(In)Effective Visual Encoding

DSC 106: Data Visualization

Sam Lau

UC San Diego

Join at

slido.com

#3892 640

Announcements

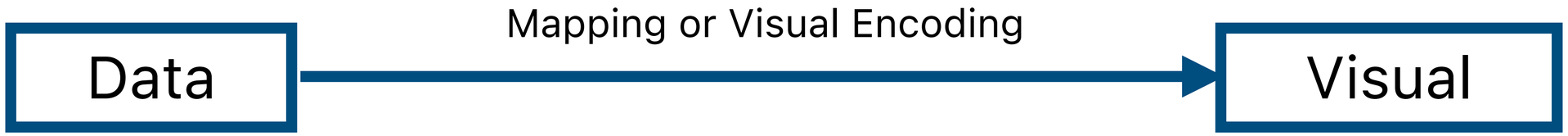
Lab 2 out, due this Friday, 1/19.

Project 1 also due this Friday, 1/19.

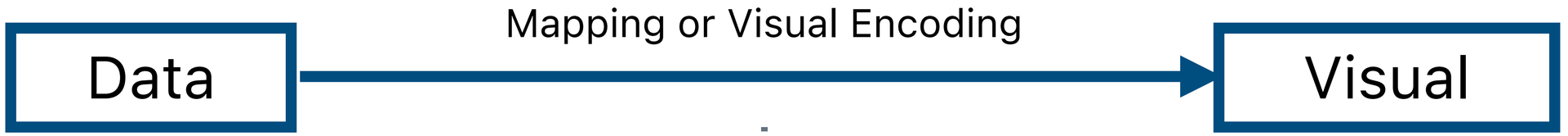
No lab checkoffs during Sam's OH, on Thurs now instead of Fri

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.



Mackinlay, Jock. "Automating the design of graphical presentations of relational information." *Acm Transactions On Graphics (Tog)* 5.2 (1986): 110-141.



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

Mackinlay, Jock. "Automating the design of graphical presentations of relational information." *Acm Transactions On Graphics (Tog)* 5.2 (1986): 110-141.

Mapping or Visual Encoding

Data → Visual

Nominal
=, ≠
Labels or categories.
E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal
=, ≠, **<, >**
Ordered.
E.g., Quality of eggs: Grade AA, A, B

Quantitative (Interval)
=, ≠, <, >, **–**
Interval (zero can be arbitrarily located).
E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09)
Only differences can be calculated (e.g., distances or spans).

Quantitative (Ratio)
=, ≠, <, >, –, **%**
Ratio (fixed zero / meaningful baseline).
E.g., Physical measurement: length, mass, temperature
Counts and amounts. Can measure ratios or proportions.

Visual Variables

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered 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)

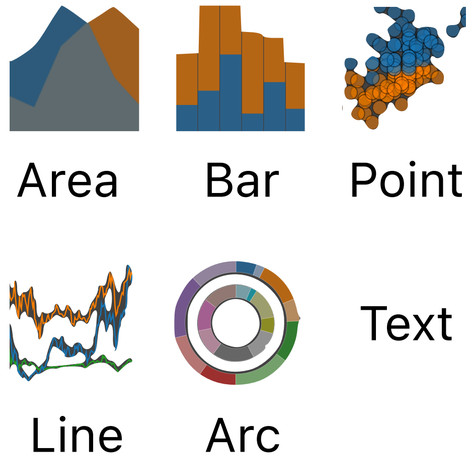
Same (Color luminance, Color saturation)
Same (Curvature, Volume (3D size))

Most ▲ Effectiveness ▼ Least

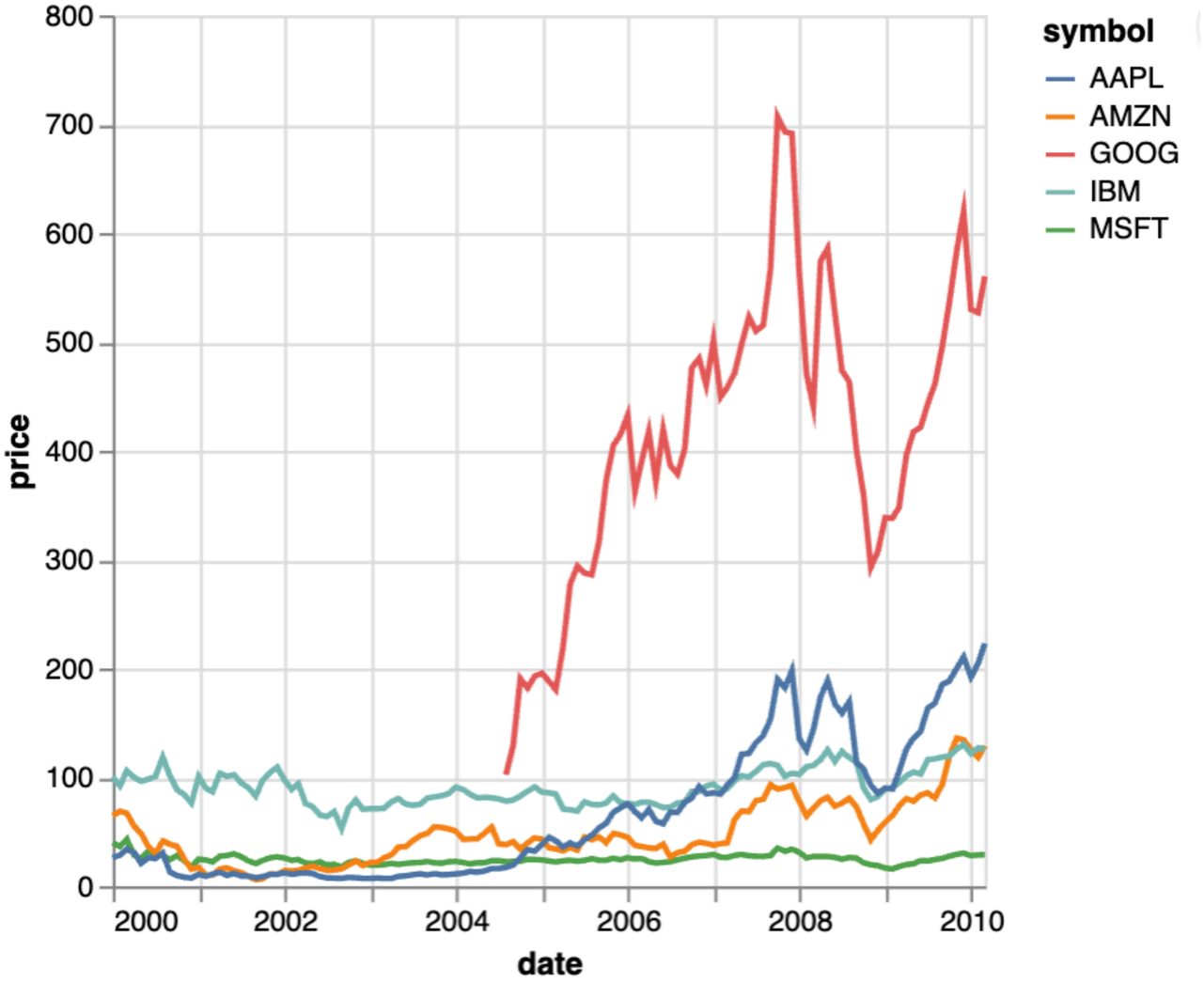
Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

Marks



Example from Lab 1



Mark: `line`

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

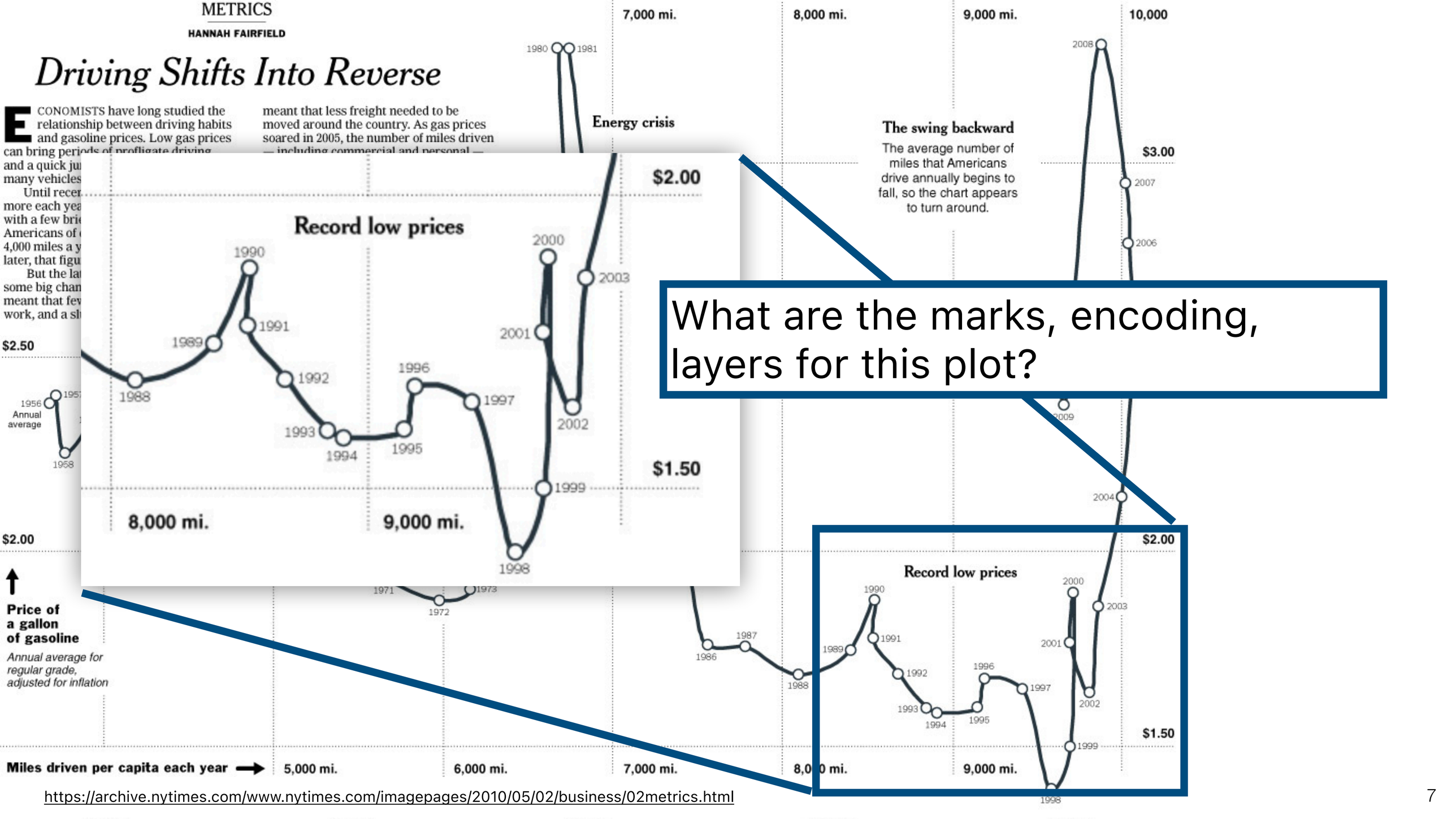
METRICS

HANNAH FAIRFIELD

Driving Shifts Into Reverse

ECONOMISTS have long studied the relationship between driving habits and gasoline prices. Low gas prices can bring periods of profligate driving and a quick ju many vehicles Until recer more each yea with a few bri Americans of 4,000 miles a y later, that figu But the la some big chan meant that fe work, and a sl

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 —



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

https://archive.nytimes.com/www.nytimes.com/imagepages/2010/05/02/business/02metrics.html

METRICS

HANNAH FAIRFIELD

Driving Shifts Into Reverse

ECONOMISTS have long studied the relationship between driving habits and gasoline prices. Low gas prices can bring periods of profligate driving and a quick ju many vehicles Until recer more each yea with a few bri Americans of 4,000 miles a y later, that figu But the la some big chan meant that fe work, and a sl

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 —

Join at

slido.com

#3892 640

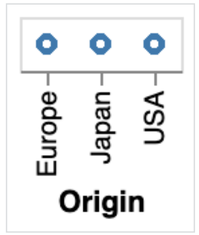
https://ar tml

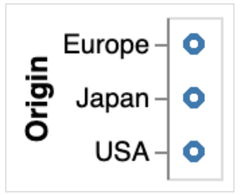
A Design Space of Visual Encodings

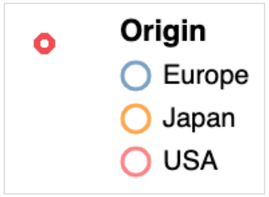
Visual Encoding = Combinatorial Design Space

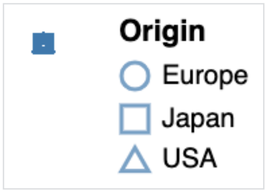
1D nominal data (N, O)

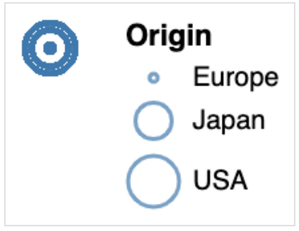
raw



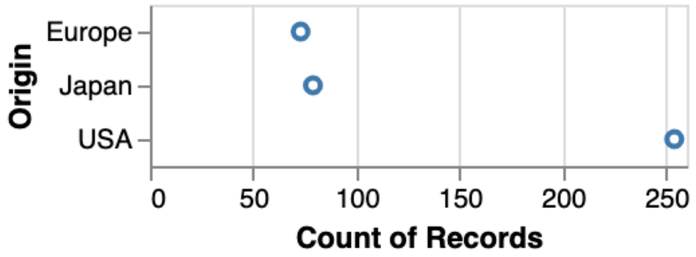


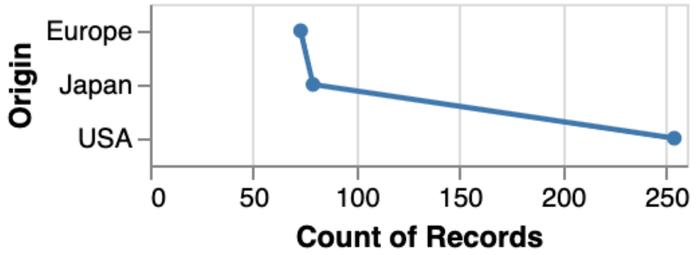


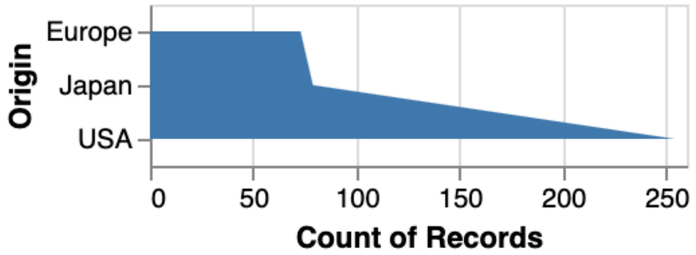


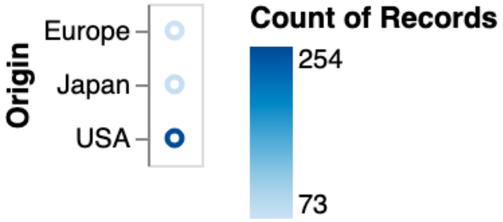


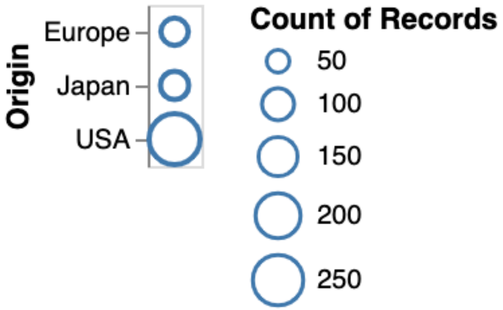
aggregate (count)

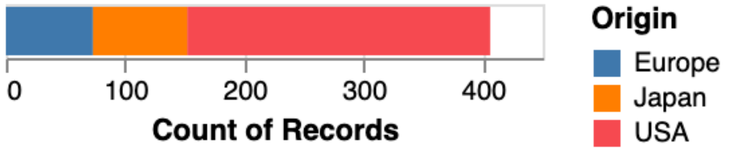


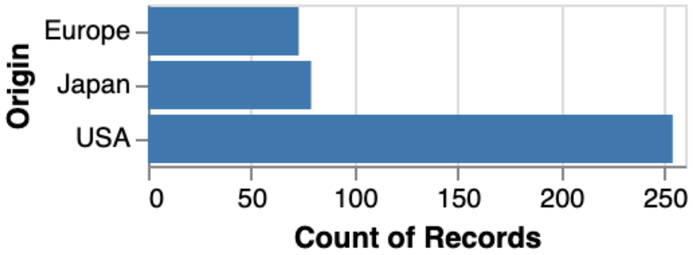










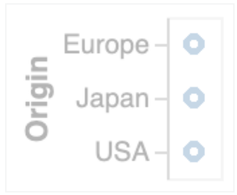


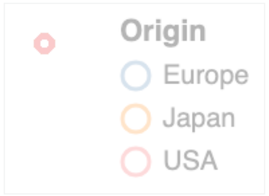
Visual Encoding = Combinatorial Design Space

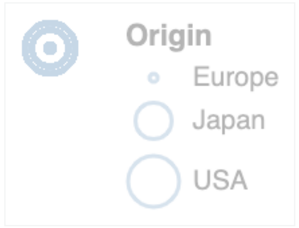
1D nominal data (N, O)

Expressive?

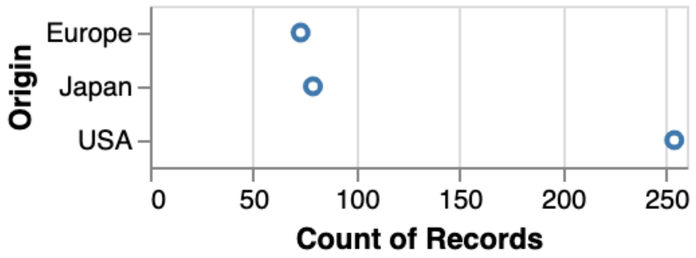
raw

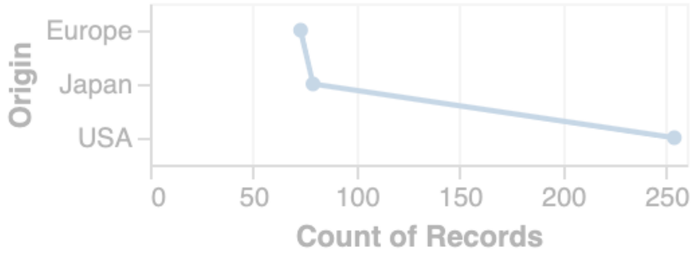


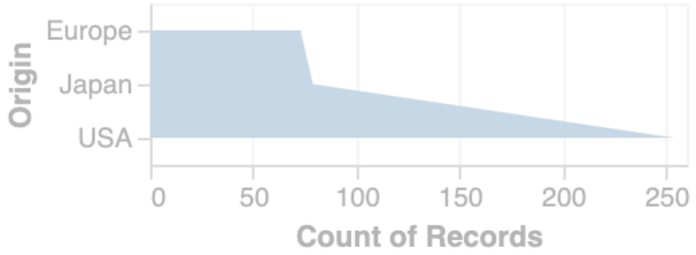


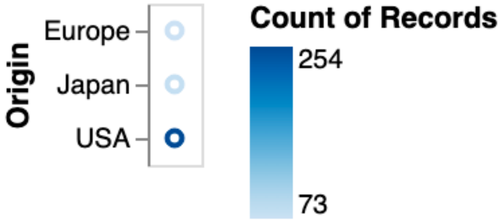


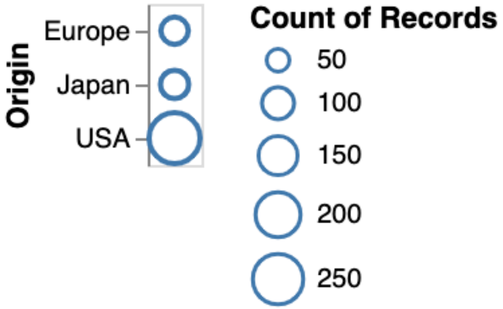
aggregate (count)

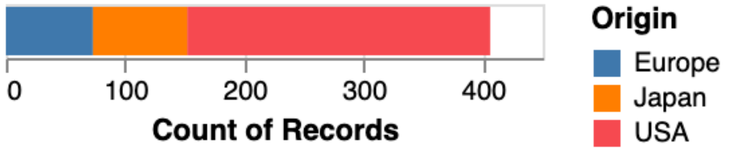


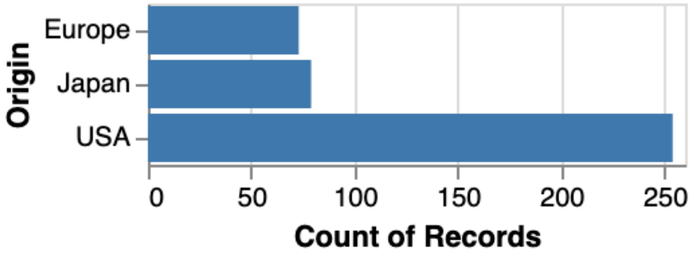








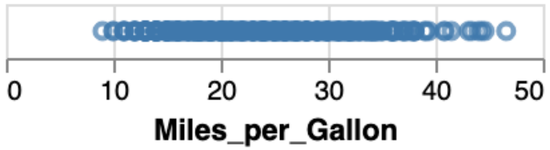


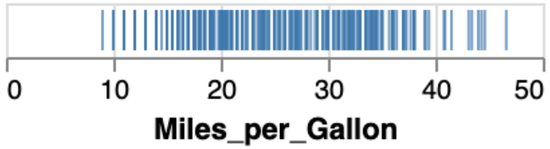


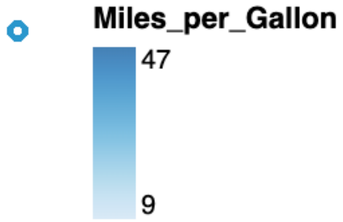
Visual Encoding = Combinatorial Design Space

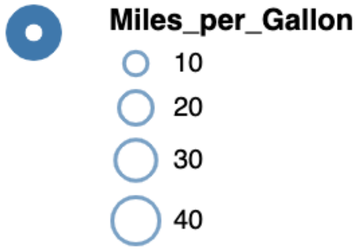
1D quantitative data (Q)

raw

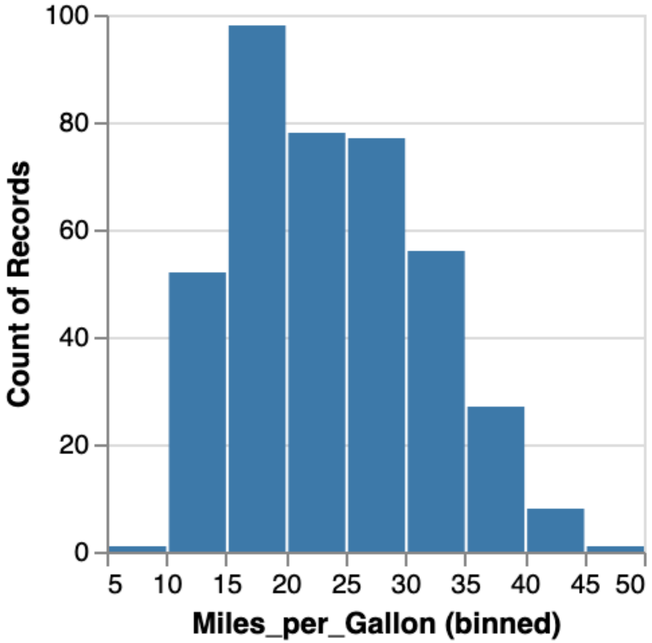


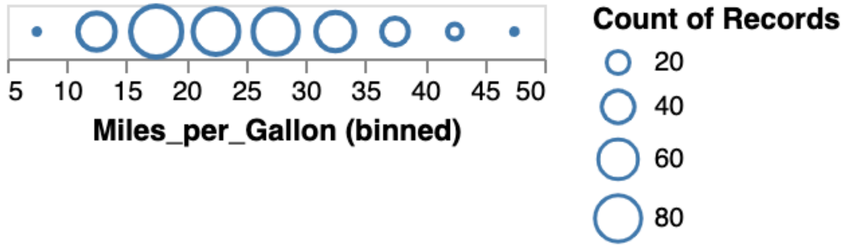


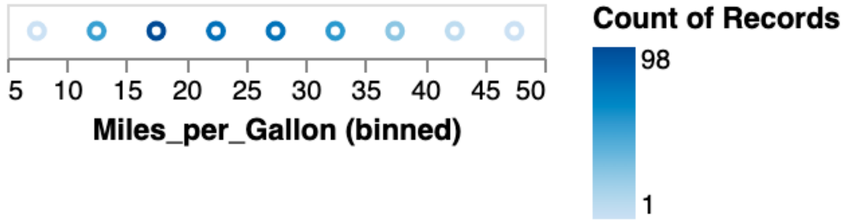




aggregate (count)





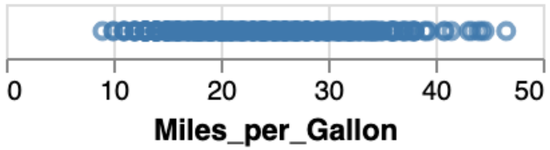


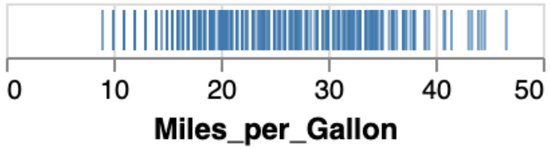
Visual Encoding = Combinatorial Design Space

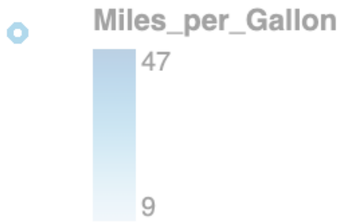
1D quantitative data (Q)

Expressive?

