

(In)Effective Visual Encoding

DSC 106: Data Visualization

Sam Lau

UC San Diego

Announcements

Lab 1 due today.

Project 1 checkpoint due next Tuesday.

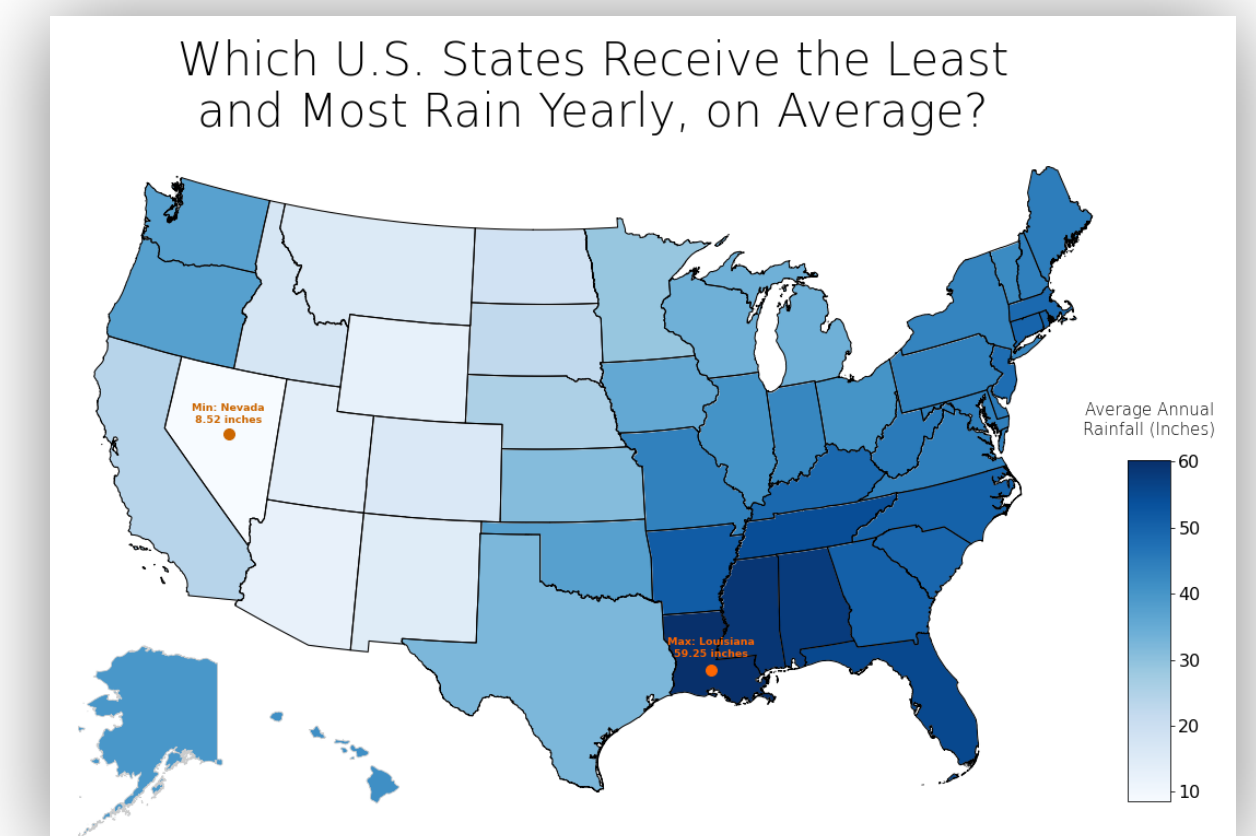
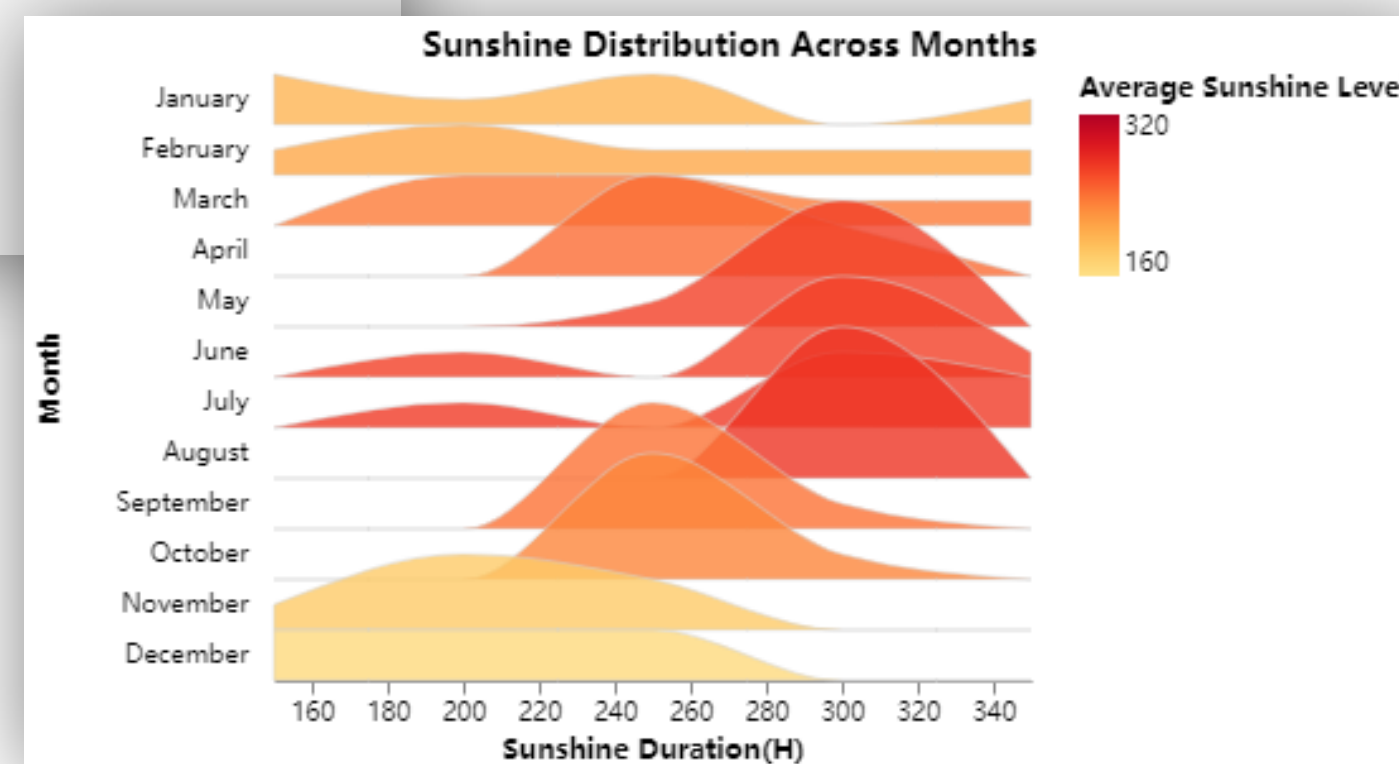
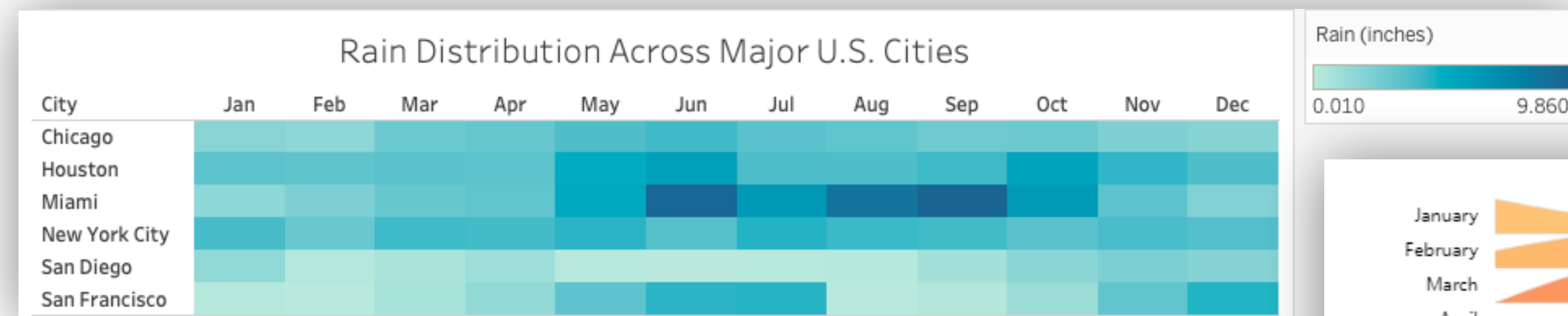
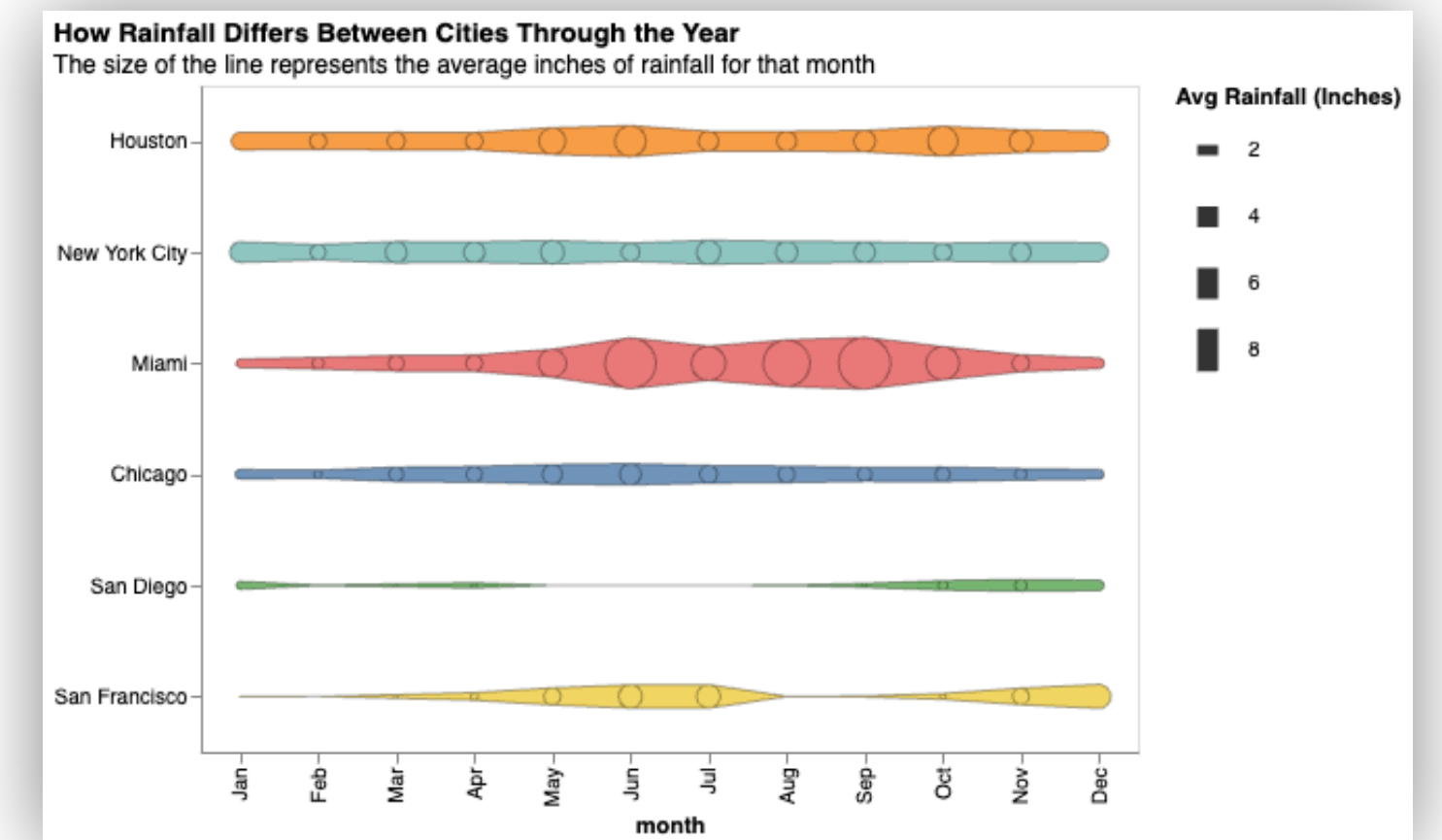
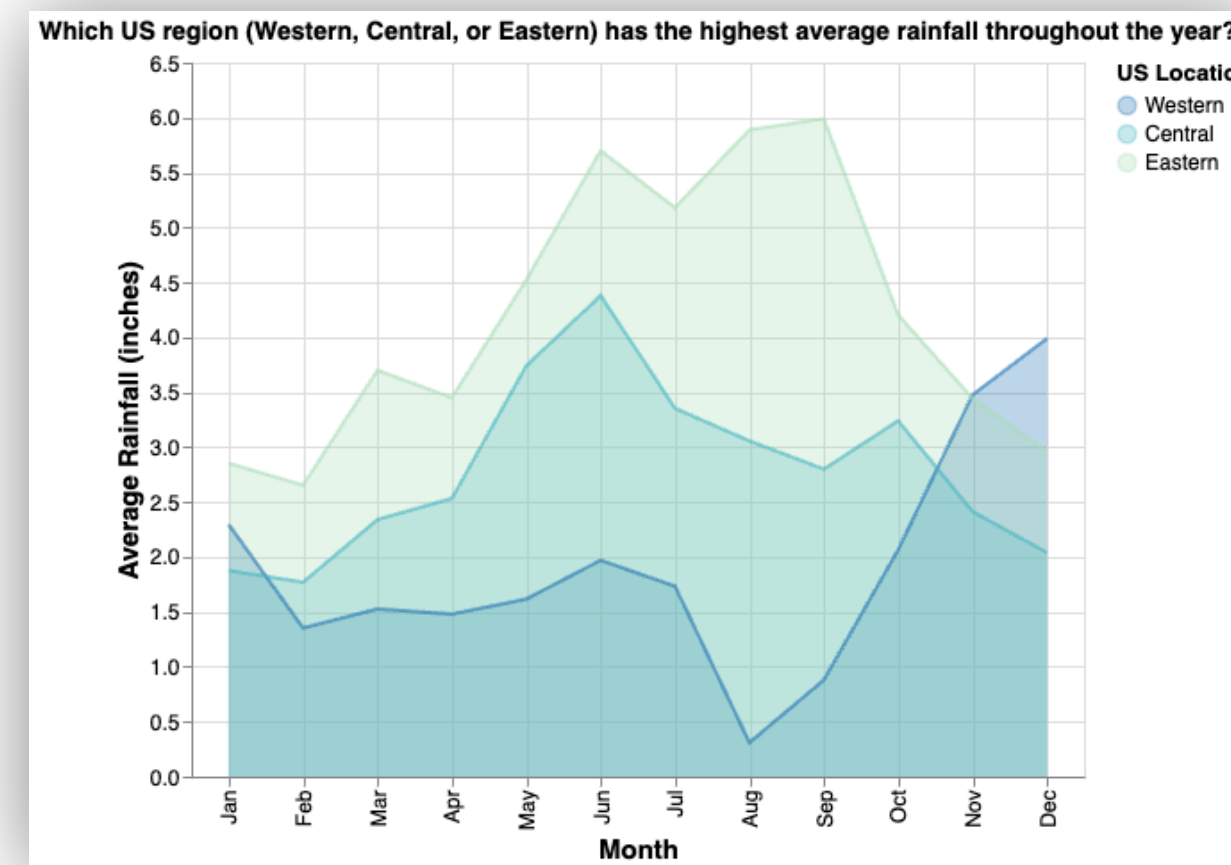
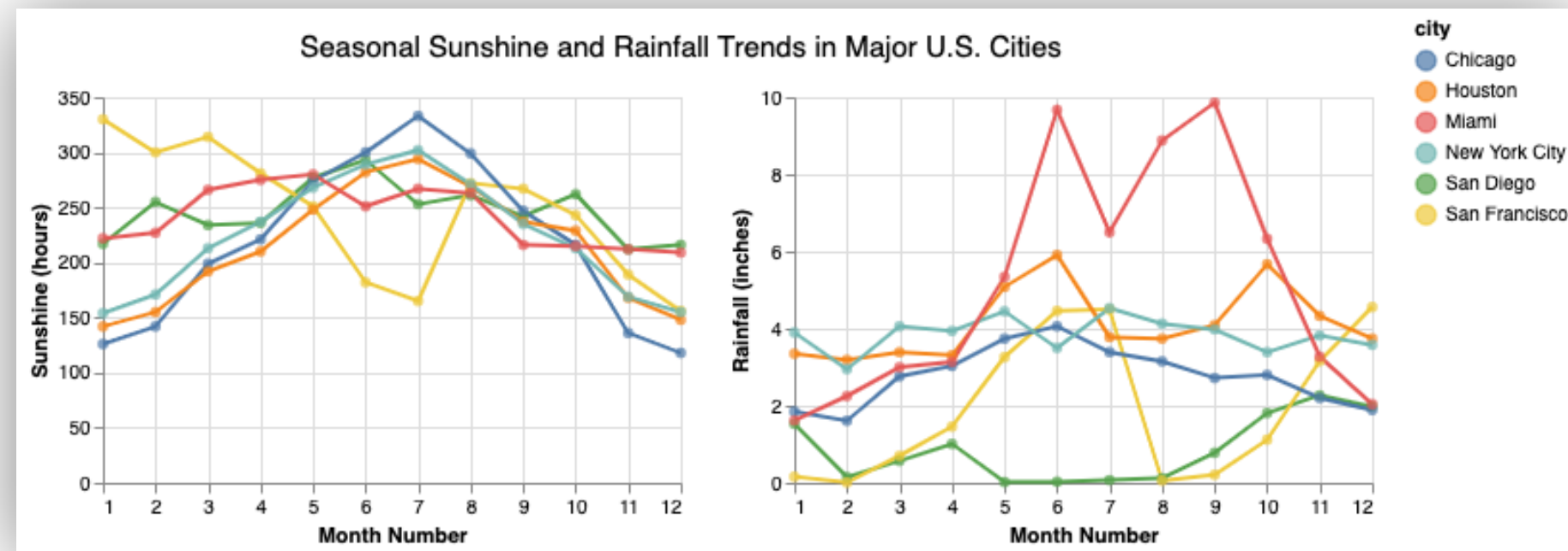
No lectures on Wed and Fri next week, coming to discussion on Monday will give you full participation that week.

FAQs:

1. What if I joined the class late? We'll provide extensions for assignments, you'll use participation drops for weeks missed.

How to ace Project 1

Try out a LOT of possible visualizations!

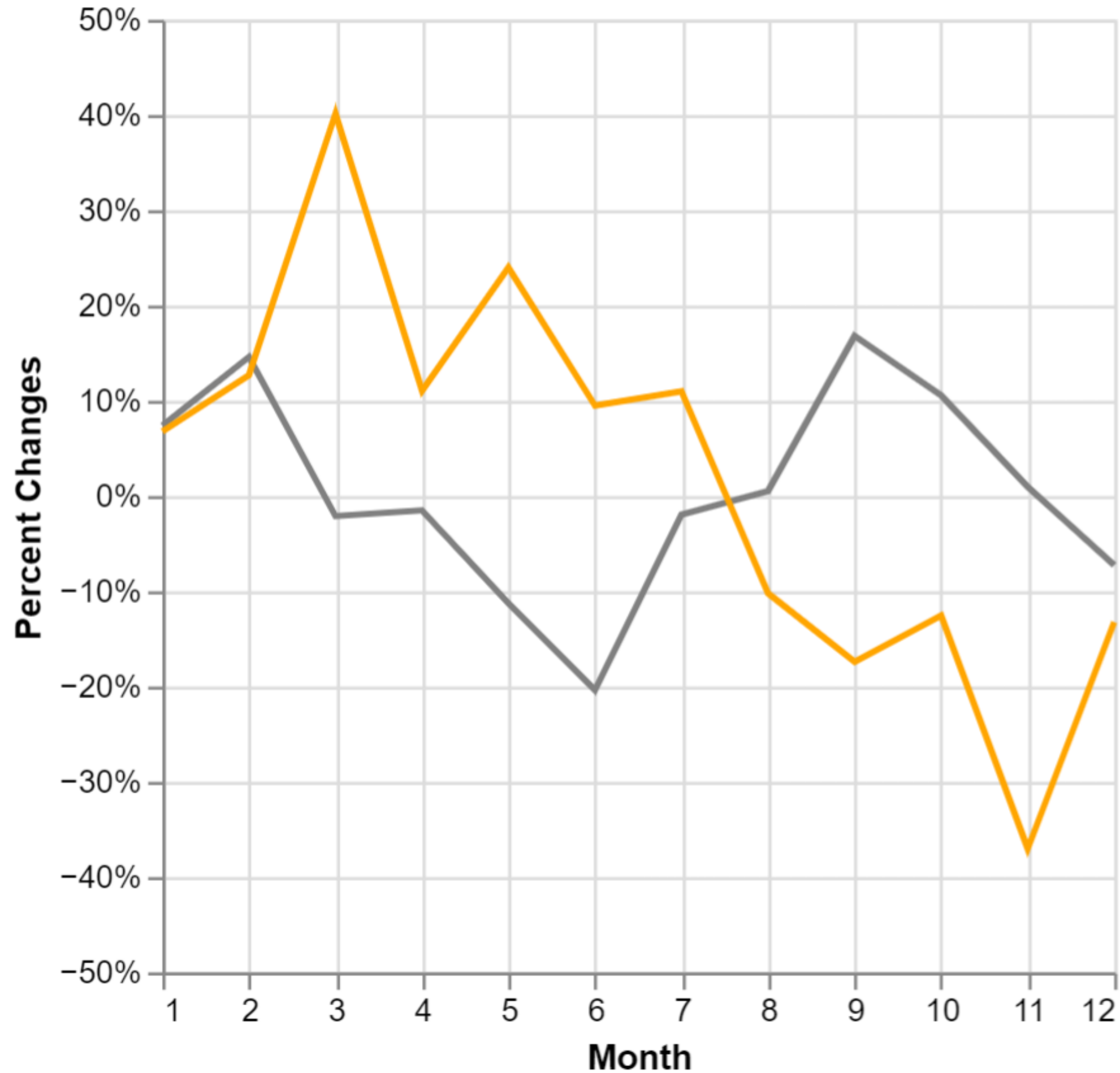


Then, tell a story

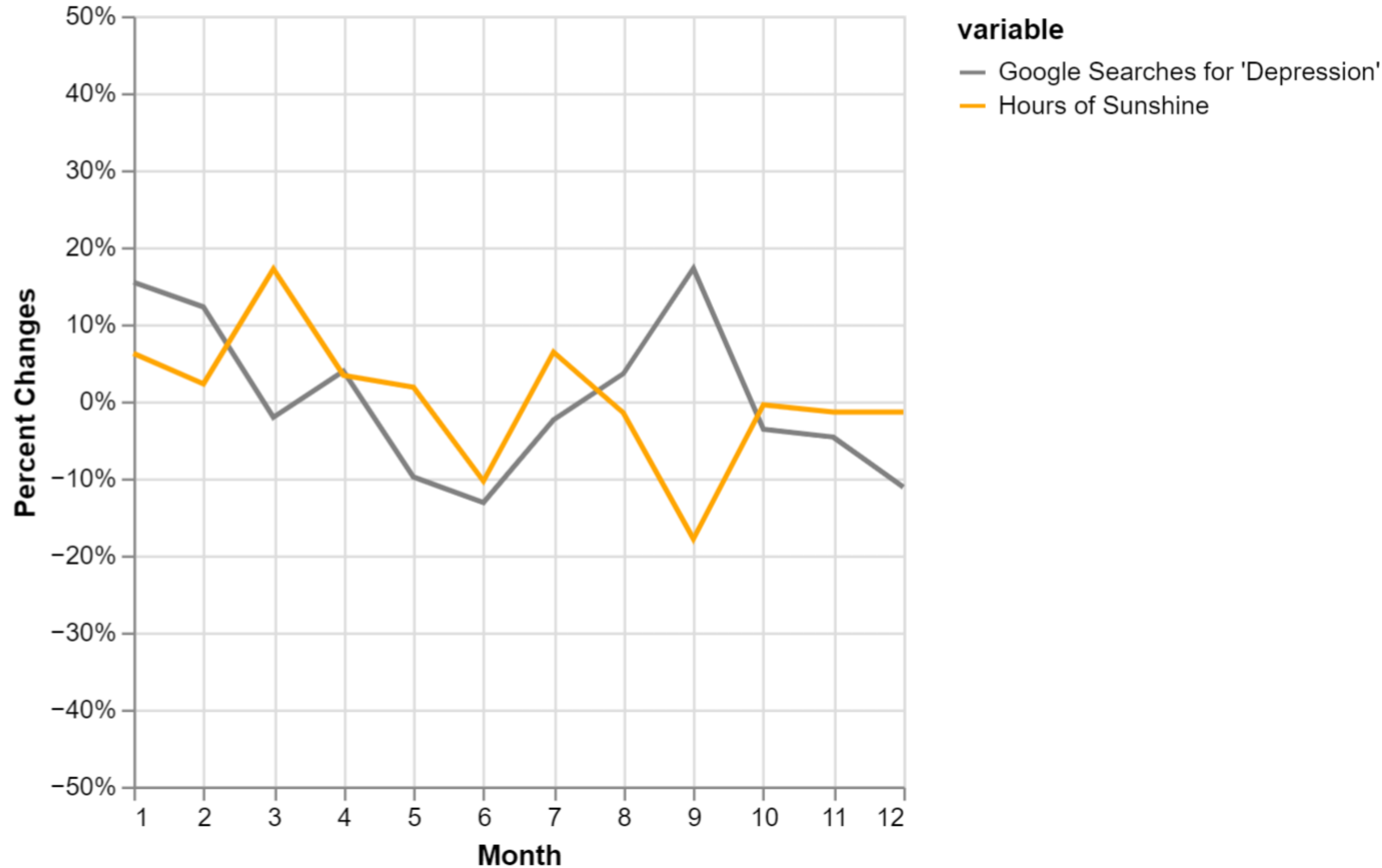
A Tale of Two Cities: How Dramatic are the Effects of Seasonal Depression?

Compared to a city like Miami, where sunshine hardly changes year round, Google searches for depression fluctuates much more in Chicago.

Chicago



Miami

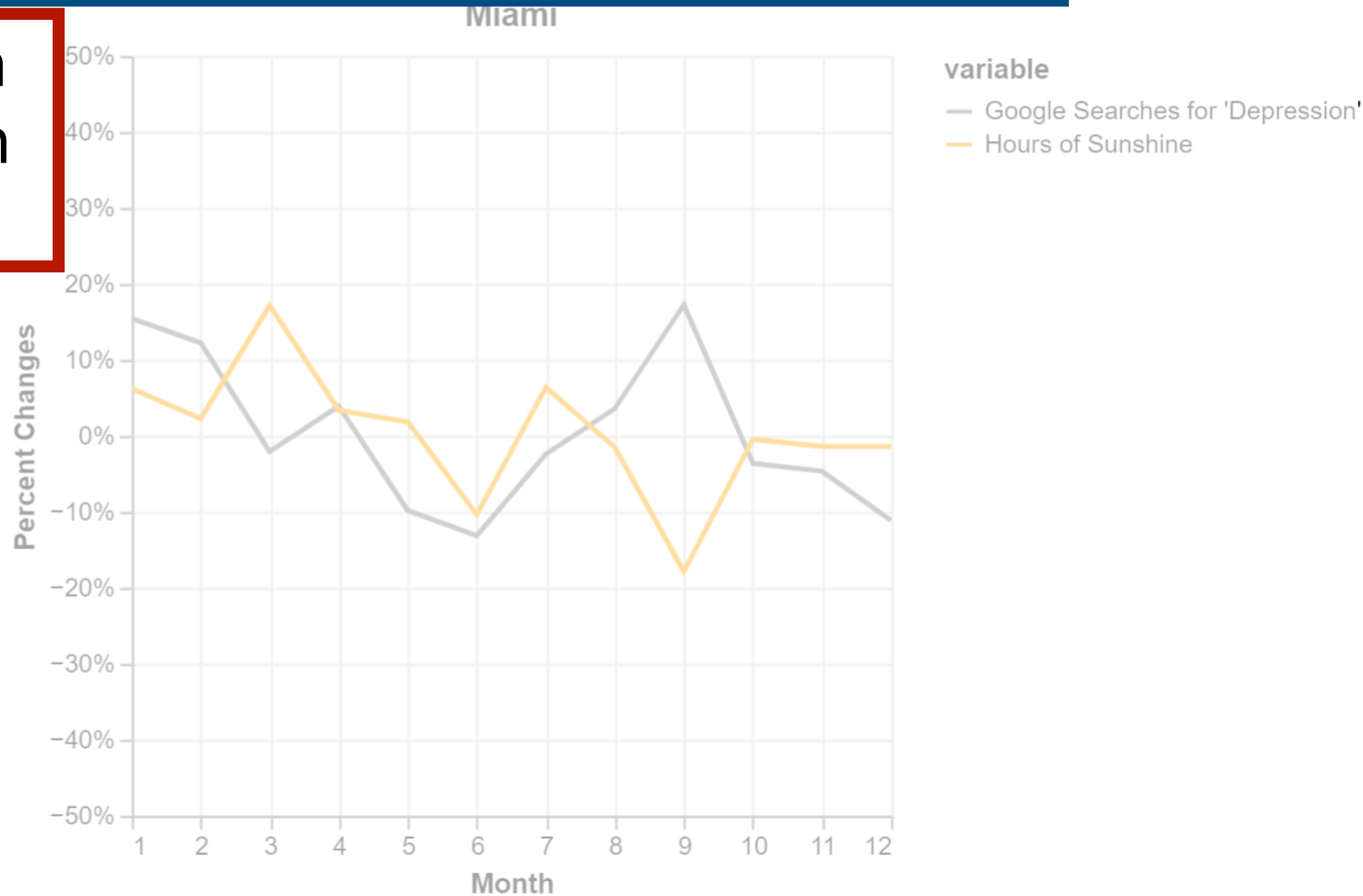
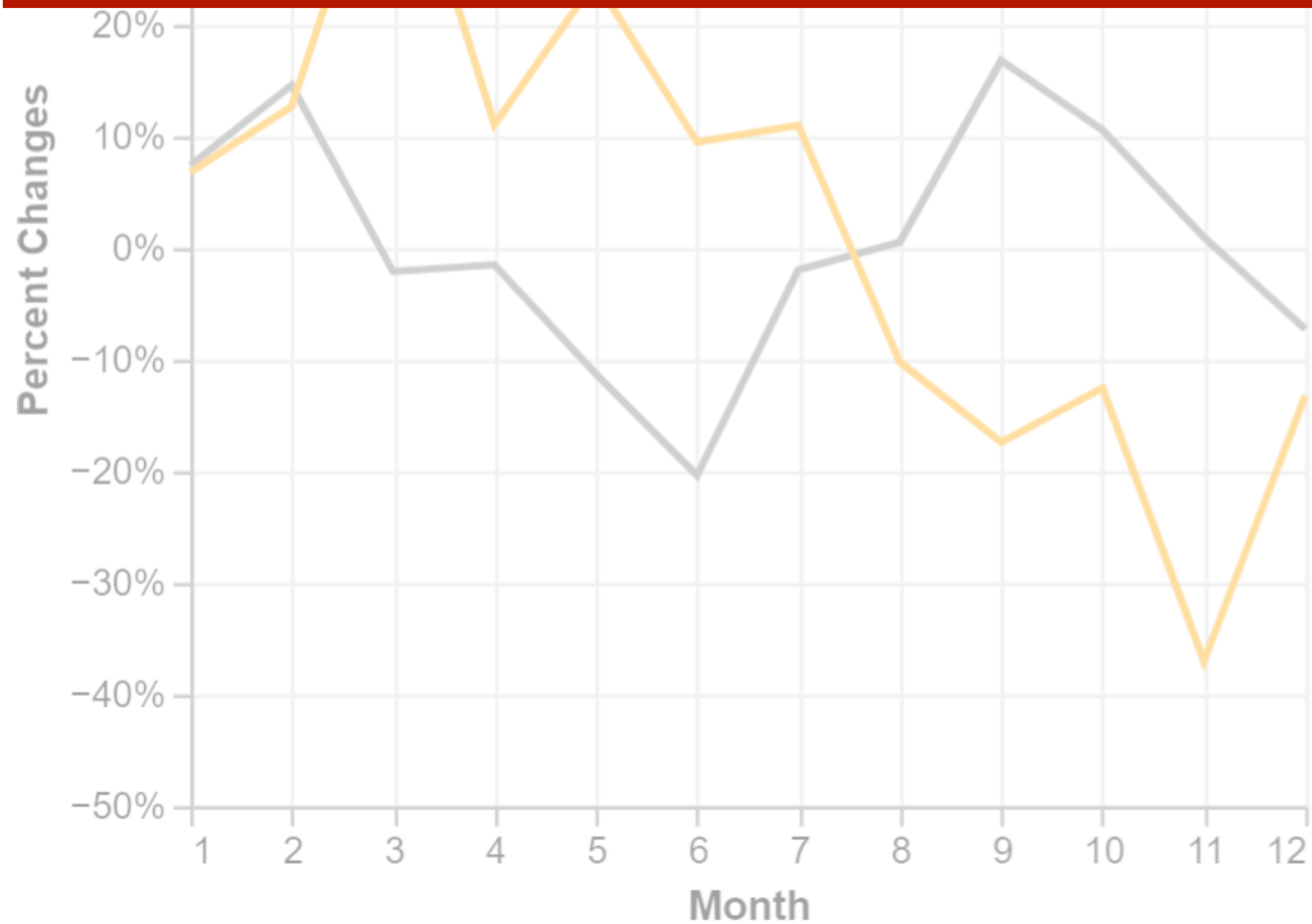


Then, tell a story

A Tale of Two Cities: How Dramatic are the Effects of Seasonal Depression?

Compared to a city like Miami, where sunshine hardly changes year round, Google searches for depression fluctuates much more in Chicago.

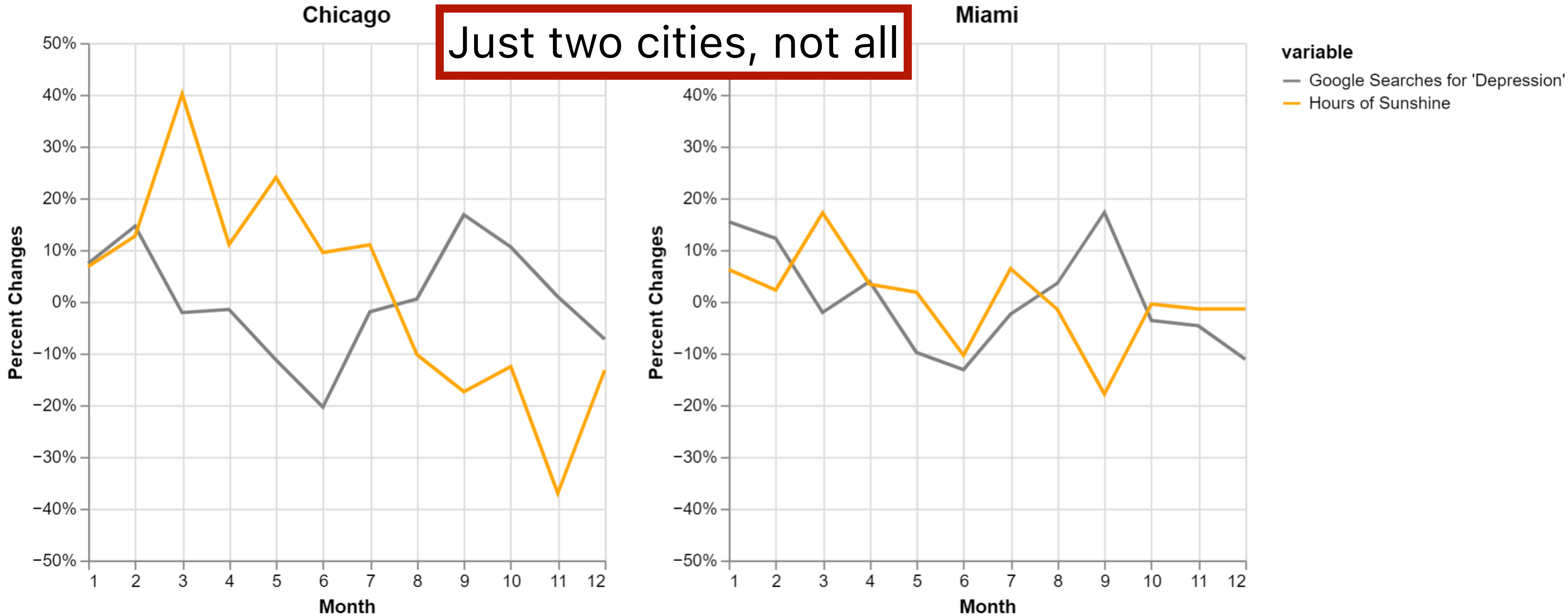
Missing takeaway / question in title was most common reason for docking points last year



Then, tell a story

A Tale of Two Cities: How Dramatic are the Effects of Seasonal Depression?

Compared to a city like Miami, where sunshine hardly changes year round, Google searches for depression fluctuates much more in Chicago.

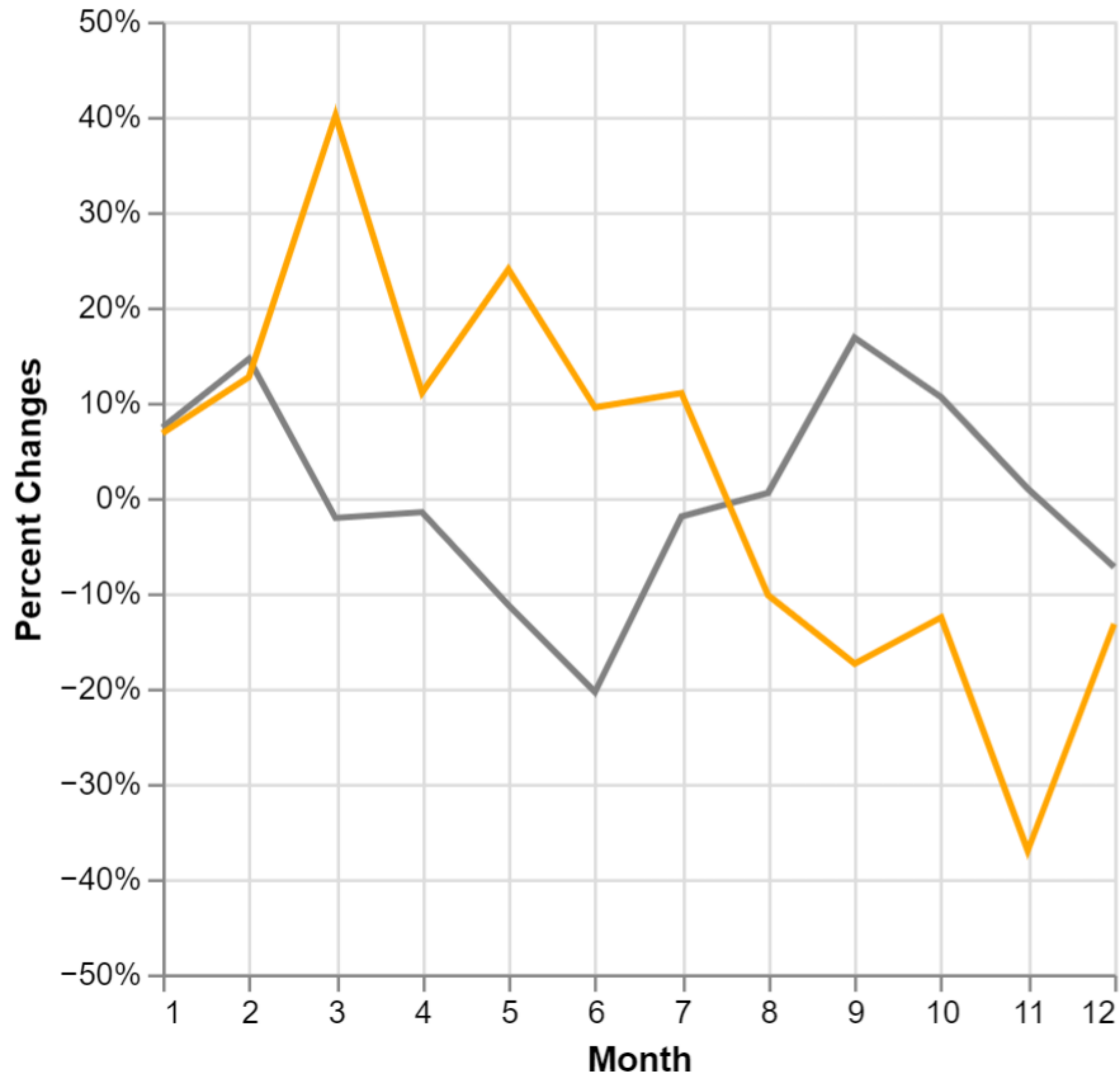


Then, tell a story

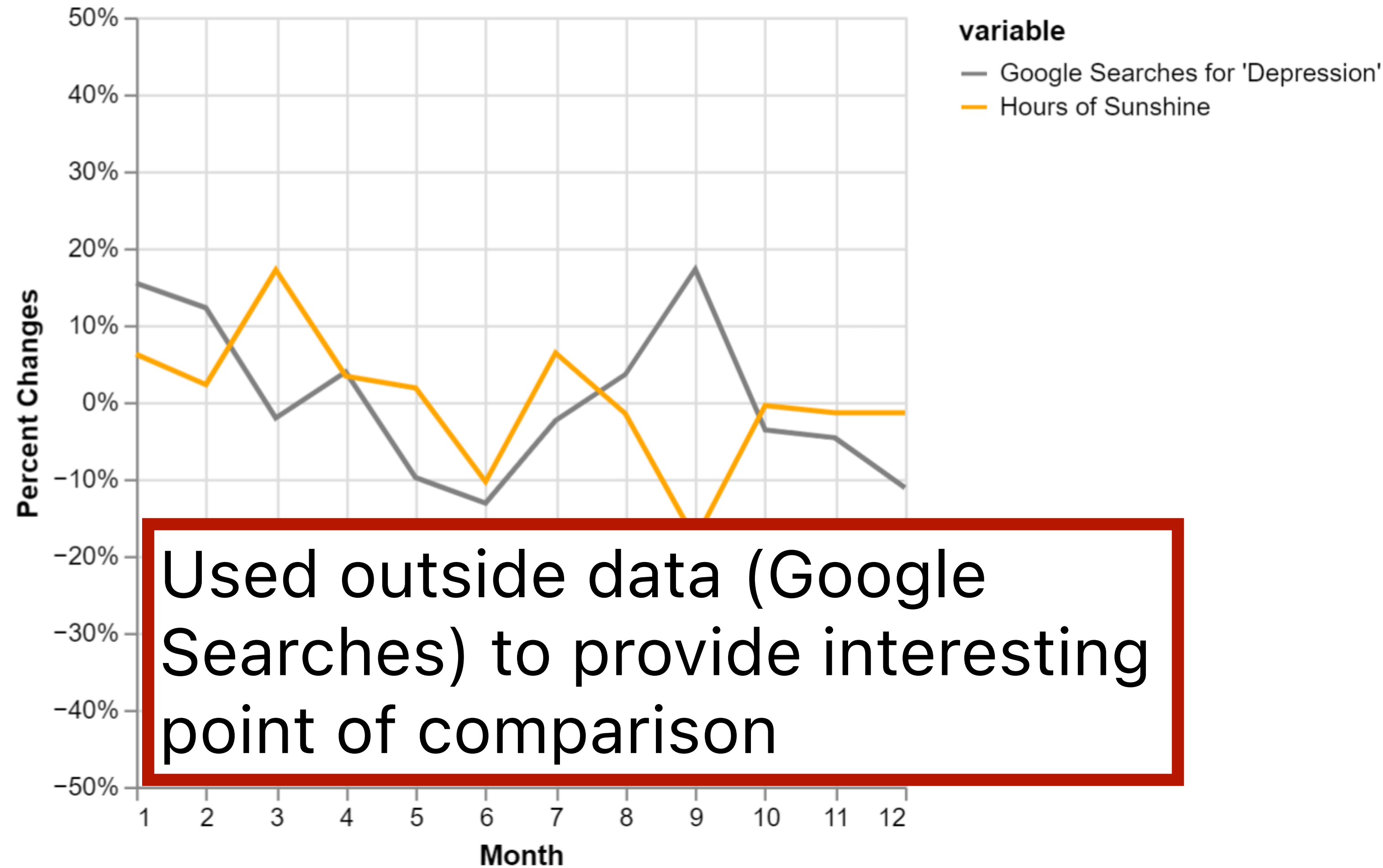
A Tale of Two Cities: How Dramatic are the Effects of Seasonal Depression?

