(In)Effective Visual Encoding

DSC 106: Data Visualization

Sam Lau

UC San Diego

Announcements

Lab 2 out, due this Friday.

Project 1 due next Tuesday.

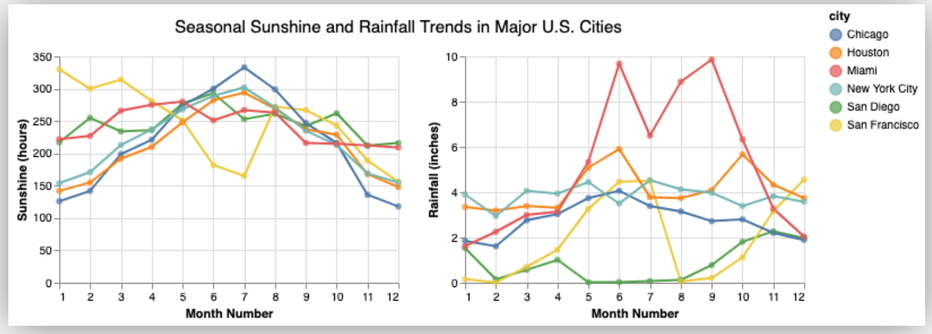
Sam's OH are in-person on Friday 3-4pm (but will change next week)

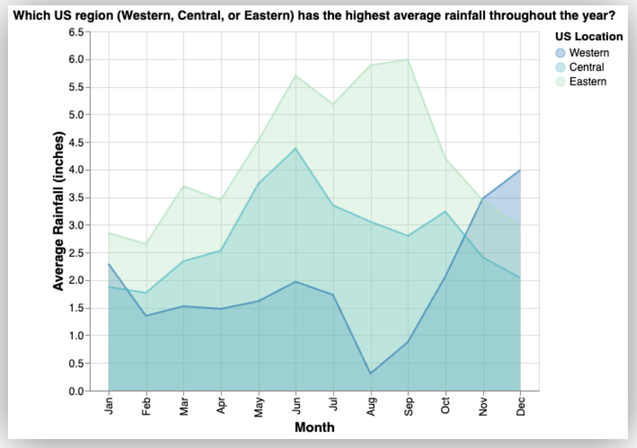
FAQs:

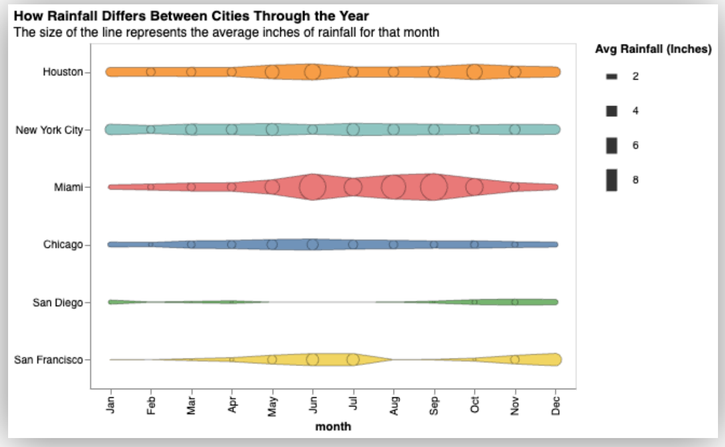
1. What if I joined the class late? We'll provide extensions for assignments, you'll use participation drops for weeks missed.
2. What should I submit for the Lab videos? One mp4 file, max 2 minutes. See the Lab 2 page for guidelines.

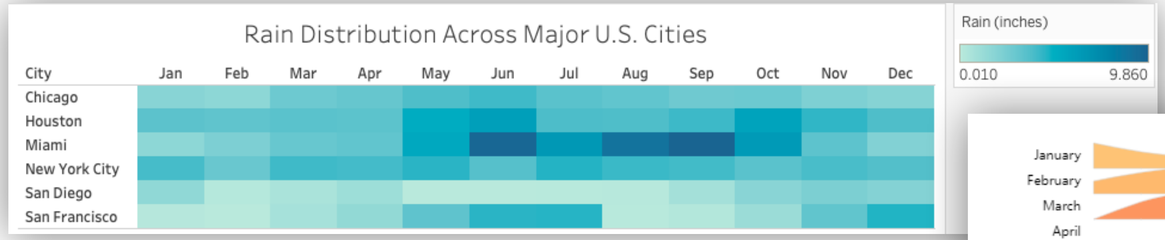
How to ace Project 1

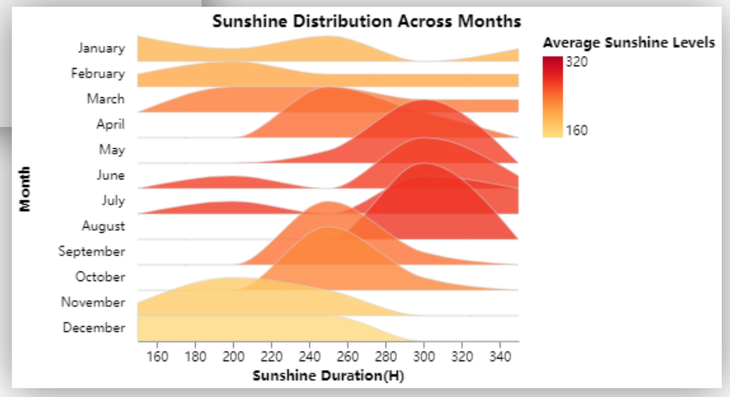
Try out a LOT of possible visualizations!

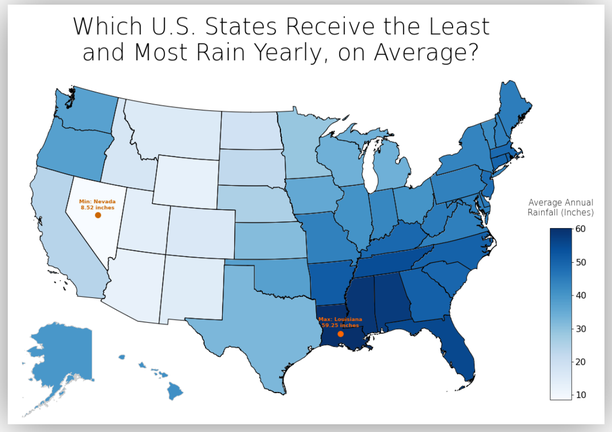










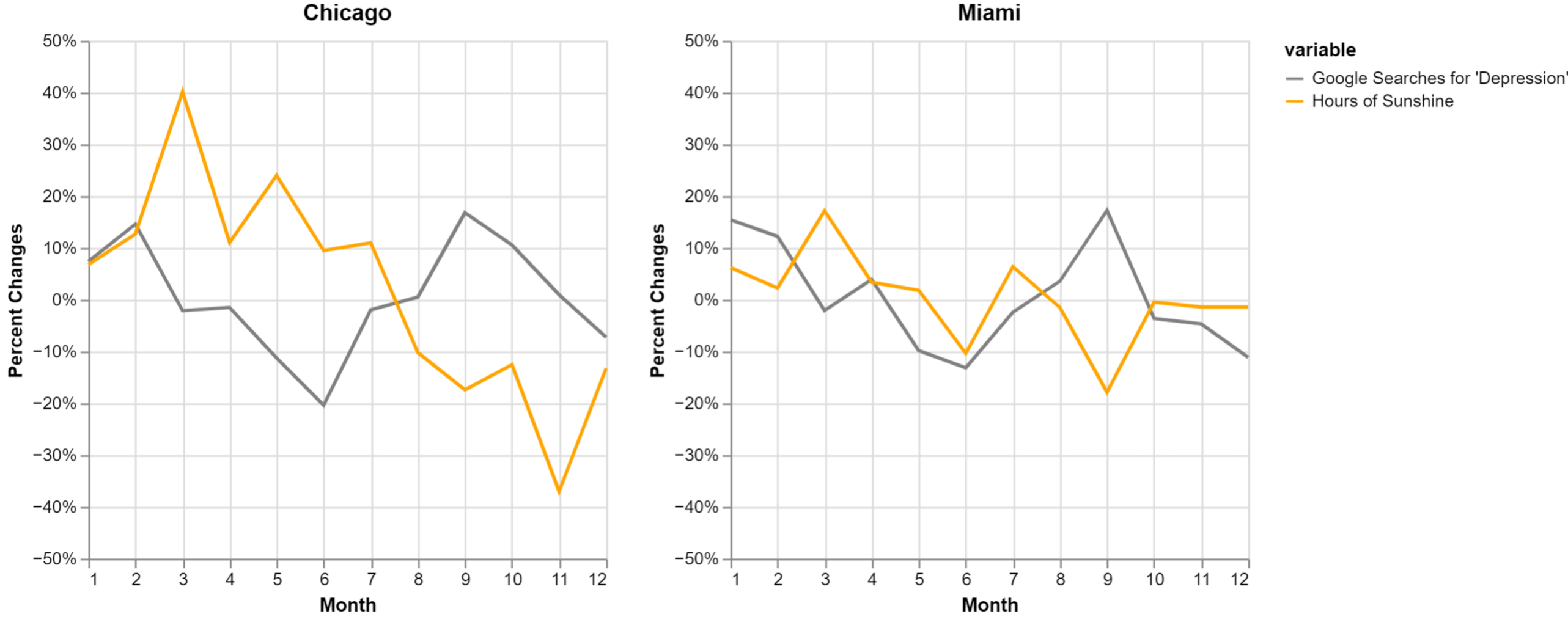


Dow, Steven P., et al. "Parallel prototyping leads to better design results, more divergence, and increased self-efficacy." ACM Transactions on Computer-Human Interaction (TOCHI) 17.4 (2010): 1-24.

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 flucates much more in Chicago.

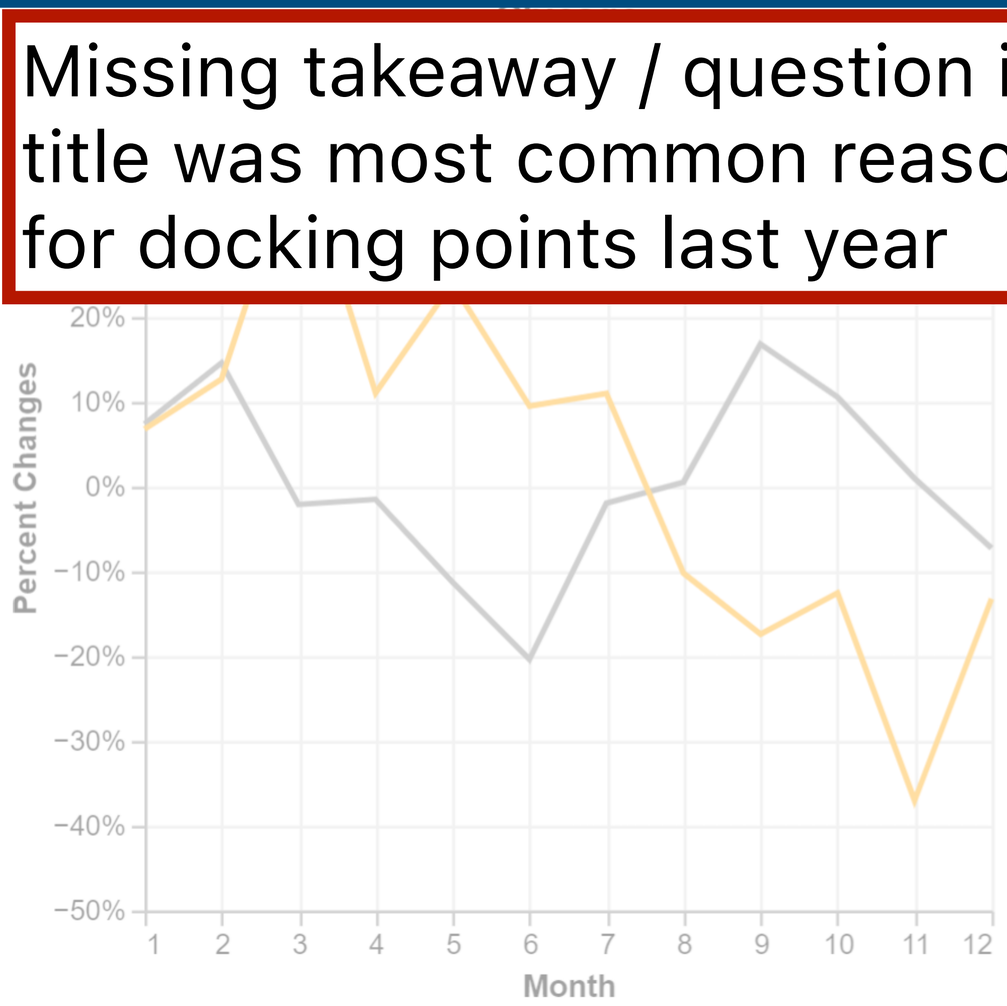


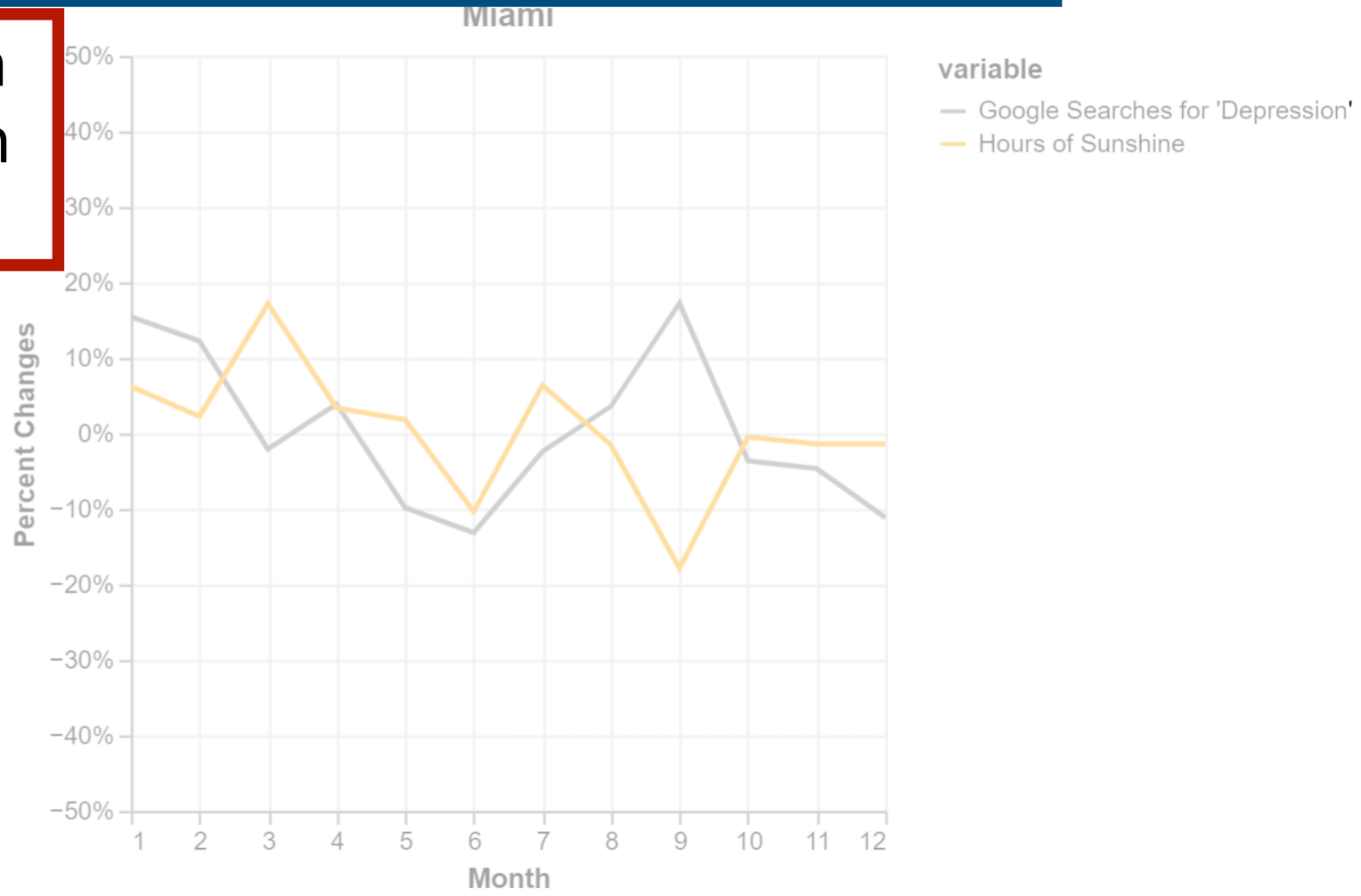
Then, tell a story

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 flucates much more in Chicago.

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

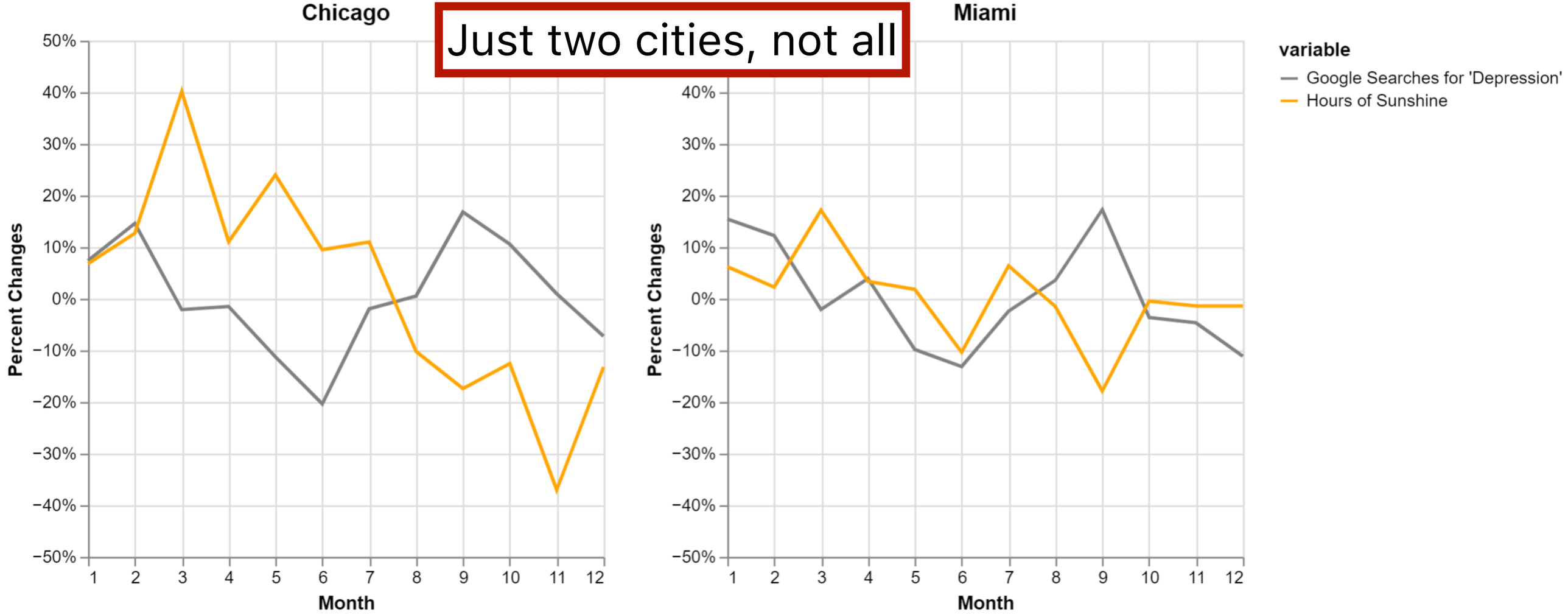




Then, tell a story

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

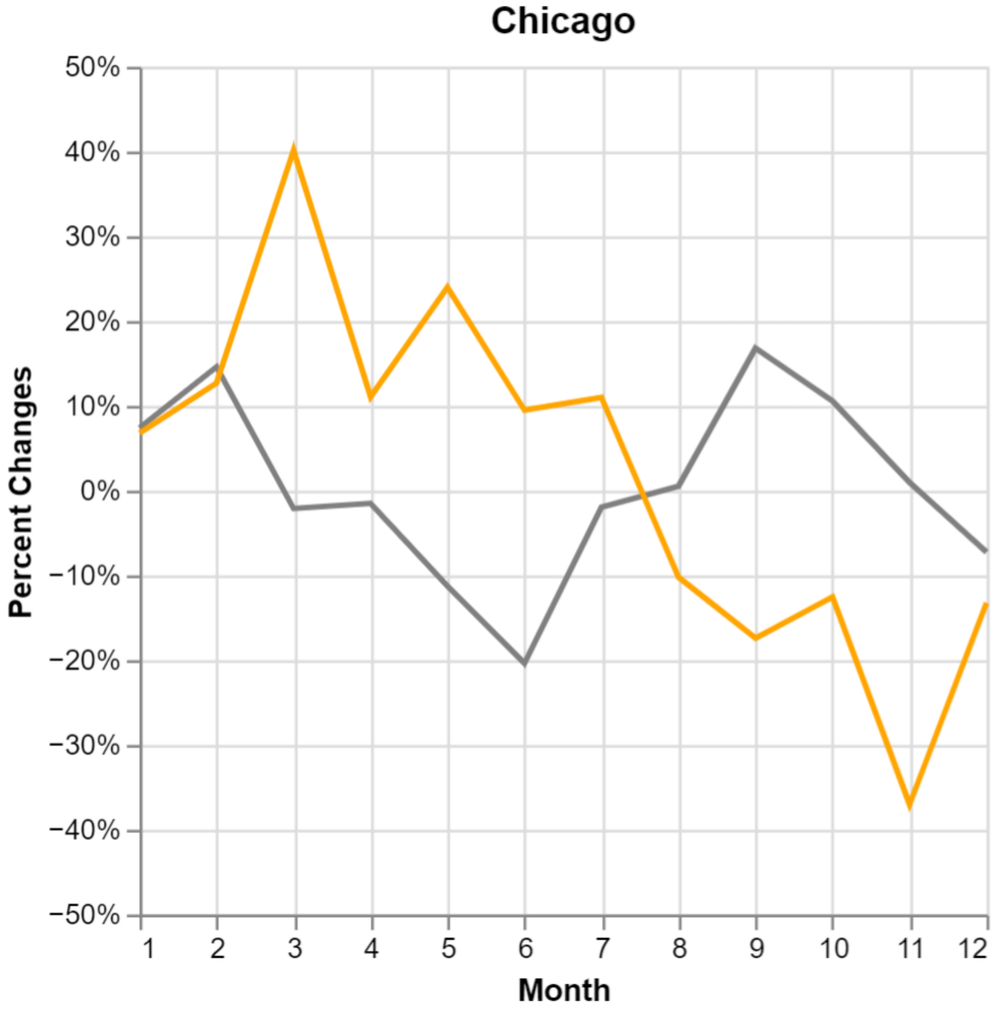
Compared to a city like Miami, where sunshine hardly changes year round, Google searches for depression flucates much more in Chicago.

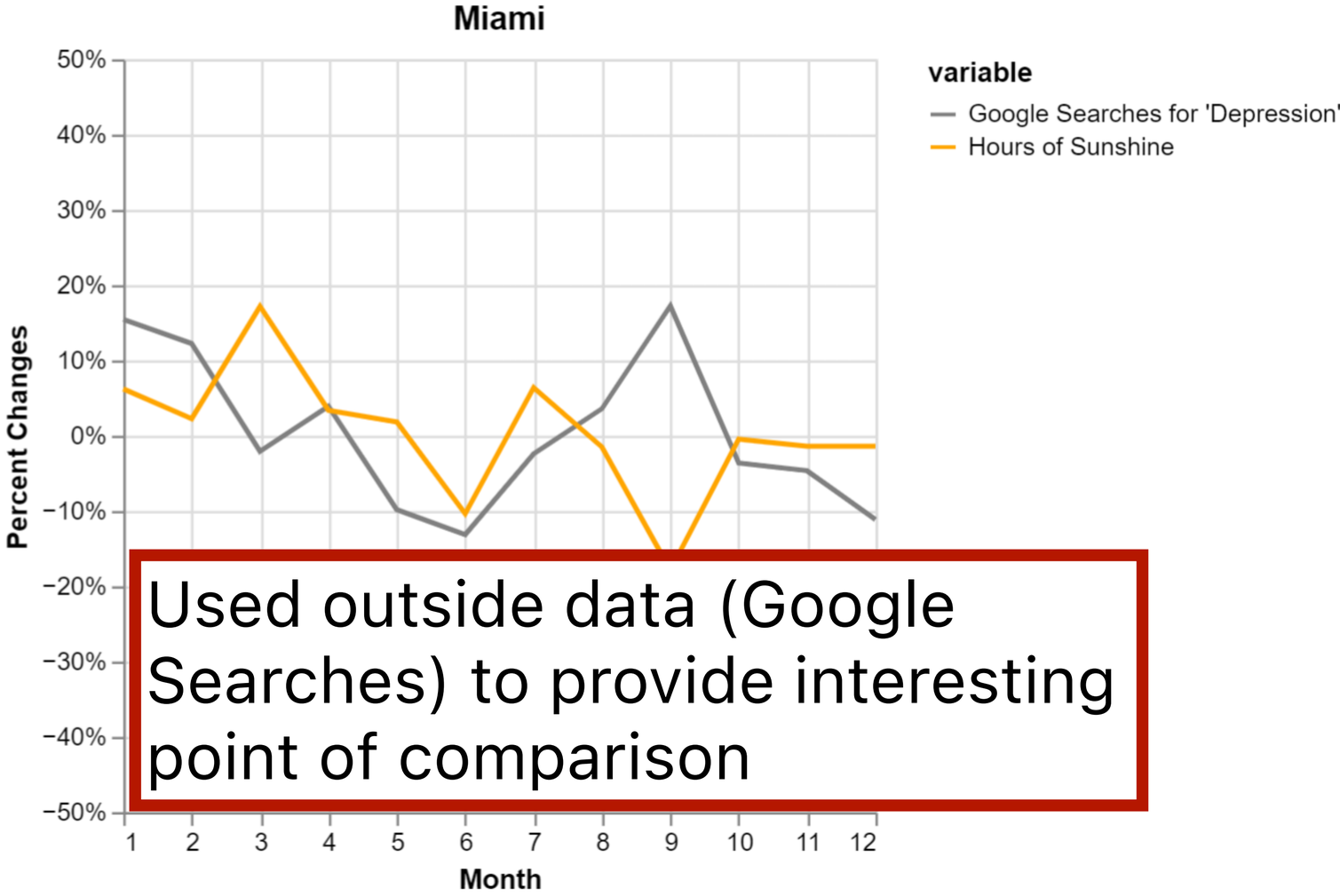


Then, tell a story

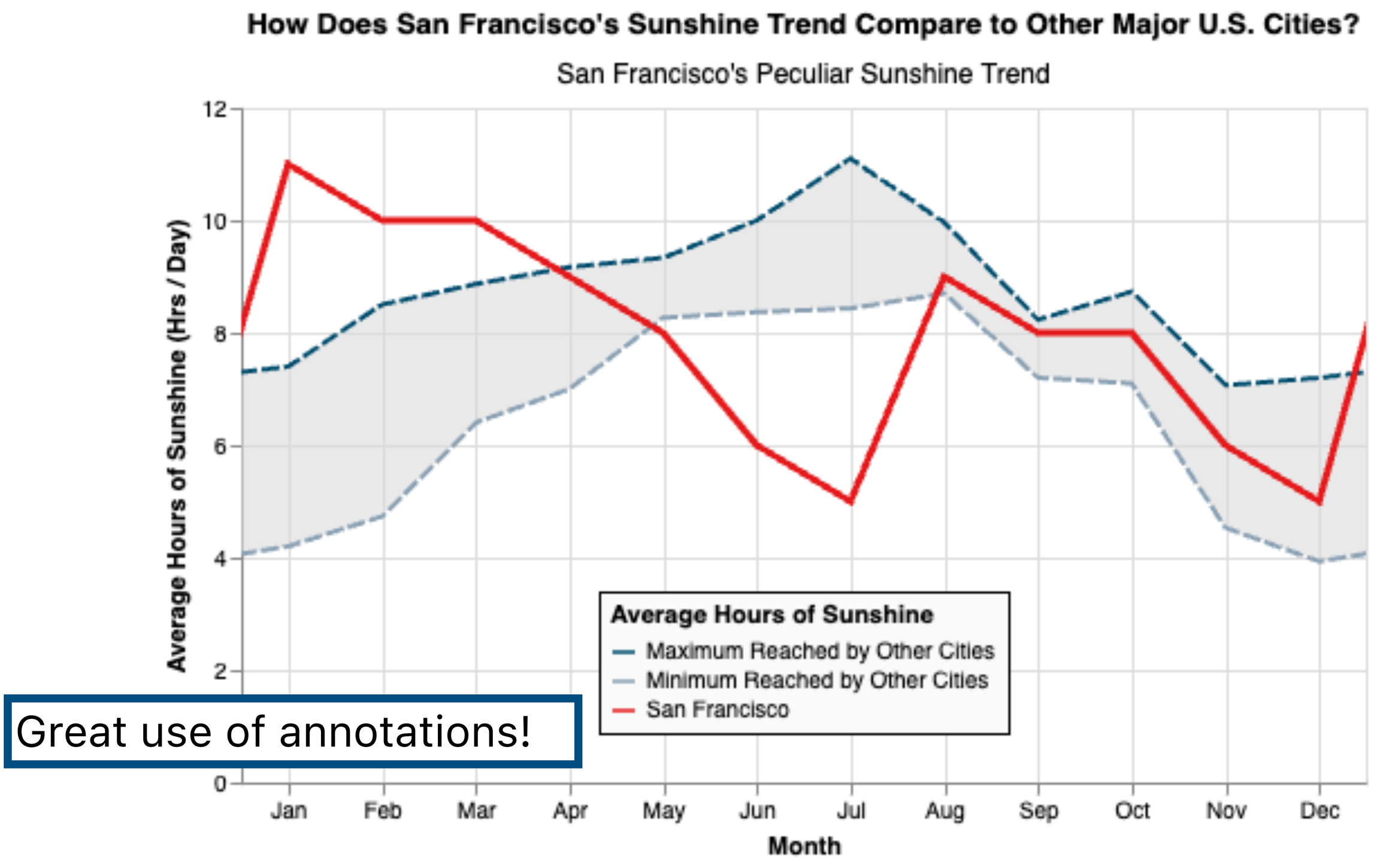
A Tale of Two Cities: How Dramatic are the Effects of Seasonal Depression?

