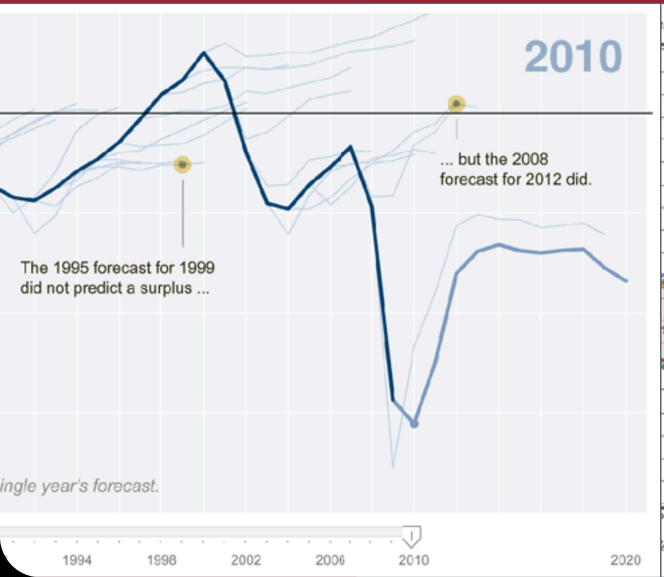
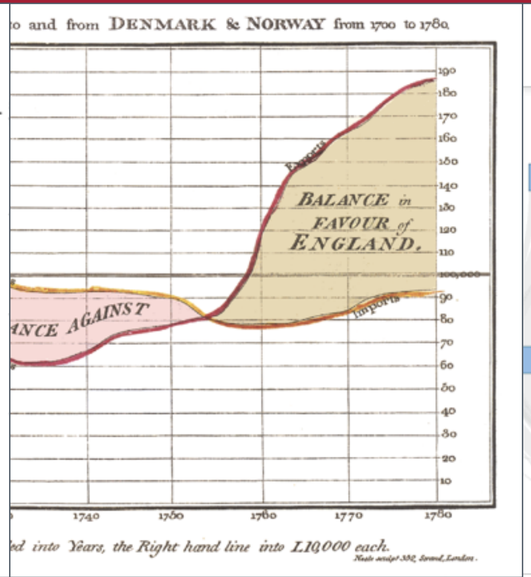
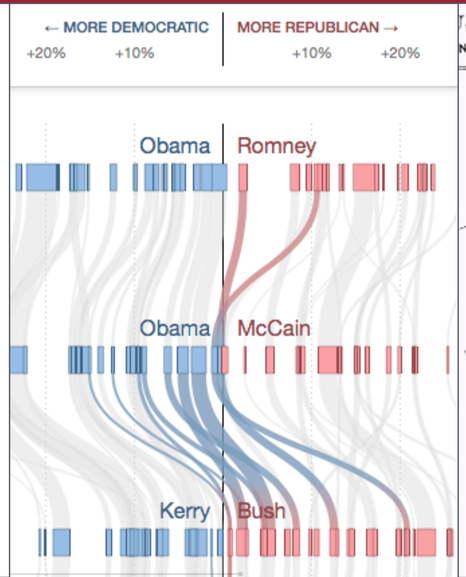
6.859: Interactive Data Visualization

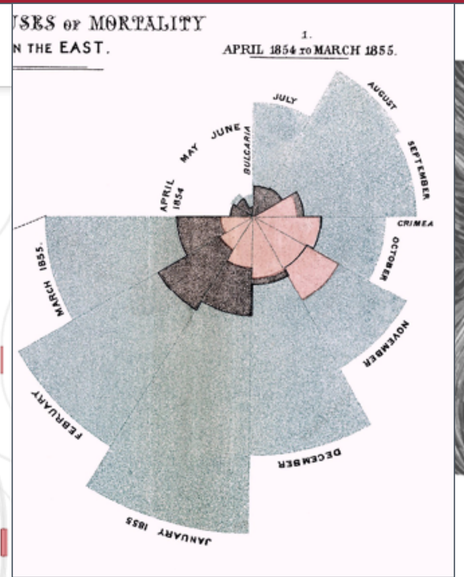
Effective Encoding & Dark Patterns

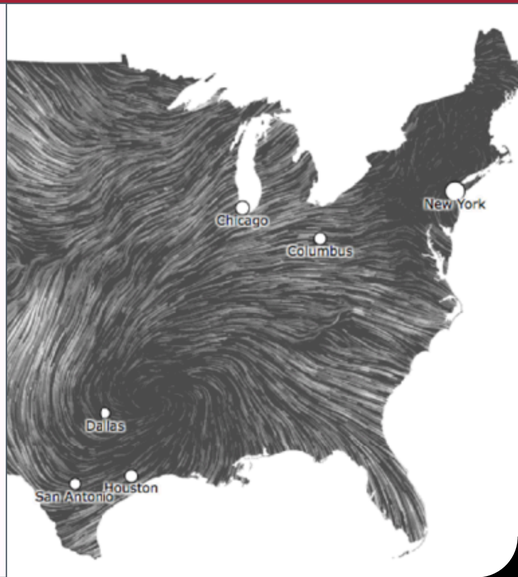
Arvind Satyanarayan











Mapping or Visual Encoding

Data → **Visual**

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*.

Data models give us a way of talking about this.

[Mackinlay 1986]

Effectiveness

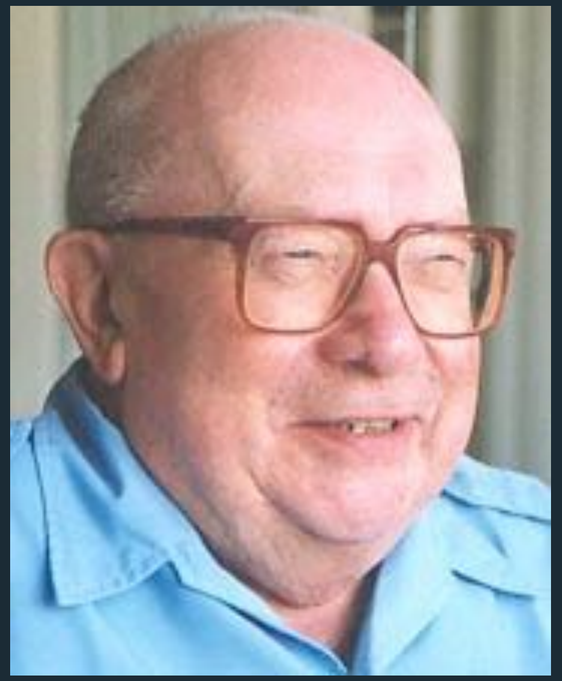
A visualization is more *effective* than another if the information it conveys *is more readily perceived* than the information in the other visualization

Image models give us a way of talking about this.

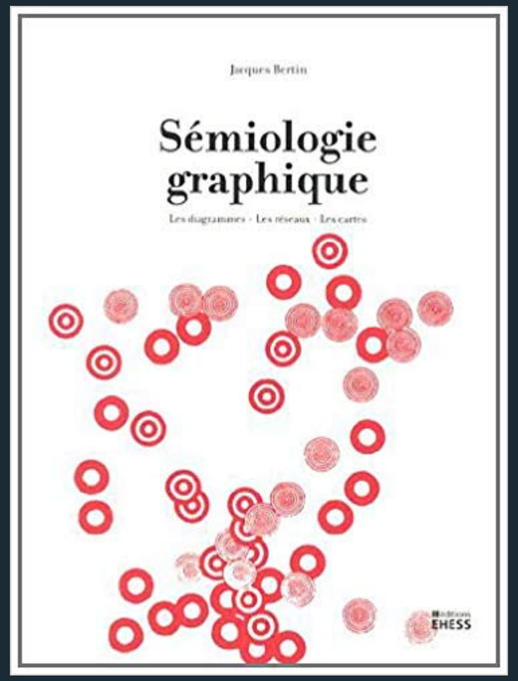
The Semiology of Graphics (1967)

*Study of **signs** and how cultures use them.*

Anything that stands for something other than itself.



Jacques Bertin (1918 – 2010)
French cartographer

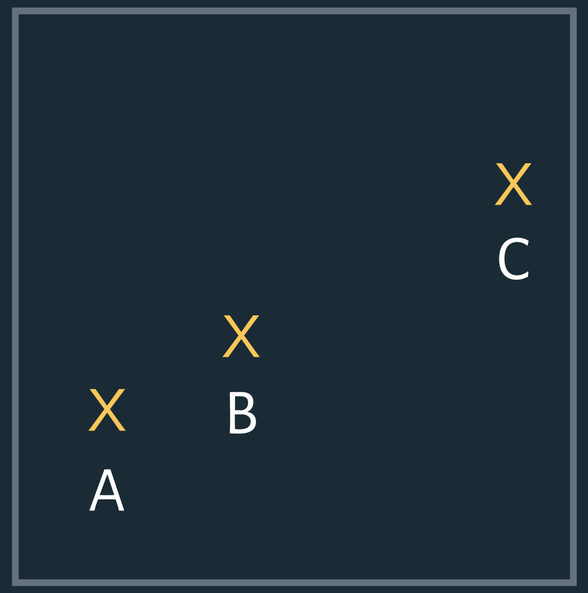


Images are perceived as a set of signs.

Sender encodes information in signs.

Through visual perception, the receiver decodes the signs for information:

1. What are the elements in question?
2. What are the relationships between them?



Sender encodes information in signs.

Through visual perception, the receiver decodes the signs for information:

1. What are the elements in question?
2. What are the relationships between them?

What do these signs signify?

1. A, B, C are distinguishable.
2. B is between A and C.
3. BC is twice as long as AB.

}

"Resemblance, order, and proportional are the three signfields in graphics."

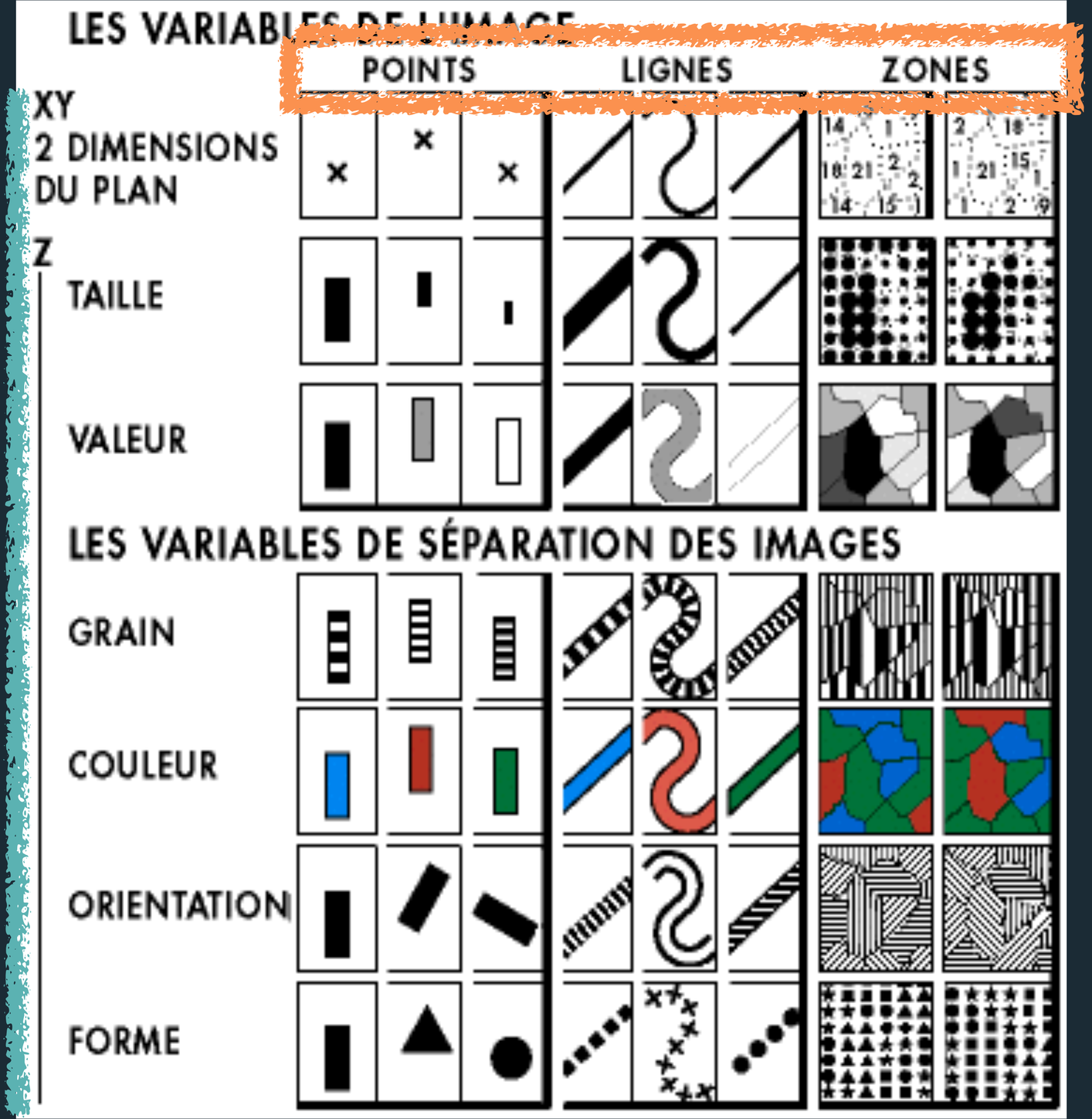
—Bertin

Visual Variables

Also called visual *channels*.

Used to encode data values as characteristics of marks.

** From 1967, so Bertin only accounted for visualizations that were printable, white paper.*



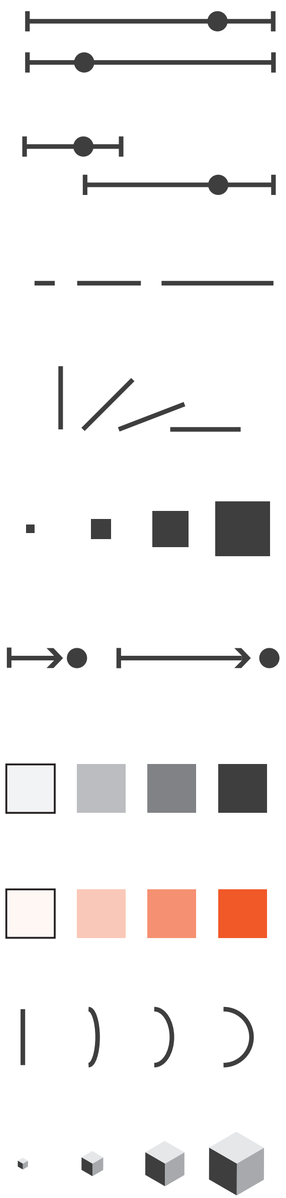
Marks

Basic graphical elements that represent data items.

Channels: Expressiveness Types and Effectiveness Ranks

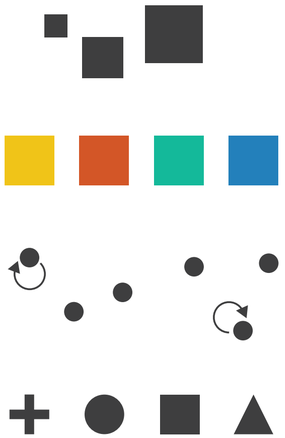
Magnitude Channels: O or Q attributes

- 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)



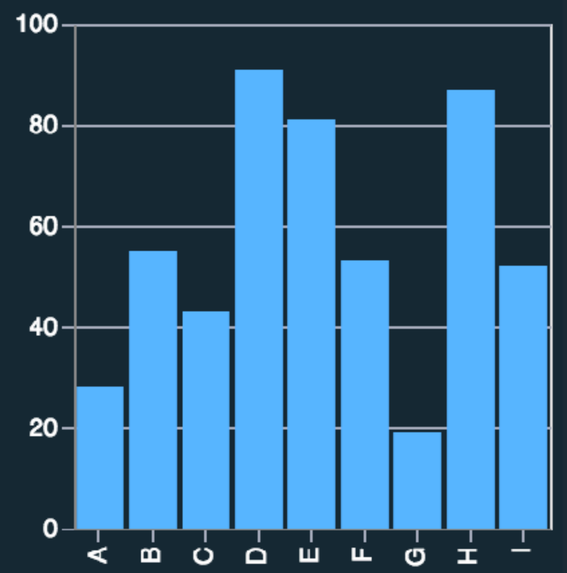
Identity Channels: N attributes

- Spatial region
- Color hue
- Motion
- Shape



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

Visual Encoding: 1 Nominal, 1 Quantitative

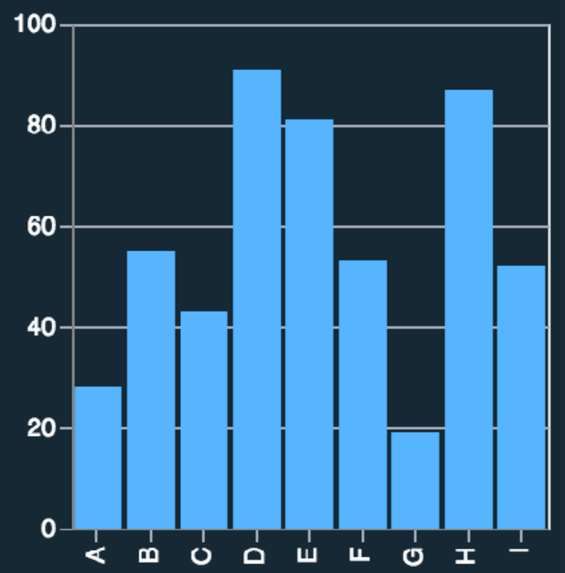


Mark: Bar

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

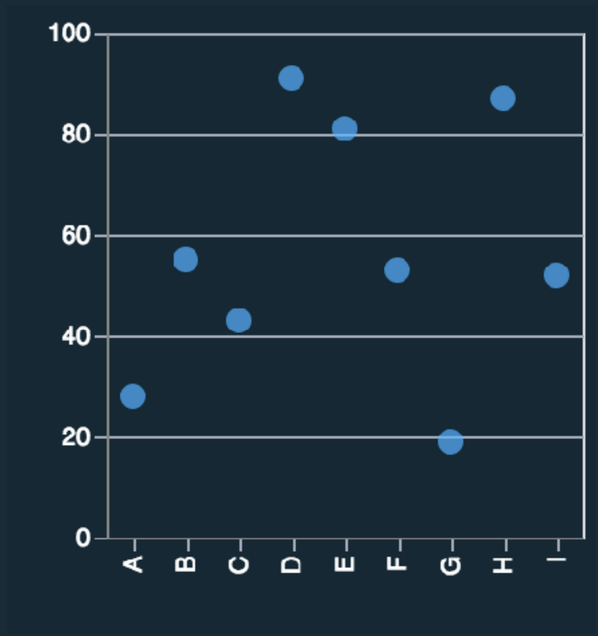
Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

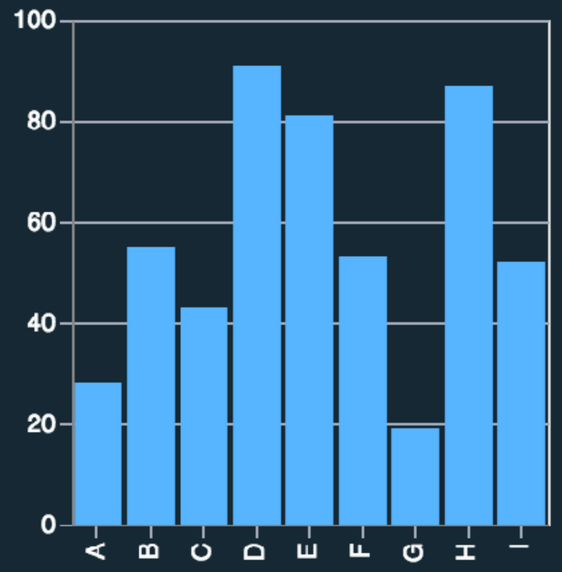


Mark: Point

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

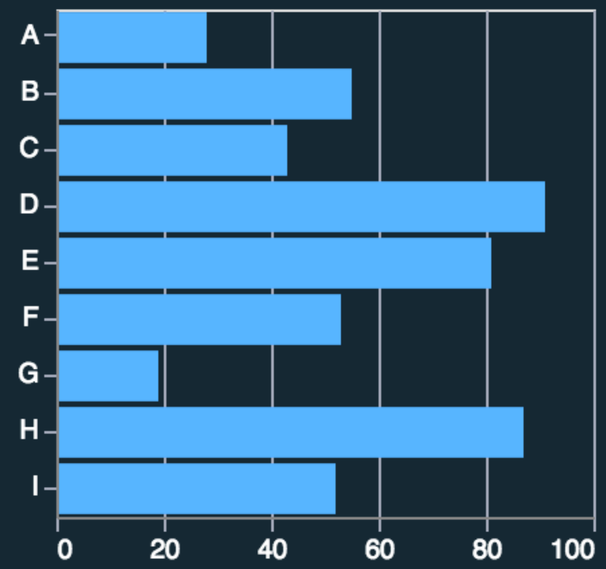
Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar

$d_{nominal} \rightarrow x$

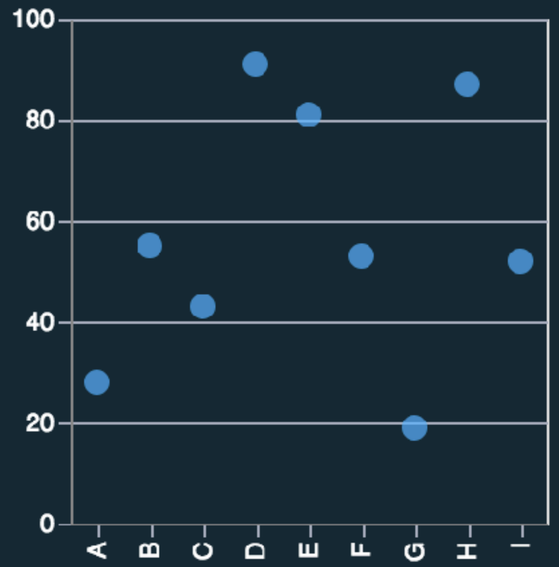
$d_{quantitative} \rightarrow y$



Mark: Bar

$d_{nominal} \rightarrow y$

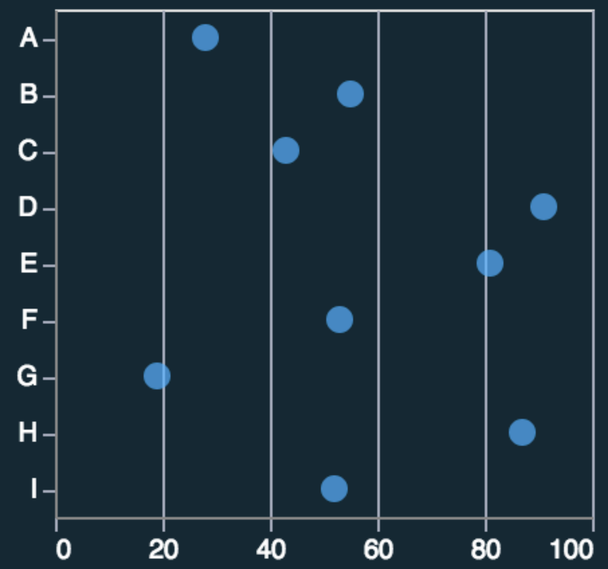
$d_{quantitative} \rightarrow x$



Mark: Point

$d_{nominal} \rightarrow x$

$d_{quantitative} \rightarrow y$

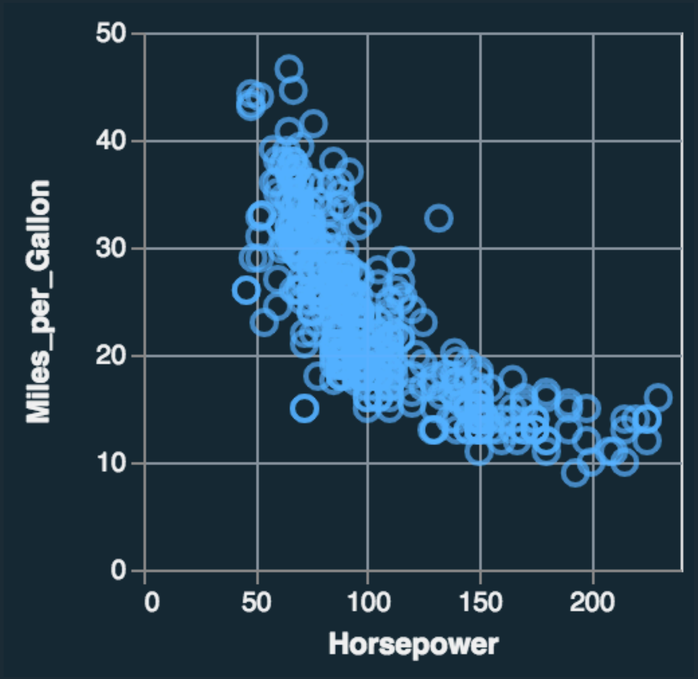


Mark: Point

$d_{nominal} \rightarrow y$

$d_{quantitative} \rightarrow x$

Visual Encoding

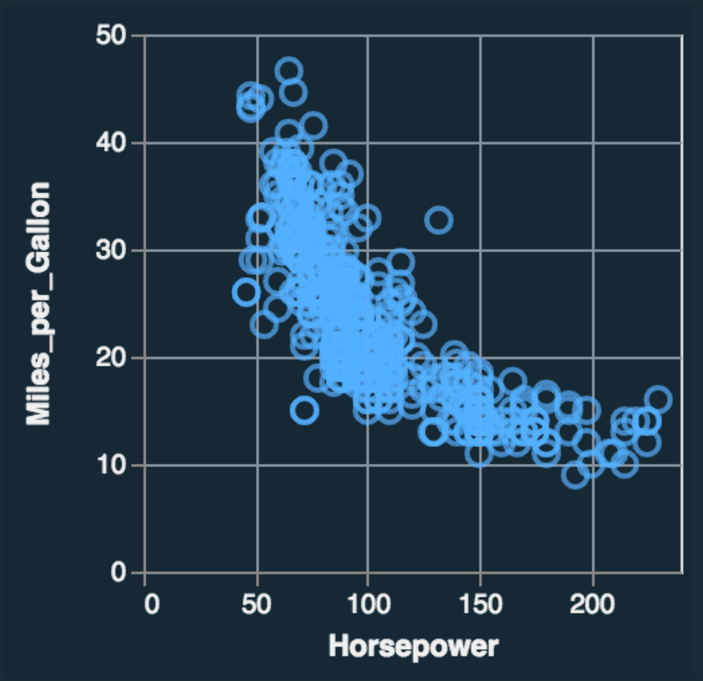


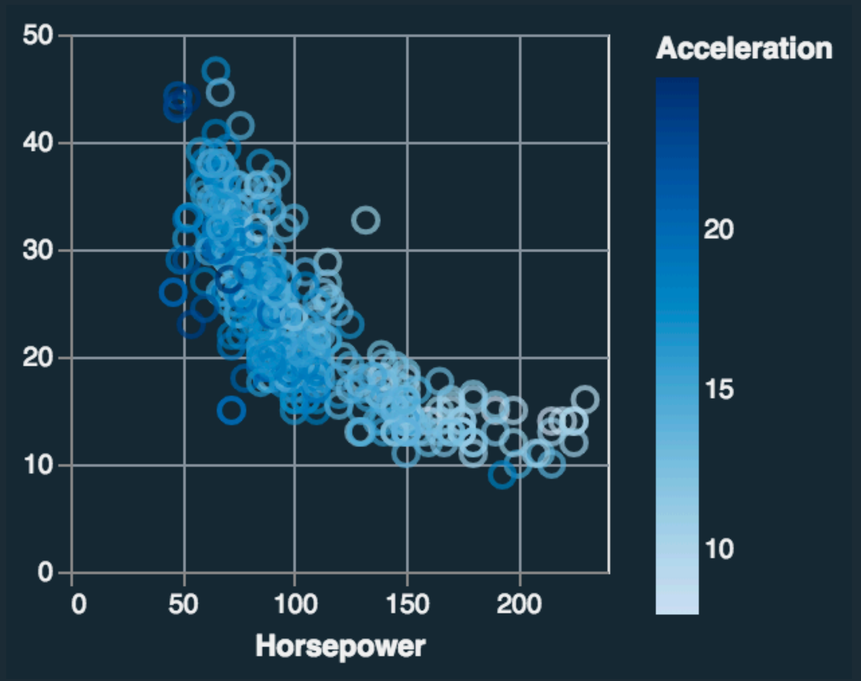
Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