Compared to a city like Miami, where sunshine hardly changes year round, Google searches for depression fluctuates much more in Chicago.

Chicago



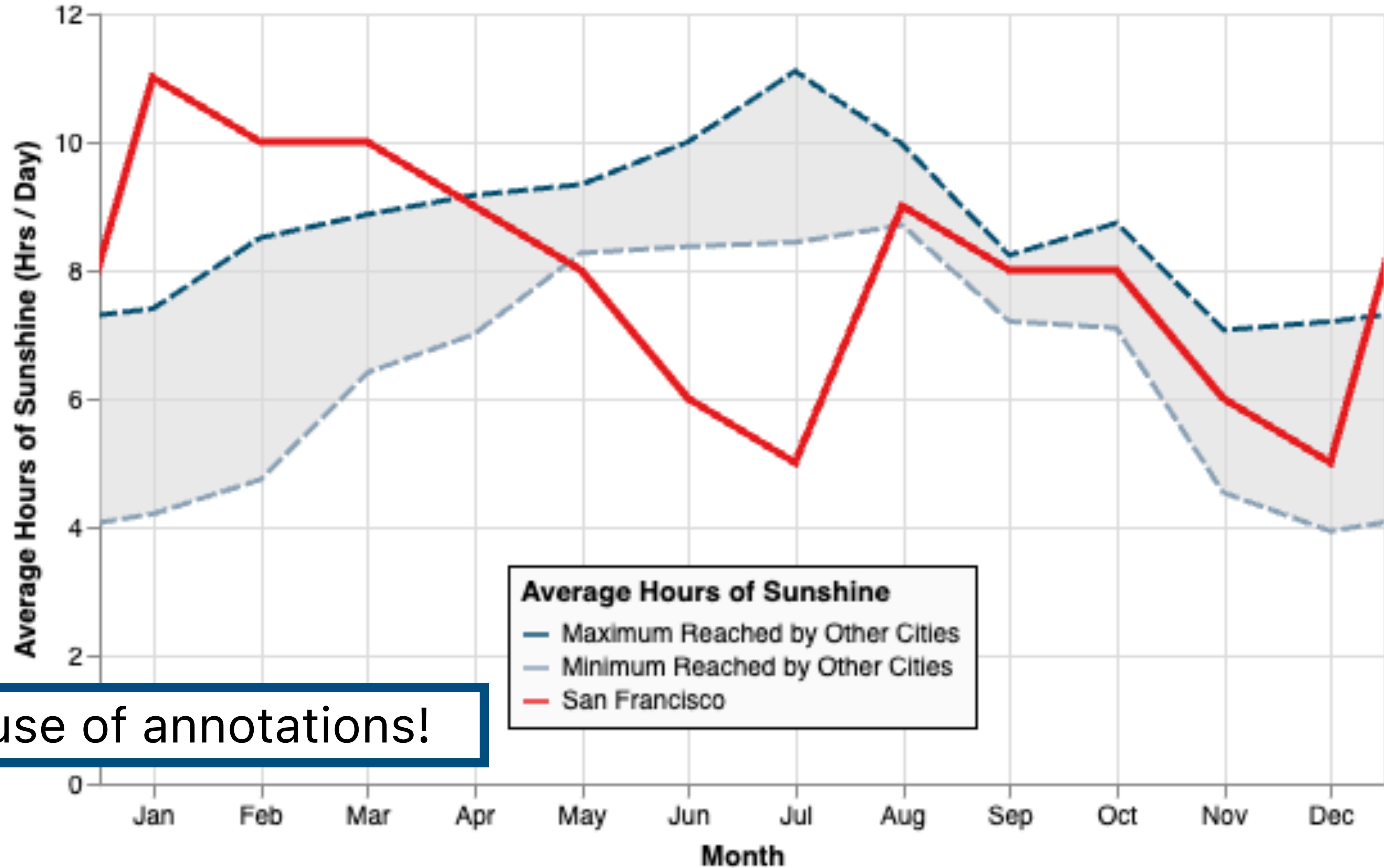
Miami



Used outside data (Google Searches) to provide interesting point of comparison

How Does San Francisco's Sunshine Trend Compare to Other Major U.S. Cities?

San Francisco's Peculiar Sunshine Trend

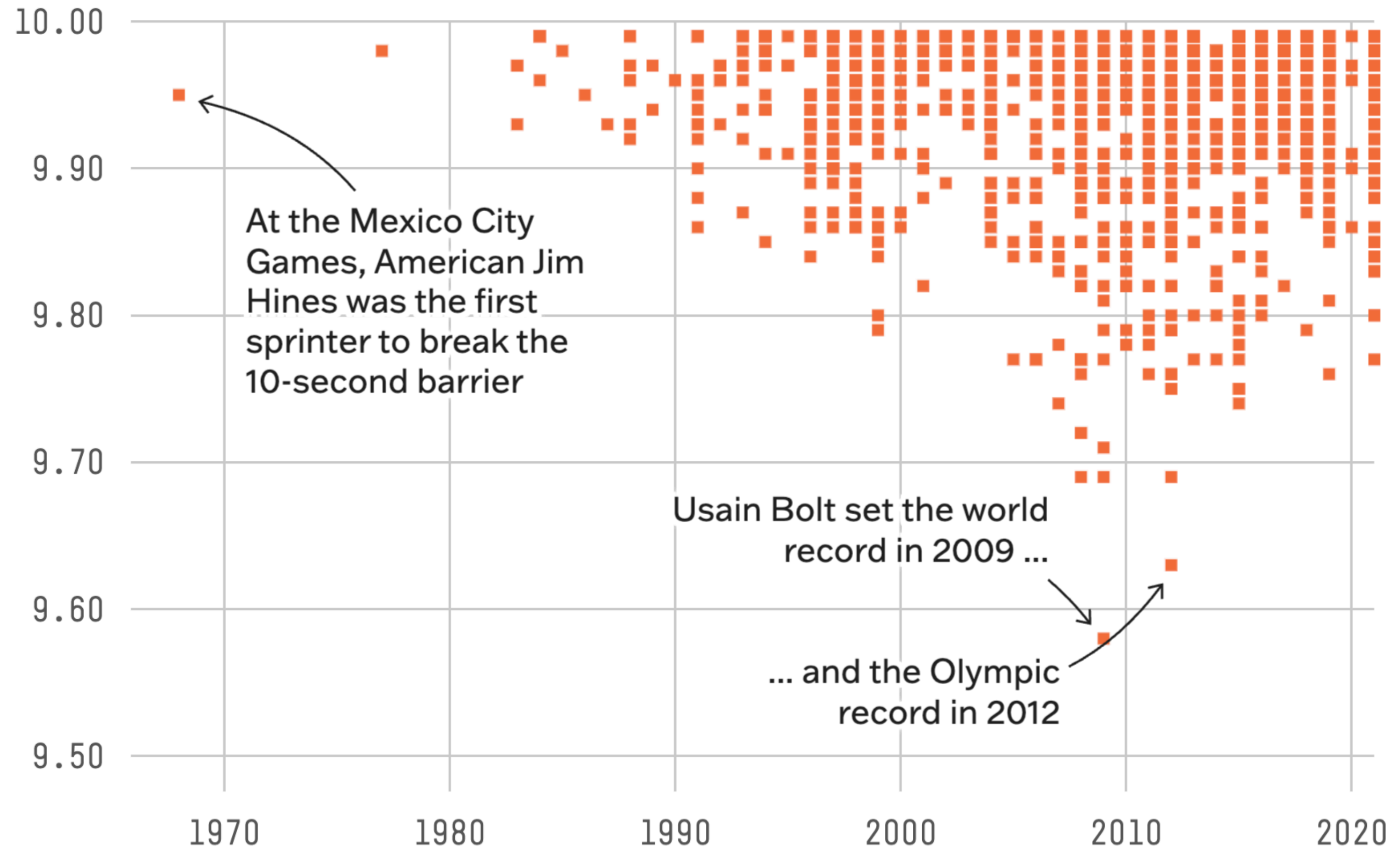


Great use of annotations!

Exemplar: Usain Bolt

No one is coming close to Usain Bolt's best times

All times under 10 seconds in the outdoor men's 100-meter sprint, using only electronic readings and under regular wind conditions

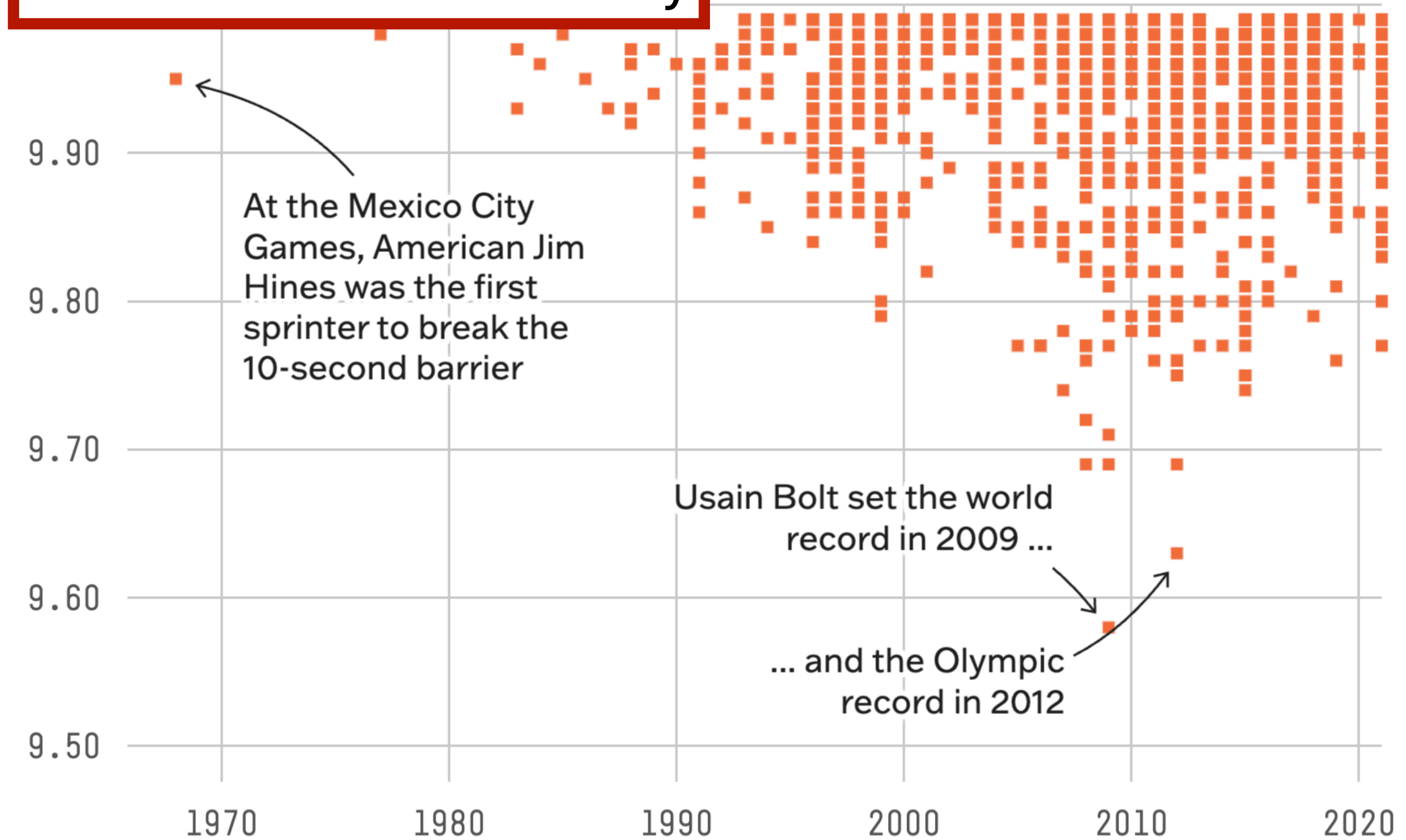


Exemplar: Usain Bolt

No one is coming close to Usain Bolt's best times

Time versus year for 100-meter sprints

Grade: Not satisfactory



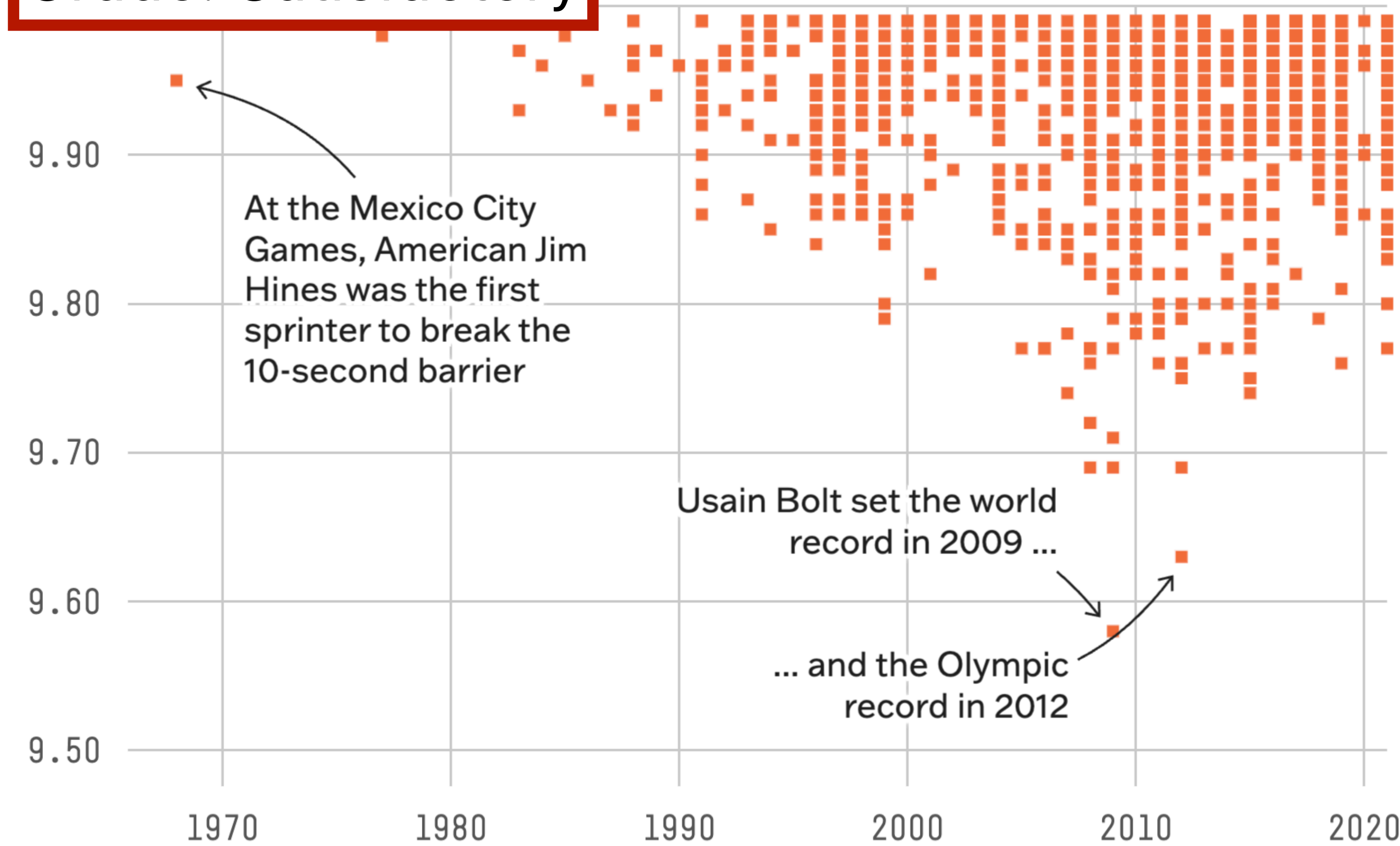
Exemplar: Usain Bolt

No one is coming close to Usain Bolt's best times

How have 100m sprint times changed over time?

only electronic readings and under regular wind conditions

Grade: Satisfactory

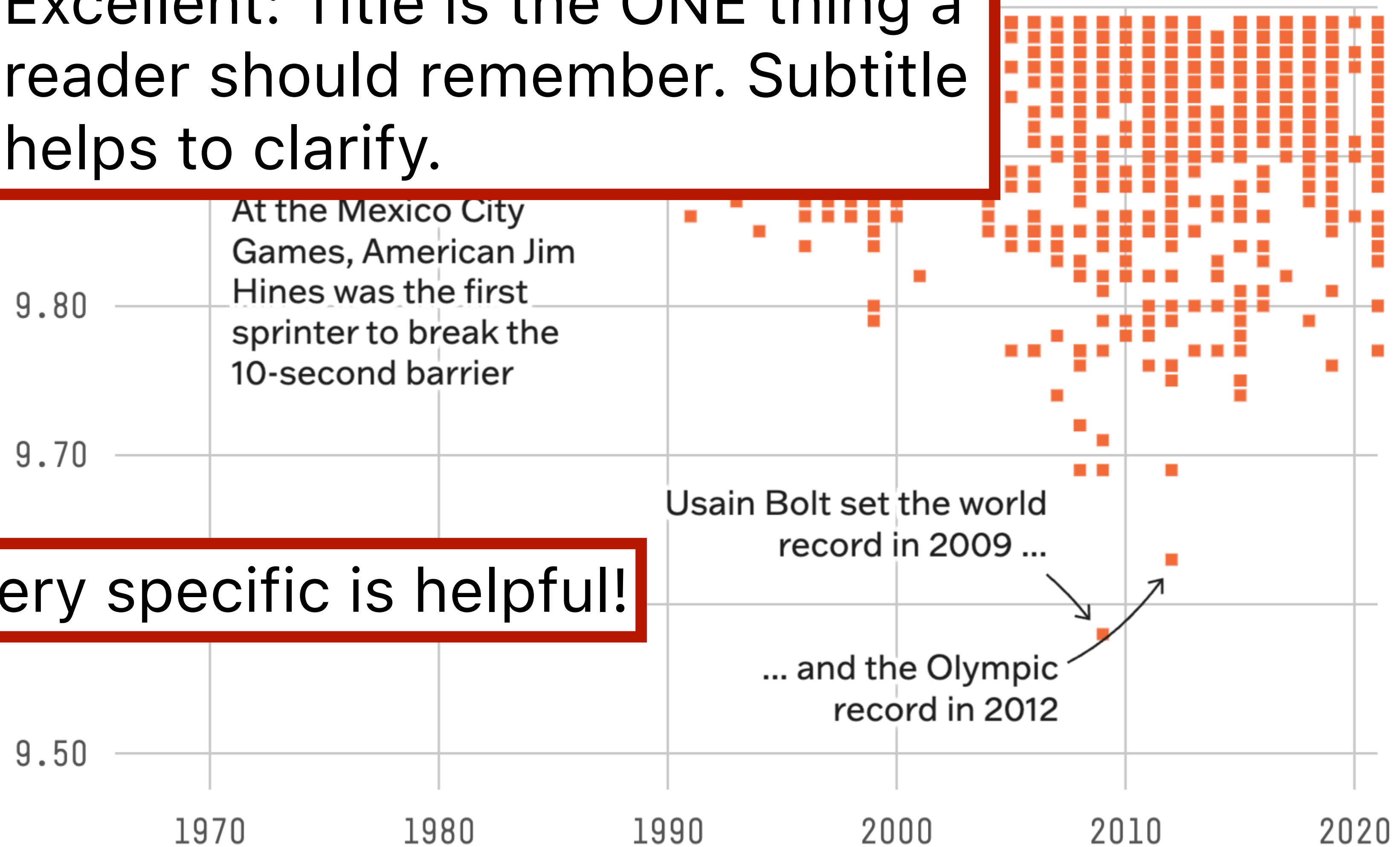


Exemplar: Usain Bolt

No one is coming close to Usain Bolt's best times

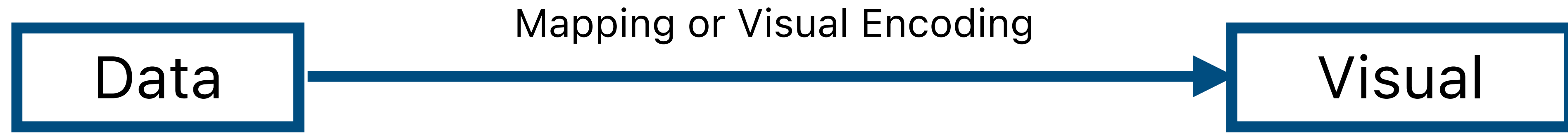
All times under 10 seconds in the outdoor men's 100-meter sprint, using only electronic readings and under regular wind conditions

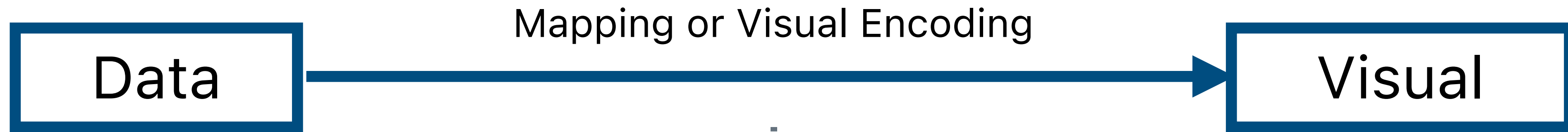
Excellent: Title is the ONE thing a reader should remember. Subtitle helps to clarify.



Often, being very specific is helpful!

Visual Encodings





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

Mapping or Visual Encoding

Data

Visual



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)

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape



Marks

Area Bar Point

Line Arc Text

- 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.

Name that ~~chart!~~

Visual Encoding!

Percent of working-age people who said they had "serious difficulty" with ...



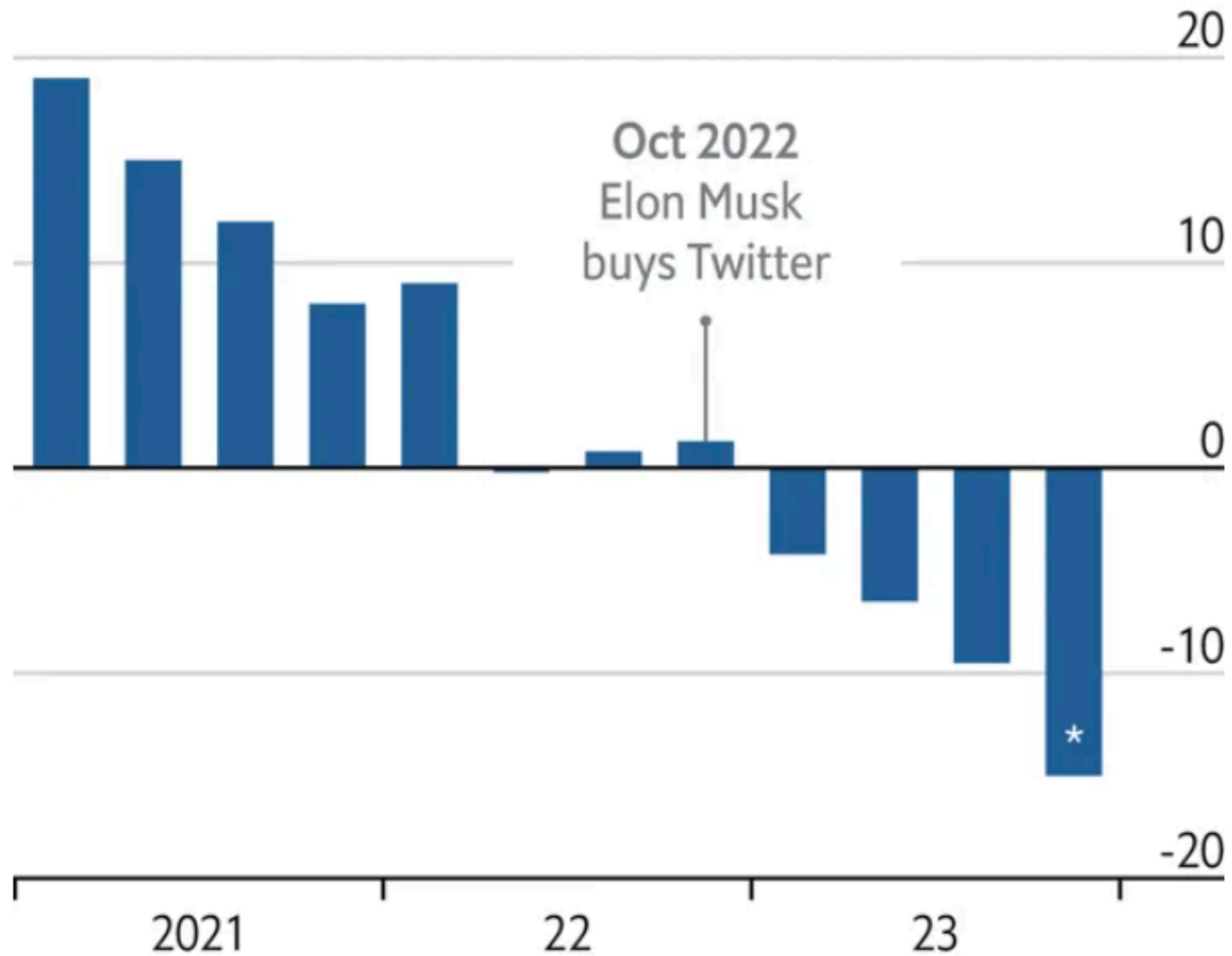
Mark: line
X-axis: date (Q-interval)
Y-axis: percent (Q-ratio)

What about color?

Drop off

Estimated monthly active Twitter/X users

% change on a year earlier



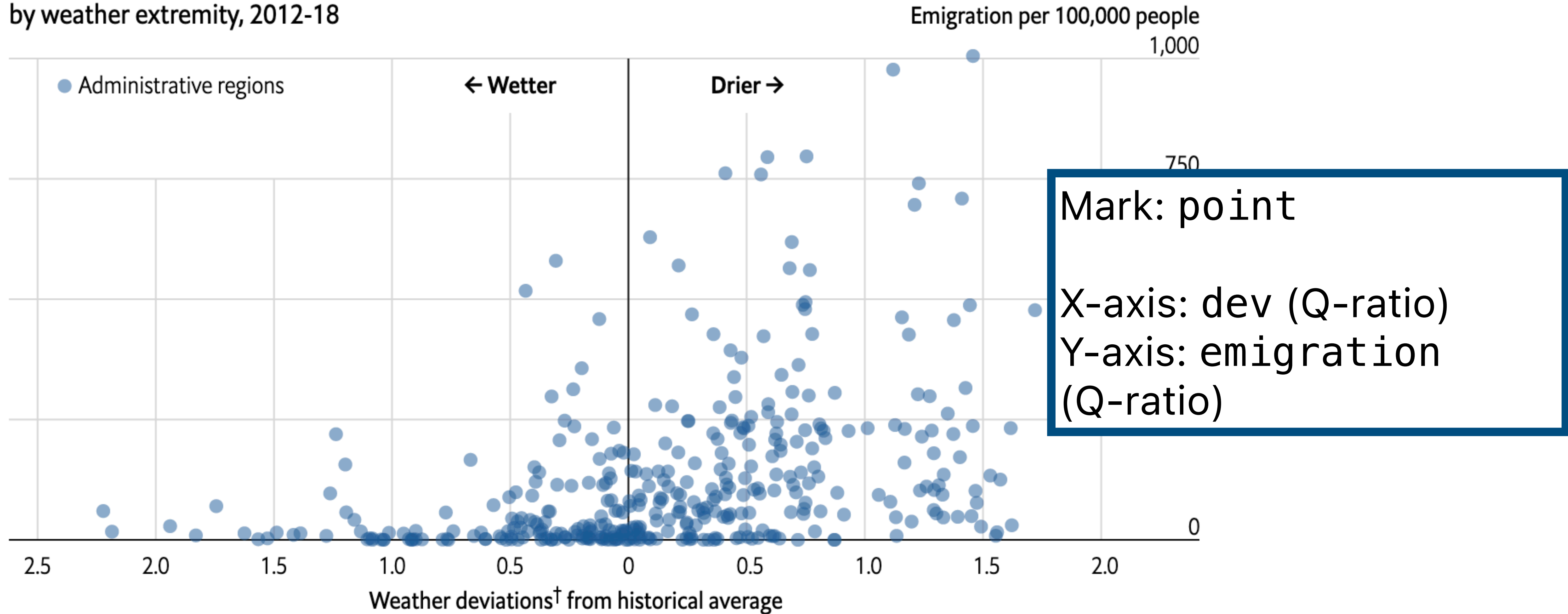
Mark: bar
X-axis: date (Q-interval)
Y-axis: percent (Q-ratio)

*To December 5th

Source: Sensor Tower

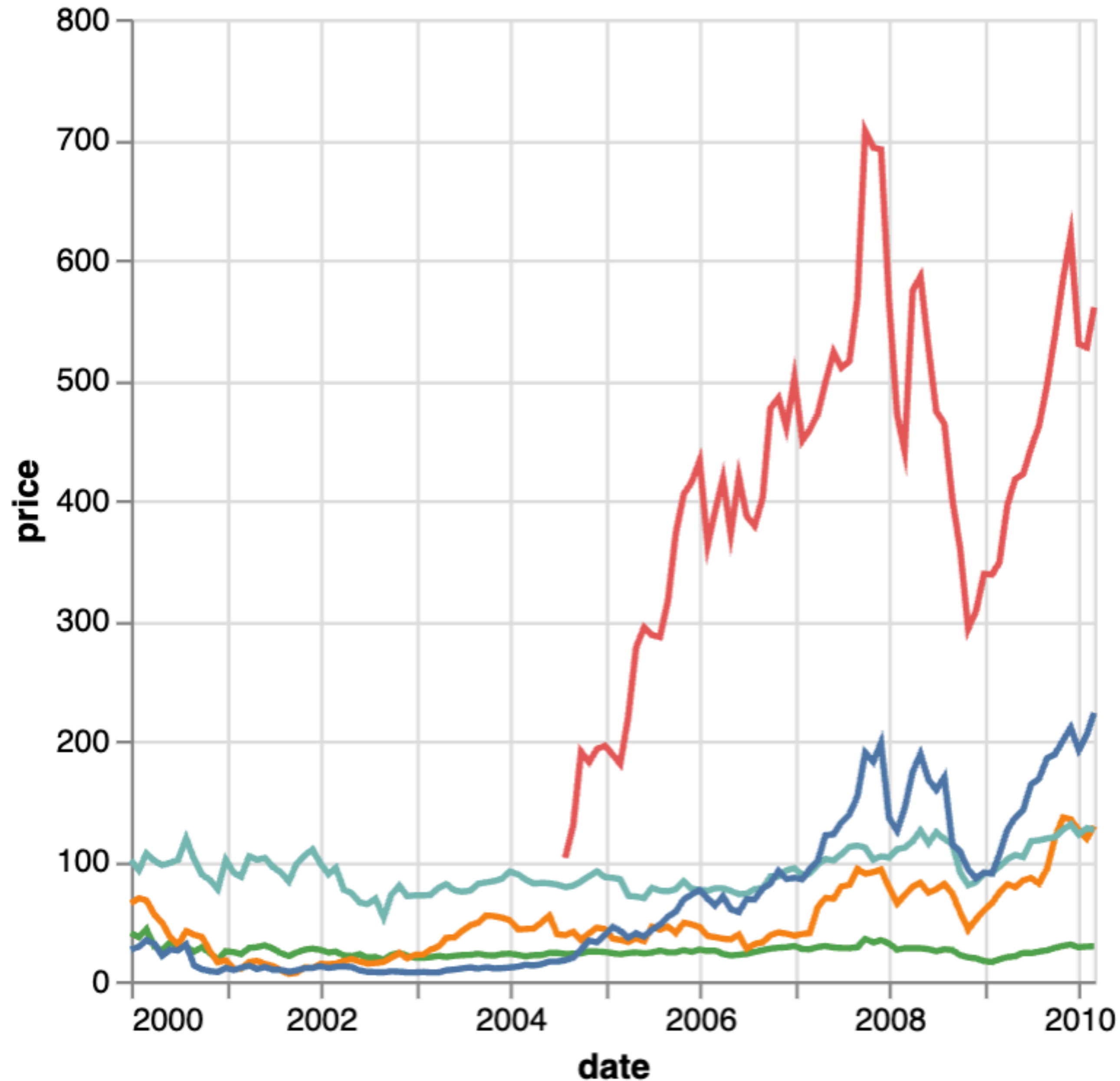
Spotting a trend

Emigration from the Northern Triangle* to United States, by weather extremity, 2012-18



*El Salvador, Guatemala and Honduras †Using the Standardised Precipitation-Evapotranspiration Index three-month average

Source: "Dry growing seasons predicted Central American migration to the US from 2012 to 2018", by A. Linke et al., 2023



symbol

— AAPL
— AMZN
— GOOG
— IBM
— MSFT

Mark: line

X-axis: date (Q-interval)

Y-axis: price (Q-ratio)

Color: symbol (N)

Notice the parallel with
plotly express syntax!

```
px.line(  
    stocks_df,  
    x='date',  
    y='price',  
    color='symbol',  
)
```

Actual win percentage

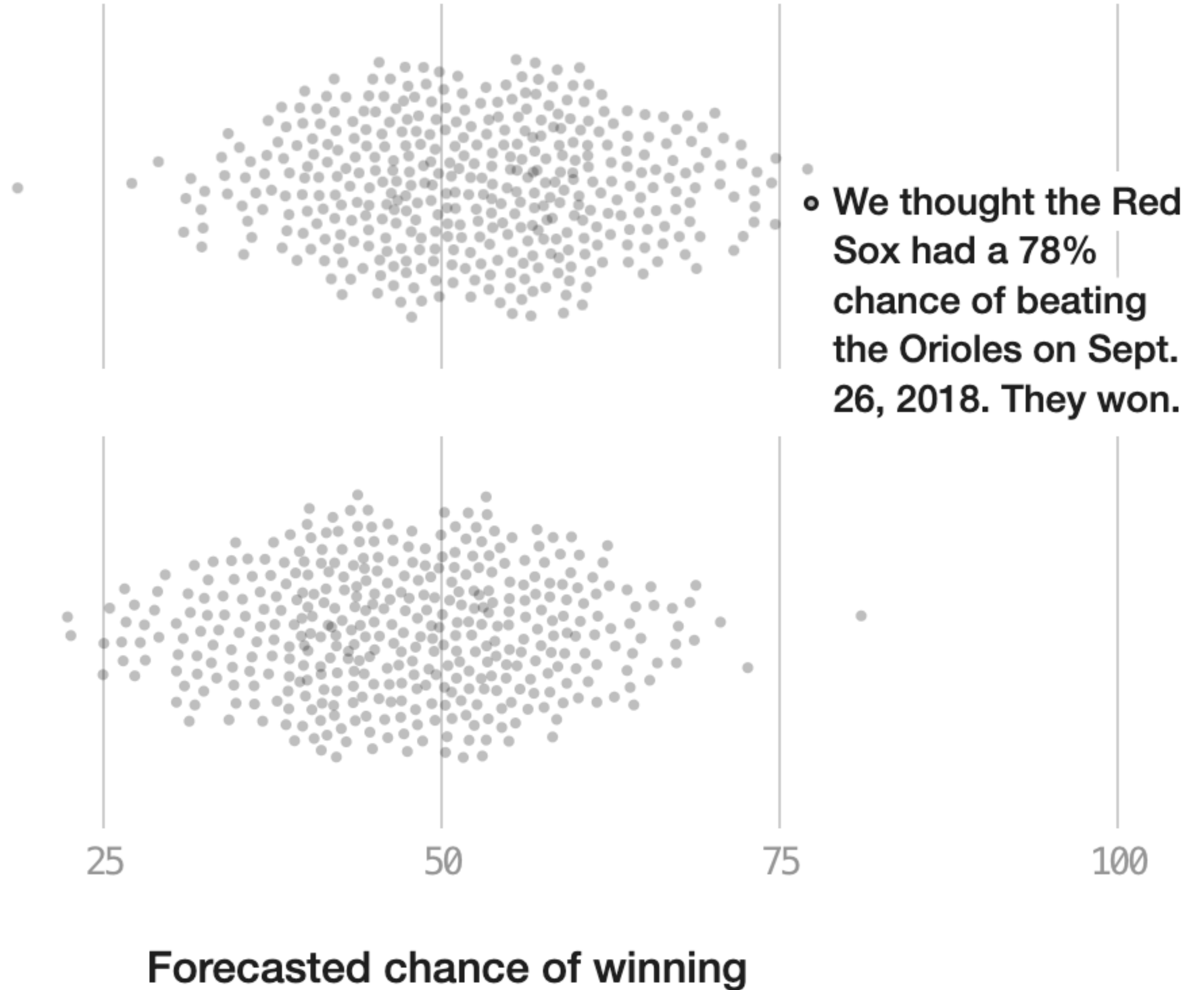
Team won
100%

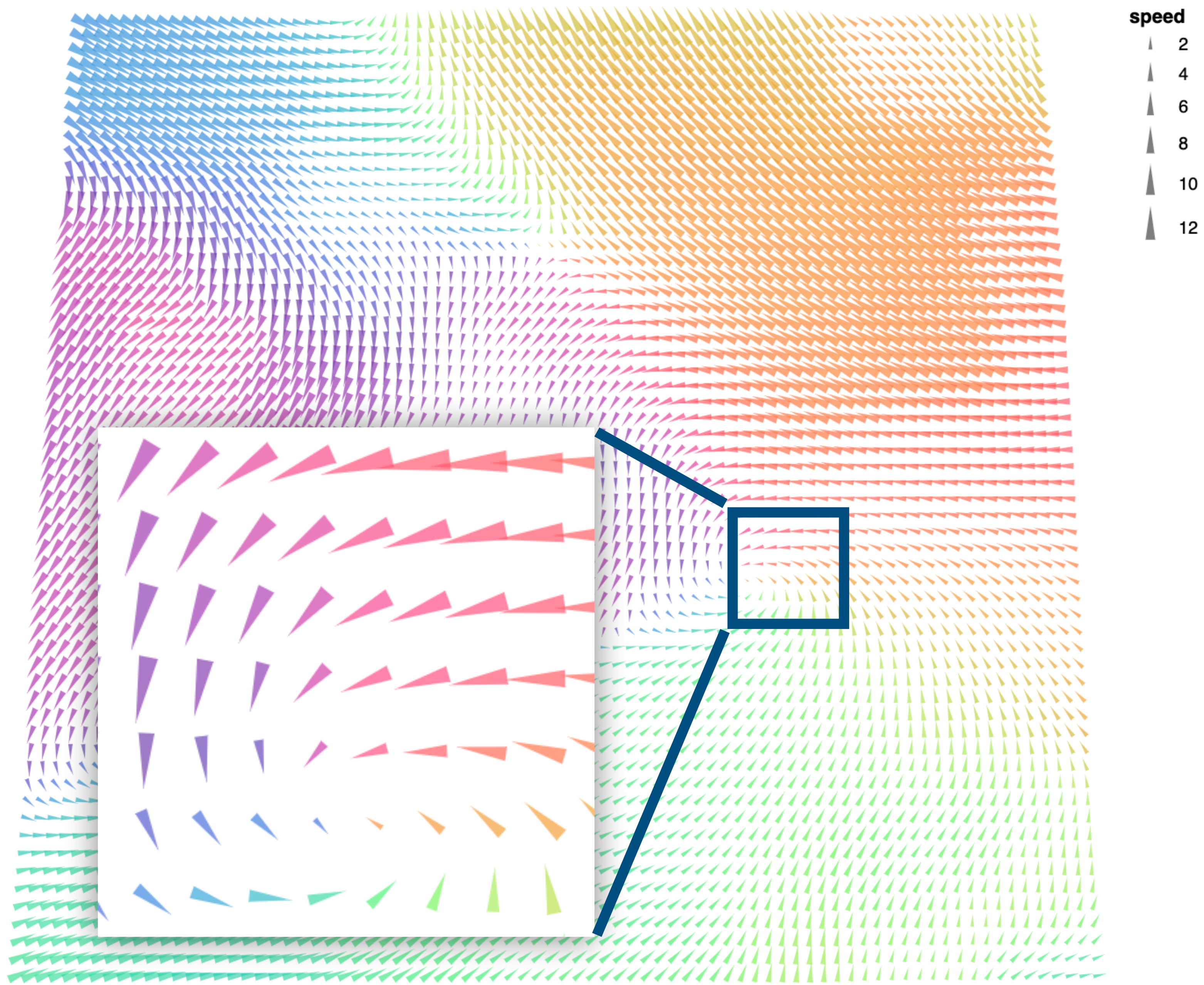
Team lost
0%

Mark: point

X-axis: chance (Q-ratio)

Y-axis: ?? (nothing!)





