

# (In)Effective Visual Encoding

---

**DSC 106: Data Visualization**

Sam Lau  
UC San Diego

Join at  
[slido.com](https://slido.com)  
**#3892 640**



# Announcements

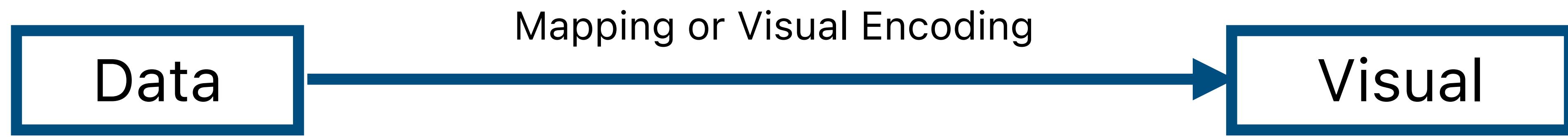
Lab 2 out, due this Friday, 1/19.

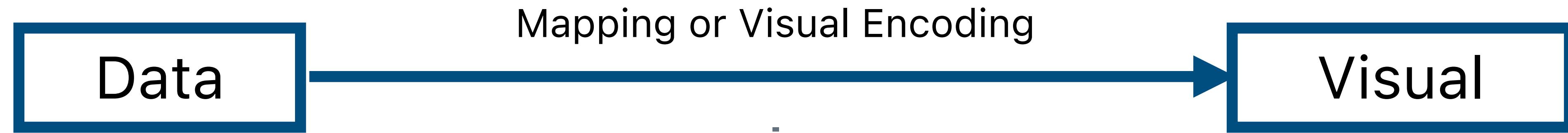
Project 1 also due this Friday, 1/19.

No lab checkoffs during Sam's OH, on Thurs now instead of Fri

## FAQs:

1. How does project grading work? You get 9/10 points if you follow all the project requirements. Can get more if your project goes above and beyond requirements (see project page for more details).
2. OH now have signup forms to distribute checkoffs, see Ed for more details.





## Expressiveness

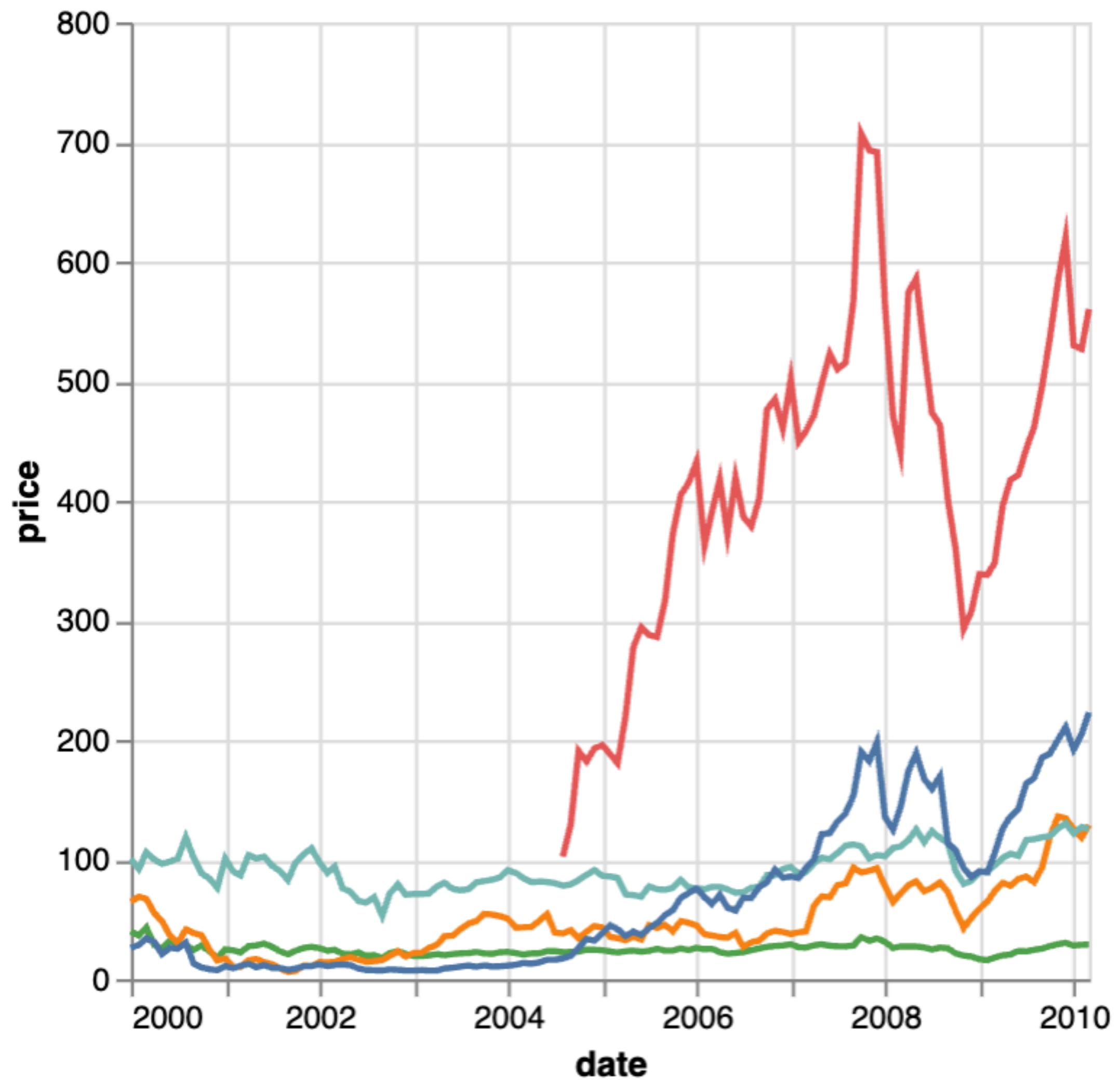
A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express *all the facts in the set of data, and only the facts in the data.*

## Effectiveness

A visualization is more *effective* than another if the information it conveys *is more readily perceived* than the information in the other visualization



# Example from Lab 1



**symbol**

- AAPL
- AMZN
- GOOG
- IBM
- MSFT

Mark: line  
X-axis: date (Q-interval)  
Y-axis: price (Q-ratio)  
Color: symbol (N)

# *Driving Shifts Into Reverse*

**E**CONOMISTS have long studied the relationship between driving habits and gasoline prices. Low gas prices can bring periods of profligate driving and a quick jump in the number of vehicles

Until recent

more each year with a few brief Americans of 4,000 miles a year later, that figure

But the last few years have seen some big changes. The recession meant that few people could afford to buy a house, and a sharp

THE BOSTONIAN

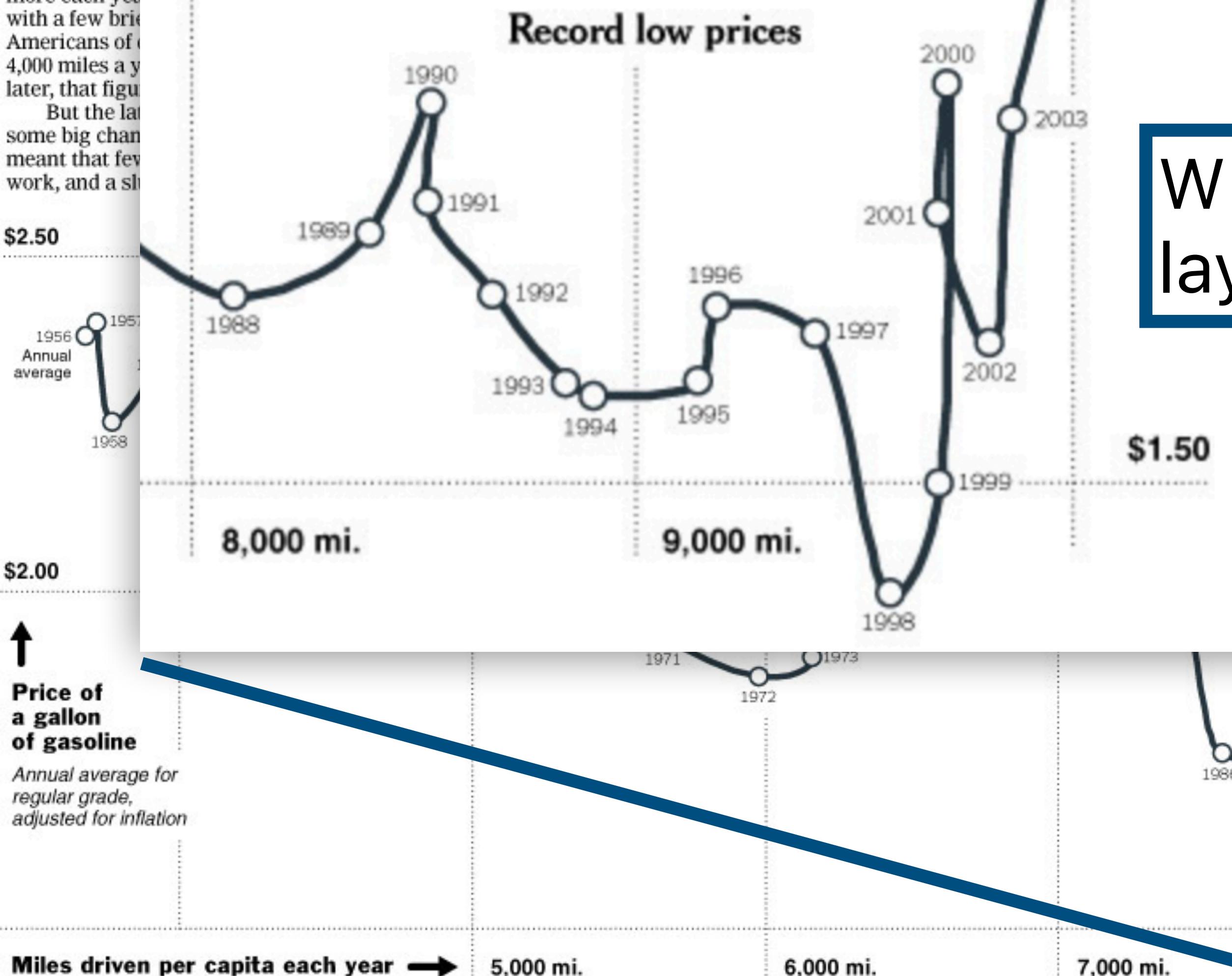
10.1007/s00332-010-9000-0

meant that less freight needed to be moved around the country. As gas prices soared in 2005, the number of miles driven — including commercial and personal —

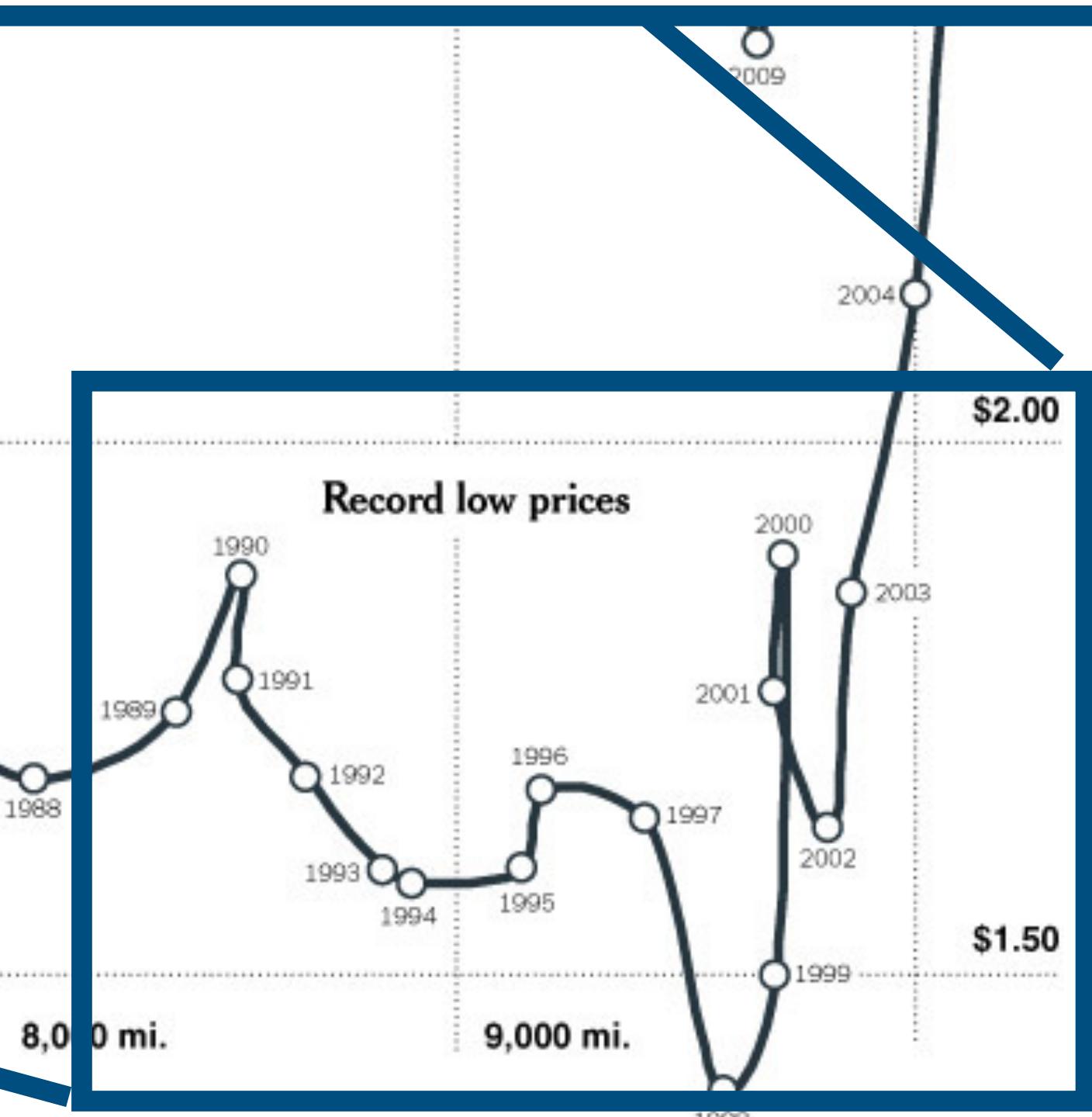
## Energy crisis

### The swing backward

The average number of miles that Americans drive annually begins to fall, so the chart appears to turn around.



What are the marks, encoding, layers for this plot?



# Driving Shifts Into Reverse

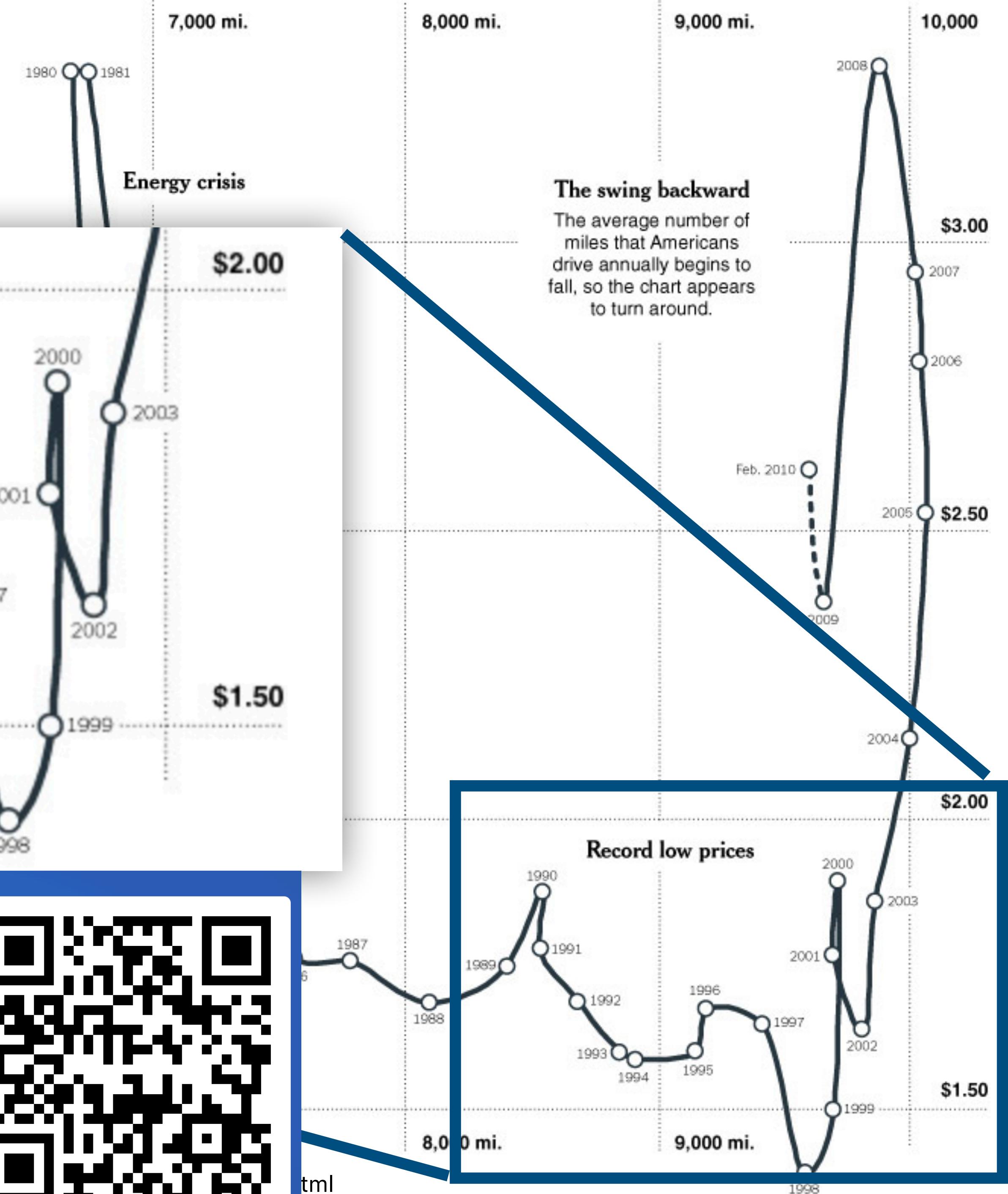
**E**CONOMISTS have long studied the relationship between driving habits and gasoline prices. Low gas prices can bring periods of profligate driving and a quick jump in the number of miles driven.

Until recently,

more each year with a few brief Americans of 4,000 miles a year later, that figure

But the last some big changes meant that few work, and a sh

meant that less freight needed to be moved around the country. As gas prices soared in 2005, the number of miles driven — including commercial and personal —



Price of a gallon of gasoline

Annual average for regular grade, adjusted for inflation

Miles driven per year

Join at

slido.com

#3892 640

<https://ar>



html

# A Design Space of Visual Encodings

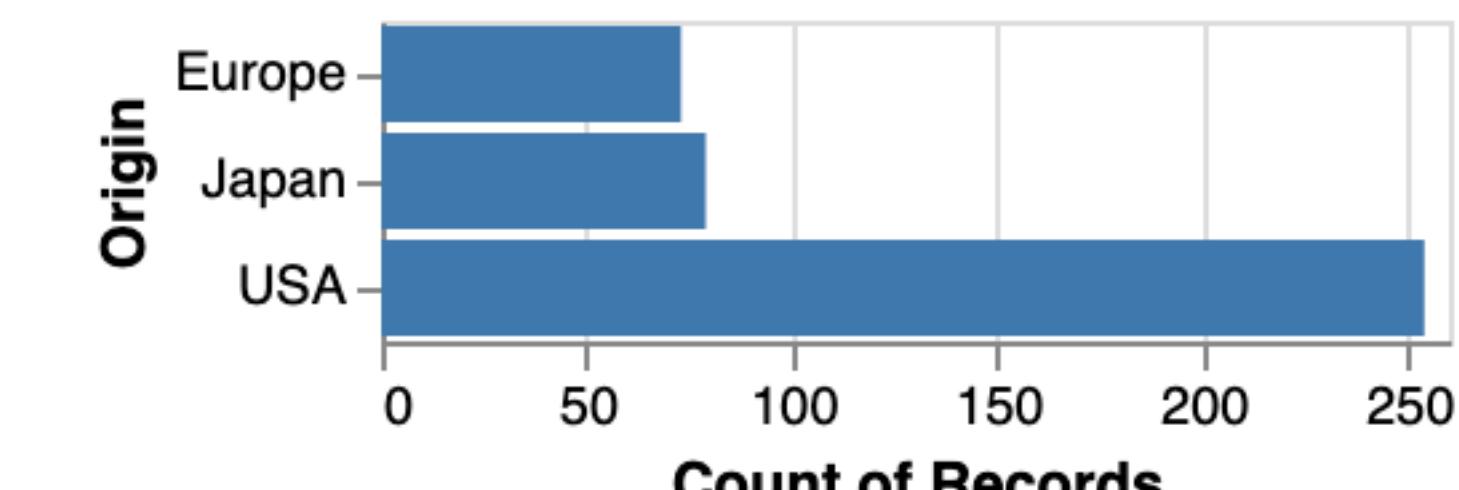
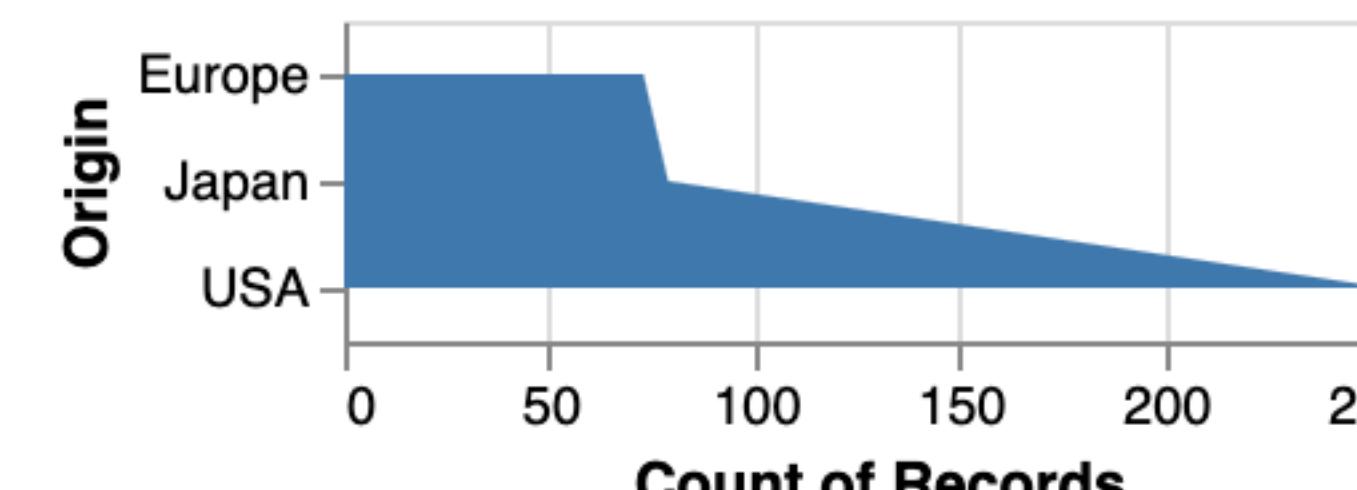
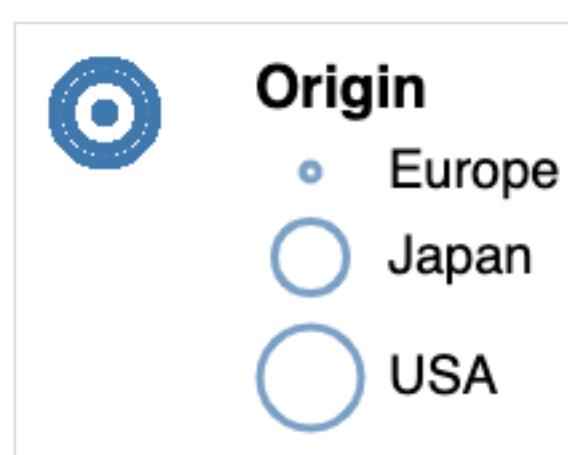
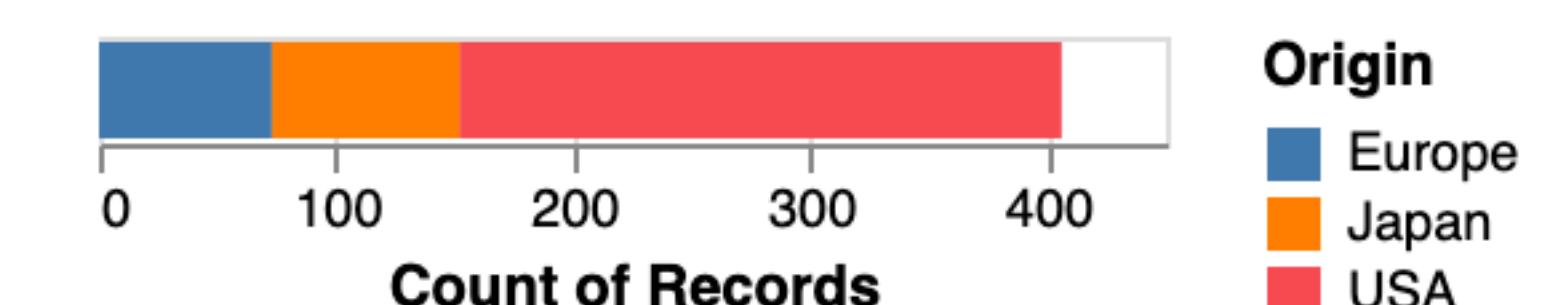
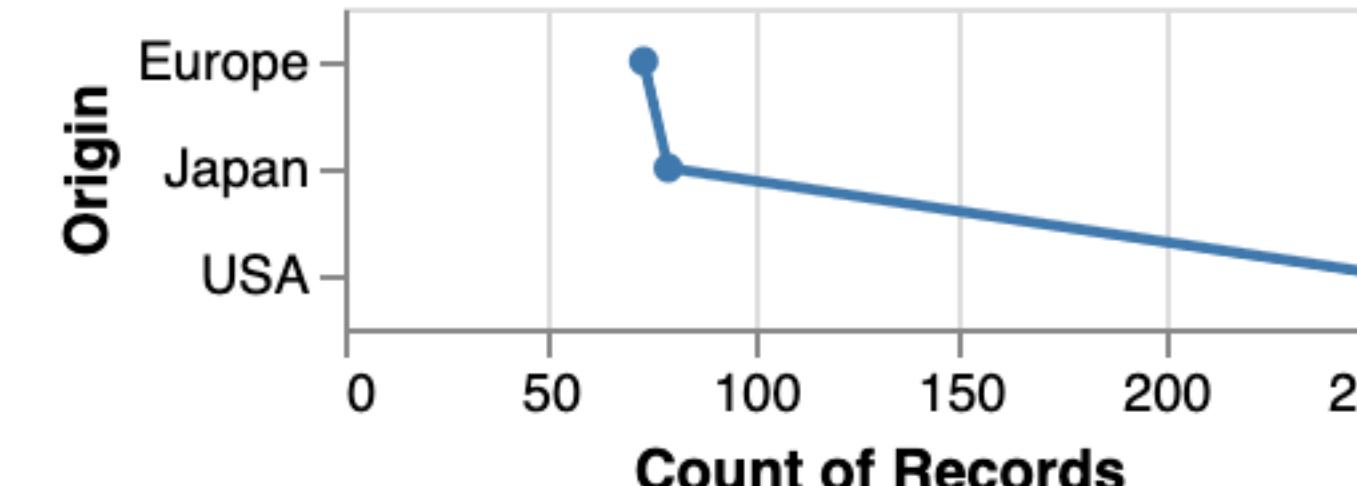
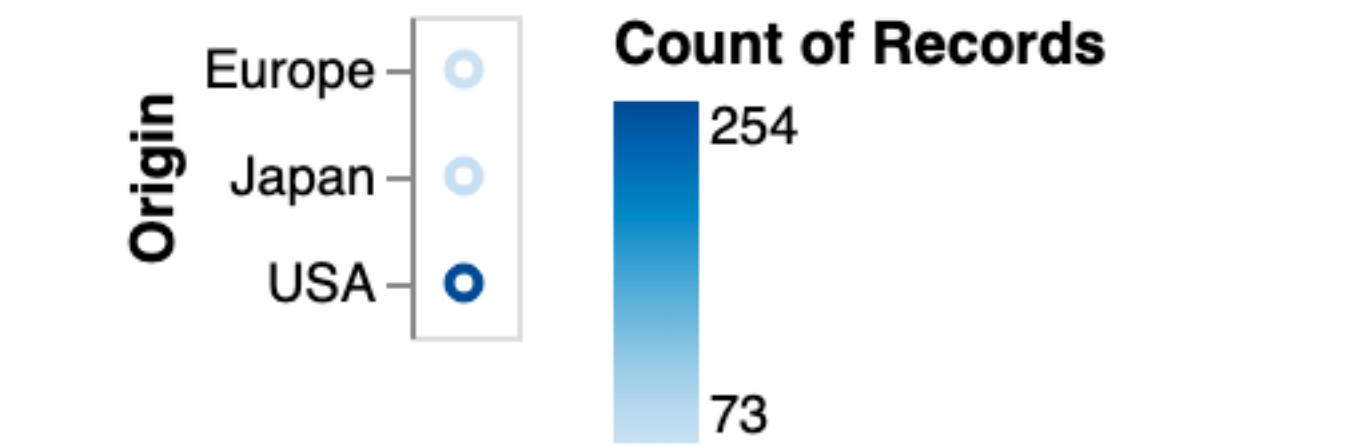
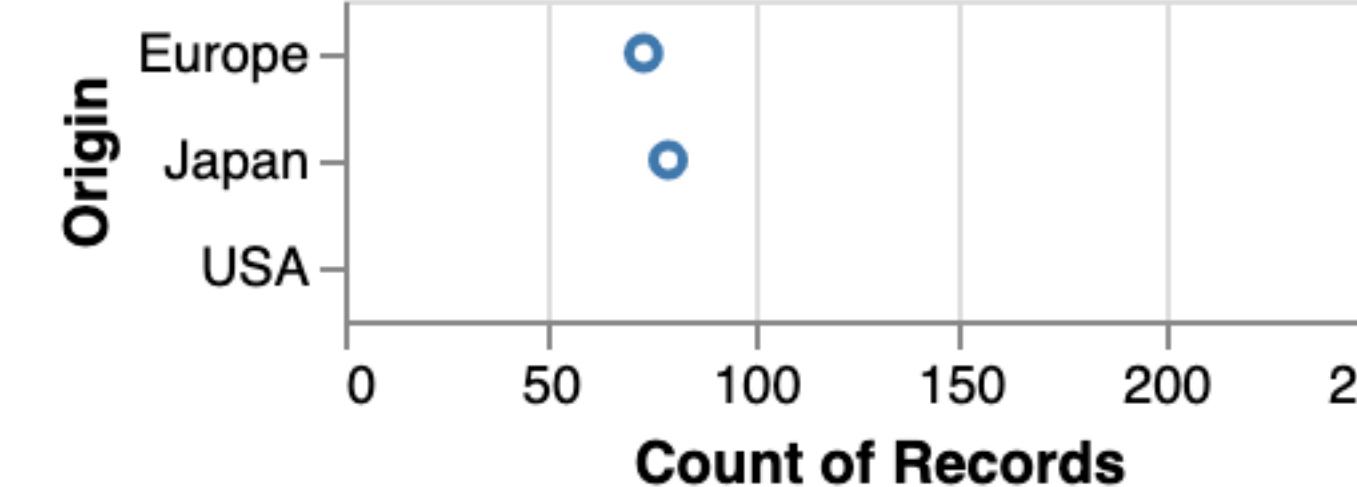
# Visual Encoding = Combinatorial Design Space

1D nominal data (N, O)

raw



aggregate (count)

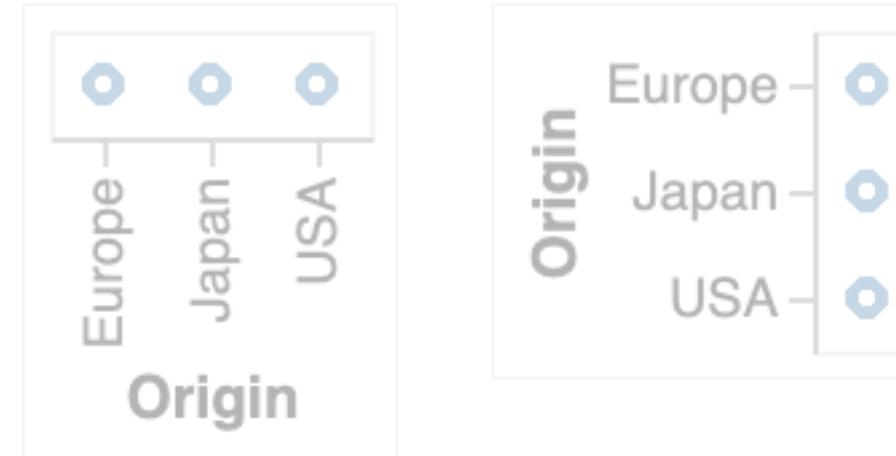


# Visual Encoding = Combinatorial Design Space

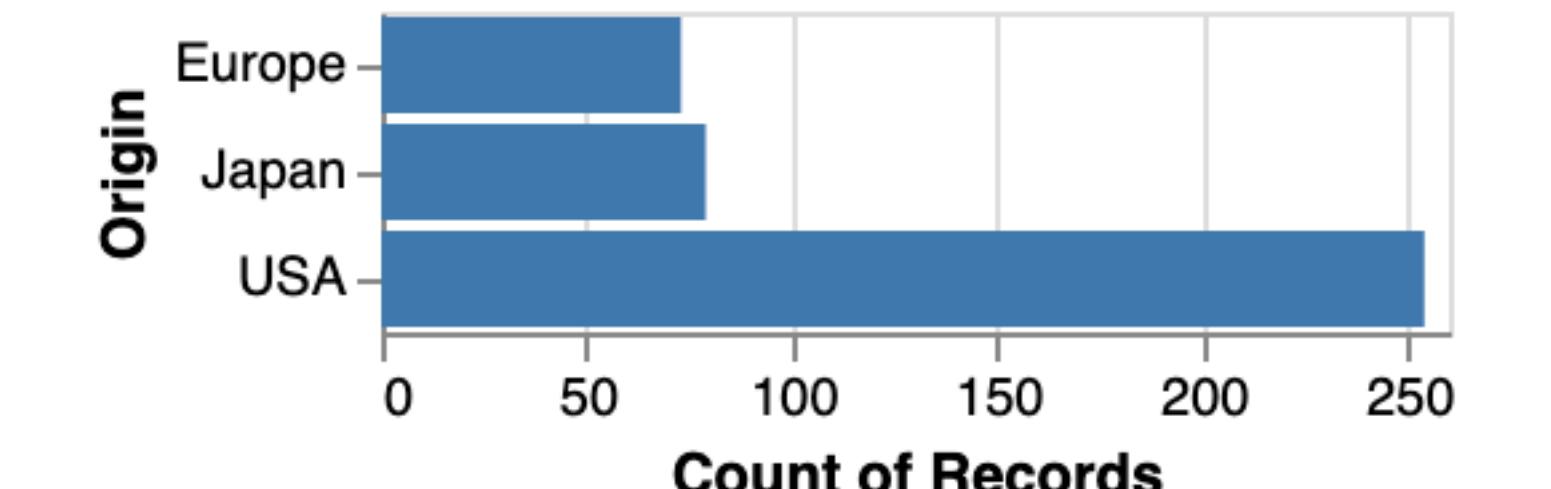
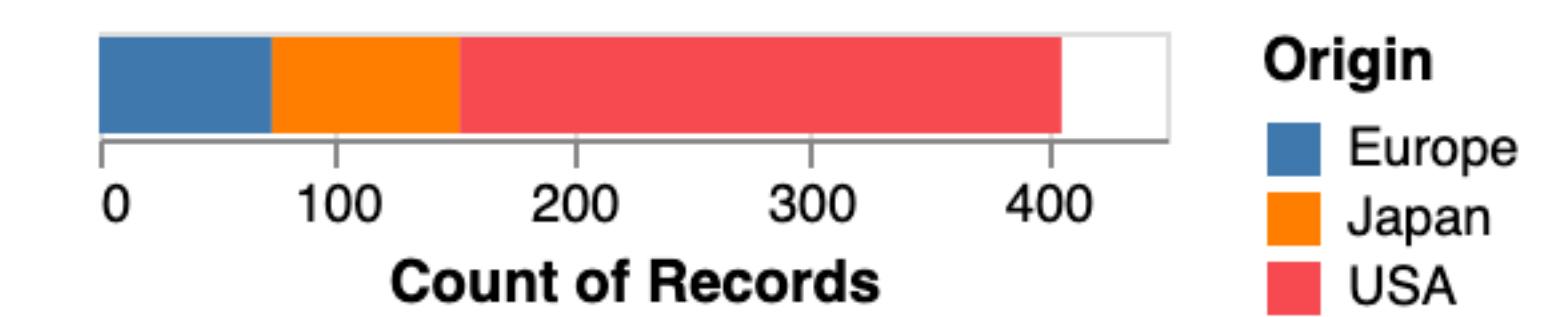
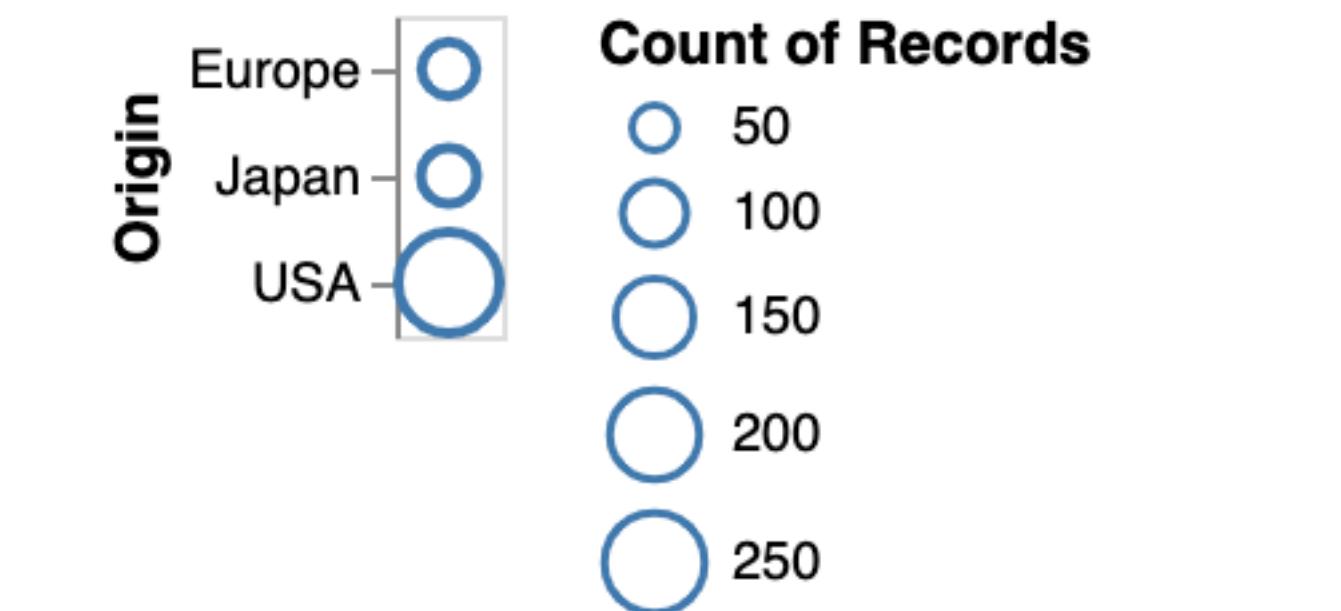
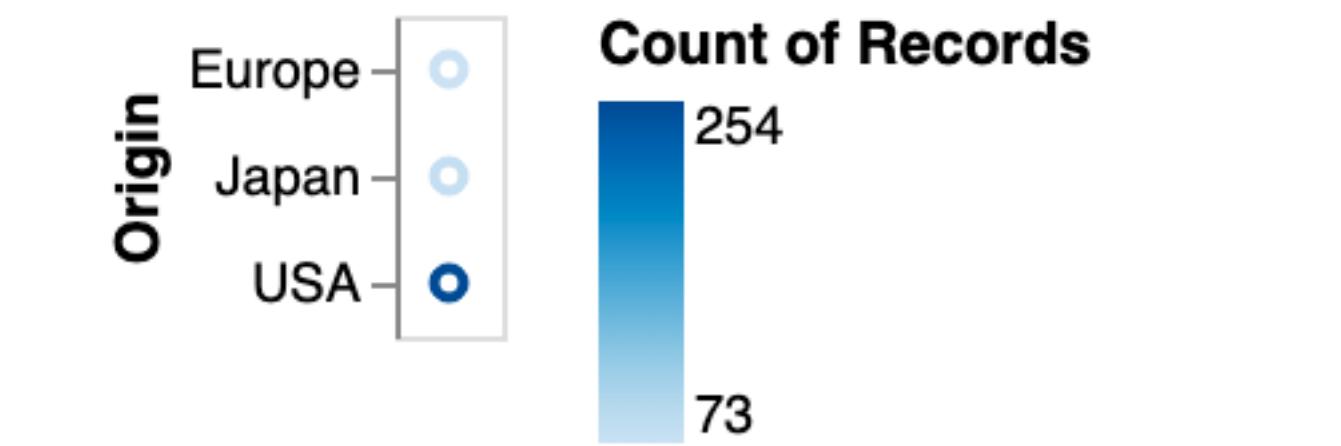
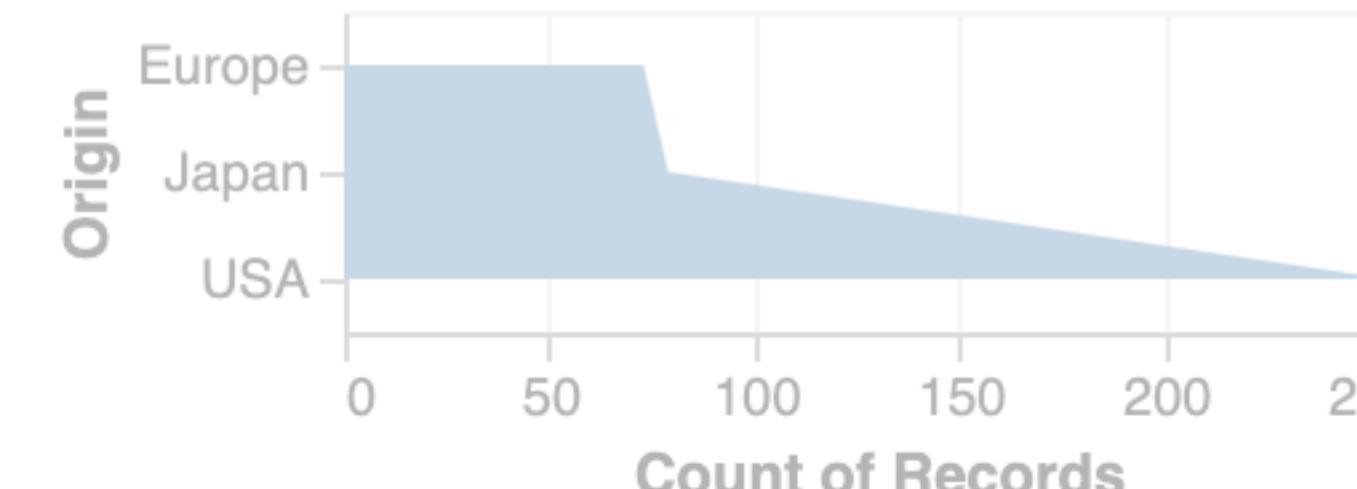
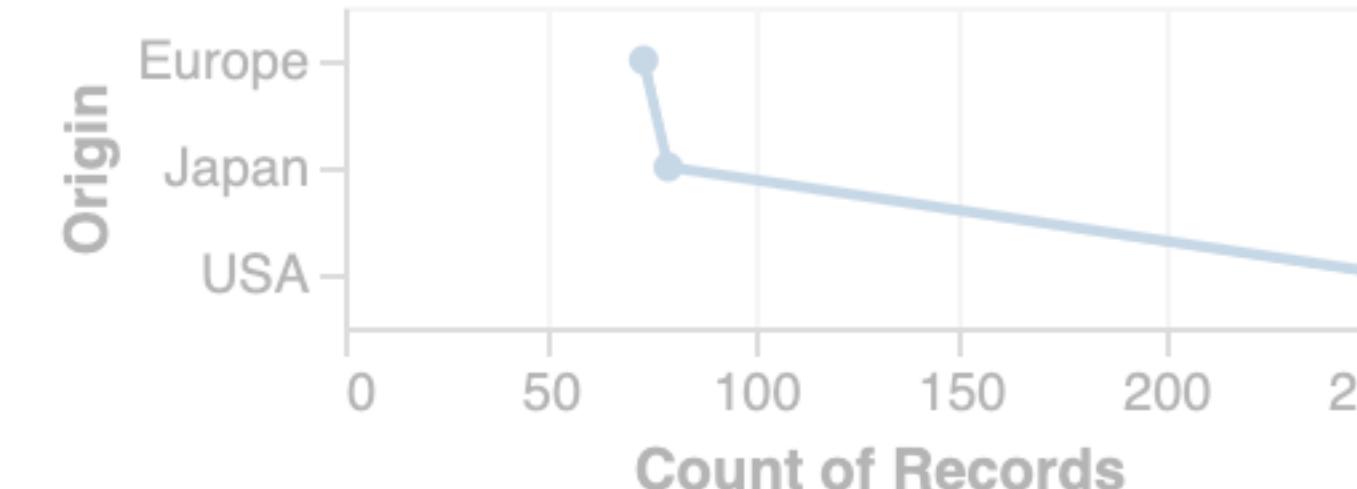
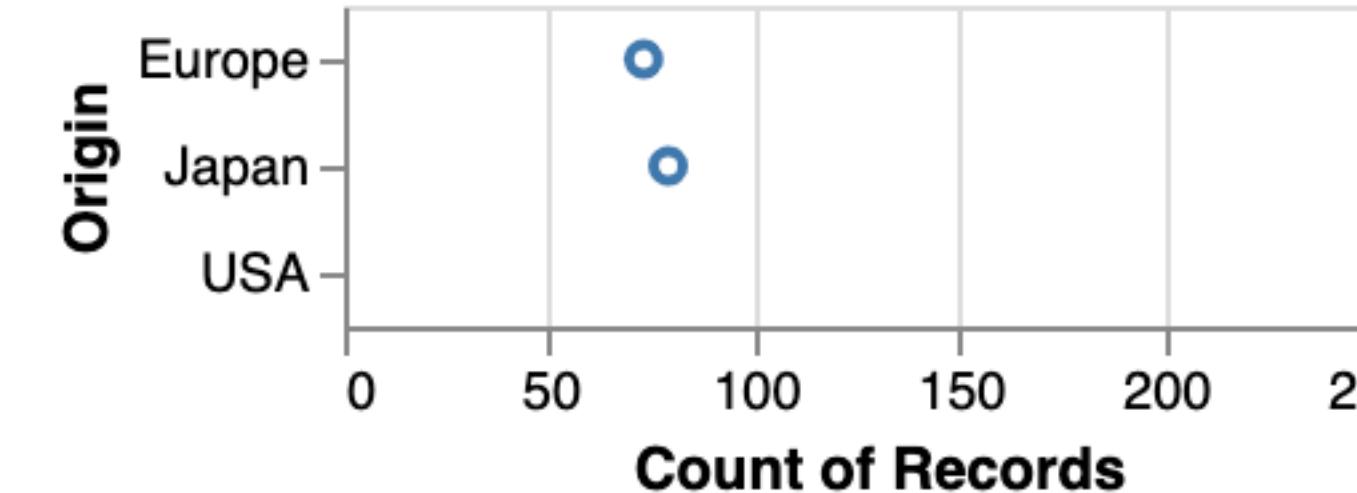
1D nominal data (N, O)