Visual Encoding





Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

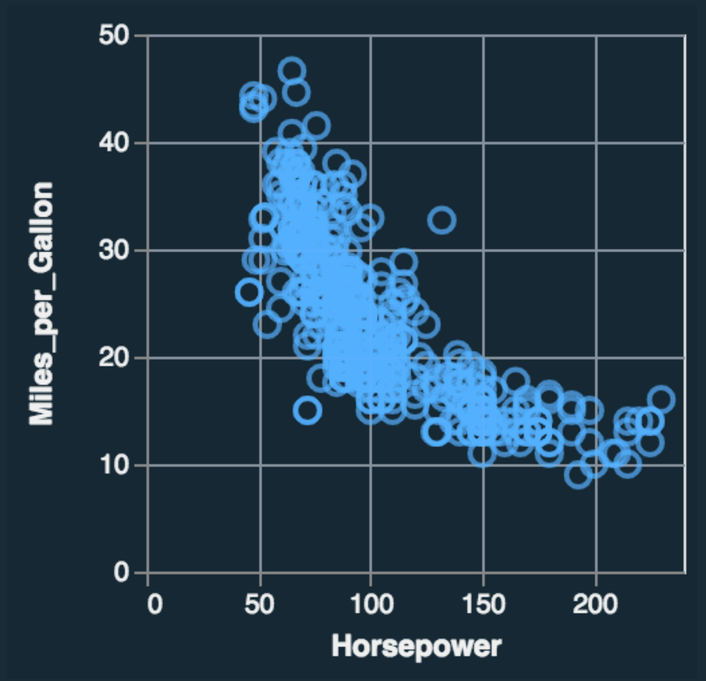
Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{quantitative} \rightarrow color$

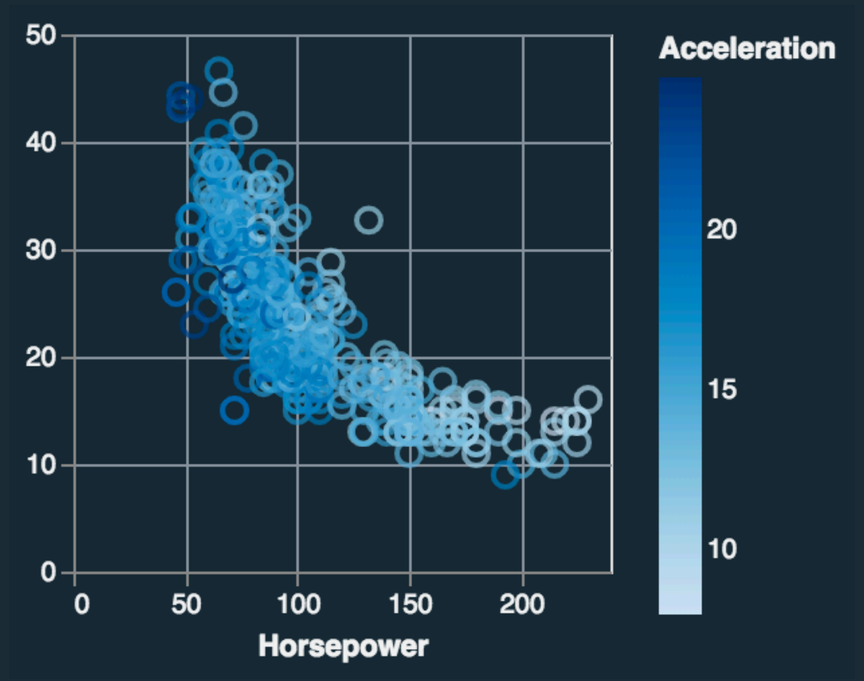
Visual Encoding



Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

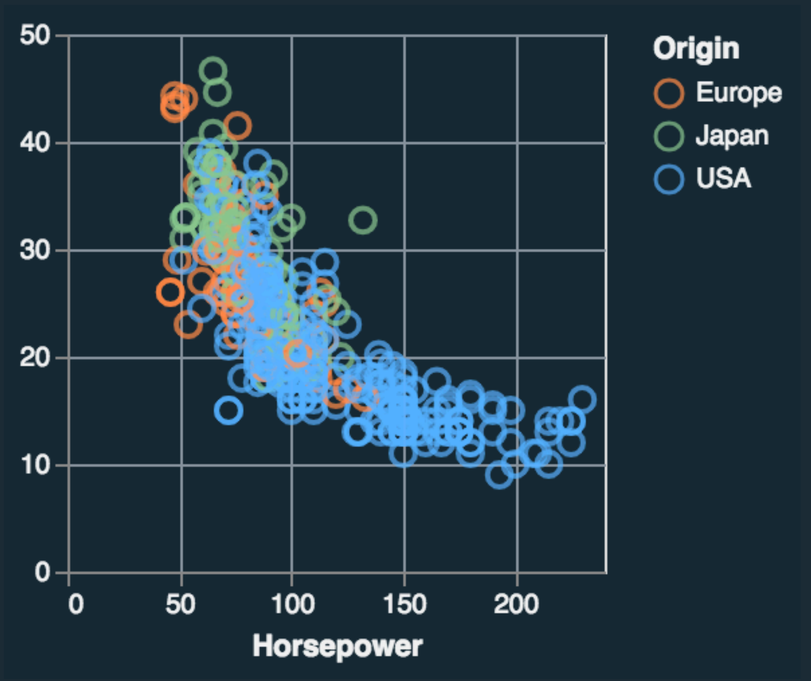


Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{quantitative} \rightarrow color$



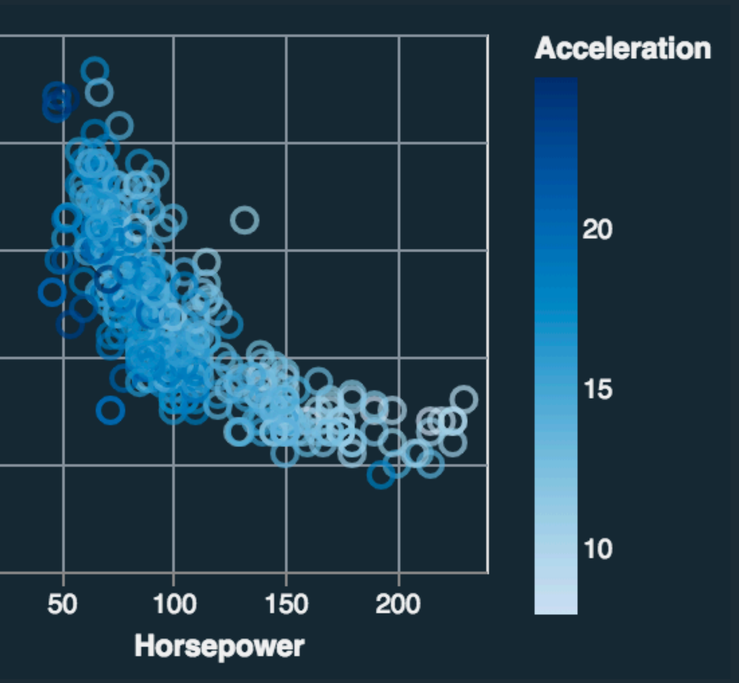
Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{nominal} \rightarrow color$

Visual Encoding

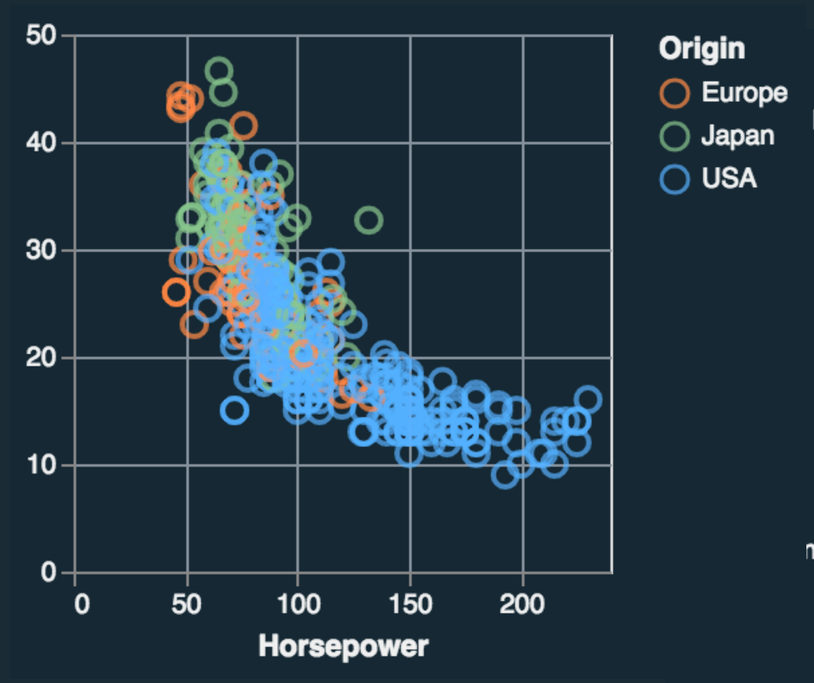


Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{quantitative} \rightarrow color$

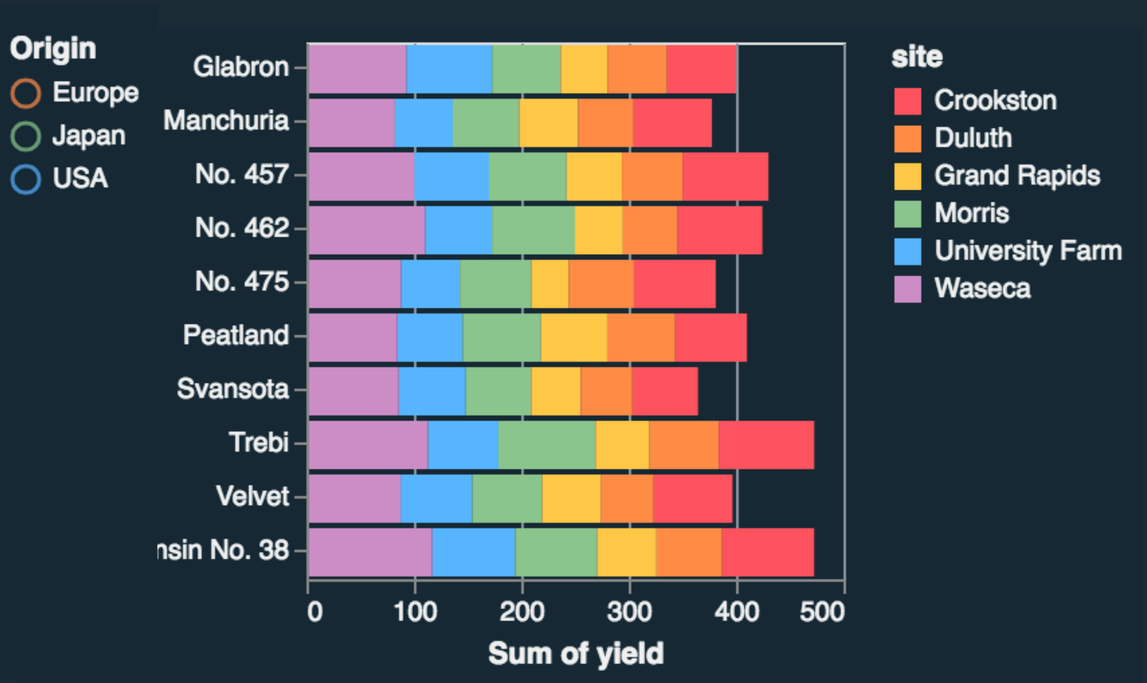


Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{nominal} \rightarrow color$



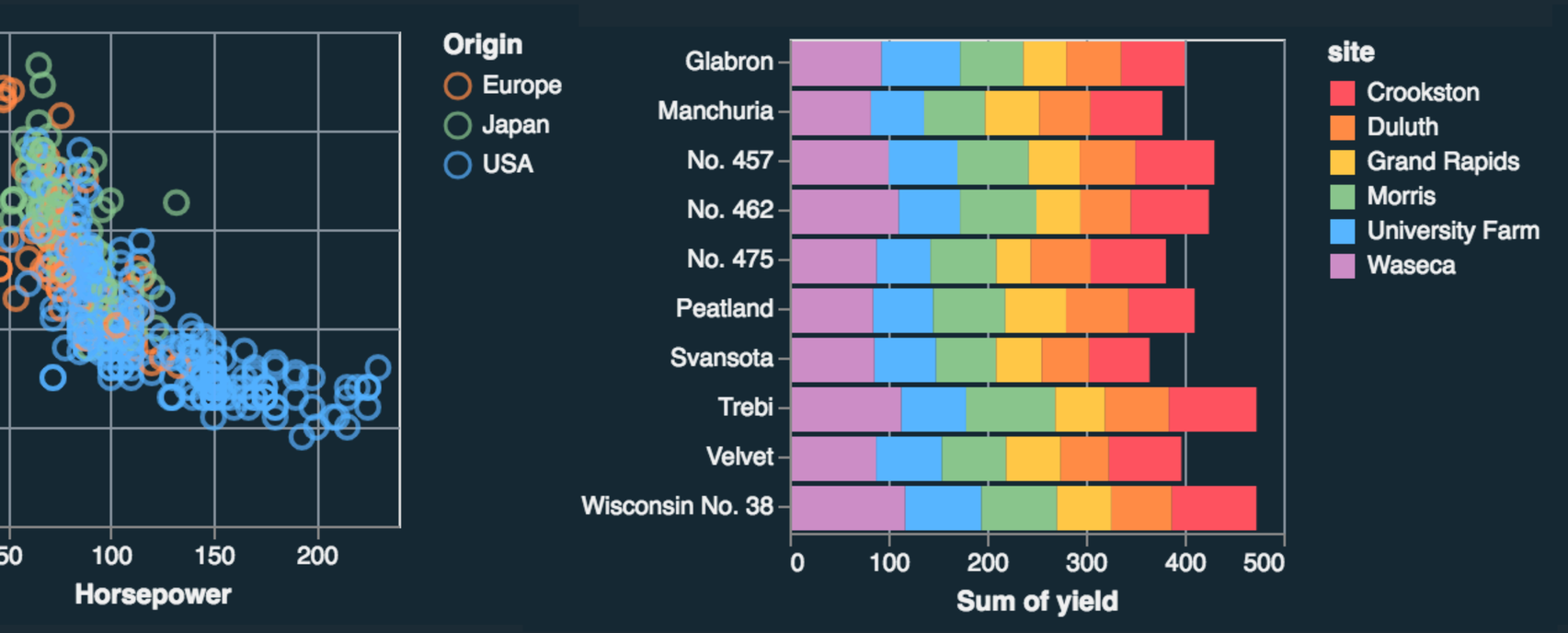
Mark: Bar

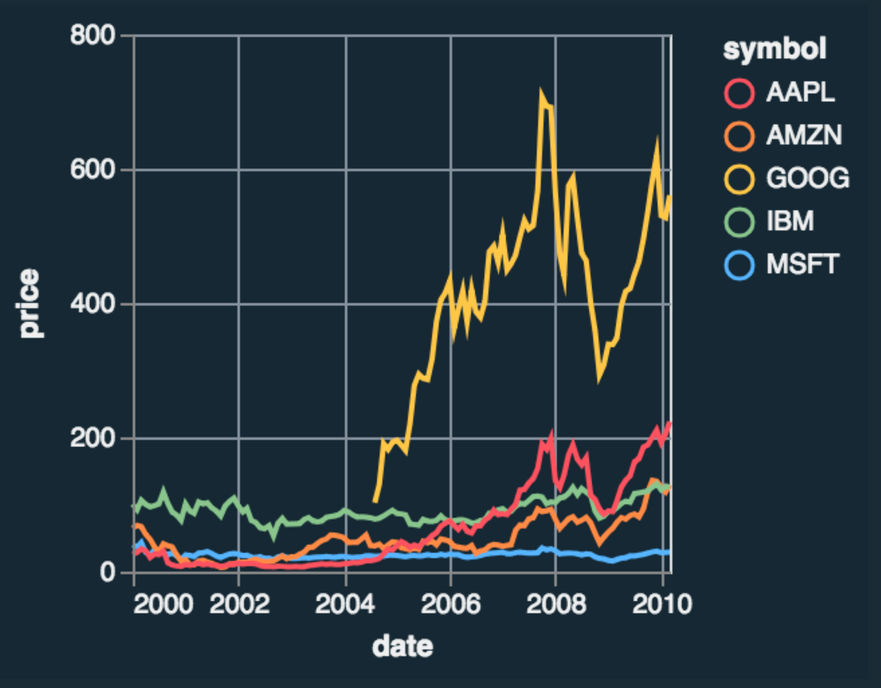
$d_{quantitative} \rightarrow x$

$d_{nominal} \rightarrow y$

$d_{nominal} \rightarrow color$

Visual Encoding





Mark: Point

$d_{quantitative} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{nominal} \rightarrow color$

Mark: Bar

$d_{quantitative} \rightarrow x$

$d_{nominal} \rightarrow y$

$d_{nominal} \rightarrow color$

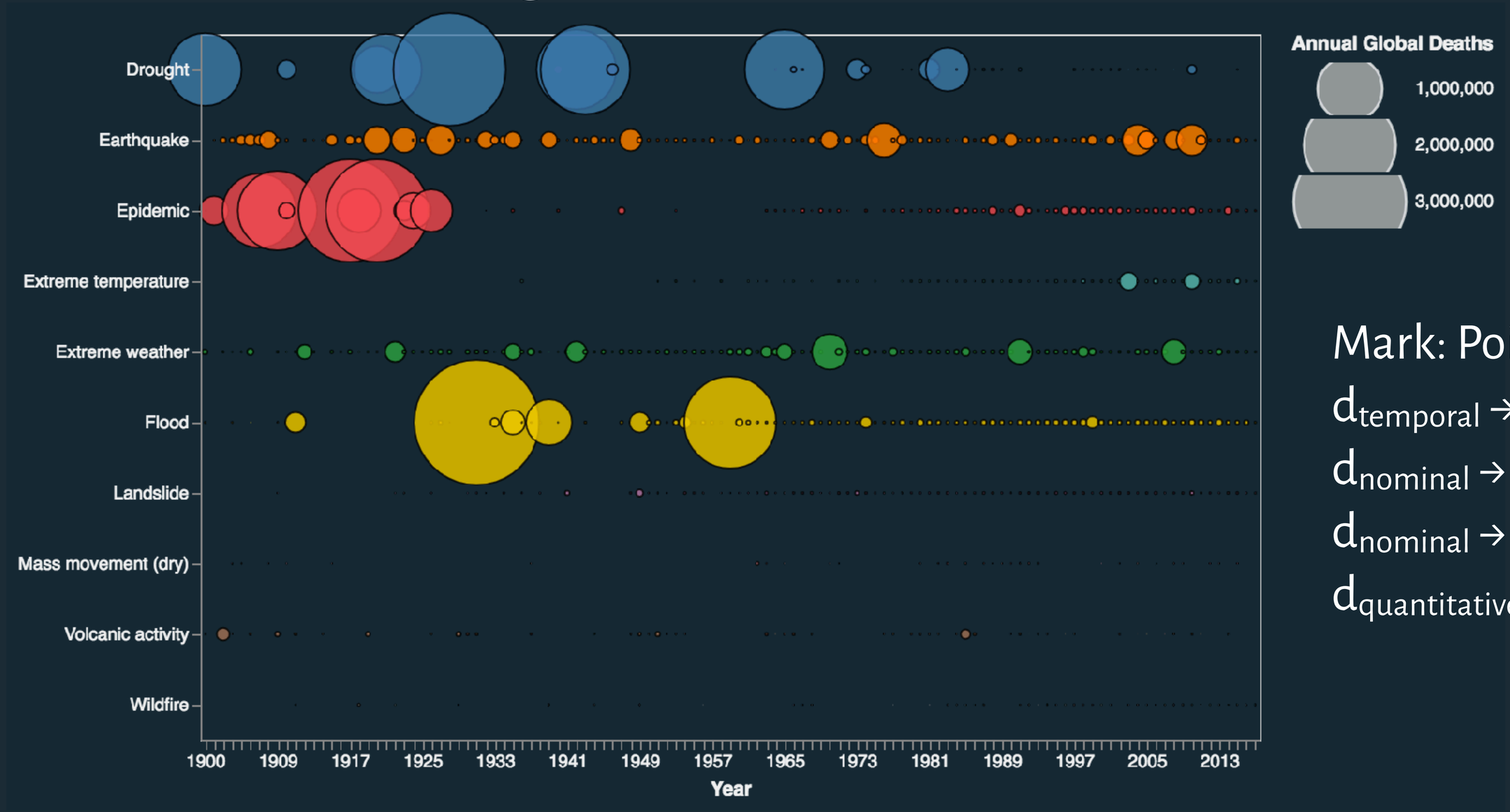
Mark: Line

$d_{temporal} \rightarrow x$

$d_{quantitative} \rightarrow y$

$d_{nominal} \rightarrow color$

Visual Encoding



Mark: Point

$d_{temporal} \rightarrow x$

$d_{nominal} \rightarrow y$

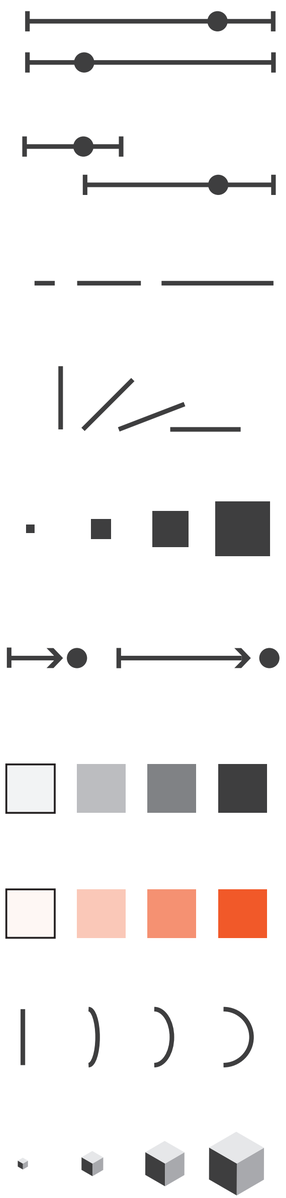
$d_{nominal} \rightarrow$ color

$d_{quantitative} \rightarrow$ size

Channels: Expressiveness Types and Effectiveness Ranks

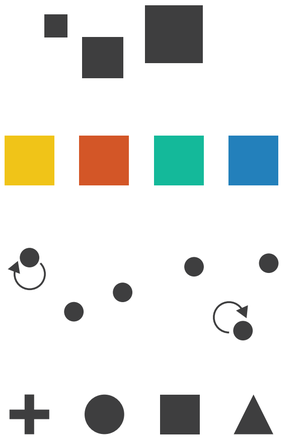
Magnitude Channels: O or Q attributes

- 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)



Identity Channels: N attributes

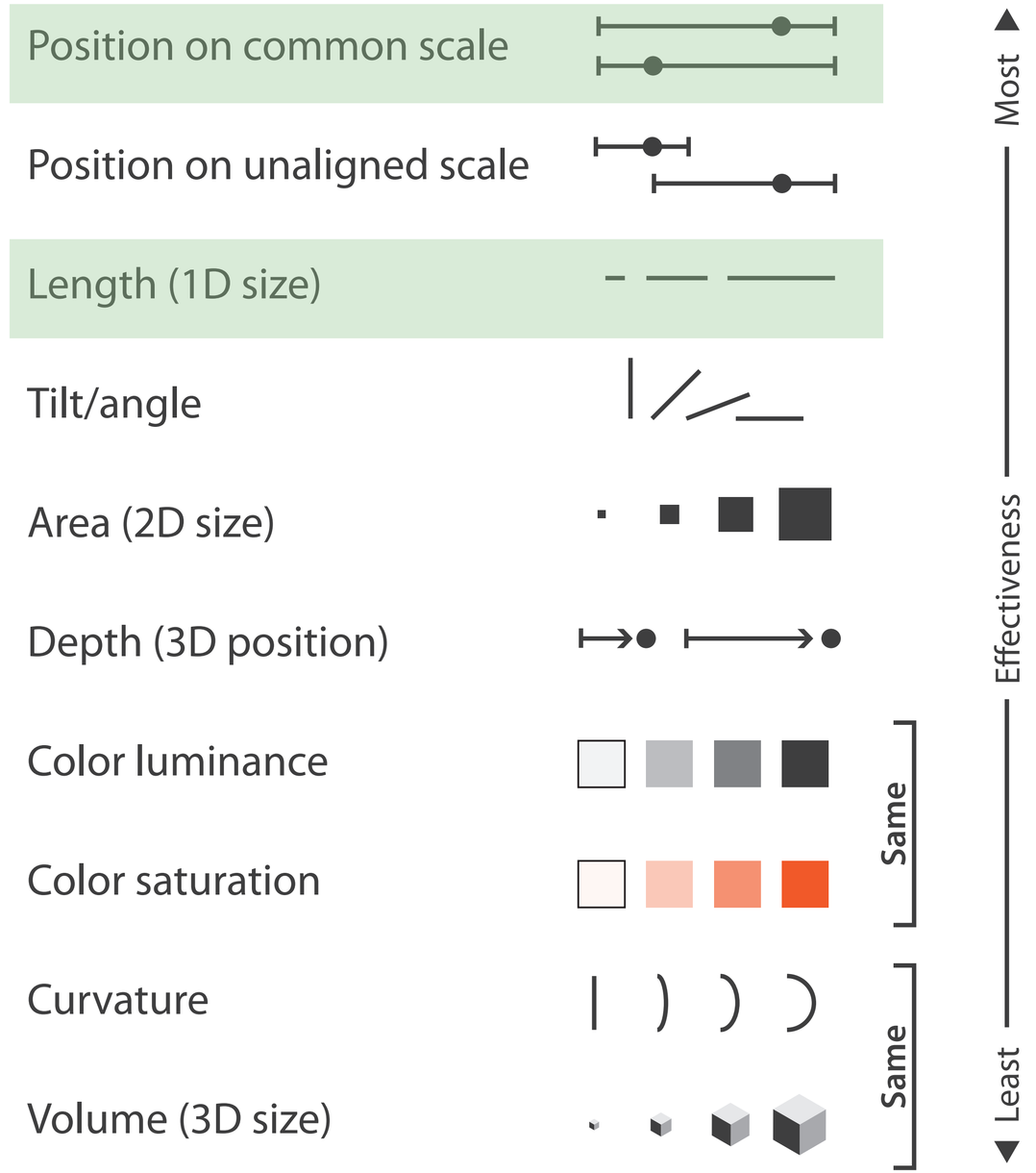
- Spatial region
- Color hue
- Motion
- Shape

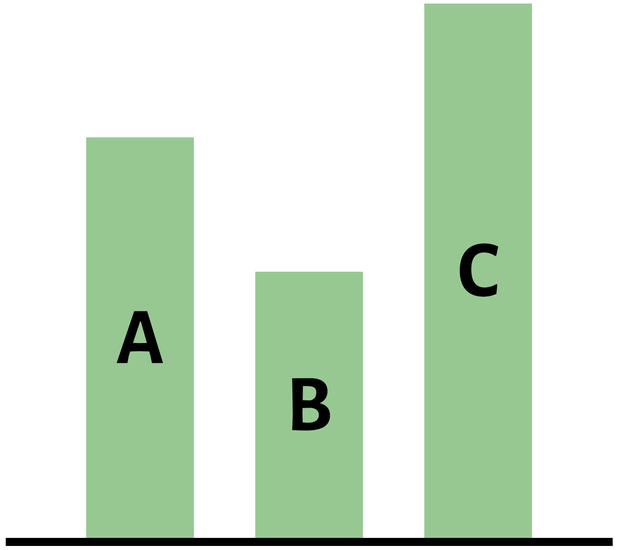


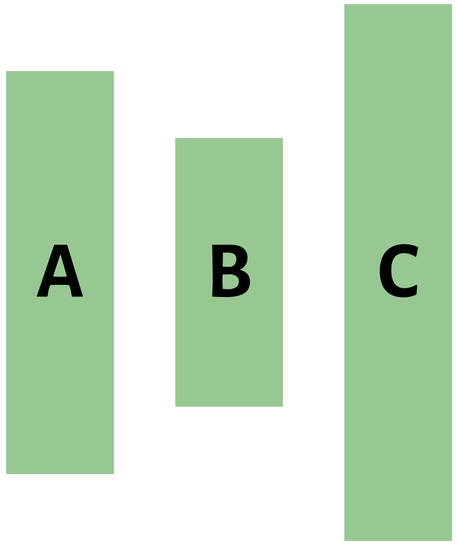
Tamara Munzner, *Visualization Analysis and Design* (2014).