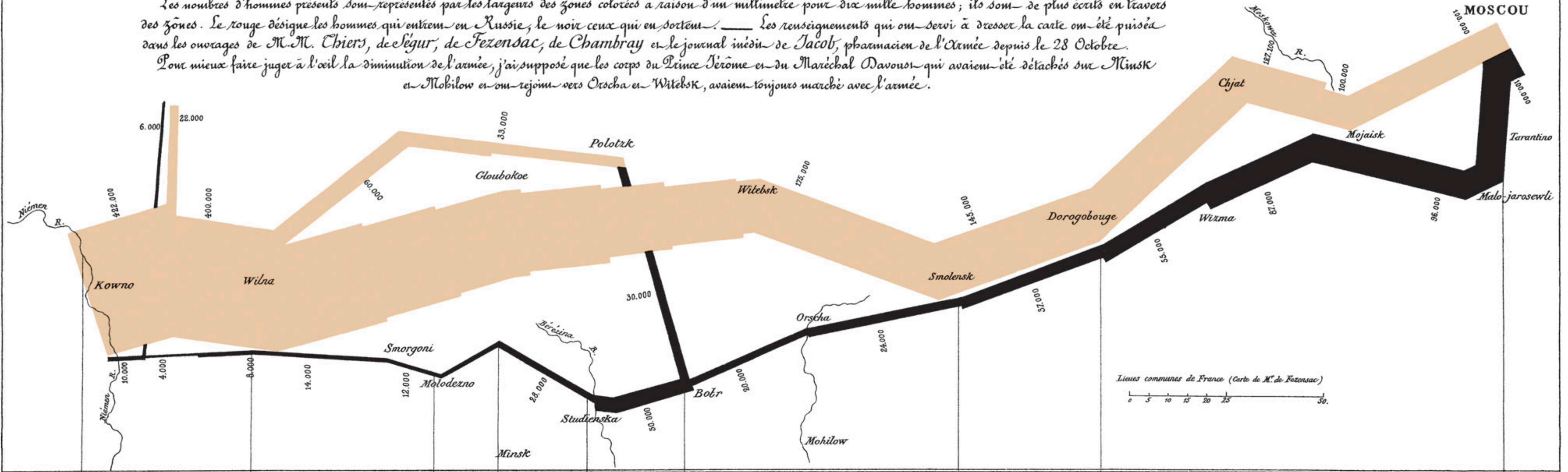
tryclassbuzz.com:
wind

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

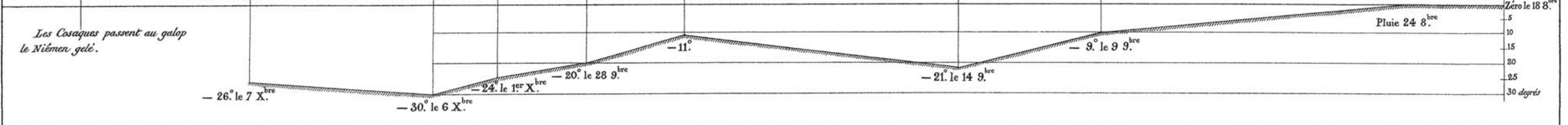
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et qui rejoignent vers Orscha et Witebsk, avaient toujours marché avec l'armée.



Lieux communs de France (Carte de M. de Fezensac)
0 5 10 15 20 25 30

TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



Les Cosaques passent au galop le Niemen gelé.

Autog. par Regnier, 8. Par. 5^{te} Marie 5^{te} G^{ne} à Paris.

Imp. Lith. Regnier et Dourdet.

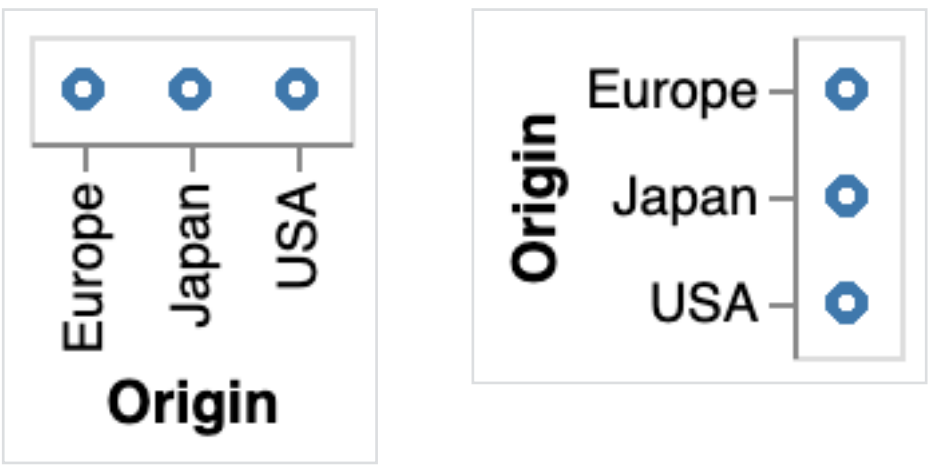
tryclassbuzz.com:
minard

A Design Space of Visual Encodings

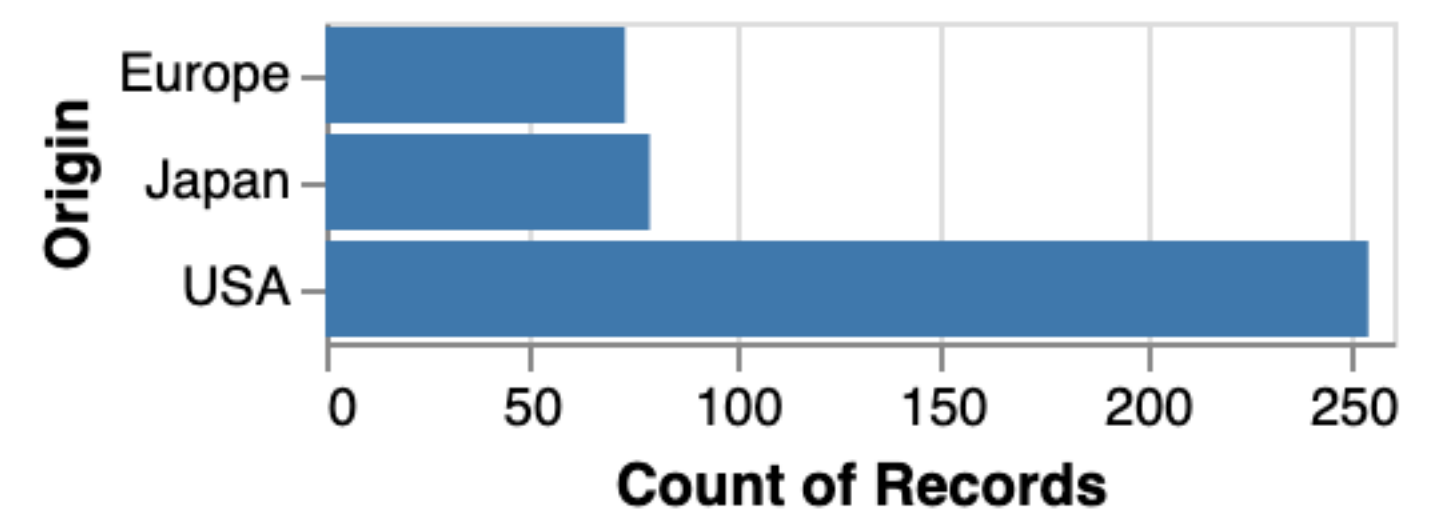
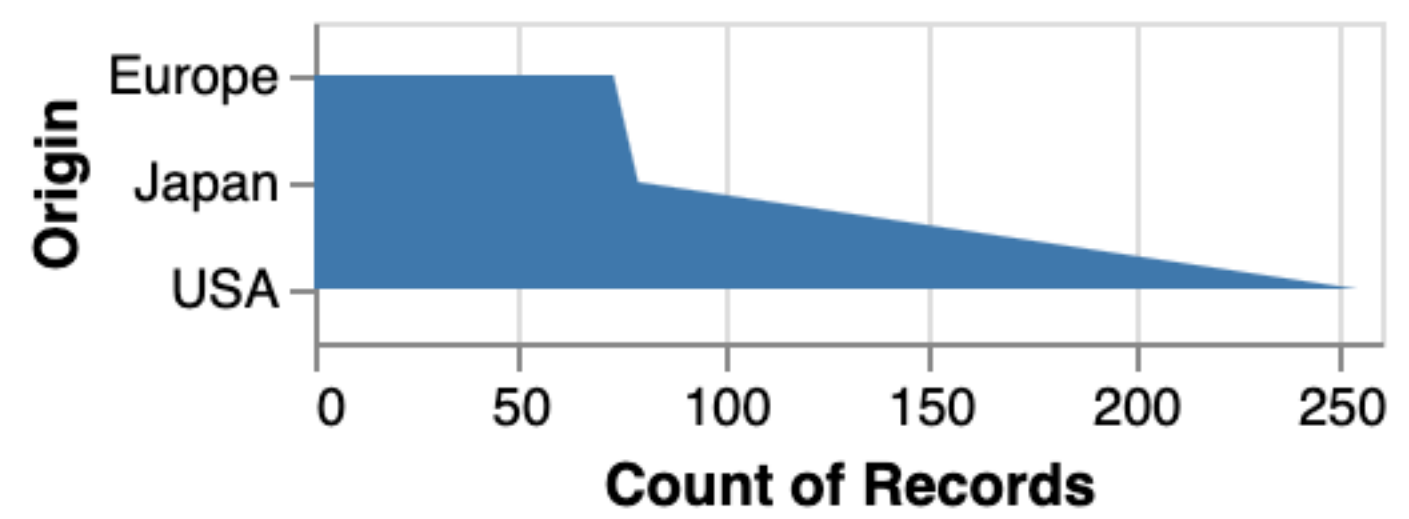
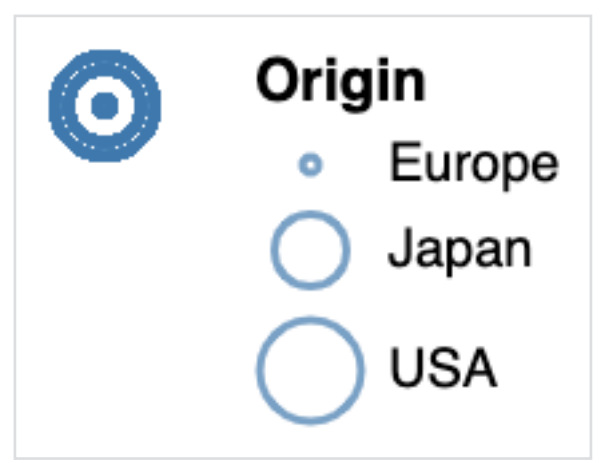
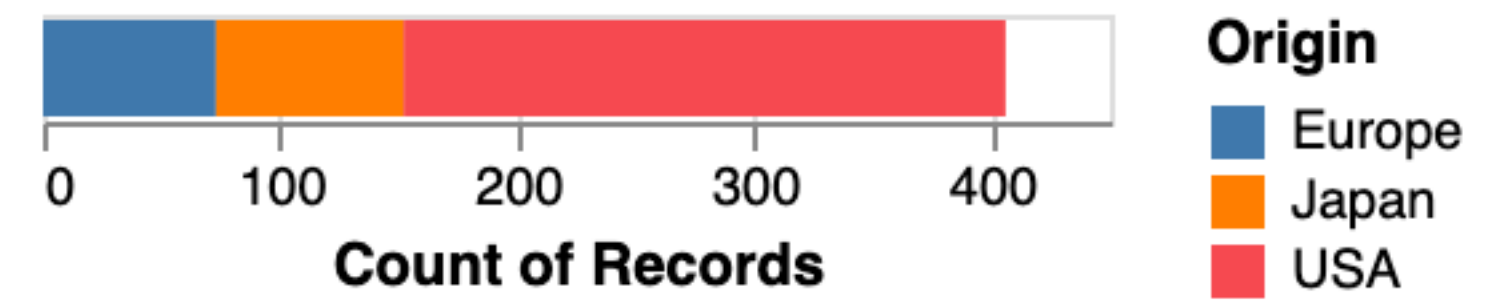
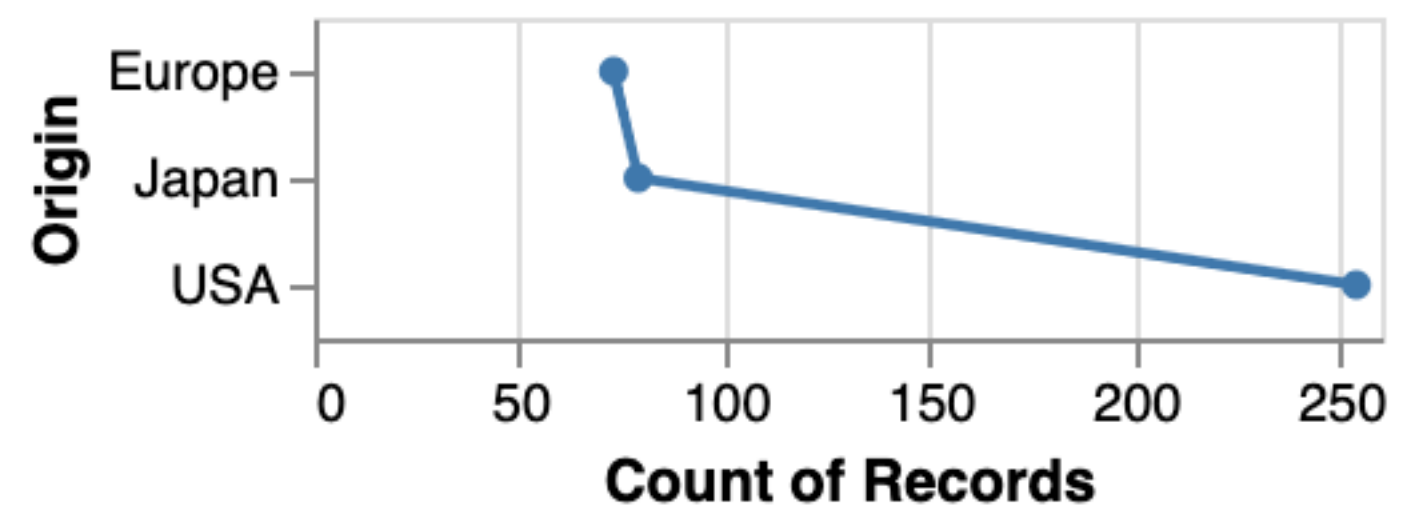
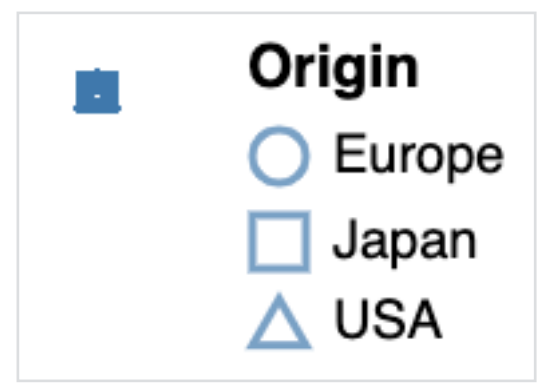
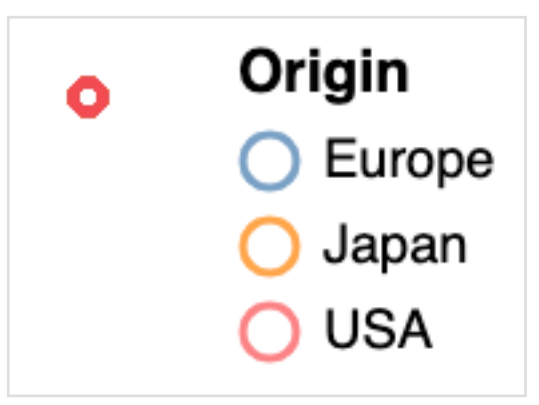
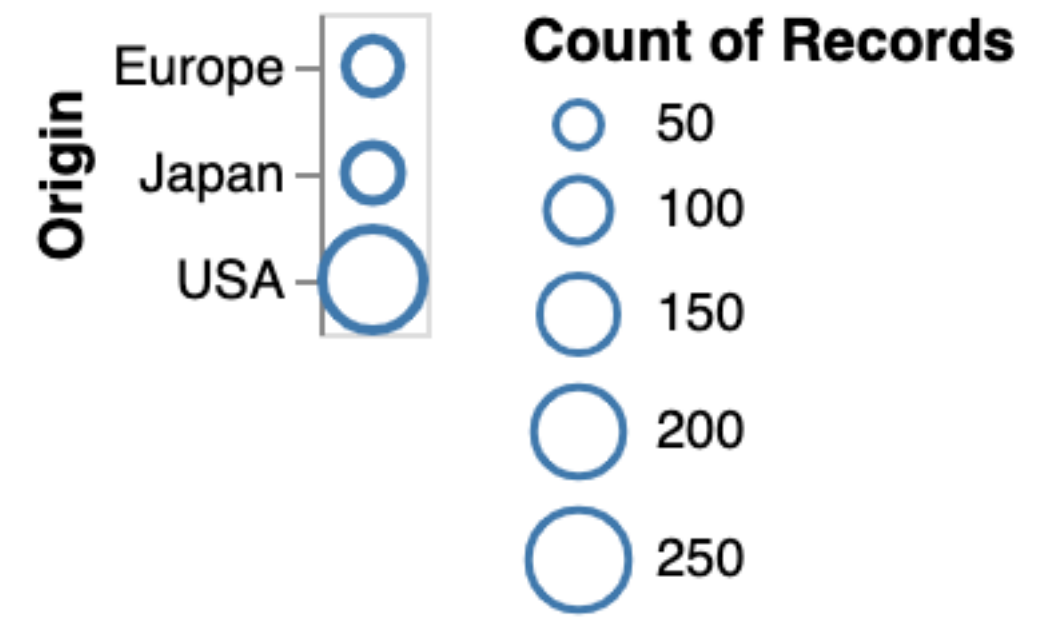
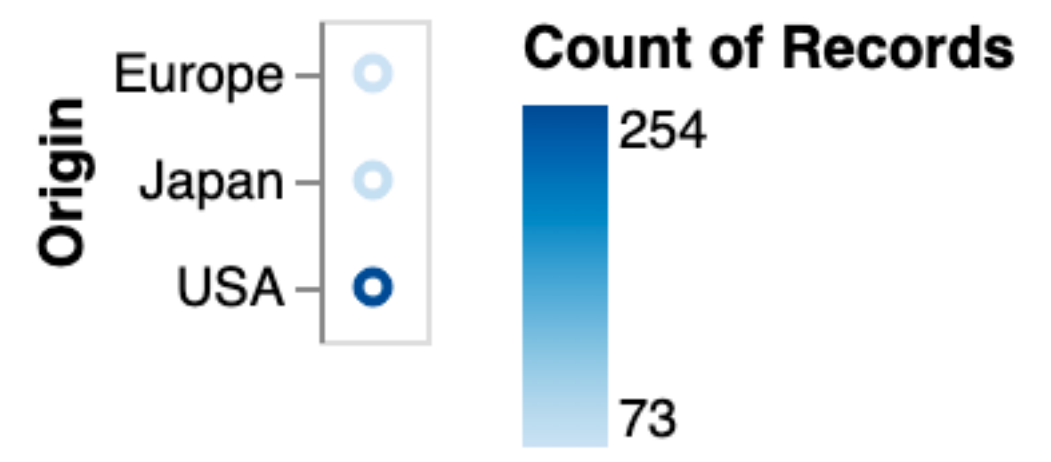
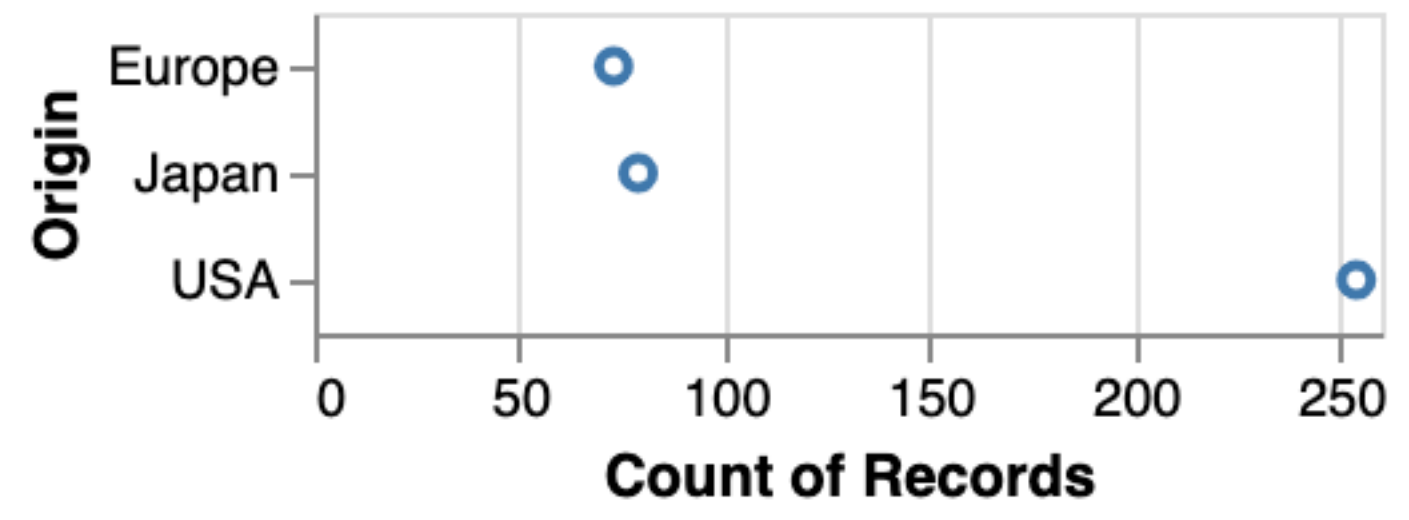
Visual Encoding = Combinatorial Design Space

1D nominal data (N, O)

raw



aggregate (count)



Visual Encoding = Combinatorial Design Space

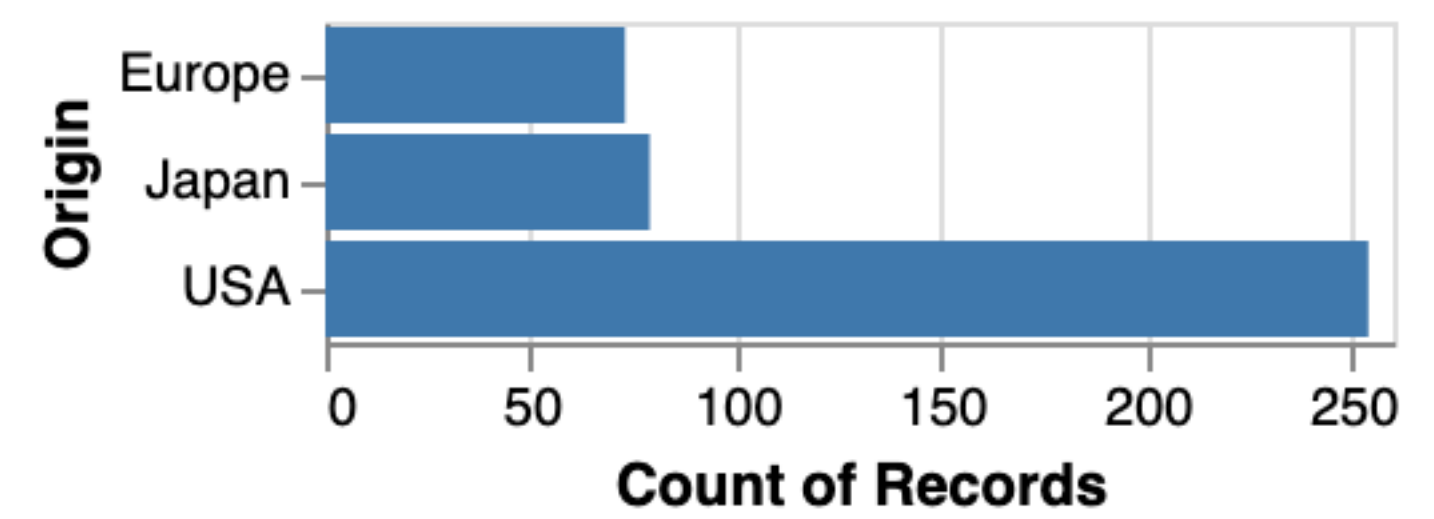
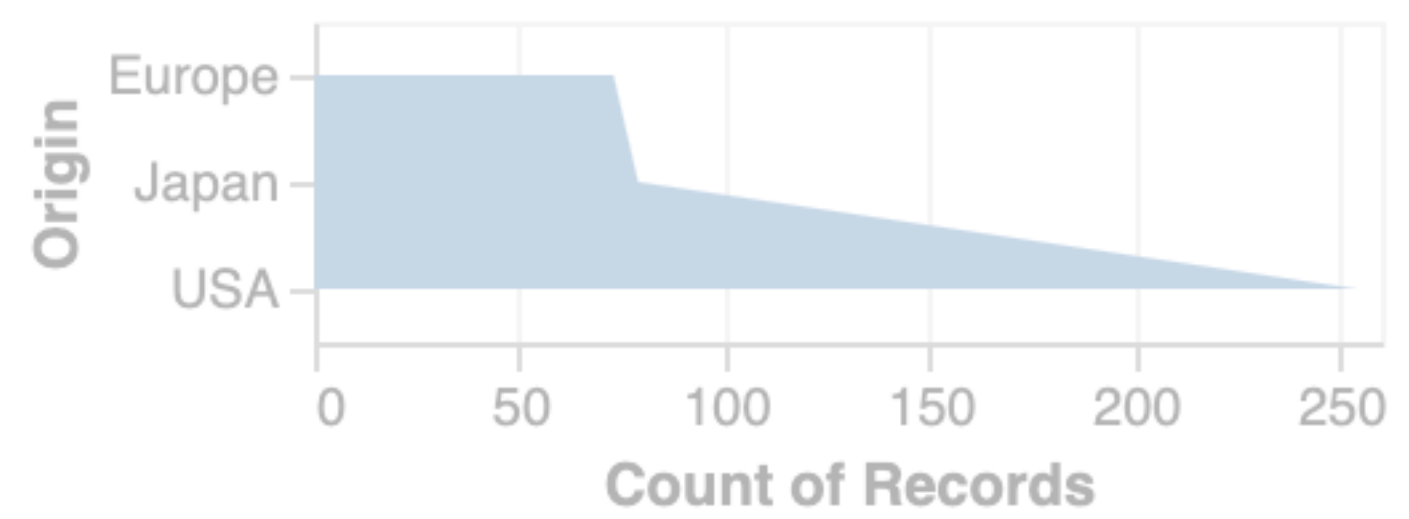
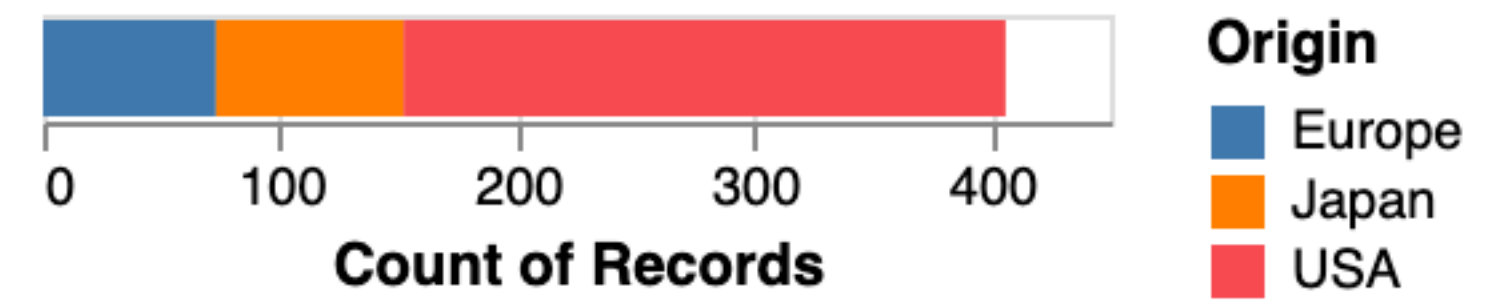
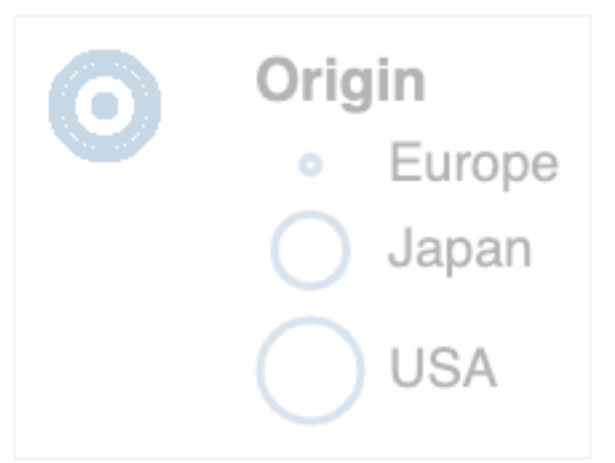
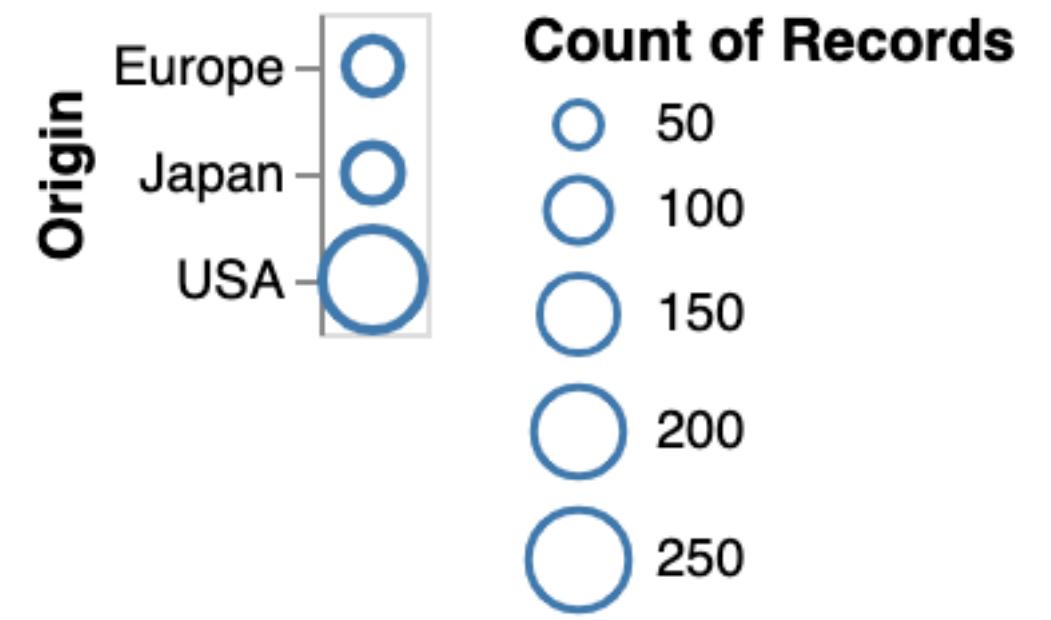
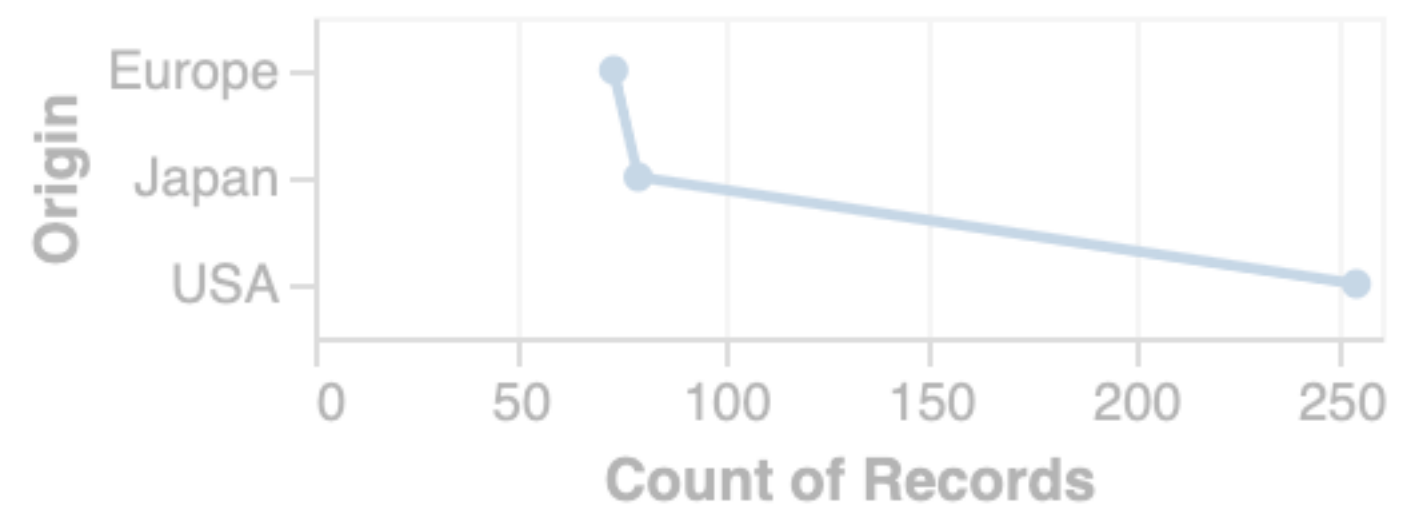
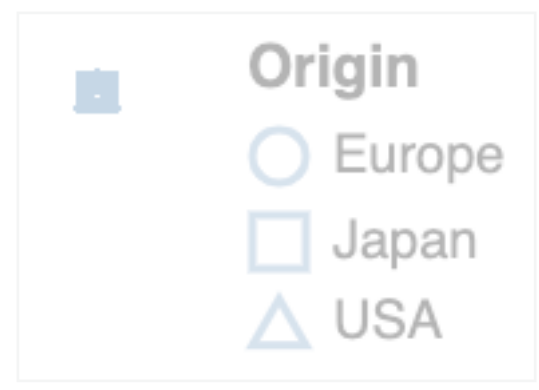
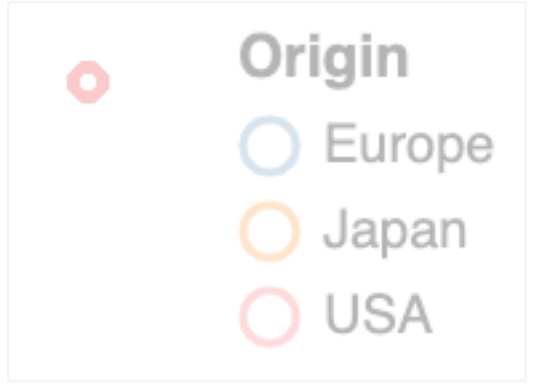
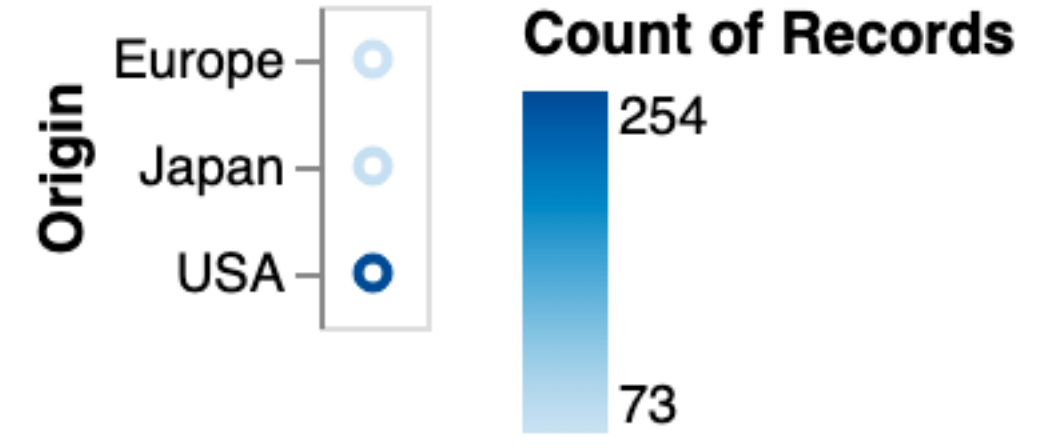
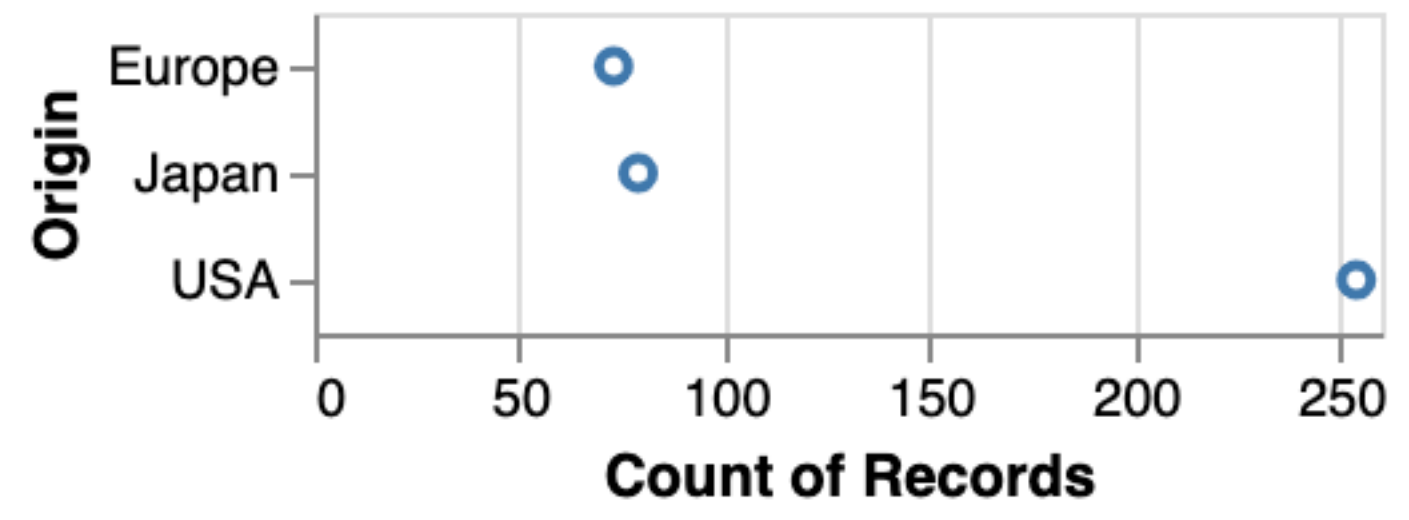
1D nominal data (N, O)

Expressive?

raw



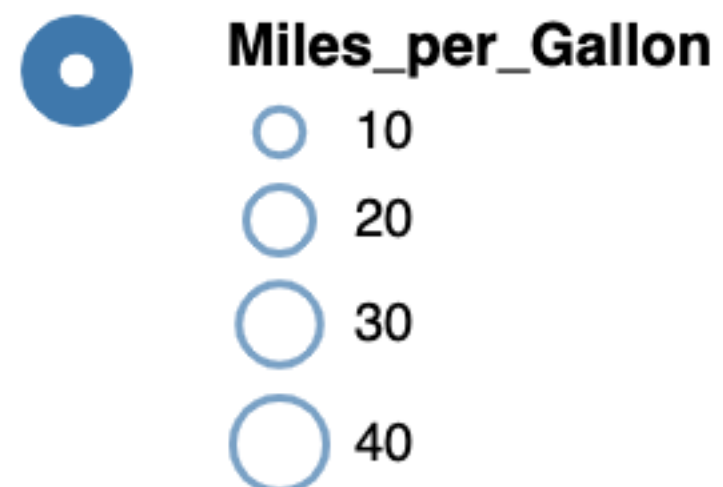
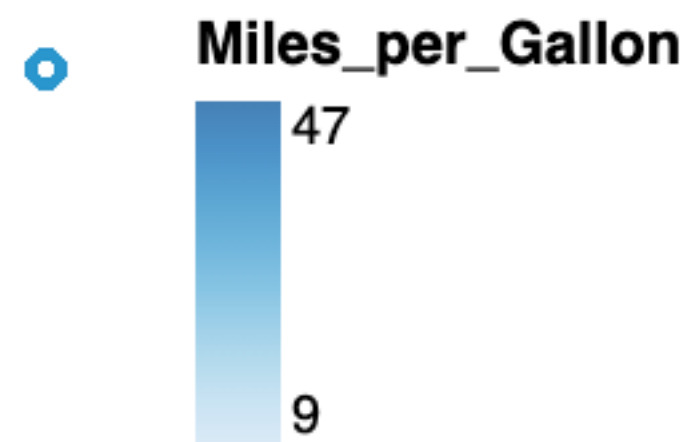
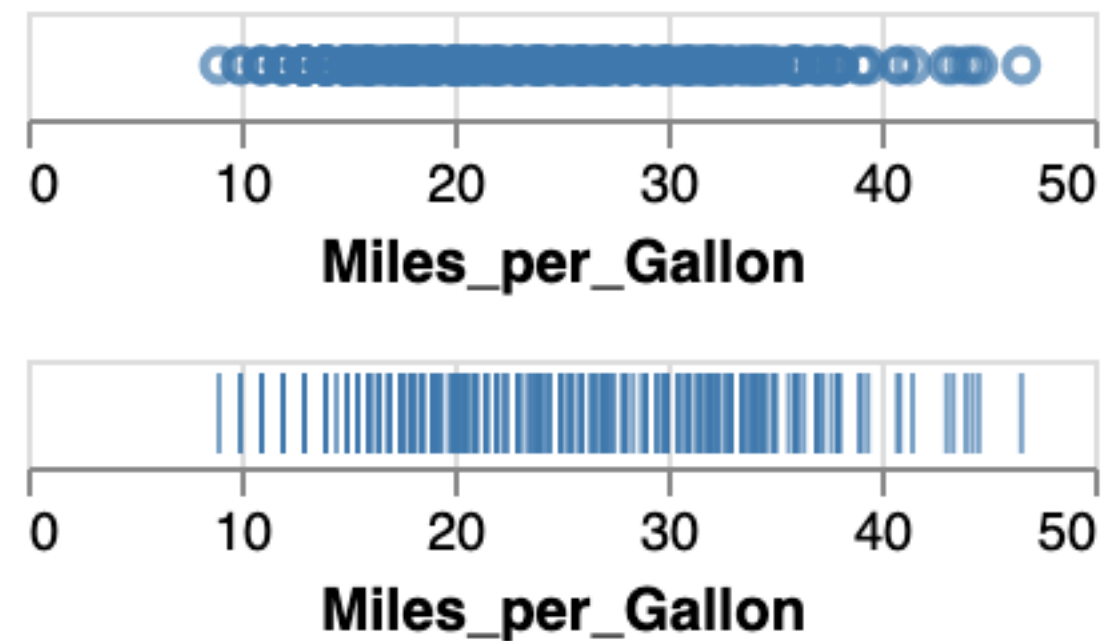
aggregate (count)