Expressive?

raw



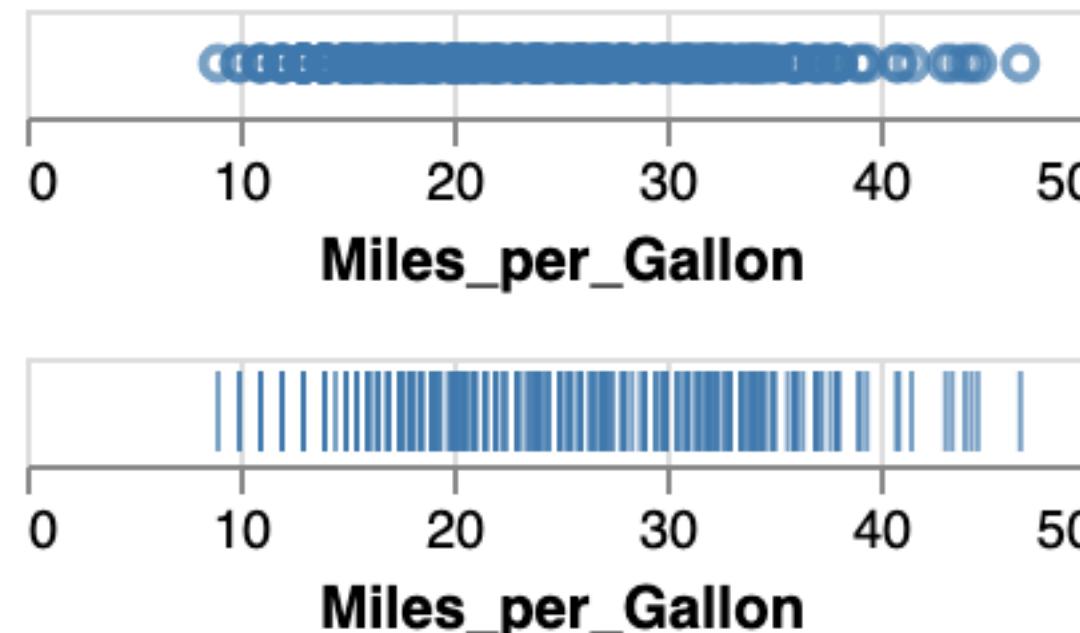
aggregate (count)



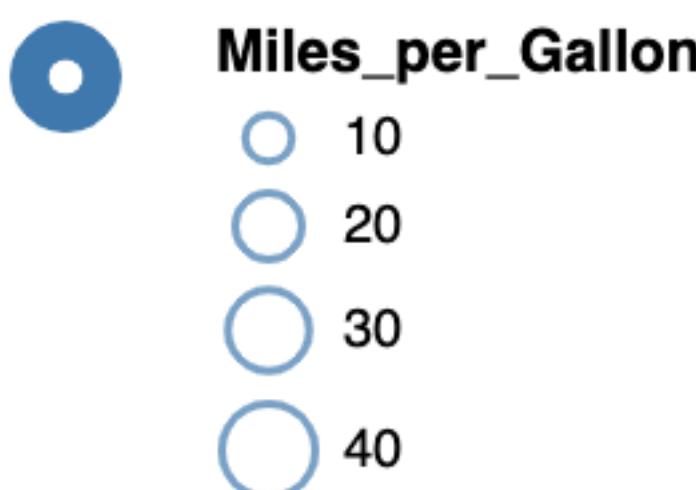
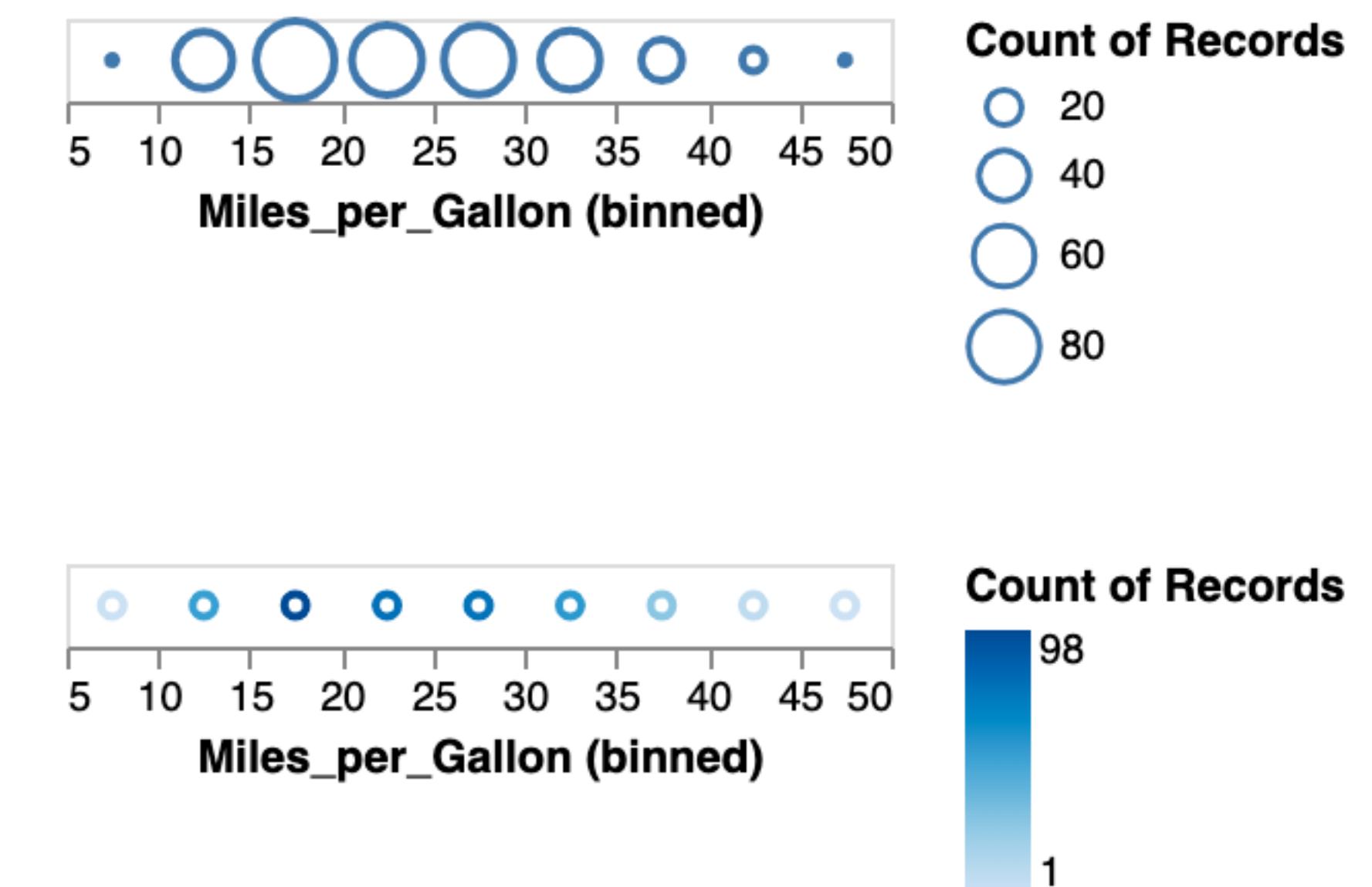
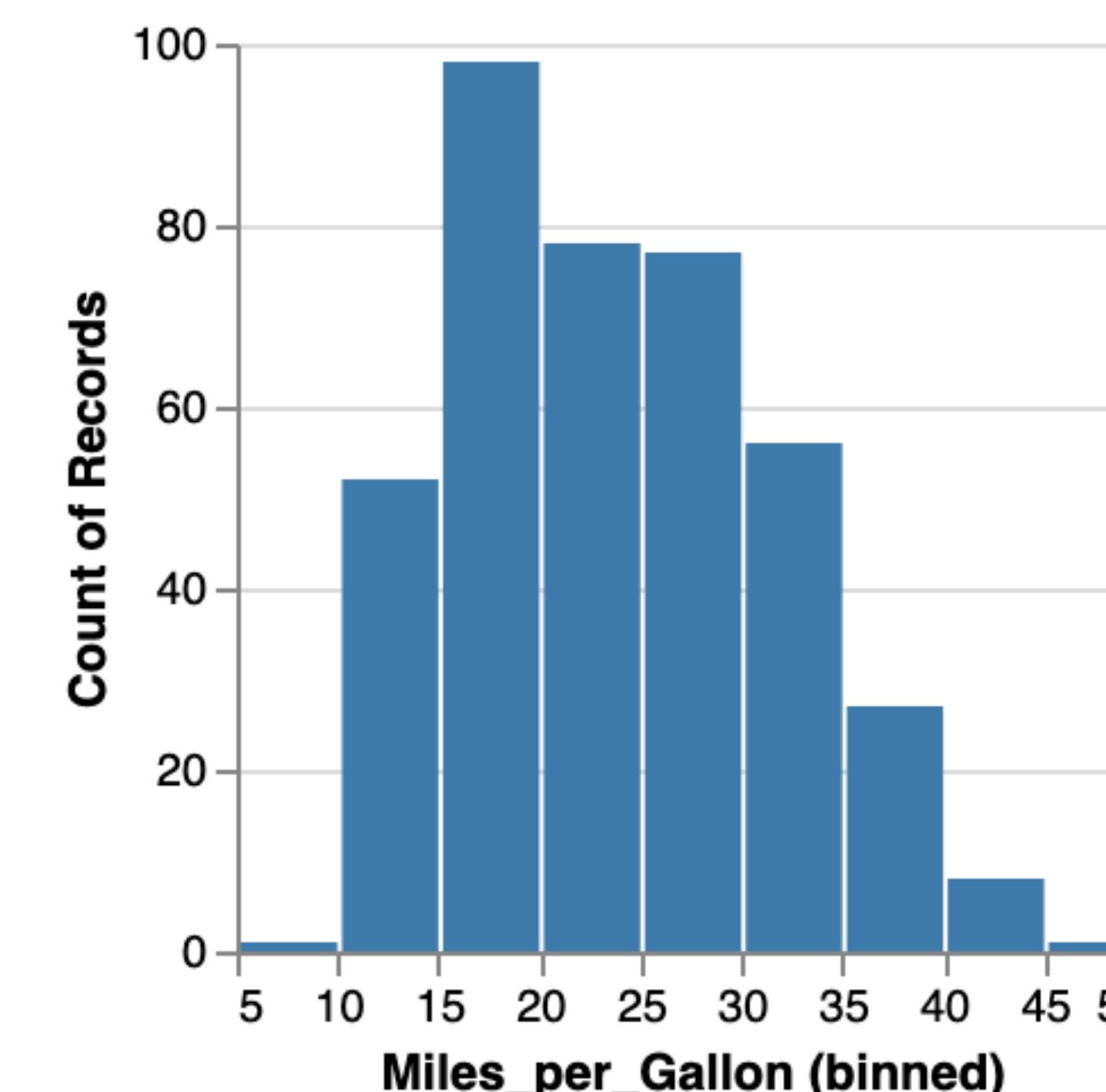
# Visual Encoding = Combinatorial Design Space

1D quantitative data (Q)

raw



aggregate (count)

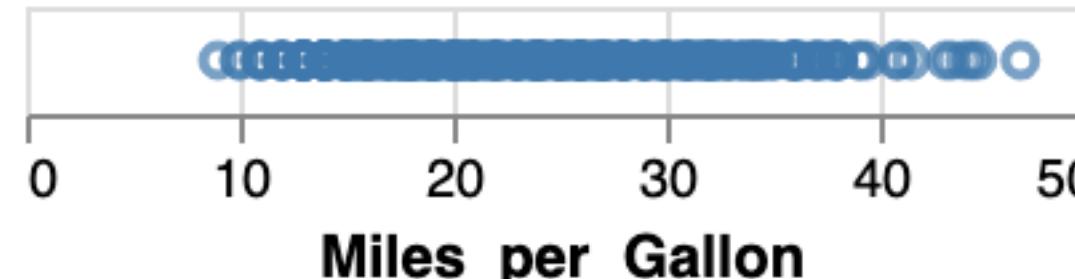


# Visual Encoding = Combinatorial Design Space

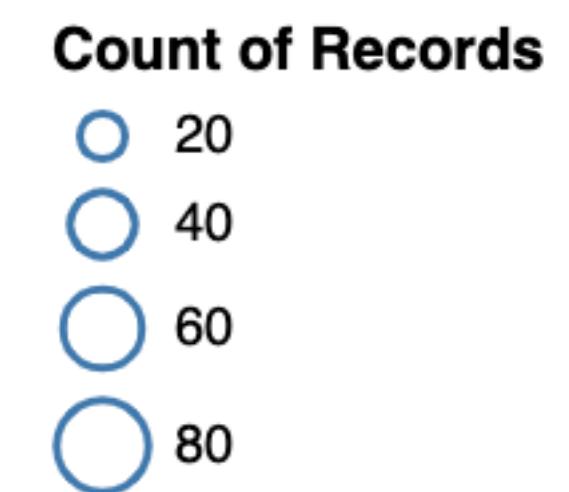
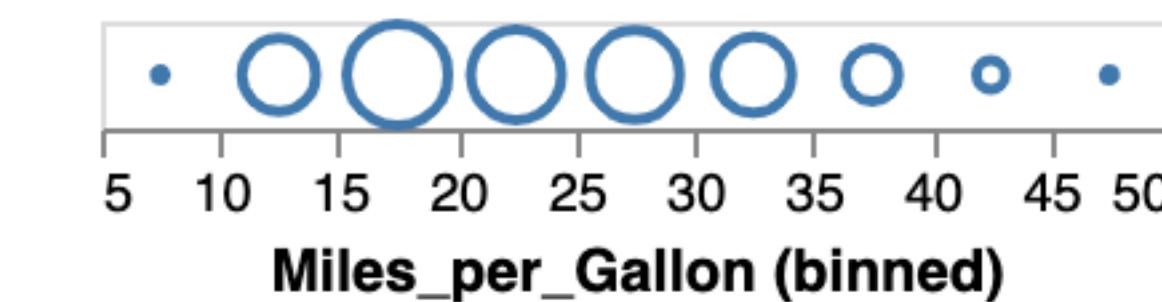
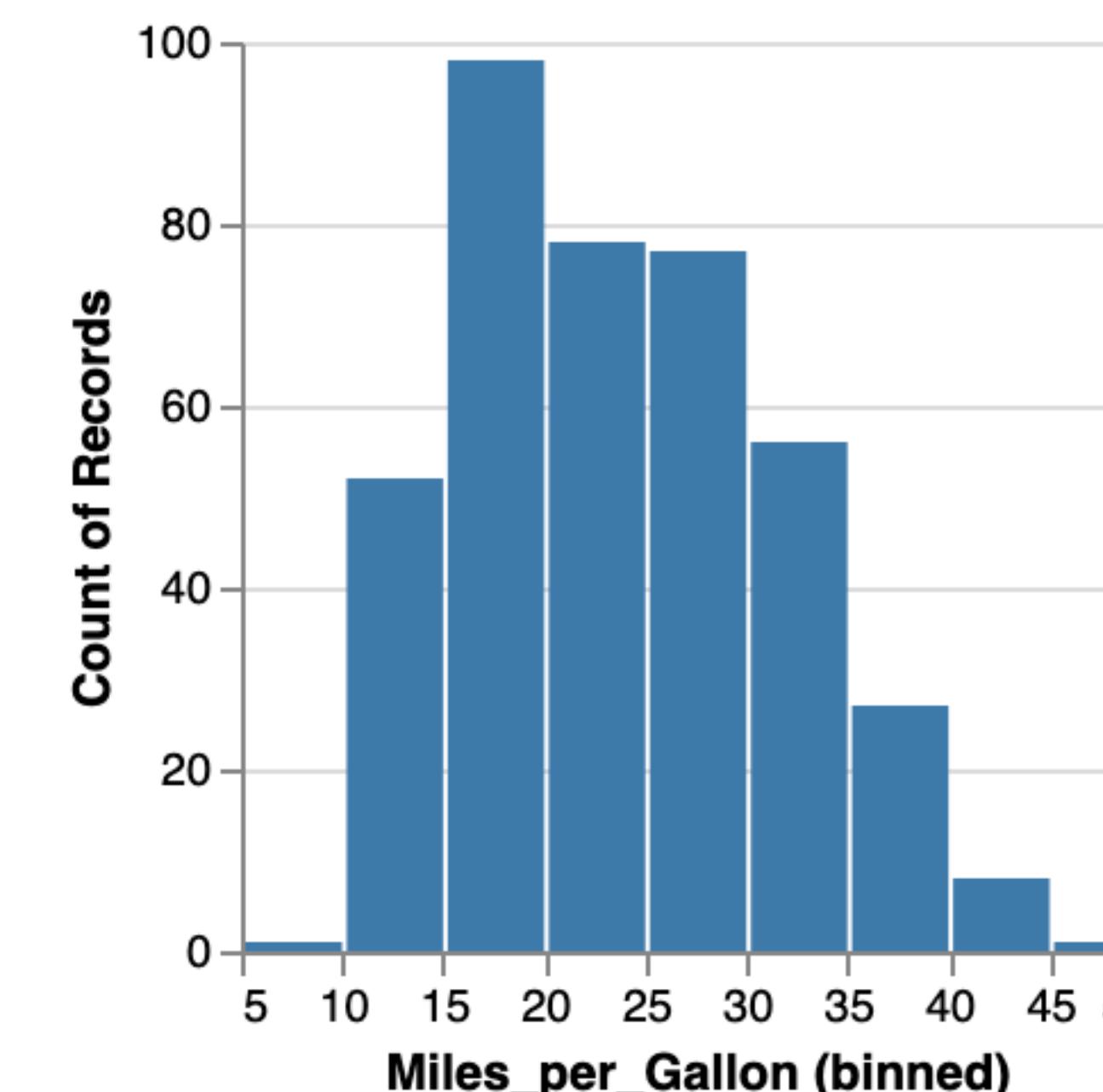
1D quantitative data (Q)

Expressive?

raw



aggregate (count)



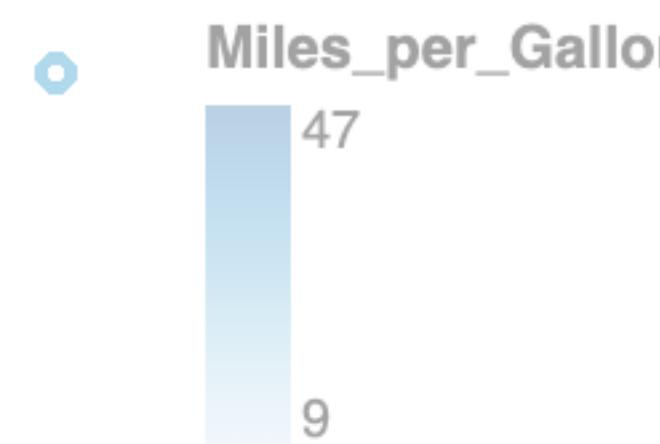
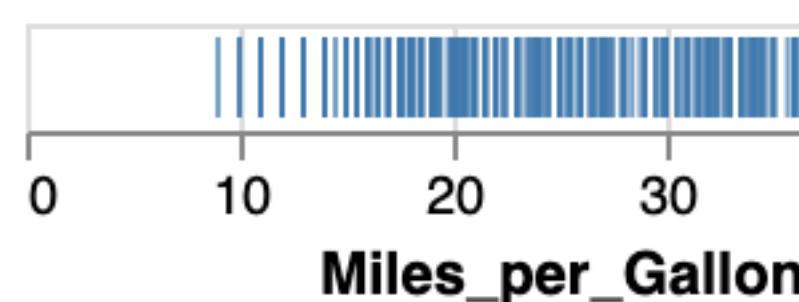
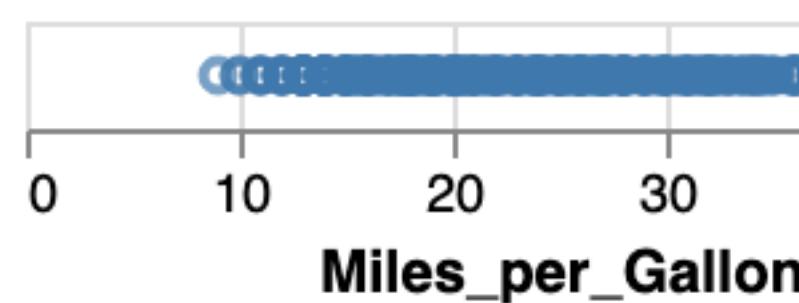
# Visual Encoding = Combinatorial Design Space

# 1D quantitative data (Q)

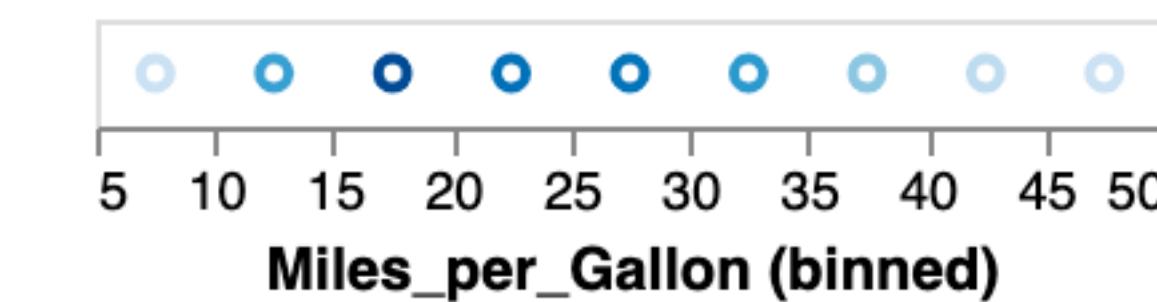
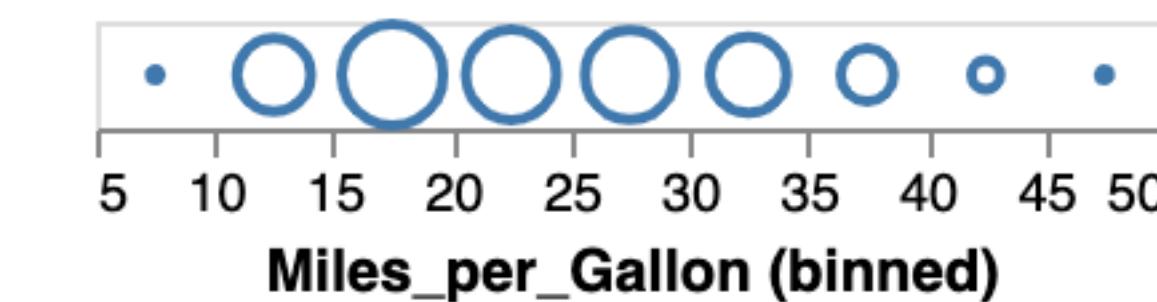
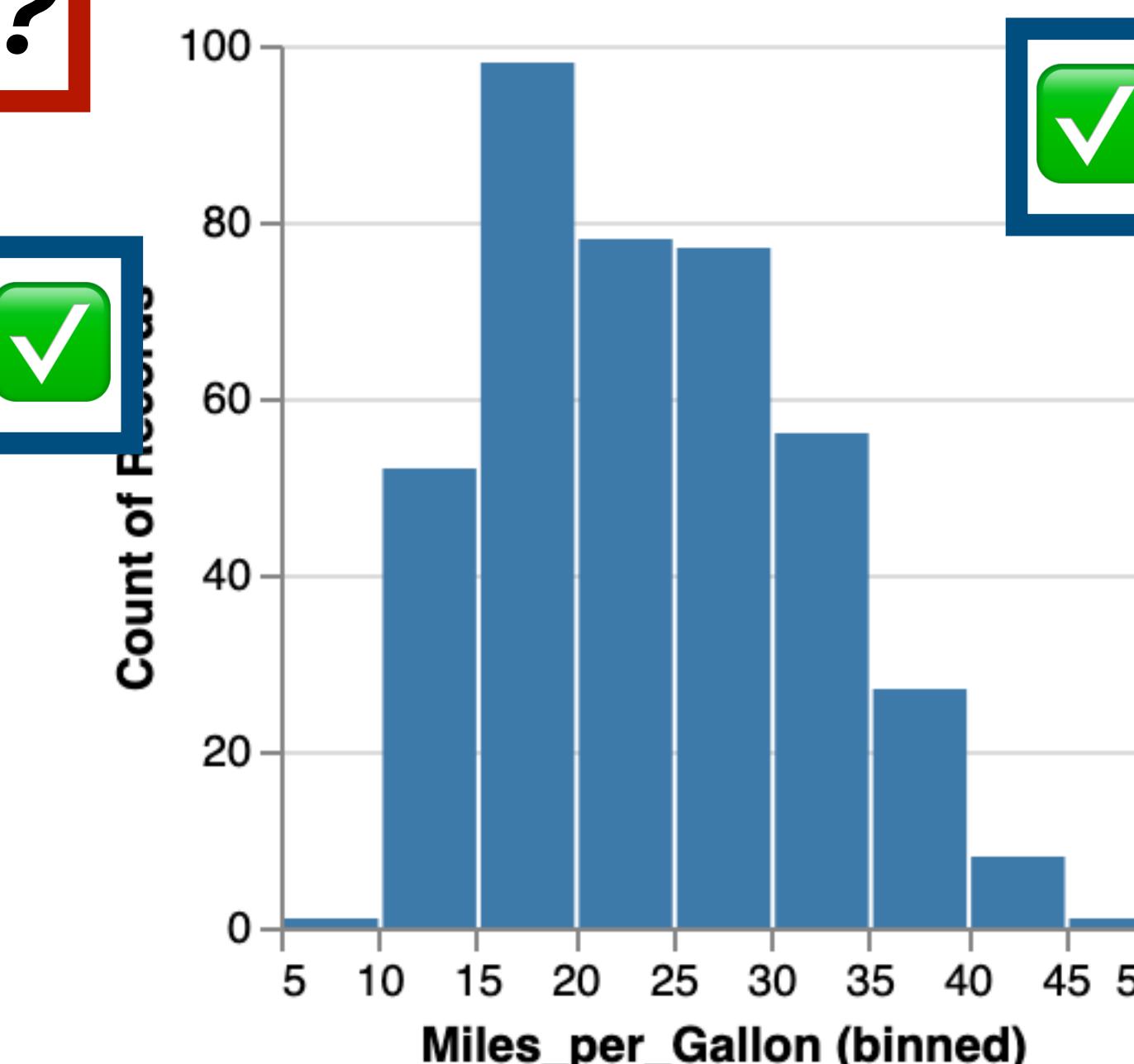
# Expressive?

# Effective\*

raw



# aggregate (count



Count of Records

- 20
- 40
- 60
- 80

!?

# Count of Records

98

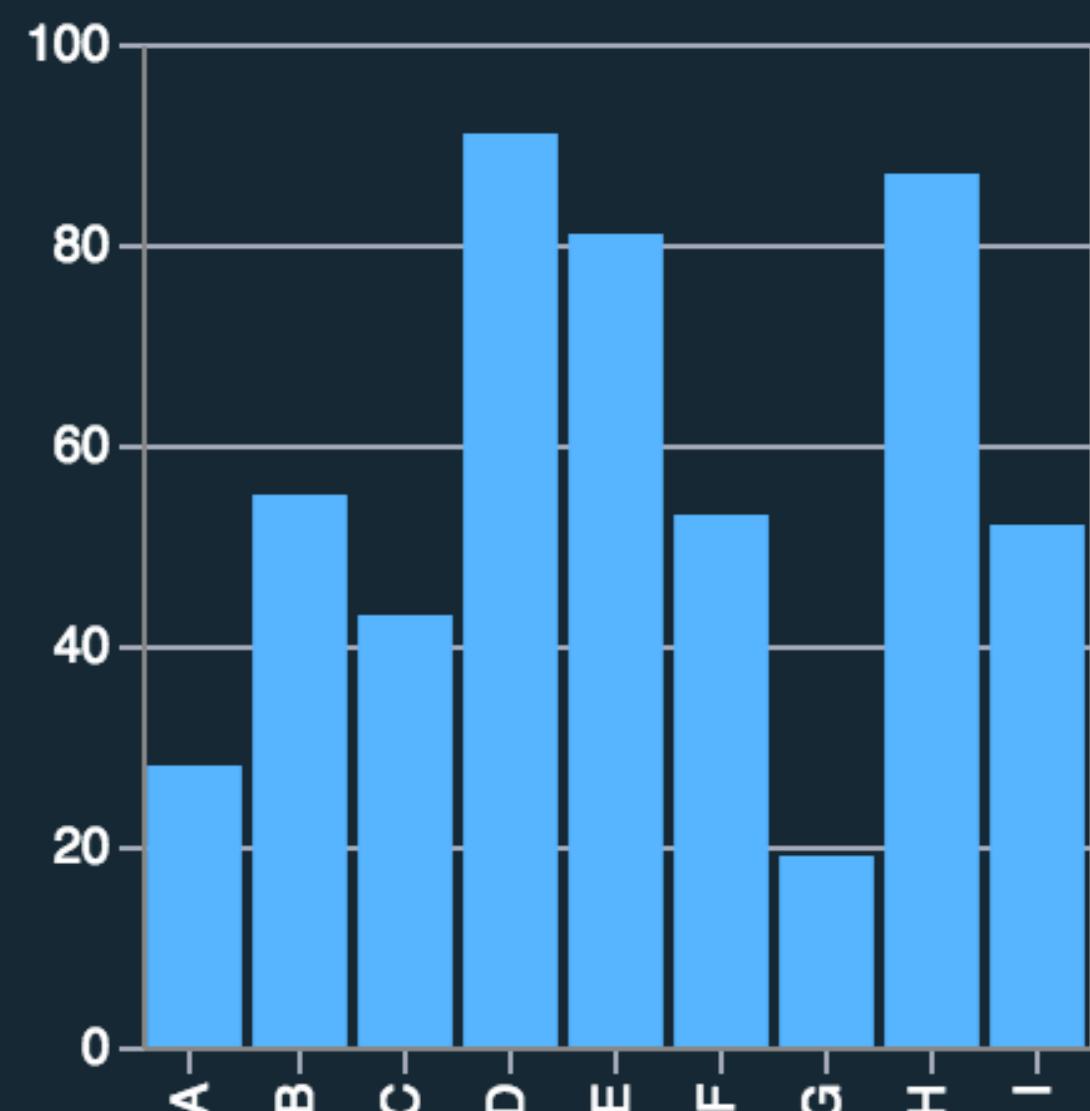
1

!?

