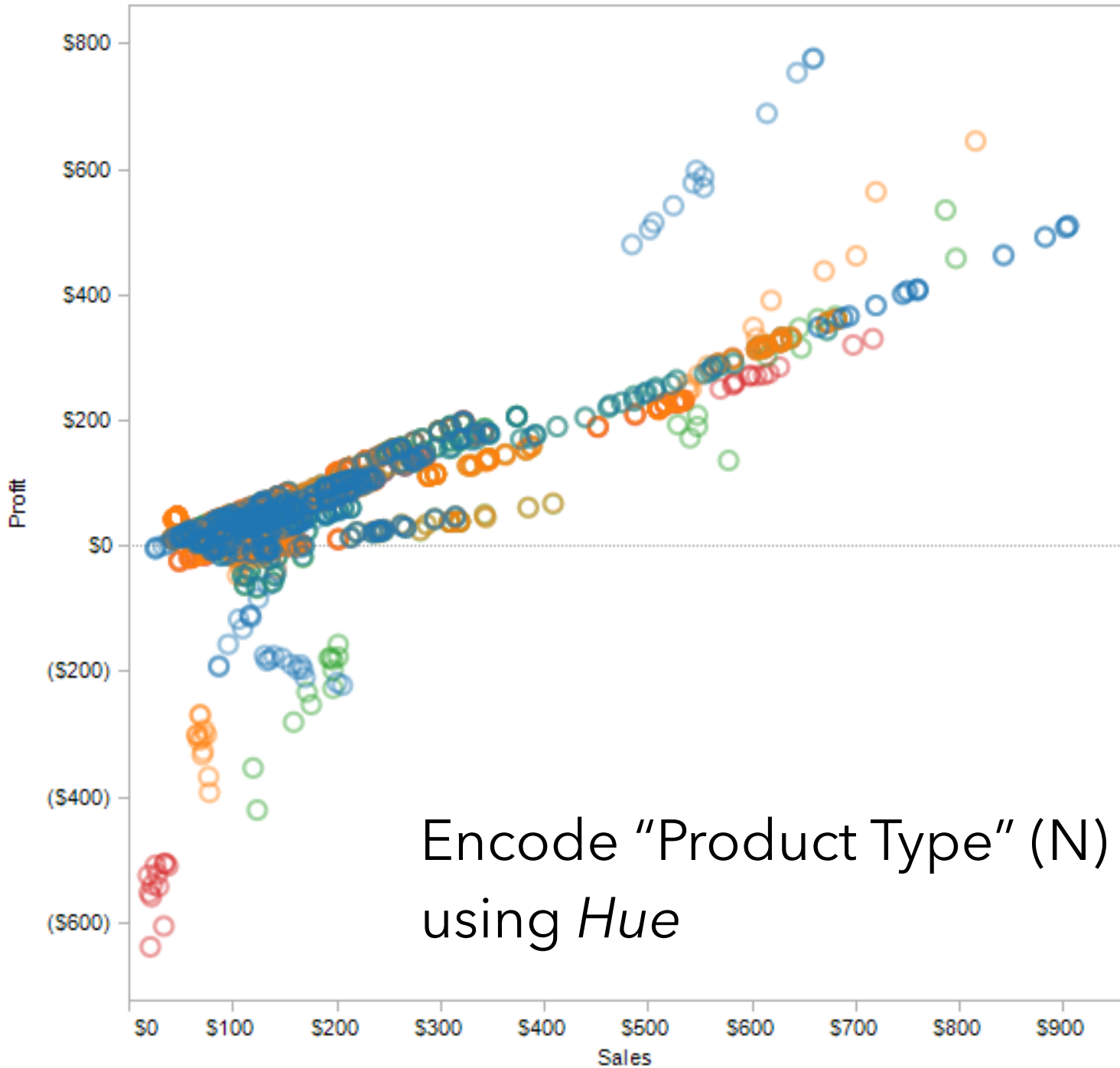
Color

Size

Level of Detail

Product Type

- Coffee
- Espresso
- Herbal Tea
- Tea



Filters

YEAR(Date): 2010

Marks

Automatic

Shape Market

Label Market

Color Product Type

Size

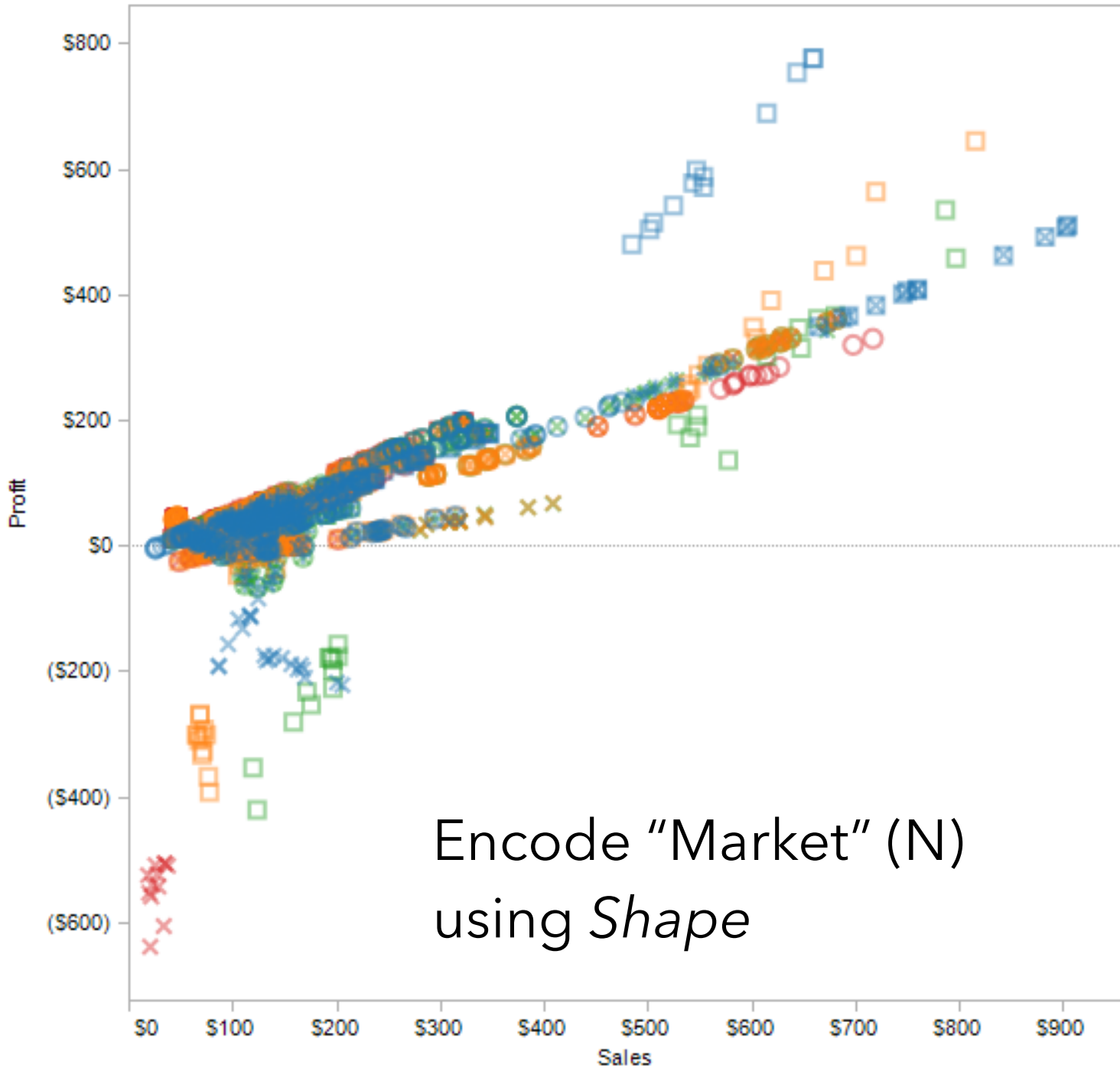