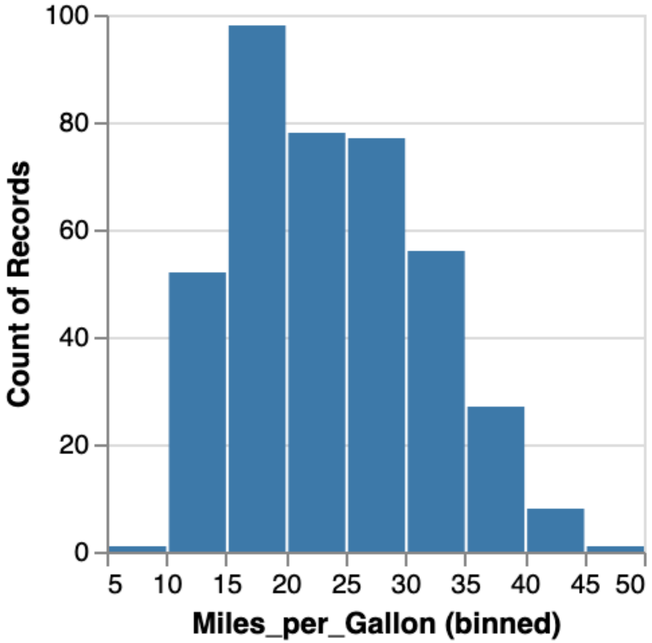
raw

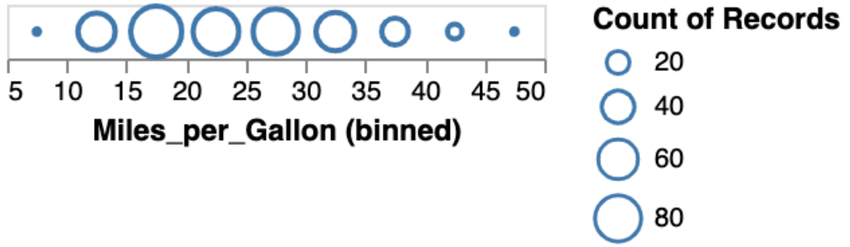


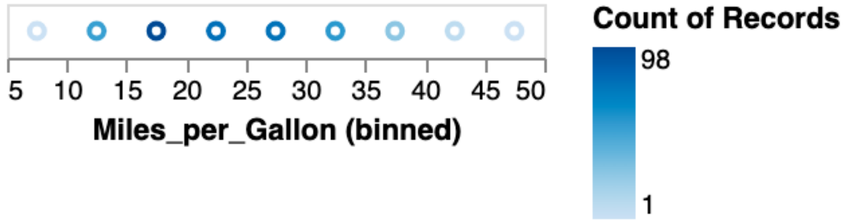




aggregate (count)



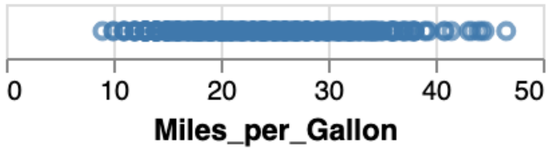




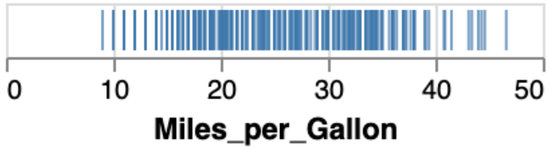
Visual Encoding = Combinatorial Design Space

1D quantitative data (Q) | Expressive? | **Effective?**

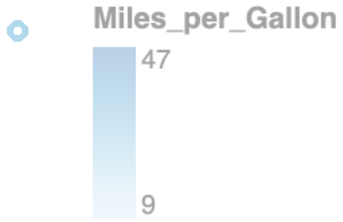
raw



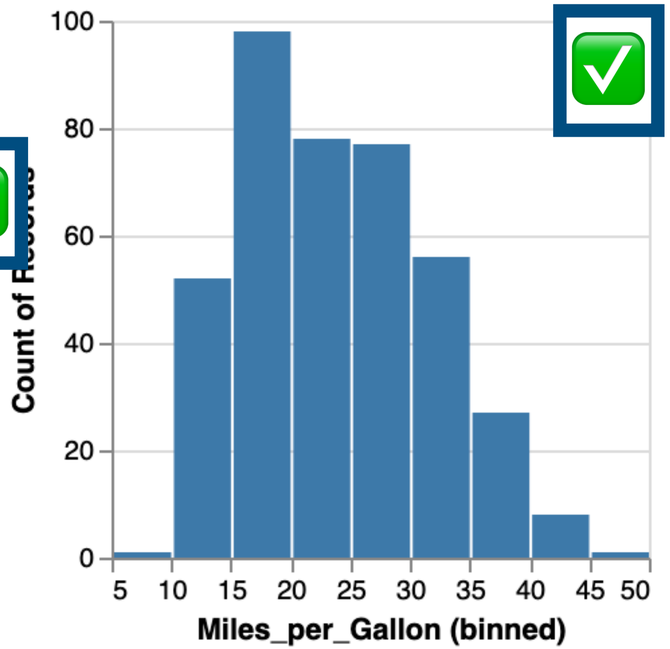
?



✅

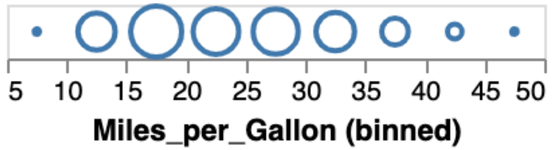


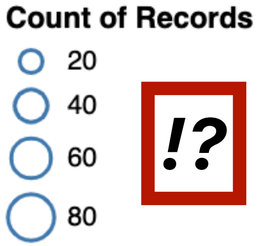
aggregate (count)



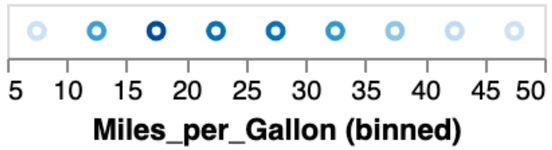
✅

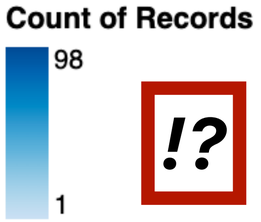
https://vega.github.io/vega/examples/histogram/





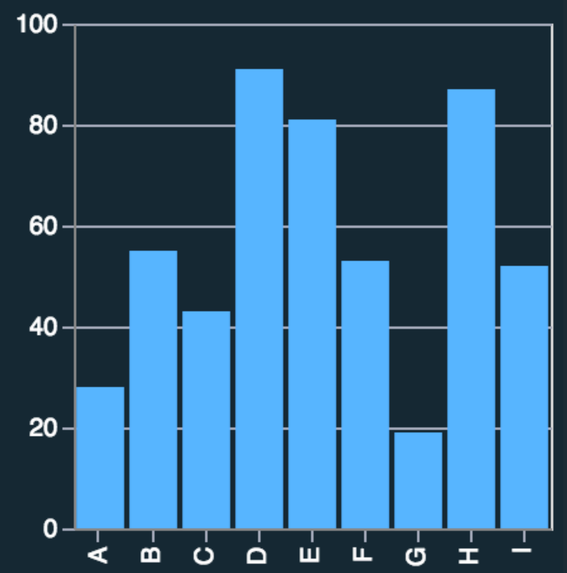
!?





!?

Visual Encoding: Nimble Design Moves

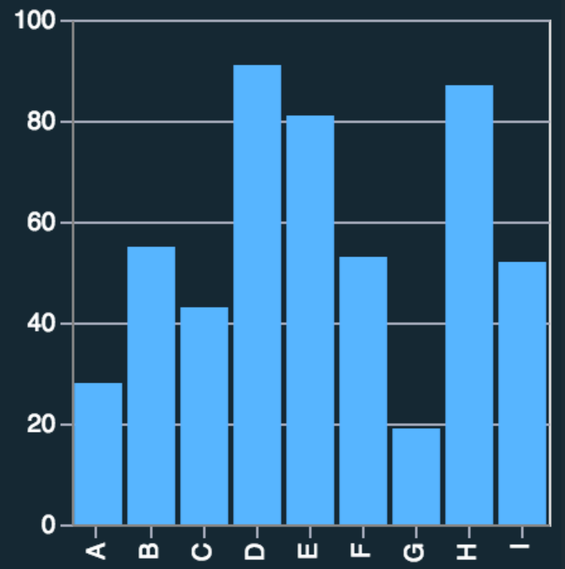


Mark: Bar

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

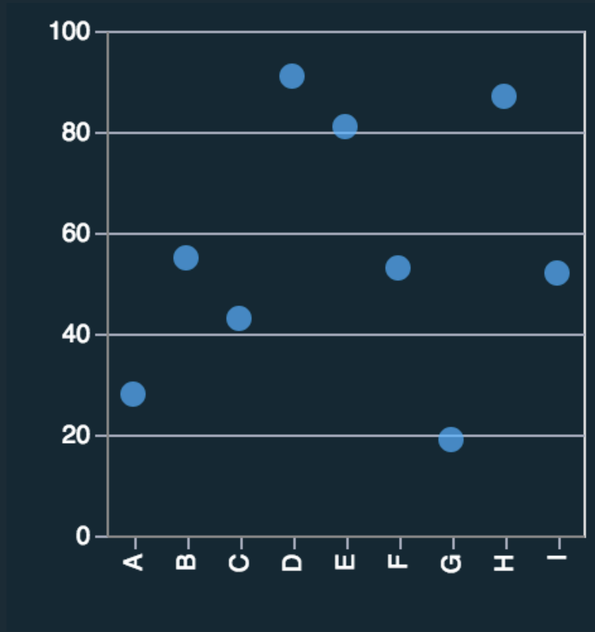
Visual Encoding: Nimble Design Moves



Mark: Bar

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

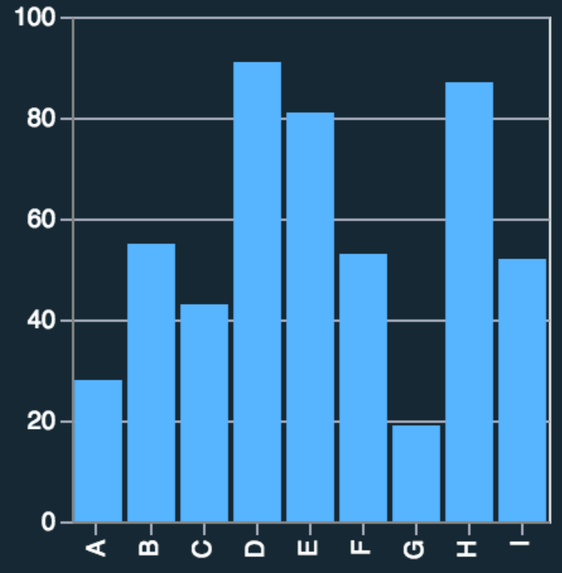


Mark: Point

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

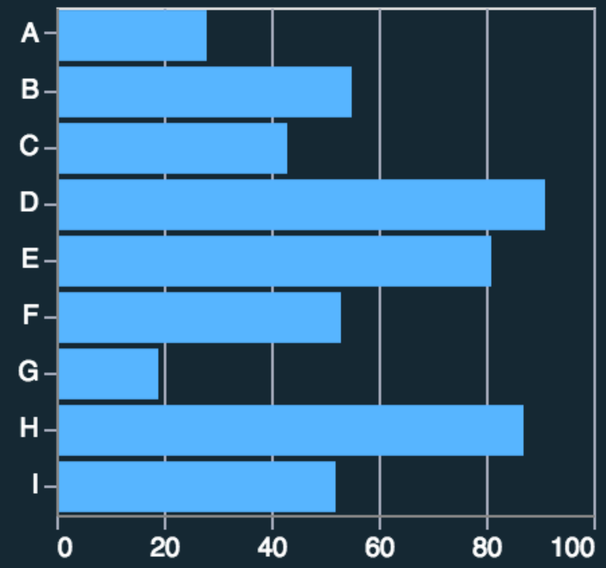
Visual Encoding: Nimble Design Moves



Mark: Bar

$d_{nominal} \rightarrow x$

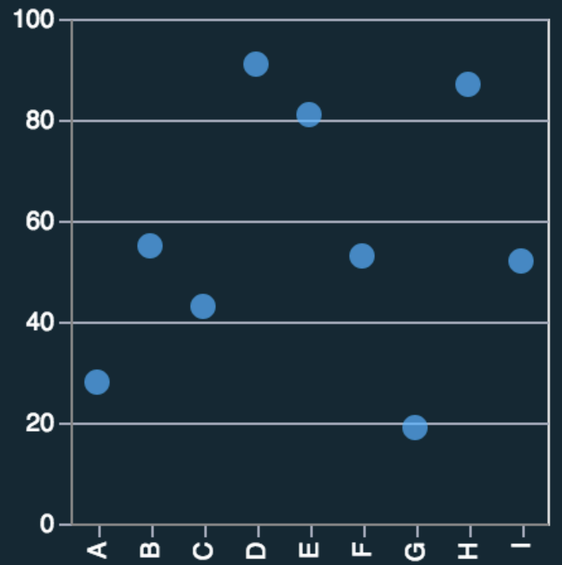
$d_{quantitative} \rightarrow y$



Mark: Bar

$d_{nominal} \rightarrow y$

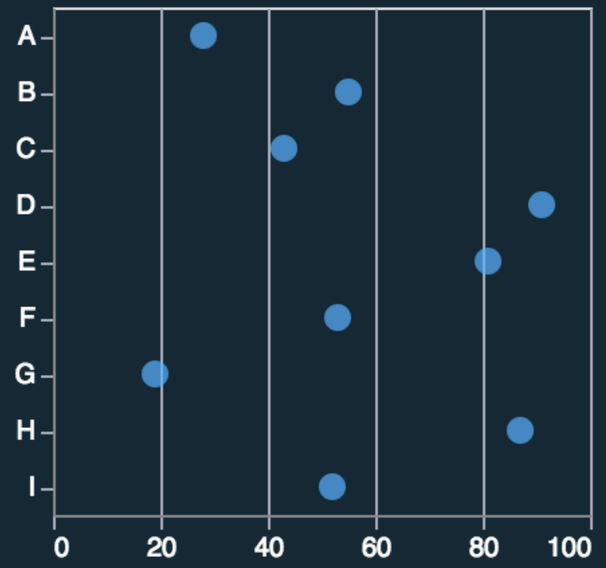
$d_{quantitative} \rightarrow x$



Mark: Point

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

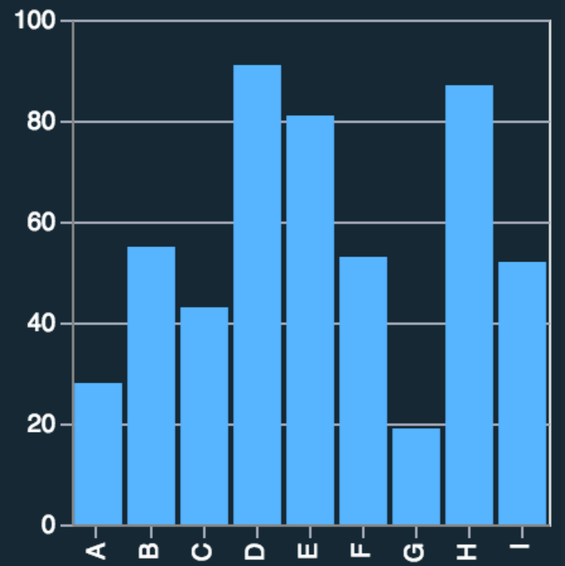


Mark: Point

$d_{nominal} \rightarrow y$

$d_{quantitative} \rightarrow x$

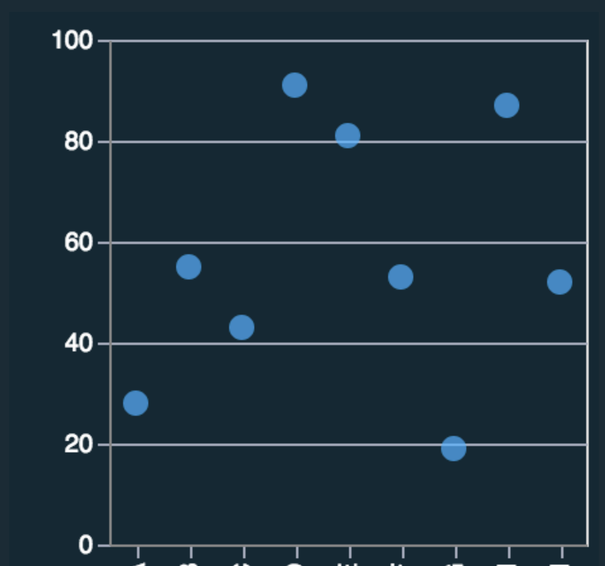
Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar

$d_{nominal} \rightarrow x$

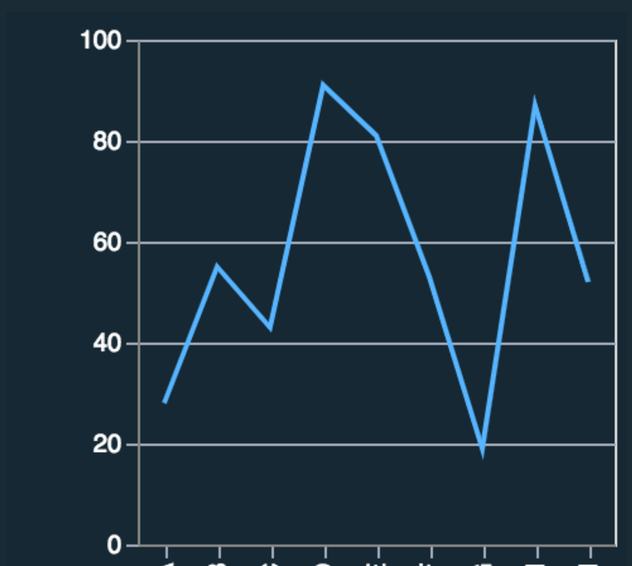
$d_{quantitative} \rightarrow y$



Mark: Point

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

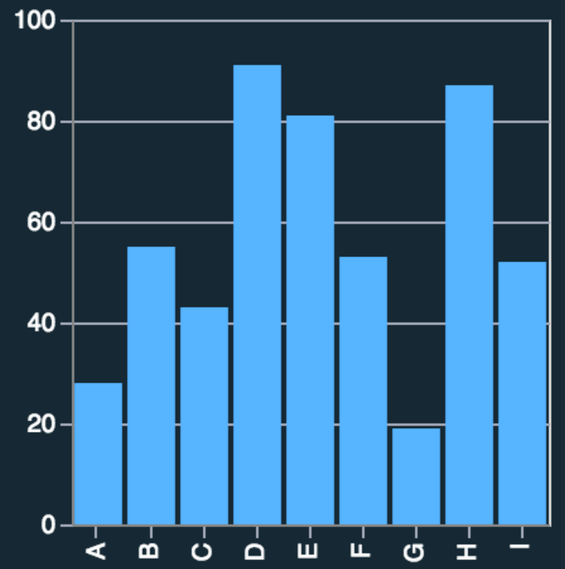


Mark: Line

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

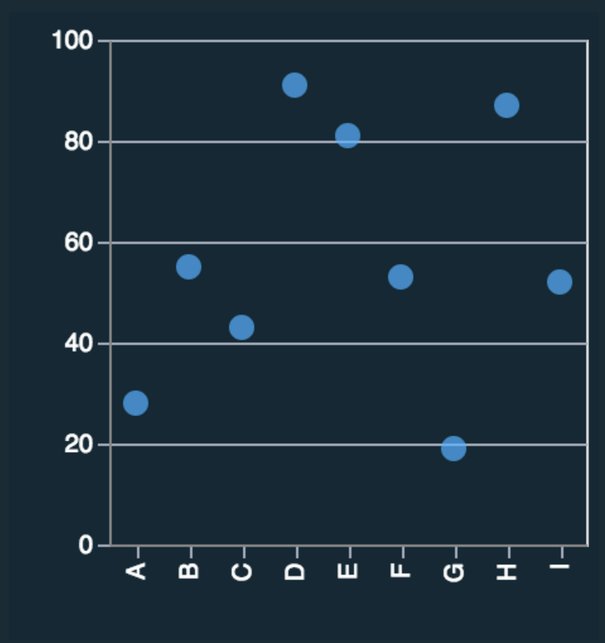
Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar

$d_{nominal} \rightarrow x$

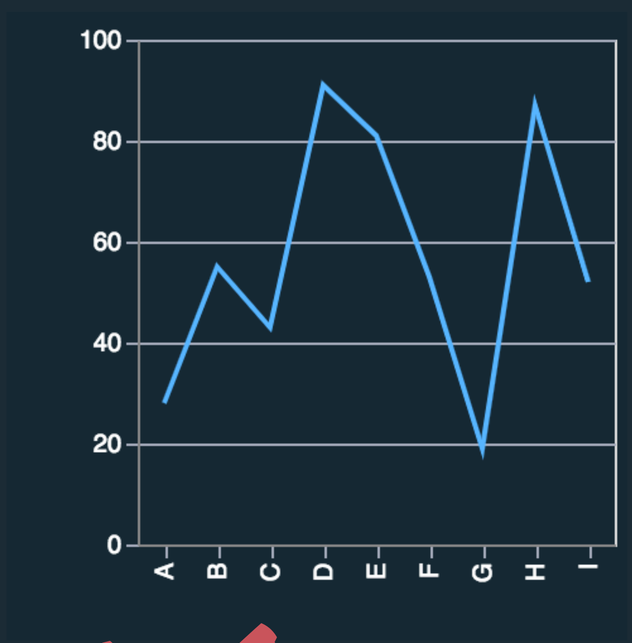
$d_{quantitative} \rightarrow y$



Mark: Point

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$



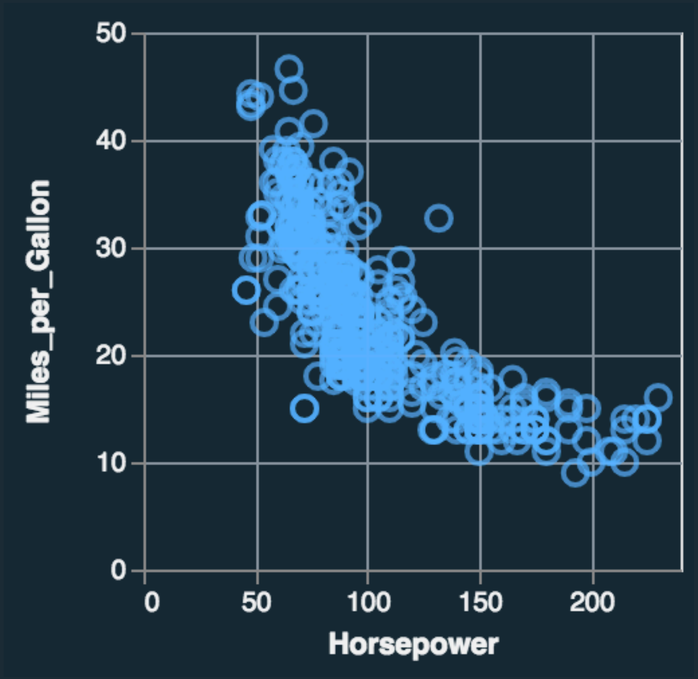
Mark: Line

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

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

Visual Encoding

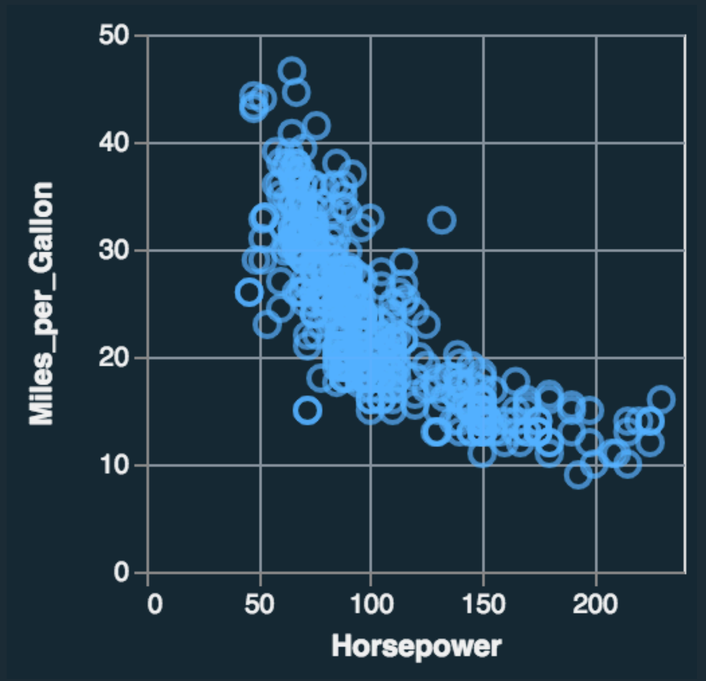


Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

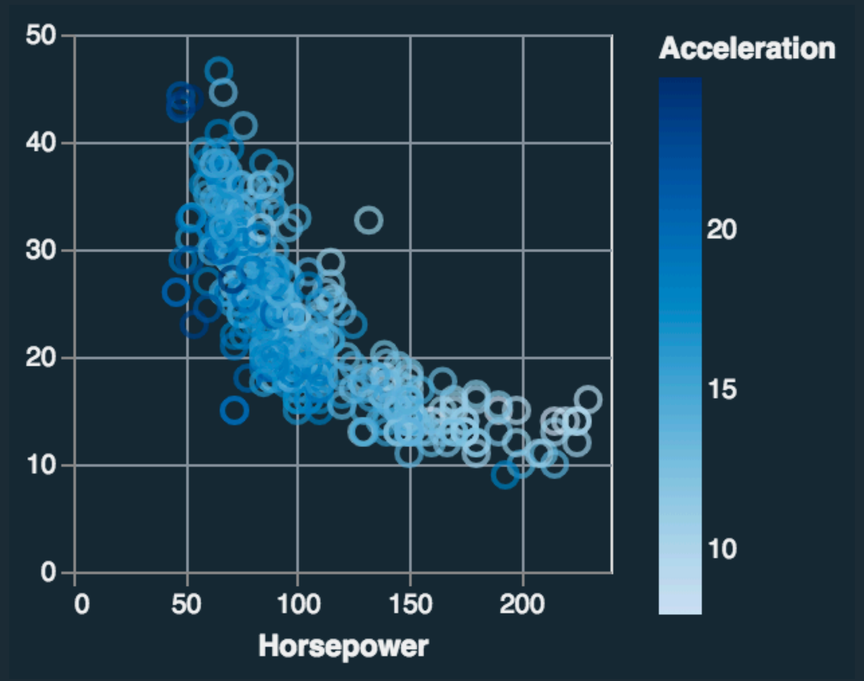
Visual Encoding



Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$



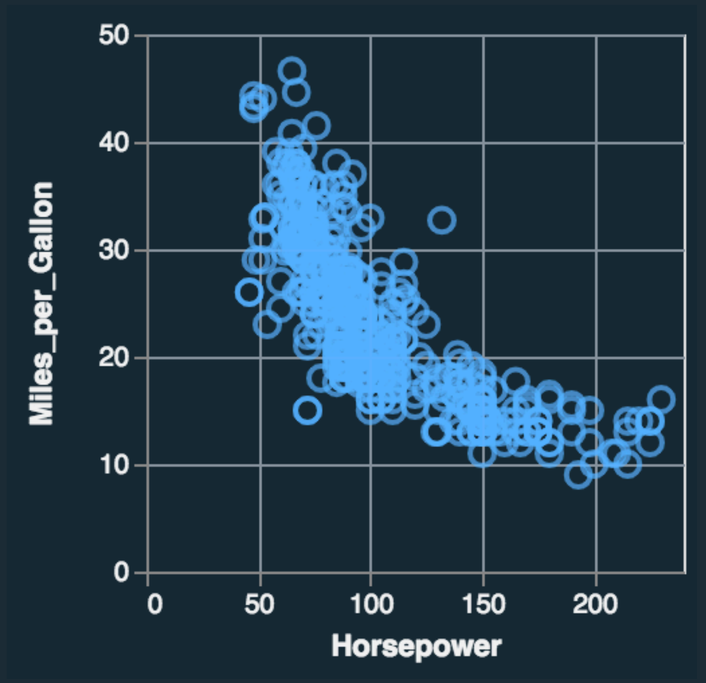
Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{quantitative} \rightarrow color$