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





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

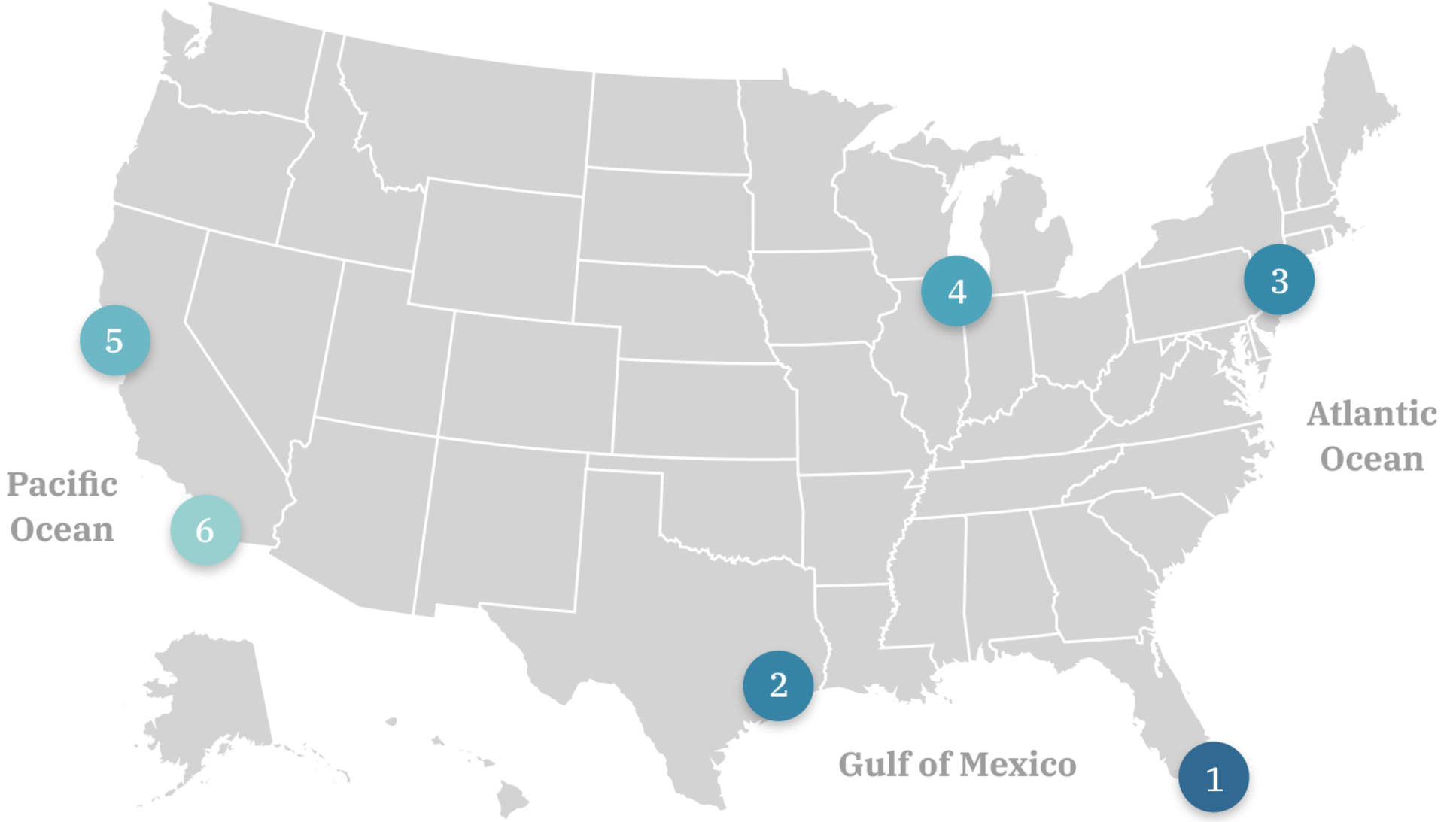


Great use of annotations!

More Project 1 examples (won't cover in lecture, but here as a resource for you)

Most of these examples are imperfect, but I included because of their creativity.

How Does Location Impact the Average Annual Rainfall of Major U.S. Cities?



Rainfall

1. Miami — 61.90in

2. Houston — 49.58in

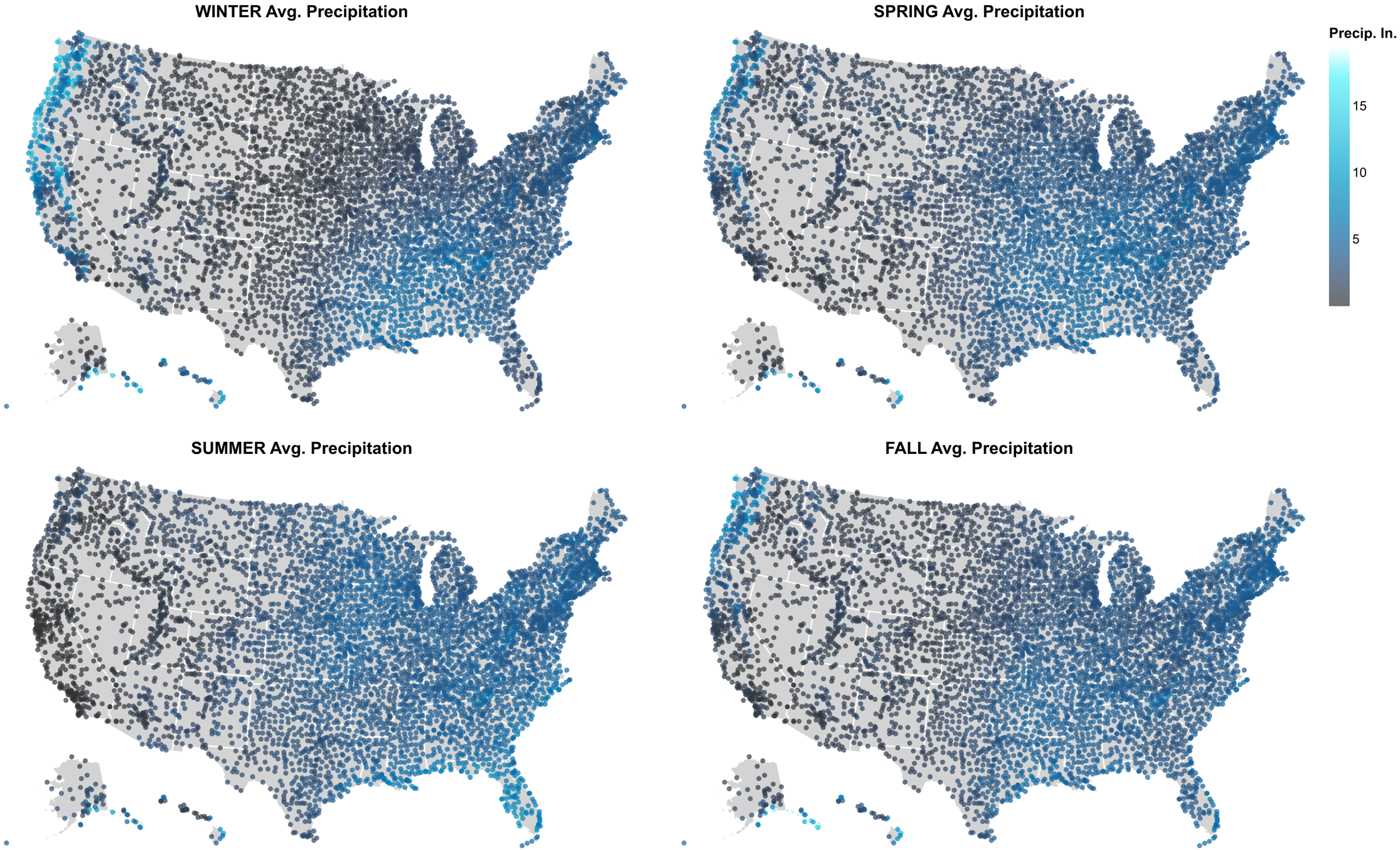
3. New York City — 46.23in

4. Chicago — 33.20in

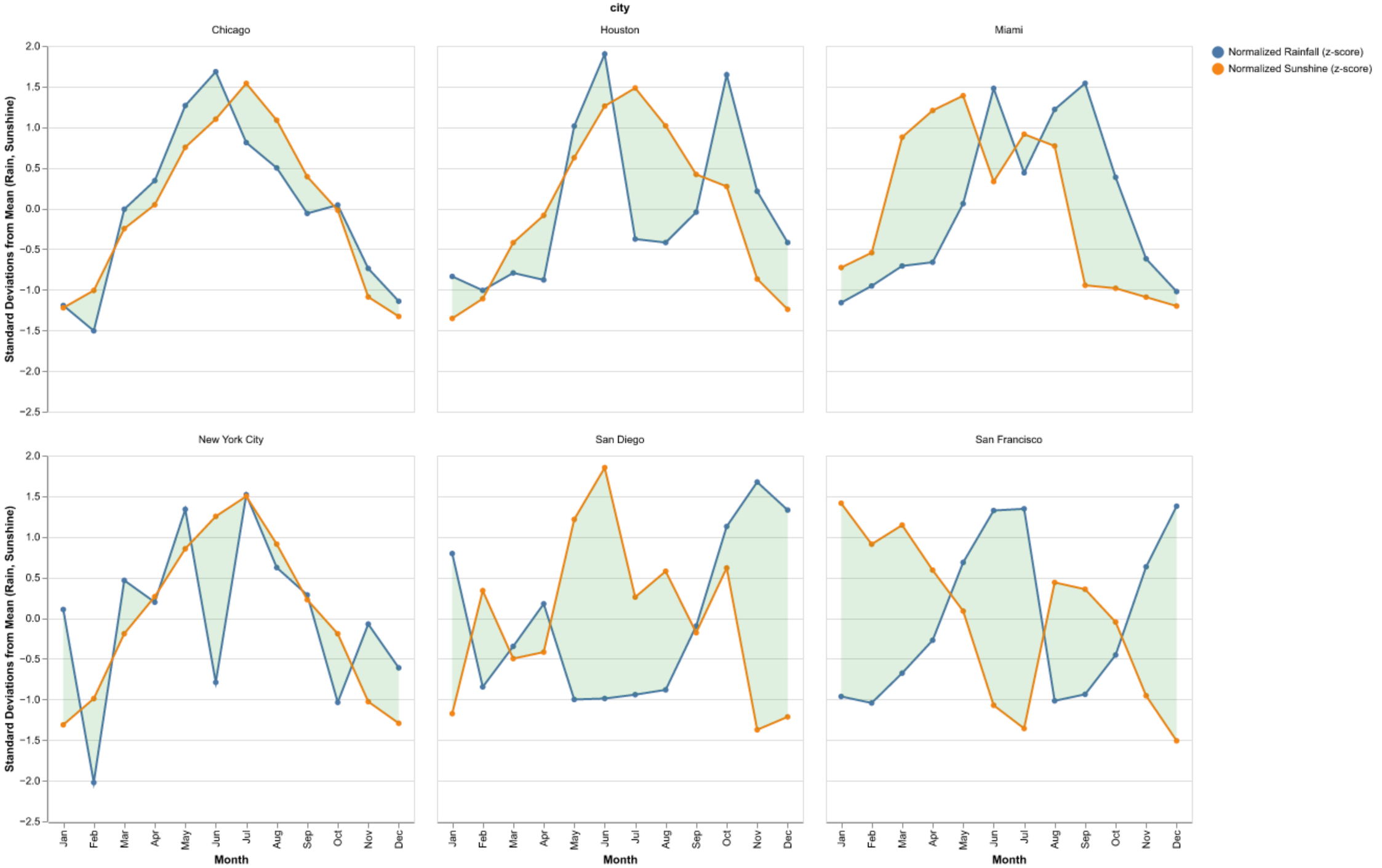
5. San Francisco — 23.66in

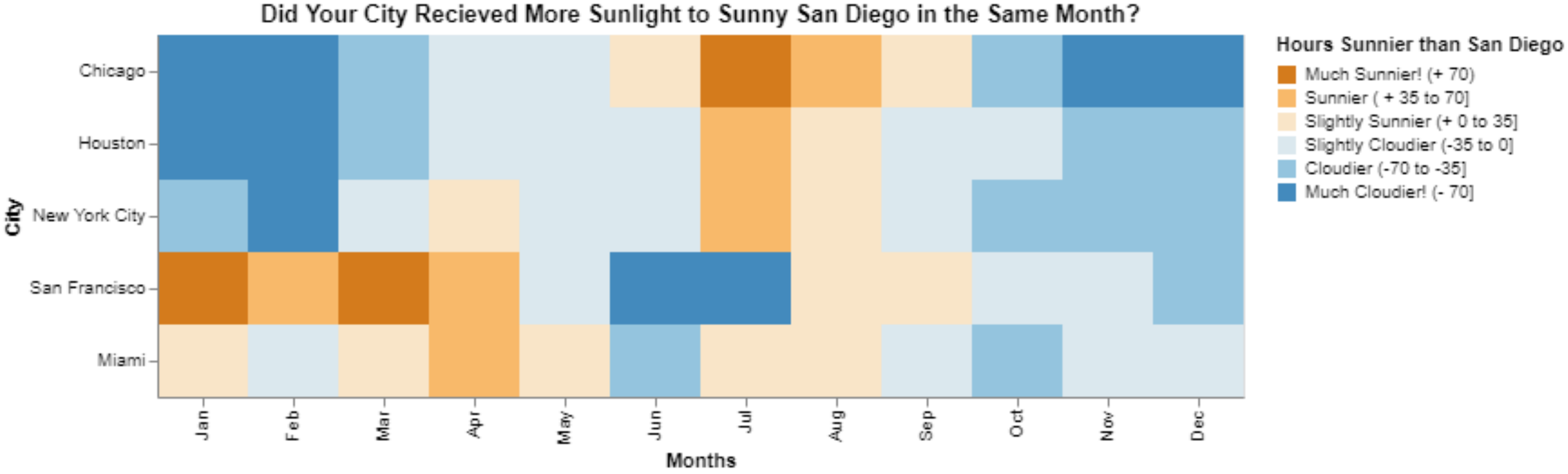
6. San Diego — 10.34in

"My Husband Hates Rain. I Love Rain and Hate My Husband. Where's the Best Place to Live in the US?"

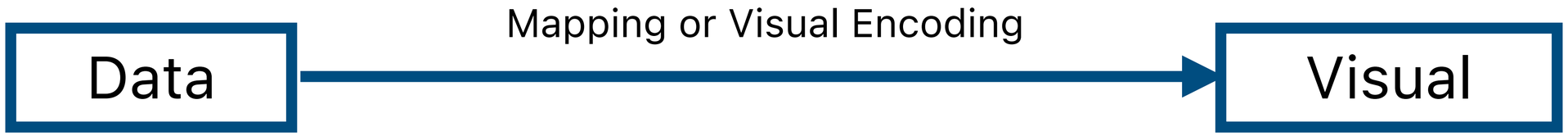


What is the relationship between sunshine and rainfall by month for each city?

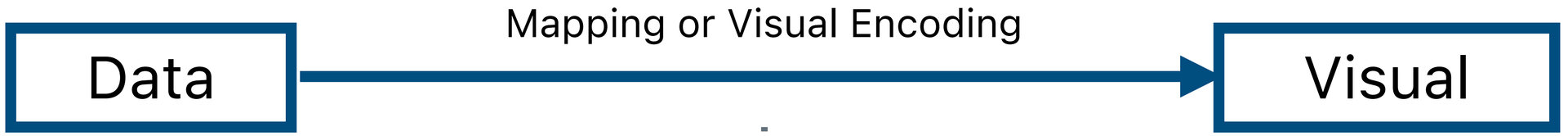



Did Your City Recieved More Sunlight to Sunny San Diego in the Same Month?
City
Chicago
Houston
New York City
San Francisco
Miami
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months
Hours Sunnier than San Diego
Much Sunnier! (+ 70)
Sunnier (+ 35 to 70]
Slightly Sunnier (+ 0 to 35]
Slightly Cloudier (-35 to 0]
Cloudier (-70 to -35]
Much Cloudier! (- 70]

Visual Encodings



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, ...

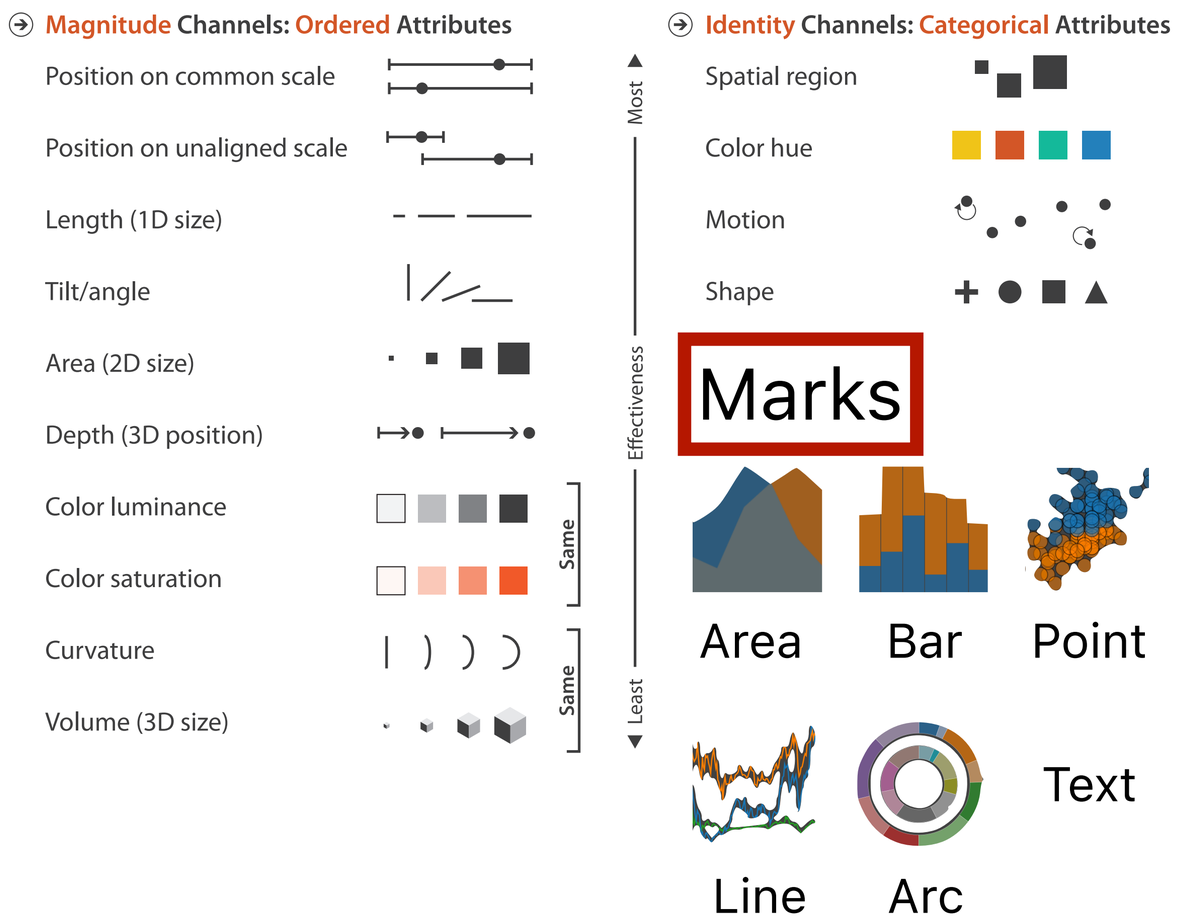
Ordinal
=, ≠, **<, >**
Ordered.
E.g., Quality of eggs: Grade AA, A, B

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

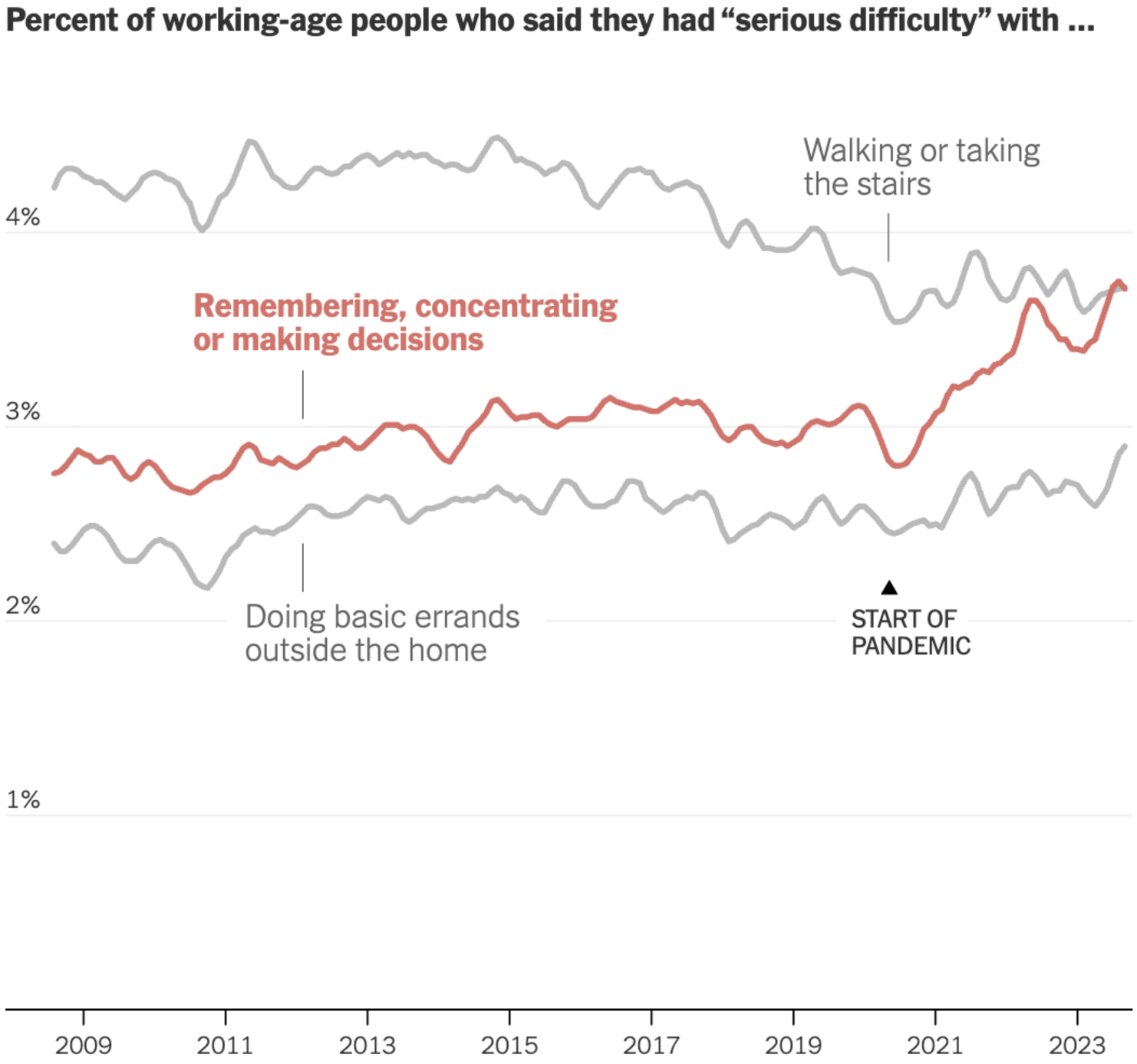
Channels: Expressiveness Types and Effectiveness Ranks



Marks

Name that ~~chart!~~

Visual Encoding!