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: O or Q attributes







Channels: Expressiveness Types and Effectiveness Ranks

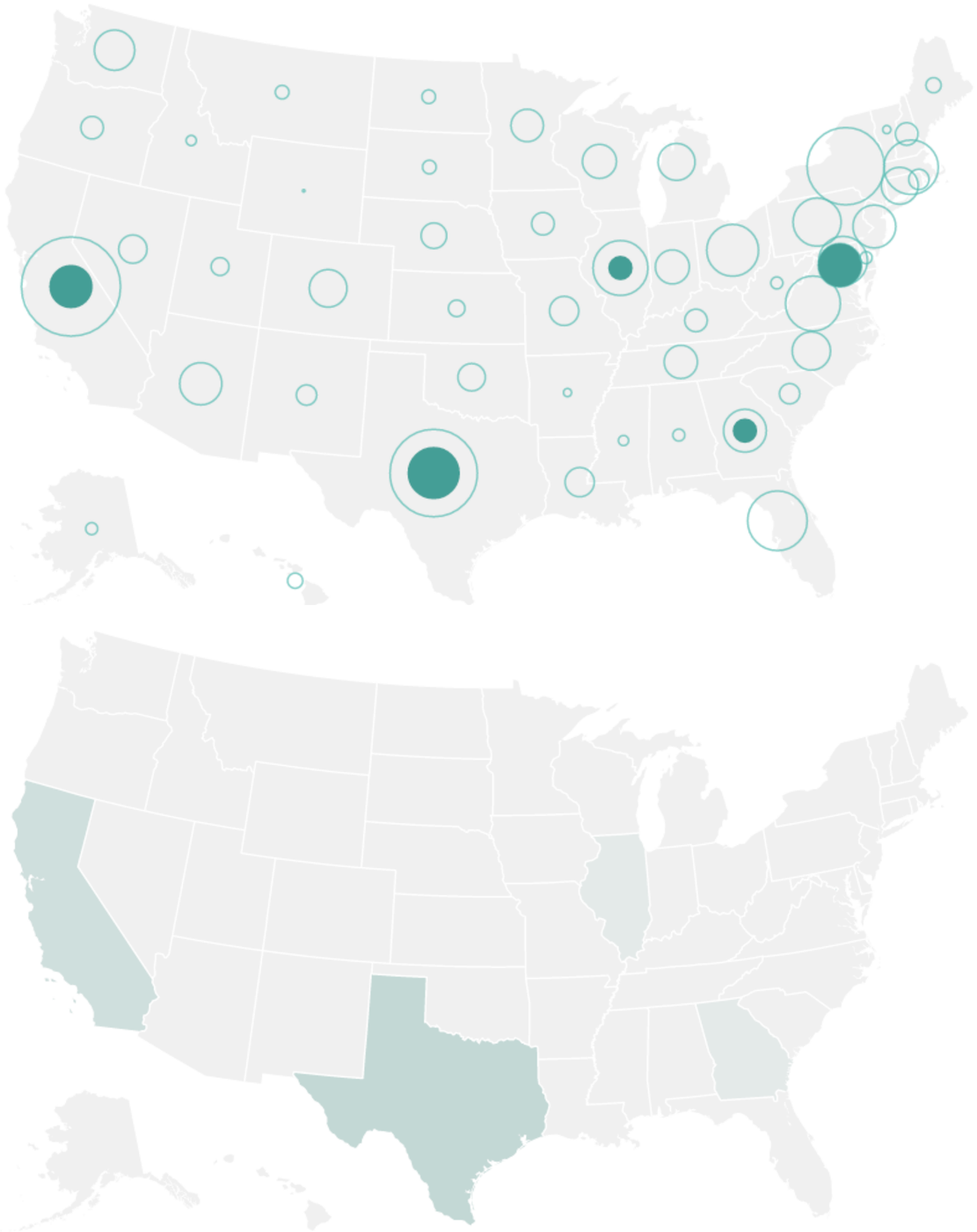
Magnitude Channels: O or Q attributes

- 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

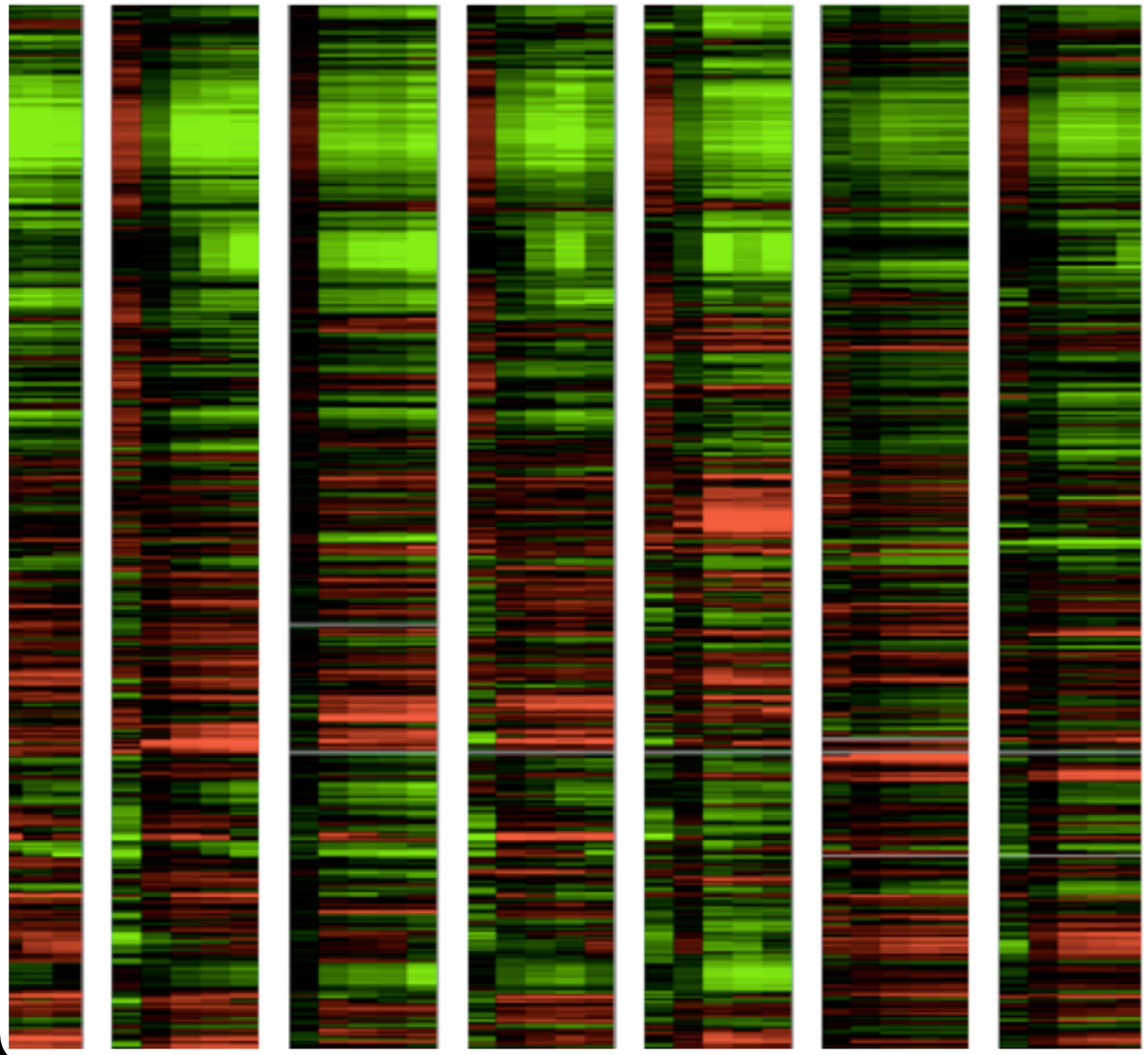
Same

Same

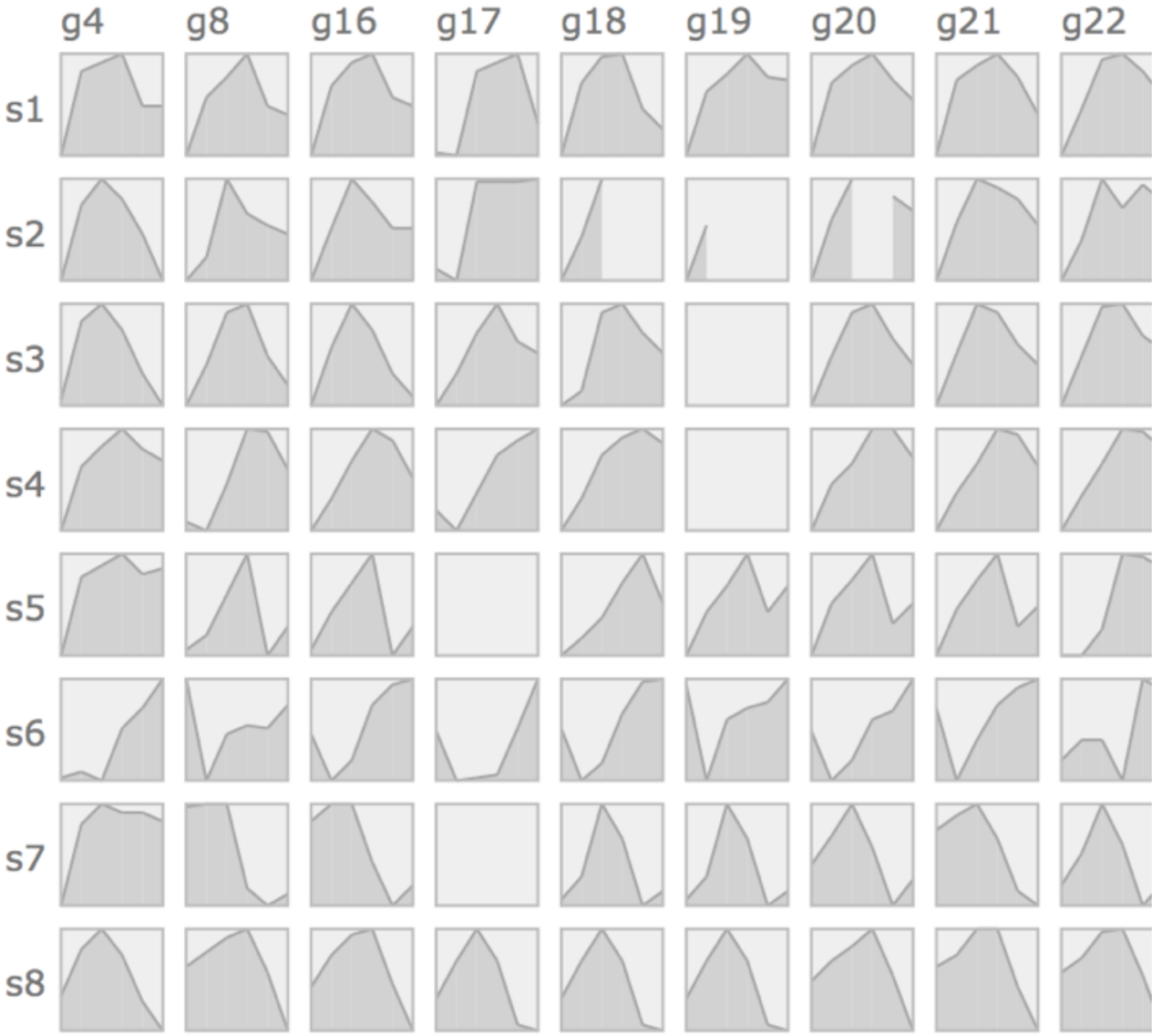


Gene Expression Time-Series [Meyer et al. '10]

Color Encoding



Position Encoding



Artery Visualization [Borkin et al ’11]

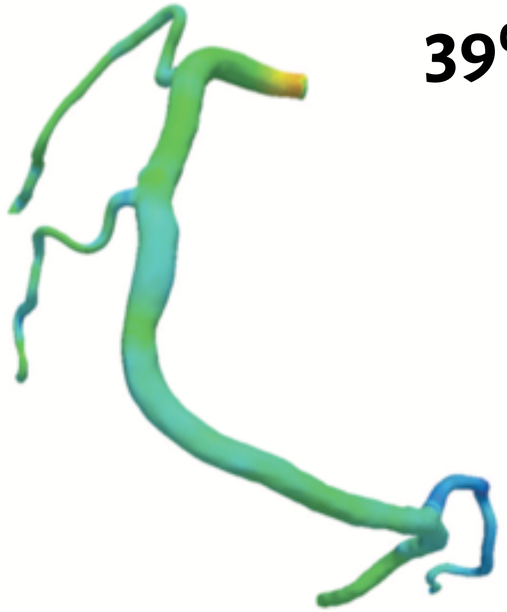
Rainbow Palette

62%

2D

39%

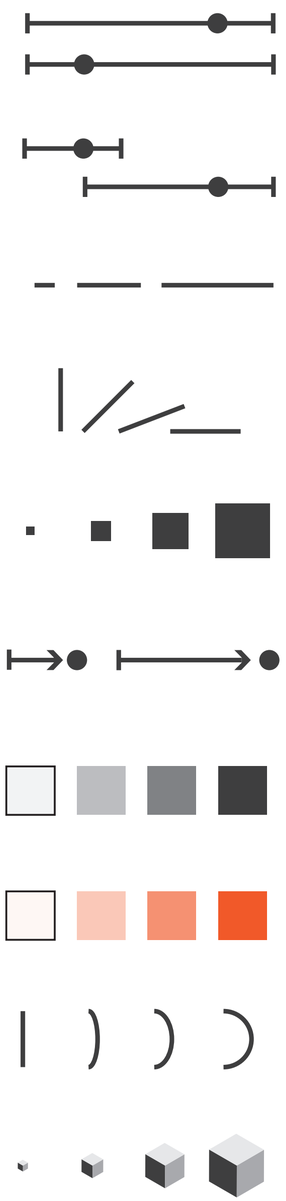
3D



Channels: Expressiveness Types and Effectiveness Ranks

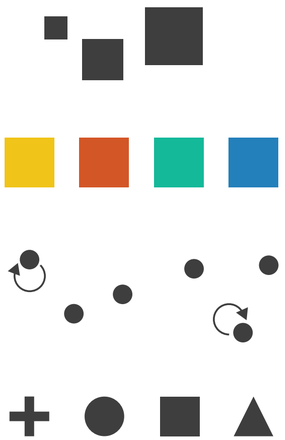
Magnitude Channels: O or Q attributes

- 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)



Identity Channels: N attributes

- Spatial region
- Color hue
- Motion
- Shape



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

Using space (in)effectively

(De-)Obfuscating data

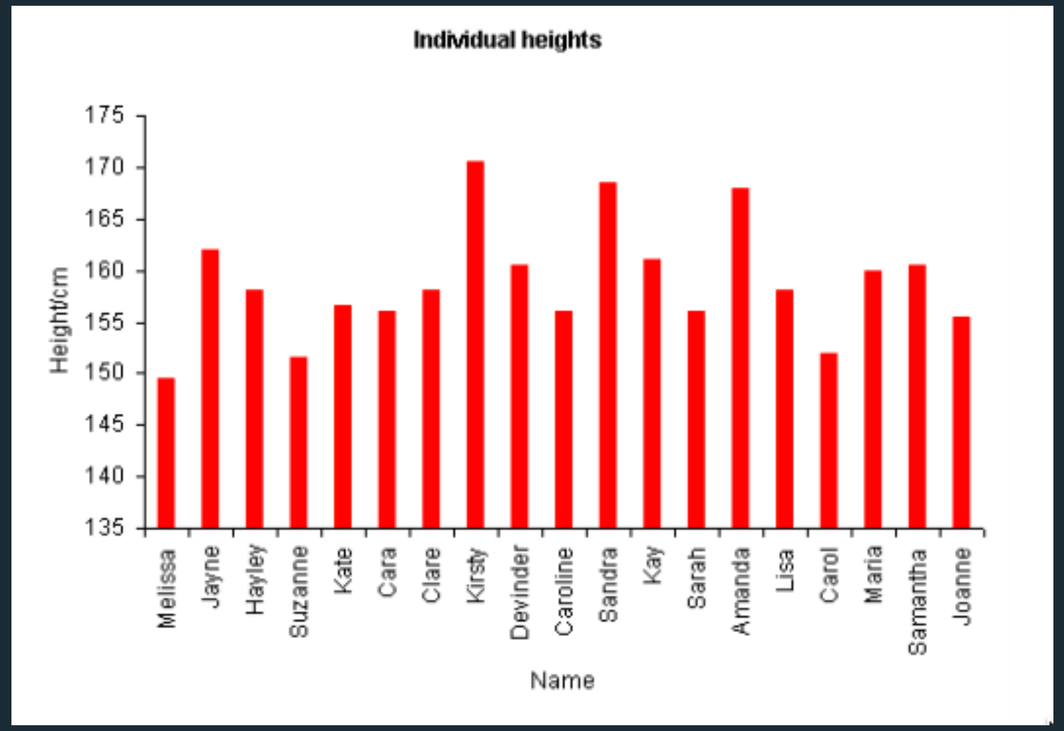
(Mis)leading the witness

Using space (in)effectively

(De-)Obfuscating data

(Mis)leading the witness

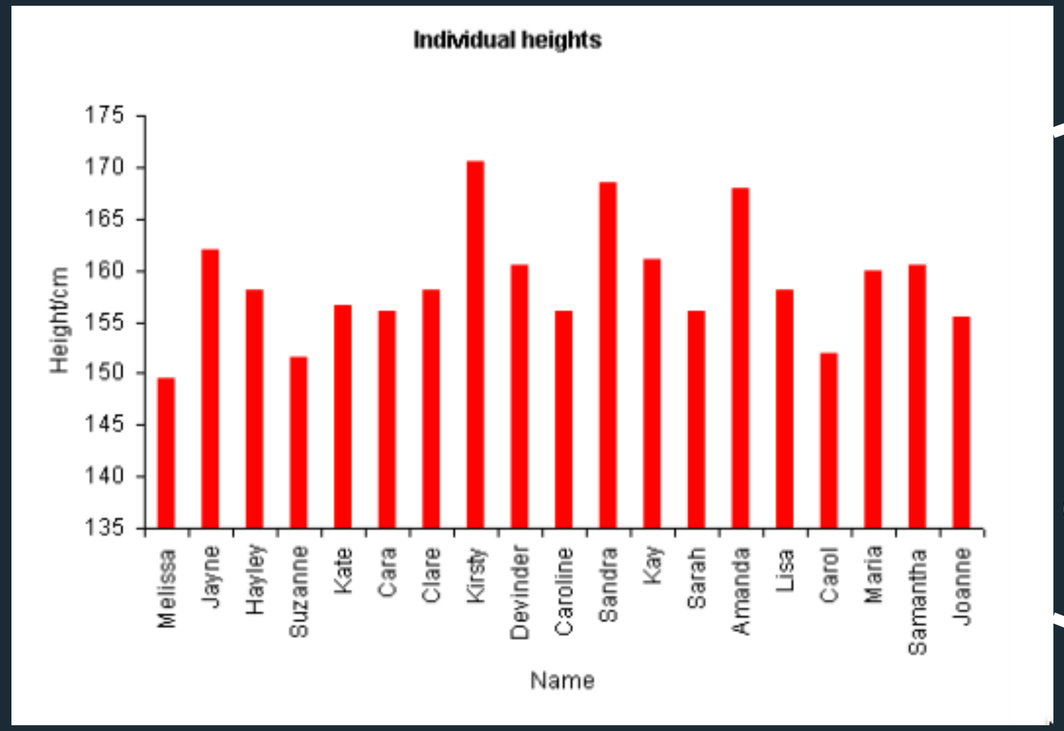
Why might this visualization be inexpressive?



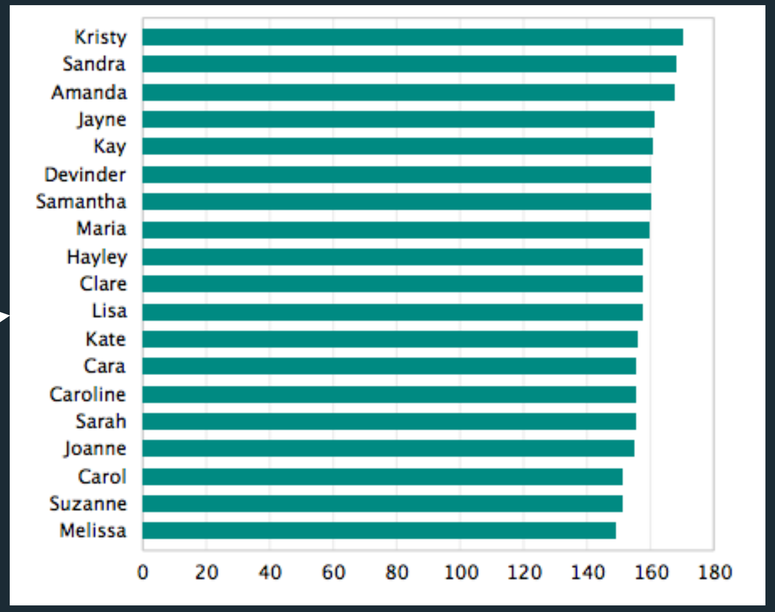
Raise your hand

Post in the chat

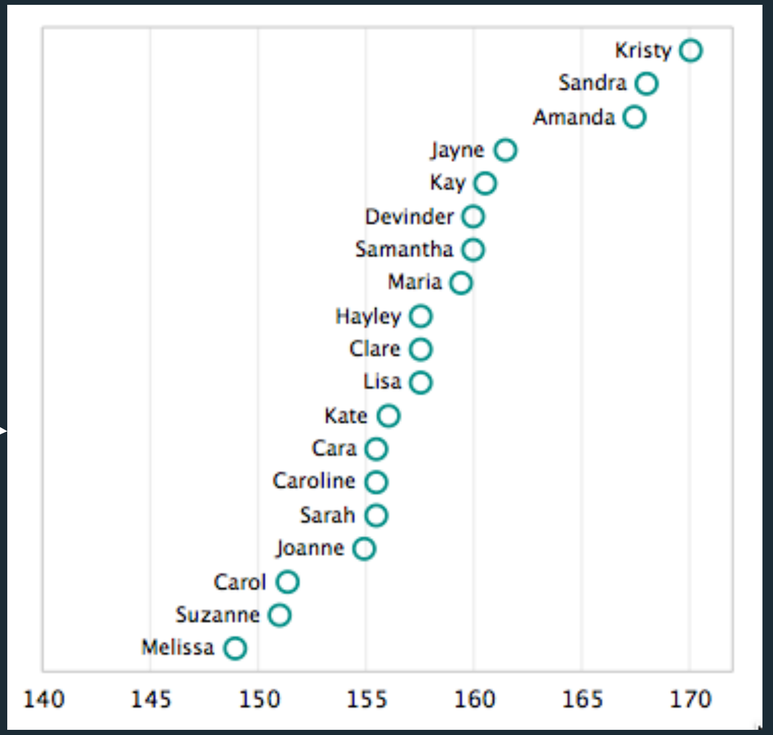
Zero Baseline?