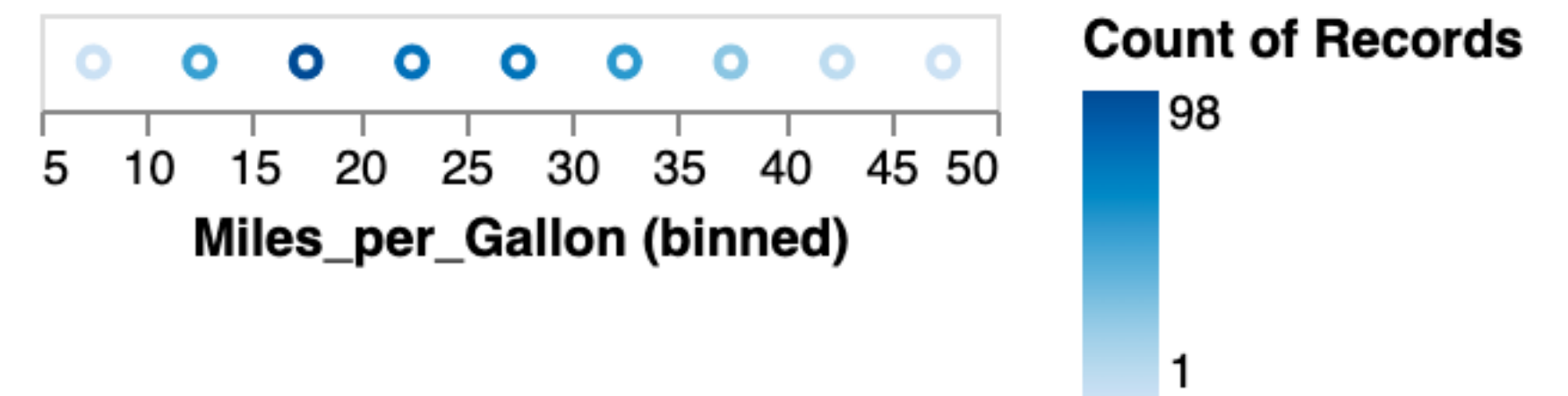
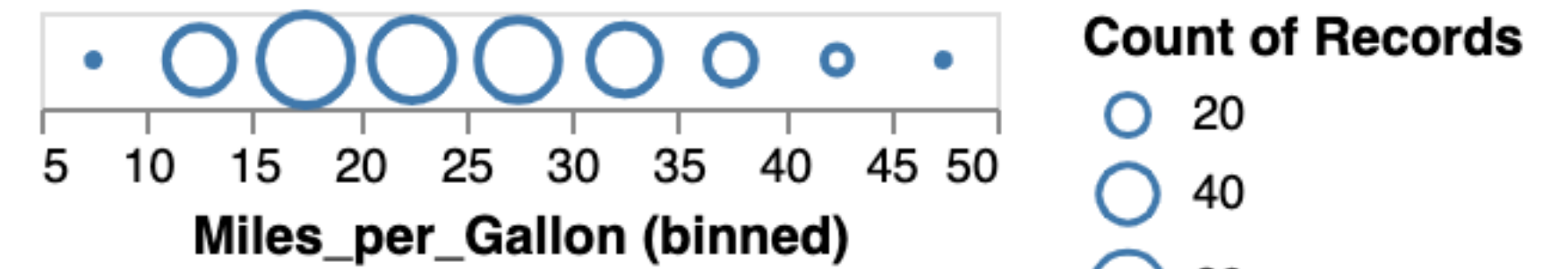
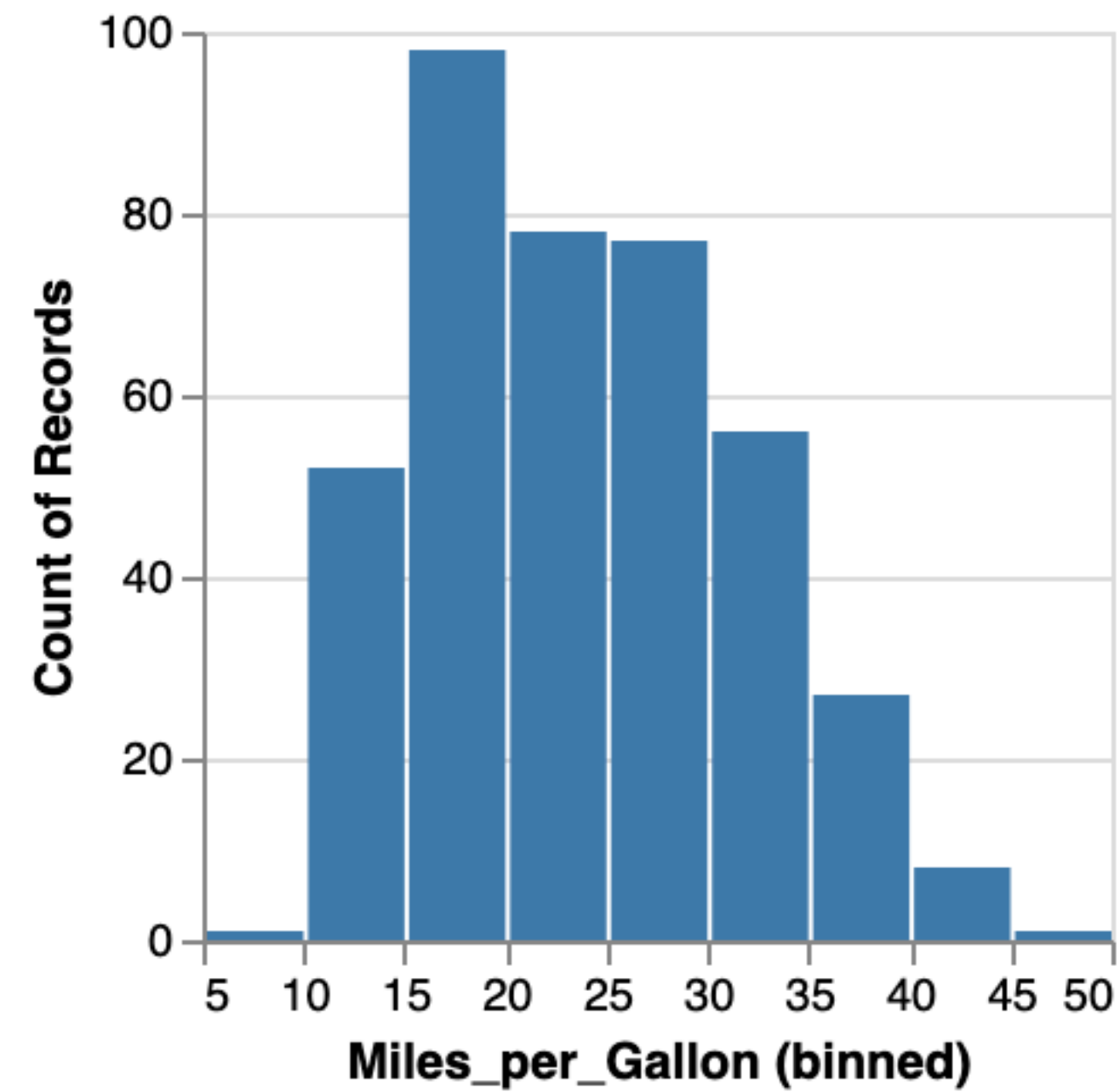
Visual Encoding = Combinatorial Design Space

1D quantitative data (Q)

raw



aggregate (count)

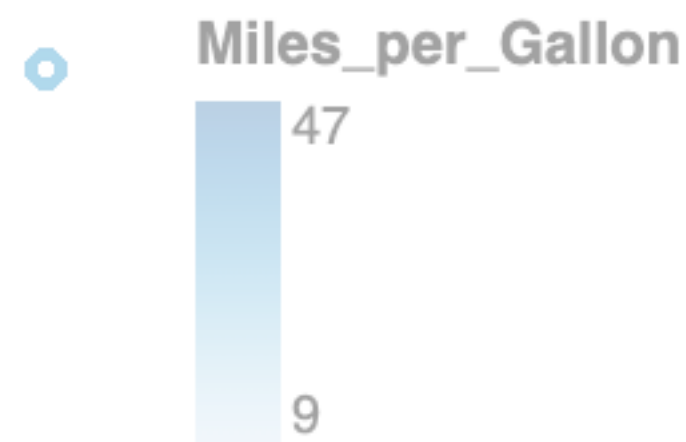
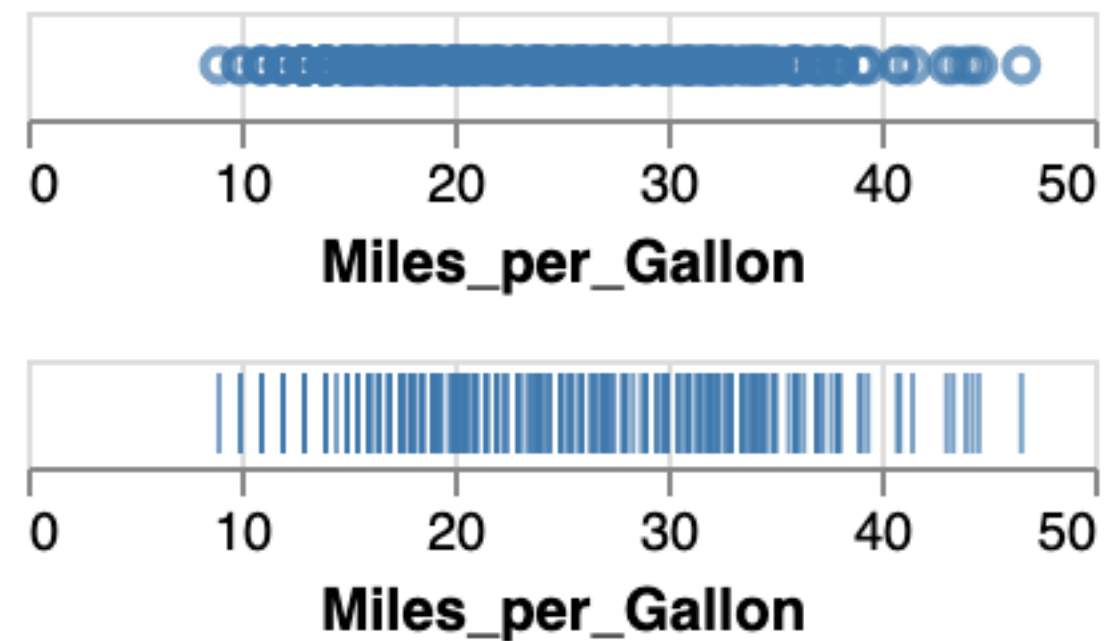


Visual Encoding = Combinatorial Design Space

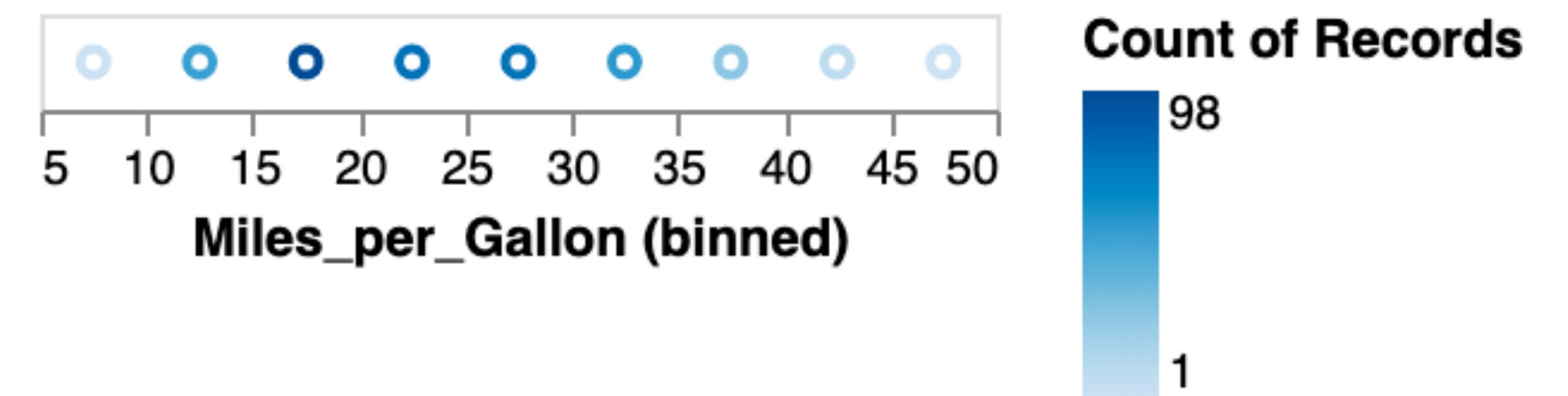
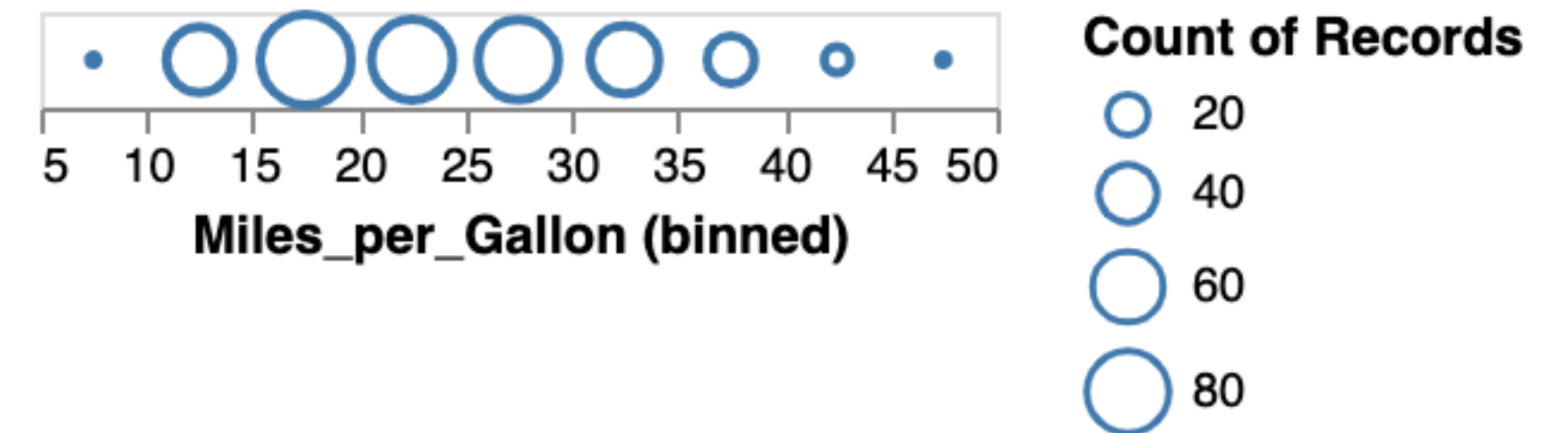
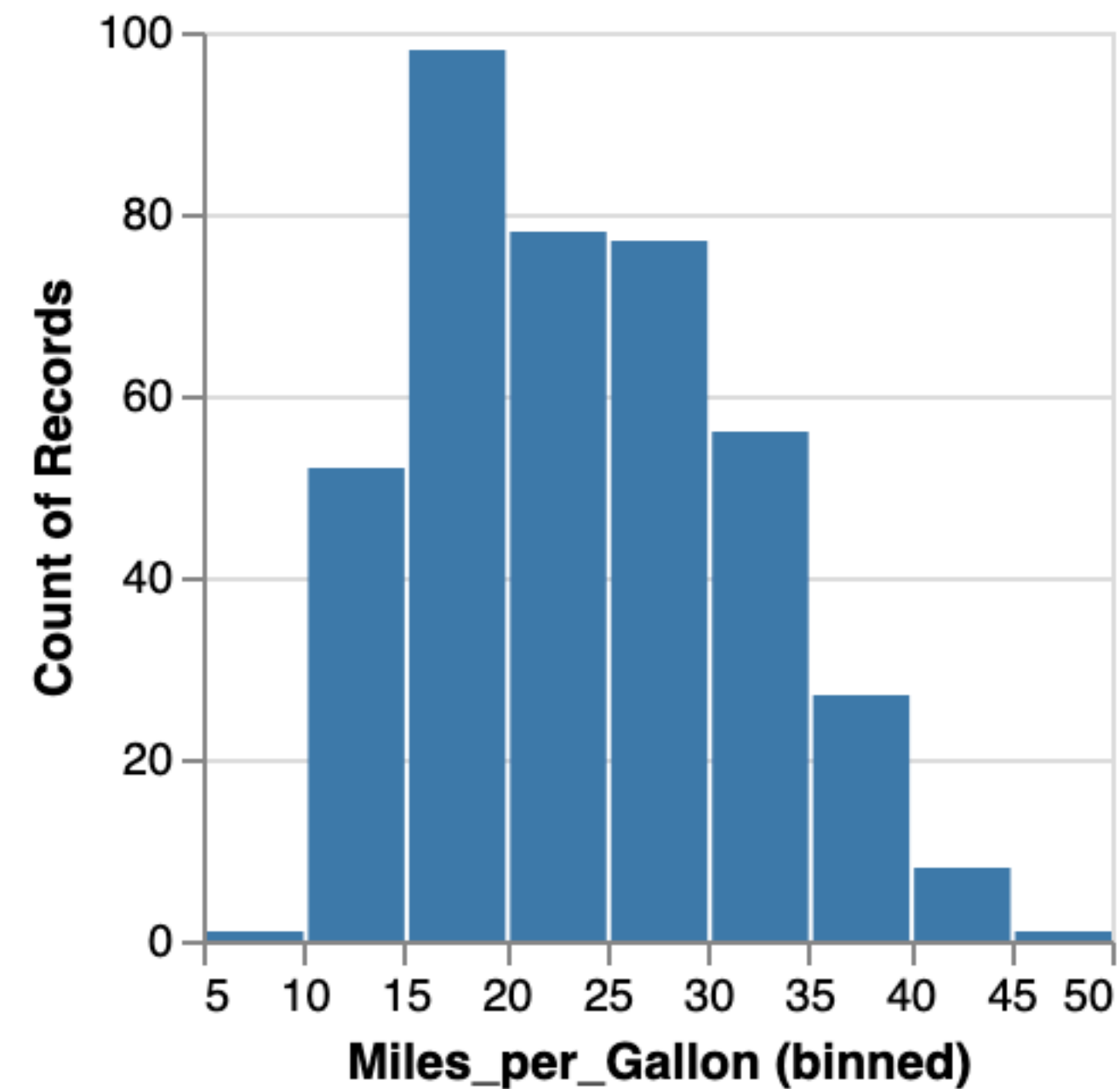
1D quantitative data (Q)

Expressive?

raw



aggregate (count)



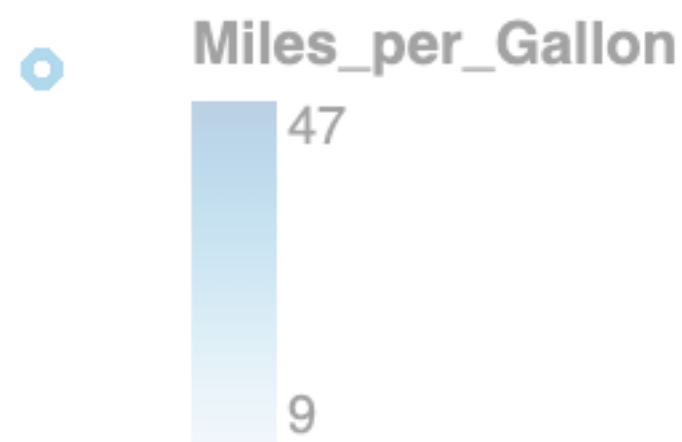
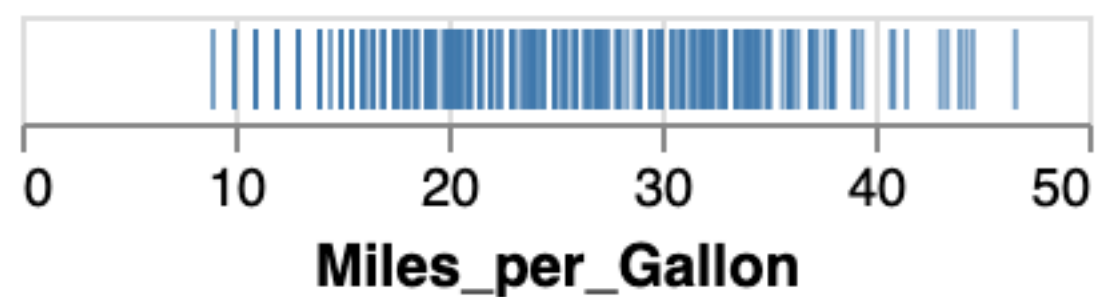
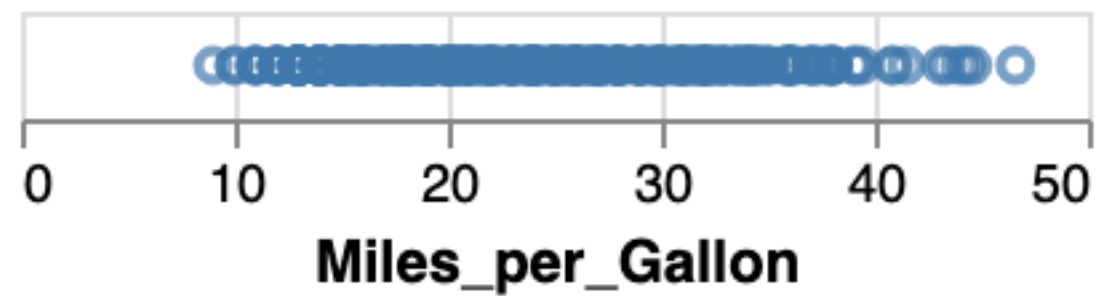
Visual Encoding = Combinatorial Design Space

1D quantitative data (Q)

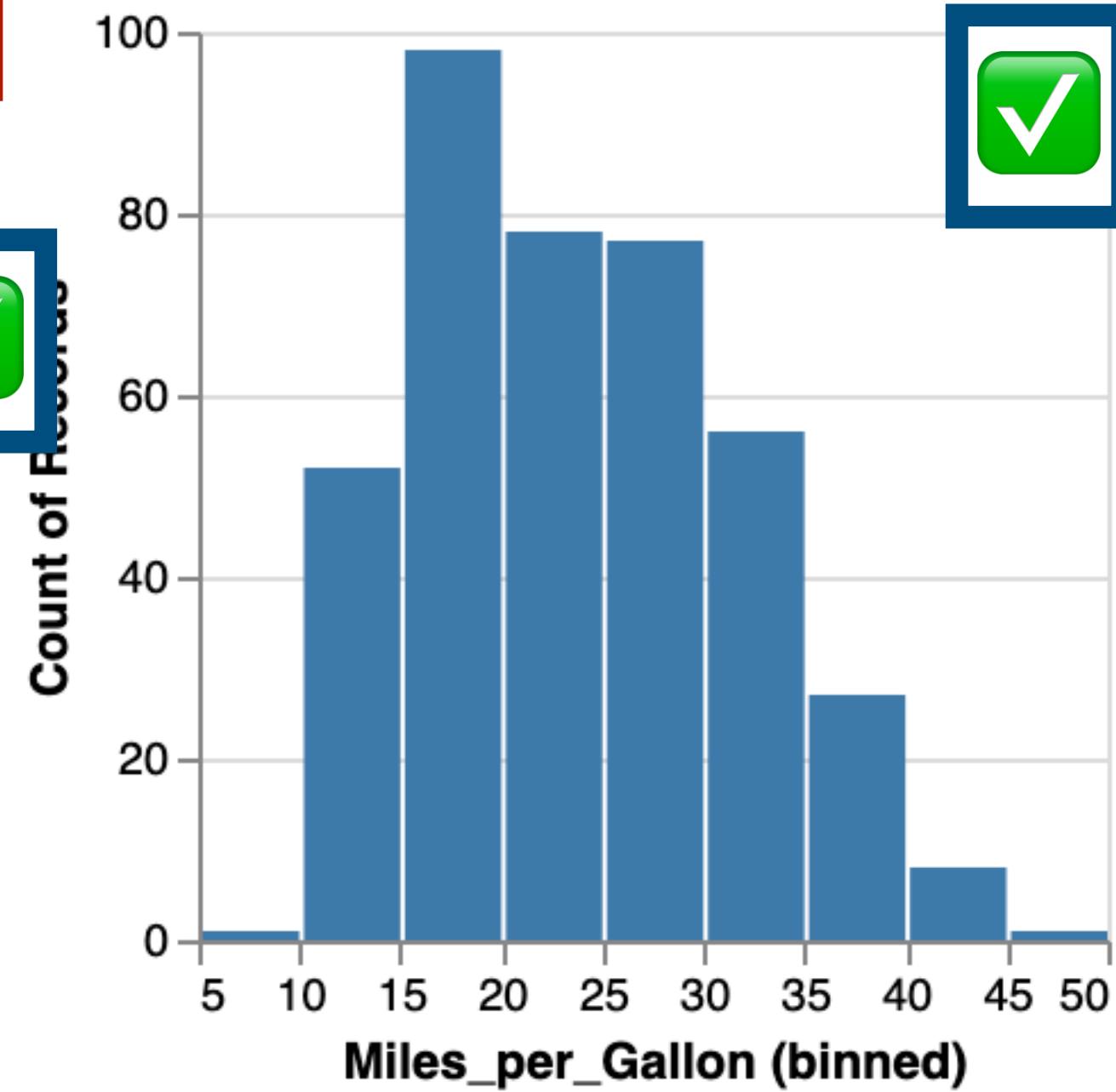
Expressive?

Effective?

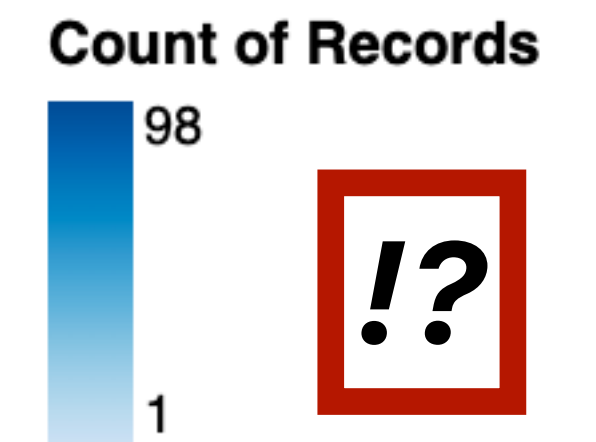
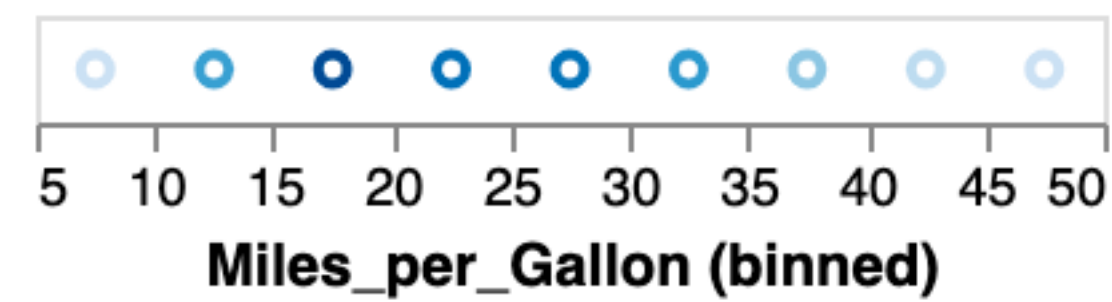
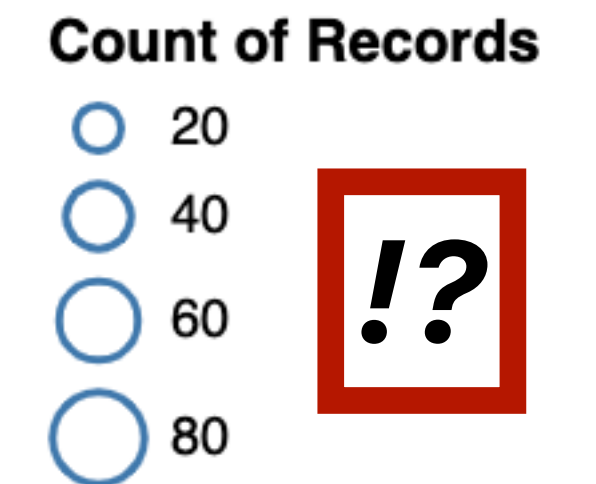
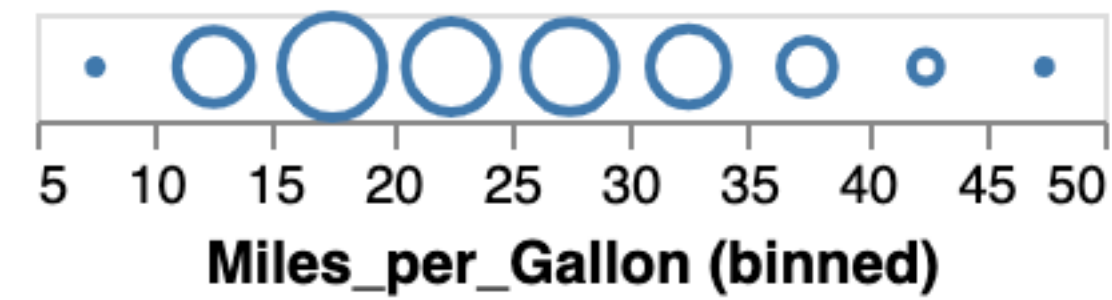
raw



aggregate (count)



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



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



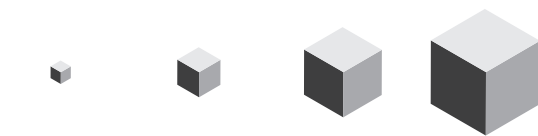
Color saturation



Curvature



Volume (3D size)



Same

Most Effectiveness Least

➔ Identity Channels: Categorical Attributes

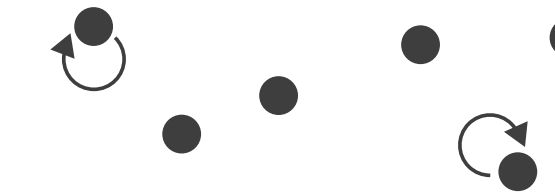
Spatial region



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) 

Perceptual work

Cognitive work

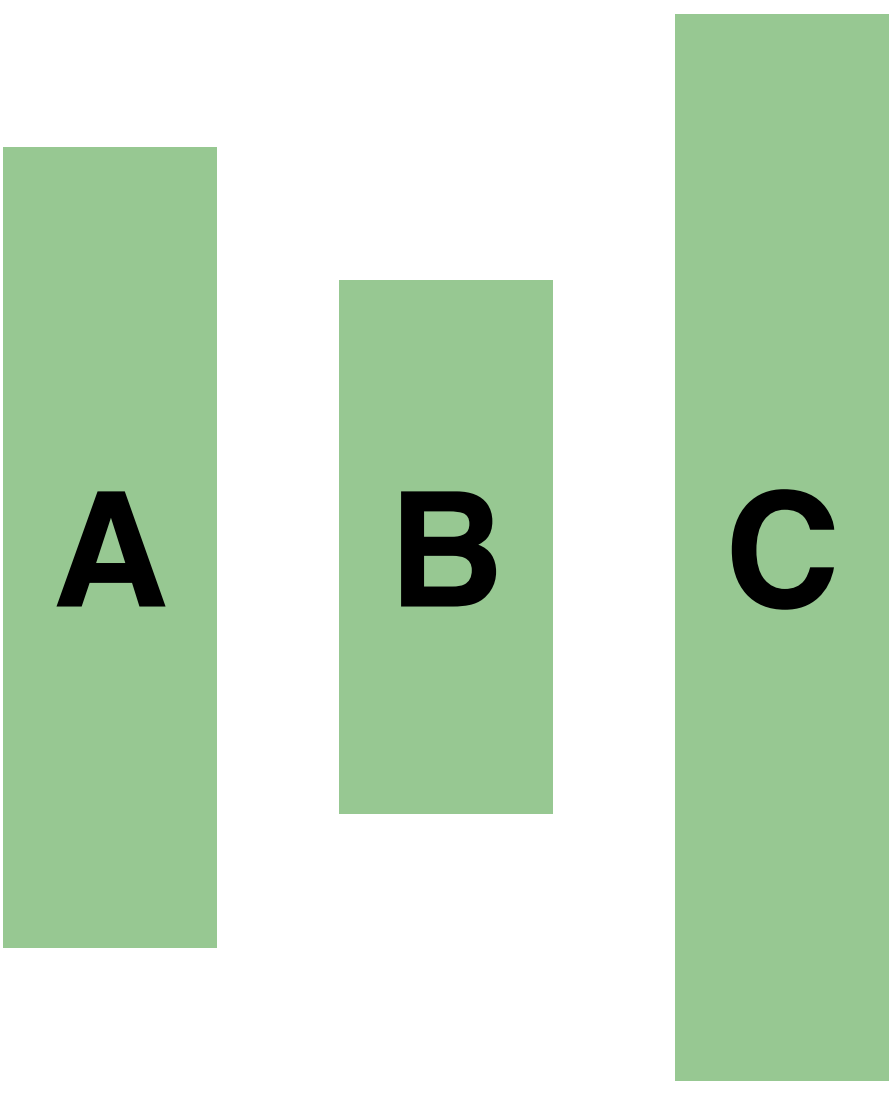
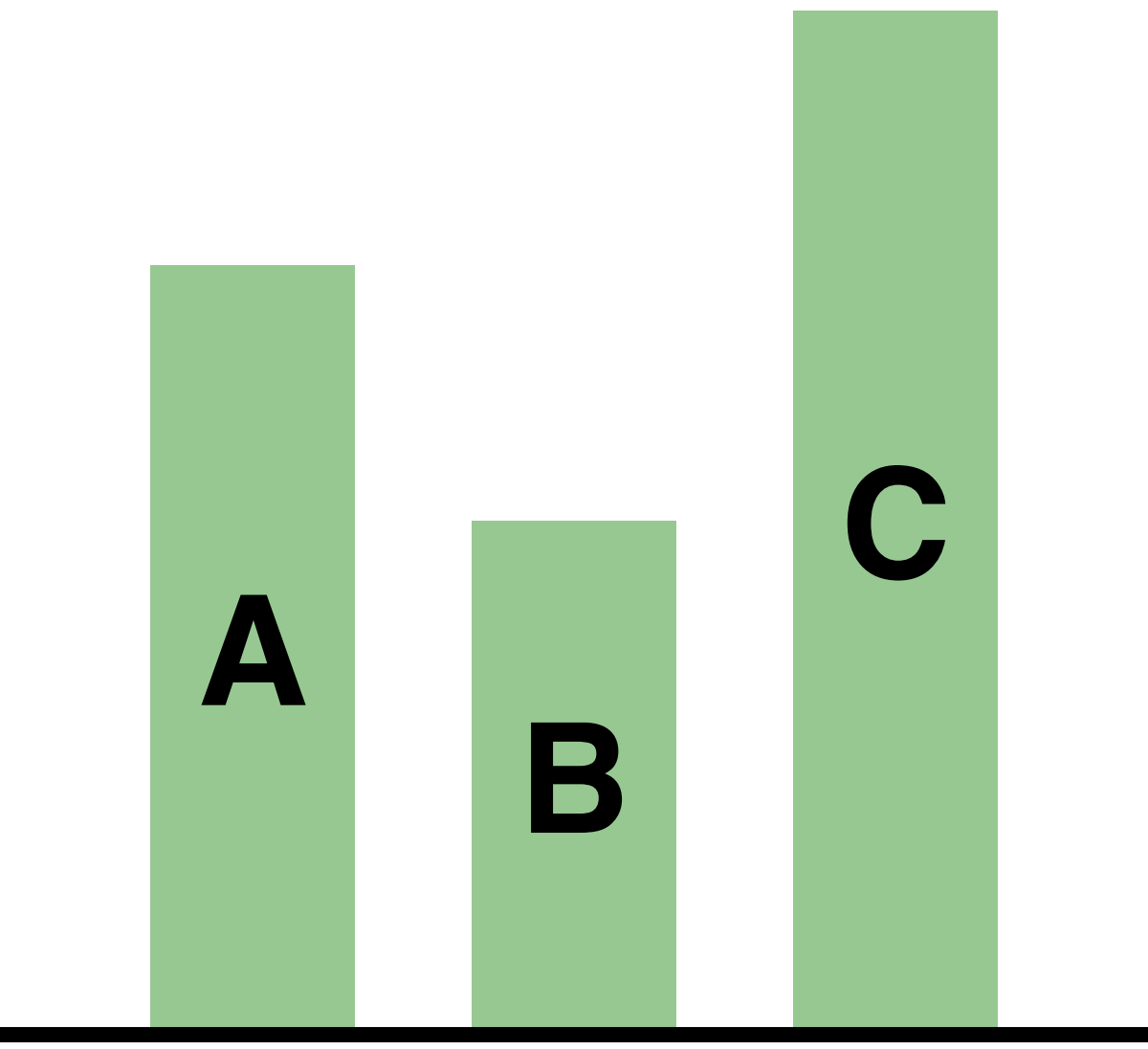
Most

Effectiveness

Least

Same

Same



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)



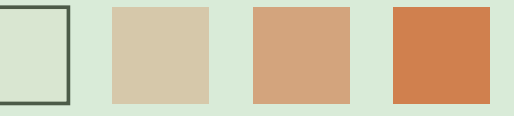
Depth (3D position)



Color luminance



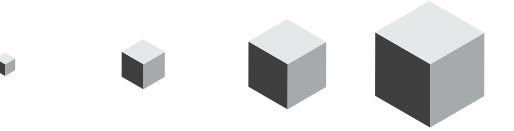
Color saturation



Curvature



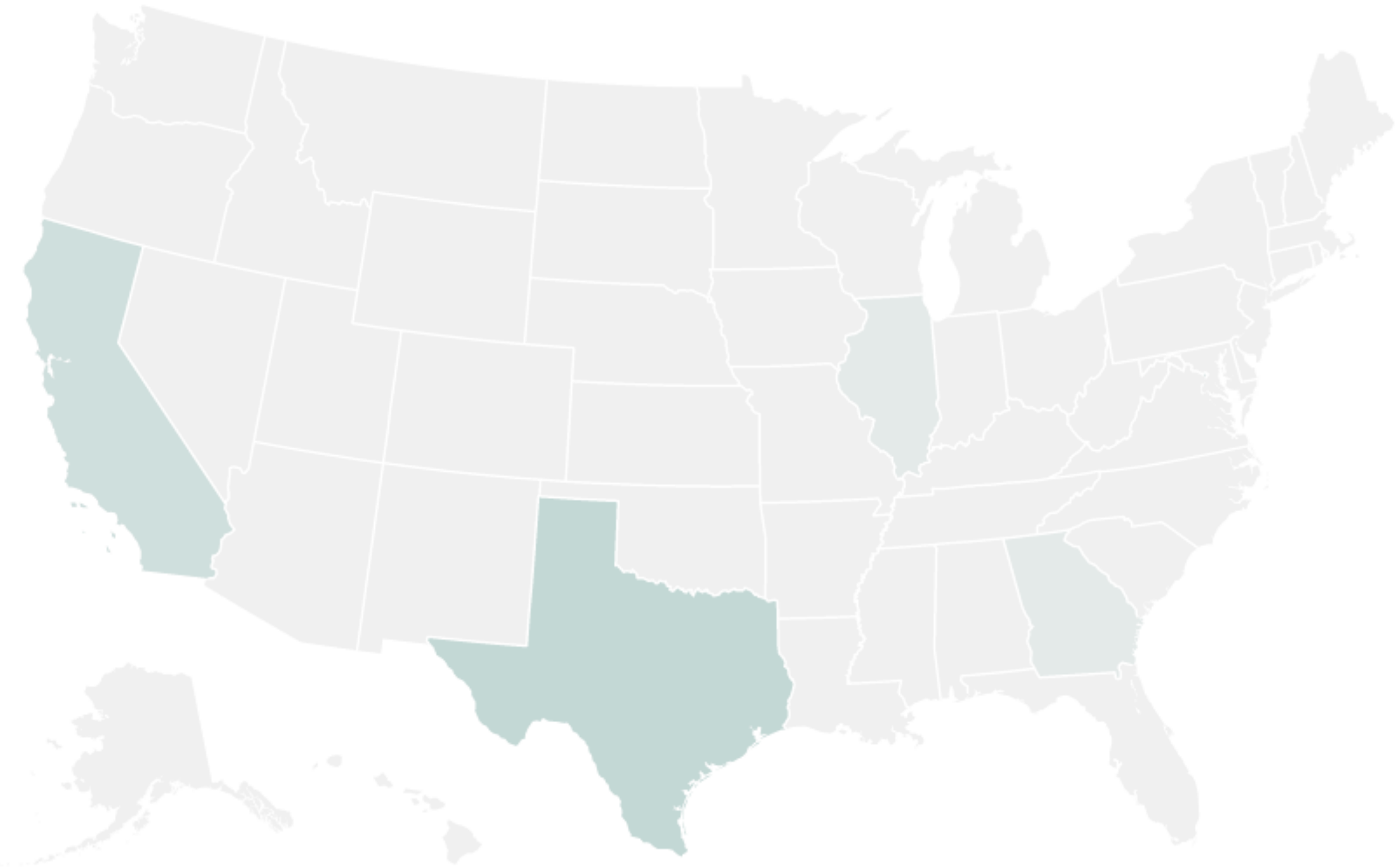
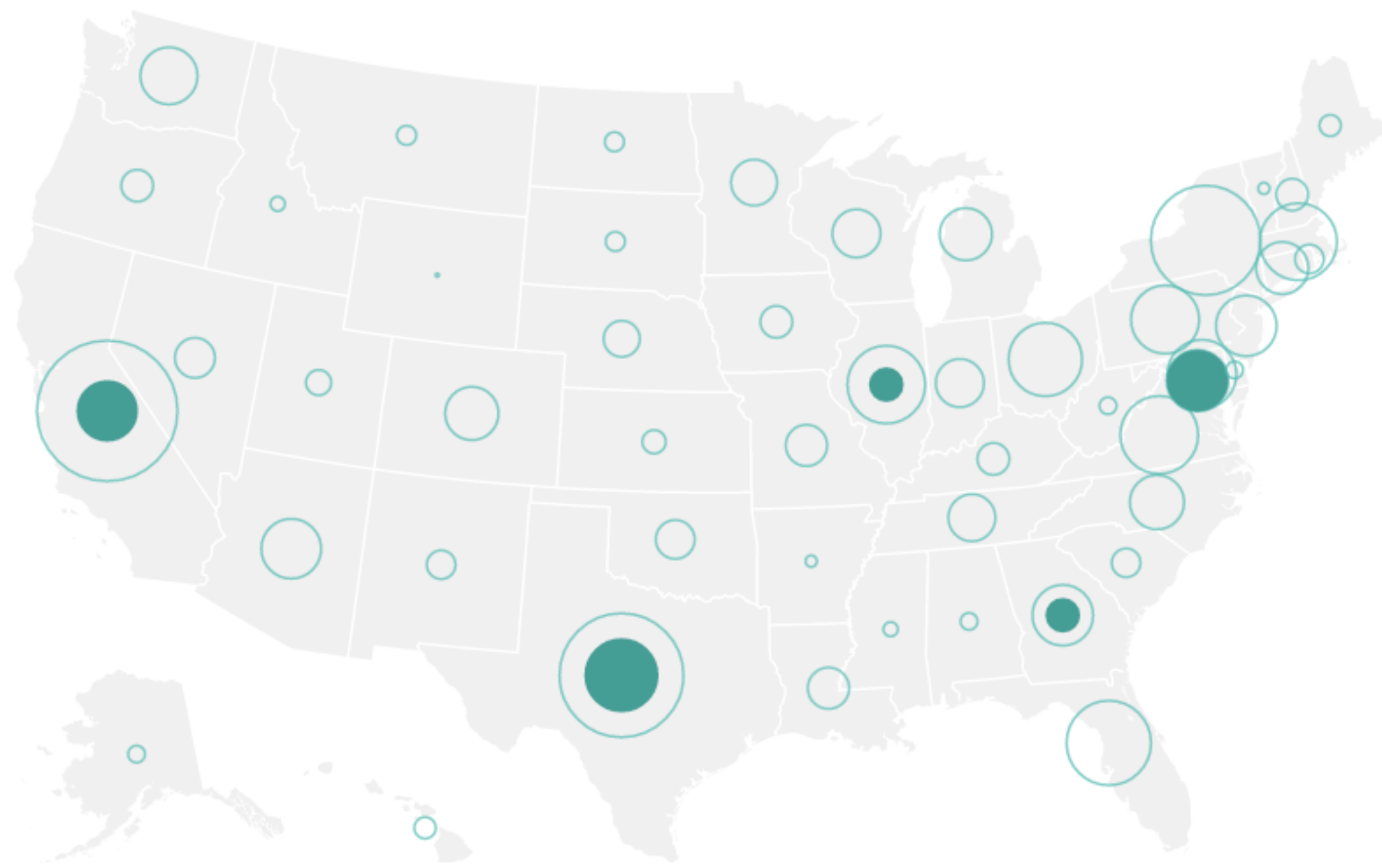
Volume (3D size)