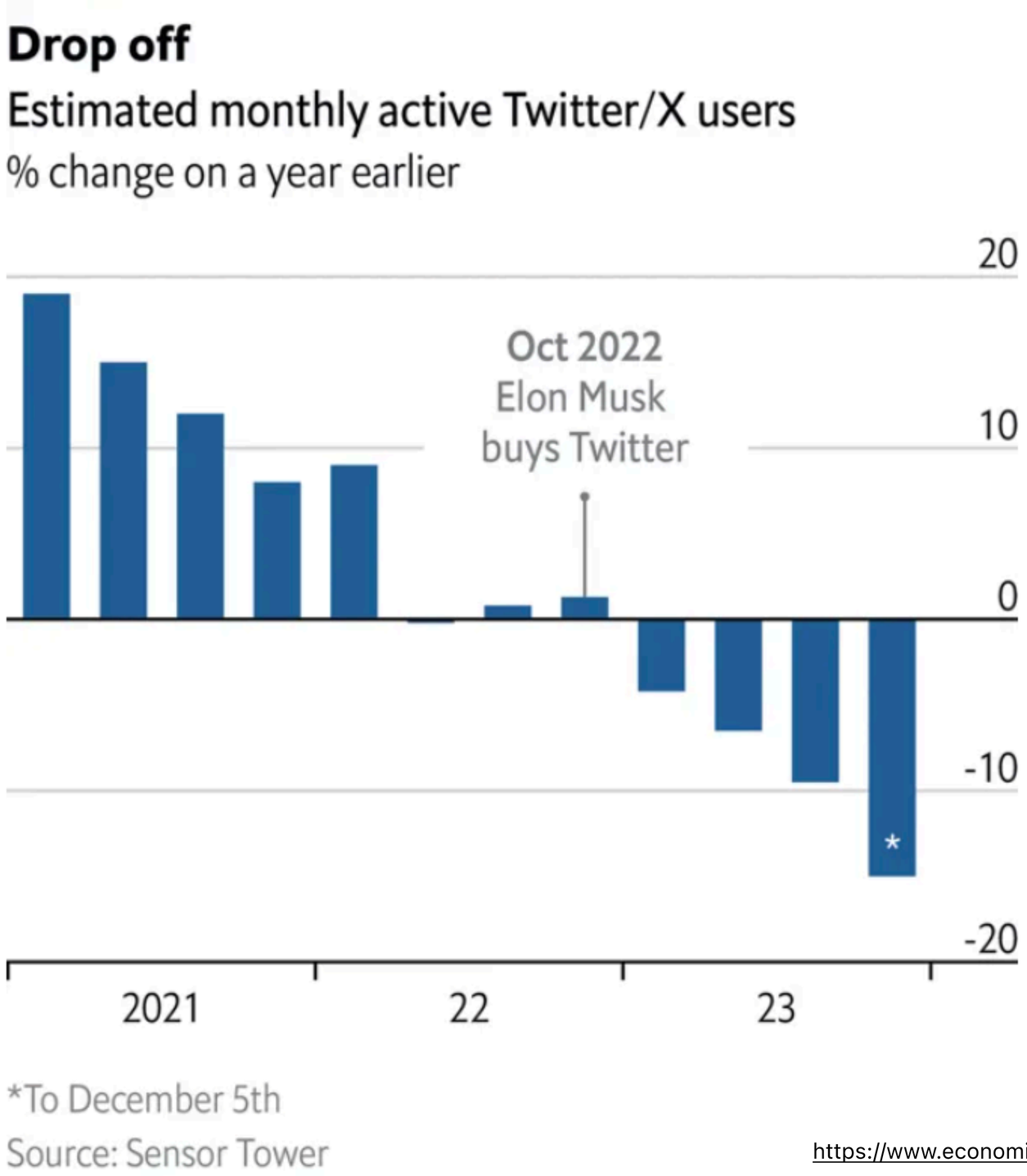


Mark: `line`

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

What about color?

https://www.nytimes.com/2024/01/10/upshot/covid-pandemic-wave.html



Mark: `bar`

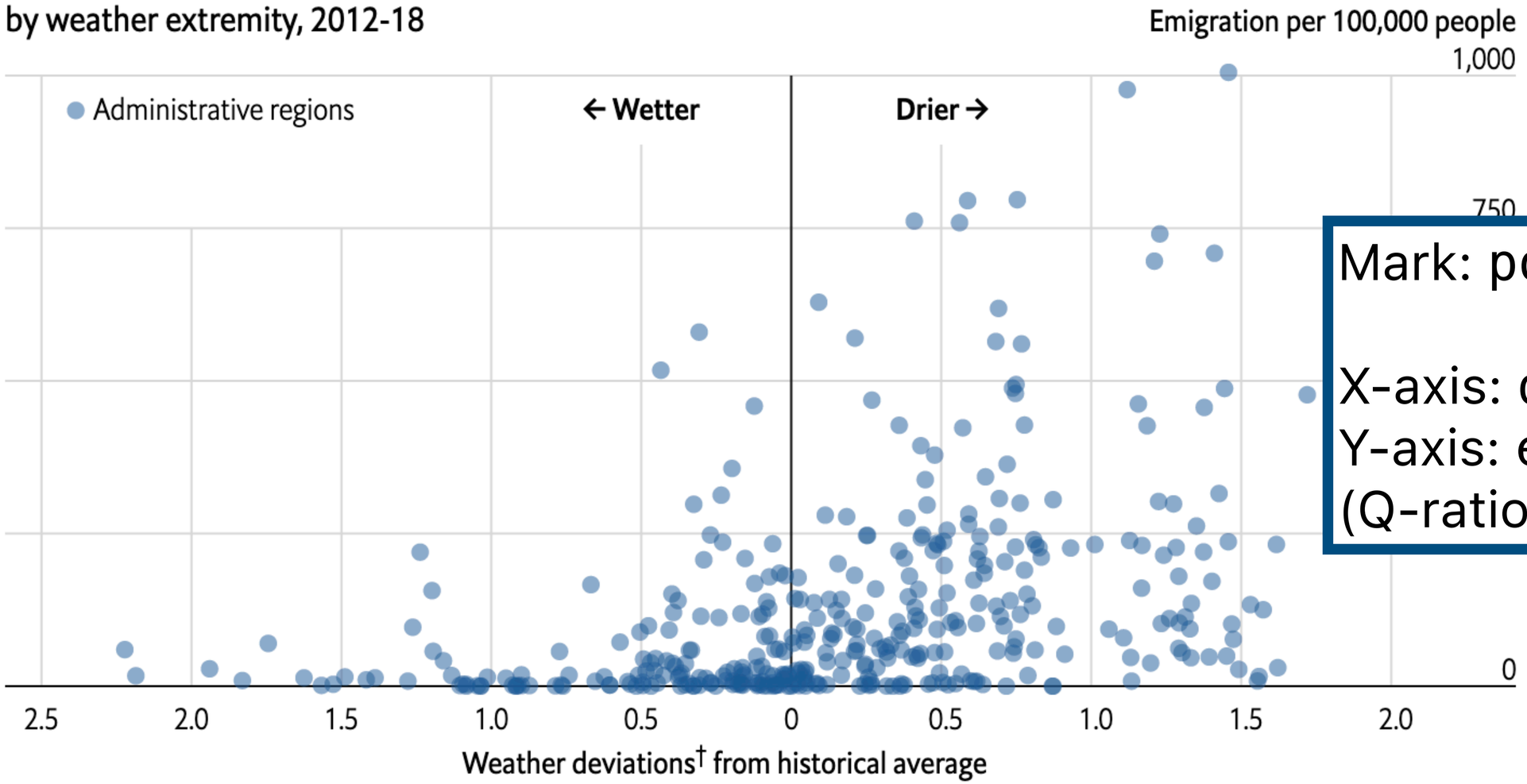
X-axis: `date` (Q-interval)
Y-axis: `percent` (Q-ratio)

Same encodings as line plot, just different mark!

https://www.economist.com/graphic-detail/2023/12/20/has-twitter-now-x-become-more-right-wing

Spotting a trend

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



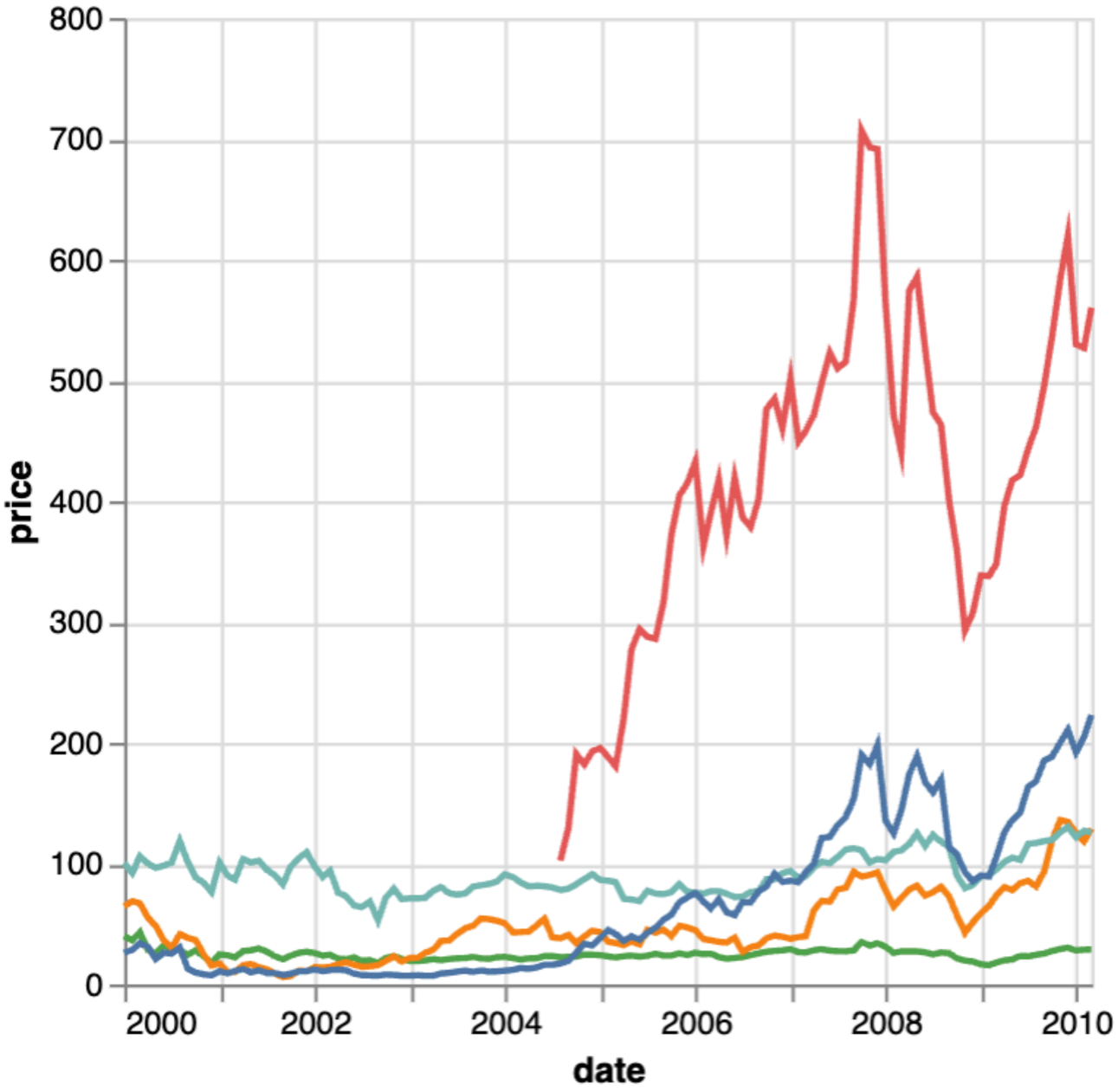
Mark: `point`

X-axis: `dev` (Q-ratio)
Y-axis: `emigration` (Q-ratio)

*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

https://www.economist.com/graphic-detail/2023/11/17/why-central-americans-migrate-to-the-united-states-when-they-do https://www.nature.com/articles/s41598-023-43668-9

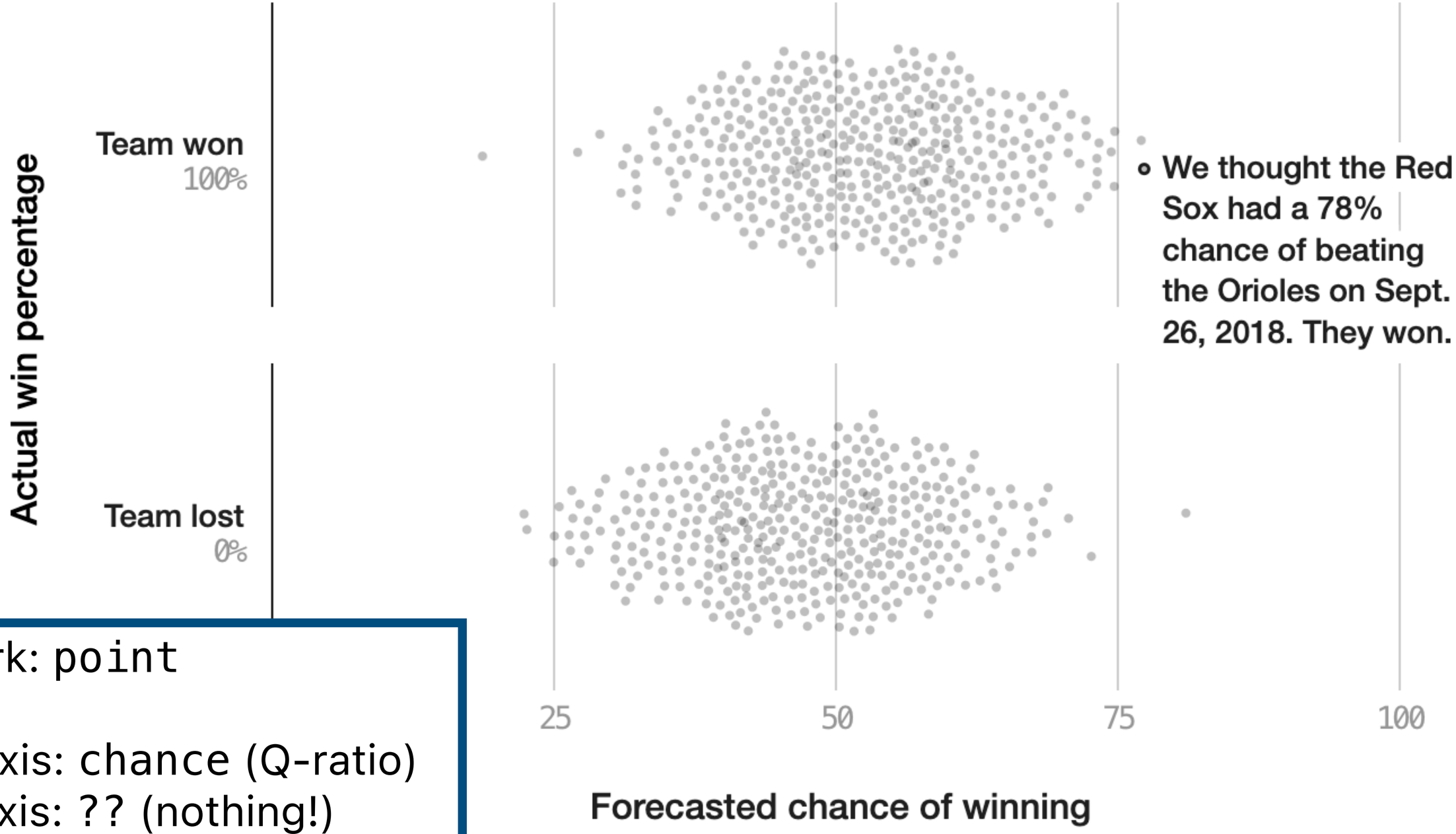


Mark: `line`

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

Notice the parallel with plotly express syntax!

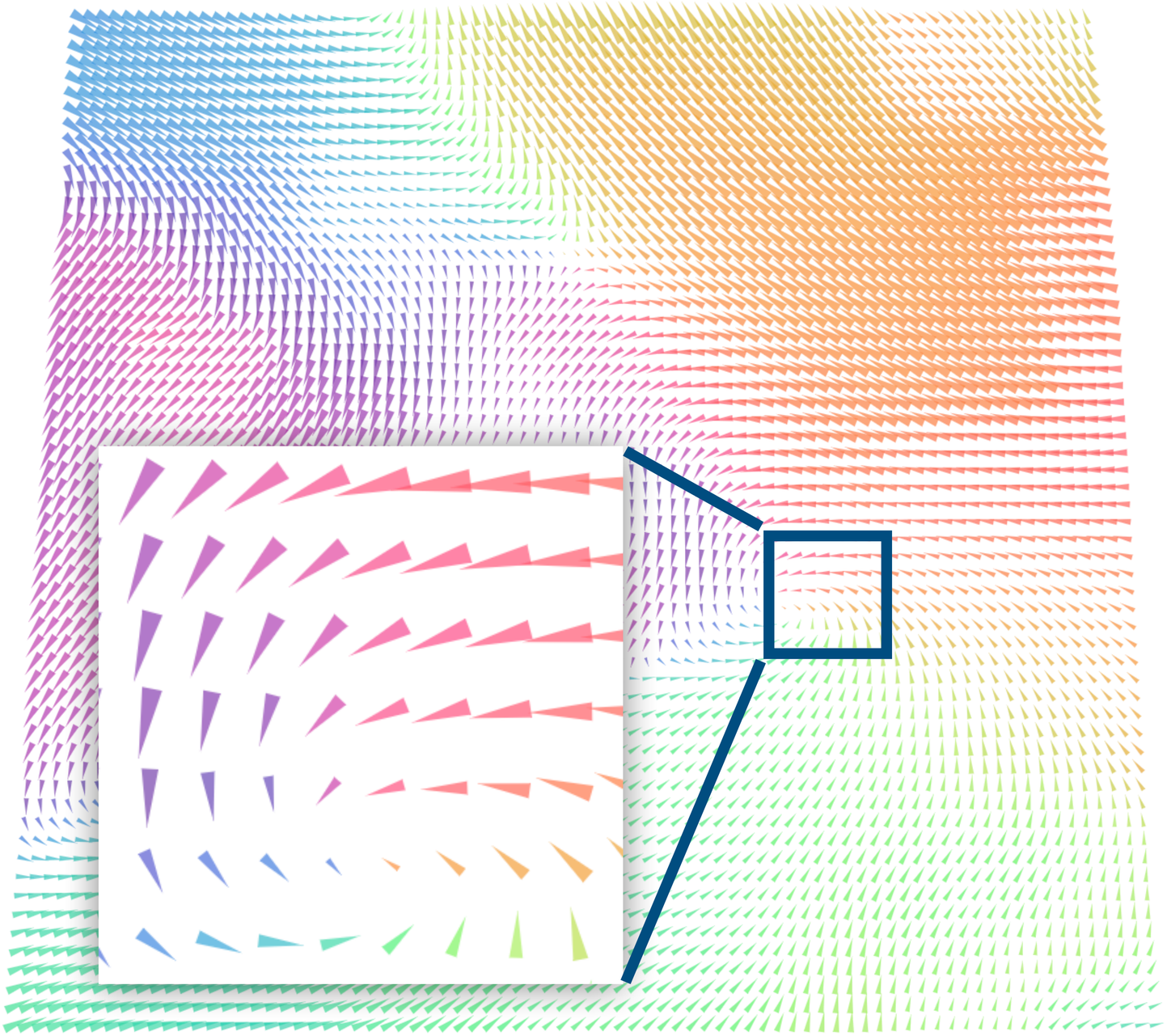
```
px.line(
    stocks_df,
    x='date',
    y='price',
    color='symbol',
)
```



Mark: `point`

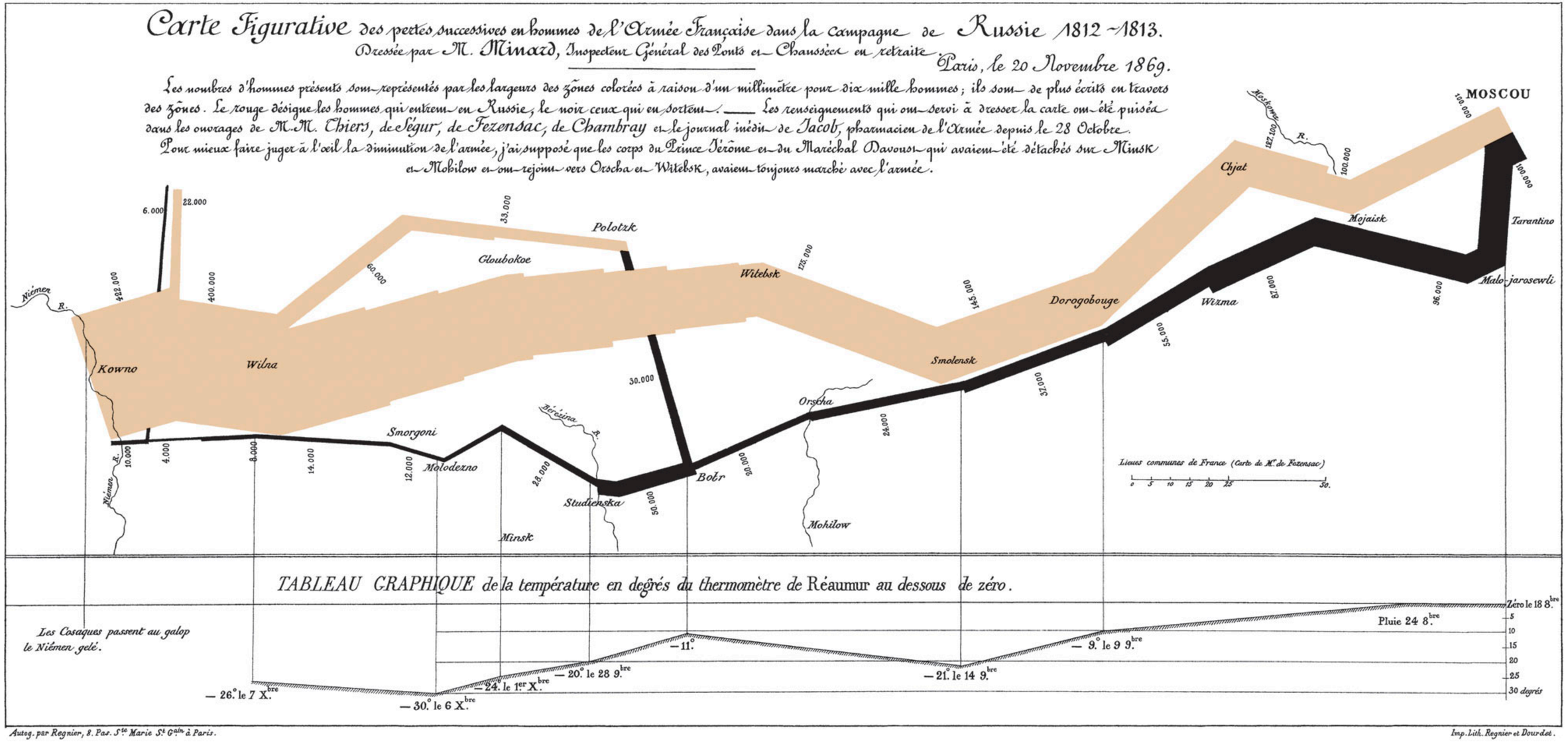
X-axis: `chance` (Q-ratio)
Y-axis: ?? (nothing!)

https://projects.fivethirtyeight.com/checking-our-work/



tryclassbuzz.com:
wind

https://vega.github.io/vega-lite/examples/point_angle_windvector.html



tryclassbuzz.com:
minard

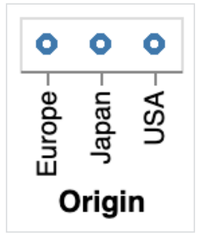
[Charles Minard, 1869]

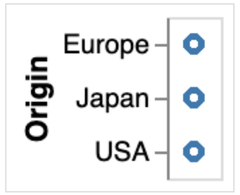
A Design Space of Visual Encodings

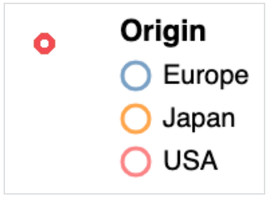
Visual Encoding = Combinatorial Design Space

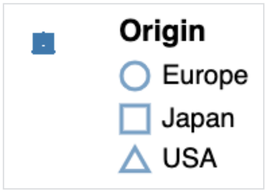
1D nominal data (N, O)

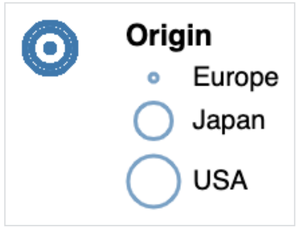
raw



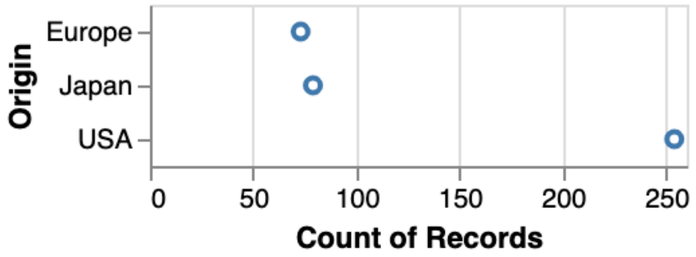


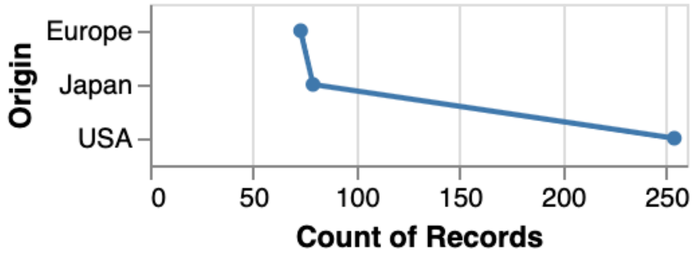


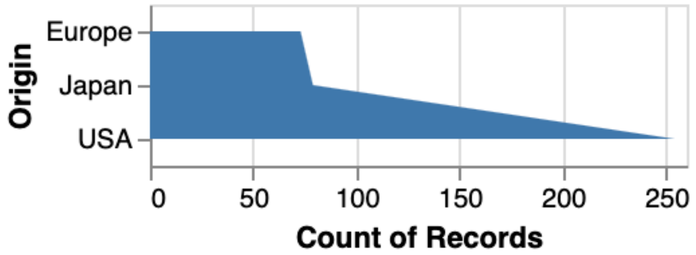


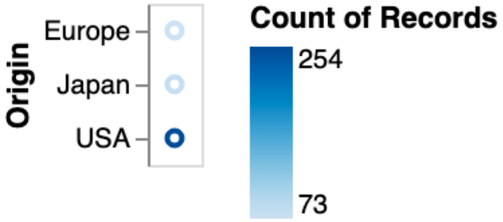


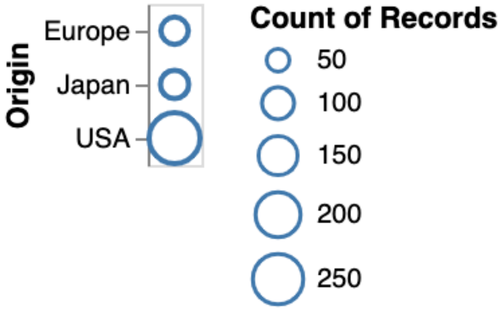
aggregate (count)

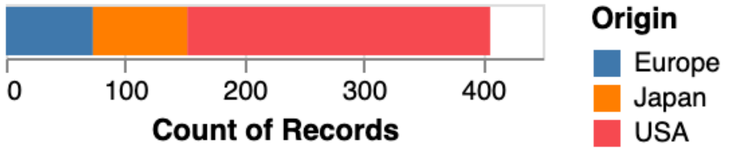


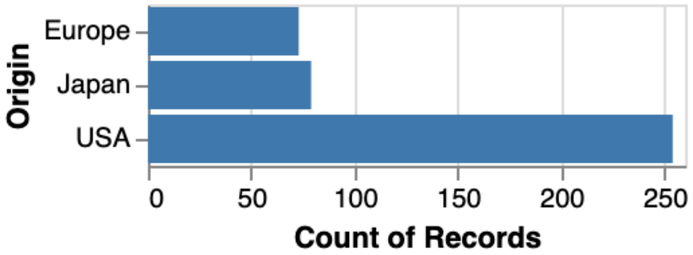










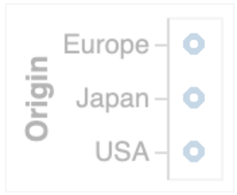


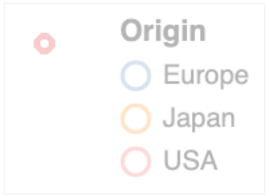
Visual Encoding = Combinatorial Design Space

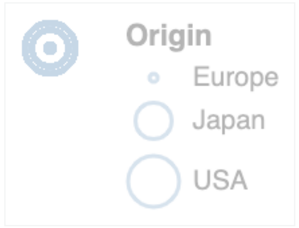
1D nominal data (N, O)

Expressive?