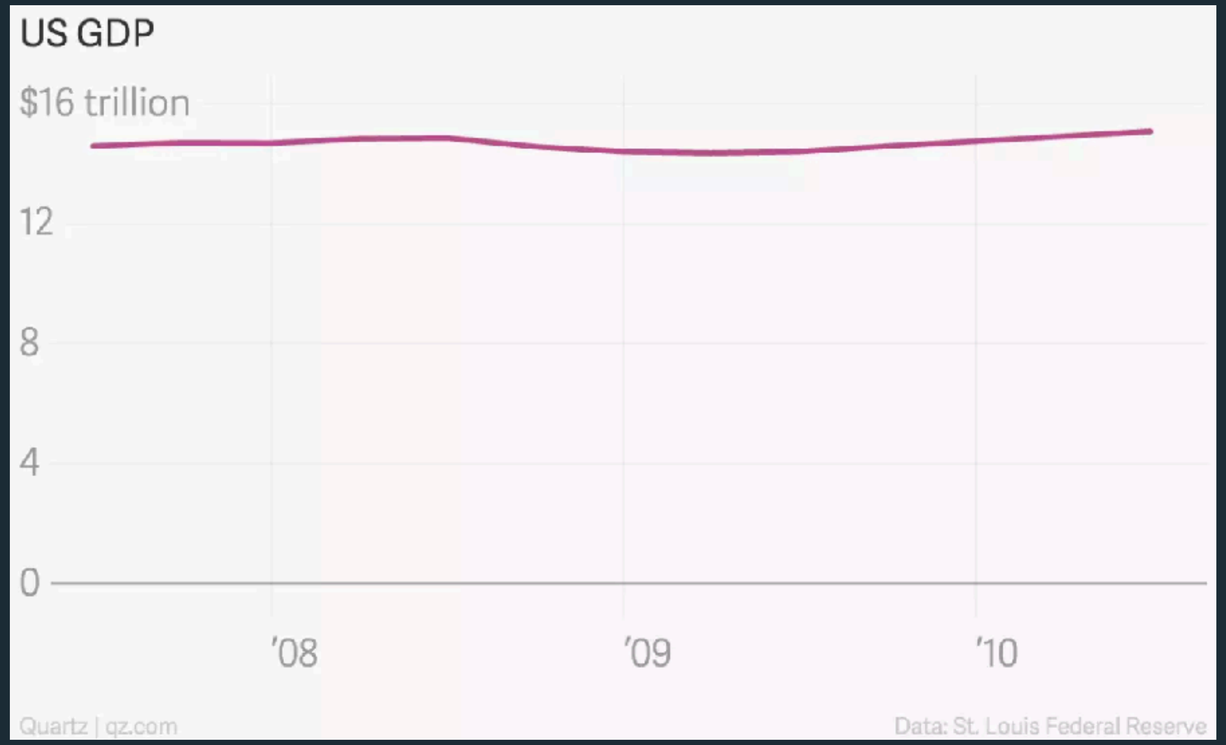
Compare proportions (Q-Ratio)

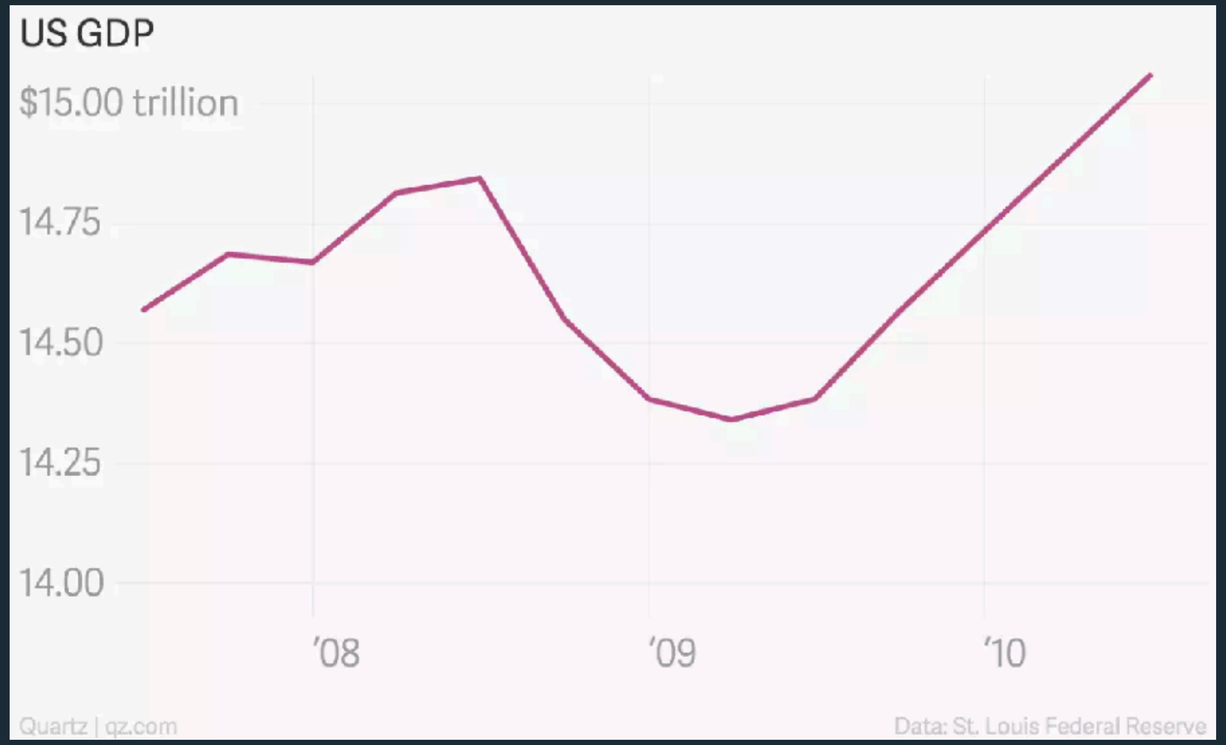


Compare relative positions (Q-Interval)



Zero Baseline?

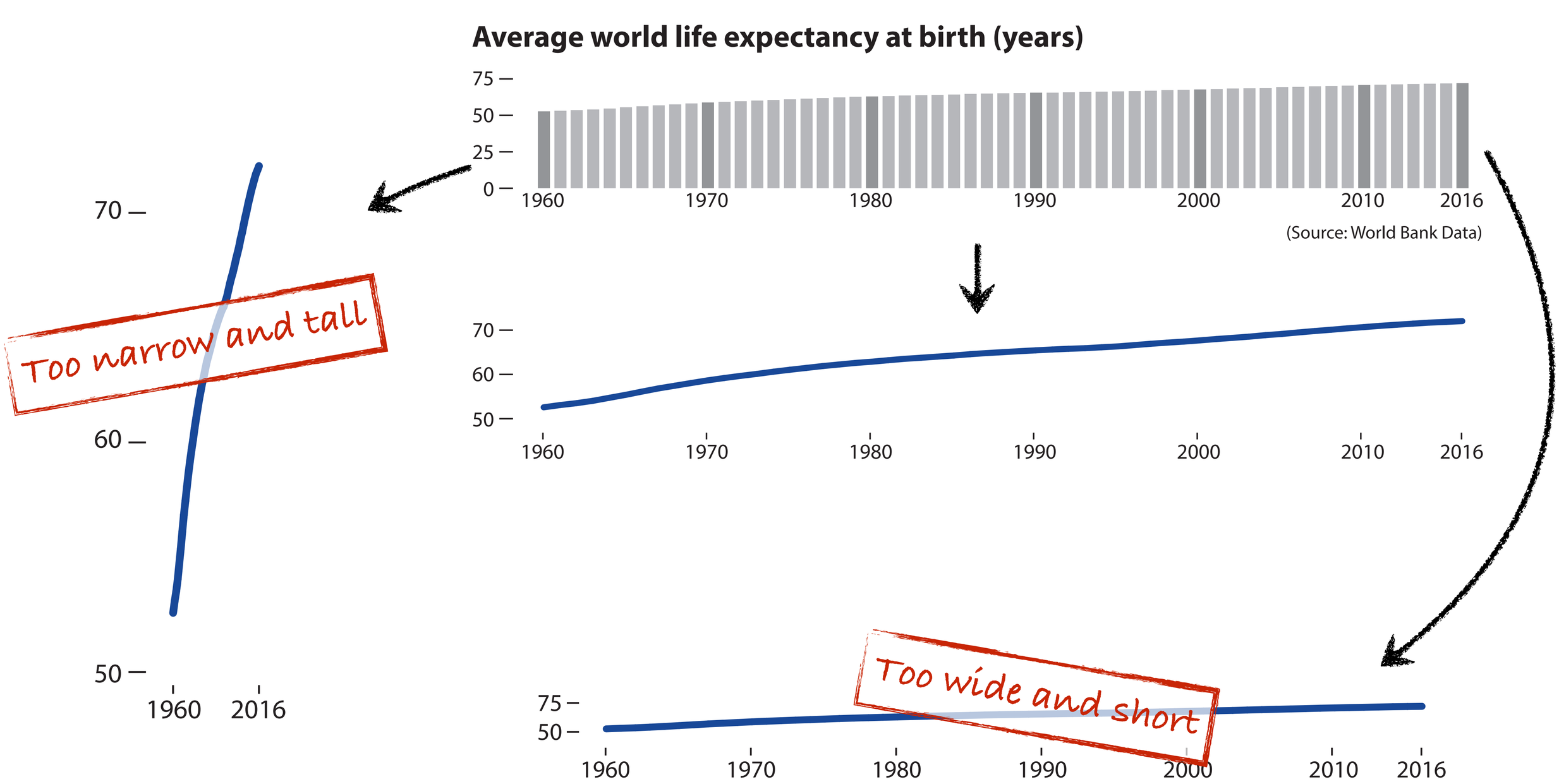




[David Yanofsky. *Quartz*, 2015]

Truncate the y-axis:

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

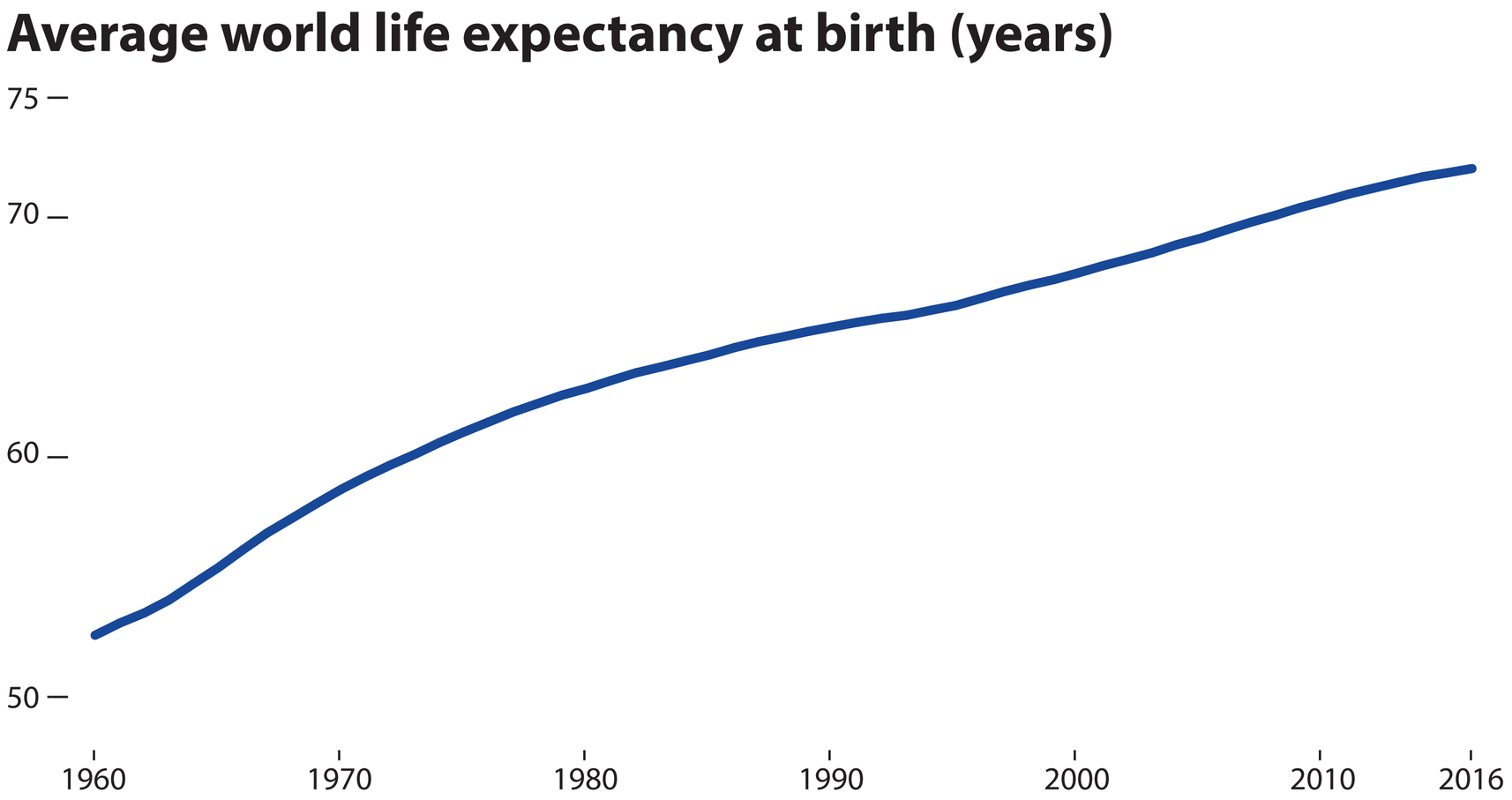


[Alberto Cairo. *How Charts Lie*, 2019]

Aspect Ratio

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

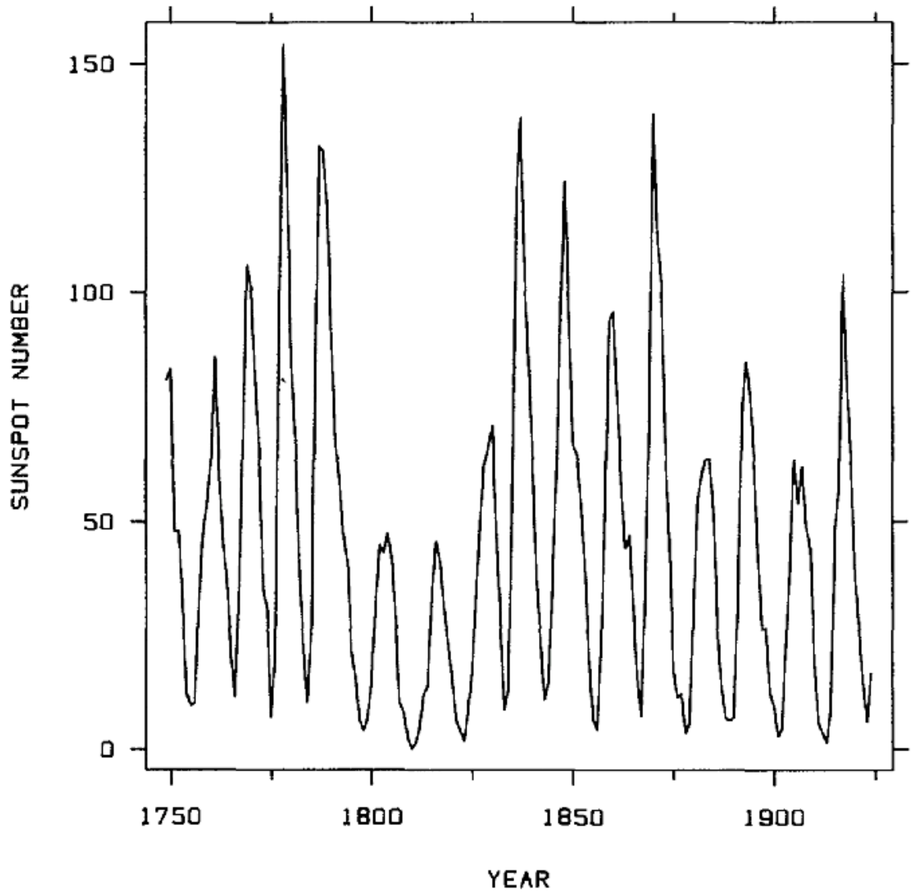
35% increase ≈ 1/3rd
≈ 3:1 aspect ratio

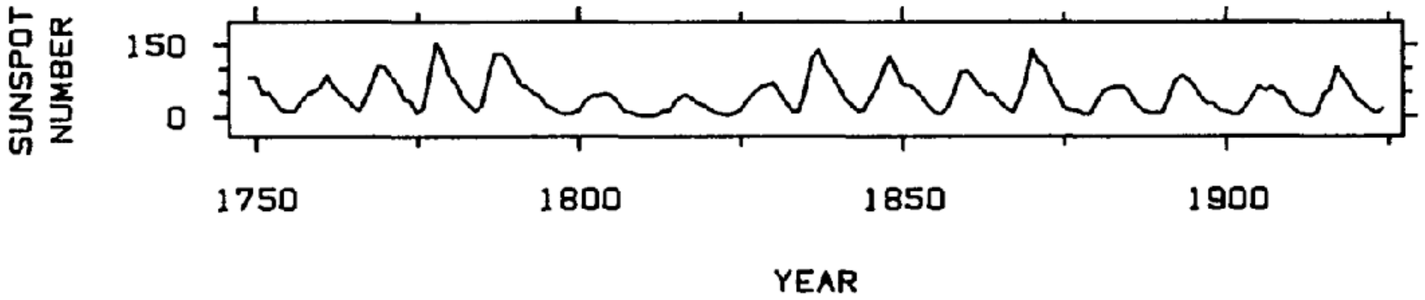


[Alberto Cairo. *How Charts Lie*, 2019]

Aspect Ratio

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



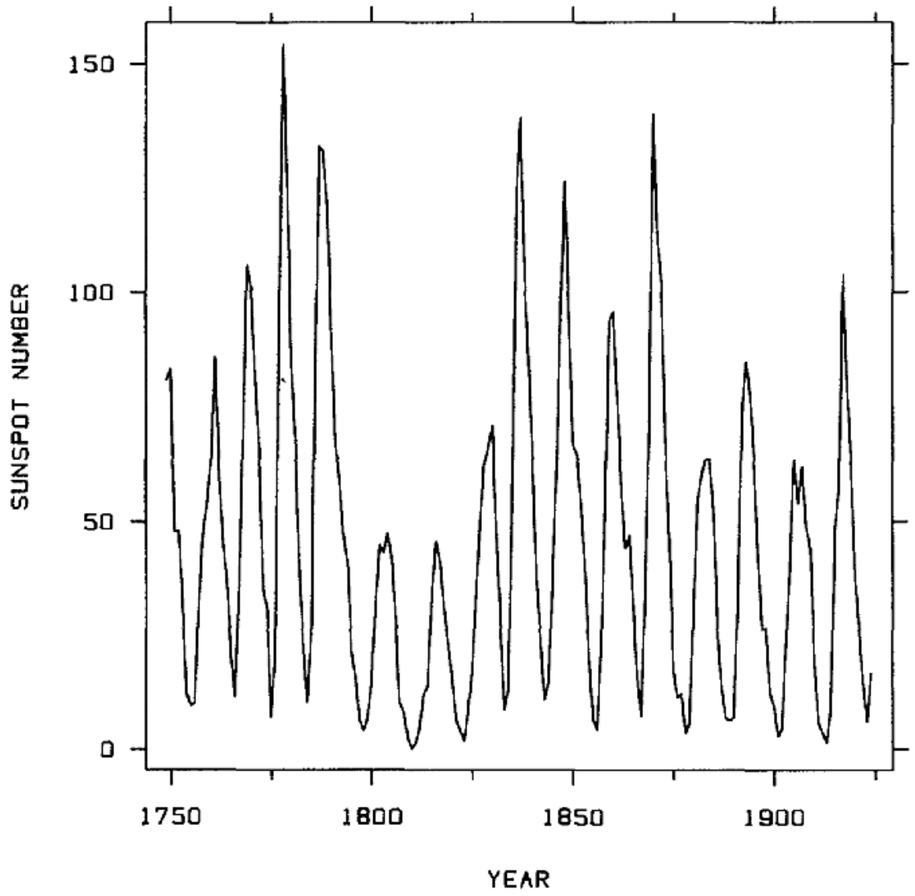


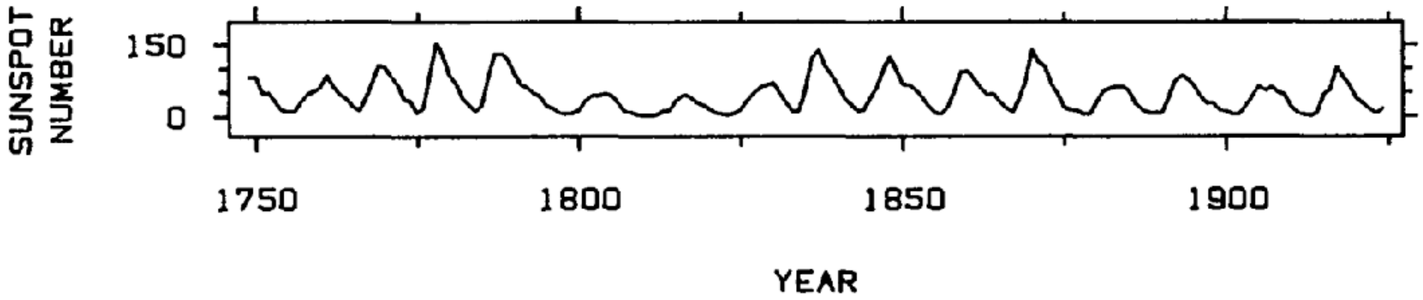
[Cleveland & McGill, 1987]

Aspect Ratio

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

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



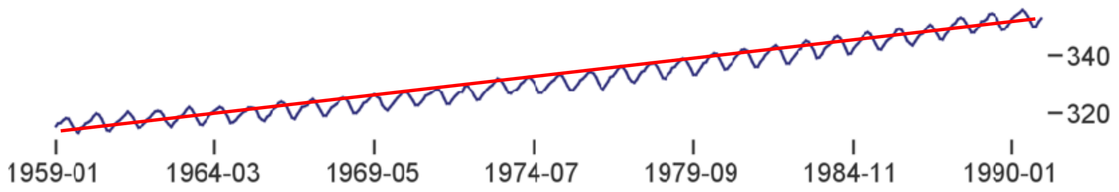


[Cleveland & McGill, 1987]

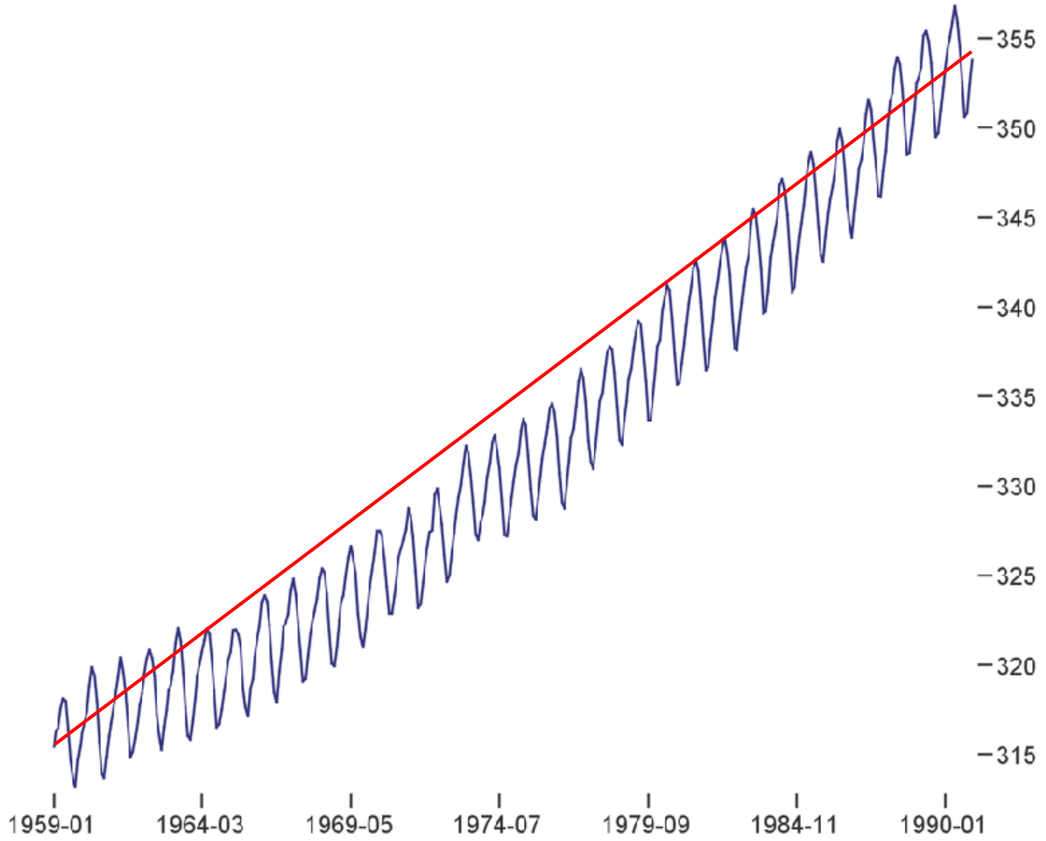
Aspect Ratio

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

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



Aspect ratio = 7.87



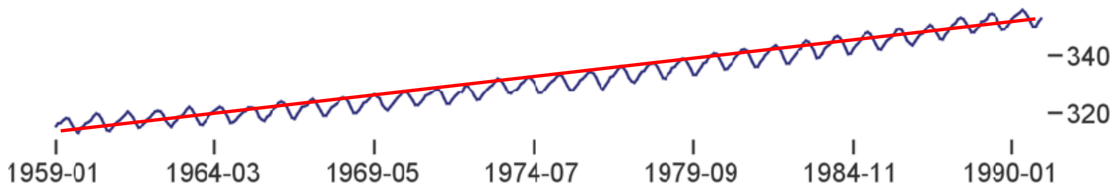
Aspect ratio = 1.17

[Heer & Agrawala, 2006]