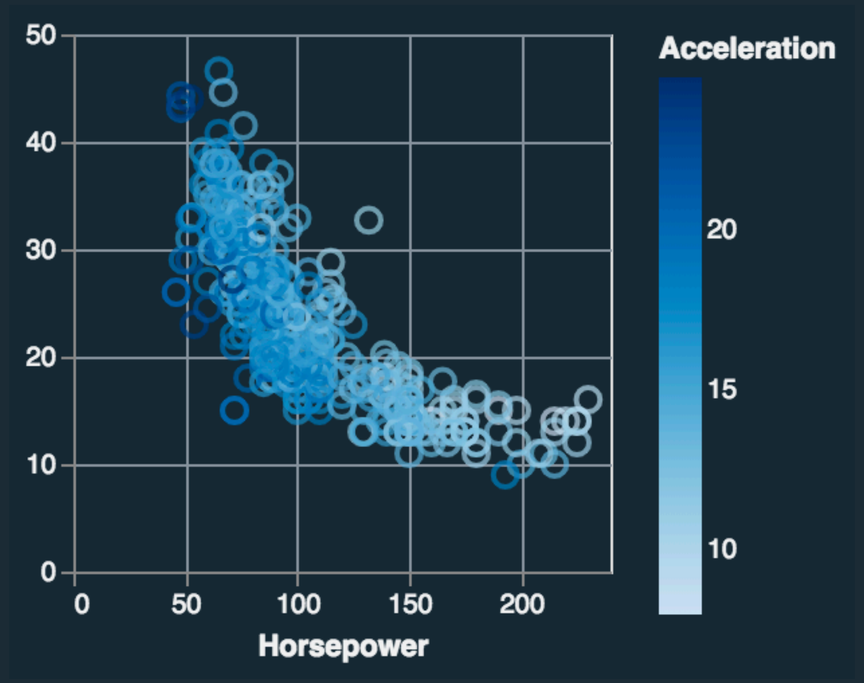
Visual Encoding



Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

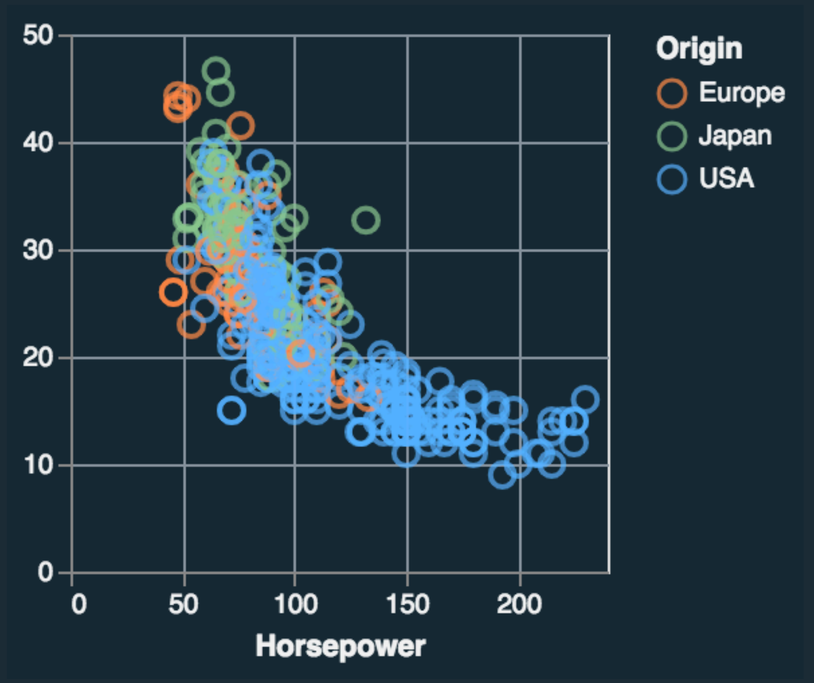


Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{quantitative} \rightarrow color$



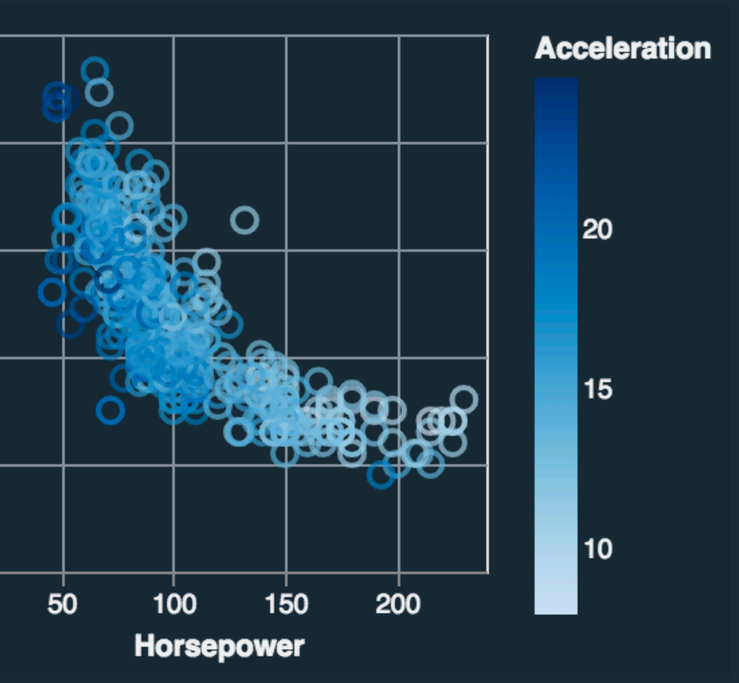
Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{nominal} \rightarrow color$

Visual Encoding

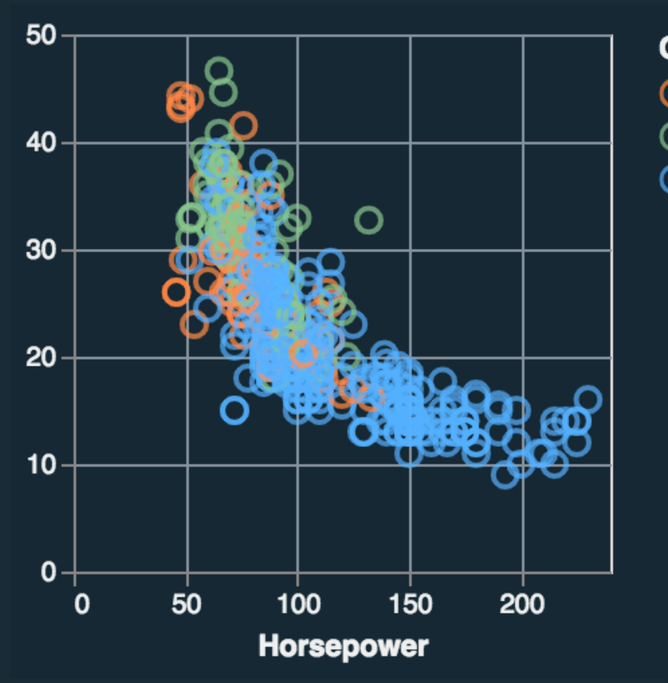


Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{quantitative} \rightarrow color$

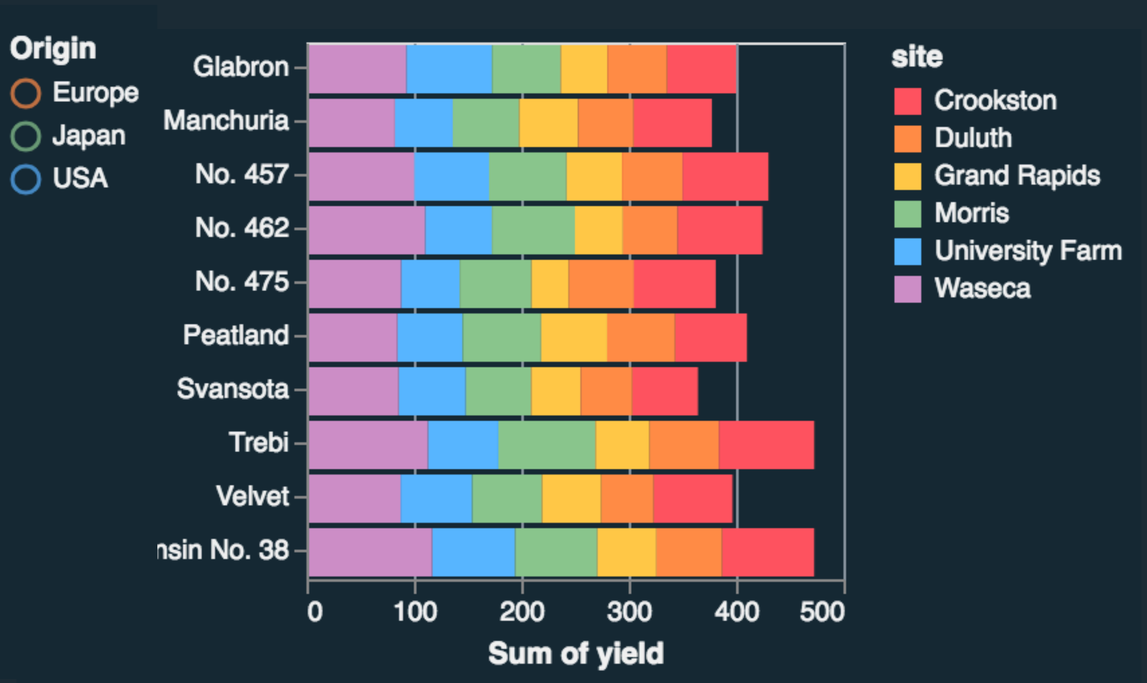


Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{nominal} \rightarrow color$



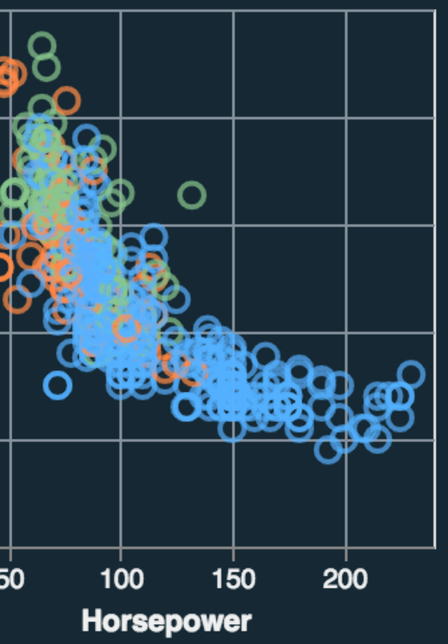
Mark: Bar

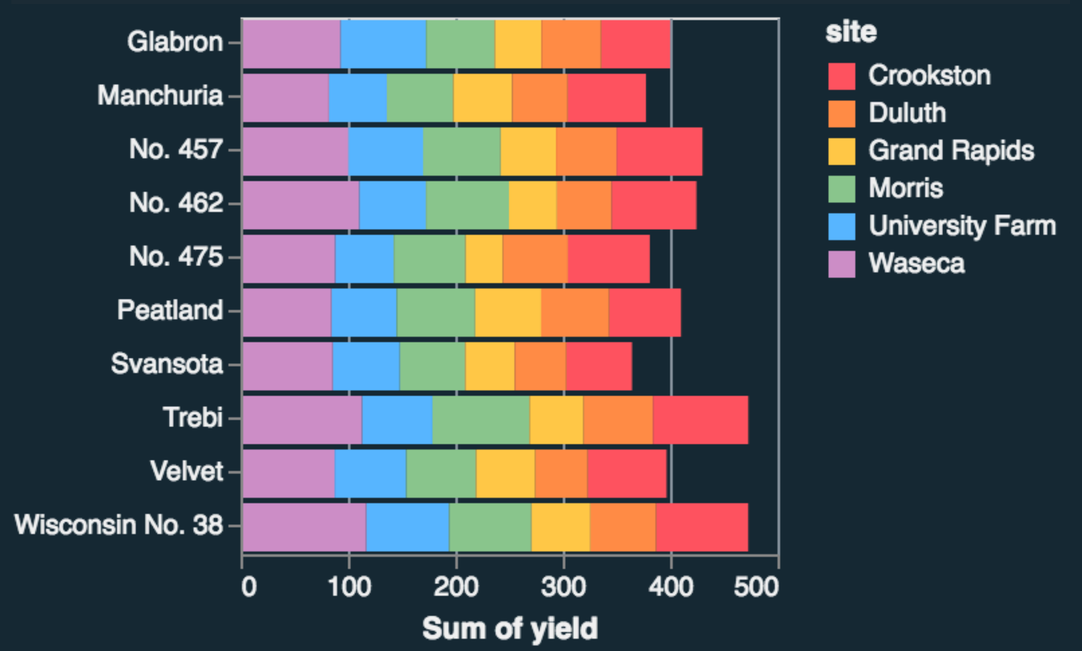
$d_{quantitative} \rightarrow x$

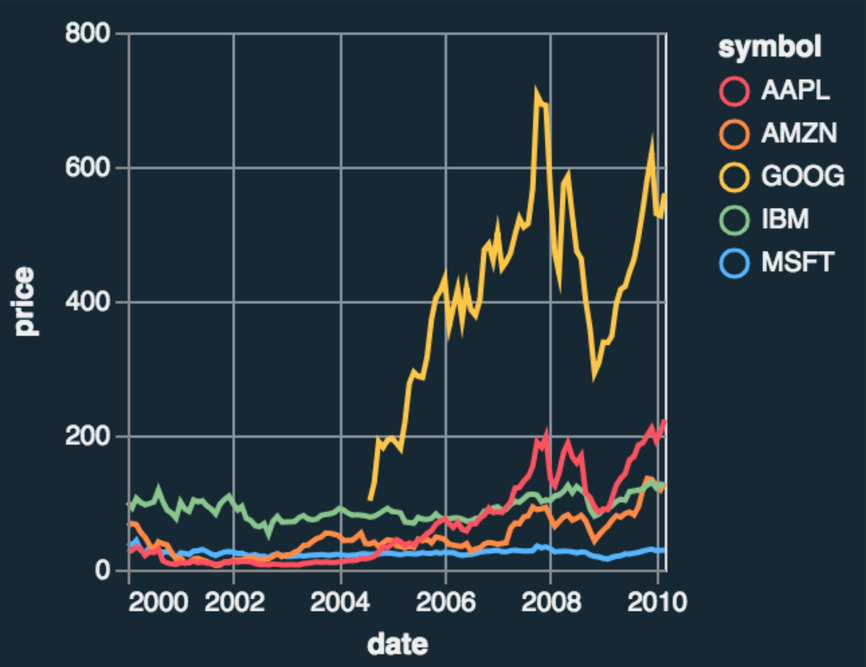
$d_{nominal} \rightarrow y$

$d_{nominal} \rightarrow color$

Visual Encoding







Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{nominal} \rightarrow color$

Mark: Bar

$d_{quantitative} \rightarrow x$

$d_{nominal} \rightarrow y$

$d_{nominal} \rightarrow color$

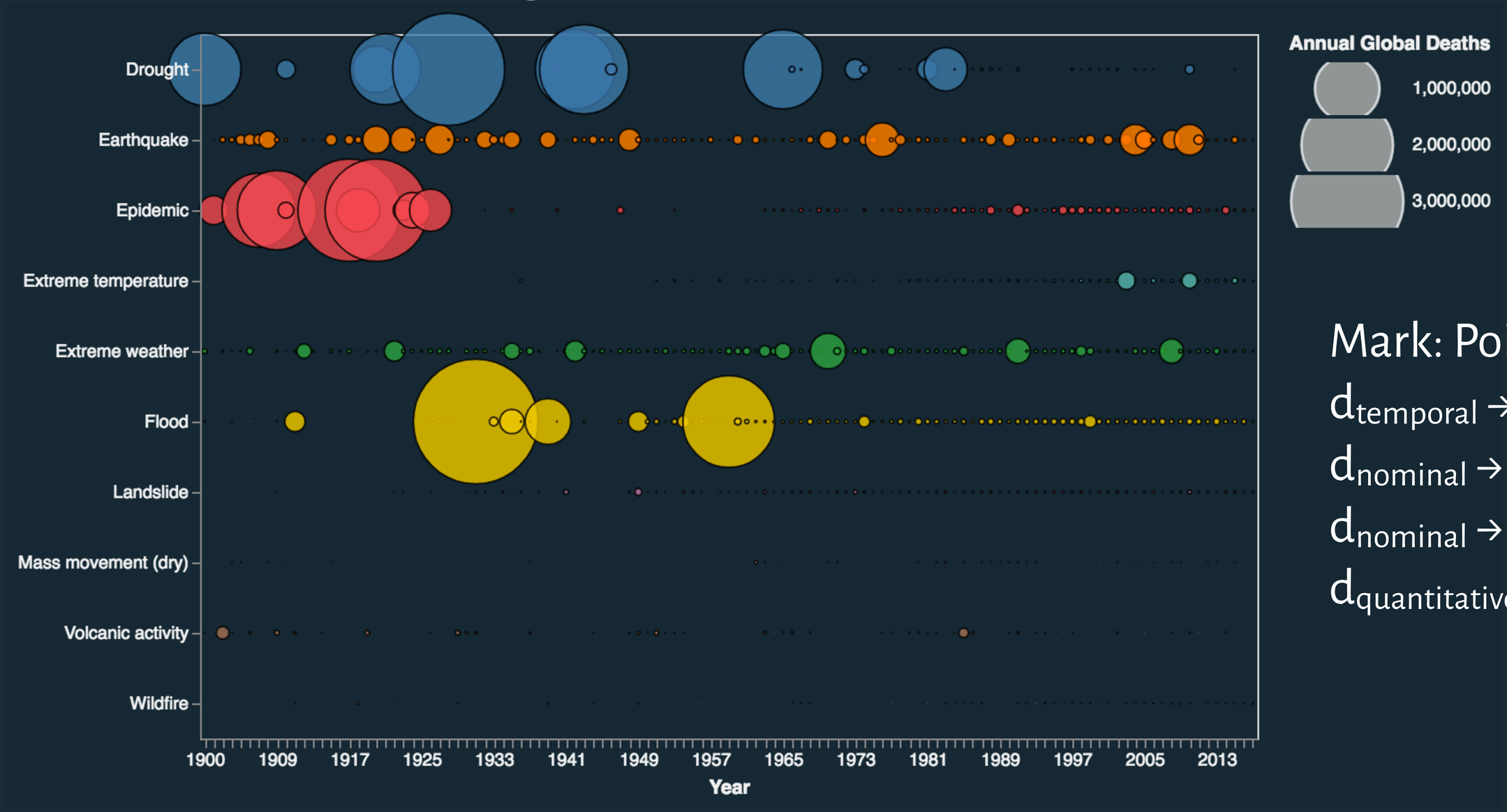
Mark: Line

$d_{temporal} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{nominal} \rightarrow color$

Visual Encoding



Mark: Point

$d_{temporal} \rightarrow x$

$d_{nominal} \rightarrow y$

$d_{nominal} \rightarrow color$

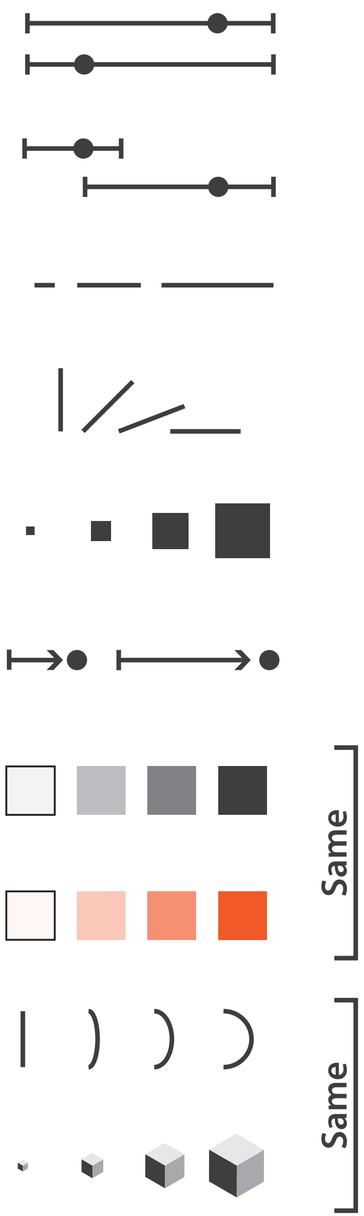
$d_{quantitative} \rightarrow size$

Effective Visual Encodings

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered 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)



Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

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

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

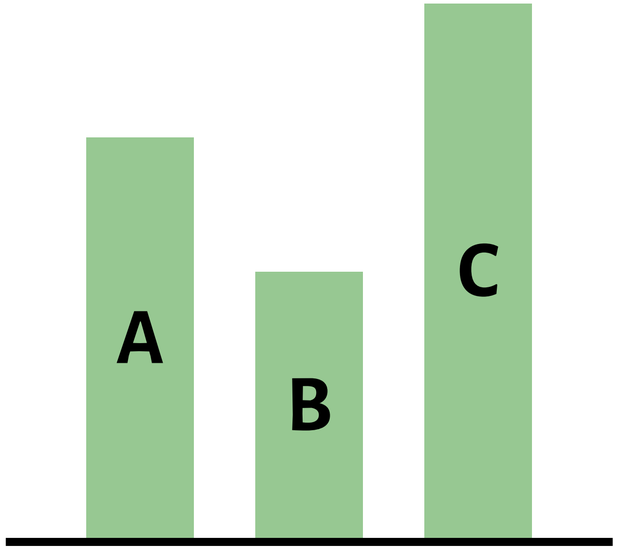
Least

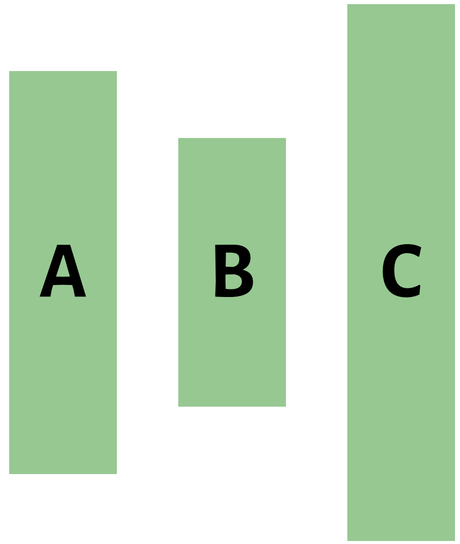
Same

Same

Perceptual work

Cognitive work





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)