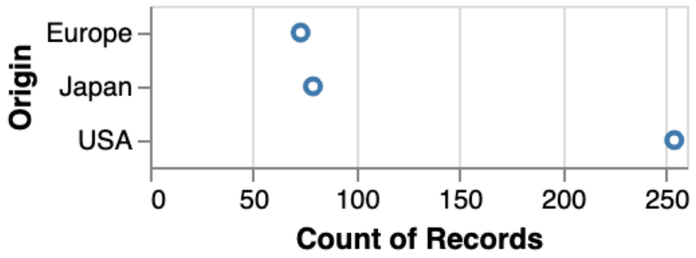
raw

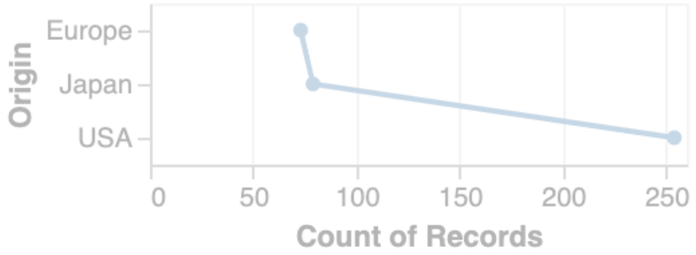


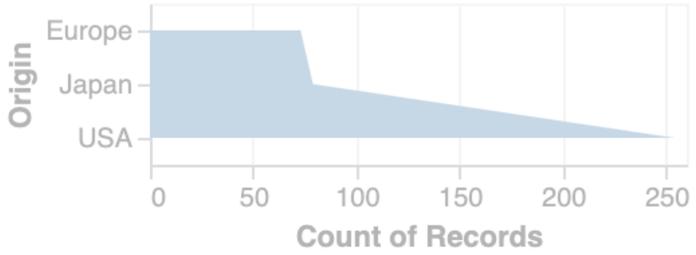


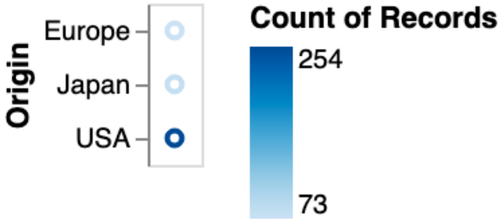


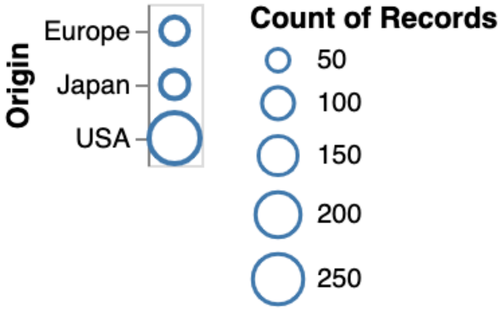
aggregate (count)

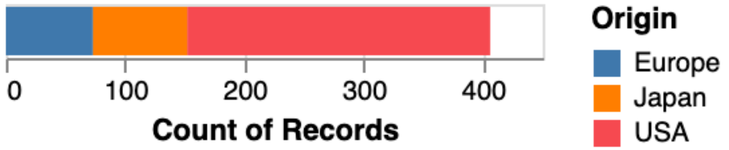


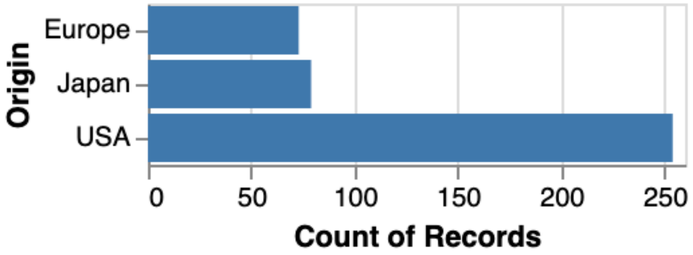








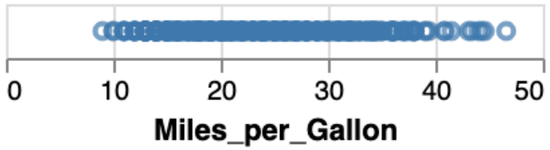


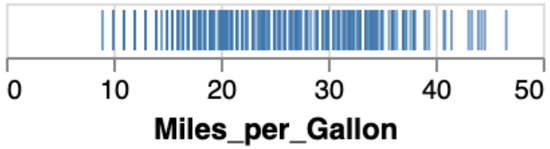


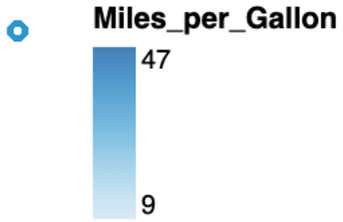
Visual Encoding = Combinatorial Design Space

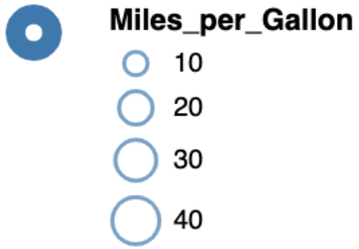
1D quantitative data (Q)

raw

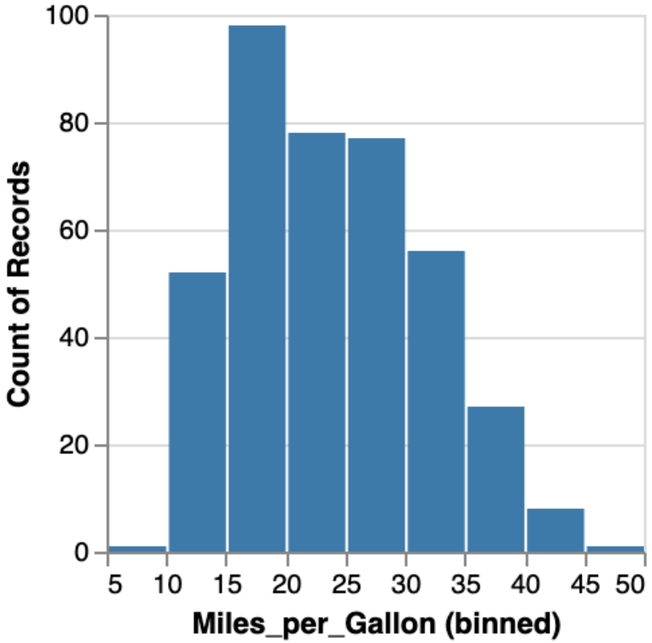


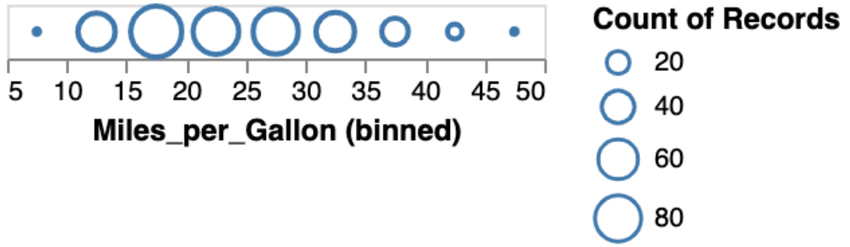


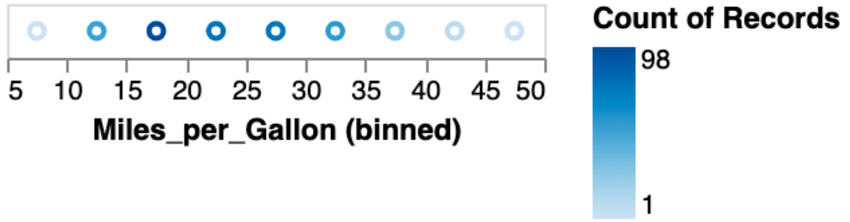




aggregate (count)





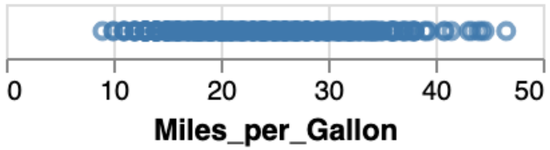


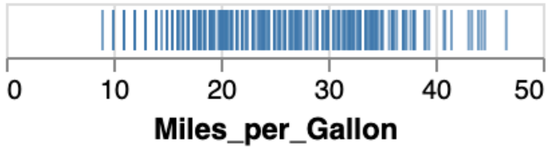
Visual Encoding = Combinatorial Design Space

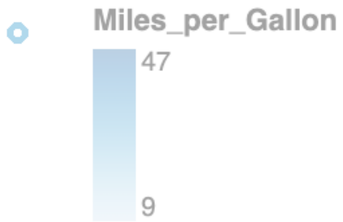
1D quantitative data (Q)

Expressive?

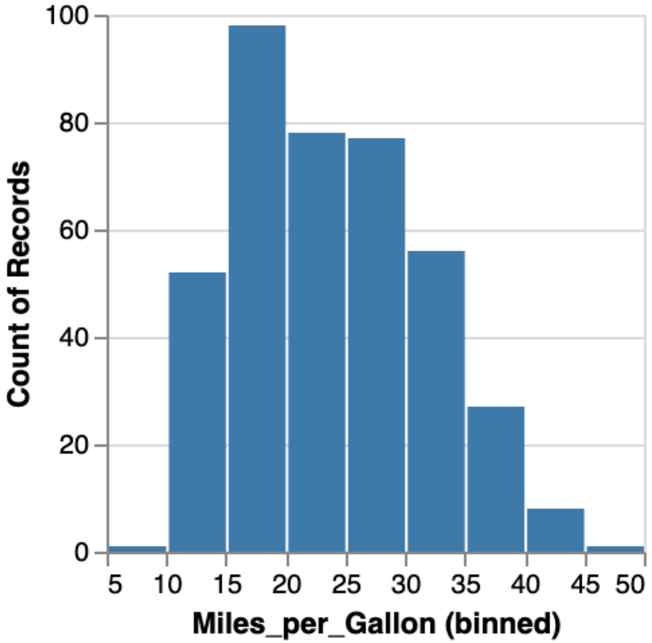
raw

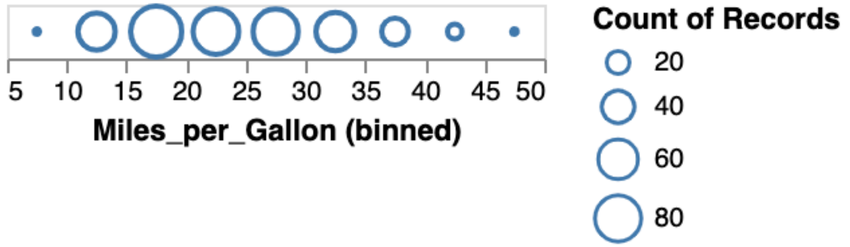


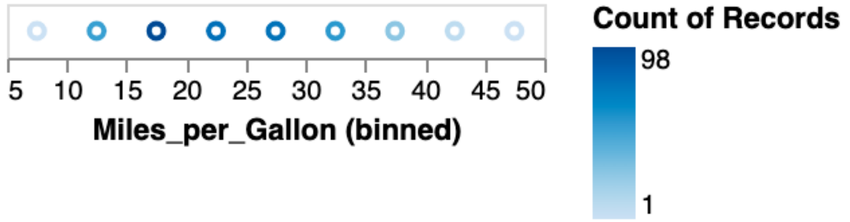




aggregate (count)







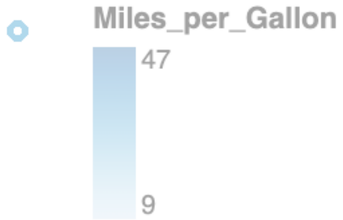
Visual Encoding = Combinatorial Design Space

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

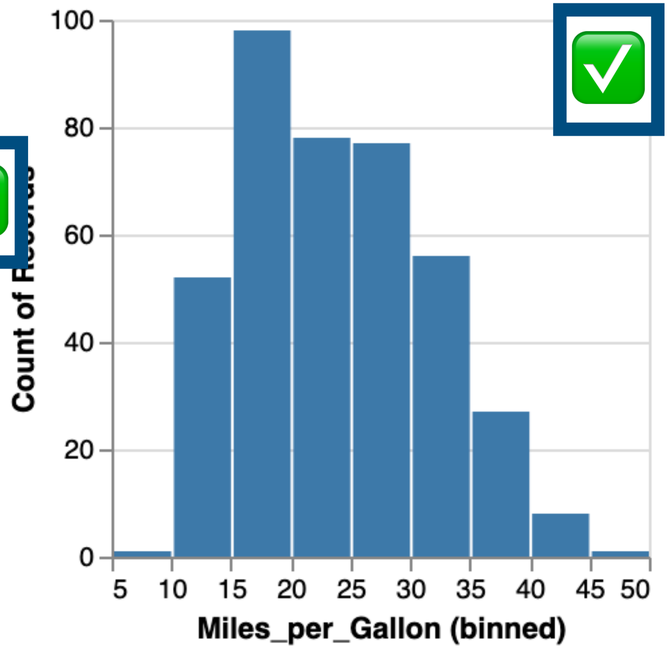
raw

?

✅

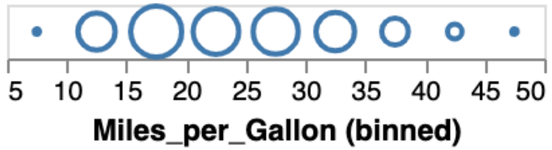


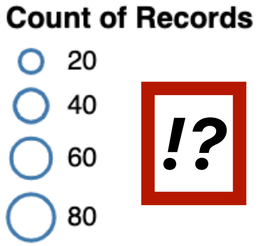
aggregate (count)



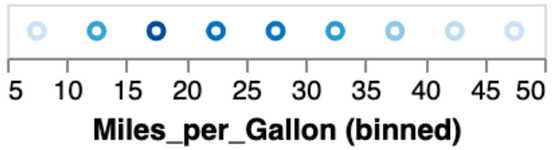
✅

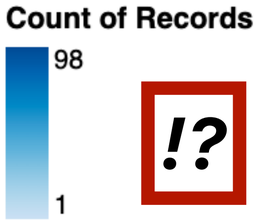
https://vega.github.io/vega/examples/histogram/





!?





!?

Visual Encoding:
Nimble Design Moves

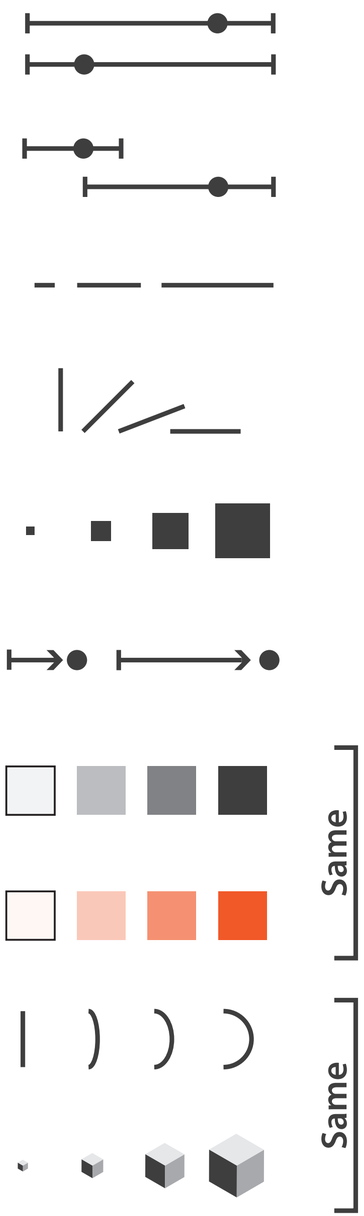
demo

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

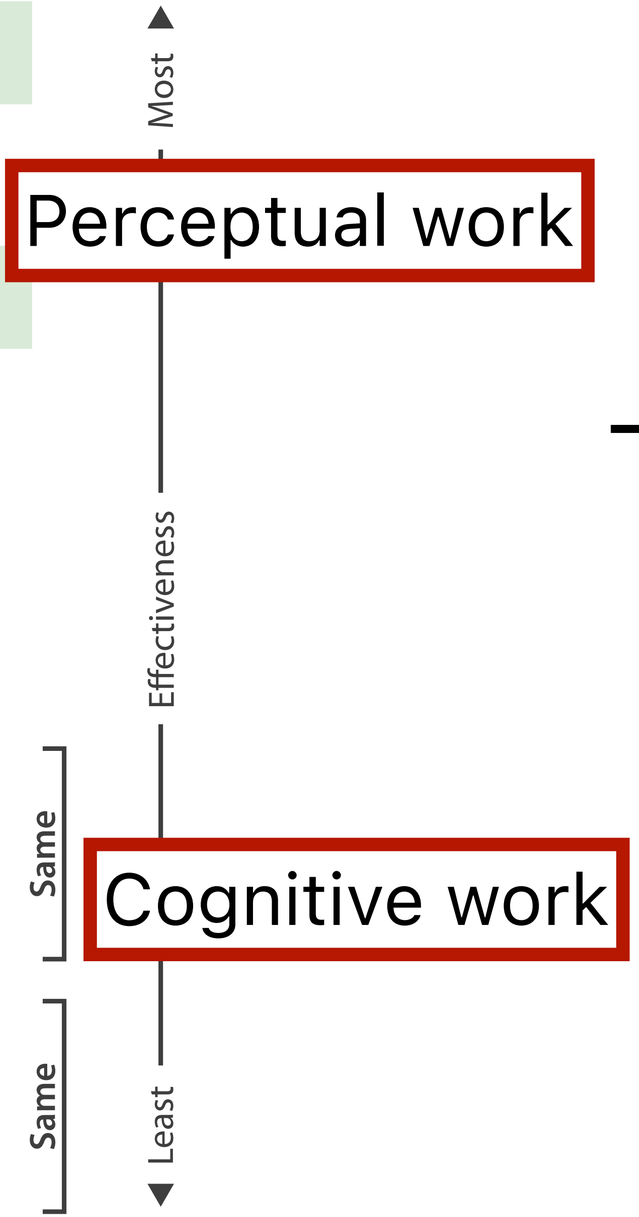
- Spatial region
- Color hue
- Motion
- Shape

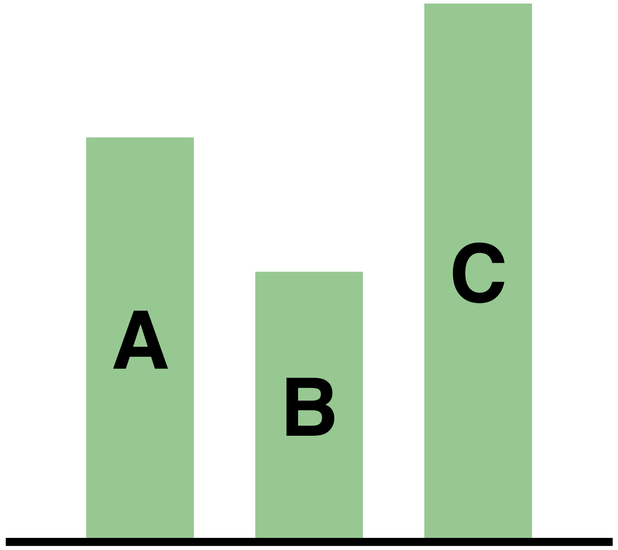
Tamara Munzner, *Visualization Analysis and Design* (2014).

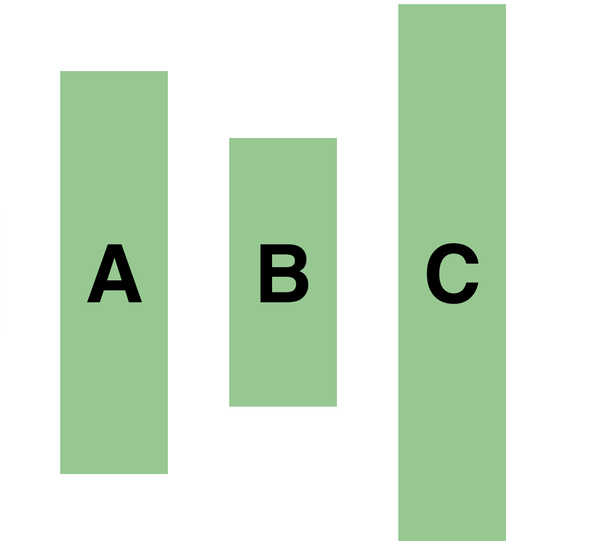
Channels: Expressiveness Types and Effectiveness Ranks

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)







Channels: Expressiveness Types and Effectiveness Ranks

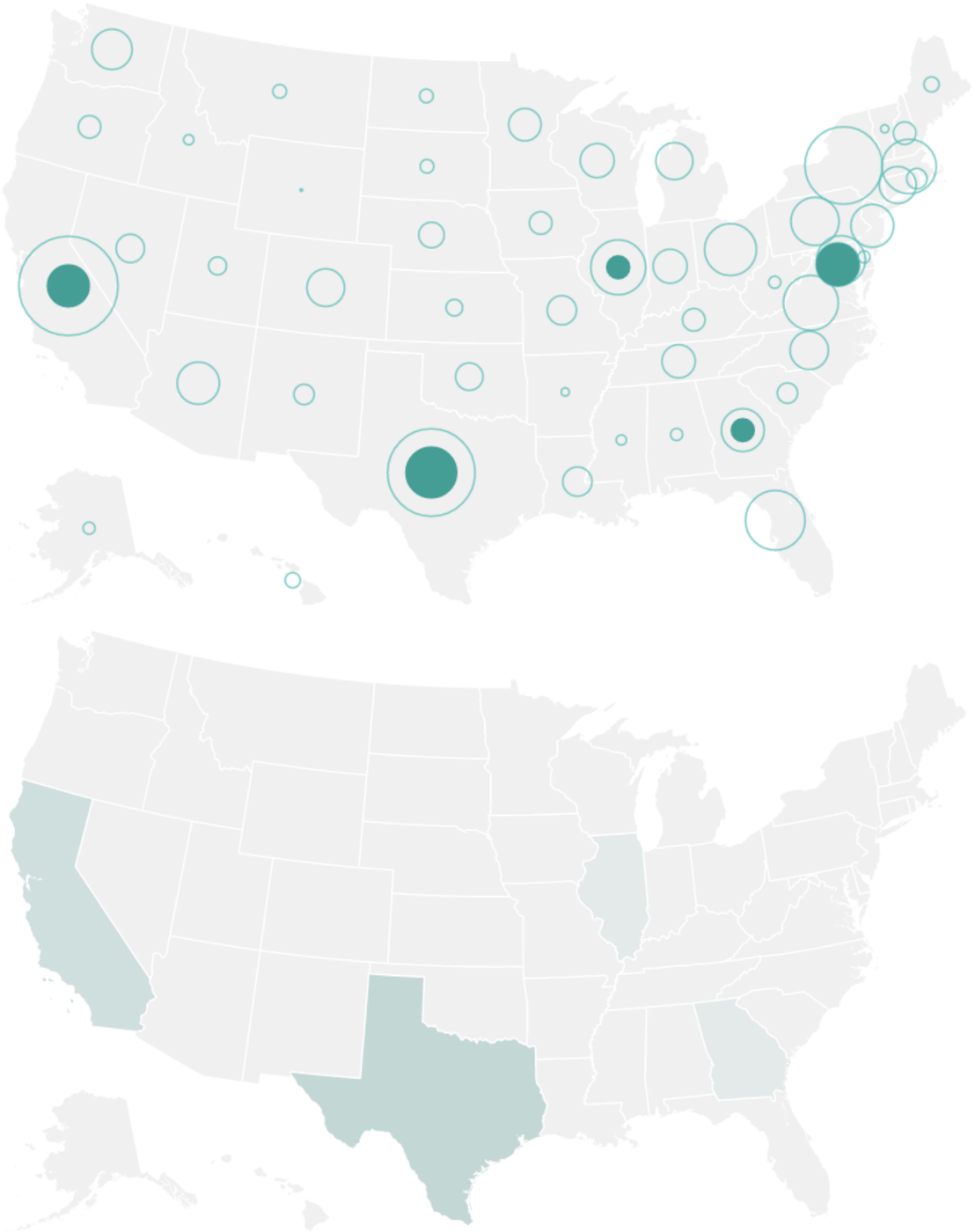
Magnitude Channels: O or Q attributes

- 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

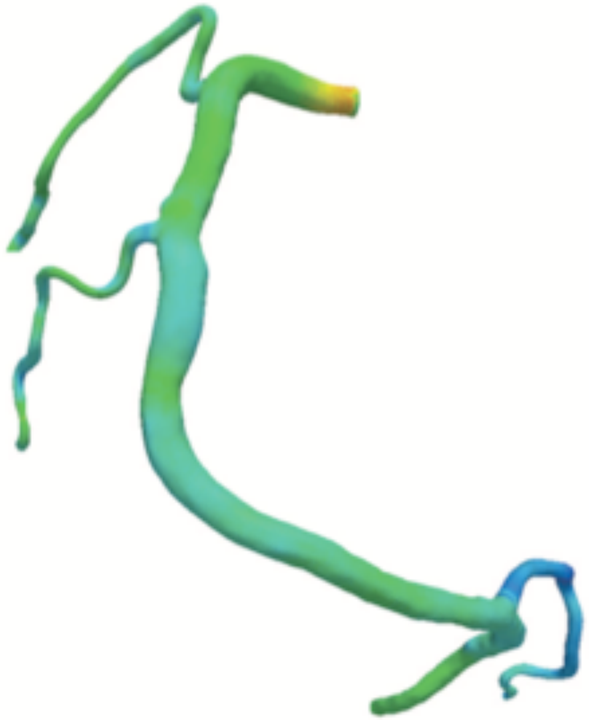
Same

Same



Artery Visualization

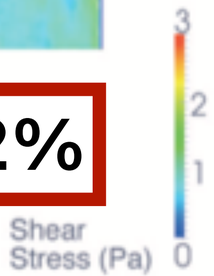
Rainbow Palette



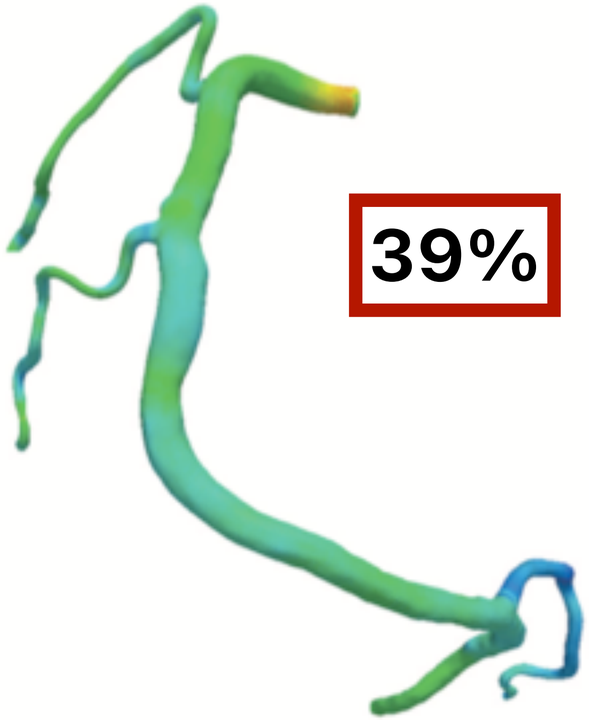
Borkin, Michelle, et al. "Evaluation of artery visualizations for heart disease diagnosis." *2011*