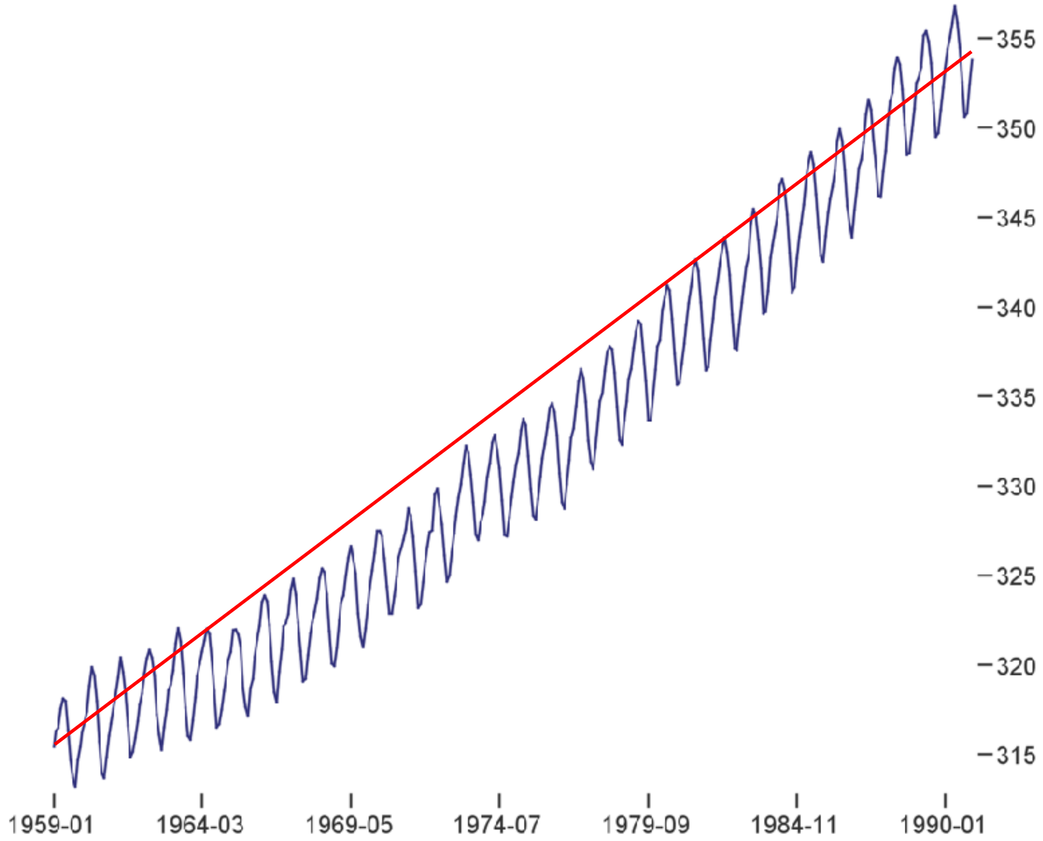
Aspect Ratio

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

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



Aspect ratio = 7.87



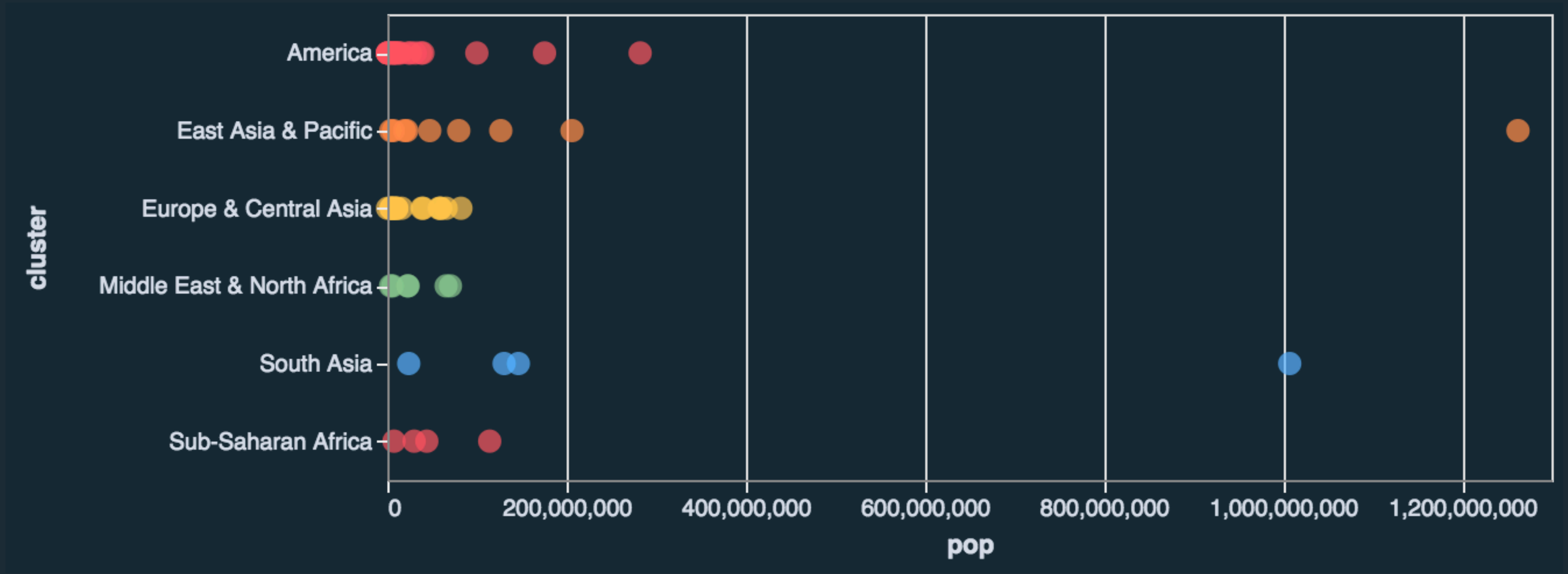
Aspect ratio = 1.17

[Heer & Agrawala, 2006]

Scaling Axes: Outliers & Skew

Options:

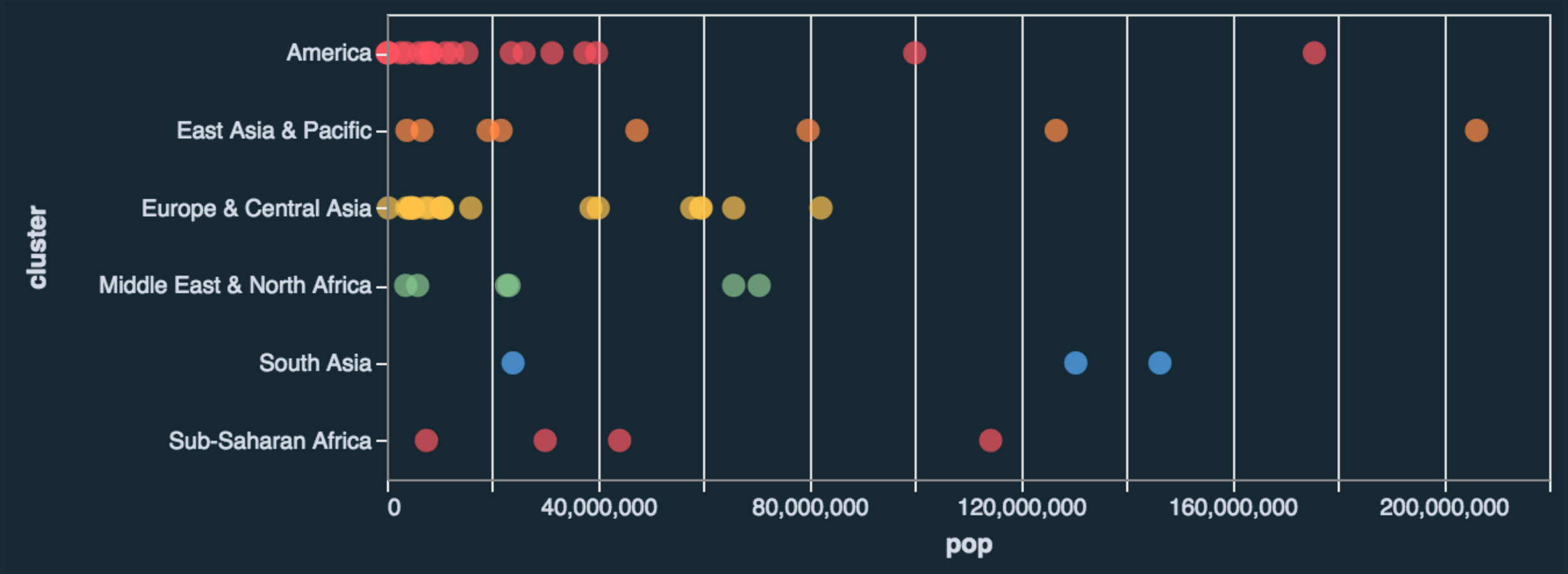
1. Clip them out.



Scaling Axes: Outliers & Skew

Options:

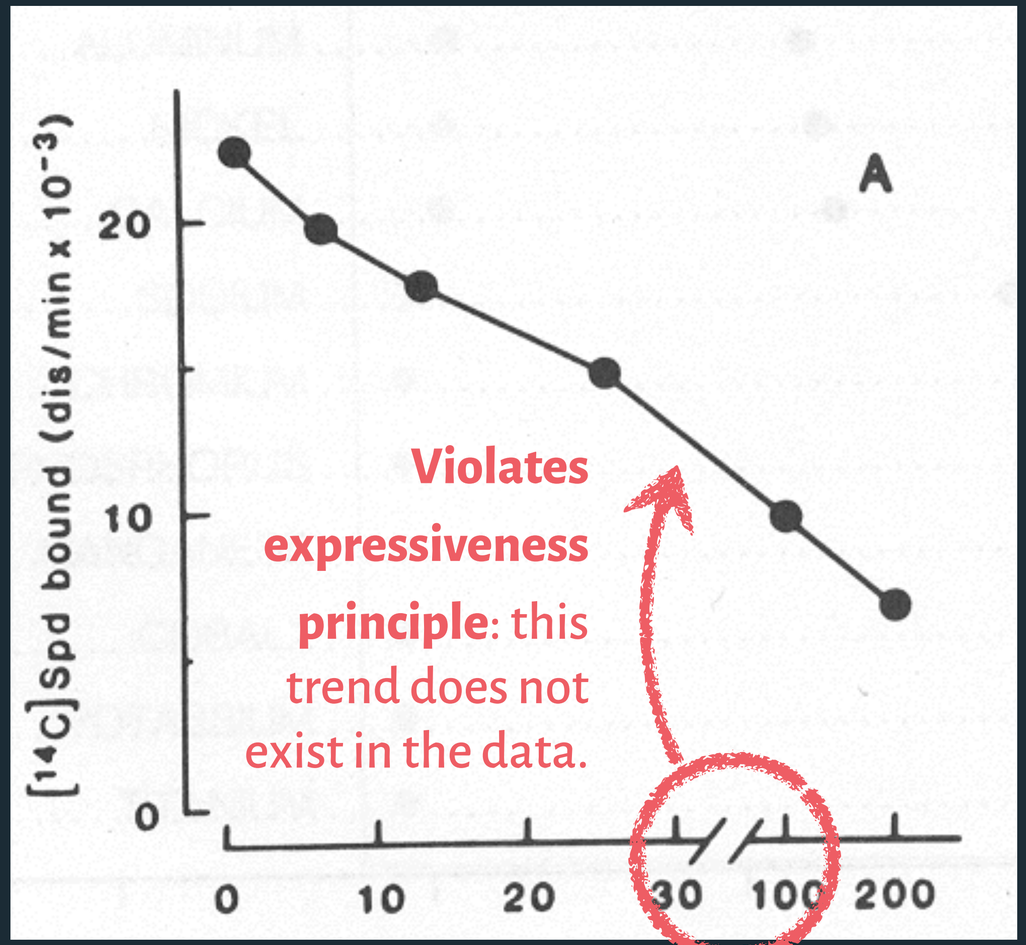
1. Clip them out.



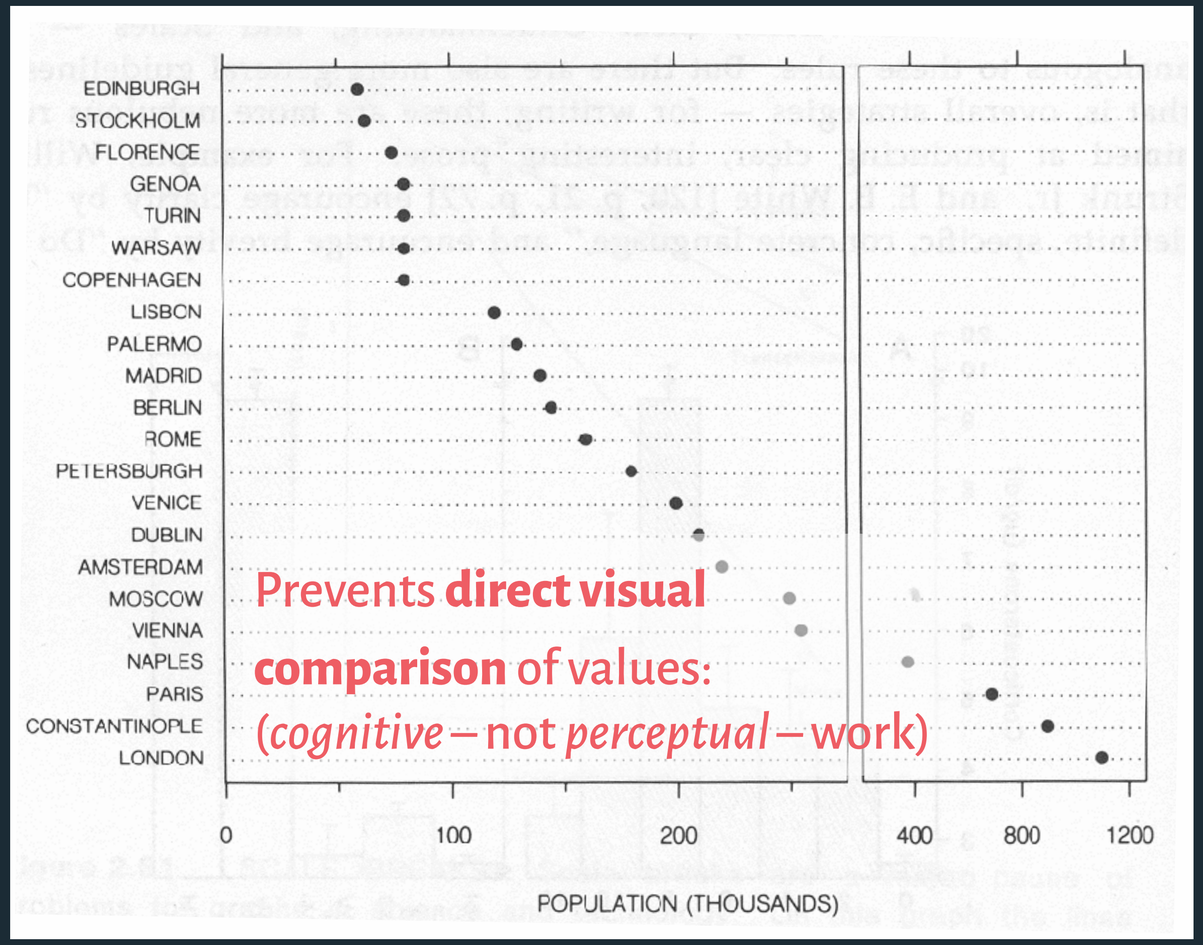
Scaling Axes: Outliers & Skew

Options:

1. Clip them out.
2. Scale Breaks



[Cleveland 85]

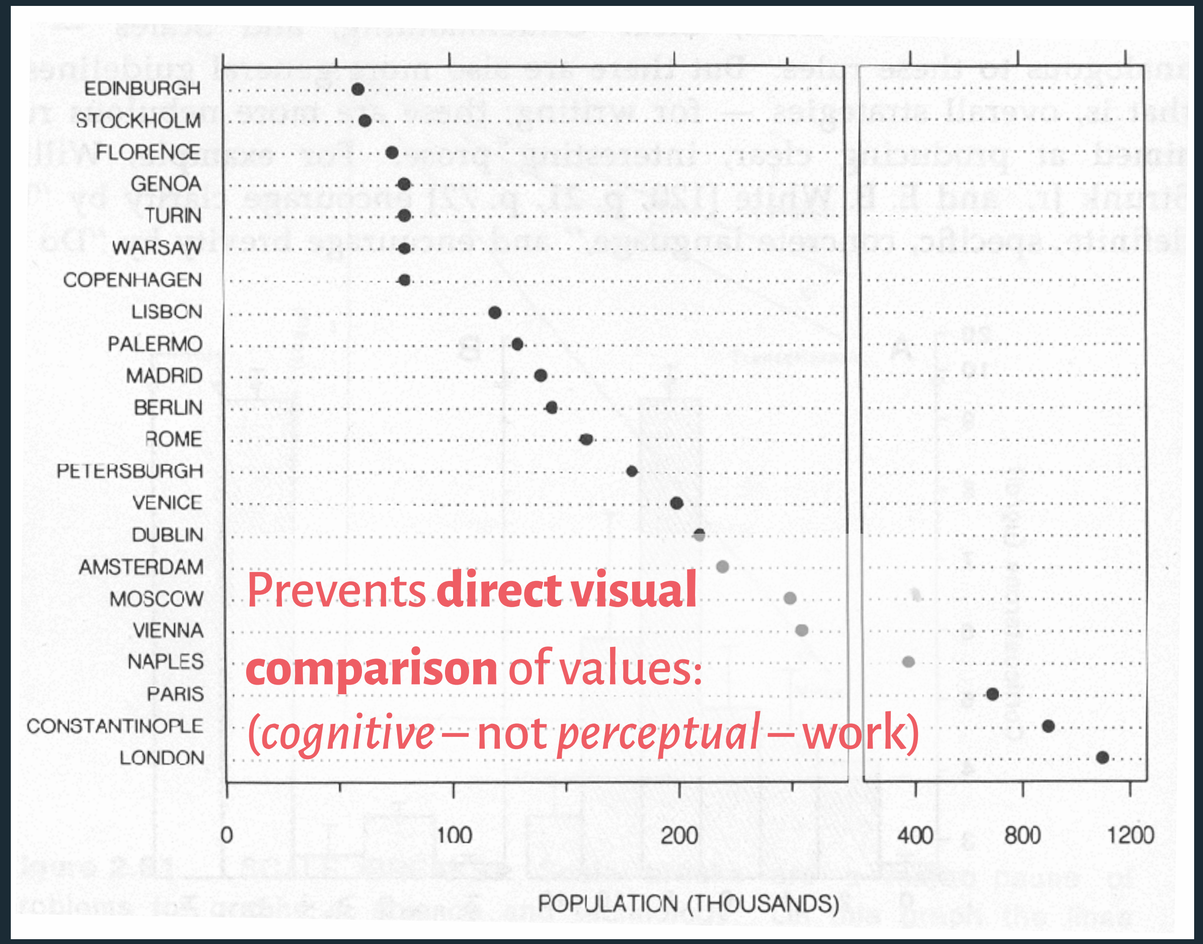


[Cleveland 85]

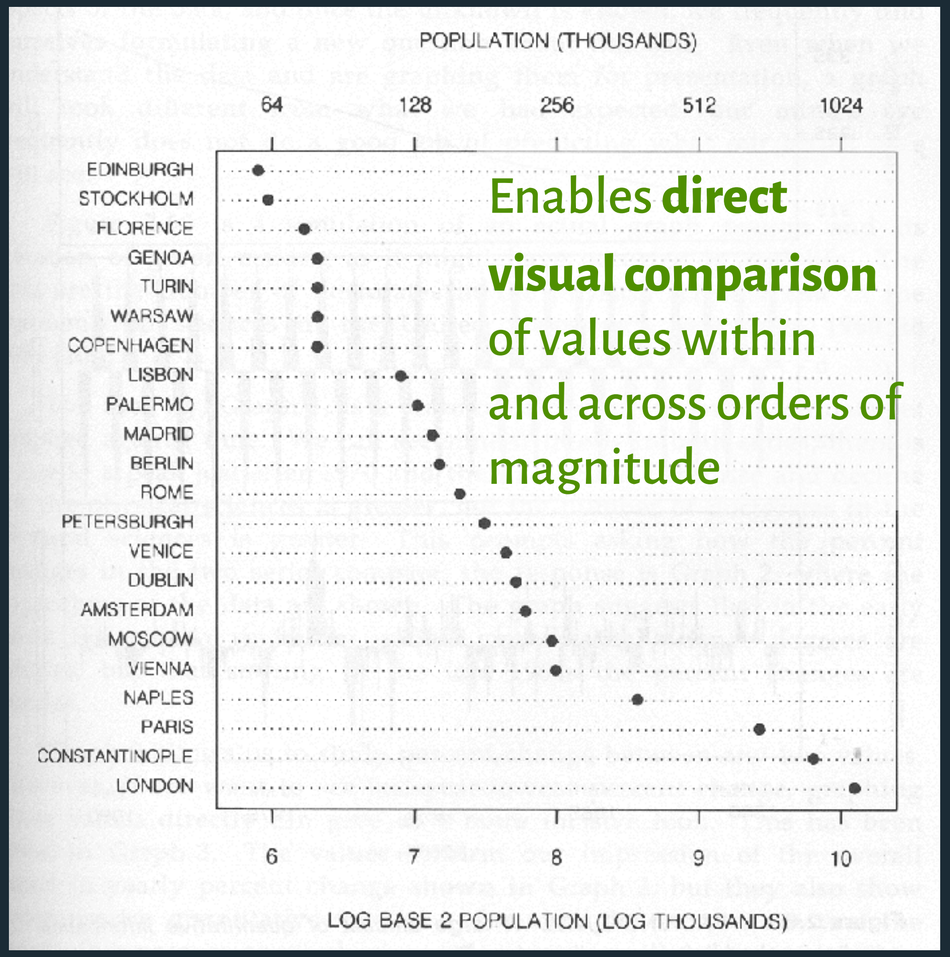
Scaling Axes: Outliers & Skew

Options:

1. Clip them out.
2. Scale Breaks
3. Log Scales



[Cleveland 85]

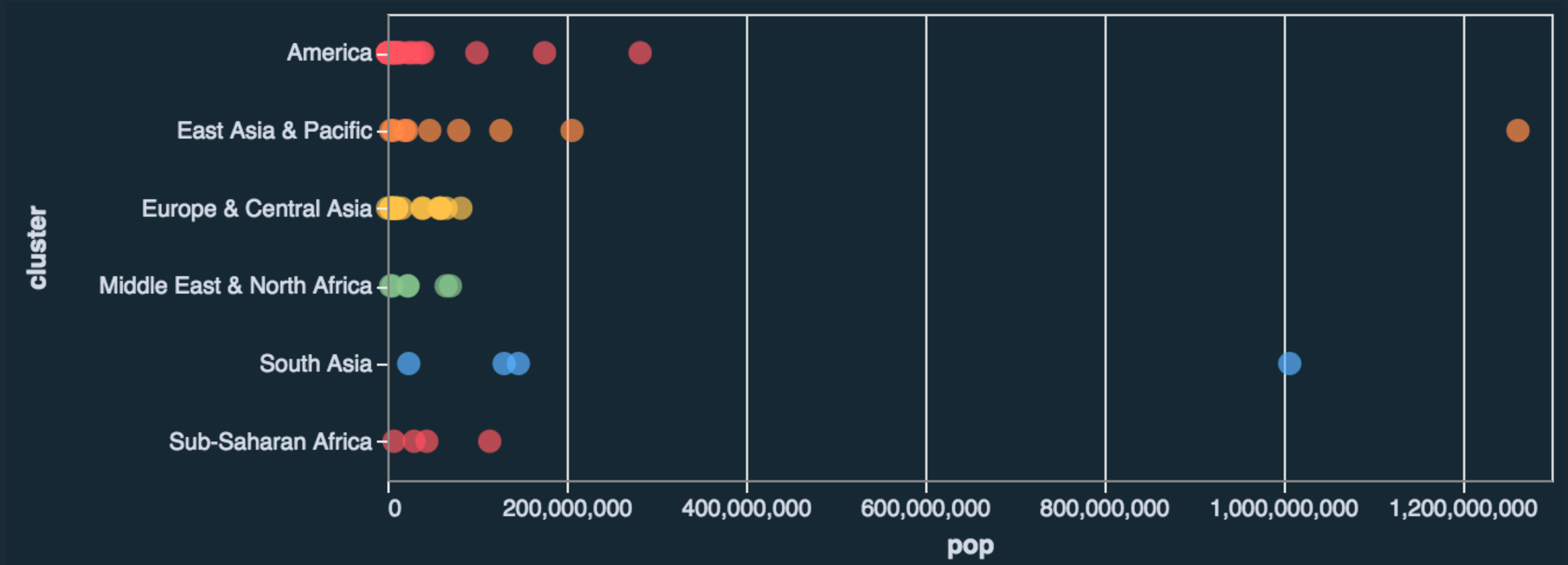


[Cleveland 85]

Scaling Axes: Outliers & Skew

Options:

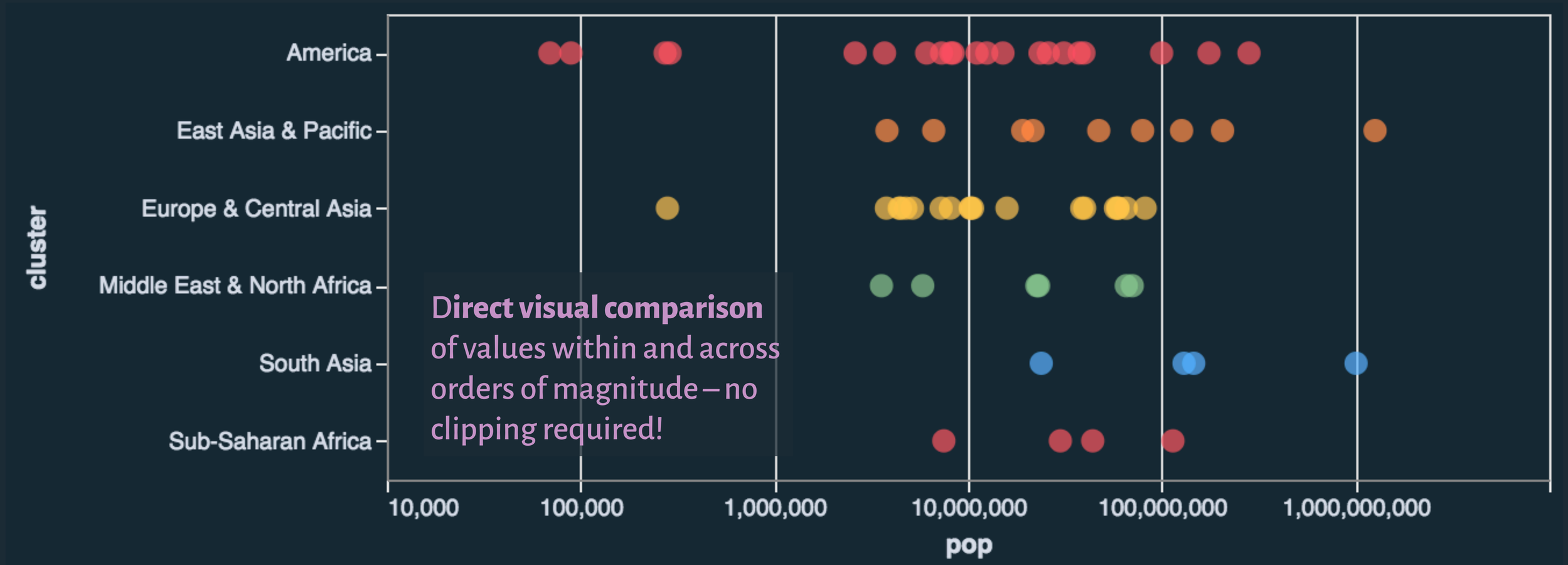
1. Clip them out.
2. Scale Breaks
3. Log Scales



Scaling Axes: Outliers & Skew

Options:

1. Clip them out.
2. Scale Breaks
3. Log Scales



Scaling Axes: Linear vs. Log

Linear Scale

Absolute change

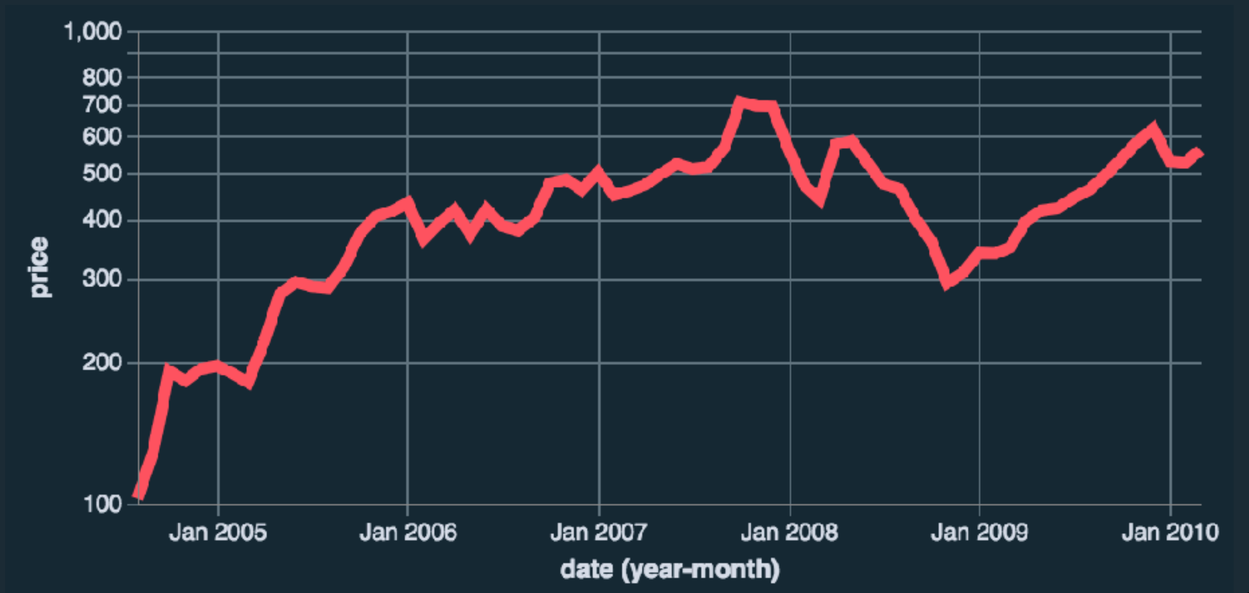
10 visual units (pixels) = 10 additional data units

Log Scale

Percentage change

10 visual units = multiplication of 10 data units

```
log(u) + log(v) = log(u*v)
d(100, 200) = d(300, 600)
```



Scaling Axes: Linear vs. Log

Log Scale

Percentage change

10 visual units = multiplication of 10 data units

```
log(u) + log(v) = log(u*v)
d(100, 200) = d(300, 600)
```

Constraints

Positive, non-zero values

Audience familiarity?