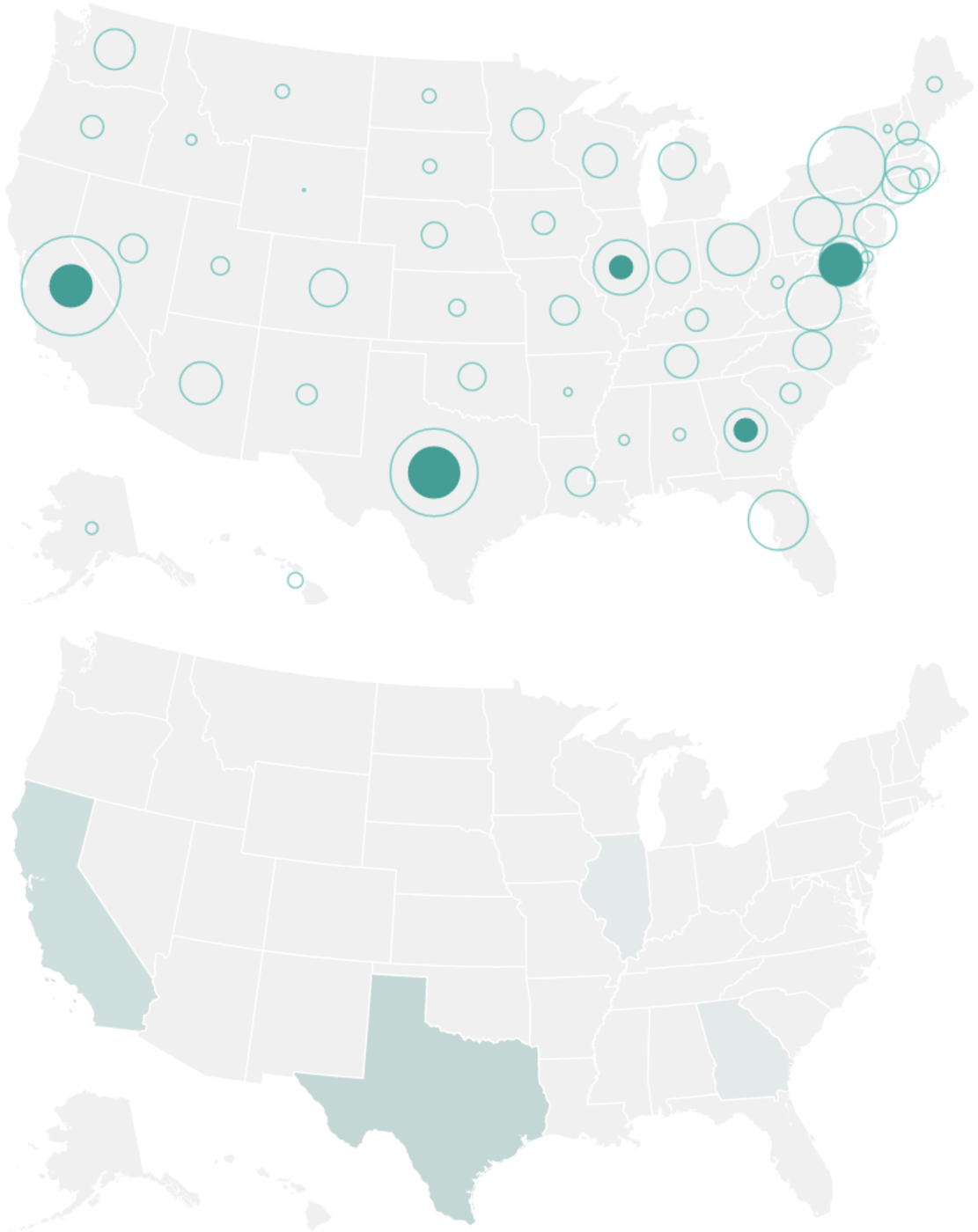
Same

Same

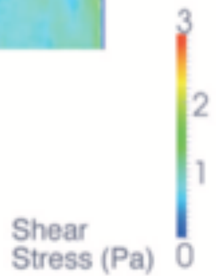
Most

Effectiveness

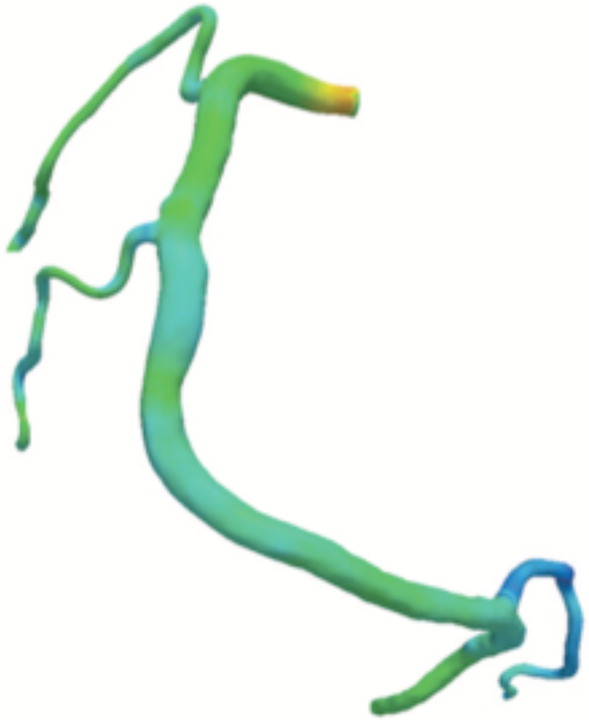
Least



Artery Visualization



Rainbow Palette

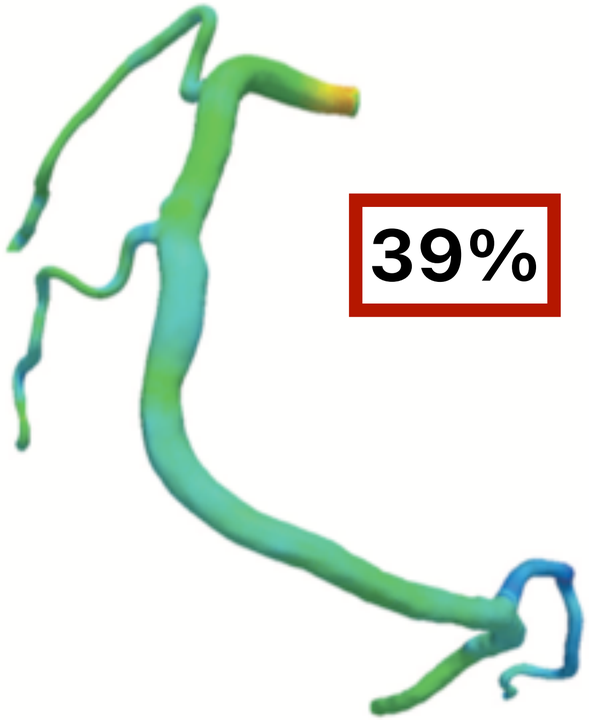


Borkin, Michelle, et al. "Evaluation of artery visualizations for heart disease diagnosis." *2011*

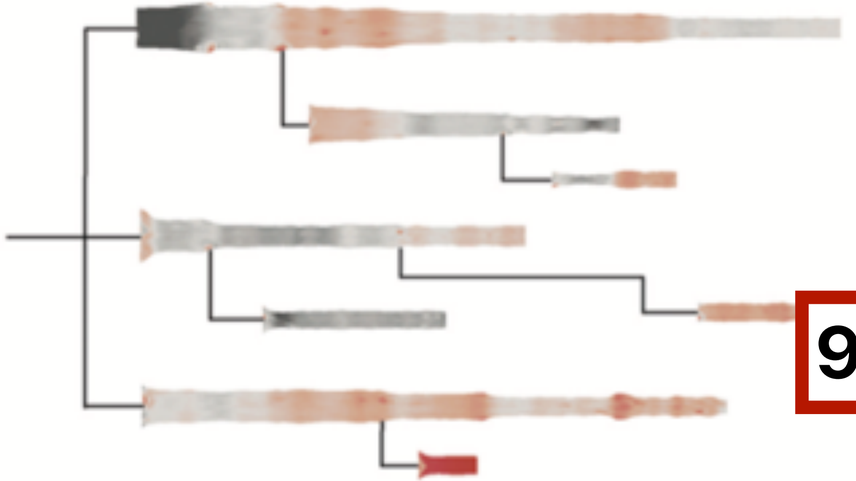
Artery Visualization

62%

Rainbow Palette

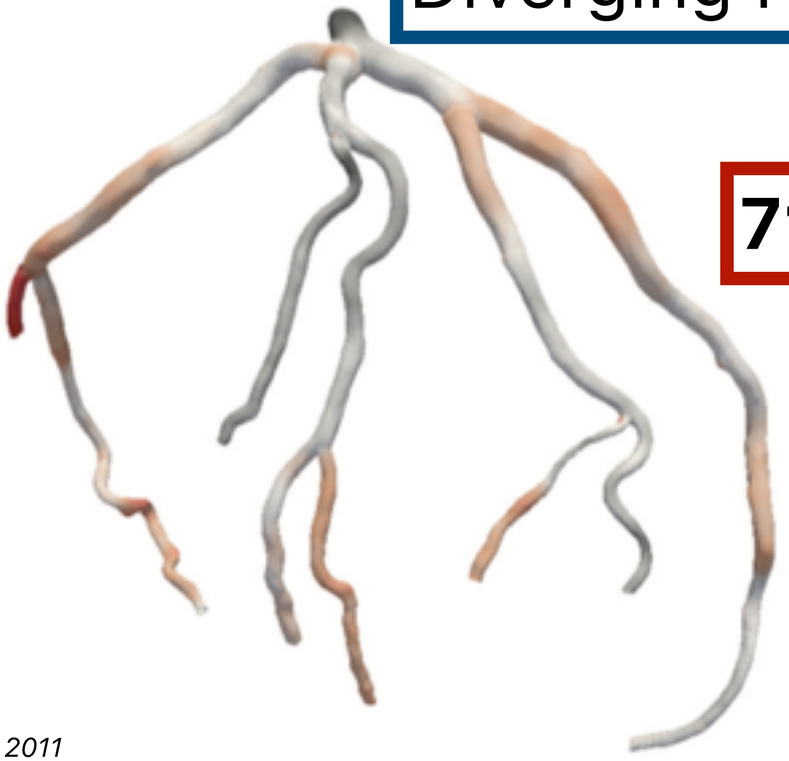


39%



92%

Diverging Palette



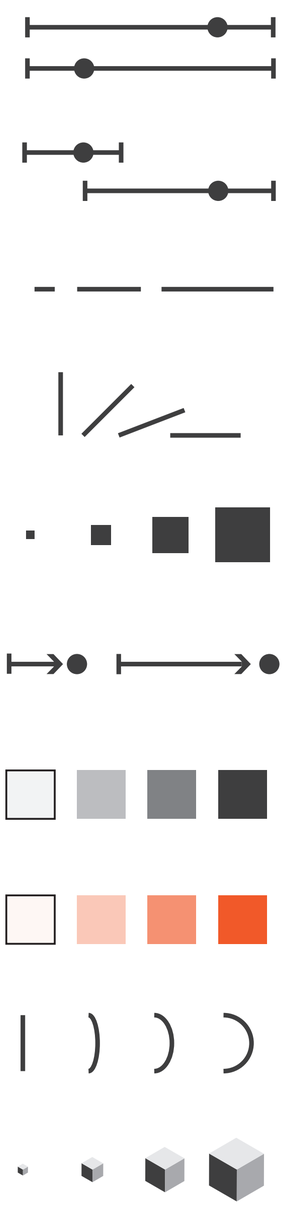
71%

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

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered 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)



Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

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

Using space (in)effectively

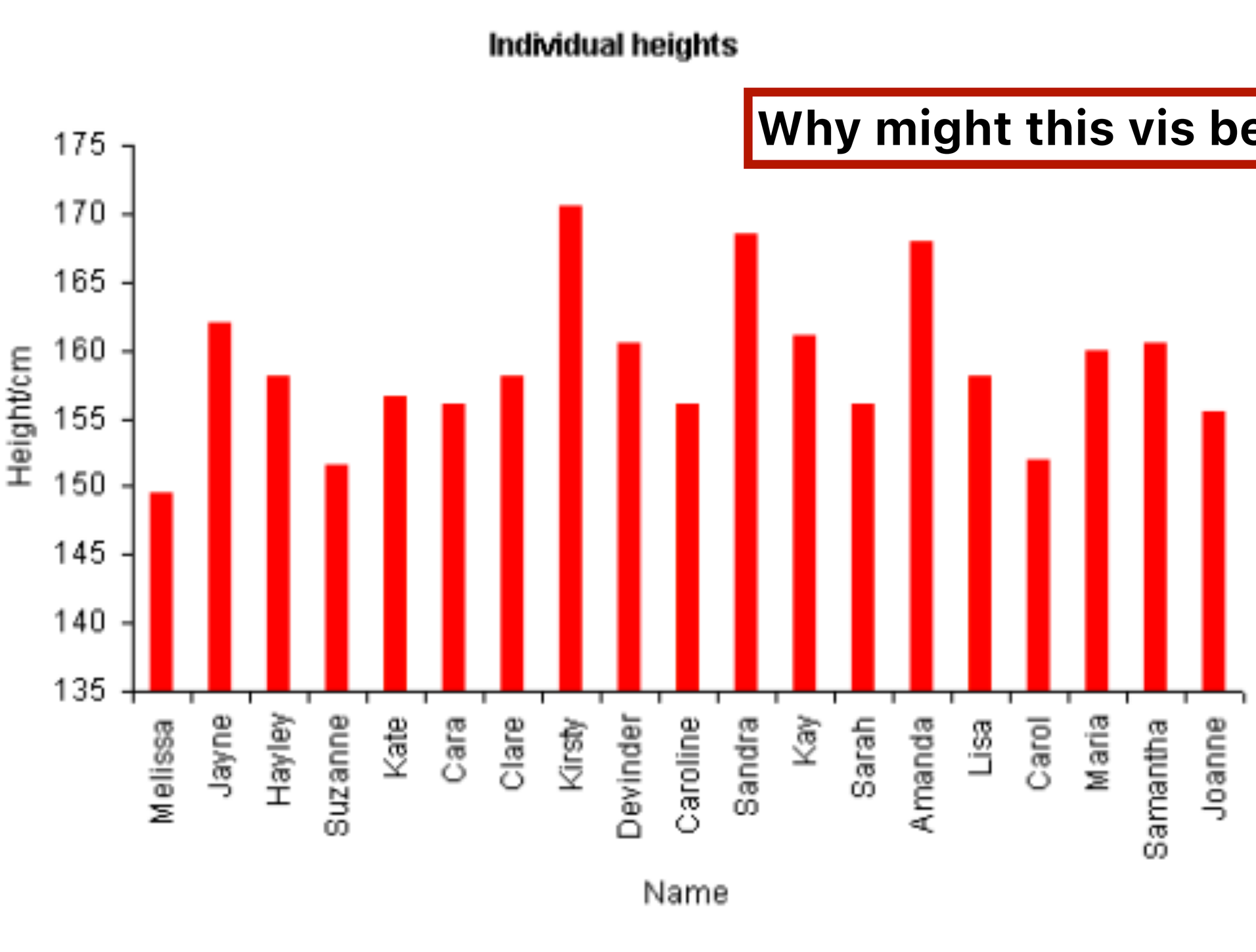
(De-)Obfuscating data

(Mis)leading the witness

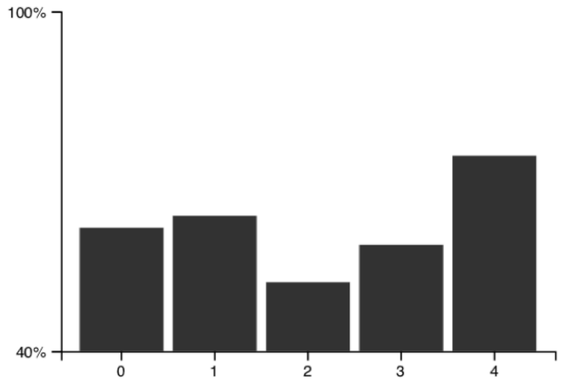
Using space (in)effectively

(De-)Obfuscating data

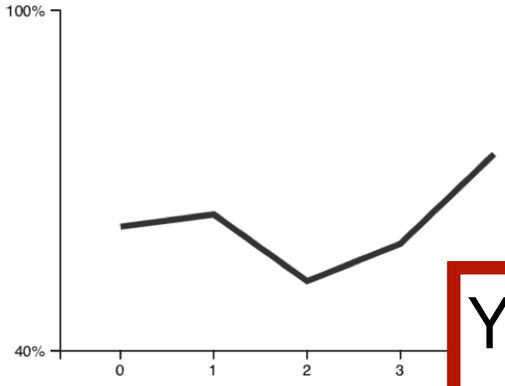
(Mis)leading the witness



Why might this vis be inexpressive?



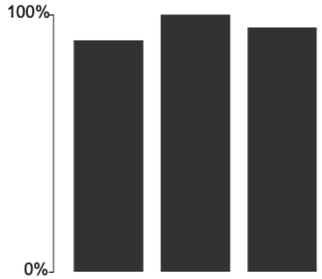
(a) Bar Chart



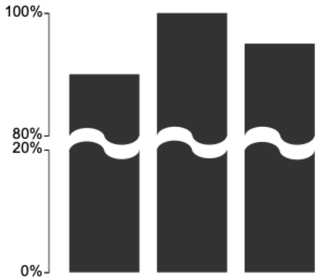
(b) Line Chart

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

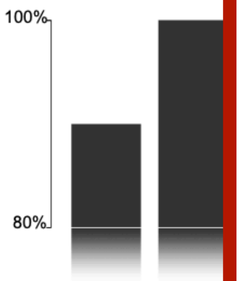
Interventions did not make a difference.



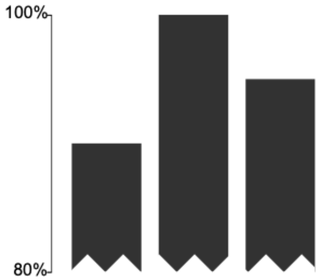
(a) Bar Chart



(b) Broken Axes

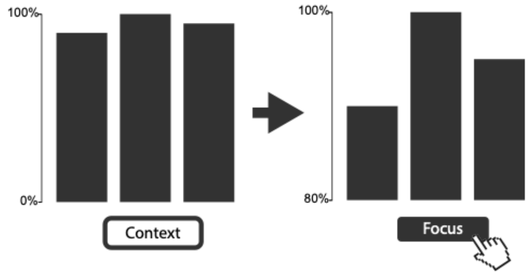


(c) Gradient Ba

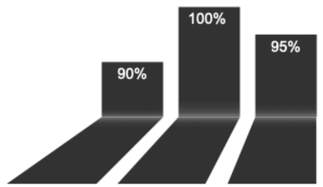


(d) Torn Paper Chart

(e) Panel Chart



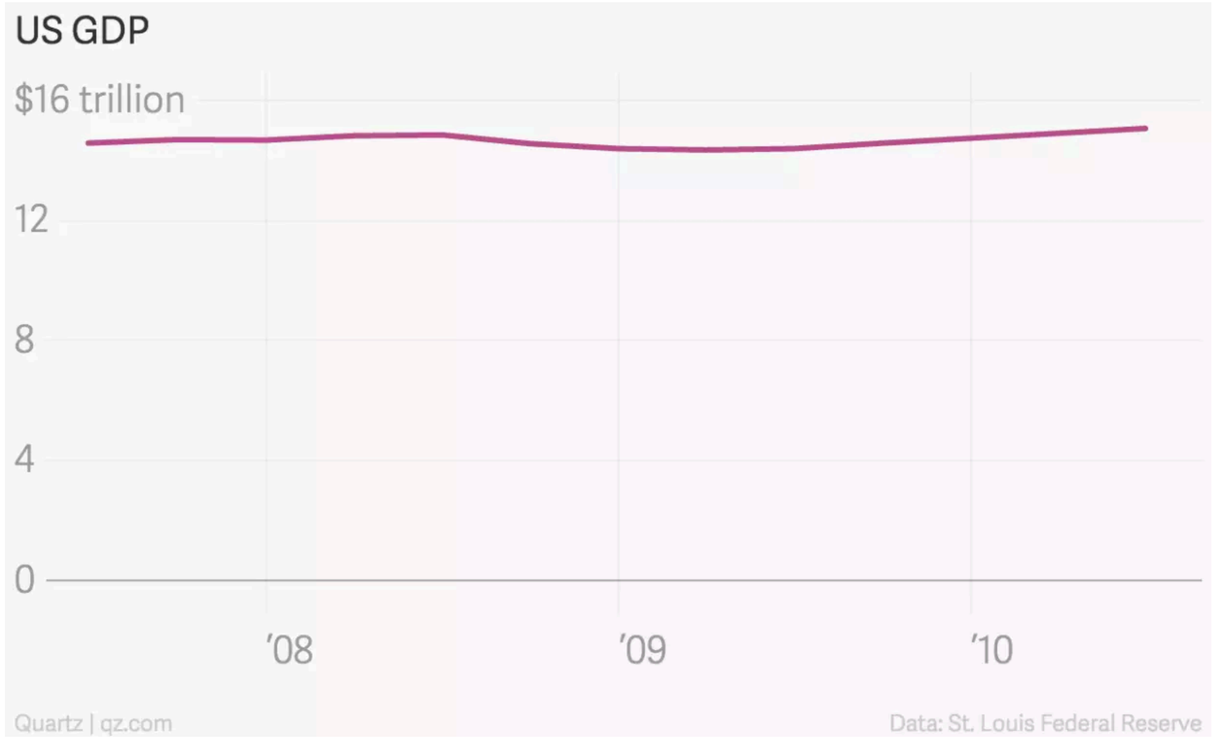
(f) Interactive Focus+Context

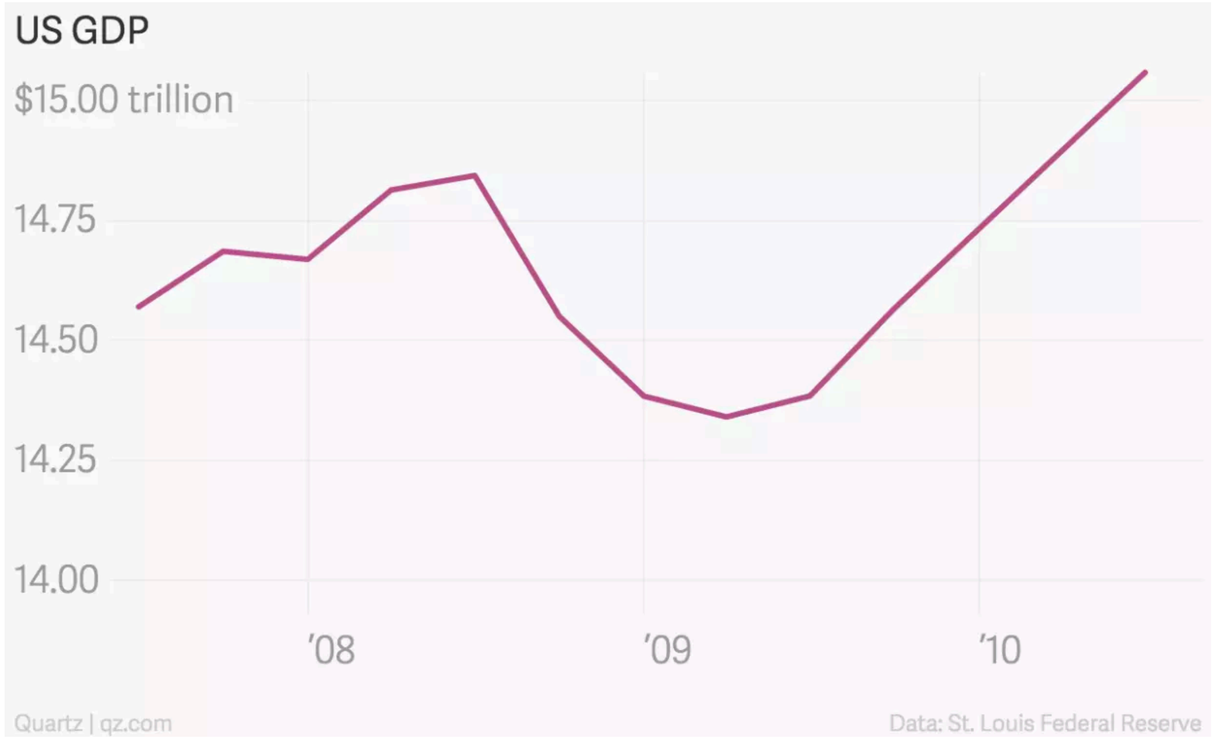


(g) Bent Bar Chart

Correll, Michael, Enrico Bertini, and Steven Franconeri. "Truncating the y-axis: Threat or menace?." *CHI 2020*.

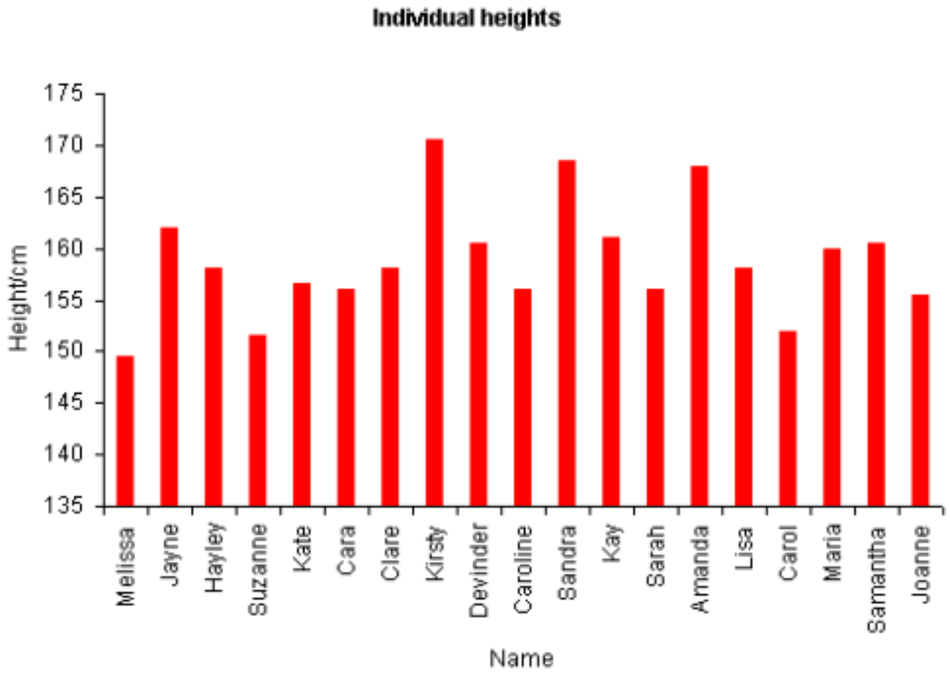
Always start at zero?