<https://vega.github.io/vega/examples/histogram.html>



# Visual Encoding: Nimble Design Moves

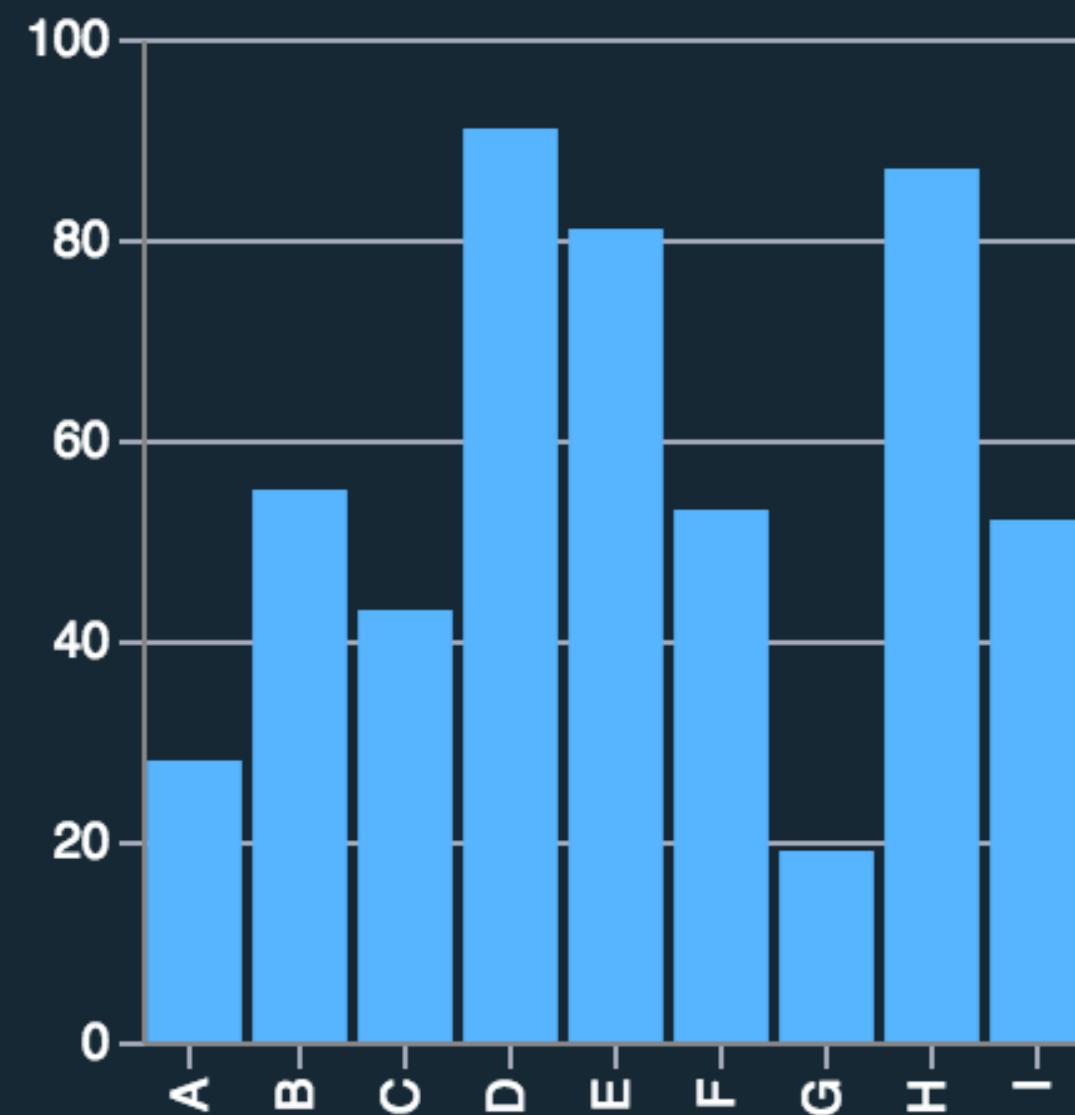


Mark: Bar

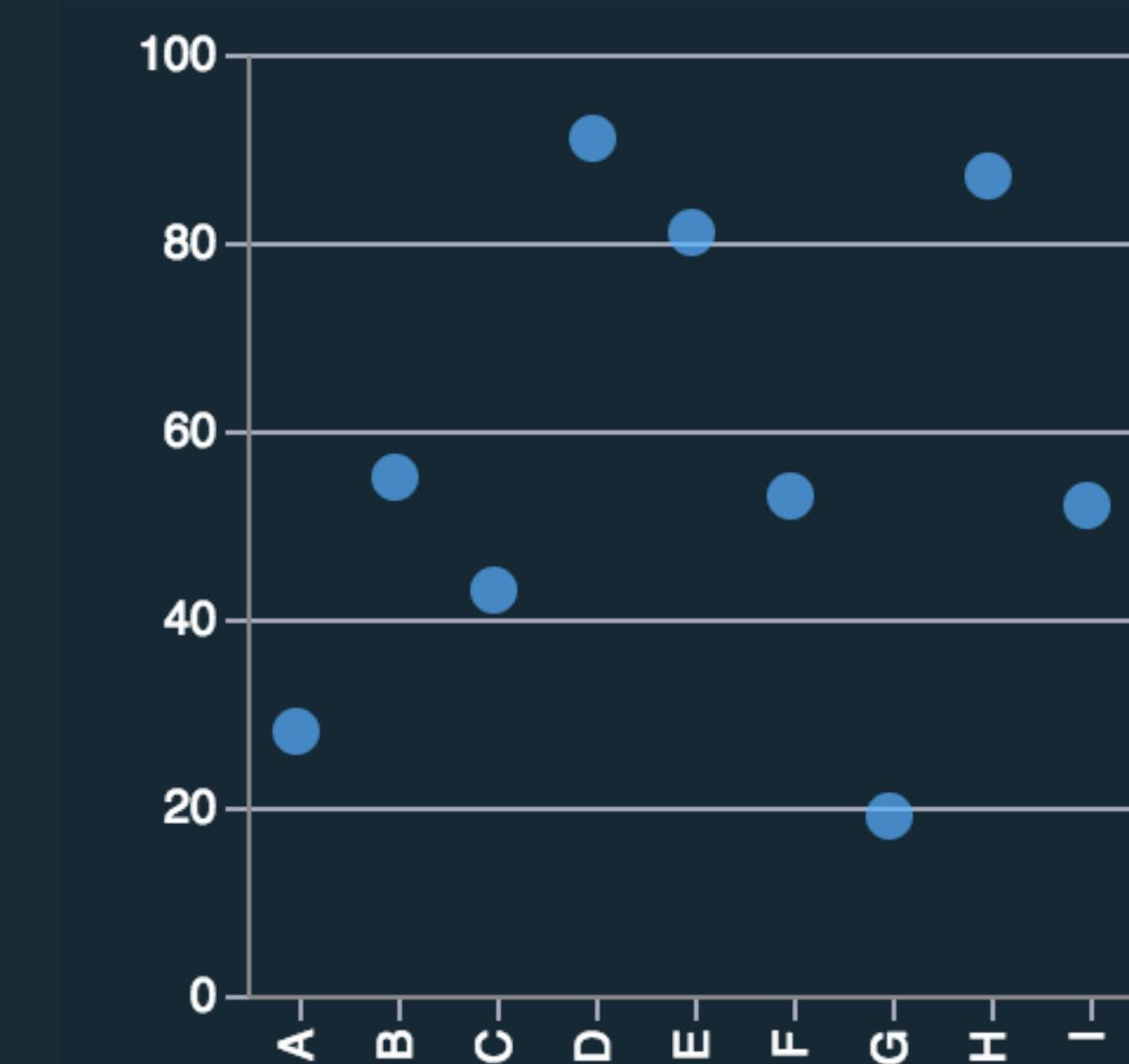
$d_{nominal} \rightarrow X$

$d_{quantitative} \rightarrow y$

# Visual Encoding: Nimble Design Moves

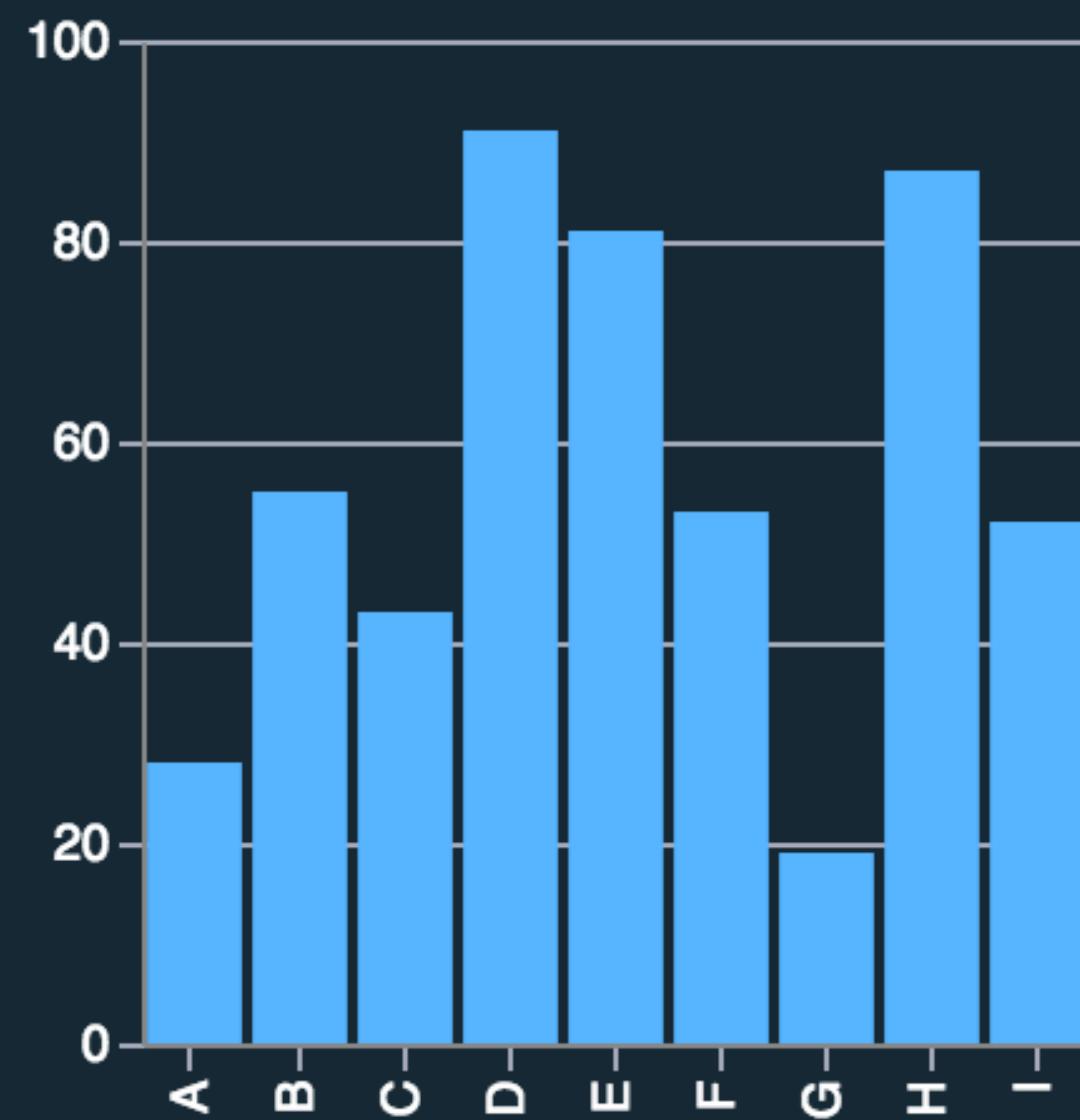


Mark: Bar  
 $d_{nominal} \rightarrow X$   
 $d_{quantitative} \rightarrow y$

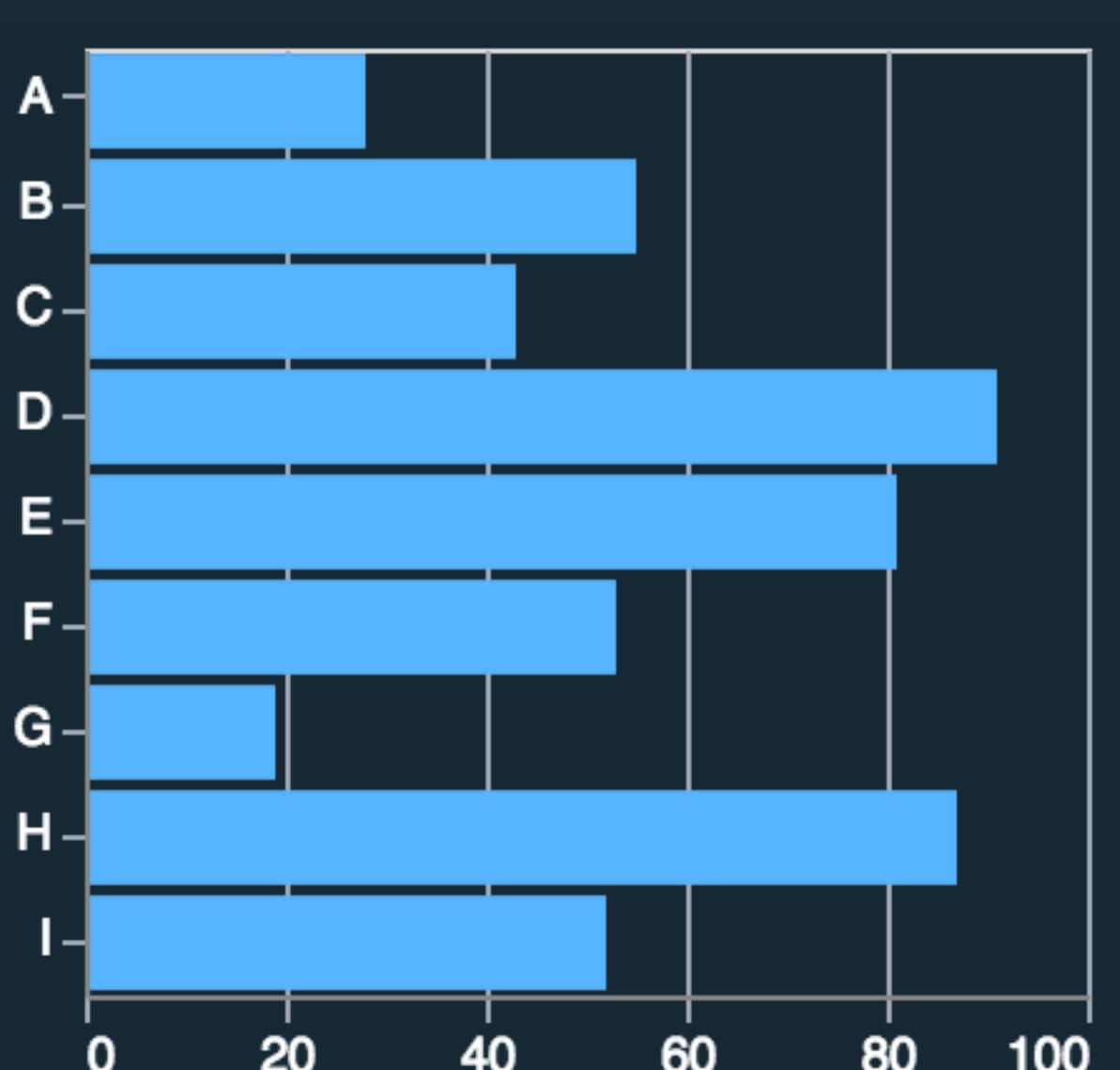


Mark: Point  
 $d_{nominal} \rightarrow X$   
 $d_{quantitative} \rightarrow y$

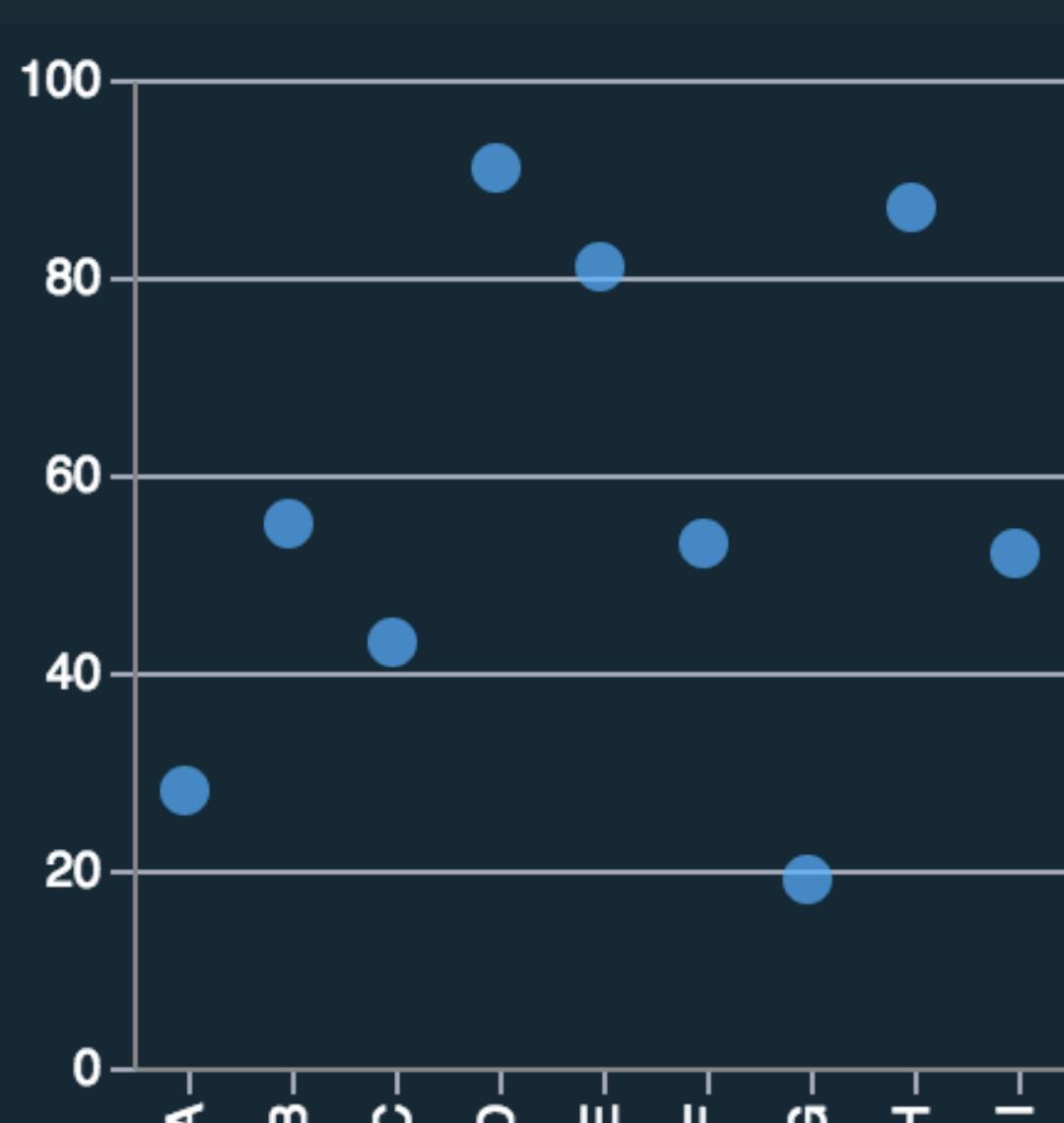
# Visual Encoding: Nimble Design Moves



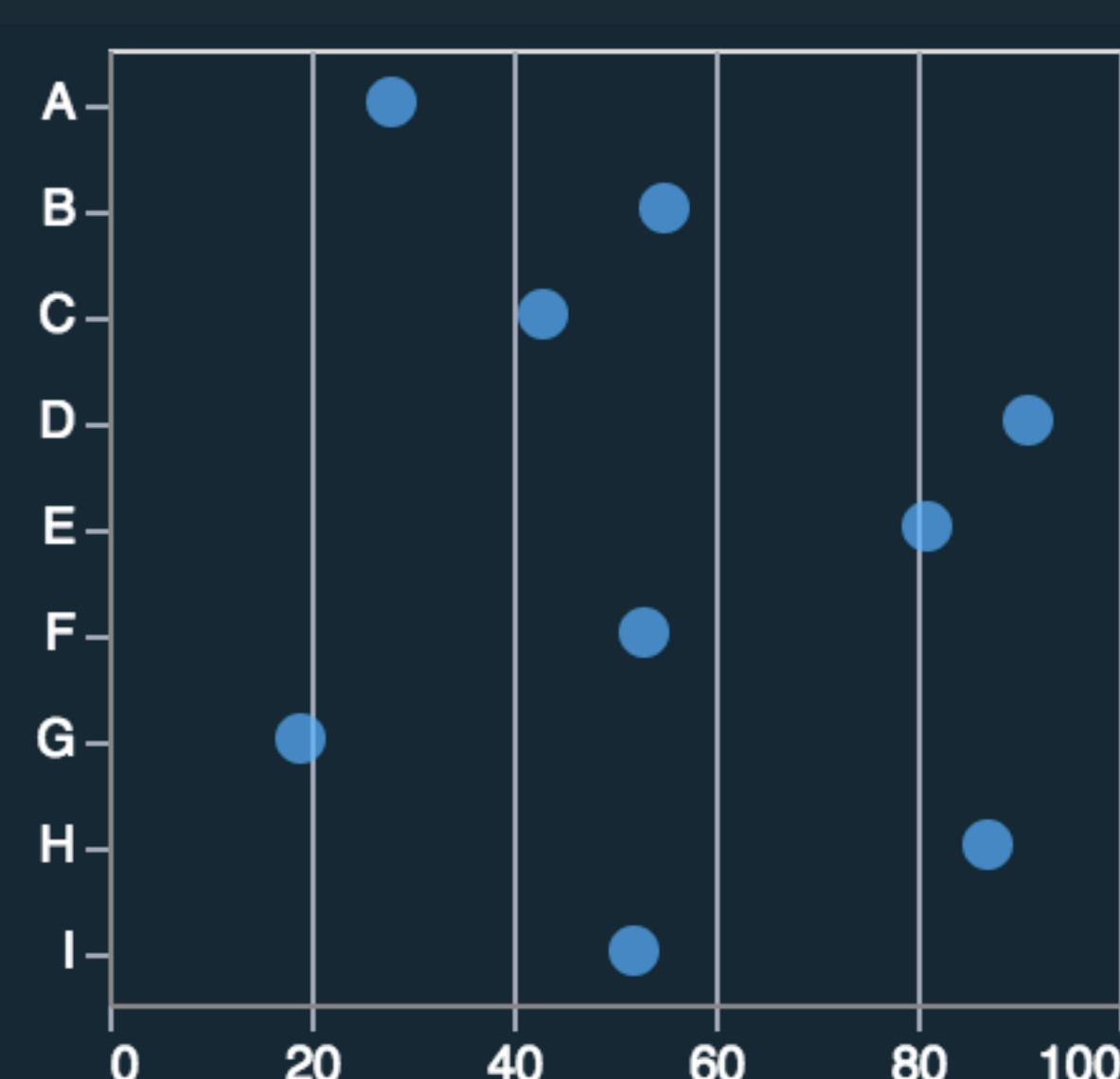
Mark: Bar  
 $d_{nominal} \rightarrow X$   
 $d_{quantitative} \rightarrow y$



Mark: Bar  
 $d_{nominal} \rightarrow y$   
 $d_{quantitative} \rightarrow X$

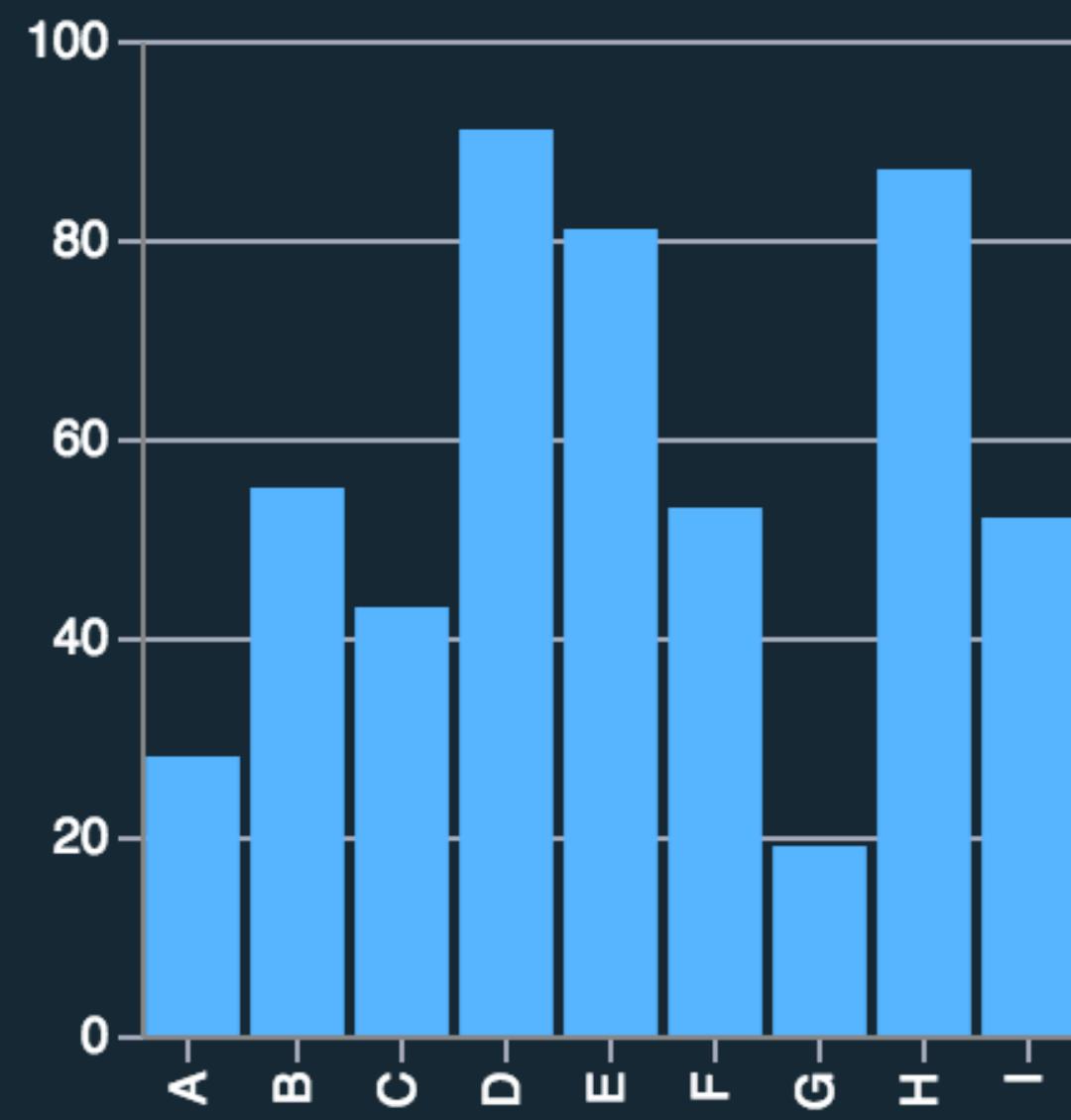


Mark: Point  
 $d_{nominal} \rightarrow X$   
 $d_{quantitative} \rightarrow y$

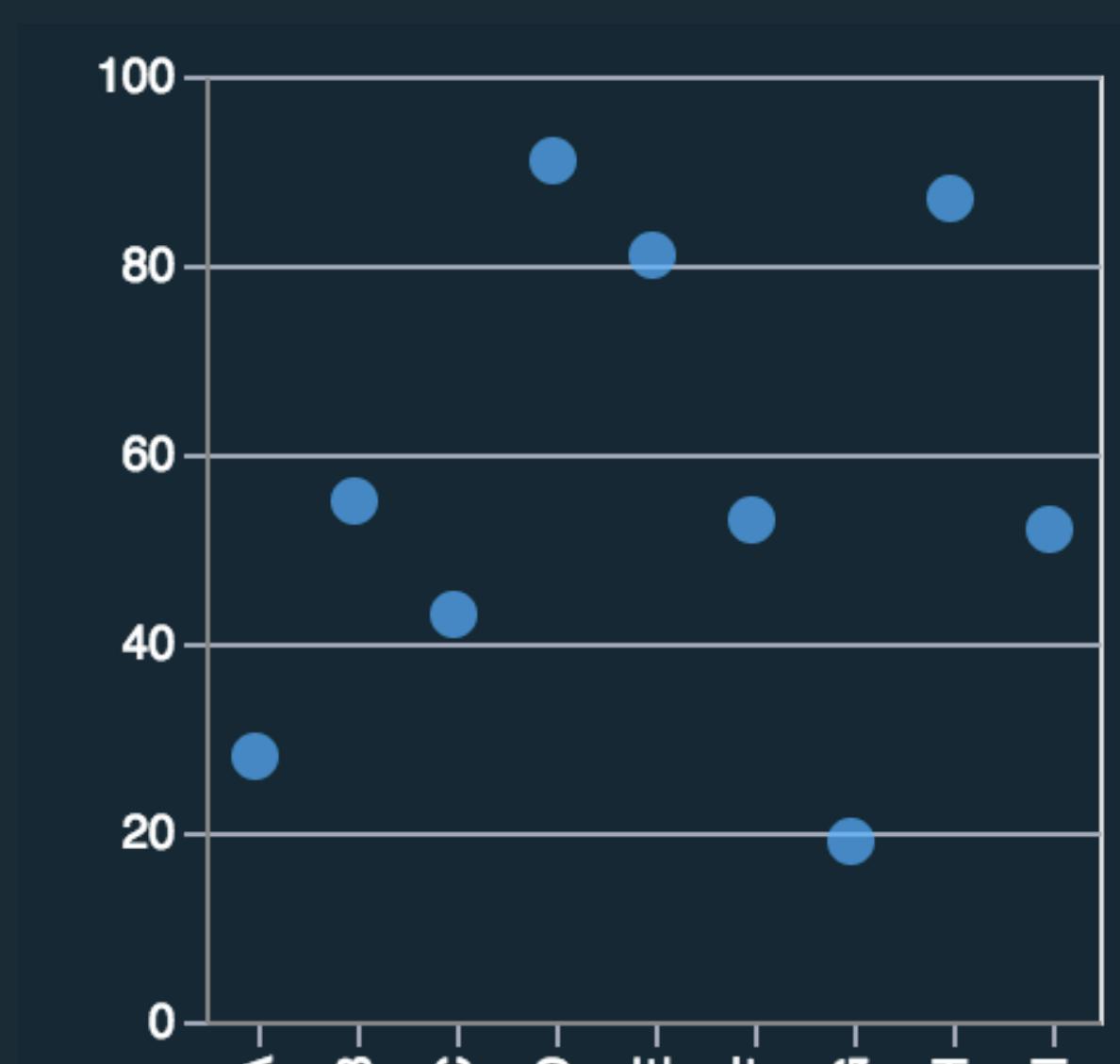


Mark: Point  
 $d_{nominal} \rightarrow y$   
 $d_{quantitative} \rightarrow X$

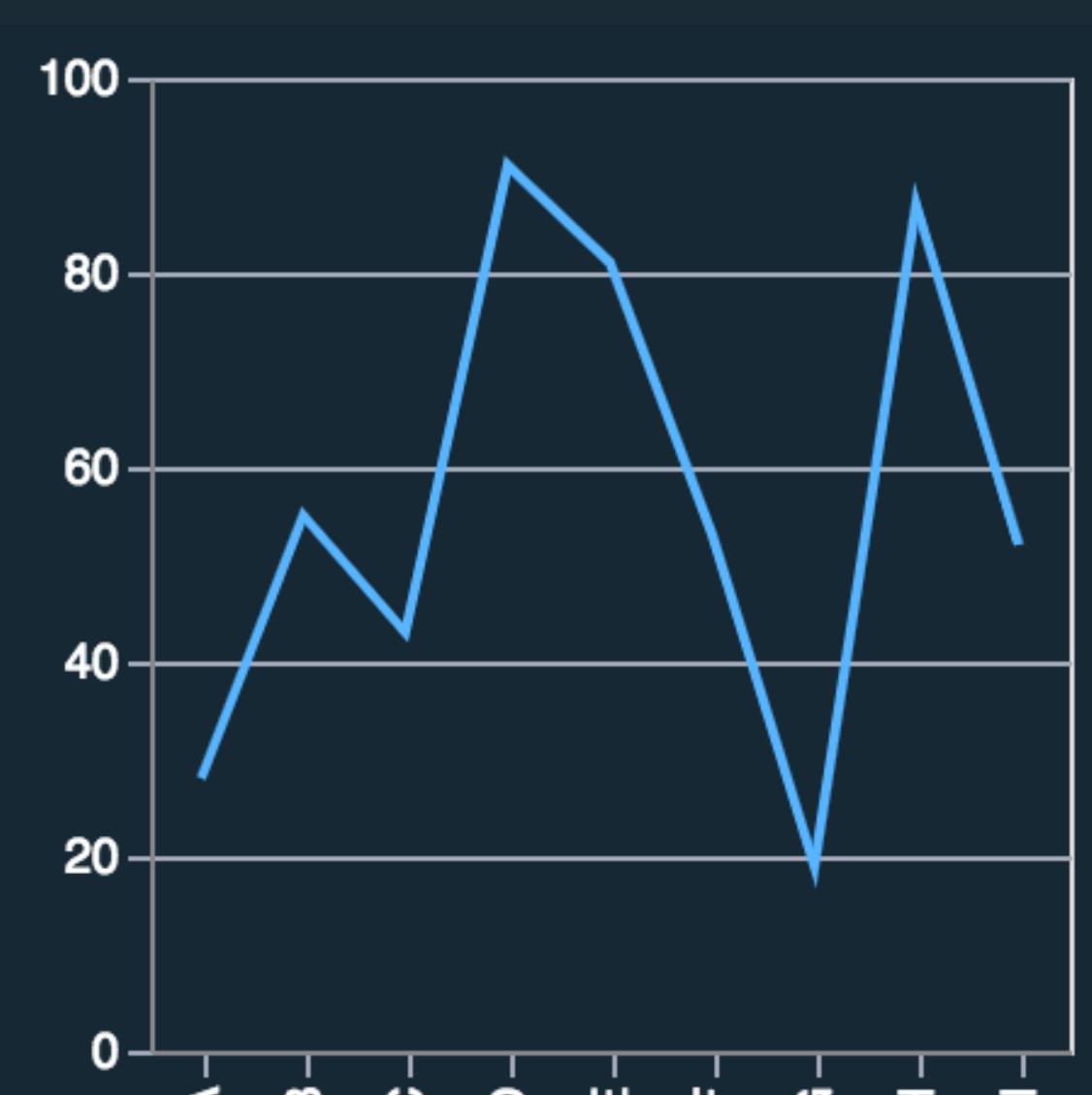
# Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar  
 $d_{\text{nominal}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow Y$

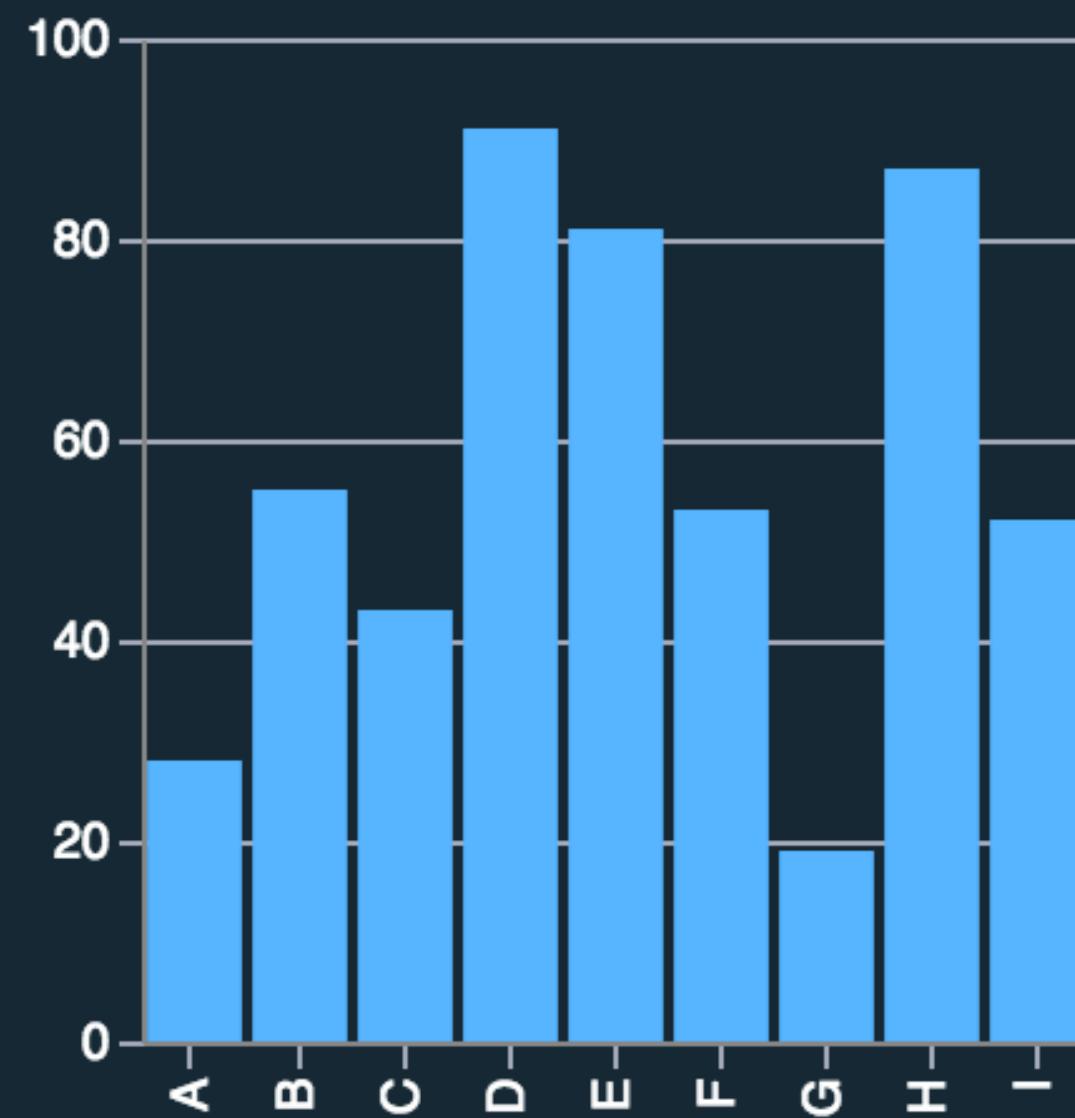


Mark: Point  
 $d_{\text{nominal}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow Y$

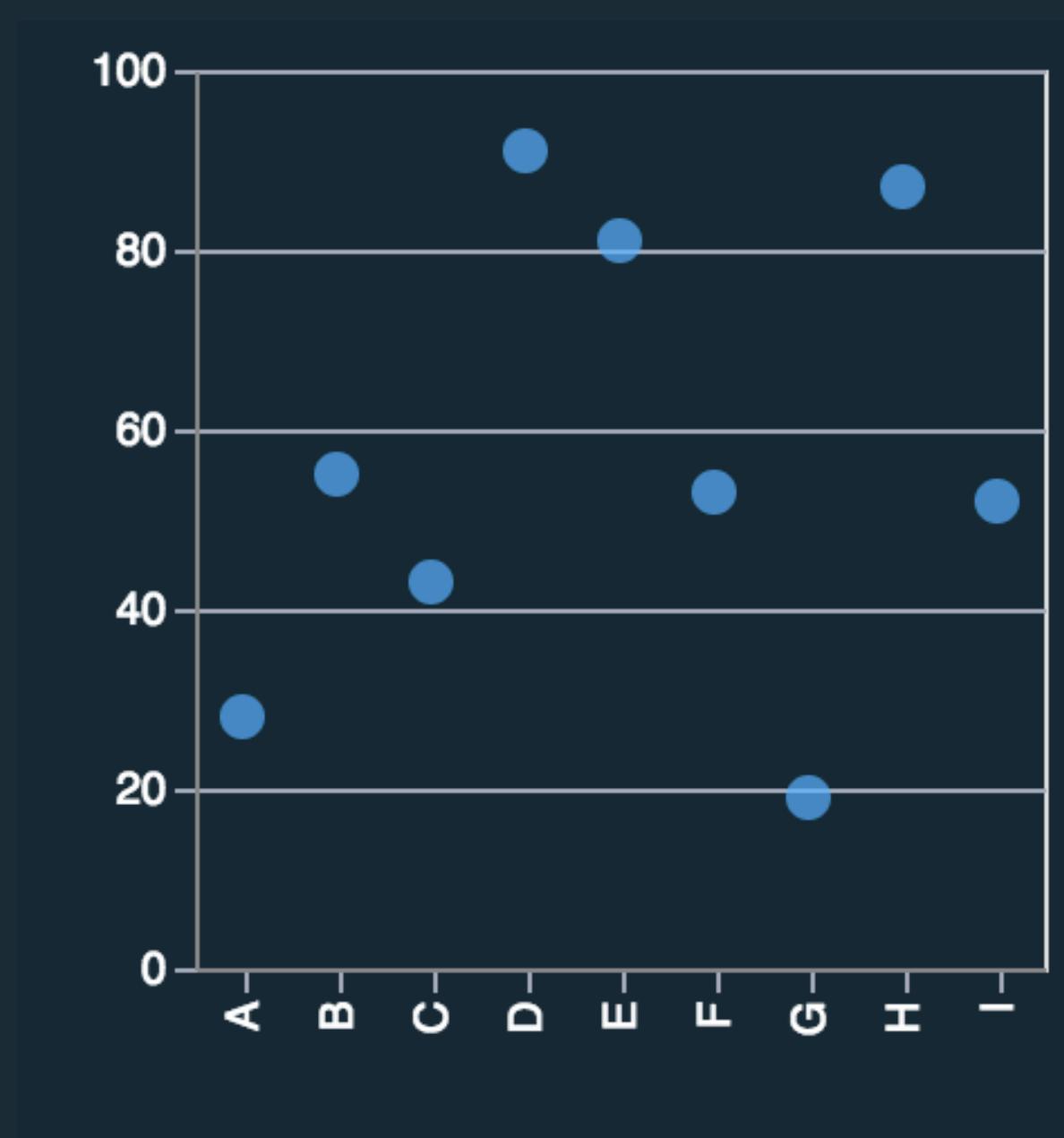


Mark: Line  
 $d_{\text{nominal}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow Y$

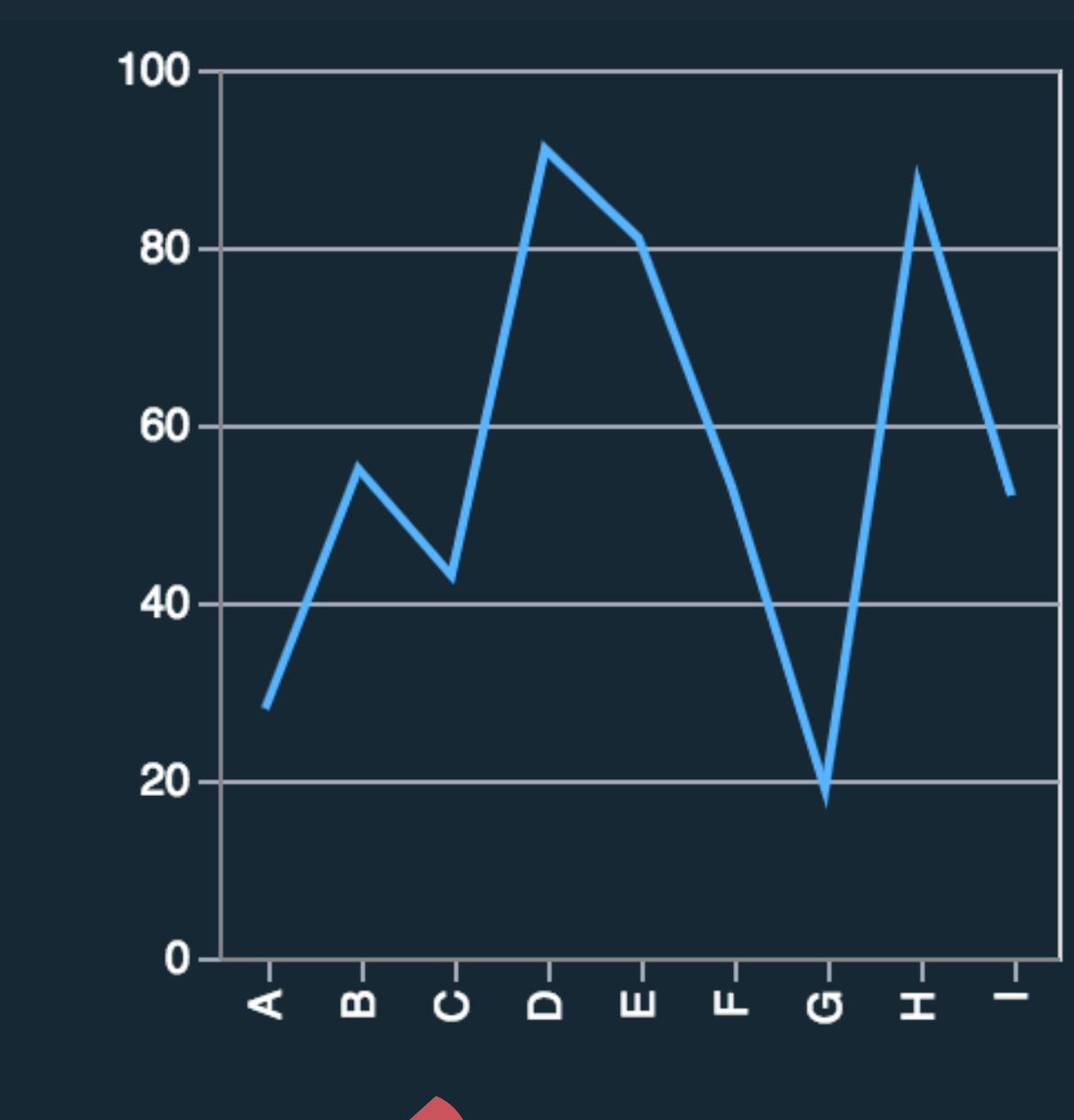
# Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar  
 $d_{\text{nominal}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow y$



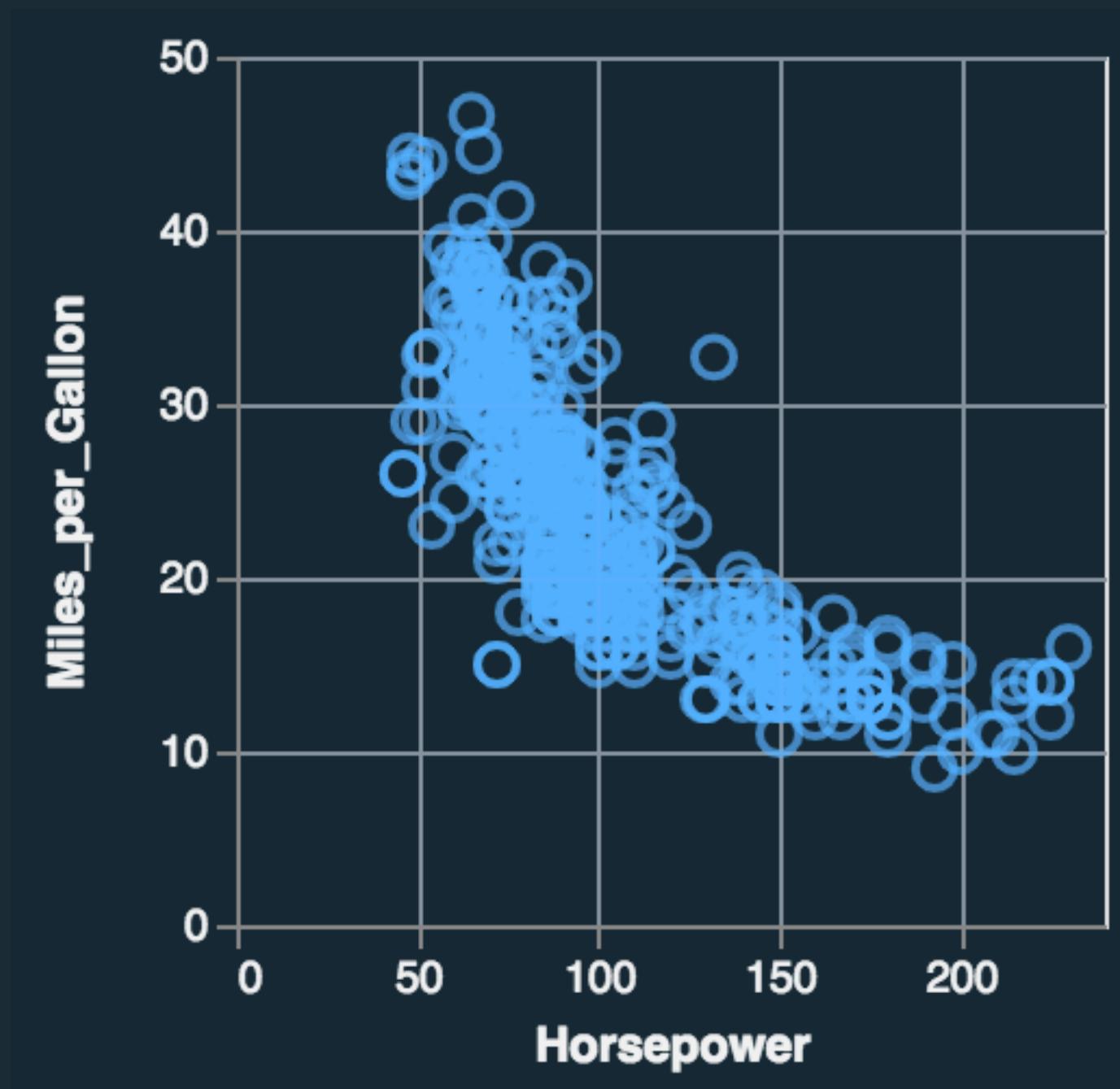
Mark: Point  
 $d_{\text{nominal}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow y$



Mark: Line  
 $d_{\text{nominal}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow y$

**Violates expressiveness:** the line mark implies a trend across the various categories.

# Visual Encoding

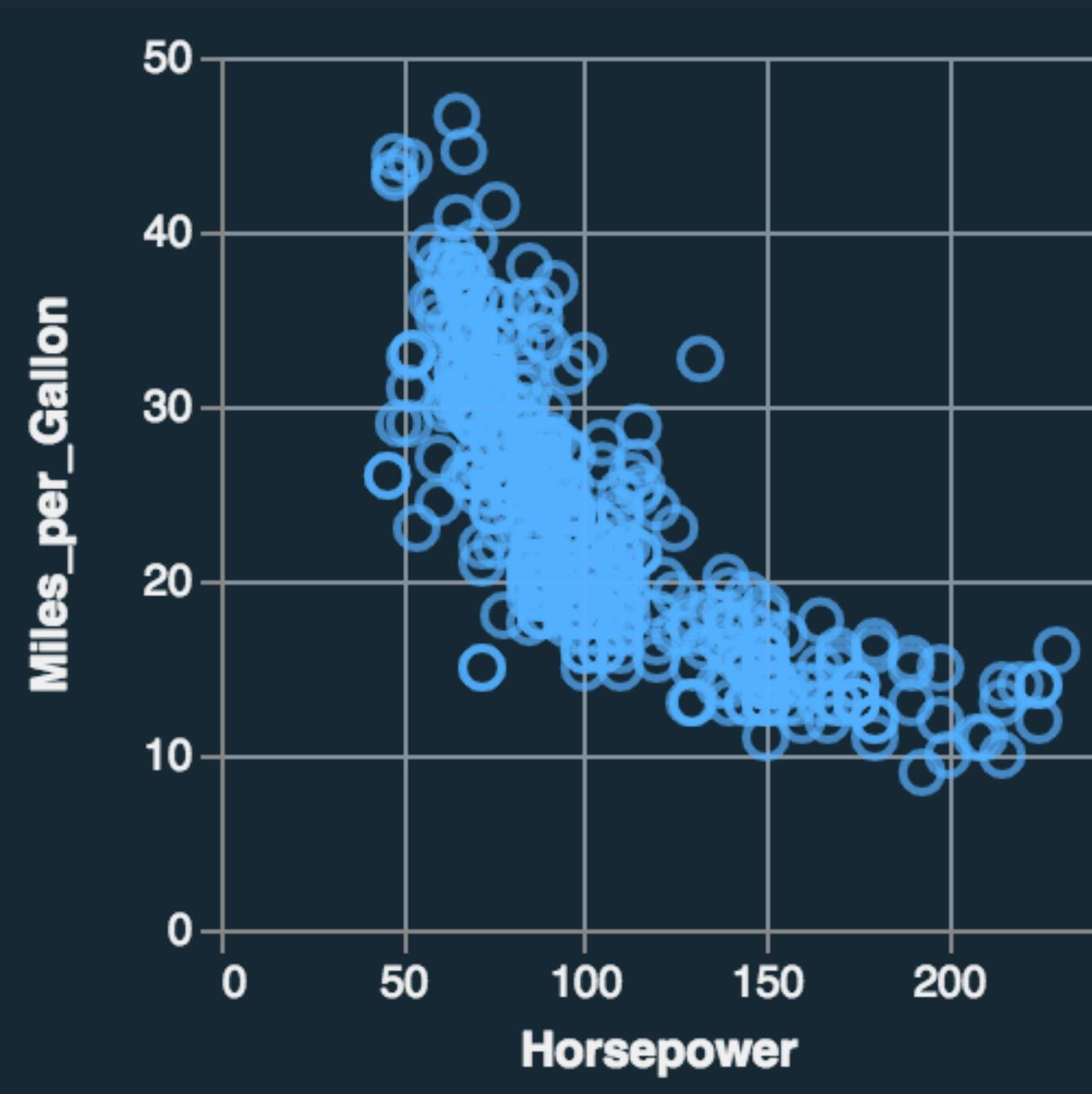


Mark: Point

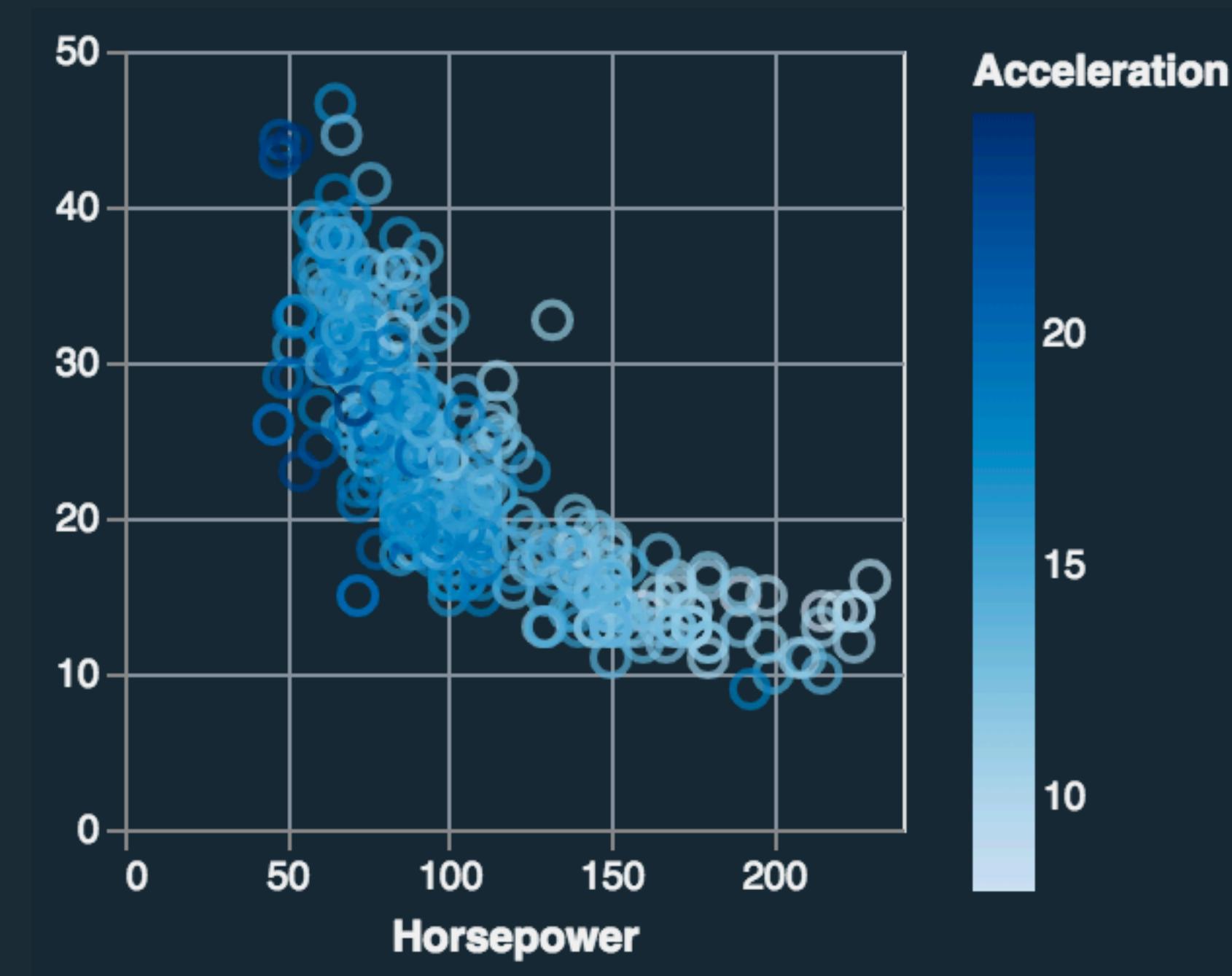
$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

# Visual Encoding

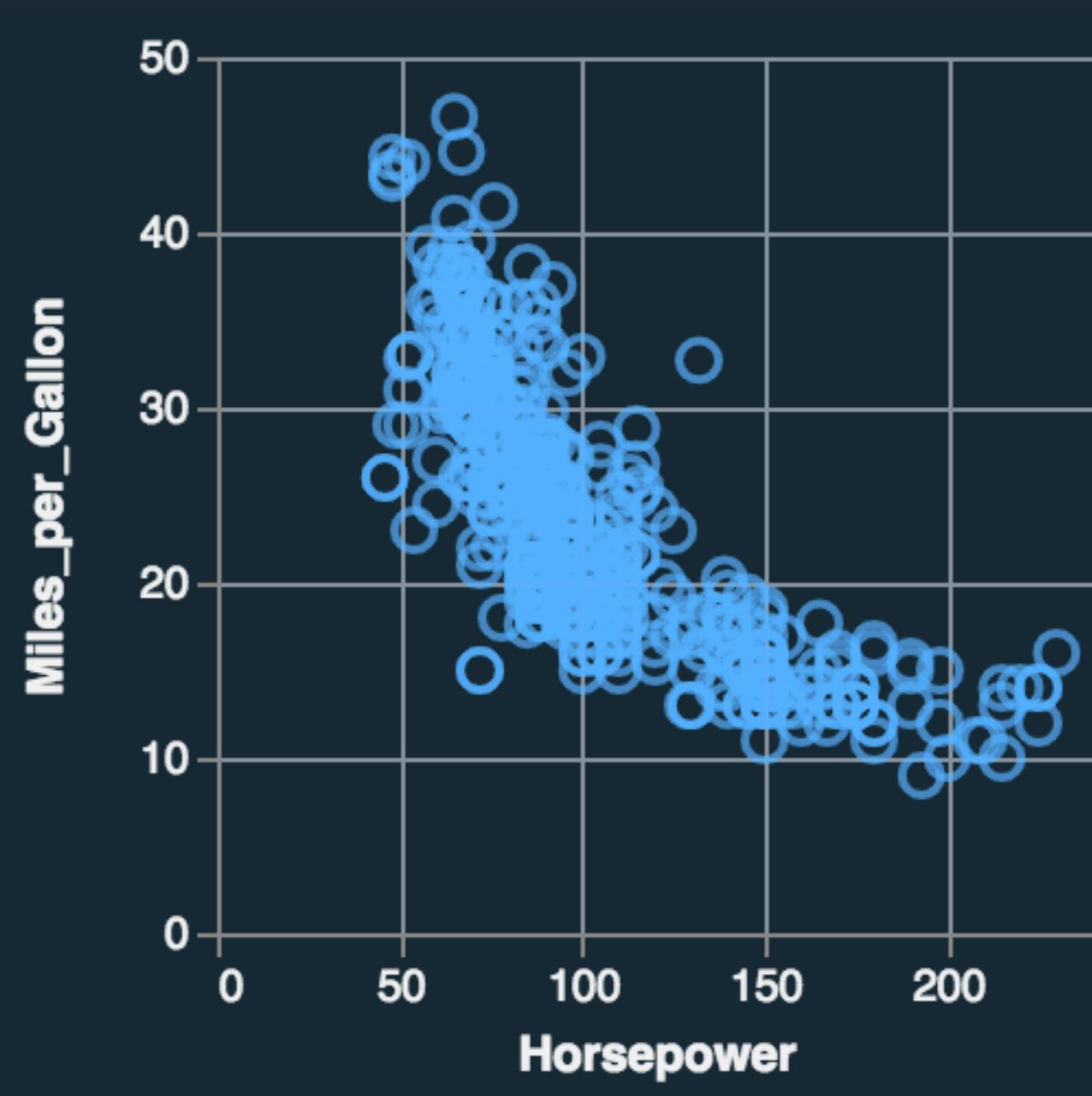


Mark: Point  
 $d_{\text{quantitative}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow y$