Artery Visualization

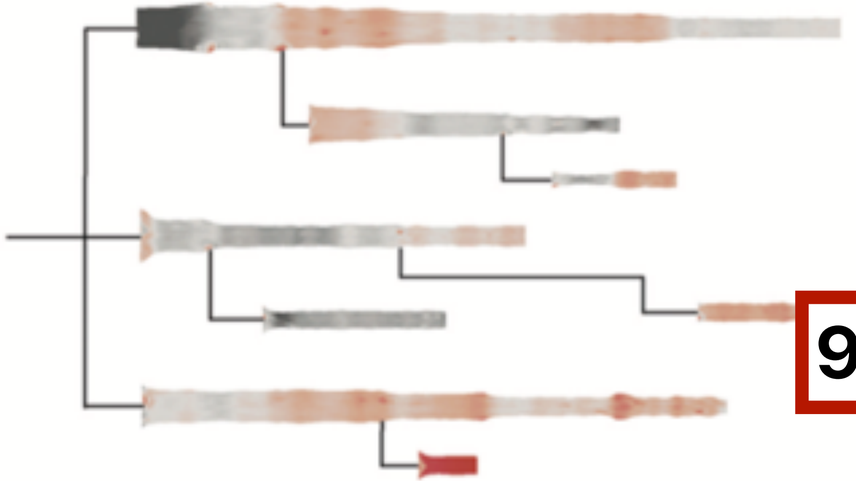
62%



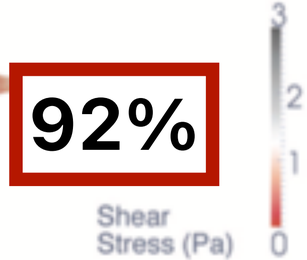
Rainbow Palette



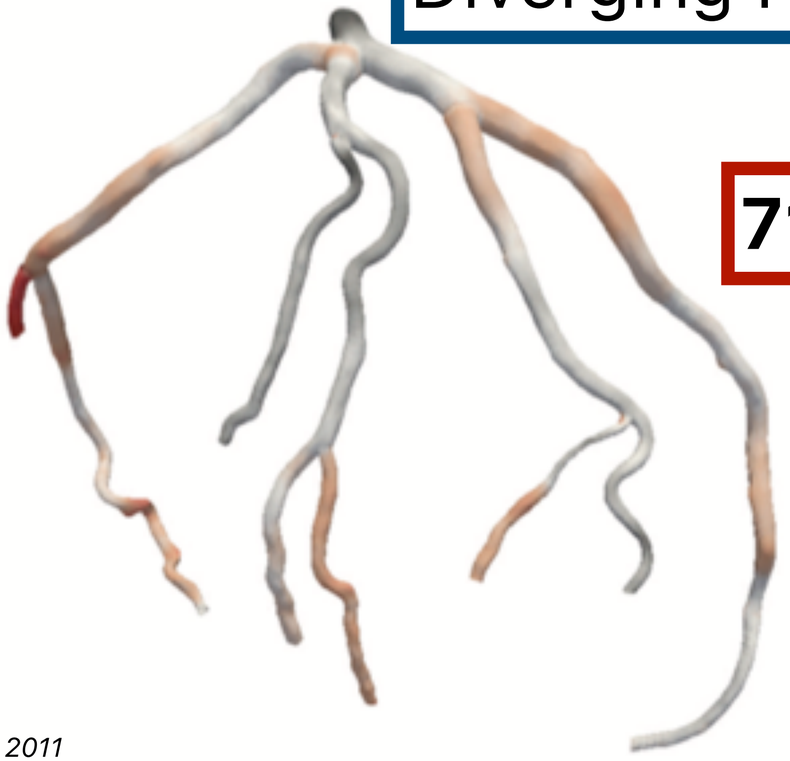
39%



92%



Diverging Palette



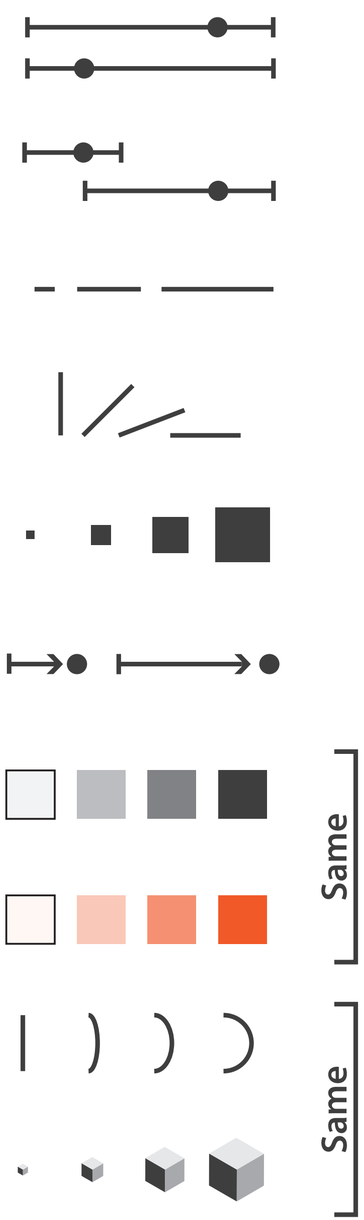
71%

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

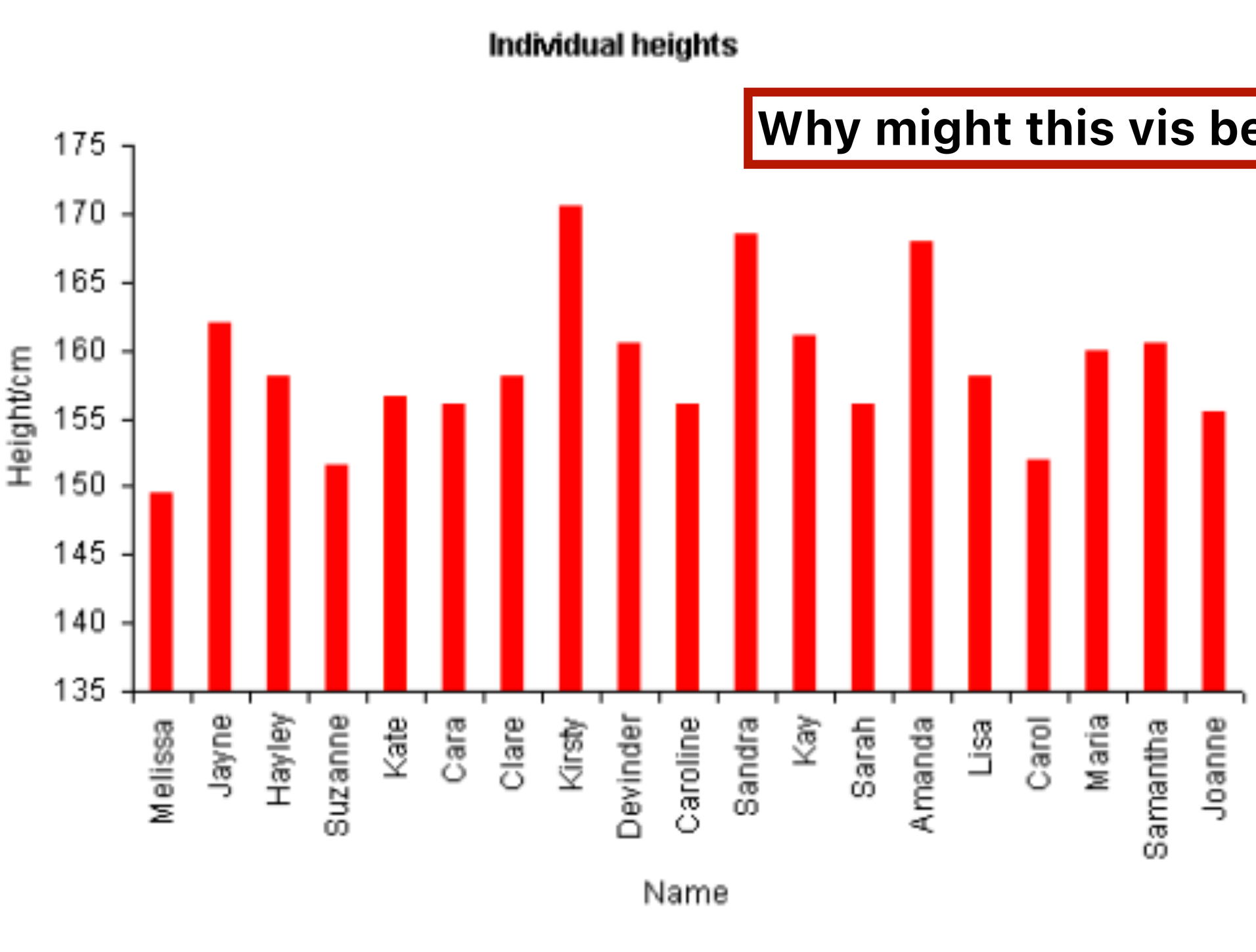
(De-)Obfuscating data

(Mis)leading the witness

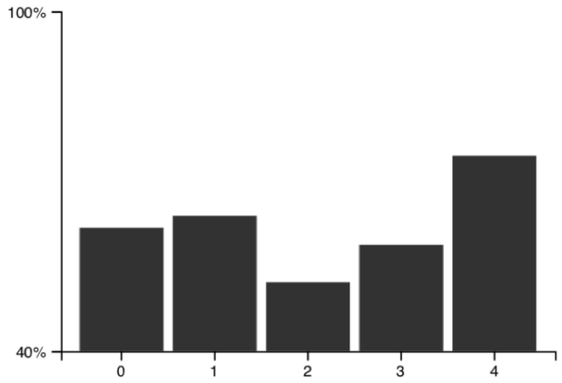
Using space (in)effectively

(De-)Obfuscating data

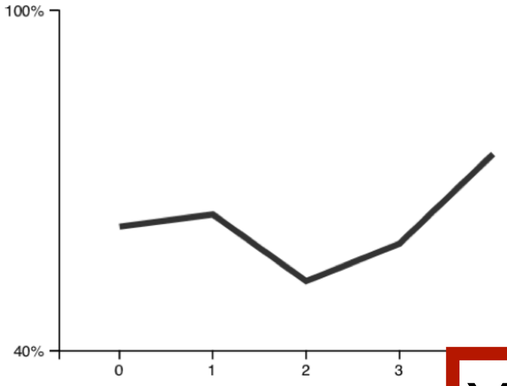
(Mis)leading the witness



Why might this vis be inexpressive?



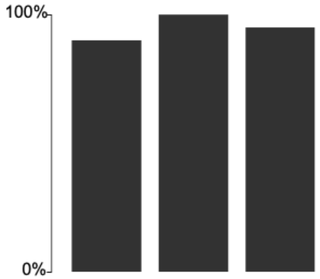
(a) Bar Chart



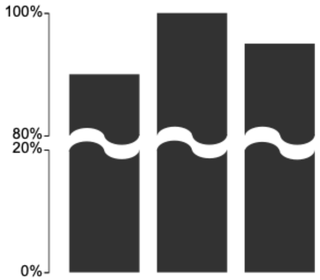
(b) Line Chart

Y-axis truncation impacts perception...

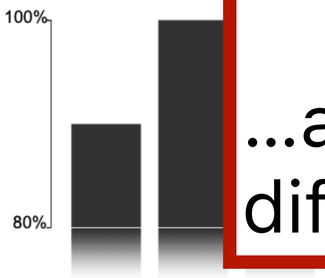
...and interventions did not make a difference.



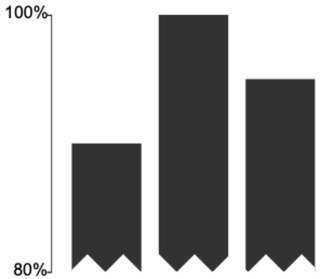
(a) Bar Chart



(b) Broken Axes

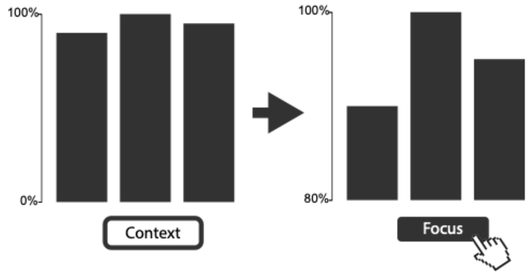


(c) Gradient Bar Chart

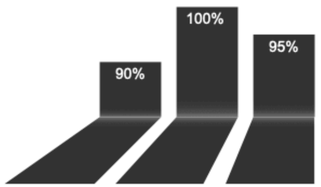


(d) Torn Paper Chart

(e) Panel Chart



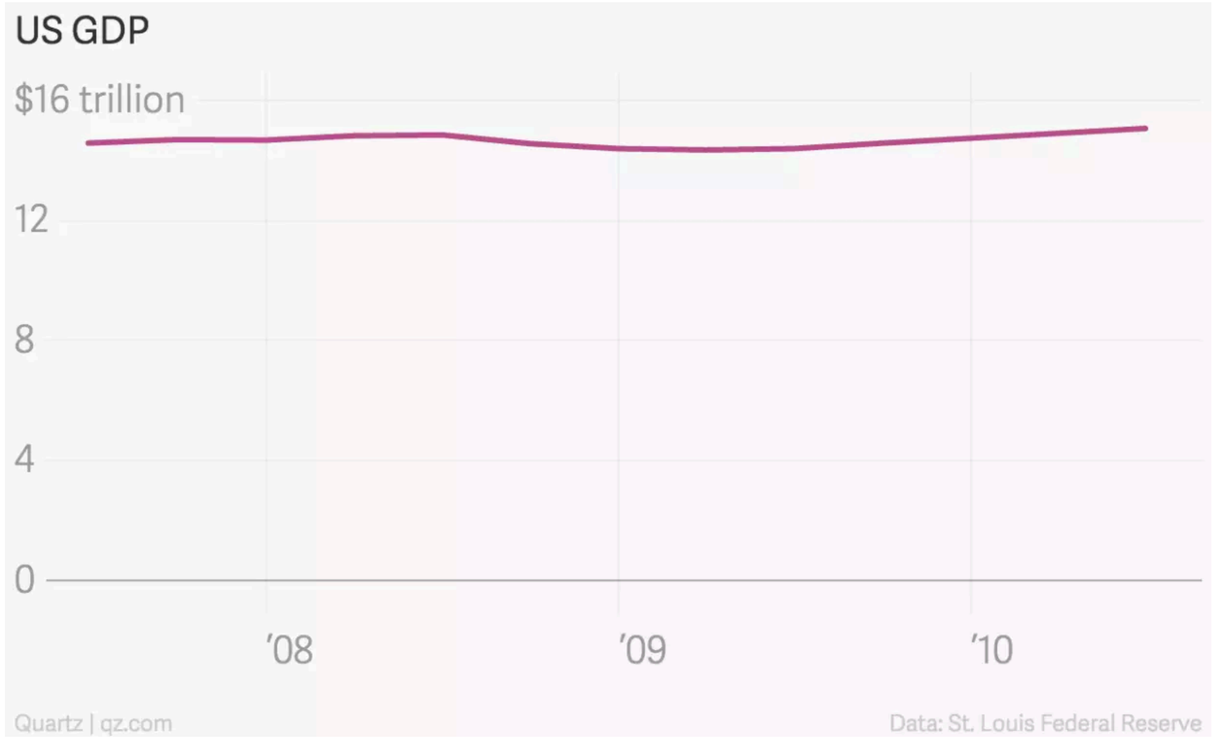
(f) Interactive Focus+Context

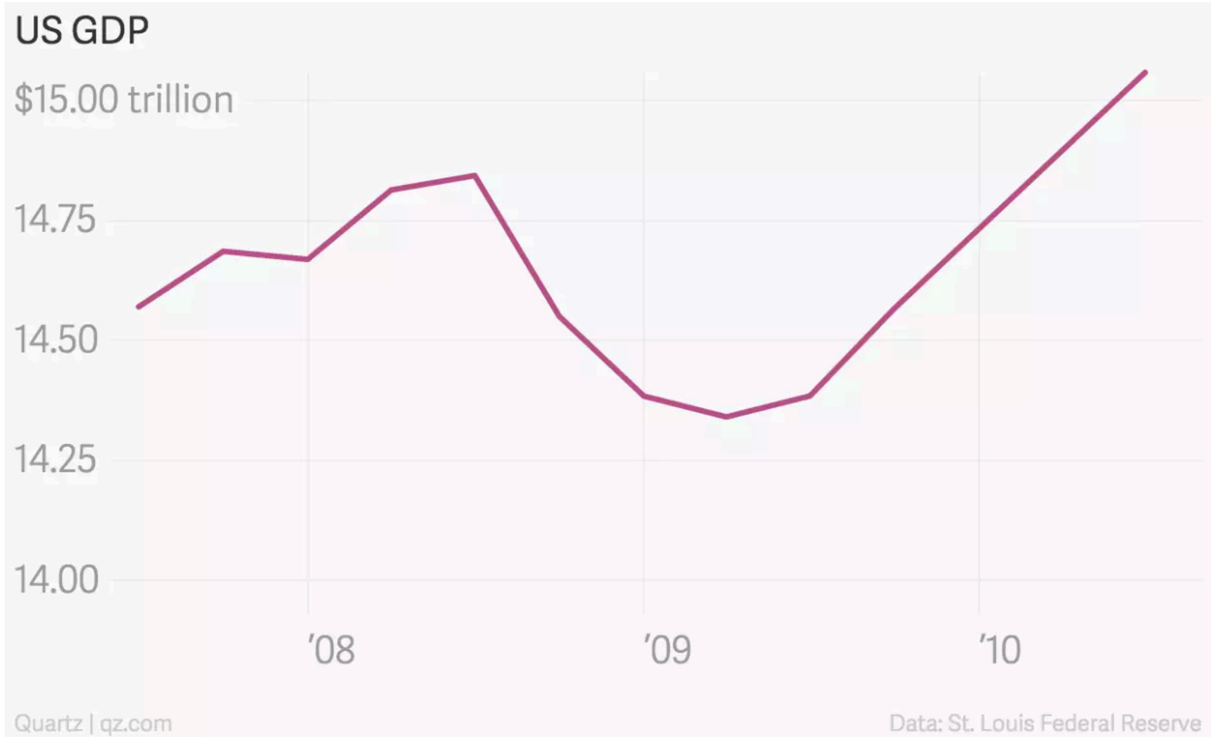


(g) Bent Bar Chart

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

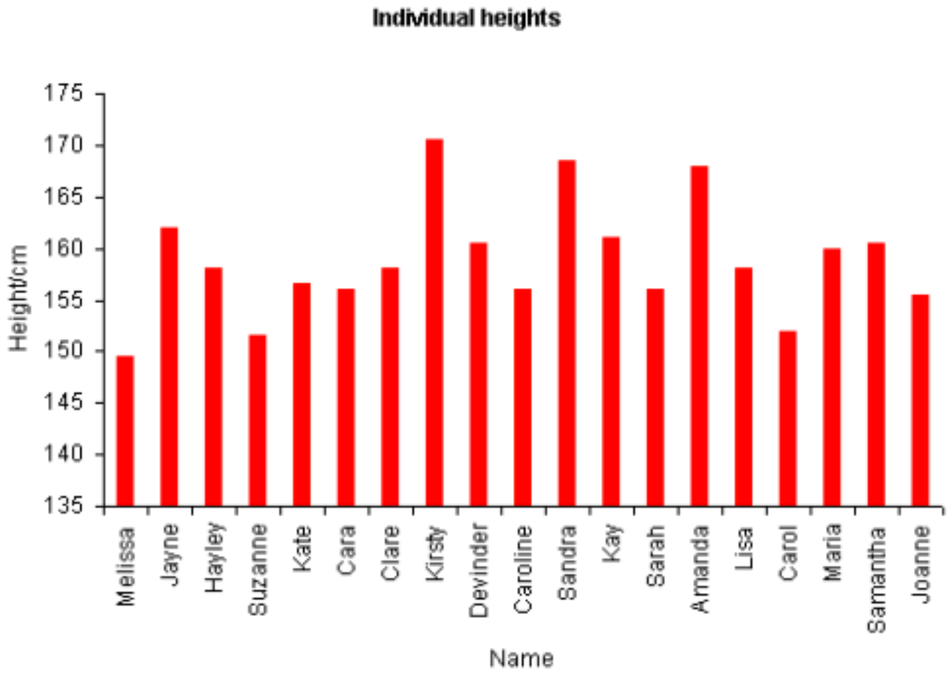
Always start at zero?



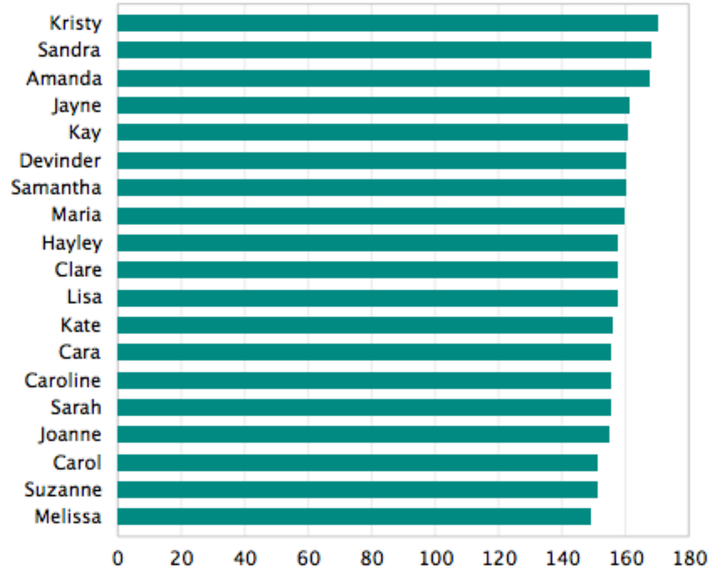


https://qz.com/418083/its-ok-not-to-start-your-y-axis-at-zero

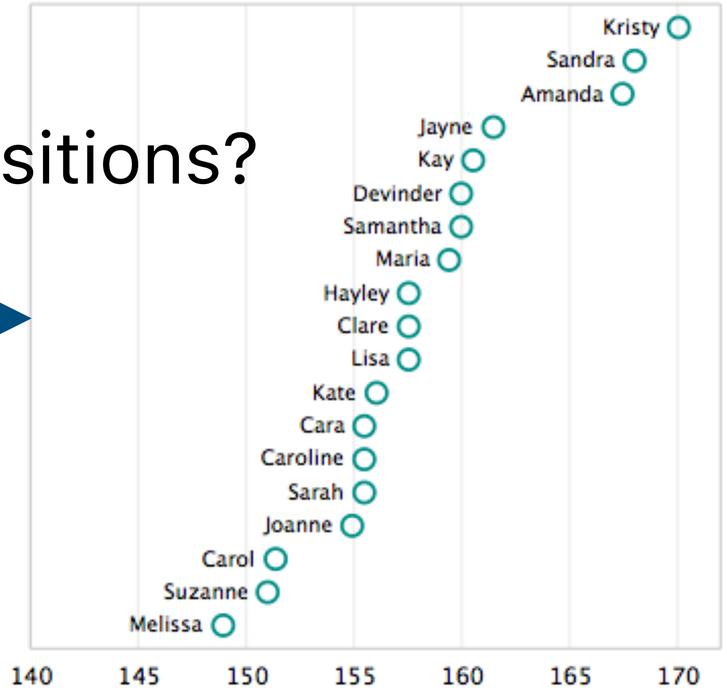
Truncating the y-axis?



Compare proportions?
(Q-ratio)



Compare relative positions?
(Q-interval)



Truncating the y-axis?

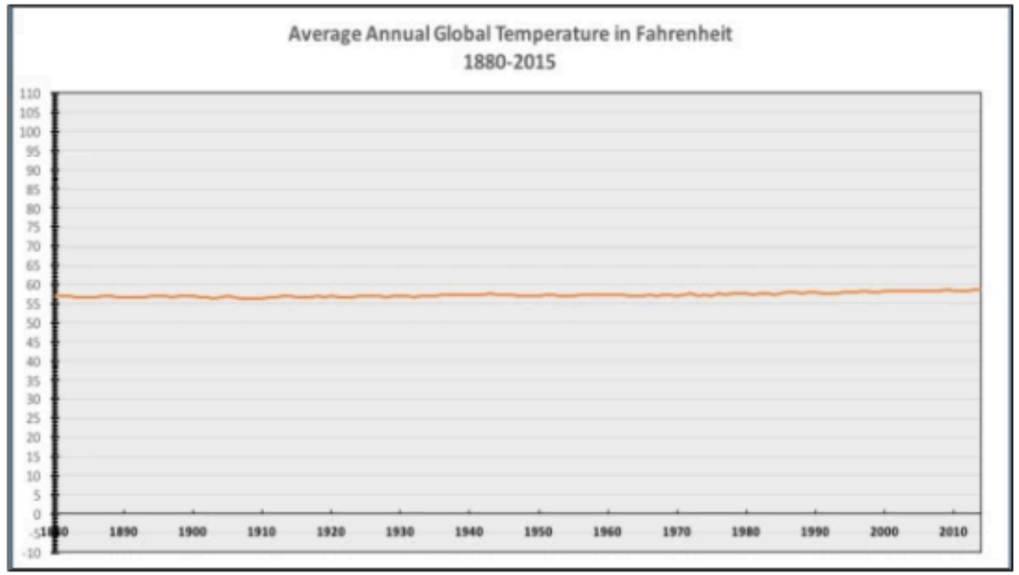
To emphasize Q-interval (vs. Q-ratio)
If the zero value doesn't make much sense.
If it is the norm (e.g., stock charts).

National Review ✓
@NRO

Follow

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

(h/t @powerlineUS)