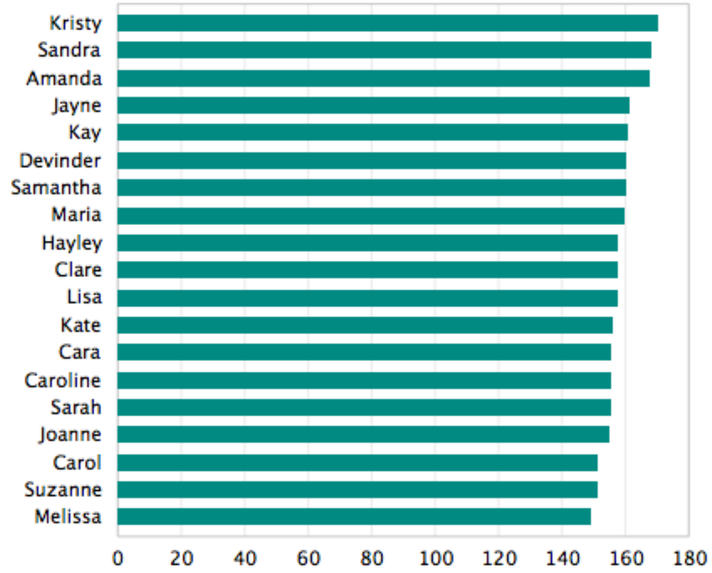


https://qz.com/418083/its-ok-not-to-start-your-y-axis-at-zero

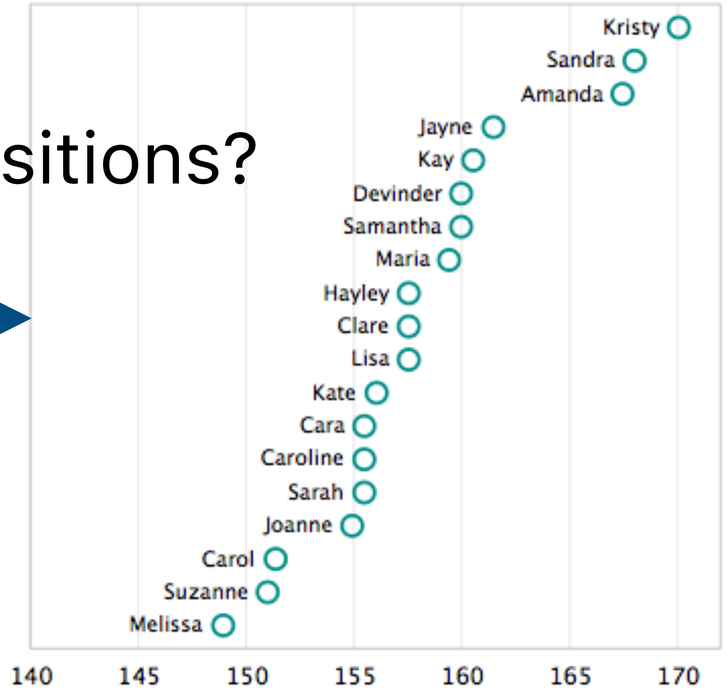
Truncating the y-axis?



Compare proportions?
(Q-ratio)



Compare relative positions?
(Q-interval)



Truncating the y-axis?

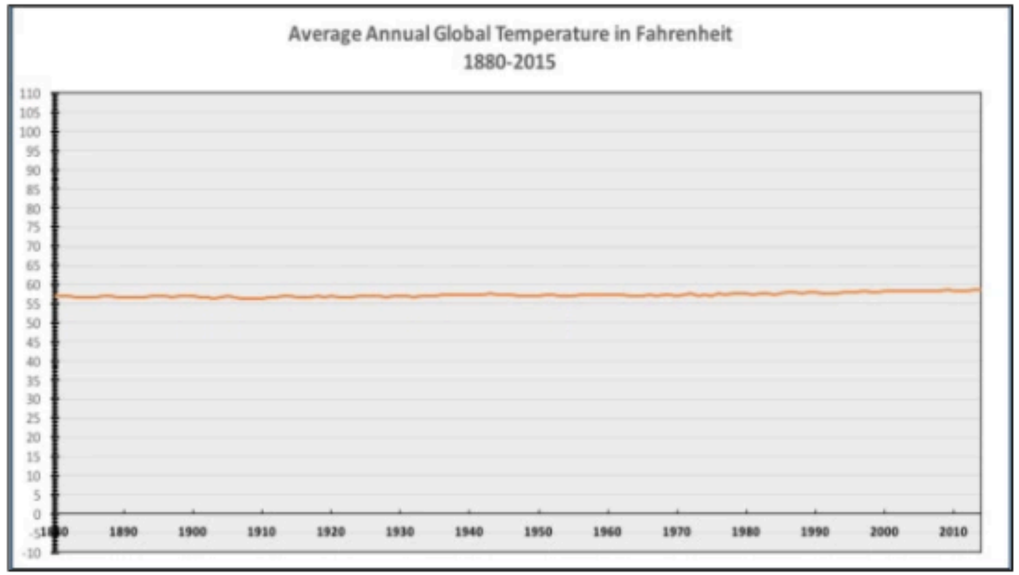
To emphasize Q-interval (vs. Q-ratio)
If the zero value doesn't make much sense.
If it is the norm (e.g., stock charts).

National Review ✓
@NRO

Follow

The only #climatechange chart you need to see. natl.re/wPKpro

(h/t @powerlineUS)

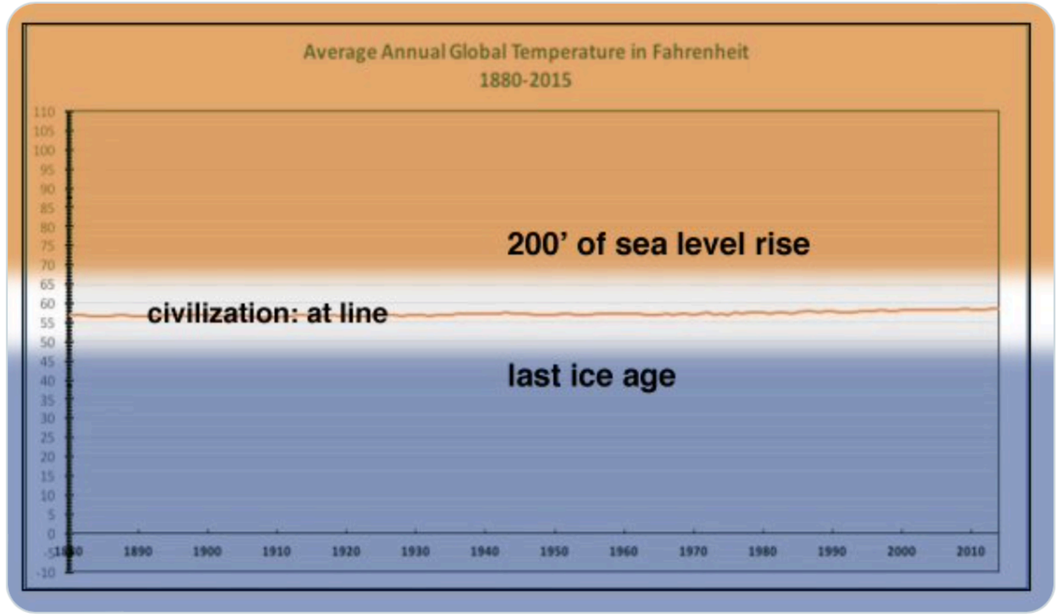


12:36 PM - 14 Dec 2015

City Atlas
@cityatlas

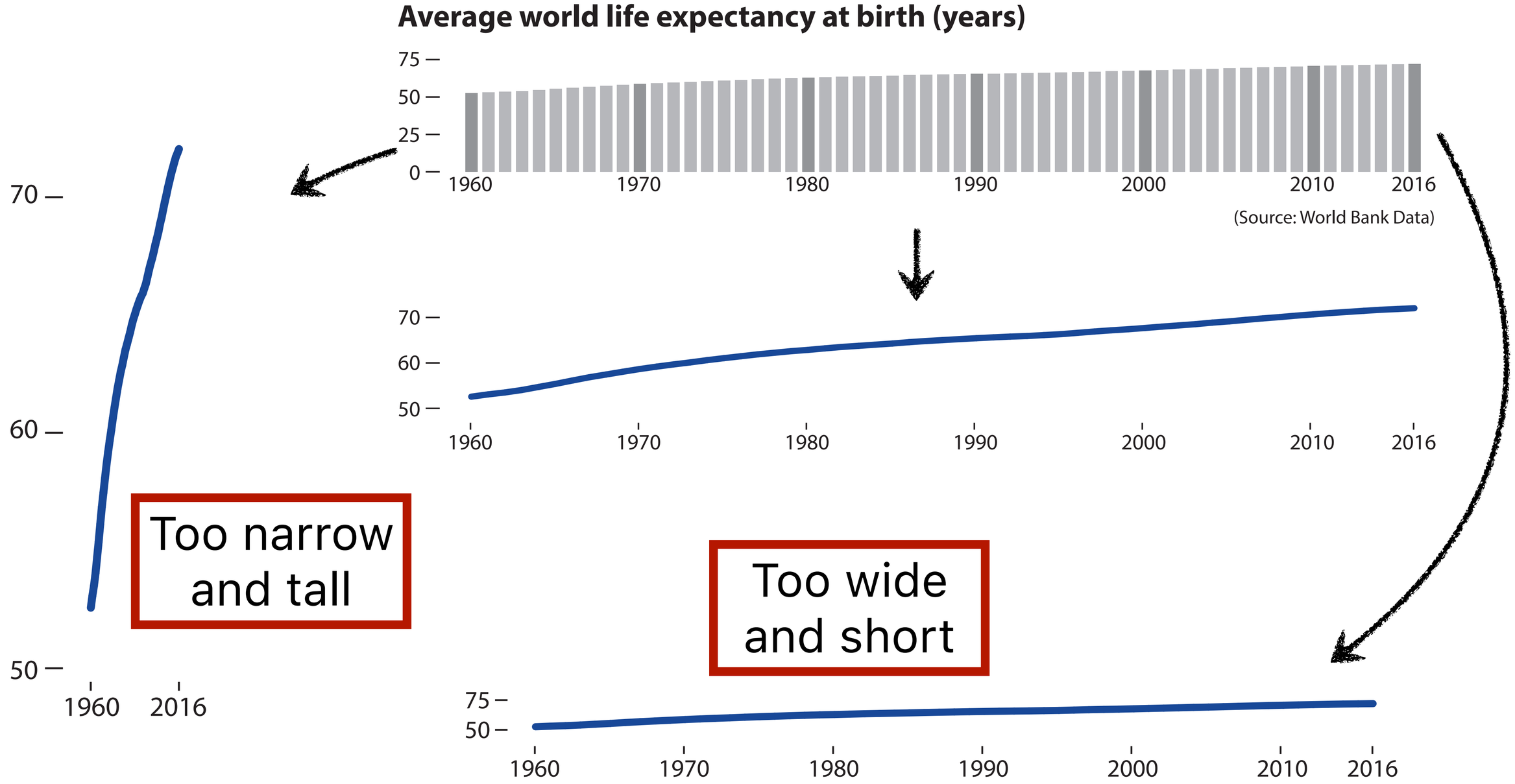
Replying to @NRO

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



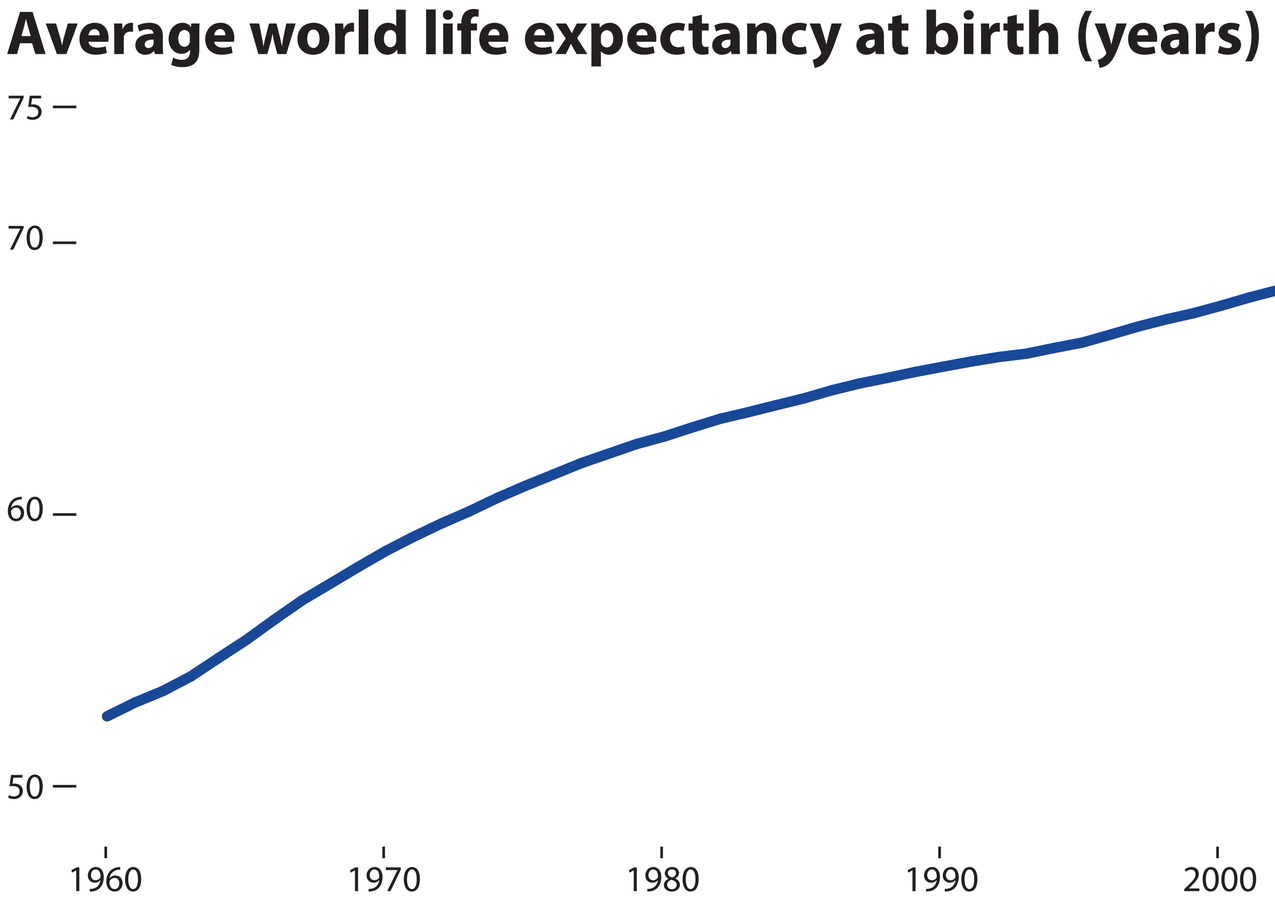
5:28 PM · Dec 14, 2015

78 Retweets **1** Quote Tweet **208** Likes



Cairo, Alberto. How charts lie: Getting smarter about visual information. 2019

Aspect Ratio

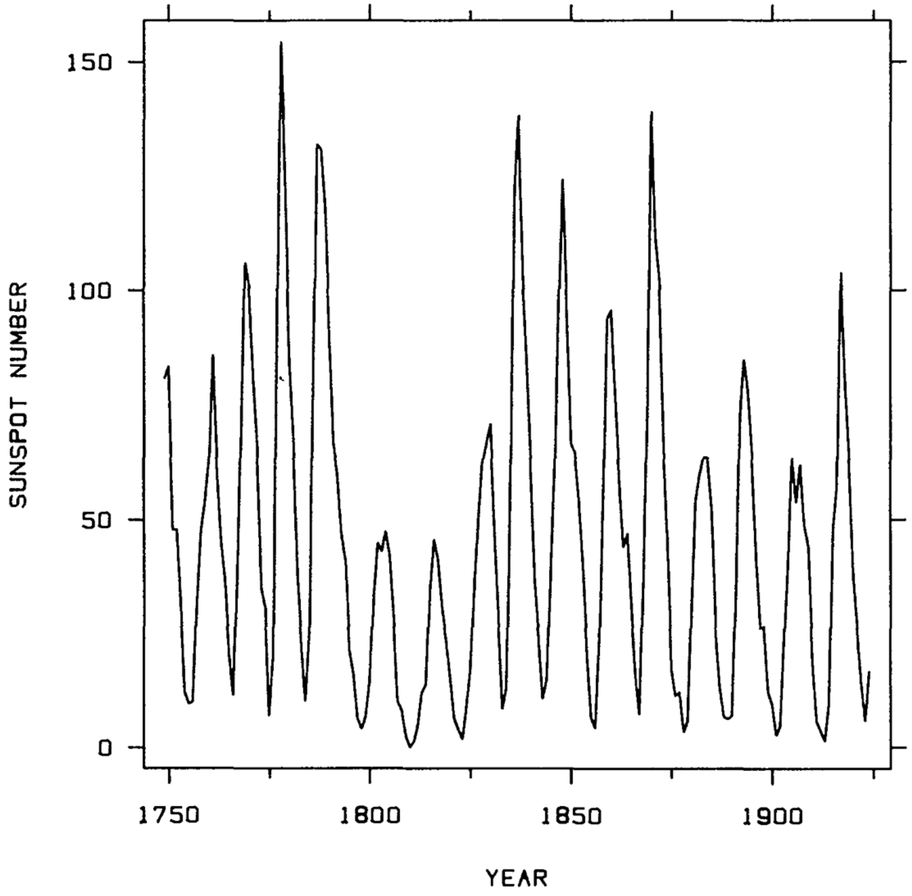


Approximate the proportion of the chart to match the depicted trend.

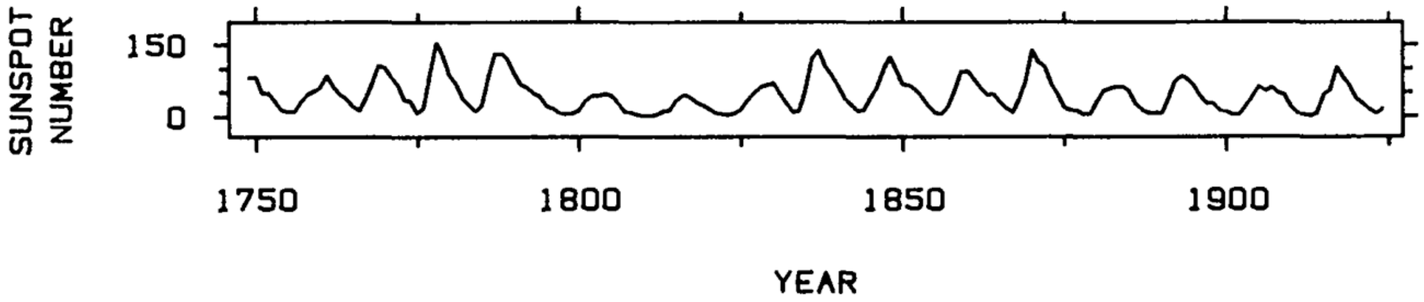
35% increase ≈ 1/3rd
≈ 4:3 aspect ratio

Cairo, Alberto. How charts lie: Getting smarter about visual information. 2019

Aspect Ratio

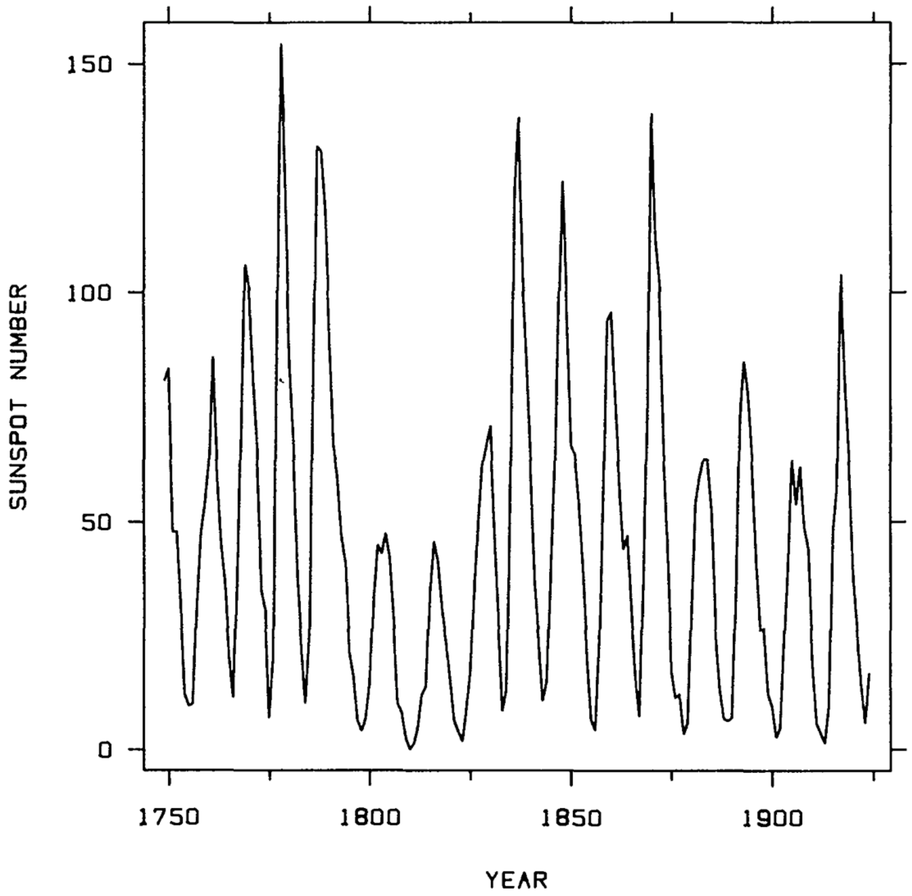


1. Approximate the proportion of the chart to match the depicted trend.

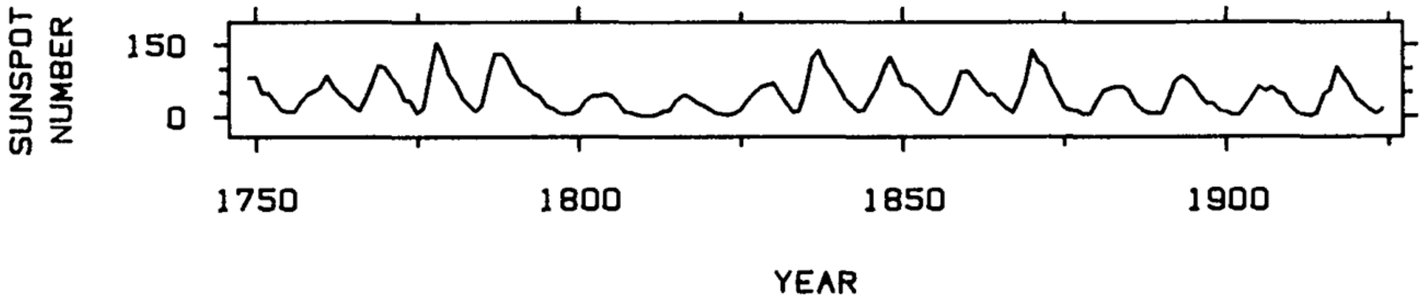


Cleveland, William S., and Robert McGill. "Graphical perception: The visual decoding of quantitative information on graphical displays of data." *1987*

Aspect Ratio



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

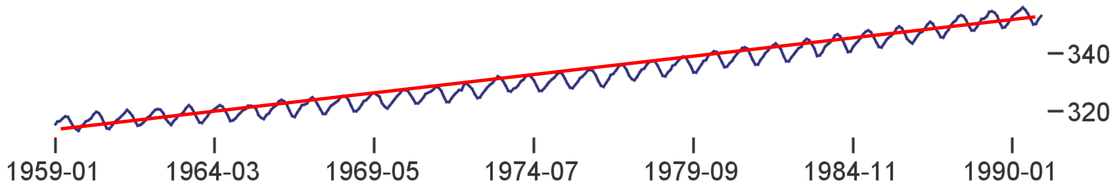


2. Bank to 45°: aspect ratios with 45° avg. line segment orientation.

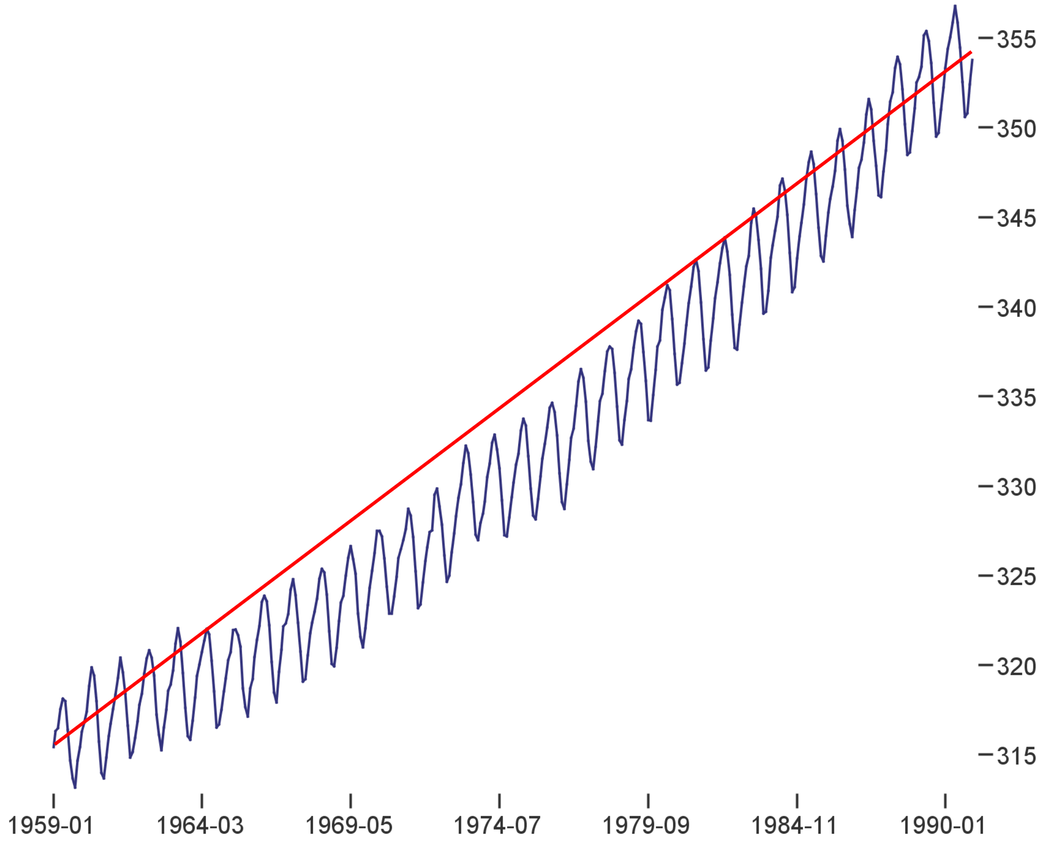
Cleveland, William S., and Robert McGill. "Graphical perception: The visual decoding of quantitative information on graphical displays of data." *1987*