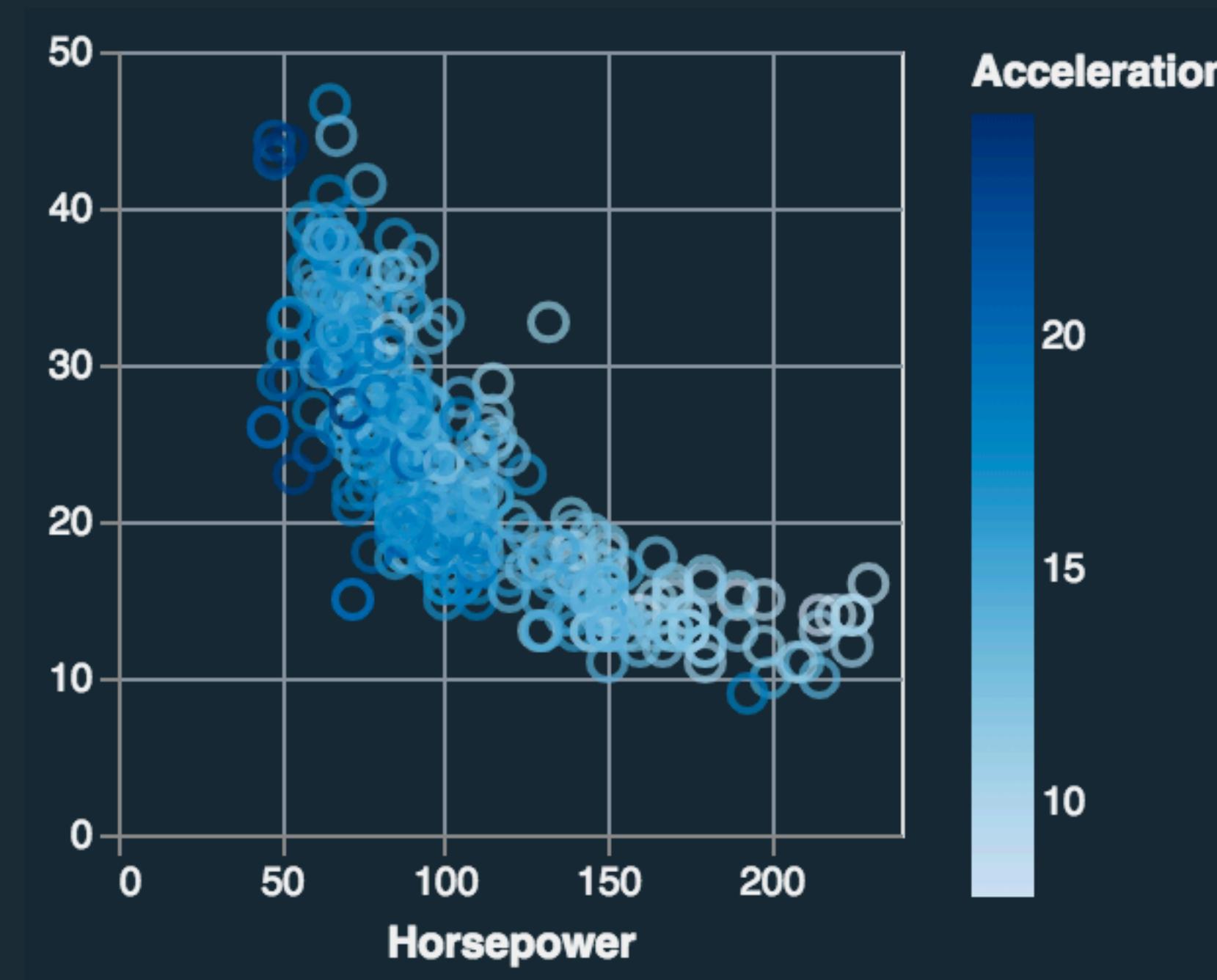


Mark: Point  
 $d_{\text{quantitative}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow y$   
 $d_{\text{quantitative}} \rightarrow \text{color}$

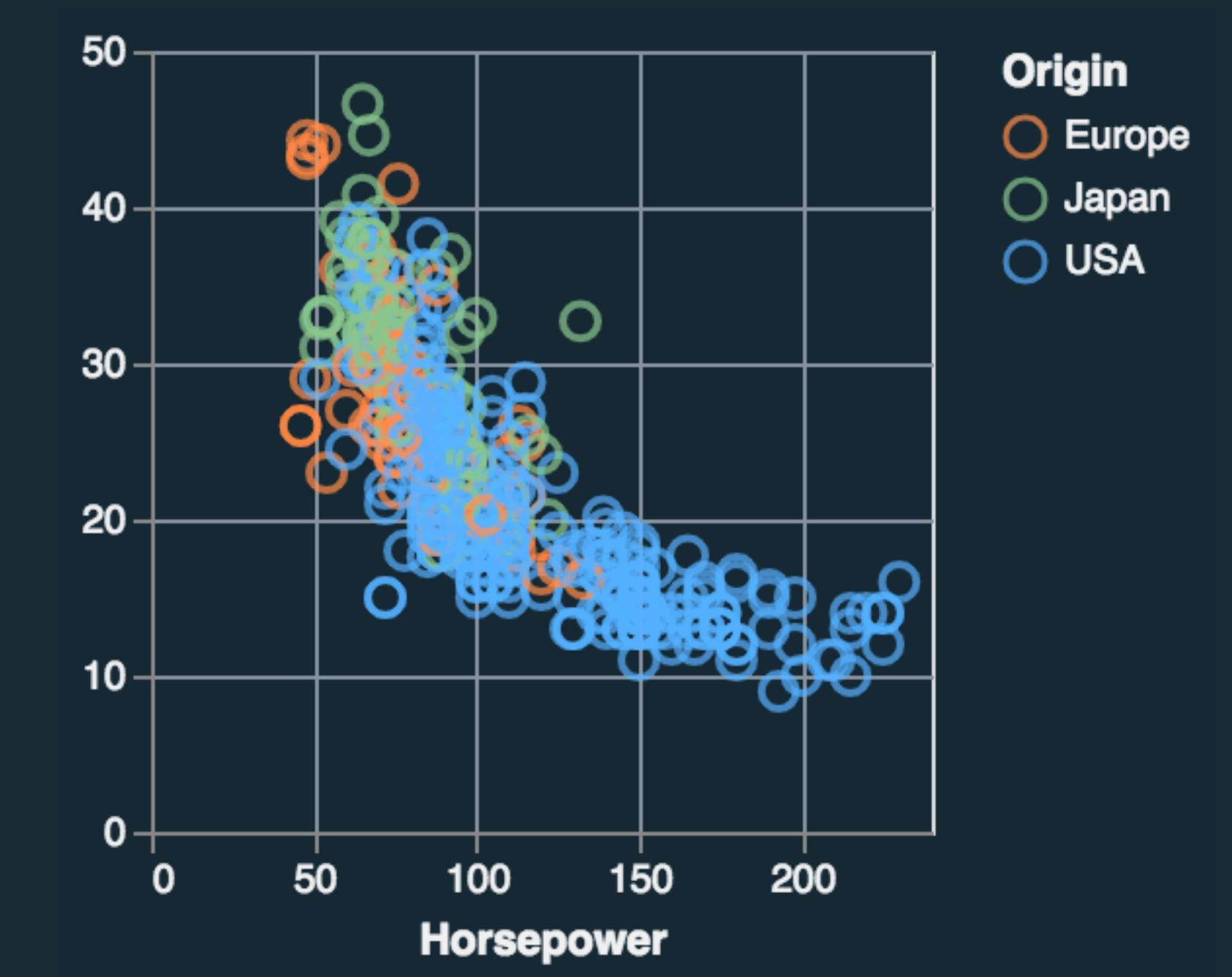
# Visual Encoding



Mark: Point  
 $d_{\text{quantitative}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow y$

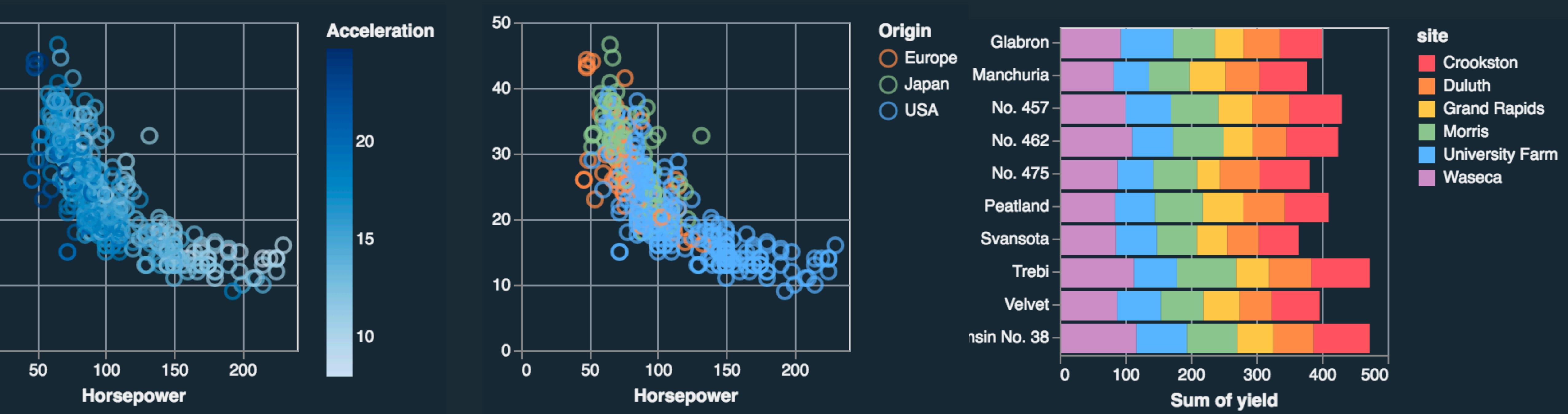


Mark: Point  
 $d_{\text{quantitative}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow y$   
 $d_{\text{quantitative}} \rightarrow \text{color}$



Mark: Point  
 $d_{\text{quantitative}} \rightarrow X$   
 $d_{\text{quantitative}} \rightarrow y$   
 $d_{\text{nominal}} \rightarrow \text{color}$

# Visual Encoding



Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{quantitative}} \rightarrow \text{color}$

Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

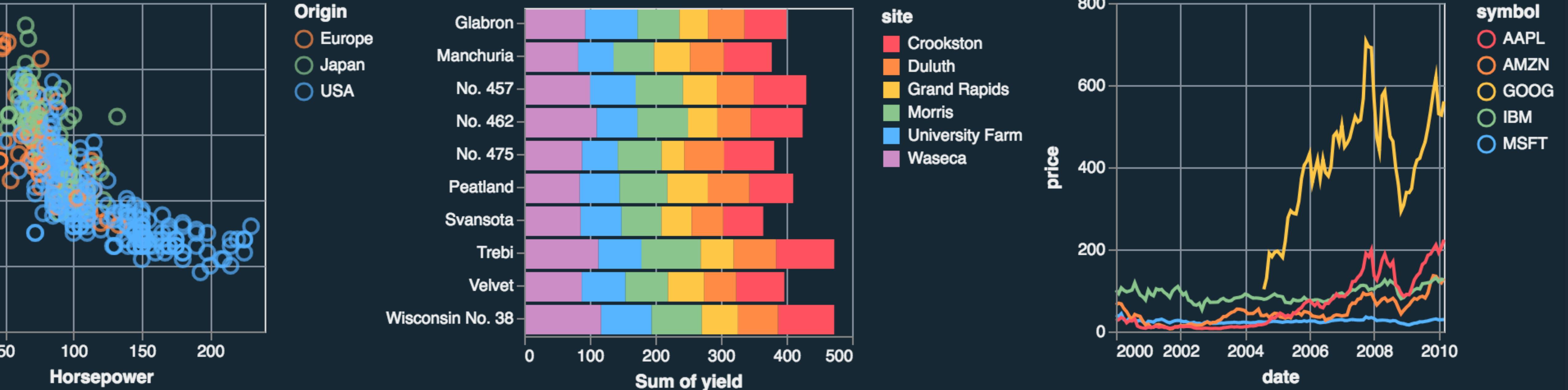
Mark: Bar

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{nominal}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

# Visual Encoding



Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

Mark: Bar

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{nominal}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

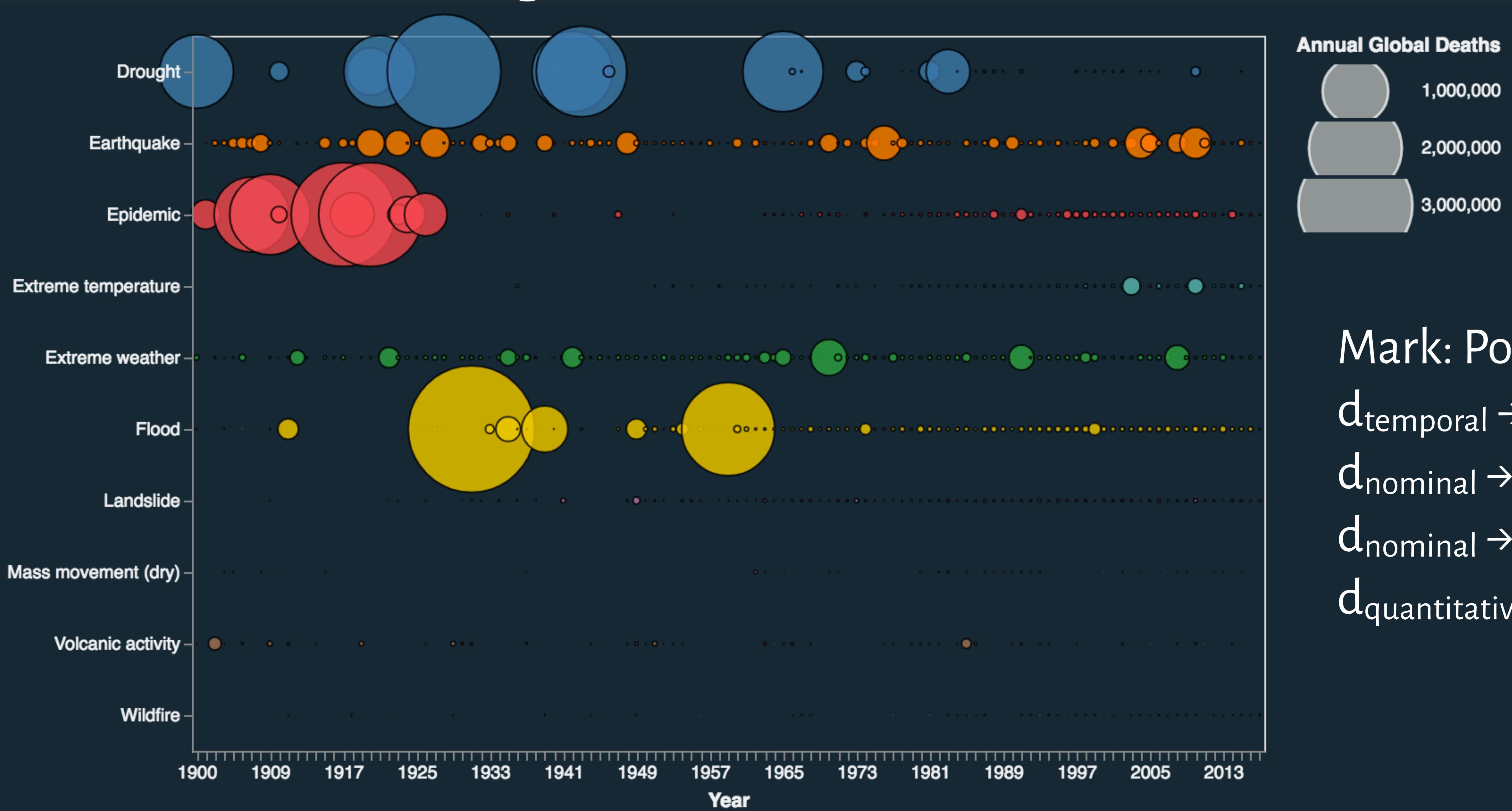
Mark: Line

$d_{\text{temporal}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

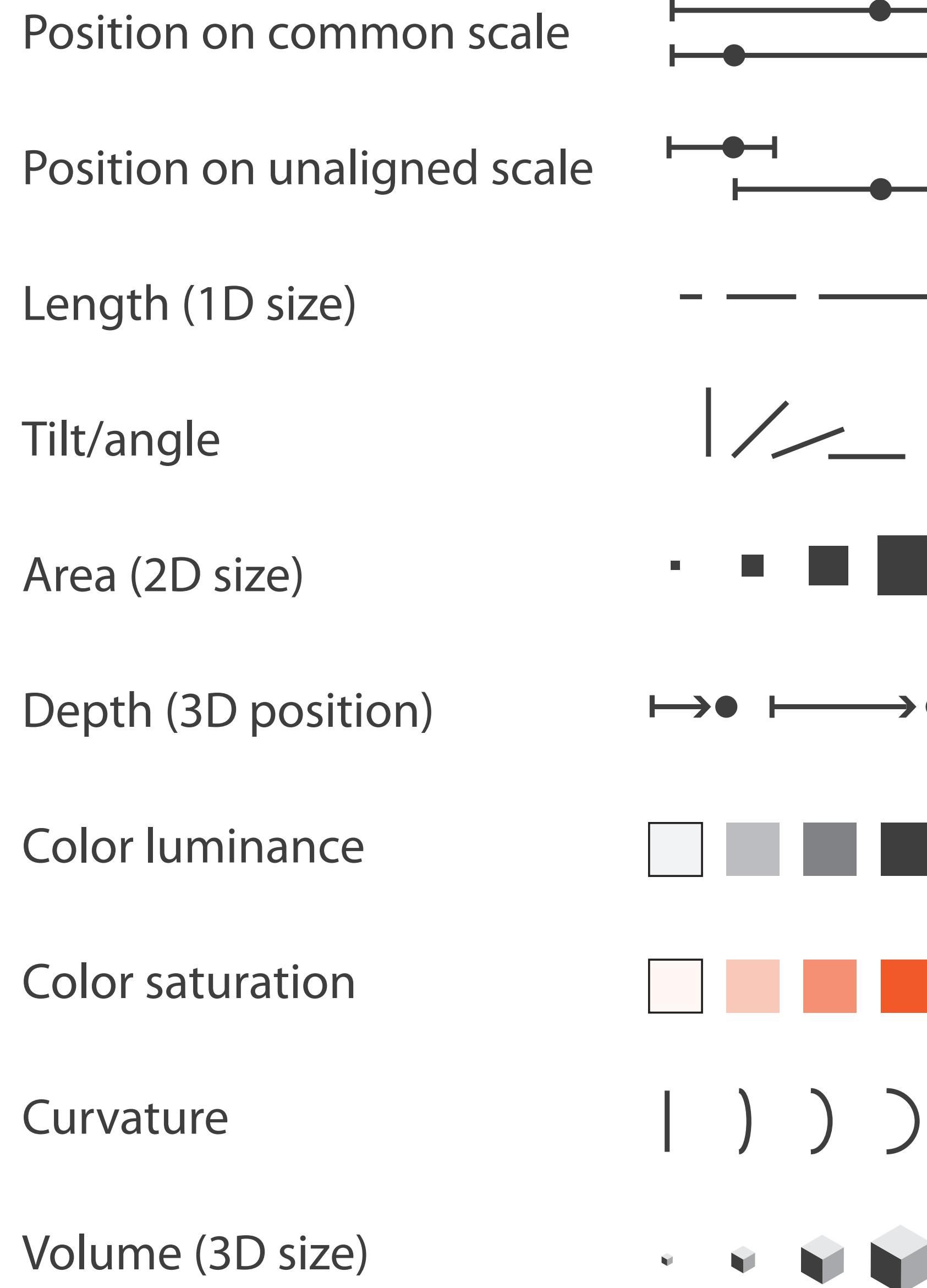
# Visual Encoding



# Effective Visual Encodings

## Channels: Expressiveness Types and Effectiveness Ranks

### → Magnitude Channels: Ordered Attributes



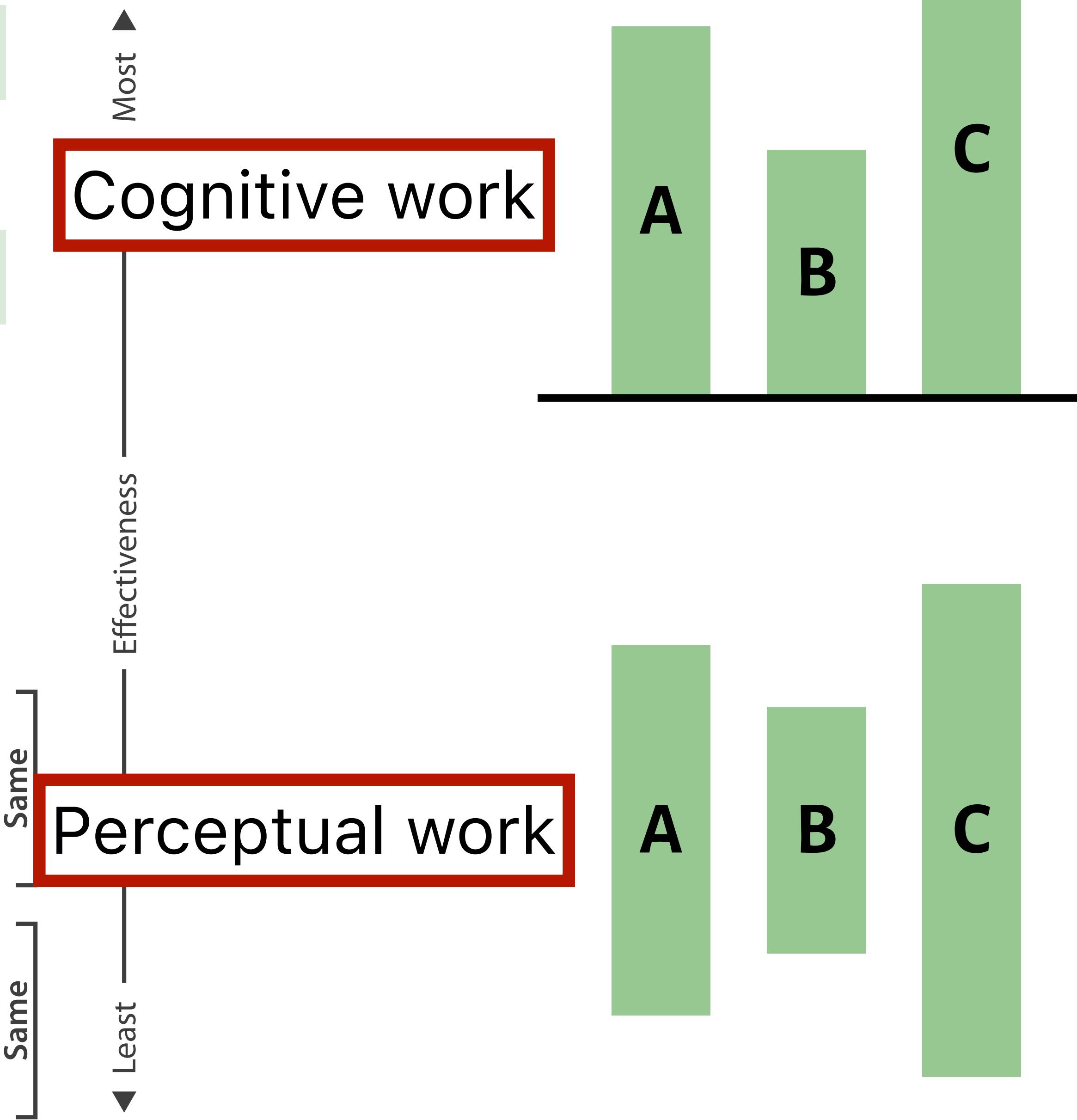
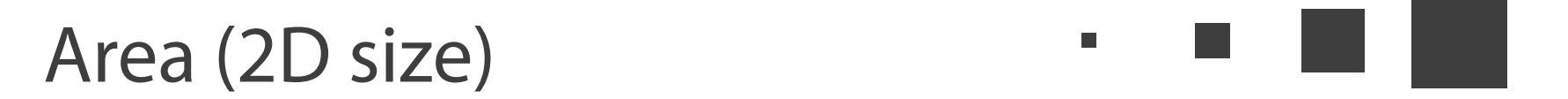
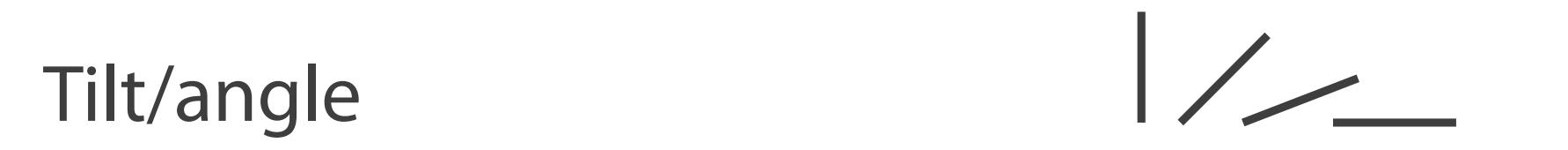
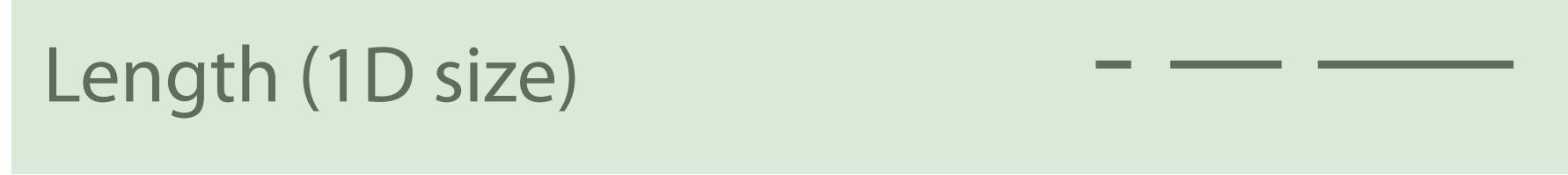
### → Identity Channels: Categorical Attributes



Tamara Munzner, *Visualization Analysis and Design* (2014).

## Channels: Expressiveness Types and Effectiveness Ranks

### → **Magnitude** Channels: **O** or **Q** attributes



# Channels: Expressiveness Types and Effectiveness Ranks

## → **Magnitude** Channels: **O** or **Q** attributes

Position on common scale



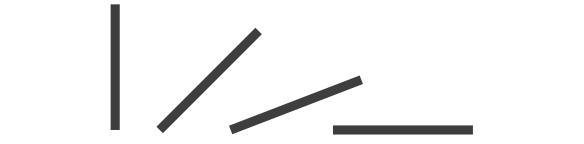
Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



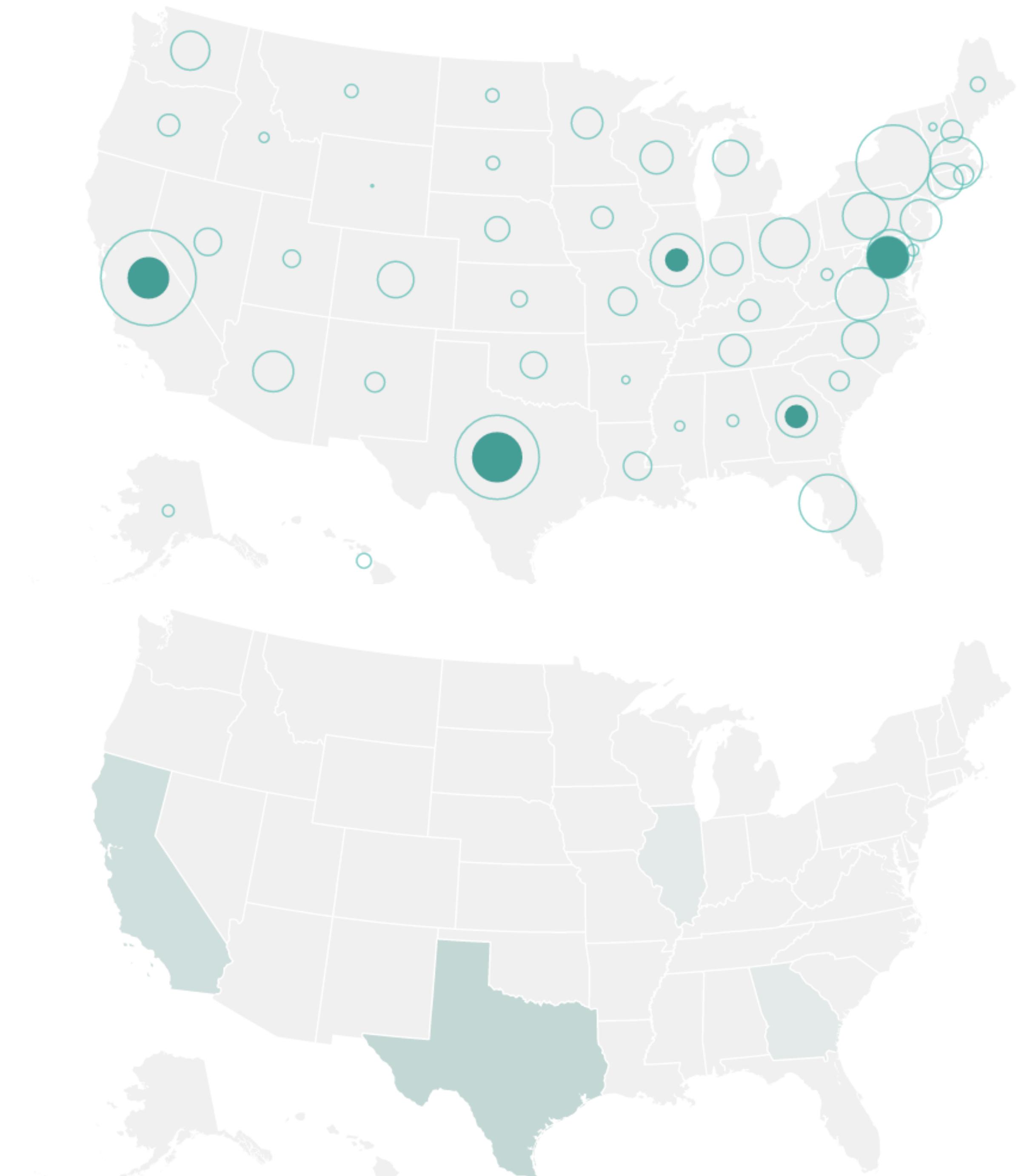
Curvature



Volume (3D size)



▲ Most Effectiveness ▼ Least Effectiveness



# Artery Visualization

Rainbow Palette



Shear  
Stress (Pa)

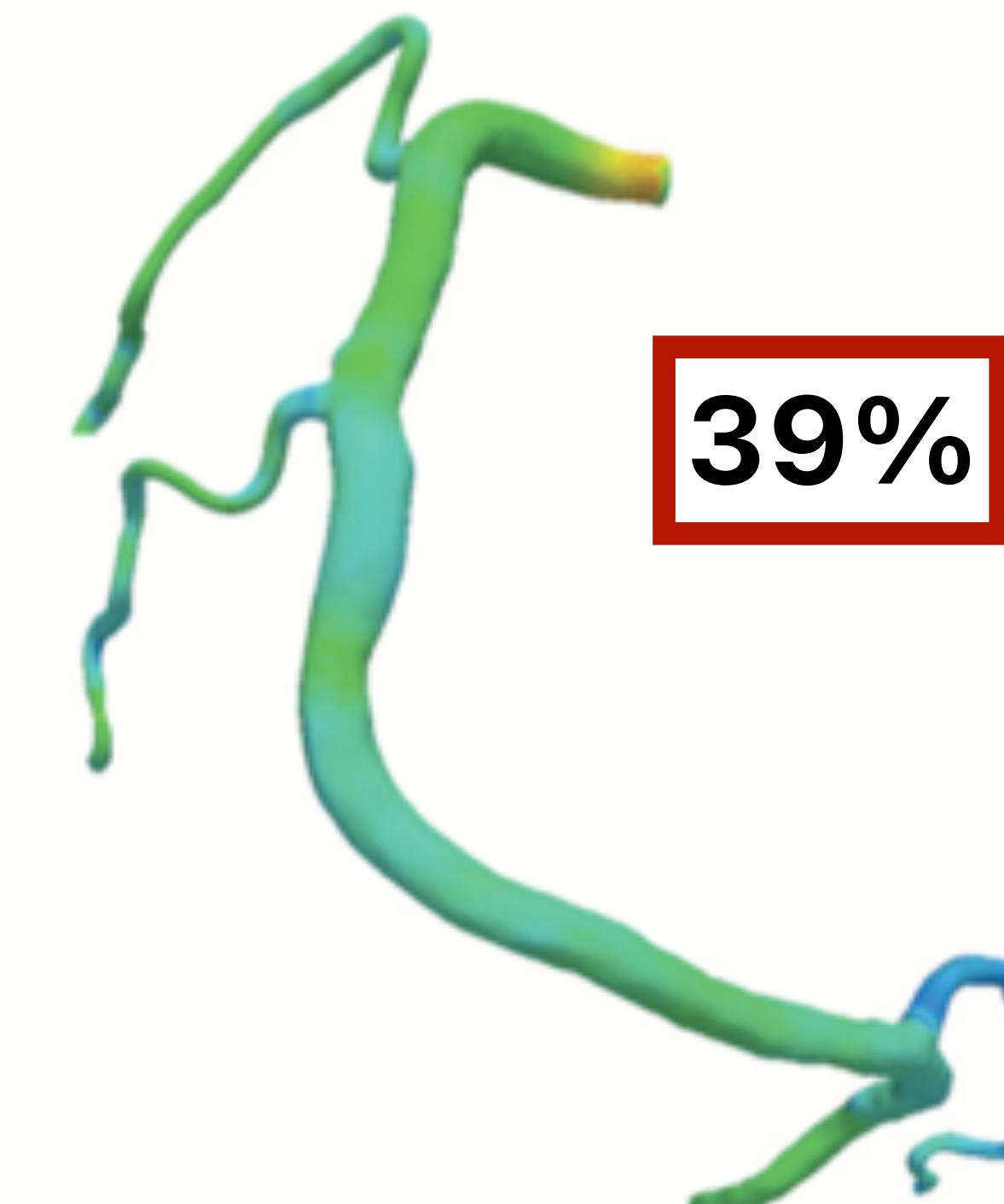
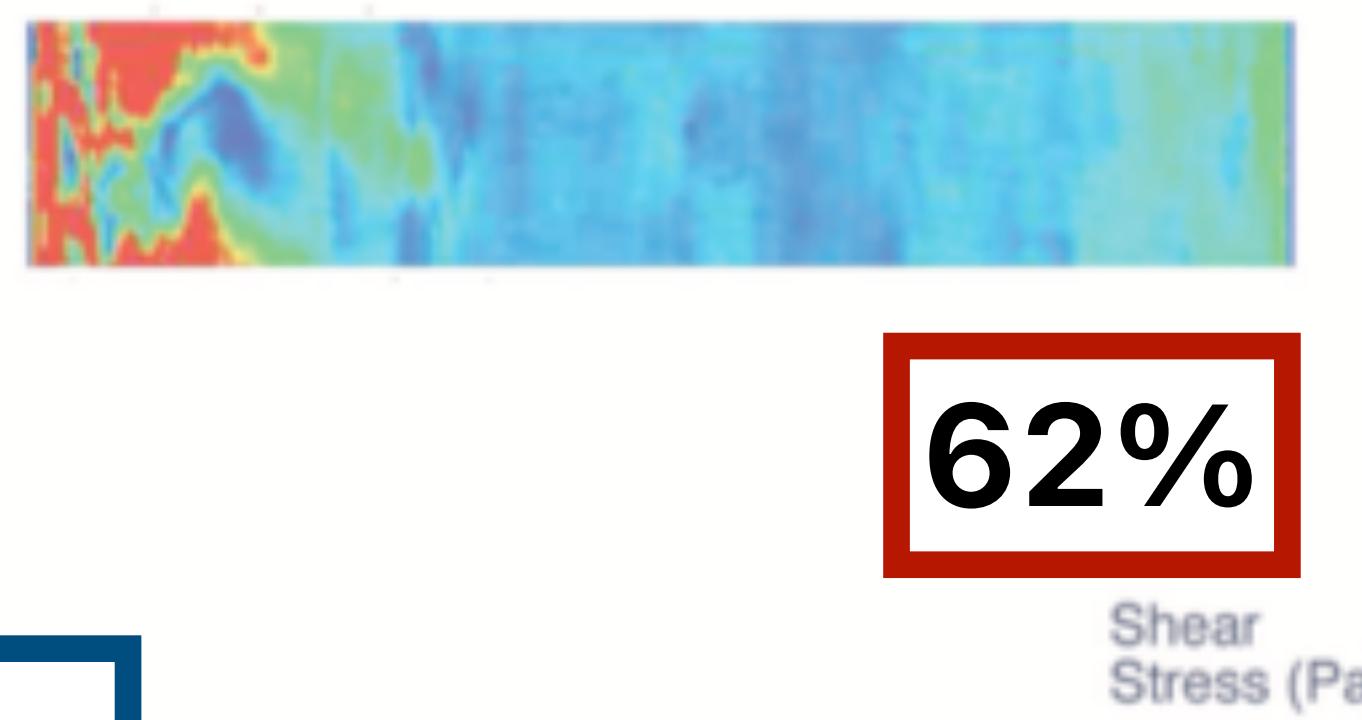
3  
2  
1  
0



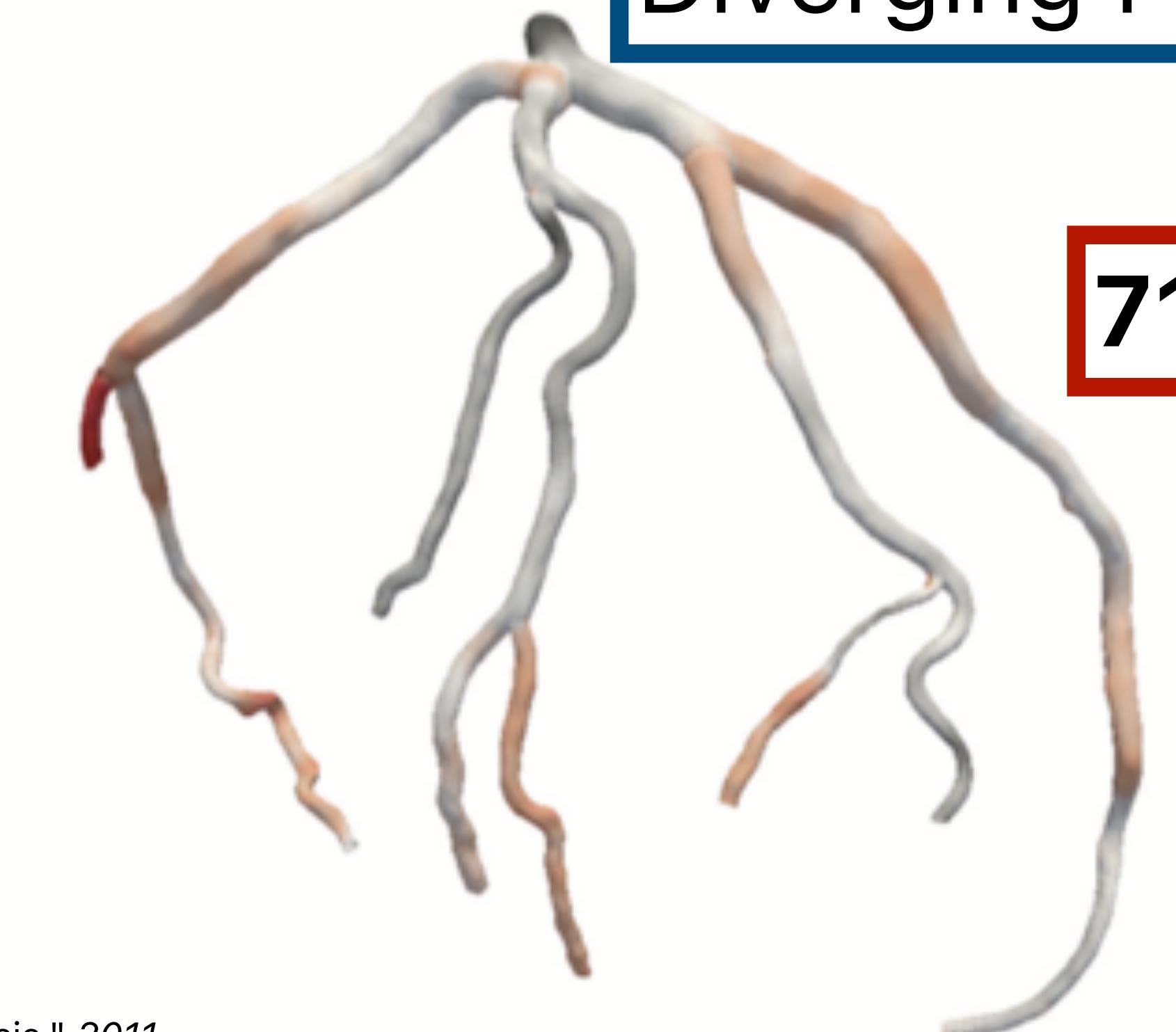
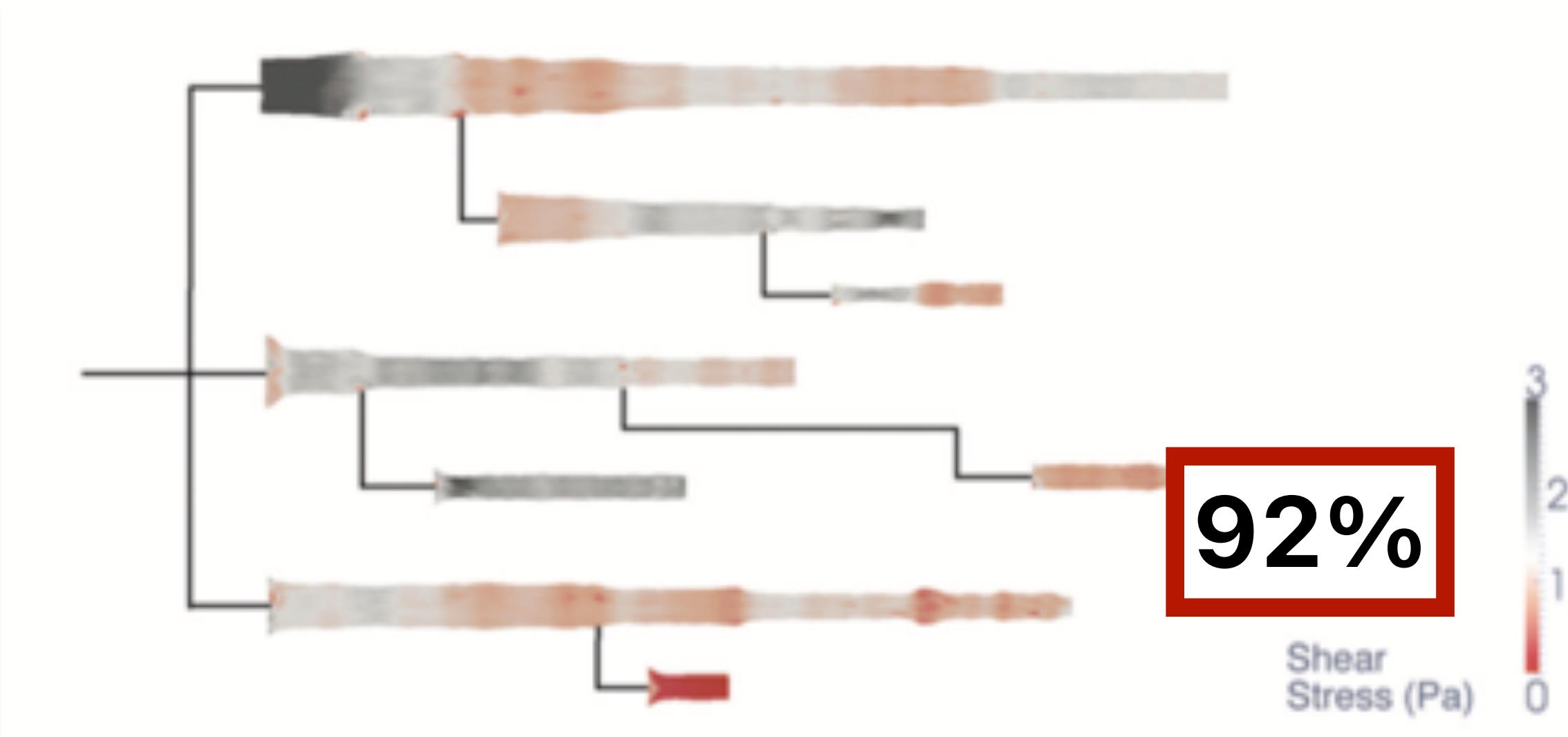
Borkin, Michelle, et al. "Evaluation of artery visualizations for heart disease diagnosis." 2011

# Artery Visualization

Rainbow Palette

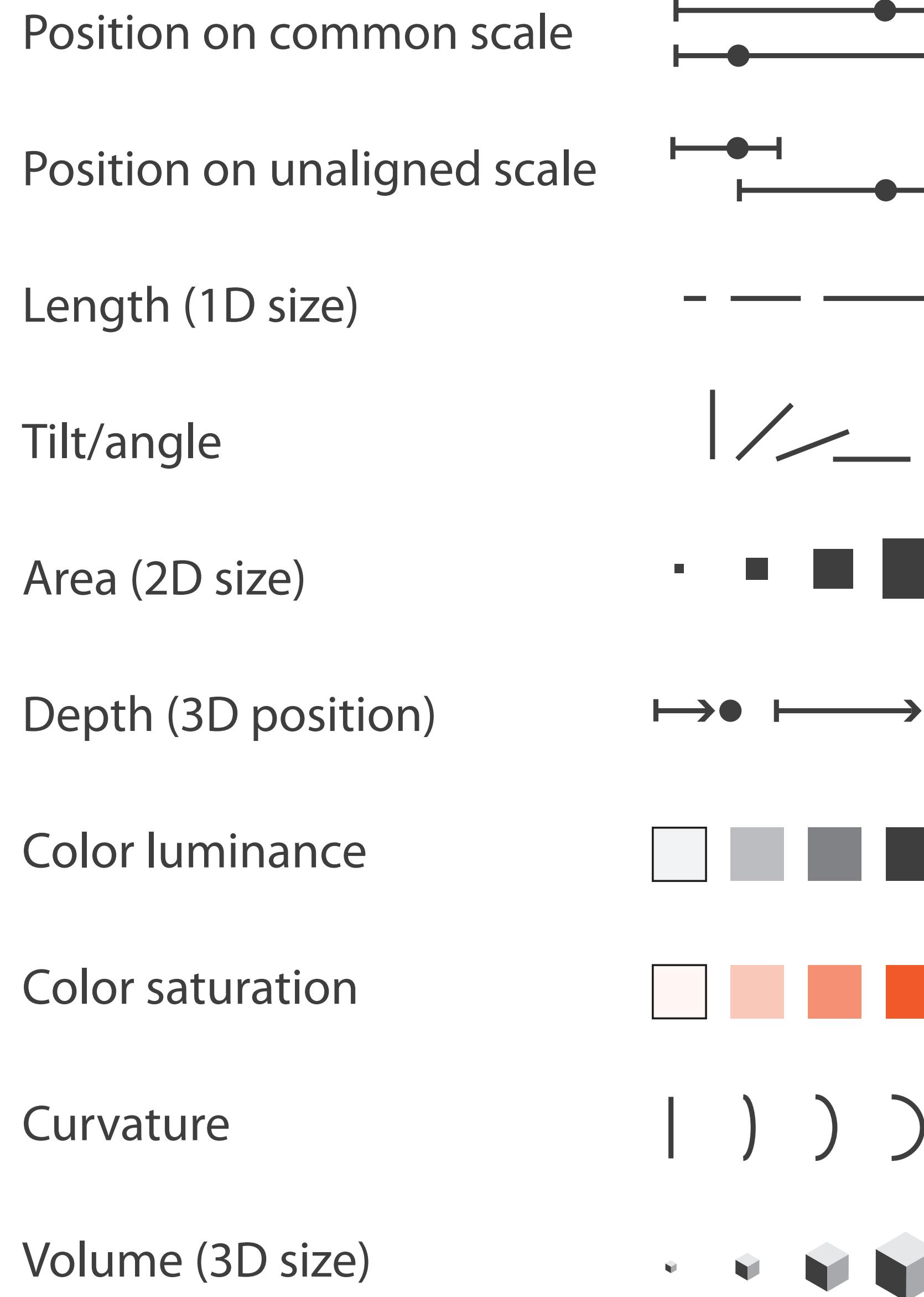


Diverging Palette



## Channels: Expressiveness Types and Effectiveness Ranks

### → Magnitude Channels: Ordered Attributes



### → Identity Channels: Categorical Attributes



Tamara Munzner, *Visualization Analysis and Design* (2014).