Using space (in)effectively

(De-)Obfuscating data

(Mis)leading the witness

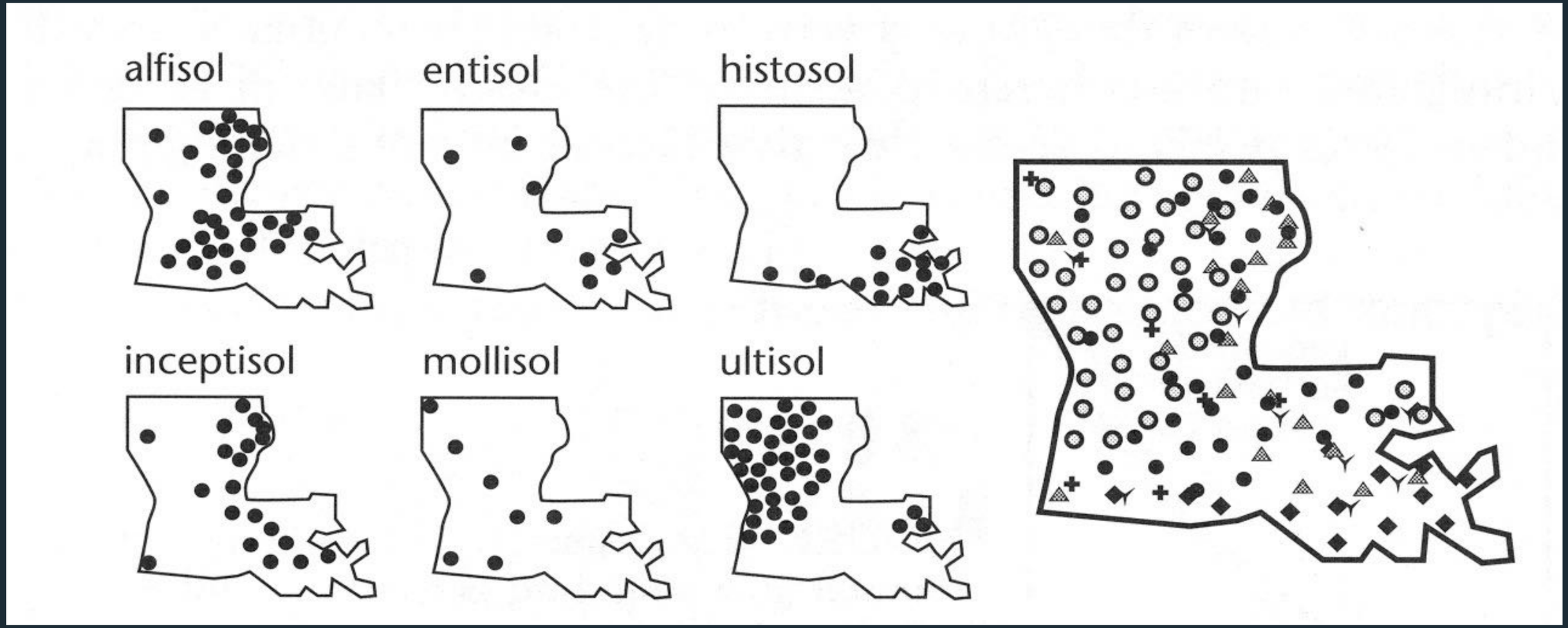
Using space (in)effectively

(De-)Obfuscating data

(Mis)leading the witness

[MacEachren ’95, Figure 2.11, p. 38]

Small Multiples



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

Trellis Plots

Subdivide space to enable comparison across multiple plots.

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

[Becker, Cleveland, and Shyu. 1996]

Data-Ink Ratio

$$= \frac{\text{Data ink}}{\text{Total ink use in graphic}}$$

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

= 1.0 – proportion of graphic that can be erased.

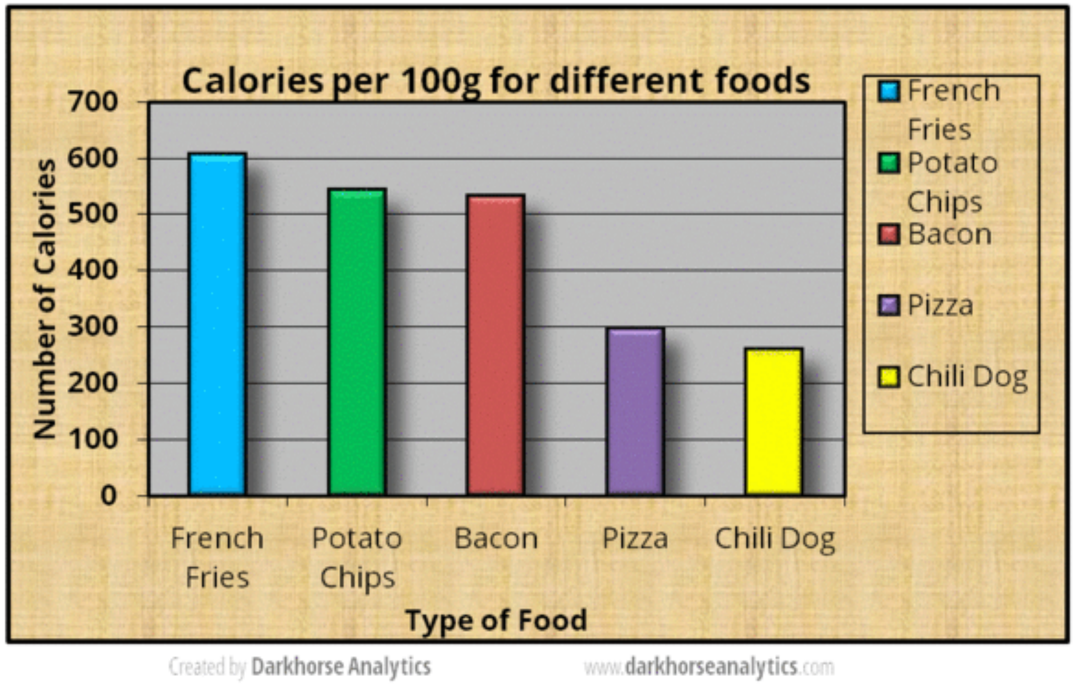
Remove

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?



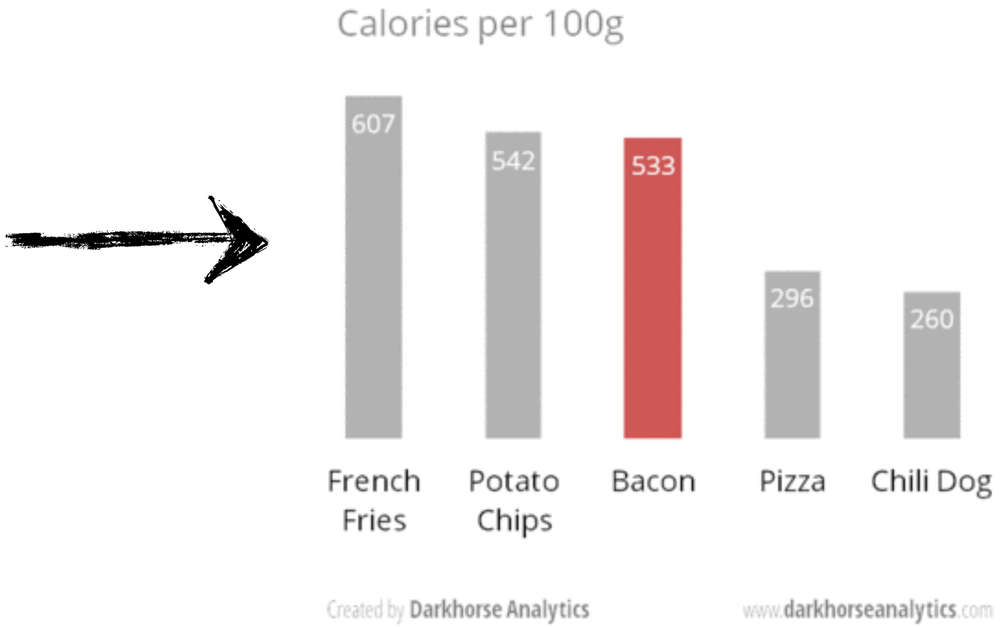
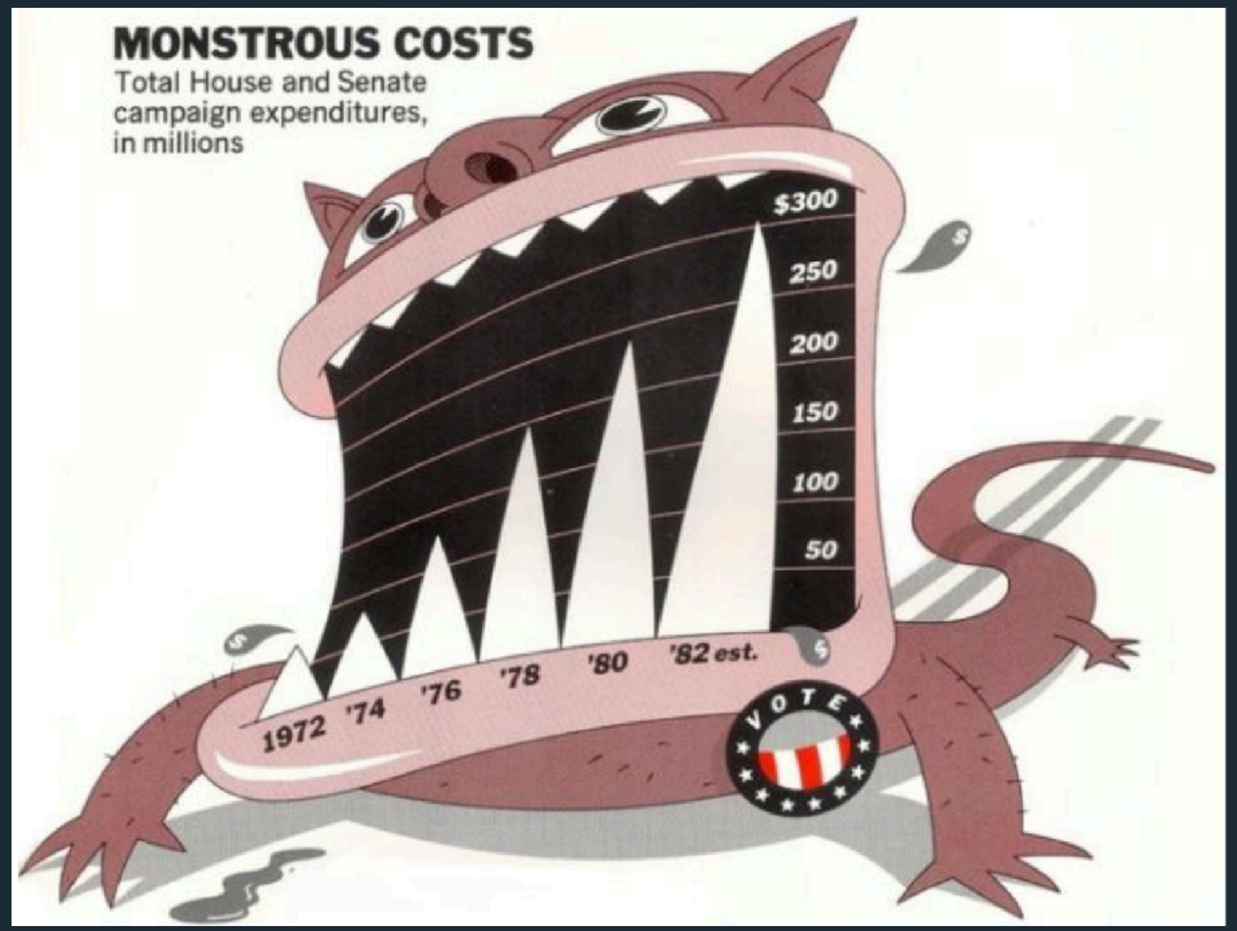
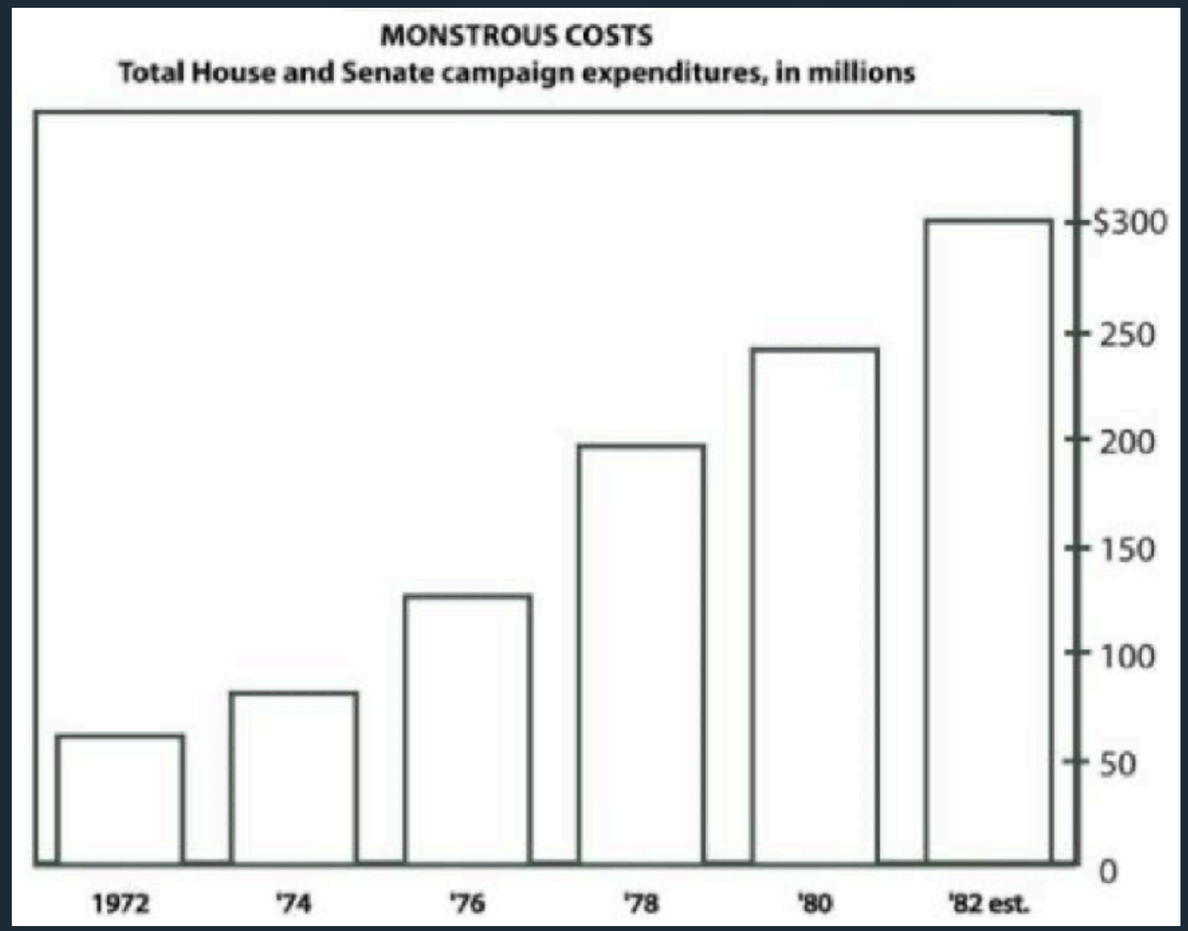


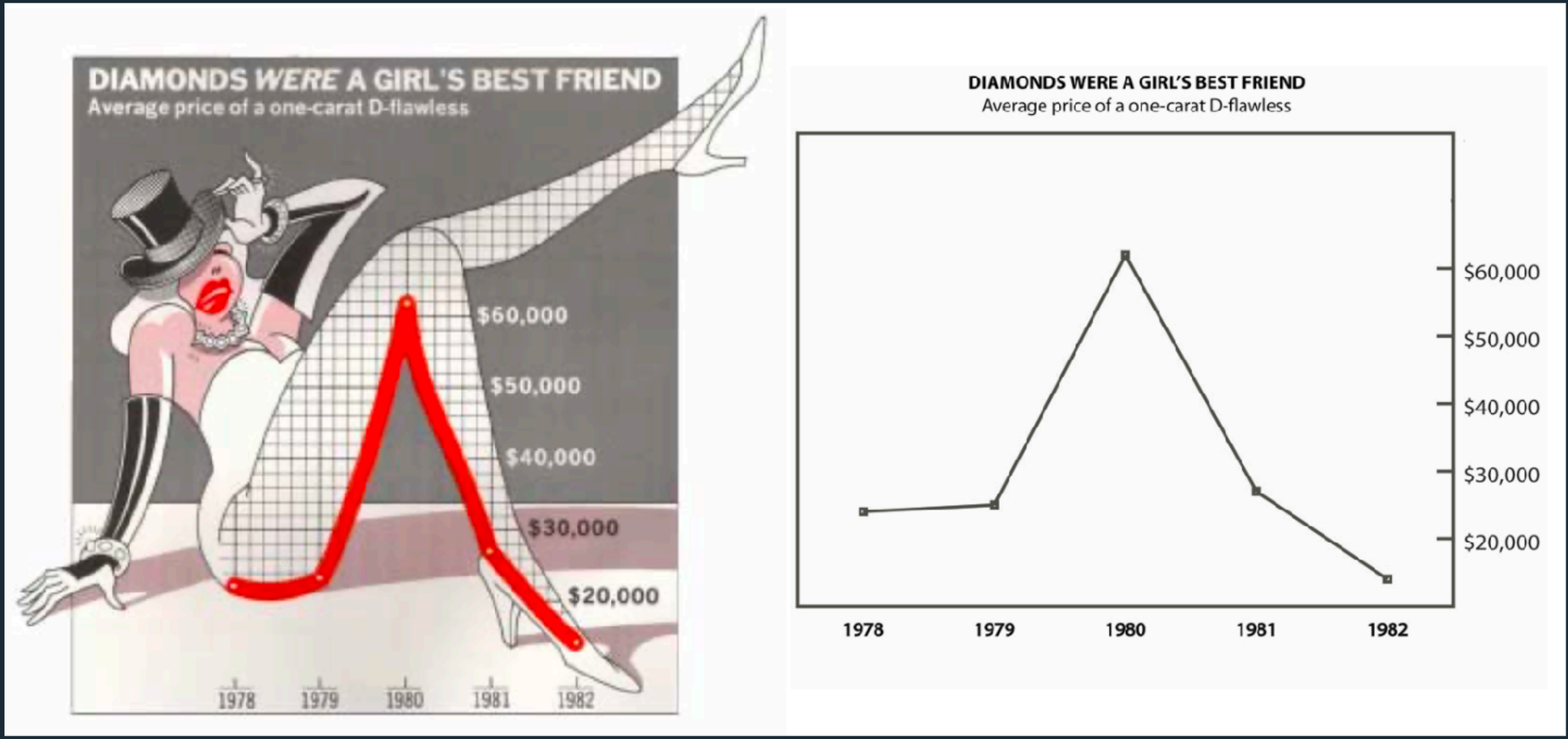
Chart Junk





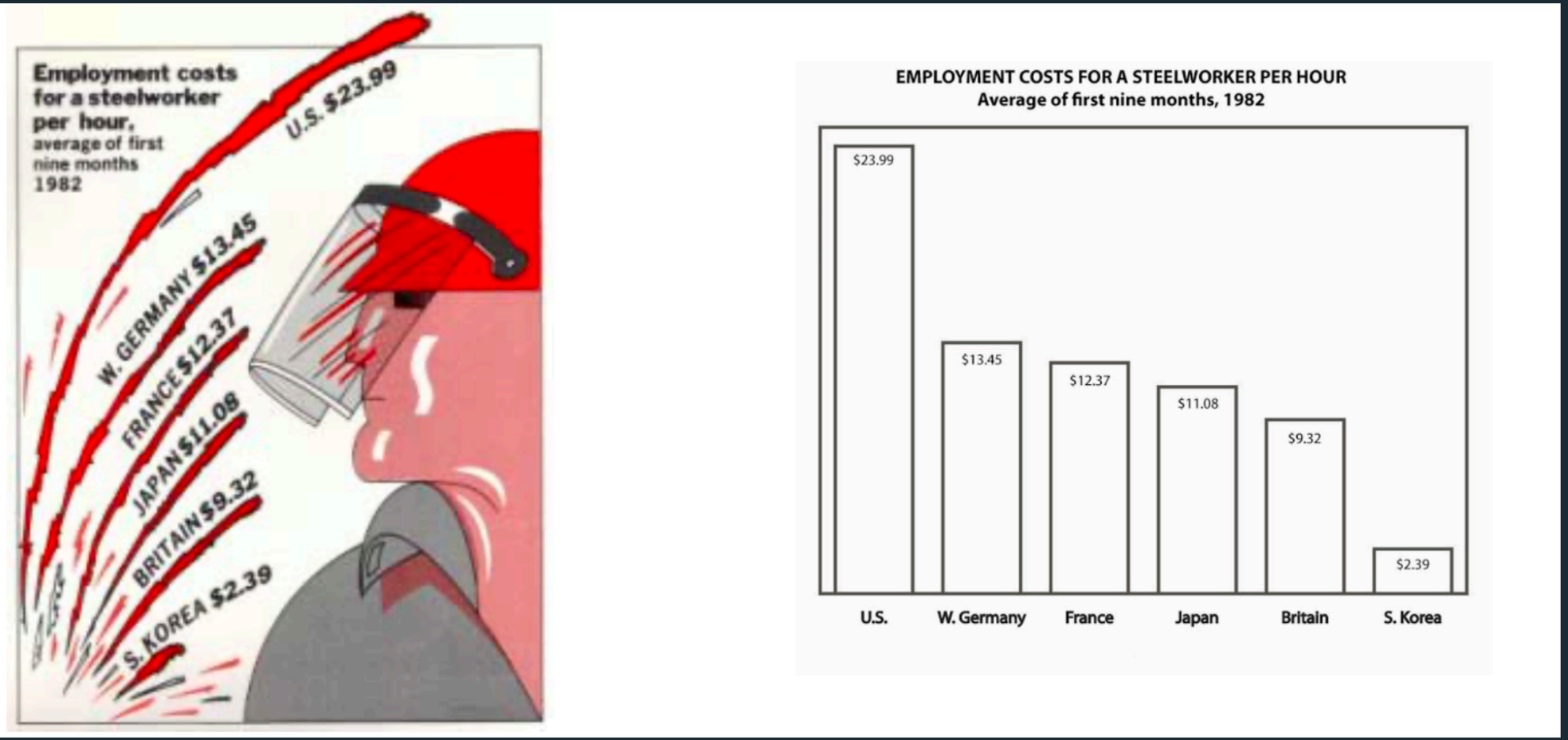
[Bateman et al., *CHI 2010*]

Chart Junk



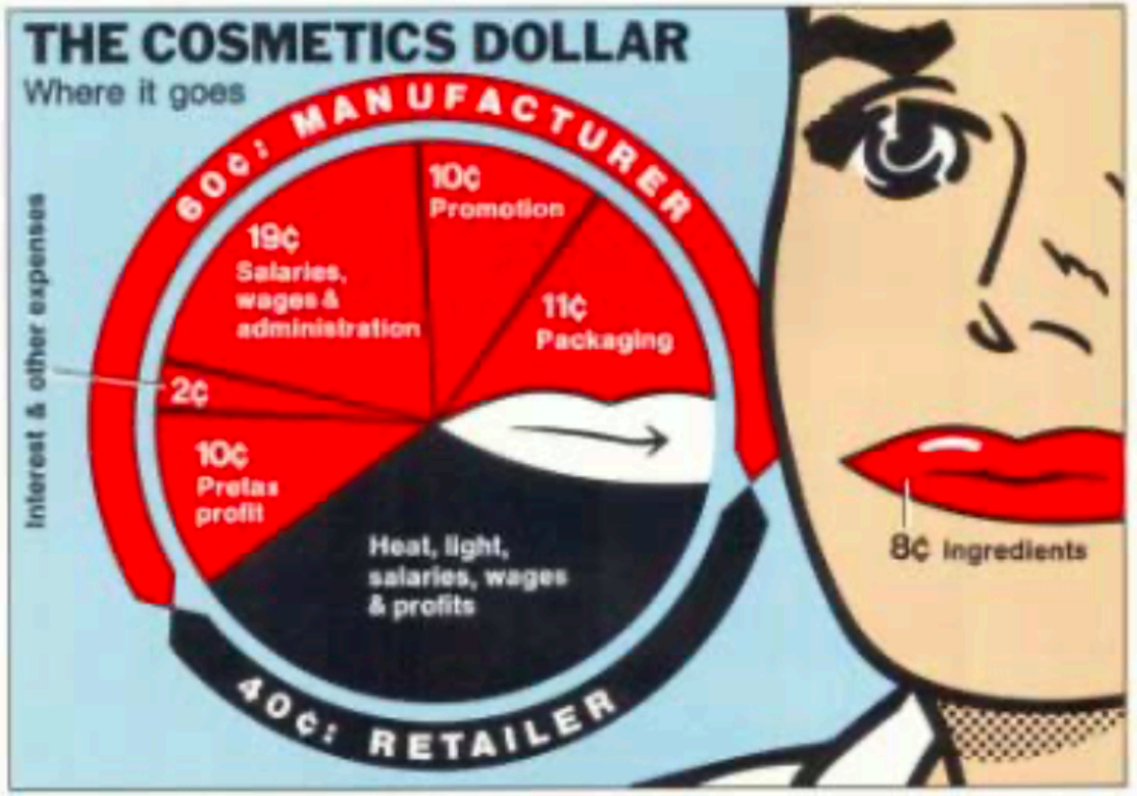
[Bateman et al., *CHI 2010*]