Same

Same

Most Effectiveness Least



Artery Visualization



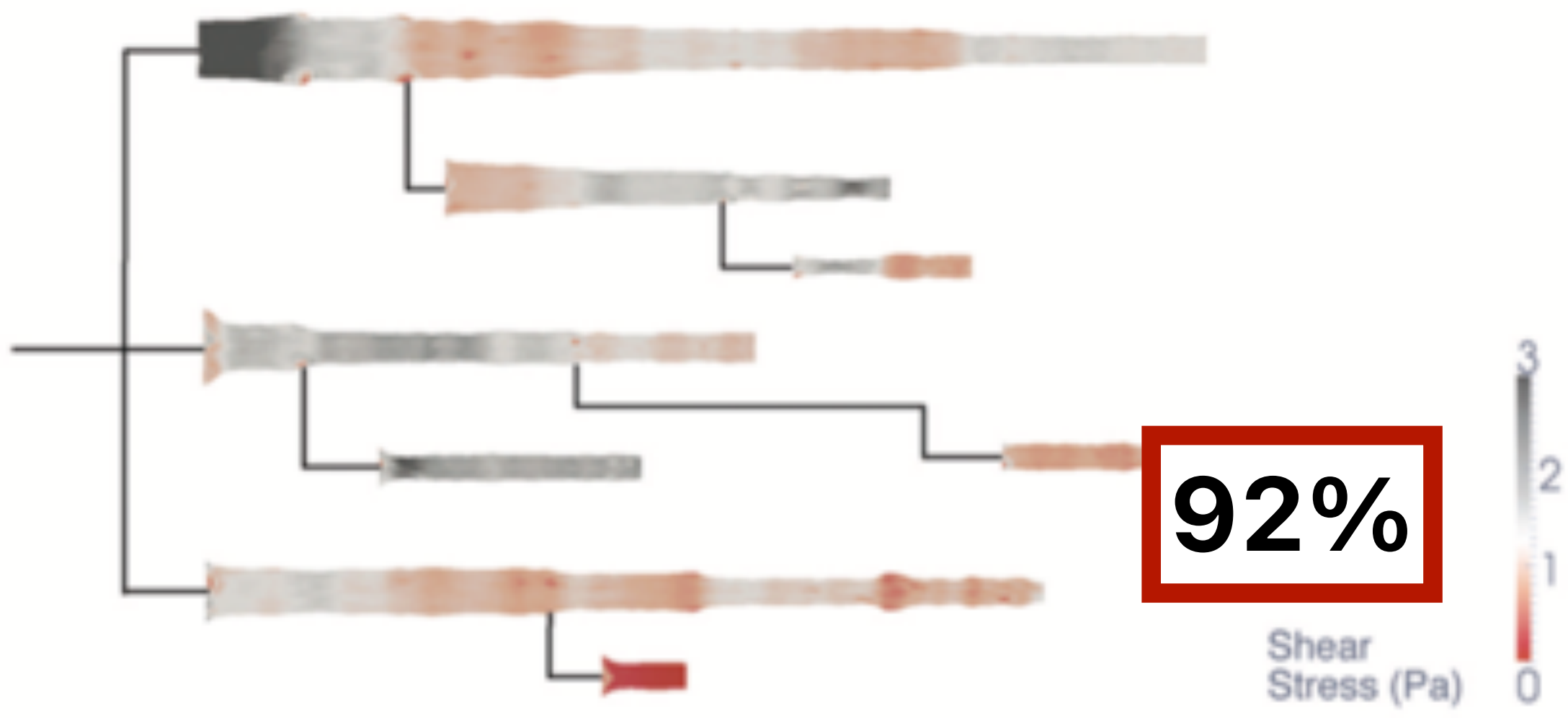
Rainbow Palette



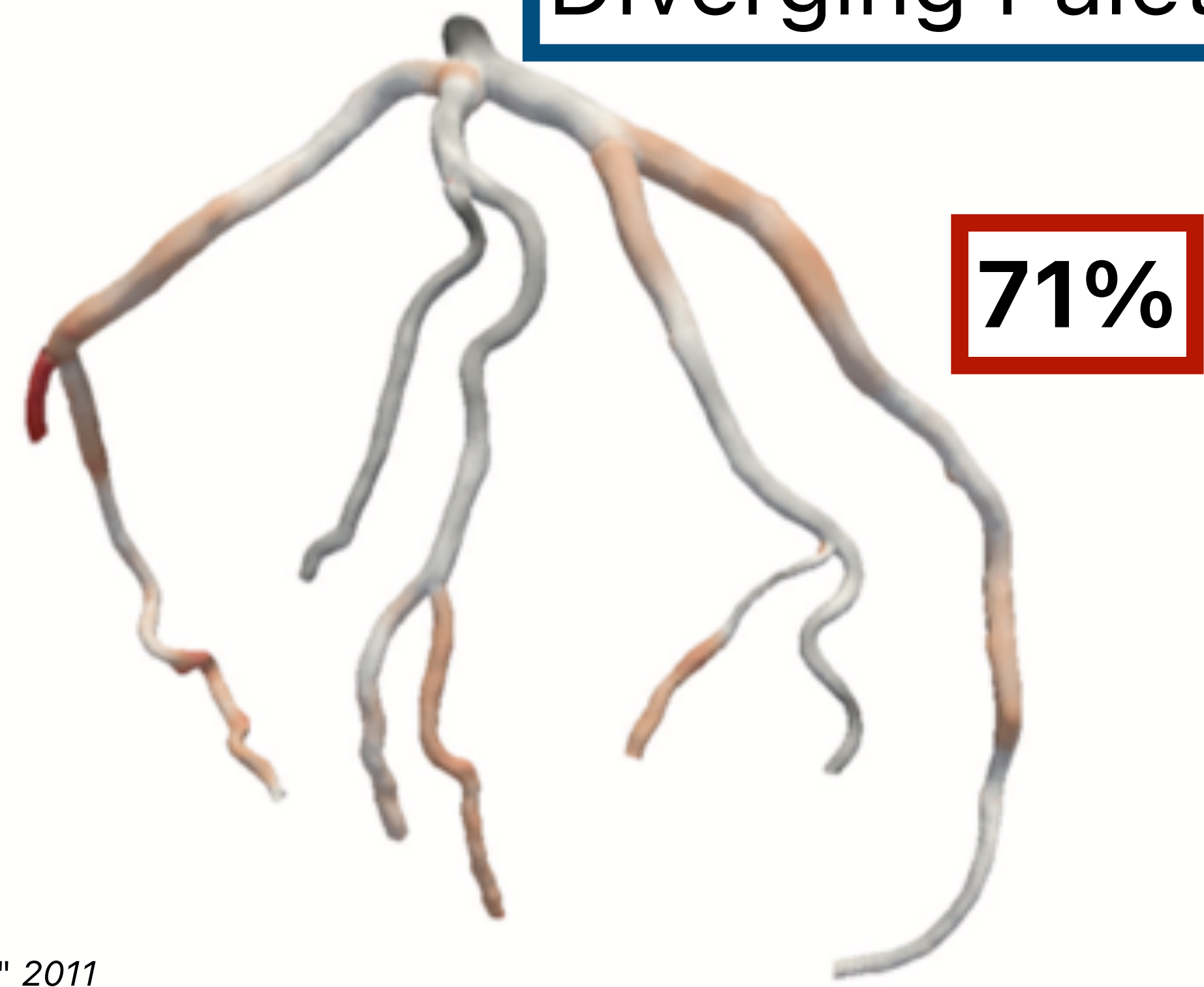
Artery Visualization



Rainbow Palette



Diverging Palette



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



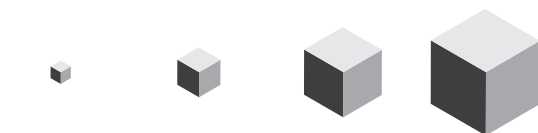
Color saturation



Curvature



Volume (3D size)



Same

Same

Same

Most Effectiveness Least

➔ Identity Channels: Categorical Attributes

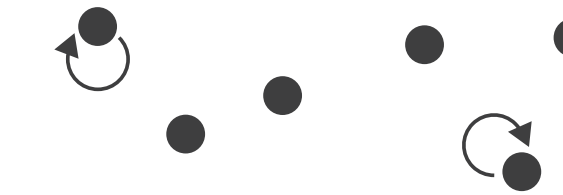
Spatial region



Color hue



Motion



Shape



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

Using space (in)effectively

(De-)Obfuscating data

(Mis)leading the witness

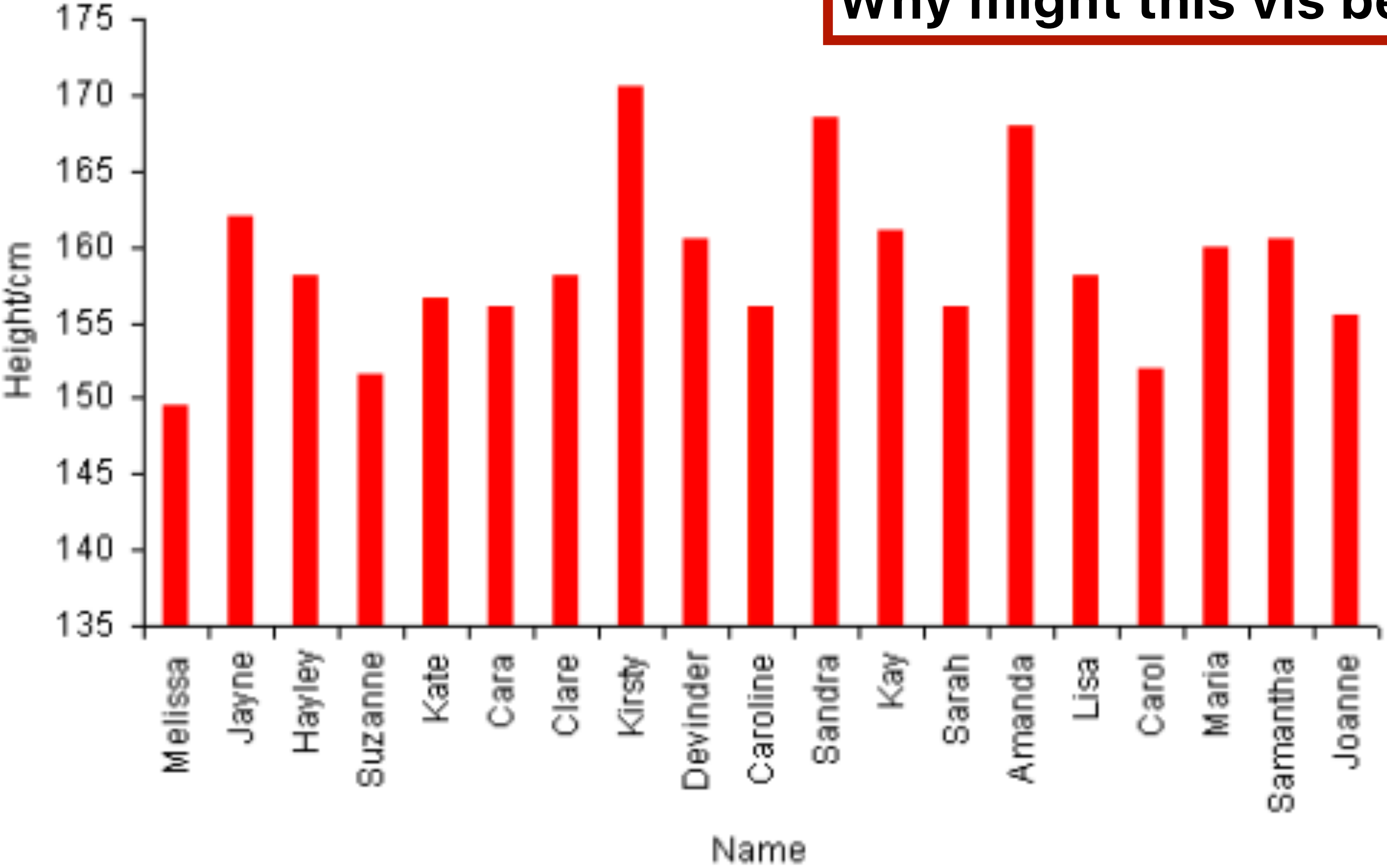
Using space (in)effectively

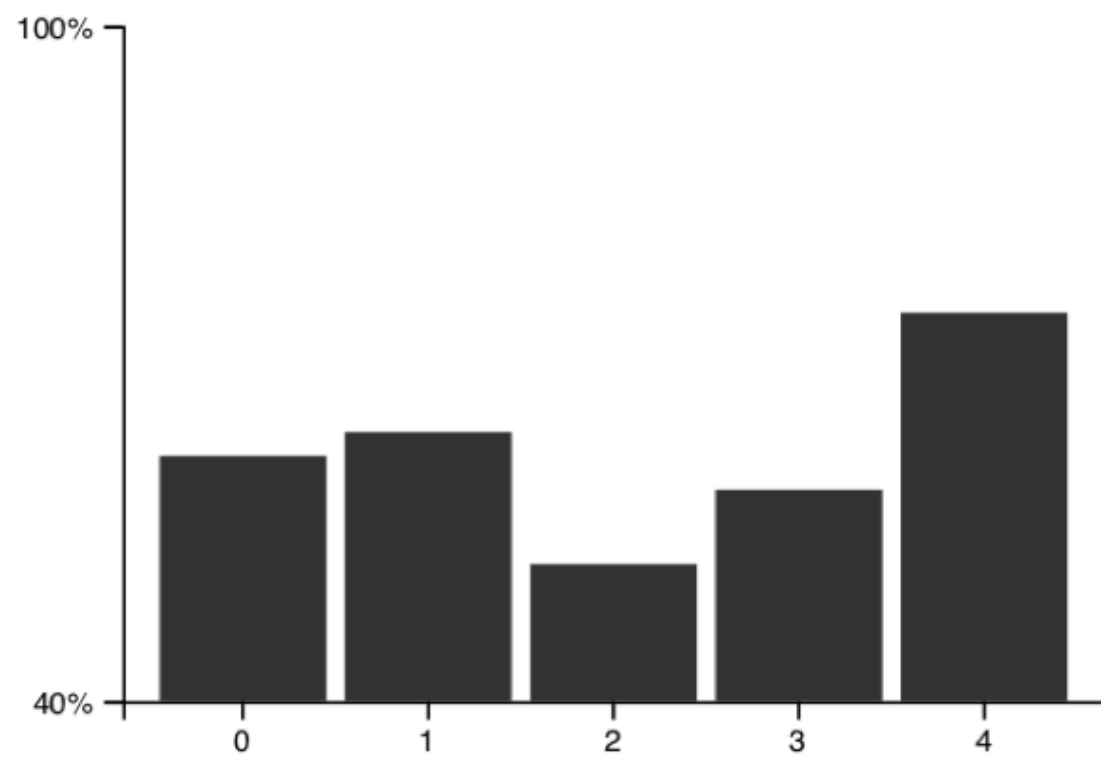
(De-)Obfuscating data

(Mis)leading the witness

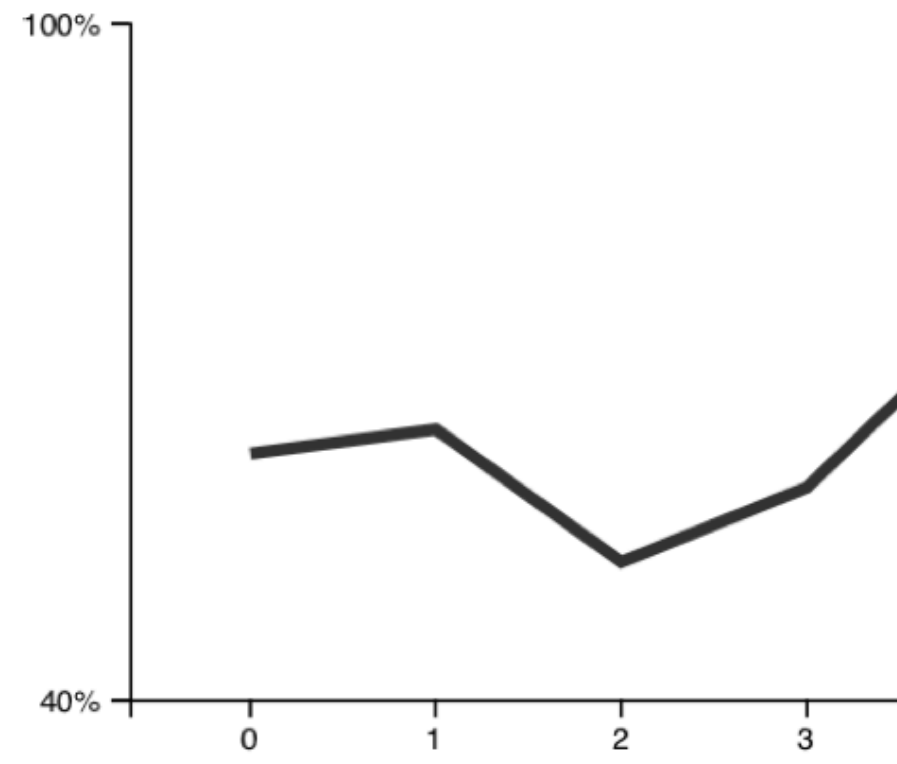
Individual heights

Why might this vis be inexpressive?



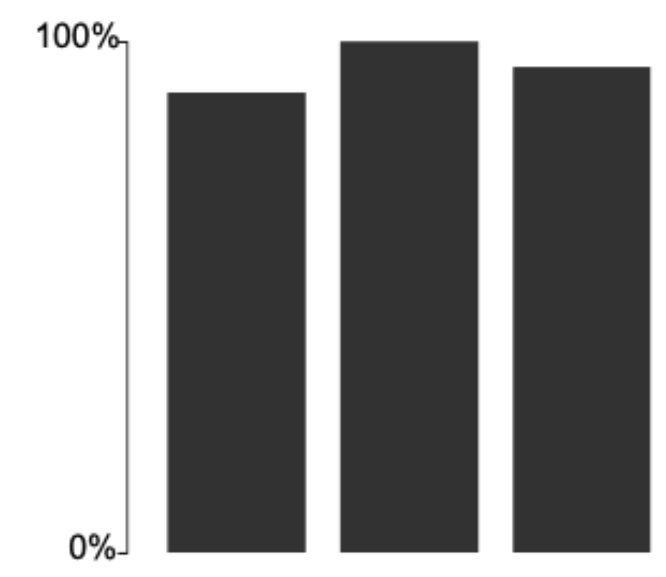


(a) Bar Chart

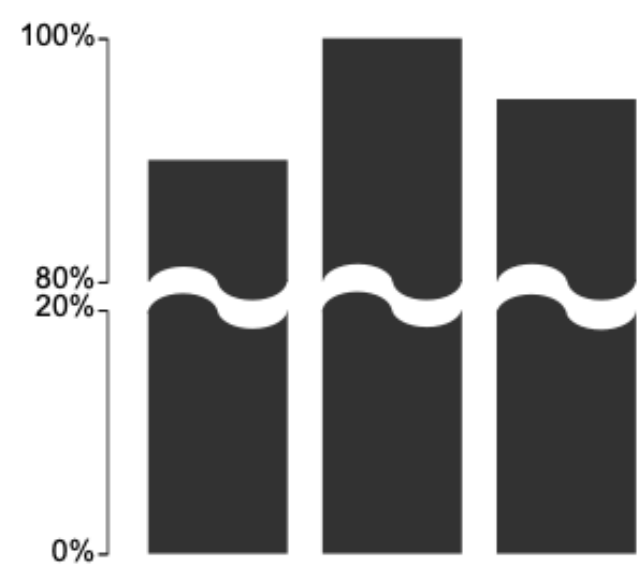


(b) Line Chart

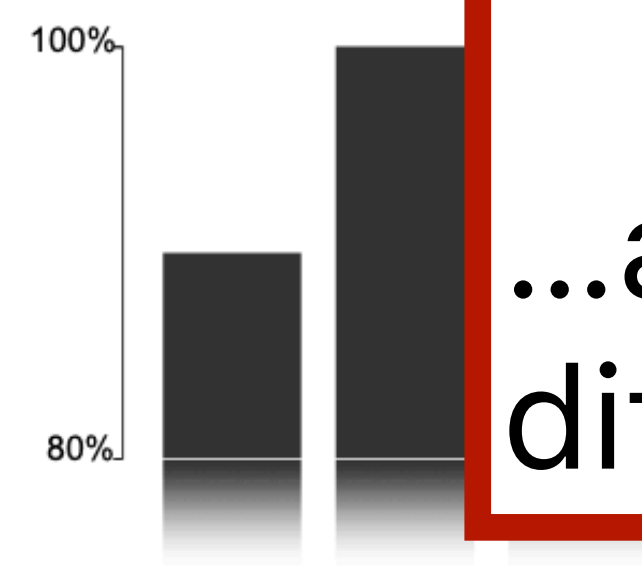
Y-axis truncation impacts perception...
 ...and interventions did not make a difference.



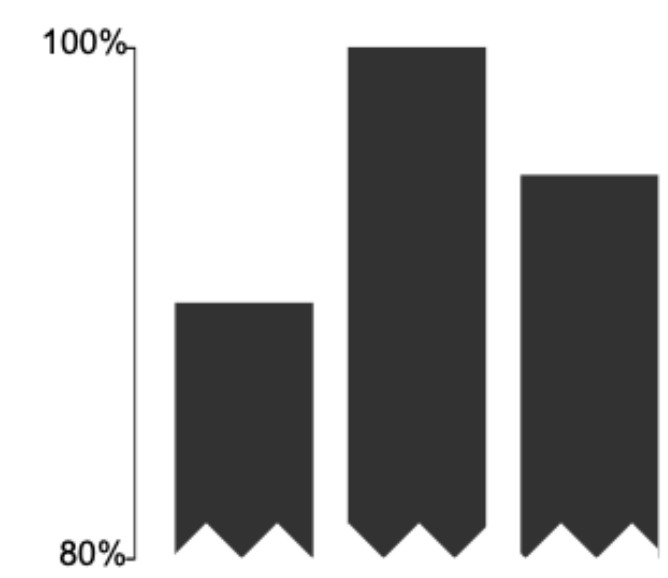
(a) Bar Chart



(b) Broken Axes



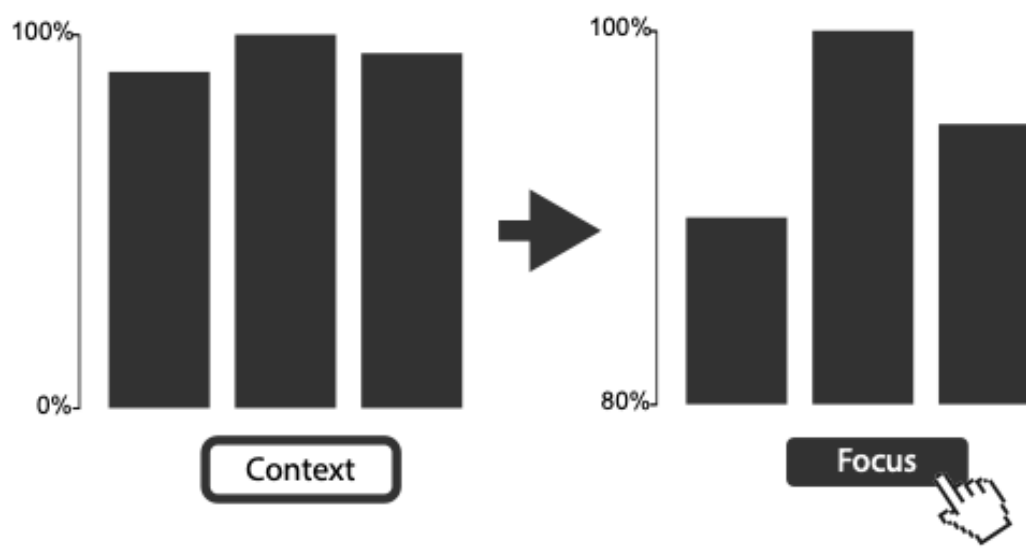
(c) Gradient Bar Chart



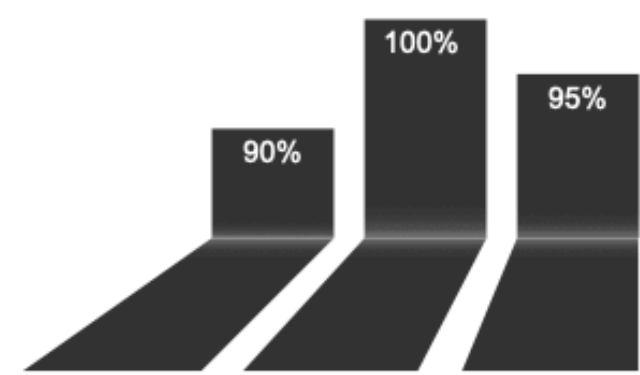
(d) Torn Paper Chart



(e) Panel Chart



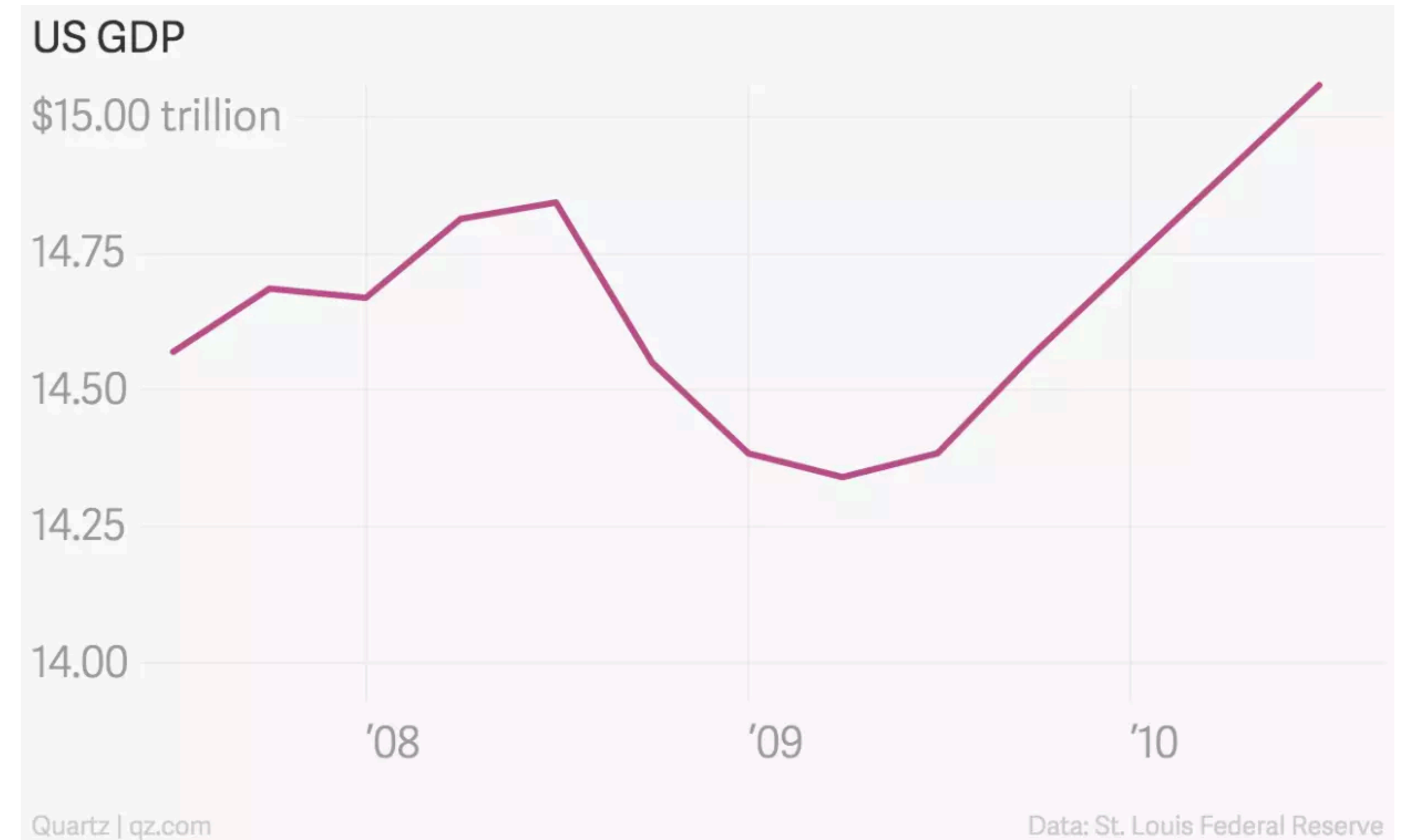
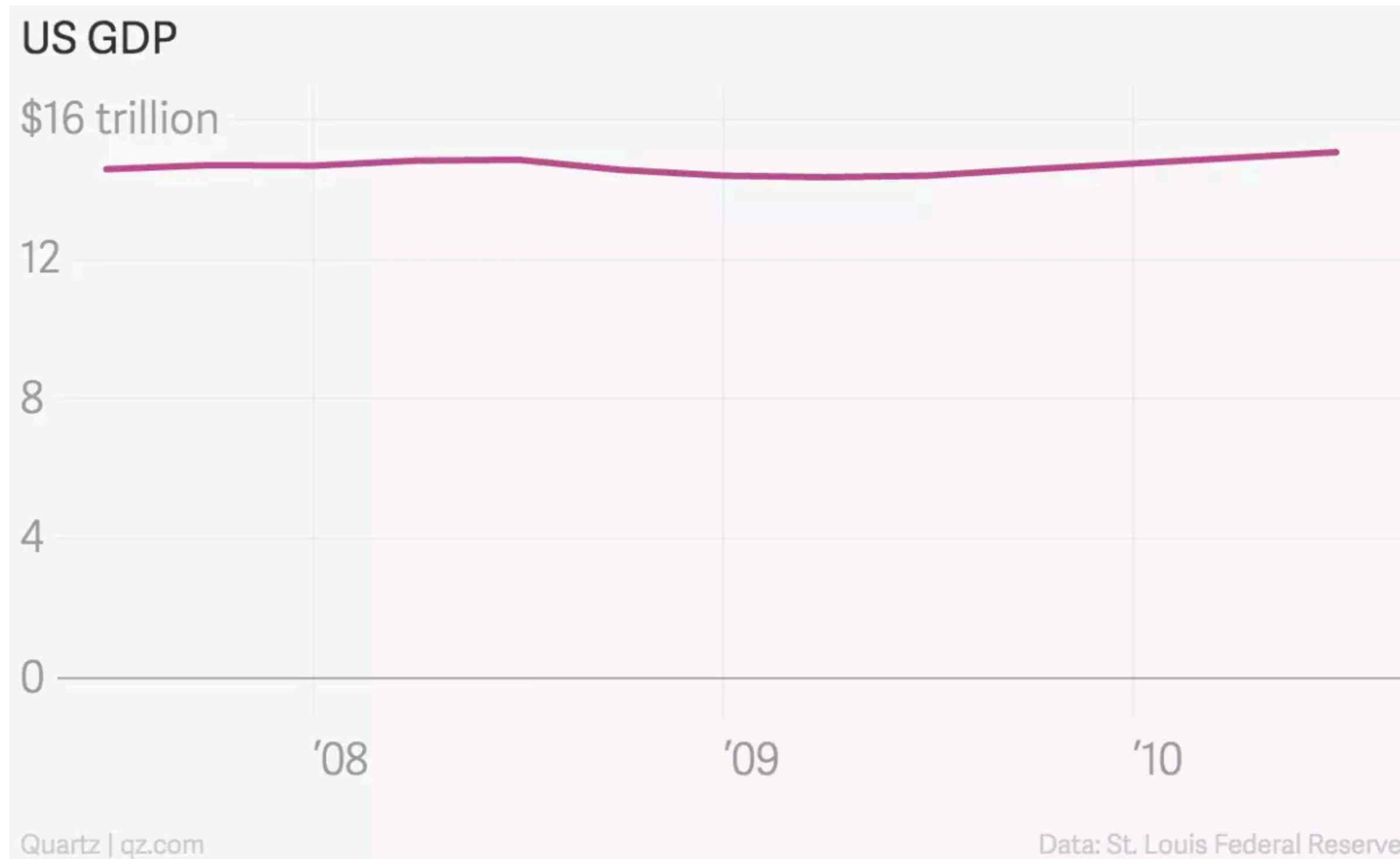
(f) Interactive Focus+Context



(g) Bent Bar Chart

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

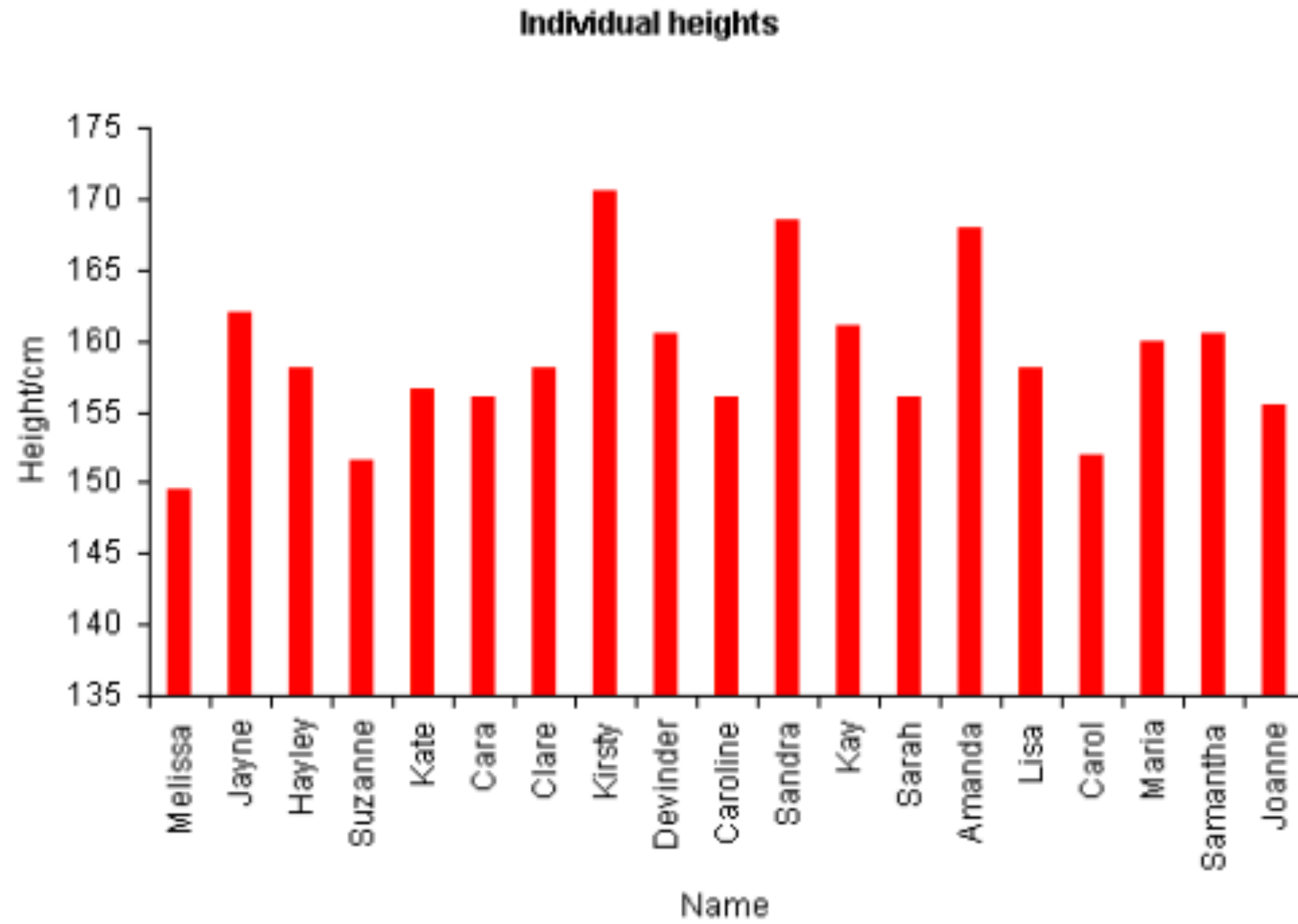
Always start at zero?



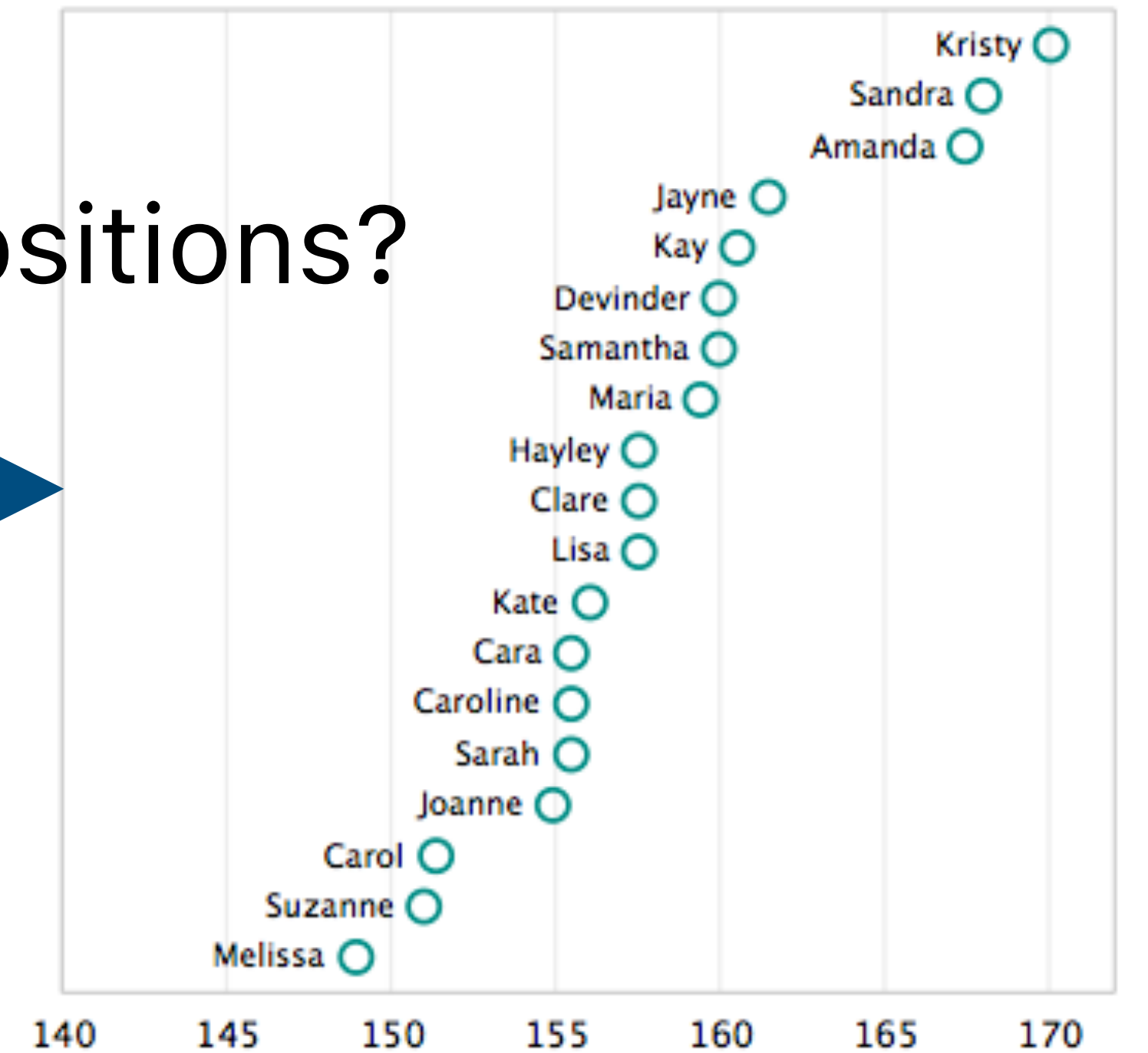
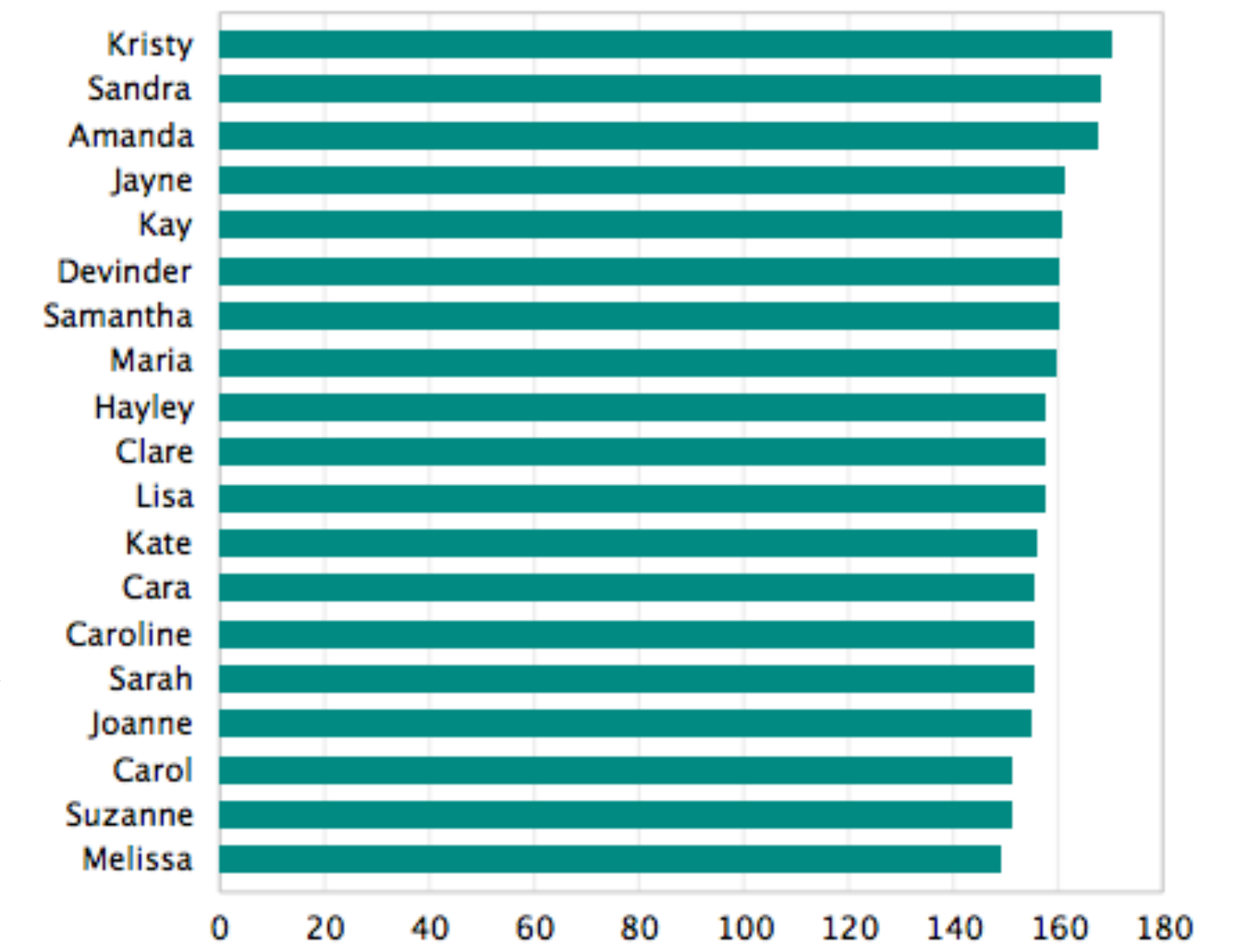
<https://qz.com/418083/its-ok-not-to-start-your-y-axis-at-zero>

Truncating the y-axis?