# Using space (in)effectively

## (De-)Obfuscating data

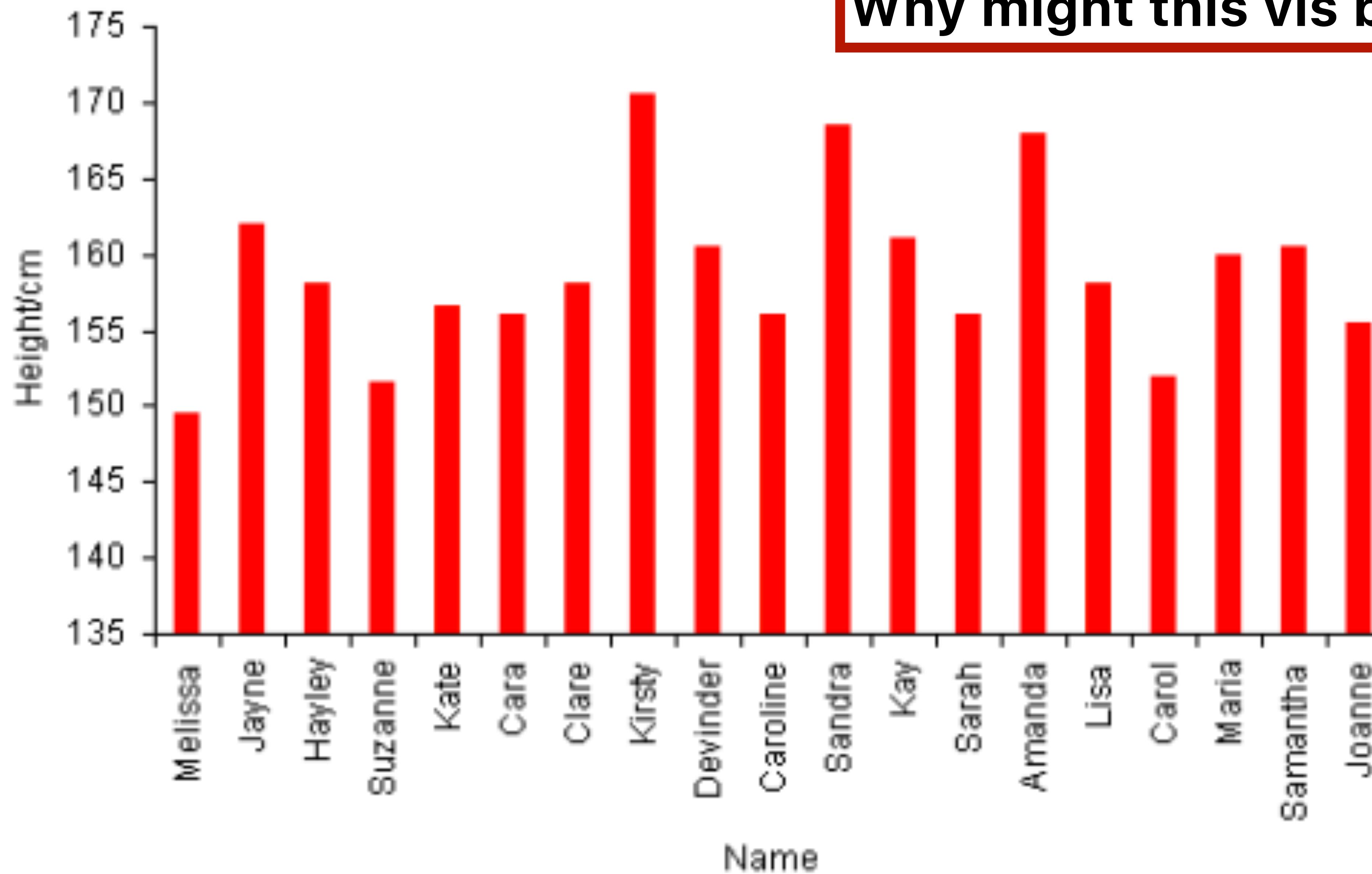
## (Mis)leading the witness

# Using space (in)effectively

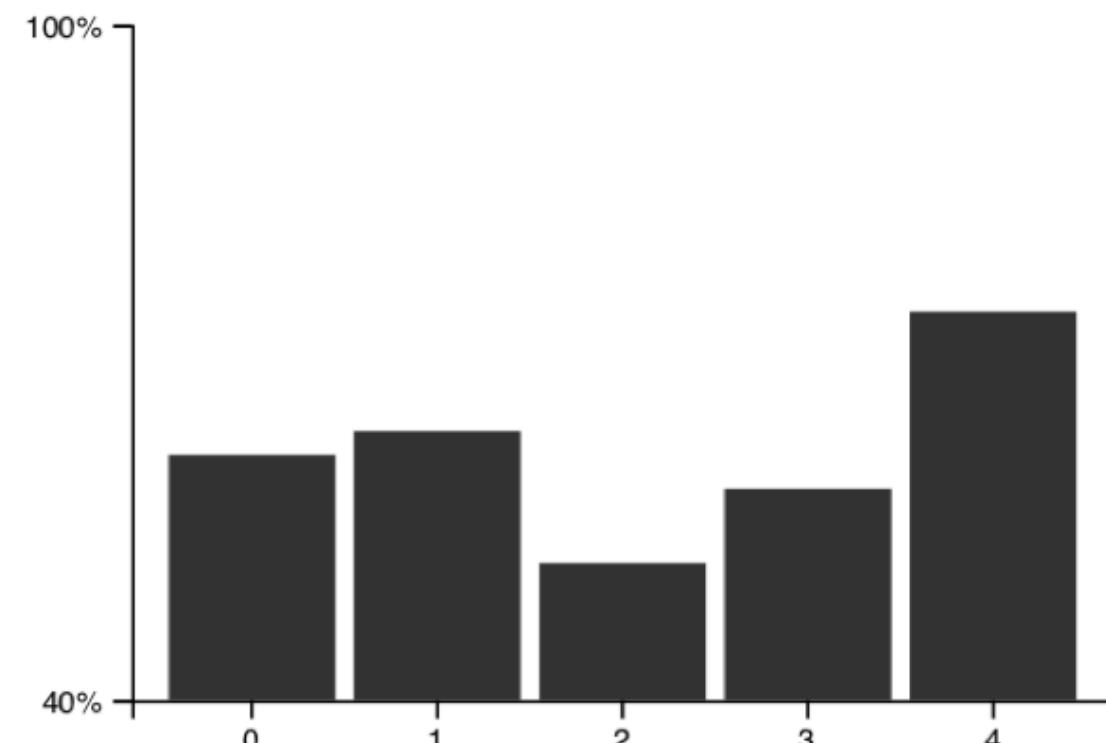
(De-)Obfuscating data

(Mis)leading the witness

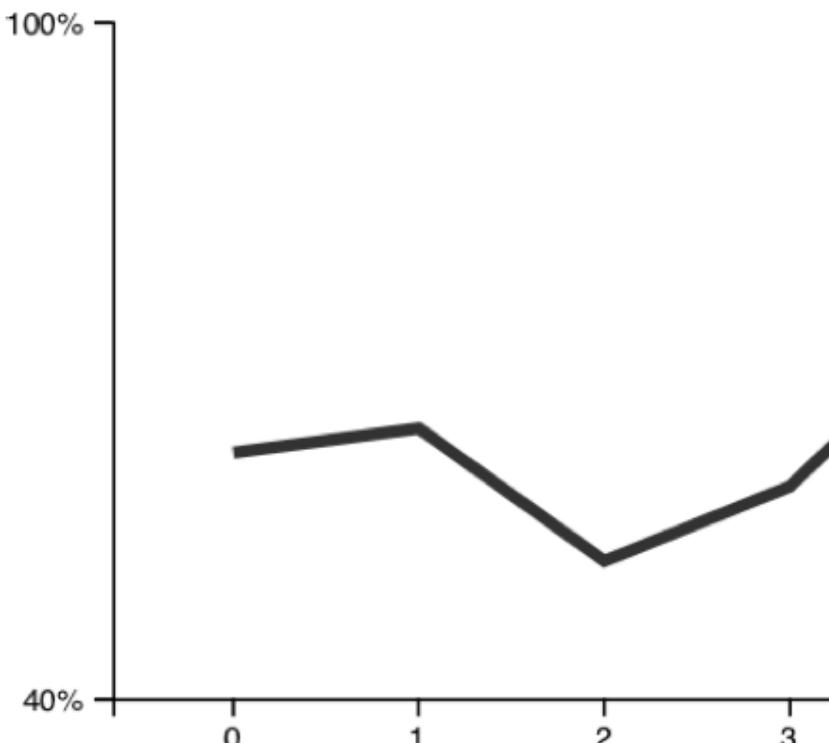
## Individual heights



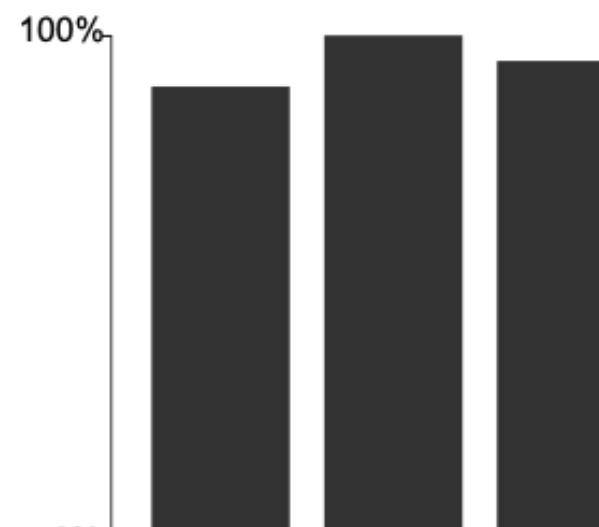
Why might this vis be inexpressive?



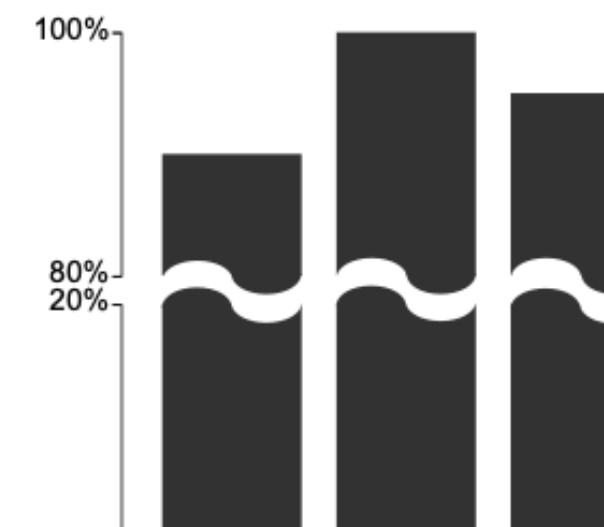
(a) Bar Chart



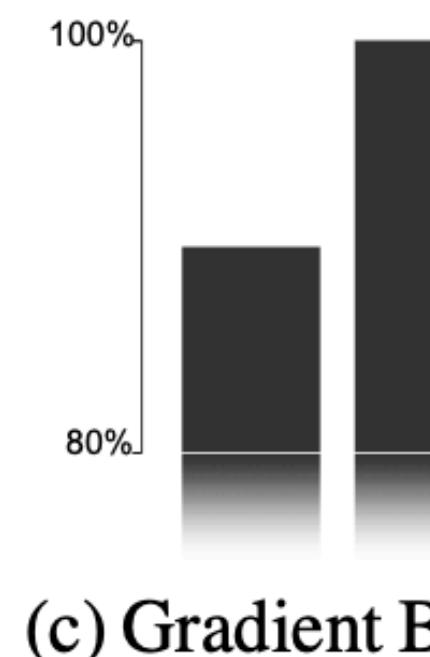
(b) Line Chart



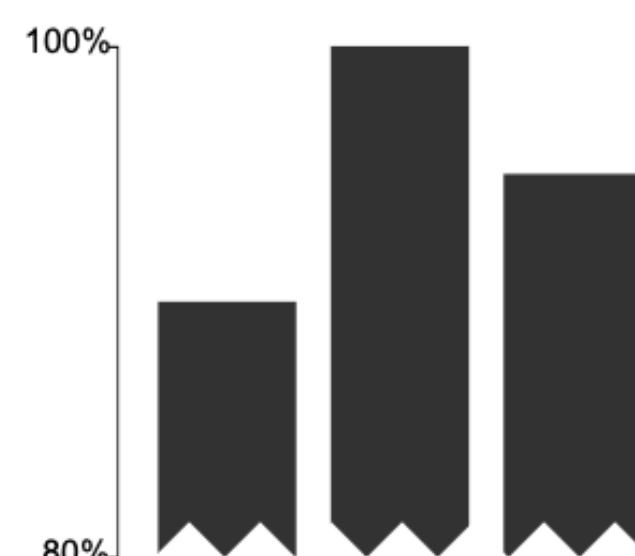
(a) Bar Chart



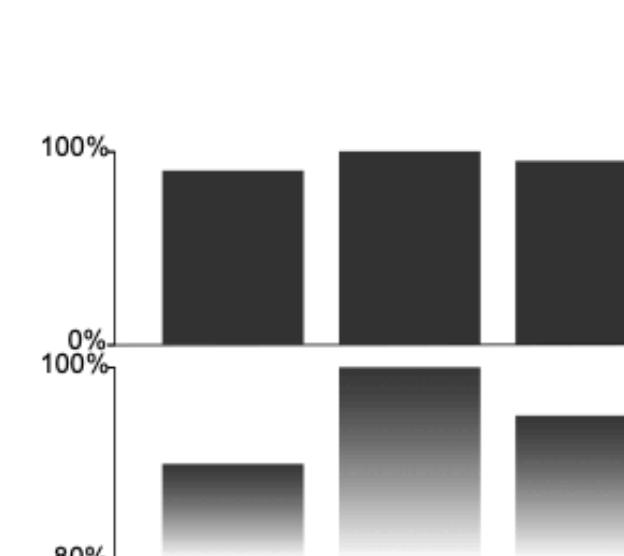
(b) Broken Axes



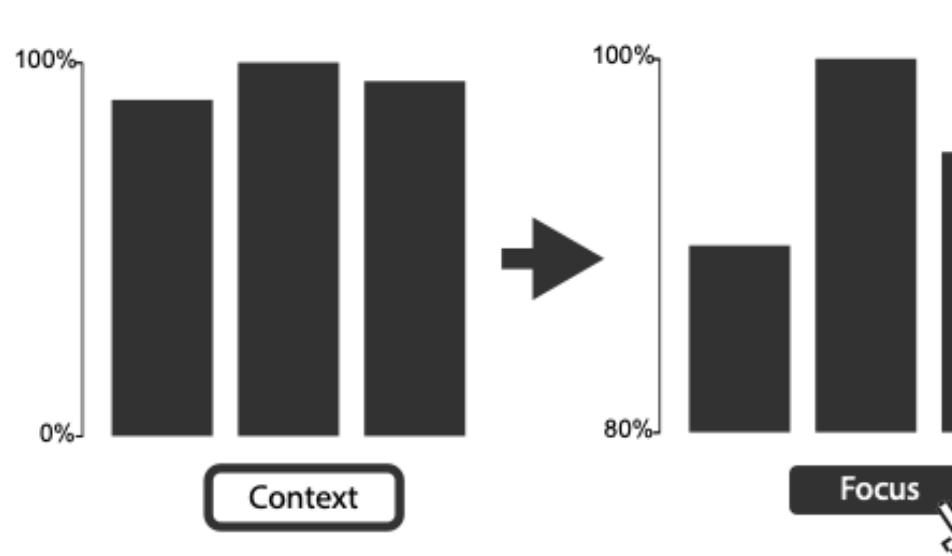
(c) Gradient Bar



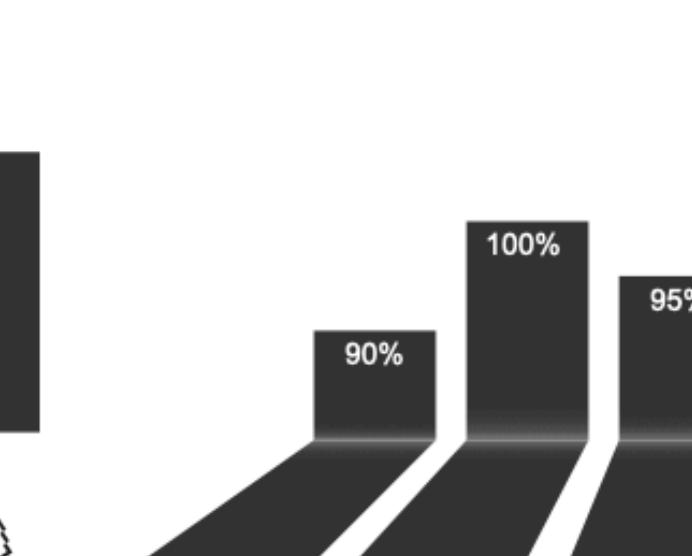
(d) Torn Paper Chart



(e) Panel Chart



(f) Interactive Focus+Context

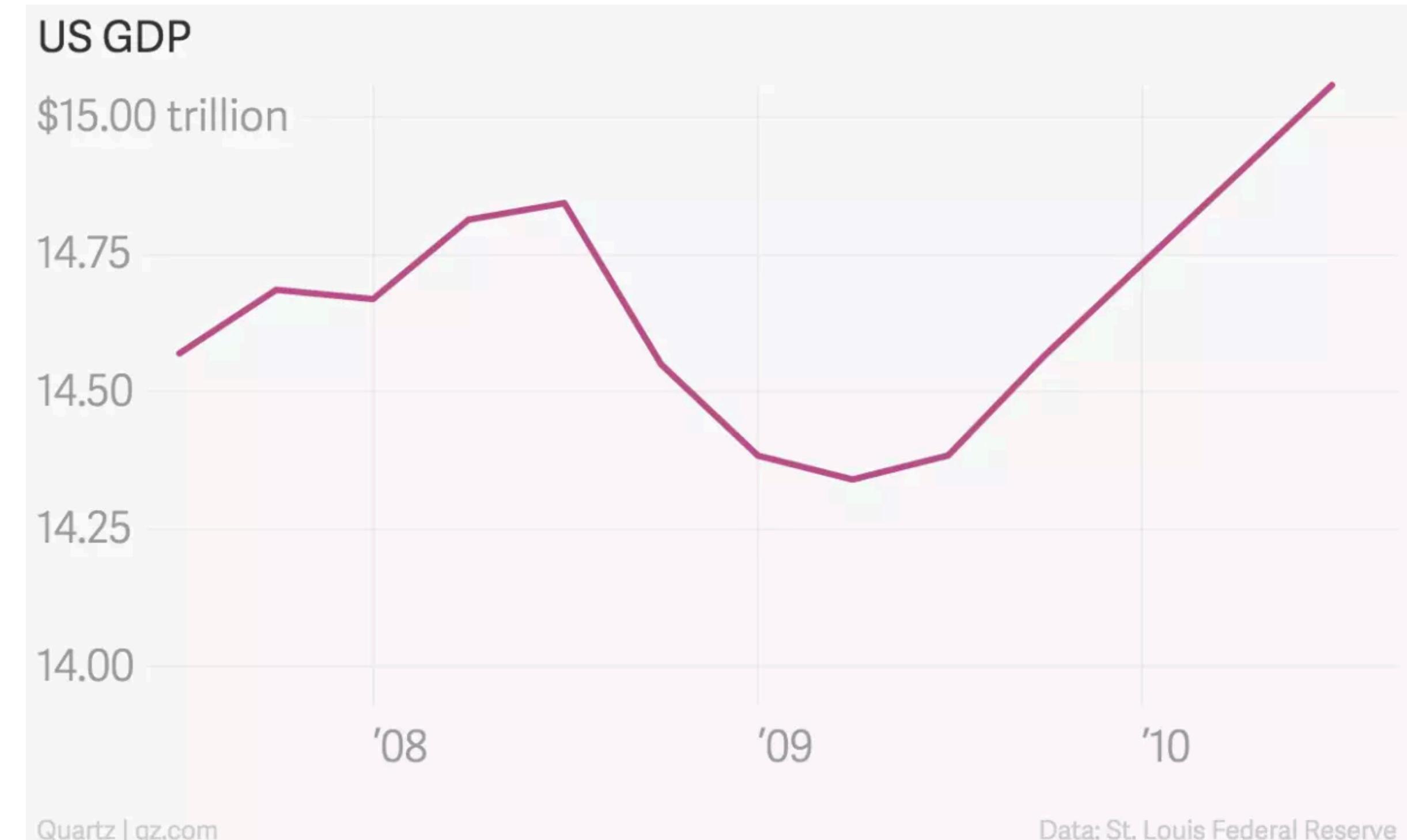
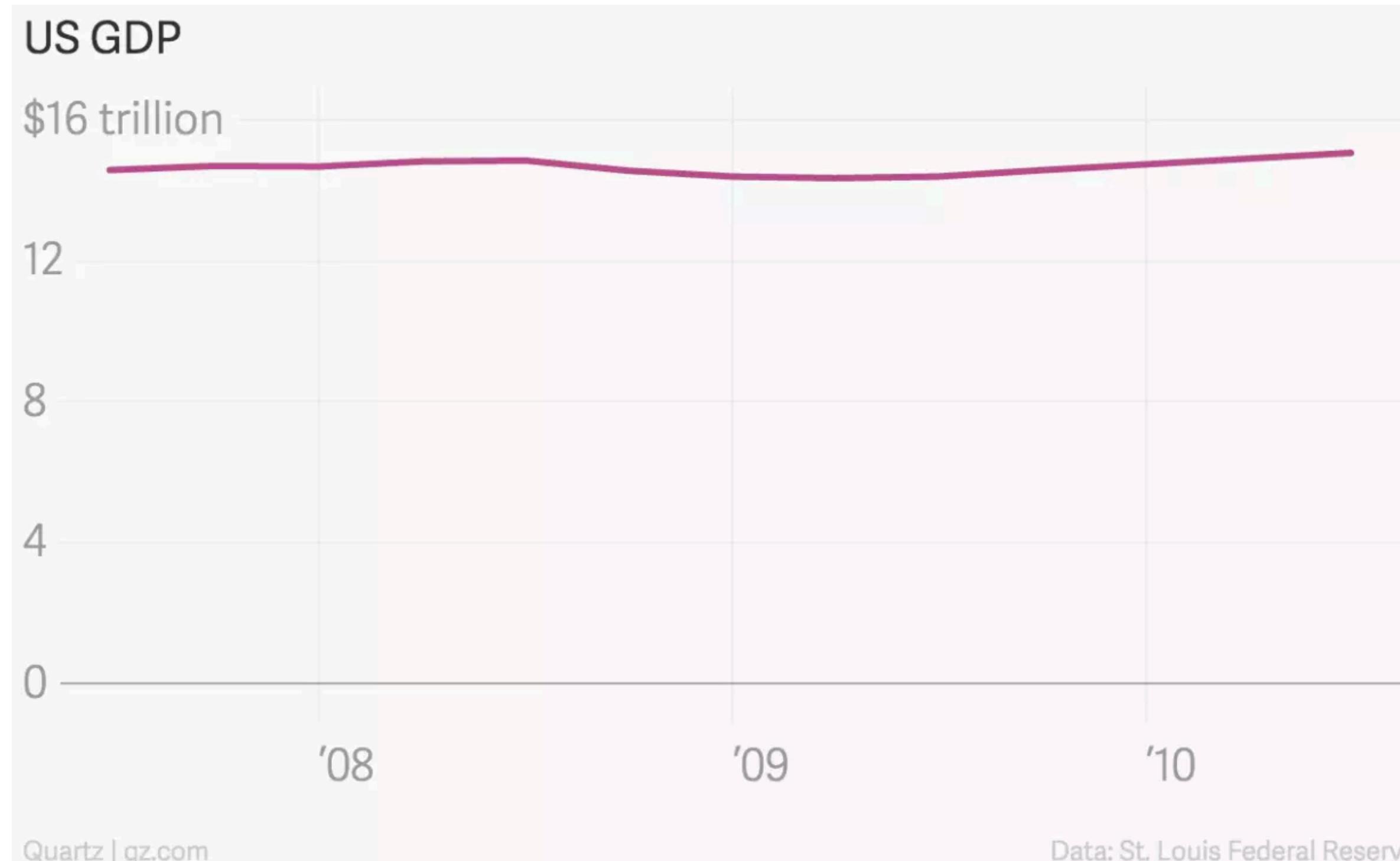


(g) Bent Bar Chart

Y-axis truncation has a consistent and significant impact on perceived effect size for both line and bar charts.

Interventions did not make a difference.

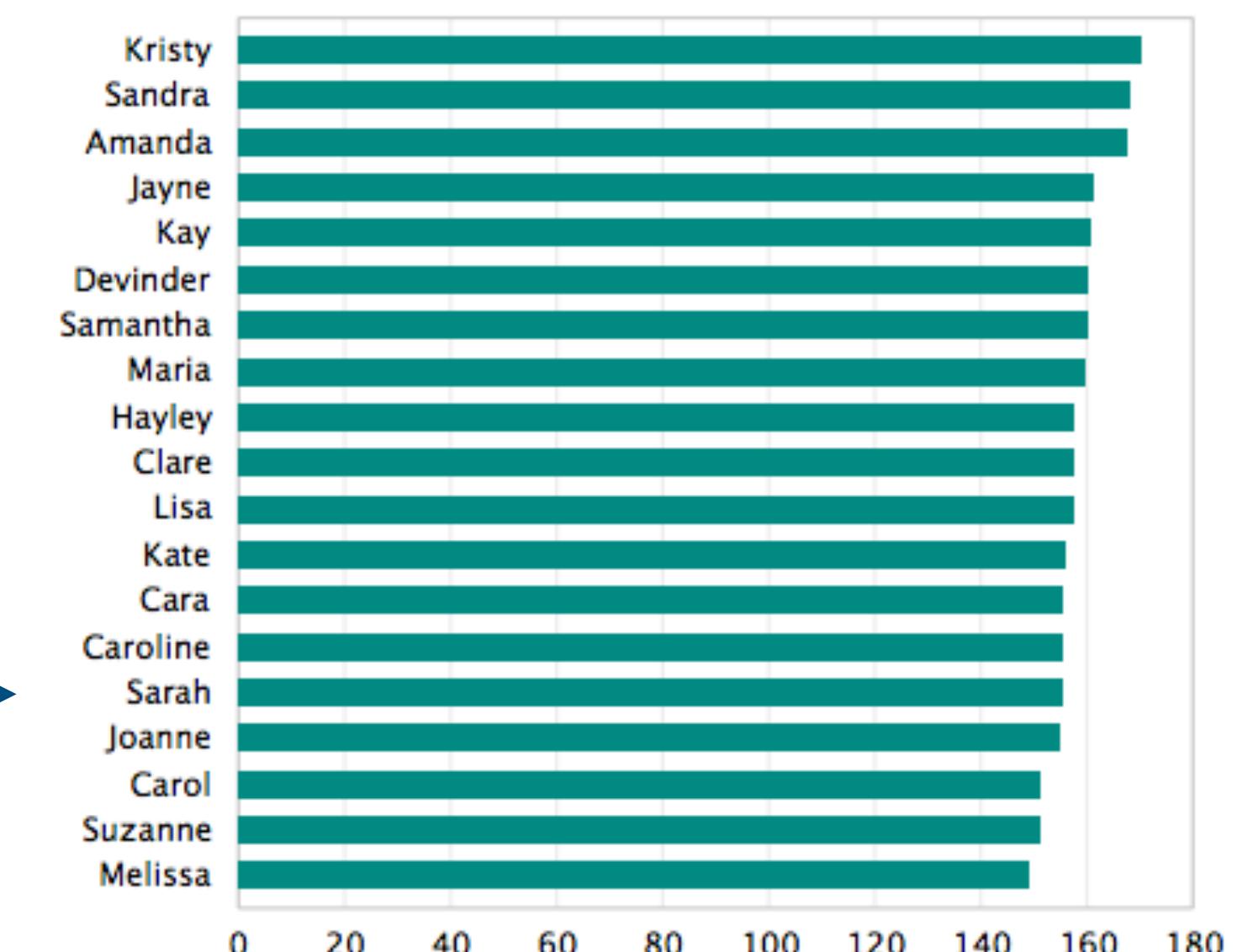
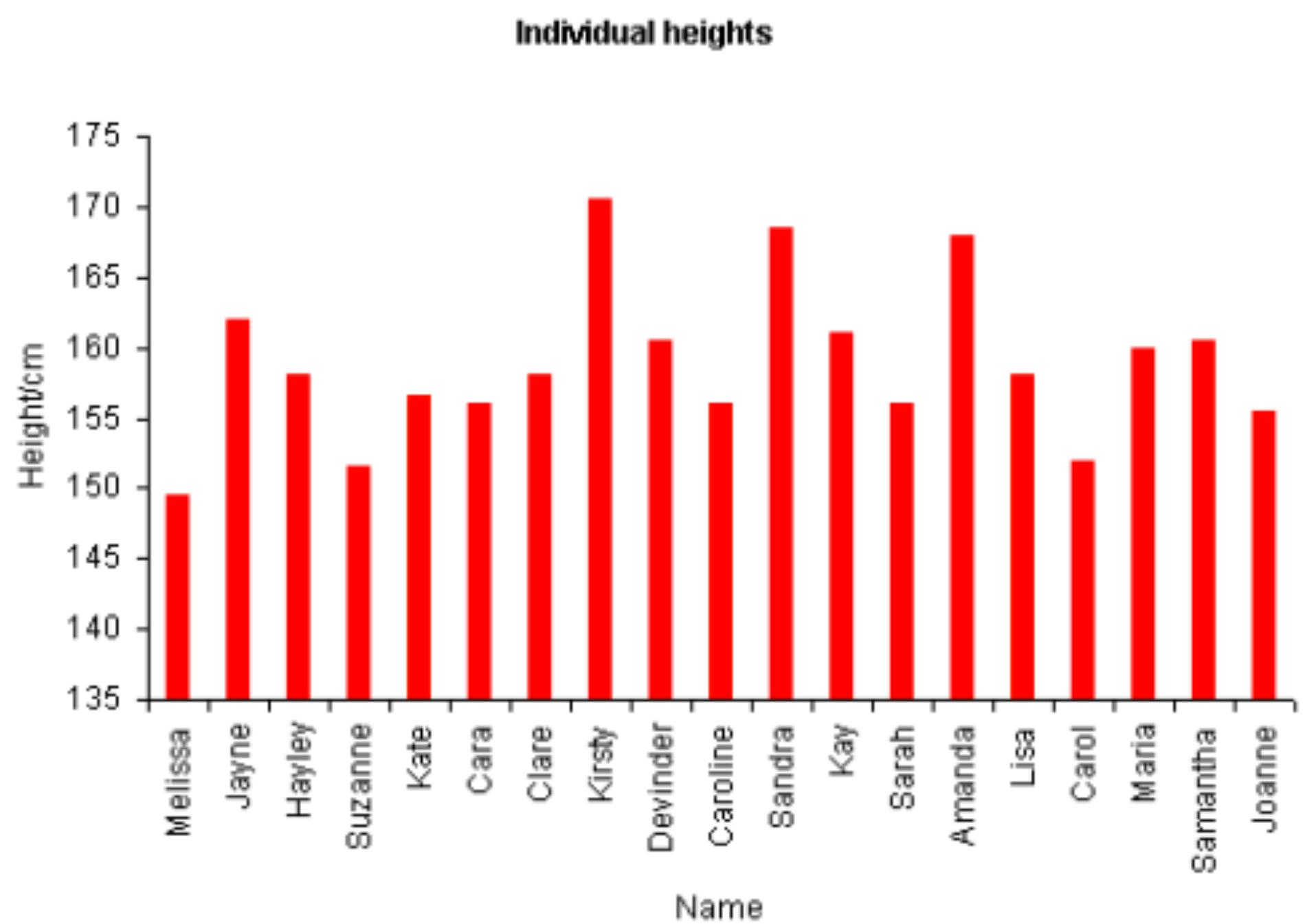
# Always start at zero?



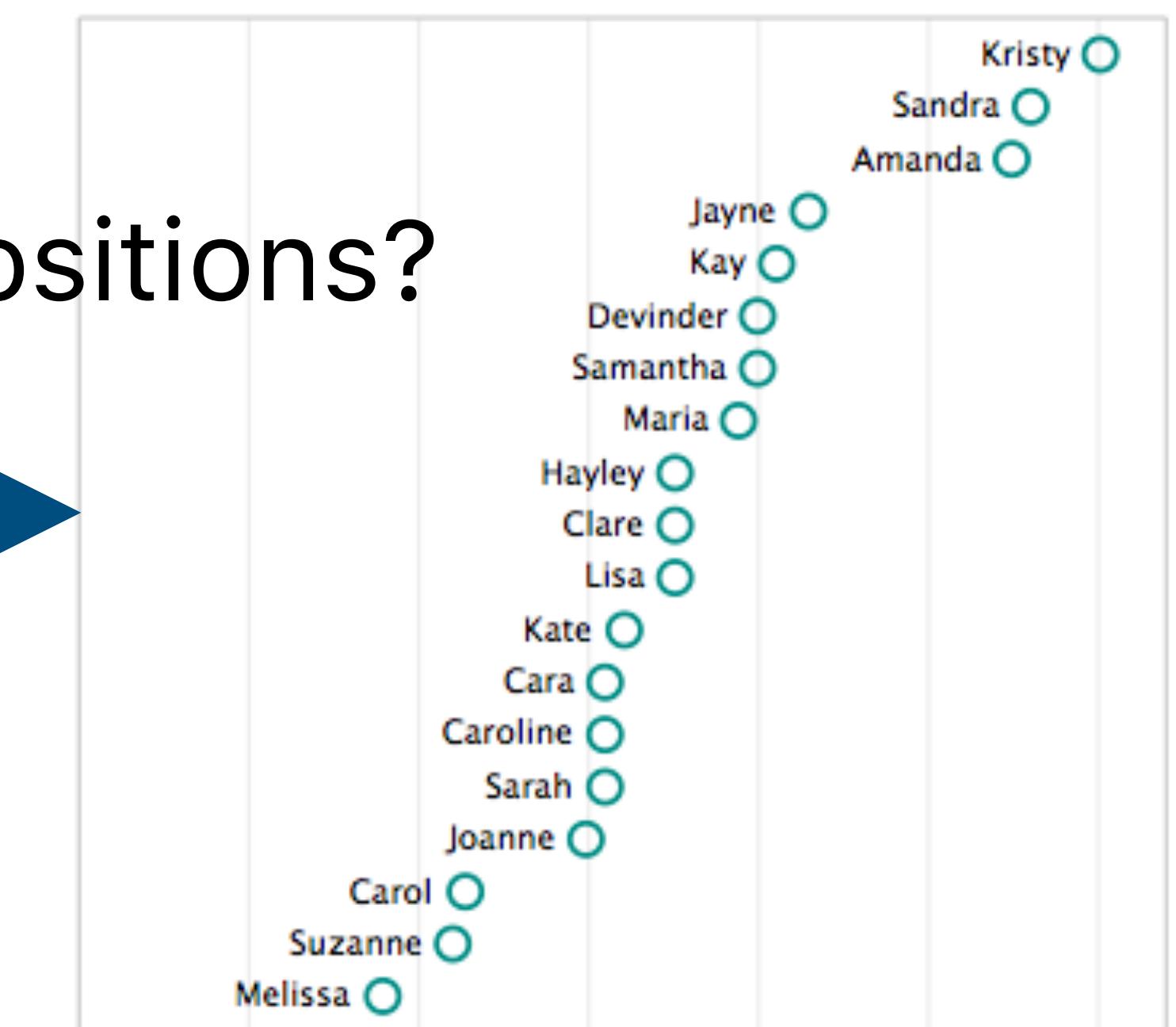
<https://qz.com/418083/its-ok-not-to-start-your-y-axis-at-zero>

# Truncating the y-axis?

Compare proportions?  
(Q-ratio)



Compare relative positions?  
(Q-interval)



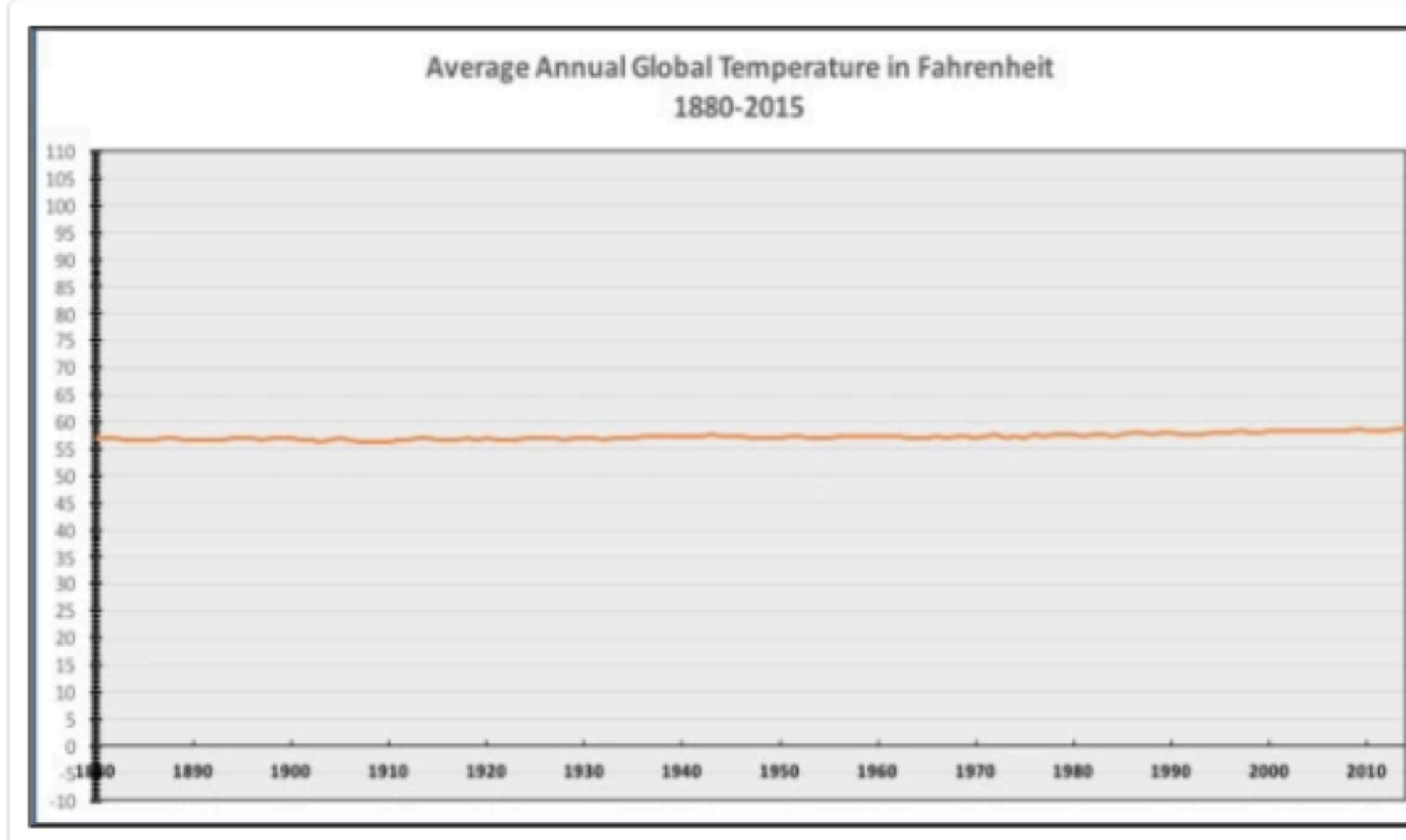
# Truncating the y-axis?

To emphasize Q-interval (vs. Q-ratio)  
If the zero value doesn't make much sense.  
If it is the norm (e.g., stock charts).