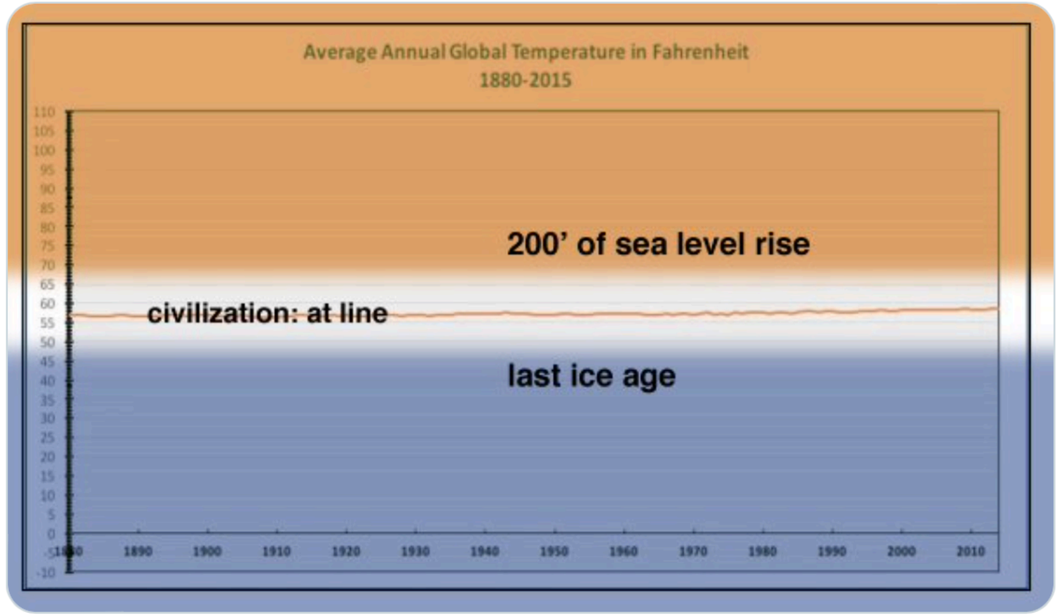


12:36 PM - 14 Dec 2015

City Atlas
@cityatlas

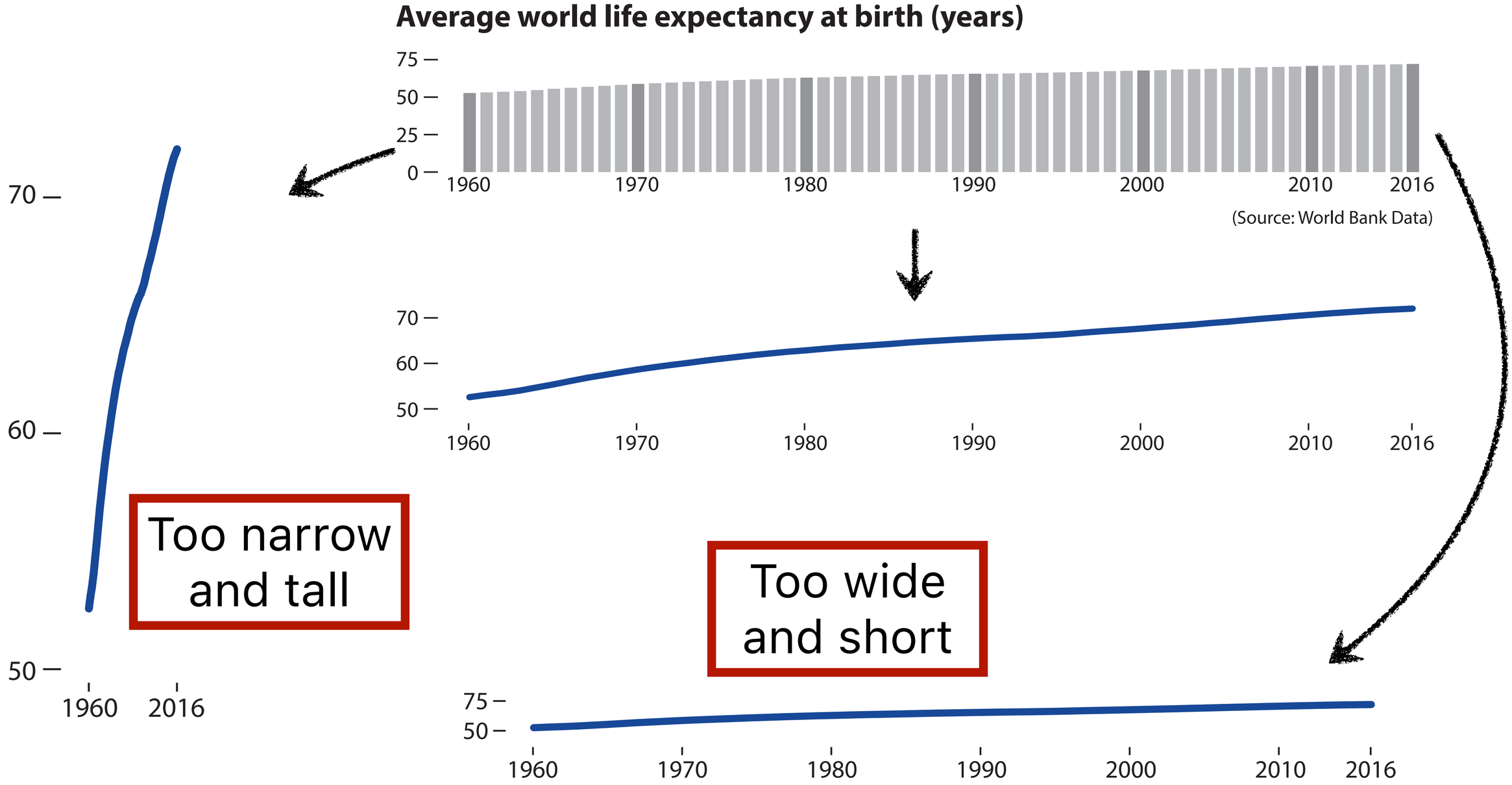
Replying to @NRO

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



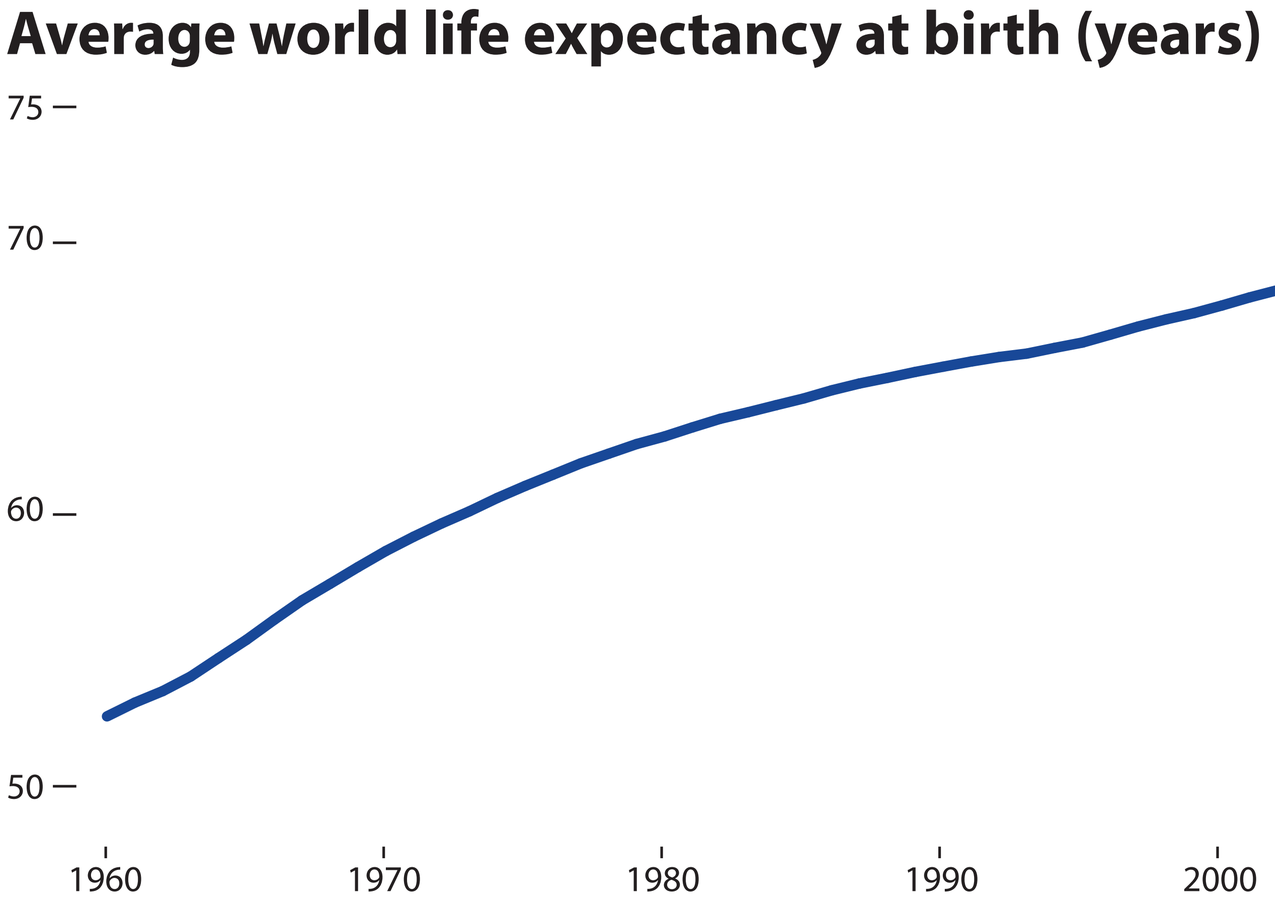
5:28 PM · Dec 14, 2015

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



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

Aspect Ratio

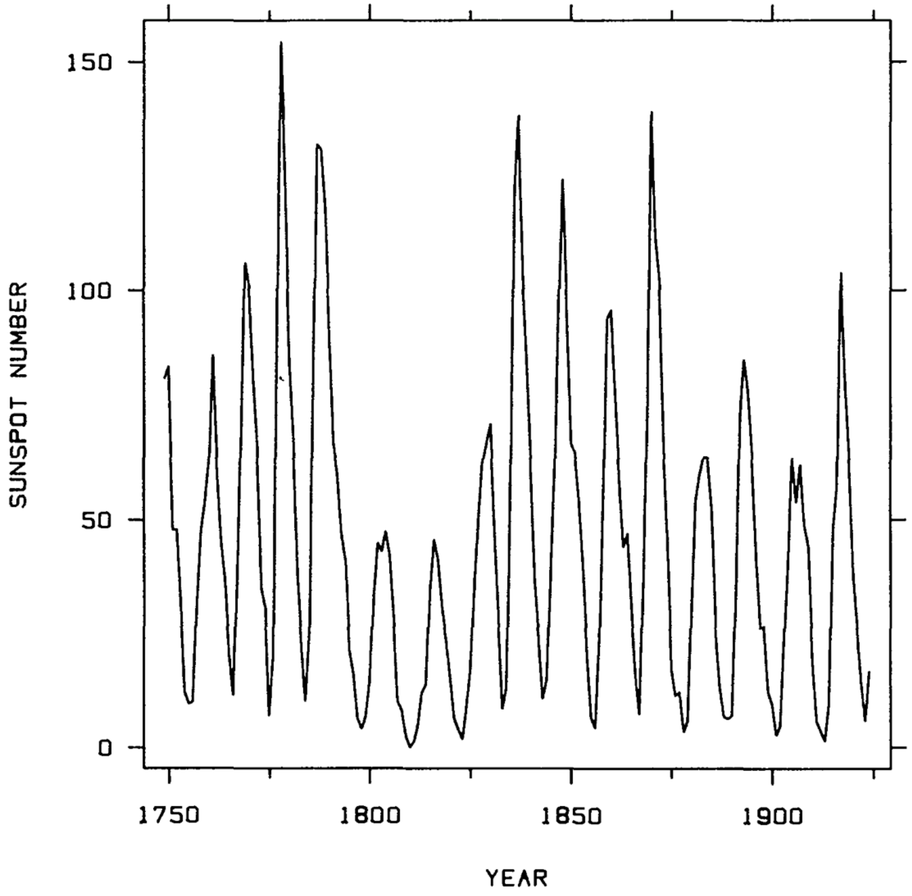


Approximate the proportion of the chart to match the depicted trend.

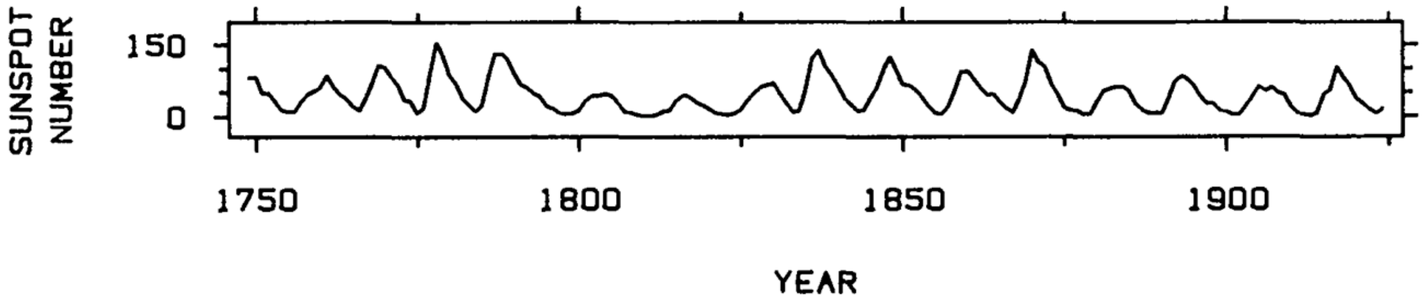
35% increase ≈ 1/3rd
≈ 4:3 aspect ratio

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

Aspect Ratio

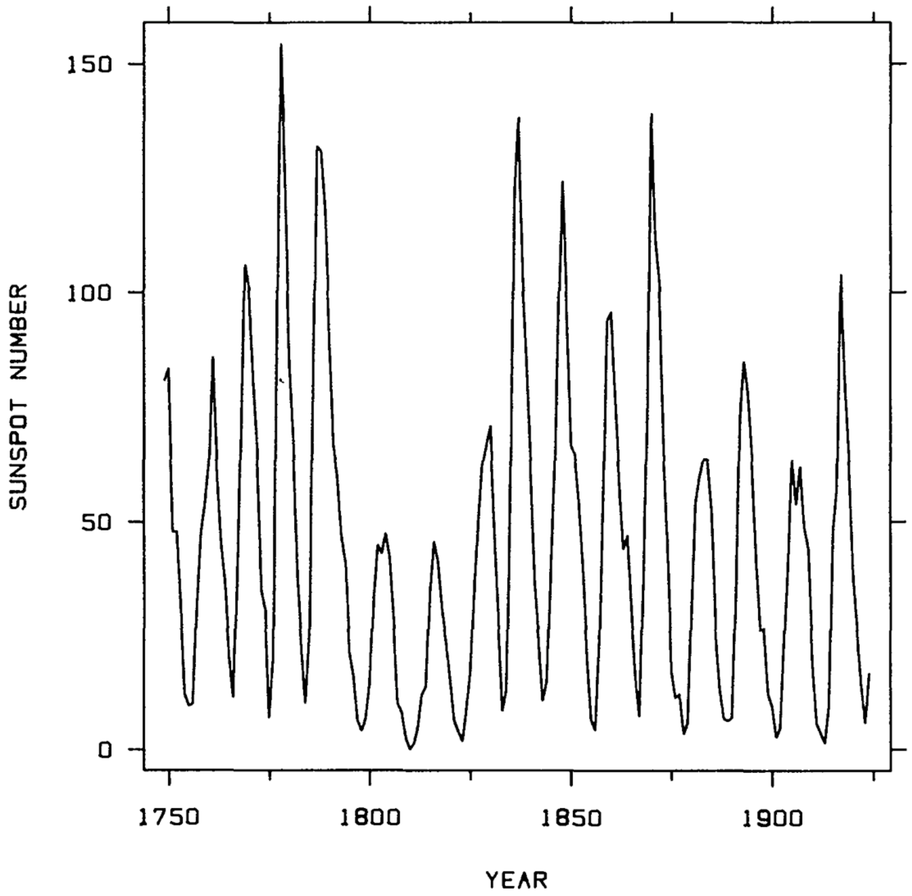


1. Approximate the proportion of the chart to match the depicted trend.

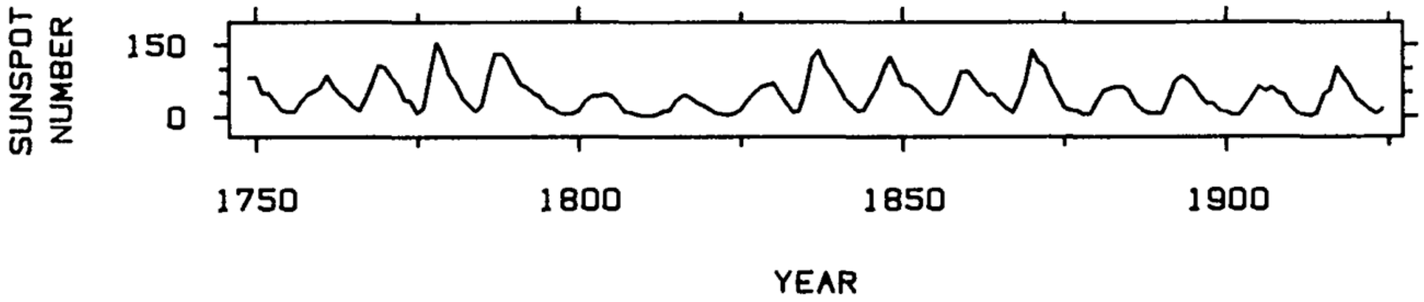


Cleveland, William S., and Robert McGill. "Graphical perception: The visual decoding of quantitative information on graphical displays of data." *1987*

Aspect Ratio



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



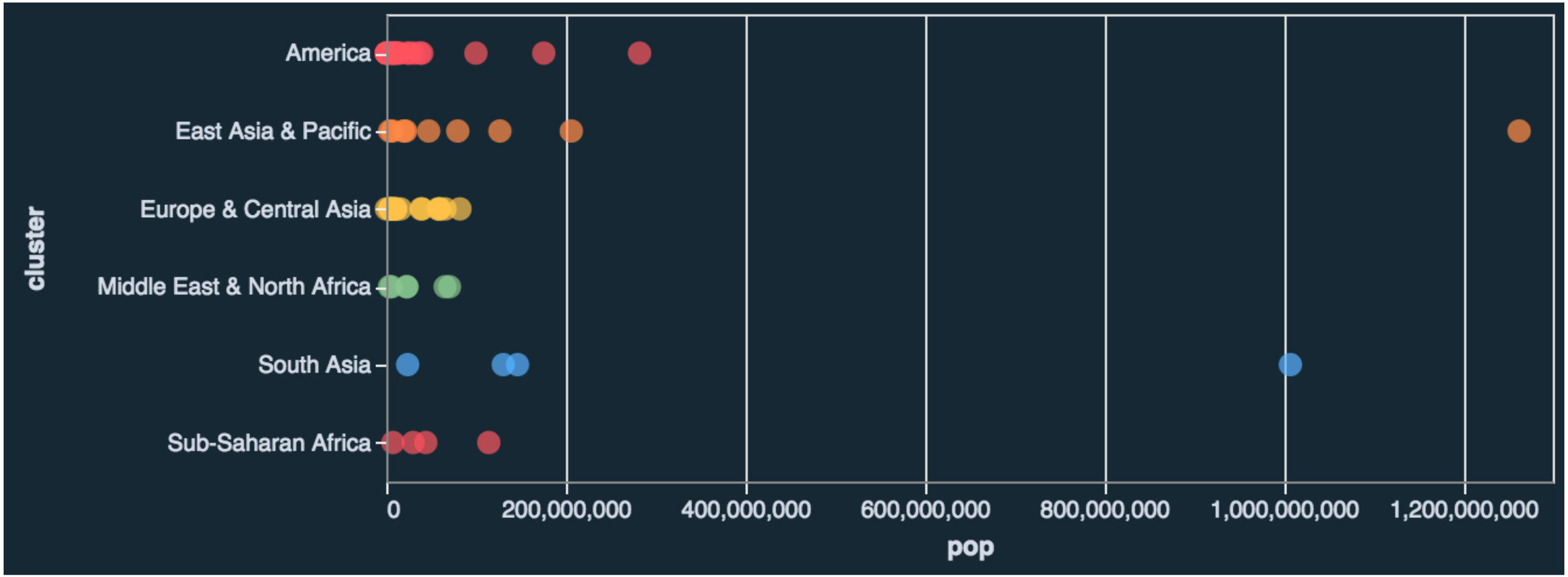
2. Bank to 45°: aspect ratios with 45° avg. line segment orientation.

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

Scaling Axes: Outliers and Skew

Options:

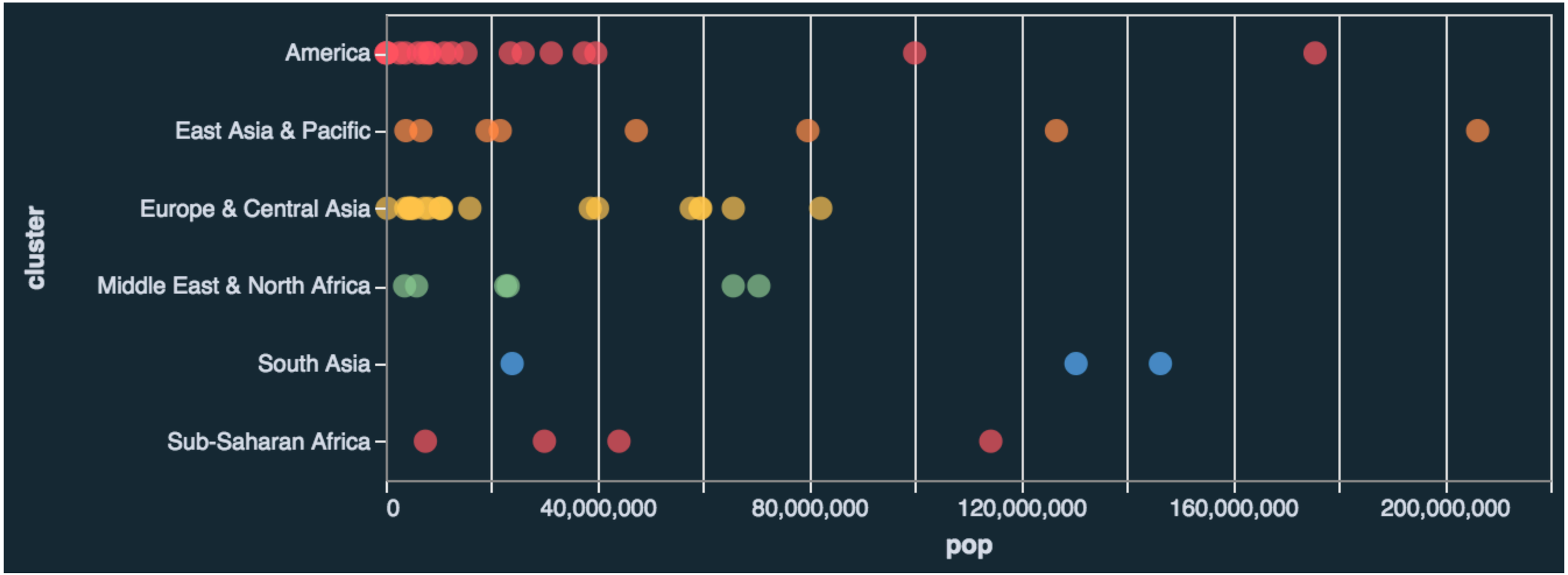
1. Clip them out.



Scaling Axes: Outliers and Skew

Options:

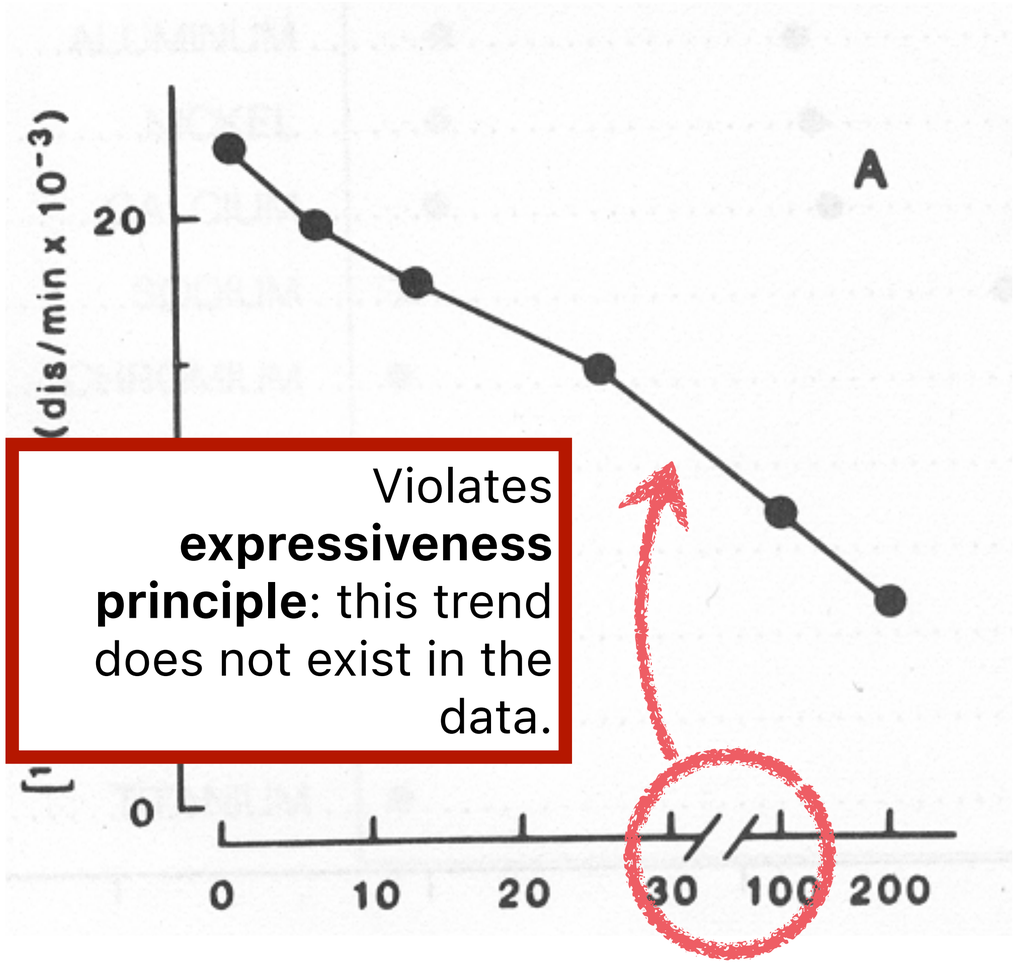
1. Clip them out.



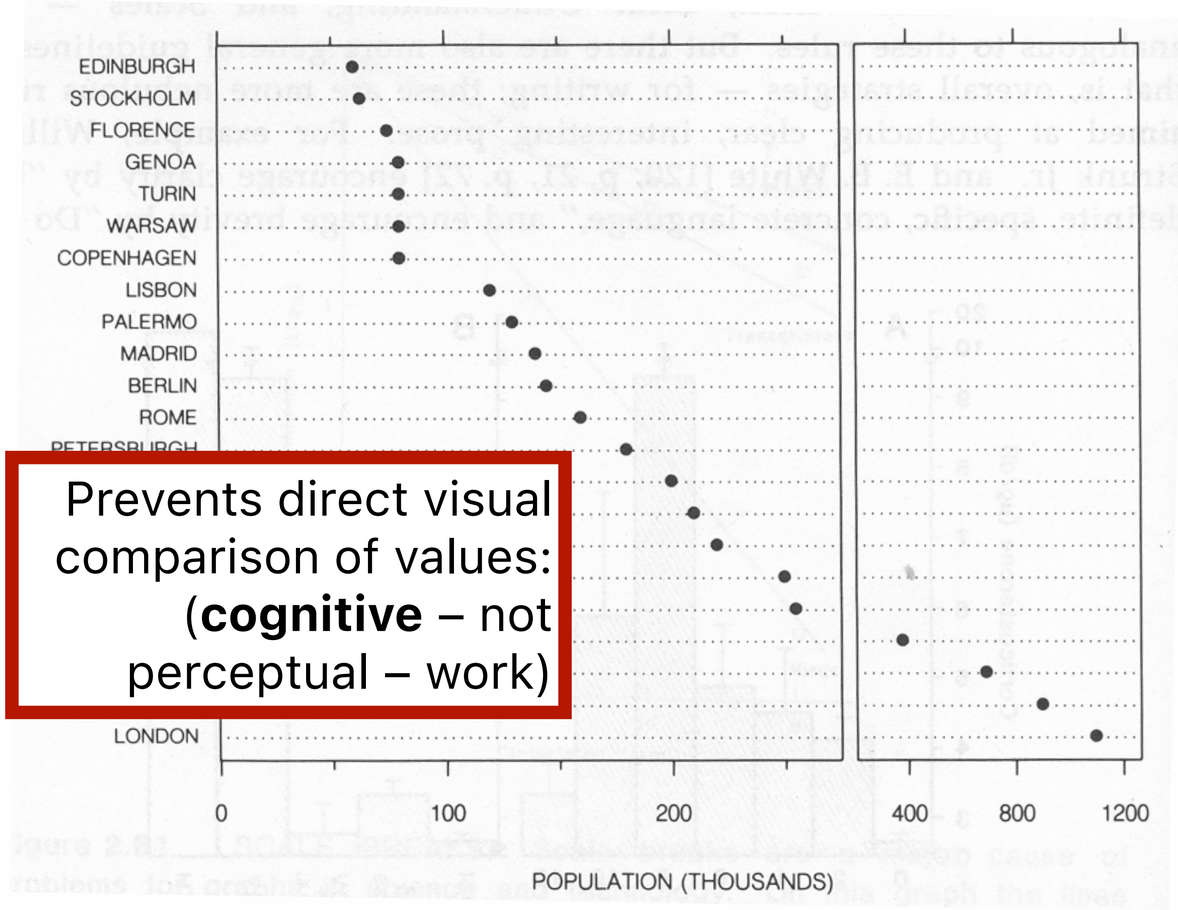
Scaling Axes: Outliers and Skew

Options:

1. Clip them out.
2. Scale breaks



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

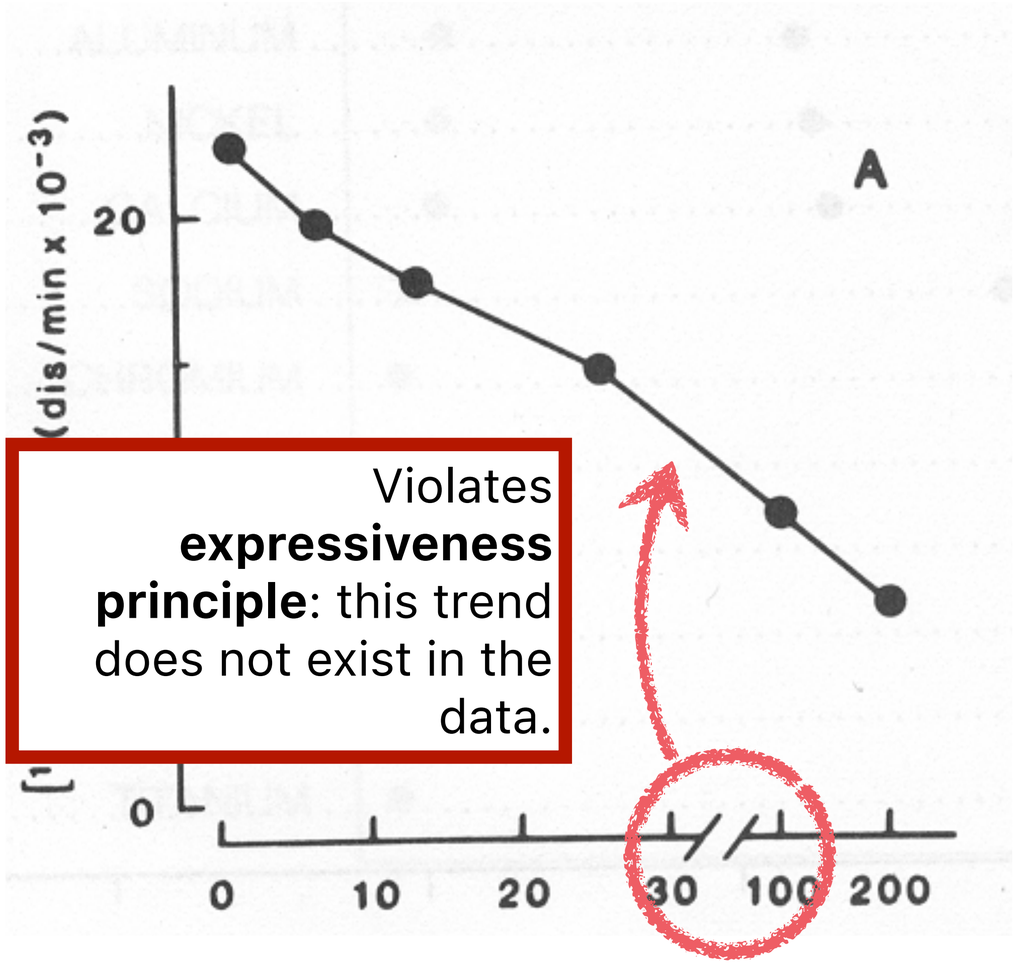


Prevents direct visual comparison of values: (**cognitive** – not perceptual – work)