Compare proportions?
(Q-ratio)

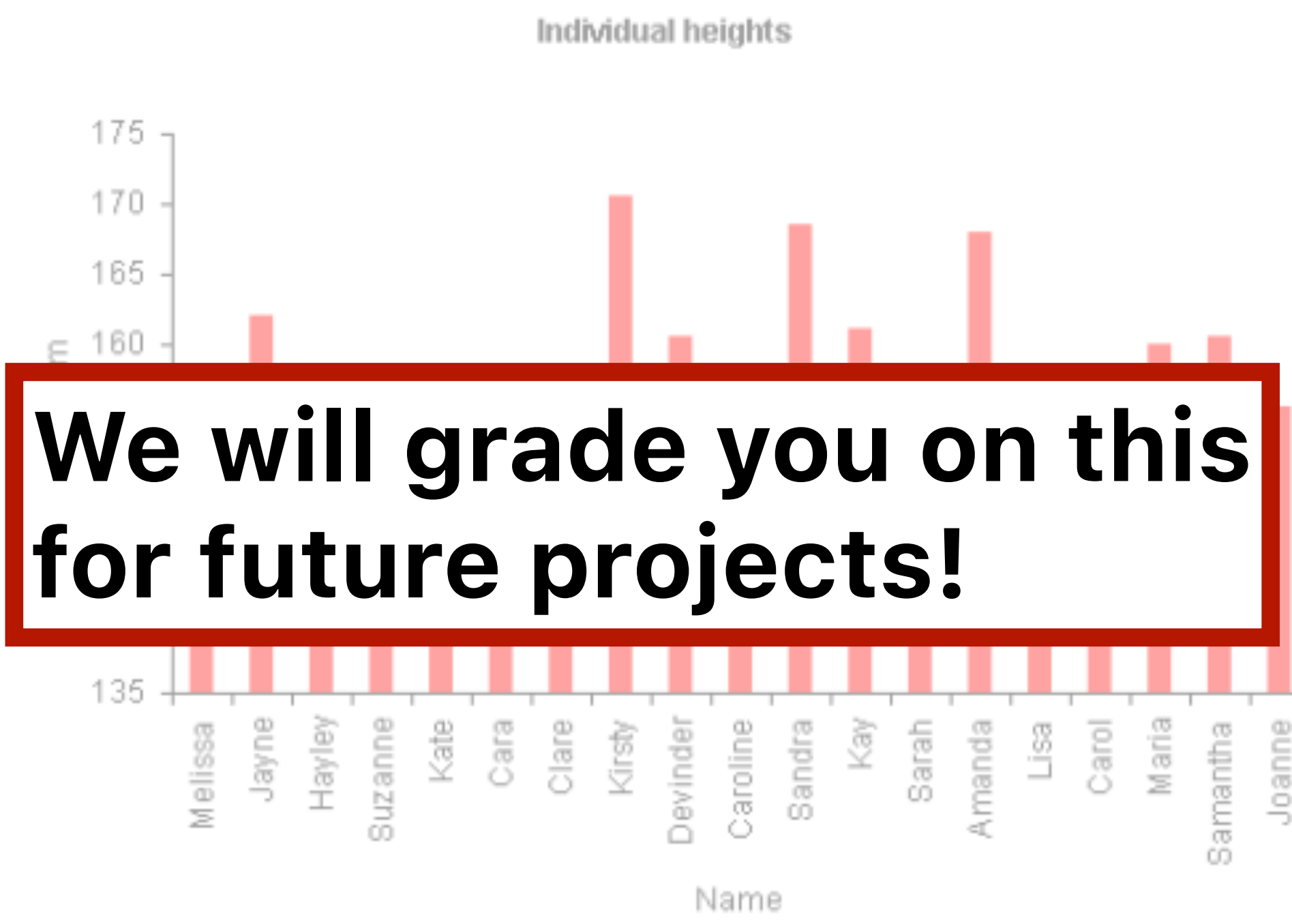
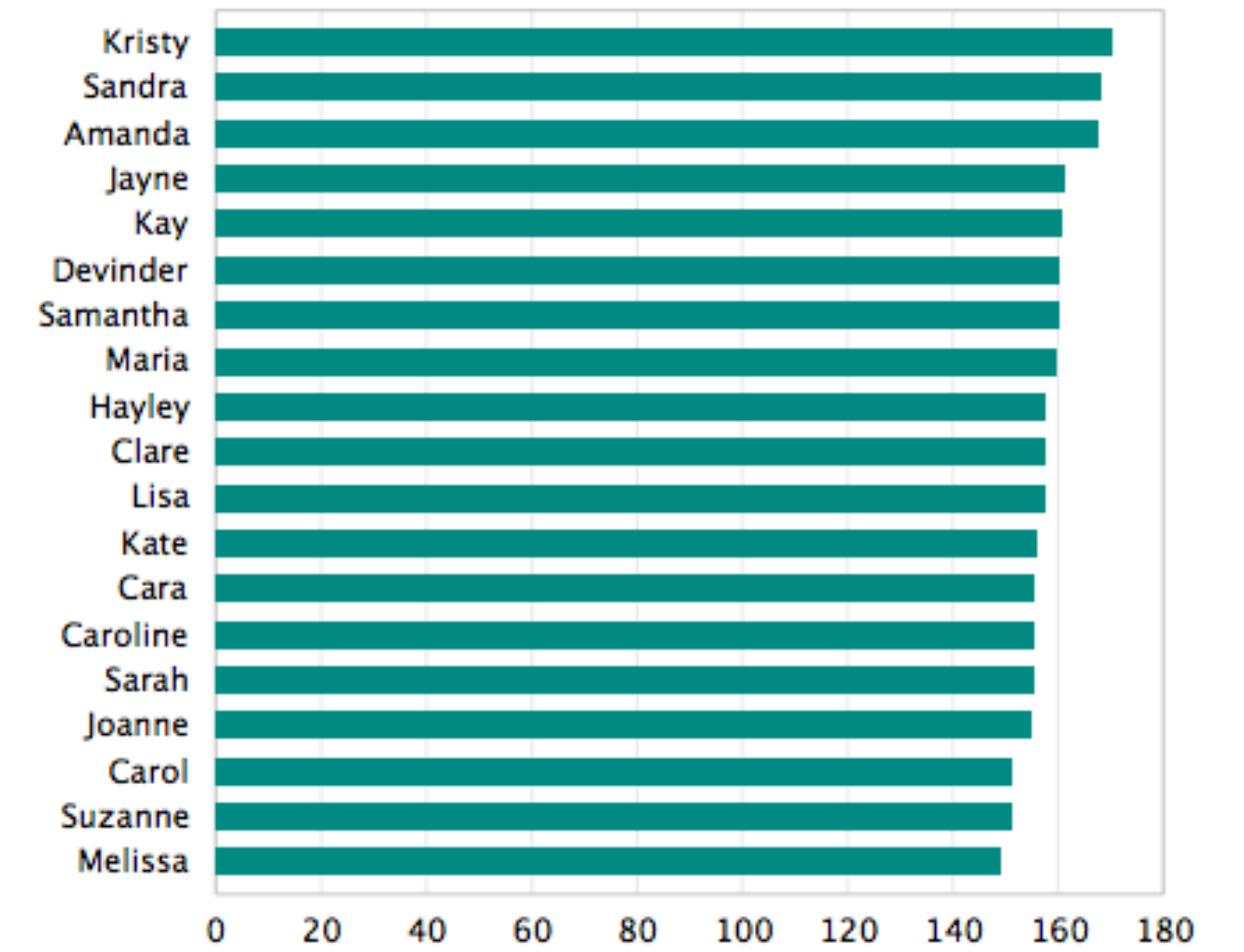


Compare relative positions?
(Q-interval)

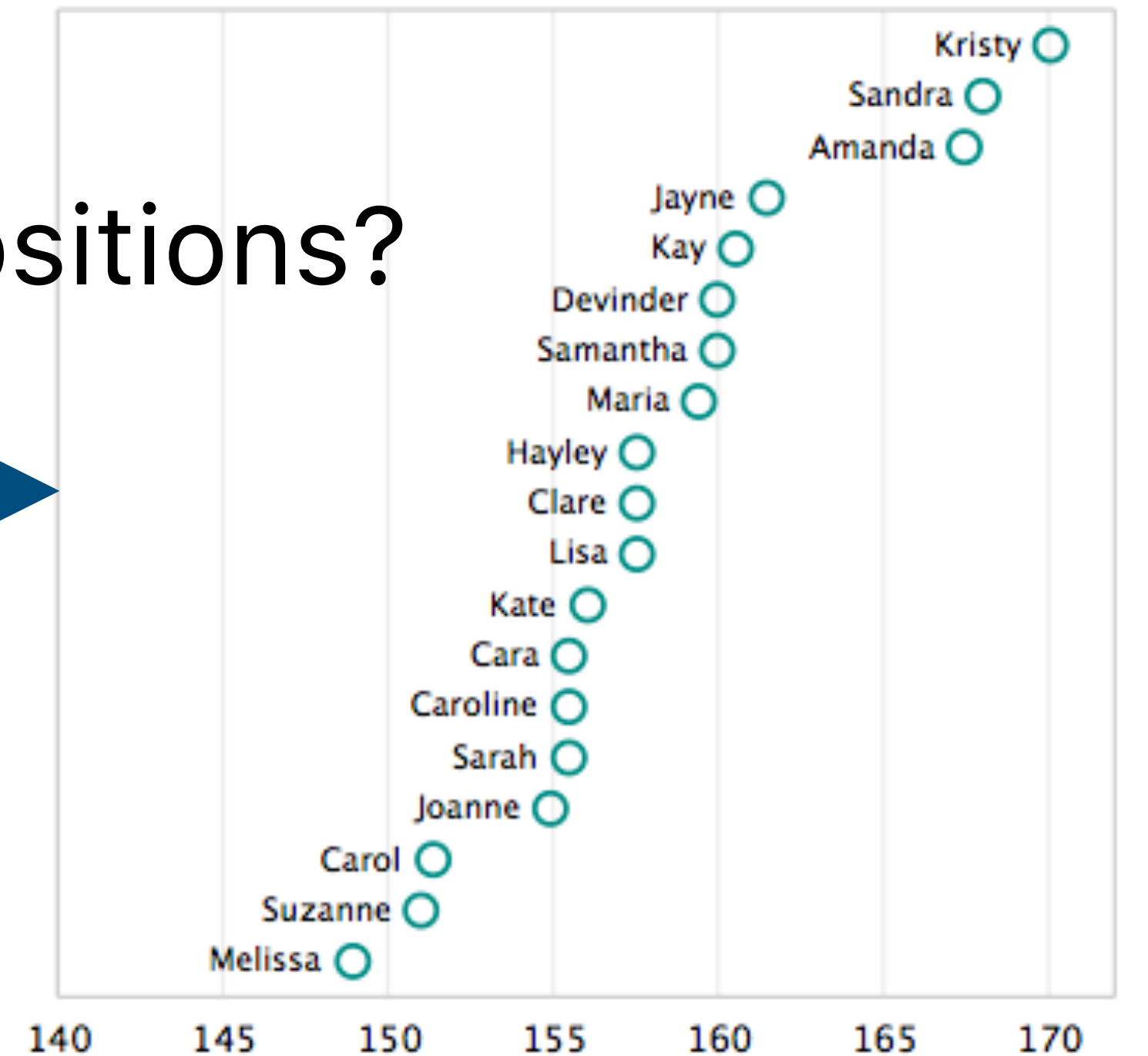


Truncating the y-axis?

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

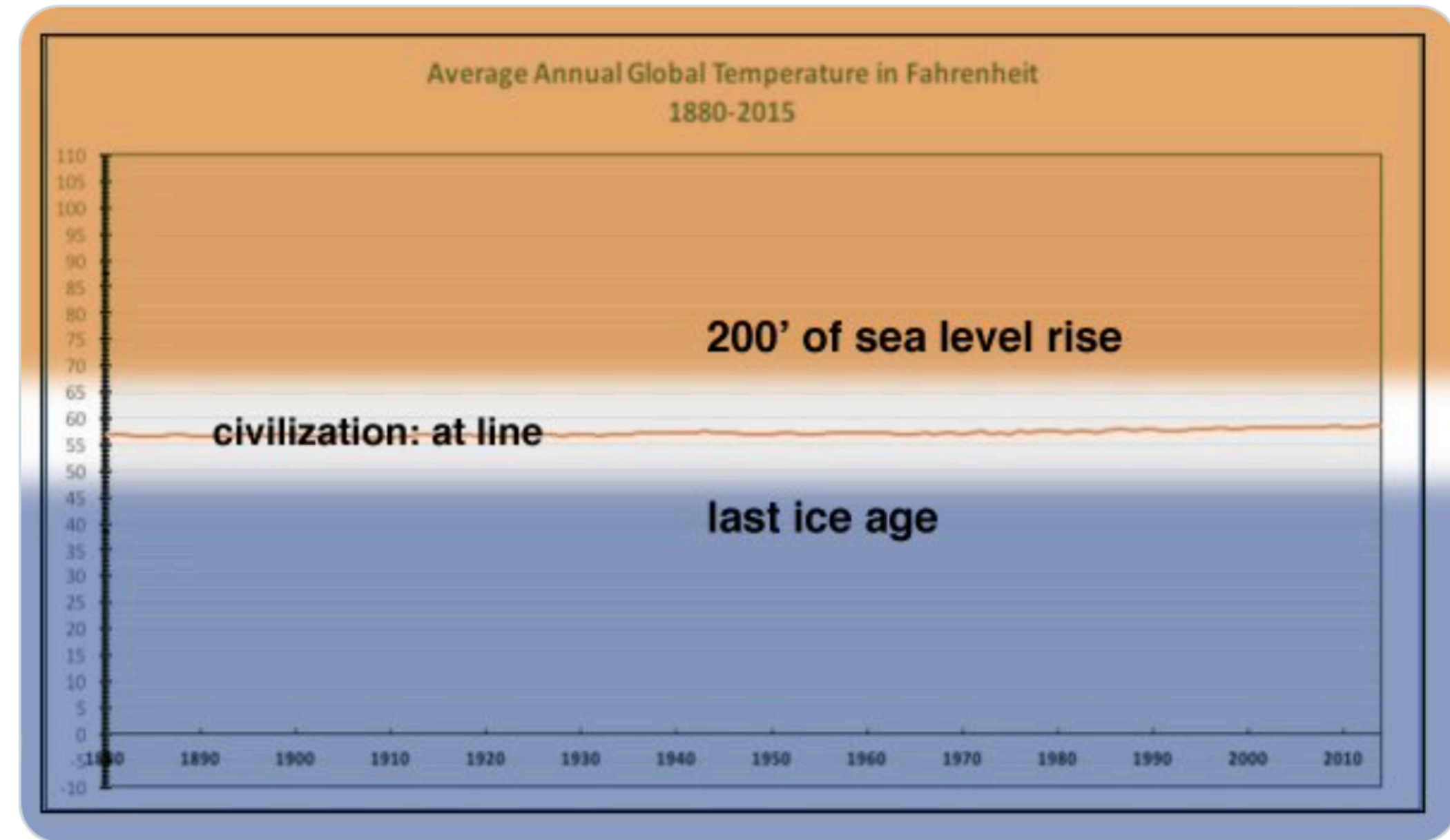
(h/t @powerlineUS)



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 --

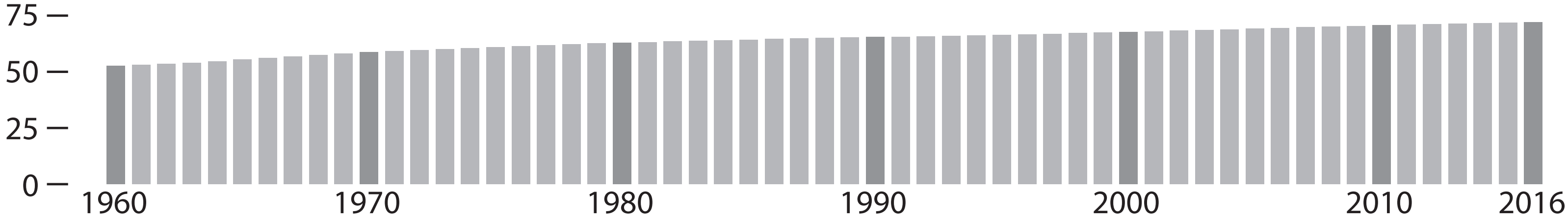


5:28 PM · Dec 14, 2015

78 Retweets 1 Quote Tweet 208 Likes

12:36 PM - 14 Dec 2015

Average world life expectancy at birth (years)



(Source: World Bank Data)

70 —

60 —

50 —

1960

2016

Too narrow and tall

70 —

60 —

50 —

1960

1970

1980

1990

2000

2010

2016

Too wide and short

75 —

50 —

1960

1970

1980

1990

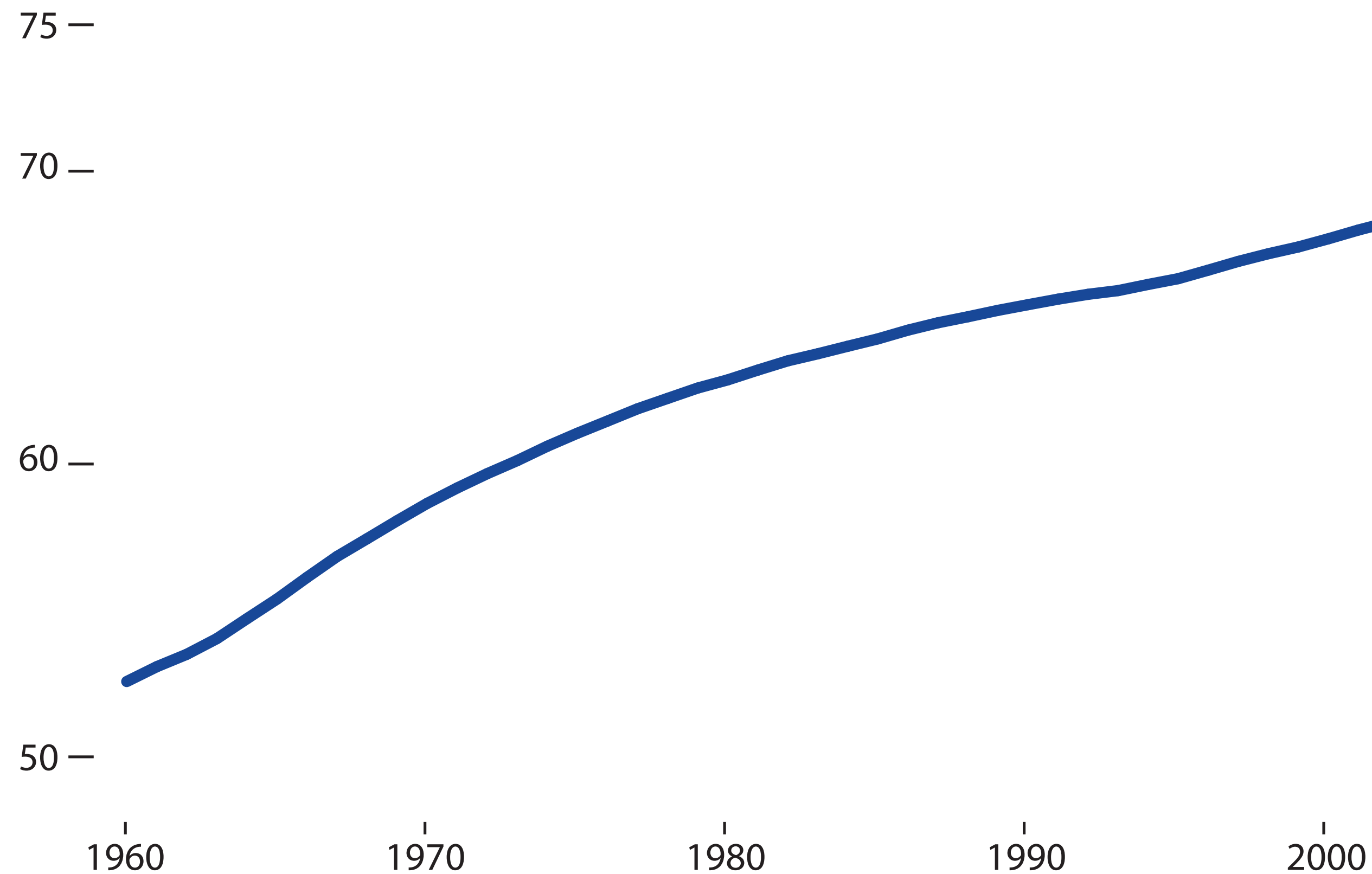
2000

2010

2016

Aspect Ratio

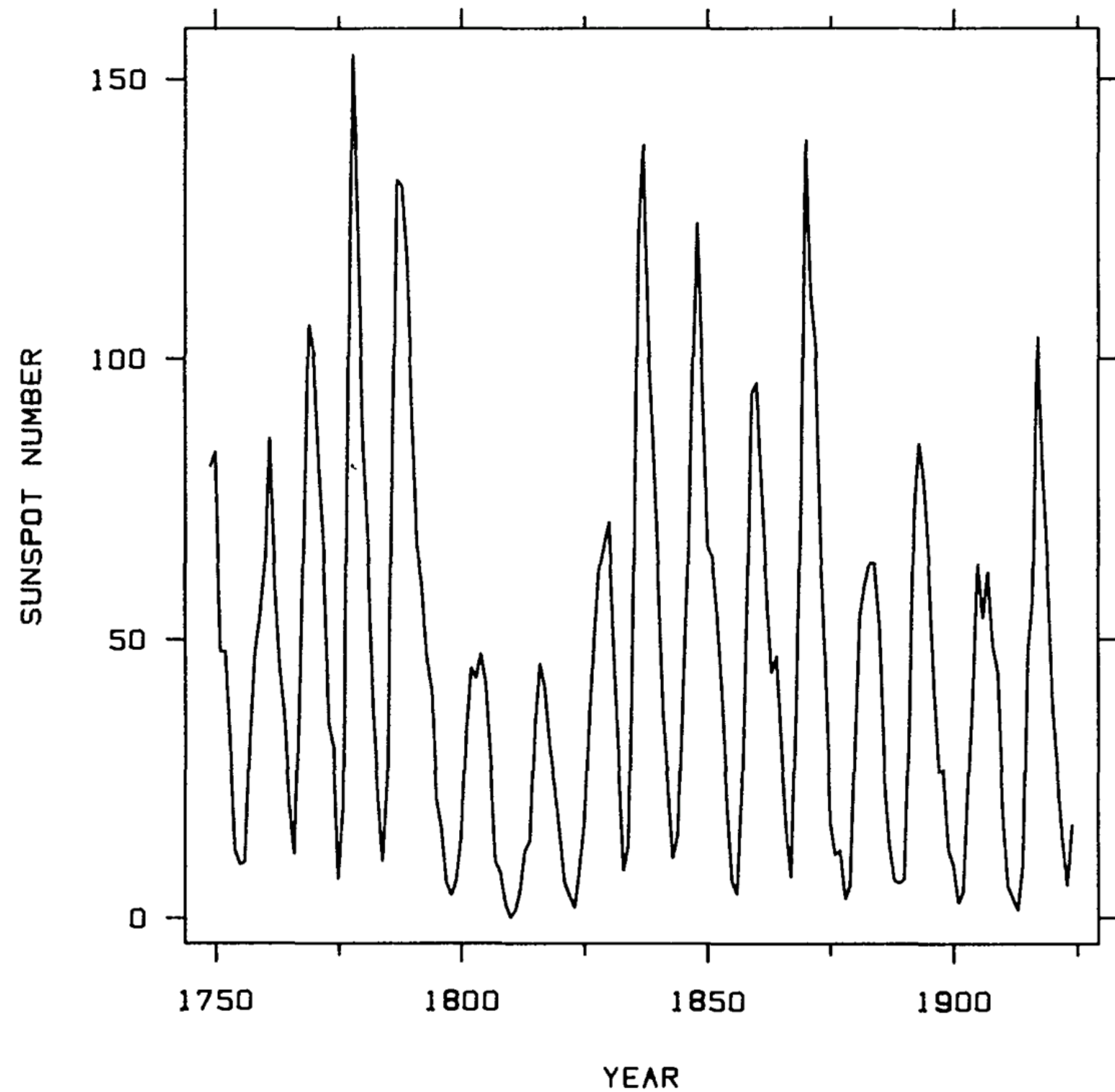
Average world life expectancy at birth (years)



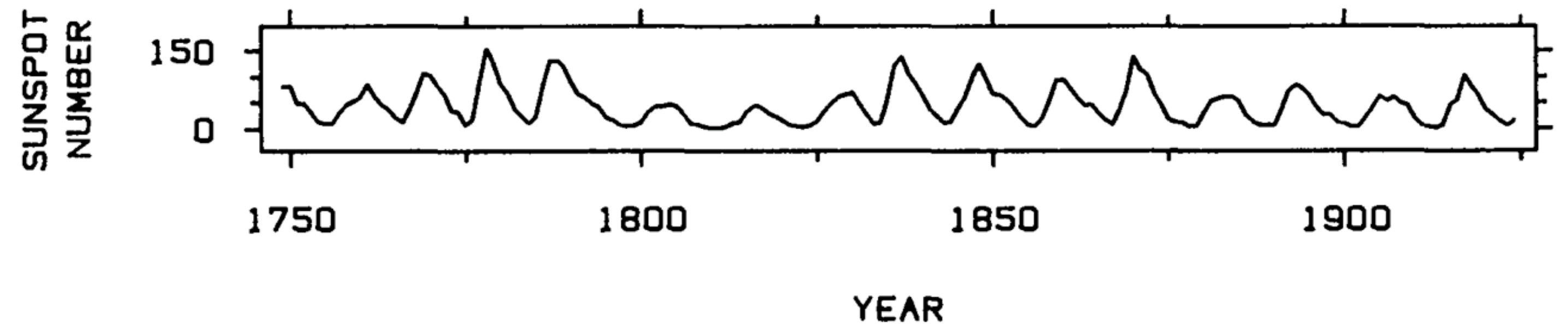
Approximate the proportion of the chart to match the depicted trend.

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

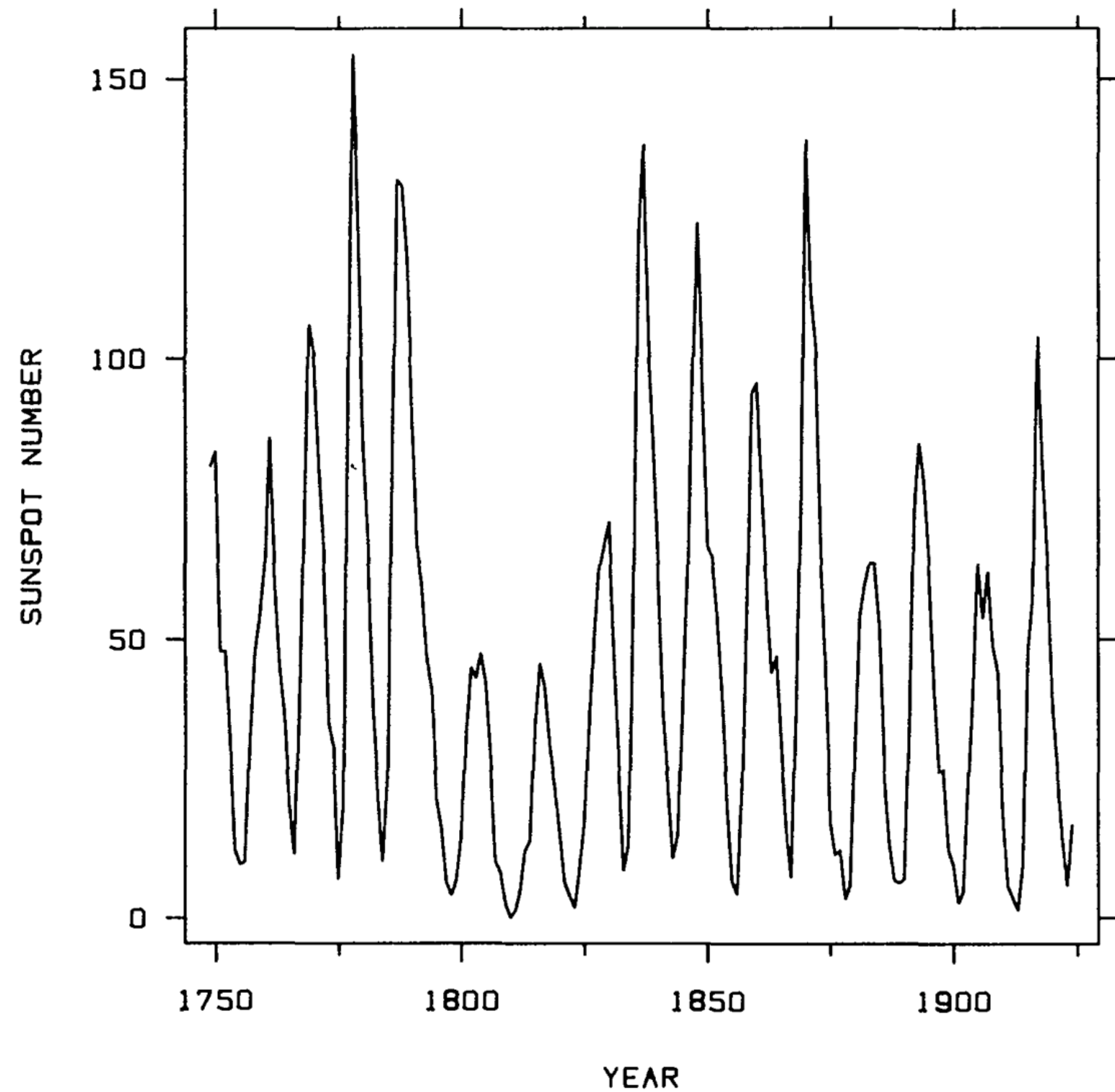
Aspect Ratio



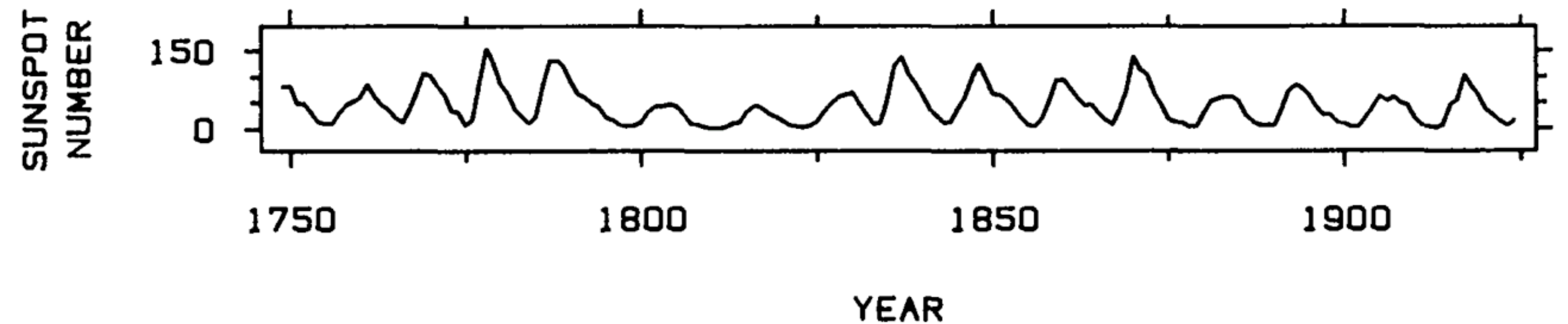
1. Approximate the proportion of the chart to match the depicted trend.



Aspect Ratio



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

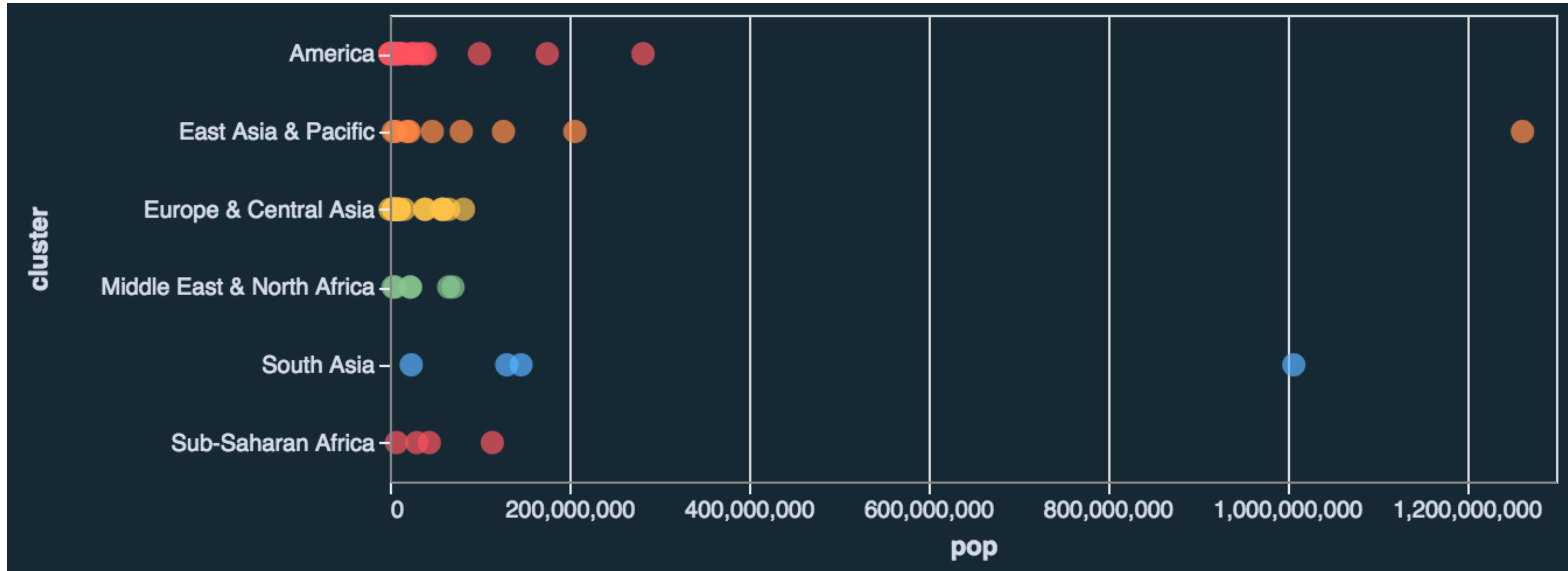


2. Bank to 45°: aspect ratios with 45° avg. line segment orientation.

Scaling Axes: Outliers and Skew

Options:

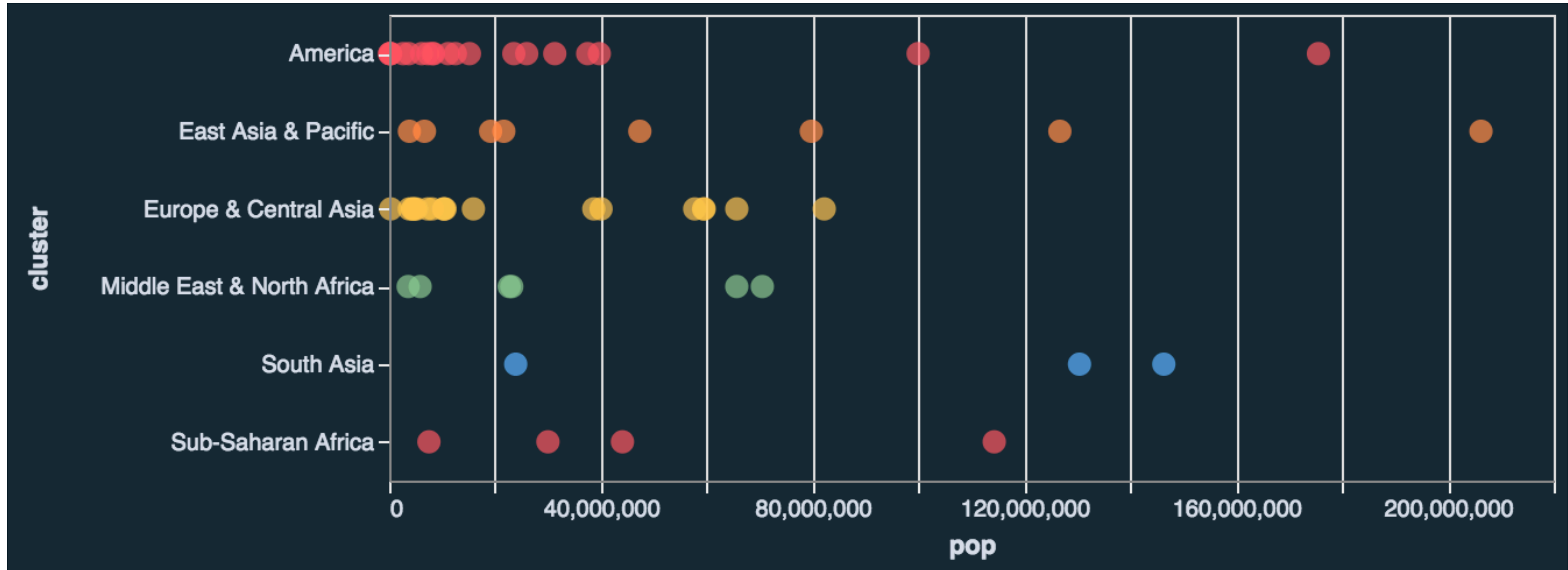
1. Clip them out.