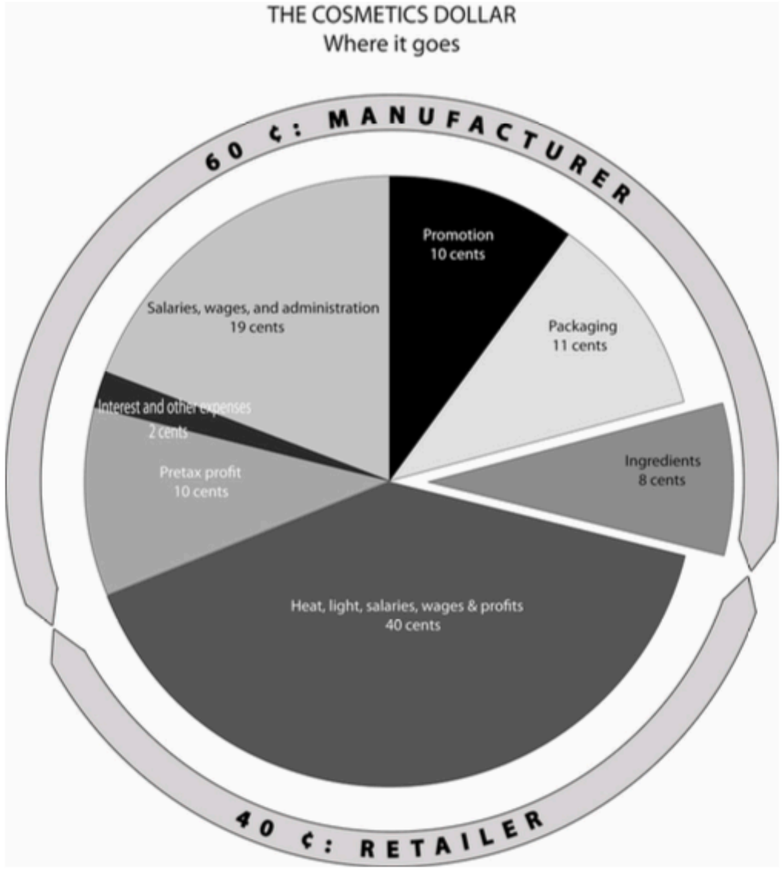
Chart Junk



[Bateman et al., *CHI 2010*]

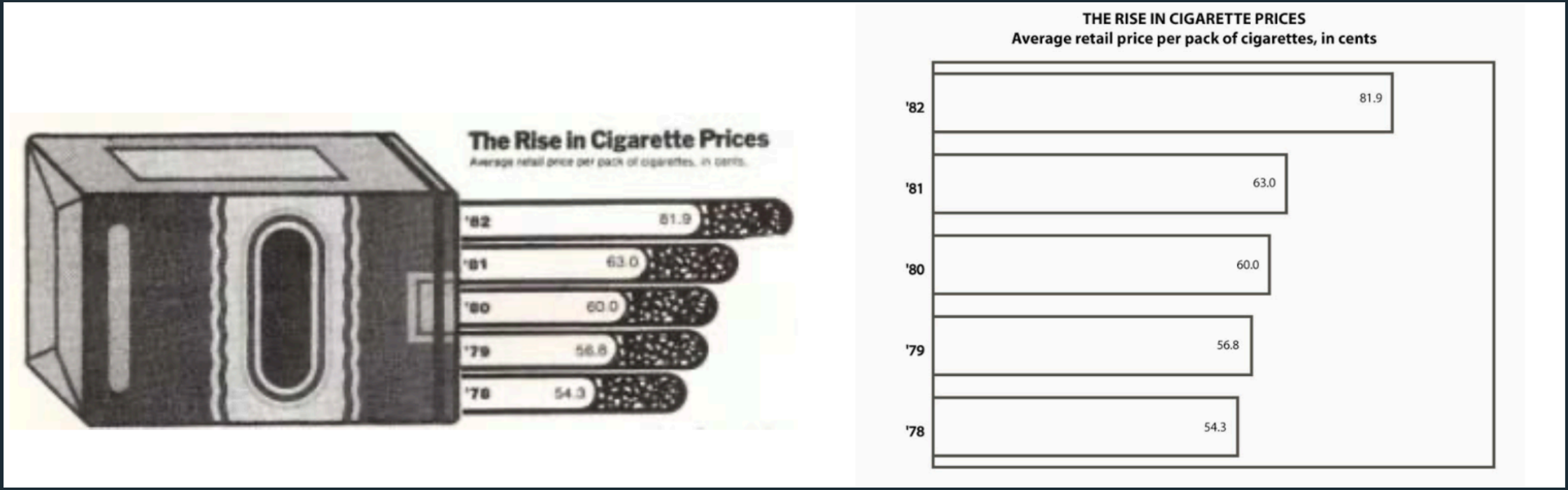
Chart Junk





[Bateman et al., *CHI 2010*]

Chart Junk



[Bateman et al., *CHI 2010*]

Using space (in)effectively

(De-)Obfuscating data

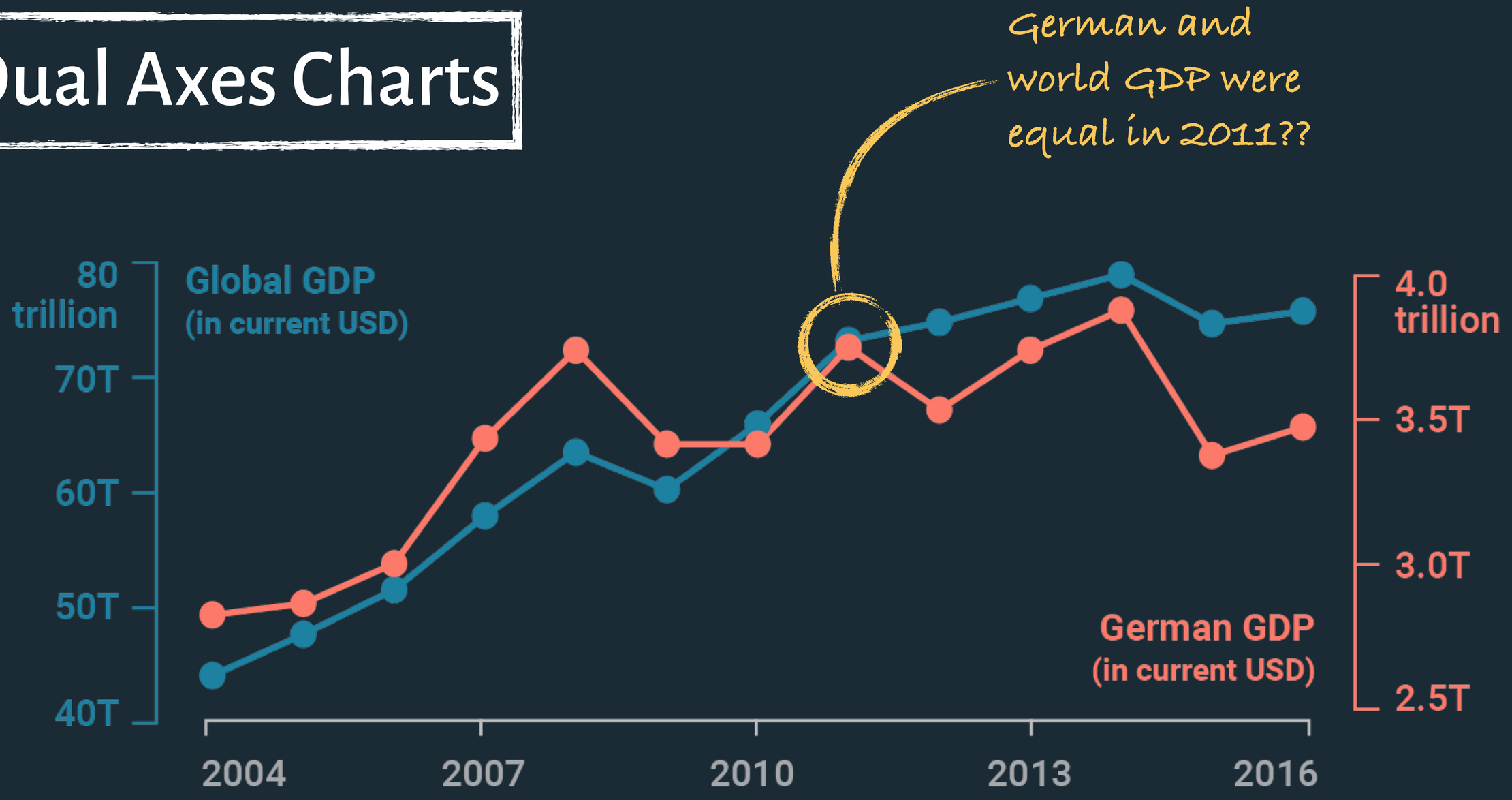
(Mis)leading the witness

Using space (in)effectively

(De-)Obfuscating data

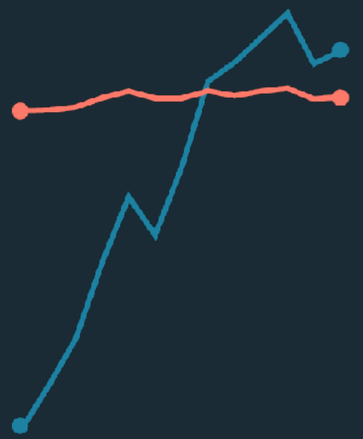
(Mis)leading the witness

Dual Axes Charts



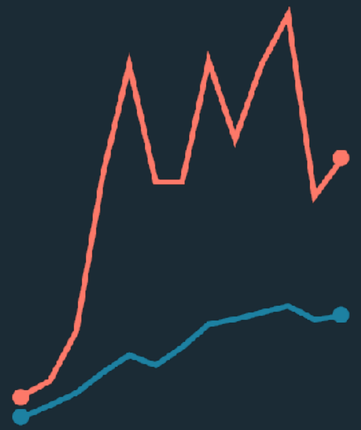
Lisa Charlotte Rost. *DataWrapper*, May 2018.

Dual Axes Charts

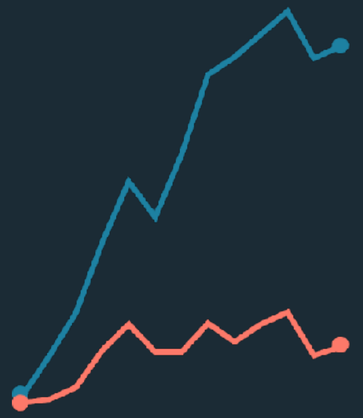


Orange steady,
Blue massively increasing.

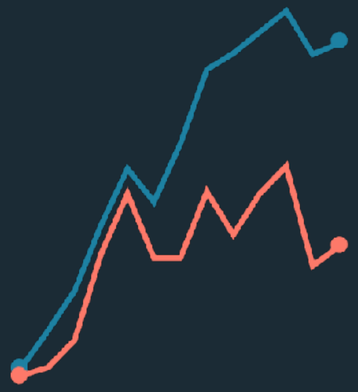
Blue steady,
Orange increasing.



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.

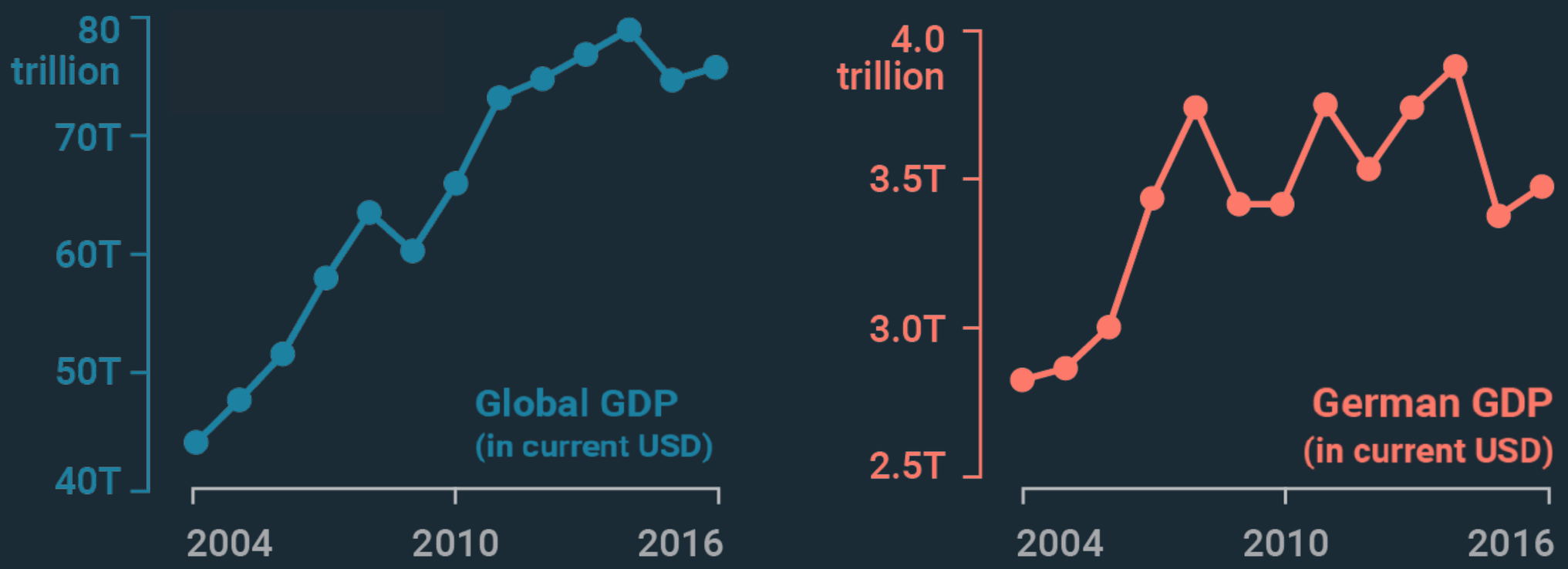


Both started with the same increase, then Blue raced to the top.

Both steady.

Lisa Charlotte Rost. *DataWrapper*, May 2018.

Dual Axes Charts



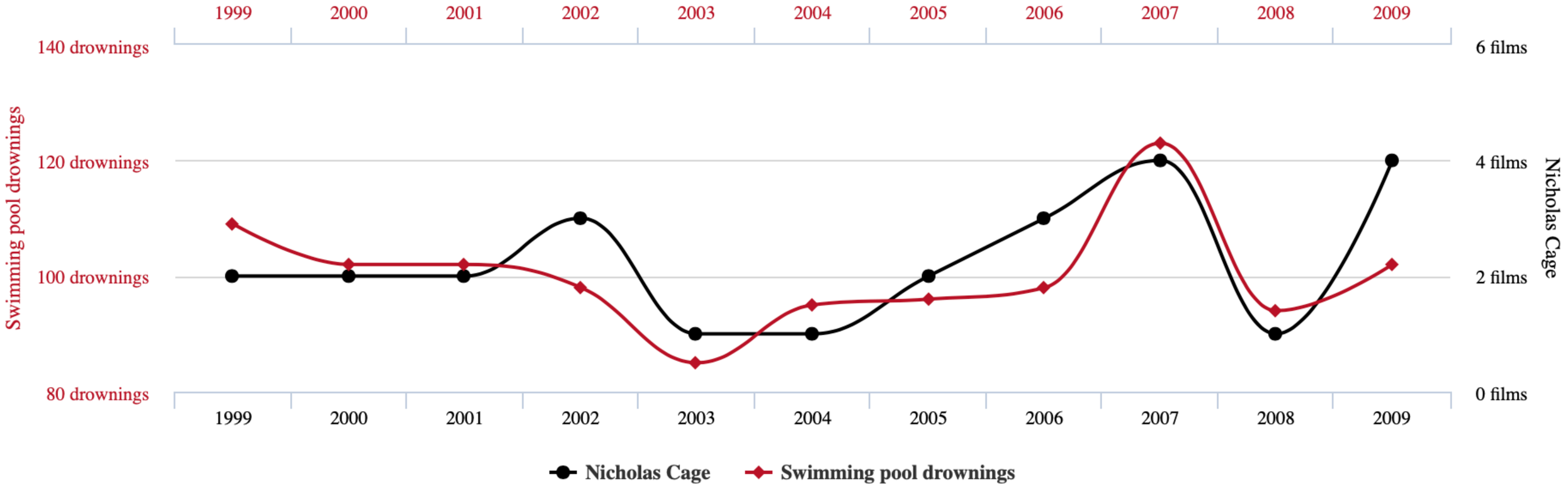
Lisa Charlotte Rost. *DataWrapper*, May 2018.

Number of people who drowned by falling into a pool

correlates with

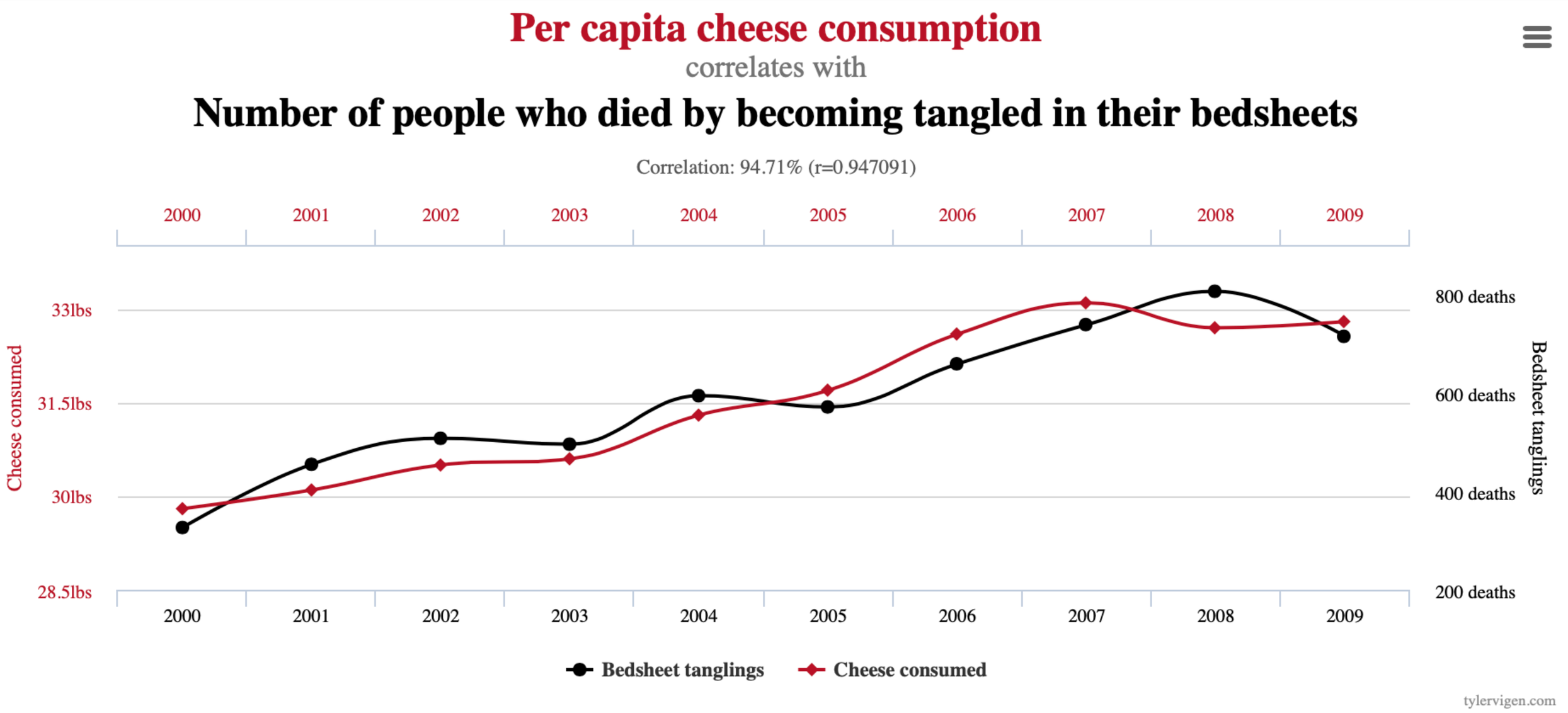
Films Nicolas Cage appeared in

Correlation: 66.6% (r=0.666004)

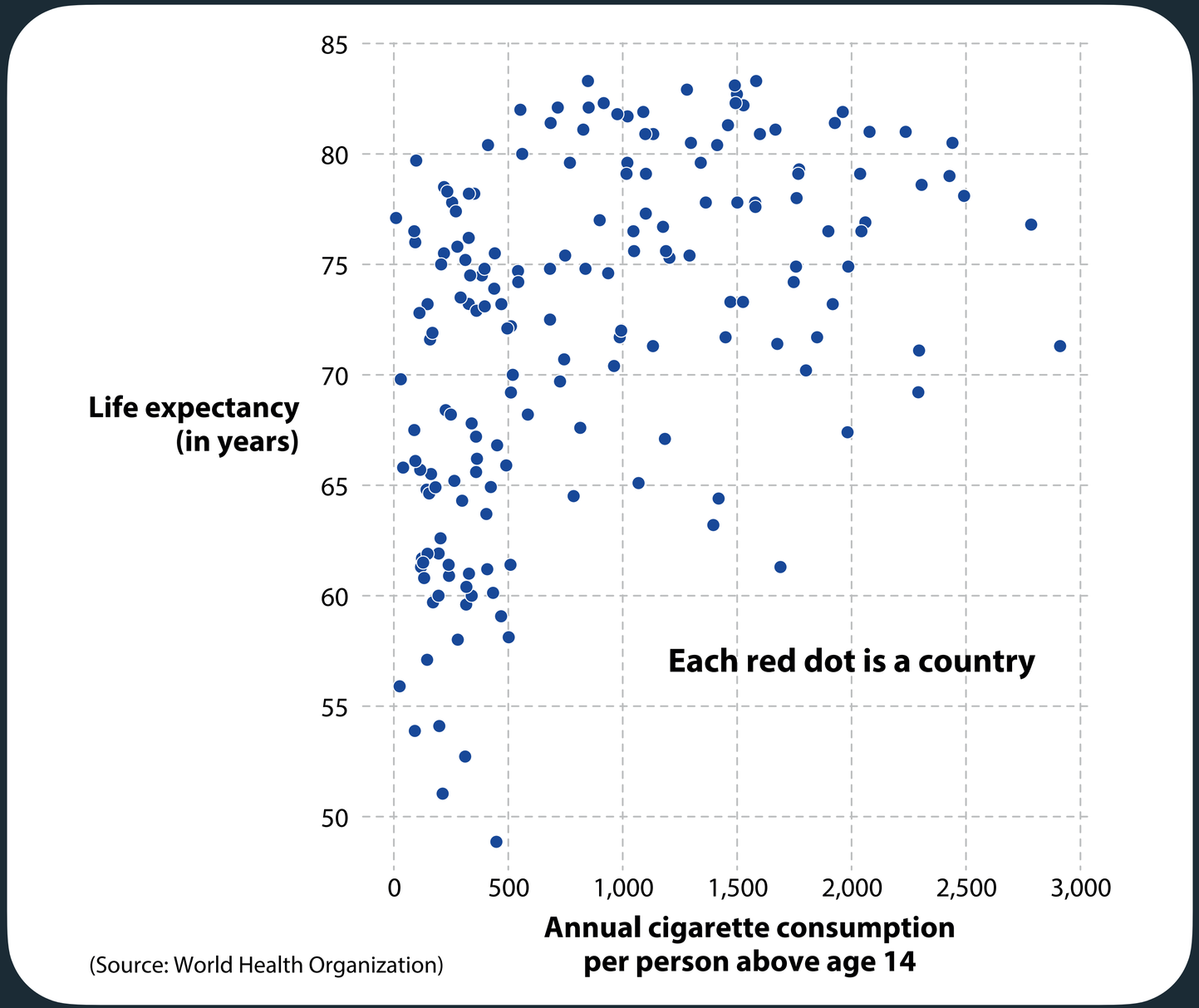


tylervigen.com

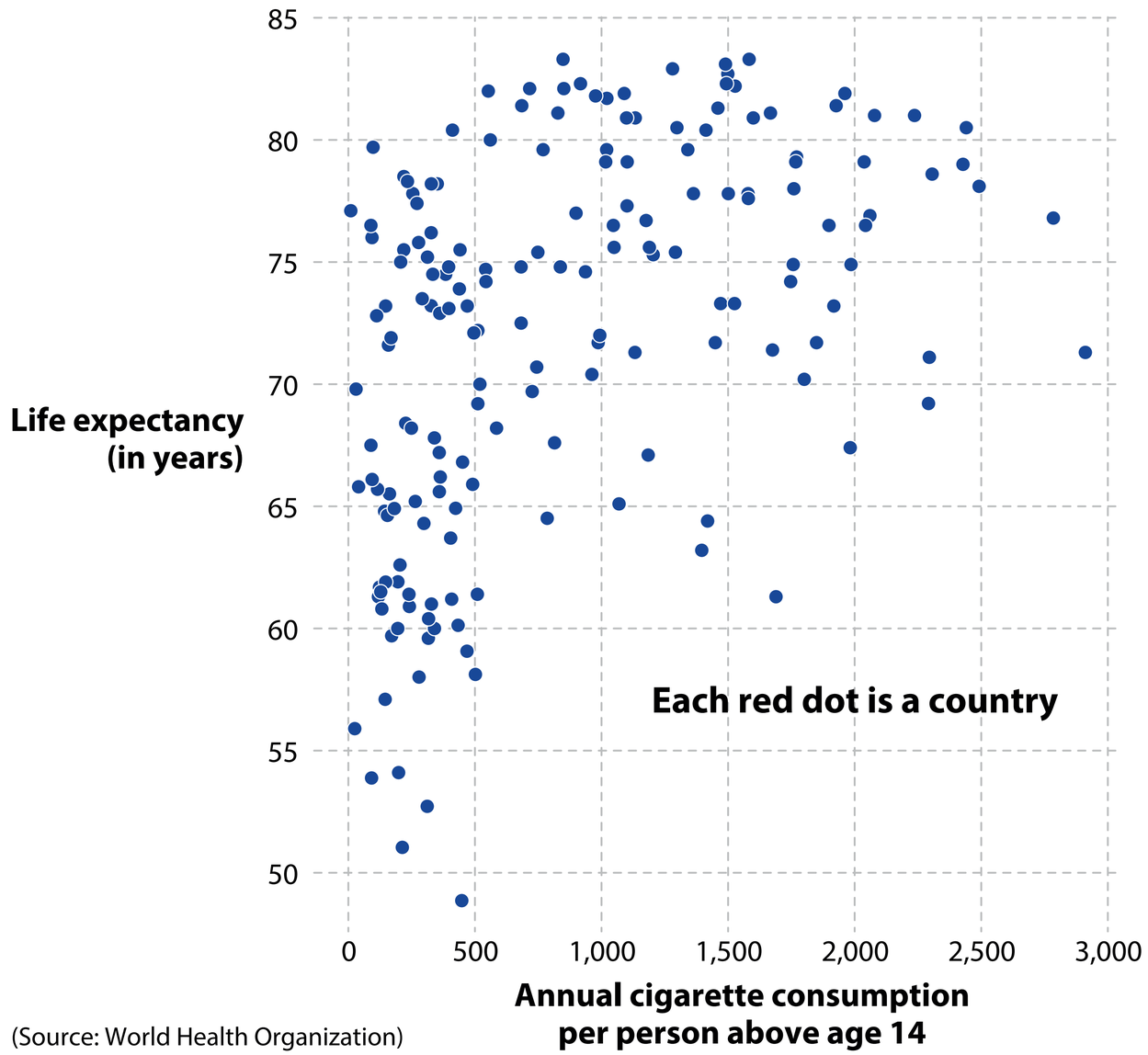
Data sources: Centers for Disease Control & Prevention and Internet Movie Database



Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention



[Alberto Cairo. *How Charts Lie*, 2019]

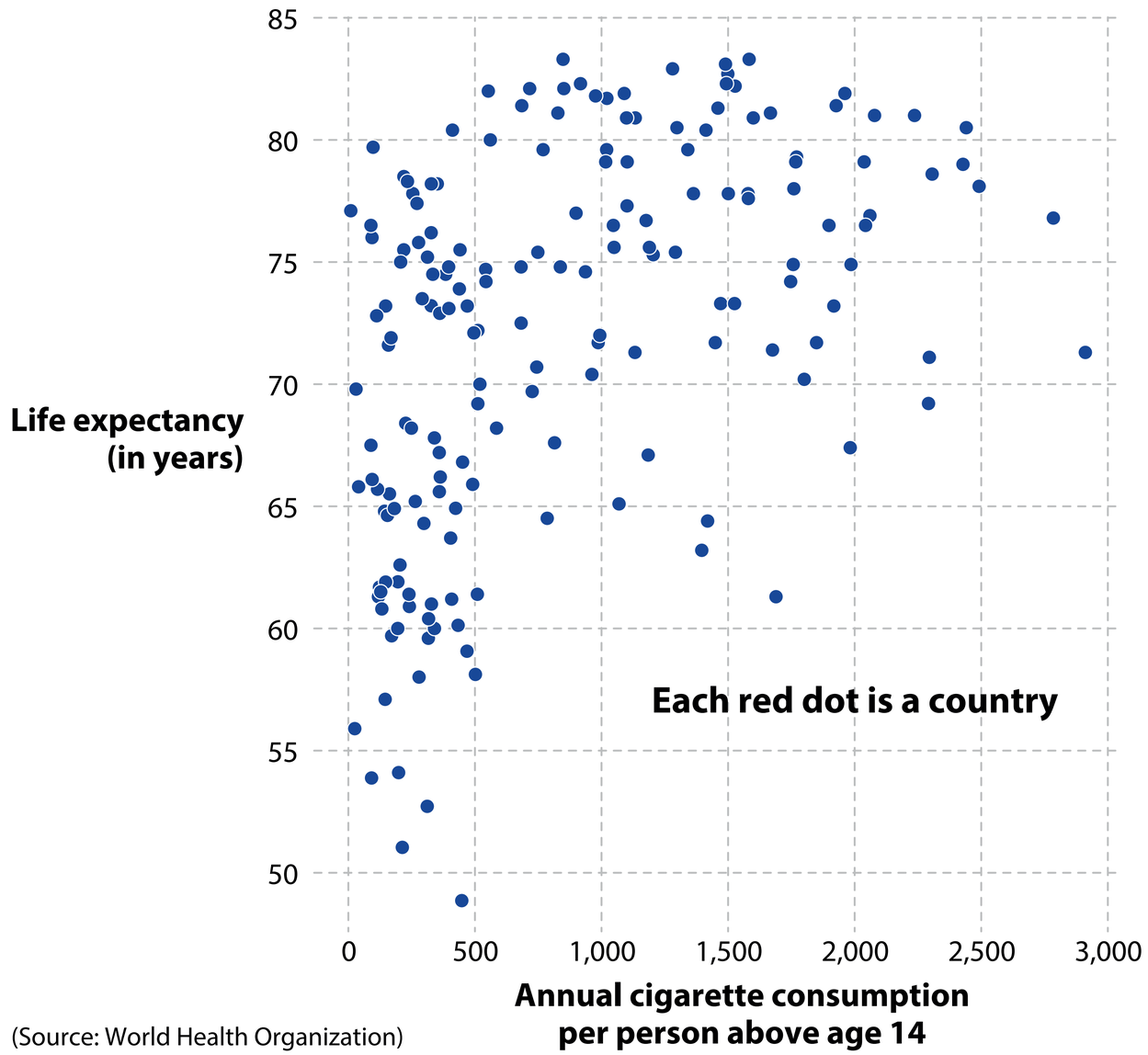


Describes a causal relationship.

~~*"The more cigarettes we consume, the more we live!"*~~

"There is a positive relationship between cigarette consumption and life expectancy at a country-by-country level!"

[Alberto Cairo. *How Charts Lie*, 2019]

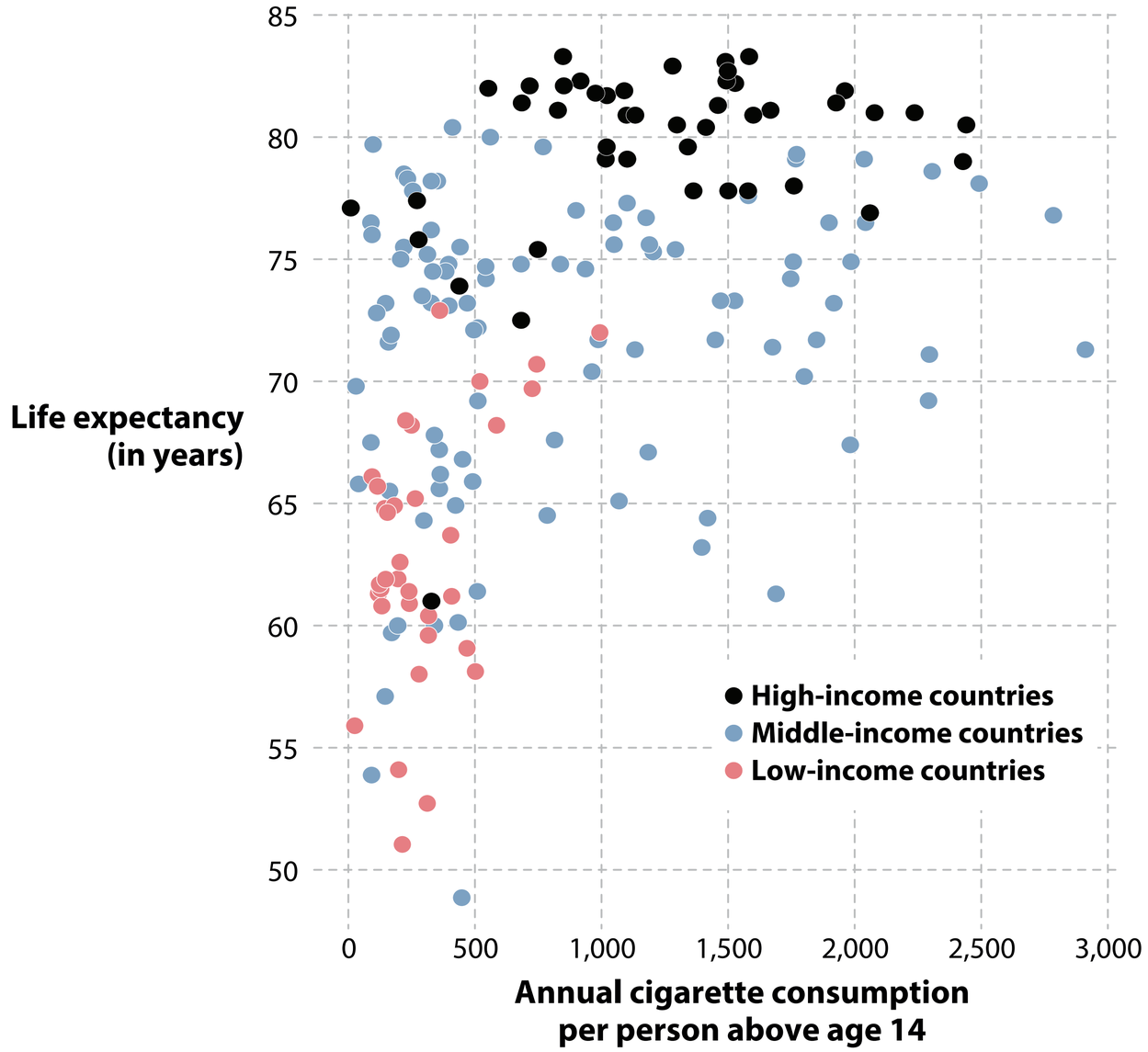


Describes a causal relationship.

~~*"The more cigarettes we consume, the more we live!"*~~

"There is a positive relationship between cigarette consumption and life expectancy at a country-by-country level!"

[Alberto Cairo. *How Charts Lie*, 2019]

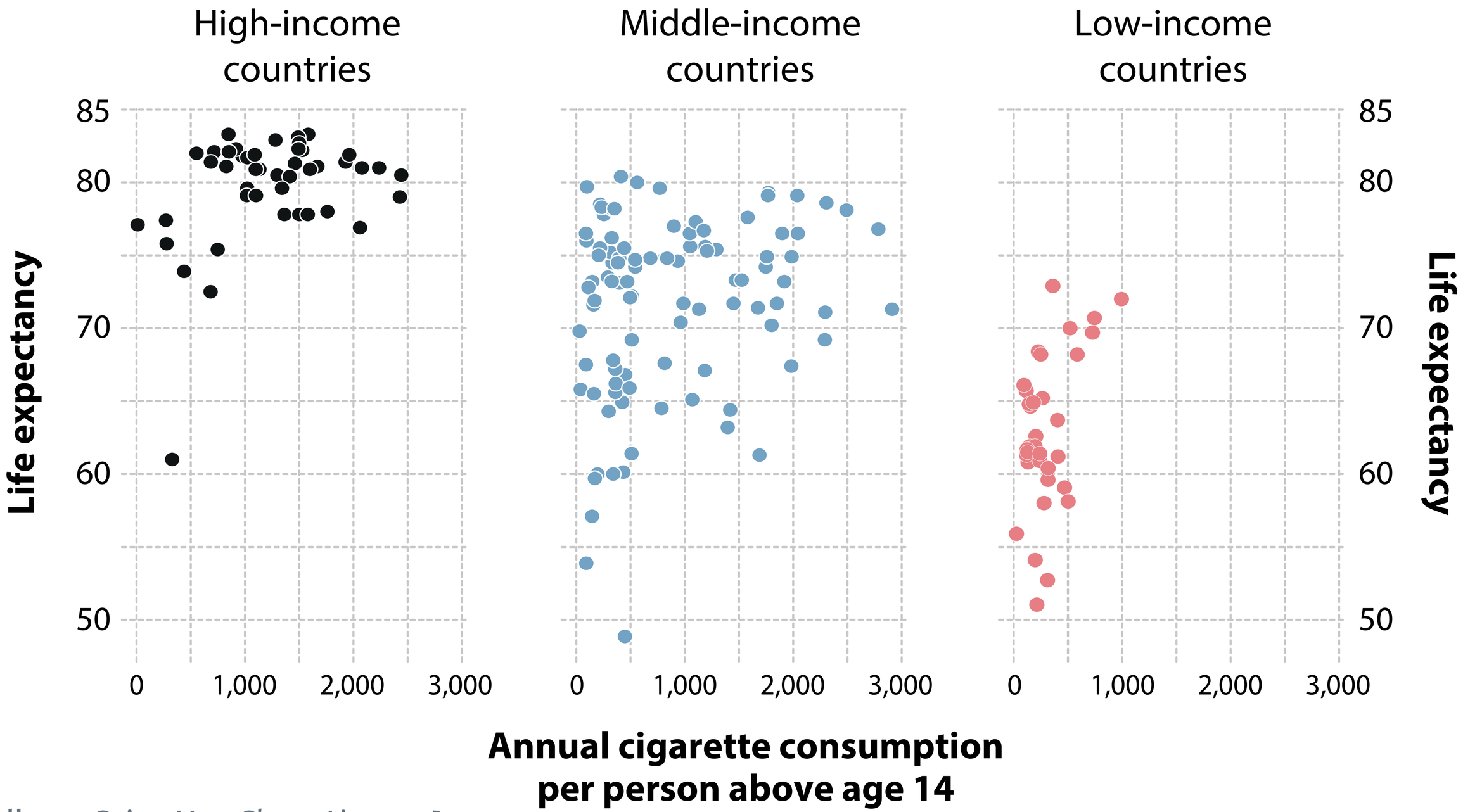


Describes a causal relationship.

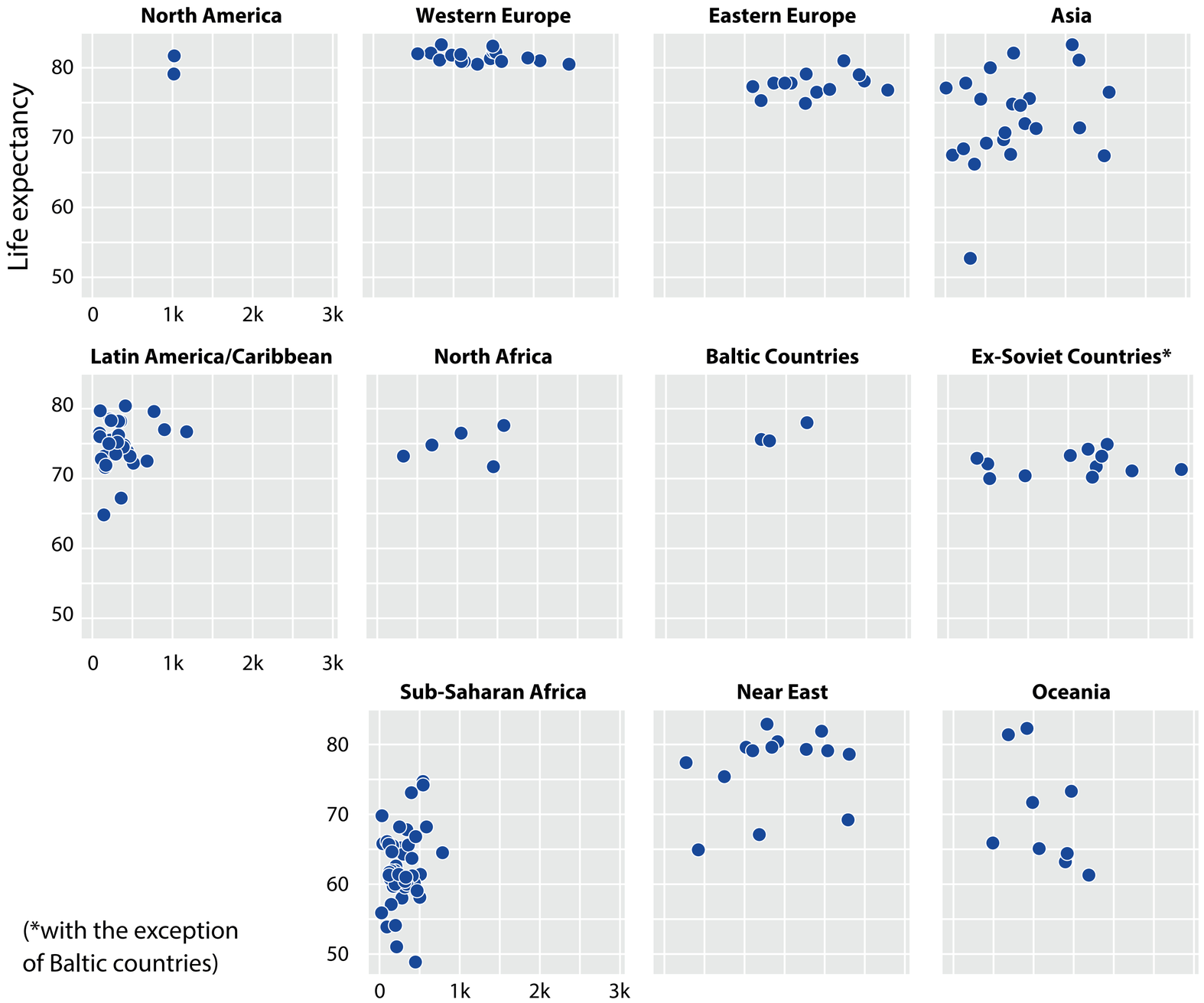
~~"The more cigarettes we consume, the more we live!"~~

"There is a positive relationship between cigarette consumption and life expectancy *at a country-by-country level*!"

[Alberto Cairo. *How Charts Lie*, 2019]



[Alberto Cairo. *How Charts Lie*, 2019]



[Alberto Cairo. *How Charts Lie*, 2019]

Simpson's Paradox

	Men		Women	
	Applicants	Admitted	Applicants	Admitted
Total	8442	**44%**	4321	35%

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	***825***	62%	108	**82%**
B	***560***	63%	25	**68%**
C	325	**37%**	***593***	34%
D	417	33%	375	**35%**
E	191	**28%**	***393***	24%
F	373	6%	341	**7%**

Table 1: Change in Median Wage by Education from 2000 to 2013

Segment	Change in Median Wage (%)
Overall	+0.9%
No degree	-7.9%
HS, no college	-4.7%
Some college	-7.6%
Bachelor's +	-1.2%

Economix

Explaining the Science of Everyday Life

Can Every Group Be Worse Than Average? Yes.

BY FLOYD NORRIS MAY 1, 2013 12:17 PM

Table 2: Number Employed (in millions) by Education: 2000, 2013

Segment	Employed 2000	Employed 2013	Change (%)
Overall	89.4	95.0	+6.4%
No degree	8.8	7.0	-21.3%
HS, no college	28.0	25.0	-10.6%
Some college	24.7	26.0	+5.4%
Bachelor's +	27.8	37.0	+33.0%

Titles & Labels

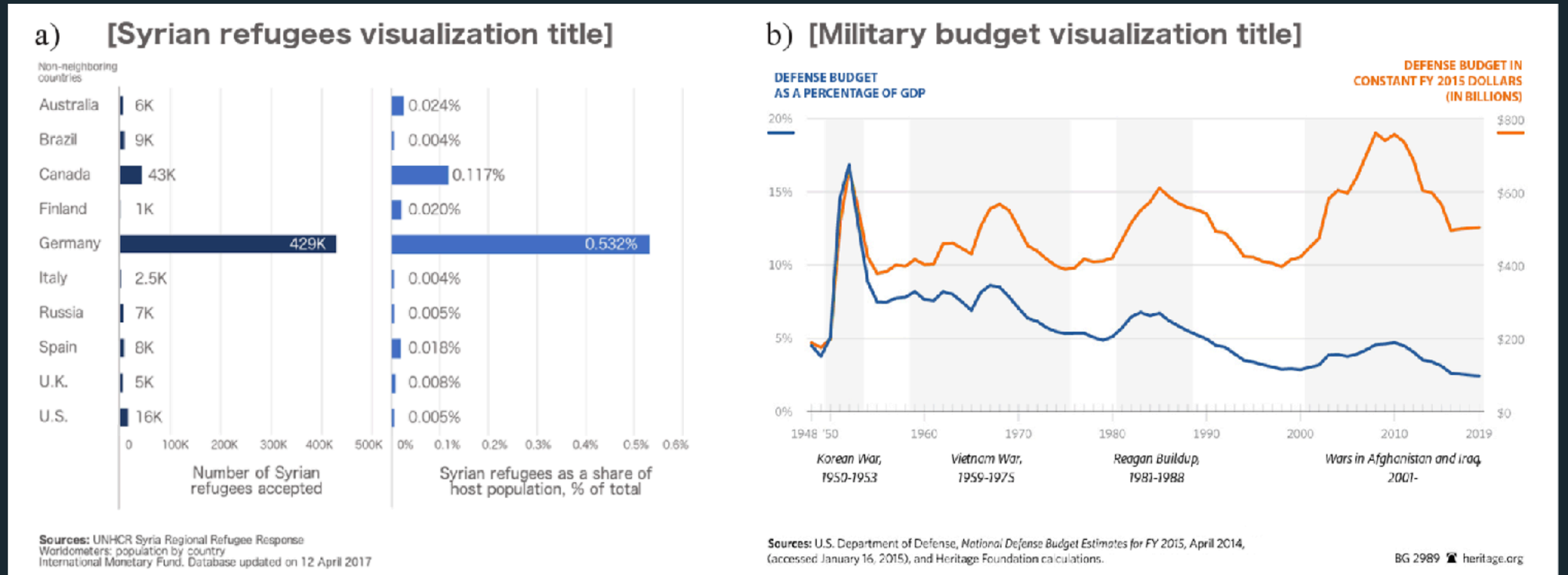


Figure 1: Two visualizations on popular foreign policies that show two variables supporting opposite sides of the policy.
a) shows the number of Syrian refugees accepted in each country and Syrian refugees as a share of host population
Supporting title: U.S. has accepted a lower percentage of Syrian refugees than the U.K., Spain, Finland, and Australia
Non-supporting title: U.S. has accepted more Syrian refugees than the U.K., Italy, Russia, and Finland combined
b) shows the U.S. defense budget in the last 50 years in constant fiscal year 2015 dollars and as a percentage of GDP
Supporting title: Defense budget on a steady decrease as a percentage of GDP over the past 50 years
Non-supporting title: Defense budget on an increase in constant dollars heading towards $500 billion by 2019

[Kong et al. ACM *CHI* 2018]

Titles & Labels

Generic	Issue-specific/ Subcategory	Example	Syrian refugees d	pro	con	n	Military budget d	pro	con	n	Total
Conflict	risk	Letting in potential killers?	3	1	27	7	0	0	0	1	40
	imperative	Keep them out	2	0	14	0	1	14	2	1	34
	priority	We can't feed our own	1	1	11	2	2	0	13	3	31
Economic	economic	U.S. can't afford to house refugees	0	2	5	0	9	21	27	12	76
	imperative	No more increase for war	0	0	0	0	0	2	6	1	9
Human interest	human	Syrians need homes too	5	17	2	7	0	1	1	1	34
	safety	Safety against terrorism	0	0	0	0	0	18	3	2	23
Morality	morality	We need to do our share	0	4	0	6	0	10	0	0	20
	US identity	We must remember lady liberty	0	7	1	1	0	0	0	0	9
Responsibility	responsibility	Other countries need to do more for Syrian refugees than US	0	2	11	3	0	1	5	1	23
Statistics	variable	Comparison of defense budget to GDP	6	2	1	7	7	2	0	6	31
	trend	Decrease in defense spending	1	37	24	6	8	40	34	14	164
	value	16K Syrian Refugees in the US	11	2	4	9	0	0	0	0	26
	balanced	Defense budget in billions of dollars and as a percentage of GDP	8	0	0	5	7	0	0	2	22
Open-ended	topic	Syrian refugee acceptance	73	4	4	42	74	4	6	49	256
	undecided	Syrian refugees, what to do	0	0	0	10	0	1	1	8	20

[Kong et al. ACM *CHI* 2018]

Titles & Labels

Experiment 2 (slant of titles)	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**
refugee-pro	0	0	1	2	30	9	5	2	2	0
refugee-con	0	1	4	5	28	9	3	1	0	0
budget-pro	1	1	3	13	32	12	3	1	0	0
budget-con	0	0	2	9	21	2	2	0	1	1

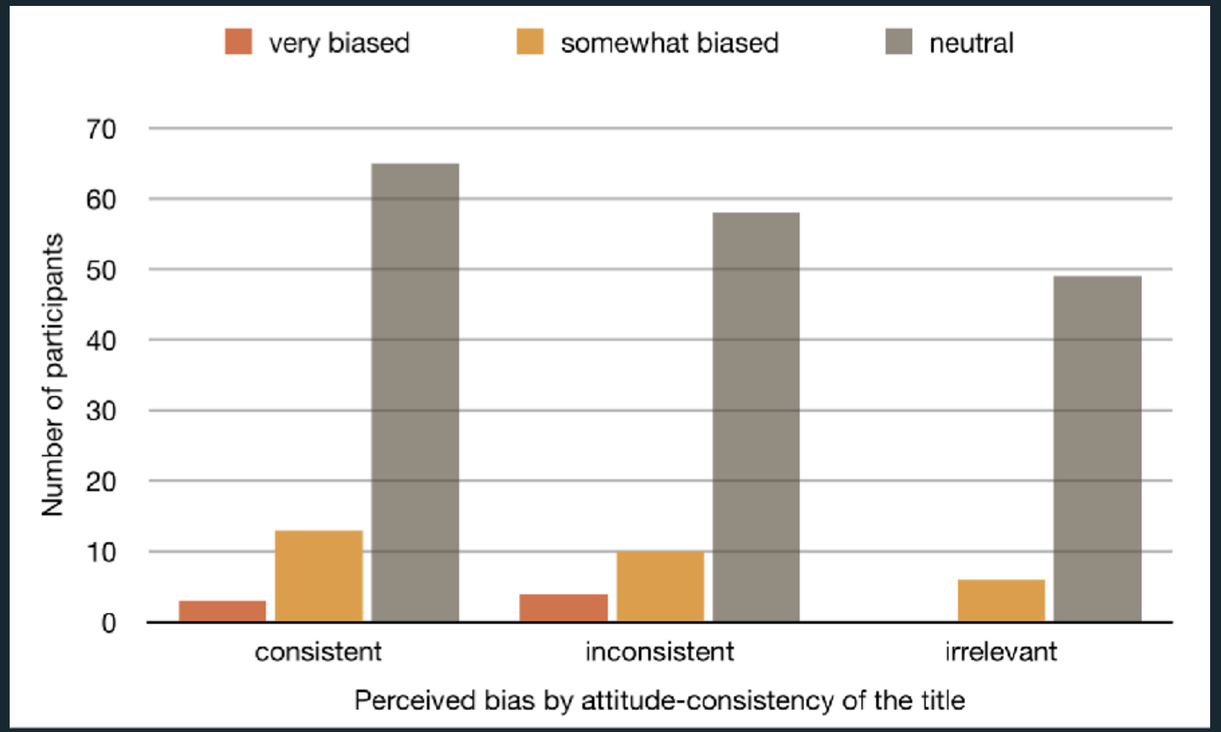
Table 3: The number of participants who indicated an attitude change for each topic and title slant. Attitude change is measured by the difference between the post-survey attitude and the pre-survey attitude. Each attitude is on a 7-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree." Expected directions of change are marked in pale blue.

[Kong et al. ACM *CHI* 2018]

Titles & Labels

Experiment 2 (slant of titles)	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**
refugee-pro	0	0	1	2	30	9	5	2	2	0
refugee-con	0	1	4	5	28	9	3	1	0	0
budget-pro	1	1	3	13	32	12	3	1	0	0
budget-con	0	0	2	9	21	2	2	0	1	1

Table 3: The number of participants who indicated an attitude change for each topic and title slant. Attitude change is measured by the difference between the post-survey attitude and the pre-survey attitude. Each attitude is on a 7-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree." Expected directions of change are marked in pale blue.



[Kong et al. ACM *CHI* 2018]