Level of Detail

Product Type

- Coffee
- Espresso
- Herbal Tea
- Tea

Market

- Central
- East
- South
- West



Filters

YEAR(Date): 2010

Marks

Automatic

Shape Market

Label

Color Product Type

Size Marketing

Marketing

Level of Detail

Product Type

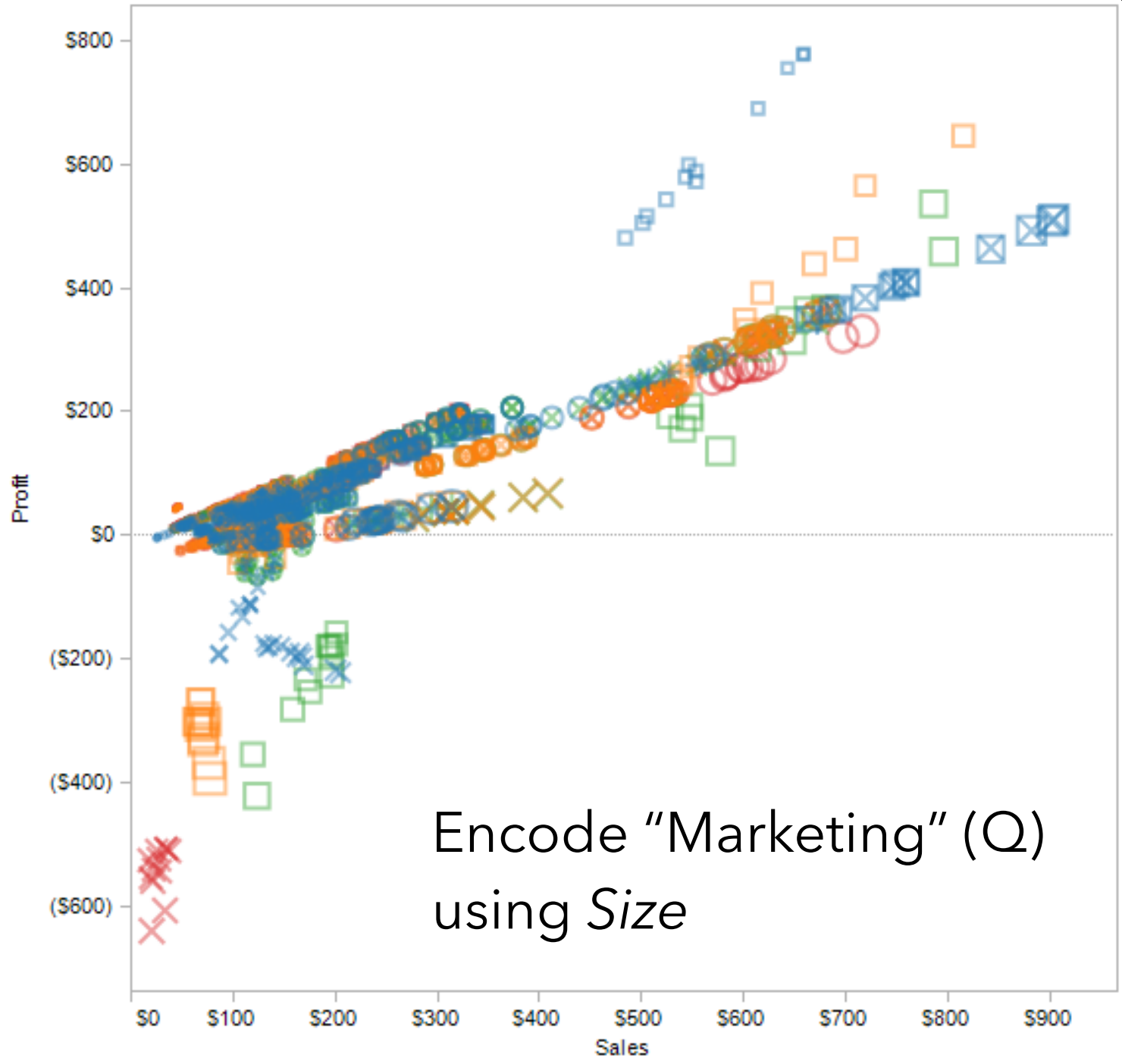
- Coffee
- Espresso
- Herbal Tea

Market

- Central
- East
- South

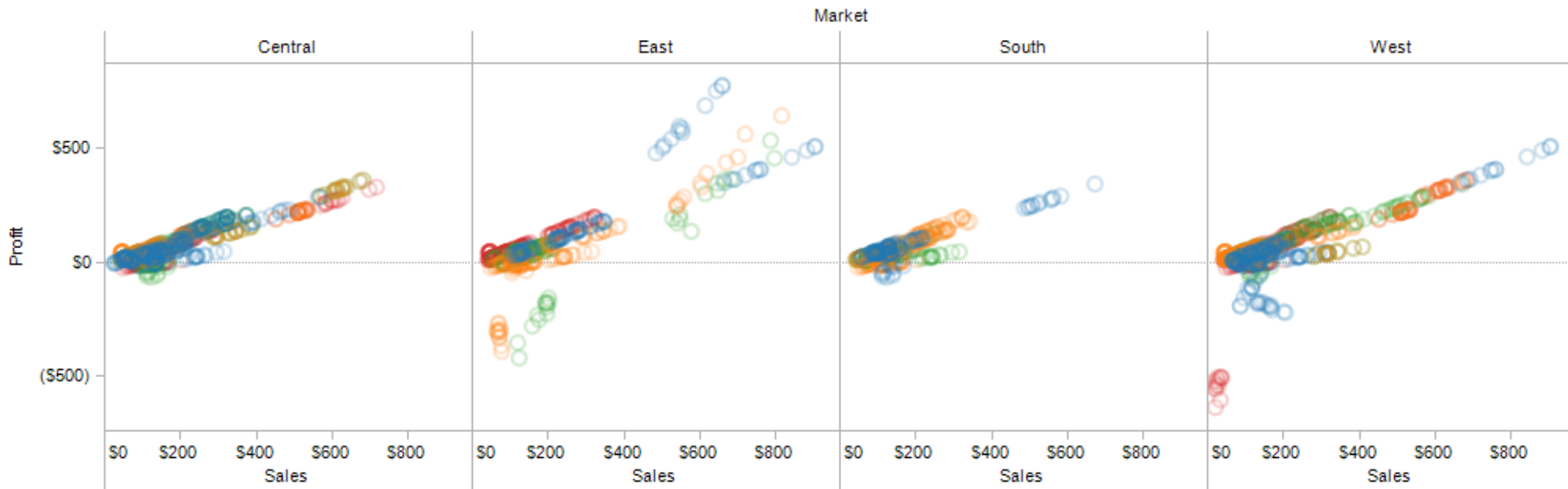
Marketing

- \$0
- \$50
- \$100



Encode "Marketing" (Q) using *Size*

Trellis Plots



A *trellis plot* subdivides space to enable comparison across multiple plots.

Typically nominal or ordinal variables are used as dimensions for subdivision.

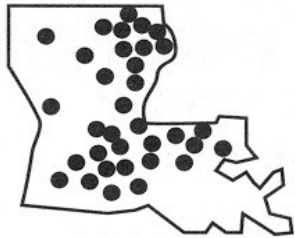
Small Multiples



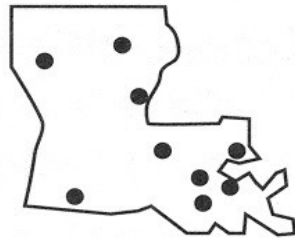
[MacEachren '95, Figure 2.11, p. 38]