Aspect Ratio

2. Bank to 45°: original data **or** fitted lines



Aspect ratio = 7.87



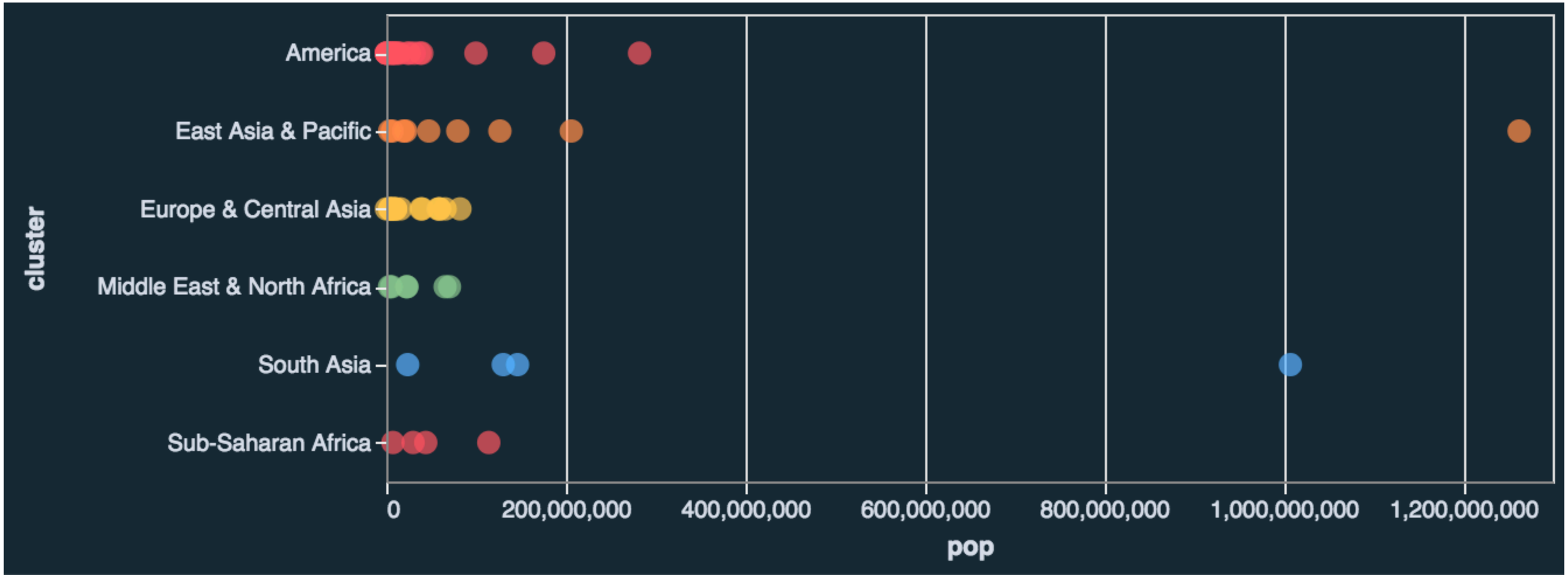
Aspect ratio = 1.17

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

Scaling Axes: Outliers and Skew

Options:

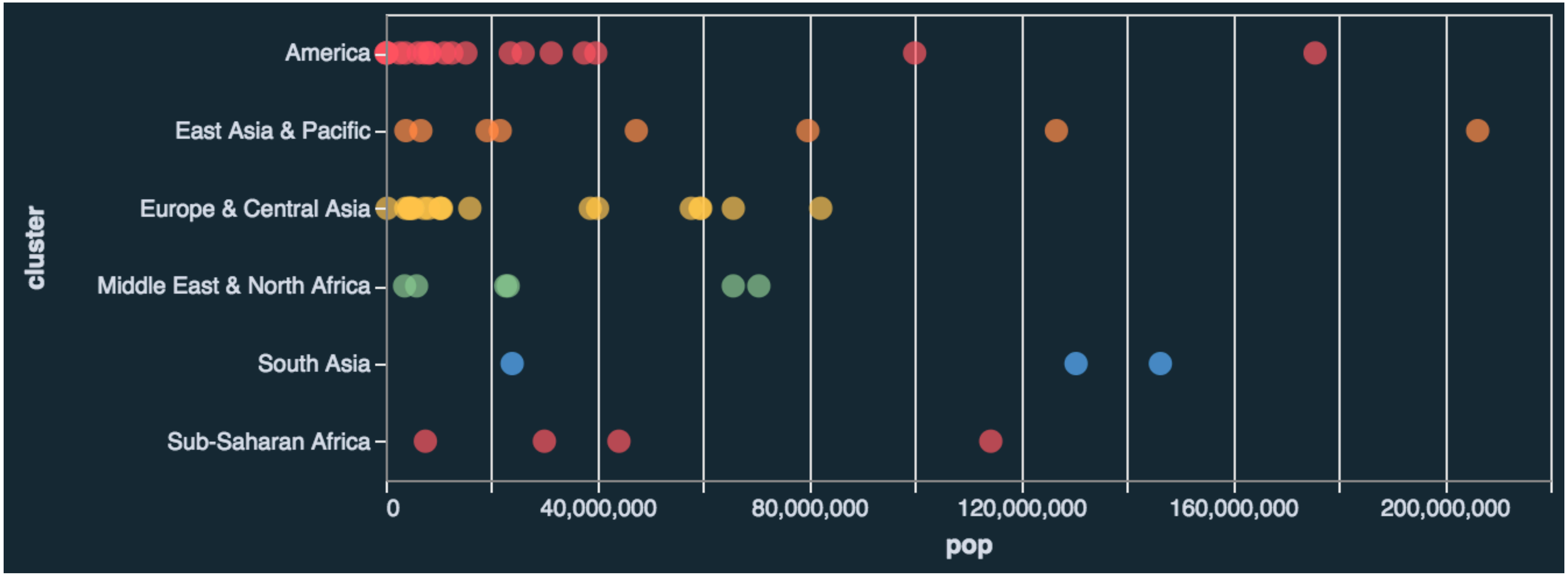
1. Clip them out.



Scaling Axes: Outliers and Skew

Options:

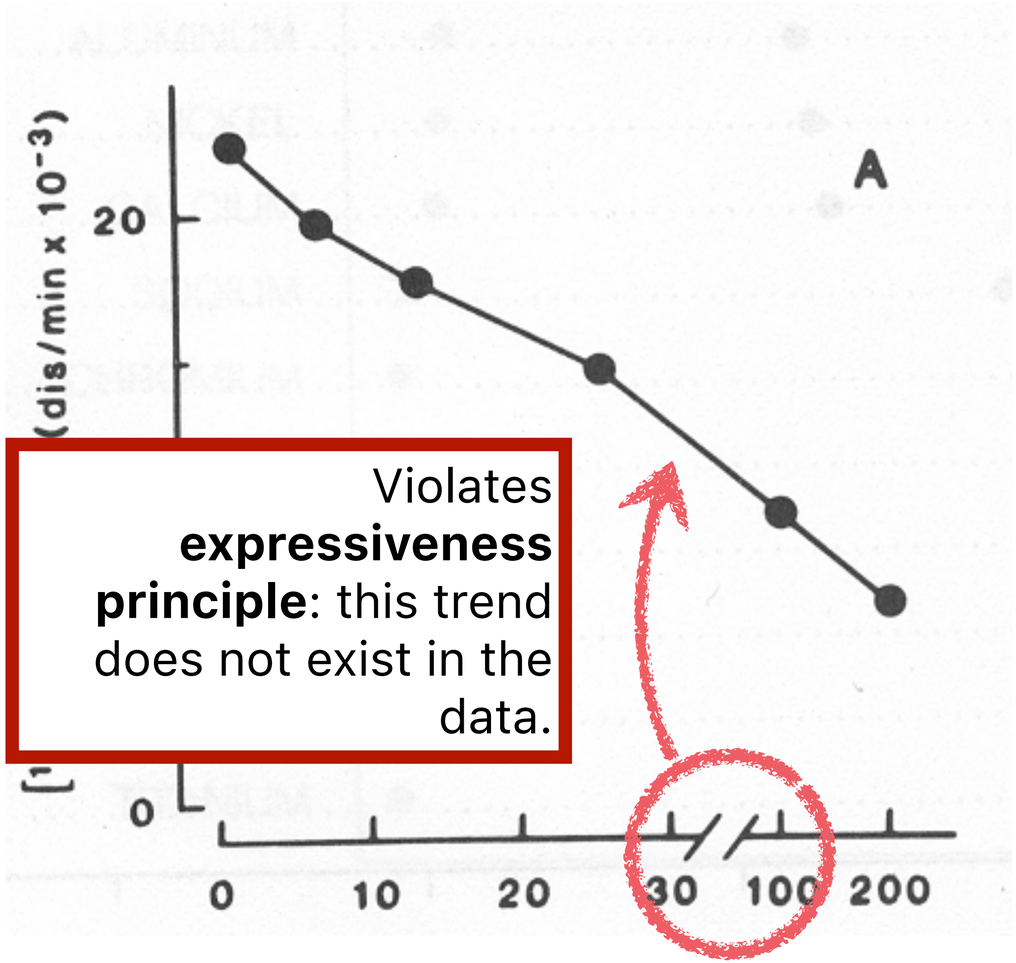
1. Clip them out.



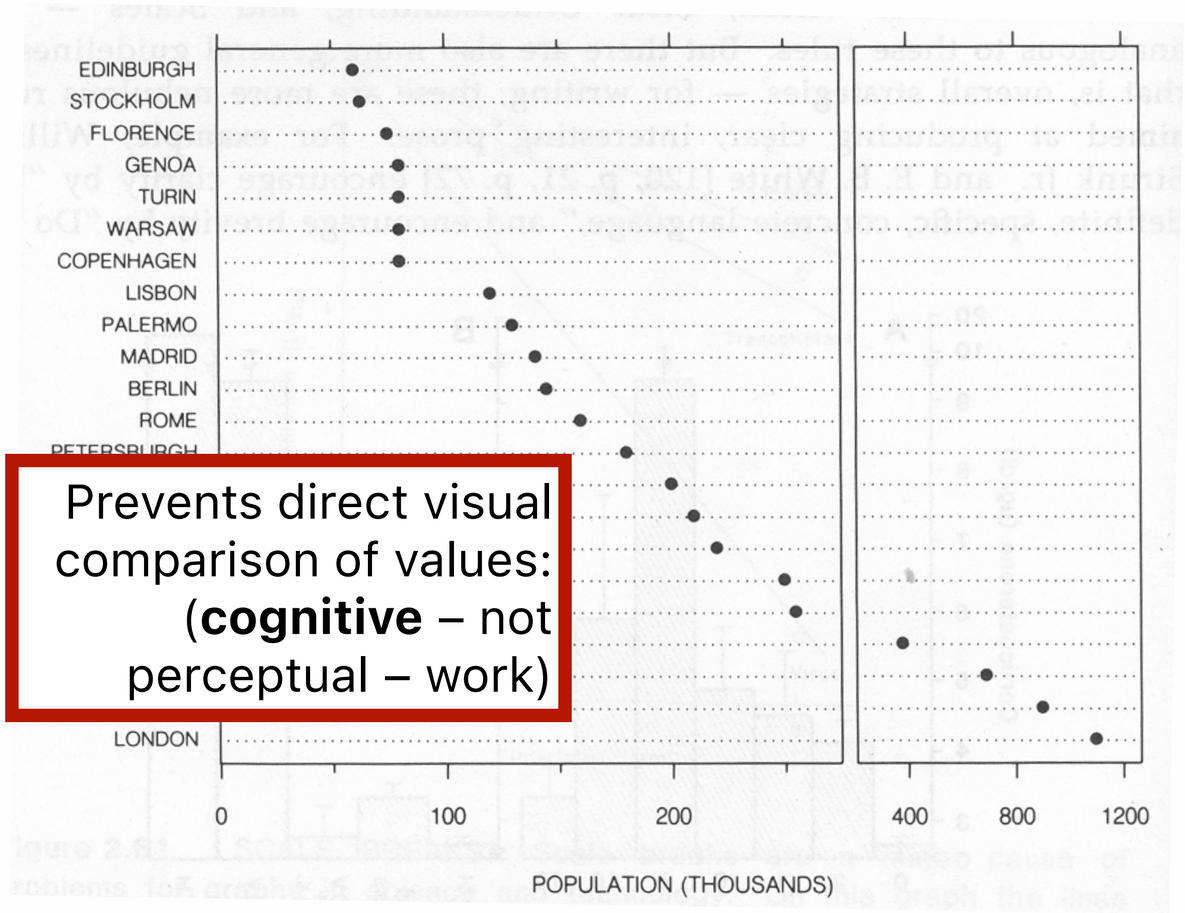
Scaling Axes: Outliers and Skew

Options:

1. Clip them out.
2. Scale breaks



Violates **expressiveness principle**: this trend does not exist in the data.

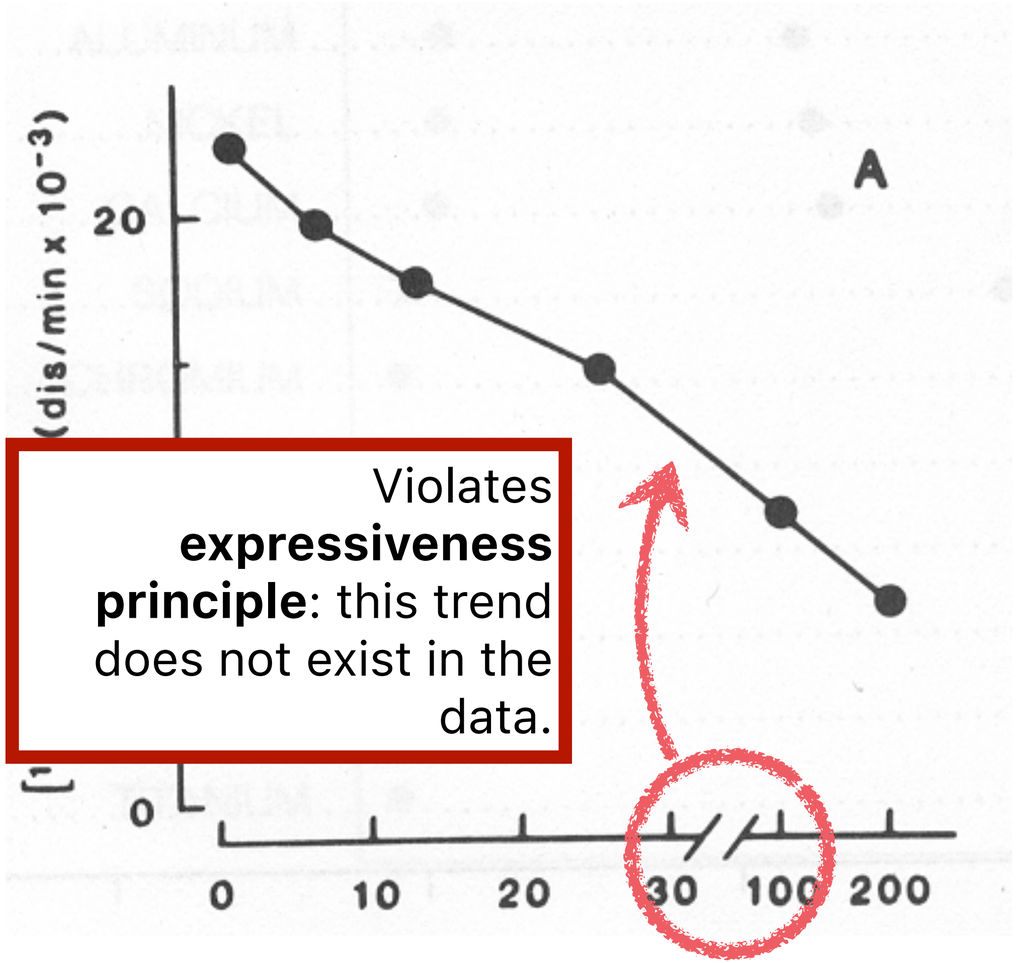


Prevents direct visual comparison of values: (**cognitive** – not perceptual – work)

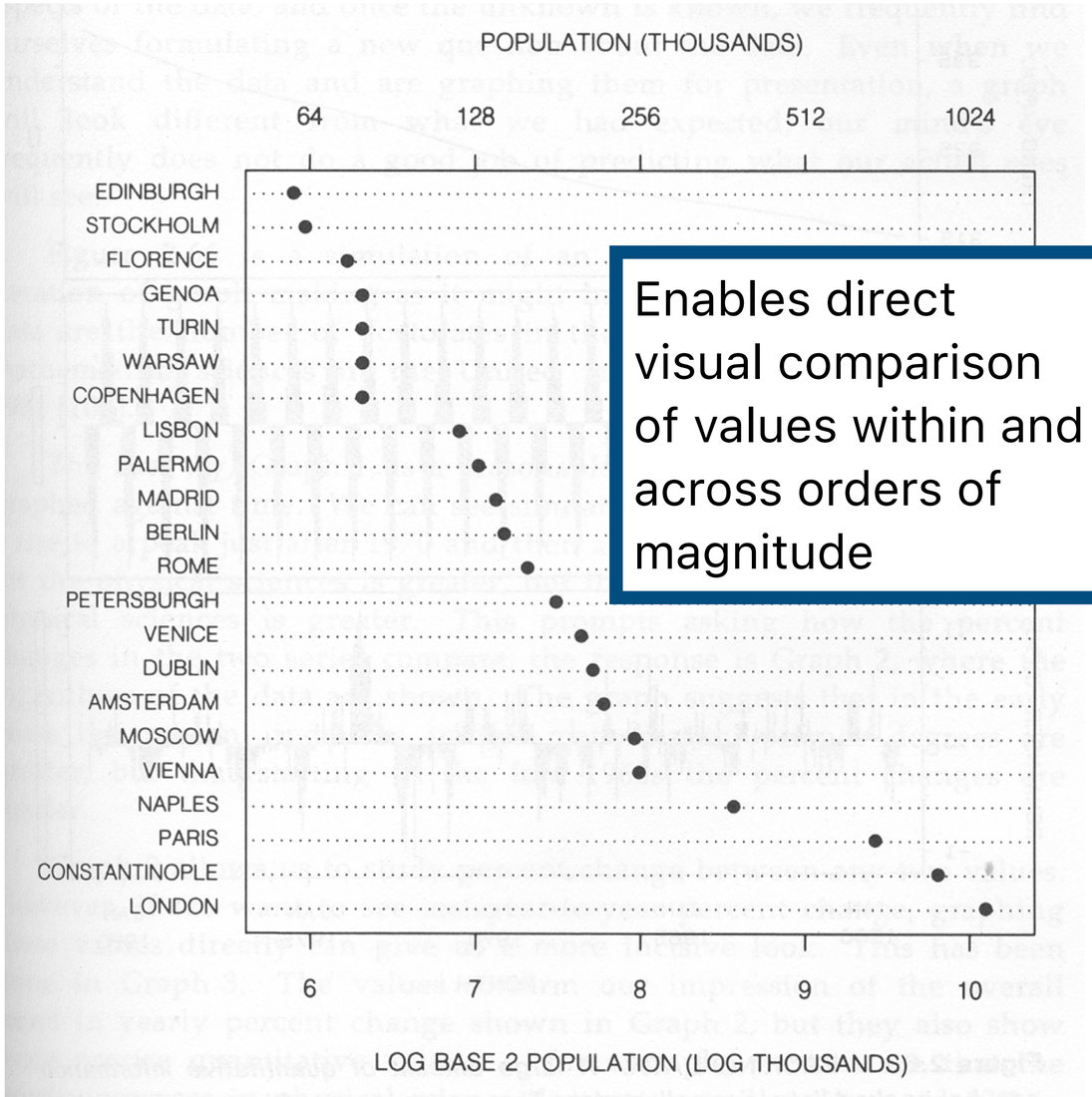
Scaling Axes: Outliers and Skew

Options:

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



Violates **expressiveness principle**: this trend does not exist in the data.

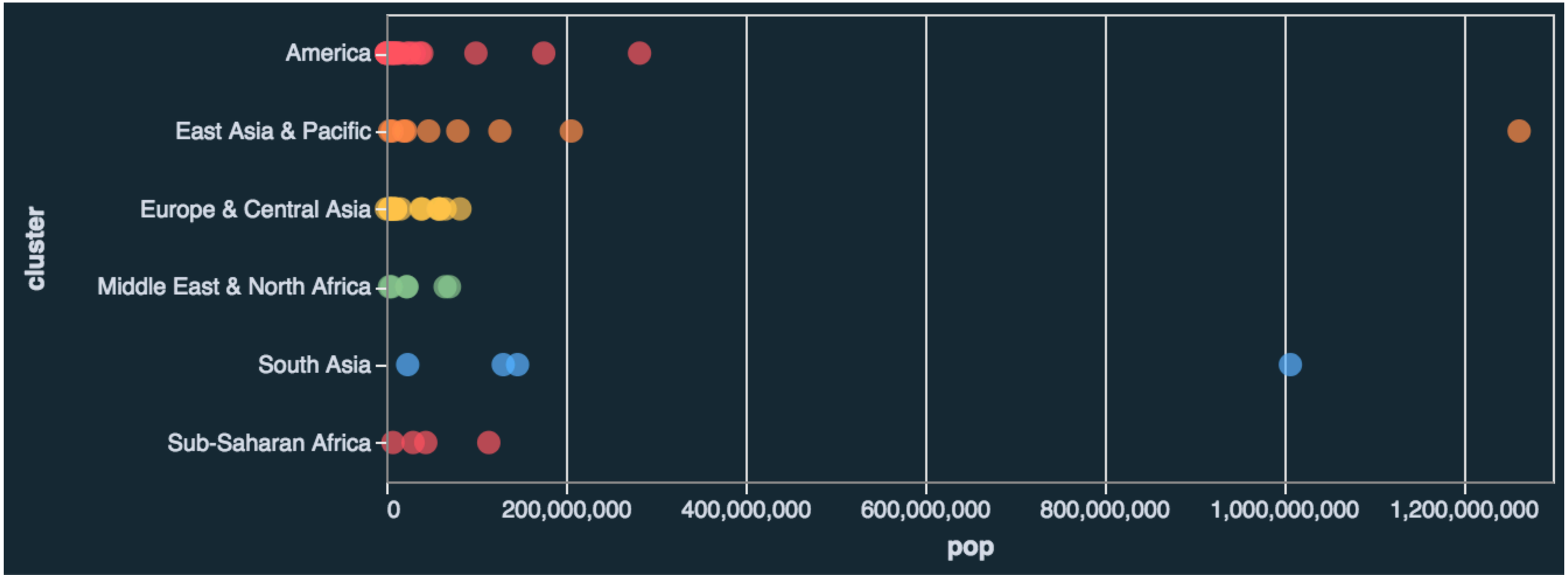


Enables direct visual comparison of values within and across orders of magnitude

Scaling Axes: Outliers and Skew

Options:

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



Scaling Axes: Outliers and Skew

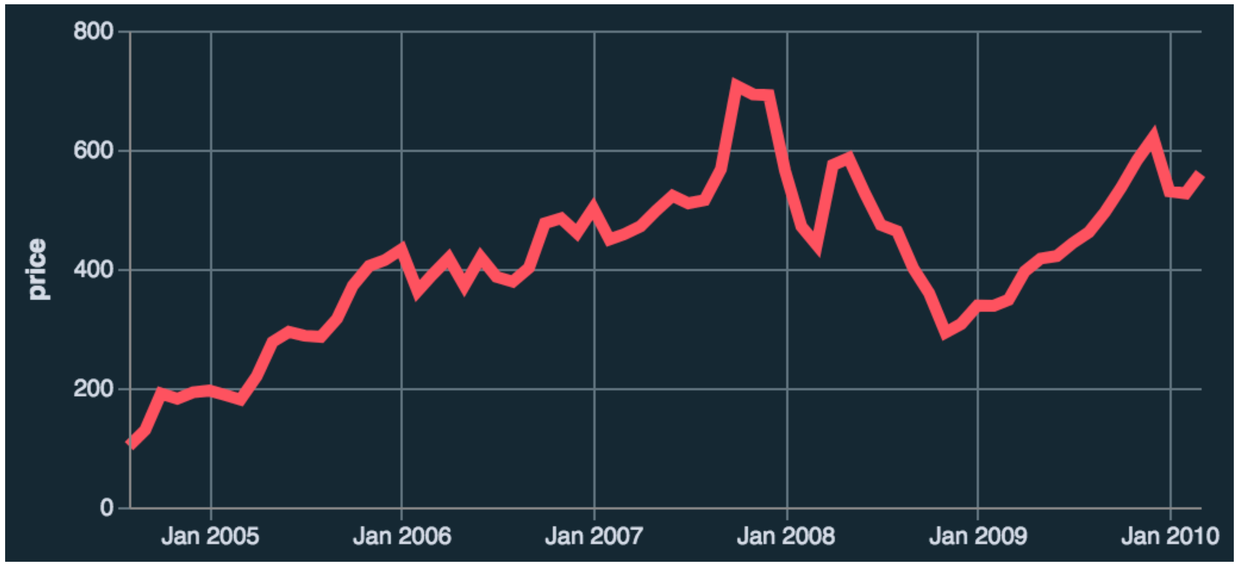
Options:

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

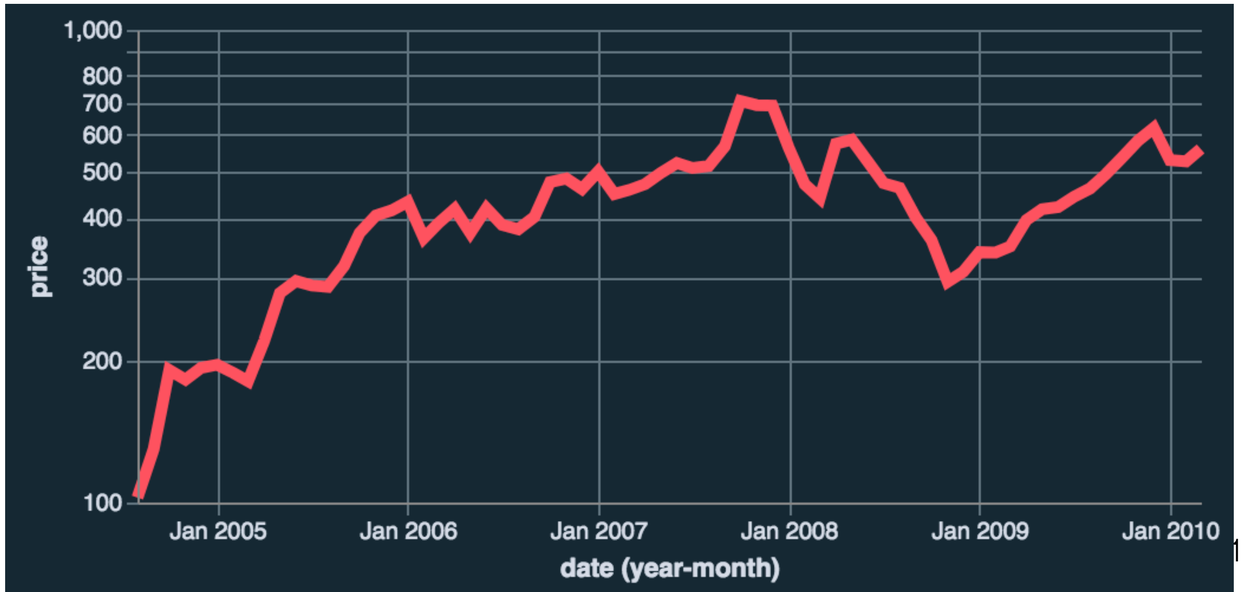
Direct visual comparison of values within and across orders of magnitude – no clipping required!