Scaling Axes: Outliers and Skew

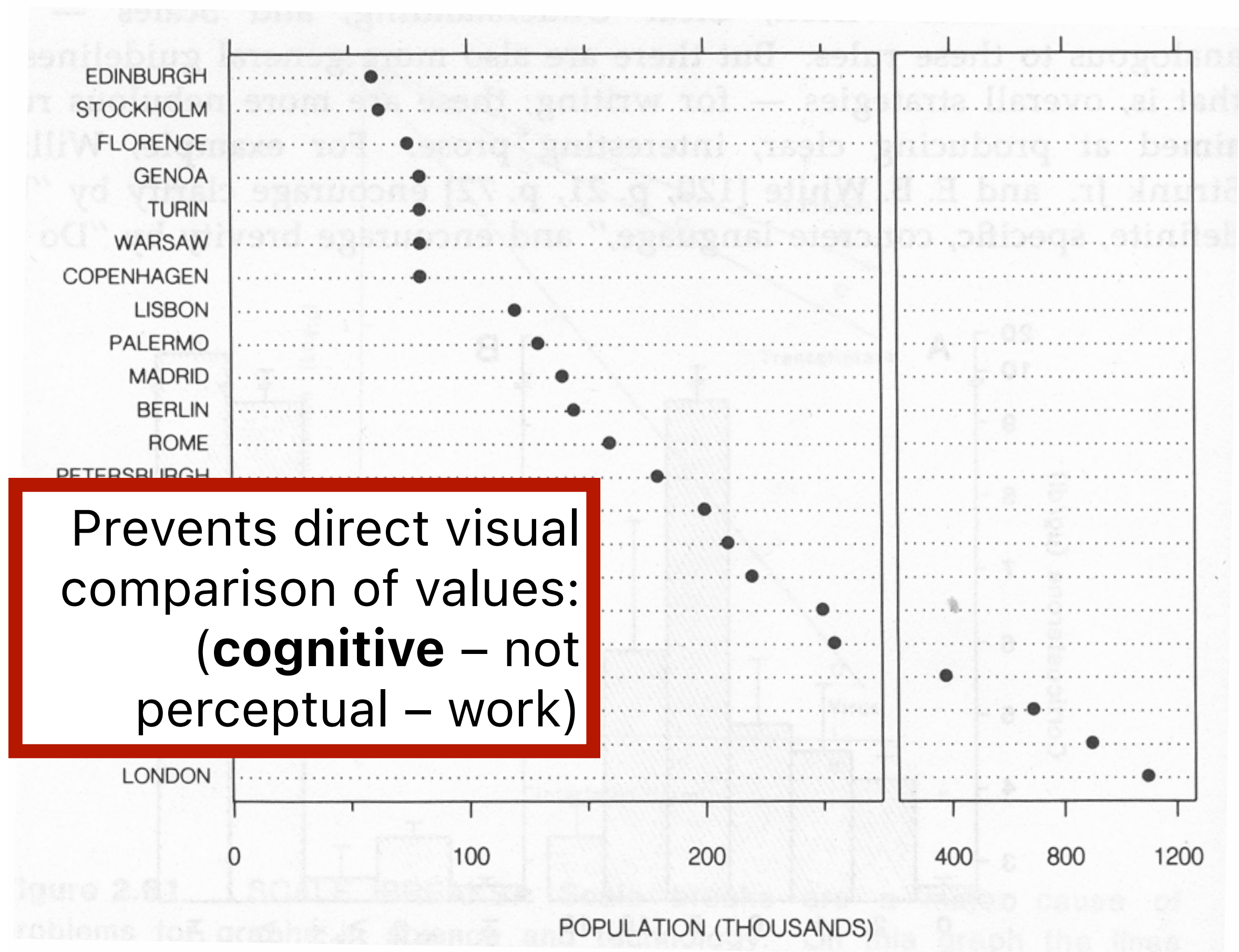
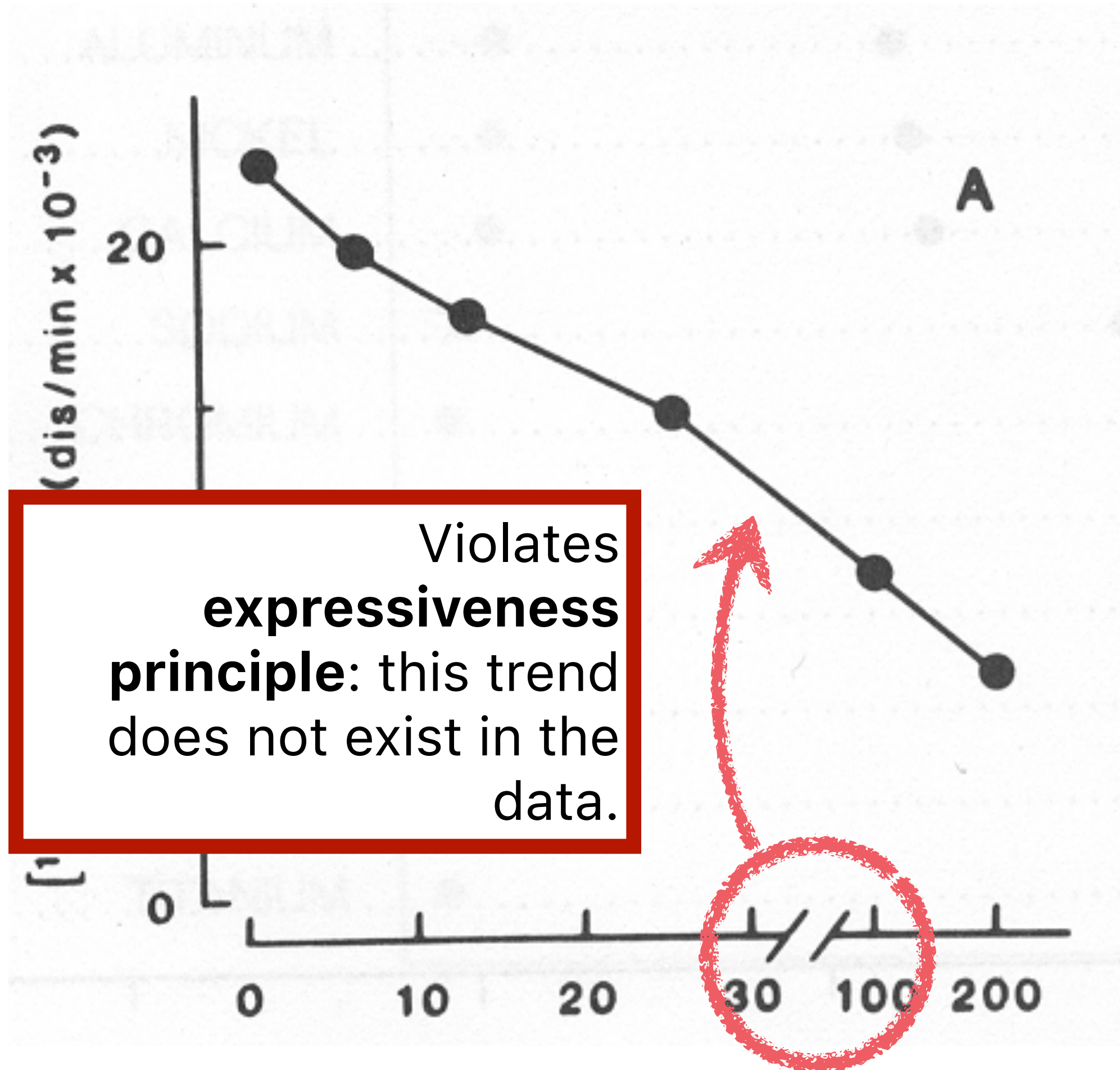
Options:

1. Clip them out.



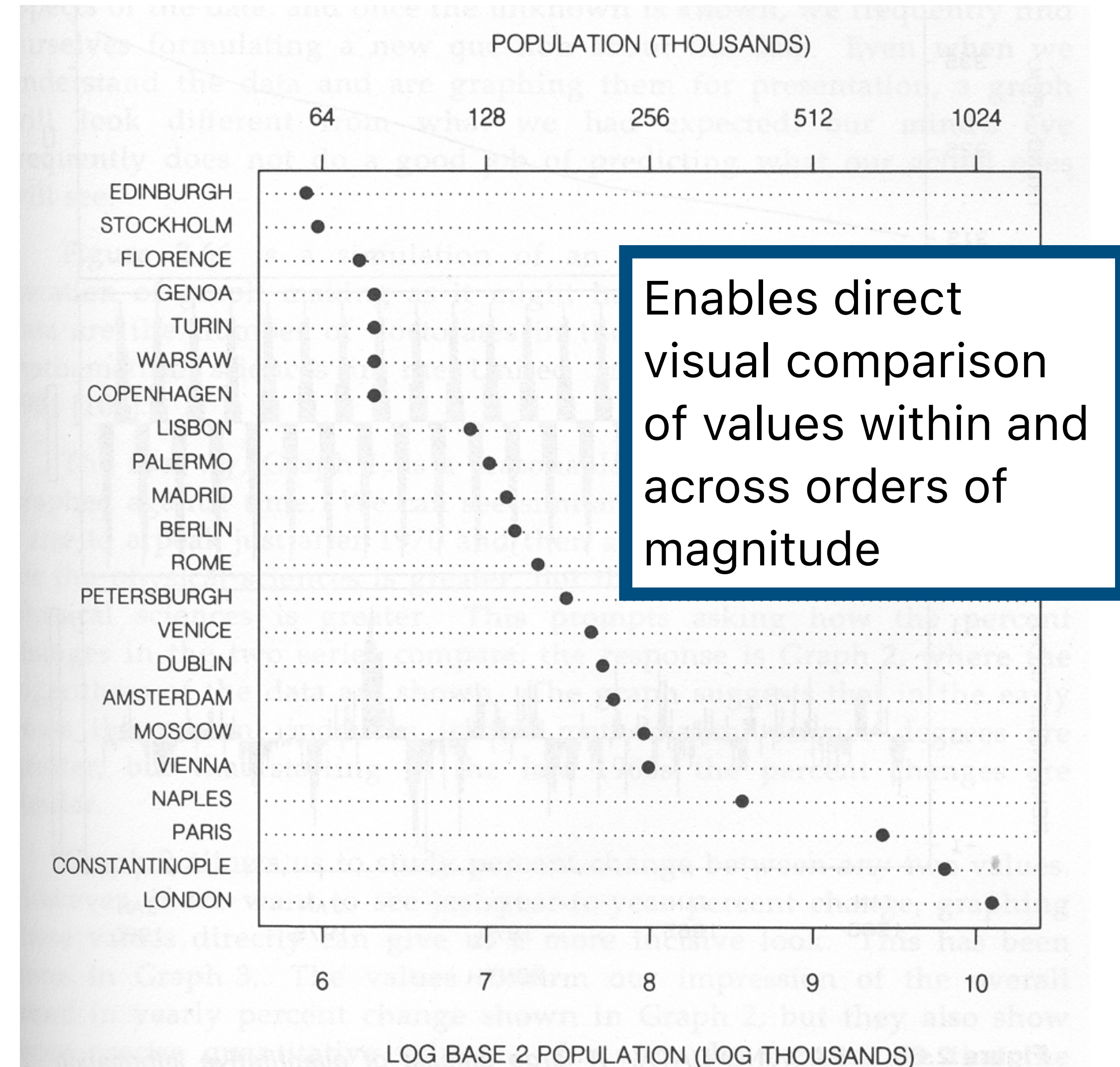
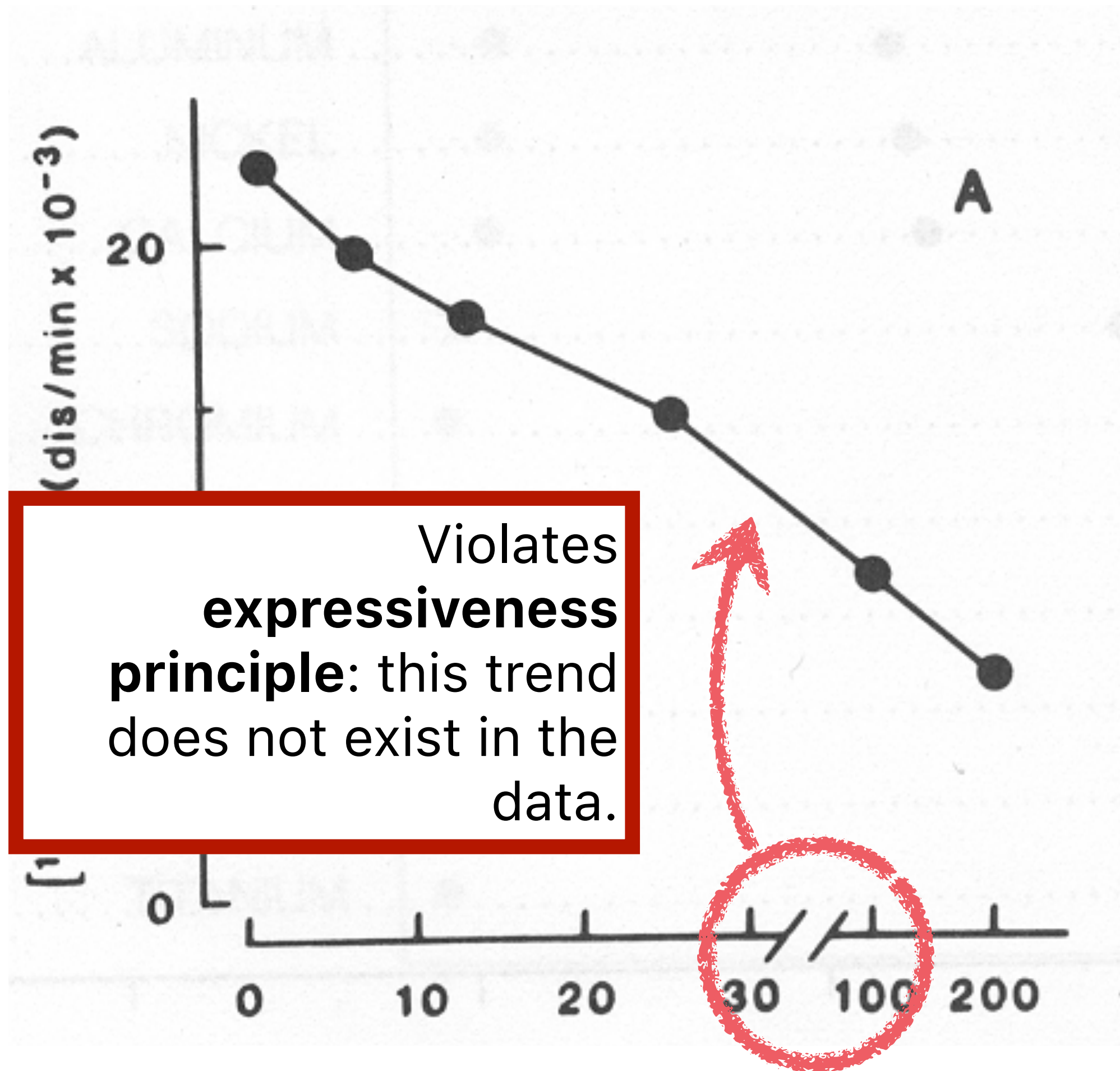
Scaling Axes: Outliers and Skew

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



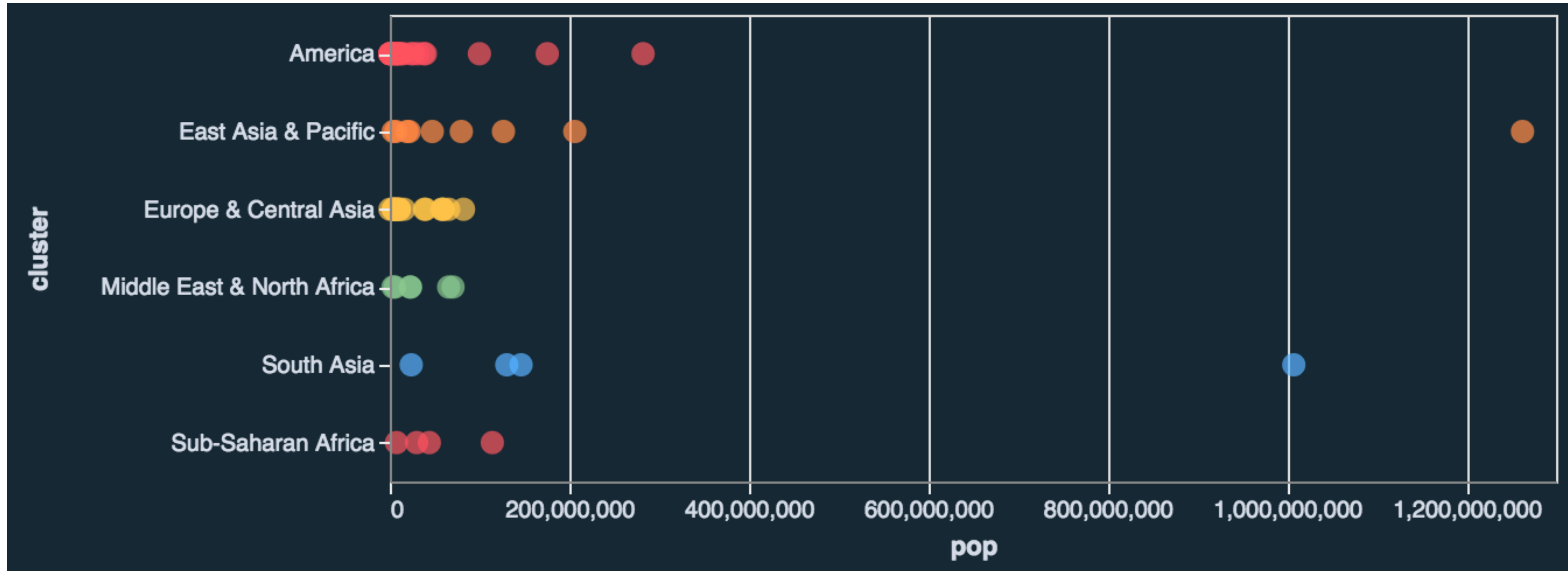
Scaling Axes: Outliers and Skew

- 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

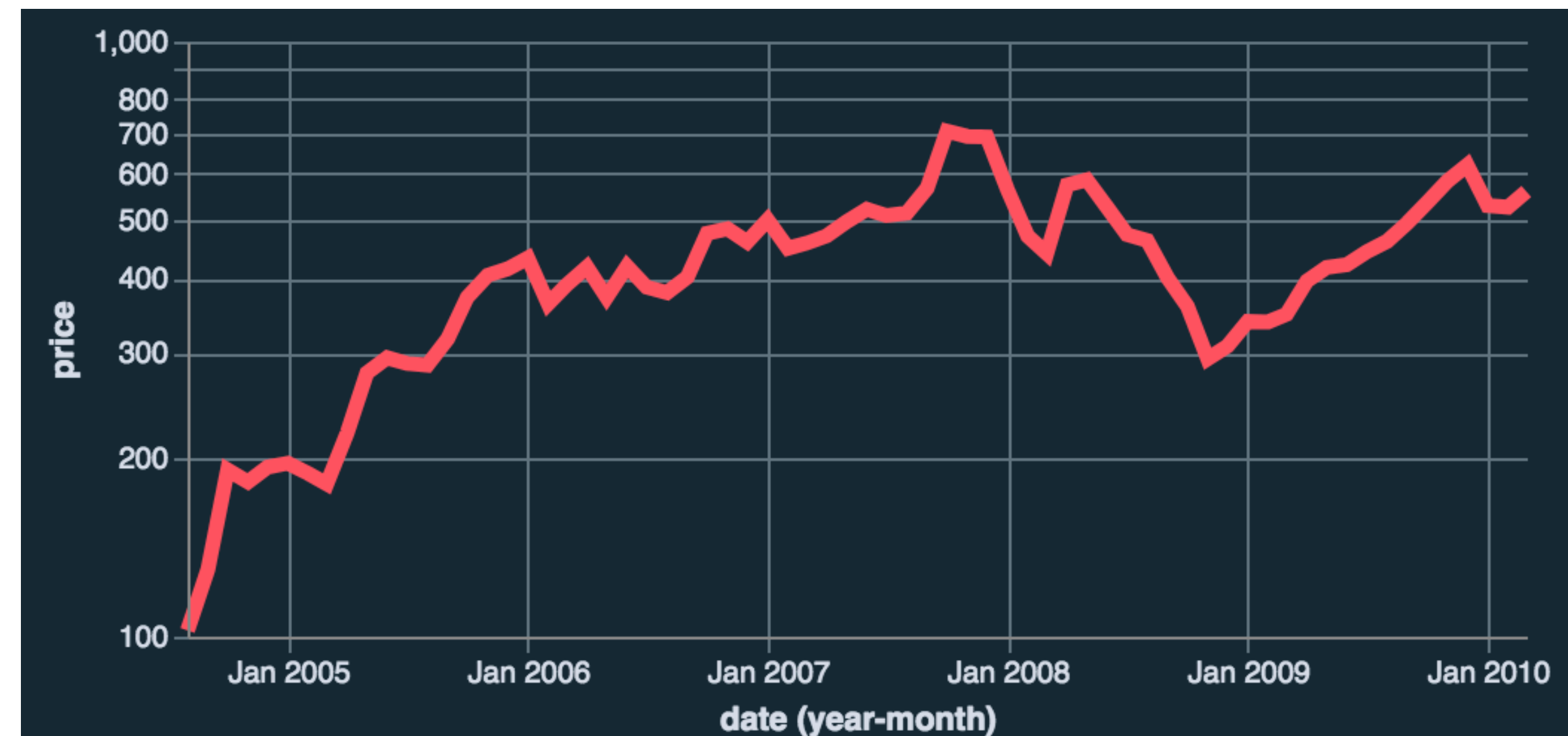
Linear Scale

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



Log Scale

Percentage change
10 visual units =
multiplication of 10 data units



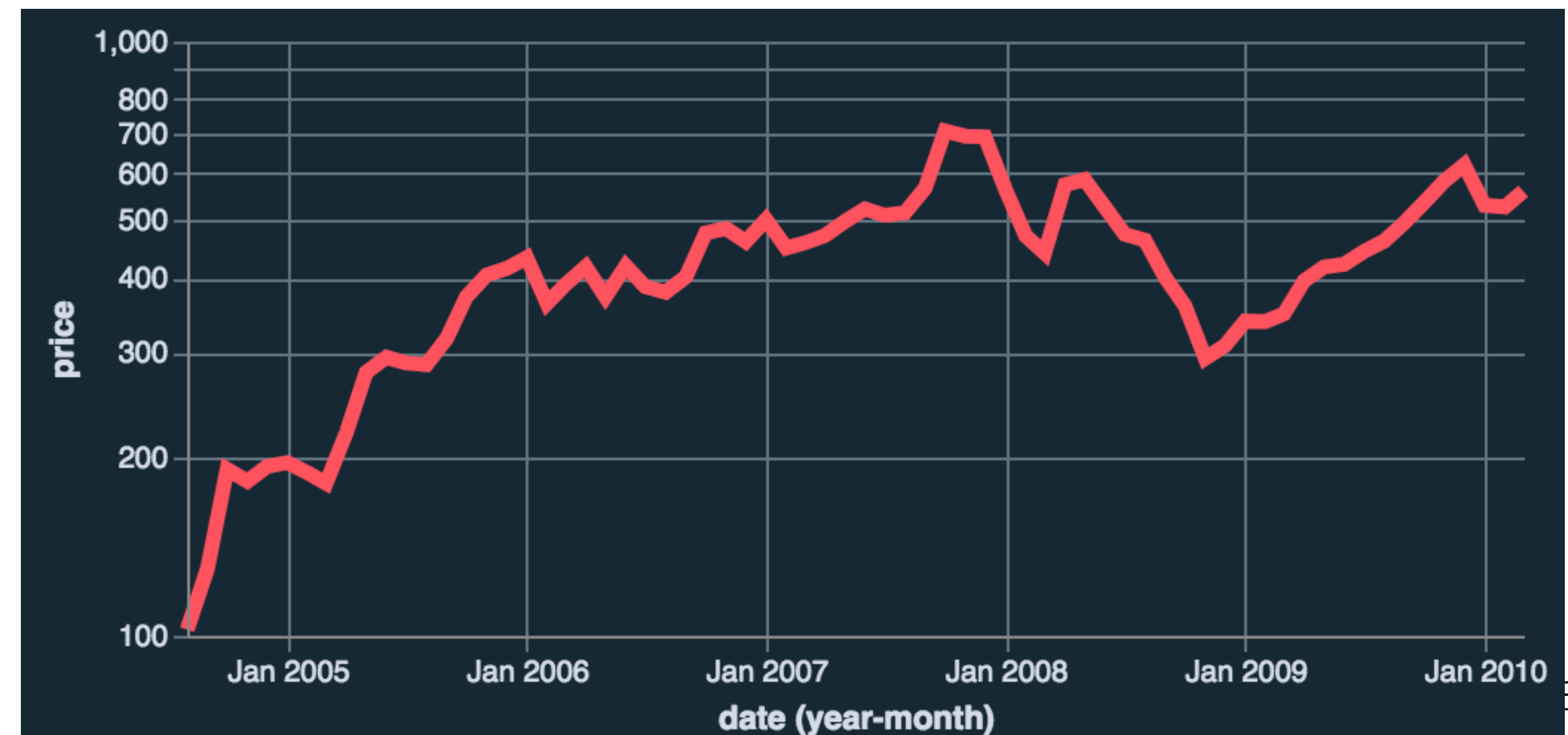
Scaling Axes: Linear vs Log

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

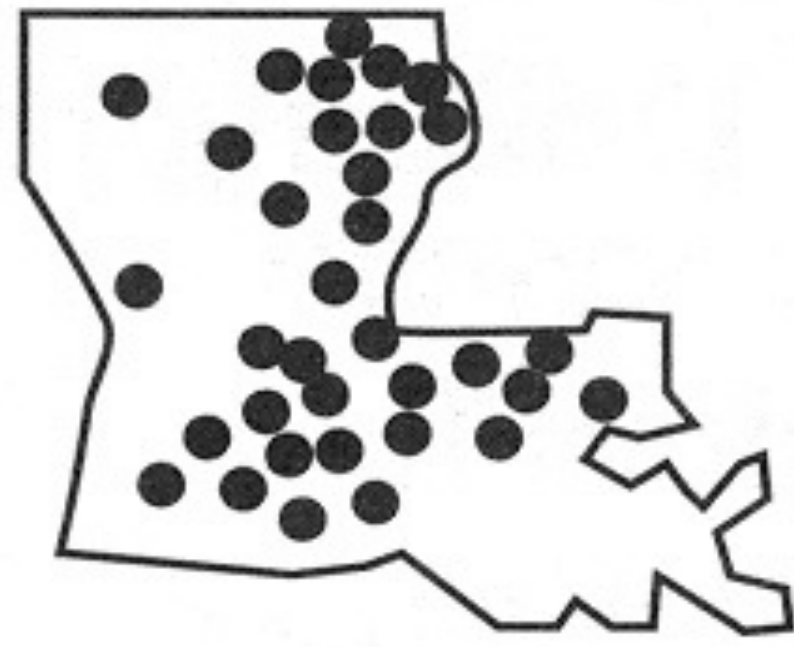
(De-)Obfuscating data

(Mis)leading the witness

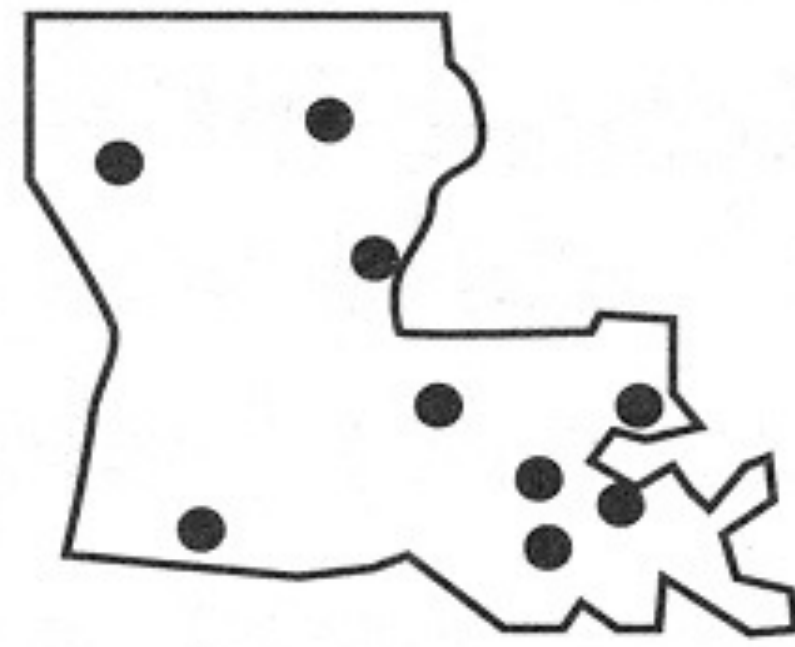


We will grade you on this for future projects!

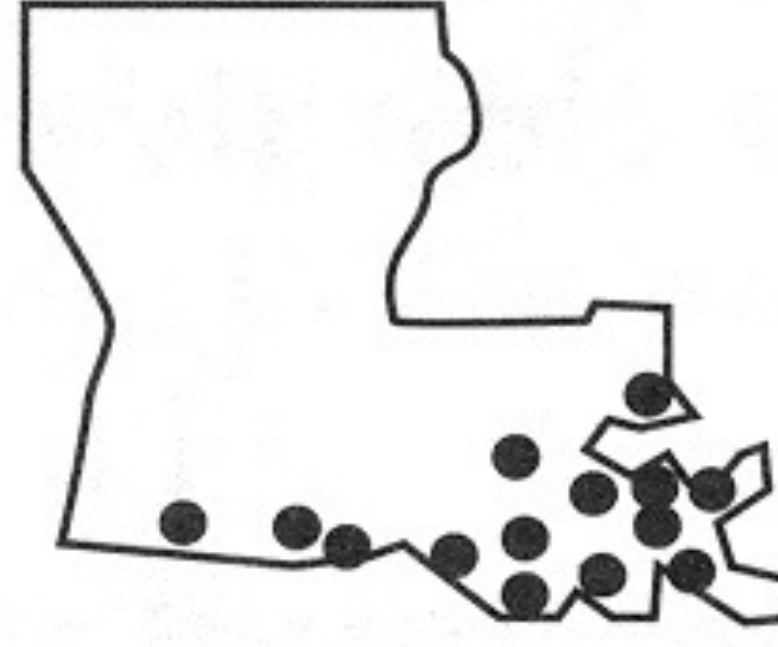
alfisol



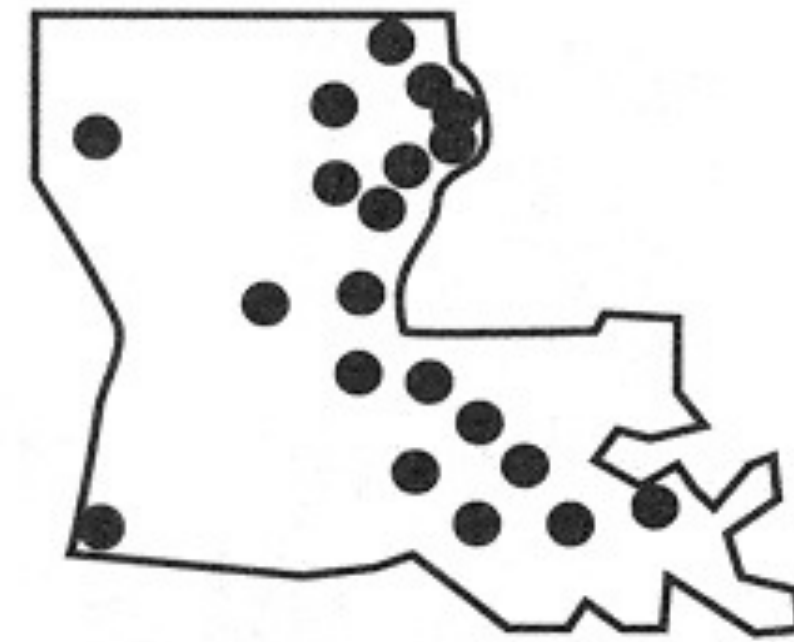
entisol



histosol



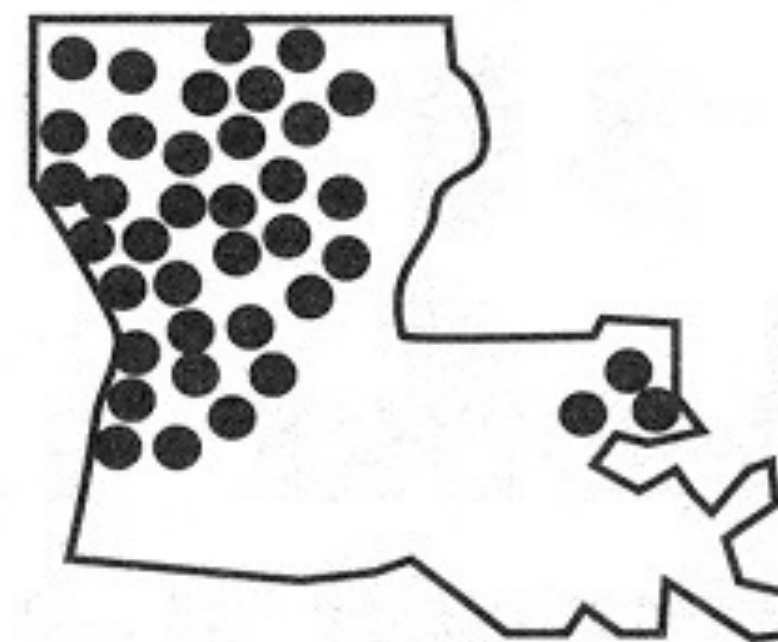
inceptisol



mollisol



ultisol



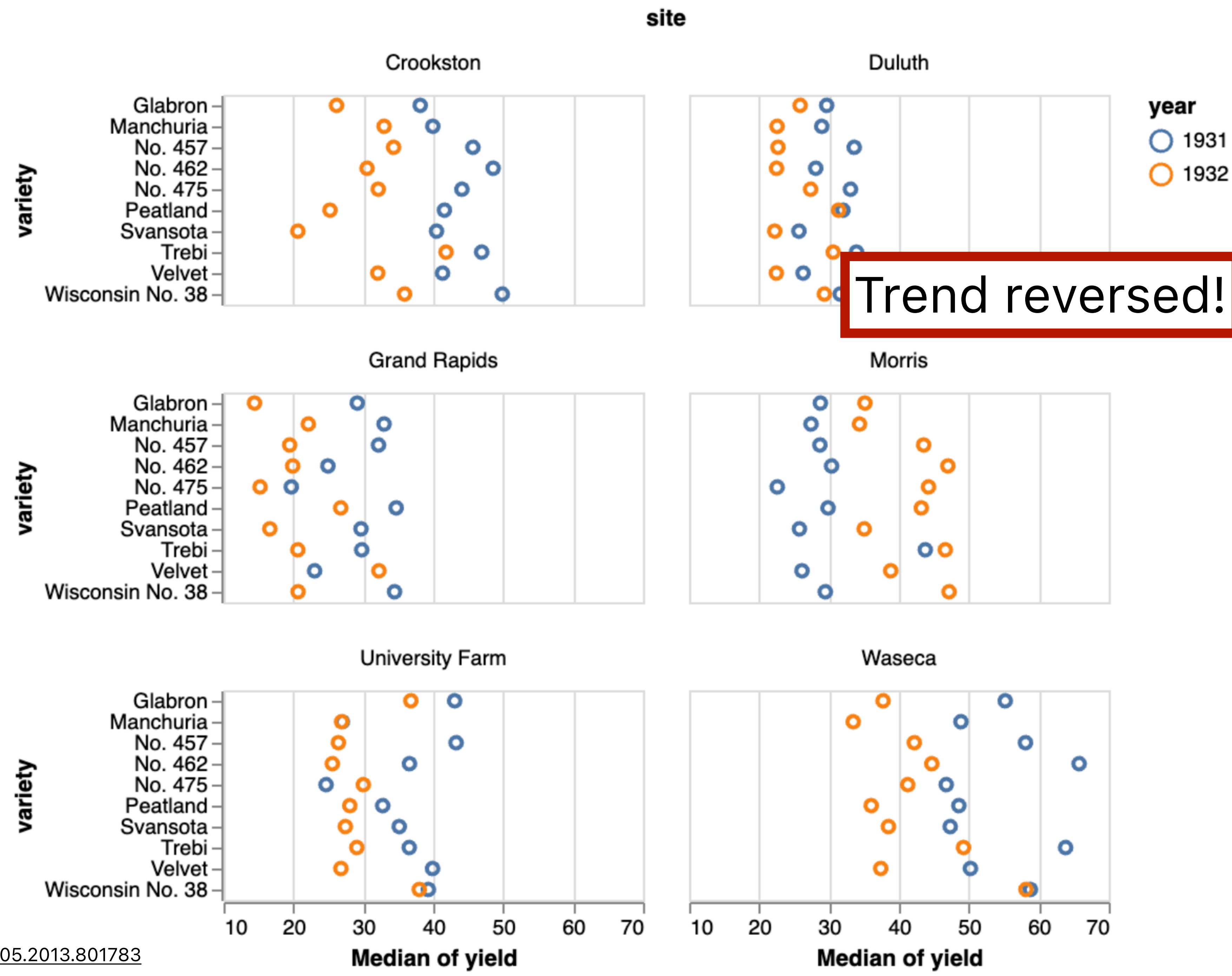
↑ Better!

↑ Bad

Small Multiples

Subdivide space to enable comparison across multiple plots.

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



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

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?

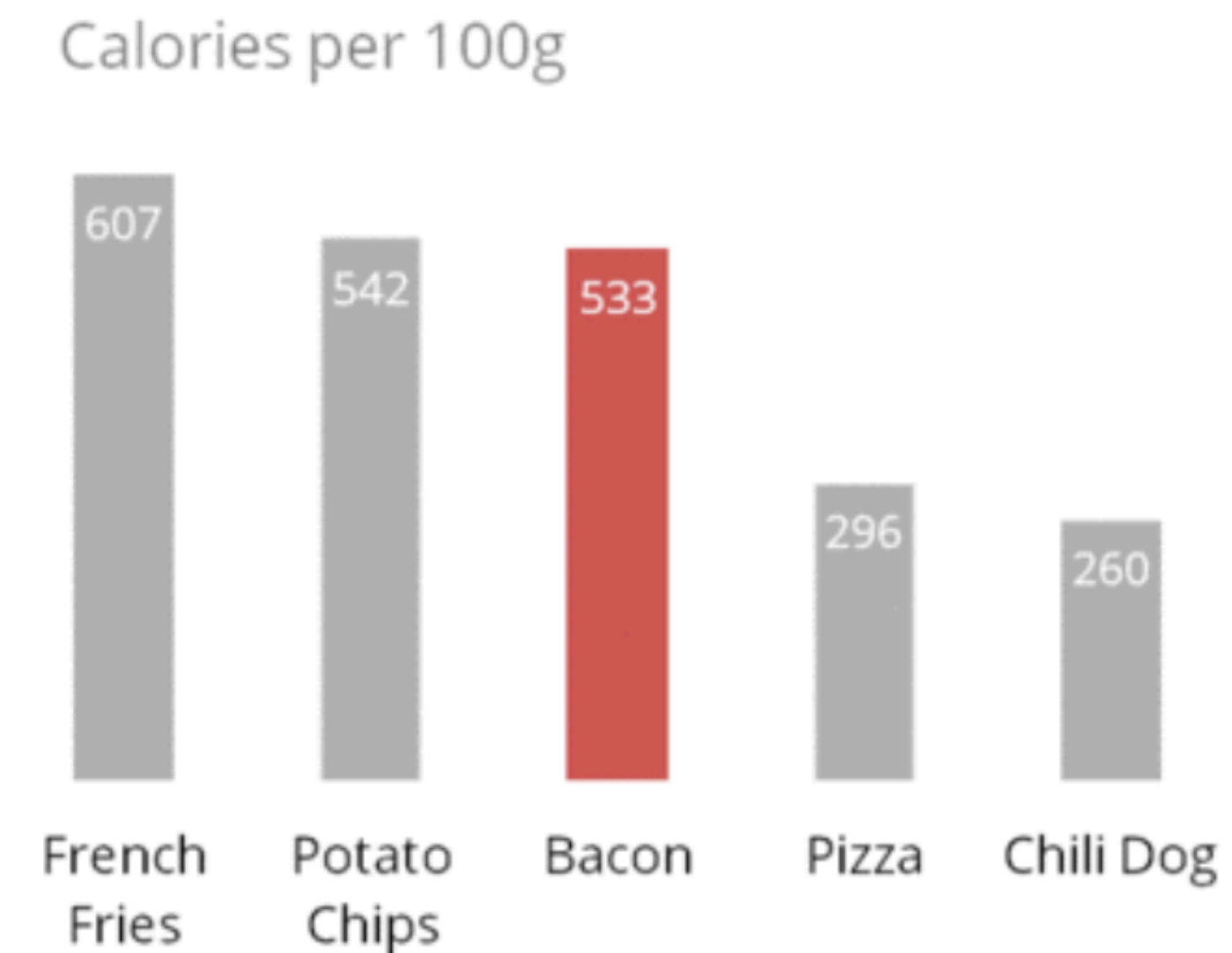
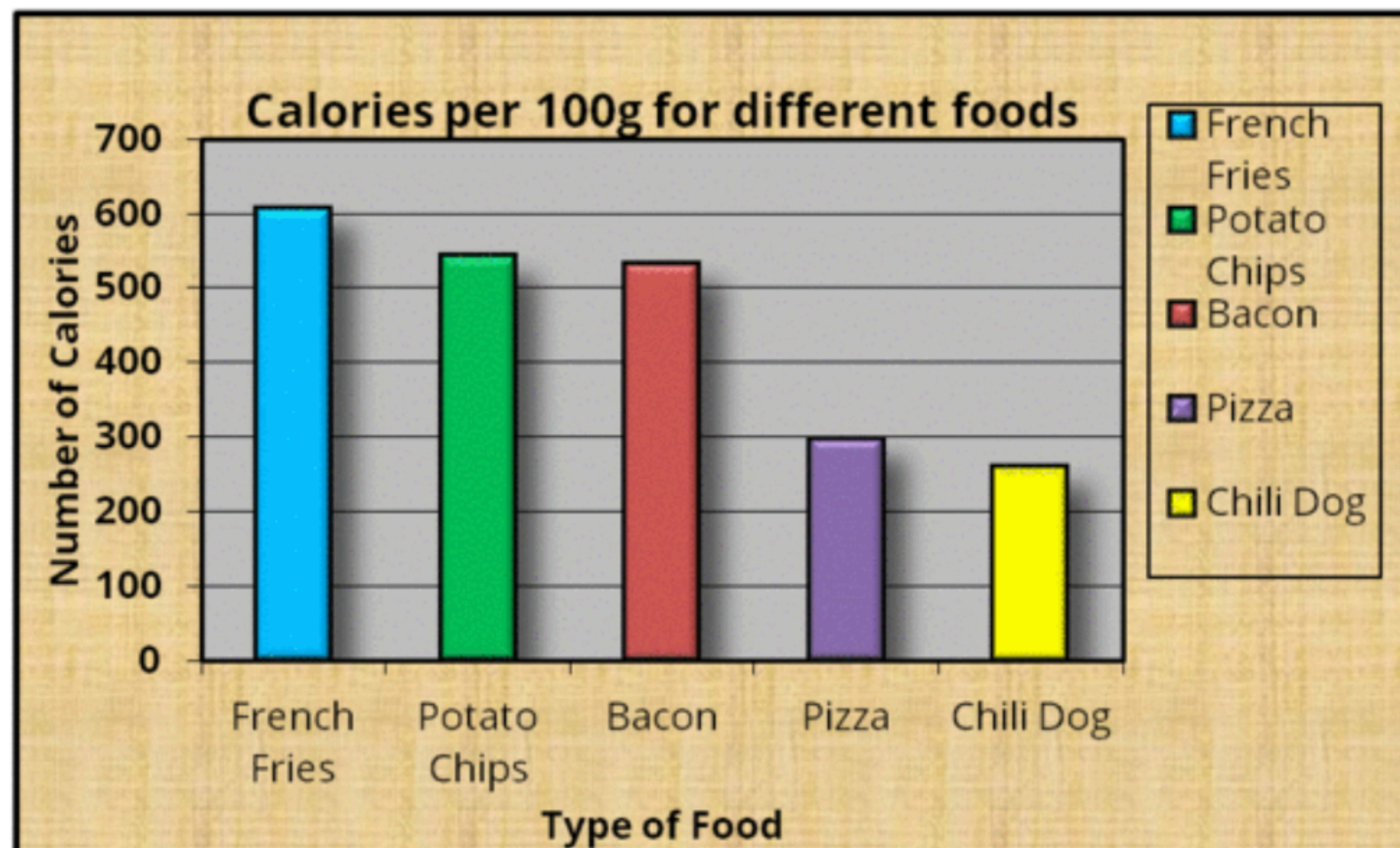


Chart "Junk"