Small Multiples

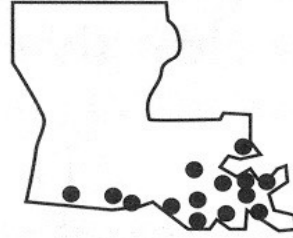
alfisol



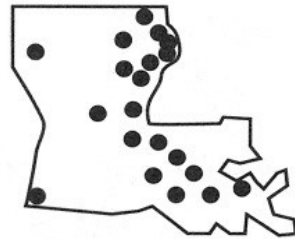
entisol



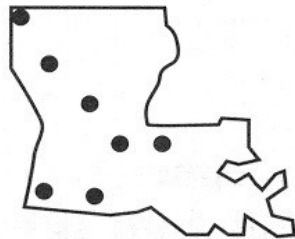
histosol



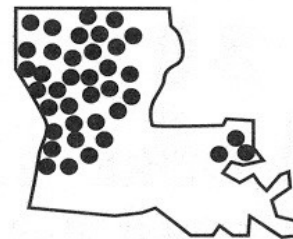
inceptisol



mollisol

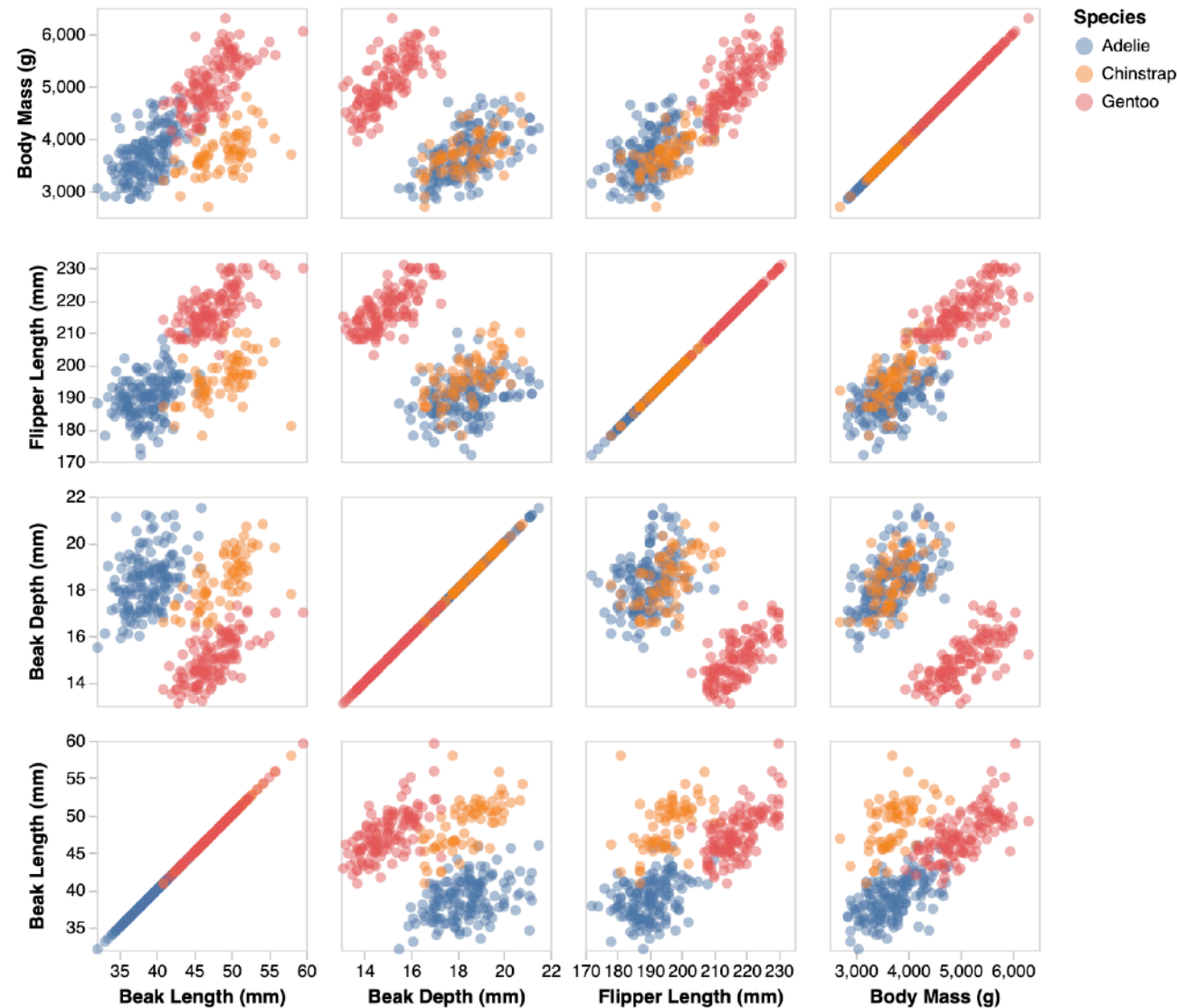


ultisol



[MacEachren '95, Figure 2.11, p. 38]