Scaling Axes: Linear vs Log

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



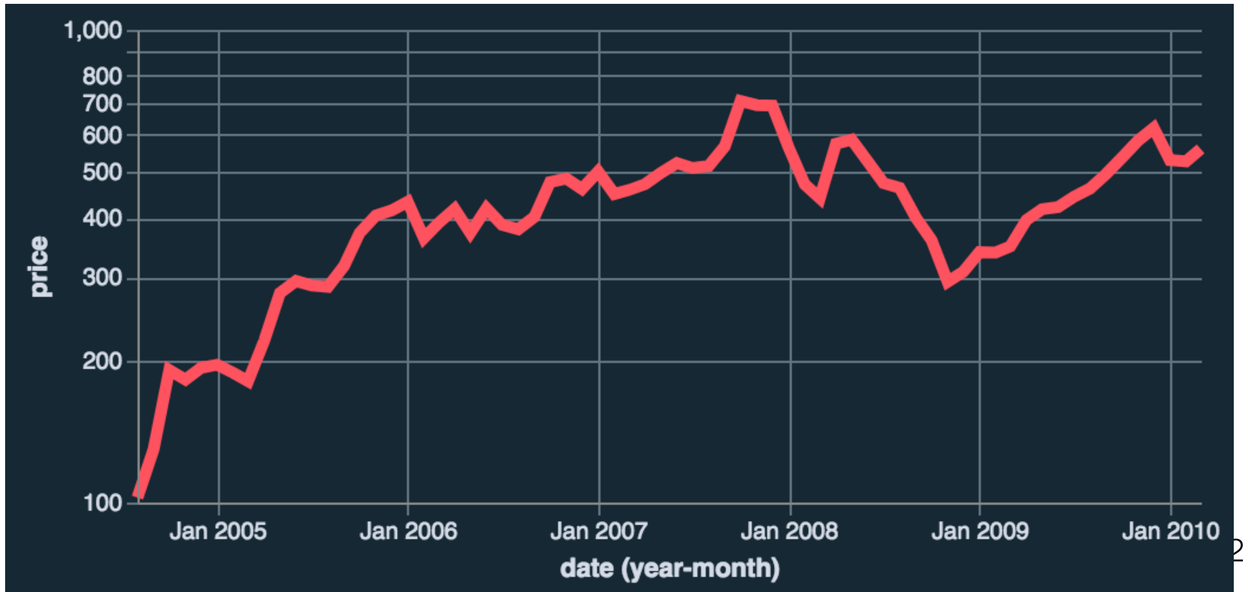
Log Scale
Percentage change
10 visual units =
multiplication of 10 data units



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

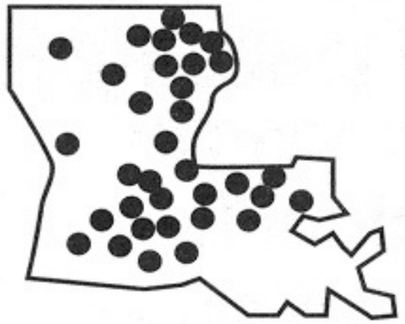
(Mis)leading the witness

Using space (in)effectively

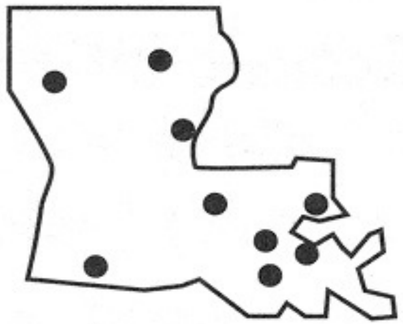
(De-)Obfuscating data

(Mis)leading the witness

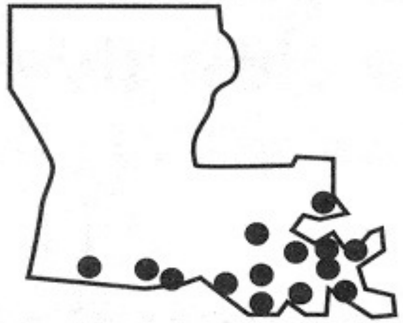
alfisol



entisol



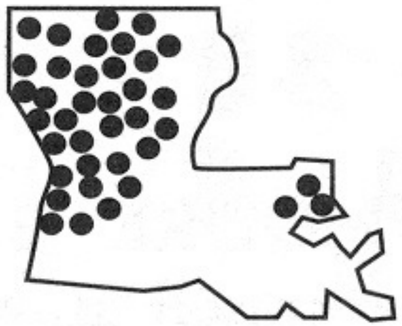
histosol



inceptisol

mollisol

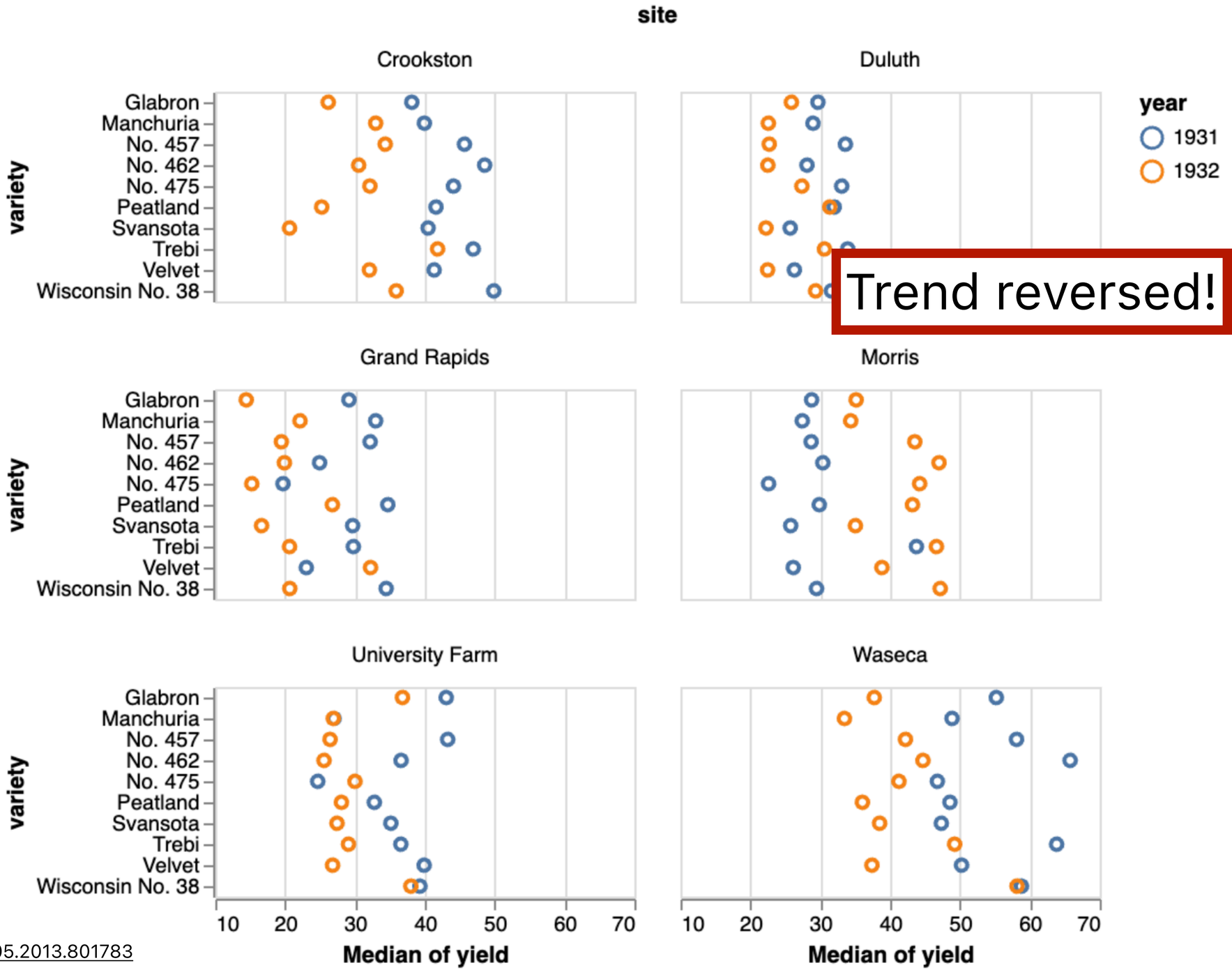
ultisol



Trellis Plots

Subdivide space to enable comparison across multiple plots.

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



https://www.tandfonline.com/doi/abs/10.1080/00031305.2013.801783

Data-ink Ratio

$$= \frac{\text{Data Ink}}{\text{Ink used in graphic}}$$

= Proportion of a graphic's ink devoted to non-redundant display of data.

= 1.0 – proportion of graphic that can be erased.

Remove

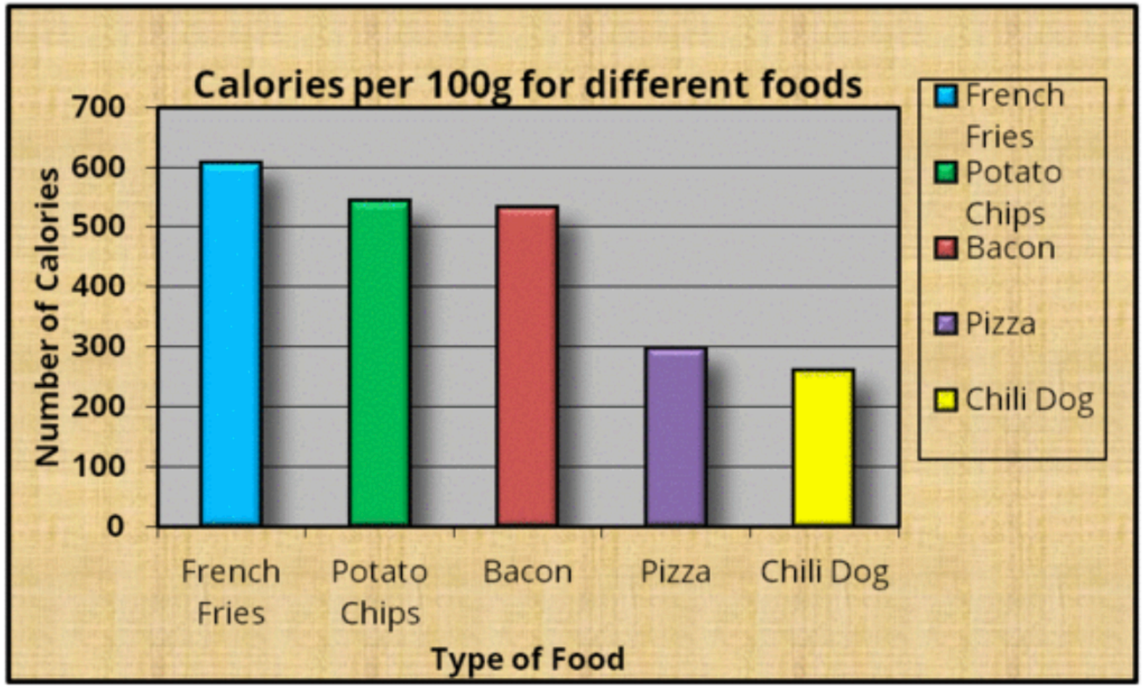
to improve

(the **data-ink** ratio)

Data-ink Ratio

Join at
slido.com
#3892 640

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?



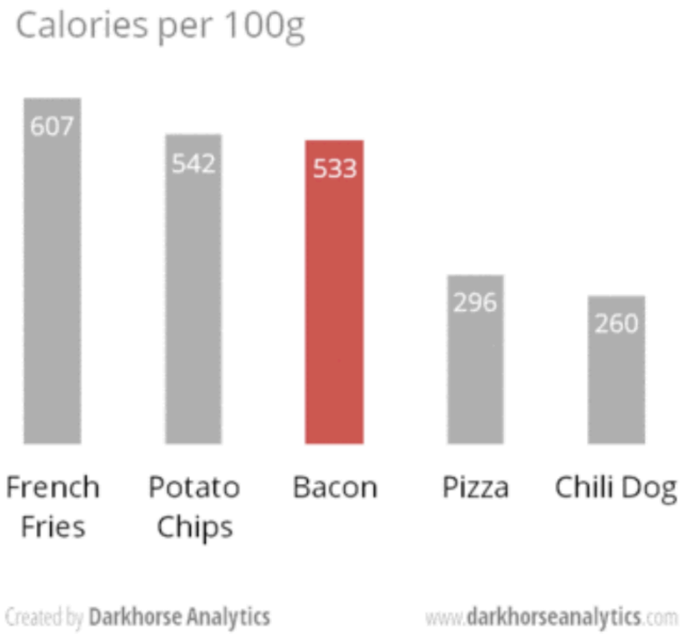
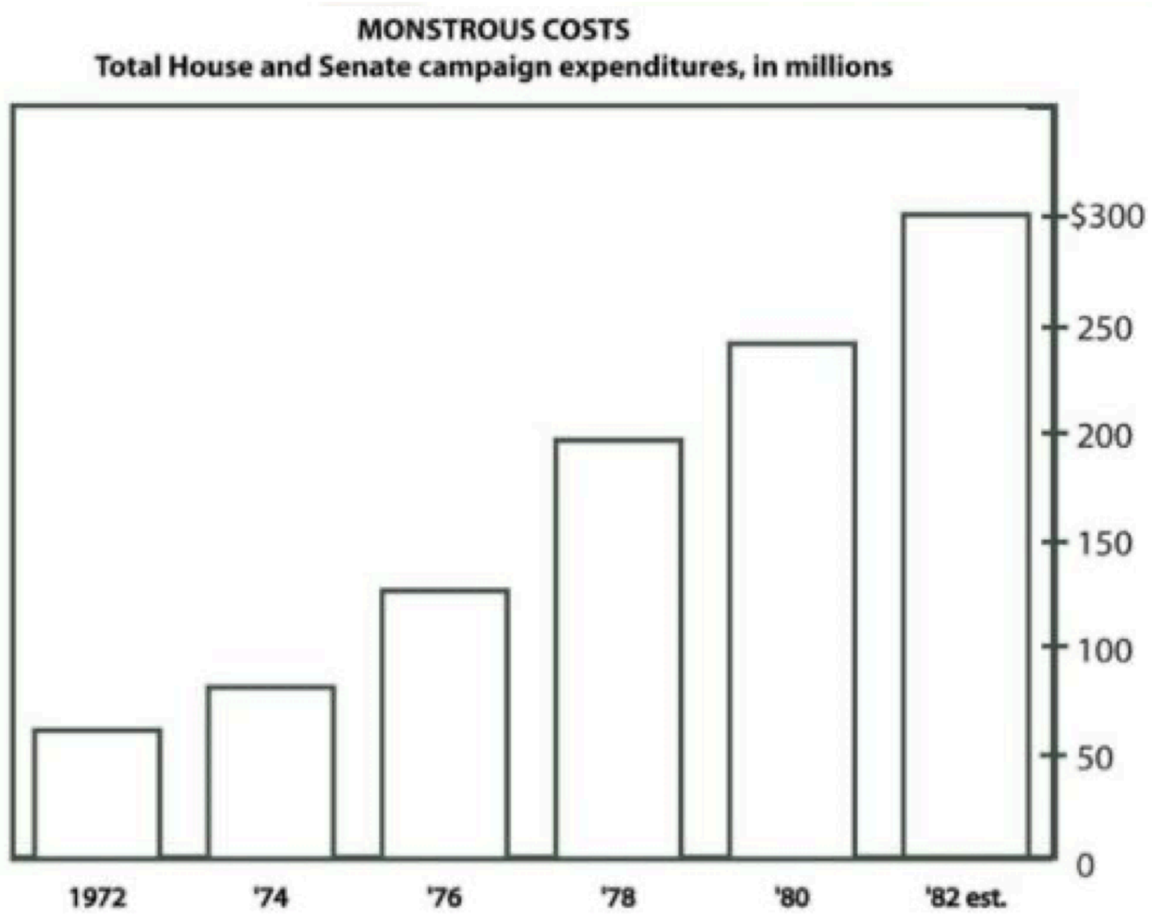
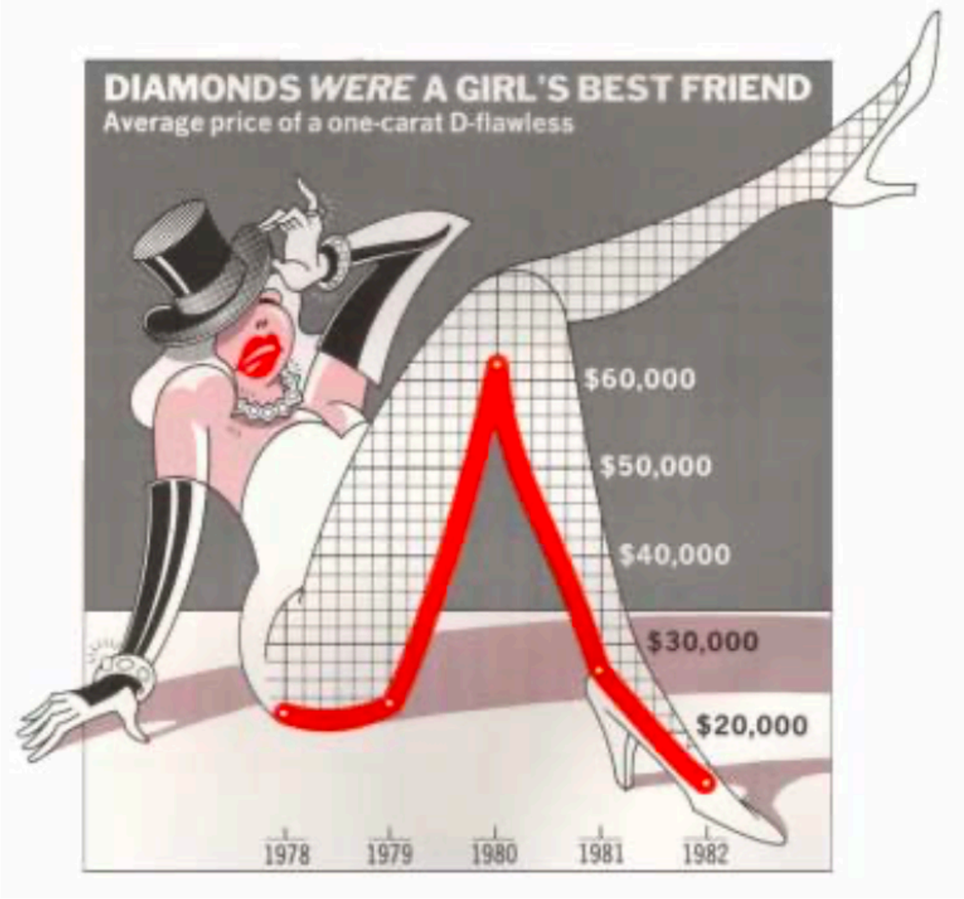


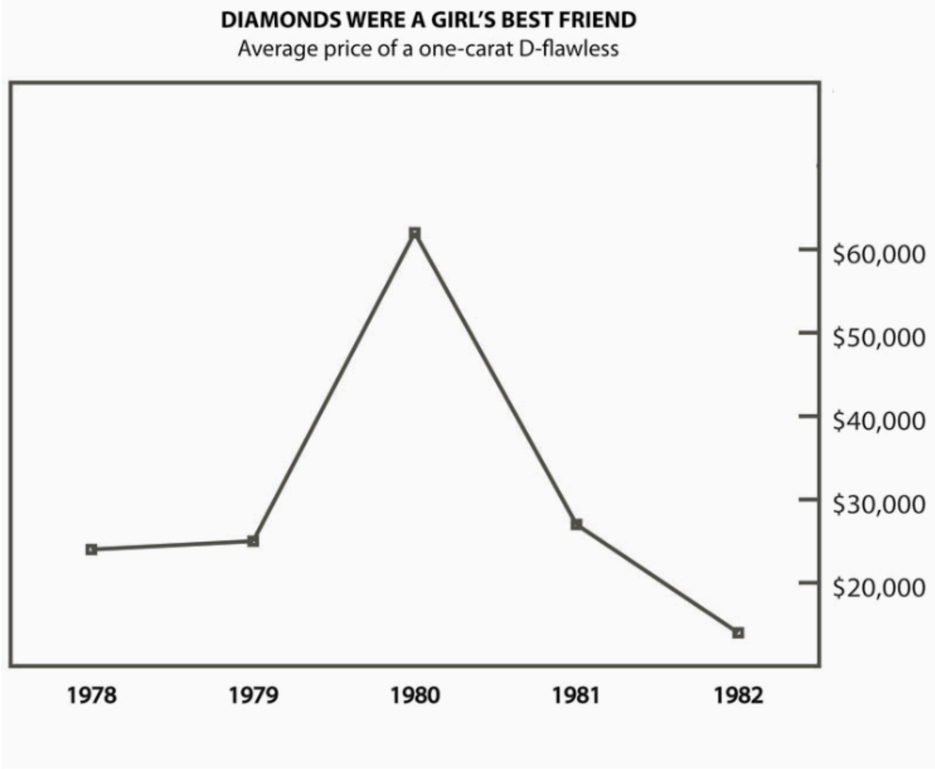
Chart "Junk"



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

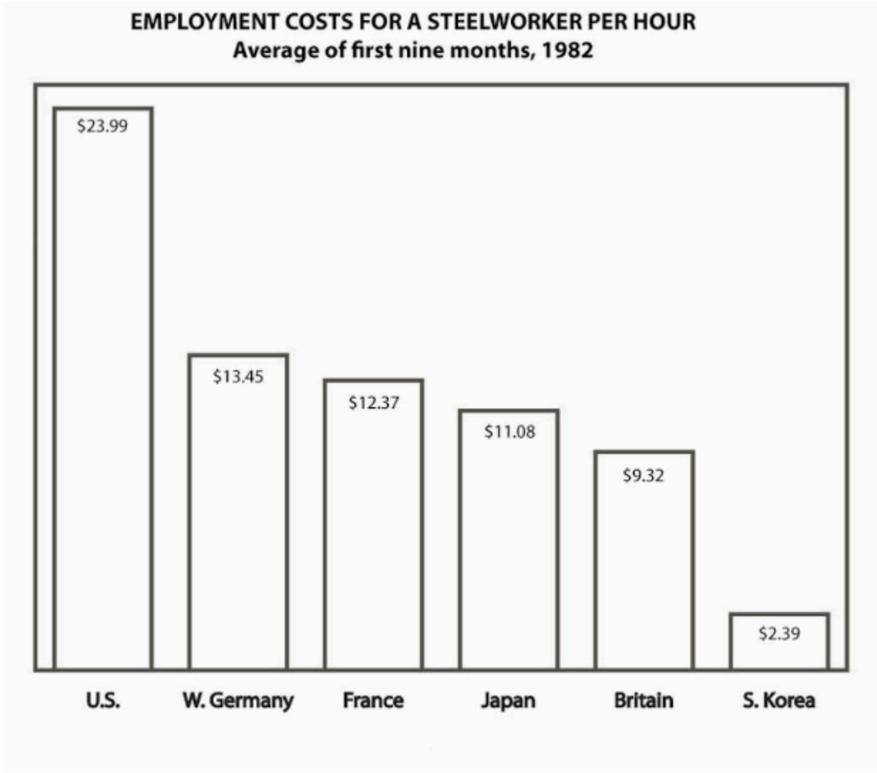
Chart "Junk"





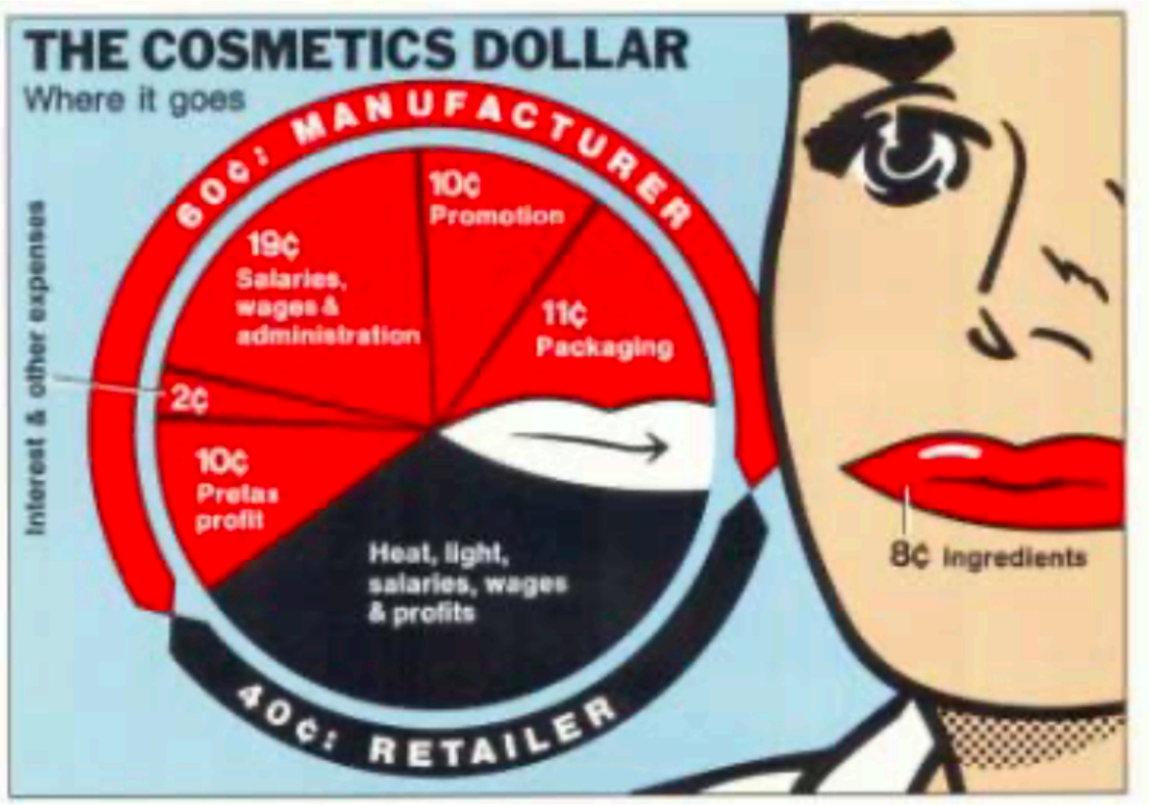
Bateman, Scott, et al. "Useful junk? The effects of visual embellishment on comprehension and memorability of charts." *CHI 2010*.