The only **#climatechange** chart you need to see. [natl.re/wPKpro](http://natl.re/wPKpro)

(h/t [@powerlineUS](#))

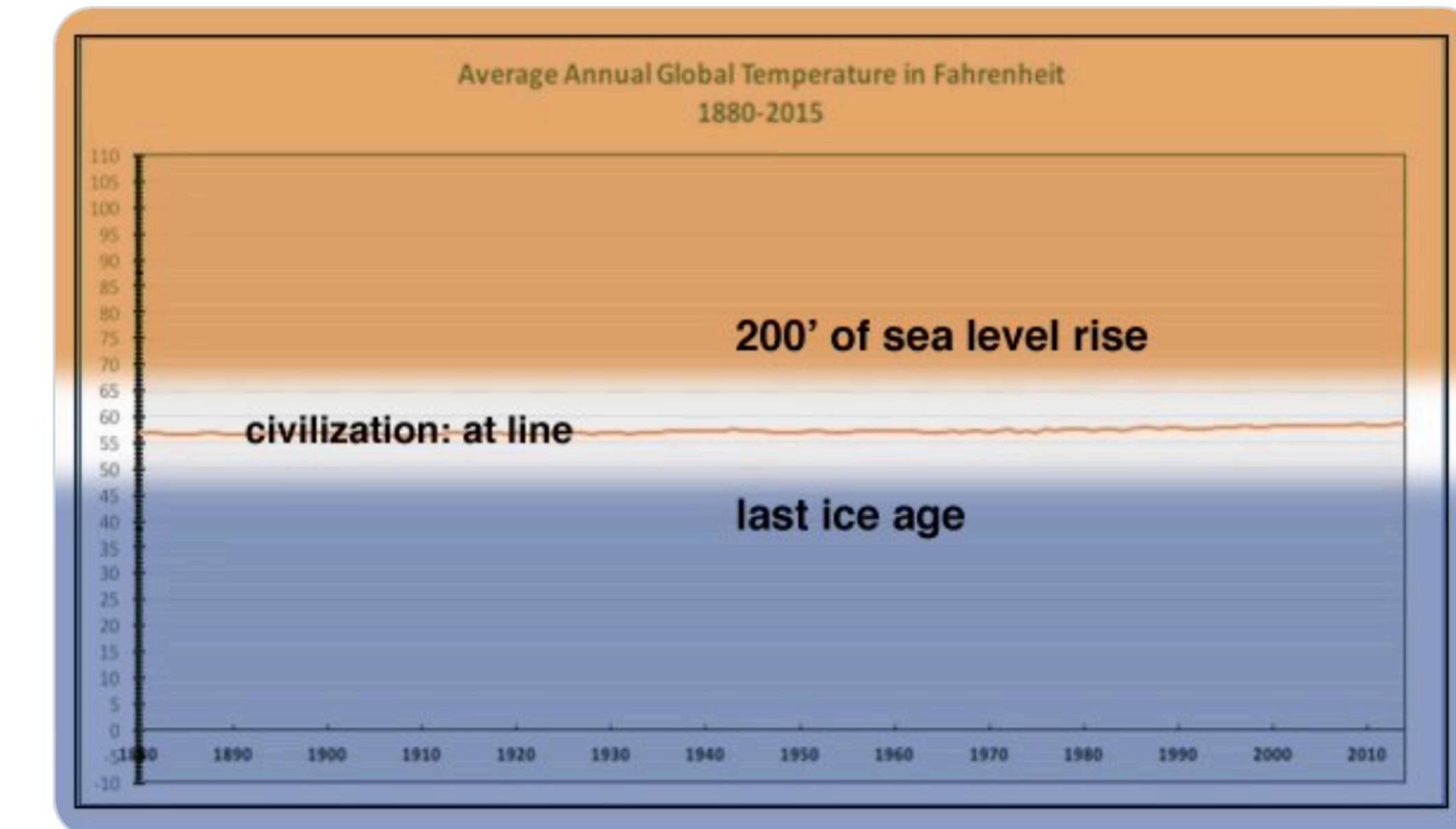


12:36 PM - 14 Dec 2015



Replying to [@NRO](#)

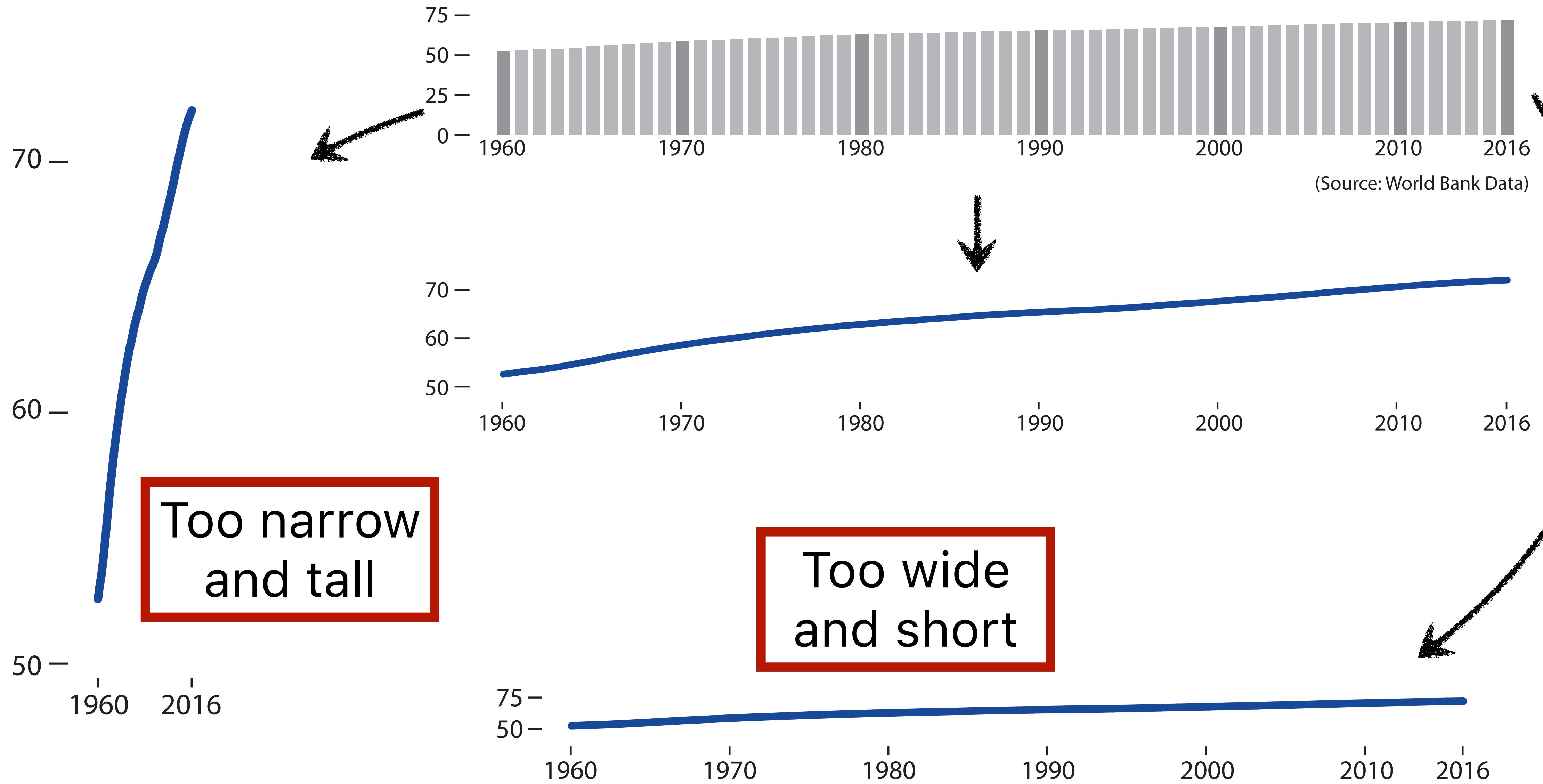
[@NRO](#) [@powerlineUS](#) [@bradplumer](#) I'm sure someone else has fixed this for you, but here you go. Great idea, thx --



5:28 PM · Dec 14, 2015

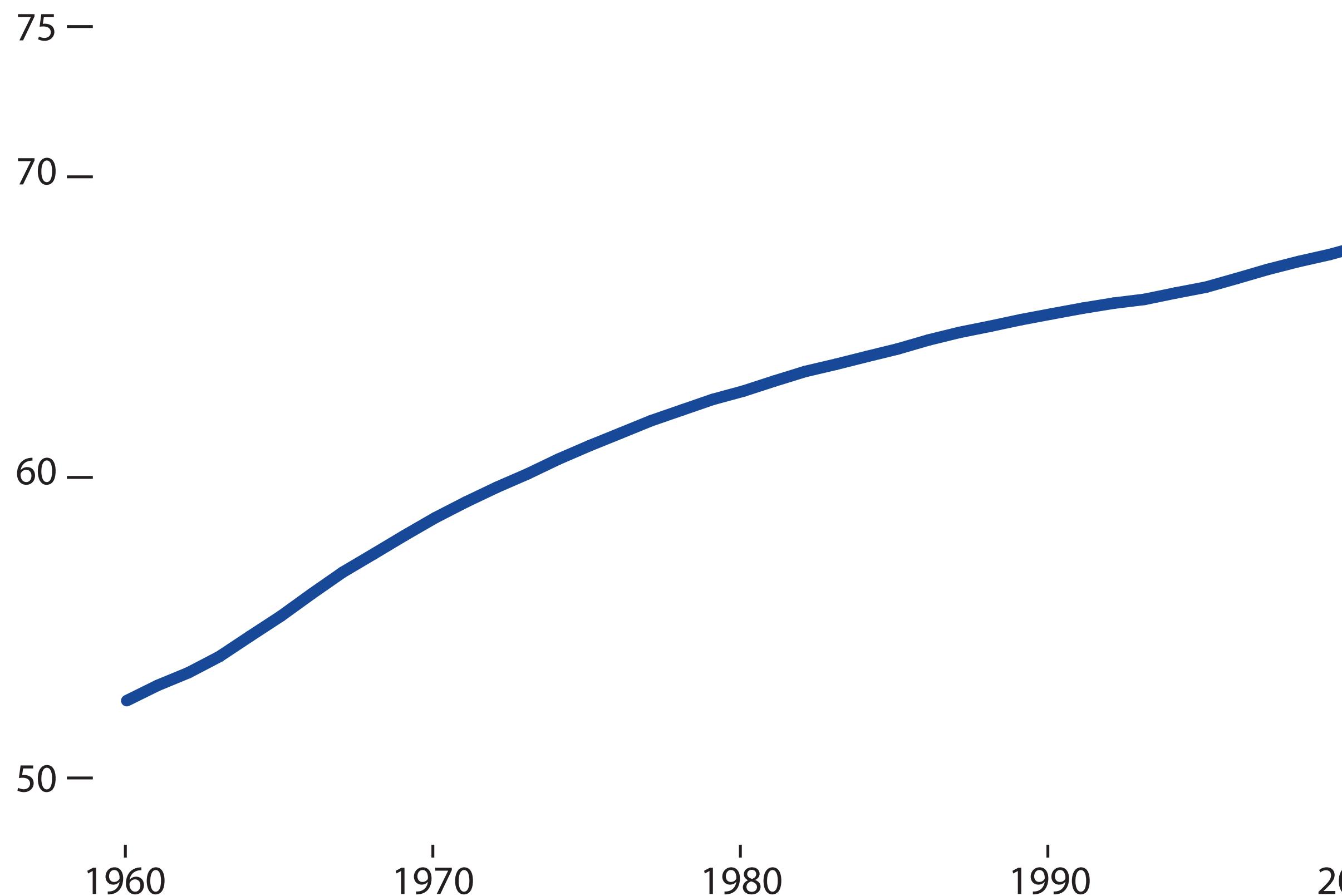
78 Retweets 1 Quote Tweet 208 Likes

## Average world life expectancy at birth (years)



# Aspect Ratio

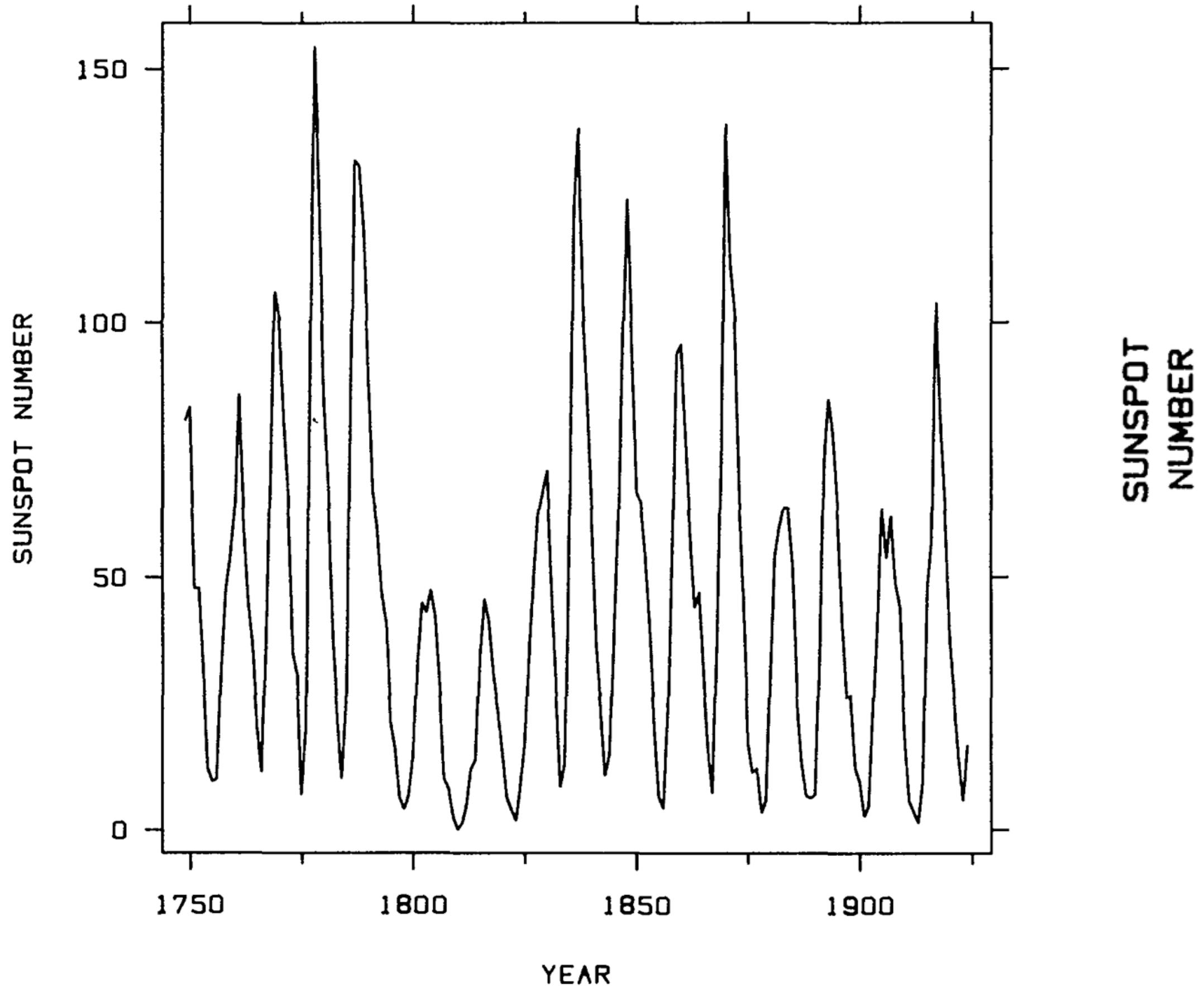
Average world life expectancy at birth (years)



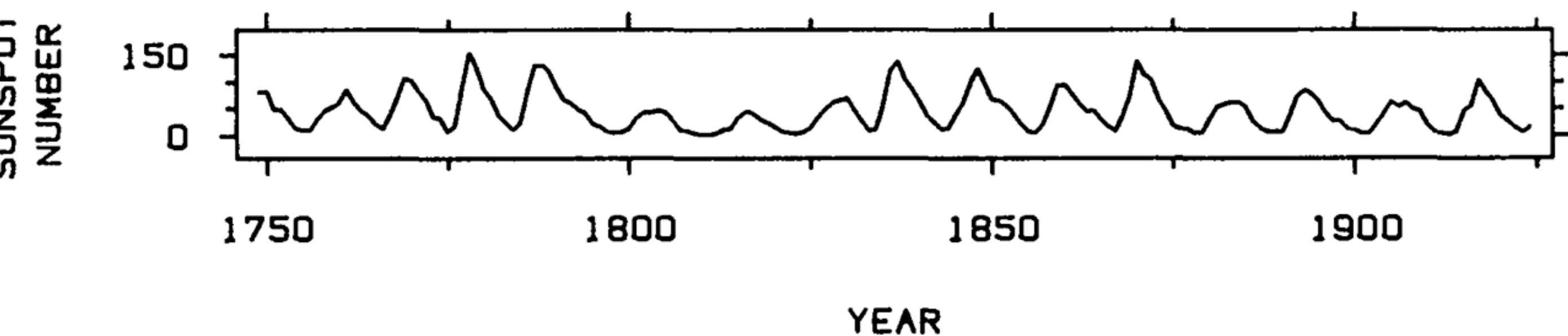
Approximate the proportion of the chart to match the depicted trend.

35% increase  $\approx 1/3$ rd  
 $\approx 4:3$  aspect ratio

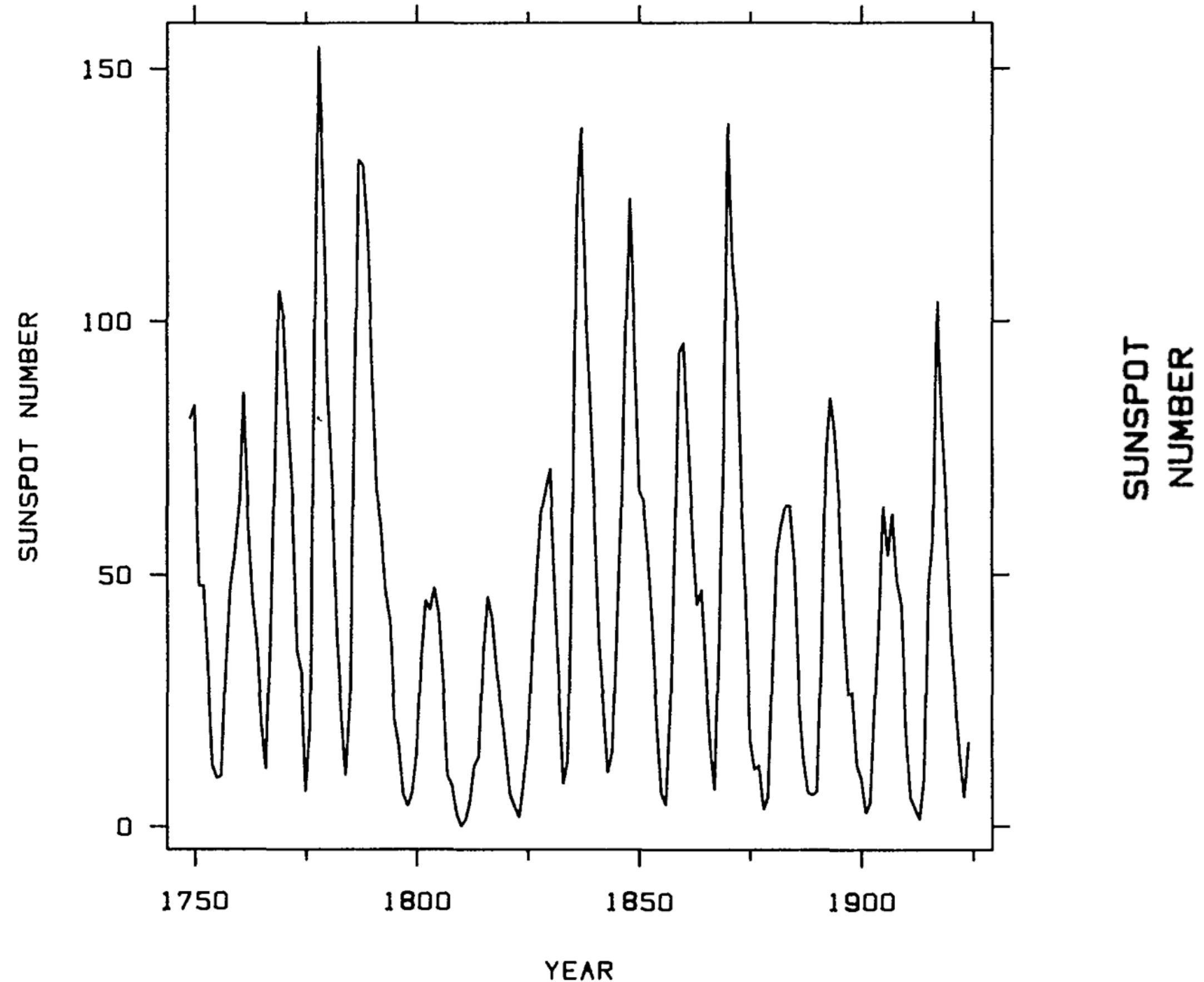
# Aspect Ratio



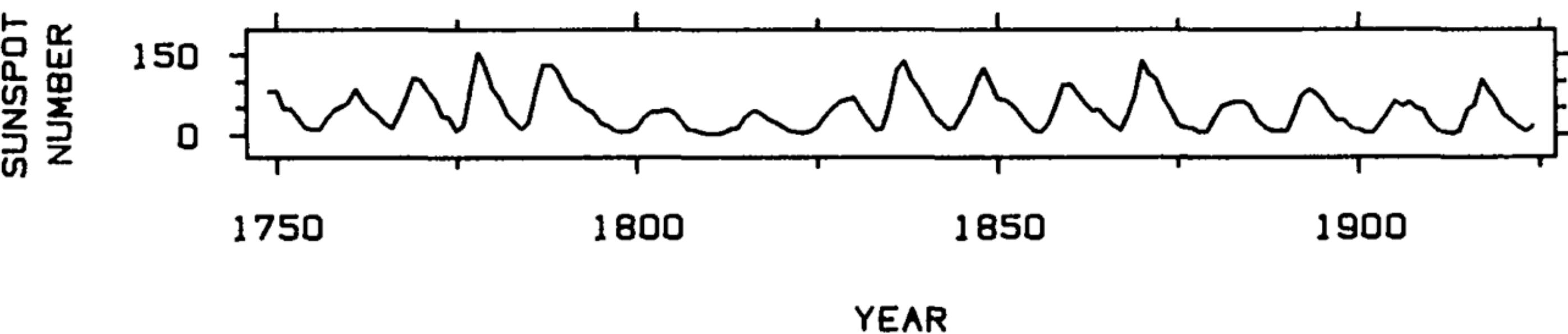
1. Approximate the proportion of the chart to match the depicted trend.



# Aspect Ratio



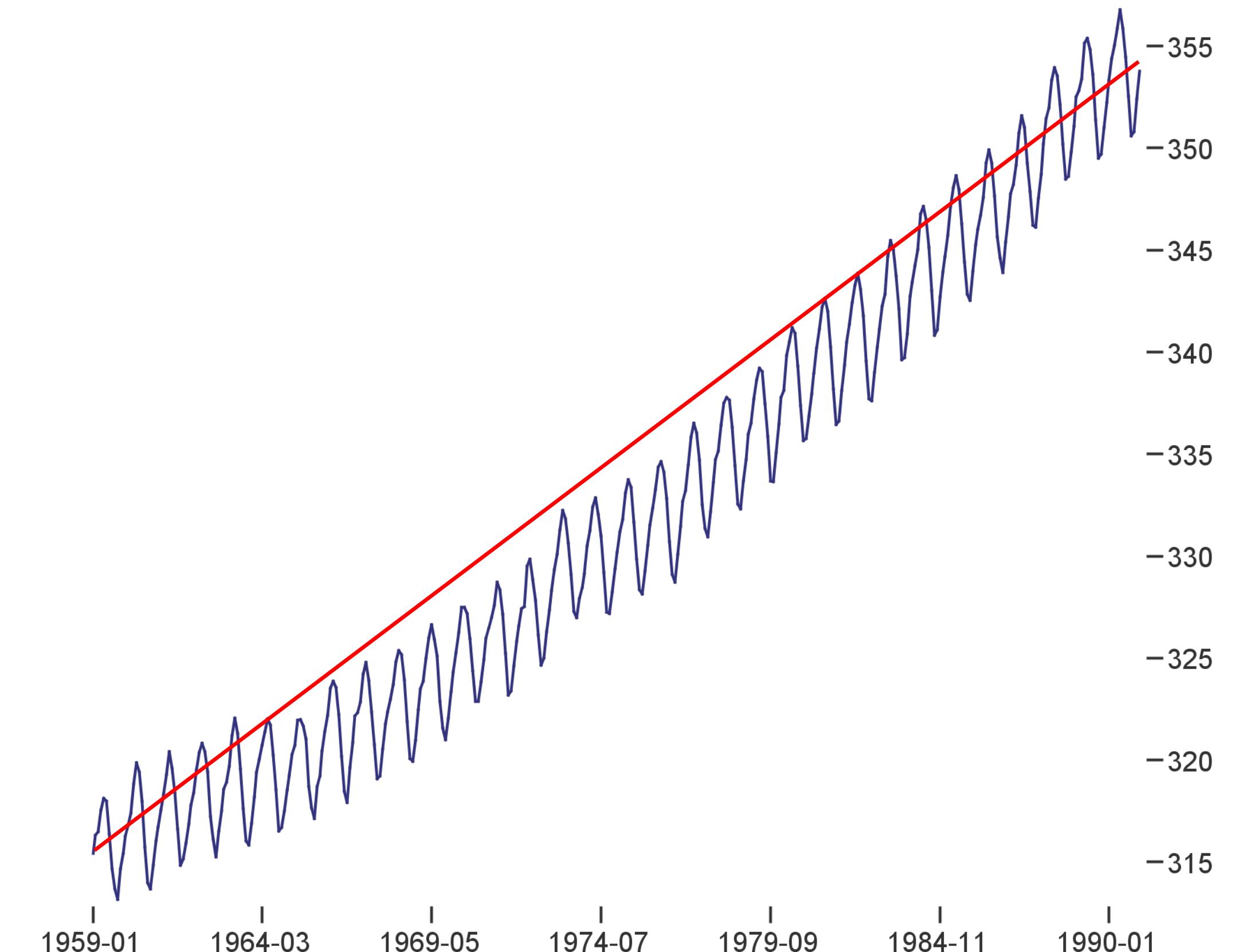
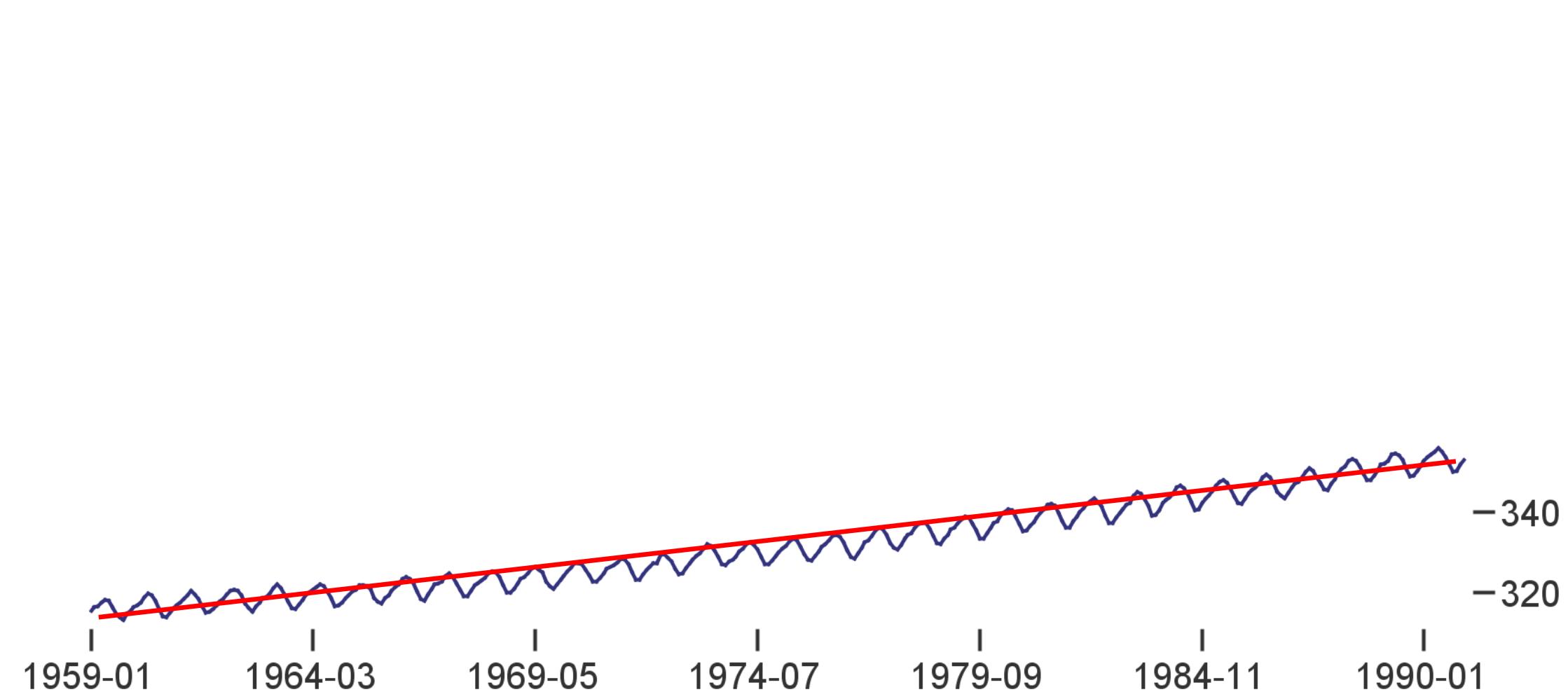
1. Approximate the proportion of the chart to match the depicted trend.



2. Bank to  $45^\circ$ : aspect ratios with  $45^\circ$  avg. line segment orientation.

# Aspect Ratio

2. Bank to  $45^\circ$ : original data or fitted lines

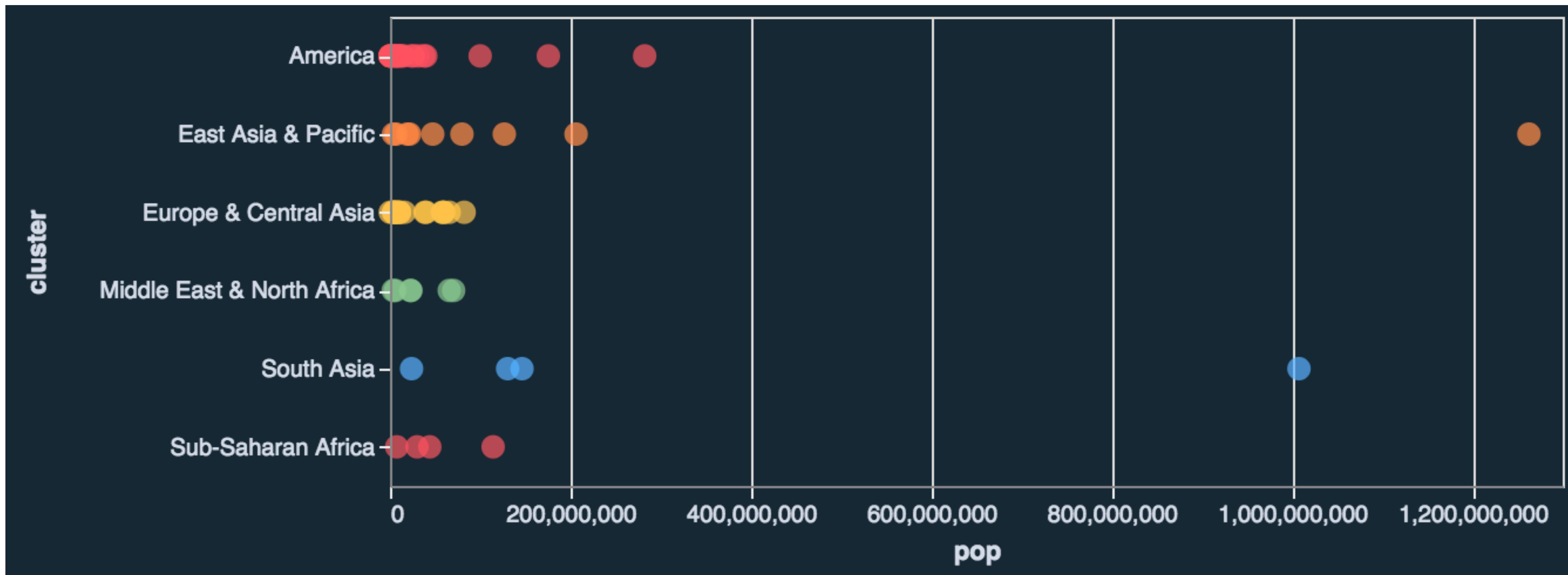


Heer, Jeffrey, and Maneesh Agrawala. "Multi-scale banking to 45 degrees." *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006): 701-708.

1. Clip them out.

# Scaling Axes: Outliers and Skew

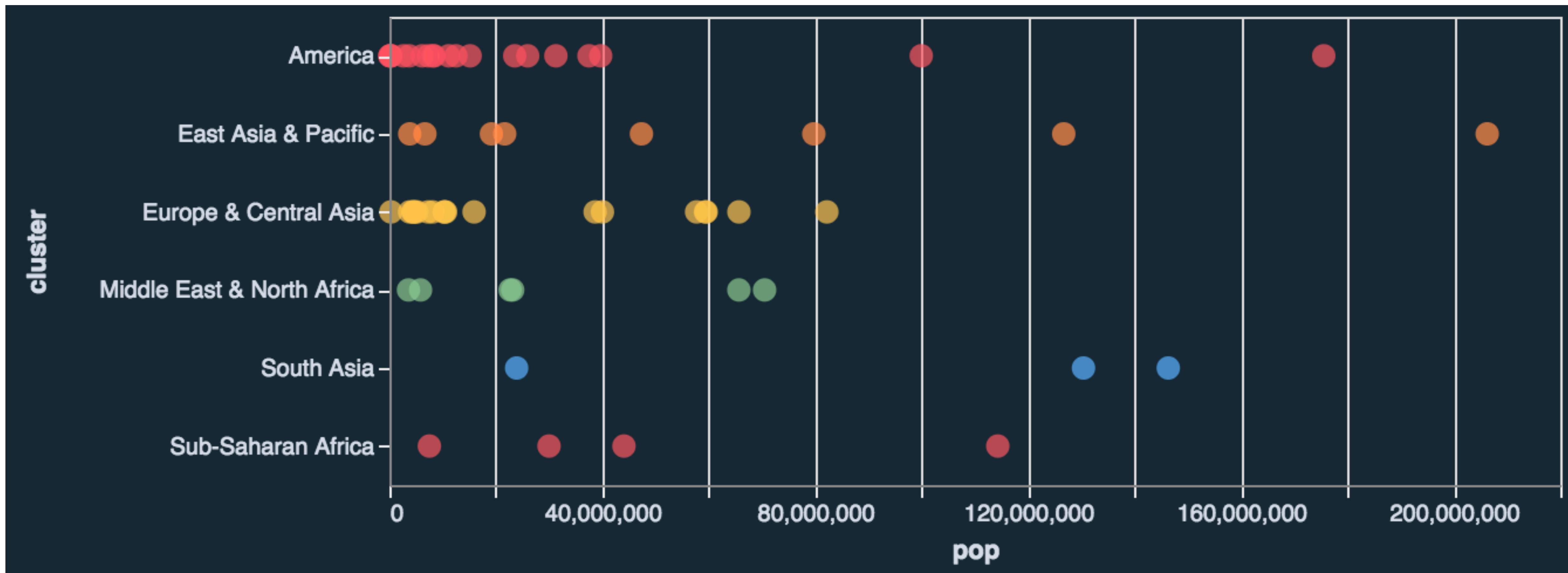
Options:



1. Clip them out.

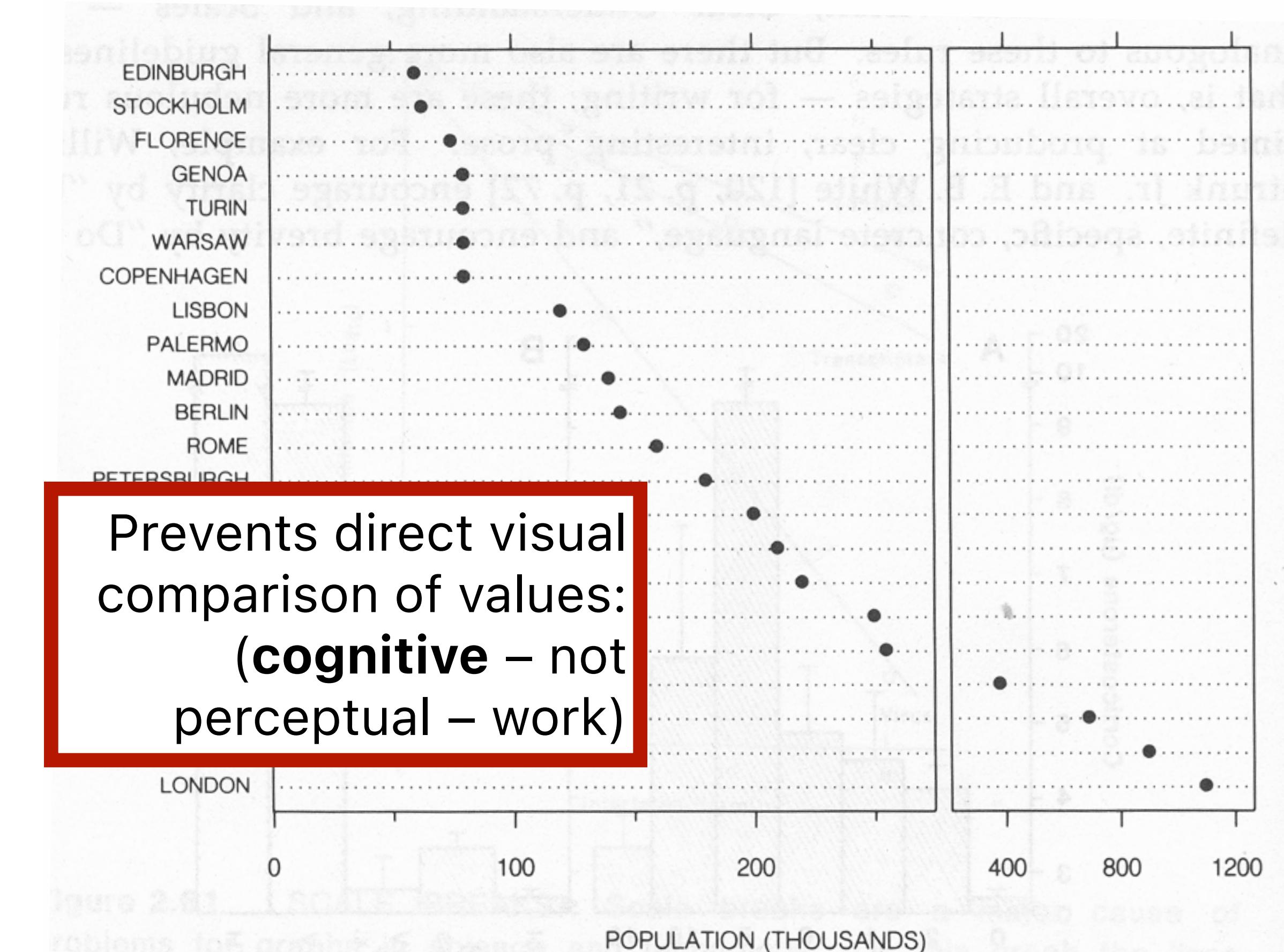
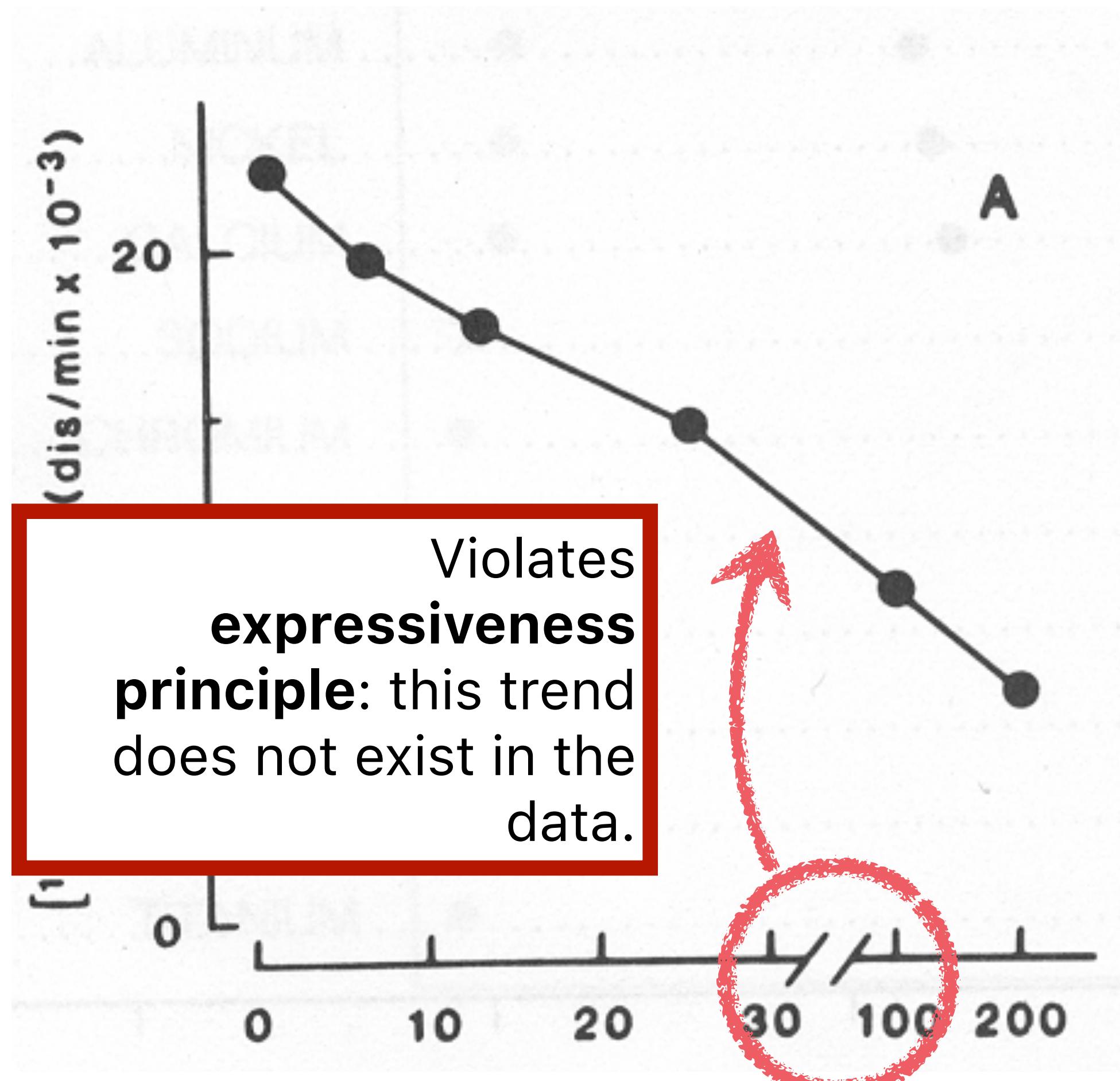
# Scaling Axes: Outliers and Skew

Options:

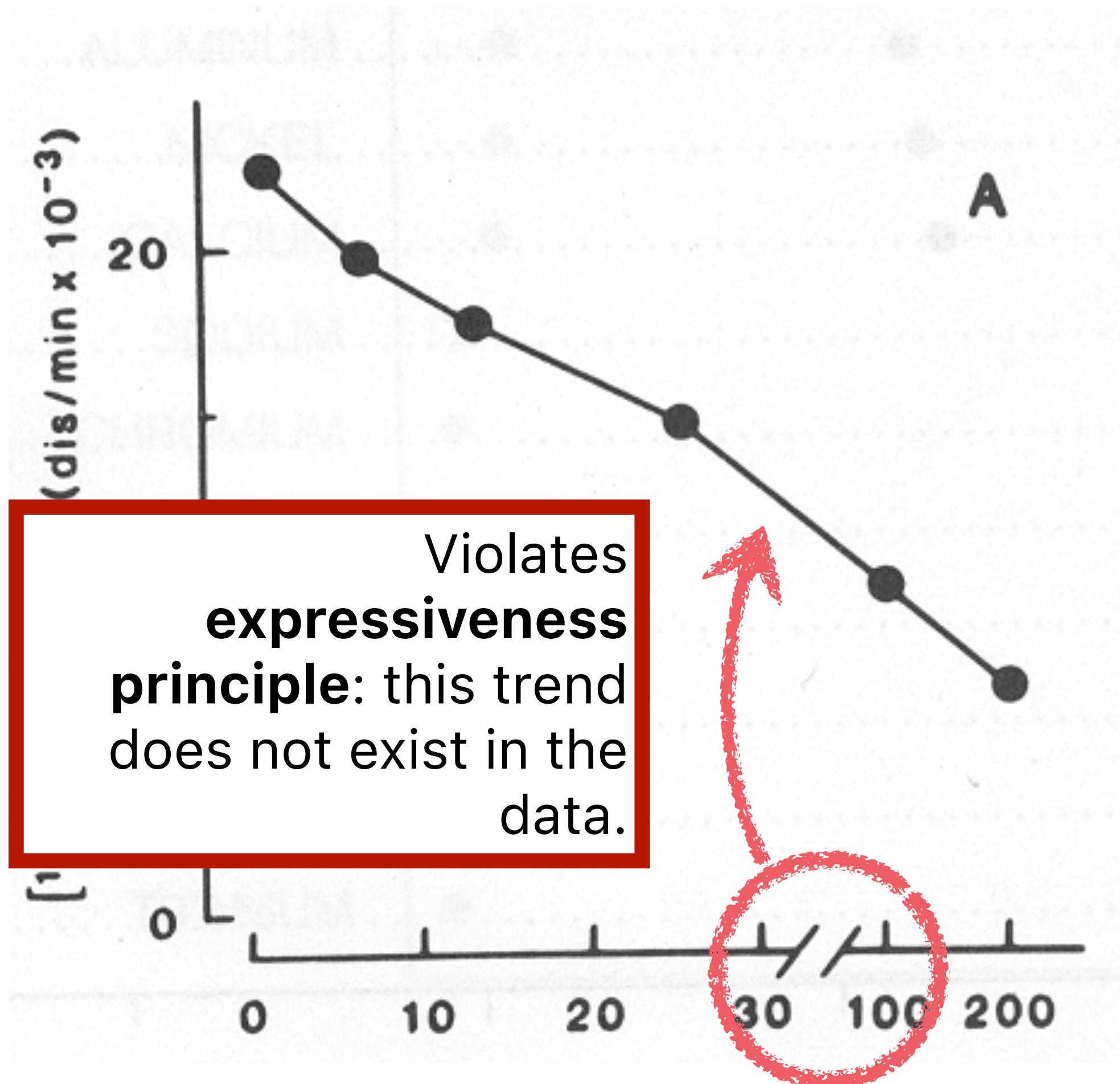


# Scaling Axes: Outliers and Skew

- Options:
1. Clip them out.
  2. Scale breaks

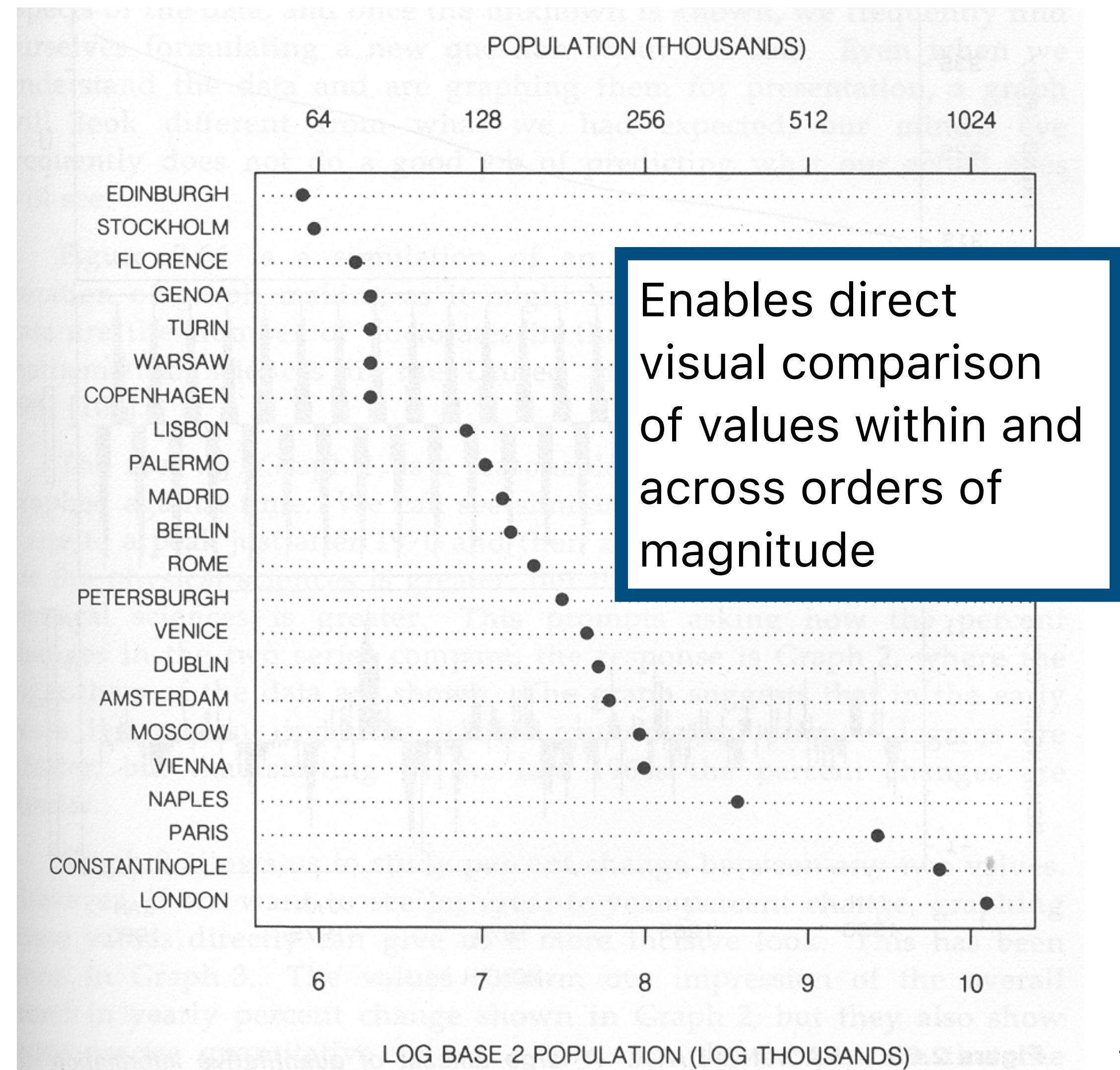


# Scaling Axes: Outliers and Skew



Options:

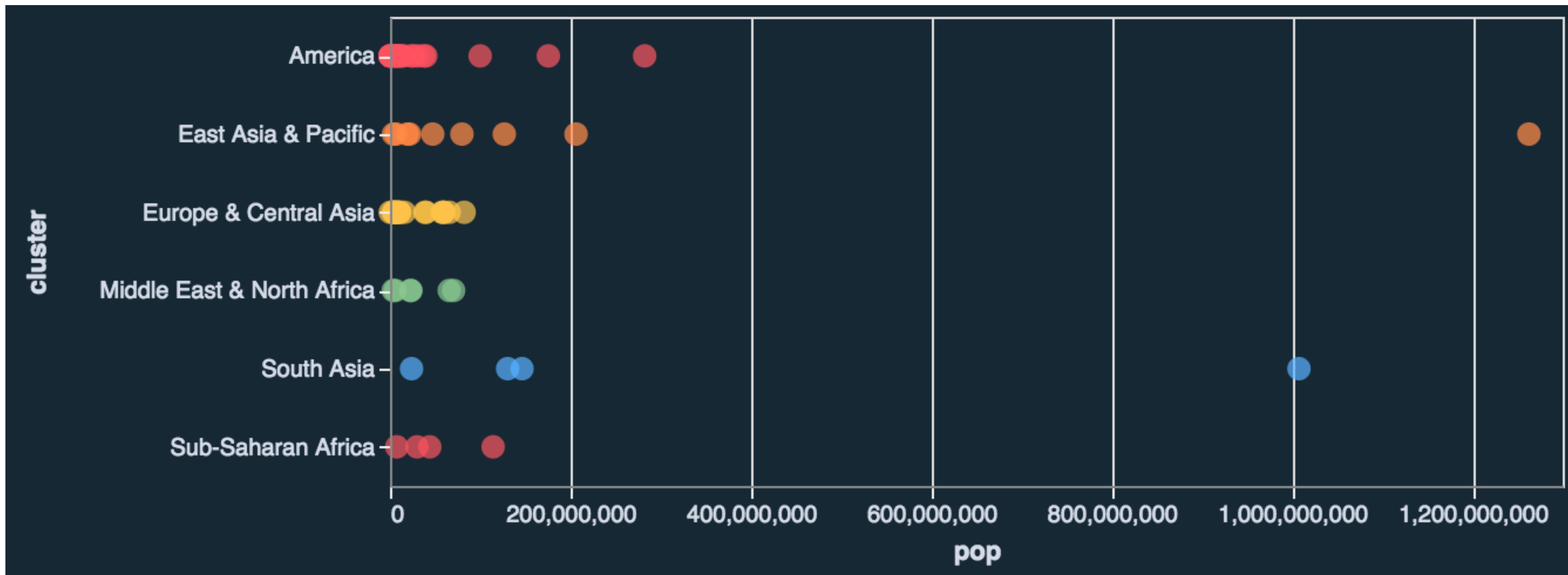
1. Clip them out.
2. Scale breaks
3. Log scale



# Scaling Axes: Outliers and Skew

Options:

1. Clip them out.
2. Scale breaks
3. Log scale



# Scaling Axes: Outliers and Skew

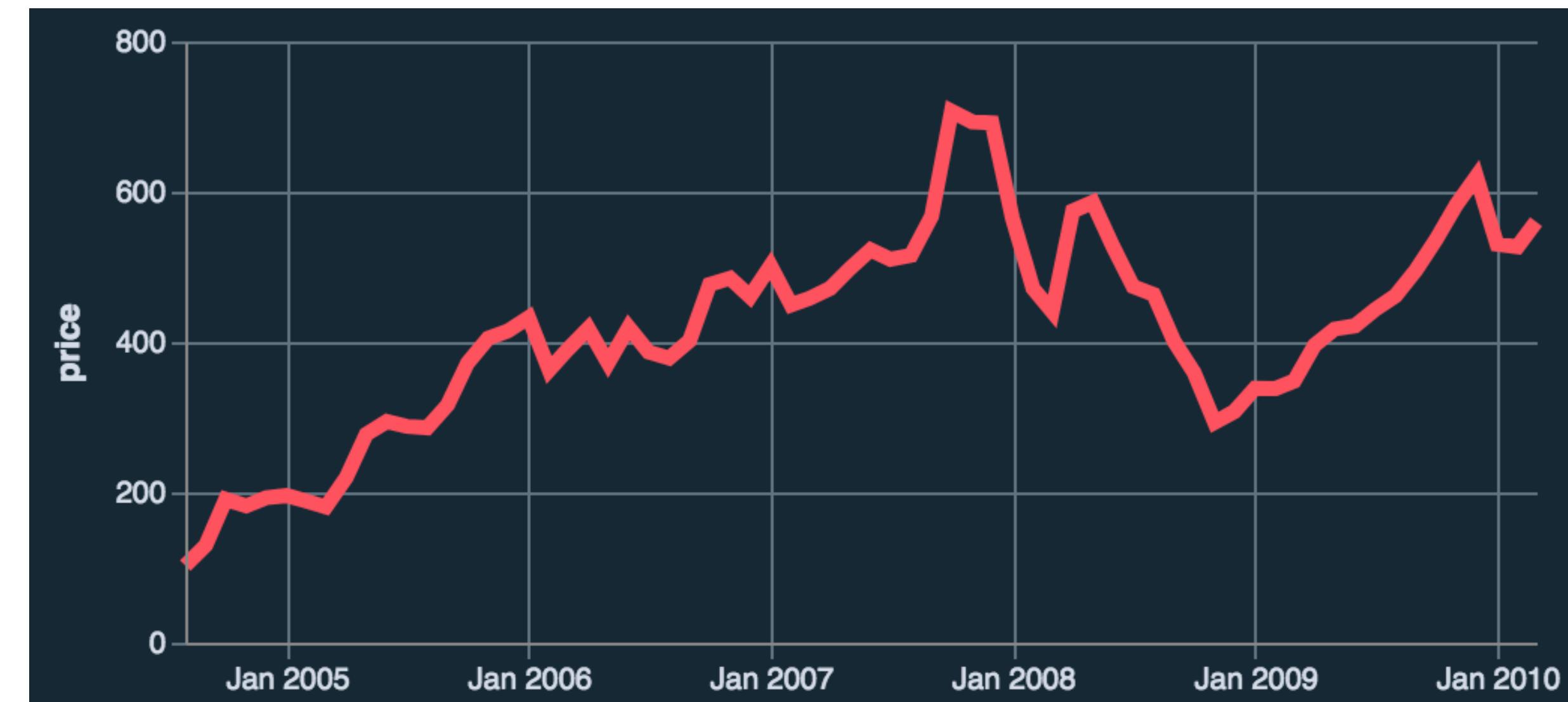
Options:

1. Clip them out.
2. Scale breaks
3. Log scale

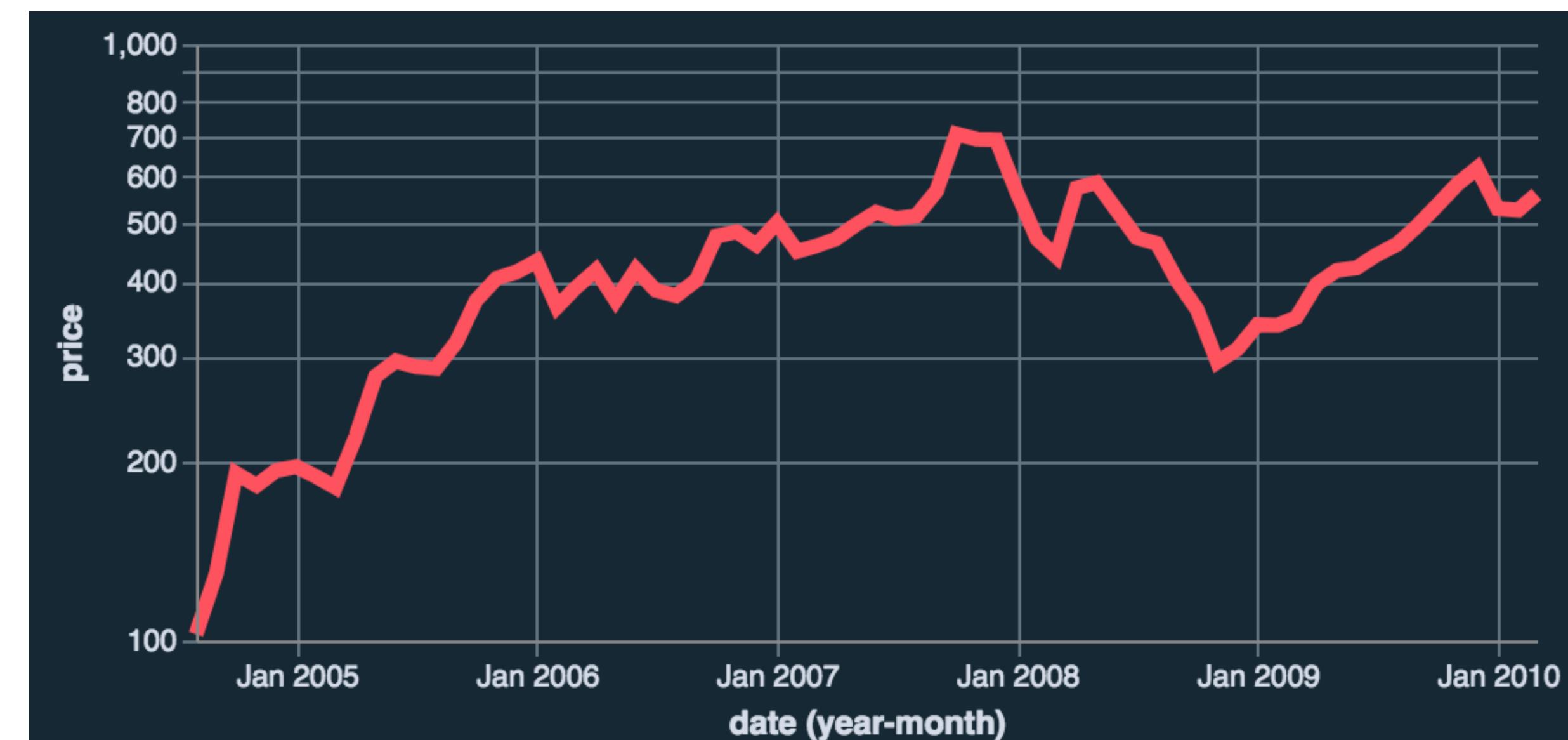


# Scaling Axes: Linear vs Log

**Linear Scale**  
Absolute change  
10 visual units (pixels) =  
10 additional data units



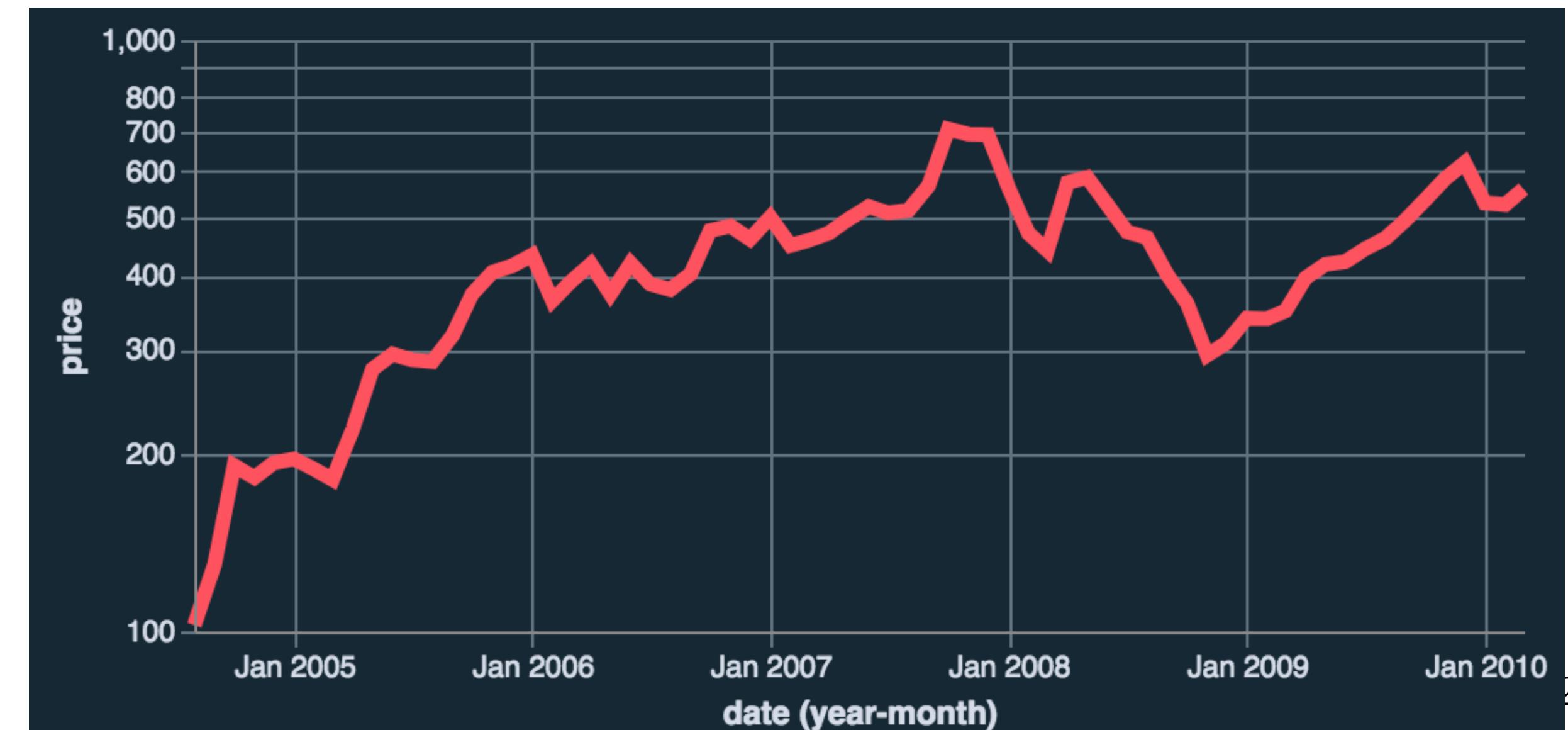
**Log Scale**  
Percentage change  
10 visual units =  
multiplication of 10 data units



# Scaling Axes: Linear vs Log

**Constraints**  
Positive, non-zero values  
Audience familiarity?

**Log Scale**  
Percentage change  
10 visual units =  
multiplication of 10 data units



# Using space (in)effectively

(De-)Obfuscating data

(Mis)leading the witness

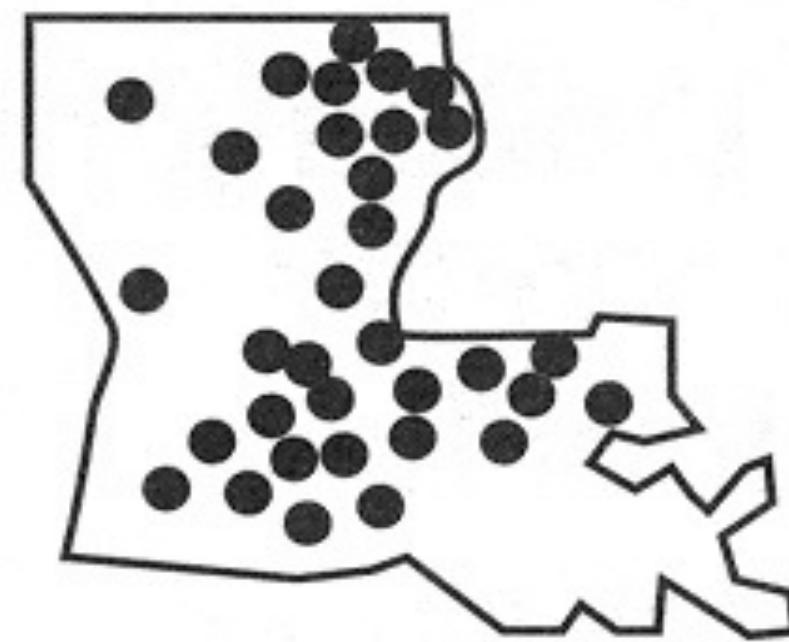
# Using space (in)effectively

## **(De-)Obfuscating data**

## (Mis)leading the witness



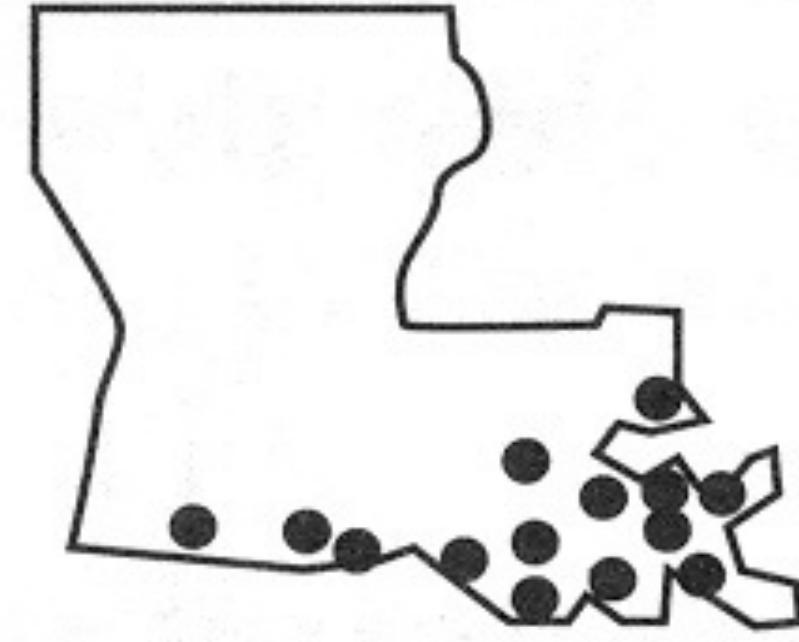
alfisol



entisol



histosol



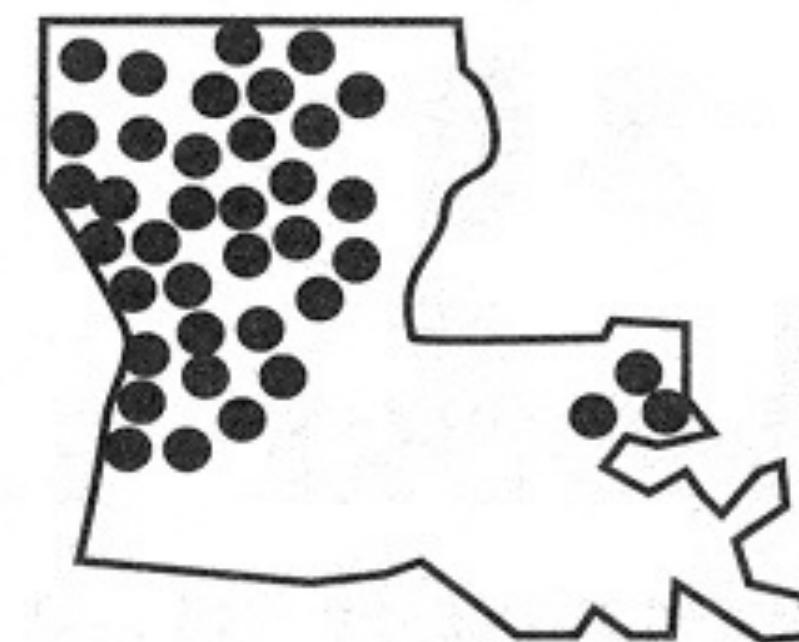
inceptisol



mollisol



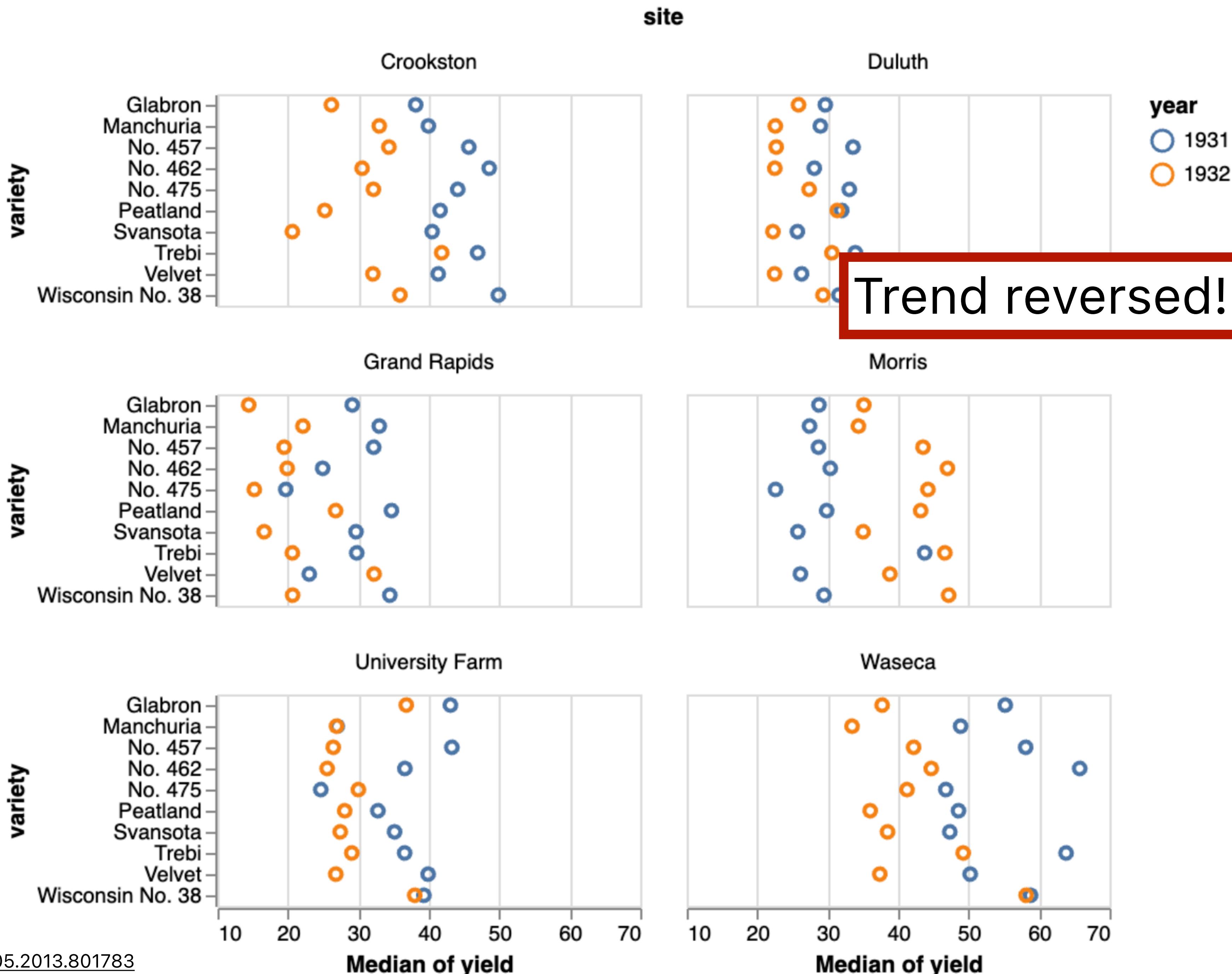
ultisol



# Trellis Plots

Subdivide space to enable comparison across multiple plots.

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



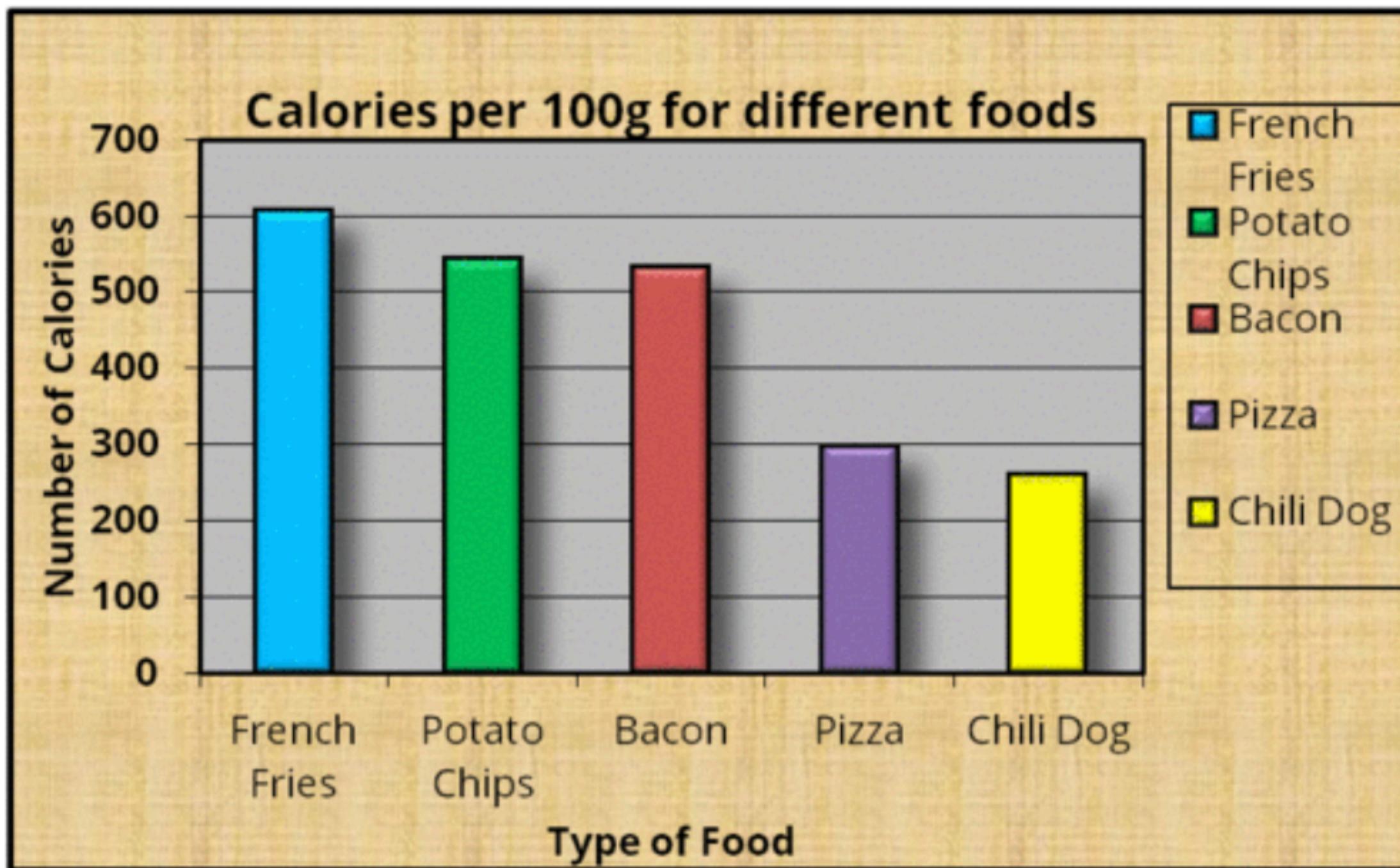
# Data-ink Ratio

- $$\text{Data Ink} \over \text{Ink used in graphic}$$
- = Proportion of a graphic's ink devoted to non-redundant display of data.
- =  $1.0 - \text{proportion of graphic that can be erased.}$

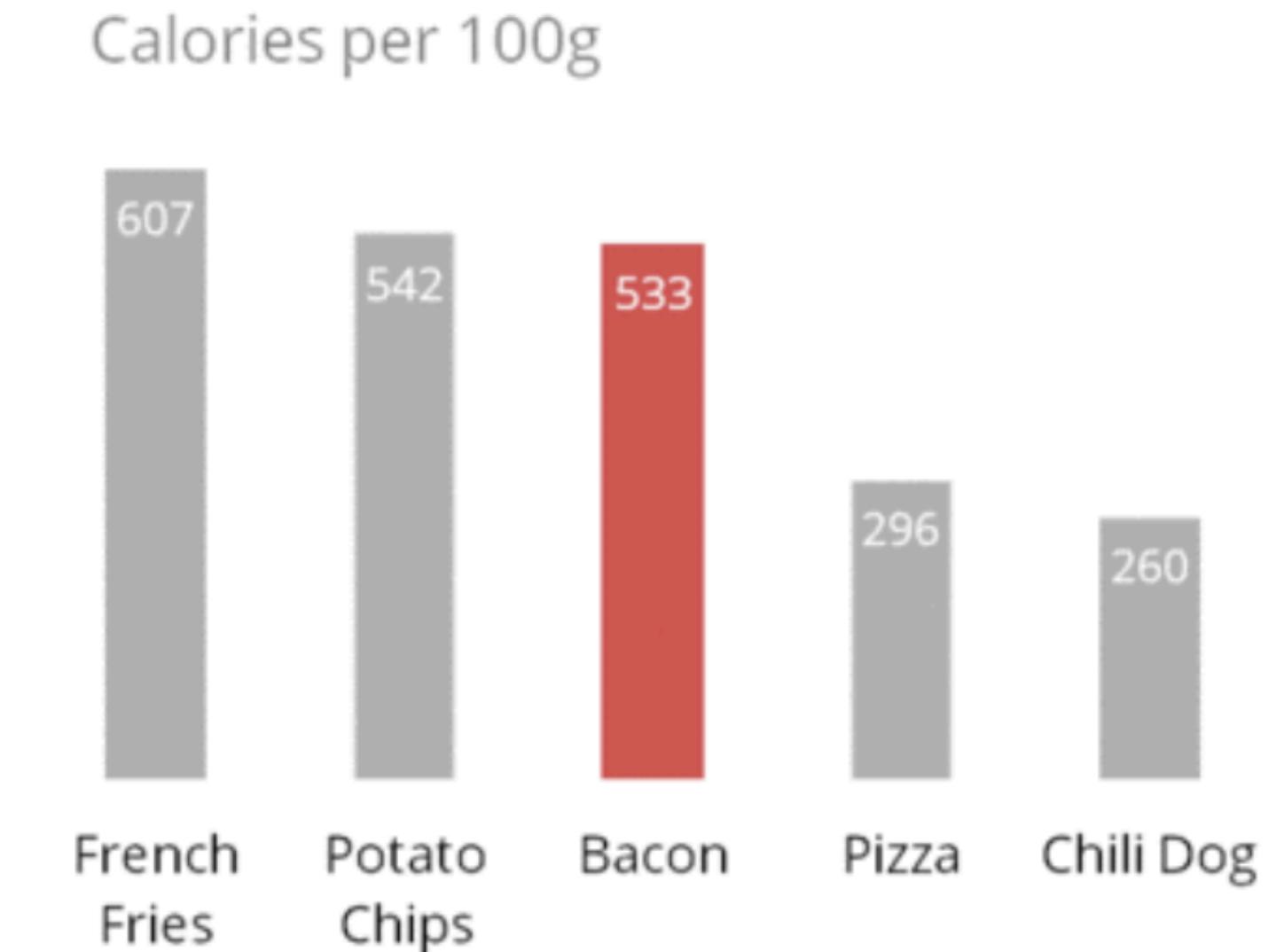
**Remove**  
to improve  
(the **data-ink** ratio)

# Data-ink Ratio

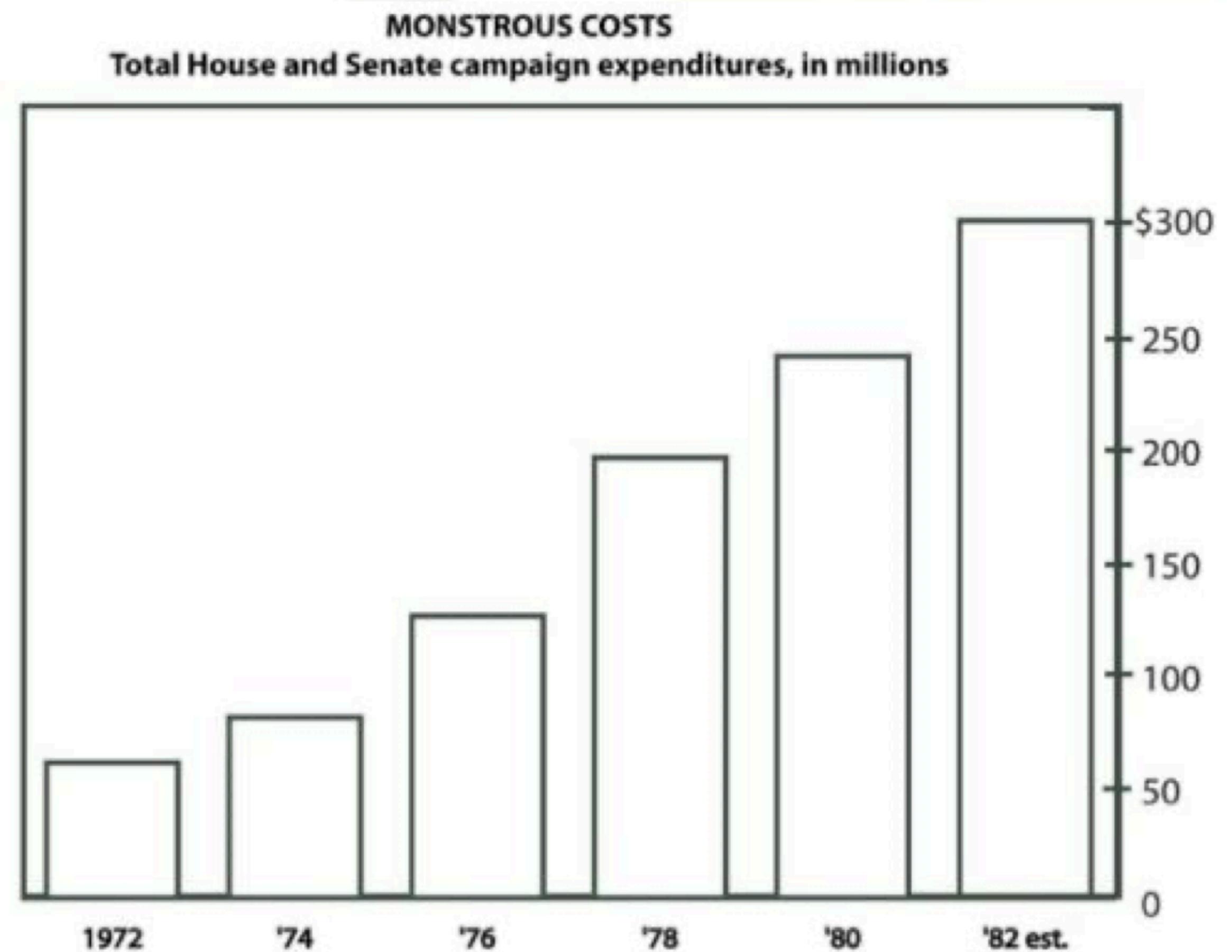
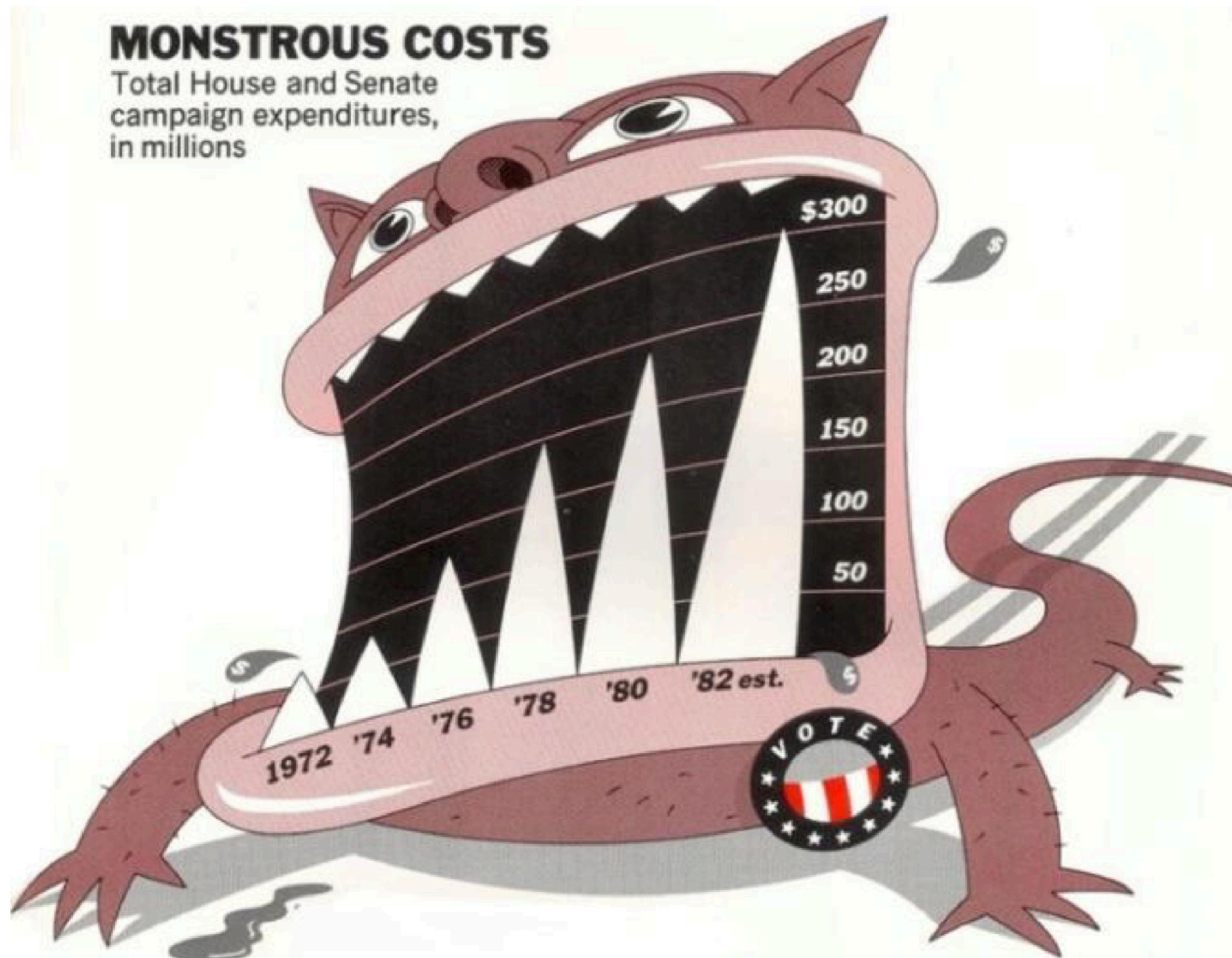
When is the data-ink ratio helpful?  
Does it have limitations?  
Might it ever be harmful?  
Is there benefit in using ink for non-data?



Join at  
[slido.com](https://slido.com)  
#3892 640

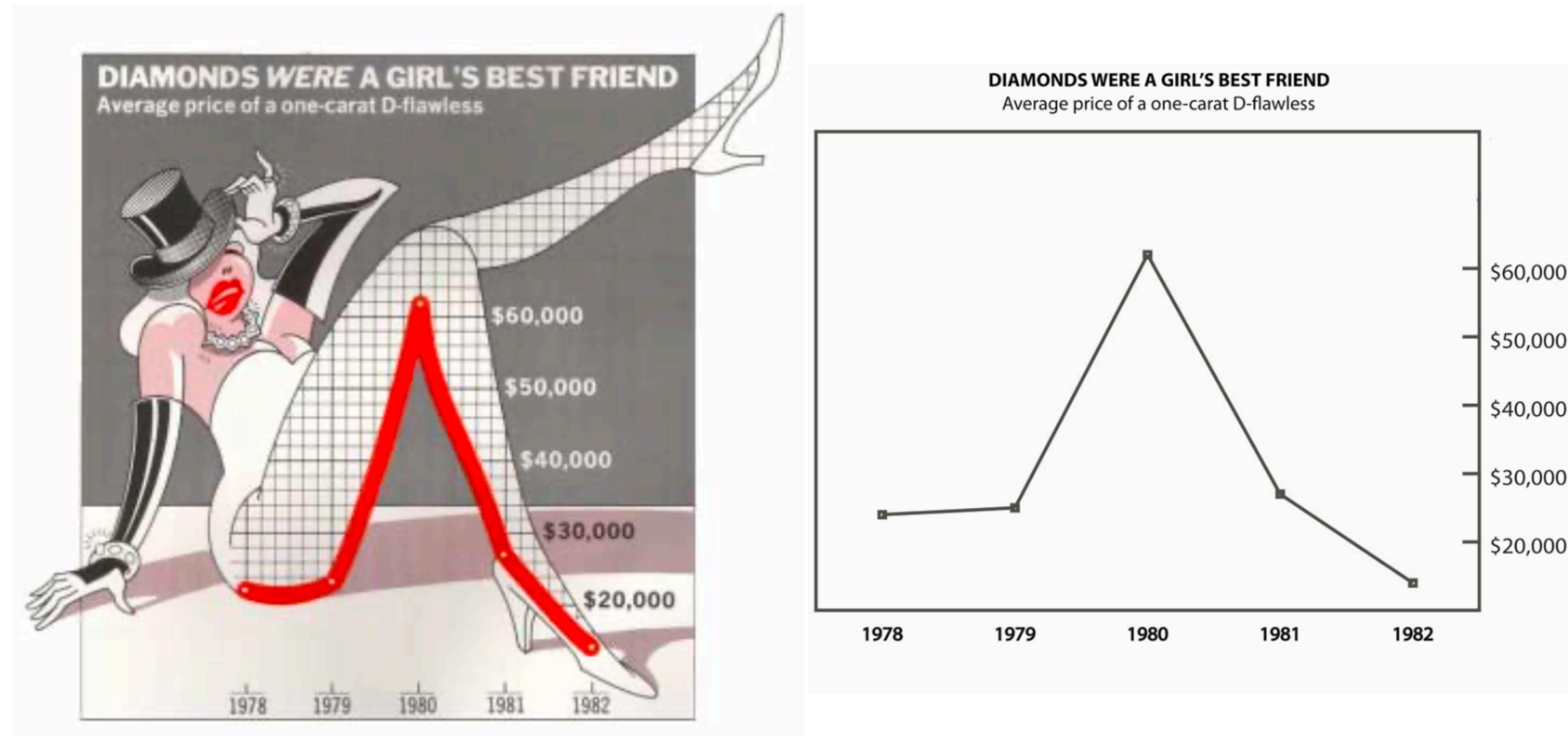


# Chart "Junk"



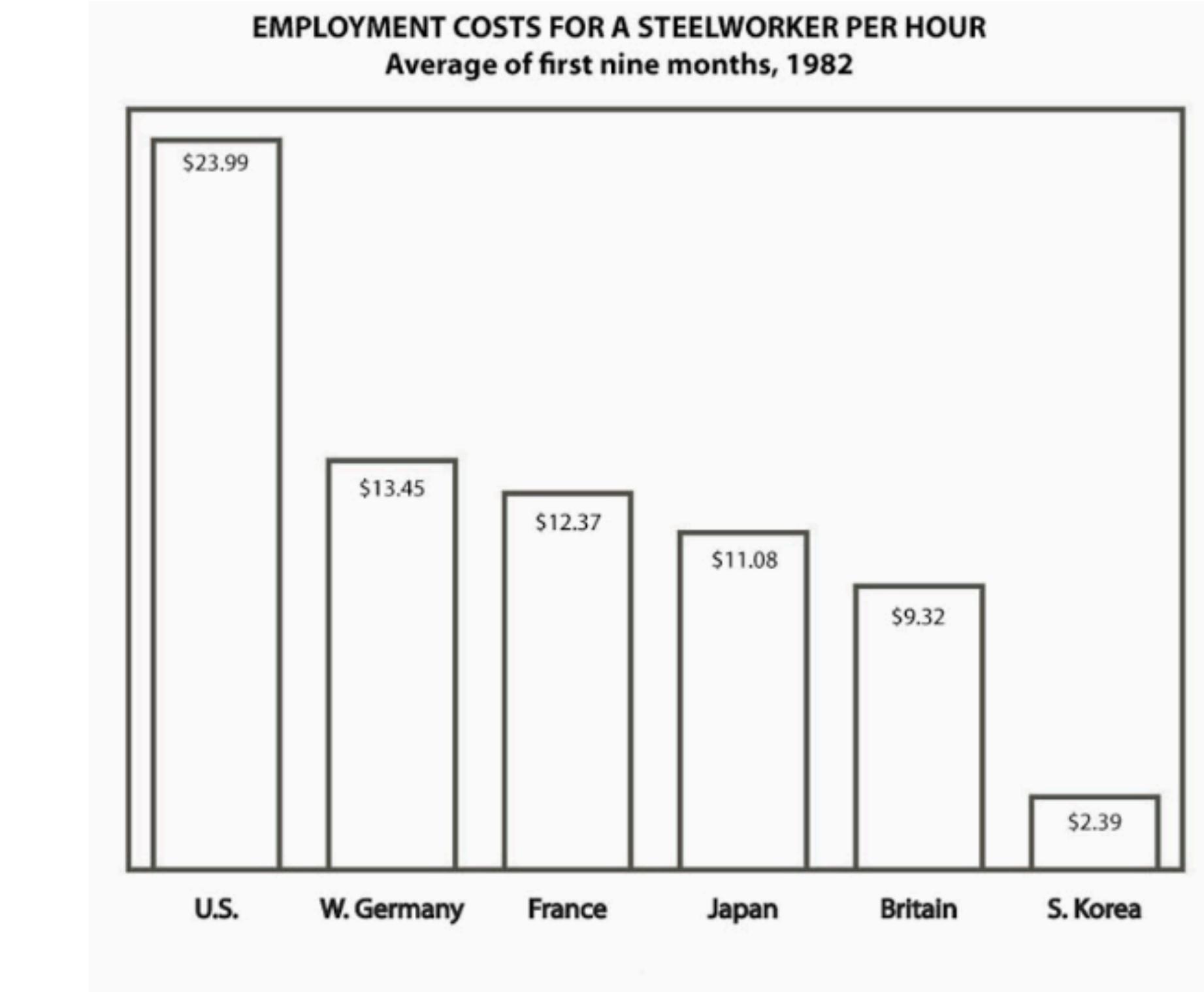
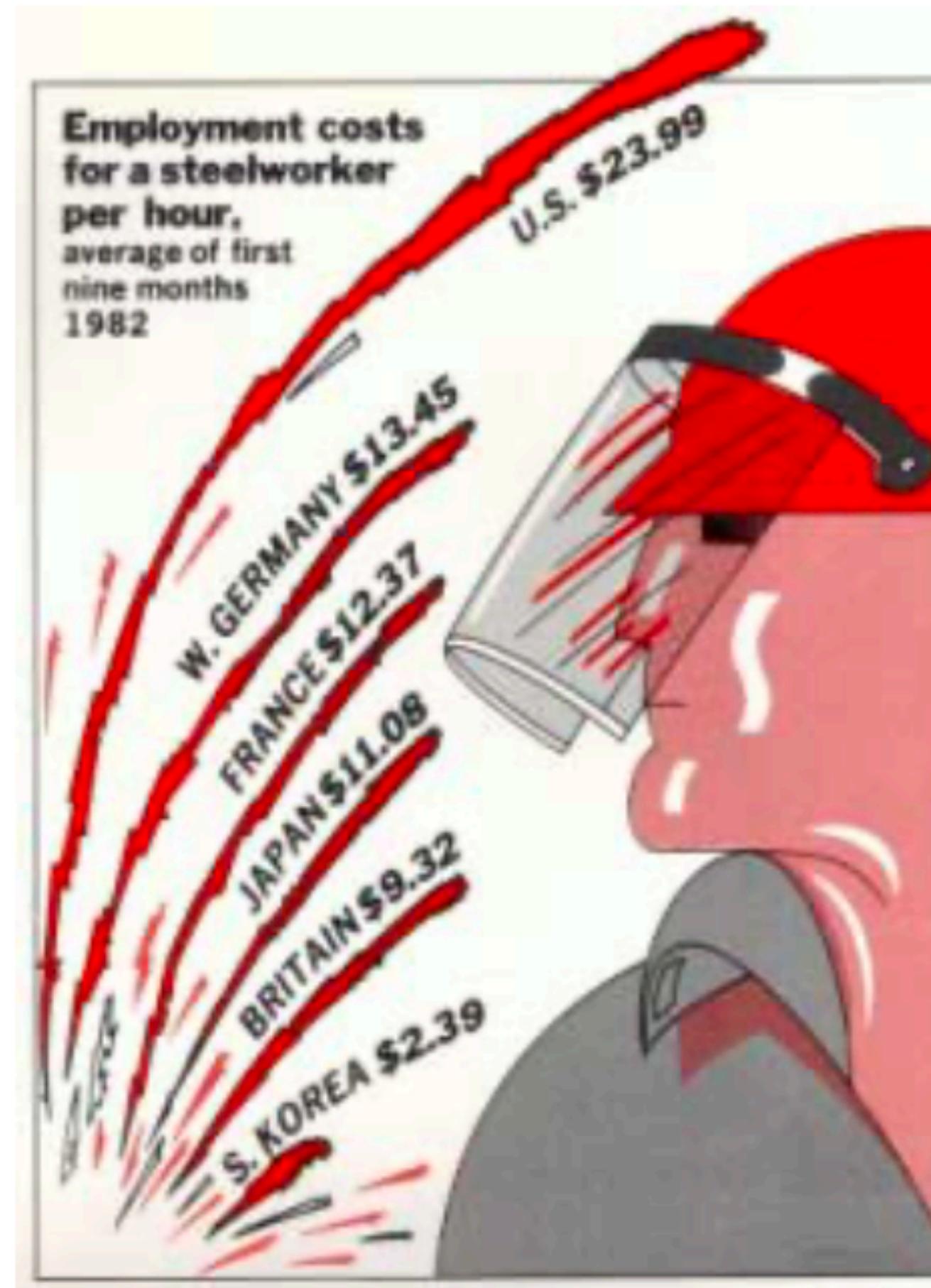
Bateman, Scott, et al. "Useful junk? The effects of visual embellishment on comprehension and memorability of charts." CHI 2010.