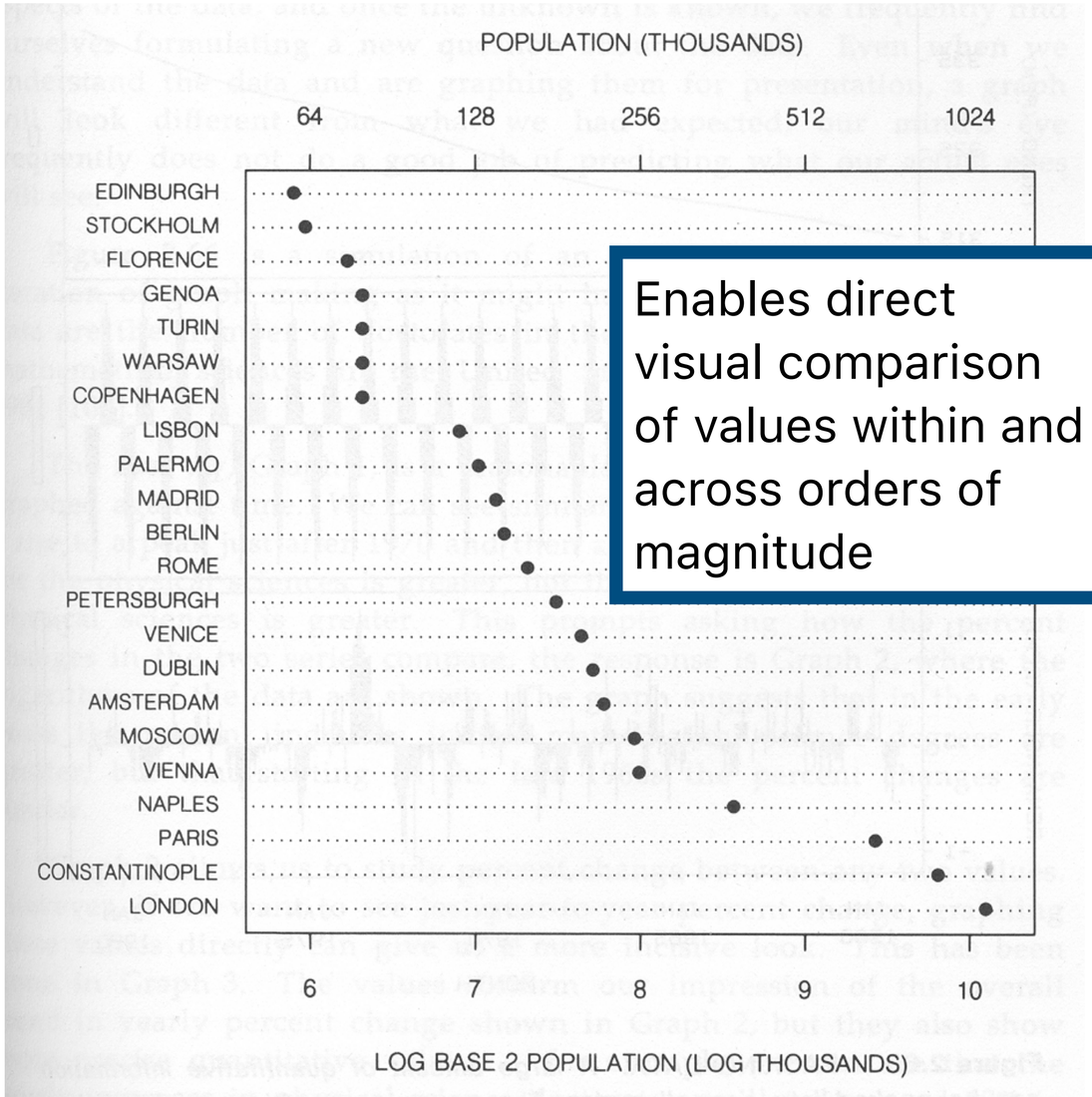
Scaling Axes: Outliers and Skew

Options:

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



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

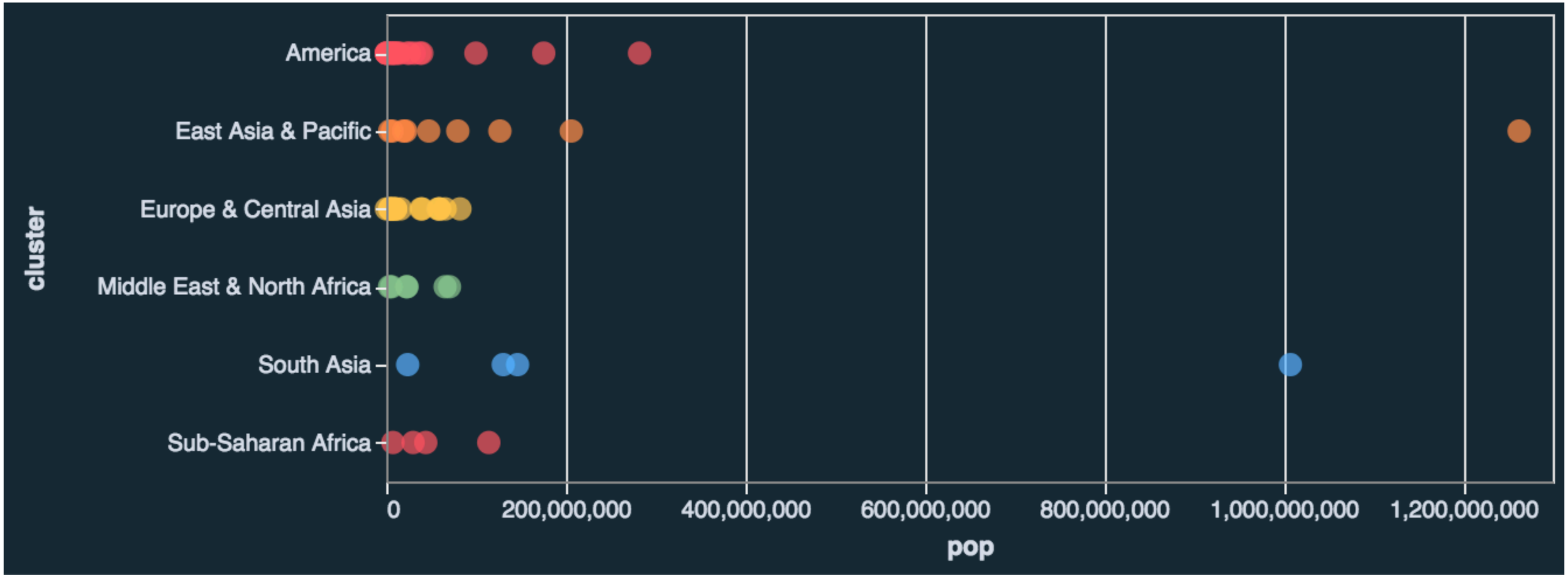


Enables direct visual comparison of values within and across orders of magnitude

Scaling Axes: Outliers and Skew

Options:

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



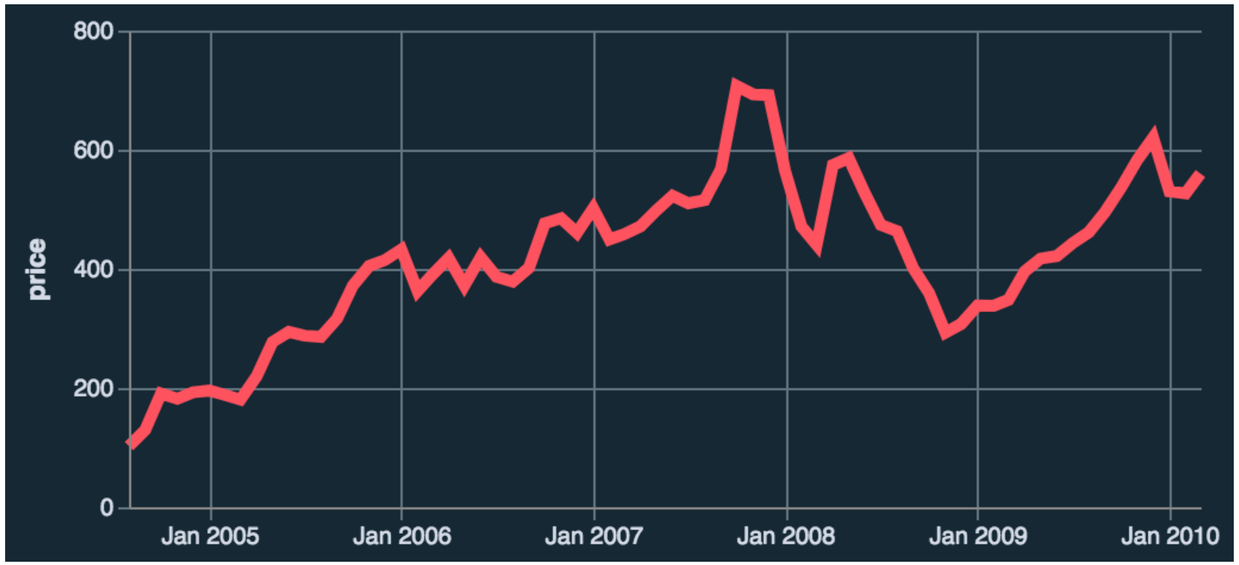
Scaling Axes: Outliers and Skew

Options:

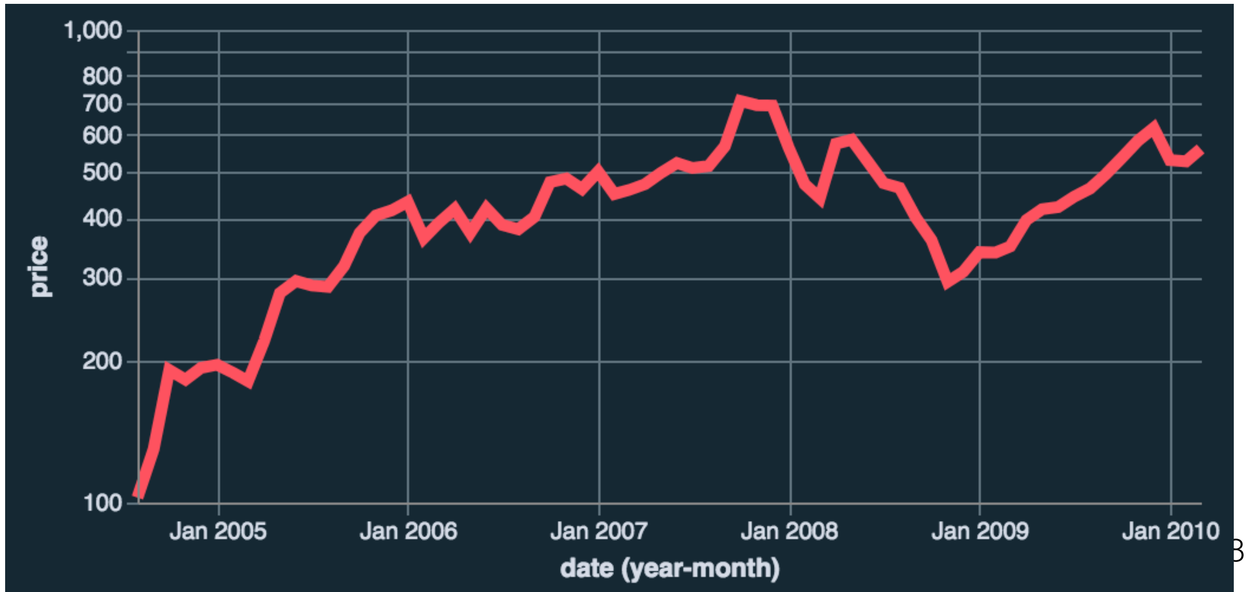
1. Clip them out.
2. Scale breaks
3. Log scale

Scaling Axes: Linear vs Log

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



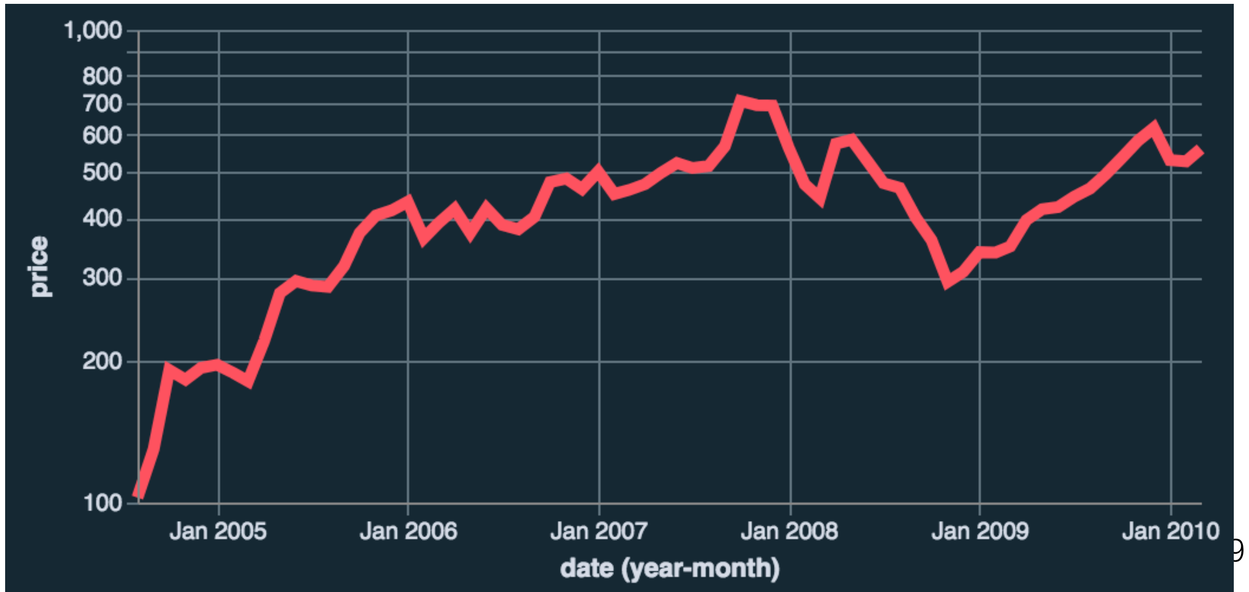
Log Scale
Percentage change
10 visual units =
multiplication of 10 data units



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

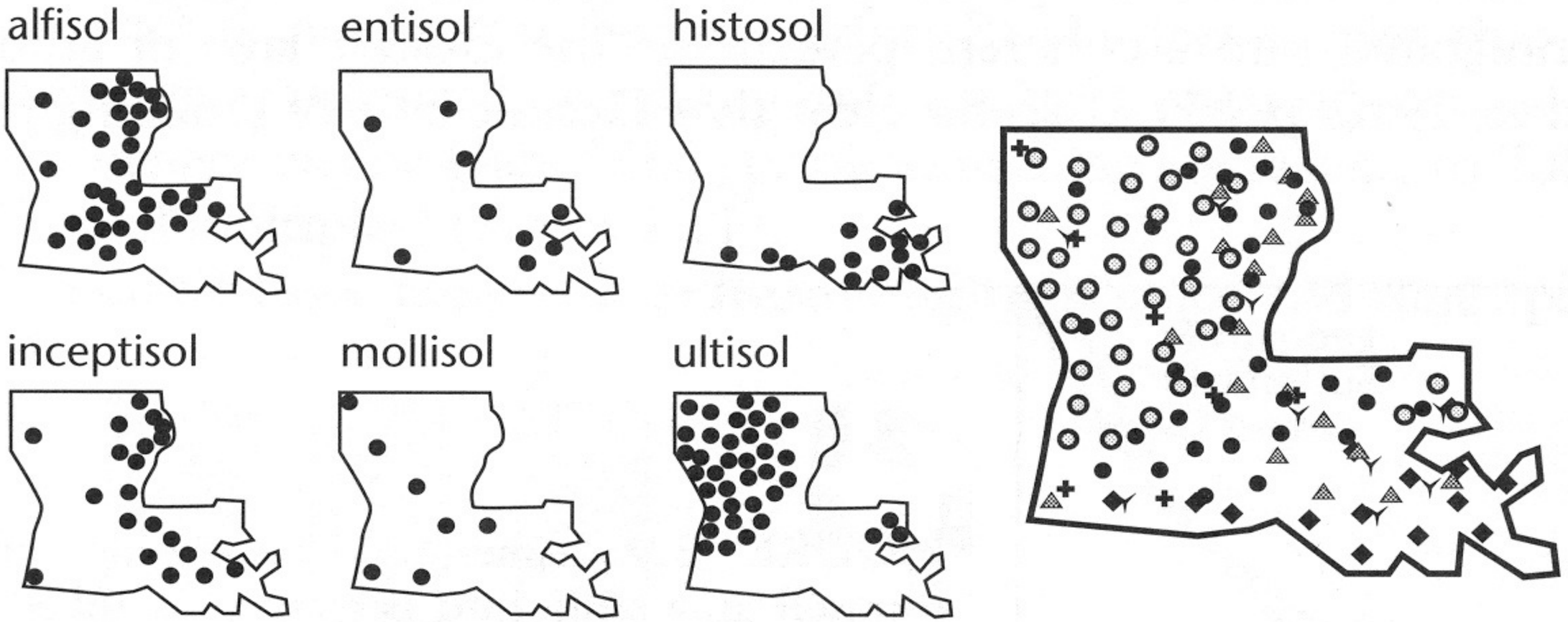
(De-)Obfuscating data

(Mis)leading the witness

Using space (in)effectively

(De-)Obfuscating data

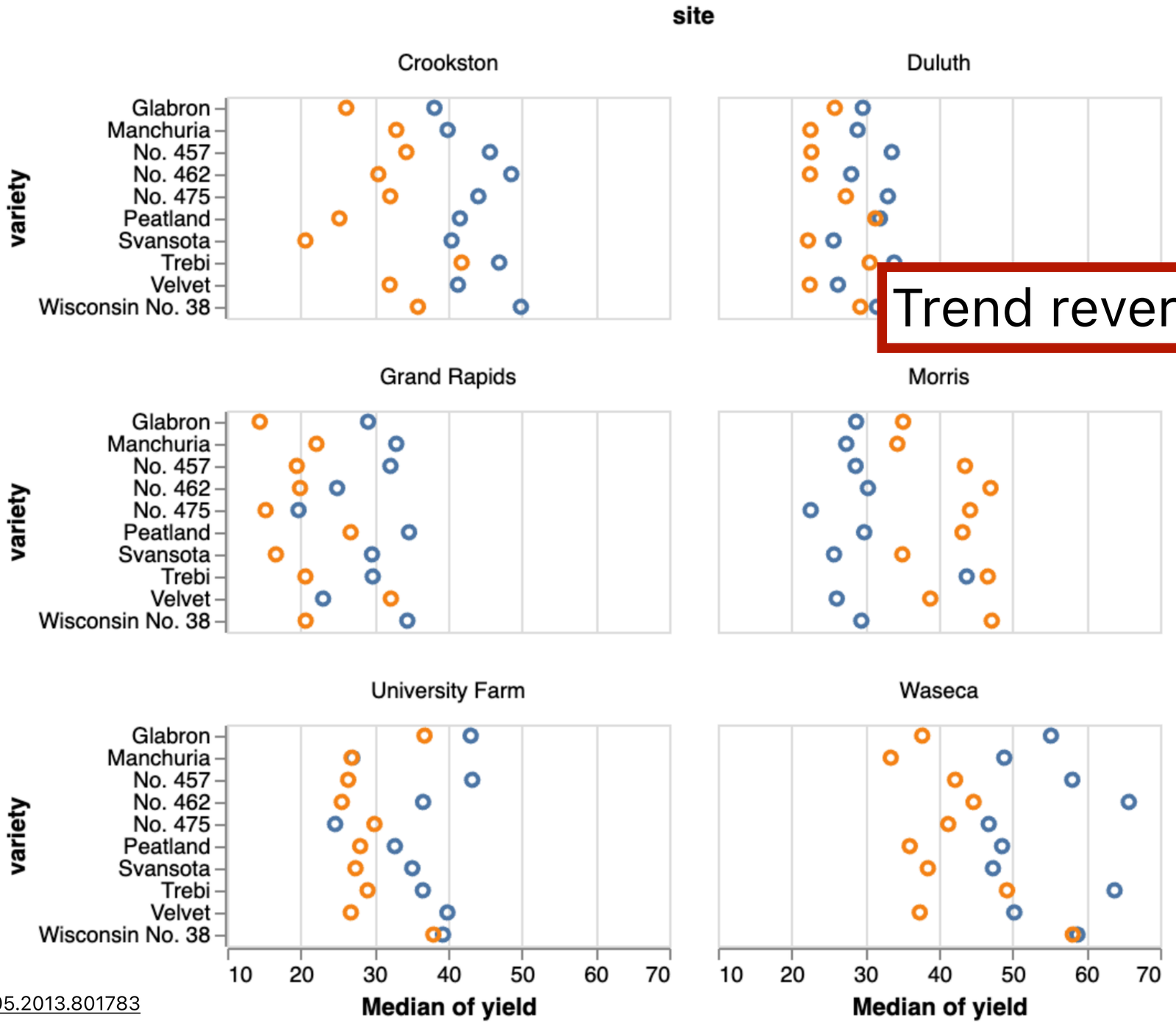
(Mis)leading the witness


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

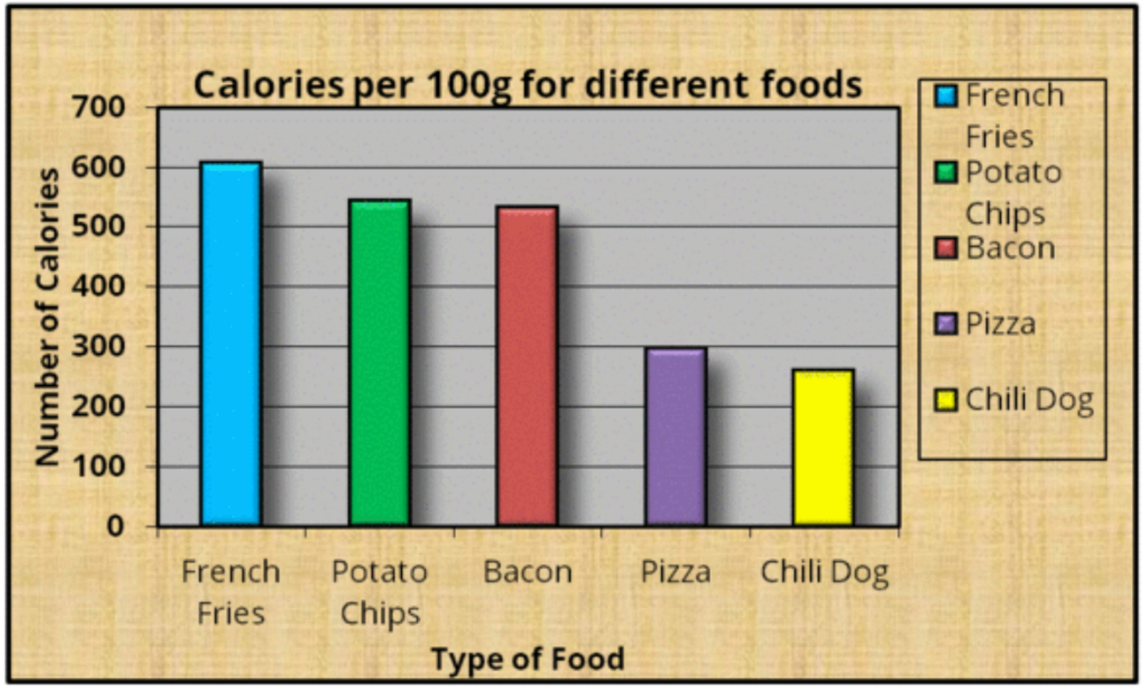
to improve

(the **data-ink** ratio)

Data-ink Ratio

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?

tryclassbuzz.com:
dataink



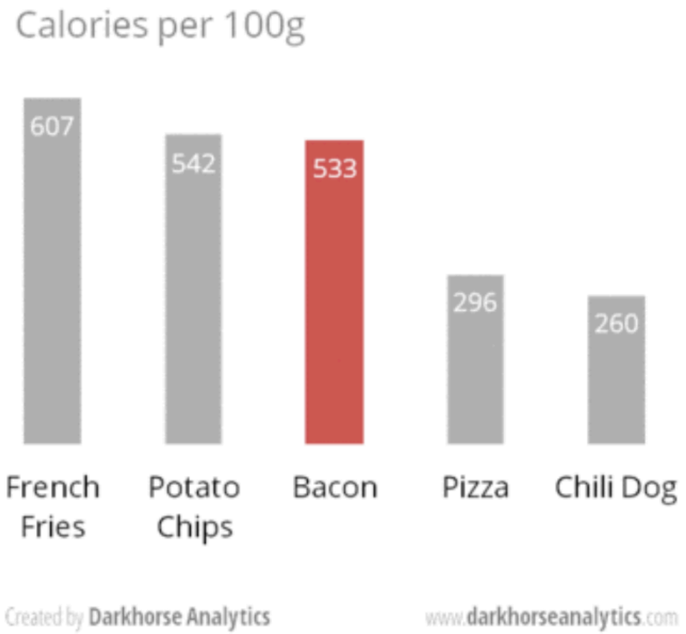
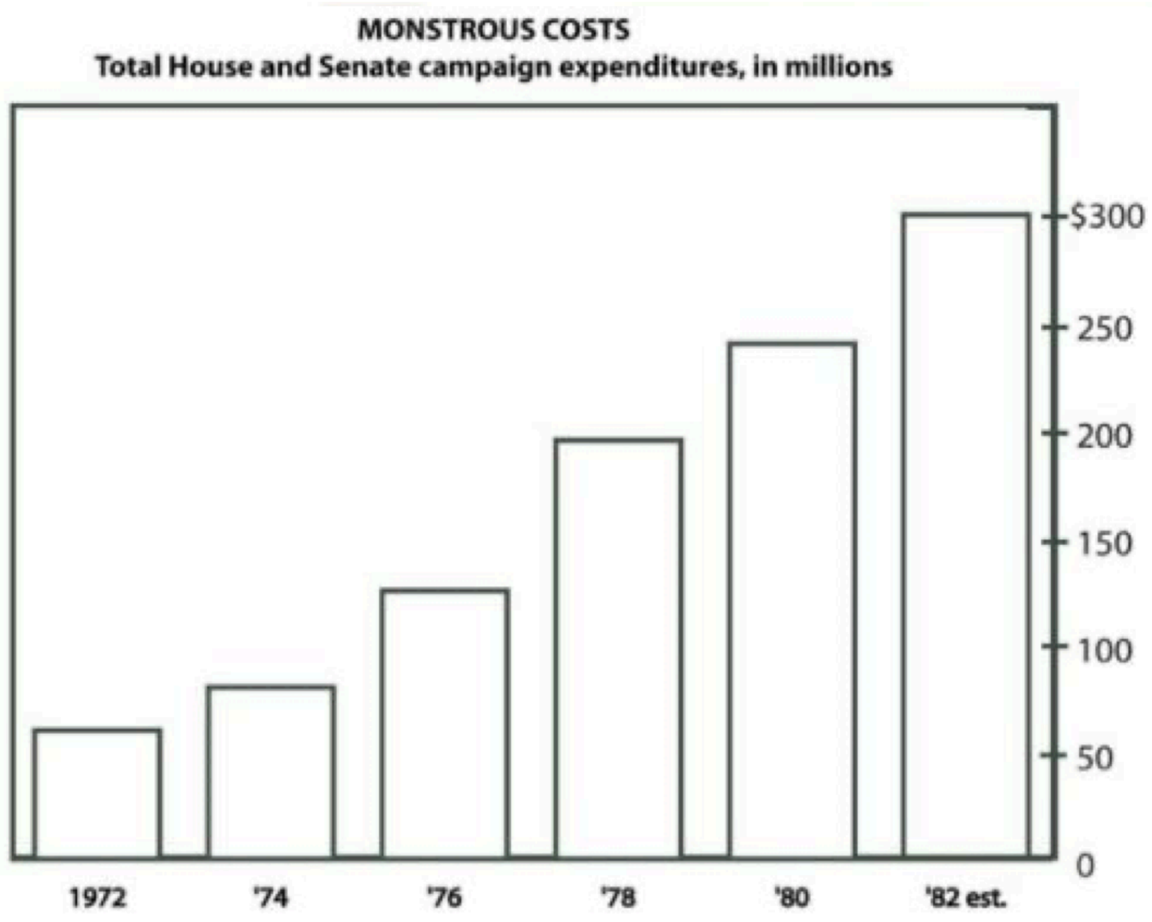
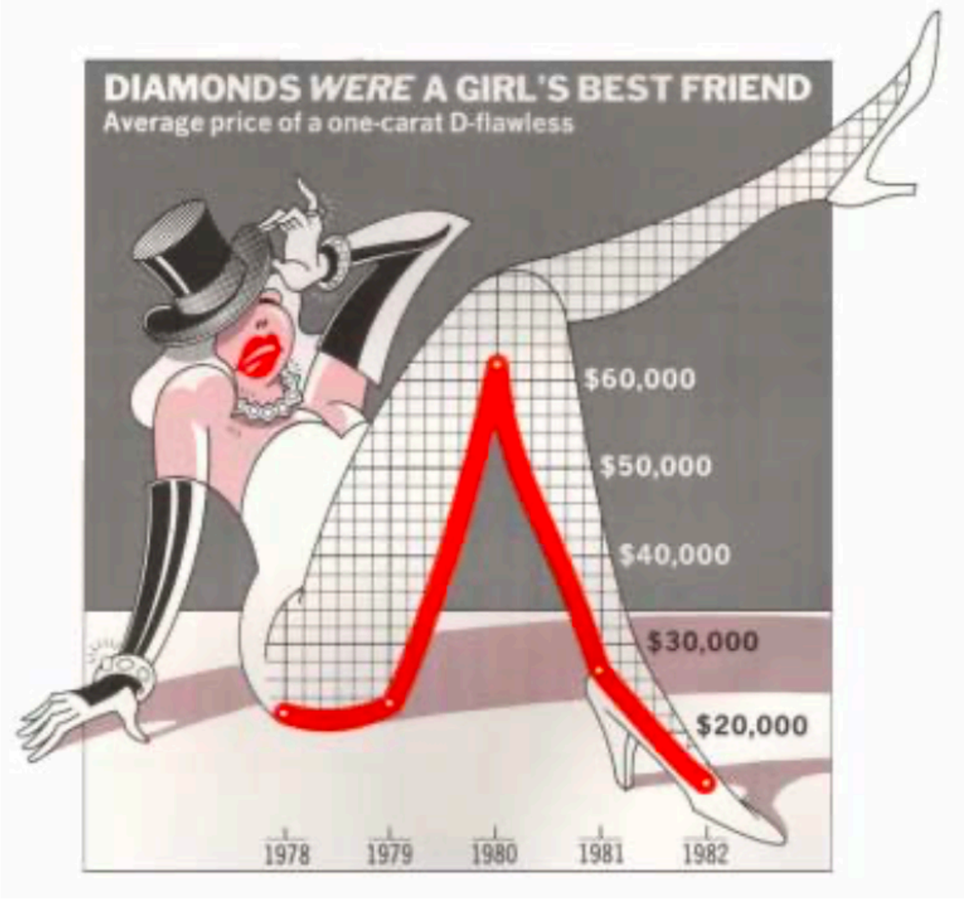