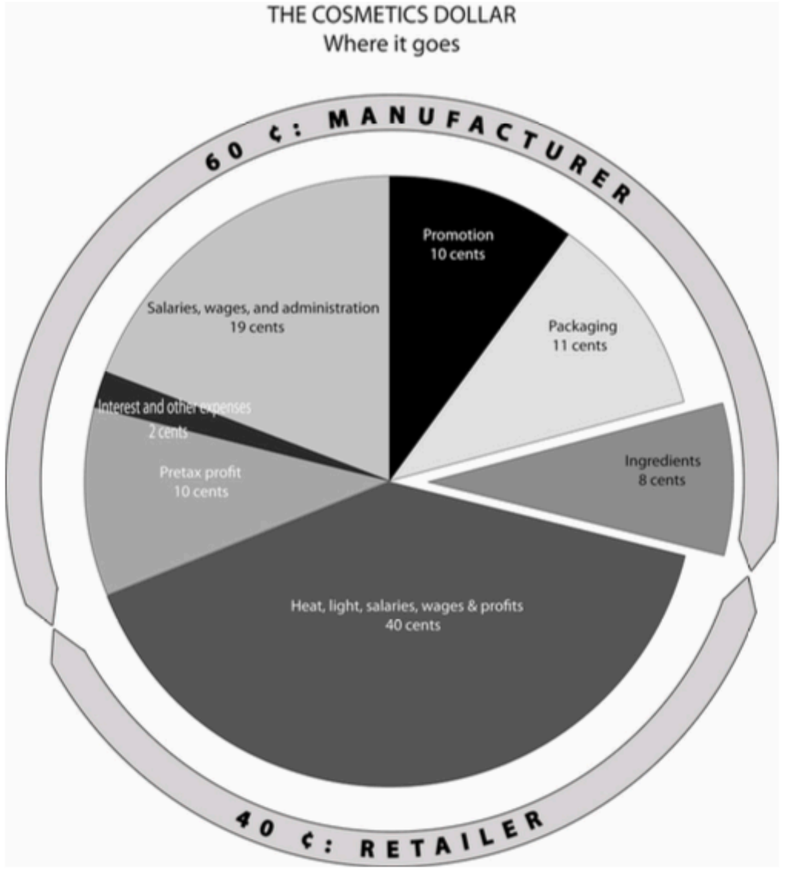
Chart "Junk"



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

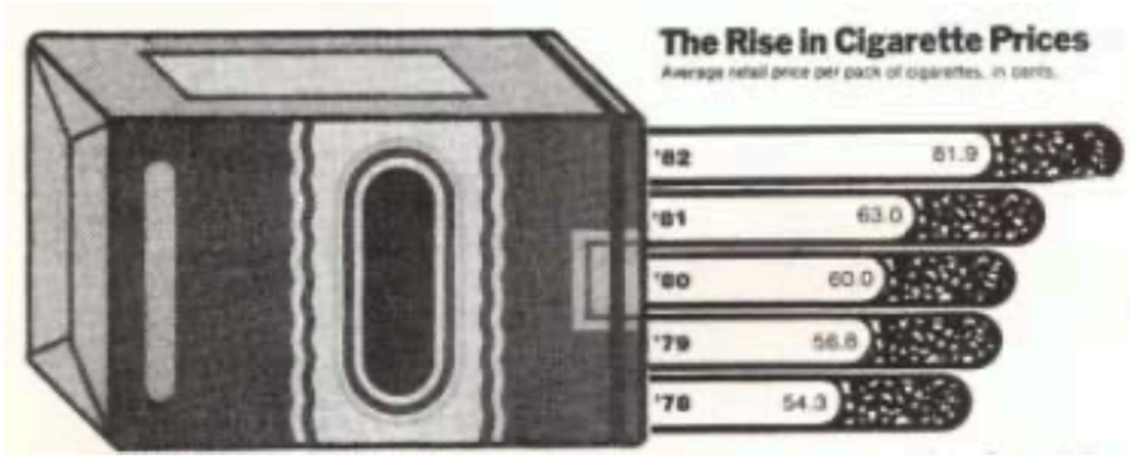
Chart "Junk"

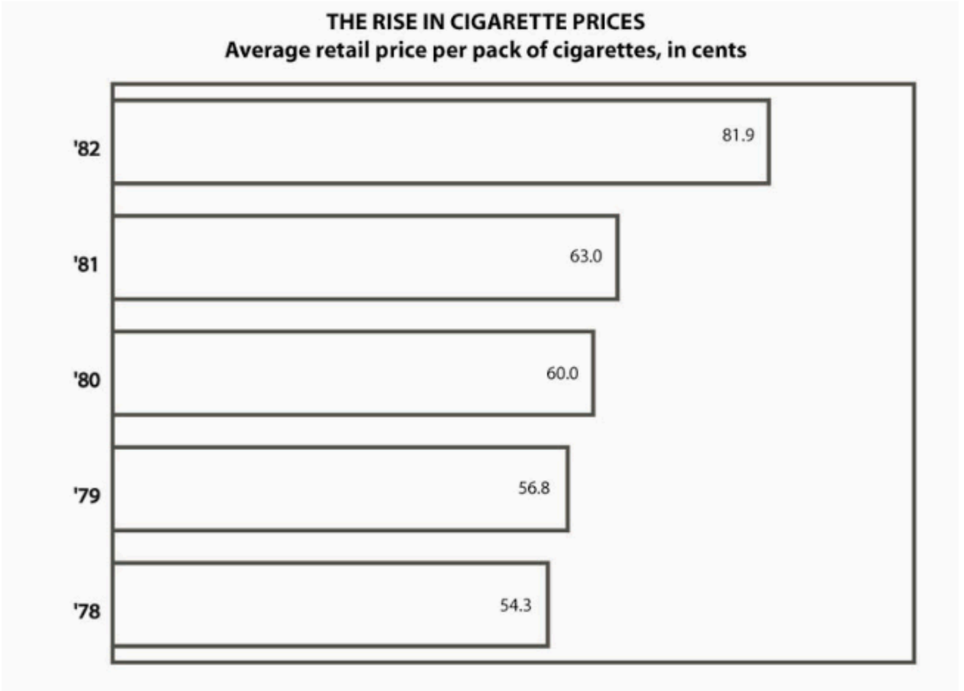




Bateman, Scott, et al. "Useful junk? The effects of visual embellishment on comprehension and memorability of charts." *CHI 2010*.

Chart "Junk"





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

Using space (in)effectively

(De-)Obfuscating data

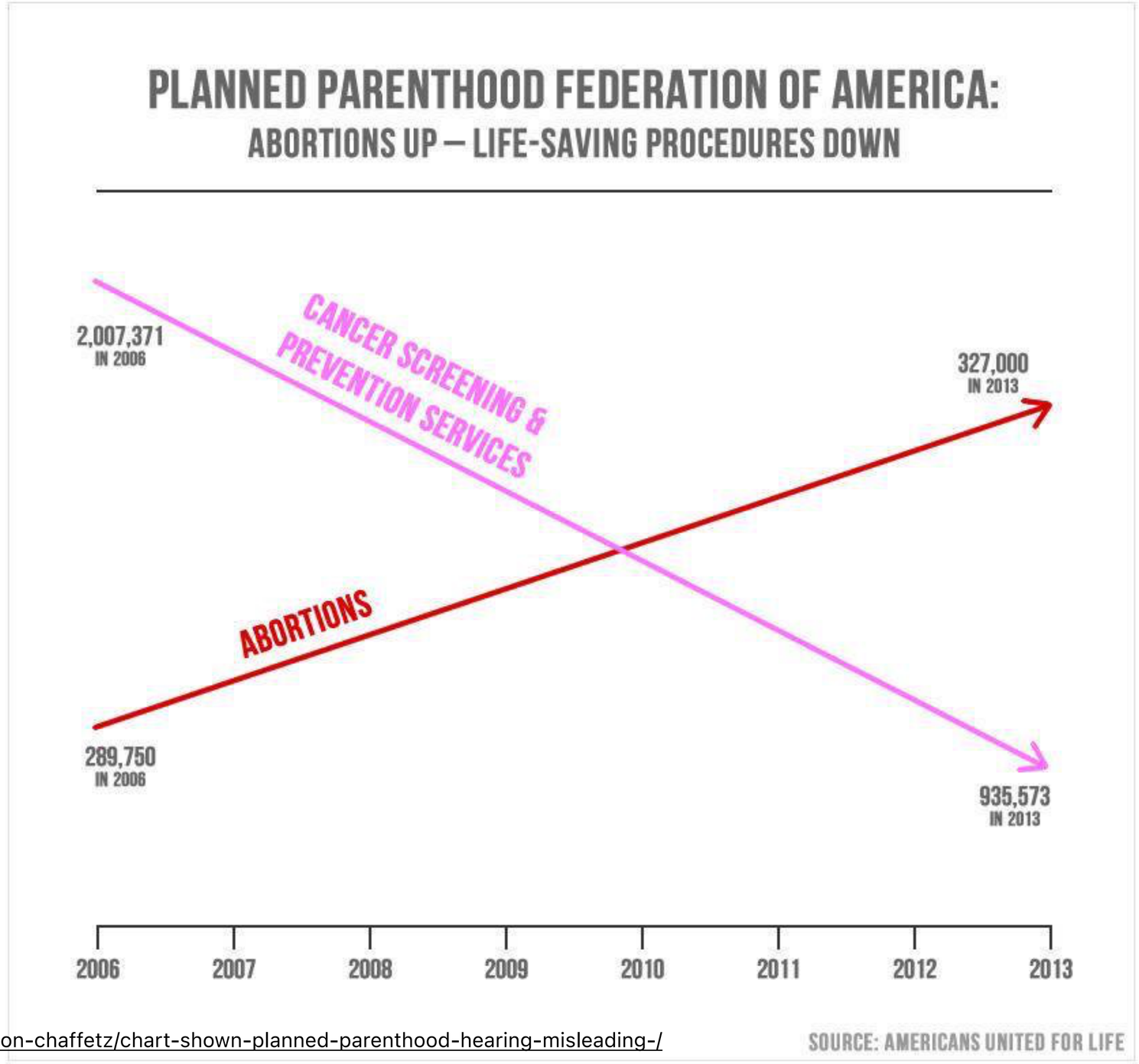
(Mis)leading the witness

Using space (in)effectively

(De-)Obfuscating data

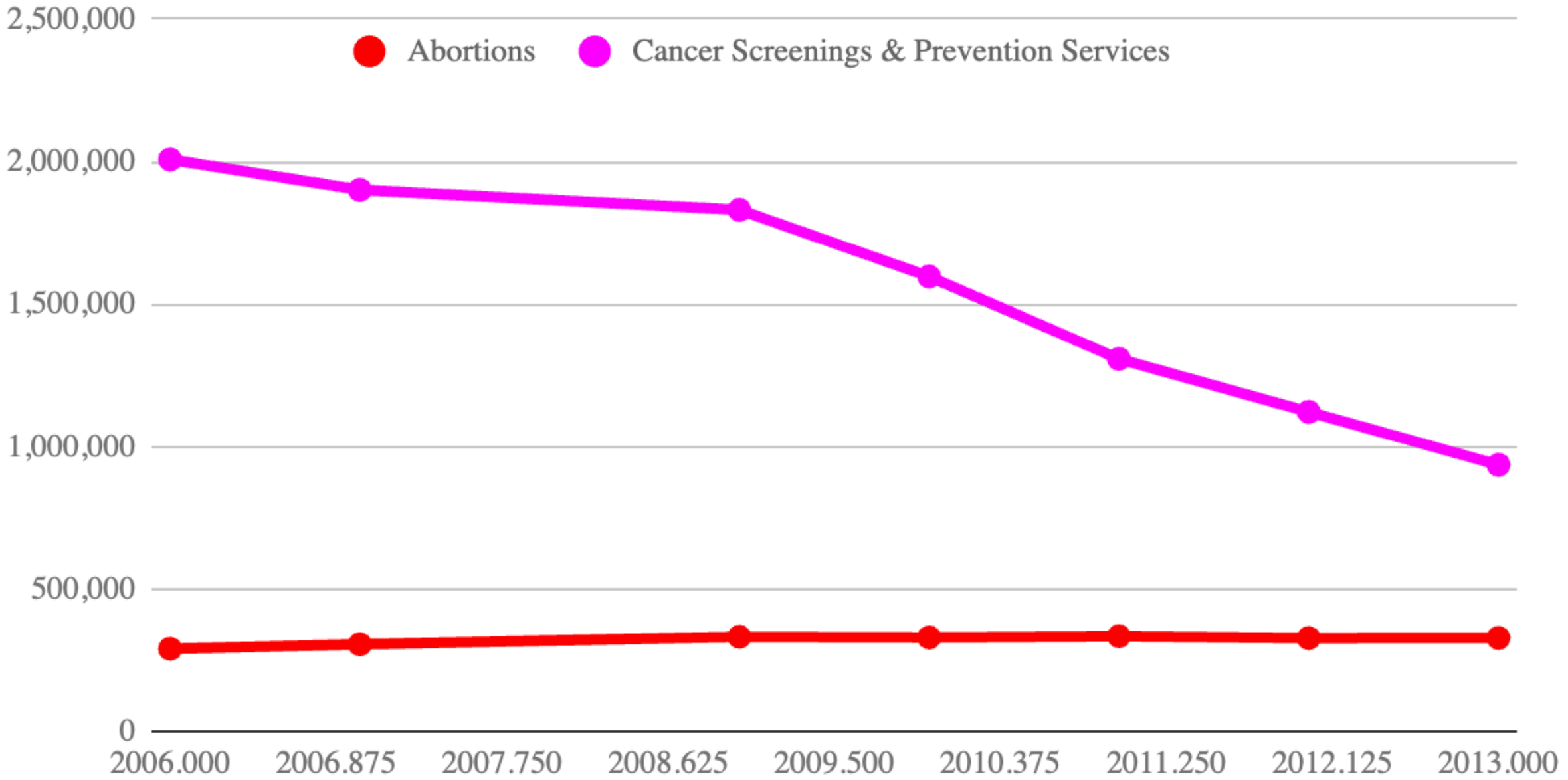
(Mis)leading the witness

Join at
slido.com
#3892 640



https://www.politifact.com/factchecks/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading-/

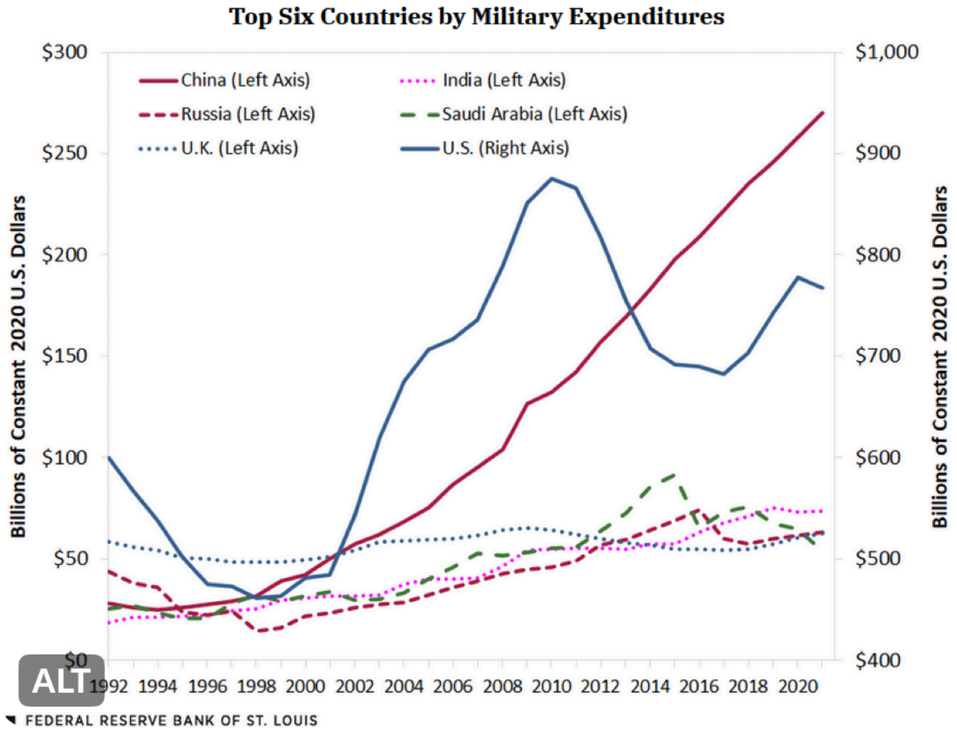
Planned Parenthood Federation of America: Abortions vs. Cancer and Prevention Services



https://www.politifact.com/factchecks/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading-/

St. Louis Fed
@stlouisfed

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



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...

Do you find this helpful? Rate it

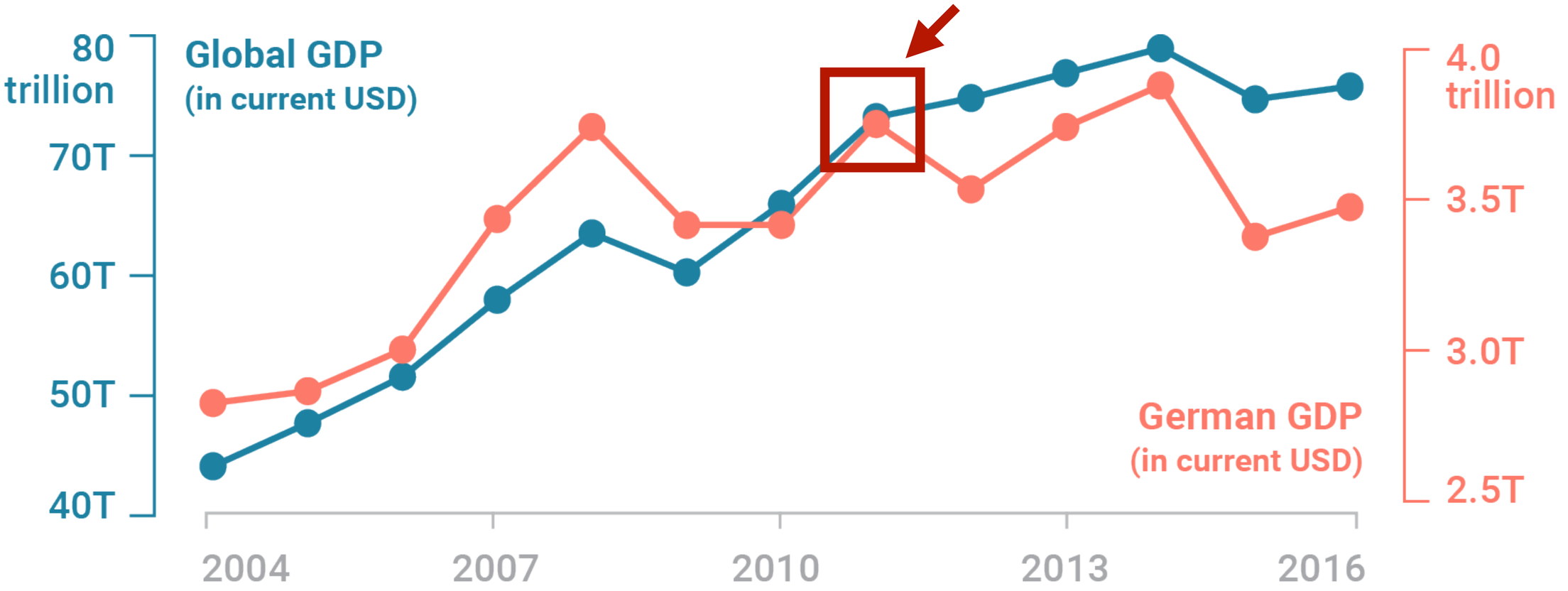
Context is written by people who use Twitter, and appears when rated helpful by others. Find out more.

4:00 PM · 1/22/23 · **7.3M** Views

1,128 Likes **157** Retweets **2,281** Quotes

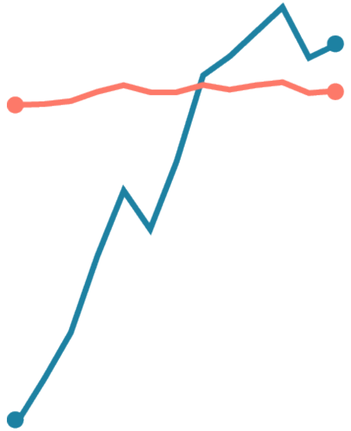
Dual Axes Charts

German and world GDP were equal in 2011??

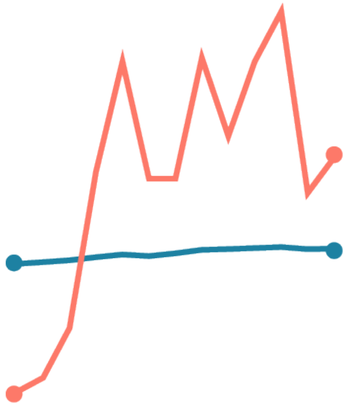


https://blog.datawrapper.de/dualaxis/

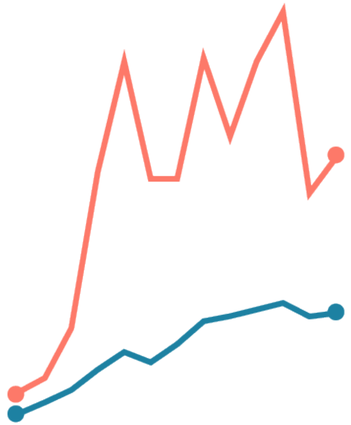
Dual-Axes Charts



Orange steady,
Blue massively increasing.



Blue steady,
Orange increasing.



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

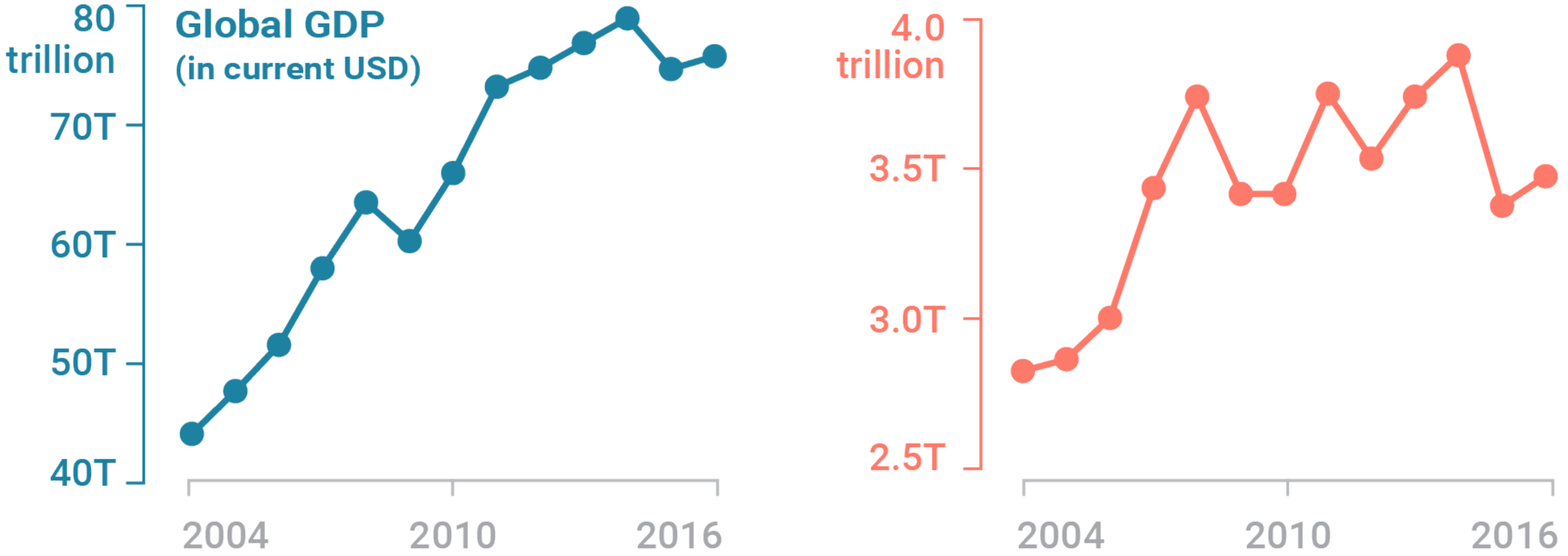
Both started at the same level, but Blue increased far more than Orange.

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

Both steady.

https://blog.datawrapper.de/dualaxis/

Dual-Axes Charts



https://blog.datawrapper.de/dualaxis/

Using space (in)effectively

(De-)Obfuscating data

(Mis)leading the witness

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

Visualization draws upon both science and art!

Next Time: Perception