# Chart "Junk"



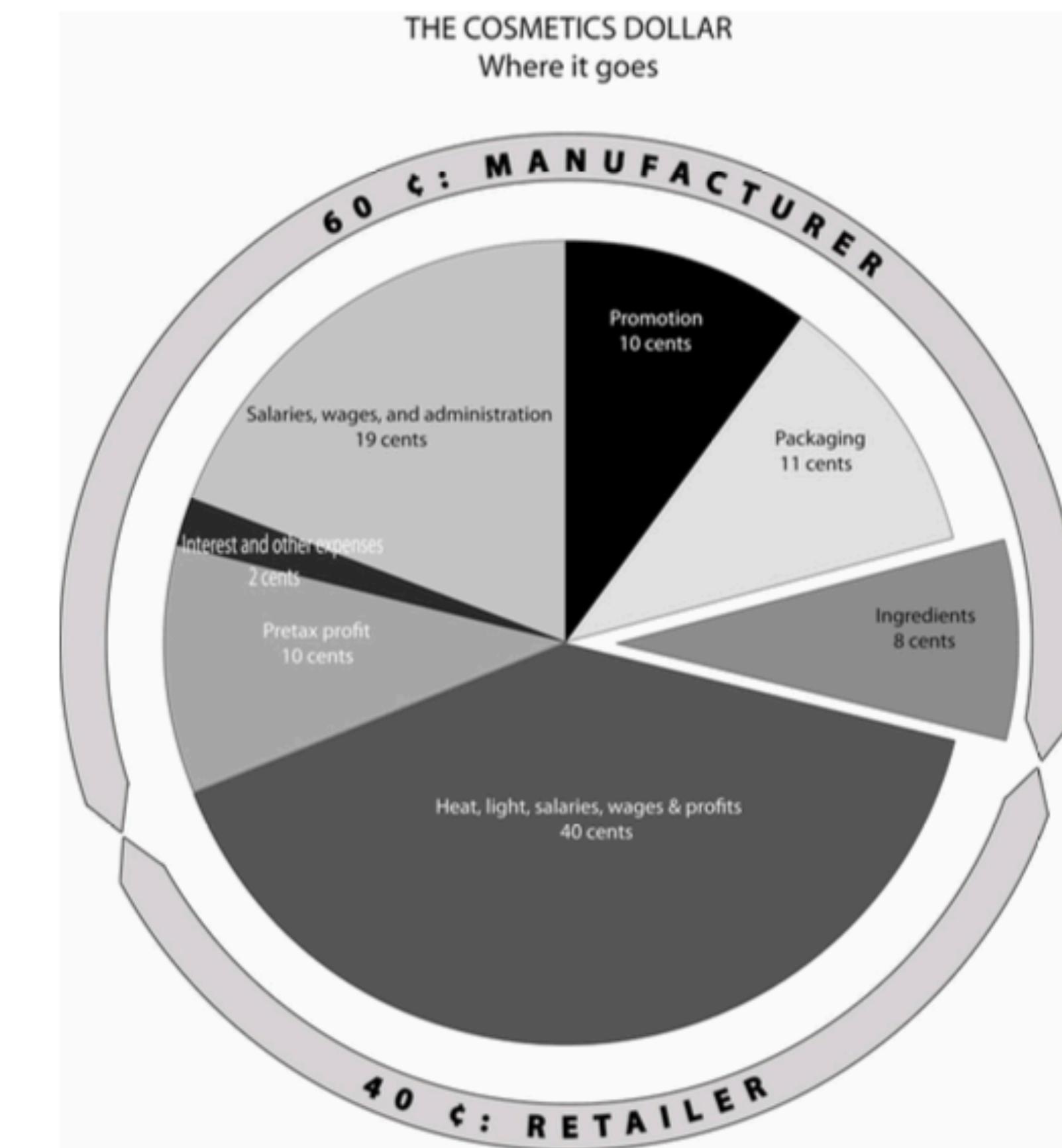
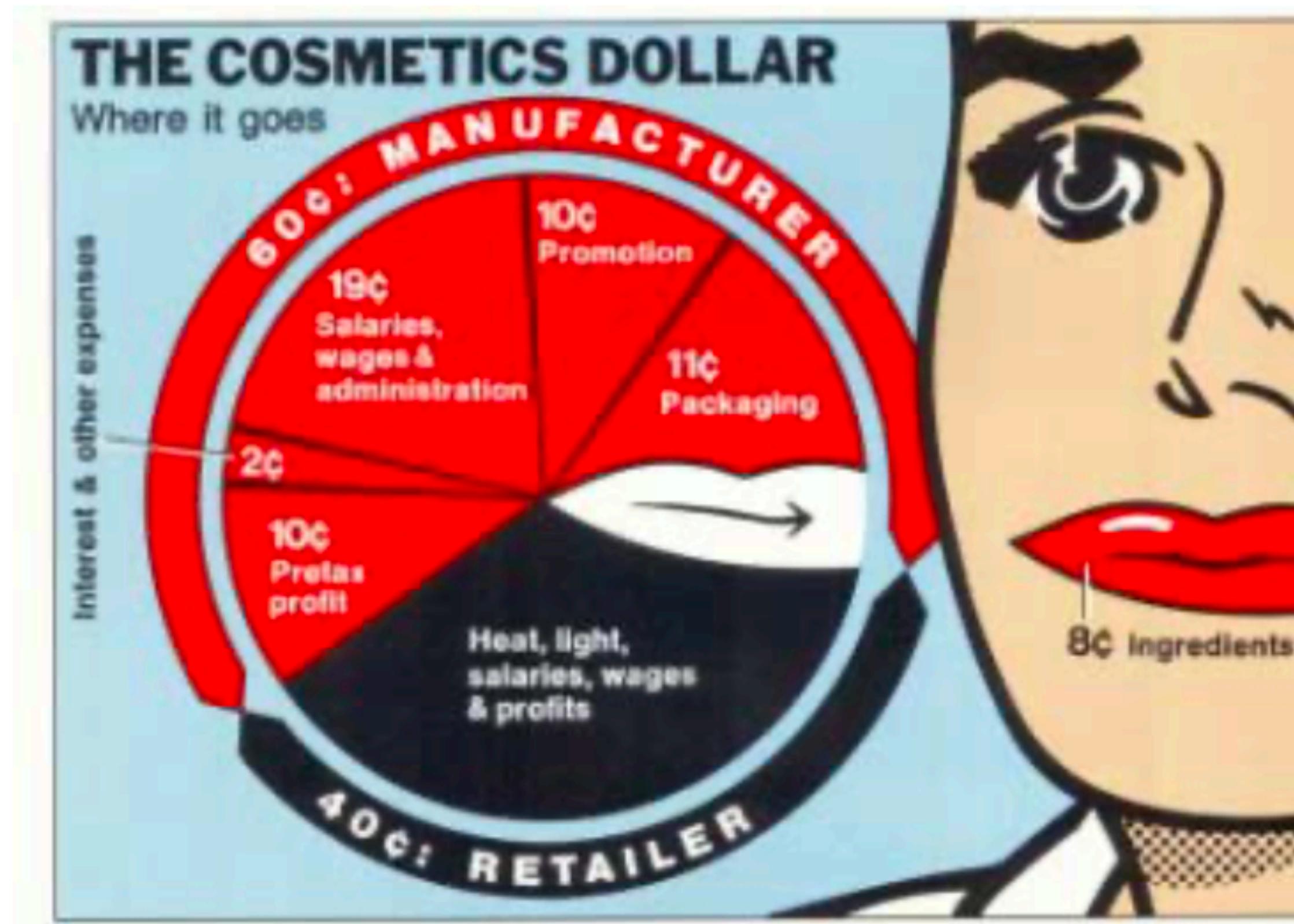
Bateman, Scott, et al. "Useful junk? The effects of visual embellishment on comprehension and memorability of charts." *CHI 2010*.

# Chart "Junk"

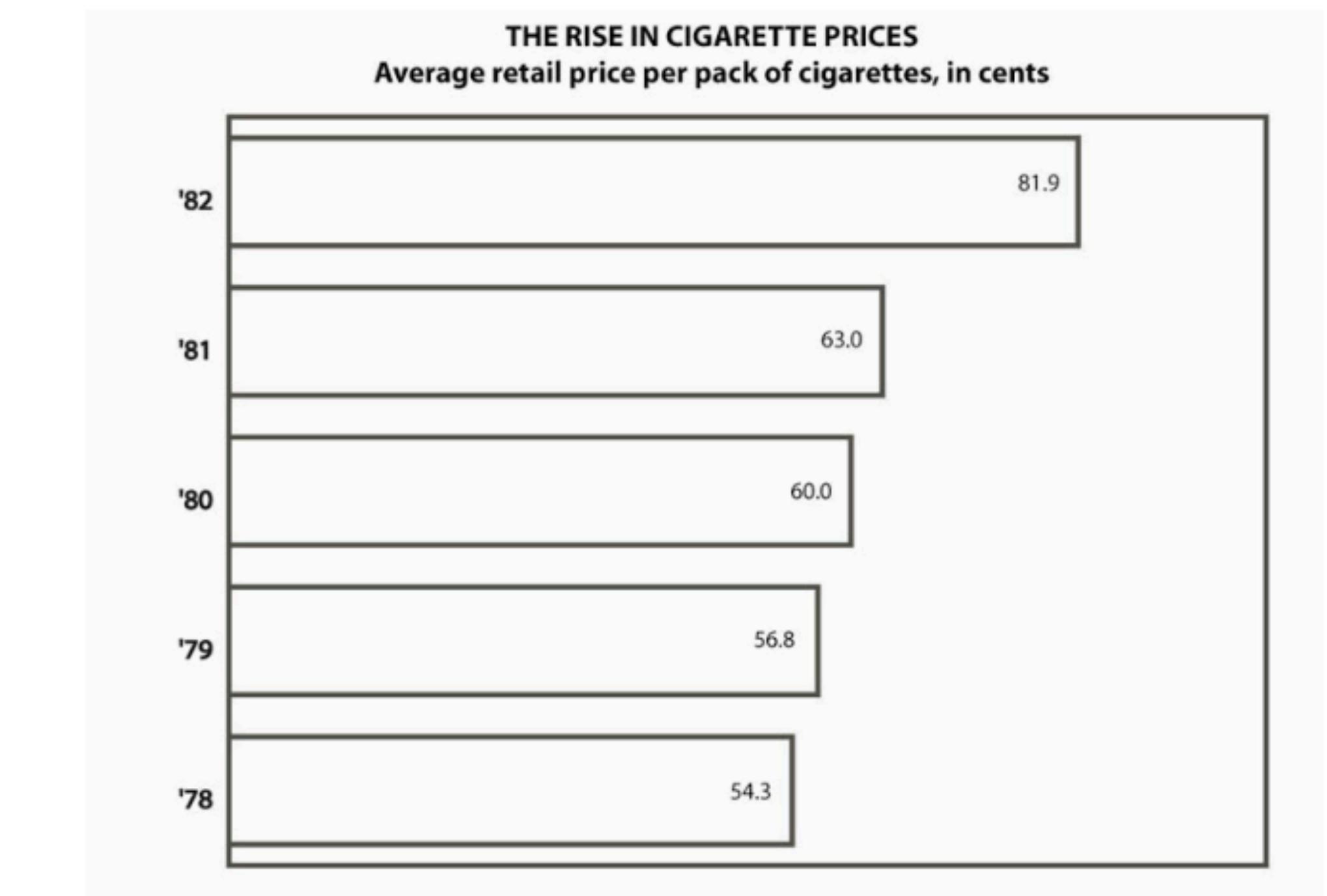
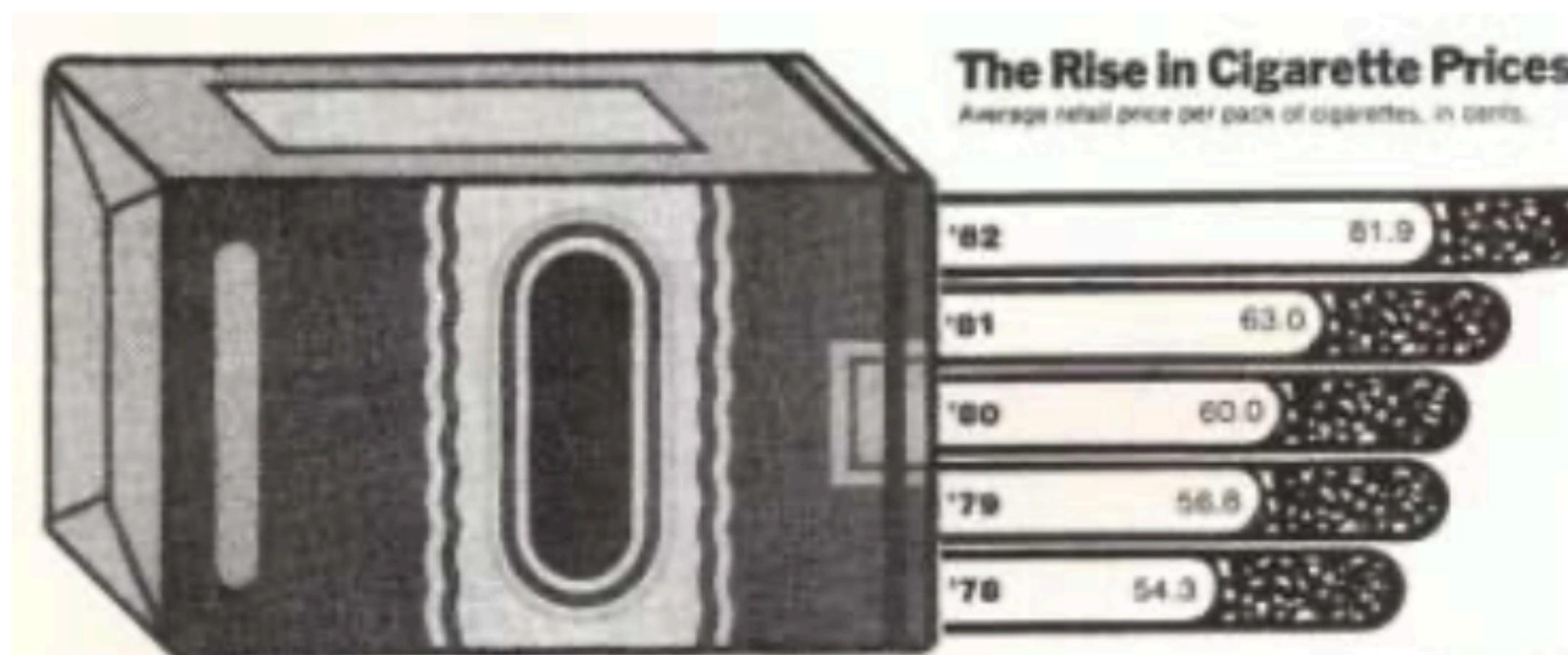


Bateman, Scott, et al. "Useful junk? The effects of visual embellishment on comprehension and memorability of charts." *CHI 2010*.

# Chart "Junk"



# Chart "Junk"



# Using space (in)effectively

## (De-)Obfuscating data

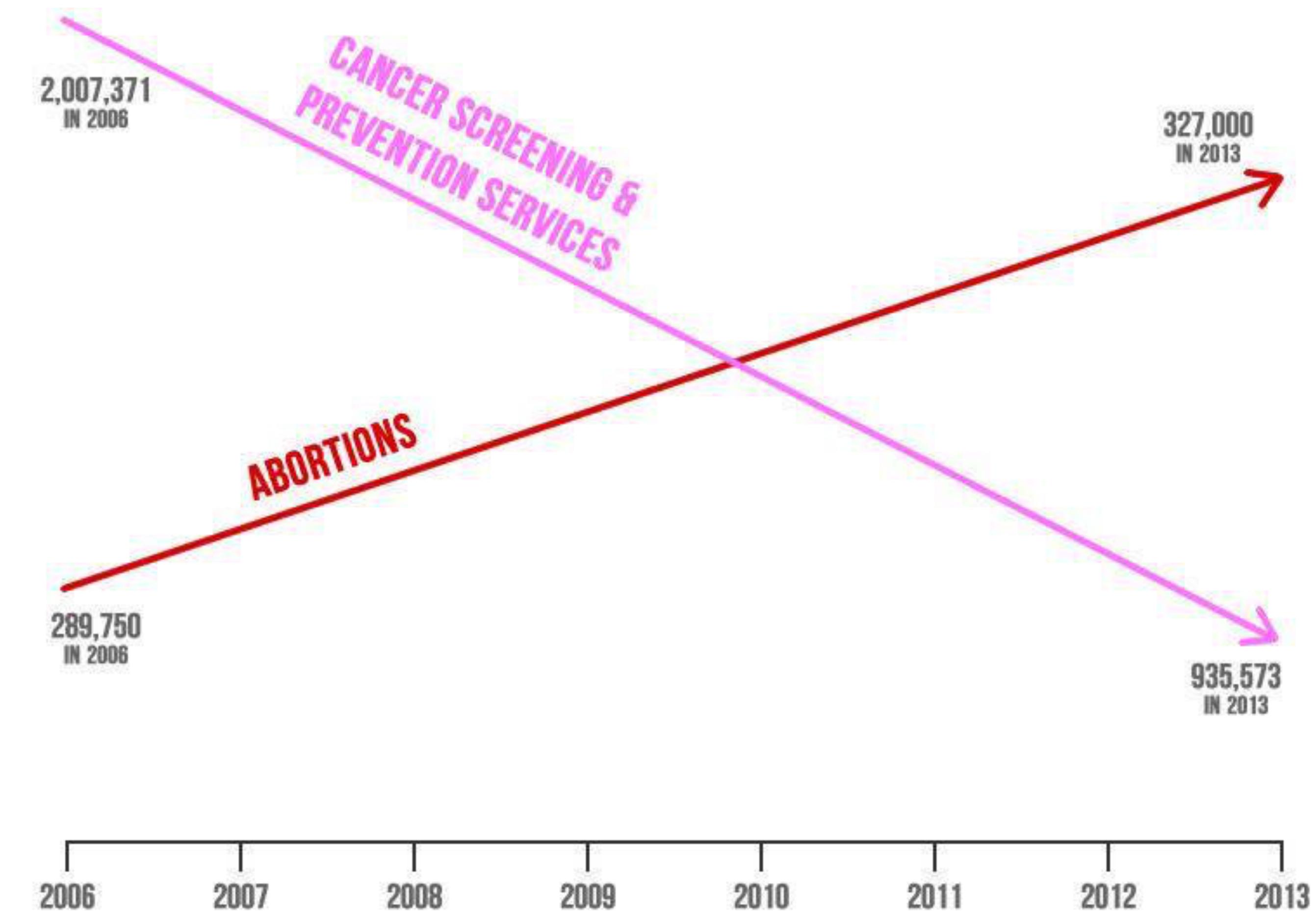
## (Mis)leading the witness

# Using space (in)effectively

## (De-)Obfuscating data

## (Mis)leading the witness

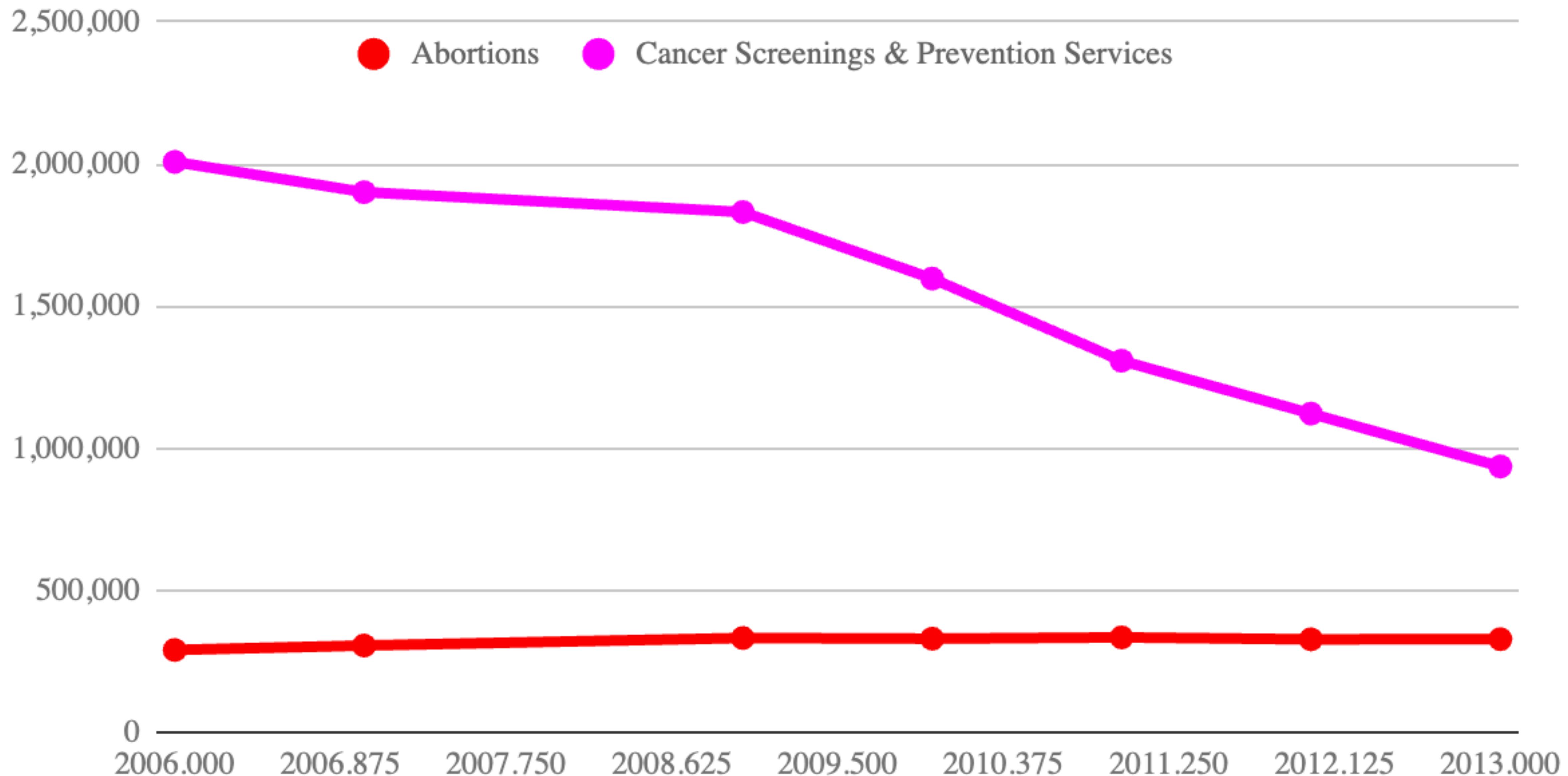
## PLANNED PARENTHOOD FEDERATION OF AMERICA: ABORTIONS UP – LIFE-SAVING PROCEDURES DOWN



Join at  
**slido.com**  
**#3892 640**



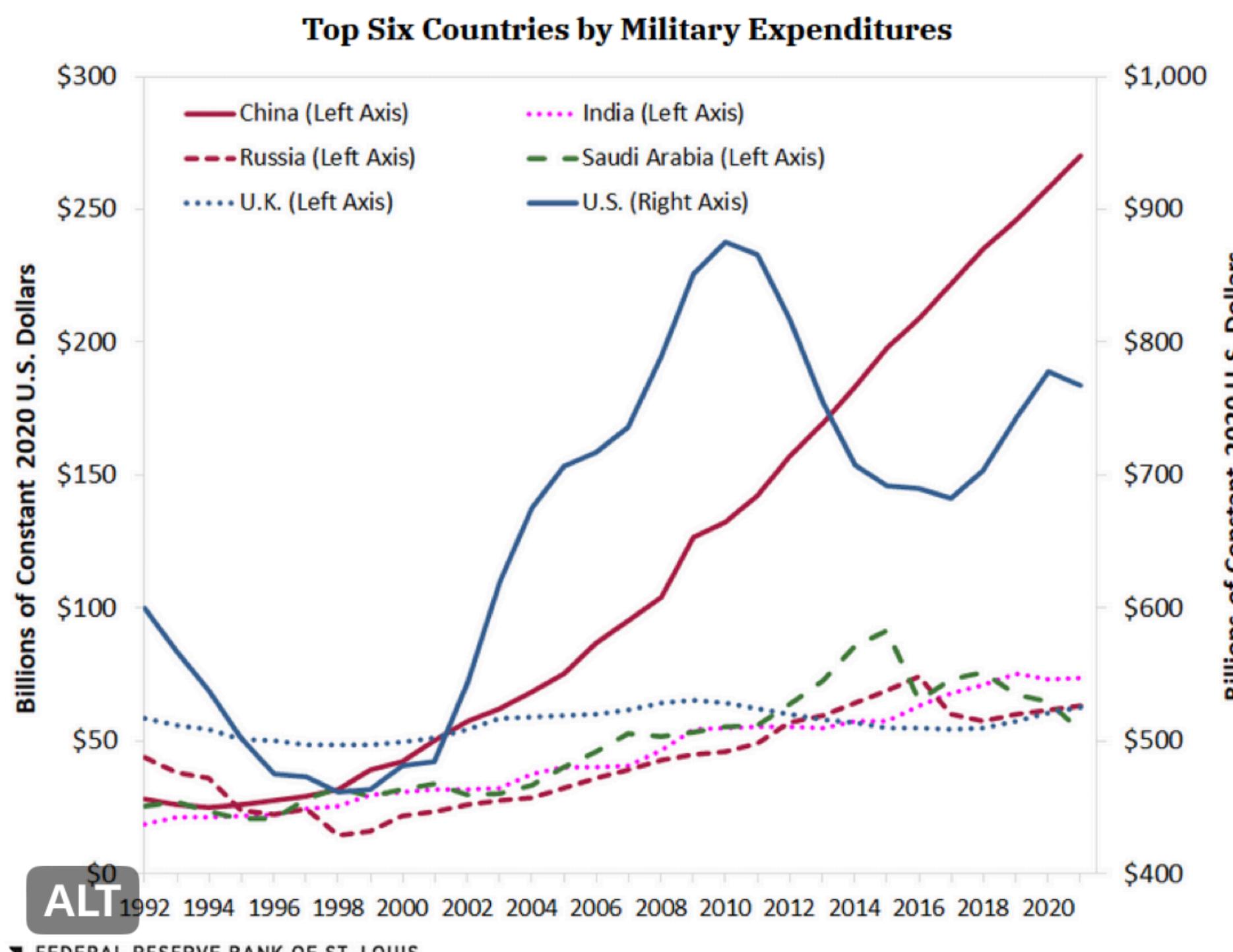
# Planned Parenthood Federation of America: Abortions vs. Cancer and Prevention Services





...

An analysis looks at how defense spending among the nations with the highest expenditures has changed since 1992 and what may have driven the changes [ow.ly/MyOx50MwEyF](http://ow.ly/MyOx50MwEyF)



Readers added context they thought people might want to know

While this information is correct, the graph is poorly formatted, with a separate Y-axis on the right-hand side which only applies to the US budget. This may make it seem like China has a higher military budget than the US, when the reverse is true.

[data.worldbank.org/indicator/MS.M...](http://data.worldbank.org/indicator/MS.M...)

Do you find this helpful?

Rate it

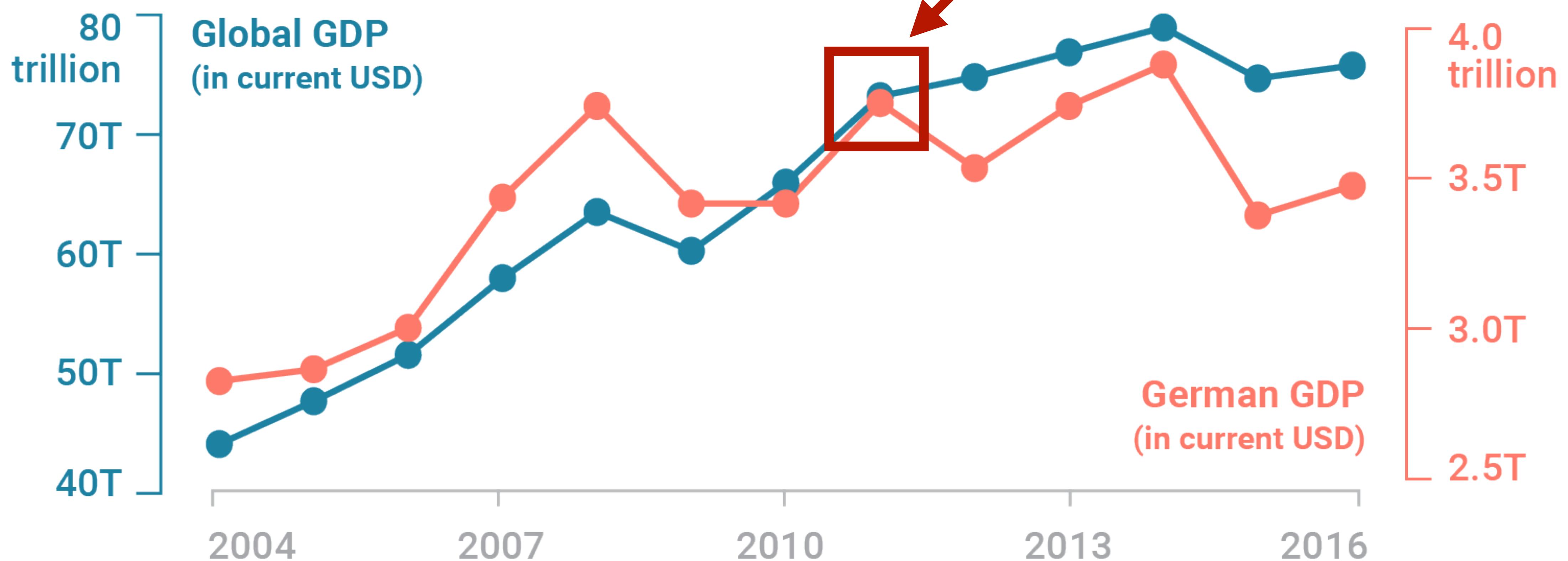
Context is written by people who use Twitter, and appears when rated helpful by others. [Find out more.](#)

4:00 PM · 1/22/23 · 7.3M Views

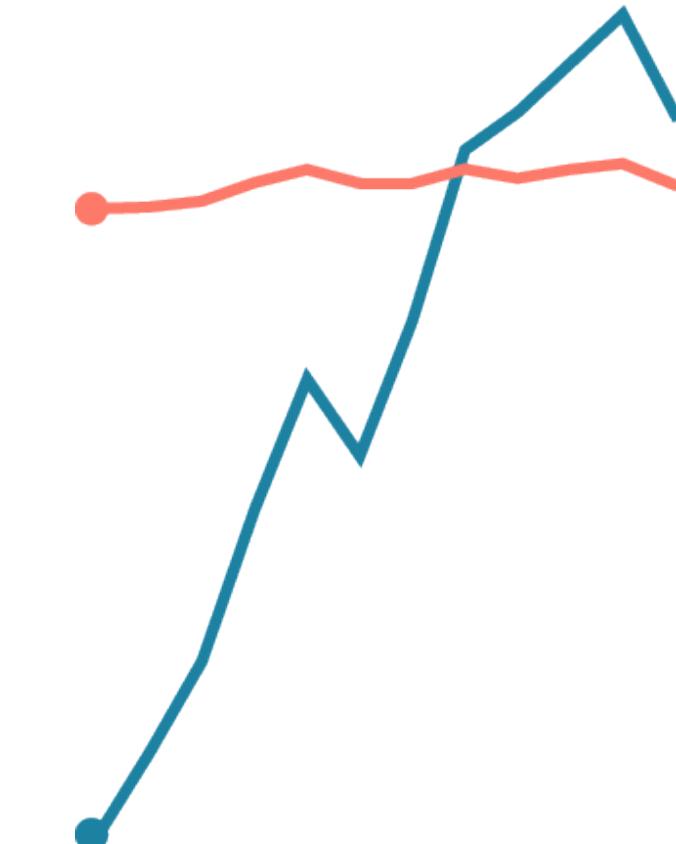
1,128 Likes 157 Retweets 2,281 Quotes

# Dual Axes Charts

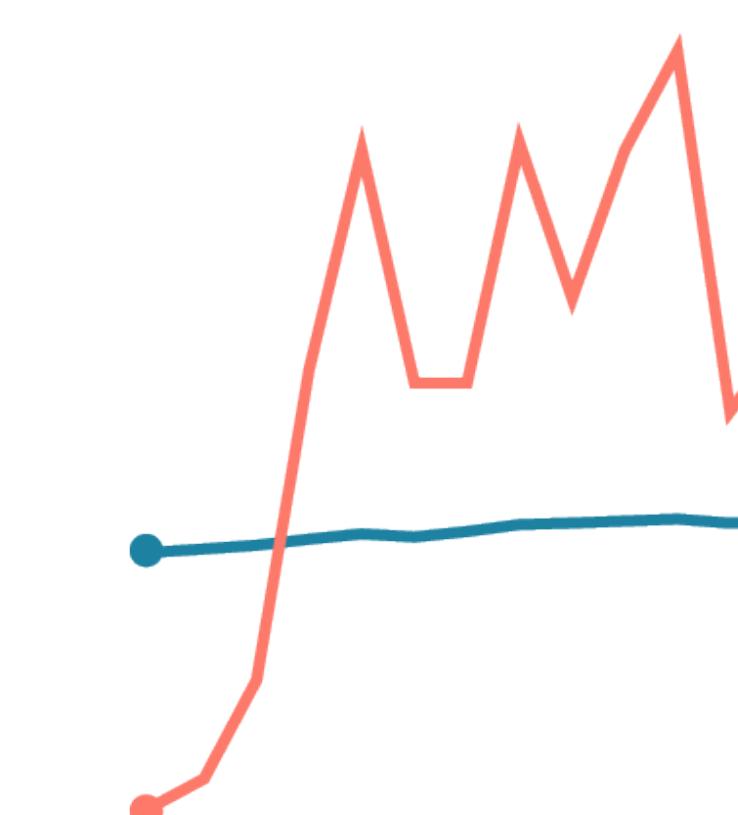
German and world GDP  
were equal in 2011??



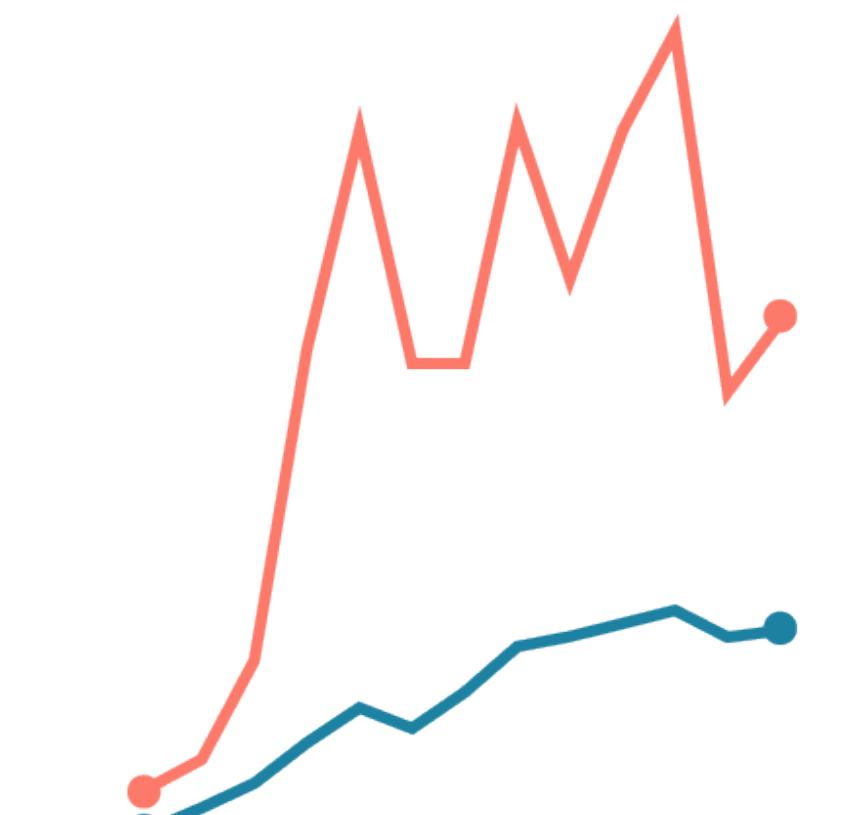
# Dual-Axes Charts



Orange steady,  
Blue massively increasing.



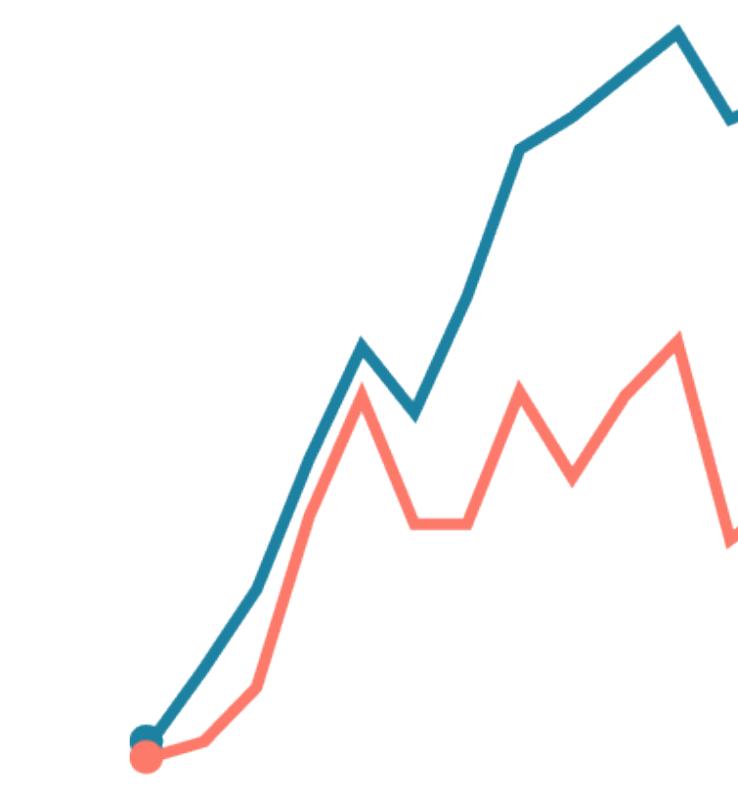
Blue steady,  
Orange increasing.



Both started at the same level, but Orange increased far more than Blue.

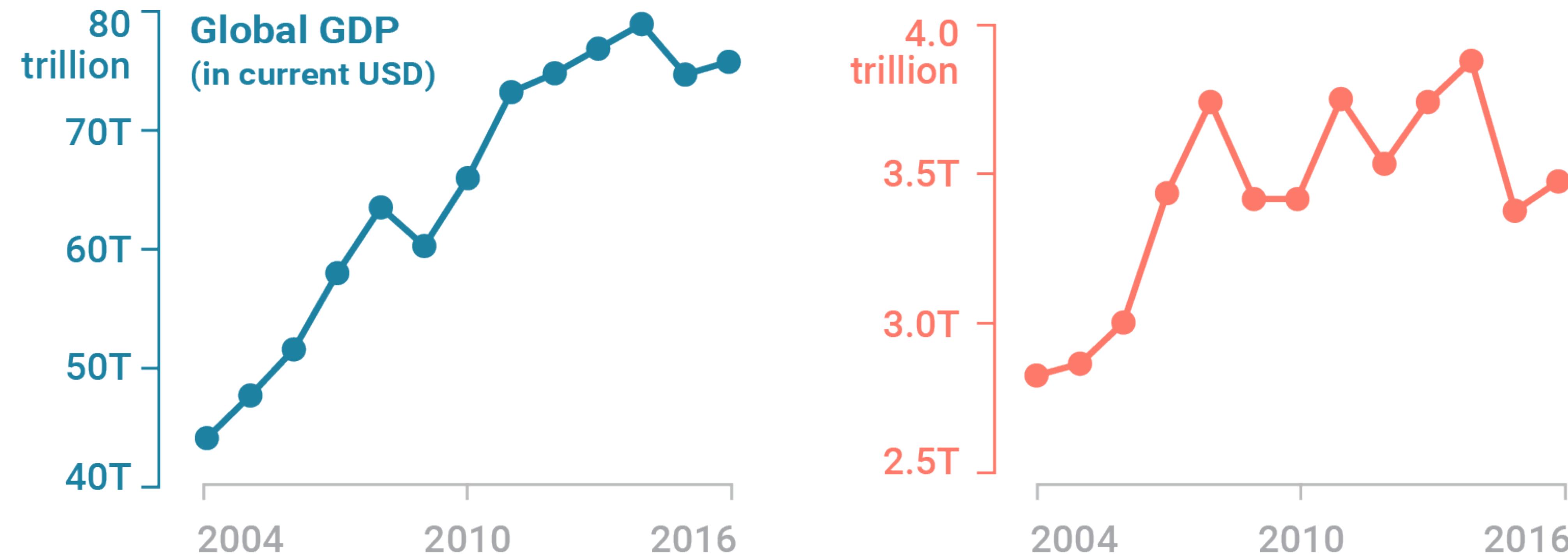


Both started at the same level, but Blue increased far more than Orange.



Both started with the same increase, then Blue raced to the top.

# Dual-Axes Charts



# Using space (in)effectively

## (De-)Obfuscating data

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

## (Mis)leading the witness

Visualization draws upon both science and art!

# **Next Time: Perception**