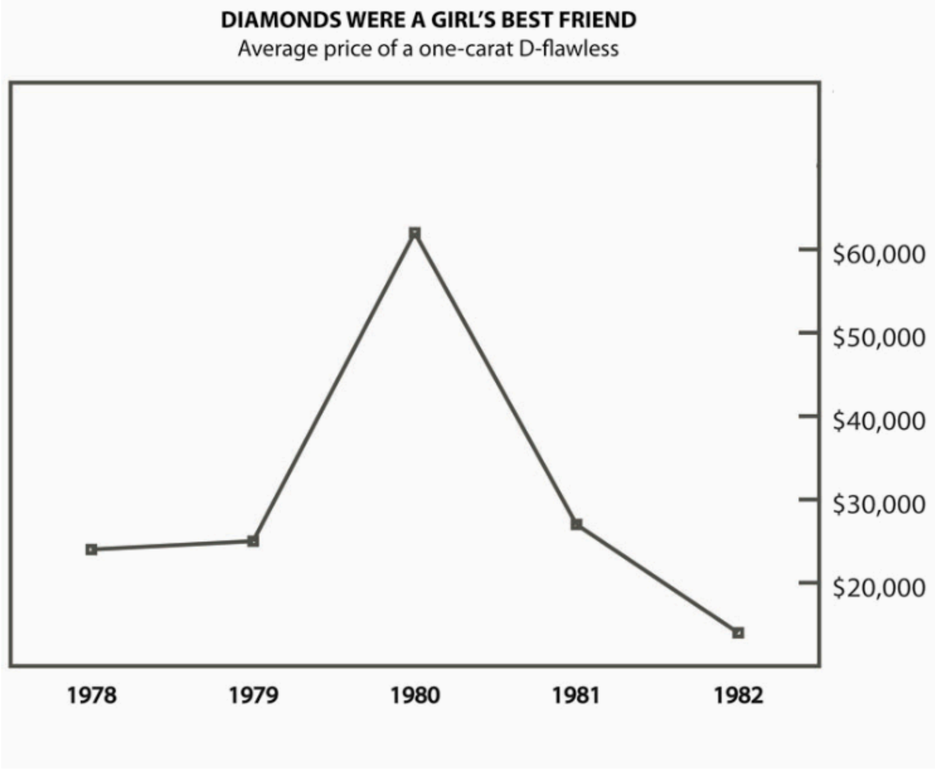
Chart "Junk"



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

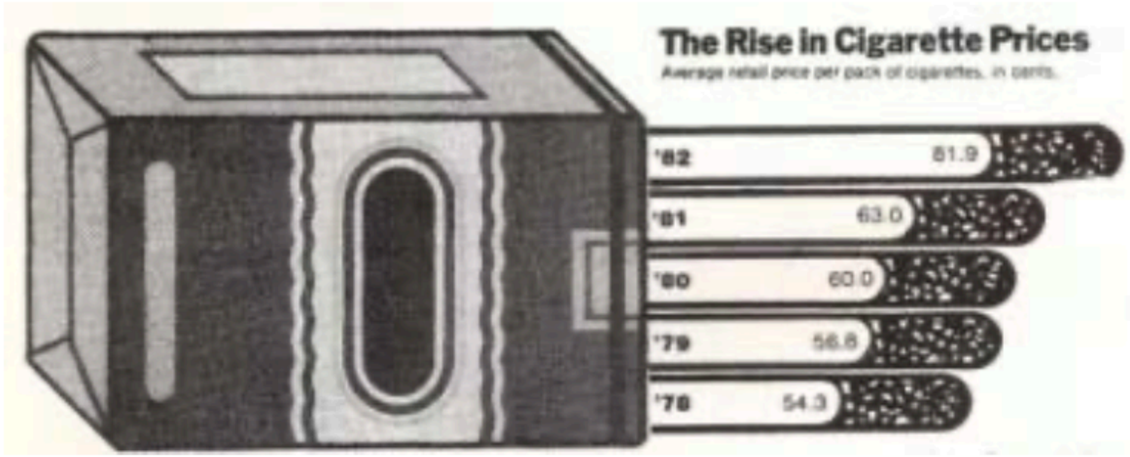
Chart "Junk"

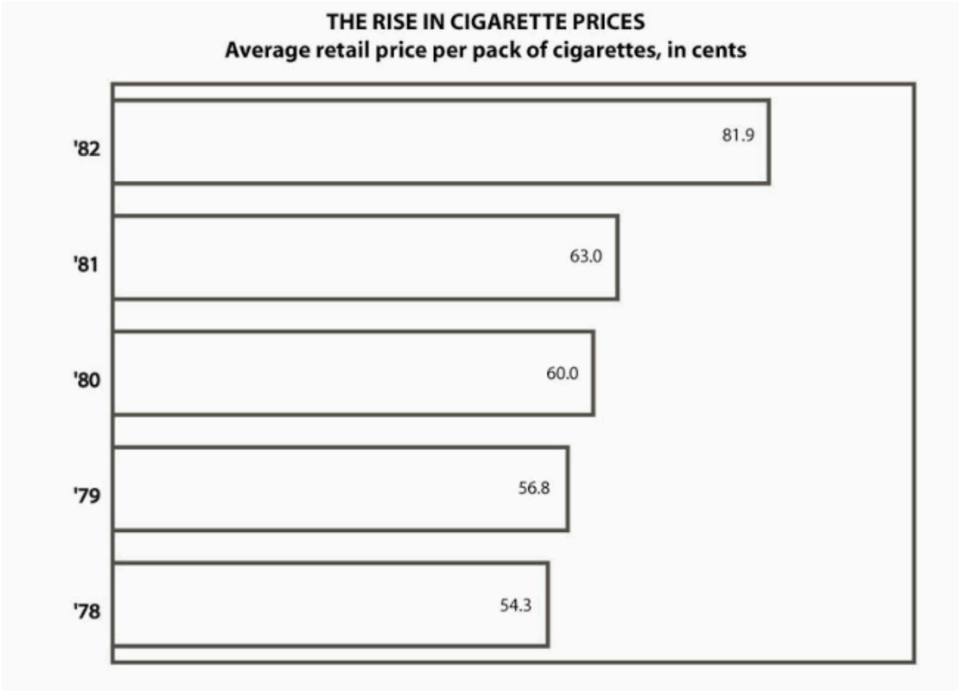




Bateman, Scott, et al. "Useful junk? The effects of visual embellishment on comprehension and memorability of charts." *CHI 2010*.

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

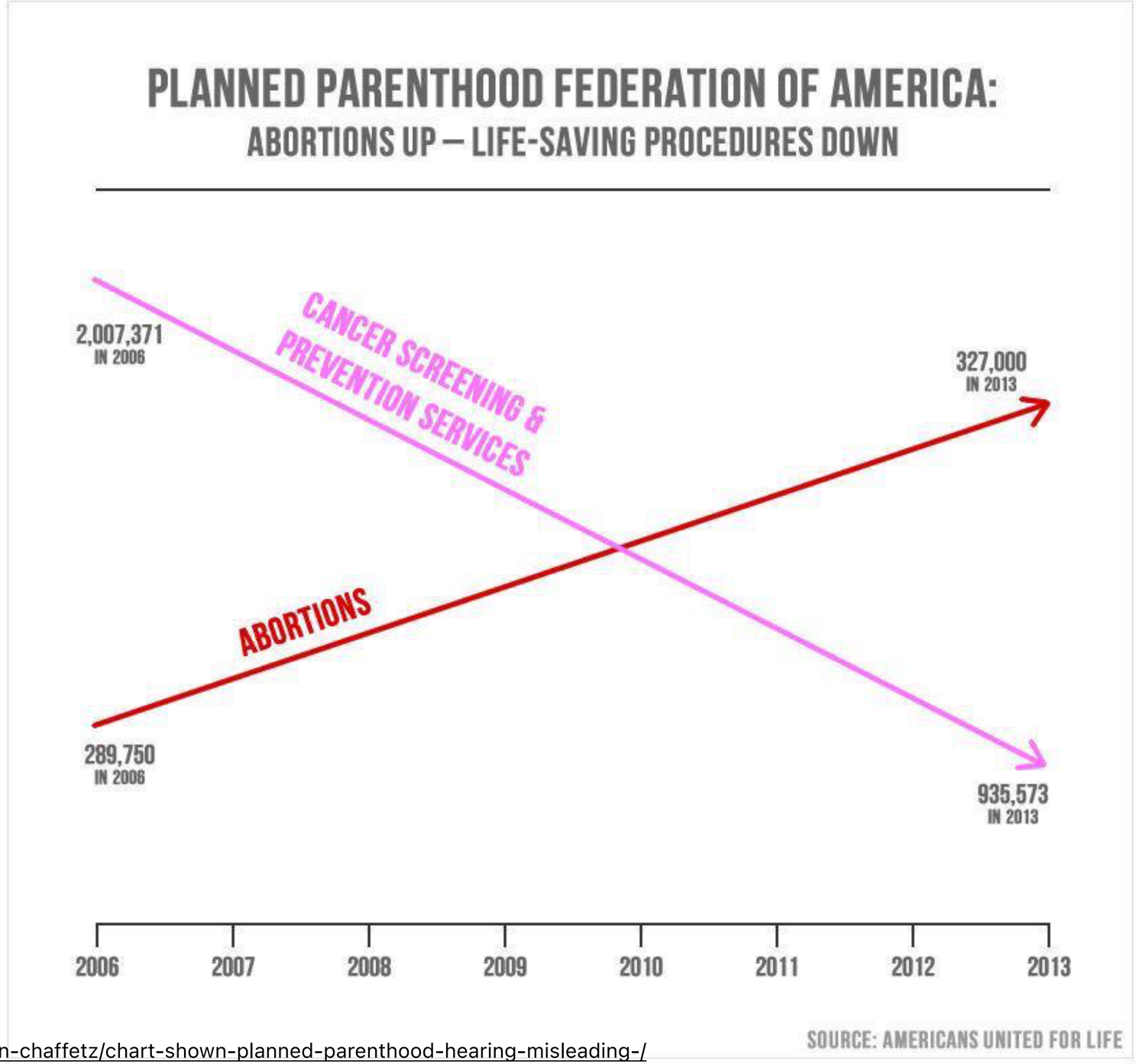
(Mis)leading the witness

Using space (in)effectively

(De-)Obfuscating data

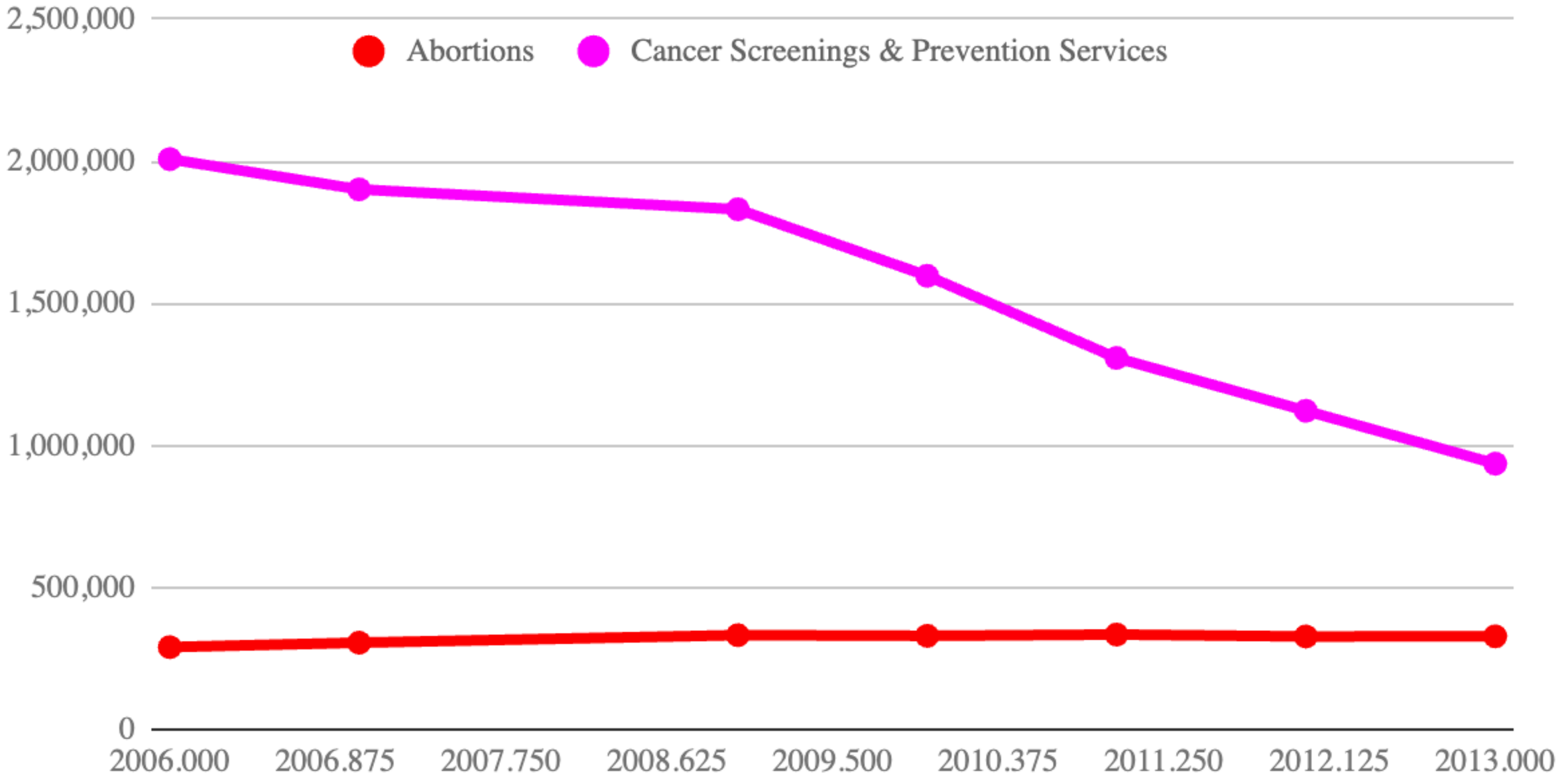
(Mis)leading the witness

What are the issues with this chart?



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

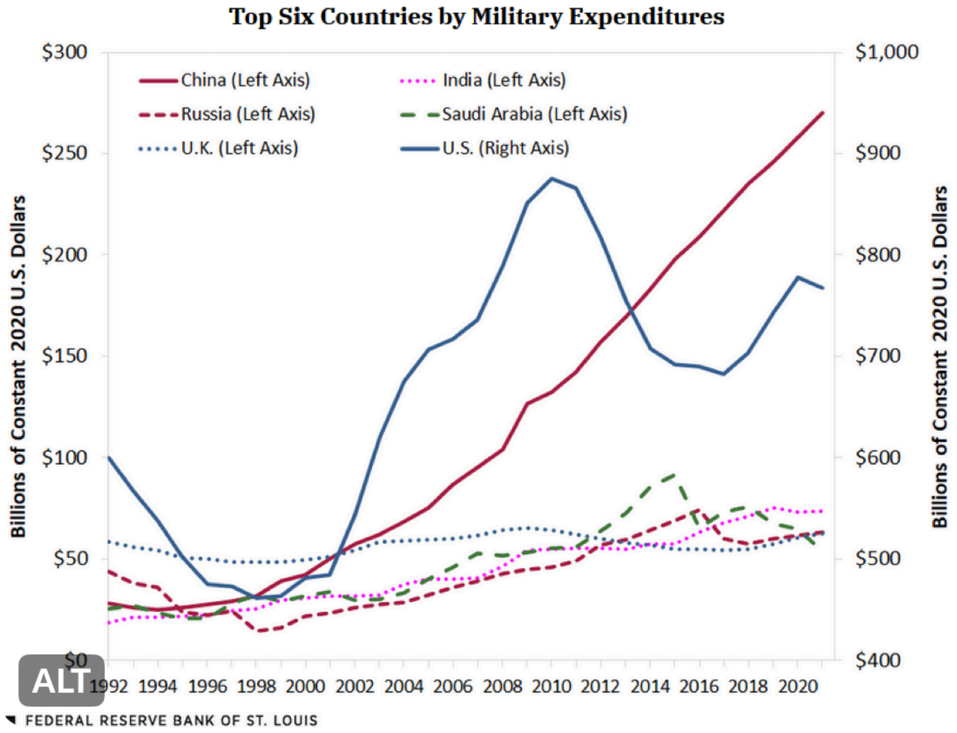
Planned Parenthood Federation of America: Abortions vs. Cancer and Prevention Services



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...

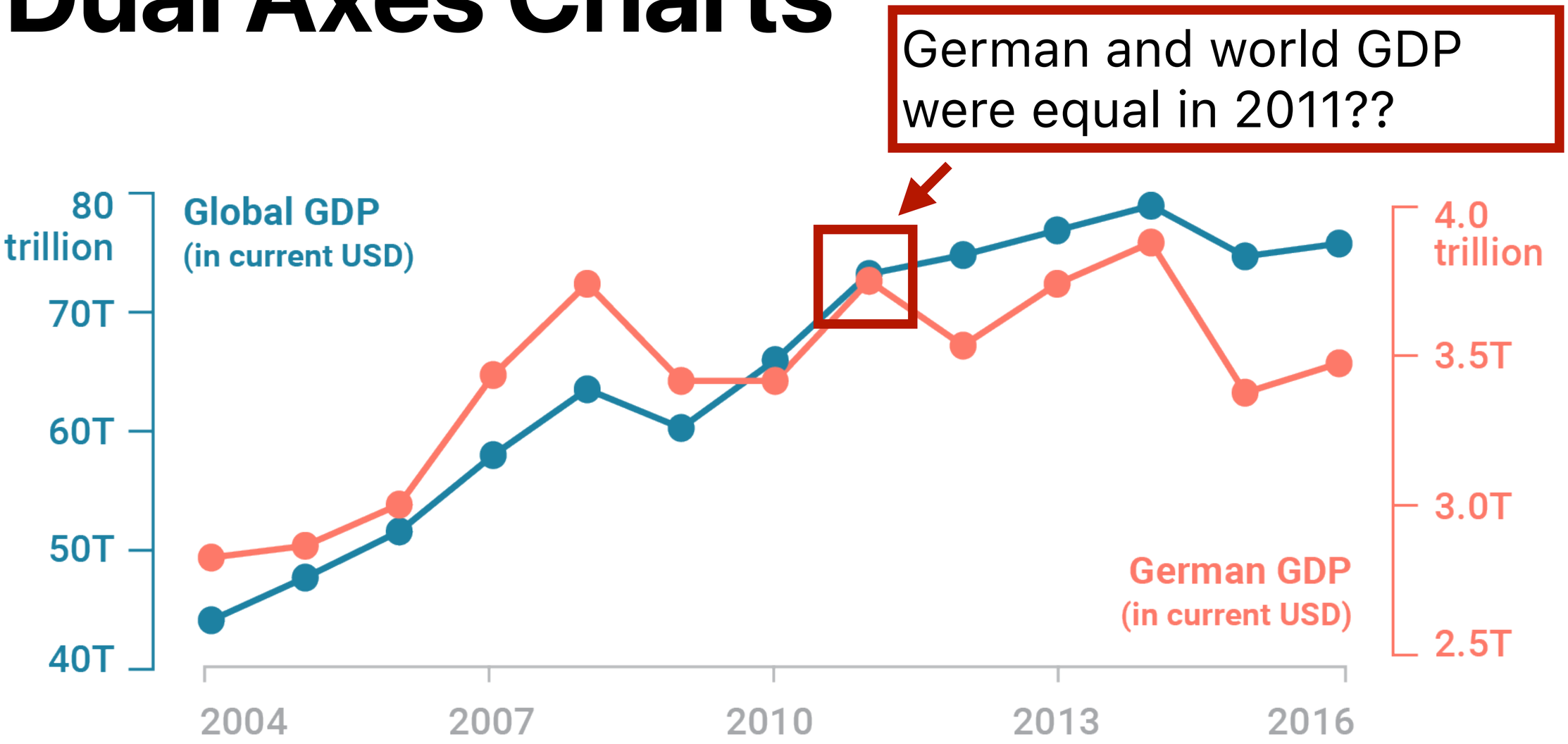
Do you find this helpful? Rate it

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

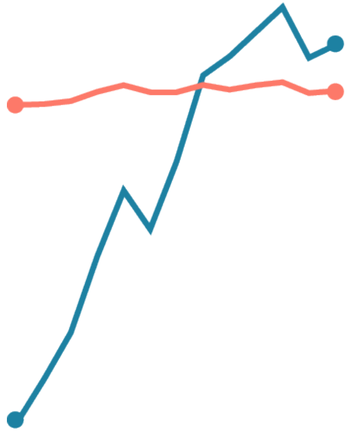
1,128 Likes **157** Retweets **2,281** Quotes

Dual Axes Charts

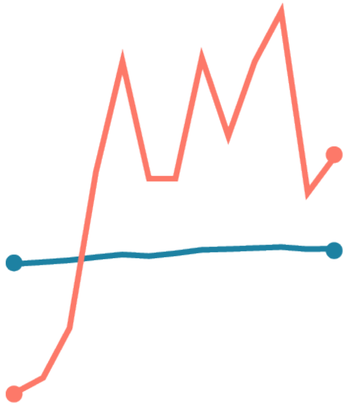


https://blog.datawrapper.de/dualaxis/

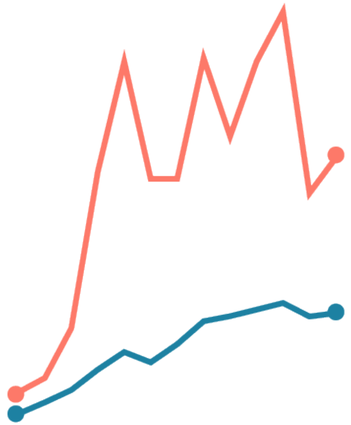
Dual-Axes Charts



Orange steady,
Blue massively increasing.



Blue steady,
Orange increasing.



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

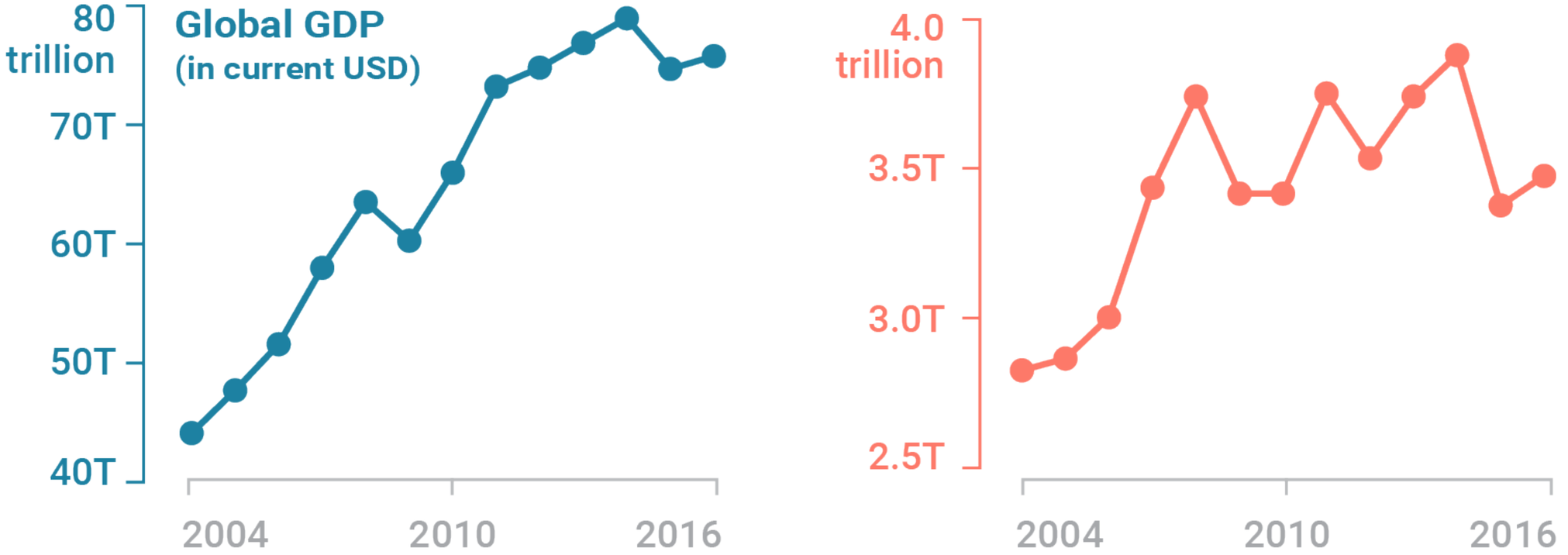
Both started at the same level, but Blue increased far more than Orange.

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