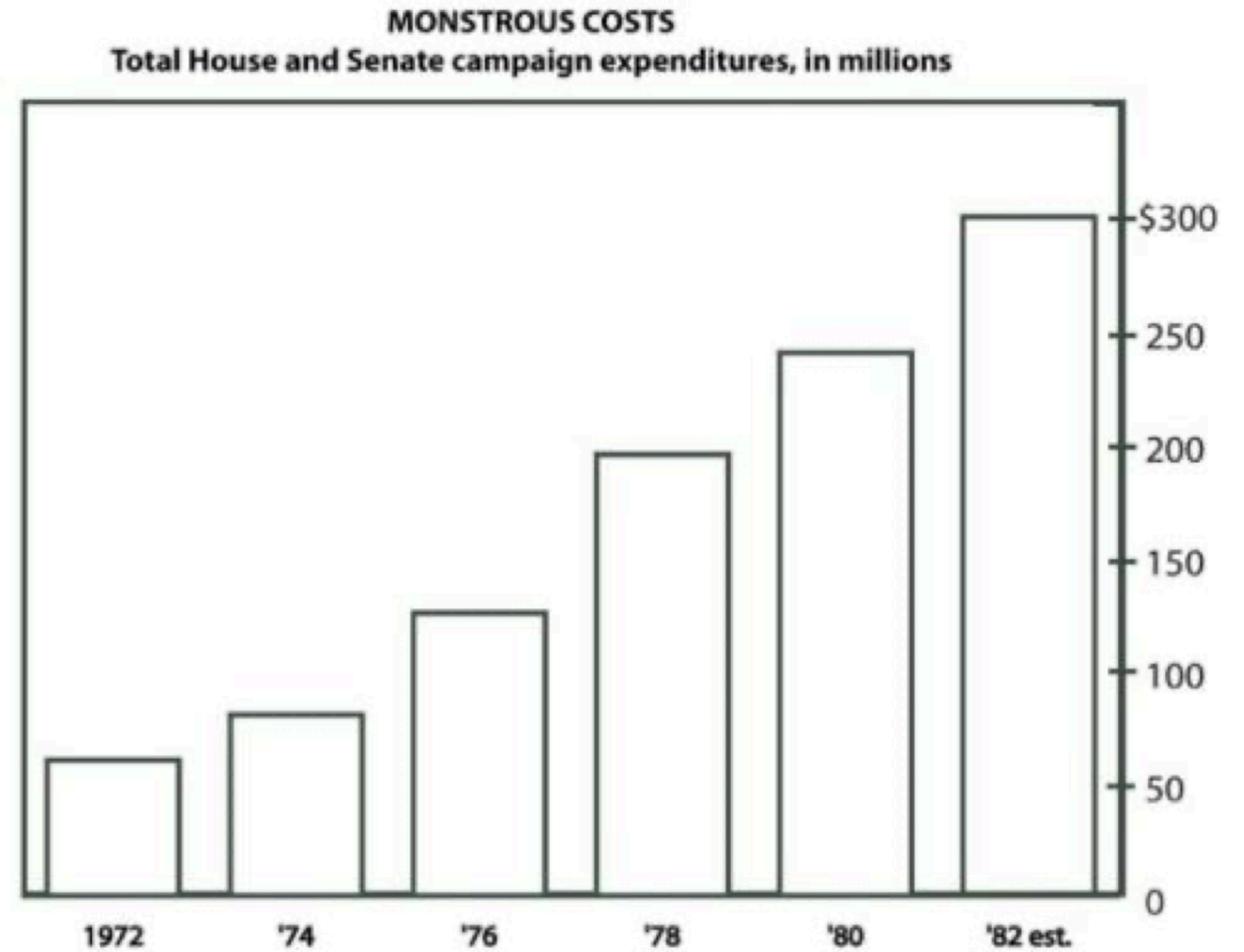
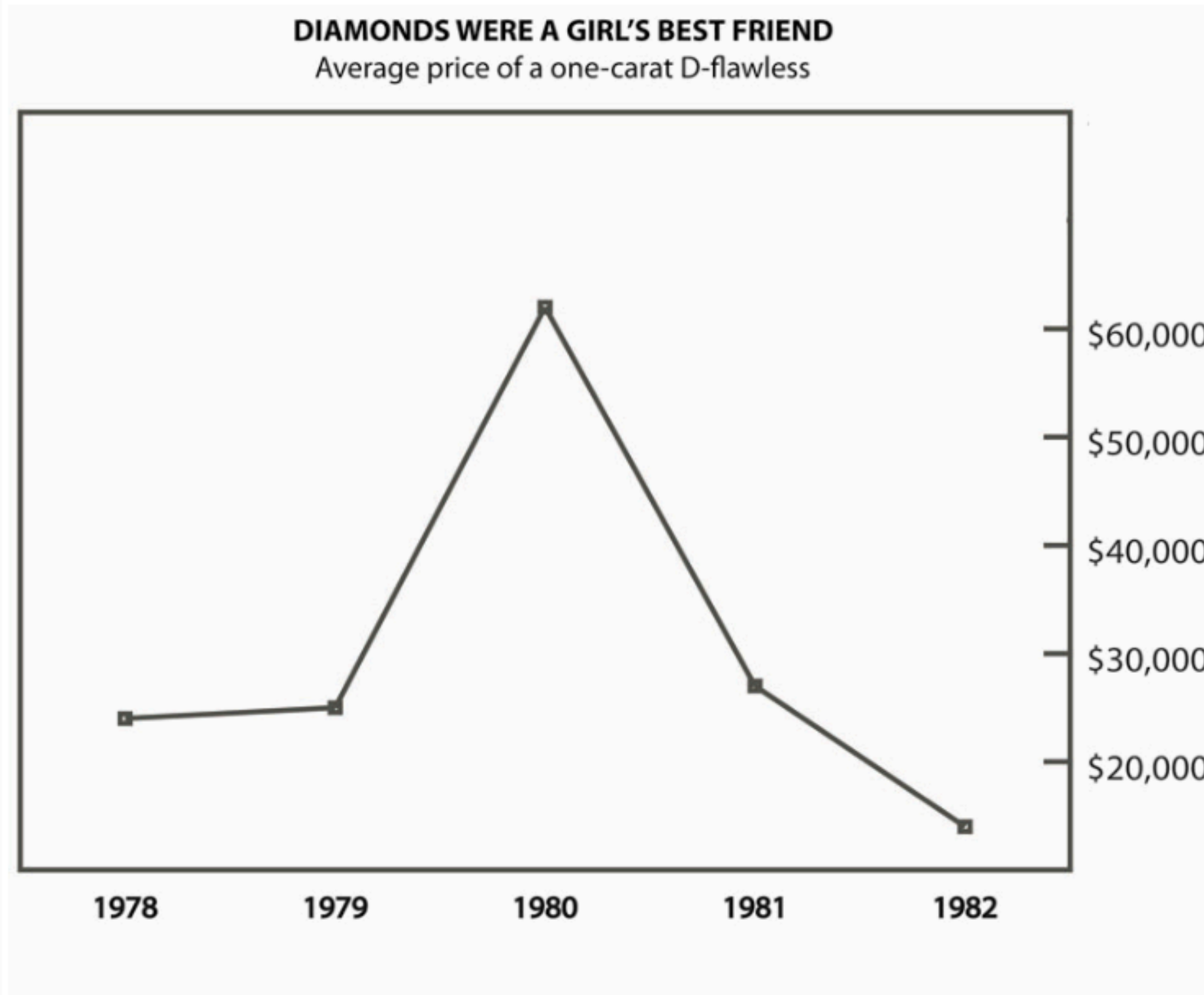
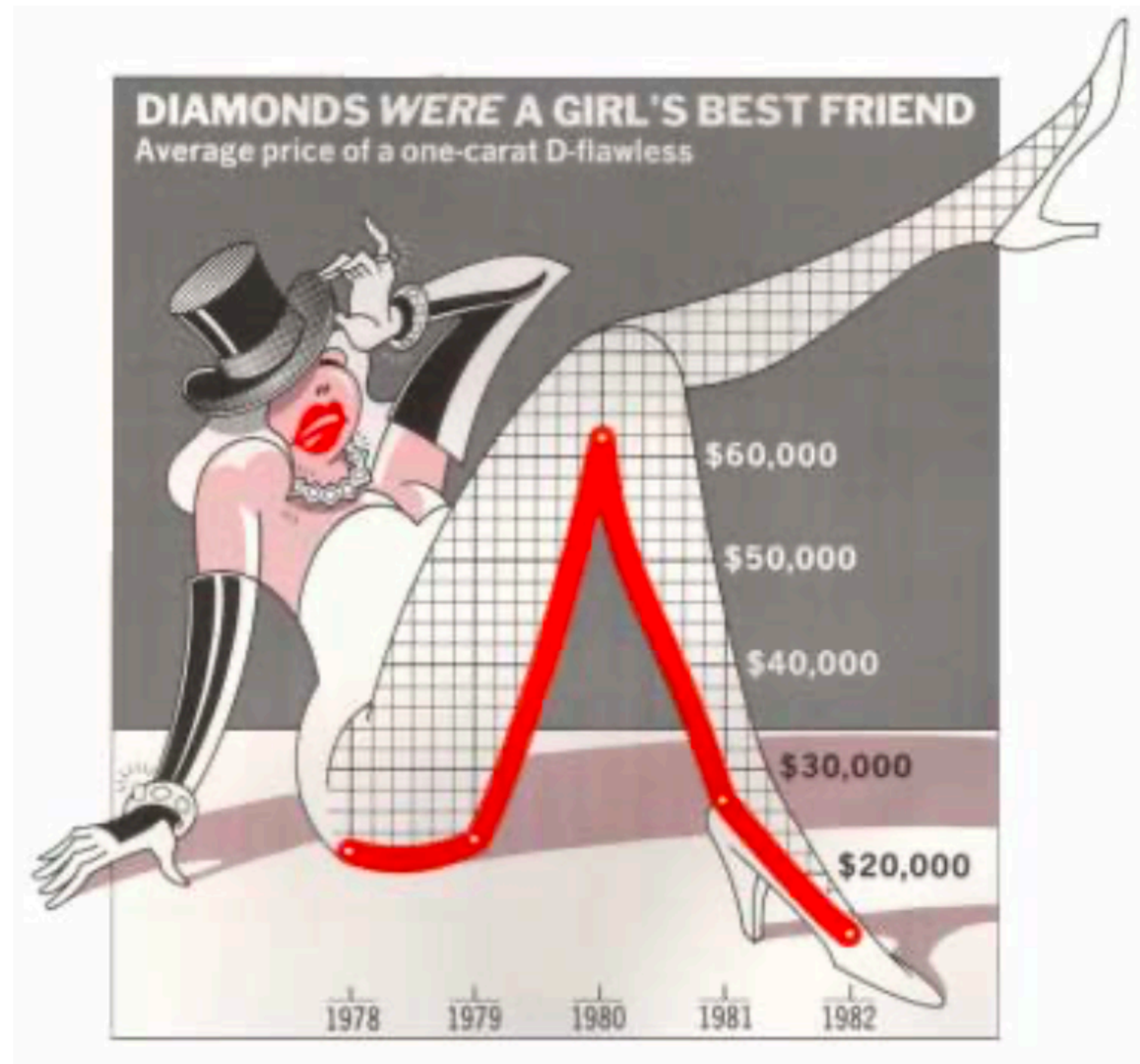
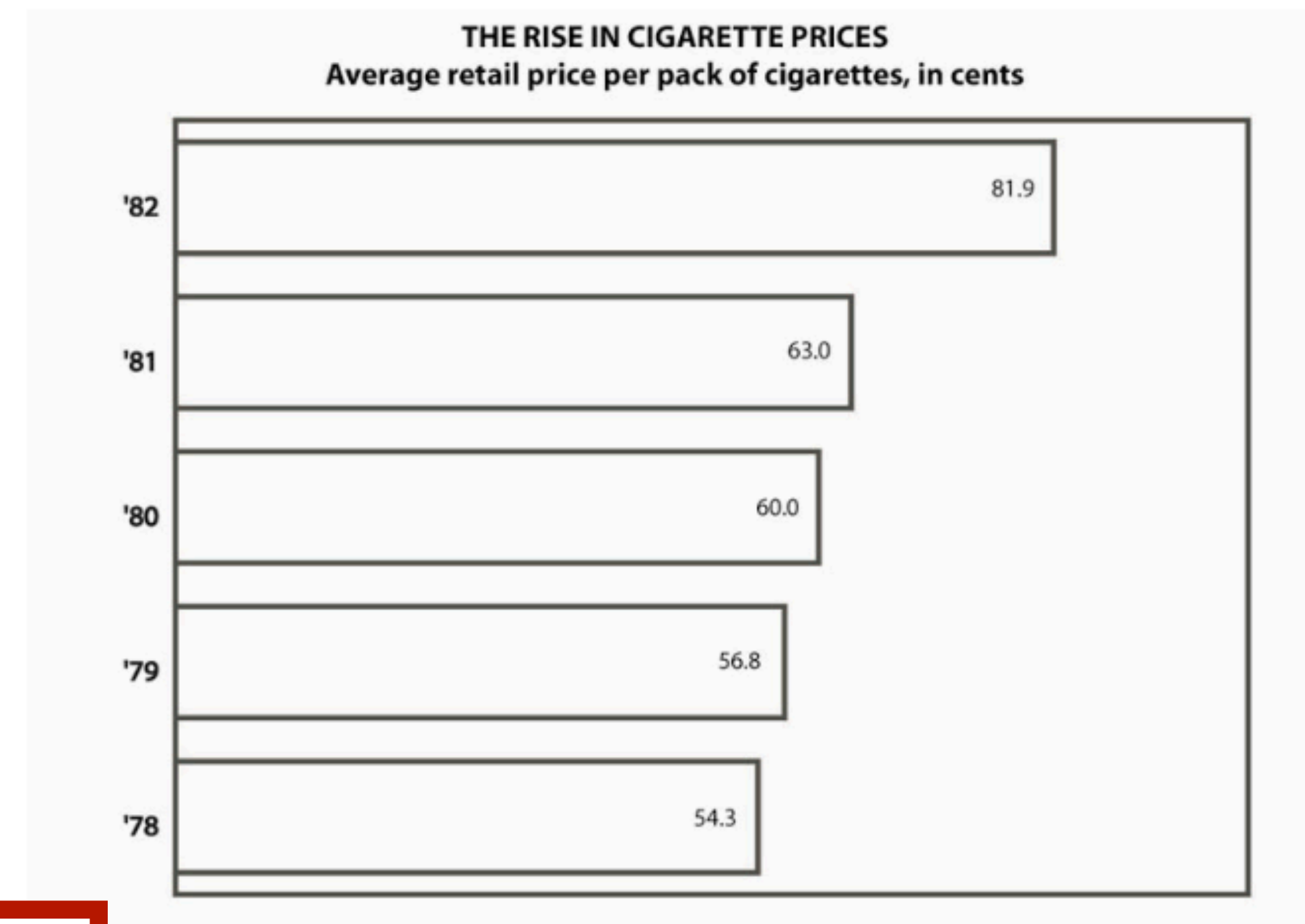
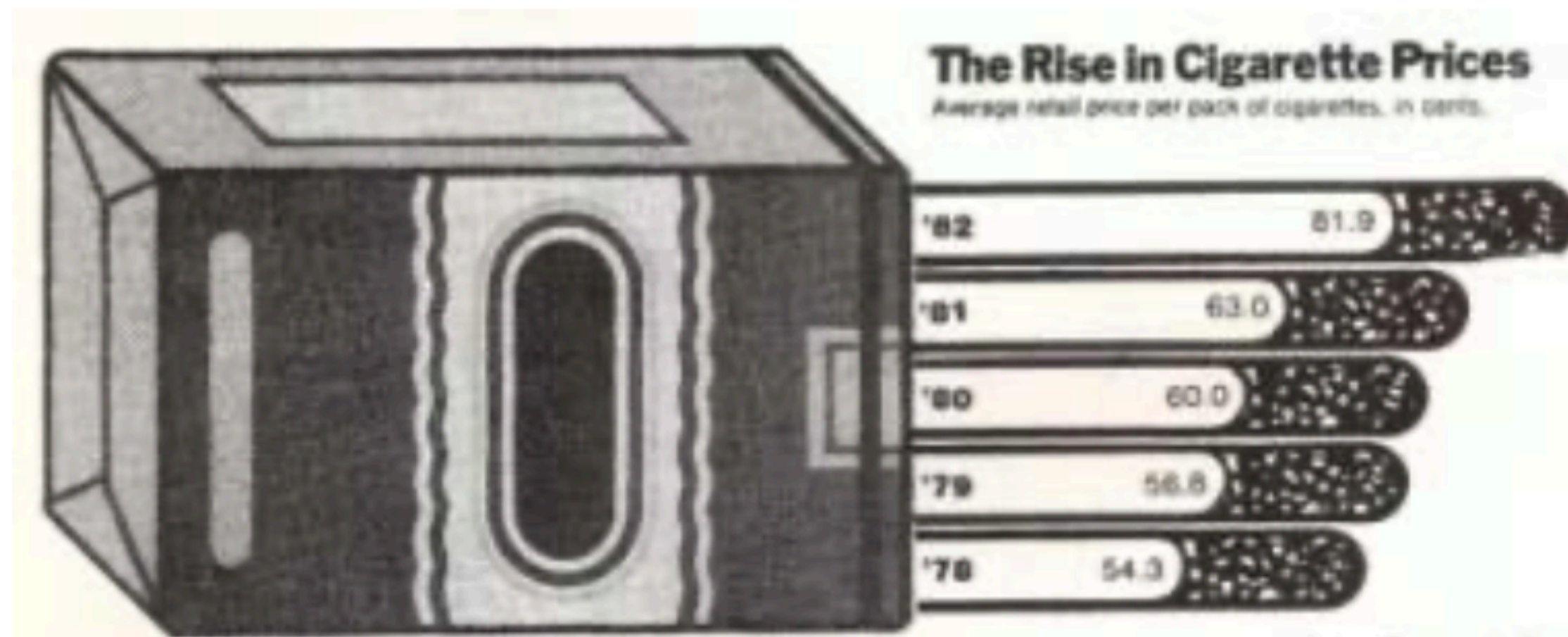


Chart "Junk"



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

Chart "Junk"



Which visualizations do you think were more effective, and why?

Using space (in)effectively

(De-)Obfuscating data

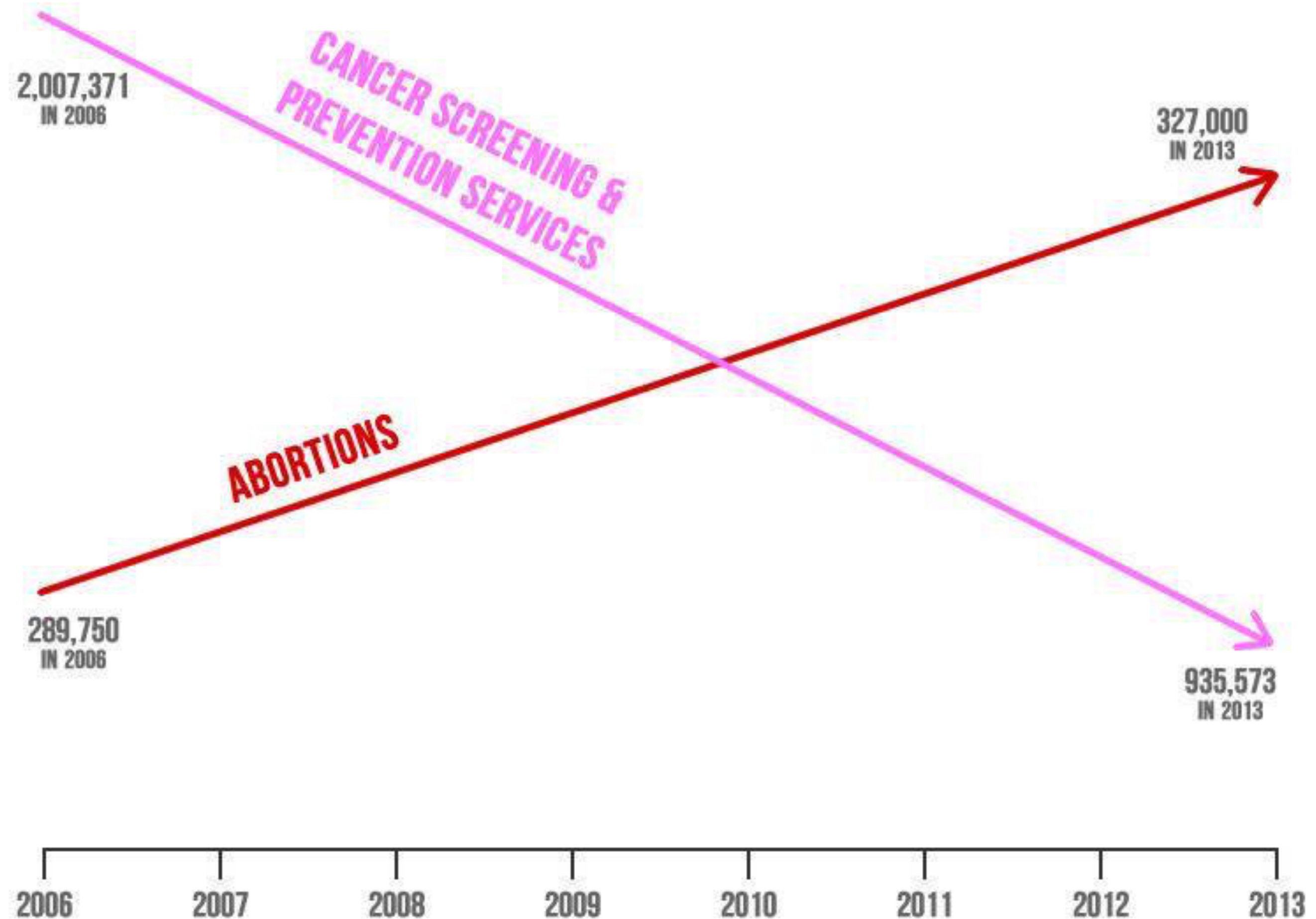
(Mis)leading the witness

Using space (in)effectively

(De-)Obfuscating data

(Mis)leading the witness

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

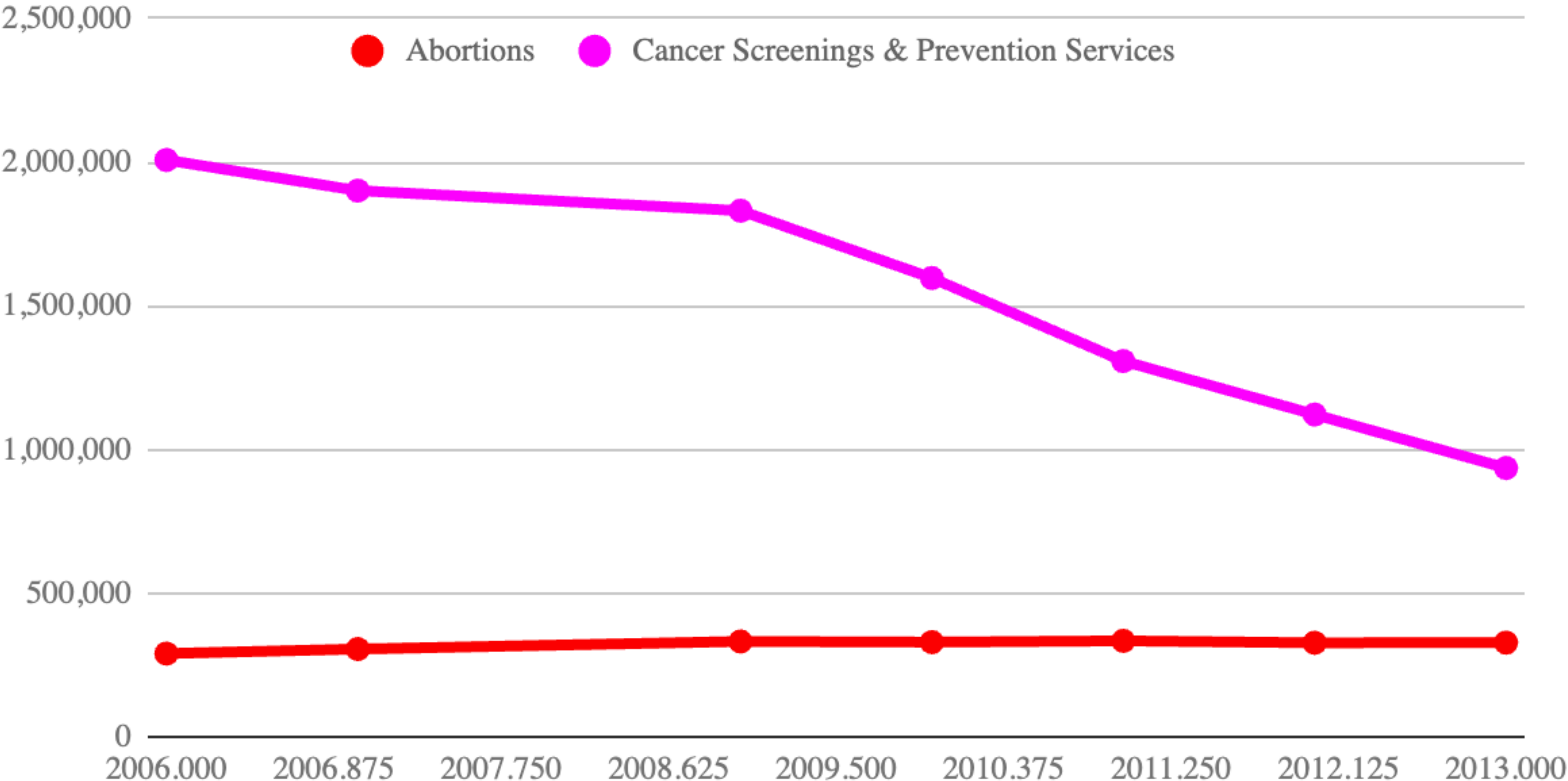


SOURCE: AMERICANS UNITED FOR LIFE

What are the issues with this chart?

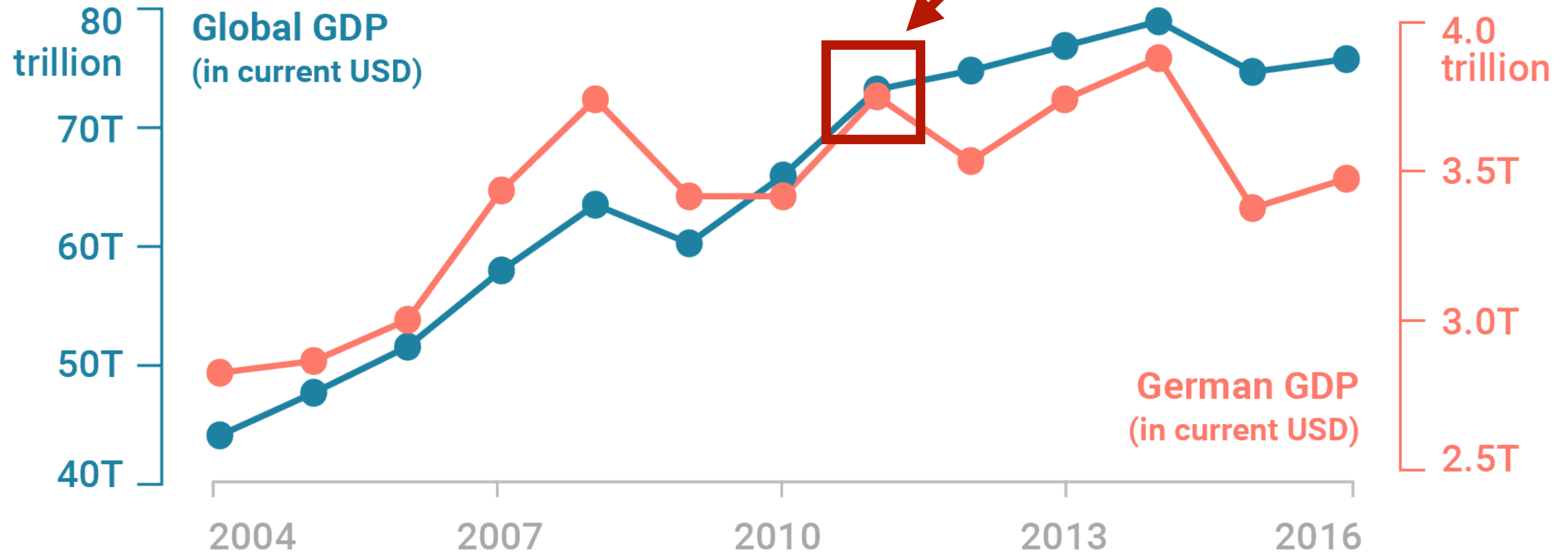
tryclassbuzz.com:
senate

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



Dual Axes Charts

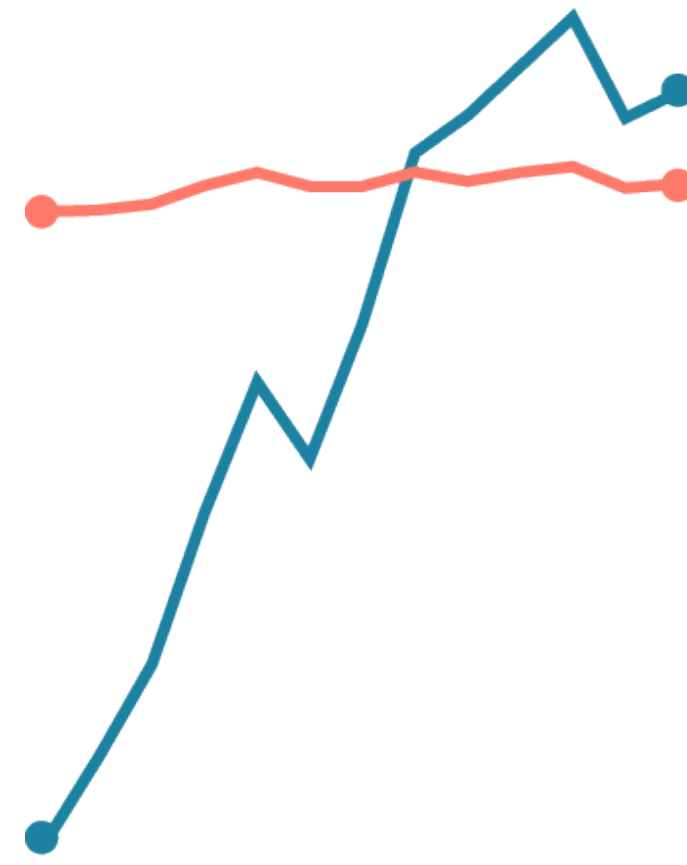
German and world GDP were equal in 2011??



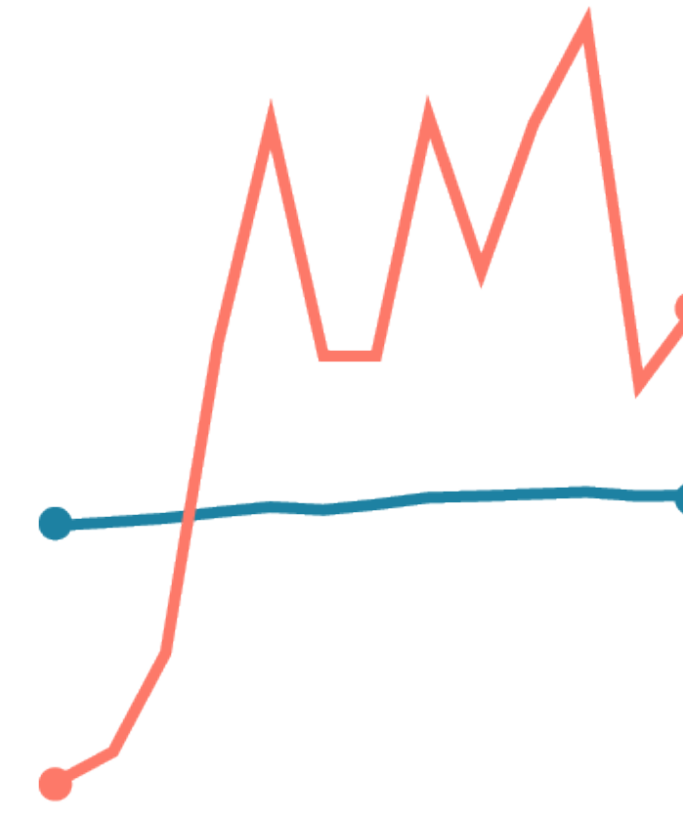
We will grade you on this for future projects!

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

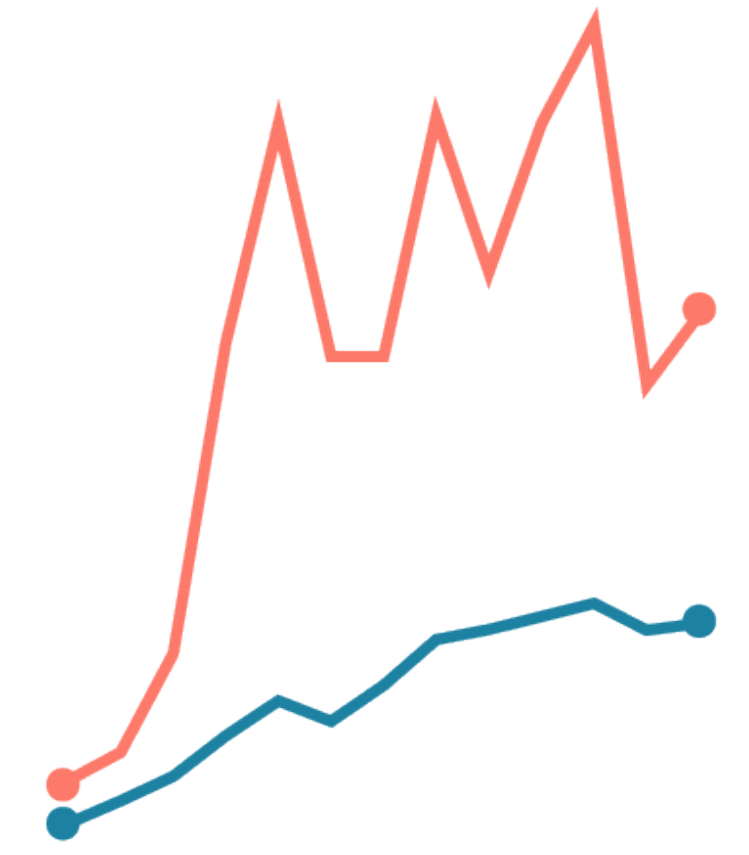
Dual-Axes Charts



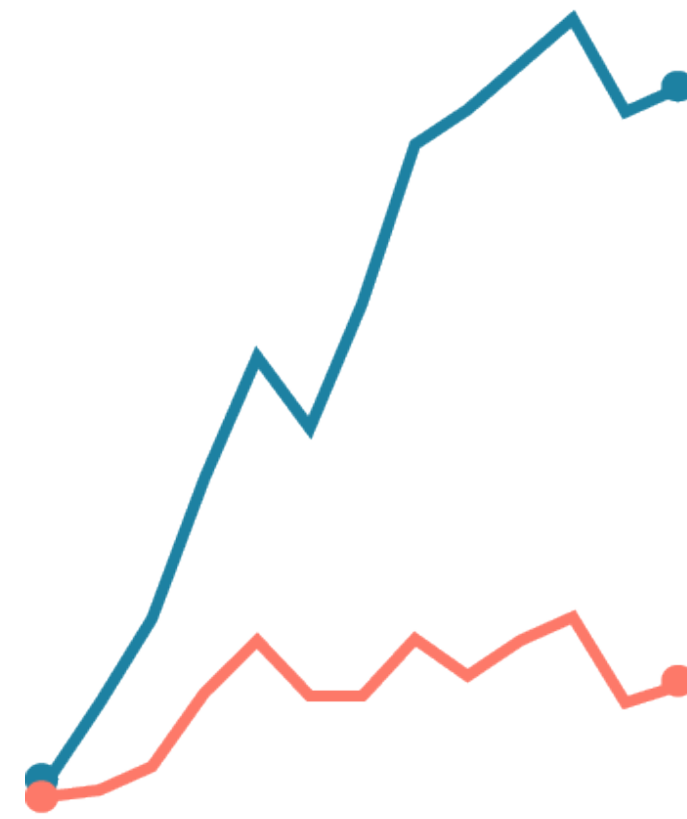
Orange steady,
Blue massively increasing.



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.

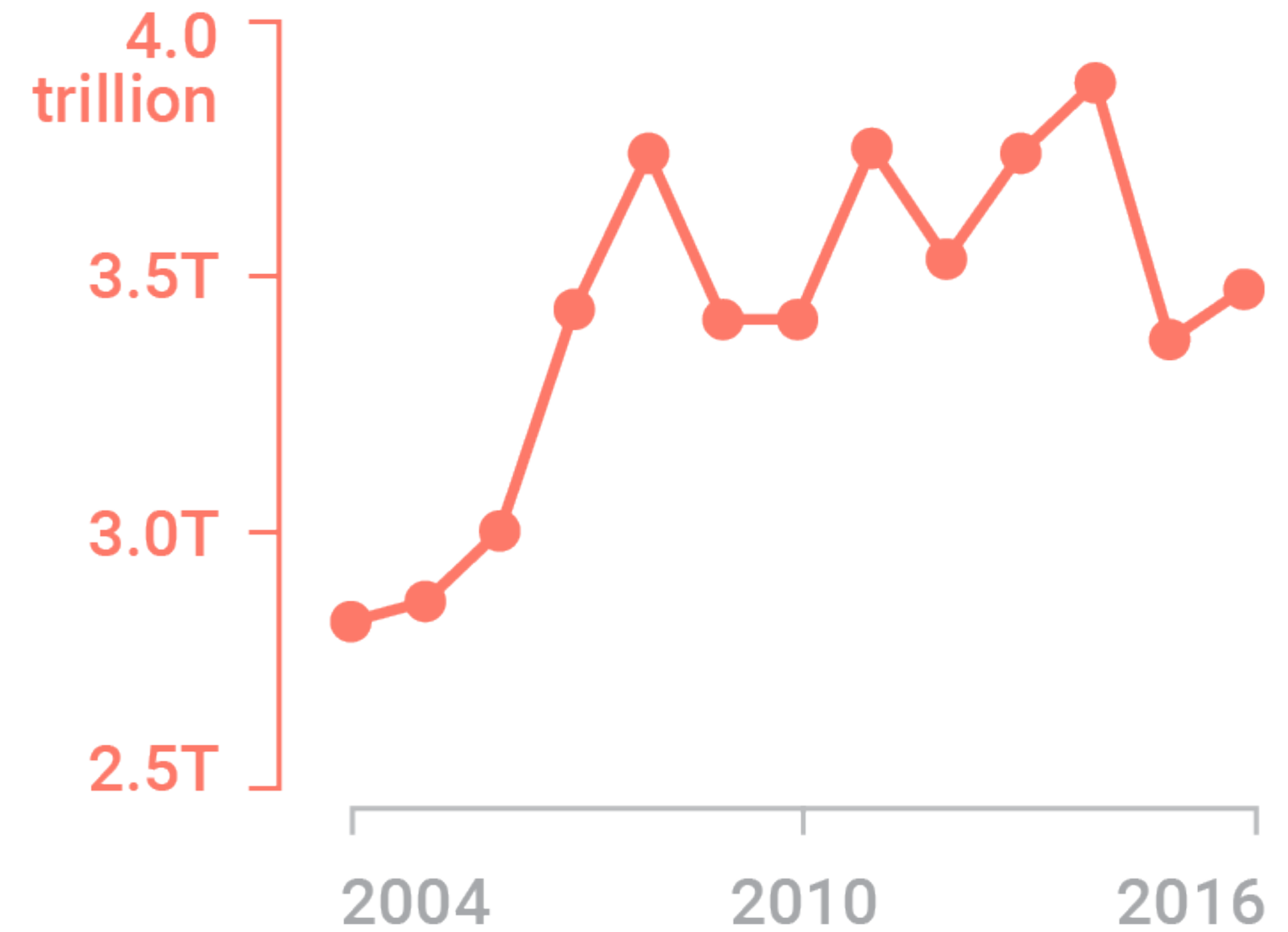
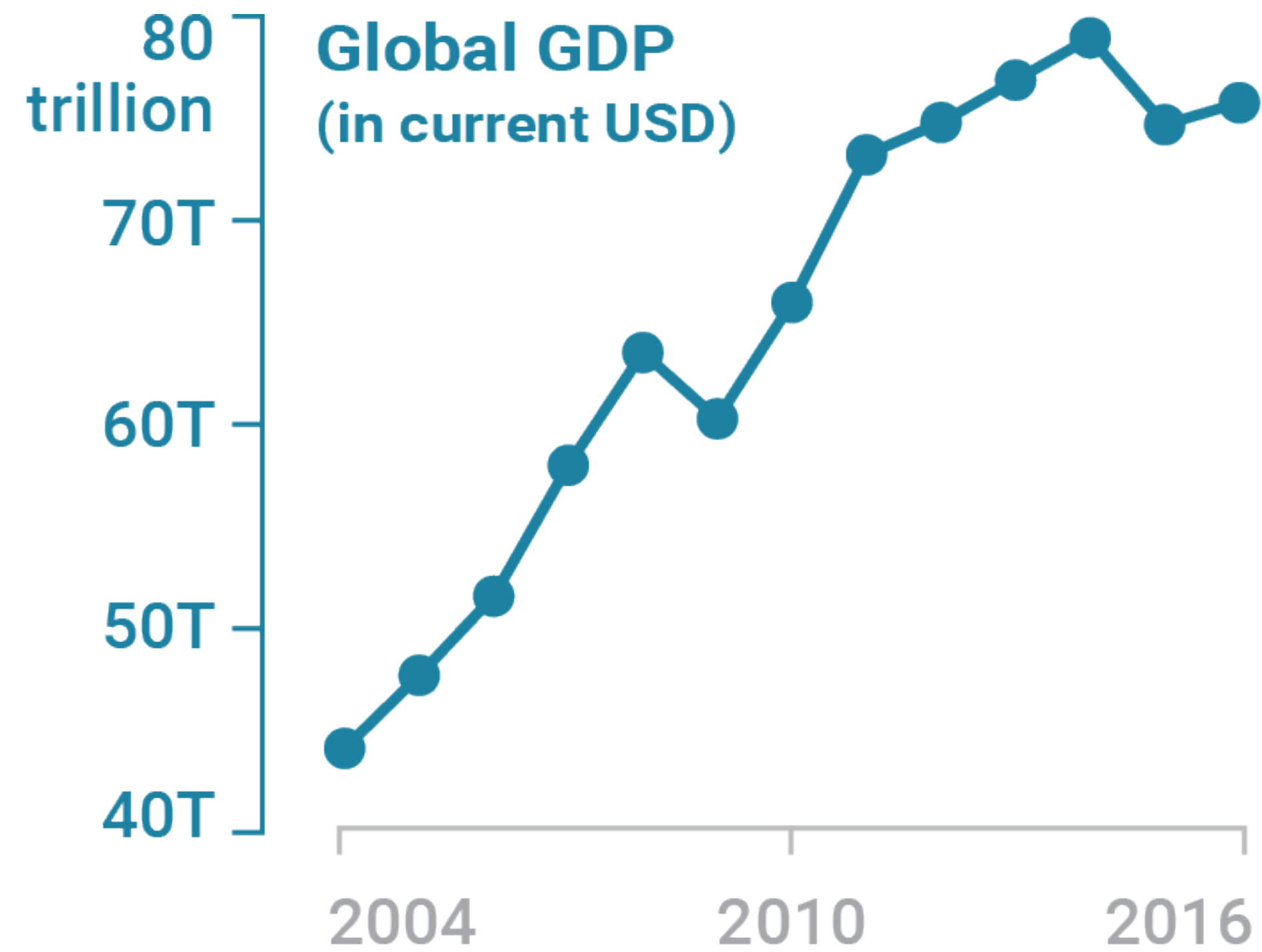


Both started with the
same increase, then Blue
raced to the top.



Both steady.

Dual-Axes Charts



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!