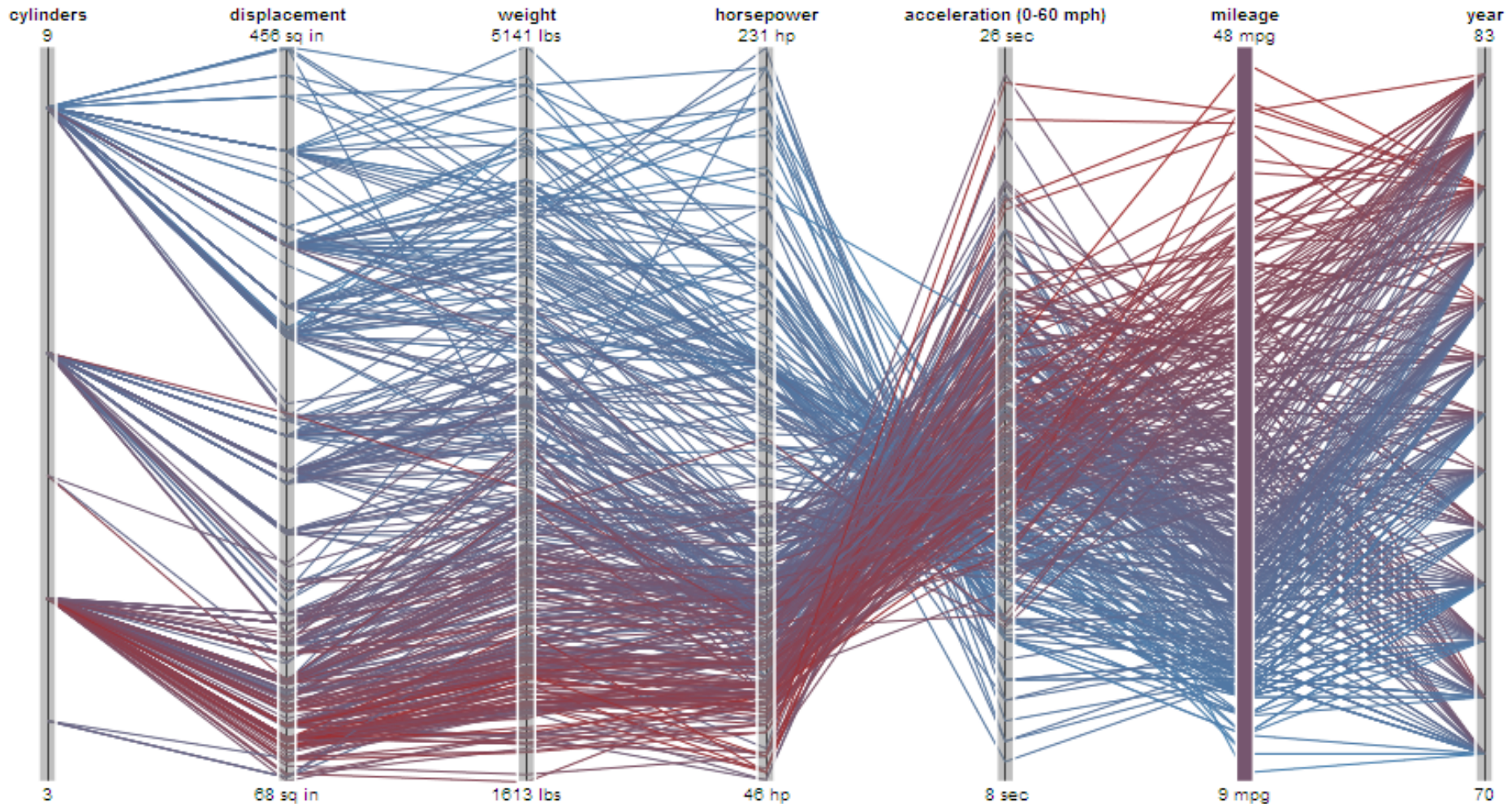
Scatterplot Matrix (SPLOM)



Scatter plots for pairwise comparison of each data dimension.

Parallel Coordinates

Parallel Coordinates [Inselberg]



Parallel Coordinates [Inselberg]

Visualize up to ~two dozen dimensions at once

1. Draw parallel axes for each variable
2. For each tuple, connect points on each axis

Between adjacent axes: line crossings imply neg. correlation, shared slopes imply pos. correlation.

Full plot can be cluttered. **Interactive selection** can be used to assess multivariate relationships.

Highly sensitive to axis **scale** and **ordering**.

Expertise required to use effectively!

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!